Report on a workshop on AR5 science

Nov 17-19, 2008, New York

Background

The increasing sophistication of multi-model climate simulation databases has created important opportunities for new scientific advances but also problems in data handling and in the appropriate use of the archives. There is also a lag in the understanding of the theoretical underpinning of using multi-model ensembles. This workshop was convened in order to assess the potential for further advances in the use of these datasets prior to the accumulation of model simulations for the IPCC Fifth Assessment Report (AR5).

Executive Summary

The coordination of coupled general circulation model simulations in support of the IPCC by the WGCM committee and their curation at PCMDI has been a tremendous boost to the study of climate change for assessing the future likelihood of further change.

In particular, the simulations made in support of IPCC AR4 (the CMIP-3 data) have been the most successful data sharing activity in climate science (with thousands of registered users, 500+ papers written, and hundreds of terabytes of data shared).

The outreach achieved by the public accessibility of these simulations has exceeded all expectations and opened out the analysis of climate models to a much wider academic community than has been seen before.

Despite the multiple successes of CMIP3, the workshop identified a number of areas where improved database support, software infrastructure and coordination of techniques to interpret diverse model outputs could improve science outcomes:

-More sophisticated database services (server side analyses, automatic calculation of specific diagnostics) would reduce download times and provide more useful services, which would enable an expansion and enhancement of science projects associated with the database. The plans for the Earth System Grid (ESG) go some way towards addressing this.
-Documentation of the contents of the archive and the meta-data associated with the models could be greatly improved (c.f. the EU METAFOR project).
-New models for communications between users and originators of data should be developed to encourage more effective integration of new results into model development.
-The theoretical or practical basis for dealing with multiple models of varying skill is as yet at a very primitive level of development. Nevertheless, as we move to a situation where the details of future climate change become more important in the policy and planning arenas (for example, in adaptation to unavoidable climate change) then such

methods will become increasingly important.
-Some CMIP3/AR4 studies have developed methods of rejecting, ranking and weighting different models using algorithms of varying degrees of complexity. More effort should be expended on synthesising 'best practice' before the AR5 (CMIP5) analyses start in earnest.
-Issues with originator initiated restrictions on access have slowed wider use of the data, even for groups (such as for the US modelling centers) with no nominal access restrictions (ESG should include this functionality). Access should instead be determined on a center-by-center basis, and a revisiting of what restrictions are actually required is needed.
-Data products derived from the underlying data, and toolkits for these analyses would be very useful for the wider community but curating these derived data, ensuring that proper credit is acknowledged, and validating their provenance and accuracy needs to be carefully thought through.

The size of the CMIP5 archive will be perhaps two orders of magnitude larger than the CMIP3 archive, necessitating new distributed infrastructure which will create both problems and opportunities for the data centers.

Many of the above issues have been anticipated in the planning for CMIP5 and are being addressed in the current design of the next generation system for collecting and distributing model output (ESG, Williams et al, 2009), but others require fresh ideas.

For CMIP5, we recommend a new Coupled Model Ensemble Project (CMEP) funding effort to ensure that key diagnostics can be calculated and analysed in plenty of time for AR5. As one of the conditions for new funding, it might be useful to request archiving of all these basic diagnostics and toolkits for wider use in a central repository, perhaps at PCMDI.

Workshop report

A workshop addressing science issues arising from the multi-model archives created in support of the IPCC 4th Assessment Report and future reports was held at NASA GISS, New York in November 2008. It gathered model developers, database curators, users of the data and program managers to assess the successes of the CMIP3 archive and the prospects for future improvements and expansion in preparation for CMIP5 in support of the next IPCC report.

Introduction:

Since the advent of multiple climate modelling groups in the early 1990s there has been an interest in comparing diagnostics and sensitivities of the various models. The first model intercomparison project was AMIP (Gates et al, 1999) for atmospheric models, and this was followed a few years later by the first coupled model intercomparison project (CMIP) (Covey et al, 2003). These projects, generated and curated by the community of climate modellers, were designed to provide benchmarks of model performance and tried to provide insight into the varying sensitivities of different models. However, the projects were somewhat limited by the few diagnostics available and the partial and intermittent participation of the main groups. This allowed the identification of different behaviours among the models at particular points in time, but often did not allow for a thorough examination of the underlying reasons for the differences.

At around the same time, the Intergovernmental Panel on Climate Change (IPCC) began assessing the state of our understanding of the climate system and its sensitivity to anthropogenic forcings. Results from the model intercomparison studies were discussed in the 1995 SAR and were extensively used in the TAR (2001). In particular, the results from CMIP2 figured highly in the model assessment, detection and attribution and future projection chapters. Other modelling results appeared throughout the reports but it was soon appreciated that the robustness of climate model results was greatly enhanced if there were multiple models that had the same basic sensitivity.

In the run-up to the IPCC AR4, the committee that facilitates interactions among modelling groups (WGCM) realized that there was an opportunity and a need to greatly expand the scope of the coupled model archives (Meehl et al, 2007). The opportunity arose from the greatly expanded data storage and bandwidth capabilities coming on line and the understanding among model developers that these archives were providing an ever more important source of climate model results to the wider community. The need to provide more extensive diagnostics over a wider range of experiments came from the increasing complexity of the models and their forcings and wider sets of observational data that were becoming available for comparison to model outputs. No longer were model experiments only run with a single change in CO_2 or the total solar irradiance, but included changes in ozone, aerosols, land use, volcanic effects and multiple greenhouse gases. Much greater fidelity to the observational record was possible with that generation of models, and so the usefulness of comparing model output to the actual transient behaviour of the climate increased enormously. Thus the specifications for the CMIP3 experimental protocol were much more sophisticated than previously (including historical hindcasts as well as idealised benchmark simulations) , the data uploaded was greatly in excess of previous efforts and participation among modelling groups was almost universal - including many groups that participated in such efforts for the first time.

Standards put in place by WGCM (common data formats, quality control at the post-processing stage and at PCMDI) helped ensure that available data were free of obvious problems. Due in no small part to the tireless efforts of the group at PCMDI, the archive was developed and made available for researchers in late 2004. It contained some 36 Terabytes in 84,000 separate files.

To encourage some initial analyses, the US funding agencies (NSF, NASA, DOE and NOAA) provided small seed-money grants (CMEP) to two dozen investigators and held a workshop in May 2005 to discuss results. Even at that point it was clear that the interest in the archive from the wider academic community (i.e. not just among people associated with the contributing groups, and not only those directly funded to look at it) was very wide and indeed much more so than for any previous model intercomparison project. Over 500 Tb of data have been downloaded to date.

More remarkable perhaps is that while a large impetus for organising the CMIP3 archive was to provide assessments that could be used by the IPCC AR4, downloads actually increased dramatically (by roughly a factor of two) *after* the report was published and have remained at that high level subsequently (about half a terabyte a day).

As of early 2009, the number of published studies using CMIP3 data has reached more than 500, with thousands of registered users of the data, and multiple portals for the data have been developed. It is clear that the limited investments that were made to directly encourage use of this data and its curation have been one of the most cost-efficient ways to leverage the large investments made in model development into useful science (Meehl et al, 2007).

Issues arising:

Given the much wider use made of CMIP3 and its wider scope, a number of problems arose that had not been so prevalent or evident in earlier projects. These fell into roughly 4 categories: the quality of the documentation and metadata; access to and magnitude of the data downloads; the feedback between users, other users and the originating groups; and scientific strategies for dealing with multi-model ensembles themselves. Many of these issues will be tackled in the current design of the next generation system (Williams et al., 2009).

The 20th Century climate change model simulations (20C3M) were the most complex of the archived simulations, but since each modelling group had designed their simulations independently, different forcings were used across models. Some forcings were found in some simulations and not others (stratospheric ozone decline for instance). The actual magnitude of forcings (a function both of the input data, the model code and the background simulated climate) were not mandated, and inevitably there were some mistakes in the configuration of these forcings (one modelling group placed their volcanic aerosols much higher in the stratosphere than was appropriate, another had initially incorrectly specified the trend in ozone depletion). Furthermore, errors in the some actual diagnostics (incorrect variables, varying definitions) have also been found. Documentation of these issues has been piecemeal and not systematically archived. No methodology existed to notify users of potentially flawed data that updates were available, or that issues had been raised. Versioning of corrected data (whether due to corrupted diagnostics or incorrectly specified simulations) was rather ad hoc.

Only the raw data are available from the original PCMDI archive; global or zonal means, standard indices, climatologies etc. had to be derived separately by each user. This is a waste of time and downloading bandwidth if the user is only interested in the derived index. Obvious manipulations of the simulations (stringing together the 20th Century simulations to the continuations under various future scenarios, or removing any control run drift, for instance) was left to individual users, increasing the chances for errors and unreported differences.

Feedback between users and model groups was also lacking. Many results using the archive were only made known to the originating groups as a final product in the technical literature, sometimes years after the models were run. As well as the difficulty in keeping up with the sheer number of papers published, these results often came too late to influence model development prior to the settling on the AR5 configurations. The inability to tap users' expertise for similar follow-up analyses on newer versions of the models was also lacking.

Most importantly from a scientific view point, the methodology and framework for dealing with a non-random ensemble of opportunity is still at a primitive stage. While the overall multi-model mean has been shown to give a better climatology than any individual model (Gleckler et al., 2008; Reichler and Kim, 2008), it is not clear that the multi-model mean gives more accurate projections. For instance, many skill scores based on current climate or its variability are uncorrelated with changes in future projections, and thus are not useful in reducing prediction uncertainty. Model selection algorithms based on cross-validation or sensitivity studies have up until now been rather ad hoc, while the development of model weighting strategies is still at the experimental stage. While we recognized that it will not be possible to agree on an internationally agreed set of standards for rejecting, ranking, or weighting models on the time scales of AR5, this should not deter us from starting this endeavor. Bringing together researchers to discuss these issues is a relatively simple way of ensuring that common pitfalls are avoided.

Data Access and Downloads:

One additional issue that has been subject to much discussion is the issue of data access. The diverse bodies which provide policy guidance for individual model centers have very different structures and missions. This has led to a somewhat confusing and occasionally frustrating process for accessing data. For instance, the US modelling centers are all run by US Government entities (NASA, NSF, NOAA) and thus all of their output is in the public domain. On the other hand, some centers such as the UK Hadley Centre have very specific missions to commercialize their model output if it is used for 'non-scientific' purposes. The initial users of the PCMDI archive thus had to go through a vetting process to determine whether their aim in examining the data was 'for research purposes', even for US Govt. data that was actually provided with no restrictions. Additionally, 'research purpose' was defined almost exclusively as the preparation of scientific papers in the technical literature. It did not include data analysis for public education, for presentation on Wikipedia, or any other kind of casual use.

Subsequent changes to the PCMDI registration, and the transposition of some of the data to new data archives (in particular ClimateExplorer available at the Royal Dutch Meteorological Office (KNMI), and Dapper available at NOAA PMEL) have partially alleviated this problem, but similar issues will arise as the AR5 data come online. There appears to us to be no good reason to restrict access to data that are nominally already in the public domain, nor why the center with the most restrictive access policies should determine the access levels for any other center's data.

We strongly recommend that the data access restrictions be reduced as far as possible and that all uses of the data be implicitly allowed except where expressly restricted – especially the potential use of this data for education, outreach and other public, but non-academic, purposes. No usage restrictions should be placed on information already in the public domain. Early clarification from modelling groups what restrictions need to be in place and center specific controls put in place needs to be prioritized before the data start being delivered.

Derived products:

One of the most important principles in science is that other scientists should be able to 'stand on the shoulders of giants' in order to further progress and understanding. For this to happen in this case, there needs to be a mechanism by which analyses that are done once and published are made available either as a recipe (or code) for others to apply to new sets of model output, or as data sets that would save others the necessity of calculating derived quantities from first principles. This would include straightforward diagnostics such as the Niño 3.4 index for each model run, or a frequency analysis of a model's tropical variability, or a simulated satellite diagnostic or a drift-corrected temperature change. None of these are necessarily trivial or without some ambiguity, but we consider it appropriate to at least provide a first cut which can subsequently be built on. However, in allowing access to derived data (or the programs to derive the data) a number of principles need to be upheld. Firstly, the intellectual property associated with the analysis needs to be clearly associated with the originator and credit and citations to the methodology need to be correctly assigned. Secondly, there needs to be a site or mechanism where such derived quantities can be stored and verified. If any derived quantities are going to be the basis for further work, ways of validating the correctness of the original work must be available.

Suggestions for future data archives:

In order to alleviate data download bottlenecks, to allow for efficient building on previous work and to provide maximum and timely results to the modelling centers, we propose a number of features for future databases or portals. Many of these are already under consideration or development (i.e through the Earth System Grid), but may require further resources to fulfill.

- Databases and portals need to be able to provide significant server-side analyses. This reduces download requests in the situations where only a global mean or regional analysis is required. Ideally, some scripting environment (nco (NetCdf Operators) or 'R' would be obvious choices) and the possibility of allowing both public and private scripts to be stored would be highly useful.

- Documentation, commentary and corrections of data should be an ongoing process (though one that is moderated). Explorations of ideas such as a wiki, or rss feeds for model data for all users should be explored. These tools could provide automatic notifications of any corrections, or any descriptions of the data that would be useful for others (whether it was a description of forcings used or an assessment of the magnitude of the model's ENSO variability). This could provide a much more timely notification of data analyses than is currently available (using the standard publication system).

- Data versioning should be made standard. Whether this uses a doi or another system, distinct versions of data (or derived data) need to be web-accessible, along with pointers to updated or corrected data and forward citation to publications that used it.

- Databases could save a lot of time by pre-computing a small number of representative standard indices (global means, ENSO or NAO indices). Similarly, facilities for subtracting suitable control runs and linking different runs to their initial conditions would be enormously useful.

Priorities for funders:

We strongly support funding for CMEP-like activity for AR5 as was done for AR4. In addition, we recommend that supported projects be required to archive their derived data or key scripts upon acceptance of papers.

We also recognise that post-processing of the amount of data being requested by CMIP5 is not necessarily a core competency of the model development groups. Funders need to be ready to help out with support so that the data originators can appropriately serve their data.

We especially recognise that many users and potential users of the climate model output are not traditional climate modellers. Special efforts need to be made for these users to ensure that key diagnostics are available in, for instance, Excel-ready formats, or that straightforward mechanisms for downscaling to local or regional scales are available.

We also note that many of the issues raised here are common to a number of existing products on the Internet. Issues of 'trust', appropriate citation, data version control, security, mirroring integrity are common in many applications. More effort could certainly be made to tap into existing experience within the wider Internet community.

Overall, we are extremely excited for the prospects for CMIP5 and we anticipate that significant improvements will occur in data analysis, model-observation comparisons and community building as a result. Funders and scientific working groups have a major role to play in setting the priorities and standards for this effort.

References:

Covey, C., K. M. AchutaRao, U. Cubasch, P. Jones, S. J. Lambert, M. E. Mann, T. J. Phillips, and K. E. Taylor, 2003. An overview of results from the Coupled Model Intercomparison Project. *Global Planet. Change,* **37**, 103–133.

Gates, W. L., and Coauthors, 1999. An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bull. Amer. Meteor. Soc.*, **80**, 29-55.

Gleckler, P. J., K. E. Taylor, and C. Doutriaux (2008). Performance metrics for climate models. *J. Geophys. Res*., **113**, D06104, doi:10.1029/2007JD008972.

Houghton, J.T., L. G. Meira Filho, B. A. Callander, N. Harris, A. Kattenburger and K. Maskell, 1995. *Climate Change 1995: The science of climate change*, Cambridge Univ. Press, New York, 572pp

Houghton, J.T, Y. Ding, D. J. Griggs, M. Nouger, P. J. van der Linden, X. Dai, K. Maskell and C. A. Johnson, 2001. *Climate Change 2001: The scientific basis*, Cambridge Univ. Press, New York, 881pp

Meehl G.A., C. Covey, T. Delworth, M. Latif, B. McAvaney, J. F. B. Mitchell, R. J. Stouffer, and K. E. Taylor, 2007. The WCRP CMIP3 Multimodel Dataset: A New Era in Climate Change Research, *Bull. Amer. Meteor. Soc.*, **88**, 1383-1394.

Reichler, T., and J. Kim, 2008. How Well do Coupled Models Simulate Today's Climate? *Bull. Amer. Meteor. Soc,* **89**, 303-311.

Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor and H. L. Miller, 2007. *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, Cambridge, United Kingdom and New York, NY, USA.

Williams, D.N., R. Ananthakrishnan, D.E. Bernholdt, S. Bharathi, D. Brown, M. Chen, A.L. Chervenak, L. Cinquini, R. Drach, I.T. Foster, P. Fox, D. Fraser, J. Garcia, S. Hankin, P. Jones, D.E. Middleton, J. Schwidder, R. Schweitzer, R. Schuler, A. Shoshani, F. Siebenlist, A. Sim, W.G. Strand, M. Su, and N. Wilhelmi, 2009: The Earth System Grid:

Enabling Access to Multimodel Climate Simulation Data. *Bull. Amer. Meteor. Soc.*, **90**, 195–205.

Websites:

Climate Explorer, Royal Dutch Meteorological Institute: `http://climexp.knmi.nl/`
NOAA Dapper (PMEL), `http://dapper.pmel.noaa.gov/`
Program for Climate Model Diagnosis and Intercomparison (PCMDI), `http://www-pcmdi.llnl.gov/`

Authors:

Gavin Schmidt (NASA GISS) with contributions from Karl Taylor (PCMDI) and many other workshop participants.