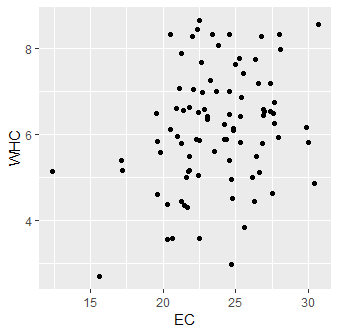
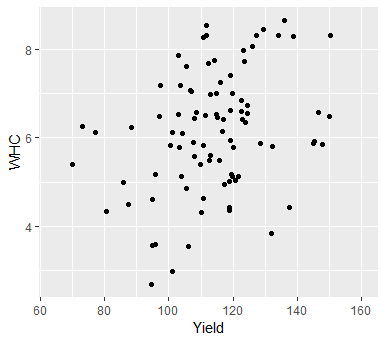
Homework 5

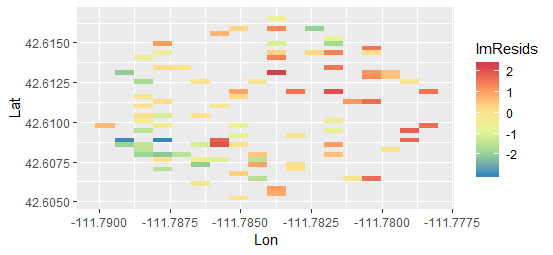
Jonah Meherg

1. Exploratory Graphs

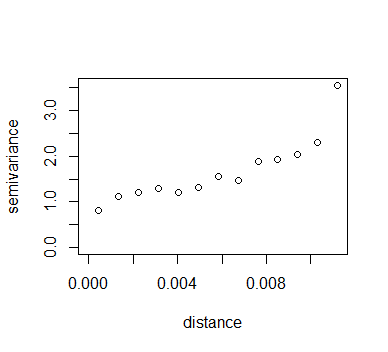


There appears to be some form of a linear relationship between WHC and Yield and EC. Although there does appear a need for a modification to the linear model.

2. Map of Residuals



Plot of Variogram



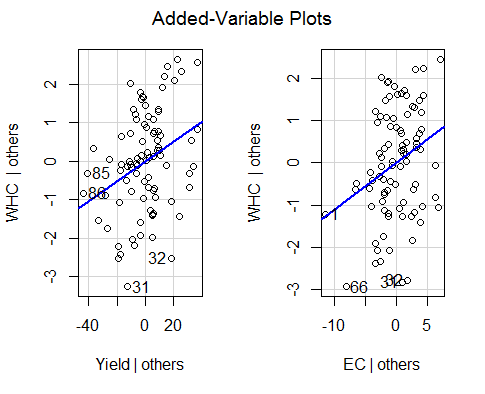
3. After fitting the tree different models using the exponential, spherical, and gaussian correlation models the AICs are 272.3653, 272.9623, and 273.4355 respectively. This suggests the use of an exponential correlation matrix to account for spatial correlation.

4. The model used to fit is a spatial MLR model that is shown by:

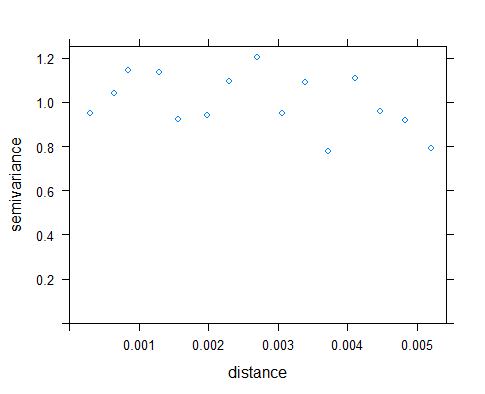
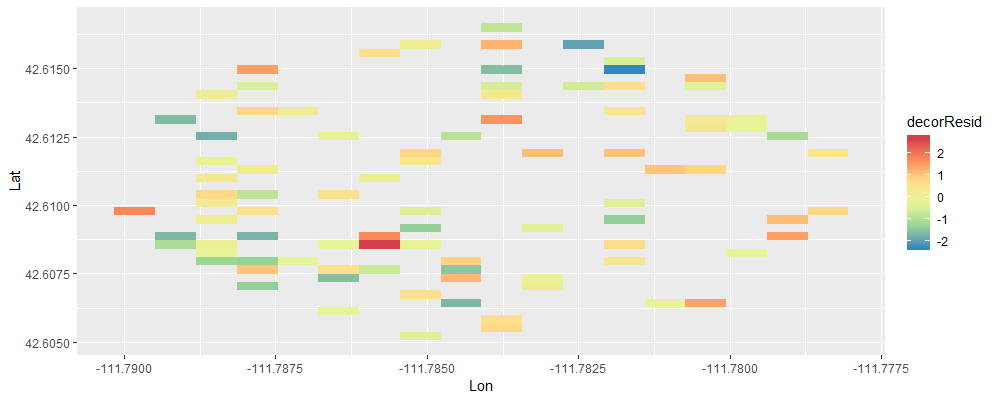
y ~ *N* (Xß, σ2R)

where y is the response variable distributed normally with a mean of Xß and covariance σ2R. Where X is the n x 3 design matrix including a term for the intercept, yield, and HC values. ß are the linear regression coefficients for intercept, yield, and HC. σ2 is the variance and R is the n x n exponential correlation matrix with 1 down the diagonal and ρ(si, sj) where it is the exponential correlation at location i and j where ρ(si, sj) = exp(-||si – sj|| / Φ) where Φ is the range parameter.

5. Linearity seen in added variable plots.

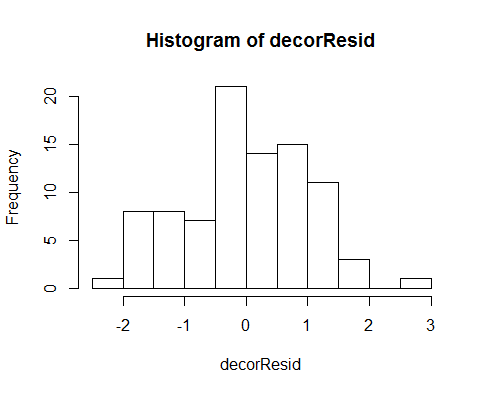


Independence seen from a mapping of the decorrelated residuals and variogram.

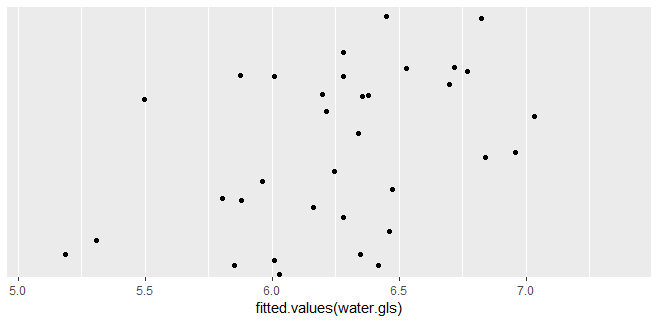
 

The variogram appears to be linear enough to assume that independence has been reached.

Normality is seen from the normality of the decorrelated residuals as evidenced by the histogram of the decorrelated residuals.

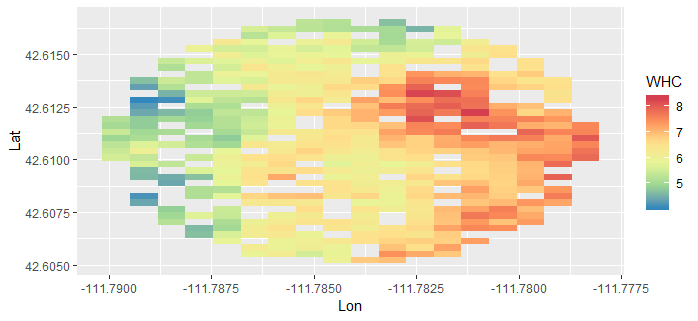


Equal variance is seen through a plot of the fitted values against the decorrelated residuals.

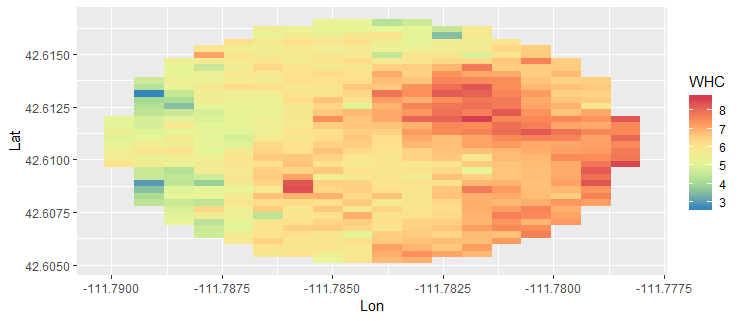


6. A hypothesis test is performed by using a GLHT using a linear factor of (0, 1, 0) to look at just the affect that crop yield has on the WHC. The test results in an estimate of 0.02578 for the beta value of Yield with a 95% confidence interval between 0.00724 and 0.04432. The confidence interval not crossing zero shows that higher crop yield results in higher water holding capacity.

7. The following are WHC predictions at all locations with missing values.



The following shows the predictions with the original values together.



Appendix

Code:

library(ggplot2)

library(geoR)

library(tidyverse)

library(nlme)

library(car)

library(gstat)

library(multcomp)

waterData <- read.table("WaterHoldingCapacity.txt", header=TRUE)

#1

#Exploratory Graphs

ggplot(data=waterData, mapping=aes(x=Yield, y=WHC)) + geom\_point()

ggplot(data=waterData, mapping=aes(x=EC, y=WHC)) + geom\_point()

#There appears to be some form of a linear relationship between WHC and Yield and EC.

#2

waterpredict <- waterData %>% filter(is.na(WHC))#fix NAs dropping

waterDataNoNA <- waterData %>% filter(!is.na(WHC))

water.lm <- lm(WHC~Yield + EC, data=waterDataNoNA)

lmResids <- as.numeric(resid(water.lm))

ggplot(data=waterDataNoNA, mapping=aes(x=Lon, y=Lat, fill=lmResids)) + geom\_raster() + scale\_fill\_distiller(palette="Spectral",na.value=NA)

coordinates <- cbind.data.frame(waterDataNoNA$Lon, waterDataNoNA$Lat)

plot(variog(coords= coordinates, data=lmResids))

#There appears to be some spatial correlation that is seen through a non-linear variogram.

#3

model1 <- gls(model=WHC~Yield + EC, data=waterDataNoNA, correlation=corExp(form=~Lon+Lat, nugget=TRUE), method="ML")

model2 <- gls(model=WHC~Yield + EC, data=waterDataNoNA, correlation=corSpher(form=~Lon+Lat, nugget=TRUE), method="ML")

model3 <- gls(model=WHC~Yield + EC, data=waterDataNoNA, correlation=corGaus(form=~Lon+Lat, nugget=TRUE), method="ML")

summary(model1) #AIC 272.3653

summary(model2) #AIC 272.9623

summary(model3) #AIC 273.4355

#Best coorelation model to use is an exponential.

#4 Write out model.

#5

water.gls <- gls(model=WHC~Yield + EC, data=waterDataNoNA, correlation=corExp(form=~Lon+Lat, nugget=TRUE), method="ML")

avPlots(water.lm)

source("stdres.gls.R")

decorResid <- stdres.gls(water.gls)

residDF <- data.frame(Lon=waterDataNoNA$Lon, Lat=waterDataNoNA$Lat, decorrResid=decorResid)

residVariogram <- variogram(object=decorrResid~1, locations=~Lon+Lat, data=residDF)

plot(residVariogram)

ggplot(data=waterDataNoNA, mapping=aes(x=Lon, y=Lat, fill=decorResid)) + geom\_raster() + scale\_fill\_distiller(palette="Spectral",na.value=NA)

hist(decorResid)

ggplot() + geom\_point(mapping=aes(x=fitted.values(water.gls)), y=decorResid)

#6

a <- c(1, 1, 0) - c(1, 0, 0)

a.transpose <- t(as.matrix(a))

my.test <- glht(water.gls, linfct=a.transpose, alternative="two.sided")

confint(my.test, 0.95)

#A 95% confidence interval between 0.00724 and 0.04432 with an estimate of 0.02578. This shows that locations with higher yield had higher WHC.

#7

source("predictgls.R")

tempPreds <- predictgls(water.gls, newdframe = waterpredict)

waterpredict$WHC <- tempPreds$Prediction

ggplot(data=waterpredict, mapping=aes(x=Lon, y=Lat, fill=WHC)) + geom\_raster() + scale\_fill\_distiller(palette="Spectral") #Predictions only.

waterFinal <- rbind.data.frame(waterpredict, waterDataNoNA)

ggplot(data=waterFinal, mapping=aes(x=Lon, y=Lat, fill=WHC)) + geom\_raster() + scale\_fill\_distiller(palette="Spectral") #Predictions with given values.