Sensitivity Analysis of Monte Carlo Estimates from Computer Models in the Presence of Epistemic and Aleatory Uncertainties

Bernard Krzykacz-Hausmann

Gesellschaft für Anlagen- und Reaktorsicherheit (GRS)
85748 Garching
Germany
E-mail: krb@grs.de
Fax: ++49-89-32004 301

Abstract: Results from complex computer models are often subject to both *aleatory* and *epistemic* uncertainty. The natural straightforward procedure to analyze these uncertainties by Monte Carlo simulation is a double-loop nested sampling: the epistemic parameters are sampled in the outer loop and the aleatory variables are sampled in the nested inner loop. For time-demanding codes, however, the computational effort of this procedure may be prohibitive. Therefore a method of an approximate sensitivity analysis ("sensitivity" in the sense of "uncertainty importance") has been suggested which is based on a single-loop sampling procedure with epistemic parameters and aleatory variables being sampled "simultaneously" from their respective distributions. From the results of such sampling one can obtain approximate estimates of many of the commonly used sensitivity measures for the aleatory probability distributions of model outcomes of interest with respect to the underlying epistemic parameters. The reliability of these estimates depends on the relative contribution of epistemic uncertainties **U** to the overall joint epistemic & aleatory uncertainty in the outcome Y expressed by the quantity $c^2 = \text{var}E[Y|\mathbf{U}]/\text{var}Y$. This quantity can be estimated in several ways depending on the feasibility of additional sampling and model computations.

Keywords: sensitivity analysis, aleatory and epistemic uncertainty, uncertainty importance, conditional expectation.

1. INTRODUCTION

The effect of model input variables subject to aleatory uncertainty ("random behavior") on the results of a complex model can be analyzed by Monte Carlo simulation. To this end the aleatory variables are sampled according to their random laws and the results of the corresponding model runs are summarized in form of empirical distributions which represent the aleatory uncertainty of the model outcomes. From these empirical distributions statistical estimates of the probabilities of the process states of interest and other useful probabilistic quantities like expectations etc. may be obtained.

Often, however, the exact types of the random laws, their distributional parameters, the model formulations, the values of model parameters, the input data of the model application etc are not known precisely, i.e. they are subject to epistemic ("lack-of-knowledge") uncertainty. These uncertainties, denoted as epistemic input uncertainties, are quantified by probability distributions representing the respective subjective state of knowledge.

The aim of epistemic sensitivity analysis ("uncertainty importance analysis") in this case is to quantify the effect of the epistemic input uncertainties on the epistemic uncertainty of the probabilistic quantities representing aleatory output uncertainty, e.g. probabilities, expectations etc.

It is widely recognized and accepted that these two types of uncertainty must very carefully be distinguished and therefore it wouldn't make sense to perform a "simultaneous" Monte Carlo simulation of both types of variables and a sensitivity analysis of a direct model outcome with respect to the variables of both types.

It is intuitively clear and has often been pointed out by many authors, e.g. [1], that the natural method to appropriately account for both types of uncertainty by Monte Carlo simulation is a "double-loop" nested sampling procedure (also called "two-stage" or "two-dimensional" sampling,). It consists of (1) an "outer loop" where the values of the epistemic parameters are sampled according to their epistemic marginal probability distributions and (2) a nested "inner loop" where the values of the aleatory variables are sampled according to their aleatory conditional probability distributions given the values of the epistemic variables chosen in the outer loop. Each "inner loop" provides an empirical conditional aleatory distribution of the process outcome of interest such that finally a sample of empirical distributions is obtained. This sample could be used for a standard epistemic sensitivity analysis for various (aleatory) probabilistic quantities.

However, for complex and computationally expensive models, as used e.g. in probabilistic safety analysis of nuclear power plants, the computational effort for the double-loop procedure will be prohibitive. In such cases the consequence would be to do without an uncertainty and sensitivity analysis.

Therefore, an approach of an approximate epistemic sensitivity analysis is suggested in the following sections. Instead of the nested double-loop sampling procedure the above-mentioned simple single-loop sampling procedure is employed with both types of variables being sampled "simultaneously" according to their joint probability distribution. From the results of this sampling appropriate sensitivity measures can be computed.

2. FUNDAMENTALS

Being subject to both epistemic and aleatory uncertainties, any scalar process variable or model outcome Y may be represented as

$$Y = h(\mathbf{U},\mathbf{V})$$

with

- **U** = set of all epistemic uncertainties (**u**ncertain parameters),
- **V** = set of all aleatory uncertainties (random **v**ariables),
- h = the computational model considered as a deterministic function of both aleatory and epistemic uncertainties **U** and **V**.

When holding the epistemic variables **U** fixed at a value **u** , i.e. **U**=**u**, the resulting outcome Y is a function of the aleatory uncertainties **V**, solely. Its probability distribution, i.e. the conditional distribution $F(y|\mathbf{U}=\mathbf{u})$ of Y given **U**=**u**, quantifies the corresponding (conditional) aleatory uncertainty in Y. Its expectation

$$E[Y|\mathbf{U}=\mathbf{u}]$$

taken over all aleatory variables **V** conditionally on **U**=**u** may be considered as a scalar quantity representing this conditional aleatory uncertainty of the outcome Y.

Using expectation to represent conditional aleatory uncertainty must not be considered very restrictive since many of the standard distributional parameters characterizing aleatory uncertainty can be viewed as expectations of appropriately chosen outcome functions Y'. E.g. the value $F_Y(y)$ of a distribution function of a random variable Y at any given point y may be represented as expectation of the indicator variable $Y' = I_{\{Y \le y\}}$, i.e. Y'=1 if Y≤y and Y'=0 otherwise, from which follows that $EY' = F_Y(y)$.

In the following the standard concise notation

$$E[Y|\mathbf{U}]$$

will be used to denote the above conditional expectation $E[Y|\mathbf{U}=\mathbf{u}]$ considered as function of the epistemic uncertainties **U**, i.e. as a quantity subject to epistemic uncertainty from **U** alone.

The principal aim of an approximate epistemic sensitivity analysis of results from models subject to both epistemic and aleatory uncertainties will therefore be to determine appropriate sensitivity indices of the conditional expectation $E[Y|\mathbf{U}]$ with respect to the components $U_1,\ldots,U_n$ of **U** avoiding the time-consuming double-loop Monte Carlo sampling.

The following fact is the basis of the proposed method:

Many of the standard sensitivity measures of $E[Y|\mathbf{U}]$ *with respect to* $U_1,\ldots U_n$ *are uniformly proportional to the corresponding sensitivity measures of* $Y=h(\mathbf{U},\mathbf{V})$ *with respect to* $U_1,\ldots U_n$. *The proportionality constant* c *is, in most cases, given by*

$$c = \sqrt{\frac{\text{var}\, EY\,|\,\mathbf{U}}{\text{var}Y}} .$$

I.e. if SM_i denotes the (population) sensitivity measure of $E[Y|\mathbf{U}]$ with respect to epistemic parameter U_i, and SM'_i denotes the corresponding sensitivity measure of $Y= h(\mathbf{U},\mathbf{V})$ with respect to the same parameter, then

$$SM'_i = c \cdot SM_i$$

for all i=1,…,n. This holds for many types of sensitivity measures with the same constant c.

Consequently, this property implies that the sensitivity indices for
(a) the conditional expectation $E[Y|\mathbf{U}]$ and for
(b) the direct outcome $Y=h(\mathbf{U},\mathbf{V})$
provide *the same uncertainty importance ranking* with respect to parameters $U_1,\ldots U_n$.

This result holds for the sensitivity measures
- Correlation Coefficient (CC)
- Standardized Regression Coefficient (SRC)
- Correlation Ratio CR (="main effect" sensitivity index)

and with slight modifications also for
- Partial Correlation Coefficient (PCC)
- "total effect" sensitivity index ST
- "linearized"(or R^2-) Version of the "total effect" sensitivity index STL.

The proof of this fact becomes very simple if the concept of conditional expectation $E[Y|\mathbf{U}]$ is employed. It is worthwhile mentioning that the notion of conditional expectation is very useful also in the context of sensitivity analysis. Many results from the standard sensitivity analysis which look rather complex and difficult can very effectively be represented, very clearly interpreted and very easily proved with the aid of the concept of conditional expectation.

The following basic properties of conditional expectation are useful in this context. They can be found in many textbooks and can also very easily be proved:

(1) $E(E[Y|\mathbf{U}]) = EY$
(2) $var(E[Y|\mathbf{U}]) = varY - E(var[Y|\mathbf{U}])$
(3) $E(E^2[Y|\mathbf{U}]) = E(Y \cdot E[Y|\mathbf{U}])$
(4) $E(E[Y|\mathbf{U}]|U_i) = E[Y|U_i]$
(5) $E[E[Y|\mathbf{U}] \cdot U_i] = E[E[Y|U_i] \cdot U_i] = E[Y \cdot U_i]$
(6) $cov(E[Y|\mathbf{U}],U_i)=cov(Y,U_i)$
(7) the linear regression of $E[Y|\mathbf{U}]$ with respect to $\mathbf{U}$ and the linear regression of Y with respect to $\mathbf{U}$ are identical, i.e. $RC(E[Y|\mathbf{U}],U_i) = RC(Y,U_i)$ with RC(…) being the corresponding regression coefficients.

Using these properties the above result can easily be proved. Here, e.g., the proofs for the correlation coefficient CC and the correlation ratio CR ("main effect" sensitivity index):

$$CC(E[Y|\mathbf{U}],U_i) = \frac{cov(E[Y|\mathbf{U}],U_i)}{\sqrt{var\,E[Y|\mathbf{U}] \cdot varU_i}} = \frac{cov(Y,U_i)}{\sqrt{var\,E[Y|\mathbf{U}] \cdot varU_i}} =$$

$$= \frac{cov(Y,U_i)}{\sqrt{varY \cdot varU_i}} \cdot \sqrt{\frac{var\,Y}{var\,E[Y|\mathbf{U}]}} = CC(Y, U_i) \cdot \sqrt{\frac{var\,Y}{var\,E[Y|\mathbf{U}]}} =$$

$$= CC(Y, U_i) \cdot 1/c \,.$$

$$CR^2(E[Y|\mathbf{U}],U_i) = \frac{varE[E[Y|\mathbf{U}]|U_i]}{varE[Y|\mathbf{U}]} = \frac{varE[Y|U_i]}{varE[Y|\mathbf{U}]} = \frac{varE[Y|U_i]}{varY} \cdot \frac{varY}{varE[Y|\mathbf{U}]} =$$

$$= CR^2(Y,U_i) \cdot 1/c^2 \,.$$

The proofs for the other sensitivity measures are similar.

3. SAMPLING METHOD FOR AN APPROXIMATE SENSITIVITY ANALYSIS

Owing to the preceding result it seems natural and reasonable to replace the above-mentioned but often impracticable double-loop sample-based sensitivity analysis for the conditional expectation $E[Y|\mathbf{U}]$ by the corresponding sensitivity analysis for the direct outcome $Y=h(\mathbf{U},\mathbf{V})$ with respect to the components $U_1,\ldots U_n$ of $\mathbf{U}$, alone. The Monte Carlo sampling procedure appropriate for such sensitivity analysis for the direct outcome Y, however, is a simple single-loop sampling with the epistemic parameters $\mathbf{U}$ and the aleatory variables $\mathbf{V}$ being sampled "simultaneously" according to their joint probability distribution $f(\mathbf{u},\mathbf{v})$. This joint probability distribution is given by the product of the marginal distribution $f(\mathbf{u})$ of $\mathbf{U}$ and the conditional distribution $f(\mathbf{v}|\mathbf{U}=\mathbf{u})$ of $\mathbf{V}$ given $\mathbf{U}=\mathbf{u}$, i.e. by the expression

$$f(\mathbf{u},\mathbf{v}) = f(\mathbf{v}|\mathbf{U}=\mathbf{u}) \cdot f(\mathbf{u}) \,.$$

In most applications the marginal distribution $f(\mathbf{u})$ of the epistemic parameters $\mathbf{U}$ will be given directly, while the conditional distribution $f(\mathbf{v}|\mathbf{U}=\mathbf{u})$ of the aleatory variables $\mathbf{V}$ may also be given in terms of intermediate results from the computational model.

Thus, the "simultaneous" sampling procedure with sample size N generates N joint epistemic & aleatory sample values

$$(\mathbf{u}_1,\mathbf{v}_1),\ldots\ldots,(\mathbf{u}_N,\mathbf{v}_N)$$

from which, eventually, the corresponding sample values

$$y_1,\ldots\ldots,y_N$$

of the direct outcome Y=h(**U**,**V**) are calculated via the computer code.

From all these sample values the above mentioned standard sensitivity measures with respect to the parameters $U_1,\ldots U_n$ for the outcome Y=h(**U,V**) can be computed. Since the proportionality constant $c=\sqrt{(\mathrm{var}E[Y|\mathbf{U}]/\mathrm{var}Y)}$ is usually not known one cannot directly derive the sensitivity indices for E[Y|**U**] from the sensitivity indices for Y. However, according to the preceding section, the sample based parameter importance ranking obtained for Y may approximately be used as the importance ranking for the conditional expectation E[Y|**U**] asked for. Methods for approximating/estimating the proportionality constant c will be presented in section 5.

It is also clear that Simple Random Sampling (SRS) as well Latin Hypercube Sampling (LHS) or any other sampling method appropriate for the selected type of sensitivity measure may be used for such sample-based approximate sensitivity analysis.

4. ACCURACY CONSIDERATIONS

The accuracy of the approximate sensitivity analysis for the outcome Y depends on the (usually) unknown value of the (squared) proportionality constant

$$c^2 = \frac{\mathrm{var}\, E[Y\,|\,\mathbf{U}]}{\mathrm{var}\, Y}$$

which relates the sensitivity measures for E[Y|U] to the sensitivity measures for Y.

Clearly, $0 \le c^2 \le 1$ since varE[Y|**U**] ≤ varY due to the above property (2) of the conditional expectation. From the proportionality $SM_i=1/c\cdot SM_i'$ (i=1,…,n) it follows that the values of the sensitivity measures SM_i' for Y are uniformly lower than the corresponding sensitivity measures SM_i for E[Y|**U**]. If this constant c^2 is small, the sample-based approximate sensitivity analysis for Y may produce small or even statistically not significant values of the sensitivity measure for a parameter although the sensitivity of E[Y|**U**] with respect to this parameter one is actually interested in may be high. Nevertheless, c^2 is unknown and therefore it is important to analyze it more closely.

By definition, c^2 is easily identified as squared multiple correlation ratio (or "main effect" sensitivity index) [2],[3] of Y with respect to the whole parameter vector **U**. It can therefore be interpreted in several ways, e.g.

- as an indicator of the accuracy of the approximation of Y=h(**U**,**V**) by E[Y|**U**] as a function of **U** alone,
- as an indicator of the relative contribution of the epistemic uncertainties from **U** to the overall "joint" uncertainty in Y=h(**U**,**V**) from **U** and **V**,
- as the extent to which the overall "joint" uncertainty in Y coming from **U** and **V** is dominated by the epistemic uncertainty coming from **U** alone.

Consequently, the more "dominant" the epistemic uncertainties the higher the c^2 value, and, consequently, the higher the dependability of the proposed approximate sensitivity analysis.

In practical applications it may sometimes be immediately clear which type of uncertainty is dominant such that the reliability of the approximate sensitivity results may also be judged immediately. Nevertheless, an approximation of c^2 is needed on the basis of the reduced sampling effort without employing the impracticable double-loop approach.

5. APPROXIMATING THE PROPORTIONALITY CONSTANT c^2

Three alternative procedures are proposed to approximate res. to estimate the (squared) proportionality constant $c^2 = \frac{\text{var } E[Y \mid \mathbf{U}]}{\text{var } Y}$. Below the three procedures are ordered according to the amount of the additional computational effort necessary to determine the corresponding approximated res. estimated value of c^2.

(1) Procedure No.1 to approximate c^2 is based solely on the underlying sample values from the "joint" sampling of **U** and **V**, i.e. without additional model computations. It is given by

$$\hat{c}^2 = \frac{R^2(Y, \mathbf{U})}{R^2(Y, (\mathbf{U}, \mathbf{V}))} ,$$

with

$R^2(Y,(\mathbf{U},\mathbf{V}))$ = multiple sample correlation coefficient of outcome Y with respect to the joint sample of $(\mathbf{U},\mathbf{V})$.

$R^2(Y,\mathbf{U}))$ = multiple sample correlation coefficient of outcome Y with respect to the sample of **U** alone.

Both multiple correlation coefficients can easily be computed from the available sample values $(\mathbf{u}_1,\mathbf{v}_1),\ldots,(\mathbf{u}_N,\mathbf{v}_N)$ and $y_1,\ldots,y_N$ according to the well-known formulae:

$$R^2(Y,\mathbf{U}) = \rho^t_{Y,\mathbf{U}} \mathbf{R_U}^{-1} \rho_{Y,\mathbf{U}}$$

$$R^2(Y,(\mathbf{U},\mathbf{V})) = \rho^t_{Y,(\mathbf{U},\mathbf{V})} \mathbf{R}_{\mathbf{U},\mathbf{V}}^{-1} \rho_{Y,(\mathbf{U},\mathbf{V})}$$

with

$\rho_{Y,\mathbf{U}}$ = vector of empirical correlation coefficients between Y and the components of **U**

$\rho_{Y,(\mathbf{U},\mathbf{V})}$ = vector of empirical correlation coefficients between Y and the components of **U**,**V**

$\mathbf{R_U}^{-1}$ = inverse of the empirical correlation matrix $\mathbf{R_U}$ between the components of **U**

$\mathbf{R}_{\mathbf{U},\mathbf{V}}^{-1}$ = inverse of the empirical correlation matrix $\mathbf{R}_{\mathbf{U},\mathbf{V}}$ between the components of **U**,**V**.

All these quantities are computed from the underlying sample values $(\mathbf{u}_1,\mathbf{v}_1),\ldots,(\mathbf{u}_N,\mathbf{v}_N)$ and $y_1,\ldots,y_N$ generated by the single-loop joint sampling of **U,V** and the corresponding model computations of Y. The sample size N must exceed the joint number of variables in **U,V**.

The motivation behind this method is simply to approximate the conditional expectation (= regression of the 1st kind) by the linear regression (= regression of the 2nd kind).

(2) Procedure No.2 of approximating c^2 is based on two samples: (a) the underlying sample values $y_1,\ldots,y_N$ from the same "joint" sampling of **U** and **V** and (b) sample values from an additional (single-loop) sampling of aleatory variables **V** alone with the values of epistemic parameters **U** held fixed at their nominal values $\mathbf{u}_0$. It is defined by

$$\hat{c}^2 = \frac{s^2(Y) - s^2(Y \mid \mathbf{U} = \mathbf{u}_0)}{s^2(Y)} \;.$$

where

$$s^2(Y) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{y})^2$$

is the variance from the underlying sample values $y_1,\ldots,y_N$, and

$$s^2(Y|\mathbf{U}=\mathbf{u}_0) = \frac{1}{N'}\sum_{i=1}^{N'}(y'_i - \overline{y}')^2 \,.$$

is the variance from the other sample values $y'_1,\ldots,y'_{N'}$ generated by sampling the aleatory variables **V** alone while the epistemic parameters **U** are held fixed at their nominal values $\mathbf{u}_0$. In many applications this additional sample may already be available as the "nominal result" computed before starting uncertainty and sensitivity analysis.

The motivation behind this method is to approximate the term Evar[Y|**U**] appearing in the expression varE[Y|**U**] = varY - Evar[Y|**U**] for the numerator of c^2 by the term var[Y|**U**=$\mathbf{u}_0$].

(3) Procedure No. 3: While the first two methods should rather be considered as numerical approximations to the constant c^2=varE[Y|**U**]/varY, the third method may be viewed as an estimate of c^2 in the full statistical sense. It is based on the following basic and easy to prove property of conditional expectation:

If **V** *and* **V'** *are identically distributed and conditionally independent given* **U**, *i.e. the joint conditional distribution of* **V** *and* **V'** *given* **U** *is the product if the two marginal conditional distributions, formally:* f(**v**,**v**'|**U**=**u**) = f(**v**|**U**=**u**)·f(**v**'|**U**=**u**) *and if* Y = h(**U**,**V**) *and* Y' = h(**U**,**V**')*), then the (squared) proportionality constant* c^2 *can be expressed by:*

$$c^2 = \frac{\mathrm{var}\,E[Y \mid \mathbf{U}]}{\mathrm{var}\,Y} = \frac{\mathrm{cov}(Y, Y')}{\sqrt{\mathrm{var}Y\ \mathrm{var}Y'}} = \rho(Y,Y')\ ,$$

i.e. c^2 *is the correlation coefficient between the variables* Y *und* Y'.

Consequently, one can estimate the proportionality constant c^2 by the sample correlation coefficient r(**y**,**y**') from the two-dimensional sample $(y_1,y_1'),\ldots,(y_N,y_N')$ from the bivariate distribution of (Y,Y'). The corresponding well-known formula is

$$\hat{c}^2 = r(\mathbf{y},\mathbf{y}') = \frac{\sum_{i=1}^{N}(y_i - \overline{y})\cdot(y_i' - \overline{y}')}{\sqrt{\sum_{i=1}^{N}(y_i - \overline{y})^2 \cdot \sum_{i=1}^{N}(y_i' - \overline{y}')^2}}\ ,$$

where

$y_1,\ldots,y_N$ are the sample values of Y from the underlying "joint" sample of **U** and **V**, i.e. y_i=h($\mathbf{u}_i$,$\mathbf{v}_i$),), i=1,…,N and

$y_1',\ldots,y_N'$ are the sample values of Y' from the "joint" sample of **U** and **V'** generated by independently sampling the aleatory variables **V'** alone, according to the conditional distribution with the epistemic parameters **U** held fixed at the same values as in the 1st sample, i.e. y_i'=h($\mathbf{u}_i$,$\mathbf{v}_i'$), i=1,…,N.

The additional computational effort for this statistical estimate of c^2 is therefore N additional model computations (= 2nd single-loop sample of size N).

Obviously, these two single-loop samples may also be viewed as a realization of the above-mentioned nested double-loop sampling with the "inner" loop sample size being 2.

Remark 1:

The above statistical estimate $\hat{c}^2$ may also be considered as an extension of the familiar procedure [4] to estimate the so-called "main effect" and "total effect" sensitivity indices SM and ST in the case of *not independent* variables. Changing the notation and replacing **U** by $\mathbf{X}_2$ and **V** by $\mathbf{X}_1$ the "total effect" sensitivity index ST_1 for Y with respect to $\mathbf{X}_1$ may be defined as

$$ST_1 = \frac{E\,\mathrm{var}[Y\,|\,\mathbf{X}_2]}{\mathrm{var}Y} = \frac{\mathrm{var}Y - \mathrm{var}\,E[Y\,|\,\mathbf{X}_2]}{\mathrm{var}Y}.$$

It can be interpreted as "the relative amount of variance of Y that is expected to remain if the values of all variables except variables $\mathbf{X}_1$ will become known". Analogously, the "main effect" sensitivity index SM_1 for Y with respect to $\mathbf{X}_2$ may be defined as

$$SM_2 = \frac{\mathrm{var}\,E\,[Y\,|\,\mathbf{X}_2]}{\mathrm{var}Y}$$

and interpreted as "the relative amount of variance of Y that is expected to be removed if the values of all variables $\mathbf{X}_2$ will become known". This representation holds for independent as well as for dependent variables $\mathbf{X}_1$ and $\mathbf{X}_2$ and is equivalent to the representation given in [4] in the case of *independent* variables (e.g. ST_1 := sum of all terms containing $\mathbf{X}_1$ of the "Sobol decomposition" of $Y=h(\mathbf{X}_1,\mathbf{X}_2)$ into a sum of uncorrelated terms of increasing dimensionality [2], [4]). It is immediately seen that

$$ST_1 = 1 - c^2$$
$$SM_2 = c^2$$

with $\mathbf{X}_1,\mathbf{X}_2$ playing the role of **V**,**U** in the above representation of c^2. It can also be easily seen that for independent variables the estimate presented in this paper and the estimate presented in [4] are nearly equivalent. Consequently, in the procedure [4] to compute the "main effect" and the "total effect" sensitivity indices it is not necessary to assume the input variables be independent. This procedure can be used for dependent variables, as well, provided the two samples of $\mathbf{X}_1$ are generated conditionally independently given $\mathbf{X}_2$.

Remark 2:

According to the above procedure a 2nd sample is generated to estimate (together with the 1st sample) the constant c^2 while to estimate the sensitivity indices only the 1st sample is needed. It appears, and is intuitively clear, too, that using the mean sample values $y_i^* = (y_i+y_i')/2$, i=1,…,N from both samples an improvement of the accuracy of the sensitivity results can be achieved compared to the results obtained with the values y_i , i=1,…,N, from a single sample. As before, since $E[Y^*|\mathbf{U}]=E[Y|\mathbf{U}]$ and $\mathrm{var}Y^*=(\mathrm{var}Y+\mathrm{var}E[Y|\mathbf{U}])/2$, it can easily be shown that a similar proportional relationship holds between the sensitivity measures SM_i of $E[Y|\mathbf{U}]$ and the corresponding sensitivity measures SM^*_i of $Y^*=(Y+Y')/2$ with respect to parameter U_i , i.e.

$$SM^*_i = c^* \cdot SM_i$$

with the new proportionality constant c^{*2} given by

$$c^{*2} = \frac{\text{var } E[Y^* | \mathbf{U}]}{\text{var } Y^*} = \frac{2 \text{ var } E[Y | \mathbf{U}]}{\text{var } Y + \text{var } E[Y | \mathbf{U}]} = \frac{2c^2}{1+c^2} > c^2$$

As stated above, since $c^{*2} > c^2$, the uncertainty importance ranking (sensitivity results) based on the y_i^* values will provide a more reliable approximation to the importance ranking for $E[Y|\mathbf{U}]$ than the importance ranking based on the y_i values from a single sample.

A straightforward generalization to K conditionally independent samples, i.e.

$$Y^* = 1/K \sum Y^{(k)}$$

provides an improvement with the proportionality constant

$$c^* = \frac{K \text{ var } E[Y | \mathbf{U}]}{\text{var } Y + (K-1) \text{ var } E[Y | \mathbf{U}]} = \frac{K c^2}{1+(K-1)c^2}.$$

This, obviously, is equivalent to the above-mentioned nested double-loop sampling with the “inner” loop sample size being K.

6. SIMPLE ANALYTICAL EXAMPLE (LINEAR NORMAL CASE)

To illustrate some of the preceding results a simple (artificial) numerical example is presented where all quantities of interest can be determined analytically and compared with the results from the sampling procedures presented above. In this example a simple linear independent normal case is considered, i.e.

$$Y = h(\mathbf{U},\mathbf{V}) = \sum_{i=1}^{n} a_i U_i + \sum_{j=1}^{m} b_j V_j$$

where all epistemic parameters $\mathbf{U}=(U_1,\ldots,U_n)$ and all aleatory variables $\mathbf{V}=(V_1,\ldots,V_m)$ are independent and have the standard Normal distribution N(0,1). The coefficients a_i , b_j are assumed to be known. Then it can easily be shown that

- $\text{var}Y = \sum a_i^2 + \sum b_j^2$,
- $E[Y|\mathbf{U}] = \sum a_i U_i$,
- $\text{var}E[Y|\mathbf{U}] = \sum a_i^2$,
- $c^2 = \dfrac{\text{var } E[Y | \mathbf{U}]}{\text{var } Y} = \dfrac{\sum a_i^2}{\sum a_i^2 + \sum b_j^2}$,
- $SM_i' = SM(Y,U_i) = \dfrac{a_i}{\sqrt{\sum a_i^2 + \sum b_j^2}}$, (i=1,…,n) ,
- $SM_i = SM(E[Y|\mathbf{U}], U_i) = \dfrac{a_i}{\sqrt{\sum a_i^2}}$, (i=1,…,n) ,

where SM denotes any type of sensitivity measure, since, due to linearity and independence all standard sensitivity measures of Y or of $E[Y|\mathbf{U}]$ with respect to U_i are equal.

Here it can directly be seen: the higher the contribution of the epistemic uncertainties from **U** to the overall joint uncertainty in Y, expressed by the constant c^2, the more precise the proposed approximation of the sensitivity measures for $E[Y|\mathbf{U}]$ by the sensitivity measures for Y.

For numerical calculations it was assumed that n=m=5, **a**=**b**=(1,2,3,4,5). Consequently varY=110, $\text{var}E[Y|\mathbf{U}]=55$, $c^2 = 1/2$.

The following table summarizes the results obtained analytically and with the sampling methods described above. It shows the values of the sensitivity measures (Standardized Regression Coefficient, SRC) for E[Y|**U**] with respect to all five parameters $U_1,\ldots,U_5$ obtained in four different ways:

(1) analytically,
(2) from double-loop simple random sampling with sample size 100x100
(3) from single-loop simple random sampling with sample size 500
(4) from single-loop simple random sampling with sample size 200.

Standardized Regression Coefficients (SRC) for E[Y|**U**]

Index of Parameter	(1) analytic	(2) two-loop ss=100x100	(3) one-loop ss=500	(4)one-loop ss=200
1	0.1348	0.1369	0.140	0.071
2	0.2696	0.2946	0.259	0.265
3	0.4044	0.4262	0.387	0.397
4	0.5392	0.5340	0.584	0.629
5	0.6740	0.7225	0.703	0.658

The three alternative methods for approximating/estimating the proportionality constant c provide the results:

The proportionality constant c

sample size	method 1	method 2	method 3	exact value
500	0.7027	0.7032	0.7003	0.7071 =$\sqrt{0.5}$
200	0.7437	0.6895	0.6920	0.7071 =$\sqrt{0.5}$

Conclusion: The results of this simple example look promising and suggest that in real situations with complex and computationally expensive models where the double-loop sampling is prohibitive, the approximate sensitivity analysis presented in this paper may provide reasonable results. It may therefore be preferred to the alternative of not performing any sensitivity analysis.

REFERENCES

[1] Helton J.C. et al., 1998. Uncertainty and sensitivity results obtained in the 1996 performance assessment for the waste isolation pilot plant. SAND98-0365, Albuquerque: Sandia National Laboratories; 1998

[2] Saltelli A., Chan K., Scott M. (eds.). Sensitivity Analysis. John Wiley & Sons, 2000.

[3] McKay M.D., 1995. Evaluating prediction uncertainty. Tech. Rep. NUREG/CR-6311. U.S. Nuclear Regulatory Commission and Los Alamos National Laboratory.

[4] Homma, T., Saltelli A., 1996. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering and System Safety* Vol. 52, 1-17