# Fitting genotype by environment models in sommer

Giovanny Covarrubias-Pazaran

2022-09-02

The sommer package was developed to provide R users a powerful and reliable multivariate mixed model solver. The package is focused on problems of the type p > n (more effects to estimate than observations) and its core algorithm is coded in C++ using the Armadillo library. This package allows the user to fit mixed models with the advantage of specifying the variance-covariance structure for the random effects, specifying heterogeneous variances, and obtaining other parameters such as BLUPs, BLUEs, residuals, fitted values, variances for fixed and random effects, etc.

The purpose of this vignette is to show how to fit different genotype by environment (GxE) models using the sommer package:

1) Single environment model
2) Multienvironment model: Main effect model
3) Multienvironment model: Diagonal model (DG)
4) Multienvironment model: Compund symmetry model (CS)
5) Multienvironment model: Unstructured model (US)
6) Multienvironment model: Random regression model (RR)
7) Multienvironment model: Other covariance structures for GxE
8) Finlay-Wilkinson regression
9) Two stage analysis

When the breeder decides to run a trial and apply selection in a single environment (whether because the amount of seed is a limitation or there's no availability for a location) the breeder takes the risk of selecting material for a target population of environments (TPEs) using an environment that is not representative of the larger TPE. Therefore, many breeding programs try to base their selection decision on multi-environment trial (MET) data. Models could be adjusted by adding additional information like spatial information, experimental design information, etc. In this tutorial we will focus mainly on the covariance structures for GxE and the incorporation of relationship matrices for the genotype effect.

## 1) Single environment model

A single-environment model is the one that is fitted when the breeding program can only afford one location, leaving out the possible information available from other environments. This will be used to further expand to GxE models.

```
library(sommer)
data(DT_example)
DT <- DT_example
A <- A_example

ansSingle <- mmer(Yield~1,
              random= ~ vsr(Name, Gu=A),
              rcov= ~ units,
```

```r
             data=DT, verbose = FALSE)
summary(ansSingle)
```

```
## ============================================================
##          Multivariate Linear Mixed Model fit by REML
## **********************  sommer 4.1  **********************
## ============================================================
##          logLik      AIC      BIC Method Converge
## Value -78.80875 159.6175 162.8378     NR     TRUE
## ============================================================
## Variance-Covariance components:
##                    VarComp VarCompSE Zratio Constraint
## u:Name.Yield-Yield   6.529     2.202  2.965   Positive
## units.Yield-Yield   13.868     1.633  8.494   Positive
## ============================================================
## Fixed effects:
##   Trait      Effect Estimate Std.Error t.value
## 1 Yield (Intercept)    11.74    0.4876   24.07
## ============================================================
## Groups and observations:
##        Yield
## u:Name    41
## ============================================================
## Use the '$' sign to access results and parameters
```

```r
# or
Ai <- as(solve(A), Class="sparseMatrix")
ansSingle <- mmec(Yield~1,
              random= ~ vsc(isc(Name), Gu=Ai),
              rcov= ~ units,
              data=DT, verbose = FALSE)
summary(ansSingle)
```

```
## ============================================================
##          Multivariate Linear Mixed Model fit by REML
## **********************  sommer 4.1  **********************
## ============================================================
##          logLik      AIC      BIC Method Converge
## Value -359.0031 720.0062 723.2265     AI     TRUE
## ============================================================
## Variance-Covariance components:
##                 VarComp VarCompSE Zratio Constraint
## Name:Ai:isc:isc   6.497     1.479  4.392   Positive
## units:isc:isc    13.868     1.798  7.711   Positive
## ============================================================
## Fixed effects:
##             Estimate Std.Error t.value
## (Intercept)    11.74    0.4862   24.14
## ============================================================
## Use the '$' sign to access results and parameters
```

In this model, the only term to be estimated is the one for the germplasm (here called `Name`). For the sake of example we have added a relationship matrix among the levels of the random effect `Name`. This is just a diagonal matrix with as many rows and columns as levels present in the random effect `Name`, but any other non-diagonal relationship matrix could be used.

## 2) MET: main effect model

A multi-environment model is the one that is fitted when the breeding program can afford more than one location. The main effect model assumes that GxE doesn't exist and that the main genotype effect plus the fixed effect for environment is enough to predict the genotype effect in all locations of interest.

```
ansMain <- mmer(Yield~Env,
              random= ~ vsr(Name, Gu=A),
              rcov= ~ units,
              data=DT, verbose = FALSE)
summary(ansMain)
```

```
## ============================================================
##          Multivariate Linear Mixed Model fit by REML
## **********************  sommer 4.1  **********************
## ============================================================
##          logLik      AIC      BIC Method Converge
## Value -32.59421 71.18842 80.84949     NR     TRUE
## ============================================================
## Variance-Covariance components:
##                      VarComp VarCompSE Zratio Constraint
## u:Name.Yield-Yield     4.856    1.5233  3.188   Positive
## units.Yield-Yield      8.109    0.9615  8.434   Positive
## ============================================================
## Fixed effects:
##   Trait      Effect Estimate Std.Error t.value
## 1 Yield (Intercept)   16.385    0.5849  28.012
## 2 Yield  EnvCA.2012   -5.688    0.5741  -9.908
## 3 Yield  EnvCA.2013   -6.218    0.6107 -10.182
## ============================================================
## Groups and observations:
##        Yield
## u:Name    41
## ============================================================
## Use the '$' sign to access results and parameters
```

```
# or

Ai <- as(solve(A), Class="sparseMatrix")
ansMain <- mmec(Yield~Env,
              random= ~ vsc(isc(Name), Gu=Ai),
              rcov= ~ units,
              data=DT, verbose = FALSE)
summary(ansMain)
```

```
## ============================================================
##          Multivariate Linear Mixed Model fit by REML
## **********************  sommer 4.1  **********************
## ============================================================
##          logLik      AIC      BIC Method Converge
## Value -313.3005 632.6011 642.2621     AI     TRUE
## ============================================================
## Variance-Covariance components:
##                   VarComp VarCompSE Zratio Constraint
## Name:Ai:isc:isc   4.852     1.449  3.349   Positive
```

```
## units:isc:isc     8.109     1.807  4.487   Positive
## ============================================================
## Fixed effects:
##           Estimate Std.Error t.value
## Intercept   16.385    0.5847  28.021
## CA.2012     -5.688    0.5740  -9.909
## CA.2013     -6.219    0.6107 -10.183
## ============================================================
## Use the '$' sign to access results and parameters
```

## 3) MET: diagonal model (DG)

A multi-environment model is the one that is fitted when the breeding program can afford more than one location. The diagonal model assumes that GxE exists and that the genotype variation is expressed differently at each location, therefore fitting a variance component for the genotype effect at each location. The main drawback is that this model assumes no covariance among locations, as if genotypes were independent (despite the fact that is the same genotypes). The fixed effect for environment plus the location-specific BLUP is used to predict the genotype effect in each locations of interest.

```
ansDG <- mmer(Yield~Env,
              random= ~ vsr(dsr(Env),Name, Gu=A),
              rcov= ~ units,
              data=DT, verbose = FALSE)
summary(ansDG)
```

```
## ============================================================
##          Multivariate Linear Mixed Model fit by REML
## **********************  sommer 4.1  **********************
## ============================================================
##          logLik      AIC      BIC Method Converge
## Value -21.04157 48.08315 57.74421     NR     TRUE
## ============================================================
## Variance-Covariance components:
##                           VarComp VarCompSE Zratio Constraint
## CA.2011:Name.Yield-Yield  17.493    6.1099  2.863   Positive
## CA.2012:Name.Yield-Yield   5.337    1.7662  3.022   Positive
## CA.2013:Name.Yield-Yield   7.884    2.5526  3.089   Positive
## units.Yield-Yield          4.381    0.6493  6.747   Positive
## ============================================================
## Fixed effects:
##   Trait      Effect Estimate Std.Error t.value
## 1 Yield (Intercept)   16.621     0.948  17.532
## 2 Yield  EnvCA.2012   -5.958     1.045  -5.699
## 3 Yield  EnvCA.2013   -6.662     1.098  -6.067
## ============================================================
## Groups and observations:
##              Yield
## CA.2011:Name    41
## CA.2012:Name    41
## CA.2013:Name    41
## ============================================================
## Use the '$' sign to access results and parameters
```

```r
# or
Ai <- as(solve(A), Class="sparseMatrix")
ansDG <- mmec(Yield~Env,
              random= ~ vsc(dsc(Env),isc(Name), Gu=Ai),
              rcov= ~ units,
              data=DT, verbose = FALSE)
summary(ansDG)
```

```
## ============================================================
##           Multivariate Linear Mixed Model fit by REML
## **********************  sommer 4.1  **********************
## ============================================================
##          logLik      AIC     BIC Method Converge
## Value -301.9224 609.8449 619.506     AI     TRUE
## ============================================================
## Variance-Covariance components:
##                            VarComp VarCompSE Zratio Constraint
## Env:Name:Ai:CA.2011:CA.2011  15.792     3.405  4.638   Positive
## Env:Name:Ai:CA.2012:CA.2012   5.192     2.846  1.825   Positive
## Env:Name:Ai:CA.2013:CA.2013   7.717     2.947  2.618   Positive
## units:isc:isc                 4.744     2.362  2.009   Positive
## ============================================================
## Fixed effects:
##           Estimate Std.Error t.value
## Intercept   16.622     0.911  18.245
## CA.2012     -5.962     1.012  -5.888
## CA.2013     -6.664     1.067  -6.249
## ============================================================
## Use the '$' sign to access results and parameters
```

## 4) MET: compund symmetry model (CS)

A multi-environment model is the one that is fitted when the breeding program can afford more than one location. The compound symmetry model assumes that GxE exists and that a main genotype variance-covariance component is expressed across all location. In addition, it assumes that a main genotype-by-environment variance is expressed across all locations. The main drawback is that the model assumes the same variance and covariance among locations. The fixed effect for environment plus the main effect for BLUP plus genotype-by-environment effect is used to predict the genotype effect in each location of interest.

```r
E <- diag(length(unique(DT$Env)))
rownames(E) <- colnames(E) <- unique(DT$Env)
EA <- kronecker(E,A, make.dimnames = TRUE)
ansCS <- mmer(Yield~Env,
              random= ~ vsr(Name, Gu=A) + vsr(Env:Name, Gu=EA),
              rcov= ~ units,
              data=DT, verbose = FALSE)
summary(ansCS)
```

```
## ===========================================================
##          Multivariate Linear Mixed Model fit by REML
## **********************  sommer 4.1  **********************
## ===========================================================
##          logLik      AIC      BIC Method Converge
## Value -20.14538 46.29075 55.95182     NR     TRUE
```

```
## ============================================================
## Variance-Covariance components:
##                          VarComp VarCompSE Zratio Constraint
## u:Name.Yield-Yield         3.682     1.691  2.177   Positive
## u:Env:Name.Yield-Yield     5.173     1.495  3.460   Positive
## units.Yield-Yield          4.366     0.647  6.748   Positive
## ============================================================
## Fixed effects:
##   Trait      Effect Estimate Std.Error t.value
## 1 Yield (Intercept)   16.496    0.6855  24.065
## 2 Yield  EnvCA.2012   -5.777    0.7558  -7.643
## 3 Yield  EnvCA.2013   -6.380    0.7960  -8.015
## ============================================================
## Groups and observations:
##            Yield
## u:Name        41
## u:Env:Name   123
## ============================================================
## Use the '$' sign to access results and parameters
```

```
## or
E <- diag(length(unique(DT$Env)));rownames(E) <- colnames(E) <- unique(DT$Env)
Ei <- solve(E)
Ai <- solve(A)
EAi <- kronecker(Ei,Ai, make.dimnames = TRUE)
Ei <- as(Ei, Class="sparseMatrix")
Ai <- as(Ai, Class="sparseMatrix")
EAi <- as(EAi, Class="sparseMatrix")
ansCS <- mmec(Yield~Env,
              random= ~ vsc(isc(Name), Gu=Ai) + vsc(isc(Env:Name), Gu=EAi),
              rcov= ~ units,
              data=DT, verbose = FALSE)
summary(ansCS)
```

```
## ============================================================
##          Multivariate Linear Mixed Model fit by REML
## **********************  sommer 4.1  **********************
## ============================================================
##          logLik      AIC      BIC Method Converge
## Value -300.8632 607.7264 617.3875     AI     TRUE
## ============================================================
## Variance-Covariance components:
##                        VarComp VarCompSE Zratio Constraint
## Name:Ai:isc:isc          3.703     1.884  1.966   Positive
## Env:Name:EAi:isc:isc     5.132     2.428  2.114   Positive
## units:isc:isc            4.466     2.295  1.946   Positive
## ============================================================
## Fixed effects:
##             Estimate Std.Error t.value
## Intercept     16.496    0.6863  24.035
## CA.2012       -5.777    0.7564  -7.637
## CA.2013       -6.380    0.7967  -8.008
## ============================================================
## Use the '$' sign to access results and parameters
```

## 5) MET: unstructured model (US)

A multi-environment model is the one that is fitted when the breeding program can afford more than one location. The unstructured model is the most flexible model assuming that GxE exists and that an environment-specific variance exists in addition to as many covariances for each environment-to-environment combinations. The main drawback is that is difficult to make this models converge because of the large number of variance components, the fact that some of these variance or covariance components are zero, and the difficulty in choosing good starting values. The fixed effect for environment plus the environment specific BLUP (adjusted by covariances) is used to predict the genotype effect in each location of interest.

```
ansUS <- mmer(Yield~Env,
              random= ~ vsr(usr(Env),Name, Gu=A),
              rcov= ~ units,
              data=DT, verbose = FALSE)
summary(ansUS)
```

```
## ===================================================================
##             Multivariate Linear Mixed Model fit by REML
## *************************  sommer 4.1  *************************
## ===================================================================
##          logLik      AIC      BIC Method Converge
## Value -14.20951 34.41901 44.08008     NR     TRUE
## ===================================================================
## Variance-Covariance components:
##                                  VarComp VarCompSE Zratio Constraint
## CA.2011:Name.Yield-Yield          15.994     5.381  2.972   Positive
## CA.2012:CA.2011:Name.Yield-Yield   6.172     2.503  2.465   Unconstr
## CA.2012:Name.Yield-Yield           5.273     1.750  3.013   Positive
## CA.2013:CA.2011:Name.Yield-Yield   6.366     3.069  2.074   Unconstr
## CA.2013:CA.2012:Name.Yield-Yield   0.376     1.535  0.245   Unconstr
## CA.2013:Name.Yield-Yield           7.689     2.490  3.088   Positive
## units.Yield-Yield                  4.386     0.650  6.748   Positive
## ===================================================================
## Fixed effects:
##   Trait      Effect Estimate Std.Error t.value
## 1 Yield (Intercept)   16.341    0.8141  20.072
## 2 Yield  EnvCA.2012   -5.696    0.7406  -7.692
## 3 Yield  EnvCA.2013   -6.286    0.8202  -7.664
## ===================================================================
## Groups and observations:
##                     Yield
## CA.2011:Name           41
## CA.2012:CA.2011:Name    82
## CA.2012:Name           41
## CA.2013:CA.2011:Name    82
## CA.2013:CA.2012:Name    82
## CA.2013:Name           41
## ===================================================================
## Use the '$' sign to access results and parameters
```

```
# adjust variance BLUPs by adding covariances
# ansUS$U[1:6] <- unsBLUP(ansUS$U[1:6])

# or
Ai <- solve(A)
```

```r
Ai <- as(Ai, Class="sparseMatrix")
ansUS <- mmec(Yield~Env,
              random= ~ vsc(usc(Env),isc(Name), Gu=Ai),
              rcov= ~ units,
              data=DT, verbose = FALSE)
summary(ansUS)
```

```
## ============================================================
##           Multivariate Linear Mixed Model fit by REML
## **********************  sommer 4.1  **********************
## ============================================================
##          logLik      AIC      BIC Method Converge
## Value -302.6944 611.3888 621.0499     AI     TRUE
## ============================================================
## Variance-Covariance components:
##                             VarComp VarCompSE Zratio Constraint
## Env:Name:Ai:CA.2011:CA.2011  14.002     3.305 4.2366   Positive
## Env:Name:Ai:CA.2011:CA.2012   4.951     1.785 2.7728   Unconstr
## Env:Name:Ai:CA.2012:CA.2012   4.394     2.202 1.9956   Positive
## Env:Name:Ai:CA.2011:CA.2013   6.145     2.427 2.5315   Unconstr
## Env:Name:Ai:CA.2012:CA.2013   0.604     1.759 0.3433   Unconstr
## Env:Name:Ai:CA.2013:CA.2013   7.946     2.774 2.8646   Positive
## units:isc:isc                 2.663     1.770 1.5040   Positive
## ============================================================
## Fixed effects:
##           Estimate Std.Error t.value
## Intercept   16.341    0.7432  21.987
## CA.2012     -5.685    0.6740  -8.435
## CA.2013     -6.265    0.7418  -8.446
## ============================================================
## Use the '$' sign to access results and parameters
```

## 6) MET: random regression model

A multi-environment model is the one that is fitted when the breeding program can afford more than one location. The random regression model assumes that the environment can be seen as a continuous variable and therefore a variance component for the intercept and a variance component for the slope can be fitted. The number of variance components will depend on the order of the Legendre polynomial fitted.

```r
library(orthopolynom)
DT$EnvN <- as.numeric(as.factor(DT$Env))
ansRR <- mmer(Yield~Env,
              random= ~ vsr(leg(EnvN,1),Name),
              rcov= ~ units,
              data=DT, verbose = FALSE)
summary(ansRR)
```

```
## ===========================================================
##          Multivariate Linear Mixed Model fit by REML
## **********************  sommer 4.1  **********************
## ===========================================================
##          logLik      AIC      BIC Method Converge
## Value -27.70318 61.40636 71.06743     NR     TRUE
## ===========================================================
```

```
## Variance-Covariance components:
##                         VarComp VarCompSE Zratio Constraint
## leg0:Name.Yield-Yield  10.392    3.1473  3.302   Positive
## leg1:Name.Yield-Yield   2.079    0.9792  2.123   Positive
## units.Yield-Yield       6.297    0.8442  7.459   Positive
## ============================================================
## Fixed effects:
##   Trait      Effect Estimate Std.Error t.value
## 1 Yield (Intercept)   16.541    0.6770  24.432
## 2 Yield  EnvCA.2012   -5.832    0.6425  -9.078
## 3 Yield  EnvCA.2013   -6.472    0.8239  -7.854
## ============================================================
## Groups and observations:
##           Yield
## leg0:Name    41
## leg1:Name    41
## ============================================================
## Use the '$' sign to access results and parameters
```

```r
# or

ansRR <- mmec(Yield~Env,
              random= ~ vsc(dsc(leg(EnvN,1)),isc(Name)),
              rcov= ~ units,
              data=DT, verbose = FALSE)
summary(ansRR)
```

```
## ============================================================
##          Multivariate Linear Mixed Model fit by REML
## **********************  sommer 4.1  **********************
## ============================================================
##          logLik      AIC      BIC Method Converge
## Value -308.4122 622.8243 632.4854     AI     TRUE
## ============================================================
## Variance-Covariance components:
##                       VarComp VarCompSE Zratio Constraint
## EnvN:Name:leg0:leg0  10.276     2.313  4.442   Positive
## EnvN:Name:leg1:leg1   2.086     1.756  1.188   Positive
## units:isc:isc         6.430     1.998  3.218   Positive
## ============================================================
## Fixed effects:
##           Estimate Std.Error t.value
## Intercept   16.542    0.6745  24.523
## CA.2012     -5.833    0.6405  -9.107
## CA.2013     -6.473    0.8217  -7.877
## ============================================================
## Use the '$' sign to access results and parameters
```

In addition, an unstructured, diagonal or other variance-covariance structure can be put on top of the polynomial model:

```r
library(orthopolynom)
DT$EnvN <- as.numeric(as.factor(DT$Env))
ansRR <- mmer(Yield~Env,
              random= ~ vsr(usr(leg(EnvN,1)),Name),
              rcov= ~ units,
```

```r
              data=DT, verbose = FALSE)
summary(ansRR)
```

```
## ============================================================
##          Multivariate Linear Mixed Model fit by REML
## **********************  sommer 4.1  **********************
## ============================================================
##          logLik      AIC      BIC Method Converge
## Value -25.56967 57.13935 66.80042     NR     TRUE
## ============================================================
## Variance-Covariance components:
##                            VarComp VarCompSE Zratio Constraint
## leg0:Name.Yield-Yield       10.791    3.2745  3.295   Positive
## leg1:leg0:Name.Yield-Yield  -2.428    1.3699 -1.772   Unconstr
## leg1:Name.Yield-Yield        2.286    1.0404  2.197   Positive
## units.Yield-Yield            6.260    0.8421  7.434   Positive
## ============================================================
## Fixed effects:
##   Trait      Effect Estimate Std.Error t.value
## 1 Yield (Intercept)   16.501    0.7778  21.216
## 2 Yield  EnvCA.2012   -5.791    0.6704  -8.638
## 3 Yield  EnvCA.2013   -6.476    0.8554  -7.570
## ============================================================
## Groups and observations:
##                 Yield
## leg0:Name          41
## leg1:leg0:Name     82
## leg1:Name          41
## ============================================================
## Use the '$' sign to access results and parameters
```

```r
# or

ansRR <- mmec(Yield~Env,
              random= ~ vsc(usc(leg(EnvN,1)),isc(Name)),
              rcov= ~ units,
              data=DT, verbose = FALSE)
summary(ansRR)
```

```
## ============================================================
##          Multivariate Linear Mixed Model fit by REML
## **********************  sommer 4.1  **********************
## ============================================================
##          logLik     AIC      BIC Method Converge
## Value -309.2045 624.409 634.0701     AI     TRUE
## ============================================================
## Variance-Covariance components:
##                     VarComp VarCompSE Zratio Constraint
## EnvN:Name:leg0:leg0 10.5273     2.262 4.6532   Positive
## EnvN:Name:leg0:leg1  0.1493     1.468 0.1017   Unconstr
## EnvN:Name:leg1:leg1  2.0889     1.954 1.0691   Positive
## units:isc:isc        7.2282     2.126 3.4003   Positive
## ============================================================
## Fixed effects:
```

```
##           Estimate Std.Error t.value
## Intercept   16.541    0.6916  23.916
## CA.2012     -5.836    0.6707  -8.703
## CA.2013     -6.471    0.8498  -7.614
## ============================================================
## Use the '$' sign to access results and parameters
```

## 7) Other GxE covariance structures

Although not very commonly used in GxE models, the autoregressive of order 1 (AR1) and other covariance structures could be used in the GxE modeling. Here we show how to do it (not recommending it).

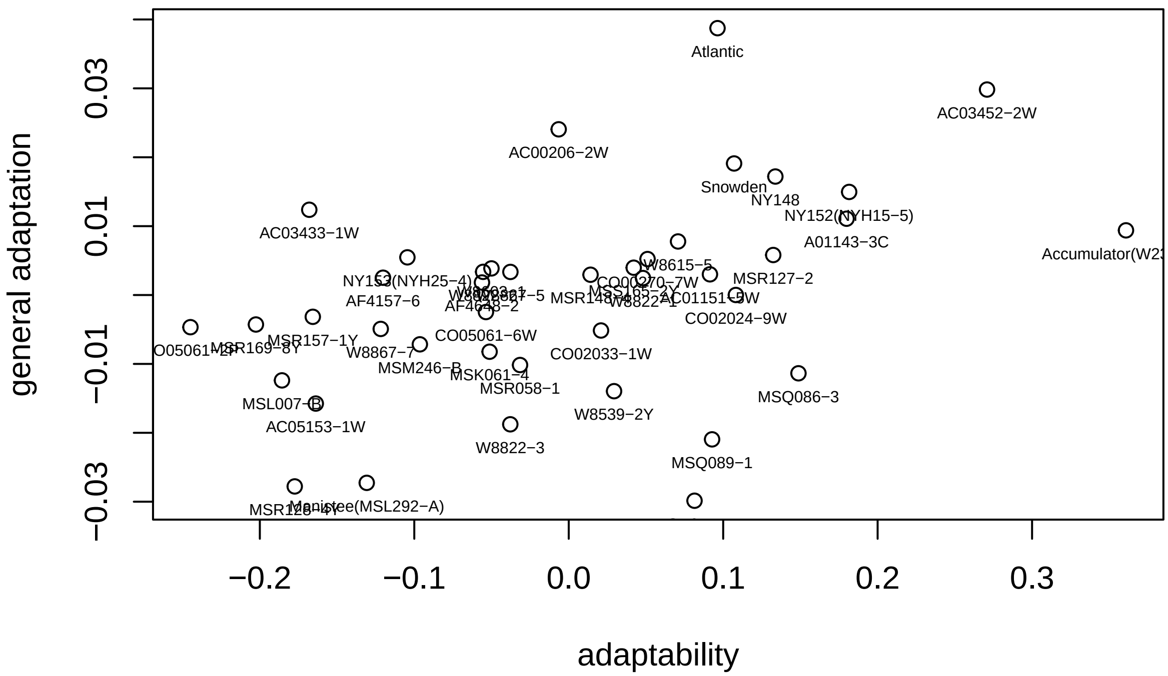
```
E <- AR1(DT$Env) # can be AR1() or CS(), etc.
rownames(E) <- colnames(E) <- unique(DT$Env)
EA <- kronecker(E,A, make.dimnames = TRUE)
ansCS <- mmer(Yield~Env,
              random= ~ vsr(Name, Gu=A) + vsr(Env:Name, Gu=EA),
              rcov= ~ units,
              data=DT, verbose = FALSE)
summary(ansCS)
```

```
## ============================================================
##          Multivariate Linear Mixed Model fit by REML
## **********************  sommer 4.1  **********************
## ============================================================
##          logLik      AIC     BIC Method Converge
## Value -19.39067 44.78134 54.4424     NR     TRUE
## ============================================================
## Variance-Covariance components:
##                         VarComp VarCompSE Zratio Constraint
## u:Name.Yield-Yield        2.225    1.7536  1.269   Positive
## u:Env:Name.Yield-Yield    6.424    1.8293  3.512   Positive
## units.Yield-Yield         4.334    0.6418  6.752   Positive
## ============================================================
## Fixed effects:
##   Trait      Effect Estimate Std.Error t.value
## 1 Yield (Intercept)   16.484    0.6735  24.474
## 2 Yield  EnvCA.2012   -5.780    0.7365  -7.848
## 3 Yield  EnvCA.2013   -6.372    0.7799  -8.170
## ============================================================
## Groups and observations:
##           Yield
## u:Name       41
## u:Env:Name  123
## ============================================================
## Use the '$' sign to access results and parameters
```

## 8) Finlay-Wilkinson regression

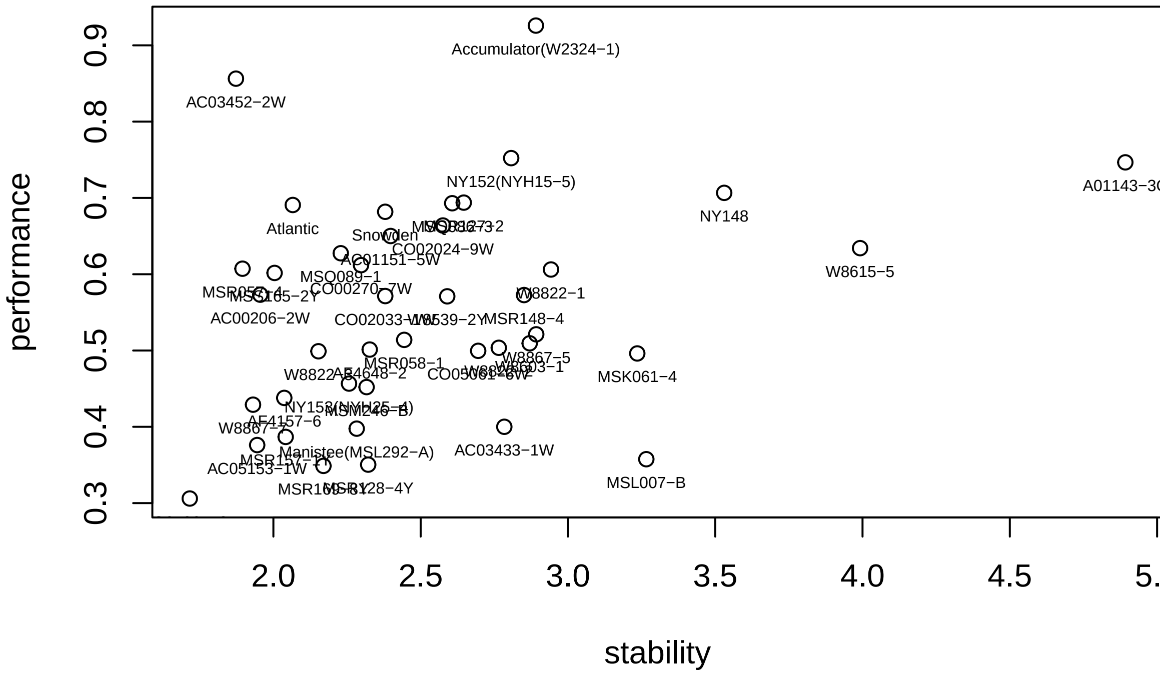
```
data(DT_h2)
DT <- DT_h2
```

```r
## build the environmental index
ei <- aggregate(y~Env, data=DT,FUN=mean)
colnames(ei)[2] <- "envIndex"
ei <- ei[with(ei, order(envIndex)), ]

## add the environmental index to the original dataset
DT2 <- merge(DT,ei, by="Env")

# numeric by factor variables like envIndex:Name can't be used in the random part like this
# they need to come with the vsc() structure
DT2 <- DT2[with(DT2, order(Name)), ]
mix2 <- mmec(y~ envIndex,
             random=~ Name + vsc(dsc(envIndex),isc(Name)), data=DT2,
             rcov=~vsc(dsc(Name),isc(units)),
             tolParConvNorm = .0001,
             nIters = 50, verbose = FALSE
)
# summary(mix2)$varcomp

b=mix2$uList$`vsc(dsc(envIndex), isc(Name))` # adaptability (b) or genotype slopes
mu=mix2$uList$`vsc( isc( Name ) )` # general adaptation (mu) or main effect
e=sqrt(summary(mix2)$varcomp[-c(1:2),1]) # error variance for each individual

## general adaptation (main effect) vs adaptability (response to better environments)
plot(mu[,1]~b[,1], ylab="general adaptation", xlab="adaptability")
text(y=mu[,1],x=b[,1], labels = rownames(mu), cex=0.5, pos = 1)
```




```r
## prediction across environments
Dt <- mix2$Dtable
Dt[1,"average"]=TRUE
Dt[2,"include"]=TRUE
Dt[3,"include"]=TRUE
```

```r
pp <- predict(mix2,Dtable = Dt, D="Name")
preds <- pp$pvals
# preds[with(preds, order(-predicted.value)), ]
## performance vs stability (deviation from regression line)
plot(preds[,2]~e, ylab="performance", xlab="stability")
text(y=preds[,2],x=e, labels = rownames(mu), cex=0.5, pos = 1)
```




## 9) Two stage analysis

It is common then to fit a first model that accounts for the variation of random design elements, e.g., locations, years, blocks, and fixed genotype effects to obtain the estimated marginal means (EMMs) or best linear unbiased estimators (BLUEs) as adjusted entry means. These adjusted entry means are then used as the phenotype or response variable in GWAS and genomic prediction studies.

```r
##########
## stage 1
## use mmer for dense field trials
##########
data(DT_h2)
DT <- DT_h2
head(DT)
```

```
##                 Name     Env Loc Year      Block y
## 1             W8822-3 FL.2012  FL 2012 FL.2012.1 2
## 2             W8867-7 FL.2012  FL 2012 FL.2012.2 2
## 3            MSL007-B MO.2011  MO 2011 MO.2011.1 3
## 4          CO00270-7W FL.2012  FL 2012 FL.2012.2 3
## 5 Manistee(MSL292-A) FL.2013  FL 2013 FL.2013.2 3
## 6            MSM246-B FL.2012  FL 2012 FL.2012.2 3
```

```r
envs <- unique(DT$Env)
BLUEL <- list()
```

```r
XtXL <- list()
for(i in 1:length(envs)){
  ans1 <- mmer(y~Name-1,
                random=~Block,
                verbose=FALSE,
                data=droplevels(DT[which(DT$Env == envs[i]),]
               )
  )
  ans1$Beta$Env <- envs[i]

  BLUEL[[i]] <- ans1$Beta
  XtXL[[i]] <- ans1$VarBeta
}

DT2 <- do.call(rbind, BLUEL)
OM <- do.call(adiag1,XtXL)

##########
## stage 2
## use mmec for sparse equation
##########
m <- matrix(1/var(DT2$Estimate, na.rm = TRUE))
ans2 <- mmec(Estimate~Env,
             random=~Effect + Env:Effect,
             rcov=~vsc(isc(units,thetaC = matrix(3), theta = m)),
             W=OM,
             verbose=FALSE,
             data=DT2
             )
```

```
## Using the weights matrix
```

```r
summary(ans2)$varcomp
```

```
##                       VarComp VarCompSE     Zratio Constraint
## Effect:isc:isc       2.463610 0.2275963 10.8244720   Positive
## Env:Effect:isc:isc 5.754576 0.6055289  9.5033881   Positive
## units:m:             1.000000 1.8910611  0.5288036      Fixed
```

## Literature


Covarrubias-Pazaran G. 2016. Genome assisted prediction of quantitative traits using the R package sommer. PLoS ONE 11(6):1-15.

Covarrubias-Pazaran G. 2018. Software update: Moving the R package sommer to multivariate mixed models for genome-assisted prediction. doi: https://doi.org/10.1101/354639

Bernardo Rex. 2010. Breeding for quantitative traits in plants. Second edition. Stemma Press. 390 pp.

Gilmour et al. 1995. Average Information REML: An efficient algorithm for variance parameter estimation in linear mixed models. Biometrics 51(4):1440-1450.

Henderson C.R. 1975. Best Linear Unbiased Estimation and Prediction under a Selection Model. Biometrics vol. 31(2):423-447.

Kang et al. 2008. Efficient control of population structure in model organism association mapping. Genetics 178:1709-1723.

Lee, D.-J., Durban, M., and Eilers, P.H.C. (2013). Efficient two-dimensional smoothing with P-spline ANOVA mixed models and nested bases. Computational Statistics and Data Analysis, 61, 22 - 37.

Lee et al. 2015. MTG2: An efficient algorithm for multivariate linear mixed model analysis based on genomic information. Cold Spring Harbor. doi: http://dx.doi.org/10.1101/027201.

Maier et al. 2015. Joint analysis of psychiatric disorders increases accuracy of risk prediction for schizophrenia, bipolar disorder, and major depressive disorder. Am J Hum Genet; 96(2):283-294.

Rodriguez-Alvarez, Maria Xose, et al. Correcting for spatial heterogeneity in plant breeding experiments with P-splines. Spatial Statistics 23 (2018): 52-71.

Searle. 1993. Applying the EM algorithm to calculating ML and REML estimates of variance components. Paper invited for the 1993 American Statistical Association Meeting, San Francisco.

Yu et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Genetics 38:203-208.

Tunnicliffe W. 1989. On the use of marginal likelihood in time series model estimation. JRSS 51(1):15-27.