

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)
library(ggthemes)
library(gridExtra)
library(lme4)
```

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

```
library(lmerTest)
```

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

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

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

```
source('CODE/fixn_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)
timeline<-read.table("DATA/e8_timeline.txt", header=TRUE, sep="\t", stringsAsFactors = FALSE)
```

0. Clean and merge

A. Add identifiers, remove extra columns, merge raw datasets

```
#soilData
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")

#add vegData
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
soilVegData<-merge(soilData_cast, vegData.pruned) #merge soilData_cast and vegData

#add 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_ECM')]
data<-merge(soilVegData,tmp)
```

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

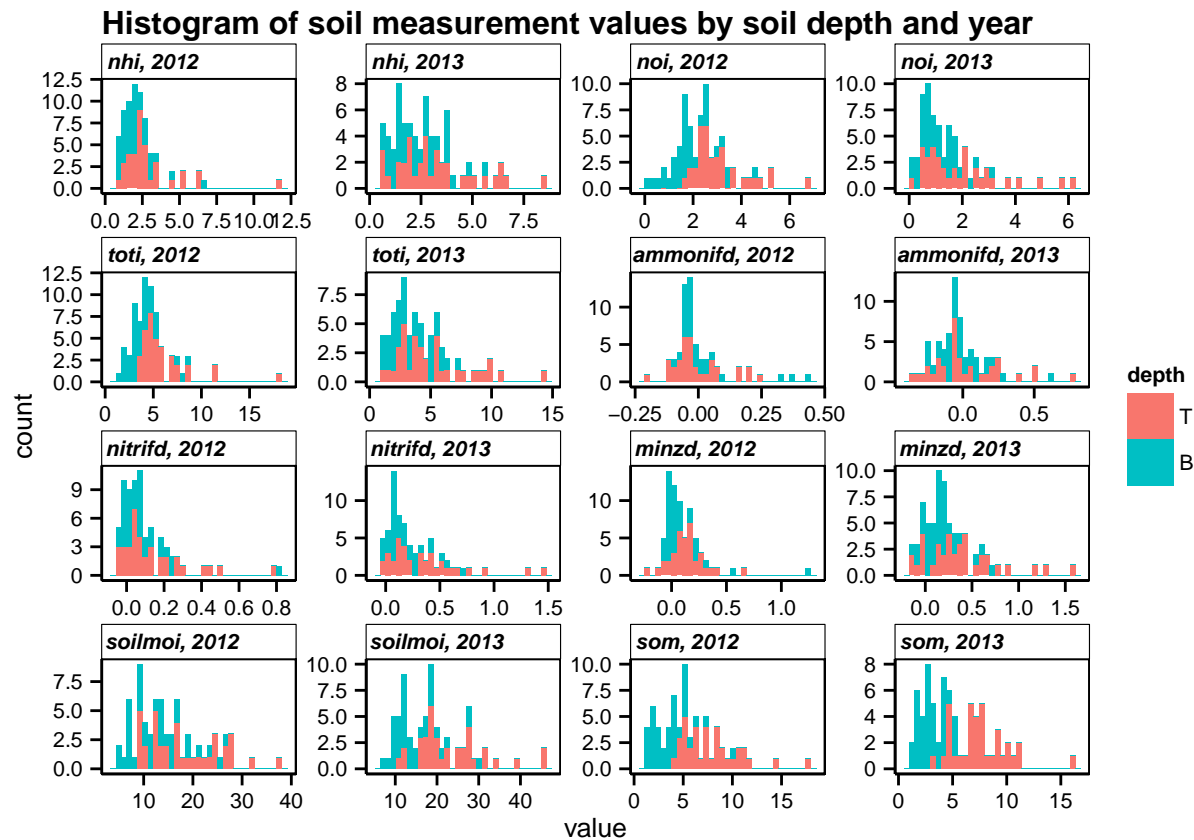
i. Plot histogram of values by depth

```
#subset by year
data.sub<-subset(data, year %in% c(2012, 2013))
#reshape
data.sub_melted <- melt(data.sub,id.vars=c("plotid","plothalfid1","inv","year"))
data.sub_melted$value<-as.numeric(data.sub_melted$value)
#subset by _B and _T
data.sub1<-data.sub_melted[grepl("_B",data.sub_melted$variable) | grepl("_T",data.sub_melted$variable),]
#add back the separate measCat and depth columns
tmp<-ldply(strsplit(as.character(data.sub1$variable),"_", fixed=T))
colnames(tmp)<-c('measCat','depth')
data.sub2<-cbind(data.sub1, tmp)
#correct the dataframe structure
```

```

data.sub2$measCat<-factor(data.sub2$measCat, levels=measCat_order)
data.sub2$depth<-factor(data.sub2$depth, levels=depth_order)
data.sub2$inv<-factor(data.sub2$inv, levels=inv_order)
#plot
pHist.depth<-ggplot(data.sub2, 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=fig.res)
pHist.depth
dev.off()

```

```

## pdf
## 2

```

- ii. DATASET 'data_a.txt' Aggregate values, put aggregated values into full dataset and replace empty _F rows

```

#aggregate by depth
summ.data.sub2<-ddply(data.sub2, ~plotid+plothalid1+inv+year+measCat, summarize,
  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
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.sub2<-rename(summ.data.sub2, replace=c("new.value" = "value")) #rename new.value -> value
summ.data.sub2$variable<-paste(summ.data.sub2$measCat, '_F',sep="")#measCat items should be ammended wi
summ.data.sub3<-summ.data.sub2[,c('plotid','plothalfid1','inv','year','variable','value','note')].#get r

#merge
data_a<-rbind(data_m_woBT_e,summ.data.sub3)
#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$variable %in% c('mv','nat','litter','total'),'varType']<-'understoryBiom'
data_a[data_a$variable %in% c('percpaz','soiltemp','soilBasin'),'varType']<-'environParam'
data_a[data_a$variable %in% c('nTrees','BA_total','PercBA_ECM'),'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_ECM')
plothallevel.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% plothallevel.vars,'measLevel']<-'plothalf'

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

```

- ii. DATASET 'data_d.txt' Keep 2012 and 2013 values dis-aggregated and remove 2011 __F values, remove biomass values from 2011 since time point was past peak biomass

```

#remove 2011 data
data.sub<-subset(data, year != 2011)

#reshape
data_melted <- melt(data.sub,id.vars=c("plotid","plothalfid1","inv","year")) #melt down the full dataset

#remove variables that were only measured in 2011
tmp1<-data_melted[!grepl("_F", data_melted$variable),]
tmp2<-tmp1[tmp1$variable != 'soiltemp',]

```

```

data_d<-tmp2

#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','nat','litter','total'),'varType']<-'understoryBiom'
data_d[data_d$variable %in% c('percpair'),'varType']<-'environParam'
data_d[data_d$variable %in% c('nTrees','BA_total','PercBA_ECM'),'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_ECM')
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'

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

```

1. Q1: Does invader abundance correlate with whole-plot and reference plot environmental conditions? Model: $mvI_logt \sim \text{variable} + (1|\text{year})$

```

#decide which dataset to use:
data<-data_d
#dataframe needs...rows by plot x year
#plotid
#year
#mv_I (invaded area biomass)
#plot-based variables
#native plot half-based variables (measurement type)
#value_N (measurement value in native area)
#value_I (measurement value in invaded area)

#subset mv biomass in invaded area plots by year
data_mv.I<-subset(data, variable=='mv' & inv=='I')
data_mv.I<-rename(data_mv.I, replace=c("value"="mv_I"))
data_mv.I<-data_mv.I[,c('plotid','year','mv_I')]

#subset plot variables
data_plot<-subset(data, measLevel=='plot')

```

```

summ.data_plot<-ddply(data_plot, ~plotid + year + variable, summarize,
  uniq.value=paste(unique(value),collapse = "_"))
data_plot.cast<-dcast(summ.data_plot, plotid + year ~ variable, value.var='uniq.value')

#subset plothalf variables, reshape by inv
data_plothalf<-subset(data, measLevel=='plothalf')
data_plothalf.IN<-dcast(data_plothalf, plotid + year + variable ~ inv, value.var='value')
data_plothalf.N<-data_plothalf.IN[,c('plotid','year','variable','N')]
data_plothalf.N<-rename(data_plothalf.N, replace=c("N"="value_N"))

#merge dataframes
tmp1<-merge(data_mv.I,data_plot.cast)
tmp2<-merge(tmp1, data_plothalf.N)

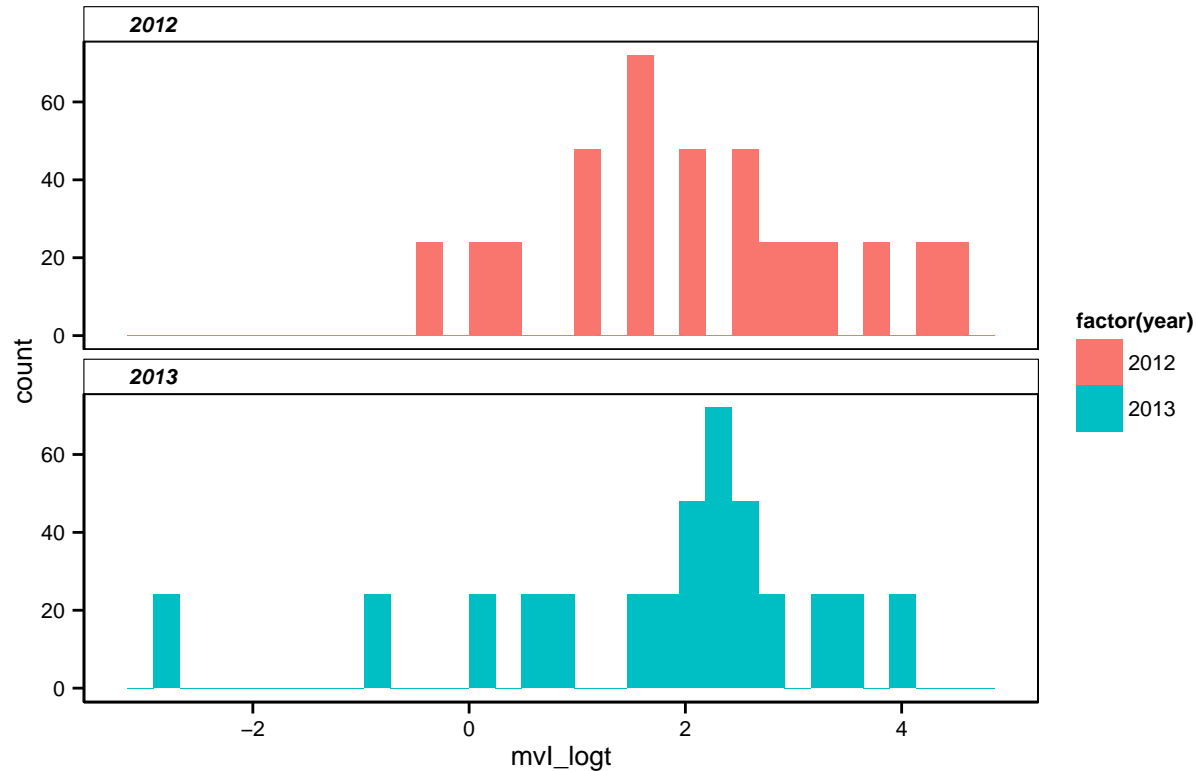
#make plothalf variables wide and then melt everything down
tmp2_cast<-dcast(tmp2, plotid + year + mv_I + nTrees + BA_total + PercBA_ECM ~ variable, value.var='value')
data_mvI <- melt(tmp2_cast,id.vars=c("plotid","year","mv_I")) #melt down the full dataset
data_mvI$value<-as.numeric(data_mvI$value)

#Factor labels
vars<-c(measCat_T, measCat_B, otherPH_vars, P_vars)
var_names<-c(paste(measCat_names,"T", sep="_"), paste(measCat_names,"B", sep="_"),otherPH_names, P_names)
factor.indx<-data.frame(variable=vars, var_names)
data_mvI<-merge(data_mvI, factor.indx)
data_mvI$variable<-factor(data_mvI$variable, levels=vars)
data_mvI$var_names<-factor(data_mvI$var_names, levels=var_names)

#look at distribution of mv biomass by year
pHist.mv<-ggplot(data_mvI,aes(x=mv_I, fill=factor(year))) +
  geom_histogram() + mytheme + ggtitle('Histogram of Mv biomass by year') +
  facet_wrap(~year, scales='fixed', ncol=1)
# #get rid of mv biomass outlier... this is only in 2011
# data_mvI[data_mvI$mv_I>150, 'mv_I']<-NA
# #again, look at distribution of mv biomass by year
# pHist.mv.OR<-ggplot(data_mvI,aes(x=mv_I, fill=factor(year))) +
#   geom_histogram() + mytheme + ggtitle('Histogram of Mv biomass by year\nOutlier in 2011 removed') +
#   facet_wrap(~year, scales='fixed', ncol=1)
# pHist.mv.OR
#log-transform mv biomass
data_mvI$mvI_logt<-log(data_mvI$mv_I)
#plot again
pHist.mv.OR.logt<-ggplot(data_mvI,aes(x=mvI_logt, fill=factor(year))) +
  geom_histogram() + mytheme + ggtitle('Histogram of Log-transformed Mv biomass by year\nOutlier removed') +
  facet_wrap(~year, scales='fixed', ncol=1)
pHist.mv.OR.logt

```

Histogram of Log-transformed Mv biomass by year Outlier removed



```
#save plot
newfilename<-'pHist_mvI_logt.png'
png(paste(figuresPath,newfilename, sep='/'),
    units='in', width = fig.width*1.5, height = fig.height*2, res=fig.res)
pHist.mv.OR.logt
dev.off()
```

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

```
#prep dataset
data_mvI1<-subset(data_mvI, variable != 'total') #this is typically the same as nat in reference plots
data_mvI2<-subset(data_mvI1, variable != 'mv') #there are only 3 observations of Mv biomass in referenc
data_mvI2$variable<-droplevels(data_mvI2$variable)
vars1<-vars[!vars %in% c('total','mv')]
factor.indx1<-factor.indx[!factor.indx$variable %in% c('total','mv'),]
var_names1<-factor.indx1$var_names
data_mvI2$plotid<-factor(data_mvI2$plotid)
data_mvI2$year<-factor(data_mvI2$year)
data_mvI2$measCat<-NA
data_mvI2$depth<-NA
i<-0
for(i in 1:length(measCat_order)){
  data_mvI2[data_mvI2$variable==measCat_B[i], 'measCat']<-measCat_order[i]
  data_mvI2[data_mvI2$variable==measCat_T[i], 'measCat']<-measCat_order[i]
  data_mvI2[data_mvI2$variable==measCat_B[i], 'depth']<- 'B'
```

```

    data_mvI2[data_mvI2$variable==measCat_T[i], 'depth']<-'T'
  }
data_mvI2$measCat<-factor(data_mvI2$measCat, levels=measCat_order)
data_mvI2$depth<-factor(data_mvI2$depth)

#run a linear model to determine the role of year and variable on mvI_logt
#mod1 <- lm(mvI_logt ~ year + value + year:value, data=df)
#mod2 <- lm(mvI_logt ~ value + year + value:year, data=df)
MvYrMods<-FitPlot.MvYr(data_mvI2)
MvYrMods_table<-ldply(MvYrMods[['mod1L']])
MvYrMods_table$.id<-factor(MvYrMods_table$.id, levels = vars)
MvYrMods_table<-dcast(MvYrMods_table, .id ~ terms, value.var = 'pvals.a') #these are anova p-values from
MvYrMods_table #no effect of year or year:value

```

```

##      .id value  year year:value
## 1     nhi_T 0.335 0.446      0.618
## 2     noi_T 0.372 0.448      0.933
## 3     toti_T 0.297 0.445      0.675
## 4 ammonifd_T 0.893 0.436      0.115
## 5  nitrifd_T 0.224 0.444      0.911
## 6   minzd_T 0.362 0.445      0.514
## 7  soilmoi_T 0.577 0.450      0.629
## 8     som_T 0.505 0.450      0.657
## 9     nhi_B 0.952 0.448      0.357
## 10    noi_B 0.888 0.454      0.801
## 11    toti_B 0.914 0.449      0.418
## 12 ammonifd_B 0.167 0.564      0.243
## 13  nitrifd_B 0.026 0.360      0.723
## 14   minzd_B 0.457 0.450      0.804
## 15  soilmoi_B 0.793 0.447      0.336
## 16     som_B 0.308 0.446      0.817
## 17     nat 0.002 0.384      0.689
## 18    litter 0.002 0.384      0.739
## 19  percpa 0.000 0.457      0.615
## 20    nTrees 0.000 0.358      0.961
## 21   BA_total 0.617 0.452      0.808
## 22 PercBA_ECM 0.039 0.442      0.527

```

```

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

#run a mixed effects model to determine the role of year and measCat on mvI_logt, with depth as a random
#mod <- lmer(mvI_logt ~ year + value + year:value + (1|depth), data=df)
MvYrDpMods<-FitPlot.MvYrDp(data_mvI2)
MvYrDpMods_table<-ldply(MvYrDpMods[['results']])
MvYrDpMods_table$.id<-factor(MvYrDpMods_table$.id, levels = measCat_order)
MvYrDpMods_table<-dcast(MvYrDpMods_table, .id ~ terms, value.var = 'pvals.a') #these are anova p-values
MvYrDpMods_table #no effects

```

```

##      .id value  year year:value
## 1     nhi 0.736 0.648      0.899

```



```
## 2      noi 0.909 0.648      0.936
## 3      toti 0.774 0.704      0.954
## 4 ammonifd 0.382 0.275      0.475
## 5      nitrifd 0.729 0.105      0.916
## 6      minzd 0.924 0.191      0.754
## 7      soilmoi 0.545 0.740      0.416
## 8      som 0.426 0.998      0.575
```

#save and export

```
newfilename<-'MvYrDpMods_table.txt'
```

```
write.table(MvYrDpMods_table, file=paste(synthdataPath,newfilename, sep='/'), sep='\t')
```

#run mixed effects models where year is a random effect and each variable is a fixed effect

```
#mod <- lmer(mvI_logt ~ value + (1/year), data = df)
```

```
MvMods<-FitPlot.Mv(data_mvI2)
```

```
MvMods_table<-ldply(MvMods[['results']])
```

```
MvMods_table
```

```
##      .id terms pvals.a
## 1      nhi_T value 0.355
## 2      noi_T value 0.249
## 3      toti_T value 0.239
## 4 ammonifd_T value 0.978
## 5      nitrifd_T value 0.352
## 6      minzd_T value 0.519
## 7      soilmoi_T value 0.456
## 8      som_T value 0.468
## 9      nhi_B value 0.911
## 10     noi_B value 0.808
## 11     toti_B value 0.974
## 12 ammonifd_B value 0.254
## 13      nitrifd_B value 0.035
## 14      minzd_B value 0.585
## 15      soilmoi_B value 0.708
## 16      som_B value 0.226
## 17      nat value 0.002
## 18      litter value 0.002
## 19      percpar value 0.000
## 20      nTrees value 0.000
## 21      BA_total value 0.610
## 22 PercBA_ECM value 0.036
```

#save and export

```
newfilename<-'MvMods_table.txt'
```

```
write.table(MvMods_table, file=paste(synthdataPath,newfilename, sep='/'), sep='\t')
```

```
MvFE_table<-ldply(MvMods[['coef']])
```

```
MvFE_table
```

```
##      .id int est pVal.est df.est pr2
## 1      nhi_T 1.47 0.14 0.35 34.00 -0.01
## 2      noi_T 1.37 0.22 0.25 34.00 0.00
## 3      toti_T 1.27 0.12 0.24 34.00 -0.01
## 4 ammonifd_T 1.92 0.04 0.98 34.00 0.02
```

```
## 5   nitrifd_T 1.68  0.82      0.35  20.62  0.02
## 6   minzd_T  1.79  0.47      0.52  28.09  0.01
## 7   soilmoi_T 2.39 -0.02     0.46  34.00 -0.03
## 8   som_T    1.40  0.07      0.47  34.00 -0.02
## 9   nhi_B    1.97 -0.02     0.91  34.00 -0.01
## 10  noi_B    1.81  0.07      0.81  34.00  0.00
## 11  toti_B   1.90  0.00      0.97  34.00 -0.02
## 12 ammonifd_B 1.77 -2.44     0.25  32.00  0.04
## 13 nitrifd_B 1.35  4.41      0.04  31.75  0.05
## 14 minzd_B   1.82  1.11      0.58  34.00  0.03
## 15 soilmoi_B 2.17 -0.02     0.71  34.00 -0.03
## 16 som_B     1.28  0.18      0.23  34.00  0.00
## 17 nat       1.32  0.06      0.00  33.46  0.03
## 18 litter    3.17 -0.01      0.00  33.96 -0.01
## 19 percpar   1.21  0.03      0.00  27.39  0.06
## 20 nTrees    3.33 -0.34      0.00  34.00  0.08
## 21 BA_total  2.09 -0.78      0.61  34.00  0.02
## 22 PercBA_ECM 1.25  0.01      0.04  30.00 -0.04
```

```
#save and export
newfilename<-'MvFE_table.txt'
write.table(MvFE_table, file=paste(synthdataPath,newfilename, sep='/'), sep='\t')
#save plot
newfilename<-'pScat_mvI_logt.png'
png(paste(figuresPath,newfilename, sep='/'),
    units='in', width = fig.width*5, height = fig.height*5, res=fig.res)
grid.arrange(
  textGrob('Microstegium biomass (log-transformed)', y=0.5, x=unit(1,'lines'), rot=90), # ylabel to
  do.call(arrangeGrob, MvMods[['figures']]),
  textGrob(" "), #bottom left
  textGrob("Variable value", x=0.5, y=unit(1,'lines')), #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
```

```
# #check out the relationships between nTrees, BA_total, PercBA_ECM, percpar
# data.sub<-subset(data_mvI3, variable %in% c('nTrees','BA_total','percBA_ECM','percpar'))
# colnames(data.sub)
```

2. Q2: How does reference nutrient availability correlate with invaded nutrient availability across sites?

A. Piece-meal approach. ...Model: $\text{invVar} \sim \text{natVar} + (1|\text{plotid}+\text{year})$.. not really sure how to specify this...
or Model: $\text{esVar} \sim 1 + (1|\text{year})$

B. Ordination approach... Ordinate plot halfids in soil N space and look whether inv corresponds to variation and how soil N components load onto the axes

3. Q3: Do the following factors correspond to changes in soil N effect sizes: i) invader abundance, ii) time, iii) reference plot som?

A. Piece-meal approach... Model: $esVar \sim \text{variable} + (1|\text{year})$

B. Ordination approach... ordinate plotids in effect size soil N space and look whether any of the factors are important