18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

GCM simulations of a future climate: How does the skill of GCM precipitation simulations compare to temperature simulations?

Johnson, F.M. [1] **and A. Sharma** [1]

[1] *School of Civil and Environmental Engineering, University of New South Wales, Sydney, Australia*
Email: fiona.johnson@student.unsw.edu.au

Abstract: Which General circulation model (GCM) is more accurate? This is a question that has been addressed by many, using a range of assessment criteria and specified regions and time periods. The question we seek to address in this paper is not related to the relative skills of individual GCMs, but their collective skill at simulating a range of commonly used GCM hydroclimatic variable outputs. Hence, we seek to answer questions such as – Are GCM simulated temperatures more accurate than surface level pressures? How poor is the GCM skill at simulating rainfall compared to more stable variables such as temperature or wind speed? And how does this skill vary with region and distance from the coast?

The Variable Convergence Score (VCS) was used to rank hydroclimatic variables based on the coefficient of variation of the ensemble of all models. The VCS is a simple methodology that allows a quantitative assessment of the performance of the models for different hydroclimatic variables. The skill score methodology has been applied to the outputs of multiple GCMs for a range of hydroclimatic variables and future emission scenarios to provide a relative ranking of the performance of the models over Australia. The methodology would be applicable for any region or any variable of interest available as a GCM output.

The variation of model convergence with distance from the coast was examined. It was found for some variables such as temperature, specific humidity and precipitable water that the agreement of the GCMs in their future projections decreases for areas that are further inland. For other variables such as longwave radiation and wind speed, distance from the coast is not a good indicator of model agreement. For these variables there is a strong north-south gradient for model convergence.

The effects of spatial averaging on model convergence were also assessed using the VCS. As expected, the spread of model projections lies closer to the multi-model ensemble mean for increasing levels of spatial averaging. This improvement in skill is more pronounced for variables such as wind speed that show pronounced regional variations. Variables for which the models consistently agree (e.g temperature, surface pressure) or disagree (precipitation) do not show as strong improvement in model convergence for larger spatial scales.

The VCS has been shown to provide information to researchers and policy makers on how much agreement from GCMs we can expect in time and space.

Keywords: *Climate change, model skill, general circulation model (GCM)*

1. INTRODUCTION

General circulation models (GCMs) are useful tools for understanding the dynamics of climate change and assessing likely impacts on infrastructure and natural ecosystems. However with 23 GCMs and multiple ensemble run currently available, we need to understand how predicted impacts may vary depending on the choice of GCM and climatic variables used in a particular assessment. For example, we may want to know if we can have as much confidence in future projections of precipitation as we do in temperature. Do models agree more in their projections in coastal areas or inland areas? What affect does spatial averaging have on the results?

This paper applies the previously developed Variable Convergence Score (VCS) (Johnson and Sharma, 2009) to answer these questions. Climate model evaluation is discussed in the following section, with a summary of the VCS methodology presented in Section 3. Results and applications of the VCS are shown in Sections 4 and 5 respectively. Finally conclusions are drawn with areas for future research highlighted.

2. CLIMATE MODEL EVALUATION

To evaluate the skill of GCM projections, Dessai et al. (2005) suggest that we need to answer two questions:

- Model Performance – how well does a model simulate the observed climate record?
- Model Convergence – how consistent are the predictions from a range of models in space and time?

Model performance has been assessed by many researchers using a variety of techniques, such as the comparison of observed and modelled probability density functions for daily precipitation and temperature (Perkins et al. 2007).

Model convergence is assessed by examining the spread of predictions and is calculated based on the future projections of the models. Research using ensembles of multiple ensembles has generally found that the ensemble mean gives a better forecast than any individual model, particularly if we consider multiple variables (Lambert and Boer, 2001, Tebaldi and Knutti, 2007). When the spread of predictions around the mean is small, then the model convergence is good and we have confidence that the predictions are reasonably insensitive to the choice of model (Tang et al., 2008). However when the model convergence is poor, then the predictions that we get using a particular model could vary markedly from those of a second model.

This is demonstrated in Figure 1 where the projections in 2050 of the mean climate state from 18 realisations of 9 different GCMs are compared for temperature, precipitation and specific humidity at one grid cell. The projections shown are for the SRESA2 emissions scenario, which is reflects a high CO_2 emissions profile for the future (IPCC, 2000). The temperature projections show good consistency between the different models, so our results should not vary too much if we were to use any one GCM for an impact assessment. In contrast, the precipitation projections vary by up to 100% from the smaller to largest values for the future.

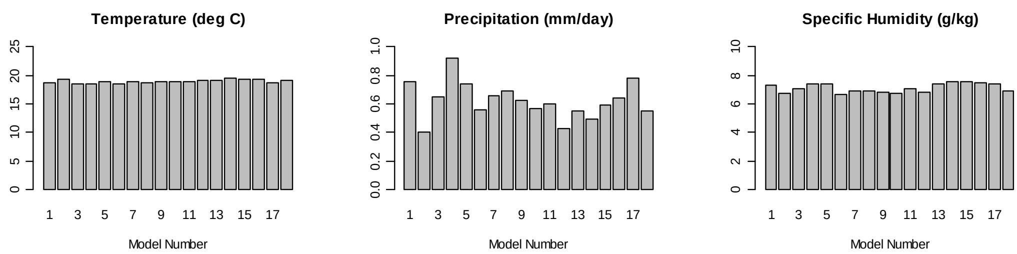

Figure 1. Convergence of 18 different GCM realizations for three climatic variables in 2050

This could have significant implications for water resources infrastructure assessed using different models' projections.

Intercomparison studies have generally shown that different variables are simulated with varying degrees of success by different models and that no particular model is best for all variables and/or all regions (Lambert and Boer, 2001, Gleckler et al., 2008). Although the metrics developed to date are useful, there remain several problems with model evaluation methods, as listed below:

- There is no widely accepted metric for assessing climate models as a whole, due to the number of variables and a lack of observational data (Raisanen, 2007).

- It is unclear how performance in simulating the observed climate translates into future simulations, particularly since calibration of models could potentially hide deficiencies in the modelling of physical processes.
- Observed data sets that are used to evaluate model performance also have uncertainties (Gleckler et al., 2008).

Few studies have compared the reliability of variables rather than individual models. Xu (1999) qualitatively assessed GCM skill in predicting variables relevant to hydrologic impact assessment, concluding that as the importance of a variable to hydrologic impact assessment increases, the ability of the GCMs to simulate the variable decreases. Gleckler et al. (2008) compares the errors in modelling precipitation, mean sea level pressure and surface air temperature in the GCMs which were submitted to the 3rd Coupled Model Intercomparison Projects (CMIP3). They found that the models simulate temperature best, followed by pressure and finally precipitation.

To address these problems, Johnson and Sharma (2009) developed the Variable Convergence Score (VCS) to provide a quantitative measure that can be used to compare GCM skill in predicting a range of variables. The VCS compares the modelling of meteorological variables using GCM projections of future changes. The VCS highlights how well models converge in their projections of each climatic variables, assuming that the multi model ensemble mean provides a reasonable estimate of the likely future climate state.

3. VCS METHODOLOGY

The methodology and rationale for the development of the VCS was presented in Johnson and Sharma (2009). Key features of the methodology are summarised below, and readers are directed to Johnson and Sharma (2009) for the full details.

3.1. VCS calculations

The VCS is designed to allow different climatic variables to be compared across time and space. It there needs to be insensitive to the absolute values of the variables. The coefficient of variation (CV) was therefore used in deriving the skill score. The CV value for a particular grid cell is calculated for the set of mean annual values for a particular climatic variable from the GCM results. The VCS is calculated using the cumulative distribution of the CV values from all locations, variables and time periods. The detailed steps in calculating the VCS are shown below (Johnson and Sharma, 2009):

1. Combine the results from each model for all ensembles for a particular variable to give estimates at each grid cell for each 10 year window as shown in (1)

$$X_{i,i} = \lfloor x_{i,t,1}, x_{i,t,2} \ldots x_{i,t,n} \rfloor \quad (1)$$

where $x_{i,t,n}$ refers to variable v at grid cell i at time t from model n.

2. Calculate the mean and standard deviation and hence the coefficient of variation (CV) at each cell, as in (2)

$$CV_{i,t,x} = \frac{\sigma_{i,t}}{\bar{X}_{i,t}} \quad (2)$$

where $\sigma_{i,t}$ is the standard deviation of $X_{,t}$ and $\bar{X}_{i,t}$ is the mean of $X_{i,t}$ for variable x.

3. Pool the CV values from all 128 grid cells for eight variables, two emission scenarios and three 10 year windows, giving a set of 6144 CV values.

4. Calculate the empirical CDF of CV values, assuming that the pooled CV values come from a common distribution which characterizes the variability of climatic variables, where i is the rank of an individual CV value, and n is the total number of CV values, as shown in (3).

$$F(x) = \frac{1}{n} \times i \quad (3)$$

5. Calculate the skill score, *VCS,* for a particular variable or grid cell according to (4).

$$VCS_{x,i} = 100 \times \left(1 - F(x)\right) \quad (4)$$

Johnson and Sharma, GCM simulations of a future climate: How does the skill of GCM precipitation simulations compare to temperature simulations?

Johnson and Sharma (2009) assessed the sensitivity of the VCS to the use of an empirical distribution for the CV values. They found that the relative performance of different variables was unchanged when different combinations of CV values were used to construct the empirical distribution.

3.2. GCM and variable selection

In theory, the VCS can be calculated using any number of the available GCMs to provide the CV values for the skill score calculations. Johnson and Sharma (2009) demonstrated that for the VCS values to be relatively insensitive to the choice of specific GCMs included in the calculations at least 15 GCM integrations should be used. For the results presented below, we have used integrations/ensembles from 9 GCMs available from the World Climate Research Programme's (WCRP's) Coupled Model Intercomparison Project Phase 3 (CMIP3) multi-model dataset. The 9 GCMs used in the calculations were chosen as they have been shown to perform well over Australia (Perkins et al. 2007, CSIRO and BOM, 2007). Details of the models used in the calculations are provided in Table 1. The GCM results from two emission scenarios, SRESA2 and SRESB1, have been used. In total we have used 21 ensembles from the SRESB1 scenario and 18 sets of results for the SRESA2 scenario from the 9 different GCMs. Future work will extend the analysis to include all 23 available GCMs and ensemble results.

Table 1. Details of GCMs used for calculations

Full Model Name	Avail. SRES Scenarios	Available Ensembles	Approx Grid Resolution
bccr_bccm2_0	A2, B1*	1	175 km
cccma_cgcm3_1_t47	A2, B1	5	250 km
cnrm_cm3	A2, B1	1	175 km
csiro_mk3_5	A2, B1	1	175 km
iap_fgoals1_o_g	B1	3	300 km
inmcm3_0	A2, B1	1	400 km
ipsl_cm4	A2, B1	1	275 km
miroc3_2_medres	A2, B1	3	250 km
mri_cgcm2_3_2a	A2,B1	5	250 km

* A2 refers to SRESA2 and B1 to SRESB1 as defined in IPCC (2000)

The climatic variables chosen for analysis are those important for hydrologic impact studies including precipitation rate and precipitable water; often used as inputs to stochastic downscaling studies. Variables used in evaporation estimation – temperature, net radiation, wind speed, pressure and specific humidity were also analysed. One of the benefits of the VCS methodology is that it can be easily expanded to compare any variables of interest for a particular study or region. The analysed variables and their CMIP3 abbreviations are listed below:

- Surface specific humidity **- huss**
- Precipitation rate **- pr**
- Precipitable water **- prw**
- Surface pressure **- ps**
- Surface air temperature **- tas**
- Net longwave radiation **- rls** (combined downwelling (rlds) and upwelling (rlus) longwave radiation)
- Net longwave radiation **- rss** (combined downwelling (rsds) and upwelling (rsus) shortwave radiation)
- Surface wind speed **- was** (combined zonal (uas) and meridional (vas) wind speeds)

Prior to the skill score analyses, GCM outputs were pre-processed using several steps. The GCM outputs were interpolated to a common 2.5° x 2.5° grid, so that the predictions from each model can be compared. Using the 2.5 degree grid, there are 128 land cells covering Australia. To smooth the variable predictions across Australia each grid cell value was replaced by the average of the 9 nearest cells.

The GCM outputs were also bias corrected so that the outputs match the mean and variance of the recorded historical data, which was taken from NCEP reanalysis (Kalnay et al., 1996). The bias correction was then applied to future projections, assuming that the bias will not change over time. The bias correction was carried out for each pixel on a monthly basis. The bias correction was undertaken by standardizing the variables at each grid cell by the modelled monthly means and standard deviations over the period 1961 to 1990. The standardized variables are then rescaled by the observed (in this case NCEP reanalysis) monthly means and standard deviations for the period from 1961 to 1990 for each variable.

Mean annual values for 10 year periods from 2030, 2050 and 2070 were calculated for each variable from each model, for all grid cells within Australia (128 in total) and for both of the emission scenarios considered for the modelling. These values were then used in calculating the CV values and hence VCS as outlined in Section 3.1.

4. RESULTS

Figure 2 shows the VCS values for different climatic variables from 2050 for the two emission scenarios. The results presented in the boxplots show the median, interquartile range and outliers in the VCS values for each variable across Australia. Each boxplot represents 128 VCS values (one for each land cell). Looking at the median values (denoted by the heavy black line in each boxplot), we can compare the relative performance of each climatic variable. The predictions from the 21 model realisations are very consistent for surface pressure (ps), whilst the low VCS for precipitation (pr) shows that there is considerable variation in the predictions from different GCMs. Results, not shown here, are similar for 2030 and 2070.

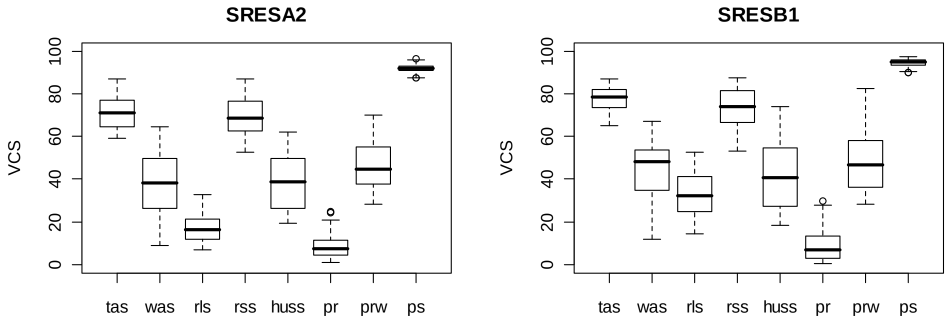

Figure 2. Comparison of VCS values for different climatic variables in 2050 for a) SRESA2 and b) SRESB1

The SRESB1 VCS values are generally slightly higher than the SRESA2, which indicates that there is more agreement in the projections from different models for this lower emission scenario than for the SRESA2 scenario. However, the variation between the different climatic variables is much larger than the differences between the two emission scenarios. Of particular interest in Figure 2 is the range of variation in the VCS values for some climatic variables. For example, the convergence of wind speed predictions across Australia varies markedly. In some areas, the models show as little agreement in their predictions as they do for precipitation, whilst in other areas of Australia we find wind speed predictions show as much agreement as temperature or short wave radiation, both of which generally show good performance.

4.1. Regional variation

Johnson and Sharma (2009) used Koeppen climate zones in an attempt to understand these regional variations in the VCS values. It was found that best convergence occurs in the tropical and temperate regions of Australia. The desert region shows low variable convergence, particularly for the moisture related variables of precipitation rate, specific humidity and precipitable water. In this study, we instead examine the variations with respect to the distance from the coast. Figure 3 shows the VCS values plotted against the minimum distance of the grid cell to any point on the coast. A local linear regression has been fit to the data in each panel to highlight the relationship between VCS and distance in each case.

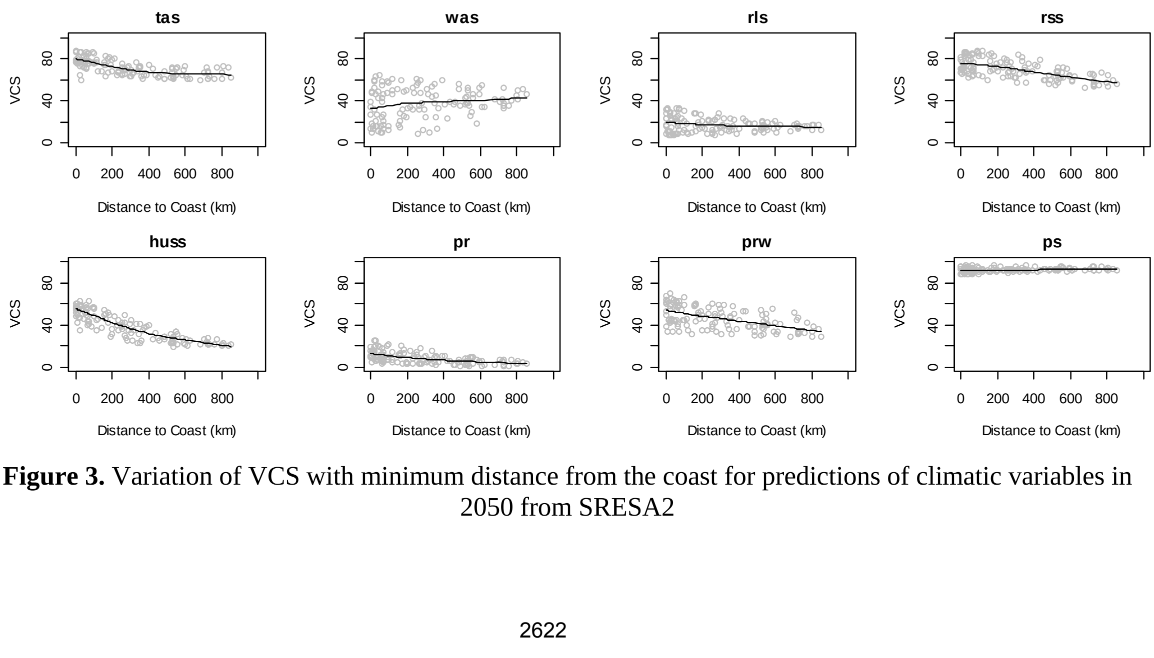

Figure 3. Variation of VCS with minimum distance from the coast for predictions of climatic variables in 2050 from SRESA2

Based on the results of Johnson and Sharma (2009), we expected to find that the VCS decreased with increasing distance from the coast for all variables. This was found to be the case for short wave radiation, precipitable water and specific humidity. Temperature projections show the best convergence closest to the coast, with convergence then fairly consistent once we are approximately 300 km inland. However for the other variables, there is no strong relationship between minimum distance from the coast and model convergence. Precipitation and surface pressure show consistently poor and good convergence respectively. Wind speed shows a large variation of VCS values along coastal areas, with more consistency in VCS values as we move away from the coast.

These relationships have been examined by plotting the VCS values in space to understand the regional variation of model convergence for these variables. Figure 4 shows the spatial variations of wind speed and long wave radiation convergence. The spatial patterns of convergence are almost opposite for these two variables. The GCMs show less agreement for wind speed predictions in the southern parts of Australia, with increasing agreement as we move north. This explains the wide range of VCS values compared to distance from the coast, with high values coming from northern coastal regions and the lowest values from the Victorian coast. Long wave radiation shows better convergence between models in the south and decreasing agreement in the north.

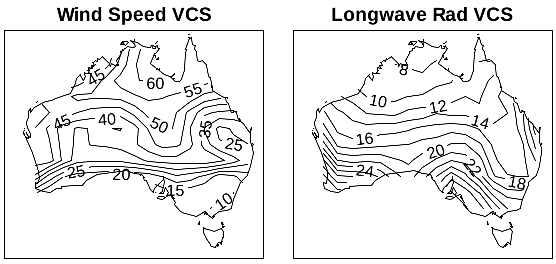

Figure 4. Spatial distribution of VCS for wind speed and long wave radiation for SRESA2 in 2050

4.2. Spatial averaging

We are also interested how model convergence changes with spatial scale. It is generally considered that GCMs are more reliable at larger scales e.g. continental to sub-continental projections. The VCS can be used to quantitatively examine how the projections may change for different levels of averaging. We start by averaging the projections from each model over 3 different scales; firstly, the original 2.5 x 2.5° grid, a 7.5 x 7.5 ° grid (equivalent to 9 grid cells) and finally a 12.5 x 12.5° grid (equivalent to 25 grid cells). The VCS was then calculated for each of these new grid sizes across Australia. Boxplots showing the variation of the VCS values for three key climatic variables are presented in Figure 5.

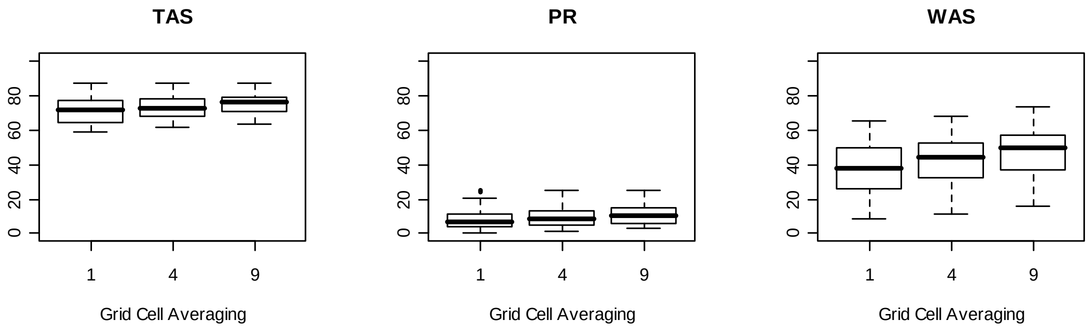

Figure 5. Effect of increasing levels of spatial averaging on model convergence in 2050 for SRESA2

For the temperature and precipitation projections where there was a relatively small range of VCS scores across Australia at the original 2.5° degree grid, the grid cell averaging does not make too much difference to the projections. However for windspeed where regional variations were more pronounced, as previously demonstrated by Figure 4, averaging over larger spatial scales improves the model convergence, as shown by the increasing median values.

5. APPLICATIONS

We have demonstrated in the previous section the use of the VCS in diagnosing how well a set of GCMs converge in their projections of future conditions for a range of hydroclimatic variables. The VCS provides a quantitative measure of the reliability of different variables, and therefore can help to indicate to researchers how much the results of an impact study may vary based on the choice of a single GCM or limited subset of GCMs.

Johnson and Sharma, GCM simulations of a future climate: How does the skill of GCM precipitation simulations compare to temperature simulations?

It is envisaged that a potential application of the VCS could involve variable selection for stochastic downscaling studies. For downscaling studies the variables of interest are often upper atmosphere variables, which are generally assumed to be modelled more reliably than surface variables. The VCS allows these assumptions to be tested and allows, for example, comparisons of the model convergence of mean sea level pressure, the 700 h Pa geopotential height and its gradient, all of which were identified as plausible atmospheric indicators of rainfall in Sydney by Mehrotra and Sharma (2006). We can therefore use the VCS to pre-screen variables for convergence before they are incorporated into the downscaling calculations.

6. CONCLUSIONS

It is important to understand how the estimates of uncertainty of climate change impacts are affected by the selection of a particular subset of GCMs or choice of climatic variables. The VCS has been demonstrated to provide information to researchers and policy makers on how much agreement between GCMs we can expect in time and space, which can provide some bounds on the uncertainty estimates. We found for some variables a strong relationship between model convergence and distance from the coast. In other cases, a north-south gradient of model convergence is present, which can be used to understand model strengths and weaknesses.

ACKNOWLEDGMENTS

Funding for this research came from the Australian Research Council and the Sydney Catchment Authority. Their support for this work is gratefully acknowledged. We acknowledge the modelling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modelling (WGCM) for their roles in making available the WCRP CMIP3 multi-model dataset. Support of this dataset is provided by the Office of Science, U.S. Department of Energy. NCEP Reanalysis data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.cdc.noaa.gov/

REFERENCES

CSIRO and Bureau of Meteorology (2007), Climate change in Australia - Technical Report 2007.

Dessai, S., Lu, X. F., and Hulme, M., (2005), Limited sensitivity analysis of regional climate change probabilities for the 21st century. *Journal of Geophysical Research-Atmospheres*, 110.

Gleckler, P. J., Taylor, K. E., and Doutriaux, C. (2008), Performance metrics for climate models. *Journal of Geophysical Research-Atmospheres*, 113.

Intergovernmental Panel on Climate Change (IPCC) (2000), IPCC Special Report Emissions Scenarios, Summary for Policymakers, *A Special Report of IPCC Working Group III,* Intergovernmental Panel on Climate Change

Johnson, F and Sharma, A. (2009), Measurement of GCM skill in predicting variables relevant for hydroclimatological assessments. *Journal of Climate,* DOI: 10.1175/2009JCLI2681.1.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D. (1996), The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77, 437-471.

Lambert, S. J. and Boer, G. J. (2001), CMIP1 evaluation and intercomparison of coupled climate models. *Climate Dynamics*, 17, 83-106.

Mehrotra, R. and Sharma, A. (2006), Conditional resampling of hydrologic time series using multiple predictor variables: A K-nearest neighbour approach. *Advances in Water Resources*, 29, 987-999.

Perkins, S. E., Pitman, A. J., Holbrook, N. J., and McAneney, J. (2007), Evaluation of the AR4 Climate Models' Simulated Daily Maximum Temperature, Minimum Temperature, and Precipitation over Australia Using Probability Density Functions. *Journal of Climate*, 20, 4356-4376.

Raisanen, J. (2007), How reliable are climate models? *Tellus Series a-Dynamic Meteorology and Oceanography*, 59, 2-29.

Tang, Y., H. Lin, and A. M. Moore (2008), Measuring the potential predictability of ensemble climate predictions. *Journal of Geophysical Research*, 113.

Tebaldi, C. and Knutti, R. (2007), 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.

Xu, C. Y. (1999), Climate change and hydrologic models: A review of existing gaps and recent research developments. *Water Resources Management*, 13, 369-382.