



Guest Editors' Introduction to the Special Issue on "Recent Advances in Design and Analysis of Experiments and Observational Studies in Agriculture"

Hans-Peter PIEPHO, Robert J. TEMPELMAN, and Emlyn R. WILLIAMS

The Journal of Agricultural, Biological and Environment Statistics (JABES) special issue on Recent Advances in Design and Analysis of Experiments and Observational Studies in Agriculture covers a select set of topics currently of primary importance in the field. Efficient use of resources in agricultural research, as well as valid statistical inference, requires good designs, and this special issue boasts seven papers providing both review and cutting-edge methodology for the purpose. A broad range of methods for analysis of data arising in different branches agricultural research is covered in another five exciting papers. This special issue highlights the importance of and opportunities for applied statistics in agriculture.

Key Words: Design of experiments; Generalized linear mixed models; Structural equation models; Michaelis–Menten kinetics; Generalized additive models; Spatial modelling.

Agriculture is a vibrant field of science, requiring up-to-date statistical methodology both at the design and analysis stage of experiments and studies. This special issue brings together recent developments in the field. There are a total of twelve papers, seven of which focus primarily on design and five on analysis. This balance between design and analysis is intentional as we believe both are equally important. Whilst we cover agriculture quite broadly, there are seven papers focused on plant breeding and variety testing, giving this area some prevalence; most of these papers also have a focus on design.

A natural starting point for design of experiments in agriculture is the excellent review by Verdooren (2020) who tracks development from ancient times through to the advent of Sir R. A. Fisher in 1926, which represents the start of modern statistical design. The paper then highlights some important advances in the design of agricultural experiments over the ensu-

Hans-Peter Piepho (✉), Institute of Crop Science, University of Hohenheim, 70 593 Stuttgart, Germany (E-mail: hans-peter.piepho@uni-hohenheim.de). Robert J. Tempelman, Department of Animal Science, Michigan State University, East Lansing, MI 48824-1225, USA. Emlyn R. Williams, Statistical Consulting Unit, Australian National University, Canberra, ACT 2601, Australia.

© 2020 The Author(s)

Journal of Agricultural, Biological, and Environmental Statistics, Volume 25, Number 4, Pages 453–456
<https://doi.org/10.1007/s13253-020-00417-z>

ing nearly 100 years. Included is the discussion of various designs with incomplete blocking structures, factorial and spatial designs and software for design generation. Among the other papers with an emphasis on design, five deal with designs that are most commonly used in plant breeding and variety testing, even though the concepts and methods are clearly more broadly applicable. Bailey et al. (2020) construct efficient alternatives to the non-existent square lattice designs for 36 varieties and for up to eight replicates and study their properties. Whilst square lattice designs provide an optimal arrangement of v^2 varieties in resolvable incomplete blocks of size v , for $v = 6$ such designs cannot be constructed when there are more than three replicates. Designs are produced using both mathematical theory and design generation software. Edmondson (2020) proposes an interesting extension of classical blocked designs that allows multiple nested levels of blocking, thus addressing the common challenge that optimal block size is seldom known exactly beforehand. Several analysis strategies are also considered, including smoothing methods based on generalized additive models. Hoefler et al. (2020) describe a large-scale simulation study to investigate the performance of spatial designs relative to designs with more traditional blocking structures such as alpha, row-column and partially replicated designs. Designs are analysed and compared across multiple locations. The results provide important information for researchers when designing field experiments. Two further papers consider designs when population structure and genetic relatedness among breeding lines are taken into account. Whereas classical design approaches are based on linear models with fixed treatment effects, these two papers model treatments (genotypes) as random. Cullis et al. (2020) propose a novel approach to optimize design for a given set of genotypes in early-stage plant breeding trials when genetic relatedness is modeled using a marker-based kinship matrix. The authors also provide freely available software to generate designs using their proposed methods, which should be of broad interest to plant breeders and agronomists. Heslot and Feoktistov (2020) tackle the problem of selecting a subset of genotypes for phenotyping when the objective is to do genomic prediction for a larger related set of genotypes. Design optimization is done using an evolutionary algorithm. This key problem in the optimization of breeding programmes is solved for three cases, i.e. (i) optimization of selective phenotyping of available individuals, (ii) optimization of hybrid testing and (iii) optimization of designs for genetically connected crosses. Finally, Huang et al. (2020) explore strategies for the design of experiments to estimate nonlinear regression models suitable to assess extended Michaelis–Menten kinetics in biochemical reactions involving enzymes and substrates. The authors focus on applications with several controllable inputs and consider optimal designs based on multifactor hybrid nonlinear models, which are computationally rather challenging. Among other things, they study a compound design criterion for discriminating between two candidate models, which they recommend for design of advanced kinetic studies.

Structural equation models (SEM) have been increasingly utilized for data analyses involving agricultural production systems to infer causality or directional relationships between economically important traits. However, the typical assumption of SEM is that these relationships are uniform across production environments. Chitakasempornkul et al. (2020) propose a Bayesian model to allow for heterogeneity in these relationships between six economically important reproduction traits in a swine production system. They demonstrate that these heterogeneous specifications provide substantial improvements in model fit

in their application relative to the typical SEM specifications. A broad class of models popular among agricultural statisticians are generalized linear mixed models (GLMM). These are routinely used to analyse many different observational studies and designed experiments involving discrete outcomes. A predominant consensus that has emerged has been that integral approximation methods (e.g. Gauss–Hermite quadrature) should be generally preferred to linearization based (e.g. pseudo-likelihood) methods for estimating variance components in GLMM. Using extensive simulation representative of agricultural experimental designs, Stroup and Claassen (2020) demonstrate the folly of that general recommendation and provide some guidelines in making a suitable choice between the two GLMM estimation strategies. Lewis-Beck et al. (2020) look at the use of remote sensing data for monitoring crop phenology and development within and across seasons. They nicely illustrate the use of a functional data approach in a spatio-temporal setting, accounting for spatial dependence between locations through the functional curve coefficients. Modeling across multiple growing years, and including growing degree days as a covariate, the authors estimate the timing for when crops reach their peak each season in the US corn belt. The remaining two papers deal with the analysis of field trials using spatial methods, a field of application that also featured prominently in several of the design-related papers. Boer et al. (2020) explore the links between three popular models for spatial analysis, i.e. the linear variance model, the random walk model and P-splines with first-differences penalties. They discuss in which settings these three models are equivalent and when they differ. The comparison provides new perspectives for further developments based on P-splines. Continuing on the same theme, Mao et al. (2020) propose a strategy for accurate estimation of genetic effects for genomic prediction in plant breeding populations whilst accounting for spatial (co)variability within fields. They additionally model subpopulation effects, not fully accounted for by genetic markers, as a third source of (co)variability. Their Gaussian random field model jointly accounts for these three sources of (co)variability using Gaussian kernel specifications. They demonstrate the competitiveness of their proposed method with other recently developed methods on publicly available datasets involving comparisons of maize and wheat varieties. We note that the latter two papers consider first differences among neighbouring plots in their spatial variance–covariance structures, but in different ways as regards the (i) estimation or definition of variance parameters, (ii) handling the singularity in the associated precision matrix, and whether the matrix related to first differences enters the variance matrix assumed for the data or the precision matrix (generalized inverse of the variance matrix).

In summary, the manuscripts provide excellent examples of the importance of statistical research related to applications in the agricultural sciences. It goes without saying that this special issue can provide but a very small glimpse of the diverse and interesting opportunities and challenges that agriculture provides for statisticians. We are hopeful that the exciting work presented in this special issue spawns further contributions to the development and adaptation of statistical methodology for agricultural research.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give

appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

[Published Online October 2020.]

REFERENCES

- Bailey, R.A., Cameron, P.J., Soicher, L.H., and Williams, E.R. (2020), "Substitutes for the Non-existent Square Lattice Designs for 36 Varieties". *Journal of Agricultural, Biological, and Environmental Statistics*, <https://doi.org/10.1007/s13253-020-00388-1>.
- Boer, M., Piepho, H.P., and Williams, E.R. (2020), "Linear Variance, P-Splines and Neighbour Differences for Spatial Adjustment in Field Trials: How are They Related?" *Journal of Agricultural, Biological, and Environmental Statistics*.
- Chitakasempornkul, K., G. J. M. Rosa, G.J.M., Jager, A., and Bello, N.M. (2020), "Hierarchical Modeling of Structural Coefficients for Heterogeneous Networks with an Application to Animal Production Systems". *Journal of Agricultural, Biological, and Environmental Statistics*. <https://doi.org/10.1007/s13253-020-00389-0>.
- Cullis, B.R., Smith, A.B., Cocks, N.A., and Butler, D.G. (2020), "The Design of Early-Stage Plant Breeding Trials Using Genetic Relatedness". *Journal of Agricultural, Biological, and Environmental Statistics*. <https://doi.org/10.1007/s13253-020-00403-5>.
- Edmondson, R.N. (2020), "Multi-Level Block Designs for Comparative Field Experiments". *Journal of Agricultural, Biological, and Environmental Statistics*.
- Heslot, N., and Feoktistov, V. (2020), "Optimization of selective phenotyping and population design for genomic prediction". *Journal of Agricultural, Biological, and Environmental Statistics*.
- Hoefler, R., González-Barrios, P., Bhatta, M., Nunes, J.A.R., Berro, I., Nalin, R.S., Borges, A., Covarrubias, E., Diaz-Garcia, L., Quincke, M., and Gutierrez, L. (2020), "Do Spatial Designs Outperform Classic Experimental Designs?" *Journal of Agricultural, Biological, and Environmental Statistics*. <https://doi.org/10.1007/s13253-020-00406-2>
- Huang, Y., Gilmour, S.G., Mylona, K., and Goos, P. (2020), "Optimal Design of Experiments for Hybrid Nonlinear Models, with Applications to Extended Michaelis–Menten Kinetics". *Journal of Agricultural, Biological, and Environmental Statistics*. <https://doi.org/10.1007/s13253-020-00405-3>
- Lewis-Beck, C., Zhu, Z., Walker, V.A., and Hornbuckle, B.K. (2020), "Modeling Crop Phenology in the U.S. Corn Belt using Spatially Referenced SMOS Satellite Data". *Journal of Agricultural, Biological, and Environmental Statistics*.
- Mao, X., Dutta, S., Wong, R.K.W., and Nettleton, D. (2020), "Adjusting for Spatial Effects in Genomic Prediction". *Journal of Agricultural, Biological, and Environmental Statistics*. <https://doi.org/10.1007/s13253-020-00396-1>.
- Stroup, W., and Claassen, E. (2020), "Pseudo-Likelihood or Quadrature? What We Thought We Knew, What We Think We Know, and What We Are Still Trying to Figure Out". *Journal of Agricultural, Biological, and Environmental Statistics*. <https://doi.org/10.1007/s13253-020-00402-6>
- Verdooren, L.R. (2020), "History of the Statistical Design of Agricultural Experiments". *Journal of Agricultural, Biological, and Environmental Statistics*. <https://doi.org/10.1007/s13253-020-00394-3>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.