

Spatial Modeling

Zach Proux

February 25, 2018

Collaborators: Nick Weber and Megan Sporre

Spatial Modeling Assignment

```
library(vegan)

## Warning: package 'vegan' was built under R version 3.3.3

## Loading required package: permute

## Warning: package 'permute' was built under R version 3.3.3

## 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))
```

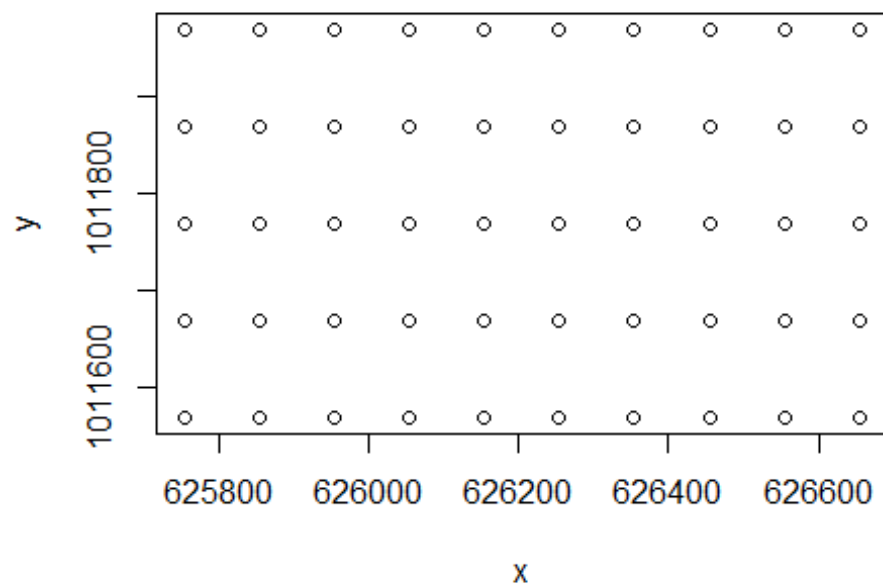
- 1) Examine if there is evidence of spatial dependence in a rare and a common species in the BCI tree dataset

```
abu <- colSums(BCI)
quantile(abu)

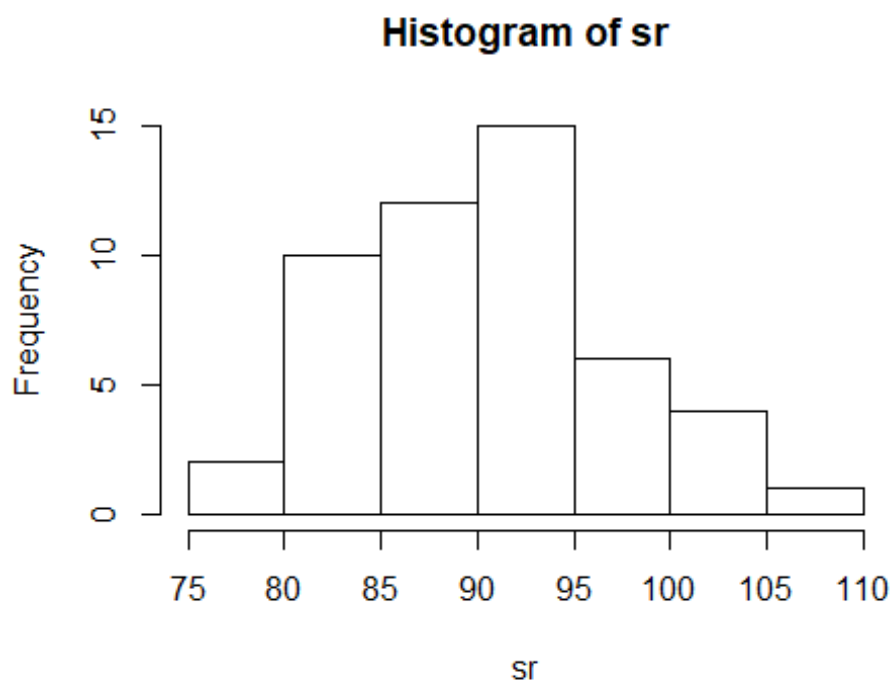
##    0%   25%   50%   75%  100%
##     1     7    25    82  1717

bci_rare = BCI[ , abu <= 25]
bci_comm = BCI[ , abu > 25]

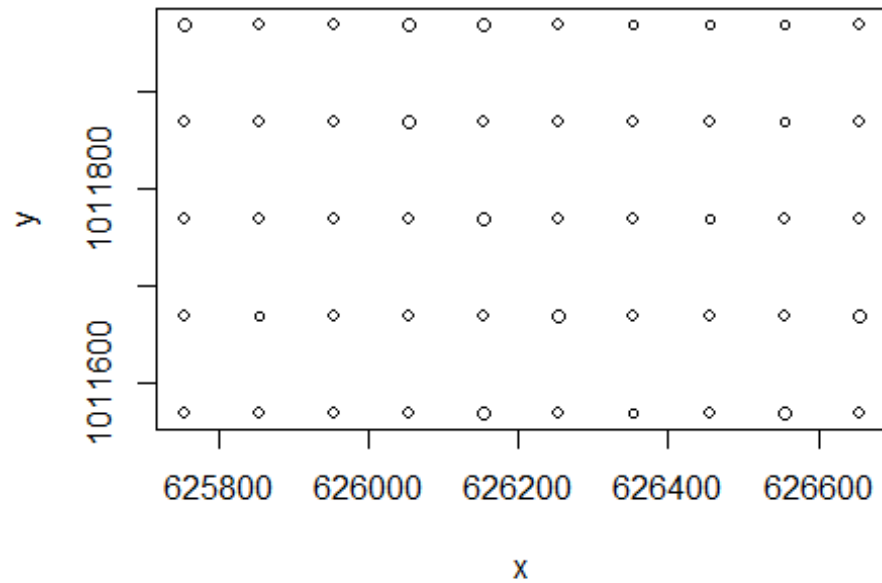
plot(BCI_xy)
```



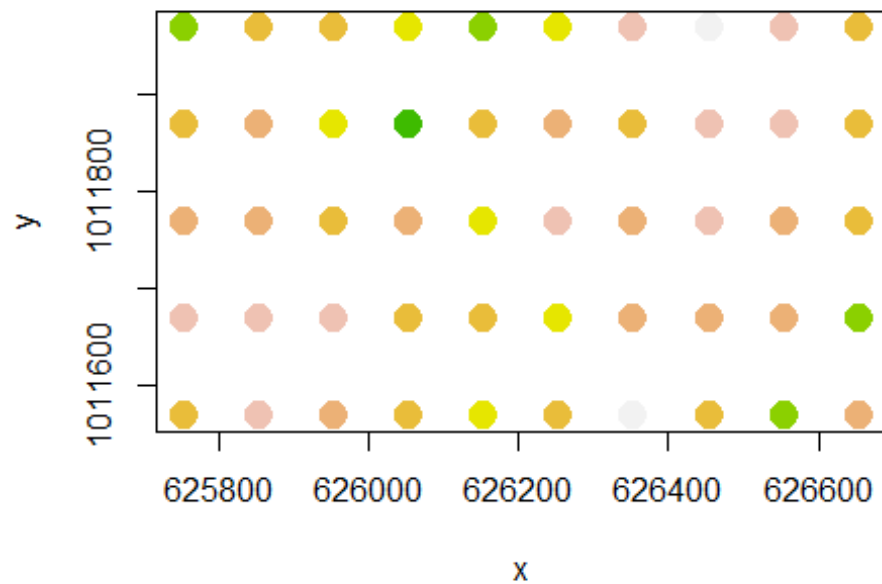
```
sr = apply(BCI, 1, function(x) sum(x > 0))  
hist(sr)
```



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



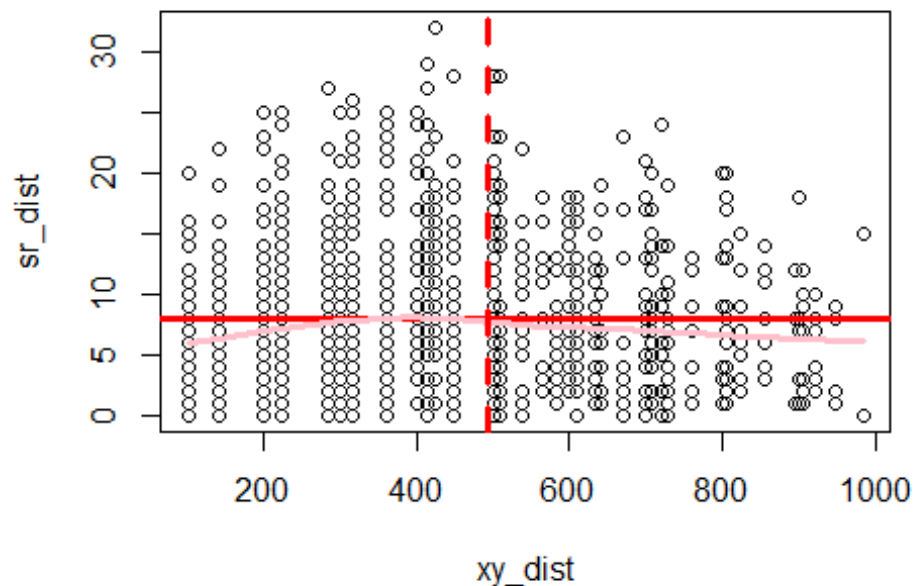
```
col_brks = hist(sr, plot=F)$breaks
col_indices = as.numeric(cut(sr, col_brks))
cols = rev(terrain.colors(length(col_brks)))
plot(BCI_xy, cex=2, pch=19, col=cols[col_indices])
```



```
sr_dist = dist(sr)
xy_dist = dist(BCI_xy)

max_dist = max(xy_dist) / 2

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



```
# compute correlation
obs_cor = cor(xy_dist, sr_dist)
obs_cor

## [1] 0.004339913

# carry out a permutation test for significance:
nperm = 1000
null_cor = obs_cor
for (i in 2:nperm) {
  # shuffle the rows of the spatial coordinates
  tmp_xy = BCI_xy[sample(nrow(BCI_xy)), ]
  # correlation between the shuffled spatial coordinates and sr_dist
  null_cor[i] = cor(dist(tmp_xy), sr_dist)
}
# compute the p-value
sum(null_cor >= obs_cor) / nperm

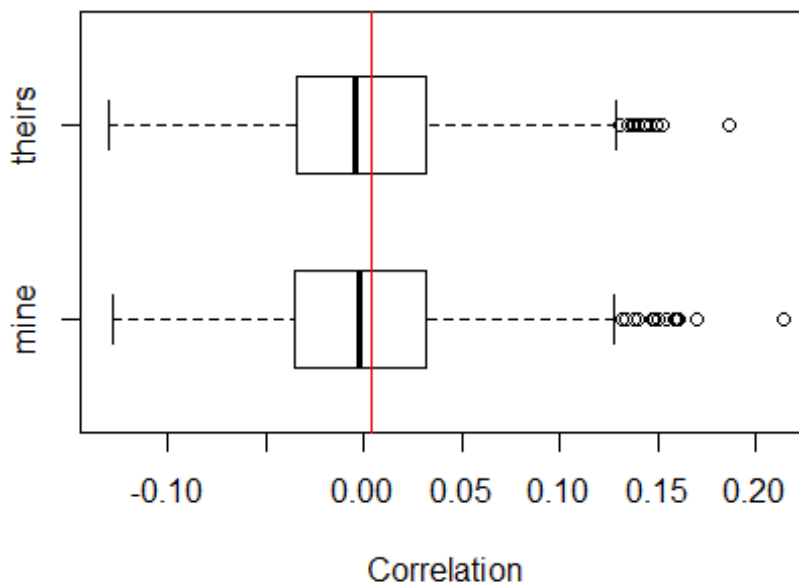
## [1] 0.445

# carry out the same analysis using the function mantel()
sr_mantel = mantel(xy_dist, sr_dist)
sr_mantel

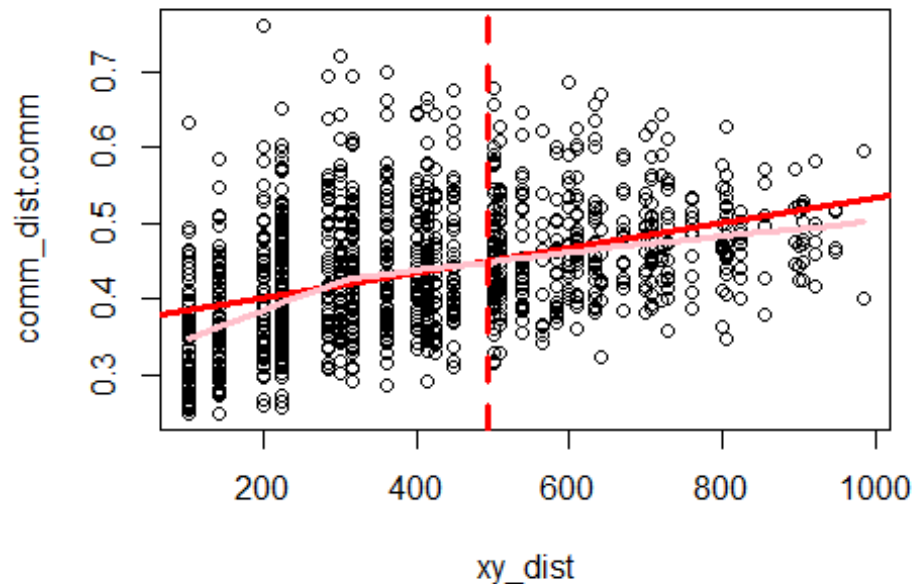
##
## Mantel statistic based on Pearson's product-moment correlation
##
## Call:
```

```
## mantel(xdis = xy_dist, ydis = sr_dist)
##
## Mantel statistic r: 0.00434
##      Significance: 0.438
##
## Upper quantiles of permutations (null model):
##      90%    95%   97.5%   99%
## 0.0669 0.0870 0.1063 0.1307
## Permutation: free
## Number of permutations: 999

# compare the two approaches graphically using stacked boxplots
boxplot(list(null_cor, sr_mantel$perm), horizontal = T, boxwex = 0.5,
        names = c('mine', 'theirs'), xlab='Correlation')
abline(v=obs_cor, col='red')
```



```
## compute bray curtis distance for the community matrix - common species
comm_dist.comm = vegdist(bci_comm)
plot(xy_dist, comm_dist.comm)
abline(lm(comm_dist.comm ~ xy_dist), lwd=3, col='red')
lines(lowess(xy_dist, comm_dist.comm), lwd=3, col='pink')
abline(v = max_dist, col='red', lwd=3, lty=2)
```



#mantel Test

```
comm_mantel = mantel(xy_dist, comm_dist.comm)
comm_mantel
```

```
##
```

```
## Mantel statistic based on Pearson's product-moment correlation
```

```
##
```

```
## Call:
```

```
## mantel(xdis = xy_dist, ydis = comm_dist.comm)
```

```
##
```

```
## Mantel statistic r: 0.4045
```

```
##      Significance: 0.001
```

```
##
```

```
## Upper quantiles of permutations (null model):
```

```
##   90%   95%  97.5%  99%
```

```
## 0.0697 0.0859 0.1040 0.1247
```

```
## Permutation: free
```

```
## Number of permutations: 999
```

```
## compute bray curtis distance for the community matrix - rare species
```

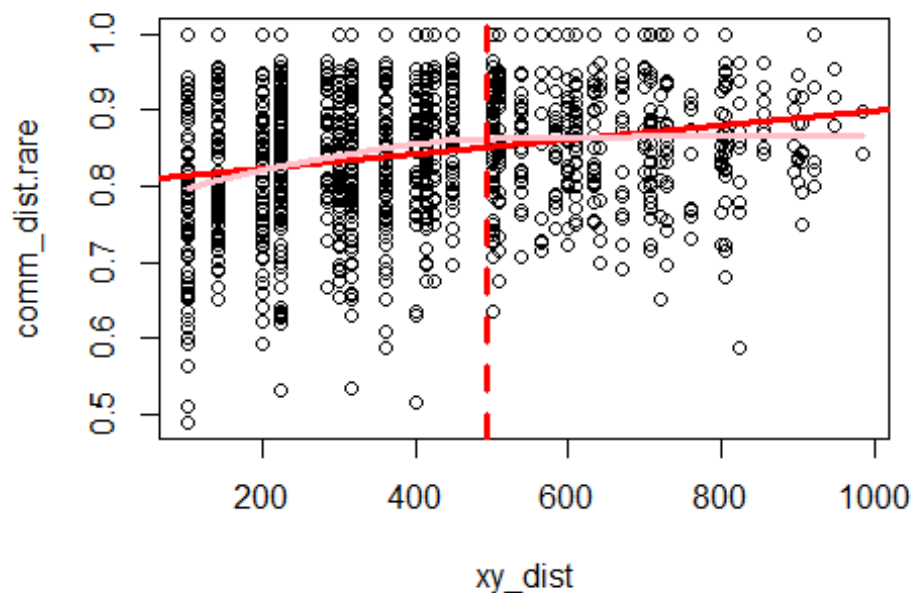
```
comm_dist.rare = vegdist(bci_rare)
```

```
plot(xy_dist, comm_dist.rare)
```

```
abline(lm(comm_dist.rare ~ xy_dist), lwd=3, col='red')
```

```
lines(lowess(xy_dist, comm_dist.rare), lwd=3, col='pink')
```

```
abline(v = max_dist, col='red', lwd=3, lty=2)
```



#mantel Test

```
comm_mantel.rare = mantel(xy_dist, comm_dist.rare)
comm_mantel.rare

##
## Mantel statistic based on Pearson's product-moment correlation
##
## Call:
## mantel(xdis = xy_dist, ydis = comm_dist.rare)
##
## Mantel statistic r: 0.2232
##      Significance: 0.001
##
## Upper quantiles of permutations (null model):
##   90%   95%  97.5%   99%
## 0.0488 0.0685 0.0782 0.0943
## Permutation: free
## Number of permutations: 999
```

- 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")
```



```

sp_a <- BCI[, 'Drypetes.standleyi']
sp_b <- BCI[, 'Cordia.lasiocalyx']
sp_c <- BCI[, 'Hirtella.triandra']
sp_d <- BCI[, 'Picramnia.latifolia']
sp_e <- BCI[, 'Quassia.amara']
sp_f <- BCI[, 'Tabernaemontana.arborea']
sp_g <- BCI[, 'Trattinnickia.aspera']
sp_h <- BCI[, 'Xylopia.macrantha']

```

```
BCI_abc <- cbind(sp_a, sp_b, sp_c, sp_d, sp_e, sp_f, sp_g, sp_h)
```

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

```
library(nlme)
```

```
## Warning: package 'nlme' was built under R version 3.3.3
```

```
# one species
```

```
abundance <- apply(BCI_abc, 1, function(x) sum(x > 0))
```

```
abundance_dat = data.frame(abundance, BCI_abc, BCI_xy)
```

```
single_lm = gls(sp_a ~ sp_b, data = abundance_dat)
```

```
summary(single_lm)
```

```
## Generalized least squares fit by REML
```

```
## Model: sp_a ~ sp_b
```

```
## Data: abundance_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
```

```
## sp_b 1.165220 0.2292289 5.083217 0.0000
```

```
##
```

```
## Correlation:
```

```
## (Intr)
```

```
## sp_b -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
```

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

```

# all species
all_lm = gls(sp_a ~ sp_b + sp_c + sp_d + sp_e + sp_f + sp_g + sp_h, data =
abundance_dat)
summary(all_lm)

## Generalized least squares fit by REML
## Model: sp_a ~ sp_b + sp_c + sp_d + sp_e + sp_f + sp_g + sp_h
## Data: abundance_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
## sp_b         0.428920 0.2039316   2.103255  0.0415
## sp_c         0.122279 0.0802638   1.523462  0.1351
## sp_d         0.662259 0.6358905   1.041468  0.3036
## sp_e         4.085661 2.2842770   1.788602  0.0809
## sp_f        -0.249725 0.1491192  -1.674667  0.1014
## sp_g         1.349323 0.7147412   1.887848  0.0660
## sp_h         0.548832 0.1468772   3.736672  0.0006
##
## Correlation:
##      (Intr) sp_b   sp_c   sp_d   sp_e   sp_f   sp_g
## sp_b -0.618
## sp_c -0.212 -0.354
## sp_d  0.025 -0.019 -0.381
## sp_e  0.163 -0.378  0.307 -0.302
## sp_f -0.708  0.245  0.163 -0.113  0.148
## sp_g -0.139  0.187 -0.311  0.308 -0.708 -0.144
## sp_h -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

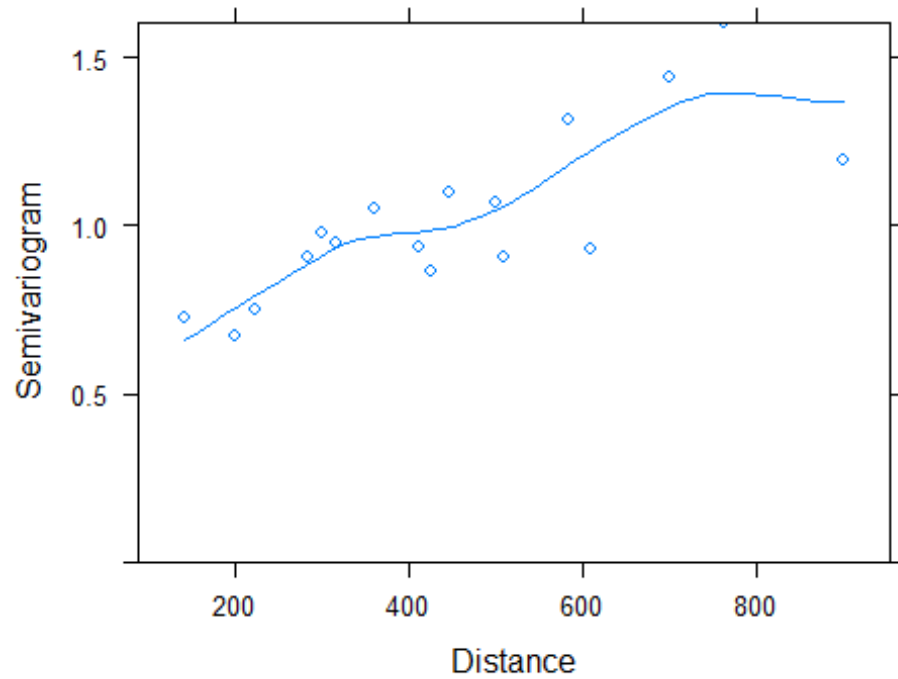
```

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.

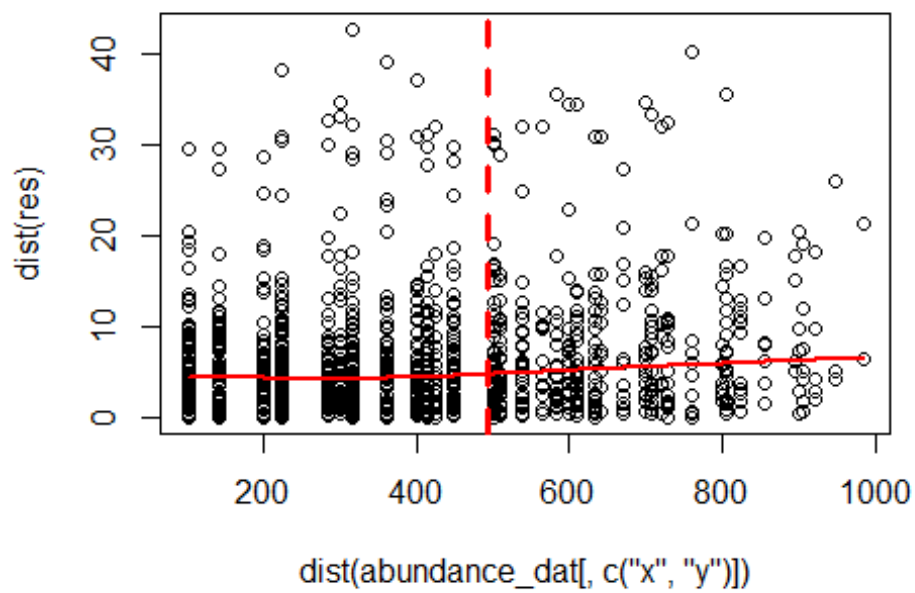
```

# single species
par(mfrow=c(1,1))
plot(Variogram(single_lm, form= ~ x + y))

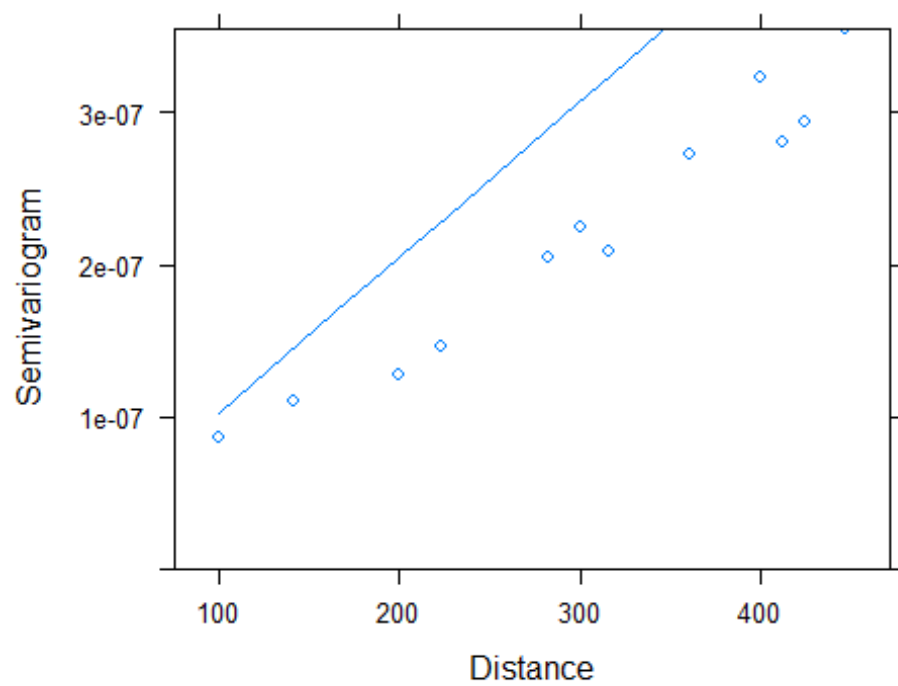
```



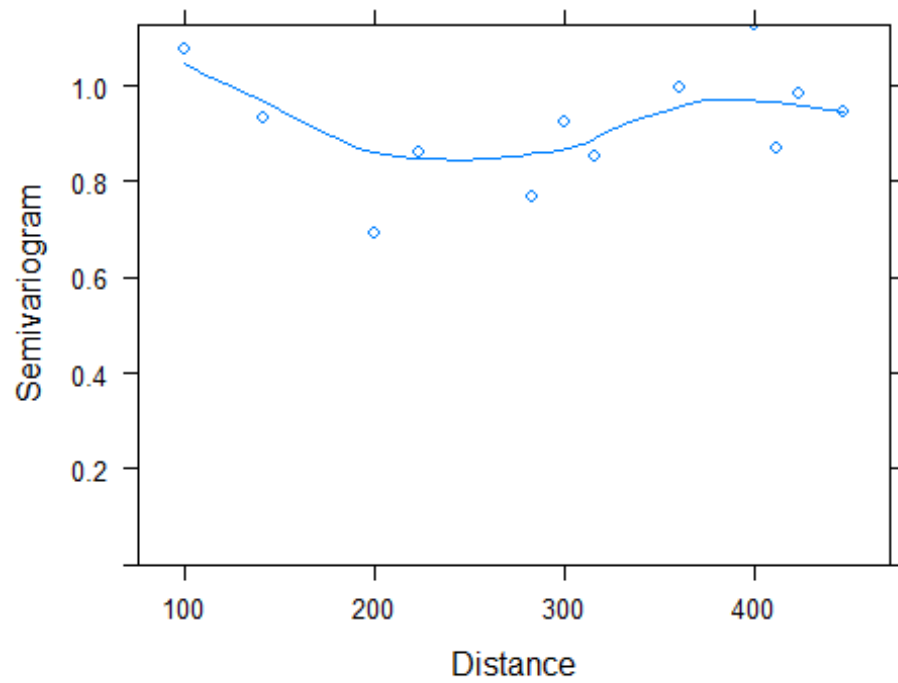
```
res = residuals(single_lm)
plot(dist(abundance_dat[, c('x', 'y')]), dist(res))
lines(lowess(dist(abundance_dat[, c('x', 'y')]), dist(res)), col='red',
lwd=2)
abline(v = max_dist, col='red', lwd=3, lty=2)
```



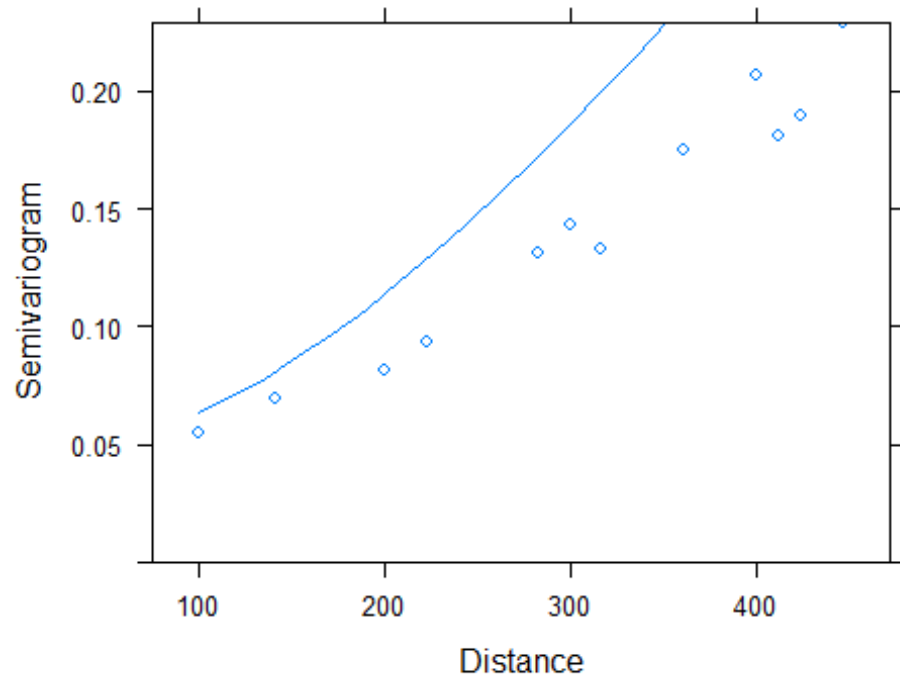
```
single_exp = update(single_lm, corr=corExp(form=~x + y))
plot(Variogram(single_exp, maxDist = max_dist))
```



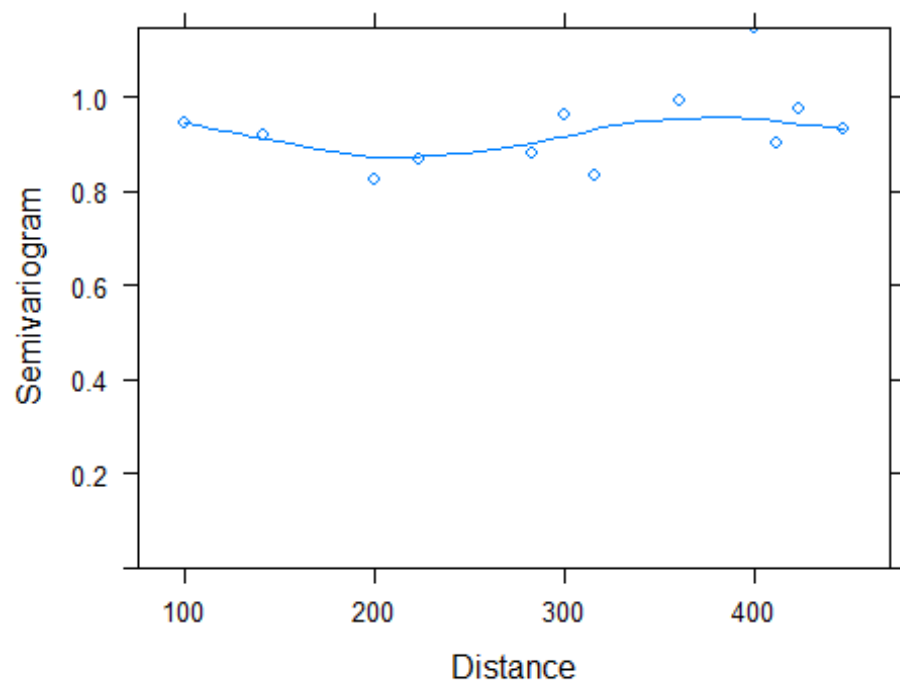
```
plot(Variogram(single_exp, resType='normalized', maxDist = max_dist))
```



```
single_rat_nug = update(single_lm, corr=corRatio(form=~x + y, nugget=T))  
plot(Variogram(single_rat_nug, maxDist = max_dist))
```



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



```
anova(single_lm, single_exp, single_rat_nug, test=F)
```

```
##           Model df      AIC      BIC    logLik
## single_lm      1  3 335.1246 340.7382 -164.5623
## single_exp      2  4 310.6438 318.1286 -151.3219
## single_rat_nug   3  5 308.6589 318.0149 -149.3295
```

`summary(single_lm)`

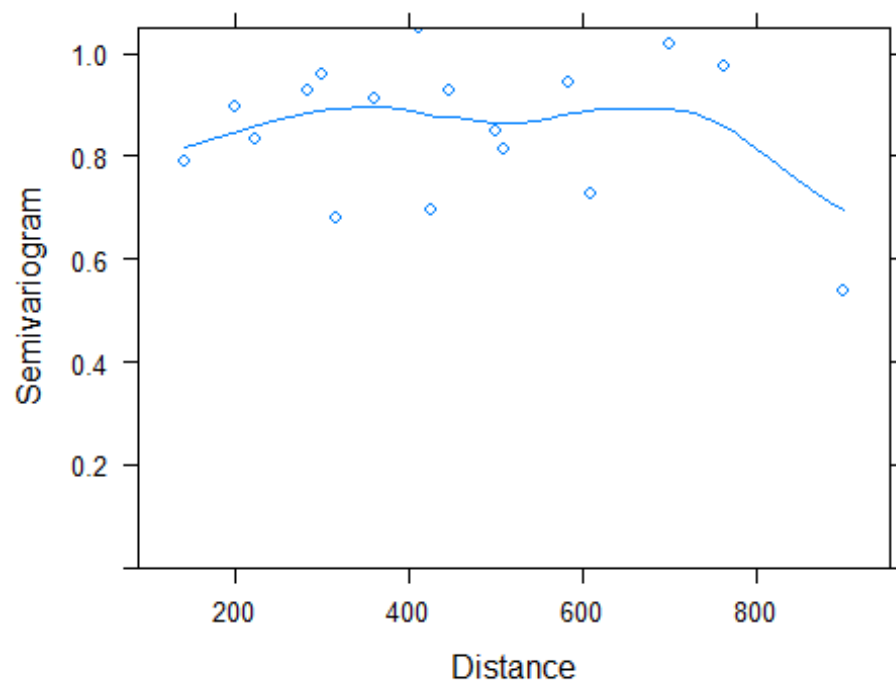
```
## Generalized least squares fit by REML
##   Model: sp_a ~ sp_b
##   Data: abundance_dat
##           AIC      BIC    logLik
##   335.1246 340.7382 -164.5623
##
## Coefficients:
##              Value Std.Error   t-value p-value
## (Intercept) -2.782803 1.9173184 -1.451404  0.1532
## sp_b         1.165220 0.2292289  5.083217  0.0000
##
## Correlation:
##   (Intr)
## sp_b -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((single_rat_nug))`

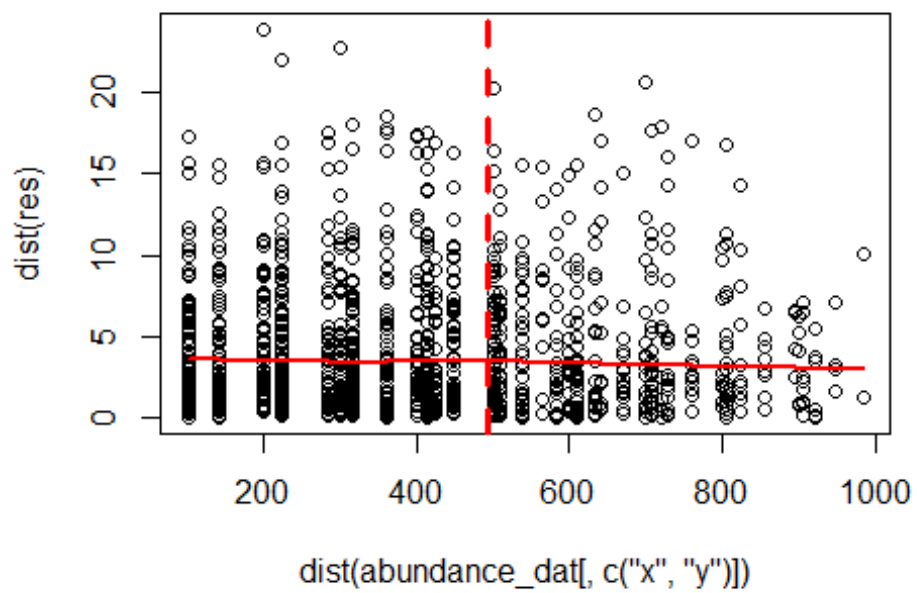
```
## Generalized least squares fit by REML
##   Model: sp_a ~ sp_b
##   Data: abundance_dat
##           AIC      BIC    logLik
##   308.6589 318.0149 -149.3295
##
## Correlation Structure: Rational quadratic spatial correlation
## Formula: ~x + y
## Parameter estimate(s):
##           range      nugget
## 721.93442068  0.04545983
##
## Coefficients:
##              Value Std.Error   t-value p-value
## (Intercept) 12.879120 14.012990  0.9190843  0.3626
## sp_b         0.104606  0.211529  0.4945249  0.6232
##
## Correlation:
##   (Intr)
## sp_b -0.157
```

```
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -0.7668581 -0.7168281 -0.6175156 -0.3685195  1.3705711
##
## Residual standard error: 18.29516
## Degrees of freedom: 50 total; 48 residual

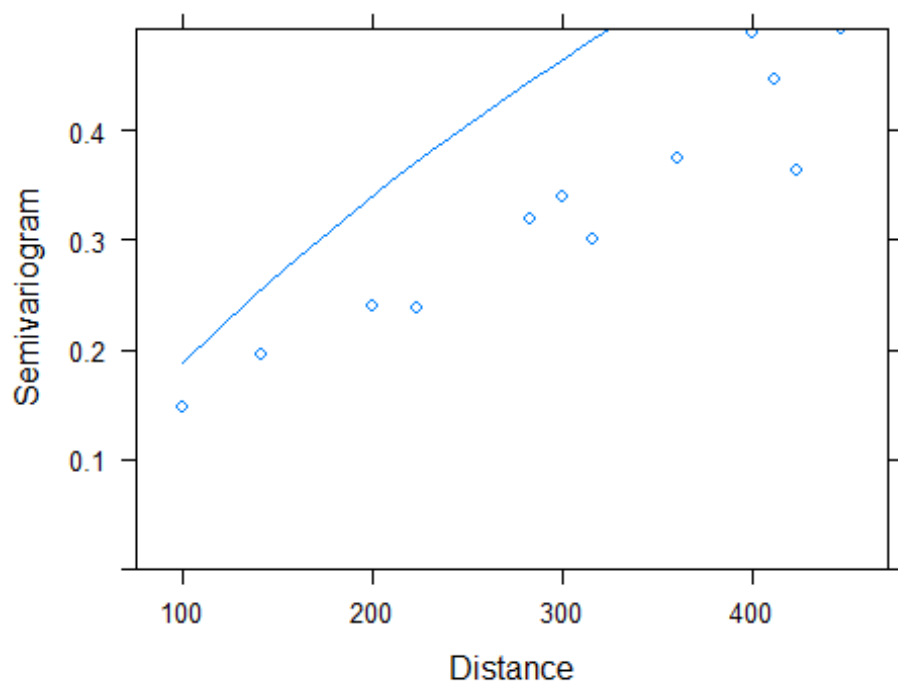
# all species
par(mfrow=c(1,1))
plot(Variogram(all_lm, form= ~ x + y))
```



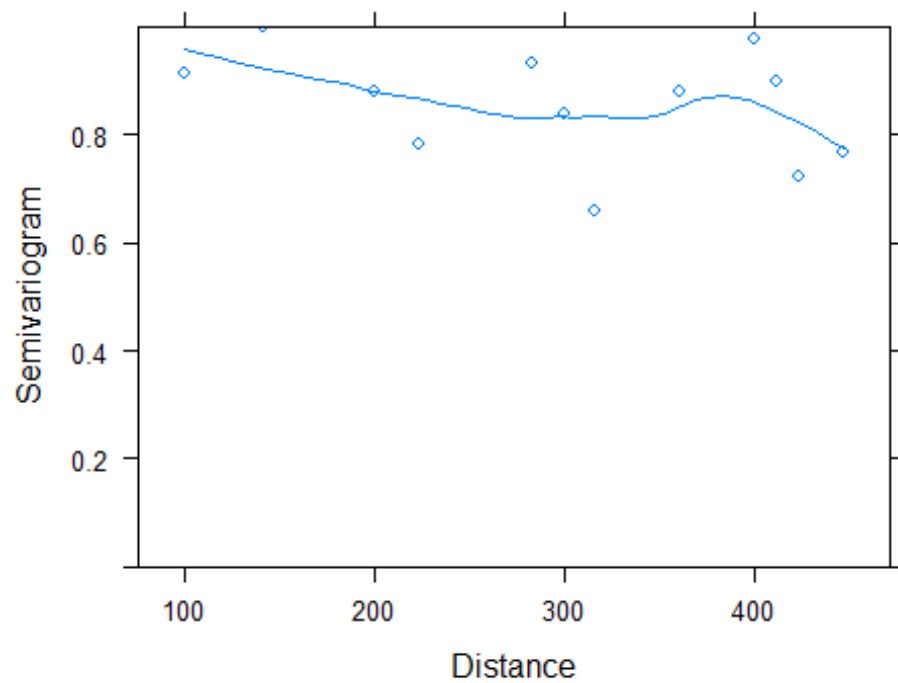
```
res = residuals(all_lm)
plot(dist(abundance_dat[, c('x', 'y')]), dist(res))
lines(lowess(dist(abundance_dat[, c('x', 'y')])), dist(res)), col='red',
      lwd=2)
abline(v = max_dist, col='red', lwd=3, lty=2)
```

```
all_exp = update(all_lm, corr=corExp(form=~x + y))
plot(Variogram(all_exp, maxDist = max_dist))
```

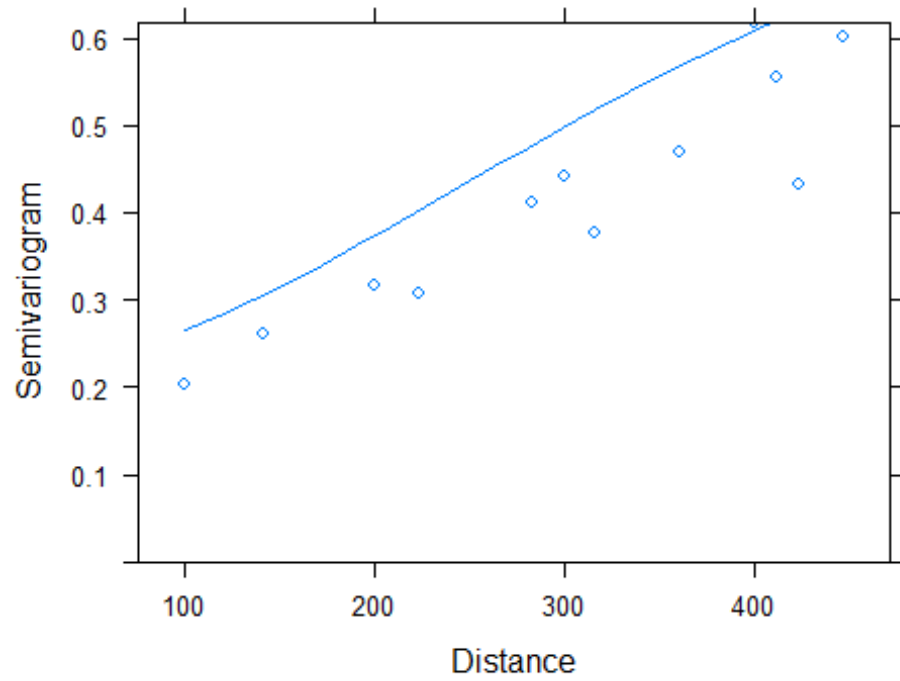


```
plot(Variogram(all_exp, resType='normalized', maxDist = max_dist))
```

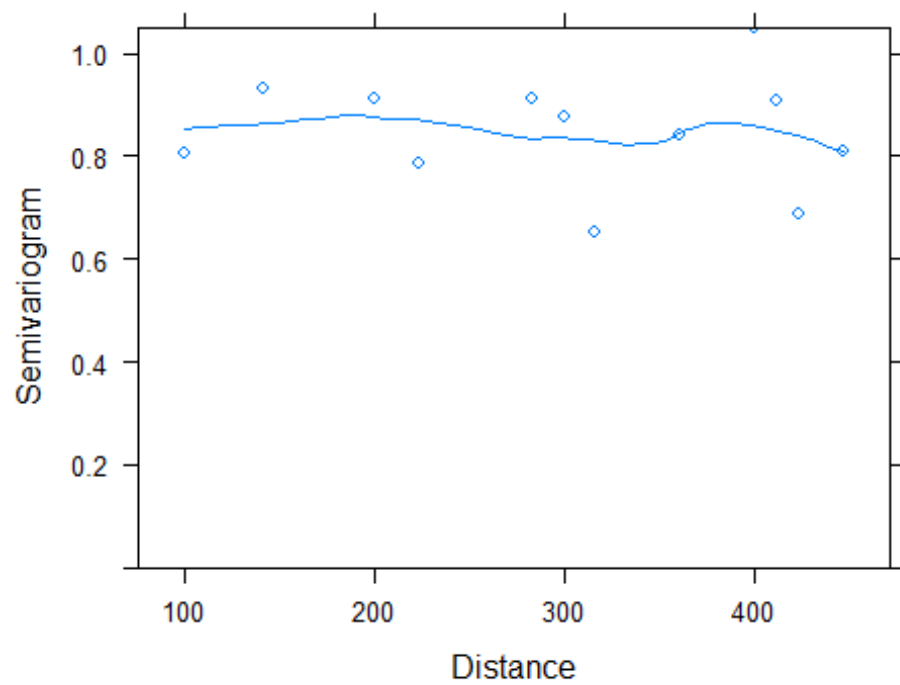


```
all_rat_nug = update(all_lm, corr=corRatio(form=~x + y, nugget=T))
```

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



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



```
anova(all_lm, all_exp, all_rat_nug, test=F)
```

```
##           Model df      AIC      BIC    logLik
## all_lm        1  9 307.1163 322.7554 -144.5582
## all_exp        2 10 301.6062 318.9829 -140.8031
## all_rat_nug    3 11 303.1486 322.2630 -140.5743
```

`summary(all_lm)`

```
## Generalized least squares fit by REML
##   Model: sp_a ~ sp_b + sp_c + sp_d + sp_e + sp_f + sp_g + sp_h
##   Data: abundance_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
## sp_b         0.428920  0.2039316   2.103255  0.0415
## sp_c         0.122279  0.0802638   1.523462  0.1351
## sp_d         0.662259  0.6358905   1.041468  0.3036
## sp_e         4.085661  2.2842770   1.788602  0.0809
## sp_f        -0.249725  0.1491192  -1.674667  0.1014
## sp_g         1.349323  0.7147412   1.887848  0.0660
## sp_h         0.548832  0.1468772   3.736672  0.0006
##
## Correlation:
##   (Intr) sp_b   sp_c   sp_d   sp_e   sp_f   sp_g
## sp_b -0.618
## sp_c -0.212 -0.354
## sp_d  0.025 -0.019 -0.381
## sp_e  0.163 -0.378  0.307 -0.302
## sp_f -0.708  0.245  0.163 -0.113  0.148
## sp_g -0.139  0.187 -0.311  0.308 -0.708 -0.144
## sp_h -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(all_rat_nug)`

```
## Generalized least squares fit by REML
##   Model: sp_a ~ sp_b + sp_c + sp_d + sp_e + sp_f + sp_g + sp_h
##   Data: abundance_dat
##           AIC      BIC    logLik
##   303.1486 322.263 -140.5743
##
## Correlation Structure: Rational quadratic spatial correlation
## Formula: ~x + y
```

```
## Parameter estimate(s):
##      range      nugget
## 402.2077831    0.2194023
##
## Coefficients:
##              Value Std.Error   t-value p-value
## (Intercept) 2.0306920  5.171732  0.3926522  0.6966
## sp_b         0.1508099  0.194940  0.7736210  0.4435
## sp_c         0.0076692  0.091987  0.0833720  0.9340
## sp_d         0.2509289  0.539635  0.4649976  0.6443
## sp_e         1.5049423  1.960799  0.7675147  0.4471
## sp_f         0.0322219  0.142012  0.2268964  0.8216
## sp_g         1.7698936  0.583930  3.0310015  0.0042
## sp_h         0.4058061  0.161181  2.5177087  0.0157
##
## Correlation:
##      (Intr) sp_b    sp_c    sp_d    sp_e    sp_f    sp_g
## sp_b -0.273
## sp_c -0.272 -0.122
## sp_d  0.017  0.038 -0.387
## sp_e -0.039 -0.304  0.337 -0.213
## sp_f -0.242 -0.029  0.166 -0.201  0.106
## sp_g -0.090  0.163 -0.272  0.271 -0.646 -0.036
## sp_h -0.095 -0.055 -0.073 -0.035  0.295  0.143 -0.164
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -1.0645964 -0.5625847 -0.3564599  0.2563989  2.6599508
##
## Residual standard error: 7.348772
## Degrees of freedom: 50 total; 42 residual
```

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

Yes, the models improved when including spatial error and the coefficients were affected.

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

Yes.

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

Including spatial error accounts the potential non-independence of samples that are closer together.