

Comment



SUPRATIM BHATTACHARJEE/ZUMA WIRE

Island communities in the Ganges–Brahmaputra delta in south Asia are at risk of flooding as a result of climate warming.

Climate simulations: recognize the ‘hot model’ problem

Zeke Hausfather, Kate Marvel, Gavin A. Schmidt, John W. Nielsen-Gammon & Mark Zelinka

The sixth and latest IPCC assessment weights climate models according to how well they reproduce other evidence. Now the rest of the community should do the same.

Computer models that project future climates are widely used for adaptation, mitigation and resilience planning. More than 50 such models were assessed and compared in the latest round of the Coupled Model Intercomparison Project, phase 6 (CMIP6), run by the World Climate Research Programme¹. It is crucial that researchers know the best way to use those outputs to provide consistent information for climate science and policy.

We are climate modellers and analysts who develop, distribute and use these projections. We know scientists must treat them with great

care. Users beware: a subset of the newest generation of models are ‘too hot’² and project climate warming in response to carbon dioxide emissions that might be larger than that supported by other evidence^{3–7}. Some suggest that doubling atmospheric CO₂ concentrations from pre-industrial levels will result in warming above 5 °C, for example. This was not the case in previous generations of simpler models.

Earth is a complicated system of interconnected oceans, land, ice and atmosphere, and no computer model could ever simulate every aspect of it exactly. Models vary in their

complexity, and each makes different assumptions about and approximations of processes that happen on small scales, such as cloud formation.

The CMIP6 models include more sophisticated treatments of ice, water and clouds than earlier ones did, including those in phase 5 (CMIP5). The latest models also include a wider variety of physical processes than before. As models become more realistic, they are expected to converge. In the meantime, individual improvements can affect how sensitive the models are to certain warming processes, in ways that are often impossible to predict.

The Intergovernmental Panel on Climate Change (IPCC), to its credit, has recognized this ‘hot model’ problem. Scientists contributing to the main sections of its Sixth Assessment Report (AR6; published over the past few months) reconciled the newest climate models with key observational constraints on global mean warming, sea-level rise and ocean heat content, and other analyses. They applied statistics to determine the most reasonable projections, consistent with many lines of evidence, which they call ‘assessed warming’.

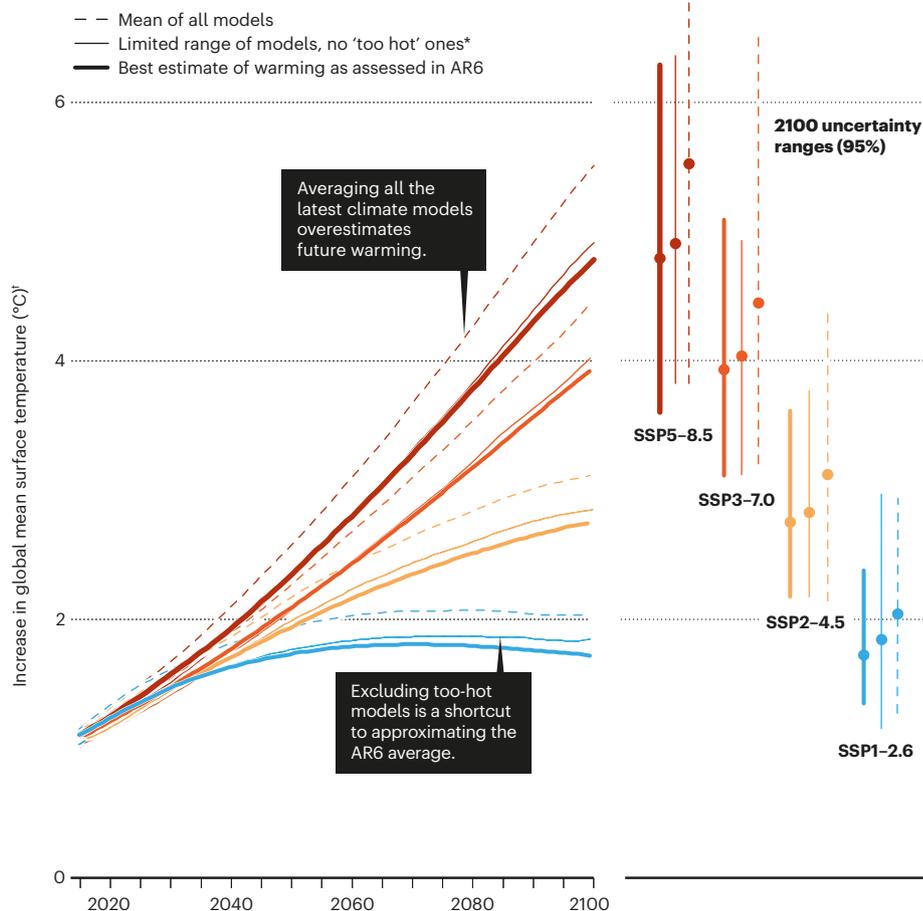
Unfortunately, little guidance was made available for scientists wishing to study projections in other contexts. We are concerned that in the absence of such guidance, much of the scientific literature is at risk of reporting projections that are inconsistent with the approach taken by the IPCC, and that are overly influenced by the hot models.

Studies that cover monthly or daily extremes or regional climate impacts, for example, are instead left to use the full set of CMIP6 models. And simply taking an average of those leads to higher projections of warming than the IPCC’s assessed-warming averages. As a result, some studies have reported projections that might be inconsistent with AR6 assessments. Findings that show projected climate change will be ‘worse than we thought’ are often attributable to the hot models in CMIP6.

It is important to emphasize that, whereas unduly hot outcomes might be unlikely, this does not mean that global warming is not a serious threat. Multiple lines of evidence establish that the planet is more than 1 °C warmer than it was before the Industrial Revolution, and that further warming poses severe risks to society and the natural world. There are many aspects of climate change we do not yet understand, hence the continued necessity of climate science. But there is no serious disagreement that continued emissions will lead to dangerous levels of warming.

CLIMATE MODELS: CHOICE MATTERS

The IPCC’s Sixth Assessment Report (AR6) assessed dozens of computer models to project global temperature change (four scenarios shown). Some of these projections were ‘too hot’ when compared with other lines of evidence for climate warming in response to carbon dioxide emissions⁸. Researchers using all these models without the AR6 statistical adjustments could end up overestimating future temperature change.



*Using the transient climate response (TCR) metric in the range 1.4–2.2 °C deemed as “likely” in AR6. (TCR is the amount of global warming in the year in which atmospheric CO₂ concentrations have doubled after having steadily increased by 1% each year.)

†Global mean surface temperatures are relative to a 1850–99 baseline.

IPCC, Intergovernmental Panel on Climate Change; SSP, Shared Socioeconomic Pathway.

The IPCC came up with a solution for global mean projections. Now researchers, communities and policymakers need more information. To inform better practice, we outline here what the IPCC has done differently in AR6, and offer some suggestions on how best to address these gaps.

Hot tail

The largest source of uncertainty in global temperatures 50 or 100 years from now is the volume of future greenhouse-gas emissions, which are largely under human control. However, even if we knew precisely what that volume would be, we would still not know exactly how warm the planet would get. This is because human-caused global warming is an

enormous experiment that has no precedent, and feedback processes, such as changes to cloud cover, will affect the pace and magnitude of warming.

To quantify the influence of these effects, climate modellers define standardized metrics. One is the transient climate response (TCR), or the amount of global warming in the year in which atmospheric CO₂ concentrations have finally doubled after having steadily increased by 1% every year. A second metric is equilibrium climate sensitivity (ECS), the eventual long-term temperature response to CO₂ concentrations that have doubled and remain doubled. The two metrics are distinct but related: ECS measures a long-term equilibrium climate response, whereas TCR



A conservationist helps to tackle a fire in Brazil's Pantanal region as it burns in September 2021.

measures a climate that has not yet had time to fully adjust². Models with a high TCR tend to have a high ECS⁴.

In previous generations of climate models in CMIP5, no model had an ECS of higher than 4.7 °C. In CMIP6, more than one-quarter of models have sensitivities that are greater than this, and around one-fifth show warming of at least 5 °C in response to a doubling of atmospheric CO₂ concentrations, according to our analysis. Numerous studies have found that these high-sensitivity models do a poor job of reproducing historical temperatures over time^{4–7} and in simulating the climates of the distant past⁸. Specifically, they often show no warming over the twentieth century and then a sharp warming spike in the past few decades³, and some simulate the last ice age as being much colder than palaeoclimate evidence indicates⁷.

At the time these new models were being developed, climate scientists were also trying to improve understanding of the range of climate sensitivity, and to narrow it. A

2020 community review (that four of us co-authored)⁸ combined lines of evidence from palaeoclimate, observations of surface temperatures and ocean heat content, and models of physical processes. It concluded⁸ that the ECS is likely (with a 66% chance) to be in the range of 2.6–3.9 °C, and very likely (with a 90% chance) to lie between 2.3 and 4.7 °C.

On the basis of that review and other recent findings, the AR6 authors decided to narrow the climate sensitivity they considered 'likely' to a similar range, of between 2.5 and 4 °C, and to a 'very likely' range of between 2 °C and 5 °C.

Beyond model democracy

The climate community has debated what to do about the hot models since results began to appear in 2019. Before then, the IPCC and many other assessments simply used the mean and spread of models to estimate impacts and their uncertainties. Such 'model democracy' assumed that each model is independent and equally valid. Other methods of combining model projections did not yield

results that were more consistent or credible⁹.

In AR6, such simple methods no longer work: the high-sensitivity models are not as equally valid as others for estimating global temperature. AR6 authors decided to apply weights to each model before averaging them, to produce 'assessed global warming' projections. Specifically, the AR6 report used various published statistical weighting methods^{4–6} to combine the projections of different climate models, giving more weight to those that agreed with historical temperature observations.

They also used a climate model 'emulator' – a simpler model requiring less computing power – that incorporated the latest estimates of the sensitivity of the climate to CO₂ emissions, based on lines of evidence beyond climate models. This approach provides a more realistic range of future warming projections, which are better constrained by observations than the raw CMIP6 model output, but are difficult for non-specialists to reproduce.

CARLDES SOUZA/AFP VIA GETTY



the previous model generation, caused largely by the inclusion of unrealistic high-sensitivity models.

What to do

The broad and diverse community studying climate change and its impacts urgently needs guidance on best practices for combining the outputs of multiple climate models. One key message: the multi-model mean and spread of the new ensemble (CMIP6) should not simply be used like the old one (CMIP5).

We suggest that climate researchers consider the following options.

First, follow the lead of the AR6 to base analyses on global warming levels rather than on time¹⁰. For example, instead of assessing changes in rainfall by the year 2100, research-

“Reporting that severe risks are projected to occur at a particular time can give a false sense of inevitability.”

ers could report changes at global warming levels of 1.5, 2, 3 and 4 °C. This has several advantages. It mirrors the policy discourse surrounding the Paris agreement targets of 1.5 °C and ‘well below 2 °C’. It is also largely independent of the choice of future emissions scenario – despite some differences related to the rate of warming and aerosol forcing, the world largely looks the same at 2 °C, no matter how we get there. And, to a certain extent, using global warming levels bypasses the need to select or weight CMIP6 models. Each model has something to offer at a given temperature, so the full CMIP6 ensemble can be used. The IPCC Working Group I Interactive Atlas is a good tool for calculating multi-model means at a particular level of global warming (see <https://interactive-atlas.ipcc.ch>).

Global warming levels force a simple question: when will the world reach a given level of warming? The answer, of course, is that it’s up to us. Reporting that severe risks and catastrophic outcomes are projected to occur at a particular time can give a false sense of inevitability and obscure the role of human choice in determining the future. In situations for which policymakers require information on timing, we suggest using AR6 assessed warming to map projections for global warming levels onto emissions scenarios, ensuring consistency between regional studies and AR6.

Second, if the warming trajectory – rather than just the global warming level – is important for a particular climate outcome, focus on the subset of CMIP6 models that is most consistent with AR6 assessed-warming projections. We recommend screening out models with a TCR that lies outside the ‘likely’

(66% likelihood range) of 1.4–2.2 °C. The AR6 assessed-warming constraints are correlated with the TCR, so this gives a good approximation to the assessed warming⁴. This approach allows for an assessment of regional changes over time. Alternatively, using a ‘likely’ 2.5–4 °C ECS screen also reproduces AR6 results well, although at the expense of discarding 60% of the models in the CMIP6 ensemble, compared with 40% in the TCR screened subset.

Third, pick models that are best suited to the task at hand. The problem is not that high sensitivity models exist, but rather that the preponderance of them in the CMIP6 ensemble biases the mean and uncertainty range upwards. If there is a real need to examine ‘hot tail’ risks – because there is still a more than 5% chance of ECS exceeding 5 °C (ref. 8) – use a high-sensitivity subset. Ask whether changes in average conditions or extreme events in the region of interest scale with the global mean temperature. In cases in which model spread is not clearly related to the spread in climate sensitivity, alternative metrics might be appropriate.

Using the latest generation of models in a way that is consistent with AR6 requires both an awareness of the problem, and easy-to-use alternatives such as those we highlight here.

The authors

Zeke Hausfather is climate research lead at Stripe, South San Francisco, California, USA. **Kate Marvel** is a research scientist at the Center for Climate Systems Research, Columbia University; and at the NASA Goddard Institute for Space Studies, New York, USA. **Gavin A. Schmidt** is director of the NASA Goddard Institute for Space Studies, New York, USA. **John W. Nielsen-Gammon** is a professor and Texas State Climatologist in the Department of Atmospheric Sciences, Texas A&M University, College Station, Texas, USA. **Mark Zelinka** is a staff scientist at Lawrence Livermore National Laboratory, Livermore, California, USA.
e-mail: hausfath@stripe.com

1. Eyring, V. et al. *Geosci. Model Dev.* **9**, 1937–1958 (2016).
2. Zelinka, M. D. et al. *Geophys. Res. Lett.* **47**, e2019GL085782 (2020).
3. Nijse, F. J. M. M., Cox, P. M. & Williamson, M. S. *Earth Syst. Dyn.* **11**, 737–750 (2020).
4. Tokarska, K. B. et al. *Sci. Adv.* **6**, eaaz9549 (2020).
5. Ribes, A., Qasmi, S. & Gillett, N. P. *Sci. Adv.* **7**, eabc0671 (2021).
6. Liang, Y., Gillett, N. P. & Monahan, A. H. *Geophys. Res. Lett.* **47**, e2019GL086757 (2020).
7. Zhu, J., Poulsen, C. J. & Otto-Bliesner, B. L. *Nature Clim. Change* **10**, 378–379 (2020).
8. Sherwood, S. C. et al. *Rev. Geophysics* **58**, e2019RG000678 (2020).
9. Knutti, R. *Clim. Change* **102**, 395–404 (2010).
10. James, R., Washington, R., Schleussner, C.-F., Rogelj, J. & Conway, D. *Wires Clim. Change* **8**, e457 (2017).

The authors declare no competing interests. Supplementary information accompanies this Comment online (see go.nature.com/3rwizsh).

The IPCC’s assessed-warming projections produce only annual average global changes. Researchers looking to study regional climate impacts, daily extremes or other climate variables have had to pick their own path. Many analysts have defaulted to the pre-AR6 approach of treating each model the same. This leads to exaggerated projections: global average surface temperatures in 2100 that are 0.2–0.7 °C warmer than those with AR6 assessed warming (see ‘Climate models: choice matters’; underlying data are available in Supplementary information). The assessed-warming projections, by contrast, are broadly consistent with those from CMIP5.

Results using the raw CMIP6 models are already entering the climate impacts literature. In our experience, few climate researchers outside those directly involved in the creation of models are aware of the assessed-warming approach taken in AR6. In recent months, we’ve seen numerous papers highlighting how much worse regional and global climate outcomes are in CMIP6 than in