Lecture 13
Validation and Verification

Making Sure that the Theory is Practical!

Agenda
- General Overview of Model Validation
- Some Comment and Thoughts on the Practical Use of Model Output (vis-à-vis, Validation and Applications)

Misquote of the Day
“A Modeler is someone who goes into the restaurant and eats the menu.”

Why Validate?
A Model that is not reviewed or tested against reality cannot be trusted to simulate it.

The Challenge of Validating a Model
- Any Model being used for decision support, pedagogical support needs to go through an evaluation phrase.
- We will call this Validation and Verification.
- These two V’s can be fightin’ words!

From the beginning...
- A model is an idealized representation of a system or entity that encapsulates select key attributes of that system.
  - Not all elements are included, nor can, nor should be.
  - Those elements that are included may not be represented explicitly or even in a manner that is physically correct!

From the beginning...
- A model is an idealized representation of a system or entity that encapsulates select key attributes of that system.
  - Not all elements are included, nor can, nor should be.
  - Those elements that are included may not be represented explicitly or even in a manner that is physically correct!
Enter the Problem

- OK then...
  - We have an approximation of a complex system in which we have, by necessity
  - Dropped certain physical components
  - Isolate Others
  - Discredited from a Continuous Framework
  - Truncated (in a fourier/wave sense)
  - Simplified
  - Scaled
  - ...And otherwise removed it from a precise match of the real universe

So all we want to know is...

What bloody good is it, Then?

What we may want to know tomorrow?

- Can I use it for something else?

So, how does one evaluate a model then?

- Given the broad spectrum of applications and needs, there is no universal procedure for model validation.
- There ARE, however, a set of confidence building tests that many users perform.

First, what is your model supposed to do?

- Capture general trends in behavior?
  - (e.g., a pedagogical example)
- Predict specific events within a real-time/real-world spatio-temporal threshold?
  - (e.g., operational models)
- Diagnose current events (no time dependence)
  - (e.g., diagnostic models)

The General Goal

- A model should "inspire" confidence for its particular application in its first-level users.
- You are the first-level users
  - The developers and primary executors of the models.
  - Those responsible for interpreting the direct model output into "products" and "decisions."
Jargon Alert

Some of the terms you see here do not meet with the approval of all communities
- I don’t like all of the specific terms for the groups we’ll now discuss but I do like the groups themselves.

Jargon Alert

Excellent Example!
- Verify
- Validate
- See the Konicko paper

Jargon Alert

Best advice:
- Always define your assessment criteria
- Always be willing to account for differences between cliques, communities and disciplines.

General Model Evaluations

- "Verification." (jargon alert)
- "External Benching"
- "Face Validity." (jargon alert)
- "Proper Behavior/Structure"
- "Historical Behavior."
- "Past Predictability"
- "Extreme Behavior."
- Robustness
- "Detailed Model Check."
- Model "Intercomparisons"
- Car and Driver Road Tests

“Verification.”

- Independent Evaluation
  - External Review
  - Did the model behave as it should have?
  - Did it satisfy the requirements of independent reviewers
  - Cross Platform Compatibility?
  - Chip Portability?
  - Examples
  - Comparing our models to a bench
  - Comparing against a "Demo."
“Face Validity.”

- Does it look right?
  - Output: Does it go up when it’s supposed to go up?
  - Structure:
    - Are the model components physically reasonable?
    - Are the explicit processes accurately represented for that scale?
    - Are the implicit processes (parameterizations and kludges*) tolerably represented for that scale?
    - There are those who consider this to "invalidate" face validity!

“Verification” vs “Validation”

- Validation vs Verification... (We use a different context)
  - A model can “verify” but NOT match a physical situation
  - Beware that a model is right for the wrong reasons
  - Example: An empirical model vs a physically-based model.

Dueling Jargon

- Meteorology and Forecasting
  - “Verification” often refers to overall modeling and forecast performance
- Groundwater Modeling
  - Verification = Veritas = Truth
  - Validation = “it works enough for me.”
  - (Your prof actually likes the later distinction but lives with the former)

The Money Papers!

- Verify: from Veritas (Truth)
  - Many consider “verification” to be a universal match of model output with and structure with factual data.

“Historical Behavior.”

- How does the model match output from previous cases?
- Often can be said to apply to models that are constructed based on historical data*
- Important if you have a lot of endogenous values (because it’s often driven by them)
- Comparing a model using the same datasets that helped produce it’s working behavior? AVOID!

“Extreme Behavior”

- Robustness
  - Very important for models based on empirical foundations.
    - A model derived from a limited range of case data may be vulnerable to extremes
  - This can be expanded to other concepts
    - Can include numerical and physical issues
    - Extreme temperatures and climates
    - Courant/CFL conditions
"Detailed Model Check"

- Comparisons of Models against each Other
  - Example: Project for Intercomparison for Land-Surface Process Schemes (PILPS)

Need Metrics?

- Examples of Validation Metrics
  - BIAS: \( (x_{\text{obs}} - x_m) \) let = \( x' \)
  - RMSE: Root Mean Square Error
    \[
    RMSE(t) = \sqrt{\frac{\sum_{i=1}^{N} [x_{1,i}(t) - x_{2,i}(t)]^2}{N}}
    \]

Need Metrics?

- Examples of Validation Metrics
  - J: Cost Function \( \rightarrow \ln\{P(x)\} \)
    \[
    J(\vec{x}) = \frac{1}{2}(\vec{x} - \vec{x}^b)^T B^{-1}(\vec{x} - \vec{x}^b) + \frac{1}{2}(\vec{y} - \vec{y}^o)^T (O + F)^{-1}(\vec{y} - \vec{y}^o)
    \]

Where does this put you?

- How would you demonstrate the effectiveness of a model in your problem domain?