**FOREST MODELS – homework 8**

**Set of models to project and initialize stand basal area of eucalyptus stands in Portugal**

Data file: “Ec\_StandGrowthData\_WithS.xlsx”

1. First you must read and prepare the data. The difference between this data file and the one used in the classes is that there is an additional sheet with the information needed to estimate site index (S) and also that there is less data points as the information needed to estimate site index was not available for some plots. S must be estimated from the measurement closest to 10 years using the GLOBULUS site index curves:

$$hdom\_{2}=\left(29.0669+0.2880 NdaysPrec\right)\left(\frac{hdom\_{1}}{29.0669+0.2880 NdaysPrec}\right)^{\left(\frac{t\_{1}}{t\_{2}}\right)^{0.4890}}$$

1. Once the data are prepared, you will have to use the *lag()* function to “create” the data pairs (see an example how this can be done in the code below, but you may find a better way!). Just like you did in homework\_3 with EXCEL.
2. When the data will be prepared, start by fitting the Lundqvist-k function (that was used in the GLOBULUS model)
3. Check if expressing the parameters as a function of some stand and/or site variables (*e.g.* number of trees at plating, site index) will improve the fitting and prediction abilities of the model
4. Check the regression assumptions
5. Evaluate the prediction ability of the models obtained (bias and precision) using several statistics based on the press residuals
6. Evaluate now the prediction ability of the models obtained (bias and precision) using several statistics based on cross validation
7. Check if there are tendencies of the residuals with some stand and/or site variables (stand density, stand age, site index, Region)
8. Now fit the integral form of the Lundqvist function, fixing the parameters to the values and parameters obtained for the difference equation form. This means that you just have to find the value and expression for the k parameter.

Code for data preparation (just a suggestion):

################################################################################

# read data into several dataframes ("dados")one per worksheet)

# join and merge them in order to obtain a unique data set

dados <- "Ec\_StandGrowthData\_WithS.xlsx"

AV<-read.xlsx(dados, header= TRUE, sheetIndex = "AV\_data")

AV\_plot<-read.xlsx(dados, header= TRUE, sheetIndex = "AV\_plot")

EPE<-read.xlsx(dados, header= TRUE, sheetIndex = "EPE\_data")

EPE\_plot<-read.xlsx(dados, header= TRUE, sheetIndex = "EPE\_plot")

str(AV)

str(AV\_plot)

str(EPE)

str(EPE\_plot)

# combine data vertically

# join two files one over the other

Ec<-rbind(AV,EPE)

length(Ec$hdom)

Ec\_plot<-rbind(AV\_plot,EPE\_plot)

# combine data horizontally

Ec <- inner\_join(x=Ec,y=Ec\_plot,

 by=c("ID\_plot"="ID\_plot","ID\_rot"="ID\_rot"))

# calculute stand age from the date of regeneration and the date of measurement

Ec$t<-as.numeric((Ec$Dt-Ec$Dt\_regenera))/365

################################################################################

# read information needed to estimate site index and merge it into Ec

Sdata<-read.xlsx(dados, header= TRUE, sheetIndex = "DataForS")

str(Sdata)

Ec <- inner\_join(x=Ec,y=Sdata,by=c("ID\_plot"="id\_parc"))

# note that the merge was made with variables that do not have the same name

################################################################################

# use the age closest to 10 years to estimate site index of each plot

# and add this information into the datafile Ec

Ec$dif10 <- abs(Ec$t-10)

Ec\_dif10 <- aggregate(dif10 ~ ID\_plot+ID\_rot, data=Ec, FUN=min)

names(Ec\_dif10) <- c("ID\_plot","ID\_rot","dif10\_min")

Ec <- merge (x=Ec, y=Ec\_dif10, by=c("ID\_plot","ID\_rot"))

Ec\_S <- filter(Ec,dif10==dif10\_min)

Ec\_S$S <- with(Ec\_S,(29.0669+0.2880\*NdaysPrec)\*(hdom/(29.0669+0.2880\*NdaysPrec))\*\*((t/10)\*\*0.4890))

Ec\_S <- select(Ec\_S,ID\_plot,ID\_rot,S)

Ec <- merge (x=Ec, y=Ec\_S, by=c("ID\_plot","ID\_rot"))

###############################################################################

# select just some variables

Ec <- select(Ec,c("ID\_plot","ID\_rot","t","hdom","N","G","N0","Reg\_Globulus","S"))

head(Ec,n=20)

################################################################################

# create variables that are lagged with the lag function

# first we need to sort the file by the plot and rot identifiers and age

Ec<-with(Ec,Ec[order(ID\_plot,ID\_rot,t),])

Ec\_lag <- lag(Ec)

Ec\_lag <- select (Ec\_lag,-c("Reg\_Globulus","N0","S"))

Ec\_lag <- rename (Ec\_lag,"ID\_plot1"="ID\_plot","ID\_rot1"="ID\_rot","t1"=t,"hdom1"=hdom,"G1"=G,"N1"=N)

Ec <- rename (Ec,"t2"=t,"hdom2"=hdom,"G2"=G,"N2"=N)

Ec <- cbind(Ec,Ec\_lag)

Ec <- filter(Ec,ID\_plot==ID\_plot1 & ID\_rot==ID\_rot1)

Ec <- select (Ec, -c("ID\_plot1","ID\_rot1"))

Ec <- with(Ec,data.frame(ID\_plot,ID\_rot,Reg\_Globulus,S,N0,t1,t2,hdom1,hdom2,N1,N2))