**Article:**

Yang, Jiannan orcid.org/0000-0001-8323-7406 (Accepted: 2025) Derivative based global sensitivity analysis and its entropic link. Technometrics. ISSN 1537-2723 (In Press)

https://doi.org/10.1080/00401706.2025.2455143


Technometrics

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/utch20

# Derivative based global sensitivity analysis and its entropic link

Jiannan Yang

**To cite this article:** Jiannan Yang (21 Jan 2025): Derivative based global sensitivity analysis and its entropic link, Technometrics, DOI: 10.1080/00401706.2025.2455143

**To link to this article:** https://doi.org/10.1080/00401706.2025.2455143

# Derivative based global sensitivity analysis and its entropic link

Jiannan Yang*

School of Physics, Engineering and Technology, University of York, Heslington, York, YO10 5DD, UK

*jiannan.yang@york.ac.uk



## Abstract

Variance-based Sobol' sensitivity is one of the most well-known measures in global sensitivity analysis (GSA). However, uncertainties with certain distributions, such as highly skewed distributions or those with a heavy tail, cannot be adequately characterised using the second central moment only. Entropy-based GSA can consider the entire probability density function, but its application has been limited because it is difficult to estimate. Here we present a novel derivative-based upper bound for conditional entropies, to efficiently rank uncertain variables and to work as a proxy for entropy-based total effect indices. To overcome the non-desirable issue of negativity for differential entropies as sensitivity indices, we discuss an exponentiation of the total effect entropy and its proxy. Numerical verifications demonstrate that the upper bound is tight for monotonic functions and it provides the same input variable ranking as the entropy-based indices for about three-quarters of the 1000 random functions tested. We found that the new entropy proxy performs similarly to the variance-based proxies for a river flood physics model with 8 inputs of different distributions, and these two proxies are equivalent in the special case of linear functions with Gaussian inputs. We expect the new entropy proxy to increase the variable screening power of derivative-based GSA and to complement Sobol'-indices proxy for a more diverse type of distributions.

*Keywords:* sensitivity proxy; sensitivity inequality; conditional entropy; exponential entropy; entropy power; derivative-based GSA

## 1 Introduction

This research is motivated by applications of global sensitivity analysis (GSA) towards mathematical models. The uncertain inputs of a mathematical model induce uncertainties in the output. GSA helps to identify the influential inputs and is becoming an integral part of mathematical modelling.

The most common GSA approach examines variability using the output variance. Variance-based methods, also called Sobol' indices, decompose the function output into a linear combination of input and interaction of increasing dimensionality, and estimate the contribution of each input factor to the variance of the output (Soboĺ, 1993; Saltelli, 2008). As only the 2nd-order moments are considered, it was pointed out that the variance-based sensitivity measure is not well suited for heavy-tailed or multimodal distributions (Auder and Iooss, 2008; Pianosi and Wagener, 2015; Liu et al., 2006). Entropy is a measure of uncertainty similar to variance: higher entropy tends to indicate higher variance (for Gaussian, entropy is proportional to log variance). Nevertheless, entropy is moment-independent as it is based on the entire probability density function of the model output. It was shown in Auder and Iooss (2008) that entropy-based methods and variance-based methods can sometimes produce significantly different results.

Both variance-based and entropy-based global sensitivity analysis (GSA) can provide quantitative contributions of each input variable to the output quantity of interest. However, the estimation of variance and entropy based sensitivity indices can become expensive in terms of the number of model evaluations. For example, the computational cost using sampling based estimation for variance-based total effect indices is $N(d+1)$ (Saltelli, 2008; Puy et al., 2020), where $N$ is the base sample number and $d$ is the input dimension. Large values of $N$, normally in the order of thousands or tenths of thousands, are needed for more accurate estimate, and the computational cost has been noted as one of the main drawbacks of the variance-based GSA in Saltelli (2008). In addition, it was noted in Auder and Iooss (2008) that although both variance-based and entropy-based sensitivity analysis take long computational time, the convergence for entropy-based indices is even slower.

In contrast, for a differentiable function, derivative-based methods can be more efficient. For example, Morris' method (Morris, 1991) constructs a global sensitivity measure by computing a weighted mean of the finite difference approximation to the partial derivatives, and it requires only a few model evaluations. The computational time required can be many orders of magnitude lower than that for estimation of Sobol' sensitivity indices as demonstrated by Kucherenko et al. (2009) and it is thus often used for screening a large number of input variables.

Previous studies have found a link between the derivative-based measures and variance-based total effect indices. In Campolongo et al. (2007), a sensitivity measure $\mu*$ is proposed based on the absolute values of the partial derivatives. It is empirically demonstrated that for some practical problems, $\mu*$ is similar to the variance-based total indices. In Wainwright et al. (2014), the variance-based sensitivity indices are interpreted as difference-based measures, where the total sensitivity index is equivalent to taking a difference in the output when perturbing one of the parameters with the other parameters fixed. The similarity to partial derivatives helps to explain why the mean of absolute elementary effects from Morris' method can be a good proxy for the total sensitivity index for detecting unessential variables.

Observing the empirical success of the modified Morris' measure, Sobol' and Kucherenko (2009) have proposed the so-called derivative-based global sensitivity measures (DGSM). This importance criterion is similar to the modified Morris' measure, except that the squared partial derivatives are used instead of their absolute values. In addition, an inequality link between variance-based global sensitivity indices and the DGSM is established in the case of independent Uniform or Gaussian input variables. This inequality between DGSM and variance-based GSA has been extended to input variables belonging to the large class of Boltzmann probability measures in Lamboni et al. (2013). This link is via the Poincaré inequality and an optimal value for the scaling Poincaré constant has been developed in Roustant et al. (2017).

Inspired by the success of derivative-based proxies for Sobol' indices, in this paper, we present a novel derivative-based upper bound for entropy-based total effect indices. The key idea here is to make use of a well-known inequality between the entropy of a continuous random variable and its deterministic transformation. This inequality can be seen as a version of the information processing inequality and is shown here to provide an upper bound for the total effect entropy sensitivity measure. The entropy upper bounds are demonstrated to efficiently rank uncertain variables and can thus potentially be used as a proxy for entropy-based total effect indices for screening purposes. And that is the main contribution of this paper.

In addition, via exponentiation, we extend the upper bound to the widely used DGSM for a total sensitivity measure based on entropy power (also known as effective variance). In the special case with Gaussian inputs and linear functions, the proposed new proxy for entropy GSA is found to be equivalent to the proxy for variance-based total effect sensitivity. Furthermore, unlike the variance proxy, the inequality link between derivatives and entropy does not require the random inputs to be independent. The new entropy proxy is thus expected to not only increase the variable screening power of derivative-based GSA, but can also complement variance proxies for a more diverse type of distributions.

Note that we focus on the standard derivative-based sensitivity measures in this paper. A closely related derivative-based sensitivity analysis technique is active subspace, which makes use of the leading eigenspaces of the second moment matrix of the function partial derivatives. DGSM indices are the diagonal of this second moment matrix. Sensitivity indices based on active subspace have been found to bound DGSM and Sobol' total effect indices for scalar-valued outputs (Constantine and Diaz, 2017), and a generalization to vector-valued functions with Gaussian inputs has been discussed by Zahm et al. (2020). Instead of the function derivatives, active sensitivity directions can also be obtained from the derivatives of the output distributions using the leading eigenvectors the Fisher Information Matrix (FIM) (Yang, 2023). Use of the leading symplectic eigenspace of the FIM has also been proposed for decision oriented sensitivity analysis (Yang, 2024).

It should also be noted that, in addition to entropy-based measures, there are many other moment-independent sensitivity measures, such as the $\delta$-indicator (Borgonovo, 2007), maximum mean discrepancy (Da Veiga, 2021), Kolmogorov–Smirnov statistic (Pianosi and Wagener, 2015) and Kullback–Leibler (KL) divergence (Krzykacz-Hausmann, 2001; Liu et al., 2006). More discussions can be found in (Borgonovo and Plischke, 2016).

Many of the above mentioned measures can be seen as special cases of the Csiszar f-divergence between the conditional and unconditional output densities (Borgonovo and

Plischke, 2016; Da Veiga, 2015). The expected f-divergence can be re-expressed in terms of the joint probability density function (PDF) of the input and output distributions and the product of their marginals. This reformulation enables a more general framework in which a “distance” between the joint PDF and the product of the marginal PDFs is considered. Such a framework serves as a foundation for deriving certain dependence measures for sensitivity analysis, such as the Distance Correlation based on characteristic functions and the Hilbert–Schmidt Independence Criterion (HSIC) that generalizes the notion of covariance between two random variables (Da Veiga, 2015).

In what follows, we will first review global sensitivity measures in Section 2, where the motivation for the entropy-based measure and its proxy is discussed with an example. In Section 3, we first establish the inequality relationship between the total effect entropy measure and the partial derivatives of the function of interest, where mathematical proofs are provided. An exponential version of the inequality link is further proposed in Section 3. In Section 4, we provide numerical illustrations with analytical functions. Additional numerical tests using a randomised meta-function is given in the *Supplementary Material*. In Section 5, a river flood physics model is used to demonstrate the effectiveness of the new entropy proxy. Concluding remarks are given in Section 6.

# 2 Global sensitivity measures

In our context, an important question for GSA is: ‘Which model inputs can be fixed anywhere over its range of variability without affecting the output?’ (Saltelli, 2008). In this section, we first review both variance-based and derivative-based sensitivity measures, which can provide answers to the above ‘screening’ question. We then use a simple example to motivate the use of entropy-based sensitivity indices.

## 2.1 Total variance effect and its link with derivative-based GSA

The variance-based total effect measure accounts for the total contribution of an input to the output variation, and is often a preferred approach due to its intuitive interpretation and quantitative nature.

Let us denote $\mathbf{x} = (x_1, x_2, \ldots, x_d)$ as independent random input variables, and $y$ being the output of our computational model represented by a function $g$, such that $y = g(\mathbf{x})$.

The variance-based GSA decompose the output variance $V(Y)$ into conditional terms (Hoeffding, 1948; Soboĺ, 1993):

$$V(Y) = \sum_i V_i + \sum_i \sum_{i<j} V_{ij} + \cdots + V_{1,2,\ldots,d} \quad (1)$$

where

$$V_i = V[\mathbb{E}(Y \mid X_i)]; \quad V_{ij} = V[\mathbb{E}(Y \mid X_i, X_j)] - V[\mathbb{E}(Y \mid X_i)] - V[\mathbb{E}(Y \mid X_j)]$$

and so on for the higher interaction terms. $V_i$ measures the first order effect variance and $V_{ij}$ for a second order effect variance, where their contributions to the unconditional model output variance can be quantified as $V_i / V(Y)$ and $V_{ij} / V(Y)$ respectively. Analogous formulas can be written for higher-order terms, enabling the analyst to quantify the higher-order interactions.

The total order sensitivity index is then defined as:

$$S_{T_i} = \frac{\mathbb{E}[V(Y \mid \mathbf{X}_{\sim i})]}{V(Y)} = \frac{V_{T_i}}{V(Y)} \quad (2)$$

where $\mathbf{X}_{\sim i}$ is the set of all inputs except $X_i$, and $\mathbb{E}[V(Y \mid \mathbf{X}_{\sim i})] = V(Y) - V[\mathbb{E}(Y \mid \mathbf{X}_{\sim i})]$ is the remaining variance if the true values of $\mathbf{X}_{\sim i}$ can be determined. The total order sensitivity index measures the total contribution of the input $X_i$ to the output variance, including its first order effect and its interactions of any order with other inputs.

When the function $g(\cdot)$ is differentiable, local sensitivity can be measured using the square integrable partial derivatives $g_{x_i} = \partial g / \partial x_i$ which can be seen as a limiting version of Morris' elementary effect when the incremental step tends to zeros (Sobol' and Kucherenko, 2009). The partial derivative depends on a nominal point. For global sensitivity analysis, an average of the partial derivatives can be taken over the input parameter space:

$$\mu_i = \mathbb{E}\left[\left|\frac{\partial g(\mathbf{X})}{\partial x_i}\right|\right] = \int \left|\frac{\partial g(\mathbf{x})}{\partial x_i}\right| f_X(\mathbf{x}) d\mathbf{x} \quad (3)$$

for $i = 1, 2, \ldots, d$ and $f_X(\mathbf{x})$ is the PDF of $\mathbf{x}$. As pointed out by Sobol' and Kucherenko (2009), for uniformly distributed inputs, the measure $\mu_i$ can be seen as a limiting version of the modified Morris' index $\mu^*$.

Based on that observation, the Derivative-based Global Sensitivity Measure (DGSM):

$$\nu_i = \mathbb{E}\left[\left(\frac{\partial g(\mathbf{X})}{\partial x_i}\right)^2\right] = \int \left(\frac{\partial g(\mathbf{x})}{\partial x_i}\right)^2 f_X(\mathbf{x}) d\mathbf{x} \quad (4)$$

has been proposed to be used as a proxy for $S_{T_i}$ (Sobol' and Kucherenko, 2009) to detect un-influential input variables. In particular, the total sensitivity variance $V_{T_i} = \mathbb{E}[V(Y \mid \mathbf{X}_{\sim i})]$ is upper bounded by DGSM via the following inequality:

$$V_{T_i} \leq C_i \nu_i \quad (5)$$

based on Poincaré inequality and the Poincaré constants were found to be optimal for $C_i$ (Lamboni et al., 2013; Roustant et al., 2017). Note that independent input variables are required for the DGSM-based upper bound as it is based on variance decomposition.

Eq 5 thus provides a screening method using the upper bound, which is typically computationally faster compared to a direct estimation of the Sobol' indices. Tighter the upper bound, more effective the low cost screening is.

Divide Eq 5 by the output variance $V(Y)$ from both sides, we then have the upper bound $S_{T_i} \le C_i \nu_i / V(Y)$ which can be used as a proxy for $S_{T_i}$ for variable screening. For Gaussian inputs with variance $\sigma_i^2$, $C_i = \sigma_i^2$ and the inequality in Eq 5 becomes $S_{T_i} \le \sigma_i^2 \nu_i / V(Y)$ as given in Sobol' and Kucherenko (2009).

## 2.2 A motivating example and entropy-based sensitivity

Pianosi and Wagener (2015) pointed out that a major limitation of variance-based sensitivity indices is that they implicitly assume that output variance is a sensible measure of the output uncertainty. However, if the output distribution is multi-modal or if it is highly skewed, using variance as a proxy of uncertainty may lead to contradictory results.

To illustrate this point, we look at the simple function $y = x_1 / x_2$. In this case, the two inputs both follow the chi-squared $\chi^2$ distribution with $x_1 \sim \chi^2(10)$ and $x_2 \sim \chi^2(13.978)$, and are assumed to be independent. This results in a positively skewed distribution of *Y* with a heavy tail. This example has been used by Liu et al. (2006) to demonstrate the limitation of variance-based sensitivity indices, where they propose a Kullback-Leibler (KL) divergence based metric:

$$KL_{T_i} = \int f_1(y(x_1,\ldots,\bar{x}_i,\ldots,x_d)) \ln \frac{f_1(y(x_1,\ldots,\bar{x}_i,\ldots,x_d))}{f_0(y(x_1,\ldots,x_i,\ldots,x_d))} dy \quad (6)$$

In Eq 6, $f_1(y)$ and $f_0(y)$ are the PDFs of the output, depending on whether $x_i$ is fixed, usually at its mean. The larger the $KL_{T_i}$, the more important $X_i$ is. It was found in Liu et al. (2006) that the effect of $X_1$ is higher in terms of divergence of the output distribution, but the variance-based total index shows that $X_1$ and $X_2$ are equally important. The higher influence of $X_1$ has also been confirmed in Pianosi and Wagener (2015) where the sensitivity is characterised by the change of cumulative distribution function of the output.

We reproduce the sensitivity results of $S_{T_i}$ and $KL_{T_i}$ in Table 1 from Liu et al. (2006). In addition, we also compare the results with the entropy-based total sensitivity index (ETSI) (Kala, 2021):

$$\eta_{T_i} = \frac{\mathbb{E}[H(Y \mid \mathbf{X}_{\sim i})]}{H(Y)} = \frac{H_{T_i}}{H_Y} \quad (7)$$

which measures the remaining entropy of *Y* if the true values of $\mathbf{X}_{\sim i}$ can be determined, in analogy to the variance-based total effect index $S_{T_i}$. *H* is the differential entropy, that is:

$$H_Y = H(Y) = -\int f(y) \ln f(y) dy$$

and the conditional differential entropy is defined accordingly as:

$$\mathbb{E}[H(Y \mid \mathbf{X}_{\sim i})] = -\int f(y, \mathbf{x}_{\sim i}) \ln f(y \mid \mathbf{x}_{\sim i}) dy d\mathbf{x}_{\sim i}$$

where the integral is with respect to the support set of the random variables and where $\sim i$ indicates the index ranges from 1 to $d$ excluding $i$. The conditional PDF can in general be written as $f(y \mid \mathbf{x}_{\sim i}) = f(y, \mathbf{x}_{\sim i}) / f(y)$, except for cases where the differential entropy becomes infinite.

The total effect entropy $H_{T_i}$ is a global measure of uncertainty as the expectation is with respect to all possible values of $\mathbf{X}_{\sim i}$. $\eta_{T_i} \leq 1$ because $H_{T_i} = \mathbb{E}[H(Y \mid \mathbf{X}_{\sim i})] \leq H(Y)$ with equality if and only if $\mathbf{X}_{\sim i}$ and $Y$ are independent (Cover, 1999). However, $\eta_{T_i}$ can be negative as it is defined using differential entropy. This is undesirable as sensitivity indices and later in Section 3 we propose an exponentiation to overcome this issue.

$\eta_{T_i}$ has been estimated numerically based on the histogram method given in *Supplementary Material: S1*, where $10^7$ samples are used. Table 1 shows that the entropy-based $\eta_{T_i}$ is able to effectively identify the higher influence of $X_1$. Note that different from $KL_{T_i}$ which is conditional on the value of $x_i$ ($x_i$ are set at their mean values in Table 1), $\eta_{T_i}$ is an un-conditional sensitivity measure as all possible values of the inputs are averaged out.

We note in passing that analogously to the variance based sensitivity indices, a first order entropy index can also be defined as $\eta_i = (H(Y) - \mathbb{E}[H(Y \mid X_i)]) / H(Y) = I(X_i, Y) / H(Y)$ (Krzykacz-Hausmann, 2001). $I(X_i, Y)$ is the mutual information which measures how much knowing $X_i$ reduces uncertainty of $Y$ or vice versa. The index $\eta_i$ can thus be regarded as a measure of the expected reduction in the entropy of of $Y$ by fixing $X_i$.

## 2.3 Summary

From the motivating example, it became clear that $S_{T_i}$ might not be very indicative for variable rankings with outputs of general distribution shapes. This is especially the case for highly skewed or multi-modal distributions. This limitation is overcome by the entropy-based measures which are applicable independent of the underlying shape of the distribution.

However, the entropy-based ETSI from Eq 7 has limited application in practice, mainly due to the heavy computational burden where the knowledge of conditional probability distributions are required. Both histogram and kernel based estimation methods have computational challenges for entropy-based sensitivity indices (Pianosi and Wagener, 2015).

Motivated by the above issues and inspired by the low-cost sensitivity screening proxy for variance-based measures, in the next section, we will propose a computationally efficient upper bound for the entropy-based total sensitivity measure. We then extend the upper bound to the DGSM indices via exponentiation, and show that in the special case with Gaussian inputs and linear functions, the proposed new proxy for entropy GSA is equivalent to the proxy for variance-based total effect sensitivity.

An overview of the above mentioned relationship between entropy and variance proxies is summarised in Figure 1. The sensitivity proxies are represented by the upper bounds of the inequalities in Figure 1, where $l_i$ and $v_i$ are derivative-based quantities. The entropy proxies to be developed in the next sections are highlighted in the box with dash lines.

# 3 Link between the partial derivatives and the total effect entropy measure

In this section we will consider a differentiable function $y = g(\mathbf{x})$. Recall that our interest here is for applications of global sensitivity analysis (GSA) towards mathematical models, where a physical phenomenon is typically studied with a complex numerical code. The computation of the partial derivatives can then be obtained via the companion adjoint code, or numerically estimated by a finite difference method. For example, the derivative-based DGSM can be estimated by finite difference, and this can be performed efficiently via Monte Carlo sampling as discussed in Kucherenko et al. (2009).

## 3.1 An upper bound for the total effect entropy

For a general vector transformation $\mathbf{Y} = \mathbf{g}(\mathbf{X})$, the differential entropy of the output is related to the input via (Papoulis and Pillai, 2002, p.660):

$$H(\mathbf{Y}) \quad \leq H(\mathbf{X}) + \int f_X(\mathbf{x}) \ln \left|\det \mathbb{J}\right| d\mathbf{x} \quad (8)$$

where $\mathbb{J}$ is the Jacobian matrix with $\mathbb{J}_{ij} = \partial g_i / \partial x_j$ and $f_X(\mathbf{x})$ is the probability density function (PDF) of $\mathbf{X}$. $\mathbf{g}(\mathbf{X})$ is assumed to be differentiable and the partial derivatives are assumed to be square integrable. The above inequality becomes an equality if the transform is a bijection, i.e. an invertible transformation. Note that there is no independence assumption for the inputs of the inequality above.

As shown in Papoulis and Pillai (2002), Eq 8 can be proved substituting the transformed PDF $f_Y(\mathbf{y}) = f_X(\mathbf{x}) / \det \mathbb{J}$ into the expression of $H(\mathbf{Y}) = -\int f(\mathbf{y}) \ln f(\mathbf{y}) d\mathbf{y}$ and note there will be a reduction of entropy if the transformation is not one-to-one. Following this line of thought, Eq 8 can also be seen as one version of the data processing inequality, where the transformation does not increase information (Geiger and Kubin, 2011).

Given the data processing inequality in Eq 8, we have the following theorem to upper bound the total effect entropy:

**Theorem** .

*For a differentiable deterministic function* $y = g(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}$ *with continuous random inputs, there exists an inequality for the total effect entropy:*

$H_{T_i} \leq H(X_i) + l_i$ (9)

*where* $H_{T_i}$ *is the total effect entropy which is an expected conditional entropy* $H_{T_i} = \mathbb{E}[H(Y \mid \mathbf{X}_{\sim i})]$, *where* $\sim i$ *indicates the index ranges from 1 to d excluding i.* $H(X_i)$ *is the differential entropy of the input variable* $X_i$ *and* $l_i$ *is the expected log-derivatives* $l_i = \mathbb{E}\left[\ln\left|\partial g(\mathbf{x}) / \partial x_i\right|\right]$.

**Proof.**

Set $y_1 = g_1(\mathbf{x}) = g(\mathbf{x})$ where $\mathbf{x} = (x_1, x_2, \ldots, x_d)$, and introduce dummy variables $y_i = g_i(\mathbf{x}) = x_i$ with $i = 2, \ldots, d$. In this setting, the Jacobian matrix from the 2nd row onwards, i.e. $i \geq 2$, $\partial g_i / \partial x_j = 1$ when $i = j$ and $\partial g_i / \partial x_j = 0$ when $i \neq j$. Therefore, the Jacobian matrix in this case is a triangular matrix. As a result, the Jacobian determinant is the product of the diagonal entries:

$$\det \mathbb{J} = \left| \frac{\partial g_1}{\partial x_1} \times \underbrace{1 \times \ldots \times 1}_{2\,\mathrm{to}\,d} \right| = \left| \frac{\partial g_1}{\partial x_1} \right|$$

The information processing inequality from Eq 8 can thus be expressed as:

$$H(\mathbf{Y}) \leq H(\mathbf{X}) + \int f_X(\mathbf{x}) \ln \left| \frac{\partial g_1(\mathbf{x})}{\partial x_1} \right| d\mathbf{x}$$

where $\mathbf{Y} = \{Y, X_2, X_3, \ldots, X_d\}$.

On the left hand side of the above inequality, the joint entropy of $\mathbf{Y}$ can be expressed using the conditional entropies as $H(\mathbf{Y}) = \mathbb{E}[H(Y \mid \mathbf{X}_{\sim 1})] + H(\mathbf{X}_{\sim 1})$ using the chain rule for differential entropies (Cover, 1999). On the right hand side, we have $H(\mathbf{X}) \leq H(X_1) + H(\mathbf{X}_{\sim 1})$ using the subadditivity property of the joint entropy of the input variables. The joint entropy $H(\mathbf{X})$ becomes additive if the input variables are independent.

Putting these together, the above inequality using the variable $x_1$ then becomes:

$$\mathbb{E}[H(Y \mid \mathbf{X}_{\sim 1})] \leq H(X_1) + l_1$$

The reasoning above uses the first variable $x_1$ as an example. However, the results hold for any variables via simple row/column exchanges, which only affects the sign of the determinant but not its modulus. ▪

The total effect entropy $H_{T_i} = \mathbb{E}[H(Y \mid \mathbf{X}_{\sim i})]$ is a global measure of uncertainty as the expectation is with respect to all possible values of $\mathbf{X}_{\sim i}$. It measures the remaining entropy of *Y* if the true values of $\mathbf{X}_{\sim i}$ can be determined, in analogy to the total effect variance $V_{T_i}$.

The total effect entropy inequality from Eq 9 thus demonstrates that, for a differentiable function $y = g(\mathbf{x})$, the entropy-based total sensitivity is bounded by the expectation of log partial derivatives of the function, with the addition of the entropy of the input variable of interest. And this inequality becomes an equality if the input variables are independent and the transformation $g(\cdot)$ has a unique inverse.

The inequality in Eq 9 is one of the main contributions of this paper. It establishes an upper bound for the total effect entropy $H_{T_i}$ for ETSI given in Eq 7 using computationally efficient partial-derivative based functionals. As smaller $l_i + H(X_i)$ tends to indicate smaller total effect entropy, it can thus be used to screen un-influential variables and work as a low cost proxy for entropy-based indices.

## 3.2 Exponential entropy based total sensitivity measure

The use of $l_i + H(X_i)$ as a screening proxy for the total effect entropy is similar to the DGSM-based upper bound for the variance-based $S_{T_i}$ described in Eq 5. However, there are two issues with the differential entropy based sensitivity measure: 1) differential entropy can become negative and this is undesirable for sensitivity analysis (Kala, 2021). More importantly, the inequality in Eq 9 is not valid when normalised by a negative $H(Y)$; 2) the interpretation of conditional entropy is not as intuitive as variance based sensitivity indices. This is partly due to the fact that variance-based methods is firmly anchored in variance decomposition, but also because entropy measures the average information or non-uniformity of a distribution as compared to variance which measures the spread of data around the mean. Although non-uniformity can be seen as a suitable measure for epistemic uncertainties (Krzykacz-Hausmann, 2001), its interpretation for GSA in a general setting is less intuitive.

To overcome these two issues, we propose to use exponential entropy as an entropy-based measure for global sensitivity analysis. Although not directly investigated, studies in Auder and Iooss (2008) have noted that an exponentiation of the standard entropy-based sensitivity measures may improve its discrimination power.

We take an exponentiation of the total effect entropy inequality in Eq 9:

$$e^{H_{T_i}} \le e^{H(X_i)} e^{l_i} \quad (10)$$

where we recall that $l_i = \mathbb{E}\left[\ln\left|\partial g(\mathbf{x}) / \partial x_i\right|\right]$. Divide both sides of Eq 10 by $e^{H(Y)}$:

$$\kappa_{T_i} = \frac{e^{H_{T_i}}}{e^{H(Y)}} \le \frac{e^{H(X_i)}}{e^{H(Y)}} e^{l_i} \quad (11)$$

where $\kappa_{T_i}$ can be considered as the exponential entropy based total sensitivity indices (eETSI), and the upper bound can then be used as a proxy for $\kappa_{T_i}$ to detect less influential input variables. As the total effect entropy $H_{T_i} = \mathbb{E}[H(Y \mid \mathbf{X}_{\sim i})] \le H(Y)$, we then have $0 < \kappa_{T_i} \le 1$ which is desirable as sensitivity indices.

eETSI $\kappa_{T_i}$, and its un-normalised $e^{H_{T_i}}$, have a more intuitive interpretation as GSA indices as compared to ETSI, because exponential entropy can be seen as a measure for the effective spread or extent of a distribution (Campbell, 1966). As the total effect entropy $H_{T_i}$ measures the remaining entropy in average if the true values of $\mathbf{X}_{\sim i}$ can be determined, $e^{H_{T_i}}$ can thus be regarded as the effective remaining range of the output distribution conditioning on that $\mathbf{X}_{\sim i}$ are known. The normalised indices eETSI $\kappa_{T_i}$ then measure the ratio of the effective range before and after $\mathbf{X}_{\sim i}$ are fixed, and larger $\kappa_{T_i}$ thus indicate a higher influence of $X_i$. We provide additional discussion on its intuitive interpretation in *Supplementary Material: S3* with several concrete examples.

In addition to its non-negativity and a more intuitive interpretation for GSA, exponential entropy is also closely linked to variance-based GSA indices and their corresponding bounds. To demonstrate this, we first note that the three different derivative-based sensitivity indices are closely related as:

$$e^{l_i} \le \mu_i \le \sqrt{\nu_i} \quad (12)$$

where we recall $l_i = \mathbb{E}\left[\ln\left|\partial g(\mathbf{x}) / \partial x_i\right|\right]$, $\mu_i = \mathbb{E}\left[\left|\partial g(\mathbf{x}) / \partial x_i\right|\right]$ and $\nu_i = \mathbb{E}\left[\left|\partial g(\mathbf{x}) / \partial x_i\right|^2\right]$.

It is evident that $\mu_i \le \sqrt{\nu_i}$ based on Cauchy-Schwarz inequality. In addition, we have $e^{l_i} \le \mu_i$ using Jensen's inequality as the exponential function is convex. So the inequality for the exponential entropy based eETSI from Eq 11 can be further associated with DGSM as:

$$\kappa_{T_i}^2 = \frac{e^{2H_{T_i}}}{e^{2H(Y)}} \le \frac{e^{2H(X_i)}}{e^{2H(Y)}} \nu_i \quad (13)$$

where we recall that $\nu_i$ are the derivative-based DGSM indices.

Eq 13 already looks remarkably similar to the variance-DGSM inequality given in Eq 5. In the special case with independent Gaussian inputs and a linear function, the $\nu_i$-based entropy upper bound from Eq 13 is equivalent to the $\nu_i$-based variance upper bound relationship given in Eq 5. See *Supplementary Material: S4* for additional discussion.

# 4 Numerical illustrations

In this section, we first examine the special equality case with monotonic functions, and then provide assessment of the total effect entropy inequality with general nonlinear functions. A physical example will be considered in Section 5. Note that although the inequality in Eq 9 makes no assumptions of independence, for simplicity the input variables are assumed to be independent in these numerical examples.

Additional illustrations using a randomised meta-function is given in *Supplementary Material: S6*, where the derivative-based proxy provides the same input variable ranking as the entropy-based indices for about three-quarters of the 1000 random functions tested. The upper bound (UB) can also be adjusted to work with groups of input variables, and this is illustrated briefly in *Supplementary Material: S7*.

## 4.1 Analytical verifications for equality cases with monotonic functions

In this section, we verify analytically that the total effect entropy inequality from Eq 9 is tight for a monotonic function. A function $g:\mathbb{R}^d \to \mathbb{R}$ is monotonically increasing if $x_i \le x'_i$ for all $i$ implies $g(\mathbf{x}) \le g(\mathbf{x}')$. Based on the analytical results, we present numerical experiments to investigate the required sample size for the tight bound.

Five monotonic functions are considered. These are listed in Table 2 and Figure 2 shows plots of examples 1 - 3 which are non-decreasing in the domain of interest. All input variables are assumed to have the same uniform distribution for examples 1 - 4, i.e. $x_i \sim \mathbb{U}(0,1)$, while Gaussian distributions are used for example 5. For verification purposes, all the examples in this section are chosen to have tractable expressions for both the integral of derivatives and the conditional entropies.

From table 2 we can see that $H_{T_i} = H(X_i) + l_i$ for the monotonic examples considered. These analytical results not only verify that the inequality from Eq 9 is tight for monotonic functions with independent inputs, but also provide benchmark for convergence test of numerical estimations.

For examples 1 - 3, the total effect entropies $H_{T_i}$ are also numerically estimated using the method given in *Supplementary Material: S1*. We estimate $H_{T_i}$ with Monte Carlo sampling, with number of samples ranging from $10^3$ to $10^8$ as shown in Figure 3. Also shown are the standard deviations (std) from 10 repeated estimations and the analytical values from Table 2.

It can be seen from Figure 3 that the estimation of $H_{T_i}$ converges to the exact values with increasing number of samples, and the relative error with $10^8$ samples is less than 1% for all functions. However, large number of samples is required.

In comparison, a smaller number of samples is sufficient for the estimation of derivative-based sensitivity measures. Numerical error of $l_i$ with 100 samples is less than 1% for the low dimensional cases considered. Note that the number of samples required is case dependent and a typical cost of derivative-based indices is in the order $10^3$ or $10^4$ (Kucherenko et al., 2009). For the estimation of $l_i$, the finite difference method for approximating the partial derivative is used with a fixed increment step of $10^{-5}$ following Kucherenko et al. (2009) for DGSM estimation. Note that the numerical results for $l_i$ are not shown in Figure 3 as it is indistinguishable from the analytical values in comparison to $H_{T_i}$.

## 4.2 Illustrations with Ishigami function and G-function

In this section, we use Ishigami function (Figure 4a) and G-function (Figure 4b) for illustrations with general nonlinear functions. Both functions are commonly used test functions for global sensitivity analysis, due to the presence of strong interactions.

### 4.2.1 Assessment of total effect entropy inequality

These two functions, each with three input variables, are used in this section to demonstrate the inequality relationship derived in Eq 9. The conditional entropies are estimated numerically using Monte Carlo sampling as described in *Supplementary Material: S1*. Different numbers of samples are used, ranging from $1e6$ to $1e8$. For each estimation, the computation is repeated 20 times and both the mean value and the standard deviation (std) are reported in Table 3 and 4. For the estimation of derivative-based $l_i$, we use the finite-difference based approach with 1000 samples. As the analytical expressions of the partial derivatives are readily available in this case, the derivative-based results are further verified with direct integration using Matlab's inbuilt numerical integrator "integral" with default tolerance setting.

The sensitivity results for the Ishigami function are listed in Table 3, where it is clear that inequality from Eq 9 is satisfied. It is clear the standard deviation is small. However, the convergence of conditional entropy estimation is slow as large number of samples is needed, as we saw for monotonic functions in Figure 3.

The sensitivity results for the G-function, in the same format as Table 3, are reported in Table 4. It is clear that the inequality relationship in Eq 9 is also satisfied. The results in Table 4 also highlights the issue that the differential entropy based results can be negative. The relative amplitudes of $H_{T_i}$ can still indicate the relative importance of the input variables for the output entropy, but the negative amplitudes are undesirable for sensitivity analysis.

### 4.2.2 Ranking with eETSI

In this section, we take the exponentiation of the total effect entropy, and discuss the exponential entropy based total sensitivity index (eETSI) $\kappa_{T_i}$ and its $l_i$-based upper bound (UB). $\kappa_{T_i}$ are obtained using the mean values of the total effect entropy $H_{T_i}$ from Table 3 and 4 with $10^8$ samples, and the corresponding $e^{H(Y)}$ for the output.

The results of the sensitivity indices are shown in Figure 5 for both Ishigami function and G-function. It can be seen that $l_i$-based upper bounds provide the same variable ranking as $\kappa_{T_i}$, although the bound can be loose for these non-linear functions.

In addition, we have also calculated the variance based $S_{T_i}$, using $10^5$ samples with 20 repetitions, where the mean values are shown in Figure 5. It can be seen that the results from $\kappa_i$ are generally consistent with $S_{T_i}$, especially for G-function where the sensitivity ranking are similar both qualitatively and quantitatively.

For the Ishigami function, both indices have successfully identified the contribution of $X_3$ which is the lowest. However, the relative importance of $X_1$ and $X_2$ are opposite from $\kappa_{T_i}$ and $S_{T_i}$. We explain in the *Supplementary Material: S5* that the interaction between $x_1$ and $x_3$ is more influential for variance due to the squaring effect, as compared to the entropy operation which takes logarithm of the interaction. This difference increases towards the boundary as the interaction between $x_1$ and $x_3$ gets stronger towards $-\pi$ and $\pi$. And this helps to explain why $x_1$ is the most influential variable for the variance-based $S_{T_i}$. This example highlights that, despite many similarities, entropy and variance are fundamentally different, for example, the variable interactions are processed differently between them. Note that the difference between variance-based and entropy-based ranking for Ishigami function was also noted in Auder and Iooss (2008).

# 5 A flood model case study

The numerical examples in the previous section have demonstrated that the log-derivative $l_i$ based upper bound can be potentially used as a screening proxy for entropy-based sensitivity indices, but only limited types of input uncertainty distributions were considered.

To demonstrate for practical problems with a wide range of input distributions, a simple river flood physics model is considered. This model has been used by Lamboni et al. (2013) and Roustant et al. (2017) for demonstration of the use of Poincaré inequality for factor prioritization with DGSM, and as an example in GSA review (Iooss and Lemaître, 2015).

This model simulates the height of a river, and flooding occurs when the river height exceeds the height of a dyke that protects industrial facilities. It is based on simplification of the 1D hydro-dynamical equations of SaintVenant under the assumptions of uniform and constant flow rate and large rectangular sections. The quantity of interest in this case is the maximal annual overflow *Y*:

$$Y = Z_v + D_m - D_d - C_b \quad \text{with} \quad D_m = \left( \frac{Q}{BK_s\sqrt{(Z_m - Z_v)/L}} \right)^{0.6} \quad (14)$$

where the distributions of the independent input variables are listed in Table 5. We have also added the exponential entropy $e^{H(X_i)}$ value for each variable in Table 5. The analytical expressions for differential entropy of most probability distributions are readily available and well documented. A comprehensive list can be found in Lazo and Rathie (1978).

For a truncated distribution with the interval *[a, b]*, its differential entropy can be found as: (Moharana and Kayal, 2020):

$$H_{\text{truncated}}(X) = -\int_a^b \frac{f(x)}{\Delta F} \ln \frac{f(x)}{\Delta F} dx \quad (15)$$

where $f(x)$ and $F(x)$ are the original probability density function (PDF) and cumulative distribution function (CDF) of the random variable *X* respectively. $f(x)/\Delta F$ is the PDF of the truncated distribution where $\Delta F = F(b) - F(a)$. As both $f(x)$ and $F(x)$ are analytically known, Eq 15 can then integrated numerically for the truncated entropy.

The results for the sensitivity indices are shown in Figure 6, where the variance-based $S_{T_i}$ are directly obtained from Table 5 of Lamboni et al. (2013). According to Lamboni et al. (2013), the variance-based index was based on $2 \times 10^7$ model evaluations. The total effect variance indices $S_{T_i}$ have upper bounds that are proportional to the derivative-based DGSM $\nu_i$. The $\nu_i$-variance upper bounds (UB) are also shown in Figure 6, where the optimal Poincaré constants are used as the proportional constants (Roustant et al., 2017).

In comparison, the exponential entropy based total effect sensitivity indices (eETSI) $\kappa_{T_i}$ are also shown in Figure 6, together with the derivative-based upper bounds for eETSI. It can be seen that, similar to the upper bounds for $S_{T_i}$, both $l_i$ and $\nu_i$ based entropy upper bounds are relatively close to $\kappa_{T_i}$, thus providing an efficient proxy for entropy-based total effect sensitivity indices. In this case, we can see that four input variables, $Q, K_s, Z_v, D_d$, have been identified as the most important variables for maximal annual overflow, with *L* and *B* of negligible influence. And this conclusion is consistent from both variance-based $S_{T_i}$ and entropy-based $\kappa_{T_i}$, and their upper bounds.

Note that as it is computationally difficult to estimate $\kappa_{T_i}$ accurately for this eight dimensional problems, the four least influential variables $Z_m, C_b, L, B$ have been set at their mean values, thus reducing the problem to four dimensions, for the estimation of $\kappa_{T_i}$. Both derivative-based measures have been estimated using all set of input random variables, using the finite difference method with 1000 samples and a fixed increment step of $10^{-5}$.

From this physics example, we can see that the derivative-based upper bounds can be used as a proxy for the total effect entropy sensitivity analysis for models with a variety of input distributions. Similar to the variance case where the optimal Poincaré constants can be estimated numerically using the R sensitivity package (Da Veiga et al., 2021), it is also straightforward to calculate the normalization constant $e^{H(X_i)}$ for entropy-bounds. Better still, $H(X_i)$ can often be analytically computed as entropy of many distribution functions are known in closed-form. The main computational cost for estimation of the total effect sensitivity proxies is thus the calculation of the partial derivatives, which is much more affordable than a direct estimation of the conditional variance or conditional entropy.

# 6 Conclusions

A novel global sensitivity proxy for entropy-based total effect has been developed in this paper. We have made use of the inequality between the entropy of the model output and its inputs, which can be seen as an instance of data processing inequality, and established an upper bound for the total effect entropy. This upper bound is tight for monotonic functions. It also provides similar input rankings for about three quarters of the 1000 random functions and can thus be regarded as a proxy for entropy-based total effect measure. Applied to a simplified flood analysis, the new proxy shows good ranking capability for physics problems with a variety of input distributions.

The resulting log-derivative $l_i$-based proxy is computationally cheap to estimate. If the the derivatives are available, e.g. as the output of a computational code, the proxy would be readily available. Even if a Monte Carlo based approach is used, the computational cost is typically in the order $10^3$ or $10^4$, as compared to $10^7$ or $10^8$ for entropy-based indices. This computational advantage would be even more important for high dimensional problems. However, the numerical assessment has been limited to low dimensional examples, because it becomes prohibitively inefficient to compute conditional entropies with the Monte Carlo based histogram approach adopted in this paper. In subsequent works, we will explore neural network based density estimation techniques for efficient approximation of information-theoretic quantities and extend the numerical assessment of the upper bound to high dimensions.

In many physical applications, the input variables often have a dependence structure due to physical constraints. Unlike DGSM-based proxy, the inequality link between derivatives and entropy presented in this paper does not require the random inputs to be independent. This point is subject to further research, where numerical assessment needs to conducted to explore the screening power of $l_i$-based upper bound for dependent inputs.

Drawing on the criticism of differential entropy based sensitivity indices, we propose to use its natural exponentiation and the resulting sensitivity measure $\kappa_{T_i}$ possesses many desirable properties for GSA, such as quantitative, moment independent and easy to interpret. The G-function example shows that $\sum \kappa_{T_i}$ is close to one for this product function, as opposed to the variance-based indices where the sum of sensitivity indices is equal to one for additive functions. As exponential entropy can be seen as a geometric mean of the underlying distribution, i.e. $e^{H(X)} = e^{\mathbb{E}[-\ln f(x)]}$, one of the future research is to examine the unique properties of GSA indices based on exponential entropy, and explore its decomposition characteristics for sensitivity analysis of different interaction orders.

# Supplementary Material

**Supplementary Manuscript.**

The supplementary materials contain eight sections: S1 for numerical estimation of entropy, S2 for analytical derivations for monotonic examples, S3 and S4 on interpretation of exponential entropy and its link to variance for sensitivity analysis, S5 for interpretation of Ishigami results, S8 for flood model data; S6 provides additional numerical illustrations of the upper bound using a randomised meta-function, while S7 provides examples with application to group input variables.

**Supplementary Code:**

Matlab codes to reproduce Figure 3.

## Acknowledgement

The authors are grateful to Dr Sergei Kucherenko from Imperial College London and Dr Marco De Angelis from University of Strathclyde for comments on an early draft of this article, and Dr Bertrand Iooss from EDF and Dr Eric Hall from University of Dundee for useful discussions and reviewing of this work.

## Data availability statement

The authors confirm that the data supporting the findings of this study are available within the article.

## Disclosure Statement

The authors report there are no competing interests to declare.

## References

Auder, B. and B. Iooss (2008). Global sensitivity analysis based on entropy. *Proceedings of the ESREL 2008 Conference*, 2107–2115.

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

Borgonovo, E. and E. Plischke (2016). Sensitivity analysis: A review of recent advances. *European Journal of Operational Research 248*(3), 869–887.

Campbell, L. L. (1966). Exponential entropy as a measure of extent of a distribution. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 5*(3), 217–225.

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.

Constantine, P. G. and P. Diaz (2017). Global sensitivity metrics from active subspaces. *Reliability Engineering & System Safety 162*, 1–13.

Cover, T. M. (1999). *Elements of information theory*. John Wiley & Sons.

Da Veiga, S. (2015). Global sensitivity analysis with dependence measures. *Journal of Statistical Computation and Simulation 85*(7), 1283–1305.

Da Veiga, S. (2021). Kernel-based anova decomposition and shapley effects–application to global sensitivity analysis. *arXiv preprint arXiv:2101.05487*.

Da Veiga, S., F. Gamboa, B. Iooss, and C. Prieur (2021). *Basics and Trends in Sensitivity Analysis*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Geiger, B. C. and G. Kubin (2011). On the information loss in memoryless systems: The multivariate case. *arXiv preprint arXiv:1109.4856*.

Hoeffding, W. (1948). A Class of Statistics with Asymptotically Normal Distribution. *The Annals of Mathematical Statistics 19*(3), 293 – 325.

Iooss, B. and P. Lemaître (2015). A review on global sensitivity analysis methods. *Uncertainty management in simulation-optimization of complex systems: algorithms and applications*, 101–122.

Kala, Z. (2021). Global sensitivity analysis based on entropy: From differential entropy to alternative measures. *Entropy 23*(6), 778.

Krzykacz-Hausmann, B. (2001). Epistemic sensitivity analysis based on the concept of entropy. *Proceedings of SAMO2001*, 31–35.

Kucherenko, S., M. Rodriguez-Fernandez, C. Pantelides, and N. Shah (2009). Monte carlo evaluation of derivative-based global sensitivity measures. *Reliability Engineering & System Safety 94*(7), 1135–1148.

Lamboni, M., B. Iooss, A.-L. Popelin, and F. Gamboa (2013). Derivative-based global sensitivity measures: General links with sobol' indices and numerical tests. *Mathematics and Computers in Simulation 87*, 45–54.

Lazo, A. V. and P. Rathie (1978). On the entropy of continuous probability distributions (corresp.). *IEEE Transactions on Information Theory 24*(1), 120–122.

Liu, H., W. Chen, and A. Sudjianto (2006). Relative entropy based method for probabilistic sensitivity analysis in engineering design. *Journal of Mechanical Design 128*(2), 326–336.

Moharana, R. and S. Kayal (2020). Properties of shannon entropy for double truncated random variables and its applications. *Journal of Statistical Theory and Applications 19*(2), 261–273.

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics 33*(2), 161–174.

Papoulis, A. and S. Pillai (2002). *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill series in electrical and computer engineering. McGraw-Hill.

Pianosi, F. and T. Wagener (2015). A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environmental Modelling & Software 67*, 1–11.

Puy, A., W. Becker, S. L. Piano, and A. Saltelli (2020). The battle of total-order sensitivity estimators. *arXiv preprint arXiv:2009.01147*.

Roustant, O., F. Barthe, and B. Iooss (2017). Poincaré inequalities on intervals – application to sensitivity analysis. *Electronic Journal of Statistics 11*(2), 3081 – 3119.

Saltelli, A. (2008). *Global sensitivity analysis: the primer*. John Wiley.

Soboĺ, I. (1993). Sensitivity estimates for nonlinear mathematical models. *Math. Model. Comput. Exp. 1*, 407.

Sobol', I. 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.

Wainwright, H. M., S. Finsterle, Y. Jung, Q. Zhou, and J. T. Birkholzer (2014). Making sense of global sensitivity analyses. *Computers & Geosciences 65*, 84–94. TOUGH Symposium 2012.

Yang, J. (2023). A general framework for probabilistic sensitivity analysis with respect to distribution parameters. *Probabilistic Engineering Mechanics 72*, 103433.

Yang, J. (2024). Decision-oriented two-parameter fisher information sensitivity using symplectic decomposition. *Technometrics 66*(1), 28–39.

Zahm, O., P. G. Constantine, C. Prieur, and Y. M. Marzouk (2020). Gradient-based dimension reduction of multivariate vector-valued functions. *SIAM Journal on Scientific Computing 42*(1), A534–A558.

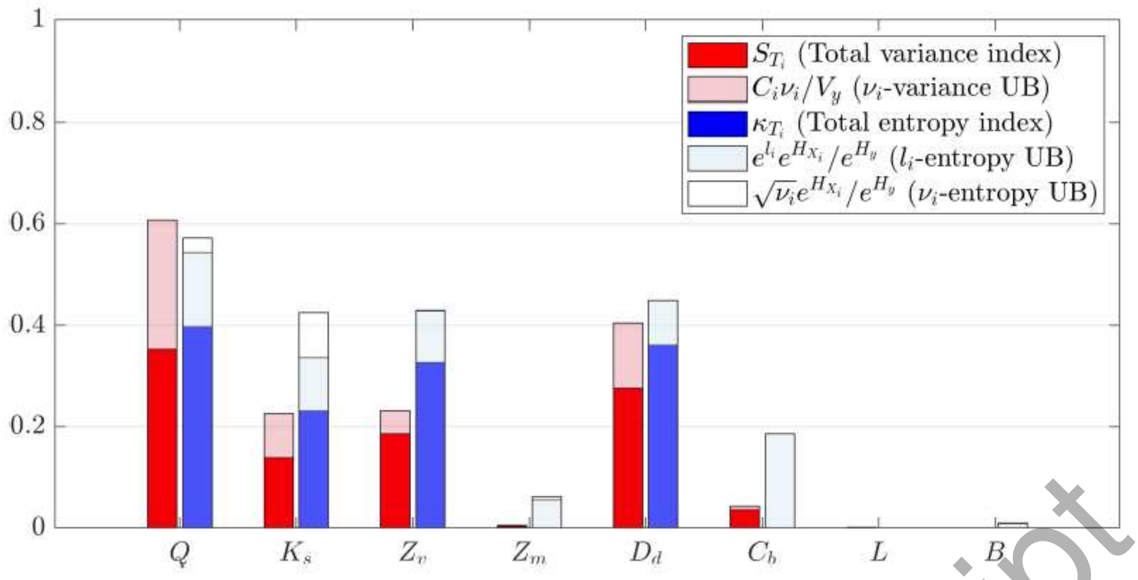



Figure 1: An overview of the relationship between entropy and variance proxies where the entropy proxies developed in this paper are highlighted in the box with dash lines.

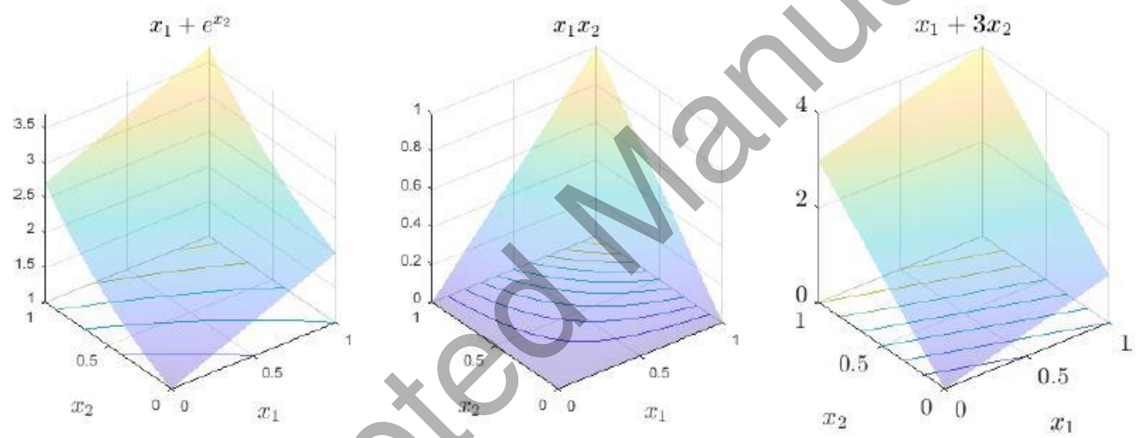



Figure 2: Surface plots for the monotonic functions in examples 1 – 3

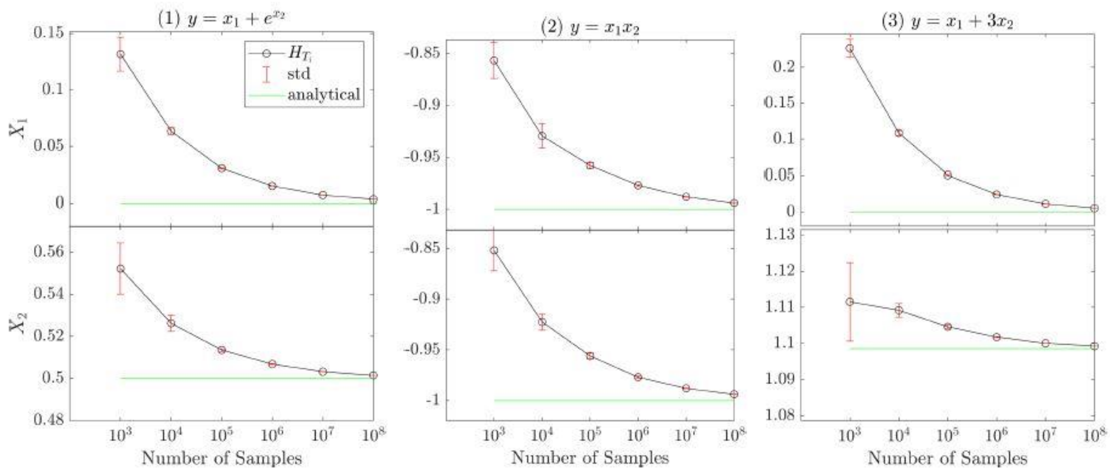
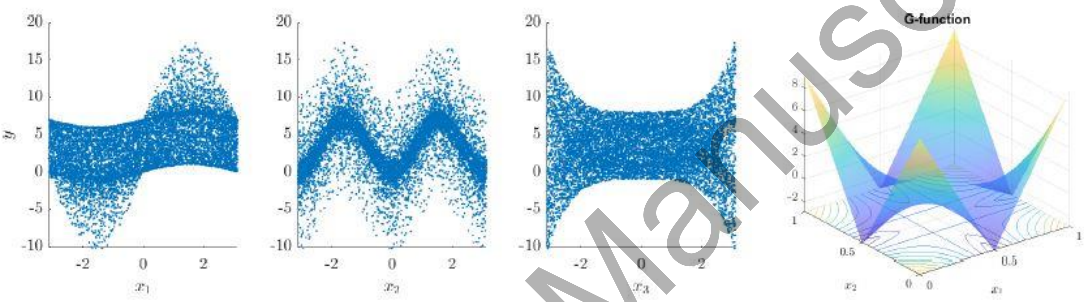
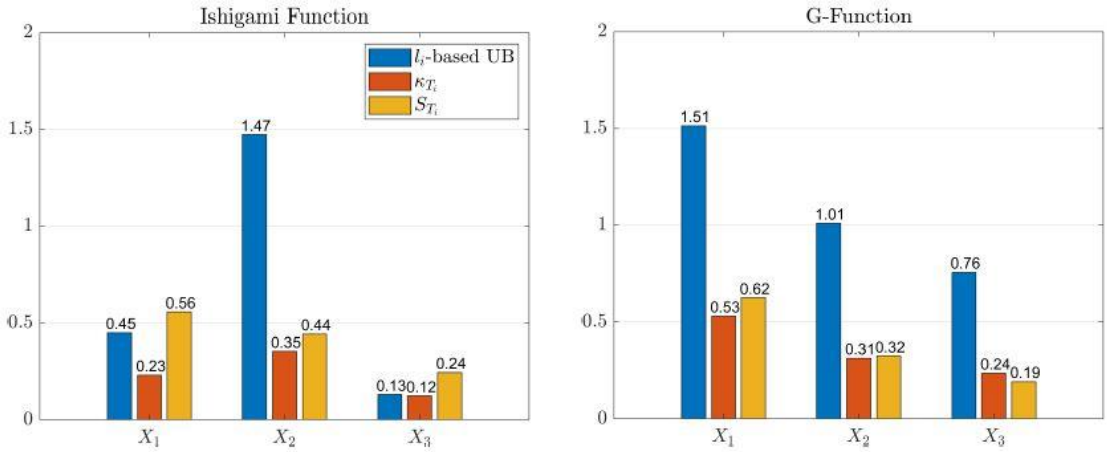



Figure 3: Convergence of the numerical estimated total effect entropy for monotonic examples 1 - 3.




(a) Ishigami function (b) G-function

Figure 4: Example plot for the general functions considered. a) Scatter plot for the Ishigami function, $y = \sin(x_1) + 7\sin^2(x_2) + 0.1x_3^4 \sin(x_1)$, $x_i \sim \mathbb{U}(-\pi, \pi)$ for $i = 1,2,3$. ; b) Surface plots for the G-function (2 variable plot), $y = \prod_{i=1}^{3} (|4x_i - 2| + a_i)/(1 + a_i)$, $x_i \sim \mathbb{U}(0,1)$ for $i = 1,2,3$. In this case, $a_i = (i-2)/2$, for $i = 1,2,3$. A lower value of $a_i$ indicates a higher importance of the input variable $x_i$, i.e., $x_1$ is the most important, while $x_3$ is the least important in this case




(a) Ishigami function

(b) G-function

Figure 5: Sensitivity indices for Ishigami function and G-function

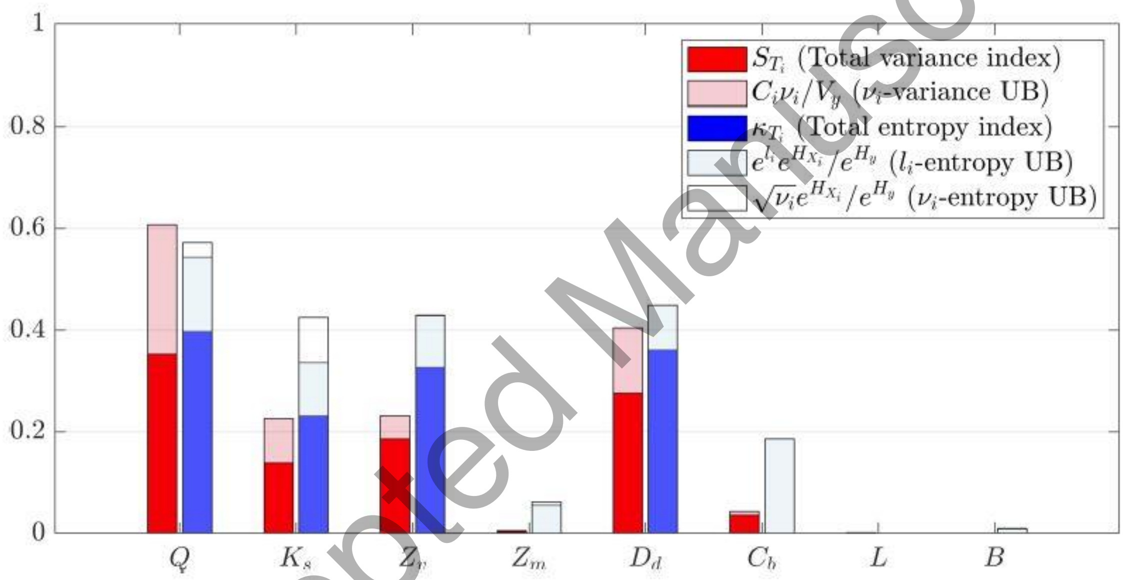



Figure 6: Total effect sensitivity indices for the flood model, from both variance-based $S_{T_i}$ and entropy-based $\kappa_{T_i}$ and their upper bounds (UB). Recall $l_i$ is the log-derivative sensitivity measure and $\nu_i$ is DGSM. Note that for entropy upper bound, we use $\sqrt{\nu_i}$ for a direct comparison with $\kappa_{T_i}$ as seen from Eq 13. Numerical data of the results presented here can be found in *Supplementary Material: S8*.

Table 1: Sensitivity results for $y = x_1 / x_2$.

| | Variance-based | K-L divergence based | Entropy-based |
|---|---|---|---|
| Variable | $S_{T_i}$ | $KL_{T_i}$ | $\eta_{T_i}$ |
| $x_1 \sim \chi^2(10)$ | 0.546 | 0.1571 | 0.510 |
| $x_2 \sim \chi^2(13.978)$ | 0.547 | 0.0791 | 0.213 |

Table 2: Analytical results for five different monotonic functions, where derivations are given in *Supplementary Material: S2*. Note that $H_{T_i} = \mathbb{E}[H(Y \mid \mathbf{X}_{\sim i})]$ is the total effect entropy and $l_i = \mathbb{E}\left[\ln\left|\partial g(\mathbf{x}) / \partial x_i\right|\right]$

| | | $X_1$ | | $X_2$ | |
|---|---|---|---|---|---|
| Examples | | $H_{T_1}$ | $H(X_1)+l_1$ | $H_{T_2}$ | $H(X_2)+l_2$ |
| Ex-1 | $y = x_1 + e^{x_2}$ | 0 | 0 | 1/2 | 1/2 |
| Ex-2 | $y = x_1 x_2$ | -1 | -1 | -1 | -1 |
| Ex-3 | $y = x_1 + 3x_2$ | 0 | 0 | $\ln 3$ | $\ln 3$ |
| Ex-4 | $y = x_1 x_2^r$ | $-r$ | $-r$ | $\ln r - r$ | $\ln r - r$ |
| Ex-5 | $y = \sum_{i=1}^{d} a_i x_i$ | $\ln\lvert a_i\rvert$ | $\ln\lvert a_i\rvert$ | for all *i* | |

Table 3: Total effect entropy results for the Ishigami function, which are obtained for different number of samples. This is repeated for 20 times and the mean and standard deviation (std) are given. The results from $10^8$ samples are compared to $H(X_i)+l_i$, for which the inequality in Eq 9 is clearly satisfied.

| | Ishigami function | | | | | |
|---|---|---|---|---|---|---|
| | $y=\sin(x_1)+7\sin^2(x_2)+0.1x_3^4\sin(x_1)$ | | | | | |
| Number of Samples | $H_{T_1}$ | | $H_{T_2}$ | | $H_{T_3}$ | |
| | mean | std | mean | std | mean | std |
| 1.00E+06 | 1.3902 | 0.0007 | 1.7614 | 0.0006 | 0.9701 | 0.0013 |
| 1.00E+07 | 1.2978 | 0.0003 | 1.7023 | 0.0001 | 0.7693 | 0.0004 |
| 1.00E+08 | 1.2335 | 0.0001 | 1.6609 | 0.0001 | 0.6066 | 0.0002 |
| | $X_1$ | $X_2$ | $X_3$ | | | |
| $H_{T_i}=\mathbb{E}[H(Y\mid\mathbf{X}_{\sim i})]$ | 1.2335 | 1.6609 | 0.6066 | | | |
| $H(X_i)+l_i$ | 1.9024 | 3.0906 | 0.6626 | | | |

Table 4: Sensitivity results for the G-function, where the inequality in Eq 9 is clearly satisfied. Same key as Table 3

| | G-function $y=\prod_{i=1}^{3}\frac{\lvert 4x_i-2\rvert+a_i}{1+a_i}$ with $a_i=(i-2)/2$ | | | | | |
|---|---|---|---|---|---|---|
| Number of Samples | $H_{T_1}$ | | $H_{T_2}$ | | $H_{T_3}$ | |
| | mean | std | mean | std | mean | std |
| 1.00E+06 | 0.3477 | 0.0009 | -0.1376 | 0.0013 | -0.3988 | 0.0015 |
| 1.00E+07 | 0.3398 | 0.0006 | -0.1737 | 0.0005 | -0.4482 | 0.0006 |
| 1.00E+08 | 0.3378 | 0.0003 | -0.1917 | 0.0002 | -0.4738 | 0.0002 |
| | $X_1$ | $X_2$ | $X_3$ | | | |
| $H_{T_i}=\mathbb{E}[H(Y\mid\mathbf{X}_{\sim i})]$ | 0.3378 | -0.1917 | -0.4738 | | | |
| $H(X_i)+l_i$ | 1.3863 | 0.9808 | 0.6931 | | | |

Table 5: Entropy and distribution of the flood model input variables

| | Variable | Description | Distribution Function | Exponential Entropy $e^{H(X_i)}$ |
|---|---|---|---|---|
| $x_1$ | *Q* | Maximal annual flowrate [m$^3$/s] | Truncated Gumbel $\mathcal{G}$ (1013,558) on [500,3000] | 2051 |
| $x_2$ | $K_s$ | Strickler coefficient [ - ] | Truncated Normal $\mathcal{N}$ (30, $8^2$) on [15, $\infty$] | 30 |
| $x_3$ | $Z_v$ | River downstream level [m] | Triangular $\mathcal{T}$ (49,50,51) | 1.65 |
| $x_3$ | $Z_m$ | River upstream level [m] | Triangular $\mathcal{T}$ (54,55,56) | 1.65 |
| $x_5$ | $D_d$ | Dyke height [m] | Uniform $\mathcal{U}$ [7,9] | 2 |
| $x_6$ | $C_b$ | Bank level [m] | Triangular $\mathcal{T}$ (55,55.5,56) | 0.825 |
| $x_7$ | *L* | Length of the river stretch [m] | Triangular $\mathcal{T}$ (4990,5000,5010) | 16.5 |
| $x_8$ | *B* | River width [m] | Triangular $\mathcal{T}$ (295,300,305) | 8.24 |