Statistical Meta-Modeling for Complex System Simulations: Kriging, Alternatives and Design

C. F. Jeff Wu
School of Industrial and Systems Engineering
Georgia Institute of Technology

- Statistical meta modeling of computer experiments
- Kriging as the main tool: two examples of computer experiments
  - two levels of accuracy
  - with qualitative/quantitative factors
- Alternatives to kriging (to avoid numerical instability)
- Some novel design issues
Why computer experiments?

✓ No need for expensive lab equipments and materials, less costly than physical experiments.

✓ Not affected by human and environmental factors.

✓ Study dangerous or infeasible physical experiments, such as ammunition detonation.
Some examples

Computer Experiments/Simulations

Chemical & Biology: nanoparticle and polymer synthesis…

Mechanical: machining, assembling…

Aerospace: aircraft design, dynamics…
Statistical Meta-Modeling of Computer Experiments

Robustness, optimization

Surrogate model (Kriging)

Noise simulations, error propagation

More FEA runs

Computer modeling (finite-element simulation)

Physical experiment or observations
Kriging as an interpolator

• Kriging is an interpolation method: \( \hat{y}(x_i) = y_i, x \in \mathbb{R}^p. \)
• Originated in geostatistics; proposed for computer experiments by Sacks, Welch et al.
Kriging Modeling

- Model assumption: Gaussian process.

\[ y(x) \sim GP(\mu(x), \sigma^2 \phi(\cdot)). \]

- \( \mu(x) \) — mean, linear model \( \mu(x) = f(x)' \beta \), where
  \[ f(x) = (f_1(x), \ldots, f_k(x))'. \]

- \( \phi(\cdot) \) — correlation function, \( \phi(x_1 - x_2, \theta) \), e.g., Gaussian Correlation Function,
  \[ \phi(x_1 - x_2, \theta) = \exp\left(-\sum_{i=1}^{p} \theta_i (x_{1,i} - x_{2,i})^2\right), \quad \theta \in \mathbb{R}_+^p \]

- \( \sigma^2 \) — variance.
Best Linear Unbiased Predictor

- Best Linear Unbiased Predictor (BLUP):

\[ \hat{y}(x) = f(x)\hat{\beta} + r(x)'R^{-1}(y - F\hat{\beta}), \]

- \( \hat{\beta} = (F' R^{-1} F)^{-1} F' R^{-1} y \) is the generalized least squares estimation. \( F = n \times k \) model matrix.
- \( R \) is the correlation matrix, \( R_{i,j} = \phi(x_i - x_j, \theta) \).
- \( r(x) = (\phi(x - x_1, \theta), \ldots, \phi(x - x_n, \theta))' \) = vector of correlation between prediction points and observed points.
- \( \hat{y}(x_i) = y_i \), interpolating property.
Kriging Modeling: Estimation

• Maximum Likelihood Estimation:

\[
\min_{\theta \geq 0} \left\{ n \log (\hat{\sigma}^2) + \log (\det(R)) \right\}.
\]

where \( \hat{\sigma}^2 = (y - F\hat{\beta})' R^{-1} (y - F\hat{\beta}) / n. \)

• Numerical Stability: R is ill-conditioned, especially for large \( n \) and/or \( k \) (=input dimension) (Peng and Wu, 2010b).

• Computational Complexity: matrix inversion \( O(n^3) \); many optimization iterations are needed; in each iteration matrix inversion is computed.
Examples of Kriging Models

• Computer experiments with two levels of accuracy
• Computer experiments with quantitative and qualitative factors
• Simulation codes with calibration parameters
Example: Designing Cellular Heat Exchangers for an Electronic Cooling Application

Important Factors:
- Flow-rate of Air
- Inlet Temp of Air
- Conductivity of Solid
- Temp of Upper Wall

Response: Total Heat Transfer Rate from Solid to Air

Courtesy of Systems Realization Lab at Georgia Tech
Heat Transfer Analysis

**HE: Detailed Computer Simulation – Finite Element Analysis (FEA) Method**

- Using the computational fluid dynamic solver FLUENT

- Problem domain is divided into thousands or millions of elements.

- Each run requires hours to days to complete.
Heat Transfer Analysis

LE: Approximate Computer Simulation — Finite Difference Method

- The finite difference technique is a numerical technique for solving 2- or 3-D steady state heat transfer problems.
- Temperature distribution approximated via numerical solution of 3D heat transfer equations using forward or central difference methods.
- Each run takes minutes to complete.
- Less accurate than FEA.
High-accuracy and Low-accuracy Experiments

• **Paradigm shift**: single experiment $\rightarrow$ multiple experiments with different levels of accuracy.

• A generic pair: **high-accuracy experiment** (HE) and **low-accuracy experiment** (LE).

• HE is more **accurate** but more **expensive** than LE.

• Examples:
  – physical experiment vs. computer simulation
  – detailed computer simulation vs. approximate computer simulation

• 1. **Modeling**: How to model and analyze data from HE and LE?

  2. **Experimental design**: How to plan HE and LE?
Frequentist Approach in Qian et al. (2006, ASME)


- \( \mathbf{x} = (x_1, \ldots, x_k) \): design variables.
  - \( D_l \) and \( D_h \): sets of design points \((\mathbf{x}_i)\) for LE and HE with \( D_h \subset D_l \).
  - \( y_h \) and \( y_l \): outputs from HE and LE.

- **Base Surrogate Model**: \( y_l = \beta_{l0} + \sum_j \beta_{lj} \mathbf{x}_i + \varepsilon_l(\mathbf{x}_i), \; \mathbf{x}_i \in D_l \),
  \( \varepsilon_l \sim \text{GP}(0, \sigma_l^2, \phi_l) \).

- **Adjustment Model**: \( y_h(\mathbf{x}_i) = \rho(\mathbf{x}_i)y_l(\mathbf{x}_i) + \delta(\mathbf{x}_i), \; \mathbf{x}_i \in D_h \),
  - scale adjustment: \( \rho(\mathbf{x}_i) = \rho_0 + \sum_j \rho_j x_{ij} \);
  - location adjustment: \( \delta \sim \text{GP}(\delta_0, \sigma_\delta^2, \phi_\delta) \).

- **Fitting**:
  \[
  \begin{align*}
  \text{Base surrogate model: } & \widehat{y}_l \\
  \text{Adjustment model: } & \widehat{\rho} \text{ and } \widehat{\delta}
  \end{align*}
  \]
  \( \implies \text{Final surrogate model: } \widehat{y}_h = \widehat{\rho}\widehat{y}_l + \widehat{\delta} \).

- **Desirable property**: \( \widehat{y}_h \) interpolates \( y_h(\mathbf{x}_i), \mathbf{x}_i \in D_h \) if HE is deterministic.
Bayesian Approach in Qian and Wu (2008, Technometrics)

- Use flexible Bayesian hierarchical Gaussian process model (BHGP).
- Provide more flexible adjustment. Can accommodate parameter uncertainty and measurement error of HE.
- BHGP:
  - Base Surrogate Model: \( y_l = \beta_{l0} + \sum_j \beta_{lj} x_i + \varepsilon_l(x_i), \ x_i \in D_l, \)
    \( \varepsilon_l \sim GP(0, \sigma^2_l, \phi_l). \)
  - Flexible Adjustment Model:
    \[
    y_h(x_i) = \rho(x_i)y_l(x_i) + \delta(x_i) + \varepsilon(x_i), \ x_i \in D_h,
    \]
    scale adjustment \( \rho \sim GP(\rho_0, \sigma^2_{\rho}, \phi_{\rho}), \)
    location adjustment \( \delta \sim GP(\delta_0, \sigma^2_{\delta}, \phi_{\delta}), \)
    \( \varepsilon(x) \sim N(0, \sigma^2_{\varepsilon}). \ \varepsilon = 0 \iff \) no measurement error.
Model Parameters and Priors for the BHGP Model

- **Model parameters**
  - Mean parameters $\theta_1 = (\beta_l, \rho_0, \delta_0)$.
  - Variance parameters $\theta_2 = (\sigma^2_l, \sigma^2_\rho, \sigma^2_\delta, \sigma^2_\varepsilon)$.
  - Correlation parameters $\theta_3 = (\phi_l, \phi_\rho, \phi_\delta)$;
    complexity *increases* with the no. of input factors ($\therefore$ dim = $3k$).

- **Priors**
  - Normal for $\theta_1$.
  - Inverse-gamma for $\theta_2$.
  - Gamma for $\theta_3$. 
Computational Algorithm

1: Fitting Correlation Parameters $\theta_3$

Solve a *stochastic program*

$$\max_{\phi_\rho, \phi_\delta} L_2 = E_{p(\tau_1, \tau_2)} f(\tau_1, \tau_2)$$

by the Sample Average Approximation (SAA) algorithm.

2: Markov Chain Monte Carlo (MCMC) Sampling from $p(\theta_1, \theta_2 | y_l, y_h, \theta_3)$

- Rationale: Conditional distributions for $\theta_1, \theta_2$ are in closed form, but not for $\theta_3$ given $\theta_1, \theta_2$.
- It is not a *fully* Bayesian procedure (to save computation for $\theta_3$).
Posterior density of $\rho_0$, $\delta_0$, $\sigma^2_\rho$, $\sigma^2_\delta$
Location and Scale Change

- **Symmetric** location change $\delta_0$ with center at 0.85.
- Scale adjustment $\rho_0$ has multi modes, which may be explained by *multiple* laws.
Calibration of Computer Models

• Calibration = process of fitting a computer model to observed data by adjusting input parameters of the model (Kennedy and O’Hagan, 2001).

• Computer model: \( y_c = \eta(x, \theta) \),

• Physical model: \( y_p = \rho \eta(x, \theta) + \delta(x) + \varepsilon \),
  \(- \theta = \text{calibration parameters}, \delta(x) = \text{model inadequacy}, \rho = \text{adjustment term}.\)

• Example 1-Nuclear Accident: Use data on reactor fire to calibrate a plume model of radioactive deposition. Source term and deposition velocity = calibration parameters.

GP with quanti/quali factors: Data Center Thermal Distribution

Courtesy of IBM T. J. Watson Research Center
Configuration Variables for the Data Center Example

- Five quanti factors: rack temperature rise, rack power, diffuser angle, diffuser flow rate, ceiling height.
- Three quali factors: diffuser location, hot-air return-vent location, power allocation.
GP Models with Quantitative and Qualitative Factors
(Qian, H. Wu, C. F. J. Wu 2008, Technometrics)

- Input factors: \( w = (x, z) \); \( I \) quantitative factors: \( x = (x_1, \ldots, x_I) \); \( J \) qualitative factors: \( z = (z_1, \ldots, z_J) \); response value: \( y(w) \).

- Build a single GP model for both \( x \) and \( z \). Borrow strengths from all the observations:

\[
y(w) = \sum_{m} \beta_m f_m(w) + \epsilon(w).
\]

- How to specify \( f_m \)? Use regression modeling involving \( x \) and \( z \).
- How to specify \( \epsilon \) (especially its correlation structure)?
- Related work: Bayesian hierarchical GP models (Han et al., 2009).
- Corresponding designs: Sliced space-filling designs (Qian and Wu, 2009, Biometrika).
Construction of Correlation Functions for $\varepsilon(w)$

- Consider one quali factor $z_1$ with $m_1$ levels 1, $\ldots$, $m_1$.
  For $u = 1, \ldots, m_1$, $\varepsilon_u(x) = \varepsilon(x, u)$.

- Idea: envision a mean-zero multivariate process $(\varepsilon_1(x), \ldots, \varepsilon_{m_1}(x))' = A\eta(x)$.

- $A$: an $m_1 \times m_1$ matrix with unit row vectors.

- Elements of $\eta(x)$: $m_1$ independent processes with a common variance $\sigma^2$.

- For two input values $w_1 = (x_1, z_{11})$ and $w_2 = (x_2, z_{12})$,

$$\text{cor}(\varepsilon(w_1), \varepsilon(w_2)) = \tau_{z_{11}, z_{12}} K_\phi(x_1, x_2)$$

  - $K_\phi(x_1, x_2)$: correlation between $x_1$ and $x_2$.
  - $T_1 = AA^t$: an $m_1 \times m_1$ correlation matrix for $z_1$ (i.e., positive definite matrix with unit diagonal elements).
Alternatives to Kriging

1. Tweaking of kriging:
   – covariance matrix tapering (Kaufman et al., 2008),
   – rank reduction (Cressie-Johannesson, 2008),
   – regularized kriging (Peng-Wu, 2010a)

2. Adding penalty to likelihood (Li-Sudjianto, 2005): another approach to regularization.
Alternatives to Kriging (cont.)

3. Completely different ideas:

- Radial basis interpolation functions (Buhmann, 2004)

- Regression-based inverse distance weighting (Joseph and Kang, 2009): a fast interpolator

- Overcomplete basis surrogate method (OBSM) (Chen, Wang and Wu, 2010): an approximator, not interpolator
Inverse Distance Weighting (IDW)

- Inverse Distance Weighting (Shepard, 1968):

\[
\hat{y}(\mathbf{x}) = \frac{\sum_{k=1}^{n} w_k(\mathbf{x}) y_k}{\sum_{i=1}^{n} w_i(\mathbf{x})}.
\]

- \( w_i(\mathbf{x}) = \frac{1}{d(\mathbf{x}, \mathbf{x}_i)^2} \).
- \( d(\mathbf{x}, \mathbf{x}_i) = \left\{ \sum_{j=1}^{p} (x_j - x_{i,j})^2 \right\}^{1/2} \).

- Simple computation but poor prediction.
Regression-Based Inverse Distance Weighting (RIDW)

- Add regression part to IDW (Joseph and Kang, 2009):

\[ \hat{y}(\mathbf{x}) = \mu(\mathbf{x}; \boldsymbol{\beta}) + \frac{\sum_{k=1}^{n} w_k(\mathbf{x}) e_k}{\sum_{i=1}^{n} w_i(\mathbf{x})} \]

- \( \mu(\mathbf{x}_k; \boldsymbol{\beta}) \) = mean part; can be linear, nonlinear, nonparametric.

- \( e_k = y_k - \mu(\mathbf{x}_k; \boldsymbol{\beta}) = y_k - \mu_k \).

- \( w_i(\mathbf{x}) = \frac{\exp\{-d^2(\mathbf{x}, \mathbf{x}_i)\}}{d^2(\mathbf{x}, \mathbf{x}_i)} \). (faster convergence than IDW)

- \( d(\mathbf{x}, \mathbf{x}_i) = \sqrt{\sum_{j=1}^{p} \theta_j (x_j - x_{i,j})^2} \).
Comparisons Between RIDW and Kriging

Standardized RMSPE

CPU time in simulation
Design of Experiments with HE and LE

• Key to efficiently allocate resources and acquire information from HE and LE.

• $\mathbf{x} = (x_1, \ldots, x_k)$: design variables in $[0, 1]^k$.

  $D_l$: set of design points ($\mathbf{x}_i$) for LE. $D_h$: set of design points ($\mathbf{x}_i$) for HE.

• Three principles for constructing $D_l$ and $D_h$:
  
  **Economy:** The size of $D_h$ is less than the size of $D_l$.

  **Nested relationship:** $D_h$ is a subset of $D_l$.

  **Space-filling:** Points in $D_h$ and $D_l$ achieve uniformity in low dimensions.

• How to construct multiple experiments with respect to multiple requirements?

• New issue in design of experiments: traditional methods deal almost exclusively with experiments with one level of accuracy.
Nested Space-Filling Designs


Ideas:

1. Use a *special* orthogonal array to construct an OA-based Latin hypercube design for $D_l$.

2. Choose $D_h$ to be a *carefully* selected subset of $D_l$ with two-dimensional balance.

Need to use simple Galois field theory.
Nested Orthogonal Array

Let $A$ be an $OA(N_1, k, s_1, t)$. Suppose $A$ has a subarray with $N_2$ rows, denoted by $A_1$, and there is a projection $\phi$ that collapse the $s_1$ levels of $A$ into $s_2$ levels. Further suppose $A_1$ is an $OA(N_2, k, s_2, t)$ if the $s_1$ levels of its entries are collapsed to $s_2$ levels according to $\phi$. Then an array with this structure is called a nested orthogonal array.
Construction of Nested Space-Filling Design

• The nested orthogonal array $A_2 \subset A_1$ does not automatically generate nested space-filling designs if the $s_1$ levels are arbitrarily labeled.

• In constructing OA-based Latin hypercube $D_1$ using $A_1$, the $s_1$ levels of $A_1$, represented by the polynomials of degree $u_1 - 1$ or lower, have to be first labeled as $1, \ldots, s_1$.

• The subset $D_2$ of $D_1$ corresponding to $A_2$ may not have good space-filling properties.

• Care should be taken in labeling the levels to ensure that $D_2$ achieves stratification in two dimensions.

• The $s_1$ levels of $A_1$ must be labeled in such a way that the group of levels that are mapped to the same level should form a consecutive subset of $1, \ldots, s_1$. 
An Example of Nested Space-Filling Design

- Five factors \( \mathbf{x} = (x_1, x_2, x_3, x_4, x_5) \) taking values in the unit hypercube \([0, 1]^5\).
- Use the orthogonal array in Table 1 to construct an OA-based Latin hypercube \( D_1 \) for \( x_1-x_5 \).
- Choose a subarray \( D_2 \) of \( D_1 \) consisting of the 16 points in \( D_1 \) corresponding to runs 1-4, 9-12, 17-20, 25-28 of the array in Table 1.
The Underlying Nested Orthogonal Array

<table>
<thead>
<tr>
<th>Run #</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$x_4$</th>
<th>$x_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Run #</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$x_4$</th>
<th>$x_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>8</td>
<td>1</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>34</td>
<td>8</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>35</td>
<td>8</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>36</td>
<td>8</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>37</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>38</td>
<td>8</td>
<td>6</td>
<td>3</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>39</td>
<td>8</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>40</td>
<td>8</td>
<td>2</td>
<td>7</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>41</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>42</td>
<td>6</td>
<td>3</td>
<td>8</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>43</td>
<td>6</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>44</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>45</td>
<td>6</td>
<td>8</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>46</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>47</td>
<td>6</td>
<td>4</td>
<td>7</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>48</td>
<td>6</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Run #</td>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_3$</td>
<td>$x_4$</td>
<td>$x_5$</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>17</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>18</td>
<td>5</td>
<td>3</td>
<td>7</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>19</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>20</td>
<td>5</td>
<td>7</td>
<td>3</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>21</td>
<td>5</td>
<td>8</td>
<td>4</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>22</td>
<td>5</td>
<td>6</td>
<td>2</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>23</td>
<td>5</td>
<td>4</td>
<td>8</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>24</td>
<td>5</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>25</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>26</td>
<td>7</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>27</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>28</td>
<td>7</td>
<td>7</td>
<td>1</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>29</td>
<td>7</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>30</td>
<td>7</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>31</td>
<td>7</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>32</td>
<td>7</td>
<td>2</td>
<td>8</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Run #</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$x_4$</th>
<th>$x_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>49</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>50</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>51</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>52</td>
<td>4</td>
<td>7</td>
<td>6</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>53</td>
<td>4</td>
<td>8</td>
<td>5</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>54</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>55</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>56</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>57</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>58</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>59</td>
<td>2</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>60</td>
<td>2</td>
<td>7</td>
<td>8</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>61</td>
<td>2</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>62</td>
<td>2</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>63</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>64</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 1. An $OA(64,5,8,2)$ that contains an $OA(16,5,4,2)$, i.e., the submatrix consisting of runs 1-4, 9-12, 17-20, 25-28 becomes an $OA(16,5,4,2)$ after some suitable level collapsing
2-D Projection of $D_1$
2-D Projection of $D_2$
Conclusions

• Computer simulations have become popular in studying complex systems.
• Statistics-based meta (surrogate, approximate) models are useful companions to simulations.
• A rich class of problems: modeling, estimation, prediction, calibration, computations, design.
• Kriging is popular for building meta models but can have problems with numerical instability.
• Effective alternatives to kriging have emerged in recent work. Cross-disciplinary in nature.
Step 1: Fitting Correlation Parameters $\theta_3$

- Obtain posterior mode $\hat{\theta}_3 = (\hat{\phi}_l, \hat{\phi}_\rho, \hat{\phi}_\delta)$ by solving

\[(P) : \max_{\phi_l, \phi_\rho, \phi_\delta} p(\theta_3 | y_h, y_l).\]

- $(P)$ is equivalent to two separable problems:
  - $(P_1)$: a non-linear program for $\phi_l$,
  - $(P_2)$: a problem for $\phi_\rho$ and $\phi_\delta$. Its objective function involves an integral.

- $(P_2) \iff$ a stochastic program $(P'_2)$: \( \max_{\phi_\rho, \phi_\delta} L_2 = E_{p(\tau_1, \tau_2)} f(\tau_1, \tau_2). \)

- Solve $P'_2$ by the Sample Average Approximation (SAA) algorithm:
  1. Generate random samples $(\tau^s_1, \tau^s_2)$ from $p(\tau_1, \tau_2), s = 1, \cdots, S$.
  2. Solve the approximate problem:

\[(\tilde{\phi}_\rho, \tilde{\phi}_\delta) = \arg\max_{\phi_\rho, \phi_\delta} [\tilde{L}_2 = \frac{1}{S} \sum_{s=1}^{S} f(\tau^s_1, \tau^s_2)].\]
Step 2: Markov Chain Monte Carlo (MCMC)
Sampling from $p(\theta_1, \theta_2| y_l, y_h, \theta_3)$

- Some full conditional distributions for $\theta_1, \theta_2$ are not regular:
  - $p(\beta_l|y_l, y_h, \theta_3, \bar{\beta}_l) \sim \text{Normal}$,
  - $p(\rho_0|y_l, y_h, \theta_3, \bar{\rho}_0) \sim \text{Normal}$,
  - $p(\delta_0|y_l, y_h, \theta_3, \bar{\delta}_0) \sim \text{Normal}$,
  - $p(\sigma^2_l|y_l, y_h, \theta_3, \bar{\sigma}^2_l) \sim \text{IG}$,
  - $p(\sigma^2_\rho|y_l, y_h, \theta_3, \bar{\sigma}^2_\rho) \sim \text{IG}$,
  - $p(\tau_1, \tau_2|y_l, y_h, \tau_1, \tau_2) \propto \frac{1}{\tau_1^{\alpha_\delta + \frac{3}{2}}} \frac{1}{\tau_2^{\alpha_\epsilon + 1}} \exp\left\{ - \frac{1}{\tau_1} \left( \frac{\gamma_\delta}{\sigma^2_\rho} + \frac{(\delta_0 - u_\delta)^2}{2v_\delta \sigma^2_\rho} \right) - \frac{\gamma_\epsilon}{\tau_2 \sigma^2_\rho} \right\} \frac{1}{|M|^{1/2}} \cdot \exp\left\{ - \frac{(y_h - \rho_0 y_{l1} - \delta_0 1_{n_1})^t M^{-1} (y_h - \rho_0 y_{l1} - \delta_0 1_{n_1})}{2 \sigma^2_\rho} \right\}$
  - $\tau_1 = \sigma^2_\delta / \sigma^2_\rho$, $\tau_2 = \sigma^2_\delta / \sigma^2_\epsilon$

- Use the Metropolis-within-Gibbs algorithm.
The General Case

- Consider $J$ qualitative factors $z_1, \ldots, z_J$ with $z_j$ having $m_j$ levels $1, \ldots, m_j$.
- For two input values $w_1 = (x_1, z_{11})$ and $w_2 = (x_2, z_{12})$,

$$\text{cor}(\varepsilon(w_1), \varepsilon(w_2)) = \left[ \prod_{j=1}^{J} \tau_{j,z_{j1},z_{j2}} \right] \exp \left\{ -\sum_{i=1}^{I} \phi_i (x_{i1} - x_{i2})^2 \right\}.$$

- $T_j$: an $m_j \times m_j$ correlation matrix for $z_j$ (i.e., a positive definite matrix with unit diagonal elements).
- This correlation function has a product form.
- Assume the elements of $T_j \geq 0$ for deterministic computer experiments.
Some Restrictive Forms of $T_j$

Consider two input values $w_1 = (x_1, z_1)$ and $w_2 = (x_2, z_2)$ with responses $y(w_1)$ and $y(w_2)$. Recall that $\tau_{j,z_1,z_2}$ is the correlation between $z_{j1}$ and $z_{j2}$.

1. Isotropy correlation function: $\tau_{j,z_1,z_2} = \exp\{-\theta_j I[z_{j1} \neq z_{j2}]\}$.

   For $w_1$ and $w_2$,
   
   $$\text{cor}(\varepsilon(w_1), \varepsilon(w_2)) = \exp \left\{ -\sum_{i=1}^{I} \phi_i (x_{i1} - x_{i2})^2 - \sum_{j=1}^{J} \theta_j I[z_{j1} \neq z_{j2}] \right\}$$

   Euclidean distance for $x_i$; 0-1 distance for $z_j$.

2. Multiplicative correlation function:

   $$\tau_{r,s} = \exp\{-\theta_{r,s}\} = \exp\{-\theta_r + \theta_s I[r \neq s]\}. \quad (1)$$


4. Correlation functions for ordinal qualitative factors.
Estimation

• Model parameters:
  – mean parameters $\beta = (\beta_1, \ldots, \beta_p)$.
  – variance parameter $\sigma^2$.
  – correlation parameters $\phi = (\phi_1, \ldots, \phi_I)^t$, and $T = \{T_1, \ldots, T_J\}$.

• The estimation iterates between

  **Regression fitting:** Given $\hat{\phi}$ and $\hat{T}$, estimate $\beta$ and $\sigma^2$.
  
  Simple!

  **Correlation fitting:** Given $\hat{\beta}$ and $\hat{\sigma}^2$, let $U = (u_1, \ldots, u_n)$ with
  
  $u_i = [y_i - \hat{\beta}^t f(w_i)]/\hat{\sigma}$ and then fit a GP with mean zero and variance one to
  the transformed data $U$. 