Code of Fallacies in the Global Water Crisis statistics

R code

Arnald Puy

Contents

```r
# PRELIMINARY ----------------------------------------------------------------

sensobol::load_packages(c("sensobol", "tidyverse", "data.table", "cowplot"))


theme_AP <- function() {
  theme_bw() +
    theme(panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.background = element_rect(fill = "transparent",
                                           color = NA),
          legend.margin=margin(0, 0, 0, 0),
          legend.box.margin=margin(-5,-5,-5,-5),
          legend.key = element_rect(fill = "transparent",
                                    color = NA),
          strip.background = element_rect(fill = "white"),
          axis.title = element_text(size = 9),
          legend.text = element_text(size = 9),
          legend.title = element_text(size = 9),
          legend.key.width = unit(0.3, "cm"),
          legend.key.height = unit(0.3, "cm"))
}
```

```r
# SETTINGS ###########################################################################

# Values used in the paper
precipitation_estimate <- 120000
precipitation_min <- precipitation_estimate - (precipitation_estimate * 0.1)
precipitation_max <- precipitation_estimate + (precipitation_estimate * 0.1)

land_runoff_estimate <- 46000
land_runoff_min <- land_runoff_estimate - (land_runoff_estimate  * 0.1)
land_runoff_max <- land_runoff_estimate + (land_runoff_estimate  * 0.1)

# SAMPLE MATRIX ######################################################################

N <- 2^12
params <- c("precipitation", "green_water_consumption_crops",
            "global_green_water_consumption", "global_consumption",
            "planetary_boundary")

mat <- sobol_matrices(N = N, params = params)

mat[, "precipitation"] <- floor(qunif(mat[, "precipitation"], precipitation_min, precipitation_
mat[, "green_water_consumption_crops"] <- floor(qunif(mat[, "green_water_consumption_crops"], 5
mat[, "global_green_water_consumption"] <- floor(qunif(mat[, "global_green_water_consumption"]
mat[, "global_consumption"] <- floor(qunif(mat[, "global_consumption"], 3391, 5349))
```

```r
mat[, "planetary_boundary"] <- floor(qunif(mat[, "planetary_boundary"], 4000, 6000))

# RUN MODELS ##############################################################

land_runoff <- mat[, "precipitation"] - mat[, "global_green_water_consumption"] -
  mat[, "green_water_consumption_crops"]

exceedance <-  mat[, "planetary_boundary"] - mat[, "global_consumption"]

# ARRANGE DATA ############################################################

dt <- cbind(land_runoff, exceedance) %>%
  data.table() %>%
  .[1:(2 * N)] %>%
  .[, outside:= ifelse(land_runoff < land_runoff_min |
                         land_runoff > land_runoff_max, "Yes", "No")] %>%
  .[, accessible_water_runoff:= land_runoff - 7800 - 20400] %>%
  .[, outside_runoff:= ifelse(accessible_water_runoff < 12500 |
                                accessible_water_runoff > 18500, "Yes", "No")] %>%
  .[, water_deficit:= ifelse(exceedance < 0, "Yes", "No")]


# SOME STATS ##############################################################

cols <- c("land_runoff", "accessible_water_runoff", "exceedance")
summary_fun = function(x) list(min = min(x), max = max(x))
dt[, lapply(.SD, summary_fun), .SDcols = (cols)]
```

```
##    land_runoff accessible_water_runoff exceedance
## 1:       33442                    5242      -1334
## 2:       58554                   30354       2563
```

```r
tmp <- melt(dt, measure.vars = c("outside", "outside_runoff", "water_deficit")) %>%
  .[, .N, .(variable, value)]

tmp[, total:= (2^12 * 2)] %>%
  .[, prop:= N / total] %>%
  print()
```

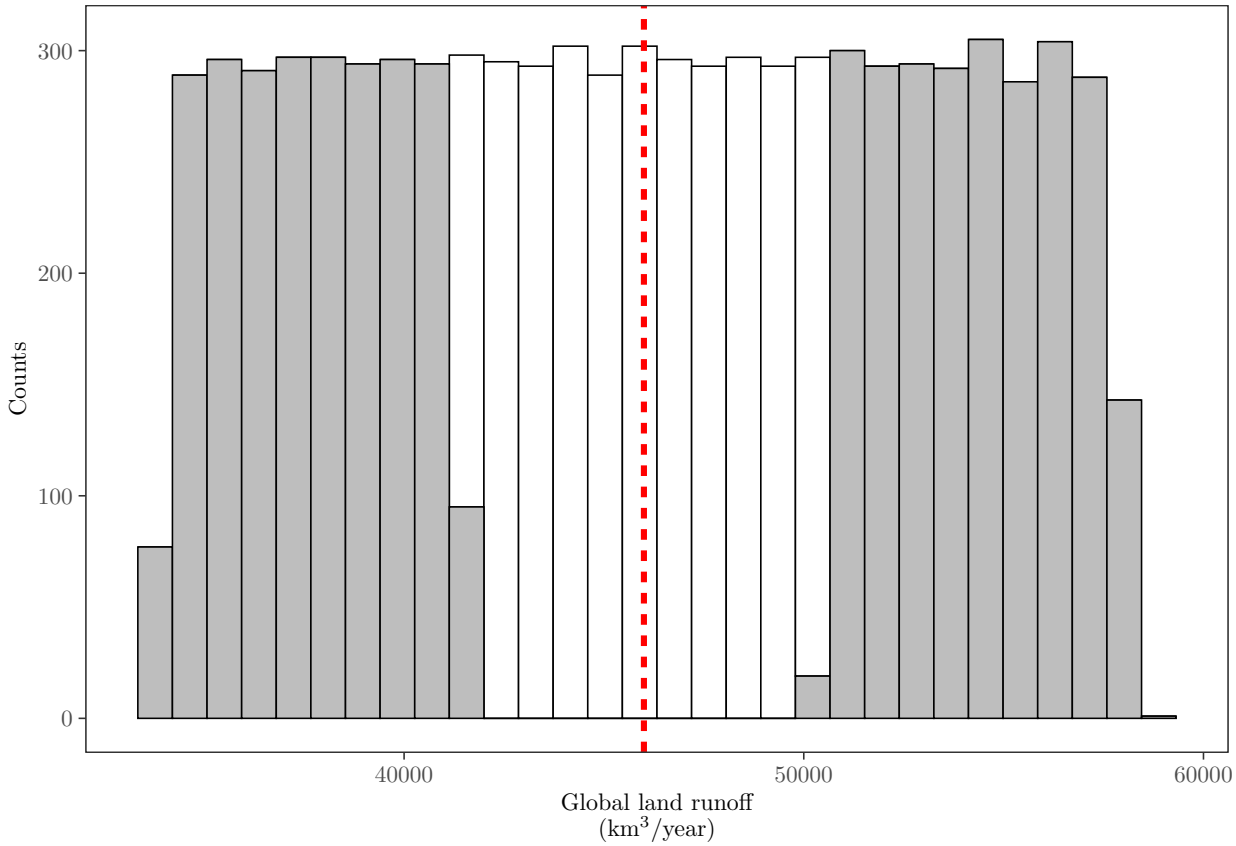
```
##          variable value    N total      prop
## 1:        outside    No 3141  8192 0.3834229
## 2:        outside   Yes 5051  8192 0.6165771
## 3: outside_runoff    No 2047  8192 0.2498779
## 4: outside_runoff   Yes 6145  8192 0.7501221
## 5:  water_deficit    No 6272  8192 0.7656250
## 6:  water_deficit   Yes 1920  8192 0.2343750
```

```
# PLOT LAND RUNOFF DISTRIBUTION ###################################################

plot_land_runoff <- ggplot(dt, aes(land_runoff, fill = outside)) +
  geom_histogram(colour = "black") +
  scale_fill_manual(values = c("white", "grey")) +
  theme_AP() +
  geom_vline(xintercept = land_runoff_estimate, color = "red", lty = 2, size = 2) +
  labs(x = "Global land runoff \n (km$^3$/year)", y = "Counts") +
  theme(legend.position = "none")
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
plot_land_runoff
```

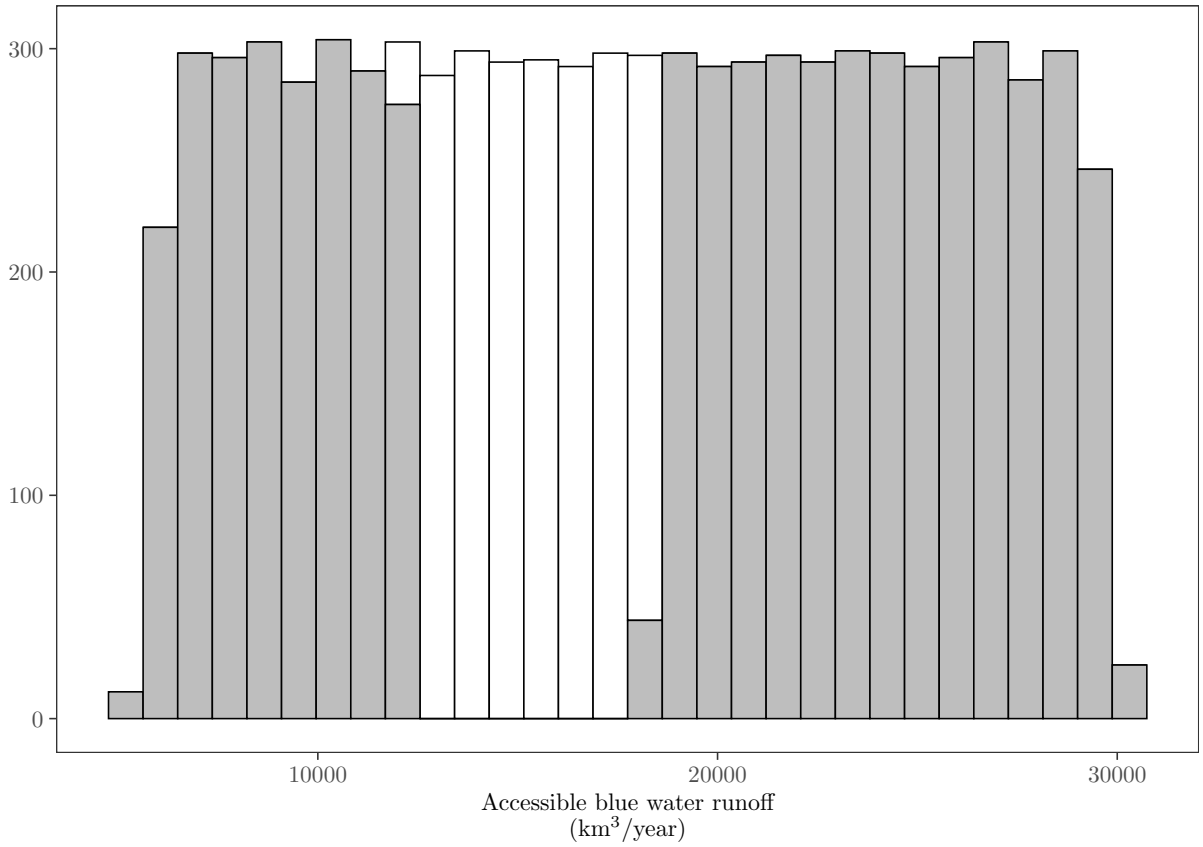
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
plot_accessible_runoff <- ggplot(dt, aes(accessible_water_runoff, fill = outside_runoff)) +
  geom_histogram(colour = "black") +
  scale_fill_manual(values = c("white", "grey")) +
  theme_AP() +
  labs(x = "Accessible blue water runoff \n (km$^3$/year)", y = "") +
```

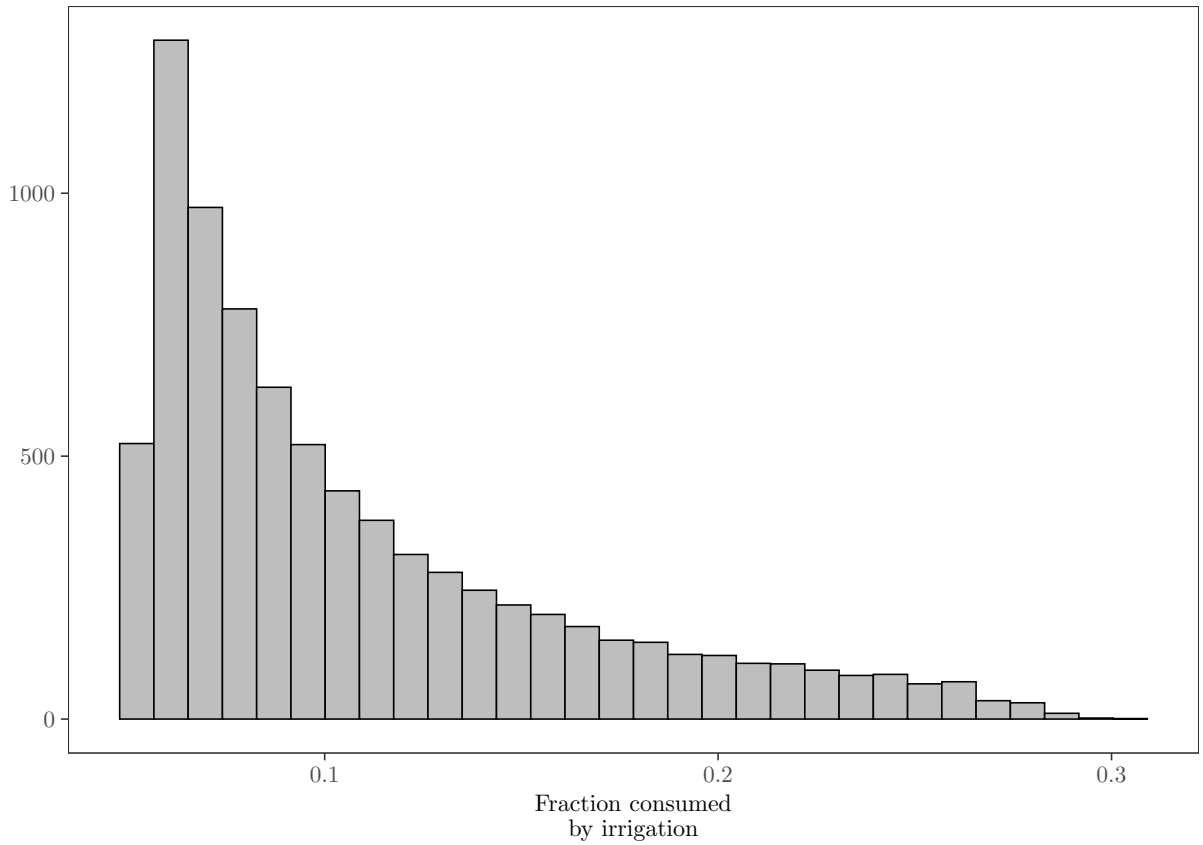
```
  theme(legend.position = "none")

plot_accessible_runoff
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



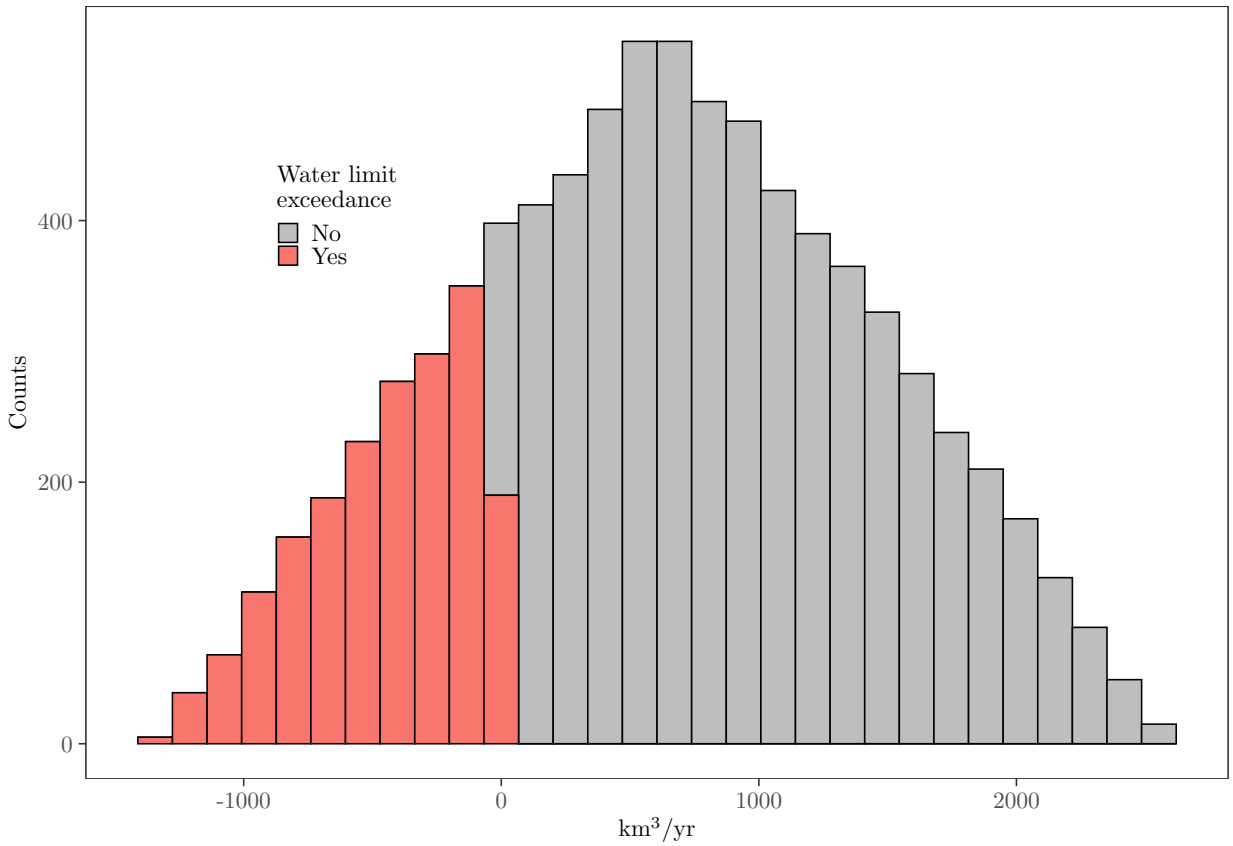

```
vec <- data.table(fraction.irrig = 1600 / dt$accessible_water_runoff)

plot.irrigation <- ggplot(vec, aes(fraction.irrig)) +
  geom_histogram(fill = "grey", color = "black") +
  labs(x = "Fraction consumed \n by irrigation", y = "") +
  theme_AP()

plot.irrigation
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
plot.exceedance <- ggplot(dt, aes(exceedance, fill = water_deficit)) +
  geom_histogram(color = "black") +
  scale_fill_manual(values = c("grey", "#F8766D"),
                    name = "Water limit \n exceedance") +
  labs(x = "km$^3$/yr", y = "Counts") +
  theme_AP() +
  theme(legend.position = c(0.22, 0.73))

plot.exceedance
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



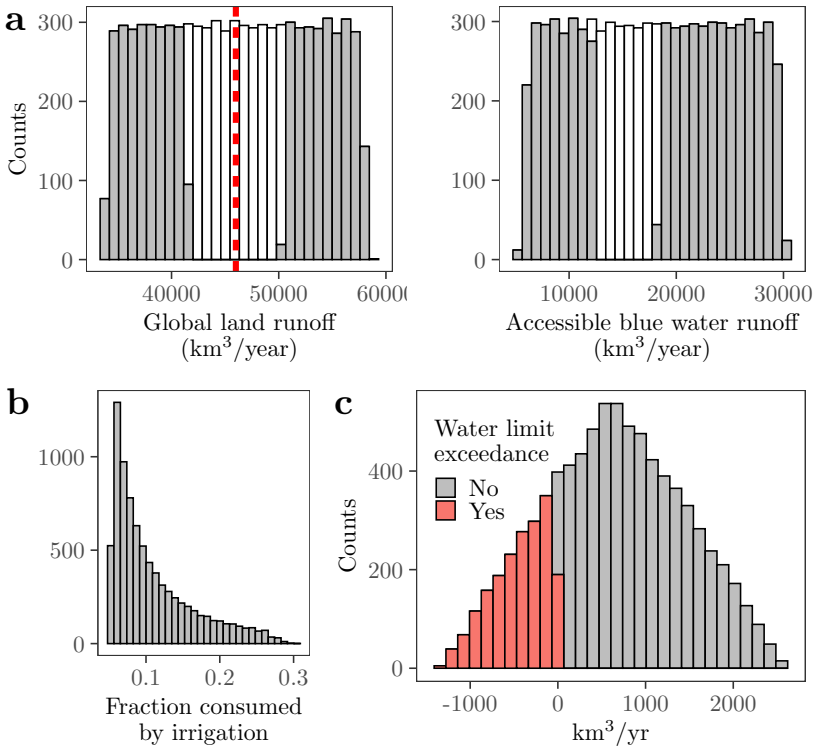

```
# MERGE PLOTS ##################################################################

top <- plot_grid(plot_land_runoff, plot_accessible_runoff, labels = c("a", ""))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
bottom <- plot_grid(plot.irrigation, plot.exceedance, labels = c("b", "c"),
                    rel_widths = c(0.4, 0.6))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
plot_grid(top, bottom, ncol = 1)
```


a
300
200
100
0
Counts
40000
50000
60000
Global land runoff
(km3/year)
300
200
100
0
10000
20000
30000
Accessible blue water runoff
(km3/year)
b
1000
500
0
0.1
0.2
0.3
Fraction consumed
by irrigation
c
Water limit
exceedance
No
Yes
400
200
0
Counts
-1000
0
1000
2000
km3/yr