# 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

# 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), $cm^2/g$

- 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

| Exp No | Run Order | API | MgSt | LubT | Hard | Diss |
|---|---|---|---|---|---|---|
| 1 | 1 | 0.5 | 3000 | 1 | 60 | 101.24 |
| 2 | 14 | 1.5 | 3000 | 1 | 60 | 87.99 |
| 3 | 22 | 0.5 | 12000 | 1 | 60 | 99.13 |
| 4 | 8 | 1.5 | 3000 | 10 | 60 | 86.03 |
| 5 | 18 | 0.5 | 12000 | 10 | 60 | 94.73 |
| 6 | 9 | 1.5 | 12000 | 10 | 60 | 83.04 |
| 7 | 15 | 0.5 | 3000 | 1 | 110 | 98.07 |
| 8 | 2 | 0.5 | 12000 | 1 | 110 | 97.68 |
| 9 | 6 | 1.5 | 12000 | 1 | 110 | 85.47 |
| 10 | 16 | 0.5 | 3000 | 10 | 110 | 95.81 |
| 11 | 20 | 1.5 | 3000 | 10 | 110 | 84.38 |
| 12 | 3 | 1.5 | 12000 | 10 | 110 | 81 |
| 13 | 10 | 0.5 | 7500 | 5.5 | 85 | 96.85 |
| 14 | 17 | 1.5 | 7500 | 5.5 | 85 | 85.13 |
| 15 | 19 | 1 | 3000 | 5.5 | 85 | 91.87 |
| 16 | 21 | 1 | 12000 | 5.5 | 85 | 90.72 |
| 17 | 7 | 1 | 7500 | 1 | 85 | 91.95 |
| 18 | 4 | 1 | 7500 | 10 | 85 | 88.9 |
| 19 | 5 | 1 | 7500 | 5.5 | 60 | 92.37 |
| 20 | 11 | 1 | 7500 | 5.5 | 110 | 90.95 |
| 21 | 12 | 1 | 7500 | 5.5 | 85 | 91.95 |
| 22 | 13 | 1 | 7500 | 5.5 | 85 | 90.86 |
| 23 | 23 | 1 | 7500 | 5.5 | 85 | 89 |

# Case Study 1: Problem Description

**Objective**: Evaluate the response variance in the predictive model

**Prediction algorithm:**
$$\text{Diss} = 108.9 - 11.96 \times \text{API} - 7.556\times10^{-5} \times \text{MgSt} - 0.1849 \times \text{LubT} - 3.783\times10^{-2} \times \text{Hard} - 2.557\times10^{-5} \times \text{MgSt} \times \text{LubT}$$

- 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

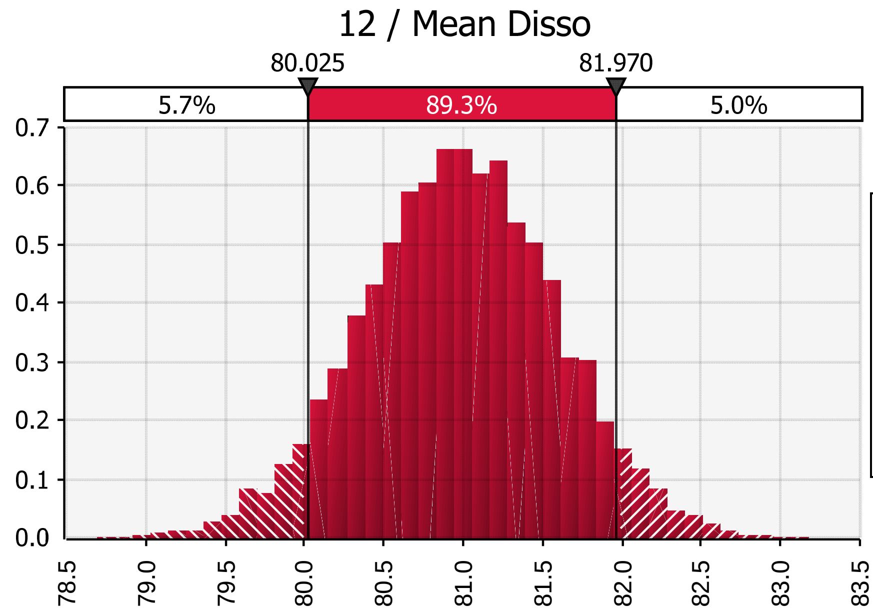
# Case Study 1: Approach for Uncertainty Estimation

**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, <u>no</u> 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




Minimum = 78.7

Maximum = 83.2

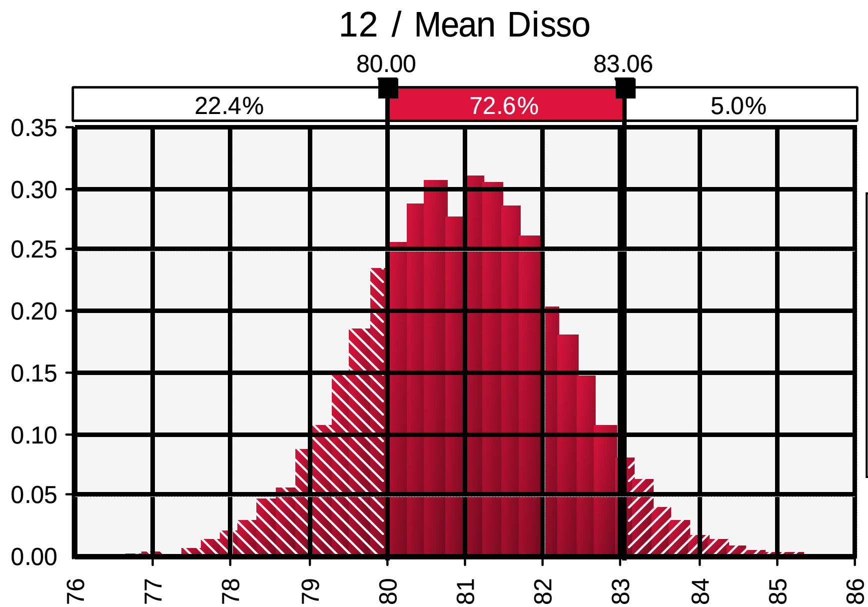
Mean = 81.0

St. Dev = 0.60

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




Minimum = 76.1

Maximum = 85.8

Mean = 81.0

St. Dev = 1.3

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

# 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
Covariance matrix

$$\mathrm{Cov}(\mathbf{B}) = [\mathbf{D}^T\mathbf{D}]^{-1}\delta_R^2$$

$$\mathrm{Cov}(\mathbf{B}) = \mathbf{D}^{\dagger}[\mathrm{Cov}(\mathbf{R})]\mathbf{D}^{\dagger T}$$

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

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

$J^{th}$ Model coefficient variance = $J^{th}$ 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: $\mathbf{R = DB}$

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

Solving for the Model Coefficients: $\mathbf{D^{\dagger}R = B}$

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**

$$D45 = 68.35 - 1.34(GS) - 2.88(P) - 8.95(CF) + 2.43(GS)^2$$

* **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

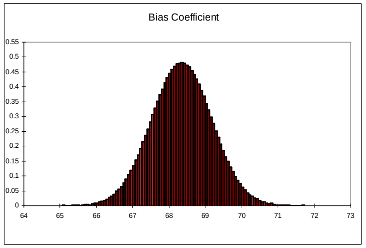
| | D45(%) | | |
|---|---|---|---|
| **Experiment** | **Measured Mean** | **Model Prediction** | **Measured St. Dev.** |
| 1 | 69.0 | 69.2 | 3.1 |
| 2 | 73.5 | 72.3 | 3.1 |
| 3 | 71.9 | 72.1 | 1.9 |
| 4 | 67.1 | 65.5 | 2.9 |
| 5 | 69.8 | 71.2 | 1.9 |
| 6 | 75.5 | 75.0 | 0.8 |
| 7 | 65.4 | 66.6 | 2.7 |
| 8 | 75.8 | 77.3 | 1.1 |
| 9 | 56.1 | 59.4 | 3.7 |
| 10 | 61.0 | 59.4 | 3.7 |
| 11 | 67.2 | 68.3 | 2.4 |
| 12 | 77.0 | 77.3 | 1.1 |
| 13 | 72.7 | 68.3 | 2.4 |
| **Standard Error of Prediction** | | **2.4** | |
| **Standard Error (RMS Measurement Standard Deviation)** | | | **2.6** |

Benchmark

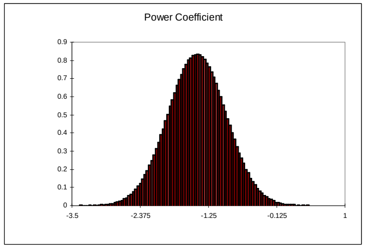
# Case Study 2: How Do Variances in Process Parameters Influence Model Coefficients?

Regression

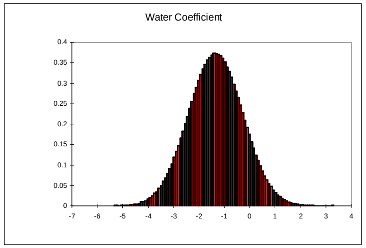
"Bias"



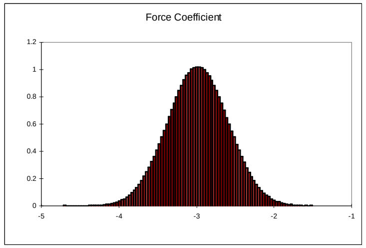

Power (P)



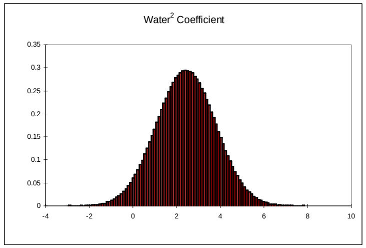

Water (GS)




Force (CF)




Water$^2$

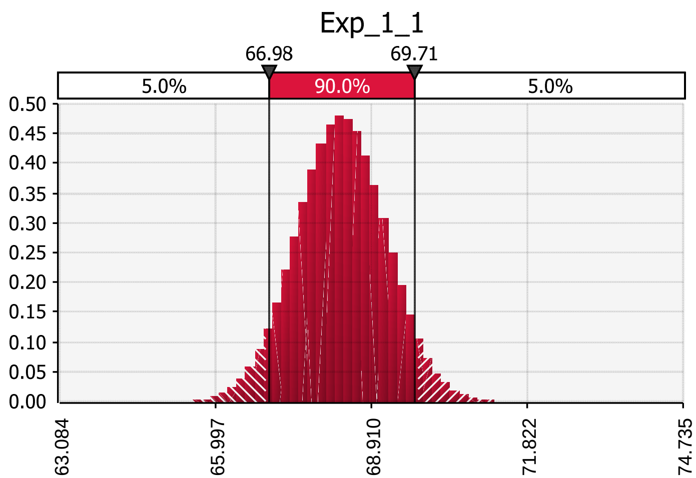


# Estimated Model Coefficient Uncertainties from Monte Carlo Simulation

| | | Monte Carlo Simulation Parameters | | | |
|---|---|---|---|---|---|
| | **Regression** | **#1** | **#2** | **#3** | **#4** |
| **P Std. Dev.** | - | - | 1 | 1 | 0.5 |
| **GS Std. Dev.** | - | - | 0.25 | 0.25 | 0.25 |
| **CF Std. Dev.** | - | - | 1 | 1 | 0.5 |
| **R Std. Dev.** | - | measured | - | measured | 2.5% |
| $\mathbf{B_0}$ | 68.35 ± 0.83 | 68.35 ± 0.93 | 68.41 ± 1.06 | 68.39 ± 1.44 | 68.39 ± 0.9 |
| $\mathbf{B_{(GS)}}$ | -1.34 ± 1.07 | -1.34 ± 1.21 | -1.23 ± 1.49 | -1.22 ± 1.96 | -1.24 ± 1.28 |
| $\mathbf{B_{(P)}}$ | -2.88 ± 0.96 | -2.88 ± 1.04 | -2.13 ± 1.18 | -2.11 ± 1.53 | -2.63 ± 1.06 |
| $\mathbf{B_{(CF)}}$ | -8.95 ± 1.17 | -8.95 ± 1.38 | -7.37 ± 1.35 | -7.37 ± 1.87 | -8.52 ± 1.2 |
| $\mathbf{B_{(GS\hat{}2)}}$ | 2.44 ± 1.35 | 2.44 ± 1.52 | 2.13 ± 1.84 | 2.14 ± 2.43 | 2.15 ± 1.57 |

# Regression versus Monte Carlo Distributions for D45

## Regression



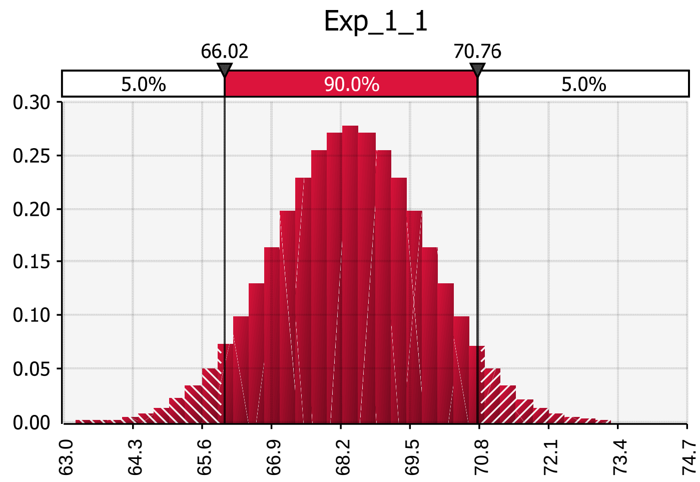

Minimum = 65.2

Maximum = 71.9

Mean = 68.3

St. Dev = 0.83

## Monte Carlo (random coeff.)




Minimum = 62.6

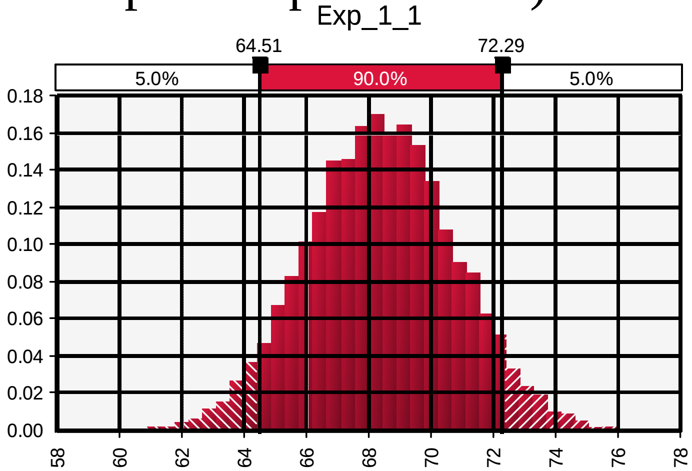
Maximum = 74.5

Mean = 68.4

St. Dev = 1.44

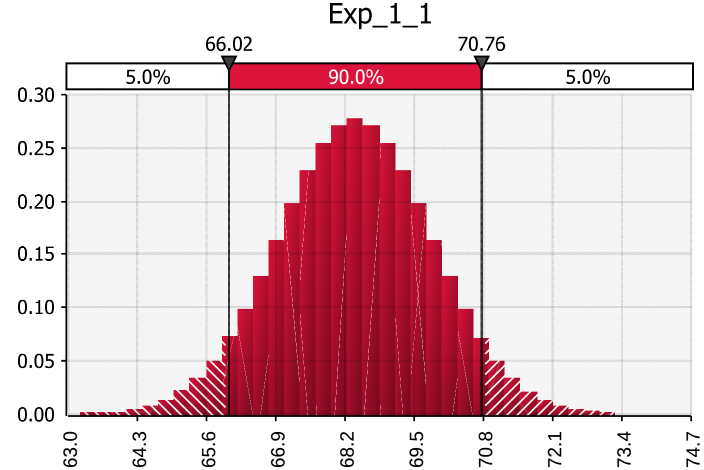
# Monte Carlo with and without measurement error

## 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

| | | | |
|---|---|---|---|
| $D45_{exp\ 1}$ | $GS_{exp\ 1}$ | $P_{exp\ 1}$ | $CF_{exp\ 1}$ |
| $D45_{exp\ 2}$ | $GS_{exp\ 2}$ | $P_{exp\ 2}$ | $CF_{exp\ 2}$ |
| $D45_{exp\ 3}$ | $GS_{exp\ 3}$ | $P_{exp\ 3}$ | $CF_{exp\ 3}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |

# Experimentation and Process Modeling

$$D45_{pred\ 1} = B_0 + B_1 \cdot GS_{exp\ 1} + B_2 \cdot P_{exp\ 1} + B_3 \cdot CF_{exp\ 1} + B_4 \cdot GS^2_{exp\ 1}$$

**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

$$\begin{pmatrix} D45_{pred\ 1} \\ D45_{pred\ 2} \\ D45_{pred\ 3} \\ \vdots \end{pmatrix} = \begin{pmatrix} 1 & GS_{exp\ 1} & P_{exp\ 1} & CF_{exp\ 1} & GS^2{}_{exp\ 1} \\ 1 & GS_{exp\ 2} & P_{exp\ 2} & CF_{exp\ 2} & GS^2{}_{exp\ 2} \\ 1 & GS_{exp\ 3} & P_{exp\ 3} & CF_{exp\ 3} & GS^2{}_{exp\ 3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} B_0 \\ B_1 \\ B_2 \\ B_3 \\ B_4 \end{pmatrix}$$

**R**esponse matrix | **D**esign matrix | **B** matrix

**R = DB**

# Estimating Variance in Prediction: The Basis for Uncertainty in Design Space

Response Covariance matrix

$$\mathrm{Cov}(\mathbf{R}) = \mathbf{B}[\mathrm{Cov}(\mathbf{D})]\mathbf{B}^{\mathbf{T}}$$

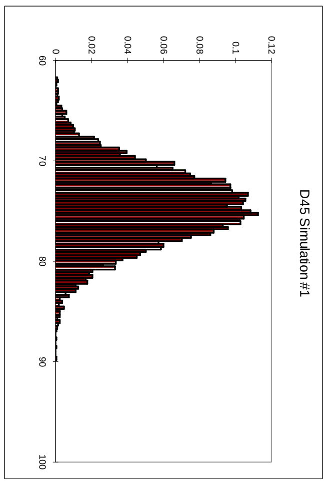
$J^{th}$ experimental variance = $J^{th}$ diagonal element of $\mathrm{Cov}(\mathbf{R})$

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**)

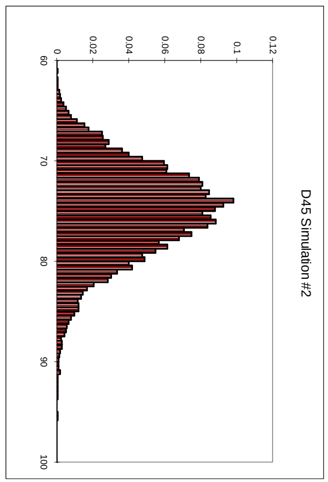
# Example: Simulations 1-4

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



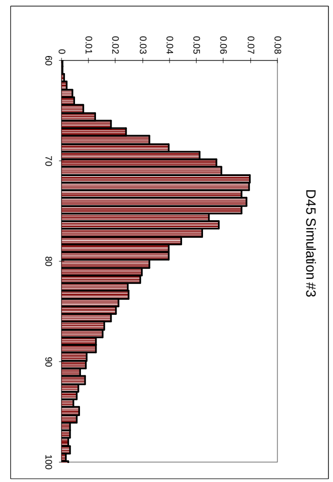

**GS Std. Dev.=0.25 kg**
P Std. Dev.=1 kW
CF Std. Dev.= 1 kN
D45 Std. Dev.= 0%

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



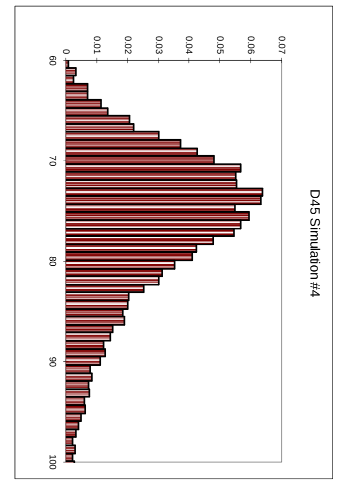

**GS Std. Dev.=0.5 kg**
P Std. Dev.=1 kW
CF Std. Dev.= 1 kN
D45 Std. Dev.= 0%

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




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

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



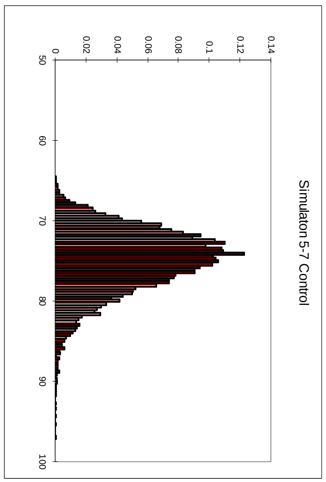

GS Std. Dev.=1 kg
**P Std. Dev.=2 kW**
CF Std. Dev.= 1 kN
D45 Std. Dev.= 0%

# 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.

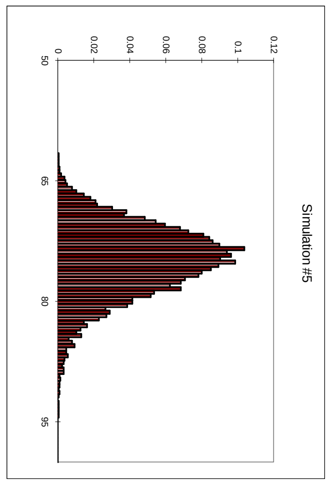
# Example: Simulations 5-7

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



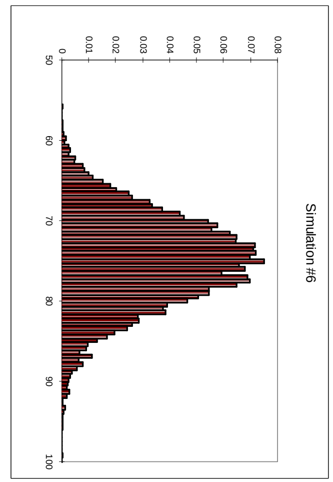

GS Std. Dev=0.5 kg
P Std. Dev=1 kW
CF Std. Dev=0.5 kN
**D45 Std. Dev=0%**
(Control)

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



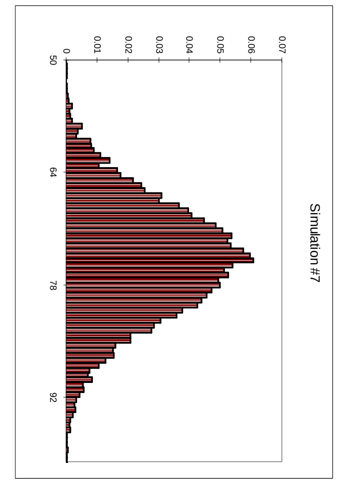

GS Std. Dev=0.5 kg
P Std. Dev=1 kW
CF Std. Dev=0.5 kN
**D45 Std. Dev=2%**

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




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

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=6%**

# Propagation of Uncertainty in Process Modeling

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

Solving for the Model Coefficients: $\mathbf{D^{\dagger}R=B}$

$$B_1 = D^{\dagger}_{11}\cdot R_{\mathrm{Exp\ 1}} + D^{\dagger}_{12}\cdot R_{\mathrm{Exp\ 2}} + D^{\dagger}_{13}\cdot R_{\mathrm{Exp\ 3}} + \ldots$$

$$B_2 = D^{\dagger}_{21}\cdot R_{\mathrm{Exp\ 1}} + D^{\dagger}_{22}\cdot R_{\mathrm{Exp\ 2}} + D^{\dagger}_{23}\cdot R_{\mathrm{Exp\ 3}} + \ldots$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

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

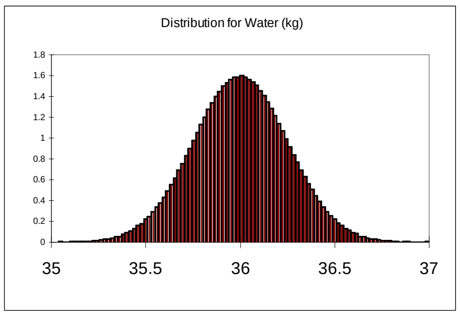
2. Assign random variables to **Process Parameters (D)** and use Monte Carlo simulations to propagate error to **B**.

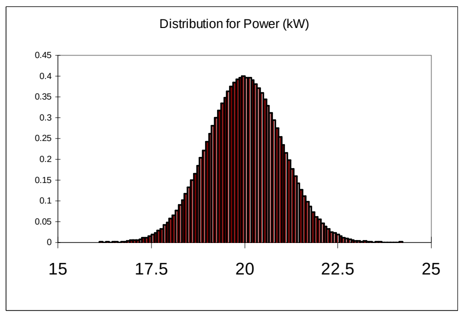
# How Do Variances in Process Parameters Influence Model Coefficients?

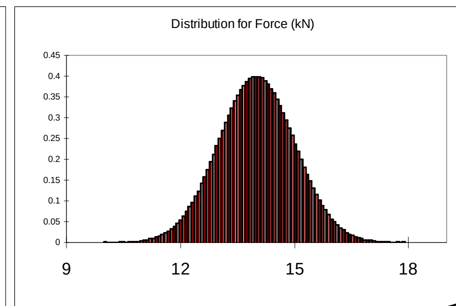
- Simulation # 1 (1-0.25-1)
  - Measured D45 means and **standard deviations**.
  - P 19-23 kW **± 1 kW**
  - GS 36-38 kg **± 0.25 kg**
  - CF 12-18 **± 1 kN**

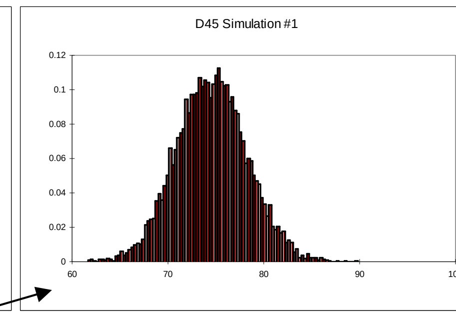
- Compare to regression distributions
  - Model coefficient means
  - Model coefficient standard deviations

# 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%

$$\mathbf{D45 = 68.35 - 1.34(GS) - 2.88(P) - 8.95(CF) + 2.43(GS)^2}$$

# 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

| Experiment | D45(%) Measured Mean | Model Prediction | Measured St. Dev. |
|---|---|---|---|
| 1 | 69.0 | 69.2 | 3.1 |
| 2 | 73.5 | 72.3 | 3.1 |
| 3 | 71.9 | 72.1 | 1.9 |
| 4 | 67.1 | 65.5 | 2.9 |
| 5 | 69.8 | 71.2 | 1.9 |
| 6 | 75.5 | 75.0 | 0.8 |
| 7 | 65.4 | 66.6 | 2.7 |
| 8 | 75.8 | 77.3 | 1.1 |
| 9 | 56.1 | 59.4 | 3.7 |
| 10 | 61.0 | 59.4 | 3.7 |
| 11 | 67.2 | 68.3 | 2.4 |
| 12 | 77.0 | 77.3 | 1.1 |
| 13 | 72.7 | 68.3 | 2.4 |
| **Standard Error of Prediction** | | **2.4** | |
| **Standard Error (RMS Measurement Standard Deviation)** | | | **2.6** |

Benchmark