NEWS

PHOTO: GUY EDWARDES/NPL/MINDEN PICTURES

METEOROLOGY

# 30-day forecasts? Weather prediction has room to run

AI models suggest a 2-week limit linked to the "butterfly effect" may not be definite

PAUL VOOSEN

It's a truism almost as old as modern weather prediction: Any forecast beyond 2 weeks will fall apart because of the way tiny perturbations compound in the atmosphere. The 2-week limit, grounded in chaos theory and notions of the "butterfly effect" from the 1960s, has been handed down from generation to generation, says Peter Dueben, head of earth system modeling at the European Centre for Medium-Range Weather Forecasts, the world's leading forecaster. "It's basically a God-given rule."

But even the gods can be wrong.

Using an artificial intelligence (AI) weather model developed by Google, atmospheric scientists have found that forecasts of 1 month or more into the future might be possible. "We haven't found a limit to how far you can go out," says Trent Vonich, a doctoral student at the University of Washington (UW) who led the work, released late last month as a preprint on arXiv. "We ran out of memory first."

The result has caused a stir ever since Vonich and Gregory Hakim, his adviser, spoke this year at the annual meeting of the American Meteorological Society, says Amy McGovern, a computer scientist and meteorologist at the University of Oklahoma. Using powerful computer models, researchers have already pushed meaningful forecasts out to about 10 days, coming ever closer to the 2-week limit. Showing this limit can in principle be broken "means that AI will be able to do this someday, which is really exciting," she says.

The paper has caveats. For one thing, it does not make actual forecasts beyond 2 weeks, points out Tobias Selz, an atmospheric scientist at the Ludwig Maximilian University of Munich. So far, he says, the UW researchers have only tested their long-term forecasts with reconstructed snapshots of past weather. Moreover, as Selz and colleagues demonstrated in a 2023 study in *Geophysical Research Letters*, the AI models ignore the small-scale atmospheric processes—effects as small as a butterfly flapping—that are thought to snowball and drive the predictability limit. "I'm really reluctant to use these models to make statements about atmospheric predictability."

The notion of an intrinsic forecasting limit goes back to Edward Lorenz, the famed mathematician and meteorologist at the Massachusetts Institute of Technology (MIT). In a 1963 paper, he pointed out that even a small difference in rendering the initial state of the atmosphere or a similarly chaotic system would ultimately cause forecasts to diverge wildly. Then, in a 1969 paper, he suggested that, even if these initial conditions were known almost perfectly, the system would still have a predictability limit driven by the rapid error growth at small scales.

Lorenz did not, however, actually specify a 2-week limit. According to a recent historical study led by Bo-wen Shen, a mathematician at San Diego State University, Lorenz put forward a variety of possible limits but never settled on one. The 2-week figure came instead from MIT's Jule Charney and other pioneers who were gauging the capabilities of the world's first numerical weather models at about the same time. Shen also notes that Lorenz's 1969 modeling exercise relied on equations that were highly sensitive to their input data, which has caused Shen to wonder whether the butterfly effect is an artifact. Either way, there's no reason

Chaos theory in the 1960s suggested modeling errors from atmospheric processes as small as a butterfly flapping would snowball, limiting forecasts.

to think the 2-week limit is a rule, he says. "It's not a physically based law. It's an empirical assumption."

In their new work, Vonich and Hakim relied on Google's GraphCast, an AI model trained on 40 years of "reanalysis data"—high-resolution snapshots of the planet's weather based on observations and short-term model forecasts. The duo wanted to see how well GraphCast would work if they could somehow radically boost the accuracy of the initial conditions, the starting snapshot.

They did this by comparing the final state of the atmosphere from reanalysis data with GraphCast's forecasts. Shortcomings in a forecast could then be used to adjust the initial conditions, potentially bringing them closer to the atmosphere's true state. Operational weather models can also be tuned backward in this way, as subsequent observations are amassed. But the calculations needed to look back more than 12 hours in time quickly grow overwhelming. The structure of GraphCast, by contrast, makes such analyses easy to run thousands of times over and further back in time, allowing the model to home in on a near-perfect starting snapshot for the atmosphere, Hakim says. "Basically they were handing this to us on a silver platter."

With the trained initial conditions, GraphCast's accuracy for its 10-day forecast improved by 86% on average—"absolutely massive" in weather terms, Vonich says. Even more surprising, the model showed skill at predicting weather more than 33 days in the future. It was hard for Hakim to believe at first given what he had learned. "It's almost like a disconnect from reality," he says. "Yet here are the results. You can repeat this calculation."

The duo also looked at how the model was altering initial conditions, fearing it was doing something unrealistic. They found the model was making small adjustments to parameters such as temperature across large scales. It also seemed to be strengthening certain wind patterns that traditional weather models have been known to dampen. It just goes to show there are ways for AI, if it has enough data, to overcome the approximations and errors that get baked into traditional models, says Animashree Anandkumar, a computer scientist at the California Institute of Technology. "Once you throw everything away, you have a chance to rethink things."

Selz, however, says there is no evidence that the adjusted initial conditions are actually closer to the true reality seen in the atmosphere, he says. The adjustments could simply be creating a starting point that's ideal for GraphCast forecasts, in a sort of self-fulfilling prophecy. If that perfect version is perturbed at all, Selz suspects, the lengthened forecast window could close again. "And that's exactly what the butterfly effect tells you."

Regardless, the work is raising a lot of questions about the received wisdom, says Dueben, who has always been a little skeptical of how well the butterfly effect applies to weather. "It's probably too narrow a view to say it's only small scales moving upward and destroying prediction limits," he says. That view is echoed by James Doyle, a research meteorologist at the Naval Research Laboratory. Lorenz wasn't wrong that the small errors can proliferate, he says. "But maybe it's not as critical."

For now, a monthlong forecast is still aspirational, as it would require a far more refined view of the atmosphere than is currently possible with satellites and weather balloons. But if the new forecast horizon continues to beckon, Doyle says, now is not the time to draw back from weather research. "It tells us there is more to be gained by pushing the models out further." □