# Simulation: Solving Dynamic Models

**ABE 5646**  
**Week 12, Spring 2009**

<table>
<thead>
<tr>
<th>Week</th>
<th>Description</th>
<th>Reading Material</th>
</tr>
</thead>
</table>
| 12   | Uncertainty and Sensitivity Analysis  
• Two forms of crop models  
• Random sampling for stochastic models  
• Terminology, Definitions  
• Objectives for Uncertainty, Sensitivity Analyses  
• Ingredients  
• Notation with Example  
• Describing Uncertainty, Sensitivity Analysis Factors  
• Methods Overview (Sensitivity and Uncertainty)  
• Sensitivity Analysis  
  – Local sensitivity analysis (absolute, relative)  
  – Global Sensitivity Analysis  | Wallach et al. (2006)  
Chapters 1, 3  
Also see chapters 14, 15, 16 for examples |

- Analysis of output variance (ANOVA)  
- Monte Carlo sampling
Weeks 12 Objectives

1) Learn basic methods for uncertainty and sensitivity analysis using dynamic biophysical models
Shifting Focus

• From:
  – System modeling and simulation methods
• To:
  – Working with dynamic models (with crop examples, but methods apply to other biological and agricultural system models)

This shift means that we take a model that has already been developed and work with it

In our class, we will focus on model 1) evaluation (comparison with data from the real system), 2) analyzing uncertainty and sensitivity of a model to various factors, and 3) estimating parameters and their uncertainty.

We will not attempt to cover all methods in the book in the remaining time, but instead cover 2 or 3 of the most important methods and concepts for each topic that we do cover.
Emphasis on Statistical Notions

- Review the Appendix in Wallach et al.
  - Particularly pp. 429, 430, 431, and 433
- This review will help you better follow and understand the notation in the chapters that we will cover
Previous Lecture

• Notation for models in discrete time form
• A simple crop model example (3 state variables)
• Dynamic models viewed as response functions
• Random components – stochastic models
Random Model Components: Dynamic Equations & Response Function

**Dynamic Equations:**

\[ U_i(t+\Delta t) = U_i(t) + g_i[U(t),X(t); \theta] \Delta t + \eta_i \]

Where \( \eta_i \) is a random variable for rate equation I

Also, parameters \( \theta \) may be uncertain as could inputs \( X(t) \)

**Response functions:**

\[ Y = f(X,\theta) + \varepsilon \]

Where \( \varepsilon \) is a random variable
How to deal with these random variables in solutions

• Monte Carlo Sampling
  – Random variables of model equations
  – Random variables of responses
  – Random variables associated with individual parameters (not mentioned in Chapter 1)
    • i.e., $\theta_i$ uncertainty could be represented by a mean value and variance if it is normally distributed. Monte Carlo simulation could also be used to sample possible values from its distribution, as in the model and response equations. But, if parameters are not independent, random sampling would need to be done from a multivariate distribution
Monte Carlo Sampling

• Inverse transformation method
  – Cumulative probability distribution
  – Pseudo random number [0,1] assumed to be cumulative probability
  – Solve for the value of the variable for this particular random number

• This is the “random” sample
• This is repeated n times to produce n random samples from the population with the probability distribution given
Inverse Sampling Method

\[ F_X(x) = P[X \leq x] = \int_{-\infty}^{x} f(x)dx \]

With \( f(x) \) the probability density and 
\( F_X(x) \) the cumulative probability 
of the random variable \( X \)
Example
Uniform Probability Distribution

\[ f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x \leq b \\ 0 & \text{else} \end{cases} \]

and

\[ F(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x - a}{b-a} & \text{if } a < x \leq b \\ 1.0 & \text{if } x > b \end{cases} \]

Set r.n. = F(x) and solve for x
If r.n. = 0.71, then the sample of x is 5.84
Example
Uniform Probability Distribution

\[ f(x) = \begin{cases} 
\frac{1}{b-a} & \text{if } a < x \leq b \\
0 & \text{else} 
\end{cases} \]

\[ F(x) = \begin{cases} 
(x-a)/(b-a) & \text{if } a < x \leq b \\
1.0 & \text{if } x > b 
\end{cases} \]

Set r.n. = \( F(x) \) and solve for \( x \)
If r.n. = 0.71, then the sample of \( x \) is 5.84
Example
Normal Distribution

Mean = 5.0
Std Dev = 1.0
r.n. = 0.71
Random x sample = 5.55
Chapter 3 – Uncertainty Analysis (UA) & Sensitivity Analysis (SA)
Uncertainty and Sensitivity Analysis Terminology

- Uncertainty Analysis – quantify uncertainty in model outputs in the form of distributions due to uncertainties in model components
- Risk Analysis – probability of model output exceeding a threshold value (or dropping below a threshold); associated with uncertainty analysis
- Sensitivity Analysis – quantify how the outputs of a model vary when model components are varied
- Model Error Propagation – similar to sensitivity analysis for dynamic models, particularly how output variability changes vs. time
- Computer Experiments – set of simulation runs designed to explore model output responses when inputs vary

Is “Sensitivity Analysis” the same as “Uncertainty Analysis”?
Objectives for UA & SA

• Explore model behavior
• Identify parameters with large (or small) influence on outputs
• Identify parameters that need to be estimated more accurately
• Identify inputs that need to be estimated more accurately
• Characterize interactions between parameters, inputs
• Determine ways to simplify model
Ingredients in UA & SA

• The model
• Input factors
  – Set of alternative model structures
  – Uncertain or variable parameter ($\theta_j$)
  – Uncertain or variable input variable ($X_i$)
  – Series of several related input variables (such as annual weather)
  – Note that a subset of parameters or inputs may be selected for a particular UA or SA
• Experiment design
• Simulation of experiment
• Analysis of outputs
Notation

Input factors: $Z_1, \ldots, Z_s$, where $s$ is the number of input factors

Input Scenario: combination of levels $\mathbf{z} = (z_1, \ldots, z_s)$
There will be several or many ($k$) input scenarios $\mathbf{z}_k = (z_{k,1}, \ldots, z_{k,s})$

Model response: $f(\mathbf{x}, \theta)$ or expressed vs scenario, $f(\mathbf{z}) = f(z_1, \ldots, z_s)$

Noting that:
Bold letters designate a vector of values (inputs, $\mathbf{x}$, and factors, $\mathbf{z}$)
Factors may include parameters and/or inputs, and may also include categories as well as quantitative components
A winter wheat dry matter model

A simple crop model will be used in this chapter to illustrate the different methods of uncertainty and sensitivity analysis. The model has a single state variable, the above-ground winter wheat dry matter, denoted by $U(t)$ with $t$ the day number since sowing. This state variable is calculated on a daily basis as a function of cumulative degree-days $T(t)$ (above a baseline of 0°C) and of daily photosynthetically active radiation PAR(t). The model equation is:

$$U(t + 1) = U(t) + E_b E_{imax} \left[ 1 - e^{-K \cdot LAI(t)} \right] \text{PAR}(t) + \varepsilon(t),$$

with $E_b$ the radiation use efficiency, $E_{imax}$ the maximal value of the ratio of intercepted to incident radiation, $K$ the coefficient of extinction, LAI($t$) the leaf area index on day $t$, and $\varepsilon(t)$ a random term representing the model error. In this chapter, we consider the deterministic part of the model only, so this model error will be assumed null in the simulations. LAI($t$) is calculated as a function of cumulative degree-days $T(t)$, as follows (Baret, 1986):

$$LAI(t) = L_{max} \left\{ \frac{1}{1 + e^{-A[T(t)-T_1]}} - e^{B[T(t)-T_2]} \right\}.$$

The dry matter at sowing ($t = 1$) is set equal to zero: $U(1) = 0$. In addition, the constraint $T_2 = \frac{1}{B} \log[1 + \exp(A \times T_1)]$ is applied, so that LAI(1) = 0.

We will assume that the dry matter at harvest $U(t_H)$ is the main output variable of interest, and denote

$$\hat{Y} = U(t_H) = \sum_{t=1}^{t_H-1} E_b E_{imax} \left[ 1 - e^{-K \cdot LAI(t)} \right] \text{PAR}(t).$$
Example Model Showing Notation

While presenting sensitivity analysis, it is convenient to consider the model in the form

$$\hat{Y} = f(X; \theta).$$

In this expression, $X = (T(1), \ldots, T(t_H), \text{PAR}(1), \ldots, \text{PAR}(t_H))$ denotes the daily climate input variables, and $\theta = (E_b, E_{imax}, K, L_{max}, A, B, T_1)$ denotes the vector of parameters, with $L_{max}$ the maximal value of LAI, $T_1$ a temperature threshold and $A$ and $B$ two additional.

In the winter wheat dry matter model, the seven parameters have associated uncertainty and so they represent seven input factors for the uncertainty and sensitivity analyses. The other source of uncertainty to be considered in this example is that related to the input variables of the model. Instead of considering each input variable PAR(t) and T(t) at each time t as a separate sensitivity input factor, a set of fourteen annual series of climate measurements in the region of interest will constitute the eighth factor of the sensitivity analysis. Thus, there are eight factors: the seven parameters $E_b$, $E_{imax}$, $K$, $L_{max}$, $A$, $B$, $T_1$ and the climate factor $C$. An input scenario is a vector

$$z = (z_{E_b}, z_{E_{imax}}, z_K, z_{L_{max}}, z_A, z_B, z_{T_1}, z_C)$$

specifying a combination of values of the input parameters. As this example shows, a factor may be quantitative – the seven parameters – or categorical – the climate series.
Setting Factors for UA &SA

- Nominal value of $z_{0,i}$ – standard setting for factor and may be your expectation of the most likely value
- Control Scenario $z_0$ – nominal values for all of the input factors
- Uncertainty Range – around nominal values
  - $[\theta_{min(j)}, \theta_{max(j)}]$ for $\theta_j$
  - $[x_{min(l)}, x_{max(l)}]$ for $x_l$
  - $[C]$ for set of categorical values

What are the $z_i$ factor limits?
Coding Input Factors

Coding or normalizing factors to all vary [-1,1] or [0,1]

\[ z_j^C = \frac{z_i - (z_{\min(i)} + z_{\max(i)})}{(z_{\max(i)} - z_{\min(i)})} / 2 \quad \text{for } [-1,+1] \text{ range} \]

or

\[ z_j^C = \frac{z_i - z_{\min(i)}}{z_{\max(i)} - z_{\min(i)}} \quad \text{for } [0,1] \text{ range} \]
Methods Overview

• Define distribution of factors
  – Uniform, Normal, Log-Normal, Beta, Triangular, …

• Generate N scenarios of input factors
  \[ z_k = (z_{k,1}, …, z_{k,s}), \, k = 1, …, N \]
  – Monte Carlo, Latin Hypercube, Correlations among \( z_{k,i} \)?

• Compute model output for each scenario \( f(z_k), \, k = 1, …, N \)

• Analyze output distributions
  – compute means, variances, etc.
An Uncertainty Analysis Example

A winter wheat dry matter model (continued)

The chosen nominal values and uncertainty ranges are given in Table 1 for the parameters. These values come from past experiments, bibliography and expert knowledge. For the climate factor, a set of 14 annual series observed in the region of Grignon (France) was chosen. Note that for such a factor, there is no obvious nominal value.

Table 1. Uncertainty intervals for the parameters of the winter wheat dry matter models.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unit</th>
<th>Nominal value</th>
<th>Uncertainty range</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_b$</td>
<td>g/MJ</td>
<td>1.85</td>
<td>0.9</td>
</tr>
<tr>
<td>$E_{imax}$</td>
<td>–</td>
<td>0.94</td>
<td>0.9</td>
</tr>
<tr>
<td>$K$</td>
<td>–</td>
<td>0.7</td>
<td>0.6</td>
</tr>
<tr>
<td>$L_{max}$</td>
<td>–</td>
<td>7.5</td>
<td>3</td>
</tr>
<tr>
<td>$T_{1}$</td>
<td>C</td>
<td>900</td>
<td>700</td>
</tr>
<tr>
<td>$A$</td>
<td>–</td>
<td>0.0065</td>
<td>0.003</td>
</tr>
<tr>
<td>$B$</td>
<td>–</td>
<td>0.00205</td>
<td>0.0011</td>
</tr>
</tbody>
</table>

To illustrate the coding of factors, let us consider the parameter $E_b$. The values $z^c_{E_b}$ of $E_b$ vary in the uncertainty range $[0.9, 2.8]$. By setting $z^c_{E_b} = (z_{E_b} - 1.85)/0.95$, we get coded values $z^c_{E_b}$ which vary in $[-1, +1]$. 
Analysis of Wheat Model Results

Figure 2. Density functions of the standard uniform (a) and Beta(5, 5) distributions (b). Histograms of the winter wheat model output (biomass at harvest) from samples of size 5000, generated assuming the uniform (c) or Beta(5, 5) (d) distributions.
Summary of Wheat Model Results

- Each scenario was generated assuming independence
- Comparison of statistics of results

<table>
<thead>
<tr>
<th>Sampling</th>
<th>Minimum</th>
<th>1st Quartile</th>
<th>Median</th>
<th>3rd Quartile</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform</td>
<td>0</td>
<td>1119</td>
<td>1636</td>
<td>2257</td>
<td>3785</td>
</tr>
<tr>
<td>Beta(5,5)</td>
<td>134</td>
<td>1665</td>
<td>1955</td>
<td>2233</td>
<td>3305</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform</td>
<td>1669</td>
<td>785</td>
</tr>
<tr>
<td>Beta(5,5)</td>
<td>1946</td>
<td>420</td>
</tr>
</tbody>
</table>
Sensitivity Analysis

- Local Sensitivity
  - Approximates derivative of output relative to input of chosen factor, holding all other factors constant
  - Vary depending on settings of other factors
  - May miss important interactions, factors

- Global Sensitivity
  - Output analyzed for the range of variability or uncertainty in input factor
  - More realistic view of model response
  - Refer also to:
Local Sensitivity

Measures of sensitivity, local sensitivity, one factor at a time

Derivative
Range of Response
Slope
Variance

Figure 3. Bases for defining sensitivity criteria of model output $\hat{Y}$ with respect to input factor
One At a Time (OAT) Input Factor

Derivative Approximation

• Local Sensitivity:

\[ S_{i}^{\text{local}}(z_{k}) = \left. \frac{\partial f(Z)}{\partial Z_i} \right|_{z_k} \approx \frac{f(z_{k,i}) - f(z_{k,i} - \Delta z_{k,i})}{\Delta z_{k,i}} \]

• Local Standardized Sensitivity:

\[ S_{i}^{\text{local}}(z_{k}) = \left. \frac{\partial f(Z)}{\partial Z_i} \right|_{z_k} \cdot \frac{z_{k,i}}{f(z_{k})} \]
Multiple Factors, OAT

Sampled points when three factors are studied through one-at-a-time sensitivity profiles
OAT Example Results, Wheat Analysis

Analyses can also focus on model outputs, as shown here.

Note variations in wheat yield response.

Why?
Extending OAT for Several Input Factors
2 factors, 3 levels each

Figure 4. Two-factor interactions graphics: the output $\hat{Y}$ is represented as a function of the input factor $Z_1$, for three distinct values of $Z_2$. 
## Analysis of Variance: Effects of Factors on Output

*Table 3.* Output values $\hat{Y}$ for two interacting factors $Z_1$ and $Z_2$ and calculation of variance-based criteria for the first factor.

| $Z_1$ | $Z_2$ | $\hat{Y}$ | $E(\hat{Y} | Z_1)$ | $\text{var}(\hat{Y} | Z_1)$ | $\text{var}[E(\hat{Y} | Z_1)]$ | $E[\text{var}(\hat{Y} | Z_1)]$ |
|-------|-------|----------|------------------|----------------|-------------------------|-------------------------|
| 1     | 1     | 3        |                  |                |                         |                         |
| 1     | 2     | 9        | 9                | 24             |                         |                         |
| 1     | 3     | 15       |                  |                |                         |                         |
| 2     | 1     | 5        |                  |                |                         |                         |
| 2     | 2     | 7        | 7                | 8/3            | 26/3 $\approx$ 8.67    | 142/9 $\approx$ 15.78   |
| 2     | 3     | 9        |                  |                |                         |                         |
| 3     | 1     | 4        |                  |                |                         |                         |
| 3     | 2     | 2        | 2                | 8/3            |                         |                         |
| 3     | 3     | 0        |                  |                |                         |                         |

$$\text{Var}(\hat{Y} | Z_1) = E[\hat{Y} | Z_1] - E(\hat{Y} | Z_1)]^2$$

*for each $Z_1$ level*

See chapter 3 for other terms; the last term is tricky.
Global Sensitivity Analysis

• One At a Time
  – Standardized scenario as basis – limited utility
  – Morris (1991) method to select model parameters/inputs that show strongest effects on outputs, i.e., a screening method

• Analysis of Variance, factorial design

• Monte Carlo – Regression method

• Generating scenarios considering correlated input factors

• FAST (Fourier amplitude sensitivity test)

• etc.
Toward Global Sensitivity Analysis

- What factors should be included? There may be tens or hundreds of possibilities
- Need systematic way to determine factors to include
- Local sensitivity, by itself, is not adequate (why?)
- Morris (1991) method
Morris Method

- Extends OAT to explore more combinations
- Create a series of scenarios based on perturbed values of the factor levels, one at a time
- But do not hold all other variables at their nominal value; let them change also.
- If there are $s=7$ factors and each is allowed to take on any of 5 levels, then there are $7^5$ scenarios ($=16,807$)
- If a random sample of 1000 scenarios are selected, then one can compute $d_i(z_0)$, a local sensitivity index as:

$$d_i(z_0) = \frac{f(z_{0,1} \ldots z_{0,i-1}, z_{0,i} + \Delta, z_{0,i+1} \ldots z_{0,s}) - f(z_{0,1} \ldots z_{0,i-1}, z_{0,i}, z_{0,i+1} \ldots z_{0,s})}{\Delta}$$
Morris Method

\[ d_i(z_0) = \frac{f(z_{0,1} \ldots z_{0,i-1}, z_{0,i} + \Delta, z_{0,i+1} \ldots z_{0,s}) - f(z_{0,1} \ldots z_{0,i-1}, z_{0,i}, z_{0,i+1} \ldots z_{0,s})}{\Delta} \]

- Each \( d_i(z_0) \) is a local sensitivity for \( z_0 \) at a particular point
- One can compute the mean and variance of \( d_i \) for each factor \( (i=1, \ldots s) \) across all points & compare them to determine factors that are important.
- Factors with high mean \( d_i \) indicate an important influence
- Factors with high variance indicate may indicate that the \( i^{th} \) factor interacts with other factors or that the response to this factor in non-linear – or both.
Basic idea of the Morris Method

- $x$ = vector of factors (k of them)
- $k$ = no. of factors (or parameters) in the model
- $p$ = no. of discrete coded factor levels
- no of possibilities for each factor in discrete space = $1/p$,
- $\{0, 1/(p-1), 2/(p-1), \ldots, 1\}$ are possible discrete factor levels
- $y(x)$ = model predictions for the $x$ vector of factors
- $\Delta$ = step size (in $p$) for computing "gradient"; no units.

1. Set up combinations for factors in first $x$ set to be simulated (taking each from 0, 1/5, 2/5 in this case since $\Delta = p/(2(p-1))$)
   - i.e. (0, 2/5, 0, 1/5, 1/5, 2/5, 1/5, 0, 1/5) if there are 8 factors
2. Increment one of the factors by $\Delta = 3/5$ in this example
3. Compute local effect for factor $i$ by perturbing only one of the factor levels, say the first.
   - Use the equation to compute $x$ for the original set AND for the perturbed one.
   - Compute $d_i$, the local effect, which has units of the original $y$
     - i.e., $(3/5, 2/5, 0, 1/5, 1/5, 2/5, 1/5, 0, 1/5)$ for the 8 factors
     - evaluate $y$ for this set, then compute $d_i$
4. Now, perturb one other factor, but not the ith one, using all values for the ith one except on.
   - Then, compute $d_j$ for the jth factor using predictions for $i$ and $j$.
   - For example, perturb j=2 for this run. This will produce a vector of $x$ with only the jth factor different from the last run
     - i.e., $(3/5, 5/5, 0, 1/5, 1/5, 2/5, 1/5, 0, 1/5)$ for the 8 factors
5. Repeat. This will result in $k+1$ vectors and one local evaluation of $d_i$ for each of the $k$ factors

This process can be repeated to get a number of combinations of vectors. Then one can analyze the $d_i$ for each i=1,2,... $k$ factors

- using $r$ starting vectors and repeats. This results in $r*k$ local effects
- for example, let the first vector of the second set be $\{0, 0, 0, 0, 0, 0\}$
- and let the first vector of the third set of runs be $\{1/5, 1/5, 1/5, 1/5, 1/5, 1/5, 1/5\}$

However, it is better to use a random selection process instead of setting combinations ourselves.

- See Alam, McNaught, and Ringrose
**Morris Screening-Monte Carlo Sampling**

1. Generate $s$ random numbers (uniform) to select a sample of the discretized, coded factors or parameters, $z_0$ ($s = \# \text{ parameters}$)
2. Choose one parameter as your nominal parameter
3. Simulate the model with the vector $z_0$ of generated parameter values
4. Increment the nominal parameter code by $1 \, \Delta\text{code}$ (i.e., $1/9$ if there you discretize the parameter range into 9 intervals), keeping all other randomly-generated parameter values from step 1 above
5. Simulate the model with this new vector in which only one parameter was changed
6. Compute $d_i(z_0)$ for this pair of outputs using the Morris equation shown earlier
7. Repeat this process again starting at step 1, keeping the same nominal parameter as in step 2. If there are $p$ coded values per parameter, then
8. Each $d_i(z_0)$ is a local sensitivity for $z_0$ at a particular point
9. One can compute the mean and variance of $d_i$ for each factor ($i=1,\ldots,s$) across all points & compare them to determine factors that are important.
10. Factors with high mean $d_i$ indicate an important influence
11. Factors with high variance indicate may indicate that the $i^{th}$ factor interacts with other factors or that the response to this factor in non-linear – or both.

$$d_i(z_0) = \frac{f(z_{0,1}, \ldots, z_{0,i-1}, z_{0,i} + \Delta, z_{0,i+1}, \ldots, z_{0,s}) - f(z_{0,1}, \ldots, z_{0,i-1}, z_{0,i}, z_{0,i+1}, \ldots, z_{0,s})}{\Delta}$$
Example: Morris Screening

_Alam, McNaught, and Ringrose_

Table 2: Elementary Effects of Various Factors for Six Different Trajectories

<table>
<thead>
<tr>
<th>Factors</th>
<th>$r = 1$</th>
<th>$r = 2$</th>
<th>$r = 3$</th>
<th>$r = 4$</th>
<th>$r = 5$</th>
<th>$r = 6$</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.07</td>
<td>0.03</td>
<td>0.07</td>
<td>0.1</td>
<td>0.15</td>
<td>0.07</td>
<td>0.08</td>
<td>0.002</td>
</tr>
<tr>
<td>2</td>
<td>0.58</td>
<td>0.5</td>
<td>0.62</td>
<td>0.22</td>
<td>0.33</td>
<td>0.45</td>
<td>0.45</td>
<td>0.023</td>
</tr>
<tr>
<td>3</td>
<td>0.07</td>
<td>0.08</td>
<td>0.1</td>
<td>0.02</td>
<td>0.05</td>
<td>0.13</td>
<td>0.08</td>
<td>0.001</td>
</tr>
<tr>
<td>4</td>
<td>0.05</td>
<td>0.08</td>
<td>0</td>
<td>0.05</td>
<td>0.12</td>
<td>0.03</td>
<td>0.06</td>
<td>0.002</td>
</tr>
<tr>
<td>5</td>
<td>0.05</td>
<td>0.2</td>
<td>0.05</td>
<td>0.1</td>
<td>0.08</td>
<td>0.180</td>
<td>0.11</td>
<td>0.004</td>
</tr>
<tr>
<td>6</td>
<td>0.17</td>
<td>0.28</td>
<td>0.05</td>
<td>0.12</td>
<td>0.17</td>
<td>0.080</td>
<td>0.15</td>
<td>0.007</td>
</tr>
<tr>
<td>7</td>
<td>0.03</td>
<td>0.02</td>
<td>0.03</td>
<td>0.08</td>
<td>0.02</td>
<td>0.020</td>
<td>0.03</td>
<td>0.001</td>
</tr>
<tr>
<td>8</td>
<td>0.28</td>
<td>0.23</td>
<td>0.1</td>
<td>0.3</td>
<td>0.47</td>
<td>0.050</td>
<td>0.24</td>
<td>0.023</td>
</tr>
<tr>
<td>9</td>
<td>0.07</td>
<td>0.12</td>
<td>0.08</td>
<td>0.08</td>
<td>0.02</td>
<td>0.020</td>
<td>0.07</td>
<td>0.002</td>
</tr>
<tr>
<td>10</td>
<td>0.45</td>
<td>0.18</td>
<td>0.53</td>
<td>0.42</td>
<td>0.43</td>
<td>0.370</td>
<td>0.40</td>
<td>0.014</td>
</tr>
<tr>
<td>11</td>
<td>0.15</td>
<td>0.12</td>
<td>0.12</td>
<td>0.05</td>
<td>0.33</td>
<td>0.180</td>
<td>0.16</td>
<td>0.009</td>
</tr>
<tr>
<td>12</td>
<td>0.1</td>
<td>0.17</td>
<td>0.17</td>
<td>0.03</td>
<td>0.2</td>
<td>0.050</td>
<td>0.12</td>
<td>0.005</td>
</tr>
<tr>
<td>13</td>
<td>0.03</td>
<td>0.32</td>
<td>0.05</td>
<td>0.37</td>
<td>0.2</td>
<td>0.170</td>
<td>0.19</td>
<td>0.019</td>
</tr>
</tbody>
</table>
Factorial Experiment Design: 
after selecting factors to include

![Factorial experiment design diagram]

*Figure 7.* Sampled points when three factors are studied through a complete factorial design with two modalities per factor (large dots) or three modalities per factor (small and large dots).

*Table 4.* Number of runs for complete $m^s$ factorial designs.

<table>
<thead>
<tr>
<th>$s$</th>
<th>$m = 2$</th>
<th>$m = 3$</th>
<th>$m = 4$</th>
<th>$m = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>32</td>
<td>243</td>
<td>1024</td>
<td>3125</td>
</tr>
<tr>
<td>10</td>
<td>1024</td>
<td>59049</td>
<td>1048576</td>
<td>9765625</td>
</tr>
<tr>
<td>20</td>
<td>1.05e + 06</td>
<td>3.49e + 09</td>
<td>1.10e + 12</td>
<td>9.54e + 13</td>
</tr>
</tbody>
</table>
Analysis of Variance of Responses

The response variability can be decomposed into factorial terms as follows:

$$\sum_{ab} (\hat{Y}_{ab} - \mu)^2 = m \sum_{a} \alpha_a^2 + m \sum_{b} \beta_b^2 + \sum_{a,b} \gamma_{ab}^2,$$

where $SS_T$ measures the total variability in the model responses, $SS_1$ is the sum of squares associated with the main effect of $Z_1$, $SS_2$ is the sum of squares associated with the main effect of $Z_2$, and $SS_{12}$ is the sum of squares associated with the interaction between $Z_1$ and $Z_2$.

With $s$ factors at $m$ levels, the complete ANOVA decomposition is a sum of $(2^s - 1)$ factorial terms:

$$SS_T = \sum_{i} SS_i + \sum_{i<j} SS_{ij} + \cdots + SS_{1...s},$$

including main effects ($SS_i$) and interactions between up to $s$ factors ($SS_{1...s}$). The number
Table 5. Analysis of variance table of the complete factorial design applied to the winter wheat dry matter model; the table was calculated for the model including main effects and two-factor interactions. Sensitivities smaller than 0.01 are not displayed.

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>Sensitivity index</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_b$</td>
<td>777593320</td>
<td>0.33</td>
</tr>
<tr>
<td>$E_{\text{max}}$</td>
<td>6686674</td>
<td></td>
</tr>
<tr>
<td>$K$</td>
<td>3662758</td>
<td></td>
</tr>
<tr>
<td>$L_{\text{max}}$</td>
<td>80732881</td>
<td>0.03</td>
</tr>
<tr>
<td>$A$</td>
<td>520104586</td>
<td>0.22</td>
</tr>
<tr>
<td>$B$</td>
<td>309742948</td>
<td>0.13</td>
</tr>
<tr>
<td>$T_1$</td>
<td>551495</td>
<td></td>
</tr>
<tr>
<td>$\text{YEAR}$</td>
<td>7230246</td>
<td></td>
</tr>
<tr>
<td>$E_b \cdot E_{\text{max}}$</td>
<td>1763250</td>
<td></td>
</tr>
<tr>
<td>$E_b \cdot K$</td>
<td>965855</td>
<td></td>
</tr>
<tr>
<td>$E_b \cdot L_{\text{max}}$</td>
<td>21288948</td>
<td></td>
</tr>
<tr>
<td>$E_b \cdot A$</td>
<td>137149566</td>
<td>0.06</td>
</tr>
<tr>
<td>$E_b \cdot B$</td>
<td>81678016</td>
<td>0.04</td>
</tr>
<tr>
<td>$E_b \cdot T_1$</td>
<td>145427</td>
<td></td>
</tr>
<tr>
<td>$E_b \cdot \text{YEAR}$</td>
<td>1932958</td>
<td></td>
</tr>
<tr>
<td>$E_{\text{max}} \cdot K$</td>
<td>8306</td>
<td></td>
</tr>
<tr>
<td>$E_{\text{max}} \cdot L_{\text{max}}$</td>
<td>183068</td>
<td></td>
</tr>
<tr>
<td>$E_{\text{max}} \cdot A$</td>
<td>1179375</td>
<td></td>
</tr>
<tr>
<td>$E_{\text{max}} \cdot B$</td>
<td>702365</td>
<td></td>
</tr>
<tr>
<td>$E_{\text{max}} \cdot T_1$</td>
<td>1251</td>
<td></td>
</tr>
<tr>
<td>$E_{\text{max}} \cdot \text{YEAR}$</td>
<td>16622</td>
<td></td>
</tr>
<tr>
<td>$K \cdot L_{\text{max}}$</td>
<td>823631</td>
<td></td>
</tr>
<tr>
<td>$K \cdot A$</td>
<td>82704</td>
<td></td>
</tr>
<tr>
<td>$K \cdot B$</td>
<td>635271</td>
<td></td>
</tr>
<tr>
<td>$K \cdot T_1$</td>
<td>395</td>
<td></td>
</tr>
<tr>
<td>$K \cdot \text{YEAR}$</td>
<td>6643</td>
<td></td>
</tr>
<tr>
<td>$L_{\text{max}} \cdot A$</td>
<td>60448</td>
<td></td>
</tr>
<tr>
<td>$L_{\text{max}} \cdot B$</td>
<td>17116469</td>
<td></td>
</tr>
<tr>
<td>$L_{\text{max}} \cdot T_1$</td>
<td>35467</td>
<td></td>
</tr>
<tr>
<td>$L_{\text{max}} \cdot \text{YEAR}$</td>
<td>145584</td>
<td></td>
</tr>
<tr>
<td>$A \cdot B$</td>
<td>193147537</td>
<td>0.08</td>
</tr>
<tr>
<td>$A \cdot T_1$</td>
<td>28101635</td>
<td>0.01</td>
</tr>
<tr>
<td>$A \cdot \text{YEAR}$</td>
<td>2586798</td>
<td></td>
</tr>
<tr>
<td>$B \cdot T_1$</td>
<td>1425195</td>
<td></td>
</tr>
<tr>
<td>$B \cdot \text{YEAR}$</td>
<td>1178019</td>
<td></td>
</tr>
<tr>
<td>$T_1 \cdot \text{YEAR}$</td>
<td>2471694</td>
<td></td>
</tr>
<tr>
<td>Residuals</td>
<td>128829265</td>
<td></td>
</tr>
</tbody>
</table>

Analysis of Variance Table:

Complete Factorial design creates many combinations of simulations, terms for analysis.

You can do this in statistical software, like SAS, S+, R, Minitab, etc.
Wheat Model Variance Components

What does this tell us?

Figure 8. The eight largest factorial sensitivity indices based on the $2^5 \times 14$ factorial design and its analysis of variance with a complete factorial model, for the winter wheat crop model. The upper bars show cumulative indices.
Main Effects Variances

Figure 9. Main-effect (first part of the bars) and total (full bars) sensitivity indices based on the $2^3 \times 14$ factorial design and its analysis of variance, for the winter wheat model.
Discussion

- Monte Carlo methods
- Dealing with correlated factors
- Other methods
- Computer resources needed
- Design of experiment
- Interpretation of results
Homework

- Retrieve files from Class Web site
  - WORD document with problems to work
  - EXCEL Spreadsheet with data
- Due on April 17 or any time before then