

E8: Part1

Marissa Lee

October 23, 2015

Filename: E8_1.Rmd This markdown file does the following tasks: 0. Clean and merge 1. Q1: How does invader biomass vary across resource availability in the reference plots?

```
knitr::opts_chunk$set(cache=TRUE)

library(plyr)
library(reshape2)
library(ggplot2) #plotting
library(ggthemes) #ggplot accessories
library(GGally) #????
library(gridExtra) #for grid.arrange fxn
library(lme4) #for mixed-effects models
```

```
## Loading required package: Matrix
```

```
library(lmerTest) #for lmer p-values
```

```
##
## Attaching package: 'lmerTest'
##
## The following object is masked from 'package:lme4':
##
##     lmer
##
## The following object is masked from 'package:stats':
##
##     step
```

```
library(ggbiplot) #for pretty biplots
```

```
## Loading required package: scales
## Loading required package: grid
```

```
library(vegan) #for adonis fxn to do perMANOVAs
```

```
## Loading required package: permute
## Loading required package: lattice
## This is vegan 2.3-1
```

```
source('CODE/mytheme.R')
source('CODE/fxn_FitPlot.R')
```

```
figuresPath<-file.path(getwd()[1], "FIGURES_TABLES") #where to put the saved plots
```

```

fig.height<-2.5 #inches
fig.width<- 2.5 #inches
fig.res<-300

synthdataPath<-file.path(getwd()[1], "DATA", "DATA_SYNTHESIZED") #where to put the clean dataframes

soilData<-read.table("DATA/e8_plothalfSoilData.txt", header=TRUE, sep="\t", stringsAsFactors = FALSE)
vegData<-read.table("DATA/e8_plothalfVegData.txt", header=TRUE, sep="\t", stringsAsFactors = FALSE)
plotLoc<-read.table("DATA/e8_plotLoc.txt", header=TRUE, sep="\t", stringsAsFactors = FALSE)
plotTrees<-read.table("DATA/e8_plotTrees.txt", header=TRUE, sep="\t", stringsAsFactors = FALSE)

soilData.dict<-read.table("DATA/e8_plothalfSoilData_dictionary.txt", header=TRUE, sep="\t", stringsAsFactors = FALSE)
vegData.dict<-read.table("DATA/e8_plothalfVegData_dictionary.txt", header=TRUE, sep="\t", stringsAsFactors = FALSE)
plotLoc.dict<-read.table("DATA/e8_plotLoc_dictionary.txt", header=TRUE, sep="\t", stringsAsFactors = FALSE)
plotTrees.dict<-read.table("DATA/e8_plotTrees_dictionary.txt", header=TRUE, sep="\t", stringsAsFactors = FALSE)
timeline<-read.table("DATA/e8_timeline.txt", header=TRUE, sep="\t", stringsAsFactors = FALSE)

```

0. Clean and merge

Add identifiers, remove extra columns, merge raw datasets Deal with soil N data in years where 0-5cm and 5-15cm layers were measured separately.

```

#####
#A. Clean and merge
#####

# SOIL DATA
soilData$plothalfid1<-paste(soilData$inv,soilData$plotid, sep="_") #add new identifiers
removeCols<-c('plothalfid','site','rep') #prune columns
soilData.pruned<-soilData[,!colnames(soilData) %in% removeCols]
#reshape so that each depth x meas has its own column
soilData_melted <- melt(soilData.pruned,
                        id.vars=c("plotid","plothalfid1",
                                "inv","depth","year"),
                        variable.name = "measCat")
soilData_melted$measCat2<-paste(soilData_melted$measCat, soilData_melted$depth, sep = '_')
soilData_cast <- dcast(soilData_melted, plotid + plothalfid1 + inv + year ~ measCat2, value.var="value")
soilData_c<-soilData_cast

# VEG DATA
vegData$plothalfid1<-paste(vegData$inv,vegData$plotid, sep="_") #add new identifiers
vegData.pruned<-vegData[,!colnames(vegData) %in% removeCols] #prune columns
vegData.pruned$total<-vegData.pruned$mv+vegData.pruned$nat #make a total understory biomass variable
vegData.dict[c(2,4),c('v7','v8','v9')]

```

```

##          v7          v8          v9
## 2          mv          nat        litter
## 4 g/.1875m2 g/.1875m2 g/.1875m2

```

```
paste('Dry biomass values are currently in units of g/.1875m2')
```

```
## [1] "Dry biomass values are currently in units of g/.1875m2"
```

```
ConvertBiom<-function(currVal){newVal <- currVal / 0.1875}  
vegData.pruned$mv_g.m2<-ConvertBiom(vegData.pruned$mv)  
vegData.pruned$nat_g.m2<-ConvertBiom(vegData.pruned$nat)  
vegData.pruned$litter_g.m2<-ConvertBiom(vegData.pruned$litter)  
vegData.pruned$total_g.m2<-ConvertBiom(vegData.pruned$total)  
colsOldUnits<-c('mv','nat','litter','total')  
vegData_c<-vegData.pruned[,!colnames(vegData.pruned) %in% colsOldUnits] #prune columns
```

```
#add vegData
```

```
soilVegData<-merge(soilData_c, vegData_c) #merge soilData_cast and vegData
```

```
# TREE DATA
```

```
plotTrees$basalArea.m2<-(plotTrees$dbh * plotTrees$dbh) * 0.00007854 #calculate basal area/m2 from each  
#summarize the total basal area/m2 per plot and that which is made up by either AM- or ECM-associated t
```

```
plotTrees.summ<-ddply(plotTrees, ~plotid, summarize,  
  nTrees=length(plotid),  
  BA_total=sum(basalArea.m2, na.rm=T),  
  BA_AM=sum(basalArea.m2[myc=='A'], na.rm=T),  
  BA_ECM=sum(basalArea.m2[myc=='E'], na.rm=T),  
  PercBA_AM=(BA_AM/BA_total)*100,  
  PercBA_ECM=(BA_ECM/BA_total)*100)
```

```
#update the number of trees (there was a cell that was counted even for plots where there were no trees
```

```
plotTrees.summ[plotTrees.summ$plotid %in% c(12,15),'nTrees']<-0  
plotTrees.summ[plotTrees.summ$plotid %in% c(12,15),c('PercBA_AM','PercBA_ECM')]<-NA  
tmp<-plotTrees.summ[,c('plotid','nTrees','BA_total','PercBA_AM')]  
trees_c<-tmp
```

```
#add tree data
```

```
data<-merge(soilVegData, trees_c)
```

```
#get rid of the plots without trees
```

```
data<-data[data$nTrees != 0,]
```

```
#str(data)
```

```
#####
```

```
#B. Deal with soil N data in years where 0-5cm and 5-15cm layers were measured separately
```

```
#####
```

```
# Plot histogram of values by depth
```

```
#subset by year
```

```
data.1213<-subset(data, year %in% c(2012, 2013))
```

```
#reshape
```

```
data.1213_melted <- melt(data.1213,id.vars=c("plotid","plothalfid1","inv","year"))  
data.1213_melted$value<-as.numeric(data.1213_melted$value)
```

```
#subset by _B and _T
```

```
data.1213_melted_soilVars<-data.1213_melted[grep("_B",data.1213_melted$variable) | grep("_T",data.1213_melted$variable)]
```

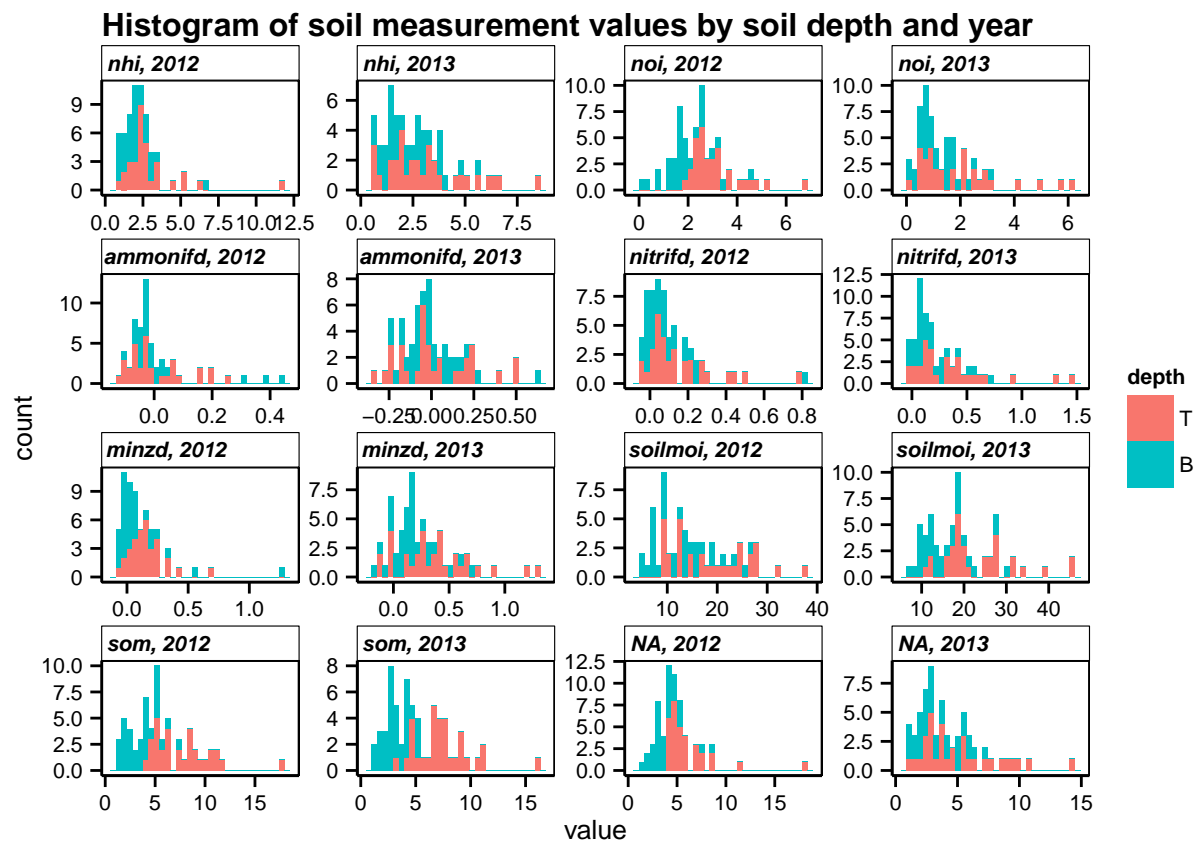
```

#add back the separate measCat and depth columns
tmp<-ldply(strsplit(as.character(data.1213_melted_soilVars$variable),"_", fixed=T))
colnames(tmp)<-c('measCat','depth')
data.1213_melted_soilVars<-cbind(data.1213_melted_soilVars, tmp)

#correct the dataframe structure
data.1213_melted_soilVars$measCat<-factor(data.1213_melted_soilVars$measCat, levels=measCat_order)
data.1213_melted_soilVars$depth<-factor(data.1213_melted_soilVars$depth, levels=depth_order)
data.1213_melted_soilVars$inv<-factor(data.1213_melted_soilVars$inv, levels=inv_order)

#plot
pHist.depth<-ggplot(data.1213_melted_soilVars, aes(x=value, fill=depth)) +
  geom_bar() + mytheme + facet_wrap(measCat~year, scales='free') +
  ggtitle("Histogram of soil measurement values by soil depth and year")
pHist.depth

```



```

#save plot
newfilename<-'pHist_depth.png'
png(paste(figuresPath,newfilename, sep='/'), units='in', width = fig.width*4, height = fig.height*4, res=
pHist.depth
dev.off()

```

```

## pdf
## 2

```

```
#####
#C. DATASET A : Aggregate values, put aggregated values into full dataset and replace empty _F rows
#####

#aggregate by depth
summ.data.1213_melted_soilVars<-ddply(data.1213_melted_soilVars, ~plotid+plothalfid1+inv+year+measCat,
                                     new.value=mean(value, na.rm=T),
                                     n=length(plothalfid1),
                                     note='aggregated by depth')

#make a dataset without the dis-aggregated data
data_melted <- melt(data,id.vars=c("plotid","plothalfid1","inv","year")) #melt down the full dataset, i
data_melted$value<-as.numeric(data_melted$value) #make values numeric again
uniqWo.B<-unique(data_melted$variable)[!grepl("_B",unique(data_melted$variable))] #isolate only the var
uniqWo.BT<-uniqWo.B[!grepl("_T",uniqWo.B)] #ditto
data_m_woBT<-data_melted[data_melted$variable %in% uniqWo.BT,] #subset data by 'correct' variables
selectvars<-unique(data_m_woBT$variable)[grepl("_F",unique(data_m_woBT$variable))] #select measCat vari
data_m_woBT_e<-data_m_woBT[!(data_m_woBT$year %in% c(2012,2013) & data_m_woBT$variable %in% selectvars)
data_m_woBT_e$note<-NA

#make a dataset with the correctly-aggregated data and comparable column names to data_m_woBT
summ.data.1213_melted_soilVars<-rename(summ.data.1213_melted_soilVars, replace=c("new.value" = "value"),
summ.data.1213_melted_soilVars$variable<-paste(summ.data.1213_melted_soilVars$measCat, '_F',sep="")#mea
summ.data.1213_melted_soilVars<-summ.data.1213_melted_soilVars[,c('plotid','plothalfid1','inv','year','

#merge
data_a<-rbind(data_m_woBT_e, summ.data.1213_melted_soilVars)
#ddply(data_a, ~year+variable, summarize, n=length(plotid)) #check the dataframe structure
#ddply(data_a, ~inv+variable, summarize, n=length(plotid)) #check the dataframe structure

#add a column that categorizes the variable types
data_a$varType<-NA
data_a[grepl("_F", data_a$variable),'varType']<-'measCat'
data_a[data_a$variable %in% c('mv_g.m2','nat_g.m2','litter_g.m2','total_g.m2'),'varType']<-'understoryB
data_a[data_a$variable %in% c('percpair','soiltemp','soilBasin'),'varType']<-'environParam'
data_a[data_a$variable %in% c('nTrees','BA_total','PercBA_AM'),'varType']<-'overstoryParam'

#add a column that identifies whether the variable type is measured on the plot half or the plot level
plotlevel.vars<-c('soilBasin','nTrees','BA_total','PercBA_AM')
plothalflevel.vars<-unique(data_a$variable)[!unique(data_a$variable) %in% plotlevel.vars]
data_a$measLevel<-NA
data_a[data_a$variable %in% plotlevel.vars,'measLevel']<-'plot'
data_a[data_a$variable %in% plothalflevel.vars,'measLevel']<-'plothalf'

#Ammend variable names with "_N" or "_I" depending on the plothalf
data_a$variable1<-paste(data_a$variable, data_a$inv, sep="_")
data_a[data_a$measLevel=='plot','variable1']<-as.character(data_a[data_a$measLevel=='plot','variable'])

#save and export
newfilename<-'data_a.txt'
write.table(data_a, file=paste(synthdataPath,newfilename, sep='/'), sep='\t')
```

```
#####
#D. DATASET D : Keep 2012 and 2013 values dis-aggregated and remove 2011 _F values, remove biomass values
#####
#start with data.1213_melted
data_d<-data.1213_melted[!grepl("_F", data.1213_melted$variable),] #remove variables that were only measured in 2011
data_d<-data_d[data_d$variable != 'soiltemp',] #remove variables that were only measured in 2011

#add a column that categorizes the depth types
data_d$depth<-NA
data_d[grepl("_B", data_d$variable), 'depth']<-'B'
data_d[grepl("_T", data_d$variable), 'depth']<-'T'

#add a column that categorizes the variable types
data_d$varType<-NA
data_d[grepl("_B", data_d$variable), 'varType']<-'measCat'
data_d[grepl("_T", data_d$variable), 'varType']<-'measCat'
data_d[data_d$variable %in% c('mv_g.m2', 'nat_g.m2', 'litter_g.m2', 'total_g.m2'), 'varType']<-'understoryParam'
data_d[data_d$variable %in% c('percpa'), 'varType']<-'environParam'
data_d[data_d$variable %in% c('nTrees', 'BA_total', 'PercBA_AM'), 'varType']<-'overstoryParam'

#add a column that identifies whether the variable type is measured on the plot half or the plot level
plotlevel.vars<-c('nTrees', 'BA_total', 'PercBA_AM')
plothalflevel.vars<-unique(data_d$variable)[!unique(data_d$variable) %in% plotlevel.vars]
data_d$measLevel<-NA
data_d[data_d$variable %in% plotlevel.vars, 'measLevel']<-'plot'
data_d[data_d$variable %in% plothalflevel.vars, 'measLevel']<-'plothalf'

#Amend variable names with "_N" or "_I" depending on the plothalf
data_d$variable1<-paste(data_d$variable, data_d$inv, sep="_")
data_d[data_d$measLevel=='plot', 'variable1']<-as.character(data_d[data_d$measLevel=='plot', 'variable'])

#save and export
newfilename<-'data_d.txt'
write.table(data_d, file=paste(synthdataPath, newfilename, sep='/'), sep='\t')

#####
# DATA CHOICE : decide which dataset to use
data.choice<-data_a
data.choice<-subset(data.choice, year != 2011)
#####
```

1. Q1: Do soil N pools and fluxes shift in response to the presence of Microstegium?

A. Do an ordination of soil N pools and fluxes. B. Test the role of invasion status with permMANOVA C. Investigate individual relationships using mixed effects models with year+site as a random effect. D. Identify N variables that increase/decrease across sites.

```
#####
```

```
#A. Do an ordination of soil N pools and fluxes.
```

```
#####
```

```
#1) Name the plot variable to ordinate
```

```
vars<-c('nhi_F','noi_F','ammonifd_F','nitrifd_F','minzd_F') #make vector of terms
```

```
varNames<-c('NH4+', 'NO3-', 'Ammonif.', 'Nitrif.', 'Mineraliz.')
```

```
#2) Subset and reshape
```

```
data.vars<-data.choice[data.choice$variable %in% vars, c('plotid','plothalfid1','inv','year','variable')  
data.vars$variable<-factor(data.vars$variable, levels=vars)
```

```
data.vars.wide<-dcast(data.vars, plothalfid1 + inv+ year ~ variable, value.var='value')
```

```
data.vars.wide$plotInvYear<-paste(data.vars.wide$plothalfid1, data.vars.wide$year, sep="_") #use this to block
```

```
data.vars.wide$plotYear<-paste(data.vars.wide$plotid, data.vars.wide$year, sep="_") #use this to block
```

```
row.names(data.vars.wide)<-data.vars.wide$plotInvYear
```

```
NonValCols<-c('plothalfid1','plotInvYear', 'plotYear','inv','year')
```

```
ValCols<-colnames(data.vars.wide)[!(colnames(data.vars.wide) %in% NonValCols)]
```

```
#3) Remove rows with missing data and look at correlation matrix
```

```
data.vars.wide.rm<-data.vars.wide[!rowSums(is.na(data.vars.wide))>0,]
```

```
cor(data.vars.wide.rm[,ValCols])
```

```
##              nhi_F      noi_F ammonifd_F   nitrifd_F    minzd_F  
## nhi_F      1.00000000  0.51023042 -0.1704322  0.23675304  0.06605252  
## noi_F      0.51023042  1.00000000 -0.0651940  0.04729761 -0.04983418  
## ammonifd_F -0.17043223 -0.06519400  1.00000000 -0.31729409  0.31270435  
## nitrifd_F   0.23675304  0.04729761 -0.3172941  1.00000000  0.79373804  
## minzd_F     0.06605252 -0.04983418  0.3127044  0.79373804  1.00000000
```

```
#4) Ordinate and save plot
```

```
df.ord<-data.vars.wide.rm[,ValCols]
```

```
colnames(df.ord)<-varNames
```

```
df.ord.scaled<-scale(df.ord)
```

```
PCA.res<-prcomp(df.ord.scaled)
```

```
pPCA<-ggbiplot(PCA.res, groups=data.vars.wide.rm$inv, ellipse=T) + mytheme
```

```
newfilename<-'pPCA_invPlot.png'
```

```
png(paste(figuresPath,newfilename, sep='/'), units='in', width = fig.width*2, height = fig.height*2, res=300)
```

```
pPCA
```

```
dev.off()
```

```
## pdf
```

```
## 2
```

```
#5) Make a dataframe of the pc scores
```

```
data.PCA<-data.frame(plotInvYear=rownames(PCA.res$x), PCA.res$x)
```

```
data.vars.PCA<-merge(data.vars.wide.rm, data.PCA[,c(1:3)])
```

```
#####
```

```
#B. Test the role of invasion status with permMANOVA
```

```
#####
```

```
sum(rownames(data.vars.wide.rm) != rownames(df.ord.scaled)) # if this is 0, then rownames match
```

```
## [1] 0
```

```
permMANOVA<-adonis(df.ord.scaled ~ inv, data = data.vars.wide.rm, method='eu')
aov.tab.permmanova<-data.frame(permMANOVA$aov.tab)
aov.tab.permmanova
```

```
##           Df SumsOfSqs MeanSqs F.Model      R2 Pr..F.
## inv         1   9.419705 9.419705 1.911189 0.02990382 0.104
## Residuals  62 305.580295 4.928714      NA 0.97009618    NA
## Total      63 315.000000      NA      NA 1.00000000    NA
```

```
newfilename<-'permMANOVA_invPlot.txt'
write.table(aov.tab.permmanova, file=paste(figuresPath,newfilename, sep='/'), sep='\t')
```

```
#####
```

```
#C. Investigate individual relationships using mixed effects models with year+site as a random effect.
```

```
#####
```

```
#1) Make a fcn to run the models with invasion status predicting different y variables
```

```
InvEffect<-function(Yvars, YvarNames, df){
```

```
  anova.list<-list()
```

```
  figure.list<-list()
```

```
  i<-0
```

```
  for(i in 1:length(Yvars)){
```

```
    curr.Yvar<-Yvars[i]
```

```
    y<-df[,curr.Yvar]
```

```
    #regression model
```

```
    mod<-lmer(y~inv + (1|plotYear), data=df)
```

```
    anova.list[[i]]<-anova(mod)
```

```
    #plot
```

```
    df$year<-factor(df$year)
```

```
    p<-ggplot(df, aes_string(x='inv', y=curr.Yvar, color='year')) + geom_boxplot() +
```

```
      mytheme + xlab(NULL) + ylab(NULL) + ggtitle(YvarNames[i])
```

```
    figure.list[[i]]<-p
```

```
  }
```

```
  names(anova.list)<-Yvars
```

```
  names(figure.list)<-Yvars
```

```
  #anova summary
```

```
  anovaSummary<-ldply(lapply(anova.list, data.frame))
```

```
  anovaSummary[, -1]<-round(anovaSummary[, -1], digits=4)
```

```
  anovaSummary<-rename(anovaSummary, c(.id='yVar', Pr..F='pValue'))
```

```
  results<-list(anovaSummary=anovaSummary, figures=figure.list)
```

```
  return(results)
```

```
}
```



```

#2) Run models
result<-InvEffect(Yvars=c('noi_F','nitrifd_F','ammonifd_F'),
                  YvarNames=c('Nitrate (ug/G)','Nitrification (ug/G*d)','Ammonification (ug/G*d)'),
                  df=data.vars.PCA)

#3) Save anova summary table
data.frame(result[['anovaSummary']])

##           yVar Sum.Sq Mean.Sq NumDF   DenDF F.value pValue
## 1      noi_F 0.3584  0.3584     1 61.0000  0.3832 0.5382
## 2 nitrifd_F 0.0049  0.0049     1 61.0000  0.1424 0.7072
## 3 ammonifd_F 0.1215  0.1215     1 62.0001  8.0397 0.0062

newfilename<-'anovaSummary_inv.txt'
write.table(data.frame(result[['anovaSummary']]), file=paste(figuresPath,newfilename, sep='/'), sep='\t')

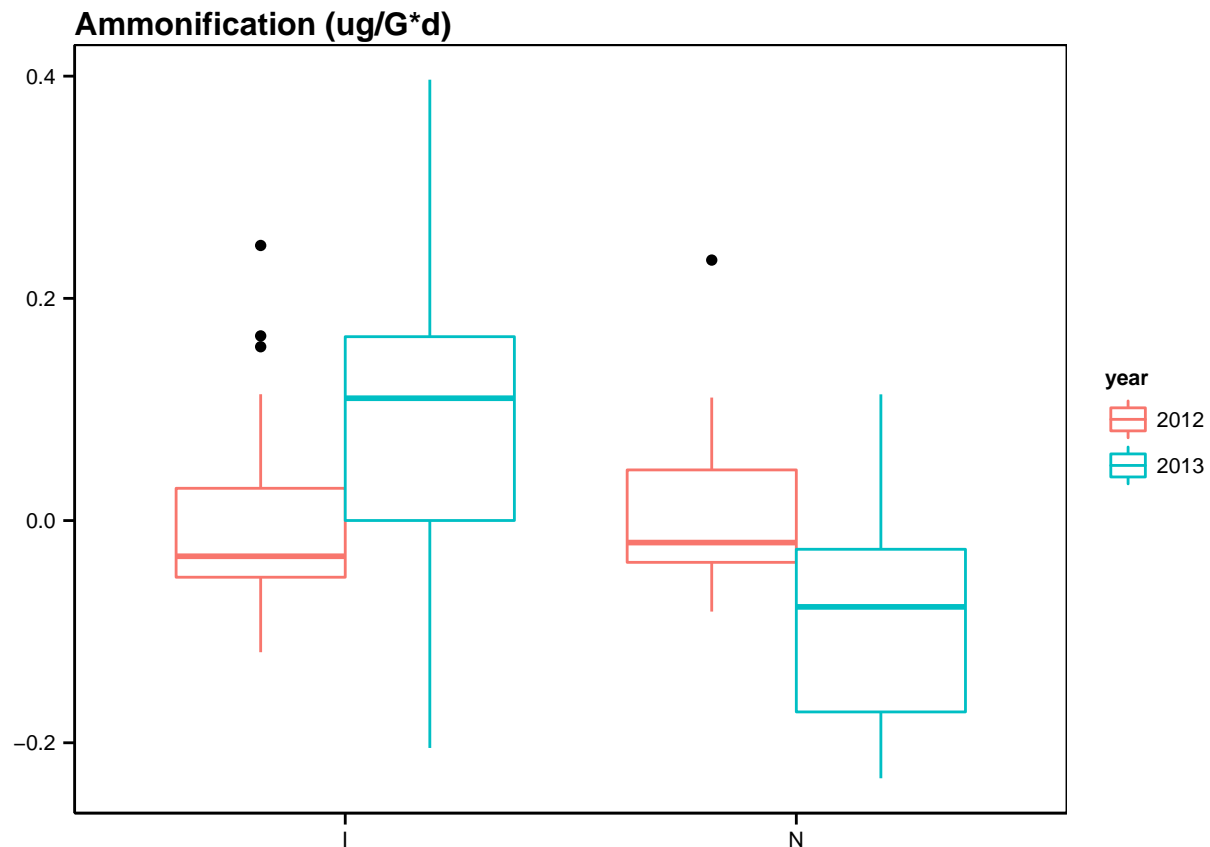
#4) Plot
#grid.arrange(do.call(arrangeGrob, result[['figures']])) #all panels
ylabel<-textGrob("Plot variable", rot=90)
xlabel<-textGrob("Invasion status")
newfilename<-'pScat_inv.png'
png(paste(figuresPath,newfilename, sep='/'), units='in', width = fig.width*1.5, height = fig.height*3,
grid.arrange(
  ylabel, # topleft
  do.call(arrangeGrob, result[['figures']]),
  textGrob(" "), #bottom left
  xlabel, #bottom right
  widths = unit.c(unit(2.5, "lines"), unit(1, "npc") - unit(2.5, "lines")),
  heights = unit.c(unit(1, "npc") - unit(2.5, "lines"), unit(2.5, "lines")),
  nrow=2, ncol=2
)
dev.off()

## pdf
## 2

#####
#D. Identify N variables that increase/decrease across sites.
#####

result[['figures']]$ammonifd_F

```



```
paste('Net potential ammonification increases in invaded plots so that more ammonium is being released')
```

```
## [1] "Net potential ammonification increases in invaded plots so that more ammonium is being released"
```

```
paste('Although net ammonification increases in invaded plots, ammonium pools do not differ. This may be')

```

```
## [1] "Although net ammonification increases in invaded plots, ammonium pools do not differ. This may be"
```

2. Q2: Do reference plot conditions and/or Microstegium biomass predict impact magnitudes?

A. Calculate impact magnitude for each target variable (nitrate, nitrification, other variables detected in Q1 ordination) B. Test the role of reference plot conditions and Microstegium biomass on impacts individually using mixed effects models with year as a random effect

```
#####
#A. Calculate impact magnitude for each target variable (nitrate, nitrification, ammonification)
#####

#reshape dataframe
data.vars.wideIN <- dcast(data.vars, plotid + year + variable ~ inv, value.var="value")
```

```

#calculate diff
data.vars.wideIN$Diff<-data.vars.wideIN$I - data.vars.wideIN$N

#subset select diff variables and reshape
diffVars<-c('noi_F','nitrifd_F','ammonifd_F')
data.diffVars<-data.vars.wideIN[data.vars.wideIN$variable %in% diffVars, c('plotid','year','variable','value')]
data.diffVars.wide<-dcast(data.diffVars, plotid + year ~ variable, value.var='Diff')
colnames(data.diffVars.wide)[-c(1:2)]<-paste('Diff', colnames(data.diffVars.wide)[-c(1:2)], sep="_")

#####
#B. Test the role of reference plot conditions and Microstegium biomass on impacts individually using mixed models
#####

#1. Identify reference plot condition variables
refVars<-c('nhi_F', 'noi_F', 'ammonifd_F', 'nitrifd_F', 'minzd_F','soilmoi_F','som_F',
           'nat_g.m2','litter_g.m2','percpar','nTrees','BA_total','PercBA_AM')
data.refVars<-data.choice[data.choice$variable %in% refVars, c('plotid','plothalfid1','inv','year','value')]
data.refVars$variable<-factor(data.refVars$variable, levels=refVars)
data.refVars<-subset(data.refVars, inv == 'N')
data.refVars.wide<-dcast(data.refVars, plotid + year ~ variable, value.var='value')

#2. Identify microstegium biomass
mvVar<-c('mv_g.m2')
data.mvVar<-data.choice[data.choice$variable %in% mvVar, c('plotid','plothalfid1','inv','year','value')]
data.mvVar<-subset(data.mvVar, inv == 'I')
data.mvVar.wide<-dcast(data.mvVar, plotid + year ~ variable, value.var='value')
pHist.mv<-ggplot(data.mvVar.wide, aes(x=log(mv_g.m2)))+geom_histogram() #log-transform mv biomass to improve normality
data.mvVar.wide$mv_g.m2_logt<-log(data.mvVar.wide$mv_g.m2)

#3. Merge the reference plot condition columns and the microstegium biomass columns
data.q2.indp<-merge(data.refVars.wide, data.mvVar.wide)

#4. Merge the impact variables and the independant variables
data.q2<-merge(data.diffVars.wide, data.q2.indp)
data.q2$plotYear<-paste(data.q2$plotid,data.q2$year, sep="_")

#5. Make a fcn to run the models with reference conditions (excluding linked response variable) and Microstegium biomass

#testing
#diffVarShort='nitrifd_F'; refVars=refVars; mvVar='mv_g.m2_logt'; df=data.q2

ExplainDiffVar<-function(diffVarShort, refVars, mvVar, df){

  #identify variables
  select.diffVar<-paste('Diff',diffVarShort, sep='_')
  y<-df[,select.diffVar]
  select.refVars<-refVars[!refVars %in% diffVarShort]
  xVars<-c(select.refVars, 'mv_g.m2_logt')

  #individual model for each reference condition
  anova.list<-list()
  pBivar.list<-list()

```

```

i<-0
for(i in 1:length(xVars)){

  #identify current dataframe
  curr.Xvar<-xVars[i]
  x<-df[,curr.Xvar]
  df.i<-data.frame(x, y, year=df$year)
  df.i.rm<-df.i[!rowSums(is.na(df.i)),]
  df.i.rm$year<-factor(df.i.rm$year)

  #regression model
  mod<-lmer(y~x + (1|year), data=df.i.rm)
  anova.list[[i]]<-anova(mod)

  #bivariate plot
  p<-ggplot(df.i.rm, aes(x=x, y=y, color=year)) + geom_point() + mytheme +
    xlab(curr.Xvar) + ylab(select.diffVar)
  pBivar.list[[i]]<-p

}

#save anovaSummary
names(anova.list)<-xVars
anovaSummary<-ldply(lapply(anova.list, data.frame))
anovaSummary[, -1]<-round(anovaSummary[, -1], digits=4)
anovaSummary<-rename(anovaSummary, c(.id='yVar', Pr..F.='pValue'))
anovaSummary
#save plots
names(pBivar.list)<-xVars

#full model multivariate
select.df<-data.frame(df[,c('plotYear', 'year')], Diff=df[,select.diffVar], df[,c(select.refVars, mvVar)])
select.df.rm<-select.df[!rowSums(is.na(select.df))>0,]
select.df.rm$DiffColor<-factor(select.df.rm$Diff>0)
select.df.rm.scaled<-scale(select.df.rm[,!colnames(select.df.rm) %in% c('DiffColor', 'plotYear', 'year')])
row.names(select.df.rm.scaled)<-select.df.rm$plotYear
#make ordination
PCA.res<-prcomp(select.df.rm.scaled)
pPCA<-ggbiplot(PCA.res, group=select.df.rm$DiffColor, ellipse = T) + mytheme + ggtitle(diffVarShort)
#test permanova
data.PCA<-data.frame(plotYear=row.names(PCA.res$x), PCA.res$x)
data.vars.PCA<-merge(select.df.rm, data.PCA[,c(1:3)])
permMANOVA<-adonis(select.df.rm.scaled ~ Diff, data = data.vars.PCA, method='eu')
aov.tab.permmanova<-data.frame(permMANOVA$aov.tab)

results<-list(anovaSummary=anovaSummary, permAOV=aov.tab.permmanova, pPCA=pPCA, pBivar.list=pBivar.list)

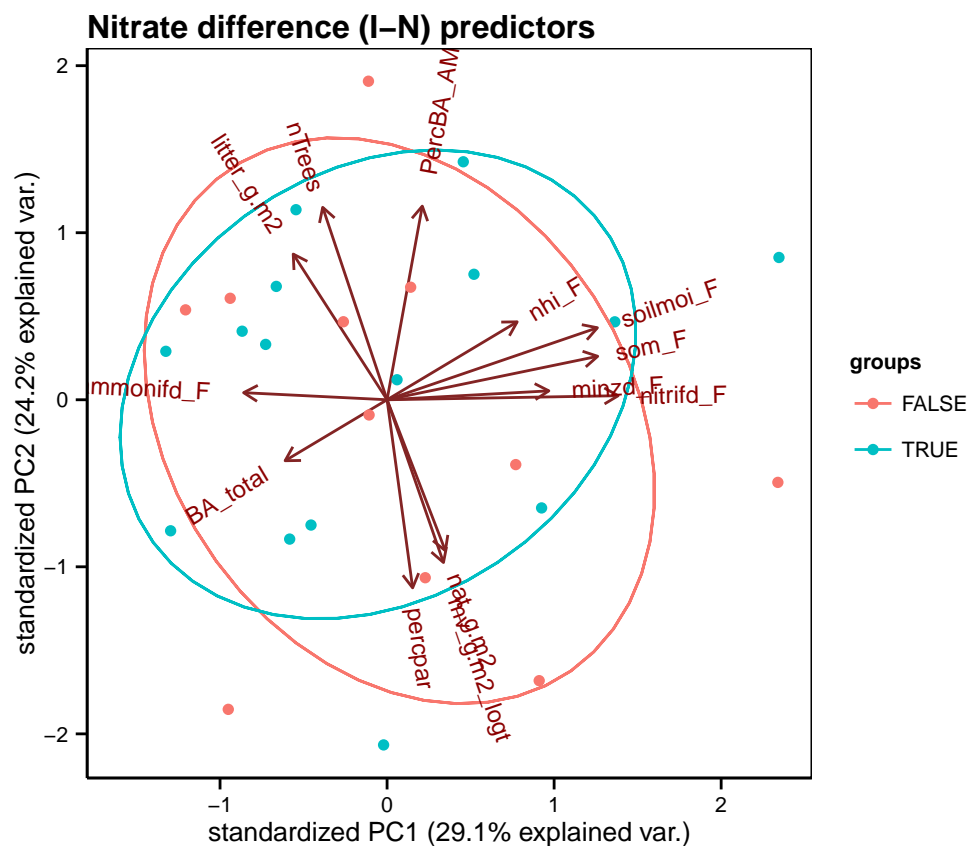
return(results)
}

#A. Predict nitrate diff
result<-ExplainDiffVar(diffVarShort='noi_F', refVars=refVars, mvVar='mv_g.m2_logt', df=data.q2)
result[['anovaSummary']] #save

```

	yVar	Sum.Sq	Mean.Sq	NumDF	DenDF	F.value	pValue
## 1	nhi_F	0.0812	0.0812	1	30	0.1395	0.7114
## 2	ammonifd_F	0.0839	0.0839	1	30	0.1441	0.7069
## 3	nitrifd_F	0.0811	0.0811	1	30	0.1394	0.7115
## 4	minzd_F	0.2143	0.2143	1	30	0.3711	0.5470
## 5	soilmoi_F	0.1172	0.1172	1	30	0.2018	0.6565
## 6	som_F	0.0074	0.0074	1	30	0.0126	0.9114
## 7	nat_g.m2	0.0703	0.0703	1	30	0.1207	0.7307
## 8	litter_g.m2	0.0474	0.0474	1	30	0.0812	0.7776
## 9	percpair	0.0071	0.0071	1	24	0.0105	0.9190
## 10	nTrees	0.8314	0.8314	1	30	1.4929	0.2313
## 11	BA_total	0.1882	0.1882	1	30	0.3254	0.5726
## 12	PercBA_AM	0.0637	0.0637	1	30	0.1093	0.7432
## 13	mv_g.m2_logt	0.0170	0.0170	1	30	0.0291	0.8658

```
#save anova table
newfilename<-'anovaTable_noiDiff.txt'
write.table(result[['anovaSummary']], file=paste(figuresPath,newfilename, sep='/'), sep='\t')
#save permMANOVA table
newfilename<-'permMANOVA_noiDiff.txt'
write.table(result[['permaOV']], file=paste(figuresPath,newfilename, sep='/'), sep='\t')
result[['pPCA']] + ggtitle('Nitrate difference (I-N) predictors')
```



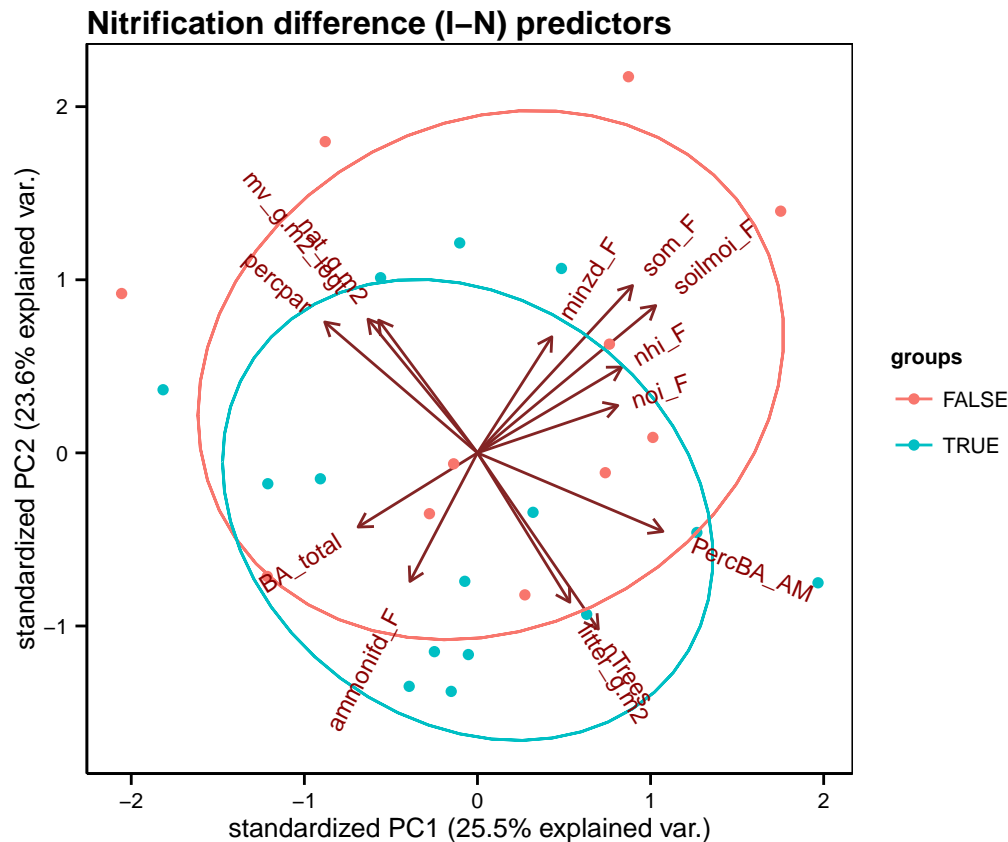
```
paste('Neither reference conditions nor Microstegium biomass explain variation in nitrate')
```

```
## [1] "Neither reference conditions nor Microstegium biomass explain variation in nitrate"
```

```

#B. Predict nitrification diff
result<-ExplainDiffVar(diffVarShort='nitrifd_F', refVars=refVars, mvVar='mv_g.m2_logt', df=data.q2)
#save anova table
newfilename<-'anovaTable_nitrifDiff.txt'
write.table(result[['anovaSummary']], file=paste(figuresPath,newfilename, sep='/'), sep='\t')
#save permANOVA table
newfilename<-'permANOVA_nitrifDiff.txt'
write.table(result[['permaOV']], file=paste(figuresPath,newfilename, sep='/'), sep='\t')
result[['pPCA']] + ggtitle('Nitrification difference (I-N) predictors')

```



```

paste('Reference mineralization rate, soil moisture, and soil organic matter explain variation in nitrification difference')

```

```

## [1] "Reference mineralization rate, soil moisture, and soil organic matter explain variation in nitrification difference"

```

```

#plot signif vars (x) vs nitrif diff (y)
#save plot
ylabel<-textGrob("Nitrification difference (I-N)", rot=90)
xlabel<-textGrob("Reference condition")
pMinzd<-result[['pBivar.list']]$minzd_F + geom_smooth(method='lm') + guides(color=FALSE) +
  xlab(NULL) + ylab(NULL) + ggtitle('Mineralization (ug/G*d)')
pSoilmoi<-result[['pBivar.list']]$soilmoi_F + geom_smooth(method='lm') + guides(color=FALSE) +
  xlab(NULL) + ylab(NULL) + ggtitle('Soil moisture (%)')
pSom<-result[['pBivar.list']]$som_F + geom_smooth(method='lm') + guides(color=FALSE) +
  xlab(NULL) + ylab(NULL) + ggtitle('Soil organic matter (%)')
newfilename<-'pScat_noiDiff.png'

```

```

png(paste(figuresPath,newfilename, sep='/'), units='in', width = fig.width*1, height = fig.height*3, res=
grid.arrange(
  ylabel, # topleft
  arrangeGrob(pMinz, pSoilmoi, pSom),
  textGrob(" "), #bottom left
  xlabel, #bottom right
  widths = unit.c(unit(2.5, "lines"), unit(1, "npc") - unit(2.5, "lines")),
  heights = unit.c(unit(1, "npc") - unit(2.5, "lines"), unit(2.5, "lines")),
  nrow=2, ncol=2
)
dev.off()

```

```

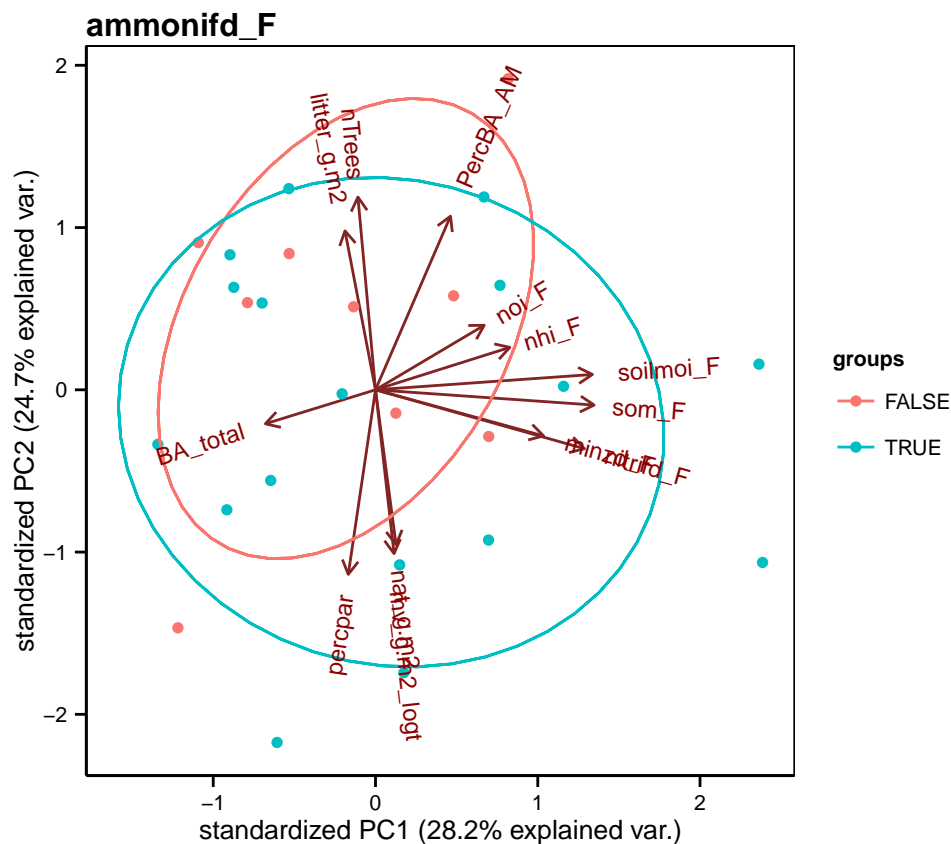
## pdf
## 2

```

```

#C. Predict ammonification diff
result<-ExplainDiffVar(diffVarShort='ammonifd_F', refVars=refVars, mvVar='mv_g.m2_logt', df=data.q2)
#save anova table
newfilename<-'anovaTable_ammonifDiff.txt'
write.table(result[['anovaSummary']], file=paste(figuresPath,newfilename, sep='/'), sep='\t')
#save permMANOVA table
newfilename<-'permMANOVA_ammonifDiff.txt'
write.table(result[['permAOV']], file=paste(figuresPath,newfilename, sep='/'), sep='\t')
result[['pPCA']]

```



```

paste('Microstegium biomass explains variation in ammonification')

## [1] "Microstegium biomass explains variation in ammonification"

#plot signif vars (x) vs nitrid diff (y)
newfilename<-'pScat_ammonifDiff.png'
png(paste(figuresPath,newfilename, sep='/'), units='in', width = fig.width*1, height = fig.height*1, res=300)
result[['pBivar.list']]$mv_g.m2_logt + geom_smooth(method='lm') + guides(color=FALSE) +
  xlab('Microstegium') + ylab('Ammonification difference (I-N)')
dev.off()

## pdf
## 2

```

3. Q3: Do reference plot conditions predict Microstegium biomass?

A. Do an ordination of reference conditions. B. Test the relationship between Microstegium biomass and reference conditions using permMANOVA. C. Investigate individual relationships using mixed effects models with year as a random effect.

```

#1. Identify dataframe with (x) reference plot condition variables and (y) Microstegium biomass
#refVars
data.q3<-data.q2.indp
data.q3$plotYear<-paste(data.q3$plotid, data.q3$year, sep="_")

#2. Ordinate reference conditions
data.q3.rm<-data.q3[!rowSums(is.na(data.q3))>0,]
tmp<-data.q3.rm[,refVars]
cor(tmp) #look at correlation matrix

```

```

##              nhi_F          noi_F  ammonifd_F  nitrid_F  minzd_F
## nhi_F          1.00000000  0.228928023 -0.40703592  0.35909440  0.10965729
## noi_F          0.22892802  1.000000000 -0.15042511  0.13248114  0.03498599
## ammonifd_F    -0.40703592 -0.150425113  1.00000000 -0.54414358  0.13400374
## nitrid_F      0.35909440  0.132481143 -0.54414358  1.00000000  0.75840269
## minzd_F       0.10965729  0.034985991  0.13400374  0.75840269  1.00000000
## soilmoi_F     0.36263779  0.432467789 -0.44254127  0.72619591  0.51254044
## som_F         0.50381419  0.369062109 -0.24036856  0.68003593  0.61671610
## nat_g.m2      -0.13698505 -0.032439353 -0.11123402  0.18133659  0.13060454
## litter_g.m2   0.20273076  0.219876083  0.50310183 -0.40893547 -0.09327104
## percpair      0.01268812 -0.303452291 -0.05176894  0.02578535 -0.00206067
## nTrees        -0.07118420  0.002642215  0.11352848 -0.09210334 -0.02830722
## BA_total      -0.42345093 -0.271278802  0.23082878 -0.23115841 -0.09471401
## PercBA_AM     0.19137324  0.251179780 -0.17740683  0.16487359  0.05245243
##              soilmoi_F          som_F          nat_g.m2  litter_g.m2  percpair
## nhi_F          0.362637788  0.503814194 -0.13698505  0.202730757  0.01268812
## noi_F          0.432467789  0.369062109 -0.03243935  0.219876083 -0.30345229
## ammonifd_F    -0.442541267 -0.240368561 -0.11123402  0.503101827 -0.05176894

```



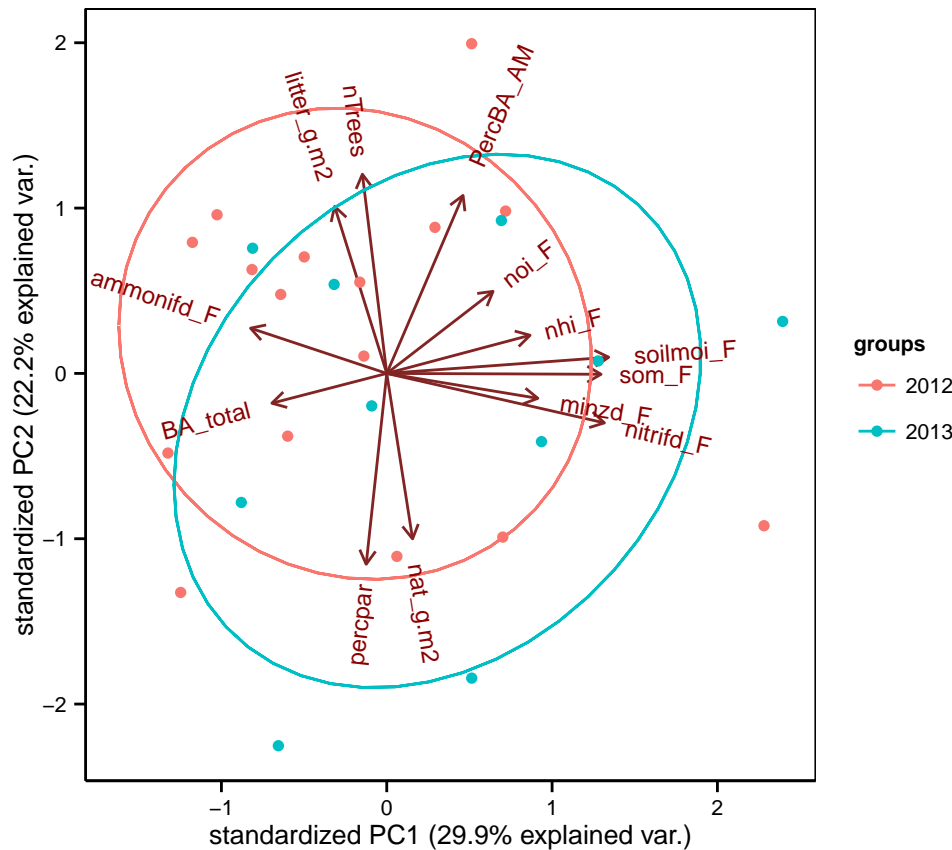
```
## nitrifd_F      0.726195912  0.680035926  0.18133659 -0.408935465  0.02578535
## minzd_F       0.512540440  0.616716099  0.13060454 -0.093271039 -0.00206067
## soilmoi_F     1.000000000  0.786122334  0.19823184 -0.138069439 -0.21633329
## som_F         0.786122334  1.000000000  0.14148840 -0.008757706 -0.04707648
## nat_g.m2      0.198231838  0.141488398  1.000000000 -0.357597640  0.19719854
## litter_g.m2   -0.138069439 -0.008757706 -0.35759764  1.000000000 -0.38023773
## percpair      -0.216333289 -0.047076483  0.19719854 -0.380237733  1.00000000
## nTrees        -0.005851712 -0.115328018 -0.57268724  0.391036473 -0.63121063
## BA_total      -0.322432174 -0.349825442  0.45247833  0.030849064 -0.21459259
## PercBA_AM     0.318111795  0.137143941 -0.36251300  0.206796770 -0.63141277
##              nTrees  BA_total  PercBA_AM
## nhi_F         -0.071184204 -0.42345093  0.19137324
## noi_F          0.002642215 -0.27127880  0.25117978
## ammonifd_F    0.113528478  0.23082878 -0.17740683
## nitrifd_F     -0.092103335 -0.23115841  0.16487359
## minzd_F       -0.028307217 -0.09471401  0.05245243
## soilmoi_F     -0.005851712 -0.32243217  0.31811180
## som_F         -0.115328018 -0.34982544  0.13714394
## nat_g.m2      -0.572687240  0.45247833 -0.36251300
## litter_g.m2   0.391036473  0.03084906  0.20679677
## percpair      -0.631210635 -0.21459259 -0.63141277
## nTrees         1.000000000  0.09391344  0.50260561
## BA_total       0.093913441  1.00000000 -0.15152227
## PercBA_AM     0.502605609 -0.15152227  1.00000000
```

```
row.names(tmp)<-data.q3.rm$plotYear
tmp.scaled<-scale(tmp)
PCA.res<-prcomp(tmp.scaled)
summary(PCA.res)
```

```
## Importance of components:
```

```
##              PC1    PC2    PC3    PC4    PC5    PC6    PC7
## Standard deviation  1.9712 1.6979 1.2651 1.1627 1.0726 0.86978 0.65346
## Proportion of Variance 0.2989 0.2218 0.1231 0.1040 0.0885 0.05819 0.03285
## Cumulative Proportion 0.2989 0.5206 0.6438 0.7478 0.8363 0.89444 0.92729
##              PC8    PC9    PC10    PC11    PC12    PC13
## Standard deviation  0.57207 0.45718 0.42108 0.37217 0.30516 0.006864
## Proportion of Variance 0.02517 0.01608 0.01364 0.01065 0.00716 0.000000
## Cumulative Proportion 0.95246 0.96854 0.98218 0.99283 1.00000 1.000000
```

```
ggbiplot(PCA.res, groups=factor(data.q3.rm$year), ellipse = T) + mytheme
```



```
df.PCA<-data.frame(plotYear=row.names(PCA.res$x), PCA.res$x)
data.q3.PCA<-merge(data.q3.rm, df.PCA[,c(1:3)])

#3. Do reference area and plot characteristics explain variation in Mv biomass?
#a) perMANOVA on mv biomass
perMANOVA<-adonis(tmp.scaled ~ mv_g.m2_logt, data = data.q3.PCA, method='eu')
aov.tab.permanova<-data.frame(perMANOVA$aov.tab)
#save anova table
newfilename<-'permMANOVA_RefxMv.txt'
write.table(aov.tab.permanova, file=paste(figuresPath,newfilename, sep='/'), sep='\t')

#b) mixed effects model that uses 1st two PC scores to predict mv biomass
data.q3.PCA$year<-factor(data.q3.PCA$year)
mod<-lmer(mv_g.m2_logt~PC1*PC2 + (1|year), data=data.q3.PCA)
anovaTable_mv.PC<-data.frame(round(anova(mod),digits=4))
#save anova table
newfilename<-'anovaTable_mvPC.txt'
write.table(anovaTable_mv.PC, file=paste(figuresPath,newfilename, sep='/'), sep='\t')
#plot
pScat_PC2.mv<-ggplot(data.q3.PCA, aes(x=PC2, y=mv_g.m2_logt, color=year)) +
  geom_point() + mytheme + geom_smooth(method='lm') + ylab('Microstegium (log trans. g/m2)')
#save plot
newfilename<-'pScat_RefPC2xMv.png'
png(paste(figuresPath,newfilename, sep='/'), units='in', width = fig.width*1.5, height = fig.height*1,
pScat_PC2.mv
dev.off()
```

```
## pdf
## 2
```

```
#c) Look at how Mv biomass corresponds with variables associated with PC2
```

```
ExplainMvVar<-function(sub.refVars, mvVar, df){
  anova.list<-list()
  figure.list<-list()
  i<-0
  for(i in 1:length(sub.refVars)){

    #identify current data frame
    curr.refVar<-sub.refVars[i]
    x<-df[,curr.refVar]
    y<-df[,mvVar]
    df.i<-data.frame(x=x, y=y, year=df$year)
    df.i$year<-factor(df.i$year)

    #regression model
    mod<-lmer(y~x + (1|year), data=df)
    anova.list[[i]]<-anova(mod)

    #plot
    p<-ggplot(df.i, aes(y=y, x=x, color=year)) + geom_point() +
      mytheme + xlab(NULL) + ylab(NULL) + ggtitle(sub.refVars[i]) + guides(color=FALSE)
    figure.list[[i]]<-p
  }
  names(anova.list)<-sub.refVars
  anovaSummary<-ldply(lapply(anova.list, data.frame))
  anovaSummary[, -1]<-round(anovaSummary[, -1], digits=4)
  anovaSummary<-rename(anovaSummary, c(.id='yVar', Pr..F.='pValue'))

  names(figure.list)<-sub.refVars

  results<-list(anovaSummary=anovaSummary, figures=figure.list)
  return(results)
}

result<-ExplainMvVar(sub.refVars=c('percpair', 'nTrees', 'PercBA_AM', 'nat_g.m2', 'litter_g.m2'),
  mvVar='mv_g.m2_logt',
  df=data.q3.PCA)

#save anova table
newfilename<-'anovaSummary_RefINDxMv.txt'
write.table(result[['anovaSummary']], file=paste(figuresPath, newfilename, sep='/'), sep='\t')

#plot
#grid.arrange(do.call(arrangeGrob, result[['figures']])) #all panels
ylabel<-textGrob("Microstegium biomass (log trans. g/m2)", rot=90)
xlabel<-textGrob("Reference area variable")
#save plot
newfilename<-'pScat_RefxMv.png'
png(paste(figuresPath, newfilename, sep='/'), units='in', width = fig.width*1.5, height = fig.height*3,
  grid.arrange(
```

```

ylabel, # opleft
arrangeGrob(result[['figures']]$percpair + geom_smooth(method='lm'),
            result[['figures']]$nTrees + geom_smooth(method='lm'),
            result[['figures']]$litter_g.m2 + geom_smooth(method='lm'),
            nrow=3, ncol=1),
textGrob(" "), #bottom left
xlabel, #bottom right
widths = unit.c(unit(2.5, "lines"), unit(1, "npc") - unit(2.5, "lines")),
heights = unit.c(unit(1, "npc") - unit(2.5, "lines"), unit(2.5, "lines")),
nrow=2, ncol=2
)
dev.off()

```

```

## pdf
## 2

```

Trends over time

- a. Does *Microstegium* biomass increase over time?
 - b. Do soil N pools and fluxes in reference (and invaded) plots shift over time?
 - c. Do impact magnitudes shift over time?
 - d. Does the influence of reference conditions on *Microstegium* biomass shift over time?
-

Invasion front study design

- a. What environmental factors besides soil N pools and fluxes differ among paired invaded and reference plots?
-

%AM

- a. Does *Microstegium* biomass vary with %AM?
- b. Does %AM vary with multivariate soil N conditions in reference plots?