ASReml workshop

3.1 Variance structures

Arthur Gilmour

NSW DEPARTMENT OF
PRIMARY INDUSTRIES

Overview

- Traditional variance models assume independent effects: $\sigma^2 \boldsymbol{I}$
- General variance structures
  - Unstructured - every variance and covariance is a separate parameter
  - Structured - variances and covariances are functions of parameters
- Spatial models
  - correlation based on distance
  - paramerterized in terms of correlation and variance

Overview

- Traditional variance models
- General variance structures
  - Unstructured- Structured
- Spatial models
  - correlation based on distance
  - parameterized in terms of correlation and variance
- Compound variance structures
  - formed as a direct product

General Variance structures

- Unstructured (US) is parameterised directly as variances and covariances
- Symmetric Lower triangle rowwise

$$\begin{matrix} V_{11} & & \\ V_{21} & V_{22} & \\ V_{31} & V_{32} & V_{33} \end{matrix}$$

Reduced parameterization

- Diagonal (DIAG) has zero covariances
  Factor Analytic (FACV, XFA): $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}$
  Cholesky (CHOLn, CHOLnC): $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}'$
  where $\boldsymbol{L}$ is unit lower triangle
  Antedependence (ANTEn): $\boldsymbol{\Sigma}^{-1} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}'$
  where $\boldsymbol{U}$ is unit lower triangle

Reduced parameterization

- Aim in using alternate forms is
  –to accomodate the variance heterogeneity adequately while minimising the number of parameters
  –force a positive definite structure.
- ANTE (a generalization of AR) is suited to ordered levels (e.g. times)
- CHOL, XFA, FACV are suited to unordered levels (e.g. sites, traits)

General variance structures

- DIAG - off diagonal is zero
- CHOLi - $\Sigma = \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}'$
  - $\boldsymbol{L}$ is lower triangle unit matrix with i off-diagonal bands
  - $\boldsymbol{D}$ is diagonal matrix of conditional variances.

CHOL1 of order 4

- e.g. in CHOL1 $\boldsymbol{L} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ a & 1 & 0 & 0 \\ 0 & b & 1 & 0 \\ 0 & 0 & c & 1 \end{pmatrix}$

$\boldsymbol{D} = \text{diag}(A \;\; B \;\; C \;\; D)$ so that

$$\boldsymbol{\Sigma} = \begin{pmatrix} A & aA & 0 & 0 \\ aA & aAa + B & bB & 0 \\ 0 & bB & bBb + C & cC \\ 0 & 0 & cC & cCc + D \end{pmatrix}$$

CHOL1C of order 4

- e.g. in CHOL1C $\boldsymbol{L} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ a & 1 & 0 & 0 \\ b & 0 & 1 & 0 \\ c & 0 & 0 & 1 \end{pmatrix}$

$\boldsymbol{D} = \text{diag}(A \;\; B \;\; C \;\; D)$ so that

$$\boldsymbol{\Sigma} = \begin{pmatrix} A & aA & bA & cA \\ aA & aAa + B & bAa & cAa \\ bA & bAa & bAb + C & bAc \\ cA & aAc & cAb & cAc + D \end{pmatrix}$$

Antedependence

- is a generalized form of Autoregressive
- ANTEi - $\Sigma^{-1} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}'$
  - $\boldsymbol{U}$ is upper triangle unit matrix with i off-diagonal bands
  - $\boldsymbol{D}$ is diagonal matrix of conditional inverse variances.
- Since parameterization is obtuse for CHOL and ANTE, you may supply an unstructured matrix as starting values and ASReml will factorize it.

Factor Analytic

- Correlation Form: FAi
  $\Sigma = \boldsymbol{D}(\boldsymbol{L}\boldsymbol{L}' + \boldsymbol{E})\boldsymbol{D}'$
  Parameters are elements of $p \times i$ matrix $\boldsymbol{L}$ and diag($\boldsymbol{\Sigma}$) $= \boldsymbol{D}\boldsymbol{D}$; $\boldsymbol{E}$ is defined such that diag($\boldsymbol{L}\boldsymbol{L}' + \boldsymbol{E}$) is Identity.
- Variance Form: FACVi
  $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}$
  Paramaters are $\boldsymbol{\Lambda} = \boldsymbol{D}\boldsymbol{L}$ and $\boldsymbol{\Psi} = \boldsymbol{D}\boldsymbol{E}\boldsymbol{D}$

Extended Factor Analytic

- Same parameterization as FACV but in order (Ψ) vec$(\mathbf{\Lambda})$
- Elements of $\mathbf{\Psi}$ may be zero (making $\mathbf{\Sigma}$ singular)
- Requires use of `xfa`(T, i) model term which inserts i columns of zeros into the design matrix corresponding to the i factors.
- Much faster than FAi and FACVi when more than 10 levels in term.

Extended Factor Analytic

- ... xfa(Trait,1).dam ...

```
 xfa(Trait,1).dam 2
 xfa(Trait,1) 0 XFA1
2*0
1.1 0.9
dam
 Covariance/Variance/Correlation Mat
  1.550       1.000       1.000
  1.437       1.332       1.000
  1.245       1.154       1.000
```

Other structures

- US - unstructured
- OWNi - user supplies program to calculate G and the derivatives of G
- AINV - Use fixed relationship matrix
- GIVi - Use user defined fixed relationship matrix (see .giv, .grm)

Spatial structures

- ID - Identity
  CORU - uniform correlation
  AR1 $1\ \rho\ \rho^2\ \rho^3\ \rho^4\ \rho^5 \ldots$
  AR2, MA1, MA2, ARMA, SAR1, SAR2,
  CORU, CORB, CORH
  EXP, GAU
  IEXP, AEXP, IGAU, AGAU, IEUC, LVR, ISP,
  SPH, MAT
  one or two dimensional distance

Variances

- Equal variance correlation
  append V to code e.g. AR1V, CORUV
- Unequal (Heterogeneous) variance correlation
  append H to code e.g. AR1H, CORUH
- If D is the diagonal matrix of variances, and C is a correlation matrix, $\Sigma = D^{0.5} C D^{0.5}$

ASReml workshop

3.2 Spatial Analysis

Arthur Gilmour

NSW DEPARTMENT OF
PRIMARY INDUSTRIES

Two basic kinds

- Regular grid e.g. field trial
  - interest is in adjusting for other effects

Two basic kinds

- Regular grid e.g. field trial
  - interest is in adjusting for other effects
- Irregular grid e.g. survey
  - interest is in modelling the spatial pattern
  - kriging

Two basic kinds

- Regular grid e.g. field trial
  - interest is in adjusting for other effects
- Irregular grid e.g. survey
  - interest is in modelling the spatial pattern
  - kriging
- ASReml is regularly used for former
  - developing capability for latter

Single field trial

- Slate Hall Farm - Barley 1976
  - Balanced Incomplete block design
  - 25 varieties, 6 replicates
  - layout 10 rows by 15 columns
- BIB Model
  fixed: treatments
  random: rep block
- Spatial Model
  Autoregressive error model $\boldsymbol{R} = \boldsymbol{\Sigma}_R \otimes \boldsymbol{\Sigma}_C$

Slate Hall base

- Slate Hall 1976 Cereal trial
  rep 6 latrow 30 latcol 30
  fldrow 10 fldcol 15
  variety 25
  yield !/100
  shf.dat !DOPART $1
  !DISPLAY 15 !SPATIAL !TWOWAY

Slate Hall - Design based

```
!PART 1   RCB Analysis
yield ~ mu var !r rep

!PART 2   # BIB analysis
yield ~ mu var !r rep latrow latcol
```

Slate Hall - Model based

```
!PART 3 #   Fitting AR1.AR1
yield ~ mu var
predict var
1 2
fldrow fldrow AR .1
fldcol fldcol AR .1
```

Slate Hall - Model + Design

- !PART 4 # Fitting AR1.AR1
  yield ~ mu var !r rep latrow latcol
  predict var
  1 2
  fldrow fldrow AR .1
  fldcol fldcol AR .1

Slate Hall - summary

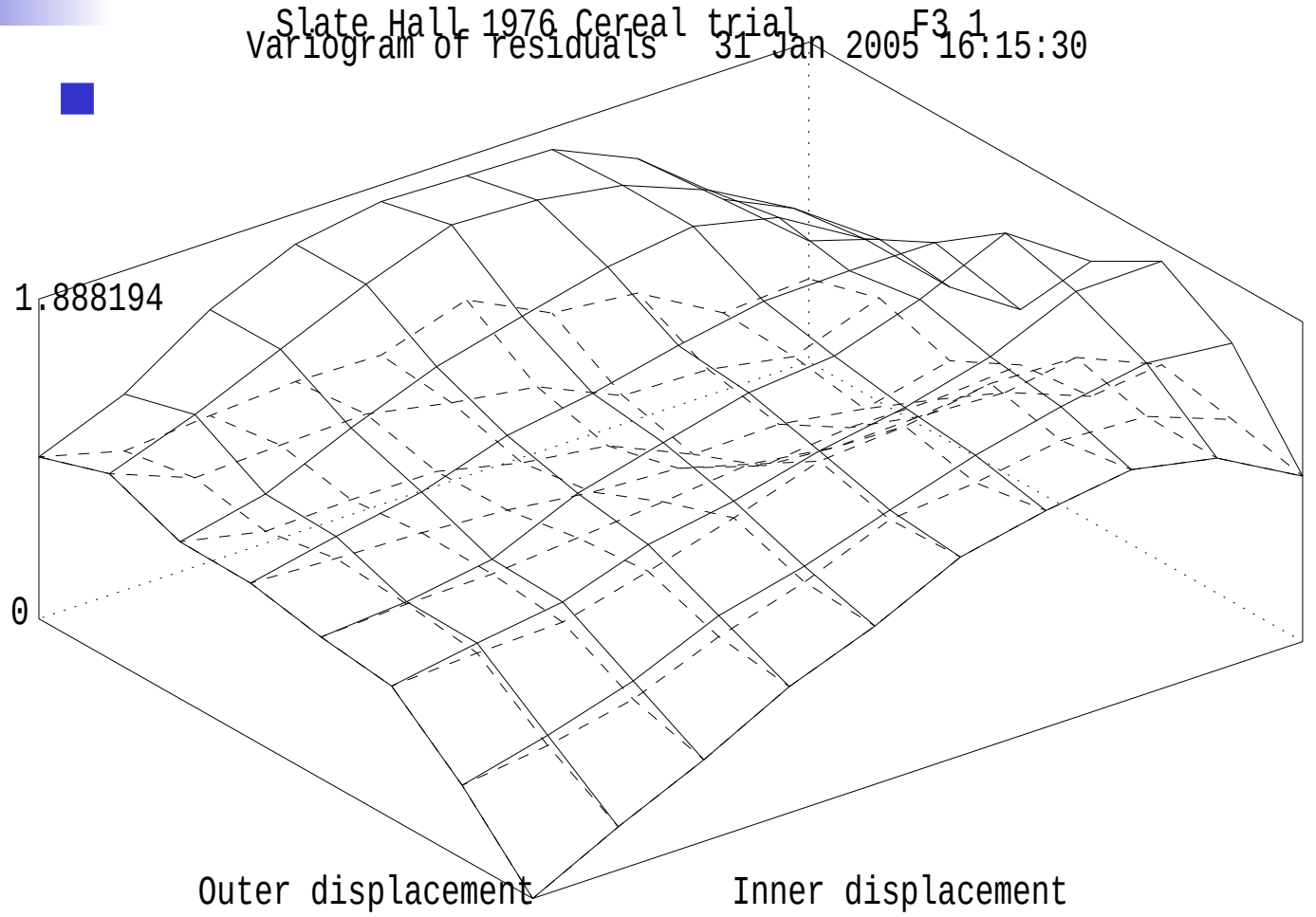
Model	LogL(l)	$-2\Delta(l)$
RCB	-167.694	2
BIB design	-132.134	4
Spatial model	-124.676	3
BIB+Spatial	-124.312	6

- Spatial correlation model fits better than the BIB model

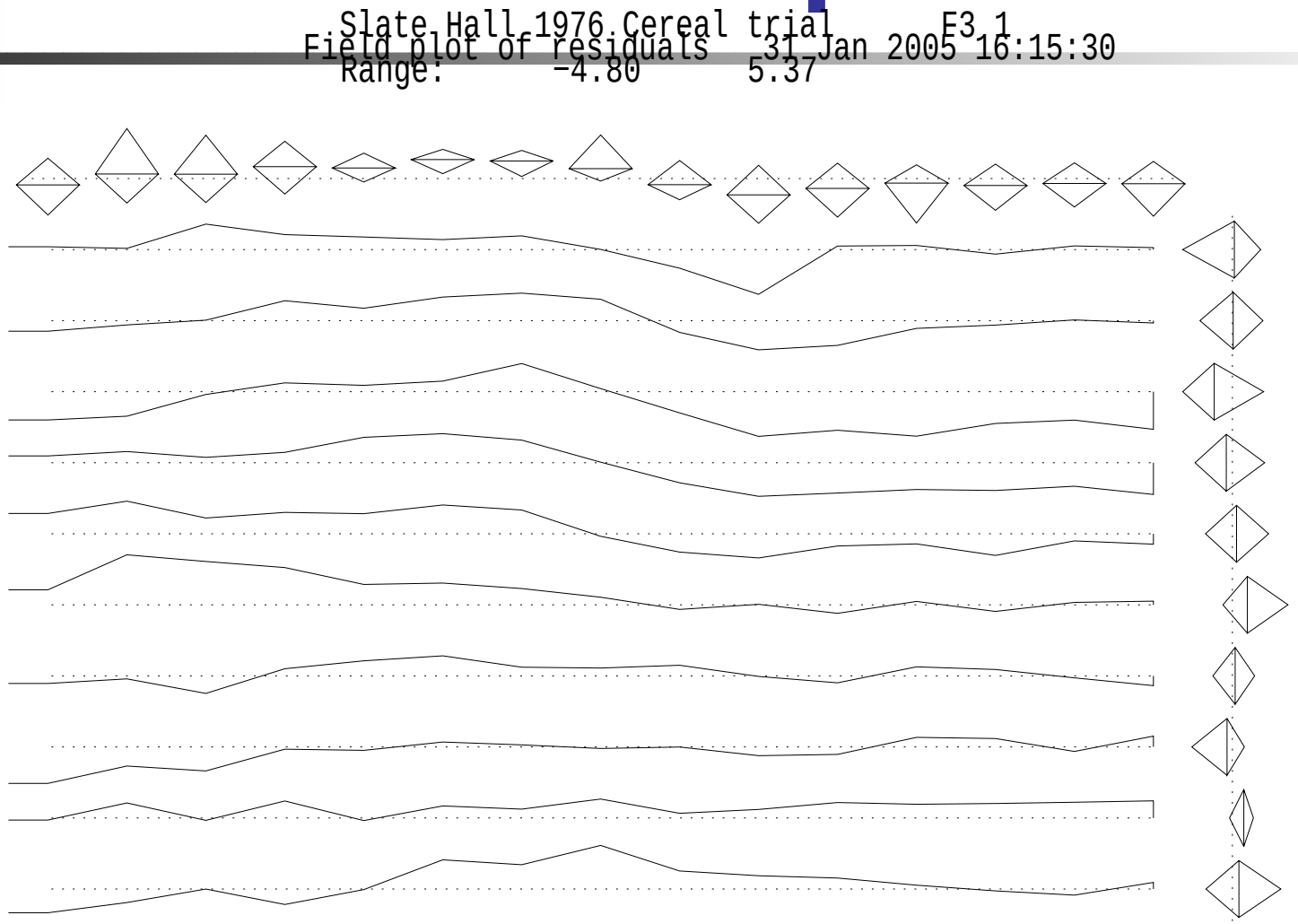
Spatial components

Source	terms		Gamma	Component	Comp/SE	%
rep	6	6	.2003E-05	.724166E-05	0.00	0
latrow	30	30	.6327E-01	.228684	0.71	0
latcol	30	30	.1608E-03	.581362E-03	0.00	0
Variance	150	125	1.000	3.61464	4.28	0
Residual AutoR	10		.4652	.465209	4.85	0
Residual AutoR	15		.6741	.674095	8.76	0

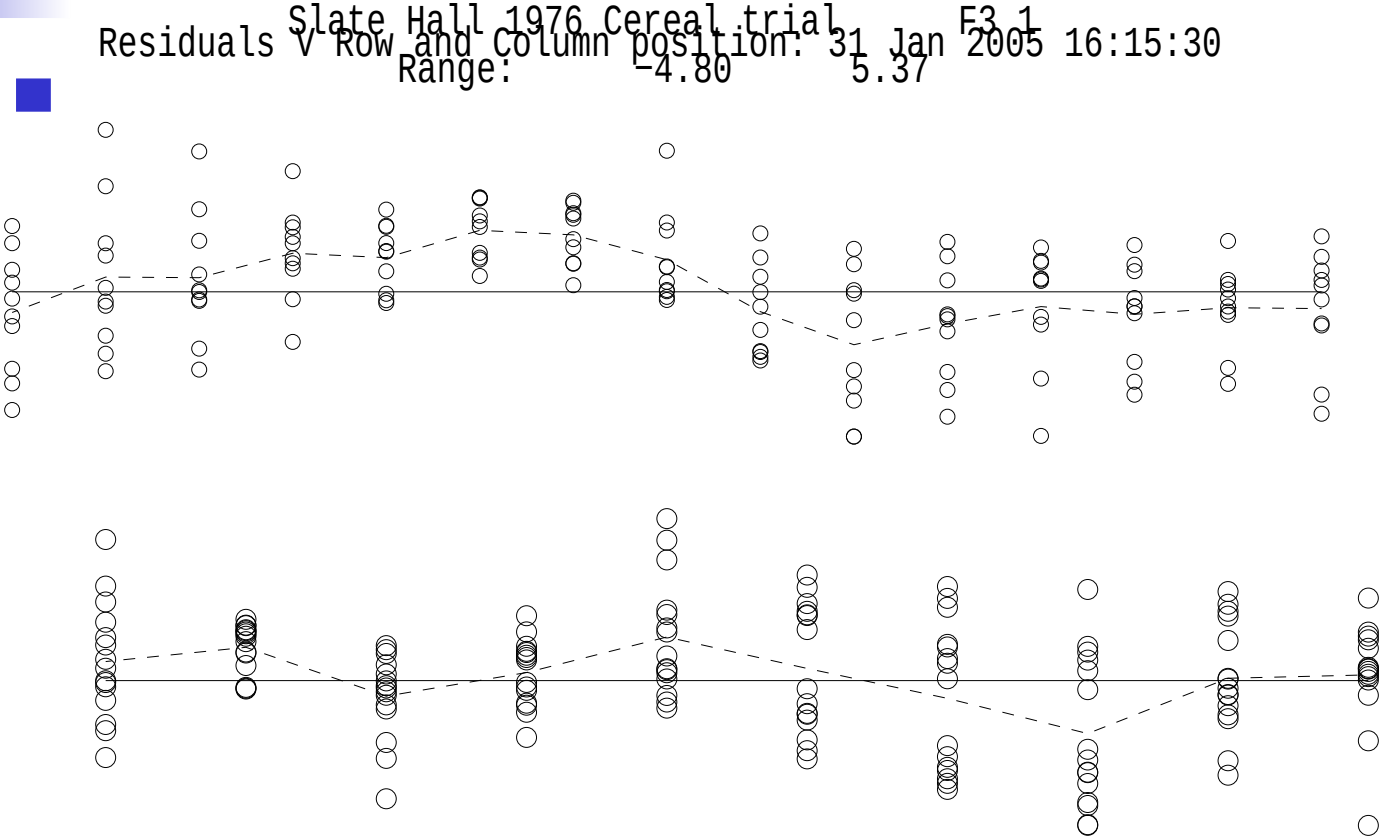
Variogram


Slate Hall 1976 Cereal trial F3 1
Variogram of residuals 31 Jan 2005 16:15:30
1.888194
0
Outer displacement
Inner displacement

Residual to plan



row/column


Slate Hall 1976 Cereal trial F3.1
Residuals V Row and Column position: 31 Jan 2005 16:15:30
Range: -4.80 5.37

Multi environment trial

- In early generational cereal breeding, run several trials with 1 or two replicates of test lines, 20 percent check lines for error estimation.
- More power from fitting as correlated effects across sites.

MET in ASReml

- Three Multi Environment Trial

```
seq
col 15      # Actually 12 12 and 15 respectivel
row 34      # Actually 34 34 and 28 respectivel
chks 7      # Check 7 is the test lines
test 336    # coded 0 for check lines
geno 337
yld    !*.01
site 3
met.dat    !section site
```

Spatial models

- yld ~ site chk.site !r at(site,3).row .02,
  at(site).col .90 .40 .036 site.test
  site 2 1
  12 col AR .1271 !S2=2.19
  34 row AR .751
  12 col AR .25 !S2=0.84
  34 row AR .56
  15 col ID !S2=0.19
  28 row AR .38

Model genetic variation

- site.test 2
  site 0 FA1
  .5 .5 .5
  .1 .1 .1
  test

Components

Source	Model	terms	Component	Comp/SE	%
Residual	1236	1213			
at(site,01).col	15	15	0.323302E-05	0.00	0
at(site,02).col	15	15	0.142114	1.32	0
at(site,03).col	15	15	0.446791E-01	1.77	0
at(site,3).row	34	34	0.241380E-01	2.80	0
Variance[1]	408	0	2.60271	5.18	0
Residual	AR=AutoR	12	0.407051	4.45	0
Residual	AR=AutoR	34	0.882580	33.50	0
Variance[2]	408	0	1.00339	8.29	0
Residual	AR=AutoR	12	0.282407	4.84	0
Residual	AR=AutoR	34	0.580701	11.37	0
Variance[3]	420	0	0.105411	5.59	0
Residual	AR=AutoR	28	0.687455	10.14	0

Factor Analytic

```
site.test  FA D(L   1   1  0.518516      5.35  0
site.test  FA D(L   1   2   1.13028      2.18  0
site.test  FA D(L   1   3  0.735010      6.04  0
site.test  FA D(L   0   1  0.991585      7.99  0
site.test  FA D(L   0   2  0.731805E-01 1.07  0
site.test  FA D(L   0   3  0.121810      7.17  0
Covariance/Variance/Correlation  FA D(LL'+E)D
0.9916     0.5865     0.3811
0.1579     0.7308E-01 0.8313
0.1325     0.7844E-01 0.1218
```

Spatial analysis in Forest Genetic trials.

- Typically not a complete rectangle
  - add missing values to complete the pattern
  - use map points (if < 5000 trees)
- With Tree model, must include Nugget variance
  - either Nugget is residual, spatial is in G
    or spatial is residual and Nugget is G,
- spatial model typically superior to 'design' model for growth/production traits
  - less so for disease and conformation traits

MicroArray

- spatial pattern

ASReml workshop

3.3 Repeated Measures

Arthur Gilmour

NSW DEPARTMENT OF
PRIMARY INDUSTRIES

Main approaches

- General variance structure
  (Multivariate approach)
  UnStructured, Autoregressive, EXPponential
- Repeated measures
  Longitudinal model
  Repeated measures.

Multivariate approach

- Suited when most animals have most measures
- Repeats are at significant standard times
  Say WWT, 200dayWT, 400dayWT, 600dayWT
- Discuss

Multivariate

- WWT WT200 WT400 WT600 ~ Trait Tr.sex
  !r Tr.animal !f Tr.cohort

```
1 2 1
0
Trait 0 US
10*0
Tr.animal 2
Tr 0 US
10*0
animal 0 AINV
```

Multivariate

- WWT WT200 WT400 WT600 ~ Trait Tr.sex,
  !r Tr.animal !f Tr.cohort

```
1 2 1
0
Trait 0 US
10*0

Tr.animal 2
Tr 0 US
10*0
animal 0 AINV
```

Multivariate

- WWT WT200 WT400 WT600 ~ Trait Tr.sex,
  !r Tr.animal !f Tr.cohort
  1 2 1
  0
  Trait 0 US
  10*0

  Tr.animal 2
  Tr 0 US
  10*0
  animal 0 AINV

Random Regression

- Appropriate when
  - there is considerable unbalance in times of measurement
  - there are varying numbers of measurements
  - all animals have multiple measures
- Concept: Regression for each individual consisting of an overall response pattern (fixed) plus an individual (random) adjustment.

RR principles

- This is a reduced parameterization model which must be well formulated
  - mean profile of higher order than random profile - random profile generally low order
- Usually formulated as polynomial but could be low order spline

RR Example

```
!WORK 150
This is random regression analysis o
  animal  !P  sire 89 !I   dam 1052
  year 2 !I !V21=V4 !==2 !*-365
  flock 5     sex 2 !A     aod
  tobr 3 !I   dob !-14800 !+V21
  age   wt    fat  emd

sdf01a.ped !SKIP 1
sdfwfml.csv !SKIP 1  !MVremove !DOPA
   !DDF !TYPEIIISS !MAXIT 20
```

RR Model

```
!PART 1  #  Linear RR
emd  ~ mu age year wt sex sex.wt flock,
     tobr aod dob year.dob year.age,
     year.sex year.flock year.tobr,
     sex.dob tobr.dob,
   !r animal animal.age,
      ide(animal) ide(animal).age,
      at(year,1,2).spl(age,20)
```

RR G structure

```
0 0 2
animal 2
2 0 US !GP  # Intercept and slope
1.3 0.01 0.01
animal 0 AINV
ide(animal) 2# Intercept and slope
2 0 US !GP
1.6 0.01 0.03
ide(animal)
```

Fitting PART 1

- Fixed terms year.age, year.sex year.tobr are NS
- variance of ide(animal).age is at boundary
- LogL after dropping 3 interactions was -726.867

Quadratic RR

```
!PART 2  # Quadratic RR using pol
emd  ~ mu age year wt sex sex.wt flo
   dob year.dob   year.flock sex.dob
   !r pol(age,2).animal  pol(age,1).
   at(year,1,2).spl(age,20)
0 0 2
pol(age,2).animal 2
3 0 US
1.6  .6 .6  .3 .3 .3
animal 0 AINV
pol(age,1).ide(animal)
2 0 US
```

PART 2 G structures

- 0 0 2
  pol(age,2).animal 2
  3 0 US
  1.6 .6 .6 .3 .3 .3
  animal 0 AINV
  pol(age,1).ide(animal)
  2 0 US
  2.1 .6 1.3
  ide(animal)

PART 2

- LogL -643.67 so significant quadratic curvature
- Obtained inital values by ignoring G structure in initial run.

Spline curvature

```
!PART 3
!SPLINE spl(age,3) 4 0 6
emd  ~ mu age year wt sex sex.wt flo
     year.dob year.age year.sex year.
 !r animal animal.age animal.spl(age
      ide(animal) ide(animal).age,
      ide(animal).spl(age,3),
      at(year,1,2).spl(age,20)
0 0 2
animal 2
3 0 US !GU  # Icept,slope,spl
1.3 0.1 0.01
```

Simpler

- !PART 4

```
emd ~ mu age year wt sex sex.wt flock tobr
  year.dob year.age -year.sex year.flock ye
  !r pol(age,2).animal  ide(animal) ,
  at(year,1).spl(age,20) at(year,2).spl(age

0 0 1
pol(age,2).animal 2
3 0 US
1.6
.6 .6
 3  3  3
```

Interpretation

- .res file has pol() coefficients. say T
  Form TGT' to get full matrix of variances (all times).