

# spatial\_assignment

*Julius Alipala*

*February 18, 2019*

## Spatial Modeling Assignment

```
library(vegan)
```

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

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

```
## This is vegan 2.5-4
```

```
library(nlme)
```

```
data(BCI)
```

```
## 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))
```

- 1) Examine if there is evidence of spatial dependence in a rare and a common species in the BCI tree dataset

```
# count of species
```

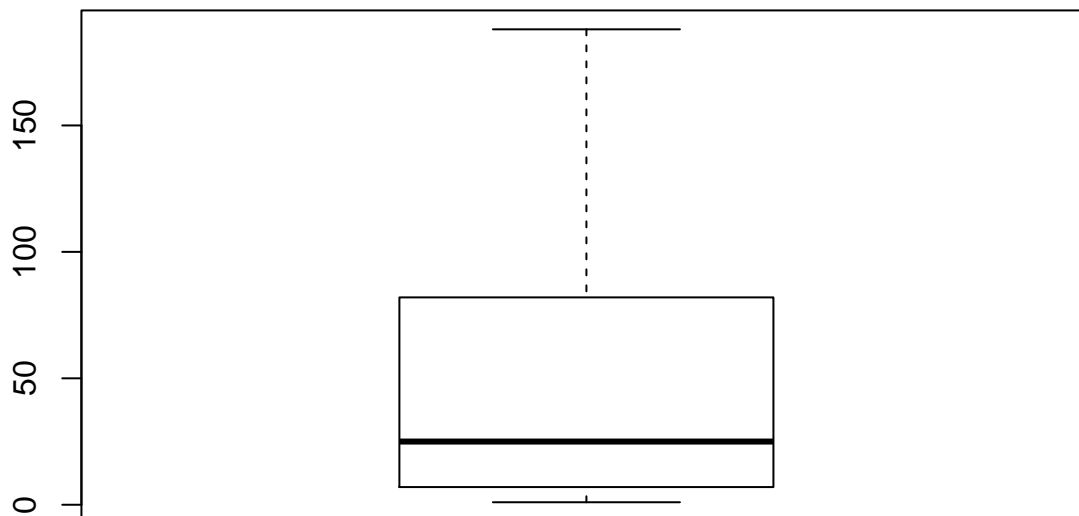
```
sp_count = colSums(BCI, na.rm=FALSE, dims=1)
```

```
summary(sp_count)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
##      1.00   7.00   25.00   95.36   82.00  1717.00
```

```
# boxplot of sp_count without outliers
```

```
boxplot(sp_count, outline = FALSE)
```



```
# rare species: <20
```

```
# return index of species with counts between 15-20
```

```
rare_sp = which(sp_count < 20 & sp_count > 15)
```

```
rare_sp
```

```
## Cavanillesia.platanifolia   Diospyros.artanthifolia
##                          33                          60
## Erythroxylum.macrophyllum   Cinnamomum.triplinerve
##                          66                          151
##      Sapium.glandulosum      Siparuna.pauciflora
##                          174                          179
```

```
# chose Cavanillesia platanifoli to investigate further
# index of Cavanillesia platanifoli
```

```
cav_plat_index = 33
sprintf("count of Cavanillesia platanifoli: %i", sp_count[cav_plat_index])
```

```
## [1] "count of Cavanillesia platanifoli: 19"
```

```
# common species: > 150
# return index of species with counts between 150-200
comm_sp = which(sp_count > 150 & sp_count < 200)
comm_sp
```

```
##      Alchornea.costaricensis      Brosimum.alicastrum
##                          7                          24
##      Eugenia.oerstediana Maquira.guianensis.costaricana
##                          70                          125
##      Ocotea.whitei      Unonopsis.pittieri
##                          143                          214
##      Virola.surinamensis
##                          217
```

```
# chose Alchornea costaricensis to investigate further
# index of Alchornea costaricensis
```

```
alc_cos_index = 7
sprintf("count of Alchornea costaricensis: %i", sp_count[alc_cos_index])
```

```
## [1] "count of Alchornea costaricensis: 156"
```

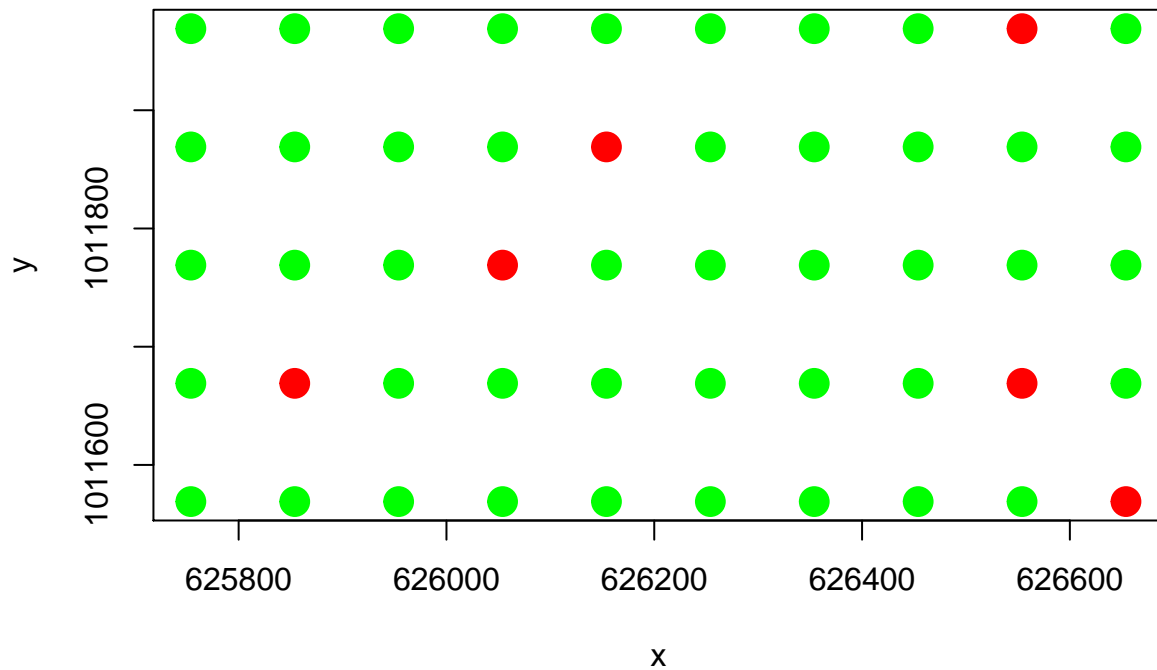
Based on the results above, I've chosen Cavanillesia platanifoli to represent my rare species and Alchornea costaricensis to represent my common species.

```
# index of sites which contain Alchornea costaricensis
```

```
alc_cos_sites = which(BCI[alc_cos_index]>0)
```

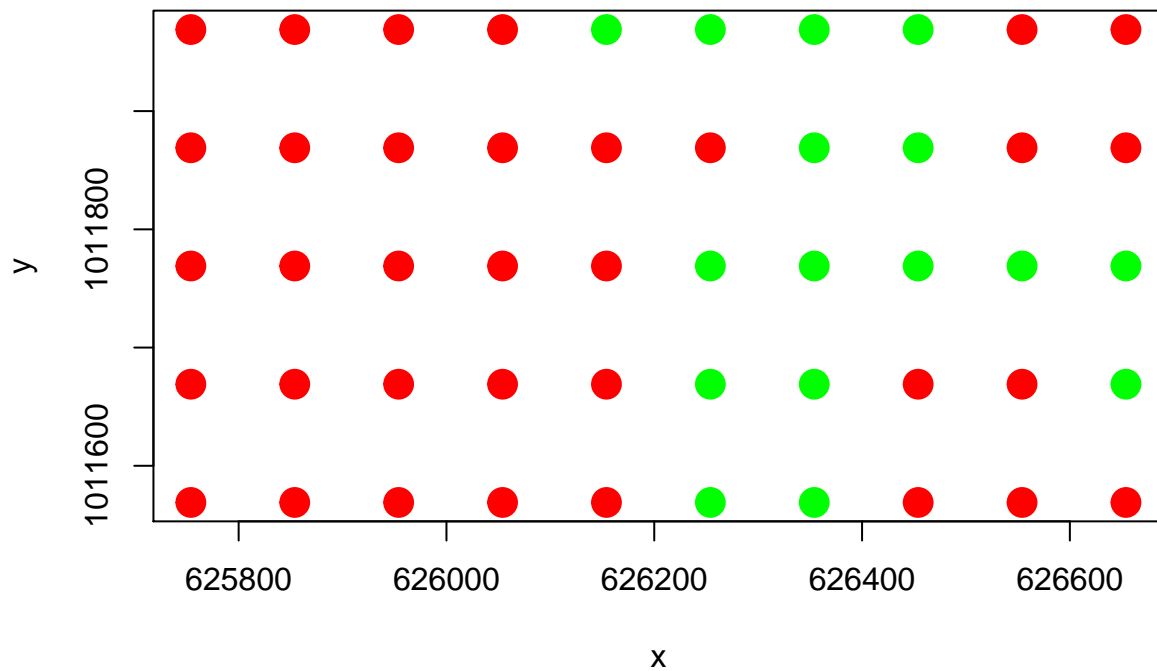
```
# sites that have Alchornea costaricensis colored green; otherwise colored red
```

```
plot(BCI_xy, cex=2, pch=19, col=ifelse(rownames(BCI_xy) %in% alc_cos_sites, 'green', 'red'))
```



The graph above shows sites which contain *Alchornea costaricensis* in green. Sites without *Alchornea costaricensis* are colored red. Based on this graph, there is no evidence of spatial dependence for the common species, *Alchornea costaricensis*.

```
# index of sites which contain Cavanillesia platanifoli
cav_plat_sites = which(BCI[cav_plat_index]>0)
# sites that have Cavanillesia platanifoli colored green; otherwise colored red
plot(BCI_xy, cex=2, pch=19, col=ifelse(rownames(BCI_xy) %in% cav_plat_sites, 'green', 'red'))
```



The graph above shows evidence of spatial dependence for the rare species, *Cavanillesia platanifoli*. The sites that contain this species are grouped in a certain area and are all adjacent to each other. No solitary sites contain the species.

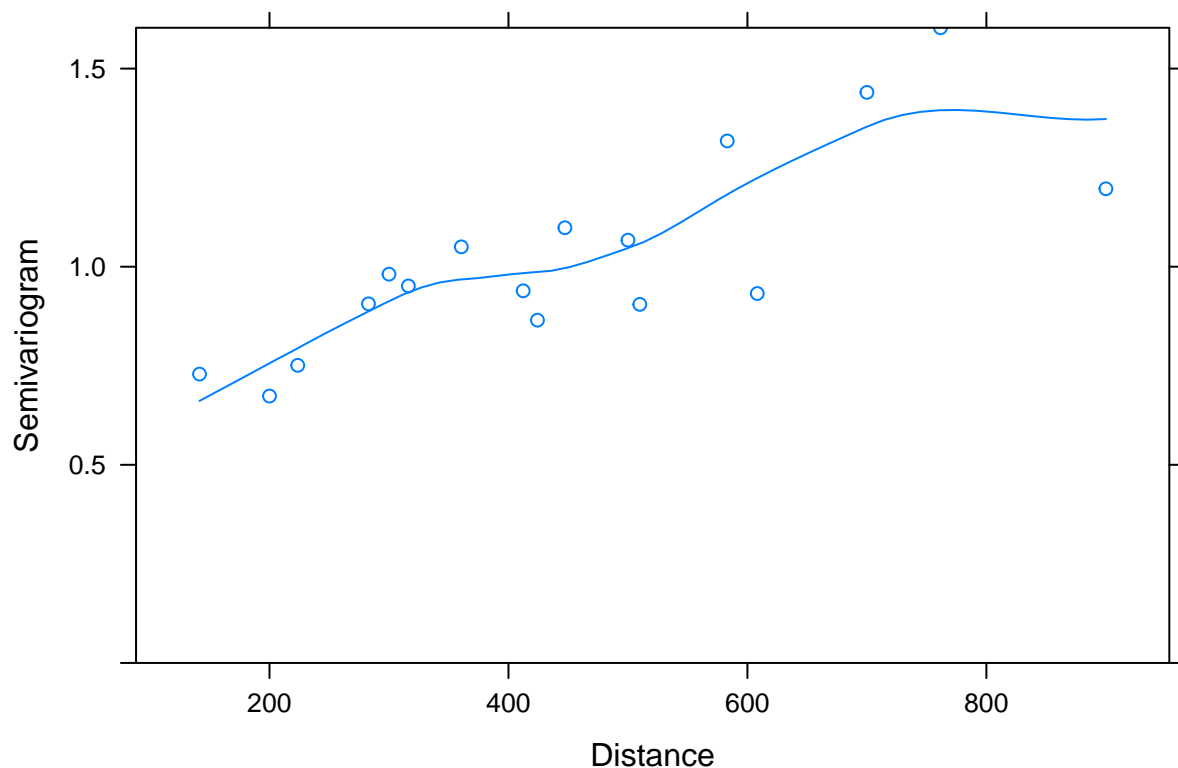
- 2) Build two generalized linear models to predict the abundance of the species *Drypetes standleyi* using the abundance of other tree species in the study site. Specifically examine the following species as predictor variables:

```
sp_ids = c("Cordia.lasiocalyx", "Hirtella.triandra",
           "Picramnia.latifolia", "Quassia.amara",
           "Tabernaemontana.arborea", "Trattinnickia.aspera",
           "Xylopia.macrantha")
x = BCI_xy$x
y = BCI_xy$y
```

Note renaming the species ids to something a little easier to work with like “sp\_a”, “sp\_b” will make model construction a little less cumbersome

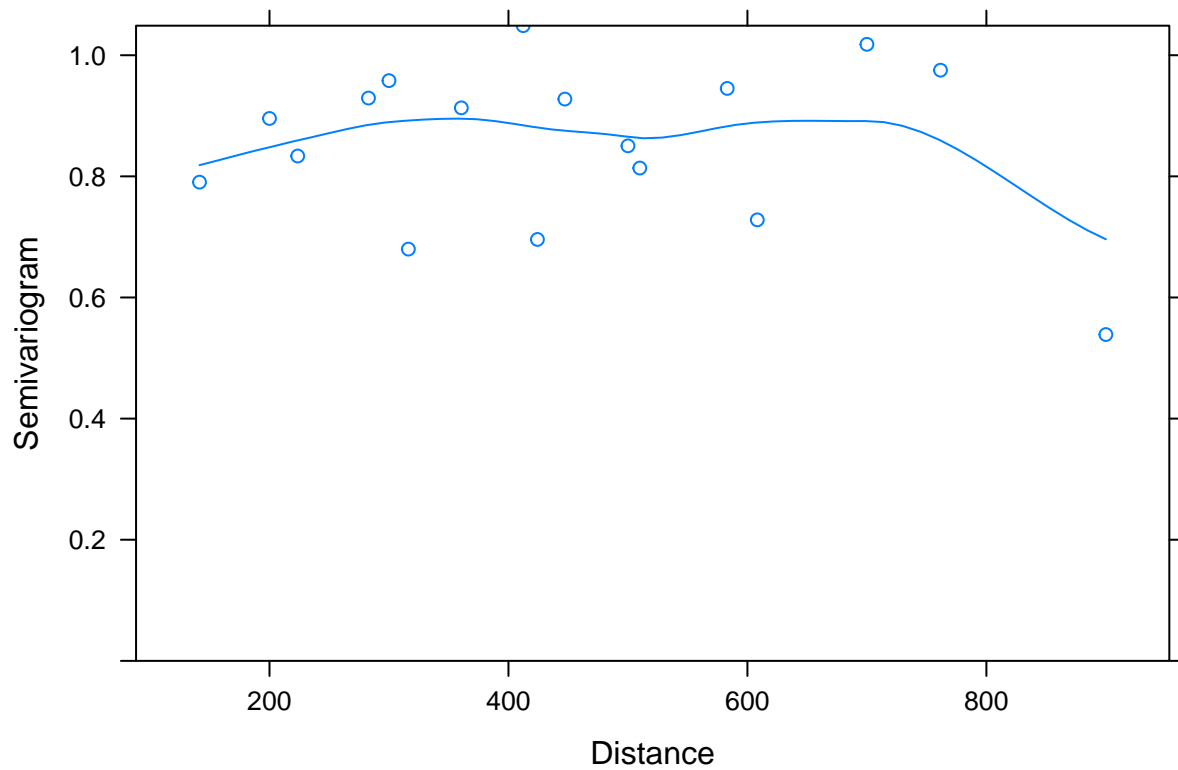
- Model 1: only include a single species as a predictor variable

```
ds_lm_single = gls(Drypetes.standleyi ~ Cordia.lasiocalyx, data=BCI)
par(mfrow=c(1,1))
plot(Variogram(ds_lm_single, form = ~ x + y))
```



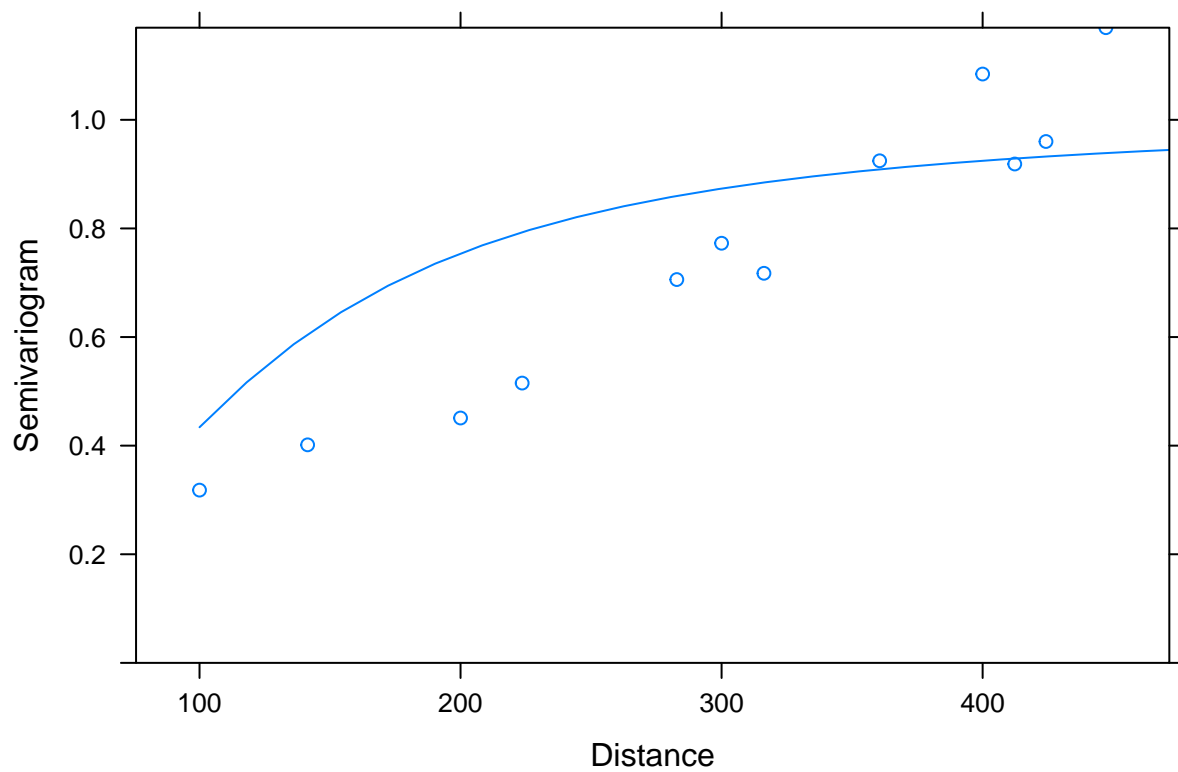
- Model 2: include all of the species as predictor variables

```
ds_lm_all = gls(Drypetes.standleyi ~ Cordia.lasiocalyx + Hirtella.triandra + Picramnia.latifolia +
               Quassia.amara + Tabernaemontana.arborea + Trattinnickia.aspera +
               Xylopia.macrantha, data=BCI)
par(mfrow=c(1,1))
plot(Variogram(ds_lm_all, form = ~ x + y))
```

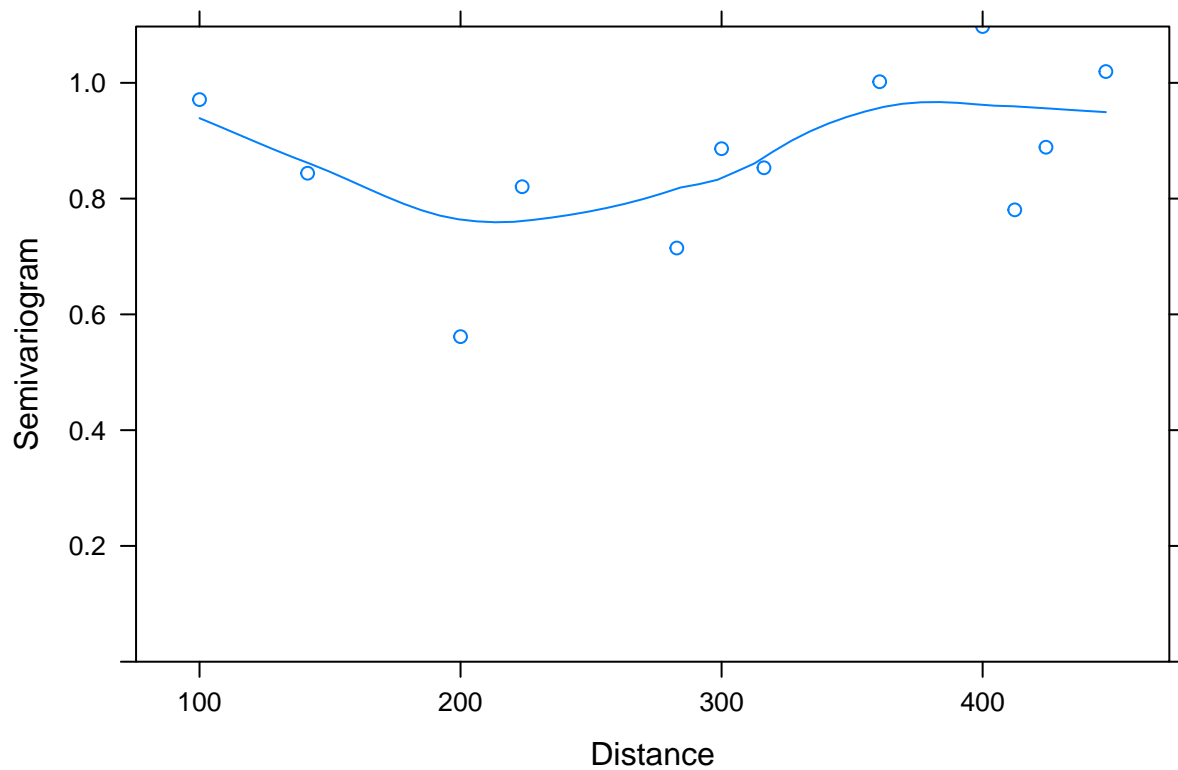


With both models examine the spatial dependence of the residuals using the function `Variogram`. Model the spatial dependence in the residuals using one of the error structures available.

```
xy_dist = dist(BCI_xy)
max_dist = max(xy_dist)/2
# rational quadratic
ds_single_rat = update(ds_lm_single, corr=corRatio(form=~x+y))
plot(Variogram(ds_single_rat, maxDist = max_dist))
```



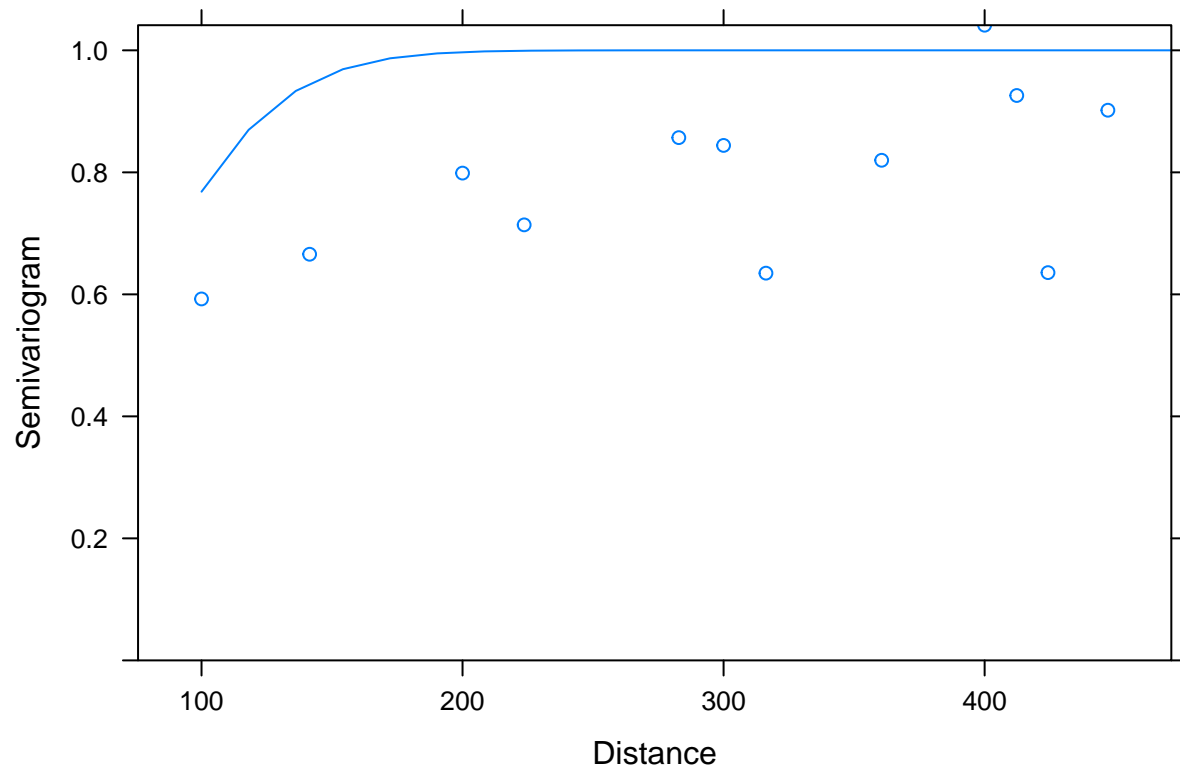
```
plot(Variogram(ds_single_rat, resType='normalized', maxDist = max_dist))
```



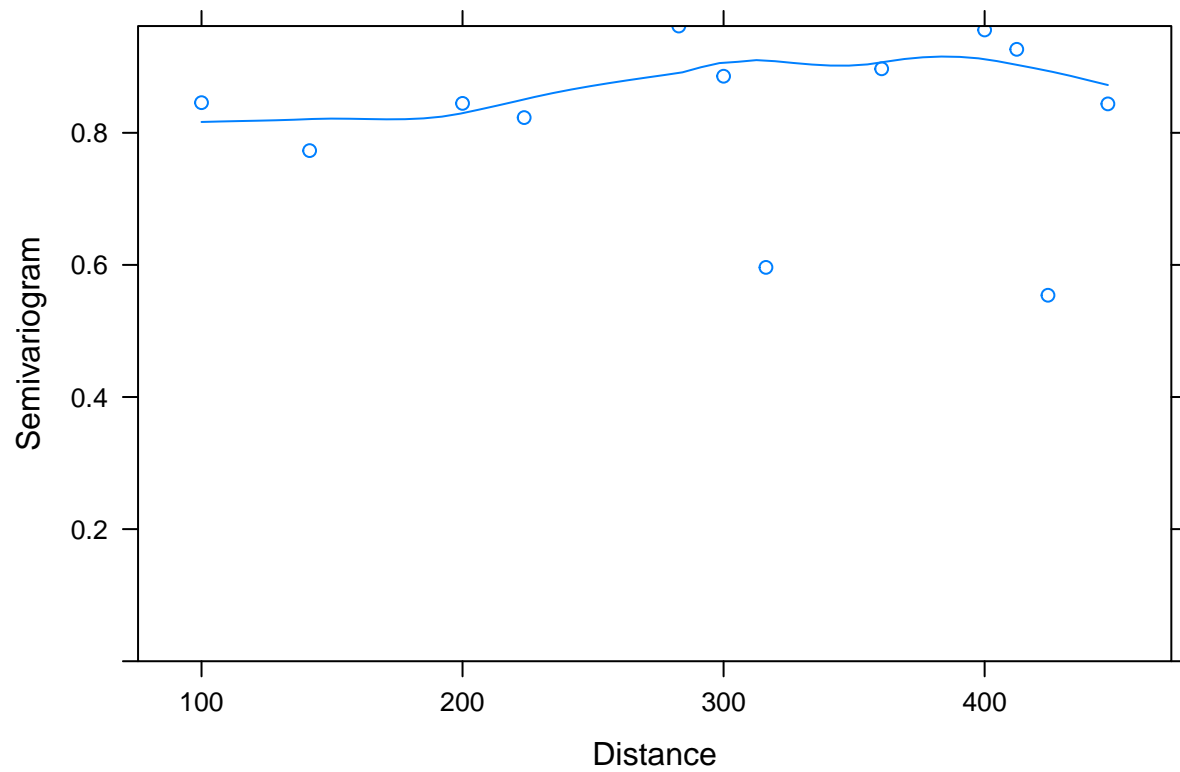
The rational quadratic model does not seem to fit very well, but the plot of residuals shows a better fit.

```
# gaussian
ds_all_gauss = update(ds_lm_all, cor=corGaus(form=-x+y))
```

```
plot(Variogram(ds_all_gauss, maxDist = max_dist))
```



```
plot(Variogram(ds_all_gauss, resType='normalized', maxDist=max_dist))
```



The Gaussian model does not seem to fit well at all. I would expect the plot for residuals to drop a little

more as the distance increases.

- Did including the spatial error term have a large impact on the coefficients of the model?

```
summary(ds_lm_single)$coefficients
```

```
##      (Intercept) Cordia.lasiocalyx
##      -2.782804      1.165220
```

```
summary(ds_single_rat)$coefficients
```

```
##      (Intercept) Cordia.lasiocalyx
##      4.970422      0.312505
```

Spatial error term has a large impact on single species predictor.

```
summary(ds_lm_all)$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
```

```
summary(ds_all_gauss)$coefficients
```

```
##      (Intercept)      Cordia.lasiocalyx      Hirtella.triandra
##      -1.0108518      0.2868254      0.1280075
##      Picramnia.latifolia      Quassia.amara Tabernaemontana.arborea
##      0.6787089      3.4613563      -0.1014413
##      Trattinnickia.aspera      Xylopia.macrantha
##      1.4863690      0.5060473
```

In general, spatial error term did not change the coefficients much for the models using 7 species as a predictor.

- Did including the spatial error terms significantly improve model fit (use function `anova` to carry out model comparison)?

```
# compare single species models
```

```
anova(ds_lm_single, ds_single_rat)
```

```
##      Model df      AIC      BIC    logLik    Test  L.Ratio p-value
## ds_lm_single      1  3 335.1246 340.7382 -164.5623
## ds_single_rat      2  4 318.9170 326.4018 -155.4585 1 vs 2 18.20762 <.0001
```

Spatial error terms improved the model for the single species as a predictor. I expected the model without spatial error terms to be a better fit from looking at the plots.

```
# compare all species models
```

```
anova(ds_lm_all, ds_all_gauss)
```

```
##      Model df      AIC      BIC    logLik    Test  L.Ratio p-value
## ds_lm_all      1  9 307.1163 322.7554 -144.5582
## ds_all_gauss      2 10 307.2070 324.5837 -143.6035 1 vs 2 1.90935 0.167
```

No significant improvements were made by including spatial error terms for the model with 7 species as a predictor.

- Explain why you did or did not observe a difference in the influence of adding the spatial error term between the two models.



For the model with a single species as a predictor, including spatial error terms showed a significant improvement in fit. Since only one species is used as a predictor, that species' features are the only variables driving the prediction. If the predictor species is very different from the target species, this may result in a large difference between predicted and actual data.

Using multiple species as a predictor may provide a more generalized prediction. This is why including spatial error terms has little impact on the model using multiple species as predictors.