All disease models are 'wrong,' but scientists are working to fix that

May 5 2020, by Daniel Strain

The Zika virus can be transmitted by mosquitos from the species Aedes aegypti Credit: CDC

An international team of researchers has developed a new mathematical tool that could help scientists to deliver more accurate predictions of how diseases, including COVID-19, spread through towns and cities around the world.

Rebecca Morrison, an assistant professor of computer science at the University of Colorado Boulder, led the research. For years, she has run a repair shop of sorts for mathematical models—those strings of equations and assumptions that scientists use to better understand the

world around them, from the trajectory of climate change to how chemicals burn up in an explosion.

As Morrison put it, "My work starts when models start to fail."

She and her colleagues recently set their sights on a new challenge: epidemiological models. What can researchers do, in other words, when their forecasts for the spread of infectious diseases don't match reality?

In a study published today in the journal *Chaos*, Morrison and Brazilian mathematician Americo Cunha turned to the 2016 outbreak of the Zika virus as a test case. They report that a new kind of tool called an "embedded discrepancy operator" might be able to help scientists fix models that fall short of their goals—effectively aligning model results with real-world data.

Morrison is quick to point out that her group's findings are specific to Zika. But the team is already trying to adapt their methods to help researchers to get ahead of a second virus, COVID-19.

"I don't think this tool is going to solve any epidemiologic crisis on its own," Morrison said. "But I hope it will be another tool in the arsenal of epidemiologists and modelers moving forward."

When models fail

The study highlights a common issue that modelers face.

"There are very few situations where a model perfectly corresponds with reality. By definition, models are simplified from reality," Morrison said. "In some way or another, all models are wrong."

Cunha, an assistant professor at Rio de Janeiro State University, and his

colleagues ran up against that very problem several years ago. They were trying to adapt a common type of disease model—called a Susceptible, Exposed, Infected or Recovered (SEIR) model—to recreate the Zika virus outbreak from start to finish. In 2015 and 2016, this pathogen ran rampant through Brazil and other parts of the world, causing thousands of cases of severe birth defects in infants.
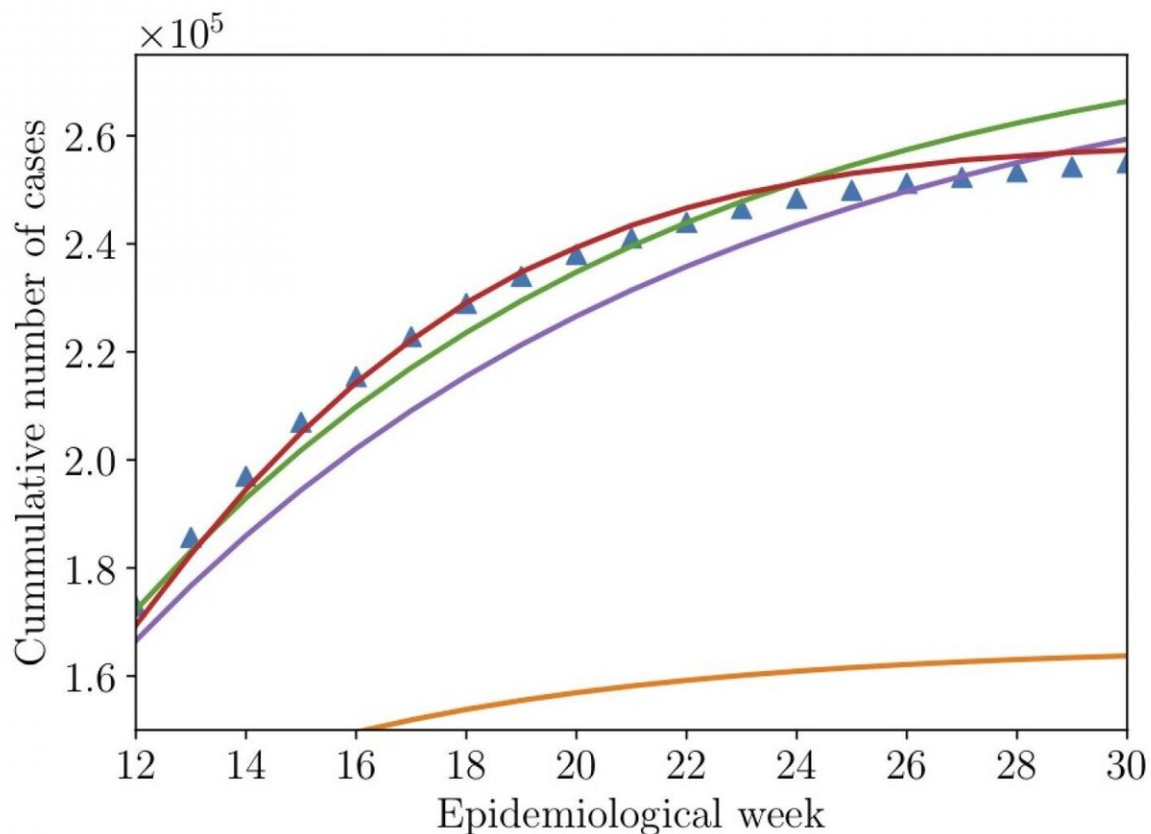
Top: A Zika virus health advisory posted near an airport security checkpoint in San Juan, Puerto Rico, in 2016; bottom: Graphic showing the success of different models in predicting recorded Zika cases in Brazil in the hundreds of thousands over the weeks since the outbreak began. Blue triangles represent the recorded cases, while the orange, green and purple lines show various versions of the original disease model. The red line corresponds to the "enriched" model that

Morrison and Cunha developed. Credit: CDC; Morrison and Cunha, 2020.

The problem: No matter what the researchers tried, their results didn't match the recorded number of Zika cases, in some cases miscalculating the number of infected people by tens of thousands.

Such a shortfall isn't uncommon, Cunha said.

"The actions you take today will affect the course of the disease," he said. "But you won't see the results of that action for a week or even a month. This feedback effect is extremely difficult to capture in a model."

Rather than abandon the project, Cunha and Morrison teamed up to see if they could fix the model. Specifically, they asked: If the model wasn't replicating real-world data, could they use that data to fashion a better model?

Enter the embedded discrepancy operator. You can picture this tool, which Morrison first developed to study the physics of combustion, as a sort of spy that sits within the guts of a model. When researchers feed data to the tool, it sees and responds to the information, then rewrites the model's underlying equations to better match reality.

"Sometimes, we don't know the correct equations to use in a model," Cunha said. "The idea behind this tool is to add a correction to our equations."

The method worked. After letting their operator do its thing, Morrison and Cunha discovered that they had nearly eliminated the gap between the model's results and public health records.

Being honest

The team isn't stopping at Zika. Morrison and Cunha are already working to deploy their same strategy to try to improve models of the coronavirus pandemic.

Morrison doubts that any disease model will ever be 100% accurate. But, she said, these tools are still invaluable for informing public health decisions—especially if modelers are up front about what their results can or can't tell you about a disease.

"This epidemic has revealed how difficult it is to model a real system," Morrison said. "But I hope that people don't take that to mean that we shouldn't trust our scientists."

More information: Rebecca E. Morrison et al, Embedded model discrepancy: A case study of Zika modeling, *Chaos: An Interdisciplinary Journal of Nonlinear Science* (2020). DOI: 10.1063/5.0005204

Provided by University of Colorado at Boulder