On the optimal design of experiments with dependent observations and correlated treatments

David Butler\textsuperscript{1}, Alison Smith\textsuperscript{2} and Brian Cullis\textsuperscript{3}
\textsuperscript{1}AgriScience Queensland, Department of Employment, Economic Development and Innovation, Toowoomba, QLD
\textsuperscript{2}University of Wollongong, Wollongong, NSW
\textsuperscript{3}University of Wollongong, Wollongong, NSW and Centre for Mathematics, Informatics and Statistics, Canberra

Phenotypic evaluation of genetic material is an important part of the selection cycle in plant genetic improvement programs. The identification of elite individuals for release to industry is but one selection objective, another equally important goal is the selection of potential parents for use in the breeding program. The same field experiments are often used for both objectives, although separate mating experiments may be conducted. In animal breeding, where replication is either not easy or impossible, known genetic relationships are routinely used to predict the breeding value, or additive genetic effect, of an individual, and consequently its value as a potential parent. The prediction of breeding values and total line genetic effects using this approach has been less prolific in plant breeding, but is of growing interest as digitized pedigrees and molecular marker information become routinely available. The overall efficiency of the trialing process depends on sound experimental design, yet design methods recognising this genetic relatedness have received little attention.

Experimental design promotes the efficient separation of genetic and environmental effects in the analysis of phenotypic observations. Commonly, designs used in plant breeding programs fall within the general class of block designs, and include complete or incomplete block, row-column and -designs. Partially replicated designs are increasingly used in early generation evaluation where resources are limiting. In more complex situations, end-product suitability is assessed in subsequent laboratory phases, where sample numbers are usually constrained by cost or capacity. In such instances methods such as embedded designs in the field have been proposed as a way to maintain statistical rigor while reducing sample numbers for expensive quality determinations. These features of the experimental landscape encountered across plant breeding and allied areas of science, including unorthodox blocking or replication strategies, correlated observations, known genetic relatedness and procedural restrictions, challenge standard design methods. Traditionally, these experiments have been designed assuming independent effects, constrained perhaps by the availability of practical methods for alternative variance models. The assumption of independent observations has attracted recent attention, for example, and numerical design methods for spatially correlated data are being increasingly adopted within Australian plant breeding programs. Optimal design provides a flexible model based framework for constructing experimental designs in novel situations.

We consider the design of plant breeding experiments where the residuals may be correlated with a presupposed process, and there is a known genetic covariance structure among treatments. A computational approach will be discussed in the context of the R statistical environment that admits the specification of a design in terms of the underlying linear model, and illustrated with several of the cases mentioned above, including partially replicated and embedded designs.