Linear Mixed Models with `st.lmm` in SelectionTools

Matthias Frisch

1 Overview **3**
1.1 Interface . . . 3
1.2 Specifying the model . . . 4
1.3 Specifying variance structures . . . 6
1.4 Multivariate models . . . 7
1.5 Model selection . . . 8
1.6 Tests of model parameters . . . 9
1.7 Least-squares means and reference grids . . . 10
1.8 Test of hypotheses on least-squares means . . . 11
1.9 Best linear unbiased prediction of random effects . . . 13
1.10 Results, printing and serialization . . . 14
1.11 Estimation procedures . . . 16

2 Cross classified and hierarchical Designs **17**
2.1 Moment estimators . . . 17
2.2 Hierarchical effects: Regions and villages . . . 26
2.3 Multi-environment trials: Rapeseed . . . 26
2.4 Crossed random effects: Years and locations . . . 28
2.5 Separate genetic variances for material groups . . . 29
2.6 Multi-environment trials: Balanced and unbalanced . . . 30
2.7 Inference spaces for machines and operators . . . 37

3 Experimental Designs **38**
3.1 Moment estimators . . . 38
3.2 Completely randomized design: Fertilizer contrasts . . . 46
3.3 Randomized complete blocks: Cotton . . . 50
3.4 Latin square: Sugar beet . . . 52
3.5 Factorial randomized blocks: Cowpea . . . 54
3.6 Incomplete blocks: Soybean lattice . . . 56
3.7 Split plots: Oats . . . 57
3.8 Nested incomplete blocks: Barley lattice . . . 60
3.9 Row and column blocks: Barley lattice . . . 61
3.10 Separable autoregressive plot residuals . . . 63
3.11 Incomplete blocks: Rapeseed . . . 64
3.12 Multi-environment trials: Barley . . . 68

4 Heterogeneous Residual Variances **70**
4.1 Maize genotype comparisons across and within counties . . . 70
4.2 Genetic and environmental variance components in soybean . . . 73
4.3 Adjusted genotype means in an unbalanced barley trial . . . 74

5 Heritability **76**
5.1 Balanced barley trial across years and locations . . . 76
5.2 Barley with missing location–year combinations . . . 78

6 Multivariate Experiments **79**
6.1 Moment estimators . . . 79
6.2 Adjusted genotype means for three traits . . . 84
6.3 Genetic variances and correlations in maize . . . 85
6.4 Three traits with heterogeneous residuals . . . 86
6.5 Unbalanced trial with seven traits and heterogeneous environments in barley . . . 88

7 Regression **90**
7.1 Moment estimators . . . 90
7.2 Random regression: Chicken growth . . . 97

8 Diallel analysis **106**
8.1 Half-diallel example . . . 106

9 Genomic covariances of random genotype effects in a RCBD **108**
9.1 Tropical maize: BLUE, BLUP and genomic BLUP . . . 108

10 Analysis of a series of trials **112**
10.1 Tropical maize: A series across environments . . . 112
10.2 Barley: Stagewise analysis across years and locations . . . 114

11 Genome wide association studies **116**
11.1 Rapeseed marker and flowering-time data . . . 116
11.2 Kinship-adjusted GWAS with P3D . . . 117
11.3 Re-estimating variance components for every marker . . . 117
11.4 Kinship and principal-component adjustment . . . 118
11.5 Resampled genome-wide significance threshold . . . 120
11.6 Regional association . . . 122

12 Genomic prediction **123**
12.1 Two-step genomic prediction with one dataset containing the unobserved phenotypic values of the prediction set as NA . . . 123
12.2 Two-step genomic prediction with one training dataset and one prediction dataset . . . 125
12.3 Genomic prediction with a multi-environment trial . . . 127
12.4 Hybrid prediction: Flint and dent maize . . . 129
12.5 GBLUP with SelectionTools marker data sets . . . 132
12.6 Hybrid prediction with SelectionTools marker data sets . . . 135
12.7 Genomic prediction with simulated DH pools . . . 140

13 References **144**
13.1 Experimental designs and analysis of variance . . . 144
13.2 Linear mixed models and estimation . . . 145
13.3 Statistical inference and model assessment . . . 146
13.4 Least-squares means, contrasts and multiple comparisons . . . 147
13.5 Plant breeding, association studies and genomic prediction . . . 148

1 Overview

1.1 Interface

The fitting function `st.lmm` accepts a fixed-effects formula, random-effects formulas, a residual covariance formula and model data. Optional arguments specify known covariance matrices, the estimation method and the inference settings.

The following example uses the rapeseed data distributed with the package. Genotypes, environments and their interaction are fixed, replicate effects are random within environments. The residual variance is common to all observations. The detailed analysis appears in Section 2.3. The returned object `m.rapeseed` stores the fitted model and its results.

```
> library(SelectionTools)
> data("v-lmm-rapeseed", package = "SelectionTools")
> m.rapeseed <- st.lmm(
+     fixed = hgt ~ geno * env,
+     random = ~env:rep,
+     data = rapeseed
+ )
> m.rapeseed
```

The function `st.lsm` calculates Least-squares means. The following call returns genotype means averaged over the environments in the rapeseed model (Searle et al., 1980).

```
> st.lsm(m.rapeseed, "geno")
```

In the tropical maize example genoype and block are random factors, only the intercept is fixed. From the marker data of the maize lines a genomic relationship matrix is calculated, which is used to model the covariance between the random genotype effects using `ucov`. The genomic analysis appears in Section 9.1.3.

```
> data("v-lmm-tropmaize", package = "SelectionTools")
> data("v-tropmaize-vcf", package = "SelectionTools")
> st.load.vcf.data(v.tropmaize.vcf, data.set = "maize.markers")
> A <- st.get.A("maize.markers")
> maize.model <- st.lmm(
+     fixed  = y ~ 1,
+     random = ~ gen + block,
+     ucov   = list(gen = A),
+     data   = tropmaize
+ )
```

The function `st.predict` returns the fixed intercept plus each predicted genotype effect, giving BLUPs of the genotypic values (Henderson, 1975).

```
> maize.predictions <- st.predict(maize.model, "gen")
> maize.predictions
```

The chicken example compares diet-specific growth slopes. The model describes chicken weight as a function of time and diet. Each diet has its own time slope, while all diets share a common fixed intercept. Each chick has a random slope deviation from its diet's slope, without a random intercept. Residual variances may differ among measurement times. The detailed chicken analyses appear in Section 7.2.

```
> data("v-lmm-chickens", package = "SelectionTools")
> chickens$weight.s <- sqrt(chickens$weight)
> chicken.model <- st.lmm(
```

```
+     fixed  = weight.s ~ Time:Diet,
+     random = ~ (0 + Time) | Chick,
+     cov    = ~ at(Time):units,
+     data   = chickens
+ )
```

The slopes for the diets are estimated with `st.lsm`.

```
> st.lsm(
+     chicken.model,
+     "Diet",
+     trend = "Time"
+ )
```

The one-sided tests assess whether each slope improvement exceeds 0.05 square-root-weight units per day. The argument `rhs` specifies this nonzero right-hand side, and `side=">"` specifies the alternative. Single-step adjustment accounts for correlations among the contrasts and controls the family-wise error rate (Hothorn et al., 2008).

```
> st.lsm(
+     chicken.model,
+     "Diet",
+     trend = "Time",
+     method = "trt.vs.ctrl",
+     ref = "d1",
+     rhs = .05,
+     side = ">",
+     adjust = "single-step"
+ )
```

The ajdusted mean weight for each diet at Time = 21 is

```
> st.lsm(
+     chicken.model,
+     "Diet",
+     at = list(Time = 21)
+ )
```

1.2 Specifying the model

Fixed effects. The fixed argument is a two-sided formula with a numeric response on the left: a single response for a univariate model, or several responses combined with `cbind()` for a multivariate model. A numeric predictor represents a regression.

These examples illustrate specifying fixed effects, using generic variable names:

```
y ~ treatment                # Treatment main effects and an intercept
y ~ treatment + site         # Additive treatment and site effects
y ~ treatment * site         # Main effects plus their interaction
y ~ treatment:site           # An interaction term
y ~ site/block               # Site plus blocks nested within site
y ~ x + I(x^2)               # Linear and quadratic regression
y ~ 0 + treatment            # Treatment columns without an intercept
y ~ .                        # Automatic fixed-predictor selection from data
```

The formula operators combine predictors and define their interactions. `a*b` expands to `a + b + a:b`; `a/b` expands to `a + a:b`. The constants `1` and `0` include and exclude the intercept. The expression `I(x^2)` is an arithmetic square. Outside `I()`, `^` specifies the highest interaction order: `(a + b + c)^2` includes the

three main effects and all two-way interactions, expanding to `a + b + c + a:b + a:c + b:c` (Wilkinson and Rogers, 1973). In a fixed formula, `.` expands to the other named predictor columns of the supplied data. For example, `mpg ~ .` includes the other columns of `mtcars`. In univariate models, additive `offset(...)` terms and the `offset` argument supply known mean contributions.

Factor coding determines how coefficients describe differences between factor levels. With the usual treatment coding, the first level is the reference. Sum coding describes deviations from an average of the level means. The choice changes coefficient interpretation; alternative codings of the same model give the same fitted values (Searle, 1987). The following argument selects sum coding for the two rapeseed factors:

```
contrasts = list(geno = "contr.sum", env = "contr.sum")
```

The list can name a contrast function or supply a coding matrix for each factor. With the usual R settings, unordered factors use treatment coding and ordered factors use polynomial coding. An explicit `contrasts` entry takes precedence over the factor's contrast attribute, followed by `options("contrasts")`. The order of the factor levels identifies the reference level for treatment coding.

Linear and quadratic regression are illustrated with fertilizer doses in Section 7.1.1. The multiple-regression example with `mtcars` in Section 7.1.4 illustrates selection among fixed predictors.

Random effects. A one-sided grouping formula adds random intercepts. The levels of each grouping term are independent unless a matching covariance matrix is supplied through `ucov`. These examples illustrate specifying random effects:

```
~gen + env + gen:env           # Genotype, environment and interaction
~block + block:wholeplot       # Blocks and whole plots within blocks
~region:village               # Villages identified within regions
~env/rep/block                # Environments, replicates and nested blocks
~(0 + Time) | Chick           # Random time slope per chick, without an intercept
~at(materialgroup):gen        # Separate genotype variances by material group
~overlay(female, male)        # Shared parental effects in both parental roles
```

The term `gen:env` has grouping levels defined by genotype by environment combinations. A shared variance component is estimated for those combinations in a univariate model. Nesting is expressed through the formula and the identifiers. For blocks nested within environments, the combination of environment and block identifies each experimental block.

The bar in `~(0 + Time) | Chick` separates the random coefficients on its left from the grouping factor on its right. Each chick has its own random time slope. The `0 +` excludes a random intercept; `~Time | Chick` includes both a random intercept and a random slope, together with their covariance. The chicken example in Section 1.1 uses the slope-only specification; further random-regression examples appear in Section 7.2.

For a univariate response, `at(materialgroup):gen` gives each material group its own genotype variance. Genotype effects in different groups are independent. The colon attaches the genotype effects to the groups selected by `at()`. The material-group example in Section 2.5 uses `grp` as the group variable.

The `overlay(female, male)` operator uses one shared set of parental effects for both parental roles. Each cross receives the sum of the effects of its female and male parents, so a parent has the same effect whichever role it occupies. The half-diallel example in Section 8.1 illustrates this specification (Sprague and Tatum, 1942).

Nested random factors, as required for incomplete blocks within replicates in a lattice design, are illustrated with barley in Section 3.8. Its `RowBlk` identifiers already distinguish blocks within replicates. Correlated random genotype effects specified through a genomic relationship matrix are illustrated with tropical maize in Section 9.1.3.

Omitting `random`, or setting it to `NULL`, gives a fixed-only Gaussian model. A list combines grouping and random-regression formulas. This example combines block effects with subject-specific random intercepts and time slopes:

```
random = list(~block, ~Time | subject)
```

The list combines these terms in one fitted model. Fixed genotype effects give adjusted genotype means; random genotype effects give BLUPs under the specified genotype covariance. The tropical-maize example in Section 1.1 illustrates genotype prediction with related random effects.

Repeated measurements and related genotypes. There are two distinct kinds of correlation. A random regression can estimate correlation between coefficients belonging to the same subject (Laird and Ware, 1982). A supplied relationship matrix can describe correlation between different subjects or genotypes. They can also be combined.

The formula `random = ~Time | Chick` gives every chick an intercept and a slope, estimating their two variances and their covariance. The formula `~0 + Time | Chick` keeps a random slope only. The formula `~diag(1 + Time) | Chick` retains both coefficients but fixes their covariance to zero. Independent chicks can therefore have correlated intercept and slope coefficients within each chick. Centering a time covariate changes the interpretation of the intercept and its covariance with the slope. The chicken example in Section 1.1 uses a random slope for each chick.

For related genotypes, `ucov = list(gen = A)` supplies their relationship matrix. The matrix names identify the random grouping terms. The tropical-maize example in Section 1.1 illustrates this use.

Residuals. Residuals are the variation left after the fixed and random terms. The default `cov = ~units` gives one residual variance for a single response. The word `units` is reserved in covariance formulas: it identifies the original data rows and does not require a column named `units` in the data. Different rows have independent residuals in these `units` structures. Shared random effects can still correlate their observations.

The formula `cov = ~at(env):units` estimates a separate residual variance for each environment. Each such term contains `units`. Random blocks describe shared block deviations; heterogeneous residual variances describe differences in the remaining variability.

Measurements from the same individual or nearby plots may remain correlated after fitting the random effects. Residual covariance formulas can describe that pattern; the repeated-measurement and spatial examples illustrate the available structures.

Multivariate responses. A matrix response, for example `cbind(GYLD, TKW, PLH)`, fits several traits jointly. Each trait has its own fixed-effect coefficients, while the random and residual blocks describe covariance across traits. The symbol `trait` is reserved in covariance formulas and indexes the response columns in their `cbind()` order. The response columns themselves supply this trait index. The joint maize analysis in Section 6.1.3 illustrates genetic and residual covariances between traits.

1.3 Specifying variance structures

Variance structures describe how variability differs between groups and which measurements or outcomes may be correlated. The fixed-effects formula still describes their expected values. A common residual variance is the default; separate variances can be fitted for locations or other groups. For several traits, `us(trait)` allows correlations between traits, whereas `diag(trait)` sets those correlations to zero in that component (Searle et al., 1992).

For one outcome recorded once per subject and visit, a residual formula such as `~ar1(0 + Visit) | Subject` models correlation across visits within a subject. The factor `Visit` defines their order. Random slopes describe differences in individual trajectories, as in the chicken example in Section 1.1.

Table 1 summarizes variance structures. Here, q is the number of random coefficients or ordered measurement occasions. AR(1) correlation decreases with the distance between ordered occasions when its correlation parameter lies between zero and one.

Separable AR(1) by AR(1) residuals are illustrated with barley in Section 3.10. The chicken analysis in Section 7.2.6 combines random regression with separate residual variances for measurement times.

Table 1: Univariate variance structures and numbers of covariance parameters.

Modifier or residual formula	Restriction	Parameters
Univariate random coefficients or residual occasions: q positions		
`us(...)`	Separate variances and unrestricted covariances	$q(q+1)/2$
`diag(...)`	Separate variances, zero covariances	q
`homdiag(...)`	Common variance, zero covariances	1
`cs(..., hom=TRUE)`	Common variance and common correlation	2
`cs(..., hom=FALSE)`	Separate variances and common correlation	$q+1$
`ar1(..., hom=TRUE)`	Common variance and AR(1) correlation	2
`ar1(..., hom=FALSE)`	Separate variances and AR(1) correlation	$q+1$
Spatial residual covariance		
`~ar1(Row):ar1(Column)`	Common variance, with separate AR(1) correlations along rows and columns	3

Different experimental designs. For a univariate response, experiment A can have spatial AR(1) by AR(1) residuals, while experiment B has random replicates and blocks nested within replicates, together with independent residuals. The factor `experiment` identifies the experiments; `Row` and `Column` identify plot coordinates, and `rep` and `block` identify the replicates and incomplete blocks. The model arguments are:

```
random = ~(rep/block):at(experiment, "B"),
cov = list(
    ~(ar1(Row):ar1(Column)) | at(experiment, "A"),
    ~units:at(experiment, "B")
)
```

The selectors restrict each structure to its experiment. Spatial residuals are illustrated in Section 3.10, and nested incomplete blocks in Section 3.8.

1.4 Multivariate models

A multivariate model analyses several recorded outcomes together, such as yield and dry-matter content. `cbind()` identifies those response columns. The reserved covariance symbol `trait` refers directly to those response columns. A joint analysis can estimate relationships between outcomes while reporting results for each outcome (Anderson, 2003).

Table 2 summarizes residual covariance structures for multivariate responses. Here, T is the number of traits, L the number of locations and E the number of environments. Bare `units` uses an unstructured trait covariance matrix.

Unstructured covariance between traits is illustrated in the maize analysis in Section 6.1.3. The three-trait barley analysis in Section 6.4 estimates a separate residual trait covariance matrix for each environment using `at(env):units`.

The following example uses the barley multi-environment data. A separate unstructured residual covariance matrix for each location allows residual correlations between traits. The detailed three-trait analysis appears in Section 6.4:

```
data("v-lmm-barley-met", package = "SelectionTools")
m.multi <- st.lmm(
    fixed = cbind(GYLD, TKW, PLH) ~ 1,
    random = ~gen + loc + year + gen:loc + gen:year + loc:year +
        gen:loc:year + year:loc:rep,
    cov = ~us(trait):at(loc):units,
```

Table 2: Multivariate residual variance structures and numbers of covariance parameters.

Residual formula	Restriction	Parameters
`~units`	One common unstructured trait matrix.	$T(T+1)/2$
`~diag(trait):units`	Separate trait variances, without trait covariance.	T
`~at(loc):units`	A full trait covariance matrix for each location.	$LT(T+1)/2$
`~diag(trait):at(loc):units`	A separate variance for every trait and location.	LT
`~at(env):units`	A full trait matrix for every defined environment.	$ET(T+1)/2$
`~diag(trait):at(env):units`	A separate variance for every trait and environment.	ET

```
    data = barley.met,
    control = list(maxit = 1000)
)
```

The setting `control = list(maxit = 1000)` allows the optimizer up to 1,000 iterations when estimating variance and covariance parameters. Fitting stops earlier when convergence is reached. A larger iteration limit permits more iterations but does not guarantee convergence.

Separate variances for groups. For a single response, `~units` has one residual variance and `~at(loc):units` has one per location. Similarly, `random = ~at(grp):gen` gives independent genotype components in the groups defined by `grp`. The fitted fixed mean still comes from `fixed`.

The operator `at(factor)` creates a separate covariance block for each observed level of the specified factor. For a location factor named `loc`, `at(loc, "HOH")` selects only location HOH, while `at(loc, c("HOH", "EWE"))` creates separate blocks for those two locations. A selected location contributes residual variation only for observations at that location.

The following specification assigns separate trait variances to HOH and EWE and unstructured trait covariance matrices to EIN, OLI and HFH.

```
cov = list(
    ~diag(trait):at(loc, c("HOH", "EWE")):units,
    ~us(trait):at(loc, c("EIN", "OLI", "HFH")):units
)
```

1.5 Model selection

Model selection compares alternative sets of fixed predictors. The `selection` list specifies the search method, the largest number of predictors considered through `nvmax`, and the comparison criterion. An exhaustive search examines all eligible subsets up to that size; forward and backward searches examine a sequence of models (Draper and Smith, 1998).

`summary(model.selection)` displays the comparison. The specified full model remains fitted. The chosen subset can be fitted with its own formula in a subsequent `st.lmm` call, as illustrated in Section 7.1.4.

The following example compares subsets of five predictors of fuel economy. The smallest `BICm` identifies the preferred retained model. The information-criterion approach is described by Akaike (1974) and Schwarz (1978).

```
> data("v-lmm-mtcars", package = "SelectionTools")
> model.selection <- st.lmm(
+     mpg ~ disp + hp + wt + qsec + am,
+     data = mtcars,
+     method = "moments",
```

```
+     selection = list(
+         search = "exhaustive",
+         nvmax = 5,
+         criterion = "BICm"
+     )
+ )
> summary(model.selection)
```

1.6 Tests of model parameters

1.6.1 Fixed parameters

Fixed-effect tests assess whether a model term contributes to the expected response after accounting for the other terms. A factor term usually involves several coefficients, which are tested jointly; a regression term can test a slope. The default Type III tests assess each term conditional on the other model terms. In a model containing interactions, the tested hypothesis also depends on those interactions and the factor coding. The call `anova(m)` returns the fitted model's fixed-term tests. Explicit coefficient hypotheses are available through `st.lsm` with `L` and `joint=TRUE` (Searle, 1987).

For complete balanced categorical designs with independent random intercepts and homogeneous errors, fixed-effect ANOVA terms use exact F-tests where an expected-mean-square denominator is available. The output labels these tests as exact, including when variance components are estimated by REML. Different factors can use different denominators, as in a split-plot experiment. For other terms, an F or Wald statistic compares the estimated effect with its uncertainty under the null hypothesis. The default Satterthwaite method then approximates denominator degrees of freedom while accounting for estimated covariance parameters (Satterthwaite, 1946). For suitable REML models, `ddf="Kenward-Roger"` also adjusts the covariance of the fixed-effect estimates and scales the F statistic (Kenward and Roger, 1997). The setting `ddf="asymptotic"` uses a large-sample Wald reference, displayed with infinite denominator degrees of freedom. These methods assess the same specified effects with different approximations to their uncertainty.

For balanced moment analyses, fixed-effect F tests use the expected mean squares to identify an appropriate error term. That denominator can differ between experimental factors, as in a split-plot experiment. For fixed-only moment models, `anova(m1, m2)` compares nested models by an F test (Cornfield and Tukey, 1956; Underwood, 1997).

1.6.2 Random parameters

A random-term test asks whether the variation represented by an entire random term is needed. For example, testing a random genotype term asks whether its variance is zero. For a random regression or a multivariate term, all associated coefficients and covariance parameters are assessed together. The call `anova(m, effects="random")` returns these tests; `effects="both"` combines fixed, random and any joint tests in one table.

For likelihood fits, the default `random.test="auto"` uses exact F-tests when a complete balanced categorical design provides an appropriate expected-mean-square denominator for every random term. The REML or ML variance estimates are retained. The explicit setting `random.test="F"` selects this exact analysis. In the regions and villages example, the village-within-region mean square is compared with the residual mean square; the fixed region effect instead uses the village-within-region mean square.

For other designs, `random.test="auto"` uses likelihood-ratio tests. The explicit setting `random.test="LRT"` selects these tests. Each compares the full model with a reduced model omitting the tested term. The statistic is twice the difference in log likelihood, with the same observations, fixed effects and estimation method in both fits. REML models therefore use REML likelihoods for this comparison. The reported chi-squared approximation uses the difference in the number of free covariance parameters as its degrees of freedom. Zero variance lies on the boundary of the parameter space, so this reference distribution is approximate (Self and Liang, 1987).

The option `random.test="bootstrap"` obtains a simulation-based reference distribution. Responses are simulated under the reduced model, and both models are fitted to each simulated response. The p-value

measures how often the simulated likelihood-ratio statistic reaches or exceeds the observed statistic. The simulation count is set by `control$bootstrap`; `set.seed()` provides a reproducible starting state. The returned results report Monte Carlo uncertainty and failed fits. This approach requires many more fits (Davison and Hinkley, 1997). For balanced moment analyses, random-term tests instead use expected mean squares and the corresponding F-test denominators. The setting `random.test="none"` omits random-term testing.

1.7 Least-squares means and reference grids

Least-squares means (in the terminology of SAS and the older R package `lsmeans`) or expected marginal means (in the terminology of the newer R packge `emmeans`) describe the expected response of a treatment under conditions specified by the fixed effects of the model (Lenth, 2016). Pairwise comparisons, treatment vs. control comparisons, or user defined contrast matrices test hypotheses on these least-squares means. The reference grid contains the factor combinations and numeric predictor values used to calculate the least-squares means. Predictor names or a one-sided formula select the reported means.

In the rapeseed example we investigated genotypes at two fixed enviroments (drought stress and no stress). The trial and its genotype means are discussed in Section 2.3.

```
library(SelectionTools)
data("v-lmm-rapeseed", package = "SelectionTools")
m.rapeseed <- st.lmm(
    fixed = hgt ~ geno * env,
    random = ~env:rep,
    data = rapeseed
)
```

`~geno` gives a reference grid of genotype means averaged equally over environments. The formula `~geno * env` gives a mean for every genotype–environment combination in a combined grid. The formula `~geno | env` gives the same means, grouped by environment.

```
st.lsm(m.rapeseed, ~geno)
st.lsm(m.rapeseed, ~geno * env)
st.lsm(m.rapeseed, ~geno | env)
```

The grouping determines how subsequent comparisons are organized. Pairwise comparisons based on `~geno * env` include pairs of genotype–environment combinations across the combined grid, with a respective multiple testing adjustment. Pairwise comparisons based on `~geno | env` compare genotypes within each environment, with multiplicity adjustment applied separately within each environment by default.

The chicken eperiment compares weight gain of four chicken diets. The diets are fixed regression slopes, and each chicken gets a random slope for its individual growth, the residuals have different variances at the times when chicken weight was measured. We investigate the square root of thicken weight to make the estimated regression lines look more linear. The growth models are discussed in Section 7.2.

```
library(SelectionTools)
data("v-lmm-chickens", package = "SelectionTools")
chickens$weight.s <- sqrt(chickens$weight)
chicken.model <- st.lmm(
    fixed = weight.s ~ Time:Diet,
    random = ~ (0 + Time) | Chick,
    cov = ~at(Time):units,
    data = chickens
)
```

In the chicken trial, `at=list(Time=21)` estimates the mean response for each diet at day 21. The reference grid fixes `Time` at 21 and includes each diet.

```
st.lsm(
    chicken.model,
    "Diet",
    at = list(Time = 21)
)
```

The option `trend="Time"` instead estimates the growth slope for each diet.

```
st.lsm(
    chicken.model,
    "Diet",
    trend = "Time"
)
```

The chicken model uses square-root weight as its response. Back-transformation expresses the least-squares means and their confidence intervals in the original weight units. The estimated square-root means and interval endpoints are truncated at zero and then squared, while standard errors are converted using the delta method. Tests remain on the square-root scale. These back-transformed least-squares means describe the fitted growth curves on the weight scale. Estimating arithmetic mean weight additionally requires a correction for variation among chicks and residual variation. The square-root growth analysis appears in Section 7.2.

```
st.lsm(
    chicken.model,
    "Diet",
    at = list(Time = 21),
    transform = "sqrt",
    type = "response"
)
```

The standard error of a least-squares mean depends on the inference space, which specifies the population the mean describes. The default `inference.space = "broad"` averages over the random effects. With `inference.space = "narrow"`, the mean describes the particular groups in the experiment for all random terms. A formula such as `inference.space = operator` gives an intermediate mean for the operators in the experiment while averaging over the other random effects. Section 2.7 compares the standard errors for these choices using the same model fit.

1.8 Test of hypotheses on least-squares means

Pairwise and control comparisons. The fertilizer experiment estimates mean yield for each treatment. The detailed comparisons appear in Section 3.2.

```
data("v-lmm-fertilizer", package = "SelectionTools")
m.fert <- st.lmm(
    yield ~ fert,
    data = fertilizer
)
```

The pairwise comparison tests yield differences between all treatment pairs. Hochberg adjustment controls the family-wise error rate under its dependence assumptions (Hochberg, 1988).

```
st.lsm(
    m.fert,
    "fert",
```

```
    method = "pairwise",
    adjust = "hochberg"
)
```

The one-sided Dunnett comparisons test whether each treatment yields more than the control. The argument `ref=1` selects the first treatment in the means grid. The multivariate t adjustment accounts for the correlations between contrasts that share the control (Dunnett, 1955 and Genz and Bretz, 2009).

```
st.lsm(
    m.fert,
    "fert",
    method = "trt.vs.ctrl",
    ref = 1,
    side = ">",
    adjust = "mvt"
)
```

User-defined comparisons. In the oats split-plot experiment, varieties occupy whole plots within blocks and nitrogen treatments occupy subplots. The detailed oats analysis appears in Section 3.7.

```
data("v-lmm-oats", package = "SelectionTools")
m.oats.contrasts <- st.lmm(
    fixed = Yield ~ Variety * Nitro,
    random = ~Block + Block:Variety,
    data = oats
)
```

The Least-squares means for nitrogen treatments average equally over varieties.

```
st.lsm(m.oats.contrasts, ~Nitro)
```

The contrast compares the average yield of the three fertilized treatments with the unfertilized treatment. Its coefficients follow the nitrogen-level order `n0.0`, `n0.2`, `n0.4`, `n0.6`.

```
st.lsm(
    m.oats.contrasts,
    ~Nitro,
    method = list(fertilized_vs_unfertilized = c(-1, 1/3, 1/3, 1/3))
)
```

Contrasts with fixed right-hand sides. The joint F test assesses two nitrogen-treatment differences simultaneously. The null hypothesis specifies that the yield difference between `n0.2` and `n0.0` equals 10 and that the difference between `n0.6` and `n0.0` equals 30, in yield units. The two values in `rhs` follow the order of the contrasts. The underlying split-plot analysis appears in Section 3.7.

```
st.lsm(
    m.oats.contrasts,
    ~Nitro,
    method = list(
        low_vs_zero = c(-1, 1, 0, 0),
        high_vs_zero = c(-1, 0, 0, 1)
    ),
    rhs = c(10, 30),
    joint = TRUE
)
```

Letter displays. A compact letter display summarizes the Tukey-adjusted pairwise comparisons. Treatments that share a letter have no significant difference at the 5% level. Treatments with no letter in common differ significantly (Piepho, 2004). The fertilizer example in Section 3.2 includes letter displays.

```
st.lsm.cld(st.lsm(m.fert, "fert"))
```

1.9 Best linear unbiased prediction of random effects

Best linear unbiased predictions (BLUPs) estimate the random effects in a mixed model. They account for the fitted variance components and the information available for each effect (Henderson, 1975).

The half-diallel model contains a shared parental effect for general combining ability (GCA) and a cross effect for specific combining ability (SCA). The overlay adds the effects of the two parents of each cross (Sprague and Tatum, 1942). The example is discussed in Section 8.1.

```
library(SelectionTools)
data("v-lmm-halfdiallel", package = "SelectionTools")
m.diallel <- st.lmm(
    fixed = sugar ~ 1,
    random = ~overlay(female, male) + cross,
    data = halfdiallel
)
```

This prediction returns each parent's GCA effect.

```
st.predict(
    m.diallel,
    "0 + overlay(female, male)"
)
```

The second returns each cross's SCA effect. The `0 +` prefix requests the random effects alone.

```
st.predict(
    m.diallel,
    "0 + cross"
)
```

In the tropical-maize trial, genotype and block effects are random. The marker-based relationship matrix `A` describes the covariance between genotype effects and enters the model through `ucov` (VanRaden, 2008). The detailed genomic analysis appears in Section 9.1.3.

```
library(SelectionTools)
data("v-lmm-tropmaize", package = "SelectionTools")
data("v-tropmaize-vcf", package = "SelectionTools")
st.load.vcf.data(v.tropmaize.vcf, data.set = "maize.markers")
A <- st.get.A("maize.markers")
maize.model <- st.lmm(
    fixed = y ~ 1,
    random = ~gen + block,
    ucov = list(gen = A),
    data = tropmaize
)
```

The prediction call returns the fixed intercept plus each predicted genotype effect, giving genomic BLUPs of the genotypic values.

Table 3: Methods and components for accessing fitted model results.

Command or component	Contents
`coef(m)`, `vcov(m)`	Fixed coefficients and their covariance.
`confint(m)`	Normal-reference intervals for fixed coefficients.
`fitted(m)`, `residuals(m)`	Matrices with observations in rows and traits in columns. The option `type="marginal"` uses the fixed-effects mean.
`formula(m)`, `model.matrix(m)`	Fixed formula and its design matrix.
`anova(m)`	Cached fixed-term tests.
`anova(m, effects="both")`	A common table of fixed, random and joint tests.
`logLik(m)`, `AIC(m)`, `BIC(m)`	Log-likelihood and tables of fit criteria with a row per model.
`nobs(m)`, `df.residual(m)`	Observed response cells and their count minus the fixed-design rank.
`st.lmm.components(m, "observations")`	Retained rows, response cells, fixed rank, and error-stratum DF when available.
`predict(m)`	A list with predictions in `fit`, a matrix with observations in rows and traits in columns.
`st.lmm.components(m, "variance")`	A matrix with a row per block and a column per trait.
`st.lmm.components(m, "covariance")`	Covariance matrices grouped into random and residual blocks.
`st.lmm.components(m, "correlation")`	Correlation matrices with the same block groups.
`st.lmm.components(m, "random")`	Random coefficients, with grouping levels in rows.
`st.lmm.components(m, "metadata")`	Convergence and computation diagnostics in `convergence`.

```
st.predict(
    maize.model,
    "gen"
)
```

Prediction output includes prediction-error standard errors and pointwise confidence intervals, with fitted variance parameters held fixed.

1.10 Results, printing and serialization

Table 3 summarizes methods and components that provide access to fitted results.

A request for a single covariance or correlation block, such as `block = "gen"`, returns that matrix directly. A request for several components returns a named list of those components.

Shared printing settings control automatic console display for fitted models, least-squares means, predictions, test tables and extracted results. The session-wide defaults are listed in Table 4. The function `st.options()` returns the current settings as a visible named list, so a call without an assignment displays them. A call such as `st.options(list(print.digits = 3))` changes the named setting and returns the complete updated list invisibly. Setting options therefore produces no automatic console output, while the returned list remains available for assignment. An individual value is available as `st.options()$print.digits`.

ANOVA tables omit the intercept test by default. Regression coefficient reports retain the intercept

Table 4: Shared printing options and their initial values.

Option	Default	Display choice
`print.digits`	`4`	Decimal places for statistics.
`print.char.perline`	`75`	Maximum line width.
`print.top`	`10`	Leading table rows.
`print.bottom`	`5`	Trailing table rows.
`print.cov.blocks.top`	`3`	Leading covariance blocks.
`print.cov.blocks.bottom`	`1`	Trailing covariance blocks.
`print.ems`	`FALSE`	Expected mean squares in moment-model output.
`print.test.construction`	`FALSE`	Numerator and denominator details for moment tests.
`print.tests.for.bare.lsmeans`	`FALSE`	Tests against zero for ordinary least-squares means.
`print.intercept.test`	`FALSE`	Stored intercept tests in ANOVA tables.
`print.LSD`	`TRUE`	LSD beneath least-squares means without displayed tests.
`print.HSD`	`TRUE`	HSD beneath least-squares means without displayed tests.

estimate, standard error and test. The setting `st.options(list(print.intercept.test = TRUE))` includes an intercept test where it is stored in the fitted model.

Balanced categorical analyses can also report the least significant difference (LSD) and the Tukey-based honestly significant difference (HSD). One or two compact lines identify common comparison strata, for example `Within env` and `Across env`, at an alpha level of one minus the confidence level. The options `print.LSD` and `print.HSD` control their display independently. For a least-squares means result `x`, `attr(summary(x), "thresholds")` returns the available thresholds with their SE and DF. Thresholds refer to the analysis scale, including when the means are displayed after back-transformation.

The following example changes the formatting and displays the model coefficients.

```
> printing.options <- st.options()
> printing.options
> st.options(list(
+     print.digits = 3,
+     print.char.perline = 60,
+     print.top = 5,
+     print.bottom = 2
+ ))
> printing.coefficients <- coef(m.rapeseed)
> printing.coefficients
```

Setting either `print.top` or `print.bottom` to `Inf` displays all rows. All covariance blocks appear when either block limit in Table 4 is set to `Inf`. The row limits apply independently. These settings affect display; stored estimates and test results retain their full values.

The following conversions return ordinary R values for further work:

```
b <- as.numeric(coef(m.rapeseed))
V <- as.matrix(vcov(m.rapeseed))
lsmeans <- as.data.frame(st.lsm(m.rapeseed, ~geno))
```

Models and their results can be saved and restored with the standard R serialization functions:

```
rapeseed.file <- tempfile(fileext = ".rds")
saveRDS(m.rapeseed, rapeseed.file)
m.saved <- readRDS(rapeseed.file)
coef(m.saved)
```

1.11 Estimation procedures

1.11.1 Estimation methods: REML, ML and moments

The argument `method` selects how model parameters are estimated. Restricted maximum likelihood (REML), the default, estimates variance and covariance parameters while accounting for estimation of the fixed effects (Patterson and Thompson, 1971). Maximum likelihood, `method="ML"`, estimates the parameters from the likelihood and provides a common basis for comparing models with different fixed effects (Harville, 1977). The setting `method="moments"` gives ordinary least squares for fixed-only regression and ANOVA-based estimates from expected mean squares following the Cornfield–Tukey rules for balanced experimental designs, including nested, split-plot and replicated multi-environment trials (Cornfield and Tukey, 1956).

1.11.2 Optimizers: Automatic selection, AI and L-BFGS-B

Likelihood estimation finds variance and covariance parameters iteratively. The default `control = list(optimizer = "auto")` chooses the optimizer. For REML models with diagonal free covariance blocks and at most 200 free covariance parameters, automatic selection uses safeguarded average-information iterations (AI-REML) (Gilmour et al., 1995). Other models, including ML fits, use L-BFGS-B. The setting `optimizer="AI"` also offers AI-REML for models with unstructured or other structured covariance blocks within the parameter limit. The setting `optimizer="L-BFGS-B"` selects L-BFGS-B directly (Byrd et al., 1995). Moment estimation calculates estimates from regression and ANOVA quantities.

The printed fit reports convergence. The option `control$maxit` sets the main iteration limit; fitting can stop earlier when convergence is reached. A larger limit permits more iterations but does not guarantee convergence.

1.11.3 Backend paths

The `control` list selects the covariance calculation through `backend`. These choices give alternative calculations of the same statistical model.

Automatic selection. The default `backend="auto"` chooses a calculation from the covariance structure and matrix sizes. It first checks for a suitable spectral model, then considers sparse mixed-model equations and covariance blocks. The dense calculation handles the remaining models. For example, repeated measurements can lead to separate subject blocks, while independent residuals can permit sparse equations. The selection uses estimated computational costs; the fastest choice can also depend on the data and BLAS installation.

Dense covariance calculation. The setting `backend="dense"` assembles the full covariance matrix of the observed responses. Cholesky factorization supplies linear solves and log determinants during estimation (Anderson et al., 1999). This gives a general calculation for complex random and residual covariance models, including correlated traits. Dense BLAS operations can be efficient for large matrices, while storage grows with the square of the number of observed response values. Automatic selection uses this calculation when the structured alternatives do not provide an eligible, sufficiently economical calculation.

Block covariance calculation. The setting `backend="block"` factors separate observation-covariance blocks, for example repeated measurements within independent subjects. Smaller factorizations reduce storage and work, and identical blocks can reuse a factorization. Suitable designs can also include random effects connecting the blocks through a smaller coefficient system. Automatic selection considers this route after spectral and sparse checks; its cost comparison favors partitions with smaller blocks. An explicit block request selects the block calculation for a suitable model.

Sparse mixed-model equations. The setting `backend="sparse"` uses the nonzero entries of the mixed-model equations for models with independent residuals and suitable random-effect designs. It is useful when many observations share relatively few random coefficients, as in replicated trials or random regressions. Sparse factorization can reduce storage and calculation when few additional nonzero entries arise during elimination (Davis, 2006). Supplied relationship terms can enter through a smaller covariance correction. Automatic selection checks model size and the estimated factorization work before choosing this route.

Spectral calculation. The setting `backend="spectral"` uses eigendecompositions for suitable relationship-kernel models with one positive diagonal residual shape. Repeated variance evaluations reuse the transformed matrices; verified product structures can use smaller factor eigensystems and tensor operations. Automatic selection checks this route first. The option `eigen.solver="dsyevd"` selects LAPACK DSYEVD and `eigen.solver="dsyevr"` selects DSYEVR within the spectral calculation. The default `eigen.solver="auto"` prefers DSYEVD when its workspace fits the LAPACK integer interface, and uses DSYEVR otherwise (Anderson et al., 1999).

The following rapeseed model uses automatic optimizer selection with the dense covariance calculation. The trial analysis appears in Section 2.3:

```
> data("v-lmm-rapeseed", package = "SelectionTools")
> model.auto <- st.lmm(
+     hgt ~ geno * env,
+     random = ~env:rep,
+     data = rapeseed,
+     control = list(optimizer = "auto", backend = "dense")
+ )
```

2 Cross classified and hierarchical Designs

2.1 Moment estimators

2.1.1 Two fixed factors: Climate and variety

CM01: RLab-AppStat/v14-u31.R

```
> data("v-lmm-climate", package = "SelectionTools")
> m.climate <- st.lmm(
+     yield ~ climate * variety,
+     data = climate,
+     method = "moments"
+ )
> m.climate
```

```
Linear model fitted by the method of moments

Model:
  Data:      climate
  Fixed:     yield ~ climate * variety

ANOVA (method of moments, exact F):
  Source           DF1  DF2          SS          MS        F        p
```

```
  climate            2   18   2800.0000   1400.0000   9.5599   0.0015
  variety            1   18   5400.0000   5400.0000  36.8741  <0.0001
  climate:variety    2   18  22800.0000  11400.0000  77.8452  <0.0001
  Residual          18        2636.0000    146.4444

Model statistics:
  Trait        N  Residual DF      R2  Adjusted R2  Residual SD
  yield  24.0000           18  0.9216       0.8999      12.1014
```

```
> st.lsm(m.climate, ~climate)
```

```
Least-squares means (exact)
  climate  lsmeans      SE  DF  lower.CL  upper.CL
  I        55.0000  4.2785  18   46.0112   63.9888
  II       80.0000  4.2785  18   71.0112   88.9888
  III      60.0000  4.2785  18   51.0112   68.9888
Confidence level: 0.9500
LSD=12.7121 HSD=15.4424 (alpha=0.05)
```

```
> st.lsm(
+     m.climate,
+     ~climate,
+     method = "pairwise",
+     adjust = "tukey"
+ )
```

```
Least-squares mean contrasts for climate (exact)
  contrast  estimate      SE  DF  lower.CL  upper.CL        t       p
  I - II    -25.0000  6.0507  18  -40.4424   -9.5576  -4.1317  0.0017
  I - III    -5.0000  6.0507  18  -20.4424   10.4424  -0.8263  0.6919
  II - III   20.0000  6.0507  18    4.5576   35.4424   3.3054  0.0104
p-value and interval adjustment: tukey; confidence level: 0.9500
```

```
> st.lsm(m.climate, ~climate * variety)
```

```
Least-squares means (exact)
  climate  variety   lsmeans      SE  DF  lower.CL  upper.CL
  I        a         80.0000  6.0507  18   67.2879   92.7121
  II       a         60.0000  6.0507  18   47.2879   72.7121
  III      a         10.0000  6.0507  18   -2.7121   22.7121
  I        b         30.0000  6.0507  18   17.2879   42.7121
  II       b        100.0000  6.0507  18   87.2879  112.7121
  III      b        110.0000  6.0507  18   97.2879  122.7121
Confidence level: 0.9500
LSD=17.9776 HSD=27.1944 (alpha=0.05)
```

2.1.2 Nested mixed model: Regions and villages

CM02: RLab-AppStat/v14-u31.R

```
> data("v-lmm-soils", package = "SelectionTools")
> m.soils <- st.lmm(
+     fixed = yield ~ region,
+     random = ~region:village,
```

```
+     data = soils,
+     method = "moments"
+ )
> m.soils
```

```
Linear mixed model fitted by the method of moments

Model:
  Data:      soils
  Fixed:     yield ~ region
  Random:    ~region:village

Design: balanced nested

ANOVA (method of moments, exact F):
  Source          DF1  DF2        SS         MS        F        p
  region            2    3  208.0000  104.0000   2.7857   0.2071
  region:village    3   18  112.0000   37.3333  12.9231  <0.0001
  Residual         18        52.0000    2.8889

Unconstrained moment variance estimates:
  Component       Estimate
  region:village    8.6111
  Residual          2.8889
```

The `region:village` row tests whether the variance among villages within regions is zero. Its mean square is compared with the residual mean square, giving an exact F-test with 3 and 18 degrees of freedom.

The shared printing options also show the expected mean squares and the construction of the F-tests after the variance component estimates.

```
> st.options(list(
+     print.ems = TRUE,
+     print.test.construction = TRUE
+ ))
> m.soils
```

```
Linear mixed model fitted by the method of moments

Model:
  Data:      soils
  Fixed:     yield ~ region
  Random:    ~region:village

Design: balanced nested

ANOVA (method of moments, exact F):
  Source          DF1  DF2        SS         MS        F        p
  region            2    3  208.0000  104.0000   2.7857   0.2071
  region:village    3   18  112.0000   37.3333  12.9231  <0.0001
  Residual         18        52.0000    2.8889

Unconstrained moment variance estimates:
  Component       Estimate
  region:village    8.6111
  Residual          2.8889

Expected mean squares (EMS):
  Convention: Underwood (1997), Cornfield-Tukey; restricted interactions
  EMS(region)         = s2 + 4 s2(region:village) + 8 theta(region)
  EMS(region:village) = s2 + 4 s2(region:village)
  EMS(Residual)       = s2

Test construction:
```

```
  region (exact F-test)
    Numerator MS: region
    Denominator MS: region:village

  region:village (exact F-test)
    Numerator MS: region:village
    Denominator MS: Residual
```

The least-squares means describe the mean yield in each region, with uncertainty accounting for the random variation among villages within regions.

```
> st.lsm(m.soils, ~region)
```

```
Least-squares means (exact)
  region  lsmeans      SE  DF  lower.CL  upper.CL
  I        9.0000  2.1602   3    2.1251   15.8749
  II      14.0000  2.1602   3    7.1251   20.8749
  III      7.0000  2.1602   3    0.1251   13.8749
Confidence level: 0.9500
LSD=9.7225 HSD=12.7663 (alpha=0.05)
```

The region effects are deviations from the equally weighted average of the three region means. A positive effect indicates a region mean above this average; a negative effect indicates a mean below it. The tests compare these deviations with zero.

```
> st.lsm(
+     m.soils,
+     ~region,
+     method = "eff"
+ )
```

```
Least-squares mean contrasts (exact)
  contrast    estimate      SE  DF  lower.CL  upper.CL        t       p
  I effect     -1.0000  1.7638   3   -6.6133    4.6133  -0.5669  0.6104
  II effect     4.0000  1.7638   3   -1.6133    9.6133   2.2678  0.1081
  III effect   -3.0000  1.7638   3   -8.6133    2.6133  -1.7008  0.1875
p-value adjustment: none; confidence level: 0.9500
```

A one-factor ANOVA compares the six villages as fixed levels. The village factor has six levels, `A`–`F`, with two villages in each region. The residual variation describes differences among fields within villages.

```
> m.villages <- st.lmm(
+     fixed = yield ~ village,
+     data = soils,
+     method = "moments"
+ )
> m.villages
```

```
Linear model fitted by the method of moments

Model:
  Data:      soils
  Fixed:     yield ~ village

ANOVA (method of moments, exact F):
  Source    DF1  DF2        SS       MS        F        p
```

```
  village     5   18  320.0000  64.0000  22.1538  <0.0001
  Residual   18        52.0000   2.8889

Model statistics:
  Trait        N  Residual DF      R2  Adjusted R2  Residual SD
  yield  24.0000           18  0.8602       0.8214       1.6997
```

The compact letter display orders the village means from smallest to largest and uses Tukey-adjusted pairwise comparisons. Villages sharing a letter have no significant pairwise difference at the 5% level (Piepho, 2004).

```
> st.lsm.cld(
+     st.lsm(m.villages, ~village),
+     sort = TRUE,
+     adjust = "tukey"
+ )
```

```
Least-squares means (exact)
  village  lsmeans      SE  DF  lower.CL  upper.CL  group
  E         6.0000  0.8498  18    3.4905    8.5095  a
  A         7.0000  0.8498  18    4.4905    9.5095  a
  F         8.0000  0.8498  18    5.4905   10.5095  ab
  B        11.0000  0.8498  18    8.4905   13.5095   b
  C        11.0000  0.8498  18    8.4905   13.5095   b
  D        17.0000  0.8498  18   14.4905   19.5095    c
Confidence level: 0.9500
Comparison adjustment: tukey (alpha=0.05)
Confidence intervals use sidak adjustment.
Shared letters indicate failure to reject a difference.
```

2.1.3 Crossed and nested effects: Rapeseed

The rapeseed trial compares plant height for 20 genotypes in two environments, with three replicates in each environment. Genotypes occur in both environments, whereas replicates are nested within environments.

```
> data("v-lmm-rapeseed", package = "SelectionTools")
```

Genotype, environment and their interaction are fixed effects. The random `rep:env` term describes variation among replicates within environments. The method of moments estimates the replicate and residual variances and provides F-tests for the model terms. The shared printing options include the expected mean squares and the construction of these tests in the model output.

```
> st.options(list(
+     print.ems = TRUE,
+     print.test.construction = TRUE
+ ))
> m.rape <- st.lmm(
+     fixed = hgt ~ geno * env,
+     random = ~rep:env,
+     data = rapeseed,
+     method = "moments"
+ )
> m.rape
```

```
Linear mixed model fitted by the method of moments
```

```
Model:
  Data:      rapeseed
  Fixed:     hgt ~ geno * env
  Random:    ~rep:env

Design: general balanced orthogonal

ANOVA (method of moments, exact F):
  Source     DF1  DF2          SS         MS        F        p
  geno        19   76  11425.8667   601.3614   6.6832  <0.0001
  env          1    4   9469.6333  9469.6333  51.5051   0.0020
  geno:env    19   76   1204.3667    63.3877   0.7045   0.8031
  env:rep      4   76    735.4333   183.8583   2.0433   0.0967
  Residual    76        6838.5667    89.9811

Unconstrained moment variance estimates:
  Component  Estimate
  env:rep      4.6939
  Residual    89.9811

Expected mean squares (EMS):
  Convention: Underwood (1997), Cornfield-Tukey; restricted interactions
  EMS(geno)     = 1 s2                  + 6 theta(geno)
  EMS(env)      = 1 s2 + 20 s2(env:rep) + 60 theta(env)
  EMS(geno:env) = 1 s2                  + 3 theta(geno:env)
  EMS(env:rep)  = 1 s2 + 20 s2(env:rep)
  EMS(Residual) = 1 s2

Test construction:

  geno (exact F-test)
    Numerator MS: geno
    Denominator MS: Residual

  env (exact F-test)
    Numerator MS: env
    Denominator MS: env:rep

  geno:env (exact F-test)
    Numerator MS: geno:env
    Denominator MS: Residual

  env:rep (exact F-test)
    Numerator MS: env:rep
    Denominator MS: Residual
```

The environment test uses the replicate-within-environment mean square as its denominator, with four denominator degrees of freedom. Genotype and genotype-by-environment interaction are tested against the residual mean square, with 76 denominator degrees of freedom. Genotype and environment have significant effects; the interaction is not significant.

The genotype least-squares means describe plant height averaged over both environments.

```
> st.lsm(m.rape, ~geno)
```

```
Least-squares means (Satterthwaite)
  geno    lsmeans      SE     DF  lower.CL  upper.CL
  l01   215.8333  3.9723  76.43  207.9225  223.7441
  l02   232.1667  3.9723  76.43  224.2559  240.0775
  l03   212.8333  3.9723  76.43  204.9225  220.7441
  l04   223.6667  3.9723  76.43  215.7559  231.5775
  l05   207.8333  3.9723  76.43  199.9225  215.7441
  l06   229.1667  3.9723  76.43  221.2559  237.0775
  l07   209.6667  3.9723  76.43  201.7559  217.5775
  l08   229.3333  3.9723  76.43  221.4225  237.2441
```

```
  l09   221.3333  3.9723  76.43  213.4225  229.2441
  l10   211.8333  3.9723  76.43  203.9225  219.7441
  [5 table rows omitted]
  l16   212.8333  3.9723  76.43  204.9225  220.7441
  l17   195.6667  3.9723  76.43  187.7559  203.5775
  l18   225.8333  3.9723  76.43  217.9225  233.7441
  l19   214.1667  3.9723  76.43  206.2559  222.0775
  l20   197.8333  3.9723  76.43  189.9225  205.7441
Confidence level: 0.9500
LSD=10.9077 HSD=20.1085 (alpha=0.05)
```

The following least-squares means describe plant height for every combination of genotype and environment. Further analysis of this trial appears in Section 2.3.

```
> st.lsm(m.rape, ~geno * env)
```

```
Least-squares means (Satterthwaite)
  geno  env   lsmeans      SE     DF  lower.CL  upper.CL
  l01   e1   209.0000  5.6177  76.43  197.8124  220.1876
  l02   e1   217.0000  5.6177  76.43  205.8124  228.1876
  l03   e1   204.3333  5.6177  76.43  193.1458  215.5209
  l04   e1   213.3333  5.6177  76.43  202.1458  224.5209
  l05   e1   200.3333  5.6177  76.43  189.1458  211.5209
  l06   e1   221.3333  5.6177  76.43  210.1458  232.5209
  l07   e1   205.0000  5.6177  76.43  193.8124  216.1876
  l08   e1   222.3333  5.6177  76.43  211.1458  233.5209
  l09   e1   215.3333  5.6177  76.43  204.1458  226.5209
  l10   e1   201.0000  5.6177  76.43  189.8124  212.1876
  [25 table rows omitted]
  l16   e2   220.6667  5.6177  76.43  209.4791  231.8542
  l17   e2   204.6667  5.6177  76.43  193.4791  215.8542
  l18   e2   230.0000  5.6177  76.43  218.8124  241.1876
  l19   e2   223.3333  5.6177  76.43  212.1458  234.5209
  l20   e2   214.3333  5.6177  76.43  203.1458  225.5209
Confidence level: 0.9500
Within env: LSD=15.4258 HSD=31.3689 (alpha=0.05)
Across env: LSD=15.8216 HSD=32.1694 (alpha=0.05)
```

The compact letter display compares the varieties within environment `e2`, with Tukey adjustment at a 10% significance level. Varieties sharing a letter have no significant pairwise difference.

```
> st.lsm.cld(
+     st.lsm(
+         m.rape,
+         ~geno | env,
+         at = list(env = "e2")
+     ),
+     alpha = 0.10,
+     adjust = "tukey"
+ )
```

```
Least-squares means (Satterthwaite)
At env = e2
  geno  env   lsmeans      SE     DF  lower.CL  upper.CL  group
  l17   e2   204.6667  5.6177  76.43  187.1469  222.1865  a
  l07   e2   214.3333  5.6177  76.43  196.8135  231.8531  ab
  l20   e2   214.3333  5.6177  76.43  196.8135  231.8531  ab
  l05   e2   215.3333  5.6177  76.43  197.8135  232.8531  ab
  l15   e2   218.0000  5.6177  76.43  200.4802  235.5198  ab
  l16   e2   220.6667  5.6177  76.43  203.1469  238.1865  ab
  l03   e2   221.3333  5.6177  76.43  203.8135  238.8531  abc
```

```
  l10   e2   222.6667  5.6177  76.43  205.1469  240.1865  abc
  l01   e2   222.6667  5.6177  76.43  205.1469  240.1865  abc
  l19   e2   223.3333  5.6177  76.43  205.8135  240.8531  abc
  [5 table rows omitted]
  l04   e2   234.0000  5.6177  76.43  216.4802  251.5198   bc
  l08   e2   236.3333  5.6177  76.43  218.8135  253.8531   bc
  l06   e2   237.0000  5.6177  76.43  219.4802  254.5198   bc
  l11   e2   237.3333  5.6177  76.43  219.8135  254.8531   bc
  l02   e2   247.3333  5.6177  76.43  229.8135  264.8531    c
Confidence level: 0.9500
Comparison adjustment: tukey (alpha=0.1)
Confidence intervals use sidak adjustment.
Shared letters indicate failure to reject a difference.
```

2.1.4 Crossed random effects: Years and locations

M23: barley-met

```
> data("v-lmm-barley-met", package = "SelectionTools")
> st.options(list(
+     print.ems = TRUE,
+     print.test.construction = TRUE
+ ))
> m.m23.moments <- st.lmm(
+     fixed = TKW ~ 1,
+     random = ~gen + loc + year + gen:loc + gen:year + loc:year +
+         gen:loc:year + rep:loc:year,
+     data = barley.met,
+     method = "moments"
+ )
> m.m23.moments
```

```
Linear mixed model fitted by the method of moments

Model:
  Data:      barley.met
  Fixed:     TKW ~ 1
  Random:    ~gen + loc + year + gen:loc + gen:year + loc:year +
               gen:loc:year + rep:loc:year

Design: general balanced orthogonal

ANOVA (method of moments, term-specific tests):
  Source          DF1     DF2          SS          MS        F        p
  gen           14.62   43.85  14582.0826  1041.5773   2.9494   0.0028
  loc            4.03   11.04  22128.2220  5532.0555  10.7590   0.0008
  year           2.09   19.36   2167.2323  1083.6161   1.5334   0.2409
  gen:loc          56     112   4357.7665    77.8173   3.4107  <0.0001
  gen:year         28     112   7925.7253   283.0616  12.4064  <0.0001
  loc:year       8.30  142.89   3507.8558   438.4820  11.1864  <0.0001
  gen:loc:year    112     630   2555.3662    22.8158   2.7626  <0.0001
  loc:year:rep     45     630    770.4118    17.1203   2.0730  <0.0001
  Residual        630          5202.9573     8.2587

Test methods:
  gen: quasi F (Satterthwaite)
  loc: quasi F (Satterthwaite)
  year: quasi F (Satterthwaite)
  gen:loc: exact F
  gen:year: exact F
  loc:year: quasi F (Satterthwaite)
  gen:loc:year: exact F
  loc:year:rep: exact F
```

```
  gen: approximate
  loc: approximate
  year: approximate
  loc:year: approximate

Unconstrained moment variance estimates:
  Component     Estimate
  gen            11.7252
  loc            27.9921
  year            1.2830
  gen:loc         4.5835
  gen:year       13.0123
  loc:year        6.7801
  gen:loc:year    3.6393
  loc:year:rep    0.5908
  Residual        8.2587

Expected mean squares (EMS):
  Convention: Underwood (1997), Cornfield-Tukey; restricted interactions
  EMS(gen) = 1 s2 + 4 s2(gen:loc:year) + 20 s2(gen:year) + 12 s2(gen:loc)
    + 60 s2(gen)
  EMS(loc) = 1 s2 + 15 s2(loc:year:rep) + 4 s2(gen:loc:year)
    + 60 s2(loc:year) + 12 s2(gen:loc) + 180 s2(loc)
  EMS(year) = 1 s2 + 15 s2(loc:year:rep) + 4 s2(gen:loc:year)
    + 60 s2(loc:year) + 20 s2(gen:year) + 300 s2(year)
  EMS(gen:loc)      = 1 s2 + 4 s2(gen:loc:year) + 12 s2(gen:loc)
  EMS(gen:year)     = 1 s2 + 4 s2(gen:loc:year) + 20 s2(gen:year)
  EMS(loc:year) = 1 s2 + 15 s2(loc:year:rep) + 4 s2(gen:loc:year)
    + 60 s2(loc:year)
  EMS(gen:loc:year) = 1 s2 + 4 s2(gen:loc:year)
  EMS(loc:year:rep) = 1 s2 + 15 s2(loc:year:rep)
  EMS(Residual)     = 1 s2

Test construction:

  gen (approximate F-test)
    Numerator MS: gen + gen:loc:year
    Denominator MS: gen:year + gen:loc

  loc (approximate F-test)
    Numerator MS: loc + gen:loc:year
    Denominator MS: loc:year + gen:loc

  year (approximate F-test)
    Numerator MS: year + gen:loc:year
    Denominator MS: loc:year + gen:year

  gen:loc (exact F-test)
    Numerator MS: gen:loc
    Denominator MS: gen:loc:year

  gen:year (exact F-test)
    Numerator MS: gen:year
    Denominator MS: gen:loc:year

  loc:year (approximate F-test)
    Numerator MS: loc:year + Residual
    Denominator MS: loc:year:rep + gen:loc:year

  gen:loc:year (exact F-test)
    Numerator MS: gen:loc:year
    Denominator MS: Residual

  loc:year:rep (exact F-test)
    Numerator MS: loc:year:rep
    Denominator MS: Residual
```

2.2 Hierarchical effects: Regions and villages

M01: RLab-AppStat/v14-u31.R

```
> data("v-lmm-soils", package = "SelectionTools")
> st.lmm(
+     fixed = yield ~ region,
+     random = ~region:village,
+     data = soils
+ )
```

```
Linear mixed model fitted by REML

Model:
  Data:       soils
  Fixed:      yield ~ region
  Random:     ~region:village
  Residual:   ~units

Convergence:
  converged: projected score tolerance met

Fixed effects (exact F):
  Term    DF1  DF2       F       p
  region    2    3  2.7857  0.2071

Random effects (exact F-tests):
  Term            DF1  DF2        F        p
  region:village    3   18  12.9231  <0.0001

Variance components:
  Component       Variance
  region:village    8.6111
  Residual          2.8889

Fit and model statistics:
  Observations                24    R-squared (conditional)  0.8594
  Log-likelihood         -47.8945    RMSE (conditional)       1.4814
  AIC                    105.7891    Variance: fixed          9.0435
  BIC                    111.6794    Variance: random         8.6111
  R-squared (marginal)     0.4402    Variance: residual       2.8889
```

2.3 Multi-environment trials: Rapeseed

M02: RLab-AppStat/v14-u31.R

```
> data("v-lmm-rapeseed", package = "SelectionTools")
> m <- st.lmm(
+     fixed = hgt ~ geno * env,
+     random = ~env:rep,
+     data = rapeseed
+ )
> m
```

```
Linear mixed model fitted by REML

Model:
  Data:       rapeseed
  Fixed:      hgt ~ geno * env
  Random:     ~env:rep
  Residual:   ~units
```

```
Convergence:
  converged: projected score tolerance met

Fixed effects (exact F):
  Term      DF1  DF2       F        p
  geno       19   76  6.6832  <0.0001
  env         1    4 51.5051   0.0020
  geno:env   19   76  0.7045   0.8031

Random effects (exact F-tests):
  Term     DF1  DF2      F       p
  env:rep    4   76 2.0433  0.0967

Variance components:
  Component  Variance
  env:rep      4.6939
  Residual    89.9811

Fit and model statistics:
  Observations               120    R-squared (conditional)    0.6791
  Log-likelihood        -316.9005    RMSE (conditional)         7.6456
  AIC                    717.8009    Variance: fixed          185.7132
  BIC                    834.8756    Variance: random           4.6939
  R-squared (marginal)     0.6623    Variance: residual        89.9811
```

2.3.1 Genotype means and standard errors

The following calculations use the rapeseed model fitted above. `M03:  RLab-AppStat/v14-u31.R`

```
> lsmeans.G <- st.lsm(m, "geno")
> lsmeans.G
```

```
Least-squares means (Satterthwaite)
  geno   lsmeans      SE     DF  lower.CL  upper.CL
  101   215.8333  3.9723  76.43  207.9225  223.7441
  102   232.1667  3.9723  76.43  224.2559  240.0775
  103   212.8333  3.9723  76.43  204.9225  220.7441
  104   223.6667  3.9723  76.43  215.7559  231.5775
  105   207.8333  3.9723  76.43  199.9225  215.7441
  106   229.1667  3.9723  76.43  221.2559  237.0775
  107   209.6667  3.9723  76.43  201.7559  217.5775
  108   229.3333  3.9723  76.43  221.4225  237.2441
  109   221.3333  3.9723  76.43  213.4225  229.2441
  110   211.8333  3.9723  76.43  203.9225  219.7441
  [5 table rows omitted]
  116   212.8333  3.9723  76.43  204.9225  220.7441
  117   195.6667  3.9723  76.43  187.7559  203.5775
  118   225.8333  3.9723  76.43  217.9225  233.7441
  119   214.1667  3.9723  76.43  206.2559  222.0775
  120   197.8333  3.9723  76.43  189.9225  205.7441
Confidence level: 0.9500
LSD=10.9077 HSD=20.1085 (alpha=0.05)
```

```
> data.frame(lsmeans.G)$SE[1]
```

```
[1] 3.9723
```

```
> data.frame(lsmeans.G)$DF[1]
```

```
[1] 76.43048
```

```
> st.lsm(
+     m,
+     "env",
+     by = "geno"
+ )
```

```
Least-squares means (Satterthwaite)
  env  geno   lsmeans      SE     DF  lower.CL  upper.CL
  e1   101   209.0000  5.6177  76.43  197.8124  220.1876
  e2   101   222.6667  5.6177  76.43  211.4791  233.8542
  e1   102   217.0000  5.6177  76.43  205.8124  228.1876
  e2   102   247.3333  5.6177  76.43  236.1458  258.5209
  e1   103   204.3333  5.6177  76.43  193.1458  215.5209
  e2   103   221.3333  5.6177  76.43  210.1458  232.5209
  e1   104   213.3333  5.6177  76.43  202.1458  224.5209
  e2   104   234.0000  5.6177  76.43  222.8124  245.1876
  e1   105   200.3333  5.6177  76.43  189.1458  211.5209
  e2   105   215.3333  5.6177  76.43  204.1458  226.5209
  [25 table rows omitted]
  e2   118   230.0000  5.6177  76.43  218.8124  241.1876
  e1   119   205.0000  5.6177  76.43  193.8124  216.1876
  e2   119   223.3333  5.6177  76.43  212.1458  234.5209
  e1   120   181.3333  5.6177  76.43  170.1458  192.5209
  e2   120   214.3333  5.6177  76.43  203.1458  225.5209
Confidence level: 0.9500
Within geno: LSD=15.8216 HSD=15.8216 (alpha=0.05)
```

2.4 Crossed random effects: Years and locations

M23: barley-met

```
> data("v-lmm-barley-met", package = "SelectionTools")
> m.m23.reml <- st.lmm(
+     fixed = TKW ~ 1,
+     random = ~gen + loc + year + gen:loc + gen:year + loc:year +
+         gen:loc:year + rep:loc:year,
+     data = barley.met
+ )
> m.m23.reml
```

```
Linear mixed model fitted by REML

Model:
  Data:      barley.met
  Fixed:     TKW ~ 1
  Random:    ~gen + loc + year + gen:loc + gen:year + loc:year +
               gen:loc:year + rep:loc:year
  Residual:  ~units

Convergence:
  converged: projected score tolerance met

Random effects (likelihood-ratio tests):
```

```
  Term          DF       LR        p
  gen            1   7.1922   0.0073
  loc            1   9.1164   0.0025
  year           1   0.1945   0.6592
  gen:loc        1  30.3730  <0.0001
  gen:year       1  95.8428  <0.0001
  loc:year       1  48.2637  <0.0001
  gen:loc:year   1  61.2318  <0.0001
  loc:year:rep   1  13.8315   0.0002

Variance components:
  Component     Variance
  gen            11.7252
  loc            27.9921
  year            1.2830
  gen:loc         4.5835
  gen:year       13.0123
  loc:year        6.7801
  gen:loc:year    3.6393
  loc:year:rep    0.5908
  Residual        8.2587

Fit and model statistics:
  Observations                900    R-squared (conditional)    0.8939
  Log-likelihood        -2481.2768    RMSE (conditional)         2.5332
  AIC                    4982.5535    Variance: fixed            0.0000
  BIC                    5030.5775    Variance: random          69.6061
  R-squared (marginal)      0.0000    Variance: residual         8.2587
```

2.5 Separate genetic variances for material groups

M30: RLab-PopGen/q046-05-s1.R

```
> data("v-lmm-material-groups", package = "SelectionTools")
> st.lmm(
+     fixed = yield ~ 1,
+     random = ~at(grp):gen + env,
+     data = material.groups
+ )
```

```
Linear mixed model fitted by REML

Model:
  Data:      material.groups
  Fixed:     yield ~ 1
  Random:    ~at(grp):gen + env
  Residual:  ~units

Convergence:
  converged: projected score tolerance met

Random effects (likelihood-ratio tests):
  Term          DF        LR        p
  at(grp):gen    2  166.8162  <0.0001
  env            1  151.1080  <0.0001

Variance components:
  Component           Variance
  at(grp):gen[local]    7.8923
  at(grp):gen[ref]     49.5978
  env                  13.5615
  Residual             25.0890

Fit and model statistics:
```

```
  Observations                 464    R-squared (conditional)    0.5782
  Log-likelihood         -1467.9796   RMSE (conditional)         4.7164
  AIC                     2945.9592   Variance: fixed            0.0000
  BIC                     2966.6586   Variance: random          34.3968
  R-squared (marginal)       0.0000   Variance: residual        25.0890
```

2.6 Multi-environment trials: Balanced and unbalanced

M37: barley-met

The supplied barley data form a balanced reference for genotype means. The table counts usable TKW records by location and year. Each of the 15 environments contains 15 genotypes in four replicates, so the means can be estimated across all 15 environments.

```
> data("v-lmm-barley-met", package = "SelectionTools")
> xtabs(~loc + year, data = barley.met, subset = !is.na(TKW))
```

```
     year
loc   y2016 y2017 y2018
  HOH    60    60    60
  EWE    60    60    60
  EIN    60    60    60
  OLI    60    60    60
  HFH    60    60    60
```

```
> m <- st.lmm(
+     fixed = TKW ~ gen + loc + year + loc:year + year:loc:rep,
+     random = ~gen:loc + gen:year + gen:loc:year,
+     data = barley.met
+ )
> m
```

```
Linear mixed model fitted by REML

Model:
  Data:      barley.met
  Fixed:     TKW ~ gen + loc + year + loc:year + year:loc:rep
  Random:    ~gen:loc + gen:year + gen:loc:year
  Residual:  ~units

Convergence:
  converged: projected score tolerance met

Fixed effects (term-specific tests):
  Term          DF1    DF2        F        p  Method
  gen            14  38.42   3.0810   0.0029  Satterthwaite
  loc             4     56  71.0903  <0.0001  exact F
  year            2     28   3.8282   0.0339  exact F
  loc:year        8    112  19.2184  <0.0001  exact F
  loc:year:rep   45    630   2.0730  <0.0001  exact F

Random effects (exact F-tests):
  Term          DF1  DF2        F        p
  gen:loc        56  112   3.4107  <0.0001
  gen:year       28  112  12.4064  <0.0001
  gen:loc:year  112  630   2.7626  <0.0001

Variance components:
  Component     Variance
  gen:loc         4.5835
```

```
  gen:year       13.0123
  gen:loc:year    3.6393
  Residual        8.2587

Fit and model statistics:
  Observations                 900    R-squared (conditional)    0.8934
  Log-likelihood         -2321.7376    RMSE (conditional)         2.4930
  AIC                     4799.4753    Variance: fixed           48.0042
  BIC                     5174.0621    Variance: random          21.2350
  R-squared (marginal)       0.6194    Variance: residual         8.2587
```

The option `print.test.construction` adds details about each test to the model printout. The exact F-tests identify their numerator and denominator mean squares. The test for `gen` reports the Satterthwaite denominator DF.

```
> st.options(list(print.test.construction = TRUE))
> m
```

```
Linear mixed model fitted by REML

Model:
  Data:      barley.met
  Fixed:     TKW ~ gen + loc + year + loc:year + year:loc:rep
  Random:    ~gen:loc + gen:year + gen:loc:year
  Residual:  ~units

Convergence:
  converged: projected score tolerance met

Fixed effects (Type 3, term-specific tests):
  Term          DF1    DF2        F        p  Method
  gen            14  38.42   3.0810   0.0029  Satterthwaite
  loc             4     56  71.0903  <0.0001  exact F
  year            2     28   3.8282   0.0339  exact F
  loc:year        8    112  19.2184  <0.0001  exact F
  loc:year:rep   45    630   2.0730  <0.0001  exact F

Random effects (exact F-tests):
  Term          DF1  DF2        F        p
  gen:loc        56  112   3.4107  <0.0001
  gen:year       28  112  12.4064  <0.0001
  gen:loc:year  112  630   2.7626  <0.0001

Variance components:
  Component     Variance
  gen:loc         4.5835
  gen:year       13.0123
  gen:loc:year    3.6393
  Residual        8.2587

Test construction:

  gen (Satterthwaite)
    Statistic: Wald F for the fixed-effect hypothesis.
    Denominator DF: 38.42

  loc (exact F)
    Numerator MS: loc
    Denominator MS: gen:loc

  year (exact F)
    Numerator MS: year
    Denominator MS: gen:year

  loc:year (exact F)
```

```
    Numerator MS: loc:year
    Denominator MS: gen:loc:year

  loc:year:rep (exact F)
    Numerator MS: loc:year:rep
    Denominator MS: Residual

  gen:loc (exact F)
    Numerator MS: gen:loc
    Denominator MS: gen:loc:year

  gen:year (exact F)
    Numerator MS: gen:year
    Denominator MS: gen:loc:year

  gen:loc:year (exact F)
    Numerator MS: gen:loc:year
    Denominator MS: Residual

Fit and model statistics:
  Observations                 900    R-squared (conditional)    0.8934
  Log-likelihood         -2321.7376    RMSE (conditional)         2.4930
  AIC                     4799.4753    Variance: fixed           48.0042
  BIC                     5174.0621    Variance: random          21.2350
  R-squared (marginal)       0.6194    Variance: residual         8.2587
```

```
> st.lsm(m, "gen")
```

```
Least-squares means (Satterthwaite)
  gen    lsmeans      SE     DF  lower.CL  upper.CL
  lne01  34.3150  2.3737  38.42   29.5115   39.1185
  lne02  44.9733  2.3737  38.42   40.1698   49.7769
  lne03  47.3033  2.3737  38.42   42.4998   52.1069
  lne04  40.3278  2.3737  38.42   35.5242   45.1313
  lne05  44.2333  2.3737  38.42   39.4298   49.0369
  lne06  42.0650  2.3737  38.42   37.2615   46.8685
  lne07  41.9983  2.3737  38.42   37.1948   46.8019
  lne08  41.9000  2.3737  38.42   37.0965   46.7035
  lne09  44.4067  2.3737  38.42   39.6031   49.2102
  lne10  36.5867  2.3737  38.42   31.7831   41.3902
  lne11  44.4083  2.3737  38.42   39.6048   49.2119
  lne12  45.3417  2.3737  38.42   40.5381   50.1452
  lne13  35.6533  2.3737  38.42   30.8498   40.4569
  lne14  45.7200  2.3737  38.42   40.9165   50.5235
  lne15  48.0467  2.3737  38.42   43.2431   52.8502
Confidence level: 0.9500
LSD=6.7932 HSD=12.1498 (alpha=0.05)
```

2.6.1 Environment model for genotype means

`M39: RLab-PopGen/q046-10-s3.R`

Two location–year combinations are absent. Each of the ten observed environments has usable yields for 15 genotypes in three replicates. The model uses one `env` level per observed environment, so the genotype means are estimable across those ten environments.

```
> data("v-lmm-met-unbalanced", package = "SelectionTools")
> xtabs(~loc + year, data = met.unbalanced, subset = !is.na(yld))
```

```
     year
loc   2015 2016
  ASE   45   45
  BOV    0   45
  MOS   45   45
  NIE    0   45
  RHH   45   45
  ROS   45   45
```

```
> m <- st.lmm(
+     fixed = yld ~ gen + env + rep:env,
+     random = ~gen:env,
+     data = met.unbalanced
+ )
> m
```

```
Linear mixed model fitted by REML

Model:
  Data:      met.unbalanced
  Fixed:     yld ~ gen + env + rep:env
  Random:    ~gen:env
  Residual:  ~units

Convergence:
  converged: projected score tolerance met

Fixed effects (exact F):
  Term     DF1  DF2        F        p
  gen       14  126  15.8462  <0.0001
  env        9  126  71.0107  <0.0001
  env:rep   20  280   5.0004  <0.0001

Random effects (exact F-tests):
  Term     DF1  DF2       F        p
  gen:env  126  280  2.3078  <0.0001

Variance components:
  Component  Variance
  gen:env      0.0418
  Residual     0.0959

Fit and model statistics:
  Observations                450    R-squared (conditional)  0.8356
  Log-likelihood        -215.8535    RMSE (conditional)       0.2670
  AIC                    523.7070    Variance: fixed          0.4456
  BIC                    712.7324    Variance: random         0.0418
  R-squared (marginal)     0.7640    Variance: residual       0.0959
```

```
> st.lsm(m, "gen")
```

```
Least-squares means (Satterthwaite)
  gen  lsmeans      SE   DF  lower.CL  upper.CL
  g01   4.6640  0.0859  126    4.4941    4.8339
  g10   4.1390  0.0859  126    3.9691    4.3089
  g11   4.7310  0.0859  126    4.5611    4.9009
  g12   3.7853  0.0859  126    3.6154    3.9553
  g14   4.7610  0.0859  126    4.5911    4.9309
  g15   4.0917  0.0859  126    3.9217    4.2616
  g02   4.6883  0.0859  126    4.5184    4.8583
```

```
  g20   4.5367  0.0859  126    4.3667    4.7066
  g21   4.6240  0.0859  126    4.4541    4.7939
  g23   4.3810  0.0859  126    4.2111    4.5509
  g03   4.0033  0.0859  126    3.8334    4.1733
  g05   4.1937  0.0859  126    4.0237    4.3636
  g07   4.1563  0.0859  126    3.9864    4.3263
  g08   4.5823  0.0859  126    4.4124    4.7523
  g09   4.9897  0.0859  126    4.8197    5.1596
Confidence level: 0.9500
LSD=0.2403 HSD=0.4202 (alpha=0.05)
```

2.6.2 Missing environment combinations

M41: RLab-PopGen/q046-10-s3.R

The table counts usable yields. Zeros mark absent trials, and counts of 79 indicate a missing yield among 80 plots. With separate location and year terms, the default grid includes absent combinations, so the genotype means are non-estimable. The genotype test remains estimable.

```
> data("v-lmm-met-missing", package = "SelectionTools")
> xtabs(~loc + year, data = met.missing, subset = !is.na(YLD))
```

```
      year
loc    y2017 y2018 y2019
  loc1     0     0    80
  loc2    80    80    80
  loc3    80    80    79
  loc4     0     0    80
  loc5     0     0    79
```

```
> m <- st.lmm(
+     fixed = YLD ~ gen + loc + year + loc:year + year:loc:rep,
+     random = ~gen:loc + gen:year + gen:loc:year,
+     data = met.missing
+ )
> m
```

```
Linear mixed model fitted by REML

Model:
  Data:      met.missing
  Fixed:     YLD ~ gen + loc + year + loc:year + year:loc:rep
  Random:    ~gen:loc + gen:year + gen:loc:year
  Residual:  ~units

Convergence:
  converged: projected score tolerance met

Fixed effects (Satterthwaite):
  Term          DF1     DF2         F        p
  gen            19   12.49    2.4448   0.0550
  loc             1   18.85  579.9438  <0.0001
  year            0      NA        NA       NA
  loc:year        2   53.46   25.5857  <0.0001
  loc:year:rep   27  511.51    7.9383  <0.0001
  loc: partially estimable
  year: non-estimable
  loc:year: partially estimable
  loc:year:rep: partially estimable
```

```
Random effects (likelihood-ratio tests):
  Term          DF      LR       p
  gen:loc        1  0.0300  0.8625
  gen:year       1  0.0200  0.8876
  gen:loc:year   1  9.6138  0.0019

Variance components:
  Component        Variance
  gen:loc         8181.0797
  gen:year        4473.0448
  gen:loc:year  151627.4081
  Residual      631781.8069

Fit and model statistics:
  Observations                718    R-squared (conditional)        0.7441
  Log-likelihood        -5508.1201    RMSE (conditional)           718.2940
  AIC                   11134.2403    Variance: fixed          1672388.4135
  BIC                   11404.2520    Variance: random          164281.5326
  R-squared (marginal)      0.6775    Variance: residual        631781.8069
```

```
> st.lsm(m, "gen")
```

```
Least-squares means (Satterthwaite)
  gen    lsmeans  SE  DF  lower.CL  upper.CL
  lne01       NA  NA  NA        NA        NA
  lne02       NA  NA  NA        NA        NA
  lne03       NA  NA  NA        NA        NA
  lne04       NA  NA  NA        NA        NA
  lne05       NA  NA  NA        NA        NA
  lne06       NA  NA  NA        NA        NA
  lne07       NA  NA  NA        NA        NA
  lne08       NA  NA  NA        NA        NA
  lne09       NA  NA  NA        NA        NA
  lne10       NA  NA  NA        NA        NA
  [5 table rows omitted]
  lne16       NA  NA  NA        NA        NA
  lne17       NA  NA  NA        NA        NA
  lne18       NA  NA  NA        NA        NA
  lne19       NA  NA  NA        NA        NA
  lne20       NA  NA  NA        NA        NA
Confidence level: 0.9500
```

2.6.3 Combined field factor

M42: RLab-PopGen/q046-10-s3.R

The same yield data are analysed with one fixed factor, REP, for the observed location–year–replicate blocks. Averaging over these blocks gives estimable genotype means. Both models give the same genotype test.

```
> data("v-lmm-met-missing", package = "SelectionTools")
> xtabs(~loc + year, data = met.missing, subset = !is.na(YLD))
```

```
      year
loc    y2017 y2018 y2019
  loc1     0     0    80
  loc2    80    80    80
  loc3    80    80    79
  loc4     0     0    80
  loc5     0     0    79
```

```
> met.missing$REP <- met.missing$rep:met.missing$year:met.missing$loc
> m <- st.lmm(
+     fixed = YLD ~ gen + REP,
+     random = ~gen:loc + gen:year + gen:loc:year,
+     data = met.missing
+ )
> m
```

```
Linear mixed model fitted by REML

Model:
  Data:      met.missing
  Fixed:     YLD ~ gen + REP
  Random:    ~gen:loc + gen:year + gen:loc:year
  Residual:  ~units

Convergence:
  converged: projected score tolerance met

Fixed effects (Satterthwaite):
  Term  DF1     DF2        F        p
  gen    19   12.49   2.4448   0.0550
  REP    35  160.94  27.7702  <0.0001
  REP: partially estimable

Random effects (likelihood-ratio tests):
  Term          DF      LR       p
  gen:loc        1  0.0300  0.8625
  gen:year       1  0.0200  0.8876
  gen:loc:year   1  9.6138  0.0019

Variance components:
  Component        Variance
  gen:loc         8181.0797
  gen:year        4473.0448
  gen:loc:year  151627.4081
  Residual      631781.8069

Fit and model statistics:
  Observations                 718    R-squared (conditional)        0.7441
  Log-likelihood       -5508.1201    RMSE (conditional)           718.2940
  AIC                  11134.2403    Variance: fixed          1672388.4135
  BIC                  11404.2520    Variance: random          164281.5326
  R-squared (marginal)     0.6775    Variance: residual        631781.8069
```

```
> st.lsm(m, "gen")
```

```
Least-squares means (Satterthwaite)
  gen      lsmeans        SE     DF   lower.CL   upper.CL
  lne01  3669.6035  195.7737  12.43  3244.6707  4094.5364
  lne02  2777.6565  195.7737  12.43  2352.7237  3202.5893
  lne03  3184.8884  195.7737  12.43  2759.9556  3609.8213
  lne04  3293.0763  195.7737  12.43  2868.1435  3718.0091
  lne05  3444.3160  195.7737  12.43  3019.3832  3869.2488
  lne06  3220.0430  195.7737  12.43  2795.1102  3644.9758
  lne07  3490.8089  195.7737  12.43  3065.8761  3915.7417
  lne08  3338.0336  195.7737  12.43  2913.1008  3762.9665
  lne09  3361.2773  197.1876  13.13  2935.7029  3786.8517
  lne10  3449.1551  195.7737  12.43  3024.2222  3874.0879
  [5 table rows omitted]
  lne16  3625.0485  195.7737  12.43  3200.1157  4049.9813
  lne17  3402.6488  195.7737  12.43  2977.7159  3827.5816
```

```
  lne18  2822.5368  195.7737  12.43  2397.6040  3247.4696
  lne19  3140.4759  195.7737  12.43  2715.5430  3565.4087
  lne20  2946.4855  195.7737  12.43  2521.5527  3371.4183
Confidence level: 0.9500
```

2.7 Inference spaces for machines and operators

Two machine types were evaluated twice by each of three operators drawn from a population of operators. The scores come from the example presented by Sanders (1991). The model includes fixed machine effects, random operator effects and a random machine-by-operator interaction.

```
> data("v-lmm-machines", package = "SelectionTools")
> m.machines <- st.lmm(
+     fixed = score ~ machine,
+     random = ~operator + machine:operator,
+     data = machines
+ )
```

The inference space determines which population each machine mean describes (McLean et al., 1991). The default `inference.space = "broad"` gives a mean for each machine type averaged over the operator population and machine-by-operator interactions.

```
> st.lsm(m.machines, ~machine)
```

```
Least-squares means (Satterthwaite)
  machine  lsmeans      SE    DF  lower.CL  upper.CL
  1        50.9483  0.2467  2.97   50.1591   51.7376
  2        51.9567  0.2467  2.97   51.1674   52.7459
Confidence level: 0.9500
LSD=0.9638 HSD=0.9638 (alpha=0.05)
```

The first intermediate space describes the three operators in the experiment and averages over the machine-by-operator interactions. The second describes the machine-by-operator combinations in the experiment and averages over the operator main effects.

```
> st.lsm(
+     m.machines,
+     ~machine,
+     inference.space = ~operator
+ )
```

```
Least-squares means (Satterthwaite)
Inference space: intermediate
  machine  lsmeans      SE  DF  lower.CL  upper.CL
  1        50.9483  0.1584   2   50.2668   51.6299
  2        51.9567  0.1584   2   51.2751   52.6382
Confidence level: 0.9500
LSD=0.9638 HSD=0.9638 (alpha=0.05)
```

```
> st.lsm(
+     m.machines,
+     ~machine,
+     inference.space = ~machine:operator
+ )
```

```
Least-squares means (Satterthwaite)
Inference space: intermediate
  machine  lsmeans      SE    DF  lower.CL  upper.CL
  1        50.9483  0.2094  1.53   49.7229   52.1737
  2        51.9567  0.2094  1.53   50.7313   53.1821
Confidence level: 0.9500
LSD=0.3112 HSD=0.3112 (alpha=0.05)
```

Narrow inference gives each machine's mean score for these three operators, including how each operator performed with that machine. In this balanced experiment, the SE depends on the residual variance and the six observations per machine.

```
> st.lsm(
+     m.machines,
+     ~machine,
+     inference.space = "narrow"
+ )
```

```
Least-squares means (Satterthwaite)
Inference space: narrow
  machine  lsmeans      SE  DF  lower.CL  upper.CL
  1        50.9483  0.0899   6   50.7283   51.1684
  2        51.9567  0.0899   6   51.7366   52.1767
Confidence level: 0.9500
LSD=0.3112 HSD=0.3112 (alpha=0.05)
```

All four estimates of the first machine mean are about 50.95 in this balanced experiment. The corresponding published SEs are 0.2467 for broad inference, 0.1584 for the intermediate space including operators, 0.2095 for the intermediate space including interactions, and 0.0899 for narrow inference (Sanders, 1991). The fitted interaction SE rounds to 0.2094 when calculated from the full precision variance estimates.

3 Experimental Designs

3.1 Moment estimators

3.1.1 Completely randomized design: Fertilizers

CM04: RLab-AppStat/v14-u22_empty.R

```
> data("v-lmm-fertilizer", package = "SelectionTools")
> m.fertilizer <- st.lmm(
+     yield ~ fert,
+     data = fertilizer,
+     method = "moments"
+ )
> m.fertilizer
```

```
Linear model fitted by the method of moments

Model:
  Data:      fertilizer
  Fixed:     yield ~ fert

ANOVA (method of moments, exact F):
  Source    DF1  DF2         SS        MS       F       p
  fert        4   20  3381.4623  845.3656  5.9979  0.0024
  Residual   20       2818.8939  140.9447
```

```
Model statistics:
  Trait        N  Residual DF      R2  Adjusted R2  Residual SD
  yield  25.0000           20  0.5454       0.4544      11.8720
```

```
> st.lsm(m.fertilizer, ~fert)
```

```
Least-squares means (exact)
  fert     lsmeans      SE  DF  lower.CL  upper.CL
   CCr-1   18.2820  5.3093  20    7.2069   29.3571
   CCr-2   31.6480  5.3093  20   20.5729   42.7231
   Man     51.6420  5.3093  20   40.5669   62.7171
   Min-A   22.1580  5.3093  20   11.0829   33.2331
   Min-B   27.4200  5.3093  20   16.3449   38.4951
Confidence level: 0.9500
LSD=15.6625 HSD=22.4683 (alpha=0.05)
```

```
> st.lsm(
+     m.fertilizer,
+     ~fert,
+     method = "pairwise",
+     adjust = "none"
+ )
```

```
Least-squares mean contrasts for fert (exact)
  contrast          estimate     SE DF lower.CL upper.CL       t      p
   CCr-1  -  CCr-2  -13.3660 7.5085 20 -29.0285   2.2965 -1.7801 0.0903
   CCr-1  -  Man    -33.3600 7.5085 20 -49.0225 -17.6975 -4.4430 0.0002
   CCr-1  -  Min-A   -3.8760 7.5085 20 -19.5385  11.7865 -0.5162 0.6114
   CCr-1  -  Min-B   -9.1380 7.5085 20 -24.8005   6.5245 -1.2170 0.2378
   CCr-2  -  Man    -19.9940 7.5085 20 -35.6565  -4.3315 -2.6628 0.0149
   CCr-2  -  Min-A    9.4900 7.5085 20  -6.1725  25.1525  1.2639 0.2208
   CCr-2  -  Min-B    4.2280 7.5085 20 -11.4345  19.8905  0.5631 0.5796
   Man  -  Min-A     29.4840 7.5085 20  13.8215  45.1465  3.9267 0.0008
   Man  -  Min-B     24.2220 7.5085 20   8.5595  39.8845  3.2259 0.0042
   Min-A  -  Min-B   -5.2620 7.5085 20 -20.9245  10.4005 -0.7008 0.4915
p-value adjustment: none; confidence level: 0.9500
```

```
> st.lsm(
+     m.fertilizer,
+     ~fert,
+     method = "pairwise",
+     adjust = "tukey"
+ )
```

```
Least-squares mean contrasts for fert (exact)
  contrast          estimate     SE DF lower.CL upper.CL       t      p
   CCr-1  -  CCr-2  -13.3660 7.5085 20 -35.8343   9.1023 -1.7801 0.4115
   CCr-1  -  Man    -33.3600 7.5085 20 -55.8283 -10.8917 -4.4430 0.0021
   CCr-1  -  Min-A   -3.8760 7.5085 20 -26.3443  18.5923 -0.5162 0.9847
   CCr-1  -  Min-B   -9.1380 7.5085 20 -31.6063  13.3303 -1.2170 0.7419
   CCr-2  -  Man    -19.9940 7.5085 20 -42.4623   2.4743 -2.6628 0.0961
   CCr-2  -  Min-A    9.4900 7.5085 20 -12.9783  31.9583  1.2639 0.7152
   CCr-2  -  Min-B    4.2280 7.5085 20 -18.2403  26.6963  0.5631 0.9789
   Man  -  Min-A     29.4840 7.5085 20   7.0157  51.9523  3.9267 0.0066
   Man  -  Min-B     24.2220 7.5085 20   1.7537  46.6903  3.2259 0.0307
   Min-A  -  Min-B   -5.2620 7.5085 20 -27.7303  17.2063 -0.7008 0.9539
p-value and interval adjustment: tukey; confidence level: 0.9500
```

3.1.2 Fixed complete blocks: Cotton

CM05: RLab-AppStat/v14-u41.R

```
> data("v-lmm-cotton", package = "SelectionTools")
> m.cotton <- st.lmm(
+     Strength ~ Blk + Fert,
+     data = cotton,
+     method = "moments"
+ )
> m.cotton
```

```
Linear model fitted by the method of moments

Model:
  Data:      cotton
  Fixed:     Strength ~ Blk + Fert

ANOVA (method of moments, exact F):
  Source    DF1  DF2      SS      MS       F       p
  Blk         2    8  0.0971  0.0486  1.1116  0.3750
  Fert        4    8  0.7324  0.1831  4.1916  0.0404
  Residual    8       0.3495  0.0437

Model statistics:
  Trait           N  Residual DF      R2  Adjusted R2  Residual SD
  Strength  15.0000            8  0.7036       0.4813       0.2090
```

```
> st.lsm(m.cotton, ~Fert)
```

```
Least-squares means (exact)
  Fert  lsmeans      SE  DF  lower.CL  upper.CL
  f1     7.8500  0.1207   8    7.5717    8.1283
  f2     8.0533  0.1207   8    7.7751    8.3316
  f3     7.7433  0.1207   8    7.4651    8.0216
  f4     7.5133  0.1207   8    7.2351    7.7916
  f5     7.4500  0.1207   8    7.1717    7.7283
Confidence level: 0.9500
LSD=0.3935 HSD=0.5896 (alpha=0.05)
```

```
> st.lsm(
+     m.cotton,
+     ~Fert,
+     method = "pairwise",
+     adjust = "tukey"
+ )
```

```
Least-squares mean contrasts for Fert (exact)
  contrast  estimate      SE  DF  lower.CL  upper.CL       t       p
  f1 - f2    -0.2033  0.1707   8   -0.7929    0.3862  -1.1915  0.7565
  f1 - f3     0.1067  0.1707   8   -0.4829    0.6962   0.6250  0.9667
  f1 - f4     0.3367  0.1707   8   -0.2529    0.9262   1.9728  0.3563
  f1 - f5     0.4000  0.1707   8   -0.1896    0.9896   2.3439  0.2246
  f2 - f3     0.3100  0.1707   8   -0.2796    0.8996   1.8165  0.4263
  f2 - f4     0.5400  0.1707   8   -0.0496    1.1296   3.1643  0.0743
  f2 - f5     0.6033  0.1707   8    0.0138    1.1929   3.5354  0.0448
  f3 - f4     0.2300  0.1707   8   -0.3596    0.8196   1.3477  0.6729
  f3 - f5     0.2933  0.1707   8   -0.2962    0.8829   1.7189  0.4740
  f4 - f5     0.0633  0.1707   8   -0.5262    0.6529   0.3711  0.9951
p-value and interval adjustment: tukey; confidence level: 0.9500
```

3.1.3 Fixed Latin square: Sugar beet

CM06: RLab-AppStat/v14-u41.R

```
> data("v-lmm-sugar", package = "SelectionTools")
> m.sugar <- st.lmm(
+     Yield ~ Row + Col + Variety,
+     data = sugar,
+     method = "moments"
+ )
> m.sugar
```

```
Linear model fitted by the method of moments

Model:
  Data:      sugar
  Fixed:     Yield ~ Row + Col + Variety

ANOVA (method of moments, exact F):
  Source     DF1  DF2       SS      MS        F        p
  Row          5   20   2.2922  0.4584   2.5453   0.0614
  Col          5   20   1.2556  0.2511   1.3942   0.2687
  Variety      5   20  12.6222  2.5244  14.0160  <0.0001
  Residual    20        3.6022  0.1801

Model statistics:
  Trait         N  Residual DF      R2  Adjusted R2  Residual SD
  Yield  36.0000           20  0.8178       0.6812       0.4244
```

```
> st.lsm(m.sugar, ~Variety)
```

```
Least-squares means (exact)
  Variety  lsmeans      SE  DF  lower.CL  upper.CL
  v1       16.4333  0.1733  20   16.0719   16.7947
  v2       17.0167  0.1733  20   16.6553   17.3781
  v3       16.4333  0.1733  20   16.0719   16.7947
  v4       16.2333  0.1733  20   15.8719   16.5947
  v5       17.9833  0.1733  20   17.6219   18.3447
  v6       16.5333  0.1733  20   16.1719   16.8947
Confidence level: 0.9500
LSD=0.5111 HSD=0.7702 (alpha=0.05)
```

```
> st.lsm(
+     m.sugar,
+     ~Variety,
+     method = "pairwise",
+     adjust = "none"
+ )
```

```
Least-squares mean contrasts for Variety (exact)
  contrast  estimate      SE  DF  lower.CL  upper.CL         t        p
  v1 - v2    -0.5833  0.2450  20   -1.0944   -0.0722   -2.3807   0.0273
  v1 - v3   >-0.0001  0.2450  20   -0.5111    0.5111  >-0.0001   1.0000
  v1 - v4     0.2000  0.2450  20   -0.3111    0.7111    0.8162   0.4240
  v1 - v5    -1.5500  0.2450  20   -2.0611   -1.0389   -6.3259  <0.0001
  v1 - v6    -0.1000  0.2450  20   -0.6111    0.4111   -0.4081   0.6875
  v2 - v3     0.5833  0.2450  20    0.0722    1.0944    2.3807   0.0273
  v2 - v4     0.7833  0.2450  20    0.2722    1.2944    3.1970   0.0045
  v2 - v5    -0.9667  0.2450  20   -1.4778   -0.4556   -3.9452   0.0008
```

```
  v2 - v6     0.4833  0.2450  20   -0.0278    0.9944    1.9726   0.0625
  v3 - v4     0.2000  0.2450  20   -0.3111    0.7111    0.8162   0.4240
  v3 - v5    -1.5500  0.2450  20   -2.0611   -1.0389   -6.3259  <0.0001
  v3 - v6    -0.1000  0.2450  20   -0.6111    0.4111   -0.4081   0.6875
  v4 - v5    -1.7500  0.2450  20   -2.2611   -1.2389   -7.1421  <0.0001
  v4 - v6    -0.3000  0.2450  20   -0.8111    0.2111   -1.2244   0.2350
  v5 - v6     1.4500  0.2450  20    0.9389    1.9611    5.9178  <0.0001
p-value adjustment: none; confidence level: 0.9500
```

3.1.4 Blocked factorial experiment: Cowpea

CM07: RLab-AppStat/v14-u41.R

```
> data("v-lmm-peas", package = "SelectionTools")
> m.peas <- st.lmm(
+     Yield ~ Block + Spacing + Variety + Spacing:Variety,
+     data = peas,
+     method = "moments"
+ )
> m.peas
```

```
Linear model fitted by the method of moments

Model:
  Data:      peas
  Fixed:     Yield ~ Block + Spacing + Variety + Spacing:Variety

ANOVA (method of moments, exact F):
  Source            DF1  DF2         SS        MS        F        p
  Block               3   24   255.6389   85.2130   4.8221   0.0091
  Spacing             2   24   155.0556   77.5278   4.3872   0.0238
  Variety             2   24  1027.3889  513.6944  29.0694  <0.0001
  Spacing:Variety     4   24   765.4444  191.3611  10.8289  <0.0001
  Residual           24        424.1111   17.6713

Model statistics:
  Trait        N  Residual DF      R2  Adjusted R2  Residual SD
  Yield  36.0000           24  0.8386       0.7646       4.2037
```

```
> st.lsm(m.peas, ~Spacing | Variety)
```

```
Least-squares means (exact)
  Spacing  Variety  lsmeans      SE  DF  lower.CL  upper.CL
  s04      v1       47.5000  2.1019  24   43.1620   51.8380
  s08      v1       50.7500  2.1019  24   46.4120   55.0880
  s12      v1       55.7500  2.1019  24   51.4120   60.0880
  s04      v2       62.2500  2.1019  24   57.9120   66.5880
  s08      v2       58.5000  2.1019  24   54.1620   62.8380
  s12      v2       52.2500  2.1019  24   47.9120   56.5880
  s04      v3       56.0000  2.1019  24   51.6620   60.3380
  s08      v3       64.2500  2.1019  24   59.9120   68.5880
  s12      v3       73.0000  2.1019  24   68.6620   77.3380
Confidence level: 0.9500
Within Variety: LSD=6.1349 HSD=7.4231 (alpha=0.05)
```

```
> st.lsm(
+     m.peas,
+     ~Spacing | Variety,
```

```
+     method = "pairwise",
+     adjust = "none"
+ )
```

```
Least-squares mean contrasts for Spacing (exact)
  contrast  Variety estimate     SE DF lower.CL upper.CL       t       p
  s04 - s08 v1       -3.2500 2.9725 24  -9.3849   2.8849 -1.0934  0.2851
  s04 - s12 v1       -8.2500 2.9725 24 -14.3849  -2.1151 -2.7755  0.0105
  s08 - s12 v1       -5.0000 2.9725 24 -11.1349   1.1349 -1.6821  0.1055
  s04 - s08 v2        3.7500 2.9725 24  -2.3849   9.8849  1.2616  0.2192
  s04 - s12 v2       10.0000 2.9725 24   3.8651  16.1349  3.3642  0.0026
  s08 - s12 v2        6.2500 2.9725 24   0.1151  12.3849  2.1026  0.0462
  s04 - s08 v3       -8.2500 2.9725 24 -14.3849  -2.1151 -2.7755  0.0105
  s04 - s12 v3      -17.0000 2.9725 24 -23.1349 -10.8651 -5.7191 <0.0001
  s08 - s12 v3       -8.7500 2.9725 24 -14.8849  -2.6151 -2.9437  0.0071
p-value adjustment: none; confidence level: 0.9500
```

3.1.5 Split-plot moments: Alfalfa

CM08: RLab-AppStat/v14-u41.R

```
> data("v-lmm-alfalfa", package = "SelectionTools")
> m.alfalfa <- st.lmm(
+     fixed = Yield ~ Variety * Date,
+     random = ~Block + Block:Variety,
+     data = alfalfa,
+     method = "moments"
+ )
> m.alfalfa
```

```
Linear mixed model fitted by the method of moments

Model:
  Data:      alfalfa
  Fixed:     Yield ~ Variety * Date
  Random:    ~Block + Block:Variety

Design: balanced split plot

ANOVA (method of moments, exact F):
  Source         DF1  DF2      SS      MS        F        p
  Variety          2   10  0.1780  0.0890   0.6534   0.5412
  Date             3   45  1.9625  0.6542  23.3897  <0.0001
  Block            5   45  4.1498  0.8300  29.6758  <0.0001
  Variety:Date     6   45  0.2106  0.0351   1.2548   0.2973
  Variety:Block   10   45  1.3623  0.1362   4.8711  <0.0001
  Residual        45       1.2585  0.0280

Unconstrained moment variance estimates:
  Component      Estimate
  Block            0.0668
  Variety:Block    0.0271
  Residual         0.0280
```

```
> st.lsm(
+     m.alfalfa,
+     ~Variety,
+     method = "pairwise",
+     adjust = "none"
+ )
```

```
Least-squares mean contrasts for Variety (exact)
  contrast         estimate     SE DF lower.CL upper.CL       t      p
  Cossack - Ladack  -0.0946 0.1065 10  -0.3320   0.1428 -0.8877 0.3956
  Cossack - Ranger   0.0192 0.1065 10  -0.2182   0.2566  0.1799 0.8608
  Ladack - Ranger    0.1137 0.1065 10  -0.1237   0.3512  1.0676 0.3108
p-value adjustment: none; confidence level: 0.9500
```

```
> st.lsm(
+     m.alfalfa,
+     ~Date,
+     method = "pairwise",
+     adjust = "none"
+ )
```

```
Least-squares mean contrasts for Date (exact)
  contrast  estimate      SE  DF  lower.CL  upper.CL        t        p
  A - B       0.4406  0.0557  45    0.3283    0.5528   7.9030  <0.0001
  A - C       0.2067  0.0557  45    0.0944    0.3189   3.7073   0.0006
  A - D       0.0900  0.0557  45   -0.0223    0.2023   1.6145   0.1134
  B - C      -0.2339  0.0557  45   -0.3462   -0.1216  -4.1957   0.0001
  B - D      -0.3506  0.0557  45   -0.4628   -0.2383  -6.2885  <0.0001
  C - D      -0.1167  0.0557  45   -0.2289   -0.0044  -2.0929   0.0420
p-value adjustment: none; confidence level: 0.9500
```

```
> st.lsm(
+     m.alfalfa,
+     ~Date | Variety,
+     method = "pairwise",
+     adjust = "none"
+ )
```

```
Least-squares mean contrasts for Date (exact)
  contrast Variety estimate     SE DF lower.CL upper.CL       t       p
  A - B    Cossack   0.4633 0.0966 45   0.2689   0.6578  4.7987 <0.0001
  A - C    Cossack   0.1883 0.0966 45  -0.0061   0.3828  1.9506  0.0574
  A - D    Cossack   0.1217 0.0966 45  -0.0728   0.3161  1.2601  0.2141
  B - C    Cossack  -0.2750 0.0966 45  -0.4695  -0.0805 -2.8482  0.0066
  B - D    Cossack  -0.3417 0.0966 45  -0.5361  -0.1472 -3.5386  0.0009
  C - D    Cossack  -0.0667 0.0966 45  -0.2611   0.1278 -0.6905  0.4934
  A - B    Ladack    0.5683 0.0966 45   0.3739   0.7628  5.8862 <0.0001
  A - C    Ladack    0.2117 0.0966 45   0.0172   0.4061  2.1922  0.0336
  A - D    Ladack    0.0550 0.0966 45  -0.1395   0.2495  0.5696  0.5718
  B - C    Ladack   -0.3567 0.0966 45  -0.5511  -0.1622 -3.6940  0.0006
  [3 table rows omitted]
  A - C    Ranger    0.2200 0.0966 45   0.0255   0.4145  2.2785  0.0275
  A - D    Ranger    0.0933 0.0966 45  -0.1011   0.2878  0.9666  0.3389
  B - C    Ranger   -0.0700 0.0966 45  -0.2645   0.1245 -0.7250  0.4722
  B - D    Ranger   -0.1967 0.0966 45  -0.3911  -0.0022 -2.0369  0.0476
  C - D    Ranger   -0.1267 0.0966 45  -0.3211   0.0678 -1.3119  0.1962
p-value adjustment: none; confidence level: 0.9500
```

```
> st.lsm(m.alfalfa, ~Variety | Date)
```

```
Least-squares means (Satterthwaite)
  Variety  Date  lsmeans      SE     DF  lower.CL  upper.CL
  Cossack  A      1.7650  0.1371  12.51    1.4675    2.0625
```

```
  Ladack   A      1.8750  0.1371  12.51    1.5775    2.1725
  Ranger   A      1.7033  0.1371  12.51    1.4059    2.0008
  Cossack  B      1.3017  0.1371  12.51    1.0042    1.5991
  Ladack   B      1.3067  0.1371  12.51    1.0092    1.6041
  Ranger   B      1.4133  0.1371  12.51    1.1159    1.7108
  Cossack  C      1.5767  0.1371  12.51    1.2792    1.8741
  Ladack   C      1.6633  0.1371  12.51    1.3659    1.9608
  Ranger   C      1.4833  0.1371  12.51    1.1859    1.7808
  Cossack  D      1.6433  0.1371  12.51    1.3459    1.9408
  Ladack   D      1.8200  0.1371  12.51    1.5225    2.1175
  Ranger   D      1.6100  0.1371  12.51    1.3125    1.9075
Confidence level: 0.9500
Within Date: LSD=0.2795 HSD=0.3382 (alpha=0.05)
```

```
> c.variety.date <- st.lsm(
+     m.alfalfa,
+     ~Variety | Date,
+     method = "pairwise",
+     adjust = "none"
+ )
> c.variety.date
```

```
Least-squares mean contrasts for Variety (Satterthwaite)
  contrast Date estimate     SE    DF lower.CL upper.CL       t      p
  C - L    A     -0.1100 0.1354 24.08  -0.3895   0.1695 -0.8121 0.4247
  C - R    A      0.0617 0.1354 24.08  -0.2178   0.3412  0.4553 0.6530
  L - R    A      0.1717 0.1354 24.08  -0.1078   0.4512  1.2674 0.2171
  C - L    B     -0.0050 0.1354 24.08  -0.2845   0.2745 -0.0369 0.9709
  C - R    B     -0.1117 0.1354 24.08  -0.3912   0.1678 -0.8245 0.4178
  L - R    B     -0.1067 0.1354 24.08  -0.3862   0.1728 -0.7875 0.4386
  C - L    C     -0.0867 0.1354 24.08  -0.3662   0.1928 -0.6399 0.5283
  C - R    C      0.0933 0.1354 24.08  -0.1862   0.3728  0.6891 0.4973
  L - R    C      0.1800 0.1354 24.08  -0.0995   0.4595  1.3290 0.1963
  C - L    D     -0.1767 0.1354 24.08  -0.4562   0.1028 -1.3044 0.2044
  C - R    D      0.0333 0.1354 24.08  -0.2462   0.3128  0.2461 0.8077
  L - R    D      0.2100 0.1354 24.08  -0.0695   0.4895  1.5505 0.1341
  Variety: C = Cossack, L = Ladack, R = Ranger
p-value adjustment: none; confidence level: 0.9500
```

```
> st.lsm.test(c.variety.date)
```

```
Joint tests (Satterthwaite)
  Date  DF1    DF2       F       p  status
  A       2  24.08  0.8244  0.4505  reduced rank
  B       2  24.08  0.4338  0.6530  reduced rank
  C       2  24.08  0.8835  0.4263  reduced rank
  D       2  24.08  1.3886  0.2687  reduced rank
```

```
> # Mean order: Cossack/A, Ladack/A, Cossack/B, Ladack/B.
> C.alf <- rbind(
+     within.plot = c(1, 0, -1, 0),
+     across.plots = c(1, 0, 0, -1)
+ )
> # Within Cossack: A - B. Across varieties: Cossack/A - Ladack/B.
> h.alf.cells <- st.lsm(
+     m.alfalfa,
+     ~Variety * Date,
+     at = list(
```

```
+         Variety = c("Cossack", "Ladack"),
+         Date = c("A", "B")
+     ),
+     method = C.alf,
+     adjust = "none"
+ )
> h.alf.cells
```

```
Least-squares mean contrasts (row-specific DF)
At Variety = Cossack, Ladack
At Date = A, B
  contrast   estimate   SE    DF lower.CL upper.CL    t       p DF method
  within.p~1     0.46 0.10    45     0.27     0.66 4.80 <0.0001 exact
  across.p~2     0.46 0.14 24.08     0.18     0.74 3.38  0.0024 Satterthw~1
  within.p~1 [contrast] = within.plot
  across.p~2 [contrast] = across.plots
  Satterthw~1 [DF method] = Satterthwaite
p-value adjustment: none; confidence level: 0.9500
```

```
> # Joint test of the two contrasts, allowing their different error strata.
> st.lsm(
+     h.alf.cells,
+     joint = TRUE
+ )
```

```
Satterthwaite linear hypothesis
At Variety = Cossack, Ladack
At Date = A, B
  DF1    DF2        F        p  F.scaling  method
    2  41.95  13.1180  <0.0001     1.0000  Satterthwaite
```

3.2 Completely randomized design: Fertilizer contrasts

L01: RLab-AppStat/v14-u22_empty.R

```
> data("v-lmm-fertilizer", package = "SelectionTools")
> m <- st.lmm(
+     yield ~ fert,
+     data = fertilizer
+ )
> m
```

```
Linear model fitted by REML

Model:
  Data:      fertilizer
  Fixed:     yield ~ fert
  Random:    None
  Residual:  ~units

Convergence:
  converged: projected score tolerance met

Fixed effects (exact F):
  Term  DF1  DF2       F       p
  fert    4   20  5.9979  0.0024

Variance components:
```

```
  Component  Variance
  Residual   140.9447

Fit and model statistics:
  Observations               25    R-squared (conditional)    0.4999
  Log-likelihood        -81.8860    RMSE (conditional)        10.6187
  AIC                   175.7721    Variance: fixed          140.8943
  BIC                   183.0853    Variance: random           0.0000
  R-squared (marginal)    0.4999    Variance: residual       140.9447
```

```
> st.lsm(m, ~fert)
```

```
Least-squares means (Satterthwaite)
  fert     lsmeans      SE  DF  lower.CL  upper.CL
   CCr-1   18.2820  5.3093  20    7.2069   29.3571
   CCr-2   31.6480  5.3093  20   20.5729   42.7231
   Man     51.6420  5.3093  20   40.5669   62.7171
   Min-A   22.1580  5.3093  20   11.0829   33.2331
   Min-B   27.4200  5.3093  20   16.3449   38.4951
Confidence level: 0.9500
LSD=15.6625 HSD=22.4683 (alpha=0.05)
```

```
> st.lsm(
+     m,
+     ~fert,
+     method = "pairwise",
+     adjust = "none"
+ )
```

```
Least-squares mean contrasts for fert (Satterthwaite)
  contrast          estimate     SE DF lower.CL upper.CL       t      p
   CCr-1  -  CCr-2  -13.3660 7.5085 20 -29.0285   2.2965 -1.7801 0.0903
   CCr-1  -  Man    -33.3600 7.5085 20 -49.0225 -17.6975 -4.4430 0.0002
   CCr-1  -  Min-A   -3.8760 7.5085 20 -19.5385  11.7865 -0.5162 0.6114
   CCr-1  -  Min-B   -9.1380 7.5085 20 -24.8005   6.5245 -1.2170 0.2378
   CCr-2  -  Man    -19.9940 7.5085 20 -35.6565  -4.3315 -2.6628 0.0149
   CCr-2  -  Min-A    9.4900 7.5085 20  -6.1725  25.1525  1.2639 0.2208
   CCr-2  -  Min-B    4.2280 7.5085 20 -11.4345  19.8905  0.5631 0.5796
   Man  -  Min-A     29.4840 7.5085 20  13.8215  45.1465  3.9267 0.0008
   Man  -  Min-B     24.2220 7.5085 20   8.5595  39.8845  3.2259 0.0042
   Min-A  -  Min-B   -5.2620 7.5085 20 -20.9245  10.4005 -0.7008 0.4915
p-value adjustment: none; confidence level: 0.9500
```

```
> st.lsm(
+     m,
+     ~fert,
+     method = "pairwise",
+     adjust = "tukey"
+ )
```

```
Least-squares mean contrasts for fert (Satterthwaite)
  contrast          estimate     SE DF lower.CL upper.CL       t      p
   CCr-1  -  CCr-2  -13.3660 7.5085 20 -35.8343   9.1023 -1.7801 0.4115
   CCr-1  -  Man    -33.3600 7.5085 20 -55.8283 -10.8917 -4.4430 0.0021
   CCr-1  -  Min-A   -3.8760 7.5085 20 -26.3443  18.5923 -0.5162 0.9847
   CCr-1  -  Min-B   -9.1380 7.5085 20 -31.6063  13.3303 -1.2170 0.7419
   CCr-2  -  Man    -19.9940 7.5085 20 -42.4623   2.4743 -2.6628 0.0961
```

```
   CCr-2  -  Min-A    9.4900 7.5085 20 -12.9783  31.9583  1.2639 0.7152
   CCr-2  -  Min-B    4.2280 7.5085 20 -18.2403  26.6963  0.5631 0.9789
   Man  -  Min-A     29.4840 7.5085 20   7.0157  51.9523  3.9267 0.0066
   Man  -  Min-B     24.2220 7.5085 20   1.7537  46.6903  3.2259 0.0307
   Min-A  -  Min-B   -5.2620 7.5085 20 -27.7303  17.2063 -0.7008 0.9539
p-value and interval adjustment: tukey; confidence level: 0.9500
```

```
> st.lsm.cld(st.lsm(m, ~fert))
```

```
Least-squares means (Satterthwaite)
  fert     lsmeans      SE  DF  lower.CL  upper.CL  group
   CCr-1   18.2820  5.3093  20    3.2240   33.3400  a
   Min-A   22.1580  5.3093  20    7.1000   37.2160  a
   Min-B   27.4200  5.3093  20   12.3620   42.4780  a
   CCr-2   31.6480  5.3093  20   16.5900   46.7060  ab
   Man     51.6420  5.3093  20   36.5840   66.7000   b
Confidence level: 0.9500
Comparison adjustment: tukey (alpha=0.05)
Confidence intervals use sidak adjustment.
Shared letters indicate failure to reject a difference.
```

```
> st.lsm(
+     m,
+     ~fert,
+     method = "trt.vs.ctrl",
+     ref = 3,
+     side = "<",
+     adjust = "mvt"
+ )
```

```
Least-squares mean contrasts (Satterthwaite)
  contrast        estimate     SE DF lower.CL upper.CL       t      p
   CCr-1  -  Man  -33.3600 7.5085 20     -Inf -16.0573 -4.4430 0.0005
   CCr-2  -  Man  -19.9940 7.5085 20     -Inf  -2.6913 -2.6628 0.0244
   Min-A  -  Man  -29.4840 7.5085 20     -Inf -12.1813 -3.9267 0.0015
   Min-B  -  Man  -24.2220 7.5085 20     -Inf  -6.9193 -3.2259 0.0073
p-value and interval adjustment: mvt; confidence level: 0.9500
Alternative: estimate < null
```

```
> C1 <- list(
+     CCr1_CCr2 = c(1, -1, 0, 0, 0),
+     MinA_MinB = c(0, 0, 0, 1, -1),
+     Org_Min = c(1/3, 1/3, 1/3, -1/2, -1/2),
+     CCr_Man = c(1/2, 1/2, -1, 0, 0)
+ )
> st.lsm(
+     m,
+     ~fert,
+     method = C1
+ )
```

```
Least-squares mean contrasts (Satterthwaite)
  contrast   estimate      SE  DF  lower.CL  upper.CL       t       p
  CCr1_CCr2  -13.3660  7.5085  20  -29.0285    2.2965  -1.7801  0.0903
  MinA_MinB   -5.2620  7.5085  20  -20.9245   10.4005  -0.7008  0.4915
  Org_Min      9.0683  4.8467  20   -1.0418   19.1784   1.8710  0.0760
  CCr_Man    -26.6770  6.5026  20  -40.2411  -13.1129  -4.1025  0.0006
p-value adjustment: none; confidence level: 0.9500
```

```
> C2 <- list(
+     Man_CCr1 = c(-1, 0, 1, 0, 0),
+     Man_CCr2 = c(0, -1, 1, 0, 0),
+     Man_MinA = c(0, 0, 1, -1, 0),
+     Man_MinB = c(0, 0, 1, 0, -1)
+ )
> st.lsm(
+     m,
+     ~fert,
+     method = C2,
+     adjust = "hochberg"
+ )
```

```
Least-squares mean contrasts (Satterthwaite)
  contrast  estimate      SE  DF  lower.CL  upper.CL       t       p
  Man_CCr1   33.3600  7.5085  20   12.7538   53.9662  4.4430  0.0010
  Man_CCr2   19.9940  7.5085  20   -0.6122   40.6002  2.6628  0.0149
  Man_MinA   29.4840  7.5085  20    8.8778   50.0902  3.9267  0.0025
  Man_MinB   24.2220  7.5085  20    3.6158   44.8282  3.2259  0.0085
p-value adjustment: hochberg; confidence level: 0.9500
Confidence intervals use bonferroni adjustment.
```

```
> st.lsm(
+     m,
+     ~fert,
+     method = C2,
+     side = ">",
+     adjust = "hochberg"
+ )
```

```
Least-squares mean contrasts (Satterthwaite)
  contrast  estimate      SE  DF  lower.CL  upper.CL       t       p
  Man_CCr1   33.3600  7.5085  20   15.1660       Inf  4.4430  0.0005
  Man_CCr2   19.9940  7.5085  20    1.8000       Inf  2.6628  0.0075
  Man_MinA   29.4840  7.5085  20   11.2900       Inf  3.9267  0.0013
  Man_MinB   24.2220  7.5085  20    6.0280       Inf  3.2259  0.0042
p-value adjustment: hochberg; confidence level: 0.9500
Alternative: estimate > null
Confidence intervals use bonferroni adjustment.
```

```
> C3 <- list(
+     MinA_CCr1 = c(-1, 0, 0, 1, 0),
+     MinA_CCr2 = c(0, -1, 0, 1, 0),
+     MinA_Man = c(0, 0, -1, 1, 0),
+     MinB_CCr1 = c(-1, 0, 0, 0, 1),
+     MinB_CCr2 = c(0, -1, 0, 0, 1),
+     MinB_Man = c(0, 0, -1, 0, 1)
+ )
> st.lsm(
+     m,
+     ~fert,
+     method = C3,
+     adjust = "fdr"
+ )
```

```
Least-squares mean contrasts (Satterthwaite)
  contrast   estimate      SE  DF  lower.CL  upper.CL       t       p
```

```
  MinA_CCr1    3.8760  7.5085  20  -18.1023   25.8543   0.5162  0.6114
  MinA_CCr2   -9.4900  7.5085  20  -31.4683   12.4883  -1.2639  0.3566
  MinA_Man   -29.4840  7.5085  20  -51.4623   -7.5057  -3.9267  0.0050
  MinB_CCr1    9.1380  7.5085  20  -12.8403   31.1163   1.2170  0.3566
  MinB_CCr2   -4.2280  7.5085  20  -26.2063   17.7503  -0.5631  0.6114
  MinB_Man   -24.2220  7.5085  20  -46.2003   -2.2437  -3.2259  0.0127
p-value adjustment: fdr; confidence level: 0.9500
Confidence intervals use bonferroni adjustment.
```

```
> # Numeric contrast rows act on the five displayed fertilizer means.
> C.fert <- do.call(rbind, C1)
> h.fert <- st.lsm(
+     m,
+     ~fert,
+     method = C.fert,
+     adjust = "single-step"
+ )
> h.fert
```

```
Least-squares mean contrasts (Satterthwaite)
  contrast   estimate      SE  DF  lower.CL  upper.CL        t       p
  CCr1_CCr2  -13.3660  7.5085  20  -33.7999    7.0679  -1.7801  0.3022
  MinA_MinB   -5.2620  7.5085  20  -25.6959   15.1719  -0.7008  0.9258
  Org_Min      9.0683  4.8467  20   -4.1217   22.2583   1.8710  0.2601
  CCr_Man    -26.6770  6.5026  20  -44.3732   -8.9808  -4.1025  0.0022
p-value and interval adjustment: single-step; confidence level: 0.9500
Simultaneous inference uses one common reference DF per family (the minimum
  marginal DF unless DF is supplied); mixed-model results are approximate.
```

```
> # The four independent contrasts jointly test equality of all five means.
> st.lsm(
+     h.fert,
+     joint = TRUE
+ )
```

```
Fixed-effect linear hypothesis test
  DF1  DF2       F       p  F.scaling  method
    4   20  5.9979  0.0024     1.0000  Satterthwaite
```

3.3 Randomized complete blocks: Cotton

L02: RLab-AppStat/v14-u41.R

```
> data("v-lmm-cotton", package = "SelectionTools")
> m <- st.lmm(
+     Strength ~ Blk + Fert,
+     data = cotton
+ )
> m
```

```
Linear model fitted by REML

Model:
  Data:      cotton
  Fixed:     Strength ~ Blk + Fert
  Random:    None
```

```
  Residual:  ~units

Convergence:
  converged: projected score tolerance met

Fixed effects (exact F):
  Term  DF1  DF2       F       p
  Blk     2    8  1.1116  0.3750
  Fert    4    8  4.1916  0.0404

Variance components:
  Component  Variance
  Residual     0.0437

Fit and model statistics:
  Observations               15    R-squared (conditional)  0.5756
  Log-likelihood        -2.6352    RMSE (conditional)       0.1526
  AIC                   21.2703    Variance: fixed          0.0593
  BIC                   26.9347    Variance: random         0.0000
  R-squared (marginal)   0.5756    Variance: residual       0.0437
```

```
> st.lsm.cld(
+     st.lsm(m, ~Fert),
+     adjust = "none"
+ )
```

```
Least-squares means (Satterthwaite)
  Fert  lsmeans      SE  DF  lower.CL  upper.CL  group
  f5     7.4500  0.1207   8    7.1717    7.7283  a
  f4     7.5133  0.1207   8    7.2351    7.7916  ab
  f3     7.7433  0.1207   8    7.4651    8.0216  abc
  f1     7.8500  0.1207   8    7.5717    8.1283   bc
  f2     8.0533  0.1207   8    7.7751    8.3316    c
Confidence level: 0.9500
Comparison adjustment: none (alpha=0.05)
Shared letters indicate failure to reject a difference.
```

```
> st.lsm(
+     m,
+     ~Fert,
+     method = "pairwise",
+     adjust = "none"
+ )
```

```
Least-squares mean contrasts for Fert (Satterthwaite)
  contrast  estimate      SE  DF  lower.CL  upper.CL       t       p
  f1 - f2    -0.2033  0.1707   8   -0.5969    0.1902 -1.1915  0.2676
  f1 - f3     0.1067  0.1707   8   -0.2869    0.5002  0.6250  0.5494
  f1 - f4     0.3367  0.1707   8   -0.0569    0.7302  1.9728  0.0840
  f1 - f5     0.4000  0.1707   8    0.0065    0.7935  2.3439  0.0471
  f2 - f3     0.3100  0.1707   8   -0.0835    0.7035  1.8165  0.1068
  f2 - f4     0.5400  0.1707   8    0.1465    0.9335  3.1643  0.0133
  f2 - f5     0.6033  0.1707   8    0.2098    0.9969  3.5354  0.0077
  f3 - f4     0.2300  0.1707   8   -0.1635    0.6235  1.3477  0.2147
  f3 - f5     0.2933  0.1707   8   -0.1002    0.6869  1.7189  0.1240
  f4 - f5     0.0633  0.1707   8   -0.3302    0.4569  0.3711  0.7202
p-value adjustment: none; confidence level: 0.9500
```

```
> st.lsm.cld(st.lsm(m, ~Fert))
```

```
Least-squares means (Satterthwaite)
  Fert  lsmeans      SE  DF  lower.CL  upper.CL  group
  f5     7.4500  0.1207   8    7.0468    7.8532  a
  f4     7.5133  0.1207   8    7.1101    7.9166  ab
  f3     7.7433  0.1207   8    7.3401    8.1466  ab
  f1     7.8500  0.1207   8    7.4468    8.2532  ab
  f2     8.0533  0.1207   8    7.6501    8.4566   b
Confidence level: 0.9500
Comparison adjustment: tukey (alpha=0.05)
Confidence intervals use sidak adjustment.
Shared letters indicate failure to reject a difference.
```

```
> st.lsm.cld(
+     st.lsm(m, ~Fert),
+     alpha = 0.1
+ )
```

```
Least-squares means (Satterthwaite)
  Fert  lsmeans      SE  DF  lower.CL  upper.CL  group
  f5     7.4500  0.1207   8    7.0468    7.8532  a
  f4     7.5133  0.1207   8    7.1101    7.9166  a
  f3     7.7433  0.1207   8    7.3401    8.1466  ab
  f1     7.8500  0.1207   8    7.4468    8.2532  ab
  f2     8.0533  0.1207   8    7.6501    8.4566   b
Confidence level: 0.9500
Comparison adjustment: tukey (alpha=0.1)
Confidence intervals use sidak adjustment.
Shared letters indicate failure to reject a difference.
```

```
> st.lsm(
+     m,
+     ~Fert,
+     method = "pairwise",
+     adjust = "tukey"
+ )
```

```
Least-squares mean contrasts for Fert (Satterthwaite)
  contrast  estimate      SE  DF  lower.CL  upper.CL        t       p
  f1 - f2    -0.2033  0.1707   8   -0.7929    0.3862  -1.1915  0.7565
  f1 - f3     0.1067  0.1707   8   -0.4829    0.6962   0.6250  0.9667
  f1 - f4     0.3367  0.1707   8   -0.2529    0.9262   1.9728  0.3563
  f1 - f5     0.4000  0.1707   8   -0.1896    0.9896   2.3439  0.2246
  f2 - f3     0.3100  0.1707   8   -0.2796    0.8996   1.8165  0.4263
  f2 - f4     0.5400  0.1707   8   -0.0496    1.1296   3.1643  0.0743
  f2 - f5     0.6033  0.1707   8    0.0138    1.1929   3.5354  0.0448
  f3 - f4     0.2300  0.1707   8   -0.3596    0.8196   1.3477  0.6729
  f3 - f5     0.2933  0.1707   8   -0.2962    0.8829   1.7189  0.4740
  f4 - f5     0.0633  0.1707   8   -0.5262    0.6529   0.3711  0.9951
p-value and interval adjustment: tukey; confidence level: 0.9500
```

3.4 Latin square: Sugar beet

L03: RLab-AppStat/v14-u41.R

```
> data("v-lmm-sugar", package = "SelectionTools")
> m <- st.lmm(
```

```
+     Yield ~ Row + Col + Variety,
+     data = sugar
+ )
> m
```

```
Linear model fitted by REML

Model:
  Data:      sugar
  Fixed:     Yield ~ Row + Col + Variety
  Random:    None
  Residual:  ~units

Convergence:
  converged: projected score tolerance met

Fixed effects (exact F):
  Term     DF1  DF2        F        p
  Row        5   20   2.5453   0.0614
  Col        5   20   1.3942   0.2687
  Variety    5   20  14.0160  <0.0001

Variance components:
  Component  Variance
  Residual     0.1801

Fit and model statistics:
  Observations               36    R-squared (conditional)  0.7195
  Log-likelihood        -23.7793    RMSE (conditional)       0.3163
  AIC                    81.5585    Variance: fixed          0.4620
  BIC                   108.4784    Variance: random         0.0000
  R-squared (marginal)    0.7195    Variance: residual       0.1801
```

```
> st.lsm.cld(
+     st.lsm(m, ~Variety),
+     adjust = "none"
+ )
```

```
Least-squares means (Satterthwaite)
  Variety  lsmeans      SE  DF  lower.CL  upper.CL  group
  v4       16.2333  0.1733  20   15.8719   16.5947  a
  v3       16.4333  0.1733  20   16.0719   16.7947  a
  v1       16.4333  0.1733  20   16.0719   16.7947  a
  v6       16.5333  0.1733  20   16.1719   16.8947  ab
  v2       17.0167  0.1733  20   16.6553   17.3781   b
  v5       17.9833  0.1733  20   17.6219   18.3447    c
Confidence level: 0.9500
Comparison adjustment: none (alpha=0.05)
Shared letters indicate failure to reject a difference.
```

```
> st.lsm(
+     m,
+     ~Variety,
+     method = "pairwise",
+     adjust = "none"
+ )
```

```
Least-squares mean contrasts for Variety (Satterthwaite)
  contrast  estimate      SE  DF  lower.CL  upper.CL       t       p
```

```
  v1 - v2    -0.5833  0.2450  20   -1.0944   -0.0722  -2.3807   0.0273
  v1 - v3    <0.0001  0.2450  20   -0.5111    0.5111  <0.0001   1.0000
  v1 - v4     0.2000  0.2450  20   -0.3111    0.7111   0.8162   0.4240
  v1 - v5    -1.5500  0.2450  20   -2.0611   -1.0389  -6.3259  <0.0001
  v1 - v6    -0.1000  0.2450  20   -0.6111    0.4111  -0.4081   0.6875
  v2 - v3     0.5833  0.2450  20    0.0722    1.0944   2.3807   0.0273
  v2 - v4     0.7833  0.2450  20    0.2722    1.2944   3.1970   0.0045
  v2 - v5    -0.9667  0.2450  20   -1.4778   -0.4556  -3.9452   0.0008
  v2 - v6     0.4833  0.2450  20   -0.0278    0.9944   1.9726   0.0625
  v3 - v4     0.2000  0.2450  20   -0.3111    0.7111   0.8162   0.4240
  v3 - v5    -1.5500  0.2450  20   -2.0611   -1.0389  -6.3259  <0.0001
  v3 - v6    -0.1000  0.2450  20   -0.6111    0.4111  -0.4081   0.6875
  v4 - v5    -1.7500  0.2450  20   -2.2611   -1.2389  -7.1421  <0.0001
  v4 - v6    -0.3000  0.2450  20   -0.8111    0.2111  -1.2244   0.2350
  v5 - v6     1.4500  0.2450  20    0.9389    1.9611   5.9178  <0.0001
p-value adjustment: none; confidence level: 0.9500
```

3.5 Factorial randomized blocks: Cowpea

L04: RLab-AppStat/v14-u41.R

```
> data("v-lmm-peas", package = "SelectionTools")
> m <- st.lmm(
+     Yield ~ Block + Spacing * Variety,
+     data = peas
+ )
> m
```

```
Linear model fitted by REML

Model:
  Data:      peas
  Fixed:     Yield ~ Block + Spacing * Variety
  Random:    None
  Residual:  ~units

Convergence:
  converged: projected score tolerance met

Fixed effects (exact F):
  Term             DF1  DF2        F        p
  Block              3   24   4.8221   0.0091
  Spacing            2   24   4.3872   0.0238
  Variety            2   24  29.0694  <0.0001
  Spacing:Variety    4   24  10.8289  <0.0001

Variance components:
  Component  Variance
  Residual    17.6713

Fit and model statistics:
  Observations                36    R-squared (conditional)   0.7808
  Log-likelihood         -77.3588    RMSE (conditional)        3.4323
  AIC                    180.7177    Variance: fixed          62.9579
  BIC                    201.3034    Variance: random          0.0000
  R-squared (marginal)     0.7808    Variance: residual       17.6713
```

```
> st.lsm(m, ~Spacing:Variety)
```

```
Least-squares means (Satterthwaite)
  Spacing  Variety  lsmeans      SE  DF  lower.CL  upper.CL
  s04      v1       47.5000  2.1019  24   43.1620   51.8380
  s08      v1       50.7500  2.1019  24   46.4120   55.0880
  s12      v1       55.7500  2.1019  24   51.4120   60.0880
  s04      v2       62.2500  2.1019  24   57.9120   66.5880
  s08      v2       58.5000  2.1019  24   54.1620   62.8380
  s12      v2       52.2500  2.1019  24   47.9120   56.5880
  s04      v3       56.0000  2.1019  24   51.6620   60.3380
  s08      v3       64.2500  2.1019  24   59.9120   68.5880
  s12      v3       73.0000  2.1019  24   68.6620   77.3380
Confidence level: 0.9500
LSD=6.1349 HSD=10.1034 (alpha=0.05)
```

```
> st.lsm.cld(
+     st.lsm(m, ~Spacing | Variety),
+     adjust = "none",
+     sort = FALSE
+ )
```

```
Least-squares means (Satterthwaite)
  Spacing  Variety  lsmeans      SE  DF  lower.CL  upper.CL  group
  s04      v1       47.5000  2.1019  24   43.1620   51.8380  a
  s08      v1       50.7500  2.1019  24   46.4120   55.0880  ab
  s12      v1       55.7500  2.1019  24   51.4120   60.0880   b
  s04      v2       62.2500  2.1019  24   57.9120   66.5880   b
  s08      v2       58.5000  2.1019  24   54.1620   62.8380   b
  s12      v2       52.2500  2.1019  24   47.9120   56.5880  a
  s04      v3       56.0000  2.1019  24   51.6620   60.3380  a
  s08      v3       64.2500  2.1019  24   59.9120   68.5880   b
  s12      v3       73.0000  2.1019  24   68.6620   77.3380    c
Confidence level: 0.9500
Comparison adjustment: none (alpha=0.05)
Shared letters indicate failure to reject a difference.
```

```
> st.lsm(
+     m,
+     ~Spacing | Variety,
+     method = "pairwise",
+     adjust = "none"
+ )
```

```
Least-squares mean contrasts for Spacing (Satterthwaite)
  contrast  Variety estimate     SE DF lower.CL upper.CL       t       p
  s04 - s08 v1       -3.2500 2.9725 24  -9.3849   2.8849 -1.0934  0.2851
  s04 - s12 v1       -8.2500 2.9725 24 -14.3849  -2.1151 -2.7755  0.0105
  s08 - s12 v1       -5.0000 2.9725 24 -11.1349   1.1349 -1.6821  0.1055
  s04 - s08 v2        3.7500 2.9725 24  -2.3849   9.8849  1.2616  0.2192
  s04 - s12 v2       10.0000 2.9725 24   3.8651  16.1349  3.3642  0.0026
  s08 - s12 v2        6.2500 2.9725 24   0.1151  12.3849  2.1026  0.0462
  s04 - s08 v3       -8.2500 2.9725 24 -14.3849  -2.1151 -2.7755  0.0105
  s04 - s12 v3      -17.0000 2.9725 24 -23.1349 -10.8651 -5.7191 <0.0001
  s08 - s12 v3       -8.7500 2.9725 24 -14.8849  -2.6151 -2.9437  0.0071
p-value adjustment: none; confidence level: 0.9500
```

```
> st.lsm.cld(
+     st.lsm(m, ~Variety | Spacing),
```

```
+     adjust = "none",
+     sort = FALSE
+ )
```

```
Least-squares means (Satterthwaite)
  Variety  Spacing  lsmeans      SE  DF  lower.CL  upper.CL  group
  v1       s04      47.5000  2.1019  24   43.1620   51.8380  a
  v2       s04      62.2500  2.1019  24   57.9120   66.5880    c
  v3       s04      56.0000  2.1019  24   51.6620   60.3380   b
  v1       s08      50.7500  2.1019  24   46.4120   55.0880  a
  v2       s08      58.5000  2.1019  24   54.1620   62.8380   b
  v3       s08      64.2500  2.1019  24   59.9120   68.5880   b
  v1       s12      55.7500  2.1019  24   51.4120   60.0880  a
  v2       s12      52.2500  2.1019  24   47.9120   56.5880  a
  v3       s12      73.0000  2.1019  24   68.6620   77.3380   b
Confidence level: 0.9500
Comparison adjustment: none (alpha=0.05)
Shared letters indicate failure to reject a difference.
```

```
> st.lsm(
+     m,
+     ~Variety | Spacing,
+     method = "pairwise",
+     adjust = "none"
+ )
```

```
Least-squares mean contrasts for Variety (Satterthwaite)
  contrast Spacing estimate     SE DF lower.CL upper.CL       t       p
  v1 - v2  s04     -14.7500 2.9725 24 -20.8849  -8.6151 -4.9622 <0.0001
  v1 - v3  s04      -8.5000 2.9725 24 -14.6349  -2.3651 -2.8596  0.0086
  v2 - v3  s04       6.2500 2.9725 24   0.1151  12.3849  2.1026  0.0462
  v1 - v2  s08      -7.7500 2.9725 24 -13.8849  -1.6151 -2.6072  0.0154
  v1 - v3  s08     -13.5000 2.9725 24 -19.6349  -7.3651 -4.5417  0.0001
  v2 - v3  s08      -5.7500 2.9725 24 -11.8849   0.3849 -1.9344  0.0649
  v1 - v2  s12       3.5000 2.9725 24  -2.6349   9.6349  1.1775  0.2506
  v1 - v3  s12     -17.2500 2.9725 24 -23.3849 -11.1151 -5.8032 <0.0001
  v2 - v3  s12     -20.7500 2.9725 24 -26.8849 -14.6151 -6.9807 <0.0001
p-value adjustment: none; confidence level: 0.9500
```

3.6 Incomplete blocks: Soybean lattice

M04: RLab-AppStat/v14-u41.R

```
> data("v-lmm-soybean-blocks", package = "SelectionTools")
> soybean.blocks$Block.in.Rep <- soybean.blocks$Block:soybean.blocks$Rep
> m <- st.lmm(
+     fixed = Yield ~ Variety,
+     random = ~Rep + Block.in.Rep,
+     data = soybean.blocks
+ )
> m
```

```
Linear mixed model fitted by REML

Model:
  Data:      soybean.blocks
  Fixed:     Yield ~ Variety
  Random:    ~Rep + Block.in.Rep
```

```
  Residual:  ~units

Convergence:
  converged: projected score tolerance met

Fixed effects (Satterthwaite):
  Term     DF1    DF2       F       p
  Variety   24  17.14  1.9670  0.0765

Random effects (likelihood-ratio tests):
  Term          DF      LR       p
  Rep            1  0.2068  0.6493
  Block.in.Rep   1  6.7030  0.0096

Variance components:
  Component     Variance
  Rep             4.0150
  Block.in.Rep   19.6300
  Residual       13.6550

Fit and model statistics:
  Observations                50    R-squared (conditional)   0.7567
  Log-likelihood         -84.2847    RMSE (conditional)        2.2052
  AIC                    224.5695    Variance: fixed          18.8316
  BIC                    278.1061    Variance: random         23.6450
  R-squared (marginal)     0.3355    Variance: residual       13.6550
```

```
> st.lsm(m, ~Variety)
```

```
Least-squares means (Satterthwaite)
  Variety  lsmeans      SE    DF  lower.CL  upper.CL
  v01      19.0681  3.5366  7.17   10.7447   27.3914
  v02      16.9728  3.5366  7.17    8.6495   25.2962
  v03      14.6463  3.5366  7.17    6.3230   22.9696
  v04      14.7687  3.5366  7.17    6.4454   23.0921
  v05      12.8470  3.5366  7.17    4.5236   21.1703
  v06      13.1701  3.5366  7.17    4.8467   21.4934
  v07       9.0748  3.5366  7.17    0.7515   17.3981
  v08       6.7483  3.5366  7.17   -1.5751   15.0716
  v09       8.3707  3.5366  7.17    0.0474   16.6941
  v10       8.4489  3.5366  7.17    0.1256   16.7723
  [10 table rows omitted]
  v21      11.6259  3.5366  7.17    3.3025   19.9492
  v22      18.5306  3.5366  7.17   10.2073   26.8539
  v23      12.2041  3.5366  7.17    3.8807   20.5274
  v24      17.3265  3.5366  7.17    9.0032   25.6499
  v25      15.4048  3.5366  7.17    7.0814   23.7281
Confidence level: 0.9500
```

3.7 Split plots: Oats

M06: RLab-AppStat/v14-u51.R

```
> data("v-lmm-oats", package = "SelectionTools")
> m <- st.lmm(
+     fixed = Yield ~ Variety + Nitro + Variety:Nitro,
+     random = ~Block + Block:Variety,
+     data = oats
+ )
> m
```

```
Linear mixed model fitted by REML

Model:
  Data:      oats
  Fixed:     Yield ~ Variety + Nitro + Variety:Nitro
  Random:    ~Block + Block:Variety
  Residual:  ~units

Convergence:
  converged: projected score tolerance met

Fixed effects (exact F):
  Term           DF1  DF2        F        p
  Variety          2   10   1.4853   0.2724
  Nitro            3   45  37.6856  <0.0001
  Variety:Nitro    6   45   0.3028   0.9322

Random effects (exact F-tests):
  Term           DF1  DF2       F       p
  Block            5   10  5.2801  0.0124
  Block:Variety   10   45  3.3957  0.0023

Variance components:
  Component      Variance
  Block          214.4771
  Block:Variety  106.0618
  Residual       177.0833

Fit and model statistics:
  Observations                 72    R-squared (conditional)    0.7812
  Log-likelihood         -264.5143    RMSE (conditional)        10.8906
  AIC                     559.0285    Variance: fixed          311.6706
  BIC                     593.1785    Variance: random         320.5389
  R-squared (marginal)      0.3851    Variance: residual       177.0833
```

```
> st.lsm.cld(
+     st.lsm(m, ~Nitro),
+     adjust = "hochberg"
+ )
```

```
Least-squares means (Satterthwaite)
  Nitro   lsmeans      SE    DF  lower.CL  upper.CL  group
  n0.0    79.3889  7.1747  6.79   55.2226  103.5552  a
  n0.2    98.8889  7.1747  6.79   74.7226  123.0552   b
  n0.4   114.2222  7.1747  6.79   90.0559  138.3885    c
  n0.6   123.3889  7.1747  6.79   99.2226  147.5552     d
Confidence level: 0.9500
Comparison adjustment: hochberg (alpha=0.05)
Confidence intervals use bonferroni adjustment.
Shared letters indicate failure to reject a difference.
```

```
> st.lsm(
+     m,
+     ~Nitro,
+     method = "pairwise",
+     adjust = "hochberg"
+ )
```

```
Least-squares mean contrasts for Nitro (Satterthwaite)
  contrast     estimate      SE  DF  lower.CL  upper.CL       t        p
```

```
  n0.0 - n0.2  -19.5000  4.4358  45  -31.7424   -7.2576  -4.3961   0.0002
  n0.0 - n0.4  -34.8333  4.4358  45  -47.0757  -22.5910  -7.8529  <0.0001
  n0.0 - n0.6  -44.0000  4.4358  45  -56.2424  -31.7576  -9.9194  <0.0001
  n0.2 - n0.4  -15.3333  4.4358  45  -27.5757   -3.0910  -3.4568   0.0024
  n0.2 - n0.6  -24.5000  4.4358  45  -36.7424  -12.2576  -5.5233  <0.0001
  n0.4 - n0.6   -9.1667  4.4358  45  -21.4090    3.0757  -2.0665   0.0446
p-value adjustment: hochberg; confidence level: 0.9500
Confidence intervals use bonferroni adjustment.
```

```
> st.lsm.cld(
+     st.lsm(m, ~Variety),
+     adjust = "hochberg"
+ )
```

```
Least-squares means (Satterthwaite)
  Variety       lsmeans      SE    DF  lower.CL  upper.CL  group
  Victory       97.6250  7.7975  8.87   74.6779  120.5721  a
  Golden Rain  104.5000  7.7975  8.87   81.5529  127.4471  a
  Marvellous   109.7917  7.7975  8.87   86.8445  132.7388  a
Confidence level: 0.9500
Comparison adjustment: hochberg (alpha=0.05)
Confidence intervals use bonferroni adjustment.
Shared letters indicate failure to reject a difference.
```

```
> st.lsm(
+     m,
+     ~Variety,
+     method = "pairwise",
+     adjust = "hochberg"
+ )
```

```
Least-squares mean contrasts for Variety (Satterthwaite)
  contrast  estimate      SE  DF  lower.CL  upper.CL        t       p
  G - M      -5.2917  7.0789  10  -25.6086   15.0253  -0.7475  0.4720
  G - V       6.8750  7.0789  10  -13.4420   27.1920   0.9712  0.4720
  M - V      12.1667  7.0789  10   -8.1503   32.4836   1.7187  0.3492
  Variety: G = Golden Rain, M = Marvellous, V = Victory
p-value adjustment: hochberg; confidence level: 0.9500
Confidence intervals use bonferroni adjustment.
```

```
> st.lsm.cld(
+     st.lsm(m, ~Variety:Nitro),
+     adjust = "hochberg",
+     alpha = 0.1,
+     sort = FALSE
+ )
```

```
Least-squares means (Satterthwaite)
  Variety      Nitro   lsmeans      SE     DF  lower.CL  upper.CL  group
  Golden Rain  n0.0     80.0000  9.1070  16.08   49.6210  110.3790  ab
  Marvellous   n0.0     86.6667  9.1070  16.08   56.2877  117.0457  abcd
  Victory      n0.0     71.5000  9.1070  16.08   41.1210  101.8790  a
  Golden Rain  n0.2     98.5000  9.1070  16.08   68.1210  128.8790  abcdef
  Marvellous   n0.2    108.5000  9.1070  16.08   78.1210  138.8790   bcdefg
  Victory      n0.2     89.6667  9.1070  16.08   59.2877  120.0457  abc e
  Golden Rain  n0.4    114.6667  9.1070  16.08   84.2877  145.0457    cdefg
  Marvellous   n0.4    117.1667  9.1070  16.08   86.7877  147.5457      efg
```

```
  Victory      n0.4   110.8333  9.1070  16.08   80.4543  141.2123   bcdefg
  Golden Rain  n0.6   124.8333  9.1070  16.08   94.4543  155.2123        g
  Marvellous   n0.6   126.8333  9.1070  16.08   96.4543  157.2123       fg
  Victory      n0.6   118.5000  9.1070  16.08   88.1210  148.8790     d fg
Confidence level: 0.9500
Comparison adjustment: hochberg (alpha=0.1)
Confidence intervals use bonferroni adjustment.
Shared letters indicate failure to reject a difference.
```

```
> st.lsm(
+     m,
+     ~Variety:Nitro,
+     method = "pairwise",
+     adjust = "hochberg"
+ )
```

```
Least-squares mean contrasts for Variety, Nitro (Satterthwaite)
  contrast        estimate     SE    DF lower.CL upper.CL       t       p
  G n0.0 - M n0.0  -6.6667 9.7150 30.23 -43.0563  29.7230 -0.6862  0.8917
  G n0.0 - V n0.0   8.5000 9.7150 30.23 -27.8897  44.8897  0.8749  0.8917
  G n0.0 - G n0.2 -18.5000 7.6830    45 -46.2648   9.2648 -2.4079  0.7273
  G n0.0 - M n0.2 -28.5000 9.7150 30.23 -64.8897   7.8897 -2.9336  0.2852
  G n0.0 - V n0.2  -9.6667 9.7150 30.23 -46.0563  26.7230 -0.9950  0.8917
  G n0.0 - G n0.4 -34.6667 7.6830    45 -62.4314  -6.9019 -4.5122  0.0027
  G n0.0 - M n0.4 -37.1667 9.7150 30.23 -73.5563  -0.7770 -3.8257  0.0311
  G n0.0 - V n0.4 -30.8333 9.7150 30.23 -67.2230   5.5563 -3.1738  0.1585
  G n0.0 - G n0.6 -44.8333 7.6830    45 -72.5981 -17.0686 -5.8354 <0.0001
  G n0.0 - M n0.6 -46.8333 9.7150 30.23 -83.2230 -10.4437 -4.8207  0.0023
  [51 table rows omitted]
  V n0.4 - M n0.6 -16.0000 9.7150 30.23 -52.3897  20.3897 -1.6469  0.8917
  V n0.4 - V n0.6  -7.6667 7.6830    45 -35.4314  20.0981 -0.9979  0.8917
  G n0.6 - M n0.6  -2.0000 9.7150 30.23 -38.3897  34.3897 -0.2059  0.8917
  G n0.6 - V n0.6   6.3333 9.7150 30.23 -30.0563  42.7230  0.6519  0.8917
  M n0.6 - V n0.6   8.3333 9.7150 30.23 -28.0563  44.7230  0.8578  0.8917
  Variety: G = Golden Rain, M = Marvellous, V = Victory
p-value adjustment: hochberg; confidence level: 0.9500
Confidence intervals use bonferroni adjustment.
```

3.8 Nested incomplete blocks: Barley lattice

M07: RLab-AppStat/v14-u51.R

The original lattice analysis includes random effects for replicates and incomplete blocks nested within replicates. In `barley.lattice`, `RowBlk` identifies these incomplete blocks.

```
> data("v-lmm-barley-lattice", package = "SelectionTools")
> m.barley.lattice <- st.lmm(
+     fixed = yield ~ Variety,
+     random = ~Rep + RowBlk,
+     data = barley.lattice
+ )
> m.barley.lattice
```

```
Linear mixed model fitted by REML

Model:
  Data:      barley.lattice
  Fixed:     yield ~ Variety
  Random:    ~Rep + RowBlk
```

```
  Residual:  ~units

Convergence:
  converged: projected score tolerance met

Fixed effects (Satterthwaite):
  Term     DF1     DF2       F        p
  Variety   24  104.85  4.3087  <0.0001

Random effects (likelihood-ratio tests):
  Term    DF       LR        p
  Rep      1   2.6635   0.1027
  RowBlk   1  19.8950  <0.0001

Variance components:
  Component    Variance
  Rep         6882.1337
  RowBlk     14384.7680
  Residual   22677.3303

Fit and model statistics:
  Observations               150    R-squared (conditional)      0.6340
  Log-likelihood        -848.2596    RMSE (conditional)         124.7949
  AIC                  1752.5192    Variance: fixed          18022.3224
  BIC                  1836.8170    Variance: random         21266.9017
  R-squared (marginal)    0.2908    Variance: residual       22677.3303
```

3.9 Row and column blocks: Barley lattice

M08: RLab-AppStat/v14-u51.R

```
> data("v-lmm-barley-lattice", package = "SelectionTools")
> m <- st.lmm(
+     fixed = yield ~ Variety,
+     random = ~Rep + RowBlk + ColBlk,
+     data = barley.lattice
+ )
> m
```

```
Linear mixed model fitted by REML

Model:
  Data:      barley.lattice
  Fixed:     yield ~ Variety
  Random:    ~Rep + RowBlk + ColBlk
  Residual:  ~units

Convergence:
  converged: projected score tolerance met

Fixed effects (Satterthwaite):
  Term     DF1    DF2       F        p
  Variety   24  79.58  8.8441  <0.0001

Random effects (likelihood-ratio tests):
  Term    DF       LR        p
  Rep      1   0.6651   0.4148
  RowBlk   1  53.5160  <0.0001
  ColBlk   1  51.2132  <0.0001

Variance components:
  Component    Variance
  Rep         4262.3862
  RowBlk     15595.0614
```

```
  ColBlk     14811.5486
  Residual    8061.8060

Fit and model statistics:
  Observations               150    R-squared (conditional)      0.8637
  Log-likelihood        -822.6530    RMSE (conditional)          64.6426
  AIC                  1703.3059    Variance: fixed          16438.1595
  BIC                  1790.6144    Variance: random         34668.9962
  R-squared (marginal)    0.2778    Variance: residual        8061.8060
```

```
> st.lsm.cld(
+     st.lsm(m, ~Variety),
+     adjust = "hochberg"
+ )
```

```
Least-squares means (Satterthwaite)
  Variety    lsmeans       SE     DF   lower.CL   upper.CL  group
  va10     1193.2236  60.1994  19.46   978.5339  1407.9133  a
  va01     1283.5870  60.1994  19.46  1068.8973  1498.2767  ab
  va09     1298.8592  60.1994  19.46  1084.1695  1513.5489  abc
  va14     1326.6454  60.1994  19.46  1111.9557  1541.3351  abcd
  va11     1327.2452  60.1994  19.46  1112.5555  1541.9349  abcd
  va23     1329.1088  60.1994  19.46  1114.4192  1543.7985  abcd
  va16     1346.1480  60.1994  19.46  1131.4583  1560.8377  abcd
  va07     1400.7283  60.1994  19.46  1186.0386  1615.4180  abcde
  va03     1420.9307  60.1994  19.46  1206.2410  1635.6204  abcdef
  va04     1451.8554  60.1994  19.46  1237.1657  1666.5451   bcdefg
  [10 table rows omitted]
  va13     1619.0431  60.1994  19.46  1404.3534  1833.7328      efg
  va25     1630.6285  60.1994  19.46  1415.9389  1845.3182      efg
  va20     1639.9458  60.1994  19.46  1425.2561  1854.6355      efg
  va22     1644.3814  60.1994  19.46  1429.6917  1859.0711       fg
  va19     1669.5507  60.1994  19.46  1454.8610  1884.2403        g
Confidence level: 0.9500
Comparison adjustment: hochberg (alpha=0.05)
Confidence intervals use bonferroni adjustment.
Shared letters indicate failure to reject a difference.
```

```
> st.lsm(
+     m,
+     ~Variety,
+     method = "pairwise",
+     adjust = "hochberg"
+ )
```

```
Least-squares mean contrasts for Variety (Satterthwaite)
  contrast     estimate      SE    DF  lower.CL upper.CL       t      p
  va01 - va02 -265.4263 62.0193 79.58 -510.5387 -20.3140 -4.2797 0.0130
  va01 - va03 -137.3438 62.0193 79.58 -382.4561 107.7686 -2.2145 0.9980
  va01 - va04 -168.2684 62.0193 79.58 -413.3808  76.8439 -2.7132 0.9980
  va01 - va05 -249.6879 62.0193 79.58 -494.8002  -4.5755 -4.0260 0.0315
  va01 - va06 -243.8203 62.0193 79.58 -488.9327   1.2920 -3.9314 0.0432
  va01 - va07 -117.1413 62.0193 79.58 -362.2537 127.9710 -1.8888 0.9980
  va01 - va08 -173.7866 62.0193 79.58 -418.8990  71.3258 -2.8021 0.9980
  va01 - va09  -15.2723 62.0193 79.58 -260.3846 229.8401 -0.2462 0.9980
  va01 - va10   90.3633 62.0193 79.58 -154.7490 335.4757  1.4570 0.9980
  va01 - va11  -43.6582 62.0193 79.58 -288.7706 201.4541 -0.7039 0.9980
  [285 table rows omitted]
  va22 - va24   97.9115 62.0193 79.58 -147.2008 343.0239  1.5787 0.9980
  va22 - va25   13.7529 62.0193 79.58 -231.3595 258.8652  0.2218 0.9980
  va23 - va24 -217.3611 62.0193 79.58 -462.4734  27.7513 -3.5047 0.1696
```

```
  va23 - va25 -301.5197 62.0193 79.58 -546.6321 -56.4073 -4.8617 0.0015
  va24 - va25  -84.1586 62.0193 79.58 -329.2710 160.9537 -1.3570 0.9980
p-value adjustment: hochberg; confidence level: 0.9500
Confidence intervals use bonferroni adjustment.
```

```
> st.lsm(
+     m,
+     ~Variety,
+     method = "trt.vs.ctrl",
+     ref = 4,
+     side = ">",
+     adjust = "none"
+ )
```

```
Least-squares mean contrasts (Satterthwaite)
  contrast     estimate      SE    DF  lower.CL upper.CL       t      p
  va01 - va04 -168.2684 62.0193 79.58 -271.4828      Inf -2.7132 0.9959
  va02 - va04   97.1579 62.0193 79.58   -6.0565      Inf  1.5666 0.0606
  va03 - va04  -30.9247 62.0193 79.58 -134.1390      Inf -0.4986 0.6903
  va05 - va04   81.4194 62.0193 79.58  -21.7949      Inf  1.3128 0.0965
  va06 - va04   75.5519 62.0193 79.58  -27.6625      Inf  1.2182 0.1134
  va07 - va04  -51.1271 62.0193 79.58 -154.3415      Inf -0.8244 0.7939
  va08 - va04    5.5181 62.0193 79.58  -97.6962      Inf  0.0890 0.4647
  va09 - va04 -152.9962 62.0193 79.58 -256.2105      Inf -2.4669 0.9921
  va10 - va04 -258.6318 62.0193 79.58 -361.8461      Inf -4.1702 1.0000
  va11 - va04 -124.6102 62.0193 79.58 -227.8246      Inf -2.0092 0.9760
  [9 table rows omitted]
  va21 - va04   41.5819 62.0193 79.58  -61.6324      Inf  0.6705 0.2523
  va22 - va04  192.5260 62.0193 79.58   89.3117      Inf  3.1043 0.0013
  va23 - va04 -122.7466 62.0193 79.58 -225.9609      Inf -1.9792 0.9744
  va24 - va04   94.6145 62.0193 79.58   -8.5999      Inf  1.5256 0.0655
  va25 - va04  178.7731 62.0193 79.58   75.5588      Inf  2.8825 0.0025
p-value adjustment: none; confidence level: 0.9500
Alternative: estimate > null
```

3.10 Separable autoregressive plot residuals

The field spans ten rows and fifteen columns. The spatial model uses their field-wide positions, including adjacency across replicate boundaries. The colon in the covariance formula specifies the product of the row and column AR(1) correlation matrices.

```
> data("v-lmm-barley-lattice", package = "SelectionTools")
> m.barley.ar1xar1 <- st.lmm(
+     fixed = yield ~ Variety,
+     cov = ~ar1(Row):ar1(Column),
+     data = barley.lattice
+ )
> m.barley.ar1xar1
```

```
Linear model fitted by REML

Model:
  Data:      barley.lattice
  Fixed:     yield ~ Variety
  Random:    None
  Residual:  ~ar1(Row):ar1(Column)

Convergence:
  converged: projected score tolerance met
```

```
Fixed effects (Satterthwaite):
  Term     DF1    DF2        F        p
  Variety   24  77.75  13.0434  <0.0001

Covariance components:
  Residual: ar1(Row):ar1(Column)
    Separable AR1 x AR1
  Parameter      Estimate
  Variance     38757.0944
  rho(Row)         0.4586
  rho(Column)      0.6838

Fit and model statistics:
  Observations                150    R-squared (conditional)      0.3112
  Log-likelihood        -815.1898    RMSE (conditional)         200.1672
  AIC                  1686.3795    Variance: fixed          17509.7379
  BIC                  1770.6773    Variance: random             0.0000
  R-squared (marginal)    0.3112    Variance: residual       38757.0944
```

For plots at positions (r,c) and (r',c'), the residual covariance is

$$\mathrm{Cov}(e_{rc}, e_{r'c'}) = \sigma_e^2 \rho_r^{|r-r'|} \rho_c^{|c-c'|}.$$

Ordering the ten rows within each of the fifteen columns gives

$$\mathbf{R} = \sigma_e^2 \left\{ \mathbf{A}_{15}(\rho_c) \otimes \mathbf{A}_{10}(\rho_r) \right\}, \qquad [\mathbf{A}_m(\rho)]_{ij} = \rho^{|i-j|}.$$

There is one residual variance and one correlation parameter for each direction. Correlation decreases with separation in both directions; plots in different rows and different columns can therefore have nonzero residual covariance.

Both models use the same 150 plots and fixed variety effects. The REML AIC favors the spatial model (1686.38 versus 1703.31 for the block model in Section 3.9), as does BIC (1770.68 versus 1790.61).

3.11 Incomplete blocks: Rapeseed

M09: RLab-AppStat/v14-u51.R

```
> data("v-lmm-rapeseed-blocks", package = "SelectionTools")
> m <- st.lmm(
+     fixed = hgt ~ geno + env + geno:env,
+     random = ~rep:env + blk:rep:env,
+     data = rapeseed.blocks
+ )
> m
```

```
Linear mixed model fitted by REML

Model:
  Data:      rapeseed.blocks
  Fixed:     hgt ~ geno + env + geno:env
  Random:    ~rep:env + blk:rep:env
  Residual:  ~units

Convergence:
  converged: projected score tolerance met

Fixed effects (Satterthwaite):
  Term      DF1    DF2        F        p
  geno       19  53.35   6.4449  <0.0001
  env         1      4  51.5051   0.0020
  geno:env   19  53.35   0.7785   0.7204
```

```
Random effects (likelihood-ratio tests):
  Term         DF      LR       p
  rep:env       1  0.2977  0.5854
  rep:env:blk   1  2.9732  0.0847

Variance components:
  Component    Variance
  rep:env        2.9625
  rep:env:blk   32.5104
  Residual      59.5880

Fit and model statistics:
  Observations               120    R-squared (conditional)    0.7798
  Log-likelihood        -315.4139    RMSE (conditional)         5.2737
  AIC                    716.8277    Variance: fixed          175.5267
  BIC                    836.6899    Variance: random          35.4729
  R-squared (marginal)     0.6487    Variance: residual        59.5880
```

```
> st.lsm(m, ~geno)
```

```
Least-squares means (Satterthwaite)
  geno   lsmeans      SE     DF  lower.CL  upper.CL
  l01   217.3993  3.8136  68.37  209.7901  225.0084
  l02   230.9399  3.8136  68.37  223.3308  238.5491
  l03   213.0651  3.8136  68.37  205.4559  220.6742
  l04   223.5202  3.8136  68.37  215.9110  231.1293
  l05   209.9625  3.8136  68.37  202.3534  217.5717
  l06   228.4188  3.8136  68.37  220.8096  236.0280
  l07   209.6484  3.8136  68.37  202.0393  217.2576
  l08   228.4379  3.8136  68.37  220.8288  236.0471
  l09   220.8363  3.8136  68.37  213.2271  228.4454
  l10   211.6777  3.8136  68.37  204.0686  219.2869
  [5 table rows omitted]
  l16   213.3490  3.8136  68.37  205.7398  220.9581
  l17   195.8451  3.8136  68.37  188.2359  203.4542
  l18   226.2422  3.8136  68.37  218.6331  233.8514
  l19   215.1494  3.8136  68.37  207.5403  222.7586
  l20   198.0943  3.8136  68.37  190.4851  205.7034
Confidence level: 0.9500
```

```
> st.lsm(m, ~geno | env)
```

```
Least-squares means (Satterthwaite)
  geno  env   lsmeans      SE     DF  lower.CL  upper.CL
  l01   k    223.6745  5.3932  68.37  212.9135  234.4355
  l02   k    244.8627  5.3932  68.37  234.1017  255.6236
  l03   k    220.8769  5.3932  68.37  210.1160  231.6379
  l04   k    232.1803  5.3932  68.37  221.4193  242.9413
  l05   k    217.3885  5.3932  68.37  206.6275  228.1495
  l06   k    236.8973  5.3932  68.37  226.1363  247.6583
  l07   k    215.0957  5.3932  68.37  204.3347  225.8566
  l08   k    234.8665  5.3932  68.37  224.1055  245.6274
  l09   k    226.7914  5.3932  68.37  216.0305  237.5524
  l10   k    221.1931  5.3932  68.37  210.4321  231.9541
  [25 table rows omitted]
  l16   s    206.6806  5.3932  68.37  195.9196  217.4416
  l17   s    184.4065  5.3932  68.37  173.6455  195.1674
  l18   s    221.0634  5.3932  68.37  210.3025  231.8244
  l19   s    204.6029  5.3932  68.37  193.8419  215.3639
  l20   s    180.8415  5.3932  68.37  170.0805  191.6025
Confidence level: 0.9500
```

```
> st.lsm(
+     m,
+     ~geno:env,
+     at = list(env = "s")
+ )
```

```
Least-squares means (Satterthwaite)
At env = s
  geno  env   lsmeans      SE     DF  lower.CL  upper.CL
  l01   s     211.1240  5.3932  68.37  200.3631  221.8850
  l02   s     217.0171  5.3932  68.37  206.2562  227.7781
  l03   s     205.2532  5.3932  68.37  194.4923  216.0142
  l04   s     214.8600  5.3932  68.37  204.0991  225.6210
  l05   s     202.5365  5.3932  68.37  191.7755  213.2975
  l06   s     219.9403  5.3932  68.37  209.1793  230.7013
  l07   s     204.2011  5.3932  68.37  193.4402  214.9621
  l08   s     222.0094  5.3932  68.37  211.2485  232.7704
  l09   s     214.8811  5.3932  68.37  204.1201  225.6421
  l10   s     202.1623  5.3932  68.37  191.4013  212.9233
  [5 table rows omitted]
  l16   s     206.6806  5.3932  68.37  195.9196  217.4416
  l17   s     184.4065  5.3932  68.37  173.6455  195.1674
  l18   s     221.0634  5.3932  68.37  210.3025  231.8244
  l19   s     204.6029  5.3932  68.37  193.8419  215.3639
  l20   s     180.8415  5.3932  68.37  170.0805  191.6025
Confidence level: 0.9500
```

```
> st.lsm(
+     m,
+     ~geno:env,
+     at = list(env = "s"),
+     method = "pairwise",
+     adjust = "none",
+     level = 0.9
+ )
```

```
Least-squares mean contrasts for geno, env (Satterthwaite)
At env = s
  contrast       estimate     SE    DF lower.CL upper.CL       t       p
  l01 s - l02 s  -5.8931 7.4965 71.82 -18.3848   6.5986 -0.7861  0.4344
  l01 s - l03 s   5.8708 7.4957 71.72  -6.6198  18.3614  0.7832  0.4361
  l01 s - l04 s  -3.7360 7.0205 49.81 -15.5026   8.0306 -0.5322  0.5970
  l01 s - l05 s   8.5875 7.4933 71.44  -3.8998  21.0748  1.1460  0.2556
  l01 s - l06 s  -8.8163 7.4933 71.44 -21.3036   3.6710 -1.1766  0.2433
  l01 s - l07 s   6.9229 7.4961 71.77  -5.5683  19.4141  0.9235  0.3588
  l01 s - l08 s -10.8854 7.3841 63.24 -23.2118   1.4410 -1.4742  0.1454
  l01 s - l09 s  -3.7570 7.4961 71.77 -16.2483   8.7342 -0.5012  0.6178
  l01 s - l10 s   8.9617 7.4763 69.89  -3.5009  21.4243  1.1987  0.2347
  l01 s - l11 s  -4.0193 7.4389 67.73 -16.4249   8.3863 -0.5403  0.5908
  [175 table rows omitted]
  l17 s - l19 s -20.1964 7.4961 71.77 -32.6876  -7.7052 -2.6943  0.0088
  l17 s - l20 s   3.5650 7.0137 48.88  -8.1945  15.3244  0.5083  0.6135
  l18 s - l19 s  16.4605 7.0205 49.81   4.6939  28.2271  2.3446  0.0231
  l18 s - l20 s  40.2219 7.4893 71.15  27.7406  52.7033  5.3706 <0.0001
  l19 s - l20 s  23.7614 7.4957 71.72  11.2708  36.2520  3.1700  0.0022
p-value adjustment: none; confidence level: 0.9000
```

```
> st.lsm.cld(
+     st.lsm(
```

```
+         m,
+         ~geno:env,
+         at = list(env = "s")
+     ),
+     adjust = "none",
+     alpha = 0.1
+ )
```

```
Least-squares means (Satterthwaite)
At env = s
  geno  env   lsmeans      SE     DF  lower.CL  upper.CL  group
  l20   s    180.8415  5.3932  68.37  170.0805  191.6025  a
  l17   s    184.4065  5.3932  68.37  173.6455  195.1674  a
  l10   s    202.1623  5.3932  68.37  191.4013  212.9233   b
  l15   s    202.1983  5.3932  68.37  191.4373  212.9593   b
  l05   s    202.5365  5.3932  68.37  191.7755  213.2975   bc
  l07   s    204.2011  5.3932  68.37  193.4402  214.9621   bcd
  l19   s    204.6029  5.3932  68.37  193.8419  215.3639   bcd
  l03   s    205.2532  5.3932  68.37  194.4923  216.0142   bcde
  l13   s    205.9807  5.3932  68.37  195.2197  216.7416   bcde
  l16   s    206.6806  5.3932  68.37  195.9196  217.4416   bcde
  [5 table rows omitted]
  l11   s    215.1433  5.3932  68.37  204.3824  225.9043     def
  l02   s    217.0171  5.3932  68.37  206.2562  227.7781      ef
  l06   s    219.9403  5.3932  68.37  209.1793  230.7013       f
  l18   s    221.0634  5.3932  68.37  210.3025  231.8244       f
  l08   s    222.0094  5.3932  68.37  211.2485  232.7704       f
Confidence level: 0.9500
Comparison adjustment: none (alpha=0.1)
Shared letters indicate failure to reject a difference.
```

```
> st.lsm.cld(
+     st.lsm(
+         m,
+         ~geno:env,
+         at = list(env = "s")
+     ),
+     adjust = "FDR",
+     alpha = 0.1
+ )
```

```
Least-squares means (Satterthwaite)
At env = s
  geno  env   lsmeans      SE     DF  lower.CL  upper.CL  group
  l20   s    180.8415  5.3932  68.37  163.9102  197.7728  a
  l17   s    184.4065  5.3932  68.37  167.4751  201.3378  a
  l10   s    202.1623  5.3932  68.37  185.2310  219.0936   b
  l15   s    202.1983  5.3932  68.37  185.2670  219.1296   b
  l05   s    202.5365  5.3932  68.37  185.6052  219.4678   b
  l07   s    204.2011  5.3932  68.37  187.2698  221.1324   bc
  l19   s    204.6029  5.3932  68.37  187.6716  221.5342   bc
  l03   s    205.2532  5.3932  68.37  188.3219  222.1845   bcd
  l13   s    205.9807  5.3932  68.37  189.0494  222.9120   bcd
  l16   s    206.6806  5.3932  68.37  189.7493  223.6119   bcd
  [5 table rows omitted]
  l11   s    215.1433  5.3932  68.37  198.2120  232.0747   bcd
  l02   s    217.0171  5.3932  68.37  200.0858  233.9484   bcd
  l06   s    219.9403  5.3932  68.37  203.0090  236.8716    cd
  l18   s    221.0634  5.3932  68.37  204.1321  237.9947     d
  l08   s    222.0094  5.3932  68.37  205.0781  238.9407     d
Confidence level: 0.9500
Comparison adjustment: FDR (alpha=0.1)
```

```
Confidence intervals use bonferroni adjustment.
Shared letters indicate failure to reject a difference.
```

3.12 Multi-environment trials: Barley

L05-TKW: barley-met

```
> data("v-lmm-barley-met", package = "SelectionTools")
> m <- st.lmm(
+     TKW ~ gen * loc * year + year:loc:rep,
+     data = barley.met
+ )
> m
```

```
Linear model fitted by REML

Model:
  Data:      barley.met
  Fixed:     TKW ~ gen * loc * year + year:loc:rep
  Random:    None
  Residual:  ~units

Convergence:
  converged: projected score tolerance met

Fixed effects (exact F):
  Term          DF1  DF2         F        p
  gen            14  630  126.1194  <0.0001
  loc             4  630  669.8488  <0.0001
  year            2  630  131.2096  <0.0001
  gen:loc        56  630    9.4225  <0.0001
  gen:year       28  630   34.2745  <0.0001
  loc:year        8  630   53.0936  <0.0001
  gen:loc:year  112  630    2.7626  <0.0001
  loc:year:rep   45  630    2.0730  <0.0001

Variance components:
  Component  Variance
  Residual     8.2587

Fit and model statistics:
  Observations                 900    R-squared (conditional)   0.8865
  Log-likelihood        -1765.4710    RMSE (conditional)        2.4044
  AIC                    4072.9421    Variance: fixed          64.5102
  BIC                    5374.3911    Variance: random          0.0000
  R-squared (marginal)      0.8865    Variance: residual        8.2587
```

```
> st.lsm(m, "gen")
```

```
Least-squares means (Satterthwaite)
  gen    lsmeans      SE   DF  lower.CL  upper.CL
  lne01  34.3150  0.3710  630   33.5864   35.0436
  lne02  44.9733  0.3710  630   44.2448   45.7019
  lne03  47.3033  0.3710  630   46.5748   48.0319
  lne04  40.3278  0.3710  630   39.5992   41.0563
  lne05  44.2333  0.3710  630   43.5048   44.9619
  lne06  42.0650  0.3710  630   41.3364   42.7936
  lne07  41.9983  0.3710  630   41.2698   42.7269
  lne08  41.9000  0.3710  630   41.1714   42.6286
  lne09  44.4067  0.3710  630   43.6781   45.1352
```

```
  lne10  36.5867  0.3710  630   35.8581   37.3152
  lne11  44.4083  0.3710  630   43.6798   45.1369
  lne12  45.3417  0.3710  630   44.6131   46.0702
  lne13  35.6533  0.3710  630   34.9248   36.3819
  lne14  45.7200  0.3710  630   44.9914   46.4486
  lne15  48.0467  0.3710  630   47.3181   48.7752
Confidence level: 0.9500
LSD=1.0303 HSD=1.7865 (alpha=0.05)
```

```
> st.lsm.cld(
+     st.lsm(m, "gen"),
+     reversed = TRUE,
+     alpha = 0.1,
+     adjust = "none"
+ )
```

```
Least-squares means (Satterthwaite)
  gen    lsmeans      SE   DF  lower.CL  upper.CL  group
  lne15  48.0467  0.3710  630   47.3181   48.7752  a
  lne03  47.3033  0.3710  630   46.5748   48.0319  a
  lne14  45.7200  0.3710  630   44.9914   46.4486   b
  lne12  45.3417  0.3710  630   44.6131   46.0702   b
  lne02  44.9733  0.3710  630   44.2448   45.7019   bc
  lne11  44.4083  0.3710  630   43.6798   45.1369    c
  lne09  44.4067  0.3710  630   43.6781   45.1352    c
  lne05  44.2333  0.3710  630   43.5048   44.9619    c
  lne06  42.0650  0.3710  630   41.3364   42.7936     d
  lne07  41.9983  0.3710  630   41.2698   42.7269     d
  lne08  41.9000  0.3710  630   41.1714   42.6286     d
  lne04  40.3278  0.3710  630   39.5992   41.0563      e
  lne10  36.5867  0.3710  630   35.8581   37.3152       f
  lne13  35.6533  0.3710  630   34.9248   36.3819        g
  lne01  34.3150  0.3710  630   33.5864   35.0436         h
Confidence level: 0.9500
Comparison adjustment: none (alpha=0.1)
Shared letters indicate failure to reject a difference.
```

3.12.1 One environment and one trait

M21/y2016:HOH/TKW: barley-met

```
> data("v-lmm-barley-met", package = "SelectionTools")
> barley.met$env <- barley.met$year:barley.met$loc
> barley.met <- subset(barley.met, env == "y2016:HOH")
> st.lmm(
+     fixed = TKW ~ 1,
+     random = ~gen + rep,
+     data = barley.met
+ )
```

```
Linear mixed model fitted by REML

Model:
  Data:      barley.met
  Fixed:     TKW ~ 1
  Random:    ~gen + rep
  Residual:  ~units

Convergence:
  converged: projected score tolerance met
```

```
Random effects (exact F-tests):
  Term  DF1  DF2        F        p
  gen    14   42  25.5607  <0.0001
  rep     3   42   2.6254   0.0629

Variance components:
  Component  Variance
  gen         38.0529
  rep          0.6716
  Residual     6.1974

Fit and model statistics:
  Observations                60    R-squared (conditional)   0.8620
  Log-likelihood        -163.7115    RMSE (conditional)        2.1243
  AIC                    335.4231    Variance: fixed           0.0000
  BIC                    343.8004    Variance: random         38.7245
  R-squared (marginal)     0.0000    Variance: residual        6.1974
```

4 Heterogeneous Residual Variances

4.1 Maize genotype comparisons across and within counties

M14, M15, M16: RLab-AppStat/v14-u51.R

```
> data("v-lmm-maize-counties", package = "SelectionTools")
> m <- st.lmm(
+     fixed = yield ~ gen + county + gen:county,
+     random = ~county:rep + county:rep:block,
+     cov = ~at(county):units,
+     data = maize.counties
+ )
> m
```

```
Linear mixed model fitted by REML

Model:
  Data:      maize.counties
  Fixed:     yield ~ gen + county + gen:county
  Random:    ~county:rep + county:rep:block
  Residual:  ~at(county):units

Convergence:
  converged: projected score tolerance met

Fixed effects (Satterthwaite):
  Term         DF1     DF2        F        p
  gen           63  597.85   4.2206  <0.0001
  county         5   11.85  30.5549  <0.0001
  gen:county   315  223.31   1.2205   0.0557

Random effects (likelihood-ratio tests):
  Term              DF       LR        p
  county:rep         1  59.1262  <0.0001
  county:rep:block   1  60.4058  <0.0001

Variance components:
  Component                      Variance
  county:rep                     119.1287
  county:rep:block                58.5633
  Residual: at(county):units[C1]  149.1453
  Residual: at(county):units[C2]  200.0959
  Residual: at(county):units[C3]  146.5867
```

```
  Residual: at(county):units[C4]  238.5624
  Residual: at(county):units[C5]  122.2418
  Residual: at(county):units[C6]  331.9068

Fit and model statistics:
  Observations                1152    R-squared (conditional)     0.8760
  Log-likelihood        -3394.3420    RMSE (conditional)         10.8245
  AIC                    7572.6840    Variance: fixed          1221.5982
  BIC                    9551.9919    Variance: random          177.6920
  R-squared (marginal)      0.7648    Variance: residual        198.0898
```

```
> st.lsm.cld(
+     st.lsm(m, ~gen),
+     adjust = "none",
+     alpha = 0.1
+ )
```

```
Least-squares means (Satterthwaite)
  gen  lsmeans     SE    DF lower.CL upper.CL group
  G57  88.3565 4.3708 79.51  79.6576  97.0553 a
  G29  90.7582 4.3708 79.51  82.0593  99.4570 ab
  G34  92.9218 4.3708 79.51  84.2229 101.6206 abc
  G59  93.1192 4.3708 79.51  84.4203 101.8181 abcd
  G10  94.7329 4.3708 79.51  86.0340 103.4318 abcde
  G33  95.4036 4.3708 79.51  86.7047 104.1025 abcdef
  G31  96.4081 4.3708 79.51  87.7092 105.1070 abcdef
  G60  97.6059 4.3708 79.51  88.9070 106.3048  bcdefg
  G37  98.7320 4.3708 79.51  90.0331 107.4308  bcdefgh
  G23  99.3291 4.3708 79.51  90.6302 108.0280   cdefghi
  [49 table rows omitted]
  G51 116.0963 4.3708 79.51 107.3974 124.7952                    wxyzA
  G38 116.6934 4.3708 79.51 107.9945 125.3923                     yzA
  G61 118.4486 4.3708 79.51 109.7497 127.1475                      zA
  G54 119.5977 4.3708 79.51 110.8988 128.2966                       A
  G05 119.6098 4.3708 79.51 110.9109 128.3087                       A
Confidence level: 0.9500
Comparison adjustment: none (alpha=0.1)
Shared letters indicate failure to reject a difference.
```

```
> st.lsm(m, ~gen)
```

```
Least-squares means (Satterthwaite)
  gen   lsmeans      SE     DF  lower.CL  upper.CL
  G01  110.6986  4.3708  79.51  101.9997  119.3975
  G02  107.9999  4.3708  79.51   99.3010  116.6988
  G03   99.4994  4.3708  79.51   90.8005  108.1983
  G04  115.4992  4.3708  79.51  106.8003  124.1981
  G05  119.6098  4.3708  79.51  110.9109  128.3087
  G06  106.8559  4.3708  79.51   98.1570  115.5548
  G07  107.0087  4.3708  79.51   98.3098  115.7076
  G08  113.7642  4.3708  79.51  105.0653  122.4631
  G09  109.5279  4.3708  79.51  100.8290  118.2268
  G10   94.7329  4.3708  79.51   86.0340  103.4318
  [49 table rows omitted]
  G60   97.6059  4.3708  79.51   88.9070  106.3048
  G61  118.4486  4.3708  79.51  109.7497  127.1475
  G62  110.0632  4.3708  79.51  101.3643  118.7620
  G63  109.4353  4.3708  79.51  100.7364  118.1342
  G64  109.2622  4.3708  79.51  100.5633  117.9611
Confidence level: 0.9500
```

```
> st.lsm(
+     m,
+     ~gen,
+     method = "pairwise",
+     adjust = "none",
+     level = 0.9
+ )
```

```
Least-squares mean contrasts for gen (Satterthwaite)
  contrast   estimate      SE      DF  lower.CL  upper.CL        t       p
  G01 - G02    2.6987  4.9461  630.82   -5.4488   10.8463   0.5456  0.5855
  G01 - G03   11.1992  4.9461  630.82    3.0516   19.3467   2.2643  0.0239
  G01 - G04   -4.8006  4.9461  630.82  -12.9482    3.3469  -0.9706  0.3321
  G01 - G05   -8.9112  4.9461  630.82  -17.0588   -0.7637  -1.8017  0.0721
  G01 - G06    3.8427  4.9461  630.82   -4.3049   11.9902   0.7769  0.4375
  G01 - G07    3.6899  4.9461  630.82   -4.4577   11.8374   0.7460  0.4559
  G01 - G08   -3.0656  4.9461  630.82  -11.2132    5.0819  -0.6198  0.5356
  G01 - G09    1.1707  4.8627  609.83   -6.8399    9.1813   0.2407  0.8098
  G01 - G10   15.9657  4.9461  630.82    7.8181   24.1132   3.2279  0.0013
  G01 - G11    7.2374  4.8627  609.83   -0.7731   15.2480   1.4884  0.1372
  [2001 table rows omitted]
  G61 - G63    9.0133  4.9461  630.82    0.8657   17.1608   1.8223  0.0689
  G61 - G64    9.1864  4.9461  630.82    1.0388   17.3339   1.8573  0.0637
  G62 - G63    0.6278  4.9461  630.82   -7.5197    8.7754   0.1269  0.8990
  G62 - G64    0.8010  4.9461  630.82   -7.3466    8.9485   0.1619  0.8714
  G63 - G64    0.1731  4.9461  630.82   -7.9744    8.3207   0.0350  0.9721
p-value adjustment: none; confidence level: 0.9000
```

```
> st.lsm(
+     m,
+     ~gen,
+     at = list(county = "C5")
+ )
```

```
Least-squares means (Satterthwaite)
At county = C5
  gen   lsmeans      SE     DF  lower.CL  upper.CL
  G01  107.5500  9.3649  43.92   88.6753  126.4248
  G02  118.7182  9.3649  43.92   99.8435  137.5930
  G03   98.3491  9.3649  43.92   79.4743  117.2238
  G04  110.8950  9.3649  43.92   92.0203  129.7698
  G05  138.7260  9.3649  43.92  119.8513  157.6008
  G06  110.1536  9.3649  43.92   91.2788  129.0283
  G07  112.8860  9.3649  43.92   94.0113  131.7608
  G08  122.8093  9.3649  43.92  103.9346  141.6841
  G09  133.1815  9.3649  43.92  114.3068  152.0563
  G10   86.2014  9.3649  43.92   67.3267  105.0762
  [49 table rows omitted]
  G60  117.3856  9.3649  43.92   98.5108  136.2603
  G61  118.0595  9.3649  43.92   99.1847  136.9342
  G62  118.6259  9.3649  43.92   99.7512  137.5006
  G63  113.3607  9.3649  43.92   94.4859  132.2354
  G64  120.9077  9.3649  43.92  102.0330  139.7825
Confidence level: 0.9500
```

```
> st.lsm(
+     m,
+     ~gen,
+     at = list(county = "C5"),
```

```
+     method = "pairwise",
+     adjust = "none",
+     level = 0.9
+ )
```

```
Least-squares mean contrasts for gen (Satterthwaite)
At county = C5
  contrast   estimate      SE      DF  lower.CL  upper.CL        t       p
  G01 - G02  -11.1682  9.6165  117.53  -27.1116    4.7752  -1.1614  0.2478
  G01 - G03    9.2010  9.6165  117.53   -6.7424   25.1443   0.9568  0.3406
  G01 - G04   -3.3450  9.6165  117.53  -19.2884   12.5984  -0.3478  0.7286
  G01 - G05  -31.1760  9.6165  117.53  -47.1194  -15.2327  -3.2419  0.0015
  G01 - G06   -2.6036  9.6165  117.53  -18.5469   13.3398  -0.2707  0.7871
  G01 - G07   -5.3360  9.6165  117.53  -21.2794   10.6073  -0.5549  0.5800
  G01 - G08  -15.2593  9.6165  117.53  -31.2027    0.6840  -1.5868  0.1152
  G01 - G09  -25.6315  9.4242  115.21  -41.2586  -10.0044  -2.7197  0.0075
  G01 - G10   21.3486  9.6165  117.53    5.4052   37.2919   2.2200  0.0283
  G01 - G11   -5.7331  9.4242  115.21  -21.3603    9.8940  -0.6083  0.5442
  [2001 table rows omitted]
  G61 - G63    4.6988  9.6165  117.53  -11.2445   20.6422   0.4886  0.6260
  G61 - G64   -2.8482  9.6165  117.53  -18.7916   13.0951  -0.2962  0.7676
  G62 - G63    5.2653  9.6165  117.53  -10.6781   21.2086   0.5475  0.5851
  G62 - G64   -2.2818  9.6165  117.53  -18.2252   13.6615  -0.2373  0.8129
  G63 - G64   -7.5471  9.6165  117.53  -23.4904    8.3963  -0.7848  0.4341
p-value adjustment: none; confidence level: 0.9000
```

```
> st.lsm.cld(
+     st.lsm(
+         m,
+         ~gen,
+         at = list(county = "C5")
+     ),
+     adjust = "none",
+     alpha = 0.1
+ )
```

```
Least-squares means (Satterthwaite)
At county = C5
  gen    lsmeans      SE     DF  lower.CL  upper.CL  group
  G10    86.2014  9.3649  43.92   67.3267  105.0762  a
  G33    90.0151  9.3649  43.92   71.1403  108.8898  ab
  G12    93.6103  9.3649  43.92   74.7356  112.4851  abc
  G29    95.6117  9.3649  43.92   76.7370  114.4865  abcd
  G34    96.3081  9.3649  43.92   77.4334  115.1829  abcde
  G57    97.4050  9.3649  43.92   78.5302  116.2797  abcdef
  G03    98.3491  9.3649  43.92   79.4743  117.2238  abcdefg
  G31   104.1480  9.3649  43.92   85.2732  123.0227   bcdefgh
  G18   104.8013  9.3649  43.92   85.9265  123.6760   bcdefgh
  G37   106.2197  9.3649  43.92   87.3450  125.0945    cdefghi
  [49 table rows omitted]
  G54   132.3220  9.3649  43.92  113.4473  151.1968                  rstuv
  G09   133.1815  9.3649  43.92  114.3068  152.0563                   stuv
  G15   135.0739  9.3649  43.92  116.1991  153.9486                    tuv
  G45   136.4691  9.3649  43.92  117.5944  155.3438                     uv
  G05   138.7260  9.3649  43.92  119.8513  157.6008                      v
Confidence level: 0.9500
Comparison adjustment: none (alpha=0.1)
Shared letters indicate failure to reject a difference.
```

4.2 Genetic and environmental variance components in soybean

M26: RLab-PopGen/q046-05-s1.R

```
> data("v-lmm-soybean", package = "SelectionTools")
> st.lmm(
+     fixed = yield ~ 1,
+     random = ~gen + env + gen:env,
+     cov = ~at(env):units,
+     data = soybean
+ )
```

```
Linear mixed model fitted by REML

Model:
  Data:      soybean
  Fixed:     yield ~ 1
  Random:    ~gen + env + gen:env
  Residual:  ~at(env):units

Convergence:
  converged: projected score tolerance met

Random effects (likelihood-ratio tests):
  Term     DF        LR        p
  gen       1   17.5889  <0.0001
  env       1  101.2509  <0.0001
  gen:env   1   16.4664  <0.0001

Variance components:
  Component                      Variance
  gen                             4.0333
  env                            42.9413
  gen:env                         4.4857
  Residual: at(env):units[A80]   10.7902
  Residual: at(env):units[A81]   16.1877
  Residual: at(env):units[A82]    6.7761
  Residual: at(env):units[G80]   15.6793
  Residual: at(env):units[G81]   14.5873
  Residual: at(env):units[G82]   10.7371
  Residual: at(env):units[L80]   18.5159
  Residual: at(env):units[L81]    4.9037
  Residual: at(env):units[L82]    5.0177
  Residual: at(env):units[V80]   11.1649
  Residual: at(env):units[V81]   18.6497
  Residual: at(env):units[V82]   14.2764

Fit and model statistics:
  Observations                306    R-squared (conditional)    0.8067
  Log-likelihood        -870.3352    RMSE (conditional)         3.1937
  AIC                  1772.6705    Variance: fixed            0.0000
  BIC                  1832.2478    Variance: random          51.4603
  R-squared (marginal)    0.0000    Variance: residual        12.3317
```

4.3 Adjusted genotype means in an unbalanced barley trial

M48: RLab-PopGen/q046-30-s1.R

The barley data with model factors and complete trait measurements are available through `data("v-lmm-barley-seven-traits", package = "SelectionTools")`.

```
> m <- st.lmm(
+     fixed = HLW ~ gen,
+     random = ~env + gen:env + env:rep,
+     cov = ~at(env):units,
+     data = barley.seven.traits
```

```
+ )
> m
```

```
Linear mixed model fitted by REML

Model:
  Data:      barley.seven.traits
  Fixed:     HLW ~ gen
  Random:    ~env + gen:env + env:rep
  Residual:  ~at(env):units

Convergence:
  converged: projected score tolerance met

Fixed effects (Satterthwaite):
  Term  DF1    DF2       F       p
  gen    27  14.06  6.3989  0.0004

Random effects (likelihood-ratio tests):
  Term     DF       LR        p
  env       1  35.8902  <0.0001
  env:gen   1   1.2517   0.2632
  env:rep   1  <0.0001   1.0000

Variance components:
  Component                            Variance
  env                                   12.6535
  env:gen                                0.3696
  env:rep                                0.0000
  Residual: at(env):units[y2019loc1]     2.1181
  Residual: at(env):units[y2019loc2]     9.1668
  Residual: at(env):units[y2019loc3]     4.6407
  Residual: at(env):units[y2020loc1]     1.2217
  Residual: at(env):units[y2020loc2]    24.5283

Fit and model statistics:
  Observations                 291    R-squared (conditional)    0.6550
  Log-likelihood        -636.0887    RMSE (conditional)         2.7740
  AIC                  1344.1775    Variance: fixed            2.8790
  BIC                  1476.4171    Variance: random          13.0231
  R-squared (marginal)    0.1186    Variance: residual         8.3759
```

```
> st.lsm(m, "gen")
```

```
Least-squares means (Satterthwaite)
  gen    lsmeans      SE    DF  lower.CL  upper.CL
  lne01  70.1363  1.6924     5   65.7870   74.4856
  lne02  68.0981  1.6924     5   63.7488   72.4474
  lne03  69.0730  1.6924     5   64.7237   73.4222
  lne04  66.3690  1.7805  6.10   62.0295   70.7086
  lne05  69.5026  1.6924     5   65.1533   73.8518
  lne06  68.6917  1.6924     5   64.3424   73.0410
  lne07  70.1422  1.6924     5   65.7930   74.4915
  lne08  65.1853  1.6924     5   60.8360   69.5345
  lne09  65.0133  1.6924     5   60.6640   69.3626
  lne10  69.1793  1.6924     5   64.8300   73.5285
  [13 table rows omitted]
  lne24  66.7102  1.7710  5.98   62.3727   71.0477
  lne25  66.1400  1.7710  5.98   61.8025   70.4776
  lne26  69.3345  1.8157  6.40   64.9586   73.7105
  lne27  69.2464  1.6924     5   64.8972   73.5957
  lne28  65.1226  1.8223  6.47   60.7406   69.5046
Confidence level: 0.9500
```

5 Heritability

5.1 Balanced barley trial across years and locations

RLab-PopGen/q046-10-s1.R

5.1.1 Heritability from variance components

```
> data("v-lmm-barley-met", package = "SelectionTools")
> xtabs(~loc + year, data = barley.met)
```

```
     year
loc   y2016 y2017 y2018
  HOH    60    60    60
  EWE    60    60    60
  EIN    60    60    60
  OLI    60    60    60
  HFH    60    60    60
```

```
> m.barley.reml <- st.lmm(
+     fixed = TKW ~ 1,
+     random = ~gen + loc + year + gen:loc + gen:year + loc:year +
+         gen:loc:year + rep:loc:year,
+     data = barley.met
+ )
> v <- st.lmm.components(m.barley.reml, "variance")[, "TKW"]
> v.g <- as.numeric(v[["gen"]])
> v.g
```

```
[1] 11.72524
```

```
> L <- nlevels(barley.met$loc)
> Y <- nlevels(barley.met$year)
> R <- nlevels(barley.met$rep)
> v.masking <- v[["gen:loc"]] / L +
+     v[["gen:year"]] / Y +
+     v[["gen:loc:year"]] / (L * Y) +
+     v[["units"]] / (L * Y * R)
> v.masking
```

```
[1] 5.634385
```

```
> heritability <- v.g / (v.g + v.masking)
> heritability
```

```
[1] 0.6754316
```

5.1.2 Heritability from the SE of genotype differences

```
> m.barley.means <- st.lmm(
+     fixed = TKW ~ gen + loc + year + loc:year + year:loc:rep,
+     random = ~gen:loc + gen:year + gen:loc:year,
+     data = barley.met
+ )
```

The pairwise comparisons supply the SE of every genotype difference (SED). Half their mean square estimates the masking variance in the BLUE formula discussed by Piepho and Möhring (2007, Eq. 19). It agrees with the variance-component calculation in this balanced trial.

```
> barley.comparisons <- as.data.frame(st.lsm(
+     m.barley.means,
+     ~gen,
+     method = "pairwise",
+     adjust = "none"
+ ))
> v.masking <- mean(barley.comparisons$SE^2) / 2
> v.masking
```

```
[1] 5.634385
```

```
> heritability <- v.g / (v.g + v.masking)
> heritability
```

```
[1] 0.6754316
```

5.1.3 Heritability from prediction errors of BLUP differences

The random-genotype fit also supplies BLUPs and their joint prediction-error covariance. The following calculation uses the mean prediction-error variance of their pairwise differences. The Cullis measure (Cullis et al., 2006), discussed by Piepho and Möhring (2007, Eq. 20), is one minus this variance divided by twice the genetic variance.

```
> PEV <- vcov(st.predict(m.barley.reml, "0 + gen"))
> PEV.differences <- outer(diag(PEV), diag(PEV), "+") - 2 * PEV
> v.diff <- mean(PEV.differences[upper.tri(PEV.differences)])
> v.diff
```

```
[1] 7.611283
```

```
> heritability <- 1 - v.diff / (2 * v.g)
> heritability
```

```
[1] 0.6754316
```

5.2 Barley with missing location–year combinations

```
> data("v-lmm-met-missing", package = "SelectionTools")
> xtabs(~loc + year, data = met.missing, subset = !is.na(YLD))
```

```
      year
loc    y2017 y2018 y2019
  loc1     0     0    80
  loc2    80    80    80
  loc3    80    80    79
  loc4     0     0    80
  loc5     0     0    79
```

```
> met.vc <- st.lmm(
+     fixed = YLD ~ 1,
+     random = ~gen + loc + year + gen:loc + gen:year + loc:year +
+         gen:loc:year + year:loc:rep,
+     data = met.missing
+ )
> v <- st.lmm.components(met.vc, "variance")
> v.g <- as.numeric(v["gen", "YLD"])
> v.g
```

```
[1] 56531.11
```

The factor `REP` identifies the location–year–replicate blocks present in the trial. With `rep*year*loc` in the fixed part, the default reference grid also includes combinations for trials that were never conducted, making the genotype means non-estimable. With `REP`, the means are averaged over the observed blocks and their SEs can be calculated (Section 2.6.3).

```
> met.missing$REP <- with(met.missing, rep:year:loc)
> met.means <- st.lmm(
+     fixed = YLD ~ gen + REP,
+     random = ~gen:loc + gen:year + gen:loc:year,
+     data = met.missing
+ )
```

Half the mean squared SE of the genotype differences estimates the masking variance in this unbalanced trial.

```
> met.comparisons <- as.data.frame(st.lsm(
+     met.means,
+     ~gen,
+     method = "pairwise",
+     adjust = "none"
+ ))
> v.masking <- mean(met.comparisons$SE^2) / 2
> v.masking
```

```
[1] 38385.1
```

```
> heritability <- v.g / (v.g + v.masking)
> heritability
```

```
[1] 0.5955896
```

The same BLUP prediction-error calculation uses the random-genotype model `met.vc` and its estimate of genetic variance.

```
> PEV <- vcov(st.predict(met.vc, "0 + gen"))
> PEV.differences <- outer(diag(PEV), diag(PEV), "+") - 2 * PEV
> v.diff <- mean(PEV.differences[upper.tri(PEV.differences)])
> v.diff
```

```
[1] 44574.21
```

```
> heritability <- 1 - v.diff / (2 * v.g)
> heritability
```

```
[1] 0.6057551
```

6 Multivariate Experiments

6.1 Moment estimators

6.1.1 Two species and two traits: Hotelling test

`CM09: RLab-BioStat/v19-u36.R`

The Hotelling–Lawley trace gives the exact F test for these two species. The other multivariate statistics give the same F and p for this comparison.

```
> data("v-lmm-iris-two", package = "SelectionTools")
> m.hotelling <- st.lmm(
+     cbind(SepalLength, SepalWidth) ~ Species,
+     data = iris.two,
+     method = "moments"
+ )
> anova(m.hotelling, effects = "manova", test = "Hotelling-Lawley")
```

```
MANOVA for Species
  Method           Statistic  DF1  DF2       F       p  Reference
  Hotelling-Lawley     0.6194    2   17  5.2650  0.0166  exact F
```

6.1.2 Three species and four traits: MANOVA

`CM10: RLab-BioStat/v19-u36.R`

The MANOVA table reports the four multivariate statistics for species and the reference calculation for each F statistic.

```
> data("v-lmm-iris", package = "SelectionTools")
> m.iris <- st.lmm(
+     cbind(SepalLength, SepalWidth, PetalLength, PetalWidth) ~ Species,
+     data = iris.manova,
+     method = "moments"
+ )
> anova(m.iris, effects = "manova")
```

```
MANOVA for Species
  Method            Statistic  DF1  DF2         F        p  Reference
  Pillai               1.1919    8  290   53.4665  <0.0001  approximate F
  Wilks                0.0234    8  288  199.1453  <0.0001  approximate F
  Hotelling-Lawley    32.4773    8  286  580.5321  <0.0001  approximate F
  Roy                 32.1919    4  145     >1000  <0.0001  F upper bound
```

The `sscp.H` and `sscp.E` components give the sums-of-squares and cross-products matrices for the species hypothesis and the error.

```
> st.lmm.components(m.iris, "sscp.H", term = "Species")
```

```
             SepalLength  SepalWidth  PetalLength  PetalWidth
  SepalLength     63.2121    -19.9527     165.2484     71.2793
  SepalWidth     -19.9527     11.3449     -57.2396    -22.9327
  PetalLength    165.2484    -57.2396     437.1028    186.7740
  PetalWidth      71.2793    -22.9327     186.7740     80.4133
```

```
> st.lmm.components(m.iris, "sscp.E")
```

```
             SepalLength  SepalWidth  PetalLength  PetalWidth
  SepalLength     38.9562     13.6300      24.6246      5.6450
  SepalWidth      13.6300     16.9620       8.1208      4.8084
  PetalLength     24.6246      8.1208      27.2226      6.2718
  PetalWidth       5.6450      4.8084       6.2718      6.1566
```

6.1.3 Multivariate maize: Genetic covariances and correlations

`M60: RLab-BioStat/v19-u36.R, section 4`

Each genotype has one observation in every environment and replicate. The ANOVA strata separate the sources of variation in this balanced design. The sums of squares and cross-products in each stratum, divided by its DF, form a mean-product matrix. These matrices provide the information used to estimate the variance and covariance components.

```
> data("v-lmm-maize-met", package = "SelectionTools")
> m.maize.three <- st.lmm(
+     fixed = cbind(yld, dmc, pht) ~ 1,
+     random = ~gen + env + gen:env + rep:env,
+     data = maize.met,
+     method = "moments",
+     random.test = "F",
+     control = list(design = "general")
+ )
> m.maize.three
```

```
Multivariate linear mixed model fitted by the method of moments

Model:
  Data:      maize.met
  Fixed:     cbind(yld, dmc, pht) ~ 1
  Random:    ~gen + env + gen:env + rep:env
  Residual:  ~units
```

```
Design: general balanced orthogonal

ANOVA strata:
  Source     DF
  gen        24
  env        23
  gen:env   552
  env:rep    24
  Residual  576

Random-effect F tests by trait (term-specific tests):
  Source   Trait    DF1    DF2           SS          MS        F        p
  gen      yld       24    552   41507.1168   1729.4632  25.4559  <0.0001
  gen      dmc       24    552    3174.7852    132.2827  26.6975  <0.0001
  gen      pht       24    552  163618.6250   6817.4427  66.5983  <0.0001
  env      yld    23.23  41.15  202635.2081   8810.2264  31.0079  <0.0001
  env      dmc    23.14  41.44   23058.9308   1002.5622  48.8285  <0.0001
  env      pht    23.15  48.32  471250.9167  20489.1703  59.8423  <0.0001
  gen:env  yld      552    576   37502.6863     67.9396   1.5790  <0.0001
  gen:env  dmc      552    576    2735.0886      4.9549   1.6418  <0.0001
  gen:env  pht      552    576   56506.3750    102.3666   1.5471  <0.0001
  env:rep  yld       24    576    5221.8438    217.5768   5.0569  <0.0001
  env:rep  dmc       24    576     375.3424     15.6393   5.1820  <0.0001
  env:rep  pht       24    576    5787.0000    241.1250   3.6441  <0.0001

Test methods:
  gen: exact F
  env: quasi F (Satterthwaite)
  gen:env: exact F
  env:rep: exact F
  env [yld]: approximate
  env [dmc]: approximate
  env [pht]: approximate

Covariance components:
  gen:
          yld      dmc       pht
  yld  34.6151  -8.7208   35.6158
  dmc  -8.7208   2.6527   -6.7633
  pht  35.6158  -6.7633  139.8974

  env:
           yld      dmc       pht
  yld  171.3547  18.3650  168.6739
  dmc   18.3650  19.6997   35.9687
  pht  168.6739  35.9687  404.2369

  gen:env:
          yld     dmc      pht
  yld  12.4569  0.2046   5.4491
  dmc   0.2046  0.9684   0.5616
  pht   5.4491  0.5616  18.0991

  [1 covariance block omitted]

  Residual:
          yld      dmc      pht
  yld  43.0258   2.5554  15.4257
  dmc   2.5554   3.0180  -0.2425
  pht  15.4257  -0.2425  66.1684

  Diagonal entries are variances; off-diagonal entries are covariances.

Fit and model statistics:

  Trait: yld
  Observations  1200    Residual SD  6.5594
  Residual DF    576
```

```
Trait: dmc
Observations  1200    Residual SD  1.7372
Residual DF    576

Trait: pht
Observations  1200    Residual SD  8.1344
Residual DF    576
```

The argument `random.test = "F"` requests an F test for each random effect and trait. The genotype test uses the genotype-by-environment mean square as its denominator. The genotype-by-environment and replicate-within-environment tests use the residual mean square. The environment test combines mean squares from several strata and uses Satterthwaite DF.

The `control` list entry `print.mean.products = TRUE` adds the mean-product matrices to the model summary.

The covariance block for `gen` describes how the random genotype effects vary together across traits.

```
> st.lmm.components(m.maize.three, "covariance", block = "gen")
```

```
          yld      dmc       pht
yld   34.6151  -8.7208   35.6158
dmc   -8.7208   2.6527   -6.7633
pht   35.6158  -6.7633  139.8974
```

The genetic correlations put these associations on a common scale.

```
> st.lmm.components(m.maize.three, "correlation", block = "gen")
```

```
          yld      dmc      pht
yld    1.0000  -0.9101   0.5118
dmc   -0.9101   1.0000  -0.3511
pht    0.5118  -0.3511   1.0000
```

6.1.4 Multivariate maize: Adjusted genotype means

The model treats genotype as fixed and accounts for environments, genotype-by-environment effects and replicates within environments as random effects. The method of moments estimates covariances between traits under the default residual formula. For this complete balanced trial, the 75 least-squares mean estimates agree with the REML results in Section 6.2.

```
> data("v-lmm-maize-met", package = "SelectionTools")
> m.maize.moments.means <- st.lmm(
+     fixed = cbind(yld, dmc, pht) ~ gen,
+     random = ~env + gen:env + rep:env,
+     data = maize.met,
+     method = "moments",
+     control = list(design = "general")
+ )
> maize.moments.lsmeans <- as.data.frame(st.lsm(m.maize.moments.means, ~gen))
> maize.moments.means <- xtabs(lsmeans ~ gen + trait,
+     data = maize.moments.lsmeans)[, c("yld", "dmc", "pht")]
> maize.moments.means
```

```
     trait
gen         yld       dmc       pht
  G01  55.94583  60.38292 167.29167
  G02  63.92854  58.28313 177.08333
  G03  65.31937  58.20021 159.06250
  G04  57.96688  61.20521 168.22917
  G05  56.48250  61.41000 166.66667
  G06  71.13375  56.76271 177.70833
  G07  76.85604  56.31000 177.18750
  G08  59.39438  60.51542 176.14583
  G09  60.42375  61.87375 169.16667
  G10  63.60792  60.46000 166.66667
  G11  57.47250  62.19396 165.72917
  G12  65.40354  59.32750 181.77083
  G13  55.43938  62.09354 174.47917
  G14  58.23500  60.57417 180.93750
  G15  53.52208  61.54167 165.93750
  G16  56.98000  62.10979 172.60417
  G17  60.00771  60.10313 173.85417
  G18  62.31188  60.69292 179.06250
  G19  68.03646  58.06188 187.18750
  G20  61.84729  59.91625 184.16667
  G21  66.80687  58.50875 206.87500
  G22  66.32625  59.61979 198.33333
  G23  61.65354  60.87521 194.37500
  G24  72.69021  59.38333 198.64583
  G25  71.03896  58.04979 185.62500
```

The phenotypic covariance matrix describes how the 25 adjusted genotype means vary together across genotypes.

```
> round(cov(maize.moments.means), 4)
```

```
        yld     dmc      pht
yld 36.0305 -8.6590  36.1642
dmc -8.6590  2.7559  -6.7449
pht 36.1642 -6.7449 142.0301
```

The phenotypic correlations express these associations on a common scale.

```
> round(cor(maize.moments.means), 4)
```

```
        yld     dmc     pht
yld  1.0000 -0.8690  0.5055
dmc -0.8690  1.0000 -0.3409
pht  0.5055 -0.3409  1.0000
```

The means for G01 and G02 combine several error strata, giving Satterthwaite DF. For their difference, the shared environment effects cancel, and the genotype-by-environment stratum gives the exact DF for each trait.

```
> st.lsm(
+     m.maize.moments.means,
+     ~gen,
+     at = list(gen = c("G01", "G02"))
+ )
```

```
Least-squares means (Satterthwaite)
At gen = G01, G02
  gen  trait   lsmeans      SE     DF  lower.CL  upper.CL
  G01  yld     55.9458  2.9497  32.26   49.9394   61.9523
  G02  yld     63.9285  2.9497  32.26   57.9221   69.9350
  G01  dmc     60.3829  0.9667  28.76   58.4050   62.3608
  G02  dmc     58.2831  0.9667  28.76   56.3052   60.2610
  G01  pht    167.2917  4.3728  28.83  158.3459  176.2374
  G02  pht    177.0833  4.3728  28.83  168.1376  186.0291
Confidence level: 0.9500
```

```
> st.lsm(
+     m.maize.moments.means,
+     ~gen,
+     at = list(gen = c("G01", "G02")),
+     method = "pairwise",
+     adjust = "none"
+ )
```

```
Least-squares mean contrasts for gen (exact)
At gen = G01, G02
  contrast  trait estimate     SE  DF lower.CL upper.CL       t       p
  G01 - G02 yld    -7.9827 1.6825 552 -11.2876  -4.6778 -4.7445 <0.0001
  G01 - G02 dmc     2.0998 0.4544 552   1.2073   2.9923  4.6213 <0.0001
  G01 - G02 pht    -9.7917 2.0653 552 -13.8484  -5.7349 -4.7411 <0.0001
p-value adjustment: none; confidence level: 0.9500
```

6.2 Adjusted genotype means for three traits

In the maize trial, grain yield, dry-matter content and plant height were recorded on the same 1200 plots. A three-trait REML model treats genotype as fixed. The random terms account for environments, genotype-by-environment effects and replicates within environments. The residual covariance allows the three measurements from the same plot to be correlated.

The least-squares means give one adjusted mean per genotype and trait. The table places the three trait means side by side for all 25 genotypes. Column indexing sets the trait order to `yld`, `dmc`, and `pht`.

```
> data("v-lmm-maize-met", package = "SelectionTools")
> m.maize.means <- st.lmm(
+     fixed = cbind(yld, dmc, pht) ~ gen,
+     random = ~env + gen:env + rep:env,
+     data = maize.met
+ )
> maize.lsmeans <- as.data.frame(st.lsm(m.maize.means, ~gen))
> maize.means <- xtabs(lsmeans ~ gen + trait,
+     data = maize.lsmeans)[, c("yld", "dmc", "pht")]
> maize.means
```

```
     trait
gen         yld       dmc       pht
  G01  55.94583  60.38292 167.29167
  G02  63.92854  58.28313 177.08333
  G03  65.31937  58.20021 159.06250
  G04  57.96687  61.20521 168.22917
  G05  56.48250  61.41000 166.66667
  G06  71.13375  56.76271 177.70833
  G07  76.85604  56.31000 177.18750
  G08  59.39437  60.51542 176.14583
```

```
  G09  60.42375  61.87375 169.16667
  G10  63.60792  60.46000 166.66667
  G11  57.47250  62.19396 165.72917
  G12  65.40354  59.32750 181.77083
  G13  55.43937  62.09354 174.47917
  G14  58.23500  60.57417 180.93750
  G15  53.52208  61.54167 165.93750
  G16  56.98000  62.10979 172.60417
  G17  60.00771  60.10313 173.85417
  G18  62.31187  60.69292 179.06250
  G19  68.03646  58.06188 187.18750
  G20  61.84729  59.91625 184.16667
  G21  66.80687  58.50875 206.87500
  G22  66.32625  59.61979 198.33333
  G23  61.65354  60.87521 194.37500
  G24  72.69021  59.38333 198.64583
  G25  71.03896  58.04979 185.62500
```

The phenotypic covariance matrix describes how the adjusted genotype means vary together across genotypes.

```
> round(cov(maize.means), 4)
```

```
        yld     dmc      pht
yld 36.0305 -8.6590  36.1642
dmc -8.6590  2.7559  -6.7449
pht 36.1642 -6.7449 142.0301
```

The phenotypic correlations express these associations on a common scale.

```
> round(cor(maize.means), 4)
```

```
        yld     dmc     pht
yld  1.0000 -0.8690  0.5055
dmc -0.8690  1.0000 -0.3409
pht  0.5055 -0.3409  1.0000
```

6.3 Genetic variances and correlations in maize

In a second three-trait REML model, genotype is random. The diagonal of the genotype covariance matrix gives the genetic variance for each trait, and the off-diagonal entries give the genetic covariances. Section 6.1.3 presents the moment estimates for the same trial.

```
> data("v-lmm-maize-met", package = "SelectionTools")
> m.maize.genetic <- st.lmm(
+     fixed = cbind(yld, dmc, pht) ~ 1,
+     random = ~gen + env + gen:env + rep:env,
+     data = maize.met
+ )
> st.lmm.components(m.maize.genetic, "covariance", block = "gen")
```

```
          yld      dmc       pht
  yld  34.6151  -8.7208   35.6158
  dmc  -8.7208   2.6527   -6.7633
  pht  35.6158  -6.7633  139.8974
```

Table 5: Three-trait maize fits in R 4.6.1 on an AMD EPYC 9V74 processor.

	Reference BLAS		OpenBLAS, four threads	
Fit	Time (s)	Peak RAM (MiB)	Time (s)	Peak RAM (MiB)
REML, adjusted genotype means	51.2	107.4	10.1	110.4
REML, genetic covariance	30.3	96.6	15.8	99.3
Moments, adjusted genotype means	0.019	74.1	0.025	75.6
Moments, genetic covariance	0.004	69.4	0.004	70.4

The genetic correlations describe the associations on a common scale.

```
> st.lmm.components(m.maize.genetic, "correlation", block = "gen")
```

```
          yld      dmc      pht
yld    1.0000  -0.9101   0.5118
dmc   -0.9101   1.0000  -0.3511
pht    0.5118  -0.3511   1.0000
```

Table 5 compares the three-trait fits on the 1200-plot maize trial. Times cover the `st.lmm()` call, including its default tests. Peak memory is the resident memory of a fresh R process. The reference BLAS figures come from one run with one SelectionTools thread. The OpenBLAS figures are medians of three runs with four threads each for OpenBLAS and SelectionTools.

6.4 Three traits with heterogeneous residuals

M22: barley-met

```
> data("v-lmm-barley-met", package = "SelectionTools")
> barley.met$env <- factor(paste(barley.met$year, barley.met$loc))
> m.barley <- st.lmm(
+     fixed = cbind(GYLD, TKW, PLH) ~ 1,
+     random = ~gen + loc + year + gen:loc + gen:year + loc:year +
+         gen:loc:year + year:loc:rep,
+     cov = ~at(env):units,
+     data = barley.met,
+     control = list(maxit = 1000)
+ )
> m.barley
```

```
Multivariate linear mixed model fitted by REML

Model:
  Data:      barley.met
  Fixed:     cbind(GYLD, TKW, PLH) ~ 1
  Random:    ~gen + loc + year + gen:loc + gen:year + loc:year +
               gen:loc:year + year:loc:rep
  Residual:  ~at(env):units

Convergence:
  converged: projected score tolerance met

Random effects (joint likelihood-ratio tests):
  Term          DF        LR        p
  gen            6   55.3057  <0.0001
  loc            6   22.3887   0.0010
  year           6    2.6896   0.8467
  gen:loc        6   44.6793  <0.0001
  gen:year       6  152.3380  <0.0001
```

```
  loc:year       6  224.1904  <0.0001
  gen:loc:year   6  104.6506  <0.0001
  loc:year:rep   6   86.4042  <0.0001
  Chi-squared p values are approximate.

Covariance components:
  gen:
              GYLD       TKW        PLH
  GYLD  33118.5034  251.2933  -398.3863
  TKW     251.2933   11.3569   -13.8987
  PLH    -398.3863  -13.8987    45.6279

  loc:
               GYLD        TKW        PLH
  GYLD  1295334.4978  3811.5880  7434.0024
  TKW      3811.5880    28.7503    52.8801
  PLH      7434.0024    52.8801    97.4887

  year:
              GYLD      TKW        PLH
  GYLD  48694.6640  44.9780  2049.1070
  TKW      44.9780   1.6052    -2.8263
  PLH    2049.1070  -2.8263   100.4697

  [19 covariance blocks omitted]

  Residual: at(env):units[y2018 OLI]
               GYLD       TKW        PLH
  GYLD  159606.1375  242.5695  -391.3643
  TKW      242.5695    5.0352    -0.5914
  PLH     -391.3643   -0.5914    21.1576

  Diagonal entries are variances; off-diagonal entries are covariances.

Fit and model statistics:
  Observations           2700    AIC  24842.1440
  Log-likelihood  -12280.0720    BIC  25674.1860

  Trait: GYLD
  Observations                 900    Variance: fixed           0.0000
  R-squared (marginal)      0.0000    Variance: random    1920236.1832
  R-squared (conditional)   0.8723    Variance: residual   281082.3148
  RMSE (conditional)       486.6261

  Trait: TKW
  Observations                 900    Variance: fixed      0.0000
  R-squared (marginal)      0.0000    Variance: random    69.5423
  R-squared (conditional)   0.8917    Variance: residual   8.4422
  RMSE (conditional)        2.6227

  Trait: PLH
  Observations                 900    Variance: fixed       0.0000
  R-squared (marginal)      0.0000    Variance: random    461.8351
  R-squared (conditional)   0.9171    Variance: residual   41.7365
  RMSE (conditional)        5.9652
```

The variance matrix has one column per trait and one row per random term or residual environment.

```
> st.lmm.components(m.barley, "variance", traits = c("GYLD", "TKW", "PLH"))
```

```
                                GYLD      TKW       PLH
  gen                     33118.5034  11.3569   45.6279
  loc                   1295334.4978  28.7503   97.4887
  year                    48694.6640   1.6052  100.4697
  gen:loc                 72689.7183   4.4969    4.1910
```

```
gen:year                     48316.1541  13.4033    6.7910
loc:year                    297634.7000   6.1589  193.8146
gen:loc:year                 79773.3218   3.3618    7.1016
loc:year:rep                 44674.6240   0.4091    6.3507
at(env):units[y2016 EIN]    281758.2800   2.5008   15.4647
at(env):units[y2016 EWE]    263157.7974   5.3767   74.8042
[8 table rows omitted]
at(env):units[y2018 EIN]    146430.1057  21.8838   18.2708
at(env):units[y2018 EWE]    264446.6657   5.5689   17.0405
at(env):units[y2018 HFH]    390998.5616   8.3880   36.9109
at(env):units[y2018 HOH]    274022.2270   7.6614   81.4252
at(env):units[y2018 OLI]    159606.1375   5.0352   21.1576
```

The `gen` block contains the covariance matrix of genotype effects across traits.

```
> st.lmm.components(m.barley, "covariance", block = "gen")
```

```
             GYLD       TKW        PLH
GYLD  33118.5034  251.2933  -398.3863
TKW     251.2933   11.3569   -13.8987
PLH    -398.3863  -13.8987    45.6279
```

6.5 Unbalanced trial with seven traits and heterogeneous environments in barley

`M47: RLab-PopGen/q046-30-s1.R`

The call `data("v-lmm-barley-seven-traits", package = "SelectionTools")` loads `barley.seven.traits`, which contains the model factors and 291 plots with measurements for all seven traits, for a total of 2,037 response values. The first table gives the number of plots in each replicate within each environment.

```
> addmargins(xtabs(~ env + rep, data = barley.seven.traits))
```

```
           rep
env          r1  r2  r3 Sum
  y2019loc1  19  18  20  57
  y2019loc2  20  18  20  58
  y2019loc3  18  19  20  57
  y2020loc1  20  20  20  60
  y2020loc2  19  20  20  59
  Sum        96  95 100 291
```

The second table gives the number of plots for each genotype in each environment. A zero indicates that there are no retained plots for that combination.

```
> addmargins(xtabs(~ gen + env, data = barley.seven.traits))
```

```
       env
gen     y2019loc1 y2019loc2 y2019loc3 y2020loc1 y2020loc2 Sum
  lne01         3         3         3         3         3  15
  lne02         3         3         3         3         3  15
  lne03         3         3         3         3         3  15
  lne04         3         2         3         0         0   8
  lne05         3         3         3         3         3  15
  lne06         3         3         3         3         3  15
  lne07         3         3         3         3         3  15
  lne08         3         3         3         3         3  15
```

```
lne09     3     3     3     3     3  15
lne10     3     3     3     3     3  15
lne11     3     3     3     3     3  15
lne12     3     3     3     3     3  15
lne13     0     0     0     3     3   6
lne14     0     0     0     3     3   6
lne15     0     0     0     3     3   6
lne16     0     0     0     3     3   6
lne17     0     0     0     3     3   6
lne18     0     0     0     3     3   6
lne19     3     3     3     0     0   9
lne20     1     3     3     0     0   7
lne21     2     2     1     0     0   5
lne22     3     3     2     0     0   8
lne23     3     3     3     0     0   9
lne24     3     3     3     0     0   9
lne25     3     3     3     0     0   9
lne26     0     0     0     3     3   6
lne27     3     3     3     3     3  15
lne28     0     0     0     3     2   5
Sum      57    58    57    60    59 291
```

The model includes fixed environment effects and fixed replicate effects within environments, separately for each trait. The random genotype and genotype-by-environment effects each have a full covariance matrix across the seven traits. A separate full residual covariance matrix is estimated for each environment. The genotype covariance matrix describes the relationships among genotype effects shared across environments.

```
> m.met.seven <- st.lmm(
+     fixed = cbind(YLD, HLW, TKW, PLH, DH, DM, SC) ~ env + rep:env,
+     random = ~gen + gen:env,
+     cov = ~at(env):units,
+     data = barley.seven.traits,
+     random.test = "none",
+     control = list(backend = "block")
+ )
> st.lmm.components(m.met.seven, "covariance", block = "gen")
```

```
          YLD      HLW      TKW       PLH        DH       DM       SC
YLD   60.3784   2.8817  24.9291  143.3567  -50.6606 -34.4759  -4.3092
HLW    2.8817   2.5724   2.7317   17.6852   -4.2309  -2.7620  -0.3568
TKW   24.9291   2.7317  17.3686   68.3571  -21.1640 -14.0188  -2.1689
PLH  143.3567  17.6852  68.3571  433.0834 -120.8631 -83.6531 -11.2377
DH   -50.6606  -4.2309 -21.1640 -120.8631   69.2974  48.1663   1.9682
DM   -34.4759  -2.7620 -14.0188  -83.6531   48.1663  36.0201   1.2429
SC    -4.3092  -0.3568  -2.1689  -11.2377    1.9682   1.2429   0.6694
```

The fitted genotype covariance is positive definite. The genotype-by- environment covariance has rank six, with a numerically zero eigenvalue.

The genetic correlation matrix expresses the same relationships on a common scale.

```
> st.lmm.components(m.met.seven, "correlation", block = "gen")
```

```
          YLD      HLW      TKW      PLH       DH       DM       SC
YLD    1.0000   0.2312   0.7698   0.8865  -0.7832  -0.7393  -0.6778
HLW    0.2312   1.0000   0.4087   0.5299  -0.3169  -0.2869  -0.2719
TKW    0.7698   0.4087   1.0000   0.7882  -0.6100  -0.5605  -0.6361
PLH    0.8865   0.5299   0.7882   1.0000  -0.6977  -0.6698  -0.6600
DH    -0.7832  -0.3169  -0.6100  -0.6977   1.0000   0.9641   0.2890
DM    -0.7393  -0.2869  -0.5605  -0.6698   0.9641   1.0000   0.2531
SC    -0.6778  -0.2719  -0.6361  -0.6600   0.2890   0.2531   1.0000
```

A second fit treats genotype as fixed and retains random genotype-by- environment effects. The least-squares means use the same reference grid of environments and replicate blocks for every genotype. The model estimates an expected value for combinations without retained plots. The table gives one adjusted mean per genotype and trait.

```
> m.met.seven.means <- st.lmm(
+     fixed = cbind(YLD, HLW, TKW, PLH, DH, DM, SC) ~ gen + env + rep:env,
+     random = ~gen:env,
+     cov = ~at(env):units,
+     data = barley.seven.traits,
+     random.test = "none",
+     control = list(backend = "block", optimizer = "AI")
+ )
> barley.lsmeans <- as.data.frame(st.lsm(m.met.seven.means, ~gen))
> barley.lsmeans.table <- round(xtabs(lsmeans ~ gen + trait, data =
barley.lsmeans)[
+     , c("YLD", "HLW", "TKW", "PLH", "DH", "DM", "SC")
+ ], 2)
> barley.lsmeans.table
```

```
       trait
gen        YLD    HLW    TKW    PLH     DH     DM     SC
  lne01  37.48  70.17  50.25 113.20  80.12 137.30   6.00
  lne02  37.08  67.74  42.49  97.24  77.88 133.62   7.05
  lne03  39.30  69.09  46.83 109.49  81.82 137.90   5.00
  lne04  27.28  65.76  38.69  59.19  83.59 141.78   7.88
  lne05  36.71  69.72  46.98 112.99  82.24 141.17   5.98
  lne06  39.66  68.75  47.48 107.09  84.78 137.47   4.78
  lne07  45.53  70.24  48.68 102.27  87.61 141.38   6.09
  lne08  40.74  65.71  53.52 102.69  82.45 136.61   5.49
  lne09  49.20  64.76  45.43 105.17  81.19 140.85   4.64
  lne10  37.42  69.01  45.22 101.88  85.84 140.25   5.59
  lne11  43.43  67.41  48.23  86.26  71.39 132.89   5.81
  lne12  43.36  67.76  50.71  94.75  71.01 133.44   5.60
  lne13  40.05  68.67  47.86  89.69  73.74 130.57   6.65
  lne14  43.42  66.74  45.68 102.65  76.18 134.00   5.70
  lne15  39.83  68.43  44.59  99.59  75.01 134.29   7.55
  lne16  38.69  67.29  49.69 103.58  76.10 134.22   6.13
  lne17  43.59  65.28  46.73  98.70  80.95 133.08   7.59
  lne18  37.35  69.63  46.26  98.56  77.97 138.09   3.88
  lne19  26.93  67.09  42.66  64.20  98.25 149.81   5.46
  lne20  18.50  65.22  38.83  52.40  95.63 147.12   6.95
  lne21  22.44  67.74  48.27  62.00  94.67 145.81   6.52
  lne22  20.36  68.11  38.07  51.27  93.28 143.03   7.32
  lne23  26.06  65.69  39.14  58.47  97.53 148.48   6.38
  lne24  24.97  66.98  42.82  61.15 100.96 162.27   6.84
  lne25  22.38  66.03  36.37  60.73  94.93 146.46   7.40
  lne26  40.86  69.37  46.76 106.25  78.30 135.26   7.06
  lne27  39.94  69.44  50.42 117.69  81.79 140.07   5.29
  lne28  43.08  64.62  46.15  89.63  88.67 144.13   5.45
```

7 Regression

7.1 Moment estimators

7.1.1 Fixed regression: Fertilizer dose and yield

CM11: RLab-BioStat/v19-u19.R; RLab-AppStat/v14-u61.R

```
> data("v-lmm-fertilizer-regression", package = "SelectionTools")
> m.regression <- st.lmm(
```

```
+     ertrag ~ dueng,
+     data = fertilizer.regression,
+     method = "moments"
+ )
> m.regression
```

```
Linear model fitted by the method of moments

Model:
  Data:      fertilizer.regression
  Fixed:     ertrag ~ dueng

ANOVA (method of moments, exact F):
  Source    DF1  DF2        SS        MS       F       p
  dueng       1    7  132.0167  132.0167  8.5185  0.0224
  Residual    7       108.4833   15.4976

Fixed-effect regression estimates (exact t):
  Coefficient  Estimate      SE  DF       t       p
  (Intercept)   12.3167  4.7585   7  2.5883  0.0360
  dueng          2.9667  1.0165   7  2.9186  0.0224

Model statistics:
  Trait        N  Residual DF      R2  Adjusted R2  Residual SD
  ertrag  9.0000            7  0.5489       0.4845       3.9367
```

7.1.2 Replicated fertilizer doses: Lack of fit

CM14: RLab-BioStat/v19-u19.R, section 2

```
> data("v-lmm-fertilizer-replicated", package = "SelectionTools")
> m.replicated <- st.lmm(
+     ertrag ~ dueng + factor(dueng),
+     data = fertilizer.replicated,
+     method = "moments"
+ )
> m.replicated
```

```
Linear model fitted by the method of moments

Model:
  Data:      fertilizer.replicated
  Fixed:     ertrag ~ dueng + factor(dueng)

ANOVA (method of moments, exact F):
  Source         DF1  DF2        SS       MS       F       p
  dueng            0    7    0.0000       NA      NA      NA
  factor(dueng)    2    7   95.5585  47.7792  1.4464  0.2980
  Residual         7       231.2250  33.0321
  dueng: not estimable

Model statistics:
  Trait         N  Residual DF      R2  Adjusted R2  Residual SD
  ertrag  11.0000            7  0.8394       0.7705       5.7474
```

7.1.3 Multiple regression: Partial tests and term order

The default Type III tests assess each predictor conditional on the other. Reversing the predictor order preserves these tests and the fitted model. CM12: RLab-BioStat/v19-u21.R

```
> data("v-lmm-rats", package = "SelectionTools")
> m.rats <- st.lmm(
+     y ~ x.1 + x.2,
+     data = rats,
+     method = "moments"
+ )
> m.rats
```

```
Linear model fitted by the method of moments

Model:
  Data:      rats
  Fixed:     y ~ x.1 + x.2

ANOVA (method of moments, exact F):
  Source    DF1  DF2         SS         MS        F       p
  x.1         1   32   291.0441   291.0441   3.0239  0.0917
  x.2         1   32  1060.1284  1060.1284  11.0145  0.0023
  Residual   32       3079.9400    96.2481

Fixed-effect regression estimates (exact t):
  Coefficient  Estimate       SE  DF       t       p
  (Intercept)   34.6257  20.6344  32  1.6781  0.1031
  x.1            0.4387   0.2523  32  1.7389  0.0917
  x.2            0.2173   0.0655  32  3.3188  0.0023

Model statistics:
  Trait        N  Residual DF      R2  Adjusted R2  Residual SD
  y      35.0000           32  0.3564       0.3162       9.8106
```

```
> m.rats.reverse <- st.lmm(
+     y ~ x.2 + x.1,
+     data = rats,
+     method = "moments"
+ )
> m.rats.reverse
```

```
Linear model fitted by the method of moments

Model:
  Data:      rats
  Fixed:     y ~ x.2 + x.1

ANOVA (method of moments, exact F):
  Source    DF1  DF2         SS         MS        F       p
  x.2         1   32  1060.1284  1060.1284  11.0145  0.0023
  x.1         1   32   291.0441   291.0441   3.0239  0.0917
  Residual   32       3079.9400    96.2481

Fixed-effect regression estimates (exact t):
  Coefficient  Estimate       SE  DF       t       p
  (Intercept)   34.6257  20.6344  32  1.6781  0.1031
  x.2            0.2173   0.0655  32  3.3188  0.0023
  x.1            0.4387   0.2523  32  1.7389  0.0917

Model statistics:
  Trait        N  Residual DF      R2  Adjusted R2  Residual SD
  y      35.0000           32  0.3564       0.3162       9.8106
```

```
> L.rats <- matrix(
+         c(0, 1, 0), nrow = 1,
+         dimnames = list("initial weight", names(coef(m.rats)))
+     )
> st.lsm(
+     m.rats,
+     L = L.rats,
+     joint = TRUE
+ )
```

```
Exact linear hypothesis
  DF1  DF2       F       p  F.scaling  method
    1   32  3.0239  0.0917     1.0000  exact
```

```
> m.rats.reduced <- st.lmm(
+     y ~ x.2,
+     data = rats,
+     method = "moments"
+ )
> anova(m.rats.reduced, m.rats)
```

```
Exact nested-model comparisons
  model          residual.DF      RSS DF DF2      SS     F      p partial.R2
  y ~ x.2                 33 3370.984
  y ~ x.1 + x.2           32 3079.940  1  32 291.044 3.024 0.0917      0.086
```

7.1.4 Multiple regression: Subset selection with mtcars

CM17: Motor Trend car measurements; RLab-StatLrn/q049-15-e1-s.R

```
> data("v-lmm-mtcars", package = "SelectionTools")
> m.mtcars.subsets <- st.lmm(
+     mpg ~ .,
+     data = mtcars,
+     method = "moments",
+     selection = list(
+         search = "exhaustive",
+         nvmax = 7,
+         nbest = 1,
+         criterion = "adjr2"
+     )
+ )
```

```
> summary(m.mtcars.subsets)
```

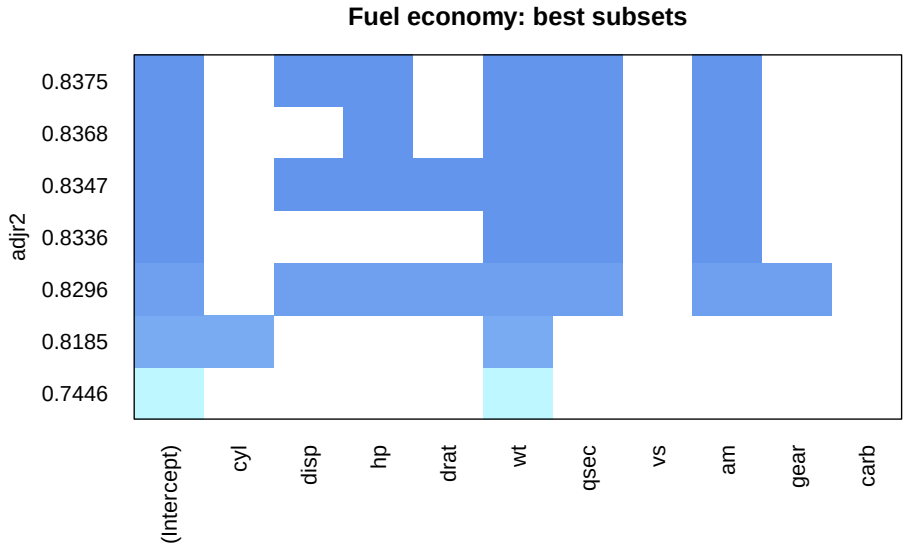
```
Fixed-effect subset comparison
Selection unit: one original data-frame column (complete factors).
Model estimation: moments
Subset comparisons: least squares
Search: exhaustive; observations: 32; retained models: 7
Candidates evaluated: 968; failed fits: 0

Included variables (*):
  ID  Size  Rank  cyl  disp  hp  drat  wt  qsec  vs  am  gear  carb
   1     1     2                       *
```

```
   2     2     3  *                    *
   3     3     4                       *   *         *
   4     4     5              *        *   *         *
   5     5     6        *     *        *   *         *
   6     6     7        *     *   *    *   *         *
   7     7     8        *     *   *    *   *         *   *

Comparison statistics:
  ID Size Rank     RSS    R2 Adj.R2     Cp    BICm     AIC     BIC  logLik
   1    1    2 278.322 0.753  0.745 11.627 -37.795 166.029 170.427 -80.015
   2    2    3 191.172 0.830  0.819  1.219 -46.348 156.010 161.873 -74.005
   3    3    4 169.286 0.850  0.834  0.103 -46.773 154.119 161.448 -72.060
   4    4    5 160.066 0.858  0.837  0.790 -45.099 154.327 163.122 -71.164
   5    5    6 153.438 0.864  0.838  1.846 -42.987 154.974 165.234 -70.487
   6    6    7 150.093 0.867  0.835  3.370 -40.227 156.269 167.995 -70.134
   7    7    8 148.528 0.868  0.830  5.147 -37.096 157.933 171.125 -69.967
Size counts selected data columns; Rank counts estimable fixed
  coefficients, including the intercept.
The full fitted model is unchanged. The chosen formula defines a subsequent
  st.lmm call.
```

```
> blues <- c(
+     "#bef7ff", "#b4ecfd", "#aae1fb", "#a0d6f9", "#96cbf7",
+     "#8cc1f5", "#82b6f3", "#78abf1", "#6ea0ef", "#6495ed"
+ )
> blues <- blues[length(blues):1]
> par(mar = c(6, 5.5, 3, 1), mgp = c(4, 1, 0))
> plot(m.mtcars.subsets, scale = "adjr2", col = blues,
+      main = "Fuel economy: best subsets")
```



Filled cells identify included predictors. Rows are ordered by adjusted R^2, with the best retained model at the top.

Refit with the chosen predictors

```
> m.mtcars <- st.lmm(
+     mpg ~ disp + hp + wt + qsec + am,
+     data = mtcars,
+     method = "moments"
+ )
```

```
> m.mtcars
```

```
Linear model fitted by the method of moments

Model:
  Data:      mtcars
  Fixed:     mpg ~ disp + hp + wt + qsec + am

ANOVA (method of moments, exact F):
  Source     DF1  DF2        SS       MS        F       p
  disp         1   26    6.6287   6.6287   1.1232  0.2990
  hp           1   26   12.5721  12.5721   2.1303  0.1564
  wt           1   26   69.0430  69.0430  11.6993  0.0021
  qsec         1   26   26.4698  26.4698   4.4853  0.0439
  am           1   26   32.1975  32.1975   5.4559  0.0275
  Residual    26       153.4378   5.9015

Fixed-effect regression estimates (exact t):
  Coefficient  Estimate      SE  DF        t       p
  (Intercept)   14.3619  9.7408  26   1.4744  0.1524
  disp           0.0112  0.0106  26   1.0598  0.2990
  hp            -0.0212  0.0145  26  -1.4596  0.1564
  wt            -4.0843  1.1941  26  -3.4204  0.0021
  qsec           1.0069  0.4754  26   2.1179  0.0439
  am             3.4705  1.4858  26   2.3358  0.0275

Model statistics:
  Trait        N  Residual DF      R2  Adjusted R2  Residual SD
  mpg    32.0000           26  0.8637       0.8375       2.4293
```

7.1.5 Cell means and a linear hypothesis: Tomatoes

CM13: RLab-BioStat/v19-u21.R

```
> data("v-lmm-tomato", package = "SelectionTools")
> m.tomato <- st.lmm(
+     y ~ 0 + x.v,
+     data = tomato,
+     method = "moments"
+ )
> m.tomato
```

```
Linear model fitted by the method of moments

Model:
  Data:      tomato
  Fixed:     y ~ 0 + x.v

ANOVA (method of moments, exact F):
  Source     DF1  DF2          SS          MS      F        p
  x.v          3    4  43997.0000  14665.6667  >1000  <0.0001
  Residual     4          58.0000     14.5000

Model statistics:
  Trait        N  Residual DF      R2  Adjusted R2  Residual SD
  y       7.0000            4  0.9987       0.9977       3.8079
```

```
> L.tomato <- matrix(
+         c(1, -1, 0), nrow = 1,
```

```
+         dimnames = list("group 1 - group 2", names(coef(m.tomato)))
+     )
> st.lsm(
+     m.tomato,
+     L = L.tomato,
+     joint = TRUE
+ )
```

```
Exact linear hypothesis
  DF1  DF2       F       p  F.scaling  method
    1    4  2.0690  0.2237     1.0000  exact
```

```
> st.lsm(m.tomato, ~x.v)
```

```
Least-squares means (exact)
  x.v  lsmeans      SE  DF  lower.CL  upper.CL
  1    73.0000  2.1985   4   66.8960   79.1040
  2    78.0000  2.6926   4   70.5242   85.4758
  3    89.0000  2.6926   4   81.5242   96.4758
Confidence level: 0.9500
```

```
> # Illustrative null: each difference from group 1 equals five response units.
> C.tomato <- rbind(
+     group2_group1 = c(-1, 1, 0),
+     group3_group1 = c(-1, 0, 1)
+ )
> st.lsm(
+     m.tomato,
+     ~x.v,
+     method = C.tomato,
+     rhs = c(5, 5),
+     adjust = "none"
+ )
```

```
Least-squares mean contrasts (exact)
  contrast        null estimate     SE DF lower.CL upper.CL      t      p
  group2_group1 5.0000   5.0000 3.4761  4  -4.6512  14.6512 0.0000 1.0000
  group3_group1 5.0000  16.0000 3.4761  4   6.3488  25.6512 3.1645 0.0340
p-value adjustment: none; confidence level: 0.9500
```

```
> # L columns refer to fitted coefficients; method columns refer to means.
> # This no-intercept cell-means model makes the two matrices identical.
> L.tomato <- C.tomato
> colnames(L.tomato) <- names(coef(m.tomato))
> h.tomato.L <- st.lsm(
+     m.tomato,
+     L = L.tomato,
+     rhs = c(5, 5),
+     adjust = "none"
+ )
> h.tomato.L
```

```
Fixed-effect linear hypotheses (exact)
  hypothesis      null estimate     SE DF lower.CL upper.CL      t      p
  group2_group1 5.0000   5.0000 3.4761  4  -4.6512  14.6512 0.0000 1.0000
  group3_group1 5.0000  16.0000 3.4761  4   6.3488  25.6512 3.1645 0.0340
p-value adjustment: none; confidence level: 0.9500
```

```
> # Jointly test both five-unit null differences.
> st.lsm(
+     h.tomato.L,
+     joint = TRUE
+ )
```

```
Exact linear hypothesis
  DF1  DF2       F       p  F.scaling  method
    2    4  5.9606  0.0631     1.0000  exact
```

7.2 Random regression: Chicken growth

M52: Additional/v14-u51-21.R

```
> data("v-lmm-chickens", package = "SelectionTools")
> chickens <- subset(chickens, Time == 21)
> m <- st.lmm(
+     fixed = weight ~ Diet,
+     data = chickens
+ )
> m
```

```
Linear model fitted by REML

Model:
  Data:      chickens
  Fixed:     weight ~ Diet
  Random:    None
  Residual:  ~units

Convergence:
  converged: projected score tolerance met

Fixed effects (Satterthwaite):
  Term  DF1  DF2       F       p
  Diet    3   41  4.6547  0.0069

Variance components:
  Component   Variance
  Residual   4093.6444

Fit and model statistics:
  Observations                 45    R-squared (conditional)     0.2409
  Log-likelihood         -233.4664    RMSE (conditional)         61.0718
  AIC                     476.9328    Variance: fixed          1299.1869
  BIC                     485.9661    Variance: random            0.0000
  R-squared (marginal)      0.2409    Variance: residual       4093.6444
```

```
> st.lsm(
+     m,
+     "Diet",
+     method = "pairwise",
+     side = "two.sided",
+     adjust = "tukey"
+ )
```

```
Least-squares mean contrasts for Diet (Satterthwaite)
  contrast  estimate       SE  DF   lower.CL  upper.CL        t       p
  d1 - d2   -36.9500  25.7918  41  -106.0106   32.1106  -1.4326  0.4868
  d1 - d3   -92.5500  25.7918  41  -161.6106  -23.4894  -3.5883  0.0047
  d1 - d4   -60.8056  26.6590  41  -132.1882   10.5771  -2.2809  0.1193
  d2 - d3   -55.6000  28.6134  41  -132.2159   21.0159  -1.9431  0.2264
  d2 - d4   -23.8556  29.3975  41  -102.5709   54.8598  -0.8115  0.8487
  d3 - d4    31.7444  29.3975  41   -46.9709  110.4598   1.0798  0.7036
p-value and interval adjustment: tukey; confidence level: 0.9500
```

7.2.1 Square-root growth: residual variance only

M53: Additional/v14-u51-21.R

```
> data("v-lmm-chickens", package = "SelectionTools")
> chickens$weight.s <- sqrt(chickens$weight)
> st.lmm(
+     fixed = weight.s ~ Time:Diet,
+     data = chickens
+ )
```

```
Linear model fitted by REML

Model:
  Data:      chickens
  Fixed:     weight.s ~ Time:Diet
  Random:    None
  Residual:  ~units

Convergence:
  converged: projected score tolerance met

Fixed effects (Satterthwaite):
  Term       DF1  DF2         F        p
  Time:Diet    4  573  690.1918  <0.0001

Fixed-effect regression estimates (Satterthwaite):
  Coefficient  Estimate      SE   DF        t        p
  (Intercept)    6.2885  0.1007  573  62.4565  <0.0001
  Time:Dietd1    0.3342  0.0097  573  34.3027  <0.0001
  Time:Dietd2    0.3965  0.0113  573  34.9971  <0.0001
  Time:Dietd3    0.4757  0.0113  573  41.9855  <0.0001
  Time:Dietd4    0.4531  0.0115  573  39.2714  <0.0001
  Factor coding for Diet: level indicators

Variance components:
  Component  Variance
  Residual     1.6646

Fit and model statistics:
  Observations                578    R-squared (conditional)  0.8271
  Log-likelihood        -981.6347    RMSE (conditional)       1.2846
  AIC                  1975.2693    Variance: fixed          7.9645
  BIC                  2001.4268    Variance: random         0.0000
  R-squared (marginal)    0.8271    Variance: residual       1.6646
```

7.2.2 Common intercept and random chick intercept

M54: Additional/v14-u51-21.R

```
> data("v-lmm-chickens", package = "SelectionTools")
> chickens$weight.s <- sqrt(chickens$weight)
> st.lmm(
+     fixed = weight.s ~ Diet:Time,
+     random = ~Chick,
+     data = chickens
+ )
```

```
Linear mixed model fitted by REML

Model:
  Data:      chickens
  Fixed:     weight.s ~ Diet:Time
  Random:    ~Chick
  Residual:  ~units

Convergence:
  converged: projected score tolerance met

Fixed effects (Satterthwaite):
  Term       DF1     DF2      F        p
  Diet:Time    4  546.82  >1000  <0.0001

Random effects (likelihood-ratio tests):
  Term   DF        LR        p
  Chick   1  277.0146  <0.0001

Fixed-effect regression estimates (Satterthwaite):
  Coefficient  Estimate      SE      DF        t        p
  (Intercept)    6.2998  0.1491   69.11  42.2531  <0.0001
  Dietd1:Time    0.3287  0.0088  572.02  37.1712  <0.0001
  Dietd2:Time    0.3936  0.0115  560.88  34.3612  <0.0001
  Dietd3:Time    0.4873  0.0115  560.88  42.5381  <0.0001
  Dietd4:Time    0.4388  0.0117  560.36  37.5505  <0.0001
  Factor coding for Diet: level indicators

Variance components:
  Component  Variance
  Chick        0.8570
  Residual     0.8307

Fit and model statistics:
  Observations                578    R-squared (conditional)  0.9136
  Log-likelihood        -843.1274    RMSE (conditional)       0.8715
  AIC                  1700.2547    Variance: fixed          7.9227
  BIC                  1730.7718    Variance: random         0.8570
  R-squared (marginal)    0.8244    Variance: residual       0.8307
```

7.2.3 Diet intercepts and random chick intercept

M55: Additional/v14-u51-21.R

```
> data("v-lmm-chickens", package = "SelectionTools")
> chickens$weight.s <- sqrt(chickens$weight)
> st.lmm(
+     fixed = weight.s ~ 0 + Diet + Diet:Time,
+     random = ~Chick,
+     data = chickens
+ )
```

```
Linear mixed model fitted by REML

Model:
  Data:      chickens
  Fixed:     weight.s ~ 0 + Diet + Diet:Time
  Random:    ~Chick
  Residual:  ~units

Convergence:
  converged: projected score tolerance met

Fixed effects (Satterthwaite):
  Term       DF1     DF2         F        p
  Diet         4   64.52  433.7926  <0.0001
  Diet:Time    4  526.29     >1000  <0.0001

Random effects (likelihood-ratio tests):
  Term   DF        LR        p
  Chick   1  278.1688  <0.0001

Fixed-effect regression estimates (Satterthwaite):
  Coefficient  Estimate      SE      DF        t        p
  Dietd1:Time    0.3290  0.0093  531.64  35.3909  <0.0001
  Dietd2:Time    0.3930  0.0123  524.35  31.9927  <0.0001
  Dietd3:Time    0.4909  0.0123  524.35  39.9608  <0.0001
  Dietd4:Time    0.4351  0.0125  524.90  34.7140  <0.0001
  Factor coding for Diet: level indicators
  Categorical coefficients omitted; st.lsm reports least-squares means.

Variance components:
  Component  Variance
  Chick        0.8897
  Residual     0.8312

Fit and model statistics:
  Observations               578    R-squared (conditional)  0.9136
  Log-likelihood        -842.6593    RMSE (conditional)       0.8711
  AIC                  1705.3187    Variance: fixed          7.9033
  BIC                  1748.9144    Variance: random         0.8897
  R-squared (marginal)    0.8212    Variance: residual       0.8312
```

7.2.4 Common intercept and random intercept/slope

M56: Additional/v14-u51-21.R

```
> data("v-lmm-chickens", package = "SelectionTools")
> chickens$weight.s <- sqrt(chickens$weight)
> st.lmm(
+     fixed = weight.s ~ Diet:Time,
+     random = ~Time | Chick,
+     data = chickens
+ )
```

```
Linear mixed model fitted by REML

Model:
  Data:      chickens
  Fixed:     weight.s ~ Diet:Time
  Random:    ~Time | Chick
  Residual:  ~units

Convergence:
```

```
  converged: projected score tolerance met

Fixed effects (Satterthwaite):
  Term       DF1    DF2         F        p
  Diet:Time    4  46.09  123.5094  <0.0001

Random effects (likelihood-ratio tests):
  Term   DF        LR        p
  Chick   3  997.4029  <0.0001

Fixed-effect regression estimates (Satterthwaite):
  Coefficient  Estimate      SE     DF        t        p
  (Intercept)    6.3415  0.0680  49.03  93.2119  <0.0001
  Dietd1:Time    0.3177  0.0233  62.72  13.6617  <0.0001
  Dietd2:Time    0.3930  0.0293  61.43  13.4212  <0.0001
  Dietd3:Time    0.4338  0.0293  61.43  14.8127  <0.0001
  Dietd4:Time    0.4764  0.0293  61.42  16.2711  <0.0001
  Factor coding for Diet: level indicators

Covariance components:
  Chick:
              (Intercept)     Time
  (Intercept)      0.1732  -0.0426
  Time            -0.0426   0.0157

  Diagonal entries are variances; off-diagonal entries are covariances.

Variance components:
  Component  Variance
  Residual     0.1868

Fit and model statistics:
  Observations                578    R-squared (conditional)  0.9806
  Log-likelihood        -482.9332    RMSE (conditional)       0.4015
  AIC                    983.8665    Variance: fixed          7.6373
  BIC                   1023.1026    Variance: random         1.7794
  R-squared (marginal)     0.7953    Variance: residual       0.1868
```

7.2.5 Diet intercepts and random intercept/slope

M57: Additional/v14-u51-21.R

```
> data("v-lmm-chickens", package = "SelectionTools")
> chickens$weight.s <- sqrt(chickens$weight)
> st.lmm(
+     fixed = weight.s ~ 0 + Diet + Diet:Time,
+     random = ~Time | Chick,
+     data = chickens
+ )
```

```
Linear mixed model fitted by REML

Model:
  Data:      chickens
  Fixed:     weight.s ~ 0 + Diet + Diet:Time
  Random:    ~Time | Chick
  Residual:  ~units

Convergence:
  converged: projected score tolerance met

Fixed effects (Satterthwaite):
  Term       DF1    DF2         F        p
  Diet         4  45.74     >1000  <0.0001
```

```
  Diet:Time    4  45.23  123.9744  <0.0001

Random effects (likelihood-ratio tests):
  Term   DF        LR        p
  Chick   3  998.5883  <0.0001

Fixed-effect regression estimates (Satterthwaite):
  Coefficient  Estimate      SE     DF        t        p
  Dietd1:Time    0.3116  0.0284  46.45  10.9811  <0.0001
  Dietd2:Time    0.3930  0.0393  44.80   9.9888  <0.0001
  Dietd3:Time    0.4909  0.0393  44.80  12.4766  <0.0001
  Dietd4:Time    0.4310  0.0394  44.92  10.9485  <0.0001
  Factor coding for Diet: level indicators
  Categorical coefficients omitted; st.lsm reports least-squares means.

Covariance components:
  Chick:
              (Intercept)     Time
  (Intercept)      0.1592  -0.0398
  Time            -0.0398   0.0151

  Diagonal entries are variances; off-diagonal entries are covariances.

Variance components:
  Component  Variance
  Residual     0.1868

Fit and model statistics:
  Observations               578    R-squared (conditional)  0.9807
  Log-likelihood        -482.4496    RMSE (conditional)       0.4015
  AIC                    988.8991    Variance: fixed          7.7321
  BIC                   1041.2140    Variance: random         1.7358
  R-squared (marginal)     0.8009    Variance: residual       0.1868
```

7.2.6 Random regression with time-specific residual variances

M58: Additional/v14-u51-21.R

```
> data("v-lmm-chickens", package = "SelectionTools")
> chickens$weight.s <- sqrt(chickens$weight)
> st.lmm(
+     fixed = weight.s ~ Diet:Time,
+     random = ~Time | Chick,
+     cov = ~at(Time):units,
+     data = chickens
+ )
```

```
Linear mixed model fitted by REML

Model:
  Data:      chickens
  Fixed:     weight.s ~ Diet:Time
  Random:    ~Time | Chick
  Residual:  ~at(Time):units

Convergence:
  converged: projected score tolerance met

Fixed effects (Satterthwaite):
  Term       DF1    DF2         F        p
  Diet:Time    4  31.62  125.1205  <0.0001

Random effects (likelihood-ratio tests):
  Term   DF        LR        p
```

```
  Chick   3  529.7413  <0.0001

Fixed-effect regression estimates (Satterthwaite):
  Coefficient  Estimate      SE     DF        t        p
  (Intercept)    6.1846  0.0666  22.13  92.8208  <0.0001
  Dietd1:Time    0.3373  0.0220  65.62  15.3476  <0.0001
  Dietd2:Time    0.4069  0.0258  70.02  15.7701  <0.0001
  Dietd3:Time    0.4434  0.0258  70.02  17.1867  <0.0001
  Dietd4:Time    0.4873  0.0258  70.02  18.8882  <0.0001
  Factor coding for Diet: level indicators

Covariance components:
  Chick:
              (Intercept)     Time
  (Intercept)      0.2004  -0.0534
  Time            -0.0534   0.0174

  Diagonal entries are variances; off-diagonal entries are covariances.

Variance components:
  Component                      Variance
  Residual: at(Time):units[0]      0.2297
  Residual: at(Time):units[2]      0.0439
  Residual: at(Time):units[4]      0.0216
  Residual: at(Time):units[6]      0.0743
  Residual: at(Time):units[8]      0.1423
  Residual: at(Time):units[10]     0.1896
  Residual: at(Time):units[12]     0.2381
  Residual: at(Time):units[14]     0.0915
  Residual: at(Time):units[16]     0.0490
  Residual: at(Time):units[18]     0.1938
  Residual: at(Time):units[20]     0.5817
  Residual: at(Time):units[21]     1.0066

Fit and model statistics:
  Observations                578    R-squared (conditional)  0.9772
  Log-likelihood         -410.4204    RMSE (conditional)       0.4626
  AIC                     860.8408    Variance: fixed          8.0860
  BIC                     948.0323    Variance: random         1.8530
  R-squared (marginal)      0.7950    Variance: residual       0.2321
```

7.2.7 Random slope with time-specific residual variances

M59: Additional/v14-u51-21.R

```
> data("v-lmm-chickens", package = "SelectionTools")
> chickens$weight.s <- sqrt(chickens$weight)
> m <- st.lmm(
+     fixed = weight.s ~ Diet:Time,
+     random = ~0 + Time | Chick,
+     cov = ~at(Time):units,
+     data = chickens
+ )
> m
```

```
Linear mixed model fitted by REML

Model:
  Data:      chickens
  Fixed:     weight.s ~ Diet:Time
  Random:    ~0 + Time | Chick
  Residual:  ~at(Time):units

Convergence:
```

```
  converged: projected score tolerance met

Fixed effects (Satterthwaite):
  Term       DF1    DF2         F        p
  Diet:Time    4  41.52  161.2052  <0.0001

Random effects (likelihood-ratio tests):
  Term   DF        LR        p
  Chick   1  502.1194  <0.0001

Fixed-effect regression estimates (Satterthwaite):
  Coefficient  Estimate      SE     DF         t        p
  (Intercept)    6.3915  0.0121  56.62  528.2660  <0.0001
  Dietd1:Time    0.3075  0.0246  42.42   12.5180  <0.0001
  Dietd2:Time    0.3902  0.0341  41.23   11.4506  <0.0001
  Dietd3:Time    0.4597  0.0341  41.23   13.4897  <0.0001
  Dietd4:Time    0.4532  0.0341  41.24   13.2976  <0.0001
  Factor coding for Diet: level indicators

Variance components:
  Component                      Variance
  Chick                            0.0115
  Residual: at(Time):units[0]      0.0081
  Residual: at(Time):units[2]      0.1065
  Residual: at(Time):units[4]      0.1808
  Residual: at(Time):units[6]      0.2252
  Residual: at(Time):units[8]      0.1898
  Residual: at(Time):units[10]     0.1582
  Residual: at(Time):units[12]     0.1502
  Residual: at(Time):units[14]     0.0395
  Residual: at(Time):units[16]     0.1386
  Residual: at(Time):units[18]     0.4488
  Residual: at(Time):units[20]     0.9327
  Residual: at(Time):units[21]     1.4106

Fit and model statistics:
  Observations                578    R-squared (conditional)  0.9671
  Log-likelihood         -424.2314    RMSE (conditional)       0.5533
  AIC                     884.4627    Variance: fixed          7.5970
  BIC                     962.9351    Variance: random         1.8469
  R-squared (marginal)      0.7779    Variance: residual       0.3217
```

```
> st.lsm(
+     m,
+     "Diet",
+     trend = "Time",
+     method = "trt.vs.ctrl",
+     ref = "d1",
+     side = ">",
+     adjust = "fdr"
+ )
```

```
Least-squares trend contrasts (Satterthwaite)
Trend: Time
  contrast  estimate      SE     DF  lower.CL  upper.CL       t       p
  d2 - d1     0.0828  0.0420  41.55   -0.0097       Inf  1.9711  0.0277
  d3 - d1     0.1522  0.0420  41.55    0.0598       Inf  3.6262  0.0009
  d4 - d1     0.1457  0.0420  41.55    0.0533       Inf  3.4706  0.0009
p-value adjustment: fdr; confidence level: 0.9500
Alternative: estimate > null
Confidence intervals use bonferroni adjustment.
```

```
> # Illustrative minimum improvement: 0.05 sqrt(weight) units per day.
> # The object retains the null, direction, adjustment and confidence level.
> chicken.minimum <- st.lsm(
+     m,
+     "Diet",
+     trend = "Time",
+     method = "trt.vs.ctrl",
+     ref = "d1",
+     rhs = .05,
+     side = ">",
+     adjust = "single-step",
+     level = .90
+ )
> chicken.minimum
```

```
Least-squares trend contrasts (Satterthwaite)
Trend: Time
  contrast   null estimate     SE    DF lower.CL upper.CL      t      p
  d2 - d1  0.0500   0.0828 0.0420 41.55   0.0067      Inf 0.7801 0.4530
  d3 - d1  0.0500   0.1522 0.0420 41.55   0.0762      Inf 2.4353 0.0267
  d4 - d1  0.0500   0.1457 0.0420 41.55   0.0696      Inf 2.2797 0.0381
p-value and interval adjustment: single-step; confidence level: 0.9000
Alternative: estimate > null
Simultaneous inference uses one common reference DF per family (the minimum
  marginal DF unless DF is supplied); mixed-model results are approximate.
```

```
> # Null values retained for the three slope comparisons.
> coef(
+     chicken.minimum,
+     rhs = TRUE
+ )
```

```
  Row   Value
  1    0.0500
  2    0.0500
  3    0.0500
```

7.2.8 Implemented features and limitations

Implemented features.

- Gaussian ML and REML random-coefficient models with random intercepts, slopes and expanded coefficient bases, as in `random=~Time | Chick`. An unmodified coefficient block estimates its full covariance. `diag()` or the double-bar form specifies independent expanded coefficients.
- Univariate coefficient covariance families `us()`, `diag()`, `homdiag()`, compound symmetry and AR1, with the documented homogeneous or heterogeneous scales. Named relationships can connect grouping levels; other supported random and residual terms can accompany the regression.
- Time-specific residual variances, fitted-trajectory predictions, refits, updates and parametric simulation through their supported interfaces. Satterthwaite and asymptotic inference are available when the covariance information is usable; covariance and boundary diagnostics are retained.

Limitations.

- AR1 on random coefficients uses their positions in the expanded design basis. It does not automatically model elapsed-time distances or within-subject residual correlation. Factor coding, intercept inclusion and coefficient order therefore define the covariance model; residual correlation requires its own `cov` specification.
- The structured coefficient modifiers described here apply to univariate likelihood fits, not moment estimation. Multivariate random regressions use trait-by-coefficient covariance and do not inherit every univariate modifier.
- Kenward–Roger adjustment is available only for supported linear covariance families. Heterogeneous compound symmetry and AR1 with more than two coefficients reject it; Satterthwaite and asymptotic inference are supported alternatives. A zero or singular covariance block does not certify ordinary regular inference, and local boundary checks do not prove a global optimum.
- The genetic-level `st.predict` interface does not select random-slope blocks; conditional `predict` returns trajectories. Whole-term random-effect tests delete the coefficient block together, rather than performing automatic stepwise slope selection. Extrapolation beyond observed times remains a model assumption, not an automatically validated forecast.

8 Diallel analysis

8.1 Half-diallel example

This example fits a half-diallel model to maize sugar measurements. Shared parental effects describe general combining ability (GCA), while unordered cross effects describe specific combining ability (SCA).

The maize dataset contains sugar measurements for 21 crosses among seven parents, with two technical repetitions per cross. Its 42 observations retain the source response values and repetition labels. The data are used in the `lme4breeding` quantitative-genetics vignette and distributed with `enhancer` (Covarrubias-Pazaran, 2016).

The packaged data already contain parental factors with the same seven levels P1 through P7 and an unordered cross factor labelled P1:P2, for example, for SCA. The redundant source identifier `geno` is omitted; `cross` identifies the cross and supplies its SCA term. The following call loads these data directly.

```
> data("v-lmm-halfdiallel", package = "SelectionTools")
> head(halfdiallel, 5)
```

```
  rep male female cross     sugar
1   1   P1     P2 P1:P2 13.950509
2   2   P1     P2 P1:P2  9.756918
3   1   P1     P3 P1:P3 13.906355
4   2   P1     P3 P1:P3  9.119455
5   1   P1     P4 P1:P4  5.174483
```

The `overlay(female, male)` term uses one shared parental effect vector and one GCA variance. Each cross receives the sum of its two parental effects; a self cross would have loading two. The separate cross term models SCA, following Sprague and Tatum (1942). The fixed model contains only an intercept, matching the `lme4breeding` half-diallel example. The technical-repetition labels are retained in the data but are not fitted as a blocking factor.

The fitted model's variance-component table reports the shared parental GCA, cross SCA and residual variances directly.

```
> m.diallel <- st.lmm(
+     fixed = sugar ~ 1,
+     random = ~overlay(female, male) + cross,
+     data = halfdiallel
```

```
+ )
> m.diallel
```

```
Linear mixed model fitted by REML

Model:
  Data:      halfdiallel
  Fixed:     sugar ~ 1
  Random:    ~overlay(female, male) + cross
  Residual:  ~units

Convergence:
  converged: projected score tolerance met

Random effects (likelihood-ratio tests):
  Term                    DF       LR       p
  overlay(female, male)    1  11.4690  0.0007
  cross                    1   2.5752  0.1086

Variance components:
  Component              Variance
  overlay(female, male)    5.5088
  cross                    1.8160
  Residual                 3.1174

Fit and model statistics:
  Observations                42    R-squared (conditional)    0.8046
  Log-likelihood         -97.7232    RMSE (conditional)         1.4356
  AIC                    203.4464    Variance: fixed            0.0000
  BIC                    210.3971    Variance: random          12.8335
  R-squared (marginal)     0.0000    Variance: residual         3.1174
```

These estimates show close numerical agreement with the published `lme4breeding` example: GCA 5.5088, SCA 1.8160 and residual variance 3.1173. This is a comparison with its displayed output, rather than a fresh run of that package. A known parental relationship can be supplied with `ucov=list(overlay=A)`, with both axes named by parent identifier. Separate female and male random terms would not impose the shared-parent model. Reciprocal and maternal effects require separate specification. The current overlay is univariate and stores a dense parent basis.

Predict each parent's GCA effect without the intercept, with prediction-error standard errors and pointwise intervals.

```
> diallel.parents <- st.predict(
+     m.diallel,
+     "0 + overlay(female, male)"
+ )
> diallel.parents
```

```
Shared parental effects from overlay(female, male)
Predicted random effects
  Parent  prediction      SE   DF  lower.CL  upper.CL
  P2          0.9190  1.1412  Inf   -1.3177    3.1557
  P3          2.8773  1.1412  Inf    0.6406    5.1140
  P4         -3.8482  1.1412  Inf   -6.0849   -1.6115
  P5          0.6570  1.1412  Inf   -1.5797    2.8937
  P6          1.6020  1.1412  Inf   -0.6347    3.8387
  P7         -1.3027  1.1412  Inf   -3.5394    0.9340
  P1         -0.9043  1.1412  Inf   -3.1410    1.3324
Confidence level: 0.9500
SE and intervals use prediction-error variance; fitted variances held
  fixed.
```

The `0 +` prefix requests GCA effects alone. Omitting it returns the fixed intercept plus GCA. Their joint prediction-error covariance is returned by `vcov(diallel.parents)`.

Predict SCA effects alone, then complete cross means. The latter combine the intercept, both shared parental effects and SCA. Each distinct target is reported once; standard errors include their joint covariance.

```
> st.predict(m.diallel, "0 + cross")
```

```
Predicted random effects
  0 + cross  prediction      SE   DF  lower.CL  upper.CL
  P1:P2          0.8103  1.0649  Inf   -1.2768    2.8975
  P1:P3         -0.4269  1.0649  Inf   -2.5140    1.6603
  P1:P4          0.6634  1.0649  Inf   -1.4238    2.7505
  P1:P5         -1.0038  1.0649  Inf   -3.0910    1.0834
  P1:P6         -0.0611  1.0649  Inf   -2.1483    2.0261
  P1:P7         -0.2800  1.0649  Inf   -2.3672    1.8071
  P2:P3          0.9789  1.0649  Inf   -1.1083    3.0660
  P2:P4         -0.8414  1.0649  Inf   -2.9285    1.2458
  P2:P5         -0.4064  1.0649  Inf   -2.4936    1.6808
  P2:P6          1.1915  1.0649  Inf   -0.8957    3.2786
  [6 table rows omitted]
  P4:P6         -1.3922  1.0649  Inf   -3.4794    0.6950
  P4:P7         -0.4820  1.0649  Inf   -2.5692    1.6052
  P5:P6          1.3386  1.0649  Inf   -0.7486    3.4258
  P5:P7          0.3350  1.0649  Inf   -1.7521    2.4222
  P6:P7          0.4268  1.0649  Inf   -1.6604    2.5140
Confidence level: 0.9500
SE and intervals use prediction-error variance; fitted variances held
  fixed.
```

```
> st.predict(m.diallel, "male + female + cross")
```

```
Predicted means: fixed intercept + random effects
  male + female + cross  prediction      SE   DF  lower.CL  upper.CL
  P1 | P2 | P1:P2           11.1582  1.0278  Inf    9.1438   13.1726
  P1 | P3 | P1:P3           11.8793  1.0278  Inf    9.8649   13.8936
  P1 | P4 | P1:P4            6.2440  1.0278  Inf    4.2296    8.2583
  P1 | P5 | P1:P5            9.0820  1.0278  Inf    7.0676   11.0964
  P1 | P6 | P1:P6           10.9697  1.0278  Inf    8.9554   12.9841
  P1 | P7 | P1:P7            7.8461  1.0278  Inf    5.8317    9.8605
  P2 | P3 | P2:P3           15.1083  1.0278  Inf   13.0940   17.1227
  P2 | P4 | P2:P4            6.5626  1.0278  Inf    4.5482    8.5770
  P2 | P5 | P2:P5           11.5028  1.0278  Inf    9.4884   13.5171
  P2 | P6 | P2:P6           14.0456  1.0278  Inf   12.0313   16.0600
  [6 table rows omitted]
  P4 | P6 | P4:P6            6.6948  1.0278  Inf    4.6804    8.7091
  P4 | P7 | P4:P7            4.7003  1.0278  Inf    2.6859    6.7146
  P5 | P6 | P5:P6           13.9307  1.0278  Inf   11.9164   15.9451
  P5 | P7 | P5:P7           10.0225  1.0278  Inf    8.0081   12.0368
  P6 | P7 | P6:P7           11.0593  1.0278  Inf    9.0449   13.0736
Confidence level: 0.9500
SE and intervals use prediction-error variance; fitted variances held
  fixed.
```

9 Genomic covariances of random genotype effects in a RCBD

9.1 Tropical maize: BLUE, BLUP and genomic BLUP

9.1.1 BLUE: Fixed genotype effects

```
> data("v-lmm-tropmaize", package = "SelectionTools")
> maize.field <- st.lmm(
+     fixed = y ~ gen,
+     random = ~block,
+     data = tropmaize
+ )
> maize.field
```

```
Linear mixed model fitted by REML

Model:
  Data:      tropmaize
  Fixed:     y ~ gen
  Random:    ~block
  Residual:  ~units

Convergence:
  converged: projected score tolerance met

Fixed effects (exact F):
  Term  DF1  DF2        F        p
  gen   263  526  91.7330  <0.0001

Random effects (exact F-tests):
  Term   DF1  DF2        F        p
  block    2  526  87.8085  <0.0001

Variance components:
  Component  Variance
  block         0.0108
  Residual      0.0327

Fit and model statistics:
  Observations                792    R-squared (conditional)  0.9686
  Log-likelihood           4.2712    RMSE (conditional)       0.1474
  AIC                    523.4575    Variance: fixed          0.9974
  BIC                   1766.8909    Variance: random         0.0108
  R-squared (marginal)     0.9583    Variance: residual       0.0327
```

```
> st.lsm(maize.field, "gen")
```

```
Least-squares means (Satterthwaite)
  gen  lsmeans      SE     DF  lower.CL  upper.CL
  1     2.5160  0.1204  30.87    2.2705    2.7615
  2     1.6410  0.1204  30.87    1.3955    1.8865
  3     0.4160  0.1204  30.87    0.1705    0.6615
  4     1.3390  0.1204  30.87    1.0935    1.5845
  5     1.7300  0.1204  30.87    1.4845    1.9755
  6     2.7040  0.1204  30.87    2.4585    2.9495
  7     1.9720  0.1204  30.87    1.7265    2.2175
  8     1.8440  0.1204  30.87    1.5985    2.0895
  9     2.0820  0.1204  30.87    1.8365    2.3275
  10    2.8980  0.1204  30.87    2.6525    3.1435
  [249 table rows omitted]
  260   1.6460  0.1204  30.87    1.4005    1.8915
  261   0.7340  0.1204  30.87    0.4885    0.9795
  262   2.7490  0.1204  30.87    2.5035    2.9945
  263   1.1020  0.1204  30.87    0.8565    1.3475
  264   0.8740  0.1204  30.87    0.6285    1.1195
Confidence level: 0.9500
LSD=0.2901 HSD=0.7006 (alpha=0.05)
```

9.1.2 BLUP: Independent random genotype effects

```
> maize.random <- st.lmm(
+     fixed = y ~ 1,
+     random = ~gen + block,
+     data = tropmaize
+ )
> maize.random
```

```
Linear mixed model fitted by REML

Model:
  Data:      tropmaize
  Fixed:     y ~ 1
  Random:    ~gen + block
  Residual:  ~units

Convergence:
  converged: projected score tolerance met

Random effects (exact F-tests):
  Term   DF1  DF2        F        p
  gen    263  526  91.7330  <0.0001
  block    2  526  87.8085  <0.0001

Variance components:
  Component  Variance
  gen           0.9891
  block         0.0108
  Residual      0.0327

Fit and model statistics:
  Observations                 792    R-squared (conditional)  0.9683
  Log-likelihood        -371.6918    RMSE (conditional)       0.1478
  AIC                    751.3837    Variance: fixed          0.0000
  BIC                    770.0819    Variance: random         0.9998
  R-squared (marginal)     0.0000    Variance: residual       0.0327
```

```
> maize.blup <- st.predict(maize.random, "gen")
> maize.blup
```

```
Predicted means: fixed intercept + random effect
  gen  prediction      SE   DF  lower.CL  upper.CL
  1        2.5114  0.1199  Inf    2.2765    2.7464
  2        1.6460  0.1199  Inf    1.4111    1.8809
  3        0.4343  0.1199  Inf    0.1994    0.6693
  4        1.3473  0.1199  Inf    1.1123    1.5822
  5        1.7340  0.1199  Inf    1.4991    1.9689
  6        2.6974  0.1199  Inf    2.4625    2.9323
  7        1.9734  0.1199  Inf    1.7384    2.2083
  8        1.8468  0.1199  Inf    1.6118    2.0817
  9        2.0822  0.1199  Inf    1.8472    2.3171
  10       2.8893  0.1199  Inf    2.6544    3.1242
  [249 table rows omitted]
  260      1.6509  0.1199  Inf    1.4160    1.8858
  261      0.7489  0.1199  Inf    0.5139    0.9838
  262      2.7419  0.1199  Inf    2.5070    2.9768
  263      1.1129  0.1199  Inf    0.8779    1.3478
  264      0.8873  0.1199  Inf    0.6524    1.1223
Confidence level: 0.9500
SE and intervals use prediction-error variance; fitted variances held
  fixed.
```

9.1.3 Genomic BLUP: Related random genotype effects

```
> data("v-tropmaize-vcf", package = "SelectionTools")
> st.load.vcf.data(v.tropmaize.vcf, data.set = "maize.lmm")
```

```
M (data set 'maize.lmm'): New data set generated 'maize.lmm'
M (data set 'maize.lmm'): No. of individuals: 264, no. of markers: 1076
```

```
> maize.Z <- gs.get.Z("maize.lmm", names = TRUE)
> maize.A <- gs.make.A(maize.Z)
```

```
> maize.genomic <- st.lmm(
+     fixed = y ~ 1,
+     random = ~gen + block,
+     ucov = list(gen = maize.A),
+     data = tropmaize
+ )
> maize.genomic
```

```
Linear mixed model fitted by REML

Model:
  Data:      tropmaize
  Fixed:     y ~ 1
  Random:    ~gen + block
  Residual:  ~units
  Covariance: gen (264 levels; supplied matrix)

Convergence:
  converged: projected score tolerance met

Random effects (likelihood-ratio tests):
  Term   DF         LR        p
  gen     1  1498.4437  <0.0001
  block   1   138.2560  <0.0001

Variance components:
  Component  Variance
  gen          1.0523
  block        0.0108
  Residual     0.0333

Fit and model statistics:
  Observations                792    R-squared (conditional)  0.9821
  Log-likelihood         -386.0421    RMSE (conditional)       0.1494
  AIC                     780.0842    Variance: fixed          0.0000
  BIC                     798.7825    Variance: random         1.8261
  R-squared (marginal)      0.0000    Variance: residual       0.0333
```

```
> maize.gblup <- st.predict(maize.genomic, "gen")
> maize.gblup
```

```
Predicted means: fixed intercept + random effect
  gen  prediction      SE   DF  lower.CL  upper.CL
  1        2.5782  0.1133  Inf    2.3562    2.8002
  2        1.6500  0.1205  Inf    1.4139    1.8861
  3        0.4070  0.1205  Inf    0.1708    0.6432
  4        1.3300  0.1207  Inf    1.0934    1.5666
```

```
  5        1.7273  0.1205  Inf    1.4911    1.9636
  6        2.7088  0.1203  Inf    2.4731    2.9445
  7        1.9770  0.1207  Inf    1.7404    2.2135
  8        1.8566  0.1205  Inf    1.6205    2.0928
  9        2.0810  0.1204  Inf    1.8449    2.3171
  10       2.8778  0.1204  Inf    2.6417    3.1138
  [249 table rows omitted]
  260      1.6462  0.1203  Inf    1.4104    1.8820
  261      0.7464  0.1207  Inf    0.5098    0.9829
  262      2.7201  0.1204  Inf    2.4840    2.9562
  263      1.1134  0.1204  Inf    0.8774    1.3495
  264      0.8841  0.1207  Inf    0.6476    1.1206
Confidence level: 0.9500
SE and intervals use prediction-error variance; fitted variances held
  fixed.
```

10 Analysis of a series of trials

10.1 Tropical maize: A series across environments

This example analyses tropical-maize lattice trials in stages. Each environment first has its own ordinary mixed model, with fixed genotype effects and random replicates and incomplete blocks within replicates. The series analysis then uses the adjusted genotype means and their full covariance matrices. This follows the fully efficient stagewise approach of Piepho et al. (2012) and Endelman (2023).

The package, plot data and marker data are loaded independently of the other examples. The first three environment labels select three independent trials; all their observed phenotypes are used.

```
> library(SelectionTools)
> st.set.num.threads(1)
> data("v-lmm-tropmaize-lattice", package = "SelectionTools")
> data("v-tropmaize-vcf", package = "SelectionTools")
> st.load.vcf.data(v.tropmaize.vcf, data.set = "maize.series")
> envs <- unique(tropmaize.lattice$env)
> G <- st.get.A("maize.series")
```

```
M (data set 'maize.series'): New data set generated 'maize.series'
M (data set 'maize.series'): No. of individuals: 264, no. of markers: 1076
```

Each environment is fitted separately. Genotypes remain fixed so that these adjusted observations have not already undergone genetic shrinkage.

```
> fit.env1 <- st.lmm(y ~ gen, random = ~ rep + rep:block,
+     data = droplevels(subset(tropmaize.lattice, env == envs[1])))
```

```
> fit.env2 <- st.lmm(y ~ gen, random = ~ rep + rep:block,
+     data = droplevels(subset(tropmaize.lattice, env == envs[2])))
```

```
> fit.env3 <- st.lmm(y ~ gen, random = ~ rep + rep:block,
+     data = droplevels(subset(tropmaize.lattice, env == envs[3])))
```

The next `st.lmm` call receives the fitted models. The first entry of `series` names the environment factor; its labels are recovered from the earlier fits. The common fixed genotype factor identifies the means to extract. The genomic matrix supplies relationships for the random main genotype effect, while `gen:env` specifies independent additional genotype-by-environment variation.

```
> fit.series.G <- st.lmm(y ~ env,
+     random = ~ gen + gen:env,
+     ucov = list(gen = G),
+     series = list("env", fit.env1, fit.env2, fit.env3))
> fit.series.G
```

```
Linear mixed model fitted by REML

Model:
  Data:      series of earlier fits
  Fixed:     y ~ env
  Random:    ~gen + gen:env
  Residual: supplied covariance; no extra units variance
  Covariance: gen (264 levels; supplied matrix)

Convergence:
  converged: projected score tolerance met

Fixed effects (Satterthwaite):
  Term  DF1      DF2        F        p
  env     2  4219.23  13.4339  <0.0001

Random effects (likelihood-ratio tests):
  Term     DF         LR        p
  gen       1  1423.0027  <0.0001
  gen:env   1   901.6339  <0.0001

Variance components:
  Stage-1 estimation covariance: full; scale fixed at 1
  This covariance is retained, not estimated in this fit.
  Component  Variance
  gen          1.0311
  gen:env      0.0287

Fit and model statistics:
  Observations               792    R-squared (conditional)  0.9947
  Log-likelihood        -418.7741    RMSE (conditional)       0.0389
  AIC                    847.5481    Variance: fixed          0.0058
  BIC                    870.9209    Variance: random         1.8074
  R-squared (marginal)     0.0032    Variance: residual       0.0096
```

Stage-one model and adjusted observations. For independent trial e, write the plot-level model as

$$y_e = X_e\beta_e + Z_e b_e + \varepsilon_e, \qquad R_e = Z_e B_e Z_e^{\mathsf{T}} + R_{\varepsilon,e}.$$

Genotypes are fixed in β_e; replicates and incomplete blocks are trial-specific random effects. Each trial estimates its own covariance components, so differences in plot error variance between environments are retained. Conditional on these components, the adjusted genotype means and their estimation-error covariance are

$$\widehat{m}_e = L_e\widehat{\beta}_e, \qquad C_e = L_e(X_e^{\mathsf{T}}R_e^{-1}X_e)^{-}L_e^{\mathsf{T}}.$$

Here L_e defines estimable genotype means, and the superscript $-$ denotes a generalized inverse when needed. The series interface uses the marginal-means reference grid: nuisance factors are averaged equally and numeric covariates use their default reference values. Only genotype means are forwarded; covariate slopes and their joint covariance are not transferred by this interface. First-stage genotype BLUPs are rejected because they have already undergone shrinkage.

Full covariance propagation and stage-two model. Stack the means and their covariances in the same trial and genotype order. Independence of the earlier analyses gives

$$\widehat{m} = (\widehat{m}_1^{\mathsf{T}}, \ldots, \widehat{m}_E^{\mathsf{T}})^{\mathsf{T}}, \qquad C = C_1 \oplus \cdots \oplus C_E.$$

Off-diagonal entries within each C_e are retained. They reflect shared block adjustments and other dependence between the estimated means; independent trials have zero cross-trial blocks. The second-stage model is

$$\widehat{m} = X\alpha + Z_g a + Z_{ge} u + \delta, \qquad a \sim N(0, \sigma_a^2 G), \quad u \sim N(0, \sigma_{ge}^2 I), \quad \delta \sim N(0, C),$$

with mutually independent random contributions. $X\alpha$ contains fixed environment means, a is the genomic genotype main effect, and u is additional genotype-by-environment variation. Thus

$$V = C + \sigma_a^2 Z_g G Z_g^{\mathsf{T}} + \sigma_{ge}^2 Z_{ge} Z_{ge}^{\mathsf{T}}.$$

There is one adjusted observation per retained genotype–environment combination, so $Z_{ge} Z_{ge}^{\mathsf{T}} = I$ in this example. The complete C is supplied on its absolute response scale with coefficient fixed at one. Stage two estimates genetic and G×E components, not a new multiplier for the first-stage error. Explicit `gen:env` therefore suppresses an additional units variance: with one observation per combination it would duplicate the same independent covariance component. The printed inherited covariance describes a fixed contribution, not an estimated variance of one for every mean.

Weighting, estimation and prediction. The precision of the means is used through the full mixed-model covariance V. With $K = \mathrm{Var}[(a^{\mathsf{T}}, u^{\mathsf{T}})^{\mathsf{T}}]$ and $Z = (Z_g, Z_{ge})$,

$$\widehat{\alpha} = (X^{\mathsf{T}} V^{-1} X)^{-1} X^{\mathsf{T}} V^{-1} \widehat{m}, \qquad \widehat{b} = K Z^{\mathsf{T}} P \widehat{m},$$

$$P = V^{-1} - V^{-1} X (X^{\mathsf{T}} V^{-1} X)^{-1} X^{\mathsf{T}} V^{-1}, \qquad \mathrm{PEV}(b) = K - K Z^{\mathsf{T}} P Z K.$$

These expressions assume an estimable full-rank fixed parameterization; the implementation handles aliases through its estimability machinery. No separate location weights are needed. In a simple trial with independent replicated observations, a mean has error variance σ_e^2 / r_{ie}, not σ_e^2. In an incomplete-block analysis, C_e additionally carries unequal precisions and correlations. Replacing C_e by its diagonal is an approximation; replacing all blocks by an average error variance discards still more information. The across-environment prediction averages fixed environment means equally and adds the genotype BLUP. This defines the prediction target; equal target weights do not imply equal precision in fitting. Prediction-error SEs include uncertainty in that complete target, including fixed–random covariance. Genetic reliability concerns only a and, conditional on covariance components, is $1 - \mathrm{PEV}(a_i) / \mathrm{Var}(a_i)$.

10.2 Barley: Stagewise analysis across years and locations

This independent example uses thousand-kernel weight (`TKW`) in the packaged barley multi-environment trial: 15 genotypes, five locations, three years and four replicates per location–year trial. Locations recur across years, so persistent genotype-by-location effects can be separated from genotype-by-year and genotype-by-location-by-year effects. No genomic relationship matrix is supplied for these data; genotype effects are independent. The preceding covariance-propagation theory applies unchanged.

Give each location–year trial a unique environment label.

```
> library(SelectionTools)
> st.set.num.threads(1)
> data("v-lmm-barley-met", package = "SelectionTools")
> barley.met$env <- interaction(barley.met$loc, barley.met$year,
+     drop = TRUE)
```

The 15 trials are fitted separately with fixed genotypes and random replicates. The list retains its environment names. Each trial's data also retains `loc` and factor `year`; they are constant within that trial and are carried into the second stage. Genotype and replicate levels unused in an individual trial are dropped.

```
> barley.trials <- split(barley.met, barley.met$env)
> barley.first <- lapply(barley.trials, function(trial) {
+     st.lmm(TKW ~ gen, random = ~ rep,
+         data = droplevels(trial))
+ })
```

The ordinary fitted models enter the combined analysis directly. Fixed `env` gives each location–year combination its own mean. The four random terms represent an overall genotype effect, persistent genotype-by-location effects, genotype-by-year effects shared across locations, and transient genotype-by-location-by-year effects, respectively.

```
> barley.series <- st.lmm(TKW ~ env,
+     random = ~ gen + gen:loc + gen:year + gen:env,
+     series = c(list("env"), barley.first))
> barley.series
```

```
Linear mixed model fitted by REML

Model:
  Data:      series of earlier fits
  Fixed:     TKW ~ env
  Random:    ~gen + gen:loc + gen:year + gen:env
  Residual: supplied covariance; no extra units variance

Convergence:
  converged: projected score tolerance met

Fixed effects (Satterthwaite):
  Term  DF1     DF2        F        p
  env    14  101.72  26.8215  <0.0001

Random effects (likelihood-ratio tests):
  Term      DF        LR        p
  gen        1    7.4767   0.0063
  gen:loc    1   28.7302  <0.0001
  gen:year   1   96.6672  <0.0001
  gen:env    1  118.2846  <0.0001

Variance components:
  Stage-1 estimation covariance: full; scale fixed at 1
  This covariance is retained, not estimated in this fit.
  Component  Variance
  gen         12.1324
  gen:loc      4.3073
  gen:year    13.1456
  gen:env      3.8547

Fit and model statistics:
  Observations               225    R-squared (conditional)   0.9668
  Log-likelihood        -597.7058    RMSE (conditional)        0.6634
  AIC                  1233.4116    Variance: fixed          31.0305
  BIC                  1298.3175    Variance: random         33.4400
  R-squared (marginal)    0.4653    Variance: residual        2.2124
```

Predict the overall genotype values averaged equally over the 15 trial means. These are predictions for the observed location–year series; they are not predictions for a future year. The SE includes the fixed mean and genetic prediction-error covariance, whereas reliability refers to the genotype main effect. Both condition on the fitted variance components.

```
> barley.prediction <- st.predict(barley.series, "gen")
> barley.prediction
```

```
Predicted means: average environment + genetic effect
  gen    prediction      SE   DF  lower.CL  upper.CL  reliability
  lne01     36.8501  1.9936  Inf   32.9427   40.7574       0.6376
  lne02     44.1093  1.9936  Inf   40.2019   48.0166       0.6376
  lne03     45.8882  1.9936  Inf   41.9808   49.7955       0.6376
  lne04     40.9527  1.9936  Inf   37.0454   44.8601       0.6376
  lne05     43.6698  1.9936  Inf   39.7624   47.5771       0.6376
  lne06     42.1637  1.9936  Inf   38.2564   46.0711       0.6376
  lne07     42.0965  1.9936  Inf   38.1892   46.0039       0.6376
  lne08     42.1406  1.9936  Inf   38.2333   46.0480       0.6376
  lne09     43.8823  1.9936  Inf   39.9749   47.7896       0.6376
  lne10     38.4597  1.9936  Inf   34.5524   42.3671       0.6376
  lne11     43.8333  1.9936  Inf   39.9260   47.7407       0.6376
  lne12     44.4298  1.9936  Inf   40.5225   48.3372       0.6376
  lne13     37.7532  1.9936  Inf   33.8458   41.6605       0.6376
  lne14     44.7179  1.9936  Inf   40.8106   48.6253       0.6376
  lne15     46.3324  1.9936  Inf   42.4250   50.2397       0.6376
Confidence level: 0.9500
SE and intervals use prediction-error variance; fitted variances held
  fixed.
```

The inherited residual covariance is the block diagonal matrix of the 15 first-stage adjusted-mean covariance matrices, on their original scale. Its multiplier is fixed at one. The `gen:env` term supplies the remaining transient genetic variation; an additional independent units variance would be confounded with this term and is therefore omitted. Each trial can have its own estimated replicate and plot-error variances.

This dataset has a complete location–year grid. The same construction allows missing location–year combinations by fitting only observed trials; estimation then depends on connectivity and adequate overlap. Three years provide limited information about the genotype-by-year variance. Boundary estimates should be interpreted in that context.

11 Genome wide association studies

This example analyses rapeseed flowering time with genome-wide SNP markers. The example covers kinship-adjusted marker scans, principal-component covariates, a resampled genome-wide significance threshold and a regional plot.

11.1 Rapeseed marker and flowering-time data

The data comprise 193 rapeseed lines with observed flowering time and 22,857 SNPs, provided by Rod Snowdon (University of Giessen). All markers and lines are included in the package data.

The packaged STvcf object and phenotype table provide the inputs for a SelectionTools marker data set.

```
> data("v-rapeseed", package = "SelectionTools")
> st.set.num.threads(1)
> st.load.vcf.data(v.rapeseed.vcf,
+                  data.set = "rapeseed")
```

```
M (data set 'rapeseed'): New data set generated 'rapeseed'
M (data set 'rapeseed'): No. of individuals: 193, no. of markers: 22857
```

The physical map is imported with the STvcf object and retained in the marker data set. Association scans use these stored chromosome labels and base-pair positions automatically.

```
> head(v.rapeseed.phe)
```

```
    gen   flt
1 SW001 154.5
2 SW002 157.0
3 SW003 154.0
4 SW004 156.0
5 SW005 155.5
6 SW006 152.5
```

11.2 Kinship-adjusted GWAS with P3D

Each marker model has an additive fixed marker effect, an intercept and a random polygenic effect with covariance proportional to the genomic relationship matrix. This accounts for relatedness among lines, following the mixed-model association framework of Yu et al. (2006). With `P3D=TRUE`, the null-model covariance is estimated once and its shape is held fixed during the marker scan. This is the variance-reuse strategy underlying EMMAX and P3D (Kang et al., 2010; Zhang et al., 2010).

```
> rapeseed.k <- st.gwas(
+     fixed = flt ~ 1,
+     random = ~gen,
+     data = v.rapeseed.phe,
+     marker.data.set = "rapeseed",
+     K = "A",
+     min.MAF = 0.05,
+     P3D = TRUE)
> print(rapeseed.k, top = 3, bottom = 0)
```

```
Genome-wide additive marker tests
F test; P3D: TRUE
  Trait  Marker     MAF   Effect      SE       p  Status
  flt    Q02002  0.0751  -0.7318  0.7182  0.3095  ok
  flt    Q02016  0.2202   0.6240  0.4838  0.1987  ok
  flt    Q02023  0.1839   0.6280  0.5165  0.2256  ok
  [22854 table rows omitted]
```

Symbolic `K = "A"` constructs the genomic relationship from the complete marker panel. The marker tests retain lines with observed flowering time and require a minor allele frequency of at least 0.05 among those lines. The output table keeps filtered markers with their status. For the ordinary residual model used here, SelectionTools reports a one-degree-of-freedom marker F test with marker-model residual degrees of freedom and a re-estimated residual scale.

The `plot` method draws the Manhattan plot directly from the fitted `st.gwas` object. Chromosomes and marker positions are read from that object. Setting `threshold=FALSE` suppresses a cutoff line.

```
> plot(rapeseed.k, which = "manhattan",
+      main = "Flowering time: K model", threshold = FALSE,
+      cex = 0.42)
```

Figure 1 shows the Manhattan plot for the K model.

The QQ diagnostic for this K model is shown with the PK model below.

11.3 Re-estimating variance components for every marker

A second scan uses `P3D=FALSE`. Free variance components are now re-estimated with each marker included in the fixed model. This illustrates the marker-specific variance-estimation strategy used in EMMA (Kang et al. (2008)).

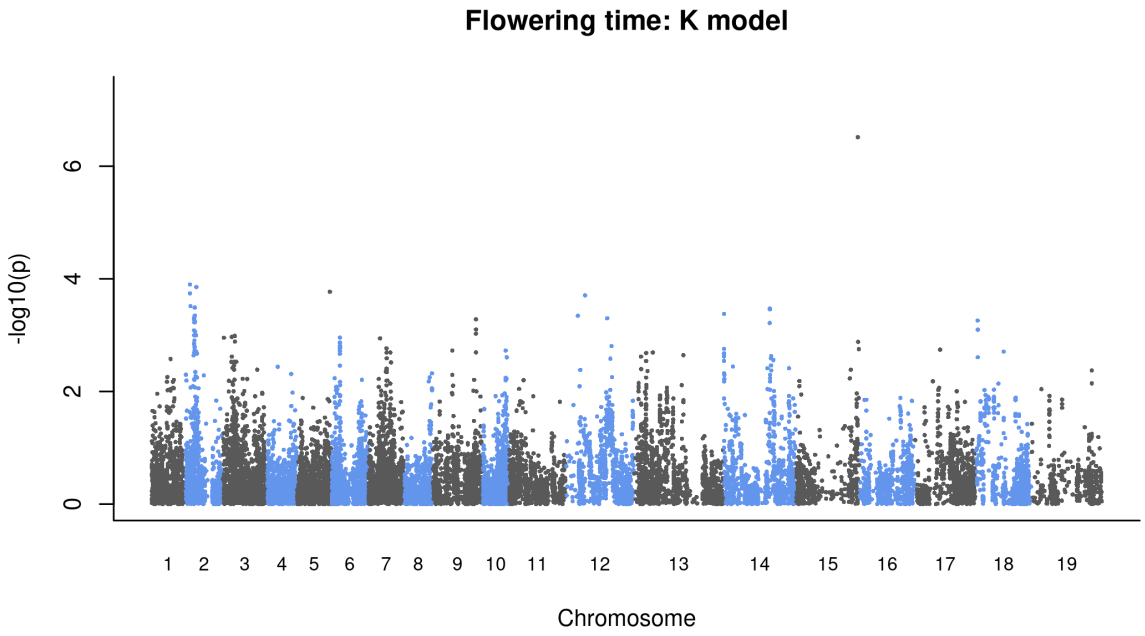


Figure 1: Genome-wide association scores for flowering time under the kinship model with P3D and without principal-component covariates. Chromosome labels and positions come from the supplied physical map.

```
> rapeseed.k.reml <- st.gwas(
+     fixed = flt ~ 1,
+     random = ~gen,
+     data = v.rapeseed.phe,
+     marker.data.set = "rapeseed",
+     K = "A",
+     P3D = FALSE)
> print(rapeseed.k.reml, top = 3, bottom = 0)
```

```
Genome-wide additive marker tests
F test; P3D: FALSE
  Trait  Marker     MAF   Effect      SE       p  Status
  flt    Q02002  0.0751  -0.7318  0.7182  0.3095  ok
  flt    Q02016  0.2202   0.6256  0.4833  0.1971  ok
  flt    Q02023  0.1839   0.6343  0.5156  0.2201  ok
  [22854 table rows omitted]
```

11.4 Kinship and principal-component adjustment

Symbolic `K = "A"` constructs the relationship matrix from the full marker panel. For these biallelic genotypes, the marker dose d_{ij} counts 0, 1 or 2 copies of the chosen allele at marker j in line i. SelectionTools retains the two allele calls and derives doses for the calculations. With allele frequency p_j, centring and scaling give

$$Z_{ij} = d_{ij} - 2p_j, \qquad K = \frac{ZZ^{\mathsf{T}}}{2\sum_j p_j(1-p_j)}.$$

Allele frequencies are calculated from observed genotypes in the complete marker panel. The sum covers polymorphic biallelic markers; missing doses are replaced by $2p_j$ for this calculation, giving a centred score of zero. Scaling uses one denominator for the matrix, following VanRaden (2008) and Endelman and Jannink (2012).

The next scan retains that matrix and adds three principal-component covariates. These covariates are unit eigenvectors of the relationship matrix, calculated internally.

```
> rapeseed.pk3 <- st.gwas(
+     fixed = flt ~ 1,
+     random = ~gen,
+     data = v.rapeseed.phe,
+     marker.data.set = "rapeseed",
+     K = "A",
+     n.PC = 3,
+     P3D = TRUE)
```

The comparison counts successful marker tests and reports FDR decisions at 0.05 using the explicit adjustment and significance-level options below. Model labels use the fitted object names. When significant markers occur on several chromosomes, the summary reports the lead marker on each chromosome.

```
> compare.gwas.fits(
+     rapeseed.k,
+     rapeseed.pk3,
+     adjust = "FDR",
+     alpha = 0.05
+ )
```

```
GWAS summary: FDR; alpha = 0.05
  Model         Trait  PCs  Tested  Significant
  rapeseed.k    flt      0   22836            1
  rapeseed.pk3  flt      3   22836            1

Chromosome leads (overall lead when none are significant)

Model: rapeseed.k; trait: flt
  Marker  Chr  Position     MAF  Effect        p     FDR
  Q46054  15   41874539  0.1503  3.0980  <0.0001  0.0070

Model: rapeseed.pk3; trait: flt
  Marker  Chr  Position     MAF  Effect        p     FDR
  Q46054  15   41874539  0.1503  3.2351  <0.0001  0.0034
```

Both models identify Q46054 as their leading association. FDR decisions refer to the complete marker scan for each model and trait.

```
> par(mfrow = c(1, 2), mar = c(4.5, 4.5, 2.2, 0.7),
+     mgp = c(2.7, 0.8, 0), cex.axis = 0.85)
> plot(rapeseed.k, which = "qq", main = "K", level = 0.95,
+      col = "cornflowerblue", cex = 0.38)
> plot(rapeseed.pk3, which = "qq", main = "PK: 3 PCs", level = 0.95,
+      col = "cornflowerblue", cex = 0.38)
```

Figure 2 compares the K model with the model including three PCs.

The following table shows the leading markers from the model with three principal components.

```
> rapeseed.summary <- summarize.gwas.fit(
+     rapeseed.pk3,
+     adjust = "FDR",
+     alpha = 0.05
+ )
> rapeseed.summary
```

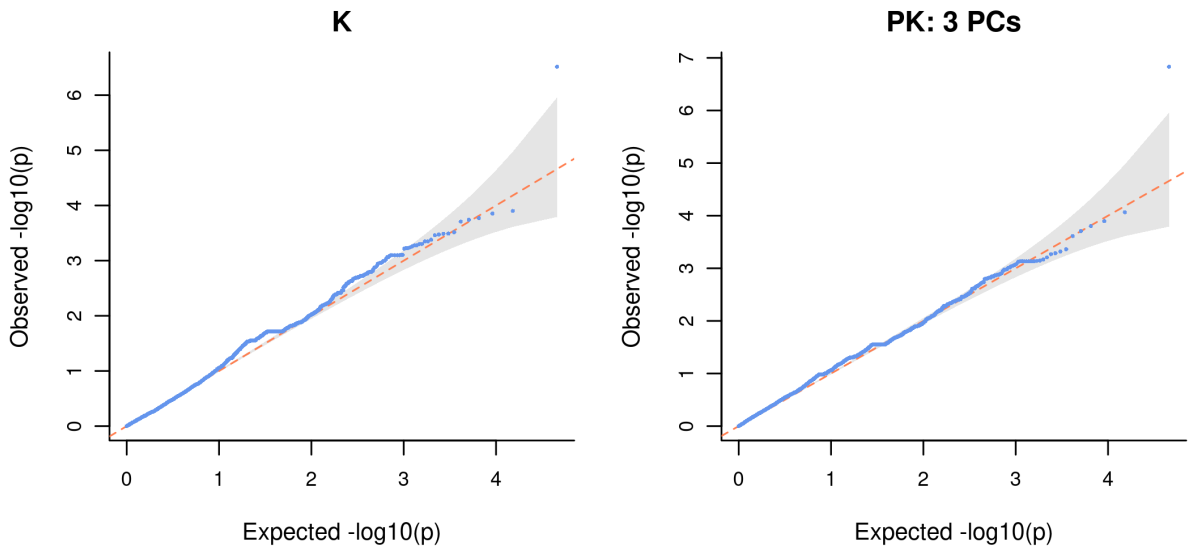


Figure 2: QQ comparison of the K model and the PK model with three principal-component covariates. The envelopes are pointwise diagnostic references.

```
GWAS summary: FDR; alpha = 0.05
  Model         Trait  PCs  Tested  Significant
  rapeseed.pk3  flt      3   22836            1

Leading markers

Model: rapeseed.pk3; trait: flt
  Marker  Chr  Position     MAF   Effect        p     FDR
  Q46054  15   41874539  0.1503   3.2351  <0.0001  0.0034
  Q02704  2     7109895  0.0570  -3.1822  <0.0001  0.7338
  Q04021  5    22302146  0.0751   2.9345   0.0001  0.7338
  Q04467  2     2767764  0.2746  -1.6811   0.0002  0.7338
  Q02720  12   13080275  0.1347  -2.1281   0.0002  0.7338
  Q04468  2     2768143  0.2772  -1.6762   0.0002  0.7338
```

11.5 Resampled genome-wide significance threshold

A genome-wide cutoff for the model with three principal components is calibrated by simulating phenotypes under its fitted marker-free null model. The marker panel, relationship matrix, principal components, fixed covariates and marker filters are retained. Each simulated phenotype has newly generated polygenic and residual effects. This preserves the fitted covariance between related lines, as required for mixed-model null simulation (Joo et al., 2016).

A nonempty `sim.threshold` list enables the calculation. Its options are `nsim`, `alpha`, `seed` and `method`; omitted options default to 999 simulations, level 0.05 and the current random-number stream. The default empty `list()` performs no simulation. The explicit seed below makes the calculation reproducible while preserving the caller's random-number state.

Here `method = "parametricFixedVC"` holds the original fitted variance components fixed across simulated scans, retaining a separate residual scale for each marker's F test. This is a conditional, plug-in calibration. The alternative `method = "parametric"` refits the null model for each simulated response. With `P3D = TRUE`, each replicate reuses its own fitted covariance shape across markers. With `P3D = FALSE`, variance components are also refitted at each marker.

```
> rapeseed.pk3.sim <- st.gwas(
+     fixed = flt ~ 1,
```

```
+     random = ~gen,
+     data = v.rapeseed.phe,
+     marker.data.set = "rapeseed",
+     K = "A",
+     n.PC = 3,
+     P3D = TRUE,
+     control = list(batch.size = 1024L),
+     sim.threshold = list(method = "parametricFixedVC",
+                          nsim = 999L, alpha = 0.05, seed = 1L))
> rapeseed.pk3.sim
```

```
Genome-wide additive marker tests
F test; P3D: TRUE
  Trait  Marker     MAF   Effect      SE       p   p.sim  Status
  flt    Q02002  0.0751  -0.7227  0.7285  0.3224  1.0000  ok
  flt    Q02016  0.2202   0.5869  0.4849  0.2277  1.0000  ok
  flt    Q02023  0.1839   0.5679  0.5182  0.2745  1.0000  ok
  flt    Q02025  0.2591   0.4758  0.4516  0.2935  1.0000  ok
  flt    Q02078  0.0803  -0.9455  0.9672  0.3296  1.0000  ok
  flt    Q02079  0.2383  -0.1076  0.4725  0.8202  1.0000  ok
  flt    Q02218  0.2280   0.2366  0.4817  0.6239  1.0000  ok
  flt    Q02222  0.1321  -0.6579  0.5712  0.2509  1.0000  ok
  flt    Q02233  0.3083  -0.0854  0.4373  0.8455  1.0000  ok
  flt    Q02247  0.2202   0.8041  0.4728  0.0907  1.0000  ok
  [22842 table rows omitted]
  flt    Q23145  0.1606  -0.3432  0.6040  0.5705  1.0000  ok
  flt    Q22687  0.2824   0.3636  0.4271  0.3957  1.0000  ok
  flt    Q39875  0.0699   0.4304  0.7167  0.5488  1.0000  ok
  flt    Q39833  0.1166  -0.2814  0.6203  0.6507  1.0000  ok
  flt    Q51873  0.0881  -0.0148  0.6696  0.9823  1.0000  ok

Simulated genome-wide thresholds (parametric null; each trait separately)
Replicates: 999; alpha: 0.05; seed: 1
  Trait  Markers  Successful  Failed   Score        p  Status
  flt      22836         999       0  5.3205  <0.0001  ok
Use p.sim <= alpha; score cutoffs require a strict exceedance.
```

For replicate b, let M_b be the largest association score $-\log_{10}(p)$ among the eligible markers. For an observed marker score s, the reported simulation-adjusted probability is

$$p_{\mathrm{sim}}(s) = \frac{1 + \#\{b : M_b \geq s\}}{B + 1}.$$

The additional count prevents a simulated probability of zero (Phipson and Smyth, 2010). With $B = 999$, the smallest reported value is 0.001. The score cutoff at level α is the kth smallest maximum, where $k = B + 1 - \lfloor \alpha(B+1) \rfloor$. For this run $k = 950$. A score must be strictly above the cutoff to pass; the count-based `p.sim` values determine decisions in the presence of ties.

```
> summarize.gwas.simulation(rapeseed.pk3.sim)
```

```
GWAS summary: simulation; alpha = 0.05
  Model             Trait  PCs  Tested  Significant
  rapeseed.pk3.sim  flt      3   22836            1

Leading markers

Model: rapeseed.pk3.sim; trait: flt
  Marker  Chr   Score        p   p.sim
  Q46054  15   6.8320  <0.0001  0.0010
  Q02704  2    4.0665  <0.0001  0.4900
  Q04021  5    3.8963   0.0001  0.6150
  Q04467  2    3.8013   0.0002  0.6970
```

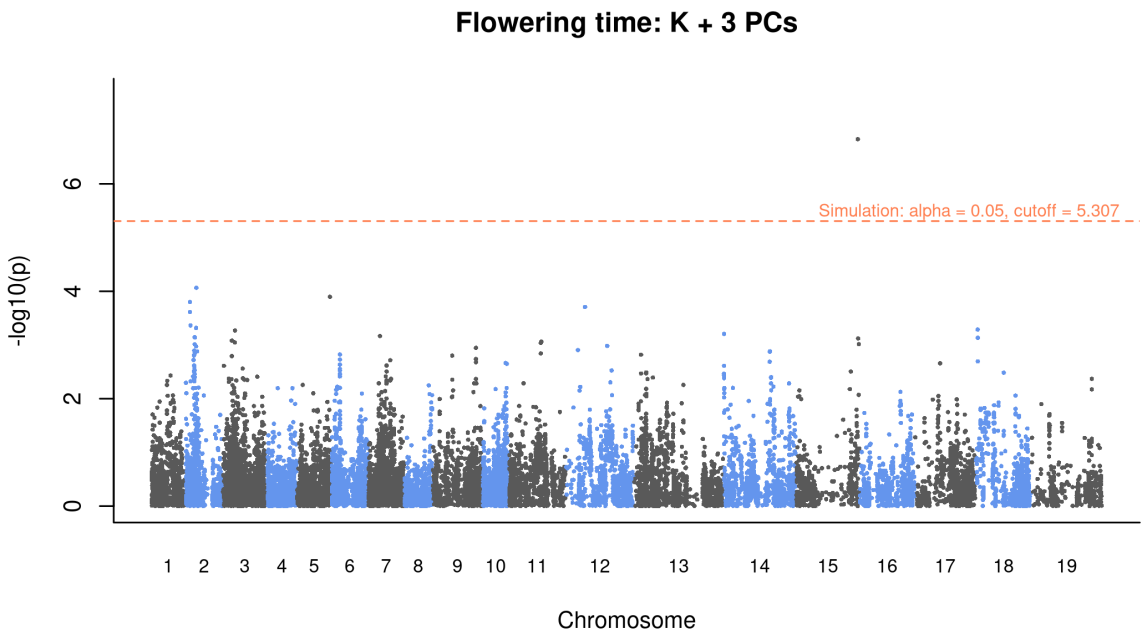


Figure 3: Genome-wide association scores for flowering time after adjustment for kinship and three principal components. The coral dashed line gives the genome-wide 0.05 cutoff from 999 null simulations.

```
Q02720  12   3.7079   0.0002  0.7700
Q04468  2    3.6147   0.0002  0.8310
```

This calibration targets the probability of at least one false positive across the eligible markers under the fitted complete-null model for this trait. It depends on that model and its estimated covariance; it does not jointly adjust the different traits or model choices. All requested simulated scans must complete. Failed fits or eligible marker tests are recorded in `simulation$diagnostics`, and an incomplete calibration does not receive a threshold or adjusted probabilities. The maxima are retained in `simulation$max.score`.

The Manhattan helper reads the saved threshold automatically. Its coral dashed line is labelled with the level and score cutoff. Plotting only uses saved results; it performs no further simulation. An explicit `threshold=FALSE`, numerical raw-p-value threshold, or `adjust` setting overrides the automatic choice.

```
> plot(rapeseed.pk3.sim, which = "manhattan", adjust = "simulation",
+      main = "Flowering time: K + 3 PCs",
+      cex = 0.42)
```

Figure 3 shows the Manhattan plot with the resampled genome-wide threshold.

11.6 Regional association

The regional plot around Q46054 on chromosome 15 highlights a 500-kb interval and the nominal $p < 10^{-5}$ cutoff.

```
> rapeseed.lead <- rapeseed.summary$markers[1, ]
```

```
> plot(rapeseed.pk3, which = "region",
+      chromosome = rapeseed.lead$chromosome,
+      xlim = rapeseed.lead$position + c(-2e6, 2e6),
+      position.scale = 1e6, position.unit = "Mb",
```

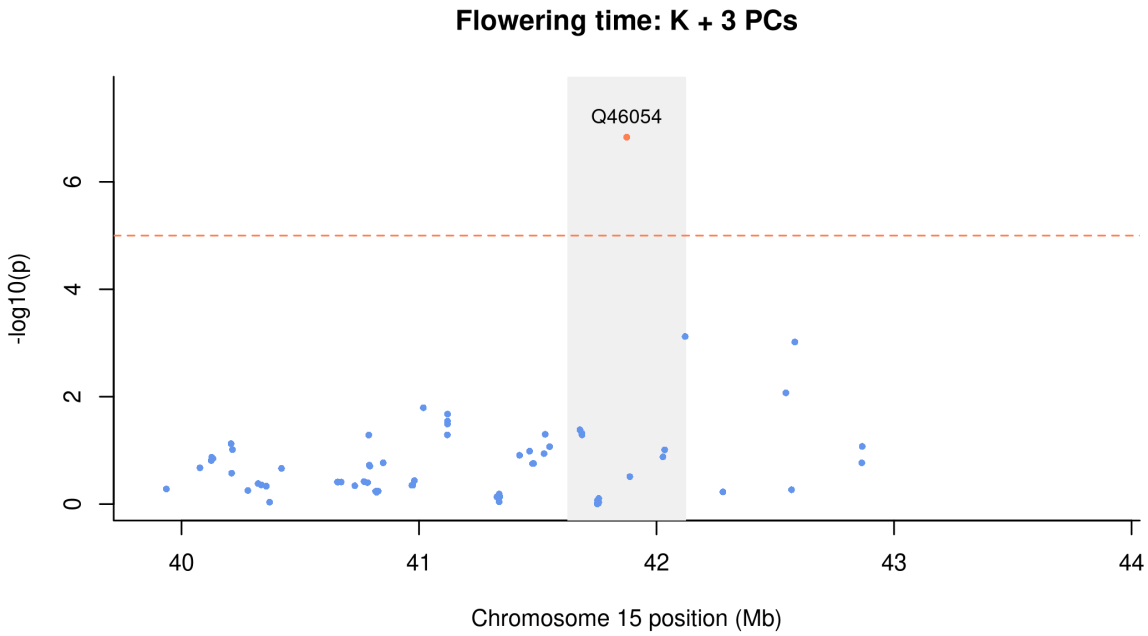


Figure 4: Regional Manhattan plot around Q46054 on chromosome 15. The shaded 500-kb interval contains 17 markers; the dashed line and highlighted point identify the nominal $p < 10^{-5}$ criterion.

```
+       shade = rapeseed.lead$position + c(-250000, 250000),
+       highlight = rapeseed.lead$marker, labels = TRUE,
+       threshold = 1e-5, main = "Flowering time: K + 3 PCs",
+       col = "cornflowerblue", cex = 0.65)
```

Figure 4 shows the regional Manhattan plot around Q46054.

12 Genomic prediction

12.1 Two-step genomic prediction with one dataset containing the unobserved phenotypic values of the prediction set as NA

The first commands attach SelectionTools and load the plot data and marker calls for the tropical maize example. The marker calls are stored in `"maize.lmm"`.

```
> library(SelectionTools)
> st.set.num.threads(1)
> data("v-lmm-tropmaize", package = "SelectionTools")
> data("v-tropmaize-vcf", package = "SelectionTools")
> st.load.vcf.data(v.tropmaize.vcf, data.set = "maize.lmm")
```

```
M (data set 'maize.lmm'): No. of individuals: 264, no. of markers: 1076
```

`st.get.A` obtains `maize.A` directly from the marker data set.

```
> maize.A <- st.get.A("maize.lmm")
```

One plot table retains all genotypes. Responses of the first ten genotypes are set to `NA`; only observed responses enter the field fit.

```
> maize.genotypes <- levels(tropmaize$gen)
> maize.masked <- tropmaize
> maize.masked$y[maize.masked$gen %in% maize.genotypes[1:10]] <- NA
```

12.1.1 Step 1: Analysis of the field trial

The field model fits genotype means with a random block effect.

```
> maize.train.field <- st.lmm(
+     fixed = y ~ gen,
+     random = ~block,
+     data = maize.masked
+ )
```

The genotype labels and least-squares means form the input to the second analysis.

```
> maize.means <- as.data.frame(st.lsm(maize.train.field, "gen"))
> maize.means <- maize.means[, c("gen", "lsmeans")]
```

12.1.2 Step 2: Genomic prediction

The second model uses these means and `maize.A`.

```
> maize.two.step <- st.lmm(
+     fixed = lsmeans ~ 1,
+     random = ~gen,
+     ucov = list(gen = maize.A),
+     data = maize.means
+ )
> maize.two.step
```

```
Linear mixed model fitted by REML

Model:
  Data:      maize.means
  Fixed:     lsmeans ~ 1
  Random:    ~gen
  Residual:  ~units
  Covariance: gen (264 levels; supplied matrix)

Convergence:
  converged: projected score tolerance met

Random effects (likelihood-ratio tests):
  Term  DF       LR        p
  gen    1  35.0225  <0.0001

Variance components:
  Component  Variance
  gen          0.3280
  Residual     0.4438

Fit and model statistics:
  Observations               254    R-squared (conditional)  0.5613
  Log-likelihood        -346.7073   RMSE (conditional)       0.4931
  AIC                    699.4146   Variance: fixed          0.0000
  BIC                    710.0266   Variance: random         0.5677
  R-squared (marginal)     0.0000   Variance: residual       0.4438
```

`st.predict` requests predictions for the `gen` term and saves the result as `maize.two.pred`. Its displayed table includes genotype labels, predictions, SEs and confidence limits.

```
> maize.two.pred <- st.predict(maize.two.step, "gen")
> maize.two.pred
```

```
Predicted means: fixed intercept + random effect
  gen  prediction      SE   DF  lower.CL  upper.CL
  1        2.4533  0.3425  Inf    1.7820    3.1246
  2        1.9477  0.6873  Inf    0.6007    3.2948
  3        1.2916  0.7218  Inf   -0.1230    2.7062
  4        1.3752  0.6972  Inf    0.0087    2.7418
  5        1.7952  0.6145  Inf    0.5909    2.9996
  6        2.3619  0.5580  Inf    1.2683    3.4556
  7        2.3133  0.6748  Inf    0.9907    3.6359
  8        2.1934  0.5645  Inf    1.0870    3.2998
  9        2.1548  0.5521  Inf    1.0727    3.2368
  10       2.0612  0.5657  Inf    0.9525    3.1698
  [249 table rows omitted]
  260      1.7935  0.4114  Inf    0.9872    2.5998
  261      1.5648  0.4786  Inf    0.6267    2.5030
  262      2.1054  0.4485  Inf    1.2263    2.9846
  263      1.5831  0.4395  Inf    0.7217    2.4446
  264      1.4003  0.4637  Inf    0.4915    2.3090
Confidence level: 0.9500
SE and intervals use prediction-error variance; fitted variances held
  fixed.
```

12.2 Two-step genomic prediction with one training dataset and one prediction dataset

This example splits the marker data into training and prediction datasets. The supplied phenotype values are loaded into the training dataset; the prediction dataset contains markers only.

```
> library(SelectionTools)
> st.set.num.threads(1)
> data("v-tropmaize-vcf", package = "SelectionTools")
> st.load.vcf.data(v.tropmaize.vcf, data.set = "maize.split")
```

```
M (data set 'maize.split'): New data set generated 'maize.split'
M (data set 'maize.split'): No. of individuals: 264, no. of markers: 1076
```

`st.marker.data.statistics` retrieves the individual identifiers from the stored marker dataset. The marker and genotype tables are omitted.

```
> maize.marker.statistics <- st.marker.data.statistics(
+     data.set = "maize.split",
+     mar = FALSE,
+     gen = FALSE
+ )
```

```
M (data set 'maize.split'): No. of individuals: 264, no. of markers: 1076
```

The first ten stored individuals define the prediction set. The remaining individuals define the training set.

```
> maize.individuals <- maize.marker.statistics$individual.list$Name
> maize.prediction.ids <- maize.individuals[1:10]
> maize.training.ids <- maize.individuals[-(1:10)]
```

`st.copy.marker.data` copies the marker dataset. The restriction retains only the selected training individuals in `"maize.training"`.

```
> st.copy.marker.data("maize.training", "maize.split")
> st.restrict.marker.data(
+     ind.list = maize.training.ids,
+     data.set = "maize.training"
+ )
```

```
M (data set 'maize.training'): New data set generated 'maize.training'
M (data set 'maize.training'): No. of individuals: 264, no. of markers: 1076
M (data set 'maize.training'): No. of individuals: 254, no. of markers: 1076
```

The second copy retains only the prediction individuals. Both datasets keep the same marker columns.

```
> st.copy.marker.data("maize.prediction", "maize.split")
> st.restrict.marker.data(
+     ind.list = maize.prediction.ids,
+     data.set = "maize.prediction"
+ )
```

```
M (data set 'maize.prediction'): New data set generated 'maize.prediction'
M (data set 'maize.prediction'): No. of individuals: 264, no. of markers: 1076
M (data set 'maize.prediction'): No. of individuals: 10, no. of markers: 1076
```

The field-data analysis results are supplied as genotype phenotype values in `v-tropmaize-phe`. `st.load.performance.data` loads only values matching the training marker identifiers.

```
> data("v-tropmaize-phe", package = "SelectionTools")
> st.load.performance.data(
+     v.tropmaize.phe,
+     data.set = "maize.training"
+ )
```

```
M (data set 'maize.training'): No. of individuals: 254, no. of markers: 1076
M (data set 'maize.training'): No. of performance data: 254
```

With a marker dataset as `data`, the model variables `y` and `gen` are supplied automatically: `y` denotes the stored phenotype values and `gen` is the factor identifying individuals. Both names appear unquoted in the formulas. The quoted `"A"` is a symbolic covariance specification that constructs the genomic relationship matrix from the training markers.

```
> maize.training.model <- st.lmm(
+     fixed = y ~ 1,
+     random = ~gen,
+     ucov = list(gen = "A"),
+     data = "maize.training"
+ )
> maize.training.model
```

```
Linear mixed model fitted by REML

Model:
  Data:      "maize.training"
  Fixed:     y ~ 1
```

```
  Random:    ~gen
  Residual:  ~units
  Covariance: gen (254 levels; supplied matrix)

Convergence:
  converged: projected score tolerance met

Random effects (likelihood-ratio tests):
  Term  DF       LR        p
  gen    1  35.0225  <0.0001

Variance components:
  Component  Variance
  gen          0.3281
  Residual     0.4438

Fit and model statistics:
  Observations               254    R-squared (conditional)  0.5612
  Log-likelihood        -346.7073    RMSE (conditional)       0.4931
  AIC                    699.4146    Variance: fixed          0.0000
  BIC                    710.0266    Variance: random         0.5676
  R-squared (marginal)     0.0000    Variance: residual       0.4438
```

`st.predict` uses the fitted training model and the marker genotypes in `"maize.prediction"`. Its result contains the prediction individuals' identifiers and predicted values.

```
> maize.predictions <- st.predict(
+     maize.training.model,
+     newdata = "maize.prediction"
+ )
> maize.predictions
```

```
      gen  prediction
   1  1        2.4533
   2  2        1.9477
   3  3        1.2916
   4  4        1.3752
   5  5        1.7952
   6  6        2.3619
   7  7        2.3133
   8  8        2.1934
   9  9        2.1548
  10  10       2.0612
```

12.3 Genomic prediction with a multi-environment trial

12.3.1 Joint analysis of plot data

The multi-environment plot table and its marker calls are loaded for this example. `st.load.vcf.data` stores the marker data under `"maize.lattice"`.

```
> st.set.num.threads(1)
> data("v-lmm-tropmaize-lattice", package = "SelectionTools")
> data("v-tropmaize-vcf", package = "SelectionTools")
> st.load.vcf.data(v.tropmaize.vcf, data.set = "maize.lattice")
```

```
M (data set 'maize.lattice'): New data set generated 'maize.lattice'
M (data set 'maize.lattice'): No. of individuals: 264, no. of markers: 1076
```

The relationship matrix is extracted from that marker data set. The first ten genotype levels identify the entries whose responses will be hidden in every environment.

```
> maize.lattice.A <- st.get.A("maize.lattice")
> maize.lattice.missing.gen <- levels(tropmaize.lattice$gen)[seq_len(10)]
```

The copy `maize.lattice.masked` retains every plot row. Responses for the selected genotypes are replaced by `NA`, including their records in different environments.

```
> maize.lattice.masked <- tropmaize.lattice
> maize.lattice.masked$y[
+     maize.lattice.masked$gen %in% maize.lattice.missing.gen
+ ] <- NA_real_
```

The fixed formula gives each environment its own mean. The random formula includes genotype, replicate within environment, block within replicate and environment, and genotype-by-environment terms. `ucov` supplies the matrix for `gen`; `data` selects the masked plot table.

```
> maize.lattice.one.step <- st.lmm(
+     fixed = y ~ env,
+     random = ~gen + env:rep + env:rep:block + gen:env,
+     ucov = list(gen = maize.lattice.A),
+     data = maize.lattice.masked
+ )
> maize.lattice.one.step
```

```
Linear mixed model fitted by REML

Model:
  Data:       maize.lattice.masked
  Fixed:      y ~ env
  Random:     ~gen + env:rep + env:rep:block + gen:env
  Residual:   ~units
  Covariance: gen (264 levels; supplied matrix)

Convergence:
  converged: projected score tolerance met

Fixed effects (Satterthwaite):
  Term  DF1    DF2        F        p
  env     4  19.26  61.9321  <0.0001

Random effects (likelihood-ratio tests):
  Term           DF         LR        p
  gen             1  2950.7508  <0.0001
  env:rep         1     2.0260   0.1546
  env:rep:block   1   355.7666  <0.0001
  gen:env         1   793.7083  <0.0001

Variance components:
  Component      Variance
  gen              1.0986
  env:rep          0.0004
  env:rep:block    0.0082
  gen:env          0.0281
  Residual         0.0252

Fit and model statistics:
  Observations              3810    R-squared (conditional)  0.9873
  Log-likelihood       -199.1523    RMSE (conditional)       0.1300
  AIC                   418.3046    Variance: fixed          0.0195
```

```
  BIC                    480.7584    Variance: random         1.9380
  R-squared (marginal)     0.0099    Variance: residual       0.0252
```

`st.predict` extracts the genotype predictions from the fitted model. The `"gen"` request returns the fixed intercept plus the genotype effect; the additional environment, replicate, block and interaction terms remain in the fitted model.

```
> maize.lattice.one.pred <- st.predict(maize.lattice.one.step, "gen")
> maize.lattice.one.pred
```

```
Predicted means: fixed intercept + random effect
  gen  prediction      SE   DF  lower.CL  upper.CL
  1        2.6996  0.2394  Inf    2.2304    3.1689
  2        1.7111  1.1043  Inf   -0.4533    3.8754
  3       -0.2781  1.1493  Inf   -2.5307    1.9744
  4       -0.0483  1.1286  Inf   -2.2603    2.1637
  5        1.3228  0.9658  Inf   -0.5701    3.2158
  6        2.7753  0.7815  Inf    1.2436    4.3070
  7        2.0736  1.0873  Inf   -0.0575    4.2048
  8        2.4038  0.9047  Inf    0.6306    4.1771
  9        1.6077  0.8942  Inf   -0.1449    3.3604
  10       1.3819  0.8842  Inf   -0.3511    3.1149
  [249 table rows omitted]
  260      1.4438  0.0880  Inf    1.2714    1.6162
  261      0.5373  0.0881  Inf    0.3645    0.7101
  262      2.5289  0.0880  Inf    2.3564    2.7014
  263      0.9073  0.0880  Inf    0.7347    1.0798
  264      0.6770  0.0882  Inf    0.5042    0.8498
Confidence level: 0.9500
SE and intervals use prediction-error variance; fitted variances held
  fixed.
```

12.4 Hybrid prediction: Flint and dent maize

12.4.1 Parental markers and relationship matrices

The two marker objects contain the flint and dent parental pools, and `v.hybmaize.phe` supplies the hybrid responses and parental identifiers. The first import stores the flint markers under the name `"flintmaize"`.

```
> st.set.num.threads(1)
> data("v-flintmaize-vcf", package = "SelectionTools")
> data("v-dentmaize-vcf", package = "SelectionTools")
> data("v-hybmaize-phe", package = "SelectionTools")
> st.load.vcf.data(v.flintmaize.vcf, data.set = "flintmaize")
```

```
M (data set 'flintmaize'): New data set generated 'flintmaize'
M (data set 'flintmaize'): No. of individuals: 86, no. of markers: 35478
```

The second import stores the dent markers separately as `"dentmaize"`, preserving the identifiers of that parental pool.

```
> st.load.vcf.data(v.dentmaize.vcf, data.set = "dentmaize")
```

```
M (data set 'dentmaize'): New data set generated 'dentmaize'
M (data set 'dentmaize'): No. of individuals: 123, no. of markers: 35478
```

The flint marker filter retains markers with at most two alleles and positive expected heterozygosity. The small positive `ExHet.MIN` threshold removes markers that are monomorphic in this pool.

```
> st.restrict.marker.data(ExHet.MIN = 1e-8, NoAll.MAX = 2,
+                         data.set = "flintmaize")
```

```
M (data set 'flintmaize'): No. of individuals: 86, no. of markers: 28240
```

The same filter is applied within the dent pool. Each pool therefore uses markers that remain informative among its own parents.

```
> st.restrict.marker.data(ExHet.MIN = 1e-8, NoAll.MAX = 2,
+                         data.set = "dentmaize")
```

```
M (data set 'dentmaize'): No. of individuals: 123, no. of markers: 29108
```

The two `st.get.A` calls return the parental relationship matrices. Their row names identify the parents used in the subsequent model.

```
> Aflint <- st.get.A("flintmaize")
> Adent <- st.get.A("dentmaize")
```

The performance frame is copied to `hybmaize`. Its parental factor levels are aligned with the corresponding relationship-matrix names; `flint` and `dent` then hold the parent identifier for each row.

```
> hybmaize <- v.hybmaize.phe
> hybmaize$flint <- factor(hybmaize$flint, levels = rownames(Aflint))
> hybmaize$dent <- factor(hybmaize$dent, levels = rownames(Adent))
> flint <- as.character(hybmaize$flint)
> dent <- as.character(hybmaize$dent)
```

Each parental matrix is indexed by the parents of the observed hybrids. Elementwise multiplication forms `Asca` for those crosses, and colon-separated parental identifiers label its rows and columns. `hybrid.covariance.dimensions` reports its dimensions.

```
> Asca <- Aflint[flint, flint] * Adent[dent, dent]
> hybrid.names <- paste(flint, dent, sep = ":")
> dimnames(Asca) <- list(hybrid.names, hybrid.names)
> hybrid.covariance.dimensions <- dim(Asca)
> hybrid.covariance.dimensions
```

```
[1] 1254 1254
```

12.4.2 GCA plus SCA with ten missing hybrid phenotypes

The first ten hybrid records have both response columns set to missing. The selected rows of `hybrids.without.responses` show the hybrid identifiers and the masked grain-yield values.

```
> hybmaize[seq_len(10), c("GY", "GM")] <- NA_real_
> hybrids.without.responses <- hybmaize[seq_len(10), c("hybrid", "GY")]
> hybrids.without.responses
```

```
    hybrid GY
1  518.298 NA
2  518.305 NA
3  518.306 NA
4  518.316 NA
5  518.323 NA
6  518.327 NA
7  518.437 NA
8  518.439 NA
9  518.444 NA
10 545.298 NA
```

The model uses grain yield `GY` and an intercept in the fixed part. The random formula contains separate flint and dent effects and their interaction. The named `ucov` entries assign `Aflint`, `Adent` and `Asca` to those three terms.

```
> hybrid.fit <- st.lmm(
+     fixed = GY ~ 1,
+     random = ~flint + dent + flint:dent,
+     ucov = list(
+         flint = Aflint,
+         dent = Adent,
+         "flint:dent" = Asca
+     ),
+     data = hybmaize
+ )
> hybrid.fit
```

```
Linear mixed model fitted by REML

Model:
  Data:      hybmaize
  Fixed:     GY ~ 1
  Random:    ~flint + dent + flint:dent
  Residual:  ~units
  Covariance: flint (86 levels; supplied matrix)
  Covariance: dent (123 levels; supplied matrix)
  Covariance: flint:dent (1254 levels; supplied matrix)

Convergence:
  converged: projected score tolerance met

Random effects (likelihood-ratio tests):
  Term        DF        LR        p
  flint        1  529.4587  <0.0001
  dent         1  693.3686  <0.0001
  flint:dent   1   41.2301  <0.0001

Variance components:
  Component   Variance
  flint        11.1037
  dent         15.9394
  flint:dent    1.2223
  Residual     13.8906

Fit and model statistics:
  Observations               1244    R-squared (conditional)   0.7997
  Log-likelihood        -3748.7430    RMSE (conditional)        3.1598
  AIC                    7507.4860    Variance: fixed           0.0000
  BIC                    7533.1164    Variance: random         55.4643
  R-squared (marginal)      0.0000    Variance: residual       13.8906
```

12.4.3 Hybrid response predictions

The prediction call evaluates every row of `hybmaize`, including rows with missing yield. `type = "conditional"` includes the fitted fixed and random contributions; the `fit` component is saved as `hybrid.predicted`.

```
> hybrid.predicted <- predict(
+     hybrid.fit,
+     newdata = hybmaize,
+     type = "conditional"
+ )$fit
> hybrid.predicted
```

```
             GY
1       -9.2068
2       -9.8837
3      -11.5460
4        4.2884
5        0.2847
6      -12.6736
7       -5.0252
8       -7.4678
9        4.9095
10      -3.4079
[1239 table rows omitted]
1250    13.6018
1251    13.1619
1252    10.7008
1253    13.0992
1254    11.1308
```

The number of missing yield observations is stored separately in `hybrid.missing.phenotypes`.

```
> hybrid.missing.phenotypes <- sum(is.na(hybmaize$GY))
> hybrid.missing.phenotypes
```

```
[1] 10
```

The corresponding count for the prediction matrix is stored in `hybrid.missing.predictions`.

```
> hybrid.missing.predictions <- sum(is.na(hybrid.predicted))
> hybrid.missing.predictions
```

```
[1] 0
```

12.5 GBLUP with SelectionTools marker data sets

This example stores markers and phenotypes together under one data-set name. The supplied tropical-maize marker and phenotype objects are loaded, and the marker calls are imported as `"maize"`.

```
> st.set.num.threads(1)
> data("v-tropmaize-vcf", package = "SelectionTools")
> data("v-tropmaize-phe", package = "SelectionTools")
> st.load.vcf.data(v.tropmaize.vcf, data.set = "maize")
```

```
M (data set 'maize'): New data set generated 'maize'
M (data set 'maize'): No. of individuals: 264, no. of markers: 1076
```

The `yy` values for the first ten marker identifiers are set to missing before the phenotype table is attached to `"maize"`. The marker identifiers remain available in that data set.

```
> v.tropmaize.phe$yy[
+     v.tropmaize.phe$ii %in% v.tropmaize.vcf$individual[seq_len(10)]
+ ] <- NA_real_
> st.load.performance.data(v.tropmaize.phe, data.set = "maize")
```

```
M (data set 'maize'): No. of individuals: 264, no. of markers: 1076
M (data set 'maize'): No. of performance data: 264
```

The response values are retrieved with their genotype names by `gs.get.y`. The first ten values are saved in `maize.missing.responses` to display the masked entries.

```
> maize.y <- gs.get.y("maize", names = TRUE)
> maize.missing.responses <- head(maize.y, 10)
> maize.missing.responses
```

```
 1  2  3  4  5  6  7  8  9 10
NA NA NA NA NA NA NA NA NA NA
```

Here `data = "maize"` supplies the stored response and genotype identifiers directly to `st.lmm`. The quoted covariance request `gen = "A"` constructs the genotype relationship matrix from that marker data set.

```
> maize.two.step <- st.lmm(
+     fixed = y ~ 1,
+     random = ~gen,
+     ucov = list(gen = "A"),
+     data = "maize"
+ )
> maize.two.step
```

```
Linear mixed model fitted by REML

Model:
  Data:       "maize"
  Fixed:      y ~ 1
  Random:     ~gen
  Residual:   ~units
  Covariance: gen (264 levels; supplied matrix)

Convergence:
  converged: projected score tolerance met

Random effects (likelihood-ratio tests):
  Term  DF       LR        p
  gen    1  35.0225  <0.0001

Variance components:
  Component  Variance
  gen          0.3280
  Residual     0.4438
```

```
Fit and model statistics:
  Observations                254    R-squared (conditional)  0.5613
  Log-likelihood        -346.7073    RMSE (conditional)       0.4931
  AIC                    699.4146    Variance: fixed          0.0000
  BIC                    710.0266    Variance: random         0.5677
  R-squared (marginal)     0.0000    Variance: residual       0.4438
```

The call to `st.predict` requests genotype response predictions from `maize.two.step`. The returned object retains the genotype identifiers, predictions and their uncertainty.

```
> maize.two.pred <- st.predict(maize.two.step, "gen")
> maize.two.pred
```

```
Predicted means: fixed intercept + random effect
  gen  prediction      SE   DF  lower.CL  upper.CL
  1        2.4533  0.3425  Inf    1.7820    3.1246
  2        1.9477  0.6873  Inf    0.6007    3.2948
  3        1.2916  0.7218  Inf   -0.1230    2.7062
  4        1.3752  0.6972  Inf    0.0087    2.7418
  5        1.7952  0.6145  Inf    0.5909    2.9996
  6        2.3619  0.5580  Inf    1.2683    3.4556
  7        2.3133  0.6748  Inf    0.9907    3.6359
  8        2.1934  0.5645  Inf    1.0870    3.2998
  9        2.1548  0.5521  Inf    1.0727    3.2368
  10       2.0612  0.5657  Inf    0.9525    3.1698
  [249 table rows omitted]
  260      1.7935  0.4114  Inf    0.9872    2.5998
  261      1.5648  0.4786  Inf    0.6267    2.5030
  262      2.1054  0.4485  Inf    1.2263    2.9846
  263      1.5831  0.4395  Inf    0.7217    2.4446
  264      1.4003  0.4637  Inf    0.4915    2.3090
Confidence level: 0.9500
SE and intervals use prediction-error variance; fitted variances held
  fixed.
```

The length of the retrieved response vector gives the number of marker genotypes represented in `maize.y`.

```
> maize.marker.genotypes <- length(maize.y)
> maize.marker.genotypes
```

```
[1] 264
```

The missing response entries are counted in `maize.missing.phenotypes`.

```
> maize.missing.phenotypes <- sum(is.na(maize.y))
> maize.missing.phenotypes
```

```
[1] 10
```

The variable `maize.missing.predictions` counts missing entries in the prediction result.

```
> maize.missing.predictions <- sum(is.na(coef(maize.two.pred)))
> maize.missing.predictions
```

```
[1] 0
```

12.6 Hybrid prediction with SelectionTools marker data sets

The next example uses stored parental data sets rather than supplying relationship matrices as R objects. The two parental marker objects and the hybrid phenotype table are loaded, and the flint markers are stored under the name `"flint"`.

```
> st.set.num.threads(1)
> data("v-flintmaize-vcf", package = "SelectionTools")
> data("v-dentmaize-vcf", package = "SelectionTools")
> data("v-hybmaize-phe", package = "SelectionTools")
> st.load.vcf.data(v.flintmaize.vcf, data.set = "flint")
```

```
M (data set 'flint'): New data set generated 'flint'
M (data set 'flint'): No. of individuals: 86, no. of markers: 35478
```

The dent marker calls are imported as `"dent"`. These two stored names match the parental columns in the hybrid performance table.

```
> st.load.vcf.data(v.dentmaize.vcf, data.set = "dent")
```

```
M (data set 'dent'): New data set generated 'dent'
M (data set 'dent'): No. of individuals: 123, no. of markers: 35478
```

The flint marker filter removes markers that are monomorphic within that pool and retains at most two alleles per marker.

```
> st.restrict.marker.data(ExHet.MIN = 1e-8, NoAll.MAX = 2,
+                         data.set = "flint")
```

```
M (data set 'flint'): No. of individuals: 86, no. of markers: 28240
```

The dent marker filter applies the same criteria independently to the second parental pool.

```
> st.restrict.marker.data(ExHet.MIN = 1e-8, NoAll.MAX = 2,
+                         data.set = "dent")
```

```
M (data set 'dent'): No. of individuals: 123, no. of markers: 29108
```

A copy of the hybrid performance table is retained in `hybrid.performance`. Both responses in its first ten rows are set to missing, leaving their parental identifiers unchanged.

```
> hybrid.performance <- v.hybmaize.phe
> hybrid.performance[seq_len(10), c("GY", "GM")] <- NA_real_
```

The constructor combines that performance frame with references to `"flint"` and `"dent"` under the name `"maizehybrids"`. The parental marker calls are shared with those existing data sets.

```
> st.hybrid.data.set(
+     markerdata1 = "flint",
+     markerdata2 = "dent",
+     hybrid.performance = hybrid.performance,
+     hybrid.data.set = "maizehybrids"
+ )
```

The masked records are displayed through `hybrid.missing.rows`, which contains their hybrid identifiers and grain-yield values.

```
> hybrid.missing.rows <- hybrid.performance[seq_len(10), c("hybrid", "GY")]
> hybrid.missing.rows
```

```
    hybrid GY
1  518.298 NA
2  518.305 NA
3  518.306 NA
4  518.316 NA
5  518.323 NA
6  518.327 NA
7  518.437 NA
8  518.439 NA
9  518.444 NA
10 545.298 NA
```

12.6.1 GCA only and prediction of all parental combinations

The fixed formula specifies an intercept for `GY`; the random formula contains the two parental effects. The two quoted `"A"` entries obtain their relationships from the linked parental marker data. `prediction = "Full"` stores predictions for every parental combination.

```
> hybrid.gca <- st.lmm(
+     fixed = GY ~ 1,
+     random = ~flint + dent,
+     ucov = list("flint" = "A", "dent" = "A"),
+     data = "maizehybrids",
+     prediction = "Full"
+ )
> hybrid.gca
```

```
Linear mixed model fitted by REML

Model:
  Data:       "maizehybrids"
  Fixed:      GY ~ 1
  Random:     ~flint + dent
  Residual:   ~units
  Covariance: flint (86 levels; supplied matrix)
  Covariance: dent (123 levels; supplied matrix)

Convergence:
  converged: projected score tolerance met

Random effects (likelihood-ratio tests):
  Term    DF        LR        p
  flint    1  550.6730  <0.0001
  dent     1  900.4038  <0.0001

Variance components:
  Component  Variance
  flint       10.8579
  dent        16.7089
  Residual    17.8369

Fit and model statistics:
  Observations                1244    R-squared (conditional)    0.7450
  Log-likelihood       -3769.3580    RMSE (conditional)         3.9353
  AIC                   7546.7161    Variance: fixed            0.0000
```

```
BIC                    7567.2204    Variance: random         52.1114
R-squared (marginal)      0.0000    Variance: residual       17.8369
```

This call separately evaluates conditional predictions for the rows in `hybrid.performance`. Its `fit` matrix is saved as `gca.predictions`, in the order of that performance frame.

```
> gca.predictions <- predict(
+     hybrid.gca,
+     newdata = hybrid.performance,
+     type = "conditional"
+ )$fit
> gca.predictions
```

```
             GY
1      -9.3700
2     -10.8441
3     -12.1808
4       3.3502
5      -0.0304
6     -11.6418
7      -3.2593
8      -7.2542
9       4.1474
10     -2.1485
[1239 table rows omitted]
1250   14.7874
1251   14.1453
1252    9.2118
1253   13.6140
1254   10.6454
```

The accessor retrieves the stored full grid of parental combinations. `gca.crosses.preview` shows its first six flint–dent pairs and their predicted responses.

```
> gca.crosses <- st.lmm.components(hybrid.gca, "predictions")
> gca.crosses.preview <- head(gca.crosses[, c("flint", "dent", "prediction")])
> gca.crosses.preview
```

```
   flint  dent  prediction
1  316    513       6.7632
2  316    518       3.3502
3  316    536       9.9896
4  316    545      10.5717
5  316    546      -0.8049
6  316    548      -2.3638
```

The number of rows in `gca.crosses` gives the number of parental combinations in that full grid.

```
> gca.cross.count <- nrow(gca.crosses)
> gca.cross.count
```

```
[1] 10578
```

The masked yield values in the performance frame are counted in `gca.missing.phenotypes`.

```
> gca.missing.phenotypes <- sum(is.na(hybrid.performance$GY))
> gca.missing.phenotypes
```

```
[1] 10
```

The prediction matrix for those performance rows is checked for missing entries, with the result saved in `gca.missing.predictions`.

```
> gca.missing.predictions <- sum(is.na(gca.predictions))
> gca.missing.predictions
```

```
[1] 0
```

12.6.2 GCA plus SCA and prediction of missing responses

The interaction `flint:dent` adds a third random term to the model. Its quoted `"kro(a,a)"` request derives the interaction relationships from the two parental matrices. `prediction = "NAonly"` stores predictions for the missing responses in the performance frame.

```
> st.set.num.threads(1)
> hybrid.gca.sca <- st.lmm(
+     fixed = GY ~ 1,
+     random = ~flint + dent + flint:dent,
+     ucov = list(
+         "flint" = "A",
+         "dent" = "A",
+         "flint:dent" = "kro(a,a)"
+     ),
+     data = "maizehybrids",
+     prediction = "NAonly"
+ )
> hybrid.gca.sca
```

```
Linear mixed model fitted by REML

Model:
  Data:       "maizehybrids"
  Fixed:      GY ~ 1
  Random:     ~flint + dent + flint:dent
  Residual:   ~units
  Covariance: flint (86 levels; supplied matrix)
  Covariance: dent (123 levels; supplied matrix)
  Covariance: flint:dent (1254 levels; supplied matrix)

Convergence:
  converged: projected score tolerance met

Random effects (likelihood-ratio tests):
  Term         DF         LR         p
  flint         1   529.4587   <0.0001
  dent          1   693.3686   <0.0001
  flint:dent    1    41.2301   <0.0001

Variance components:
  Component   Variance
  flint        11.1037
  dent         15.9394
  flint:dent    1.2223
  Residual     13.8906

Fit and model statistics:
```

```
Observations                1244    R-squared (conditional)    0.7997
Log-likelihood        -3748.7430    RMSE (conditional)         3.1598
AIC                    7507.4860    Variance: fixed            0.0000
BIC                    7533.1164    Variance: random          55.4643
R-squared (marginal)      0.0000    Variance: residual        13.8906
```

The accessor returns those stored predictions with both parent identifiers. The `source.row` column identifies the corresponding row in the retained performance frame.

```
> sca.predictions <- st.lmm.components(hybrid.gca.sca, "predictions")
> sca.predictions
```

```
      flint  dent  source.row  prediction
   1  298    518             1     -9.2068
   2  305    518             2     -9.8837
   3  306    518             3    -11.5460
   4  316    518             4      4.2884
   5  323    518             5      0.2847
   6  327    518             6    -12.6736
   7  437    518             7     -5.0252
   8  439    518             8     -7.4678
   9  444    518             9      4.9095
  10  298    545            10     -3.4079
```

The number of masked yield observations is saved in `sca.missing.phenotypes`.

```
> sca.missing.phenotypes <- sum(is.na(hybrid.performance$GY))
> sca.missing.phenotypes
```

```
[1] 10
```

The saved prediction column is checked for missing entries, and that count is displayed through `sca.missing.predictions`.

```
> sca.missing.predictions <- sum(is.na(sca.predictions$prediction))
> sca.missing.predictions
```

```
[1] 0
```

12.6.3 Hybrid data inspection

The listing returns information about the stored hybrid data sets. `hybrid.status.summary` retains the data-set and parental names, record and unique-hybrid counts, and the SCA-matrix indicator.

```
> hybrid.status <- st.hybrid.data.set.list()
> hybrid.status.summary <- hybrid.status[
+     , c("hybrid.data.set", "markerdata1", "markerdata2",
+         "no.records", "no.hybrids", "SCA.matrix")
+ ]
> hybrid.status.summary
```

```
    hybrid.data.~1 markerdata1 markerdata2 no.records no.hybrids SCA.matrix
  1 maizehybrids   flint       dent           1254.00    1254.00 TRUE
  hybrid.data.~1 [hybrid.data.set] = hybrid.data.set
```

12.7 Genomic prediction with simulated DH pools

The mapped maize data contain 86 flint and 123 dent parents, 8072 shared markers polymorphic in both pools, and 1254 hybrid performance records (Technow et al., 2014; Melchinger and Frisch, 2023). This example first fits the observed hybrid records, then creates simulated DH parents and crosses, assigns phenotypes to those crosses, and fits the simulated records. The package data provide the parental marker calls and observed hybrid responses.

```
> st.set.num.threads(1)
> data("v-flintmaize2-vcf")
> data("v-dentmaize2-vcf")
> data("v-hybmaize2-phe")
```

The three removal calls release hybrid data-set names used in preceding examples. The parental marker data are then loaded under the names `"flint"` and `"dent"`.

```
> st.hybrid.data.set.remove("maizehybrids")
> st.hybrid.data.set.remove("maizeTRUE")
> st.hybrid.data.set.remove("maizeTS")
```

Each `data.set` argument names the marker data set created by the load call. The two data sets retain their separate parental identifiers.

```
> st.load.vcf.data(v.flintmaize2.vcf, data.set = "flint")
> st.load.vcf.data(v.dentmaize2.vcf, data.set = "dent")
```

```
M (data set 'flint'): No. of individuals: 86, no. of markers: 8072
M (data set 'dent'): No. of individuals: 123, no. of markers: 8072
```

`st.hybrid.data.set` associates the observed hybrid records with the two parental marker data sets. Its last argument gives this hybrid data set the name `"maizeTRUE"`.

```
> st.hybrid.data.set("flint", "dent", v.hybmaize2.phe, "maizeTRUE")
```

The model uses `GY` as its response. The terms `flint` and `dent` identify parental effects, and `flint:dent` identifies the cross-specific effect. The `ucov` entries request parental relationship matrices with `"A"` and their cross relationship with `"kro(a,a)"`. The fitted object is stored in `maize.founder.model`; the hybrid data set also retains the information needed for later prediction.

```
> maize.founder.model <- st.lmm(
+     GY ~ 1,
+     random = ~flint + dent + flint:dent,
+     ucov = list(flint = "A", dent = "A", "flint:dent" = "kro(a,a)"),
+     data = "maizeTRUE"
+ )
> maize.founder.model
```

```
Linear mixed model fitted by REML

Model:
  Data:       "maizeTRUE"
  Fixed:      GY ~ 1
  Random:     ~flint + dent + flint:dent
  Residual:   ~units
  Covariance: flint (86 levels; supplied matrix)
```

```
  Covariance: dent (123 levels; supplied matrix)
  Covariance: flint:dent (1254 levels; supplied matrix)

Convergence:
  converged: projected score tolerance met

Random effects (likelihood-ratio tests):
  Term        DF        LR        p
  flint        1  551.5060  <0.0001
  dent         1  700.7613  <0.0001
  flint:dent   1   45.8906  <0.0001

Variance components:
  Component   Variance
  flint        10.8961
  dent         15.3697
  flint:dent    1.3144
  Residual     13.4673

Fit and model statistics:
  Observations               1254    R-squared (conditional)    0.8029
  Log-likelihood        -3772.9671    RMSE (conditional)         3.0813
  AIC                    7555.9342    Variance: fixed            0.0000
  BIC                    7581.6047    Variance: random          54.8625
  R-squared (marginal)      0.0000    Variance: residual        13.4673
```

The four variables specify the total marker count after map extension, the number of DH lines in each pool, the number of simulated hybrid records, and the target `h2` used to generate phenotypes.

```
> n.markers <- 9000L
> n.dh      <- 100L
> n.hybrids <- 200L
> h2        <- 0.8
```

The thread setting and seed reproduce the simulation configuration. `reset.all()` clears existing simulation populations and genome settings before the marker data are transferred.

```
> st.set.num.threads(1)
> rng.init(seed = 20261024L)
> reset.all()
```

`st.set.sim.gp` defines the simulation genome from the flint map, and `st.set.sim.mp` copies its mapped loci. The two `st.set.sim.pp` calls create founder populations from the parental marker data. `extend.marker.map` increases the total to 9000 loci while retaining the original marker identifiers and positions.

```
> st.set.sim.gp("flint")
> st.set.sim.mp("flint")
> st.set.sim.pp("simFlint", "flint")
> st.set.sim.pp("simDent", "dent")
> extend.marker.map(n.markers)
```

```
M: 10 chromosomes defined
M: 8072 loci defined
```

Each `cross` call creates an F1 population by random mating within one founder pool. The setting `mode = 0` excludes selection of the same individual as both parents. Each `dh` call samples parents from that F1 population and produces 100 DH lines. The names distinguish the F1 and DH populations in the simulation.

```
> cross("simFlintF1", "simFlint", "simFlint", n.dh, mode = 0)
> cross("simDentF1", "simDent", "simDent", n.dh, mode = 0)
> dh("simflintDH", "simFlintF1", n.dh)
> dh("simdentDH", "simDentF1", n.dh)
```

`st.get.simpop` exports the simulated marker calls, identifiers and pedigrees to marker data sets. The resulting names `"TSflint"` and `"TSdent"` are used by the subsequent hybrid analysis. The added marker calls follow the inherited chromosome segments.

```
> st.get.simpop("simflintDH", data.set = "TSflint")
> st.get.simpop("simdentDH", data.set = "TSdent")
```

```
M (data set 'TSflint'): New data set generated 'TSflint'
M (data set 'TSflint'): No. of individuals: 100, no. of markers: 9000
M (data set 'TSdent'): New data set generated 'TSdent'
M (data set 'TSdent'): No. of individuals: 100, no. of markers: 9000
```

The pedigree accessor reads the exported flint marker data set. `flint.pedigree` contains its first three identifier and pedigree records.

```
> flint.pedigree <- head(
+     population.pedigree("TSflint", source = "marker"),
+     3
+ )
> flint.pedigree
```

```
                          name        pedigree
1 simflintDH.410((338_x_430)DH) ((338_x_430)DH)
2 simflintDH.411((332_x_354)DH) ((332_x_354)DH)
3 simflintDH.412((347_x_316)DH) ((347_x_316)DH)
```

`population.cross` creates 200 random crosses and returns their hybrid and parental identifiers. The renamed columns match the two exported marker data sets. `st.hybrid.data.set` then makes this crossing design available as `"maizeTS"`.

```
> hybrid.design <- population.cross(
+     "simHybrids", "simflintDH", "simdentDH", n.hybrids
+ )
> names(hybrid.design) <- c("hybrid", "TSflint", "TSdent")
> st.hybrid.data.set("TSflint", "TSdent", hybrid.design, "maizeTS")
```

Prediction uses the model retained with `"maizeTRUE"` for the crosses named in `"maizeTS"`. It matches the 8072 markers used in the original fit. The prediction table supplies the values and parental identifiers to `st.generate.phenotypes`, which adds random residuals and stores the generated response as `GY` in `"maizeTS"`. With `h2 = 0.8`, the added residual variance is one quarter of the sample variance of the supplied predictions.

```
> hybrid.predictions <- as.data.frame(
+     st.predict("maizeTRUE", newdata = "maizeTS")
+ )
> st.generate.phenotypes("maizeTS", hybrid.predictions, h2 = h2, trait = "GY")
```

The second fit uses all 200 generated responses. Its parental and cross-specific random terms refer to the exported DH marker data sets. Their symbolic relationship specifications now use the complete 9000-marker panel. `maize.simulated.model` contains this fitted model.

```
> maize.simulated.model <- st.lmm(
+     GY ~ 1,
+     random = ~TSflint + TSdent + TSflint:TSdent,
+     ucov = list(
+         TSflint = "A", TSdent = "A",
+         "TSflint:TSdent" = "kro(a,a)"
+     ),
+     data = "maizeTS"
+ )
> maize.simulated.model
```

```
Linear mixed model fitted by REML

Model:
  Data:       "maizeTS"
  Fixed:      GY ~ 1
  Random:     ~TSflint + TSdent + TSflint:TSdent
  Residual:   ~units
  Covariance: TSflint (100 levels; supplied matrix)
  Covariance: TSdent (100 levels; supplied matrix)
  Covariance: TSflint:TSdent (196 levels; supplied matrix)

Convergence:
  converged: projected score tolerance met

Random effects (likelihood-ratio tests):
  Term            DF        LR        p
  TSflint          1   77.4058  <0.0001
  TSdent           1  110.7390  <0.0001
  TSflint:TSdent   1    0.6904   0.4060

Variance components:
  Component       Variance
  TSflint           4.5291
  TSdent            6.5047
  TSflint:TSdent    0.5622
  Residual          4.0426

Fit and model statistics:
  Observations               200    R-squared (conditional)   0.8575
  Log-likelihood        -567.2473    RMSE (conditional)        1.1600
  AIC                  1144.4947    Variance: fixed           0.0000
  BIC                  1160.9863    Variance: random         24.3238
  R-squared (marginal)    0.0000    Variance: residual        4.0426
```

The final call predicts the records in `"maizeTS"` from its retained model. `coef` extracts their genetic values, and `first.genotypic.values` contains the first six rows for display.

```
> genotypic.values <- coef(st.predict("maizeTS"))
> first.genotypic.values <- head(genotypic.values)
> first.genotypic.values
```

```
  Row     Value
  1     -2.5119
  2     -9.3634
  3    -13.1846
  4     -0.0778
  5     -4.4069
  6      2.8736
```

13 References

13.1 Experimental designs and analysis of variance

Completely randomized designs, randomized blocks, Latin squares, factorial experiments, split plots, incomplete blocks, lattices and expected mean squares.

Cochran WG (1941) The distribution of the largest of a set of estimated variances as a fraction of their total. Annals of Eugenics 11:47–52

Cochran's C test for homogeneity of cell variances

Cornfield J, Tukey JW (1956) Average values of mean squares in factorials. Annals of Mathematical Statistics 27:907–949

The original reference for the Cornfield–Tukey rules.

Fisher RA (1925) Statistical methods for research workers. 1st edn. Oliver and Boyd, Edinburgh

Chapter VIII, §49, explicitly develops the Latin square as an experimental design and gives its ANOVA.

Fisher RA (1926) The arrangement of field experiments. Journal of the Ministry of Agriculture of Great Britain 33:503–513

A foundational original paper on randomization, blocking and factorial experimentation.

Fisher RA (1935) The design of experiments. 1st edn. Oliver and Boyd, Edinburgh

The foundational systematic treatment of experimental design, including randomized blocks, Latin squares and factorial experiments.

Fletcher DJ, Underwood AJ (2002) How to cope with negative estimates of components of variance in ecological field studies. Journal of Experimental Marine Biology and Ecology 273:89–95

Explicit pooling and the interpretation of negative component estimates

Patterson HD, Williams ER (1976) A new class of resolvable incomplete block designs. Biometrika 63:83–92

Alpha-lattice designs

Underwood AJ (1997) Experiments in ecology: their logical design and interpretation using analysis of variance. Cambridge University Press, Cambridge

Particularly relevant to the balanced crossed, nested and split-plot examples.

Yates F (1935) Complex experiments. Supplement to the Journal of the Royal Statistical Society 2:181–223

An early detailed treatment, including the oats experiment used in the vignette.

Yates F (1936) A new method of arranging variety trials involving a large number of varieties. Journal of Agricultural Science 26:424–455

Introduces the arrangements underlying standard lattice designs.

Yates F (1936) Incomplete randomized blocks. Annals of Eugenics 7:121–140

An original reference for incomplete-block construction and analysis.

Yates F (1937) The design and analysis of factorial experiments. Technical Communication No. 35, Imperial Bureau of Soil Science, Harpenden

Covers factorial arrangements, confounding and split plots.

Yates F (1940) Lattice squares. Journal of Agricultural Science 30:672–687

Develops their analysis, including recovery of inter-row and inter-column information.

13.2 Linear mixed models and estimation

Variance components, mixed-model equations, BLUE and BLUP, ML and REML, optimization algorithms and random regression.

Anderson E, Bai Z, Bischof C, Blackford S, Demmel J, Dongarra J, Du Croz J, Greenbaum A, Hammarling S, McKenney A, Sorensen D (1999) LAPACK users' guide. 3rd edn. SIAM, Philadelphia

Dense linear algebra, least squares, factorizations, eigensystems and BLAS

Byrd RH, Lu P, Nocedal J, Zhu C (1995) A limited memory algorithm for bound constrained optimization. SIAM Journal on Scientific Computing 16:1190–1208

L-BFGS-B optimization

Davis TA (2006) Direct methods for sparse linear systems. SIAM, Philadelphia

Sparse direct factorization and triangular solves

de Boor C (2001) A practical guide to splines. Revised edn. Springer, New York

Spline bases used in regression and prediction

Draper NR, Smith H (1998) Applied regression analysis. 3rd edn. Wiley, New York

Least squares, polynomial regression, pure error and lack-of-fit tests

Gilmour AR, Thompson R, Cullis BR (1995) Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. Biometrics 51:1440–1450

Average-information REML

Harville DA (1977) Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association 72:320–338

Likelihood estimation with structured covariance components

Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423–447

BLUE, BLUP and mixed-model equations

Laird NM, Ware JH (1982) Random-effects models for longitudinal data. Biometrics 38:963–974

Relevant to the chicken-growth analyses.

McLean RA, Sanders WL, Stroup WW (1991) A unified approach to mixed linear models. The American Statistician 45:54–64. doi:10.1080/00031305.1991.10475767

Linear functions of fixed and random effects and their inference spaces.

Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. Biometrika 58:545–554

REML estimation

Sanders WL (1991) The use of the SAS system in statistical training. Proceedings of the Sixteenth Annual SAS Users Group International Conference

Machine and operator data and the inference-space comparisons.

Searle SR (1987) Linear models for unbalanced data. Wiley, New York

Coding, estimability, adjusted sums of squares and general linear hypotheses

Searle SR, Casella G, McCulloch CE (1992) Variance components. Wiley, New York

Covers moment estimation, ML, REML and prediction in mixed models.

Wilkinson GN, Rogers CE (1973) Symbolic description of factorial models for analysis of variance. Applied Statistics 22:392–399

The original symbolic model-formula notation

13.3 Statistical inference and model assessment

Satterthwaite and Kenward–Roger methods, bootstrap tests, Hotelling's T^2, MANOVA, AIC, BIC and mixed-model R^2.

Akaike H (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control 19:716–723

Akaike information criterion

Anderson TW (2003) An introduction to multivariate statistical analysis. 3rd edn. Wiley, Hoboken

Covers the multivariate linear model and the Wilks, Pillai, Hotelling–Lawley and Roy criteria.

Davison AC, Hinkley DV (1997) Bootstrap methods and their application. Cambridge University Press, Cambridge

Relevant to simulation under the reduced model and refitting both models.

Fai AHT, Cornelius PL (1996) Approximate F-tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. Journal of Statistical Computation and Simulation 54:363–378

Multiple-degree-of-freedom approximate F-tests

Hotelling H (1931) The generalization of Student's ratio. Annals of Mathematical Statistics 2:360–378

Hotelling's T^2 and its multivariate F reference

Johnson PCD (2014) Extension of Nakagawa & Schielzeth's R^2_{GLMM} to random slopes models. Methods in Ecology and Evolution 5:944–946

Relevant to observation-averaged random-effect variances.

Kenward MG, Roger JH (1997) Small sample inference for fixed effects from restricted maximum likelihood. Biometrics 53:983–997

Supports covariance adjustment, F scaling and denominator degrees of freedom.

Nakagawa S, Schielzeth H (2013) A general and simple method for obtaining R^2 from generalized linear mixed-effects models. Methods in Ecology and Evolution 4:133–142

Marginal and conditional R^2

Satterthwaite FE (1946) An approximate distribution of estimates of variance components. Biometrics Bulletin 2:110–114

Satterthwaite degrees of freedom

Schwarz G (1978) Estimating the dimension of a model. Annals of Statistics 6:461–464

Bayesian information criterion

Self SG, Liang KY (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. Journal of the American Statistical Association 82:605–610

Explains why variance-boundary likelihood-ratio inference is nonstandard

Wald A (1943) Tests of statistical hypotheses concerning several parameters when the number of observations is large. Transactions of the American Mathematical Society 54:426–482

Wald tests and large-sample normal inference

Wilks SS (1938) The large-sample distribution of the likelihood ratio for testing composite hypotheses. Annals of Mathematical Statistics 9:60–62

The regular asymptotic chi-squared likelihood-ratio reference

13.4 Least-squares means, contrasts and multiple comparisons

Least-squares means, reference grids, transformations, simultaneous inference, multiplicity adjustments and compact letter displays.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B 57:289–300

Benjamini–Hochberg false discovery rate

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Annals of Statistics 29:1165–1188

Benjamini–Yekutieli adjustment

Bretz F, Hothorn T, Westfall P (2010) Multiple comparisons using R. CRC Press, Boca Raton

Weighted contrast families, simultaneous inference and step-down procedures

Dunnett CW (1955) A multiple comparison procedure for comparing several treatments with a control. Journal of the American Statistical Association 50:1096–1121

Comparisons with a control

Genz A, Bretz F (2009) Computation of multivariate normal and t probabilities. Lecture Notes in Statistics 195. Springer, Berlin, Heidelberg

Relevant to simultaneous inference using contrast correlations.

Hochberg Y (1988) A sharper Bonferroni procedure for multiple tests of significance. Biometrika 75:800–802

Hochberg adjustment

Hochberg Y, Tamhane AC (1987) Multiple comparison procedures. Wiley, New York

Bonferroni adjustment and general multiple-comparison theory

Holm S (1979) A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6:65–70

Holm adjustment

Hommel G (1988) A stagewise rejective multiple test procedure based on a modified Bonferroni test. Biometrika 75:383–386

Hommel adjustment

Hothorn T, Bretz F, Westfall P (2008) Simultaneous inference in general parametric models. Biometrical Journal 50:346–363

General linear hypotheses and simultaneous inference

Lenth RV (2016) Least-squares means: the R package least-squares means. Journal of Statistical Software 69(1):1–33

A useful methodological reference for the marginal-means interface.

Marcus R (1976) The powers of some tests of the equality of normal means against an ordered alternative. Biometrika 63:177–183

Ordered-alternative tests and Marcus contrasts

Marcus R, Peritz E, Gabriel KR (1976) On closed testing procedures with special reference to ordered analysis of variance. Biometrika 63:655–660

The closed-testing principle underlying logical multiplicity procedures

Oehlert GW (1992) A note on the delta method. The American Statistician 46:27–29

Relevant to transformed predictions and approximate bias corrections.

Piepho HP (2004) An algorithm for a letter-based representation of all-pairwise comparisons. Journal of Computational and Graphical Statistics 13:456–466

Compact letter displays

Scheffé H (1953) A method for judging all contrasts in the analysis of variance. Biometrika 40:87–110

Scheffé simultaneous inference

Searle SR, Speed FM, Milliken GA (1980) Population marginal means in the linear model: an alternative to Least-squares means. The American Statistician 34:216–221

Least-squares means

Shaffer JP (1986) Modified sequentially rejective multiple test procedures. Journal of the American Statistical Association 81:826–831

Improved step-down tests using logical relations among hypotheses

Šidák Z (1967) Rectangular confidence regions for the means of multivariate normal distributions. Journal of the American Statistical Association 62:626–633

Šidák adjustment

Tukey JW (1949) Comparing individual means in the analysis of variance. Biometrics 5:99–114

Tukey pairwise comparisons

Westfall PH (1997) Multiple testing of general contrasts using logical constraints and correlations. Journal of the American Statistical Association 92:299–306

Correlation-based step-down inference and logical constraints

Williams DA (1971) A test for differences between treatment means when several dose levels are compared with a zero dose control. Biometrics 27:103 –117

The ordered-dose foundation of Williams-type contrasts

13.5 Plant breeding, association studies and genomic prediction

Multi-environment trials, prediction of genotypic values, genomic relationships, one-stage and two-stage analyses, and GCA/SCA hybrid prediction.

Covarrubias-Pazaran G (2016) Genome-assisted prediction of quantitative traits using the R package sommer. PLoS ONE 11:e0156744

Source documentation for the small maize half-diallel data.

Crossa J et al. (2010) Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186:713–724

Source of the tropical-maize markers and observed phenotypes.

Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. Journal of Agricultural, Biological, and Environmental Statistics 11:381–393

Heritability from prediction errors of BLUP differences

Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. The Plant Genome 4:250–255

Mixed-model marker scans and ridge-regression kernels.

Endelman JB (2023) Fully efficient, two-stage analysis of multi-environment trials with directional dominance and multi-trait genomic selection. Theoretical and Applied Genetics 136:65

Full covariance propagation and genomic prediction in two stages.

Endelman JB, Jannink JL (2012) Shrinkage estimation of the realized relationship matrix. G3: Genes, Genomes, Genetics 2:1405–1413

Centring and scaling of marker-based relationship matrices.

Joo JWJ, Hormozdiari F, Han B, Eskin E (2016) Multiple testing correction in linear mixed models. Genome Biology 17:62

Parametric null simulation preserving the covariance between related individuals.

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

EMMA and marker-specific variance-component estimation.

Kang HM et al. (2010) Variance component model to account for sample structure in genome-wide association studies. Nature Genetics 42:348–354

Null-model covariance estimation for efficient marker scans.

Melchinger AE, Frisch M (2023) Genomic prediction in hybrid breeding: II. Reciprocal recurrent genomic selection with full-sib and half-sib families. Theoretical and Applied Genetics 136:203

Source of the mapped maize parental data used for the simulation example.

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Dense-marker genomic prediction

Phipson B, Smyth GK (2010) Permutation p-values should never be zero: calculating exact p-values when permutations are randomly drawn. Statistical Applications in Genetics and Molecular Biology 9:Article 39

Finite Monte Carlo tail probabilities with an additional count.

Piepho HP, Möhring J (2007) Computing heritability and selection response from unbalanced plant breeding trials. Genetics 177:1881–1888

Entry-mean heritability in unbalanced trials

Piepho HP, Möhring J, Melchinger AE, Büchse A (2008) BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161:209–228

BLUE and BLUP in plant breeding

Piepho HP, Möhring J, Schulz-Streeck T, Ogutu JO (2012) A stage-wise approach for the analysis of multi-environment trials. Biometrical Journal 54:844–860

Full covariance propagation of adjusted means between stages.

Schulz-Streeck T, Ogutu JO, Piepho HP (2013) Comparisons of single-stage and two-stage approaches to genomic selection. Theoretical and Applied Genetics 126:69–82

One-stage versus two-stage genomic prediction

Smith AB, Cullis BR, Thompson R (2005) The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. Journal of Agricultural Science 143:449–462

Mixed models for multi-environment trials

Sprague GF, Tatum LA (1942) General vs. specific combining ability in single crosses of corn. Agronomy Journal 34:923–932

The original general- and specific-combining-ability distinction

Technow F, Schrag TA, Schipprack W, Bauer E, Simianer H, Melchinger AE (2014) Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. Genetics 197:1343–1355

Relevant to parental GCA, hybrid SCA and the flint–dent example.

VanRaden PM (2008) Efficient methods to compute genomic predictions. Journal of Dairy Science 91:4414–4423

The direct reference for the genomic relationship matrix used in the maize examples.

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

Population structure and kinship adjustment in association mapping.

Zhang Z et al. (2010) Mixed linear model approach adapted for genome-wide association studies. Nature Genetics 42:355–360

Population parameters previously determined (P3D) for marker scans.