<table>
<thead>
<tr>
<th>Week</th>
<th>Description</th>
<th>Reading Material</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 Apr 4-8</td>
<td>Data Assimilation with Dynamic Models&lt;br&gt;• Principles&lt;br&gt;• Example: Ensemble Kalman Filter</td>
<td>Wallach et al. (2006) Ch 5&lt;br&gt;Jones et al. paper</td>
</tr>
<tr>
<td></td>
<td>Optimization with Dynamic Simulation Models&lt;br&gt;• Objective function&lt;br&gt;• Decision variables&lt;br&gt;• Gradient search methods&lt;br&gt;• Monte Carlo methods – Simulated Annealing</td>
<td>Wallach et al. (2006) Ch 6</td>
</tr>
</tbody>
</table>
Week 14 Objectives

1) Become familiar with concepts of optimization using dynamic models and learn about some of the basic methods
   • Concepts
   • Methods

2) Introduction to data assimilation using dynamic models
   • Concepts
   • Methods
Optimization

• Components
  – Model
  – Objective Function
  – Decision Variables
  – Constraints
  – Procedure to estimate best set of decision variables relative to criteria in Objective Function

• Decision Variables may be continuous or discrete
Model in Response Form

Deterministic: \[ Y = f(X, \theta) \]

Where \( Y \) = vector of response variables simulated by the model, \( X \) = vector of all inputs and initial conditions, \( \theta \) = the model parameters, and \( f \) = a representation of the integration of the model up to some time \( T \) (see chapter 1 in Wallach)

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

Where \( \varepsilon \) is a vector of random variables
Decision Variables

Deterministic: \( Y = f(X, \theta, D) \)

Where \( D = \text{vector of decision variables} \), which are the variables that one can manipulate to achieve a particular goal or objective. For example, one may be interested in determining the amount of N fertilizer to apply to a crop (a continuous variable) or the date to plant (a discrete variable).

And \( Y, f, X, \) and \( \theta \) are as defined in previous slide. The stochastic version can also be written following the previous slide:

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

And, the \( X \) can include uncertainty, such as the case with different weather years.
Decision Variables

• May be discrete, such as number of cars to own in a pizza delivery service or the number of times to spray a particular dose of insecticide, number of times to irrigate

• Continuous example would be the amount of water to apply to a crop when irrigating, the temperature set point in a control system

• May also be categorical, such as a set of specific scenarios

• In each case, your objective would be to find the best combination of variables to achieve a particular purpose
Objective Function

- An objective function expresses in mathematical terms one’s objective to be met through selection of decision variables.
- An example is profit, and one’s goal would be to maximize profit and thus to determine the management practices that maximize net profit. Another example is to find the combination of temperature and CO$_2$ settings in a greenhouse to maximize production of tomatoes.
Objective Function

Let \( U(Y_1, Y_2, Y_3, \ldots Y_N) \) be a function of the response variables of the model. The objective function would be:

\[
U^* = \max_D \{U(Y_1, Y_2, \ldots Y_N)\} \quad \text{over all } D
\]

Where \( U^* \) represents the OPTIMAL value of \( U \), with \( U \) depending on the vector of \( Y \) values, and \( Y \) depending on the decision variables, \( D \). This can be read as follows: Find \( D \) such that the value of \( U \) is maximized. This implies that one must “search” for values of \( D \) that satisfy this objective.
Objective Function

An example would be to find the combination of irrigation amount and N fertilizer to apply to maximize profit. In this case, the objective function may look like the following:

\[
U^* = \max_{\{D_1, D_N\}} \left[ P_1 \cdot Y_1 + P_2 \cdot Y_2 - C_I \cdot D_I - C_N \cdot D_N - C_{\text{base}} \right]
\]

Where \( U^* \) represents the maximum profit, \( Y_1 \) is grain yield, \( Y_2 \) is straw yield, \( P_1 \) and \( P_2 \) are prices for grain and straw, respectively, \( C_I \) and \( C_N \) are costs for applying \( D_I \) and \( D_N \) amounts of water and N fertilizer, respectively, and \( C_{\text{base}} \) are all other costs of production.

Note that \( U \) is usually a scalar value whereas \( Y \) may be a vector. However, one can have two objectives, for example, but it may not be possible to satisfy both at the same time.
Constraints

- Set domain of feasible decision variables
- Ranges of variables
- Combination equations
- Example
  - $X > 0$
  - $X + Y > 120$
  - $0 < Y < 45$
Objective Function, Stochastic Models

The previous equation would be modified only slightly, but its computation would typically mean that one would have to use Monte Carlo methods, for example, to generate a number of possibilities. Also, prices may be probabilistic:

\[ U^* = \max_{\{D_I D_N\}} E[P_1 \cdot Y_1 + P_2 \cdot Y_2 - C_I \cdot D_I - C_N \cdot D_N - C_{\text{base}}] \]

Where E represents expected value. This means that Yi are stochastic and that each as a probability of occurring. The expected value states that you want to maximize the mean value of the objective function.
Methods: Visual?

- Vary parameters and select the combination that gives the best visual fit to the observations
- Not a good way to do this, but this is frequently done
- Need to use some criteria and allow a procedure to select the parameter set
Methods for Selecting Optimal Decision Variables

• Brute force, search over all combinations of decision variables and select decisions that result in optimal objective function value
• Systematic search methods, such as gradient search methods, simplex method
• Monte Carlo methods, randomly choosing different combinations of decision variables to search the decision space D
Brute Force: Factorial Design

- Set a range over which to search for each decision variable
- Select m equally-spaced values within the range for each decision variable
- Simulate model output for all combinations of decision variable values
- Compute U for each
- Select decision variable set that maximizes U
- This is the same as a complete factorial design in sensitivity analysis
- Too costly to be practical in most cases (i.e., when number of decision variables is larger than about 3 to 5.)
Simplex Method

With n decision variables, generate n+1 combinations and simulate Y for each

Compute U, the objective function used to determine “best set” of decision variables

Find the minimum U and replace the set that gave this U with another

Still have n+1 sets, repeat the above, reducing the changes as iterations increase

Procedure illustrated here

Advantage: no derivatives are needed compared with gradient search methods

See Numerical Recipes book..
Originally developed by Nelder and Mead (1961)
Gradient Search Methods

$U[Y(X, \theta, D)]$ is the objective function, for example, and $D$ is a vector of decision variables.

Have starting vector, $D^0$ and compute $U^0$.

Compute $\delta U(D)/\delta D$, the gradient of $U$ or $\nabla U$, for each $D_i$ at the $D^0$ point.

Choose a new point, $D^1 = D^0 - \alpha (\nabla U)$

compute $U^1$ at the point $D^1$

If $U^1$ is less than $U^0$, continue along gradient computed earlier.

But, if $U^1$ is greater than $U^0$, compute gradient again at $D^j$.

Continue until $U^{j+1}$ is within a tolerance of $U^j$.

The decision variable set $D^j$ is the best combination, and $U^j$ is the maximum value of the objective function that is possible for the feasible ranges of $D_i$. 
Gradient Search

This is shown schematically here (2 decision variables)
The contour lines are lines of equal U, the maximum value is in the center
Vertical axis shows values of one variable and the horizontal shows the other decision variable
The arrows are in the direction of the gradient, which is computed 3 times in this example
Gradient Search Problems

Local minimums can result in a set of decision variables that are not optimal.

Note what happens when initial set of decision variables was A vs. B.

This is particularly true when the number of decision variables to select are more than 3 or 4, but it is true in all cases.
Simulated Annealing

• Bouncing ball idea, random jumps to different parts of the search space to avoid getting trapped (Monte Carlo method)
• Size of jumps decreases with number of iterations so that the search converges to a minimum (or maximum) point
• Solution keeps record of m points (say m = 100) that are the lowest (highest) values obtained so far as the solution proceeds
• At the end, one has m solutions for D, including the “optimal” one; one can think of these as a cloud of points, each with the same U to a desired level of accuracy
• The initial jump bounds (temperature) and the rate of decrease of this temperature (annealing) is the origin of the name, like in annealing metals
• If D is composed of 7 variables (i.e., D₁ …, D₇), then one can estimate the expected value from the m possibilities that are “optimal” or within a small tolerance of the optimal point.
• Also, correlations among decision variables can be estimated from the m best solutions
Simulated Annealing Algorithm

The structure of the simulated annealing algorithm
Simulated Annealing Example

Two parameter Rosenbrock function with 2 parameters \((x_1 \text{ and } x_2):\)

\[
f(x) = [(1 - x_1)^2 + 100(x_2 - x_1^2)]^2
\]

Find \(x_1\) and \(x_2\) that minimize \(f\), with \(x_1\) and \(x_2\) varying between -1.5 and 1.5 each.
Simulated Annealing Example (Continued)

Progression of reduction in objective criteria $U$ vs. number of trials

![Graph showing progression of reduction in objective criteria $U$ vs. number of trials](image)

Objective function reduction
Examples?
Data Assimilation

• Concepts
  – Measurements provide estimates of system states, but measurement error may be large (uncertainty)
  – Measurements can only be taken infrequently (over time and space) relative to the need for information
  – Measurements may be costly
  – Direct measurements of variable of interest may not be possible or practical
Example – Replace State Variable
What about modifying parameter & updating state variable?
Another View – Larger Scale
Data Assimilation

• Models may provide estimates of system states, but uncertainties are inherent
• Need to improve the accuracy of model-predicted estimates over time
  – Reinitialize states when they are observed
  – Refine parameters for better predictions
• Why not combine measurements with model-based predictions for estimation?
Data Assimilation

• The name is based on the incorporation of measured data over time (i.e., assimilated)
• Taking into account errors in measurements and models, one can obtain an optimal estimate
• Bayesian estimation methods
• Kalman Filter is one good example of a widely-used data assimilation method
• Method can also assimilate data over space instead of time, or over time and space (if model and measurement vary over space and time)
• Model must be stochastic (i.e., include uncertainty)
Kalman Filter

- Originally developed by R. Kalman, who retired from UF in the 1990s.
- Developed for linear dynamic models.
- Modifications by others later showed that non-linear dynamic models could be approximately linearized and used (i.e., Extended Kalman Filter – see Chapter 18).
- Monte Carlo method may be the most practical method for non-linear dynamic models (the Ensemble Kalman Filter, Chapters 5, 18, & handout by Jones et al.).
Ensemble Kalman Filter (EnKF)

Figure 1. Schematic of the Ensemble Kalman filter procedure. An ensemble of variables, created at time 0 (a), is simulated over time. The model is used to predict the state variables (for each ensemble replicate) for time $t+1$ (b) using values at time $t$. Then measurements are used to update the estimates of variables for time $(t+1)$. These updated values (c) are used as initial values for the next prediction/update step. The heavy line in the figure shows the true value of soil C.
Example Ensemble Kalman Filter

- Soil carbon dynamics model
- Discrete time formulation
- Stochastic
- EnKF has 2 states (X and R)

\[
X_t = X_{t-1} - R \cdot X_{t-1} + b \cdot U_{t-1} + \varepsilon_t
\]
\[
R = R_0 + \eta
\]

where
- \(X_t\) = soil organic carbon in year \(t\) (kg[C]/ha)
- \(R\) = rate of decomposition of existing soil C (1/yr)
- \(R_0\) = initial estimate of soil C decomposition rate (1/yr)
- \(b\) = fraction of fresh organic C that is added to the soil in year \(t\) that remains after one year
- \(U_t\) = amount of C in crop residue that is added to the soil in year \(t\)
- \(\varepsilon_t\) = model error for soil C (kg[C]/ha)
- \(\eta\) = error in initial estimate of decomposition rate \(R\) (1/yr).

\[
\varepsilon_t \sim N(0, \sigma_{\varepsilon}^2)
\]
\[
\eta \sim N(0, \sigma_{\eta}^2)
\]

where
- \(\sigma_{\varepsilon}^2\) = variance of model error for soil C
- \(\sigma_{\eta}^2\) = variance of error for estimate of soil C decomposition rate \(R\).
Representing Measurements

\[ Z_t \] = measurement made at time \( t \), one variable in this example (could be vector of measurements)

\[ \sigma_{Z,t}^2 \] = variance of measurement made at time \( t \) (this could be a covariance matrix if more than one measurement is made)

Variance can change with time, usually measurement at time \( t \) is independent from measurement at \( t-1 \)

Measurements may be independent from each other, if more than one variable is measured at time \( t \)
The equations for the update step are given by:

\[
X_{t+1|z,t+1}^j = X_{t+1|z,t}^j + K_{X,t+1} \left( Z_{t+1}^j + \varepsilon_{z,t+1}^j - X_{t+1|z,t}^j \right) \\
R_{t+1|z,t+1}^j = R_{t+1|z,t}^j + K_{R,t+1} \left( Z_{t+1}^j + \varepsilon_{z,t+1}^j - X_{t+1|z,t}^j \right)
\]

where \( K_{X,t+1} \) and \( K_{R,t+1} \) are Kalman gain values for updating \( X \) and \( R \), respectively. In the equations above, the superscript \( j \) indicates a random sample of measurement for Monte Carlo simulation of measurements; thus, simulated measurement in each realization of ensemble has uncertainty.

And \( \varepsilon_{z,t+1}^j \) = random sample of measurement for Monte Carlo simulation of measurements; thus, simulated measurement in each realization of ensemble has uncertainty.
Updating States using EnKF

The equations for the update step are given by:

\[
X_{i+1|z,t+1}^j = X_{t+1|z,t}^j + K_{X,t+1}^j \left( Z_{t+1}^j + \varepsilon_{Z,t+1}^j - X_{t+1|z,t}^j \right)
\]

\[
R_{i+1|z,t+1}^j = R_{t+1|z,t}^j + K_{R,t+1}^j \left( Z_{t+1}^j + \varepsilon_{Z,t+1}^j - X_{t+1|z,t}^j \right)
\]

(4)

where \( K_{X,t+1} \) and \( K_{R,t+1} \) are Kalman gain values for updating \( X \) and \( R \), respectively. In the equations above, the superscript \( j \) indicates an individual realization. In a Monte Carlo approach, this process is repeated for each realization to account for uncertainty.

Example when \( K_{X,t+1} = 0 \)

\[
X_{t+1|z,t+1}^j = X_{t+1|z,t}^j \quad \text{(model-predicted value)}
\]

and when \( K_{X,t+1} = 1 \)

\[
X_{t+1|z,t+1}^j = Z_{t+1}^j + \varepsilon_{Z,t+1}^j \quad \text{(Observed value of i^{th} realization)}
\]

Similar calculation for \( R \)
Derivation of $K$ (Kalman Gain)

For our model, which has two equations and two variables, the Kalman gain matrix can be written as:

$$
\begin{bmatrix}
K_{X,t} \\
K_{R,t}
\end{bmatrix} = 
\begin{bmatrix}
\sigma_{X,t}^2 & \sigma_{XR,t} \\
\sigma_{XR,t} & \sigma_{R,t}^2
\end{bmatrix} \cdot
\begin{bmatrix}
1 \\
0
\end{bmatrix} \cdot
\begin{bmatrix}
\sigma_{X,t}^2 & \sigma_{XR,t} \\
\sigma_{XR,t} & \sigma_{R,t}^2
\end{bmatrix}^{-1} \cdot
\begin{bmatrix}
1 \\
0
\end{bmatrix} + \sigma_Z^2
$$

(5)

where the terms of the covariance matrix are the variance of soil C predictions at time $t$ ($\sigma_{X,t}^2$), the variance of estimates of soil C decomposition rate at time $t$ ($\sigma_{R,t}^2$), and the

$$
\left[\sigma_{X,t}^2 + \sigma_Z^2\right]^{-1}
$$

Thus, after matrix multiplication, equation 5 can be simplified and written as two terms, as follows:

$$
K_{X,t} = \frac{\sigma_{X,t}^2}{\sigma_{X,t}^2 + \sigma_Z^2}
$$

$$
K_{R,t} = \frac{\sigma_{XR,t}}{\sigma_{X,t}^2 + \sigma_Z^2}
$$

(6)

Note that although $R$ is not measured, the measurement of soil C provides information for refining the estimate of $R$ via the covariance term. Note also that $K$ varies with time; it is recalculated each time a measurement is made using
Calculation of Variances of X, R

• Based on all members of the ensemble of estimates of X and R on day t

\[ \sigma_{Z,t}^2 \] is computed from updated values of Z each day t (variance of all \( Z_t \) estimates)

\[ \sigma_{R,t}^2 \] is computed from updated values of R each day t (variance of all \( R_t \) estimates)

\[ \sigma_{ZR,t} \] is computed from updated values of Z and R each day t (covariance of all \( R_t \) and \( Z_t \) estimates)
## Numerical Example

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>Units</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_0$</td>
<td>True value of soil C at time 0</td>
<td>kg[C]/ha</td>
<td>16,000</td>
</tr>
<tr>
<td>$R$</td>
<td>True value of mineralization parameter</td>
<td>1/yr</td>
<td>0.010</td>
</tr>
<tr>
<td>$\sigma^2_2$</td>
<td>Variance of measurement, constant over time</td>
<td>(kg[C]/ha)$^2$</td>
<td>500,000</td>
</tr>
<tr>
<td>$\sigma^2_\varepsilon$</td>
<td>Variance in model estimates of soil C, each year time step</td>
<td>(kg[C]/ha)$^2$</td>
<td>20,000</td>
</tr>
<tr>
<td>$R_0$</td>
<td>Initial value of soil C decomposition parameter</td>
<td>1/yr</td>
<td>0.015</td>
</tr>
<tr>
<td>$\sigma^2_\eta$</td>
<td>Variance of decomposition rate parameter</td>
<td>(1/yr)$^2$</td>
<td>0.0001</td>
</tr>
<tr>
<td>$U_t$</td>
<td>Input of C to the soil each year (assumed constant)</td>
<td>kg[C]/ha</td>
<td>2,000</td>
</tr>
<tr>
<td>$b$</td>
<td>Proportion of annual soil C that remains after one year</td>
<td>--</td>
<td>0.20</td>
</tr>
<tr>
<td>nrcps</td>
<td>Number of ensemble replicates used</td>
<td>--</td>
<td>1000</td>
</tr>
<tr>
<td>Zfreq</td>
<td>Measurement frequency</td>
<td>1/yr</td>
<td>1</td>
</tr>
</tbody>
</table>
Measurements and Updated EnKF Estimates for 25 Years

Figure 2. Effect of different levels of fresh organic C input on soil C estimates from measurements and from the EnKF. Starting with the lowest curve, C inputs were 0, 1000, 2000, and 4000 kg[C]/ha each year, respectively. Solid lines are EnKF estimates, points are measurements, and the dashed line is the time course of true soil C values for the base case.
Comparing Estimated Annual Changes in Soil Carbon ($Z_{t+1} - Z_t$)

Figure 3. Annual estimates of soil C changes, comparing those made from measurements and from the EnKF with true values.
Variance of Soil C Estimates for Different Frequencies of Measurements

Figure 6. Variance of EnKF estimates of soil C vs. time for different frequencies of measurement (ranging from one per year to one every five years).
Discussion of the EnKF

• General use for any model
• May need many random members in the generated ensemble
• The ensemble of realizations evolves using the model and updates on each realization (member of ensemble)
• Parameters can be assumed to be fixed and model state variables can be estimated
• Assumptions about independence of errors is very important and must be checked
• Can be used to estimate initial conditions, similar to the R parameter
• Can use it (in a sub-optimal approximation) for complex models when only a few of the model state variables are considered uncertain.