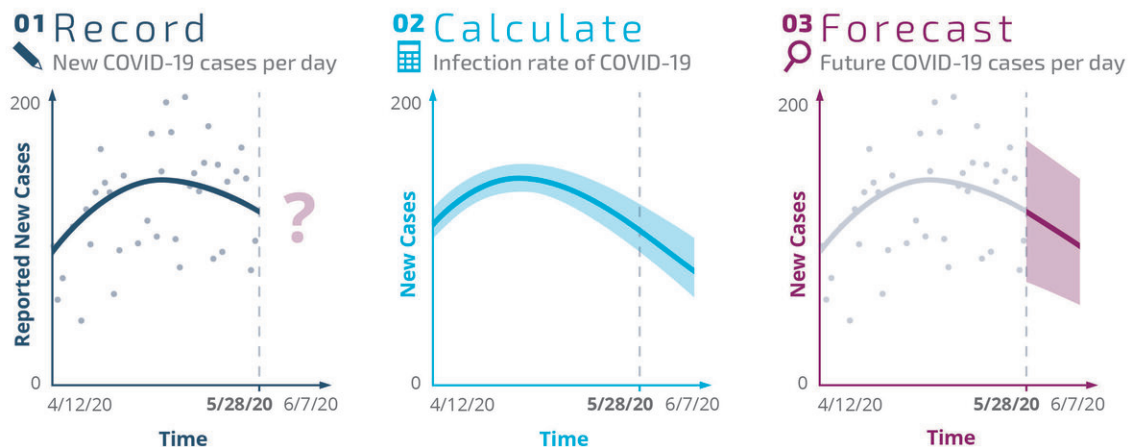# Researchers use public data to forecast new coronavirus cases

June 30 2020

Jaideep Ray and Cosmin Safta use recorded data and a calculated infection rate to predict future cases of the coronavirus. This example is based off of data from New Mexico from April 12 to May 28 which was then used to forecast new COVID-19 cases between May 28 and June 7. Credit: Sydney Spruiell

Global data networks that connect people through their devices have made it possible to create accurate short-term forecasts of new COVID-19 cases, using a method pioneered by two researchers at Sandia National Laboratories.

Jaideep Ray and Cosmin Safta used a model developed by Ray more than a decade ago to track plague epidemics using statistics. For

COVID-19 they also drew upon the advice of their Sandia co-workers with expertise in modeling, mathematics and software engineering.

"I first started using this method in 2008-09. Cosmin and I adapted it in 2010 to track influenza-like illnesses," Ray said. "When COVID-19 began to spread so rapidly, we knew we could use the same method to help forecast the outbreak."

Ray and Safta use publicly available data from the Centers for Disease Control and Prevention, The New York Times Data Repository, Johns Hopkins University and various state departments of health. Within minutes, and without the need for high-performance computing resources, the researchers can forecast new cases in a region or nationally for the next seven to 10 days. Since April, the number of new cases have roughly followed the trends predicted by Ray and Safta.

"This method is a relatively easy and inexpensive way to get short-term forecasts about new coronavirus cases that decision-makers can use to allocate health care resources and response," Safta explained. "This method is much easier and cheaper to do than methods that require more robust computers and manpower."

The range of accuracy for the predictions varies with the number of days out Safta and Ray are trying to forecast. So, while the number of cases have generally followed the trends predicted in the model within seven to 10 days, the method is not useful to predict more than 10 days out.

"The forecasts come with a range within which users can expect reality to lie," Ray said. "The range changes daily depending on the data, but the model ensures that the user can have 95% confidence that reality will fall within the range."

The project, which was funded through Sandia's Lab Directed Research

and Development program, provided national results to the National Virtual Biotechnology Laboratory team for publication on a DOE-run dashboard (funded by the U.S. Department of Energy Office of Science) for federal decision-makers. Specific results were also provided to the New Mexico Department of Health to guide regional responses throughout the state.

The data revealed by the forecasts can also gauge the impact of interventions over time. Ray and Safta said responding quickly to provide data on emerging outbreaks would not have even been possible five years ago.

"Since we are so connected today, it's possible to get an accurate number of COVID-19 cases in a day and get it to everyone in the world within a 24-hour period," Ray said. "Ten years ago, even five years ago, you could not get this data. In 2015, with the Ebola outbreak, by the time they got data it was pointless to try and make a forecast because it was already out of date and useless to decision-makers."

"For the current COVID-19 situation, having more sources of data dramatically assists our ability to create short-term forecasts to inform public health decisions," Safta concluded.

Provided by Sandia National Laboratories

Citation: Researchers use public data to forecast new coronavirus cases (2020, June 30) retrieved 2 May 2024 from https://medicalxpress.com/news/2020-06-coronavirus-cases.html