sainfoin seed imaging power analysis

Bo Meyering

2023-08-22

Experimental Set-up

Implementation

Analysis

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(data.table)
```

```
##
## Attaching package: 'data.table'
##
## The following objects are masked from 'package:lubridate':
##
##     hour, isoweek, mday, minute, month, quarter, second, wday, week,
##     yday, year
##
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
##
## The following object is masked from 'package:purrr':
##
##     transpose
```

```r
library(readxl)

# names = read_xlsx("../data/power_analysis/SainfoinDepoddingExp.xlsx",
#                   sheet = "Sheet1",
#                   col_types = c("text",
#                                 "numeric",
#                                 "text",
#                                 "numeric",
#                                 "text",
#                                 "numeric",
#                                 "numeric",
#                                 "skip"),
#                   na = "NA")

names = read_excel("../data/power_analysis/SainfoinDepoddingExp.xlsx",
                   sheet = "Sheet1",
                   col_types = c("text",
                                 "numeric",
                                 "text",
                                 "numeric",
                                 "text",
                                 "numeric",
                                 "numeric",
                                 "skip"),
                   na = "NA")

raw_data <- fread("../data/power_analysis/results.csv")

clean_names <- names
clean_names$QR_rand[clean_names$QR_rand=='4.6E+043'] <- '46e43'
clean_names$QR_rand[clean_names$QR_rand=='7.2E+068'] <- '72e68'

clean_data <- raw_data %>%
        separate(img_id, into = c(NA, NA, NA, 'img_id'), sep = '/') %>%
        mutate(across(img_id, \(x) str_replace(x, '.jpg', '')),
               across(img_id, \(x) str_replace(x, '.JPG', '')))
clean_data$img_id[clean_data$img_id=='4.6e+43'] <- '46e43'
clean_data$img_id[clean_data$img_id=='7.2e+68'] <- '72e68'

combined <- clean_names %>%
        full_join(clean_data, by = c("QR_rand"="img_id")) %>%
        rename('img_id' = 'QR_rand',
               'variety' = 'Variety',
               'sample_mass' = 'Sample Mass',
               'method' = 'Method',
               'rep' = 'Rep')

comb_sum <- combined %>%
        group_by(variety, sample_mass, method, rep, img_id, class) %>%
        summarise(count = n()) %>%
        spread(key=class, value=count) %>%
        mutate(split = replace_na(split, 0)) %>%
        replace_na(replace = list(0)) %>%
```

```
        mutate(sum = rowSums(across(pod:split)),
               p_pod = pod/sum,
               p_seed = seed/sum,
               p_split = split/sum) %>%
        ungroup()
```

`summarise()` has grouped output by 'variety', 'sample_mass', 'method', 'rep',
'img_id'. You can override using the `.groups` argument.

```
estimates = comb_sum %>%
        gather(key = 'class', value = 'prop', p_pod:p_split) %>%
        group_by(variety, sample_mass, method, class) %>%
        summarise(est = mean(prop),
                  est_sd = sd(prop),
                  cv = est_sd/est) %>%
        arrange(variety, method, sample_mass)
```
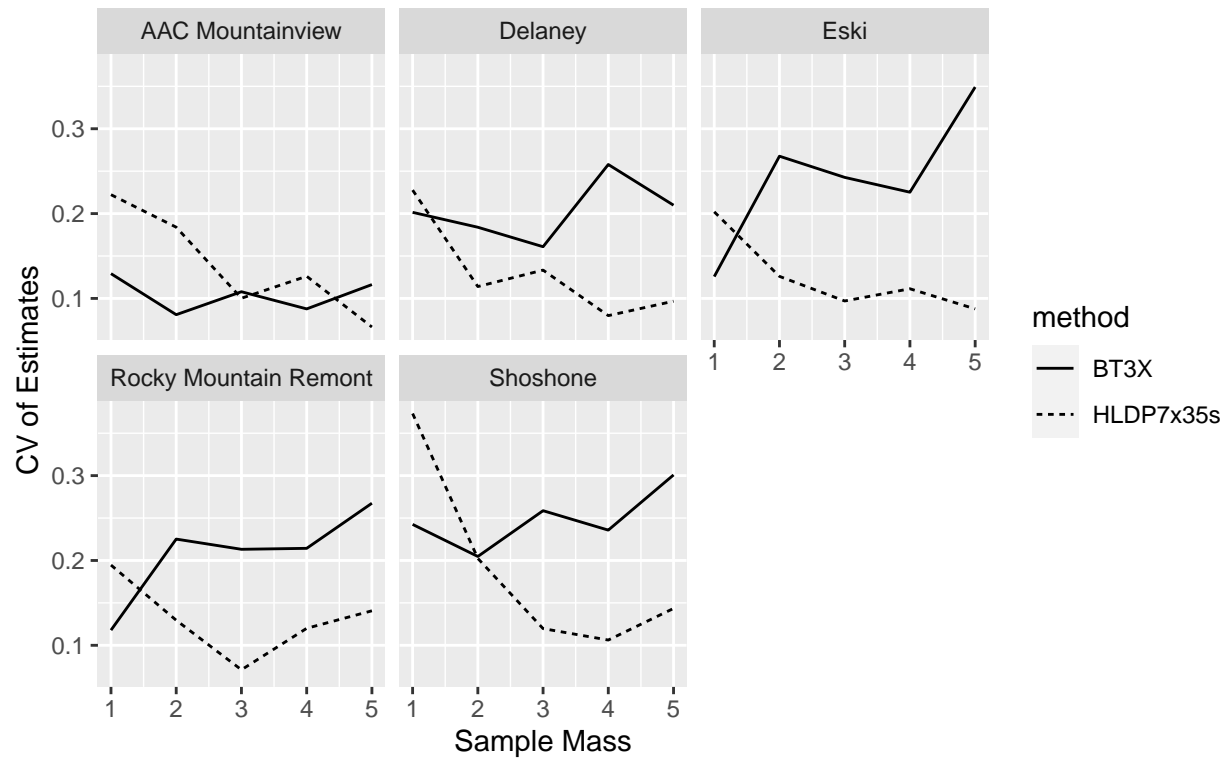
`summarise()` has grouped output by 'variety', 'sample_mass', 'method'. You can
override using the `.groups` argument.

```
seed_estimates <- estimates %>%
        filter(class=='p_seed')

pod_estimates <- estimates %>%
        filter(class=='p_pod')

split_estimates <- estimates %>%
        filter(class=='p_split')
```
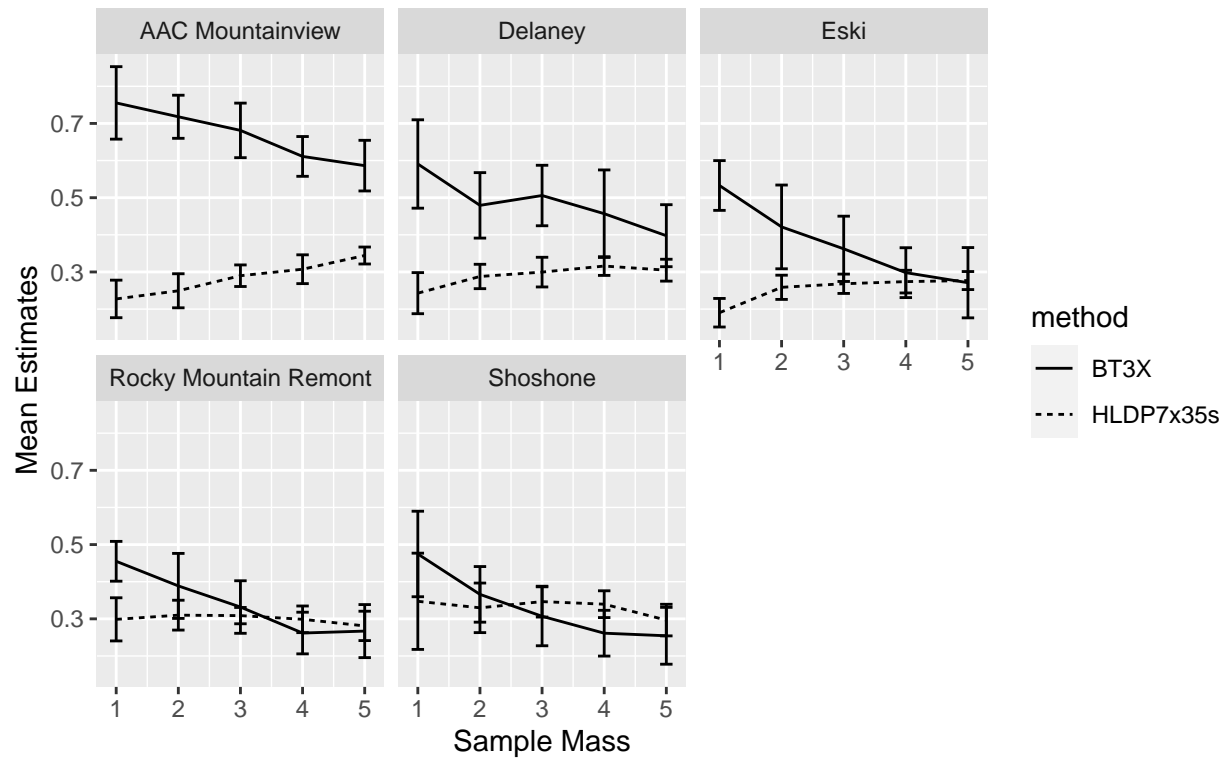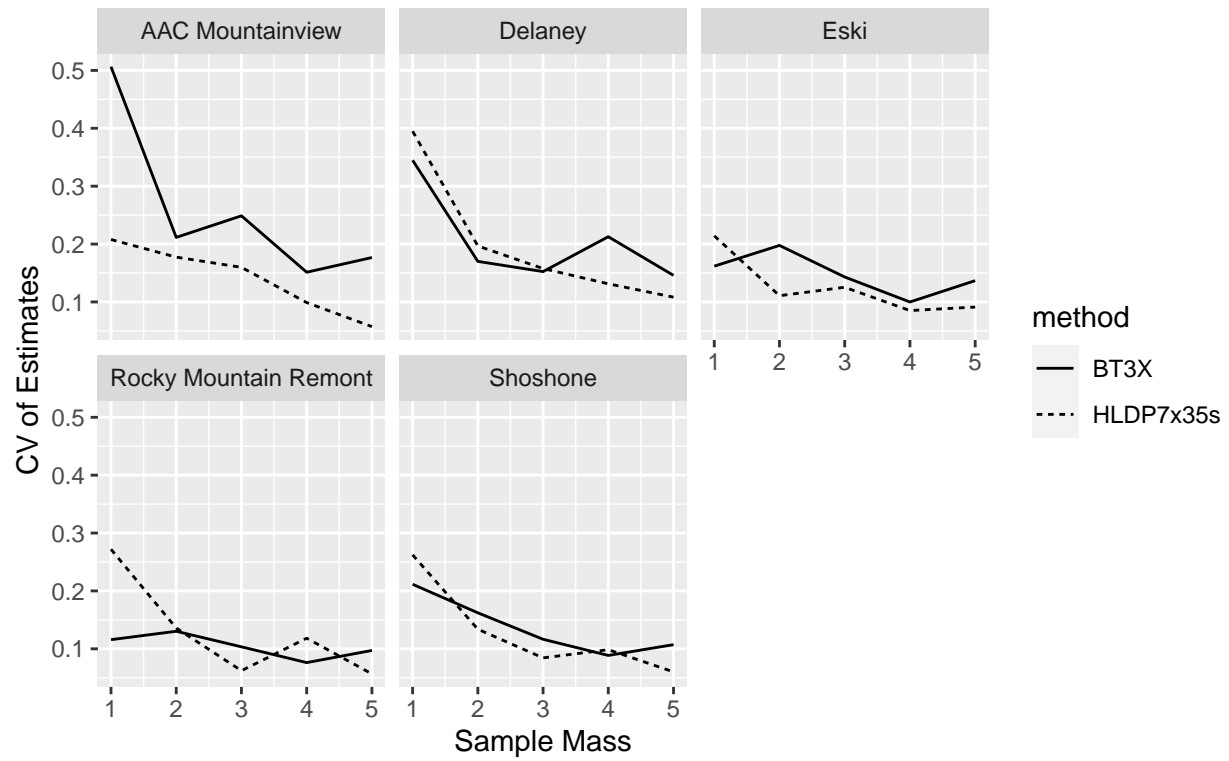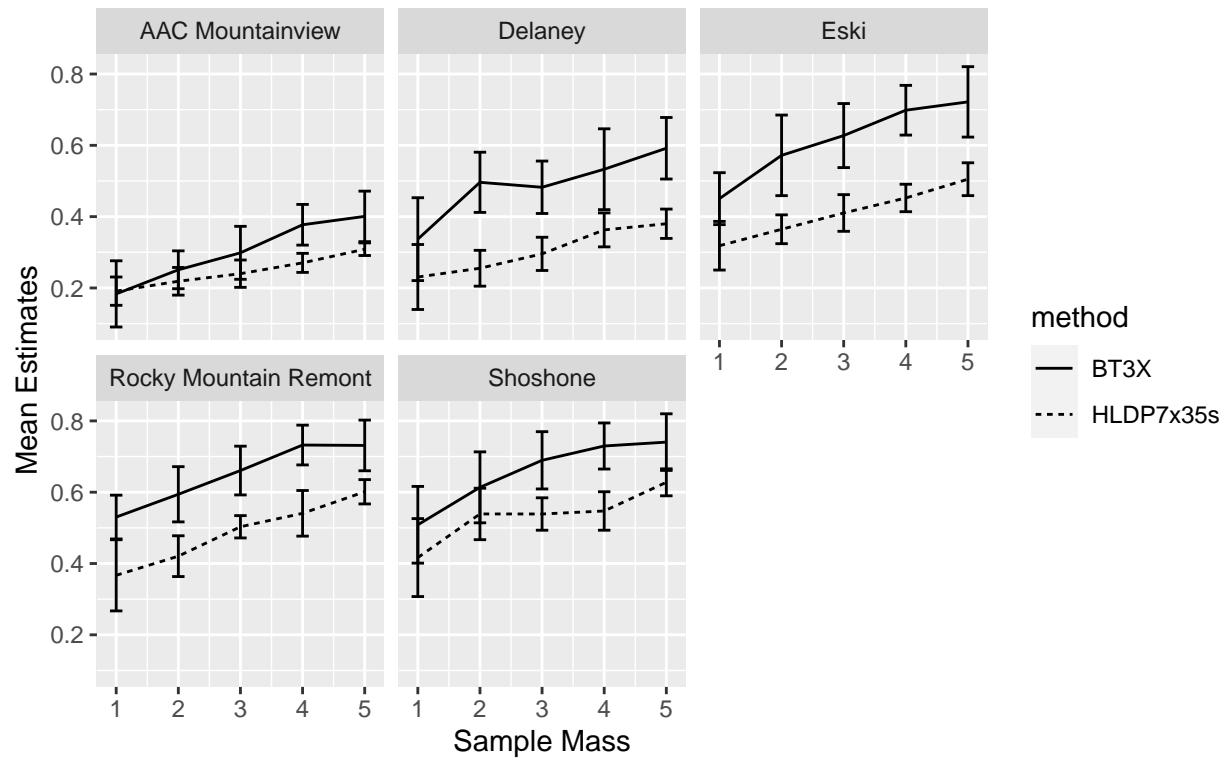
Seed Proportions
Coefficient of Variation by Variety
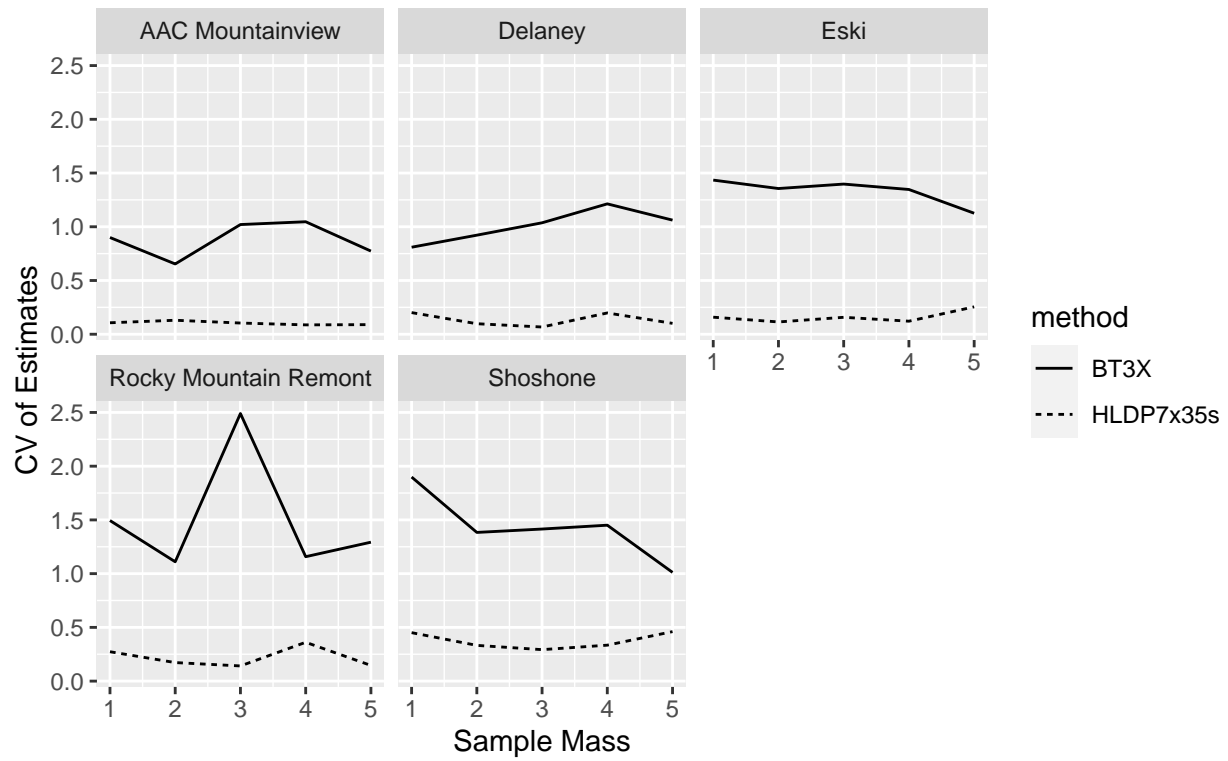
Seed Proportions
Estimates and SD by Variety

Pod Proportions
Coefficient of Variation by Variety

Pod Proportions
Estimates and SD by Variety

Split Proportions
Coefficient of Variation by Variety

Split Proportions
Estimates and SD by Variety