



Editorial for the special issue for mODa 13: model-oriented data analysis and optimum design

David C. Woods¹ · Stefanie Biedermann² · Chiara Tommasi³

Published online: 18 July 2023

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

The mODa workshops bring together researchers and practitioners to present and discuss recent advances and applications in the field of optimum experimental design. Early career researchers are a particular focus of the workshops, providing them an opportunity to establish professional networks with leading researchers in the field. Substantial opportunities are always provided for individual research discussions and mentoring. mODa is (generally) organised every three years, starting from 1987 (Eisenach, former GDR). An important feature of the early workshops was that they brought together researchers from either side of the (collapsing) Iron Curtain. The meetings have in the past been held in a variety of both Eastern and Western European countries. After a short Covid-induced delay, the 13th edition of the workshop series will be held in Southampton, UK, 9–15 July 2023. The UK has a strong tradition in DoE dating from the development of methods for randomised experiments at the Rothamsted Experimental Station in the 1920s. mODa 13 will be the first mODa workshop hosted in the UK.

Formal statistical methodology for designing experiments has benefitted experimenters for over a century in ensuring accurate and precise conclusions can be drawn from experimental data for minimum cost. It is still an active, modern research field, particularly valued in industrial settings where experimental costs are high (e.g. manufacturing), statistical rigour is a regulatory requirement (e.g. clinical trials) or where causal reasoning is necessary (e.g. A/B testing in the technology industry). Even in fields where high-throughput or “big data” is available, clarity in conclusions can be provided by statistical design thinking on fundamentals such as randomisation, replication and stratification. Published proceedings of the mODa workshops have traditionally been made available (see the list in the References section). The 17 papers in the proceedings for the 13th workshop reflect many modern trends in optimum

✉ David C. Woods
d.woods@soton.ac.uk

¹ University of Southampton, Southampton, UK

² Open University, Milton Keynes, UK

³ Università degli Studi di Milano, Milan, Italy

design: analytic and numerical construction of designs under different, possible multiple criteria; big data subsampling; connections to machine learning and uncertainty quantification; and clinical studies and trials.

1. Radloff and Schwabe present analytical results on D -optimal designs when the experimental region is a k -dimensional sphere. For a class of models with symmetrical intensity functions, which includes binary response models with logit and probit links, their results simplify design search considerably. They further propose three construction methods for minimally supported exact designs and show through examples that these are highly efficient. The results can be generalized to arbitrary k -dimensional ellipsoidal design regions.
2. Tommasi, Rodríguez-Díaz and López-Fidalgo investigate multi-objective design problems by casting them in a maximin design framework. The minimum efficiency of a design, with respect to a finite number of different optimality criteria, is to be maximized. They prove an equivalence theorem which states that the maximin optimal design is Bayesian optimal for a specific prior distribution on the set of the individual optimality criteria, thus transforming an optimization problem on a non-differentiable function into one on a differentiable function. In addition, they provide an analytic method to find this prior distribution, which facilitates the applicability of the equivalence theorem for design search.
3. Prus focuses on designs for random coefficient regression models in the case where there is one observation per individual. Design optimality conditions are established for prediction of the random effects in a subset of the individuals. Explicit optimal designs are established for the random intercept and random slope linear models under certain restrictions on the design space and variance components.
4. Yu, Liu and Wang describe a new algorithm for selecting informative subdata from massive data for a wide class of models. They show the connection between their selection method and some optimality criteria, such as A -, D -, E - and T -criteria. They provide theoretical justifications for the proposed algorithm, and develop numerical simulations to compare the performance of their algorithm with other commonly used subsampling and subdata selection procedures.
5. To deal with the increasing availability of large datasets and the consequent need of subsampling, Reuter and Schwabe propose a subsampling method based on D -optimality for polynomial regression in one covariate that has been generated by an invariant distribution. The quadratic regression is more deeply studied for specific distributions of the covariate. In particular, they make statements on the shape of the resulting optimal subsampling designs and the effect of the subsample size on the design.
6. Deldossi, Pesce and Tommasi provide methods for subsampling big data that use (i) characterization of the design matrix or the covariates (referred to as non-informative sampling) along with (ii) the knowledge about the responses (referred to as informative sampling). The objective is to discount non-representative points which lie near the boundary of the design space. Such points are routinely chosen by standard criteria such as D -optimality. To characterise such points, the

authors use leverage values and Cook's distance, and combine these measures with exchange algorithms for finding optimal subsets.

7. Mahendran, Thompson and McGree propose a model-robust approach to finding sampling probabilities for the big data subsampling problem. The approach overcomes the dependence of the optimal subsample on a single assumed model for the response. Sampling probabilities are constructed as weighted combinations of probabilities for individual models, with weights chosen to reflect a priori beliefs about the model set. The methods are illustrated using simulation studies and real examples assuming generalised linear models for the response.
8. The design problem addressed by Müller and Schorning is motivated by a problem from electrical engineering where electrical power distribution grids of medium and low-voltage levels are studied. In such a distribution grid, the question arises where measurements of the electrical power should be taken, and how precise these measurements should be, in order to obtain a precise estimation of the state of the grid. Müller and Schorning analytically derive A -optimal designs for a simple network with star configuration and find conditions which simplify the numerical calculation of A -optimal designs in more complicated scenarios.
9. The paper by Pronzato and Zhigljavsky uncovers the connections between (a) simple and ordinary kriging with kernel K for prediction of values of a random field indexed by a set X , (b) energy minimization for K , and (c) parameter estimation with the Ordinary Least Squares Estimator and the Best Linear Unbiased Estimator in the location model with observations whose correlation is defined by K . They emphasize the special role of the constant function in their theoretical results and provide several illustrative examples.
10. Strouwen, Nicolaï and Goos explore optimal model-based designs for dynamic systems where the Kalman filter can be used and both measurement and process noise are present. The Fisher information matrix for appropriate models is characterised, and robust or "pseudo-Bayesian" designs are found. The authors also study adaptive, or sequential, designs. Two case studies, on a mass-spring-damper system and a two-compartment model, are studied.
11. Maruri-Aguilar and Wynn generalise the connection between computational algebraic geometry and the theory and practice of design of experiments. The focus is on sparse grids and polynomial interpolators, widely used in numerical methods. The grid is described via an inclusion-exclusion formula, with Betti numbers used as a computation tool to reduce the number of terms in the formula.
12. Fontana, Molena, Pegoraro and Salmaso review the application of machine learning and design of experiments approaches in the active learning (sequential design) setting. A simulation study assessed the effectiveness of the combination of various designs (including fractional factorials and space-filling designs) and machine learning methods (including artificial neural nets and support vector machines). The ALPERC active learning framework is then described, which combines sequential design with variable selection and model-fitting. A further simulation study shows the benefits of this approach over competing active learning strategies and non-sequential designs.

13. Yousefi, Pronzato, Hainy, Müller and Wynn discuss the design and analysis of experiments to discriminate between Gaussian process (GP) models with different covariance kernels. Adaptive and static design methods are described. The former uses Kullback–Leibler divergence between two GPs or the mean squared error of competing models; the latter uses loglikelihood ratios or the Fréchet distance between the covariance functions. Designs are compared on numerical examples using hit rate for the correct model (or power). Necessary conditions are given for a design to be optimal under a criterion using differences in expected log-likelihood ratios.
14. To address the challenges caused by moving towards personalized medicine, innovative study designs with increasing complexity have been proposed. In particular, adaptive enrichment designs are becoming more attractive for their flexibility. Baldi Antognini, Frieri and Zagoraiou present a comprehensive review of adaptive enrichment studies with a focus on design considerations. They discuss multiple aspects involved in adaptive enrichment designs, including clinical, ethical, scientific and statistical considerations, that contribute to their advantages and disadvantages.
15. Flournoy considers Bayesian design for studies that have the possibility of early stopping after an interim analysis. She raises concerns that, in experiments with informative interim stopping decisions, standard practice is not to condition the sampling density on interim decisions that are made, and thus not to utilize information about the decision. Flournoy introduces, and examines the consequences of, a Bayesian design approach that accommodates interim decisions.
16. Chen, Fries and Leonov introduce a longitudinal model for the change from baseline for a clinical endpoint, to describe the disease progression for different dose groups. They build a nonlinear mixed effects model using the techniques which have been more frequently applied in the design and analysis of population pharmacokinetic/pharmacodynamics studies in the last two decades. To evaluate operating characteristics of the proposed design, they derive the Fisher information matrix and validate analytical results via simulations.
17. Tarima and Flournoy consider group sequential tests powered for multiple ordered alternative hypotheses with a predetermined α -spending function. They prove that if a parametric distribution of the data is either known or assumed, then MLE-based group sequential tests powered for multiple ordered alternatives are most powerful for this set of hypotheses in either finite or local asymptotic settings. They provide some examples of how their theory applies.

References

- Atkinson AC, Pronzato L, Wynn HP (eds) (1998) *mODa 5: advances in modeloriented data analysis and experimental design, contributions to statistics*. Physica. <https://doi.org/10.1007/978-3-642-58988-1>
- Atkinson AC, Hackl P, Müller WG (eds) (2001) *mODa 6: advances in modeloriented design and analysis, contributions to statistics*. Physica. <https://doi.org/10.1007/978-3-642-57576-1>
- Di Bucchianico A, Läuter H, Wynn HP (eds) (2004) *mODa 7: advances in modeloriented design and analysis, contributions to statistics*. Physica. <https://doi.org/10.1007/978-3-7908-2693-7>

- Fedorov V, Läuter H (eds) (1987) Model-oriented data analysis: proceedings of an IIASA workshop on data analysis held at Eisenach, GDR, in March 1987, Lecture notes in economics and mathematical systems (Book 297). Springer, Berlin
- Fedorov VV, Müller WG (1992) Model oriented data-analysis: a survey of recent methods: proceedings of the 2nd IIASA-workshop in St. Kyrik, Bulgaria, May 28–June 1, 1990. Springer, Heidelberg
- Giovagnoli A, Atkinson AC, Torsney B, May C (eds) (2010) mODa 9: advances in model-oriented design and analysis, contributions to statistics. *Physica*. <https://doi.org/10.1007/978-3-7908-2410-0>
- Harman R, Müller WG, Woods DC (eds) (2019) Special issue on mODa12: advances in model-oriented design and analysis. *Stat Pap*
- Kitsos CP, Müller WG (eds) (1995) mODa4: advances in model-oriented data analysis, contributions to statistics. *Physica*. <https://doi.org/10.1007/978-3-662-12516-8>
- Kunert J, Müller CH, Atkinson AC (eds) (2016) mODa 11: advances in modeloriented design and analysis, contributions to statistics. Springer, Cham
- López-Fidalgo J, Rodríguez-Díaz JM, Torsney B (eds) (2007) mODa 8: advances in model-oriented design and analysis, contributions to statistics. *Physica*. <https://doi.org/10.1007/978-3-7908-1952-6>
- Müller WG, Wynn HP, Zhigljavsky AA (1993) Model-oriented data analysis: proceedings of the 3rd international workshop in Petrodvoretz, Russia, May 25–30, 1992. *Physica*
- Uciński D, Atkinson AC, Patan M (eds) (2013) mODa 10: advances in modeloriented design and analysis, contributions to statistics. Springer, Heidelberg

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