Yield gap decomposition: Theory & practice

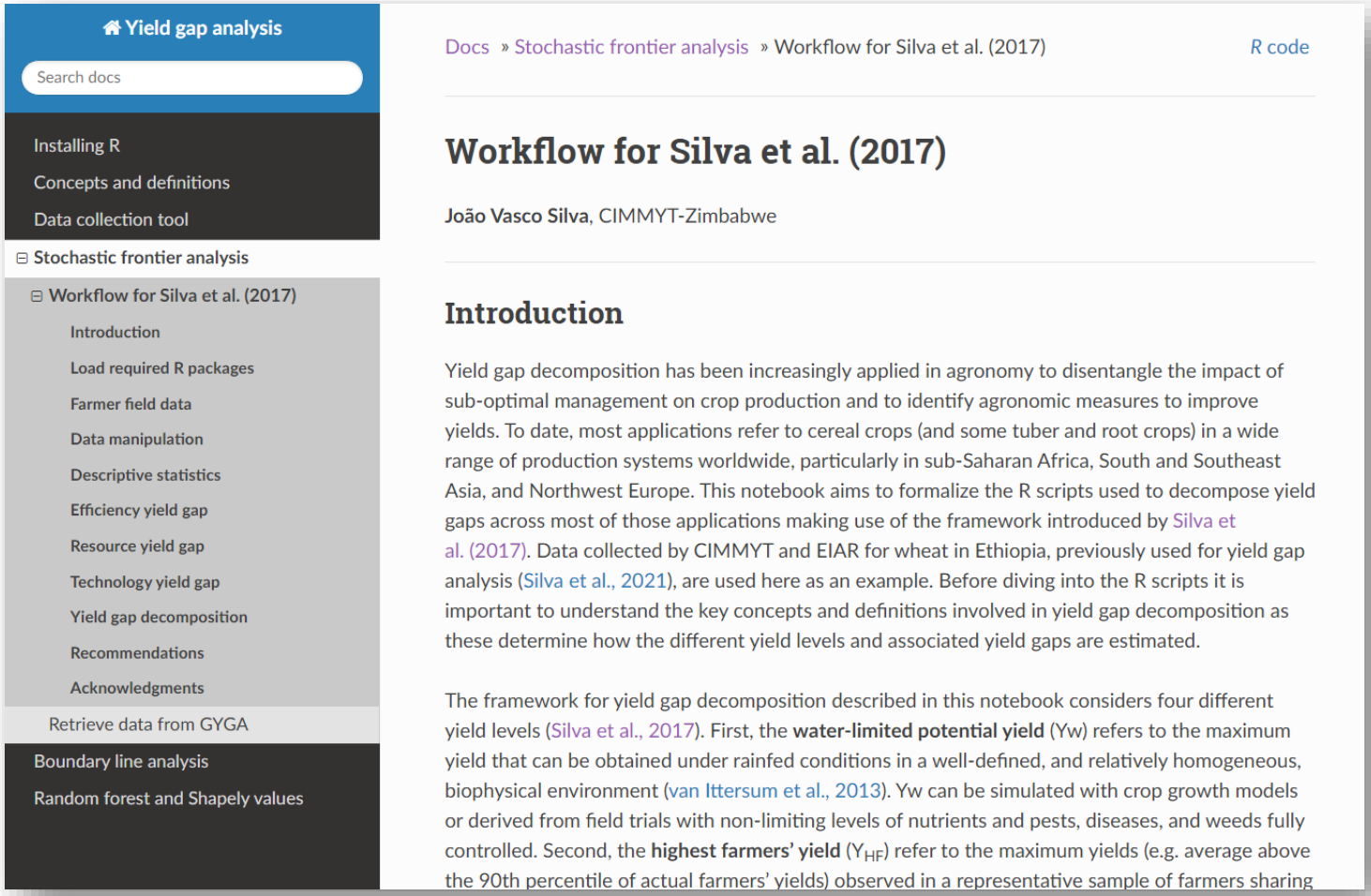
João Vasco Silva (PhD)
j.silva@cgiar.org
Agronomy-at-scale Data Scientist
CIMMYT-Zimbabwe



Addis Ababa, May 24th 2023

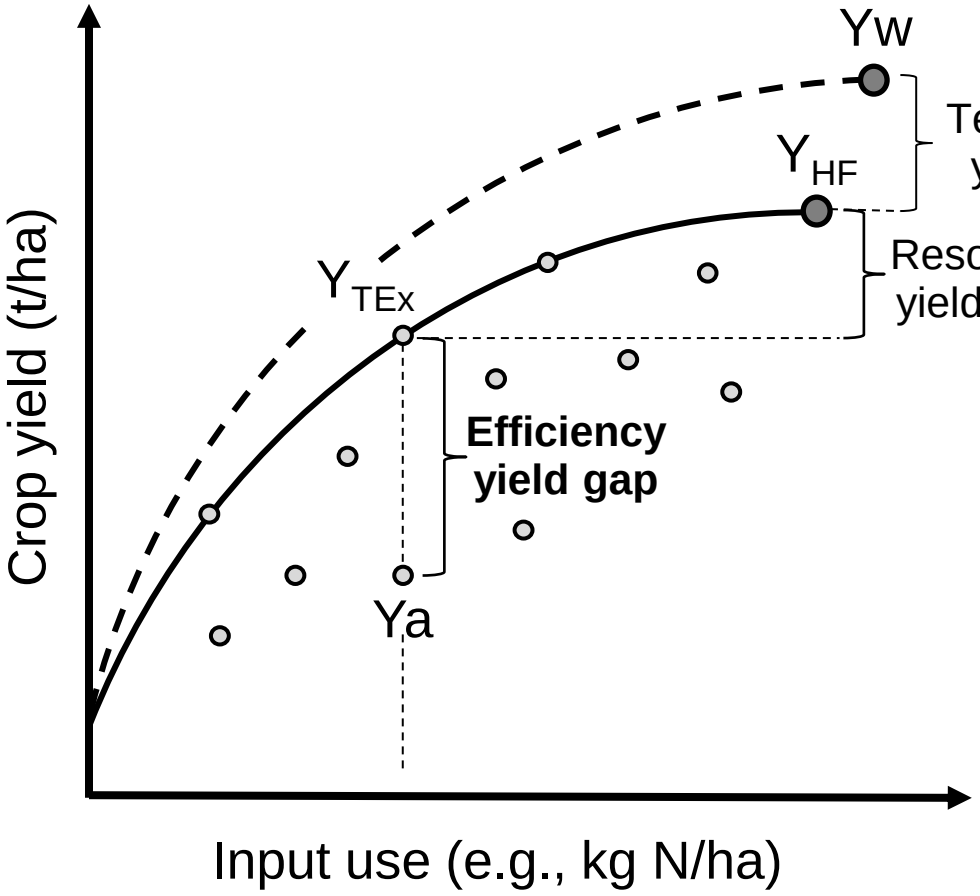
CIMMYT

https://jvasco323.github.io/eia-yg-training



CIMMYT

Efficiency yield gap and Y_{TEx}



Ya = actual farmers' yields

- Farm household surveys

Y_{TEx} = technical efficient yields

- Stochastic frontier analysis

Y_{HF} = highest farmers' yields

- Top 10th percentile of Ya

Yp = water limited potential yield

- Global Yield Gap Atlas

Silva et al. (2017) *Eur. J. Agronomy*

CIMMYT

Stochastic frontier analysis

- SFA is a **parametric** method for technical efficiency analysis that differentiates **two error terms** (random noise, *v*, and technical inefficiency, *u*).

$$\ln y_{it} = \alpha_0 + \sum_{k}^{K} \beta_k \ln x_{kit} + \frac{1}{2} \sum_{k}^{K} \sum_{j}^{K} \theta_{kj} \ln x_{kit} \times \ln x_{jit} + \delta T + \lambda T^2 + v_{it} - u_{it}$$

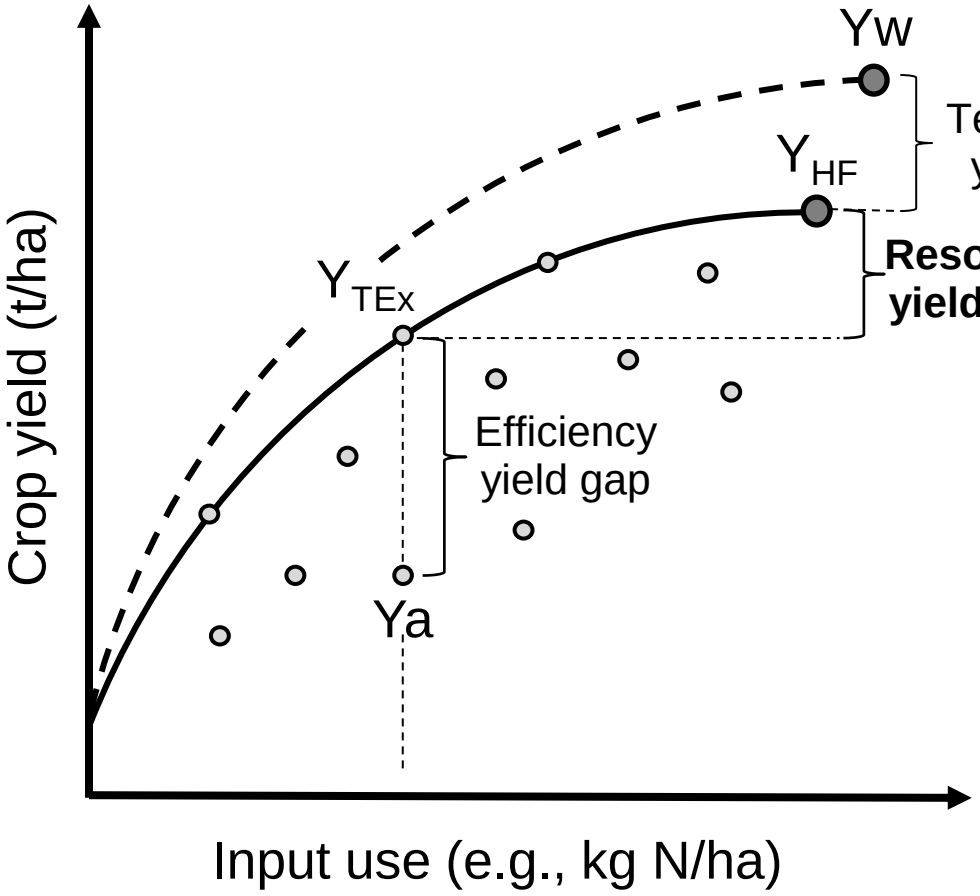
- **Technical efficiency** measures the effectiveness of converting inputs to outputs. TE = 1 implies that maximum output is produced per unit input.
- **Econometric** method developed and widely applied by economists. But also used in many other disciplines.
- Applied to **individual firms** (e.g., farms, factories, banks, etc.) to understand the scope for improving performance and efficiency of production.

Stochastic frontier analysis in R

- Start **always** with lm() and check R^2, residuals and multi-collinearity (VIF).
- sfa() function of the R package *frontier.*
- Battese & Coelli models:
  - Single-output multiple-input frontier;
  - 1992 model: Cross-section vs. panel data;
  - 1995 model: Single step estimation of frontier & inefficiency effects;
- Functional forms: Cobb-Douglas vs translog.
- Control for climate, soil, and varieties, so yield gaps are due to management only.
- Decide on variables for production frontier vs. inefficiency effects.

```
# fit cobb-douglas stochastic frontier
sfa_cd <-
  sfa(yield_tha ~
    season_year + gyga_gdd + gyga_tseas + seed_kgha + variety +
    gyga_ai + gyga_av_water + soil_depth + soil_fertility + waterlogging_yn + drought_yn + soilwatercons
    nfert_kgha + manure_yn + residues_yn + previous_crop + oxplough_freq_cat +
    herb_lha + handweeding_persdayha + weeding_yn + pesticide_yn + disease_incidence_yn + pest_incidence
    data=data_new)
```

Resource yield gap and Y_{HF}



Ya = actual farmers' yields

- Farm household surveys

Y_{TEx} = technical efficient yields

- Stochastic frontier analysis

Y_{HF} = highest farmers' yields

- Top 10th percentile of Ya

Yp = water limited potential yield

- Global Yield Gap Atlas

Silva et al. (2017) *Eur. J. Agronomy*

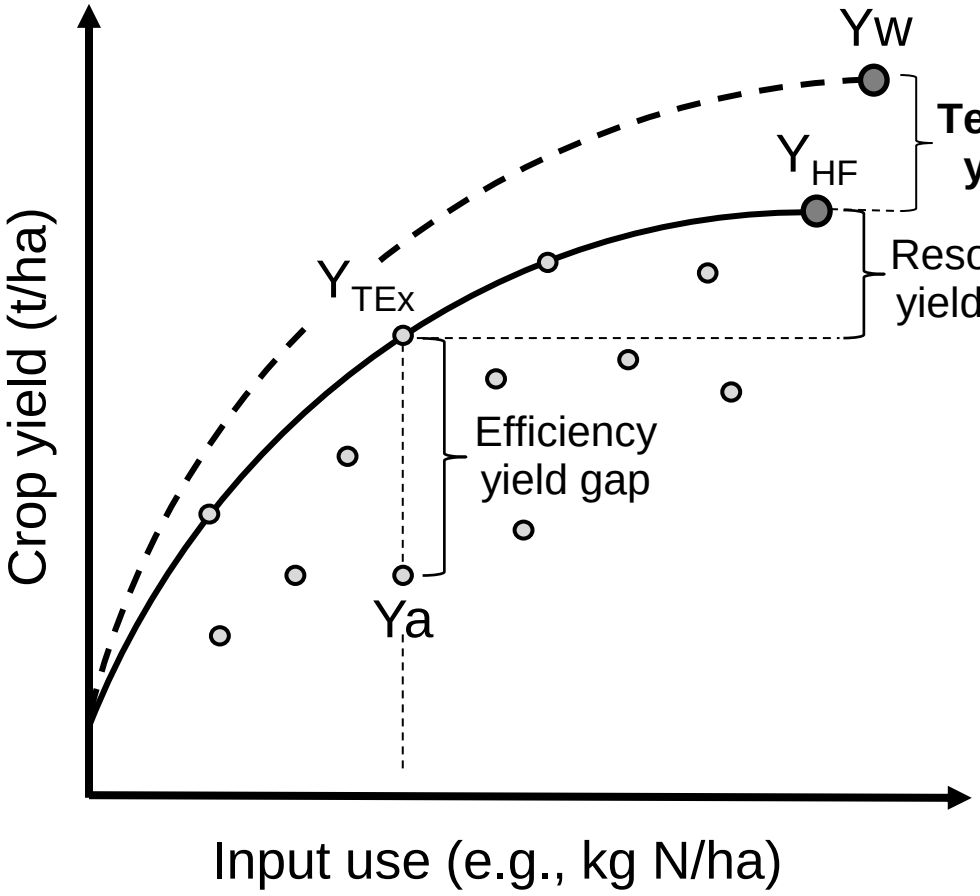
CIMMYT

Quantiles within 'for loop' in R

```
# create an empty data frame
data_final <- data.frame()
#
# create loop per year
for(yr in unique(data$year)){
  subset_year <- subset(data, year == yr)
  #
  # create loop per climate zone
  for(cz in unique(subset_year$gyga_cz)){
    subset_cz <- subset(subset_year, gyga_cz == cz)
    #
    # create loop per soil type
    for(soil in unique(subset_cz$soil_fertility)){
      subset_soil <- subset(subset_cz, soil_fertility == soil)
      # create column with field class based on yield distribution
      subset_soil$field_class <- ifelse(subset_soil$yield_tha >= quantile(subset_soil$yield_tha, 0.90),
                                        'YHF', '')
      subset_soil$field_class <- ifelse(subset_soil$yield_tha <= quantile(subset_soil$yield_tha, 0.10),
                                        'YLF', subset_soil$field_class)
      subset_soil$field_class <- ifelse(subset_soil$yield_tha > quantile(subset_soil$yield_tha, 0.10) &
                                        subset_soil$yield_tha < quantile(subset_soil$yield_tha, 0.90),
                                        'YAF', subset_soil$field_class)
      #
      # subset highest yielding fields only
      yhf <- subset(subset_soil, field_class == 'YHF')
      #
      # add column with yhf in t/ha to data frame
      subset_soil['yhf_tha'] <- mean(yhf$yield_tha, na.rm=T)
      #
      # bind all individual fields into single data frame
      data_final <- rbind(data_final, subset_soil)
}}}
```

CIMMYT

Technology yield gap and Yw



Ya = actual farmers' yields

- Farm household surveys

Y_{TEx} = technical efficient yields

- Stochastic frontier analysis

Y_{HF} = highest farmers' yields

- Top 10th percentile of Ya

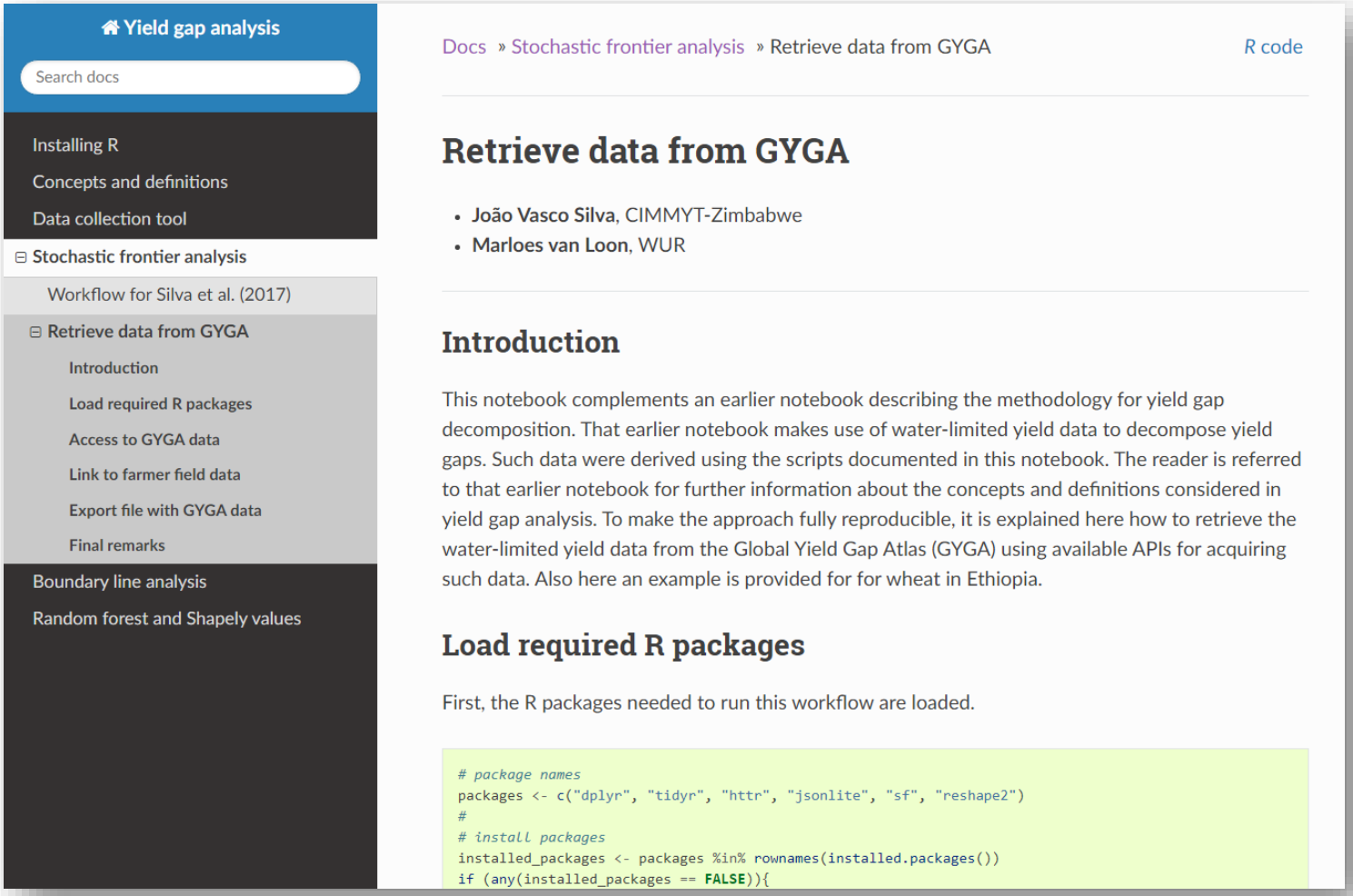
Yp = water limited potential yield

- Global Yield Gap Atlas

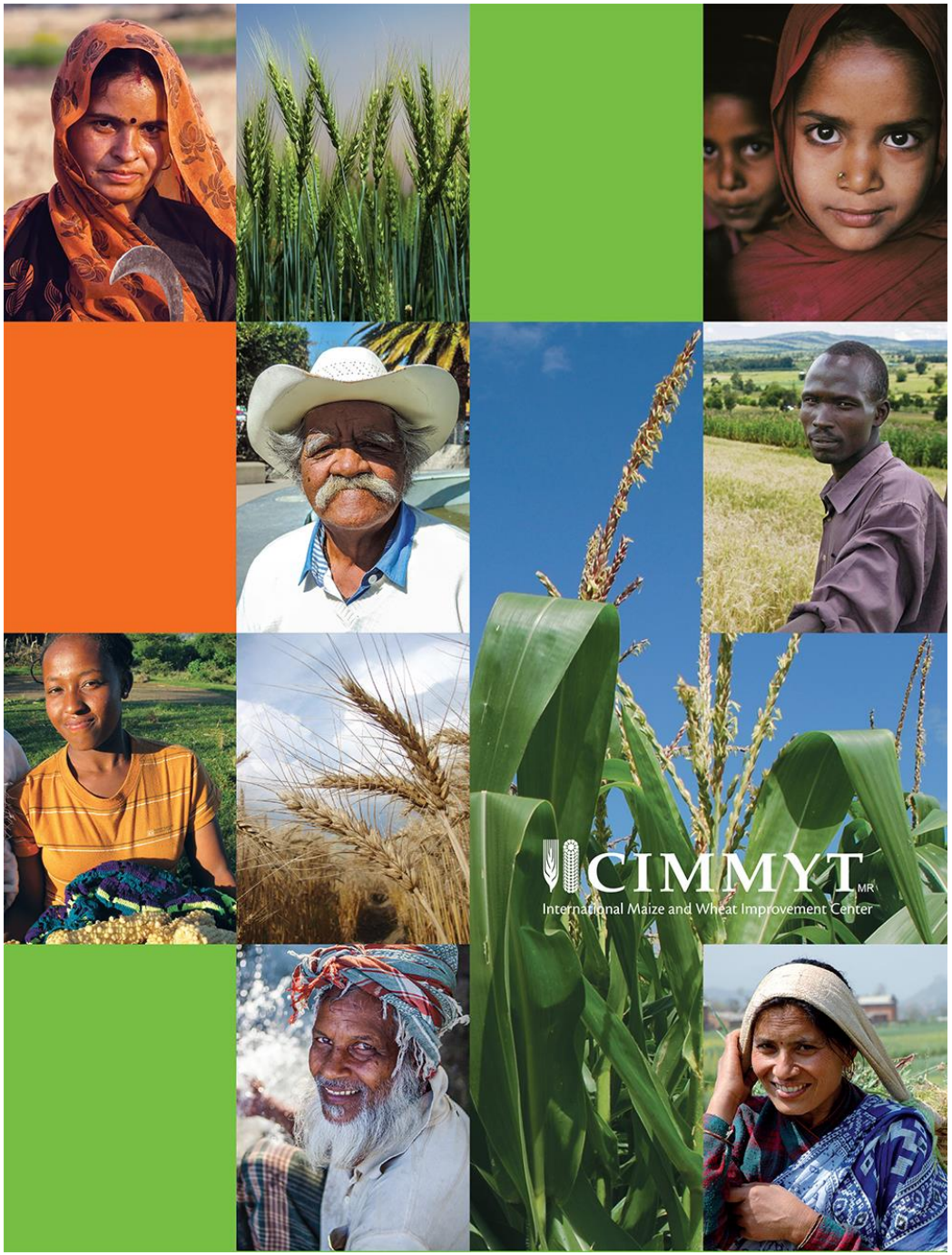
Silva et al. (2017) *Eur. J. Agronomy*

CIMMYT

Global Yield Gap Atlas in R



CIMMYT



Thank you for your interest!