Assessing the Reliability of Climate Models, CMIP5

Bart van den Hurk, Pascale Braconnot, Veronika Eyring,
Pierre Friedlingstein, Peter Gleckler, Reto Knutti, and Joao Teixeira

Abstract In spite of the yet incomplete subsample of the 5th phase of the Coupled Model Intercomparison Project (CMIP5) model ensemble to date, evaluation of these models is underway. Novel diagnostics and analysis methods are being utilized in order to explore the skill of particular processes, the degree to which models have improved since CMIP3, and particular features of the hindcasts, decadal and centennial projections. These assessments strongly benefit from the increasing availability of state-of-the-art data sets and model output processing techniques. Also paleo-climate analysis proves to be useful for demonstrating the ability of models to simulate climate conditions that are different from present day. The existence of an increasingly wide ensemble of model simulations re-emphasizes the need to carefully consider the implications of model spread. Disparity between projected results does imply that model uncertainty exists, but not necessarily reflects a true estimate of this uncertainty.

B. van den Hurk (✉)
Royal Netherlands Meteorological Institute (KNMI), AE De Bilt,
Post Box 201 NL-3730, The Netherlands
e-mail: hurkvd@knmi.nl

P. Braconnot
Insitut Pierre Simon Laplace/Laboratoire des Sciences du Climat et de l'Environnement,
unité mixte de recherches CEA-CNRS-UVSQ, Bât.712, Orme des Merisiers CE-Saclay,
Gif sur Yvette Cedex F-91191, France
e-mail: pascale.braconnot@lsce.ipsl.fr

V. Eyring
Deutsches Zentrum fuer Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre
(IPA), Oberpfaffenhofen, Wessling 82234, Germany
e-mail: Veronika.Eyring@dlr.de

P. Friedlingstein
College of Engineering, Mathematics and Physical Sciences, University of Exeter,
Harrison Building, Streatham Campus, North Park Road, Exeter EX4 4QF, UK
e-mail: P.Friedlingstein@exeter.ac.uk

G.R. Asrar and J.W. Hurrell (eds.), *Climate Science for Serving Society: Research, Modeling and Prediction Priorities*, DOI 10.1007/978-94-007-6692-1_9,

Projections generated by models with a similar origin or utilizing parameter perturbation techniques generally show more mutual agreement than models with different development histories. Weighting results from different models is a potentially useful technique to improve projections, if the purpose of the weighting is clearly identified. However, there is yet no consensus in the community on how to best achieve this.

These findings, discussed at the session “Assessing the reliability of climate models: CMIP5” of the World Climate Research Program (WCRP) Open Science Conference (OSC), illustrate the need for comprehensive and coordinated model evaluation and data collection. The role that WCRP can play in this coordination is summarized at the end of this chapter.

Keywords Climate model assessment • Evaluation • Model ensembles • Process verification • CMIP5 • WCRP coordinations

List of Acronymns

AMIP	Atmospheric Model Intercomparison Project
CMIP3	CMIP5 3rd, 5th Coupled Model Intercomparison Project
ENSO	El Nino Southern Oscillation
ESM	Earth System Model
GCM	General Circulation Model
IGBP	International Geosphere-Biosphere Program
IHDP	International Human Dimensions Program
IPCC	Intergovernmental Panel on Climate Change
ISCCP	International Satellite Cloud Climatology Project
MME	Multi Model Ensemble
OSC	Open Science Conference
PCMDI	Program for Climate Model Diagnosis and Intercomparison
PPE	Perturbed Physics Ensemble
WCRP	World Climate Research Program

P. Gleckler
Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, USA
e-mail: gleckler1@llnl.gov

R. Knutti
Institute for Atmospheric and Climate Science, CHN N 12.1,
Universitätstrasse 16, ETH Zurich, CH-8092, Switzerland
e-mail: reto.knutti@env.ethz.ch

J. Teixeira
Jet Propulsion Laboratory, M/S 169-237, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
e-mail: Joao.Teixeira@jpl.nasa.gov

1 Introduction

The assessment of the reliability of climate models is needed to have confidence in the information about future development of the climate system generated with these models. It is generally applied by confronting climate model output with observations over a past period, and interpreting the performance of the model to replicate observed trends, spatial and temporal variability patterns, mean seasonal cycles, responses to perturbation and mutual relationships between relevant quantities. However, this assessment is subject to a large number of aspects:

- the quality and representativity of the reference observational data set
- the knowledge on the initial and time-varying boundary conditions needed to force the climate model
- the comparability between the observed and modeled quantities
- interpretation of discrepancies in terms of model or observational deficiencies, etc.

Yet, this assessment is rapidly evolving and improving. In the context of the 5th Coupled Model Intercomparison Project (CMIP5) an increasing number of climate model simulations is becoming available, and a wide range of analyses currently based on a sub-set of the anticipated model ensemble is being undertaken. During the WCRP Open Science Conference (OSC) held in Denver, October 2011, a selection of studies dedicated to the assessment of the reliability of these climate models was presented in the parallel session B7. Many studies referred back to results from the earlier CMIP3 project, which likewise benefited from the public availability of a large set of model results, leading to a revolution of model evaluation tools, observations and diagnostics. This revolution is ongoing as CMIP5 is running ahead, but important new findings can already be noted. Here we provide an overview of the main topics that emerged during the OSC, which reflect the current state-of-the-art assessments for the reliability of climate models. In particular we summarize what has been implied from the spread in model results, provide examples of novel observations and diagnostics, and give a set of examples of ongoing process evaluation studies that have been discussed as part of the session. Recommendations for WCRP to the governance of this important activity are given at the end of this document.

2 The Implications (and Usefulness) of Model Spread

CMIP5 is clearly more ambitious than its predecessors (in particular CMIP3): although it is still under development, more experiments and associated research questions, more participating models, more model fields, a better documentation of models, and more data storage are becoming available (CLIVAR 2011; Taylor et al. 2012). As model data are submitted to the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the several storage nodes that are linked together, it becomes evident that model spread will still be substantial. Part of the difference between model results can be attributed to unforced variability, originating from the nonlinear nature of the variable climate system. The impact of this unforced variability reduces as the projection horizon, spatial scale or averaging period increases (Hawkins and Sutton 2009), although natural variability may still be pronounced at the small spatial scales at the end of the twenty-first century. At short time scales (i.e., less than 5 years) unforced uncertainty may potentially be reduced by a realistic initialization of the forecasts, a procedure at the heart of the decadal projections contained in CMIP5.

The spread between equally forced models at longer projection time scales or averaging intervals can be considered to be related to the total model uncertainty. However, this spread does not reflect systematic biases in the models, it assumes that the model sample is representative across the model space, and may be limited due to model formulations that are mutually similar. In general, model spread does imply that model uncertainty exists, but it may not reflect what we think the "true" model uncertainty is, because models are related by taking some observations as reference for the model tuning (Masson and Knutti 2011) and may not sample all possible uncertainties.

Although an increased model ensemble does not capture model deficiencies common to all models (like missing small-scale processes), it is informative about our ability to reliably simulate the climate and its response to external forcings. The design of CMIP5 allows assessment of the importance of many processes in the climate system (see examples of process analyses below), and can help set research priorities in order to reduce this aspect of uncertainty (Dufresne and Bony 2008). Some of these process uncertainties can potentially be reduced by making use of observations showing variability due to comparable forcings at seasonal or interannual time scales. Examples include the evaluation of the snow-albedo feedback (Hall and Qu 2006) and the evaluation of the terrestrial biosphere in the carbon-climate feedback (see below). A paleo-modeling experiment is explicitly included in CMIP5, focusing on the mid Holocene, the Last Glacial Maximum (LGM, where large ice sheets and low greenhouse gas levels were present) and the Past Millennium are considered specifically. Also paleo-observations of SST during the LGM from the MARGO synthesis (Margo Project Members 2009) allow "out-of-sample" evaluation of climate models, that is, evaluation of models under climate conditions different from present day. In spite of a fair amount of uncertainty of these observations, a reliable model ensemble should encompass the observed observation range, showing as a preferably uniform rank histogram of model-observation differences. Comparisons between MARGO data and a Perturbed Physics Ensemble (PPE) generated by perturbing physical parameters clearly showed this PPE to be incapable of capturing the large SST-responses of the LGM relative to the current climate. The available CMIP5 Multi-Model Ensemble (MME) was shown to be able to encompass the LGM observed global mean SST-response, although spatial patterns of this response and individual model results were not as reliable (Hargreaves et al. 2011).

Model spread can also be utilized to diagnose the inherent predictability of the climate at decadal time scales. Climate states can be considered to be predictable if their probability of occurrence conditional to the initial state is significantly different from the climatological probability of occurrence. An ensemble of initialized model simulations will at some stage diverge to their climatological probability distribution, preferably at the same rate as nature. The predictability of the natural climate system given an initial state cannot be assessed, as we have only a single realization of the near future. But long integrations of climate models can be used to infer their inherent predictability, for instance by calculating the rate of divergence from an ensemble of episodes that have analogous start conditions (Branstator and Teng 2010). This procedure does assume the absence of model error and thus maps the inherent predictability in the modeled climate, which can be considered as an upper limit of predictability under the assumption that models are free of systematic errors. Analysis of the predictability of the 5 year low-pass filtered ocean heat content in the upper 300 m from six climate models for which a long unforced integration was available revealed that the time range in which useful predictions could be made varied between 5 and 20 year for the North Atlantic basin and somewhat shorter for the Pacific. An evaluation with 10 CMIP5 models shows comparable results, albeit that the results varied widely across the ensemble. Assessment of the inherent predictability should be carried out for every model participating in the decadal prediction simulations (e.g. Matei et al. 2012).

The existence of an MME and their varying degree of consistence with a wide range of observations raises the question as whether a probability distribution of future climate conditions could be constructed by weighting the models using performance metrics. Model quality metrics obtained by combining multi-variable performance metrics such as those presented by Reichler and Kim (2008) demonstrate an increase in skill of climate models over time, but their interpretations are not clear, and not very useful for any particular purpose. For example, climate change assessments in the Arctic regions will be tempted to give a stronger weight to metrics that represent sea ice conditions, whereas climate change assessments for the Sahel will need other variables to be represented well (Knutti 2008). The increased skill of climate models does not imply that the spread in future projections is reducing. On the contrary, preliminary analysis of a small subset of CMIP5 (10 models) shows that the spread in twenty-first century global mean temperature is similar to the CMIP3 ensemble, despite considerable model development. New observational analyses put extra constraints on the range in modeled climate sensitivity (e.g. the analysis of land-ocean contrasts in longwave radiation by Huber et al. 2011), but the probability distribution of this climate sensitivity is still wide.

The Intergovernmental Panel on Climate Change (IPCC) Expert Meeting on "Assessing and Combining Multi Model Climate Projections[1]" held in Boulder in January 2010 (Knutti et al. 2010) gave a list of properties of performance metrics to be useful. In particular:

- they should be simple to interpret
- they should be related to the prediction purpose
- they should reflect known processes

[1] https://www.ipcc-wg1.unibe.ch/publications/supportingmaterial/IPCC_EM_MultiModelEvaluation_MeetingReport.pdf

- relevant observations with sufficiently low uncertainty should be available
- they should be robust against their exact definition and aggregation procedure.

When metrics comply with these criteria, they could be used to generate a model weighting or selection procedure, provided that the rationale and implementation of the weighting are clearly defined and documented.

The existence of a range of models is a prerequisite of generating useful climate change assessment. No model is perfect, and even a model based on “the best available knowledge” cannot easily be defined since the “best knowledge” concerning a particular process or regime cannot easily be defined. The variability in the multi-model ensemble should be utilized and tailored to the application at hand.

3 New Observations and Diagnostics

Since the CMIP3 era a wealth of new observations, diagnostics and analysis methods have evolved, tailored to the evaluation of physical processes, their interactions, prediction skill and reliability for describing climate change. A full review of these developments is out of the scope of this report, but a few noticeable developments were discussed at the session, which are summarized here.

Use of observations is often implicit in the development and tuning of climate models: no model can be constructed without them. Model output evaluation using observations implies an explicit use of these. This evaluation supports the further development of the models, and gives credibility to the projections produced. In practice model evaluation and generation of “operational” climate model projections are parallel processes, where the model versions that usually have some inertia between upgrades. It preferably should be designed to highlight concrete model components that should be changed or replaced in order to improve the model’s skill. An assessment of the model uncertainty in quantities that are subject to many processes (such as near surface temperature or precipitation) is in itself useful (see previous section), but often does not reveal the necessary adjustments to models with limited skill. In the context of model evaluation, observations should be as much as possible analogous to the model variables that are generated (see the CMIP5 protocol of Taylor et al. 2012), and readily be available to the research community. For this a strong collaboration between model developers and data collectors (including satellite mission teams) is mandatory, not only concerning the technical infrastructure that allows model-to-observation comparisons, but also in the area of defining comparison metrics and skill thresholds.

A promising initiative in this respect is the presence of a “Obs4MIPS” tab on the PCMDI website (Teixeira et al. 2011) that discloses a number of satellite products designed to evaluate model cloud, precipitation and radiation characteristics. For instance, Jiang et al. (2012) compared A-train ice/water cloud and integrated water vapor observations to a range of CMIP3 and CMIP5 models, generally showing an improvement of the modeled cloud characteristics over the recent past. A traditional approach to compare model output to satellite data is to transfer the observed radiances into physical fields using retrieval algorithms. However, using this approach error propagation and aggregation are difficult to assess. Therefore, significant progress has been made in CMIP5 by building in satellite simulators into the GCMs, allowing evaluation of radiances instead (e.g. Bodas-Salcedo et al. 2011). However, the translation of radiance errors back to model improvement is often concealed by the many processes considered in the forward radiance modeling. A new stage in model-to-observation comparison is to map radiances back to model fields that takes the observational postprocessing and aggregation into account, but yet allows a model evaluation in geophysical units.

A powerful analysis method is to conditionally sample observations and model data in order to obtain quantities that are representative for a certain climate regime. A Cloud Regime Error Metric is derived by Williams and Webb (2009) by decomposing cloud regimes from the International Satellite Cloud Climatology Project (ISCCP) data archive discriminating classes of cloud top temperature and optical thickness. Utilizing this conditional sampling approach model biases can be specified for particular cloud regimes, enabling discerning errors due to a misrepresentation of cloud radiative forcing for particular classes and signals attributable to changes in the relative frequency of occurrence of particular cloud regimes. Using transpose-AMIP simulations from a single CMIP5 model for which all necessary data were available it was shown that biases in cloud radiative properties develop very fast in the forecast (already present 1 day after the initialization). Also a persistent problem of undersampled frequency of mid-level clouds is evident from this analysis.

Advanced statistical techniques are also being utilized to detect the scale dependence of skill of GCMs (Sakaguchi et al. 2012). The skill of GCM-generated surface temperature trends over the past decades obviously varies over temporal and spatial scales: global mean and long term trends are more easily reproduced than similar trends at smaller scales. The detection of the spatial and temporal scale of model skill is strongly relevant for the confidence in model climate projections. Global mean temperature trends from a sub-set of CMIP5 climate models are shown to be accurate: the uncertainty is smaller than the observational uncertainty. At smaller spatial scales the CMIP5 subset outperforms the earlier CMIP3 ensemble, at least for longer time scales.

Also for evaluation of land surface processes more data sets have become available. Jung et al. (2009) used a regression tree analysis to extrapolate Fluxnet site observations of surface evaporation and gross primary production (GPP) to all land areas by means of a set of climate data and vegetation indices satellite products. This data set is useful for evaluation of global patterns of mean GPP and evaporation, but by nature of its construction trend analysis cannot be applied. Leaf area index (LAI) data from MODIS show that simulations by the CMIP5 Earth System models show a fair correlation for the northern hemisphere (Anav et al. 2013).

For paleo-studies an increasing number of observations becomes available. Mutually independent data sets exist that reveal information on vegetation (pollen and other tracers), fires (charcoal deposition), regional hydrology (lake level marks) and aerosol level (dust deposition). Schmittner et al. (2011) used the Univ. of Victoria climate model to constrain the likelihood range of the climate sensitivity using LGM temperature reconstructions. Similar exercises are currently being undertaken using CMIP5 model output. In general, climate models seem to reproduce first order responses to different climate conditions fairly well, but do not reconstruct the regional signature of these responses and the various feedbacks at the millennium time scale (such as vegetation feedback) very well.

4 Examples of Process Evaluations Currently in Progress

Many process evaluations are currently underway, exploring the (yet limited) CMIP5 data archive with sophisticated analysis methods. Here we give a small number of examples of such studies that were presented during the session, not attempting to give a complete overview of the ongoing analyses.

An important source of uncertainty is the degree to which terrestrial and oceanic fluxes of CO_2 respond to future climate change. In the C4MIP experiment, Friedlingstein et al. (2006) showed that uncertainty in this response, represented in an ensemble of Earth System Models (ESMs) representing the carbon cycle and its interactions with the climate system, has a strong impact on the projected global temperature, due to pronounced feedbacks between the climate and the carbon cycle. Increased ecosystem release of carbon under warmer climate conditions may imply a strong positive carbon-climate feedback. Determination of the strength of this feedback is one of the outstanding problems in climate research.

Hall and Qu (2006) used the pronounced seasonal cycle in observed snow cover to determine the strength of the snow-albedo feedback (where reduced snow cover leads to higher radiative absorption which in turn promotes snow melt), and compared this to climate change projections from a range of GCMs. The physical mechanism of the snow-albedo feedback is fairly well understood and operates similarly at the seasonal and centennial time scale, thus allowing to determine the optimal feedback strength that should be present in the model simulations. Cox et al. (2012) similarly utilize observations collected at time scales covered in the current data record to infer an estimate of the carbon-climate feedback strength. Notifying that the observed interannual variability of atmospheric CO_2 concentration is primarily due to terrestrial biosphere responses to (ENSO-modulated) temperature fluctuations, the terrestrial carbon loss per degree warming can be derived from the observational record. During ENSO years the carbon uptake by vegetation is much weaker than during non-ENSO years. CMIP5 models can similarly be evaluated on such interannual time scales and tested against the observed CO2-climate sensitivity.

Using a similar approach, Mahlstein and Knutti (JGR submitted) use the relation between Arctic temperature and sea ice extend to estimate future ice-cover area as a function of regional temperature projections from CMIP5. According to this simple extrapolation the Arctic will be free of ice during summer when global mean temperature increases by 2K above present, whereas the uncalibrated CMIP3 models suggest that this does not occur until a 3K global mean warming.

Analysis of feedbacks between processes is the key in evaluating the ocean component of GCMs. This feedback analysis requires advanced processing of available observations and the use of informative conceptual frameworks. The CMIP5 archive and individual models participating to this experiment are currently explored intensively to diagnose the complex physical feedbacks between the ocean and atmosphere. Guilyardi et al. (2011) revisit the classical ENSO theory of the interplay between the dynamical Bjerkness feedback and the heat flux feedback, and conclude that the disparity between a subset of 6 CMIP5 models is largest in the strength and sign of the Bjerkness feedback. Spatial patterns of the shortwave radiative forcing play a key role. Evaluation of the CCSM4 hindcasts by Bates et al. (2012) reveal the existence of compensating errors between solar and evaporative fluxes. Using the Common Ocean-ice Reference Experiments data set the atmospheric feedbacks are further disentangled in a heat flux equation decomposition, and errors in the surface wind fields explain a significant portion of the model disparity. The wind driven forcing also tends to play a role in explaining model errors in Atlantic meridional heat transport at 26°N, where a trade off between overturning and wind driven gyre transport takes place. Msadek (2011) found that the slope of the wind driven gyre transport as function of the Atlantic Meriodinal Overturning Circulation strength (another example of an advanced feedback diagnostic) has the wrong sign in the GFDL model. These diagnostic studies are crucial to gain confidence in decadal predictions which critically depend on a right initialization and propagation of anomalies in the ocean component of coupled AOGCMs, and on centennial time scale in which the ocean mixing properties play a crucial role in determining the time scale of the transient climate response to the changed radiative forcing.

The recent trends in surface solar radiation, attributed to global dimming and brightening, provide a useful testbed for evaluation of the representation of the clear sky direct aerosol radiative forcing. CMIP3 models underestimate the amplitude of the dimming/brightening signals particularly over China, Europe and India (Dwyer et al. 2010), which was partly attributed to incorrect aerosol emission scenarios. Allen et al. (2012) revisit this analysis for CMIP5, where in contrast to CMIP3 only a single aerosol emission inventory was used. 14 CMIP5 models with a total of 54 ensemble members were available, and compared to observations from the Global Energy Balance Archive, ISCCP, and surface data sets. Clear sky radiation was calculated by removing radiation variability explained by cloud cover. In spite of the increased consistency in the aerosol fields, the dimming trend was not well captured by CMIP5: the timing of the reversion from dimming to brightening in Europe was about right, but the amplitude both in Europe and China is still too small. The conclusions are robust after correcting for cloud cover in the observations and models. It is of interest to explore the ability of detailed radiation process models in simulating the dimming and brightening features.

5 Summary and Recommendations for WCRP

Since CMIP3 significant progress has been made in the design of multi-model experiments, the interpretation of model spread, the availability and usage of observations, and the diagnostics of complex processes and their interactions. As the

CMIP5 archive is filling up these analyses will further develop. Many other studies, not reported here, are underway along the lines sketched above. The coordinating role of WCRP in these developments has been very beneficial. But where can WCRP play a further role?

The evident increase in the level of sophistication in the practice of model evaluation, data collection and development of diagnostics and experimental design is clearly reflected in a number of recent WCRP coordination meetings and documents, such as the world modeling summit (2008), and the WCRP modeling coordination meeting (2010). Many recommendations documented in the workshop reports call for enhanced collaboration, particular focus areas and promotion of development of e.g. better observational data bases. Here we will review these recommendations in the light of the discussions and developments reported previously.

The current overall organizational design of WCRP is well targeted to establish the required improved collaboration of experts with a different disciplinary background. Specifically, improved links should be encouraged between observationalists and modelers, NWP and climate model developers, and physical and statistical experts. Links between observation and model experts should be organized around the development of agreed observable model evaluation diagnostics and performance metrics (see e.g. the "Good practice paper" by Knutti et al. 2010). Frequents meetings between model developers and application experts can benefit from a cross-fertilization of common practices in these communities. Frequent evaluation of climate models using data assimilation and routine observations, as commonly applied in NWP, can help target the most important biases and their causes in climate models. Long climate integrations can highlight systematic shortcomings in NWP systems normally masked by routine application and model state adjustments. The concept of seamless prediction is a fruitful research area where climate and NWP applications are joined. And finally, the involvement of statisticians is important to improve the detection and evaluation of extreme events in the model suite.

Another step forward is the identification of a number of key model deficiencies. An inventory over >100 experts, carried out in 2010, revealed a number of persistent shortcomings in model performance, that urgently need improvement. The issues mentioned most frequently were:

- tropical variability and biases
- moist processes (clouds, convection, precipitation)
- carbon cycle and land/ocean–atmosphere coupling
- troposphere-stratosphere interaction
- formulation of physics in high resolution models.

The first three topics are well covered in the studies reported previously in this paper, while the agreed need to improve the predictability of the atmospheric circulation, and the representation of extremes are reflected by the last two topics. Improvements in these areas require an increased investment in model development capacity (Jacob 2011), but also require improved experimental design (e.g. initialized forecasts, specific feedback experiments, experiments aimed at describing specific (extreme) events) and diagnostics. The WCRP working group structure is well capable for designing these focused studies.

Finally, WCRP can continue to play its role as ambassador aiming at targeting funding resources, improving interdisciplinary links and engaging experts and students. It can do so by organizing targeted conferences and sessions, and provide input to circuits where decisions are being made. Important overarching targets are for instance:

- the continued need to close the gaps between observations and models
- the continuation of the collection and storage of high-quality homogenized observation records
- the design of focused field observation studies
- the involvement of the NASA and ESA climate initiatives
- the call for focused modeling studies
- the promotion of development of comprehensive Earth System Models including components of e.g. the biosphere, cryosphere, and human dimensions, requiring strengthened links with IGBP and IHDP.

Most recommendations require efforts from the researchers in the fields: submit targeted research proposal, commit to coordinated activities, maintain or improve the interdisciplinary network. By its organizational design with its working groups and conference sessions, WCRP can synchronize the activities of the wide range of involved researchers, and as such help in improving the important understanding of our environment.

References

Allen RJ, Norris JR, Wild M (2012) Evaluation of multidecadal variability in CMIP5 surface solar radiation and inferred underestimation of aerosol direct effects. Submitted to J Geophys Res

Anav A, Friedlingstein P, Kidston M, Bopp L, Ciais P, Cox P, Jones C, Jung M, Myneni R, Zhu Z (2013) Evaluating the land and ocean components of the carbon cycle in the CMIP5 Earth System Models. J Climate. doi:10.1175/JCLI-D-12-00417.1 (in press)

Bates SC, Fox-Kemper B, Jayne SR, Large WG, Stevenson S, Yeager SG (2012) Mean biases, variability, and trends in air-sea fluxes and SST in the CCSM4. J Climate 25:7781–7801. doi:10.1175/JCLI-D-11-00442.1

Bodas-Salcedo A and Coauthors (2011) COSP: satellite simulation software for model assessment. Bull Am Meteorol Soc 92:1023–1043. doi:10.1175/2011BAMS2856.1

Branstator G, Teng H (2010) Two limits of initial-value decadal predictability in a CGCM. J Clim 23(23):6292–6311. doi:10.1175/2010JCLI3678.1

CLIVAR (2011) WCRP Coupled Model Intercomparison Project – Phase 5 – CMIP5 –, CLIVAR exchanges, Special issue no 56, vol 16(2), May 2011

Cox PM, Pearson D, Booth BB, Friedlingstein P, Huntingford C, Jones CD, Luke CM (2012) Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability. Nature 494:341–344. doi:10.1038/nature11882

Dufresne J-L, Bony S (2008) An assessment of the primary sources of spread of global warming estimates from coupled atmosphere–ocean models. J Clim 21:5135–5144

Dwyer JG, Norris JR, Ruckstuhl C (2010) Do climate models reproduce observed solar dimming and brightening over China and Japan? J Geophys Res 115:D00K08. doi:10.1029/2009JD012945

Friedlingstein P et al (2006) Climate–carbon cycle feedback analysis: results from the C4MIP model intercomparison'. J Clim 19(15):3337–3353

Guilyardi E, Cai W, Collins M, Fedorov A, Jin F-F, Kumar A, Sun D-Z, Wittenberg A (2011) New strategies for evaluating ENSO processes in climate models. BAMS. doi:10.1175/BAMS-D-11-00106.1

Hall A, Qu X (2006) Using the current seasonal cycle to constrain snow albedo feedback in future climate change. Geophys Res Lett 33:L03502. doi:10.1029/2005GL025127

Hargreaves HC, Paul A, Ohgait R, Abe-Ouchi A, Annan JD (2011) Are paleoclimate model ensembles consistent with the MARGO data synthesis? Clim Past Discuss 7:775–807. doi:10.5194/cpd-7-775-2011

Hawkins E, Sutton RT (2009) The potential to narrow uncertainty in regional climate predictions. BAMS 90:1095. doi:10.1175/2009BAMS2607.1

Huber M, Mahlstein I, Wild M, Fasullo J, Knutti R (2011) Constraints on climate sensitivity from radiation patterns in climate models. J Clim 24:1034–1052. doi:10.1175/2010JCLI3403.1

Jacob C (2011) From regional weather to global climate; oral presentation at OSC. http://conference2011.wcrp-climate.org/abstracts/jackob_A4.pdf

Jiang JH, Su H, Zhai C, Perun VS et al (2012) Evaluation of cloud and water vapor simulations in CMIP5 climate models using NASA A-train satellite observations. J Geophys Res 117(D1410):24 pp. doi:10.1029/2011JD017237

Jung M, Reichstein M, Bondeau A (2009) Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model. Biogeosciences 6:2001–2013

Knutti R (2008) Should we believe model predictions of future climate change? Trienn Issue Earth Sci Philos Trans R Soc A 366:4647–4664. doi:10.1098/rsta.2008.0169

Knutti R et al (2010) Good practice guidance paper on assessing and combining multi model climate projections. In: Stocker TF, Qin D, Plattner G.-K, Tignor M, Midgley PM (eds) Meeting report of the Intergovernmental Panel on Climate Change expert meeting on assessing and combining multi model climate Projections, IPCC Working Group I Technical Support Unit, University of Bern, Bern, Switzerland

MARGO Project Members (2009) Constraints on the magnitude and patterns of ocean cooling at the Last Glacial Maximum. Nat Geosci 2:127–132. doi:10.1038/ngeo411

Masson D, Knutti R (2011) Climate model genealogy. Geophys Res Lett 38:L08703. doi:10.1029/2011GL046864

Matei D, Baehr J, Jungclaus JH, Haak H, Müller WA, Marotzke J (2012) Multiyear prediction of monthly mean atlantic meridional overturning circulation at 26.5°N. Science 335:76–79. doi:10.1126/science.1210299

Msadek R (2011) Comparing the meridional heat transport at 26.5ºN and its relationship with the MOC in two CMIP5 coupled models and in RAPID-array observations (oral presentation WCRP OSC Denver, Oct 2011)

Reichler T, Kim J (2008) How well do coupled models simulate today's climate? Bull Am Meteorol Soc 89:303–311

Sakaguchi K, Xubin Z, Brunke MA (2012) Temporal- and spatial-scale dependence of three CMIP3 climate models in simulating the surface temperature trend in the twentieth century. J Clim 25:2456–2470. doi:10.1175/JCLI-D-11-00106.1, http://dx.doi.org/

Schmittner A, Urban NM, Shakun JD, Mahowald NM, Clark PU, Bartlein PJ, Mix AC, Rosell-Melé A (2011) Climate ensitivity estimated from temperature reconstructions of the last glacial maximum. Science 334(6061):1385–1388. doi:10.1126/science.1203513

Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. Bull Am Meteorol Soc 93:485–498

Teixeira J, Waliser D, Ferraro R, Gleckler P, Potter G (2011) Satellite observations for CMIP5 simulations. CLIVAR Exchanges No. 56, 16(2) May 2011

Williams KD, Webb MJ (2009) A quantitative performance assessment of cloud regimes in climate models. Clim Dyn 33:141–157. doi:10.1007/s00382-008-0443-1