####################################################################

## L4b preparation codes (Sruthi & Vero)

####################################################################

Codes are found in the following folder (Sruthi’s original): /gpfs/data1/duncansongp/sruthikp/GEDI\_PA/final\_codes/l4b\_prep/

STEP1: Run "**merge\_country\_level\_RDS.R**" to get the PA and control files merged at the country level. Resulting files are stored here: /gpfs/data1/duncansongp/GEDI\_global\_PA/Country\_level\_RDS/

STEP2: Run "**get\_country\_biome\_level\_merged\_pas\_and\_Ctrls.R**" to get shapefiles merged at the country-biome level (eg. BRA\_Mangroves\_PA.shp refers to all the protected areas within the mangrove biome in Brazil for which matches could be found and BRA\_Mangroves\_Ctrl.shp refers to all the matched control cells for the PAs. Besides, if there are some PAs within mangrove biomes without any matches, that would be stored separately as BRA\_Mangroves\_unmatched\_PA.shp).

This script calls within itself the following function: "getBiomeLevelPAsPerCountryFunction.R".

Resulting shapefiles are stored here: /gpfs/data1/duncansongp/GEDI\_global\_PA/WDPA\_AGB\_Input\_Polygons/Country\_Biome\_PAs\_and\_Ctrls/

STEP3: Run "**dissolve\_pa\_polygon.R**" to merge/aggregate the neighboring 1 sq.km cells within PAs and control .shp files into a single cell of larger area. This is necessary for running the l4b biomass estimation codes of John. The dissolved polygons are stored here: /gpfs/data1/duncansongp/GEDI\_global\_PA/WDPA\_AGB\_Input\_Polygons/Country\_Biome\_PAs\_and\_Ctrls/Dissolved/

\*\*\*Vero update: The original dissolved polygons are now stored here: /gpfs/data1/duncansongp/GEDI\_global\_PA/WDPA\_AGB\_Input\_Polygons/Country\_Biome\_PAs\_and\_Ctrls/Dissolved\_veryold/

\*\*\*Vero update: run "**Dissolved\_replace\_filenames.R**" script to modify filenames, replacing any special characters or spaces with underscore. The resulting shapefiles are stored here: /gpfs/data1/duncansongp/GEDI\_global\_PA/WDPA\_AGB\_Input\_Polygons/Country\_Biome\_PAs\_and\_Ctrls/Dissolved\_old/

######### Codes listed below are found in: /gpfs/data1/duncansongp/leitoldv/l4b\_prep\_V2/

STEP4: run "**get\_COUNTRYbiome\_level\_shapefiles\_V2.R**" to (1) correct unmatched\_PA shapefiles with buffer = 0 around them and to (2) create COUNTRY x Biome for USA (from USA east, west, pcfc). Input shapefiles are called in from /gpfs/data1/duncansongp/GEDI\_global\_PA/WDPA\_AGB\_Input\_Polygons/Country\_Biome\_PAs\_and\_Ctrls/Dissolved\_new/ and output shapefiles are stored in /gpfs/data1/duncansongp/GEDI\_global\_PA/WDPA\_AGB\_Input\_Polygons/Country\_Biome\_PAs\_and\_Ctrls/Dissolved\_new\_V2

STEP5: run "**get\_COUNTRY\_level\_shapefiles\_V2\_with\_readShapePoly.R**" to aggregate PA, Ctrl, and unmatched shapefiles to the country level. Input shapefiles are in /gpfs/data1/duncansongp/GEDI\_global\_PA/WDPA\_AGB\_Input\_Polygons/Country\_Biome\_PAs\_and\_Ctrls/Dissolved\_new\_V2 and output aggregated shapedfiles are in /gpfs/data1/duncansongp/GEDI\_global\_PA/WDPA\_AGB\_Input\_Polygons/Country\_PAs\_and\_Ctrls\_V2

STEP6: run "**get\_CONTINENTbiome\_level\_shapefiles\_V2\_with\_readShapePoly.R**" to aggregate PA, Ctrl, and unmatched shapefiles to CONTINENT x BIOME level from Country x Biome files V2. Input shapefiles are in /gpfs/data1/duncansongp/GEDI\_global\_PA/WDPA\_AGB\_Input\_Polygons/Country\_Biome\_PAs\_and\_Ctrls/Dissolved\_new\_V2 and the aggregated output shapefiles are in /gpfs/data1/duncansongp/GEDI\_global\_PA/WDPA\_AGB\_Input\_Polygons/Continent\_Biome\_PAs\_and\_Ctrls\_V2

STEP7: run "**get\_CONTINENT\_level\_shapefiles\_V2\_with\_readShapePoly.R**" to generate CONTINENT level shapefiles from Continent x Biome files V2

STEP8: run "**get\_CONTINENT\_level\_shapefiles\_V2\_with\_readShapePoly\_fromCountryBiome.R**" to generate CONTINENT level shapefiles from Country x Biome files V2 (for comparison with the above method)

####################################################################

####################################################################

####################################################################

####################################################################

####################################################################

## L4B processing for PA vs CTRL areas using "bigregion" approach

####################################################################

#Step0a. IF NEEDED, set up issgedi\_umd conda environment following instructions here:

### https://bitbucket.org/umdgedi/iss\_gedi/wiki/Getting%20setup

#Step0b. IF NEEDED, clone gedi\_l4a from https://github.com/armstonj/gedi\_l4a

cd /gpfs/data1/duncansongp/leitoldv/git/ ##this is my local git directory

git clone https://github.com/armstonj/gedi\_l4a.git

[username]

[token]

## OR copy over most up-to-date code on the cluster:

cp -R /gpfs/data1/vclgp/software/umdgedi/gedi\_l4a/umd\_cluster /gpfs/data1/duncansongp/leitoldv

cp -R /gpfs/data1/duncansongp/leitoldv/umd\_cluster /gpfs/data1/duncansongp/leitoldv/gedi\_l4b

####################################################################

## L4B processing for PA vs CTRL areas using "bigregion" approach

####################################################################

#Step1. TO RUN TEST CODE using issgedi\_l4b\_bigregion.py

conda activate issgedi\_umd

module unload gdal

module load umdgedi

#Step2a. EITHER run from my own directory (where some necessary input files are located)

cd /gpfs/data1/duncansongp/leitoldv/gedi\_l4b ##this is where some input files are located

issgedi\_l4b\_bigregion.py -f Shapefile -E 6933 -T l4b\_pa -v -R /gpfs/data1/duncansongp/gedi\_data/l4b\_tiles/ -P 1 -G 4 --estimator\_data issgedi\_l4a\_r001\_estimators.json --estimator\_selection issgedi\_l4a\_r001\_estimator\_selection.json -i /gpfs/data1/duncansongp/GEDI\_global\_PA/WDPA\_AGB\_Input\_Polygons/Country\_Biome\_PAs\_and\_Ctrls/Dissolved/ALB/ALB\_Mediterranean\_Forests\_Woodlands\_Scrub\_PA.shp

issgedi\_l4b\_bigregion.py -f Shapefile -E 6933 -T l4b\_pa -v -R /gpfs/data1/duncansongp/gedi\_data/l4b\_tiles/ -P 1 -G 4 --estimator\_data issgedi\_l4a\_r001\_estimators.json --estimator\_selection issgedi\_l4a\_r001\_estimator\_selection.json -i /gpfs/data1/duncansongp/GEDI\_global\_PA/WDPA\_AGB\_Input\_Polygons/Country\_Biome\_PAs\_and\_Ctrls/Dissolved/ALB/ALB\_Temperate\_Broadleaf\_Mixed\_Forests\_PA.shp

#Step2b. OR run from any directory (complete path to the necessary input files included here)

issgedi\_l4b\_bigregion.py -f Shapefile -E 6933 -T l4b\_pa -v -R /gpfs/data1/duncansongp/gedi\_data/l4b\_tiles/ -P 1 -G 4 --estimator\_data /gpfs/data1/duncansongp/leitoldv/gedi\_l4b/issgedi\_l4a\_r001\_estimators.json --estimator\_selection /gpfs/data1/duncansongp/leitoldv/gedi\_l4b/issgedi\_l4a\_r001\_estimator\_selection.json -i /gpfs/data1/duncansongp/GEDI\_global\_PA/WDPA\_AGB\_Input\_Polygons/Country\_Biome\_PAs\_and\_Ctrls/Dissolved/ALB/ALB\_Temperate\_Broadleaf\_Mixed\_Forests\_CTRL.shp

##-----------------------------------------------------------------------------------

## PARALLEL PROCESSING multiple files within a selected country

#python /gpfs/data1/duncansongp/leitoldv/gedi\_l4b/run\_issgedi\_l4b\_regionTEST.py ALB

#python /gpfs/data1/duncansongp/leitoldv/gedi\_l4b/run\_issgedi\_l4b\_bigregionTEST.py ALB

#nohup python /path/to/test.py > output.log &

####################################################################

## L4B processing code to run at different levels of aggregation

####################################################################

All scripts are in “/gpfs/data1/duncansongp/leitoldv/gedi\_l4b/”

##-----------------------------------------------------------------------------------

COUNTRYbiome V2: “run\_issgedi\_l4b\_bigregionCOUNTRYbiome\_V2.py”

filepath = "/gpfs/data1/duncansongp/GEDI\_global\_PA/WDPA\_AGB\_Input\_Polygons/Country\_Biome\_PAs\_and\_Ctrls/Dissolved\_new\_V2/" + iso3 + "/\*.shp"

Example code to run from terminal:

conda activate issgedi\_umd

module unload gdal

module load umdgedi

cd /gpfs/data1/duncansongp/leitoldv/gedi\_l4b

python /gpfs/data1/duncansongp/leitoldv/gedi\_l4b/run\_issgedi\_l4b\_bigregionCOUNTRYbiome\_V2.py AUS

###############################

## AUS & BRA re-processing in smaller chunks

nohup python /gpfs/data1/duncansongp/leitoldv/gedi\_l4b/run\_issgedi\_l4b\_bigregionCOUNTRYbiome\_V2.py BRA2 &

##98621 @gsapp7

nohup python /gpfs/data1/duncansongp/leitoldv/gedi\_l4b/run\_issgedi\_l4b\_bigregionCOUNTRYbiome\_V2.py AUS2 &

##161723 @gsapp7

###############################

##-----------------------------------------------------------------------------------

COUNTRY V2: ”run\_issgedi\_l4b\_bigregionCOUNTRY\_V2.py”

filepath = "/gpfs/data1/duncansongp/GEDI\_global\_PA/WDPA\_AGB\_Input\_Polygons/Country\_PAs\_and\_Ctrls\_V2/" + iso3 + "/\*.shp"

Example code to run from terminal:

conda activate issgedi\_umd

module unload gdal

module load umdgedi

cd /gpfs/data1/duncansongp/leitoldv/gedi\_l4b

python /gpfs/data1/duncansongp/leitoldv/gedi\_l4b/run\_issgedi\_l4b\_bigregionCOUNTRY\_V2.py AUS

##-----------------------------------------------------------------------------------

CONTINENTbiome V2: “run\_issgedi\_l4b\_bigregionCONTINENTbiome\_V2.py”

filepath = "/gpfs/data1/duncansongp/GEDI\_global\_PA/WDPA\_AGB\_Input\_Polygons/Continent\_Biome\_PAs\_and\_Ctrls\_V2/" + continent + "/\*.shp"

Example code to run from terminal:

conda activate issgedi\_umd

module unload gdal

module load umdgedi

cd /gpfs/data1/duncansongp/leitoldv/gedi\_l4b

python /gpfs/data1/duncansongp/leitoldv/gedi\_l4b/run\_issgedi\_l4b\_bigregionCONTINENTbiome\_V2.py Af

##-----------------------------------------------------------------------------------old version

CONTINENT: “run\_issgedi\_l4b\_bigregionCONTINENT.py”

filepath = "/gpfs/data1/duncansongp/GEDI\_global\_PA/WDPA\_AGB\_Input\_Polygons/Continent\_PAs\_and\_Ctrls/" + continent + "/\*.shp"

####################################################################

## To export/access the OUTPUTS from the L4B processing

####################################################################

conda activate issgedi\_umd

module unload gdal

cd /gpfs/data1/duncansongp/leitoldv ##navigate to the directory where you want to save the results

ogr2ogr -f "GPKG" gedi\_l4b\_unequal\_protected-areas\_001\_04.gpkg PG:"host='gsapp11.clust.gshpc.umd.edu' user=sruthikp dbname=gedicalval password=abc123" "gedi\_roi.l4b\_pa" -where "rid\_type = 'user'"

##-------------------------------------------------------------------

## open an R session

ssh leitoldv@gsapp8.umd.edu

module load R

R

R

library(sp)

library(sf)

library(rgdal)

library(tidyverse)

data <- st\_read("/gpfs/data1/duncansongp/leitoldv/gedi\_l4b\_unequal\_protected-areas\_001\_04.gpkg", layer="gedi\_roi.l4b\_pa")

data.df <- data.frame(data)

nogeom.df <- st\_set\_geometry(data, NULL)

reprocessed <- c(grep(pattern="\_PA",nogeom.df[,"rid"]), grep(pattern="\_Ctrl",nogeom.df[,"rid"]))

output.df <- nogeom.df[reprocessed,]

output.df <- output.df[order(output.df$rid),]

nrow(output.df)

tests <- c(grep(pattern="AND",output.df[,"rid"]),grep(pattern="BGR",output.df[,"rid"]),

grep(pattern="GEO",output.df[,"rid"]),grep(pattern="ISR",output.df[,"rid"]),

grep(pattern="ZMB",output.df[,"rid"]))

tests.df <- output.df[c(tests),]

tests.df <- tests.df[grep(pattern="unmatched",tests.df[,"rid"]),]

#write.csv(tests.df, "/gpfs/data1/duncansongp/leitoldv/gedi\_l4b\_AGB\_results\_2022FEB11\_Country\_V2\_test\_for\_area\_mismatches.csv")

remove.rows <- c(grep(pattern="test",output.df[,"rid"]),

grep(pattern="\_PA2",output.df[,"rid"]),

grep(pattern="\_Ctrl2",output.df[,"rid"]),

grep(pattern="\_sample",output.df[,"rid"]),

grep(pattern="\_v2",output.df[,"rid"]),

which(output.df$rid=="BGR\_Temperate\_Conifer\_Forests\_unmatched\_PA"))

length(remove.rows) ##43

output.df <- output.df[-c(remove.rows),]

nrow(output.df) ##4563 (MAY/25)

all\_PAs\_dissolved <- output.df[grep(pattern="all\_PAs",output.df[,"rid"]),]

nrow(all\_PAs\_dissolved) ##169

analyzed\_PAs\_dissolved <- output.df[grep(pattern="analyzed\_PAs",output.df[,"rid"]),]

nrow(analyzed\_PAs\_dissolved) ##116

##COUNTRY

Country.df <- output.df[FALSE,]

iso3 <- unique(substr(output.df$rid, 1, 3))

iso3 <- grep("^[A-Z ]+$", iso3, value = TRUE)

for(i in 1:length(iso3)){

PA <- output.df[grep(pattern=paste(iso3[i],"\_PA",sep=""), output.df[,"rid"]),]

Ctrl <- output.df[grep(pattern=paste(iso3[i],"\_Ctrl",sep=""), output.df[,"rid"]),]

UM <- output.df[grep(pattern=paste(iso3[i],"\_unmatched\_PA",sep=""), output.df[,"rid"]),]

Country.df <- rbind(Country.df, PA, Ctrl, UM)

}

USA <- rbind(output.df[grep(pattern="USA\_east\_PA",output.df[,"rid"]),],

output.df[grep(pattern="USA\_east\_Ctrl",output.df[,"rid"]),],

output.df[grep(pattern="USA\_east\_unmatched\_PA",output.df[,"rid"]),],

output.df[grep(pattern="USA\_pcfc\_PA",output.df[,"rid"]),],

output.df[grep(pattern="USA\_pcfc\_Ctrl",output.df[,"rid"]),],

output.df[grep(pattern="USA\_pcfc\_unmatched\_PA",output.df[,"rid"]),],

output.df[grep(pattern="USA\_west\_PA",output.df[,"rid"]),],

output.df[grep(pattern="USA\_west\_Ctrl",output.df[,"rid"]),],

output.df[grep(pattern="USA\_west\_unmatched\_PA",output.df[,"rid"]),])

Country.df <- rbind(Country.df, USA)

Country.df <- Country.df[order(Country.df$rid),]

Country.df$rid

nrow(Country.df) ##956 (MAY/25)

##CONTINENT

continent\_names <- c("Africa", "Asia", "Antarctica", "Australia", "Europe", "NAmerica", "SAmerica")

continent\_ids <- c("Af", "As", "At", "Au", "Eu", "NAm", "SAm")

Continent.df <- output.df[FALSE,]

for(i in 1:length(continent\_ids)){

continent <- rbind(output.df[grep(pattern=paste(continent\_names[i]), output.df[,"rid"]),],

output.df[grep(pattern=paste(continent\_ids[i],"\_PA",sep=""), output.df[,"rid"]),],

output.df[grep(pattern=paste(continent\_ids[i],"\_Ctrl",sep=""), output.df[,"rid"]),],

output.df[grep(pattern=paste(continent\_ids[i],"\_unmatched\_PA",sep=""), output.df[,"rid"]),])

Continent.df <- rbind(Continent.df, continent)

}

Continent.df <- Continent.df[-grep(pattern="all\_PAs\_", Continent.df[,"rid"]),]

Continent.df <- Continent.df[-grep(pattern="analyzed\_PAs\_", Continent.df[,"rid"]),]

Continent.df <- Continent.df[order(Continent.df$rid),]

Continent.df$rid

nrow(Continent.df) ##22

##CONTINENTxBIOME

ContinentBiome.df <- output.df[FALSE,]

for(i in 1:length(continent\_ids)){

continent <- output.df[grep(pattern=continent\_ids[i], output.df[,"rid"]),]

ContinentBiome.df <- rbind(ContinentBiome.df, continent)

}

ContinentBiome.df <- ContinentBiome.df[-grep(pattern="all\_PAs\_", ContinentBiome.df[,"rid"]),]

ContinentBiome.df <- ContinentBiome.df[-grep(pattern="analyzed\_PAs\_", ContinentBiome.df[,"rid"]),]

ContinentBiome.df <- generics::setdiff(ContinentBiome.df, Continent.df)

ContinentBiome.df <- ContinentBiome.df[order(ContinentBiome.df$rid),]

ContinentBiome.df$rid

nrow(ContinentBiome.df) ##364 (MAY/25)

##COUNTRYxBIOME

CountryBiome.df <- generics::setdiff(output.df, rbind(Country.df, ContinentBiome.df, Continent.df, all\_PAs\_dissolved, analyzed\_PAs\_dissolved))

CountryBiome.df$rid

nrow(CountryBiome.df) ##2936 (MAY/25)

write.csv(CountryBiome.df, "/gpfs/data1/duncansongp/leitoldv/gedi\_l4b\_AGB\_results\_2022MAY25\_CountryBiome\_V2.csv")

write.csv(Country.df, "/gpfs/data1/duncansongp/leitoldv/gedi\_l4b\_AGB\_results\_2022MAY25\_Country\_V2.csv")

write.csv(ContinentBiome.df, "/gpfs/data1/duncansongp/leitoldv/gedi\_l4b\_AGB\_results\_2022MAY25\_ContinentBiome\_V2.csv")

##saves the results as a csv table into the directory & filename specified above