

# The Uncertain Future in How a Virus Spreads

A new model helps clarify the limits of pandemic predictions, which are notoriously difficult for the near future and impossible for longer timescales.

By **Susanna Manrubia**

On April 26th, 3000, a total solar eclipse will trace a path over South America, the Atlantic Ocean, and North Africa, a prediction no one disputes. While astronomers have long astonished humankind with amazingly precise forecasts, scientists trying to describe the future spread of pathogens must wrestle with far greater uncertainty. Researchers lack reliable data on how contagions spread, and small differences in their infectability, for example, can exponentially amplify the propagation of a virus, making it hard to predict the severity of epidemics and pandemics. A new study from Nigel Goldenfeld and Sergei Maslov, from the University of Illinois at Urbana–Champaign, and colleagues could help make those predictions a little easier [1]. Taking the state of Illinois as a case example, the team developed a model that assesses the short-term effect of nonpharmaceutical interventions on the spread of the virus that causes COVID-19.



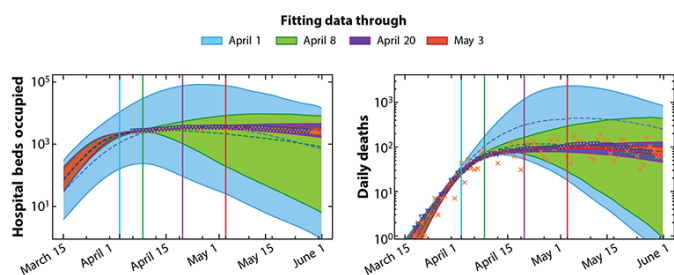
**Figure 1:** “Compartmental” models have been effective at reproducing the spread of COVID-19. These models group individuals into compartments, such as hospitalized (left), confined (middle), and infected (right), and then predict how the virus will spread depending on how people move between compartments. Credit: APS/Carin Cain

The researchers also explore and quantify the model’s limitations, clarifying what it cannot foretell. Their work is an exemplary illustration of the use of models to inform and guide citizens, clinicians, and policy makers.

Fewer questions have more immediate importance than that of when the COVID-19 pandemic will be contained. Is the curve flattening? Can the healthcare system withstand current and future demand? Will another wave of infections hit, and if so when? The need to answer these questions prompted an unparalleled response from the scientific community. Never before have we seen such an effort to predict the future. Nor have we before witnessed such a large number of failures [2, 3].

The vast majority of the models used to predict the spread of COVID-19 are so-called compartmental models, which group the individuals of a population into compartments [4]. The most commonly used compartments are susceptible (S), infected (I), and removed (R), a group that accounts for people who have either died or have recovered. Other compartments can include asymptomatic, confined (noninfected individuals with restricted mobility), quarantined, hospitalized, and vaccinated. More complex COVID-19 models include additional rules on top of the SIR dynamics that can account for geographical-dependent definitions of compartments and for the mobility of individuals, such as the likelihood that individuals will move from one place to another.

A compelling reason for the wide use of SIR models for studying the path of COVID-19 is that the models have proved astoundingly effective at reproducing the past [5]. But reproducing what has already happened does not mean a



**Figure 2:** The availability of accurate real-world data is key to training SIR-type models, such as that developed by Goldenfeld, Maslov, and their colleagues. The team shows that the spread of possible future outcomes significantly narrows when more data is input into the model.

Credit: G. N. Wong et al. [1]; adapted by APS/Carin Cain

model can accurately predict the future—the parameter values obtained to get the model to fit past data are not necessarily those needed to fit future data. SIR models have a number of limitations in that regard, which come from the inability of the models to represent the complex, distributed, and heterogeneous nature of the real world, and from the lack of accurate data to train the models [6, 7]. While these two issues can potentially be fixed, a third issue cannot: the predictions made using SIR models are particularly sensitive to small variations in the initial values assigned to model parameters. This sensitivity unavoidably restricts forecast capabilities to the near future [8]. There is also a fourth crucial player, the reactions of individuals and groups to the spread of COVID-19. These reactions are one of the main unknowns in all virus models but have been particularly problematic with COVID-19 [9].

Goldenfeld, Maslov, and their collaborators were well aware of the strengths and limitations of SIR models when they started constructing their own model, which they tailored to predicting COVID-19 propagation in Illinois [1]. Their model is a so-called age-of-infection variant, which is one that “remembers” when an individual arrives in a compartment and then uses that time to calculate the likelihood that it moves to the next compartments after some time delay. This feature is absent in most SIR models.

To obtain the model parameters, of which there are 22, the

team used numerical “Markov chain Monte Carlo” methods. These methods evaluate the likelihood that a given set of parameters is compatible with the input data, something that allows them to account for the inherent uncertainty in the model’s predictions. This compatibility check meant that their model did not yield a singular “best-fit” prediction but, instead, output a variety of compatible future scenarios (Fig. 2). To train the model, the team input data from the Illinois Department of Public Health, which included the daily number of COVID-caused deaths and the daily number of intensive-care beds occupied by COVID-positive patients. Using this reliable data added robustness to the predictions.

Using their model, the team made predictions about how the spread of COVID-19 would change under the implementation of various interventions, including a Stay-at-Home order, which confined people to their homes unless they were performing some essential activity. They predicted that exponential growth in the number of infections would occur if this order was not quickly implemented, even though, when they first started work on the model, Illinois had only 19 cases. Their prediction strongly suggested the need to issue such an order, which is what the Illinois Governor did after reading their results [10].

The team also predicted how infection numbers might change for various release scenarios of the Stay-at-Home order. Those predictions infamously had less success. A necessary condition for the predictions of SIR models to hold is that rules remain fixed, matching those that the model incorporates, and that individuals follow the rules. As reported in *The New York Times*, the predictions made by Goldenfeld, Maslov, and colleagues broke down when university students, who had tested positive for COVID-19, went to parties and classes, rather than quarantining in their dorms as they were meant to [9].

This apparent failure, however, is as important as the model’s successes: theories are of no use if their applicability limits are unknown. Most people understand that Solar System dynamics can be predicted a thousand years out, but that weather predictions are only reliable a week into the future. The validity timescale for pandemic modeling remains an open question. To answer that, we need improved virus-modeling infrastructure, such as a global data collection system for tracking virus infections and models that run in real time. That infrastructure could lead to better—global—forecasts for disease spreading,

the models for which will likely include elements from those of Goldenfeld, Maslov, and colleagues. These models will have to account for the ability of a population to adapt their behavior in response to the crisis and for the specific biological characteristics of the pathogen, something that will influence intervention measures. This new study is a step in the right direction, but there is much to do, and we need to do it before the next pandemic arrives.

**Susanna Manrubia:** National Center for Biotechnology (CSIC), Madrid, Spain

## REFERENCES

1. G. N. Wong *et al.*, “Modeling COVID-19 dynamics in Illinois under nonpharmaceutical interventions,” *Phys. Rev. X* **10**, 041033 (2020).
2. J. P. A. Ioannidis *et al.*, “Forecasting for COVID-19 has failed,” *Int. J. Forecast.* (2020).
3. I. Holmdahl and C. Buckee, “Wrong but useful — What Covid-19 epidemiologic models can and cannot tell us,” *N. Engl. J. Med.* **383**, 303 (2020).
4. W. O. Kermack and A. G. McKendrick, “A contribution to the mathematical theory of epidemics,” *Proc. R. Soc. London A* **115**, 700 (1927).
5. T. Carletti *et al.*, “COVID-19: The unreasonable effectiveness of simple models,” *Chaos, Solitons & Fractals: X* **5**, 100034 (2020).
6. W. C. Roda *et al.*, “Why is it difficult to accurately predict the COVID-19 epidemic?” *Infect. Dis. Model.* **5**, 271 (2020).
7. F. Cecconi *et al.*, “Predicting the future from the past: An old problem from a modern perspective,” *Am. J. Phys.* **80**, 1001 (2012).
8. M. Castro *et al.*, “The turning point and end of an expanding epidemic cannot be precisely forecast,” *Proc. Natl. Acad. Sci. U.S.A.* **117**, 26190 (2020).
9. K. Chang, A university had a great coronavirus plan, but students partied on, *New York Times*, Sept 10, 2020.
10. S. Chen, Physicists rise to the challenge, *APS News* **29**, No. 5 (2020).