IMPROVING PANDEMIC FORECASTS

ASSIMILATING OBSERVATIONS AND SIMULATIONS

At the height of the COVID-19 pandemic, an international team of mathematicians borrowed techniques from the geosciences to predict the complex and shifting dynamics of the virus' spread. The technique known as data assimilation combines numerical model data with fresh observational data to deliver more accurate forecasts. Validating the method in eight distinct countries, the team demonstrated the potential to reasonably and accurately predict the short-term impacts of various reopening measures on virus transmission. This method can provide critical information to policymakers to make informed decisions and design effective policies to mitigate the pandemic's impacts.

KEY MESSAGES

- ☑ An international team of mathematicians employed geosciences-derived data assimilation methods to enhance the prediction accuracy of traditional epidemiological models.
- ☑ The team introduced a variant of the susceptible-exposed-infected-recovered, SEIR, model, a standard mathematical approach to forecast infectious disease transmission. By combining the model information with new observational data for death and hospitalization numbers, they were able to produce realistic predictions of the pandemic's evolution, with quantified uncertainty estimates, for eight distinct countries.
- ☑ Mathematical data assimilation methods can play an important role in predicting a pandemic's evolution.
- ☑ The method can provide real-time, data-informed forecasts to policymakers to implement effective interventions to control a pandemic's spread and mitigate its impacts.

COMBINING THE BEST OF BOTH

As scientists scrambled to understand SARS-CoV-2, decision-makers and the public reasonably asked: How many more people are likely to die, and what effect will governmental containment policies have on the virus' spread? To answer these questions, scientists turn to numerical models, which describe the relationships between epidemic parameters or variables, and observational data, such as the number of individuals who died or were hospitalized due to the virus.

Neither numerical models nor observational data alone can accurately answer such questions, but by objectively combining the two, scientists can utilize the best parts of each while minimizing their respective flaws. The technique, known as data assimilation, is used routinely in geosciences, for example, in modern-day weather forecasting, arguably the best known and most successful application. With improved numerical models and data assimilation, today's 5-day weather forecast is as accurate as a 1-day forecast was in 1980.

In addition to improved forecast accuracy, data assimilation provides a robust assessment of the uncertainties of the output, a significant benefit over simpler, free-running models. In this regard, it offers predictions of the worst, best, and most likely situations.

AN INTERNATIONAL INITIATIVE

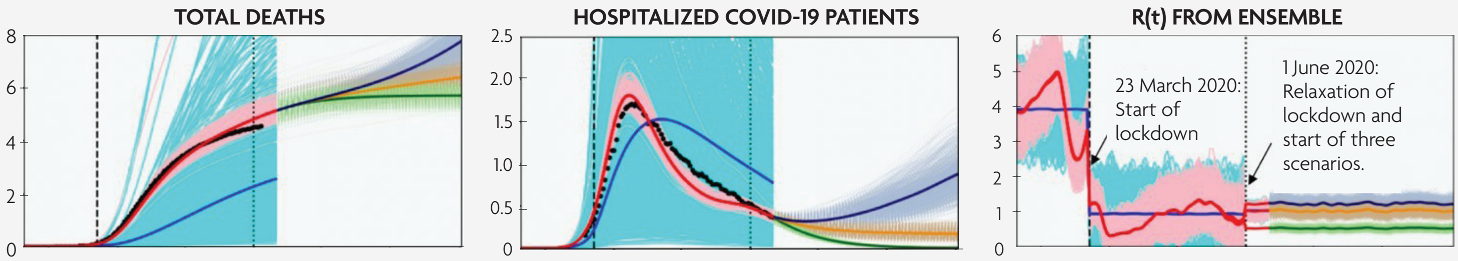
In the spring of 2020, an international team of data assimilation scientists representing eight countries diverted their attention from their work in Earth science fields to explore whether data assimilation tools could be applied to pandemic modeling in each of their countries. As the evolution of the epidemic varied widely between countries. Hence, the team began by identifying the factors that potentially contributed to virus transmission, such as geography, population density, social habits, healthcare systems, and, importantly, governmental policies and mitigation strategies, including lockdowns.

The team found that they could use data assimilation to explain reported deaths and hospitalizations using a classic metapopulation model, a type of spatial model that explores interactions of subpopulations across time and space. Their model is a version of the susceptible-exposed-infected-recovered (SEIR) compartment model, adapted to COVID-19 by including age-stratification and additional compartments for quarantined, hospitalized, and deceased.

Using this approach, they successfully represented the impact of the various interventions taken in each of their eight countries: Visualizing the rapid drop-off in person-to-person transmission at different points in each country's lockdown. Given the success of data assimilation to explain the reported deaths, they went on to develop predictions

CASE STUDY: EVOLUTION OF THE COVID-19 EPIDEMIC IN ENGLAND

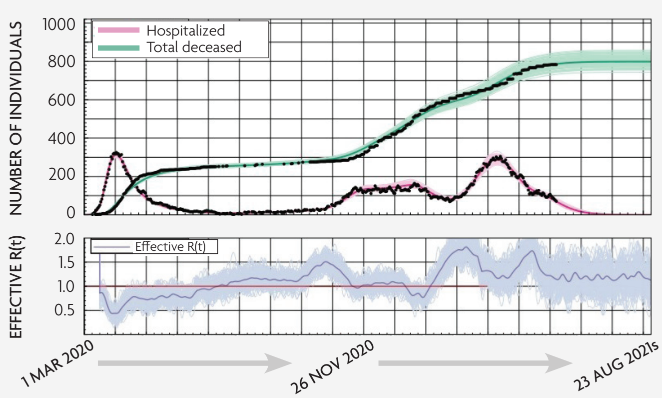
For England, simulations start on February 20th and assimilate data on a daily cycle beginning March 5th. Daily data for total deaths are assimilated from March 5th to May 29th, together with daily hospitalization numbers from March 20th to May 29th. Over the assimilation period, data are assimilated sequentially as they become available at a daily rate. The final assimilation solution over the whole data period is thus informed on all observations. Data after May 29th are not used in the assimilation, and the simulations run unconstrained. Three possible scenarios were considered beginning June 1st of 2020 when lockdown restrictions started to ease. Scenarios were defined in terms of the R naught value, which represents the number of people an infected person can be expected to transmit the virus to. They chose three values: 1) R equals 0.5, where cases reduce over time; 2) R equals 1.0, where case numbers remain steady over time; and 3) R equals 1.2, where case numbers increase over time. On June 1, approximately 45,000 deaths in England were attributed to COVID-19 in all settings. The team's analyses under the three scenarios projected that by the 1st of September 2020, total deaths would equal 57,000 00 ffl 1,900 (R=0.5), 63,600 ffl 2,700 (R=1), and 76,400 ffl 4,900 (R=1.2). These results highlight the potential to save tens of thousands of lives by using containment measures that reduce a significant amount of person-to-person contact.



The black dots depict reported values up to June 5th for deaths and up to June 12th for the number hospitalized. The bright blue lines indicate the initial estimates, and the red lines indicate the values after assimilation, with the bold line indicating the most likely value. After the 1st of June, three predictions are made based on three different R values: R=1.2 (navy), R=1 (yellow), and R=0.5 (green).

CASE STUDY: EFFECTS OF THE VACCINATION CAMPAIGN IN NORWAY

Using more sophisticated data assimilation methods, the team modeled the predicted effect of Norway's vaccination campaign. This new version of the model illustrates the power to forecast the pandemic's evolution over a longer period. In this case, the predictions' uncertainties reflect the uncertainties in both the simple model and the reported values for deaths, hospitalizations, and positive cases.



Top: hospitalized and total dead with observations (black dots). Bottom: Reproductive number R. In both panels shading shows the estimated uncertainties.

under different possible scenarios, such as reopening strategies and vaccination campaigns.

CONCLUSIONS

A popular framework in geosciences, numerical and mathematical data assimilation methods have proven extraordinarily versatile and provide a dynamical and statistically-sound way to combine multiple pandemic measurements with numerical models of its evolution. By quantifying model and observation uncertainties and including new epidemic data, the team provided reliable short-term predictions of pandemic indicators — deaths, infections, and hospitalizations — and estimates of the accuracy of these numbers. Across eight different countries and states, the team demonstrated the method's capability to detect the impact of each region's governmental interventions and assess their effects on the SARS-CoV-2 pandemic evolution.

REFERENCES

Carrassi, A., Bocquet, M., Bertino, L. and Evensen, G. 2018. Data assimilation in the Geosciences: An overview on methods, issues and perspectives. *WIREs Climate Change*, Vol. 9, 50 pp.

Evensen, G., Amezcua, J., Bocquet, M., Carrassi, A. Farchi, A., Fowler, A., et al. 2021. An international initiative of predicting the SARS-CoV-2 pandemic using ensemble data assimilation. *Foundations of Data Science*, Vol. 3, No. 3, pp. 413-477.

AUTHORS

Geir Evensen
NORCE and NERSC, Bergen, Norway

Javier Amezcua
University of Reading and NCEO, United Kingdom

Alberto Carrassi
University of Bologna, Italy

Alison Fowler
University of Reading and NCEO, United Kingdom