Human Influence in SDMs: Literature Review (Part V)

Veronica F. Frans (email: verofrans@gmail.com)

February 3, 2024

Contents

1 Summary

This is the fifth R script of the literature review and synthesis for the article entitled, “Gaps and opportunities in modeling human influence on species distributions in the Anthropocene,” by Veronica F. Frans and Jianguo Liu.

Here, in Part V of the synthesis, we use the CSV files of the predictor list from Part III to evaluate whether some of the human predictors used by authors are relevant for monitoring the United Nations Sustainable Development Goals (SDGs). This assessment is appended to the predictor list for access in the Supporting Information.

Thus, the following is accomplished:

(1) Matching of predictor names with the 17 SDGs
(2) Summary of predictor use across SDGs and papers
(3) Plots of SDGs by paper and predictor
(4) Final export of predictor list with SDGs included

This is the final R script for this study.

2 R Setup

We are using R version 4.3.0 (R Core Team 2023).

2.1 Libraries

Load libraries

```
# load libraries
  library("dplyr")          # for table manipulations
  library("scales")         # for scales and formatting
  library("kableExtra")     # for table viewing in Rmarkdown
  library("tidyr")          # for table manipulations
  library("plyr")           # for table manipulations
  library("tidyverse")      # for graphics/table management
  library("ggplot2")        # for graphics
  library("RColorBrewer")   # for graphics
  library("ggforce")        # for graphics (speeds up ggplot)
  library("ggalluvial")     # for graphics
  library("ggbreak")        # for graphics
  library("patchwork")      # for graphics
  library("ggExtra")        # for graphics
  library("ggrepel")        # for labeling
  library("plotfunctions")  # for data visualization
  library("svglite")        # for saving graphics in svg format
  library("SDGdetector")    # for text-mining SDGs
  library("text2sdg")       # for text-mining SDGs
```

2.2 Directories

The primary directory is the folder where the `hum_sdm_litrv_r.Rproj` is stored.

```
# create image folder and its directory
  dir.create(paste0("images"))
  image.dir <- paste0("images\\")

# create data folder and its directory
  dir.create(paste0("data"))
  data.dir <- paste0("data\\")
```

2.3 Load data

Upload the data table from the abstract screening and review, and subset to only the articles that are accepted. We will also need a few saved CSV files from Part II.

```
# study domain, taxa, and focus
  domtaxfoc.df <- read.csv(paste0(data.dir,"domain_taxa_focus_count_papers.csv"),
                           header=T, sep=",")

# predictor list
  preds.list.export <- read.csv(paste0(data.dir,"predictor_list_summary_FINAL.csv"),
                                header=T, sep=",")
```

3 Classify SDGs across predictors

We use the `text2sdg` package to match predictors with the 17 SDGs. Because this packages normally works with an entire corpus, the analysis is instead done row by row (i.e., per predictor), treating each predictor as a corpus. Row by row, names and sums of SDGs are extracted. Note that this takes at least 9 hours to run.

```
# Record start time
  start_time <- proc.time()

# Run SDG scan for all predictor rows
  output <- lapply(preds.list.export$predictor, detect_sdg, verbose=FALSE)

# Combine results and add predictor column
  preds.sdgs <- data.frame(
                           predictor = preds.list.export$predictor,
                           sdg = sapply(output, function(x)paste(unique(x$sdg),
                                                                   collapse = "; ")),
                           count_sdgs = sapply(output,
                                               function(x)length(unlist(strsplit(x$sdg,";"))))
                          )

# Optionally, if you need to order the rows by the predictor column
  preds.sdgs <- preds.sdgs[order(preds.sdgs$predictor), ]

# Reset row names if desired
  row.names(preds.sdgs) <- NULL

# Output the combined result
  preds.sdg.export <- preds.sdgs
```

```r
# save as CSV
  write.csv(preds.sdg.export, paste0(data.dir,"predictors_and_SDGs.csv"),row.names=FALSE)

# Record end time
  end_time <- proc.time()

# Calculate the runtime
  runtime <- end_time - start_time
  runtime/60/60
```

```
       user      system     elapsed 
 7.03329444  0.09487778 11.69226111 
```

3.1 Plot SDG relevance across human predictors

Next, we calculate the sum and percents of predictors per SDG.

```r
# extend list of SDGs per predictor
  preds.sdg.long <- separate_rows(preds.sdg.export, sdg, sep="; ",convert = TRUE)

# summary list of predictor counts per SDG
  sdg_sums <- ddply(preds.sdg.long, .(sdg),
                    summarize,
                    # count of predictors
                    num_preds=length(unique(predictor)),
                    # percent of predictors
                    perc_preds=round(length(unique(
                      predictor))/nrow(preds.list.export)*100,4),
                    # list of predictors
                    predictors=paste(unique(predictor),collapse="; ")
                    )

# show summary here (predictors not shown)
  kableExtra::kbl(sdg_sums[,1:3], booktabs=T, longtable=T) %>%
    kable_styling(latex_options = c("striped","repeat_header"))
```

sdg	num_preds	perc_preds
	1625	70.4378
SDG-01	41	1.7772
SDG-02	65	2.8175
SDG-03	3	0.1300
SDG-04	2	0.0867
SDG-06	253	10.9666
SDG-07	9	0.3901
SDG-08	2	0.0867
SDG-09	9	0.3901
SDG-11	282	12.2237
SDG-12	4	0.1734
SDG-13	4	0.1734

(continued)

sdg	num_preds	perc_preds
SDG-14	35	1.5171
SDG-15	246	10.6632

Add missing SDGs to the list.

```
# remove the NA field
  sdg_only <- sdg_sums[!sdg_sums$sdg=='',]

# add SDGs that are missing
  missing_sdgs <- data.frame(sdg=c("SDG-05","SDG-10","SDG-16","SDG-17"),
                             num_preds=0,
                             perc_preds=0,
                             predictors="")

# combine and re-sort
  sdg_only <- rbind(sdg_only,missing_sdgs)
  sdg_only <- sdg_only[order(sdg_only$sdg),]

# show summary here (predictors not shown)
  kableExtra::kbl(sdg_only[,1:3], booktabs=T, longtable=T) %>%
    kable_styling(latex_options = c("striped","repeat_header"))
```

	sdg	num_preds	perc_preds
2	SDG-01	41	1.7772
3	SDG-02	65	2.8175
4	SDG-03	3	0.1300
5	SDG-04	2	0.0867
1	SDG-05	0	0.0000
6	SDG-06	253	10.9666
7	SDG-07	9	0.3901
8	SDG-08	2	0.0867
9	SDG-09	9	0.3901
21	SDG-10	0	0.0000
10	SDG-11	282	12.2237
11	SDG-12	4	0.1734
12	SDG-13	4	0.1734
13	SDG-14	35	1.5171
14	SDG-15	246	10.6632
31	SDG-16	0	0.0000
41	SDG-17	0	0.0000

Plot summary

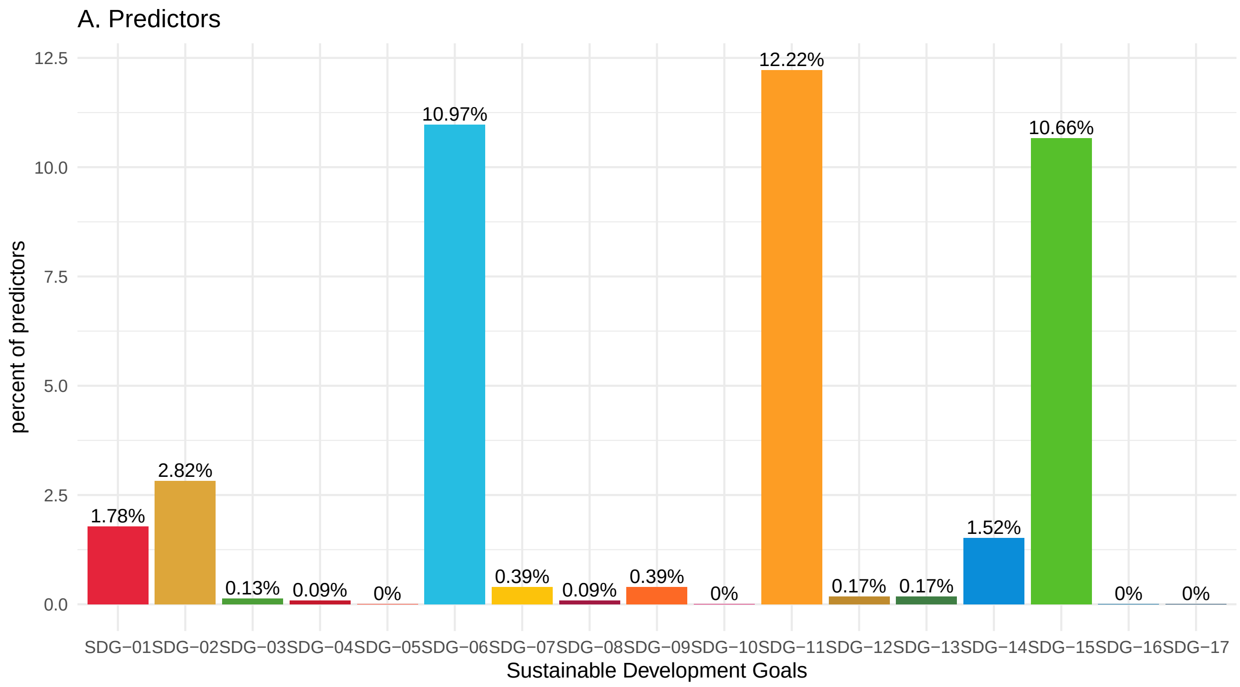
```
# SDG colors (taken from SDGdetector package)
  color_rgb <- data.frame(
      R = c(229, 221, 76, 197, 255, 38, 252, 162, 253, 221, 253, 191, 63, 10, 86, 0, 25),
      G = c(36, 166, 159, 25, 58, 189, 195, 25, 105, 19, 157, 139, 126, 141, 192, 104, 72),
      B = c(59, 58, 56, 45, 33, 226, 11, 66, 37, 103, 36, 46, 68, 217, 43, 157, 106)
    )

    ### HEX
    color_hex <- rgb(color_rgb, maxColorValue = 255)
    names(color_hex) <- sdg_only$sdg
    sdg_color <- function(x) {
                              color <- color_hex[x]
                              return(color)
                              }

# plot
  sdg.plt <- ggplot(data=sdg_only, aes(x=sdg, y=perc_preds, fill=sdg)) +
                    scale_fill_manual(values = sdg_color(x = 1:17)) +
                    geom_bar(stat="identity") +
                    geom_text(aes(label=paste0(round(perc_preds,2),'%')),
                              vjust=-0.3, size=3.5) +
                    ylab('percent of predictors') +
                    xlab('Sustainable Development Goals')+
                    theme_minimal() +
                    theme(legend.position = "none") +
                    ggtitle('A. Predictors')

# save
  ggsave(plot=sdg.plt, filename = paste0(image.dir,'sdg_bar_plot.png'),
         height = 5, width = 9, units = 'in', dpi = 600)
  ggsave(plot=sdg.plt, filename = paste0(image.dir,'sdg_bar_plot.svg'),
         height = 5, width = 9, units = 'in')

# view
  sdg.plt
```



## Total predictors related to SDGs

```
# get summary
  paste(nrow(preds.sdg.export[preds.sdg.export$count_sdgs>0,]),
        "out of",
        nrow(preds.sdg.export),
        "predictors",
        "are related to Sustainable Development Goals")
```

```
## [1] "682 out of 2307 predictors are related to Sustainable Development Goals"
```

3.2 Summarize predictor use compared to number of articles

Next, we calculate the sum and percents of papers per SDG.

```
# append article numbers to predictors
  preds.sdg.papers <- left_join(preds.list.export, preds.sdg.export,
                                by='predictor')

# extend list by SDGs per predictor
  preds.sdg.papers <- separate_rows(preds.sdg.papers, sdg, sep="; ", convert = TRUE)

# extend list by paper per predictor
  preds.sdg.papers <- separate_rows(preds.sdg.papers, uid, sep="; ", convert = TRUE)

# summary list of predictor counts per SDG
  sdg_paper_sums <- ddply(preds.sdg.papers, .(sdg),
                          summarize,
                          # count of papers
                          num_papers=length(unique(uid)),
                          # percent of papers
                          perc_papers=round(
                            length(unique(uid))/length(unique(
                              preds.sdg.papers$uid))*100,4),
                          # list of predictors
                          predictors=paste(unique(predictor),collapse="; ")
                          )

# show summary here (predictors not shown)
  kableExtra::kbl(sdg_paper_sums[,1:3], booktabs=T, longtable=T) %>%
    kable_styling(latex_options = c("striped","repeat_header"))
```

sdg	num_papers	perc_papers
	1113	77.8866
SDG-01	39	2.7292
SDG-02	52	3.6389
SDG-03	3	0.2099
SDG-04	2	0.1400
SDG-06	299	20.9237
SDG-07	8	0.5598

(continued)

sdg	num_papers	perc_papers
SDG-08	2	0.1400
SDG-09	9	0.6298
SDG-11	444	31.0707
SDG-12	3	0.2099
SDG-13	4	0.2799
SDG-14	51	3.5689
SDG-15	580	40.5878

Add missing SDGs to the list.

```
# remove the NA field
  sdg_paper_only <- sdg_paper_sums[!sdg_paper_sums$sdg=='',]

# add SDGs that are missing
  missing_sdgs <- data.frame(sdg=c("SDG-05","SDG-10","SDG-16","SDG-17"),
                             num_papers=0,
                             perc_papers=0,
                             predictors="")

# combine and re-sort
  sdg_paper_only <- rbind(sdg_paper_only,missing_sdgs)
  sdg_paper_only <- sdg_paper_only[order(sdg_paper_only$sdg),]

# show summary here (predictors not shown)
  kableExtra::kbl(sdg_paper_only[,1:3], booktabs=T, longtable=T) %>%
    kable_styling(latex_options = c("striped","repeat_header"))
```

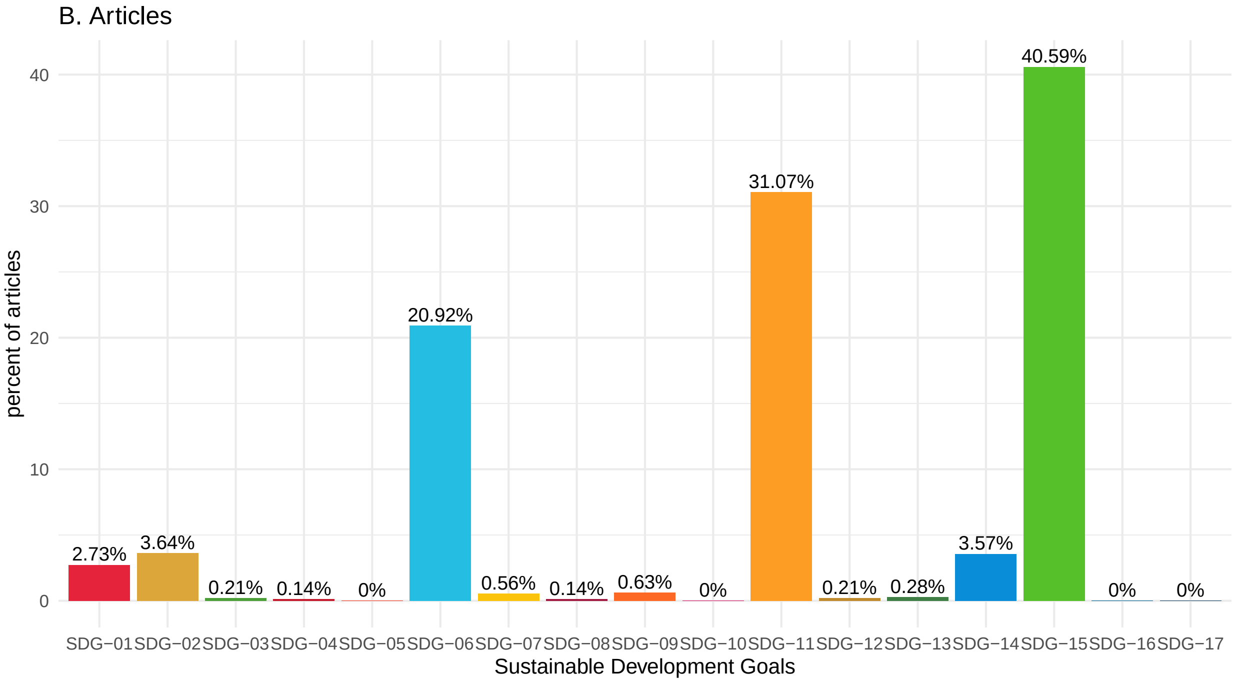
	sdg	num_papers	perc_papers
2	SDG-01	39	2.7292
3	SDG-02	52	3.6389
4	SDG-03	3	0.2099
5	SDG-04	2	0.1400
1	SDG-05	0	0.0000
6	SDG-06	299	20.9237
7	SDG-07	8	0.5598
8	SDG-08	2	0.1400
9	SDG-09	9	0.6298
21	SDG-10	0	0.0000
10	SDG-11	444	31.0707
11	SDG-12	3	0.2099
12	SDG-13	4	0.2799
13	SDG-14	51	3.5689
14	SDG-15	580	40.5878
31	SDG-16	0	0.0000
41	SDG-17	0	0.0000

Plot summary

```
# plot
  sdgp.plt <- ggplot(data=sdg_paper_only, aes(x=sdg, y=perc_papers, fill=sdg)) +
                     scale_fill_manual(values = sdg_color(x = 1:17)) +
                     geom_bar(stat="identity") +
                     geom_text(aes(label=paste0(round(perc_papers,2),'%')),
                               vjust=-0.3, size=3.5) +
                     ylab('percent of articles') + xlab('Sustainable Development Goals')+
                     theme_minimal() +
                     theme(legend.position = "none") +
                     ggtitle('B. Articles')

# save
  ggsave(plot=sdgp.plt, filename = paste0(image.dir,'sdg_papers_bar_plot.png'),
         height = 5, width = 9, units = 'in', dpi = 600)
  ggsave(plot=sdgp.plt, filename = paste0(image.dir,'sdg_papers_bar_plot.svg'),
         height = 5, width = 9, units = 'in')

# view
  sdgp.plt
```



## Total papers related to SDGs

```
# get summary
  paste(length(unique(preds.sdg.papers$uid[preds.sdg.papers$count_sdgs>0])),
        "out of",
        length(unique(preds.sdg.papers$uid)),
        "papers",
        "use human predictors related to Sustainable Development Goals")
```

```
## [1] "924 out of 1429 papers use human predictors related to Sustainable Development Goals"
```

4 Combined plot

We turn these plots into a scatterplot, with the number of predictors related to SDGs on the x-axis, and the number of papers using those predictors on the y-axis.

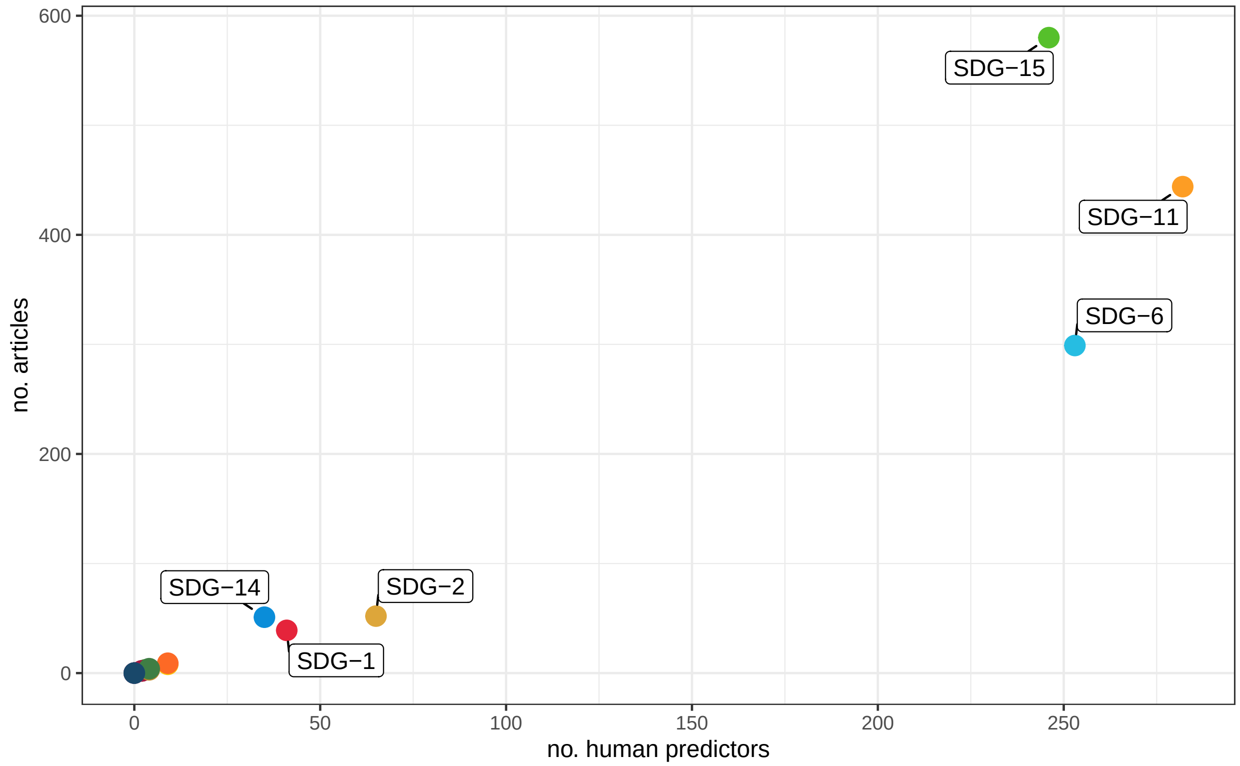
```r
# make a new dataframe
  sdg_summary <- data.frame(sdg=sdg_only$sdg,
                            num_preds=sdg_only$num_preds,
                            perc_preds=sdg_only$perc_preds,
                            num_papers=sdg_paper_only$num_papers,
                            perc_papers=sdg_paper_only$perc_papers)

# label the top 6 SDGs
  sdg_summary$labs <- ''
  sdg_summary$labs[sdg_summary$sdg=='SDG-01'] <- 'SDG-1'
  sdg_summary$labs[sdg_summary$sdg=='SDG-02'] <- 'SDG-2'
  sdg_summary$labs[sdg_summary$sdg=='SDG-06'] <- 'SDG-6'
  sdg_summary$labs[sdg_summary$sdg=='SDG-11'] <- 'SDG-11'
  sdg_summary$labs[sdg_summary$sdg=='SDG-14'] <- 'SDG-14'
  sdg_summary$labs[sdg_summary$sdg=='SDG-15'] <- 'SDG-15'

# plot
  sdg_pts <- ggplot(sdg_summary,
                    aes(x=num_preds, y=num_papers, color=sdg, label=labs)) +
                    geom_point(color=sdg_color(x = 1:17), size = 4, guide = FALSE) +
                    scale_fill_manual(values = sdg_color(x = 1:17)) +
                    geom_label_repel(aes(label = labs),
                                     box.padding   = 0.35,
                                     point.padding = 1,
                                     segment.color = 'black',
                                     segment.size = 0.5,
                                     color = 'black',
                                     min.segment.length = 0.1) +
                    ylab('no. articles') + xlab('no. human predictors') +
                    scale_x_continuous(breaks = seq(0,300,50)) +
                    theme_bw()

# save
  ggsave(plot=sdg_pts, filename = paste0(image.dir,'sdg_scatterplot.png'),
         height = 4, width = 6, units = 'in', dpi = 600)
  ggsave(plot=sdg_pts, filename = paste0(image.dir,'sdg_scatterplot.svg'),
         height = 4, width = 6, units = 'in')

# view
  sdg_pts
```



5 Append SDGs to predictor list dataset

We add the SDG names and counts to the existing predictor list dataset, for export and use in Supporting Information.

```
# append article numbers to predictors
  preds.sdg.papers.export <- left_join(preds.list.export, preds.sdg.export,
                                       by='predictor')

# reorder columns
# use colname = newcolname notation for renaming
  preds.sdg.papers.export <- relocate(preds.sdg.papers.export,
                                      sdg,
                                      .after = uid)
  preds.sdg.papers.export <- relocate(preds.sdg.papers.export,
                                      count_sdgs,
                                      .after = sdg)

# rename column
  colnames(preds.sdg.papers.export)[10] <- "count_papers"

# save
  write.csv(preds.sdg.papers.export,paste0(data.dir,"predictor_SDG_list_summary_FINAL.csv"),
            row.names = FALSE)
```

6 Save

```r
# save progress
  save.image("SDMs_human_lit_review_V.RData")
```

THIS IS THE END OF THE SCRIPT.
