

# SpacialModels\_\_HW5\_\_CSO

*Caroline Oliver*

*2/20/2018*

```
library(vegan)
```

```
## Loading required package: permute
```

```
## Loading required package: lattice
```

```
## This is vegan 2.4-6
```

```
data(BCI)
```

```
data(BCI.env)
```

```
## UTM Coordinates (in metres)
```

```
BCI_xy = data.frame(x = rep(seq(625754, 626654, by=100), each=5),  
                    y = rep(seq(1011569, 1011969, by=100), len=50))
```

Split BCI data into rare and common datasets

```
abu = colSums(BCI)
```

```
BCIrare = BCI[ , abu<=7]
```

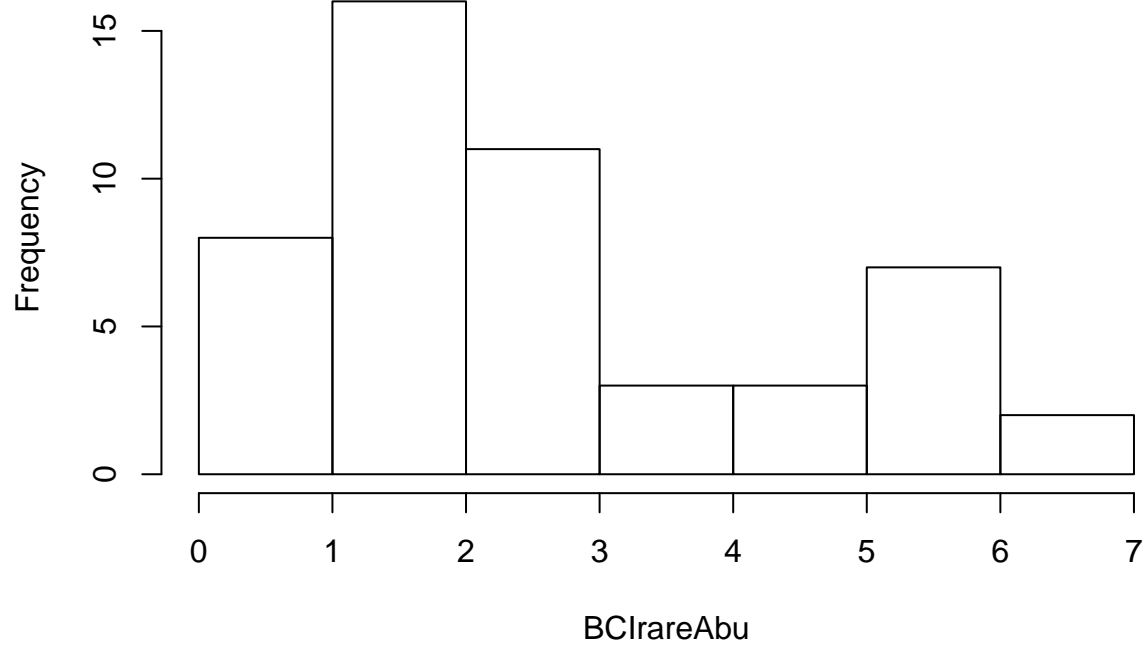
```
BCIcommon = BCI[ , abu>80]
```

Rare BCI

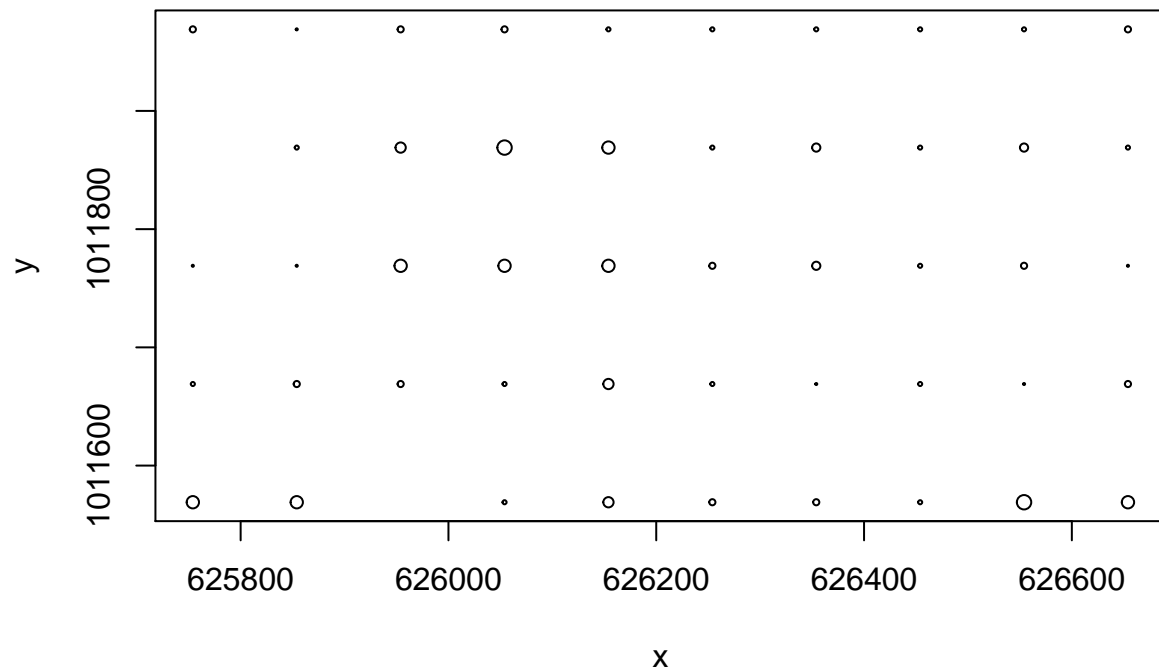
```
BCIrareAbu = apply(BCIrare, 1, function(x) sum(x > 0))
```

```
hist(BCIrareAbu)
```

**Histogram of BCIrareAbu**

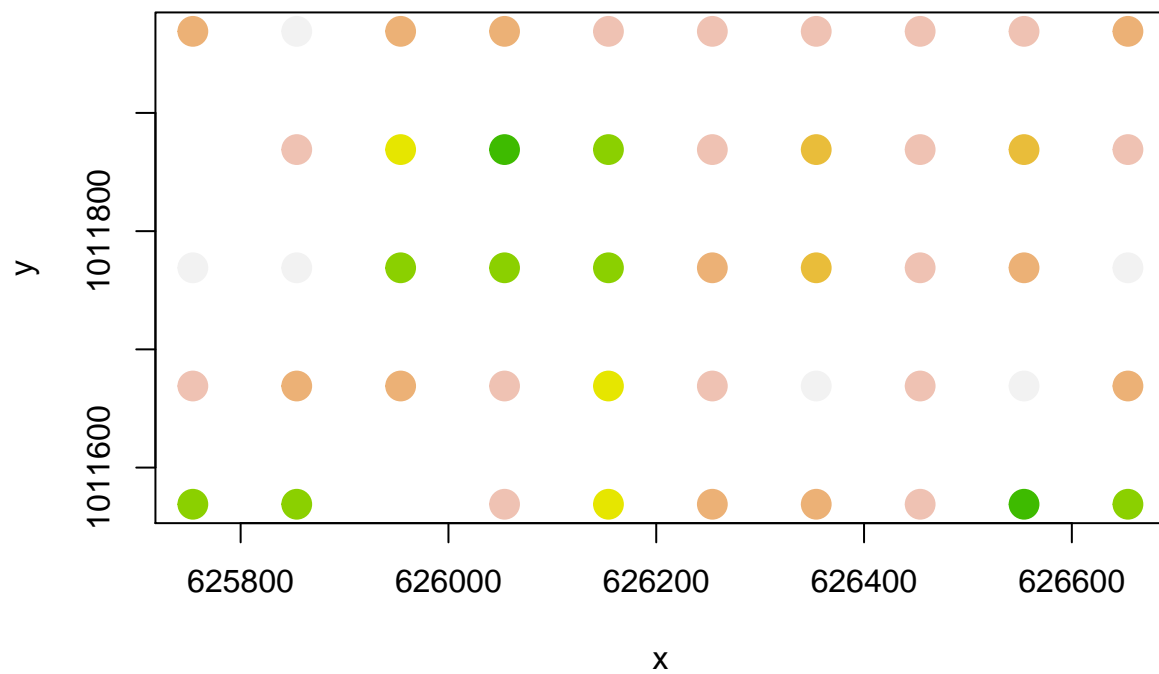


```
plot(BCI_xy, cex=BCIrareAbu/max(BCIrareAbu))
```



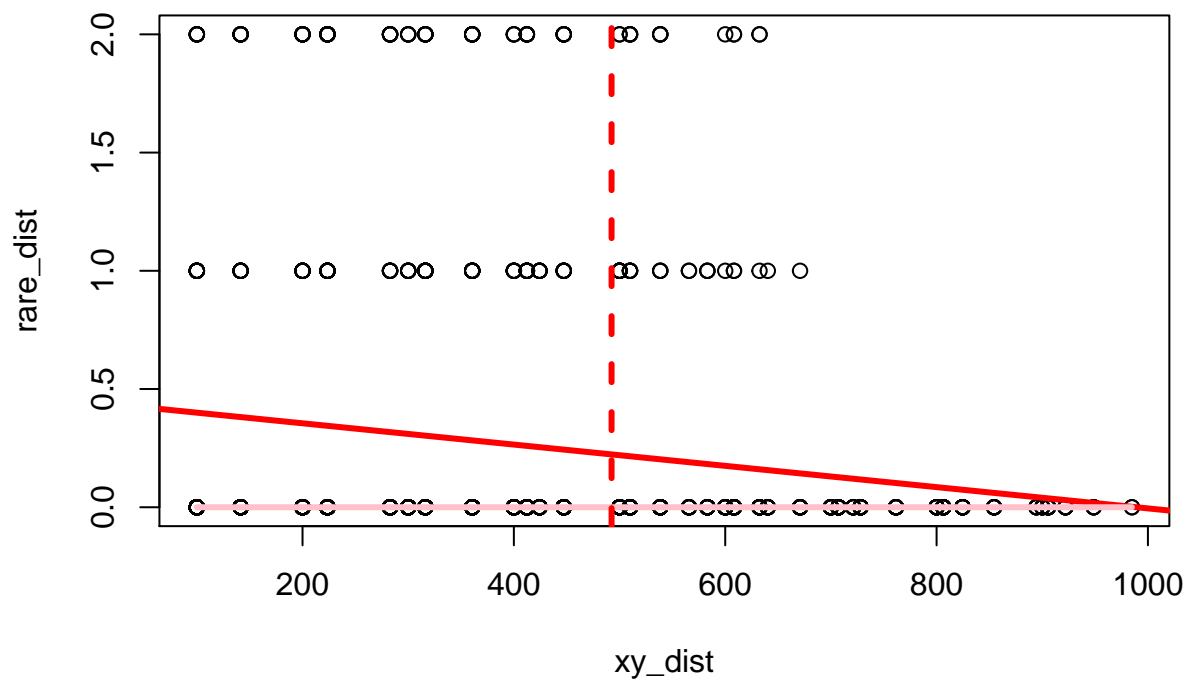
```
rare_species = BCIrare$Tetrathylacium.johansenii

col_brks = hist(BCIrareAbu, plot=F)$breaks
col_indices = as.numeric(cut(BCIrareAbu, col_brks))
cols = rev(terrain.colors(length(col_brks)))
plot(BCI_xy, cex=2, pch=19, col=cols[col_indices])
```



```
xy_dist = dist(BCI_xy)
max_dist = max(xy_dist) / 2
rare_dist = dist(rare_species)

plot(xy_dist, rare_dist)
abline(lm(rare_dist ~ xy_dist), lwd=3, col='red')
lines(lowess(xy_dist, rare_dist), lwd=3, col='pink')
abline(v = max_dist, col='red', lwd=3, lty=2)
```

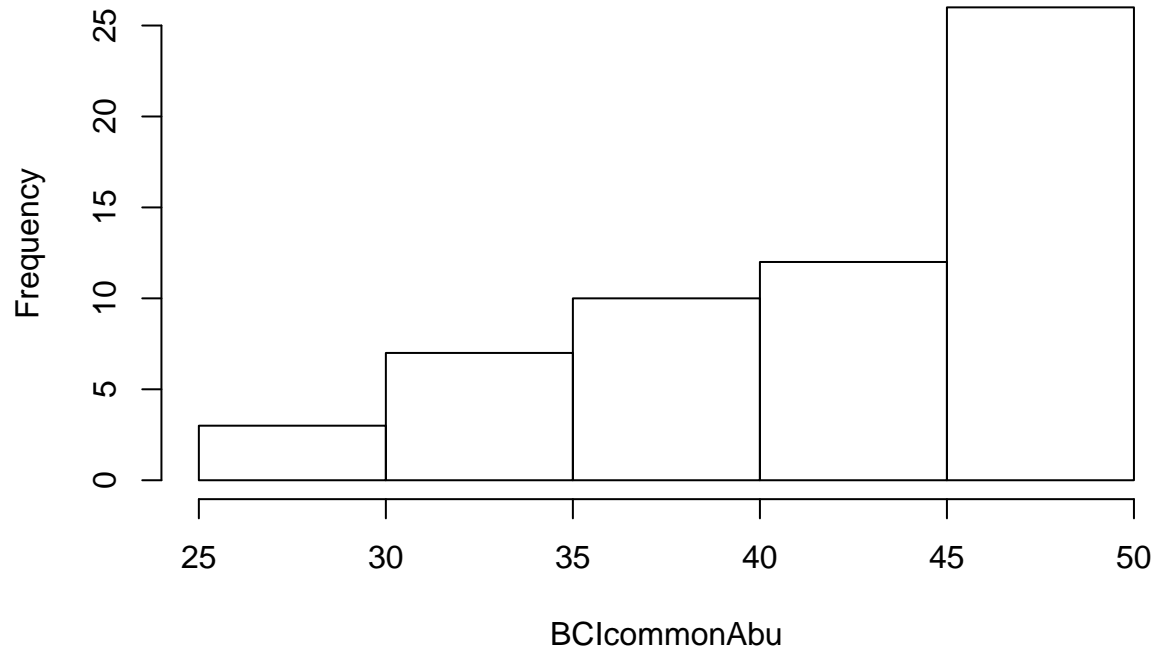


#### The rare species I choose from the BCIrare Dataset is *Tetrathylacium johansenii*. As you can see from the solid red line in the graph above, there is evidence of over dispersal of this species. This is determined from the fact that the solid red line has a downward slope.

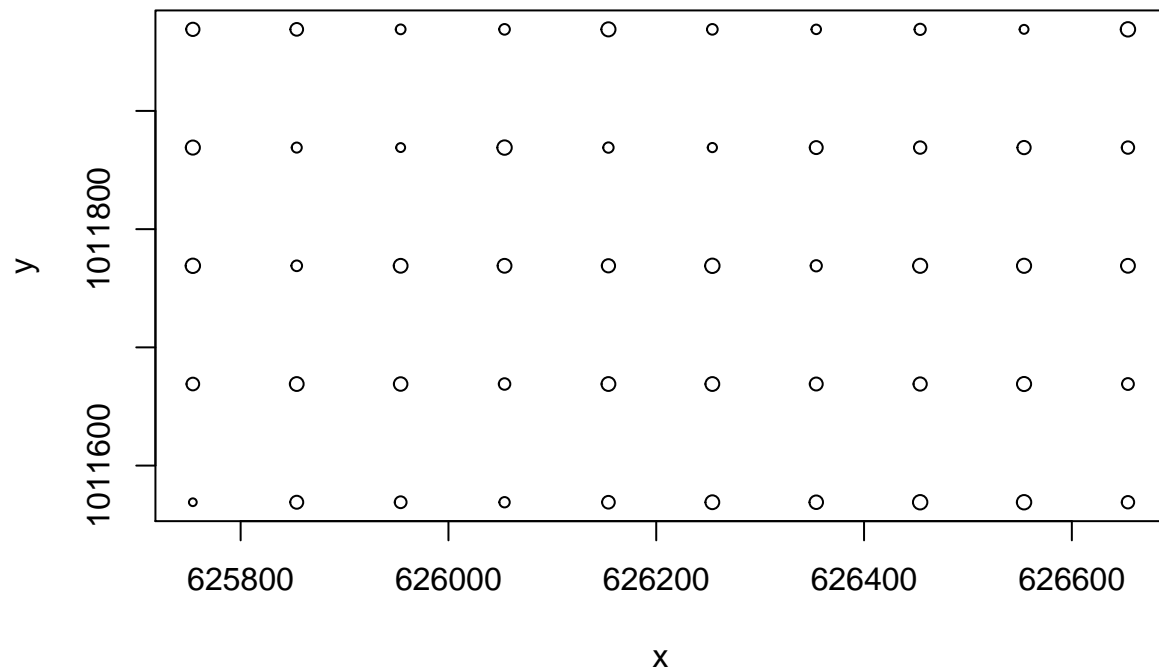
### Common BCI

```
BCIcommonAbu = apply(BCIcommon, 2, function(x) sum(x > 0))
hist(BCIcommonAbu)
```

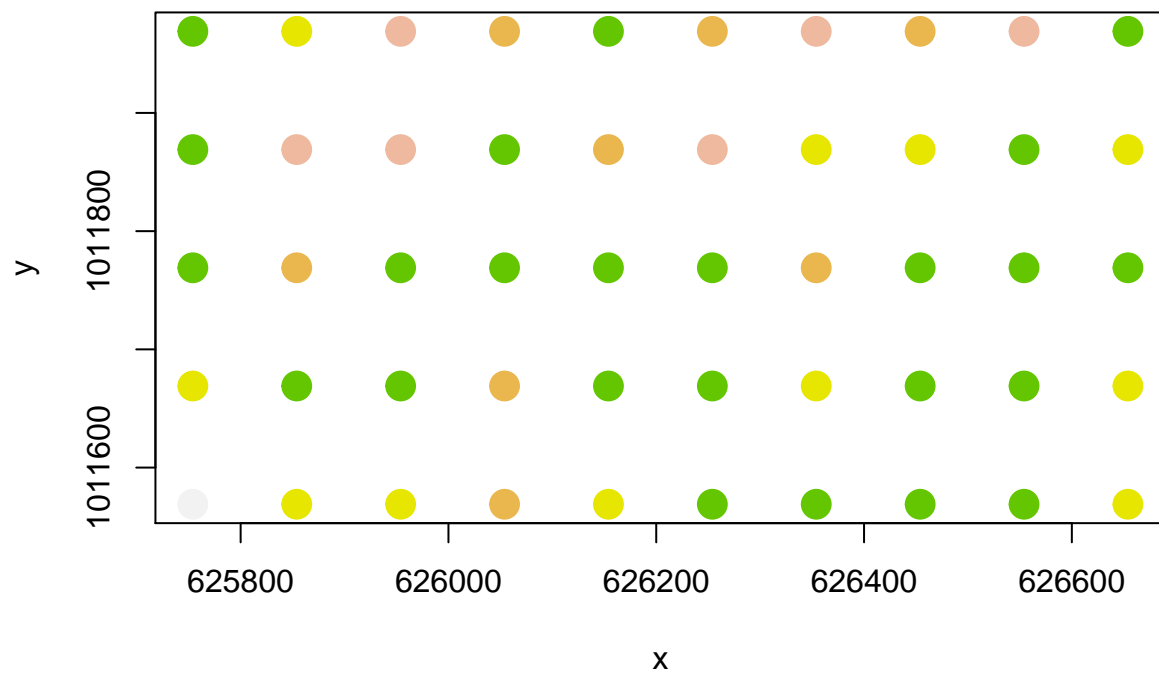
## Histogram of BCIcommonAbu



```
common_species = BCIcommon$Triplaris.cumingiana  
plot(BCI_xy, cex=BCIcommonAbu/max(BCIcommonAbu))
```



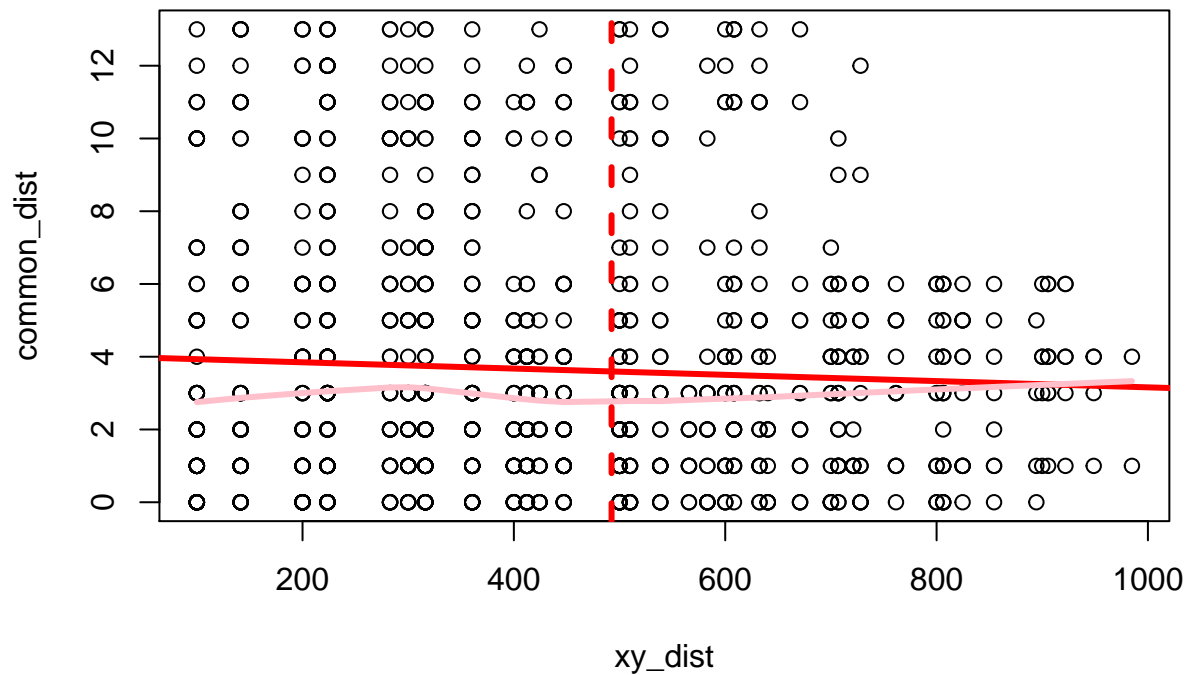
```
col_brks = hist(BCIcommonAbu, plot=F)$breaks
col_indices = as.numeric(cut(BCIcommonAbu, col_brks))
cols = rev(terrain.colors(length(col_brks)))
plot(BCI_xy, cex=2, pch=19, col=cols[col_indices])
```



```
xy_dist = dist(BCI_xy)
max_dist = max(xy_dist) / 2
common_dist = dist(common_species)

plot(xy_dist, common_dist)
abline(lm(common_dist ~ xy_dist), lwd=3, col='red')
lines(lowess(xy_dist, common_dist), lwd=3, col='pink')
abline(v = max_dist, col='red', lwd=3, lty=2)
```





#### The common species I choose to single out from the BCIcommon Dataset is *Triplaris cumingiana*. As you can see from the solid, mostly horizontal red line in the graph above, there is little to no evidence of spacial dependence in this common species.

## Question 2

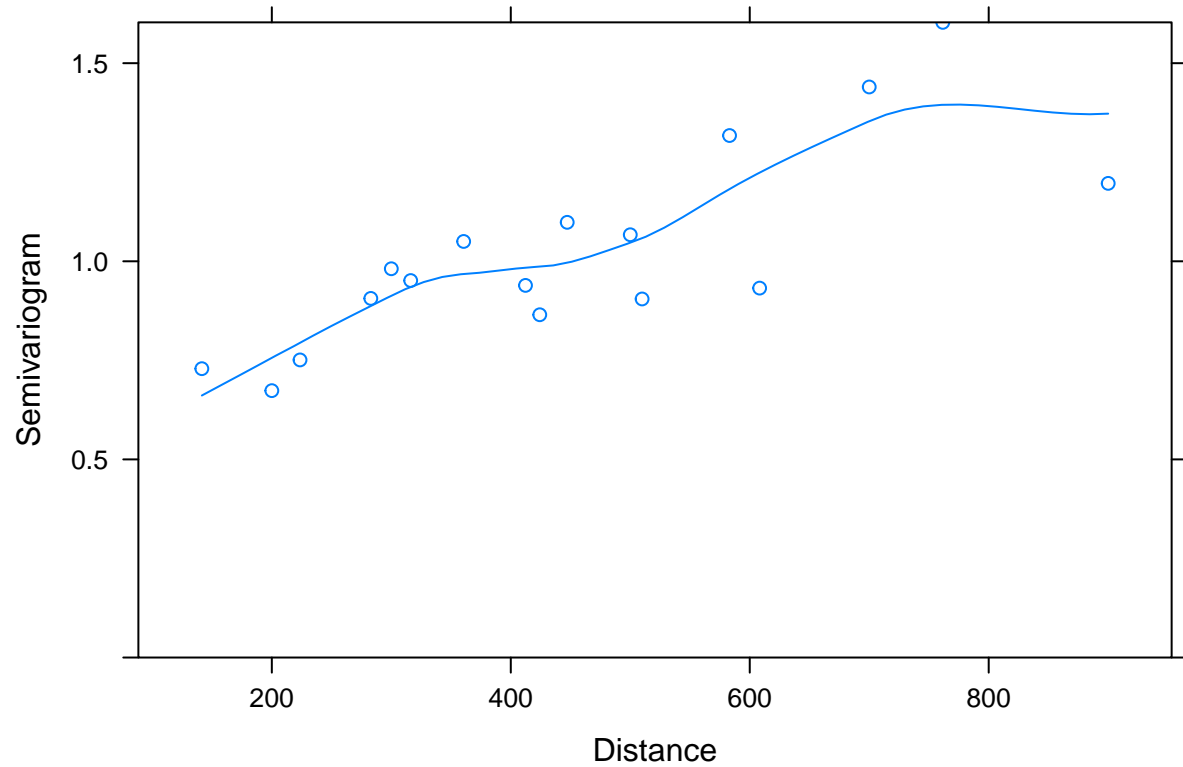
### Set Up

```
library(gstat)
library(nlme)
sp_ids = c("Drypetes.standleyi", "Cordia.lasiocalyx", "Hirtella.triandra",
           "Picramnia.latifolia", "Quassia.amara",
           "Tabernaemontana.arborea", "Trattinnickia.aspera",
           "Xylopia.macrantha")
```

### One Species Model

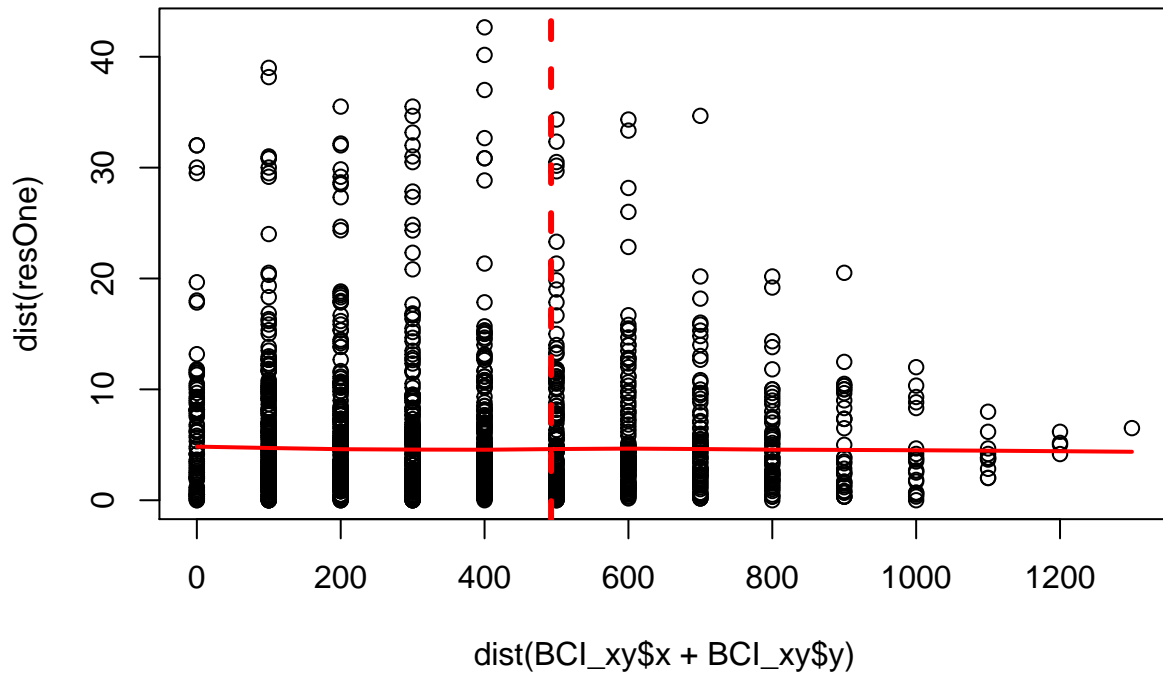
#### Base Model

```
lm_oneSpecies = gls(Drypetes.standleyi ~ Cordia.lasiocalyx, data = BCI[, sp_ids[1:2]])
plot(Variogram(lm_oneSpecies, form = ~ BCI_xy$x + BCI_xy$y))
```



### Residuals for One Species Model

```
resOne = residuals(lm_oneSpecies)
plot(dist(BCI_xy$x + BCI_xy$y), dist(resOne))
lines(lowess(dist(BCI_xy$x + BCI_xy$y), dist(resOne)), col='red', lwd=2)
abline(v = max_dist, col='red', lwd=3, lty=2)
```

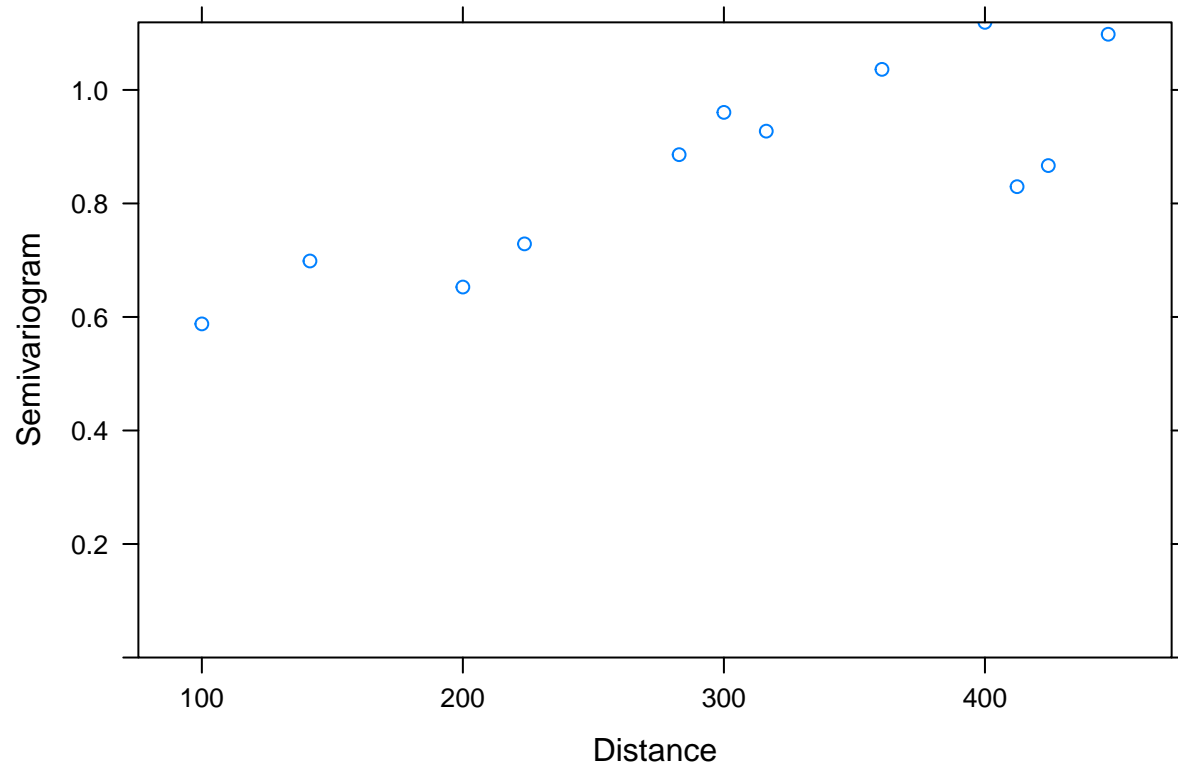


```
#### Update of One Species Model using Nuggets
```

```
oneSpecies_nug = update(lm_oneSpecies, corr=corExp(nugget=T))
oneSpecies_nug
```

```
## Generalized least squares fit by REML
## Model: Drypetes.standleyi ~ Cordia.lasiocalyx
## Data: BCI[, sp_ids[1:2]]
## Log-restricted-likelihood: -164.2575
##
## Coefficients:
## (Intercept) Cordia.lasiocalyx
## -2.357142 1.105904
##
## Correlation Structure: Exponential spatial correlation
## Formula: ~1
## Parameter estimate(s):
## range nugget
## 4.682511e-01 2.901272e-06
## Degrees of freedom: 50 total; 48 residual
## Residual standard error: 6.701558
```

```
plot(Variogram(oneSpecies_nug, form = ~ BCI_xy$x + BCI_xy$y, maxDist = max_dist))
```



### Coefficient Comparison

lm\_oneSpecies

```
## Generalized least squares fit by REML
## Model: Drypetes.standleyi ~ Cordia.lasiocalyx
## Data: BCI[, sp_ids[1:2]]
## Log-restricted-likelihood: -164.5623
##
## Coefficients:
## (Intercept) Cordia.lasiocalyx
## -2.782804 1.165220
##
## Degrees of freedom: 50 total; 48 residual
## Residual standard error: 6.675561
```

oneSpecies\_nug

```
## Generalized least squares fit by REML
## Model: Drypetes.standleyi ~ Cordia.lasiocalyx
## Data: BCI[, sp_ids[1:2]]
## Log-restricted-likelihood: -164.2575
##
## Coefficients:
## (Intercept) Cordia.lasiocalyx
## -2.357142 1.105904
##
## Correlation Structure: Exponential spatial correlation
```

```
## Formula: ~1
## Parameter estimate(s):
##      range      nugget
## 4.682511e-01 2.901272e-06
## Degrees of freedom: 50 total; 48 residual
## Residual standard error: 6.701558
```

The coefficients for both the intercept and the species are lower for the model that includes nuggets than the base model. A higher coefficient means that the predictor variable has a greater impact on the dependant variable so the model without the nuggets and with the greater coefficients would be the better model.

### ANOVA Comparison of One Species Models

```
anova(lm_oneSpecies, oneSpecies_nug, test = F)
```

	Model	df	AIC	BIC	logLik
## lm_oneSpecies	1	3	335.1246	340.7382	-164.5623
## oneSpecies_nug	2	5	338.5150	347.8710	-164.2575

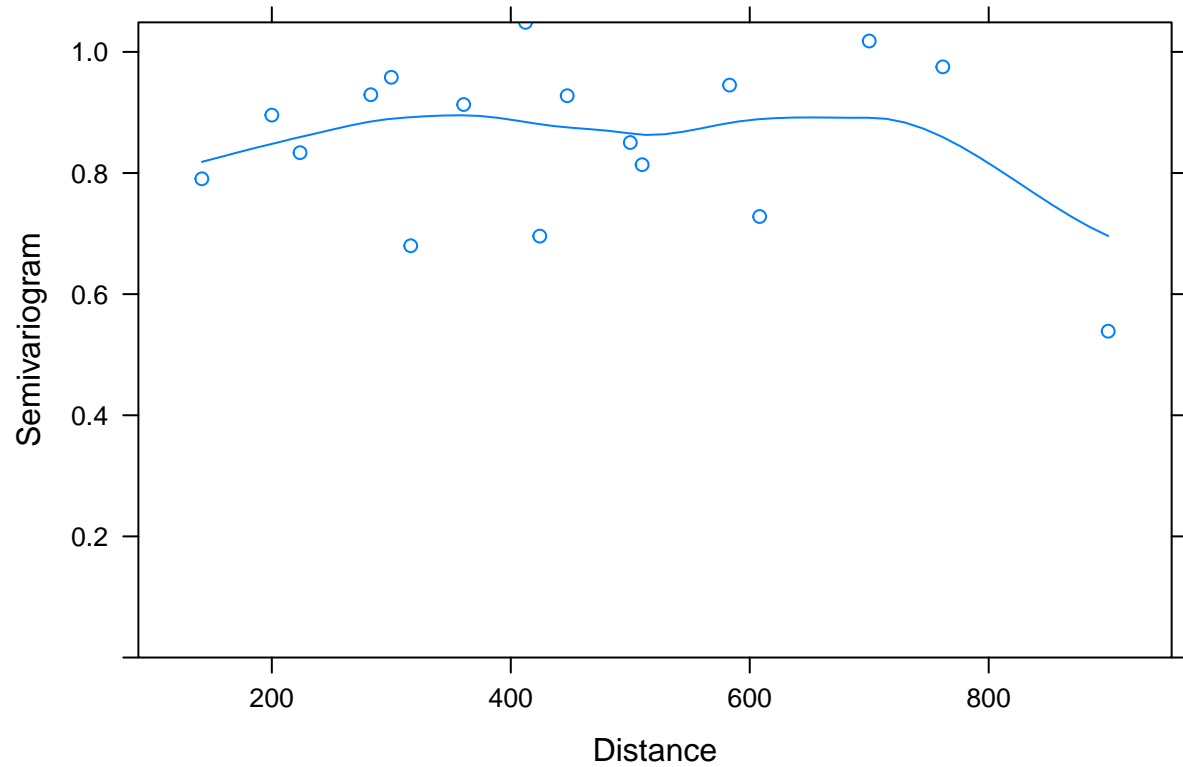
### Discussion

As you can see from the ANOVA results, the One Species GLS Base Model did better than the updated GLS Model using nuggets. Both the AIC and BIC is greater for the model that included the nuggets compared to the model without the nuggets which suggests that the base model is better since it has a lower AIC and BIC score.

### All Species Model

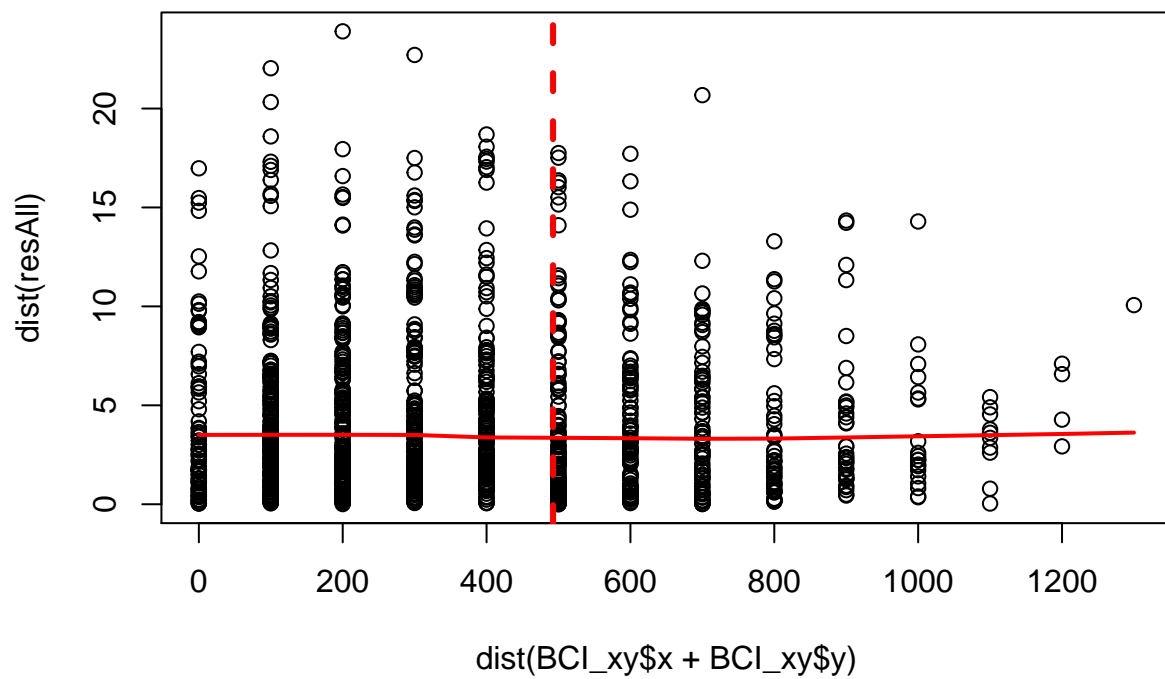
#### Base Model

```
lm_allSpecies = gls(Drypetes.standleyi ~ Cordia.lasiocalyx + Hirtella.triandra + Picramnia.latifolia + (1|BCI_xy), data = BCI_xy)
plot(Variogram(lm_allSpecies, form = ~ BCI_xy$x + BCI_xy$y))
```



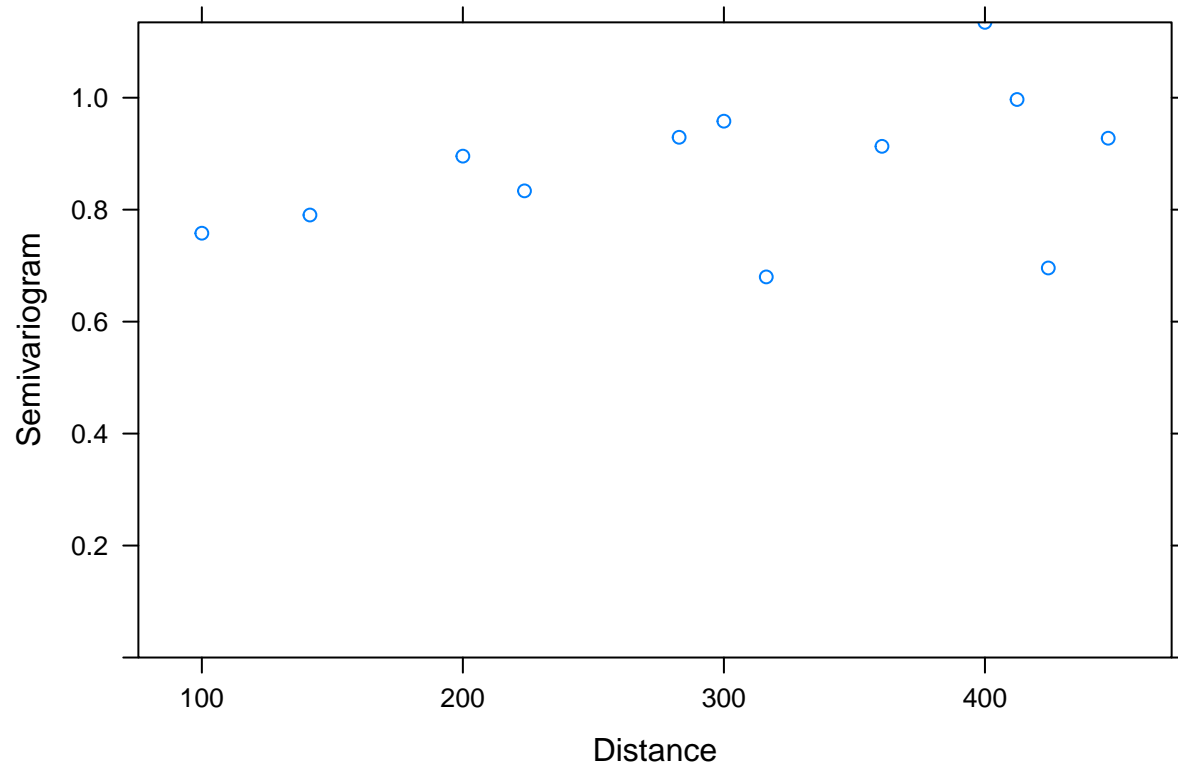
#### Residuals for All Species Model

```
resAll = residuals(lm_allSpecies)
plot(dist(BCI_xy$x + BCI_xy$y), dist(resAll))
lines(lowess(dist(BCI_xy$x + BCI_xy$y), dist(resAll)), col='red', lwd=2)
abline(v = max_dist, col='red', lwd=3, lty=2)
```



Update of All Species Model using Nuggets

```
allSpecies_nug = update(lm_allSpecies, corr=corExp(nugget=T))
plot(Variogram(allSpecies_nug, form = ~ BCI_xy$x + BCI_xy$y, maxDist = max_dist))
```



### Coefficient Comparison

lm\_allSpecies

## Generalized least squares fit by REML

## Model: Drypetes.standleyi ~ Cordia.lasiocalyx + Hirtella.triandra + Picramnia.latifolia + Quassia.amara

## Data: BCI[, sp\_ids]

## Log-restricted-likelihood: -144.5582

##

## Coefficients:

## (Intercept) Cordia.lasiocalyx Hirtella.triandra

## -1.0517523 0.4289202 0.1222789

## Picramnia.latifolia Quassia.amara Tabernaemontana.arborea

## 0.6622595 4.0856614 -0.2497250

## Trattinnickia.aspera Xylopia.macrantha

## 1.3493226 0.5488320

##

## Degrees of freedom: 50 total; 42 residual

## Residual standard error: 4.539713

allSpecies\_nug

## Generalized least squares fit by REML

## Model: Drypetes.standleyi ~ Cordia.lasiocalyx + Hirtella.triandra + Picramnia.latifolia + Quassia.amara

## Data: BCI[, sp\_ids]

## Log-restricted-likelihood: -144.5582

##

## Coefficients:



```
##           (Intercept)          Cordia.lasiocalyx          Hirtella.triandra
##           -1.0517526              0.4289202              0.1222789
##   Picramnia.latifolia          Quassia.amara Tabernaemontana.arborea
##           0.6622596              4.0856610              -0.2497250
##   Trattinnickia.aspera          Xylopia.macrantha
##           1.3493227              0.5488319
##
## Correlation Structure: Exponential spatial correlation
## Formula: ~1
## Parameter estimate(s):
##   range      nugget
## 0.06294492 0.14610588
## Degrees of freedom: 50 total; 42 residual
## Residual standard error: 4.539713
```

The coefficients show to be equal between the two models for all variables included so the spatial error term did not have any impact what so ever on the coefficients.

### ANOVA Comparison of All Species Models

```
anova(lm_allSpecies, allSpecies_nug, test = F)
```

```
##           Model df      AIC      BIC    logLik
## lm_allSpecies    1  9 307.1163 322.7554 -144.5582
## allSpecies_nug    2 11 311.1163 330.2307 -144.5582
```

### Discussion

From the ANOVA we can see that the GLS base model performed better than the updated model using nuggets. Similarly to the One Species Model compariosn AVOVA, both the AIC and BIC scores are lower for the base model suggesting that it will perform better than the model including nuggets.

### Adding of Spacial Data Discussion

For both models, one using only one species as a predictor and the second using a subset of species as predictors, the inclusion of spacial data did not seem to positively effect the outcome of the model. In both cases the ANOVA showed that the base model that did not incluide spacial data performed better. One reason for the spacial model performing worse could be that there was not a strong enough correlation between abundance of a species and its spacial location in the grid for it to be of any help in predicting the dependent species variable. Space/location may simply not play a role for the certain trees ability to grow/abundance.