Frequently Asked Questions

FAQ 9.1 | Are Climate Models Getting Better, and How Would We Know?

Climate models are extremely sophisticated computer programs that encapsulate our understanding of the climate system and simulate, with as much fidelity as currently feasible, the complex interactions between the atmosphere, ocean, land surface, snow and ice, the global ecosystem and a variety of chemical and biological processes.

The complexity of climate models—the representation of physical processes like clouds, land surface interactions and the representation of the global carbon and sulphur cycles in many models—has increased substantially since the IPCC First Assessment Report in 1990, so in that sense, current Earth System Models are vastly 'better' than the models of that era. This development has continued since the Fourth Assessment, while other factors have also contributed to model improvement. More powerful supercomputers allow current models to resolve finer spatial detail. Today's models also reflect improved understanding of how climate processes work—understanding that has come from ongoing research and analysis, along with new and improved observations.

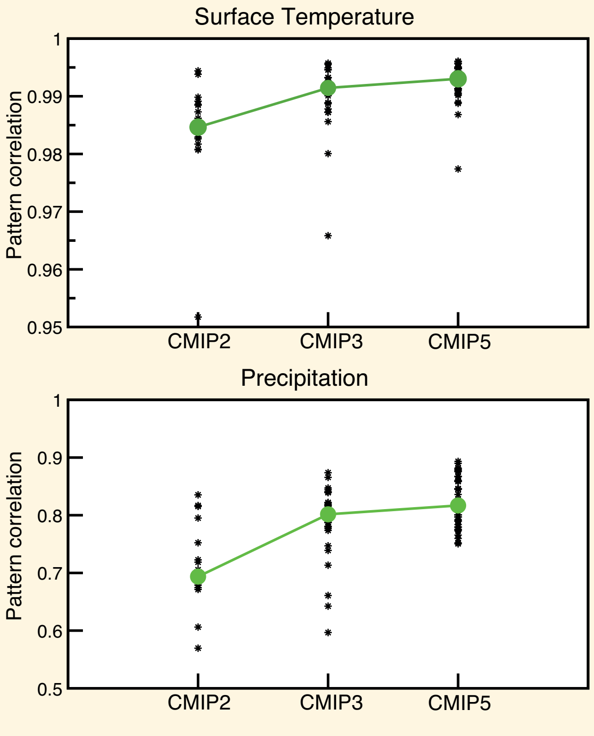
Climate models of today are, in principle, better than their predecessors. However, every bit of added complexity, while intended to improve some aspect of simulated climate, also introduces new sources of possible error (e.g., via uncertain parameters) and new interactions between model components that may, if only temporarily, degrade a model's simulation of other aspects of the climate system. Furthermore, despite the progress that has been made, scientific uncertainty regarding the details of many processes remains.

An important consideration is that model performance can be evaluated only relative to past observations, taking into account natural internal variability. To have confidence in the future projections of such models, historical climate—and its variability and change—must be well simulated. The scope of model evaluation, in terms of the kind and quantity of observations available, the availability of better coordinated model experiments, and the expanded use of various performance metrics, has provided much more quantitative information about model performance. But this alone may not be sufficient. Whereas weather and seasonal climate predictions can be regularly verified, climate projections spanning a century or more cannot. This is particularly the case as anthropogenic forcing is driving the climate system toward conditions not previously observed in the instrumental record, and it will always be a limitation.

Quantifying model performance is a topic that has featured in all previous IPCC Working Group I Reports. Reading back over these earlier assessments provides a general sense of the improvements that have been made. Past reports have typically provided a rather broad survey of model performance, showing differences between model-calculated versions of various climate quantities and corresponding observational estimates.

Inevitably, some models perform better than others for certain climate variables, but no individual model clearly emerges as 'the best' overall. Recently, there has been progress in computing various performance metrics, which synthesize model performance relative to a range of different observations according to a simple numerical score. Of course, the definition of such a score, how it is computed, the observations used (which have their

(continued on next page)



FAQ 9.1, Figure 1 | Model capability in simulating annual mean temperature and precipitation patterns as illustrated by results of three recent phases of the Coupled Model Intercomparison Project (CMIP2, models from about year 2000; CMIP3, models from about 2005; and CMIP5, the current generation of models). The figure shows the correlation (a measure of pattern similarity) between observed and modelled temperature (upper panel) and precipitation (lower panel). Larger values indicate better correspondence between modelled and observed spatial patterns. The black symbols indicate correlation coefficient for individual models, and the large green symbols indicate the median value (i.e., half of the model results lie above and the other half below this value). Improvement in model performance is evident by the increase in correlation for successive model generations.

FAQ 9.1 (continued)

own uncertainties), and the manner in which various scores are combined are all important, and will affect the end result.

Nevertheless, if the metric is computed consistently, one can compare different generations of models. Results of such comparisons generally show that, although each generation exhibits a range in performance, the average model performance index has improved steadily between each generation. An example of changes in model performance over time is shown in FAQ 9.1, Figure 1, and illustrates the ongoing, albeit modest, improvement. It is interesting to note that both the poorest and best performing models demonstrate improvement, and that this improvement comes in parallel with increasing model complexity and an elimination of artificial adjustments to atmosphere and ocean coupling (so-called 'flux adjustment'). Some of the reasons for this improvement include increased understanding of various climate processes and better representation of these processes in climate models. More comprehensive Earth observations are also driving improvements.

So, yes, climate models are getting better, and we can demonstrate this with quantitative performance metrics based on historical observations. Although future climate projections cannot be directly evaluated, climate models are based, to a large extent, on verifiable physical principles and are able to reproduce many important aspects of past response to external forcing. In this way, they provide a scientifically sound preview of the climate response to different scenarios of anthropogenic forcing.

9.8.2 Implications of Model Evaluation for Climate Change Detection and Attribution

The evaluation of model simulations of historical climate is of direct relevance to detection and attribution (D&A) studies (Chapter 10) since these rely on model-derived patterns (or 'fingerprints') of climate response to external forcing, and on the ability of models to simulate decadal and longer-time scale internal variability (Hegerl and Zwiers, 2011). Conversely, D&A research contributes to model evaluation through estimation of the amplitude of modeled response to various forcings (Section 10.3.1.1.3). The estimated fingerprint for some variables such as water vapor is governed by basic physical processes that are well represented in models and are rather insensitive to model uncertainties (Santer et al., 2009). Figure 9.44 illustrates slight improvements in the representation of some of the modes of variability and climate phenomena discussed in Sections 9.5.2 and 9.5.3, suggesting with *medium confidence* that models now better reproduce internal variability. On the other hand, biases that affect D&A studies remain. An example is the warm bias of lower-stratosphere temperature trends during the satellite period (Section 9.4.1.4.5) that can be linked to uncertainties in stratospheric ozone forcing (Solomon et al., 2012; Santer et al., 2013). Recent studies of climate extremes (Section 9.5.4) also provide evidence that models have reasonable skill in these important attributes of a changing climate; however, there is an indication that models have difficulties in reproducing the right balance between historical changes in cold and warm extremes. They also confirm that resolution affects the confidence that can be placed in the analyses of extreme in precipitation. D&A studies focussed on extreme events are therefore constrained by current model limitations. Lastly, some D&A studies have incorporated model quality results by repeating a multi-model analysis with only the models that agree best with observations (Santer et al., 2009). This model discrimination or weighting is less problematic for D&A analysis than it is for model projections of future climate (Section 9.8.3), because D&A research is focussed on historical and control-run simulations which can be directly evaluated against observations.

9.8.3 Implications of Model Evaluation for Model Projections of Future Climate

Confidence in climate model projections is based on physical understanding of the climate system and its representation in climate models, and on a demonstration of how well models represent a wide range of processes and climate characteristics on various spatial and temporal scales (Knutti et al., 2010b). A climate model's credibility is increased if the model is able to simulate past variations in climate, such as trends over the 20th century and palaeoclimatic changes. Projections from previous IPCC assessments can also be directly compared to observations (see Figures 1.4 and 1.5), with the caveat that these projections were not intended to be predictions over the short time scales for which observations are available to date. Unlike shorter lead forecasts, longer-term climate change projections push models into conditions outside the range observed in the historical period used for evaluation.

In some cases, the spread in climate projections can be reduced by weighting of models according to their ability to reproduce past observed climate. Several studies have explored the use of unequally weighted means, with the weights based on the models' performance in simulating past variations in climate, typically using some performance metric or collection of metrics (Connolley and Bracegirdle, 2007; Murphy et al., 2007; Waugh and Eyring, 2008; Pierce et al., 2009; Reifen and Toumi, 2009; Christensen et al., 2010; Knutti et al., 2010b; Raisanen et al., 2010; Abe et al., 2011; Shiogama et al., 2011; Watterson and Whetton, 2011; Tsushima et al., 2013). When applied to projections of Arctic sea ice, averages in which extra weight is given