Global Uncertainty and Sensitivity Analysis

Extended Fourier Amplitude Sensitivity test - eFAST

The Fourier amplitude sensitivity test (FAST) was first developed by Cukier *et al.*(1973, 1975, 1978), Schaibly and Shuler (1973) then improved to the extended FAST (eFAST) by Saltelli *et al.*(1999) to provide a way to estimate the expected value and variance of the model output and the contribution of parameter input to this variance.

Let us consider the model $y = f(\vec{x})$. The output y is determined through a model f using a set of K input parameters $\vec{x} = (x_1, x_2, \ldots, x_K)$.

The main idea behind eFAST is to 1) convert the k-dimensional function $f(\vec{x})$ into a one-dimensional sampling function $f(\vec{x}) = f(\vec{x}(s))$ by defining a sinusoidal function for each input parameter $x_i = G_i(\sin(\omega_i s))$ then 2) estimate the partial output variance (V_{x_i}) accounted by each parameter x_i by decomposing it into Fourier series given by:

$$V_{x_i} = 2\sum_{p=1}^{M}(A^2_{p\omega_i} + B^2_{p\omega_i}) \tag{1}$$

where the Fourier coefficients $A_{p\omega_i}$ and $B_{p\omega_i}$ are defined as:

$$\begin{aligned} A^2_{p\omega_i} &= \frac{1}{N_s}\sum_{k=1}^{N_s} f(s_k)\,\cos(p\omega_i s_k)\,ds \\ B^2_{p\omega_i} &= \frac{1}{N_s}\sum_{k=1}^{N_s} f(s_k)\,\sin(p\omega_i s_k)\,ds \end{aligned} \tag{2}$$

The angular frequency ω_i is used as a parameter identifier to compute partial variance associate with each parameter x_i and the computation can be estimated using N_s number of sampling points. Since the Fourier amplitude decreases as p increases, for simplicity, we approximate the Fourier series of the partial variance by considering first M harmonic components, usually taken to be 4. Lastly, eFAST algorithm will use the partial variance estimated in 1 to compute the *first order sensitivity index* ($S_i^{x_i}$) and total order sensitivity index ($S_{Ti}^{x_i}$) to quantify the first and second order interaction effects contribution of x_i to the output variance as following:

$$S_i^{x_i} = \frac{V_i}{V} \tag{3}$$

$$S_{Ti}^{x_i} = 1 - S_i^{\sim x_i} \tag{4}$$

where $V = \frac{\sum_{k=1}^{N_S}(f(s_k) - E(y))^2}{N_S}$ is the total variance of model outcome as the result of simultaneously varying all parameter inputs and $S_i^{\sim x_i}$ is the sum of all the first order sensitivity of input parameters that do not include x_i.

Sensitivity indices derivation were well documented in [1, 2]. For the sake of completeness, we just briefly review the procedure here.

Choice of search curve. The sinusoidal search curve G_i must provide a uniformly distributed sample for each factor x_i over parameter space as s varies. Among many functions that were proposed and studied in [1], the transformation introduced by Salteli(1999) has proved its prominence in accomplish such goal, which is defined as:

$$x_i(s) = \frac{1}{2} + \frac{1}{\pi}\arcsin(\sin(\omega_i s + \phi_i)) \tag{5}$$

where $\{\omega_i\}$ is a set of integer angular frequencies and ϕ_i is a random phase shift chosen uniformly in $[0, 2\pi)$.

The domain of $\sin(\omega_i s + \phi)$ is defined for every real number s and the range of $\sin(\omega_i s + \phi)$ is $[-1, 1]$ which lies within the domain of $\arcsin(\cdot)$ which is $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. So we conclude that the function 5 is defined for every value of s. For simplicity, we assume the domain of s to be $[-\pi, \pi]$.

The choice of function defined in 5 works for any x_i from a uniform distribution over domain $[0, 1]$. For a more general case where x_i is from a non-uniform distribution over a domain (x_i^{min}, x_i^{max}) the periodic search function defined in 5 can be built into the parameter's inverse cumulative distribution $(F_i^{-1}(\cdot))$ and the search function in 5 can be re-defined as:

$$x_i = F_i^{-1}\left(\frac{1}{2} + \frac{1}{\pi}\arcsin(\sin(\omega_i s + \phi_i))\right) \tag{6}$$

Choice of number of sampling points. Aliasing is an unwanted process when the model spectrum frequency that lies outside the range that we consider is

incorrectly aliased with the frequency within that range. To avoid the aliasing effects, by the Nyquist criteria, the minimum number of sampling points N_S is required to be [1]:

$$N_s = 2M\omega_{max} + 1 \tag{7}$$

where M is previously defined as the maximum harmonic we consider, usually taken to be 4 and $\omega_{max} = \max\{\omega_i\}_i^K$ as the maximum angular frequency of all the input parameters.

Due to symmetry properties of trigonometric functions, the sinusoidal function will eventually repeat the same samples. A re-sampling scheme is implemented to avoid this inefficiency [1] by introducing the random phase shift ϕ_i to the search curve G_i as seen in equation 5. Particularly, eFAST algorithm is repeated N_R (the re-sampling size) times with different search curves specified by ϕ_i. In the re-sampling scheme, the sample size is re-defined to be [1]:

$$N_s = (2M\omega_{max} + 1)N_R \tag{8}$$

Choice of frequency.

The fundamental frequency ω_i assigned to each parameter search function is critical as it is used to evaluate the partial variance of model outcome V_i as seen in expression 1 and 2. Not only the set of frequency $\{\omega_i\}_i^K$ must be less than the Nyquist critical frequency to avoid aliasing effects which was discussed previously but they also must be linearly independent up to the range that we consider to avoid interference effects, which is a problem when the information provided by the harmonic components of V_i and $V_{\sim i}$ are mixed. For example, in the case with only two parameter inputs x_1 and x_2 and their associated frequencies ω_1 and ω_2. There exists at least a combination of harmonic indices such that $p_1\omega_1 = p_2\omega_2$. The spectrum at frequency $p_1\omega_1$ then contains mixed information about the contributions due to both x_1 and x_2. However, the problem of choosing frequencies becomes more difficult as the number of parameter input increase in the model. Details in choosing frequencies for eFAST algorithm was well documented in [1, 2]. In summary, the algorithm in choosing frequency for a K-parameter model can be summarized below:

- A maximum frequency ω_{max} is chosen to be at least 8.
- In calculating the first order and total order sensitivity indice of parameter

x_i, assigning $\omega_i = \omega_{max}$ while all other frequencies can be chosen within the range $\left[1, \frac{\omega_{max}}{2M}\right]$ where M is the maximum harmonics that we consider.

- The step between two consecutive frequencies must be as large as possible
- The number of factors to which the same frequency is assigned must be as low as possible.

Since the minimum allowable frequency is 1, $\frac{\omega_{max}}{2M} \geq 1$, where M is usually chosen to be 4, this implies ω_{max} must be at least 8 and therefore the minimum number of sampling points must be at least 65 according to equation 7.

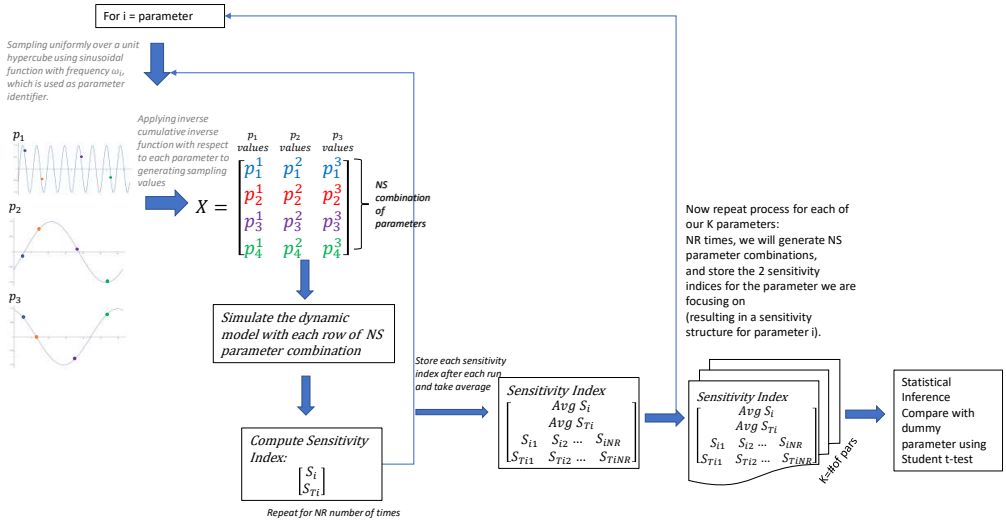

Figure 1: Diagram of eFAST algorithm.

References

[1] Andrea Saltelli, Stefano Tarantola, and KP-S Chan. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, 41(1):39–56, 1999.

[2] Andrea Saltelli, Karen Chan, M Scott, et al. Sensitivity analysis. probability and statistics series. *John and Wiley & Sons, New York*, 2000.