HW5_MargaretWalker

Margaret Walker

February 21, 2016

Spatial Modeling Assignment

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

```
library(vegan)
```

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

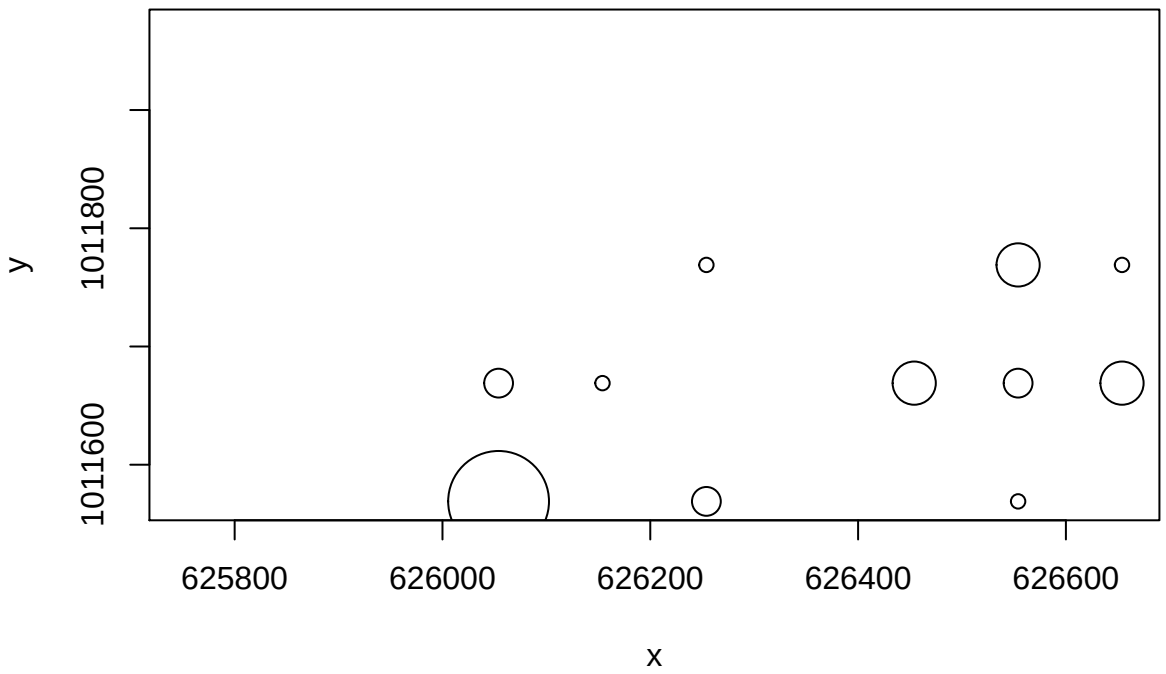
```
## Warning: package 'permute' was built under R version 3.1.3
```

```
## Loading required package: lattice
## This is vegan 2.2-1
```

```
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))
bci_dat <- data.frame(BCI_xy, BCI)
max_dist = max(BCI_xy) / 2
```

I chose *Inga acuminata* as my rare species because there were only 26 individuals for that species. First, I plotted the abundance of *Inga acuminata*. It looks like there are only a few areas with inga. I then used the mantel test to test whether there is spatial dependence in the data. The mantel test returned a test statistic of 0.01753 and a significance of 0.414. Due to the large significance value we can conclude that there is no relationship between the spatial distance and the abundance of *Inga acuminata*. That is, there is no spatial dependence for the abundance of *Inga acuminata*. See code below:

```
plot(y~x, cex=Inga.acuminata, data=bci_dat)
```

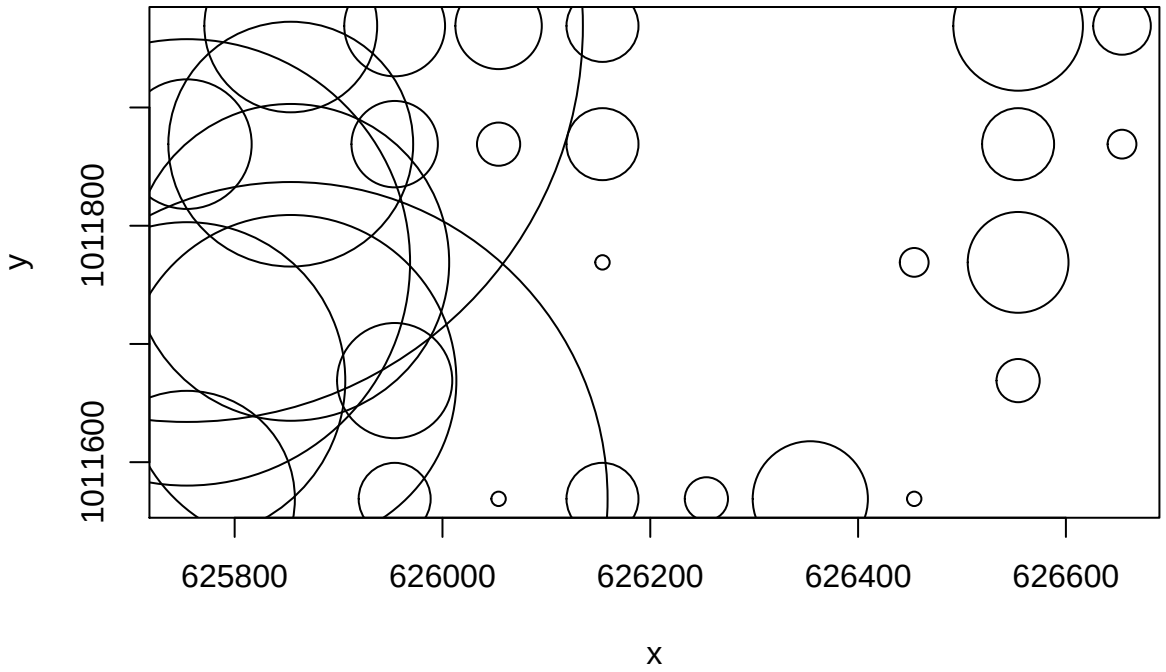



```
inga <- subset(BCI, select=Inga.acuminata)
inga_dist <- dist(inga)
xydist <- dist(BCI_xy)
inga_mantel <- mantel(xydist, inga_dist)
inga_mantel
```

```
##
## Mantel statistic based on Pearson's product-moment correlation
##
## Call:
## mantel(xdis = xydist, ydis = inga_dist)
##
## Mantel statistic r: 0.01753
##       Significance: 0.367
##
## Upper quantiles of permutations (null model):
##    90%    95%  97.5%    99%
## 0.0917 0.1151 0.1384 0.1521
## Permutation: free
## Number of permutations: 999
```

For my common species I chose *Socratea exorrhiza* because there were 346 individuals. Once again I started out by plotting the abundance of *Socratea exorrhiza*. The plot is kind of hard to interpret because of the size of the circles but it appears that all of the large circles are very close to one another. Next, I ran the mantel test on socratea. The mantel statistics came out to be 0.337 and the significance value was 0.001. Due to the significance value we can conclude that there is a significant relationship between the spatial distance and the abundance of *Socratea exorrhiza*. That is, there is spatial dependence in the abundance of *Socratea exorrhiza*.

```
plot(y~x, cex=Socratea.exorrhiza, data=bci_dat)
```




```
socratea <- subset(BCI, select=Socratea.exorrhiza)
socratea_dist <- dist(socratea)
socratea_mantel <- mantel(xydist, socratea_dist)
socratea_mantel
```

```
##
## Mantel statistic based on Pearson's product-moment correlation
##
## Call:
## mantel(xdis = xydist, ydis = socratea_dist)
##
## Mantel statistic r: 0.337
##       Significance: 0.001
##
## Upper quantiles of permutations (null model):
##   90%   95% 97.5%   99%
## 0.085 0.107 0.128 0.150
## Permutation: free
## Number of permutations: 999
```

So, it appears that there is no spatial dependence for the rare species but there is for the common species. This makes some sense becuase with the rare species there are fewer sites with that species compared to the common species. So, there are less chances for nearby sites to be similar.

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", "Drypetes.standleyi")
bci_sp <- subset(bci_dat, select=sp_ids)
```

- Model 1: only include a single species as a predictor variable
- Model 2: include all of the species as predictor variables

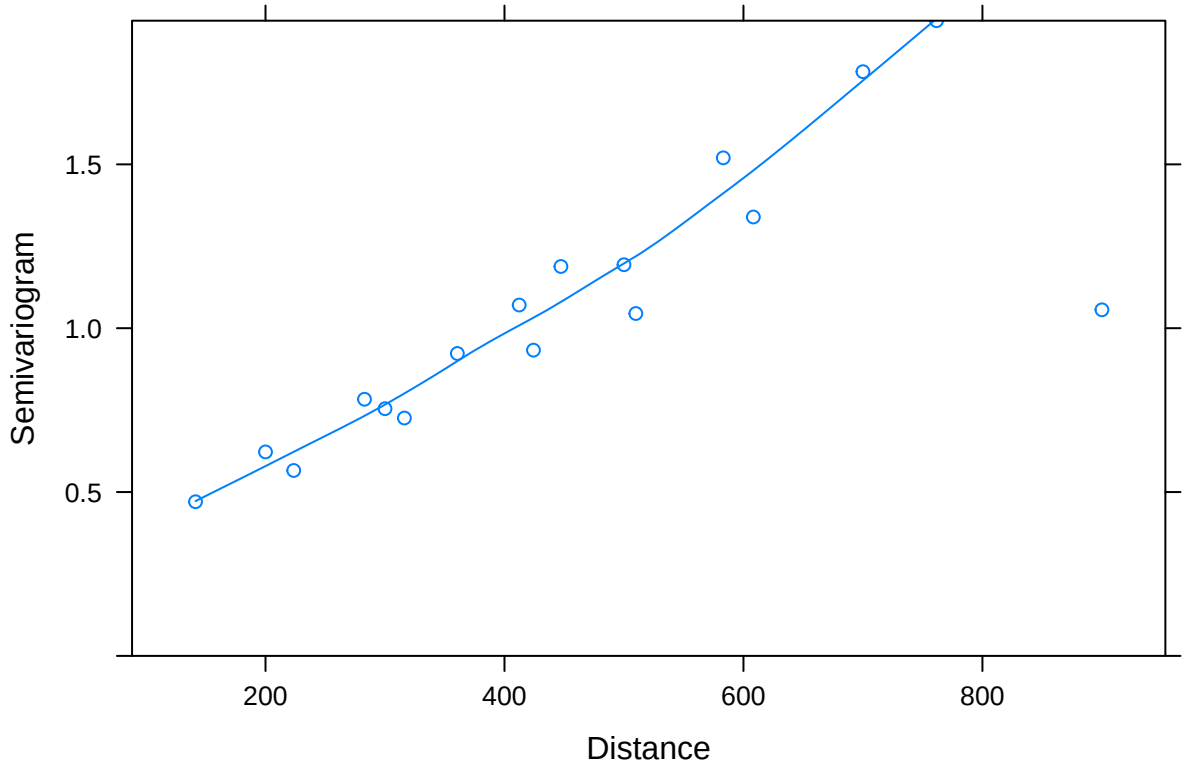
With both models examine the spatial dependence of the residuals using the function `Variogram`. Model the spatial dependence in the residuals using one of the error structures available.

Starting with the single species model. I randomly chose *Quassia amara* to use as my single species to model the abundance of *Drypetes standleyi*. So, first I started out by creating the gls model and the variogram of the model without using one of the error structures. There appears to be spatial dependence in the residuals since there is a linear relationship on the variogram. Furthermore, the normalized residuals have a pattern. If there was no spatial dependence we would expect no pattern on the variogram.

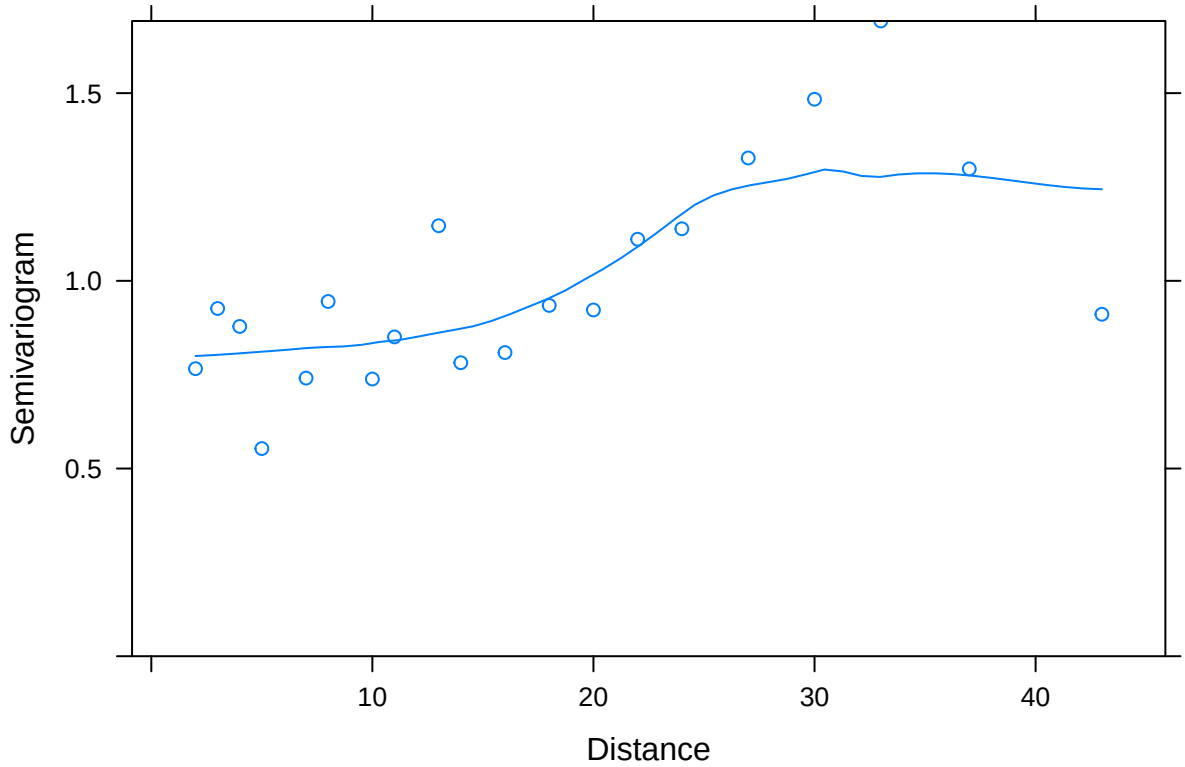
```
one_species <- gls(Drypetes.standleyi~Quassia.amara, data=bci_dat)
summary(one_species)
```

```
## Generalized least squares fit by REML
##   Model: Drypetes.standleyi ~ Quassia.amara
##   Data: bci_dat
##        AIC      BIC    logLik
##   337.5287 343.1423 -165.7644
##
## Coefficients:
##                  Value Std.Error  t-value p-value
## (Intercept)   4.962810  1.030804 4.814503   0e+00
## Quassia.amara 9.214876  2.304948 3.997867   2e-04
##
##  Correlation:
##               (Intr)
## Quassia.amara -0.179
##
## Standardized residuals:
##        Min         Q1        Med         Q3        Max
## -0.6920362 -0.6920362 -0.4131473  0.2840748  4.7462963
##
## Residual standard error: 7.171316
## Degrees of freedom: 50 total; 48 residual
```

```
plot(Variogram(one_species, form=~x+y))
```




```
plot(Variogram(one_species, resType='n'))
```



Since there appears to be spatial dependence I will try to use the exponential error model and see if that helps. The variogram of the normalized residuals look better, but not perfect. It does appear that modeling the error helps with the spatial dependence.

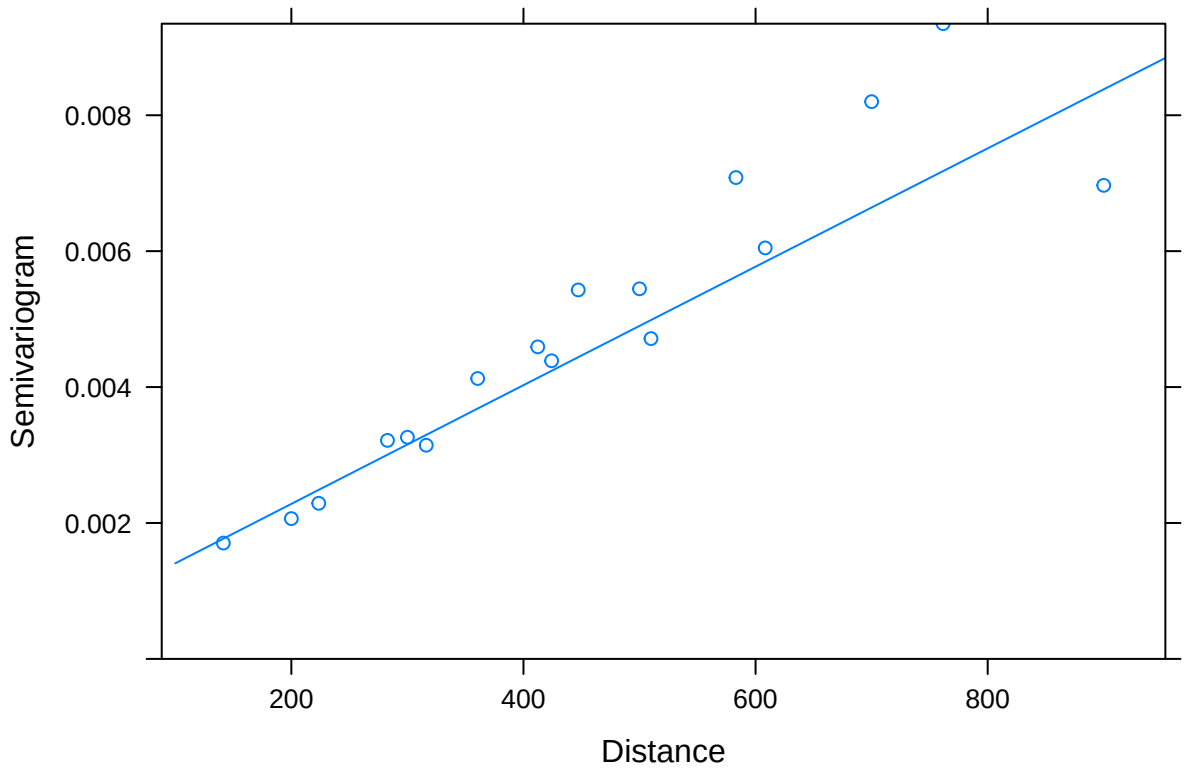
```
one_exp <-  gls(Drypetes.standleyi ~ Quassia.amara, data=bci_dat,
                corr=corExp(form=~x + y, nugget=T),
                control = glsControl(opt='optim',  msVerbose=T))
```

```
## initial  value 178.348772
## iter  10 value 170.860669
## final  value 170.839207
## converged
```

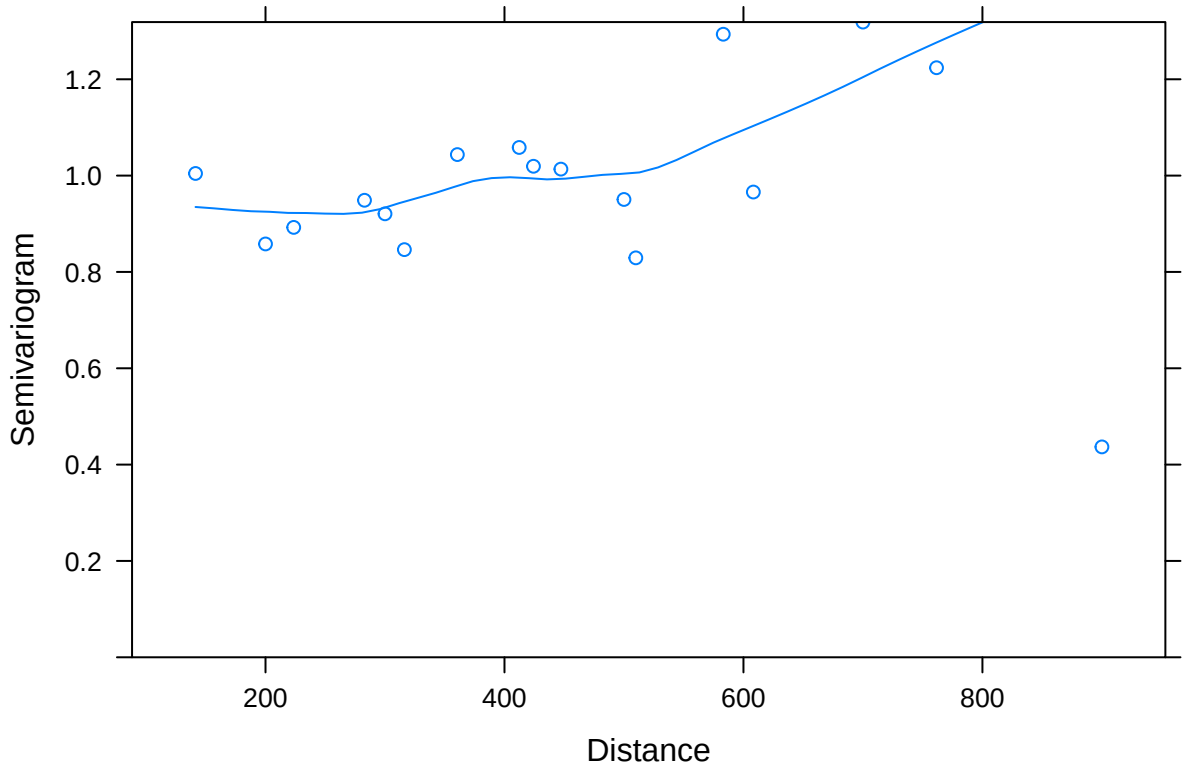
```
summary(one_exp)
```

```
## Generalized least squares fit by REML
##   Model: Drypetes.standleyi ~ Quassia.amara
##   Data: bci_dat
##        AIC      BIC    logLik
##   302.0789 311.4349 -146.0394
##
## Correlation Structure: Exponential spatial correlation
##  Formula: ~x + y
##  Parameter estimate(s):
##        range       nugget
## 1.141308e+05 5.309808e-04
##
## Coefficients:
##                   Value Std.Error   t-value p-value
## (Intercept)   10.161124 113.03730 0.0898918  0.9287
## Quassia.amara  3.690981   1.51215 2.4408904  0.0184
##
##  Correlation:
##               (Intr)
## Quassia.amara -0.009
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -0.08964134 -0.08964134 -0.07199736 -0.02127091  0.25441628
##
## Residual standard error: 113.3531
## Degrees of freedom: 50 total; 48 residual
```

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




```
plot(Variogram(one_exp, resType='n', maxDist = max_dist))
```



```
anova(one_species, one_exp)
```

```
##             Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## one_species     1  3 337.5287 343.1423 -165.7644
## one_exp         2  5 302.0789 311.4349 -146.0394 1 vs 2 39.44988  <.0001
```

Furthermore, when looking at the coefficients, the intercept and beta coefficients changed slightly, and the p-values of the intercept and beta coefficient changed significantly. The correlation coefficient went from -0.179 in the original model to -0.009 in the model with the error structure. This leads me to believe there is less of a correlation between the two variables when the spatial dependnece is removed. That is, most of the correlation before was due to the fact that closer sites are more related. Finally, when looking at the anova() function of the two models we see that the AIC value is significantly lower for the model with the error structure compared to the model with out. That is, the AIC value for the regular model is 337.5 while the model with the error structure is 302.1. So, it appears that the model with the error strucutre is a better overall fit than the model without it.

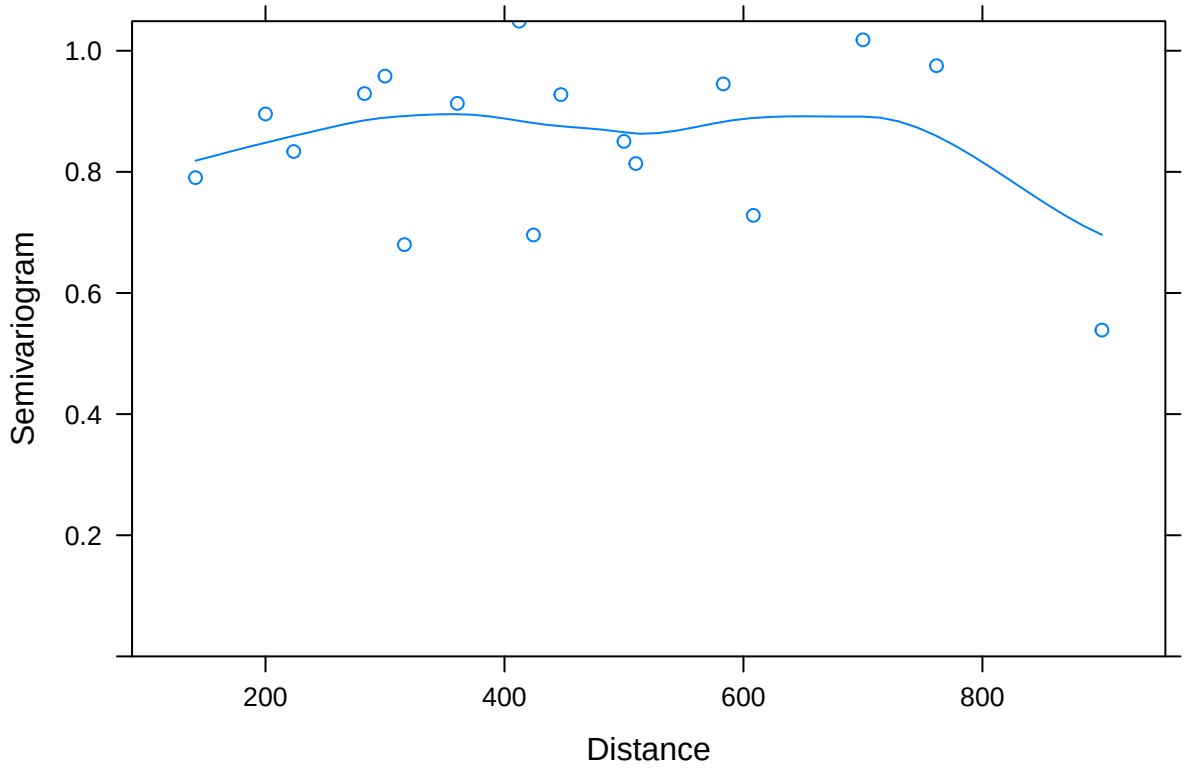
Now for the model with all of the species. Once again I started out by creating the model with all of the species and looking at the variogram of the residuals. There appears to be less spatial dependence in the model with all species than with the one species. I say this due to the fact that the variogram of the residuals has no real pattern. See code below:

```
all_species <- gls(Drypetes.standleyi~Quassia.amara +
                     Cordia.lasiocalyx + Hirtella.triandra +
                     Picramnia.latifolia + Tabernaemontana.arborea +
                     Trattinnickia.aspera + Xylopia.macrantha, data=bci_dat)
summary(all_species)
```

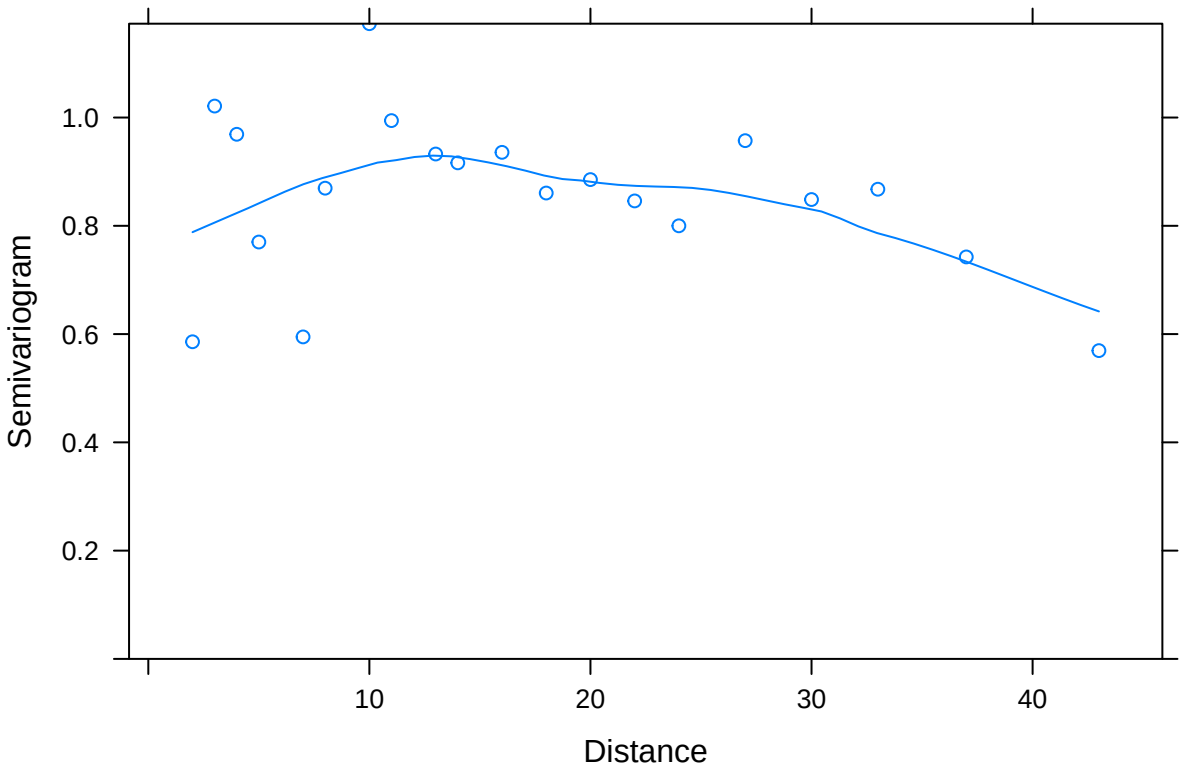
```
## Generalized least squares fit by REML
##   Model: Drypetes.standleyi ~ Quassia.amara + Cordia.lasiocalyx + Hirtella.triandra +      Picramnia
##   Data: bci_dat
##        AIC      BIC    logLik
##   307.1163 322.7554 -144.5582
##
## Coefficients:
##                              Value Std.Error   t-value p-value
## (Intercept)             -1.051752 2.1175346 -0.496687  0.6220
## Quassia.amara            4.085661 2.2842770  1.788602  0.0809
## Cordia.lasiocalyx        0.428920 0.2039316  2.103255  0.0415
## Hirtella.triandra        0.122279 0.0802638  1.523462  0.1351
## Picramnia.latifolia      0.662259 0.6358905  1.041468  0.3036
## Tabernaemontana.arborea -0.249725 0.1491192 -1.674667  0.1014
## Trattinnickia.aspera     1.349323 0.7147412  1.887848  0.0660
## Xylopia.macrantha        0.548832 0.1468772  3.736672  0.0006
##
##  Correlation:
##                         (Intr) Qss.mr Crd.ls Hrtll. Pcrmn. Tbrnm. Trttn.
## Quassia.amara            0.163
## Cordia.lasiocalyx       -0.618 -0.378
## Hirtella.triandra       -0.212  0.307 -0.354
## Picramnia.latifolia      0.025 -0.302 -0.019 -0.381
## Tabernaemontana.arborea -0.708  0.148  0.245  0.163 -0.113
## Trattinnickia.aspera    -0.139 -0.708  0.187 -0.311  0.308 -0.144
## Xylopia.macrantha       -0.140  0.314 -0.125  0.156 -0.463  0.279 -0.294
```

```
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -1.87708765 -0.42701500 -0.04032793  0.23615609  3.38768871
##
## Residual standard error: 4.539713
## Degrees of freedom: 50 total; 42 residual
```

```
plot(Variogram(all_species, form = ~x+y))
```




```
plot(Variogram(all_species, resType='n'))
```



However, I will use the exponential error structure once again to see how that effects the model with all of the species. There appears to be more of a pattern for the normalized residuals in the model with the error structure than without. That is, modeling the error structure had no positive impact on the spatial dependence of the model.

```
all_exp <- gls(Drypetes.standleyi~Quassia.amara +
                 Cordia.lasiocalyx + Hirtella.triandra +
                 Picramnia.latifolia + Tabernaemontana.arborea +
                 Trattinnickia.aspera + Xylopia.macrantha,
               data=bci_dat, corr=corExp(form = ~ x+y, nugget=T),
               control = glsControl(opt='optim', msVerbose=T))
```

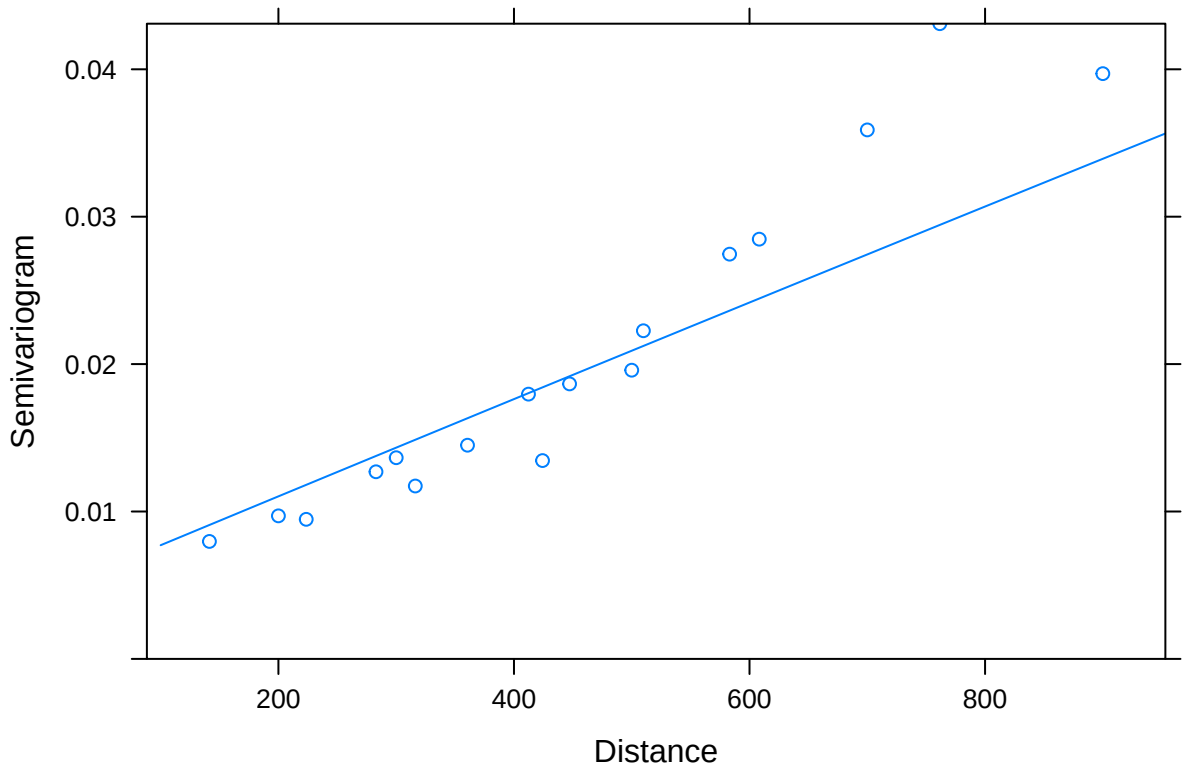
```
## initial  value 161.334238
## iter  10 value 158.897416
## iter  20 value 158.881464
## final  value 158.879778
## converged
```

```
summary(all_exp)
```

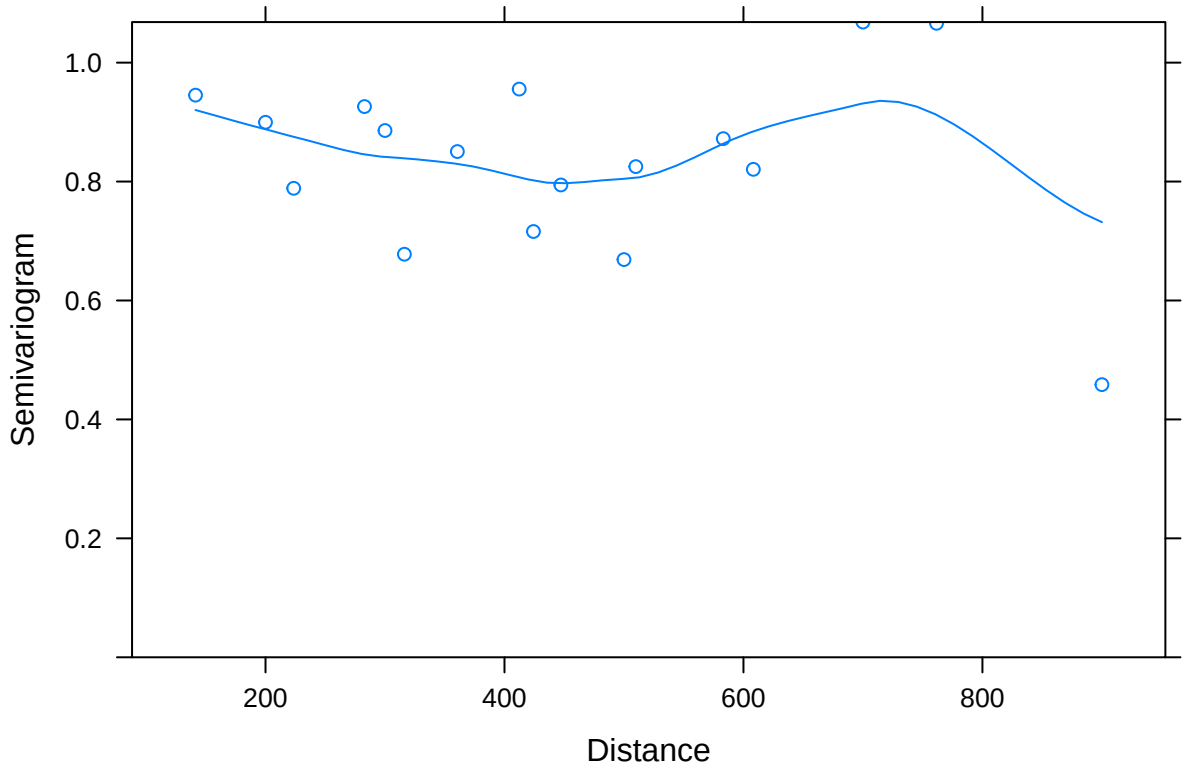
```
## Generalized least squares fit by REML
##   Model: Drypetes.standleyi ~ Quassia.amara + Cordia.lasiocalyx + Hirtella.triandra +      Picramnia
##   Data: bci_dat
##        AIC      BIC    logLik
##   301.9683 321.0826 -139.9841
##
## Correlation Structure: Exponential spatial correlation
##  Formula: ~x + y
##  Parameter estimate(s):
```

```
##        range       nugget
## 2.987794e+04 4.384773e-03
##
## Coefficients:
##                              Value Std.Error   t-value p-value
## (Intercept)              3.0176507  41.75315  0.072274  0.9427
## Quassia.amara            1.3311569   1.93645  0.687420  0.4956
## Cordia.lasiocalyx        0.1426934   0.18954  0.752855  0.4557
## Hirtella.triandra       -0.0014346   0.09043 -0.015863  0.9874
## Picramnia.latifolia      0.2863453   0.52745  0.542883  0.5901
## Tabernaemontana.arborea  0.0407405   0.13953  0.291994  0.7717
## Trattinnickia.aspera     1.8165108   0.57285  3.171032  0.0028
## Xylopia.macrantha        0.4085102   0.15374  2.657119  0.0111
##
##  Correlation:
##                          (Intr) Qss.mr Crd.ls Hrtll. Pcrmn. Tbrnm. Trttn.
## Quassia.amara            -0.008
## Cordia.lasiocalyx        -0.040 -0.292
## Hirtella.triandra        -0.044  0.344 -0.098
## Picramnia.latifolia       0.004 -0.193  0.016 -0.360
## Tabernaemontana.arborea -0.031  0.088 -0.020  0.160 -0.197
## Trattinnickia.aspera     -0.014 -0.655  0.165 -0.276  0.255 -0.036
## Xylopia.macrantha        -0.008  0.306 -0.066 -0.038 -0.047  0.140 -0.183
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -0.20840999 -0.12258058 -0.08687395  0.02426672  0.44015116
##
## Residual standard error: 42.12227
## Degrees of freedom: 50 total; 42 residual
```

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




```
plot(Variogram(all_exp, resType='n', maxDist=max_dist))
```



```
anova(all_species, all_exp)
```

```
##             Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## all_species     1  9 307.1163 322.7554 -144.5582
## all_exp         2 11 301.9683 321.0826 -139.9841 1 vs 2 9.148073  0.0103
```

Furthermore, when looking at the coefficients of the models, the intercept and beta coefficients did not change that significantly. The correlation coefficients are hard to decipher since there is so many, but looking at them it doesn't appear that any change drastically. And finally, when using the anova() function we see that the AIC values for the model with and without the error structure are pretty close together. Therefore, modeling the error structure didn't have a significant impact on the overall model fit.

So, in conclusion, it appears that the model with one species had significant spatial dependence while the model with all of the species did not. Furthermore, using the exponential error structure in the one species model seemed to improve the overall model fit while that was not the case for the model with all of the species. So, with the one species model you are using only one species to model the abundance of the other. However, with the other model you are using multiple species to model the abundance of that one species. So, with the multiple species model there is less chance that any one site will be more related to the others since you are combining the abundance of all of the species. That is, having similar abundance of all of the different species in multiple sites is unlikely. In other words, in order for a sites to be similar you would have to have similar abundances for all of the species in the model which is less likely than with the single species model. On the other hand, with one species there is a much higher chance that nearby sites will have similar abundances.

When trying to use the shortcut for the model with all of the species I had a difficult time getting the variograms to work. The distances used for x and y did not seem right. So I wrote out the whole model from a dataframe with BCI and BCI_xy.