

How a quest for mathematical truth and complex models can lead to useless scientific predictions

November 5 2022, by Arnald Puy



Credit: AI-generated image ([disclaimer](#))

A dominant view in science is that there is a mathematical truth structuring the universe. It is assumed that the scientist's job is to decipher these mathematical relations: once understood, they can be translated into mathematical models. Running the resulting "silicon

reality" in a computer may then provide us with useful insights into how the world works.

Since science keeps on revealing secrets, models keep getting bigger. They integrate discoveries and newly found mechanisms to better reflect the world around us. Many scholars assume that more detailed models [produce sharper estimates](#) and better predictions because they are closer to reality. But our new research, [published in *Science Advances*](#), suggests they may have the opposite effect.

The assumption that "more detail is better" cuts across disciplinary fields. The ramifications are enormous. Universities get more and more powerful computers because they want to run bigger and bigger models, requiring an increasing amount of computing power. Recently, the European Commission invested €8bn euros (£6.9bn) to create a very detailed simulation of the Earth (with humans), [dubbed a "digital twin,"](#) hoping to better address current social and ecological challenges.

In our latest research, we show that the pursuit of ever more complex models as tools to produce more accurate estimates and predictions may not work. Based on statistical theory and mathematical experiments, we ran hundreds of thousands of models with different configurations and measured how uncertain their estimations are.

We discovered that more complex models tended to produce more uncertain estimates. This is because new parameters and mechanisms are added. A new parameter, say the effect of chewing gum on the spread of a disease, needs to be measured—and is therefore subject to measurement errors and uncertainty. Modelers may also use different equations to describe the same phenomenon mathematically.

Once these new additions and their associated uncertainties are integrated into the [model](#), they pile on top of the uncertainties already

there. And uncertainties keep on expanding with every model upgrade, making the model output fuzzier at every step of the way—even if the model itself becomes more faithful to reality.

This affects all models that do not have appropriate validation or training data against which to check the accuracy of their output. This includes global models of climate change, hydrology ([water flow](#)), food production and epidemiology alike, as well as all models predicting future impacts.

Fuzzy results

In 2009, engineers created an algorithm called Google Flu Trends for predicting the proportion of flu-related doctor visits across the US. Despite being based on 50 million queries that people had typed into Google, the model wasn't able to predict the 2009 swine flu outbreak. The engineers then made the model, which is no longer operating, even more complex. But it still wasn't all that accurate. [Research led by German psychologist Gerd Gigerenzer](#) showed it consistently overestimated doctor visits in 2011–13, in some cases by more than 50%.

Gigerenzer discovered that a much simpler model could produce better results. His model predicted weekly flu rates based only on one teeny piece of data: how many people had seen their GP the previous week.

Another example is global hydrological models, which track how and where water moves and is stored. They started simple in the 1960s based on "evapotranspiration processes" (the amount of water that could evaporate and transpire from a landscape covered in plants) and soon got extended, taking into account domestic, industrial and agricultural water uses at the global scale. The next step for these models is to simulate water demands on Earth for every kilometer each hour.

And yet one wonders if this extra detail will not just make them even more convoluted. We [have shown](#) that estimates of the amount of water used in irrigation produced by eight global hydrological models can be calculated with a single parameter only—the extent of the irrigated area.

Ways forward

Why has the fact that more detail can make a model worse been overlooked until now? Many modelers do not submit their models to uncertainty and sensitivity analysis, methods that tell researchers how uncertainties in the model affect the final estimation. Many keep on adding detail without working out which elements in their model are most responsible for the uncertainty in the output.

It is concerning as modelers are interested in developing ever larger models—in fact, entire careers are built on complex models. That's because they are harder to falsify: their complexity intimidates outsiders and complicates understanding what is going on inside the model.

There are remedies, however. We suggest ensuring that models don't keep getting larger and larger for the sake of it. Even if scientists do perform an uncertainty and sensitivity analysis, their estimates risk getting so uncertain that they become useless for science and policymaking. Investing a lot of money in computing just to run models whose estimate is completely fuzzy makes little sense.

Modelers should instead ponder how uncertainty expands with every addition of detail into the model—and find the best trade-off between the level of model detail and uncertainty in the estimation.

To find this trade-off, one can use the concept of "effective dimensions"—a measure of the number of parameters which add uncertainty to the final output, taking into account how these parameters

interact with each other—which we define in our paper.

By calculating a model's effective dimensions after each upgrade, modelers can appraise whether the increase in [uncertainty](#) still makes the model suitable for policy—or, in contrast, if it makes the model's output so uncertain as to be useless. This increases transparency and helps scientists design [models that better serve science and society](#).

Some modelers may still argue that the addition of [model detail can lead to more accurate estimates](#). The burden of proof now lies with them.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation

Citation: How a quest for mathematical truth and complex models can lead to useless scientific predictions (2022, November 5) retrieved 26 April 2024 from <https://phys.org/news/2022-11-quest-mathematical-truth-complex-useless.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--