

```
# Forest Biomass in Bangladesh: An historical review of forest inventories to assist national estimation
```

```
# Gael Sola, FAO, gael.sola@fao.org,
# Liam Costello, liam.costello@fao.org
# 04/2016
```

```
# Main script for ready and concatenation of existing forest inventories.
# Part 01b: tree corrections (dbh, scale factor, h, species and wd)
# Contains:
# 0. Read table
# 1. DBH correction
# 2. Scale factor correction
# 3. H correction
# 4. Species and wood density
# 5. DBH classes
# 6. Basal area
# 7. Write tables
```

```
#####
# Erase memory
rm(list=ls())
```

```
# remove graphs
dev.off()
```

```
# Set working directory
# setwd("//uom-file.unimelb.edu.au/220/Users/lcostello1/Documents/BFI/Harmo/Analysis/data")
setwd("/media/gs/OSDisk/Users/solag/Documents/Work/Bangladesh article harmo/Analysis/data")
```

```
#install.packages("ggrepel")
library(tidyverse)
library(ggthemes)
library(ggrepel)
# Set theme for printing
source("R-corrections/theme_gs_print.R")
```

```
#-----#
#           0. read data tables
#-----#
```

```
table_all <- read.csv("results/table_all_01a_plot_corr.csv", stringsAsFactors = F)
nb_plot_tree <- read.csv("results/nb_plot_tree01a.csv")
```

```
#-----#
#           1. DBH correction
#-----#
```

```
t_save1 <- table_all
table_all <- t_save1
```

```
# First check
summary(table_all$dbh_cm)
```

```
# dbh distribution
ggplot(table_all) +
  geom_boxplot(aes(x=entity_ab, y=dbh_cm, colour = entity_ab)) +
  coord_flip() +
  theme(legend.position = "none")
```

```
#####
##### REMOVING ALL TREES WITH DBH > 700
#####
table_all <- table_all[which(table_all$dbh_cm < 700),]
```

```
# dbh distribution
ggplot(table_all) +
  geom_boxplot(aes(x=entity_ab, y=dbh_cm, colour = entity_ab)) +
```

```

coord_flip() +
theme(legend.position = "none")

#####
##### two options: 1. all trees below 10 cm are removed -> consistency between inventories
##### 2. adapting to forest inventories:
##### - FRMP-97: dbh>10cm, 2 levels (10-15 et 15+)
##### - FD-10-SundRF: dbh>2.5cm, 2 levels (2.5-10 et 10+)
##### - FD-07-NFA: dbh>10cm, 1 level (small trees not counted)
##### - FD-01-SAL: dbh>10cm, 1 level
##### - CREL-14: dbh>5cm, 3 levels (5-20, 20-50, 50+)
##### - CREL-10: dbh>10cm, 1 level
#####

# Set the option 1: TRUE or option 2: FALSE, option 3: ALL
remove_dbh_less_than_10 <- TRUE

# Update graph title
if (remove_dbh_less_than_10 == T) {
  plot_title <- "_sup10"
} else if (remove_dbh_less_than_10 == F) {
  plot_title <- "_alltrees"
} else if (remove_dbh_less_than_10 == "ALL") {
  plot_title <- "_forhdmmodel"
} else {
  stop("Option not correctly defined")
}

#####
##### Correct DBH values
#####

if (remove_dbh_less_than_10 == TRUE){

  table_all <- table_all[table_all$dbh_cm >= 10,]

} else if (remove_dbh_less_than_10 == FALSE){

  table_all[table_all$entity_ab == "FD-10-SundRF" & table_all$dbh_cm < 2.5, "dbh_cm"] <- 0
  table_all[table_all$entity_ab == "FD-10-SundRF" & table_all$species_local == "LIANA", "dbh_cm"] <- 0
  table_all[table_all$entity_ab == "CREL-14" & table_all$dbh_cm < 5, "dbh_cm"] <- 0
  table_all[table_all$entity_ab != "CREL-14" & table_all$entity_ab != "FD-10-SundRF" & table_all$dbh_cm < 0,]
  table_all <- table_all[table_all$dbh_cm > 0,]

} else {

  table_all[table_all$entity_ab == "FD-10-SundRF" & table_all$species_local == "LIANA", "dbh_cm"] <- 0
  table_all <- table_all[table_all$dbh_cm >0,]

}

summary(table_all$dbh_cm)

# dbh distribution
ggplot(table_all[table_all$dbh_cm<40,]) +
  geom_boxplot(aes(x=entity_ab, y=dbh_cm, colour = entity_ab)) +
  coord_flip() +
  theme(legend.position = "none") +
  ylim(0,40)

#-----#
#-----# update nb trees and plots
nb_tree <- table_all[,c("id","entity_ab")]
nb_plot <- unique(nb_tree)

# Update the table nb tree/plots
nb_plot_tree <- cbind(nb_plot_tree,
  DBH_cleaning = paste(c(table(nb_plot$entity_ab),

```

```

        dim(nb_plot)[1]),
        " (",
        c(table(nb_tree$entity_ab),
          dim(nb_tree)[1]),
        ")",
        sep = "")
    )
nb_plot_tree
#-----#
#-----#

#-----#
#                               2. Scale factor correction
#-----#

t_save2 <- table_all
table_all <- t_save2

if (remove_dbh_less_than_10 != "ALL"){

# Correcting scale factors for
# - FRMP-97: trees > 15cm in 11m radius and tree > 10 cm in 5m radius (5 plots per cluster)
table_all[table_all$entity_ab == "FRMP-97" , "plot_size_ha"] <- (5*11^2*pi
table_all[table_all$entity_ab == "FRMP-97" & table_all$dbh_cm < 15, "plot_size_ha"] <- (5* 5^2*pi

# - FD-10-SundRF: if dbh < 10 considered small trees plot size needs to be corrected
# trees > 10 cm in 10m radius and trees < 10 cm in 3m radius (5plots per clusters)
vec <- with(table_all, plot_size_ha[entity_ab == "FD-10-SundRF"])
summary(vec)
10^2*pi*5/10000
3^2*pi*5/10000 # -> not found so small trees need to be corrected
# only if small trees considered

if (remove_dbh_less_than_10 == FALSE){

    table_all[table_all$entity_ab == "FD-10-SundRF" & table_all$dbh_cm <10, "plot_size_ha"] <-
        table_all[table_all$entity_ab == "FD-10-SundRF" & table_all$dbh_cm <10, "plot_size_ha"]*3^2/10

}

# - FD-07-NFA: dbh>10cm, 1 level (small trees not counted)
# Plot size in m2 instead of ha
vec <- table_all[table_all$entity_ab == "FD-07-NFA", "plot_size_ha"]
summary(vec)
if (max(vec) > 0.5){

    table_all <- within(table_all,
        plot_size_ha[entity_ab == "FD-07-NFA"] <-
        plot_size_ha[entity_ab == "FD-07-NFA"]/10000
    )
}

vec <- table_all[table_all$entity_ab == "FD-07-NFA", "plot_size_ha"]
summary(vec)

# - FD-01-SAL: dbh>10cm, 1 level but 2 plot sizes based on location
vec <- table_all[table_all$entity_ab == "FD-01-SAL", "plot_size_ha"]
summary(vec)
table(vec) # -> all ok

# - CREL-14: dbh>5cm, 3 levels (5-20, 20-50, 50+)
# the three levels are kept in both options
vec <- table_all[table_all$entity_ab == "CREL-14", "plot_size_ha"]
summary(vec)
3*17.84^2*pi/10000 # -> needs correction
table_all$plot_size_ha <- ifelse(table_all$entity_ab != "CREL-14",
    table_all$plot_size_ha,
    ifelse(table_all$dbh_cm > 50,

```

```

        (3*17.84^2*pi)/10000,
        ifelse(table_all$dbh_cm > 20,
              (3*10^2*pi)/10000,
              (3*4^2*pi)/10000
        )
    )
)

vec <- table_all[table_all$entity_ab == "CREL-14", "plot_size_ha"]
table(vec)

# - CREL-10: dbh>10cm, 1 level
summary(table_all[table_all$entity_ab == "CREL-10", "dbh_cm"])
unique(table_all[table_all$entity_ab == "CREL-10", "plot_size_ha"])
5*10^2*pi/10000 # -> all ok

# Round plot size
table_all$plot_size_ha <- round(table_all$plot_size_ha, 6)
summary(table_all$plot_size_ha)
unique(table_all[-which(table_all$entity_ab %in% c("FD-07-NFA", "FD-10-SundRF")), c("entity_ab", "p
table(table_all[-which(table_all$entity_ab %in% c("FD-07-NFA", "FD-10-SundRF")), c("entity_ab", "p

# Update scale factor
table_all$scale_factor <- round(1/table_all$plot_size_ha, 6)

}

#-----#
#                               3. Correction H
#-----#

t_save3 <- table_all
table_all <- t_save3

# first check
summary(table_all$h_m)

# !!! REPLACING h greater than 60 m or smaller than 3m with NA
table_all[which(table_all$h_m > 60 | table_all$h_m < 3), ]$h_m <- NA

# Control (use tidyverse to bring FRMP in the back of the graph)
t_h <- table_all[which(!is.na(table_all$h)), ]
t_h %>%
  arrange(desc(entity_ab)) %>%
  ggplot() +
  geom_point(aes(x=dbh_cm, y=h_m, colour = entity_ab, pch = entity_ab), cex=0.8) +
  facet_wrap(~bfi_zone) +
  theme(legend.position = c(0.8, 0.2))

# !!! REPLACING with NA unrealistic height based on expert judgement
table_all[which(table_all$bfi_zone == "Villages" & table_all$h_m >= 35 & table_all$dbh_cm < 60), ]$h_m
table_all[which(table_all$bfi_zone == "Sundarbans" & table_all$h_m >= 35 & table_all$dbh_cm <= 35), ]$h_m

# Control (use tidyverse to bring FRMP in the back of the graph)
t_h <- table_all[which(!is.na(table_all$h)), ]
t_h %>%
  arrange(desc(entity_ab)) %>%
  ggplot() +
  geom_point(aes(x=dbh_cm, y=h_m, pch = entity_ab, color = entity_ab), cex = 1) +
  facet_wrap(~bfi_zone, ncol = 3) +
  labs(x = "Diameter at breast height (cm)", y = "tree total height (m)",
       color = "Project code", pch = "Project code") +
  theme(legend.position = c(0.8, 0.2))

# save plot
ggsave(paste("results/gr_H-D per BFI zone", plot_title, ".png", sep=""), width = 15, height = 12, unit

```

```

#-----#
#-----# for table data correction: nb trees with height measurement
nb_tree <- table_all[!is.na(table_all$h_m),]
nb_plot <- unique(nb_tree[,c("id","entity_ab", "fao_biome", "bfi_zone", "lccs6")])

# Update the table nb tree/plots
nb_plot_tree <- cbind(nb_plot_tree,
                      h_meas = paste(c(table(nb_plot$entity_ab),
                                         dim(nb_plot)[1]),
                                     " (",
                                     c(table(nb_tree$entity_ab),
                                         dim(nb_tree)[1]),
                                     ") ",
                                     sep = ""))
)
nb_plot_tree
#-----#
#-----#

#-----#
#           4. Correcting species and adding wood density
#-----#

t_save4 <- table_all
table_all <- t_save4

if (remove_dbh_less_than_10 != "ALL"){

  # Run the code
  source("R-corrections/corr05_species_and_wd_10nov.R")

  # Check numbers of corrections
  dim(table_all[which(is.na(table_all$scientific_name)),])
  dim(table_all[which(is.na(table_all$scientific_name_corr)),])
  length(unique(table_all$scientific_name))
  length(unique(table_all$scientific_name_corr))

  #####
  ##### Remove Poaceae and liana if any
  #####
  table_all <- table_all[-which(table_all$family_corr == "Poaceae"),]
  table_all <- table_all[-which(table_all$genus == "Bambusa"),]
  table_all <- table_all[-which(table_all$species_local == "Liana"),]

  # Control
  summary(table_all$wd)
  ggplot(table_all, aes(x=entity_ab, y=wd)) +
    geom_hline(aes(yintercept = 0.578), col = 'grey50', cex = 1.2) +
    geom_hline(aes(yintercept = 0.731), col = 'grey75', cex = 0.8) +
    geom_boxplot() +
    labs(x = NULL, y = "Wood density (g/cm3)") +
    facet_grid(~bfi_zone, scales = "free_x", space = "free_x") + #activate only if class == mix
    theme(axis.text.x = element_text(angle = 90),
          axis.text.y = element_text(angle = 90),
          axis.title.x = element_text(angle = 180))

  # save the plot
  ggsave(paste("results/gr_boxplot wd per project and BFI zone",plot_title,".png", sep=""), width =

  # Check main species per zone: check species with more than 5% of the trees
  vec <- unique(table_all$bfi_zone)
  main_sp_all <- NA
  for (i in 1:length(vec)){
    test <- table_all[table_all$bfi_zone == vec[i],]
    main_sp <- sort(table(test$scientific_name_corr), decreasing = T)
    main_sp <- data.frame(main_sp)
  }
}

```

```

names(main_sp) <- c("sp_name", "nb_tree")
main_sp$perc <- round(main_sp$nb_tree/sum(main_sp$nb_tree)*100,0)
main_sp$bfi_zone <- vec[i]
main_sp <- main_sp[main_sp$perc > 5,]
main_sp_all <- rbind(main_sp_all, main_sp)

}

main_sp_all <- main_sp_all[-1,]
main_sp_all <- merge(main_sp_all,
                    unique(table_all[,c("scientific_name_corr", "wd")]),
                    by.x = "sp_name", by.y = "scientific_name_corr", all.x =T)

# Check main species per zone: Add nb of trees with no species
no_sp <- table_all[is.na(table_all$family_corr), c("bfi_zone","wd")]
no_sp$count <-1
no_sp <- aggregate(count~bfi_zone, data = no_sp, FUN = sum)
no_sp <- data.frame(sp_name = "no species info",
                   nb_tree = no_sp$count,
                   perc = NA,
                   bfi_zone = no_sp$bfi_zone)

no_sp$wd <- ifelse(no_sp$bfi_zone == "Sundarbans", 0.731, 0.58)

main_sp_all <- rbind(main_sp_all,no_sp)

ggplot(main_sp_all,aes(x = bfi_zone, y = wd)) +
  geom_point(aes(size = nb_tree)) +
  geom_text_repel(aes(label = sp_name), box.padding = 0.35, point.padding = 0.5) +
  coord_flip() +
  labs(x = "BFI zone", y = "Wood density (in g/cm3)", size = "Number of \ntrees per ha")

# save plot
ggsave(paste("results/gr_wd for main species per zone",plot_title,".png", sep=""), width = 15, height = 10)

#-----#
#-----# for table data correction: nb trees after removing species
nb_tree <- table_all
nb_plot <- unique(nb_tree[,c("id","entity_ab", "fao_biome", "bfi_zone", "lccs6")])

# Update the table nb tree/plots
nb_plot_tree <- cbind(nb_plot_tree,
                    species_rm = paste(c(table(nb_plot$entity_ab),
                                          dim(nb_plot)[1]),
                                       " (",
                                       c(table(nb_tree$entity_ab),
                                          dim(nb_tree)[1]),
                                       ")",
                                       sep = ""))
)
nb_plot_tree

#-----# for table data correction: nb trees with species
nb_tree <- table_all[!is.na(table_all$scientific_name_corr),]
nb_plot <- unique(nb_tree[,c("id","entity_ab", "fao_biome", "bfi_zone", "lccs6")])

# Update the table nb tree/plots
nb_plot_tree <- cbind(nb_plot_tree,
                    with_species = paste(c(table(nb_plot$entity_ab),
                                          dim(nb_plot)[1]),
                                       " (",
                                       c(table(nb_tree$entity_ab),
                                          dim(nb_tree)[1]),
                                       ")",
                                       sep = ""))
)
nb_plot_tree
#-----#
#-----#

```

```

}

#-----#
#           5. Create dbh classes and check per entity
#-----#

t_save5  <- table_all
table_all <- t_save5

if (remove_dbh_less_than_10 != "ALL"){

  # Create diameter class
  table_all$dbh_class2 <- round((table_all$dbh_cm - 4.99)/10,0)*10
  table_all$dbh_class2 <- ifelse(table_all$dbh_class2 > 50, 50, table_all$dbh_class2)

  if (remove_dbh_less_than_10 == TRUE) {

    table_all$dbh_class2 <- factor(table_all$dbh_class2, labels = c("10-20", "20-30", "30-40", "40-50"))

  } else {

    table_all$dbh_class2 <- factor(table_all$dbh_class2, labels = c("0-10", "10-20", "20-30", "30-40"))

  }

  summary(table_all$dbh_class2)

  # Calculate the number of stem per dbh class and per ha for each bfi_zone.
  test <- aggregate(scale_factor~dbh_class2+id+bfi_zone, data=table_all, FUN=sum)
  test2 <- aggregate(scale_factor~dbh_class2+bfi_zone, data = test, FUN = mean)
  ggplot(test2[test2$dbh_class2 != "0-10",]) +
    geom_col(aes(x = dbh_class2, y = scale_factor)) +
    facet_wrap(~bfi_zone, ncol = 3) +
    labs(x = "DBH Class", y = "Average number of stems per ha")

  # save plot
  ggsave(paste("results/gr_nb of stem per ha per BFI zone",plot_title,".png", sep=""), width = 15, height = 10)

}

#-----#
#           6. Calculate the basal area per plot and per project/zone/lccs
#-----#

t_save6  <- table_all
table_all <- t_save6

if (remove_dbh_less_than_10 != "ALL"){

  # Calc basal area
  table_all$ba_ha <- (table_all$dbh_cm/200)^2*pi*table_all$scale_factor

  #####
  ##### Remove outliers (run the code without removing to see them on graph)
  #####
  table_all <- table_all[table_all$id != "FD-10-SundRF_451",]
  table_all <- table_all[table_all$id != "FD-07-NFA_311211201",]

  # Sum at plot level
  plot_test <- Reduce(function(...)merge(..., by = "id"),
    list(a = aggregate(ba_ha~id, data = table_all, FUN=sum),
          b = aggregate(scale_factor~id, data = table_all, FUN=sum),
          c = unique(table_all[,c("id","entity_ab", "bfi_zone", "lccs6", "district")])
    )
  )
  names(plot_test)[3] <- "nstem_ha"
}

```

```

# overall check
summary(plot_test$nstem_ha)
summary(plot_test$ba_ha)

# Plot ba against nb stem
test <- plot_test
test %>%
  arrange(desc(entity_ab)) %>%
  ggplot(aes(x = nstem_ha, y = ba_ha)) +
  geom_point(aes(colour=entity_ab, pch = entity_ab)) +
  facet_wrap(~bfi_zone) +
  labs(x = "Number of stems per ha ", y = "Basal area (m2/ha)",
       color = "Project code", pch = "Project code") +
  theme(legend.position = c(0.8,0.2))

# save plot
ggsave(paste("results/gr_basal area against nstem per bfi zone", plot_title, ".png", sep=""),
       width = 15, height = 12, units = "cm", dpi = 300)

# # check frmp-97 per district (in line with FRMP report)
# test <- plot_test[plot_test$entity_ab == "FRMP-97",]
#
# vec <- unique(test$district)
# for(i in 1:length(vec)){
#
#   print(vec[i])
#   print(summary(test[test$district == vec[i], c("nstem_ha", "ba_ha")]))
# }

# Check nb stem per entity and bfi zone
myplot <- arrangeGrob(plot1 = ggplot(plot_test) +
  geom_boxplot(aes(x = entity_ab, y = ba_ha)) +
  labs(x = NULL, y = "Basal area (m2/ha)") +
  facet_grid(~bfi_zone, scales = "free_x", space = "free_x") + #activate on
  theme(axis.text.x = element_blank(),
        axis.ticks.x = element_blank(),
        axis.text.y = element_text(angle = 90)),
  plot2 = ggplot(plot_test) +
  geom_boxplot(aes(x = entity_ab, y = nstem_ha)) +
  labs(x = NULL, y = "Number of stem per ha") +
  facet_grid(~bfi_zone, scales = "free_x", space = "free_x") + #activate on
  theme(strip.background = element_blank(),
        strip.text.x = element_blank(),
        axis.text.x = element_text(angle = 90),
        axis.text.y = element_text(angle = 90)),
  ncol = 1)

grid.arrange(myplot)

# save plot
ggsave(paste("results/gr_boxplot basal area and nstem per ha",plot_title,".png", sep=""),
       myplot, width = 15, height = 18, units = "cm", dpi = 300)

#-----#
#-----# for table data correction: nb trees/plots after removing ba outliers
nb_tree <- table_all[,c("id","entity_ab", "fao_biome", "lccs6")]
nb_plot <- unique(nb_tree)

# Update the table nb tree/plots
nb_plot_tree <- cbind(nb_plot_tree,
  ba_out = paste(c(table(nb_plot$entity_ab),
                    dim(nb_plot)[1]),
                 " (",
                 c(table(nb_tree$entity_ab),
                    dim(nb_tree)[1]),
                 ")"),
  sep = "")

```



```

)
nb_plot_tree
#-----#
#-----#
}

#-----#
# 7. Write the modified table and a table with plot coordinates
#-----#

# Sub tables
t_gps <- unique(table_all[,c("id","x_corr","y_corr","entity_ab","id_plot",
                             "fao_biome","bfi_zone", "lccs1", "lccs2","lccs3",
                             "lccs4","lccs5","lccs6")])

t_h <- table_all[which(!is.na(table_all$h_m)),]

if (remove_dbh_less_than_10 == T) {

  write.csv(table_all      , file="results/table_all_01b_treesandplots_sup10.csv", row.names=FALSE)
  write.csv(t_gps         , file="results/t_gps_sup10.csv"                      , row.names=FALSE)
  write.csv(t_h           , file="results/t_hd_sup10.csv"                      , row.names=FALSE)
  write.csv(nb_plot_tree , file="results/nb_plot_tree_01b_sup10.csv"          , row.names=FALSE)

} else if (remove_dbh_less_than_10 == F) {

  write.csv(table_all      , file="results/table_all_01b_treesandplots_alltrees.csv", row.names=FALSE)
  write.csv(t_gps         , file="results/t_gps_alltrees.csv"                    , row.names=FALSE)
  write.csv(t_h           , file="results/t_hd_alltrees.csv"                  , row.names=FALSE)
  write.csv(nb_plot_tree , file="results/nb_plot_tree_01b_alltrees.csv"        , row.names=FALSE)

} else if (remove_dbh_less_than_10 == "ALL") {

  write.csv(t_h           , file="results/t_hd_forhdmmodel.csv"                , row.names=FALSE)
  write.csv(nb_plot_tree , file="results/nb_plot_tree_01b_forhdmmodel.csv"    , row.names=FALSE)

}

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

```