A new derivative based importance criterion for groups of variables and its link with the global sensitivity indices.

I.M. Sobol ${ }^{\text {ª }}$, S. Kucherenko ${ }^{\text {b1 }}$<br>${ }^{a}$ Institute for Mathematical Modelling of the Russian Academy of Sciences, 4 Miusskaya Square, Moscow 125047, Russia<br>${ }^{b}$ Centre for Process Systems Engineering, Imperial College London, London, SW7 2AZ, UK


#### Abstract

A new derivative based criterion $\tau_{y}$ for groups of input variables is presented. It is shown that there is a link between global sensitivity indices and the new derivative based measure. It is proved that small values of derivative based measures imply small values of total sensitivity indices. However, for highly nonlinear functions the ranking of important variables using derivative based importance measures can be different from that based on the global sensitivity indices. The computational costs of evaluating global sensitivity indices and derivative based measures, are compared and some important tests are considered.


Keywords: Global sensitivity analysis, Global sensitivity index; Monte Carlo method; Derivative based global sensitivity measure

## 1. Introduction

The quality of a model depends on a variety of aspects such as accuracy of experimental data, choice of an appropriate model and reliable identification of the unknown model parameters. With regard to these aspects, sensitivity analysis offers a generalized approach for

[^0]identification of functional dependencies, selection of a model structure from a set of known competing models, effective and reliable identification of important model parameters and input variables and subsequent reduction of model complexity.

Let $x=\left(x_{1}, \ldots, x_{n}\right)$ be a point in the $n$-dimensional unit hypercube $H^{n}$ with Lebesgue measure $d x=d x_{1} \cdots d x_{n}$. Consider a model function $f\left(x_{1}, \ldots, x_{n}\right)$ defined on this hypercube $H^{n}$. It can be a black box model and not necessarily an analytical expression.

For $f(x) \in L_{2}$ global sensitivity indices provide adequate estimates for the impact upon $f(x)$ from individual factors $x_{i}$ or from groups of these factors. Such indices can be efficiently computed by Monte Carlo or quasi-Monte Carlo methods that include values of $f(x)$ at special random or quasi-random points. However, if the number of model evaluations is too high application of global sensitivity indices can become unpractical.

Derivative based importance criteria provide an alternative approach to the same problem: the impact upon $f(x)$ from the factor $x_{i}$ is estimated by a functional that depends on the derivative $\partial f / \partial x_{i}$. A link between both approaches was established in [1]. It was proved that derivative based estimates can be successfully used for identifying non important factors, but ranking important variables according to their derivative based estimates can be different from that based on the global sensitivity indices. Numerical experiments show that in certain situations (e.g. if the variation of $\partial f / \partial x_{i}$ is small) derivative based importance estimates can be computed much faster than the corresponding sensitivity indices [2]. Although global sensitivity indices are superior to derivative based importance criteria in that they provide more detailed information about models accounting for both individual effects and interactions between variables, derivative based importance criteria can be seen as a good practice for factors screening purposes in place of the modified Morris method $\mu^{*}$ measure [3]. The Morris method uses random sampling of points from the fixed grid (levels) for averaging elementary effects which are calculated as finite differences with the increment delta comparable with the range of uncertainty. For this reason it can not correctly account for the effects with characteristic dimensions much less than delta. Calculation of the Morris measures is not supported by the convergence monitoring procedure and therefore Morris measures can be unreliable.

In the present paper a new derivative based criterion $\tau_{y}$ is introduced, that is regarded as a possible estimate of the impact upon $f(x)$ from a group of factors $y=\left(x_{i_{i}}, \ldots, x_{i_{s}}\right)$. It is proved, that if $f(x)$ is linear with respect to $x_{i_{1}}, \ldots, x_{i_{s}}$, then the performance of $\tau_{y}$ is equivalent to the performance of the sensitivity index $S_{y}^{\text {tot }}$ defined in section 2.3 below. In the case when $y$ consists of one factor $x_{i}$, the corresponding criterion $\tau_{i}$ is a slight improvement of the criteria studied in [1].

This paper is organized as follows: The next section contains definitions and main properties of global sensitivity indices. Section 3 introduces the new derivatives based importance criterion $\tau_{y}$. Section 4 considers the one-factor case criterion $\tau_{i}$. Sections 5, 6 and 7 contain some simple but significant examples illustrating theoretical results. In section 8 the case of independent random factors $x_{1}, \ldots, x_{n}$ is briefly discussed. Finally, conclusions are presented in the last section. The Appendix contains a simple test for arbitrary importance criteria that estimate the impact of groups of variables. It compares the results for global sensitivity indices, the criterion $\tau_{y}$ and the generalization of the Morris measure for groups of variables $\mu^{*}$ proposed in [3].

Note that global sensitivity indices were introduced in [4]. Their main theoretical properties were developed in [5-10]. Applications of these techniques are presented in e.g. [11-13]. Morris elementary effect measures were introduced in [14] and further developed in [3]. Derivative based criteria were introduced in [2]. Their theoretical properties and the link with global sensitivity indices were established in [1]. Numerical examples are given in [2] and [15]. It was shown in [15] that this technique becomes especially efficient if automatic calculation of derivatives is used.

## 2. Global sensitivity indices

As mentioned before, let $x=\left(x_{1}, \ldots, x_{n}\right)$ be a point in the $n$-dimensional unit hypercube with Lebesgue measure $d x=d x_{1} \cdots d x_{n}$. Consider an arbitrary subset of the variables $y=\left(x_{i_{1}}, \ldots, x_{i_{s}}\right), 1 \leq s<n$, and the set of remaining complementary variables $z$, so that $x=(y, z), d x=d y d z$. All the integrals in the paper are written without integration limits: we assume that each integration variable varies independently from 0 to 1 .

### 2.1 ANOVA-like decomposition

Consider a model function $f(x) \in L_{2}$. Denote

$$
\begin{aligned}
& f_{0}=\int f(x) d x \\
& h_{1}(y)=\int f(x) d z-f_{0} \\
& h_{2}(z)=\int f(x) d y-f_{0} \\
& h_{12}(x)=f(x)-\int f(x) d z-\int f(x) d y+f_{0}
\end{aligned}
$$

Then we obtain a decomposition of $f(x)$

$$
\begin{equation*}
f(x)=f_{0}+h_{1}(y)+h_{2}(z)+h_{12}(x) \tag{2.1}
\end{equation*}
$$

One can easily verify that

$$
\int h_{1}(y) d y=\int h_{2}(z) d z=\int h_{12}(x) d y=\int h_{12}(x) d z=0 .
$$

These relations imply the orthogonality of the terms in (2.1):

$$
\int h_{1}(y) h_{2}(z) d x=\int h_{1}(y) h_{12}(x) d x=\int h_{2}(z) h_{12}(x) d x=0
$$

### 2.2 Variances

The following integrals are called partial variances: $D_{y}=\int h_{1}^{2}(y) d y, D_{z}=\int h_{2}^{2}(z) d z$, $D_{y z}=\int h_{12}^{2}(x) d x$ and $D$ is the total variance: $D=\int f^{2}(x) d x-f_{0}^{2}$. Squaring (2.1) and integrating over $d x$, we obtain $D=D_{y}+D_{z}+D_{y z}$. Two total partial variances are defined as $D_{y}^{\text {tot }}=D_{y}+D_{y z}, D_{z}^{\text {tot }}=D_{z}+D_{y z}$. Note, that if $x_{1}, \ldots, x_{n}$ were treated as independent random variables uniformly distributed in the unit interval, then all the terms in (2.1) would be random variables and $D, D_{y}, D_{z}, D_{y z}$ their variances.

### 2.3 Global sensitivity indices

The global sensitivity indices are defined as ratios with common denominator $D$ :

$$
S_{y}=\frac{D_{y}}{D}, S_{z}=\frac{D_{z}}{D}, S_{y z}=\frac{D_{y z}}{D}, S_{y}^{\text {tot }}=\frac{D_{y}^{\text {tot }}}{D}, S_{z}^{\text {tot }}=\frac{D_{z}^{\text {tot }}}{D} .
$$

Obviously: $S_{y}+S_{z}+S_{y z}=1, S_{y}^{\text {tot }}=1-S_{z}, S_{z}^{\text {tot }}=1-S_{y}$. These indices have the following main properties:
$1^{\circ} 0 \leq S_{y} \leq S_{y}^{\text {tot }} \leq 1$.
$2^{\circ} S_{y}=S_{y}^{\text {tot }}=0$ means that $f(x)$ does not depend on $y$.
$3^{\circ} S_{y}=S_{y}^{\text {tot }}=1$ means that $f(x)$ depends only on $y$.
$4^{0}$ If the set of variables $z$ is somehow fixed, $z=z_{0}$, and $f(x)$ is approximated by $f\left(y, z_{0}\right)$, then the approximation error depends strongly on $S_{z}^{\text {tot }}$.

Here is an exact formulation of property $4^{\circ}$ from [8] and [9]. Denote by $\delta\left(z_{0}\right)$ the approximation error in a scaled $L_{2}$ metric:

$$
\delta\left(z_{0}\right)=\frac{1}{D} \int\left[f(x)-f\left(y, z_{0}\right)\right]^{2} d x .
$$

Then $\delta\left(z_{0}\right) \geq S_{z}^{\text {tot }}$. But if $z_{0}$ is random and uniformly distributed, then the expectation is

$$
\boldsymbol{E} \delta\left(z_{0}\right)=2 S_{z}^{\text {tot }} .
$$

### 2.4 Integral formulas

The following integral formulas show that global sensitivity indices can be computed without knowing the terms in (2.1):

$$
\begin{align*}
& D_{y}^{\text {tot }}=\frac{1}{2} \int\left[f\left(y^{\prime}, z\right)-f(x)\right]^{2} d x d y^{\prime},  \tag{2.2}\\
& D_{y}=\int f\left(x^{\prime}\right)\left[f\left(y^{\prime}, z\right)-f(x)\right] d x d x^{\prime} . \tag{2.3}
\end{align*}
$$

The integrals on the right hand side can be computed by crude Monte Carlo or quasi-Monte Carlo methods, the point $x^{\prime}$ is an independent realization of the point $x$, so that $x^{\prime}=\left(y^{\prime}, z^{\prime}\right)$ and $d x^{\prime}=d y^{\prime} d z^{\prime}$. The total variance $D$ can be computed by traditional statistical methods. If simultaneously $m$ different subset $y$ are considered, then according to [6] for each Monte Carlo trial $m+2$ model estimates are necessary.

Formula (2.2) is given in [5], while formula (2.3) was derived in [9]. The variance of the (2.2) estimator was investigated in [5]; the variance of the (2.3) estimator was studied in [16].

## 3. Importance criterion $\tau_{y}$

Assume that $f(x)$ is differentiable. Consider the Taylor expansion

$$
f\left(y^{\prime}, z\right)-f(y, z)=\sum_{p=1}^{s} \frac{\partial f(x)}{\partial x_{i_{p}}}\left(x_{i_{p}}^{\prime}-x_{i_{p}}\right)+\ldots
$$

The omitted terms depend on second order derivatives. If these derivatives (or more accurately, if certain integrals that include these derivatives) are small, then the last sum will be a fair approximation for the expression on the left hand side. Substituting this sum into formula (2.2), we construct the following derivative based importance criterion:

$$
\hat{\tau}_{y}=\frac{1}{2} \int\left[\sum_{p=1}^{s} \frac{\partial f(x)}{\partial x_{i_{p}}}\left(x_{i_{p}}^{\prime}-x_{i_{p}}\right)\right]^{2} d x d y^{\prime} .
$$

In certain circumstances $\hat{\tau}_{y}$ will be near to $D_{y}^{\text {tot }}$. In the formula for $\hat{\tau}_{y}$, the integration over $d y^{\prime}$ can be carried out. Indeed,

$$
\hat{\tau}_{y}=\frac{1}{2} \int \sum_{p=1}^{s}\left(\frac{\partial f(x)}{\partial x_{i_{p}}}\right)^{2}\left(x_{i_{p}}^{\prime}-x_{i_{p}}\right)^{2} d x d y^{\prime}+\int \sum_{p<q} \frac{\partial f(x)}{\partial x_{i_{p}}} \frac{\partial f(x)}{\partial x_{i_{q}}}\left(x_{i_{p}}^{\prime}-x_{i_{p}}\right)\left(x_{i_{q}}^{\prime}-x_{i_{q}}\right) d x d y^{\prime} .
$$

The final expression for $\hat{\tau}_{y}$ is

$$
\hat{\tau}_{y}=\sum_{p=1}^{s} \int\left(\frac{\partial f(x)}{\partial x_{i_{p}}}\right)^{2} \frac{1-3 x_{i_{p}}+3 x_{i_{p}}^{2}}{6} d x+\sum_{p<q} \int \frac{\partial f(x)}{\partial x_{i_{p}}} \frac{\partial f(x)}{\partial x_{i_{q}}}\left(x_{i_{p}}-\frac{1}{2}\right)\left(x_{i_{q}}-\frac{1}{2}\right) d x .
$$

The sum over $p<q$ means that $1 \leq p<q \leq s$.
$\tau_{y}$ is a further simplification of $\hat{\tau}_{y}$ : only one main term is retained, namely

$$
\begin{equation*}
\tau_{y}=\sum_{p=1}^{s} \int\left(\frac{\partial f(x)}{\partial x_{i_{p}}}\right)^{2} \frac{1-3 x_{i_{p}}+3 x_{i_{p}}^{2}}{6} d x \tag{3.1}
\end{equation*}
$$

This simplification is justified by the following properties of $\tau_{y}$ :

## Theorem 1

If $f(x)$ is linear with respect to $x_{i_{1}}, \ldots, x_{i_{s}}$, then $D_{y}^{\text {tot }}=\tau_{y}$, or in other words $S_{y}^{\text {tot }}=\frac{\tau_{y}}{D}$.

## Theorem 2

A general inequality holds: $D_{y}^{\text {tot }} \leq\left(24 / \pi^{2}\right) \tau_{y}$ or in other words $S_{y}^{\text {tot }} \leq \frac{24}{\pi^{2}} \frac{\tau_{y}}{D}$.

The first theorem shows that if the model $f(x)$ is near to linear, the performance of $\tau_{y}$ will be near to the performance of global sensitivity indices.

The second theorem shows that small values of $\tau_{y}$ imply small values of $S_{y}^{\text {tot }}$ and this allows identification of a set of unessential factors $y$ (usually defined by a condition of the type $\left.S_{y}^{\text {tot }}<\delta\right)$.

## Proof of Theorem 1

Assume that $f(x)=\sum_{p=1}^{s} a_{p}(z) x_{i_{p}}+b(z)$. Formula (2.2) then implies

$$
D_{y}^{\text {tot }}=\frac{1}{2} \int\left\{\sum_{p=1}^{s} a_{p}(z)\left(x_{i_{p}}^{\prime}-x_{i_{p}}\right)\right\}^{2} d x d y^{\prime}=\frac{1}{2} \sum_{p=1}^{s} \int a_{p}^{2}(z)\left(x_{i_{p}}^{\prime}-x_{i_{p}}\right)^{2} d z d y^{\prime}=\frac{1}{12} \int \sum_{p=1}^{s} a_{p}^{2}(z) d z
$$

On the other hand, from (3.1)

$$
\tau_{y}=\sum_{p=1}^{s} \int a_{p}^{2}(z) d z \int \frac{1-3 x_{i_{p}}+3 x_{i_{p}}^{2}}{6} d x_{i_{p}}=\frac{1}{12} \sum_{p=1}^{s} \int a_{p}^{2}(z) d z . \square
$$

The proof of Theorem 2 will be given at the end of the next section, because this proof needs some results derived in the latter section.

## 4. Importance criterion $\tau_{i}$

Consider now the one dimensional case when the subset $y$ consists of only one variable $y=\left(x_{i}\right)$. From (3.1) we then obtain a criterion

$$
\begin{equation*}
\tau_{i}=\int\left(\frac{\partial f(x)}{\partial x_{i}}\right)^{2} \frac{1-3 x_{i}+3 x_{i}^{2}}{6} d x \tag{4.1}
\end{equation*}
$$

that is very close to the criterion $v_{i}$, discussed in [1]:

$$
\begin{equation*}
v_{i}=\int\left(\frac{\partial f(x)}{\partial x_{i}}\right)^{2} d x \tag{4.2}
\end{equation*}
$$

In fact, $1-3 t+3 t^{2}$ for $0 \leq t \leq 1$ is bounded: $1 / 4 \leq\left(1-3 t+3 t^{2}\right) \leq 1$. Therefore $v_{i} / 24 \leq \tau_{i} \leq v_{i} / 6$. In [1] a general inequality was proved: $S_{i}^{\text {tot }} \leq \frac{1}{\pi^{2}} \frac{v_{i}}{D}$. From this inequality we immediately obtain a general inequality for $\tau_{i}$ :

$$
\begin{equation*}
S_{i}^{\text {tot }} \leq \frac{24}{\pi^{2}} \frac{\tau_{i}}{D} . \tag{4.3}
\end{equation*}
$$

Thus small values of $\tau_{i}$ imply small values of $S_{i}^{\text {tot }}$, that are characteristic for non important variables $x_{i}$.

At the same time from Theorem 1 we obtain a corollary: if $f(x)$ depends linearly on $x_{i}$, then $S_{i}^{\text {tot }}=\frac{\tau_{i}}{D}$. Thus $\tau_{i}$ is closer to $D_{i}^{\text {tot }}$ than $v_{i}$.

Note that the constant factor $1 / \pi^{2}$ in the general inequality for $v_{i}$ is the best possible. But in the general inequality for $\tau_{i}$ the best possible constant factor is unknown.

### 4.1 Example

Consider the so-called $g$-function that is often used in sensitivity analysis for numerical experiments $g(x)=\prod_{i=1}^{n} \frac{\left|4 x_{i}-2\right|+a_{i}}{1+a_{i}}$, where the non negative parameters $a_{i}$ define the importance of $x_{i}$ : the larger $a_{i}$, the less important the input variable $x_{i}$ is. For the $g$-function $D_{i}^{\text {tot }}=d_{i} \prod_{k \neq i}\left(1+d_{k}\right)$, where $d_{k}=\frac{1}{3}\left(1+a_{k}\right)^{-2}$ (details can be found in [1]). It can be proved that $\tau_{i}=4 D_{i}^{\text {tot }}$.

Here are two comments in connection with this example.
First, if $\partial^{2} f / \partial x_{i}^{2} \equiv 0$, then $f(x)$ is a linear function of $x_{i}$, and according to the corollary, $\tau_{i}=D_{i}^{\text {tot }}$. For the $g$-function $\partial^{2} g / \partial x_{i}^{2} \equiv 0$ for all values of $x_{i}$ except $x_{i}=1 / 2$, so in this case $\tau_{i} \neq D_{i}^{\text {tot }}$.

Second, in this example the $\tau_{i}$ are proportional to $S_{i}^{\text {tot }}$ for all $x_{i}$; these $x_{i}$ can be either nonimportant or very important. This example suggests that the $\tau_{i}$ could be used in the same
way as $S_{i}^{\text {tot }}$ for ranking arbitrary variables. The example in the next section shows that this is not always so: ranking important variables according to the values of $\tau_{i}$ and $S_{i}^{\text {tot }}$ can be different.

## Proof of Theorem 2

The inequality declared in Theorem 2 in Section 3 follows immediately from the one dimensional inequality (4.3) and the following two relations:

$$
\begin{equation*}
\tau_{y}=\sum_{p=1}^{s} \tau_{i_{p}} \tag{4.4}
\end{equation*}
$$

and

$$
\begin{equation*}
D_{y}^{\text {tot }} \leq \sum_{p=1}^{s} D_{i_{p}}^{\text {tot }} . \tag{4.5}
\end{equation*}
$$

Indeed $D_{y}^{\text {tot }} \leq \sum_{p=1}^{s} D_{i_{p}}^{\text {tot }} \leq \frac{24}{\pi^{2}} \sum_{p=1}^{s} \tau_{i_{p}}=\frac{24}{\pi^{2}} \tau_{y}$.
Equality (4.4) follows from the definitions of $\tau_{y}$ and $\tau_{i}$. Here is a sketch of a proof of the inequality (4.5). We recall the ANOVA decomposition of $f(x)$ into a sum of orthogonal terms of different dimensions and the corresponding decomposition of the total variance $D$ :

$$
D=\sum_{i} D_{i}+\sum_{i<j} D_{i j}+\cdots+D_{12 \cdots n} .
$$

The total partial variance $D_{y}^{\text {tot }}$ is a sum of all $D_{k_{1} \cdot k_{m}}$ where at least one of the indices $k_{j} \in\left(i_{1}, \ldots, i_{s}\right)$. The partial variance $D_{i_{p}}^{\text {tot }}$ is the sum of all $D_{k_{1} \cdot k_{m}}$ where one of the indices $k_{j}$ is equal to $i_{p}$. Obviously, an arbitrary term $D_{k_{1} \cdots k_{m}}$ from $D_{y}^{\text {tot }}$ has an index equal to sum $i_{p}$, and is included into $D_{i_{p}}^{\text {tot }}$. .

## 5. Counterexample

Consider a function $f$ which has the following ANOVA decomposition:

$$
f=\sum_{i=1}^{4} c_{i}\left(x_{i}-\frac{1}{2}\right)+c_{12}\left(x_{1}-\frac{1}{2}\right)\left(x_{2}-\frac{1}{2}\right)^{5},
$$

where $c_{i}=1,1 \leq i \leq 4, c_{12}=50$. For this function all $S_{i}=0.237,1 \leq i \leq 4, S_{12}=0.0523$ and $S_{1}^{\text {tot }}=S_{2}^{\text {tot }}=0.289, S_{3}^{\text {tot }}=S_{4}^{\text {tot }}=0.237$, so variables 1,2 have the same importance, and so do
variables 3, 4. However, for derivative based importance criteria, variables 1 and 2 have different importance: $\tau_{1}=0.101, \tau_{2}=0.409$, while variables 3 and 4 still have equal importance $\tau_{3}=\tau_{4}=0.083$. Moreover, $\tau_{2}>\tau_{1}+\tau_{3}+\tau_{4}$.

One can see that $\tau_{i} /\left(\pi^{2} D\right)$ is much higher than $S_{i}^{\text {tot }}$ only for variable 2 . It is caused by the strong nonlinearity of the term $f_{1,2}\left(x_{1}, x_{2}\right)$ with respect to $x_{2}$. This example shows that ranking of influential variables based on $\tau_{i}$ may result in false conclusions: in our example $x_{2}$ seems more important than all the other variables together.
6. Variances: $\tau_{i}$ versus $D_{i}^{\text {tot }}$

In this Section Monte Carlo algorithms for computing $\tau_{i}$ and $D_{i}^{\text {tot }}$ are compared and the advantage of the algorithm for computing $\tau_{i}$ is shown. Consider a model function $f(x)$ that is linear with respect to $x_{i}$ :

$$
\begin{equation*}
f(x)=a(z) x_{i}+b(z) \tag{6.1}
\end{equation*}
$$

It follows from Section 4 that for this function $\tau_{i}=D_{i}^{\text {tot }}=\frac{1}{12} \int a^{2}(z) d z$.
Assume that the integral representation (2.2) is applied for computing $D_{i}^{\text {tot }}$

$$
D_{i}^{\text {tot }}=\frac{1}{2} \int a^{2}(z)\left(x_{i}^{\prime}-x_{i}\right)^{2} d x d x_{i}^{\prime} .
$$

If this integral is estimated by crude Monte Carlo, the estimator's variance is

$$
V_{1}=\frac{1}{4} \int a^{4}(z)\left(x_{i}^{\prime}-x_{i}\right)^{4} d x d x_{i}^{\prime}-\left(D_{i}^{t o t}\right)^{2} .
$$

Assume now that formula (4.1) is used for computing $\tau_{i}$ :

$$
\tau_{i}=\int a^{2}(z) \frac{1-3 x_{i}+3 x_{i}^{2}}{6} d x
$$

The crude Monte Carlo estimator's variance of $\tau_{i}$ is

$$
V_{2}=\int a^{4}(z)\left(\frac{1-3 x_{i}+3 x_{i}^{2}}{6}\right)^{2} d x-\left(\tau_{i}\right)^{2}
$$

Denote by $N_{1}$ and $N_{2}$ the numbers of Monte Carlo trials (or the sample sizes) required to achieve a prescribed relative standard error:

$$
\delta=\frac{1}{D_{i}^{\text {tot }}} \sqrt{\frac{V_{1}}{N_{1}}}=\frac{1}{\tau_{i}} \sqrt{\frac{V_{2}}{N_{2}}} .
$$

It follows from these relations that $\frac{N_{1}}{N_{2}}=\frac{V_{1}}{V_{2}}$.

## Assertion 1

The function (6.1) implies $2<\frac{N_{1}}{N_{2}} \leq 7$, where the maximum 7 corresponds to $a(z) \equiv$ Const.

## Proof

Let us introduce an auxiliary parameter $\lambda=\frac{\int a^{4}(z) d z}{\left(\int a^{2}(z) d z\right)^{2}}$. Obviously, $\lambda \geq 1$. The integrals over $d x_{i}$ and $d x_{i}^{\prime}$ in $V_{1}$ and $V_{2}$ can be easily computed. Eliminating $\int a^{4}(z) d z$, one obtains: $\frac{N_{1}}{N_{2}}=\frac{V_{1}}{V_{2}}=\frac{\lambda / 60-1 / 144}{\lambda / 120-1 / 144}$. One can easily verify that the derivative $d\left(N_{1} / N_{2}\right) / d \lambda$ is negative in the interval $1<\lambda<\infty$. Therefore, the maximum value is attained at $\lambda=1$ and the minimum value is approached as $\lambda \rightarrow \infty$.

It follows from Assertion 1, that the sample size for $D_{i}^{\text {tot }}$ must be several times larger than that for $\tau_{i}$.
$6.1 v_{i}$ versus $D_{i}^{\text {tot }}$

Assume now that the criterion $v_{i}$ is used rather than $\tau_{i}$ with

$$
v_{i}=\int a^{2}(z) d z
$$

If this integral is estimated by crude Monte Carlo, the related variance $V_{3}$ will be $V_{3}=(\lambda-1)\left(\int a^{2}(z) d z\right)^{2}$. Denote by $N_{3}$ the sample size that provides the same relative standard error

$$
\delta=\frac{1}{v_{i}} \sqrt{\frac{V_{3}}{N_{3}}} .
$$

Then we obtain:

$$
\frac{N_{1}}{N_{3}}=\frac{144 V_{1}}{V_{3}}=\frac{2.4 \lambda-1}{\lambda-1} .
$$

## Assertion 2

The function (6.1) implies $2.4<\frac{N_{1}}{N_{3}} \leq \infty$ where the maximum $\infty$ corresponds to $a(z) \equiv$ Const.

## Proof

In the interval $1 \leq \lambda<\infty$ the ratio $N_{1} / N_{3}$ decreases from $\infty$ at $\lambda=1$ to 2.4 as $\lambda \rightarrow \infty$.
It follows from Assertion 2, that if the criterion $v_{i}$ is used and $\lambda$ is close to 1 , the required sample size for estimating $D_{i}^{\text {tot }}$ will be by several orders of magnitude larger than the sample size for estimating $v_{i}$. Such an effect has also been numerically observed in [2].

## Remark

We have tried to generalize Assertion 1 to nonlinear functions of the type

$$
f(x)=a(z) x_{i}^{m}+b(z)
$$

where $m>1 / 2$. However, we have found that at all $m \geq 2$, the ratio $N_{1} / N_{2}<1$. Thus $D_{i}^{\text {tot }}$ is superior to $v_{i}$ in terms of the computational effort.

## 7. Nonlinear example

Consider the function

$$
\begin{equation*}
f(x)=\prod_{i=1}^{n} \varphi_{i}\left(x_{i}\right), \tag{7.1}
\end{equation*}
$$

where $\varphi_{i}\left(x_{i}\right)=a_{i}+b_{i}\left(x_{i}-\frac{1}{2}\right)$. Obviously, $\int \varphi_{i} d x_{i}=a_{i}, \int \varphi_{i}^{2} d x_{i}=a_{i}^{2}+\varepsilon_{i}$, where the variance $\varepsilon_{i}=b_{i}^{2} / 12$. Denote $R_{i}=\prod_{t \neq i}\left(a_{t}^{2}+\varepsilon_{t}\right), R_{i j}=\prod_{t \neq i, j}\left(a_{t}^{2}+\varepsilon_{t}\right), R_{i j k}=\prod_{t \neq i, j, k}\left(a_{t}^{2}+\varepsilon_{t}\right)$.

### 7.1 Individual variables

If $y=\left(x_{i}\right)$, then according to Theorem $1, \tau_{y}=D_{y}^{\text {tot }}:$

$$
D_{i}^{t o t}=\tau_{i}=\varepsilon_{i} R_{i} .
$$

Two dimensional set $y=\left(x_{i}, x_{j}\right)$
In this case $\tau_{y}>D_{y}^{\text {tot }}$ :

$$
\begin{aligned}
& D_{y}^{\text {tot }}=\left(a_{i}^{2} \varepsilon_{j}+a_{j}^{2} \varepsilon_{i}+\varepsilon_{i} \varepsilon_{j}\right) R_{i j}, \\
& \tau_{y}=\left(a_{i}^{2} \varepsilon_{j}+a_{j}^{2} \varepsilon_{i}+2 \varepsilon_{i} \varepsilon_{j}\right) R_{i j} .
\end{aligned}
$$

Three dimensional set $y=\left(x_{i}, x_{j}, x_{k}\right)$
Here the difference $\tau_{y}-D_{y}^{t o t}$ increases:

$$
\begin{aligned}
& D_{y}^{\text {tot }}=\left(a_{i}^{2} a_{j}^{2} \varepsilon_{k}+a_{i}^{2} a_{k}^{2} \varepsilon_{j}+a_{j}^{2} a_{k}^{2} \varepsilon_{i}+a_{i}^{2} \varepsilon_{j} \varepsilon_{k}+a_{j}^{2} \varepsilon_{i} \varepsilon_{k}+a_{k}^{2} \varepsilon_{i} \varepsilon_{j}+\varepsilon_{i} \varepsilon_{j} \varepsilon_{k}\right) R_{i j k} ; \\
& \tau_{y}=\left(a_{i}^{2} a_{j}^{2} \varepsilon_{k}+a_{i}^{2} a_{k}^{2} \varepsilon_{j}+a_{j}^{2} a_{k}^{2} \varepsilon_{i}+2 a_{i}^{2} \varepsilon_{j} \varepsilon_{k}+2 a_{j}^{2} \varepsilon_{i} \varepsilon_{k}+2 a_{k}^{2} \varepsilon_{i} \varepsilon_{j}+3 \varepsilon_{i} \varepsilon_{j} \varepsilon_{k}\right) R_{i j k} .
\end{aligned}
$$

One can see that as the dimension of the set $y$ increases, the difference $\tau_{y}-D_{y}^{\text {tot }}$ increases too. However, if all $\varepsilon_{i}$ « $a_{i}^{2}$, then the main terms in $\tau_{y}$ and $D_{y}^{\text {tot }}$ are the same. One can expect that if in (7.1) the variances $\operatorname{Var}\left(\varphi_{i}\right)$ are small, then $\tau_{y}$ will be near to $D_{y}^{\text {tot }}$.

## 8. Normally distributed random variables

Consider independent normal random variables $x_{1}, \ldots, x_{n}$ with parameters $\left(a_{i} ; \sigma_{i}\right)$. Then instead of the importance criterion $\tau_{i}$ given in (4.1) we now get the expression

$$
\begin{equation*}
\tau_{i}=\frac{1}{2} \boldsymbol{E}\left[\left(\frac{\partial f(x)}{\partial x_{i}}\right)^{2}\left(x_{i}^{\prime}-x_{i}\right)^{2}\right] . \tag{8.1}
\end{equation*}
$$

The expectation over $x_{i}^{\prime}$ can be computed analytically. Then we obtain

$$
\begin{equation*}
\tau_{i}=\frac{1}{2} \boldsymbol{E}\left[\left(\frac{\partial f(x)}{\partial x_{i}}\right)^{2} \frac{\left(x_{i}-a_{i}\right)^{2}+\sigma_{i}^{2}}{2}\right] . \tag{8.2}
\end{equation*}
$$

Suppose $f(x)$ is linear with respect to $x_{i}: f(x)=a(z) x_{i}+b(z)$. Then $\tau_{i}=D_{i}^{\text {tot }}=\sigma_{i}^{2} \boldsymbol{E}\left(a^{2}(z)\right)$.

From (8.2) we can obtain the following inequality

$$
\tau_{i} \geq \frac{\sigma_{i}^{2}}{2} \boldsymbol{E}\left[\left(\frac{\partial f(x)}{\partial x_{i}}\right)^{2}\right]
$$

or

$$
\tau_{i} \geq \frac{\sigma_{i}^{2}}{2} v_{i}
$$

Using Theorem 4 from [1] stating that $D_{i}^{\text {tot }} \leq \sigma_{i}^{2} v_{i}$, we obtain

$$
D_{i}^{\text {tot }} \leq 2 \tau_{i}
$$

or

$$
S_{i}^{\text {tot }} \leq \frac{2 \tau_{i}}{D} .
$$

Using this inequality we can easily prove the following theorem.

Theorem 3. If $x_{1}, \ldots, x_{n}$ are independent normal random variables, then for an arbitrary subset $y$ of these variables

$$
D_{y}^{\text {tot }} \leq 2 \tau_{y}
$$

or in other words

$$
S_{y}^{\text {tot }} \leq \frac{2 \tau_{y}}{D} .
$$

## 9. Conclusions

A new derivative based criterion $\tau_{y}$ for groups of variables is derived. We also introduced a new criterion for a single variable $\tau_{i}$, which is a modification of the criteria studied in our previous work. A link between global sensitivity indices and new derivative based measures is established. It is shown that for functions linear with respect to a group of variables the performance of $\tau_{y}$ for the group is equivalent to the performance of the sensitivity index $S_{y}^{\text {tot }}$ for the same group. It is proved that small values of derivative based measures imply small values of total sensitivity indices. However, for highly nonlinear functions the ranking of important factors for variance based and derivative based measures can be different.

The computational costs of evaluating global sensitivity indices and derivative based measures for one class of functions were compared. It was shown that for linear and quasi-linear functions derivative based measures require fewer function evaluations, which confirms earlier numerical findings. For highly non-linear functions, however, evaluations of global sensitivity indices can be cheaper.

## Practical Recommendations

For estimating the impact upon $f(x)$ from a group of variables $y$, we recommend to use the global sensitivity index $S_{y}^{\text {tot }}$. If the computation of this index seems too expensive, one can try to apply the derivative based criterion $\tau_{y}$, taking into account the following facts:

1) If the dependence of $f(x)$ on the variables from $y$ is nearly linear, then the values of $\tau_{y} / D$ are near to the values of $S_{y}^{\text {tot }}$.
2) In the strongly nonlinear case, small values of $\tau_{y}$ are always significant: they imply small values of $S_{y}^{\text {tot }}$. At the same time, large values of $\tau_{y}$ are less informative: the ratio $\tau_{y} / D$ can be arbitrarily large, while $S_{y}^{\text {tot }}$ does not exceed 1 .

## Acknowledgements

The authors would like to thank A. Sobol' for his help in preparation of this manuscript, N . Shah for his support and interest in this work, and anonymous referees for their valuable comments. One of the authors (SK) gratefully acknowledges the financial support by the EPSRC grant EP/D506743/1.

## Appendix. A simple test for importance measures

Consider a function $f(x)$ where all the factors $x_{1}, \ldots, x_{n}$ are equivalent and do not interact: $f(x)=\sum_{i=1}^{n} u\left(x_{i}\right)$. Here $u(t)$ is an arbitrary function, $0 \leq t \leq 1$. From general symmetry
considerations one can expect that a reasonable importance measure for the set $y=\left(x_{1}, \ldots, x_{s}\right)$ must be proportional to $s$.

## Global sensitivity indices

Assume that $u(t) \in L_{2}$. One can easily find out, that $S_{y}=S_{y}^{\text {tot }}=\frac{s}{n}$.

## Importance criterion $\tau_{y}$

Assume that $u^{\prime}(t) \in L_{2}$. From (3.1)

$$
\tau_{y}=s \alpha,
$$

where

$$
\alpha=\int_{0}^{1}\left[u^{\prime}(t)\right]^{2} \frac{1-3 t+3 t^{2}}{6} d t
$$

## Importance measure $\mu^{*}$

The very first attempt to define a derivative based importance measure for groups of factors was made in [3].

The importance measure $\mu^{*}$ for the set $y=\left(x_{1}, \ldots, x_{s}\right)$ is defined as $\mu^{*}=\frac{1}{N} \sum_{k=1}^{N} \eta\left(x^{(k)}\right)$, where the $x^{(k)}$ are independent trial points, and $\eta(x)$ (called "elementary effect") is $\eta(x)=\frac{|f(\tilde{x})-f(x)|}{\Delta}, \tilde{x}=\left(x_{1} \pm \Delta, \ldots, x_{s} \pm \Delta, x_{s+1}, \ldots, x_{n}\right) ;$ the signs + and - are random and equiprobable.

Consider the simplest test: $u(t)=t$ (which means that $f=x_{1}+\cdots+x_{n}$ ). In this case $\eta(x)$ depends neither on $x$, nor on $\Delta$ : it is simply a random variable $\eta=| \pm 1 \pm 1 \cdots \pm 1|$, where the number of units is equal to $s$. Obviously if $N \rightarrow \infty$ the value $\mu^{*} \xrightarrow{\boldsymbol{P}} \boldsymbol{E} \eta$. Here are the distributions and expectations of $\eta$ at $s=1,2,3,4,5$.

$$
s=1: \eta \equiv 1, E \eta=1,
$$

$$
\begin{aligned}
& s=2: \eta \equiv\left(\begin{array}{cc}
2 & 0 \\
\frac{1}{2} & \frac{1}{2}
\end{array}\right), E \eta=1, \\
& s=3: \eta \equiv\left(\begin{array}{cc}
3 & 1 \\
\frac{1}{4} & \frac{3}{4}
\end{array}\right), E \eta=\frac{3}{2}, \\
& s=4: \eta \equiv\left(\begin{array}{ccc}
4 & 2 & 0 \\
\frac{1}{8} & \frac{4}{8} & \frac{3}{8}
\end{array}\right), E \eta=\frac{3}{2}, \\
& s=5: \eta \equiv\left(\begin{array}{ccc}
5 & 3 & 1 \\
\frac{1}{16} & \frac{5}{16} & \frac{10}{16}
\end{array}\right), E \eta=\frac{15}{8} .
\end{aligned}
$$

Clearly, the simplest test for $\mu^{*}$ fails.

## References

[1] I.M. Sobol', S. Kucherenko, Math. Computer Simulat. 79 (2009), 3009
[2] S. Kucherenko, M.Rodriguez-Fernandez, C.Pantelides, N.Shah, Rel.Eng. \& Syst. Safet. 94 (2009) 1135
[3] F. Campolongo, J. Cariboni, A. Saltelli, Environ Model Software 22 (2007) 1509
[4] I.M. Sobol', Matem. Modelirovanie 2(1990), 112 (in Russian). English translation: Math.
Modelling and Comput. Experiment 4 (1993) 407
[5] I.M. Sobol', Math. Computer Simulat, 55(2001) 271
[6] A. Saltelli, Comput. Phys. Commun. 145 (2002) 280
[7] I.M. Sobol, S. Kucherenko, Monte Carlo Methods Applicat. 11 (2005) 83
[8] I.M. Sobol, Matem. Modelirovanie 17 (2005) 43. Additonal: Matem. Modelirovanie 19 (2007) 23 (in Russian)
[9] I.M. Sobol', S. Tarantola, D. Gatelli, S. Kucherenko, W. Mauntz, Rel.Eng. \& Syst.Safet 92 (2007) 957.
[10] I.M. Sobol', B.V. Shukhman, Monte Carlo Methods Applicat. 13 (2007) 89
[11] A. Saltelli, K. Chan, E.M. Scott, Eds, Sensitivity Analysis ( John Wiley \& Sons) 2000
[12] A. Saltelli, S. Tarantola, F. Campolongo, M. Ratto, Sensitivity Analysis in Practice (John Wiley \& Sons) 2004
[13] S. Kucherenko, N. Shah, Wilmott, July (2007) 82
[14] M.D. Morris, Technometrics 33 (1991) 161
[15] A. Kiparissides, S. Kucherenko, A. Mantalaris, E.N. Pistikopoulos, Ind. Eng. Chem. Res. 48 (2009) 7168
[16] I.M. Sobol', E. Myshetskaya, Monte Carlo Methods and Applicat. 13 (2007) 455


[^0]:    ${ }^{1}$ Corresponding author
    E-mail address: s.kucherenko@ic.ac.uk

