What modeling data-rich systems taught me about validating epidemiological models

Stephen Eubank

University of Virginia Biocomplexity Institute & Initiative

Technical Report BI-2021-62

May 10, 2021

Verification, Validation, and Uncertainty Quantification Across Disciplines

NSF #1918656 Expeditions: Collaborative Research: Global Pervasive Computational Epidemiology

NSF #2033390 RAPID: Improving Computational Epidemiology with Higher Fidelity Models of Human Behavior

What is the question?

???

The obvious question is how the highly developed techniques for V&V [and UQ] in the data-rich (principally engineering) environments can nevertheless make contact with the far more constrained modeling environments defined by disciplines ranging from astrophysics to the social sciences.

Data-rich (more or less) systems I have simulated

- The "standard model" of particle theory
- Experiments in nonlinear systems, e.g., fluid dynamics
- Financial market time series
- Environmental time series (CO_2-temperature-insolation)
- Natural language part-of-speech tagging and parsing
- Regional transportation systems
- Mitigation after a WMD attack
- Infectious disease epidemics:
  influenza, Ebola, cholera, malaria, Zika, COVID-19

Simulating agent based models of socio-technical systems

- Synthesize populations (geographic + demographic distribution) from administrative data: census, land-use, etc.
- Model individual behaviors using active and passive, longitudinal and snapshot, academic and proprietary observations:
  - American Community Survey, Pew Research surveys,
  - Behavioral Risk Factor Surveillance System, National Health and Nutrition Examination Survey, Framingham Heart Study
  - CMU/Facebook COVIDcast, Twitter, cell phone mobility
- Place individuals in context (interaction networks)
- Vary resolution, scale, and fidelity

Goal

An evidence base to inform policy-makers' decisions about complex systems.

The general's question

(apologies to any generals in the audience)

The general's question

Have you validated this model?

(apologies to any generals in the audience)

The modeler's answer

(No apologies – let's get this out of the way for the week)

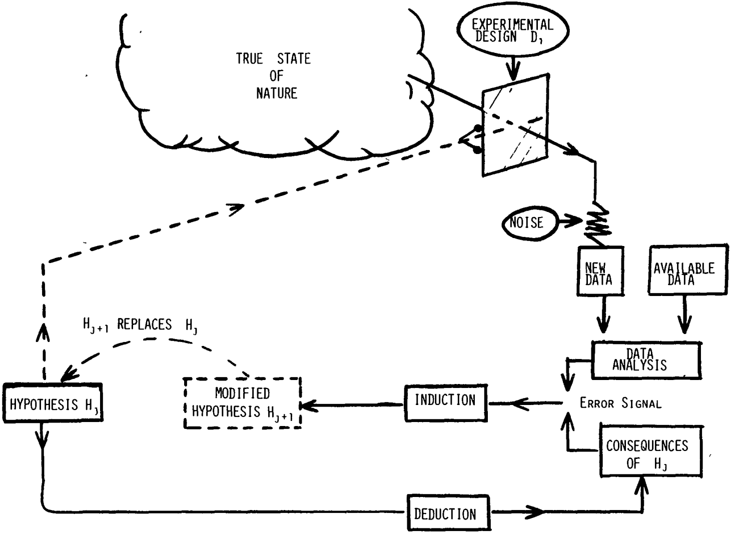
The modeler's answer

All models are wrong
but some are useful

– G.E.P. Box

(No apologies – let's get this out of the way for the week)

B. Data Analysis and Data Getting in the Process of Scientific Investigation[a]

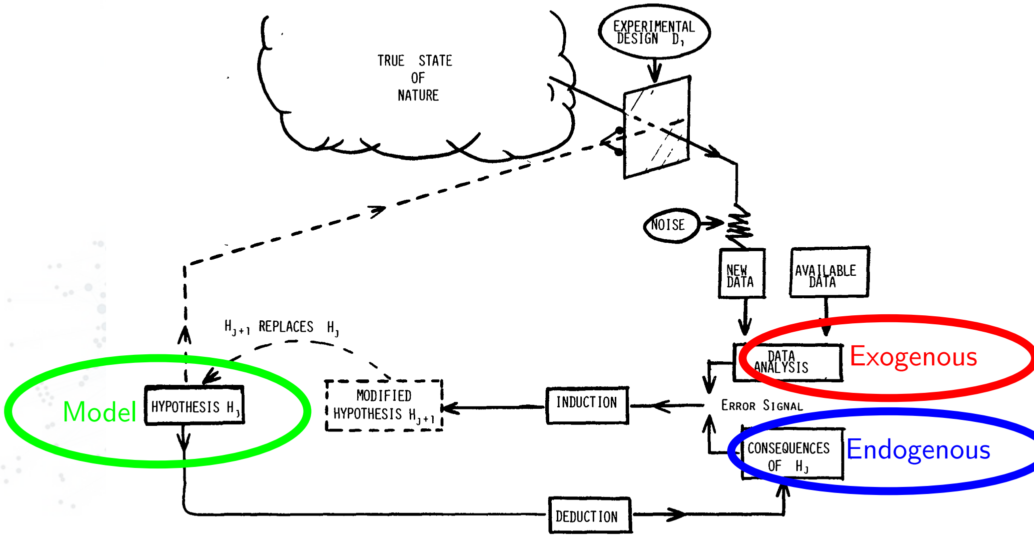


[a] The experimental design is here shown as a movable window looking onto the true state of nature. Its positioning at each stage is motivated by current beliefs, hopes, and fears.

Statistics and the Environment, GEP Box, Int'l Statistical Review, **43** (3) p. 279-287 (1975)

Models

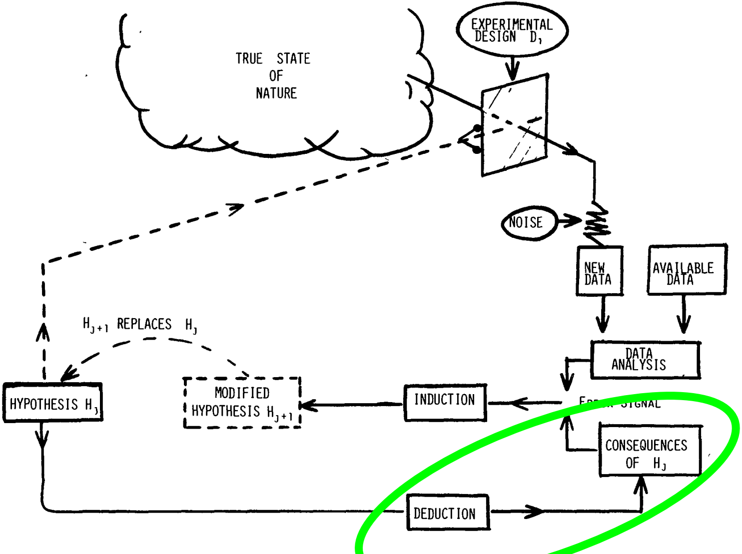
B. Data Analysis and Data Getting in the Process of Scientific Investigation[a]



[a] The experimental design is here shown as a movable window looking onto the true state of nature. Its positioning at each stage is motivated by current beliefs, hopes, and fears.

Simulation

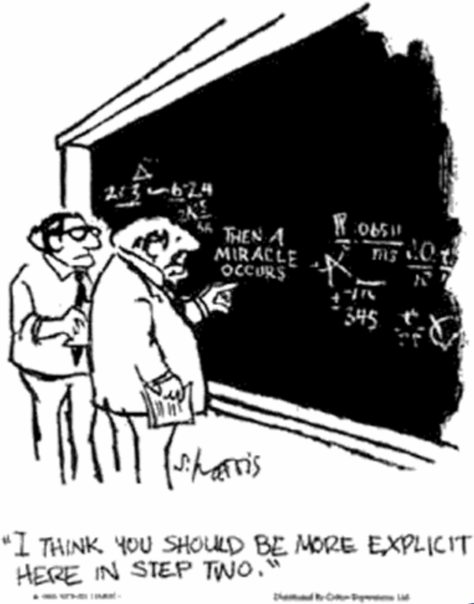
B. *Data Analysis and Data Getting in the Process of Scientific Investigation*[a]



[a] The experimental design is here shown as a movable window looking onto the true state of nature. Its positioning at each stage is motivated by current beliefs, hopes, and fears.

Does the simulation correctly deduce consequences of H_J?

Further remarks assume a well-specified model and simulation.



Ensuring internal consistency

- Decompose into small modules with well-defined interactions
  - small class of possible interactions
  - edge cases grow linearly vs exponentially
- Use synthetic data (e.g., synthetic population) as interface between modules
  - natural interface for most models
  - calibrate each module to available data
- Test against examples with known solutions
- Test at scale
- Replicability vs indeterminacy (?)

What could possibly go wrong?

- Continuous vs discrete math
- Shortcuts for scaling, e.g., mean field approximation, parallel computation
- Structure of the interaction network
- Order of interaction (especially important for parallelism)
- Do sources of randomness in the simulation model sources in the world?
- Mission creep: Is H_{J+1} in the same class of models as H_J?

Are data rich environments really data rich?

- Model complexity expands to fill the available data.
- N $\sim$ 1 for important systems that are extensively modeled
  – there is no Planet B
- What we see is almost certainly not a maximum likelihood outcome.

(Why) is it easier for physics?

Consider a spherical cow in isolation from the universe in the tail-free limit ...

Ensuring generalizability, extrapolation quality, out-of-sample performance, ...

Ensuring generalizability, extrapolation quality, out-of-sample performance, ...

Occam's razor, e.g., MDL, AIC, BIC, χ^2

It can be scarcely denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience.

"On the Method of Theoretical Physics" The Herbert Spencer Lecture, delivered at Oxford (10 June 1933),

as cited by en.wikiquote.org/wiki/Albert_Einstein

Ensuring generalizability, extrapolation quality, out-of-sample performance, ...

Occam's razor, e.g., MDL, AIC, BIC, χ^2

It can be scarcely denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience.

"On the Method of Theoretical Physics" The Herbert Spencer Lecture, delivered at Oxford (10 June 1933),

as cited by en.wikiquote.org/wiki/Albert_Einstein

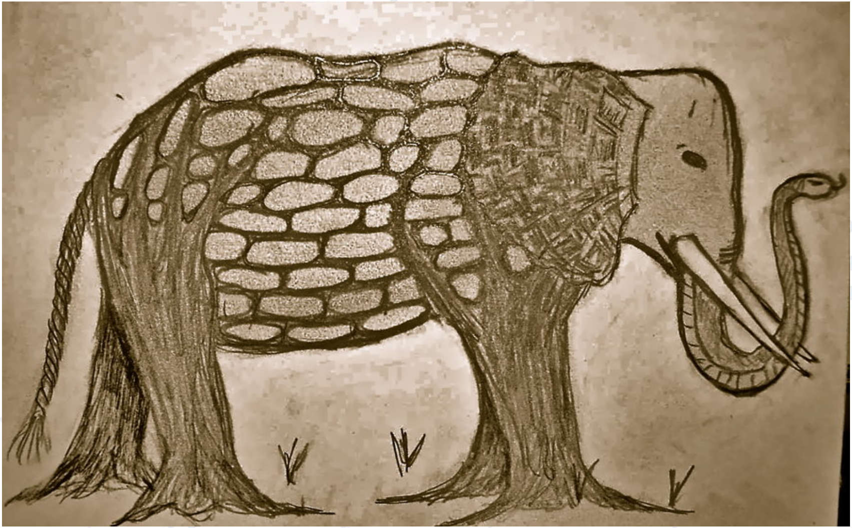
or, as rendered by composer Roger Sessions:

Everything should be made as simple as possible, but no simpler.

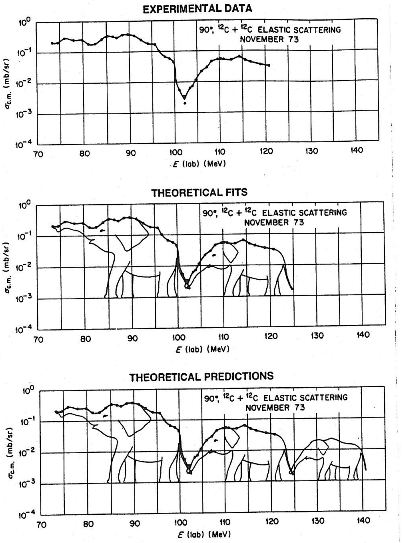
"How a 'Difficult' Composer Gets That Way", New York Times, 8 January 1950,

as cited by quoteinvestigator.com/2011/05/13/einstein-simple/

Simple models of complex systems are not necessarily better.



Simple models of complex systems are not necessarily better.



The process of fitting data, as seen by Subramanian Raman in *Science with a Smile.*

What is the real question?

???

The general wants to know:

"If I take actions based on this evidence, will I regret it?"

Will the resulting course of action lead to bad consequences?
Will the consequences be worse than for other courses of action?
Will others believe that other courses of action would have been worse?
Will my name end up in headlines as a fool or as a sage?

What is the real question?

???

The general wants to know:

"If I take actions based on this evidence, will I regret it?"

Will the resulting course of action lead to bad consequences?
Will the consequences be worse than for other courses of action?
Will others believe that other courses of action would have been worse?
Will my name end up in headlines as a fool or as a sage?

"Validation" is shorthand for "building confidence in models."

– Naomi Oreskes(?)

What is the real question?

???

The general wants to know:

"If I take actions based on this evidence, will I regret it?"

Will the resulting course of action lead to bad consequences?
Will the consequences be worse than for other courses of action?
Will others believe that other courses of action would have been worse?
Will my name end up in headlines as a fool or as a sage?

"Validation" is shorthand for "building confidence in models."

– Naomi Oreskes(?)

Answers may not even lie in the realm of statistics.

The validation commutative diagram

$$\begin{tikzcd}
\textit{Condition 1} \arrow[r, "\textit{time / space}", "\textit{demography}"'] \arrow[d] & \textit{Condition 2} \arrow[d] \\
\textit{System state} \arrow[d, "\textit{observation}"', "\textit{analysis}"] & \textit{System state} \\
\textit{Model representation} \arrow[r, "\textit{modeled \& simulated}", "\textit{interactions}"'] & \textit{Model representation} \arrow[u, "\textit{implications}"']
\end{tikzcd}$$

What could possibly go wrong?

- State assessment – observation / analysis errors.
- Model does not provide faithful representation of system state.
- Misunderstood model: incorrect implications from distribution of outcomes.
- Mis-specified model: model class doesn't contain true dynamics.
- Model is over-fit, especially with machine-learning.
- Parameters are not identifiable: the flip side of universality.
- For extrapolation, $N = 0$.
- Machine learning is incredibly powerful

These are all common in social science, with
many confounders,
few randomized controlled trials,
and biased samples.

Machine learning is incredibly powerful

Cautionary tale of a gold model:

- Data sources
- Handling outliers
- Train + Cross-validate + out-of-sample
- Choose when you want to forecast
- Complexity control vs power of combinatorial search

State representation ∼ "the chicken or the egg"

As for the taxonomy of large sociograms [social networks], this apparently involves problems of great complexity. It would seem offhand that a taxonomy of "nets" [. . .] would arise naturally from the consideration of the statistical parameters, e.g. as a continuum of nets in the parameter space. But the statistical parameters themselves are singled out on the basis of taxonomic considerations, which have yet to be clarified.

– Rapaport and Horvath, Behav Sci., **6**, p. 279–291 (1961))

Does consistency even matter? It depends on how results will be used.

"Track before detect" in surveillance systems, e.g. Sherlock Holmes, Feynman:

- Hypothesize possible tracks
- Discard hypothetical tracks as observations invalidate them
- Add new tracks consistent with what's been seen

What remains, however unlikely, is ...

the truth? our best estimate? an evidence base for policy formation?

Weakening confidence in models

- Adaptive modeling – is non-stationarity in the eye of the beholder?
- Chasing statistical significance (p-value) at the expense of importance.
- Spurious robustness – model is insensitive to relevant parameters.
- Model extrapolated far from testing conditions
- Model is not transparent, assumptions are hidden
- Dueling models – given 10 contradictory models, choose the magic talisman
- Consensus results due to groupthink
- Restating results – does the "g-2" experiment portend the death of the Standard Model or just a higher-order perturbation?
- The scientific process – dietary guidelines, mask-wearing guidance
- 

Building confidence in models

- Modularity, revisited
  - Interactions are more believable than state estimates?? Why?
  - Modules should be extended, not replaced, e.g. Newton $\rightarrow$ Einstein.
  - Understand variability in each module
- Well-designed simulation experiments to model natural experiments.
- Observe unexpected (emergent) phenomena.
- Ensembling (in practice, but I'm not convinced)
- Simplify in ways that reflect [Public Health] understanding, not mathematical / computational convenience.

Food for thought

LeVerrier [...] died famous for discovering two planets. He used Newton's Laws to predict the location of Neptune based on 'irregularities' in the observed time series of Uranus's orbit, and that planet was duly observed. He also analysed 'irregularities' in the orbit of Mercury, and again told observers where to find another new planet. And they did: the new planet, named Vulcan, was very near the Sun and difficult to see, but it was observed for decades.

–L. Smith, "Chaos: A Very Short Introduction", Oxford University Press, p. 57 (2007)