Computational Statistics and Data Analysis 55 (2011) 184–198

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Understanding and comparisons of different sampling approaches for the Fourier Amplitudes Sensitivity Test (FAST)

Chonggang Xu [a], George Gertner [b,*]

[a] Department of Entomology and Center for Quantitative Sciences in Biomedicine, North Carolina State University, Raleigh, NC 27695, USA
[b] Department of Natural Resources and Environmental Sciences, University of Illinois, Urbana, W523 Turner Hall 1102, South Goodwin Avenue, IL 61801, USA

ARTICLE INFO

Article history:
Received 12 August 2009
Received in revised form 23 June 2010
Accepted 23 June 2010
Available online 3 August 2010

Keywords:
Uncertainty analysis
Sensitivity analysis
Fourier Amplitude Sensitivity Test
Simple random sampling
Random balance design
Interactions

ABSTRACT

Fourier Amplitude Sensitivity Test (FAST) is one of the most popular uncertainty and sensitivity analysis techniques. It uses a periodic sampling approach and a Fourier transformation to decompose the variance of a model output into partial variances contributed by different model parameters. Until now, the FAST analysis is mainly confined to the estimation of partial variances contributed by the main effects of model parameters, but does not allow for those contributed by specific interactions among parameters. In this paper, we theoretically show that FAST analysis can be used to estimate partial variances contributed by both main effects and interaction effects of model parameters using different sampling approaches (i.e., traditional search-curve based sampling, simple random sampling and random balance design sampling). We also analytically calculate the potential errors and biases in the estimation of partial variances. Hypothesis tests are constructed to reduce the effect of sampling errors on the estimation of partial variances. Our results show that compared to simple random sampling and random balance design sampling, sensitivity indices (ratios of partial variances to variance of a specific model output) estimated by search-curve based sampling generally have higher precision but larger underestimations. Compared to simple random sampling, random balance design sampling generally provides higher estimation precision for partial variances contributed by the main effects of parameters. The theoretical derivation of partial variances contributed by higher-order interactions and the calculation of their corresponding estimation errors in different sampling schemes can help us better understand the FAST method and provide a fundamental basis for FAST applications and further improvements.

1. Introduction

Models are popular tools to help us understand and predict potential behaviors of different systems in physics, chemistry, biology, environmental sciences and social sciences. For a better use of the model and a better understanding of modeled systems, it is important that the model users quantify the overall amount of uncertainty in a model output (referred to as uncertainty analysis) and the importance of parameters in their contributions to uncertainty in the model output (referred to as sensitivity analysis). Sensitivity analysis techniques can be divided into two groups (Saltelli et al., 2000; Borgonovo et al., 2003): local sensitivity analysis methods and global sensitivity analysis methods. The local sensitivity analysis techniques examine the response of model output by varying model parameters one at a time around a local neighborhood of their

* Corresponding author. Tel.: +1 217 333 9346.
E-mail addresses: xuchongang@gmail.com (C. Xu), gertner@illinois.edu (G. Gertner).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2010.06.028

central values. The global sensitivity techniques examine the global response (i.e., response averaged over variations of all parameters) of model output by exploring a finite (or even an infinite) region. The local sensitivity analysis is easy to implement, but dependent on central values of parameters. In addition, the local sensitivity analysis is not able to estimate the amount of uncertainty in the model output. Thus, global sensitivity analysis methods are generally preferred over local sensitivity analysis.

The global sensitivity analysis generally encompasses two processes: (1) sampling of parameters values from defined probability density functions for parameters; and (2) quantification of uncertainties in the model output contributed by different parameters. Many uncertainty and sensitivity analysis techniques are now available (Saltelli et al., 2000, 2005). They include Fourier Amplitude Sensitivity Test (FAST) (Cukier et al., 1973; Schaibly and Shuler, 1973; Cukier et al., 1975, 1978); the screening methods (Morris, 1991; Saltelli et al., 1995; Henderson-Sellers and Henderson-Sellers, 1996; Cryer and Havens, 1999; Beres and Hawkins, 2001; Saltelli et al., 2009); regression-based methods (Helton, 1993; Helton and Davis, 2002, 2003; Helton et al., 2005, 2006); Sobol's method (Sobol, 1993); McKay's one-way ANOVA method (McKay, 1997); and moment independent approaches (Park and Ahn, 1994; Chun et al., 2000; Borgonovo, 2006, 2007; Borgonovo and Tarantola, 2008). The global sensitivity analysis techniques differ in their algorithms for sampling or uncertainty quantifications.

FAST is one of the most popular global sensitivity analysis techniques. It uses a periodic sampling approach and a Fourier transformation to decompose the variance of a model output into partial variances contributed by different model parameters. Ratios of partial variances to model output variance are used to measure the parameters' importance in their contributions to uncertainty in the model output. The theory of FAST was first proposed by Cukier et al. (1973, 1975, 1978). The traditional FAST analysis uses a periodic sampling approach to generate a search curve in the parameter space. The periodic sample of each parameter is assigned with a characteristic frequency (i.e., a distinct integer), which is used to determine the parameter's contribution to the variance of a model output based on a Fourier transformation. Koda et al. (1979) and McRae et al. (1982) provided the computational codes for the traditional FAST analysis. To reduce the estimation errors in the sensitivity indices, characteristic frequencies need to be selected based on certain criteria (see Section 2.2 for details), which could be difficult for models with many parameters. In view of that, Tarantola et al. (2006) introduced a random balance design sampling method to avoid the difficulty of selecting characteristic frequencies. FAST analysis is originally developed for models with independent parameters. In order to extend FAST for models with dependent parameters, Xu and Gertner (2007, 2008) introduced a random reordering approach to account for rank correlations among parameters.

FAST is computationally efficient and can be used for nonlinear and non-monotonic models. Thus, it has been widely applied in the uncertainty and sensitivity analysis of different models, such as chemical reaction models (Haaker and Verheijen, 2004); atmospheric models (Collins and Avissar, 1994; Rodriguez-Camino and Avissar, 1998; Kioutsioukis et al., 2004); nuclear waste disposal models (Lu and Mohanty, 2001); soil erosion models (Wang et al., 2001); hydrological models (Francos et al., 2003); matrix population models (Xu and Gertner, 2009); and forest landscape models (Xu et al., 2009).

Although the FAST method has been widely applied to different models, it is mainly confined to the estimation of partial variances contributed by the main effects of model parameters. For the traditional search-curve based sampling, it has been heuristically shown for simple test models that FAST can be used to estimate partial variances contributed by parameter interactions using linear combination of characteristic frequencies through the exploration of Fourier amplitudes at different frequencies (Saltelli et al., 1999). Based on that, Saltelli et al. (1999) proposed a frequency selection method to estimate the sum of partial variances contributed by a special type of interactions (i.e., all interactions involving a parameter of interest) using the traditional search-curve based sampling. However, there is a lack of theoretical understanding and no proof for the calculation of partial variances contributed by the interactions among parameters, which may hinder future development of FAST. Furthermore the heuristic understanding does not allow for the estimation of partial variances contributed by the interactions among specific parameters due to a lack of knowledge of potential errors and biases for the estimation of partial variances. Finally, the heuristic understanding is only based on the traditional search-curve based sampling approach. It is important that we can also calculate the partial variances contributed by interactions for new sampling approaches (e.g., the random balance design sampling), in view that it would be difficult to apply the traditional sampling to modern models with many parameters (e.g., 50 parameters).

In this paper, we provide a theoretical derivation of FAST for higher-order sensitivity indices and compare three sampling approaches for FAST (i.e., traditional search-curve based sampling, simple random sampling, and random balance design sampling). We also analytically calculate the potential errors and biases in the estimation of partial variances with different sampling approaches. Finally, we compare the performance of the three sampling approaches for a simple test model. The theoretical derivation of partial variances contributed by higher order interactions and the calculation of their corresponding estimation errors in different sampling schemes can help us better understand the FAST method and extend FAST to applications where interactions among model parameters are concerned.

The paper is organized as follows. Section 2.1 introduces the background of FAST and provides a theoretical derivation of FAST for first-order and higher-order sensitivity indices. Section 2.2 provides algorithms for calculating partial variances contributed by the main effects and interaction effects of parameters for the traditional search-curve based sampling. Section 2.3 proposes a simple random sampling approach and hypothesis tests to reduce estimation errors and biases for partial variances. Section 2.4 introduces random balance design sampling and determines potential estimation errors and biases for partial variance calculations. Section 3 compares the estimation errors and biases in three sampling approaches and provides a summary of procedures for FAST analysis. Section 4 shows the results of a comparison of FAST using different