

# Employing `asremlPlus`, in conjunction with `asreml`, to calculate and use information criteria

Chris Brien

11 July, 2020

This vignette illustrates the facilities in `asremlPlus` (Brien, 2020), in conjunction with `asreml` (Butler et al., 2018), for calculating and using information. Here, `asremlPlus` and `asreml` are packages for the R Statistical Computing environment (R Core Team, 2020).

It is divided into the following main sections:

1. Set up the maximal model for this experiment
2. Obtaining information criteria for separate models
3. Obtaining information criteria for a prescribed sequence of model changes
4. Using information criteria to decide model changes

## 1. Set up the maximal model for this experiment

```
library(knitr)
opts_chunk$set("tidy" = FALSE, comment = NA)
suppressMessages(library(asreml, quietly=TRUE))
packageVersion("asreml")
```

```
## [1] '4.1.0.130'
```

```
suppressMessages(library(asremlPlus))
packageVersion("asremlPlus")
```

```
## [1] '4.2.21'
```

```
options(width = 100)
```

### Get data available in `asremlPlus`

The data are from a 1976 spring wheat experiment and are taken from Gilmour et al. (1995). An analysis is presented in the `asreml` manual by Butler et al. (2018, Section 7.6), although they suggest that it is a barley experiment.

```
data(Wheat.dat)
```

### Fit the maximal model

In the following a model is fitted that has the terms that would be included for a balanced lattice. In addition, a term `WithinColPairs` has been included to allow for extraneous variation arising between pairs of adjacent lanes. Also, separable `ar1` residual autocorrelation has been included. This model represents the maximal anticipated model,

```
max.asr <- asreml(yield ~ WithinColPairs + Variety,
  random = ~ Rep/(Row + Column) + units,
  residual = ~ ar1(Row):ar1(Column),
  data=Wheat.dat)
```

License check Sat Jul 11 10:27:55 2020

Model fitted using the gamma parameterization.

ASReml 4.1.0 Sat Jul 11 10:27:55 2020

	LogLik	Sigma2	DF	wall	cpu
1	-724.121	23034.14	124	10:27:55	0.0
2	-717.415	9206.93	124	10:27:55	0.0 (2 restrained)
3	-694.875	26492.99	124	10:27:55	0.0 (2 restrained)
4	-694.160	33101.80	124	10:27:55	0.0 (1 restrained)
5	-692.002	36912.26	124	10:27:55	0.0 (1 restrained)
6	-691.789	46701.51	124	10:27:55	0.0 (2 restrained)
7	-691.834	46208.51	124	10:27:55	0.0 (1 restrained)
8	-691.775	47698.26	124	10:27:55	0.0
9	-691.771	47041.85	124	10:27:55	0.0

Warning in asreml(yield ~ WithinColPairs + Variety, random = ~Rep/(Row + : Some components changed by more than 1% on the last iteration.

The warning from asreml is probably due to a bound term.

## Initialize a testing sequence by loading the current fit into an asrtests object

```
max.asrt <- as.asrtests(max.asr, NULL, NULL)
```

Calculating denominator DF

## Check for and remove any boundary terms

```
max.asrt <- rmboundary(max.asrt)
summary(max.asrt$asreml.obj)$varcomp
```

	component	std.error	z.ratio	bound	%ch
Rep:Row	4.293282e+03	3.199458e+03	1.3418779	P	0.0
Rep:Column	1.575689e+02	1.480357e+03	0.1064398	P	0.7
units	5.742689e+03	1.652457e+03	3.4752438	P	0.0
Row:Column!R	4.706787e+04	2.515832e+04	1.8708669	P	0.0
Row:Column!Row!cor	7.920301e-01	1.014691e-01	7.8056280	U	0.0
Row:Column!Column!cor	8.799559e-01	7.370402e-02	11.9390486	U	0.0

```
print(max.asrt, which = "testsummary")
```

#### Sequence of model investigations

(If a row has NA for p but not denDF, DF and denDF relate to fixed and variance parameter numbers)

terms	DF	denDF	p	AIC	BIC	action
1 Rep	1	NA	NA	NA	NA	Boundary

Rep has been removed because it has been constrained to zero. Following the recommendation of Littell et al. (2006, p. 150), the bound on all variance components is set to unconstrained (U) using

setvariances.asreml so as to avoid bias in the estimate of the residual variance. Alternatively, one could move Rep to the fixed model.

## Unbind Rep, Row and Column components and reload into an asrtests object

```
max.asr <- setvarianceterms(max.asr$call,
                           terms = c("Rep", "Rep:Row", "Rep:Column"),
                           bounds = "U")
```

Model fitted using the gamma parameterization.

ASReml 4.1.0 Sat Jul 11 10:27:56 2020

	LogLik	Sigma2	DF	wall	cpu
1	-724.121	23034.14	124	10:27:56	0.0
2	-717.415	9206.93	124	10:27:56	0.0 (2 restrained)
3	-694.875	26492.99	124	10:27:56	0.0 (2 restrained)
4	-693.974	33129.65	124	10:27:56	0.0 (1 restrained)
5	-692.886	39662.12	124	10:27:56	0.0
6	-691.428	53103.83	124	10:27:56	0.0
7	-691.239	48092.17	124	10:27:56	0.0
8	-691.181	47278.94	124	10:27:56	0.0
9	-691.171	46850.98	124	10:27:56	0.0
10	-691.170	46690.46	124	10:27:56	0.0

Warning in asreml(fixed = yield ~ WithinColPairs + Variety, random = ~Rep/(Row + : Some components changed by more than 1% on the last iteration.

```
max.asrt <- as.asrtests(max.asr, NULL, NULL)
```

Calculating denominator DF

```
max.asrt <- rmboundary(max.asrt)
summary(max.asrt$asreml.obj)$varcomp
```

	component	std.error	z.ratio	bound	%ch
Rep	-2462.3785855	1.191435e+03	-2.066734	U	0.2
Rep:Row	5012.4021413	3.396848e+03	1.475604	U	0.1
Rep:Column	920.5936388	1.704008e+03	0.540252	U	1.1
units	5964.9099374	1.608792e+03	3.707695	P	0.1
Row:Column!R	46690.4620355	2.731906e+04	1.709080	P	0.0
Row:Column!Row!cor	0.8152180	9.988929e-02	8.161216	U	0.1
Row:Column!Column!cor	0.8857252	7.487875e-02	11.828793	U	0.0

```
print(max.asrt, which = "testsummary")
```

#### Sequence of model investigations

(If a row has NA for p but not denDF, DF and denDF relate to fixed and variance parameter numbers)

```
[1] terms DF denDF p AIC BIC action
<0 rows> (or 0-length row.names)
```

Now the Rep component estimate is negative.

The test.summary output shows that no changes have been made to the model loaded using as.asrtests. The pseudo-anova table shows that Varieties are highly significant ( $p < 0.001$ )

## 2. Obtaining information criteria for separate models

The method `infoCriteria` has two methods for calculating information criteria. One, `infoCriteria.asreml`, is a method for `asreml` objects and the other, `infoCriteria.list`, is for 'listobjects', the components of `thelistbeingasreml` objects.

### Single models

Firstly, `infoCriteria` is called with the default `IClikelihood`, which is `REML`. Then it is called with `IClikelihood` set to `full` (Verbyla, 2019).

```
infoCriteria(max.asr)
```

	fixedDF	varDF	NBound	AIC	BIC	loglik
1	0	7	0	1396.34	1416.082	-691.17

```
infoCriteria(max.asr, IClikelihood = "full")
```

```
Model fitted using the gamma parameterization.  
ASReml 4.1.0 Sat Jul 11 10:27:57 2020
```

	LogLik	Sigma2	DF	wall	cpu
1	-691.170	46641.98	124	10:27:57	0.0

```
Warning in asreml(fixed = yield ~ WithinColPairs + Variety, random = ~Rep/(Row + : Log-likelihood  
not converged
```

	fixedDF	varDF	NBound	AIC	BIC	loglik
1	26	7	0	1647.193	1746.544	-790.5967

### A list of models

Now, a second model, from which the `withinColPairs` term has been omitted, is fitted; to be consistent, the variance components are unconstrained using `setvariances.asreml`. Then the `asreml` objects for this model and the maximal model are combined into a list and a `data.frame` produced that includes their information criteria.

```
m1.asr <- asreml(yield ~ Variety,  
               random = ~ Rep/(Row + Column) + units,  
               residual = ~ ar1(Row):ar1(Column),  
               data=Wheat.dat)
```

```
Model fitted using the gamma parameterization.  
ASReml 4.1.0 Sat Jul 11 10:27:57 2020
```

	LogLik	Sigma2	DF	wall	cpu
1	-727.774	22898.99	125	10:27:57	0.0
2	-721.097	9190.30	125	10:27:57	0.0 (2 restrained)
3	-698.313	26671.76	125	10:27:57	0.0 (2 restrained)
4	-697.517	32677.28	125	10:27:57	0.0 (1 restrained)
5	-695.419	36662.27	125	10:27:57	0.0 (1 restrained)
6	-695.208	46263.96	125	10:27:57	0.0 (2 restrained)
7	-695.198	46156.63	125	10:27:57	0.0
8	-695.191	46630.21	125	10:27:57	0.0

```
Warning in asreml(yield ~ Variety, random = ~Rep/(Row + Column) + units, : Some components changed  
by more than 1% on the last iteration.
```

```
m1.asr <- setvarianceterms(m1.asr$call,  
                          terms = c("Rep", "Rep:Row", "Rep:Column"),  
                          bounds = "U")
```

Model fitted using the gamma parameterization.

ASReml 4.1.0 Sat Jul 11 10:27:57 2020

	LogLik	Sigma2	DF	wall	cpu
1	-727.774	22898.99	125	10:27:57	0.0
2	-721.097	9190.30	125	10:27:57	0.0 (2 restrained)
3	-698.313	26671.76	125	10:27:57	0.0 (2 restrained)
4	-697.333	32689.33	125	10:27:57	0.0 (1 restrained)
5	-697.016	39975.97	125	10:27:57	0.0
6	-695.070	54825.30	125	10:27:57	0.0
7	-694.757	47637.20	125	10:27:57	0.0
8	-694.644	46775.41	125	10:27:57	0.0
9	-694.618	46175.06	125	10:27:57	0.0
10	-694.615	45940.69	125	10:27:57	0.0

Warning in asreml(fixed = yield ~ Variety, random = ~Rep/(Row + Column) + : Some components changed by more than 1% on the last iteration.

```
mods <- list(max = max.asr, m1 = m1.asr)
ic <- infoCriteria(mods, ICLikelihood = "full")
print(ic)
```

	fixedDF	varDF	NBound	AIC	BIC	loglik
max	26	7	0	1647.193	1746.544	-790.5967
m1	25	7	0	1645.326	1741.666	-790.6629

### 3. Obtaining information criteria for a prescribed sequence of model changes

The use of `changeTerms.asrtests` is demonstrated for a sequence of models, starting with the maximal model.

#### Drop the term for within Column pairs (a post hoc factor)

```
current.asrt <- as.asrtests(max.asrt$asreml.obj, NULL, NULL,
                           label = "Maximal model", ICLikelihood = "full")
```

Warning in asreml(fixed = yield ~ WithinColPairs + Variety, random = ~Rep/(Row + : Log-likelihood not converged

Calculating denominator DF

```
current.asrt <- changeTerms(current.asrt, dropFixed = "WithinColPairs",
                             label = "Drop withinColPairs", ICLikelihood = "full")
```

Warning in asreml(fixed = yield ~ Variety, random = ~Rep + units + Rep:Row + : Some components changed by more than 1% on the last iteration.

Calculating denominator DF

Calculating denominator DF

```
print(current.asrt, which = "testsummary", omit.columns = "p")
```

#### Sequence of model investigations

(If a row has NA for p but not denDF, DF and denDF relate to fixed and variance parameter numbers)

	terms	DF	denDF	AIC	BIC	action
1	Maximal model	26	7	1647.193	1746.544	Starting model
2	Drop withinColPairs	25	7	1645.326	1741.666	Changed fixed

So the same values of the information criteria have been obtained as when `infoCriteria.list` was used on a `list` containing the `asreml` objects for the two models. The difference is that here there is ultimately only one fitted model, the model stored in the `asreml` object in the `asrtests` object named `current.asrt`: this is the model with `withinColPairs` omitted.

Note this use of the `omit.columns` argument from `print.test.summary` to omit the irrelevant column `p` from the `test.summary`.

## Drop nugget term

```
current.asrt <- changeTerms(current.asrt, dropRandom = "units",
                           label = "Drop units", ICLikelihood = "full")
```

Warning in `asreml(fixed = yield ~ Variety, random = ~Rep + Rep:Row + Rep:Column, : Some components changed by more than 1% on the last iteration.`

```
Calculating denominator DF
Calculating denominator DF
```

## Check Row autocorrelation

```
current.asrt <- changeTerms(current.asrt, newResidual = "Row:ar1(Column)",
                           label="Row autocorrelation", ICLikelihood = "full")
```

```
Calculating denominator DF
Calculating denominator DF
```

```
print(current.asrt, which = "testsummary", omit.columns = "p")
```

```
#### Sequence of model investigations
```

(If a row has NA for `p` but not `denDF`, `DF` and `denDF` relate to fixed and variance parameter numbers)

	terms	DF	denDF	AIC	BIC	action
1	Maximal model	26	7	1647.193	1746.544	Starting model
2	Drop withinColPairs	25	7	1645.326	1741.666	Changed fixed
3	Drop units	25	6	1650.126	1743.456	Changed random
4	Row autocorrelation	25	5	1660.882	1751.201	Changed residual

## 4. Using information criteria to decide model changes

This section illustrates the use of `changeModelOnIC.asrtests` to decide between consecutive models in a sequence of models. The default information criterion to use for this is the AIC. However, `which.IC` can be used to specify the use of the BIC or `both`. Here we use the AIC and the full likelihood.

### Check the term for within Column pairs (a post hoc factor)

As before, we start with the maximal model, in which the variance components have been unconstrained and look to decide whether or not to drop the `withinColPairs` term.

```
current.asrt <- as.asrtests(max.asrt$asreml.obj, NULL, NULL,
                          label = "Maximal model", IClkelihood = "full")
```

Warning in asreml(fixed = yield ~ WithinColPairs + Variety, random = ~Rep/(Row + : Log-likelihood not converged

Calculating denominator DF

```
current.asrt <- iterate(current.asrt)
```

Calculating denominator DF

```
current.asrt <- changeModelOnIC(current.asrt, dropFixed = "WithinColPairs",
                                label = "withinColPairs",
                                IClkelihood = "full", which.IC = "AIC",
                                allow.unconverged = FALSE)
```

Warning in asreml(fixed = yield ~ Variety, random = ~Rep + units + Rep:Row + : Some components changed by more than 1% on the last iteration.

Calculating denominator DF

Calculating denominator DF

```
print(current.asrt, which = "testsummary", omit.columns = "p")
```

#### Sequence of model investigations

(If a row has NA for p but not denDF, DF and denDF relate to fixed and variance parameter numbers)

	terms	DF	denDF	AIC	BIC	action
1	Maximal model	26	7	1647.193455	1746.544420	Starting model
2	withinColPairs	-1	0	-1.867556	-4.878191	Swapped

Given the warning about a lack of convergence, we use `iterate.asrtests` to perform additional iterations of the fitting process. It seems that it was successful.

It can be seen from the `test.summary` that the term has been swapped out and this has the effect of reducing the number of fixed parameters by one and makes no change to the variance parameters.

## Check the nugget term

```
current.asrt <- changeModelOnIC(current.asrt, dropRandom = "units",
                                label = "units", IClkelihood = "full",
                                allow.unconverged = FALSE)
```

Warning in asreml(fixed = yield ~ Variety, random = ~Rep + Rep:Row + Rep:Column, : Some components changed by more than 1% on the last iteration.

Calculating denominator DF

Calculating denominator DF

## Check Row autocorrelation

```
current.asrt <- changeModelOnIC(current.asrt, newResidual = "Row:ar1(Column)",
                                label="Row autocorrelation", IClkelihood = "full",
                                allow.unconverged = FALSE)
```

Warning in asreml(fixed = yield ~ Variety, random = ~Rep + units + Rep:Row + : Log-likelihood not converged

Warning in asreml(fixed = yield ~ Variety, random = ~Rep + units + Rep:Row + : Some components changed by more than 1% on the last iteration.

Warning in newfit.asreml(asreml.obj, fixed. = fix.form, random. = ran.form, :

Calculating denominator DF

Warning in asreml(fixed = yield ~ Variety, random = ~Rep + units + Rep:Row + : Some components changed by more than 1% on the last iteration.

Calculating denominator DF

Warning in asreml(fixed = yield ~ Variety, random = ~Rep + units + Rep:Row + : Some components changed by more than 1% on the last iteration.

## Check Column autocorrelation (depends on whether Row autocorrelation retained)

```
{
  last.action <- current.asrt$test.summary$action[current.asrt$test.summary$terms ==
                                                    "Row autocorrelation"]
  if (grepl("Unswapped", last.action, fixed = TRUE) |
      grepl("Unchanged", last.action, fixed = TRUE))
    current.asrt <- changeModelOnIC(current.asrt, newResidual = "ar1(Row):Column",
                                   label="Col autocorrelation", ICLikelihood = "full",
                                   allow.unconverged = FALSE)
  else
    current.asrt <- testresidual(current.asrt, newResidual = "Row:Column",
                                 label="Col autocorrelation", ICLikelihood = "full",
                                 allow.unconverged = FALSE)
}
```

Calculating denominator DF

Warning in infoCriteria.asreml(asreml.obj, ICLikelihood = ic.lik, bound.exclusions = bound.exclusions):  
Row:Column!Row!cor

Calculating denominator DF

Warning in infoCriteria.asreml(new.asrttests.obj\$asreml.obj, ICLikelihood = ic.lik, : The following bound  
Row:Column!Row!cor

## Output the results

```
print(current.asrt, which = "test", omit.columns = "p")
```

#### Sequence of model investigations

(If a row has NA for p but not denDF, DF and denDF relate to fixed and variance parameter numbers)

	terms	DF	denDF	AIC	BIC	action
1	Maximal model	26	7	1647.193455	1746.544420	Starting model
2	withinColPairs	-1	0	-1.867556	-4.878191	Swapped



```

3          units  0   -1   4.801145   1.790510          Unswapped
4 Row autocorrelation  0   -1  17.819174  14.808538 Unchanged - new unconverged
5 Col autocorrelation  0   -2  19.479776  13.458506          Unswapped

```

```
summary(current.asrt$asreml.obj)$varcomp
```

	component	std.error	z.ratio	bound	%ch
Rep	-2391.9489939	1.194581e+03	-2.0023338	U	0.4
Rep:Row	5035.5311054	3.406006e+03	1.4784269	U	0.3
Rep:Column	761.9535622	1.612103e+03	0.4726458	U	1.2
units	5933.2133794	1.610805e+03	3.6833848	P	0.1
Row:Column!R	45970.8383027	2.635124e+04	1.7445415	P	0.0
Row:Column!Row!cor	0.8101615	9.995498e-02	8.1052641	U	0.1
Row:Column!Column!cor	0.8846970	7.503039e-02	11.7911827	U	0.0

The `test.summary` shows us that the model without the autocorrelation failed to converge and so no change was made to the model. It, and the messages from checking the Column autocorrelation, also show us that the omission of the Column autocorrelation resulted in the Row autocorrelation becoming bound. That is, dropping the Column autocorrelation resulted in the dropping of two variance parameters

The function `printFormulae.asreml` is used to display the fitted model.

```
printFormulae(current.asrt$asreml.obj)
```

```
#### Formulae from asreml object
```

```

fixed: yield ~ Variety<environment: 0x000000001a2aca48>
random: ~ Rep + units + Rep:Row + Rep:Column<environment: 0x000000001a2aca48>
residual: ~ ar1(Row):ar1(Column)<environment: 0x000000001a2aca48>

```

## References

- Brien, C. J. (2020) `asremlPlus`: *Augments ASReML-R in fitting mixed models and packages generally in exploring prediction differences*. Version 4.2-15. <http://cran.r-project.org/package=asremlPlus/> or <http://chris.brien.name/rpackages/>.
- Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J. and Thompson, R. (2018). *ASReML-R Reference Manual Version 4*. VSN International Ltd, <http://asreml.org>.
- Gilmour, A. R., Thompson, R., & Cullis, B. R. (1995). Average Information REML: An Efficient Algorithm for Variance Parameter Estimation in Linear Mixed Models. *Biometrics*, **51**, 1440–1450.
- Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberger, O. (2006). *SAS for Mixed Models* (2nd ed.). Cary, N.C.: SAS Press.
- R Core Team (2020) *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.r-project.org>.
- Verbyla, A. P. (2019). A note on model selection using information criteria for general linear models estimated using REML. *Australian & New Zealand Journal of Statistics*, **61**, 39-50. <https://doi.org/10.1111/anzs.12254>.