Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-259, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 5 July 2016

Hydrology and Earth System Sciences Discussions
Open Access

The new importance measures based on vector projection for multivariate output: application on hydrological model

Xu Liyang

School of Aeronautics, Northwestern Polytechnical University, Xi'an, 710072, P. R. China

Lu Zhenzhou

Corresponding author

School of Aeronautics, Northwestern Polytechnical University, Xi'an, 710072, P. R. China

e-mail: zhenzhoulu@nwpu.edu.cn

Tel: 0086-29-88460480

Xiao Sinan

School of Aeronautics, Northwestern Polytechnical University, Xi'an, 710072, P. R. China

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-259, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 5 July 2016

Hydrology and Earth System Sciences Discussions
Open Access

Abstract:

Analyzing the effects of the inputs on the correlated multivariate output is important to assess risk and make decisions in Hydrological processes. However, the existing methods, such as output decomposition approach and covariance decomposition approach, can't provide sufficient information of the effects of the inputs on the multivariate output, since these methods only measure the influence of input variables on the magnitudes of variances of the dimensionalities in the multiple output space and ignore the effects on the dimensionality directions of output variances. In this paper, a new kind of sensitivity indices based on vector projection for the multivariate output is proposed. By the projection of the conditional vectors on the unconditional vector in the dimensionless multiple output space, the new sensitivity indices measure the influence of the input variables on the magnitudes of variances and directions of the dimensionalities simultaneously. The mathematical properties of the proposed index are discussed, and its link with the Sobol indices is derived. And Polynomial Chaos Expansion (PCE) is used to estimate the proposed sensitivity indices. The results for two numerical examples and a hydrological model indicate the validity and potential benefits of the vector projection index and the efficiency of estimation approach.

Keywords:

Sensitivity analysis; Multivariate output; Vector projection; Dimension; HBV;

1. Introduction

Models with multivariate output are widely used in the field of engineer and science, and the multivariate output is correlated in some degree. For example, output of multiple elicitation surveys are applied to the cost of key low-carbon energy technology (Bosetti, Marangoni et al. 2015), and many dynamic models used to study risk assessment and decision support in ecology and crop science generate time-dependent model predictions, with time being either discretized in a finite number of time steps or considered as continuous(Lamboni, Monod et al. 2011). Traditional methods for Global Sensitivity Analysis (GSA), including the elementary effect method (Campolongo, Cariboni et al. 2007, Campolongo, Saltelli et al. 2011), variance based method (Homma and Saltelli 1996, Sobol' 2001), derivative based method (Sobol' and Kucherenko 2009, Sobol' and Kucherenko 2010) and moment dependent method (Borgonovo 2007, Cui, Lü et al. 2010, Luyi, Zhenzhou et al. 2012), were designed for scalar output. And the direct way to perform sensitivity analysis for models with multivariate output is to perform sensitivity analysis for each

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-259, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 5 July 2016

output separately. However, this way is just a repetition of the traditional GSA and it ignores the correlations among the multivariate output. Thus, it may be insufficient to perform sensitivity analysis on each output separately or on a few context specific scalar functions of the output (Lamboni, Monod et al. 2011). A high degree of redundant sensitivity indices can be obtained when the correlation in the outputs is strong. In such a case, it is difficult to interpret the result (Garcia-Cabrejo and Valocchi 2014). Saltelli and Tarantola proposed to define a scalar index of interest to apply the GSA to simplify the original problem(Saltelli, Tarantola et al. 2000). It is recommended to apply sensitivity analysis to the multivariate output as a whole, and criteria and methods need to be developed for the sensitivity analysis of multivariate output.

Campbell (Campbell, McKay et al. 2006) proposed the output decomposition method for sensitivity analysis, which consists in (i) performing an orthogonal decomposition of the multivariate output, and (ii) applying sensitivity analysis on most informative components separately. This method gives more attention to a few components rather than the whole output. To summarize the sensitivity over the whole output, Lamboni (Lamboni, Monod et al. 2011) proposed a new synthetic sensitivity criterion and extended the criterion to the continuous case. Generalized Sobol' sensitivity indices for multivariate output based on the decomposition of covariance matrix of model outputs was defined by Gamboa et al (Gamboa, Janon et al. 2013), and it is more computational efficient since it doesn't need spectral decomposition compared to the output decomposition method (Lamboni, Monod et al. 2011).

These sensitivity analysis methods for multivariate output only considered the sum of variance of each output, which implicitly assumes that the relationship between outputs is simple and additive. However, there are different dimensions of measurement and orders of magnitude among outputs, which make them not be directly used for the comprehensive analysis. Therefore, it is necessary to have a dimensionless process for the outputs before the comprehensive analysis. Besides, for multivariate output space, each output represents one dimensionality of the multivariate output space. The variance of each output can represent the uncertainty of each dimensionality, which can be regarded as the magnitude of each variance dimensionality. The covariance decomposition method compares the importance of the model inputs by the influence of the inputs on the variance of each output. For the output decomposition method, the original outputs are transformed into a new set of outputs, which form a transformed space. Then, the influence of the model inputs on the variance of the new outputs tells the importance of each input. The sensitivity methods above can tell the influence of the model inputs on the variance of model outputs, which can be regarded as an influence on the magnitude of all the variance dimensionalities. However, they can't tell the influence of the model inputs on the directions of all the variance dimensionalities, i.e., the direction of the variance vector of the output space, which can reflect another character of the multivariate output uncertainty space. Thus, these methods are not

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-259, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 5 July 2016

Hydrology and Earth System Sciences
Open Access
Discussions

sufficient to tell the importance of model inputs.

In this work, we introduce a new sensitivity index based on vector projection which contains the influence of the input uncertainties on the magnitudes and directions simultaneously of all the multivariate output variance vector. Through dimensionless process, the influence of dimension of outputs can be eliminated. Then the conditional variances and unconditional variances of each output are set as vectors, the influence of the input uncertainties can be reflected by the similarity between the conditional variances vector and the unconditional variances vector, which can be measured by the vector projection.

The rest of this paper is organized as follows. The next section briefly reviews the global sensitivity indices based on the variance for scalar output and multiple outputs, then the definition and properties of the vector projection method. In Section 3, Polynomial Chaos Expansion (PCE) is applied to estimate the new sensitivity index. In Section 4, the new sensitivity index is illustrated by two numerical examples and HBV model, which gives the hydrological forecasts and predicts the potential climate changes or floods. Conclusions come at the end of paper.

2. Methodology

2.1 The traditional importance measures

2.1.1 Variance based method

The importance measure (IM) is defined as "the study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input" (Saltelli, Tarantola et al. 2004), and Sobol decomposition is one of the main methodologies for IM.

Let $\mathbf{Y}=(Y^{(1)},...,Y^{(m)})$ denote the m-dimensional model output vector, where $Y^{(r)}=g^{(r)}(\boldsymbol{X})$ $(r=1,...,m)$, and $\boldsymbol{X}=(X_1,X_2,...,X_n)^T$ is the vector of n-dimensional independent input variables. In the case of scalar output $r=1$, the Sobol decomposition of the function $Y^{(1)}=g^{(1)}(X_1,X_2,...X_n)$ is given by

$$Y^{(1)}=g_0^{(1)}+\sum_{i=1}^{n}g_i^{(1)}+\sum_{i_1=1}^{n}\sum_{i_2=1+i_1}^{n}g_{i_1,i_2}^{(1)}(X_{i_1},X_{i_2})+\cdots+g_{1,2,\cdots,n}^{(1)}(X_1,\ldots,X_n) \quad (1)$$

where $g_0^{(1)}=E(Y^{(1)})$, $g_i^{(1)}=E(Y^{(1)}\mid X_i)-g_0^{(1)}$ and $g_{i_1,i_2}^{(1)}=E(Y^{(1)}\mid X_{i_1},X_{i_2})-g_{i_1}^{(1)}-g_{i_2}^{(1)}-g_0^{(1)}$. It is shown by Sobol [2] that $g_i^{(1)}$ is the variation of $Y^{(1)}$ due to change of X_i only when the mean $g_0^{(1)}$ has been considered, and similarly $g_{i_1,i_2}^{(1)}$ is the variation of $Y^{(1)}$ when X_{i_1} and X_{i_2} are interacting. All the terms in Eq.(1) are orthogonal. Taking variances to both sides of Eq.(1),the following Eq.(2) can be obtained:

$$V^{Y^{(1)}}=\sum_{i=1}^{n}V_{X_i}^{Y^{(1)}}+\sum_{i_1=1}^{n}\sum_{i_2=1+i_1}^{n}V_{X_{i_1}X_{i_2}}^{Y^{(1)}}+\cdots+V_{X_1X_2\ldots X_n}^{Y^{(1)}} \quad (2)$$

where $V^{Y^{(1)}}=\mathrm{var}(Y^{(1)})=\int(g^{(1)}(\mathbf{x}))^2 f_{\mathbf{x}}(\mathbf{x})d\mathbf{x}-(g_0^{(1)})^2$, $V_{X_i}^{Y^{(1)}}=V(g_i^{(1)})=V(E(Y^{(1)}\mid X_i))$ and

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-259, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 5 July 2016

Hydrology and Earth System Sciences
Open Access
Discussions

$V_{X_{i_1}X_{i_2}}^{Y^{(1)}} = V(g_{i_1,i_2}^{(1)}) = V(E(Y^{(1)} \mid X_{i_1}, X_{i_2})) - V_{X_{i_1}}^{Y^{(1)}} - V_{X_{i_2}}^{Y^{(1)}}$. The first order partial variance $V_{X_i}^{Y^{(1)}}$ can be explained as the average reduction of model output variance resulting from fixing X_i, which measures the individual contribution of X_i to the total variance $V^{Y^{(1)}}$. The second order partial variance $V_{X_{i_1}X_{i_2}}^{Y^{(1)}}$ represents the interaction effect between X_{i_1} and X_{i_2}. In the same way explanation can be given to the higher order partial variances. And the total partial variance $V_{T_i}^{Y^{(1)}}$ contributed by X_i is defined as the summation of all terms in Eq. (2) with subscripts including i :

$$V_{T_i}^{Y^{(1)}} = V_{X_i}^{Y^{(1)}} + \sum_{i_1 \neq i}^{n} V_{X_i X_{i_1}}^{Y^{(1)}} + \cdots + V_{X_1 \ldots X_i \ldots X_n}^{Y^{(1)}} \tag{3}$$

So $V_{T_i}^{Y^{(1)}}$ consists of the individual effect of X_i and its interaction effects with all the other $n-1$ input variables $\mathbf{X}_{\sim i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$. The measure $V_{T_i}^{Y^{(1)}} = V^{Y^{(1)}} - V(E(Y^{(1)} \mid \mathbf{X}_{\sim i}))$ is the average residual variance of the model output when all the inputs but X_i are fixed over their full ranges.

Dividing $V_{X_i}^{Y^{(1)}}$ and $V_{T_i}^{Y^{(1)}}$ by the total variance $V^{Y^{(1)}}$, the main effect index S_i and total effect index S_{Ti} for X_i on the first output can be obtained as follows.

$$S_i^{Y^{(1)}} = \frac{V_{X_i}^{Y^{(1)}}}{V^{Y^{(1)}}} \tag{4}$$

and

$$S_{T_i}^{Y^{(1)}} = \frac{V_{T_i}^{Y^{(1)}}}{V^{Y^{(1)}}} = \frac{V^{Y^{(1)}} - V(E(Y^{(1)} \mid \boldsymbol{X}_{\sim i}))}{V^{Y^{(1)}}} \tag{5}$$

2.1.2 Covariance decomposition approach

In the case of multivariate output $r = m$ $(m > 1)$, taking the covariance matrices for both sides of $\mathbf{Y}$, Gamboa (Garcia-Cabrejo and Valocchi 2014) obtained

$$\begin{aligned} \boldsymbol{C}(Y^{(1)}, \ldots, Y^{(m)}) = & \sum_{i=1}^{n} \boldsymbol{C}_i(Y^{(1)}, \ldots, Y^{(m)}) + \sum_{1 \le i \le j \le n} \boldsymbol{C}_{i,j}(Y^{(1)}, \ldots, Y^{(m)}) + \sum_{1 \le i < j < k \le n} \boldsymbol{C}_{i,j,k}(Y^{(1)}, \ldots, Y^{(m)}) \\ & + \cdots + \boldsymbol{C}_{1,2,\ldots,n}(Y^{(1)}, \ldots, Y^{(m)}) \end{aligned} \tag{6}$$

The expression implies that the covariance matrix $\boldsymbol{C}$ of the multivariate output can be partitioned into the sum of covariance matrices that comes from changes in single $\boldsymbol{C}_i$, pairs $\boldsymbol{C}_{i,j}$, triples $\boldsymbol{C}_{i,j,k}$ and so on of input variables.

When $m = 1$, Eq.(6) regresses to Eq.(2) the decomposition of the variance for the scalar output Take the trace about the both sides of Eq.(6), Eq.(7) can be obtained,

$$\begin{aligned} Tr[\mathbf{C}(Y^{(1)}, \ldots, Y^{(m)})] = & \sum_{i=1}^{n} Tr[\mathbf{C}_i(Y^{(1)}, \ldots, Y^{(m)})] + \sum_{1 \le i \le j \le n} Tr[\mathbf{C}_{i,j}(Y^{(1)}, \ldots, Y^{(m)})] + \\ & \sum_{1 \le i < j < k \le n} Tr[\mathbf{C}_{i,j,k}(Y^{(1)}, \ldots, Y^{(m)})] + \cdots + Tr[\mathbf{C}_{1,2,\ldots,n}(Y^{(1)}, \ldots, Y^{(m)})] \end{aligned} \tag{7}$$

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-259, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 5 July 2016

According to Eq.(7), the multivariate single effect index $S1_i^M$ of the input variable X_i is given by

$$S1_i^M(Y^{(1)},\ldots,Y^{(m)}) = \frac{Tr[\mathbf{C}_i]}{Tr[\mathbf{C}]} \tag{8}$$

While the multivariate total effect index ST_i^M can be defined as

$$ST_i^M(Y^{(1)},\ldots,Y^{(m)}) = \frac{Tr[\mathbf{C}_i] + \sum_{1\le i\le j\le n} Tr[\mathbf{C}_{i,j}] + \sum_{1\le i<j<k\le n} Tr[\mathbf{C}_{i,j,k}] + \cdots + Tr[\mathbf{C}_{1,2,\ldots,n}]}{Tr[\mathbf{C}]} \tag{9}$$

The trace $Tr[\mathbf{C}]$ is the sum of the variances of all outputs $Y^{(r)}(r=1,\ldots,m)$. The $S1_i^M$ or ST_i^M can be interpreted as the sum of the variances associated with input variable X_i and $\boldsymbol{X}_{\sim i}$ of all the outputs $\boldsymbol{Y}$. Garcia-Cabrejo et al. (Garcia-Cabrejo and Valocchi 2014) pointed out that the output decomposition method and the covariance decomposition method are equivalent if the first K eigenvectors in the principle component decomposition preserve the original variance of outputs. The output and covariance decomposition methods mainly focus on the sum of the variances of the multivariate output. However, the comprehensive effect for the input variable on the multiple output may not be equal to the sum of each input contribution to the scalar output. If the correlation is in the output, the traditional sensitivity measure for the multivariate output is difficult to be interpreted. Furthermore, these methods ignore the influence of the dimension of the output variable. If some outputs have higher order of magnitude than others, they will make larger contribution improperly over the whole outputs (Szirtes and Rózsa 2007). To solve these problems, an alternative measure has been proposed for the importance measure of the multivariate output. It is called the vector projection approach.

2.2 The Definitions and properties of the new importance measure for multivariate output based on the vector projection

2.2.1 Preliminaries

Assume random input variables $\mathbf{X} = (X_1,\ldots,X_n)$ be independent defined on some probability space（Ω,P）, and $Y^{(r)} = g^{(r)}(\mathbf{X})(r=1,\ldots,m)$.

2.2.2 Definition of Vector Projection approach

Transformed by Eq.(10), the output $Y^{(r)}$ can be dimensionless

$$\hat{Y}^{(r)} = \frac{Y^{(r)}}{E(|Y^{(r)}|)} \tag{10}$$

where $E(\bullet)$ is the expectation operator and $|\bullet|$ is the absolute operator.

From Eq.(2), the variance decomposition of $\hat{\mathbf{Y}} = (\hat{Y}^{(1)},\ldots,\hat{Y}^{(m)})$ can be obtained:

$$\mathbf{V}^{\hat{\mathbf{Y}}} = \sum_{i=1}^{n} \mathbf{V}_{X_i}^{\hat{\mathbf{Y}}} + \sum_{i_1=1}^{n} \sum_{i_2=1+i_1}^{n} \mathbf{V}_{X_{i_1}X_{i_2}}^{\hat{\mathbf{Y}}} + \cdots + \mathbf{V}_{X_1X_2\ldots X_n}^{\hat{\mathbf{Y}}} \tag{11}$$

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-259, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 5 July 2016

where m -dimension vectors $\mathbf{V}_{i_1,i_2,\ldots,i_r} = [V_{X_{i_1,i_2,\ldots,i_r}}^{\hat{Y}^{(1)}}, V_{X_{i_1,i_2,\ldots,i_r}}^{\hat{Y}^{(2)}}, \ldots, V_{X_{i_1,i_2,\ldots,i_r}}^{\hat{Y}^{(m)}}]^T, (1 \le i_r (r=1,\ldots,n) \le n)$ and $\mathbf{V}^{\hat{\mathbf{Y}}} = [V^{\hat{Y}^{(1)}}, V^{\hat{Y}^{(2)}}, \ldots, V^{\hat{Y}^{(m)}}]^T$ are the rth,$(r \in (1,2,\ldots,n))$ conditional variances and the unconditional variances of the multivariate output respectively. The above equation can be simplified as following by ignoring the superscript $\hat{\mathbf{Y}}$.

$$\mathbf{V} = \sum_{i=1}^{n} \mathbf{V}_i + \sum_{i_1=1}^{n} \sum_{i_2=1+i_1}^{n} \mathbf{V}_{i_1,i_2} + \cdots + \mathbf{V}_{1,2,\ldots,n} \tag{12}$$

The vector projection can be used to generalize the important measure. According to the definition of inner product (Durier 1994), the vector projection Q_i of the vector $\mathbf{V}_i$ on the vector $\mathbf{V}$ can be given by

$$Q_i = \|\mathbf{V}_i\| \cos \boldsymbol{\theta}_i = \frac{<\mathbf{V}_i, \mathbf{V}>}{\|\mathbf{V}\|} = \frac{\sum_{k=1}^{m} V_{X_i}^{\hat{Y}^{(k)}} V^{\hat{Y}^{(k)}}}{\sqrt{\sum_{k=1}^{m} (V^{\hat{Y}^{(k)}})^2}} \tag{13}$$

where $\boldsymbol{\theta}_i$ is the angle from the vector $\mathbf{V}$ to the vector $\mathbf{V}_i$, $\langle \cdot, \cdot \rangle$ represents the inner product of two vectors, and $\|\cdot\|$ represents the magnitude of a vector. Then, normalize the projection by dividing the norm of vector $\mathbf{V}$ and the new main effect index P_i is defined as following:

$$P_i = \frac{Q_i}{\|\mathbf{V}\|} = \frac{\sum_{k=1}^{m} V_{X_i}^{\hat{Y}^{(k)}} V^{\hat{Y}^{(k)}}}{\sum_{k=1}^{m} (V^{\hat{Y}^{(k)}})^2} \tag{14}$$

Similarly, the vector projection $Q_{i_1,i_2,\ldots,i_r}$ $((r \in 1,\ldots,n), 1 \le i_r \le n)$ of the interaction effect is given by:

$$Q_{i_1,i_2,\ldots,i_r} = \|\mathbf{V}_{i_1,i_2,\ldots,i_r}\| \cos \boldsymbol{\theta}_{i_1,i_2,\ldots,i_r} = \frac{<\mathbf{V}_{i_1,i_2,\ldots,i_r}, \mathbf{V}>}{\|\mathbf{V}\|} = \frac{\sum_{k=1}^{m} V_{X_{i_1} X_{i_2} \ldots X_{i_r}}^{\hat{Y}^{(k)}} V^{\hat{Y}^{(k)}}}{\sqrt{\sum_{k=1}^{m} (V^{\hat{Y}^{(k)}})^2}} \tag{15}$$

where $\boldsymbol{\theta}_{i_1,i_2,\ldots,i_r}$ is the angle from vector $\mathbf{V}$ to vector $\mathbf{V}_{i_1,i_2,\ldots,i_r}$. And the interaction effect index is:

$$P_{i_1,i_2,\ldots,i_r} = \frac{Q_{i_1,i_2,\ldots,i_r}}{\|\mathbf{V}\|} = \frac{\sum_{k=1}^{m} V_{X_{i_1} X_{i_2} \ldots X_{i_r}}^{\hat{Y}^{(k)}} V^{\hat{Y}^{(k)}}}{\sum_{k=1}^{m} (V^{\hat{Y}^{(k)}})^2} \tag{16}$$

Like Eq.(3) the total effect can be expressed as Eq.(17):

$$\mathbf{V}_{T_i} = \mathbf{V}_i + \sum_{i_1 \ne i}^{n} \mathbf{V}_{i,i_1} + \cdots + \mathbf{V}_{1,\ldots,i,\ldots n} \tag{17}$$

$\mathbf{V}_{T_i}$ includes the individual effect of X_i and its interaction effects with all the other $n-1$ input variables $\boldsymbol{X}_{\sim i}$. Let $\mathbf{V}_{\sim i} = [V_{\mathbf{X}_{\sim i}}^{Y^{(1)}}, V_{\mathbf{X}_{\sim i}}^{Y^{(2)}}, \ldots, V_{\mathbf{X}_{\sim i}}^{Y^{(m)}}]^T$ and the vector $\mathbf{V}_{T_i} = \mathbf{V} - \mathbf{V}_{\sim i}$ contains the average

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-259, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 5 July 2016

Hydrology and Earth System Sciences Open Access Discussions

remaining variances of all model outputs when all the inputs but X_i are fixed over their full ranges.

The vector projection of the total effect Q_{T_i} can be expressed as:

$$Q_{T_i} = \frac{\mathbf{V}_{T_i}{}^T \cdot \mathbf{V}}{\|\mathbf{V}\|} = \|\mathbf{V}\| - \frac{<\mathbf{V}_{\sim i}, \mathbf{V}>}{\|\mathbf{V}\|} = \sqrt{\sum_{k=1}^{m}(V^{\hat{Y}^{(k)}})^2} - \frac{\sum_{k=1}^{m} V_{\mathbf{X}_{\sim i}}^{\hat{Y}^{(k)}} V^{\hat{Y}^{(k)}}}{\sqrt{\sum_{k=1}^{m}(V^{\hat{Y}^{(k)}})^2}} \tag{18}$$

And the total effect index is:

$$P_{T_i} = \frac{Q_{T_i}}{\|\mathbf{V}_n\|} = 1 - \frac{\sum_{k=1}^{m} V_{\mathbf{X}_{\sim i}}^{\hat{Y}^{(k)}} V^{\hat{Y}^{(k)}}}{\sum_{k=1}^{m}(V^{\hat{Y}^{(k)}})^2} \tag{19}$$

Lemma 2.2.1 The vector projection measures sum up to 1, i.e,

$$\sum_{i=1}^{n} P_i + \sum_{i_1=1}^{n} \sum_{i_2=1+i_1}^{n} P_{i_1,i_2} + \cdots + P_{1,2,\ldots,n} = 1 \tag{20}$$

Proof.

$$\begin{aligned}
\sum_{i=1}^{n} P_i + \sum_{i_1=1}^{n} \sum_{i_2=1+i_1}^{n} P_{i_1,i_2} + \cdots + P_{1,2,\ldots,n} &= \frac{\sum_{i=1}^{n} Q_i + \sum_{i_1=1}^{n} \sum_{i_2=1+i_1}^{n} Q_{i_1,i_2} + \cdots + Q_{1,2,\ldots,n}}{\|\mathbf{V}\|} \\
&= \frac{\sum_{i=1}^{n} \|\mathbf{V}_i\| \cos\boldsymbol{\theta}_i + \sum_{i_1=1}^{n} \sum_{i_2=1+i_1}^{n} \|\mathbf{V}_{i_1,i_2}\| \cos\boldsymbol{\theta}_{i_1,i_2} + \cdots + \|\mathbf{V}_{1,2,\ldots,n}\| \cos\boldsymbol{\theta}_{1,2,\ldots,n}}{\|\mathbf{V}\|} \\
&= \frac{\sum_{i=1}^{n} \sum_{k=1}^{m} V_{X_i}^{\hat{Y}^{(k)}} V^{\hat{Y}^{(k)}} + \sum_{i_1=1}^{n} \sum_{i_2=1+i_1}^{n} \sum_{k=1}^{m} V_{X_{i_1}X_{i_2}}^{\hat{Y}^{(k)}} V^{\hat{Y}^{(k)}} + \cdots + \sum_{k=1}^{m} V_{X_1X_2\ldots X_n}^{\hat{Y}^{(k)}} V^{\hat{Y}^{(k)}}}{\sum_{k=1}^{m}(V^{\hat{Y}^{(k)}})^2} \\
&= \frac{\sum_{k=1}^{m} V^{\hat{Y}^{(k)}} \left(\sum_{i=1}^{n} V_{X_i}^{\hat{Y}^{(k)}} + \sum_{i_1=1}^{n} \sum_{i_2=1+i_1}^{n} V_{X_{i_1}X_{i_2}}^{\hat{Y}^{(k)}} + \cdots + V_{X_1X_2\ldots X_n}^{\hat{Y}^{(k)}}\right)}{\sum_{k=1}^{m}(V^{\hat{Y}^{(k)}})^2} \\
&= \frac{\sum_{k=1}^{m} V^{\hat{Y}^{(k)}} V^{\hat{Y}^{(k)}}}{\sum_{k=1}^{m}(V^{\hat{Y}^{(k)}})^2} \\
&= 1
\end{aligned} \tag{21}$$

Proportion 2.2.1 For all input terms i the vector projection indices satisfy:

(i) $0 \le \boldsymbol{\theta}_i \le 2\pi, 0 \le P_i \le 1$

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-259, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 5 July 2016

(ii) If X_i has no effect on $\boldsymbol{Y}$, then $P_i = 0$

(iii) If X_i has no effect on $\boldsymbol{Y}$ but X_j has effect on $\boldsymbol{Y}$, then $P_{i,j} = P_j$

(iv) $P_i = \sum_{k=1}^{m} w_k S_i^{Y^{(k)}}, w_k = \frac{(V^{\hat{Y}^{(k)}})^2}{\sum_{k=1}^{m}(V^{\hat{Y}^{(k)}})^2}$

(v) $P_i = \sum_{k=1}^{m} w_k corr(\hat{Y}^{(k)}, E(\hat{Y}^{(k)} \mid X_i))$

where $corr(\cdot,\cdot)$ is the correlation coefficient between two random variables.

Proof. Point(i):positivity is clear, as $V_{X_i}^{\hat{Y}^{(k)}}$ and $V^{\hat{Y}^{(k)}}$ are positive; $P_i \le 1$ follows from Eq.(19).Point(ii)and point(iii)are easy to be proved by the definition. For (iv) and (v), more details are given in section 2.3.

2.3 The link between the vector projection index and the Sobol index

A comparable definition of S_i proposed by Sobol' (Sobol 1996) is based on the correlation between the kth output $\hat{Y}^{(k)}$ and the conditional expectation $E(\hat{Y}^{(k)} \mid X_i)$ of the kth output.

$$S_i^{\hat{Y}^{(k)}} = corr(\hat{Y}^{(k)}, E(\hat{Y}^{(k)} \mid X_i)) \tag{21}$$

And P_i is given by

$$P_i = \frac{\sum_{k=1}^{m} V_{X_i}^{\hat{Y}^{(k)}} V^{\hat{Y}^{(k)}}}{\sum_{k=1}^{m}(V^{\hat{Y}^{(k)}})^2} = \frac{\sum_{k=1}^{m} \frac{V_{X_i}^{\hat{Y}^{(k)}}}{V^{\hat{Y}^{(k)}}}(V^{\hat{Y}^{(k)}})^2}{\sum_{k=1}^{m}(V^{\hat{Y}^{(k)}})^2} = \frac{\sum_{k=1}^{m} S_i^{\hat{Y}^{(k)}}(V^{\hat{Y}^{(k)}})^2}{\sum_{k=1}^{m}(V^{\hat{Y}^{(k)}})^2}$$
$$= \sum_{k=1}^{m} w_k S_{k,i} = \sum_{k=1}^{m} w_k corr(\hat{Y}^{(k)}, E(\hat{Y}^{(k)} \mid X_i)) \tag{22}$$

where $w_k = \frac{(V^{\hat{Y}^{(k)}})^2}{\sum_{k=1}^{m}(V^{\hat{Y}^{(k)}})^2}$ is the weight of the kth correlation coefficient of the output $Y^{(k)}$ and the conditional expectation $E(\hat{Y}^{(k)} \mid X_i)$ .When $m=1$, the weight $w_k = 1$ and the index P_i degrades into the main effect index S_i of the scalar output. Similarly P_{Ti} is given by

$$P_{T_i} = \frac{\sum_{k=1}^{m} S_{T_i}^{Y^{(k)}}(V^{\hat{Y}^{(k)}})^2}{\sum_{k=1}^{m}(V^{\hat{Y}^{(k)}})^2} = \sum_{k=1}^{m} w_k S_{T_i}^{Y^{(k)}} = \sum_{k=1}^{m} w_k corr(Y^{(k)}, E(Y^{(k)} \mid \boldsymbol{X}_{\sim i})) \tag{23}$$

And when $m=1$, the index P_{Ti} degrades into the total effect index S_{Ti} of the scalar output.

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-259, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 5 July 2016

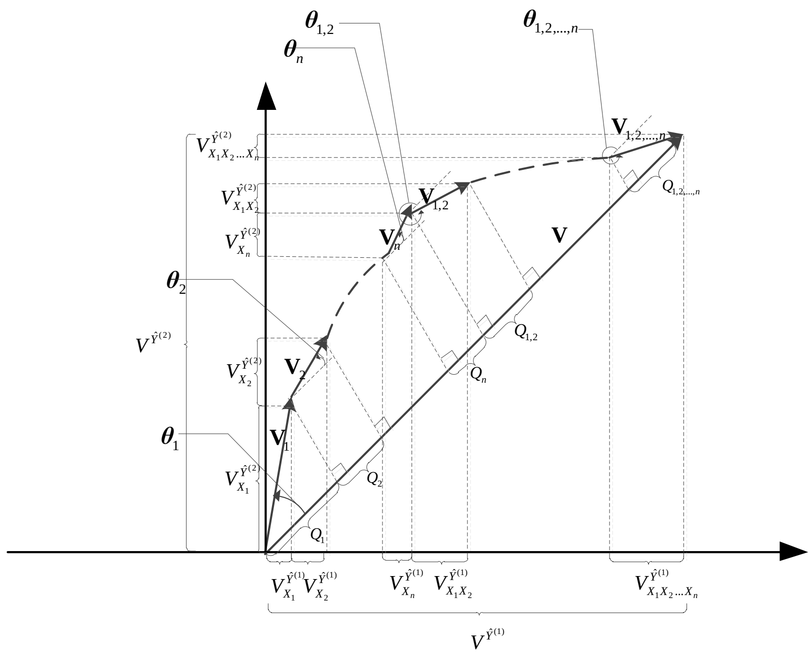

Fig 1.The relationship among $\mathbf{V}, \mathbf{V}_i, \mathbf{V}_{i_1,i_2}, \cdots, \mathbf{V}_{1,\ldots,i,\ldots n}$ for two outputs.

For the scalar output, the Sobol index measures the ratio of magnitude of the conditional variance to the magnitude of the unconditional variance. For the multivariate output case, multiple parameters can be expressed as a vector and the inner product can measure the similarity between the vectors. The degree of the coincidence, between the vector $\mathbf{V}_i$ included all conditional variances and the vector $\mathbf{V}$ included all unconditional variances implies the main effect contribution of each input factor to the multivariate variance of all the outputs. Fig.1 shows the geometric interpretation of the vector projection measure for multivariate output ($m=2$). In Fig.1, we can see that the vector angle θ_i reflects the difference of the direction between two vectors, the smaller is, the closer the overlapping between vectors is. Except for vector angle, the magnitude of vector $\mathbf{V}_i$ also influences the coincidence degree. The new sensitivity index is the ratio of the vector projection that from the vectors $\mathbf{V}_i$ to the vector $\mathbf{V}$ to the norm of the unconditional variance vector.

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-259, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 5 July 2016

3. Estimation of the new sensitivity indices

3.1 Polynomial Chaos Expansion

The Polynomial Chaos expansion (PCE) of 2-nd order random variable is a decomposition of the form (Wiener 1938, Ghanem and Spanos 1991)

$$Y = \sum_{j=0}^{\infty} \alpha_j \Psi_j(\boldsymbol{x}) \tag{24}$$

where α_j is the jth $(j = 1,\ldots,\infty)$ coefficient, Ψ_j $(j = 1,\ldots,\infty)$ are orthogonal to each other with respect the corresponding PDF (Xiu and Karniadakis 2002, Xiu 2010) and $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ are independent standard normal random variables. To be used in engineering models, Eq.(24) needs to be truncated. The order of the polynomials is M and the number of input variables is n, then the total number of terms $P+1$ with order less than or equal to M is given by

$$P+1 = \frac{(M+n)!}{M!n!} \tag{25}$$

There are two approaches for the estimation of the coefficients: projection and regression. The projection can take advantage of the orthogonal nature of the polynomial Ψ and the coefficients are estimated using multidimensional numerical integration (Ghanem and Spanos 1991), but it requires a large number of the model evaluations to compute integration (Xiu 2010). In the regression approach (Berveiller, Sudret et al. 2006, Sudret 2008), the coefficients are estimated by minimizing the sum of squares of the difference between a set of model evaluations $\mathbf{Y}_{\mathrm{N}}$. Assume a set of realizations to be $\{\boldsymbol{\xi}^{(1)}, \boldsymbol{\xi}^{(2)}, \ldots, \boldsymbol{\xi}^{(N_0)}\}$ for $\mathbf{X}$, where N_0 is the base number of model realizations. Then $\mathbf{Y}_{\mathrm{N}} = \{y^{(1)}(\boldsymbol{\xi}^{(1)}), y^{(2)}(\boldsymbol{\xi}^{(2)}), \ldots, y^{(N_0)}(\boldsymbol{\xi}^{(N_0)})\}$, $\boldsymbol{\xi}^{(k)} = (\xi_1^{(k)}, \ldots, \xi_n^{(k)}), (k = 1,\ldots,N_0)$ and we define $\boldsymbol{\Psi}$ the matrix whose coefficients are given by $\boldsymbol{\Psi}_{kj} = \Psi_j(\boldsymbol{\xi}^{(k)}), k = 1,\ldots,N_0; j = 0,\ldots,P+1$ with evaluation of the orthogonal polynomials at the collocation points $\boldsymbol{\xi}^{(k)}$, from these we can obtain the coefficients α_j as

$$\boldsymbol{\alpha} = (\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^{\mathrm{T}}\mathbf{Y}_{\mathrm{N}} \tag{26}$$

Due to the orthogonality of the basis, it is easy to show that the statistical moments of the random variable Y such as the mean and variance respectively read:

$$\begin{aligned} \overline{Y} &= E[Y] = \alpha_0 \\ \sigma_Y^2 &= Var[Y] = \sum_{j=1}^{P} \alpha_j^2 E[\Psi_j^2] \end{aligned} \tag{27}$$

3.2 Estimating the new sensitivity indices for multivariate output by the PCE

The Sobol decomposition of multivariate output $\hat{Y}^{(k)}$ $(k = 1,\ldots,m)$ (see Eq.(2)) can be

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-259, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 5 July 2016

Hydrology and Earth System Sciences Open Access Discussions

expressed from a reorganization of its PCE (Sudret 2008) as

$$\begin{aligned}\hat{Y}^{(k)} &= \alpha_0^{(k)} + \sum_{i=1}^{d}\sum_{j\in I_i}\alpha_j^{(k)}\Psi_{k,j}(x_i) + \sum_{1\le i_1\le i_2\le d}\ \sum_{j\in I_{i_1,i_2}}\alpha_j^{(k)}\Psi_j^{(k)}(x_{i_1},x_{i_2}) + \cdots \\ &+ \sum_{1\le i_1\le\cdots\le i_s\le d}\ \sum_{j\in I_{i_1,\ldots,i_s}}\alpha_j^{(k)}\Psi_j^{(k)}(x_{i_1},\ldots,x_{i_s}) + \cdots\end{aligned} \tag{28}$$

where $I_{i_1,\ldots,i_s}$ is the set of multiple tuples and indices $(i_1,\ldots,i_s)$ are nonzero:

$$I_{i_1,\ldots,i_s} = \begin{Bmatrix} a_i^{(k)} > 0 & \forall i \in \{i_1,\ldots,i_s\} \\ a_i^{(k)} = 0 & \forall i \notin \{i_1,\ldots,i_s\} \end{Bmatrix} \tag{29}$$

where $a_i^{(k)} \ge 0$ and $\sum_{i=1}^{M} a_i^{(k)} \le P$ is an integer set used to correspond each term in Eq.(24) to the orthogonal polynomials (Sudret 2008). The variance of $\hat{Y}^{(k)}$ can be obtained using Eq.(27), and therefore the main effect of the new sensitivity index for an input variable X_i can be given by:

$$\hat{P}_i = \frac{\sum_{k=1}^{m}\left(\sum_{j\in I_i}(\alpha_j^{(k)})^2 E((\Psi_j^{(k)})^2)\sum_{i=1}^{P}(\alpha_i^{(k)})^2 E((\Psi_i^{(k)})^2)\right)}{\sum_{k=1}^{m}\left(\sum_{i=1}^{P}(\alpha_i^{(k)})^2 E((\Psi_i^{(k)})^2)\right)^2} \tag{30}$$

And the interaction effect of any group of input variables $X_{i_1,\ldots,i_r}$ can be estimated as follows by PCE

$$\hat{P}_{i_1,\ldots,i_s} = \frac{\sum_{k=1}^{m}\left(\sum_{j\in I_{i_1,\ldots,i_s}}(\alpha_j^{(k)})^2 E((\Psi_j^{(k)})^2)\sum_{i=1}^{P}(\alpha_i^{(k)})^2 E((\Psi_i^{(k)})^2)\right)}{\sum_{k=1}^{m}\left(\sum_{i=1}^{P}(\alpha_i^{(k)})^2 E((\Psi_i^{(k)})^2)\right)^2} \tag{31}$$

Similarly, the total effect index of X_i can be compactly expressed as

$$\hat{P}_{T_i} = \frac{\sum_{k=1}^{m}\left(\sum_{j\in I_{i^+}}(\alpha_j^{(k)})^2 E((\Psi_j^{(k)})^2)\sum_{i=1}^{P}(\alpha_i^{(k)})^2 E((\Psi_i^{(k)})^2)\right)}{\sum_{k=1}^{m}\left(\sum_{i=1}^{P}(\alpha_i^{(k)})^2 E((\Psi_i^{(k)})^2)\right)^2} \tag{32}$$

where $I_{i+} = I_i + \sum_{i_1\ne i}^{n} I_{i,i_1} + I_{1,\ldots,i,\ldots n}$.

Therefore, the proposed new sensitivity indices for the multivariate output have analytical expressions which are estimated from the coefficients of the PCE of the output variables. There is no additional cost for obtaining the new sensitivity indices once the coefficients of the PCE are available.

4. Example

In this section the new sensitivity index is applied to two numerical examples and a hydrological model to analyze the influence of the input variables on the multivariate output. And the results of the new sensitivity index based on the vector projection index will be compared with

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-259, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 5 July 2016

Hydrology and Earth System Sciences
Open Access
Discussions

the sensitivity index based on the covariance decomposition method. To ensure the convergence of the computational results, the sample size of Monte Carlo Simulation (MCS) for all the sensitivity indices is taken as $N = 100000$. The results of the PCE in different orders (M) to estimate the new sensitivity indices are compared with the results of the MCS to verify the effectiveness of the PCE based method.

4.1 Numerical examples

Example 4.1 Consider a linear model with multivariate output

$$\begin{cases} y^{(1)} = x_1 + 9\times x_2 + x_3 + x_4 \\ y^{(2)} = 100\times(x_1 + x_2 + 5\times x_3 + x_4) \end{cases} \quad (33)$$
$$x_1 \sim N(0,1), x_2 \sim N(0,1), x_3 \sim N(0,1), x_4 \sim N(0,1)$$

The sensitivity results obtained by the Sobol index S_i for each output, the covariance decomposition method $S1_i^M$ and the proposed index P_i are listed in Table 1. For the above equations, we magnify the second function y_2 100 times to simulate the influence of dimension. Since the input variables are standardized normal random variables, it is straightforward to find the important measure ranking is $X_2 > X_3 > X_1 = X_4$ through qualitative analysis.

Table 1 Sensitivity indices for Example 1

Indices	Function evaluation number	X_1	X_2	X_3	X_4	Time
S_i-$y^{(1)}$-MCS	$(4+2)\times 10^6$	0.0134 (3)	0.9643 (2)	0.0132 (1)	0.0129 (4)	2.2158s
S_i-$y^{(2)}$-MCS	$(4+2)\times 10^6$	0.0355(3)	0.0357(2)	0.8930(1)	0.0354(4)	2.2158s
$S1_i^M$-MCS	$(4+2)\times 10^6$	0.0360(3)	0.0365(2)	0.8925(1)	0.0359(4)	2.2158s
P_i-*MCS*	$(4+2)\times 10^6$	0.0231(3)	0.4999 (2)	0.4517 (1)	0.0229 (4)	2.1460s
P_i-PCE	45(M=2)	0.0239(4)	0.5221 (1)	0.4314 (2)	0.0241 (3)	0.0632s
P_i-PCE	105(M=3)	0.0238(4)	0.4997 (1)	0.4526 (2)	0.0238 (3)	0.0662s
P_i-PCE	210(M=4)	0.0233(4)	0.5001 (1)	0.4519 (2)	0.0228 (3)	0.0731s

Since there are no interaction term in this example, just the main sensitivity index of the original sensitivity indices is calculated. From Table 1, it can be observed the following aspects. Firstly, the importance rankings and the sensitivity values obtained by S_i for $\mathrm{y}^{(2)}$ and $S1_i^M$ are the same. This is easy to explain by the fact that $S1_i^M$ is influenced by the high order of magnitude of dimension of the second function $\mathrm{y}^{(2)}$ .Therefore, $S1_i^M$ can't describe output variability comprehensively during the dimensions of outputs are different, and the magnitude orders of the $\mathrm{y}^{(1)}$ and $\mathrm{y}^{(2)}$ have too big discrepancy. Secondly, the ranking result of the vector projection index from the quantitative analysis is equal to the ranking of qualitative analysis, which denotes that the importance measure based on the new sensitivity index is more applicable than the traditional indices for the multivariate output. Thirdly, there is less computational cost for PCE to obtain the convergent values. So the accuracy and efficiency for estimating the new sensitivity index can be improved by PCE based method.

Example 4.2 Consider the following nonlinear model used in (Luyi, Zhenzhou et al. 2016)

$$\begin{aligned} g^{(1)}(\boldsymbol{X}) &= 0.03 - \frac{1.905X_1X_2^2}{X_3X_7} - \frac{0.565X_1X_2^2}{X_4X_8} \\ g^{(2)}(\boldsymbol{X}) &= X_5X_3 - 1.185X_1X_2 \\ g^{(3)}(\boldsymbol{X}) &= X_6X_4 - 0.75X_1X_2 \end{aligned} \quad (34)$$

The input variables follow normal distribution, and their distribution parameters are shown in Table 2. The sensitivity results are listed in Table 3. Since this example has interaction terms, the main effect indexes and the total effect indexes based on the covariance decomposition and the vector projection are both presented in Table 3. To compare other differency between the new indices with the traditional method for multivariate output expect dimension, we calculate the results of $\hat{S}1_i^M$ and $\hat{S}T_i^M$ which the influence of the dimension of the outputs is eliminated.

Table 2 the distribution parameters of Example 2

Variables	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Mean	20000	12	0.04	9.82×10^{-4}	1.34×10^{7}	3.35×10^{8}	2×10^{10}	1×10^{11}
Standard deviation	1400	0.12	0.0048	5.892×10^{-5}	2.412×10^{6}	4.02×10^{7}	1.2×10^{9}	6×10^{9}

Table 3 Sensitivity indices for Example 2

indices	Function

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-259, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 5 July 2016

Hydrology and Earth System Sciences Discussions Open Access EGU

	evaluation number	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	Time
$S1_i^M$-MCS	$(8+2)\times10^6$	0.0333 (4)	0.0003 (6)	0.2558 (2)	0.0230 (5)	0.5788 (1)	0.0959 (3)	0 (7)	0 (8)	5.4545s
ST_i^M-MCS	$(8+2)\times10^6$	0.0338 (4)	0.0003 (6)	0.2653 (2)	0.0241 (5)	0.5874 (1)	0.0963 (3)	0 (7)	0 (8)	5.4545s
$\hat{S1}_i^M$-MCS	$(8+2)\times10^6$	0.1758 (3)	0.0122 (7)	0.1807 (2)	0.0920 (5)	0.3147 (1)	0.1506 (4)	0.0092 (8)	0.0547 (6)	5.5632s
$\hat{ST}_i^M$-MCS	$(8+2)\times10^6$	0.1775 (3)	0.0130 (7)	0.1870 (2)	0.0940 (5)	0.3193 (1)	0.1521 (4)	0.0100 (8)	0.0553 (6)	5.5632s
P_i-*MCS*	$(8+2)\times10^6$	0.1535 (3)	0.0105 (7)	0.2149 (2)	0.0690 (5)	0.4034 (1)	0.0836 (4)	0.0079 (8)	0.0474 (6)	5.4323s
P_{Ti}-*MCS*	$(8+2)\times10^6$	0.1535 (3)	0.0105 (7)	0.2149 (2)	0.0696 (5)	0.4034 (1)	0.0836 (4)	0.0079 (8)	0.0474 (6)	5.4323s
P_i-*PCE*	315 M=2	0.1531 (3)	0.0109 (7)	0.2148 (2)	0.0687 (5)	0.4043 (1)	0.0852 (4)	0.0083 (8)	0.0485 (6)	0.0694s
P_{Ti}-*PCE*	315 M=2	0.1537 (3)	0.0111 (7)	0.2209 (2)	0.0695 (5)	0.4101 (1)	0.0855 (4)	0.0085 (8)	0.0479 (6)	0.0694s
P_i-*PCE*	1155 M=3	0.1542 (3)	0.0111 (7)	0.2151 (2)	0.0691 (5)	0.4026 (1)	0.0845 (4)	0.0084 (8)	0.0479 (6)	0.1458s
P_{Ti}-*PCE*	1155 M=3	0.1550 (3)	0.0112 (7)	0.2212 (2)	0.0698 (5)	0.4084 (1)	0.0848 (4)	0.0085 (8)	0.0483 (6)	0.1458s
P_i-*PCE*	3465 M=4	0.1539 (3)	0.0110 (7)	0.2153 (2)	0.0688 (5)	0.4035 (1)	0.0841 (4)	0.0084 (8)	0.0478 (6)	0.7412s
P_{Ti}-*PCE*	3465 M=4	0.1546 (3)	0.0111 (7)	0.2215 (2)	0.0696 (5)	0.4093 (1)	0.0844 (4)	0.0086 (8)	0.0473 (6)	0.7412s

We note from Table 3 that the importance rankings of the nonlinear model of the input variables obtained by P_i and P_{Ti} are same with the rankings obtained by $\hat{S1}_i^M$ and $\hat{ST}_i^M$, but the values of the importance measure of them are different. That is because that $\hat{S1}_i^M$ and $\hat{ST}_i^M$ only contain the magnitudes of the variances in the multiple output space, whereas P_i and P_{Ti} include the magnitudes of variances and the directions in the multiple output space. This indicates that the importance measures based on the vector projection are more comprehensive than the generalized Sobol indexes. In addition, although the results of P_i and P_{Ti} estimated by MCS and PCE are approximately equal, the computation cost of PCE is much less than that of MCS. Once the coefficients of the PCE are estimated, the multivariate sensitivity indices can be obtained without additional computational cost shown in Eqs.(30)-(32). Therefore, the proposed measure P_i

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-259, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 5 July 2016

Hydrology and Earth System Sciences
Discussions
Open Access

provides an efficient alternative for the sensitivity analysis for multivariate output space by taking both of its dimension, magnitudes and directions of the multivariate variances into account simultaneously.

4.2 The hydrological model: HBV model

The HBV model is a conceptual model for rainfall-runoff simulation and takes the precipitation, temperature and potential evaporation as the inputs. The model consists of a degree-day snow model, soil-moisture accounting model and a runoff response model (Kollat, Reed et al. 2012). A sketch map of the HBV model is shown in Fig 2.

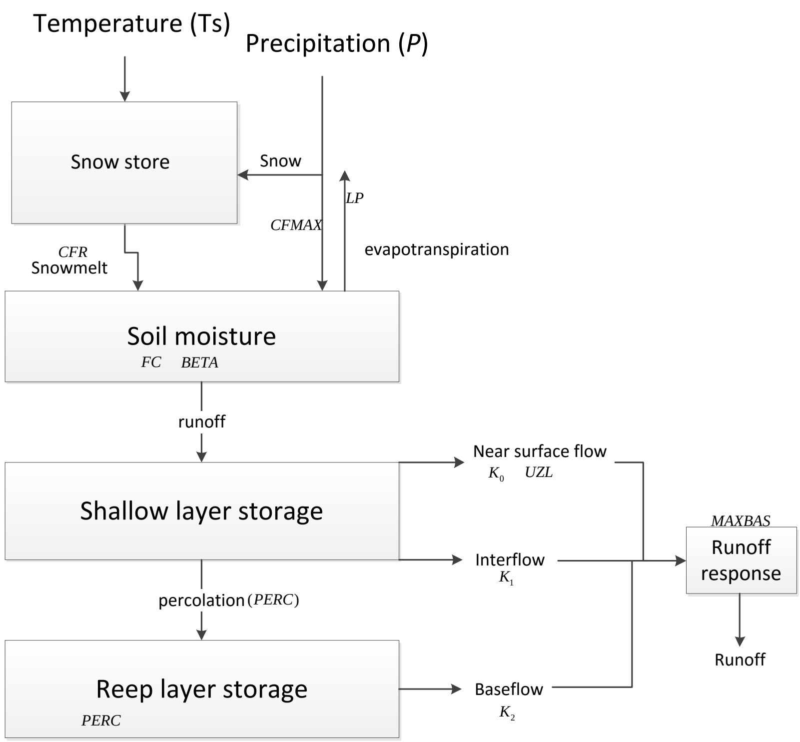

Fig.2 Sketch map of the HBV model

There are 13 parameters that should be calibrated for the HBV model. The parameters and the corresponding ranges are shown in Table 1 (the first four parameters are related to degree-day snow module, next three parameters are related to soil-moisture accounting model, and the last six ones are related to the runoff response model). The ranges of the parameters are based on prior studies (Kollat, Reed et al. 2012).

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-259, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 5 July 2016

Hydrology and Earth System Sciences Discussions
Open Access

Table 4 The parameters of HBV model and the corresponding ranges

Parameters	Meaning	Units	Ranges
Ts	Threshold temperature	$^{\circ}\mathrm{C}$	[-3.0,3.0]
$CFMAX$	Degree day factor	$\mathrm{mm}\cdot{}^{\circ}\mathrm{C}^{-1}\cdot\mathrm{d}^{-1}$	[0.0,20.0]
CFR	Refreezing factor	-	[0.0,1.0]
CWH	Water holding capacity factor of snow	-	[0.0,0.8]
$BETA$	Shape parameter	-	[0.0,7.0]
LP	limiting soil moisture at which potential evaporation occurs	-	[0.3,1.0]
FC	Maximum soil moisture content	mm	[0.0,2000.0]
$PERC$	percolation rate into deep layer	$\mathrm{mm}\cdot\mathrm{d}^{-1}$	[0.0,100.0]
K_0	near-surface flow recession coefficient	d^{-1}	[0.05,2]
K_1	interflow recession coefficient	d^{-1}	[0.01,1]
K_2	base flow recession coefficient	d^{-1}	[0.05,0.1]
UZL	Near surface flow threshold	mm	[0.0,100.0]
$MAXBAS$	Base length for transformation	d	[1,6]

There are a variety of criterions for the calibration of HBV model (Diskin and Simon 1977, van Werkhoven, Wagener et al. 2009). Here we consider three metrics, which are Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe 1970, Kollat, Reed et al. 2012), Transformed Root-Mean Square Error (TRMSE) (Kollat, Reed et al. 2012) and Slope of the Flow Duration Curve (SDFCE). Jan suggested that the combination of different functions is suitable to judge different parameter sets which may perform more or less similarly well (Seibert 1997).

Nash Sutcliffe Efficiency(NSE)

The first objective emphasizes peak flow errors using the Nash-Sutcliffe Efficiency as shown in Eq.(34),

$$NSE = 1 - \frac{\sum_{t=1}^{N}(Q_{s,t} - Q_{o,t})^2}{\sum_{t=1}^{N}(Q_{o,t} - \overline{Q_o})^2} \tag{34}$$

where $Q_{s,t}$ is the simulated runoff at time t , $Q_{o,t}$ is the observed runoff at time t ,and $\overline{Q_o}$ is the mean observed flow over the calibration period. N is the summation, which performs over $t=1$ through the number of time steps on the calibration period.NSE is most often used as a hydrologic model calibration objective,which ranges from 1 to $-\infty$

Transformed Root-mean-square-error (RMSE)

The second objective emphasizes low flow errors using the Box-Cox transformed root-mean-

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-259, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 5 July 2016

square-error as shown in Eq.(35)

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(Q_{s,t} - Q_{o,t})^2} \quad (35)$$

Slope of the Flow Duration Curve(SDFCE)

The third objective emphasizes the flashiness of a watershed's response by minimizing in simulating the slope of the flow duration curve(SFDCE) as shown in Eq.(35)

$$SFDCE = \left|\frac{Q_{s,67\%} - Q_{s,33\%}}{Q_{o,67\%} - Q_{o,33\%}} - 1\right| \times 100\% \quad (35)$$

About the three response outputs, they have different dimensions and the third function's dimension has the largest orders of magnitude.

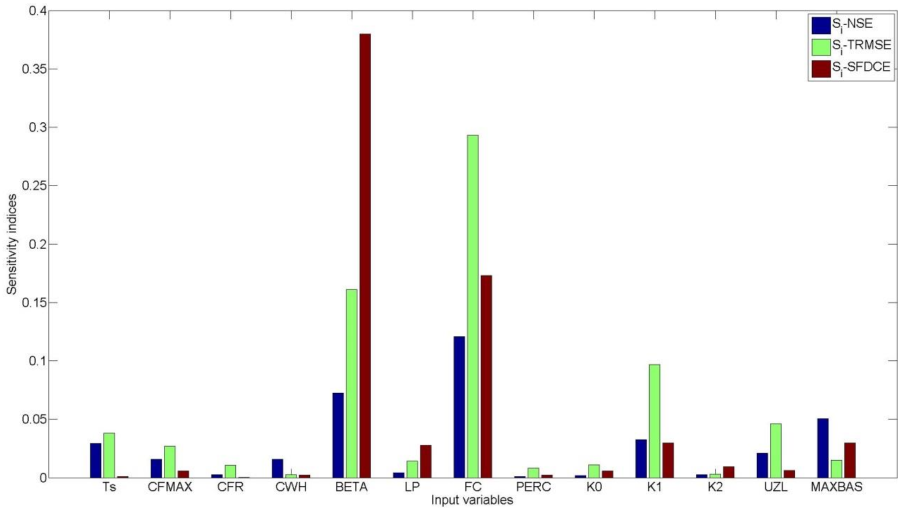

Fig.3.the main effect of Sobol index of three outputs

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-259, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 5 July 2016

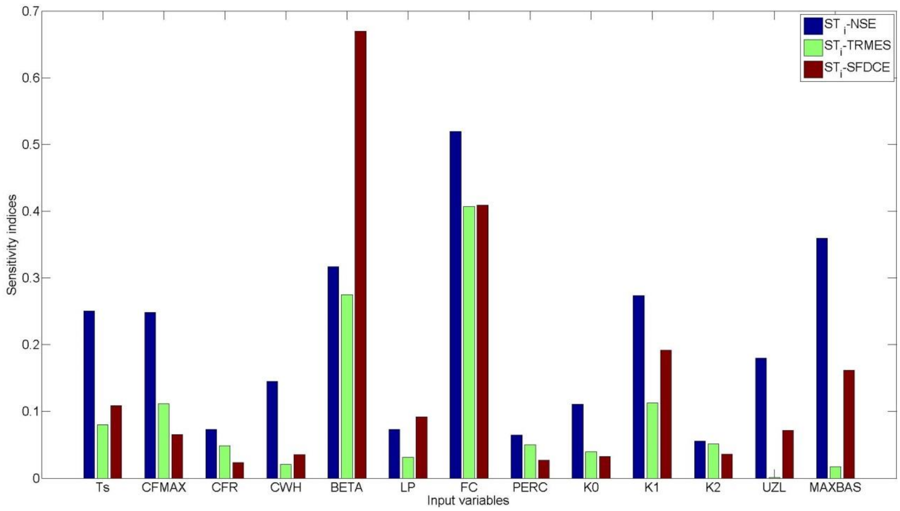

Figure 4.the total effect of Sobol index of three outputs

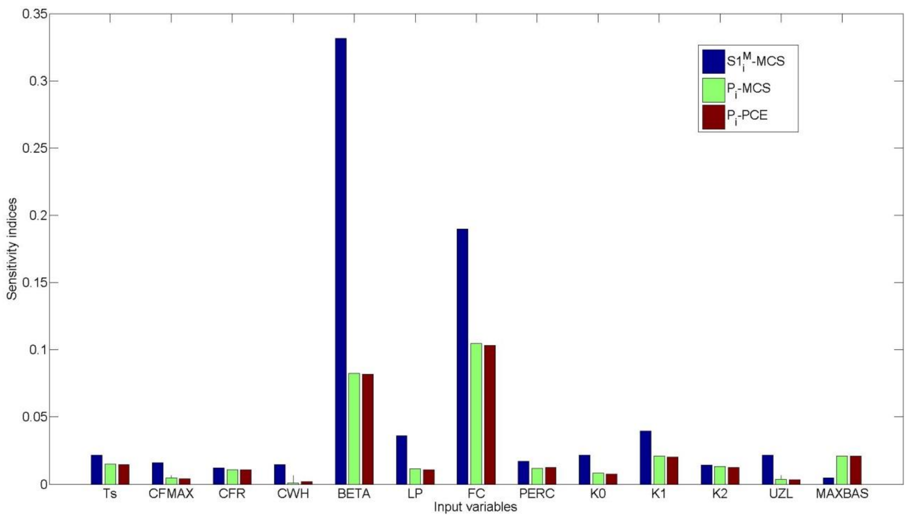

Fig.5.The main effect indices of multivariate output of the HBV model

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-259, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 5 July 2016

Hydrology and Earth System Sciences
Open Access
Discussions

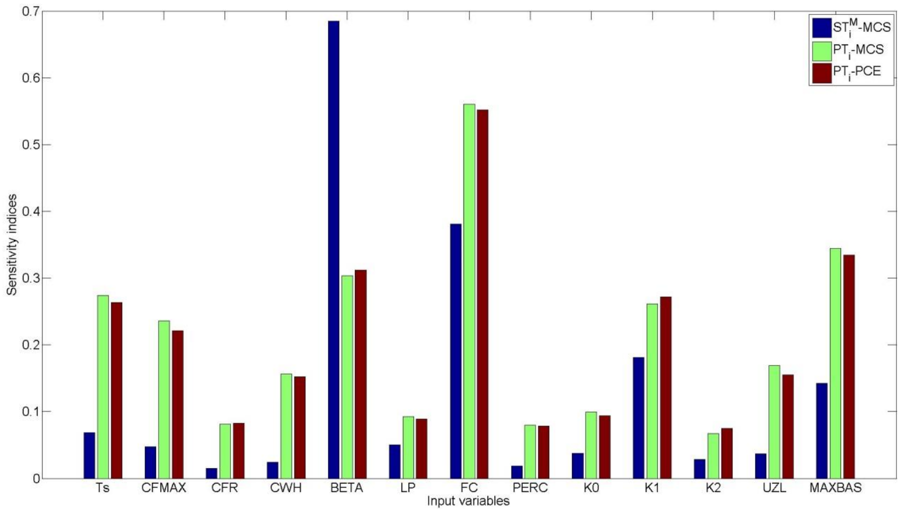

Fig.6.The total effect indices of multivariate output of the HBV model

In Figs.3 and 4, the results of the Sobol index S_i and ST_i of three outputs are presented respectively. And the sensitivity analysis results of the multivariate output of the HBV model, which are obtained by the vector projection indices P_i and P_{Ti} and the covariance decomposition method $S1_i^M$ and ST_i^M , are shown in Figs.5 and 6. The MCS with $(13+2)\times10^5$ model evaluations and the PCE with 6720 model evaluations ($M=3$)are used to get the convergent results, which verifies the efficiency of the PCE.

As for this hydrological model, we can find For SFDCE, it can be found that the ranking obtained by S_i shown in Fig.3 is same as that by S_i^M shown in Fig.5, and the ranking obtained by S_{T_i} shown in Fig.4 is also same as that by $S_{T_i}^M$ shown in Fig.6,which is caused by the influence of dimension of SFDCE. This suggests that in multivariate output case, the magnitude orders of the dimension has great impact on ranking results. For the main effect, Fig.5 shows that although both P_i and $S1_i^M$ identify the same important variables BATE and FC, the rankings they are obtained are not same. P_i indicates that FC is more important than BATE based on the vector projection, while $S1_i^M$ indicates BATE has the largest importance for the multivariate output, followed by FC. For the total effect, Fig6 shows that the rankings obtained by PT_i and ST_i^M are totally different, and PT_i considers more interaction effect between the input variables, since PT_i which includes magnitudes of the variances and the directions in the dimensionless multiple output space, but the traditional sensitivity index just includes magnitudes of variances in the multiple output space. In addition, Fig.5 and 6 show that results of PCE are similar to those of MSC. The PCE is able to evaluate the proposed index, with 6720 model evaluations ($M=3$) which is much lower than

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-259, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 5 July 2016

Hydrology and Earth System Sciences
Discussions

MCS with $(13+2)\times 10^5$ model evaluations.

Based on the above results, it can be concluded that the parameters BETA and FC have much more importance for 3 outputs represented by NSE, TRMSE and SFDCE among 13 inputs, the following importance inputs are the parameters CFMAX, TS,CWH,K1,UZL,MAXBAS since they have large interaction effects. For the rest parameters CFR, LP, PERC, K0 and K2, they have less contribution to the multivariate output.

5. Conclusions

The vector projection importance measure is proposed in this paper to evaluate the comprehensive effect of the inputs on the magnitudes of variances and directions of the multiple output space. The mathematical properties of the new sensitivity index are derived and its geometric significance is discussed. Two numerical examples and a hydrological model are employed to verify the effectiveness of the proposed method. Comparison with the covariance decomposition method shows that the new sensitivity index based on the vector projection can measure the effect of the inputs on the whole uncertainty of the multivariate output synthetically. The rankings of the input variables obtained by the generalized sensitivity indices are not necessarily the same with the proposed index. This is easy to understand by the fact that the vector projection based method additionally considers the effects on the dimension and directions which are ignored by traditional indices. Thus, only measuring the effects of the input variables on the magnitudes of variances is not enough to reflect the relative importance of the input variables comprehensively. In addition, the Polynomial Chaos Expansion method is used to estimate the new sensitivity indices for the multivariate output, and the main computational cost of the PCE based method is the estimation of the coefficients of the expansions. Thus the PCE based method for estimating the new sensitivity index is efficient compared with the Monte Carlo Simulation.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. NSFC 51505382) and Natural Science Foundation of Shaan Xi Province (Grant No. 2016JQ1034).

Reference

Berveiller, M., B. Sudret and M. Lemaire (2006). "Stochastic finite element: a non intrusive approach by regression." Revue Européenne De Mécanique Numérique **15**(1): 3671-3677.

Borgonovo, E. (2007). "A new uncertainty importance measure." Reliability Engineering & System Safety **92**(6): 771-784.

Bosetti, V., G. Marangoni, E. Borgonovo, L. D. Anadon, R. Barron, H. C. Mcjeon, S. Politis and P. Friley (2015). "Sensitivity to energy technology costs: A multi-model comparison analysis." Energy Policy **80**: 244-263.

Campbell, K., M. D. McKay and B. J. Williams (2006). "Sensitivity analysis when model outputs are

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-259, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 5 July 2016

functions." Reliability Engineering & System Safety **91**(10–11): 1468-1472.

Campolongo, F., J. Cariboni and A. Saltelli (2007). "An effective screening design for sensitivity analysis of large models." Environmental Modelling & Software **22**(10): 1509-1518.

Campolongo, F., A. Saltelli and J. Cariboni (2011). "From screening to quantitative sensitivity analysis. A unified approach." Computer Physics Communications **182**(4): 978-988.

Cui, L., Z. Lü and X. Zhao (2010). "Moment-independent importance measure of basic random variable and its probability density evolution solution." Science China Technological Sciences **53**(4): 1138-1145.

Diskin, M. H. and E. Simon (1977). "A procedure for the selection of objective functions for hydrologic simulation models." Journal of Hydrology **34**(1–2): 129-149.

Durier, R. (1994). "Inner-Product Spaces." All Publications **78**(2): 161–173.

Gamboa, F., A. Janon, T. Klein and A. Lagnoux (2013). "Sensitivity indices for multivariate outputs." Comptes Rendus Mathematique **351**(7–8): 307-310.

Garcia-Cabrejo, O. and A. Valocchi (2014). "Global Sensitivity Analysis for multivariate output using Polynomial Chaos Expansion." Reliability Engineering & System Safety **126**: 25-36.

Ghanem, R. G. and P. D. Spanos (1991). "Stochastic Finite Elements: A Spectral Approach." Springer Berlin **224**.

Homma, T. and A. Saltelli (1996). "Importance measures in global sensitivity analysis of nonlinear models." Reliability Engineering & System Safety **52**(1): 1-17.

Kollat, J. B., P. M. Reed and T. Wagener (2012). "When are multiobjective calibration trade-offs in hydrologic models meaningful?" Water Resources Research **48**(3): n/a-n/a.

Lamboni, M., H. Monod and D. Makowski (2011). "Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models." Reliability Engineering & System Safety **96**(4): 450-459.

Luyi, L., L. Zhenzhou, F. Jun and W. Bintuan (2012). "Moment-independent importance measure of basic variable and its state dependent parameter solution." Structural Safety **38**: 40-47.

Luyi, L., L. Zhenzhou and D. Wu (2016). "A new kind of sensitivity index for multivariate output." Reliability Engineering & System Safety **147**: 123-131.

Nash, J. E. and J. V. Sutcliffe (1970). "River flow forecasting through conceptual models part I — A discussion of principles." Journal of Hydrology **10**(3): 282-290.

Saltelli, A., S. Tarantola and F. Campolongo (2000). "Sensitivity Analysis as an ingredient of modelling." Statistical Science **15**(4): 377-395.

Saltelli, A., S. Tarantola, F. Campolongo and M. Ratto (2004). Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models. New York, John Wiley.

Seibert, J. (1997). "Estimation of Parameter Uncertainty in the HBV Model." Nordic Hydrology **28**: 247-262.

Sobol', I. M. (2001). "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates." Mathematics and Computers in Simulation **55**(1–3): 271-280.

Sobol', I. M. and S. Kucherenko (2010). "A new derivative based importance criterion for groups of variables and its link with the global sensitivity indices." Computer Physics Communications **181**(7): 1212-1217.

Sobol, I. M. (1996). "On 'freezing' unessential variables." Vestnik Moskovskogo Universiteta,Serija Matematika **6**: 92-94.

Sobol′, I. M. and S. Kucherenko (2009). "Derivative based global sensitivity measures and their link with global sensitivity indices." Mathematics and Computers in Simulation **79**(10): 3009-3017.

Sudret, B. (2008). "Global sensitivity analysis using polynomial chaos expansions." Reliability Engineering & System Safety **93**(7): 964-979.

Szirtes, T. and P. Rózsa (2007). "- Applied Dimensional Analysis and Modeling (Second Edition)." Applied Dimensional Analysis & Modeling(5): 659–665.

van Werkhoven, K., T. Wagener, P. Reed and Y. Tang (2009). "Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models." Advances in Water Resources **32**(8): 1154-1169.

Wiener, N. (1938). "The Homogeneous Chaos." American Journal of Mathematics **60**(1): 897-936.

Xiu, D. (2010). Numerical Methods for Stochastic Computations: A Spectral Method Approach, Princeton University Press.

Xiu, D. and G. E. Karniadakis (2002). "The Wiener-Askey polynomial chaos for stochastic di eren-tial equations." Siam Journal on Scientific Computing **24**(2): 619-644.