Pacific Sardines, NOAA

A PRIMER ON EMPIRICAL DYNAMIC MODELING FOR FISHERIES

The Lenfest Ocean Program has supported several projects to improve predictive power in fisheries science, using an innovative technique called empirical dynamic modeling (EDM). This primer is intended to provide a readily understandable explanation of the theory behind EDM, its basic operation, and why it may be able to improve population forecasting in fisheries.

Mathematical models are critical to fisheries management. They give managers an overall picture of a fish population, helping them set catch limits and make other rules intended to sustain the ecosystem.

But virtually every model used in fisheries management makes a fundamental assumption: that the inner workings of these populations, and the external factors that affect them, can be approximated with simple equations. For example, many models incorporate an equation that describes how the abundance of adult fish affects "recruitment"—the number of young fish added to the catchable population. Such equations are useful because they put data into a mathematical framework to create a consistent, coherent picture of what's happening in the water.

However, because these equations attempt to define the relationship between variables in advance, this approach runs the risk of detecting only the patterns that the model builders have already identified. Empirical dynamic modeling (EDM), by contrast, starts with no equations and no assumptions, relying primarily on observational data to detect the complicated, shifting relationships that are common in nature. It has been shown to improve forecasting in several studies, and scientists and managers are now working to translate that success into more accurate predictions of recruitment and other data that are useful in management.

SOME BACKGROUND ON CONVENTIONAL FISHERIES MODELING

Figures 1 and 2 depict two conventional stock-recruitment models. These models predict the number of young fish that reach catchable age ("recruitment") based on the size of the adult population ("stock"). Each model is a single equation, based on a general observation: small populations grow slowly, somewhat larger populations grow more quickly, and the largest populations do not grow at all because they have reached an upper limit.

Figure 1

An example of a Ricker "stock-recruit" relationship, showing how an increase in the abundance of adult fish leads to changes in recruitment.

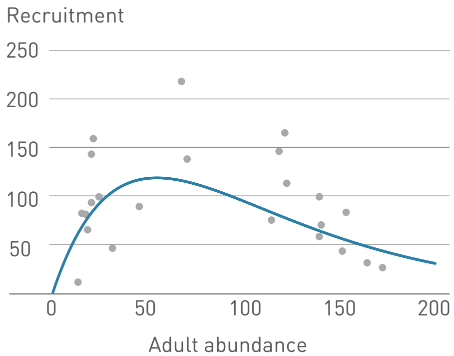
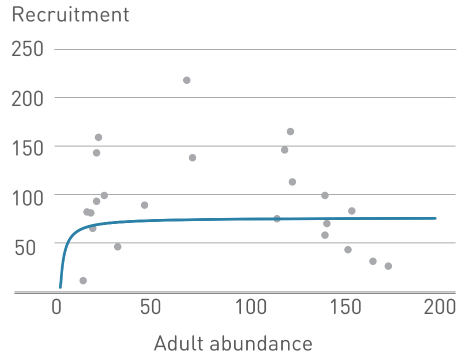


Figure 2

A Beverton-Holt stock-recruitment curve. Unlike the Ricker model, this model posits that recruitment continues to grow with adult abundance.



When modeling a specific population, scientists look for the best fit to the data for that population by adjusting parameters—numbers intended to represent a characteristic of the population (or of the larger system it is part of).

For example, in the "Ricker model" above, the "steepness" parameter represents the initial rate of increase, and the "peak recruitment" parameter sets a ceiling on growth. Changing the parameters changes the slope and height of the curve, but the general shape stays the same because it is dictated by the equation.

Once they have determined the best-fitting curve, scientists can use it to forecast population size in future years. In the simplest version, they would just enter this year's adult abundance, and the model would return a prediction for recruitment.

For the formal process of fisheries "stock assessment," scientists build larger models that incorporate more detail. For example, they include the effect of adult mortality and the additive effect of several years of recruitment on overall abundance and the distribution of ages in the population. Some stock assessments use non-modeling methods to estimate recruitment, such as a long-term average. But all stock assessment models retain the fundamental idea that equations can reasonably describe what happens in nature.

WHEN MODEL ASSUMPTIONS ARE WRONG

The equations used in conventional modeling are powerful because they integrate a wealth of data to explain and predict how a system will behave. But they also contain assumptions about that behavior. These assumptions are, by necessity, simplifications that are incomplete and never entirely correct.

For example, what if recruitment does not decline at higher levels of abundance? The Ricker model will still predict a decline, regardless of what the data say. In this situation, one solution is to try another model. For example, the Beverton-Holt model, shown in Figure 2, assumes recruitment does not decline at high abundance.

With both the Ricker and Beverton-Holt models, there is quite a bit of unexplained "scatter." To some extent, this is inevitable because no measurement is perfect, and nature is variable. But it is also possible that the model does not adequately represent reality. There are at least two possible explanations for this mismatch.

First, there could be other variables affecting recruitment, such as temperature and predators that eat fish eggs. A model based only on abundance cannot capture such variables.

Second, there could be interactions among variables, a situation known as "state dependence" or "nonlinearity." For example, temperature might have a large effect on recruitment when fish abundance is low but little to no effect when abundance is high.

Conventional models could in theory account for both issues. But to do so, they would need to use far more complicated equations, and these equations would have to contain the correct variables and make the correct assumptions about how the variables interact. This may be possible, but in practice it is extraordinarily difficult.

Modelers are well aware that they cannot account for all of nature's complexity. They often acknowledge this fact with the insightful cliché, "all models are wrong, but some are useful." They address the problem by continually improving their models, and by consulting multiple models in hopes that different equations will converge on similar predictions.

By contrast, EDM addresses the problem by doing away with model equations altogether.

FINDING HIDDEN PATTERNS USING A SINGLE VARIABLE

Instead, EDM makes predictions based primarily on patterns in the data. It does this using the idea that the rules governing natural systems can be described with a geometric shape called an "attractor."

For example, Figure 3 shows three variables and how they change together over time. The resulting attractor reveals that the relationship among these variables is state dependent. Specifically, when the variable X is low (on the left side of the graph X), an increase in X means a decrease in Z. But the opposite is true on the right side: an increase in X means an increase in Z. Thus, the effect of X on Z is not consistent; it depends on the state of X.

Figure 3 comes from a model used in physics, in which the three axes represent properties of a fluid. But one could imagine a biological version, in which the X-axis could be fish abundance, the Y-axis temperature, and the Z-axis recruitment.

State dependence is common in nature. For example, in one population of sockeye salmon, higher temperature appears to reduce recruitment when the adult population is low, but to have no effect when population is high.

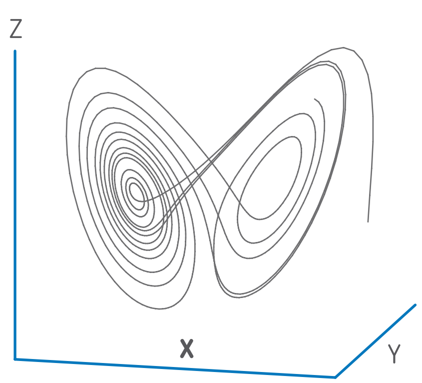
To create a conventional model of such a system, one would need to know in advance what variables are influencing recruitment and how they fit together mathematically.

With EDM, the attractor can be reconstructed directly from observational data, with no advance knowledge of how the system works, and even without any data on some of the key factors influencing recruitment. This reconstruction is possible thanks to a theorem of the Dutch mathematician Floris Takens.

Takens' theorem shows how to recover information about the entire system from a series of observations on a single variable. This is possible because when one variable influences another, information about the first variable is encoded in the second. This theorem is at the heart of the process of forecasting with EDM.

Figure 3

This example from fluid dynamics shows how three variables change together over time. This "butterfly attractor" shows a system that is "state dependent" because the effect one variable has on another changes as the variables change.



The first step in the process is to use one variable to create multiple dimensions. For example, you could take the value of recruitment at three different points in time and make those your three dimensions.

Let's say your dataset ran from 1981 to 2000. Your first point in the graph would be given by recruitment in 1981 on the X-axis, recruitment in 1982 on the Y-axis, and recruitment in 1983 on the Z-axis. The next point would be 1982 on the X-axis, 1983 on the Y-axis, and 1984 on the Z-axis, and so on.

By continuing this process through 2000, you would plot out 18 oints in a three-dimmensional space, even though your data have only one dimension. By connecting these points together chronologically, you would generate a shape. If you have enough good data, you can use this shape to reconstruct the underlying attractor.

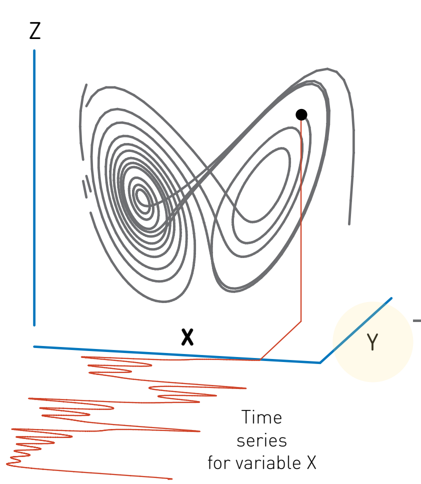
Box 1, in panels 2 and 3, shows an example of how this reconstruction works. Panel 1 shows the underlying attractor, which has three variables. Panel two shows one variable, X, shifted in time, as described above, to generate the reconstructed attractor in panel 3. The result is not identical to the original attractor, but it generates a shape that offers the same predictive power. (This video shows the process in motion.)

Box 1: Takens' Theorem

In mathematics, an "attractor" is a geometric shape that constrains how a system changes over time. Panel 1 below shows a plot of three variables over time from in a model of atmospheric convection. Plotting them together results in a "butterfly" attractor. Takens' theorem says that the signature of the other variables are often "embeded" in the time series of a single variable, an that variable can be used to reconstruct the dynamics of the entire system: the attractor.

1

The graph below shows the path traced by the variables X, Y, and Z as they change over time. Below that, in red, is a projection onto the X-axis, resulting in a plot of X against time.



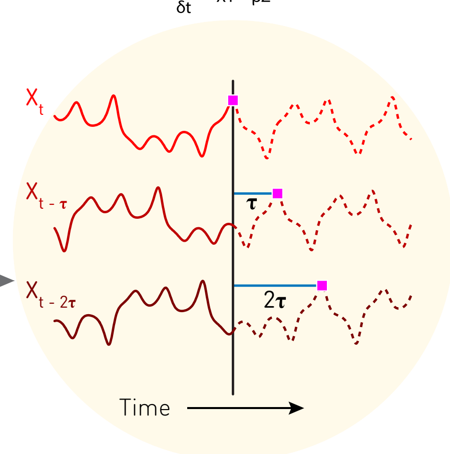
2

This graph shows the plot of X against time from the bottom of the previous graph. To reconstruct the attractor using only variable X, this time series is shifted by the value tau (τ), and again by 2τ to produce the three plots below.

$$\frac{\delta X}{\delta t} = -\sigma Y + \sigma X$$

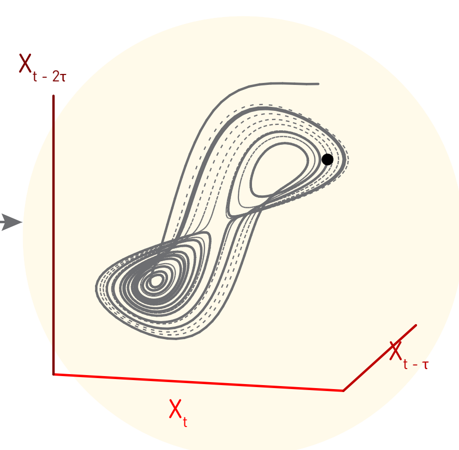
$$\frac{\delta Y}{\delta t} = -XZ + \rho X - Y$$

$$\frac{\delta Z}{\delta t} = XY - \beta Z$$



3

Using these three time series as the three dimensions in a new state space generates the following graph—a reconstruction of the attractor using only the Y variable.



To make a forecast, one would look at the system's current position on the reconstructed attractor and ask what points from past data are nearest. These points are then projected forward by one time step. The average of these projected points is the forecast. (The most straightforward method to get this average is called "simplex." It uses three "nearest-neighbor" points and is thoroughly explained here).

A few studies have applied Takens' theorem to fisheries. For example, in a 2013 study, George Sugihara's laboratory at the Scripps Institution of Oceanography successfully forecasted the abundance of Pacific sardine.

MAKING EDM USABLE FOR MANAGEMENT

Although such research is promising, EDM is not yet being used in fisheries management. This is partly because of its novelty, and partly because it is not always clear how to integrate its results into the current management system.

For example, conventional models are designed to readily produce "reference points"—benchmark values of abundance and fishing intensity that serve as triggers for management action. A common basis for reference points is the concept of "maximum sustainable yield," which makes the commonsense assumption that if a fish population were harvested at an ideal constant rate, it would produce the highest long-term yield. EDM does not assume this, so there is no simple way to use it to generate reference points.

A first step to incorporate EDM into management would be to use its predictions of recruitment in a "projection analysis," which uses stock assessment models to predict the effects of various management options. The Sugihara laboratory is now working to provide the technical tools to do this analysis for Atlantic and Gulf menhaden. These tools include software code and guidelines for applying EDM to new data. The Canadian government also plans to examine the possibility of incorporating EDM into the management of Fraser River sockeye salmon.

Another project, led by fisheries ecologist Steve Munch, is using EDM to better forecast recruitment for hundreds of species. Munch is also asking whether EDM can be used to measure how harvesting a prey species affects its predators, and to generate harvest strategies that optimize long-term yield.

OTHER APPLICATIONS OF EDM

EDM has been used in a variety of other intriguing studies that show the wide applicability of this method to complex, nonlinear systems that have proven difficult to understand with conventional models:

- **Understanding the impact of fishing**. A 2016 study led by ecologist Emily Klein (supported in part by a Lenfest Ocean Program grant to ecologist Adrian Jordaan) examined data from Canada's Bay of Fundy in the 1800s and 1900s and found evidence that intense fishing pressure may increase variability and uncertainty in fish population dynamics.
- **Finding causal connections**. A tool derived from EDM, known as convergent cross mapping, can provide evidence of a causal connection between two variables, even when they are not correlated. A 2016 study from the Sugihara laboratory applied this technique to show that tropical flu outbreaks in humans were linked to the interplay of humidity and temperature.
- **Making non-fishery predictions**: EDM has also proven promising for the prediction of measles outbreaks, Canadian lynx populations, and "red tides," which are blooms of ocean algae that can be toxic to marine life and people.
- **Applications in biomedicine and finance**: EDM has been used experimentally for early detection of Alzheimer's, for proprietary trading at Deutsche Bank to manage $2 billion of daily notional risk from 1996 to 2002, as a diagnostic for cardiac disease in infants, and to study coupling between brain regions. In addition, the predictability of heart rhythms using EDM may be an indicator of human health.

Research is ongoing to connect empirical dynamic modeling to real-world applications. It will take time to make this connection, but there is reason to believe it can result in better tools, in fisheries management and many other areas.