

A Climate for Math

ITHE SPACE OF ONE WEEK in June 2021, a small town in British Columbia called Lytton unexpectedly turned into a laboratory for climate change. From June 27 to 29, an unprecedented “heat dome” settled over the Pacific Northwest. Temperatures skyrocketed to 116 degrees (Fahrenheit) in Portland and 108 degrees in Seattle. On the Canadian side of the border, it was even hotter. For three days in a row, Lytton set all-time records for the entire country of Canada, topping out at an amazing 121 degrees Fahrenheit (49.6 Celsius) on June 29.

And then, on June 30, Lytton burned to the ground.

For meteorologists, as well as the inhabitants of Lytton, it was an eye-opening introduction to what the world might look like in a few decades. Temperature records are usually beaten by a degree or two, but Lytton’s record smashed the previous record for all of Canada by eight degrees. Hundreds of people died in the heat wave in Canada and the U.S., because most people in the Pacific Northwest had no air conditioners. The cool climate normally makes them unnecessary.

Events like the 2021 heat dome make it seem as if we are living in a strange new world, yet such events seem to be happening more and more often. In 2020, California wildfires injected so much smoke into the atmosphere that cities hundreds of miles away were shrouded in twilight-like gloom (see Figure 1). Hurricane Harvey in 2017 dumped more than 30 inches of rain on the



Figure 1. *In September 2020, orange skies due to distant wildfires enveloped much of California in a crepuscular gloom, even in midafternoon. (Figure courtesy of Dana Mackenzie.)*

Houston area, an inundation that should happen only once in 10,000 years, according to some climate scientists. (See Figure 2.)

Such extreme events beg us to ask questions like these: Was this heat wave, or this storm, a result of climate change? How much more often can we expect these extremes to happen in a warming world? Where will they occur? What other changes can we expect? And also—what changes have already happened? These questions demand mathematical answers. In a world that is so different from the one we grew up in, mathematics is almost our only guide to what will happen next.

Historically, climate science has been dominated by two kinds of mathematics, which have sometimes been at odds with one another. “I like to make an analogy to medical science,” says Francis Zwiers, a climatologist at the University of Victoria. “When you go to your doctor, you might get two kinds of advice. One might be to get a biopsy, and have the pathologist analyze a sample from your body. Or you might be given advice from epidemiology, which tells you the risk factors that affect populations like you.”

In climate science, the “pathologists” are climate modelers, who use partial differential equation models based on physics to identify specific processes happening in specific places. The “epidemiologists” are statisticians, who look for patterns in data. In addition, a third way has recently emerged—machine learning, which has unique strengths and weaknesses that complement the traditional approaches. We need to use every possible mathematical tool because in the end it isn’t about numbers. As Lytton showed us, it’s about human lives and livelihoods.

Most articles about climate change, understandably, place the climate news in the foreground and the mathematics far in the background, if it is mentioned at all. This chapter will take the opposite approach, and shine a spotlight on four mathematical and statistical techniques: extreme value analysis, differential equations, changepoint analysis, and machine learning.



Figure 2. *Flooding in Houston due to Hurricane Harvey. (Figure courtesy of Win McNamee/Staff, via Getty Images.)*

The Math of Extremes

The 2021 heat dome is a perfect example of a phenomenon that can be studied in more than one way. If you are a statistician, extreme temperatures call to mind the generalized extreme value distribution, which was first discovered by Ronald Fisher and Leonard Tippett in 1928, and independently rediscovered by Boris Gnedenko in 1943.

Imagine that each day the high temperature is drawn by a weather genie from a climate hat. The hat contains temperatures that come from a fixed probability distribution with a fixed mean and variance. The static nature of this distribution captures the idea that climate is stable; the fact that each day's high is drawn at random captures the idea that weather is variable.

Of course, high temperatures aren't actually determined this way, because there are day-to-day correlations and seasonal effects. But mathematicians like to start with simple assumptions, in order to build their intuition. Fisher, Tippett, and Gnedenko proved that in this simple model, the distribution of the maximum temperature over a large number of draws will (if it converges to anything) converge to one of three distributions. If the "climate hat" has an absolute maximum temperature, then the maximum of a large number of draws from that hat will follow the reverse Weibull distribution, shown in Figure 3 (next page, top). Here the blue curve represents the probability distribution in the climate hat. Notice that the most likely high temperature has been normalized to 0.0 and the absolute maximum temperature has been normalized to 1.0. The orange curve represents the likely hottest temperature over a 100-day period, which peaks around 0.8. This means that you would expect the hottest temperature in 100 days to be about 0.8, and you would be surprised if it was as high as 0.99. Over a 1000-day period (green curve) you would expect to see at least one day with a temperature of 0.95. Both the orange and green curves are skewed to the right but have very weak tails on the right-hand side.

On the other hand, if the probability distribution in the climate hat looks like Figure 3 (middle, blue), with a long right-hand tail that follows a power law, then the distribution of maximum temperatures over a 100-day period (orange) or a 1000-day period (green) is much more widely spread out. This type of distribution is called a Fréchet distribution, and it is "heavy-tailed" in the sense that extreme events are much more likely than in the reverse Weibull.

Finally, a tricky but important class of climate hats have exponentially decreasing tails; for example, the climate hat may have a Gaussian or normal distribution (Figure 3 (bottom), blue curve). The corresponding distributions for the extreme temperature, called Gumbel distributions, are heavier-tailed than the Weibull but lighter than the Fréchet. The Gumbel distribution has a high likelihood of "2-sigma" events happening in a 100-day period (orange) and "3-sigma" events in a 1000-day period (green) but 10-sigma events, for example, are wildly unlikely. Unlike the Weibull distributions, the Gumbel distributions are skewed to the left.

In a world that is so different from the one we grew up in, mathematics is almost our only guide to what will happen next.

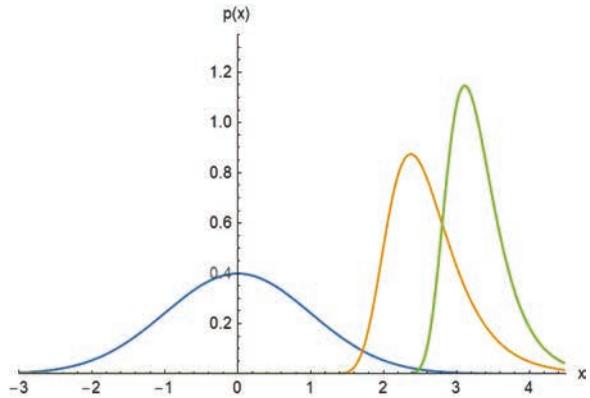
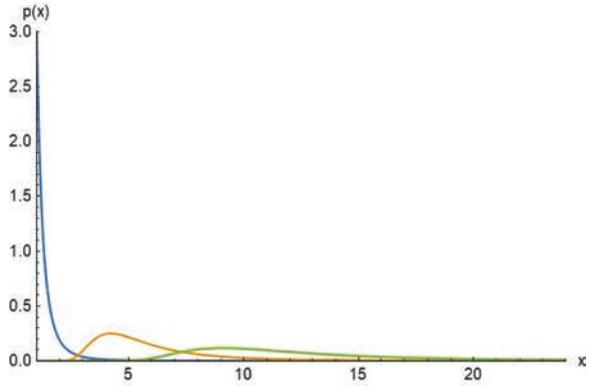
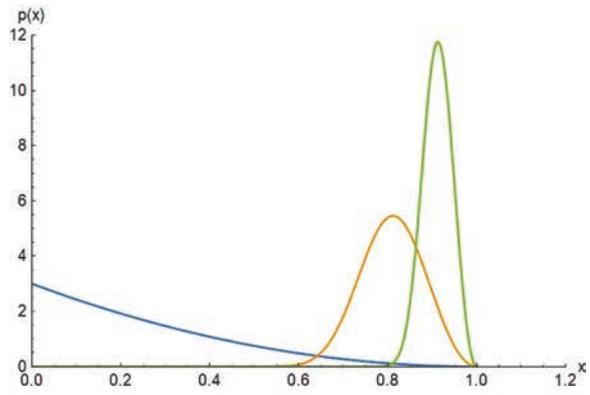


Figure 3. Extreme value distributions. Blue curves represent three different models for the probability distribution of the high temperature in a single day (in each case, with temperature rescaled so that 0.0 is the most likely high). Orange and green curves show the corresponding probability distributions for the maximum temperature over 100 days (orange) or 1000 days (green). (Figure Credit: Alex Hansen, *The Three Extreme Value Distributions: An Introductory Review*, *Frontiers in Physics*, 10 December 2020, Copyright ©2020 Hansen. Licensed under Creative Commons Attribution 4.0 International (CC BY 4.0) license, <https://creativecommons.org/licenses/by/4.0/>.)

The Fisher-Tippett-Gnedenko theorem paints a very optimistic picture of the predictability of extreme temperatures (and many other kinds of extremes). “That’s one reason that climatologists have glommed onto it,” Zwiers says. “It has been heavily used for 20 years now.”

Unfortunately, the 2021 heat dome was so extreme that it broke the model. “Usually, the extreme value distributions for temperature end up being light-tailed,” Zwiers says—in other words, they follow a Weibull distribution. “That makes sense from a physical point of view because there is only so much energy available, and that sets the fixed upper bound for the fitted distribution.” However, the inferred upper bound from historical data is significantly below 121 degrees, the temperature that was actually observed in Lytton. “So, if you use historical data, you can’t estimate the probability of this event,” Zwiers says.

A week after the heatwave, a group of scientists called the World Weather Attribution initiative said basically the same thing: the heat dome was “virtually impossible without human-caused climate change.” Even *with* the existing amount of global warming, they estimated that it was a once-in-a-thousand-year event, but in a future world with 2 degrees (Celsius) of global warming, it would become a once every 5–10 year event. However, even these calculations fail to convey the idea that the probabilistic model of drawing out of a climate hat simply fails to replicate this event. An alternative strategy is to try explaining what happened in physical terms—a strategy that is possibly more enlightening but also considerably more controversial.

In the short term, the heat dome was caused by a feature called an “omega block” in the jet stream. A region of high pressure forms a barrier that the jet stream can’t go through, so instead it loops around, forming a shape like the Greek letter Ω . If that omega stays in place for an extended period of time, the hot and dry weather beneath it will build and build.

Several studies in recent years, employing both mathematical analysis and computer models, have suggested that climate change may make this sort of stationary weather pattern more frequent. In climate models (see next section), the mid-latitudes can become a waveguide for traveling waves, called Rossby waves, which are meanders in the jet stream on the scale of thousands of kilometers. In Rossby wave conditions, the jet stream typically makes 6 to 8 north and south oscillations in one circuit around the globe.

In the northern hemisphere, Rossby waves propagate from east to west, against the prevailing winds. However, their overall velocity (taking into account the prevailing winds) can either be eastward, westward... or stationary. As Michael Mann, a climate scientist at Pennsylvania State University, wrote in a 2018 article, in the latter case “a pronounced amplification of waves that are excited by orographic or thermal forcing can occur.” (Lytton was nestled at the base of the Pacific Coast Range—an orographic forcing—and was subjected to several days of building heat—a thermal forcing.)

A stationary Rossby wave may seem like just a fluke event that has nothing to do with global climate change, but it’s not necessarily so. Climate change has weakened the temperature gradient between the poles and the equator, and that reduces

...they estimated that [the heat dome] was a once-in-a-thousand-year event, but in a future world with 2 degrees (Celsius) of global warming, it would become a once every 5–10 year event.



Andrew Majda. (Photo courtesy of Gerta Keller.)

the mean velocity of the mid-latitude winds. The conditions for a stationary or “quasi-resonant” Rossby wave are still not frequent, but they occur more often, and Mann’s group has even identified a temperature profile—a “quasi-resonant amplification fingerprint”—that is most conducive to their formation.

Climatologists from the statistical camp, on the other hand, contend that there is no evidence yet in the data for an increased frequency of omega-block conditions. And even the physical models are not unanimous. “From a basic science viewpoint, it’s hard to tell because there are competing effects,” says Edwin Gerber, a mathematician at New York University who specializes in climate. For example, the reduction in the north-south temperature gradient decreases the amount of energy available to drive winds and storms. On the other hand, a warming planet will have more water vapor—and water vapor itself traps more energy in the atmosphere. “We have to figure out which effect wins,” Gerber says. “My expert opinion is that we don’t know yet.”

A Tropical Wave

If the mid-latitude climate is hard to figure out, the tropics are even harder. They are hotter than the mid-latitudes, so more water evaporates and forms clouds. These clouds have two opposing effects on the climate: they cool Earth by reflecting sunlight into space, and they warm Earth by absorbing infrared radiation from the surface. (Water vapor is actually a more potent greenhouse gas than carbon dioxide.) Computerized global climate models (GCMs) have trouble reproducing cloud effects because clouds are smaller than the 50-kilometer grid scale used in the most fine-grained models. In the mid-latitudes this is not as much of a problem because larger-scale features, like warm and cold fronts, Rossby waves and the jet stream, dominate the weather.

But in the tropics, all of the scales are connected. One mathematical model that elegantly captures this property is the “skeleton model” proposed by Andrew Majda, an applied mathematician at New York University who died in 2021. Majda was perhaps the first mathematician to cross over and make almost as big a mark in meteorology as in mathematics. “Some people told him that the problems he was trying to address were too hard for an outsider like him and that he should not waste his time and energy. Many others admired his original and unconventional approach,” wrote his collaborator Boualem Khouider in an obituary for *Bulletin of the AMS*. That is, the *Bulletin of the American Meteorological Society*, not the American Mathematical Society.

Majda was not discouraged by the detractors. Beginning with a paper published in 2009, he investigated one of the “holy grails” of meteorology—an explanation of the Madden Julian Oscillation (MJO), discovered in the early 1970s. This is a planetary-scale wave, with wavelength roughly half of Earth’s circumference, that travels west to east at a brisk jogging pace (about 5 meters per second), lasts for 20 to 60 days (much longer than weather systems in the mid-latitudes) and usually contains several storms within it. Because of its size and long duration, “modelers call it a phenomenon that lies at the intersection between climate and weather,” says Khouider. (See Figure 4)

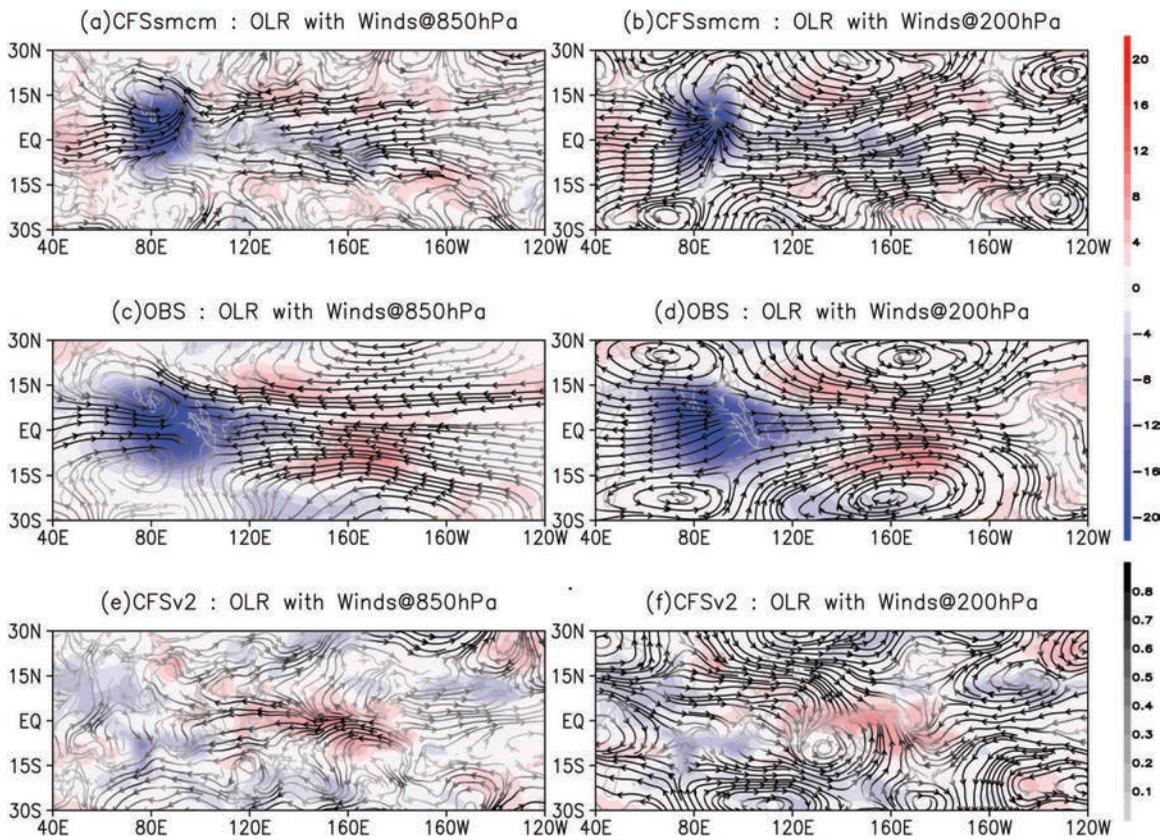


Figure 4. Madden-Julian Oscillation in simulations and in real data. (a), (b)—Stochastic simulations at two altitudes, incorporating small-scale processes of convection and cloud formation. (c), (d)—Real data at the same two altitudes. (e), (f)—Deterministic simulation without small-scale effects. Blue, red—Reflectivity (blue correlates with precipitation). The stochastic simulation accurately predicts storm location and predicts some aspects of the wind field, including three of the four gyres seen in part (d). The deterministic simulation bears very little resemblance to reality. (Figure courtesy of B. B. Goswami et al., “Improved Tropical Modes of Variability in the NCEP Climate Forecast System (Version 2) via a Stochastic Multicloud Model,” *Journal of the Atmospheric Sciences*, volume 74, issue 10, 3339–3364. Figure 5. ©American Meteorological Society. Used with permission.)

The great mystery of the MJO was, and still is, that atmospheric models based on pure dynamics (such as the Navier-Stokes equations of fluid mechanics) fail to predict its existence. Majda made it his goal to find the simplest set of equations beyond Navier-Stokes that could successfully account for it. The missing ingredient, he said, is water vapor.

To see why, it helps to start with a simple version of the Navier-Stokes equations called the primitive equations, first written down by Vilhelm Bjerknes in the early twentieth century. They take into account two peculiarities of Earth’s atmosphere: its vertical thickness is much less than its horizontal extent, and it sits on the surface of a spinning sphere. There are four “dry” equations (ignoring the role of water) and two that involve water. Here is what they look like:

$$\frac{\partial u}{\partial t} - \beta y v = -\frac{\partial \Phi}{\partial x} - \epsilon u,$$

$$\frac{\partial v}{\partial t} + \beta y u = -\frac{\partial \Phi}{\partial y} - \epsilon v.$$

The first two equations express conservation of momentum. The coordinates are x (longitude), y (latitude), and t (time). The unknowns are u (the east-west component of wind velocity) and v (the north-south component). The two beta terms are quite interesting: they represent the Coriolis force, which is perpendicular to the wind velocity and caused by Earth's rotation. Note that the Coriolis force drops to zero at the equator ($y = 0$), which makes tropical air masses behave differently from those in the mid-latitudes. The variable Φ represents gravitational potential or more precisely the *geopotential*, a hybrid of gravity and centrifugal force.

$$\frac{RT}{p} = -\frac{\partial \Phi}{\partial p}.$$

This is the vertical momentum or hydrostatic balance equation. The left-hand side is buoyancy, the right-hand side is the gradient of the geopotential. Notice that meteorologists do not use z as the vertical coordinate: they use p (air pressure), which is roughly a linear function of z , but not exactly. Cold air is heavier, and this is why temperature (T) enters into the balance equation.

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial \omega}{\partial p} = 0.$$

This relatively simple equation expresses conservation of mass. For consistency, ω , the vertical velocity, is computed in isobaric (x, y, p) coordinates, rather than (x, y, z) coordinates.

$$C_p \frac{\partial T}{\partial t} - \omega \left(C_p \frac{\partial T}{\partial p} - \frac{RT}{p} \right) = Q_1.$$

This ultra-important equation expresses the conservation of energy. Here C_p represents the specific heat of air, which is a function of p , and Q_1 represents all of the sources and sinks of heat. This includes sunlight and also latent heat generated by the condensation or evaporation of water. Therefore Q_1 is a function of q , the humidity, and this is how the water cycle enters into the fifth equation.

$$\frac{Dq}{Dt} = -Q_2.$$

The sixth and last primitive equation simply expresses the conservation of mass for water. The function Q_2 represents all the sinks and sources of water in the atmosphere.

The primitive equations give six equations in six unknowns: u, v, ω, Φ, T , and q . They lie at the heart of any weather model, whether it is solved by supercomputers or human beings. From here you can construct an almost limitless number of variations. One popular idea is to simplify them by ignoring effects that you expect to be small. For example, in the mid-latitudes the effect of

Because of its size and long duration, “modelers call [the Madden Julian Oscillation (MJO)] a phenomenon that lies at the intersection between climate and weather”...

latent heat from evaporation or condensation of water is small compared to the Coriolis force. So the earliest weather models, in the 1950s, could get away with ignoring Q_1 (as well as several other simplifying assumptions). In the tropics, though, where $y \approx 0$ and the Coriolis force is small, this assumption is no longer valid.

Another approach is to make the primitive equations more complex, in order to capture real observed phenomena that they cannot. That is what Andrew Majda did to model the Madden-Julian Oscillation. He added one new variable, called the wave activity function, and one more equation depending on the humidity q :

$$\frac{\partial a}{\partial t} = \Gamma qa.$$

The idea is to include an interaction between the wave activity and the humidity, which establishes a sort of predator-prey dynamics. Khouider explains: “The idea is that the activity function, a , feeds on the moisture provided by the MJO wave, q . They interact with each other, the moisture is depleted, and when the moisture is gone the wave action disappears. But it will move somewhere else where there is moisture.” That “somewhere else” is east of the wave. That is because the remaining primitive equations generate something called a Kelvin wave that travels east (opposite to Rossby waves) and creates favorable conditions for convection and cloud formation. “It happens automatically,” Khouider says.

Majda’s theory has several very attractive features. First, it is consistent with the observed phenomena. The MJO in his model drifts slowly east, just like the real MJO. It produces a quadrupole of counter-rotating storms, which are also seen in the real world. (This quadrupole shows up very clearly in Figure 4d.) It also has the very nice mathematical property of being neutrally stable. This means that it does not either grow or dissipate, but just cycles through wet and dry phases forever—much like a classical predator-prey model in the theory of dynamical systems.

This advantage is also a disadvantage, though: Majda’s theory requires a kick to get the MJO started and another kick to stop it. This is why he called it a “skeleton” theory: he felt that it would need to be completed with another theory to provide the “muscle” that starts and stops the oscillation. “He never actually had a model that combines the two,” Khouider says.

However, recent work of Khouider and others shows that the kicks do not require a separate model. You can simply replace the differential equation for the activity function by a stochastic equation, in which the activity function has a certain probability of increasing by one unit and a certain probability of decreasing. Because the model is neutrally stable, small random kicks can add up and generate a new wave out of nowhere. In this way, the stochastic skeleton model links the small scale of clouds to the planetary scale. More elaborate stochastic models can reproduce more detailed features of the MJO, as seen in Figure 4.

Majda’s skeleton model was indeed unconventional, introducing an abstract new variable (the activity function) that can’t

Majda’s theory has several very attractive features...it was unconventional...and showed that the life cycle of convection is important...



Robert Lund. (Photo courtesy of Robert Lund.)

actually be measured. But in this way, it gives us an idea what our data are missing. “Ten years ago, our best climate models didn’t get the MJO at all,” says Gerber. “Andy’s model showed that the life cycle of convection is important, and helps us improve the model. The memory on small scales gives you the global scale. Our parametrizations need to have that memory.”

How will climate change affect the MJO, and vice versa? The answer is so far unknown; some people predict more MJO activity and some predict less. Two things are certain. First, you can’t predict what will happen to the MJO until you can model the MJO. That box has now been checked. Second, the MJO is closely connected to many of the other big uncertainties of the global warming era. “It controls climate variables like clouds and moisture, which are important for the heating and cooling of Earth’s surface,” Bouider says. “It is associated with tropical cyclone [hurricane] activity. Your models cannot capture the probability of more hurricanes if you don’t get the tropical waves right.”

Changepoints

The whole concept of climate change presupposes that you know what “change” is. How can you tell when a time series of data has an abrupt change, either in mean or variance or trend line? That is a question that Robert Lund of the University of California at Santa Cruz has been studying for 15 years.

Lund likes to start with his favorite example: the annual average temperature, year by year, in Tuscaloosa, Alabama. (See Figure 5.) It appears that Tuscaloosa has had three periods of gradually rising temperature, interrupted by two sudden decreases in the 1930s and 1950s. According to Lund, there is a simple reason: In 1939, the temperature gauge was moved from downtown to a cooler site close to the river. Then, in the mid-1950s, a new gauge was installed that tended to give lower readings. If you took the data at face value, you would conclude that the temperature in Tuscaloosa had gone down by 0.1 degree per 100 years. But if you take into account the changes in equipment and location, the temperature has been increasing at a rate of 2 degrees per century. Quite a difference!

“In this temperature record, I actually know what is going on,” Lund says. “But there are 50 thousand temperature stations around the world, and at most of them we don’t have any record of what is going on.” So the question is: without knowing the history of the temperature station, how can we tell when the equipment changed? If we can find these “changepoints” and compensate for them, then we can meaningfully compare the 1930s to the 2000s.

If we are merely testing for the existence of one changepoint, a simple and beautiful solution was discovered by statistician Ian MacNeill in 1974. For every year up to year k , work out the difference between the average temperature before year k and the average temperature afterward. Scale the difference by dividing by the standard deviation of the measurements; let’s call this scaled statistic $B(k)$. Assuming the null hypothesis holds, i.e., that there was no changepoint, then the graph of B should look like a Brownian motion that is tied down at both ends so that $B(0) = B(n) = 0$. This kind of process is called a *Brownian bridge*.

The graph of B gives you two pieces of information. First, it identifies the most likely changepoint if there is one: namely, the point k where $B(k)$ is maximized. Second, it gives you a parameter-free test of your confidence in the null hypothesis. If the maximum of B is greater than 1.31, you can reject the null hypothesis with 95 percent confidence and conclude that there must have been a changepoint. As an example, for the data shown in Figure 6 (temperatures in central England) the statistical test strongly favors the hypothesis of a changepoint around 1988, which is likely due to climate change.

Unfortunately, the theory for multiple changepoints is by no means so clear. You could get a perfect fit to your data by making every year a changepoint, but that would be highly uninformative. Therefore, most methods impose a penalty for too many breakpoints, but they do not agree on what penalty to use. Lund likes one called the minimum description length, which is derived from information theory. He computes the most likely changepoint if there is one, or if there are two, or if there are three, and so on. The penalty term identifies which of these possibilities is most likely. The method is pretty good but not infallible. For example, out of 1000 simulated datasets that had three changepoints, it correctly identified 631 as having three changepoints, but in 288 cases it was too conservative and identified only two changepoints.

Lund has mostly used this approach to clean up records that have instrumental artifacts. But perhaps more interestingly, changepoint analysis can also be used to detect real changes in climate that are not instrumental artifacts. For example, he has studied the historical record of the number of hurricanes in the north Atlantic each year. His method picks up some known artifacts. Prior to the 1930s, fewer hurricanes were detected because the only ones we knew about were the ones that made landfall, plus the ones that happened to cross shipping lanes. As ship traffic increased before World War II, the number of reported hurricanes increased. In the 1960s, thanks to satel-

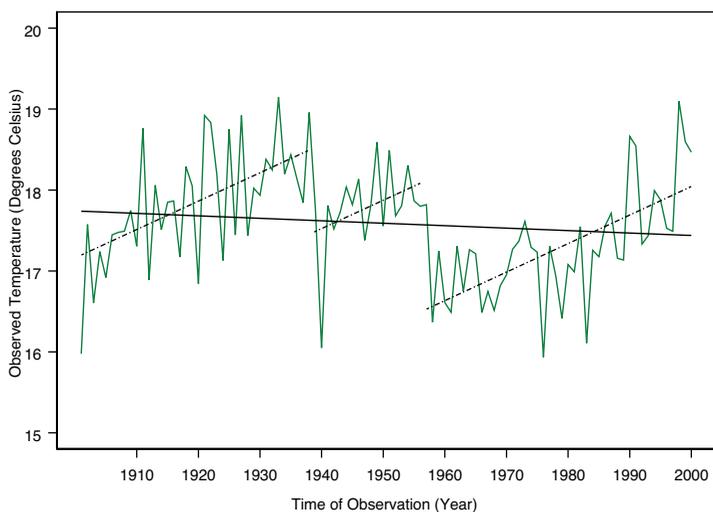


Figure 5. *Temperature graph for Tuscaloosa. (Figure courtesy of Robert Lund.)*

lites, we became able to detect all hurricanes, and the reported number jumped again.

But then in 1995, Lund says, “All hell broke loose.” It doesn’t matter how you parse the data. You can look for 2 changepoints, or 3, or 4; you can look at the count of hurricanes or their total energy. In all of these versions, 1995 shows up as the strongest changepoint. The number of hurricanes went up that year and it has stayed high ever since. In 2006 and 2020, the weather service even ran out of names for all the hurricanes and had to resort to Greek letters.

At present, climatologists disagree on the question of whether climate change will lead to more hurricanes or stronger ones. But, Lund says, “Their opinion is based mainly on computer models.” Changepoint analysis answers the question without models, albeit with an answer that is retrospective rather than prospective. Something happened around 1995 that made hurricanes more frequent, and we should try to figure out what it was.

Changepoint analysis can also be used in the reverse way, to rule out changes. Ten years ago, climate-change deniers were fond of claiming that there had been a “hiatus” in global warming beginning around 1998. In 2015, Niamh Cahill and Andrew Parnell of the University of Ireland, with Stefan Rahmstorf of the Potsdam Institute for Climate Impact Research, tested this claim with a slightly different changepoint model

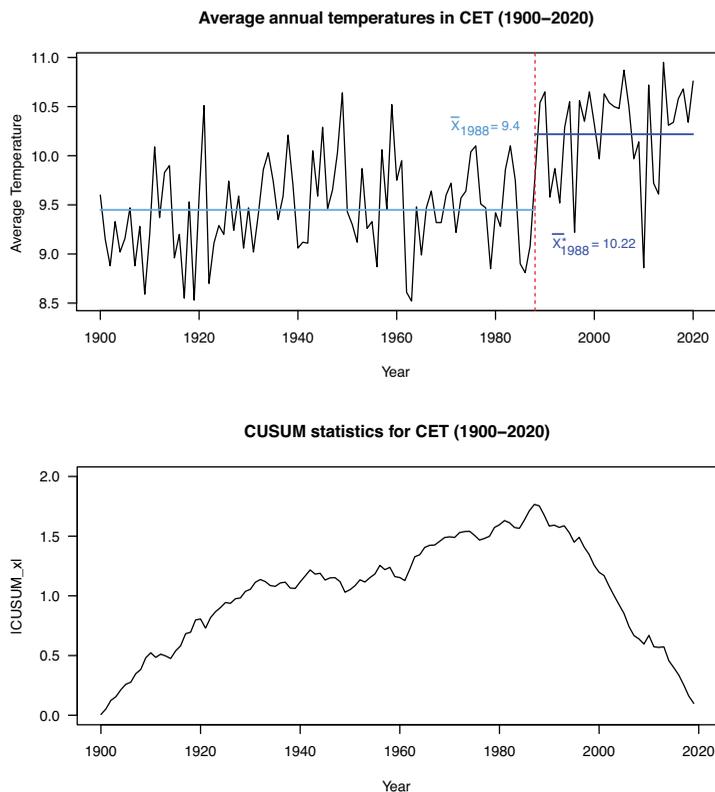


Figure 6. *Central England temperature and Brownian bridge process, indicating a significant change point around 1988. (Figure courtesy of Robert Lund.)*

(one that allows linear trends between changepoints, rather than step functions). They found that the data on world average temperatures best fit a curve with three changepoints in 1912, 1940, and 1970. There was no improvement when they forced the model to have a fourth changepoint in 1998. Furthermore, simulations showed that if there had been a hiatus beginning in 1998, they would have detected it with 92 percent probability. (Indeed, there was a hiatus between 1940 and 1970, which their method detects just fine.) Conclusion: the “hiatus” was fake news.

Machine Learning in Climate Change

In recent years, neural networks have given the whole field of machine learning a shot in the arm, with amazing and unexpected successes in tasks like image recognition, natural language processing, and even playing human games like chess and go. Neural networks are to some extent inspired by the architecture of the human brain—they can be thought of as networks of “neurons” that are connected in layers.

But this analogy obscures the mathematical essence of neural networks, which can be thought of as a very elaborate outgrowth of traditional statistics. The basic methods of statistics teach us how to estimate linear functions that approximate data. But some systems, such as weather and climate, are not linear. The “universal approximation theorem,” proven originally by George Cybenko of Dartmouth University in 1989, says that any nonlinear function, no matter how complicated, can be approximated by a composition of very simple nonlinear functions. The number of compositions corresponds to the number of layers in a neural network.

Thus, for example, a neural network can in theory “learn” a nonlinear function that will input an image and output a judgment of whether that image contains a face, or a zebra, or a hurricane. In the 2010s, computer scientists have made rapid and constant progress in constructing algorithms that can efficiently find these approximations that are promised by theory.

The most radical approach to machine learning in climate science would be to ditch the differential equation models that have been developed over decades, replace them with neural networks, and let the computer learn for itself how to predict weather and climate from observed data. There are several reasons why no one is very eager to do that. First, neural nets are notoriously hard to interpret. We can look at the equation for conservation of energy and understand what it is saying; but if we look at a neural network, we have no idea whether it has a neuron or layer of neurons that enforces conservation of energy. Also, neural nets that are trained on data from the current climate have no guarantee of extrapolating correctly to a different climate. “Usually in machine learning and artificial intelligence, you have a system that is stationary,” says Peer Nowack of the University of East Anglia. (For example, the rules of chess don’t change, and zebras will always look like zebras.) “But we don’t have this luxury in climate science. You might have to go back to simpler methods to make machine learning work.”

One simpler method is to use the machine to emulate other machines—in other words, to predict the output of global models that already obey the law of physics, predict the weather very

In recent years, neural networks have given the whole field of machine learning a shot in the arm, with amazing and unexpected successes in tasks like image recognition, natural language processing, and even playing human games like chess and go.

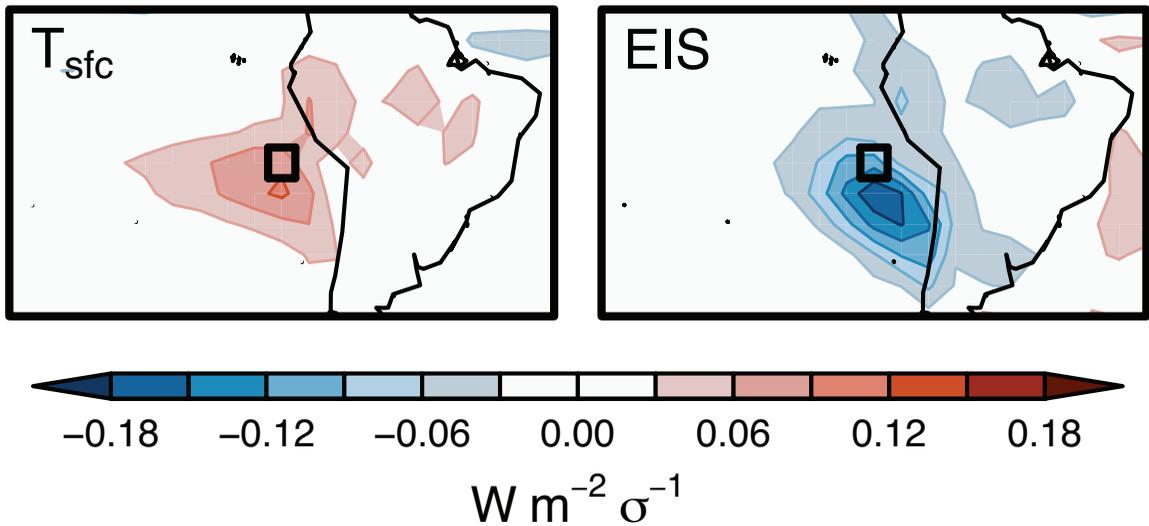


Figure 7. Teleconnections between cloud-producing regions and satellite reflectivity data. Reflectivity (a measure of cloud cover) is observed by satellite in the small square. Surface temperature (T_{sfc}) and “estimated inversion strength” (EIS) are estimated from climate models. In the left map, the reflectivity in the small square is most sensitive to the surface temperature in the shaded regions. (Figure courtesy of Paulo Ceppi and Peer Nowack, “Observational evidence that cloud feedback amplifies global warming,” Proceedings of the National Academy of Sciences, Vol. 118 No. 30, July 27, 2021.)

well in the short term, and that (we believe) have some ability to extrapolate to future climates. “It’s extremely expensive to run a climate model right now,” says Tapio Schneider of the Climate Modeling Alliance (CLIMA) at Caltech. “We are using tools from machine learning to accelerate learning from data by about 1000 times.” With an approach called “calibrate-emulate-sample,” they construct a simpler model with a couple hundred parameters that emulates the existing GCM’s. As new data come in, they update the parameters. This is a process called data assimilation. In fact, the “calibrate” step of calibrate-emulate-sample uses exactly the same mathematical tools as weather forecasting, such as Kalman filters.

“We are betting on retaining as much of the fundamental science as we can, and using data only when we can’t go any further,” says Schneider.

There is yet a third level of machine learning in climate science, which is conceptually even simpler. This approach is what Schneider calls “post-processing” the results of global climate models. Instead of emulating the GCM’s, you simply compare the predictions of climate models with what is observed in real time, through the lens of what is identified as important by a machine learning algorithm. This approach blurs the line between conventional statistical analysis and machine learning. It’s the approach that Nowack and his collaborator Paulo Ceppi take, and they call it “statistical learning.”

In a paper published in August 2021, Ceppi and Nowack used this approach to answer a question that has been controversial in climate science: will the change in cloud cover amplify global warming or retard it? As mentioned above, either answer is possible *a priori*. Clouds reflect sunlight (shortwave radiation) into

Reducing the uncertainty over the climate response to greenhouse gases will be crucial for adaptation and mitigation measures.

space and cool the earth; they also trap infrared light (longwave radiation) and warm the earth. Under global warming, the atmosphere changes, and so do the characteristics of the cloud cover. Either of the above effects could weaken or strengthen, and climate scientists have not been able to say definitively what the net effect would be.

Ceppi and Nowack used 20 years of data on the reflectivity and longwave absorption of clouds, recorded by satellite, to identify a variety of “cloud-controlling factors.” They used a linear machine-learning algorithm to give an estimate for $\lambda_C(r)$, the response in the shortwave (or longwave) radiation at each point r resulting from an increase in global temperature T :

$$\lambda_C(r) = \frac{\partial C(r)}{\partial T} \approx \sum_{\text{ccf}} \Theta_{\text{ccf}}(r) \cdot \frac{\partial(\text{ccf}(r))}{\partial T}.$$

Here “ccf” refers to an individual cloud-controlling factor, Θ_{ccf} denotes the sensitivity of the radiation to that factor, and the sum is over the cloud-controlling factors. The most important of these factors was the local sea surface temperature, $T_{\text{sfc}}(r)$. Figure 7 shows the relation between the satellite-observed radiation in the dark square and $T_{\text{sfc}}(r)$ in the Southern Hemisphere, with darker colors representing stronger correlations. Note that there are thousands of variables in this linear regression, because the reflectivity in each grid square is a function of the cloud-controlling factors at more than 1000 grid points surrounding that square.

With so many variables, there is a danger of overfitting, so Ceppi and Nowack used a modified version of regression called “ridge regression,” which penalizes nonzero correlations. This is similar in spirit, though not in detail, to the idea of penalizing changepoint models with too many changepoints, discussed in the previous section. The idea of penalizing an overly complex model with a “loss function” is a well-established paradigm in machine learning.

Another important point is that part of the *output* of the global climate models—namely, the anticipated changes in the cloud-controlling factors—becomes an *input* to their linear model. In effect, they trust the GCM’s to predict the cloud-controlling factors correctly. Their model is intended to focus on the one ingredient that the GCM’s cannot predict well: the response of the clouds, which is the greatest uncertainty in existing GCM’s. (See Figure 8, next page)

In the end, Ceppi and Nowack obtained a strong correlation (0.87) between the cloud-radiation feedback in their statistical learning model and the feedback simulated by 52 GCM’s. By extension, they argued that the same statistical learning function should also be able to predict the “true” cloud feedback. “We identify the most likely value, or range of values, for λ_C , and compare this with the models’ much wider range,” says Ceppi. In the previous state-of-the-art estimate of the cloud-radiation feedback (the fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC), released in 2013) the feedback was estimated at -0.2 to +2.0 watts per square meter per degree Celsius. In other words, it was most likely positive, but a weak negative feedback could not be ruled out. The new estimate by Ceppi and Nowack was 0.08 to 0.78, which implies that

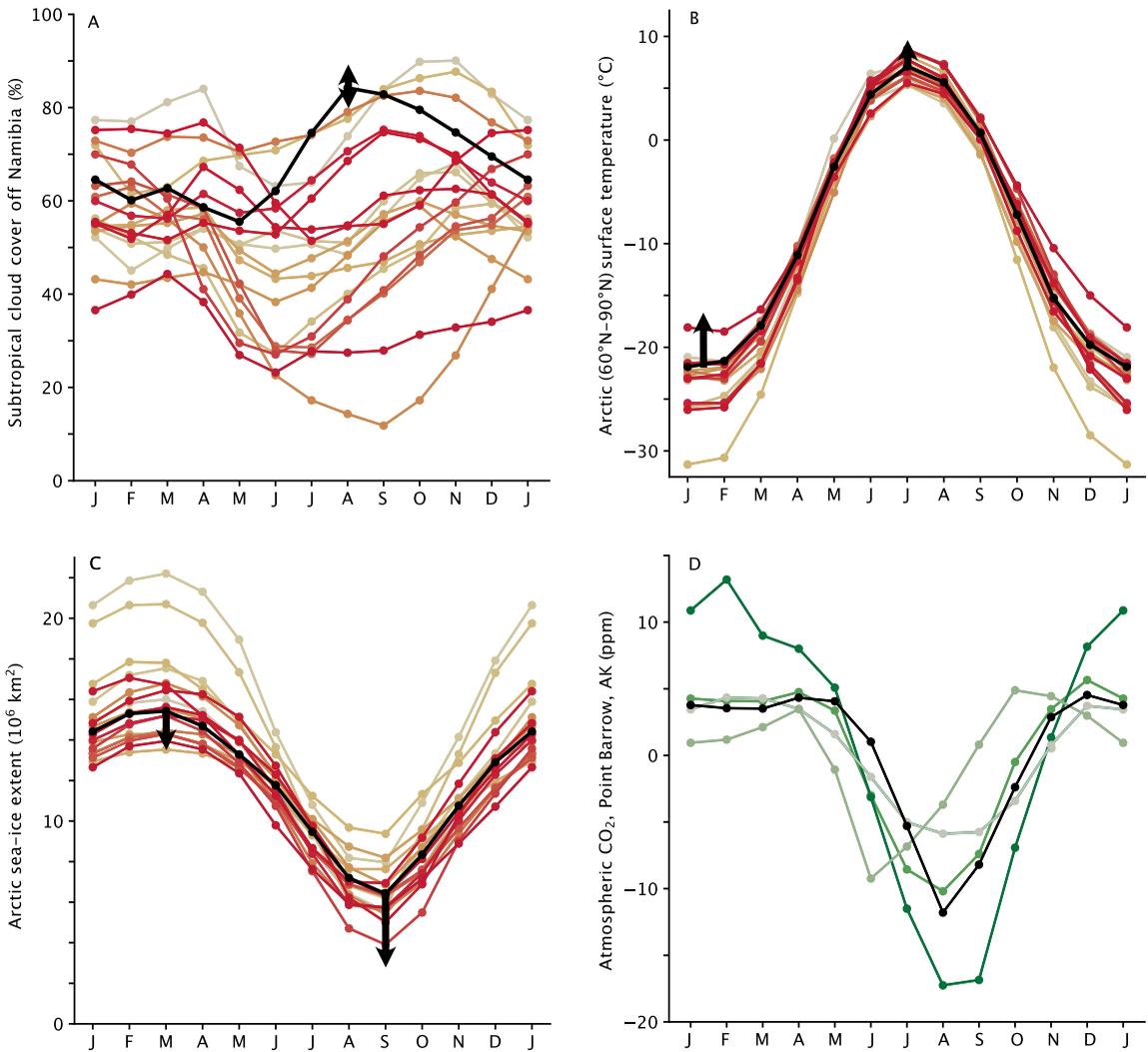


Figure 8. Difficulty of modeling cloud cover. Global circulation models (GCMs) reasonably well predict seasonal variations in temperature (b), ice cover (c) and carbon dioxide (d). But their projections of cloud cover (a) are little better than chance. (Figure courtesy of Tapio Schneider, Nadir Jeevanjee, Robert Socolow, “Accelerating Progress in Climate Science,” *Physics Today*, June 2021, 45–51. Figure 5.)

the uncertainty could be reduced by a factor of three and, for the first time, shows that the cloud feedback is unambiguously positive. (See Figure 9, page 111) “It is a major step forward that we could achieve that [reduction] using only the most recent and highest-quality satellite data as a source of information,” Nowack says.

Reducing the uncertainty over the climate response to greenhouse gases will be crucial for adaptation and mitigation measures. In general, the uncertainty has remained remarkably steady over the years. In 1979, Jule Charney estimated that if atmospheric carbon dioxide were to double, the worldwide average temperature would climb by 1.5 to 4.0 degrees Celsius at its new equilibrium point. This number is called the equilibrium climate sensitivity (ECS). The fifth IPCC assessment report,

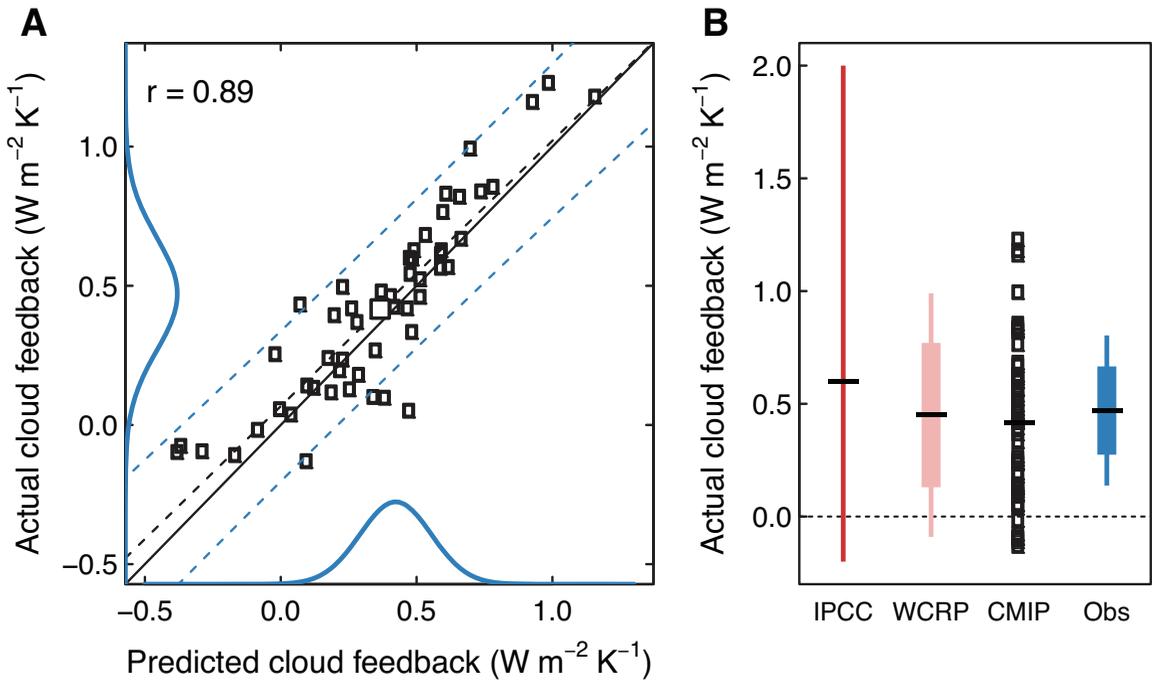
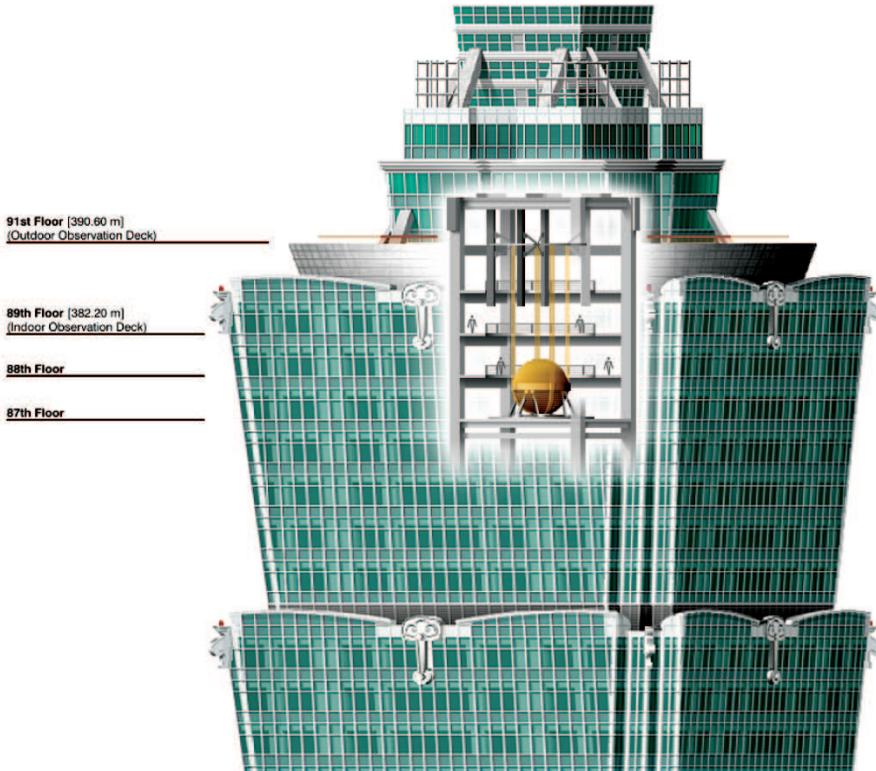


Figure 9. (a) High correlation between predicted and observed cloud cover feedback. (b) Ceppi and Nowack’s estimate of cloud cover feedback (blue) for the first time is unambiguously positive after taking into account all known sources of uncertainty. IPCC, WCRP, and CMIP are prior estimates. (Figure courtesy of Paulo Ceppi and Peer Nowack, “Observational evidence that cloud feedback amplifies global warming,” *Proceedings of the National Academy of Sciences*, Vol. 118 No. 30, July 27, 2021.)

more than thirty years later, gave almost the same confidence interval for the ECS: 1.5 to 4.5 degrees. “What happened over the years was that there were many unknown unknowns that became known,” says Ceppi. So even though climate scientists were learning more, it looked as if they weren’t.

The sixth assessment report, released in 2021, for the first time narrows the ECS down to an interval of 2.5 to 4.0 degrees, a factor of two reduction in the uncertainty. Ironically, the range of predictions of the global climate models themselves has not shrunk. Instead, the reduction in uncertainty came about through better understanding of individual processes, such as the cloud-radiative feedback studied by Ceppi and Nowack.

Plenty of uncertainties remain in climate forecasting: clouds, aerosols (particles in the atmosphere that promote formation of water drops), the melting of sea ice. In addition, certain low-probability but high-impact events, such as the collapse of the Antarctic ice sheet or changes to ocean circulation, are hard to forecast. However, the uncertainties are decreasing, thanks to a variety of mathematical methods, and the overall message is the same: Change ahead.



Good vibrations. *The resonant frequencies of a skyscraper are related to its engineered structures (beams and loads) by an eigenvalue equation. “Zero forcing” is an elegant and simple mathematical technique to constrain the number of repeated eigenvalues. In practice, a “tuned mass damper” like this one in the Taipei 101 skyscraper achieves a similar purpose. (Top figure courtesy of Wikimedia Commons, author: Armand du Plessis, licensed under the Creative Commons Attribution 3.0 Unported License. Bottom figure courtesy of Wikimedia Commons, author: somekindofhuman, licensed under the Creative Commons Attribution-ShareAlike 4.0 International License.)*