Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2016-259-RC3, 2016

Hydrology and Earth System Sciences
Discussions

Interactive comment on "The new importance measures based on vector projection for multivariate output: application on hydrological model" *by* L. Xu et al.

Anonymous Referee #3

Received and published: 11 December 2016

General comments

This paper deals with global sensitivity analysis for numerical simulator providing a multivariate output. This kind of problem is very popular in the domain of computer experiments and for many years, many authors have proposed different solutions. Most of them generalize the variance-based sensitivity indices, commonly known as Sobol' indices, using an orthogonal decomposition of the model function with respect to the input probability distribution. Precisely, this orthogonal decomposition is applied either to the output components, either to the whole output or to its principal components coming from a PCA. Then, the variance or covariance is considered and a normalization lead to Sobol' indices or generalized Sobol' indices.

In this paper, the authors propose to normalize the outputs by their absolute means in order to get rid of the output dimensions. Their objective is to avoid difficulties of interpretation when we look at the results of a sensitivity analysis for a model where the outputs have strongly different amplitude, which mask the less spread outputs. The authors also propose a sensitivity index based on the projection of the output vector variance conditioned by a given input onto the unconditioned output vector variance normalized by square euclidean norm of the latter. They show that this sensitivity index is linked to the Sobol' indices associated to this input and to the different outputs. They apply it to the normalized outputs. Lastly, the authors proposed closed-form estimators for their sensitivity indices; these estimators are built from the coefficients of polynomial chaos expansions obtained from a learning sample. All of these elements of methodology are applied to toy functions and to an hydrological model.

From my point of view, this work could have been more interesting if the authors were essentially preoccupied by the idea consisting in projecting the conditional output variance vectors onto the unconditional ones. The first reason is that even if the projection technique is a smart approach, the proposed sensitivity index is not really a novelty. Indeed, this weighted sum of Sobol' indices over the different outputs is very similar to the one proposed by Gamboa et al (2013), the only difference consisting in squaring the output variance terms. The second reason is that the authors mix a normalization step with the definition of a so-called new importance measure and the reader runs the risks of not understanding which is the key element of this article. Lastly, the authors replace the simulator by a polynomial chaos expansion in order to estimate the proposed sensitivity index without a computational constraint. I think this last point is out of the scope of this work whose objective is to propose a new importance measure.

To conclude, I advise against manuscript this article which does not provide relevant technical or methodological innovations in sensitivity analysis and whose common thread is not enough motivated by hydrological modeling. Moreover, the manuscript is very poorly written (grammar, sentences, english, ...) and the argumentation needs

to be largely reworked. Maybe, if major revisions and pertinent technical points are provided, this manuscript could be accepted.

Specific comments

Abstract: - l. 31 : The “output decomposition approach” does not provide any information about the effect of an input on the model output. You have to specify on which elements the output should be decomposed. Do you mean the Hoeffding decomposition leading to the Sobol’ indices? - l. 31: Same remark concerning the “covariance decomposition approach”: do you mean the Hoeffding decomposition leading to the generalized Sobol’ indices (Gamboa et al., 2013)? - l. 33: The expression “the influence of input variables on the effects on the magnitudes of variances of the dimensionalities in the multiple output space” is poorly written. A reformulation is needed; e.g. “the contribution of input variables to the variance of the different output components”. - l. 34: “dimensionality direction” is not a common expression; what do you mean by “the effects on the dimensionality directions of output variance”? - l. 36: You deal with “conditional vectors”and “unconditional vectors” but these vector are conditional or unconditional relatively to which variable(s)? You do not indicate the nature of these vectors; I suppose these vectors are output vectors. - l. 37: If the concept of “dimensionless” is important in your work, you need to clarify its presence in the abstract because readers cannot understand why and how you consider dimensionless output. - l. 38-39: You need to clarify your concept of “directions of the dimensionalities” which is not widespread as far as I know, notably in the sensitivity analysis community. - References: You should add other references dealing with sensitivity analysis in a multivariate output context and providing some applications or sensitivity indices which are not based on output (co)variance decomposition. E.g. : Dependence and variance-based measures for sensitivity analysis with multidimensional variables, M. De Lozzo, A. Marrel, Stochastic Environmental Research & Risk Assessment, 2016 ; Screening and metamodeling of computer experiments with functional outputs - Application to thermal-hydraulic computations, Auder, B., Crecy, A.D., Iooss, B., Marquès,

M., Reliability Engineering & System Safety, 107, 122 – 131, 2012 ; Global sensitivity analysis for models with spatially dependent outputs, Marrel, A., Iooss, B., Jullien, M., Laurent, B., Volkova, E., Environmetrics, 22, 383–397, 2011.

Section 1: - l. 74: Replace “the continuous case” by “functional outputs” - l. 79: Firstly, you cannot reduced sensitivity analysis to variance-based methods. Secondly, the variance-based approaches do not only consider the model outputs; some of them consider the principal components of the output space. Lastly, both kinds of methods do not assume that the relationship between the different outputs is simple and additive. The additive aspect comes from the Hoeffding decomposition which can be applied to any square integrable function and this decomposition is the fundamental step to define the Sobol’ indices (the next steps consisting of taking the trace of the covariance on both sides of the decomposition and normalizing by the trace of the output covariance). - l. 81: What do you mean by “dimensions of measurement” ? Moreover, having outputs with different orders of magnitude is not necessarily a drawback which requires a normalization of these outputs, e.g. when the output is a quantity of interest discretized over a spatial grid or over a temporal interval. Consequently, you have to explain and justify when the dimensionless process is required, and in return, when the output dimensions have to be kept. - l. 84-86: The sentence “The variance [. . .] each variance dimensionality” is extremely poorly written. The variance associated to an output is a form of output uncertainty representation among others (probability density function, quantiles, . . .). It can be viewed as a spread measure associated to this output. But “the magnitude of each variance dimensionality” does not mean anything. - l. 88-89: A “transformed space” is not a particular space; you have to remove this expression. The output decomposition method simply orthogonalizes a set of output observation vectors and then, the variance-based sensitivity analysis consider these orthogonal vectors (or those having the most important variance) rather than the original ones. - l. 91: You cannot use the expression “outputs” for both model outputs and transformed model outputs; this could create a misunderstanding. - l. 92-95:- What do you mean by “the directions of all the variance dimensionalities” ? Sobol’ indices

explained the contribution of the different inputs on the different output variance. Similarly, the generalized Sobol' indices obtained from an orthogonal decomposition of the output space explained the contribution of the different inputs on the different output variance directions. Moreover, you propose to consider the vector containing the variances of the different outputs but this is not a new approach. Indeed, the generalized Sobol' indices are based on this vector; more precisely, these indices decompose the covariance of this vector using the Hoeffding decomposition applied to the model.

Section 2: - l. 113: Your definition is not the definition of the importance measure but the definition of sensitivity analysis (in Saltelli et al., 2014: "The definition of sensitivity analysis that matches the contentof this primer better is 'The study of how the uncertainty in theoutput of a model (numerical or otherwise) can be apportioned todifferent sources of uncertainty in the model input'."). Moreover, importance measure is not a scientific domain but another expression for sensitivity measure; consequently, Sobol' indices belong to the sensitivity analysis toolbox and are not "methodologies for IM" (this expression is unsuitable). - l. 118: this is the Hoeffding decomposition and this decomposition requires some conditions on the function g^{(1)}. Moreover, you should defined the mathematical expression (1) in function of the superscript (i), i in 1,...,m, rather than in function of the superscript (1). - l. 120: (X_i) is missing after g_i^{(i)}. - l. 122: g_i^{(1)} is not a variation but a function of the input X_i with zero mean. By definition of the Hoeffding decomposition, the terms have zero mean and are mutually orthogonal in a probabilistic point of view. - l. 127: Your writing of V^{Y^{(1)}} is pointless. - l. 145: Review you citation because you write Gamboa without completing the citation and add another citation in parenthesis. - l. 147-149: You need to provide the expressions of the covariance matrix C_i, C_{ij}, ... in the same way than for the standard variance decomposition (2). - l. 157-159: These indices are not the sum of variances. - l. 161: Where do these eigenvectors come from? Moreover, the "principle component decomposition" does not mean anything. - l. 163-164: These sentence is not at all clear. A reformulation is needed. - l. 166-170: You say that these methods ignore the influence of the output dimensions and then you explain that these meth-

ods take into account these dimensions. It is a little bit confusing. In reality, these indices consider the dimensions of the outputs and if you think that this consideration is a problem, you have to expand on what your meaning is. - l. 169: Does this method come from the literature or is it a new approach that you propose? - l. 171: This title is too long. - l. 173-175: This is the general framework that you consider all along the paper. This preliminaries shall be placed at the beginning of Section 2. - l. 178: This is not a common normalization. Usually the quantity of interest is divided by its standard deviation. You have to explain accurately your choice. Moreover, I do not agree with this choice because the new outputs keep different orders of magnitude. From my point of view, normalizing your outputs by their standard deviations or something like that is the only means to prevent outputs with different spread or, in other words, different dimensions. - l. 180: This is not the variance decomposition of the vector \hat{Y} but the variance decomposition of the vector \hat{Y} term by term. - l. 186: Equation (11) should be removed because it is the same expression. The superscript does not help the understanding of you paper. - l. 188: Justify the choice of the vector projection technique in the sensitivity analysis context. - l. 194: This sensitivity index is substantially the same as the one proposed by Gamboa et al. (2013). Indeed, both Gamboa et al. (2013) and you consider a weighted of Sobol' indices associated to the different outputs and the weights are very similar: you consider squared output variance divided by the sum of the different squared output variances while Gamboa et al. (2013) consider output variance divided by the sum of the different output variances. The only difference resides in squaring the variance and consequently, I am not convinced by the superiority of your sensitivity index. Please, explain the advantage of this squaring aspect. Same remarks for line 196 and over. - l. 211: The formula is wrong. Same reasons as above. - l. 215: (i),(ii), (iii) and (iv) are obvious. Moreover, these points are properties rather than propositions. Moreover, you must reverse the logical relationships in points (ii) and (iii) and use total variance-based indices because in practice, we are interesting in fixing the inputs to nominal values when these indices are equal to zero. These indices indicate the dependence between outputs and inputs.

- l. 227: Equation (21) is nothing more than one of the usual Sobol' index definitions applied to the kth output. - l. 230: Replace by "P_i can be rewritten as follows:" - l. 225: At the beginning of this subsection, you have to explain its objective because we do not understand the introduction of a new definition of Sobol' indices and your wish to link you sensitivity indices to the correlation between the outputs and the conditional outputs. - l. 241-242: Sobol' indices associated to a particular input represent the proportions of output variance explained by this input. - l. 241-251: Why does the notion of covariance disappear? This paragraph seems to be off topic in the context of Section 2.3. Moreover, you should develop the interpretations of Figure 1.

Section 3: - l. 253 and over: What is the role of PCE in this paper? The use of PCE is not motivated in Section 3 and a simple question is: why do not you estimate Sobol' indices using a Monte-Carlo sampling from the simulator (gˆ{(1)},...,gˆ{(m)})? - l. 270: The choice of \xiˆ{(i)} is curiously chosen for an instance of the variable X ! - l. 275: Is the matrix PsiˆT*Psi invertible? - l. 282-299: Your developments are a straightforward application of Sudret (2008) who worked on PCE-based Sobol' indices for monodimensional output problem. You must mention this point.

Section 4: - l. 320: In Table 1, your comparison contains an important bias. Indeed, the lines P_i-MCS and SI_iˆM-MCS do not only differ only on the output dimension but also in the formulation. More precisely, applying the methods P_i-MCS and SI_iˆM-MCS to the initial outputs leads to similar results while applying these methods to the normalized outputs leads to strongly different results. Consequently, the differences mentioned in Table 1 come essentially from the output normalization, rather than the expressions of the sensitivity indices. - l. 326-328: Your sentence is poorly written. Moreover, I do not agree with you because we often need to keep the dimensions of the outputs, notably when these outputs have the same nature (e.g. when these outputs correspond to the value of a quantity of interest at different location). - l. 328-331: What do you mean by "quantitative analysis" and "qualitative analysis" ? Why do you conclude that an equality in terms of these analysis lead to the superiority of

your sensitivity indices for a model with multivariate output. - l. 332: What do you mean by "convergent values" ? Which "Values" ? - l. 332-333: Why can the estimation accuracy and efficiency be improved by a PCE based method, given that the PCE is an approximation of the true model? All of my previous commentaries concerning Section 4 are also valid for the second toy example and the HBV model.

Conclusion: - l. 441-443: The new sensitivity index leads to the same conclusion than the covariance decomposition method. The differences shown in this paper are essentially explained by the normalization of the model outputs, not by the type of sensitivity analysis. Moreover, the proposed sensitivity index is a weighted sum of individual Sobol' indices very similar to the one proposed by Gamboa et al (2013). - l. 445-447: It is really necessary to distinguish your method of variance vector projection and the output normalization step. The approach based on vector projection does not break the dimensions associated to the different outputs, contrarily to the normalization step, dividing by the standard deviation or by the mean absolute value in your case. - l. 447-449: This sentence is not at all clear. A reformulation is needed. - l. 449-451: You should move this part before the result description in the conclusion. Moreover, you say that the main computational cost of the PCE is associated to the coefficient estimation but this is not a useful information for the reader at this step of this paper, unless you give complementary information concerning this cost and possible costs avoided by the PCE use.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-259, 2016.