

Top-down disaggregation of Burkina Faso 2019 census - Scripts

Edith Darin, WorldPop

Contents

1	Introduction	1
2	Environment setup	1
3	Training the model	2
4	Predicting the population	2
5	Gridding the predicted pop	3
6	Disaggregating by age and sex groups	4
	References	4

1 Introduction

This document presents the data and R code used to estimate the top down model for disaggregating 2019 Burkina Faso census totals (Institut National de la Statistique et de la Démographie 2019). It is mainly based on scripts developed by Bondarenko for the WorldPop project (Bondarenko et al. 2018).

Please note that we used only their scripts aiming at estimating the model. All the pre-processing of the data input was done directly in our custom scripts as well as the prediction.

The data can be found in here [TO BE COMPLETED WHEN RELEASED].

2 Environment setup

We had first to configure the setting specifically for Burkina Faso as specified in Bondarenko (Bondarenko et al. 2018), that can be seen in the `input_BFA.R` script.

```
source(paste0(root_path, "/input_BFA.R"))
```

Then we need to load all the required functions for estimating the tRandom Forest model for population modelling.

```
source(paste0(root_path, "/config.R"))
source(paste0(root_path, "/load_Packages.R"))
source(paste0(root_path, "/internal_functions.R"))
source(paste0(root_path, "/create_dirs_for_prj.R"))
```

```
source(paste0(root_path, "/rf_functions.R"))

if (!load.Packages())
  stop("There was an error when loading R packages")
```

We create all the necessary directories for the estimation.

3 Training the model

We load the data. The data are in a table format with each row being an admin 3.

```
source(paste0(root_path, "/variable_names.R"))

train <- readRDS(paste0(data_path, "/BFA_population_v1_0_train.rds"))
predict <- readRDS(paste0(data_path, "/BFA_population_v1_0_predict.rds"))
age_sex <- read.csv(paste0(data_path, "/BFA_population_v1_0_agesex.csv"), stringsAsFactors = F)

# mastergrid to convert back the prediction into raster format
masterGrid <- raster(paste0(data_path, "/masterGrid.tif"))
```

We prepare the predictors and the response variable.

```
# Predictors
cov_names <- colnames(train)
cov_names <- cov_names[5:(length(cov_names)-19)]
cov_names <- cov_names[!grepl("settled_binary_V3|settledArea_V3|Img_date", cov_names)]

x_data <- train[, cov_names]

# Response variable: population per sum of settled pixel size
y_data <- log(as.numeric(train$pop_density_pixels))
```

We train the model.

```
## Fit the RF, removing any covariates which are not important to the model:
popfit <- get_popfit()

## Fit the final RF
popfit_final <- get_popfit_final()
```

4 Predicting the population

We prepare the prediction data set in a table format with each row being a settled pixel.

```
cov_names <- names(popfit_final$forest$xlevels)
cov_predict <- predict[, ..cov_names]
cov_predict$admin3_id <- predict$admin3_id
cov_predict$masterGrid_id <- predict$masterGrid_id
cov_predict$settledArea <- predict$settledArea_V3
```

We define the function that predict density at pixel level and convert it to a weight which is multiply by the total census count of the related admin 3.

```
predict_weights <- function(df, census, model=popfit_final){
  prediction_set <- predict(model,
                           newdata=df,
                           predict.all=TRUE)
  output <- data.table(rf_pred = exp(apply(prediction_set$individual, MARGIN=1, mean)))

  output$admin3_id <- df$admin3_id
  output$masterGrid_id <- df$masterGrid_id
  output$weight <- output$rf_pred
  output$pop <- (output$weight/sum(output$weight))*census

  fwrite(output, paste0(output_path, "/topdown/output/prediction/predictions_",
                        df$admin3_id[1], ".csv"))
}
```

We run the prediction in parallel mode, each batch containing all the pixels of a given admin 3;

```
cov_predict_list <- as.data.frame(cov_predict) %>%
  group_by(admin3_id) %>%
  group_split()

co <- detectCores()-2
tic()
cl <- makeCluster(co)
registerDoParallel(cl)
predicted <- NULL
predicted <- foreach(
  i=1:length(cov_predict_list),
  .packages=c("tidyverse", "data.table", "randomForest")) %dopar% {
  predict_weights(
    cov_predict_list[[i]],
    master %>% filter(admin3_id==i) %>% dplyr::select(population) %>% unlist()
  )
}

stopCluster(cl)
toc() #30sec
```

5 Gridding the predicted pop

We convert the prediction tables into a raster representing the gridded population.

```
# reading in the predictions
predictions_list <- list.files(
  paste0(output_path, "/topdown/output/prediction/"), pattern = ".csv")
print(length(predictions_list))

readPredictions <- function(file){
  df <- fread(paste0(output_path, "/topdown/output/prediction/", file)
  )
  return(df)
```

```

}

tic()
predictions <- do.call("rbind",
                      lapply(
                        predictions_list,
                        function(x) readPredictions(x))
                      )
toc() #20

# assign the prediction to its pixel using the mastergrid id

raster_df <- data.table(masterGrid[])
names(raster_df) <- "masterGrid_id"

raster_df <- predictions[raster_df, on="masterGrid_id"]
r <- masterGrid
r[] <- raster_df$pop

# writing the output raster

writeRaster(r,
            paste0(output_path, "/BFA_population_v1_0_gridded.tif"),
            overwrite=T)

```

6 Disaggregating by age and sex groups

To obtain age and sex disaggregation at pixel level, we multiply the predicted population count by the national age and sex proportions.

```

for(i in 1:(ncol(age_sex)-1)){
  prop <- age_sex[1,i] %>% unlist()
  print(names(prop))
  r_decomp <- r* prop
  writeRaster(r_decomp,
             paste0(output_path,
                   'BFA_population_v1_0_agesex/BFA_population_v1_0_agesex_',
                   names(prop),
                   '.tif'),
             overwrite=T)
}

```

References

- Bondarenko, Maksym, Jeremiah Nieves, Alessandro Sorichetta, Forrest R Stevens, Andrea E Gaughan, Andrew Tatem, and others. 2018. "WpgpRFPMS: WorldPop Random Forests Population Modelling R Scripts, Version 0.1. 0."
- Institut National de la Statistique et de la Démographie. 2019. *Recensement Général de La Population et de L'habitation de 2019 Du Burkina Faso - Résultats Provisoires*. INSD Ouagadougou, Burkina Faso.