Available online at www.sciencedirect.com
ScienceDirect
Mathematics and Computers in Simulation 128 (2016) 42–54
MATHEMATICS AND COMPUTERS IN SIMULATION
www.elsevier.com/locate/matcom

Original articles

Global sensitivity analysis using sparse high dimensional model representations generated by the group method of data handling

Romain S.C. Lambert[a,*], Frank Lemke[b], Sergei S. Kucherenko[a], Shufang Song[a,c], Nilay Shah[a]

[a] Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, UK
[b] KnowledgeMiner Software, 13187 Berlin, Germany
[c] School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China

Received 11 September 2014; received in revised form 14 January 2016; accepted 12 April 2016
Available online 2 May 2016

Highlights

- The group method of data handling (GMDH) is used to construct the high dimensional model representation (HDMR) to calculate Sobol's first and second order global sensitivity analysis indices.
- This methodology uses the parameter selection features of GMDH to construct a sparse HDMR expansion for high dimensional problems from a limited number of function evaluations.
- By design, the method also allows for the optimal (i.e. balancing accuracy and complexity) polynomial order selection in the HDMR expansion.

Abstract

In this paper, the parameter selection capabilities of the group method of data handling (GMDH) as an inductive self-organizing modelling method are used to construct sparse random sampling high dimensional model representations (RS-HDMR), from which the Sobol's first and second order global sensitivity indices can be derived. The proposed method is capable of dealing with high-dimensional problems without the prior use of a screening technique and can perform with a relatively limited number of function evaluations, even in the case of under-determined modelling problems. Four classical benchmark test functions are used for the evaluation of the proposed technique.

Keywords: Global sensitivity analysis; High dimensional model representations; Sobol indices; Group method of data handling

1. Introduction

Over the last decade, global sensitivity analysis (GSA) has gained considerable attention among practitioners, due to its advantages over local sensitivity analysis methods [4,8,9,21,20,29], for example in the detection of parameter

* Corresponding author. Tel.: +44 0 207 594 6611.
E-mail address: rl508@ic.ac.uk (R.S.C. Lambert).

http://dx.doi.org/10.1016/j.matcom.2016.04.005

interactions. An eminent class of GSA techniques is that of variance based methods, which includes the well-known Sobol method of global sensitivity indices [23]. Sobol sensitivity indices are used to rank input parameters and to discard unessential parameters. One way to reduce the computational expense of performing a sensitivity analysis is the use of surrogate-models or meta-models, which emulate the behaviour of the original computationally expensive models. Various surrogate modelling methods such as gaussian process modelling [16,18], polynomial chaos expansion (PCE) [2,3,25], and random sampling-high dimensional model representations (RS-HDMR) [14,15] have been proposed. RS-HDMR was originally defined as a set of quantitative tools to map the input–output behaviour of high dimensional systems [19]. Recently this technique has become popular and widely used by practitioners (e.g. [5,6,28,30–32]). RS-HDMR has also been used as an efficient way to compute first and second order Sobol global sensitivity indices. Despite improvements over the direct Sobol method, there have been attempts to create methods that can more efficiently generate (sparse) HDMR expansions. One particularly successful technique proposed by Blatman and Sudret [2,3] consists of the calculation of the polynomial chaos expansion (PCE) via a least angle regression (LAR) using cross validation schemes. Other adaptive methods to efficiently calculate HDMR expansions using machine learning techniques have also been suggested in the literature [7,28,30]. Ziehn and co-workers proposed an approach to calculate the 'optimal' polynomial order in HDMR expansions [30,31]. For completeness the reader is also invited to refer to bias correction methods [13] for the calculation of sensitivity indices [26].

In this paper we propose a new alternative method using techniques from the class of the so-called inductive modelling methods, namely the group method of data handling (GMDH). The group method of data handling (GMDH) was originally developed by Ivakhnenko and co-workers [10–12]. It is based on the principle of inductive self-organization. Unlike many other machine-learning techniques, this method is inductive which means that it does not a priori postulate the structure of the expressions. During model self-organization, GMDH generates, validates, and selects many alternative networks of growing complexity (i.e. with increasing number of parameters, interactions between these parameters and/or nonlinearity) until an 'optimally' complex model has been found (i.e. when it begins to over-fit the design data). A class of elementary expressions is used, which by making them gradually more complex, can describe every possible instance of a sought general function. GMDH has the ability to perform efficiently with a limited number of function evaluations in high dimensional spaces (under-determined systems), by selecting important parameters in an adaptive fashion (feature selection). A second key principle developed and introduced by the GMDH inductive modelling theory in the 1970s, and subsequently adopted for use in neural networks and other machine learning methods, is the principle of integrating external information into modelling to allow the objective selection of a model of optimal complexity [17,24].

We propose to use the characteristics of GMDH to efficiently calculate a sparse HDMR expansion, and to subsequently calculate Sobol first and second order global sensitivity indices.

This paper is organized as follows: the mathematical fundamentals of global sensitivity analysis and HDMR are introduced in Section 2, the GMDH method and the combined GMDH-HDMR method are presented in Section 3. The developed methodology is then applied to well-known benchmark functions in Section 4.

2. Methodology and statistical fundamentals

2.1. Sobol method of global sensitivity analysis

Consider an integrable function f defined in the unit hypercube $[0, 1]^M$. This function can be decomposed as

$$f(x) = \sum_{\alpha \subseteq \{1,\dots,M\}} f_\alpha(x_\alpha). \tag{1}$$

Here α is a subset of indices from $\{1, \dots, M\}$. A generic point of $[0, 1]^M$ is noted $x = (x_1, \dots, x_M)$. $|\alpha|$ denotes the cardinality and x_α represents the $|\alpha|$-vector of components x_j, for $j \in \alpha$ Decomposition (1) is unique if

$$\forall\, i \epsilon \alpha, \quad \int_0^1 f_\alpha(x_\alpha)\, dx_i = 0, \tag{2}$$

in which case it is called the ANOVA decomposition. It follows from condition (2) that the ANOVA decomposition is orthogonal.