In an ensemble model, forecasters run many different versions of a weather model with slightly different initial conditions. This helps account for uncertainty and shows forecasters a spread of possible outcomes.

Members in a typical ensemble:

- A universe where...
  - Rain is 0.5% more likely in some areas
  - Wind speeds are slightly lower
  - Pressure levels are randomly tweaked
  - Dogs run slightly faster
  - There's one extra cloud in the Bahamas
  - Germany won WWII
  - Snakes are wide instead of long
  - Will Smith took the lead in The Matrix instead of Wild Wild West
  - Swimming pools are carbonated
  - Sliced bread, after being banned in January 1943, was never re-legalized
Concepts

* Sensitivity Analysis
  * How does a change in X translate into a change in Y?

* Uncertainty Propagation
  * How do we forecast Y with uncertainty?
  * How does the uncertainty in X affect the uncertainty in Y?

* Uncertainty Analysis
  * which sources of uncertainty are most important?

* Optimal Design
  * How do we best reduce the uncertainty in our forecast?
Sensitivity Methods

- Local
  - Analytical: $\frac{df}{d\Theta}$
  - One-at-a-time perturbations

Saltelli et al. 2008. Global Sensitivity Analysis
Sensitivity Analysis
Global Sensitivity

Curse of Dimensionality
Sensitivity Methods

- Local
  - Analytical: $df/d\Theta$
  - One-at-a-time perturbations

- Global
  - Monte Carlo
  - Sobol
  - Emulators
  - Elementary Effects
  - Group Sampling

Saltelli et al. 2008. Global Sensitivity Analysis
Monte Carlo Sensitivity

$R^2 = 0.17$
$m = -1.2E-8$

$R^2 < 0.01$
$m \sim 0$

Free if you do MC uncertainty propagation or MCMC
UNCERTAINTY PROPAGATION
### Uncertainty Propagation

<table>
<thead>
<tr>
<th>Approach</th>
<th>Distribution</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Variable Transform</td>
<td>Analytical Moments</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Taylor Series</td>
</tr>
<tr>
<td>Analytic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Numeric</td>
<td>Monte Carlo</td>
<td>Ensemble</td>
</tr>
</tbody>
</table>

I don't know how to propagate error correctly, so I just put error bars on all my error bars.
\[ P_y[y] = P_{\theta}[f^{-1}(y)] \cdot \left| \frac{d f^{-1}(y)}{dy} \right| \]
\[ \text{Var}(aX) = a^2 \text{Var}(X) \]

\[ \text{Var}(X + b) = \text{Var}(X) \]

\[ \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y) \]

\[ \text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Cov}(X, Y) \]

\[ \text{Var} \left( \sum X \right) = \sum \text{Var}(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j) \]

\[ \text{Var}(X) = \text{Var}(E[X|Y]) + E[\text{Var}(X|Y)] \]
\[ y = m\Theta + b \]
\[ m = \frac{dy}{d\Theta} \]
\[ \text{Var}[y] = \text{Var}[m\Theta + b] \]
\[ = m^2 \text{Var}[\Theta] \]
\[ = (\frac{dy}{d\Theta})^2 \text{Var}[\Theta] \]
TAYLOR SERIES

\[ y = f(x|\Theta) \]

\[ \text{Var}(y) \sim (dy/d\Theta)^2 \text{Var}[\Theta] \]
LINEAR APPROX

\[
\text{Var}[f(x|\theta)] \approx \text{Var}\left[f(x|\bar{\theta}) + \frac{df}{d\theta}(x|\bar{\theta}) \frac{(\theta - \bar{\theta})}{1!} + \ldots \right]
\]

\[
\text{var}[f(x)] \approx \left(\frac{\partial f}{\partial \theta_i}\right)^2 \text{var}[\theta]
\]
LINEAR APPROX

\[
\text{Var} \left[ f(x|\theta) \right] \approx \text{Var} \left[ f(x|\bar{\theta}) + \frac{df}{d\theta}(x|\bar{\theta}) \frac{1}{1!} (\theta - \bar{\theta}) + \ldots \right]
\]

\[
\text{var} \left[ f(x) \right] \approx \sum \left( \frac{\partial f}{\partial \theta_i} \right)^2 \text{var} \left[ \theta_i \right] + \\
\sum_{i \neq j} \left( \frac{\partial f}{\partial \theta_i} \right) \left( \frac{\partial f}{\partial \theta_j} \right) \text{cov} \left[ \theta_i, \theta_j \right]
\]
\[ Y_{t+1} = f(Y_t, X_t | \theta) + \varepsilon \]

\[ \text{Var}[Y_{t+1}] \approx \left( \frac{df}{dY} \right)^2_{\text{stability}} \text{Var}[Y_t] + \left( \frac{df}{dX} \right)^2_{\text{IC uncert}} \text{Var}[X] + \left( \frac{df}{d\theta} \right)^2_{\text{param uncert}} \text{Var}[\theta] + \text{Var}[\varepsilon]_{\text{process error}} \]
COV & SCALING

• Scaling very dependent on spatial and temporal auto- & cross-correlation

\[
\sum \sum \frac{\partial f}{\partial X_i} \frac{\partial f}{\partial X_j} \text{COV}[X_i, X_j]
\]
## UNCERTAINTY PROPAGATION

<table>
<thead>
<tr>
<th>Approach</th>
<th>Distribution</th>
<th>Output Moments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytic</td>
<td>Variable Transform</td>
<td>Analytical Moments</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Taylor Series</td>
</tr>
<tr>
<td>Numeric</td>
<td>Monte Carlo</td>
<td>Ensemble</td>
</tr>
</tbody>
</table>
Numerical Approximation

- Monte Carlo Simulation --> Distribution
- Ensemble Analysis --> Moments
JENSEN’S INEQUALITY

\[ f(\bar{x}) \neq \bar{f(x)} \]
MONTE CARLO UNCERTAINTY

- for (i in 1:n)
  - draw random values from distributions
  - run model
  - save results
- summarize distributions
**ENSEMBLE UNCERTAINTY**

- for (i in 1:n)
  - draw random values from distributions
  - run model
  - save results

**Requires smaller N to estimate moments than to approximate full PDF**

- Fit PDF to results
- Use PDF for intervals, etc.

**Already have this from MCMC!**
Monte Carlo

\[ \mathbf{y} = f(\mathbf{x}) \]

Taylor Series

\[ \bar{\mathbf{y}} = f(\bar{\mathbf{x}}) \]

\[ P_y = A^T P_x A \]

\[ f(\bar{\mathbf{x}}) \]

Unscented Transform

\[ \mathbf{Y} = f(\mathbf{X}) \]

weighted sample mean and covariance

transformed sigma points

UT mean

UT covariance
<table>
<thead>
<tr>
<th>Approach</th>
<th>Distribution</th>
<th>Output Moments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytic</td>
<td>Variable Transform</td>
<td>Analytical Moments</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Taylor Series</td>
</tr>
<tr>
<td>Numeric</td>
<td>Monte Carlo</td>
<td>Ensemble</td>
</tr>
</tbody>
</table>
Uncertainty Analysis
VARIANCE DECOMPOSITION

SWITCHGRASS YIELD, CENTRAL ILLINOIS
How do the drivers of forecast uncertainty vary across ecological system?

Dietze (2017)

Thomas unpublished

**UNCERTAINTY:**
- Green: driver: meteorology
- Black: initial conditions (IC)
- Red: parameter
- Green: driver: meteorology downscaling
- Blue: process
Tools for model-data feedbacks

- **Power analysis**
  - Sample size needed to detect an effect size
  - Minimum effect size detectable given a size

- **Observational design**
  - What do I need to measure?
  - Where should I collect new data?
  - How do I gain new info most efficiently?
$SE \propto \frac{1}{\sqrt{n}}$
Power = f(effect size, SE)

H0

Ha

False Positives

False Negatives

Power = 1-β

Effect Size

Effect Size

False Negatives

False Positives

Power = f(effect size, SE)
Pseudo-data simulation

for(k in 1:M)
Draw random data of size N
Fit model
Save Parameters

- Nonparameteric bootstrap: resample data
- Parameteric bootstrap: assume param, sim data
- Embed in overall loop over N or different effect sizes
- Summarize distribution
Observing System Simulation Experiments

- Simulate “true” system
- Simulate pseudo-observations
- Assimilate pseudo-observations
- Assess impact on estimates

- Augment an existing network
  - Additional locations
  - New Sensors
- Common in Weather, Remote Sensing, Oceanography
Zeng et al. 2020 “Use of Observing System Simulation Experiments in the United States” BAMS https://doi.org/10.1175/BAMS-D-19-0155.1