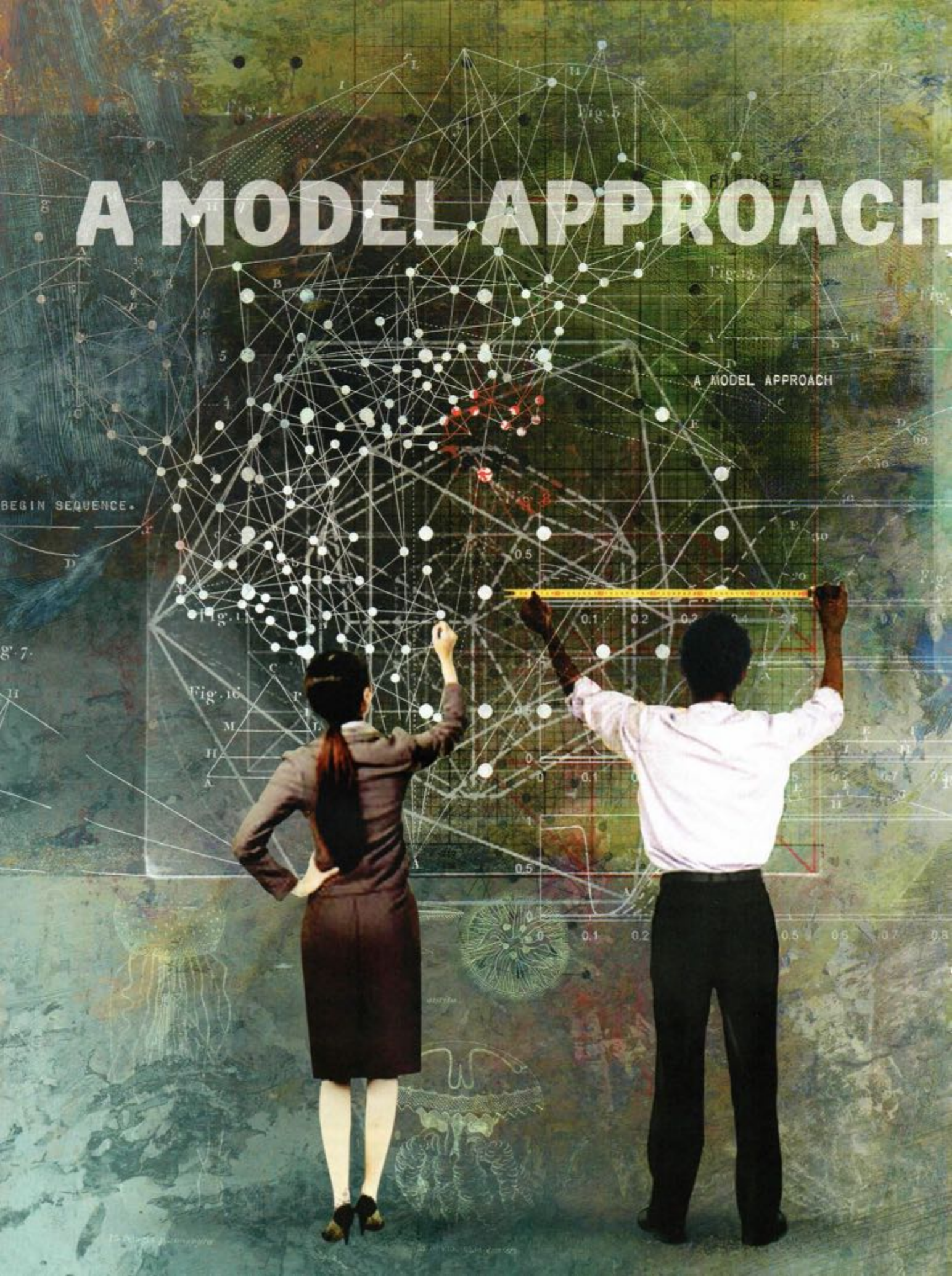


# A MODEL APPROACH





## COMPUTATIONAL TOOLS HELP SCIENTISTS SOLVE PROBLEMS AND WRANGLE BIG DATA.

Written by KOREN WETMORE

Illustration by DUNG HOANG

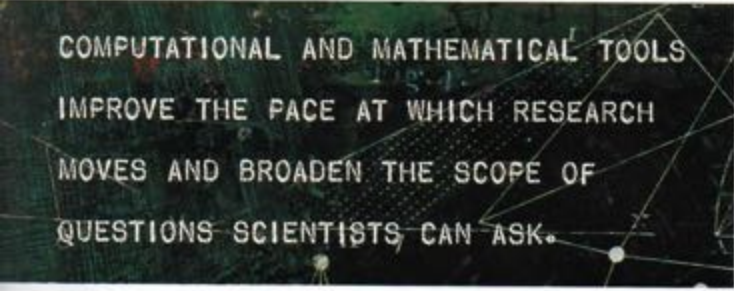
**IN THIS INTERNET AGE**, most of us can relate to the concept of information overload. But that's nothing compared to the data volume that bombards research scientists.

Faced with 60 million DNA sequences or a record containing 750,000 data points, the last thing a biologist wants to do is perform an analysis using spreadsheets. Instead, today's researcher reaches for computational and mathematical tools, producing models and scripts to process information at lightning speed.

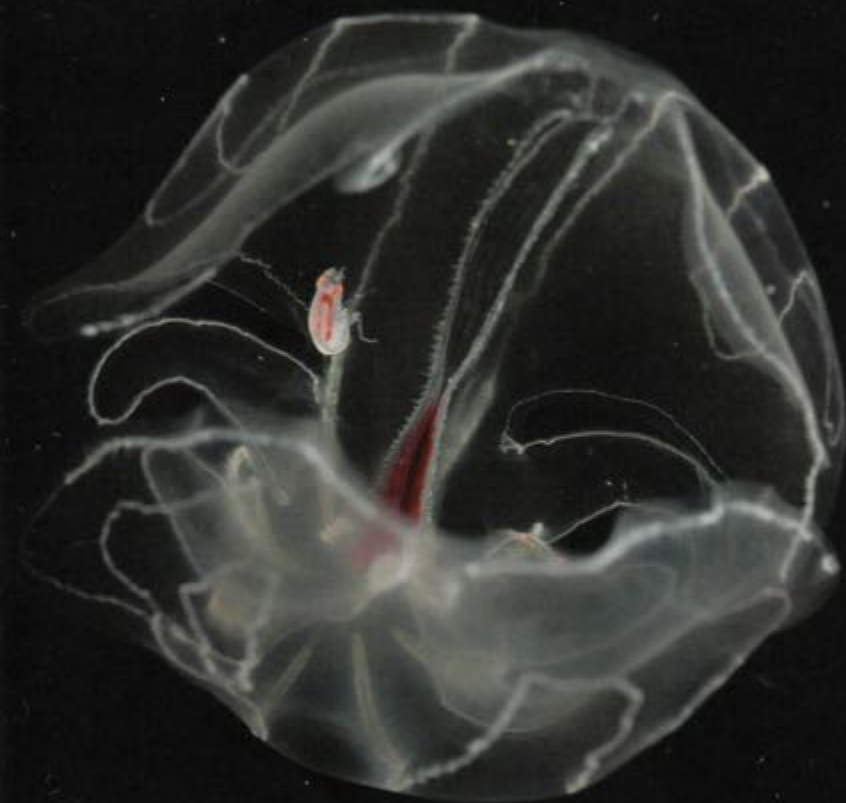
Such tools not only improve the pace at which research moves but also broaden the scope of questions scientists can ask.

"If we're looking for new photoproteins or fluorescent proteins involved in how a jellyfish makes light, we can find those easily now by sequencing an entire transcriptome of an organism," says marine biologist Steve Haddock '88. "Instead of saying, 'We think it's going to look like this, and we're going to try and fish that gene out,' we say, 'What are all the genes this animal is expressing right now?' It really allows you to think more about your science instead of your analysis."

Haddock often applies computational skills in his work at the Monterey Bay Aquarium Research Institute, where he studies how jellyfish and other deep-sea gelatinous creatures interact with light. His research involves describing species, building phylogenetic trees and working with the DNA sequences and chemicals the organisms use to produce light. "There's obviously a lot of computation," he says. "So, we'll have a program that will do each of the steps: cleaning up the sequences, pulling out those that look like they should be used



COMPUTATIONAL AND MATHEMATICAL TOOLS  
IMPROVE THE PACE AT WHICH RESEARCH  
MOVES AND BROADEN THE SCOPE OF  
QUESTIONS SCIENTISTS CAN ASK.



STEVE HADDOCK '88 USED  
COMPUTATIONAL TOOLS TO  
FIND THE PHOTOPROTEIN  
THAT HELPS THIS DEEP-SEA  
STENOPHORE MAKE LIGHT.

to assemble, assembling the sequences and then doing diagnostics on the assembly.”

By translating their questions into mathematical language and then encoding it into a computer program, researchers can build models to predict outcomes or identify potential causes of observed phenomena.

The Centers for Disease Control and Prevention, for example, uses mathematical models to determine how a disease might spread in a given population and how to potentially stop or minimize its spread. Cancer researchers use models to explore potential ways to treat or prevent various cancers.

In fact, according to a study published last August in the *Annals of Internal Medicine*, a risk-prediction model developed at the University of Liverpool’s Cancer Research Center proved more accurate at determining a person’s lung cancer risk than assessments based upon family history or years of smoking.

Harvey Mudd College mathematics Professor Lisette de Pillis, an expert in the field of tumor modeling, uses mathematical models to define variables involved in tumor growth rates and to identify the effects of the immune system and drug therapies in treating cancer patients.

One day, she was asked to work with a data set from a breast cancer patient whose treatment response puzzled her physicians. “They would measure her tumor before and after each treatment, and what they expected to see was treatment, tumor shrinks, treatment again and tumor shrinks some

more,” says de Pillis. “But what they saw was treatment, tumor shrinks, treatment again and tumor grows. The overall trend was that her tumor would shrink, but why would it grow immediately after treatment?”

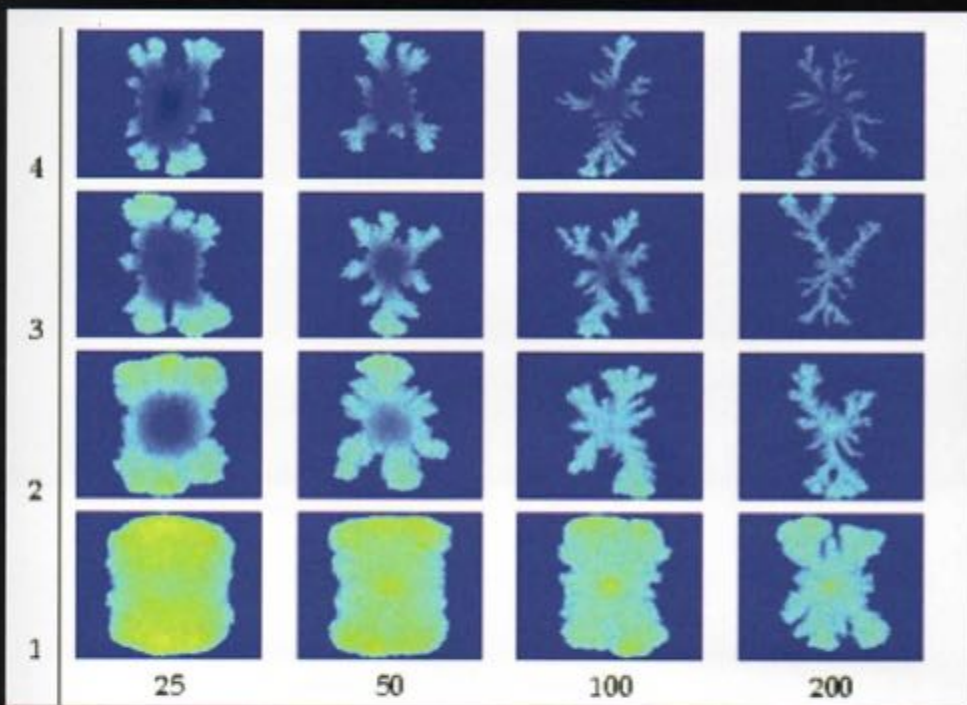
With her collaborators, she created a model that offered a possible explanation. By accounting for the action of the immune system, it showed that the chemotherapy might not only be damaging the tumor but also harming the immune system that kept the tumor in check so that, in some cases, the tumor was allowed to grow. “That model did not say it definitely is the immune system; it said this is one possible mechanism. To confirm that, we would have to develop some assays and check our results in the lab,” she says.

A core group member of the newly formed Brain Tumor Ecology Collaborative—a joint, three-year project by Washington University, Columbia University, U.C. San Diego and HMC—de Pillis will be at the forefront of research to better understand the behavior of pediatric brain and spine tumors called gliomas.

Her role will include using mathematical models to test hypotheses about what natural elements in the body might lead to the growth or the control of these tumors.

“Right now, we are working to determine distinguishing features of these cancers, such as the number of mitochondria per cell or the expression of the NF1 gene, that may influence tumor formation. The collaborative wants to use models to look at the physical space in which the cells live, to see how





The multiple images at left are models of tumor shapes being studied by mathematics Professor Lisette de Pillis. The more "gluttonous" a tumor type is, the more "branchy" (papillary) the resulting tumor shape, as shown in the upper-right corner.

they interact, to discover the ecosystem networks and to develop and evaluate ways to better treat patients who have this cancer."

Even the drugs, devices and therapies developed to treat disease rely on mathematical models to help prove their safety, efficacy and cost-effectiveness.

"It's called pharmacoeconomics," says Anita (McMorrow) Brogan '97, head of Decision Analytic Modeling in the Health Economics group at RTI Health Solutions, an independent research organization whose clients include pharmaceutical and biotech companies. "We use different types of models depending on whether we are looking at disease transmission in a population or disease progression in an individual. Once we've modeled the disease, the economics come in as we look at how a new intervention—a drug or a device—might impact the course of the disease, patient outcomes and costs."

Those projections are then compared with the outcomes associated with existing therapies. "A new drug might cost more up front but still save money because it helps prevent costly events from happening in the future. For example, if you have an expensive drug that cures hepatitis C, it may save money in the long-run because you avoid cases of liver disease and liver transplant," Brogan says.

One of the joys of Brogan's career was working on models that showed how darunavir—a new drug for HIV patients—was a cost-effective treatment option in Canada. "Our analysis helped the entire country of Canada to get access to it. The

provincial governments added it to their drug lists, so patients can get it and the cost is covered."

While the advantages of using computational and mathematical tools may be apparent, their use is not actually widespread.

"We need better bridges of communication and interaction between people in the clinics and the mathematicians," de Pillis says. "When I first went to talk to the [Brain Tumor Ecology Collaborative] members in St. Louis, they assumed a mathematical model was simply a statistical analysis of data. So when I explained what we actually mean when we say 'mathematical model,' and presented the range of mathematical models we have developed, the doctors were excited about the idea of modeling a growing tumor as a collection of heterogeneous organisms living in their own ecological system."

When Haddock spotted fellow researchers spending hours inputting data and formulas into spreadsheets, he decided to co-author a book, *Practical Computing For Biologists*, to introduce them to their computational alternatives.

"People hear 'computational biology' and automatically think of DNA and genes, yet the same skills are just as useful for someone who is an ecologist or physiologist," he says. "You have large amounts of data to process, and being able to do that repeatedly in a way that you can document for someone else is essential to any biological science now."