Spatial Model Assignment

Katie Beidler 02/22/15

```r
library(vegan)

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

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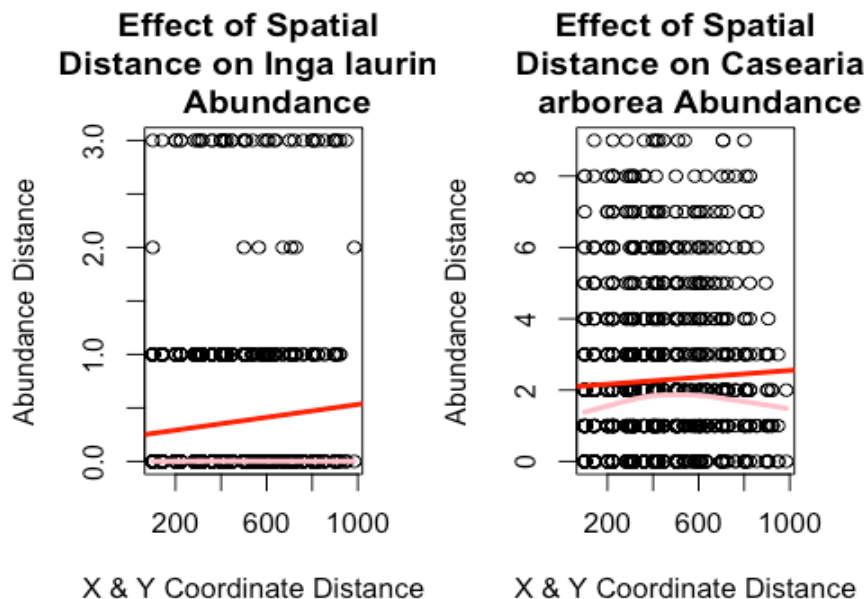
1) **Examine if there is evidence of spatial dependence in a rare and a common species in the BCI tree dataset**

```r
# Choosing a rare/ abundant sp
abu = apply(BCI, 2, sum)
summary(abu)

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

sp_order = sort(abu)
# rare species:Inga.laurina
# abundant species:Casearia.arborea
I_laur = BCI$Inga.laurina
C_arb = BCI$Casearia.arborea

## Testing for Spatial Dependence
I_dist = dist(I_laur) #how different each site pair is in abundance of Inga laurina
C_dist = dist(C_arb) #how different each site pair is in abundance of Casearia arborea
xy_dist = dist(BCI_xy) # how different each site pair is in spatial distance
```


Effect of Spatial Distance on Inga laurin Abundance — Abundance Distance vs X & Y Coordinate Distance

Effect of Spatial Distance on Casearia arborea Abundance — Abundance Distance vs X & Y Coordinate Distance

Mantel test: to test if the spatial dependence is significant

```
I_mantel = mantel(xy_dist, I_dist)
C_mantel = mantel(xy_dist, C_dist)
I_mantel

##
## Mantel statistic based on Pearson's product-moment correlation
##
## Call:
## mantel(xdis = xy_dist, ydis = I_dist)
##
## Mantel statistic r: 0.09446
##       Significance: 0.091
##
## Upper quantiles of permutations (null model):
##     90%    95%  97.5%     99%
## 0.0879 0.1129 0.1295 0.1451
## Permutation: free
## Number of permutations: 999

C_mantel

##
## Mantel statistic based on Pearson's product-moment correlation
##
## Call:
## mantel(xdis = xy_dist, ydis = C_dist)
##
## Mantel statistic r: 0.04719
##       Significance: 0.204
##
## Upper quantiles of permutations (null model):
##     90%    95%  97.5%     99%
## 0.0746 0.1053 0.1204 0.1439
## Permutation: free
## Number of permutations: 999
```

- Both Mantel tests resulted in a P value greater than 0.05 and spatial distance only explains 9% of the variation in *Inga laurina* abundance and 5% of the variation in *Casearia arborea* abundance. Inga laurina and Casearia arborea abundance are not spatially correlated, or spatial dependence is not supported in this case.

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")
BCI_sp = BCI[sp_ids]
D_stand = BCI$Drypetes.standleyi
C_las = BCI$Cordia.lasiocalyx
H_tri = BCI$Hirtella.triandra
P_lat = BCI$Picramnia.latifolia
Q_ama = BCI$Quassia.amara
```

```
T_arb = BCI$Tabernaemontana.arborea
T_asp = BCI$Trattinnickia.aspera
X_mac = BCI$Xylopia.macrantha
```

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

```
library(nlme)
BCI_dat = data.frame(BCI, BCI_xy)
D_lm1 = gls(D_stand ~ C_las, data=BCI_dat)
summary(D_lm1)

## Generalized least squares fit by REML
##   Model: D_stand ~ C_las
##   Data: BCI_dat
##        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
## C_las        1.165220 0.2292289  5.083217  0.0000
##
##  Correlation:
##       (Intr)
## C_las -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
```

 # the relationship between Drypetes standleyi and Cordia lasiocalyx abundance is significant
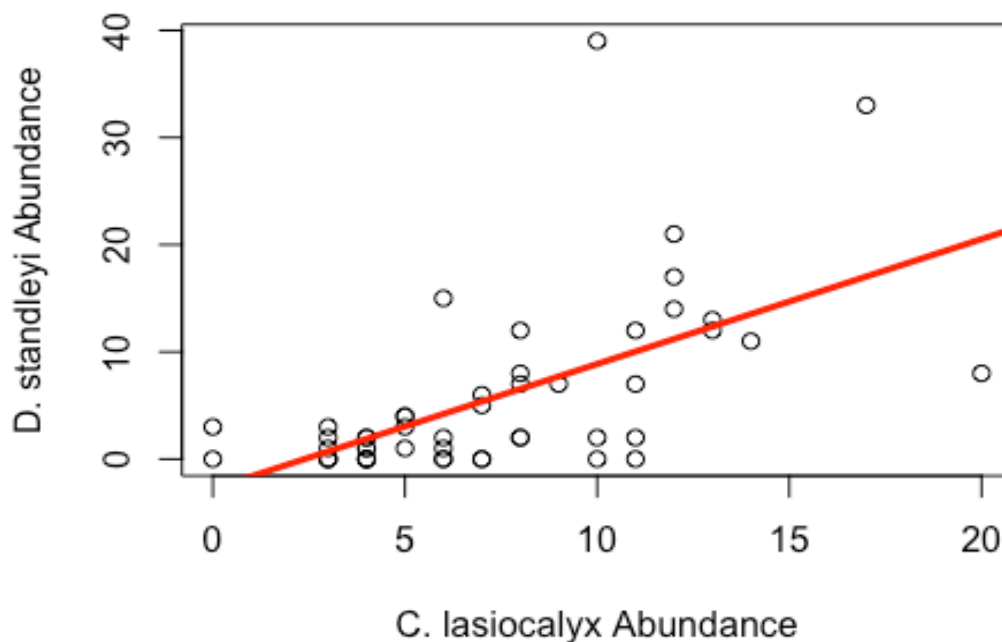
Model 2: include all of the species as predictor variables- multivariate model

```
D_lm2 = gls(D_stand ~ C_las + H_tri + P_lat + Q_ama + T_arb + T_asp + X_mac,
data=BCI_dat)
summary(D_lm2)

## Generalized least squares fit by REML
##   Model: D_stand ~ C_las + H_tri + P_lat + Q_ama + T_arb + T_asp + X_mac
##   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
## C_las        0.428920 0.2039316  2.103255  0.0415
## H_tri        0.122279 0.0802638  1.523462  0.1351
## P_lat        0.662259 0.6358905  1.041468  0.3036
## Q_ama        4.085661 2.2842770  1.788602  0.0809
## T_arb       -0.249725 0.1491192 -1.674667  0.1014
## T_asp        1.349323 0.7147412  1.887848  0.0660
## X_mac        0.548832 0.1468772  3.736672  0.0006
##
##  Correlation:
##        (Intr) C_las  H_tri  P_lat  Q_ama  T_arb  T_asp
## C_las -0.618
## H_tri -0.212 -0.354
## P_lat  0.025 -0.019 -0.381
## Q_ama  0.163 -0.378  0.307 -0.302
## T_arb -0.708  0.245  0.163 -0.113  0.148
## T_asp -0.139  0.187 -0.311  0.308 -0.708 -0.144
## X_mac -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
```

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.

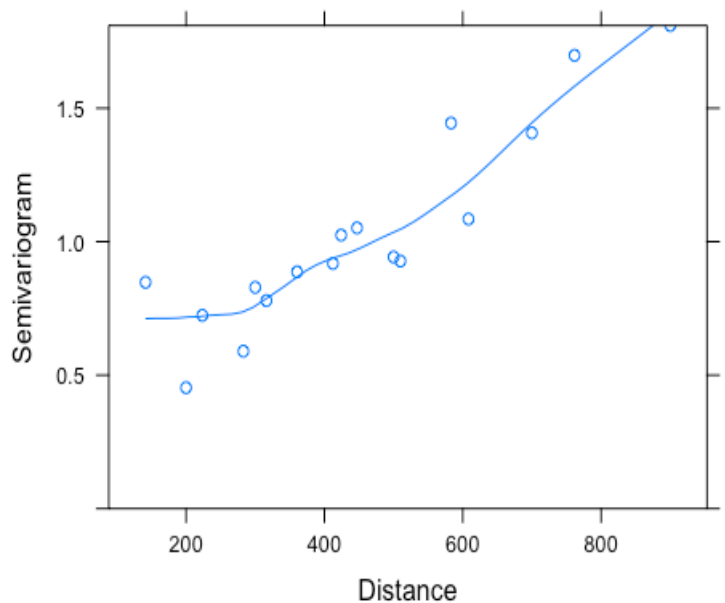- Model 1

```
plot(Variogram(D_lm1, form= ~ x + y))
```
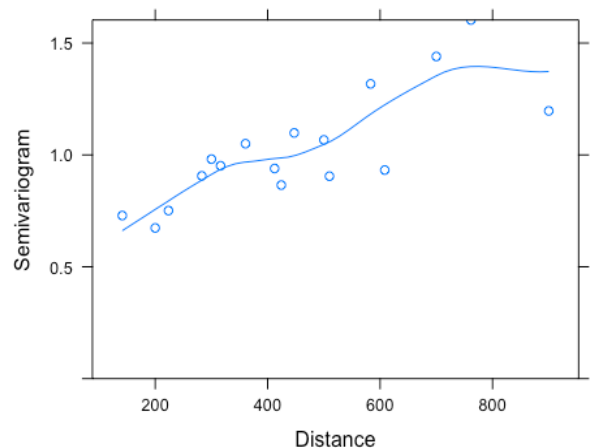
```
# The Variogram depicts spatial dependence
# Modeling the spatial dependence in the residuals to see if the relationship between
Drypetes standleyi and Cordia lasiocalyx abundance remains significant

# Picking which error structure to use
D_exp1 = update(D_lm1, corr=corExp(1, form=~x + y))
D_spher1 = update(D_lm1, corr=corSpher(c(800, 0.8), form=~x + y, nugget=T))
D_lin1 = update(D_lm1, corr=corLin(form=~x+y, nugget=F))
plot(Variogram(D_exp1, resType='n'))

plot(Variogram(D_spher1, resType='n'))

plot(Variogram(D_lin1, resType='n'))
```
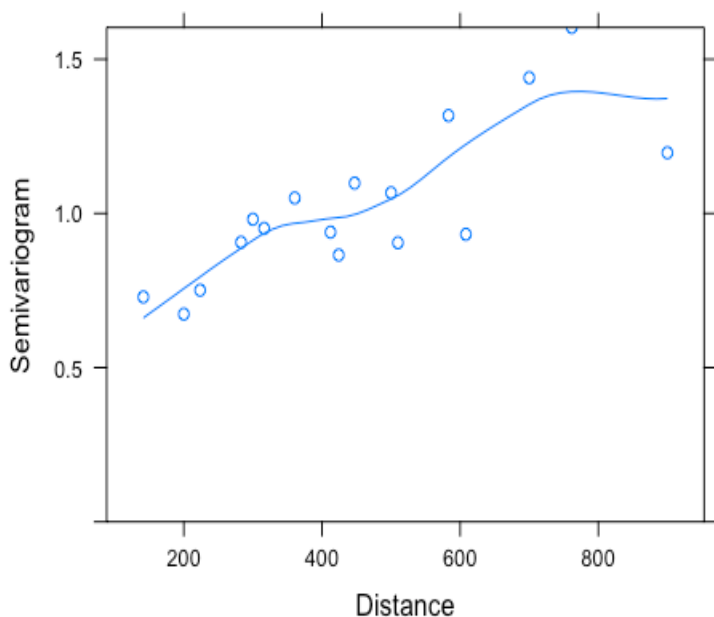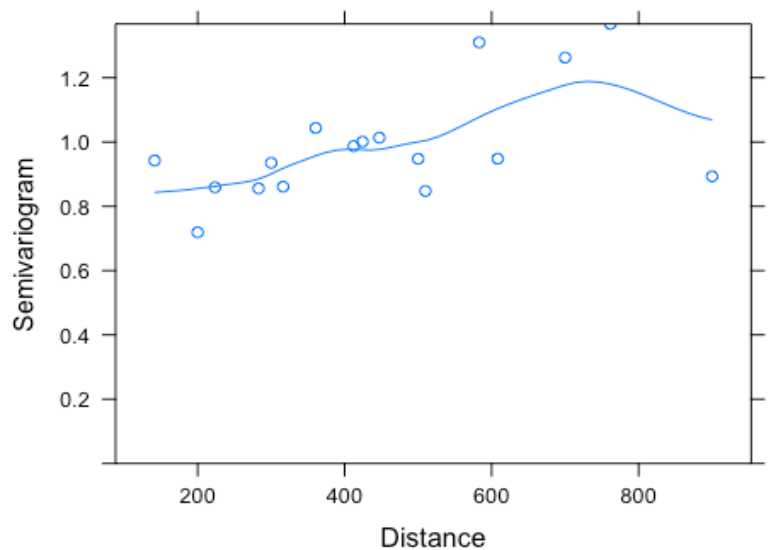


```
AIC(D_lm1)

## [1] 335.1246

AIC(D_exp1)

## [1] 337.1246

AIC(D_spher1)

## [1] 311.0547

AIC(D_lin1)

## [1] 326.0101
```

The spherical spatial correlation has the
lowest AIC/ the residual distribution is
"flatter"

- Model 2

```
#The spherical error structure also works the best for model 2
D_spher2 = update(D_lm2, corr=corSpher(c(800, 0.8), form=~x + y, nugget=T))
```

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

- Including the spatial error had a large impact on the coeffecients in both models- for ex: the intercept term went from -2.78 to 10.7 in model 1 and from -1.05 to 3.05 in model 2

```
summary(D_lm1)

## Generalized least squares fit by REML
##    Model: D_stand ~ C_las
##    Data: BCI_dat
##         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
## C_las        1.165220 0.2292289  5.083217  0.0000
##
##  Correlation:
##        (Intr)
## C_las -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

summary(D_spher1)

## Generalized least squares fit by REML
##    Model: D_stand ~ C_las
##    Data: BCI_dat
##         AIC      BIC    logLik
##    311.0547 320.4107 -150.5274
##
## Correlation Structure: Spherical spatial correlation
##  Formula: ~x + y
##  Parameter estimate(s):
##         range       nugget
## 1.676785e+06 5.531413e-05
##
## Coefficients:
##                 Value Std.Error   t-value p-value
## (Intercept) 10.70638  371.9633 0.0287834  0.9772
## C_las        0.17306    0.2146 0.8065655  0.4239
##
##  Correlation:
##        (Intr)
## C_las -0.006
##
```

```
## Standardized residuals:
##          Min          Q1         Med          Q3         Max
## -0.03389213 -0.02984798 -0.02526077 -0.01379904  0.07139370
##
## Residual standard error: 372.064
## Degrees of freedom: 50 total; 48 residual

summary(D_lm2)

## Generalized least squares fit by REML
##   Model: D_stand ~ C_las + H_tri + P_lat + Q_ama + T_arb + T_asp + X_mac
##   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
## C_las        0.428920 0.2039316  2.103255  0.0415
## H_tri        0.122279 0.0802638  1.523462  0.1351
## P_lat        0.662259 0.6358905  1.041468  0.3036
## Q_ama        4.085661 2.2842770  1.788602  0.0809
## T_arb       -0.249725 0.1491192 -1.674667  0.1014
## T_asp        1.349323 0.7147412  1.887848  0.0660
## X_mac        0.548832 0.1468772  3.736672  0.0006
##
##  Correlation:
##        (Intr) C_las  H_tri  P_lat  Q_ama  T_arb  T_asp
## C_las -0.618
## H_tri -0.212 -0.354
## P_lat  0.025 -0.019 -0.381
## Q_ama  0.163 -0.378  0.307 -0.302
## T_arb -0.708  0.245  0.163 -0.113  0.148
## T_asp -0.139  0.187 -0.311  0.308 -0.708 -0.144
## X_mac -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

summary(D_spher2)

## Generalized least squares fit by REML
##   Model: D_stand ~ C_las + H_tri + P_lat + Q_ama + T_arb + T_asp + X_mac
##   Data: BCI_dat
##        AIC      BIC    logLik
##   301.9592 321.0735 -139.9796
##
## Correlation Structure: Spherical spatial correlation
##  Formula: ~x + y
##  Parameter estimate(s):
##        range        nugget
## 2.359821e+06 8.451233e-05
##
```

```
## Coefficients:
##                  Value Std.Error    t-value p-value
## (Intercept)  3.0501002 303.95100   0.010035  0.9920
## C_las        0.1426674   0.18953   0.752755  0.4558
## H_tri       -0.0017713   0.09038  -0.019598  0.9845
## P_lat        0.2863355   0.52743   0.542883  0.5901
## Q_ama        1.3263740   1.93681   0.684823  0.4972
## T_arb        0.0407522   0.13952   0.292079  0.7717
## T_asp        1.8170747   0.57298   3.171296  0.0028
## X_mac        0.4086716   0.15367   2.659337  0.0110
##
##  Correlation:
##       (Intr) C_las  H_tri  P_lat  Q_ama  T_arb  T_asp
## C_las -0.006
## H_tri -0.006 -0.098
## P_lat  0.001  0.017 -0.360
## Q_ama -0.001 -0.292  0.344 -0.193
## T_arb -0.004 -0.020  0.160 -0.197  0.088
## T_asp -0.002  0.165 -0.276  0.255 -0.655 -0.036
## X_mac -0.001 -0.066 -0.037 -0.048  0.306  0.140 -0.183
##
## Standardized residuals:
##          Min            Q1           Med            Q3           Max
## -0.028963349 -0.017080131 -0.012138198  0.003270555  0.060889615
##
## Residual standard error: 304.0023
## Degrees of freedom: 50 total; 42 residual
```

> ***Did including the spatial error terms significantly improve model fit (use function anova to carry out model comparison)?**
> - Including the spatial error term significantly improved model 1. The AIC was lower for the version of model 2 that included the spatial error term, but the versions were not significantly different.

```
mod1_comp = anova(D_lm1, D_spher1)
mod2_comp = anova(D_lm2, D_spher2)
mod1_comp

##             Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## D_lm1           1  3 335.1246 340.7382 -164.5623
## D_spher1        2  5 311.0547 320.4107 -150.5274 1 vs 2 28.06989  <.0001

mod2_comp

##             Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## D_lm2           1  9 307.1163 322.7554 -144.5582
## D_spher2        2 11 301.9592 321.0735 -139.9796 1 vs 2 9.157175   0.010
```

Explain why modeling the spatial error has a large influence in the first model but not in the second model.

The second model includes multiple species as predictor variables, explaining more of the overall variance and in turn decreasing the amount of variance that can be attributed to spatial error. The first model only includes 1 predictor variable, so including the spatial error term greatly reduces the amount of unexplained variance in comparison to the original model (D_lm1).