

MIIN Part 3: Calculate effect sizes

Marissa Lee

June 1, 2015

Filename: MIIN_3_calcEffectSizes.Rmd

This markdown file does the following tasks: 1. Clean dataframes: A) Remove observations that do not fit meta-analysis criteria that it must have least 1 focal exotic invasive species AND at least 1 nonfocal species, B) Remove ancillary soil measurements that will not be used in the meta-analysis, C) Simplify ecosystem type factor, D) Identify the N-fixing species, E) Create a character string to identify the invasive species associated with each observation ... then, F) rename all the clean dataframes

2. Calculate invasion effect sizes
3. Create a composite dataset for analysis. Look at A) dataset structure, B) distribution of invasion effect size values, C) distribution of unit-standardized soil measurement values, and C) distribution of cwm trait values
4. Export dataframes

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

```
library(plyr)
library(reshape2)
library(ggplot2)
library(metafor)
```

```
## Loading required package: Matrix
## Loading 'metafor' package (version 1.9-7). For an overview
## and introduction to the package please type: help(metafor).
```

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

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

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

```
#data synthesized by *MIIN_1_paperData.Rmd*
```

```
papers<-read.table("DATA/DATA_SYNTHESIZED/paperData/papers_procd.txt", header=TRUE, sep="\t", quote="")
observations<-read.table("DATA/DATA_SYNTHESIZED/paperData/observations_procd.txt", header=TRUE, sep="\t")
cover<-read.table("DATA/DATA_SYNTHESIZED/paperData/cover_procd.txt", header=TRUE, sep="\t")
species<-read.table("DATA/DATA_SYNTHESIZED/paperData/species_procd.txt", header=TRUE, sep="\t")
traits<-read.table("DATA/DATA_SYNTHESIZED/paperData/traits_procd.txt", header=TRUE, sep="\t")
measures<-read.table("DATA/DATA_SYNTHESIZED/paperData/measures_procd.txt", header=TRUE, sep="\t")
```

```

#data synthesized by *MIIN_2_cwm.Rmd*
cwm<-read.table("DATA/DATA_SYNTHESIZED/cwm/cwm.txt", header=TRUE, sep="\t")
spIDcover<-read.table("DATA/DATA_SYNTHESIZED/cwm/spIDcover.txt", header=TRUE, sep="\t")
spIDtraits<-read.table("DATA/DATA_SYNTHESIZED/cwm/spIDtraits.txt", header=TRUE, sep="\t")

#read-in the list of Frankia and Rhizobia-associated plant genera
nfixGenera<-read.table("DATA/NfixingPlantGenera.txt", header=TRUE, sep="\t")
legumeGenera<-read.table("DATA/Leguminosae.csv", header=TRUE, sep=',')

```

1. Clean dataframes

1A. Remove observations that do not fit meta-analysis criteria that they must have least 1 focal exotic invasive species AND at least 1 nonfocal species. Observations will be removed from 'observations' dataframe and the corresponding related data in the following dataframes: cover, species, traits, measures, cwm, papers.

```

summ.spp <- ddply(species,~obsID,summarise,
                  numTotalspp=length(obsID),
                  numInv spp=sum(spInvasive=='invasive' & spExotic=='exotic' & spFocal=='focal'),
                  numNonFocalspp=sum(spFocal=='not focal'))
exclude.1inv<-summ.spp[summ.spp$numInv spp == 0,'obsID']
exclude.1nonfoc<-summ.spp[summ.spp$numNonFocal == 0,'obsID']
exclude.tmp<-c(exclude.1inv,exclude.1nonfoc)
exclude.obsID<-unique(exclude.tmp)

paste('Exclude',length(exclude.obsID), 'observations because there is not at least 1 species that is invasive, exotic, AND focal')

```

```
## [1] "Exclude 86 observations because there is not at least 1 species that is invasive, exotic, AND focal"
```

```

#identify the ok obsIDs
observations1<-observations[!observations$obsID %in% exclude.obsID,]
obsOK<-unique(observations1$obsID)
paste('There are',length(obsOK), 'observations remaining in the dataset that have at least 1 species that is invasive, exotic, AND focal')

```

```
## [1] "There are 404 observations remaining in the dataset that have at least 1 species that is invasive, exotic, AND focal"
```

```

#subset the remaining dfs from paperData
cover1<-subset(cover, obsID %in% obsOK)
species1<-subset(species, obsID %in% obsOK)
traits1<-subset(traits, obsID %in% obsOK)
measures1<-subset(measures, obsID %in% obsOK)

#subset the dfs from cwm
cwm1<-subset(cwm, obsID %in% obsOK)
spT0obs<-function(df){ #first, need to convert the spID to an obsID column in these dfs
  tmp<-ldply(strsplit(as.character(df[, 'spID']), "."), fixed=T)
  df[, 'obsID']<-paste(tmp[,1],tmp[,2], sep=".")
  return(df)
}

```

```

spIDcover<-spT0obs(spIDcover)
spIDtraits<-spT0obs(spIDtraits)
spIDcover1<-subset(spIDcover, obsID %in% obsOK)
spIDtraits1<-subset(spIDtraits, obsID %in% obsOK)

#annotate papers dataframe to reflect removal of observations and thus papers
#identify which observations in the exclude list come from papers that had OK'd observations
exclude.p1<-ldply(strsplit(as.character(exclude.obsID), "."), fixed=T))[,1]
include.p1<-ldply(strsplit(as.character(obsOK), "."), fixed=T))[,1]
reject.p<-unique(exclude.p1[!exclude.p1 %in% include.p1]) #if FALSE, then label these paperIDs as reject
papers[papers$paperID %in% reject.p,'reject']<-'Yes'
newRationale<-'Not at least 1 species that is invasive, exotic, AND focal'
papers$rejectRationale<-factor(papers$rejectRationale, levels=c(levels(papers$rejectRationale),newRationale))
papers[papers$paperID %in% reject.p,'rejectRationale']<-newRationale
papers1<-papers

```

1B. Remove ancillary soil measurements that will not be used in the meta-analysis.

```

summ.meas <- ddply(measures1,~measCat,summarise, numObs=length(unique(obsID)))
summ.meas

```

```

##      measCat numObs
## 1    ammonif     54
## 2      biom    125
## 3       cn     14
## 4 litterbiom    38
## 5 littercn     21
## 6 litterpercN   26
## 7 microbcn     11
## 8       nh    162
## 9    nitrif     85
## 10    nminz    128
## 11      no    177
## 12    percN     53
## 13      ph    154
## 14 plantcov      3
## 15    soilcn    126
## 16    soilmoi   108
## 17    soiln    212
## 18      som     97
## 19     toti    225

```

```

removeMeasCats<-c('biom','litterbiom','plantcov',
                  'cn','littercn','percN','litterpercN',
                  'microbcn','ph','soilmoi')
paste('Remove these measurement types:', paste(removeMeasCats, collapse=", "))

```

```
## [1] "Remove these measurement types: biom, litterbiom, plantcov, cn, littercn, percN, litterpercN, m
```

```
measOK<-summ.meas[!summ.meas$measCat %in% removeMeasCats,'measCat']
```

```
#subset the remaining dfs from paperData
```

```
measures2<-subset(measures1, measCat %in% measOK)
measures3 <- droplevels(measures2)
```

1C. Simplify ecosystem type factor The 'other' category now consists of studies that took place in a dune system, or some combination of forest, grassland, wetland

```
summ.obs.eco <- ddply(observations1,~ecosystCat,summarise,
                      numObs=length(paperID),
                      numPapers=length(unique(paperID)))
summ.obs.eco
```

```
##           ecosystCat numObs numPapers
## 1             dune      2         2
## 2          forest    123         51
## 3 forest,grassland      1          1
## 4 forest,grassland,wetland 1          1
## 5          grassland    176         53
## 6          shrubland     73         23
## 7           wetland     28         14
```

```
#limit ecosystem categories to forest, grassland, shrubland, wetland, and other
criteria<-observations1$ecosystCat == 'forest,grassland' |
observations1$ecosystCat == 'forest,grassland,wetland' |
observations1$ecosystCat == 'dune'
levels(observations1$ecosystCat) <- c(levels(observations1$ecosystCat), "other")
observations1[criteria,'ecosystCat']<-'other'
```

1D. Identify the N-fixing plant species

```
#pull the unique legume genuses and combine with the full list of Frankia and Rhizobia associated plant
LegGenera<-unique(legumeGenera$Genus)
nfixGenera.sub<-nfixGenera[!nfixGenera$plantFamily == 'Leguminosae', 'plantGenera']
nfixGenera.complete<-c(as.character(LegGenera),as.character(nfixGenera.sub))

#select rows in 'species' dataframe based on whether the species' genus name is present in nfixGenera.c
species1$nfixGenus<-'No' #fill everything in with 'No' first
species1[species1$Genus %in% nfixGenera.complete,'nfixGenus']<-'Yes'
numNfixRows<-dim(species1[species1$Genus %in% nfixGenera.complete,])[1]
numNonNfixRows<-dim(species1[!species1$Genus %in% nfixGenera.complete,])[1]
paste('There are', numNfixRows, 'and', numNonNfixRows, 'species entries (species unique to each observat.
```

```
## [1] "There are 192 and 1806 species entries (species unique to each observation) that are N-fixing and"
```

```
#identify observations based on presence/absence of N-fixer as invasive species
selection<-species1$spInvasive == 'invasive' & species1$spExotic == 'exotic' & species1$spFocal == 'focal'
df.selection<-species1[selection,]
NfixObsIDs<-unique(df.selection$obsID)
observations1$InvNfix<-'Non-N-fixing' #fill everthing with 'Non-N-fixing' first
observations1[observations1$obsID %in% NfixObsIDs,'InvNfix']<-'N-fixing'
numNfixRows<-sum(observations1$InvNfix == 'N-fixing')
numNonNfixRows<-sum(observations1$InvNfix != 'N-fixing')
paste('There are', numNfixRows, 'and', numNonNfixRows, 'observations that have N-fixing and non-N-fixing
```

```
## [1] "There are 71 and 333 observations that have N-fixing and non-N-fixing invasive species, respectively"
```

```
#identify observations based on % native N-fixing species (not cover)
df.notFocal<-species1[species1$spFocal == 'not focal',]
summ.Nfix <- ddply(df.notFocal,~obsID,summarise,
                  NatnumNfix=sum(nfixGenus=='Yes'),
                  NatnumNotNfix=sum(nfixGenus=='No'),
                  NatpercNfix=(NatnumNfix/(NatnumNfix + NatnumNotNfix))*100)
summ.Nfix$NatNfix<-'No N-fixers' #fill everthing with 'No N-fixers' first
summ.Nfix[summ.Nfix$NatnumNfix > 0,'NatNfix']<-'N-fixers present'
numNfixRows<-sum(summ.Nfix$NatNfix == 'N-fixers present')
numNonNfixRows<-sum(summ.Nfix$NatNfix != 'N-fixers present')
paste('There are', numNfixRows, 'and', numNonNfixRows, 'observations that have reference areas with N-fixers present')
```

```
## [1] "There are 58 and 346 observations that have reference areas with N-fixers present and No N-fixers present"
```

```
#create a composite Nfix factor
observations2<-merge(observations1, summ.Nfix, by='obsID')
observations2$Nfix<-paste(observations2$InvNfix, observations2$NatNfix, sep="_")
unique(observations2$Nfix)
```

```
## [1] "N-fixing_No N-fixers"          "Non-N-fixing_No N-fixers"
## [3] "Non-N-fixing_N-fixers present" "N-fixing_N-fixers present"
```

```
observations2$Nfix<-revalue(observations2$Nfix,
                           c("Non-N-fixing_No N-fixers" = "No N-fixers",
                             "N-fixing_No N-fixers" = "Invasive N-fixers only",
                             "Non-N-fixing_N-fixers present" = "Resident N-fixers only",
                             "N-fixing_N-fixers present" = "Invasive and resident N-fixers"))
```

1E. Create a character string to identify the invasive species associated with each observation

```
#create an obsID x invasive species dataframe
species.tmp<-subset(species1, spInvasive=='invasive' & spExotic=='exotic' & spFocal=='focal')
OBSID<-unique(species.tmp$obsID)
bindedrows<-numeric(0)
i<-0
for(i in 1:length(OBSID)){
  invGenera<-paste(species.tmp[species.tmp$obsID == OBSID[i],'Genus'], collapse='_')
  nspecies<-length(species.tmp[species.tmp$obsID == OBSID[i],'Genus'])
  if(nspecies > 2){
    invGenera<->2spp'
  }
  row<-data.frame(obsID=OBSID[i], invGenera)
  bindedrows<-rbind(bindedrows,row)
}
species.tmp2<-bindedrows
#View(species.tmp2)

#merge by obsID to add invasive species name to observations table
observations3<-merge(observations2, species.tmp2, by='obsID')
```

1F. Rename all the clean dataframes

```
papers.c<-papers1
observations.c<-observations3
cover.c<-cover1
species.c<-species1
traits.c<-traits1
measures.c<-measures3
cwm.c<-cwm1
spIDcover.c<-spIDcover1
spIDtraits.c<-spIDtraits1
```

2. Calculate invasion effect sizes and create a composite dataset for analyses (observations, measures, cwm)

To calculate invasion effect sizes, use measurement values that have not been unit-standardized. Calculate effect sizes using the “standard mean difference” (SMD).

```
chooseMeasType<-'nonSTD' #decide whether to use standardized/non-standardized soil measurement values
chooseESType<-'SMD' #decide whether to use ROM or SMD to calculate effect sizes

### Data that will be used #####
#observations.c
#measures.c
#cwm.c

### Calculate measurement ESs #####

#study identifiers
obsID<-measures.c$obsID
measCat<-measures.c$measCat
measQuality<-measures.c$YN

#invader impact measures - standardized units
n1i<-measures.c$inv_n
m1i<-measures.c$inv_mean_std
sd1i<-sqrt(measures.c$inv_var_std)
n2i<-measures.c$nat_n
m2i<-measures.c$nat_mean_std
sd2i<-sqrt(measures.c$nat_var_std)

dat.STD<-data.frame(obsID, measCat, measQuality,
                    n1i, m1i, sd1i, n2i, m2i, sd2i)

#invader impact measures - non-standardized units
m1i<-measures.c$inv_mean
sd1i<-sqrt(measures.c$inv_var)
m2i<-measures.c$nat_mean
sd2i<-sqrt(measures.c$nat_var)
```

```

dat.nonSTD<-data.frame(obsID, measCat, measQuality,
                      n1i, m1i, sd1i, n2i, m2i, sd2i)

### Combine the selected measurement ES values with the observation ID modifiers #####
if(chooseMeasType == 'STD'){dat<-dat.STD}
if(chooseMeasType == 'nonSTD'){dat<-dat.nonSTD}

#add obsID factor columns to measures
dat.obs<-merge(dat,observations.c, by='obsID')
#dim(dat);dim(dat.obs) #should more columns, but same number of rows
#colnames(dat.obs) #get rid of unnecessary columns
dat.obs1<-dat.obs[,c('paperID','obsID','measCat',
                    'n1i', 'm1i', 'sd1i', 'n2i', 'm2i', 'sd2i',
                    'measQuality',
                    'ecosystCat','studyType','InvNfix','NatNfix','Nfix','invGenera')]

### Re-organize cwm data and prep for merging #####
#recast cwm so that type of CWM values are in the same row
cwm.tmp<-cwm.c[,c('obsID','traitCat','invType','qualRank','cwm')]
m.cwm.tmp<-melt(cwm.tmp, id.vars=c('obsID', 'traitCat','invType'))
c.cwm.tmp<-dcast(m.cwm.tmp, obsID+traitCat~invType+variable)
#View(c.cwm.tmp)

#calculate the raw difference of invaded and native area cwm trait values
c.cwm.tmp$CWMDiff_cwm<-c.cwm.tmp$InvArea_cwm - c.cwm.tmp$NatArea_cwm

### Combine the selected measurement ES values and obsID info with CWM data #####
#add cwm data to measures
#View(dat.obs1) #each row is a unique obsID x measCat
#View(c.cwm.tmp) #each row is a unique obsID x traitCat
dat.all<-merge(dat.obs1, c.cwm.tmp, by='obsID', all=TRUE)
#dim(dat.all)
#paste(length(unique(dat.obs1$obsID)), 'observations') #check to make sure that observations did not ge
#paste(length(unique(dat.all$obsID)), 'observations')

### Calculate the effect sizes #####
dat1 <- escalc(measure=chooseESType, m1i=m1i, sd1i=sd1i, n1i=n1i, m2i=m2i, sd2i=sd2i, n2i=n2i, data=dat

```

3. Now that we have a composite dataset for analysis. Look at...

3A. Dataset structure

```

#summarize dataset by unique obsID+measCats so that data is not duplicated (multiple traits per obsID+m
summ<-ddply(dat1, ~obsID+measCat, summarize,
            uniqm1i = length(unique(m1i)),
            uniqm2i = length(unique(m2i)),
            uniqyi = length(unique(yi)),

```

```

    total = sum(uniqm1i, uniqm2i, uniqyi))
sum(summ$total != 3) # if 0, then obsID + measCat produces all unique rows

```

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

```

dat1.meas<-ddply(dat1, ~obsID+measCat, summarize,
  m1i = unique(m1i),
  m2i = unique(m2i),
  yi = unique(yi))
#head(dat1.meas) #each row is a unique obsID x measCat

#summarize dataset by unique obsID+traitCats so that data is not duplicated (multiple measures per obsID)
summ<-ddply(dat1, ~obsID+traitCat, summarize,
  uniqInvArea = length(unique(InvArea_cwm)),
  uniqInvSpInvArea = length(unique(InvSpInvArea_cwm)),
  uniqNatArea = length(unique(NatArea_cwm)),
  uniqCWMDiff = length(unique(CWMDiff_cwm)),
  total = sum(uniqInvArea, uniqInvSpInvArea, uniqNatArea, uniqCWMDiff))
sum(summ$total != 4) # if 0, then obsID + traitCat produces all unique rows

```

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

```

dat1.tr<-ddply(dat1, ~obsID+traitCat, summarize,
  InvArea = unique(InvArea_cwm),
  InvSpInvArea = unique(InvSpInvArea_cwm),
  NatArea = unique(NatArea_cwm),
  CWMDiff = unique(CWMDiff_cwm))
#head(dat1.tr) #each row is a unique obsID x traitCat

#summarize dataset by unique obsID+traitCats+variable (where variable == c(InvArea, InvSpInvArea, NatArea))
m.dat1.tr<-melt(dat1.tr, id.vars=c('obsID', 'traitCat'))
#head(m.dat1.tr) #each row is a unique obsID x traitCat x variable
#sum(is.na(m.dat1.tr$traitCat)) #combination is not fully factorial, so there are NAs here
#sum(is.na(m.dat1.tr$value)) #combination is not fully factorial, so there are NAs here
m.dat1.tr1<-m.dat1.tr[!is.na(m.dat1.tr$value),] #get rid of the NAs

```

3B. Distribution of effect size values

```

dat1.meas_tmp<-dat1.meas[!is.na(dat1.meas$yi),]

pHist_ES<-ggplot(dat1.meas_tmp, aes(x=yi)) + facet_wrap(~measCat, scales='free', ncol=3) + geom_histogram(
  mytheme + scale_y_continuous(expand = c(0,0)) + ggtitle('Histogram of effect size values'))

newfilename<-'pHist_ES.png'
png(paste(figuresPath,newfilename, sep='/'),
  units='in', width = fig.width*3, height = fig.height*6, res=fig.res)
pHist_ES
dev.off()

```

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



```
#get rid of outliers
```

```
#nh
```

```
dat1[dat1$measCat=='nh' & dat1$yi > 30 & !is.na(dat1$yi),]
```

```
##      obsID paperID measCat n1i  m1i  sd1i n2i  m2i  sd2i measQuality ecosystCat
## 1 706.01      706      nh   5 15.1  0.3   5 4.5  0.2 NoAgg.Conv grassland
## 2 706.01      706      nh   5 15.1  0.3   5 4.5  0.2 NoAgg.Conv grassland
##      studyType      InvNfix      NatNfix      Nfix invGenera traitCat
## 1 field study Non-N-fixing No N-fixers No N-fixers Solidago      cn
## 2 field study Non-N-fixing No N-fixers No N-fixers Solidago      percN
##      InvArea_qualRank InvArea_cwm InvSpInvArea_qualRank InvSpInvArea_cwm
## 1      2      28.88740      2      28.88740
## 2      2      1.67977      2      1.67977
##      NatArea_qualRank NatArea_cwm CWMDiff_cwm      yi      vi
## 1      1      31.950795 -3.0633950 37.5314 70.8301
## 2      0      1.960199 -0.2804294 37.5314 70.8301
```

```
dat1[dat1$measCat=='nh' & dat1$yi > 30 & !is.na(dat1$yi), 'yi'] <- NA #replace outlier with NA
#no
```

```
dat1[dat1$measCat=='no' & dat1$yi > 30 & !is.na(dat1$yi),]
```

```
##      obsID paperID measCat n1i  m1i      sd1i n2i  m2i      sd2i
## 1  57.02      57      no  20 0.05 4.472136e-05  20 1e-05 4.472136e-05
## 2  57.02      57      no  20 0.05 4.472136e-05  20 1e-05 4.472136e-05
## 3 693.03     693      no   2 2.00 1.414214e-05   2 1e+00 1.414214e-05
##      measQuality ecosystCat      studyType      InvNfix      NatNfix      Nfix
## 1 NoAgg.Conv      wetland field study Non-N-fixing No N-fixers No N-fixers
## 2 NoAgg.Conv      wetland field study Non-N-fixing No N-fixers No N-fixers
## 3 NoAgg.Conv      forest field study Non-N-fixing No N-fixers No N-fixers
##      invGenera traitCat InvArea_qualRank InvArea_cwm InvSpInvArea_qualRank
## 1  Lythrum      cn      3      25.4895      3
## 2  Lythrum      percN      3      1.7957      3
## 3 Asparagus      percN      2      2.9475      2
##      InvSpInvArea_cwm NatArea_qualRank NatArea_cwm CWMDiff_cwm      yi
## 1      26.9600      3      25.26340      0.22610 1095.5776
## 2      1.6400      3      1.81964      -0.02394 1095.5776
## 3      2.9475      2      2.94750      0.00000 39894.2280
##      vi
## 1      15003.7280
## 2      15003.7280
## 3 198943679.8649
```

```
dat1[dat1$measCat=='no' & dat1$yi > 30 & !is.na(dat1$yi), 'yi'] <- NA #replace outlier with NA
# #ph
```

```
# dat1[dat1$measCat=='ph' & dat1$yi > 30 & !is.na(dat1$yi),]
```

```
# dat1[dat1$measCat=='ph' & dat1$yi < -30 & !is.na(dat1$yi),]
```

```
# dat1[dat1$measCat=='ph' & dat1$yi > 30 & !is.na(dat1$yi), 'yi'] <- NA #replace outlier with NA
```

```
# dat1[dat1$measCat=='ph' & dat1$yi < -30 & !is.na(dat1$yi), 'yi'] <- NA #replace outlier with NA
```

```
#soiln
```

```
dat1[dat1$measCat=='soiln' & dat1$yi > 30 & !is.na(dat1$yi),]
```

```
##      obsID paperID measCat n1i  m1i      sd1i n2i  m2i      sd2i
```

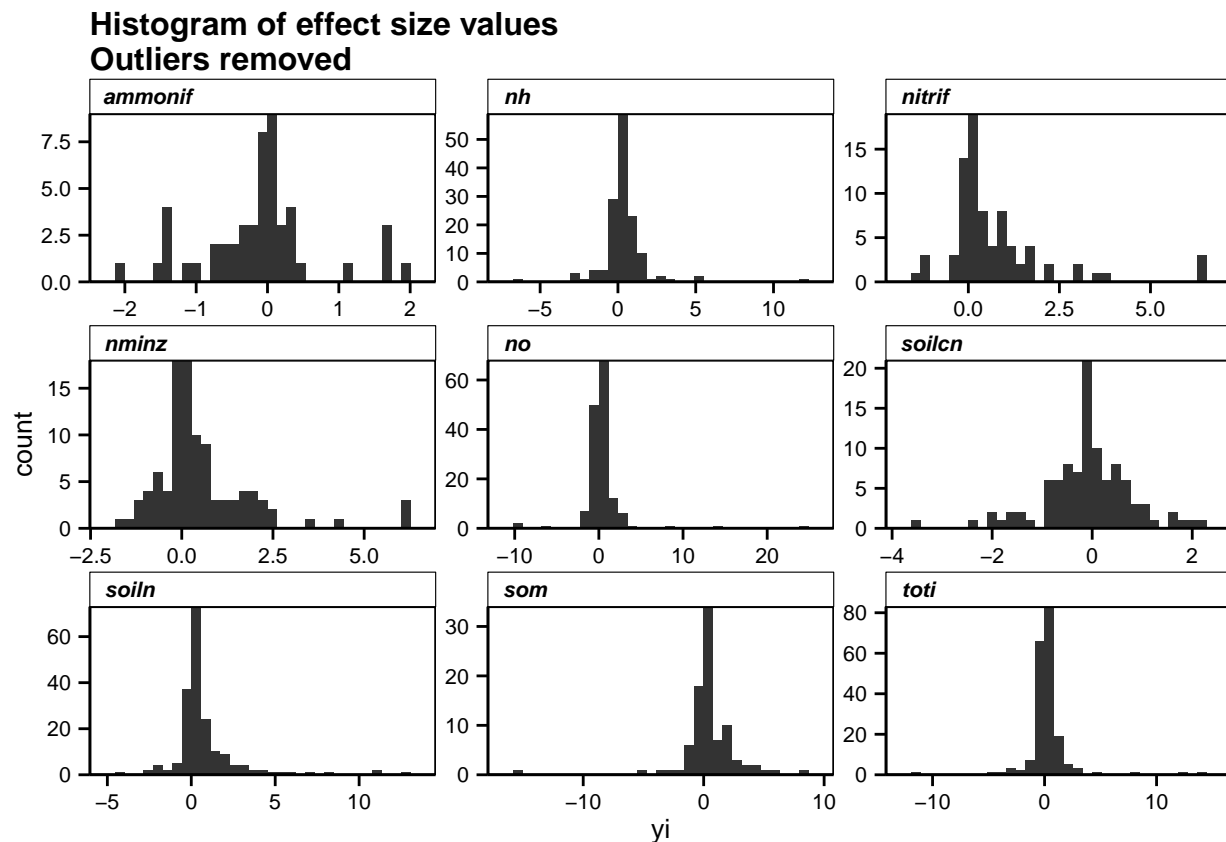
```
## 1 663.01      663    soiln  32 0.3 5.656854e-05  32 0.2 5.656854e-05
## 2 663.01      663    soiln  32 0.3 5.656854e-05  32 0.2 5.656854e-05
##   measQuality ecosystCat   studyType      InvNfix      NatNfix      Nfix
## 1 NoAgg.NoConv  shrubland field study Non-N-fixing No N-fixers No N-fixers
## 2 NoAgg.NoConv  shrubland field study Non-N-fixing No N-fixers No N-fixers
##   invGenera traitCat InvArea_qualRank InvArea_cwm InvSpInvArea_qualRank
## 1    >2spp      cn          1    27.079636                1
## 2    >2spp    percN          2     1.437404                2
##   InvSpInvArea_cwm NatArea_qualRank NatArea_cwm CWMDiff_cwm      yi
## 1          25.53359                1    29.7324372   -2.6528016 1746.2816
## 2          2.15343                2     0.8154332    0.6219708 1746.2816
##           vi
## 1 23824.2757
## 2 23824.2757
```

```
dat1[dat1$measCat=='soiln' & dat1$yi > 30 & !is.na(dat1$yi), 'yi']<-NA #replace outlier with NA
```

```
#update and re-plot
```

```
dat1.meas<-ddply(dat1, ~obsID+measCat, summarize,
  m1i = unique(m1i),
  m2i = unique(m2i),
  yi = unique(yi))
```

```
pHist_ES_OR<-ggplot(dat1.meas, aes(x=yi)) + facet_wrap(~measCat, scales='free', ncol=3) + geom_histogram(
  mytheme + scale_y_continuous(expand = c(0,0)) + ggtitle('Histogram of effect size values\nOutliers removed
pHist_ES_OR
```



```
paste('Removed outlier effect size values in nh, no, soiln.')
```

```
## [1] "Removed outlier effect size values in nh, no, soiln."
```

```
paste('Effect size values look normally-distributed-ish')
```

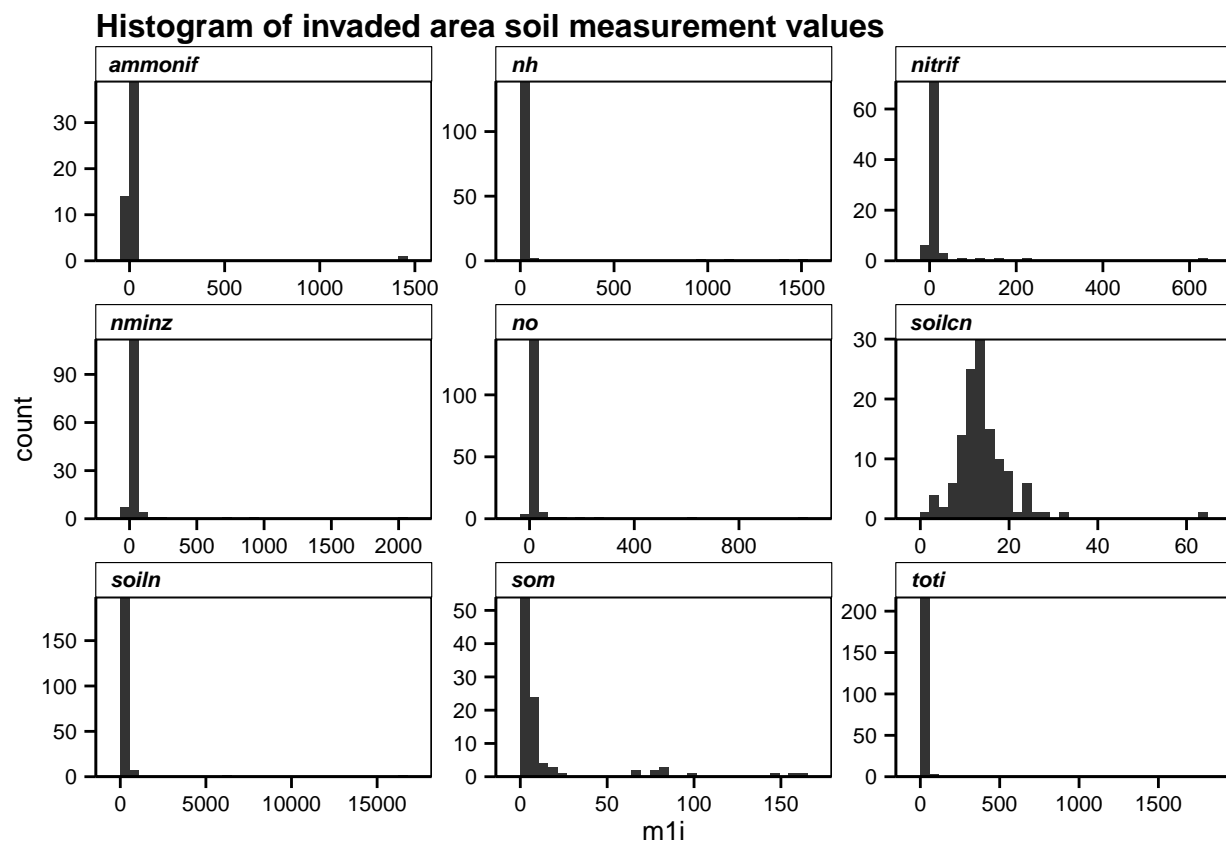
```
## [1] "Effect size values look normally-distributed-ish"
```

```
newfilename<-'pHist_ES_OR.png'
png(paste(figuresPath,newfilename, sep='/'),
    units='in', width = fig.width*3, height = fig.height*6, res=fig.res)
pHist_ES_OR
dev.off()
```

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

3C. Distribution of unit-standardized soil measurement values

```
#mli (invaded area soil measurements)
pHist_measInv<-ggplot(dat1.meas, aes(x=mli)) + facet_wrap(~measCat, scales='free', ncol=3) + geom_histogram(
  mytheme + scale_y_continuous(expand = c(0,0)) + ggtitle('Histogram of invaded area soil measurement values'))
pHist_measInv
```



```

newfilename<-'pHist_measInv.png'
png(paste(figuresPath,newfilename, sep='/'),
    units='in', width = fig.width*3, height = fig.height*6, res=fig.res)
pHist_measInv
dev.off()

```

```

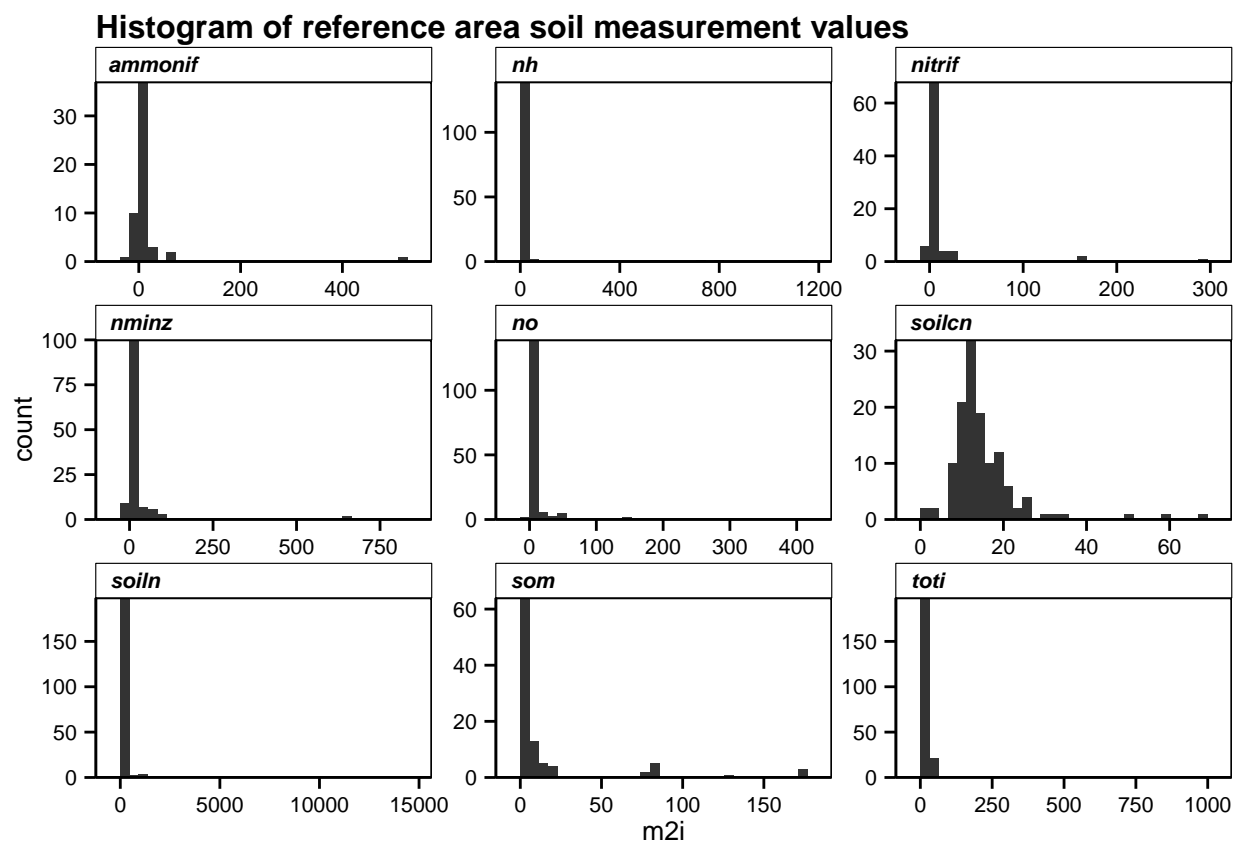
## pdf
## 2

```

```

#m2i (reference area soil measurements)
pHist_measRef<-ggplot(dat1.meas, aes(x=m2i)) + facet_wrap(~measCat, scales='free', ncol=3) + geom_histogram(
  mytheme + scale_y_continuous(expand = c(0,0)) + ggtitle('Histogram of reference area soil measurement
pHist_measRef

```



```

newfilename<-'pHist_measRef.png'
png(paste(figuresPath,newfilename, sep='/'),
    units='in', width = fig.width*3, height = fig.height*6, res=fig.res)
pHist_measRef
dev.off()

```

```

## pdf
## 2

```

```

#Log-transform some measurements
# logtMeas<-c('nh','no','toti','soilmoi','som','soiln','biom','litterbiom','littercn','litterpercN',
logtMeas<-c('nh','no','toti','som','soiln')
paste('Log-transform the following soil measurement values to improve normality:',paste(logtMeas, collapse=' '))

## [1] "Log-transform the following soil measurement values to improve normality: nh, no, toti, som, soiln, biom, litterbiom, littercn, litterpercN"

nologtMeas<-unique(dat1$measCat)[!unique(dat1$measCat) %in% logtMeas]
SD.logt<-function(meanval, sdval){ #function that corrects the variance values associated with each observation
  varval<-(sdval)^2
  sd.logt<-sqrt(log10(1+varval/(meanval)^2))
  return(sd.logt)
}
dat1$m1i_logt<-log10(dat1$m1i) #warning message about NaNs is because of negative rate values
dat1$sd1i_logt<-SD.logt(meanval=dat1$m1i, sdval=dat1$sd1i)
dat1$m2i_logt<-log10(dat1$m2i)
dat1$sd2i_logt<-SD.logt(meanval=dat1$m2i, sdval=dat1$sd2i)
#put the non-transformed data back into measures that shouldn't be transformed
dat1[dat1$measCat %in% nologtMeas, 'm1i_logt']<-dat1[dat1$measCat %in% nologtMeas, 'm1i']
dat1[dat1$measCat %in% nologtMeas, 'sd1i_logt']<-dat1[dat1$measCat %in% nologtMeas, 'sd1i']
dat1[dat1$measCat %in% nologtMeas, 'm2i_logt']<-dat1[dat1$measCat %in% nologtMeas, 'm2i']
dat1[dat1$measCat %in% nologtMeas, 'sd2i_logt']<-dat1[dat1$measCat %in% nologtMeas, 'sd2i']

#update and re-plot
dat1.meas<-ddply(dat1, ~obsID+measCat, summarize,
  m1i = unique(m1i_logt),
  m2i = unique(m2i_logt),
  yi = unique(yi))

pHist_measInv_T<-ggplot(dat1.meas, aes(x=m1i)) +
  facet_wrap(~measCat, scales='free', ncol=3) + geom_histogram() +
  mytheme + scale_y_continuous(expand = c(0,0)) + ggtitle('Histogram of InvArea unit-std soil meas. values')
newfilename<-'pHist_measInv_T.png'
png(paste(figuresPath,newfilename, sep='/'),
  units='in', width = fig.width*3, height = fig.height*6, res=fig.res)
pHist_measInv_T
dev.off()

## pdf
## 2

pHist_measRef_T<-ggplot(dat1.meas, aes(x=m2i)) +
  facet_wrap(~measCat, scales='free', ncol=3) + geom_histogram() +
  mytheme + scale_y_continuous(expand = c(0,0)) + ggtitle('Histogram of NatArea unit-std soil meas. values')
newfilename<-'pHist_measRef_T.png'
png(paste(figuresPath,newfilename, sep='/'),
  units='in', width = fig.width*3, height = fig.height*6, res=fig.res)
pHist_measRef_T
dev.off()

## pdf
## 2

```

```
#get rid of outliers
#ammonif
dat1[dat1$measCat=='ammonif' & dat1$m1i_logt > 50 & !is.na(dat1$m1i_logt),]
```

```
##      obsID paperID measCat n1i      m1i      sd1i n2i      m2i      sd2i
## 1 327.01      327 ammonif   3 1431.47 509.9504   3 517.77 105.5061
## 2 327.01      327 ammonif   3 1431.47 509.9504   3 517.77 105.5061
##      measQuality ecosystCat  studyType  InvNfix      NatNfix
## 1 NoAgg.NoConv  shrubland field study N-fixing No N-fixers
## 2 NoAgg.NoConv  shrubland field study N-fixing No N-fixers
##              Nfix invGenera    traitCat InvArea_qualRank
## 1 Invasive N-fixers only  Acacia litterpercN              1
## 2 Invasive N-fixers only  Acacia      percN              4
##      InvArea_cwm InvSpInvArea_qualRank InvSpInvArea_cwm NatArea_qualRank
## 1      1.730857              1              1.730857              0
## 2      1.682097              4              5.270581              4
##      NatArea_cwm CWMDiff_cwm      yi      vi m1i_logt sd1i_logt m2i_logt
## 1      NA              NA 1.9798 0.9933 1431.47 509.9504 517.77
## 2      1.32      0.3620972 1.9798 0.9933 1431.47 509.9504 517.77
##      sd2i_logt
## 1 105.5061
## 2 105.5061
```

```
dat1[dat1$measCat=='ammonif' & dat1$m1i_logt > 50 & !is.na(dat1$m1i_logt),c('m1i_logt','m2i_logt')]<-NA
#dat1[dat1$measCat=='ammonif' & dat1$m1i_logt > 50 & !is.na(dat1$m2i_logt),]
```

```
#nitrif
dat1[dat1$measCat=='nitrif' & dat1$m1i_logt > 50 & !is.na(dat1$m1i_logt),]
```

```
##      obsID paperID measCat n1i      m1i      sd1i n2i      m2i      sd2i
## 1   7.01      7  nitrif   6 110.133 24.94805   6 18.780 47.559293
## 2  29.01     29  nitrif   5 166.667 408.30843   5 164.815 274.879220
## 3  29.01     29  nitrif   5 166.667 408.30843   5 164.815 274.879220
## 4  29.02     29  nitrif   5 217.037 390.69407   5 164.815 274.879220
## 5  29.02     29  nitrif   5 217.037 390.69407   5 164.815 274.879220
## 6 327.01    327  nitrif   3 639.590 237.39488   3 294.420 52.753071
## 7 327.01    327  nitrif   3 639.590 237.39488   3 294.420 52.753071
## 8 704.03    704  nitrif   2  67.150 34.00052   2  26.943  9.671807
##      measQuality ecosystCat  studyType  InvNfix      NatNfix
## 1 NoAgg.NoConv  grassland field study    N-fixing No N-fixers
## 2   Agg.NoConv   wetland field study Non-N-fixing No N-fixers
## 3   Agg.NoConv   wetland field study Non-N-fixing No N-fixers
## 4   Agg.NoConv   wetland field study Non-N-fixing No N-fixers
## 5   Agg.NoConv   wetland field study Non-N-fixing No N-fixers
## 6 NoAgg.NoConv  shrubland field study    N-fixing No N-fixers
## 7 NoAgg.NoConv  shrubland field study    N-fixing No N-fixers
## 8 NoAgg.NoConv   forest field study Non-N-fixing No N-fixers
##              Nfix invGenera    traitCat InvArea_qualRank
## 1 Invasive N-fixers only Elaeagnus      percN              1
## 2      No N-fixers  Lythrum      cn              2
## 3      No N-fixers  Lythrum      percN              2
```

```

## 4      No N-fixers  Lythrum      cn      2
## 5      No N-fixers  Lythrum      percN    2
## 6 Invasive N-fixers only  Acacia litterpercN    1
## 7 Invasive N-fixers only  Acacia      percN    4
## 8      No N-fixers  Berberis      percN    2
##      InvArea_cwm InvSpInvArea_qualRank InvSpInvArea_cwm NatArea_qualRank
## 1  3.98000000      1      3.98000000      1
## 2 20.75290000      2      20.75290000      0
## 3  0.06090224      2      0.03442348      2
## 4 20.75290000      2      20.75290000      0
## 5  0.03442348      2      0.03442348      2
## 6  1.73085700      1      1.73085700      0
## 7  1.68209721      4      5.27058120      4
## 8  2.01406656      1      1.72307140      2
##      NatArea_cwm CWMDiff_cwm      yi      vi mli_logt sdli_logt m2i_logt
## 1  2.08983703  1.89016297 2.2197 0.5386 110.133  24.94805  18.780
## 2      NA      NA 0.0048 0.4000 166.667 408.30843 164.815
## 3  0.08738099 -0.02647876 0.0048 0.4000 166.667 408.30843 164.815
## 4      NA      NA 0.1396 0.4010 217.037 390.69407 164.815
## 5  0.08738099 -0.05295751 0.1396 0.4010 217.037 390.69407 164.815
## 6      NA      NA 1.6016 0.8804 639.590 237.39488 294.420
## 7  1.32000000  0.36209721 1.6016 0.8804 639.590 237.39488 294.420
## 8  1.98885721  0.02520935 0.9075 1.1030  67.150  34.00052  26.943
##      sd2i_logt
## 1  47.559293
## 2 274.879220
## 3 274.879220
## 4 274.879220
## 5 274.879220
## 6  52.753071
## 7  52.753071
## 8   9.671807

```

```

dat1[dat1$measCat=='nitrif' & dat1$mli_logt > 50 & !is.na(dat1$mli_logt),c('mli_logt','m2i_logt')]<-NA
#dat1[dat1$measCat=='nitrif' & dat1$mli_logt > 50 & !is.na(dat1$m2i_logt),]

#nminz
dat1[dat1$measCat=='nminz' & dat1$mli_logt > 100 & !is.na(dat1$mli_logt),]

```

```

##      obsID paperID measCat n1i      mli      sdli n2i      m2i      sd2i
## 1   29.01      29    nminz   5  743.3863 1391.2134   5 637.5663 1841.0016
## 2   29.01      29    nminz   5  743.3863 1391.2134   5 637.5663 1841.0016
## 3   29.02      29    nminz   5  910.0530 1202.1308   5 637.5663 1841.0016
## 4   29.02      29    nminz   5  910.0530 1202.1308   5 637.5663 1841.0016
## 5  327.01     327    nminz   3 2060.9100  553.9098   3 822.3400  131.8818
## 6  327.01     327    nminz   3 2060.9100  553.9098   3 822.3400  131.8818
## 7  592.01     592    nminz   8  210.1505  869.9009   3  85.8680  107.4344
## 8  592.01     592    nminz   8  210.1505  869.9009   3  85.8680  107.4344
## 9  592.01     592    nminz   8  210.1505  869.9009   3  85.8680  107.4344
## 10 592.02     592    nminz   5  184.0385  498.3854   3  85.8680  107.4344
## 11 592.02     592    nminz   5  184.0385  498.3854   3  85.8680  107.4344
## 12 592.02     592    nminz   5  184.0385  498.3854   3  85.8680  107.4344
##      measQuality ecosystCat  studyType      InvNfix      NatNfix
## 1    Agg.NoConv    wetland field study Non-N-fixing No N-fixers

```

## 2	Agg.NoConv	wetland	field	study	Non-N-fixing	No N-fixers
## 3	Agg.NoConv	wetland	field	study	Non-N-fixing	No N-fixers
## 4	Agg.NoConv	wetland	field	study	Non-N-fixing	No N-fixers
## 5	NoAgg.NoConv	shrubland	field	study	N-fixing	No N-fixers
## 6	NoAgg.NoConv	shrubland	field	study	N-fixing	No N-fixers
## 7	Agg.NoConv	forest	field	study	Non-N-fixing	No N-fixers
## 8	Agg.NoConv	forest	field	study	Non-N-fixing	No N-fixers
## 9	Agg.NoConv	forest	field	study	Non-N-fixing	No N-fixers
## 10	Agg.NoConv	forest	field	study	Non-N-fixing	No N-fixers
## 11	Agg.NoConv	forest	field	study	Non-N-fixing	No N-fixers
## 12	Agg.NoConv	forest	field	study	Non-N-fixing	No N-fixers
##		Nfix	invGenera		traitCat	InvArea_qualRank
## 1	No N-fixers	Lythrum		cn		2
## 2	No N-fixers	Lythrum		percN		2
## 3	No N-fixers	Lythrum		cn		2
## 4	No N-fixers	Lythrum		percN		2
## 5	Invasive N-fixers only	Acacia		litterpercN		1
## 6	Invasive N-fixers only	Acacia		percN		4
## 7	No N-fixers	Ligustrum		littercn		2
## 8	No N-fixers	Ligustrum		litterpercN		2
## 9	No N-fixers	Ligustrum		percN		2
## 10	No N-fixers	Ligustrum		littercn		2
## 11	No N-fixers	Ligustrum		litterpercN		2
## 12	No N-fixers	Ligustrum		percN		2
##	InvArea_cwm	InvSpInvArea_qualRank	InvSpInvArea_cwm	NatArea_qualRank		
## 1	20.75290000	2	20.75290000	0		
## 2	0.06090224	2	0.03442348	2		
## 3	20.75290000	2	20.75290000	0		
## 4	0.03442348	2	0.03442348	2		
## 5	1.73085700	1	1.73085700	0		
## 6	1.68209721	4	5.27058120	4		
## 7	51.35531915	2	34.60000000	2		
## 8	0.99414894	2	1.30000000	2		
## 9	1.65873742	1	1.73112500	2		
## 10	48.81875000	2	34.60000000	2		
## 11	1.01679688	2	1.30000000	2		
## 12	1.82804753	1	1.73112500	2		
##	NatArea_cwm	CWMDiff_cwm	yi	vi	m1i_logt	sd1i_logt
## 1	NA	NA	0.0585	0.4002	743.3863	1391.2134
## 2	0.08738099	-0.02647876	0.0585	0.4002	743.3863	1391.2134
## 3	NA	NA	0.1582	0.4013	910.0530	1202.1308
## 4	0.08738099	-0.05295751	0.1582	0.4013	910.0530	1202.1308
## 5	NA	NA	2.4545	1.1687	2060.9100	553.9098
## 6	1.32000000	0.36209721	2.4545	1.1687	2060.9100	553.9098
## 7	54.83548800	-3.48016885	0.1477	0.4593	210.1505	869.9009
## 8	0.81289375	0.18125519	0.1477	0.4593	210.1505	869.9009
## 9	1.99357538	-0.33483796	0.1477	0.4593	210.1505	869.9009
## 10	54.83548800	-6.01673800	0.2072	0.5360	184.0385	498.3854
## 11	0.81289375	0.20390313	0.2072	0.5360	184.0385	498.3854
## 12	1.99357538	-0.16552785	0.2072	0.5360	184.0385	498.3854
##	sd2i_logt					
## 1	1841.0016					
## 2	1841.0016					
## 3	1841.0016					

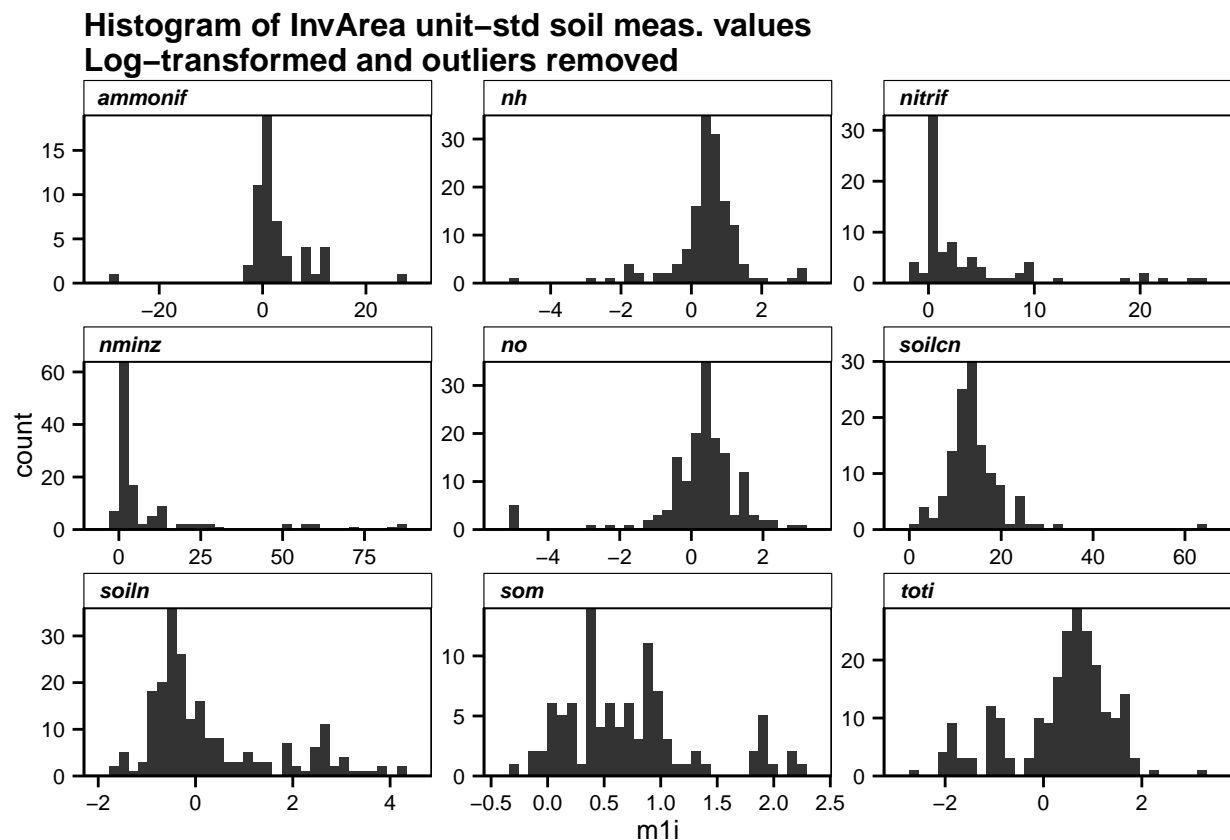

```
## 4 1841.0016
## 5 131.8818
## 6 131.8818
## 7 107.4344
## 8 107.4344
## 9 107.4344
## 10 107.4344
## 11 107.4344
## 12 107.4344
```

```
dat1[dat1$measCat=='nminz' & dat1$m1i_logt > 100 & !is.na(dat1$m1i_logt),c('m1i_logt','m2i_logt')]<-NA
#dat1[dat1$measCat=='nminz' & dat1$m1i_logt > 100 & !is.na(dat1$m2i_logt),]
```

#update and re-plot

```
dat1.meas<-ddply(dat1, ~obsID+measCat, summarize,
  m1i = unique(m1i_logt),
  m2i = unique(m2i_logt),
  yi = unique(yi))
```

```
pHist_measInv_T_OR<-ggplot(dat1.meas, aes(x=m1i)) +
  facet_wrap(~measCat, scales='free', ncol=3) + geom_histogram() +
  mytheme + scale_y_continuous(expand = c(0,0)) + ggtitle('Histogram of InvArea unit-std soil meas. values')
pHist_measInv_T_OR
```



```
newfilename<-'pHist_measInv_T_OR.png'
png(paste(figuresPath,newfilename, sep='/'),
```

```

units='in', width = fig.width*3, height = fig.height*6, res=fig.res)
pHist_measInv_T_OR
dev.off()

```

```

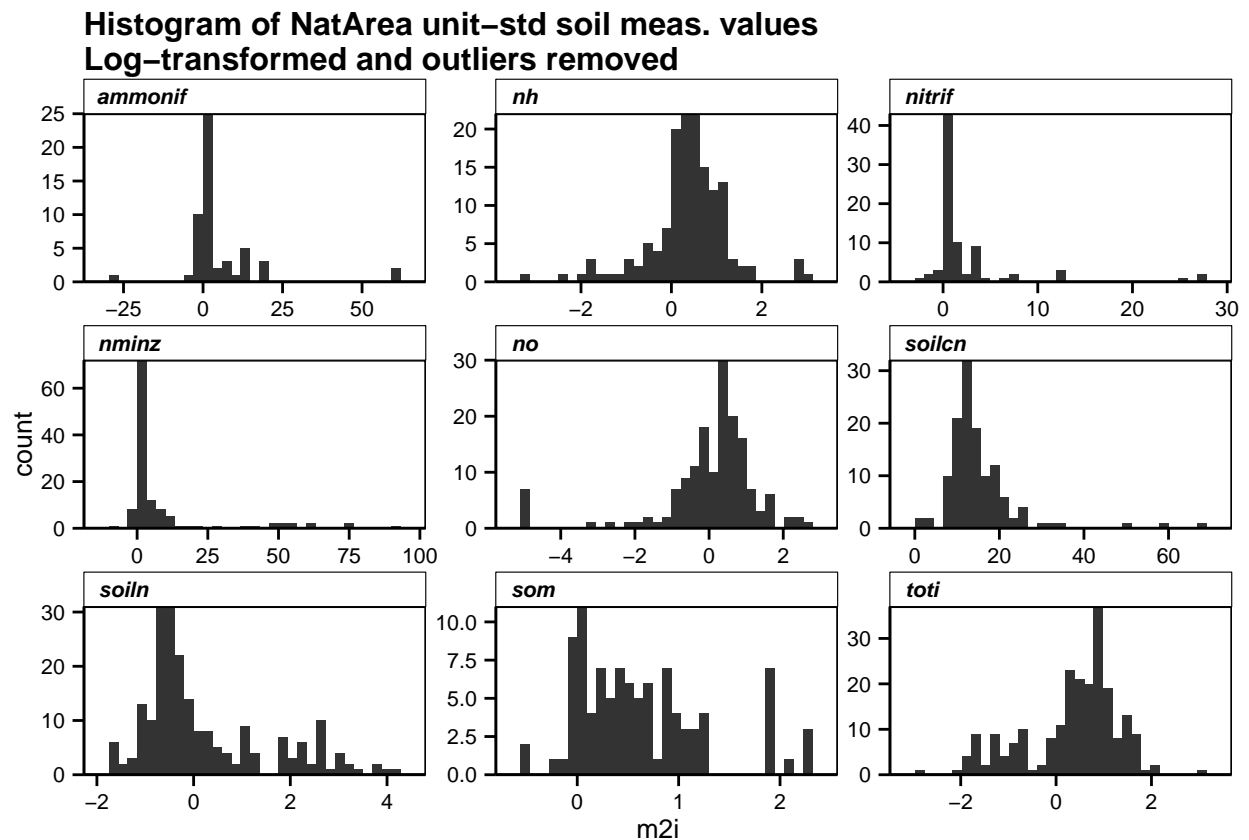
## pdf
## 2

```

```

pHist_measRef_T_OR<-ggplot(dat1.meas, aes(x=m2i)) +
  facet_wrap(~measCat, scales='free', ncol=3) + geom_histogram() +
  mytheme + scale_y_continuous(expand = c(0,0)) + ggtitle('Histogram of NatArea unit-std soil meas. val
pHist_measRef_T_OR

```



```

newfilename<-'pHist_measRef_T_OR.png'
png(paste(figuresPath,newfilename, sep='/'),
  units='in', width = fig.width*3, height = fig.height*6, res=fig.res)
pHist_measRef_T_OR
dev.off()

```

```

## pdf
## 2

```

```

paste('Removed outlier effect size values in ammonif, nitrif, nminz.')

```

```

## [1] "Removed outlier effect size values in ammonif, nitrif, nminz."

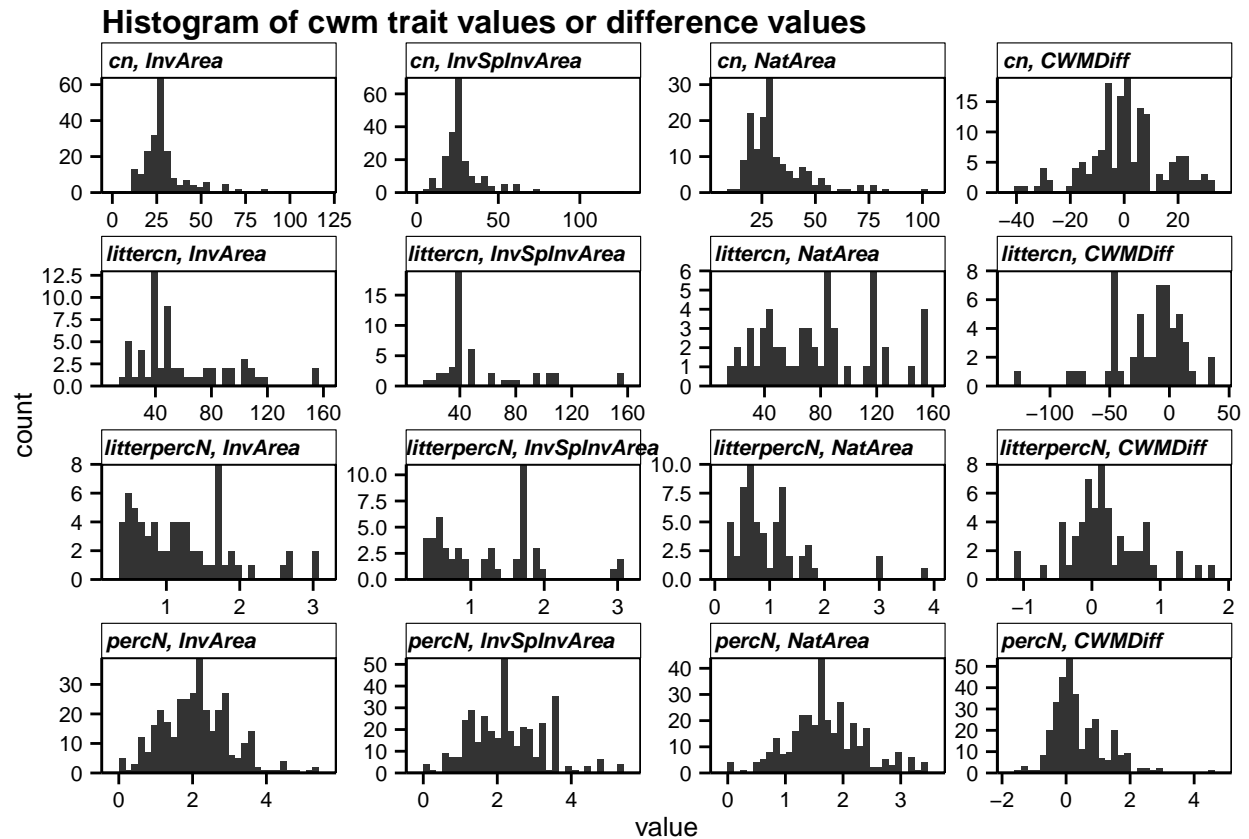
```

```
paste('Most unit-std soil meas. values look normally-distributed after log-transforming some measurements')
```

```
## [1] "Most unit-std soil meas. values look normally-distributed after log-transforming some measurements"
```

3D. Distribution of CWM trait values

```
pHist_cwm<-ggplot(m.dat1.tr1, aes(x=value)) + facet_wrap(~traitCat+variable, scales='free', ncol=4) +
  mytheme + scale_y_continuous(expand = c(0,0)) + ggtitle('Histogram of cwm trait values or difference values')
pHist_cwm
```



```
newfilename<-'pHist_cwm.png'
png(paste(figuresPath,newfilename, sep='/'),
    units='in', width = fig.width*3, height = fig.height*6, res=fig.res)
pHist_cwm
dev.off()
```

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

```
paste('No outliers removed at this point, but see MIIN_2_cwm.Rmd')
```

```
## [1] "No outliers removed at this point, but see MIIN_2_cwm.Rmd"
```

```
paste('Most cwm values look normally-distributed - except littercn, probably because there was not very
```

```
## [1] "Most cwm values look normally-distributed - except littercn, probably because there was not very
```

4. Export dataframes

```
newfilename<-'papers.txt'
write.table(papers.c, file=paste(synthdataPath,newfilename, sep='/'), sep='\t', quote=TRUE, qmethod="do

newfilename<-'observations.txt'
write.table(observations.c, file=paste(synthdataPath,newfilename, sep='/'), sep='\t')

newfilename<-'cover.txt'
write.table(cover.c, file=paste(synthdataPath,newfilename, sep='/'), sep='\t')

newfilename<-'species.txt'
write.table(species.c, file=paste(synthdataPath,newfilename, sep='/'), sep='\t')

newfilename<-'traits.txt'
write.table(traits.c, file=paste(synthdataPath,newfilename, sep='/'), sep='\t')

newfilename<-'measures.txt'
write.table(measures.c, file=paste(synthdataPath,newfilename, sep='/'), sep='\t')

newfilename<-'cwm.txt'
write.table(cwm.c, file=paste(synthdataPath,newfilename, sep='/'), sep='\t')

newfilename<-'spIDcover.txt'
write.table(spIDcover.c, file=paste(synthdataPath,newfilename, sep='/'), sep='\t')

newfilename<-'spIDtraits.txt'
write.table(spIDtraits.c, file=paste(synthdataPath,newfilename, sep='/'), sep='\t')

colnames(dat1)
```

```
## [1] "obsID"          "paperID"
## [3] "measCat"        "nli"
## [5] "mli"           "sdli"
## [7] "n2i"           "m2i"
## [9] "sd2i"          "measQuality"
## [11] "ecosystCat"     "studyType"
## [13] "InvNfix"        "NatNfix"
## [15] "Nfix"          "invGenera"
## [17] "traitCat"       "InvArea_qualRank"
## [19] "InvArea_cwm"    "InvSpInvArea_qualRank"
## [21] "InvSpInvArea_cwm" "NatArea_qualRank"
## [23] "NatArea_cwm"    "CWMDiff_cwm"
## [25] "yi"            "vi"
```

```
## [27] "m1i_logt"          "sd1i_logt"  
## [29] "m2i_logt"          "sd2i_logt"
```

```
newfilename<-'metaDataset.txt'  
write.table(dat1, file=paste(synthdataPath,newfilename, sep='/'), sep='\t')
```