nature climate change

LETTERS

PUBLISHED ONLINE: 28 OCTOBER 2012 | DOI: 10.1038/NCLIMATE1716

# Robustness and uncertainties in the new CMIP5 climate model projections

Reto Knutti* and Jan Sedláček

**Estimates of impacts from anthropogenic climate change rely on projections from climate models. Uncertainties in those have often been a limiting factor, in particular on local scales. A new generation of more complex models running scenarios for the upcoming Intergovernmental Panel on Climate Change Fifth Assessment Report (IPCC AR5) is widely, and perhaps naively, expected to provide more detailed and more certain projections. Here we show that projected global temperature change from the new models is remarkably similar to that from those used in IPCC AR4 after accounting for the different underlying scenarios. The spatial patterns of temperature and precipitation change are also very consistent. Interestingly, the local model spread has not changed much despite substantial model development and a massive increase in computational capacity. Part of this model spread is irreducible owing to internal variability in the climate system, yet there is also uncertainty from model differences that can potentially be eliminated. We argue that defining progress in climate modelling in terms of narrowing uncertainties is too limited. Models improve, representing more processes in greater detail. This implies greater confidence in their projections, but convergence may remain slow. The uncertainties should not stop decisions being made.**

Coordinated experiments, in which many climate models run a set of scenarios, have become the *de facto* standard to produce climate projections[1]. Those multi-model ensembles sample uncertainties in emission scenarios, model uncertainty and initial condition uncertainty, and provide a basis to estimate projection uncertainties[2–6]. The Coupled Model Intercomparison Project Phase 5 (CMIP5; ref. 7), coordinated by the World Climate Research Programme in support of the IPCC AR5, is the most recent of these activities, and builds on CMIP3. The efforts for CMIP5 are enormous, with a larger number of more complex models run at higher resolution, with more complete representations of external forcings, more types of scenario and more diagnostics stored. Here we perform a first comparison between projections from CMIP3 and CMIP5 and test whether the new models converge in their projections.

The change in global mean temperature over the twentieth and twenty-first century as simulated by the CMIP3 and CMIP5 models is shown in Fig. 1. The simulated twentieth-century warming is less gradual in the CMIP5 model mean, because radiative forcings are included more completely. In contrast, some CMIP3 models did not include solar and volcanic forcings, and indirect aerosol effects. Interannual variations are large in each model but suppressed in the model mean. The range of warming for the twenty-first century is not straightforward to compare, because CMIP3 used the Special Report on Emissions Scenarios (SRES) B1, A1B and A2 scenarios[8], whereas CMIP5 uses the new Representative Concentration Pathways[9] (RCPs). The overall range across the RCP scenarios is larger because for the first time a low-emission mitigation scenario is included. That does not imply a larger uncertainty in climate change, but is simply a choice of economic scenarios. The model spread relative to the model mean change for a given scenario is similar or in some cases slightly larger, implying that the models have not converged in their projections. This is consistent with the fact that the mean and range of both the climate sensitivity and transient climate response of the CMIP5 models are similar to CMIP3 (ref. 10), and consistent with the fact that climate sensitivity has been notoriously difficult to constrain[11,12] and projection uncertainties have been similar over successive IPCC reports.

The lack of a common scenario makes a direct CMIP3–CMIP5 comparison difficult. The box plots in Fig. 1 provide the best alternative by comparing the CMIP5 RCP mean and spread to that predicted by the energy balance model MAGICC for the RCPs but with model parameters calibrated to the older CMIP3 models[13,14] (see Methods). We observe a higher warming in CMIP5 than predicted by MAGICC, and larger spread for the lower scenarios, but an interpretation of that seems premature given the lack of low-emission scenarios in CMIP3 and the potential uncertainty implied in using a simple model to emulate RCP2.6. Our results are consistent with a recent probabilistic study for the SRES and RCP scenarios[15]. It is important to note that model spread is not necessarily a good estimate of uncertainty, because the distribution of models in the CMIP ensemble of opportunity is rather arbitrary and affected by interdependencies across models[2–5,16,17]. Other methods to quantify uncertainties in global temperature on the basis of observational constraints often yield larger uncertainties than those in CMIP (ref. 18). With all those caveats, CMIP5 projections seem largely consistent with CMIP3. There is, despite better process understanding, little evidence from CMIP5 that our ability to constrain the large-scale climate feedbacks has improved significantly. Differences in global temperature projections are largely attributable to different scenarios[15].

Model mean patterns of temperature and precipitation change are shown in Figs 2 and 3, respectively, for two seasons and two time periods. They also are remarkably similar in CMIP3 and CMIP5, indicating that the large-scale features of climate change are robust towards resolution and assumptions, to the extent that they are sampled in today's models. We argue that this robustness across generations of models is positive, and its consistency with simpler models, theoretical process understanding and observed changes provides strong support for the argument that climate change over the twenty-first century will probably exceed that observed over the past century[19], even for the RCP2.6 scenario in which global greenhouse-gas emissions are reduced by about 90% in 2100 compared with the present.

To show model agreement locally we use stippling in the maps, based on a new robustness measure *R*. This quantity is inspired by the signal-to-variability ratio and the ranked probability skill

Institute for Atmospheric and Climate Science, ETH Zurich, CH-8092 Zurich, Switzerland. *e-mail: reto.knutti@env.ethz.ch.

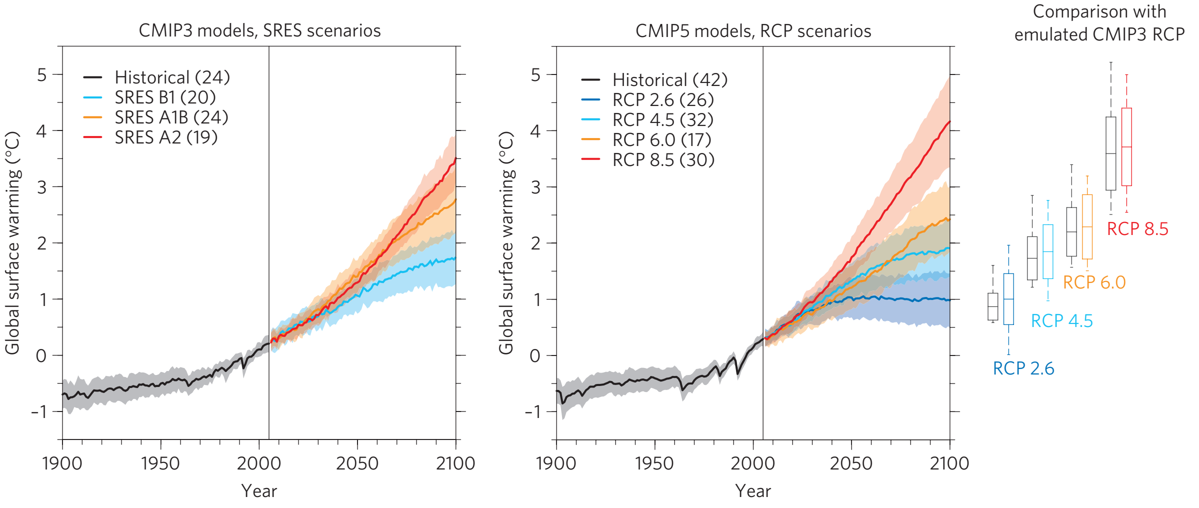


**Figure 1 | Global temperature change and uncertainty.** Global temperature change (mean and one standard deviation as shading) relative to 1986–2005 for the SRES scenarios run by CMIP3 and the RCP scenarios run by CMIP5. The number of models is given in brackets. The box plots (mean, one standard deviation, and minimum to maximum range) are given for 2080–2099 for CMIP5 (colours) and for the MAGICC model calibrated to 19 CMIP3 models (black), both running the RCP scenarios.

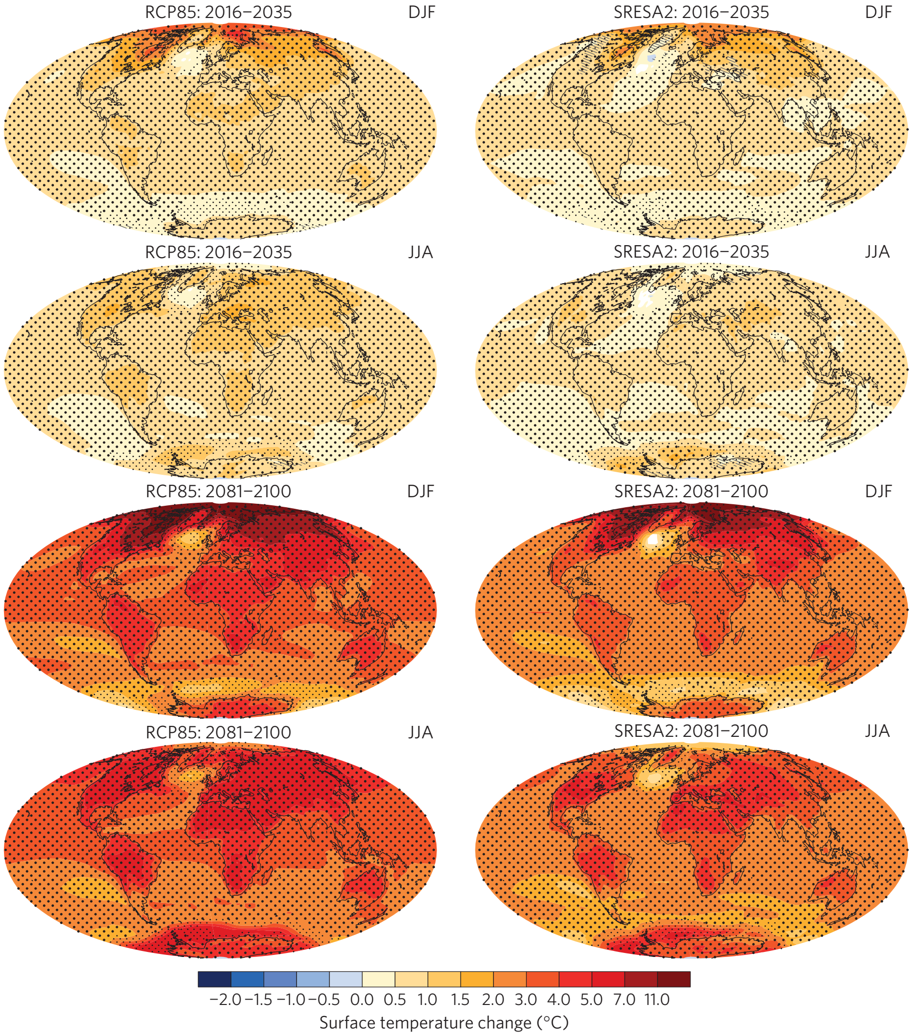


**Figure 2 | Patterns of surface warming.** Multi-model mean surface warming for two seasons (December–February, DJF, and June–August, JJA) and two 20-year time periods centred around 2025 and 2090, relative to 1986–2005, for CMIP5 (left) and CMIP3 (right). Stippling marks high robustness, hatching marks no significant change and white areas mark inconsistent model responses (see Methods and Supplementary Figs S2 and S3).

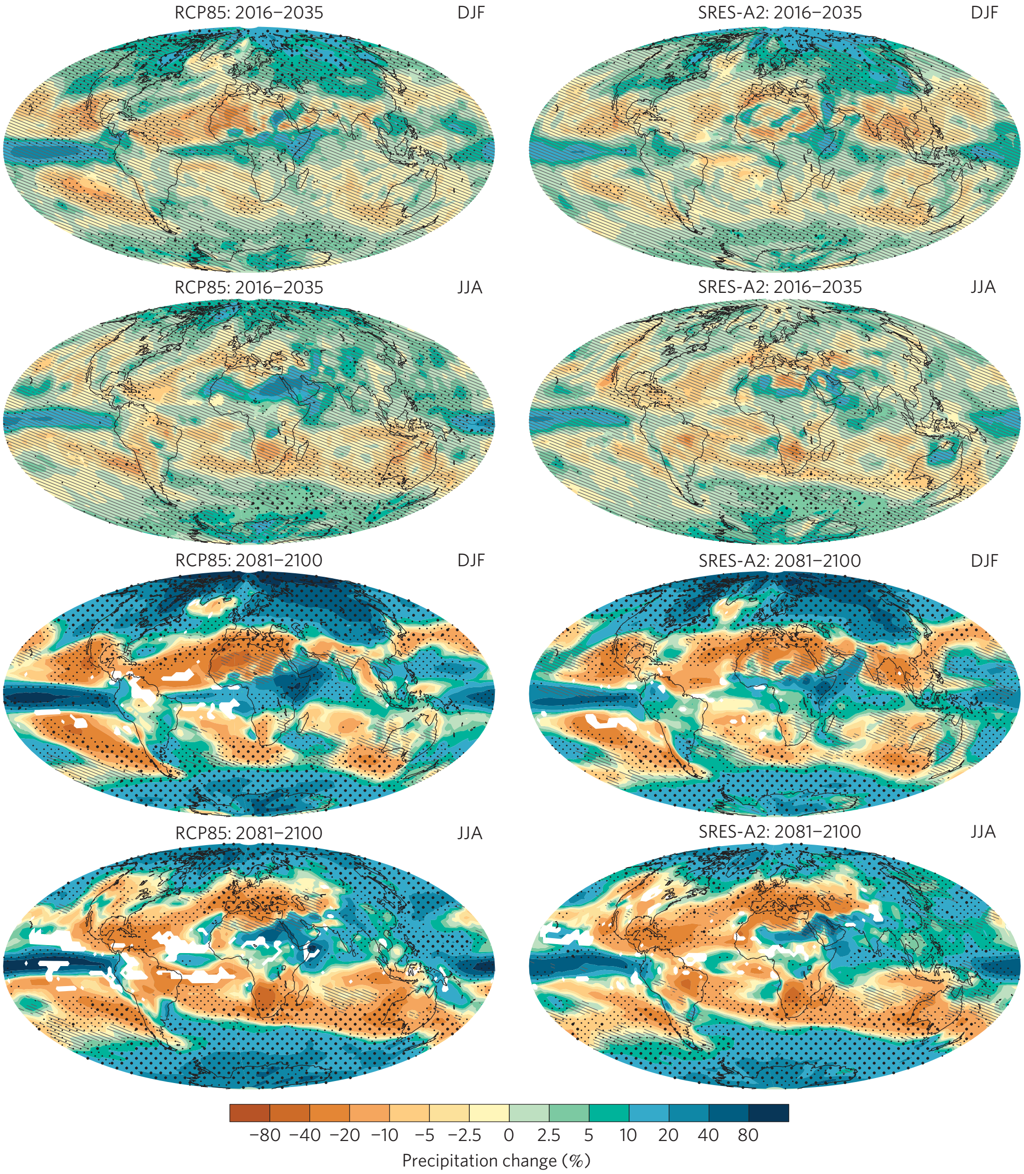


**Figure 3 | Patterns of precipitation change.** Multi-model mean relative precipitation change for two seasons (December–February, DJF, and June–August, JJA) and two 20-year time periods centred around 2025 and 2090, relative to 1986–2005, for CMIP5 (left) and CMIP3 (right). Stippling marks high robustness, hatching marks no significant change and white areas mark inconsistent model responses (see Methods and Supplementary Figs S2 and S3).

score used in weather prediction (see Methods). In contrast to other criteria[19–21], it considers the magnitude of change, the sign, natural variability and inter-model spread. The main conclusions are similar if other methods are used to measure model agreement[20,21]. Small and large dots indicate good and very good agreement between models, respectively (see Methods). Hatching marks areas where at least 80% of the models show no significant change, information that is often not highlighted yet clearly relevant for impacts and adaptations. A significant warming with high model agreement is evident already for a projection centred around 2025. Regions where most models show significant changes but do not agree well (robustness $R < 0.5$) are masked as white. Even for precipitation, the extent of those is limited, as pointed out recently[20,22]. The area of the Earth where the robustness $R$ exceeds 0.8 (fine stippling) for precipitation change is depicted in Fig. 4a (black lines). The area fraction with robust projections is increasing with global temperature as the precipitation signal emerges, but levels off at about 3 °C, where the signal further strengthens, but model differences also become pronounced. There are also large areas with no significant precipitation change (that is, 50% of the globe in boreal winter for 2 °C warming)[20,23].

Whereas the similarity of the projected precipitation change in CMIP3 and CMIP5 is reassuring, the similarity of the measure of robustness is more troublesome. The stippled area in CMIP3 and CMIP5 is nearly identical, implying little increase in model agreement in CMIP5 for precipitation changes. The corresponding results for RCP4.5 and SRES B1 are similar. Robustness over land is slightly higher but also similar in CMIP3 and CMIP5 (Fig. 4c). There are several hypotheses that potentially explain the lack of convergence and associated reduction of uncertainty. There could be (1) inherent limitations in the way models are built given limited computational resources and spatial resolution, (2) lack of process understanding, (3) lack of accurate long term observations to constrain models, (4) lack of consensus on metrics of present-day model performance that clearly separate better from worse models in terms of projection quality, (5) inherent limitation of climate change not being predictable owing to internal variability, (6) addition of dissimilar models from institutions new in CMIP5

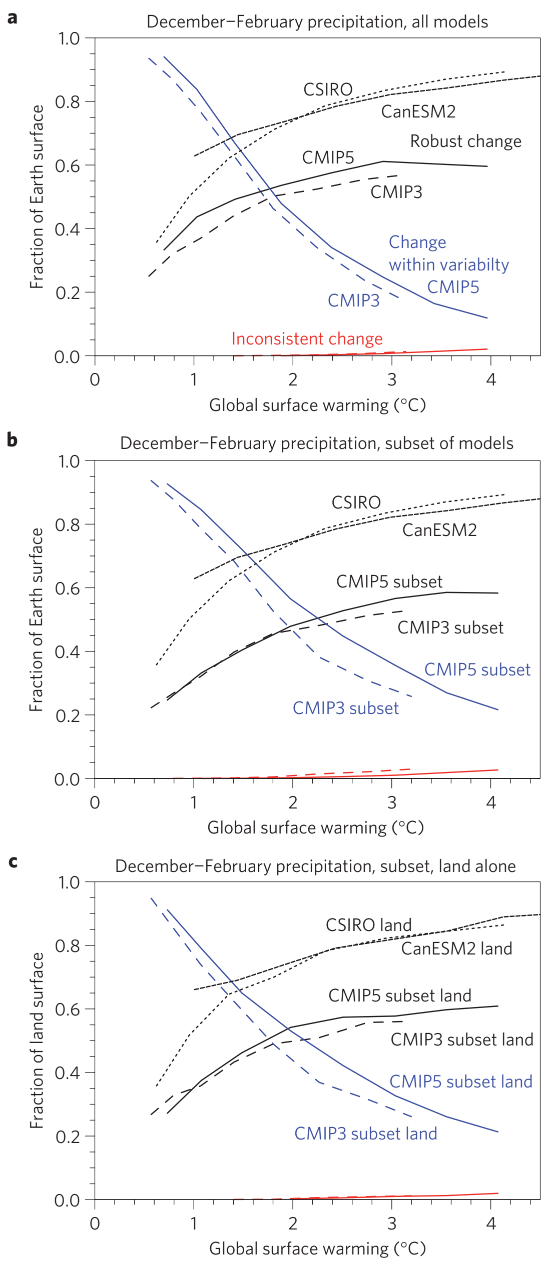


**Figure 4 | Model robustness for precipitation. a**, Fraction of the Earth surface with high robustness (black, $R > 0.8$), no significant change (blue) and inconsistent model responses (red) illustrated for December–February precipitation change for CMIP3 and CMIP5. Fractions are shown as a function of global temperature change rather than time, which largely eliminates the differences in scenarios. Maximum achievable robustness is calculated for CSIRO and CanESM2 initial condition ensembles (see the main text and Methods). **b**, The same as for **a** but for a subset of 11 models from 11 institutions participating in both CMIP3 and CMIP5 (see Supplementary Table S1). **c**, The same as for **b** but for land alone. Results for June–August look very similar.

and (7) addition of new processes, components, or forcings in CMIP5 that are not well understood, not well represented in the model, or not well constrained by observations. To pin down the contributions of each of these is difficult at this stage. We test hypothesis (6) by analysing a subset of 11 models from 11 institutions who participated in both CMIPs. Although CMIP5 has slightly higher robustness for all models (Fig. 4a), that becomes negligible for the subset (Fig. 4b). We find that new institutions contributing to CMIP5 are not to blame for the lack of convergence, and that nearly identical models from some institutions in CMIP5 are artificially (and unjustifiably) increasing the robustness. In our view the lack of consensus on relevant model performance metrics (4) is important, and not surprising given that models have to be evaluated not on their future projection but the present and past simulation[2,3,5,17]. The extra model complexity in CMIP5 (7) is likely to be an important factor. In contrast to end users, who would define model quality on the basis of prediction accuracy, climate model developers often judge their models to be better if the processes are represented in more detail. Thus, the new models are likely to be better in the sense of being physically more plausible, but it is difficult to quantify the impact of that on projection accuracy, in contrast to weather prediction where forecast skill is readily estimated through verification. To quantify the contribution of internal variability (5), we recalculated the same robustness measure for precipitation in Fig. 4 separately for two initial condition ensembles (the Commonwealth Scientific and Industrial Research Organisation (CSIRO) with 10 members; the Second Generation Canadian Earth System Model (CanESM2) with 5 members). Those demonstrate the inherent limits to maximum achievable robustness if model uncertainty was eliminated. For a 1 °C warming, only about half of the Earth surface could be stippled even if model uncertainty was eliminated entirely. Still, the area with high robustness ($R > 0.8$) is only about 60–80% of what it could potentially be if models converged, so significant progress is possible in principle. The comparison of robustness from initial condition runs and CMIP5 shows that potential improvements in precipitation projections are largest in the tropics (Central America and Africa, Middle East, Southeast Asia and Australia; see Supplementary Fig. S1). For temperature, the potentially achievable $R$ is always close to one, and the regions with the greatest potential improvement are the North Atlantic, the areas with or close to sea ice, the Southern Ocean and Southeast Asia.

The area where changes do not emerge from variability locally (hatching) and with inconsistent model responses (white) for precipitation is also shown in Fig. 4. The former is larger in CMIP5 owing to a slightly increased magnitude of internal variability in CMIP5; the latter is very small (a few per cent). Temperature changes emerge from variability much more quickly and with high robustness, consistent with earlier studies[6,23,24].

From the lack of reduction in model spread one might conclude that the models have not changed, but we believe this argument is not justified. It is clear that large efforts were made to include more complex and comprehensive representations of the processes in many models, and to build different model versions (for example, for decadal prediction, or including chemistry, or a carbon cycle). We estimate that, contrary to popular belief, a larger fraction of the increase in computational capacity is used to build more complete models rather than simply increasing spatial resolution (see Methods).

How should we interpret the lack of model convergence? Can we be more confident in a projection even if the uncertainty is unchanged? We believe this can be the case if more model data, observations and process understanding are available. It is common that more research uncovers a picture that is more complicated; thus, uncertainty can grow with time. Climate models in CMIP5 are better in the sense that they represent more of the relevant processes in more detail. Even though the model spread in CMIP3 and CMIP5 projections is similar, model developers have incorporated some of the unknown unknowns, that is bold assumptions or previously ignored factors, into the projection, so we are more confident that the models capture most of the relevant processes. Recent strategic documents suggested that a massive international effort into model development and high-performance computing could strongly increase the value of climate predictions for impacts[25]. Judging the potential success of such a project is speculative, and it may simply take a long time to succeed. However, if the past is a guide to the future then uncertainties in climate change are unlikely to decrease quickly, and may even grow temporarily. We have illustrated this for seasonal temperature and precipitation, and it is likely that impact-relevant predictions, for example of extreme weather events, may be even harder to improve. Progress in climate modelling is essential and will continue when efforts are maintained. It must not be measured only in how quickly model spread in projections decreases, but should also consider how adequate models are for specific purposes. Some uncertainties will remain, but these should not prevent those working on climate impacts, mitigation and adaptation from making decisions[26].

## Methods

The comparison between CMIP3 and CMIP5 in Fig. 1 is based on the actual CMIP5 RCP data and the simple climate model MAGICC calibrated to 19 CMIP3 models[13,14] and then run for the RCP scenarios. Those projections were made before CMIP5 data were available. For global temperature, MAGICC provides a sufficiently accurate prediction of what CMIP3 would have given for the RCPs if they had run them. Although the calibration has some uncertainty for individual models, it is very small for the model mean. Differences can however arise from the fact that CMIP3 forcings were not fully documented. In addition, MAGICC assumes a zero mean volcanic forcing in the past and future, whereas many GCMs assume volcanic forcing can only be negative. The difference between these assumptions could explain up to 0.2 °C between MAGICC and CMIP5.

The robustness $R$ used here is inspired by the ranked probability skill score[27] used in weather prediction, and by the ratio of model spread to the predicted change (noise to signal). It is defined as $R = 1 - A_1/A_2$, where $A_1$ is defined as the integral of the squared area between two cumulative density functions characterizing the individual model projections and the multi-model mean projection and $A_2$ is the integral of the squared area between two cumulative density functions characterizing the multi-model projection and the historical climate (see Supplementary Fig. S2). A value of $R$ equal to one implies perfect model agreement. Higher model spread or smaller signal decreases the value of $R$. It can therefore be interpreted as a measure of relative agreement. However, importantly, it considers the width of the initial distributions (such that the same projection has a lower $R$ value if the variability is small). In contrast to other methods[19,20], it also penalizes the case when models agree on the sign but disagree on the magnitude. A high value for $R$ is possible even if the models do not change in their mean but agree on a change in shape of the distribution of variability.

For the maps, colour contours show the multi-model average. Each model is given the same weight, and the first available initial condition ensemble member is used for each model (see Supplementary Table S1). Light stippling marks $R > 0.8$ (good agreement); strong stippling marks $R > 0.95$ (very good agreement). Hatching marks areas where at least 80% of models indicate no significant change. White indicates areas where at least half of the models show significant changes but $R < 0.5$, that is significant changes but little agreement among models (see Supplementary Fig. S3 for illustration). The maps therefore convey high model agreement (stippling) and no significant change (hatching) separately. Note that the thresholds chosen for $R$ and the fraction of models without significant change are subjective and for illustration alone, and may need to be different for different impact applications. Higher threshold implies higher confidence for a certain response, or the lack thereof. The main conclusions are similar for other thresholds for the robustness $R$.

The capacity of the fastest computers has increased annually by about a factor of 1.8, implying about a factor of 60 in computational capacity between CMIP3 and CMIP5, or a bit less if the fraction of computing for research has decreased. On average, the total number of grid cells in the atmosphere has approximately doubled from CMIP3 to CMIP5. Ocean resolution has also improved in some models. About a factor of 2–4 in computing could therefore have gone into model resolution. The number of simulated years has probably also increased by a factor of 2–3 owing to more scenarios in CMIP5. This leaves a factor of roughly 3–10 of computing that has gone into model complexity and the development of different model versions, that is at least as much as went into higher resolution.

Received 14 May 2012; accepted 13 September 2012; published online 28 October 2012

## References

1. Meehl, G. A. *et al.* The WCRP CMIP3 multimodel dataset—A new era in climate change research. *Bull. Am. Meteorol. Soc.* **88,** 1383–1394 (2007).
2. Tebaldi, C. & Knutti, R. The use of the multi-model ensemble in probabilistic climate projections. *Phil. Tran. R. Soc. A* **365,** 2053–2075 (2007).
3. Knutti, R., Furrer, R., Tebaldi, C., Cermak, J. & Meehl, G. A. Challenges in combining projections from multiple climate models. *J. Clim.* **23,** 2739–2758 (2010).
4. Knutti, R. The end of model democracy? *Climatic Change* **102,** 395–404 (2010).
5. Knutti, R. *et al. Good Practice Guidance Paper on Assessing and Combining Multi Model Climate Projections* (IPCC Working Group I Technical Support Unit, University of Bern, 2010).
6. Hawkins, E. & Sutton, R. The potential to narrow uncertainty in regional climate predictions. *Bull. Am. Meteorol. Soc.* **90,** 1095–1107 (2009).
7. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. A Summary of the CMIP5 Experiment Design. *Bull. Am. Meteorol. Soc.* **93,** 485–498 (2012).
8. IPCC *Special Report on Emissions Scenarios* (eds Nakićenović, N. & Swart, R.) (Cambridge Univ. Press, 2000).
9. Meinshausen, M. *et al.* The RCP Greenhouse Gas Concentrations and their extensions from 1765 to 2300. *Climatic Change* **109,** 213–241 (2011).
10. Andrews, T., Gregory, J. M., Webb, M. J. & Taylor, K. Forcing, feedbacks and climate sensitivity in CMIP5 atmosphere–ocean climate models. *Geophys. Res. Lett.* **39,** L09712 (2012).
11. Charney, J. G. *et al. Carbon Dioxide and Climate: A Scientific Assessment* (National Academy of Sciences, 1979).
12. Knutti, R. & Hegerl, G. C. The equilibrium sensitivity of the Earth's temperature to radiation changes. *Nature Geosci.* **1,** 735–743 (2008).
13. Meinshausen, M., Wigley, T. & Raper, S. Emulating atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6-Part 2: Applications. *Atmos. Chem. Phys.* **11,** 1457–1471 (2011).
14. Meinshausen, M., Raper, S. & Wigley, T. Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6-Part 1: Model description and calibration. *Atmos. Chem. Phys.* **11,** 1417–1456 (2011).
15. Rogelj, J., Meinshausen, M. & Knutti, R. Global warming under old and new scenarios using IPCC climate sensitivity range estimates. *Nature Clim. Change* **2,** 248–253 (2012).
16. Masson, D. & Knutti, R. Climate model genealogy. *Geophys. Res. Lett.* **38,** L08703 (2011).
17. Knutti, R. Should we believe model predictions of future climate change? *Phil. Trans. R. Soc. A* **366,** 4647–4664 (2008).
18. Knutti, R. *et al.* A review of uncertainties in global temperature projections over the twenty-first century. *J. Clim.* **21,** 2651–2663 (2008).
19. IPCC *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. *et al.*) (Cambridge Univ. Press, 2007).
20. Tebaldi, C., Arblaster, J. M. & Knutti, R. Mapping model agreement on future climate projections. *Geophys. Res. Lett.* **38,** L23701 (2011).
21. Power, S., Delage, F., Colman, R. & Moise, A. Consensus on twenty-first-century rainfall projections in climate models more widespread than previously thought. *J. Clim.* **25,** 3792–3809 (2012).
22. Schaller, N., Mahlstein, I., Cermak, J. & Knutti, R. Analyzing precipitation projections: A comparison of different approaches to climate model evaluation. *J. Geophys. Res.* **116,** D10118 (2011).
23. Mahlstein, I., Portmann, R., Daniel, J., Solomon, S. & Knutti, R. Perceptible changes in regional precipitation in a future climate. *Geophys. Res. Lett.* **39,** L05701 (2012).
24. Mahlstein, I., Knutti, R., Solomon, S. & Portmann, R. Early onset of significant local warming in low latitude countries. *Environ. Res. Lett.* **6,** 034009 (2011).
25. Shukla, J. *et al.* Revolution in climate prediction is both necessary and possible: A Declaration at the World Modelling Summit for Climate Prediction. *Bull. Am. Meteorol. Soc.* **90,** 175–178 (2009).
26. Dessai, S., Hulme, M., Lempert, R. & Pielke, R. Do we need better predictions to adapt to a changing climate? *Eos* **90,** 111–112 (2009).
27. Weigel, A., Liniger, M. & Appenzeller, C. The discrete Brier and ranked probability skill scores. *Mon. Weath. Rev.* **135,** 118–124 (2007).

## Acknowledgements

We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups for producing and making available their model output. For CMIP the US Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. We thank U. Beyerle and T. Corti for making the CMIP5 data accessible at ETH Zurich, and M. Meinshausen for the MAGICC projections.

## Author contributions

Both authors designed the study and wrote the paper. J.S. performed the CMIP5 analysis.

## Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.K.

## Competing financial interests

The authors declare no competing financial interests.