MITx Video | 12.340xClimMod07v01

A critically important step in creating climate models is evaluating them. How do we know if we've got it right? Unfortunately, there are very few tests of the model as a whole. We don't actually have very many degrees of freedom in the observed climate system to test against. We have annual and diurnal cycles. We can try to make sure that they produce reasonably good weather forecasts, when they run in that mode.

We have variability of the climate during the 20th century, where we had reasonably good instrumental records to compare against. We could hope to test climate models on much longer time scales, and their response orbital variations.

But at the end of the day, the number of degrees of freedom we realistically have to test climate models may be fewer than the number of free parameters that we can change in the model. So if we have more free parameters than tests against independent degrees of freedom, tuning the model, or optimizing it, may be mathematically ill posed.

One alternative is to take all the different components of the model, and try as best we can to test them rigorously offline. This is very difficult. Offline tests are, in fact, very arduous. Even if each of the components in fact satisfy these tests, it is not necessarily true that putting them all together in a model will result in a robust model.

The model as a whole might not work, even though all the sub-components are robust. A weak analogy to this might be a complicated electronic circuit. We can ensure, for example, that all the diodes, and transistors, capacitors and what-not are working. But if the circuit design is flawed, then putting all those components together in a circuit board may not produce a viable circuit.

One of the indications that models have been in fact optimized comes from comparing their performance during the period where we have observations-- the 20th century-- to projections by those same models. So the graph at the top of this slide shows-- in lurid colors here-- hind casts of global mean temperature during the 20th century. So each of the different-colored lines is a simulation by a different climate model. The gray line that you can see along here is the multi-model mean and that can be compared to observations of the global mean temperature, in black here. One can see, for example, the effects of individual volcanic eruptions here and here.

The fit is, in general, quite good. The multi-model global mean temperature follows the observations rather well. If we run the same set of simulations, by the way, shown at the bottom, but we omit changes in anthropogenic forcing-- mostly greenhouse gases, but also aerosols-- we can see that the fit is quite a bit poorer, particularly after about 1970. So this gap here is some evidence that the recent warming-- the warming over the last 40 years or so-- is caused mostly by anthropogenic effects.

However, if we take that or a similar suite of climate models and run them forward, all exposed to the same general emissions scenario, we get a scatter that looks like what you see in this diagram. So this is a multi-model average in black. And then each of a set of models-- shown in the table at the right in different colors-- projected from the year 2000 or so up to mid century.

One can see quite an enormous scatter among these models, more so than one saw in the hind casts of the 20th century. And this scatter shows that there is indeed quite a bit of uncertainty in forward projections, even of something like global mean temperature.

Let's look at the performance of these models weighed against observations in a different way. So what you see at the upper left here is the annual mean surface air temperature for the period 1980 to 2005. This is a mean across many different ensembles of climate models.

And the graph at the right shows the difference between the multi-model mean temperature and the observed temperature, as reported in something called a reanalysis product. It's a way of optimally estimating the state of the system, given diverse and somewhat sparse observations.

You can see that there are some systematic biases. For example, the eastern parts of subtropical ocean basins are a bit warm in the models. The central part of the North Atlantic is a bit cool here, and so forth.

The chart at the lower left shows basically the multi-model mean of the absolute error. So we're just taking the magnitude of the error and averaging it across the models. Here again, one sees a tendency for there to be more error over land than over the ocean, although there are some exceptions, such as the far southeastern Atlantic.

And the chart at the bottom right is an estimate of the uncertainty of the analysis itself, in this case made by comparing different reanalyses of the climate state by different groups of people. So there are some errors as well, but by comparing the right and left parts at the bottom of this diagram, one can see that there's more difference among different climate models than there are among different climatological data sets.

Here's a chart showing the multi-model mean bias in seasonality. This is defined as the difference between the December, January, and February mean temperature and the June, July, and August mean temperature. One can see that there is a bias over land. Essentially, in the winter it's too cold over the continents in the northern hemisphere, and a bit too warm over the oceans. So there's quite a difference in the continentality of the seasonal cycle of temperature between models and observations.

The left-hand side of this chart shows the multi-model mean precipitation on the left, and on the right, the multi-model mean bias compared to a global precipitation climatology. The blue colors here indicate that the models are producing too much precipitation. And the reddish colors indicate that the models are producing too little precipitation.

So there are all kinds of interesting patterns in the bias. Here the models are essentially shifting the intertropical convergence zone [ITCZ] too far to the south. There's too much rain near the equator, and not enough north of the equator.

In the Pacific region, what should be a single ITCZ is split into two parts. This is a well-known, systematic model bias. It tends to produce a double ITCZ in the Western Pacific, where nature usually has one.

We can also see biases in things like the short-wave cloud radiative effect, by comparing that quantity as calculated across many different climate models, with satellite-derived cloud long-wave and short-wave radiative effects. The upper left chart in this diagram shows the difference between the model-estimated short-wave cloud radiative effect, and that estimated from satellite observations. And one can see by looking at the chart at the bottom that the magnitude of the short-wave radiative error is on the order of 10s of watts per meter squared.

Likewise, the middle chart shows the difference between the model mean, and the observed long-wave cloud radiative effect. There's too much long-wave trapping in the deep tropics, and not enough in some places outside the tropics. And if you add the two together, you get the net cloud radiative effect. Again, the same color scale at the bottom. There is too little cloud albedo effect, for example, in the stratocumulus regions of the eastern subtropical oceans.

The charts on the right compare the distribution with latitude of the zonal mean-- short-wave at the top, long-wave at the bottom-- and net cloud radiative forcing with observations. The multi-model mean is shown by the thick, red line. Two estimates from observations are shown by the thick, black line and the thick, black, dashed line.

And then the individual ensemble members of the models are shown by that light gray background. So one gets an appreciation, particularly when one looks at the net cloud radiative forcing. For example, in the tropics, there's a great deal of variation from one model to the next in the magnitude of that forcing.

Looking at the ocean, one see some systematic biases. So what you see at the left is essentially the difference between a multi-model mean average ocean temperature, and observed ocean temperatures as a function of latitude and depth. So we're talking about zonally average temperatures.

This chart is split between a top half and a bottom half, simply because they're different depth scales. The top half goes from the surface to 1,000 meters, and the bottom half from 1,000 to 5,000 meters depth. So the models are effectively mixing heat down too rapidly in the upper ocean, so you have a warm bias in the models in the thermocline region between, say, 50 or 100 meters depth down to 1,000 meters.

The right-hand chart is the same thing, but for salinity biases.

So these are some ways that we have of evaluating how well the climate models are doing. But there are many different models. There are very many different switches in the models-- parameterizations, parameters that can be tuned-- and optimizing these models and evaluating against observations is actually very difficult.