

Import Fric et al. results from supplemental tables for comparisons

Elise Larsen & Vaughn Shirey

Updated 12/7/2020; initiated 12/3/2020

✱

Here we import Fric et al. results from supplemental tables.

We start with ~latitude and ~latitude|altitude+year model results

```
##Needed packages:
```

```
library(readxl)
```

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2      v purrr   0.3.4
```

```
## v tibble  3.0.3      v dplyr   1.0.1
```

```
## v tidyr   1.1.1      v stringr 1.4.0
```

```
## v readr   1.3.1      v forcats 0.5.0
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()    masks stats::lag()
```

```
##Import Fric results:
```

```
#Check order of phenometric results in tables
```

```
#Single regression
```

```
fric.metrics1<-names(read_excel("fric_supplements/ele13419-sup-0003-tables2.xlsx", sheet="~latitude", range="A3:Y3"))
```

```
## New names:
## * `` -> ...3
## * `` -> ...4
## * `` -> ...5
## * `` -> ...6
## * `` -> ...7
## * ...
```

```
fric.metrics1[grep("e",fric.metrics1)]
```

```
## [1] "Species"      "Peak"          "Onset"          "Termination"
```

#Corrected regression (Regression of residuals)

```
fric.metrics2<-names(read_excel("fric_supplements/ele13419-sup-0003-tables2.xlsx", sheet=~latitude|altitude+year", range="A2:Y2"))
```

```
## New names:
## * `` -> ...3
## * `` -> ...4
## * `` -> ...5
## * `` -> ...6
## * `` -> ...7
## * ...
```

```
fric.metrics2[grep("e",fric.metrics2)]
```

```
## [1] "Species"      "Peak"          "Onset"          "Termination"
```

Table is 1 species column and 3 sets of results columns (peak, onset, termination). There are 8 results parameters reported per model.

```
n.metrics<-8
fric.results.lat<-na.omit(read_excel("fric_supplements/ele13419-sup-0003-tables2.xlsx", sheet=~latitude", range="A4:Y113"))
```

```
## New names:
## * `` -> ...1
## * ResDFMean -> ResDFMean...2
## * ResDevMean -> ResDevMean...3
## * DFMean -> DFMean...4
## * DevMean -> DevMean...5
## * ...
```

```
#nrow(fric.results.lat)
fric.results.corr<-read_excel("fric_supplements/ele13419-sup-0003-tables2.xlsx", sheet=~latitude|altitude+year", range="A3:Y108")
```

```
## New names:
## * `` -> ...1
## * ResDFMean -> ResDFMean...2
## * ResDevMean -> ResDevMean...3
## * DFMean -> DFMean...4
## * DevMean -> DevMean...5
## * ...
```

```
fric.results.lat$model<-"lat"
fric.results.corr$model<-"corr"
fric.results<-rbind(fric.results.lat, fric.results.corr)

field.names<-c(names(read_excel("fric_supplements/ele13419-sup-0003-tables2.xlsx", sheet=~latitude", range="B4:I4"))))
field.names<-c("species",paste("peak",field.names,sep="."),paste("onset",field.names,sep="."),paste("term",field.names,sep="."),"model")

names(fric.results)<-field.names
```

Add ~year and ~year|latitude responses for supplemental table 2

```
#Check order of phenometric results in tables
#Single regression
yr.metrics1<-names(read_excel("fric_supplements/ele13419-sup-0003-tables2.xlsx", sheet=~year", range="A2:Y2"))
```

```
## New names:
## * `` -> ...3
## * `` -> ...4
## * `` -> ...5
## * `` -> ...6
## * `` -> ...7
## * ...
```

```
yr.metrics1[grep("e",yr.metrics1)]
```

```
## [1] "Species"      "Peak"         "Onset"        "Termination"
```

```
#Corrected regression (Regression of residuals)
yr.metrics2<-names(read_excel("fric_supplements/ele13419-sup-0003-tables2.xlsx", sheet=~year|latitude", range="A2:Y2"))
```

```
## New names:
## * `` -> ...3
## * `` -> ...4
## * `` -> ...5
## * `` -> ...6
## * `` -> ...7
## * ...
```

```
yr.metrics2[grep("e",yr.metrics2)]
```

```
## [1] "Species"      "Peak"         "Onset"        "Termination"
```

```
# Table is 1 species column and 3 sets of results columns (peak, onset, termination). There are 8 results parameters reported per model.
n.metrics<-8
fric.yr<-na.omit(read_excel("fric_supplements/ele13419-sup-0003-tables2.xlsx", sheet=~year", range="A3:Y108"))
```

```
## New names:
## * `` -> ...1
## * ResDFMean -> ResDFMean...2
## * ResDevMean -> ResDevMean...3
## * DFMean -> DFMean...4
## * DevMean -> DevMean...5
## * ...
```

```
#nrow(fric.results.lat)
fric.yr.lat<-read_excel("fric_supplements/ele13419-sup-0003-tables2.xlsx", sheet=~year|latitude", range="A3:Y108")
```

```
## New names:
## * `` -> ...1
## * ResDFMean -> ResDFMean...2
## * ResDevMean -> ResDevMean...3
## * DFMean -> DFMean...4
## * DevMean -> DevMean...5
## * ...
```

```
fric.yr$model<-"year"
fric.yr.lat$model<-"yearcorr"

field.names<-c(names(read_excel("fric_supplements/ele13419-sup-0003-tables2.xlsx", sheet=~year", range="B3:I3"))))
field.names<-c("species",paste("peak",field.names,sep="."),paste("onset",field.names,sep="."),paste("term",field.names,sep="."),"model")

names(fric.yr.lat)<-names(fric.yr)<-field.names

#Combine all results into one table
fric.results<-rbind(fric.results,fric.yr,fric.yr.lat)
rm(fric.results.lat, fric.results.corr, fric.yr, fric.yr.lat)
rm(field.names, fric.metrics1, fric.metrics2, n.metrics, yr.metrics1, yr.metrics2)
#Keep Fields we want for comparison
fric.results<-fric.results %>%
  select(species,peak.ResDFMean,onset.ResDFMean,onset.p_mean,onset.coef,onset.response,term.p_mean,term.ResDFMean,term.coef,
term.response, model)
```

```
#standardize names
fric.results.species<-strsplit(fric.results$species," ")
result.names<-NULL
for(i in 1:length(fric.results.species)) {
  result.names<-c(result.names,paste(fric.results.species[[i]][1],fric.results.species[[i]][2],sep=" "))
}
fric.results$name<-result.names
#verify that names match
table(order(fric.results$name)-order(fric.results$species))
```

```
##
## 0
## 420
```

#The sort order is the same. Looks good.

*

Because Fric et al. results are not tagged by region, we add a field for region. We match results to region based on relative sample sizes.

```
#Get species-region & sample sizes from fric data file
#Load occurrence data (see data curation files for how this file is created)

#Loads fricdata
load('data/occurrences_FricAnalysis.RData')
#summarize species-region-sample size
datasummary<-fricdata %>% group_by(name,region) %>% tally()

fric.results$region<-datasummary$region[match(fric.results$name,datasummary$name)]

#But now we need to correct the ones that are in both regions

#ID Species names that occur in both regions
fric.dups<-fric.results %>% group_by(name) %>% tally() %>% filter(n==8)

#Get data records with duplicates
dup.sp1<-datasummary[duplicated(datasummary$name) | duplicated(datasummary$name, fromLast=T),]

#For each species, check which region has more occurrence data
biggersample<-NULL
for(sp in sort(unique(dup.sp1$name))) {
  biggersample<-c(biggersample,ifelse(dup.sp1$n[dup.sp1$name==sp & dup.sp1$region=="Europe"]-dup.sp1$n[dup.sp1$name==sp & dup.sp1$region=="N. America"]>0,"Europe","N. America"))
}
table(biggersample)

## biggersample
## Europe
##      5
```

```

#All Europe sample sizes are bigger so we can use that to ID regions in the Fric results.
for(sp in sort(unique(dup.sp1$name))) {
  sizes<-fric.results$peak.ResDFMean[which(fric.results$name==sp)]
  fric.results$region[fric.results$name==sp & fric.results$peak.ResDFMean==max(sizes)]<- "Europe"
  fric.results$region[fric.results$name==sp & fric.results$peak.ResDFMean==min(sizes)]<- "N. America"
}

fric.results$set<-paste(fric.results$name,fric.results$region,sep="-")

summary(fric.results)

```

```

##      species      peak.ResDFMean onset.ResDFMean onset.p_mean
## Length:420      Min.   :   13      Min.   : 3.00      Min.   :0.0000
## Class :character 1st Qu.:   80      1st Qu.:13.00      1st Qu.:0.1050
## Mode  :character Median :  190      Median :18.00      Median :0.3901
##                      Mean  : 2458      Mean  :18.58      Mean  :0.4013
##                      3rd Qu.:1065      3rd Qu.:25.00      3rd Qu.:0.6788
##                      Max.   :51817     Max.   :37.00      Max.   :0.9996
## onset.coef      onset.response      term.p_mean      term.ResDFMean
## Min.   :-2.62324      Min.   :-1.00000      Min.   :0.0000006      Min.   : 3.00
## 1st Qu.: -0.17607      1st Qu.: 0.00000      1st Qu.:0.1300094      1st Qu.:12.00
## Median : 0.06153      Median : 0.00000      Median :0.3829437      Median :18.00
## Mean   : 0.23994      Mean   : 0.09524      Mean   :0.4050860      Mean   :18.29
## 3rd Qu.: 0.38807      3rd Qu.: 0.00000      3rd Qu.:0.6600517      3rd Qu.:24.00
## Max.    : 4.33128      Max.    : 1.00000      Max.    :0.9972891      Max.    :34.00
## term.coef      term.response      model      name
## Min.   :-2.90317      Min.   :-1.00000      Length:420      Length:420
## 1st Qu.: -0.18540      1st Qu.: 0.00000      Class :character  Class :character
## Median : 0.02208      Median : 0.00000      Mode  :character  Mode  :character
## Mean   : 0.08746      Mean   : 0.04048
## 3rd Qu.: 0.39264      3rd Qu.: 0.00000
## Max.    : 4.10620      Max.    : 1.00000
## region      set
## Length:420      Length:420
## Class :character  Class :character
## Mode  :character  Mode  :character
##
##
##

```



```
print("Onset response: single regression")
```

```
## [1] "Onset response: single regression"
```

```
table(filter(fric.results,model=="lat")$onset.response)
```

```
##  
## -1  0  1  
## 12 67 26
```

```
print("Termination response: single regression")
```

```
## [1] "Termination response: single regression"
```

```
table(filter(fric.results,model=="lat")$term.response)
```

```
##  
## -1  0  1  
## 11 71 23
```

```
print("Onset response: residual regression")
```

```
## [1] "Onset response: residual regression"
```

```
table(filter(fric.results,model=="corr")$onset.response)
```

```
##  
## -1  0  1  
##  2 82 21
```

```
print("Termination response: residual regression")
```

```
## [1] "Termination response: residual regression"
```

```
table(filter(fric.results,model=="corr")$term.response)
```

```
##  
## -1  0  1  
##  4 89 12
```

```
#Save to data file  
save(fric.results,file="data/fric_results.RData")
```

The # responses in each group are correct.

End File.