A Unified Framework for Uncertainty and Sensitivity Analysis of Computational Models with Many Input Parameters

C. F. Jeff Wu

H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology
Joint work with Li Gu, Yuming Sun, and Godfried Augenbroe
Outline

- Introduction
  - Uncertainty Analysis
  - Sensitivity Analysis
- Framework
  - Latin Hypercube Sampling
  - Generalized Polynomial Chaos Expansions
- Hierarchical Variable Selection Approach
- Examples
Uncertainty Analysis

Identifying and analyzing the uncertainties on complex systems are important.

Identification of uncertainties

Uncertainty Analysis (UA) studies how the uncertainties in the input parameters can be mapped to the uncertainties in the outputs.

- Monte Carlo methods (Helton and Davis, 2003).
Sensitivity Analysis

Analysis of uncertainties

*Sensitivity Analysis* (SA) studies how the total output uncertainty can be attributed to the uncertainties in each input parameter.

Direct computation:
- Fourier Amplitude Sensitivity Test (FAST, Cukier *et al.*, 1973);
- Sobol’ indices (Sobol’, 1993).

Metamodelling:
- Linear regression model (Saltelli *et al.*, 2001);
- Gaussian process model (Oakley and O’Hagan, 2004);
- Generalized polynomial chaos expansions (Sudret, 2008);
- Smoothing spline (Touzani and Busby, 2013).
SA on High Dimensions

Based on the effect sparsity principle (Wu and Hamada, 2009), only a few (but unknown) parameters are significant among many candidates. When the number of input parameters is large, variable selection is necessary.

Screening design:


Model-based variable selection:

▶ Bayesian variable selection on Gaussian process models (Linkletter et al., 2006);

▶ Bayesian variable selection on smoothing spline ANOVA models (Reich et al., 2009).
One can choose the aforementioned methods to sequentially conduct UA, variable selection, and SA.

- Too many function evaluations due to *three different* methods and *three separate* designs.
- Variable selection is computational intensive when the number of parameters is large and the computational model is complex.
Consider the computational model

\[ y = f(x), \]

where \( f \) is the computer code, \( x \) is the \( p \)-variate input, and \( y \) is the output.

- Black-box function, which can be nonlinear and non-additive.
- High dimensions, where a few of the input parameters are significant. Variable selection is challenging.
- No experimental error.
Framework

- **UA:**
  Monte Carlo simulation with Latin Hypercube Samples (LHS);

- **Variable selection:**
  a new hierarchical approach on generalized Polynomial Chaos expansions (gPC, Xiu and Karniadakis, 2002);

- **SA:**
  Sobol’ indices in gPC can be computed analytically (Sudret, 2008).

Since the approach connects UA and SA using one design, we call it *Uncertainty-Sensitivity-Analysis* (or abbreviated as USA).
The computational time in the simulation is drastically reduced.
Work efficiently for high dimensions.
Latin Hypercube Sampling/Designs
2-D Projection of a 16-run LHD
Latin Hypercube Designs: Properties

- Easy to construct:
  - randomly permute integers \(\{1, \ldots, n\}\) for each factor.
- Each input factor has \(n\) different levels.
- Guarantee one-dimensional uniformity.
- Can be poor in terms of space-filling.
Latin Hypercube Designs: Space-filling

- Maximin and minimax LHDs (Johnson et al., 1990);
- Orthogonal Array-based LHDs (Owen, 1992, and Tang, 1993);
- LHDs using correlation criteria (Tang, 1998);
- Orthogonal-maximin LHDs (Joseph and Hung, 2008);
- Multi-layer designs (Ba and Joseph, 2011).
An orthogonal polynomial basis $\phi_\alpha(x)$ ($\alpha$ corresponds to the order) satisfies

$$\int \phi_\alpha(x)\phi_\beta(x)p(x)dx = \gamma \delta_{\alpha\beta},$$

where $p(x)$ is the density function of $x$, $\gamma$ is a normalizing constant, and $\delta_{\alpha\beta} = 1$ if $\alpha = \beta$ and 0 otherwise.

For $x = (x_1, \ldots, x_p)^T$, define the $p$-dimensional basis

$$\Psi_\alpha(x) = \phi_{\alpha_1}(x_1) \cdots \phi_{\alpha_p}(x_p),$$

where $\alpha = (\alpha_1, \ldots, \alpha_p)^T$. The order of $\Psi_\alpha$ is $|\alpha| = \sum_{j=1}^p \alpha_j$. 
### gPC: Basis Table

The correspondence of the polynomials and their underlying random variables is listed below.

<table>
<thead>
<tr>
<th>Random Variables</th>
<th>Polynomials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>Hermite</td>
</tr>
<tr>
<td>Uniform</td>
<td>Legendre</td>
</tr>
<tr>
<td>Beta</td>
<td>Jacobi</td>
</tr>
<tr>
<td>Gamma</td>
<td>Laguerre</td>
</tr>
<tr>
<td>Binomial</td>
<td>Krawtchouk</td>
</tr>
<tr>
<td>Hypergeometric</td>
<td>Hahn</td>
</tr>
<tr>
<td>Negative Binomial</td>
<td>Meixner</td>
</tr>
<tr>
<td>Poisson</td>
<td>Charlier</td>
</tr>
</tbody>
</table>
The gPC is given by

\[ g(x) = \sum_{k=0}^{\infty} \beta_k \Psi_{\alpha_k}(x), \quad (1) \]

where \( \beta_k \) are unknown coefficients. \( g(x) \) in (1) with a proper choice of \( \beta_k \) converges in quadratic mean to \( f(x) \) (Soize and Ghanem, 2004).

For practical computation, consider the truncated form of the gPC

\[ g_P(x) = \sum_{k=0}^{P} \beta_k \Psi_{\alpha_k}(x). \]
gPC: Inference

To estimate the unknown coefficients $\beta_k$:

- Intrusive methods: Galerkin minimization (Ghanem and Spanos, 2003);
- Non-intrusive methods: regression method (Sudret, 2008), projection method (Crestaux et al., 2009).

**Regression method**

Minimize the following sum of squares of residuals:

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^{n} \left\{ f(x_i) - \sum_{k=0}^{P} \beta_k \Psi_{\alpha_k}(x_i) \right\}^2.$$

The solution is

$$\hat{\beta} = (\Psi^T \Psi)^{-1} \Psi^T y,$$

where $\Psi$ denotes the $n \times P$ matrix with $(i, k)$th entry $\Psi_{\alpha_k}(x_i)$, and $y = (f(x_1), \ldots, f(x_n))^T$. 

C. F. Jeff Wu

Framework for Uncertainty and Sensitivity Analysis 17 / 37
Variable selection techniques in linear regression can be applied to the gPC after some modifications:

- Stepwise regression (Blatman and Sudret, 2011);
- LARS (Blatman and Sudret, 2011);
- Bayesian variable selection (Tan, 2014).

However, they work only for a small number of input parameters. Consider the gPC with order $L$. The number of candidate bases $P = (p + L)!/p!L!$ increases rapidly.
Existing Variable Selection Methods

Variable selection methods with two levels of accuracy for linear models are needed. A quick method with low accuracy is used for ultrahigh dimensions.

Sure Independence Screening (SIS, Fan and Lv, 2008) for selecting bases

Denote \( \Psi_{\alpha_k}(X) = (\Psi_{\alpha_k}(x_1), \ldots, \Psi_{\alpha_k}(x_n))^T \).

1. Compute the correlation \( \omega_k = \text{corr}(\Psi_{\alpha_k}(X), y) \), for \( k = 1, \ldots, P \).
2. Choose first \( [\gamma n] \) bases with the largest correlations, where \( \gamma n = n - 1 \) or \( n/\log(n) \).

Alternatives:
- Forward selection (Wang, 2009);
- Correlation pursuit (Zhong et al., 2012).
A more elaborate method with high accuracy is used for moderate dimensions.

Lasso (Tibshirani, 1996) for selecting bases

\[
\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \left\{ f(x_i) - \sum_{k=0}^{P} \beta_k \Psi_{\alpha_k}(x_i) \right\}^2 + \lambda \sum_{k=0}^{P} |\beta_k|,
\]

where \( \lambda \) is a tuning parameter that controls the number of selected bases.

Alternatives:
- Smoothly Clipped Absolute Deviation (SCAD, Fan and Li, 2001);
- Adaptive Lasso (Zou, 2006).
Two principles (Wu and Hamada, 2009), originated in design of experiments, are commonly considered in variable selection.

- The effect hierarchy principle:
  Lower order effects are more likely to be important than higher order effects.

- The effect heredity principle:
  An interaction can be active only if one or all of its parent effects are also active.
Basic Idea

```
/   /
|   |
|   | hierarchy
|  /
Lasso ←→ SIS  Order 1

/  /
|  |
|  |
/   /
|   |
|   | heredity
|  /
Lasso ←→ SIS  Order 2

/  /
|  |
|  |
/   /
|   |
|   | heredity
|  /
Lasso ←→ SIS  Order L

/  /
|  |
|  |
/   /
|   |
|   | heredity
|  /
Lasso ←→ SIS  Order 1
```

C. F. Jeff Wu
Framework for Uncertainty and Sensitivity Analysis
Within the Layer

Iteratively select bases within the layer

Let \( M^{(l)} = \{ \alpha_k : k = i_{k_1}, \ldots, i_{k_i}; \text{ and } |\alpha_k| \leq l \} \) be the set of selected bases in the \( l \)th layer, and \( C^{(l)} \) a candidate set for the \( l \)th layer. In the \( l \)th layer, based on \( M^{(l-1)} \) and \( C^{(l)} \):

1. Set \( r_0 = y \) and \( S^{(l)}_0 = \emptyset \).
2. At iteration \( m \), set the residual \( r_{m-1} \), which is a vector with \( i \)th entry \( f(x_i) - \sum_{\alpha_k \in S^{(l)}_{m-1}} \hat{\beta}_k \Psi \alpha_k(x_i) \), to be the response.
3. Apply the SIS on \( C^{(l)} \setminus S^{(l)}_{m-1} \) to select a subset of bases \( A^{(l)}_m \).
4. Apply the lasso on \( S^{(l)}_{m-1} \cup A^{(l)}_m \) to obtain the set of selected bases \( S^{(l)}_m \).
5. Stop if converge; go to step 2 otherwise. Denote the last \( S^{(l)}_m \) by \( S^{(l)} \).
Between Adjacent Layers

How to generate $C^{(l)}$?

Note that it is very likely that some of the parent effects of significant interactions are not significant in lower layers.

At the beginning of each layer, $C^{(l)}$ is generated by expanding $M^{(l-1)}$ (weak heredity):

$$C^{(l)} = \{ \alpha_k : \exists \alpha' \in M^{(l-1)}, \alpha'' \in C \text{ s.t. } \alpha' + \alpha'' = \alpha_k; \text{ and } |\alpha_k| \leq l \}.$$  

(2)

Explanation: $\alpha_k$ has a single parent $\alpha'$ from $M^{(l-1)}$. For example, consider three candidate bases $C = \{ \Psi_{(1,0,0)}, \Psi_{(0,1,0)}, \Psi_{(0,0,1)} \}$ and $M^{(l-1)} = \{ \Psi_{(0,1,0)} \}$. From (2), it follows

$$C^{(l)} = \{ \Psi_{(1,1,0)}, \Psi_{(0,2,0)}, \Psi_{(0,1,1)} \}.$$
How to generate $\mathcal{M}^{(l)}$?

At the end of each layer, expand $S^{(l)}$ (strong heredity) to:

$$D^{(l)} = \{\alpha_k : \exists \alpha', \alpha'' \in S^{(l)} \text{ s.t. } \alpha' + \alpha'' = \alpha_k; \text{ and } |\alpha_k| \leq l\}. \quad (3)$$

Explanation: $\alpha_k$ has both parents $\alpha', \alpha''$ from $S^{(l)}$. For example, consider $S^{(l)} = \{\Psi_{(1,0,0)}, \Psi_{(0,1,0)}, \Psi_{(0,0,1)}\}$. From (3), it follows

$$D^{(l)} = \{\Psi_{(2,0,0)}, \Psi_{(1,1,0)}, \Psi_{(1,0,1)}, \Psi_{(0,2,0)}, \Psi_{(0,1,1)}, \Psi_{(0,0,2)}\}.$$ 

Then use the lasso on $D^{(l)}$ with the response $y$. The set of the final selected bases is denoted by $\mathcal{M}^{(l)}$.

By doing so, the interactions between significant bases, which are likely to be significant, can be re-examined.
Pseudo-code

Set $\mathcal{M}^{(0)} = \emptyset$

for $l = 1, 2, \ldots, L$

   Expand $\mathcal{M}^{(l-1)}$ by (2) to obtain $\mathcal{C}^{(l)}$

   Set $r_0 = y$ and $S_0^{(l)} = \emptyset$

   for $m = 1, 2, \ldots$

      Set $r_{m-1}$ to be the response

      Apply the SIS on $\mathcal{C}^{(l)} \setminus S_{m-1}^{(l)}$ to obtain $\mathcal{A}_m^{(l)}$

      Apply the lasso on $S_{m-1}^{(l)} \cup \mathcal{A}_m^{(l)}$ to obtain $S_m^{(l)}$

   endfor

if converge; denote the last $S_m^{(l)}$ by $S_l^{(l)}$

Expand $S_l^{(l)}$ by (3) to obtain $\mathcal{D}^{(l)}$

Set $y$ to be the response

Apply the lasso on $\mathcal{D}^{(l)}$ to obtain $\mathcal{M}^{(l)}$

endfor
Morris Function

\[ y = \beta_0 + \sum_{i=1}^{20} \beta_i w_i + \sum_{i<j}^{20} \beta_{ij} w_i w_j + \sum_{i<j<l}^{20} \beta_{ijl} w_i w_j w_l + \sum_{i<j<l<s}^{20} \beta_{ijls} w_i w_j w_l w_s, \]

where

\[ w_i = \begin{cases} 
2(1.1X_i/(X_i + 0.1) - 0.5) & \text{for } i = 3, 5, 7, \\ 
2(X_i - 0.5) & \text{otherwise,} 
\end{cases} \]

and \( X_i \sim U(0, 1) \). The \( \beta_i \) are assigned as

\[ \beta_i = 20 \quad \text{for } i = 1, \ldots, 10, \]

\[ \beta_{ij} = -15 \quad \text{for } i, j = 1, \ldots, 6, \]

\[ \beta_{ijl} = -10 \quad \text{for } i, j, l = 1, \ldots, 5, \]

\[ \beta_{ijls} = 5 \quad \text{for } i, j, l, s = 1, \ldots, 4. \]

The remaining coefficients are zero. The true model has 20 \( X_i \) input parameters, out of which the first ten are significant. 980 dummy parameters are added. So we have \( p = 1000 \) parameters in total.
<table>
<thead>
<tr>
<th>Layer</th>
<th>Stage</th>
<th>Bases</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Weak heredity</td>
<td>N/A; $C^{(1)} = 1000$ bases</td>
</tr>
<tr>
<td></td>
<td>First: SIS</td>
<td>$A_1^{(1)} = 81$ bases</td>
</tr>
<tr>
<td></td>
<td>First: Lasso</td>
<td>$S_1^{(1)} = 19$ bases: 3, 5, 7, 8, 9, 10, 93, 203, 284, 286, 362, 434, 510, 511, 623, 815, 896, 940, 941</td>
</tr>
<tr>
<td></td>
<td>Last: SIS, Lasso</td>
<td>$S^{(1)} = 17$ bases: 3, 5, 7, 8, 9, 10, 93, 203, 284, 286, 362, 434, 510, 511, 556, 896, 966</td>
</tr>
<tr>
<td></td>
<td>Strong heredity</td>
<td>N/A; $M^{(1)} = S^{(1)}$</td>
</tr>
<tr>
<td>2</td>
<td>Weak heredity</td>
<td>$C^{(2)} = 16711$ bases</td>
</tr>
<tr>
<td></td>
<td>First: SIS</td>
<td>$A_1^{(2)} = 81$ bases</td>
</tr>
<tr>
<td></td>
<td>First: Lasso</td>
<td>$S_1^{(2)} = 23$ bases: 5, 7, 8, 9, 10, 103, 1&amp;3, 1&amp;5, 2&amp;5, 3&amp;4, 3&amp;6, 4&amp;6, 161&amp;284, 163&amp;556, 167&amp;896, 198&amp;284, 198&amp;544, 238&amp;284, 284&amp;858, 284&amp;885, 510&amp;592, 510&amp;966, 673&amp;792</td>
</tr>
<tr>
<td></td>
<td>Last: SIS, Lasso</td>
<td>$S^{(2)} = 19$ bases: 5, 7, 8, 9, 10, 103, 1&amp;3, 1&amp;5, 2&amp;5, 3&amp;4, 3&amp;6, 93&amp;108, 161&amp;284, 163&amp;556, 167&amp;896, 198&amp;284, 238&amp;284, 510&amp;592, 510&amp;869</td>
</tr>
<tr>
<td></td>
<td>Strong heredity</td>
<td>$M^{(2)} = 16$ bases: 7, 8, 9, 10, 1&amp;2, 1&amp;3, 1&amp;5, 1&amp;6, 2&amp;4, 2&amp;5, 2&amp;6, 3&amp;4, 3&amp;6, 4&amp;6, 7&amp;7, 163&amp;556</td>
</tr>
<tr>
<td>3</td>
<td>Weak heredity</td>
<td>$C^{(3)} = 15744$ bases</td>
</tr>
<tr>
<td></td>
<td>First: SIS</td>
<td>$A_1^{(3)} = 81$ bases</td>
</tr>
<tr>
<td></td>
<td>First: Lasso</td>
<td>$S_1^{(3)} = 21$ bases: 7, 8, 9, 10, 1&amp;2, 1&amp;3, 1&amp;4, 1&amp;5, 1&amp;6, 2&amp;4, 2&amp;5, 2&amp;6, 3&amp;4, 3&amp;6, 4&amp;5, 4&amp;6, 5&amp;6, 7&amp;7, 3&amp;4&amp;421, 5&amp;574, 7&amp;32</td>
</tr>
<tr>
<td></td>
<td>Last: SIS, Lasso</td>
<td>$S^{(3)} = 16$ bases: 7, 8, 9, 10, 1&amp;2, 1&amp;3, 1&amp;4, 1&amp;5, 1&amp;6, 2&amp;4, 2&amp;5, 2&amp;6, 3&amp;4, 3&amp;6, 4&amp;6, 7&amp;7</td>
</tr>
<tr>
<td></td>
<td>Strong heredity</td>
<td>$M^{(3)} = 20$ bases: 7, 8, 9, 10, 1&amp;2, 1&amp;3, 1&amp;4, 1&amp;5, 1&amp;6, 2&amp;4, 2&amp;5, 2&amp;6, 3&amp;4, 3&amp;6, 4&amp;6, 7&amp;7, 10&amp;10&amp;10</td>
</tr>
</tbody>
</table>
Morris Function: Results

Table shows the average values of the numbers of true positives and false positives with 20 replications. Standard deviations are given in parentheses.

<table>
<thead>
<tr>
<th>$n$</th>
<th>True Positives</th>
<th>False Positives</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>7.11(1.66)</td>
<td>9.58(4.05)</td>
</tr>
<tr>
<td>500</td>
<td>9.26(0.87)</td>
<td>2.95(2.83)</td>
</tr>
<tr>
<td>1000</td>
<td>10(0)</td>
<td>1.36(0.56)</td>
</tr>
</tbody>
</table>
Building Energy Simulation

Background

- The building sector accounts for 40% of the total energy consumption in U.S.
- Retrofit existing buildings to reduce energy consumption and greenhouse gas emissions.
- Uncertainties in the input parameters come from a variety of sources.
Building Energy Simulation: Data

Large Offices in Prototype Buildings

- Energy Plus simulator.
- 2307 input parameters (pre-retrofit, post-retrofit, and common) regarding building physical properties, system settings, usages, and weather.
- Outputs were the energy consumed in the buildings (cooling and electricity).
- 1200 runs.
## Building Energy Simulation: Input Parameters

<table>
<thead>
<tr>
<th></th>
<th>Pre-Retrofit Only</th>
<th>Post-Retrofit Only</th>
<th>Common</th>
</tr>
</thead>
<tbody>
<tr>
<td>1313 parameters</td>
<td>978 parameters</td>
<td>16 parameters</td>
<td></td>
</tr>
<tr>
<td>Construction-Pre</td>
<td>Construction-Post</td>
<td>Urban Heat Island</td>
<td></td>
</tr>
<tr>
<td>Window-Pre</td>
<td>Window-Post</td>
<td>Local Wind Speed</td>
<td></td>
</tr>
<tr>
<td>Infiltration-Pre</td>
<td>Infiltration-Post</td>
<td>Diffuse Solar Irradiation</td>
<td></td>
</tr>
<tr>
<td>Ventilation-Pre</td>
<td>Ventilation-Post</td>
<td>Ground Reflectance</td>
<td></td>
</tr>
<tr>
<td>HVAC &amp; Lighting System-Pre</td>
<td>HVAC &amp; Lighting System-Post</td>
<td>Convective Heat Transfer</td>
<td></td>
</tr>
<tr>
<td>Electric Equipment-Pre</td>
<td>Electric Equipment-Post</td>
<td>Occupant Density</td>
<td></td>
</tr>
</tbody>
</table>
Building Energy Simulation: Results

- In cooling and electricity, the USA selected the gPC which explained 90% and 92% variation of the data, respectively. In contrast, the linear regression model can only explain 78% and 71%, respectively.

- The linear model can be viewed as the gPC with order one. The significant improvement is due to the use of higher-order gPC, which was selected in higher layers of the USA.
Building Energy Simulation: Sensitivity Indices

Figure: Cooling: 39 out of 2307

Figure: Electricity: 31 out of 2307
We have proposed a framework dubbed the USA. The USA unifies UA and SA with the use of one design, and significantly reduce the number of simulations required. In the USA, a new hierarchical variable selection approach is applied for computational models with many input parameters.

- By incorporating the effect hierarchy principle and the effect heredity principle, the number of candidate bases in each layer is reduced to an acceptable level for variable selection.
- By applying the SIS and the lasso alternately, the USA strikes a balance between computation and performance.
References


