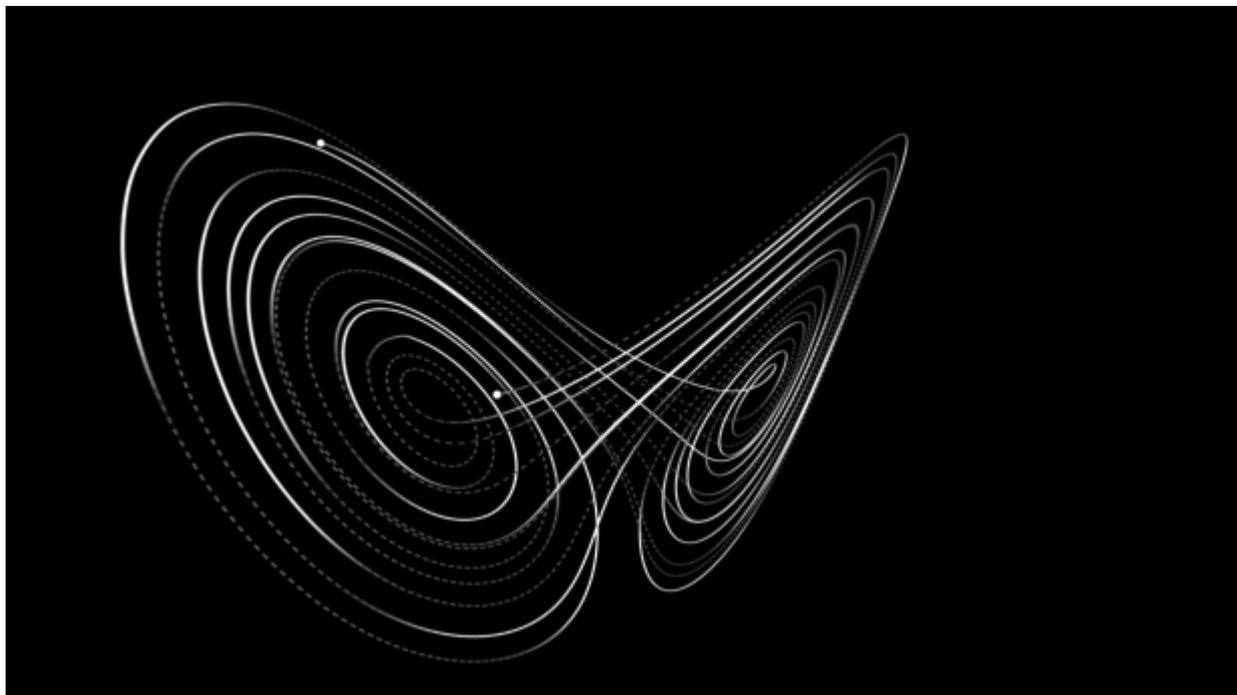




A Twisted Path to Equation-Free Prediction

Complex natural systems defy standard mathematical analysis, so one ecologist is throwing out the equations.

By Gabriel Popkin



strange attractor helps researchers predict what will happen in a chaotic system.

Sometimes [ecological data](#) just don't make sense. The sockeye salmon that spawn in British Columbia's Fraser River offer a prime example. Scientists have tracked the fishery there since 1948, through numerous upswings and downswings. At first, population numbers seemed inversely correlated with ocean temperatures: The northern Pacific Ocean surface warms and then cools again every few decades, and in the early years of tracking, fish numbers seemed to rise when sea surface temperature fell. To biologists this seemed reasonable, since salmon thrive in cold waters. Represented as an equation, the population-temperature relationship also gave fishery managers a basis for setting catch limits so the salmon population did not crash.

But in the mid-1970s something strange happened: Ocean temperatures and fish numbers went out of sync. The tight correlation that scientists thought they had found between the two variables now seemed illusory, and the salmon population appeared to fluctuate randomly.

Trying to manage a major fishery with such a primitive understanding of its biology seems like folly

to [George Sugihara](#), an ecologist at the Scripps Institution of Oceanography in San Diego. But he and his colleagues now think they have solved the mystery of the Fraser River salmon. Their crucial insight? Throw out the equations.



George Sugihara, an ecologist at the Scripps Institution of Oceanography, believes that it's foolish to model many natural systems with equations.

Sugihara's team has developed an approach based on chaos theory that they call "empirical dynamic modeling," which makes no assumptions about salmon biology and uses only raw data as input. In designing it, the scientists found that sea surface temperature can in fact help predict population fluctuations, even though the two are not correlated in a simple way. Empirical dynamic modeling, Sugihara said, can reveal hidden causal relationships that lurk in the complex systems that abound in nature.

Sugihara and his colleagues are now putting their insight to use. Earlier this year [they reported in the *Proceedings of the National Academy of Sciences*](#) (PNAS) that their method predicted the 2014 Fraser River salmon run more precisely than any other method. Sugihara's technique predicted a run of between 4.5 million and 9.1 million fish, while the Pacific Salmon Commission's models predicted anywhere from 6.9 million to 20 million fish — a forecast so broad as to be of little benefit to, for instance, a fisher wanting to know how many boats to deploy in the coming season. The final count was around 8.8 million.

This success built on an earlier result Sugihara and his colleagues had achieved with Pacific sardines, and they're working with scientists at the National Oceanographic and Atmospheric Administration (NOAA) to apply the methods to Gulf and Atlantic menhaden. Leading ecologists

hope Sugihara's methods can provide the field with some much-needed predictive power, and not just for marine fisheries but for many other ecosystems. [Don DeAngelis](#), an ecologist with the U.S. Geological Survey in Miami, calls it "a huge theoretical breakthrough."

Sugihara and others are now starting to apply his methods not just in ecology but in finance, neuroscience and even genetics. These fields all involve complex, constantly changing phenomena that are difficult or impossible to predict using the equation-based models that have dominated science for the past 300 years. For such systems, DeAngelis said, empirical dynamic modeling "may very well be the future."

A New Set of Coordinates

The roots of empirical dynamic modeling go back more than 30 years. In the late 1970s, the Dutch mathematician Floris Takens was studying chaos theory, which had begun to emerge in the 1960s as scientists recognized that many of nature's complex phenomena seem to defy prediction. In chaotic systems, small perturbations can have large and seemingly unpredictable effects, as in the archetypical example of a butterfly's flapping wings influencing the weather thousands of miles away.

It is not the world that is mysterious. Rather it is the way we view it that makes it mysterious.

Takens helped find order in the chaos. Along with the physicist David Ruelle, he developed the notion of a "strange attractor" — a set of points in a coordinate system made of the variables that influence a system, around which the system's state, plotted over time, swirls like a ball of yarn.

In many natural systems, however, the number of relevant variables that make up the coordinate system is immense. The factors that determine the weather in a certain place at a certain time are almost limitless, and some of these can be very hard to measure — the air pressure three miles above the North Pole, for example.

But let's say you could consistently and accurately measure one variable, such as the temperature in New York City. Takens found a way to use present and past measurements of that one variable to capture all the information in the system. The method involves creating an alternate coordinate system from those past measurements; in other words, one coordinate axis might be the temperature in Times Square today, a second axis might be the temperature yesterday, a third the temperature two days ago, and so on. Takens showed that the full state of a chaotic system can, in theory at least, be embedded in a time series of a single variable. He published his "embedding theorem" in 1981.

The theorem "caused a big hullabaloo," said [Timothy Sauer](#), a mathematician at George Mason University in Fairfax, Va., who has extended the original theorem so it can be applied more generally.

The next step was for people to use it in the real world, but the messiness of nature had a way of impinging on the purity of Takens' math. Despite the fact that weather provided much of the initial impetus for chaos theory, it rebuffed efforts at prediction, because too many constantly changing factors are involved, and no one variable can truly capture them all. Sauer said that Takens' theorem can be most effectively applied to systems in which the number of influencing factors is relatively

small.

Sugihara learned about Takens' theorem as a Princeton graduate student working with Robert May, a physicist by training who switched to ecology in the early 1970s. May specialized in simple and elegant theoretical studies, including one proving that the population of even a single species could fluctuate chaotically. Sugihara became interested in seeing if he could build on May's advances using real-world data. In 1986, a few years after earning his doctorate, Sugihara moved to Scripps to get his hands on plankton data that a researcher there had collected in the 1920s and '30s. "It's an amazing data set," Sugihara said. "I knew there was some way to get good information out of it."

Based on the plankton data as well as work on measles and chicken-pox cases by other researchers, Sugihara and May published a paper in *Nature* in 1990 showing how Takens' theorem could help make short-term predictions of some nonlinear systems. The essence of the method involves identifying points in a system's attractor graph that are close to the point representing the system's present state. For one or two time steps, one can then predict that the system will evolve similarly to how it did in the past. The paper has since been cited more than 1,000 times by scientists all over the disciplinary map. The paper also prompted Sugihara to make a mid-career foray into finance, as firms were very interested in forecasting stock prices using methods similar to those he had applied in ecology.

In 2002, Sugihara returned to science. He had unfinished business: convincing the world that ecosystems, though complex and chaotic, could be predicted, and that managers could use those predictions to do their jobs better. "I feel like I have a mission," he said, "to get people to understand how this all works — to begin to embrace natural systems as they are as opposed to as we hope they would be."

A Thirst for Data

Ecological modeling began almost 100 years ago, and from the start it was influenced heavily by physics and engineering, which had used differential equations to describe dynamic systems for the previous 200 years. The most commonly used fishery model, for example, is the Ricker model, developed by the Canadian biologist William Ricker in the 1950s to predict the number of new adults that an existing generation of fish is likely to produce the following year. Ricker's original equation included just two parameters: the reproductive rate of a given fish, and the number of fish the environment can sustain, known as the "carrying capacity."

Fishery managers still rely heavily on the Ricker model, along with variants that include factors like temperature, to estimate a "maximum sustainable yield" that fishers can take without causing fish stocks to crash. Such estimates are naive, Sugihara said, because they assume fish population is correlated to environmental factors in simple and static ways. "It really is almost a kind of hubris to write down an equation that guesses that temperature ought to enter in a specific way."

Environmental factors — the climate, ocean circulation, human impacts — are always changing, but parameterized models such as these are stuck in time and cannot adapt to those changes, much less incorporate them to become more accurate. "They will not necessarily improve as we get more data," Sugihara said.

Empirical dynamic modeling, by contrast, seamlessly incorporates new data and is always improving. Takens' theorem works best when there are enough data points to make a dense attractor, making it easier to find times when a system's present state is close to a past one. Any new data points will help users to see where a system is going to go next. "It's allowing the data to say what the relationships are," Sugihara said. And it succeeds, he said, "where the rubber hits the road," namely, on how well it can predict the future, and not just on how well scientists can make a

curve fit the data after the fact.

Sugihara's work is not armchair mathematics: Many fishery scientists want better forecasts, and researchers from both NOAA and Canada's equivalent agency, the Department of Fisheries and Oceans (DFO), have co-authored papers with Sugihara and his students. But so far no fishery commission has actually incorporated the methods into its management practices. One hang-up, said Jon Schnute, a retired analyst formerly at DFO, is that so far only Sugihara and his colleagues have had access to the underlying algorithms, meaning that fishery biologists must send their data to Scripps and then wait for a forecast. By contrast, all fishery biologists can use software implementing the Ricker model. Empirical dynamic modeling "hasn't reached that point of maturity," Schnute said.

That is changing. Sugihara's software is now available for researchers to use, and his students are leading workshops to teach them how to do so. DeAngelis, a lifelong user of parameterized equations, said he hopes to use Sugihara's methods in his own work predicting the population dynamics in fish populations in the Everglades.

The End of Equations

DeAngelis also goes further, writing in [a comment accompanying Sugihara's team's 2015 PNAS paper](#) that empirical dynamical modeling may be part of a broader shift away from the dominance that equations have long exerted over science. Many commentators, including DeAngelis, noted that equations have not yielded the same success in ecology that they have in the physical sciences, suggesting a new approach is needed.

Sugihara agrees. Static equilibrium equations may be useful for building a bridge, he said, but it's time to abandon the search for equilibrium in the complex, nonlinear systems that nature produces. Seductively simple correlations may appear for a period of time, he observed, but in a chaotic system such correlations do not provide true insight. "It is not the world that is mysterious," he said. "Rather, it is the way we view it that makes it mysterious."

Fellow ecologists are excited by the new method but mindful of the challenges Sugihara faces. Paucity of data remains one of the big ones. While fields like medicine and neuroscience are now spewing out huge data sets more quickly than scientists can process them, ecology is still [stumbling toward its big-data revolution](#).

A more difficult question, Sauer said, may be that of stationarity — whether the meaning of a measurement stays the same from one day or year or decade to the next. Stationarity is one of the hallmarks of laboratory science: A protein molecule or yeast cell today is the same sort of thing that it was 100 years ago. But it is less clear whether a tally of Fraser River sockeye salmon in 2015 has the same meaning as a count of the same salmon in 1950. The DFO changed how it defines salmon stocks during that time period, and even the fish themselves may have evolved.

DeAngelis adds that empirical dynamic modeling has another limitation: The method can make only short-term predictions. This goes back to the fundamental problem with chaotic systems: Two systems whose initial conditions vary only the tiniest bit will diverge over time onto totally different trajectories. In practical terms this means that even if the method does a good job of forecasting next year's salmon population, it can't say much about that population several years from now.

For these reasons and others, Sugihara is starting to push his methods beyond ecology. A few years ago Sugihara got an email from Gerald Pao, a molecular biologist in the lab of [Inder Verma](#) at the Salk Institute for Biological Studies in San Diego, just down the road from Scripps. Pao was

convinced that Sugihara's methods could be used to interpret gene expression data. Sugihara was skeptical, but once he realized how rich Pao's data were, with coordinated time series based on hourly measures of the expression of all 25,000 or so genes in human chromosomes, he realized he was wrong. Sugihara, Pao and Verma got started on yeast and mouse models and hope to publish a paper soon that will show how networks of genes can be causally linked even when their expression patterns aren't correlated.

Ideas similar to empirical dynamic modeling are also showing up in neuroscience. Neuroscientists would love to be able to predict the onset of crippling conditions like epileptic seizures, and some are modeling firing patterns of neuron networks using Takens' theorem. Sauer said neuroscientists may be further along than ecologists in bringing the theorem from the realm of theory into practice. But, he said, "the real killer app is not here yet."

Sugihara agrees with this assessment. "Takens' theorem is an amazing thing," he said, "and the potential applications remarkably have not been fully realized." He added, "I think that's just beginning to change now. ... I think we're beginning to overcome the activation energy barrier that it takes to understand this stuff."

This article was reprinted on Wired.com.