



briefing note
414
May 2020

No, senator, science can't do away with models

Foster Langbein, Chief Technology Officer, Risk Frontiers

The following article was written in response to COVID-19 pandemic modelling but has a particular resonance with why we make CAT models and how and why they change. CAT models explore some interesting territory – integrating, as they do, a myriad of sources from models of key ‘hard science’ physical processes, historical data, assumptions about geographic distribution, engineering assumptions and interpretations of building codes through to models of financial conditions from policy documents. Integrating such disparate sources becomes severely intractable mathematically when more than a few different distributions and their associated uncertainties are involved. The solution – Monte-Carlo simulation – harks back to the 1940’s and was critical in the simulations required in the Manhattan Project – in which, incidentally, a young Richard Feynman (quoted in the article) was involved. This powerful technique of random sampling a great number of times only became practical with the advent of computers - so computer models of CAT events are here to stay. But the essential point remains – they are just tools to help us understand the consequences of all the assumptions we input. When better science emerges or new data are incorporated and these assumptions are updated – changes are expected! Navigating those assumptions and helping understand the consequences and inevitable changes are part and parcel of Risk Frontiers modelling work. In what follows, Scott K. Johnson explains why U.S. Senator John Cornyn’s critique of modeling is misguided.

On Friday, Texas Senator John Cornyn took to Twitter with some advice for scientists: models aren't part of the scientific method. Scientists have responded with a mix of bafflement and exasperation. And Cornyn’s misconception is common enough - and important enough -that it’s worth exploring.

@JohnCornyn: After #COVID—19 crisis passes, could we have a good faith discussion about the uses and abuses of "modeling" to predict the future? Everything from public health, to economic to climate predictions. It isn't the scientific method, folks.

Cornyn’s beef with models echoes a talking point often brought up by people who want to reject inconvenient conclusions of systems sciences. In reality, “you can make a model say anything you want” is about as potent an argument as “all swans are white.” The latter is either a disingenuous argument, or you have an embarrassingly limited familiarity with swans.

Models aren't perfect. They can generate inaccurate predictions. They can generate highly uncertain predictions when the science is uncertain. And some models can be genuinely bad, producing useless and poorly supported predictions. But the idea that models aren’t central to science is deeply and profoundly wrong. It’s true that the criticism is usually centered on mathematical simulations, but these are just one type of model on a spectrum—and there is no science without models.

What's a model to do?

There's something fundamental to scientific thinking - and indeed most of the things we navigate in daily life: the conceptual model. This is the image that exists in your head of how a thing works. Whether studying a bacterium or microwaving a burrito, you refer to your conceptual model to get what you're looking for. Conceptual models can be extremely simplistic (turn key, engine starts) or extremely detailed (working knowledge of every component in your car's ignition system), but they're useful either way.

As science is a knowledge-seeking endeavor, it revolves around building ever-better conceptual models. While the interplay between model and data can take many forms, most of us learn a sort of laboratory-focused scientific method that consists of hypothesis, experiment, data, and revised hypothesis.

In a now-famous lecture, quantum physicist Richard Feynman similarly described to his students the process of discovering a new law of physics: "First, we guess it. Then we compute the consequences of the guess to see what... it would imply. And then we compare those computation results to nature... If it disagrees with experiment, it's wrong. In that simple statement is the key to science."

In order to "compute the consequences of the guess," one needs a model. For some phenomena, a good conceptual model will suffice. For example, one of the bedrock principles taught to young geologists is T.C. Chamberlin's "method of multiple working hypotheses." He advised all geologists in the field to keep more than one hypothesis - built out into full conceptual models - in mind when walking around making observations.

That way, instead of simply tallying up all the observations that are consistent with your favored hypothesis, the data can more objectively highlight the one that is closer to reality. The more detailed your conceptual model, the easier it is for an observation to show that it is incorrect. If you know where you expect a certain rock layer to appear and it's not there, there's a problem with your hypothesis.

There is math involved

But at some point, the system being studied becomes too complex for a human to "compute the consequences" in their own head. Enter the mathematical model. This can be as simple as a single equation solved in a spreadsheet or as complex as a multi-layered global simulation requiring supercomputer time to run.

And this is where the modeler's adage, coined by George E.P. Box, comes in: "All models are wrong, but some are useful." Any mathematical model is necessarily a simplification of reality and is thus unlikely to be complete and perfect in every possible way. But perfection is not its job. Its job is to be more useful than no model.

Consider an example from a science that generates few partisan arguments: hydrogeology. Imagine that a leak has been discovered in a storage tank below a gas station. The water table is

close enough to the surface here that gasoline has contaminated the groundwater. That contamination needs to be mapped out to see how far it has traveled and (ideally) to facilitate a cleanup.

If money and effort was no object, you could drill a thousand monitoring wells in a grid to find out where it went. Obviously, no one does this. Instead, you could drill three wells close to the tank, determining the characteristics of the soil or bedrock, the direction of groundwater flow, and the concentration of contaminants near the source. That information can be plugged into a groundwater model simple enough to run on your laptop, simulating likely flow rates, chemical reactions, and microbial activity breaking down the contaminants and so on, spitting out the probable location and extent of contamination. That's simply too much math to do all in your head, but we can quantify the relevant physics and chemistry and let the computer do the heavy lifting.

A truly perfect model prediction would more or less require knowing the position of every sand grain and every rock fracture beneath the station. But a simplified model can generate a helpful hypothesis that can easily be tested with just a few more monitoring wells - certainly more effective than drilling on a hunch.

Don't shoot the modeler

Of course, Senator Cornyn probably didn't have groundwater models in mind. The tweet was prompted by work with epidemiological models projecting the effects of COVID-19 in the United States. Recent modeling incorporating the social distancing, testing, and treatment measures so far employed is projecting fewer deaths than earlier projections did. Instead of welcoming this sign of progress, some have inexplicably attacked the models, claiming these downward revisions show earlier warnings exaggerated the threat and led to excessive economic impacts.

There is a blindingly obvious fact being ignored in that argument: earlier projections showed what would happen if we didn't adopt a strong response (as well as other scenarios), while new projections show where our current path sends us. The downward revision doesn't mean the models were bad; it means we did something.

Often, the societal value of scientific "what if?" models is that we might want to change the "if." If you calculate how soon your bank account will hit zero if you buy a new pair of pants every day, it might lead to a change in your overly ambitious wardrobe procurement plan. That's why you crunched the numbers in the first place.

Yet complaints about "exaggerating models" are sadly predictable. All that fuss about a hole in the ozone layer, and it turns out it stopped growing! (Because we banned production of the pollutants responsible.) Acid rain was supposed to be some catastrophe, but I haven't heard about it in years! (Because we required pollution controls on sulfur-emitting smokestacks.) The worst-case climate change scenario used to be over 4°C warming by 2100, and now they're projecting closer to 3°C! (Because we've taken halting steps to reduce emissions.)



briefing note
414
May 2020

These complaints seem to view models as crystal balls or psychic visions of a future event. But they're not. Models just take a scenario or hypothesis you're interested in and "compute the consequences of the guess." The result can be used to further the scientific understanding of how things work or to inform important decisions.

What, after all, is the alternative? Could science spurn models in favor of some other method? Imagine what would happen if NASA eyeballed Mars in a telescope, pointed the rocket, pushed the launch button, and hoped for the best. Or perhaps humanity could base its response to climate change on someone who waves their hands at the atmosphere and says, "I don't know, 600 parts per million of carbon dioxide doesn't sound like much."

Obviously these aren't alternatives that any reasonable individual should be seriously considering.

The spread of COVID-19 is an incredibly complex process and difficult to predict. It depends on some things that are well studied (like how pathogens can spread between people), some that are partly understood (like the characteristics of the SARS-CoV-2 virus and its lethality), and some that are unknowable (like the precise movements and actions of every single American). And it has to be simulated at fairly fine scale around the country if we want to understand the ability of hospitals to meet the local demand for care.

Without computer models, we'd be reduced to back-of-the-envelope spit-balling - and even that would require conceptual and mathematical models for individual variables. The reality is that big science requires big models. Those who pretend otherwise aren't defending some "pure" scientific method. They just don't understand science.

We can't strip science of models any more than we can strip it of knowledge.

<https://arstechnica.com/science/2020/04/no-senator-science-cant-do-away-with-models/>