J Math Chem (2014) 52:2052–2073
DOI 10.1007/s10910-014-0365-6

ORIGINAL PAPER

Analytical HDMR formulas for functions expressed as quadratic polynomials with a multivariate normal distribution

Genyuan Li · Herschel Rabitz

Received: 9 May 2014 / Accepted: 19 May 2014 / Published online: 28 May 2014

Abstract High Dimensional Model Representation (HDMR) is a general set of quantitative model assessment and analysis tools for systems with many variables. A general formulation for the HDMR component functions with independent and correlated variables was obtained previously. Since the HDMR component functions generally are coupled to one another and involve multi-dimensional integrals, explicit formulas for the component functions are not available for an arbitrary function with an arbitrary probability distribution amongst their variables. This paper presents analytical formulas for the HDMR component functions and the corresponding sensitivity indexes for the common case of a function expressed as a quadratic polynomial with a multivariate normal distribution over its variables. This advance is important for practical applications of HDMR with correlated variables.

Keywords HDMR · Correlated variables · Sensitivity analysis

1 Introduction

Many problems in science and engineering reduce to the need for efficiently constructing a map of the relationship between a set of high dimensional system inputs $\mathbf{x}$ and the system output $f(\mathbf{x})$. Let $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu)$ be a probability space, where $\mathbb{R}^n$ is a sample space, $\mathcal{B}(\mathbb{R}^n)$ denotes the the Borel σ-algebra on $\mathbb{R}^n$, and μ is a probability measure with $\mathrm{d}\mu = p(\mathbf{x})\mathrm{d}\mathbf{x}$ where $p(\mathbf{x})$ is the probability density function (pdf). Suppose $f(\mathbf{x})$

Electronic supplementary material The online version of this article (doi:10.1007/s10910-014-0365-6) contains supplementary material, which is available to authorized users.

G. Li · H. Rabitz (✉)
Department of Chemistry, Princeton University, Princeton, NJ 08544, USA
e-mail: hrabitz@princeton.edu

G. Li
e-mail: genyuan@princeton.edu

is a real square integrable function of a random vector $\mathbf{x} \in \mathbb{R}^n$. As the contributions of the multiple input variables upon the output can be independent and cooperative, it is natural to express $f(\mathbf{x})$ as a finite hierarchical expansion:

$$f(\mathbf{x}) = \sum_{u \subseteq n} f_u(\mathbf{x}_u). \tag{1.1}$$

Here we use the following multi-index notation. Given the subset $u \subseteq \{1, 2, \ldots, n\}$, we denote by $\mathbf{x}_u$ those variables in $\mathbf{x}$ whose indexes are in u, and $\mathbf{x}_{-u}$ denotes the subset of variables in $\mathbf{x}$ with indexes not in u. Note that the empty set $\emptyset$ is a subset of $\{1, 2, \ldots, n\}$ and we have $f_\emptyset = f_0$, a constant. We will also write $u \subseteq n$ in place of $u \subseteq \{1, 2, \ldots, n\}$ for simplicity. Decomposition in Eq. 1.1 was first introduced by Fisher [1] and is known as the ANOVA decomposition, which was also employed for studying U(unbiased)-statistics by Hoeffding [2].

Sobol introduced the vanishing condition for the component functions in the ANOVA decomposition [3–5]

$$\int_{\mathbb{K}^1} f_u(\mathbf{x}_u) \mathrm{d}x_i = 0, \quad i \in u, \quad \emptyset \neq u \subseteq n \tag{1.2}$$

for $\mathbb{K}^n$ being an n-dimensional hypercube $[0, 1]^n$. This condition uniquely defines the component functions as

$$f_u(\mathbf{x}_u) = \int_{\mathbb{K}^{n-|u|}} f(\mathbf{x}) \mathrm{d}\mathbf{x}_{-u} - \sum_{v \subset u} f_v(\mathbf{x}_v), \quad u \subseteq n, \tag{1.3}$$

where $|u|$ denotes the cardinality of u, which are mutually orthogonal

$$\mathbb{E}[f_u(\mathbf{x}_u) f_v(\mathbf{x}_v)] = 0, \quad u \neq v, \tag{1.4}$$

where $\mathbb{E}[\cdot]$ denotes the expectation. Since $\emptyset \subset n$, we have

$$\mathbb{E}[f_u(\mathbf{x}_u) f_0] = f_0 \mathbb{E}[f_u(\mathbf{x}_u)] = 0, \quad u \neq \emptyset. \tag{1.5}$$

Equation 1.5 is valid for an arbitrary function $f(\mathbf{x})$ and an arbitrary probability distribution and hence f_0 may be or may not be zero. Therefore, Eq. 1.5 is valid if and only if

$$\mathbb{E}[f_u(\mathbf{x}_u)] = 0, \quad u \neq \emptyset, \tag{1.6}$$

i.e., the expected value of any non-constant component function is zero. Equation 1.6 also implies that the expected value of $f(\mathbf{x})$ is f_0.

$$\mathbb{E}[f(\mathbf{x})] = \mathbb{E}\left[\sum_{u \subseteq n} f_u(\mathbf{x}_u)\right] = \mathbb{E}[f_0] = f_0. \tag{1.7}$$

Combining the mutual orthogonality and zero expectation of the non-constant component functions directly leads to the decomposition of the variance of $f(\mathbf{x})$

$$\mathrm{Var}(f(\mathbf{x})) = \mathbb{E}\left[(f(\mathbf{x}) - f_0)^2\right] = \sum_{\emptyset \neq u \subseteq n} \mathbb{E}\left[f_u(\mathbf{x}_u)^2\right] = \sum_{\emptyset \neq u \subseteq n} \mathrm{Var}(f_u(\mathbf{x}_u)) \quad (1.8)$$

and the definition of sensitivity indexes

$$1 = \sum_{\emptyset \neq u \subseteq n} \frac{\mathrm{Var}(f_u(\mathbf{x}_u))}{\mathrm{Var}(f(\mathbf{x}))} = \sum_{\emptyset \neq u \subseteq n} S_u \quad (1.9)$$

which establishes the theoretical basis for the variance-based method of sensitivity analysis [6–13].

Rabitz and Alis [14] extended Sobol's formulas to any product type measure

$$\mathrm{d}\mu = \prod_{i=1}^{n} \mathrm{d}\mu_i = \prod_{i=1}^{n} p_i(x_i)\mathrm{d}x_i \quad (1.10)$$

and used the generalized vanishing condition

$$\int_{\mathbb{K}^1} f_u(\mathbf{x}_u)\mathrm{d}\mu_i = \int_{\mathbb{K}^1} f_u(\mathbf{x}_u)p_i(x_i)\mathrm{d}x_i = 0, \quad i \in u \quad (1.11)$$

to obtain

$$f_u(\mathbf{x}_u) = \int_{\mathbb{K}^{n-|u|}} f(\mathbf{x})p_{-u}(\mathbf{x}_{-u})\mathrm{d}\mathbf{x}_{-u} - \sum_{v \subset u} f_v(\mathbf{x}_v), \quad u \subseteq n, \quad (1.12)$$

where $p_{-u}(\mathbf{x}_{-u}) = \prod_{i \notin u} p_i(x_i)$ is the marginal pdf for $\mathbf{x}_{-u}$. This definition includes Sobol's formulas as a special case for $p_i(x_i) = 1, \forall i$. As the formula of $f_u(\mathbf{x}_u)$ given in Eq. 1.12 only contains lower order component functions $f_v(\mathbf{x}_v)$, $v \subset u$, all the component functions can be obtained sequentially starting from f_0.

The extension to an arbitrary product measure makes it possible to generate various alternative forms for the component functions. Rabitz and Alis denoted Eq. 1.1 with component functions defined by Eq. 1.12 as High Dimensional Model Representation (HDMR), and called the HDMR with all $p_i(x_i) = 1$ as ANOVA-HDMR. Another type of HDMR referred to as cut-HDMR was constructed for

$$\mathrm{d}\mu = \prod_{i=1}^{n} \delta(x_i - \bar{x}_i)\mathrm{d}x_i \quad (1.13)$$

where δ is the Dirac delta function, $\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n)$ is a chosen reference point in $\mathbf{x}$ space. The component functions of cut-HDMR possess the form:

$$f_u(\mathbf{x}_u) = f(x_u, \bar{\mathbf{x}}_{-u}) - \sum_{v \subset u} f_v(\mathbf{x}_v), \quad u \subseteq n, \quad (1.14)$$

where $(x_u, \bar{\mathbf{x}}_{-u})$ denotes $\mathbf{x}$ with $x_i = \bar{x}_i$ for $i \notin u$. The cut-HDMR component functions are constructed as numerical data tables along lines, planes and other higher dimensional subvolumns through the reference point $\bar{\mathbf{x}}$. Modified algorithms of cut-HDMR to improve the accuracy and reduce the sample size were developed [15, 16]. Sobol [17] proved that cut-HDMR is actually an HDMR decomposition of the difference $f(\mathbf{x}) - f(\bar{\mathbf{x}})$, and referred to it as finite difference HDMR. Sobol [17] also discussed the influence of the choice of the reference point $\bar{\mathbf{x}}$.

Alis and Rabitz [18] proposed to approximate the HDMR component functions defined in Eq. 1.12 by a combination of linearly independent basis functions $\phi_{uk}(\mathbf{x}_u)$

$$f_u(\mathbf{x}_u) \approx \sum_{k=1}^{s} c_{uk}\phi_{uk}(\mathbf{x}_u), \quad \emptyset \neq u \subseteq n \tag{1.15}$$

where c_{uk} are constant combination coefficients, s is an integer, $\phi_{uk}(\mathbf{x}_u)$ are polynomials, orthogonal bases, splines, etc. satisfying the conditions:

$$\mathbb{E}[\phi_{uk}(\mathbf{x}_u)] = 0, \quad \forall k \tag{1.16}$$

$$\int_{\mathbb{K}^1} \phi_{uk}(\mathbf{x}_u) p_i(x_i) dx_i = 0, \quad i \in u. \tag{1.17}$$

As a special case, $\phi_{uk}(\mathbf{x}_u)$ can be a product of one variable basis functions

$$\phi_{uk}(\mathbf{x}_u) = \prod_{i \in u} \phi_{ik_i}(x_i). \tag{1.18}$$

These basis functions may be either classical orthonormal polynomials with a given measure μ or orthonormal polynomials constructed from the collected data following an implicit probability distribution by Gram–Schmidt orthogonalization. The resultant approximate HDMR component functions still satisfy the characteristic property of HDMR component functions: zero expectation and mutual orthogonality. The advantage of the basis function approximation is that using a single set of random samples of $\mathbf{x}$ generated according to the product pdf, then all the component functions for a truncated HDMR expansion can be obtained by regression to determine the combination coefficients c_{uk}. This version of HDMR was denoted as Random Sampling (RS)-HDMR [18–21].

The above work is based on the assumption that all the variables x_i are independent. In practice, very often the variables x_i are correlated or dependent, i.e., the measure μ is no longer of a product type. Thus, the theoretical basis and numerical algorithms of HDMR developed above do not apply in this circumstance. The work of Hooker [22] provides a foundation to deal with this problem by relaxing the vanishing condition, Eq. 1.11 to

$$\int f_u(\mathbf{x}_u) p(\mathbf{x}) \mathrm{d}x_i \mathrm{d}\mathbf{x}_{-u} = \int f_u(\mathbf{x}_u) p_u(\mathbf{x}_u) \mathrm{d}x_i = 0, \quad i \in u, \quad \emptyset \neq u \subseteq n, \tag{1.19}$$

where $p_u(\mathbf{x}_u)$ is the marginal pdf for $\mathbf{x}_u$, under the condition that the support of $p(\mathbf{x})$ is grid closed. For the sake of notational simplicity, we omit the specific integration dimension and range and use $\int$ to represent all integrations. Equation 1.19 includes the vanishing condition, and Eq. 1.11 as a special case for independent variables when

$$p_u(\mathbf{x}_u) = \prod_{i \in u} p_i(x_i). \tag{1.20}$$

The criterion, Eq. 1.19, is equivalent to the *hierarchical orthogonality* condition of the HDMR component functions

$$\mathbb{E}[f_u(\mathbf{x}_u) f_v(\mathbf{x}_v)] = 0, \quad v \subset u, \tag{1.21}$$

i.e., *a component function is only required to be orthogonal to all nested lower order component functions whose variables are a subset of its variables.* Hooker proved that under the relaxed vanishing condition, or equivalently the hierarchical orthogonality condition, the HDMR component functions are unique.

Based on the relaxed vanishing condition, Li and Rabitz [23] deduced the general formulas for HDMR component functions for independent and correlated variables as

$$f_0 = \int f(\mathbf{x}) p(\mathbf{x}) \mathrm{d}\mathbf{x}, \tag{1.22}$$

$$f_i(x_i) = \int f(\mathbf{x}) p_{-i}(\mathbf{x}_{-i}) \mathrm{d}\mathbf{x}_{-i} - f_0 - \sum_{\{i\} \subset v \subseteq n} \int f_v(\mathbf{x}_v) p_{-i}(\mathbf{x}_{-i}) \mathrm{d}\mathbf{x}_{-i}, \tag{1.23}$$

$$\begin{aligned} f_{ij}(x_i, x_j) = & \int f(\mathbf{x}) p_{-ij}(\mathbf{x}_{-ij}) \mathrm{d}\mathbf{x}_{-ij} - f_0 - f_i(x_i) - f_j(x_j) \\ & - \sum_{\substack{v \subseteq n \\ \{i,j\} \bigcap v \neq \emptyset}} \int f_v(\mathbf{x}_v) p_{-ij}(\mathbf{x}_{-ij}) \mathrm{d}\mathbf{x}_{-ij}, \end{aligned} \tag{1.24}$$

$$\cdots$$

or in a single formula [24],

$$\begin{aligned} f_u(\mathbf{x}_u) = & \int f(\mathbf{x}) p_{-u}(\mathbf{x}_{-u}) \mathrm{d}\mathbf{x}_{-u} - \sum_{v \subset u} f_v(\mathbf{x}_v) \\ & - \sum_{\substack{u \not\supseteq v \subseteq n \\ u \bigcap v \neq \emptyset}} \int f_v(\mathbf{x}_v) p_{-u}(\mathbf{x}_{-u}) \mathrm{d}\mathbf{x}_{-u}, \quad u \subseteq n. \end{aligned} \tag{1.25}$$

For independent variables the last term vanishes, and Eq. 1.25 reduces to Eq. 1.12.

Using the hierarchical orthogonality condition, the last term of Eq. 1.25 can be re-expressed leading to [24]

$$f_u(\mathbf{x}_u) = \int f(\mathbf{x}) p_{-u}(\mathbf{x}_{-u}) \mathrm{d}\mathbf{x}_{-u} - \sum_{v \subset u} f_v(\mathbf{x}_v)$$

$$-\sum_{\substack{u\not\supseteq v\subseteq n\\ u\bigcap v\neq\emptyset}}\int f_v(\mathbf{x}_v)p_{v\cap -u}(\mathbf{x}_{v\cap -u})\mathrm{d}\mathbf{x}_{v\cap -u},\quad u\subseteq n. \tag{1.26}$$

As shown later, we do not need to directly calculate the last term. Thus, for brevity, we will use Eq. 1.25.

Equation 1.25 implies that for correlated variables all the component functions are coupled together. The explicit form of each component function may not be obtained separately for an arbitrary function $f(\mathbf{x})$ and an arbitrary measure μ, although Li and Rabitz [23,25] and Li et al. [26] developed a numerical method to determine the HDMR component functions. They suggested to approximate the HDMR component functions as expansions for a set of suitable basis functions, and then use D-MORPH regression to determine the expansion coefficients such that the hierarchically orthogonal condition for the component functions is satisfied.

Since the HDMR component functions with correlated variables are not mutually orthogonal, the standard variance decomposition of the total output variance does not hold, and the variance-based sensitivity analysis is not valid as well for correlated variables. However, based on a covariance decomposition, a general global sensitivity analysis for independent and correlated variables, referred as structural (independent) and correlative sensitivity analysis (SCSA) was proposed by Li et al. [27]

$$\begin{aligned}
\mathrm{Var}(f(\mathbf{x})) &= \mathbb{E}\Big[(f(\mathbf{x})-f_0)^2\Big] = \mathbb{E}\Big[(f(\mathbf{x})-f_0)\sum_{\emptyset\neq u\subseteq n} f_u(\mathbf{x}_u)\Big]\\
&= \sum_{\emptyset\neq u\subseteq n}\mathbb{E}\left[(f(\mathbf{x})-f_0)(f_u(\mathbf{x}_u)-0)\right] = \sum_{\emptyset\neq u\subseteq n}\mathrm{Cov}\left(f(\mathbf{x}), f_u(\mathbf{x}_u)\right)\\
&= \sum_{\emptyset\neq u\subseteq n}\mathrm{Cov}\left(f(\mathbf{x})-f_u\left(\mathbf{x}_u\right)+f_u(\mathbf{x}_u), f_u(\mathbf{x}_u)\right)\\
&= \sum_{\emptyset\neq u\subseteq n}\left[\mathrm{Var}\big(f_u(\mathbf{x}_u)\big)+\mathrm{Cov}\big(f(\mathbf{x})-f_u(\mathbf{x}_u), f_u(\mathbf{x}_u)\big)\right]\\
&= \sum_{\emptyset\neq u\subseteq n}\left[\mathrm{Var}(f_u(\mathbf{x}_u))+\mathrm{Cov}\Big(\sum_{u\neq v\subseteq n} f_v(\mathbf{x}_v), f_u\left(\mathbf{x}_u\right)\Big)\right].
\end{aligned}\tag{1.27}$$

The general sensitivity indexes are defined by dividing both sides of Eq. 1.27 with $\mathrm{Var}(f(\mathbf{x}))$.

$$\begin{aligned}
1 &= \sum_{\emptyset\neq u\subseteq n}\left[\frac{\mathrm{Var}\left(f_u(\mathbf{x}_u)\right)}{\mathrm{Var}\left(f(\mathbf{x})\right)}+\frac{\mathrm{Cov}\left(\sum_{u\neq v\subseteq n} f_v(\mathbf{x}_v), f_u(\mathbf{x}_u)\right)}{\mathrm{Var}\left(f(\mathbf{x})\right)}\right]\\
&= \sum_{\emptyset\neq u\subseteq n}[S_u^a+S_u^b] = \sum_{\emptyset\neq u\subseteq n} S_u,
\end{aligned}\tag{1.28}$$

where S_u^a, S_u^b and S_u are respectively referred to as *structural (independent), correlative* and *total* sensitivity indexes for $\mathbf{x}_u$. For independent variables, all of the component functions are mutually orthogonal, i.e, their covariances are all zero. Then $S_u^b = 0$, and the three sensitivity indexes S_u^a, S_u^b and S_u reduce to a single index S_u utilizing the definition in Eq. 1.9 given by the variance-based method of sensitivity analysis.

The fact that explicit formulas for the HDMR component functions cannot be obtained for an arbitrary function with an arbitrary probability distribution over the variables increases the effort of applying HDMR in some practical applications. However, many systems are *exactly* described or *satisfactorily* approximated by quadratic polynomial functions $f(\mathbf{x})$, and the probability distribution over the variables is *commonly* considered to be a multivariate normal distribution. This paper presents analytical formulas for such HDMR component functions and the sensitivity indexes based on these HDMR component functions, which significantly reduces the computational effort in this class of practical applications of HDMR with correlated variables.

The remainder of the paper is organized as follows. The formulas for the HDMR component functions and their corresponding SCSA sensitivity indexes for quadratic functions are given in Sect. 2. The formulas for linear polynomials are included as a special case. Section 3 gives two illustrative examples. Finally, some concluding remarks are given in Sect. 4. The details of the mathematical derivations are given in the supplemental material.

2 A quadratic polynomial function with a multivariate normal distribution

A quadratic polynomial function with a multivariate normal distribution over its variables is commonly used in many practical circumstances. However, the construction of its HDMR component functions and the determination of the corresponding sensitivity indexes are computationally demanding tasks [28–31]. The analytical formulas for the HDMR component functions and the corresponding sensitivity indexes provided here not only significantly reduce the computational effort, but also provide clear deterministic and statistic interpretations for the sensitivity indexes.

Suppose that for $\mathbf{x} = (x_1, x_2, \ldots, x_n)$

$$y = f(\mathbf{x}) = \alpha_0 + \sum_{i=1}^{n} \alpha_i x_i + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \beta_{ij} x_i x_j + \sum_{i=1}^{n} \gamma_i x_i^2 \tag{2.1}$$

with $\beta_{ij} = \beta_{ji}$, and $\mathbf{x}$ possesses a multivariate normal distribution

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} \mid \Sigma \mid^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)\right), \tag{2.2}$$

where $\mu = (\mu_1, \mu_2, \ldots, \mu_n)$ is the expected value of $\mathbf{x}$, Σ is the covariance matrix of $\mathbf{x}$

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \cdots & \rho_{2n}\sigma_2\sigma_n \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1}\sigma_n\sigma_1 & \rho_{n2}\sigma_n\sigma_2 & \cdots & \sigma_n^2 \end{bmatrix} \quad (2.3)$$

with

$$\sigma_{ij} = \sigma_{ji}, \quad \rho_{ij} = \rho_{ji}. \quad (2.4)$$

For each x_i and each pair (x_i, x_j), their marginal distributions are also normal, i.e.,

$$p_i(x_i) = \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right], \quad (2.5)$$

$$p_{ij}(x_i, x_j) = \frac{1}{2\pi\sigma_i\sigma_j\sqrt{1-\rho_{ij}^2}} \exp\left[-\frac{1}{2(1-\rho_{ij}^2)}\left(\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2 - 2\rho_{ij}\left(\frac{x_i - \mu_i}{\sigma_i}\right)\left(\frac{x_j - \mu_j}{\sigma_j}\right) + \left(\frac{x_j - \mu_j}{\sigma_j}\right)^2\right)\right]. \quad (2.6)$$

To facilitate the treatment, all variables are first set to lie in the same range by transforming x_i to the canonical variable

$$z_i = \frac{x_i - \mu_i}{\sigma_i}, \quad (i = 1, 2, \ldots, n) \quad (2.7)$$

with zero mean and unit standard deviation. This transformation is similar to the transformation

$$z_i = \frac{x_i - a_i}{b_i - a_i} \quad (2.8)$$

for x_i with a uniform distribution, where a_i and b_i are lower and upper bounds of x_i such that all z_i's have the same range [0,1].

With new variables z_i, Eq. 2.1 becomes

$$y = f(\mathbf{z}) = a_0 + \sum_{i=1}^{n} a_i z_i + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} a_{ij} z_i z_j + \sum_{i=1}^{n} b_i z_i^2 \quad (2.9)$$

where

$$a_0 = \alpha_0 + \sum_{i=1}^{n} \alpha_i \mu_i + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \beta_{ij} \mu_i \mu_j + \sum_{i=1}^{n} \gamma_i \mu_i^2, \tag{2.10}$$

$$a_i = \alpha_i \sigma_i + \sum_{\substack{k=1 \\ k \neq i}}^{n} 2\beta_{ik} \mu_k \sigma_i + 2\gamma_i \mu_i \sigma_i, \tag{2.11}$$

$$a_{ij} = a_{ji} = \beta_{ij} \sigma_i \sigma_j, \tag{2.12}$$

$$b_i = \gamma_i \sigma_i^2. \tag{2.13}$$

The pdf $p(\mathbf{z})$ for $\mathbf{z}$ is the standard multivariate normal distribution

$$p(\mathbf{z}) = \frac{1}{(2\pi)^{n/2} \mid \Sigma \mid^{1/2}} \exp\left(-\frac{1}{2}\mathbf{z}^T \Sigma^{-1} \mathbf{z}\right), \tag{2.14}$$

with $\mu = \mathbf{0}$ and

$$\Sigma = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & 1 \end{bmatrix}. \tag{2.15}$$

2.1 Analytical formulas for the HDMR component functions

For a quadratic polynomial function with a multivariate normal distribution of $\mathbf{z}$, the HDMR expansion only contains the terms f_0, $f_i(x_i)$ and $f_{ij}(x_i, x_j)$. The analytical formulas will be given below, and in the supplemental material with further details.

1. Formula for $\mathbf{f_0}$

$$f_0 = \int f(\mathbf{z}) p(\mathbf{z}) \mathrm{d}\mathbf{z} = a_0 + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} a_{ij} \rho_{ij} + \sum_{i=1}^{n} b_i. \tag{2.16}$$

2. Analytical formulas for $\mathbf{f_i(z_i)}$ and $\mathbf{f_{ij}(z_i, z_j)}$

For correlated variables, $f_i(z_i)$ and $f_{ij}(z_i, z_j)$ are defined as

$$f_i(z_i) = \int f(\mathbf{z}) p_{-i}(\mathbf{z}_{-i}) \mathrm{d}\mathbf{z}_{-i} - f_0 - h_i(z_i), \tag{2.17}$$

$$f_{ij}(z_i, z_j) = \int f(\mathbf{z}) p_{-ij}(\mathbf{z}_{-ij}) \mathrm{d}\mathbf{z}_{-ij} - f_0 - f_i(z_i) - f_j(z_j) - h_{ij}(z_i, z_j), \tag{2.18}$$

where $h_i(z_i)$, $h_{ij}(z_i, z_j)$ denote the last terms in Eqs. 1.23, 1.24. Expressing the functions $h_i(z_i)$, $h_j(z_j)$ and $h_{ij}(z_i, z_j)$ as polynomials up to degree 2 (this is reasonable because $f(\mathbf{z})$ is a polynomial function of that degree):

$$h_i(z_i) = e_2 z_i^2 + e_1 z_i + e_0, \tag{2.19}$$

$$h_j(z_j) = e_2' z_j^2 + e_1' z_j + e_0', \tag{2.20}$$

$$h_{ij}(z_i, z_j) = e_{2i}'' z_i^2 + e_{ij}'' z_i z_j + e_{2j}'' z_j^2 + e_{1i}'' z_i + e_{1j}'' z_j + e_0'', \tag{2.21}$$

and applying the relaxed vanishing condition to $f_i(z_i)$, $f_j(z_j)$ and $f_{ij}(z_i, z_j)$ yields (see supplemental material)

$$h_i(z_i) = -\sum_{\substack{k=1\\k\neq i}}^{n} \frac{2a_{ik}\rho_{ik}}{1+\rho_{ik}^2} z_i^2 - \sum_{\substack{k=1\\k\neq i}}^{n} \frac{2a_{ik}\rho_{ik}^3}{1+\rho_{ik}^2}, \quad \forall i, \tag{2.22}$$

$$\begin{aligned} h_{ij}(z_i, z_j) = &-\sum_{\substack{k=1\\k\neq i,j}}^{n} \frac{2a_{ik}\rho_{ik}}{1+\rho_{ik}^2} z_i^2 - \sum_{\substack{k=1\\k\neq i,j}}^{n} \frac{2a_{jk}\rho_{jk}}{1+\rho_{jk}^2} z_j^2 \\ &-\sum_{\substack{k=1\\k\neq i,j}}^{n} \frac{2a_{ik}\rho_{ik}^3}{1+\rho_{ik}^2} - \sum_{\substack{k=1\\k\neq i,j}}^{n} \frac{2a_{jk}\rho_{jk}^3}{1+\rho_{jk}^2}. \end{aligned} \tag{2.23}$$

Substituting Eqs. 2.22, 2.23 into Eqs. 2.17, 2.18 we obtain

$$f_i(z_i) = \left(b_i + \sum_{\substack{k=1\\k\neq i}}^{n} \frac{2a_{ik}\rho_{ik}}{1+\rho_{ik}^2} \right) z_i^2 + a_i z_i - \left(b_i + \sum_{\substack{k=1\\k\neq i}}^{n} \frac{2a_{ik}\rho_{ik}}{1+\rho_{ik}^2} \right), \tag{2.24}$$

$$f_{ij}\left(z_i, z_j\right) = 2a_{ij} z_i z_j - \frac{2a_{ij}\rho_{ij}}{1+\rho_{ij}^2}\left(z_i^2 + z_j^2\right) - 2a_{ij}\rho_{ij}\frac{\rho_{ij}^2 - 1}{1+\rho_{ij}^2}. \tag{2.25}$$

The formulas for f_0, $f_i(z_i)$ and $f_{ij}(z_i, z_j)$ have clear meaning as explained below.

- Since the z_i variables are correlated, the variation of z_i causes a variation of z_k through the correlation coefficient ρ_{ik}. This results in the term $a_{ik} z_i z_k$ having a contribution as the quadratic form

$$\frac{2a_{ik}\rho_{ik}}{1+\rho_{ik}^2} z_i^2, \quad (k = 1, 2, \ldots, n;\, k \neq i). \tag{2.26}$$

  Thus, the structural (independent) contribution of z_i reflected by $f_i(z_i)$ not only contains $a_i z_i$ and $b_i z_i^2$ in $f(\mathbf{z})$, but also contains an extra term given by the sum of Eq. 2.26 over k.
- To assure that $f_i(z_i)$ and $f_{ij}(z_i, z_j)$ are hierarchically orthogonal, the contribution above is subtracted off in $f_{ij}(z_i, z_j)$.

- Equations 2.24, 2.25 can be rewritten in a more transparent form:

$$
\begin{aligned}
f_i(z_i) &= \left(b_i + \sum_{\substack{k=1\\k\neq i}}^{n} \frac{2a_{ik}\rho_{ik}}{1+\rho_{ik}^2} \right) z_i^2 + a_i z_i - \left(b_i + \sum_{\substack{k=1\\k\neq i}}^{n} \frac{2a_{ik}\rho_{ik}}{1+\rho_{ik}^2} \right) \\
&= (a_i z_i - 0) + \sum_{\substack{k=1\\k\neq i}}^{n} \left[\left(b_i + \frac{2a_{ik}\rho_{ik}}{1+\rho_{ik}^2} \right) z_i^2 - \left(b_i + \frac{2a_{ik}\rho_{ik}}{1+\rho_{ik}^2} \right) \right] \\
&= [a_i z_i - \mathbb{E}(a_i z_i)] + \sum_{\substack{k=1\\k\neq i}}^{n} \left[\left(b_i + \frac{2a_{ik}\rho_{ik}}{1+\rho_{ik}^2} \right) z_i^2 \right. \\
&\qquad \left. - \mathbb{E}\left(\left(b_i + \frac{2a_{ik}\rho_{ik}}{1+\rho_{ik}^2} \right) z_i^2 \right) \right] \qquad (2.27)
\end{aligned}
$$

and

$$
\begin{aligned}
f_{ij}\left(z_i, z_j\right) &= 2a_{ij} z_i z_j - \frac{2a_{ij}\rho_{ij}}{1+\rho_{ij}^2}\left(z_i^2 + z_j^2\right) - 2a_{ij}\rho_{ij} \frac{\rho_{ij}^2 - 1}{1+\rho_{ij}^2} \\
&= \left(2a_{ij} z_i z_j - 2a_{ij}\rho_{ij}\right) - \left(\frac{2a_{ij}\rho_{ij}}{1+\rho_{ij}^2}\left(z_i^2 + z_j^2\right) - \frac{4a_{ij}\rho_{ij}}{1+\rho_{ij}^2} \right) \\
&= \left[2a_{ij} z_i z_j - \mathbb{E}\left(2a_{ij} z_i z_j\right)\right] \\
&\quad - \left[\frac{2a_{ij}\rho_{ij}}{1+\rho_{ij}^2}\left(z_i^2 + z_j^2\right) - \mathbb{E}\left(\frac{2a_{ij}\rho_{ij}}{1+\rho_{ij}^2}\left(z_i^2 + z_j^2\right) \right) \right]. \qquad (2.28)
\end{aligned}
$$

Here the relation

$$
\begin{aligned}
2a_{ij}\rho_{ij} &= \frac{2a_{ij}\rho_{ij}}{1+\rho_{ij}^2}\left(1+\rho_{ij}^2\right) = \frac{2a_{ij}\rho_{ij}}{1+\rho_{ij}^2}\left(2+\rho_{ij}^2 - 1\right) \\
&= \frac{4a_{ij}\rho_{ij}}{1+\rho_{ij}^2} + \frac{2a_{ij}\rho_{ij}}{1+\rho_{ij}^2}\left(\rho_{ij}^2 - 1\right) \qquad (2.29)
\end{aligned}
$$

was used.
Taking the expected values of $f_i(z_i)$ and $f_{ij}(z_i, z_j)$ from Eqs. 2.27, 2.28 yields

$$
\mathbb{E}(f_i(z_i)) = \mathbb{E}(f_{ij}(z_i, z_j)) = 0, \qquad (2.30)
$$

consistent with the expected value of a HDMR component function being zero.
- The resultant terms f_0, $f_i(z_i)$ and $f_{ij}(z_i, z_j)$ satisfy the relaxed vanishing condition and are hierarchically orthogonal. Direct calculation can prove that the sum of f_0 and all $f_i(z_i)$ and all $f_{ij}(z_i, z_j)$ is exactly equal to $f(\mathbf{z})$, i.e.,

f_0, $f_i(z_i)$, $f_{ij}(z_i, z_j)$ is a *repartition* of $f(\mathbf{z})$, and it is a unique decomposition of $f(\mathbf{z})$ under the hierarchical orthogonality condition (see supplemental material). Thus, the quadratic polynomial function, Eq. 2.3, may be rewritten as

$$y = f(\mathbf{z}) \equiv f_0 + \sum_{i=1}^{n} f_i(z_i) + \sum_{1 \le i < j \le n} f_{ij}(z_i, z_j). \tag{2.31}$$

- When all z_i's are independent, i.e., $\rho_{ij} = 0$ for all (i, j), Eqs. 2.16, 2.24, 2.25 reduce to

$$f_0 = a_0 + \sum_{i=1}^{n} b_i, \tag{2.32}$$

$$f_i(z_i) = b_i z_i^2 + a_i z_i - b_i, \tag{2.33}$$

$$f_{ij}(z_i, z_j) = 2a_{ij} z_i z_j. \tag{2.34}$$

2.2 Analytical formulas of SCSA sensitivity indexes

According to SCSA, the structural (independent), correlative and total sensitivity indexes S^a, S^b, S are specified as

$$S_i^a = \mathrm{Var}(f_i(z_i))/\mathrm{Var}(y), \tag{2.35}$$

$$S_i^b = \mathrm{Cov}(f_i(z_i), y - f_i(z_i))/\mathrm{Var}(y), \tag{2.36}$$

$$S_i = \mathrm{Cov}(f_i(z_i), y)/\mathrm{Var}(y), \tag{2.37}$$

$$S_{ij}^a = \mathrm{Var}(f_{ij}(z_i, z_j))/\mathrm{Var}(y), \tag{2.38}$$

$$S_{ij}^b = \mathrm{Cov}(f_{ij}(z_i, z_j), y - f_{ij}(z_i, z_j))/\mathrm{Var}(y), \tag{2.39}$$

$$S_{ij} = \mathrm{Cov}(f_{ij}(z_i, z_j), y)/\mathrm{Var}(y). \tag{2.40}$$

To obtain S_i^a, S_i^b, S_i and S_{ij}^a, S_{ij}^b, S_{ij} we need to determine the variances and covariances.

Let

$$c_i = b_i + \sum_{\substack{k=1 \\ k \ne i}}^{n} \frac{2a_{ik}\rho_{ik}}{1 + \rho_{ik}^2}, \tag{2.41}$$

$$p_{ij} = \frac{2a_{ij}\rho_{ij}}{1 + \rho_{ij}^2}, \tag{2.42}$$

$$q_{ij} = 2a_{ij}\rho_{ij} \frac{\rho_{ij}^2 - 1}{1 + \rho_{ij}^2}. \tag{2.43}$$

Then we have

$$f_i(z_i) = c_i z_i^2 + a_i z_i - c_i, \tag{2.44}$$

$$f_{ij}(z_i, z_j) = 2a_{ij} z_i z_j - p_{ij}\left(z_i^2 + z_j^2\right) - q_{ij}. \tag{2.45}$$

The resultant variances and covariances are given here (see the supplemental material for further details).

$$\mathrm{Var}\,(f_i(z_i)) = a_i^2 + 2c_i^2, \tag{2.46}$$

$$\mathrm{Var}\,\left(f_{ij}(z_i, z_j)\right) = 4a_{ij}^2 - 4p_{ij}^2 + q_{ij}^2, \tag{2.47}$$

$$\begin{aligned}\mathrm{Cov}\,(f_i(z_i), y - f_i(z_i)) = &\sum_{\substack{k=1\\k\neq i}}^{n}\left(a_i a_k \rho_{ik} + 2c_i c_k \rho_{ik}^2\right)\\ &+ \sum_{\substack{k,l=1,k\neq l\\k,l\neq i}}^{n}\left[2a_{kl} c_i \rho_{ik}\rho_{il} - p_{kl} c_i\left(\rho_{ik}^2 + \rho_{il}^2\right)\right],\end{aligned} \tag{2.48}$$

$$\begin{aligned}\mathrm{Cov}\,\left(f_{ij}\left(z_i, z_j\right), y - f_{ij}\left(z_i, z_j\right)\right) = &\sum_{\substack{k=1,\\k\neq i,j}}^{n}\left[4a_{ij} c_k \rho_{ik}\rho_{jk} - 2p_{ij} c_k\left(\rho_{ik}^2 + \rho_{jk}^2\right)\right]\\ &+ \sum_{\substack{k=1\\k\neq i,j}}^{n}\left[4a_{ij}(a_{ik}\rho_{jk} + a_{jk}\rho_{ik}) - 4a_{ij}(p_{ik} + p_{jk})\rho_{ik}\rho_{jk}\right.\\ &-4p_{ij}\rho_{ij}(a_{ik}\rho_{jk} + a_{jk}\rho_{ik}) + 2p_{ij}\left(p_{ik} + p_{jk}\right)\left(\rho_{ij}^2 + \rho_{ik}^2 + \rho_{jk}^2 - 1\right)\\ &\left.+q_{ij}(q_{ik} + q_{jk})\right]\\ &+ \sum_{\substack{k,l=1,k\neq l\\k,l\neq i,j}}^{n}\left[2a_{ij}a_{kl}\left(\rho_{ik}\rho_{jl} + \rho_{il}\rho_{jk}\right) - 2a_{ij}p_{kl}\left(\rho_{ik}\rho_{jk} + \rho_{il}\rho_{jl}\right)\right.\\ &\left.-2a_{kl}p_{ij}\left(\rho_{ik}\rho_{il} + \rho_{jk}\rho_{jl}\right) + p_{ij}p_{kl}\left(\rho_{ik}^2 + \rho_{jk}^2 + \rho_{il}^2 + \rho_{jl}^2\right)\right].\end{aligned} \tag{2.49}$$

To determine $\mathrm{Var}(y)$, we set

$$a_{ii} = b_i, \quad \rho_{ii} = 1. \tag{2.50}$$

$$\begin{aligned}\mathrm{Var}(y) &= \mathrm{Var}\left[a_0 + \sum_{i=1}^{n} a_i z_i + \sum_{\substack{i,j=1\\i\neq j}}^{n} a_{ij} z_i z_j + \sum_{i=1}^{n} b_i z_i^2\right]\\ &= \mathrm{Var}\left[a_0 + \sum_{i=1}^{n} a_i z_i + \sum_{i,j=1}^{n} a_{ij} z_i z_j\right] = \mathrm{Var}\left[\sum_{i=1}^{n} a_i z_i + \sum_{i,j=1}^{n} a_{ij} z_i z_j\right]\end{aligned}$$

$$
\begin{aligned}
&= \mathrm{Var}\Big(\sum_{i=1}^{n} a_i z_i\Big) + \mathrm{Var}\Big(\sum_{i,j=1}^{n} a_{ij} z_i z_j\Big) + 2\mathrm{Cov}\Big(\sum_{i=1}^{n} a_i z_i, \sum_{i,j=1}^{n} a_{ij} z_i z_j\Big) \\
&= \mathrm{Var}\Big(\sum_{i=1}^{n} a_i z_i\Big) + \mathrm{Var}\Big(\sum_{i,j=1}^{n} a_{ij} z_i z_j\Big) = \mathrm{Var}\Big(\sum_{i=1}^{n} a_i z_i\Big) + \mathrm{Var}\Big(\mathbf{z}^T A \mathbf{z}\Big) \\
&= \mathrm{Var}\Big(\sum_{i=1}^{n} a_i z_i\Big) + 2\,\mathrm{tr}\Big(A\Sigma A\Sigma\Big) \\
&= \mathrm{Var}\Big(\sum_{i=1}^{n} a_i z_i\Big) + 2\sum_{i=1}^{n}\sum_{j=1}^{n}\Big(\sum_{k=1}^{n} a_{ik}\rho_{jk}\Big)\Big(\sum_{l=1}^{n} a_{jl}\rho_{il}\Big) \\
&= \sum_{i=1}^{n} a_i^2 + \sum_{\substack{i,j=1\\ i\neq j}}^{n} a_i a_j \rho_{ij} + \sum_{i,j=1}^{n}\sum_{k,l=1}^{n} 2a_{ik}a_{jl}\rho_{jk}\rho_{il}.
\end{aligned} \tag{2.51}
$$

Here three conditions were used: 1) the variance for a quadratic function $\mathbf{z}^T A\mathbf{z}$ [32] is

$$\mathrm{Var}(\mathbf{z}^T A\mathbf{z}) = 2\,\mathrm{tr}(A\Sigma A\Sigma) + 4\mu^T A\Sigma A\mu = 2\,\mathrm{tr}(A\Sigma A\Sigma) \tag{2.52}$$

with $A = (a_{ij})$, 2) Σ is given in Eq. 2.15, and 3) $\mu = \mathbf{0}$.

The analytical formulas for the variances and covariances show that

- $\mathrm{Var}(f_i(z_i))$ and $\mathrm{Var}(f_{ij}(z_i, z_j))$, and consequently S_i^a and S_{ij}^a, are only related to the coefficients of $f_i(z_i)$ and $f_{ij}(z_i, z_j)$. Since the formula of $f(\mathbf{z})$ is the sum of f_0, $f_i(z_i)$ and $f_{ij}(z_i, z_j)$, this implies that S_i^a and S_{ij}^a reflect the structural (independent) contributions of z_i and (z_i, z_j) in $f(\mathbf{z})$, and they are always non-negative.
- $\mathrm{Cov}(f_i(z_i), y - f_i(z_i))$ and $\mathrm{Cov}(f_{ij}(z_i, z_j), y - f_{ij}(z_i, z_j))$, and consequently, S_i^b and S_{ij}^b, are related to the products of coefficients arising from $f_i(z_i)$, $f_{ij}(z_i, z_j)$ and the corresponding correlation coefficients ρ_{ij}'s. Therefore, they reflect the correlative contributions of z_i, (z_i, z_j) with their correlated variables z_k's whose correlation coefficients ρ_{ik}, ρ_{jk} are not zero. They can be positive or negative.
- Since

  $$\mathrm{Cov}(f_i(z_i), y) = \mathrm{Var}(f_i(z_i)) + \mathrm{Cov}(f_i(z_i), y - f_i(z_i))$$

  and

  $$\mathrm{Cov}(f_{ij}(z_i, z_j), y) = \mathrm{Var}(f_{ij}(z_i, z_j)) + \mathrm{Cov}(f_{ij}(z_i, z_j), y - f_{ij}(z_i, z_j)),$$

  we have

  $$\begin{aligned} S_i &= S_i^a + S_i^b, \\ S_{ij} &= S_{ij}^a + S_{ij}^b, \end{aligned} \tag{2.53}$$

which can be positive or negative.

- Due to

$$
\begin{aligned}
&\sum_{i=1}^{n} \mathrm{Cov}(f_i(z_i), y) + \sum_{1\le i<j\le n} \mathrm{Cov}(f_{ij}(z_i, z_j), y) \\
&\quad = \mathrm{Cov}\left(\sum_{i=1}^{n} f_i(z_i) + \sum_{1\le i<j\le n} f_{ij}(z_i, z_j), y\right) \\
&\quad = \mathrm{Cov}\left(f_0 + \sum_{i=1}^{n} f_i(z_i) + \sum_{1\le i<j\le n} f_{ij}(z_i, z_j), y\right) \\
&\quad = \mathrm{Var}(y),
\end{aligned}
\tag{2.54}
$$

we have

$$
\begin{aligned}
1 &\equiv \sum_{i=1}^{n} \frac{\mathrm{Cov}(f_i(z_i), y)}{\mathrm{Var}(y)} + \sum_{1\le i<j\le n} \frac{\mathrm{Cov}(f_{ij}(z_i, z_j), y)}{\mathrm{Var}(y)} \\
&= \sum_{i=1}^{n} S_i + \sum_{1\le i<j\le n} S_{ij}.
\end{aligned}
\tag{2.55}
$$

When the parameters $\alpha_0, \alpha_i, \beta_{ij}, \gamma_i$ and σ_i, ρ_{ij} are determined from experimental data, some errors may occur due to the data error and the finite number of samples. In this case, the above relations can be used to evaluate the reliability of the SCSA.

- When all z_i's are independent, then all $\rho_{ij} = 0 (i \neq j)$. In this case $\mathrm{Cov}(f_i(z_i), y - f_i(z_i))$ and $\mathrm{Cov}(f_{ij}(z_i, z_j), y - f_{ij}(z_i, z_j))$ and consequently, S_i^b and S_{ij}^b all vanish because each term of $\mathrm{Cov}(f_i(z_i), y - f_i(z_i))$ and $\mathrm{Cov}(f_{ij}(z_i, z_j), y - f_{ij}(z_i, z_j))$ contains one or more $\rho_{ij} (i \neq j)$.
- All the statistical parameters μ_i, σ_i and ρ_{ij}, and polynomial model parameters $\alpha_0, \alpha_i, \beta_{ij}, \gamma_i$ can be easily estimated from a given set of input-output $(\mathbf{x}, f(\mathbf{x}))$ data. Using the above formulas, the SCSA sensitivity indexes may be readily obtained.

2.3 Special case: linear polynomials

As a special case of the quadratic polynomial function, the linear polynomial function can be obtained by setting $\beta_{ij} = \gamma_i = 0$ in Eq. 2.1

$$
y = f(\mathbf{x}) = \alpha_0 + \sum_{i=1}^{n} \alpha_i x_i. \tag{2.56}
$$

With new variables z_i, Eq. 2.56 becomes

$$y = f(\mathbf{z}) = a_0 + \sum_{i=1}^{n} a_i z_i \tag{2.57}$$

where

$$a_0 = \alpha_0 + \sum_{i=1}^{n} \alpha_i \mu_i, \tag{2.58}$$

$$a_i = \alpha_i \sigma_i, \quad \forall i. \tag{2.59}$$

Since there are only zeroth and first order HDMR component functions for a linear polynomial function, its HDMR expansion for $f(\mathbf{z})$ is

$$f(\mathbf{z}) = f_0 + \sum_{i=1}^{n} f_i(z_i) = a_0 + \sum_{i=1}^{n} a_i z_i, \tag{2.60}$$

where

$$f_0 = a_0, \tag{2.61}$$

$$f_i(z_i) = a_i z_i, \quad \forall i. \tag{2.62}$$

obtained from Eqs. 2.16 and 2.24 by setting $a_{ij} = a_{ik} = b_i = 0$.

Similarly, from Eqs. 2.46, 2.48, 2.51, with $c_i = c_k = a_{kl} = p_{kl} = 0$ we have

$$\mathrm{Var}(f_i(z_i)) = a_i^2, \tag{2.63}$$

$$\{\mathrm{Cov}(f_i(z_i), y - f_i(z_i)) = \sum_{\substack{j=1 \\ j \neq i}}^{n} a_i a_j \rho_{ij}. \tag{2.64}$$

$$\mathrm{Cov}(f_i(z_i), y) = a_i^2 + \sum_{\substack{j=1 \\ j \neq i}}^{n} a_i a_j \rho_{ij}, \tag{2.65}$$

$$\mathrm{Var}(y) = \sum_{i=1}^{n} a_i^2 + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} a_i a_j \rho_{ij} = \mathbf{a}^T \Sigma \mathbf{a}. \tag{2.66}$$

Then we obtain

$$S_i^a = a_i^2 / \mathbf{a}^T \Sigma \mathbf{a}, \tag{2.67}$$

$$S_i^b = \sum_{\substack{j=1 \\ j \neq i}}^{n} a_i a_j \rho_{ij} / \mathbf{a}^T \Sigma \mathbf{a}, \tag{2.68}$$

$$S_i = S_i^a + S_i^b. \tag{2.69}$$

Equations 2.67–2.69 show that:

- S_i^a is *only* proportional to a_i^2 associated with z_i in the formula of $f(\mathbf{z})$, Eq. 2.60, i.e., it reflects the *structural contribution* of z_i in $f(\mathbf{z})$ to the variance of y. For the original variable x_i, since $a_i^2 = \alpha_i^2\sigma_i^2$, S_i^a is *only* proportional to the square of coefficient α_i^2 in Eq. 2.56 and the variance of the x_i's marginal distribution. Thus, S_i^a represents the *independent* contribution of x_i to the output y's variation, and is always non-negative.
- S_i^b is proportional to the sum of covariances of z_i with other z_j's, i.e., it reflects the *correlative contribution* of z_i with other correlated z_j's whose $\rho_{ij} \neq 0$. The covariance of z_i and z_j is propositional to $a_i a_j$ and their correlation coefficient ρ_{ij}, and can be positive or negative depending on the sign of a_i, a_j and ρ_{ij}. A similar interpretation may be obtained for the original variables x_i and x_j.
- $S_i = S_i^a + S_i^b$ gives the *total contribution* of z_i to the variance of y, which can be also positive or negative.
- $\sum_i S_i \equiv 1$ and can be used to evaluate the reliability of SCSA.
- For a given set of input-output $(\mathbf{x}, f(\mathbf{x}))$ data, all the statistical parameters μ_i, σ_i and ρ_{ij} can be easily estimated, and the polynomial model parameters α_0, α_i can be obtained by linear regression. Using the formulas above, the SCSA sensitivity indexes may be readily calculated.

3 Illustrative cases

Two examples are used for illustration.

3.1 A linear polynomial

Consider an example [30]

$$y = f(\mathbf{x}) = x_1 + x_2 + x_3 \tag{3.1}$$

with $\mu_i = 0$ for all i and the covariance matrix

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \sigma\rho \\ 0 & \sigma\rho & \sigma^2 \end{bmatrix}. \tag{3.2}$$

Since all $\alpha_i = 1, \sigma_1 = \sigma_2 = 1$ and only $\sigma_3 = \sigma$, to transform x_i to the standard variable z_i we only need to set

$$a_i = \alpha_i = 1, \quad (i = 1, 2) \tag{3.3}$$

$$a_3 = \alpha_3\sigma_3 = \sigma, \tag{3.4}$$

Table 1 The analytical form of sensitivity indexes

Input	SCSA			Variance-based method
	S_i^a	S_i^b	S_i	S_i
x_1	$\frac{1}{2+\sigma^2+2\sigma\rho}$	0	$\frac{1}{2+\sigma^2+2\sigma\rho}$	$\frac{1}{2+\sigma^2+2\sigma\rho}$
x_2	$\frac{1}{2+\sigma^2+2\sigma\rho}$	$\frac{\sigma\rho}{2+\sigma^2+2\sigma\rho}$	$\frac{1+\sigma\rho}{2+\sigma^2+2\sigma\rho}$	$\frac{(1+\sigma\rho)^2}{2+\sigma^2+2\sigma\rho}$
x_3	$\frac{\sigma^2}{2+\sigma^2+2\sigma\rho}$	$\frac{\sigma\rho}{2+\sigma^2+2\sigma\rho}$	$\frac{\sigma^2+\sigma\rho}{2+\sigma^2+2\sigma\rho}$	$\frac{(\sigma+\rho)^2}{2+\sigma^2+2\sigma\rho}$
Sum	$\frac{2+\sigma^2}{2+\sigma^2+2\sigma\rho}$	$\frac{2\sigma\rho}{2+\sigma^2+2\sigma\rho}$	1	$1+\frac{2\sigma\rho+\rho^2+\sigma^2\rho^2}{2+\sigma^2+2\sigma\rho}$

and

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \rho \\ 0 & \rho & 1 \end{bmatrix}. \tag{3.5}$$

Then using Eqs. 2.67–2.69 all S_i^a, S_i^b, S_i can be obtained. The resultant sensitivity indexes are given in Table 1. For comparison, the results obtained by variance-based method [30,31] are also listed. Note that

$$\mathrm{Var}(y) = \mathbf{a}^T \Sigma \mathbf{a} = 2 + \sigma^2 + 2\sigma\rho. \tag{3.6}$$

Comparing the results given by SCSA and the variance-based method we may draw the following conclusions

- SCSA separates the structural (independent) and correlative contributions of inputs and provides a clear understanding for the various contributions of the input variance to the output variance, and the resultant sensitivity indexes are easy to interpret.
- The variance-based method mixes up the structural (independent) and correlative contributions of the input variables. Thus, from the resultant variance-based sensitivity indexes it is difficult to determine whether the influence comes from the particular variable or correlation with other input variables. This situation makes it difficult to determine the importance order of the input variables upon comparison of the sensitivity index magnitudes.
- The sum of S_i given by the SCSA method is exactly equal to 1.0, but the sum of S_i obtained from the variance-based method can be larger or smaller than 1.0 which makes it difficult to judge the reliability of the sensitivity analysis.

3.2 A quadratic polynomial

A three variable quadratic polynomial is used as an example for illustration:

$$f(\mathbf{x}) = g_1(x_1, x_2) + g_2(x_2) + g_3(x_3), \tag{3.7}$$

where

$$\begin{aligned} g_1(x_1, x_2) &= g_{1a}(x_1) g_{1b}(x_2) \\ &= [a_1(x_1 - \mu_1) + a_0][b_1(x_2 - \mu_2) + b_0], \quad (3.8) \\ g_2(x_2) &= c_2(x_2 - \mu_2)^2 + c_1(x_2 - \mu_2) + c_0, \quad (3.9) \\ g_3(x_3) &= d_2(x_3 - \mu_3)^2 + d_1(x_3 - \mu_3) + d_0 \quad (3.10) \end{aligned}$$

with a multivariate normal distribution

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{3/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)\right), \quad (3.11)$$

where $\mu = (\mu_1, \mu_2, \mu_3)$ is the expected value of $\mathbf{x}$, Σ is the covariance matrix of $\mathbf{x}$

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & 0 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix}, \quad (3.12)$$

i.e., x_1 and x_2 are correlated, but x_3 is independent. After expansion, Eq. 3.7 becomes

$$\begin{aligned} f(\mathbf{x}) = {} & a_0 b_0 + c_0 + d_0 + a_1 b_0 (x_1 - \mu_1) + (a_0 b_1 + c_1)(x_2 - \mu_2) \\ & + d_1(x_3 - \mu_3) + a_1 b_1 (x_1 - \mu_1)(x_2 - \mu_2) + c_2(x_2 - \mu_2)^2 \\ & + d_2(x_3 - \mu_3)^2. \quad (3.13) \end{aligned}$$

3.3 Formulas for the HDMR component functions

First, the transformation

$$z_i = \frac{x_i - \mu_i}{\sigma_i} \quad (3.14)$$

is performed which gives

$$f(\mathbf{z}) = \bar{a}_0 + \sum_{i=1}^{3} \bar{a}_i z_i + 2\bar{a}_{12} z_1 z_2 + \sum_{i=2}^{3} \bar{b}_i z_i^2, \quad (3.15)$$

where

$$\begin{aligned} \bar{a}_0 &= a_0 b_0 + c_0 + d_0, \quad (3.16) \\ \bar{a}_1 &= a_1 b_0 \sigma_1, \quad (3.17) \\ \bar{a}_2 &= (a_0 b_1 + c_1)\sigma_2, \quad (3.18) \\ \bar{a}_3 &= d_1 \sigma_3, \quad (3.19) \\ \bar{a}_{12} &= a_1 b_1 \sigma_1 \sigma_2 / 2, \quad (3.20) \\ \bar{b}_2 &= c_2 \sigma_2^2, \quad (3.21) \end{aligned}$$

$$\bar{b}_3 = d_2\sigma_3^2. \tag{3.22}$$

Using Eqs. 2.16, 2.24–2.25, the HDMR component functions can be readily written out for $f(\mathbf{z})$ and then substituting the original parameters a_i, b_i, a_{ij} and variables x_i gives

$$\begin{aligned}
f_0 &= \bar{a}_0 + 2\bar{a}_{12}\rho_{12} + \bar{b}_2 + \bar{b}_3 \\
&= a_1 b_1 \rho_{12}\sigma_1\sigma_2 + a_0 b_0 + c_2\sigma_2^2 + c_0 + d_2\sigma_3^2 + d_0, && (3.23)\\
f_1(z_1) &= \frac{2\bar{a}_{12}\rho_{12}}{1+\rho_{12}^2} z_1^2 + \bar{a}_1 z_1 - \frac{2\bar{a}_{12}\rho_{12}}{1+\rho_{12}^2} \\
&= a_1 b_1 \frac{\sigma_2}{\sigma_1}\frac{\rho_{12}}{\rho_{12}^2+1}(x_1-\mu_1)^2 + a_1 b_0 (x_1-\mu_1) \\
&\quad - a_1 b_1 \sigma_1\sigma_2 \frac{\rho_{12}}{\rho_{12}^2+1}, && (3.24)\\
f_2(z_2) &= \left(\bar{b}_2 + \frac{2\bar{a}_{12}\rho_{12}}{1+\rho_{12}^2}\right) z_2^2 + \bar{a}_2 z_2 - (\bar{b}_2 + \frac{2\bar{a}_{12}\rho_{12}}{1+\rho_{12}^2}) \\
&= \left[a_1 b_1 \frac{\sigma_1}{\sigma_2}\frac{\rho_{12}}{\rho_{12}^2+1} + c_2\right](x_2-\mu_2)^2 \\
&\quad + (a_0 b_1 + c_1)(x_2-\mu_2) - a_1 b_1 \sigma_1\sigma_2 \frac{\rho_{12}}{\rho_{12}^2+1} - c_2\sigma_2^2, && (3.25)\\
f_3(z_3) &= \bar{b}_3 z_3^2 + \bar{a}_3 z_3 - \bar{b}_3 \\
&= d_2(x_3-\mu_3)^2 + d_1(x_3-\mu_3) - d_2\sigma_3^2, && (3.26)\\
f_{12}(z_1, z_2) &= 2\bar{a}_{12} z_1 z_2 - \frac{2\bar{a}_{12}\rho_{12}}{1+\rho_{12}^2}(z_1^2 + z_2^2) - 2\bar{a}_{12}\rho_{12}\frac{\rho_{12}^2-1}{1+\rho_{12}^2} \\
&= -a_1 b_1 \frac{\sigma_2}{\sigma_1}\frac{\rho_{12}}{\rho_{12}^2+1}(x_1-\mu_1)^2 + a_1 b_1 (x_1-\mu_1)(x_2-\mu_2) \\
&\quad - a_1 b_1 \frac{\sigma_1}{\sigma_2}\frac{\rho_{12}}{\rho_{12}^2+1}(x_2-\mu_2)^2 - a_1 b_1 \rho_{12}\sigma_1\sigma_2 \frac{\rho_{12}^2-1}{\rho_{12}^2+1}, && (3.27)\\
f_{13}(z_1, z_3) &= 0, && (3.28)\\
f_{23}(z_2, z_3) &= 0. && (3.29)
\end{aligned}$$

3.4 Formulas for the SCSA sensitivity indexes

Using Eqs. 2.46–2.51, the variances and covariances of the HDMR component functions for $f(\mathbf{z})$ given in Eq. 3.15 were obtained below.

$$\mathrm{Var}(f_1(z_1)) = \bar{a}_1^2 + 2\left(\frac{2\bar{a}_{12}\rho_{12}}{1+\rho_{12}^2}\right)^2, \tag{3.30}$$

Table 2 SCSA sensitivity indexes

	$\rho_{12} = 0.6$			$\rho_{12} = -0.6$		
	S^a	S^b	S	S^a	S^b	S
$x_1(z_1)$	0.2851	0.1714	0.4565	0.9531	−0.5496	0.4034
$x_2(z_2)$	0.3052	0.1714	0.4766	0.9227	−0.5496	0.3731
$x_3(z_3)$	0.0594	0.0000	0.0594	0.1985	0.0000	0.1985
$x_1, x_2(z_1, z_2)$	0.0075	0.0000	0.0075	0.0250	0.0000	0.0250
Sum	0.6572	0.3428	1.0000	2.0993	−1.0993	1.0000

$$\mathrm{Var}(f_2(z_2)) = \bar{a}_2^2 + 2\left(\bar{b}_2 + \frac{2\bar{a}_{12}\rho_{12}}{1+\rho_{12}^2}\right)^2, \tag{3.31}$$

$$\mathrm{Var}(f_3(z_3)) = \bar{a}_3^2 + 2\bar{b}_3^2, \tag{3.32}$$

$$\mathrm{Var}(f_{12}(z_1, z_2)) = 4\bar{a}_{12}^2 - 4\left(\frac{2\bar{a}_{12}\rho_{12}}{1+\rho_{12}^2}\right)^2 + \left(\frac{2\bar{a}_{12}\rho_{12}(\rho_{12}^2-1)}{1+\rho_{12}^2}\right)^2, \tag{3.33}$$

$$\mathrm{Cov}(f_1(z_1), y - f_1(z_1)) = \bar{a}_1\bar{a}_2\rho_{12} + 2\left(\frac{2\bar{a}_{12}\rho_{12}}{1+\rho_{12}^2}\right)\left(\bar{b}_2 + \frac{2\bar{a}_{12}\rho_{12}}{1+\rho_{12}^2}\right)\rho_{12}^2, \tag{3.34}$$

$$\mathrm{Cov}(f_2(z_2), y - f_2(z_2)) = \bar{a}_1\bar{a}_2\rho_{12} + 2\left(\frac{2\bar{a}_{12}\rho_{12}}{1+\rho_{12}^2}\right)\left(\bar{b}_2 + \frac{2\bar{a}_{12}\rho_{12}}{1+\rho_{12}^2}\right)\rho_{12}^2, \tag{3.35}$$

$$\mathrm{Cov}\,(f_{12}(z_1, z_2), y - f_{12}(z_1, z_2)) = 0, \tag{3.36}$$

$$\mathrm{Var}(y) = \sum_{i=1}^{3} \bar{a}_i^2 + 2\left(\bar{b}_2^2 + \bar{b}_3^2\right) + 2\bar{a}_1\bar{a}_2\rho_{12} + 4\bar{a}_{12}^2\left(1+\rho_{12}^2\right) + 8\bar{b}_2\bar{a}_{12}\rho_{12}. \tag{3.37}$$

The sensitivity indexes can be obtained from the ratio of variances and covariances to the variance of y.

Setting $\sigma_1 = \sigma_2 = 0.2$, $\sigma_3 = 0.18$ and $\rho_{12} = 0.6(-0.6)$, and $a_0 = 1, a_1 = 2, b_0 = 2, b_1 = 3, c_0 = 3, c_1 = 1, c_2 = 2, d_0 = 1, d_1 = 2, d_2 = 2$, the resultant sensitivity indexes are given in Table 2.

The sensitivity indexes given in Table 2 show that (1) S_i^a, S_{ij}^a are always non-negative; (2) S_i^b, S_{ij}^b can be positive or negative depending on the sign of a_i, b_i, a_{ij} and ρ_{ij}; (3) x_3 is independent, so $S_3^b = 0$; (4) $f_{12}(x_1, x_2)$ is hierarchically orthogonal to $f_1(x_1)$, $f_2(x_2)$, and orthogonal to $f_3(x_3)(x_3$ is independent), thus $f_{12}(x_1, x_2)$ is orthogonal to all non-zero component functions, which makes $S_{12}^b = 0$; (5) the sum of all S_i, S_{ij} is exactly equal to unit.

4 Conclusions

A general formulation for the HDMR component functions with independent and correlated variables has been obtained [23]. Global sensitivity analysis, SCSA, based on the general formulas of HDMR component functions has also been established. In practice, many systems are exactly described or satisfactorily approximated by quadratic polynomial functions, and the probability distribution of the variables are commonly expressed as a multivariate normal distribution. This paper presented the analytical formulas for the HDMR component functions and the corresponding SCSA sensitivity indexes in the latter circumstances. These results should be valuable for many practical applications of HDMR with correlated variables.

Acknowledgments Support for this work was provided by ONR with account number N00014-11-1-0716.

References

1. R. Fisher, *Statistical Methods for Research Workers* (Oliver and Boyd, London, 1925)
2. W. Hoeffding, Ann. Math. Stat. **19**(3), 293–325 (1948)
3. I.M. Sobol, Nauka. (Moscow 1969) (in Russian)
4. I.M. Sobol, Math. Model. **2**, 112–118 (1990) in Russian; translated in Mathematical Modelling and Computational Experiments 1, 407–414 (1993)
5. A. Saltelli, I.M. Sobol, Reliab. Eng. Syst. Saf. **50**, 225–239 (1995)
6. A. Saltelli, M. Ratto, S. Tarantola, F. Campolongo, Chem. Rev. **105**, 2811–2827 (2005)
7. M. Ratto, A. Pagano, P. Young, Comput. Phys. Commun. **177**, 863–876 (2007)
8. T. Homma, A. Saltelli, Reliab. Eng. Syst. Saf. **52**, 1–17 (1996)
9. A. Saltelli, K. Chan, E.M. Scott (eds.), *Sensitivity Analysis* (Wiley, New York, 2000)
10. A. Saltelli, S. Tarantola, J. Am. Stat. Assoc. **97**, 702–709 (2002)
11. A. Saltelli, S. Tarantola, P.S. Chan, Technometrics **41**, 39–56 (1999)
12. M.D. McKay, R.J. Beckman, W.J. Conover, Technometrics **21**, 239–245 (1979)
13. I.M. Sobol, Math. Comput. Simul. **55**, 271–280 (2001)
14. H. Rabitz, Ö.F. Alis, J. Math. Chem. **25**, 197–233 (1999)
15. G. Li, S.W. Wang, C. Rosenthal, H. Rabitz, J. Math. Chem. **30**, 1–30 (2001)
16. G. Li, H. Rabitz, J. Schoendorf, J. Comput. Chem. **25**, 1149–1156 (2004)
17. I.M. Sobol, Reliab. Eng. Syst. Saf. **79**, 187–193 (2003)
18. O.F. Alis, H. Rabitz, J. Math. Chem. **29**, 127–142 (2001)
19. G. Li, S.-W. Wang, H. Rabitz, J. Phys. Chem. A **106**, 8721–8733 (2002)
20. G. Li, S.W. Wang, H. Rabitz, S. Wang, P.R. Jaffé, Chem. Eng. Sci. **57**, 4445–4460 (2002)
21. S.-W. Wang, P.G. Georgopoulos, G. Li, H. Rabitz, J. Phys. Chem. A **107**, 4707–4716 (2003)
22. G. Hooker, J. Comput. Graph. Stat. **16**(3), 709–732 (2007)
23. G. Li, H. Rabitz, J. Math. Chem. **50**, 99–130 (2012)
24. S. Rahman, SIAM/ASA J. Uncertain. Quantif. (2014) (submitted)
25. G. Li, H. Rabitz, J. Math. Chem. **48**(4), 1010–1035 (2010)
26. G. Li, R. Rey-de-Castro, H. Rabitz, J. Math. Chem. **50**, 1747–1764 (2012)
27. G. Li et al., J. Phys. Chem. A **114**, 6022–6032 (2010)
28. C. Xu, G.Z. Gertner, Reliab. Eng. Syst. Saf. **93**, 1563–1573 (2008)
29. C.B. Storlie et al., Reliab. Eng. Syst. Saf. **94**, 1735–1763 (2009)
30. S. Da Veiga, F. Whal, F. Gamboa, Technometrics **51**, 452–463 (2009)
31. S. Kucherenko, S. Tarantola, P. Annoni, Comput. Phys. Commun. **183**, 937–946 (2012)
32. L.R. Schaeffer, *Estimation of Variance Components in Animal Breeding* (Lecture Note, Iowa State University, Iowa, July, 2004), pp. 19–23