### BATCHELOR No Analogue assignment

#load the proper libraries:

library(analogue)

library(rioja)

library(neotoma)

data(Climate)

data(Pollen)

data(Location)

########

Doing part I: “Obtain a fossil pollen dataset for eastern North America from Neotoma. Do **standard data handling: “ensure that the taxon list and order matches exactly to the North American Pollen Dataset…”**

#got a NY fossil record from White Lake, NY:

ny.pollen.raw=read.csv("White\_Lake\_NY\_fossil.csv")

#skip the first six headers

fossil.pollen<-read.csv("White\_Lake\_NY\_fossil.csv",skip=7,header=FALSE)

#skip first five columns

fossil.pollen.clean<-fossil.pollen[-1:-5];

fossil.pollen.new=t(fossil.pollen.clean);

#assign taxa column names

colnames(fossil.pollen.new)=fossil.pollen[,1]

depths=ny.pollen.raw[2,6:49]

depths\_t=t(depths)

rownames(fossil.pollen.new)=depths\_t

fossil.pollen.corrected<- replace(fossil.pollen.new,is.na(fossil.pollen.new),0) #actually doing the replacment of NA values

counts.final.fossil=fossil.pollen.corrected;

percentages.fossil=100\*counts.final.fossil/rowSums(counts.final.fossil)

#condense the list to only a few pollen types (to match the Pollen)

colnums=c(1,19,23,31,35)

percentages.ny=percentages.fossil[,(colnums)]

###Compile modern, only chose certain Pollen lists

colnums\_Pollen=c(1,42,55,78,88)

modpoll=Pollen[,(colnums\_Pollen)]

counts.final.modern=modpoll

percentages.modern=100\*counts.final.modern/rowSums(counts.final.modern)

percentages.modern<- replace(percentages.modern,is.na(percentages.modern),0) #actually doing the replacment of NA values

colnames(percentages.modern)[4]<-'Myrica'

colnames(percentages.modern)[5]<-'Picea'

##---- ny.poll should have all the percents of pollen within the core

#loading this library because I need to use the 'pl()' function

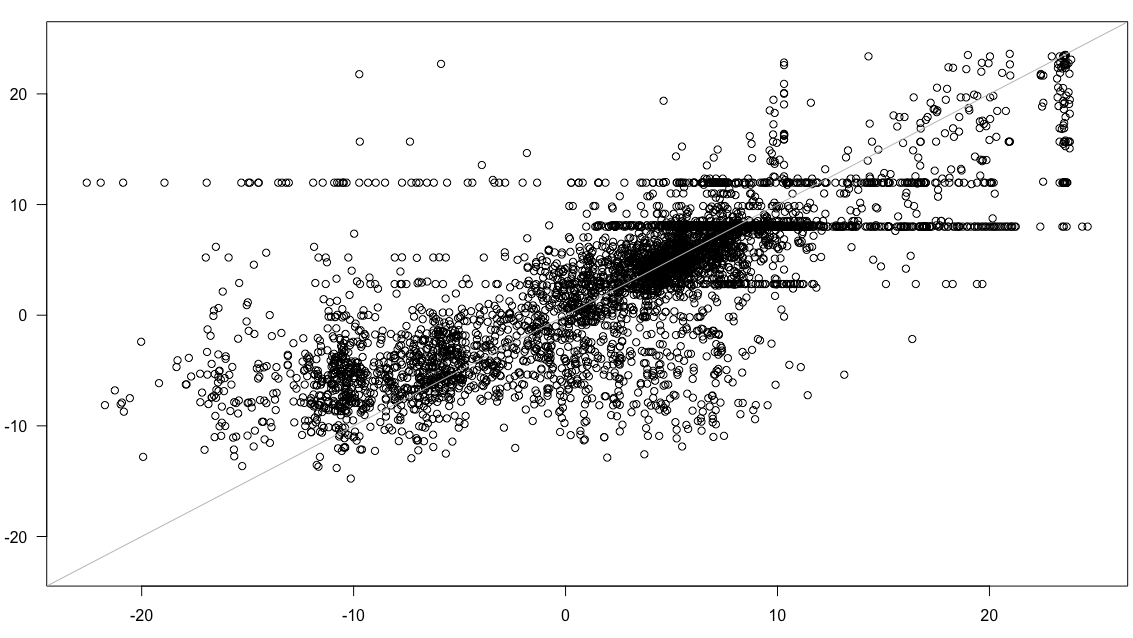
library(astrochron)

#MAT of t\_average and percentages modern

model\_ny\_tave=rioja::MAT(percentages.modern,Climate$tave,k=5,lean=FALSE)

pl()

plot(model\_ny\_pollen)

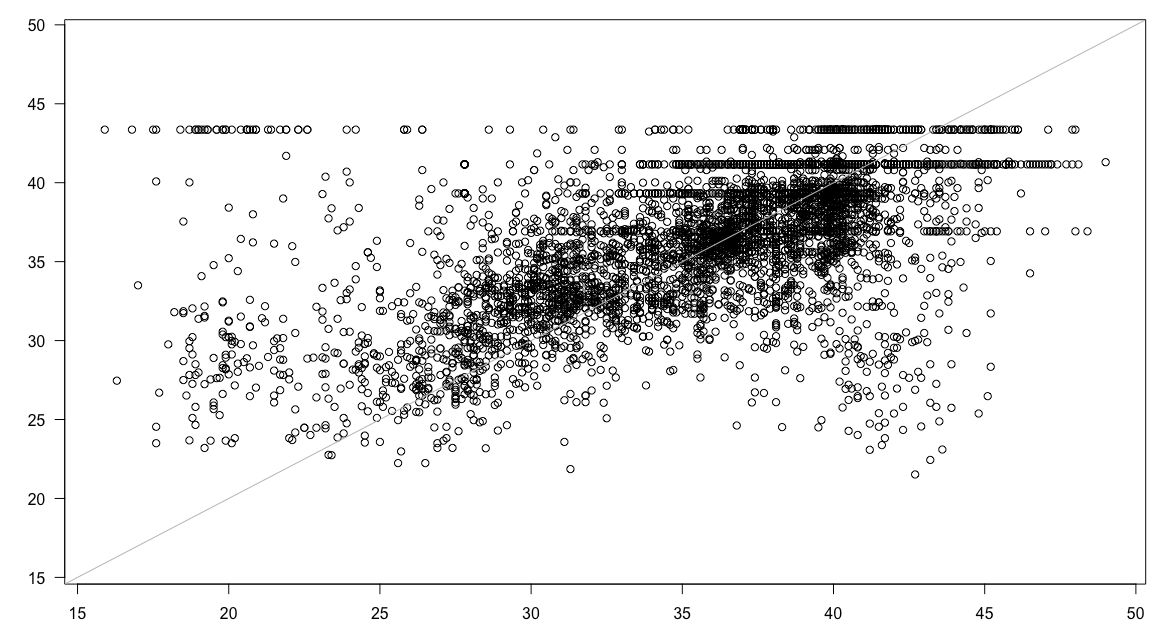


#MAT of t\_maximum and percentages modern

model\_ny\_tmax=rioja::MAT(percentages.modern,Climate$tmax,k=5,lean=FALSE)

pl()

plot(model\_ny\_tmax)

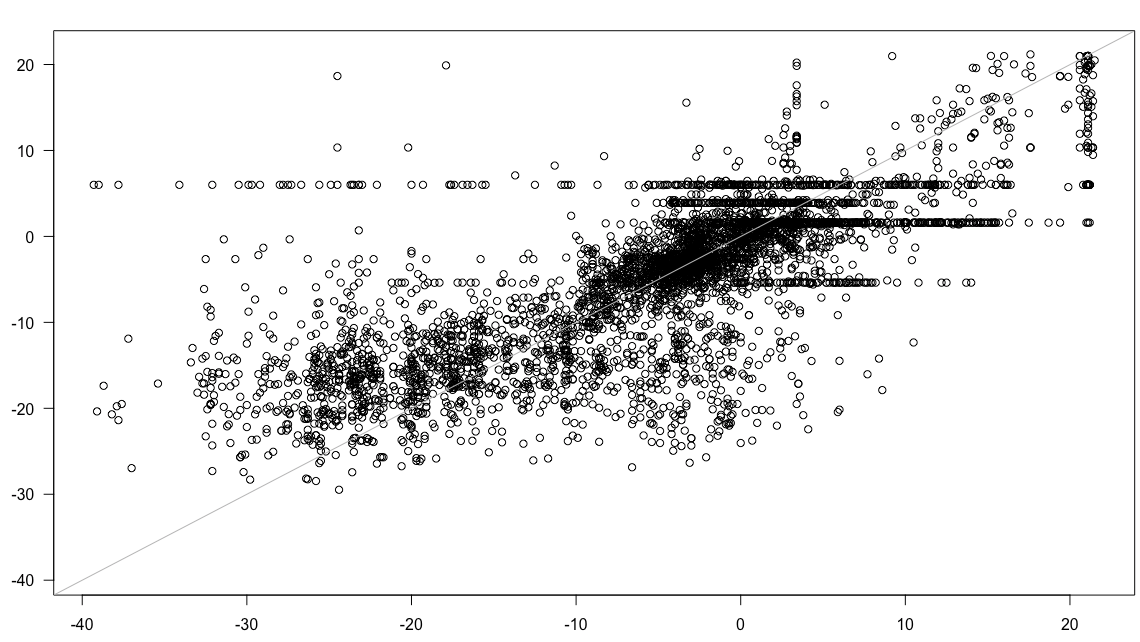


#MAT of latitude and percentages modern

model\_ny\_lat=rioja::MAT(percentages.modern,Climate$lat,k=5,lean=FALSE)

pl()

plot(model\_ny\_pollen)



temp\_ave=predict(model\_ny\_tave,percentages.fossil)

RESULT:

$RMSE0

[1] 8.239153

$object

RMSE R2 Avg.Bias Max.Bias Skill

N01 6.835022 0.4945369 -1.355503828 20.78088 31.17999

N02 5.933666 0.5240156 -0.323729223 17.65319 48.13420

N03 5.533698 0.5599193 -0.004658942 17.72467 54.89074

N04 5.492628 0.5645290 -0.115802383 18.06483 55.55784

N05 5.398532 0.5747715 -0.053533692 17.67304 57.06750

N01.wm 6.835022 0.4945369 -1.355503828 20.78088 31.17999

N02.wm 6.036093 0.5157352 -0.357507768 17.74506 46.32813

N03.wm 5.664443 0.5462519 -0.048106416 17.79990 52.73396

N04.wm 5.624772 0.5522467 -0.154107894 18.03525 53.39369

N05.wm 5.548610 0.5594204 -0.091204139 17.89637 54.64730

$crossval

RMSE R2 Avg.Bias Max.Bias Skill

N01 6.817070 0.4455490 -0.8471860 19.14363 31.54103

N02 6.292837 0.5061360 -0.9468877 18.80368 41.66518

N03 5.794426 0.5356554 -0.4390010 18.85065 50.53983

N04 5.563132 0.5565606 -0.2619245 18.38407 54.40958

N05 5.484710 0.5658449 -0.2590692 17.89559 55.68587

N01.wm 6.817070 0.4455490 -0.8471860 19.14363 31.54103

N02.wm 6.491508 0.4856198 -0.9847928 18.91159 37.92367

N03.wm 6.106469 0.4997894 -0.4824681 19.05110 45.06930

N04.wm 5.926623 0.5128366 -0.3092614 18.84466 48.25726

N05.wm 5.868851 0.5188295 -0.3025909 18.72543 49.26110

---------------

temp\_max=predict(model\_ny\_tmax,percentages.fossil)

RESULT:

$RMSE0

[1] 5.61502

$object

RMSE R2 Avg.Bias Max.Bias Skill

N01 5.703217 0.2476820 -0.9921374 14.27917 -3.166141

N02 5.330845 0.2752938 -0.8179288 13.56042 9.865814

N03 5.031969 0.2841584 -0.5130009 13.93542 19.689306

N04 5.013340 0.2871065 -0.6500828 14.31302 20.282847

N05 4.993647 0.2943742 -0.7694765 14.49583 20.907908

N01.wm 5.703217 0.2476820 -0.9921374 14.27917 -3.166141

N02.wm 5.422644 0.2719097 -0.8381716 13.64466 6.734806

N03.wm 5.130442 0.2835122 -0.5197611 13.60044 16.515282

N04.wm 5.109580 0.2913708 -0.6327715 13.87130 17.192832

N05.wm 5.094630 0.3006255 -0.7398147 14.02818 17.676719

$crossval

RMSE R2 Avg.Bias Max.Bias Skill

N01 6.174988 0.2050713 -1.1841092 13.42500 -20.939929

N02 5.546837 0.2429082 -0.9709290 13.86667 2.413820

N03 5.300807 0.2545158 -0.8348024 14.44444 10.878713

N04 5.116927 0.2678284 -0.7229826 14.56198 16.954541

N05 5.067209 0.2764508 -0.7912849 14.63708 18.560485

N01.wm 6.174988 0.2050713 -1.1841092 13.42500 -20.939929

N02.wm 5.693655 0.2312635 -1.0005256 13.90790 -2.820506

N03.wm 5.512370 0.2377108 -0.8482014 14.35577 3.622841

N04.wm 5.377539 0.2431458 -0.7346549 14.41991 8.279897

N05.wm 5.368403 0.2464117 -0.7970946 14.76037 8.591266

----------

lat=predict(model\_ny\_lat,percentages.fossil):

RESULT:

$RMSE0

[1] 10.97504

$object

RMSE R2 Avg.Bias Max.Bias Skill

N01 9.122095 0.4782768 -1.52453962 30.18000 30.91606

N02 8.017882 0.5090811 -0.31478378 25.69333 46.62876

N03 7.462159 0.5496495 0.01236637 25.80667 53.77074

N04 7.383177 0.5568731 -0.16282330 25.96833 54.74418

N05 7.272776 0.5653927 -0.04976619 25.14000 56.08748

N01.wm 9.122095 0.4782768 -1.52453962 30.18000 30.91606

N02.wm 8.171944 0.4997228 -0.35734355 25.32259 44.55802

N03.wm 7.655328 0.5347886 -0.04380096 25.50815 51.34633

N04.wm 7.583097 0.5431154 -0.20839565 25.85317 52.26013

N05.wm 7.497799 0.5483841 -0.09873443 25.57407 53.32809

$crossval

RMSE R2 Avg.Bias Max.Bias Skill

N01 9.137234 0.4317363 -1.0090834 27.31333 30.68656

N02 8.473200 0.4864778 -1.0384130 27.80000 40.39498

N03 7.890184 0.5136662 -0.4805780 28.64444 48.31528

N04 7.556411 0.5393570 -0.2857335 27.11500 52.59556

N05 7.413848 0.5528236 -0.2978316 26.53733 54.36739

N01.wm 9.137234 0.4317363 -1.0090834 27.31333 30.68656

N02.wm 8.737388 0.4663000 -1.1129690 27.42926 36.62016

N03.wm 8.289686 0.4796944 -0.5449150 28.34593 42.94889

N04.wm 8.043273 0.4956209 -0.3689036 27.36763 46.29020

N05.wm 7.977239 0.5008198 -0.3604823 27.74464 47.16848

#-------------------Doing the Cross Validation:

#tave:

cv.mat.model.tave=rioja::crossval(model\_ny\_tave,verbose=FALSE)

plot(cv.mat.model.tave)

perf.cv.tave.mat.model=rioja::performance(cv.mat.model.tave)

perf.cv.tave.mat.model

cv.mat.model.tave=rioja::crossval(model\_ny\_tave,verbose=FALSE)

per.mat.model.tave=rioja::performance(cv.mat.model.tave)

plot(cv.mat.model.tave)

depths=as.numeric(rownames(percentages.fossil))

pred=predict(model\_ny\_tave,percentages.fossil)

plot(depths,pred$fit[,'MAT'],type='l',ylab='T [C]')

#tmax

cv.mat.model.tmax=rioja::crossval(model\_ny\_tmax,verbose=FALSE)

plot(cv.mat.model.tmax)

perf.cv.tmax.mat.model=rioja::performance(cv.mat.model.tmax)

perf.cv.tmax.mat.model

cv.mat.model.tmax=rioja::crossval(model\_ny\_tmax,verbose=FALSE)

per.mat.model.tmax=rioja::performance(cv.mat.model.tmax)

plot(cv.mat.model.tmax)

depths=as.numeric(rownames(percentages.fossil))

pred=predict(model\_ny\_tave,percentages.fossil)

plot(depths,pred$fit[,'MAT'],type='l',ylab='T [C]')

#latitude

cv.mat.model.lat=rioja::crossval(model\_ny\_lat,verbose=FALSE)

plot(cv.mat.model.lat)

perf.cv.lat.mat.model=rioja::performance(cv.mat.model.lat)

perf.cv.lat.mat.model

cv.mat.model.lat=rioja::crossval(model\_ny\_tmax,verbose=FALSE)

per.mat.model.lat=rioja::performance(cv.mat.model.lat)

plot(cv.mat.model.lat)

depths=as.numeric(rownames(percentages.fossil))

pred=predict(model\_ny\_lat,percentages.fossil)

plot(depths,pred$fit[,'MAT'],type='l',ylab='T [C]')

----------------- Trying another one:

**#call fossil pollen data from a site**

fossil <- read.csv("SutherlandPond.csv")

ages<-fossil$Cal.age.linear

fossilpoll<-fossil [ , 6:81]

**#Steps to create proportions for fossil data**

pollSum<-apply(fossilpoll,1,sum)

pollSum <- ifelse(pollSum==0,1, ifelse(pollSum>0,pollSum,pollSum))

fossilpoll<-fossilpoll/pollSum

nbFosSamp=dim(fossilpoll)[1]

**#run analog**

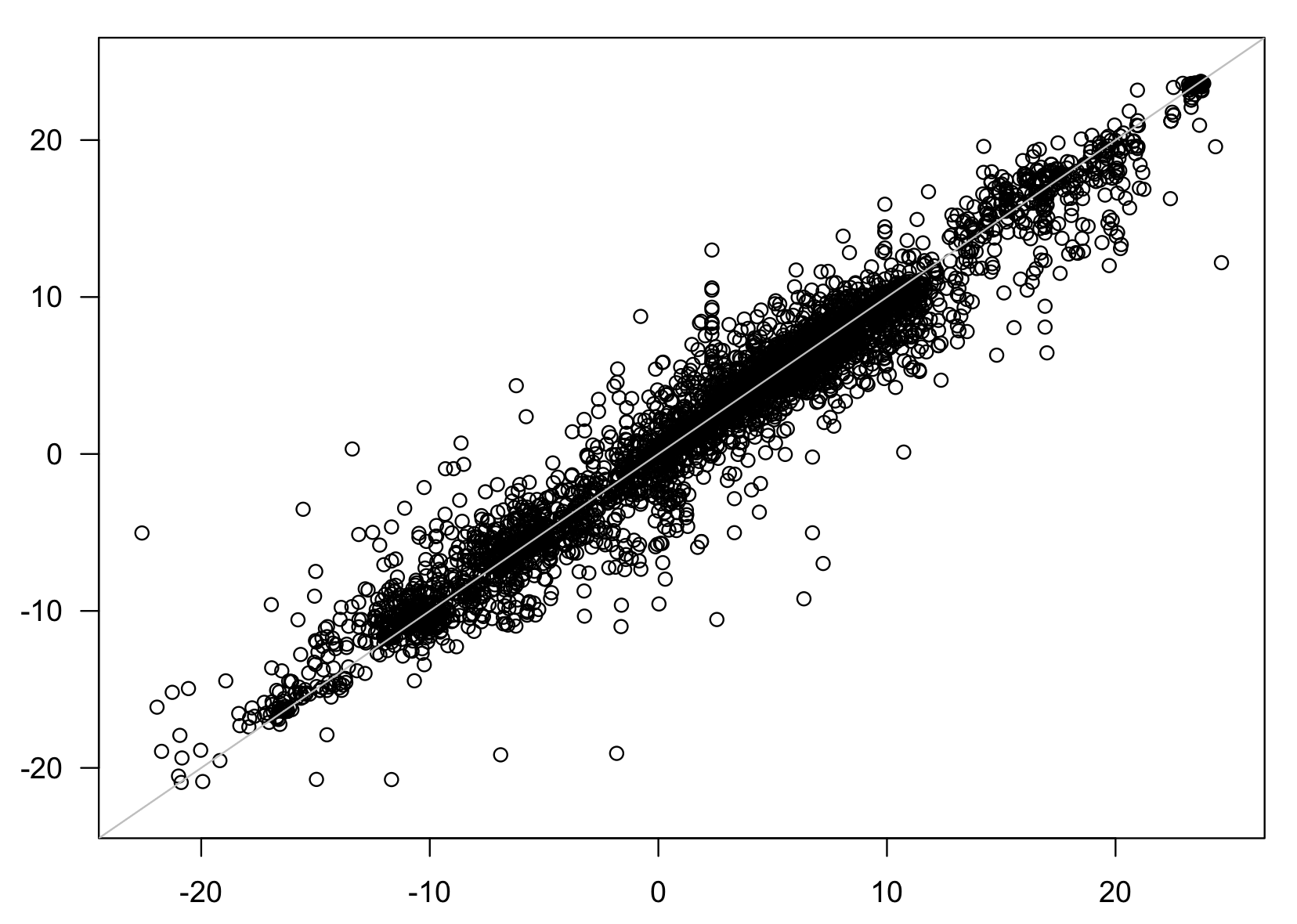
modernanalog<-analog(modpoll, fossilpoll, method = "SQchord")

#MAT of t\_average and percentages modern

model\_ny\_tave=rioja::MAT(modpoll,clim$tave,k=5,lean=FALSE)

pl()

plot(model\_ny\_tave)

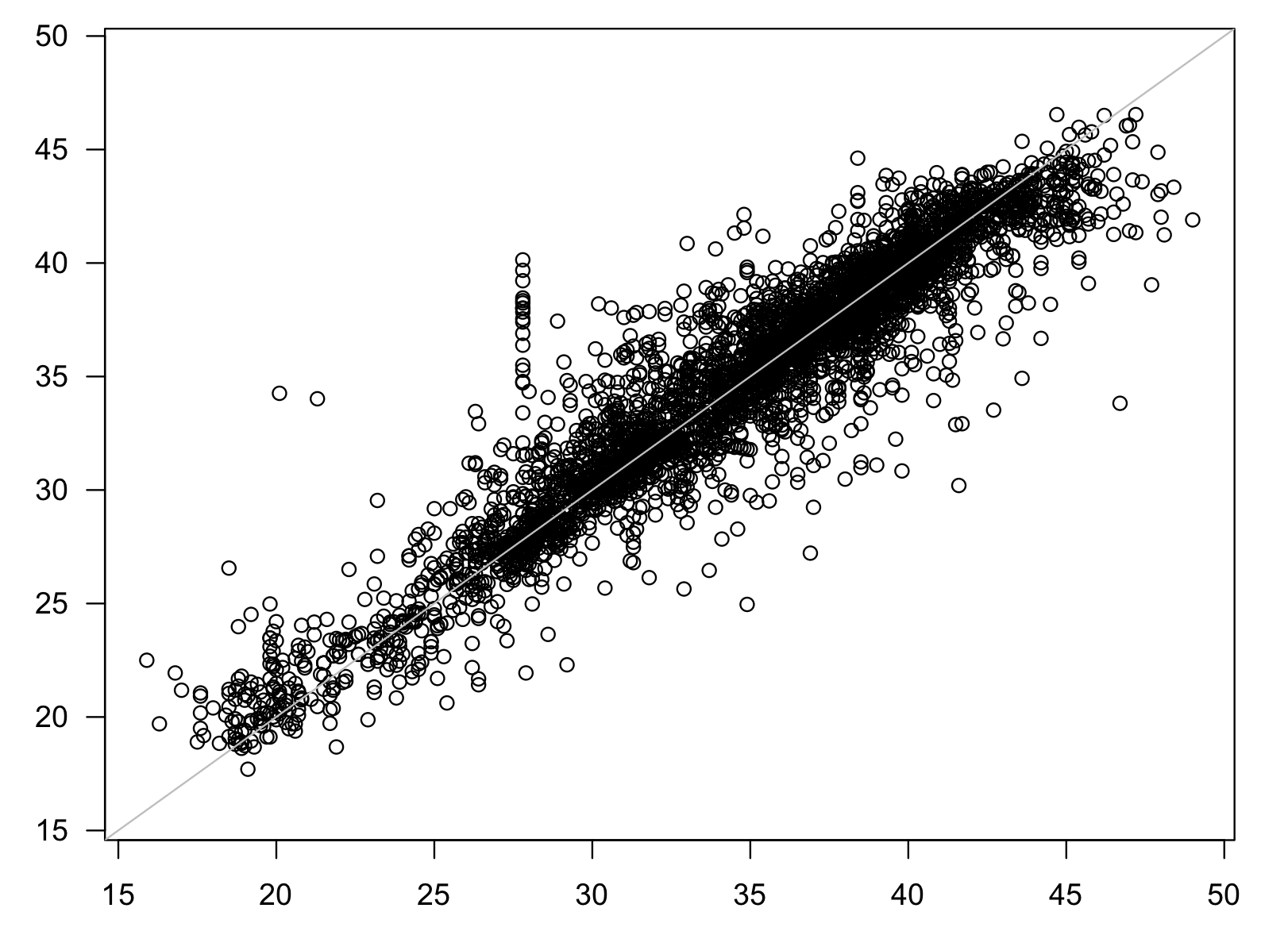


#MAT of t\_maximum and percentages modern

model\_ny\_tmax=rioja::MAT(modpoll,clim$tmax,k=5,lean=FALSE)

pl()

plot(model\_ny\_tmax)

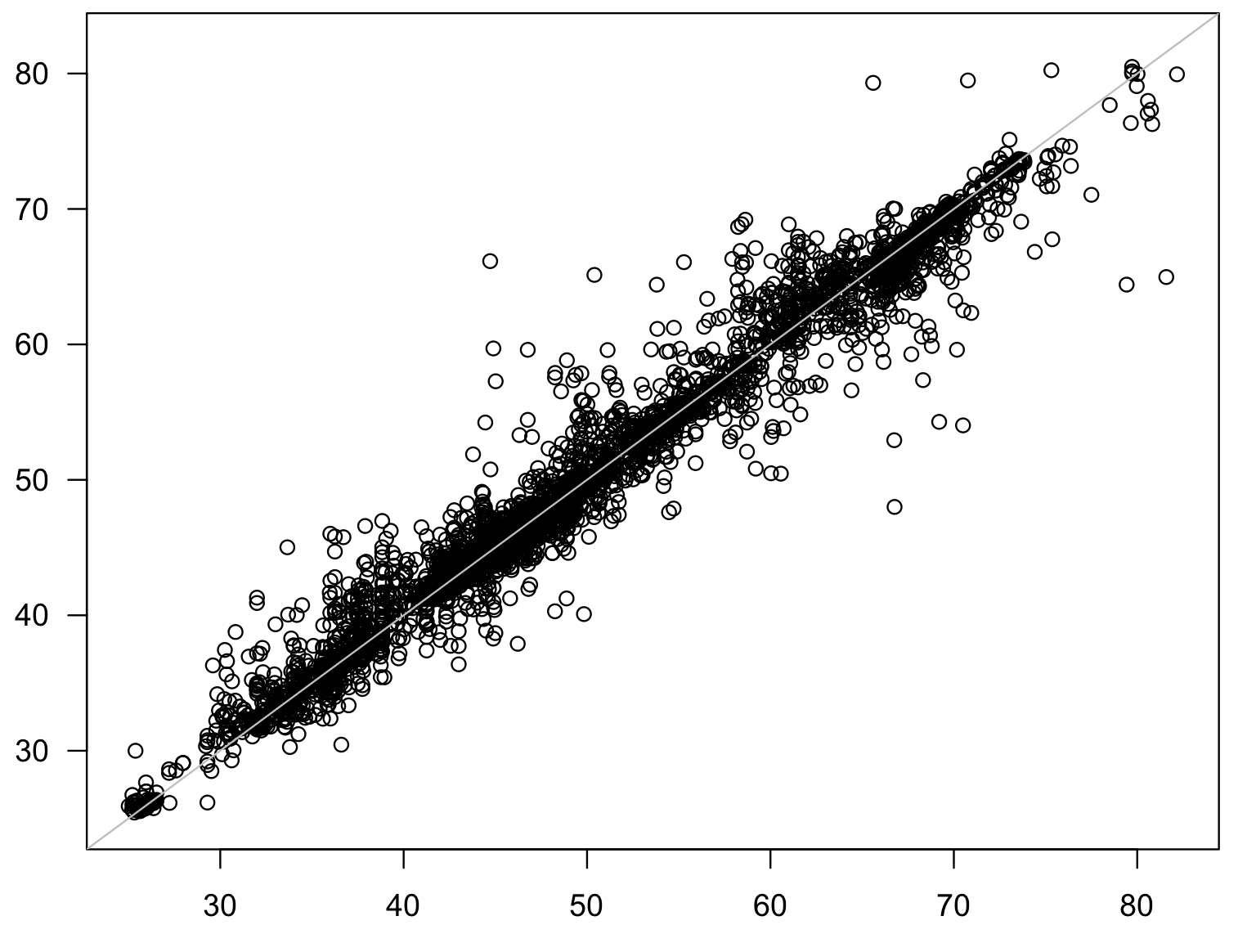


#MAT of latitude and percentages modern

model\_ny\_lat=rioja::MAT(modpoll,clim$LATDD,k=5,lean=FALSE)

pl()

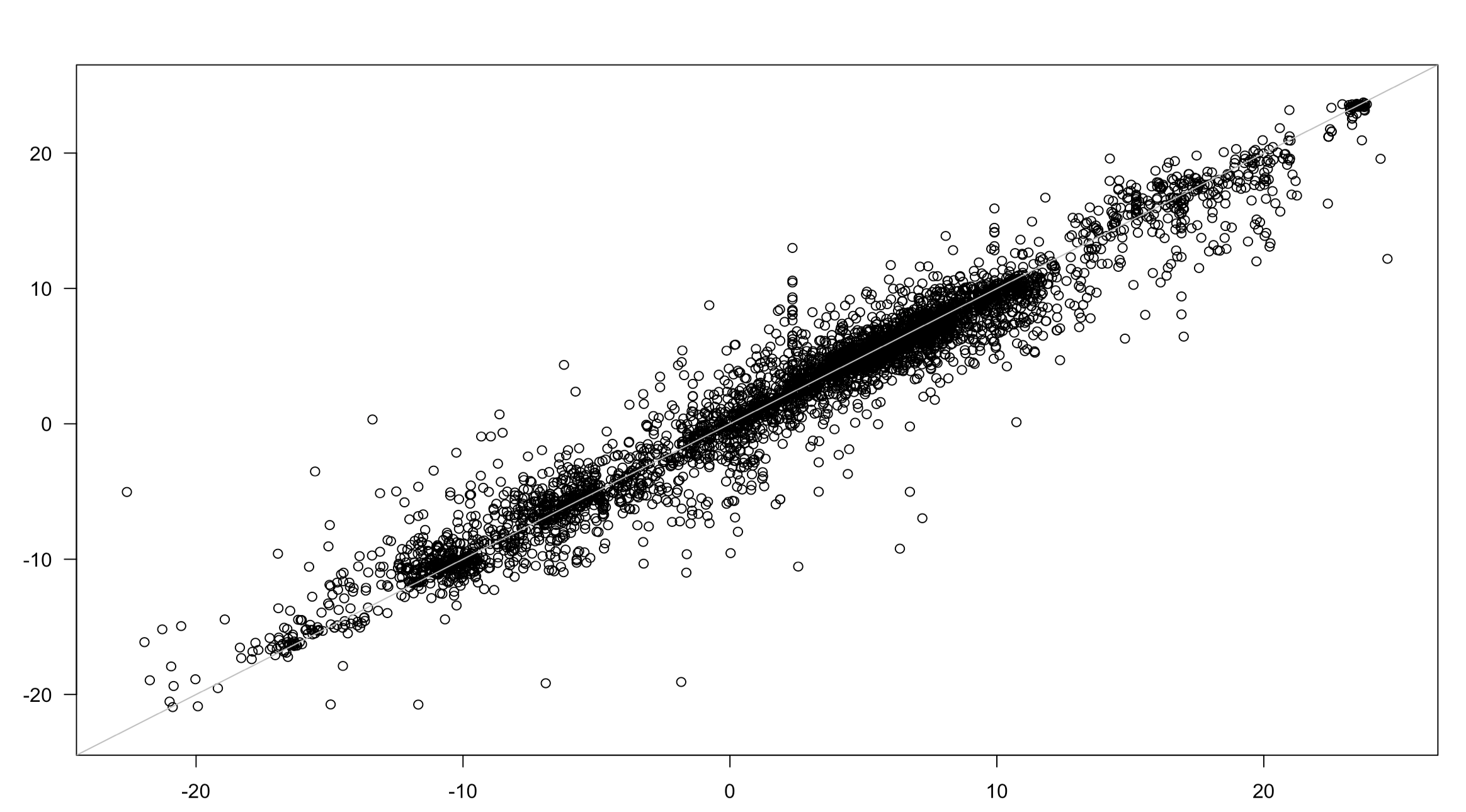
plot(model\_ny\_lat)



#tave:

cv.mat.model.tave=rioja::crossval(model\_ny\_tave,verbose=FALSE)

plot(cv.mat.model.tave)



perf.cv.tave.mat.model=rioja::performance(cv.mat.model.tave)

perf.cv.tave.mat.model

RESULT:

$RMSE0

[1] 8.239152

$object

RMSE R2 Avg.Bias Max.Bias Skill

N01 2.052123 0.9387126 0.05723660 2.080376 93.79643

N02 1.895497 0.9472936 0.08450561 2.310282 94.70725

N03 1.877812 0.9482007 0.09050603 2.826298 94.80556

N04 1.891263 0.9474760 0.10434451 2.977200 94.73087

N05 1.897593 0.9471375 0.11109888 3.027840 94.69554

N01.wm 2.052123 0.9387126 0.05723660 2.080376 93.79643

N02.wm 1.868757 0.9487452 0.07619487 2.386251 94.85553

N03.wm 1.833070 0.9506207 0.08298027 2.821659 95.05014

N04.wm 1.831426 0.9507217 0.09360579 2.906887 95.05901

N05.wm 1.828665 0.9508935 0.10231867 2.962682 95.07390

$crossval

RMSE R2 Avg.Bias Max.Bias Skill

N01 2.113093 0.9350654 0.06764771 2.232829 93.42233

N02 1.940387 0.9447494 0.07341967 2.550235 94.45359

N03 1.907076 0.9465573 0.08145374 3.012084 94.64239

N04 1.915203 0.9461019 0.09448008 3.293256 94.59663

N05 1.927414 0.9454192 0.09866082 3.376272 94.52751

N01.wm 2.113093 0.9350654 0.06764771 2.232829 93.42233

N02.wm 1.912029 0.9463345 0.07033303 2.643943 94.61452

N03.wm 1.862544 0.9490078 0.07666008 3.017888 94.88968

N04.wm 1.857569 0.9492843 0.08779693 3.248162 94.91695

N05.wm 1.861841 0.9490631 0.09311185 3.265887 94.89353

cv.mat.model.tave=rioja::crossval(model\_ny\_tave,verbose=FALSE)

per.mat.model.tave=rioja::performance(cv.mat.model.tave)

plot(cv.mat.model.tave)

depths=as.numeric(rownames(fossilpoll))

pred=predict(model\_ny\_tave fossilpoll)

plot(depths,pred$fit[,'MAT'],type='l',ylab='T [C]')

