Evaluation of Process Models with Monte Carlo Methods

John Kauffman
FDA\CDER Division of Pharmaceutical Analysis

Sharmista Chatterjee
FDA\CDER\ONDQA

Jean-Marie Geoffroy
Takeda Global Research & Development Center, Inc

IFPAC 2012

The findings and conclusions in this presentation have not been formally disseminated by the Food and Drug Administration and should not be construed to represent any Agency determination or policy.
Outline

1. Role of Monte Carlo Simulation in estimating uncertainty

2. Discussion of Case Studies

3. Conclusion
Process Models and Design Space

- Process models map process parameters and input variables onto the product attributes.
  - Design space may be defined in terms of a process model

- Process models are often based on a relatively small number of experiments
  - Designed experiments tend to be parsimonious.
  - Parsimonious experiments may have large variances.
Process Models and Probabilities

• Process models focus on predicting response means.
  – In a QbD paradigm, “What is needed is a multivariate predictive distribution for the quality responses.”
  – “If we can quantify the entire (multivariate) predictive distribution of the process quality responses as a function of [input variables and process parameters], then we can compute the probability of a future batch meeting the quality specifications.”
  – (Peterson, Pharmaceutical Manufacturing, June 25, 2010)
Advantages of Monte Carlo Simulation

1. We make no assumptions concerning sources of uncertainty or variable covariance.

2. We see the distribution of output variable values, not just a standard deviation.

3. Sensitivity analysis allows us to prioritize high risk input variables and improve process control.
Monte Carlo Methods

• Develop a mathematical model.
  – The Process Model.

• Add random variables.
  – Replace quantities of interest with random numbers selected from appropriate distribution functions that are expected to describe the variables.

• Monitor selected output variables.
  – Output variables become distributions whose properties are determined by the model and the distributions of the random variables.
Implementation of Monte Carlo Simulation

• Monte Carlo simulation offers a simple tool to explore influence of random variation in input parameters on multivariate predictive distributions

• In this presentation, we will examine the application of Monte Carlo simulation to a process model for two case studies
Case Study 1: Background Information

**Design Space** defined on the basis of Multivariate DOE results, RSM design, 20+3 center points

**DOE Inputs:**
- API particle size (API), log(D90) microns
- Magnesium Stearate specific surface area (MgSt), cm²/g
- Lubrication time (LubT), min
- Tablet hardness (Hard), N

**DOE Response:**
- % Dissolved in 20 minutes (Diss)

**Control Strategy:**
- Predictions from regression model derived from DOE data implemented as a surrogate for traditional dissolution testing

**Specification:** 80% dissolved in 20 min

Adapted from ICH IWG Training, October, 2010
## Case Study 1: DOE Data

<table>
<thead>
<tr>
<th>Exp No</th>
<th>Run Order</th>
<th>API</th>
<th>MgSt</th>
<th>LubT</th>
<th>Hard</th>
<th>Diss</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.5</td>
<td>3000</td>
<td>1</td>
<td>60</td>
<td>101.24</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>1.5</td>
<td>3000</td>
<td>1</td>
<td>60</td>
<td>87.99</td>
</tr>
<tr>
<td>3</td>
<td>22</td>
<td>0.5</td>
<td>12000</td>
<td>1</td>
<td>60</td>
<td>99.13</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>1.5</td>
<td>3000</td>
<td>10</td>
<td>60</td>
<td>86.03</td>
</tr>
<tr>
<td>5</td>
<td>18</td>
<td>0.5</td>
<td>12000</td>
<td>10</td>
<td>60</td>
<td>94.73</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>1.5</td>
<td>12000</td>
<td>10</td>
<td>60</td>
<td>83.04</td>
</tr>
<tr>
<td>7</td>
<td>15</td>
<td>0.5</td>
<td>3000</td>
<td>1</td>
<td>110</td>
<td>98.07</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>0.5</td>
<td>12000</td>
<td>1</td>
<td>110</td>
<td>97.68</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>1.5</td>
<td>12000</td>
<td>1</td>
<td>110</td>
<td>85.47</td>
</tr>
<tr>
<td>10</td>
<td>16</td>
<td>0.5</td>
<td>3000</td>
<td>10</td>
<td>110</td>
<td>95.81</td>
</tr>
<tr>
<td>11</td>
<td>20</td>
<td>1.5</td>
<td>3000</td>
<td>10</td>
<td>110</td>
<td>84.38</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>1.5</td>
<td>12000</td>
<td>10</td>
<td>110</td>
<td>81</td>
</tr>
<tr>
<td>13</td>
<td>10</td>
<td>0.5</td>
<td>7500</td>
<td>5.5</td>
<td>85</td>
<td>96.85</td>
</tr>
<tr>
<td>14</td>
<td>17</td>
<td>1.5</td>
<td>7500</td>
<td>5.5</td>
<td>85</td>
<td>85.13</td>
</tr>
<tr>
<td>15</td>
<td>19</td>
<td>1</td>
<td>3000</td>
<td>5.5</td>
<td>85</td>
<td>91.87</td>
</tr>
<tr>
<td>16</td>
<td>21</td>
<td>1</td>
<td>12000</td>
<td>5.5</td>
<td>85</td>
<td>90.72</td>
</tr>
<tr>
<td>17</td>
<td>7</td>
<td>1</td>
<td>7500</td>
<td>1</td>
<td>85</td>
<td>91.95</td>
</tr>
<tr>
<td>18</td>
<td>4</td>
<td>1</td>
<td>7500</td>
<td>10</td>
<td>85</td>
<td>88.9</td>
</tr>
<tr>
<td>19</td>
<td>5</td>
<td>1</td>
<td>7500</td>
<td>5.5</td>
<td>60</td>
<td>92.37</td>
</tr>
<tr>
<td>20</td>
<td>11</td>
<td>1</td>
<td>7500</td>
<td>5.5</td>
<td>110</td>
<td>90.95</td>
</tr>
<tr>
<td>21</td>
<td>12</td>
<td>1</td>
<td>7500</td>
<td>5.5</td>
<td>85</td>
<td>91.95</td>
</tr>
<tr>
<td>22</td>
<td>13</td>
<td>1</td>
<td>7500</td>
<td>5.5</td>
<td>85</td>
<td>90.86</td>
</tr>
<tr>
<td>23</td>
<td>23</td>
<td>1</td>
<td>7500</td>
<td>5.5</td>
<td>85</td>
<td>89</td>
</tr>
</tbody>
</table>
Case Study 1: Problem Description

Objective: Evaluate the response variance in the predictive model

Prediction algorithm:
\[
\text{Diss} = 108.9 - 11.96 \times \text{API} - 7.556 \times 10^{-5} \times \text{MgSt} - 0.1849 \times \text{LubT} - 3.783 \times 10^{-2} \times \text{Hard} - 2.557 \times 10^{-5} \times \text{MgSt} \times \text{LubT}
\]

- Predictive model derived from analysis of DOE data
- Standard Error for Diss = 0.72
- Design space defined in terms of ranges evaluated during DOE

Design Space:
API: 0.5 - 1.5
MgSt: 3000 – 12000
LubT: 1 – 10
Hard: 60 – 110
Case Study 1: Approach for Uncertainty Estimation

**Method:** Uncertainty estimated using Monte Carlo simulation, using @RISK

Following cases were studied:
- **Case A:** Assumed measurement uncertainty for each of the inputs (i.e. common cause variation), modeled as a normal distribution, **no** uncertainty assumed for the model coefficients
- **Case B:** Assumed measurement uncertainty for each of the inputs and also for the model coefficients, modeled as a normal distribution
Case Study 1: Results for Case A

For the following combination of input variables (API: 1.5, MgSt: 12000, LubT: 10, Hard: 110) there is a 5.7% probability of not meeting the dissolution specification. Thus we have ~94% ‘assurance of quality’, and, there is potential for dissolution failure at edge of design space.
Case Study 1: Results for Case B

For the same combination of input variables as in Case A, assuming uncertainty in model coefficients (s= 1% of mean value) there is a 22.4% probability of not meeting the dissolution specification.

Hence, there is an increase in potential for dissolution failure at edge of design space.

Minimum = 76.1
Maximum = 85.8
Mean = 81.0
St. Dev = 1.3
Case Study #2: Estimating Model Coefficients by Monte Carlo Simulation

• Previous Studies: Estimated model coefficient standard deviations do not predict the observed response uncertainty.

• Can we use Monte Carlo simulation to provide better estimates of model coefficient standard deviations?
  – Solve for the model coefficients using Monte Carlo simulation.
  – Model coefficients are given as distributions.
Estimating Variance of Process Model Regression Coefficients

Model coefficient variance
Covariance matrix

Response variance
(p = # model coefficients
N = # experiments)

\[ \text{Cov}(B) = [D^\dagger D][\text{Cov}(R)]D^\dagger^T \]

\[ \delta_R^2 = \sum_{i=1}^{N} \frac{(R_i - \hat{R}_i)^2}{N - p} \]

Jth Model coefficient variance = Jth diagonal element of Cov(B)

Assumptions: Only \( R \) has uncertainty;

Problems: 1.) We know that \( D \) (matrix of input variables) has uncertainty.
2.) We suspect that uncertainties may be correlated.
Least Squares Solution to a Process Model

Matrix Representation of Process Model: \( R = DB \)

Define the pseudoinverse of \( D \): \( D^\dagger = (D^T D)^{-1} D^T \)

Solving for the Model Coefficients: \( D^\dagger R = B \)

The pseudoinverse solution of a matrix equation gives the least squares best estimates of the \( B \) coefficients!
Case Study 2: Modeling 45 minute dissolution (D45) of a tableting process

• $3^2$ Factorial Experimental Design
  – Granulating Water ($GS$: 36-38 kg)
  – Granulating Power ($P$: 18.5-22.5 kW)

• Nested Compression Factors
  – Compression Force ($CF$: 11.5-17.5 kN)
  – Press Speed ($S$: 70-110 kTPH)

• Least Squares Predictive Model*
  \[ D45 = 68.35 - 1.34(GS) - 2.88(P) - 8.95(CF) + 2.43(GS)^2 \]

* Parameter values are mean-centered and range-scaled.

Publication Reference
Application of Quality by Design Knowledge (QbD) From Site Transfers to Commercial Operations Already in Progress,” J. PAT, Jan/Feb, pg. 8, 2006.
### Measurement Uncertainty and Prediction Uncertainty

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Measured Mean</th>
<th>Model Prediction</th>
<th>Measured St. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>69.0</td>
<td>69.2</td>
<td>3.1</td>
</tr>
<tr>
<td>2</td>
<td>73.5</td>
<td>72.3</td>
<td>3.1</td>
</tr>
<tr>
<td>3</td>
<td>71.9</td>
<td>72.1</td>
<td>1.9</td>
</tr>
<tr>
<td>4</td>
<td>67.1</td>
<td>65.5</td>
<td>2.9</td>
</tr>
<tr>
<td>5</td>
<td>69.8</td>
<td>71.2</td>
<td>1.9</td>
</tr>
<tr>
<td>6</td>
<td>75.5</td>
<td>75.0</td>
<td>0.8</td>
</tr>
<tr>
<td>7</td>
<td>65.4</td>
<td>66.6</td>
<td>2.7</td>
</tr>
<tr>
<td>8</td>
<td>75.8</td>
<td>77.3</td>
<td>1.1</td>
</tr>
<tr>
<td>9</td>
<td>56.1</td>
<td>59.4</td>
<td>3.7</td>
</tr>
<tr>
<td>10</td>
<td>61.0</td>
<td>59.4</td>
<td>3.7</td>
</tr>
<tr>
<td>11</td>
<td>67.2</td>
<td>68.3</td>
<td>2.4</td>
</tr>
<tr>
<td>12</td>
<td>77.0</td>
<td>77.3</td>
<td>1.1</td>
</tr>
<tr>
<td>13</td>
<td>72.7</td>
<td>68.3</td>
<td>2.4</td>
</tr>
</tbody>
</table>

**Benchmark**

<table>
<thead>
<tr>
<th>Standard Error of Prediction</th>
<th>2.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Error (RMS Measurement Standard Deviation)</td>
<td>2.6</td>
</tr>
</tbody>
</table>
Case Study 2: How Do Variances in Process Parameters Influence Model Coefficients?

Regression

"Bias"  Power (P)  Water (GS)  Force (CF)  Water^2
# Estimated Model Coefficient Uncertainties from Monte Carlo Simulation

<table>
<thead>
<tr>
<th>Regression</th>
<th>Monte Carlo Simulation Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#1</td>
</tr>
<tr>
<td>P Std. Dev.</td>
<td></td>
</tr>
<tr>
<td>GS Std. Dev.</td>
<td>-</td>
</tr>
<tr>
<td>CF Std. Dev.</td>
<td>-</td>
</tr>
<tr>
<td>R Std. Dev.</td>
<td>-</td>
</tr>
<tr>
<td>B₀</td>
<td>68.35 ± 0.83</td>
</tr>
<tr>
<td>B_(GS)</td>
<td>-1.34 ± 1.07</td>
</tr>
<tr>
<td>B_(P)</td>
<td>-2.88 ± 0.96</td>
</tr>
<tr>
<td>B_(CF)</td>
<td>-8.95 ± 1.17</td>
</tr>
<tr>
<td>B_(GS^2)</td>
<td>2.44 ± 1.35</td>
</tr>
</tbody>
</table>
Regression versus Monte Carlo Distributions for D45

**Regression**

- Minimum = 65.2
- Maximum = 71.9
- Mean = 68.3
- St. Dev = 0.83

**Monte Carlo (random coeff.)**

- Minimum = 62.6
- Maximum = 74.5
- Mean = 68.4
- St. Dev = 1.44
Monte Carlo with and without measurement error

Monte Carlo (random coeff. plus response error)

Minimum = 60.0
Maximum = 77.7
Mean = 68.4
St. Dev = 2.4

Monte Carlo (random coeff.)

Minimum = 62.6
Maximum = 74.5
Mean = 68.4
St. Dev = 1.44
Conclusion

• Uncertainty analysis is a powerful tool to estimate robustness of a process model
• Monte Carlo simulation of process models facilitates implementation of risk mitigation techniques in the control strategy
• Back up
## Experimentation and Process Modeling

<table>
<thead>
<tr>
<th>D45_{\text{exp 1}}</th>
<th>GS_{\text{exp 1}}</th>
<th>P_{\text{exp 1}}</th>
<th>CF_{\text{exp 1}}</th>
</tr>
</thead>
<tbody>
<tr>
<td>D45_{\text{exp 2}}</td>
<td>GS_{\text{exp 2}}</td>
<td>P_{\text{exp 2}}</td>
<td>CF_{\text{exp 2}}</td>
</tr>
<tr>
<td>D45_{\text{exp 3}}</td>
<td>GS_{\text{exp 3}}</td>
<td>P_{\text{exp 3}}</td>
<td>CF_{\text{exp 3}}</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
</tr>
</tbody>
</table>
Propagation of uncertainty in process model predictions:

All Model Coefficient variances and Process Variable variances contribute to each predicted Response uncertainty in a model-dependent manner.
Case Study #2: Influence of Process Parameter Variation on Prediction

• Model Conditions
  – GS mean = 36 kg
  – P mean = 20 kW
  – CF mean = 14 kN
  – Input parameter standard deviations were varied.
  – Dissolution values were predicted.
A Process Model:  
Matrix Representation

\[
\begin{bmatrix}
D_{45_{\text{pred }1}} \\
D_{45_{\text{pred }2}} \\
D_{45_{\text{pred }3}} \\
\vdots \\
\end{bmatrix}
= 
\begin{bmatrix}
1 & GS_{\exp 1} & P_{\exp 1} & CF_{\exp 1} & GS_{\exp 1}^2 \\
1 & GS_{\exp 2} & P_{\exp 2} & CF_{\exp 2} & GS_{\exp 2}^2 \\
1 & GS_{\exp 3} & P_{\exp 3} & CF_{\exp 3} & GS_{\exp 3}^2 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\end{bmatrix}
\begin{bmatrix}
B_0 \\
B_1 \\
B_2 \\
B_3 \\
B_4 \\
\end{bmatrix}
\]

Response matrix  
Design matrix  
\(\mathbf{B}\) matrix

\[
\mathbf{R} = \mathbf{DB}
\]
Estimating Variance in Prediction:
The Basis for Uncertainty in Design Space

Response Covariance matrix
Cov(R) = B[Cov(D)]B^T

Jth experimental variance = Jth diagonal element of Cov(R)

Assumptions: Only D has uncertainty.

Problems: 1.) We know that B has uncertainty.
2.) We know that uncertainties in D will be correlated, but we don’t know Cov(D)
Example: Simulations 1-4

D45 Simulation
Mean = 74.6%
Std. Dev. = 3.70%

D45 Simulation
Mean = 75.0%
Std. Dev. = 4.59%

D45 Simulation
Mean = 76.9%
Std. Dev. = 7.89%

D45 Simulation
Mean = 76.9%
Std. Dev. = 8.26%

GS Std. Dev.=0.25 kg
P Std. Dev.=1 kW
CF Std. Dev.= 1 kN
D45 Std. Dev.= 0%

GS Std. Dev.=0.5 kg
P Std. Dev.=1 kW
CF Std. Dev.= 1 kN
D45 Std. Dev.= 0%

GS Std. Dev.=1 kg
P Std. Dev.=1 kW
CF Std. Dev.= 1 kN
D45 Std. Dev.= 0%

GS Std. Dev.=1 kg
P Std. Dev.=2 kW
CF Std. Dev.= 1 kN
D45 Std. Dev.= 0%
Case Study 2: Influence of Dissolution Measurement Error

- **Model Conditions**
  - GS mean = 36 kg
  - P   mean = 20 kW
  - CF mean = 14 kN
  - Input parameter standard deviations were varied.
  - Dissolution measurement error was added.
  - Dissolution values were predicted.
Example: Simulations 5-7

D45 Simulation
Mean = 75.0%
Std. Dev. = 3.88%

D45 Simulation
Mean = 75.0%
Std. Dev. = 4.37%

D45 Simulation
Mean = 75.0%
Std. Dev. = 5.58%

D45 Simulation
Mean = 75.0%
Std. Dev. = 7.16%

GS Std. Dev=0.5 kg
P Std. Dev=1 kW
CF Std. Dev=0.5 kN
D45 Std. Dev=0% (Control)

GS Std. Dev=0.5 kg
P Std. Dev=1 kW
CF Std. Dev=0.5 kN
D45 Std. Dev=2%

GS Std. Dev=0.5 kg
P Std. Dev=1 kW
CF Std. Dev=0.5 kN
D45 Std. Dev=4%

GS Std. Dev=0.5 kg
P Std. Dev=1 kW
CF Std. Dev=0.5 kN
D45 Std. Dev=6%
Propagation of Uncertainty in Process Modeling

The pseudoinverse of $D$: \[ D^\dagger = (D^T D)^{-1} D^T \]

Solving for the Model Coefficients: \[ D^\dagger R = B \]

\[
\begin{align*}
B_1 &= D_{11}^\dagger R_{Exp\ 1} + D_{12}^\dagger R_{Exp\ 2} + D_{13}^\dagger R_{Exp\ 3} + \ldots \\
B_2 &= D_{21}^\dagger R_{Exp\ 1} + D_{22}^\dagger R_{Exp\ 2} + D_{23}^\dagger R_{Exp\ 3} + \ldots \\
&\vdots & & \vdots
\end{align*}
\]

1. Assign random variables to Dissolution values ($R$) and use Monte Carlo simulations to propagate error to the model coefficients ($B$).

2. Assign random variables to Process Parameters ($D$) and use Monte Carlo simulations to propagate error to $B$. 


How Do Variances in Process Parameters Influence Model Coefficients?

• Simulation # 1 (1-0.25-1)
  – Measured D45 means and standard deviations.
  – P 19-23 kW ± 1 kW
  – GS 36-38 kg ± 0.25 kg
  – CF 12-18 ± 1 kN

• Compare to regression distributions
  – Model coefficient means
  – Model coefficient standard deviations
Case Study 2: Simulation 1

GS
Mean = 36 kg
Std. Dev. = 0.25 kg

P
Mean = 20 kW
Std. Dev. = 1 kW

CF
Mean = 14 kN
Std. Dev. = 1 kN

D45 Simulation Result
Mean = 74.6%
Std. Dev. = 3.70%

\[ D45 = 68.35 - 1.34(GS) - 2.88(P) - 8.95(CF) + 2.43(GS)^2 \]
Case Study 2: Influence of Process Parameters Variation

• Increase in granulation water mass (GS) variance:
  – Increases predicted D45 variance.
  – Slightly shifts predicted D45 means.
  – Skews the predicted D45 distributions.

• Increase in granulator power (P) endpoint variance:
  – Increases predicted D45 variance.
  – Does not shift predicted D45 means.
  – Does not skew the predicted D45 distributions.
Case Study 2: Influence of Dissolution Measurement Error

- Increase in D45 measurement variance:
  - does not shift predicted D45 means.
  - does not appear to skew predicted D45 distributions.
  - increases predicted D45 variance.

- Advantage #2, we get the distribution, not just the standard deviation.

- Advantage #3, sensitivity analysis allows us to prioritize process improvement.
## Measurement Uncertainty and Prediction Uncertainty

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Measured Mean</th>
<th>Model Prediction</th>
<th>Measured St. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>69.0</td>
<td>69.2</td>
<td>3.1</td>
</tr>
<tr>
<td>2</td>
<td>73.5</td>
<td>72.3</td>
<td>3.1</td>
</tr>
<tr>
<td>3</td>
<td>71.9</td>
<td>72.1</td>
<td>1.9</td>
</tr>
<tr>
<td>4</td>
<td>67.1</td>
<td>65.5</td>
<td>2.9</td>
</tr>
<tr>
<td>5</td>
<td>69.8</td>
<td>71.2</td>
<td>1.9</td>
</tr>
<tr>
<td>6</td>
<td>75.5</td>
<td>75.0</td>
<td>0.8</td>
</tr>
<tr>
<td>7</td>
<td>65.4</td>
<td>66.6</td>
<td>2.7</td>
</tr>
<tr>
<td>8</td>
<td>75.8</td>
<td>77.3</td>
<td>1.1</td>
</tr>
<tr>
<td>9</td>
<td>56.1</td>
<td>59.4</td>
<td>3.7</td>
</tr>
<tr>
<td>10</td>
<td>61.0</td>
<td>59.4</td>
<td>3.7</td>
</tr>
<tr>
<td>11</td>
<td>67.2</td>
<td>68.3</td>
<td>2.4</td>
</tr>
<tr>
<td>12</td>
<td>77.0</td>
<td>77.3</td>
<td>1.1</td>
</tr>
<tr>
<td>13</td>
<td>72.7</td>
<td>68.3</td>
<td>2.4</td>
</tr>
</tbody>
</table>

Standard Error of Prediction: 2.4

Standard Error (RMS Measurement Standard Deviation): 2.6