



## Sensitivity analysis using contribution to sample variance plot: Application to a water hammer model

S. Tarantola<sup>a</sup>, V. Kopustinskas<sup>b,\*</sup>, R. Bolado-Lavin<sup>b</sup>, A. Kaliatka<sup>c</sup>, E. Ušpuras<sup>c</sup>, M. Vaišnoras<sup>c</sup>

<sup>a</sup> European Commission, Institute for Protection and Security of Citizen, Ispra, Italy

<sup>b</sup> European Commission, Institute for Energy and Transport, Petten, Netherlands

<sup>c</sup> Lithuanian Energy Institute, Kaunas, Lithuania

### ARTICLE INFO

#### Article history:

Received 29 March 2011

Received in revised form

13 October 2011

Accepted 16 October 2011

Available online 25 October 2011

#### Keywords:

Sensitivity analysis

Contribution to sample variance plot

Variance reduction

Water hammer

RELAP model

Ishigami function

### ABSTRACT

This paper presents “contribution to sample variance plot”, a natural extension of the “contribution to the sample mean plot”, which is a graphical tool for global sensitivity analysis originally proposed by Sinclair. These graphical tools have a great potential to display graphically sensitivity information given a generic input sample and its related model realizations. The contribution to the sample variance can be obtained at no extra computational cost, i.e. from the same points used for deriving the contribution to the sample mean and/or scatter-plots. The proposed approach effectively instructs the analyst on how to achieve a targeted reduction of the variance, by operating on the extremes of the input parameters’ ranges. The approach is tested against a known benchmark for sensitivity studies, the Ishigami test function, and a numerical model simulating the behaviour of a water hammer effect in a piping system.

© 2011 Elsevier Ltd. All rights reserved.

### 1. Introduction

Let us consider a simulation model represented by an input/output function  $Y=G(\mathbf{X})$ , where  $Y$  is a scalar model output and  $\mathbf{X}=(X_1, X_2, \dots, X_n)$  defines the generic vector of  $n$  input parameters. The values of the model parameters are not perfectly known; in other words, they are affected by some uncertainty. Therefore, although it is not the case, each input can be considered as a random variable characterized statistically by its probability density function  $p_i(X_i)$ . The model output  $Y$  can also be thought as a random variable; the estimation of its pdf is the objective of uncertainty analysis and is carried out by evaluating the function  $G(\mathbf{X})$  on a sample of points generated from the pdf  $p_i(X_i)$ . The analysis described here is in the context of deterministic models. Probabilities have only been introduced to represent imprecision about model parameter values.

Graphical sensitivity tools provide valuable information on the relationship between uncertain model inputs and model outputs. In 1989, Sacks et al. proposed the use of scatter plots, i.e. projections on a two-dimensional plane of the hyper-surface describing the input/output mapping [1]. A number of other graphical techniques are discussed in [2–5]. In 1993, Sinclair [6] introduced the contribution to the sample mean plot (CSM) which was further developed by Bolado-Lavin et al. [7].

The idea behind CSM is to use a given random sample of the input parameters—that is generally used for uncertainty analysis, to draw conclusions about the sensitivity of the model output. This paper presents an extension of CSM plots, called “contribution to sample variance plots” (CSV).

When investigating input–output relationship, CSV contains a considerable higher amount of information than that provided by standard global sensitivity indices [8]. Once the most important input has been detected, global sensitivity indices do not inform the analyst about how to act operatively in order to reduce the range of uncertainty of the important input for a given target reduction of the output variance. Contrarily, CSV does give us the amount of the variance reduction that would be achieved for any arbitrarily chosen restriction of the input uncertainty range.

The CSV measure is defined and its properties are described in Section 2. The interpretation of CSV is given in terms of change of variance that can be achieved by trimming the range of the input parameters in Section 3. The CSV plot can use the same sample points utilized for the CSM, for the scatter-plots, and for uncertainty analysis in general. Therefore, new information can be obtained without additional model evaluations. As the CSM is a powerful tool to identify local regions of the input space that contribute substantially to the mean value of the model output, the CSV is powerful to localize areas where the contribution to the variance of the model output is considerable.

The theoretical results are shown on the Ishigami test function in Section 4. The CSV technique is also applied to a numerical model simulating the behaviour of a water hammer in a pipeline.

\* Corresponding author. Tel.: +39 0332 786257.

E-mail address: [vytis.kopustinskas@jrc.ec.europa.eu](mailto:vytis.kopustinskas@jrc.ec.europa.eu) (V. Kopustinskas).

**Table 1**  
Notation.

$n$	Number of parameters
$N$	Sample size
$X_i$	Model parameter $i$
$Y$	Model output value
$\mathbf{V}(\cdot)$	Variance
$\mathbf{E}(\cdot)$	Mean value
$p(\cdot)$	Probability density function
$F(\cdot)$	Cumulative distribution function
$G(\cdot)$	Model function
$CSM_{X_i}(\cdot)$	Contribution to sample mean for parameter $X_i$
$CSV_{X_i}(\cdot)$	Contribution to sample variance for parameter $X_i$
$q$	quantile

The model is described in Section 5 and the results of CSV and the comparison with the results from CSM are shown in Section 6.

Table 1 presents notation used throughout the article.

**2. The contribution to the sample variance plot**

Let us recall the definition of CSM for a given input  $X_i$  [7]:

$$CSM_{X_i}(q) = \frac{1}{\mathbf{E}(Y)} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \int_{-\infty}^{F_i^{-1}(q)} \prod_{i=1}^n p_i(x_i) G(x_1, x_2, \dots, x_n) \times dx_i dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n$$

$$\mathbf{E}(Y) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \prod_{i=1}^n p_i(x_i) G(x_1, \dots, x_n) dx_1 \dots dx_n \quad (1)$$

where  $q \in [0,1]$ ,  $\mathbf{E}(Y)$  is the mean value of the model output,  $F_i^{-1}(q)$  is the inverse cumulative distribution of  $X_i$  at quantile  $q$ . The multiple integral in (1) is computed on the range  $[-\infty, \infty]$  for all input parameters except for  $X_i$ , for which the range is  $[-\infty, F_i^{-1}(q)]$ .

$CSM_{X_i}(q)$  is plotted on the  $[0,1]^2$  space:  $q$  is a point on  $x$ -axis representing a fraction of distribution range  $X_i$ , and  $CSM_{X_i}(q)$  is a fraction of the output mean corresponding to the values of  $X_i$  smaller or equal than its  $q$ -quantile. By definition,  $CSM_{X_i}(0)=0$  and  $CSM_{X_i}(1)=1$ .

The CSV for the variable  $X_i$  ( $CSV_{X_i}$ ) is defined similarly to CSM:

$$CSV_{X_i}(q) = \frac{1}{\mathbf{V}(Y)} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \int_{-\infty}^{F_i^{-1}(q)} \prod_{i=1}^n p_i(x_i) (G(x_1, x_2, \dots, x_n) - \mathbf{E}(Y))^2 dx_i dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n$$

$$\mathbf{V}(Y) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \prod_{i=1}^n p_i(x_i) (G(x_1, \dots, x_n) - \mathbf{E}(Y))^2 dx_1 \dots dx_n \quad (2)$$

where  $q \in [0,1]$ ,  $\mathbf{V}(Y)$ —variance of the model output,  $F_i^{-1}(q)$  is the inverse cumulative distribution of  $X_i$  at quantile  $q$ . The multiple integral in (2) is computed on the range  $[-\infty, \infty]$  for all input parameters except for  $X_i$ , for which the range is  $[-\infty, F_i^{-1}(q)]$ .

$CSV_{X_i}$  is also plotted on the  $[0,1]^2$  space, where  $CSV_{X_i}(q)$  is a fraction of the output variance corresponding to the values of  $X_i$  smaller or equal than its  $q$  quantile. Also for the CSV holds  $CSV_{X_i}(0)=0$  and  $CSV_{X_i}(1)=1$ . It is important to note that the CSV is defined in (2) as a contribution to variance with respect to constant mean  $\mathbf{E}(Y)$  over the full range of all parameters.

The CSM allows the analyst to identify the effect of each model input on the average of the model output for given percentiles within its uncertainty range; the CSV does the same for the variance.

Graphical tools provide useful insights into the I/O relationship using reasonably low number of model simulations. CSV plots can be placed side by side to scatter-plots and to CSM plots at no extra

computational cost, i.e. using the simulations already carried out for the scatter-plot (and the CSM).

Given a set of  $N$  sample points  $(x_{j1}, x_{j2}, \dots, x_{jn})$ ,  $j=1, 2, \dots, N$  and the corresponding model output values  $y_j$ , the CSV for input  $X_i$  can be computed using the following procedure:

1. Compute the output sample mean  $Ym$  and transform each observation  $(y_1, y_2, \dots, y_N)$  by subtracting the mean value  $Ym$ :  $y_t = y_j - Ym$ ,  $j=1, 2, \dots, N$ . The transformed output  $Yt$  has zero mean value.
2. Sort in ascending order the sample of input  $X_i$  ( $x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(N)}$ ) and the corresponding set of the transformed outputs  $y_t$ , obtaining a series  $(y_{t(i,1)}, y_{t(i,2)}, \dots, y_{t(i,N)})$ .
3. The CSV at  $q$  quantile for the input parameter  $X_i$  is computed as

$$CSV_{X_i}(q) = \frac{\sum_{j=1}^{[qN]} y_{t(i,j)}^2}{\sum_{j=1}^N y_{t(i,j)}^2}, \quad q \in [0,1] \quad (3)$$

where  $qN$  is the largest integer not greater than  $qN$ .

The CSV for input  $X_i$  is obtained by plotting the  $CSV_{X_i}(q)$  against the cumulative distribution function of  $X_i$ . The plot consists of pairs  $(F_i(x_i), CSV_{X_i}(F_i(x_i)))$  for each sample point  $x_i$  of  $X_i$ .

**3. Interpretation of the CSV plot**

Both CSM and CSV are tools to estimate the contribution to the sample mean or sample variance of a particular range of input parameter values. If CSM or CSV are close to diagonal, it indicates that the contribution to the mean or to the variance is equal throughout the range of the input parameter. The CSV value at point  $q=0.1$  provides an estimate of the model output variance due to the 10% of the smallest values of the input parameter. Where the CSV curve is steep, the contribution to the sample variance is large (variance is larger than on average). Where CSV curve is flat, contribution to the sample variance is small (variance is smaller than on average). Similar considerations are valid also for the CSM plot when referring to the contributions to the sample mean.

The CSV plot is a useful tool to analyse the effect of reduced range of an input parameter to the variance of the model output. Further, we will develop a relationship between input parameter range reduction and change of the model output variance. Let us consider a standard setting with input parameter  $X_i$  having probability density function  $p_i(x_i)$  over  $[-\infty, \infty]$ . In the new setting, the range of  $X_i$  is reduced to  $[u, z]$ ,  $-\infty < u < z < \infty$ , having probability density function  $p_i^*(x_i)$ . It can be shown that

$$p_i^*(x_i) = \frac{p_i(x_i)}{\int_u^z p_i(v) dv} \quad (4)$$

The variance  $\mathbf{V}(Y^{*[u,z]})$  in the new setting (range of the input parameter  $X_i$  is  $[u, z]$ ) is computed as

$$\mathbf{V}(Y^{*[u,z]}) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \int_u^z p_1(x_1) p_2(x_2) \dots p_i^*(x_i) \dots p_n(x_n) (G(x_1, \dots, x_n) - \mathbf{E}(Y))^2 dx_i dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n$$

$$= \frac{1}{\int_u^z p_i(v) dv} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \int_u^z p_1(x_1) p_2(x_2) \dots p_i(x_i) \dots p_n(x_n) (G(x_1, \dots, x_n) - \mathbf{E}(Y))^2 dx_i dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n \quad (5)$$

The variance (5) defines variance of the model output with the reduced range of the parameter  $X_i$ , but with respect to constant mean  $\mathbf{E}(Y)$  over the full range of all parameters. Further

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات