Design and Analysis of Simulation Experiments, Jack P. C. Kleijnen, Springer ▹, 2008.

Reviewed by Thomas Bartz-Beielstein, Faculty of Computer Science and Engineering Science, Cologne University of Applied Sciences.

Kleijnen's books, e.g. "Statistical Tools for Simulation Practitioners" from 1987 [1], are references for simulation practitioners working in industry, management, computer science, and many other disciplines. Unfortunately, these books are out of print for many years. New statistical methods, namely *design and analysis of computer experiments* (DACE) became popular [7] and new aspects such as *multivariate* simulation output were considered important over the last 20 years.

In the meantime Kleijnen continued his research and has published more than 200 articles. Now he has written a successor of his seminal books on simulation, so that a central source of information is available.

Overview

Chapter 1 introduces basic terminology used in the book and provides answers to questions like "What is simulation?" or "What is DASE?". DASE stands for *design and analysis of simulation experiments*—an acronym closely related to DACE. Chapter 2 presents basics from regression analysis and designs for experiments. It starts with a simple metamodel $\mathbf{y} = \mathbf{X}\beta + \mathbf{e}$. Most of the following sections discuss properties of the least squares estimates $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{w}$. This discussion includes consequences of violations of the normality assumption, properties of t statistics, and the treatment of replicates. Design considerations are discussed next. Design matrices are illustrated in a very comprehensive manner. Different designs types are presented in theses sections.

The third chapter entitled "classic assumptions revisited" comprehends approx. 30 pages that present the author's experiences from simulation and optimization studies. This chapter is very instructive, because it discusses problems that occur in many simulation experiments: How realistic are the classic assumptions and how can they be tested? Can the violation of these assumptions be repaired? And, if not, what should be done? Since *least squares* (LS) is a mathematical criterion, it does not require a normal distribution. However, additional statistical tests are based on certain statistical properties.

Bootstrapping, which requires a representative sample of the underlying distribution, is discussed. If simulation is very expensive and only a few runs are possible, so that a representative sample cannot be obtained, parametric bootstrapping is recommended and exemplified. Tests for constant variances are given. Variance heterogeneity is also mentioned for Kriging models (p. 147). Using a result from C.R. Rao, Kleijnen concludes (p. 77) that multiple simulation outputs can still be analyzed through *ordinary least squares* (OLS). This simplifies the situation for simulation practitioners, because they can apply classical statistical tests for the regression parameters per type.

Simulation optimization is introduced in chapter 4. It starts with the classic *response surface methodology* (RSM). Multiple outputs, constraints, and risk analysis are considered in this chapter. It contains also the description of *Latin hypercube sampling* (LHS).

Kriging basics are introduced in chapter 5. Differences between classical linear regression models and ordinary Kriging are outlined. LHS is presented as the default design technique for DACE. This is in accordance with the presentation given in [7]. Sacks et al. describe LHS as adequate for DACE, because they are computationally cheap, can cope with many variables, and provide a systematic way of discovering scientifically surprising behavior.

Screening, i.e., seeking for really important factors among the many factors, plays a significant role in simulation and optimization. Chapter 6 discusses screening methods with a focus on *sequential bifurcation* (SB).

The epilogue (chapter 7) gives a very short summary of the book.

Discussion

This book covers the relevant topics from simulation accompanied with references to authoritative publications. Since space for this review is limited, I will concentrate my discussion on the following aspects:

1. Classical linear regression vs. Kriging
2. Designs
3. Assumptions
4. Importance

Linear regression vs. Kriging. Kriging has gained a tremendous popularity in recent years. Kleijnen states (p. 140): "Although Kriging in random simulation is still rare, I strongly believe that the track record Kriging achieved in deterministic simulation holds promise for Kriging in random simulation!" Why should simulation practitioners still use classical regression? Answers to this central question are given in the book, but they are a little bit hidden in the text. Kleijnen claims that linear regression models can be used for local fitting, e.g., when searching for the optimum input combination, whereas Kriging is better suited for global fitting. Kriging and related nonlinear models may give better predictions than classical linear regression models, but "these alternatives are so complicated that they do not help the analysts better understand the underlying simulation model—except for sorting the simulation inputs in order of their importance." (p.99)

On p. 147, one can read: ". . . we give examples of Kriging predictions [. . .] that are much better than the regression predictions. (Regression metamodels may be useful for other goals such as understanding, screening, and validation [. . .])".

However, Kriging might also helpful in understanding and screening. I compare results from different models (classical regression, Kriging, classification and regression trees), whenever this is possible. Kleijnen applied Kriging to random simulation, I have also successfully applied Kriging to random optimization.

Design considerations. Kleijnen discusses optimal designs and mentions several optimality definitions. However, selection of optimal designs in practice is a "chicken or the egg causality dilemma": the optimal design depends on the regression model, but how can practitioners chose a suitable model *a priori*, i.e., before the simulation is performed and no experimental results are at hand? And, the situation is even worse: Santner et al. claim that LHS is popular in DACE, not because it is superior to other designs, but because it is easy to implement and understand [9]. Obviously, design considerations are not trivial. Unfortunately, the popular *one-factor-at-a-time* (OAT) design is not very effective and efficient. Kleijnen writes (p.7): "In practice, however, many analysts keep many inputs constant, and experiment with a few factors only. Another example of inferior practice is changing only one input at a time (while keeping all other inputs fixed at their so-called base values)." At the first sight, this statement is in contrast to Saltelli et al.'s rating of OAT designs [8]. They propose an elementary effects method which "is conceptually simple and easy to implement. It belongs to the class of OAT designs ...". But Saltelli et al. consider different goals, namely *sensitivity analysis* (SA). On page 124, DASE and (SA) are compared: "SA may use DASE, because DASE gives better answers; i.e., the common sense approach changing one factor at a time gives estimators of factor effects that have higher standard errors, and does not enable estimation of interactions among factors ..."

Kleijnen presents sequential designs as an alternative for LHS, because sequential statistical procedures are known to be more efficient and computer experiments proceed sequentially. However, he does not report problems related to sequential designs, e.g., that there may be a tendency for design sites to "pile up" [7].

Assumptions revisited. Assumptions from classic linear regression such as white noise and only univariate output "usually do not hold." (p.73). This observation is typical for many real-world simulations. While discussing sequential bifurcation, Kleijnen notes that theoretically, the SB procedure does not satisfy the classical statistical assumptions. Nevertheless, numerical results look promising (p. 165). I recommend reading Gary Klein's enlightening article [2] entitled "The Fiction of Optimization" that discusses these discrepancies between theoretical assumptions and the situation in field experiments. Klein claims that he has "not identified any decision researcher or analyst who believes that these [theoretical] assumptions will be met in any setting, with the possible exception of the laboratory or the casino."

Importance. Importance can be relative—this is mentioned on p. 31: "I point out that a factor may be significant when tested through the t statistic [...], but may be unimportant." (p.31 and also p.62). Kleijnen discusses, without explicitly mentioning, the large n problem which is well-known in philosophy of science [4]. Results that are statistically significant, e.g., results from t tests, are not automatically scientifically meaningful. Regarding *importance* also other problems might occur, e.g., factors that are statistically unimportant in the first phase of simulation might become important at a later stage.

A few additional notes

The classic 1996 publication [3] and Santner et al.'s 2003 book [9], which would be my first choices for Kriging designs discussed in section 5.4, are not mentioned in this section. For example, chapters 5 and 6 in [9] discuss space-filling designs and designs based on other criteria, e.g., maximum entropy.

Important is Kleijnen's differentiation between strategic and tactical aspects (p. 9). Tactical issues such as "How long should a simulation run be continued?" or "How accurate is the resulting value of the estimator?" arise only in stochastic simulation. Strategic issues, e.g., "How do to chose an adequate simulation model?" arise in deterministic and stochastic simulation. This book deals with strategic issues, tactical issues are discussed in the first part of the 1987's book [1].

It would be very interesting to read a summary of the most important open research questions from Kleijnen's perspective. So, I was a little bit disappointed that the book was already finished after turning page 171.

Some rather technical remarks. Roman numerals are used for the resolution. This is standard notation in *design of experiments* (DOE), but not explained in this book. A sentence like "In a resolution R design no p-factor interactions is aliased with another effect containing less than $R-p$ factors" [5] might be helpful. The concept of aliasing is introduced later (on p. 44). Minor modifications in the structure of these sections might improve the comprehensibility of DOE concepts. Although Kleijnen discusses the most important aspects of factorial and fractional factorial designs, it might be useful having a book like [5] at hand to get deeper insight into this rather technical material.

Formatting could be improved (table captions should appear above the table, figures could be scaled better). But this is only a minor point.

There are more than 400 references listed in the book. It would be nice if an *apalike* bibliography style, i.e., a style which includes some hints about the authors, would have been used. Again, this criticism is only of minor importance.

Summary

This book was written for researchers, students, and mature practitioners who get valuable hints for their projects and requires basic knowledge of simulation and mathematical statistics. It summarizes results in a very compact manner and collects material that is scattered over numerous publications. It transforms ideas from statistics to simulation and optimization.

This book is a valuable source for instructors. It contains many examples with references to both toy and real-world problems. Some material in the book was used to teach a course "Simulation for Logistics" at the Technical University Eindhoven. Course materials are available on the author's web pages. Exercises with solutions are given. Instructions for readers (recommended chapters) are given, but only very briefly.

The book complements Kleijnen's seminal books on simulation (including new topics like Kriging and multivariate output) and summarizes research results from several hundreds publications. It does not provide any rigorous proofs such as [9] or [6], but gives hands-on support. Although the *theoretical* concept of regression is rather simple, its *application* to real-world application requires an extensive knowledge and experience. This book imparts experience from one of the leading experts in this field. It is definitively a up-to-date reference for simulation and optimization practitioners. I do not know any other book which does this better.

References

[1] J. P. C. Kleijnen. *Statistical Tools for Simulation Practitioners*. Marcel Dekker, New York NY, 1987.

[2] G. Klein. The fiction of optimization. In G. Gigerenzer and R. Selten, editors, *Bounded Rationality: The Adaptive Toolbox*, pages 103–121. MIT Press, Cambridge MA, 2002.

[3] J. R. Koehler and A. B. Owen. Computer experiments. In S. Ghosh and C. R. Rao, editors, *Handbook of Statistics*, volume 13, pages 261–308. Elsevier Science, New York NY, 1996.

[4] D. G. Mayo and A. Spanos. Severe Testing as a Basic Concept in a Neyman-Pearson Philosophy of Induction. *Br J Philos Sci*, 57(2):323–357, 2006.

[5] D. C. Montgomery. *Design and Analysis of Experiments*. Wiley, New York NY, 5th edition, 2001.

[6] F. Pukelsheim. *Optimal Design of Experiments*. Wiley, New York NY, 1993.

[7] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4(4):409–435, 1989.

[8] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global Sensitivity Analysis*. Wiley, 2008.

[9] T. J. Santner, B. J. Williams, and W. I. Notz. *The Design and Analysis of Computer Experiments*. Springer, Berlin, Heidelberg, New York, 2003.