Quantitative Methods
Spatial Modeling Assignment
Lloyd Hill
2.18.2019

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

#identify a rare and common tree species in BCI dataset (I chose 'Solanum hayesii' as the rare species with 12 occurences, and 'Alseis blackiana' as the common species with 983 occurences)

sort(**colSums**(BCI))

BCI_rare <- BCI**$**Solanum.hayesii
BCI_common <- BCI**$**Alseis.blackiana

#plot these species with spatial coordinates

col_brks = **hist**(BCI_common, plot=F)**$**breaks
col_indices = **as.numeric**(**cut**(BCI_common, col_brks))
cols = **rev**(**terrain.colors**(**length**(col_brks)))
plot(BCI_xy, cex=2, pch=19, col=cols[col_indices], ylab="UTM N", xlab="UTM E",
main="Abundance of Alseis blackiana")

Abundance of Alseis blackiana
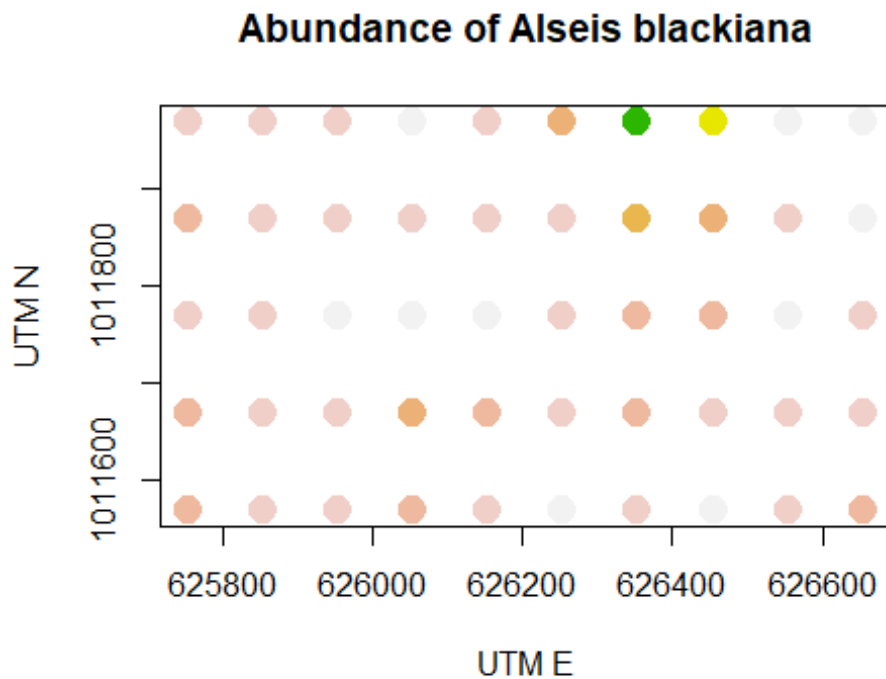
```
col_brks = hist(BCI_rare, plot=F)$breaks
col_indices = as.numeric(cut(BCI_rare, col_brks))
cols = rev(terrain.colors(length(col_brks)))
plot(BCI_xy, cex=2, pch=19, col=cols[col_indices], ylab="UTM N", xlab="UTM E",
main="Abundance of Solanum hayesii")
```

Abundance of Solanum hayesii
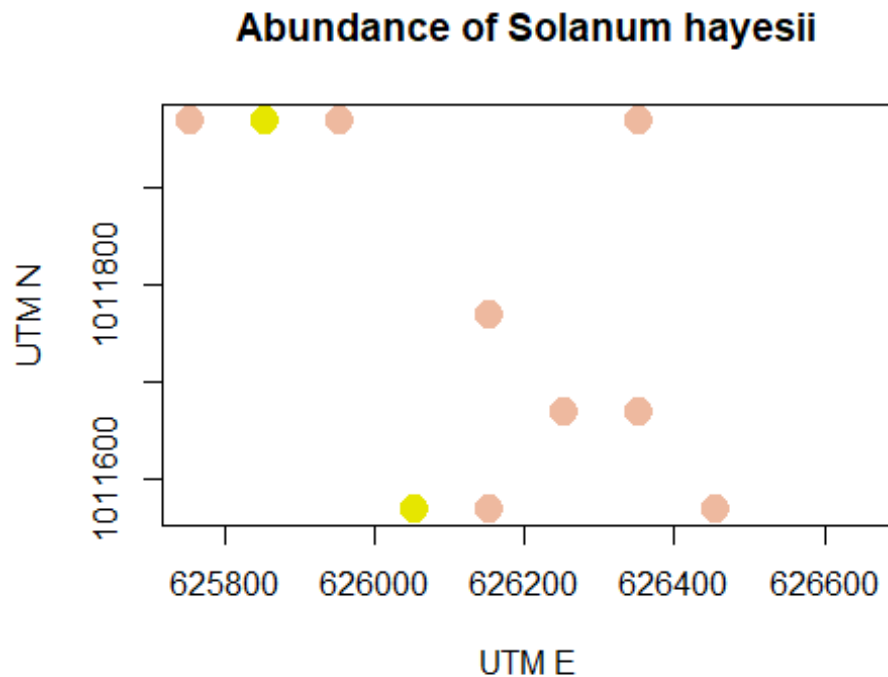

```
#use dist() function to calculate distances among occurences in common and rare species
BCI_common_dist <- dist(BCI_common)
BCI_rare_dist <- dist(BCI_rare)
BCI_xy_dist <- dist(BCI_xy)

max_dist <- max(BCI_xy_dist)/2

#plot results
plot(BCI_xy_dist, BCI_common_dist)
abline(lm(BCI_common_dist ~ BCI_xy_dist), lwd=2, col='red')
lines(lowess(BCI_xy_dist, BCI_common_dist), lwd=2, lty=2, col='red')
abline(v = max_dist, col='red', lwd=1, lty=2)
```
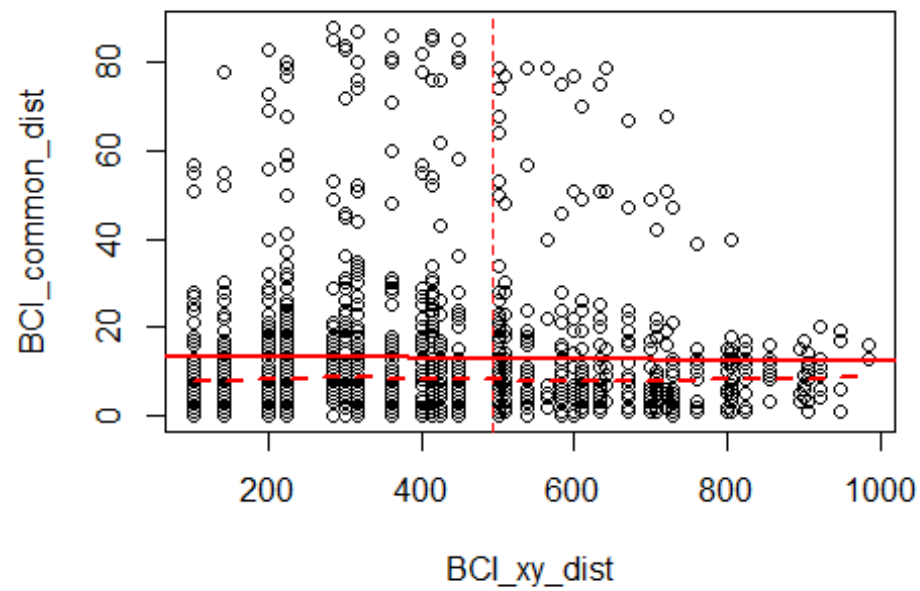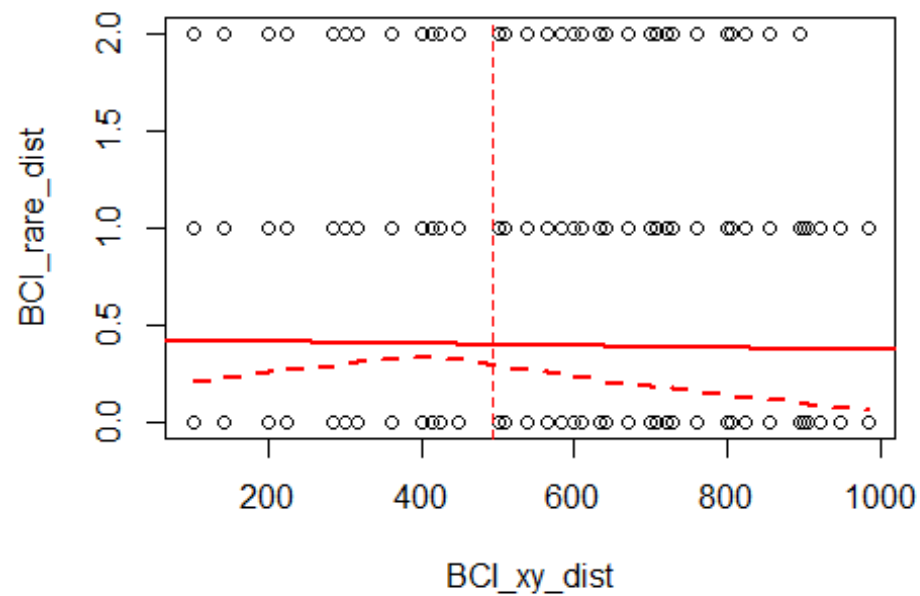
```
plot(BCI_xy_dist, BCI_rare_dist)
abline(lm(BCI_rare_dist ~ BCI_xy_dist), lwd=2, col='red')
lines(lowess(BCI_xy_dist, BCI_rare_dist), lwd=2, lty=2, col='red')
abline(v = max_dist, col='red', lwd=1, lty=2)
```

```
#use mantel() function to test for correlation
mantel(BCI_xy_dist, BCI_common_dist)
##
## Mantel statistic based on Pearson's product-moment correlation
##
## Call:
## mantel(xdis = BCI_xy_dist, ydis = BCI_common_dist)
##
## Mantel statistic r: -0.02348
##       Significance: 0.59
##
## Upper quantiles of permutations (null model):
##    90%    95%  97.5%    99%
## 0.0895 0.1162 0.1363 0.1573
## Permutation: free
## Number of permutations: 999


mantel(BCI_xy_dist, BCI_rare_dist)
##
## Mantel statistic based on Pearson's product-moment correlation
##
## Call:
## mantel(xdis = BCI_xy_dist, ydis = BCI_rare_dist)
##
## Mantel statistic r: -0.01893
##       Significance: 0.596
##
## Upper quantiles of permutations (null model):
##    90%    95%  97.5%    99%
## 0.0768 0.0987 0.1162 0.1485
## Permutation: free
## Number of permutations: 999
```

I employed a variety of techniques to test for spatial dependence in rare and common tree species from the BCI dataset. First, I plotted each species and their abundance within the spatial grid (BCI_xy) to see if any patterns immediately stood out. The common species, Alseis blackiana, appears to be evenly distributed throughout the study plots with a potential 'hotspot' near the northeastern corner. However, I did not observe any particular pattern with the abundance of the rare species, Solanum hayesii. My next step was to plot the relationship among species' occurences and the spatial grid (with distance between species on the y-axis and distance between coordinates on the x-axis). The regression line produced on both rare and common plots were virtually flat suggesting no spatial relationship. I then employed the 'Mantel Test' to test for spatial correlation within rare and common species. The resulting p-values and r statistics (0.61 and -0.023 for common; 0.63 and -0.183 for rare) support my earlier observation of no spatial correlation within rare and common tree 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")

#create separate vectors for x and y coordinates

BCI_x <- BCI_xy$x
BCI_y <- BCI_xy$y

#create vectors for each species of interest

drysta <- BCI$Drypetes.standleyi
corlas <- BCI$Cordia.lasiocalyx
hirtri <- BCI$Hirtella.triandra
piclat <- BCI$Picramnia.latifolia
quaama <- BCI$Quassia.amara
tabarb <- BCI$Tabernaemontana.arborea
traasp <- BCI$Trattinnickia.aspera
xylmac <- BCI$Xylopia.macrantha

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

BCI_gls_1 <- gls(drysta ~ corlas)

summary(BCI_gls_1)

## Generalized least squares fit by REML
##   Model: drysta ~ corlas
##   Data: NULL
##       AIC      BIC    logLik
##   335.1246 340.7382 -164.5623
##
## Coefficients:
##               Value Std.Error   t-value p-value
## (Intercept) -2.782803 1.9173184 -1.451404  0.1532
## corlas       1.165220 0.2292289  5.083217  0.0000
##
##  Correlation:
##        (Intr)
## corlas -0.87
##
## Standardized residuals:
##        Min        Q1       Med        Q3       Max
## -1.8757378 -0.4356989 -0.1061361  0.1804392  4.5135685
##
## Residual standard error: 6.675561
## Degrees of freedom: 50 total; 48 residual
```
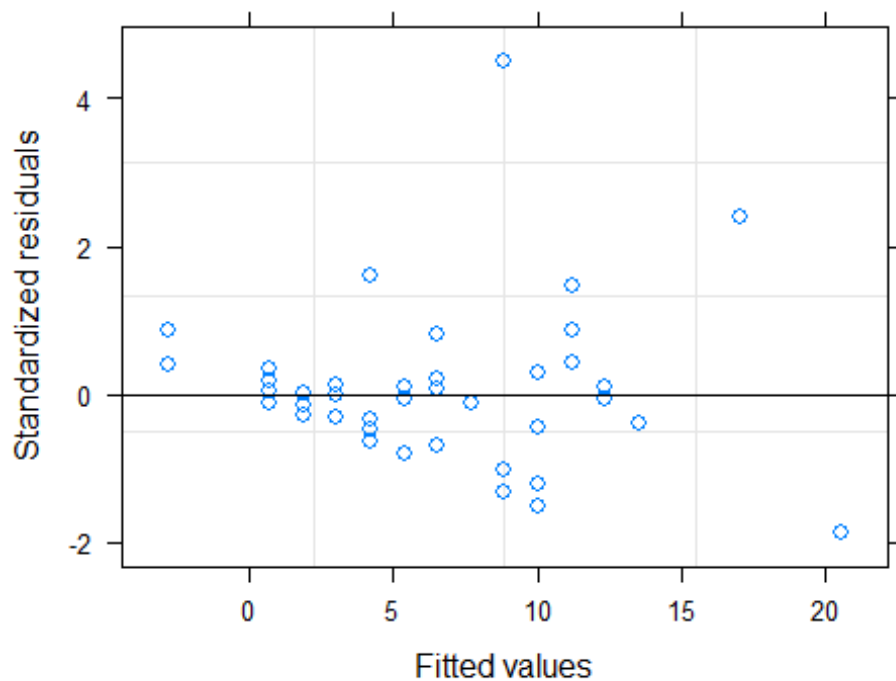
plot(BCI_gls_1)

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

BCI_gls_2 <- **gls**(drysta ~ corlas + hirtri + piclat + quaama + tabarb + traasp + xylmac)

summary(BCI_gls_2)

```
## Generalized least squares fit by REML
##   Model: drysta ~ corlas + hirtri + piclat + quaama + tabarb + traasp +    xylmac
##   Data: NULL
##      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
## corlas       0.428920 0.2039316  2.103255  0.0415
## hirtri       0.122279 0.0802638  1.523462  0.1351
## piclat       0.662259 0.6358905  1.041468  0.3036
## quaama       4.085661 2.2842770  1.788602  0.0809
## tabarb      -0.249725 0.1491192 -1.674667  0.1014
## traasp       1.349323 0.7147412  1.887848  0.0660
## xylmac       0.548832 0.1468772  3.736672  0.0006
##
```

```
##  Correlation:
##       (Intr) corlas hirtri piclat quaama tabarb traasp
## corlas -0.618
## hirtri -0.212 -0.354
## piclat  0.025 -0.019 -0.381
## quaama  0.163 -0.378  0.307 -0.302
## tabarb -0.708  0.245  0.163 -0.113  0.148
## traasp -0.139  0.187 -0.311  0.308 -0.708 -0.144
## xylmac -0.140 -0.125  0.156 -0.463  0.314  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
```
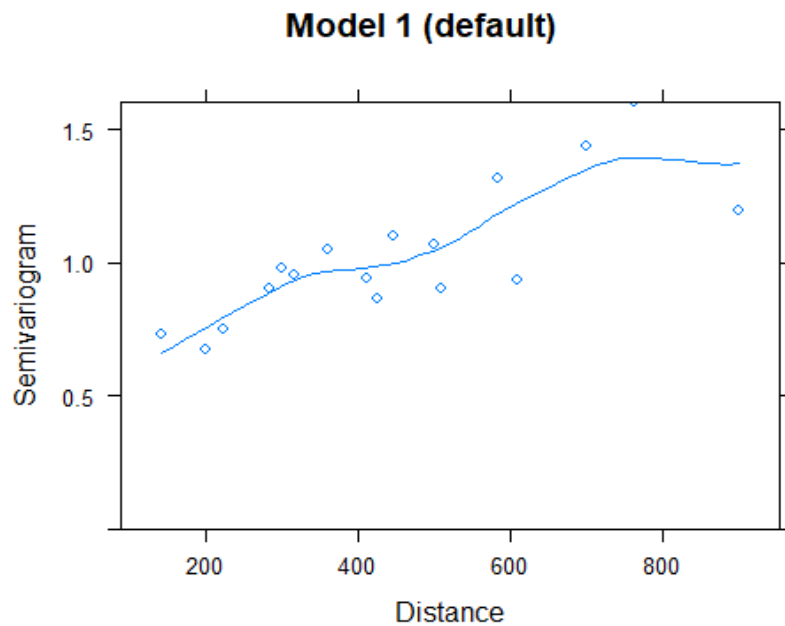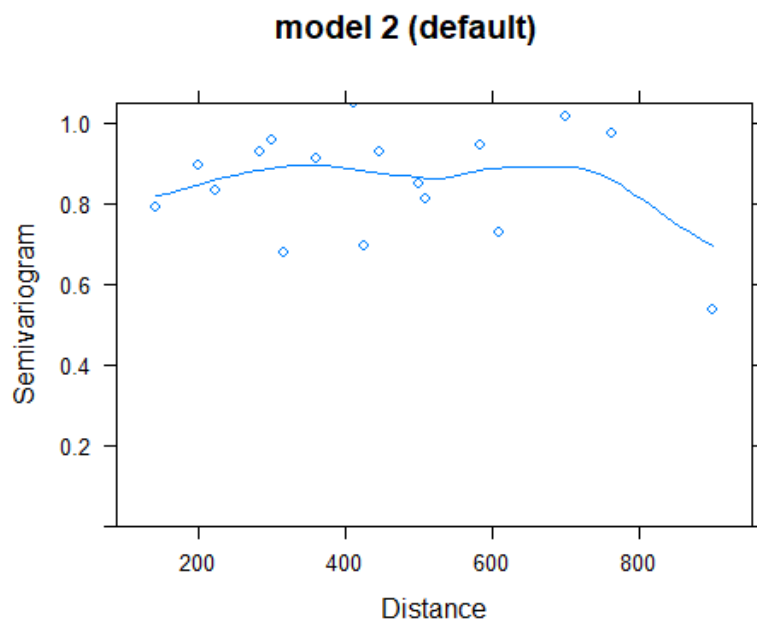
plot(BCI_gls_2)
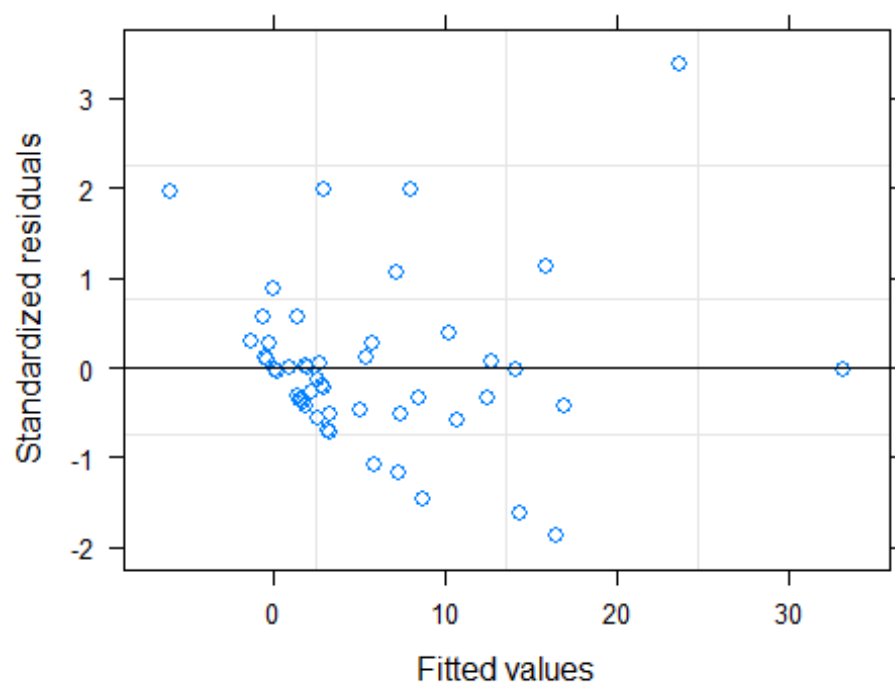
BCI_gls_1_exp <- **update**(BCI_gls_1, corr=**corExp**(form = ~ BCI_x + BCI_y))
BCI_gls_2_exp <- **update**(BCI_gls_2, corr=**corExp**(form = ~ BCI_x + BCI_y))
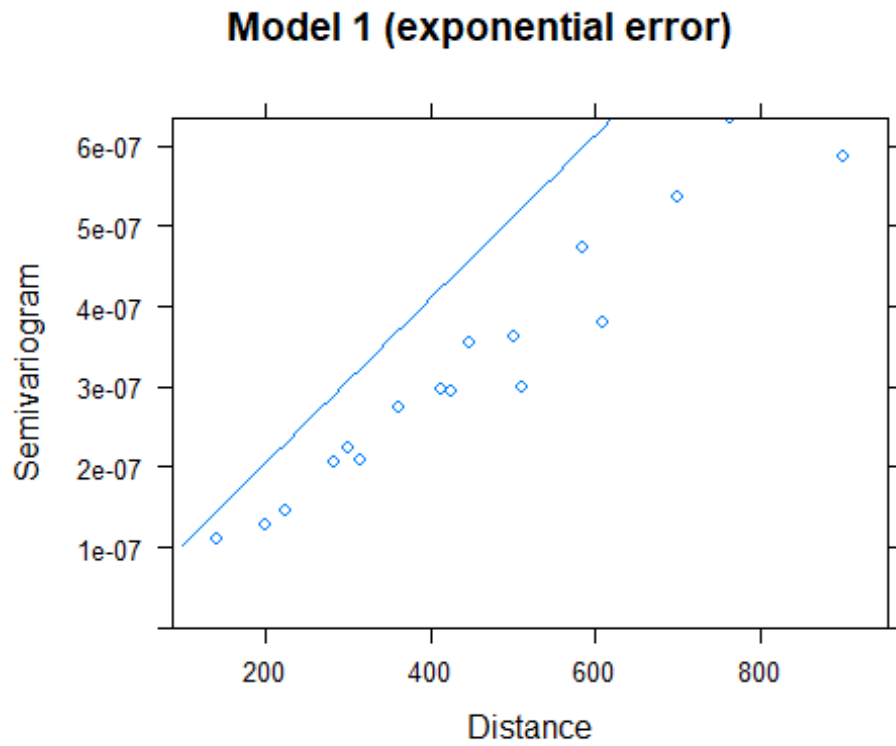
plot(**Variogram**(BCI_gls_1, form= ~ BCI_x + BCI_y), main = "Model 1 (default)")

Model 1 (default)

plot(**Variogram**(BCI_gls_2, form= ~ BCI_x + BCI_y), main = "model 2 (default)")

model 2 (default)

plot(**Variogram**(BCI_gls_1_exp, form= ~ BCI_x + BCI_y), main = "Model 1 (exponential error)")

Model 1 (exponential error)

plot(**Variogram**(BCI_gls_2_exp, form= ~ BCI_x + BCI_y), main = "Model 2 (exponential error)")

Model 2 (exponential error)
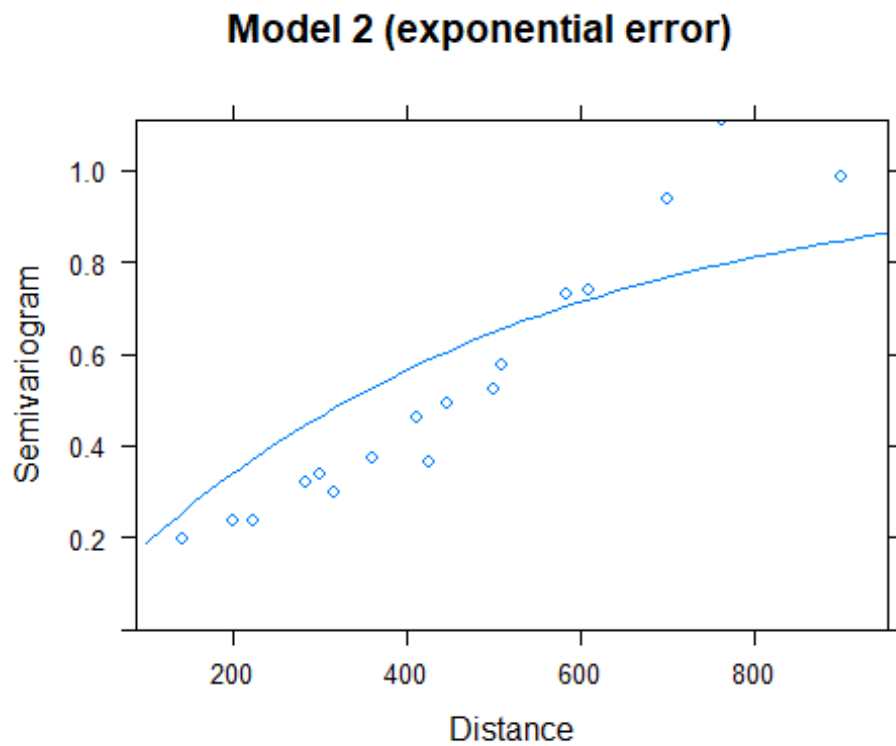
```
#compare fits of models with anova()

anova(BCI_gls_1, BCI_gls_1_exp)

##                    Model df AIC    BIC    logLik   Test    L.Ratio    p-value
## BCI_gls_1          1      3  335.12 340.74 -164.56
## BCI_gls_1_exp      2      4  310.64 318.13 -151.32  1 vs 2  26.48081  <.0001

anova(BCI_gls_2, BCI_gls_2_exp)
##                    Model df AIC    BIC    logLik   Test    L.Ratio    p-value
## BCI_gls_2          1      9  307.12 322.76 -144.56
## BCI_gls_2_exp      2     10  301.61 318.98 -140.80  1 vs 2  7.510175   0.0061
```

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

Yes, when error terms were included in the model the coefficients appeared to behave better (offer more predictability). In the resulting variograms, data points are arranged in a more linear fashion with a pronounced slope when error terms are included in the model.

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

Yes, including the spatial error terms significantly improved the fit of these models. For the first model, AIC lowered from 340 to 310 with the inclusion of exponential error terms. This effect was less pronounced in the second model (although still significant), where AIC dropped from 307 to 301 following the inclusion of exponential error terms.

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

Including the error terms resulted in a better model because these data are not truly indepenent. Spatial statistics are tricky in modeling plant distributions because of autocorrelation. For example, the plant species of interest may grow in clusters, or there may be another underlying environmental variable such as elevation. Including the error terms in these models accounted for some of this variability.