Simulation Modelling Practice and Theory 81 (2018) 85–99

Contents lists available at ScienceDirect

Simulation Modelling Practice and Theory

journal homepage: www.elsevier.com/locate/simpat

Testing the assumptions of sequential bifurcation for factor screening

Wen Shi [a,*], Jack P.C. Kleijnen [b]

[a] Research Center of Hubei Logistics Development, Hubei University of Economics, Wuhan 430205, China
[b] Department of Management/Center for Economic Research (CentER), Tilburg University, LE Tilburg 5000, The Netherlands

ARTICLE INFO

Article history:
Received 25 January 2017
Revised 5 December 2017
Accepted 6 December 2017
Available online 13 December 2017

Keywords:
Sensitivity analysis
Experimental design
Meta-modelling
Validation
Regression
Simulation

ABSTRACT

Sequential bifurcation (or SB) is an efficient and effective factor-screening method; i.e., SB quickly identifies the important factors (inputs) in experiments with simulation models that have very many factors—provided the SB assumptions are valid. The specific SB assumptions are: (i) a second-order polynomial is an adequate approximation (a valid metamodel) of the input/output function of the underlying simulation model; (ii) the directions (signs) of the first-order effects are known (so the first-order polynomial approximation is monotonic); (iii) so-called "heredity" applies; i.e., if a specific input has a "small" first-order effect, then this input has "small" second order effects. Moreover, SB assumes Gaussian simulation outputs if the simulation model is stochastic (random). A generalization of SB called "multiresponse SB" (or MSB) uses the same assumptions, but allows multiple types of simulation responses (outputs). In this article, we develop heuristic practical methods for testing whether these assumptions hold, and we evaluate these methods through Monte Carlo experiments and a case study (namely, a Chinese logistics network).

1. Introduction

By definition, *factor screening*—or briefly *screening*—means searching for the really important factors—or inputs—among the many factors that can be varied in an experiment with a given simulation model (we shall define "important" below). For example, Bettonvil and Kleijnen [2] applies the screening method called "sequential bifurcation"—abbreviated to SB—to a case study, and finds that only 15 of the 281 inputs are really important. So, screening assumes that input effects are *sparse*; i.e., only a few inputs among the many inputs are really important. Related to this sparsity are the *Pareto* principle and the *20–80* rule, which implies that roughly 20% of the inputs account for 80% of the effect on the output. The*law of parsimony* and *Occam's razor* imply that a simpler explanation with fewer factors is better than a complex explanation with many factors—all other things being equal. Altogether, we conclude that there is really a need for screening in the design and analysis of experiments with practical simulation models.

Furthermore, we assume that the number of inputs (say) k is so large that*classic designs* cannot be applied. For example, a resolution-III (R-III) design requires an experiment with at least $k + 1$ input combinations to estimate the effects in a first-order polynomial with k inputs plus an intercept, assuming this polynomial provides an adequate (valid) approximation or *metamodel* (emulator, surrogate) of the simulation model (we provide many synonyms, because simulation is used in many

* Corresponding author.
E-mail addresses: shi3wen@163.com (W. Shi), kleijnen@uvt.nl (J.P.C. Kleijnen).

https://doi.org/10.1016/j.simpat.2017.12.003
1569-190X/

scientific disciplines, which have their own terminology). Higher-order polynomials require bigger designs; e.g., a second-order polynomial may be estimated through a central composite design (CCD), which has $1 + k + k(k-1)/2 + k$ input combinations. Kleijnen [6] discusses *design of experiments* (DOE), including classic designs—such as R-III designs and CCDs—and several types of screening designs—besides SB. A recent article in this journal that applies DOE including a 2^3 design and a CCD is Vasandani et al. [16]. A recent publication that discusses screening designs is Shi et al. [15].

In this article we focus on SB and its extension to simulation models with multiple responses; this extension is called *multiresponse SB* (MSB) in Shi et al. [13]. Actually, MSB includes SB as a special case; namely, a single response. For brevity's sake we shall write "MSB"—instead of "SB or MSB" or "MSB including SB"—if the context makes confusion unlikely. The goal of MSB is to identify the inputs that have important effects on one or more output (response) types among the $n \geq 1$ output types.

We consider the following problem. The given simulation model has so many inputs that classic designs require too much computer time, so the simulation analysts apply screening. Each type of screening design has its own assumptions. Because MSB is the most efficient screening design, the analysts apply MSB. In general we emphasize that after analysts have applied a statistical method to solve a given problem, these analysts should next examine the results to verify whether these results are not conflicting with the assumptions of the method. For example, the analysts apply linear regression to analyze a data set obtained through a simulation experiment; then they should next validate the estimated (fitted) regression model through the coefficient of determination R^2 and cross-validation (R^2 and cross-validation are detailed in Kleijnen [6, p. 112–121]. More specifically, *after* the analysts have applied MSB to find important inputs, they should verify whether these results do not conflict with the assumptions of MSB. For this verification we derive and evaluate several statistical tests in this article. These *post-screening* (follow-up) tests definitely require less experimentation than MSB requires; e.g., one test requires the simulation of only two (extreme) input combinations, to validate the first-order polynomial metamodel in SB (which assumes a single response type).

SB was originally developed in Bettonvil [3]'s dissertation and summarized in Bettonvil and Kleijnen [2]. Several authors extended SB; see the many references in Kleijnen [6], and also see Han et al. [4] and Martín and Sánchez [9]. To save space, we refer to the detailed description of SB and MSB that is given in [13]; for our article it suffices to detail the following *three specific MSB assumptions*:

1. *Second-order polynomials* provide valid metamodels:

$$y^{(l)} = \beta_0^{(l)} + \sum_{j=1}^{k} \beta_j^{(l)} x_j + \sum_{j=1}^{k-1} \sum_{j'=j+1}^{k} \beta_{j;j'}^{(l)} x_j x_{j'} + \sum_{j=1}^{k} \beta_{j;j}^{(l)} x_j^2 + e^{(l)} \quad (1)$$

where $y^{(l)}$ denotes the metamodel's predictor for simulation output l with $l = 1, \ldots, n$ and $n \geq 1$ ($n = 1$ in SB), x_j the standardized (coded, scaled) simulation input j ($j = 1, \ldots, k$) so $-1 \leq x_j \leq 1$ —if an original input is qualitative and has only two levels (so it is binary), then its levels are randomly associated with the standardized values -1 and 1—$\beta_0^{(l)}$ the intercept for output l, $\beta_j^{(l)}$ the first-order (or main effect) of x_j for output l, $\beta_{j;j'}^{(l)}$ the interaction between x_j and $x_{j'}$ for output l, $\beta_{j;j}^{(l)}$ the purely quadratic effect of x_j for output l, and $e^{(l)}$ the approximation error with zero mean for output l.

2. The $\beta_j^{(l)}$ have *known signs*, so that the low bound l_j and the upper bound u_j of the original (nonstandardized) input z_j can be defined such that all k first-order effects are nonnegative for one of the n output types—(say) output type 1 (the symbol l_j is an easy mnemonic for "low", but should not be confused with the symbol l in the superscript (l); we use the symbol (l) because Shi et al. [13] uses that symbol). This assumption implies $\beta_j^{(1)} \geq 0$ (the superscript is (1), not (l)). Without assumption 2, first-order effects may cancel each other within a group of individual inputs that is used in MSB. Note that this assumption implies that the first-order polynomial is monotonically increasing in x_j ($j = 1, \ldots, k$).

3. If input j has a "small" first-order effect on simulation output l, then this input has "small" second-order effects on this output; we shall define "small" later on (e.g., above eq. (4)). Wu and Hamada [19] calls this the *heredity* assumption; this assumption is also discussed in Woods and Lewis [20].

Many publications on screening discuss the plausibility of these three assumptions. We discuss formalized statistical tests of these assumptions; such tests are often neglected in the literature.

Note: If our tests reject these MSB assumptions, then MSB may still identify the important inputs; i.e., these assumptions are sufficient but not necessary. However, we consider it to be unlikely that these assumptions do not hold, but MSB still "works". (Hasty readers may skip paragraphs that start with "Note:", and still understand this article.)

Besides the preceding three specific MSB assumptions, MSB—like many other statistical methods—assumes that the simulation outputs have *normal* (Gaussian) distributions. Our tests also assume normality.

We organize the rest of this article as follows. In Section 2 we discuss the assumed normality of the simulation outputs. In Section 3 we detail our tests for the specific three MSB assumptions. In Section 4 we compare these tests through a Monte Carlo experiment that does satisfy all MSB assumptions. In Section 5 we compare these tests through a case study concerning a logistics system in China; obviously, case studies may violate one or more MSB assumptions. In Section 6 we summarize the major conclusions and sketch future research.