Springer Proceedings in Mathematics & Statistics

Jürgen Pilz · Dieter Rasch
Viatcheslav B. Melas · Karl Moder
Editors

Statistics and Simulation

IWS 8, Vienna, Austria, September 2015

Springer

Springer Proceedings in Mathematics & Statistics

Volume 231

Springer Proceedings in Mathematics & Statistics

This book series features volumes composed of selected contributions from workshops and conferences in all areas of current research in mathematics and statistics, including operation research and optimization. In addition to an overall evaluation of the interest, scientific quality, and timeliness of each proposal at the hands of the publisher, individual contributions are all refereed to the high quality standards of leading journals in the field. Thus, this series provides the research community with well-edited, authoritative reports on developments in the most exciting areas of mathematical and statistical research today.

More information about this series at http://www.springer.com/series/10533

Jürgen Pilz · Dieter Rasch
Viatcheslav B. Melas · Karl Moder
Editors

Statistics and Simulation

IWS 8, Vienna, Austria, September 2015

Springer

Editors
Jürgen Pilz
Institute of Statistics
Alpen-Adria University of Klagenfurt
Klagenfurt
Austria

Dieter Rasch
Institute of Applied Statistics
and Computing
University of Natural Resources
and Life Sciences
Vienna
Austria

Viatcheslav B. Melas
St. Petersburg State University
St. Petersburg
Russia

Karl Moder
Institute of Applied Statistics
and Computing
University of Natural Resources
and Life Sciences
Vienna
Austria

ISSN 2194-1009 ISSN 2194-1017 (electronic)
Springer Proceedings in Mathematics & Statistics
ISBN 978-3-319-76034-6 ISBN 978-3-319-76035-3 (eBook)
https://doi.org/10.1007/978-3-319-76035-3

Library of Congress Control Number: 2018934442

Mathematics Subject Classification (2010): 62XX, 68XX, 92XX

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG
part of Springer Nature
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The present volume contains selected contributions given at the 8th International Workshop on Simulation held at the University of Natural Resources and Life Sciences, Vienna, Austria, September 21–25, 2015.

The conference was organized by the Center of Experimental Design of the Institute of Applied Statistics and Computing of the University of Natural Resources and Life Sciences, Vienna, in collaboration with the Department of Statistics of the Alpen-Adria University of Klagenfurt, the Department of Statistical Modelling of Saint Petersburg State University, and INFORMS Simulation Society (USA). This international conference was devoted to statistical techniques in stochastic simulation, data collection, and analysis of scientific experiments and studies representing broad areas of interest. The 1st–6th Workshops took place in St. Petersburg (Russia) in 1994, 1996, 1998, 2001, 2005, and 2009. The 7th International Workshop on Simulation took place in Rimini, May 21–24, 2013.

The conference in Vienna was held in memory of Luidmila Kopylova- Melas, the wife of Viatcheslav Melas who initiated this series of conferences. Luidmila passed away on September 21, 2013; she worked relentlessly as secretary of the whole series of our simulation workshops.

The Scientific Program Committee was chaired by Viatcheslav Melas (St. Petersburg, Russia), Dieter Rasch (Vienna, Austria), and Jürgen Pilz (Klagenfurt, Austria). We are indebted to the following members of the Scientific Program Committee for their fruitful help in organizing the sessions and making the Vienna Workshop a tremendous success: Aleksander Andronov (Latvia), Anthony Atkinson (UK), Narayanaswamy Balakrishnan (Canada), Russell Barton (USA), Michel Broniatowski (France), Ekaterina Bulinskaya (Russia), Holger Dette (Germany), Sergei Ermakov (Russia), Valerii Fedorov (USA), Nancy Flournoy (USA), Subir Ghosh (USA), Marie Hušková (Czech Republic), Jack Kleijnen (The Netherlands), Gennady Mikhailov (Russia), Simos Meintanis (Greece), Werner Müller (Austria), Valery Nevzorov (Russia), Michael Nikulin (France), Jordi

Ocania (Spain), Ingram Olkin (USA), Fortunato Pesarin (Italy), Luigi Salmaso (Italy), Rainer Schwabe (Germany), John Stufken (USA), Bruno Tuffin (France), Dariusz Ucinski (Poland), Henry Wynn (UK).

The Local Organizing Committee was led by Karl Moder (Vienna, Austria). We are thankful to the following members of this committee for their extremely helpful and efficient organizational work during the conference: Marianne Mansuri (Vienna), Beate Simma (Klagenfurt), Bernhard Spangl (Vienna), Gunter Spöck (Klagenfurt), and Albrecht Gebhardt (Klagenfurt).

The present proceedings volume consists of six parts; the first part contains four invited papers, and the remaining five parts deal with various applications of simulations.

The first of the invited papers, presented by Jack P. C. Kleijnen, gives an overview of the state of the art in the design and analysis of simulation experiments, with a special emphasis on simulation optimization in operation research. The second of the invited papers gives a review of simulation usage in the New Zealand electricity market: G. Zakeri and G. Pritchard demonstrate, in particular, how optimization of electricity consumption and reserves can be combined in an efficient way. In the third invited paper, Z. Prášková gives an overview of bootstrap changepoint testing procedures for dependent data. In the last one of the invited papers, C. Draxler and K. D. Kubinger review the present state and future challenges of power and sample size determination in psychometrics.

The contributed twenty-nine papers have been arranged in six parts dealing with different aspects of simulation in mathematical analysis, stochastic processes, statistical estimation and testing problems, clinical trials, design of experiments and in reliability and queueing theory models and applications.

The chapters in Part II (Simulation for Mathematical Modeling and Analysis) start with a contribution by T. M. Tovstik studying in detail the covariation matrix of solutions of linear algebraic system equations via the Monte Carlo method. O. N. Soboleva and E. P. Kurochkina consider large-scale simulation studies of acoustic waves in random multiscale media. H. S. Bhat, R. A. Madushani, and S. Rawat deal with parameter inference for stochastic differential equations with density tracking by quadrature. G. A. Mikhailov, N. V. Tracheva, and S. A. Ukhinov present a new Monte Carlo algorithm for the evaluation of outgoing polarized radiation.

Simulation models and their analysis for stochastic process applications played an important role at the 8th IWS. Contributions in this direction are collected in Part III of the present proceedings volume. E. Ermishkina and E. Yarovaya study the evolution and simulation of branching random walks. Y. Belopolskaya studies stochastic models for nonlinear cross-diffusion systems. N. Vollert, M. Ortner, and J. Pilz report on experiences with the application of tree-structured Gaussian process models for optimization in magnetic field shaping problems. The last three contributions in Part III deal with applications in actuarial science and stochastic finance: E. Bulinskaya and J. Gusak consider insurance models under incomplete

information; Ch. Quast et al. model and compare pension systems in Austria, Chile, Slovakia, and Sweden; A. Andronov and T. Yurkina study the Markowitz portfolio problem in a particular random environment.

Part IV collects contributed chapters on the use of simulation models for statistical testing and classification problems. S. Tarima et al. report on the use of signs of residuals for testing coefficients in quantile regression. B. Darkhovsky and A. Piryatinska apply their concept of ε-complexity (based on Kolmogorov's notion of complexity) to the classification of multivariate time series and give an application to the classification of EEG data. P. Langthaler et al. analyze high-dimensional data from the spectral density curves of EEG measurements on several channels to dementia classification of patients. B. Peštová and M. Pešta use simulation studies to compare ratio and non-ratio test statistics to detect structural changes in panel data. Finally, D. Rasch and T. Yanagida report on robustness results for the two-sample triangular sequential t-test against variance heterogeneity.

Part V (Clinical Trials and Design of Experiments) starts with a contribution by N. Minois et al. on the performance of the Poisson–gamma model for patients' recruitment in clinical trials when there are pauses in the recruitments procedure. N. Savy et al. detail their views on principles and good practices for simulated clinical trials, with a focus on virtual patient generation. D. Rasch et al. report on the determination of the optimal sample size of subsamples for testing a correlation coefficient by a sequential triangular test. The last two chapters in Part V deal with experimental design issues: V. B. Melas and P. V. Shpilev give explicit solutions for determining T-optimal discriminating designs for trigonometric regression models. R. Fontana and F. Rapallo perform simulation studies on the combinatorial structure of D-optimal designs.

In the final Part VI, we have collected five contributions dealing with the role of simulations for reliability and queueing models. G. Tzavelas and P. Economou investigate the consequences of model misspecification for biased samples from the Weibull distribution. D. Kurz, H. Lewitschnig, and J. Pilz give an overview on recent advances in statistical burn-in modeling for an efficient evaluation of early life failure probabilities of semiconductor devices. K. E. Samouylov, Y. V. Gaidamaka, and E. S. Sopin describe a simplified approach to the analysis of queueing systems with additional randomness due to imperfect knowledge of the exact amount of resources released by the departure of a customer. V. Rykov and D. Kozyrev compare analytic and simulation results on the sensitivity of steady-state probabilities of a cold redundant system to the shapes of life and repair time distributions of its elements. D. Efrosinin et al. perform a reliability analysis of an aging unit with a controllable repair facility activation, using a continuous-time Markov chain model for the process of gradual aging.

It is our great pleasure to thank all authors of invited and contributed chapters for carefully preparing their manuscripts and submitting them for editorial processing of the present volume. We are indebted to our reviewers from the Scientific Program Committee for critical reading and providing constructive comments.

Finally, we are indebted to the relentless secretarial work and technical help by Beate Simma and Johannes Winkler from Alpen-Adria University of Klagenfurt and to Mrs. Veronika Rosteck from Springer International Publishing.

Klagenfurt, Austria — Jürgen Pilz
Rostock, Germany — Dieter Rasch
Vienna, Austria — Viatcheslav B. Melas
St. Petersburg, Russia — Karl Moder
September 2017

Contents

Contributors

Sandrine Andrieu INSERM UMR 1027, University of Toulouse III, Toulouse, France; Epidemiology Unit of Toulouse CHU, Toulouse, France

Alexander Andronov Transport and Telecommunication Institute, Riga, Latvia

Arne C. Bathke Paris-Lodron-University Salzburg, Salzburg, Austria

Yana Belopolskaya Saint-Petersburg State University of Architecture and Civil Engineering, St. Petersburg, Russian Federation

Harish S. Bhat University of California Merced, Merced, CA, USA

Ekaterina Bulinskaya Lomonosov Moscow State University, Moscow, Russia

Boris Darkhovsky Institute for Systems Analysis, FRC CSC RAS, Higher School of Economics, Moscow, Russia

Clemens Draxler University for Health and Life Sciences, Hall, Austria

Polychronis Economou Department of Civil Engineering, University of Patras, Rio Achaia, Greece

Dmitry Efrosinin Johannes Kepler University Linz, Linz, Austria; Institute of Control Sciences, Moscow, Russia

Ekaterina Ermishkina Department of Probability Theory, Lomonosov Moscow State University, Moscow, Russia

Mais Farkhadov Institute of Control Sciences, Moscow, Russia

Roberto Fontana Department DISMA, Politecnico di Torino, Torino, Italy

Yuliya V. Gaidamaka Peoples' Friendship University of Russia (RUDN University), Moscow, Russia; Institute of Informatics Problems, FRC CSC RAS, Moscow, Russia

Julia Gusak Lomonosov Moscow State University, Moscow, Russia

Yvonne Höller Paracelsus Medical University Salzburg, Salzburg, Austria; Department of Neurology, Christian Doppler Medical Centre and Centre of Cognitive Neuroscience, Paracelsus Medical University Salzburg, Salzburg, Austria

Zuzana Hübnerová Brno University of Technology, Brno, Czech Republic

Tatjana Jurkina Transport and Telecommunication Institute, Riga, Latvia

Jozef Kiseľák Faculty of Science, Institute of Mathematics, P.J. Šafárik University in Košice, Kosice, Slovakia

Jack P. C. Kleijnen Tilburg University, Tilburg, Netherlands

Dmitry Kozyrev Peoples' Friendship University of Russia (RUDN University), Moscow, Russian Federation; V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia

Klaus D. Kubinger Division of Psychological Assessment and Applied Psychometrics, Faculty of Psychology, University of Vienna, Vienna, Austria

Ekaterina P. Kurochkina The Novosibisk State University - Baker Hughes Joint Laboratory of The Multi-Scale Geophysics and Mechanics, Novosibirsk, Russia

Daniel Kurz Department of Statistics, Alpen-Adria University of Klagenfurt, Klagenfurt, Austria

Patrick Langthaler Paris-Lodron-University Salzburg, Salzburg, Austria; Paracelsus Medical University Salzburg, Salzburg, Austria

Valérie Lauwers-Cances Epidemiology Unit, CHU Purpan, Toulouse, France

Horst Lewitschnig Infineon Technologies Austria AG, Villach, Austria

R. W. M. A. Madushani University of California Merced, Merced, CA, USA

Sébastien Marque Capionis, Paris, France; Osmose, Bordeaux, France

Viatcheslav B. Melas Department of Mathematics, St. Petersburg State University, St. Petersburg, Russia

John Meurer Institute for Health and Society, Medical College of Wisconsin, Milwaukee, WI, USA

Guillaume Mijoule University of Paris XI, Orsay, France

Gennady A. Mikhailov Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia; Novosibirsk State University, Novosibirsk, Russia

Nathan Minois INSERM UMR 1027, University of Toulouse III, Toulouse, France

Michael Ortner CTR Carinthian Tech Research AG, Villach, Austria

Michal Pešta Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Charles University, Prague, Czech Republic

Barbora Peštová Department of Medical Informatics and Biostatistics, Institute of Computer Science, The Czech Academy of Sciences, Prague, Czech Republic

Jürgen Pilz Department of Statistics, Alpen-Adria University of Klagenfurt, Klagenfurt, Austria

Alexandra Piryatinska San Francisco State University, CA, USA

Rastislav Potocký Department of Applied Mathematics and Statistics, Comenius University in Bratislava, Bratislava, Slovak Republic

Geoff Pritchard Department of Statistics, University of Auckland, Auckland, New Zealand

Zuzana Prášková Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic

Christian Quast Department of Applied Statistics, Johannes Kepler University, Linz, Austria

Fabio Rapallo Department DISIT, Università del Piemonte Orientale, Alessandria, Italy

Dieter Rasch University of Natural Resources and Life Sciences, Vienna, Austria

Shagun Rawat University of California Merced, Merced, CA, USA

Lisa Rein Institute for Health and Society, Medical College of Wisconsin, Milwaukee, WI, USA

Vladimir Rykov Peoples' Friendship University of Russia (RUDN University), Moscow, Russian Federation; Gubkin Russian State Oil and Gas University, Moscow, Russia

Konstantin E. Samouylov Peoples' Friendship University of Russia (RUDN University), Moscow, Russia; Institute of Informatics Problems, FRC CSC RAS, Moscow, Russia

Nicolas Savy Toulouse Institute of Mathematics, University of Toulouse III, Toulouse, France

Stéphanie Savy INSERM UMR 1027, University of Toulouse III, Toulouse, France

Berthold Schneider Institute for Biometry, Hannover Medical School, Hannover, Germany

Petr V. Shpilev Department of Mathematics, St. Petersburg State University, St. Petersburg, Russia

Olga N. Soboleva Novosibisk State Technical University, Novosibirsk, Russia; The Novosibisk State University - Baker Hughes Joint Laboratory of The Multi-Scale Geophysics and Mechanics, Novosibirsk, Russia

Eduard S. Sopin Peoples' Friendship University of Russia (RUDN University), Moscow, Russia; Institute of Informatics Problems, FRC CSC RAS, Moscow, Russia

Rodney Sparapani Institute for Health and Society, Medical College of Wisconsin, Milwaukee, WI, USA

Milan Stehlík Linz Institute of Technology and Department of Applied Statistics, Johannes Kepler University, Linz, Austria

Natalia Stepanova Altai Economics and Law Institute, Barnaul, Russia

Luboš Střelec Department of Statistics and Operation Analysis, Mendel University in Brno, Brno, Czech Republic

Janos Sztrik University of Debrecen, Debrecen, Hungary

Peter Tarassenko International Department of Management, Tomsk State University, Tomsk, Russia

Sergey Tarima Institute for Health and Society, Medical College of Wisconsin, Milwaukee, WI, USA

Tatiana M. Tovstik St. Petersburg State University, St. Petersburg, Russia

Natalya V. Tracheva Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia; Novosibirsk State University, Novosibirsk, Russia

Bonifride Tuyishimire Institute for Health and Society, Medical College of Wisconsin, Milwaukee, WI, USA

George Tzavelas Department of Statistics and Insurance Sciences, University of Piraeus, Piraeus, Greece

Sergey A. Ukhinov Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk State University, Novosibirsk, Russia

Vítězslav Veselý Brno University of Technology, Brno, Czech Republic

Natalie Vollert CTR Carinthian Tech Research AG, Villach, Austria; Department of Statistics, Alpen-Adria University of Klagenfurt, Klagenfurt, Austria

Takuya Yanagida University of Applied Sciences Upper Austria, Vienna, Austria; University of Vienna, Vienna, Austria

Elena Yarovaya Department of Probability Theory, Lomonosov Moscow State University, Moscow, Russia

Golbon Zakeri Department of Engineering Science, University of Auckland, Auckland, New Zealand

Part I
Invited Papers

Chapter 1
Design and Analysis of Simulation Experiments

Jack P. C. Kleijnen

Abstract This contribution summarizes the design and analysis of experiments with computerized simulation models. It focuses on two metamodel (surrogate, emulator) types, namely first-order or second-order polynomial regression, and Kriging (or Gaussian process). The metamodel type determines the design of the simulation experiment, which determines the input combinations of the simulation model. Before applying these metamodels, the analysts should screen the many inputs of a realistic simulation model; this contribution focuses on sequential bifurcation. Optimization of the simulated system may use either a sequence of first-order and second-order polynomials—so-called response surface methodology (RSM)—or Kriging models fitted through sequential designs—including efficient global optimization (EGO). Robust optimization accounts for uncertainty in some simulation inputs.

Keywords Robustness and sensitivity · Metamodel · Design · Regression Kriging

1.1 Introduction

Simulation is used in many scientific disciplines, but we focus on statistics and engineering. Moreover, we focus on stochastic (random) simulation, but parts of our contribution are also relevant for deterministic simulation. Simulation requires several steps; see [17, p. 67]. A crucial step is the design and analysis of the experiments with the computerized simulation model. This design and analysis are "intertwined": selecting an experimental design assumes a metamodel (surrogate, emulator) for the analysis of the experimental results; e.g., changing a single factor (simulation input or parameter) at a time assumes a metamodel with non-interacting factors. We focus on the two most popular metamodel types: low-order polynomial regression and Kriging.

J. P. C. Kleijnen (✉)
Tilburg University, Postbox 90153, Tilburg, Netherlands
e-mail: kleijnen@tilburguniversity.edu

J. Pilz et al. (eds.), *Statistics and Simulation*, Springer Proceedings in Mathematics & Statistics 231, https://doi.org/10.1007/978-3-319-76035-3_1

Mathematically, a *metamodel* is an explicit and relatively simple approximation of the input/output (I/O) function implicitly defined by the underlying simulation model. We define $w = f_{\text{sim}}(\mathbf{z}, \mathbf{r})$ where w is the random simulation output (response), f_{sim} the simulation I/O function, $\mathbf{z}$ the vector with the values of the k simulation inputs with the integer $k \geq 1$, and $\mathbf{r}$ the vector with pseudorandom numbers (PRNs) so $\mathbf{r}$ vanishes in deterministic simulation. Usually, $\mathbf{z}$ is standardized, so the resulting $\mathbf{d}$ has elements $-1 \leq d_j \leq 1$ ($j = 1, \ldots, k$). An input may be qualitative. If a qualitative input has more than two values (levels), then special care is needed; see [12, pp. 69–71].

We define $y = f_{\text{meta}}(\mathbf{x}) + e$ where y is the metamodel output, $\mathbf{x}$ the vector with (say) q metamodel inputs (explanatory variables), e the approximation (fitting) error; an example of f_{meta} is a second-order polynomial in d_j ($j = 1, \ldots, k$) so $\mathbf{x}$ has the components d_j, $d_j d_{j'}$ with $j \leq j'$, and the constant 1. Actually, a *polynomial* of any order is a *linear* regression (meta)model. Another type of metamodel is Kriging—or Gaussian process (GP)—metamodels, which are also explicit—but more complicated—models of d_j. Altogether, f_{meta} is explicit and much simpler than f_{sim}. We call f_{meta} "adequate" or "valid" if $E(e) = 0$.

We focus on simulation for *sensitivity analysis* (SA) and *optimization* of the underlying real system. Furthermore, we focus on global (not local) SA; e.g., in screening and Kriging, we use global metamodels (see Sects. 1.4 and 1.5). Nevertheless, we use local SA in response surface methodology (RSM) for optimization.

We base our survey on our book [12], which includes many Web site addresses for software and hundreds of additional references, and on our article [14]. However, compared with [14], our survey is half the length, corrects a mathematical error, and assumes familiarity with basic statistical design concepts (e.g., resolution and CCD) and basic operations research (OR) concepts (e.g., M/M/1); also see the more complicated queueing model in [23].

1.2 Basic Linear Regression and Designs

We define basic symbols and terminology used in the next sections, starting with *linear regression* (meta)models $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ where $\mathbf{y}$ denotes the n-dimensional vector with the dependent (explained) variable with n denoting the number of different simulated input combinations; $\mathbf{X} = (x_{i;g})$ is the $n \times q$ matrix of independent (explanatory) regression variables with $x_{i;g}$ the value of x_g in combination i ($i = 1, \ldots, n$; $g = 1, \ldots, q$), so row i of $\mathbf{X}$ is $\mathbf{x}_i = (x_{i;1}, \ldots, x_{i;q})$; $\boldsymbol{\beta}$ is the q-dimensional vector with regression parameters; $\mathbf{e}$ is the n-dimensional vector with residuals, so $\mathbf{e} = E(\mathbf{y}) - E(\mathbf{w})$ with $\mathbf{w}$ denoting the n-dimensional vector with $w_i = f_{sim}(\mathbf{z}_i, \mathbf{r}_i)$ where $\mathbf{z}_i$ denotes combination i of the k original simulation inputs that are determined by the $n \times k$ design matrix $\mathbf{D} = (d_{i;j})$, and $\mathbf{r}_i$ denotes the vector with PRNs used in combination i; row i of $\mathbf{D}$ is $\mathbf{d}_i$; x is a simple function of the original z or the standardized d.

We focus on a special case of linear regression, namely a *second-order polynomial* with k simulation inputs: $y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{j=1}^{k} \sum_{j' \geq j}^{k} \beta_{j;j'} x_j x_{j'} + e$ with the intercept β_0, the k first-order effects β_j ($j = 1, \ldots, k$), the $k(k-1)/2$ two-factor interactions (cross products) $\beta_{j;j'}$ ($j < j'$), and the k purely quadratic effects $\beta_{j;j}$. This metamodel is nonlinear in $\mathbf{x}$, but it is linear in $\boldsymbol{\beta}$; engineers call this metamodel nonlinear, whereas statisticians call it linear.

We assume that interactions among three or more inputs are unimportant; such interactions are hard to interpret. Of course, we should check this assumption; i.e., we should "validate" the estimated metamodel.

The *least squares* (LS) estimator of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{w}$. If $\mathbf{d}_i$—which determines $\mathbf{x}_i$—is simulated m_i times and m_i is a constant m, then we may replace $\mathbf{w}$ by $\overline{\mathbf{w}}$ with the n elements $\overline{w}_i = \sum_{r=1}^{m} w_{i;r}/m$ so $\mathbf{X}$ is indeed an $n \times q$ matrix. Usually, $m > 1$ in random simulation. If m_i is not a constant, then $\mathbf{x}_i$ is repeated m_i times within $\mathbf{X}$, so $\mathbf{X}$ has $\sum_{i=1}^{n} m_i$ rows and q columns.

Actually, $\hat{\boldsymbol{\beta}}$ is identical to the *maximum likelihood estimator* (MLE) if $\mathbf{e}$ is *white noise*; i.e., e_i is *normally, independently, and identically distributed* (NIID) with zero mean and constant variance σ_e^2. If the metamodel is valid, then $\sigma_e^2 = \sigma_w^2$ where σ_w^2 denotes the variance of w. The white-noise assumption implies that $\hat{\boldsymbol{\beta}}$ has the covariance matrix $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}} = (\mathbf{X}'\mathbf{X})^{-1}\sigma_w^2$. Because σ_w^2 is unknown, we estimate $\sigma_w^2 = \sigma_e^2$ through the *mean squared residuals* $\text{MSR} = (\widehat{\mathbf{y}} - \mathbf{w})'(\widehat{\mathbf{y}} - \mathbf{w})/(n-q)$ where $\widehat{\mathbf{y}} = \mathbf{X}\hat{\beta}'$ and $n - q > 0$, which gives $\widehat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}}$.

To derive *confidence intervals* (CIs) and *tests* for the individual elements of $\hat{\boldsymbol{\beta}}$, we use the estimated standard deviations $s(\widehat{\beta}_g)$ that are the square roots of $s^2(\widehat{\beta}_g)$ (estimate of $\text{Var}(\widehat{\beta}_g)$) on the main diagonal of $\widehat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}}$. This gives the following t-statistic with $n - q$ degrees of freedom: $t_{n-q} = (\widehat{\beta}_g - \beta_g)/s(\widehat{\beta}_g)$ with $g = 1, \ldots, q$.

To select a specific design matrix $\mathbf{D}$ with $\mathbf{d}_i$ in a given experimental area, we minimize $\text{Var}(\widehat{\beta}_g)$; i.e., we select an orthogonal $\mathbf{X}$. Usually, *design of experiments* (DOE) assumes that the z_j are *standardized* (scaled) such that $-1 \leq d_{i;j} \leq 1$. If z_j has only two values in the experiment with n input combinations, then this standardization uses $d_{i;j} = (z_{i;j} - \overline{z}_j)/[(H_j - L_j)/2]$ ($i = 1, \ldots, n$; $j = 1, \ldots, k$) where L_j denotes the lower value of z_j in the experiment, H_j the higher value, $\overline{z}_j$ the average value of z_j in a *balanced* experiment with z_j observed at L_j in $n/2$ combinations. If $\mathbf{X}$ is *orthogonal*, then $\mathbf{X}'\mathbf{X} = n\mathbf{I}$ so $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}} = (n\mathbf{I})^{-1}\sigma_w^2 = \mathbf{I}\sigma_w^2/n$. Hence the q estimators in $\widehat{\beta}$ are statistically independent and have the same variance. So, the $s^2(\widehat{\beta}_g)$ are constant, and we can *rank* x_g using either $\widehat{\beta}_g$ or t_{n-q} with $\beta_g = 0$ so $t_{n-q} = \widehat{\beta}_g/s(\widehat{\beta}_g)$.

Now we discuss designs of different *resolution* (R); e.g., R-III means "resolution III". Initially, we assume $m_i = 1$ ($i = 1, \ldots, n$). A *R-III or Plackett–Burman* (P–B) design gives unbiased estimators of β_j ($j = 1, \ldots, k$) if a first-order polynomial is a valid metamodel. A subclass are *fractional factorial two-level* 2_{III}^{k-p} designs with integer p such that $0 \leq p < k$ and $n = 2^{k-p} \geq 1 + k$. In a R-III design, n is a multiple of 4; e.g., $8 \leq k \leq 11$ implies $n = 12$. If $n > k + 1$, then we ignore some columns of $\mathbf{D}$. If $n = k + 1$, then $\mathbf{D}$ is *saturated*; the MSR is then undefined. To compute MSR, we may then add one or more combinations; e.g., either combinations from the 2^k

design excluding the combinations in the original saturated **D** or the combination at the *center* of the experimental area where $d_j = 0$ if d_j is quantitative, and d_j is randomly selected as -1 or 1 if d_j is qualitative with two values.

A *R-IV design* gives unbiased estimators of β_j in a first-order polynomial if *two-factor interactions* are nonzero but "higher-order" effects are zero: $\mathbf{x}_i = (1, d_{i;1}, \ldots, d_{i;k}, d_{i;1}d_{i;2}, \ldots, d_{i;k-1}d_{i;k})$. To construct a R-IV design, we apply the *foldover theorem*; i.e., we augment a R-III design **D** with its *mirror* design $-\mathbf{D}$. A R-IV design does not enable unbiased estimators of all the *individual* two-factor interactions; e.g., $k = 7$ implies $n = 2^{7-4} \times 2 = 16$ so $n < q = 1 + 7 + 21 = 29$; consequently, $\mathbf{X}'\mathbf{X}$ is singular, so the LS estimator does not exist.

A *R-V design* enables LS estimation of β_j, $\beta_{j;j'}$ with $j' > j$, and β_0 if all other effects (including $\beta_{j;j}$) are zero. Unfortunately, 2_V^{k-p} designs imply $n \gg q$. *Rechtschaffner* designs include saturated R-V designs, but they are not orthogonal; see [12, p. 62].

A *central composite design* (CCD) enables LS estimation of all the effects in a second-order polynomial if all higher-order effects are zero. A CCD consists of (i) a R-V design; (ii) the *central* combination (say) $\mathbf{0}_k'$; (iii) the $2k$ *axial* combinations, which form a *star design*; see [12, p. 63–66]. CCDs have non-orthogonal **X**, and $n \gg q$.

1.3 Assumptions Versus Practice

The *classic* statistical assumptions stipulate a single type of simulation output and white noise. A practical simulation model, however, may give *multivariate* output, and the univariate output w_i ($i = 1, \ldots, n$) may be *non-normal* with *heterogeneous* variances; w_i and $w_{i'}$ ($i, i' = 1, \ldots, n$) are correlated if the simulation uses *common random numbers* (CRN); $E(e)$ may be nonzero. In this section, we examine: (a) How realistic are the classic assumptions? (b) How can we test these assumptions? (c) Can we transform the simulation's I/O data such that the assumptions hold for the transformed data? (d) Which other statistical methods can we apply?

Multivariate Simulation Output

We assume that for r-variate simulation output with $r \geq 1$, we use r univariate linear regression metamodels, and these metamodels are polynomials of the same order (e.g., second-order):

$$\mathbf{y}^{(l)} = \mathbf{X}\beta^{(l)} + \mathbf{e}^{(l)} \text{ with } l = 1, \ldots r \tag{1.1}$$

where the various symbols are defined analogously to the univariate model (e.g., $\mathbf{y}^{(l)}$ is the dependent variable corresponding with simulation output of type l); the $\mathbf{e}^{(l)}$ have variances that may vary with l, and $e_i^{(l)}$ and $e_i^{(l')}$ ($l' = 1, \ldots, r$) are not independent. However, [21] proves that LS per output still gives the *best linear unbiased estimator* (BLUE): $\widehat{\beta}^{(l)} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{w}^{(l)}$. We can easily obtain CIs and tests for the elements in

$\widehat{\beta}^{(l)}$, using the classic formulas. We do not know any *general* designs for multivariate output; also see [11].

Non-normality

The normality assumption often holds *asymptotically*: if the simulation run is "long," then the sample average of the autocorrelated observations is "nearly" normal. Estimated quantiles, however, may be very non-normal. The t-statistic is quite insensitive to non-normality, whereas the F-statistic is not. It seems prudent to test the normality assumption as follows.

We may use various *residual plots* and *goodness-of-fit statistics* (e.g., a chi-square statistic). A basic assumption of these statistics is that the observations are identically and independently distributed (IID). We may, therefore, obtain "many" (say, 100) replications for a specific input combination (e.g., the base scenario) if the simulation is not computationally expensive; otherwise, these statistical tests lack power and the plots are too rough.

Actually, the white noise assumption concerns e, not w. Given $m_i \geq 1$ ($i = 1, \ldots, n$) replications, we obtain $\overline{w}_i = \sum_{r=1}^{m_i} w_{i;r}/m_i$ and $\widehat{\overline{e}}_i = \widehat{y}_i - \overline{w}_i$. For simplicity of presentation, we assume $m_i = m$. If $w_{i;r}$ has a constant variance σ_w^2, then $\overline{w}_i$ also has a constant variance $\sigma_{\overline{w}}^2 = \sigma_w^2/m$. Even if $\overline{w}_i$ is independent of $\overline{w}_{i'}$ with $i \neq i'$ (no CRN), then

$$\mathbf{\Sigma}_{\widehat{\overline{\mathbf{e}}}} = [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\sigma_{\overline{w}}^2, \tag{1.2}$$

so $\widehat{\overline{e}}_i$ does not have constant variance, and $\widehat{\overline{e}}_i$ and $\widehat{\overline{e}}_{i'}$ are correlated. This complicates the interpretation of the popular plot with estimated residuals.

We may apply *normalizing transformations*; e.g., $\log(w)$ may be more normally distributed than w. Unfortunately, the metamodel now explains the behavior of the transformed output (not the original output); see [12, p. 93] and [15].

Another transformation is *jackknifing*, which may (i) give CIs for non-normal observations, or (ii) reduce bias of a given estimator. Suppose we want CIs for the q elements of $\boldsymbol{\beta}$, for non-normal w. For simplicity, we assume $m_i = m > 1$. The original LS estimator is $\hat{\boldsymbol{\beta}}$; jackknifing deletes replication r for each combination i, so

$$\widehat{\boldsymbol{\beta}}_{-r} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\overline{\mathbf{w}}_{-r} \ (r = 1, \ldots, m) \tag{1.3}$$

where $\overline{\mathbf{w}}_{-r} = (\overline{w}_{i;-r})$ with $\overline{w}_{i;-r}$ denoting the average of the $m - 1$ simulation outputs excluding the output of replication r. Let us focus on β_q (last element of $\boldsymbol{\beta}$). The *pseudovalue* is

$$J_r = m\widehat{\beta}_q - (m-1)\widehat{\beta}_{q;-r}. \tag{1.4}$$

In this example, both $\widehat{\beta}_q$ and $\widehat{\beta}_{q;-r}$ are unbiased, so the m pseudovalues also remain unbiased. In general, however, possible bias is reduced by the *jackknife point estimator* $\overline{J} = \sum_{r=1}^{m} J_r/m$; an example of a biased estimator is (1.7). Jackknifing gives a CI, treating the J_r as if they were NIID. So if $t_{m-1;1-\alpha/2}$ denotes the $1 - \alpha/2$ quantile of the t_{m-1}-distribution and $\widehat{\sigma}_{\overline{J}}^2$ denotes $\sum_{r=1}^{m}(J_r - \overline{J})^2/[m(m-1)]$, then the two-sided symmetric $(1-\alpha)$ CI for β_q is $\overline{J} - t_{m-1;1-\alpha/2}\widehat{\sigma}_{\overline{J}} < \beta_q < \overline{J} +$

$t_{m-1;1-\alpha/2}\widehat{\sigma_{\overline{J}}}$. Many applications of jackknifing in simulation are given in [8] and [12, p. 95].

Another statistical method that does not assume normality is *distribution-free bootstrapping*; also see [25]. This bootstrapping may be used not only for non-normal distributions, but also for nonstandard statistics. We distinguish between the *original* w and the *bootstrapped* w^* with the usual superscript $*$ for bootstrapped observations. Standard bootstrapping assumes that the w observations are IID; indeed, $w_{i;1}, \ldots, w_{i;m}$ are IID because the m replications use non-overlapping PRN streams. We *resample with replacement* from the m original IID observations $w_{i;r}$ such that the original sample size remains m; we apply this resampling to each simulated combination, obtaining $w^*_{i;1}, \ldots, w^*_{i;m}$. This gives $\overline{\mathbf{w}}^* = (\overline{w_i^*})$, so

$$\widehat{\boldsymbol{\beta}}^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\overline{\mathbf{w}}^*.$$

To reduce sampling variation, bootstrapping repeats this resampling B times; a typical value for this *bootstrap sample size* B is 100 or 1,000. This B gives $\widehat{\boldsymbol{\beta}}^*_b$ with $b = 1, \ldots, B$. The *percentile method* gives a non-symmetric two-sided $(1-\alpha)$ CI:

$$P(\widehat{\beta}^*_{q;(B\alpha/2)} < \beta_q < \widehat{\beta}^*_{q;(B[1-\alpha/2])}) = 1 - \alpha \tag{1.5}$$

where $\widehat{\beta}^*_{q;(B\alpha/2)}$ denotes the $\alpha/2$ quantile of the *empirical density function* (EDF) of $\widehat{\beta}^*_q$ obtained through the *order statistics* denoted by the subscript $(\cdot)$ where (for simplicity) we assume that $B\alpha/2$ is integer; an analogous definition holds for $\widehat{\beta}^*_{q;(B[1-\alpha/2])}$. We shall also mention bootstrapped CIs for quantiles, R^2, and cross-validation.

Heterogeneous Variances of Simulation Outputs

In practice, $\text{Var}(w_i)$ changes as the input combination i changes. Unfortunately, $\text{Var}(w_i)$ is unknown; so if $m_i > 1$, then we compute $s_i^2 = \sum_{r=1}^{m_i}(w_{i;r} - \overline{w}_i)^2/(m_i - 1)$.This s_i^2 itself has high variance (e.g., if $w_{i;r}$ is normally distributed with $\text{Var}(w_{i;r}) = \sigma_i^2$, then $\text{Var}(s_i^2) = 2\sigma_i^4/m_i$). To compare n estimators s_i^2, we may apply various tests; see [12, p. 101].

The transformation $\log(w)$ may be used not only to obtain Gaussian output but also to obtain constant variances. Actually, this transformation is a special case of the normalizing Box–Cox power transformation; see [12, p. 93]. Anyhow, we prefer to accept variance heterogeneity, and to adapt our analysis, as follows.

If $E(\mathbf{e}) = 0$, then $\hat{\boldsymbol{\beta}}$ is still *unbiased*. However, $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}}$ then becomes

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\overline{\mathbf{w}}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \tag{1.6}$$

where $m_i = m$ so $\boldsymbol{\Sigma}_{\mathbf{w}}$ is the diagonal matrix with elements σ_i^2/m.

Alternatively, we might switch from LS to *weighted LS* (WLS), which gives $\widetilde{\boldsymbol{\beta}}$. In practice, however, $\text{Var}(w_i)$ is estimated, and using s_i^2 in WLS gives *estimated WLS* (EWLS), which gives the nonlinear estimator $\widehat{\widetilde{\boldsymbol{\beta}}}$. Obviously, $\widehat{\widetilde{\boldsymbol{\beta}}}$ is non-normally distributed and may be biased, so it is difficult to derive exact CIs. Above, we have

already discussed a simple solution, jackknifing; in *jackknifed EWLS* (JEWLS) with $m_i = m$ and without CRN, we proceed analogously to (1.3):

$$\hat{\hat{\beta}}_{-r} = (\mathbf{X}\hat{\boldsymbol{\Sigma}}^{-1}_{\overline{\mathbf{w}};-r}\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}_{\overline{\mathbf{w}};-r}\overline{\mathbf{w}}_{-r} \ (r = 1, \ldots, m) \tag{1.7}$$

where $\overline{\mathbf{w}}_{-r}$ is the vector with the n averages of the $m-1$ replications after deleting replication r, and $\hat{\boldsymbol{\Sigma}}_{\overline{\mathbf{w}};-r}$ is the diagonal matrix with $s^2_{i;-r}$ computed from the same $m-1$ replications. Using $\hat{\hat{\boldsymbol{\beta}}}$ and $\hat{\hat{\boldsymbol{\beta}}}_{-r}$, we compute the pseudovalues that give the desired CI.

The DOE literature ignores *designs* for heterogeneous output variances. We propose two-stage designs with m_i such that the resulting $\widehat{Var}(\overline{w}_i) = s_i^2/m_i$ $(i = 1, \ldots, n)$ are approximately constant; see [12, p. 105–106]. Actually, these designs use classic designs with an appropriate *relative* number of replications $\widehat{m}_i/\widehat{m}_{i'}$. To select absolute numbers $\widehat{m}$, we recommend [17, p. 505]'s rule-of-thumb with a user-specified relative estimation error r_{ee}:

$$\widehat{m} = min\left[r \geq m : \frac{t_{r-1;1-\alpha/2}\sqrt{s_i^2(m)/i}}{|\overline{w}(m)|} \leq \frac{r_{ee}}{1+r_{ee}}\right]. \tag{1.8}$$

We shall return to the selection of m_i, in Sect. 1.5.

Common Random Numbers

CRN are meant to compare the outputs of different simulation input combinations while all other "circumstances" are the same. CRN are the *default* in software for discrete event simulation. If $m_i = m$, then we can arrange $w_{i;r}$ $(i = 1, \ldots, n; r = 1, \ldots, m)$ into a matrix $\mathbf{W} = (w_{i;r}) = (\mathbf{w}_1, \ldots, \mathbf{w}_m)$ with $\mathbf{w}_r = (w_{1;r}, \ldots, w_{n;r})'$. CRN create correlation between $w_{i;r}$ and $w_{i';r}$. Two different replications use non-overlapping PRN streams, so $w_{i;r}$ and $w_{i;r'}$ with $r \neq r'$ are independent; i.e., $\mathbf{w}_r$ and $\mathbf{w}_{r'}$ are independent. The final goal of CRN is to reduce $Var(\widehat{\beta}_g)$ and $Var(\widehat{y})$; actually, CRN increase $Var(\widehat{\beta}_0)$. CRN implementation in MATLAB is discussed in [15].

If we use CRN and LS, then $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}}$ is given by (1.6) but now $\boldsymbol{\Sigma}_{\mathbf{w}}$ is not diagonal. $\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}$ is *singular* if $m \leq n$; else we may compute CIs for $\widehat{\beta}_j$ from t_{m-1}. An alternative method requires only $m > 1$:

$$\hat{\boldsymbol{\beta}}_r = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{w}_r \ (r = 1, \ldots, m)$$

where $\mathbf{w}_r$ has n elements that are correlated because of CRN and may have different variances. Furthermore, $\hat{\boldsymbol{\beta}}_r$ has q elements $\widehat{\beta}_{g;r}$ with variance $\sigma^2(\widehat{\beta}_{g;r})$ for any r. These $\widehat{\beta}_{g;r}$ give $\overline{\widehat{\beta}}_g = \sum_{r=1}^m \widehat{\beta}_{g;r}/m$ and $s^2(\overline{\widehat{\beta}}_g) = \sum_{r=1}^m (\widehat{\beta}_{g;r} - \overline{\widehat{\beta}}_g)^2/[m(m-1)]$, which give $t_{m-1} = (\overline{\widehat{\beta}}_g - \beta_g)/s(\overline{\widehat{\beta}}_g)$ with $g = 1, \ldots, q$. We cannot apply this alternative when estimating a *quantile*. We then recommend distribution-free bootstrapping; see [12, pp. 99, 110] and [16].

Validation of Metamodels

To test whether $E(e) = 0$, we may use (i) coefficients of determination; (ii) cross-validation. We explain (i) and (ii) next.

(i) R^2 is defined as

$$R^2 = \frac{\sum_{i=1}^{n}(\widehat{y}_i - \overline{\overline{w}})^2}{\sum_{i=1}^{n}(\overline{w}_i - \overline{\overline{w}})^2} = 1 - \frac{\sum_{i=1}^{n}(\widehat{y}_i - \overline{w}_i)^2}{\sum_{i=1}^{n}(\overline{w}_i - \overline{\overline{w}})^2} \tag{1.9}$$

where $\overline{\overline{w}} = \sum_{i=1}^{n}\overline{w}_i/n$ and $m_i \geq 1$. If $n = q$, then $R^2 = 1$ even if $\widehat{e}_i \neq 0$. If $n > q$ and q increases, then R^2 increases, whatever the size of $|\widehat{e}_i|$ is. Because of possible *overfitting* when q increases, we adjust R^2:

$$R^2_{\text{adj}} = 1 - \frac{n-1}{n-q}(1 - R^2). \tag{1.10}$$

Critical values for R^2 or R^2_{adj} are unknown, because these statistics do not have classic distributions. So we may use bootstrapping; see [12, p. 114].

(ii) *Leave-one-out cross-validation* may be defined as follows. For ease of presentation, we suppose that $\mathbf{X}$ has n rows: if $m_i = m \geq 1$, then we replace $\mathbf{w}$ by $\overline{\mathbf{w}}$ in the LS estimator. Now we delete I/O combination i to obtain $(\mathbf{X}_{-i}, \overline{\mathbf{w}}_{-i})$, which gives

$$\widehat{\boldsymbol{\beta}}_{-i} = (\mathbf{X}'_{-i}\mathbf{X}_{-i})^{-1}\mathbf{X}'_{-i}\overline{\mathbf{w}}_{-i} \ (i = 1, \ldots, n). \tag{1.11}$$

This gives $\widehat{y}_{-i} = \mathbf{x}'_i\widehat{\boldsymbol{\beta}}_{-i}$. We may "eyeball" the *scatterplot* with $(\overline{w}_i, \widehat{y}_{-i})$ and decide whether $E(e) = 0$. If $m_i = m > 1$, then [12, pp. 115–120] uses the *Studentized prediction error* $t^{(i)}_{m-1} = (\overline{w}_i - \widehat{y}_i)/[s^2(\overline{w}_i) + s^2(\widehat{y}_{-i})]^{1/2}$.

We may be interested not only in the predictive performance of the metamodel, but also in its *explanatory* performance; i.e., do the n estimates $\widehat{\boldsymbol{\beta}}_{-i}$ in (1.11) remain stable?

Related to cross-validation are several *diagnostic* statistics; most popular is the *prediction sum of squares* (PRESS) $\sum_{i=1}^{n}(\widehat{y}_{-i} - w_i)^2/n]^{1/2}$. Regression software uses a shortcut to avoid the n recomputations in cross-validation. We may apply bootstrapping to estimate the distribution of these validation statistics; see [3].

If $\widehat{e}$ is big, then we may consider various *transformations*. We may replace y and x_j by $\log(y)$ and $\log(x_j)$ $(j = 1, \ldots, k)$ so that the first-order polynomial approximates relative changes through *elasticity coefficients*. If we assume that f_{sim} is monotonic, then we may replace w and x_j by their ranks: *rank regression*. In the preceding subsections, we considered transformations that make w nearly normal with constant variance; unfortunately, different goals of a transformation may conflict with each other.

In Sect. 1.2, we discussed designs for low-order polynomials. If such a design does not give a valid metamodel, then we do not recommend routinely adding higher-order terms: these terms are hard to interpret. However, if the goal is not to better *understand* the simulation model but to better *predict* its output, then we may add higher-order

terms; e.g., a 2^k design enables the estimation of the interactions among two or more inputs. In the discussion of (1.10), we have already mentioned the danger of overfitting. Adding more explanatory variables is called *stepwise regression*; eliminating nonsignificant variables is called *backwards elimination*.

1.4 Factor Screening: Sequential Bifurcation

Screening means searching for the really important simulation inputs among the many inputs that can be varied in a simulation experiment. *Sparsity* means that only a few inputs among these many inputs are important. Indeed, the *Pareto* principle or *20–80* rule states that only "a few" inputs (20%) are important; e.g., [12, p. 136] presents two examples, with 281 and 92 inputs, respectively; screening finds only 15 and 11 inputs to be important.

There are several types of screening designs; see [12, pp. 137–139] and [29]. We focus on designs that treat the simulation model as a *black box*: only the I/O of the simulation model is observed. We focus on *sequential bifurcation* (SB), because SB is very efficient and effective if its assumptions are satisfied. SB selects the next input combination after analyzing the preceding I/O data, so SB is indeed sequential. SB is *customized*; i.e., SB accounts for the specific simulation model.

To explain the basic SB idea, we assume *deterministic* simulation and a valid *first-order* polynomial metamodel so $\beta_{j:j'} = 0$ with $j \leq j'$. Let γ_j denote the first-order effect of z_j (original scale). SB assumes that the *sign* of γ_j is known so that we can define the low and high bounds L_j and H_j of z_j such that $\gamma_j \geq 0$. Hence, we may rank the inputs such that the most important input has $\max_j \gamma_j$; the least important input has $\min_j \gamma_j \downarrow 0$. Changing z_j from L_j to H_j makes w change by $(H_j - L_j)\gamma_j = 2\beta_j$; also see [12, pp. 41–44] . SB calls z_j *important* if $2\beta_j \geq c_w$ where the users specify the threshold c_w (≥ 0).

In step 1, SB aggregates all k inputs into a single group and checks whether or not that group has an important effect. Let $w(\mathbf{L}_k)$ denote w with $\mathbf{z}_k = \mathbf{L}_k = (L_1, \ldots, L_k)'$ where $\mathbf{z}_k = (z_1, \ldots, z_k)'$; likewise, $w(\mathbf{H}_k)$ denotes w with $\mathbf{z}_k = \mathbf{H}_k = (H_1, \ldots, H_k)'$. So, SB obtains $w(\mathbf{L}_k)$ and $w(\mathbf{H}_k)$. If $\exists j : \beta_j > 0$, then $w(\mathbf{L}_k) < w(\mathbf{H}_k)$. It may happen that $\forall j : \beta_j < c_w/2$, but $w(\mathbf{H}_k) - w(\mathbf{L}_k) > c_w$; SB will discover this "false importance" in its next steps.

Assume that at least one input is important, so $w(\mathbf{H}_k) - w(\mathbf{L}_k) > c_w$. Then in step 2, SB splits the input group into two subgroups: *bifurcation*. Let k_1 and k_2 denote the size of subgroup 1 and subgroup 2 (so $k_1 + k_2 = k$). Then SB obtains $w(\mathbf{H}_{k_1})$. If $w(\mathbf{H}_{k_1}) - w(\mathbf{L}_k) < c_w$, then none of the individual inputs in subgroup 1 is important so SB *eliminates* this subgroup from further experimentation. If $w(\mathbf{H}_k) - w(\mathbf{H}_{k_1}) \geq c_w$, then one or more individual inputs in subgroup 2 may be important.

In each following step, SB splits important subgroups into smaller subgroups and eliminates unimportant subgroups. SB may find both subgroups to be important, so SB further experiments with two important subgroups in parallel. Obviously,

these steps give smaller subgroups; in the final steps, SB identifies and estimates all individual inputs that are not in eliminated (unimportant) subgroups.

Assuming $\beta_j \geq 0$ ensures that the β_j within an input group do not cancel each other. In practice, the users often do know the signs of β_j. Nevertheless, if in a specific case it is hard to specify the signs of a few specific inputs, then we should not group these inputs with the other inputs (with known signs). We should treat these inputs *individually* and investigate these inputs not through SB but through a classic design. This seems safer than assuming a negligible probability of cancelation within a subgroup.

The *efficiency* of SB improves if the individual inputs are labeled such that inputs are placed in increasing order of importance. Such labeling implies that the important inputs are *clustered*; i.e., these inputs are members of the same subgroup. The efficiency further improves when placing "similar" inputs within the same subgroup; e.g., place all "transportation" inputs in the same subgroup. Anyhow, splitting a group into subgroups of *equal* size is not necessarily optimal. Practical examples of SB are given in [12, pp. 136–172].

After explaining the basics of SB, we now assume *random* simulation and a *second-order* polynomial. Moreover, if $\beta_j = 0$, then $\beta_{j;j'} = 0$ ($j \leq j'$): *heredity* assumption. SB then applies the *foldover* principle (see Sect. 1.2); i.e., SB also simulates the mirror input of the original input, to estimate β_j unbiased by $\beta_{j;j'}$. In random simulation, SB may obtain a *fixed m* (number of replications) and use the t_{m-1}-statistic for a one-sided test of $\beta_j > 0$. Or SB obtains a *random m* and uses [28]'s *sequential probability ratio test* (SPRT) with user-selected thresholds $c_{w\mathrm{U}}$ and $c_{w\mathrm{I}}$ to classify inputs with $2\beta_j \leq c_{w\mathrm{U}}$ as *unimportant*, inputs with $2\beta_j \geq c_{w\mathrm{I}}$ as *important*, and remaining inputs as *intermediate*; see [12, pp. 154–159]. In practice, simulation models have *multiple response types*; see the multiresponse SB (MSB) in [12, pp.159–172]. Note that SPRTs for testing two means (instead of group effects in SB) are given in [15].

1.5 Kriging Metamodels and Their Designs

Kriging metamodels are fitted to simulation I/O data obtained for the *global* experimental areas instead of the *local* areas in RSM.

Ordinary Kriging in Deterministic Simulation

In this subsection, we focus on ordinary Kriging (OK), which is popular in deterministic simulation. OK assumes

$$y(\mathbf{x}) = \mu + M(\mathbf{x}) \tag{1.12}$$

where μ is the constant mean $E[y(\mathbf{x})]$ and $M(\mathbf{x})$ is a stationary GP with zero mean. A GP has covariances that depend only on the distance between the input combinations $\mathbf{x}$ and $\mathbf{x}'$. We call $M(\mathbf{x})$ the *extrinsic noise*, to distinguish it from the *intrinsic noise*

in stochastic simulation. Let $\mathbf{X}$ denote the $n \times k$ matrix with the n combinations $\mathbf{x}_i$ ($i = 1, \ldots, n$); in the preceding sections, we used $\mathbf{D}$, but the Kriging literature uses $\mathbf{X}$. Kriging software standardizes $\mathbf{z}_i$ to obtain $\mathbf{x}_i$ and also standardizes the simulation output w; for publications and Web sites see [12, p. 190].

OK uses the *best linear unbiased predictor* (BLUP) $\widehat{y}(\mathbf{x}_0)$ for the *new* combination $\mathbf{x}_0$:

$$\widehat{y}(\mathbf{x}_0) = \sum_{i=1}^{n} \lambda_i w_i = \boldsymbol{\lambda}' \mathbf{w}. \tag{1.13}$$

Such an "unbiased" predictor implies that if $\mathbf{x}_0 = \mathbf{x}_i$, then $\widehat{y}$ is an *exact interpolator*: $\widehat{y}(\mathbf{x}_i) = w(\mathbf{x}_i)$. This "best" predictor minimizes the *mean squared error* (MSE); because $\widehat{y}$ is unbiased, the MSE equals the variance $\text{Var}[\widehat{y}(\mathbf{x}_0)]$. Altogether, the *optimal* weight vector is

$$\lambda_o' = [\boldsymbol{\sigma}_M(\mathbf{x}_0) + \mathbf{1} \frac{1 - \mathbf{1}' \boldsymbol{\Sigma}_M^{-1} \boldsymbol{\sigma}(\mathbf{x}_0)}{\mathbf{1}' \boldsymbol{\Sigma}_M^{-1} \mathbf{1}}]' \boldsymbol{\Sigma}_M^{-1} \tag{1.14}$$

where $\boldsymbol{\Sigma}_M = (\text{cov}(y_i, y_{i'}))$ denotes the $n \times n$ matrix with the covariances between the metamodel's "old" outputs y_i, and $\boldsymbol{\sigma}_M(\mathbf{x}_0) = (\text{cov}(y_i, y_0))$ denotes the n-dimensional vector with the covariances between y_i and the new output y_0. The weight λ_i decreases with the *distance* between $\mathbf{x}_0$ and $\mathbf{x}_i$, so $\boldsymbol{\lambda}$ is not a constant vector (unlike $\boldsymbol{\beta}$ in regression). Substitution of $\boldsymbol{\lambda}_o$ into (1.13) gives

$$\widehat{y}(\mathbf{x}_0) = \mu + \boldsymbol{\sigma}_M(\mathbf{x}_0)' \boldsymbol{\Sigma}_M^{-1} (\mathbf{w} - \mu \mathbf{1}) \tag{1.15}$$

where $\mathbf{1}$ denotes an n-dimensional vector with all elements equal to 1. Obviously, $\widehat{y}(\mathbf{x}_0)$ in (1.15) varies with $\boldsymbol{\sigma}_M(\mathbf{x}_0)$, whereas μ, $\boldsymbol{\Sigma}_M$, and $\mathbf{w}$ remain fixed.

The *gradient* $\nabla(\widehat{y})$ follows from (1.15); see [19, Eq. 2.18]. We should not confuse $\nabla(\widehat{y})$ and $\nabla(w)$; sometimes we can indeed estimate $\nabla(w)$ and use $\widehat{\nabla}(w)$ to estimate a better OK model; see [12, pp. 183–184].

Defining $\tau^2 = \text{Var}(y)$ implies

$$\text{MSE}\,[\widehat{y}(\mathbf{x}_0)] = \tau^2 - \boldsymbol{\sigma}_M(\mathbf{x}_0)' \boldsymbol{\Sigma}_M^{-1} \boldsymbol{\sigma}_M(\mathbf{x}_0) + \frac{[1 - \mathbf{1}' \boldsymbol{\Sigma}_M^{-1} \boldsymbol{\sigma}_M(\mathbf{x}_0)]^2}{\mathbf{1}' \boldsymbol{\Sigma}_M^{-1} \mathbf{1}}. \tag{1.16}$$

This implies $\text{Var}[\widehat{y}(\mathbf{x}_0)] = 0$ if $\mathbf{x}_0 = \mathbf{x}_i$. So, $\text{Var}[\widehat{y}(\mathbf{x}_0)]$ has n local minima. $\text{Var}[\widehat{y}(\mathbf{x}_0)]$ has local maxima at $\mathbf{x}_0$ approximately halfway between old input combinations. Kriging gives bad extrapolations compared with interpolations (linear regression gives minimal $\text{Var}[\widehat{y}(\mathbf{x}_0)]$ when $\mathbf{x}_0 = \mathbf{0}$).

Obviously, (1.14) shows that $\boldsymbol{\lambda}_o$ is a function of $\boldsymbol{\Sigma}_M$ and $\boldsymbol{\sigma}_M(\mathbf{x}_0)$ or – switching to correlations $\boldsymbol{\Omega} = \tau^{-2} \boldsymbol{\Sigma}_M$ and $\boldsymbol{\rho}(\mathbf{x}_0) = \tau^{-2} \boldsymbol{\sigma}_M(\mathbf{x}_0)$. There are several types of correlation functions, but most popular is the *Gaussian* correlation function:

$$\rho(\mathbf{h}) = \prod_{j=1}^{k} \exp\left(-\theta_j h_j^2\right) = \exp\left(-\sum_{j=1}^{k} \theta_j h_j^2\right) \tag{1.17}$$

with distance vector $\mathbf{h} = (h_j)$ where $h_j = \left|x_{g;j} - x_{g';j}\right|$ and $g, g' = 0, 1, \ldots, n$. This $\rho(\mathbf{h})$ implies that $\boldsymbol{\lambda}_o$ assigns larger weights for $\mathbf{x}_i$ closer to $\mathbf{x}_0$. Standardization of the inputs affects $\mathbf{h}$.

When estimating the *Kriging parameters* $\boldsymbol{\psi} = (\mu, \tau^2, \boldsymbol{\theta}')'$ with $\boldsymbol{\theta} = (\theta_j)$, the MLE is most popular; yet LS (L_2 norm), cross-validation, and the L_1 norm are also used; see [13]. The estimation of $\boldsymbol{\psi}$ is challenging: different values may result from different software packages or from initializing the same package with different starting values. Anyhow, we denote the estimator of $\boldsymbol{\psi}$ by $\widehat{\boldsymbol{\psi}}$. *Plugging* $\widehat{\boldsymbol{\psi}}$ into (1.15) gives $\widehat{y}(\mathbf{x}_0, \widehat{\boldsymbol{\psi}})$. This $\widehat{y}(\mathbf{x}_0, \widehat{\psi})$ is a *nonlinear* predictor. In practice, we simply *plug* $\widehat{\boldsymbol{\psi}}$ into (1.16) to obtain MSE$[\widehat{y}(\mathbf{x}_0, \widehat{\boldsymbol{\psi}})]$; moreover, we ignore possible bias of $\widehat{y}(\mathbf{x}_0)$ so $s^2\{\widehat{y}(\mathbf{x}_0)\} =$ MSE$[\widehat{y}(\mathbf{x}_0, \widehat{\boldsymbol{\psi}})]$. To compute a CI, we use $\widehat{y}(\mathbf{x}_0, \widehat{\boldsymbol{\psi}})$, $s^2\{\widehat{y}(\mathbf{x}_0)\}$, and $z_{\alpha/2}$ ($\alpha/2$ quantile of standard normal):

$$P[w(\mathbf{x}_0) \in [\widehat{y}(\mathbf{x}_0, \widehat{\boldsymbol{\psi}}) \pm z_{\alpha/2} s\{\widehat{y}(\mathbf{x}_0)\}] = 1 - \alpha. \tag{1.18}$$

Universal Kriging (UK) replaces μ in (1.12) by a low-order polynomial. UK requires the estimation of additional parameters, besides $\beta_0 = \mu$; this may explain why UK often has a higher MSE than OK has.

Designs for Deterministic Simulation

There are several design types for Kriging in deterministic simulation; e.g., [12, p. 198] mentions orthogonal array, uniform, maximum entropy, minimax, maximin, integrated mean squared prediction error, and "optimal" designs. However, the most popular design uses *Latin hypercube sampling* (LHS). LHS assumes that an adequate metamodel is more complicated than a low-order polynomial; LHS does not assume a specific type of metamodel (e.g., an OK model), but focuses on the input space formed by x_j (standardized simulation inputs). LHS results in an $n \times k$ matrix $\mathbf{X}$. There is no strict mathematical relationship between n and k, whereas DOE may use $n = 2^{k-p}$. Nevertheless, if LHS keeps n "small" and k is "large," then "space filling" LHS covers the input space so sparsely that $E(e) \neq 0$. A rule-of-thumb for LHS in Kriging is $n = 10k$; see [18].

Mathematically, LHS divides the range of x_j into n mutually exclusive and exhaustive intervals of equal probability. The LHS design is *non-collapsing*: if an input turns out to be unimportant, then each remaining input still has one observation per interval. We conjecture that the estimation of the correlation function may benefit from this non-collapsing property. Unfortunately, projections of $\mathbf{x}$ onto more than one dimension may give "bad' designs, so there are maximin LHS, nearly orthogonal, and sliced LHS designs.

Instead of LHS with its *single-shot* design, we may use *sequential* designs that are application-driven or *customized*; i.e., they account for f_{sim}. In general, sequential procedures require fewer observations than fixed-sample procedures do, because we learn about the behavior of the underlying system as we experiment with this system and collect data (also see Sect. 1.4 on SB). Kriging, however, requires extra computer time if it re-estimates $\boldsymbol{\psi}$ when new I/O data become available.

We may use sequential Kriging designs for either *SA* (so the whole experimental area is interesting) or *optimization* (only the optimum is interesting); see [12, pp. 203–206] In a sequential design, we start with a *pilot* experiment with n_0 combinations of the k inputs selected through LHS and obtain the corresponding simulation I/O data. Next we fit a Kriging model to these data. Then we may consider—but not yet simulate—$\mathbf{X}_{cand}$ which denotes a larger matrix with *candidate* combinations selected through LHS and find the "winning" candidate. In SA, this winner has $\max_{\mathbf{x}} s^2\{\widehat{y}(\mathbf{x})\}$ with $\mathbf{x} \in \mathbf{X}_{cand}$. Next we use the winner as the input to be simulated, which gives additional I/O data. We re-fit the Kriging model to the augmented I/O data (usually re-estimating $\boldsymbol{\psi}$). We stop if either the Kriging model satisfies a given goal or the computer budget is exhausted. Altogether, the design selects relatively few combinations in subareas with an approximately linear f_{sim}.

Stochastic Kriging for Random Simulation

Stochastic Kriging (SK) was developed in [1], adding the *intrinsic noise* term $\varepsilon_r(\mathbf{x}_i)$ for replication r at combination $\mathbf{x}_i$ to (1.12), which —after averaging over replications—gives

$$\overline{y}(\mathbf{x}_i) = \mu + M(\mathbf{x}_i) + \overline{\varepsilon}(\mathbf{x}_i) \tag{1.19}$$

where $\varepsilon_r(\mathbf{x}) \in N(0, \text{Var}[\varepsilon_r(\mathbf{x})])$ and $\varepsilon_r(\mathbf{x})$ is independent of $M(\mathbf{x})$. Obviously, m_i replications without CRN make $\boldsymbol{\Sigma}_{\overline{\varepsilon}}$ diagonal with main diagonal elements $\text{Var}[\varepsilon(\mathbf{x}_i)]/m_i$; CRN and $m_i = m$ give $\boldsymbol{\Sigma}_{\overline{\varepsilon}} = \boldsymbol{\Sigma}_{\varepsilon}/m$.

To estimate $\text{Var}[\varepsilon(\mathbf{x}_i)]$, SK may use s_i^2. Alternatively, SK may use another Kriging model for $\text{Var}[\varepsilon(\mathbf{x}_i)]$ (besides the Kriging model for $E[y_r(\mathbf{x}_i)]$), which may give less volatile estimates. Because s_i^2 is not normally distributed, the GP is only a rough approximation. We might also replace s_i^2 by $\log(s_i^2)$ in the Kriging model; also see [10].

SK replaces $\boldsymbol{\Sigma}_M$ in OK by $\boldsymbol{\Sigma}_M + \boldsymbol{\Sigma}_{\overline{\varepsilon}}$ and $\mathbf{w}$ by $\overline{\mathbf{w}}$, giving $\widehat{y}(\mathbf{x}_0, \widehat{\psi})$ and $s^2\{\widehat{y}(\mathbf{x}_0)\}$; see [1, Eq. 25]. SK for a *quantile* (instead of an average) is discussed in [12, p. 208].

In our discussion of (1.18), we have already mentioned the problems caused by the randomness of $\widehat{\psi}$. If $m_i \gg 1$, then we may solve this problem through *distribution-free bootstrapping*; see [12, p. 209].

Usually SK employs the same designs as OK or UK do for deterministic simulation. So, SK often uses single-shot LHS. In random simulation, however, we also need to select m_i. Above, we discussed the analogous problem in regression metamodeling; a simple rule-of-thumb is (1.8).

In sequential designs for SA, we may select $\mathbf{x}$ that gives $\max_{\mathbf{x}} s^2\{\widehat{y}(\mathbf{x})\}$. In SK, we may find this $\mathbf{x}$ through distribution-free bootstrapping. This design selects more input values in subdomains with highly nonlinear estimated I/O functions.

More Kriging: Monotonic Kriging, and Global SA

Sometimes we know that f_{sim} is *monotonic* (e.g., if the traffic rate increases, then the mean waiting time increases); see Sect. 1.4. The Kriging predictor $\widehat{y}$, however, may be *wiggling* if the sample size n is small. To make $\widehat{y}$ monotonic, we may apply