

```

#Ensemble Creation
library(tidync)
library(data.table)
library(tidyverse)
library(lubridate)
library(zyp)
library(sf)
library(raster)
library(terra)
library(RColorBrewer)
library(rasterVis)
library(xts)

###Create Groundwater Storage Anomaly
proj_dir = "~/Dropbox/WB/GRACE_Ensemble/"

filePath = fread(paste0(proj_dir, 'FileSummary.csv'))

#Select the GLDAS Solution (Serves to subset the GRACE datasets)
sws_id = 22
validGridPoints =
  fread(filePath[Type=='GLDAS' & ID == sws_id]$FilePath) %>%
  .[order(lon,lat)] %>%
  mutate(ID2 = paste0(lon, lat))

#Select the GRACE Solutions - JPL
grace_id = 11
tw_anomaly1 =
  fread(filePath[Type=='GRACE' & ID == grace_id]$FilePath) %>%
  mutate(ID2 = paste0(lon, lat)) %>%
  dplyr::filter(ID2 %in% validGridPoints$ID2) %>%
  .[order(lon,lat)] %>%
  melt(id.vars = c("lon", "lat", "ID2"),
       measure.vars = 4:(ncol(.)-1),
       variable.name = "ym", value.name = "tw")

grace_id = 12
tw_anomaly2 =
  fread(filePath[Type=='GRACE' & ID == grace_id]$FilePath) %>%
  mutate(ID2 = paste0(lon, lat)) %>%
  dplyr::filter(ID2 %in% validGridPoints$ID2) %>%
  .[order(lon,lat)] %>%
  melt(id.vars = c("lon", "lat", "ID2"),
       measure.vars = 4:(ncol(.)-1),
       variable.name = "ym", value.name = "tw")

grace_id = 13
tw_anomaly3 =
  fread(filePath[Type=='GRACE' & ID == grace_id]$FilePath) %>%

```

```

mutate(ID2 = paste0(lon, lat)) %>%
dplyr::filter(ID2 %in% validGridPoints$ID2) %>%
.[order(lon,lat)] %>%
melt(id.vars = c("lon", "lat", "ID2"),
      measure.vars = 4:(ncol(.)-1),
      variable.name = "ym", value.name = "tw")

tw_anomaly =
  merge(tw_anomaly1, tw_anomaly2[,.(ID2, ym, tw2 = tw)], all=TRUE,
by = c('ID2', 'ym')) %>%
  merge(tw_anomaly3[,.(ID2, ym, tw3 = tw)], all=TRUE, by = c('ID2',
'ym')) %>%
  dplyr::mutate(tw_ensemble = (tw+tw2+tw3)/3)

tw_anomaly_wide =
  tw_anomaly[,.(lon, lat, cell_id = ID2, ym, tw_ensemble)] %>%
  spread(ym, tw_ensemble)

fwrite(tw_anomaly_wide,
        paste0(proj_dir, "Outputs/Ensembles/
GRACE-TWS_Ensemble_1degree_220828.csv"))

#Load the sws anomalies - 1
sws_id = 21
sws_anomaly1 =
  fread(filePath[Type=='GLDAS' & ID == sws_id]$FilePath) %>%
  .[order(lon,lat)] %>%
  mutate(ID2 = paste0(lon, lat)) %>%
  melt(id.vars = c("lon", "lat", "ID2"),
        measure.vars = 3:(ncol(.)-2),
        variable.name = "ym", value.name = "sws")

#Load the sws anomalies - 2
sws_id = 22
sws_anomaly2 =
  fread(filePath[Type=='GLDAS' & ID == sws_id]$FilePath) %>%
  .[order(lon,lat)] %>%
  mutate(ID2 = paste0(lon, lat)) %>%
  melt(id.vars = c("lon", "lat", "ID2"),
        measure.vars = 3:(ncol(.)-2),
        variable.name = "ym", value.name = "sws")

#Load the sws anomalies - 3
sws_id = 23
sws_anomaly3 =
  fread(filePath[Type=='GLDAS' & ID == sws_id]$FilePath) %>%
  .[order(lon,lat)] %>%

```

```

mutate(ID2 = paste0(lon, lat)) %>%
melt(id.vars = c("lon", "lat", "ID2"),
      measure.vars = 3:(ncol(.)-2),
      variable.name = "ym", value.name = "sws")

sws_anomaly =
  merge(sws_anomaly1, sws_anomaly2[,.(ID2, ym, sws2 = sws)], all=TRUE,
        by = c('ID2', 'ym')) %>%
  merge(sws_anomaly3[,.(ID2, ym, sws3 = sws)], all=TRUE, by = c('ID2',
    'ym')) %>%
  dplyr::mutate(sws_ensemble = (sws+sws2+sws3)/3)

sws_anomaly_wide =
  sws_anomaly[,.(lon, lat, ym, sws_ensemble, cell_id = ID2)] %>%
  spread(ym, sws_ensemble) %>%
  dplyr::select(lon, lat, 4:199, cell_id) #To ensure same format as
other GLDAS tables

fwrite(sws_anomaly_wide,
       paste0(proj_dir, "Outputs/Ensembles/
GRACE_SWS_Ensemble_1degree_220828.csv"))

#####
#Create the gws anomaly here
gws_anomaly =
  tws_anomaly[,.(ID2,lon,lat,ym,tws_ensemble)] %>%
  merge(sws_anomaly[,.(ID2,ym,sws_ensemble)], by = c('ID2', 'ym')) %>%
  mutate(gws_ensemble = tws_ensemble - sws_ensemble)

gws_anomaly_wide =
  gws_anomaly[,.(lon,lat, cell_id = ID2, ym,gws_ensemble)] %>%
  spread(ym, gws_ensemble)

```

```

#Create final GRACE dataset

###Load all the GWS datasets
proj_dir = "~/Dropbox/WB/GRACE_Ensemble/"

gws.path = list.files(paste0(proj_dir, "Outputs/GWS/"), '.csv',
full.names = T)

gws_all = lapply(gws.path, fread)

#Get the names for each of them
filePath = fread(paste0(proj_dir, 'FileSummary.csv'))
#Check filepaths in the csv and folder align before adding names
gws.path == filePath[ID %in% c(101:110)]$FilePath

names(gws_all) = filePath[ID %in% c(101:110)]$Name

#Convert each dataset into long format/add new cell_id column

for(i in 1:length(gws_all)){
  cur_df =
    gws_all[[i]] %>%
    melt(id.vars = c("lon", "lat"),
        measure.vars = 4:(ncol(.)),
        variable.name = "ym", value.name = names(gws_all[i])) %>%
    mutate(ID = paste0(lon, lat))

  if(i == 1) {
    gws_out = cur_df
  } else{
    gws_out = merge(gws_out,
                    cur_df[,c('ID', 'ym', names(gws_all[i]))],
with=FALSE], by = c('ID', 'ym') ,all=T)
  }
}

#####
#####
#Use a example gws_wide file to merge spatially with World Bank
regions
gws_spatial =
  gws_all[[1]] %>%
  mutate(ID = paste0(lon, lat)) %>%
  dplyr::select(lon, lat, ID) %>%
  st_as_sf(coords = c("lon", "lat"),
           crs = "+proj=longlat +datum=WGS84 +no_defs")

#Add the World Bank regions

```

```

####Load World Regions
wb_regions =
  st_read(paste0(proj_dir,"Spatial Files/WB_Regions/
WB_countries_Admin0_10m.shp")) %>%
  dplyr::select(WB_NAME, ISO_A2, ISO_A3, ISO_N3, TYPE) %>%
  filter(TYPE != 'Dependency') %>%
  st_make_valid()

wb_regions_ns =
  wb_regions %>% as.data.table() %>% dplyr::select(-geometry) %>%
  distinct()

#Merge country data with GRACE
gws_out_country =
  gws_spatial %>%
  st_make_valid() %>%
  st_join(wb_regions) %>%
  st_drop_geometry()

#Merge back with the gws_all long data frame
gws.final =
  gws_out %>%
  merge(gws_out_country, by = 'ID', all = T)

#####
###
# fwrite(gws.final,
#         paste0(proj_dir, "Outputs/Ensembles/
GRACE_GWS_Ensemble_1degree_220828.csv"))
#

#####Tests

gws_in = fread(paste0(proj_dir, "Outputs/Ensembles/
GRACE_GWS_Ensemble_1degree_220828.csv"))

gws_all[[4]]

View(gws.final[ID == '-178.566.5'])

```