Jun Zhang and Jin-Qiang Zhong, New York University

Monte Carlo Methods in Climate Science

John C. Baez and David Tweed

They look placid lapping against the beach on a calm day, but the oceans are actually quite dynamic. The ocean currents act as conveyor belts, transporting heat both vertically between the water's surface and the depths and laterally from one area of the globe to another. This effect is so significant that the temperature and precipitation patterns can change dramatically when currents do.

For example: Shortly after the last ice age, northern Europe experienced a shocking change in climate from 10,800 to 9,500 BC. At the start of this period, temperatures plummeted in a matter of decades. It became 7° Celsius colder, and glaciers started forming in England! The cold spell lasted for more than a thousand years, but it ended as suddenly as it had begun.

Why? The most popular theory is that a huge lake in North America formed by melting glaciers burst its bank—and in a massive torrent lasting for years, the water from this lake rushed out to the northern Atlantic Ocean. By floating atop the denser salt water, this fresh water blocked a major current: the Atlantic Meridional Overturning Circulation. This current brings warm water north and helps keep northern Europe warm. When it shut down, northern Europe was plunged into a deep freeze.

Right now, global warming is causing ice sheets in Greenland to melt and release fresh water into the North Atlantic. Could this shut down the Atlantic Meridional Overturning Circulation and make the climate of northern Europe much colder? In 2010, Keller and Urban [4] tackled this question using a simple climate model, historical data, probability theory, and lots of computing power. Their goal was to understand the spectrum of possible futures compatible with what we know today.

Let us look at some of the ideas underlying their work.

Box Models

The Earth's climate is too complex to simulate from the bottom up using basic physical principles, at least for now. The most detailed models we have use simplifying assumptions—and even so, they take days to run on very powerful computers. So, to make reasonable predictions on a laptop in a tractable time frame, geophysical modelers need to use some clever tricks.

First, it is possible to split geophysical phenomena into categories, or boxes, containing only strongly related things. For example: atmospheric gases, particulate levels, and clouds all affect each other strongly; likewise, the heat content, currents, and salinity of the oceans all interact strongly. However, the interactions between the atmosphere and the oceans are weaker, and we can approximately describe them using just a few parameters, such as the amount of atmospheric CO_2 entering or leaving the oceans. Clearly these interactions must be consistent—for example, the amount of CO_2 leaving the atmosphere box must equal the amount entering the ocean box—but breaking a complicated system into parts lets different specialists focus on different aspects; then we can combine these parts and get an approximate model of the entire planet. The box model used by Keller and Urban is shown in figure 1.

Second, it turns out that simple but effective box models can be distilled from the complicated phys-
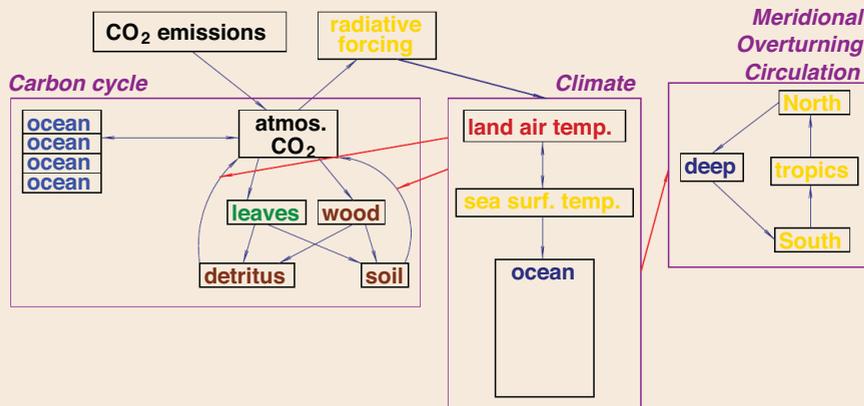
Figure 1. The box model used by Keller and Urban.

ics in terms of forcings and feedbacks. A *forcing* is a measured input to the system, such as solar radiation or CO_2 released by burning fossil fuels. As an analogy, consider a child on a swing: The adult's push every so often is a forcing. A *feedback* describes how the current state of our system influences its future state. In the swing analogy, one feedback is how the velocity will influence the child's future height. Specifying feedbacks typically uses knowledge of the detailed low-level physics to derive simple, tractable functional relationships between groups of large-scale observables, a bit like how we derive the physics of a gas by thinking about collisions of lots of particles.

However, it is often not feasible to get actual settings for the parameters in our model starting from first principles. In other words, often we can get the general form of the equations in our model, but they contain a lot of constants that we can estimate only by looking at historical data.

PROBABILITY MODELING

Suppose we have a box model that depends on some parameters S. For example, in Keller and Urban's model, S is a list of 18 numbers. To keep things simple, suppose S is an element of some finite set. Suppose we also have huge hard disk full of historical measurements, and we want to use this to find the best estimate of S. Because our data is full of "noise" from other unmodeled phenomena, we usually cannot unambiguously deduce the correct value of S. Instead we have to look at things in terms of probabilities. More precisely, we need to study the probability that S takes some value s given that the measurements take some value. Let's call the sequence of measurements M, and again let's keep things simple by saying that M takes values in some finite set of possible measurements.

The probability that $S = s$ given that M takes some value m is called the conditional probability

$P(S = s \,|\, M = m)$. How can we compute this conditional probability? This is a somewhat tricky problem.

One thing we can more easily do is repeatedly run our model with randomly chosen settings and see what measurements it predicts. By doing this, we can compute the probability that given setting values $S = s$, the model predicts measurements $M = m$. This again is a conditional probability, but now it is called $P(M = m \,|\, S = s)$.

This is *not* what we want: It's backwards! But here *Bayes' rule* comes to the rescue, relating what we want to what we can more easily compute:

$$P(S = s \,|\, M = m) = P(M = m \,|\, S = s)\frac{P(S = s)}{P(M = m)}.$$

Here $P(S = s)$ is the probability that the parameters take a specific value s, and similarly for $P(M = m)$. Bayes' rule is quite easy to prove, and it is a general rule that applies to any random variables, not just the parameters and the measurements in our problem [7]. It underpins most methods of figuring out hidden quantities from observed ones. For this reason, it is widely used in modern statistics and data analysis [5].

How does Bayes' rule help us here? We need to start with a guess about the probability $P(S = s)$; this is called our *prior*. We might assume the settings S are equally likely to take any value in some range, or we might use the opinions of experts to make a more informed guess.

Second, as mentioned, we can compute $P(M = m \,|\, S = s)$. Finally, the number $P(M = m)$ is independent of our choice of settings, so we can treat it as a constant. Thus, we can use Bayes' rule to compute $P(S = s \,|\, M = m)$ as a function of the settings s—at least up to a constant factor. Then, since probabilities must sum to 1, we can figure out this constant.

We call $P(S = s \,|\, M = m)$ the *posterior*, because it is the probability that the settings take some value *after* taking into account our measurements. Knowing this lets us do many things. It lets us find the most likely values of the parameters for our model, given our hard disk full of observed data. It also lets us find the probability that the settings lie within some set. This is important: If we're facing the possibility of a climate disaster, we don't want to know just the most likely outcome. We would like to know that with 95 percent probability, the outcome will lie in some range.

An Example

Let us look at an example much simpler than that considered by Keller and Urban. Suppose our measurements are real numbers $m_0, m_1, ..., m_T$, related by

$$m_{t+1} = sm_t - m_{t-1} + N_t.$$

Here s, a real constant, is our setting, while N_t is some noise: an independent Gaussian random variable for each time t, with mean zero and some fixed standard deviation. Then the measurements m_t will have roughly sinusoidal behavior but with irregularity added by the noise at each time step, as shown in figure 2. Note how there is no clear signal from either the curves or the differences that the green curve is at the correct setting value while the blue one has the wrong one: The noise makes it nontrivial to estimate s. This is a baby version of the problem Keller and Urban faced.

Markov Chain Monte Carlo

Having glibly said that we can compute the conditional probability $P(M = m \mid S = s)$, how do we actually do this? The simplest way would be to run our model many, many times with the settings at $S = s$, and determine the fraction of times it predicts measurements consistent with m. This gives us an estimate of $P(M = m \mid S = s)$. Then we can use Bayes' rule to work out $P(S = s \mid M = m)$, at least up to a constant factor.

Doing this by hand would be incredibly time consuming and error prone, so computers are used for this task. In our example, we do this in figure 3. As we keep running our model over and over, the curve showing $P(S = s \mid M = m)$ settles down to the right answer. However, this is computationally inefficient, as
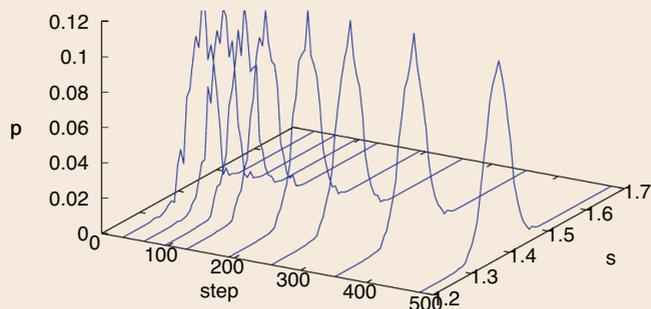
Figure 3. Estimates of the conditional probability $P(S = s \mid M = m)$, where at each step we run the model at all possible settings, and we keep a running total of the fraction of times each measurement m occurs when the setting is s.

shown in the probability distribution for small numbers of samples. This has quite a few kinks, which disappear only later. The problem is that there are lots of possible choices of s to try. And this is for a very simple model!

When dealing with the 18 settings involved in the Keller and Urban model, trying every combination would take far too long. A way to avoid this is *Markov Chain Monte Carlo* sampling. Monte Carlo is famous for its casinos, so a *Monte Carlo algorithm* is one that uses random steps to converge upon the answer. A *Markov chain* is a random walk: for example, where you repeatedly flip a coin and take one step right when you get heads, and one step left when you get tails. In Markov Chain Monte Carlo, then, we perform a random walk through the collection of all possible settings, collecting samples. From these samples we can estimate the density in the same way as before, only with greater computational efficiency.

The key to making this work is that at each step on the walk, a proposed modification s' to the current settings s is generated randomly—but it may be rejected if it does not seem to improve the estimates. By cleverly choosing the rule for doing this, in a way motivated by Bayes' rule, the random walk can be designed so that in the long run it visits each setting s with a frequency equal to $P(S = s \mid M = m)$. With some additional tricks—such as discarding the very beginning of the walk—this gives a set of samples that can be used to estimate $P(S = s \mid M = m)$. For details, try the textbooks by Bolstad [3] or Press et al. [6].

Figure 4 shows the results of using the Markov Chain Monte Carlo procedure to figure out $P(S = s \mid M = m)$ in our example. Compared with the method in figure 3, we run the Markov Chain Monte Carlo method for more steps, but each step involves running the model at just one setting, instead of all possible settings, so it is much more efficient.

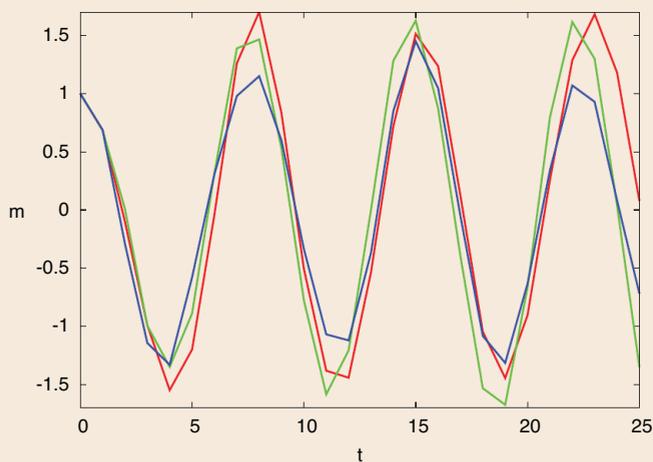The key advantage of Markov Chain Monte Carlo is

Figure 2. The example system: Red is predicted measurements for a given value of the settings, green is another simulation for the same s value, and blue is a simulation for a slightly different s.
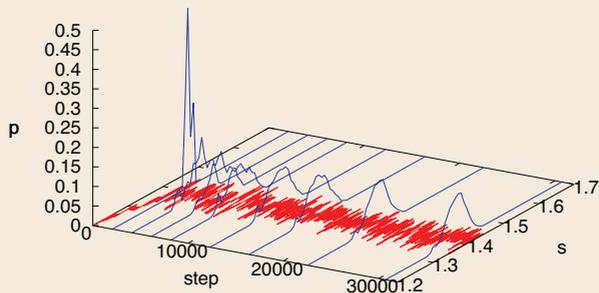
Figure 4. The estimates of $P(S = s \mid M = m)$ **using Markov Chain Monte Carlo, showing the current distribution estimate at increasing intervals. The red line shows the current position of the random walk. Again the kinks are almost gone in the final distribution.**

that it avoids performing many simulations with settings where the probability is low, as we can see from the way the red path of the random walk stays under the big peak in the probability density almost all the time. More impressively, it achieves this without us needing to know ahead of time where this peak will be!

This advantage becomes even more important as we move to dealing with more complicated models, where often the regions of high probability density are irregularly shaped multidimensional blobs, as it proves very effective at finding them efficiently.

Why is it worth doing so much work to estimate the probability distribution for settings for a climate model? One reason is that we can then estimate probabilities of events, such as the collapse of the Atlantic Meridional Ocean Current. And what's the answer? According to Keller and Urban's calculation, this current will likely weaken by about a fifth in the 21st century, but a complete collapse is unlikely before 2300. This claim needs to be checked in many ways—for example, using more detailed models. But the importance of the issue is clear, and we hope we have made the importance of good mathematical ideas for climate science clear as well.

EXPLORING THE TOPIC

The Azimuth Project is a group of scientists, engineers, and computer programmers interested in questions like this [2]. If you have questions, or want to help out, just email us. Versions of the computer programs we used in this paper available at [1].

Here are some practical issues to investigate:

- The collapse of the Atlantic Meridional Overturning Circulation would cause a temperature drop in northern Europe. Research the effects on humans and ecosystems of drops of 1°, 3°, and 7° Celsius.
- Learn in detail how Markov Chain Monte Carlo works [3 , 5 , 6]. Also learn about some other methods of parameter estimation, and try those in the example here.
- We've seen a one-dimensional system with one setting. Simulate some multidimensional and multi-setting systems. What new issues arise? ■

John C. Baez is professor of mathematics at University of California, Riverside, and a researcher for the Centre for Quantum Technologies in Singapore. He is cofounder of the Azimuth Project.
Email: baez@math.ucr.edu

David Tweed is a member of the Azimuth Project who did a Ph.D. on image recognition at the Department of Computer Science of the University of Bristol. His work makes extensive use of probability theory and statistics.
Email: david.tweed@gmail.com

FURTHER READING

1. Software available at *http://math.ucr.edu/home/baez/climate_math/*.

2. Azimuth Project, *http://azimuthproject.org.*

3. W. M. Bolstad, *Understanding Computational Bayesian Statistics*, John Wiley, New York, 2010.

4. K. Keller and N. Urban, "Probabilistic hindcasts and projections of the coupled climate, carbon cycle and Atlantic meridional overturning circulation system: a Bayesian fusion of century-scale measurements with a simple model," *Tellus A* **62** (2010), 737–750. Also available at *http://public.lanl.gov/nurban/pubs/moc-projections.pdf.*

5. J. K. Kruschke, *Doing Bayesian Data Analysis: A Tutorial with R and BUGS,* Academic Press, New York, 2010.

6. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, Section 15.8: Markov Chain Monte Carlo, Cambridge University Press, Cambridge, 2007. Also available at *http://scribd.com/doc/40633192/Cambridge-press-numerical-recipes-3rd-edition-sep.*

7. Eliezer S. Yudkowsky, "An Intuitive Explanation of Bayes Theorem," available at *http://yudkowsky.net/rational/bayes.*