ASReml tutorial

A0 Overview – Installation

Arthur Gilmour

NSW DEPARTMENT OF
PRIMARY INDUSTRIES

ASReml tutorial

A0 Overview – Installation

Arthur Gilmour

NSW DEPARTMENT OF
PRIMARY INDUSTRIES

Tutorial Overview

- Some 16 sessions
  – prepared by Arthur Gilmour
  – made available on VSNi.co.uk website
- Consisting of .mp3 audio files and slides
- covering basic ASReml use and syntax
  – and typical examples
- each session takes about 20 minutes

asrwin.exe

- The installation procedure uses InstallShield to set up the ASReml directory structure and file associations.
- Installs ASReml 2.00 as

  `C:\Program Files\ASReml2\bin\asreml.exe`
- Installs WinASReml,
  – an environment for running ASReml

Documentation

- The User Guide is supplied as a PDF file.
- A document describing the enhancements in release 2 over release 1 is supplied as a PDF file.
- Online help is supplied as HTML files.

Examples

- The examples are discussed in the User Guide.
- Suggestion: copy the Examples folder to a new folder (say `My Documents\ASReml\Exercises`) where you can modify it.

WinASReml or ConText

- WinASReml
  – a new environment for running ASReml – installed with ASReml – allows reviewing graph files – User must explicitly save required files
- ConText
  – is a popular third party text editor – well suited for running ASReml – User must externally delete superfluous files

Context

- Is an ideal editor to host the running of ASReml under Windows
- http://www.context.cx/
- Run `ContextSetup.exe`
  Copy `ASReml.chl` to
  `C:\Program Files\Context\Highlighters`

Context: Attach ASReml

- <Options> <Environment Options>
  <Execute keys>
  <Add> file extentions as, asr
  <F9>Execute
  `"C:\Program files\ASReml\bin\asreml.exe"`
  Start In %p, Parameters `%f`, Hint `Asreml`
  <Add> file extentions .pin
  <F9>Execute
  `"C:\Program files\ASReml\bin\asreml.exe"`
  Start In %p, Parameters `-p %f`, Hint `PIN file`

Context: Attach UserGuide

- <F11>Execute <ADOBE ACROBAT PATH> Start In %p, Parameters

  `"C:\Program files\ASReml\doc\UserGuide.pdf"`

- Find <ADOBE ACROBAT PATH> by right clicking the ADOBE READER icon and selecting <Properties>
- Similarly attach the Help file (say as <F12>)

Internet

- support@vsn-intl.co.uk
- http://www.asreml.co.uk/
- ASReml discussion group
  ASREML-L@dpi.nsw.gov.au
  To join, mailto:arthur.gilmour@dpi.nsw.gov.au
- Cookbook: `http://uncronopio.org/ASReml`

ASReml tutorial

A1 Getting Started with ASReml

Arthur Gilmour

NSW DEPARTMENT OF
PRIMARY INDUSTRIES

Why start?
Why am I here?

- You understand the principles of linear mixed models
- You want a flexible and comprehensive program to fit them.

The focus of this workshop

- Principles for using ASReml.
- These slides originally prepared with Animal Breeding focus
  We will adapt to plant breeding.
- Choosing appropriate variance models!

Sessions A1 to A5

- Mechanics of running ASReml
- Basic principles

Later Sessions

- Animal models
- Spatial models
- Multienvironment trials
- Multivariate Modelling
- Repeated measures
- Prediction and Testing

Getting ASReml

- http://www.asreml.co.uk
- Windows version has automatic 30day demo license.
- Other platforms (Linux, Sun Solaris, Opteron): license available on request. mailto:asreml@asreml.co.uk
- Workshop based on release 2.

Licensing

- VSN-International

Documentation

- User Guide (Release 2.00)
- ASReml update
  has summary of changes from 1 to release 2
- Help file
- Examples
- Cookbook

Installation (Windows)

- InstallShield
- `C:\Program Files\ASReml2`
  `bin` contains program
  `doc` contains pdf manual
  `examples` contains examples

A batch process

- ASReml typically runs as a batch process – with minimal user interaction
- Under Windows, run from
  – Windows Explorer - click on job file
  – Command Prompt (DOS Box) - type command – WinASReml (replaces ’Menu Mode’ of release 1) – ConText
- Under Unix, run from
  – command prompt – WinASReml

How it works.

- Identify problem
- Collect and organise data: save as ASCII (.txt, .csv, .asd) file
- Prepare .as job file (Notepad, ConText, TextEdit, vi, emacs)
- Run, Review, Revise, Rerun cycle
- Report

Zinc example

- View Data
- Fit oneway analysis of variance to SeedZn
- View running
- View output

Zinc data

- First few lines of `ZINC.DAT` are

```
Source SeedZn LeafZn
1 61 24.1
1 63 23.8
2 51 16.0
2 64 19.0
6 69 22.6
```

Zinc.as file

- Zinc concentration study

Zinc.as file

- Zinc concentration study
  Source * SeedZn LeafZn
- cf

```
Source SeedZn LeafZn
1 61 24.1
1 63 23.8
2 51 16.0
2 64 19.0
6 69 22.6
```

Zinc.as file

- Zinc concentration study
  Source * SeedZn LeafZn
  ZINC.DAT !Skip 1

Zinc.as file

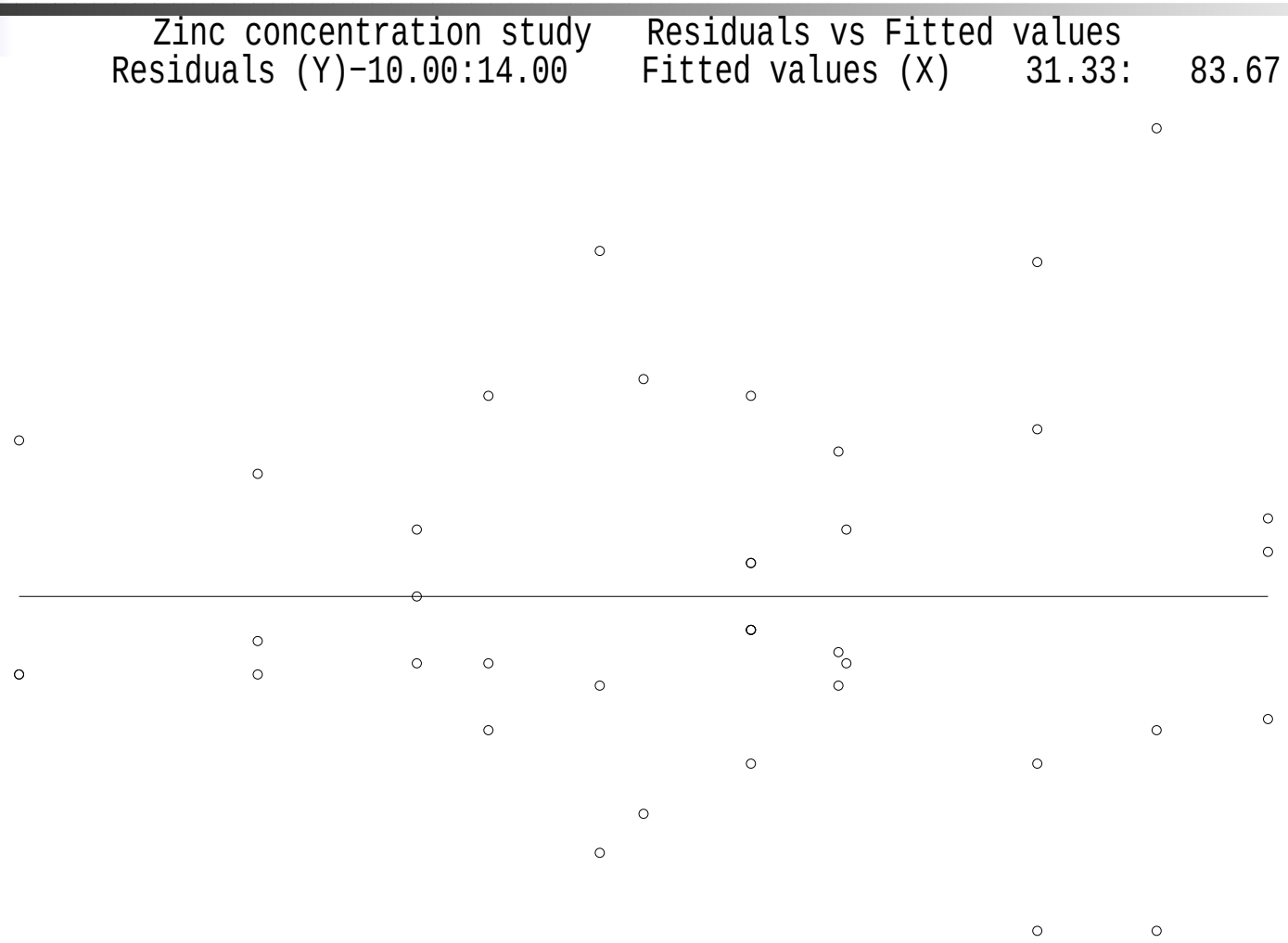
- Zinc concentration study
  Source * SeedZn LeafZn
  ZINC.DAT !Skip 1
  SeedZn ~ mu Source

Zinc.as file

- ```
  Zinc concentration study
   Source * SeedZn LeafZn
  ZINC.DAT !Skip 1
  SeedZn ~ mu Source
  ```
- Run the job from command prompt, Explorer or ConText.

  ```
  "C:\Program Files\ASReml\bin\ASReml.exe" ZINC
  ```
  ```

Automatic plot of residuals



View zinc.asr

```
ASReml 1.63i [07 Jan 2005]  Zinc concentration
17 Jan 2005 12:23:06.591   32.00 Mbyte Windows
This Beta version is valid for 5 months
Please send comments to asreml@VSN-intl.co.uk
Licensed to: Arthur Gilmour
************************************************
* SYNTAX change:  A/B  now means  A A.B        *
*                                              *
* Contact asreml@vsn-intl.com for support      *
************************************** ARG *
```

Data summary

- Folder: C:\data\Prosper\Armidale2005
  QUALIFIERS: !SKIP 1
  Reading ZINC.DAT FREE FORMAT skipping 1 lines

  Univariate analysis of SeedZn
  Using 39 records of 39 read

Model term		Size	#miss	#zero	MinNon0	Mean	MaxNon0
1	Source	24	0	0	1	11.9487	24
2	SeedZn	Variate	0	0	29.00	60.10	93.00
3	LeafZn		0	0	10.20	19.84	35.40
4	mu	1					

Iteration sequence

- ```
  Forming  25 equations:  25 dense.
  Initial updates will be shrunk by factor 0.316
  NOTICE:  10 singularities detected.
    1 LogL=-64.4770  S2=  44.160    24 df  1.000
    2 LogL=-64.4770  S2=  44.160    24 df  1.000
  Final parameter values                 1.000
  ```
  ```

Result summary

- Degrees of Freedom and Stratum Variances

```
          24.00    44.1597          1.0
Source  Model terms Gamma Component Comp/SE % C
Variance   39    241.0000   44.1597    3.46 0 P
Analysis of Var NumDF  DenDF   F-incr    Prob
  4 mu            1    24.0  3190.25   <.001
  1 Source       14    24.0    12.94   <.001
SLOPE for LOG(ABS(RES)) on LOG(PV) in Section 1
  0.50
Finished: 17 Jan 2005 12:23:08   LogL Converged
```

What's Next

- Find the `ZINC.DAT` file,
  type up the `ZINC.AS` file and run it.
  Examine the output files `zinc.asr`
  `zinc.res` `zinc.sln` `zinc.yht`.
- Four more introductory sessions:
  `A2Overview`: command line, data definition
  `A3files`: the data file
  `A4model`: model specification
  `A5exercises`: some homework.

ASReml tutorial

A2 Running ASReml

Arthur Gilmour

NSW DEPARTMENT OF
PRIMARY INDUSTRIES

A Structured .as file

- First part defines the data
- Second part defines the analysis
- A minimal job has 4 lines
- Many jobs have over 20 lines
- File is built up in stages

Definition part

- [Job qualifiers] (command line options)
- Job Title
- Data Definition
- [Pedigree and GIV Files]
- Data file name and qualifiers

```
Zinc data analysis
 Source * SeedZn LeafZn
ZINC.DAT !SKIP 1
```

Analysis part

- [Analysis qualifiers]
- [TABULATE]
- Model line
  ```
  SeedZn ~ mu Source
  ```
- [PREDICT]
- [Variance structures]
- [Component constraints]

Job qualifiers

- command line options
  `ASReml -<options> <jobname> <argumen`
- command line not easily modified under Windows
- First line of job
- Recognised by `!` character
- `!-<options> <arguments>` or `<qualifiers>`

Qualifier SYNTAX

- First character is !
- Three letters sufficient
- separate from arguments with a space
- Context specific
- Examples

```
!SKIP 1
!CONTINUE !EPS !WORKSPACE 512
```

Common Job qualifiers

- !`CONTINUE` – Use parameter estimates from a previous run as starting values
- !`FINAL` – One more iteration
- !`LOGFILE` – write .asl file
- !`DEBUG` – extra output including timing

Common Job qualifiers

- !EPS – graphics to .eps file
- !HARDCOPY – do not write graphics to screen
- !NOGRAPHICS – do not create graphics at all
- !WORKSPACE 512 – use 512 Mb workspace
- !RENAME r – rerun job with arguments
- !ARG $1 \ldots$ – job arguments

Job control continued

- P (Pinfile mode) must be specified from the command line
  `ASReml -Pmyjob mypin`
- Command line options and arguments override qualifiers and arguments on the job control line.

Arguments

- A way of resetting options within a job
- Are inserted into the job where $\$n$ appears:
  $1 is replaced by the first argument
  $2 is replaced by the second argument
- With `!RENAME` n, the first n arguments are built into the output filename, and, the job is run repeatedly after moving up any arguments after the nth

```
ASReml -r2 job alpha beta gamma
ASReml -r2 job alpha beta
ASReml -r2 job alpha gamma
```

Arguments continued

- The command line options and qualifiers in `ASReml -r2 job alpha beta gamma` may be given on the top line as
  `!-r2 alpha beta gamma`
  or as
  `!rename !arg alpha beta gamma`

Exercise

- Rerun the zinc example with
  `!DEBUG !LOGFILE`
  - look at the `zinc.asl` file
  `!FINAL`
  - see it does only one iteration
  `!RENAME !ARG aaa`
  - see output files are `zincaaa.*`
  `!EPS !HARDCOPY`
  - see no graphics displayed but `zinc*.eps` is produced

ASReml tutorial

A3 Data definition

Arthur Gilmour

NSW DEPARTMENT OF
PRIMARY INDUSTRIES

A Structured .as file

- First part defines the data
- Second part defines the analysis

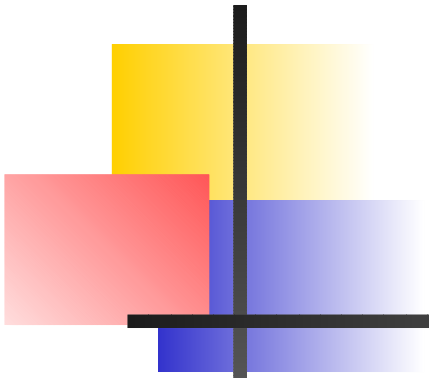
Definition part

- [Job qualifiers] (command line options)
- Job Title
- Data Definition
- [Pedigree and GIV Files]
- Data file name and qualifiers

```
Zinc data analysis
 Source * SeedZn LeafZn
ZINC.DAT !SKIP 1
```

Qualifier SYNTAX

- First character is !
- Three letters sufficient
- separate from arguments with a space
- Context specific



Title Line

- identifies the job
- must be present
- must not contain any qualifier

Comments

- On all lines, characters following # are stripped out
- Comment lines (a ! in column 1 followed by a space) are copied to the output file.
- Line length is 2000 characters
- Reserved characters: #, !, $

Data definition

- Controls reading the data file and how the data fields are used in the analysis.
- definitions should appear in the order of the data in the file
- definition lines should be indented
- transformations may alter the fields; the label will apply to the transformed field.
- all data is held as real numbers.

Basic definitions

- `label` (co)variate
- `label *` simple factor coded 1 2 ...
- `label !A [`n`]` alphabeticly coded factor
- `label !I [`n`]` numerically coded factor
- `label !G` n group of n variates
- `label !P` pedigree factor
- `label !L list` simple factor; levels named
  `SEX !L male female` 1 codes for male

Zinc example

- `SeedZn` is interpreted as a (co)variate
- `Source *` generates 24 levels 1:2, 5:8, 11:14, 17:19 21, 24
- `Source 24` generates 24 levels 1:2, 5:8, 11:14, 17:19 21, 24

Recoding levels

- `Source !I` generates 15 levels labelled 1:2, 5:8, 11:14, 17:19 21, 24
- `Source !A` generates 15 levels labelled 1:2, 5:8, 11:14, 17:19 21, 24
- Specify an indication of how many levels are expected after `!I` and `!A` if there are many (> 1000) levels

Alpha qualifiers

- !LL n – sets character length of alphabetic labels
- !SORT – puts labels in alphabetic/numeric order (current term)
- !SORTALL – puts labels in alphabetic/numeric order (current and subsequent terms)
- !SKIP s – to skip s fields

Transformations

- ASReml can transform the data as it is read in.
- applied in order of definition.
- record vector has 1000 cells
- have a elemental syntax based on qualifiers
- `yield !*100`

Arithmetic operators

- `!+`o `!-`o `!*`o `!/`o `!^`p
  o is a number, Vc or f where c is a cell position number, f is the name of an earlier column.
- `Palive !/Total # change count to proportion`

Missing value operators

- !D v – discards records with MV or v in current field
- !M v – converts values of v to missing values in current field
- !NA v – converts missing values to the value v

Exercise

- Modify the zinc job to calculate the ratio of SeedZn to LeafZn
- Modify the declaration for `Source` to use `!I`
- Review the information on transformations in the help file.

ASReml tutorial

A4 Pedigree, giv and data files

Arthur Gilmour

NSW DEPARTMENT OF
PRIMARY INDUSTRIES

Pedigree file

- Expected if `!P` data qualifier specified
- contains ID SIREID DAMID
- in birth order (parents before progeny (see !SORT))
- !ALPHA !SKIP k !DIAG !GIV !INBRED !MGS !REPEAT !SELF !SORT
- e.g.
  `mydata.ped !skip 1 !diag`

GIV files

- Generalized InVerse
- Has file extension `.giv` (`.grm` if not inverted)
- Lower triangle rowwise sparse format
  row column value
- !SKIP s

DataFile Line

- names the data file (enclosed in quotes if embedded blanks)
- Data file is typically an ASCII file
  TAB, SPACE or COMMA separated
  e.g. save from Excel as comma separated
- Missing values: *, . and NA and empty fields in .csv file are taken as missing
- !SKIP s !FILTER f !SELECT v !SUMMARY
- `zinc.dat !skip 1 !SUM`

DataFile Line Qualifiers

- Some 40 qualifiers are defined for this line or to immediately follow this line.
  !MAXIT m !EPS !CONTINUE
  !X x !Y y !JOIN !G g
  !CONTRAST t f coefficients
  !PVAL f points
  !SPLINE t points

TABULATE directive Raw tables of means

- y ~ factors
- Qualifiers
  !COUNT – numbers in each cell
  !SD – Standard deviation in each cell
  !RANGE – of values in each cell
  !STATS – same as !COUNT !RANGE !SD
- before model (or after) model line
- `TABULATE Leaf Seed ~ Source !STATS`
- Multiple statements allowed

Data exercises

- Add `!SUM` qualifier to the data line; explore `zinc.ass`
- Insert two plotting lines and view graphs

```
!X SeedZn !Y SeedZn
!X SeedZn !Y SeedZn !G Source
```

- Insert Tabulate line and view `zinc.tab` file

```
TABULATE SeedZn LeafZn ~ Source !STATS
```

ASReml tutorial

A5 Model line

Arthur Gilmour

NSW DEPARTMENT OF
PRIMARY INDUSTRIES

Model line

- Univariate
  y ~ <fixed dense>,
  !r <random sparse>,
  !f <fixed sparse>
- y is response variable
  <fixed dense> terms appear in ANOVA table
  <random sparse> and <fixed sparse> are reordered to maximize sparsity during solution.

Model terms

- Reserved terms
  `mu` – constant term
  `mv` – missing value estimates
  `units` – extra residual
- Data terms e.g. `A B X Sex Treatment`
- Functions of terms
  at(Group,1) spl(X,10) fac(X)
  log(X,1) forms log(X+1)

Model terms continued

- Combinations
  `A.B Sex.spl(X,5) at(site,3,5).row`
- Shorthand
  `A*B – A B A.B`
  `A/B – A A.B`
- Continuation of a model line is indicated by a trailing comma

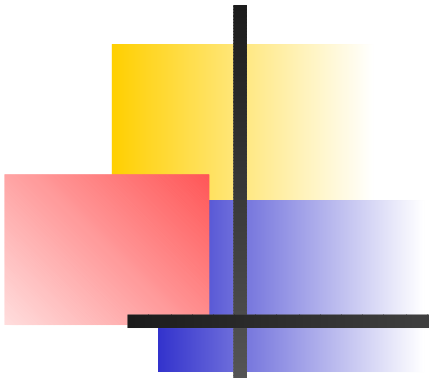
Random terms

- May be followed by an initial value for the variance component and a qualifier.
- Default initial value is 0.1
- `blocks 0.2 !GU`

  `!GP` – force positive (default)

  `!GU` – unrestricted

  `!GF` – fixed



PREDICT

- Multiple predict statements `predict A`
- More details later
- Variance structure lines
- See later

Order of processing

- Read .as file down to model
- Read data file
- Produce data summaries
- Do Plots and tabulations
- Read Variance structure lines
- Fit model
- Report results

GIGO

- Check ASReml has read the data correctly
  1. Number of records read/retained
  2. Mean and range of variables
  3. Distribution of data

     !SUM summary

     !X !Y plot

     TABULATE

ASReml tutorial

A6 Exercises

Arthur Gilmour

NSW DEPARTMENT OF
PRIMARY INDUSTRIES

Zinc data

- Use !SUM to explore data structure
- Define Source using !I
- Use !X !Y to plot SeedZn vs LeafZn
- Use sqrt() to transform to Square roots
- Fit Leaf ~ mu Seed !r Source

Explore the online help

- `ASReml.chm`

Volts data

- User Guide 15.3
- Identify outliers - assess effect of dropping two.

Oats data

- User Guide 15.1
- Split Plot design: Blocks|Variety|Nitrogen
- Use !CONTRAST to test for linear N trend

Rats data

- User Guide 15.2

Own data

- Prepare job to read and summarize the data