

ASReml: AN OVERVIEW

Rajender Parsad¹, Jose Crossa² and Juan Burgueno²
¹I.A.S.R.I., Library Avenue, New Delhi - 110 012, India
²Biometrics and statistics Unit, CIMMYT, Mexico

ASReml is a statistical package that fits linear mixed effects models using Residual Maximum Likelihood (REML). It has been under development since 1993 and is a joint venture between Biometrics Program of NSW Agriculture and the Biomathematics Unit previously the Statistics Department of Rothamsted Experimental Station.

Linear mixed effects models provide a rich and flexible tool for the analysis of many data sets commonly arising in the agricultural, biological, medical and environmental sciences. ASREML has applications in the analysis of typical applications include the analysis of

- (un)balanced longitudinal data,
- repeated measures data (multivariate analysis of variance and spline type models)
- (un)balanced designed experiments,
- multi-environment trials analysis,
- univariate and multivariate animal breeding and genetics data (involving a relationship matrix for correlated effects)
- and the analysis of regular or irregular spatial data.

Further

- The engine of ASReml forms the basis of the REML procedure in GENSTAT. An interface for S-PLUS called samm is also available.
- Handles large data sets (of 100,000 or more observations/effects).
- Supports a wide range of variance models for spatial analysis.

ASREML is distributed by VSN-International on behalf of NSW Agriculture and Rothamsted Research. Proceeds are used to support continued development. Further information can be obtained from <http://www.VSN-Intl.com> or by sending an email to info@VSN-Intl.com.

Data Analysis Using ASReml

Data File:

The data should be arranged in columns with a single line for each sampling unit. Columns must be separated by at least one blank space, TAB (ASCII file) or comma (in .CSV file). Characteristics of the data file are:

- Identifiers (headings, titles) of columns may be included at the top.
- Identifiers (headings, titles) and attributes (data, factor levels) must be alphanumeric.
- Missing data must be represented by a dot/period (.), an asterisk (*) or NA in free format data files. In .CSV a line beginning with a comma implies a preceding missing value, consecutive columns imply a missing value, etc.
- A number sign (#) and dollar sign (\$) have special meanings. Neither may appear in the data file.

Command File

By convention an ASReml command file has a .as extension. The command file consists of five sections. These are:

- I. Title
- II. Definition of data columns
- III. Name of the data file
- IV. Linear model
- V. Variance structure (when necessary)

I. Title Line: The first 40 characters of the first non-blank line in an ASReml command file is taken as the title for the job. It is used to identify the analysis for future reference.

II. Definition of Data Columns: The Labels for the data field must be given in the order in which they appear in the data file. The data field definitions must be indented and are given as

SPACE *label* [*field_type*]

Here

SPACE is a required space,

label is an alphanumeric string beginning with a letter and is of maximum of 31 characters of which only 20 are printed and

field_type indicates how a variable is interpreted if specified in the linear model.

- i. For a variate, leave *field_type* blank or specify 1.
- ii. For a model factor, use * or *n* if the data field has values from 1 to *n*; !A *n* if the data field is alphanumeric; and !I *n* if the data field is numeric.

III. Name of a Data File: A data file must always be specified after defining the data columns with complete path if it is not in the same directory as the command file. Its name must begin in the first position of the line. There are many qualifiers that can be placed on this line after specifying the name of the data file. The most commonly used are

!skip *n* indicates to skip over the first *n* lines (those containing column headings) in the data file

!maxit *m* establishes the maximum number of iterations in *m*. The default is 10 iterations.

IV. Linear Model: The linear model is a list of terms, each separated by a space in the form

<variable Y> ~ <model>

<variable Y> is the name of the data field which will be analyzed.

<model> lists the terms of the model. **<variable Y>** is separated from **<model>** using the symbol (~).

Some common model terms are:

mu	represents a constant term or the intercept
<name>	is the name of an explanatory variable or factor
<name>.<name>	is the interaction of two terms
!r	indicates that the following term is <i>random</i>
!f	indicates that the following term is <i>fixed</i>

mv if there are missing values in the response variable <variable Y>, **mv** needs to be placed among fixed terms i.e. before **!r** or after **!f**.

V. Variance structure: In a linear mixed effects models, there are two variance structures, one for the errors, known as R-Structure and one for the random effects known as G-structures. The default option is independently and identically distributed errors and random effects. For specifying a given structure, we use the variance header line as

For example:

```
Yield~ mu var !r rep
```

```
1 2 1 # Variance header line indicates that there is one R structure that involves two
      variance models and there is one G structure
```

```
col col AR1 0.1 # R-Structure indicates that there is auto-regressive structure to each
                of the dimensions row and column using 0.1 as initial correlation.
```

```
row row AR1 0.1
```

```
rep 1 # G-structure header line indicates one variance model.
```

```
rep 0 IDV 0.1 # Variance for the replicates is IDV of order 4  $\sigma_r^2 \mathbf{I}_4$ . 0.1 is a starting
              value for  $\gamma_r = \sigma_r^2 / \sigma_e^2$ .
```

Besides, the above, one can get the predicted values/BLUPs for the effects using ASReml. It gives BLUP directly but it does not do pair wise comparisons. For pairwise comparisons, there is an approximate test. Take the difference of two BLUPs, if difference is more than twice the SE of difference, then these two are significantly different.

Some examples of ASReml Command files are:

Spatial Analysis with Autoregressive error structure

```
Spatial Analysis without covariate for LTFE data # Title line
year # Data Filed Definitions
rep !I 4
trt !I 11
row !I 4
col !I 11
MzY wty ph oc sn sp sk
ltfe1981.csv !skip 1 !maxit 10 !mvinclde !Continue # Data file
MzY~mu trt rep # Model
predict trt # means for treatments
1 2 # R-structure
row row ar 0.1
col col ar 0.1

# !skip 1 skip one line from data
# !maxit maximum number of iterations is 10
# !r random effects
# !f fixed effects
# mv missing values
[End of File]
```

Combined Analysis of Data with Autoregressive structure in each environment

```

Combined Analysis of LTFE
yr !I 12
rep !I 4
trt !I 11
row !I 4
col !I 11
mzy wty ph oc sn sp sk
modipalampur19811992.csv !skip 1 !maxit 10 !continue !dense
mzy ~ mu trt !r yr rep.yr trt.yr
predicted trt
predicted yr
predicted trt yr
yr 2
4 row AR1 -0.037664 !S2==275283
11 col AR1 -0.315820
4 row AR1 -0.125133 !S2==538002
11 col AR1 -0.307042
4 row AR1 0.216599 !S2==555810
11 col AR1 -0.188775
4 row AR1 0.158081 !S2==353587
11 col AR1 0.444491
4 row AR1 -0.000232 !S2==406027
11 col AR1 -0.090648
4 row AR1 -0.0379119 !S2==621269
11 col AR1 -0.790458
4 row AR1 -0.0884483 !S2==384655
11 col AR1 -0.0133342
4 row AR1 0.375454 !S2==113983
11 col AR1 -0.0267661
4 row AR1 -0.0879242 !S2==121144
11 col AR1 0.342663
4 row AR1 -0.0675381 !S2==127423
11 col AR1 -0.163472
4 row AR1 0.177673 !S2==110175
11 col AR1 0.270297
4 row AR1 -0.0946895 !S2==98522.8
11 col AR1 -0.0117746
[End of File]

```

Analysis with Random Effects

```

NRCRM 2.5.19
loca !A 4
loc !I 4
trt !I 24
rep !I 3
yield
yldqha
c:\factor_analytic\dataset2519.csv !skip 1
yldqha ~ mu loca rep.loca !r loca.trt
predict loca
predict trt
predict loca.trt

```

ASReml: An Overview

```
0 0 1          #0 0 indicates no R structure and 1 as one G structure
loca.trt 2     # G structure is a direct product of two variance
               structure in loca.trt
loca 0 DIAG !GP # 0 denotes that the effects are in standard order
4*0.1
trt 0 ID

# !A variable is alphanumeric
# !GP attempts to keep the parameters in the theoretical parameter space.
-ve values of variances
    are replaced by small positive values
# Diag diagonal variance covariance matrix
[End of File]
```

For a detailed description on the use of ASReml one may refer to User Manual of ASReml. Some details can also be obtained in the following manual which can be downloaded from www.cimmyt.org/english/wps/biometrics.

Burgueno, J., A. Cadena, J. Crossa, M. Banziger, A.R. Gilmour and B. Cullis (2000). Users's Guide for Spatial Analysis of Field Variety Trials Using ASREML. Mexico, D.F.: CIMMYT.