Evaluating the Effect of the Nugget Variance in the Spatial Analysis of National Variety Trials

UNIVERSITY OF WOLLONGONG AUSTRALIA

Chris Lisle, David Hughes, Ky Mathews
University of Wollongong, clisle@uow.edu.au

CBB CENTRE FOR BIOINFORMATICS AND BIOMETRICS

SUPPORTED BY GRDC GRAINS RESEARCH & DEVELOPMENT CORPORATION

Key Message

- Gilmour et al. (1997) recommends the use of a separable first-order autoregressive model in two dimensions with the inclusion of the so-called "nugget" variance (i.e. measurement error) for the analyses of field experiments. In this study, we have found:
  - minimal convergence issues with fitting this model;
  - the nugget variance is not always statistically significant;
  - when the nugget variance is significant, the predicted variety ranks do not change compared to analyses without the nugget variance; and,
  - the model-based trial accuracies of variety predictions decrease when fitting the nugget variance.
- The results of this empirical study demonstrate that for field trials the use and requirement of the nugget variance is minimal, similar to Besag & Kempton (1986), Zimmerman & Harville (1991) and Stein (2012).

Motivating Example

- The National Variety Trials (NVT) program was established in 2005 by the Grains Research and Development Corporation (GRDC) of Australia. The aim of each NVT is to assess the comparative performance of new and widely adopted varieties in terms of grain yield.
- The motivating example is from the 2012-2016 Wheat South data set. A summary is provided in **Table 1**.
- Analyses were conducted using ASReml-R (Butler et al., 2018).

Table 1: *Summary of the 2012-2016 NVT Wheat South data set by year including: the overall number of trials, median number of varieties per trial. The number of trials and the median number of rows for 3-column and 6-column trials are also displayed.*

Year	Trials	Varieties	3-Column		6-Column	
			Trials	Rows	Trials	Rows
2012	37	48	30	30	7	48
2013	35	45	25	21	10	47
2014	32	31	25	17	7	31
2015	32	52	24	27	8	51
2016	34	50	24	25	10	48
Total	170		128		42	

Statistical Methodology

- The random variety effects for individual trials are predicted using a linear mixed model (LMM).
- Let $\boldsymbol{y}$ be the n-vector of yield data for a trial. The LMM for $\boldsymbol{y}$ can be written as,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e}$$

where $\boldsymbol{\tau}$ is a vector of fixed effects with associated design matrix $\boldsymbol{X}$, $\boldsymbol{u}$ is the vector of random effects with associated design matrix $\boldsymbol{Z}$, and $\boldsymbol{e}$ is the vector of residual effects.

- Three variance models for $\boldsymbol{e}$ are explored,
  - **units:** An independent structure:

$$\mathrm{var}(\boldsymbol{e}) = \sigma^2 \boldsymbol{I}_n$$

  - **AR1xAR1:** A separable autoregressive process of order 1 (note: an AR1 structure was only fitted when either rows or columns > 3):

$$\mathrm{var}(\boldsymbol{e}) = \sigma^2 \boldsymbol{\Sigma}_c(\boldsymbol{\rho}_c) \otimes \boldsymbol{\Sigma}_r(\boldsymbol{\rho}_r)$$

where $\boldsymbol{\Sigma}_c$ and $\boldsymbol{\Sigma}_r$ are the spatial correlation structures containing parameters in $\boldsymbol{\rho}_c$ and $\boldsymbol{\rho}_r$ for the column and row directions, respectively.

  - **AR1xAR1 + Nugget:** Decomposed into $\boldsymbol{\xi} + \boldsymbol{\eta}$, where $\boldsymbol{\xi}$ represents the AR1xAR1 structure, and $\boldsymbol{\eta}$ is the nugget effect (Gilmour et al., 1997):

$$\mathrm{var}(\boldsymbol{e}) = \sigma_\xi^2 \boldsymbol{\Sigma}_c(\boldsymbol{\rho}_c) \otimes \boldsymbol{\Sigma}_r(\boldsymbol{\rho}_r) + \sigma_\eta^2 \boldsymbol{I}_n$$

Results

Table 2: *Summary of the three residual models fitted sequentially to the 170 trials in the motivating data set. The number of trials that converged after 26 iterations (1 update), and the number of trials where the following model has a better fit than the preceding one using a loglikelihood ratio test (LRT) with a significance level of 0.05.*

Residual Model	Trials Converged	Trials with significant LRT to previous model
units	170	
AR1xAR1	170	161
AR1xAR1 + Nugget	167	63

Figure 1: *Boxplot of the estimated nugget variance (on the gamma scale) for the 63 trials where significant.*

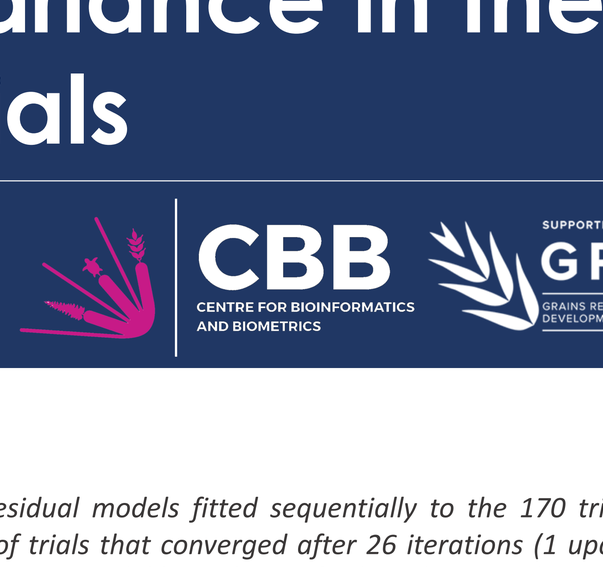

Figure 2: *Histogram of the correlation between predicted variety effects for the analyses with and without the nugget variance, for the 63 trials where significant.*

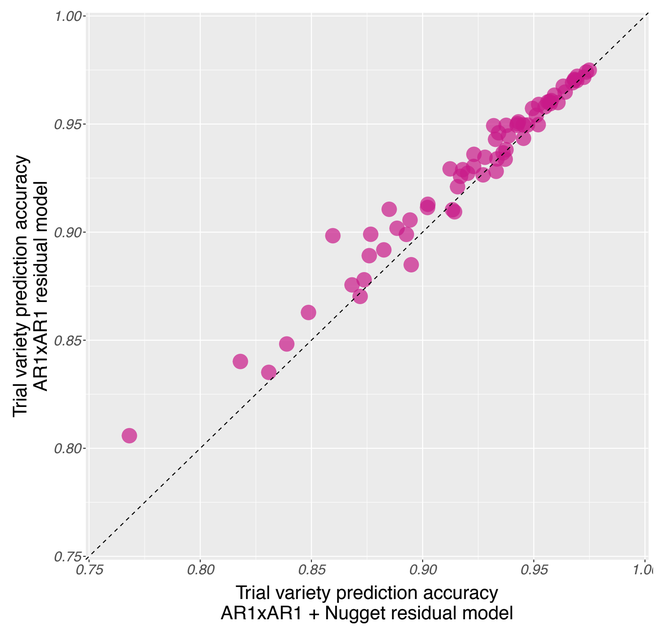

Figure 3: *Comparisons of model-based accuracies for trial variety predictions (Mrode, 2005) between the analyses without and with the nugget variance for the 63 trials where significant.*

Discussion

- There were minimal convergence issues (**Table 2**).
- The nugget variance was not significant for 58% of trials (**Table 2**).
- The median estimated nugget variance, where significant, on the gamma scale was 0.23 (**Figure 1**), which differs to the default initial starting value of 0.1 in ASReml-R.
- Fitting the nugget variance incurred negligible changes to the predicted variety effects within trials (**Figure 2**), with a median correlation of 0.995 across trials (where significant).
- Model-based accuracies for trial variety predictions are generally higher for the **AR1xAR1** analyses than the **AR1xAR1 + Nugget** analyses (**Figure 3**).
- The concepts described here are specifically for the analysis of single trials and may not translate to the analysis of multi-environment trial data sets.

References

Besag, J., & Kempton, R. (1986). Statistical analysis of field experiments using neighbouring plots. *Biometrics*, 231-251.

Butler D. G, Cullis B. R, & Thompson R (2018). ASReml: An R package to fit the linear mixed model. *Journal of Statistical Software*, in prep.

Gilmour, A. R., Cullis, B. R., & Verbyla, A. P. (1997). Accounting for natural and extraneous variation in the analysis of field experiments. *Journal of Agricultural, Biological, and Environmental Statistics*, 2(3), 269-293.

Mrode, R. A. (2005). Linear models for the prediction of animal breeding values. *CABI publishing.*

Stein, M. L. (2012). Interpolation of spatial data: some theory for kriging. *Springer Science & Business Media*.

Zimmerman, D. L., & Harville, D. A. (1991). A random field approach to the analysis of field-plot experiments and other spatial experiments. *Biometrics*, 223-239.

Acknowledgements

Thank you to the GRDC for funding this research. Project code: UW00009.