

# GlobBiomass Validation Technical Documentation

Martin Herold, Sytze de Bruin, Arnan Araza

January 2019

## 0.1 Introduction

Here we develop an automated processing chain of biomass plot and map validation. Called as “PlotToMap”, the tool is applicable for both local and global plot data using GlobBiomass as the default validation “map”. It starts by temporally adjusting plot data, if needed, using the most recent forest growth data from IPCC (2019). Plot biomass is scaled into actual forest area to address forest bias more common to small plots. Spatial aggregation of plots into larger scales is also possible to minimize random errors from forest fragmentation at map support vis a vis small plot sizes.

This documentation should guide other technical users for their own biomass validation work. Testing this prototype can also be helpful to us, the authors, in any potential improvement of our work.

## 0.2 Preliminaries

### Main folder and structure

The main folder structure is shown in Figure 1 containing 3 sub folders (data, results, and scripts) and the main script “main.R”. The data folder has the two tabular data (plot data and growth data) and 2 raster dataset (biomass map and tree cover). There can be chances that latter will not be inside “data” folder because of file size. It’s fine.

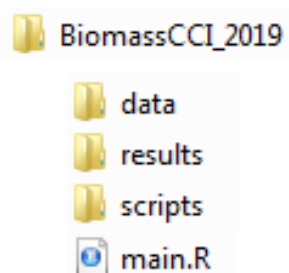


Figure 1: Main folder structure.

## Installation

Clone or extract github repository at any desired machine directory:  
<https://github.com/arnanaraza/BiomassCCI>

## Workflow

Below in Figure 2 is the schema of the validation steps. It highlights main steps which is temporal adjustment, forest scaling, and aggregation. It also emphasizes scale at “plot support” or “support unit” level.

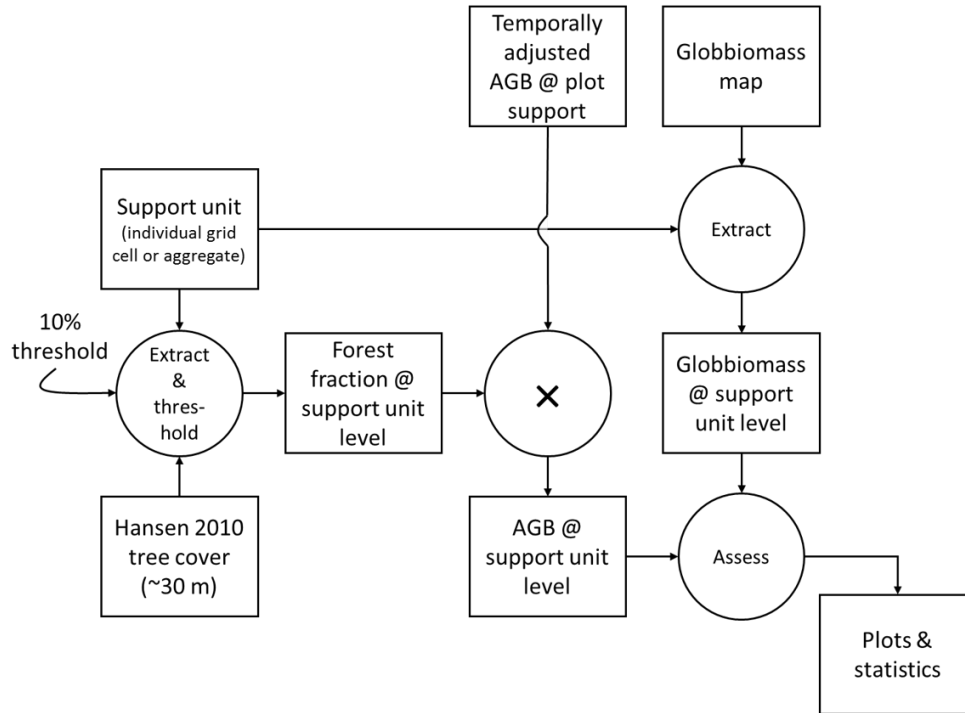


Figure 2: Schema of the validation processing chain.

## Plot data format

The plot data requires 8 columns or entities. Table 1 particularly column 1 shows the column names of a sample plot data as well as the required columns (highlighted).

Table 1: Sample and required columns and naming of plot data.

Entity/Column	Data type	Description	Example
<b>ZONE</b>	Character	Continent name	Asia
<b>CODE</b>	Character	Continent-country code	AFR_FOS
<b>WUR_ID</b>	Character	Unique ID 1	AFR5_1295
<b>PLOT_ID</b>	Character	Unique ID 2	CAP_09_0_0
<b>POINT_X</b>	Numeric	Longitude	-3.007917056
<b>POINT_Y</b>	Numeric	Latitude	6.01019559
<b>AGB_T_HA</b>	Numeric	Biomass value	106.66
<b>YEAR</b>	Character	Year range	2007_2008
<b>AVG_YEAR</b>	Numeric	Average year	2007.5
<b>SIZE_HA</b>	Numeric	Plot size (ha)	0.01
<b>SIZE_RANGE</b>	Character	Size range (ha)	0.01-0.05
<b>FAO.ecozone</b>	Character	Specific eco-zones/biomes	Tropical rainforest
<b>GEZ</b>	Character	General eco-zones/biomes	tropical

## Main raster inputs

Tiles of GlobBiomass and Hansen Tree Cover can be accessed from these respective links:

- [http://globbiomass.org/wp-content/uploads/GB\\_Maps/Globbiomass\\_global\\_dataset.html](http://globbiomass.org/wp-content/uploads/GB_Maps/Globbiomass_global_dataset.html)
- [https://earthenginepartners.appspot.com/science-2013-global-forest/download\\_v1.5.html](https://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.5.html)

## 0.3 main.R

Main script loads libraries, defines global variables, and calls 7 functions to perform the processing chain of biomass validation.

It uses 7 packages:

- 2 spatial (rgdal and raster)
- 2 data frame operations (plyr and dplyr)
- 3 for parallel processing (foreach, parallel, doParallel)
- 1 for graphing (plotrix)

## Global variables

It uses 7 packages: There are 9 global variables as follows:

- *mainDir*. Main directory
- *scriptsDir*. Directory of scripts
- *outDir*. Directory of all results
- *dataDir*. Directory of data
- *plotsFile*. Filename of plot data
- *agbTilesDir*. Directory of AGB tiles
- *treeCoverDir*. Directory of Hansen tree cover tiles
- *SRS*. Default coordinates
- *forestTHs*. Forest threshold, can be a number or vector

## 0.4 Subscripts

### TempFixed.R

Function:

- *TempFixed*. Normalizes plot biomass values to year 2010 (GlobBiomass Map year) by applying growth data to non-2010 plots from IPCC Guidelines 2019 on actual biomass growth/age relationships. Growth effect on pre-2010 plots is applied and otherwise on post-2010 plots.

Variables:

- *df*. Data frame. Default plot data.
- *domain*. Character. Continent of interest.
- *year*. Numeric. Map AGB year i.e. 2010 for GlobBiomass.

## TempEffect.R

Functions:

- *HistoShift*. Creates histogram of pre and post temporal adjustment as well as its overlap. Returns histogram and saves it at *outDir*.
- *ChangeTable*. Creates a table summarizing the temporal adjustment effect per biomass range. Returns and saves the table at *outDir*.

Variables:

- *old*. Data frame. Default plot data
- *new*. Data frame. Temporally adjusted plot data

## MakeBlockPolygon.R

Function:

- *MakeBlockPolygon*. Makes a polygon out of original plot location which aligns to AGB map pixel or larger cell over which AGB is aggregated. Returns a Spatial Polygon.

Variables:

- *x*. Numeric. Longitude of plot data
- *y*. Numeric. Latitude of plot data
- *size*. Numeric. Target aggregation scale in degrees

## TileNames.R

Function:

- *TCtileNames*. Takes up to four tree cover (TC) tile names covered by polygon. Returns 4 TC file names.
- *AGBtileNames*. Takes up to four AGB tile names covered by polygon. Returns 4 AGB file names.

Variables:

- *pol*. Spatial Polygon. The created polygon from desired aggregated scale.

## BlockMeans.R

Function:

- *sampleTreeCover*. Generates block mean mapped forest over polygon.

Variables:

- *pol*. Spatial Polygon. The created polygon from desired aggregated scale.
- *thresholds*. Numeric. Forest threshold, defaults at 10%.
- *wghts*. Boolean. Enables block mean of forest cover of polygon.

Function:

- *sampleAGBmap*. Generates block mean mapped AGB over polygon.

Variables:

- *pol*. Spatial Polygon. The created polygon from desired aggregated scale.
- *wghts*. Boolean. Enables block mean of map AGB of polygon.

## InvDasymetry.R

Function:

- *InvDasymetry*. Applies forest scaling and spatial aggregation. Returns a data frame with mean AGB from plots satisfying selection criteria of 5 plots per block polygon. Maximizes computer cores via parallel processing.

Variables:

- *clmn*. Character. Choice between “biome” or “continents”.
- *value*. Character. Continent name or biome name i.e. “boreal”
- *aggr*. Boolean. Whether to aggregate or not.
- *minPlots*. Numeric. Minimum number of plots to create a block polygon. Default is 5.
- *wghts*. Boolean. Enables block mean of forest cover of polygon.

## 0.5 Operation

### 0.5.1 Potential errors

#### Errors from not attaining 5 plots minimum

- *'names' attribute [5] must be the same length as the vector [1]*
- *NA values in coordinates*

#### Errors from limited computer memory

- *Cannot allocate a vector size of  $x$  gb*

### 0.5.2 Run time

For 100,000 plot data and a machine that has 8 gig RAM with 8 cores:

- Temporal fix = 1 minute
- Forest scaling, with aggregation = 1 hour
- Forest scaling, no aggregation = 3 hours