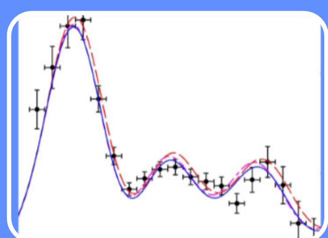
The Modeling Process

Model building

- Qualitative & archival data, expert opinion, prior theory
- Boundary & structure, dimensions, extreme conditions



Calibration & Parameter Estimation

- Quantitative data from diverse sources
- Various methods depending on the data and model structure

Policy Analysis

- Structure and behavior analysis
- Sensitivity & robustness analysis
- Various optimization methods

Agenda

- Model building
  - Data and Modeling
- Parameter Estimation
  - Calibration
  - Confidence Intervals
  - Role of noise
- Model Analysis
  - Analysis process (time step, equilibrium, causal tracing, loop knock-out, loop dominance, test understanding)
  - Controlling dynamic systems
  - Synergies with other areas of analytics

Agenda

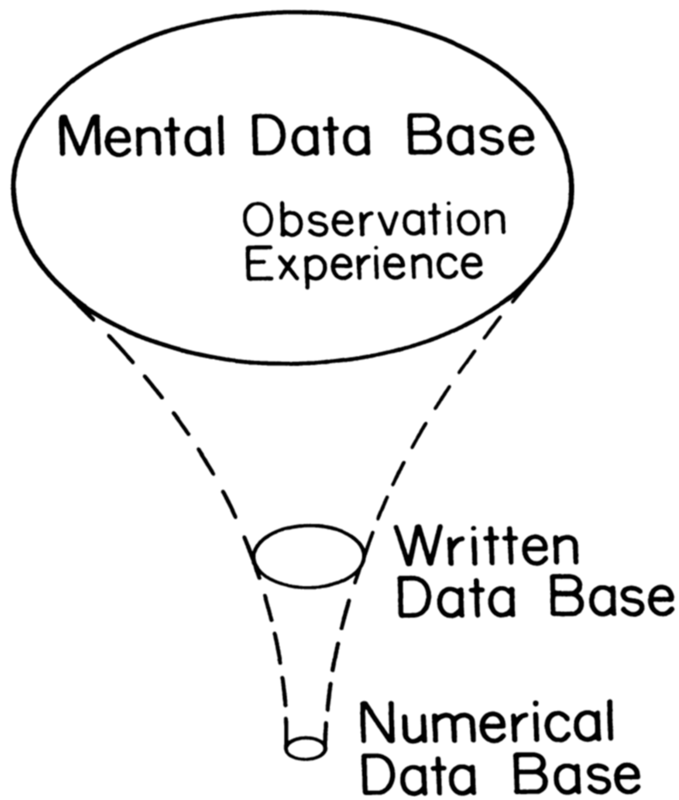
- Model building
  - Data and Modeling
- Parameter Estimation
  - Calibration
  - Confidence Intervals
  - Role of noise
- Model Analysis
  - Analysis process (time step, equilibrium, causal tracing, loop knock-out, loop dominance, test understanding)
  - Controlling dynamic systems
  - Synergies with other areas of analytics

Use of Data in Simulation Models

- Use multiple types of data for multiple purposes and different stages of modeling
  - Qualitative data is more appropriate for identification of the structure of problem and important dynamic hypotheses
  - Numerical and archival data is needed for parameter estimation and validation
  - Time series can feed into model the factors outside of model boundary

Forrester's Representation of Three Data Bases

A. Mental Data Base and Decreasing Content of Written and Numerical Data Bases



From Qualitative to Quantitative Data

- Following the SD modeling process, you can get high structural reliability through extensive use of qualitative data
- The qualitative insights allow you to tell convincing stories
  - This is good: lots of insights; convincing
  - This is bad: you can fool the client; and yourself
- Hard data can keep you straight and allow for testing of hypotheses more rigorously
- It also makes your work much stronger, both in persuasion and impact

Two Potential Misconceptions

- Misconception 1: “We can’t build models without numerical data”… BUT
  - If model structure is to follow real-world structures underlying the problem, qualitative data is needed for building it
  - Insights can be generated before starting to use numerical data
- Misconception 2: “Numerical data adds little value to the modeling” … BUT
  - Uncertainties in parameter values can only be settled with empirical data
  - Many theoretical and practical implications depend on these parameters
    - E.g. any cost benefit requires reliable estimates and estimates of reliability

Agenda

- Model building
  - Data and Modeling
- Parameter Estimation
  - Calibration
  - Confidence Intervals
  - Role of noise
- Model Analysis
  - Analysis process (time step, equilibrium, causal tracing, loop knock-out, loop dominance, test understanding)
  - Controlling dynamic systems
  - Synergies with other areas of analytics

Calibration/Parameter Estimation

- **Find "best" parameters**
- **More complex than standard OLS regression** ($y = a_0 + a_1 x_1 + \ldots. + a_n x_n + e$)
  - Dynamic model
  - Endogeneity (feedback)
  - Time delays, nonlinearities, accumulations, missing data, autocorrelation, heteroskedasticity, etc.
- **Common approaches:**
  - Nonlinear least squares, Method of Simulated Moments, Simulated Maximum likelihood, etc.

Example: VCR model

- Three data series to fit simultaneously
  - Sales, price, adopters
- Weighted nonlinear least squares (is Maximum Likelihood under some assumptions)
  - Find parameters, θ, that minimize weighted sum of squared errors between data, $x^d{}_{it}$, and model, $x^m{}_{it}$:

    $$\Sigma_{it}[x^d{}_{it} - x^m(\theta)_{it}]^2/\sigma_i^2$$

Parameters

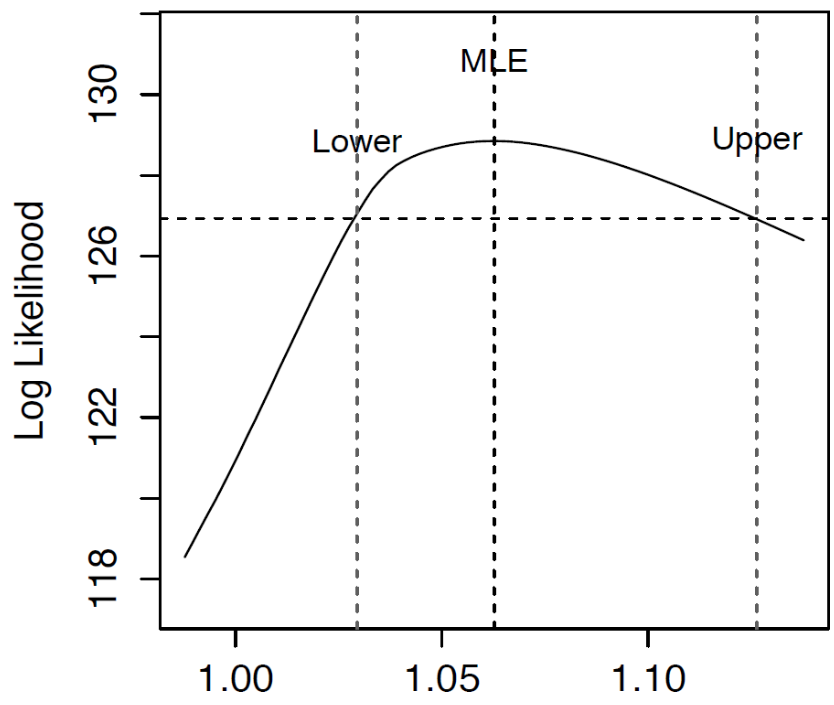
- Select parameters to be estimated
  - Only those that matter, and are not determined by physical definitions
- Select bounds for parameters based on
  - Physical constraints
  - Prior knowledge
  - Likely uncertainty range

Confidence Intervals

- Used for hypothesis testing and assessing uncertainty in parameters
- Estimated by various methods including asymptotic methods, bootstrapping, likelihood ratio, Markov Chain Monte Carlo, etc.

Basic idea:

How sensitive is the payoff function for calibration to changes in parameters?



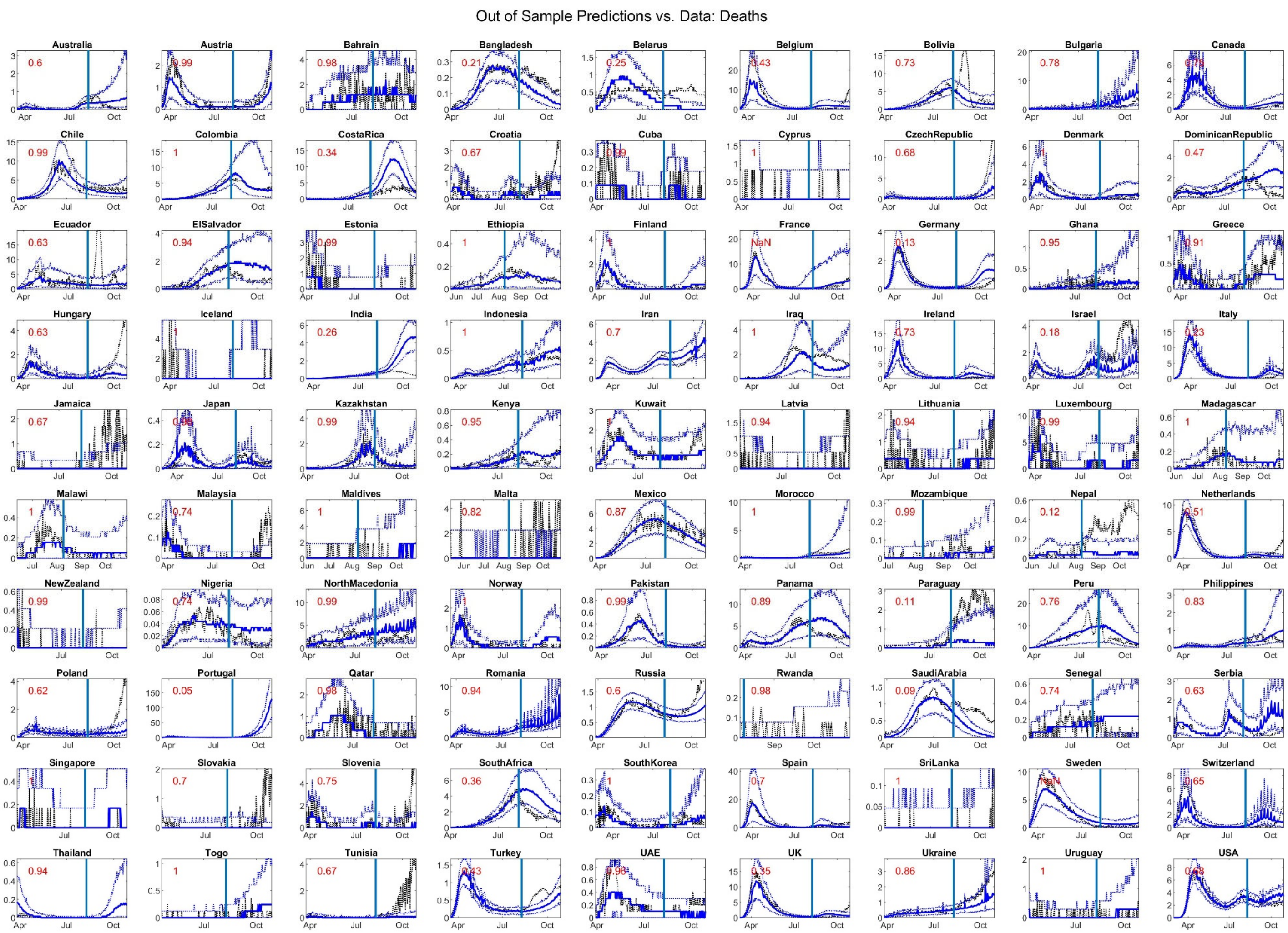
Model Calibration

- Include in calibration problem ALL knowledge available about system parameters
  - known parameters
  - physical constraints on parameters
  - likely uncertainty range
- Use the smallest calibration problems possible
  - immediacy of parameters to independent and dependent variables
- Test the hypotheses “The estimated parameter matches the observable structure of the system”
  - Does the model match the historical behavior
  - Does the model match the structure of real world system

Building Confidence in Calibration

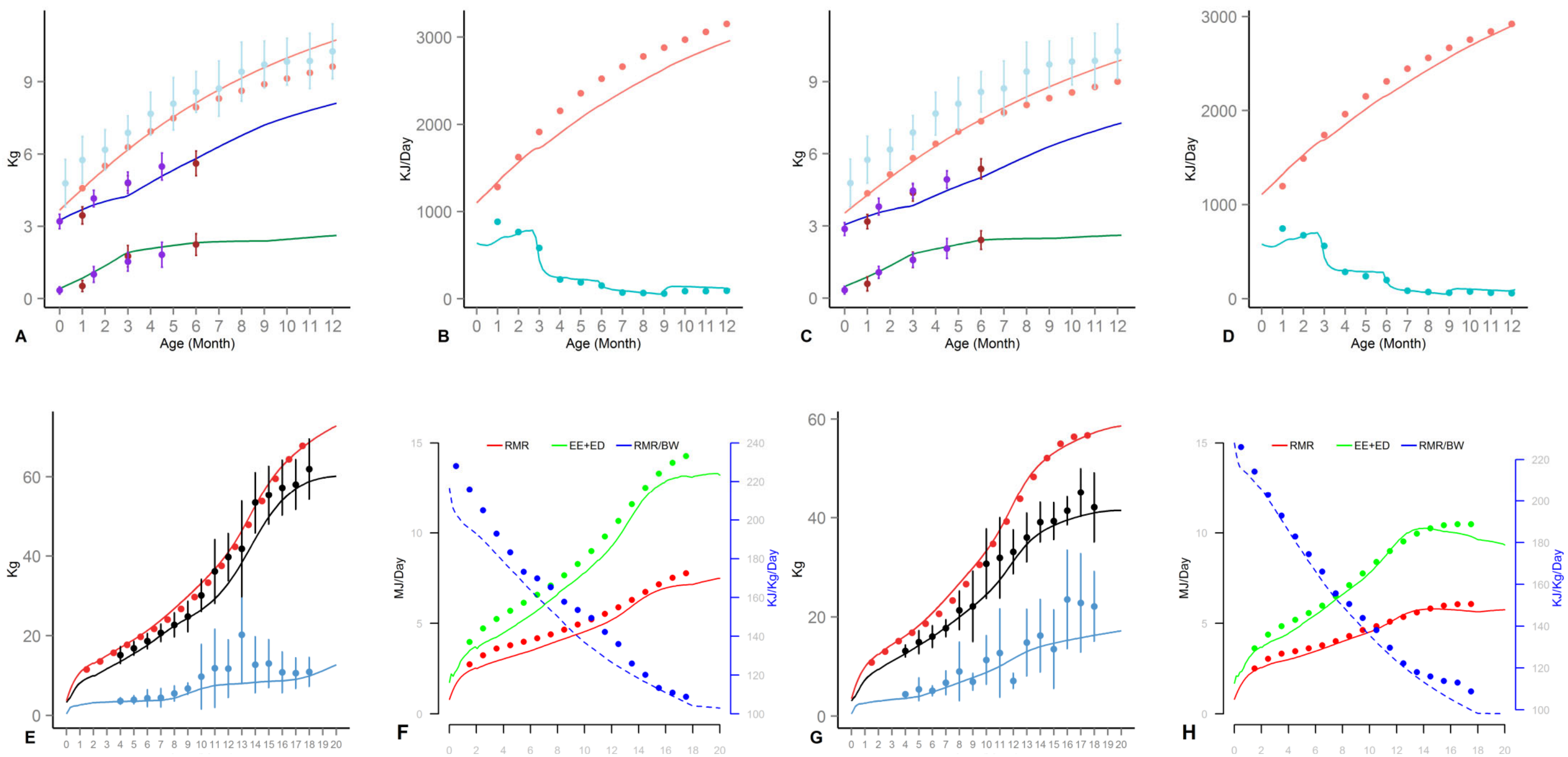
- Out of sample predictions
- Validation using other data
- Validation of estimation framework using synthetic data

Out of Sample Prediction for COVID-19



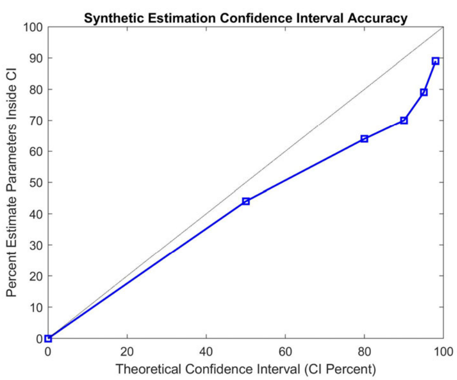
Validation: Compare model results against data not used in calibration

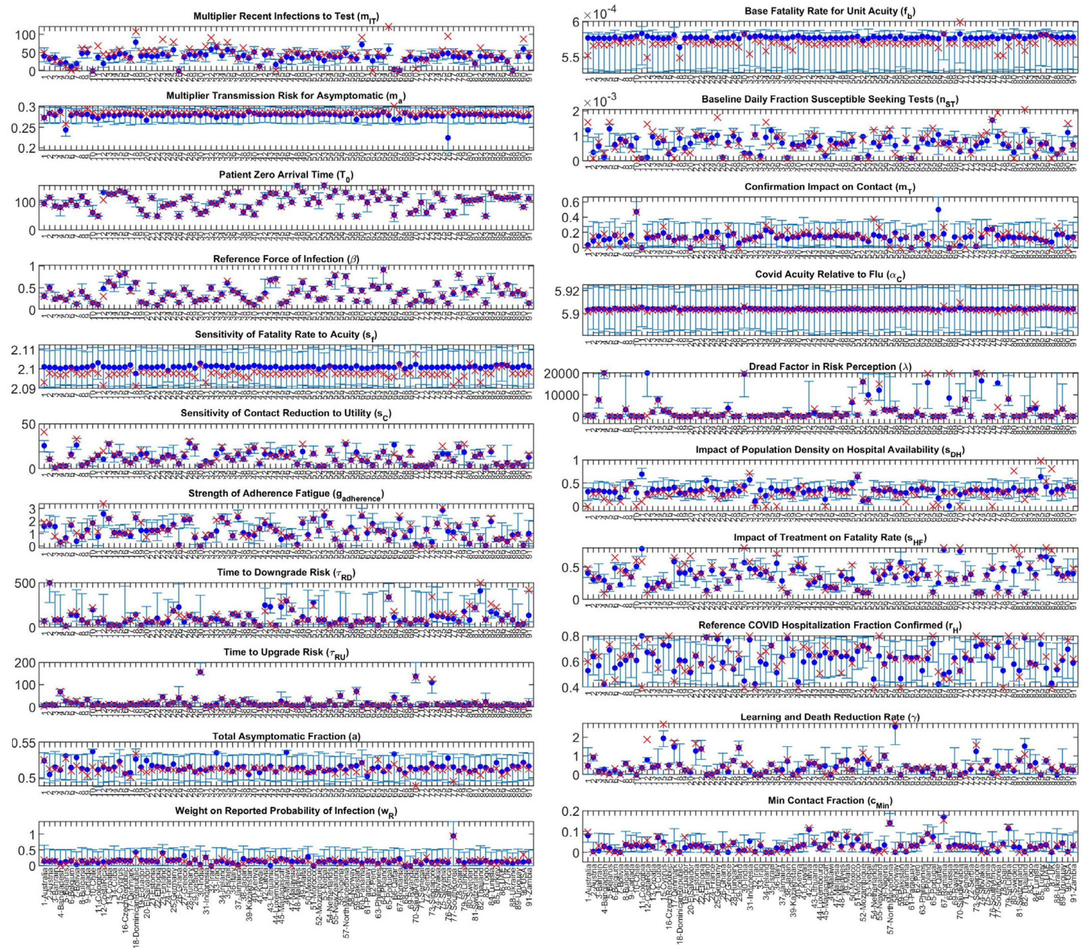
- Example: modeling human growth and weight dynamics (Rahmandad 2014)



Building confidence in estimation: **Synthetic Data**

- Median distance to true value: 12% of 95% CI

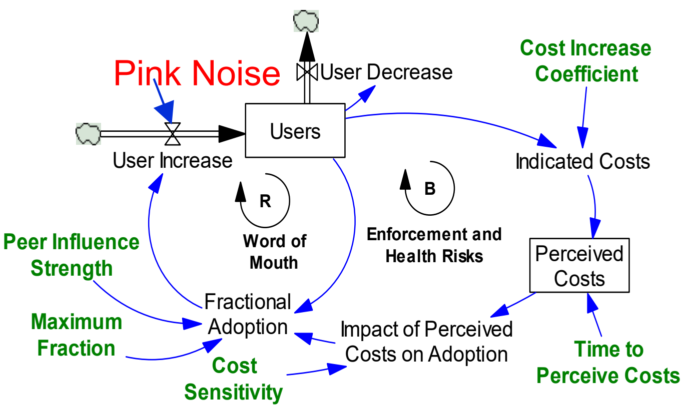


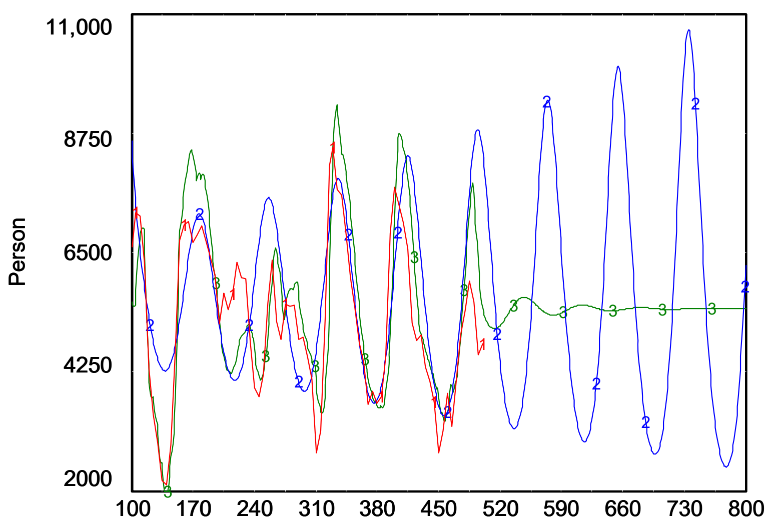


Endogeneity, Process, and Measurement Noise

- Complicate the estimation process and require more advanced tools, such as:
  - Approaches less sensitive to process noise such as method of simulated moments, indirect inference
  - Filtering and state resetting
    - Extended Kalman filter and particle filters

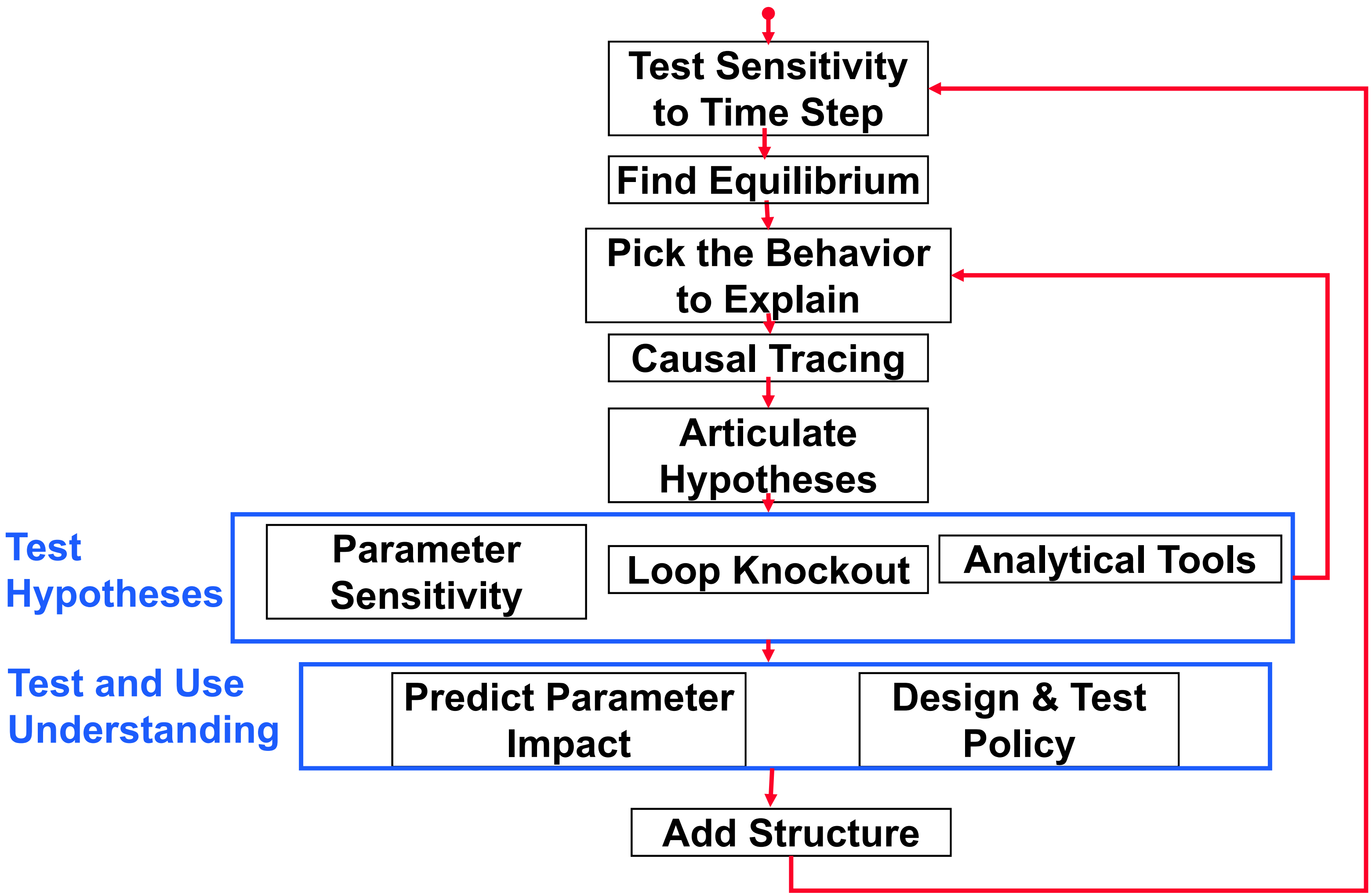
A simple oscillatory model (waves of elicit drug users over time)





- **1: Data**
- **2: Simple Calibration**
- **… and prediction**
- **3: Calibration using EKF**

Analysis Process



Test for Sensitivity to Time Step

- You don’t want to explain a behavior that is artifact of integration error. So:
  - Divide the time-step by two and see if there is any noticeable difference in behavior
  - Continue dividing until the behavior does not change
- This is very simple, but a lot of novice modelers forget it!

Find Equilibrium

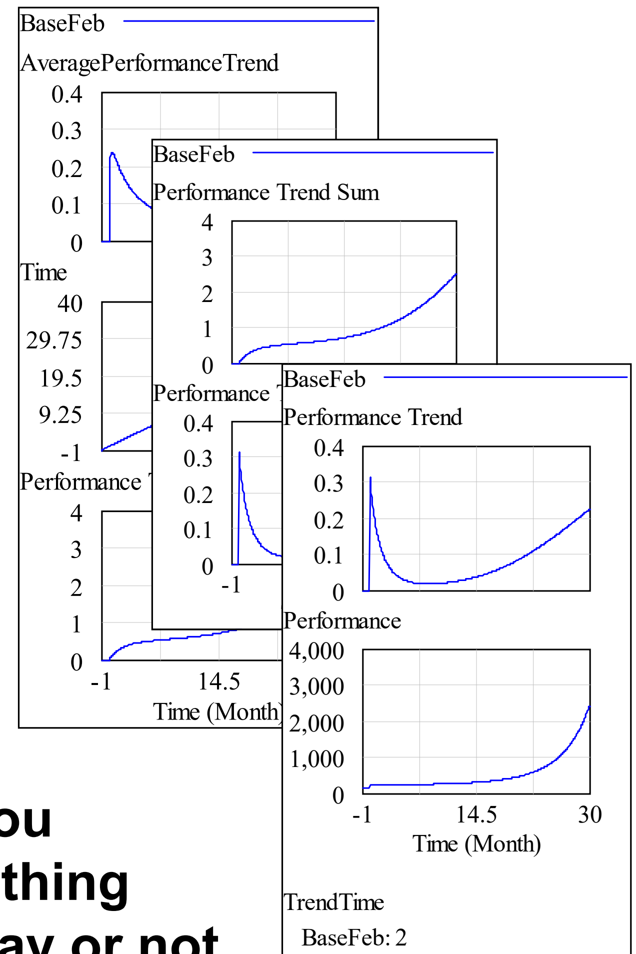
- Avoid confusing transient dynamics with inherent dynamics
- Start the model from equilibrium
  - Analytical equilibrium preferred, but numerical is OK if no better option available
- Perturb it into the mode of behavior of interest at some later time using step, pulse, or other exogenous inputs
- If no general equilibrium, put as many stocks as possible into equilibrium

Pick the Behavior to Explain

- Different variables may have different modes of behavior, and the sources of dynamics are not always the same
- Pick one variable's behavior and understand that
- Usually by doing this for 2-3 variables, you will know the model's behavior completely

Causal Tracing and Hypotheses

- Look at the causes for each variable and trace back until:
  - Single source is found (e.g. a draining, isolated stock)
  - A loop hypothesis is found (you loop back to the original variable)
  - You form other hypothesis why the behavior happens
- Design an experiment to test your hypothesis:



if hypothesis is X➔ Y, then you remove X (while changing nothing else) to see if Y also goes away or not.

Loop Knockout (if X is a loop)

- Single out the loop/causal path you hypothesize is deriving behavior
- Find parameters that change that loop's strength (and only that loop's) or create switches to shut the loop off without affecting other loops
  - Shut-off=Keep the impact as equilibrium/before
  - Example: if output=f(input) and $ouput_{EQ}=O_{EQ}$
    - Change to: $output=SW*f(input)+(1-SW)*O_{EQ}$; $0 \le SW \le 1$
- Change the strength of the loop and observe the impact on the behavior
- Sometimes multiple loops are contributing to the behavior: you may need to test more than one loop simultaneously

Structural Dominance Analysis

- Objectives of SDA
  - Articulate structural explanations for behavior
  - Support policy design

- Three flavors of SDA
  - Exploratory analysis of dominant structure
    - Systematic / exhaustive approaches
  - Formal assessment of dominant structure
    - Eigenvalue/eigenvector analysis of linearized model
    - Pathway Participation Method
      - A related tool now built into Stella Architect software

Test and Use Your Understanding

- Predict parameter impact
  - List the model parameters in a table
  - Predict what will happen if you change each (e.g. the mode increases/decreases)
  - Test and see if your prediction was correct
  - When wrong, find out what you missed
- Help build client's mental models
  - Design simple simulations to communicate the basic insights to your client
  - Be ready to explain in simple terms why the model behaves the way it does
  - Go through likely scenarios and their mechanisms
  - Engage client in explaining dynamics

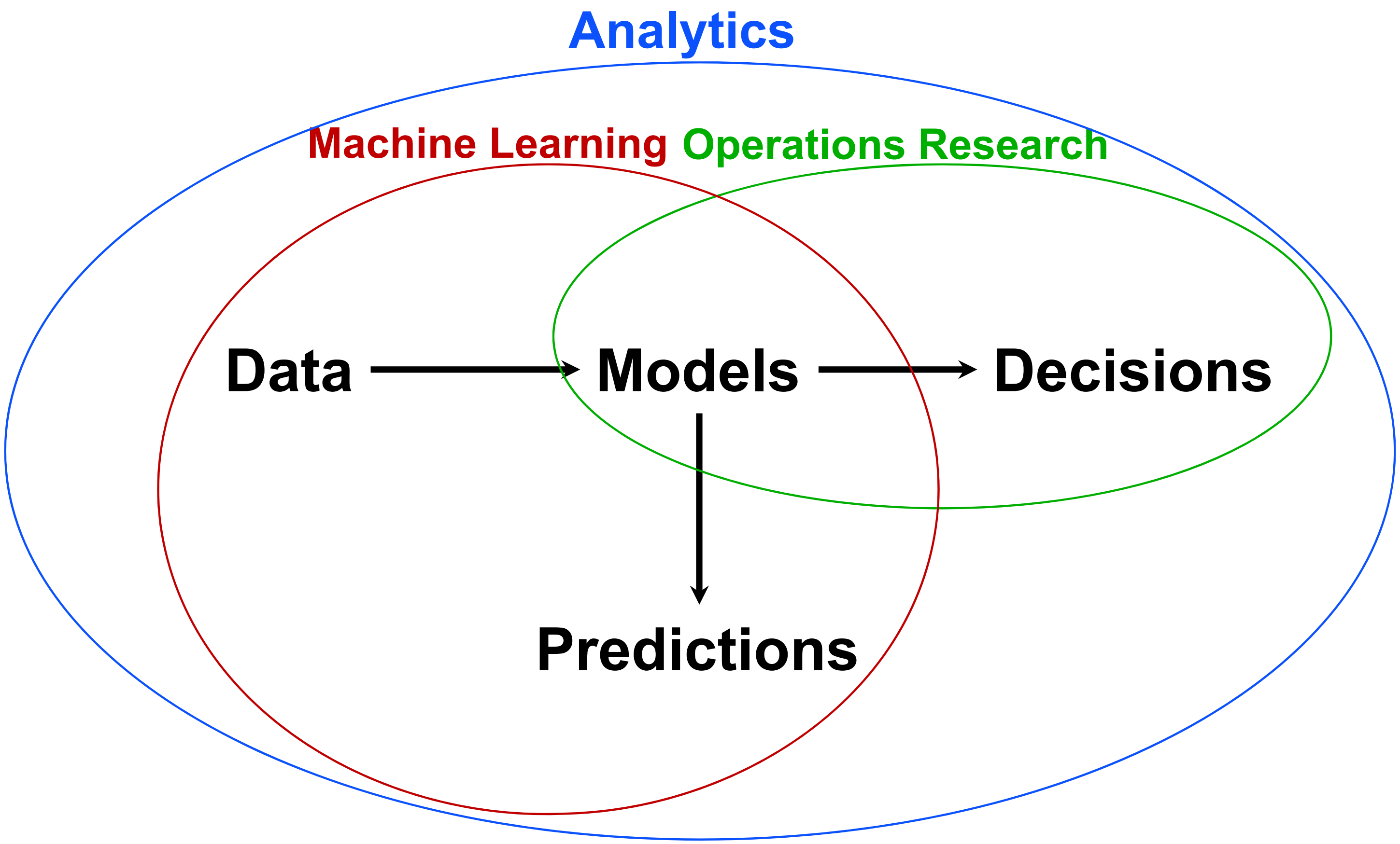
Controlling Dynamic Systems

- **Overview**: you have (a model of) a dynamic system, how can you use this model to best manage (control) the system?
  - You are the main actor➔ e.g. Managing your inventory, vehicle control, epidemic planning
    - Classical and optimal control, (approximate) dynamic programming and policy optimization
    - Decision analysis, decision trees, stochastic optimization
  - There are other (rational) actors with different or opposing goals ➔e.g. Pricing in competition, missile defense, Market entry decisions
    - Dynamical games

Analytics

- Descriptive
  – E.g. Summary statistics; visualization; clustering
- Predictive
  – E.g. Linear regression; logistic regression; CaRT; Random forests; Deep learning; …
- Prescriptive
  – Optimization; building on prescriptive

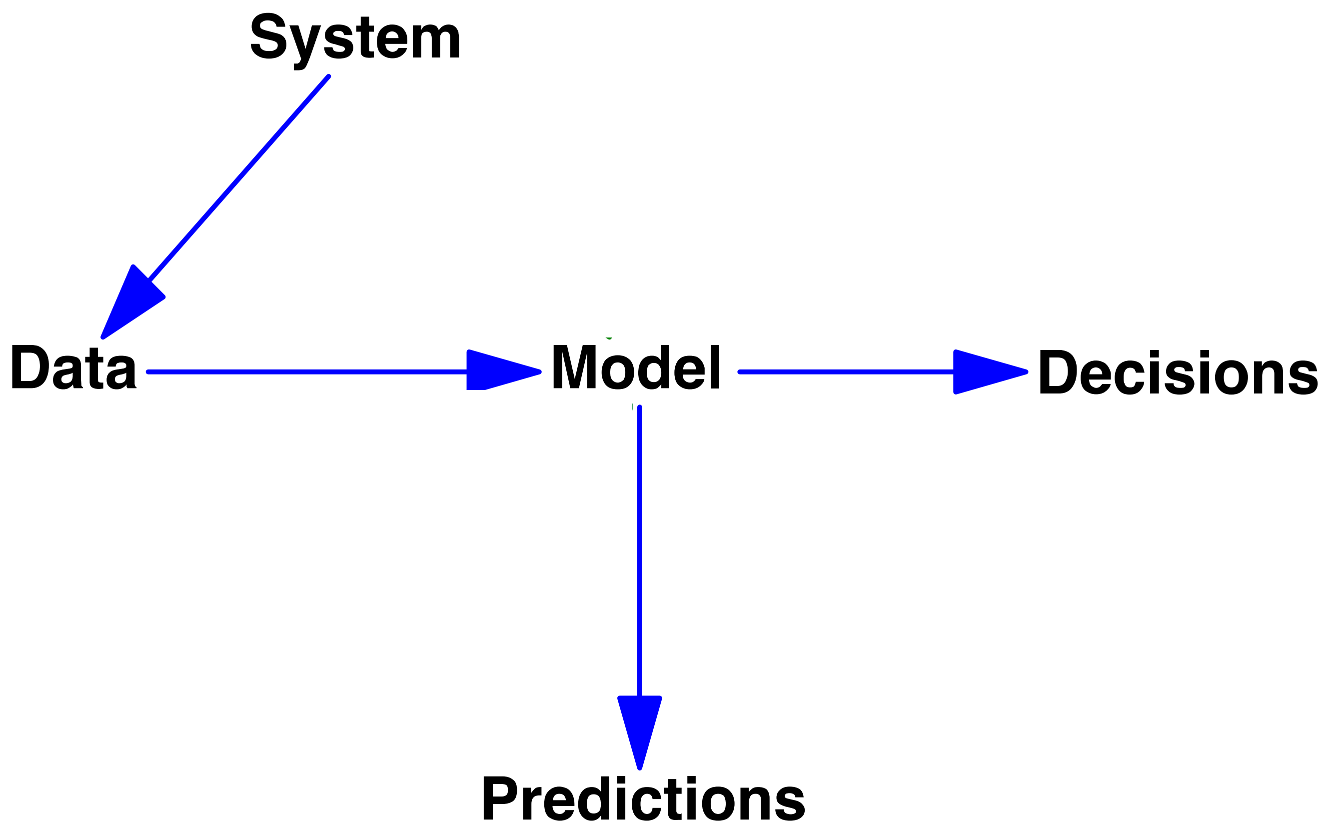
Analytics


Analytics
Machine Learning
Operations Research
Data
Models
Decisions
Predictions

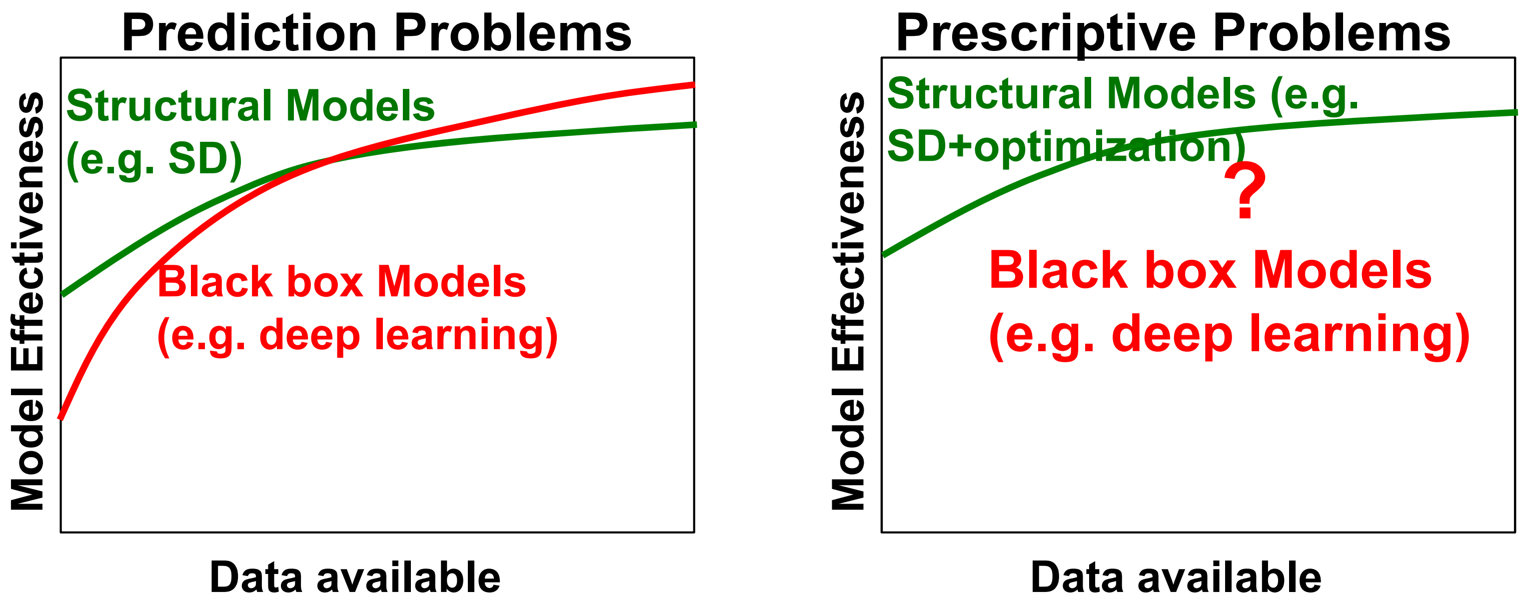
SD Models and Data

- The model is built from data beyond the numerical base
- Emphasis on operational explanations means the focus of SD is in describing the **system** that generates the data
  - Question the source of the data
  - Question the quality of the data, e.g.
    - Violations of conservation of matter
    - Biases on data collection and classification
- Basic triangulation of sources

SD and Analytics



Structural and Black-box models for Prediction, and Policy Design



SD and Analytics

- Example areas for further integration
  - Embed machine learning algorithms in the decisions ***in*** SD model
    - E.g. designing an agent that stabilizes supply chain oscillations
  - Automate cross-validated calibration with structural experiments ***on*** SD models
  - Identify/detect regularities of SD model behavior to further develop the inquiry
    - E.g. using Lasso regressions to identify simple rules for managing complex systems

SD and Analytics (continued)

- Use advances in statistical inference to calibrate SD models more effectively
- Meta-models: Calibrate a simpler/faster model (e.g. a neural net) to fit an SD model (using simulation data), then use the resulting model for optimization, or even for calibration to actual data
- Use reinforcement learning with deep NN to design control policies for SD models