A Hierarchical Statistical Sensitivity Analysis Method for Complex Engineering Systems

Xiaolei Yin¹, Wei Chen²
Integrated DEsign Automation Laboratory (IDEAL)
Department of Mechanical Engineering
Northwestern University

ABSTRACT

The method of Statistical Sensitivity Analysis (SSA) is playing an increasingly important role in engineering design, especially with the consideration of uncertainty. However, applying SSA to the design of complex engineering systems is not straightforward due to both computational and organizational difficulties. In this paper, a Hierarchical Statistical Sensitivity Analysis (HSSA) method is developed to facilitate the application of SSA to the design of complex systems especially those follow hierarchical modeling structures. A top-down strategy for HSSA is introduced to only invoke the SSA of critical submodels based on the significance of submodel performances. A simplified formulation of the Global Statistical Sensitivity Index (GSSI) is studied to represent the effect of a lower-level submodel input on a higher-level model response by aggregating the submodel SSA results across intermediate levels. A sufficient condition under which the simplified formulation provides an accurate solution is derived. To improve the accuracy of the GSSI formulation for a general situation, a modified formulation is proposed by including an Adjustment Coefficient (AC) to capture the impact of the nonlinearities of the upper level models. To save cost, the evaluation of the AC shares the same set of samplings used in the submodel SSA. The proposed HSSA method is examined through mathematical examples and a 3-level hierarchical model used in vehicle suspension systems design.

1. INTRODUCTION

Complex engineering systems design often involves multiple disciplines, a large amount of physical elements, and coupled information exchanges among subsystems. When the decomposition strategy is applied to modeling [1], appropriate submodels are first developed, and later integrated in either a hierarchical or concurrent way to predict a system’s performance. Hierarchical modeling is becoming a common modeling strategy for studying complex engineering phenomena [2-4]. For instance, in Multiscale Material Modeling (MMM) [5, 6], a material is decomposed into several individual scales of interest and the macroscale material properties are evaluated by hierarchically integrating the information from the lowest (atomic) to the highest (macroscopic) scale. Such a hierarchical modeling structure can be extended to a product design domain [7]. Figure 1 shows a hierarchical material-product system chain for vehicle design, where the hierarchical modeling bridges not only the material domain but also the product domain from parts (or components), assemblies (or subsystems), to a final product at a sequence of increasing characterization sizes. Associated with multiscale hierarchical modeling, Multiscale Design [8, 9] is an emerging research area that studies the efficient utilization of information from multiscale models that may be associated with design explorations at different scales and disciplines. The growing need for hierarchical design approaches is also evident in the development of multilevel optimization methodology [10, 11] to support hierarchical decision making in a typical industrial organization. Various sources of design uncertainty are considered in the extensions of multilevel design methods for robust design and reliability-based design [12, 13].

Figure 1. A hierarchical material–product system chain.

Our interest in this work is to develop methods that can facilitate the use of Statistical Sensitivity Analysis (SSA) in the design of hierarchical engineering systems. Built upon statistical sampling techniques, SSA studies the impact of variations in model inputs on the variations in model outputs [14]. The existing SSA methods [15-17] can assist modelers to reduce the dimension or scale of an engineering design problem by eliminating unimportant variables, and to provide efficient

¹ Graduate Student
² Corresponding Author, Associate Professor, Department of Mechanical Engineering, Northwestern University, 2145 Sheridan Rd., Evanston, IL 60208-3111; weichen@northwestern.edu.
resources allocations based on the significance of uncertainty sources [18]. However, applying SSA to the design of complex engineering systems is not straightforward. For a complex system, it is difficult or even prohibitive to carry out SSA in an All-In-One (AIO) manner [19] due to the system’s complexity, limited communications between subsystems that belong to different disciplines, and the associated high cost in simulations. Local SSA results from submodels are not sufficient to represent the global impact of a submodel input on a system performance of interest. There is a need to derive the Global Statistical Sensitivity Index (GSSI) of any input variable based on the results from submodel SSAs. Even though methods have been developed for hierarchical sensitivity analysis in multidisciplinary design optimization [20-22], these methods are to find the rate of change in a model output with respect to a submodel input based on the partial derivatives at a given design point. On the contrary, the interest of SSA is to study the impact on a model response with respect to the variation of a model input over a range.

In this paper, we propose a Hierarchical Statistical Sensitivity Analysis (HSSA) method for use in designing complex hierarchical engineering systems. Figure 2 illustrates a 3-level hierarchical model with multiple submodels at different levels. The information flow in a hierarchical model follows a bottom-up sequence, i.e., each submodel performance is passed to a submodel at a higher level. The proposed HSSA method is aimed at understanding the importance of submodel analysis in assessing the system performance, providing a global importance ranking of design variables or uncertainty sources, and releasing the computational burden of the AIO SSA by only using the results from local submodel SSAs. A top-down strategy for HSSA is introduced merely to invoke the SSA of the critical submodels based on the significance of submodel performances. By combining the local SSA results from individual submodels, formulations of the GSSI are developed to provide a comprehensive measure of the global impact of a local input variable across the submodels at intermediate levels.

There are two challenges in aggregating local SSA results. One is associated with the top-down strategy of HSSA. When applying SSA to an upper level model at a prior stage, assumptions of the distributions of upper level input variables, which are the submodel performances passed from lower levels, need to be made. Once the information of the submodel performances is available after applying SSA to the lower level submodels, the discrepancies between the real and pre-assumed submodel performance distributions need to be taken into account. The other challenge is that unlike the derivative type which input variables are ranked based on their contributions to the variance of a performance [15, 17, 25]. In our proposed Hierarchical Statistical Sensitivity Analysis (HSSA) method, the variance-based SSA methods, in particular, the Sobol’s method [27] is employed for local SSA.

In variance-based SSA, the variance of a performance (or an output) \( Y = f(X) \) is expressed in an ANOVA-like way [28]. That is, the total variance \( V \) of an output \( Y \) is decomposed into items contributed by various sources of input variations:

\[
V = \sum_{i} V_{i} + \sum_{ij} V_{ij} + \cdots + V_{12...n},
\]

where the first order term \( V_{i} \) represents the partial variance in the response due to the individual effect of a random variable \( X_{i} \), while the higher order terms represent the interaction effects between two or more random variables.

The Statistical Sensitivity Index (SSI) is defined in [15] as:

\[
SI_{i} = \frac{V_{i}}{V}, \quad SI_{ij} = \frac{V_{ij}}{V}, \text{ and so on.}
\]

Note that all the sensitivity indices are nonnegative and their sum is equal to 1. \( SI_{i} \) is the Main Sensitivity Index (MSI) of \( X_{i} \) that measures the main effect of \( X_{i} \). The Total Sensitivity Index (TSI) is defined as the sum of all the sensitivity indices that involve the input variable of interest, including both the individual effect of a variable as well as the effect caused by its interactions with other variables. To calculate the SSI, a square-integrable function \( Y = f(X) \) with independent input variables \( X \) is decomposed into an ANOVA formulation:

\[
f(X) = f_{0} + \sum f_{i}(X_{i}) + \sum f_{ij}(X_{i}, X_{j}) + \cdots + f_{12...n}(X_{1}, X_{2}, \ldots, X_{n}),
\]

The terms in Eqn. (3) are calculated via the multidimensional integrations as shown in [27]. The total variance of \( Y \) and partial contributions are calculated as:

![Figure 2. A demonstrative example of a hierarchical model.](image)
\[ V = \int f^2(X) \prod_{i=1}^{n} \rho_i dX_i - f_0^2, \]  
(4)

\[ V_{\text{cum}} = \int f^2(X) \rho_1 \cdots \rho_n dX_1 \cdots dX_n, \]  
(5)

where \( \rho_i \) is the Probability Density Function (PDF) of \( X_i \). Many efficient Monte Carlo methods are developed [27, 29, 30] to calculate the SSIs numerically. When the simulation model is computationally expensive, metamodels such as polynomial functions and Kriging models can be built first, and sensitivity indices can be derived using semi-analytical methods [28].

3. THE HSSA METHOD

3.1 Top-Down Strategy

In this paper, a top-down strategy for applying SSA to the design of hierarchical systems is presented. The top-down strategy contains three features:

a) Instead of performing SSA to a complex ‘All-In-One’ (AIO) model, SSA is applied separately, step by step, to submodels at each level of hierarchy, which allows independent or parallel executions of SSA on submodels.

b) Opposite to the bottom-up sequence of the information flow in a hierarchical model, HSSA is applied first to the top level model with the final output of interest following the top-down sequence.

c) To save cost, SSA is applied only to critical lower level submodels whose performances have a significant impact on the performance of the upper level model.

Take the hierarchical model shown in Figure 2 as an example, SSA is applied first to the top level model, Model A, whose input variables are the performance responses passed from Submodels B and C at the 2nd level. If the performance of Submodel C is found to be much more significant than the one from Submodel B, then SSA is applied only to Submodel C. Similarly, the local SSA results from Submodel C will indicate whether further SSAs are needed for Submodel E or F.

3.2 Aggregation of Local SSA Results

By applying the top-down strategy, Local Statistical Sensitivity Index (LSSI) results are obtained via the SSA study on submodels. Following that, the global impact of any submodel input variable on a final output of interest needs to be derived. In this paper, an aggregation approach is developed to combine the local SSA results to provide a global importance ranking of an input variable (or uncertain source). For demonstration purpose, a 2-level hierarchical model with one upper level model and \( N \) independent lower level submodels is considered (Figure 3). The same principle and aggregation formation can be extended to an n-level hierarchical system.

![Figure 3. A generalized two-level hierarchical model.](image)

![Figure 4. Flowchart of the HSSA method.](image)
3.3 IS Technique for Evaluating Posterior LSSI

Importance Sampling (IS) technique [31-33] is employed in Step III of the proposed HSSA method to evaluate the posterior LSSI of the upper level model given the real distribution of the submodel performance. The original idea of the importance sampling technique is to sample points subject to a prior Probability Density Function (PDF) instead of a posterior PDF so that sampling points can be concentrated in parts of the integration domain of most ‘importance’ that is determined by the prior PDF. Following the idea of importance sampling, the integral of an arbitrary integrable function in terms of a submodel performance from the lower level, \( f(Y) \), with a posterior PDF can be rewritten in terms of a prior PDF:

\[
\int f(Y) \rho_{pr}^Y dY = \int f(Y) \frac{\rho_{pr}^Y}{\rho_{st}^Y} \rho_{st}^Y dY \approx \frac{1}{M} \sum_{i=1}^{M} f(Y^i) \frac{\rho_{pr}^Y(Y^i)}{\rho_{st}^Y(Y^i)},
\]

where \( \rho_{pr}^Y \) and \( \rho_{st}^Y \) are the posterior and prior PDFs of submodel performance \( Y \), respectively; \( Y^i \) are sampled subject to the prior distribution. \( M \) is the number of sampling points. The prior PDF, \( \rho_{st}^Y \), is preset for \( Y^i \) in the SSA of the upper level model, while the posterior \( \rho_{pr}^Y \) is the real distribution of \( Y^i \) obtained from the lower level submodel SSA. If the prior PDF is preset as a uniform distribution whose density equal to a positive constant \( P \), then the integral can be approximated as a weighted average with the weights equal to the corresponding posterior probability densities at each sampling point but divided by \( P \):

\[
\int f(Y) \rho_{pr}^Y dY \approx \frac{1}{M} \sum_{i=1}^{M} f(Y^i) \frac{\rho_{pr}^Y(Y^i)}{P}.
\]

Note that Eqns. (6) and (7) can be easily extended to an \( N \)-dimensional space that includes all the submodel performances \( Y \). The posterior LSSI can be evaluated using the prior and posterior distributions information of \( Y^i \) and the sampling information used for calculating the prior LSSI. In practice, the prior PDF of submodel performance, \( \rho_{st}^Y \), is chosen to the best of our knowledge. To provide valuable posterior local SSA results, the range of \( \rho_{st}^Y \) should encompass as much as possible the real performance range \( \rho_{pr}^Y \). A larger range of \( \rho_{pr}^Y \) is more desirable than a smaller range under estimation.

3.4 Formulations of GSSI

Once local SSA results are available from both the upper and lower levels, they are aggregated to compute the Global Statistical Sensitivity Index (GSSI) of a local input variable with respect to its impact on a top level performance. As the aggregations of interaction effects and total effects are mathematically difficult, in this work, we only focus on studying the GSSI for the main effect. The main effects have been found to be the most dominating effects in typical engineering systems [34, 35]. A Simplified Formulation (SF) of the GSSI for the main effect of a local input variable \( X_i \) that belongs to Submodel \( i \) is defined as:

\[
SL_i^Y = SL_{i,0}^Y = SI_{X_i}^Y SL_{X_i}^Y, \quad \text{for } X_i \in X_i^r,
\]

where \( SI_{X_i}^Y \) is the LSSI of \( X_i \) that quantifies the contribution of \( X_i \) to the variance of the submodel performance \( Y_i \); and \( SL_{X_i}^Y \) is the upper level LSSI that measures the impact of the variation of \( Y_i \) on the final output \( Z \). In the following, we first define the condition under which the SF in Eqn. (8) can provide an accurate estimation of the GSSI and then propose an approach that uses a simple correction coefficient when the condition is not satisfied.

**Theorem:** For a model with a 2-level hierarchical structure as shown in Figure 3, if the upper level function \( h(X, Y) \) is linear with respect to \( Y_i \) i.e.,

\[
h(X, Y) = S(X, Y_i) + T(X, Y_i),
\]

where \( Y_i \) is the vector of submodel performances excluding \( Y_{i_1} \); \( S(X, Y_i) \) and \( T(X, Y_i) \) are any integrable functions in terms of \( X \) and \( Y_i \), then the GSSI for the main effect of \( X_i \) can be calculated as a product of the LSSI of \( Y_i \) from the upper level and the LSSI of \( X_i \) from the lower level.

The related proof is presented in Appendix A. Note that with the above condition, the upper level submodel function \( h \) can still be nonlinear with respect to other submodel performances \( Y_i \) but only need to be linear with respect to \( Y_i \). A simple example is \( Z = X_i + X_i^2 \cdot Y + 2Y_i \cdot Y_i \), which is linear with respect to either \( Y_i \) or \( Y_i^2 \). The GSSI of the main effect of any local input variable related to either \( Y_i \) or \( Y_i^2 \) can be accurately estimated using the SF. In addition, when the above condition is satisfied, for a subset of local input variables that belong to the same Submodel \( i \), i.e., \( U_i \subset X_i \), the GSSI for their combined effect, defined as the sum of main effects and interaction effects of all the variables from the subset \( U_i \), can also be aggregated using Eqn. (8), i.e.,

\[
SI_{e_i}^Y = SI_{X_i}^Z SI_{X_i}^Y.
\]

When the upper level model is nonlinear with respect to the linking submodel performances, it becomes a challenge to provide an accurate measure of the GSSI using the SF, because the nonlinearity will introduce interactions of lower level input variables after being propagated. For example, consider a simple upper level function \( Z = X_i + Y_i^2 \) and a lower level submodel \( Y = X_i + X_i^2 \). After plugging in \( Y \) into \( Z \), \( Z \) becomes \( Z = X_i + X_i^2 + 2X_i^2X_i + X_i^2 \), which introduces the interactions between \( X_i \) and \( X_i^2 \). To overcome this difficulty, a coefficient, named Adjustment Coefficient (AC), is introduced to approximate the accumulated nonlinear impacts of the submodel performance at the upper level. The Modified Formulation (MF) of the GSSI for main effects is expressed as:

\[
SI_{X_i}^Z = AC \cdot SI_{X_i}^Z SI_{X_i}^Z.
\]

In this paper, \( AC \) is defined as:

\[
AC = \frac{B_i^Z Y_i}{Y_i^Z},
\]
where \( V_y \) is the variance of the submodel performance of interest \( Y \); \( V_y^z \) stands for the variance contribution of \( Y \) to the variance of the upper level output \( Z \); \( B \) captures the global linear trend of the upper level output \( Z \) with respect to \( Y \). The derivation of Eqn. (11) can be found in Appendix B. When \( Z \) is exactly a linear function with respect to \( Y \), \( AC \) is equal to 1 no matter what the distribution of \( Y \) is, and the MF (Eqn. (10)) is degenerated into the SF (Eqn. (8)).

To capture the linear trend of the upper level function, a weighted linear regression [36] based on the set of samplings used for the prior upper level SSA is employed to provide a linear approximation of \( h_1(Y) \) in terms of \( Y \) as:

\[
h_1(Y) = \int h(X, Y) \rho_X \rho_Y dX dY = A + B \cdot Y , \quad (12)
\]

where constants \( A \) and linear coefficient \( B \) are obtained by using the weighted linear regression:

\[
\beta = (\xi^T W \xi)^{-1} \xi^T W Z , \quad (13)
\]

where the superscript \( T \) means a transpose of a matrix; \( \beta = [A, B] \); \( \xi = [1, Y] \), which is a matrix composed by a column vector of 1 and all the samples of submodel performance \( Y \); \( Z \) is a vector of the upper level outputs. \( W = \text{diag} \{ W_1, \ldots, W_n \} \) and each component of \( W \) is the probability density of \( Y \) at each sampling point \( Y^i \), i.e., \( W_i = \rho_X(Y^i) \) for \( k = 1, \ldots, M \). The purpose for introducing the probability density as the weight in the linear regression is to place more emphasis on the region where the submodel performance most likely occurs.

To better understand the MF, consider again the above example in which \( Z = X_1 + Y^2 \) and \( Y = X_2 + X_3 \). All \( X_i \) are subject to a uniform distribution over [0.0, 1.0]. When using the SF to evaluate the contribution of \( X_2 \) to \( Z \), we get \( SL_{x_2} = SL_{y^2} \approx 0.989 \times 0.5 = 0.449 \). Based on the All-In-One (AIO) propagated function \( Z = X_1 + X_2^2 + 2X_3X_4 + X_5^2 \), we find that the actual contribution of \( X_2 \) should be \( SL_{x_2} = 0.429 \). The discrepancy results from the fact that there exists a global interaction effect of \( X_2 \) and \( X_3 \) due to the nonlinearity of \( Z \) with respect to \( Y \). When using our proposed MF, all the samples that are in the 3-dimensional sampling space \( (X_1, Y, Z) \) are projected into the \( Y-Z \) space. \( h_1(Y) \) is linearized as \( h_1(Y) \approx -0.331 + 2.0 \cdot Y \) and \( AC \) is derived as 0.936 based on Eqn. (11). In this case, the \( AC \) introduces a reduction (< 1.0) to the magnitude of the SF solution, resulting \( SL_{y^2} = 0.936 \times 0.449 = 0.420 \), which is now very close to the correct solution (=0.429).

4. EXAMPLES

In this section, the proposed Hierarchical Statistical Sensitivity Analysis (HSSA) method is applied to a 2-level mathematical model and a 3-level hierarchical engineering model used in vehicle suspension systems design to examine the effectiveness and the accuracy of the proposed method. The Statistical Sensitivity Analysis (SSA) results via using the ‘All-In-One’ (AIO) method are treated as the ‘exact’ solutions for reference. Results from the Simplified Formulation (SF) (Eqn. (8)) and the Modified Formulation (MF) (Eqn. (10)) of the Global Statistical Sensitivity Index (GSSI) are examined with a comparison to the ‘exact’ solution.

4.1 A Mathematical Example

In the first example, a 2-level hierarchical model with only one submodel at the lower level is defined in Eqn. (14):

\[
Z = h(X_1, Y) = 2X_1 + 3X_1^2 + 2X_3 + 2X_4
\]

where the upper level model \( h \) has two input variables: a local input \( X_1 \) and a submodel performance \( Y \); \( Y \) is passed from the submodel \( g \) at the lower level which has three local input variables \{\( X_2, X_3, X_4 \}\). All \( X_i \) are assumed to be independent and uniformly distributed over [0.0, 1.0]. \( n \) is a parameter that can be controlled to change the nonlinearity of the response \( Z \) with respect to \( Y \). To study the impact of the nonlinearity on the accuracy of the SF and MF, we examine several cases in which \( n \) is assigned various values from 0.5 to 8.0.

In the upper level local SSA, the prior distribution of \( Y \) is assumed to be uniform. Its real distribution is later found to be close to a normal distribution \( N(\mu = 4.75, \sigma^2 = 2.97) \). The number of samples for SSA is set as 1.0e+5 to reduce the numerical error caused by the Monte Carlo estimation. The mean and variance of \( Z \) are obtained using the Importance Sampling (IS) technique as discussed in Section 3.3. To verify the proposed IS approach, the IS results using different pre-assumed sampling ranges of \( Y \) at the upper level are compared with one using the AIO method in Table 1 for different \( n \). It is observed that when the preset range is too small, e.g., [2.0, 8.0], which is smaller than [0.0, 10.0] that covers \pm 3\sigma of the real distribution, the importance sampling technique provides a poor estimation of the mean and variance of \( Z \) since the tails regions of the real distribution \( Y \) are not sampled. When the prior distribution of \( Y \) has a range that fully encompasses its real distribution range, e.g., [0.0 10.0] and [-5.0, 15.0], the mean and variance of \( Z \) can be accurately estimated with the IS approach. It is therefore recommended to consider a sufficiently large range of submodel performance in the prior SSA.

### Table 1. Estimation error of statistical moments of \( Z \) using the importance sampling method on various sampling ranges.

<table>
<thead>
<tr>
<th>Error %</th>
<th>Mean of ( Y ), ( \mu_y )</th>
<th>Variance of ( Y ), ( \sigma_y^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>[2.0, 8.0]</td>
<td>[0.0, 10.0]</td>
</tr>
<tr>
<td>1.0</td>
<td>[2.0, 8.0]</td>
<td>[0.0, 10.0]</td>
</tr>
<tr>
<td>4.0</td>
<td>[2.0, 8.0]</td>
<td>[0.0, 10.0]</td>
</tr>
<tr>
<td>8.0</td>
<td>[2.0, 8.0]</td>
<td>[0.0, 10.0]</td>
</tr>
</tbody>
</table>

Figure 5 illustrates the GSSIs for the main effects of the four input variables using the proposed HSSA method and the AIO method. The \( y \)-axis indicates the magnitude of GSSI. As shown in the top-right plot of Figure 5, in the case of \( n \) equal to 1.0, the SF and MF provide identical GSSI results, which coincide with the ones from the AIO method. It is noted that the nonlinearity of the upper level model with respect to the
submodel performance $Y$ has a direct impact on the importance rankings of input variables. Compared to the input variable $X_1$ at the upper level, the input variables at the lower level ($X_5$, $X_6$, $X_7$) have increasingly important contributions to the variance of $Z$ when the nonlinearity becomes higher. Both the SF and MF can capture the trend that $X_5$, $X_6$ and $X_7$ become increasingly important versus $X_1$. The contribution of $X_1$ significantly reduces, also due to the fact that the interaction effects of lower level inputs become larger because of the increasing nonlinearity of $Z$ versus $Y$ in the upper level model. For this example, the GSSI for main effects is sufficient to measure the importance of input variables since the importance ranking based on the main effects is identical to the one based on the total effects for various cases. It is also noted that both the SF and MF provide the same ranking of all four input variables as the AIO method except for $n = 4.0$, in which the MF provides the correct ranking, but the SF does not.

In comparison of the SF and MF results, it is noted that the MF provides a more accurate estimation of the GSSI especially when the magnitude of the sensitivity indices using the SF significantly deviates from the real AIO value (see $n = 4.0$ and $8.0$). The reason is that the Adjustment Coefficient ($AC$) introduced in the MF captures a global trend of the nonlinearity of the top level model with a consideration of the real distribution of the submodel performance. The prior and posterior distributions of $Y$ have a direct impact on the LSSI values at the top level and the regression coefficients ($A$ and $B$) as well. A comparison of those two distributions and their impact on the linear regression is illustrated in Figure 6 in the case of $n$ equal to 4.0. As a result, once lower level information is available, posterior LSSIs and the $AC$ based on the real distribution should be used to evaluate the GSSI.

### 4.2 An Enterprise-Driven Multilevel Vehicle Suspension Model

The enterprise-driven multilevel vehicle suspension model was originally presented in [37] to demonstrate the use of a decision-based multilevel optimization formulation in designing hierarchical systems. The 3-level hierarchical vehicle suspension model is modified slightly here by fixing the front and rear linear coil spring stiffnesses ($K_{LF}$ and $K_{LR}$) to ensure the submodels at the same level are independent (see Figure 7). The hierarchical model contains a vehicle profit model for a domestic mid-sized vehicle at the top level, two independent suspension models for the front and rear wheels respectively at the middle level, and the front and rear coil spring models at the bottom level. The local input variables to the coil spring models are the wire diameter $d$, and coil diameter $D$. The coil spring bending stiffness is calculated by the following equation:

$$
K_s = \frac{EGd^4}{16D(2G + E)},
$$

where $E$ and $G$ characterize the rigidity of the spring material. More details can be found in [37]. When applying the proposed HSSA method, all the models are treated as a black-box type of functions. All design variables are subject to uniform distributions as $d_f \in [10, 25]$, $d_r \in [10, 25]$, $D_f \in [100, 140]$, $D_r \in [100, 140]$, $L_{sf} \in [350, 400]$ and $L_{sr} \in [350, 400]$. At a prior stage, uniform distributions are preset to the submodel performances as $K_{sf} \in [3580, 190000]$, $K_{sr} \in [3580, 190000]$, $K_{SF} \in [32, 60]$ and $K_{SR} \in [28, 58]$. In this example, the normalized local Total Sensitivity Indices (TSIs), defined as the original TSI of an individual variable divided by the sum of TSIs of all input variables, are used to compare the relative importance of submodels and submodel performances. Table 2 lists both the normalized local TSIs at the prior and posterior stages for different model/submodels. It is observed that at the bottom level, $d_f$ has a dominant impact on the variance of $K_{SF}$ since its normalized TSI is close to 100%. This observation can be confirmed by the model structure as observed in Eqn. (15) where $d_f$ has a high order ($= 4$) and $D_f$ has the lower order ($= -1$).

![Figure 6. Linear regression with the consideration of submodel performance. The linear regression at the prior stage is performed with a prior uniform distribution of the submodel](image-url)
Figure 7. Hierarchical enterprise-driven multilevel vehicle suspension model.

Table 2. Application process of the top-down strategy

<table>
<thead>
<tr>
<th>Level</th>
<th>Input</th>
<th>Prior Local TSI</th>
<th>Posterior Local TSI (I)</th>
<th>Posterior Local TSI (II)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top</td>
<td>Output: Profit</td>
<td>$K_{SF}$ 54%</td>
<td>47%</td>
<td>46%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$K_{SR}$ 46%</td>
<td>53%</td>
<td>54%</td>
</tr>
<tr>
<td>Middle</td>
<td>Output: $K_{SF}$</td>
<td>31%</td>
<td>32%</td>
<td>Integration of information from all three levels but without the rear coil model</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$K_{BF}$ 69%</td>
<td>68%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Output: $L_{OF}$</td>
<td>74%</td>
<td>74%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$K_{BR}$ 26%</td>
<td>26%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bottom</td>
<td>Output: $K_{BF}$</td>
<td>Integration of information from two levels (current level and the lower level)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$d_F$ 98%</td>
<td>98%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$D_F$ 2%</td>
<td>2%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Output: $K_{BR}$</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$d_R$ --</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$D_R$ --</td>
<td>--</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

When the top-down strategy is applied, SSA is first applied to the top level profit model. Based on the local SSA results, it is found that neither $K_{SF}$ nor $K_{SR}$ can be ignored since they have a similar impact quantified as 54% and 46% respectively. SSA is further applied to both the front and rear suspension models at the middle level. Once the middle level local SSAs are implemented, the posterior local sensitivity index of the top level model is calculated by combining the SSA information from both the top level and the middle level using the importance sampling technique. We find that the impact of $K_{SF}$ decreases to 47% and the impact of $K_{SR}$ slightly increases to 56%. Based on the local SSA results at the middle level, for illustration purposes, the submodel performance $K_{BR}$ from the bottom level is considered unimportant since its normalized TSI is much smaller than that of $L_{OF}$. Further SSA is applied only to the front coil spring model at the bottom level. Once the distribution information from the bottom level is obtained, the posterior local SSA results at the middle level can be calculated. As shown in Table 2, the posterior TSI for the rear suspension model is unchanged because that is the lowest level where SSA is invoked. Given the information from both the bottom and middle levels, the posterior SSA results for the top level profit model are now recalculated by combining information from all three levels. For both the profit model and the front suspension model, the minor changes in the posterior TSIs ((I) and (II)) compared to the prior TSIs indicate that the use of prior TSI in this case is sufficient to identify critical submodels. It should be noted that the prior distribution is determined by either former experience or any available information. To reduce the error in identifying critical submodels, a small threshold value should be used so that only the factors with a very small contribution to the upper level are ignored. Furthermore, the calculation of the posterior LSSI (Step III) allows users to update/correct the LSSIs at the upper level and confirm the identification once the information from lower level is obtained.

Based on the reduced hierarchical model that excludes the rear coil submodel, the GSSIs for main effects of submodel input variables using both the HSSA method and the AIO method are shown in Figure 8. Based on the HSSA results, the rear spring free length $L_{SR}$ and the front wire diameter $d_F$ are the two dominant input variables. The MF provides a more accurate GSSI result compared to the SF. It is illustrated in Figure 9 the evolvement of the distribution of the vehicle profit in the whole process of the HSSA method when more and more information is available. The distribution of the vehicle profit that combines the information from all three levels is similar to the one using AIO. With more information available during the application process of the HSSA method, we can achieve a more accurate estimation of both the GSSI and the distribution information of submodel performances.

Figure 8. Comparison of the GSSI for main effects using the HSSA method and the AIO method over the entire design space. The GSSI of $d_R$ and $D_F$ are assumed as zero since the rear coil spring model is considered as an uncritical model. The figure shows the GSSI of all model input variables with respect to their impact on the system response – profit.

In this example, we also apply the proposed HSSA method for uncertainty analysis to identify critical uncertainty sources associated with the six local input variables. All the six input variables are assumed to follow normal distributions with mean locations at the optimal solution $[\mu_F, \mu_{L_{OF}}, \mu_{d_F}, \mu_{D_F}, \mu_{K_{BF}}, \mu_{K_{BR}}] = [384.9, 417.2, 23.2, 180, 23.7, 180]$ and the corresponding coefficients of variance equal to $[0.01, 0.01, 0.002, 0.01, 0.002, 0.01]$. The submodel performances passed from the lower level
are preset as normal distributions with their first two moments estimated using the first order Taylor expansion. The GSSIs for the main effects of the six random input variables and their importance rankings via the SF, the MF and the AIO method are illustrated in Figure 10. We find that compared to the SF, the MF provides a more accurate estimation of the magnitudes of the GSSIs as well as the importance ranking of uncertainty sources. We should point out that the importance ranking identified in uncertainty analysis is quite different from the earlier results where SSA is carried out for the entire design space. In general, the SSA results highly depend on the variation ranges chosen for each input variable.

Figure 9. Distribution of the top level response “profit” with different amount of information available from lower levels.

Figure 10. Comparison of the main effects of the uncertainty sources associated with the six local input variables. The numbers above the GSSI bars indicate the importance rankings from the most important (marked as 1) to the least important (marked as 6). The SF switches the ranking order of the first two important variables $D_F$ and $L_{OF}$.

5. CONCLUSION

In this paper, a Hierarchical Statistical Sensitivity Analysis (HSSA) method is developed to facilitate the use of SSA in designing complex engineering systems with a hierarchical structure such as multiscale design and multilevel optimization. There are two major aspects of the proposed HSSA method: 1) a top-down strategy to identify all the critical submodels whose performances have a significant impact on the top level performance of interest; 2) a method to provide an approximation of the global impact of a submodel input variable on the final output of interest by aggregating local SSA results. The HSSA method can be used for both sensitivity analysis to find critical input variables within a specified design range, and for uncertainty analysis to identify critical uncertainty sources within the range of variation.

A top-down strategy for HSSA is proposed to apply SSA first on the top level model with the end performance of interest. The local SSA results at an upper level are used to identify critical submodels at the lower level and SSA is further applied to critical submodels. By applying the top-down strategy, SSA is executed only on selected submodels instead of the whole system and the cost of SSA on insignificant submodels can be saved.

To calculate the global impact of a local input variable on the final output of interest across intermediate levels, a Simplified Formulation (SF) of the Global Statistical Sensitivity Index (GSSI) for main effects is defined as a multiplication of the relevant local statistical sensitivity indices. We illustrate that the SF provides an accurate sensitivity index when the upper level model is linear with respect to the submodel performance that links the local input variable of interest. To take into account the nonlinearity of the upper level model with respect to the submodel performance, an Adjustment Coefficient (AC) is introduced in a Modified Formulation (MF) of the GSSI. Based on our empirical study, we find that the MF provides a better estimation of the GSSI for main effects compared to the SF.

To further reduce the computational cost, the sampling information for calculating the prior Local Statistical Sensitivity Index (LSSI) with pre-assumed distributions of submodel performances is re-used to calculate the posterior LSSI with the real distributions as well as to calculate the AC in the MF. Importance Sampling (IS) technology is employed to take into account the true distributions of submodel performances in calculating posterior LSSIs. When the number of samples for integral calculation is large enough and the pre-assumed distribution encompasses the range of the true distribution, IS can provide an accurate estimation for the integrals of variance contributions. In practice, it is recommended to use a sufficiently large pre-assumed range, rather than a small one, of submodel performance for upper level SSA to achieve the desired accuracy.

In conclusion, the proposed HSSA method enables the application of SSA to a complex hierarchical engineering system especially when it is infeasible to apply SSA in an ‘All-In-One’ manner. Due to the mathematical difficulties in combining local statistical sensitivity indices to provide a global sensitivity index for interaction effects, current work only focus on the estimation of main effects with an assumption that the main effect of an input variable is sufficient to represent its importance. Furthermore, the current HSSA method is applied only to hierarchical systems with independent submodels at each level. Correlations between
submodel performances and their impacts on the importance ranking are to be considered in future work.

ACKNOWLEDGEMENTS
We would like to thank Professor Harrison Kim at UIUC for helping with the hierarchical vehicle suspension system model, which was implemented as a part of his doctoral work at University of Michigan, the Optimal Design Laboratory. NSF grant support, DMII 0522662 is also acknowledged.

REFERENCES
Appendix A – Proof of Theorem in Section 3.4

Given the upper level model function equal to $S(X_s, Y_i) + T(X_s, Y_i) \cdot Y_i$, the theorem in Section 3.4 can be proved by recalling the definition of SSI (Eqsns. (2) and (5)):

$$h_i = \int h(X_s, Y_i) dY_i = \int [S(X_s, Y_i) + T(X_s, Y_i) \cdot Y_i] dY_i,$$

where $S = \int S(X_s, Y_i) dY_i$ and $T = \int T(X_s, Y_i) dY_i$.

For a general case, $A$ and $B$ are determined by the linear weighted regression presented in Section 3.4.

$$h_i(Y_i) = \int h(X_s, Y_i) dY_i = \int [S(X_s, Y_i) + T(X_s, Y_i) \cdot Y_i] dY_i - h_0.$$

So

$$St_i = \frac{V_x^2}{V_z} = \frac{\int S(X_s, Y_i) dY_i}{\int T(X_s, Y_i) dY_i}.$$

Let

$$f(Y_i, X_s) = S(X_s, Y_i) + T(X_s, Y_i) \cdot Y_i,$$

and $X_s \subset X_i$.

So

$$f_i(X_s) = \int f(Y_i, X_s) dY_i.$$

Hence

$$SI_i = \frac{V_x^2}{V_z} = AC \cdot SI_i,$$

where $SI_i$ is the mean of the submodel performance $Y_i$. When the upper level model function is linear with respect to $Y_i$, constants $A$ and $B$ in Eqn. (12) are equivalent to $\tilde{S}$ and $\tilde{T}$, respectively. For a general case, $A$ and $B$ are determined by the linear weighted regression presented in Section 3.4.

$$h_i(Y_i) = \int h(X_s, Y_i) dY_i = \int [S(X_s, Y_i) + T(X_s, Y_i) \cdot Y_i] dY_i - h_0.$$

So

$$St_i = \frac{V_x^2}{V_z} = \frac{\int S(X_s, Y_i) dY_i}{\int T(X_s, Y_i) dY_i}.$$

Let

$$f(Y_i, X_s) = S(X_s, Y_i) + T(X_s, Y_i) \cdot Y_i,$$

and $X_s \subset X_i$.

So

$$f_i(X_s) = \int f(Y_i, X_s) dY_i.$$

Hence

$$SI_i = \frac{V_x^2}{V_z} = AC \cdot SI_i,$$

where $SI_i$ is the mean of the submodel performance $Y_i$. When the upper level model function is linear with respect to $Y_i$, constants $A$ and $B$ in Eqn. (12) are equivalent to $\tilde{S}$ and $\tilde{T}$, respectively. For a general case, $A$ and $B$ are determined by the linear weighted regression presented in Section 3.4.

$$h_i(Y_i) = \int h(X_s, Y_i) dY_i = \int [S(X_s, Y_i) + T(X_s, Y_i) \cdot Y_i] dY_i - h_0.$$

So

$$St_i = \frac{V_x^2}{V_z} = \frac{\int S(X_s, Y_i) dY_i}{\int T(X_s, Y_i) dY_i}.$$

Let

$$f(Y_i, X_s) = S(X_s, Y_i) + T(X_s, Y_i) \cdot Y_i,$$

and $X_s \subset X_i$.

So

$$f_i(X_s) = \int f(Y_i, X_s) dY_i.$$

Hence

$$SI_i = \frac{V_x^2}{V_z} = AC \cdot SI_i,$$

where $SI_i$ is the mean of the submodel performance $Y_i$. When the upper level model function is linear with respect to $Y_i$, constants $A$ and $B$ in Eqn. (12) are equivalent to $\tilde{S}$ and $\tilde{T}$, respectively. For a general case, $A$ and $B$ are determined by the linear weighted regression presented in Section 3.4.

$$h_i(Y_i) = \int h(X_s, Y_i) dY_i = \int [S(X_s, Y_i) + T(X_s, Y_i) \cdot Y_i] dY_i - h_0.$$

So

$$St_i = \frac{V_x^2}{V_z} = \frac{\int S(X_s, Y_i) dY_i}{\int T(X_s, Y_i) dY_i}.$$

Let

$$f(Y_i, X_s) = S(X_s, Y_i) + T(X_s, Y_i) \cdot Y_i,$$

and $X_s \subset X_i$.

So

$$f_i(X_s) = \int f(Y_i, X_s) dY_i.$$

Hence

$$SI_i = \frac{V_x^2}{V_z} = AC \cdot SI_i,$$

where $SI_i$ is the mean of the submodel performance $Y_i$. When the upper level model function is linear with respect to $Y_i$, constants $A$ and $B$ in Eqn. (12) are equivalent to $\tilde{S}$ and $\tilde{T}$, respectively. For a general case, $A$ and $B$ are determined by the linear weighted regression presented in Section 3.4.

$$h_i(Y_i) = \int h(X_s, Y_i) dY_i = \int [S(X_s, Y_i) + T(X_s, Y_i) \cdot Y_i] dY_i - h_0.$$

So

$$St_i = \frac{V_x^2}{V_z} = \frac{\int S(X_s, Y_i) dY_i}{\int T(X_s, Y_i) dY_i}.$$

Let

$$f(Y_i, X_s) = S(X_s, Y_i) + T(X_s, Y_i) \cdot Y_i,$$

and $X_s \subset X_i$.

So

$$f_i(X_s) = \int f(Y_i, X_s) dY_i.$$