Leveraging multi-harvest data for simultaneously selecting high-yield and witches' broom-resistant *Theobroma grandiflorum* hybrids

Saulo Fabrício da Silva Chaves   Luiz Antônio dos Santos Dias
Rodrigo Silva Alves   Jeniffer Santana Pinto Coelho Evangelista
Kaio Olimpio das Graças Dias

Contents

Required packages

```
require(asreml)
require(tidyverse)
require(patchwork)
```

Loading the data

```
data = read.csv("data.csv", header = T, sep=';')
```

Setting the factors

```
data = transform(data,
                 Harvests = factor(Harvests),
                 Plots = factor(Plots),
                 Replicates = factor(Replicates),
                 Hybrids = factor(Hybrids))
```

Linear Mixed Models

Homoscedastic model for Fruit Yield

```
fy1 = asreml(fixed = yd ~ Harvests + Replicates:Harvests,
             random = ~ Hybrids + Hybrids:Harvests + Plots,
             data = data)

sum.fy1 = summary(fy1)$varcomp[,1:2]

sum.fy1$CI = sum.fy1$std.error * 1.96

aic.fy1 = summary(fy1)$aic

predfy1_vcov = predict(fy1, classify = "Hybrids", vcov = T)
predfy1_sed = predict(fy1, classify = "Hybrids", sed = T)

PEV = mean(diag(predfy1_vcov$vcov))
MVdelta = mean((predfy1_sed$sed^2)[upper.tri(predfy1_sed$sed^2,diag = F)])

acc1 = sqrt(1-(PEV/sum.fy1[1,1]))
her1 = 1-(MVdelta/(2*sum.fy1[1,1]))
```

	Value	Std. Error	Conf. Int.
σ_g^2	14.45	5.56	10.90
σ_p^2	35.08	4.24	8.31
σ_{gh}^2	12.02	1.99	3.90
σ_e^2	71.08	2.47	4.84
r	0.82	NA	NA
H_g^2	0.68	NA	NA

Heteroscedastic model for Fruit Yield

```r
fy2 = asreml(fixed = yd ~ Harvests + Replicates:Harvests,
             random = ~ Hybrids + Hybrids:Harvests + Plots,
             residual = ~dsum(~id(units)|Harvests),
             data = data)

sum.fy2 = summary(fy2)$varcomp[,1:2]

sum.fy2$CI = sum.fy2$std.error * 1.96

aic.fy2 = summary(fy2)$aic

predfy2_vcov = predict(fy2, classify = "Hybrids", vcov = T)
predfy2_sed = predict(fy2, classify = "Hybrids", sed = T)

PEV = mean(diag(predfy2_vcov$vcov))
MVdelta = mean((predfy2_sed$sed^2)[upper.tri(predfy2_sed$sed^2,diag = F)])

acc2 = sqrt(1-(PEV/sum.fy2[1,1]))
her2 = 1-(MVdelta/(2*sum.fy2[1,1]))

# Harvest-wise heritability

her2j = NULL
for(i in 1:nlevels(data$Harvests)){
  predfy2_sed = predict(fy2, classify = "Hybrids:Harvests",
                        level=list(Harvests = i), sed = T)

  MVdelta = mean((predfy2_sed$sed^2)[upper.tri(predfy2_sed$sed^2,diag = F)])

  her2j[i] = 1-(MVdelta/(2*sum.fy2[1,1]))
}

blups.fy = coef(fy2)$random[grep("Hybrids", rownames(coef(fy2)$random))]
```

	Value	Std. Error	Conf. Int.
σ_g^2	13.88	4.93	9.67
σ_p^2	26.62	3.23	6.33
σ_{gh}^2	7.92	1.56	3.05
$\sigma_{e_1}^2$	28.98	3.43	6.72
$\sigma_{e_2}^2$	28.41	3.36	6.59
$\sigma_{e_3}^2$	40.46	4.43	8.69
$\sigma_{e_4}^2$	49.32	5.27	10.33
$\sigma_{e_5}^2$	72.59	7.46	14.62
$\sigma_{e_6}^2$	98.50	10.09	19.79
$\sigma_{e_7}^2$	167.26	16.52	32.38
$\sigma_{e_8}^2$	101.13	10.43	20.44
$\sigma_{e_9}^2$	100.17	10.05	19.69
r	0.85	0.85	0.85
$H_{g_1}^2$	0.57	0.57	0.57
$H_{g_2}^2$	0.57	0.57	0.57
$H_{g_3}^2$	0.52	0.52	0.52
$H_{g_4}^2$	0.49	0.49	0.49
$H_{g_5}^2$	0.43	0.43	0.43
$H_{g_6}^2$	0.39	0.39	0.39
$H_{g_7}^2$	0.31	0.31	0.31
$H_{g_8}^2$	0.38	0.38	0.38
$H_{g_9}^2$	0.38	0.38	0.38

Coefficients of determination

```
sig2f = NULL
cgh = NULL
cp = NULL
ce = NULL

for (i in 1:nlevels(data$Harvests)) {
  sig2f[i] = sum(sum.fy2[1:3,1]) + sum.fy2[i+3,1]
  cgh[i] = sum.fy2["Hybrids:Harvests", "component"] / sig2f[i]
  cp[i] = sum.fy2["Plots", "component"] / sig2f[i]
  ce[i] = sum.fy2[i+3,1] / sig2f[i]
}

coef.det = data.frame(
  "Harvests" = levels(data$Harvests),
  "sig2f" = sig2f,
  "chg" = cgh,
  "cp" = cp,
```

```r
  "ce" = ce
)
```

Harvests	σ_f^2	c_{gh}^2	c_p^2	c_e^2
Yr1	77.40	0.10	0.34	0.37
Yr2	76.83	0.10	0.35	0.37
Yr3	88.88	0.09	0.30	0.46
Yr4	97.74	0.08	0.27	0.50
Yr5	121.01	0.07	0.22	0.60
Yr6	146.92	0.05	0.18	0.67
Yr7	215.68	0.04	0.12	0.78
Yr8	149.56	0.05	0.18	0.68
Yr9	148.59	0.05	0.18	0.67

Generalized linear mixed model for witches' broom resistance

```r
wb = asreml(fixed = wb ~ Harvests + Replicates:Harvests,
            random = ~ Hybrids + Hybrids:Harvests + Plots,
            family=asr_binomial(link="logit"),
            data = data, maxit = 100)

sum.wb = summary(wb)$varcomp[,1:2]

sum.wb$CI = sum.wb$std.error * 1.96

predwb = predict(wb, classify = "Hybrids", sed = T)

MVdelta = mean((predwb$sed^2)[upper.tri(predwb$sed^2,diag = F)])

herwb = 1-(MVdelta/(2*sum.wb[1,1]))

blups.wb = coef(wb)$random[1:nlevels(data$Hybrids)]
```

Optimum number of harvests

```r
rho_fy = NULL
for (i in 1:nlevels(data$Harvests)) {
  rho_fy[i] = sum(sum.fy2["Hybrids","component"],sum.fy2["Plots","component"])/
    sig2f[i]
}

rho_fy = mean(rho_fy) #Fruit yield repeatability
```

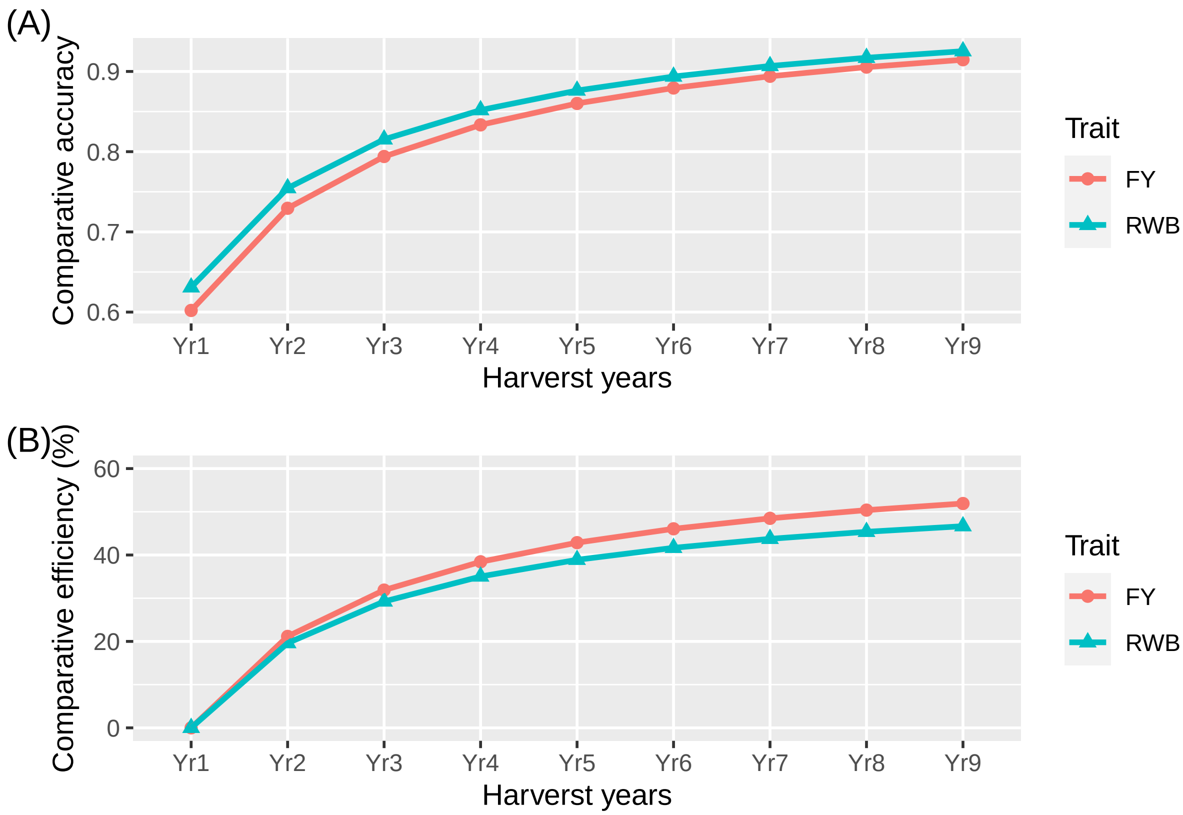
```r
rho_wb = sum(sum.wb["Hybrids","component"],sum.wb["Plots","component"])/
    sum(c(sum.wb[1:3,1],3.29)) #WB resistance repeatability


effh_fy = NULL
acch_fy = NULL
effh_wb = NULL
acch_wb = NULL
for (i in 1:nlevels(data$Harvests)) {
  effh_fy[i] = sqrt(i/(1+(i-1)*rho_fy))
  effh_wb[i] = sqrt(i/(1+(i-1)*rho_wb))
  acch_fy[i] = sqrt((i*rho_fy)/(i*rho_fy+1-rho_fy))
  acch_wb[i] = sqrt((i*rho_wb)/(i*rho_wb+1-rho_wb))
}

a = data.frame(
  "Harvests" = rep(levels(data$Harvests),2),
  "Trait" = rep(c("FY", "RWB"),each = nlevels(data$Harvests)),
  "Accuracy" = c(acch_fy,acch_wb)
) %>% ggplot()+
  geom_point(aes(x = Harvests, y = Accuracy, color = Trait, shape = Trait),
             size=2)+
  geom_line(aes(x = Harvests, y = Accuracy, color = Trait, group = Trait),
            size=1)+
  labs(y = "Comparative accuracy",x = "Harverst years",tag = "(A)")


b = data.frame(
  "Harvests" = rep(levels(data$Harvests),2),
  "Trait" = rep(c("FY", "RWB"),each = nlevels(data$Harvests)),
  "Efficiency" = c((effh_fy-1)*100,(effh_wb-1)*100)
) %>% ggplot()+
  geom_point(aes(x = Harvests, y = Efficiency, color = Trait, shape = Trait),
             size=2)+
  geom_line(aes(x = Harvests, y = Efficiency, color = Trait, group = Trait),
            size=1)+ ylim(0,60)+
  labs(y = "Comparative efficiency (%)", x = "Harverst years",tag = "(B)")

a/b
```



Selection index

Obtaining the yield persistence

```
blups.fy = left_join(data.frame(
  'Hybrids' = rep(levels(data$Hybrids),each = nlevels(data$Harvests)),
  'Harvests' = rep(levels(data$Harvests),nlevels(data$Hybrids)),
  'blup' = blups.fy[(nlevels(data$Hybrids)+1):length(blups.fy)]
), data.frame(
  'Hybrids' = levels(data$Hybrids),
  'blup' = blups.fy[1:nlevels(data$Hybrids)]
), by='Hybrids') %>% mutate(
  BLUP = blup.x + blup.y
) %>% select(Hybrids, Harvests, BLUP)


max.fy = blups.fy %>% group_by(Harvests) %>%
  summarise(max = max(BLUP))

num = NULL
```

```r
for (i in levels(data$Hybrids)) {

  dttes = blups.fy %>% filter(Hybrids == i) %>%
    select(Hybrids,Harvests,BLUP)

  num[i] = 1/sum((dttes$BLUP - max.fy$max)^2)

}

den = sum(num)^2

Per.fy = num/den
Per.fy = rownames_to_column(as.data.frame(Per.fy),'Hybrids')
```

Additive index

```r
blups.fy = blups.fy %>% group_by(Hybrids) %>%
              summarise(BLUP = mean(BLUP))

AI = cbind(Per.fy,blups.fy$BLUP,blups.wb)

colnames(AI)[2:4] = c("PERS.FY","BLUP.FY","BLUP.RWB")

AI = AI %>% mutate(
  AI = (BLUP.FY/sqrt(sum.fy2["Hybrids","component"])) -
    (BLUP.RWB/sqrt(sum.wb["Hybrids","component"]))
)
```

Hybrids	PERS.FY	BLUP.FY	BLUP.RWB	AI
H125	2.55	6.94	-0.64	2.39
H165	0.51	5.54	-0.68	2.05
H143	0.09	1.72	-1.17	1.43
H117	0.26	4.50	-0.09	1.28
H137	0.04	-0.76	-1.73	1.23
H140	0.26	4.48	0.05	1.16
H167	0.08	1.68	-0.85	1.15
H124	0.05	-0.37	-1.30	0.98
H129	0.11	2.55	-0.17	0.83
H127	0.08	1.43	-0.43	0.74
H149	0.05	-0.19	-0.89	0.69
H135	0.04	-1.60	-1.31	0.65
H157	0.04	-0.50	-0.89	0.60
H131	0.05	-0.05	-0.50	0.40
H166	0.04	-0.71	-0.65	0.35
H118	0.03	-2.09	-0.94	0.22
H130	0.06	0.39	-0.08	0.17
H132	0.03	-2.18	-0.65	-0.05
H162	0.15	3.46	1.46	-0.28
H169	0.28	4.57	1.86	-0.32
H120	0.12	2.53	1.23	-0.34
H134	0.03	-3.36	-0.65	-0.36
H133	0.03	-3.39	-0.52	-0.48
H136	0.02	-4.03	-0.68	-0.51
H144	0.08	1.39	1.10	-0.54
H163	0.08	1.44	1.38	-0.76
H138	0.05	-0.33	1.10	-1.00
H152	0.04	-1.71	0.91	-1.21
H150	0.04	-1.16	1.30	-1.39
H121	0.04	-1.60	1.29	-1.50
H161	0.06	0.40	1.96	-1.51
H128	0.02	-6.38	-0.20	-1.54
H123	0.01	-7.96	-0.07	-2.08
H126	0.02	-4.67	1.45	-2.46