

Autocorrelation 2: Spatial Autocorrelation

Michael Noonan

March 7, 2021

Biol 520C: Statistical modelling for biological data



1. Housekeeping
2. Review
3. Spatial Autocorrelation
4. Detecting Spatial Autocorrelation
5. Correcting Spatial Autocorrelation

Housekeeping



- Nothing on my end.

Review

The IID Assumption



THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

Last lecture we saw how collecting data over time can result in temporally autocorrelated data, breaking the IID assumption.

Last lecture we saw how collecting data over time can result in temporally autocorrelated data, breaking the IID assumption.

We also saw how modifying the off-diagonals of the correlation matrix can correct for temporal autocorrelation.

Last lecture we saw how collecting data over time can result in temporally autocorrelated data, breaking the IID assumption.

We also saw how modifying the off-diagonals of the correlation matrix can correct for temporal autocorrelation.

$$V = \sigma^2 \underbrace{\begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}}_{\text{correlation matrix}}$$

Last lecture we saw how collecting data over time can result in temporally autocorrelated data, breaking the IID assumption.

We also saw how modifying the off-diagonals of the correlation matrix can correct for temporal autocorrelation.

$$V = \sigma^2 \underbrace{\begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}}_{\text{correlation matrix}}$$

Finally, we learned a few ways to do this in R using functions from the `nlme` package.

Clicker question: i



The following variance-covariance will result in errors with which kind of structure?

$$V = \underbrace{\begin{bmatrix} \theta + \sigma^2 & \theta & \cdots & \theta \\ \theta & \theta + \sigma^2 & \cdots & \theta \\ \vdots & \vdots & \ddots & \vdots \\ \theta & \theta & \cdots & \theta + \sigma^2 \end{bmatrix}}_{\text{variance--covariance matrix}}$$

A — AR-1.

B — ARMA(2,1).

C — Compound symmetric.

D — Independent and identically distributed.

Clicker question: i



The following variance-covariance will result in errors with which kind of structure?

$$V = \underbrace{\begin{bmatrix} \theta + \sigma^2 & \theta & \cdots & \theta \\ \theta & \theta + \sigma^2 & \cdots & \theta \\ \vdots & \vdots & \ddots & \vdots \\ \theta & \theta & \cdots & \theta + \sigma^2 \end{bmatrix}}_{\text{variance--covariance matrix}}$$

A — AR-1.

B — ARMA(2,1).

C — Compound symmetric.

D — Independent and identically distributed.

Clicker question: ii



The following variance-covariance can be produced with the following R function

$$V = \sigma^2 \underbrace{\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^i \\ \rho & 1 & \rho & \ddots & \vdots \\ \rho^2 & \rho & 1 & \ddots & \rho^2 \\ \vdots & \ddots & \ddots & \ddots & \rho \\ \rho^1 & \cdots & \rho^2 & \rho & 1 \end{bmatrix}}_{\text{correlation matrix}}$$

A — corARMA()

C — corCompSymm()

B — corAR1()

D — Both A and B.

Clicker question: ii



The following variance-covariance can be produced with the following R function

$$V = \sigma^2 \underbrace{\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^i \\ \rho & 1 & \rho & \ddots & \vdots \\ \rho^2 & \rho & 1 & \ddots & \rho^2 \\ \vdots & \ddots & \ddots & \ddots & \rho \\ \rho^1 & \cdots & \rho^2 & \rho & 1 \end{bmatrix}}_{\text{correlation matrix}}$$

A — corARMA()

C — corCompSymm()

B — corAR1()

D — Both A and B.

Impact of autocorrelation



THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus



We touched on how, with autocorrelated data, each new datapoint does not bring a full independent datapoint worth of information.

We touched on how, with autocorrelated data, each new datapoint does not bring a full independent datapoint worth of information.

When data are autocorrelated $n_{\text{effective}} < n$, meaning SEs and CIs shrink faster than they should.

$$\text{SE} = \frac{\sigma}{\sqrt{n}}$$

$$95\% \text{ CI} = \bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

We touched on how, with autocorrelated data, each new datapoint does not bring a full independent datapoint worth of information.

When data are autocorrelated $n_{\text{effective}} < n$, meaning SEs and CIs shrink faster than they should.

$$\text{SE} = \frac{\sigma}{\sqrt{n}}$$

$$95\% \text{ CI} = \bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

This results in over-confidence in our estimates.

We touched on how, with autocorrelated data, each new datapoint does not bring a full independent datapoint worth of information.

When data are autocorrelated $n_{\text{effective}} < n$, meaning SEs and CIs shrink faster than they should.

$$\text{SE} = \frac{\sigma}{\sqrt{n}}$$

$$95\% \text{ CI} = \bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

This results in over-confidence in our estimates.

Effect is usually strongest on SEs and CIs, but autocorrelation can also impact the accuracy of parameter estimates.

Sources of autocorrelation



THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

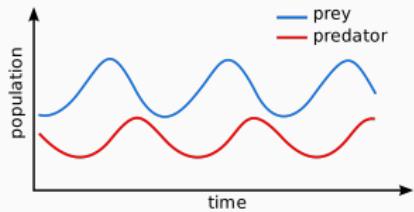


We also covered the idea that anything that causes some data points to be more similar to each other than others can result in autocorrelation.

We also covered the idea that anything that causes some data points to be more similar to each other than others can result in autocorrelation.

This included:

- **Time:** Data that are close together in time are more related.



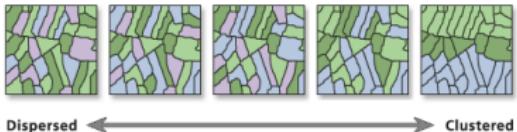
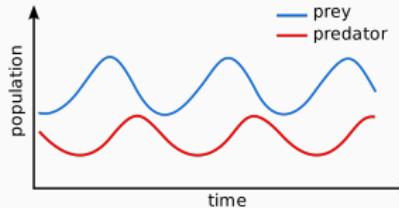
Sources of autocorrelation



We also covered the idea that anything that causes some data points to be more similar to each other than others can result in autocorrelation.

This included:

- **Time:** Data that are close together in time are more related.
- **Space:** Data that are close together in space are more related.



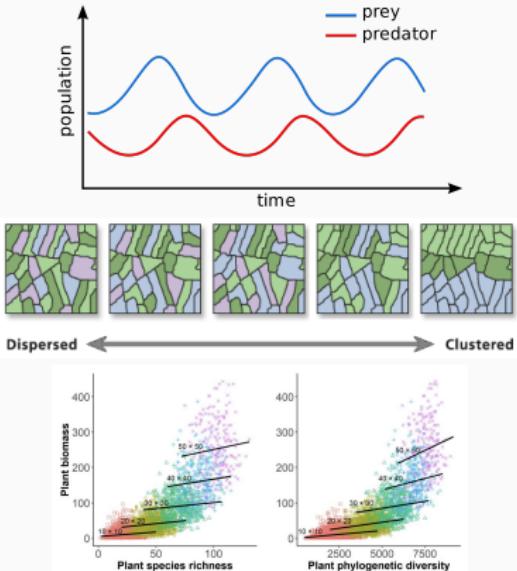
Sources of autocorrelation



We also covered the idea that anything that causes some data points to be more similar to each other than others can result in autocorrelation.

This included:

- **Time:** Data that are close together in time are more related.
- **Space:** Data that are close together in space are more related.
- **Phylogeny:** Species that are closer together on an evolutionary timescale are more related.



(Liang et al., 2019)

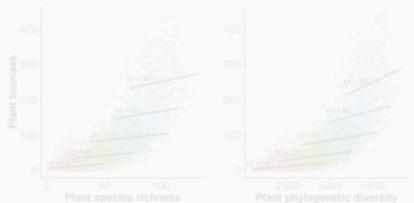
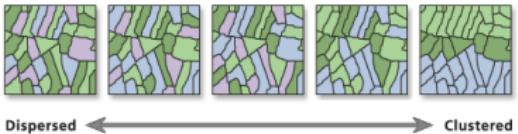
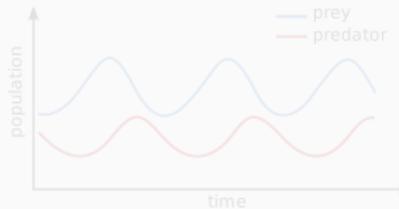
Sources of autocorrelation



We also covered the idea that anything that causes some data points to be more similar to each other than others can result in autocorrelation.

This included:

- **Time:** Data that are close together in time are more related.
- **Space:** Data that are close together in space are more related.
- **Phylogeny:** Species that are closer together on an evolutionary timescale are more related.



(Liang et al., 2019)

Spatial Autocorrelation

Spatial Autocorrelation



THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus



Biological data are often collected by measuring quantities over space (e.g., abundances, growth rates, species occurrences, etc.).

Biological data are often collected by measuring quantities over space (e.g., abundances, growth rates, species occurrences, etc.).

When this is the case, spatial autocorrelation can arise when the variation between the values of the datapoints is affected by their spatial distance.

Biological data are often collected by measuring quantities over space (e.g., abundances, growth rates, species occurrences, etc.).

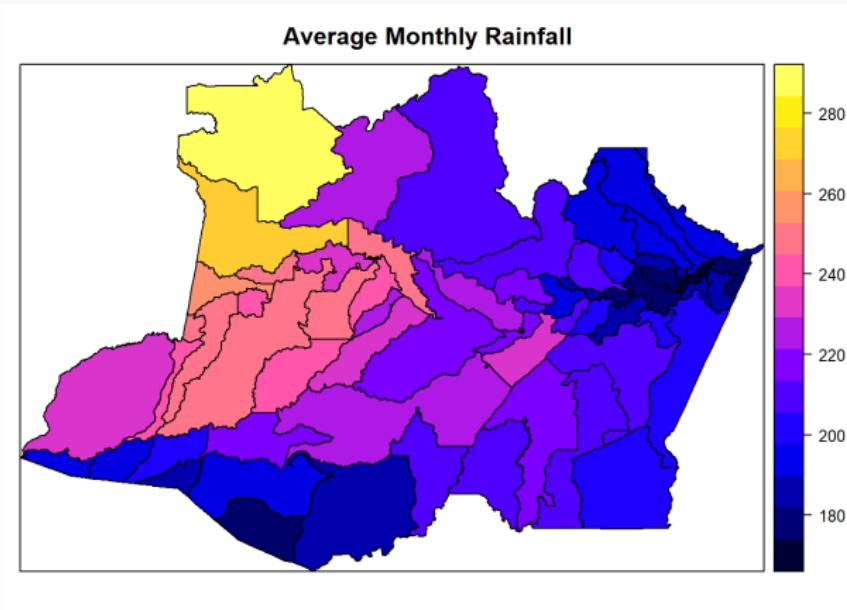
When this is the case, spatial autocorrelation can arise when the variation between the values of the datapoints is affected by their spatial distance.

The underlying reason for this is that many of the drivers of biological patterns such as environmental conditions, topography, ecosystem structure/composition act at large spatial scales, making data that are spatially close more similar than data collected further apart.

Spatial Autocorrelation cont.



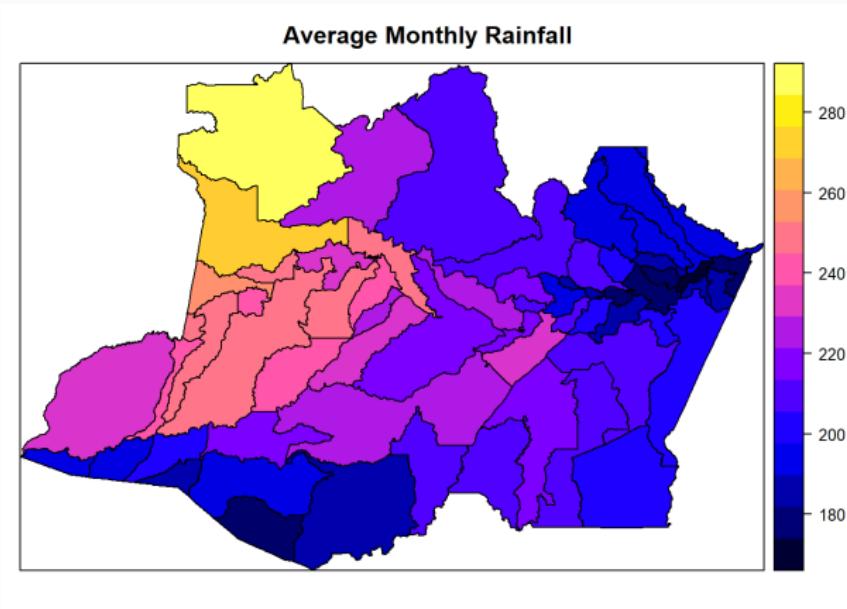
What if we're studying the effect of rainfall on spp. div. in the Amazon?



Spatial Autocorrelation cont.



What if we're studying the effect of rainfall on spp. div. in the Amazon?



Because rainfall is correlated in space, species diversity will also be correlated in space (if the relationship exists).

Detecting Spatial Autocorrelation

Bubble plots



THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

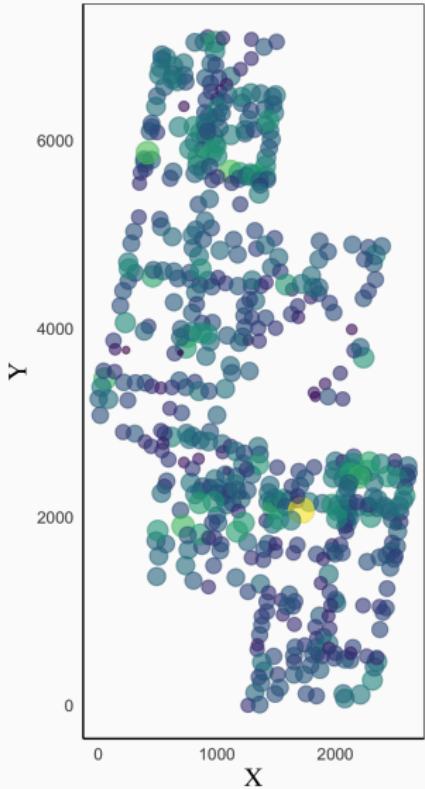
As with temporal autocorrelation, spatial autocorrelation can be difficult to see in a simple residuals vs. fitted plot (again, not designed for this purpose).

Bubble plots



As with temporal autocorrelation, spatial autocorrelation can be difficult to see in a simple residuals vs. fitted plot (again, not designed for this purpose).

'Bubble plots' are an easy tool to quickly assess the residuals for autocorrelation.



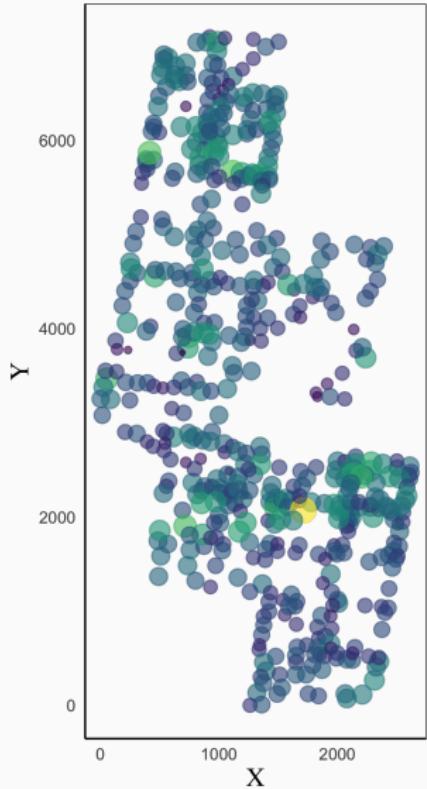
Bubble plots



As with temporal autocorrelation, spatial autocorrelation can be difficult to see in a simple residuals vs. fitted plot (again, not designed for this purpose).

'Bubble plots' are an easy tool to quickly assess the residuals for autocorrelation.

Residuals are plotted in space, and sizes/colours are proportional to their values.



Bubble plots

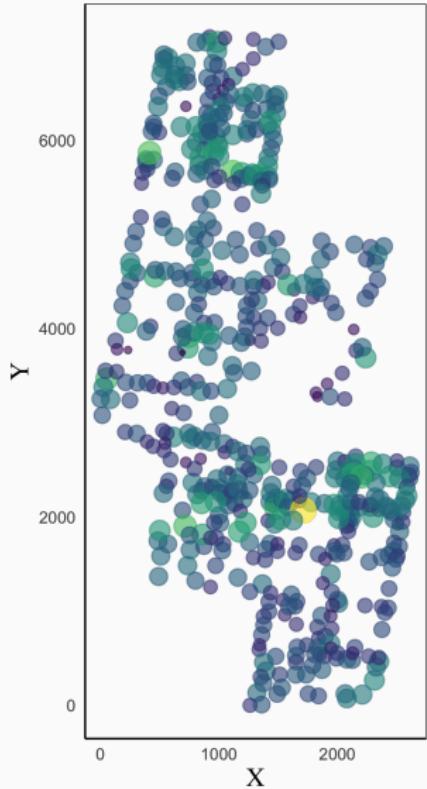


As with temporal autocorrelation, spatial autocorrelation can be difficult to see in a simple residuals vs. fitted plot (again, not designed for this purpose).

'Bubble plots' are an easy tool to quickly assess the residuals for autocorrelation.

Residuals are plotted in space, and sizes/colours are proportional to their values.

Idea is to look for patterns (absence of a pattern is good).



Bubble plots



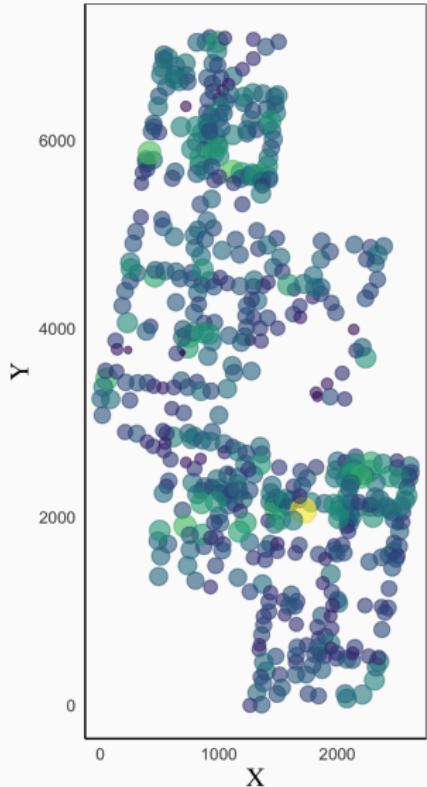
As with temporal autocorrelation, spatial autocorrelation can be difficult to see in a simple residuals vs. fitted plot (again, not designed for this purpose).

'Bubble plots' are an easy tool to quickly assess the residuals for autocorrelation.

Residuals are plotted in space, and sizes/colours are proportional to their values.

Idea is to look for patterns (absence of a pattern is good).

Bubble plots are quick and easy tools, but can be hard to read and not very formal.





Moran's I is a correlation coefficient that measures the overall spatial autocorrelation of a data set (think of it as \sim weighted covariance):

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

Moran's I is a correlation coefficient that measures the overall spatial autocorrelation of a data set (think of it as \sim weighted covariance):

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

N is the number of spatial units indexed by i and j ;

Moran's I is a correlation coefficient that measures the overall spatial autocorrelation of a data set (think of it as \sim weighted covariance):

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

N is the number of spatial units indexed by i and j ;
 x is the variable of interest and \bar{x} is the mean of x ;

Moran's I is a correlation coefficient that measures the overall spatial autocorrelation of a data set (think of it as \sim weighted covariance):

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

N is the number of spatial units indexed by i and j ;

x is the variable of interest and \bar{x} is the mean of x ;

w_{ij} is a matrix of spatial weights and W is the sum of all w_{ij} .

Moran's I is a correlation coefficient that measures the overall spatial autocorrelation of a data set (think of it as \sim weighted covariance):

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

N is the number of spatial units indexed by i and j ;

x is the variable of interest and \bar{x} is the mean of x ;

w_{ij} is a matrix of spatial weights and W is the sum of all w_{ij} .

Values of I usually range from -1 to +1.







Many R packages for calculating Moran's I

Many R packages for calculating Moran's I

```
library(ape)
library(fields)

#Vector of spatial coordinates
coords = cbind(data$x, data$y)

#Matrix of distances for the weights
w = fields::rdist(coords)

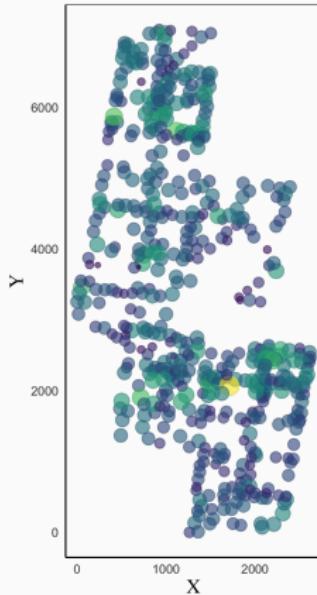
#Calculate Moran's I
ape::Moran.I(data$Bor, w = w)

$observed
[1] -0.03019649

$expected
[1] -0.001879699

$sd
[1] 0.001368412

$p.value
[1] 3.991055e-95
```



Many R packages for calculating Moran's I

```
library(ape)
library(fields)

#Vector of spatial coordinates
coords = cbind(data$x, data$y)

#Matrix of distances for the weights
w = fields::rdist(coords)

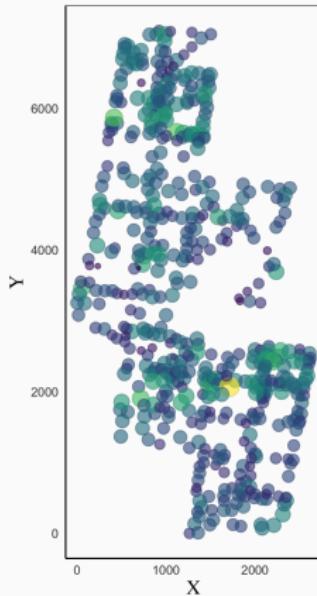
#Calculate Moran's I
ape::Moran.I(data$Bor, w = w)

$observed
[1] -0.03019649

$expected
[1] -0.001879699

$sd
[1] 0.001368412

$p.value
[1] 3.991055e-95
```



The *p*-value tells us we have significant spatial autocorrelation

Moran's I cont.



THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

Moran's I can be a useful tool for identifying the presence of autocorrelation and is quite popular.

Moran's I can be a useful tool for identifying the presence of autocorrelation and is quite popular.

The challenge is how to act on this information (i.e., lets you know if you have a problem, but doesn't help in finding a solution)?

Moran's I can be a useful tool for identifying the presence of autocorrelation and is quite popular.

The challenge is how to act on this information (i.e., lets you know if you have a problem, but doesn't help in finding a solution)?

Also very sensitive to how you define the weights:

"The idea is to construct a matrix that accurately reflects your assumptions about the particular spatial phenomenon in question. A common approach is to give a weight of 1 if two zones are neighbors, and 0 otherwise, though the definition of 'neighbors' can vary. Another common approach might be to give a weight of 1 to k nearest neighbors, 0 otherwise. An alternative is to use a distance decay function for assigning weights. Sometimes the length of a shared edge is used for assigning different weights to neighbors. The selection of spatial weights matrix should be guided by theory about the phenomenon in question." – Wikipedia

Semi-variograms



THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

Semi-variograms (or just variograms) are spatial data's equivalent of the ACF facilitate visual assessment of spatial autocorrelations in the data.

Semi-variograms (or just variograms) are spatial data's equivalent of the ACF facilitate visual assessment of spatial autocorrelations in the data.

Semi-variance is a measure of the degree of similarity between pairs of points separated by a specific distance h .

Semi-variograms (or just variograms) are spatial data's equivalent of the ACF facilitate visual assessment of spatial autocorrelations in the data.

Semi-variance is a measure of the degree of similarity between pairs of points separated by a specific distance h . Plot of semi-variance vs. separation distance is called a variogram.

Semi-variograms (or just variograms) are spatial data's equivalent of the ACF facilitate visual assessment of spatial autocorrelations in the data.

Semi-variance is a measure of the degree of similarity between pairs of points separated by a specific distance h . Plot of semi-variance vs. separation distance is called a variogram.

For residuals separated by distance h :

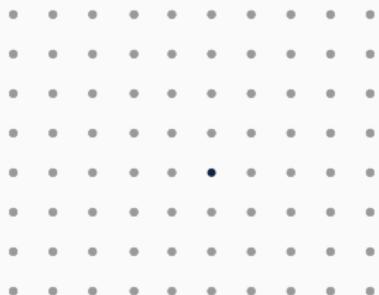
$$\hat{\gamma}(s) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i + h) - z(x_i)]^2$$

Semi-variograms (or just variograms) are spatial data's equivalent of the ACF facilitate visual assessment of spatial autocorrelations in the data.

Semi-variance is a measure of the degree of similarity between pairs of points separated by a specific distance h . Plot of semi-variance vs. separation distance is called a variogram.

For residuals separated by distance h :

$$\hat{\gamma}(s) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i + h) - z(x_i)]^2$$

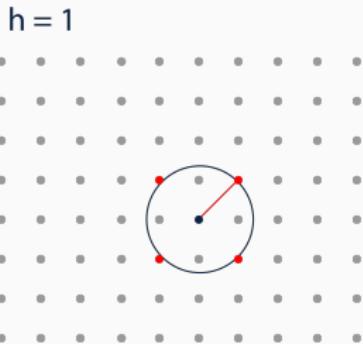


Semi-variograms (or just variograms) are spatial data's equivalent of the ACF facilitate visual assessment of spatial autocorrelations in the data.

Semi-variance is a measure of the degree of similarity between pairs of points separated by a specific distance h . Plot of semi-variance vs. separation distance is called a variogram.

For residuals separated by distance h :

$$\hat{\gamma}(s) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i + h) - z(x_i)]^2$$

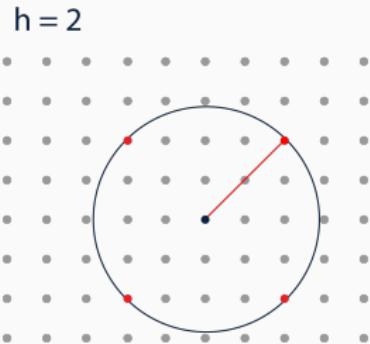


Semi-variograms (or just variograms) are spatial data's equivalent of the ACF facilitate visual assessment of spatial autocorrelations in the data.

Semi-variance is a measure of the degree of similarity between pairs of points separated by a specific distance h . Plot of semi-variance vs. separation distance is called a variogram.

For residuals separated by distance h :

$$\hat{\gamma}(s) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i + h) - z(x_i)]^2$$

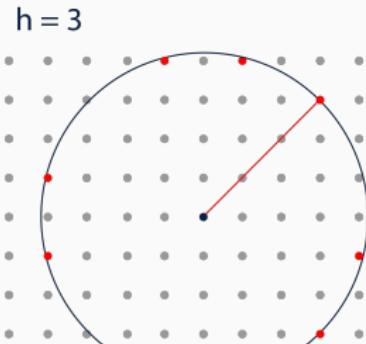


Semi-variograms (or just variograms) are spatial data's equivalent of the ACF facilitate visual assessment of spatial autocorrelations in the data.

Semi-variance is a measure of the degree of similarity between pairs of points separated by a specific distance h . Plot of semi-variance vs. separation distance is called a variogram.

For residuals separated by distance h :

$$\hat{\gamma}(s) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i + h) - z(x_i)]^2$$

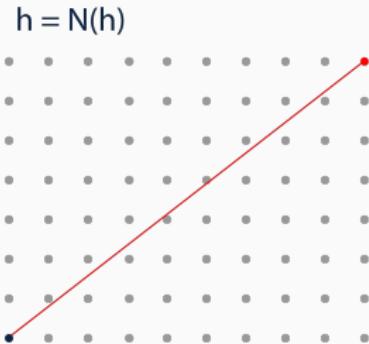


Semi-variograms (or just variograms) are spatial data's equivalent of the ACF facilitate visual assessment of spatial autocorrelations in the data.

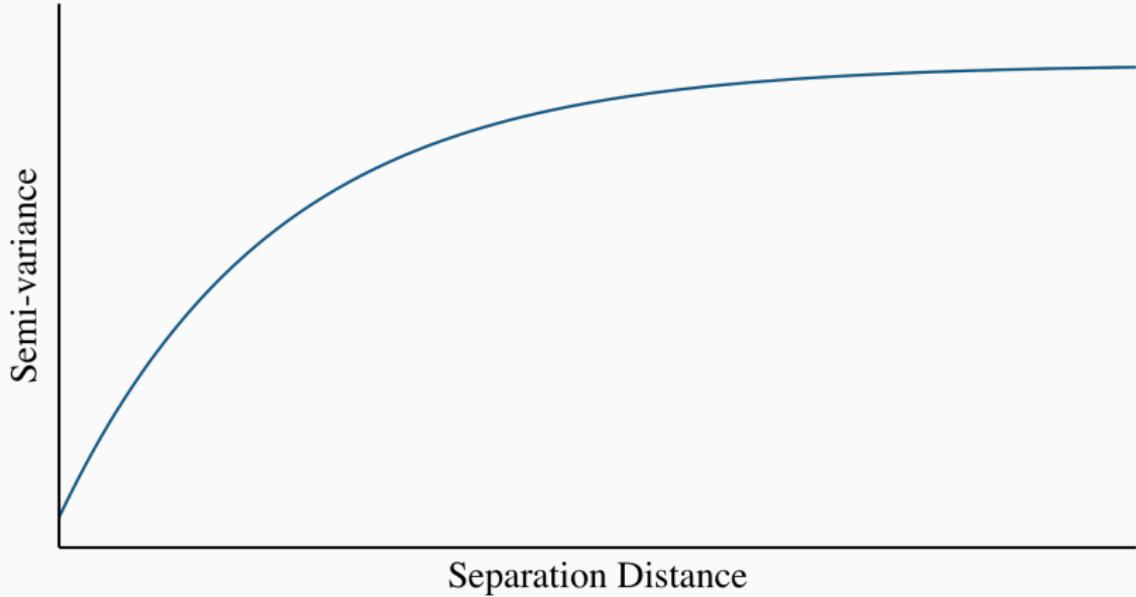
Semi-variance is a measure of the degree of similarity between pairs of points separated by a specific distance h . Plot of semi-variance vs. separation distance is called a variogram.

For residuals separated by distance h :

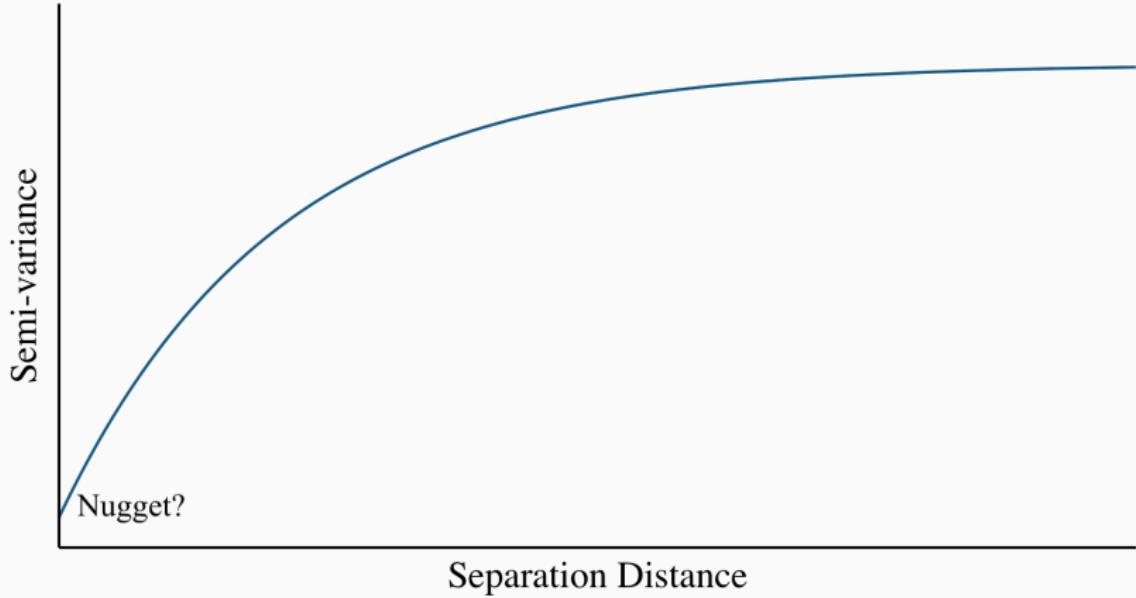
$$\hat{\gamma}(s) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i + h) - z(x_i)]^2$$



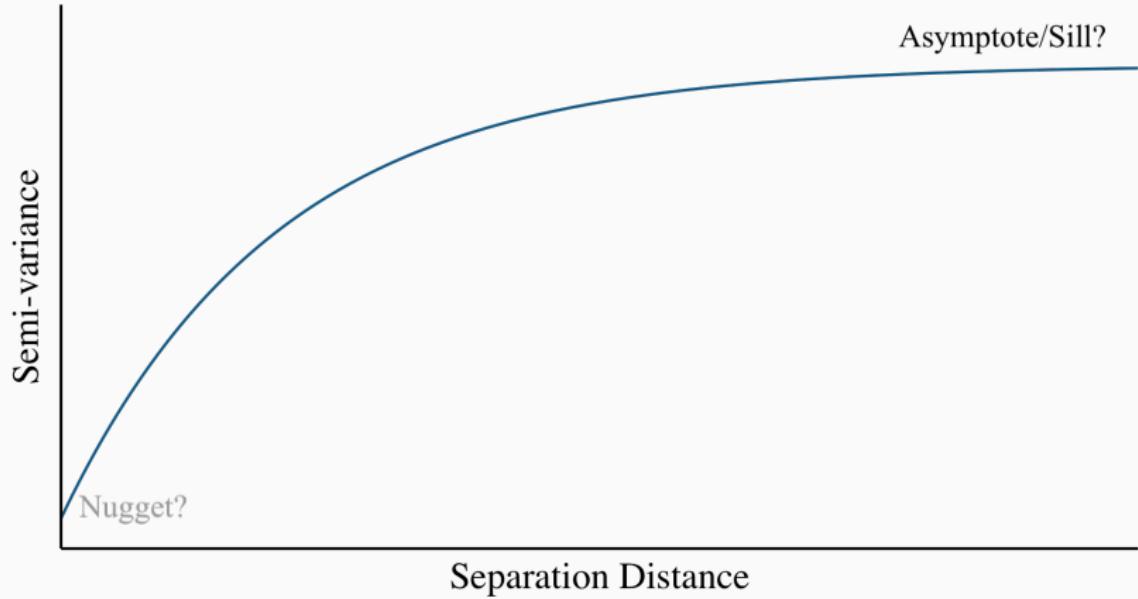
Reading a variogram



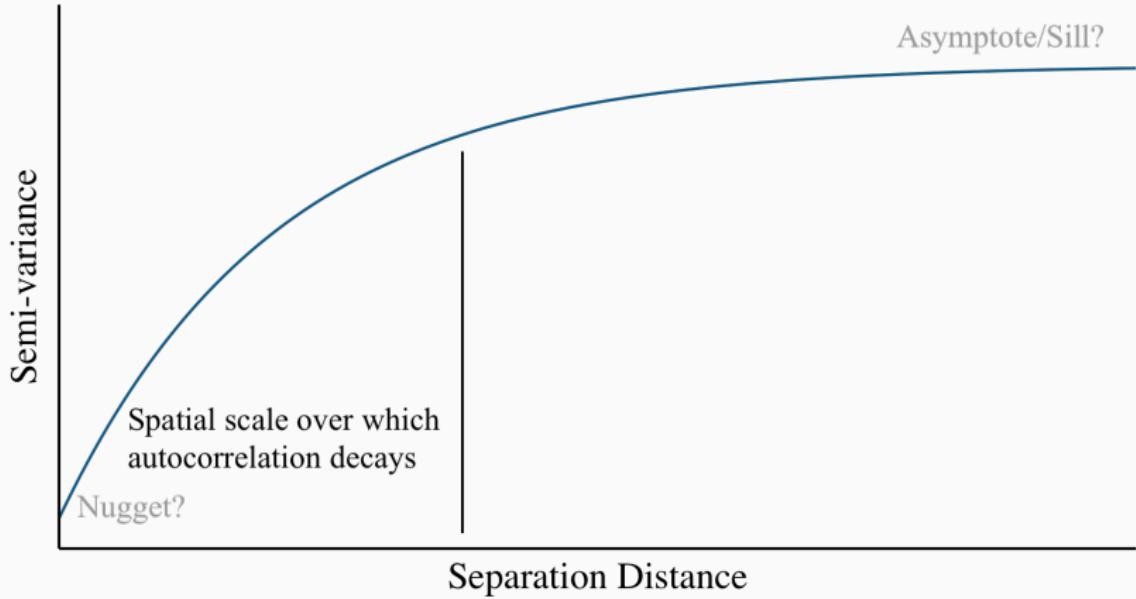
Reading a variogram



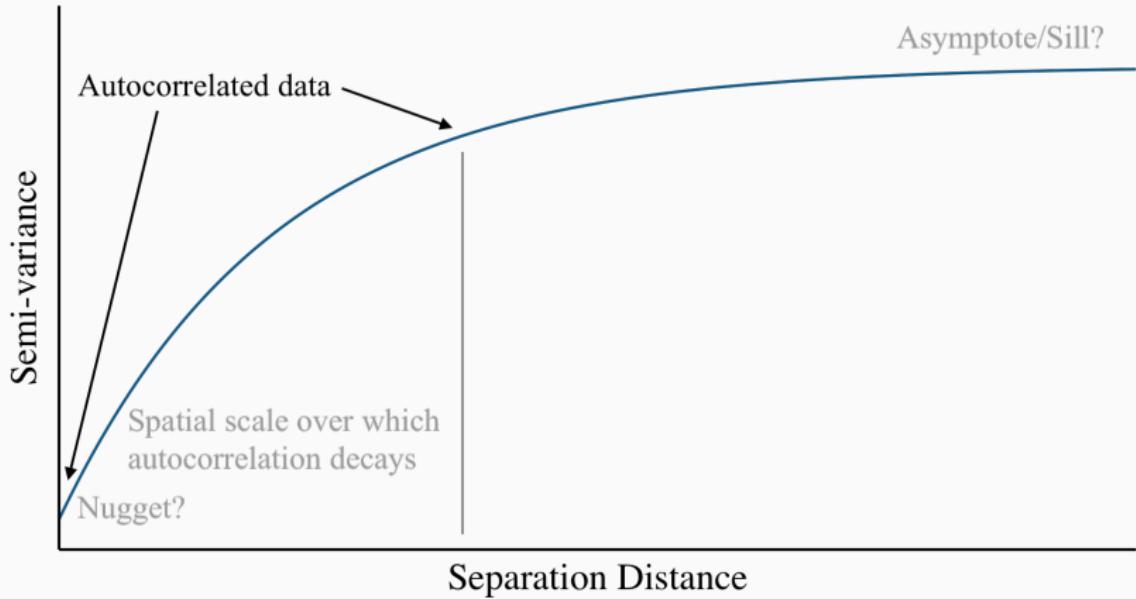
Reading a variogram



Reading a variogram



Reading a variogram



Last lecture we saw how different autocorrelation models could be used to correct for different temporal autocorrelation structures.

Last lecture we saw how different autocorrelation models could be used to correct for different temporal autocorrelation structures.

The same applies with spatial autocorrelation.

Last lecture we saw how different autocorrelation models could be used to correct for different temporal autocorrelation structures.

The same applies with spatial autocorrelation.

Usefully, the different spatial correlation models all have differently shaped theoretical variograms.

Last lecture we saw how different autocorrelation models could be used to correct for different temporal autocorrelation structures.

The same applies with spatial autocorrelation.

Usefully, the different spatial correlation models all have differently shaped theoretical variograms.

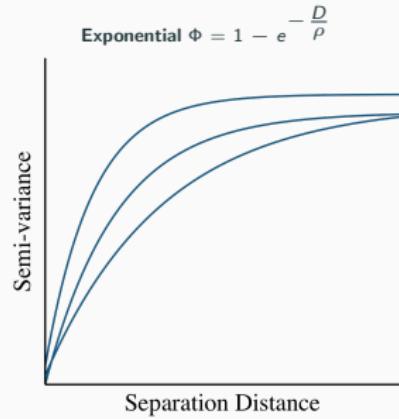
In other words the shape of a dataset's empirical variogram can provide clues on which spatial correlation model is most appropriate.

Correlation models and their variograms

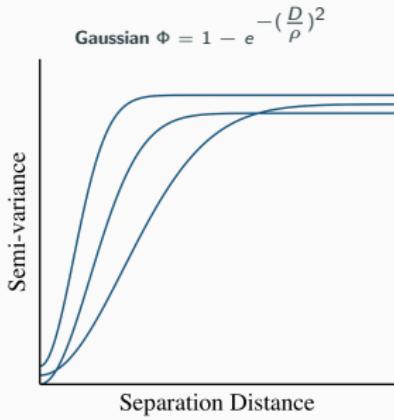
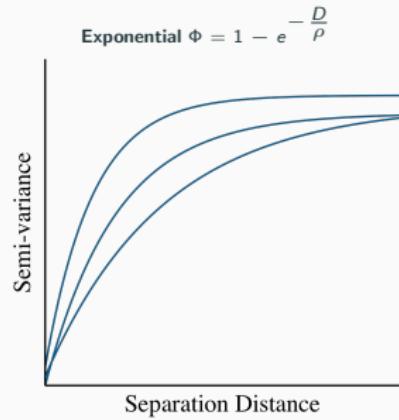


THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

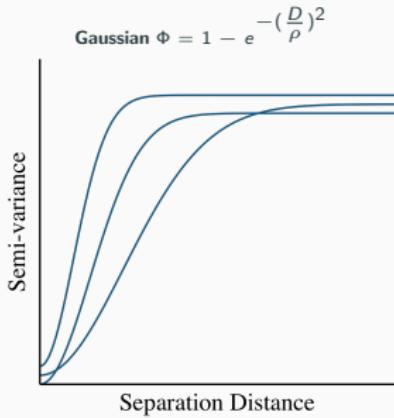
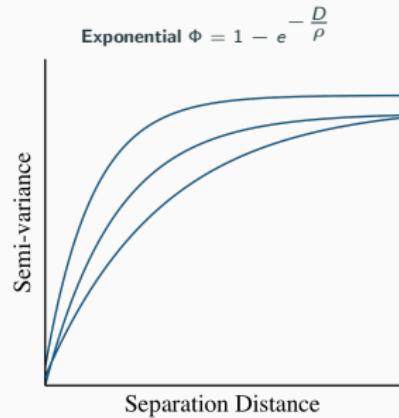
Correlation models and their variograms



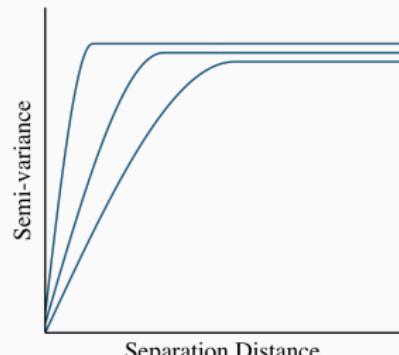
Correlation models and their variograms



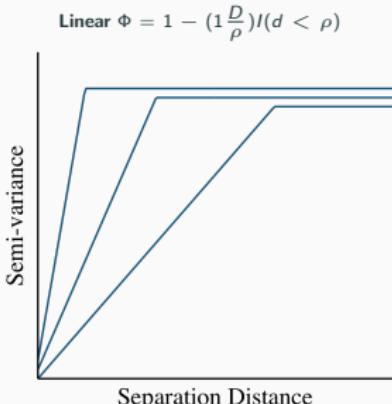
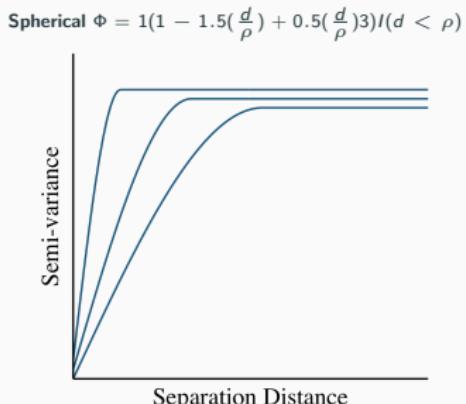
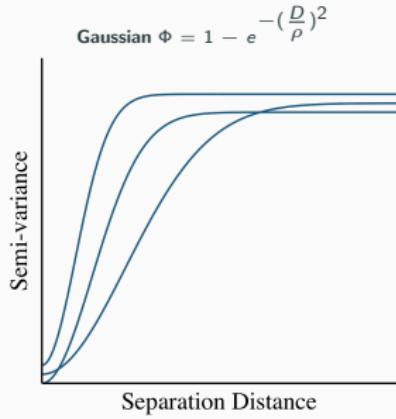
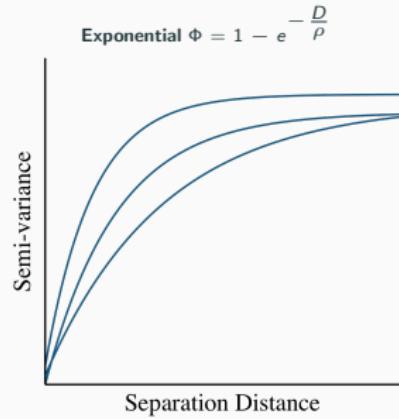
Correlation models and their variograms



Spherical $\Phi = 1(1 - 1.5(\frac{d}{\rho}) + 0.5(\frac{d}{\rho})^3)I(d < \rho)$



Correlation models and their variograms



Correcting Spatial Autocorrelation

Correcting Spatial Autocorrelation



THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus



So you find yourself with spatially autocorrelated data. What next?

Correcting Spatial Autocorrelation



THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus



Here again, 'Dealing with spatial autocorrelation' and 'analysing spatial trends' are not the same thing.



Here again, 'Dealing with spatial autocorrelation' and 'analysing spatial trends' are not the same thing. Our focus is on the former.

Here again, 'Dealing with spatial autocorrelation' and 'analysing spatial trends' are not the same thing. Our focus is on the former.

Methods for handling spatial autocorrelation in a regressional framework come primarily from the field of geostatistics.

Here again, 'Dealing with spatial autocorrelation' and 'analysing spatial trends' are not the same thing. Our focus is on the former.

Methods for handling spatial autocorrelation in a regressional framework come primarily from the field of geostatistics.

In practice, they function much like corrections for temporal autocorrelation (i.e., a modification of the model's variance-covariance matrix).

$$V = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}$$

The Russian Boreal Data

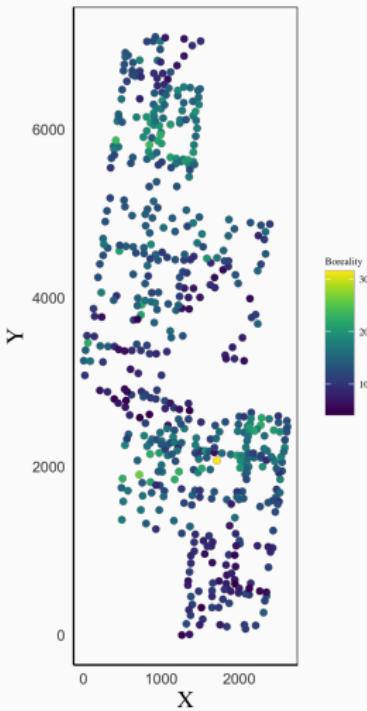


THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

The Russian Boreal Data



We're going to work with a dataset from Zuur *et al.* (2007) to examine the influence of biogeography on forest composition in Tatarstan, Russia.

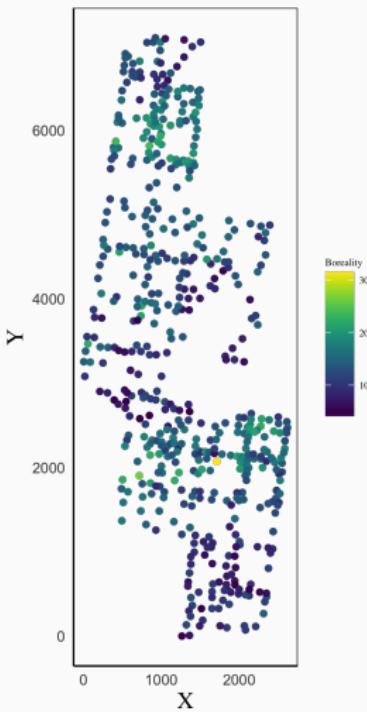


The Russian Boreal Data



We're going to work with a dataset from Zuur *et al.* (2007) to examine the influence of biogeography on forest composition in Tatarstan, Russia.

The response variable is a measure of boreality (i.e., percent boreal species at a site) and our explanatory variable is a measure of wetness.



The Russian Boreal Data

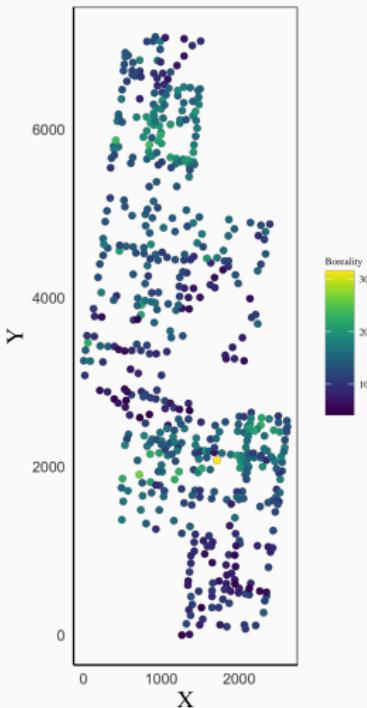


We're going to work with a dataset from Zuur *et al.* (2007) to examine the influence of biogeography on forest composition in Tatarstan, Russia.

The response variable is a measure of boreality (i.e., percent boreal species at a site) and our explanatory variable is a measure of wetness.

Today's starting point is the linear regression model:

$$\text{Boreality}_i = \beta_0 + \beta_1 \text{Wetness}_i + \varepsilon_i$$



Autocorr. in the Russian Boreality data



THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

Autocorr. in the Russian Boreality data



THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

```
library(nlme)
data <- read.csv("Boreality.csv")
data$Bor <- sqrt(1000 * (data$nBor + 1)/(data$nTot))
data$Wet <- data$Wet - mean(data$Wet)

FIT <- gls(Bor ~ Wet, data = data)

Generalized least squares fit by REML
  Model: Bor ~ Wet
  Data: data
      AIC      BIC      logLik
 2844.541 2857.365 -1419.271

Coefficients:
            Value Std.Error t-value p-value
(Intercept) 18.4880   0.378719 48.81715     0
Wet         165.8036 10.599111 15.64316     0

Correlation:
  (Intr)
Wet  0.917

Residual standard error: 3.490577
Degrees of freedom: 533 total; 531 residual
```

Autocorr. in the Russian Boreality data



```
library(nlme)
data <- read.csv("Boreality.csv")
data$Bor <- sqrt(1000 * (data$nBor + 1)/(data$nTot))
data$Wet <- data$Wet - mean(data$Wet)

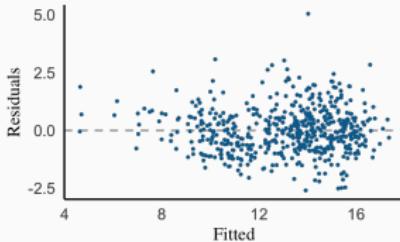
FIT <- gls(Bor ~ Wet, data = data)
```

```
Generalized least squares fit by REML
  Model: Bor ~ Wet
  Data: data
      AIC      BIC      logLik
 2844.541 2857.365 -1419.271
```

```
Coefficients:
            Value Std.Error t-value p-value
(Intercept) 18.4880   0.378719 48.81715     0
Wet         165.8036 10.599111 15.64316     0
```

```
Correlation:
  (Intr)
Wet 0.917
```

```
Residual standard error: 3.490577
Degrees of freedom: 533 total; 531 residual
```



Autocorr. in the Russian Boreality data



```
library(nlme)
data <- read.csv("Boreality.csv")
data$Bor <- sqrt(1000 * (data$nBor + 1)/(data$nTot))
data$Wet <- data$Wet - mean(data$Wet)

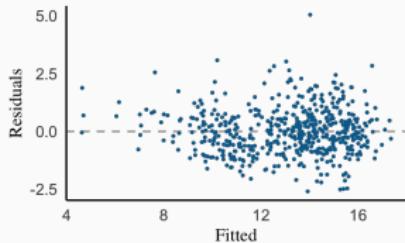
FIT <- gls(Bor ~ Wet, data = data)
```

```
Generalized least squares fit by REML
  Model: Bor ~ Wet
  Data: data
      AIC      BIC      logLik
 2844.541 2857.365 -1419.271
```

```
Coefficients:
            Value Std.Error t-value p-value
(Intercept) 18.4880   0.378719 48.81715     0
Wet         165.8036 10.599111 15.64316     0
```

```
Correlation:
  (Intr)
Wet 0.917
```

```
Residual standard error: 3.490577
Degrees of freedom: 533 total; 531 residual
```



Residuals actually look ok.

Autocorr. in the Russian Boreality data



```
library(nlme)
data <- read.csv("Boreality.csv")
data$Bor <- sqrt(1000 * (data$nBor + 1)/(data$nTot))
data$Wet <- data$Wet - mean(data$Wet)

FIT <- gls(Bor ~ Wet, data = data)
```

Generalized least squares fit by REML

Model: Bor ~ Wet

Data: data

AIC	BIC	logLik
2844.541	2857.365	-1419.271

Coefficients:

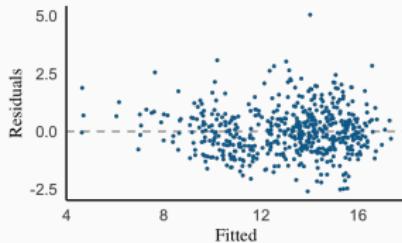
	Value	Std.Error	t-value	p-value
(Intercept)	18.4880	0.378719	48.81715	0
Wet	165.8036	10.599111	15.64316	0

Correlation:

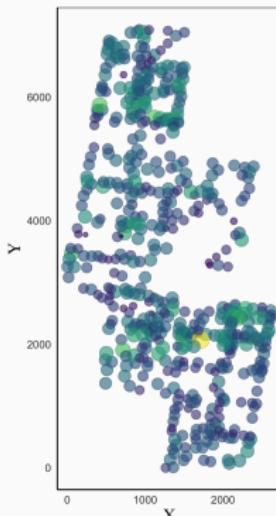
(Intr)
Wet 0.917

Residual standard error: 3.490577

Degrees of freedom: 533 total; 531 residual



Residuals actually look ok.



Autocorr. in the Russian Boreality data



THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus



```
library(ape)
library(fields)
library(gstat)
library(sp)
```

Autocorr. in the Russian Boreality data



```
library(ape)
library(fields)
library(gstat)
library(sp)

#Spatial data frame of residuals
RES <- data.frame(res = residuals(FIT,
                                    type="normalized"),
                    x = data$x,
                    y = data$y)
coordinates(RES) <- c("x","y")
```

Autocorr. in the Russian Boreality data



```
library(ape)
library(fields)
library(gstat)
library(sp)

#Spatial data frame of residuals
RES <- data.frame(res = residuals(FIT,
                                    type="normalized"),
                    x = data$x,
                    y = data$y)
coordinates(RES) <- c("x", "y")

#Matrix of distances
w = fields::rdist(cbind(data$x, data$y))

#Calculate Moran's I
ape::Moran.I(RES$res, w = w)

$observed
[1] -0.01024323

$expected
[1] -0.001879699

$sd
[1] 0.001367237

$p.value
[1] 9.52903e-10
```

Autocorr. in the Russian Boreality data



```
library(ape)
library(fields)
library(gstat)
library(sp)

#Spatial data frame of residuals
RES <- data.frame(res = residuals(FIT,
                                    type="normalized"),
                    x = data$x,
                    y = data$y)
coordinates(RES) <- c("x", "y")

#Matrix of distances
w = fields::rdist(cbind(data$x, data$y))

#Calculate Moran's I
ape::Moran.I(RES$res, w = w)

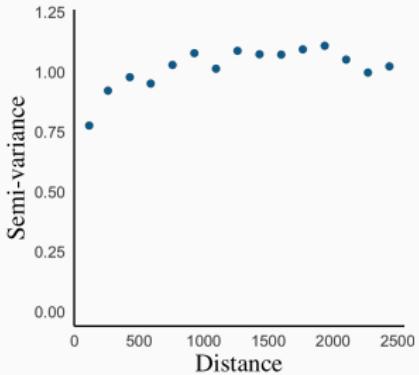
$observed
[1] -0.01024323

$expected
[1] -0.001879699

$sd
[1] 0.001367237

$p.value
[1] 9.52903e-10
```

```
#Calculate variogram
vg <- variogram(res ~ 1, data = RES)
```



Autocorr. in the Russian Boreality data



```
library(ape)
library(fields)
library(gstat)
library(sp)

#Spatial data frame of residuals
RES <- data.frame(res = residuals(FIT,
                                    type="normalized"),
                    x = data$x,
                    y = data$y)
coordinates(RES) <- c("x", "y")

#Matrix of distances
w = fields::rdist(cbind(data$x, data$y))

#Calculate Moran's I
ape::Moran.I(RES$res, w = w)

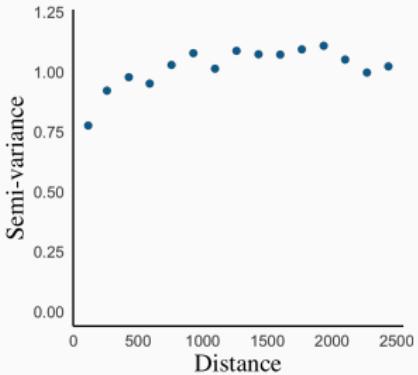
$observed
[1] -0.01024323

$expected
[1] -0.001879699

$sd
[1] 0.001367237

$p.value
[1] 9.52903e-10
```

```
#Calculate variogram
vg <- variogram(res ~ 1, data = RES)
```



Variogram shows initial curvature.

Autocorr. in the Russian Boreality data



```
library(ape)
library(fields)
library(gstat)
library(sp)

#Spatial data frame of residuals
RES <- data.frame(res = residuals(FIT,
                                    type="normalized"),
                    x = data$x,
                    y = data$y)
coordinates(RES) <- c("x", "y")

#Matrix of distances
w = fields::rdist(cbind(data$x, data$y))

#Calculate Moran's I
ape::Moran.I(RES$res, w = w)

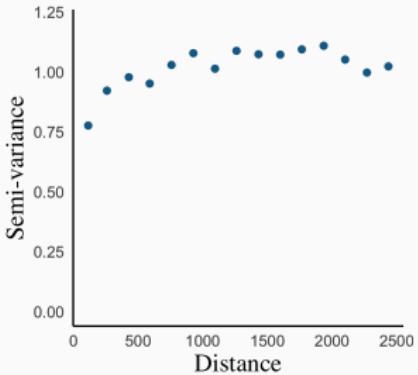
$observed
[1] -0.01024323

$expected
[1] -0.001879699

$sd
[1] 0.001367237

$p.value
[1] 9.52903e-10
```

```
#Calculate variogram
vg <- variogram(res ~ 1, data = RES)
```



Variogram shows initial curvature.

These data are clearly autocorrelated and the results can't be trusted.

Linear spatial correlation in R



THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus



A linear spatial correlation structure can be applied via the `corLin()` function.

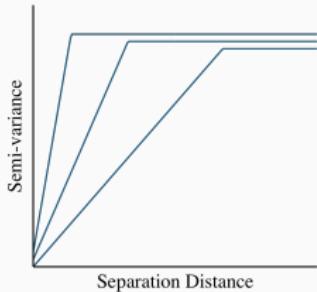
Linear spatial correlation in R



A linear spatial correlation structure can be applied via the `corLin()` function.

```
FIT_lin <- gls(Bor ~ Wet,
                 correlation = corLin(c(800, 0.75),
                                       form=~x+y,
                                       nugget = TRUE),
                 data = data)
```

$$\text{Linear } \Phi = 1 - (1 - \frac{D}{\rho})I(d < \rho)$$



Linear spatial correlation in R

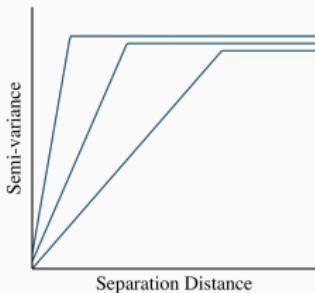


A linear spatial correlation structure can be applied via the `corLin()` function.

```
FIT_lin <- gls(Bor ~ Wet,
                 correlation = corLin(c(800, 0.75),
                                       form=~x+y,
                                       nugget = TRUE),
                 data = data)

summary(FIT_lin)
...
Correlation Structure: Linear spatial correlation
Formula: ~x + y
Parameter estimate(s):
    range      nugget
758.8801793  0.5601258
...
...
```

$$\text{Linear } \Phi = 1 - (1 - \frac{D}{\rho})I(d < \rho)$$



Linear spatial correlation in R



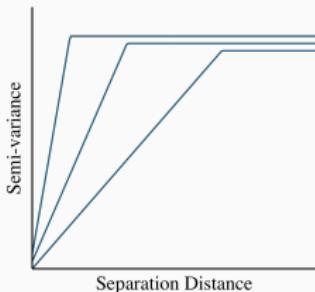
A linear spatial correlation structure can be applied via the `corLin()` function.

```
FIT_lin <- gls(Bor ~ Wet,
                 correlation = corLin(c(800, 0.75),
                                       form=~x+y,
                                       nugget = TRUE),
                 data = data)

summary(FIT_lin)
...
Correlation Structure: Linear spatial correlation
Formula: ~x + y
Parameter estimate(s):
  range      nugget
758.8801793   0.5601258
...
AIC(FIT, FIT_lin)

      df      AIC
FIT      3  2844.541
FIT_lin  5  2735.603
```

$$\text{Linear } \Phi = 1 - (1 - \frac{D}{\rho})I(d < \rho)$$



Linear spatial correlation in R



A linear spatial correlation structure can be applied via the `corLin()` function.

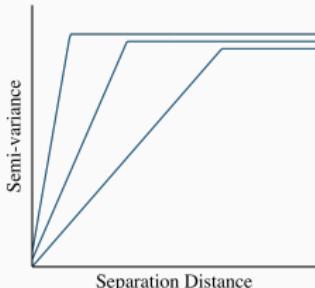
```
FIT_lin <- gls(Bor ~ Wet,
                 correlation = corLin(c(800, 0.75),
                                       form=~x+y,
                                       nugget = TRUE),
                 data = data)
```

```
summary(FIT_lin)
...
Correlation Structure: Linear spatial correlation
Formula: ~x + y
Parameter estimate(s):
  range      nugget
758.8801793   0.5601258
...
AIC(FIT, FIT_lin)
```

	df	AIC
FIT	3	2844.541
FIT_lin	5	2735.603

```
Moran.I(RES$res, w = w)
...
$p.value
[1] 0.6787855
Biol 520C: Statistical modelling for biological data
```

$$\text{Linear } \Phi = 1 - (1 - \frac{D}{\rho})I(d < \rho)$$



Linear spatial correlation in R



A linear spatial correlation structure can be applied via the `corLin()` function.

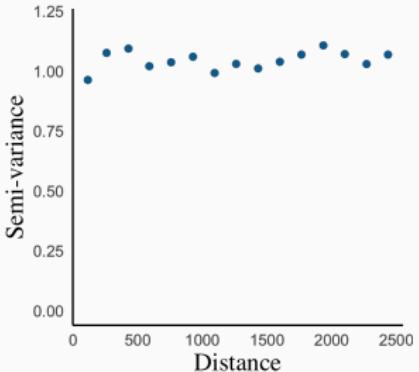
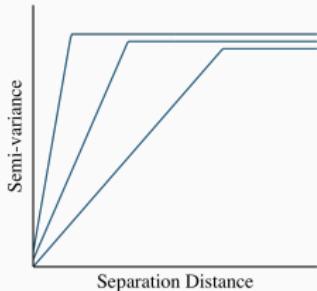
```
FIT_lin <- gls(Bor ~ Wet,
                 correlation = corLin(c(800, 0.75),
                                       form=~x+y,
                                       nugget = TRUE),
                 data = data)
```

```
summary(FIT_lin)
...
Correlation Structure: Linear spatial correlation
Formula: ~x + y
Parameter estimate(s):
  range      nugget
758.8801793   0.5601258
...
AIC(FIT, FIT_lin)
```

df	AIC
FIT	3 2844.541
FIT_lin	5 2735.603

```
Moran.I(RES$res, w = w)
...
$p.value
[1] 0.6787855
```

$$\text{Linear } \Phi = 1 - (1 - \frac{D}{\rho})I(d < \rho)$$



Spherical spatial correlation in R



THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

A spherical spatial correlation structure can be applied via the `corSpher()` function.

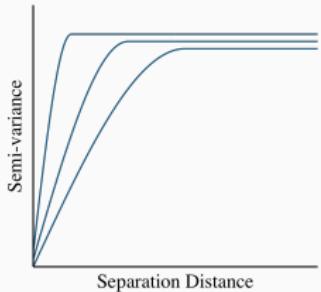
Spherical spatial correlation in R



A spherical spatial correlation structure can be applied via the `corSpher()` function.

```
FIT_Sph <- gls(Bor ~ Wet,
                 correlation = corSpher(c(800, 0.75),
                                         form=~x+y,
                                         nugget = TRUE),
                 data = data)
```

$$\text{Spherical } \Phi = 1(1 - 1.5(\frac{d}{\rho}) + 0.5(\frac{d}{\rho})^3)I(d < \rho)$$



Spherical spatial correlation in R

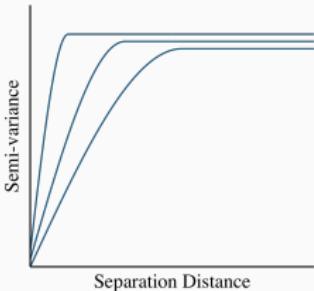


A spherical spatial correlation structure can be applied via the `corSpher()` function.

```
FIT_Sph <- gls(Bor ~ Wet,
                 correlation = corSpher(c(800, 0.75),
                                         form=~x+y,
                                         nugget = TRUE),
                 data = data)

summary(FIT_Sph)
...
Correlation Structure: Spherical spatial correlation
Formula: ~x + y
Parameter estimate(s):
    range      nugget
939.3200338   0.5211218
...
...
```

$$\text{Spherical } \Phi = 1(1 - 1.5(\frac{d}{\rho}) + 0.5(\frac{d}{\rho})^3)I(d < \rho)$$



Spherical spatial correlation in R



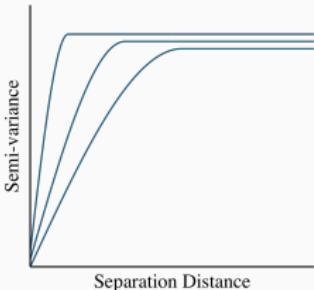
A spherical spatial correlation structure can be applied via the `corSpher()` function.

```
FIT_Sph <- gls(Bor ~ Wet,
                  correlation = corSpher(c(800, 0.75),
                                         form=~x+y,
                                         nugget = TRUE),
                  data = data)

summary(FIT_Sph)
...
Correlation Structure: Spherical spatial correlation
Formula: ~x + y
Parameter estimate(s):
    range      nugget
939.3200338   0.5211218
...
AIC(FIT, FIT_Sph)

      df      AIC
FIT     3  2844.541
FIT_Sph 5  2732.666
```

$$\text{Spherical } \Phi = 1\left(1 - 1.5\left(\frac{d}{\rho}\right) + 0.5\left(\frac{d}{\rho}\right)^3\right)I(d < \rho)$$



Spherical spatial correlation in R



A spherical spatial correlation structure can be applied via the `corSpher()` function.

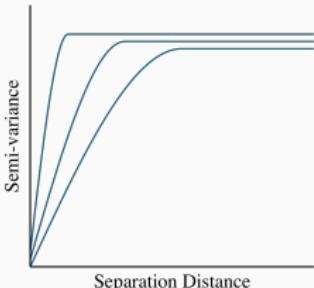
```
FIT_Sph <- gls(Bor ~ Wet,
                  correlation = corSpher(c(800, 0.75),
                                         form=~x+y,
                                         nugget = TRUE),
                  data = data)
```

```
summary(FIT_Sph)
...
Correlation Structure: Spherical spatial correlation
Formula: ~x + y
Parameter estimate(s):
    range      nugget
939.3200338   0.5211218
...
AIC(FIT, FIT_Sph)
```

df	AIC
FIT	3 2844.541
FIT_Sph	5 2732.666

```
Moran.I(RES$res, w = w)
...
$p.value
[1] 0.6511908
Biol 520C: Statistical modelling for biological data
```

$$\text{Spherical } \Phi = 1(1 - 1.5(\frac{d}{\rho}) + 0.5(\frac{d}{\rho})^3)I(d < \rho)$$



Spherical spatial correlation in R



A spherical spatial correlation structure can be applied via the `corSpher()` function.

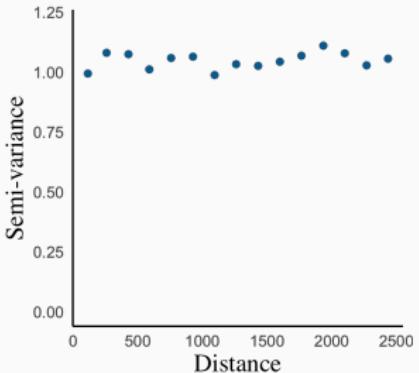
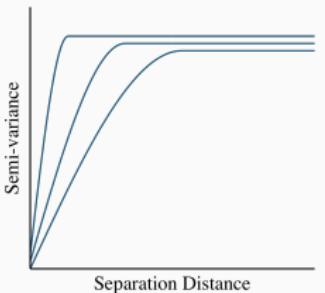
```
FIT_Sph <- gls(Bor ~ Wet,
                  correlation = corSpher(c(800, 0.75),
                                         form=~x+y,
                                         nugget = TRUE),
                  data = data)
```

```
summary(FIT_Sph)
...
Correlation Structure: Spherical spatial correlation
Formula: ~x + y
Parameter estimate(s):
  range      nugget
939.3200338   0.5211218
...
AIC(FIT, FIT_Sph)
```

df	AIC
FIT	3 2844.541
FIT_Sph	5 2732.666

```
Moran.I(RES$res, w = w)
...
$p.value
[1] 0.6511908
Biol 520C: Statistical modelling for biological data
```

$$\text{Spherical } \Phi = 1(1 - 1.5(\frac{d}{\rho}) + 0.5(\frac{d}{\rho})^3)I(d < \rho)$$



Gaussian spatial correlation in R



THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

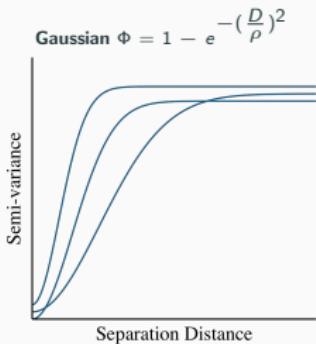
A Gaussian spatial correlation structure can be applied via the `corGaus()` function.

Gaussian spatial correlation in R



A Gaussian spatial correlation structure can be applied via the `corGaus()` function.

```
FIT_Gau <- gls(Bor ~ Wet,
                 correlation = corGaus(c(800, 0.75),
                                         form=~x+y,
                                         nugget = TRUE),
                 data = data)
```



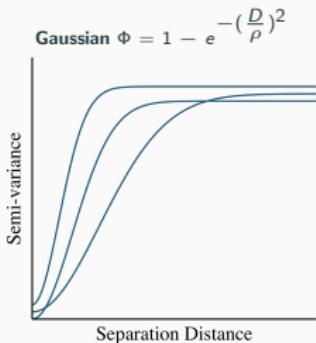
Gaussian spatial correlation in R



A Gaussian spatial correlation structure can be applied via the `corGaus()` function.

```
FIT_Gau <- gls(Bor ~ Wet,
                 correlation = corGaus(c(800, 0.75),
                                         form=~x+y,
                                         nugget = TRUE),
                 data = data)

summary(FIT_Gau)
...
Correlation Structure: Gaussian spatial correlation
Formula: ~x + y
Parameter estimate(s):
  range      nugget
460.3947751  0.6112509
...
...
```



Gaussian spatial correlation in R

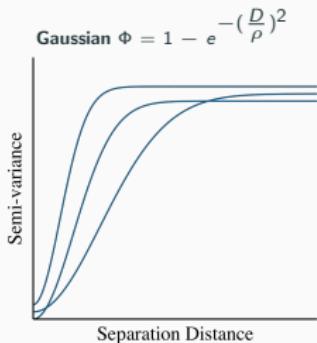


A Gaussian spatial correlation structure can be applied via the `corGaus()` function.

```
FIT_Gau <- gls(Bor ~ Wet,
                  correlation = corGaus(c(800, 0.75),
                                         form=~x+y,
                                         nugget = TRUE),
                  data = data)

summary(FIT_Gau)
...
Correlation Structure: Gaussian spatial correlation
Formula: ~x + y
Parameter estimate(s):
  range      nugget
460.3947751  0.6112509
...
AIC(FIT, FIT_Gau)

      df      AIC
FIT     3 2844.541
FIT_Gau 5 2736.292
```



Gaussian spatial correlation in R



A Gaussian spatial correlation structure can be applied via the `corGaus()` function.

```
FIT_Gau <- gls(Bor ~ Wet,
                  correlation = corGaus(c(800, 0.75),
                                         form=~x+y,
                                         nugget = TRUE),
                  data = data)
```

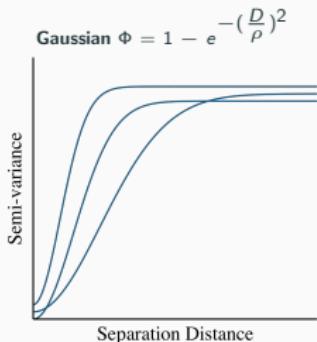
```
summary(FIT_Gau)
...
Correlation Structure: Gaussian spatial correlation
Formula: ~x + y
Parameter estimate(s):
  range      nugget
460.3947751  0.6112509
...
```

```
AIC(FIT, FIT_Gau)
```

	df	AIC
FIT	3	2844.541
FIT_Gau	5	2736.292

```
Moran.I(RES$res, w = w)
```

```
...
$p.value
[1] 0.6275071
```



Gaussian spatial correlation in R



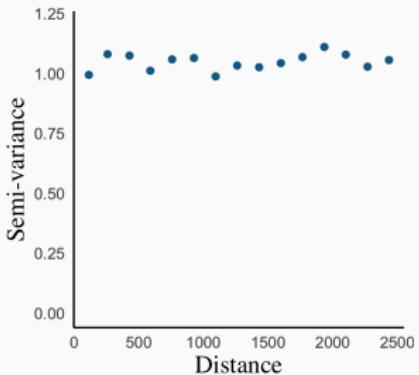
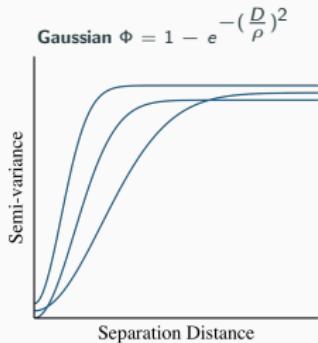
A Gaussian spatial correlation structure can be applied via the `corGaus()` function.

```
FIT_Gau <- gls(Bor ~ Wet,
                  correlation = corGaus(c(800, 0.75),
                                         form=~x+y,
                                         nugget = TRUE),
                  data = data)
```

```
summary(FIT_Gau)
...
Correlation Structure: Gaussian spatial correlation
Formula: ~x + y
Parameter estimate(s):
  range      nugget
460.3947751  0.6112509
...
AIC(FIT, FIT_Gau)
```

df	AIC
FIT	3 2844.541
FIT_Gau	5 2736.292

```
Moran.I(RES$res, w = w)
...
$p.value
[1] 0.6275071
```



Exponential spatial correlation in R



THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

Exponential spatial correlation in R



THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

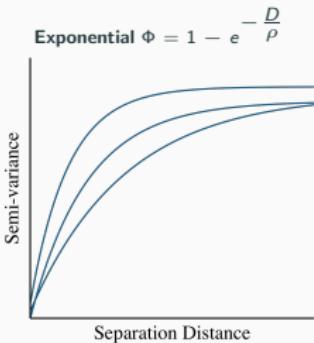
An exponential spatial correlation structure
can be applied via the `corExp()` function.

Exponential spatial correlation in R



An exponential spatial correlation structure can be applied via the `corExp()` function.

```
FIT_Exp <- gls(Bor ~ Wet,
                 correlation = corExp(c(800, 0.75),
                                       form=~x+y,
                                       nugget = TRUE),
                 data = data)
```



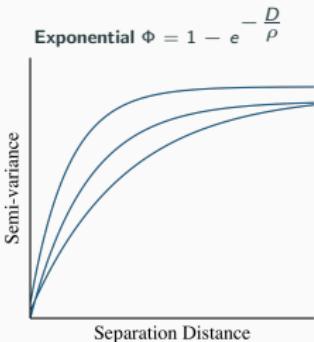
Exponential spatial correlation in R



An exponential spatial correlation structure can be applied via the `corExp()` function.

```
FIT_Exp <- gls(Bor ~ Wet,
                 correlation = corExp(c(800, 0.75),
                                       form=~x+y,
                                       nugget = TRUE),
                 data = data)

summary(FIT_Exp)
...
Correlation Structure: Exponential spatial correlation
Formula: ~x + y
Parameter estimate(s):
  range      nugget
481.1743349  0.4849357
...
...
```



Exponential spatial correlation in R

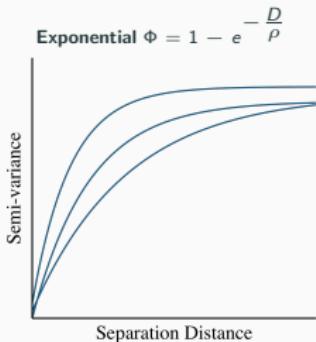


An exponential spatial correlation structure can be applied via the `corExp()` function.

```
FIT_Exp <- gls(Bor ~ Wet,
                 correlation = corExp(c(800, 0.75),
                                       form=~x+y,
                                       nugget = TRUE),
                 data = data)

summary(FIT_Exp)
...
Correlation Structure: Exponential spatial correlation
Formula: ~x + y
Parameter estimate(s):
  range      nugget
481.1743349  0.4849357
...
AIC(FIT, FIT_Exp)

      df      AIC
FIT     3 2844.541
FIT_Exp 5 2732.224
```



Exponential spatial correlation in R



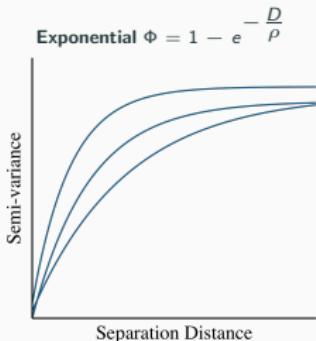
An exponential spatial correlation structure can be applied via the `corExp()` function.

```
FIT_Exp <- gls(Bor ~ Wet,
                 correlation = corExp(c(800, 0.75),
                                       form=~x+y,
                                       nugget = TRUE),
                 data = data)
```

```
summary(FIT_Exp)
...
Correlation Structure: Exponential spatial correlation
Formula: ~x + y
Parameter estimate(s):
  range      nugget
481.1743349  0.4849357
...
AIC(FIT, FIT_Exp)
```

	df	AIC
FIT	3	2844.541
FIT_Exp	5	2732.224

```
Moran.I(RES$res, w = w)
...
$p.value
[1] 0.6519741
```



Exponential spatial correlation in R



An exponential spatial correlation structure can be applied via the `corExp()` function.

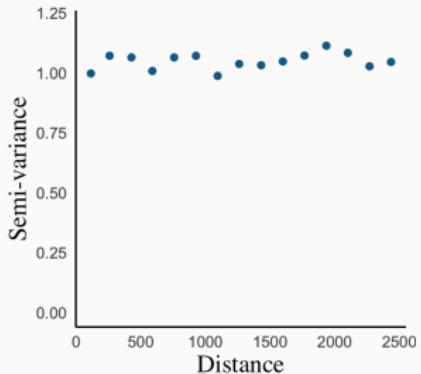
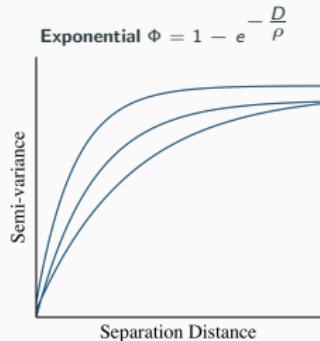
```
FIT_Exp <- gls(Bor ~ Wet,
                 correlation = corExp(c(800, 0.75),
                                       form=~x+y,
                                       nugget = TRUE),
                 data = data)
```

```
summary(FIT_Exp)
...
Correlation Structure: Exponential spatial correlation
```

```
Formula: ~x + y
Parameter estimate(s):
range      nugget
481.1743349  0.4849357
...
AIC(FIT, FIT_Exp)
```

df	AIC
FIT	3 2844.541
FIT_Exp	5 2732.224

```
Moran.I(RES$res, w = w)
...
$p.value
[1] 0.6519741
```



Rational quad. spatial correlation



THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus



A rational quadratic spatial correlation structure can be applied via the `corRatio()` function.

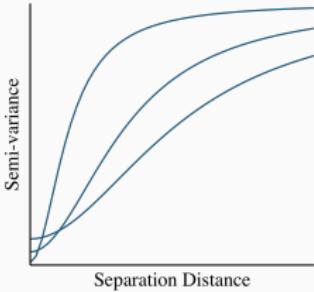
Rational quad. spatial correlation



A rational quadratic spatial correlation structure can be applied via the `corRatio()` function.

```
FIT_Rat <- gls(Bor ~ Wet,
                 correlation = corRatio(c(800, 0.75),
                                         form=~x+y,
                                         nugget = TRUE),
                 data = data)
```

$$\text{Rational quadratic } \Phi = \frac{1}{1 + (\frac{p}{d})^2}$$



Rational quad. spatial correlation

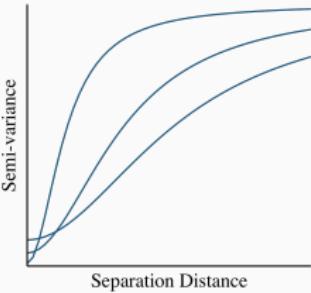


A rational quadratic spatial correlation structure can be applied via the `corRatio()` function.

```
FIT_Rat <- gls(Bor ~ Wet,
                 correlation = corRatio(c(800, 0.75),
                                         form=~x+y,
                                         nugget = TRUE),
                 data = data)

summary(FIT_Rat)
...
Correlation Structure: Rational quadratic spatial
correlation
Formula: ~x + y
Parameter estimate(s):
      range      nugget
373.0351282  0.5627731
...
...
```

$$\text{Rational quadratic } \Phi = \frac{1}{1+(\frac{p}{d})^2}$$



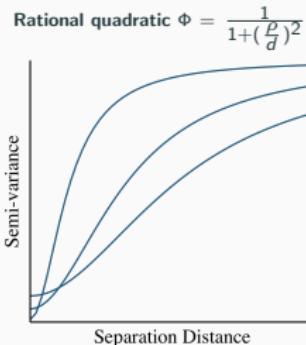
Rational quad. spatial correlation



A rational quadratic spatial correlation structure can be applied via the `corRatio()` function.

```
FIT_Rat <- gls(Bor ~ Wet,
                 correlation = corRatio(c(800, 0.75),
                                         form=~x+y,
                                         nugget = TRUE),
                 data = data)

summary(FIT_Rat)
...
Correlation Structure: Rational quadratic spatial
correlation
Formula: ~x + y
Parameter estimate(s):
      range      nugget
373.0351282  0.5627731
...
AIC(FIT, FIT_Rat)
      df      AIC
FIT      3 2844.541
FIT_Rat  5 2732.930
```



Rational quad. spatial correlation

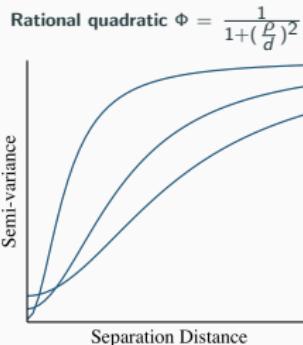


A rational quadratic spatial correlation structure can be applied via the `corRatio()` function.

```
FIT_Rat <- gls(Bor ~ Wet,
                 correlation = corRatio(c(800, 0.75),
                                         form=~x+y,
                                         nugget = TRUE),
                 data = data)
```

```
summary(FIT_Rat)
...
Correlation Structure: Rational quadratic spatial
correlation
Formula: ~x + y
Parameter estimate(s):
      range      nugget
373.0351282  0.5627731
...
AIC(FIT, FIT_Rat)
      df      AIC
FIT      3 2844.541
FIT_Rat  5 2732.930

Moran.I(RES$res, w = w)
...
$p.value
[1] 0.6508174
```



Rational quad. spatial correlation



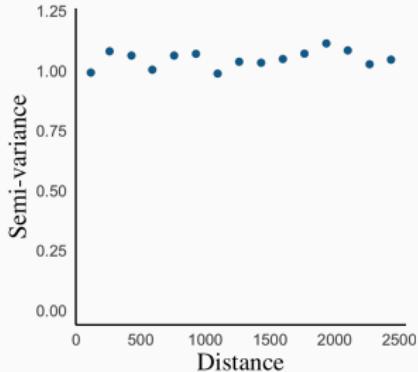
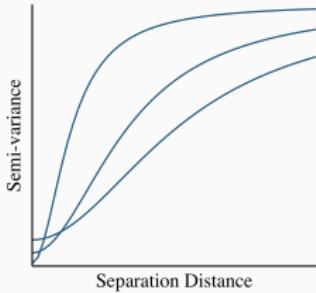
A rational quadratic spatial correlation structure can be applied via the `corRatio()` function.

```
FIT_Rat <- gls(Bor ~ Wet,
                 correlation = corRatio(c(800, 0.75),
                                         form=~x+y,
                                         nugget = TRUE),
                 data = data)
```

```
summary(FIT_Rat)
...
Correlation Structure: Rational quadratic spatial
correlation
Formula: ~x + y
Parameter estimate(s):
      range      nugget
373.0351282   0.5627731
...
AIC(FIT, FIT_Rat)
      df      AIC
FIT      3 2844.541
FIT_Rat  5 2732.930

Moran.I(RES$res, w = w)
...
$p.value
[1] 0.6508174
```

$$\text{Rational quadratic } \Phi = \frac{1}{1+(\frac{p}{d})^2}$$



Selecting the best Structure



THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

Selecting the best Structure



We just fit 6 different models, but
how do we know which correlation
structure to go with?

Selecting the best Structure



We just fit 6 different models, but
how do we know which correlation
structure to go with?

```
#Calculate AIC values
TABLE <- AIC(FIT, FIT_lin, FIT_Exp,
               FIT_Rat, FIT_Sph, FIT_Gau)

#Ordered by lowest to highest AIC
TABLE <- TABLE[order(TABLE$AIC),]

#Calculate Delta AICs
TABLE$DeltaAIC <- TABLE$AIC - TABLE$AIC[1]

#Evidence compared to AIC best model
TABLE$Evidence <- 1/exp(-TABLE$DeltaAIC/2)

TABLE
```

	df	AIC	DeltaAIC	Evidence
FIT_Exp	5	2732.224	0.0000000	1.000000e+00
FIT_Sph	5	2732.666	0.4423134	1.247519e+00
FIT_Rat	5	2732.930	0.7063538	1.423583e+00
FIT_lin	5	2735.603	3.3797773	5.418877e+00
FIT_Gau	5	2736.292	4.0685720	7.646790e+00
FIT	3	2844.541	112.3174833	2.451498e+24

Selecting the best Structure



We just fit 6 different models, but how do we know which correlation structure to go with?

```
#Calculate AIC values
TABLE <- AIC(FIT, FIT_lin, FIT_Exp,
               FIT_Rat, FIT_Sph, FIT_Gau)

#Ordered by lowest to highest AIC
TABLE <- TABLE[order(TABLE$AIC),]

#Calculate Delta AICs
TABLE$DeltaAIC <- TABLE$AIC - TABLE$AIC[1]

#Evidence compared to AIC best model
TABLE$Evidence <- 1/exp(-TABLE$DeltaAIC/2)

TABLE
```

	df	AIC	DeltaAIC	Evidence
FIT_Exp	5	2732.224	0.0000000	1.000000e+00
FIT_Sph	5	2732.666	0.4423134	1.247519e+00
FIT_Rat	5	2732.930	0.7063538	1.423583e+00
FIT_lin	5	2735.603	3.3797773	5.418877e+00
FIT_Gau	5	2736.292	4.0685720	7.646790e+00
FIT	3	2844.541	112.3174833	2.451498e+24

```
#Calculate empirical variogram
vg <- Variogram(FIT, form = ~ x + y,
                  robust = TRUE,
                  maxDist = 2500)
```

Selecting the best Structure



We just fit 6 different models, but how do we know which correlation structure to go with?

```
#Calculate AIC values
TABLE <- AIC(FIT, FIT_lin, FIT_Exp,
               FIT_Rat, FIT_Sph, FIT_Gau)

#Ordered by lowest to highest AIC
TABLE <- TABLE[order(TABLE$AIC),]

#Calculate Delta AICs
TABLE$DeltaAIC <- TABLE$AIC - TABLE$AIC[1]

#Evidence compared to AIC best model
TABLE$Evidence <- 1/exp(-TABLE$DeltaAIC/2)

TABLE

      df      AIC     DeltaAIC     Evidence
FIT_Exp  5 2732.224  0.0000000 1.000000e+00
FIT_Sph  5 2732.666  0.4423134 1.247519e+00
FIT_Rat  5 2732.930  0.7063538 1.423583e+00
FIT_lin  5 2735.603  3.3797773 5.418877e+00
FIT_Gau  5 2736.292  4.0685720 7.646790e+00
FIT      3 2844.541 112.3174833 2.451498e+24
```

Selecting the best Structure



We just fit 6 different models, but how do we know which correlation structure to go with?

```
#Calculate AIC values
TABLE <- AIC(FIT, FIT_lin, FIT_Exp,
               FIT_Rat, FIT_Sph, FIT_Gau)

#Ordered by lowest to highest AIC
TABLE <- TABLE[order(TABLE$AIC),]

#Calculate Delta AICs
TABLE$DeltaAIC <- TABLE$AIC - TABLE$AIC[1]

#Evidence compared to AIC best model
TABLE$Evidence <- 1/exp(-TABLE$DeltaAIC/2)

TABLE
```

	df	AIC	DeltaAIC	Evidence
FIT_Exp	5	2732.224	0.0000000	1.000000e+00
FIT_Sph	5	2732.666	0.4423134	1.247519e+00
FIT_Rat	5	2732.930	0.7063538	1.423583e+00
FIT_lin	5	2735.603	3.3797773	5.418877e+00
FIT_Gau	5	2736.292	4.0685720	7.646790e+00
FIT	3	2844.541	112.3174833	2.451498e+24

```
#Calculate empirical variogram
vg <- Variogram(FIT, form = ~ x + y,
                  robust = TRUE,
                  maxDist = 2500)

#Generate the fitted exp vgm
D <- data.frame(D = seq(0,2500, 1))

EXP_FIT <- corExp(c(481.1743349,
                     0.4849357),
                    form = ~D,
                    nugget = TRUE)
```

Selecting the best Structure



We just fit 6 different models, but how do we know which correlation structure to go with?

```
#Calculate AIC values
TABLE <- AIC(FIT, FIT_lin, FIT_Exp,
               FIT_Rat, FIT_Sph, FIT_Gau)

#Ordered by lowest to highest AIC
TABLE <- TABLE[order(TABLE$AIC),]

#Calculate Delta AICs
TABLE$DeltaAIC <- TABLE$AIC - TABLE$AIC[1]

#Evidence compared to AIC best model
TABLE$Evidence <- 1/exp(-TABLE$DeltaAIC/2)

TABLE
```

	df	AIC	DeltaAIC	Evidence
FIT_Exp	5	2732.224	0.0000000	1.000000e+00
FIT_Sph	5	2732.666	0.4423134	1.247519e+00
FIT_Rat	5	2732.930	0.7063538	1.423583e+00
FIT_lin	5	2735.603	3.3797773	5.418877e+00
FIT_Gau	5	2736.292	4.0685720	7.646790e+00
FIT	3	2844.541	112.3174833	2.451498e+24

```
#Calculate empirical variogram
vg <- Variogram(FIT, form = ~ x + y,
                  robust = TRUE,
                  maxDist = 2500)

#Generate the fitted exp vgm
D <- data.frame(D = seq(0,2500, 1))

EXP_FIT <- corExp(c(481.1743349,
                     0.4849357),
                    form = ~D,
                    nugget = TRUE)

EXP_FIT2 <- Initialize(EXP_FIT,data=D)
```

Selecting the best Structure



We just fit 6 different models, but how do we know which correlation structure to go with?

```
#Calculate AIC values
TABLE <- AIC(FIT, FIT_lin, FIT_Exp,
               FIT_Rat, FIT_Sph, FIT_Gau)

#Ordered by lowest to highest AIC
TABLE <- TABLE[order(TABLE$AIC),]

#Calculate Delta AICs
TABLE$DeltaAIC <- TABLE$AIC - TABLE$AIC[1]

#Evidence compared to AIC best model
TABLE$Evidence <- 1/exp(-TABLE$DeltaAIC/2)

TABLE
```

	df	AIC	DeltaAIC	Evidence
FIT_Exp	5	2732.224	0.0000000	1.000000e+00
FIT_Sph	5	2732.666	0.4423134	1.247519e+00
FIT_Rat	5	2732.930	0.7063538	1.423583e+00
FIT_lin	5	2735.603	3.3797773	5.418877e+00
FIT_Gau	5	2736.292	4.0685720	7.646790e+00
FIT	3	2844.541	112.3174833	2.451498e+24

```
#Calculate empirical variogram
vg <- Variogram(FIT, form = ~ x + y,
                  robust = TRUE,
                  maxDist = 2500)

#Generate the fitted exp vgm
D <- data.frame(D = seq(0,2500, 1))

EXP_FIT <- corExp(c(481.1743349,
                     0.4849357),
                    form = ~D,
                    nugget = TRUE)

EXP_FIT2 <- Initialize(EXP_FIT,data=D)

selected_mod <- Variogram(EXP_FIT2)
```

Selecting the best Structure



We just fit 6 different models, but how do we know which correlation structure to go with?

```
#Calculate AIC values
TABLE <- AIC(FIT, FIT_lin, FIT_Exp,
               FIT_Rat, FIT_Sph, FIT_Gau)

#Ordered by lowest to highest AIC
TABLE <- TABLE[order(TABLE$AIC),]

#Calculate Delta AICs
TABLE$DeltaAIC <- TABLE$AIC - TABLE$AIC[1]

#Evidence compared to AIC best model
TABLE$Evidence <- 1/exp(-TABLE$DeltaAIC/2)

TABLE
```

	df	AIC	DeltaAIC	Evidence
FIT_Exp	5	2732.224	0.0000000	1.000000e+00
FIT_Sph	5	2732.666	0.4423134	1.247519e+00
FIT_Rat	5	2732.930	0.7063538	1.423583e+00
FIT_lin	5	2735.603	3.3797773	5.418877e+00
FIT_Gau	5	2736.292	4.0685720	7.646790e+00
FIT	3	2844.541	112.3174833	2.451498e+24

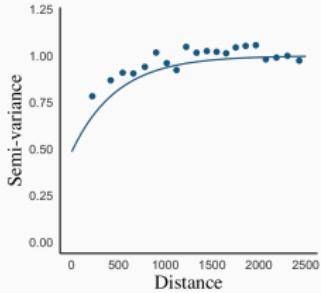
```
#Calculate empirical variogram
vg <- Variogram(FIT, form = ~ x + y,
                  robust = TRUE,
                  maxDist = 2500)

#Generate the fitted exp vgm
D <- data.frame(D = seq(0,2500, 1))

EXP_FIT <- corExp(c(481.1743349,
                     0.4849357),
                    form = ~D,
                    nugget = TRUE)

EXP_FIT2 <- Initialize(EXP_FIT,data=D)

selected_mod <- Variogram(EXP_FIT2)
```



Corrected model



THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

Corrected model



Original Model

Generalized least squares fit by REML

Model: Bor ~ Wet

Data: data

AIC	BIC	logLik
2844.541	2857.365	-1419.271

Coefficients:

	Value	Std.Error	t-value	p-value
(Intercept)	13.05622	0.151194	86.35430	0
Wet	165.80355	10.599111	15.64316	0

Correlation:

(Intr)

Wet 0

Residual standard error: 3.490577

Degrees of freedom: 533 total; 531 residual

Exponential spatial correlation model

Generalized least squares fit by REML

Model: Bor ~ Wet

Data: data

AIC	BIC	logLik
2732.224	2753.597	-1361.112

Coefficients:

	Value	Std.Error	t-value	p-value
(Intercept)	12.55226	0.678111	18.510619	0
Wet	75.43166	13.541700	5.570325	0

Correlation:

(Intr)

Wet 0.041

Residual standard error: 3.707339

Degrees of freedom: 533 total; 531 residual

Correlation Structure: Exponential spatial correlation

Formula: ~x + y

Parameter estimate(s):

range	nugget
481.1743349	0.4849357

We covered several ways to model spatially autocorrelated data:

Type	Description	R Function
IID	0	—
Linear	$\Phi = 1 - (1 - \frac{D}{\rho})I(d < \rho)$	corLin()
Spherical	$\Phi = 1(1 - 1.5(\frac{d}{\rho}) + 0.5(\frac{d}{\rho})^3)I(d < \rho)$	corSpher()
Gaussian	$\Phi = 1 - e^{-(\frac{D}{\rho})^2}$	corGaus()
Exponential	$\Phi = 1 - e^{-\frac{D}{\rho}}$	corExp()
Rational quadratic	$\Phi = \frac{1}{1+(\frac{D}{\rho})^2}$	corRatio()

We covered several ways to model spatially autocorrelated data:

Type	Description	R Function
IID	0	—
Linear	$\Phi = 1 - (1 - \frac{D}{\rho})I(d < \rho)$	corLin()
Spherical	$\Phi = 1(1 - 1.5(\frac{d}{\rho}) + 0.5(\frac{d}{\rho})^3)I(d < \rho)$	corSpher()
Gaussian	$\Phi = 1 - e^{-(\frac{D}{\rho})^2}$	corGaus()
Exponential	$\Phi = 1 - e^{-\frac{D}{\rho}}$	corExp()
Rational quadratic	$\Phi = \frac{1}{1+(\frac{D}{\rho})^2}$	corRatio()

The model structures can be difficult to interpret, but their variograms have very recognizable features.

We covered several ways to model spatially autocorrelated data:

Type	Description	R Function
IID	0	—
Linear	$\Phi = 1 - (1 - \frac{D}{\rho})I(d < \rho)$	corLin()
Spherical	$\Phi = 1(1 - 1.5(\frac{d}{\rho}) + 0.5(\frac{d}{\rho})^3)I(d < \rho)$	corSpher()
Gaussian	$\Phi = 1 - e^{-(\frac{D}{\rho})^2}$	corGaus()
Exponential	$\Phi = 1 - e^{-\frac{D}{\rho}}$	corExp()
Rational quadratic	$\Phi = \frac{1}{1 + (\frac{D}{\rho})^2}$	corRatio()

The model structures can be difficult to interpret, but their variograms have very recognizable features. Familiarising yourself with them will help you quickly narrow down what structure to use.

Technical considerations



THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus



Fitting models with complex correlation structures can be quite slow at times.



Fitting models with complex correlation structures can be quite slow at times.

```
system.time(gls(Bor ~ Wet, data = data))
  user  system elapsed
0.001   0.000   0.001
```

Fitting models with complex correlation structures can be quite slow at times.

```
system.time(gls(Bor ~ Wet, data = data))
  user  system elapsed
 0.001   0.000   0.001
```

```
CORR <- corSpher(c(800, 0.75),
                  form=~x+y,
                  nugget = TRUE)

system.time(gls(Bor ~ Wet,
               correlation = CORR,
               data = data))

  user  system elapsed
 9.781   0.109   9.893
```

Fitting models with complex correlation structures can be quite slow at times.

```
system.time(gls(Bor ~ Wet, data = data))
  user  system elapsed
0.001   0.000   0.001
```

```
CORR <- corSpher(c(800, 0.75),
                  form=~x+y,
                  nugget = TRUE)

system.time(gls(Bor ~ Wet,
               correlation = CORR,
               data = data))

  user  system elapsed
9.781   0.109   9.893
```

Took nearly 10,000 times longer!

Technical considerations



Fitting models with complex correlation structures can be quite slow at times.

```
system.time(gls(Bor ~ Wet, data = data))
  user  system elapsed
0.001   0.000   0.001
```

```
CORR <- corSpher(c(800, 0.75),
                  form=~x+y,
                  nugget = TRUE)

system.time(gls(Bor ~ Wet,
                correlation = CORR,
                data = data))

  user  system elapsed
9.781   0.109   9.893
```

Can also run into numerical issues without good starting values.

Took nearly 10,000 times longer!

Technical considerations



Fitting models with complex correlation structures can be quite slow at times.

```
system.time(gls(Bor ~ Wet, data = data))
  user   system elapsed
 0.001   0.000   0.001
```

```
CORR <- corSpher(c(800, 0.75),
                  form=~x+y,
                  nugget = TRUE)

system.time(gls(Bor ~ Wet,
                correlation = CORR,
                data = data))

  user   system elapsed
 9.781   0.109   9.893
```

Took nearly 10,000 times longer!

Can also run into numerical issues without good starting values.

```
FIT <- gls(Bor ~ Wet,
            correlation = corLin(form=~x+y,
                                  nugget=TRUE),
            data = data)
```

Technical considerations



Fitting models with complex correlation structures can be quite slow at times.

```
system.time(gls(Bor ~ Wet, data = data))
  user   system elapsed
 0.001   0.000   0.001
```

```
CORR <- corSpher(c(800, 0.75),
                  form=~x+y,
                  nugget = TRUE)
```

```
system.time(gls(Bor ~ Wet,
                correlation = CORR,
                data = data))
```

```
  user   system elapsed
 9.781   0.109   9.893
```

Can also run into numerical issues without good starting values.

```
FIT <- gls(Bor ~ Wet,
            correlation = corLin(form=~x+y,
                                   nugget=TRUE),
            data = data)
```

```
Error in gls(Bor ~ Wet, correlation = corLin(
  form = ~x + y, nugget=TRUE), :
  false convergence (8)
```

Took nearly 10,000 times longer!

Technical considerations



Fitting models with complex correlation structures can be quite slow at times.

```
system.time(gls(Bor ~ Wet, data = data))
  user   system elapsed
 0.001   0.000   0.001
```

```
CORR <- corSpher(c(800, 0.75),
                  form=~x+y,
                  nugget = TRUE)

system.time(gls(Bor ~ Wet,
                correlation = CORR,
                data = data))

  user   system elapsed
 9.781   0.109   9.893
```

Took nearly 10,000 times longer!

Can also run into numerical issues without good starting values.

```
FIT <- gls(Bor ~ Wet,
            correlation = corLin(form=~x+y,
                                   nugget=TRUE),
            data = data)
```

```
Error in gls(Bor ~ Wet, correlation = corLin(
  form = ~x + y, nugget=TRUE), :
  false convergence (8)
```

If that happens you need to provide better starting guesstimates or modify the optimiser.

Technical considerations



Fitting models with complex correlation structures can be quite slow at times.

```
system.time(gls(Bor ~ Wet, data = data))
  user   system elapsed
 0.001   0.000   0.001
```

```
CORR <- corSpher(c(800, 0.75),
                  form=~x+y,
                  nugget = TRUE)

system.time(gls(Bor ~ Wet,
                correlation = CORR,
                data = data))

  user   system elapsed
 9.781   0.109   9.893
```

Took nearly 10,000 times longer!

Can also run into numerical issues without good starting values.

```
FIT <- gls(Bor ~ Wet,
            correlation = corLin(form=~x+y,
                                   nugget=TRUE),
            data = data)
```

```
Error in gls(Bor ~ Wet, correlation = corLin(
  form = ~x + y, nugget=TRUE), :
  false convergence (8)
```

If that happens you need to provide better starting guesstimates or modify the optimiser. Neither of which are particularly enjoyable.

Spatially Autocorrelated Data



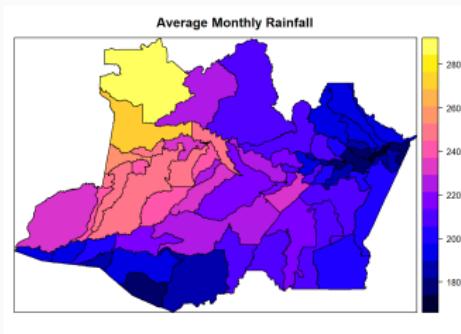
THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

Spatially Autocorrelated Data



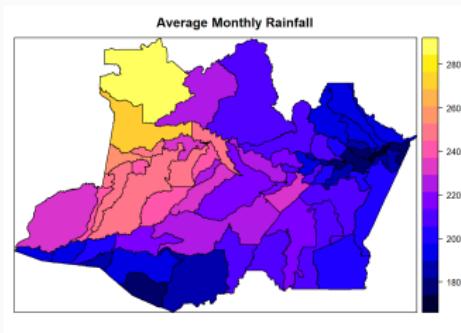
Experimental designs that do not consider spatial autocorrelation risk being subsampled.

A spatial autocorrelated data point will add little independent information, but inflate n (Dormann *et al.*, 2007).



Experimental designs that do not consider spatial autocorrelation risk being subsampled.

A spatial autocorrelated data point will add little independent information, but inflate n (Dormann *et al.*, 2007).



Source: Tadashi Fukami and Jes Coyle

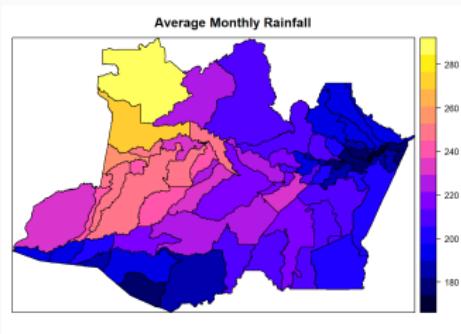
Corrections exist to deal with issues of statistical bias, but they can't inject more information into a dataset when none exists.

Spatially Autocorrelated Data



Experimental designs that do not consider spatial autocorrelation risk being subsampled.

A spatial autocorrelated data point will add little independent information, but inflate n (Dormann *et al.*, 2007).



Corrections exist to deal with issues of statistical bias, but they can't inject more information into a dataset when none exists.

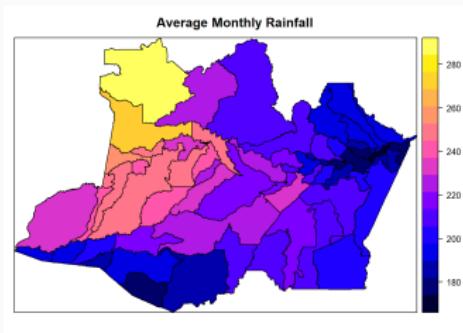
Good study design should consider spatial autocorrelation *a priori*.

Spatially Autocorrelated Data



Experimental designs that do not consider spatial autocorrelation risk being subsampled.

A spatial autocorrelated data point will add little independent information, but inflate n (Dormann *et al.*, 2007).



Corrections exist to deal with issues of statistical bias, but they can't inject more information into a dataset when none exists.

Good study design should consider spatial autocorrelation *a priori*.

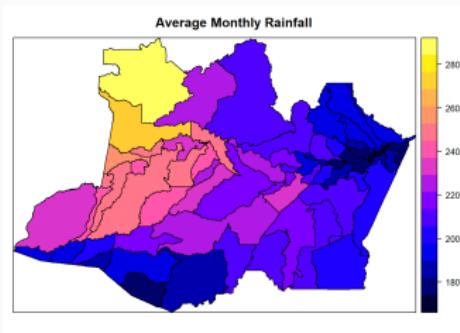
If you had to collect more data for the boreality study how far apart would you sample?

Spatially Autocorrelated Data



Experimental designs that do not consider spatial autocorrelation risk being subsampled.

A spatial autocorrelated data point will add little independent information, but inflate n (Dormann *et al.*, 2007).



Source: Tadashi Fukami and Jes Coyle

Corrections exist to deal with issues of statistical bias, but they can't inject more information into a dataset when none exists.

Good study design should consider spatial autocorrelation *a priori*.

If you had to collect more data for the boreality study how far apart would you sample? $> 500m$

References

- Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D. et al. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30, 609–628.
- Liang, M., Liu, X., Parker, I.M., Johnson, D., Zheng, Y., Luo, S., Gilbert, G.S. & Yu, S. (2019). Soil microbes drive phylogenetic diversity-productivity relationships in a subtropical forest. *Science advances*, 5, eaax5088.
- Zuur, A., Ieno, E.N. & Smith, G.M. (2007). *Analyzing ecological data*. Springer.