Quantifying Future Climate Change

Matthew Collins[1], Richard E. Chandler[2], Peter M. Cox[1], John M. Huthnance[3], Jonathan Rougier[4], David B. Stephenson[5]

[1]College of Engineering, Mathematics and Physical Sciences, University of Exeter UK
[2]Department of Statistical Science, University College London, UK
[3]National Oceanography Centre, Liverpool, UK
[4]Department of Mathematics, University of Bristol, UK

Corresponding Author: M.Collins@exeter.ac.uk

Quantitative projections of future climate are in increasing demand from the scientific community, policy makers and other stakeholders. Climate models of varying complexity are used to make projections but approximations and inadequacies or "errors" in models mean that those projections are uncertain, sometimes exploring a very wide range of possible futures. Techniques to quantify the uncertainties are described here in terms of a common framework whereby models are used to explore relationships between past climate and climate change and future projections. Model parameters may be varied to produce a range of different simulations of past climate that are then compared with observations using "metrics". If the model parameters can be constrained to a tighter range as a result of observational comparisons, projections can also be constrained to a tighter range. The strengths and weakness of different implementations are discussed.

Projections of climate change are made using climate models forced by scenarios of increasing greenhouse gases and other factors which impact on the energy balance of the climate system. The term 'projection' is used to imply a conditional dependence of a climate prediction on emission scenario, as such scenarios are derived from studies which consider multiple socio-economic factors but do not consider the relative likelihood of different pathways. Climate science in general is starting to become more quantitative, for example in attributing changes in the risk of certain weather or climate events[1] and there is a desire to be more quantitative about projections, particularly when those projections feed into assessments of the impacts of climate change[2]. Recent national assessments of climate change have moved from being qualitative to being much more quantitative, with dedicated web sites serving data to stakeholders[3] to inform decision making. Projections should be made on the basis of robust science but should also account for the uncertainties that arise because of incomplete understanding of climate change and because of limitations in models and observations.

Climate models are approximations – albeit often highly informed and sophisticated – of the real climate system and different models produce different projections of future climate change. By quantifying the uncertainty in projections, we should gain a more in-depth understanding of climate models and of the climate system and a better appreciation of the limitations of current understanding. Such an appreciation is required to also show where quantitative information cannot be provided and where science and policy should proceed more qualitatively. Uncertainty quantification also provides a benchmark so that we can measure progress and hopefully reduce uncertainties.

Much effort has been expended by climate modeling groups worldwide to coordinate simulations with the most complex climate models, to collect the outputs and make them easily available to the scientific community[4]. The third incarnation of the Coupled Model Intercomparison Projection (CMIP3) 'multi-model ensemble' or MME has been widely interrogated, resulting in an unprecedented level of scrutiny of complex climate models and their projections. The CMIP5 database of new simulations is now being populated. The quantitative interpretation of projections from a MME is extremely challenging. Reviews[5,6] highlight several techniques that have been proposed which must deal with the generic problem of trying to understand what a MME represents in terms of a statistical sample. Some studies have characterized the MME using techniques borrowed from weather forecasting in terms of the 'reliability' of present-day simulations with respect to observations[7,8] – the hypothesis that the observations can simply be regarded as one member of the MME without any special status – but those types of tests cannot be applied to future projections to assess their reliability. Others have sought to address the issue of shared approximations in model formulation and exchange of information between modeling groups[9].

Because of the difficulty in interpreting ad hoc collections of climate model projections, the climate change literature shows a range of different approaches to quantifying uncertainty in projections of future change. Some use simplified climate models, some use complex models built from 'first principles', some use multiple observational sources to evaluate those models, others take simple trends or metrics of model skill, some rely on basic understanding of the climate system, others use, what may appear to be, complex statistical techniques. Comparison of the different methods – their strengths, weaknesses and critical assumptions – is difficult because of their seemingly different formulations.

In this perspective, some of the different methods that have been used to make quantitative climate projections (including their uncertainties) are described and their assumptions, strengths and weaknesses are discussed. The work is inspired by some of the research that was discussed and undertaken during the 4-month Isaac Newton Programme on Mathematical and

Statistical Approaches to Climate Modelling and Prediction. Clearly a full explanation of the different methods would require considerable detail so the methods are only discussed at a basic level. The reader is encouraged to look at the original papers to gain further insight.

Climate Models, Errors and Uncertainties

Let us assume that any climate variable we are interested in can be described by a set of mathematical functions or model. Climate models may be simplified or complex, may be derived from physical principles or empirical relationships, or may contain elements of both. Examples range from simplified energy balance models through to complex climate or Earth System Models (ESMs). The climate variable might be the equilibrium climate sensitivity (the amount of global mean temperature change for a doubling of atmospheric CO_2), the amount of Arctic sea ice or something more complex like the amplitude and frequency of El Niño events. The model behaviour is controlled by what may be termed “internal” parameters (see the supplementary information) and by “external” forcing/boundary conditions of the climate system e.g. changes in concentrations of greenhouse gases, volcanic eruptions, orbital variations etc. The model can be used to simulate the past and the future by specifying different external forcings/boundary conditions and the behaviour of the model can be changed by varying the input parameters. In addition there are observations of past climate.

In general, simplified climate models only produce output in terms of simple or aggregate variables such as global mean temperatures, and have parameters that may similarly aggregate many physical processes. More complexity is required in the climate model to disaggregate in space and time and to simulate more complex phenomena such as precipitation or sea-ice. For simulations and projections of the smaller-scale climate variables that are required to address many policy questions, and for variables related to e.g. extreme events, requires the most complex ESMs running at high resolution.

Even the most complex climate models are approximations to the real climate system. Inadequacies or even ‘errors’ in models lead to inadequacies or errors in projections. Some inadequacies are inherent in the specification of the model (e.g. processes that are judged to be of second-order importance that are deliberately not included); others arise because limitations in computing power prevent the equations from being solved on a fine enough numerical grid, so sub-grid-scale processes must be parameterised. Complex models may simulate natural climate variability such as El Niño events (with varying degrees of success) but more simplified models may only simulate the forced response to a particular agent. For any climate projection there is both a systematic (epistemic) component of uncertainty and a random (aleatoric) component. The approximate partitioning of the range of spread of models between systematic (response and forcing) and random sources of uncertainty will depend on the variable, the spatial scale and the projection horizon of interest [10,11]. There is some potential for confusion as some studies may seek to quantify only the spread in the forced response of the climate system whereas some may seek to quantify both systematic and random components.

Quantifying Uncertainty in Projections

Ensembles of simulations of past and current climate, driven by estimates of past radiative forcing/boundary conditions, may be generated at different internal input parameter values, precise values of which are typically not known (figure 1). Observations are then used to produce a metric of the model skill in simulating selected aspects of past climate. The metric compares the model output with observed climate fields and may involve many different climate variables, trends and fields that are related to different physical processes (see supplementary information). The more realistic regions of parameter space are accepted or up-weighted, based on heuristic or more formal criteria, as those which are likely to produce the most realistic future climate projections. Less realistic regions are rejected or down-weighted. The model is calibrated by determining suitable values for the internal parameters that produce simulations of past climate consistent with the observations and their uncertainties.

Having calibrated the model, the parameters and/or their weights can be used to run an ensemble of simulations of future climate. The uncertainties in the projections are quantified in terms of probabilities. We say that both the input parameters and the projections are constrained by the observations. The climate model acts as a physically-based device to pass from historical or past climate and climate change to future projections. We expect that observations are not sufficient to constrain the parameters to single values so that multiple parameter combinations are consistent with the observations. The resulting projections will have uncertainties because of this.

The basic approach to producing projections with uncertainties is the same regardless of the complexity of the model and the climate variable of interest. Nevertheless, the implementation is affected by both factors. In general, the examples presented can all be couched in terms of a Bayesian approach with different assumptions and different techniques used in the implementation of the Bayes theorem. They are not presented in this way because that is not the way that the climate projection literature has evolved. Indeed, there has been a healthy debate within the community about the merits of such an approach and its implementation. What follows are examples of approaches drawn from different regions of the model complexity-variable complexity space.

Quantifying the Global Sensitivity of the Climate System

The climate sensitivity is a key measure of the global mean temperature response of a climate model. The equilibrium climate sensitivity may be expressed as the ratio of the radiative forcing and the climate feedback parameter. The time-dependent version of the model/formula has been exploited to compute the effective climate feedback parameter from the historical trend in ocean heat uptake (interchangeable with the top-of-atmosphere flux imbalance), the historical radiative forcing and the historical temperature change[12]. The study uses independent observations to derive distributions representing the uncertainty in global mean temperature trends and heat uptake. A distribution for radiative forcing is derived similarly, using calculations based on observed concentrations of greenhouse gases, aerosols, ozone and natural factors such as solar input and volcanic stratospheric aerosols. The internal model parameters are then sampled from these distributions and the model is evaluated to give an ensemble of climate sensitivity estimates. This is mathematically equivalent to varying the model parameters widely and then weighting the parameters using their observed and calculated estimates (with some statistical assumptions). Thus the distribution of the climate sensitivity is constrained by the observations (figure 2).

The main strength of the approach is in its simplicity in exploiting the global mean energy balance to produce a distribution of a key climate parameter, the climate sensitivity. Because of this simplicity it is relatively easy to perform sensitivity tests to see which of the model parameters is most influential in determining the relatively wide spread found in the study. This turns out to be the estimate of the radiative forcing: if, for example, the standard deviation of the forcing distribution could be halved then the 5^{th} percentile of the climate sensitivity distribution would increase from 1.6ºC to 2.5ºC.

Unfortunately the method produces a relatively weak constraint on the distribution, particularly on the upper tail. This is because the climate sensitivity estimated in this way involves a ratio of temperatures to fluxes and the denominator can get close to zero. (In fact, the distribution of the denominator in the equation for climate sensitivity admits negative values, leading to unrealistic negative climate sensitivities and a singularity which means that technically the distributions are not PDFs – a similar problem is found in[13] and is discussed in[14-18]). A further obvious drawback is that the method is only good for producing estimates of the global climate sensitivity (and feedback parameter) and such distributions can be sensitive to prior assumptions for the distributions of parameters which has been the subject of debate in the literature[16,19].

Different estimates of the probability density functions (PDFs) of the climate sensitivity have also been published[20] and other studies have used reconstructions of climate from before the observational record[21,22]. A review of palaeoclimate estimates has also been performed[23]. The climate sensitivity is one of the most studied and quantified climate projection-related variables. This is partly because model simulations suggest that it can be used to scale regional patterns of change[24] and partly because of a historical attachment of climate modelers to the doubled CO_2 experiment performed with a complex atmosphere model coupled to a thermodynamic or 'slab' ocean. This attachment may diminish as so-called slab-models fall into disuse because of technical issues with their implementation.

Large-scale Trends from Attributable Warming

The ASK [25,26] method exploits the possibility, demonstrated using energy balance climate models, that a bias in the temperature change in the future related to a particular forcing agent may be empirically related to the bias in the past change associated with that forcing agent, by a scaling factor (figure 3). The method computes a correction factor or recalibration of simulated past changes that can be used to scale future projections assuming that the empirical relationship continues to hold. The uncertain elements of the approach are the scaling factor and the component of past change related to a forcing agent. In the global mean temperature case, the scaling factor may be relatively well constrained (figure 3). The difficult parameter to assess is the past change that can be associated with a particular anthropogenic component such as CO_2, as represented by the histogram on the x-axis in fig. 3.

The observed record of global and large-scale temperature change is made of components forced by anthropogenic factors such as greenhouse gas and aerosols, external factors such as solar variability and volcanic eruptions and internally generated natural variability. Detection and attribution techniques seek to estimate these individual components of trends from the observed record, using complex climate model simulations in combination with regression techniques. Uncertainties arise because the responses to some forcing agents may correlate through time (e.g. concurrent rises in greenhouse gases and aerosols) making it hard to estimate the regression coefficients, because of uncertainties in reconstructing past forcing agents and because of potential errors in the complex model response to the forcing.

The ASK technique can therefore be thought of as generating an ensemble of future projections by sampling a large number of possible past trends that are attributable to a particular forcing agent. The parameters of the relationship between the past and the future and the attributable warming are constrained by observations and complex model studies and thus the projections are also constrained by those observations. By specifying the components of the radiative forcing separately, it is possible to make projections for combinations of radiative forcing that may occur in the future but that did not occur in the past.

Initial studies focussed on global mean temperatures[27] but have been extended to constrain continental-scale temperature changes[25]. The strengths of the approach are in the simplicity of the idea of extrapolating uncertainties in past trends. The complexity arises in the need to separate the components of the observed trends into those associated with greenhouse gases, aerosols, natural forcing factors and internal climate variability. For global mean projections, this separation is the largest source of uncertainty[26]. For regional quantities, relationships between past and future trends may be weak and for some variables and for smaller-scale regions, such relationships may not be evident in the complex models used in the detection and attribution step.

In the example highlighted here, a simple energy balance model is used to obtain the relationship between past warming and future change, hence it is tempting to conclude that the projections only quantify the uncertainty in the forced response. However, the estimate of the warming attributable to greenhouse gases is contaminated with natural variability (as we only have one realization of the real-world) so some account is taken of the random component. Limitations on computer resource also mean that results are often obtained from initial-condition ensembles from a small number of different climate models. Hence there is a potential for modeling uncertainties to be undersampled.

Emergent Constraints and Process-Based Metrics

Data archives from MMEs can also be used to link errors in simulating future and past change, in a similar spirit to the ASK technique. These data archives can be considered as representing our physical understanding of the climate system, as derived from climate models themselves. For some variables, simple relationships have been uncovered between future projection variables and past observed trends or variability. Future changes in September sea-ice extent in the Arctic have an approximately linear relationship with the past trends in the CMIP3 models[28] (figure 4). It is possible to empirically determine future trends using a simple scaling of the past trends, with some spread due to model errors and natural variability. The situation is similar to that seen in figure 3 except that the relationship is derived from complex climate model simulations rather than a simple energy balance model. By constraining the parameters of the linear relationship using the observations, it is possible to produce a calibrated projection of future September sea ice trends. Note that a different ensemble may produce a different relationship or a wider spread, but at least the sensitivity of the projections can be tested by varying such assumptions.

This Arctic-sea-ice study provides an example of what we might call an emergent constraint i.e. a relationship between past trends and future trends, developed empirically from climate model output used to make projections of the future. If the empirical relationship can be understood on simple physical grounds, belief in it is strengthened. It provides justification for attaching more credibility to models that match the observed trend well over the recent period, and hence for treating the difference between modeled and observed trends as a metric for the purposes of weighting or correcting models. Such a metric might be considered to be an example of a process-based metric i.e. a metric that is used to evaluate a process (the sensitivity of sea ice change) rather than simply a metric of how the model compares with reality in terms of the spatial distribution of sea-ice in the time average. However, a precise definition of what is process-based and what is not has not been provided in the literature and is an area that needs to be developed.

The main strength of the approach is in the simplicity and in the physical transparency. The main weakness is that it may not work in such a transparent way for all climate projection variables – although other relationships have been found[29]. Also, care must be taken to test the validity of the relationship. In the case of September sea ice, as conditions become ice free in the simulations, the trends become non-linear and the use of a simple linear regression in figure 4 would not be valid.

Bayesian Projections with Perturbed Physics Ensembles

Emergent constraints have only been found for a few climate projection variables and there is a further issue that projections of different variables produced in this way may be inconsistent with each other. Such issues have led to the development of the so-called perturbed-physics approach [30-34]. Uncertain parameters in a single climate model may be perturbed to produce alternative simulations of past and future climate and climate change (as in the case of the simplified climate model approaches described above).

In the perturbed physics approach, the input parameters are varied and the model is run using past and future radiative forcing. As in the general algorithm (figure 1) we can imagine a point in the parameter space that maps to a point in the past-climate-space that is consistent with the observations as measured by some metric i.e. is within the observational error bound. A simulation from a second point of parameter space may be less consistent with the observations. When we look at the future projections made using the model run from the first point, we may assume that these are more likely that the projections made from the second point. By running many ensemble members with the model covering the parameter space, it is possible to build up a weighted-distribution of future projections where the weights relate to the metric[35]. A key step in such analyses is to decide what observations to use: the choice is often determined by the design of the perturbed physics ensemble. In much of the work that has been conducted, a version of the atmosphere model coupled to a simple slab ocean has been used, restricting the observations to mainly time-averaged climatological fields[36,37].

In practice, running enough simulations to adequately sample a complex model parameter space and, moreover, to test the sensitivity of the projections to different assumptions about the distributions of those parameters, is computationally challenging. The burden can be eased using emulators, which are statistical models of ensembles that map input parameters to outputs, so enabling larger pseudo-ensemble calculations to be performed (albeit with loss of numerical accuracy)[38]. To combine the climate model outputs with the observations and emulators is a difficult statistical problem that is most easily handled in a Bayesian framework[35].

A further refinement is to introduce a term to represent irreducible or structural errors in climate model. If we imagine a point in parameter space at which the model produces its best simulation of both past and future climate, then, unless the model is perfect, there will still be a mismatch between model outputs and reality. Specifying the structure of this mismatch remains one of the most challenging problems in climate projection. One possibility is to take the discrepancy from the multi model ensemble as a lower bound on this 'structural error'[37].

The strengths of the perturbed-physics/Bayesian approach are that, in principle, many different observational constraints can be brought to bear on the projections, and projections of many complex climate variables (e.g. involving regional averages and extremes) may be produced[39](figure 5). Projections of several quantities simultaneously (joint projections) are also possible where the complex climate model provides the physical link between changes in those different variables. The main weakness is that, in order to use the latest, most comprehensive of climate models, the implementation is expensive in terms of computing resources and requires a very high level of technical expertise. This makes it hard to understand in simple physical terms how the observations constrain the projections.

Making Progress in Quantitative Projection

Simplified climate models (including empirical models derived from complex model output) can be easily used with formal statistical approaches to quantify uncertainty in projections but can only produce limited output: thus limited observations may be used to constrain parameters, and projections can only be made in terms of limited climate variables. As models become more complex, simulations and projections of more complex variables may be made, widening both the possible observational data that may be used to constrain parameters and the range of variables for which projections may be generated. But it becomes more expensive to produce ensembles and harder to implement and understand the projections.

The use of metrics, skill measures, model ranking and even model weighting are starting to be more widely adopted in the climate model evaluation and projection literature. This is fine when such quantitative approaches are used as a guide to future model development or as a guide to the validity of some physical understanding derived from models, although care should be taken to fully understand why that metric is a useful measure. Where metrics are used in projections, it is not safe to assume that a weighted distribution of models is superior to an unweighted distribution without demonstrating that the metric does relate, in some physically plausible way, to the projection variable of interest, and without testing the underlying assumptions[40].

There is growing use in the community of terms such as process-based metric and 'process-based' evaluation, yet it is not possible to find a formal definition of process-based in the literature. It could be argued that surface fluxes are the processes that determine the spatial variations in surface air temperature (SAT) change, so they should be used in a process-based metric of SAT changes. But clouds have a leading-order impact on surface radiation, so should cloud effects be defined as the process? It is unclear. Perhaps "process" implies rates-of-change of one variable with respect to another – under climate change or under forced or free variations on shorter time scales[29]. Is the warming attributable to greenhouse gases process-based? A better characterisation of the concept is required.

The concept of the emergent constraint is appealing because of the clear physical interpretation. However the implementation may be challenging as we have yet to produce a generic mathematical algorithm or recipe that can be used in other cases in which all the assumptions are revealed and all sources of uncertainty are considered. Perhaps the approach might be extended to account for non-linearities or even assess the impact of inadequacies that are common to all models. It is recommended that work is undertaken on both the theoretical underpinning and numerical implementation of the approach, so that it can be applied more widely.

If the behaviour of the complex models can be reproduced by fitting the parameters of a simple or intermediate models (physical or empirical) to the complex model output, then it is possible to use observations to constrain the smaller set of parameters from larger ensembles of the simple/intermediate model. We might consider this a form of "process-based emulation", without being at all rigorous about the definition of such a term. Intermediate models exist for even quite complex phenomena such as the El Nino Southern Oscillation[41,42]. They have generally been used to understand models and the real world but could also be applied to the projection problem.

To conclude, it is possible to produce quantitative projections of climate change, combining models of varying complexity and observations, expressed in terms of probabilities that measure our current uncertainty in those projections. Of course, our knowledge, as embodied in models and observations, may improve in time and thus we might be able to reduce those uncertainties. However, the possibility that new models, new observations or new theoretical research might alter the current set of projections considerably cannot be ruled out. For example, new feedbacks may be discovered or resolution thresholds are crossed so that previously parameterised process are directly resolved in models.

Acknowledgements

Financial support for the programme on “Mathematical and Statistical Approaches to Climate Modelling and Prediction” was provided by the Isaac Newton Institute and by the UK Natural Environment Research Council.

Figure Captions

Figure 1: A schematic representation of the general framework for producing projections of future climate. The climate model, M, produces output in terms of a climate variable, c, and is controlled by the model parameters, p, and the input forcing R. The model may be run with different parameter values p_1, p_2, … to produce simulations of historical climate c_{h}, and projections of future climate, c_{f} . The dark grey shaded area in the left represents the space of plausible input parameters of the model that we would consider before doing any simulations. The dark grey shaded areas on the right represent the spaces of past or historical simulated climate variables and future projections generated by running the model at that wide range of different input parameters. The simulations of historical climate may be compared with observations, *o*, using a metric, and taking into account observational errors. If one point in the climate model parameter space, p_1, produces a better simulation of historical climate than another point p_2, then the hope is that it will give a better (i.e. less error-prone) simulation of future climate. Thus we can contract the space of past or historical climate change produced by the model (light grey shading). Because there is a three-way mapping between this historical simulation space, the input parameters and the future projections, the parameter ranges are also constrained, as are the future projections, again represented by the light grey shading.

Figure 2: A PDF for the climate sensitivity obtained using a simple energy balance model approach[12]. The thick black PDF shows the curve from the original study. The thin black curve is the climate sensitivity PDF obtained if the standard deviation of the distribution of the radiative forcing input parameter is halved.

Figure 3: (a) Global mean temperature anomalies produced using an energy balance model[24,43] forced by historical changes in well-mixed greenhouse gases and future increases based on the SRES A1B scenario. The different curves are generated by varying the feedback parameter (climate sensitivity) in the EBM. (b**)** Changes in global mean temperature at the year 2000 (x-axis) vs changes in global mean temperature at 2050 obtained from the figure in the left panel showing the relationship between past changes and future temperature changes. The histogram on the x-axis represents an estimate of the 20th-century warming attributable to greenhouse gases[44]. The histogram on the y-axis uses the relationship between the past and the future to obtain a projection of future changes.

Figure 4: The modelled trend in 1979-2007 September Arctic sea-ice extent (expressed as a percentage of the total – average of 1900-1979 – x-axis) vs the 2021-2040 trend in the same variable (y-axis) computed from the CMIP3 model simulations[28] of historical climate change and future climate change under the SRES A1B scenario (solid dots) and from perturbed physics ensembles[30] (open dots). The solid black diagonal line shows the line of best fit between the historical trends and the future extents. The best estimate of the observed trend in September sea ice extent is shown by the vertical dotted line.

Figure 5: PDFs of 20-year average changes in Northern European surface air temperature (a) and precipitation (b) under the SRES A1B scenario derived using perturbed physics ensembles and a Bayesian statistical approach[39]. Changes are expressed as anomalies w.r.t. 1961-1990 period. The different PDFs correspond to different future time periods from left to right; 2000-2020, 2020-2040, 2040-2060 and 2080-2100.

References

1 Pall, P. *et al.* Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000. *Nature* **470**, 382-385, doi:DOI 10.1038/nature09762 (2011).

2 Challinor, A.J., Simelton, E.S., Fraser, E.D.G., Hemming, D. & Collins, M. Increased crop failure due to climate change: assessing adaptation options using models and socio-economic data for wheat in China. *Environmental Research Letters* **5**, doi:10.1088/1748-9326/5/3/034012 (2010).

3 Murphy, J.M. *et al. UKCP09 Climate change projections*. (2009).

4 Meehl, G.A. *et al.* The WCRP CMIP3 multimodel dataset - A new era in climate change research. *Bulletin of the American Meteorological Society* **88**, 1383-1394, doi:DOI 10.1175/BAMS-88-9-1383 (2007).

5 IPCC. Meeting Report of the Intergovernmental Panel on Climate Change Expert Meeting on Assessing and Combining Multi Model Climate Projections. 117 pp (IPCC Working Group I Technical Support Unit, University of Bern, Bern, Switzerland, 2010).

6 Tebaldi, C. & Knutti, R. The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences* **365**, 2053-2075, doi:DOI 10.1098/rsta.2007.2076 (2007).

7 Annan, J.D. & Hargreaves, J.C. Reliability of the CMIP3 ensemble. *Geophysical Research Letters* **37**, L02703, doi:10.1029/2009gl041994 (2010).

8 Yokohata, T. *et al.* Reliability of multi-model and structurally different single-model ensembles. *Climate Dynamics*, submitted (2011).

9 Chandler, R.E. Exploiting strength, discounting weakness: combining information from multiple climate simulators. *Submitted* (2011).

10 Hawkins, E. & Sutton, R. The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society* **90**, 1095-1107, doi:1010.1175/2009BAMS2607.1091, doi:DOI 10.1175/2009BAMS2607.1 (2009).

11 Hawkins, E. & Sutton, R. The potential to narrow uncertainty in projections of regional precipitation change. *Climate Dynamics* **37**, 407-418 (2011).

12 Gregory, J., Stouffer, R., Raper, S., Stott, P. & Rayner, N. An observationally based estimate of the climate sensitivity. *Journal of Climate*, 3117-3121 (2002).

13 Roe, G. & Baker, M. Why is climate sensitivity so unpredictable? *Science* **318**, 629-632, doi:DOI 10.1126/science.1144735 (2007).

14 Roe, G. & Armour, K. How sensitive is climate sensitivity? *Geophysical Research Letters* **38**, -, doi:ARTN L14708
DOI 10.1029/2011GL047913 (2011).

15 Armour, K. & Roe, G. Climate commitment in an uncertain world. G*eophysical Research Letters* 3**8,** L01707, doi:ARTN L01707
DOI 10.1029/2010GL045850 (2011).

16 Annan, J.D. & Hargreaves, J.C. On the generation and interpretation of probabilistic estimates of climate sensitivity. Cl*imatic Change 1*0**4,** 423-436, doi:10.1007/s10584-009-9715-y (2011).

17 Hannart, A., Dufresne, J. & Naveau, P. Why climate sensitivity may not be so unpredictable. Ge*ophysical Research Letters 3*6**,** -, doi:ARTN L16707
DOI 10.1029/2009GL039640 (2009).

18 Zaliapin, I. & Ghil, M. Another look at climate sensitivity. Non*linear Processes in Geophysics 17*, **1**13-122 (2010).

19 Frame, D. et *al.* Constraining climate forecasts: The role of prior assumptions. *Geophysical Research Letters 32*, doi:10.1029/2004GL022241 (2005).

20 Gregory, J. & Forster, P. Transient climate response estimated from radiative forcing and observed temperature change. *Journal of Geophysical Research-Atmospheres 113*, **D**23105, doi:ARTN D23105 DOI 10.1029/2008JD010405 (2008).

21 Hegerl, G., Crowley, T., Hyde, W. & Frame, D. Climate sensitivity constrained by temperature reconstructions over the past seven centuries. *Nature 440*, **10**29-1032, doi:DOI 10.1038/nature04679 (2006).

22 Schneider von Deimling, T., Held, H., Ganopolski, A. & Rahmstorf, S. Climate sensitivity estimated from ensemble simulations of glacial climate. *Climate Dynamics 27,* **14**9-163, doi:DOI 10.1007/s00382-006-0126-8 (2006).

23 Edwards, T., Crucifix, M. & Harrison, S. Using the past to constrain the future: how the palaeorecord can improve estimates of global warming. *Progress in Physical Geography 31,* **48**1-500, doi:DOI 10.1177/0309133307083295 (2007).

24 Harris, G.R. et *al.* Frequency distributions of transient regional climate change from perturbed physics ensembles of general circulation model simulations. *Climate Dynamics 27,* **35**7-375, doi:10.1007/s00382-006-0142-8 (2006).

25 Stott, P.A., Kettleborough, J.A. & Allen, M.R. Uncertainty in continental-scale temperature predictions. *Geophysical Research Letters 33,* **L0**2708, doi:10.1029/2005gl024423 (2006).

26 Stott, P.A. & Kettleborough, J. Origins and estimates of uncertainty in predictions of twenty-first century temperature rise (vol 416, pg 723, 2002). *Nature, 20*5-205 (2002).

27 Stott, P.A. et *al.* Observational constraints on past attributable warming and predictions of future global warming. *Journal of Climate 19,* **30**55-3069 (2006).

28 Boe, J., Hall, A. & Qu, X. September sea-ice cover in the Arctic Ocean projected to vanish by 2100. *Nature Geoscience 2,* 3**4**1-343, doi:DOI 10.1038/NGEO467 (2009).

29 Hall, A. & Qu, X. Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophysical Research Letters 33,* **L0**3502, doi:ARTN L03502 DOI 10.1029/2005GL025127 (2006).

30 Collins, M. et *al.* Climate model errors, feedbacks and forcings: a comparison of perturbed physics and multi-model ensembles. *Climate Dynamics 36,* 1**73**7-1766, doi:10.1007/s00382-010-0808-0 (2011).

31 Collins, M. et *al.* Towards quantifying uncertainty in transient climate change. *Climate Dynamics 27,* 1**27**-147, doi:10.1007/s00382-006-0121-0 (2006).

32 Murphy, J.M. et *al.* Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature 430,* **768**-772, doi:10.1038/nature02771 (2004).

33 Stainforth, D. et *al.* Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature 433,* **403**-406, doi:10.1038/nature03301 (2005).

34 Brierley, C.M., Collins, M. & Thorpe, A.J. The impact of perturbations to ocean-model parameters on climate and climate change in a coupled model. *Climate Dynamics 34,* 3**25**-343, doi:10.1007/s00382-008-0486-3 (2010).

35 Rougier, J. Probabilistic inference for future climate using an ensemble of climate model evaluations. *Climatic Change 81,* 2**47**-264, doi:10.1007/s10584-006-9156-9 (2007).

36 Murphy, J.M. et *al. A me*thodology for probabilistic predictions of regional climate change from perturbed physics ensembles. *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences 365,* **199**3-2028, doi:10.1098/rsta.2007.2077 (2007).

37 Sexton, D., Murphy, J., Collins, M. & Webb, M. Multivariate predictions using imperfect climate models: Part 1 outline of methodology. *Climate Dynamics (201*1).

38 Rougier, J., Sexton, D.M.H., Murphy, J.M. & Stainforth, D. Analyzing the Climate Sensitivity of the HadSM3 Climate Model Using Ensembles from Different but Related Experiments. *Journal of Climate 22,* 3**54**0-3557, doi:10.1175/2008jcli2533.1 (2009).

39 Harris, G.R., Collins, M., Sexton, D.M.H., Murphy, J.M. & Booth, B.B.B. Probabilistic projections for 21st century European climate. *Natural Hazards and Earth System Sciences 10,* 2**00**9-2020, doi:10.5194/nhess-10-2009-2010 (2010).

40 Knutti, R., Furrer, R., Tebaldi, C., Cermak, J. & Meehl, G. Challenges in Combining Projections from Multiple Climate Models. *Journal of Climate 23,* 2**73**9-2758, doi:10.1175/2009JCLI3361.1 (2010).

41 Philip, S. & van Oldenborgh, G. Atmospheric properties of ENSO: models versus observations. *Climate Dynamics, 107*3-1091, doi:DOI 10.1007/s00382-009-0579-7 (2010).

42 Philip, S., Collins, M., van Oldenborgh, G. & van den Hurk, B. The role of atmosphere and ocean physical processes in ENSO in a perturbed physics coupled climate model. *Ocean Science, 441*-459 (2010).

43 Huntingford, C. & Cox, P. An analogue model to derive additional climate change scenarios from existing GCM simulations. *Climate Dynamics, 575*-586 (2000).

44 Hegerl, G.C. et *al. in Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, 2007).