Assessing the reliability of climate models, CMIP5
Discussion paper for the WCRP Open Science Conference Oct 2011
Session B7

Version 29 August 2011

Bart van den Hurk, KNMI
Pascale Braconnot, IPSL
Veronika Eyring, DLR
Peter Gleckler, PCMDI

Scope of this document
Several recent meetings encapsulate the challenges facing climate modelling that are exemplified in the WCRP OSC session entitled "Assessing the reliability of climate models, CMIP5". The report of the 2008 "World Modelling Summit for Climate Prediction[1]" (ECMWF, Reading, UK) declared that "a revolution in climate modelling is necessary" (Shukla et al, 2009) and provides specific recommendations for the development and application of climate models. The more recent Intergovernmental Panel on Climate Change (IPCC) Expert Meeting on "Assessing and Combining Multi Model Climate Projections[2]" held in Boulder in Jan 2010 (Knutti et al, 2010) provides recommendations for good practice in using multi-model ensembles for detection and attribution, model evaluation and global climate projections as well as regional projections relevant for impact and adaptation studies. In parallel to these expert meetings, the WCRP is reorganizing its coordinating role to establish the WCRP Modelling[3] and Data council.. The need to create better infrastructure to facilitate access to observations for model development and evaluation was notified and among other topics, the synergism of models with observations was promoted (Eyring et al. 2010).

As introduction to the WCRP Open Science Conference session B7, this document contains summaries of the reports that were produced after these meetings. The selected statements and recommendations listed below are expected to stimulate and guide discussions associated with Session B7.

The World Modelling summit for climate prediction
The World Modelling Summit for Climate Prediction was organized around 5 themes:

- *Overview* (societal drivers; current status of weather and climate modeling; strategies for seamless prediction; crucial hypotheses)
- Prospects for *next-generation modeling systems* (balance between resolution and complexity; balance between multi-model and unified modeling framework; issues of parameterizing unresolved scales and regional models)
- Prospects for current *high-end computer systems* and implications for model code design
- *Strategies for model evaluation*, modeling experiments, and initialization for prediction of the coupled ocean-land-atmosphere climate system.
- Strategies for enhancing human and *computing resources* (requirements and possible organizational frameworks).

Shukla et al (2009) reported on this meeting in BAMS, issuing a summit statement on the Climate Prediction Project. Some elements in this summit statement that are related to the assessment of the reliability of climate models are listed below:

- *Collaboration NWP and climate research* – Advances in climate prediction will require close collaboration between the weather and climate-prediction research communities. Climate models will need to be tested in subseasonal and multiseasonal prediction mode also including use of the existing and improved data assimilation and ensemble prediction systems.
- *Process-based model evaluation* — For instance, the effects of using different cloud parameterizations and varying the parameters in those parameterizations should be evaluated not only against global-scale satellite data, but also against what is known about basic cloud processes. High-resolution model experiments are required for developing and testing parameterizations. Similar arguments hold for other parameterizations (land, turbulence, radiation, …).

[1] http://wcrp.wmo.int/documents/WCRP_WorldModellingSummit_Jan2009.pdf
[2] https://www.ipcc-wg1.unibe.ch/publications/supportingmaterial/IPCC_EM_MultiModelEvaluation_MeetingReport.pdf
[3] http://wcrp.ipsl.jussieu.fr/Workshops/ModellingMeeting/index.html

- *Data assimilation, analysis, and initialization* — The methods of data assimilation and initialization, which have been crucial for the success of numerical weather prediction, must be extended to coupled models to obtain physical consistency. Observing system experiments are required to develop rational strategies for defining the most effective and efficient data streams. Model bias must be reduced and corrected in assimilation.
- *Ensembles*—Large ensembles of high-resolution simulations are required for the quantification of forecast skill. This computing challenge is particularly acute at the decadal prediction scale.
- The initiation of a *Climate Prediction Project* coordinated by the World Climate Research Programme, in collaboration with the World Weather Research Programme and the International Geosphere–Biosphere Programme, and involving the national weather and climate centers as well as the wider research community, is highly recommended. This Climate Prediction Project should build a world climate research facility that will significantly enhance the capacity of the world's existing weather and climate research centers for prediction of weather and climate variations, particularly including the developing countries whose national capabilities need to be increased substantially. The climate research facility will bring together computer resources, sophisticated climate models (including biogeochemical components) and an archive of observations and model data.

The Expert Meeting on Assessing and Combining Multi Model Climate Projections

The Good Practice Guidance Paper of the Expert Meeting on Assessing and Combining Multi Model Climate Projections (Knutti et al. 2010) seeks to briefly summarize methods used in assessing the quality and reliability of climate model simulations and in combining results from multiple models. It provides recommendations for good practice in using multi-model ensembles for detection and attribution, model evaluation and global climate projections as well as regional projections relevant for impact and adaptation studies. It illustrates the potential for, and limitations of, combining multiple models for selected applications. Criteria for decision making concerning model quality and performance metrics, model weighting and averaging are recommended. However, specific recommendations regarding which performance metrics to use are not provided, since this will need to be decided for each application separately. A selection of important notions and recommendations is given below:

- Most multi-model model intercomparisons are not designed to yield formal error estimates but are in essence "ensembles of opportunity". The spread of such a multi-model ensemble is rarely a direct measure of uncertainty, but it can help to characterize uncertainty. When one assumes that *each ensemble member is sampled from a distribution centred around the truth* ('truth plus error' view), one expects perfect independent models in an ensemble to be random draws from a distribution centred on observations. Uncertainties in predictions should tend to zero as more models are included. However, assuming that *each member is considered to be 'exchangeable' with the other members and with the real (observed) system*, observations are viewed as a single random draw from an imagined distribution of the space of all possible outcomes of Earth's chaotic processes. A 'perfect' independent model in this case is also a random draw from the same distribution, and so is 'indistinguishable' from the observations in the statistical model. In this view uncertainties are expected to converge to a value related to the size of the distribution of all outcomes. Forming and interpreting ensembles for a particular purpose requires an understanding of the variations between model simulations and model set-up
- The reliability of projections might be improved if models are weighted according to some measure of skill and if their interdependencies are taken into account, or if only subsets of models are considered. Since there is little opportunity to verify climate forecasts on timescales of decades to centuries (except for a realization of the 20th century), the skill or performance of the models needs to be defined, for example, by comparing simulated patterns of present-day climate to observations. Such performance metrics are useful but not unique, and often it is unclear how they relate to the projection of interest. Process-based performance metrics might be derived from the analysis of multi-model simulations and/or from process studies performed in projects that complement CMIP. Researchers are encouraged to consider the different standardized model performance metrics currently being developed.
- It is problematic to regard the behavior of a weighted model ensemble as a probability density function (PDF). The range spanned by the models, the sampling within that range and the statistical interpretation of the ensemble need to be considered. Examples of performance metrics that can be used for weighting are those that are likely to be important in determining future climate change. It should be recognized that additional

forcings and feedbacks, which may not be fully represented in global models, may be important for *regional* climate change (e.g., land use change and the influence of atmospheric pollutants).

The WCRP Modelling Coordination Meeting

The WCRP Modelling Coordination Meeting was held in November 2010 in Paris, France. The purpose of this meeting was to follow-up on the recommendations of the World Modelling Summit and actions on the subject of modelling resulting from the JSC meeting that was held in Antalya (Turkey), February 2010. The JSC discussed at length the formation of a WCRP Modelling Council in the context of WCRP visioning and future directions to carry out the following functions:

- Promoting the confrontation of models with observations and results of process studies (Eyring et al. 2010);
- Promoting collaboration amongst various climate science communities (includes numerical weather prediction (NWP), seasonal to interannual prediction and climate projection communities as well as those dealing with biogeochemistry, air quality, terrestrial ecology, etc.) (Kirtman et al. 2010);
- Promoting application of models to problems of societal relevance, quantifying uncertainties and making sure they are well communicated and understood (Palmer et al. 2010);
- Promoting the model development and improvements (Jakob et al., 2010).

Possible ways to accelerate and expedite model-observation interactions within WCRP were discussed and recommendations were given that WCRP should in particular:

- Reduce the gaps between modelling, observations and assimilation communities and promote the use of multiple datasets in model development and evaluation;
- Foster collaboration among the WCRP working groups and projects;
- Reduce the gaps between NWP/seasonal/climate communities;
- Promote the development of methods to identify the key players in model error at the process level;
- Engage the potential of high resolution modelling into minimization of risks of extreme climate events and more effective coping with risks. The 'WCRP-UNESCO Workshop on metrics and methodologies of estimation of extreme climate and weather events' workshop recommended:

These bullets are detailed in Eyring et al. (2010). Specific roles of the Data Council (Gille et al., 2011) are likely to focus on coordinating access to data and meta-data, data product assessment or intercomparison, and exchange between modelers and observationalists (e.g. via the Modeling Council).

References

Eyring, V., M. Manton, D. Stammer, and K. Steffen, Promoting the synergism models with observations and results of process studies, Discussion Paper for WCRP Modelling Coordination Meeting, http://wcrp.ipsl.jussieu.fr/Workshops/ModellingMeeting/, 2010.

Gille, S., M. Manton, M. Bosilovich, A. Simmons, T. Koike, K. Taylor, C. Kummerow, C. Jones, V. Eyring, J. Key, M. Hegglin, and N. McFarlane, Report of the Task Team to develop the transition plan from WOAP to a WCRP Climate Observations, Analysis and Information Council, WCRP internal document, 2011.

Jakob, C., K. Trenberth, T. Shepherd, and D. Bromwich, What is model development and how is it approached? Discussion Paper for WCRP Modelling Coordination Meeting, http://wcrp.ipsl.jussieu.fr/Workshops/ModellingMeeting/, 2010.

Kirtman, B., V. Ramaswamy, J. Slingo, V. Kattsov, Promoting collaboration amongst various climate modeling community, Discussion Paper for WCRP Modelling Coordination Meeting, http://wcrp.ipsl.jussieu.fr/Workshops/ModellingMeeting/, 2010.

Knutti, R., G. Abramowitz, M. Collins, V. Eyring, P.J. Gleckler, B. Hewitson, and L. Mearns, Good Practice Guidance Paper on Assessing and Combining Multi Model Climate Projections. In: Meeting Report of the Intergovernmental Panel on Climate Change Expert Meeting on Assessing and Combining Multi Model Climate Projections [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, and P.M. Midgley (eds.)]. IPCC Working Group I Technical Support Unit, University of Bern, Bern, Switzerland, 2010.

Palmer, T., C. Jones, J. Hurrell, J Petch, G. Danabassoglu, M.Latif, Promoting application of models to problems of societal relevance, quantifying uncertainties and ensuring they are well communicated and understood, Discussion Paper for WCRP Modelling Coordination Meeting, http://wcrp.ipsl.jussieu.fr/Workshops/ModellingMeeting/, 2010.

Shukla, J., R. Hagedorn, M. Miller, T. N. Palmer, B. Hoskins, J. Kinter, J. Marotzke, J. Slingo, 2009: Strategies: Revolution in Climate Prediction is Both Necessary and Possible: A Declaration at the World Modelling Summit for Climate Prediction. Bull. Amer. Meteor. Soc., 90, 175–178; doi: 10.1175/2008BAMS2759.1