Design of Experiments in Ecological Systems: some methods and issues

Jorge Luis Romeu, Ph.D.
Research Professor, Syracuse University
Email: jlromeu@syr.edu
Dept. Mechanical & Aerospace Engineering, and EQS STAR Center for Environmental Quality Seminar Series.
February 27, 2009
Outline

• Problem statement
• Proposed solution
• Simulation Example
• Some applicable DOEs
• Implementation problems
  – and their consequences
• Other modeling issues
• Conclusions
Problem Statement

• Given a network of water masses
  – E.g. Finger Lakes, Great Lakes, Gulf Ports
• Optimize some performance measures:
  – Wetland Preservation; water use; exports, etc
• Subject to the Key set of economic, social, labor, political, environmental, climatic, cultural, etc. problem variables/constraints
• Maintaining employment, health etc. levels
Modeling Methods

• Theoretical (physics law, or relation)
  – But, can we come up with such equation?

• Empirical (regression)
  – But, can we find enough data to implement?

• Discrete Event Simulation
  – Don’t need to relax model assumptions
  – Can include complex interactions
  – But run time can be very long!
Simulation Modeling Problem

• Consider a Complex Simulation Model

• Main Issues:
  – Too many variables to analyze
  – Complex Dynamical System structure

• Proposed solution: Find Key Variables
  – Via Design of Experiments (DOE)
  – Derive a set of simpler Meta Models
  – Use them as proxies for the Full Model
Simulation Example

• Given a network of water masses
  – For both, civilian and industrial use

• Optimize some performance measures
  – e.g. operational, social, political, ecological

• Subject to a set of (conflicting) political, labor, socio-economic, etc. constraints
  – Maintaining levels of production, employment
  – Tax revenues, social services, economic, etc.
A Network of Interconnected Water Masses

Lakes, Rivers, etc

Domestic Consumption

Water Table
Example: River Port w/Lagoon

Schematic of the River Port and Lagoon aquatic ecosystem.
Controlled Variables: Economic

- Replenishing Levels (MIN)
- Reservoir Capacity (MAX)
- Ordering Schedule
  - Transfer Policy
  - Usage Policy
  - Shortage Policy
  - Profitability
- System’s Initial Conditions
Controlled Variables: Social

- Allocation to each sector
- Size of the Reservoirs
  - Transfer Policy
- Generation of electricity
- Hospitals and schools
  - Transportation uses
  - Recreation uses
Controlled Variables: Ecologic

- Wetland Area
- Wetland Depth
- Transfer Speed
- Water Table Use
- Pollution Level
- Fish/Foul Population
WetLand v. Level
Uncontrolled Variables

- ECONOMIC
  - Political issues
  - Labor issues
  - Water Theft
  - Water Leaks
  - Markets
  - Financial

- ECOLOGIC
  - Evaporation
  - Temperature
  - Salinity
  - Reproduction
  - Weather
  - Water Table
And Associated Costs

- Of Importing Water from other places
- Transferring from Social to Economic
- Allocation to various constituencies
- Of Water shortages and rationing
- Indirect costs (labor, political, social)
- Ecological costs (degradation, loss)
- Total costs (compound response)
Additional Model Uses

Multi-criteria (ecological, social, economic, etc.) system responses (consolidating elements in the system) can be obtained, by combining (say k) contrasting and competing individual responses into a single, complex one. The (linear) combinations formed quantify the contrasting policies and philosophies of the different constituencies. Comparisons of competing and contrasting policies, via the simulation model results, can help diverse constituencies to rationally discuss their differences, and better reach a consensus.
Simple DOE Example

Complete Factorial Experiment for the Simulation

River Port Capacity

- 636
- 680
  - One

- 581
  - Two

- 502

- 1146

- 1096

- 688
  - Rain
  - A/3
  - Water Transfer Policy
  - A/2

- 744

Seasonality

Response: Total Cost
DOE Model Limitations:

• Analyzes limited variables (here, k=3)
  – For, $2^K$ Factors/Interacts are generated
• The Effect of Interaction, when $k > 2$
  – Can affect results, if present and strong
• Need to find Robust Responses
  – Handling specific “noise variables”
• Need to Identify “significant few” variables
  – To reduce model Size, maintaining Info level.
And their Consequences …

- If large number of factors to analyze
  - Strong factor interaction may exist
  - Dependent on the model structure
  - Requires special methods for analysis
- Different objective of models derived:
  - To describe/study, forecast or control
- Robust Parameter analysis capability
  - To derive a response equation that is
  - Resilient to “noise” or uncontrolled factors
Methods for Key Variable Id

- Full Factorial Designs
- Fractional Factorial Designs
- Plackett-Burnam Designs
- Controlled Sequential Bifurcation
- Latin Hypercube Sampling
- Other modeling approaches
  - Bayesian, Hierarchical, Taguchi, PCA, etc.
Full Factorials

- Most expensive (in time and effort)
  - Prohibitive with current number of factors
- Most comprehensive information
  - Provides info on all factor interactions
- Two Examples with a $2^3$ Full Factorial
  - First case: mild interaction (AB only)
  - Second: strong and complex interaction
  - Notice how the Model-Estimations vary
2^3 Full Factorial DOE: Variables Used

- A = Replenishing Levels (MIN)
- B = Reservoir Capacity (MAX)
- C = Transfer Policy

Mild interaction assumed

- A* B only
**Full Factorial Experiment 2^3**

<table>
<thead>
<tr>
<th>Run</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>AB</th>
<th>AC</th>
<th>BC</th>
<th>ABC</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1.07</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>3.72</td>
</tr>
<tr>
<td>3</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-0.58</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>12.04</td>
</tr>
<tr>
<td>5</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>7.75</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>15.45</td>
</tr>
<tr>
<td>7</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>11.09</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>18.31</td>
</tr>
</tbody>
</table>

TotSum: 66.71

Effect: 8.08  3.75  9.62  1.84  -0.62  -0.65  -2.08

**Regression Estimations**

<table>
<thead>
<tr>
<th>RegCoef</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>AB</th>
<th>AB Estimat.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.04</td>
<td>1.88</td>
<td>4.81</td>
<td>0.92</td>
<td>8.34</td>
</tr>
<tr>
<td>TRUE</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>10</td>
</tr>
</tbody>
</table>

**Meta Model:** $Y_{ijkl} = 8.33 + 4.04A + 1.88B + 4.81C + 0.92AB$

**True Model:** $Y = 10 + 4*A + 2*B + 5*C + AB + \varepsilon$

Mild Interaction (AB only)
Meta Model Re-creation ability: mild interaction.
Fractional Factorials

• Analyzes only a Fraction of the Full
  – Reduces substantially time/effort
  – Confounding of Main Effects/Interactions
  – If Interactions present, this is a problem
  – Only for Powers of Two (no. of runs)

• Numerical Example: half fractions
  – Of the previous Full Factorial – and others
  – Assess Model-Estimation agreement
## Fractional Factorials

### First Fraction: L1

<table>
<thead>
<tr>
<th>Run</th>
<th>A</th>
<th>B</th>
<th>C=AB</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-0.33</td>
</tr>
<tr>
<td>2</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-0.33</td>
</tr>
<tr>
<td>3</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-0.33</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**TotSum** 0.00  
**Effect** 7.429  3.130  11.460  
**Signif.** No  No  Yes

### Second Fraction: L2

<table>
<thead>
<tr>
<th>Run</th>
<th>A</th>
<th>B</th>
<th>C=AB</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1.00</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>0.33</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>0.33</td>
</tr>
<tr>
<td>4</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>0.33</td>
</tr>
</tbody>
</table>

**TotSum** 0.00  
**Effect** 8.728  4.375  7.784  
**Signif.** Yes  No  Yes

### Untangling the Confounded Structure

\[
\frac{(L1+L2)}{2} = 8.079 \quad 3.753 \quad 9.622 \\
\frac{(L1-L2)}{2} = -0.649 \quad -0.623 \quad 1.838 \\
\text{Effects} = 8 \quad 4 \quad 10
\]

**Notice how, by averaging both Half Fraction results, we obtain the Full Factorial results again.**

### True Model:

\[
Y_1 = 7.3 + 3.71A + 1.57B + 5.73C^* \\
Y_2 = 8.33 + 4.36A + 2.18B + 3.89C^* \\
C^*: \text{Factor C is confounded with AB}
\]

**True Model:** \[ Y = 10 + 4*A + 2*B + 5*C + AB + \varepsilon \]
Again $2^3$ Full Factorial: Same Variables Used, but now With Strong Interaction

- $A =$ Replenishing Levels (MIN)
- $B =$ Reservoir Capacity (MAX)
- $C =$ Transfer Policy
- Strong interaction assumed
  - $A*B$, $A*C$, $B*C$
  - Overall: $A*B*C$
### Full Factorial: Complex, Strong Interaction

#### Model Parameters

<table>
<thead>
<tr>
<th>Variables</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>AB</th>
<th>AC</th>
<th>BC</th>
<th>ABC</th>
</tr>
</thead>
<tbody>
<tr>
<td>RegCoef</td>
<td>3</td>
<td>-5</td>
<td>1</td>
<td>-12</td>
<td>8</td>
<td>-10</td>
<td>-15</td>
</tr>
<tr>
<td>RegEstim</td>
<td>1.94</td>
<td>-4.38</td>
<td>1.73</td>
<td>-12.14</td>
<td>7.34</td>
<td>-10.52</td>
<td>-15.26</td>
</tr>
<tr>
<td>MainEffcts</td>
<td>6</td>
<td>-10</td>
<td>2</td>
<td>-24</td>
<td>16</td>
<td>-20</td>
<td>-30</td>
</tr>
</tbody>
</table>

| Var. of Model | 12.5173 | StdDv | 3.53799 |
| Var. of Effect | 2.0862 | StdDv | 1.44437 |
| Student T (0.025DF) | 2.47287 |
| C.I. Half Width | 3.57177 |

#### Significance

<table>
<thead>
<tr>
<th>Factor</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>AB</th>
<th>AC</th>
<th>BC</th>
<th>ABC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signific.</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

#### True Model and Estimated Meta Model:

\[ Y = 3A - 5B + C - 12AB + 8AC - 10BC - 15ABC \]

\[ \text{RegEstim} = 1.94A - 4.38B + 1.73C - 12.14AB + 7.34AC - 10.52BC - 15.26ABC \]
## Corresponding Half Fractions

<table>
<thead>
<tr>
<th>Run</th>
<th>A</th>
<th>B</th>
<th>C=AB</th>
<th>Y1</th>
<th>Y2</th>
<th>Y3</th>
<th>Avg.</th>
<th>Var</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-15.03</td>
<td>-16.54</td>
<td>-16.04</td>
<td>-15.87</td>
<td>0.59</td>
<td>-14</td>
</tr>
<tr>
<td>3</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>7.18</td>
<td>9.21</td>
<td>5.28</td>
<td>7.22</td>
<td>3.87</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-16.75</td>
<td>-19.75</td>
<td>-22.02</td>
<td>-19.51</td>
<td>6.97</td>
<td>-22</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-31.61</td>
<td>-27.62</td>
<td>-33.04</td>
<td>-30.76</td>
<td>7.89</td>
<td>-30</td>
</tr>
<tr>
<td>TotSum</td>
<td></td>
<td></td>
<td></td>
<td>-56.21</td>
<td>-54.7</td>
<td>-65.82</td>
<td>-58.91</td>
<td>19.32</td>
<td></td>
</tr>
</tbody>
</table>

For the **Effect** column:
- 1st Run: -17.17, 5.92, -20.81
- 2nd Run: 24.92, -23.44, 27.75

**Signif.**:
- Yes, Yes, Yes

For the **ModlVar.** column:
- 1st Run: 4.83
- 2nd Run: 20.2

For the **StdDev=** column:
- 1st Run: 2.2
- 2nd Run: 4.49

For the **EffVar** column:
- 1st Run: 3.49
- 2nd Run: 7.14

For the **CI-HW=** column:
- 1st Run: 2.2
- 2nd Run: 4.49

<table>
<thead>
<tr>
<th>(a+b)/2</th>
<th>Coefs</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.88</td>
<td>6</td>
</tr>
<tr>
<td>-8.76</td>
<td>-10</td>
</tr>
<tr>
<td>3.47</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(a-b)/2</th>
<th>Coefs</th>
</tr>
</thead>
<tbody>
<tr>
<td>-21.05</td>
<td></td>
</tr>
<tr>
<td>14.68</td>
<td></td>
</tr>
<tr>
<td>-24.28</td>
<td></td>
</tr>
</tbody>
</table>

**MainEff**
- "C"

**Interact**
- C=AB

**NOTE:** FRACTIONAL FACTORIAL RESULTS, GIVEN THE STRONG INTERACTIONS, ARE POOR.
Plackett-Burnam (PB) DOEs

- A Fractional Factorial (FF) DOE
- Analyses “holes” between adjacent FFs
- Reduces time/effort, considerably
- Confounding of Main Effects/Interactions
- Numerical Example: 11 main effects
  - Compare PB to a $2^{11}$ Full Factorial
  - Not all Interactions are strong/significant
- Counter Example: strong interactions
Plackett-Burnam w/o Interaction

- A=Replenishing Levels (MIN)
- B=Reservoir Capacity (MAX)
  - C=Ordering Schedule
  - D=Transfer Policy
- E=Allocation to each sector
- F=Size of the Reservoirs
- G=Generation of electricity
- H=Hospitals and schools
  - I=Wetland size
  - J=Water Table
- K=Fish/Foul Population
# Placket-Burnam Design (no interaction)

<table>
<thead>
<tr>
<th>Run</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>Avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<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>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>3</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>7</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>8</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>9</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>11</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Factors</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>Bo</th>
</tr>
</thead>
<tbody>
<tr>
<td>RegCoef</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>-4</td>
<td>-6</td>
<td>0</td>
<td>-2</td>
<td>4</td>
<td>8</td>
<td>-8</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>RegEst.</td>
<td>4.5</td>
<td>2.3</td>
<td>-0.1</td>
<td>-4.3</td>
<td>-3.6</td>
<td>1.4</td>
<td>-0.8</td>
<td>5.2</td>
<td>6.1</td>
<td>-7.6</td>
<td>-2.8</td>
<td>12.2</td>
</tr>
<tr>
<td>MainEff</td>
<td>12</td>
<td>4</td>
<td>0</td>
<td>-8</td>
<td>-12</td>
<td>0</td>
<td>-4</td>
<td>8</td>
<td>16</td>
<td>-16</td>
<td>0</td>
<td>n/a</td>
</tr>
<tr>
<td>EstimEff</td>
<td>9.1</td>
<td>4.7</td>
<td>-0.2</td>
<td>-8.6</td>
<td>-7.2</td>
<td>2.8</td>
<td>-1.5</td>
<td>10.4</td>
<td>12.3</td>
<td>-15.2</td>
<td>-5.6</td>
<td>12.2</td>
</tr>
<tr>
<td>Signific.</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Meta Model Forecasting Ability
Actual by Predicted Plot

Response Actual vs. Response Predicted

-20 -10 0 10 20 30 40 50
-20 -10 0 10 20 30 40 50

Response Predicted P<.0001
RSq=0.90 RMSE=5.1622
Plackett-Burnam with Interaction

- A=Replenishing Levels (MIN)
- B=Reservoir Capacity (MAX)
  - C=Ordering Schedule
  - D=Transfer Policy
- E=Allocation to each sector
- F=Size of the Reservoirs
- G=Generation of electricity
- H=Hospitals and schools
  - I=Wetland size
  - J=Water Table
- K=Fish/Foul Population
Model with Moderate Interaction structure:

<table>
<thead>
<tr>
<th>Factors</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>Bo</th>
</tr>
</thead>
<tbody>
<tr>
<td>RegCoef</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>-4</td>
<td>-6</td>
<td>0</td>
<td>-2</td>
<td>4</td>
<td>8</td>
<td>-8</td>
<td>0</td>
<td>12</td>
</tr>
</tbody>
</table>


Plackett-Burnam (n=12 rows) Analysis Results:

<table>
<thead>
<tr>
<th>Factors</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>MainEff</td>
<td>12</td>
<td>4</td>
<td>0</td>
<td>-8</td>
<td>-12</td>
<td>0</td>
<td>-4</td>
<td>8</td>
<td>16</td>
<td>-16</td>
<td>0</td>
</tr>
<tr>
<td>FacEstim</td>
<td>-98.6</td>
<td>61.1</td>
<td>41.3</td>
<td>-86.5</td>
<td>98.4</td>
<td>66.4</td>
<td>79.7</td>
<td>51.8</td>
<td>-26.6</td>
<td>37.6</td>
<td>-96.0</td>
</tr>
<tr>
<td>RegPar.</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>-4</td>
<td>-6</td>
<td>0</td>
<td>-2</td>
<td>4</td>
<td>8</td>
<td>-8</td>
<td>0</td>
</tr>
<tr>
<td>RegEstim</td>
<td>-49.3</td>
<td>30.5</td>
<td>20.6</td>
<td>-43.2</td>
<td>49.2</td>
<td>33.2</td>
<td>39.8</td>
<td>25.9</td>
<td>-13.3</td>
<td>18.8</td>
<td>-48.0</td>
</tr>
<tr>
<td>Signific.</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Results are seriously confounded and numerically erroneous.
Controlled Sequential Bifurcation

• Method to Identify significant Main Effects
• Requires prior knowledge of Effect signs
  – To ensure all effects are in same direction
  – Requirement is unrealistic in most cases
• Branch and Bound-like approach
  – Top-Down approach most often
• Adaptive procedure to assess estimations
  – Using the approach but not the method
There are two groups of significant variables after Plackett-Burnam:

**Positive**: B, C, E, F, G, H, J;

and

**Negative**: A, D, I, K.

We Perform a Resolution IV FF
To one of the two groups
Plackett-Burnam Result
Group of “Positive” Vars: B, C, E, F, G, H, J;

- B=Reservoir Capacity (MAX)
- C=Ordering Schedule
- E=Allocation to each sector
- F=Size of the Reservoirs
- G=Generation of electricity
- H=Hospitals and schools
  - J=Water Table
Performing a Resolution IV FF to the “Positive” group: B, C, E, F, G, H, J

<table>
<thead>
<tr>
<th>Factors</th>
<th>B</th>
<th>C</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>J</th>
<th>Bo</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRUE</td>
<td>12</td>
<td>4</td>
<td>0</td>
<td>-8</td>
<td>-12</td>
<td>0</td>
<td>-4</td>
<td>12</td>
</tr>
<tr>
<td>EffectEstim</td>
<td>12.14</td>
<td>2.53</td>
<td>1.17</td>
<td>-7.20</td>
<td>-11.82</td>
<td>0.39</td>
<td>-3.49</td>
<td>13.59</td>
</tr>
<tr>
<td>RegCoef</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>-4</td>
<td>-6</td>
<td>0</td>
<td>-2</td>
<td>12</td>
</tr>
<tr>
<td>RegEst.</td>
<td>6.07</td>
<td>1.26</td>
<td>0.59</td>
<td>-3.60</td>
<td>-5.91</td>
<td>0.19</td>
<td>-1.75</td>
<td>6.80</td>
</tr>
<tr>
<td>Signific.</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

Notice how, once all the Plackett-Burnam (erroneously estimated) variables of the “same sign” were re-analyzed as a sub-group, estimations became closer to True values, both in sign and in magnitude.
Descriptive ability of the model improves; But its Forecasting capability deteriorates.
Latin Hypercube Sampling

• Multiple regression analysis approach
  – Sampling at “best” points in sample space

• Regression selection methods
  – To obtain most efficient Meta Model set

• Provides a list of Alternative Meta Models
  – Some, not as efficient -but close enough
  – Their factors can be “controlled” by the user

• Very effective modeling approach.
Latin Hypercube Example

Assume we have a three dimensional (p = 3) problem in variables B, I, J (reservoir capacity; wetland size and water table use) and that these are respectively distributed Normal, Uniform and Exponential,. Assume that we want to draw a random sample of size n = 10. Divide each variable, according to its probability distribution, into ten equi-probable segments (Prob. = 0.1 = 1/10), identifying each segment with integers 1 through 10.

Then, draw a random variate (r.v.) from each of the ten segments, for each of the three variables B, I, J. Finally, obtain the 10! permutations of integers 1 through 10. Randomly assign one of such permutations (e.g. segments 2,1,5,4,6,9,8,10,7 for B), to each of the variables, select the corresponding segment r.v., and form the vector sample, as below:

<table>
<thead>
<tr>
<th>Sample</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>5th</th>
<th>6th</th>
<th>7th</th>
<th>8th</th>
<th>9th</th>
<th>10th</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>9</td>
<td>8</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>I</td>
<td>4</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td>5</td>
<td>9</td>
<td>10</td>
<td>8</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>J</td>
<td>8</td>
<td>6</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>9</td>
<td>10</td>
</tr>
</tbody>
</table>
Latin Hypercube Example

• Air Force Iraq Simulation Model
  – Fifty plus model variables
  – Two different responses of interest
  – Identify the Key or Relevant Few
  – Preserve as much Info as possible

• Analysis Results
  – Three Key Variables were identified
  – Ninety Percent of the Info ($R^2 = 0.9$)
Regression Selection Analyses Results

SUMMARY OUTPUT

<table>
<thead>
<tr>
<th>Regression Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>R Square</td>
</tr>
<tr>
<td>Observations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coef</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.034007</td>
</tr>
<tr>
<td>X Variable 1</td>
<td>-3.00E-05</td>
</tr>
<tr>
<td>X Variable 2</td>
<td>3.94E-05</td>
</tr>
</tbody>
</table>

SUMMARY OUTPUT:

<table>
<thead>
<tr>
<th>Regression Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>R Square</td>
</tr>
<tr>
<td>Observations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coef</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>6.158105</td>
</tr>
<tr>
<td>X Variable 1</td>
<td>0.000251</td>
</tr>
<tr>
<td>X Variable 2</td>
<td>-0.00026</td>
</tr>
<tr>
<td>X Variable 3</td>
<td>0.047911</td>
</tr>
<tr>
<td>X Variable 4</td>
<td>-1.46362</td>
</tr>
</tbody>
</table>
SUMMARY OF IMPROVED META MODELS DERIVED:

<table>
<thead>
<tr>
<th>RESPONSE</th>
<th>KEY VARIABLES</th>
<th>INDEX OF FIT</th>
<th>F-STATISTIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECONOMIC</td>
<td>AH, AK, AM, AR</td>
<td>97.6%</td>
<td>4723.7</td>
</tr>
<tr>
<td>ECONOMIC</td>
<td>AH, AK</td>
<td>97.4%</td>
<td>8488.9</td>
</tr>
<tr>
<td>VIOLENCE</td>
<td>AH, AK, AM, AR</td>
<td>85.5%</td>
<td>655.8</td>
</tr>
<tr>
<td>VIOLENCE</td>
<td>AH, AK, AR, AX</td>
<td>85.1%</td>
<td>637.5</td>
</tr>
<tr>
<td>VIOLENCE</td>
<td>AH, AK</td>
<td>72.4%</td>
<td>587.5</td>
</tr>
<tr>
<td>VIOLENCE</td>
<td>AH, AK, AM, AP, AR, AS, AT, AU, AV, AW</td>
<td>94.8%</td>
<td>800.4</td>
</tr>
</tbody>
</table>
CORRELATION MATRIX FOR THE FOUR COMMON KEY VARIABLES AND THE TWO RESPONSES:

<table>
<thead>
<tr>
<th></th>
<th>KEYVAR 1</th>
<th>KEYVAR 2</th>
<th>Column 3</th>
<th>Column 4</th>
<th>Violent</th>
<th>Econ</th>
</tr>
</thead>
<tbody>
<tr>
<td>KV 1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KV 2</td>
<td>-0.07464</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col 3</td>
<td>0.00345</td>
<td>0.088232</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col 4</td>
<td>0.032984</td>
<td>0.0929</td>
<td>0.055228</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Violence</td>
<td>0.44621</td>
<td>-0.7560</td>
<td>-0.0018</td>
<td>-0.4056</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Econ</td>
<td>-0.4637</td>
<td>0.90353</td>
<td>0.028</td>
<td>0.08433</td>
<td>-0.8671</td>
<td>1</td>
</tr>
</tbody>
</table>
Principal Components

• Can also be used with Latin Hypercube
  – When variables are strongly correlated
  – Alternative dimension reduction technique

• Main problem: how to interpret it:
  – To identify Key variables through loadings?
  – To use the PCA Main Factors, instead?
  – Alternative approaches?

• Needs evaluation and comparison w/DOE
Example of Varimax Factor Rotation:
Project Variables X1 and X2 on F1
Then, Project Variable X3 on Factor 2.
Other Approaches

• Bayesian
  – Assume a prior on Meta Model terms

• Hierarchical
  – Sub-model output yields upper level input

• Taguchi
  – Derive results resilient to “noise” parameters
  – Parameters representing “uncontrolled” vars
  – Provides many conceptual DOE ideas.
Taguchi Approach

- Analyzes both Location and Variation
  - Of the performance measure of interest
- Best combination of both these together
  - To obtain most efficient Meta Model
- Optimize Location, resilient to Variation
- Minimize Variation, resilient to Location
- Determine regions of joint optimality
- Determine Variation is Not an issue
- Can be equivalently implementing w/DOE
Examples of Taguchi’s SN Ratios

SN Ratios:
- Blue: Closer to Target
- Green: Maximize Yield
- Red: Minimize Yield
Example of Taguchi

- Response: Wet Land Size
- X1=Reservoir Capacity (MAX)
- X2=Generation of electricity
  - X3=Hospital Capacity
  - X4=Social Services
- X5=Fish/Foul Population
Comparison of *Combined* DOE and Taguchi's Approach

<table>
<thead>
<tr>
<th></th>
<th>X1</th>
<th>X3</th>
<th>X2</th>
<th>X4</th>
<th>X5</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Var</th>
<th>LnVar</th>
<th>Average</th>
<th>TagMinim</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>194</td>
<td>197</td>
<td>193</td>
<td>275</td>
<td>1616.25</td>
<td>7.39</td>
<td>214.75</td>
<td>-46.75</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>136</td>
<td>136</td>
<td>132</td>
<td>136</td>
<td>4.00</td>
<td>1.39</td>
<td>135.00</td>
<td>-42.61</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>185</td>
<td>261</td>
<td>264</td>
<td>264</td>
<td>1523.00</td>
<td>7.33</td>
<td>243.50</td>
<td>-47.81</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>47</td>
<td>125</td>
<td>127</td>
<td>42</td>
<td>2218.92</td>
<td>7.70</td>
<td>85.25</td>
<td>-39.51</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>295</td>
<td>216</td>
<td>204</td>
<td>293</td>
<td>2376.67</td>
<td>7.77</td>
<td>252.00</td>
<td>-48.15</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>234</td>
<td>159</td>
<td>231</td>
<td>157</td>
<td>1852.25</td>
<td>7.52</td>
<td>195.25</td>
<td>-45.97</td>
</tr>
</tbody>
</table>

Regression Analysis for the Main Effect influence

Coef  | Std Err  | t Stat | P-value | Lower 95 | Upper 95 |
---    |----------|--------|---------|----------|----------|
Intrcpt| 197.13    | 7.88   | 25.01   | 0.00     | 181.00   | 213.25  |
X Var 1| -27.50    | 7.88   | -3.49   | 0.00     | -43.62   | -11.38  |
X Var 2| 56.88     | 7.88   | 7.21    | 0.00     | 40.75    | 73.00   |

Regression Analysis for the Variance Influence

Coef  | Std Err  | t Stat | P-value | Lower 95 | Upper 95 |
---    |----------|--------|---------|----------|----------|
Intrcpt| 6.77     | 0.78   | 8.70    | 0.00     | 4.77     | 8.77    |
X Var 1| -0.82    | 0.78   | -1.05   | 0.34     | -2.82    | 1.18    |
X Var 2| 0.69     | 0.78   | 0.88    | 0.42     | -1.31    | 2.69    |
Optimal Solution:
Overlying both plots (for location and variation) we seek to Minimize both Yield (Errors) and Variation.

Jointly applying the two above (cols. 3 & 8).

The Optimum is around (1, -1), yielding
Estimated Minimum Output = 113; Min Variation = 5.3

Estimated Yield:
\[ Y = 197.12 - 27.5X_1 + 56.9X_2 \]
\[ Y(1, -1) = 112.72 \]

Estimated Variation:
\[ Y = 6.77 - 0.82X_1 + 0.69X_2 \]
\[ Y(1, -1) = 5.26 \]

Alternative Combined DOE Approach
Some Applications

- Model Size Reduction for:
- Evaluation of Decisions and Strategies
- Evaluation of Robust Strategies
- Trade-offs and Sensitivity analyses
- What-if, time to catastrophic fails, etc.
- Design and Optimization of Systems
- Study of key Factors on a System
- Arbitration and Conflict Resolution
Composite Objective Functions

Ecologic: $X_i$ is number of occurrences of $i$th item:

$$f(x_1, \cdots, x_p) = \sum_i v_i x_i; \text{with} \ : \sum v_i = 1$$

Economic: $Y_i = a_i X_i$ is cost of No. $i$th item occurrences:

$$g(x_1, \cdots, x_p) = \sum_i \lambda_i y_i; \text{with} \ : \sum \lambda_i = 1$$

$$l(w_1, \cdots, w_n) = \sum_i \delta_i w_i; \text{with} \ : \sum w_i = 1$$

Arbitration and Trade-Off: $\alpha$ is the preference or weight:

$$H(g, l) = \alpha g + (1 - \alpha) l; \text{with} \ : 0 < \alpha < 1$$
Example of approach use:

- Reduce Model to Key Variables to:
- Minimize Total Water Operations Cost
- Subject to:
  - Maintaining specified labor levels
  - Reducing pollution to specified levels
  - Maintaining specified social levels
  - Maintaining specified consumption levels
  - Increasing overall health indices
### Trade-Off Examples

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Ecologic</th>
<th>Health</th>
<th>Industry</th>
<th>Education</th>
<th>Recreation</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best Ecologic</td>
<td>X1</td>
<td>Y1</td>
<td>Z1</td>
<td>W1</td>
<td>L1</td>
<td>M1</td>
</tr>
<tr>
<td>Best Health</td>
<td>X2</td>
<td>Y2</td>
<td>Z2</td>
<td>W2</td>
<td>L2</td>
<td>M2</td>
</tr>
<tr>
<td>Best Industry</td>
<td>X3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best Education</td>
<td>X4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best Recreation</td>
<td>X5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best Other</td>
<td>X6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Analyze Maxi-min and Mini-max results
Conclusions

• A very complex problem
  – Size and interactions are serious issues
• Existing methods, not fully compliant
  – But a promise if worked around
• Meta Models extremely useful
  – For strategic and tactical decisions
  – In crisis, and to assess/avoid them
  – In theoretical studies.