



# Design of a regional climate modelling projection ensemble experiment – NARClIM

J. P. Evans<sup>1</sup>, F. Ji<sup>2</sup>, C. Lee<sup>2</sup>, P. Smith<sup>3</sup>, D. Argüeso<sup>1</sup>, and L. Fita<sup>1</sup>

<sup>1</sup>ARC Centre of Excellence for Climate System Science and the Climate Change Research Centre, University of New South Wales, Sydney, Australia

<sup>2</sup>Office of Environment and Heritage, New South Wales Government, Sydney, Australia

<sup>3</sup>Macquarie University, Sydney, Australia

Correspondence to: J. P. Evans (jason.evans@unsw.edu.au)

Received: 15 August 2013 – Published in Geosci. Model Dev. Discuss.: 25 September 2013

Revised: 23 February 2014 – Accepted: 11 March 2014 – Published: 16 April 2014

**Abstract.** Including the impacts of climate change in decision making and planning processes is a challenge facing many regional governments including the New South Wales (NSW) and Australian Capital Territory (ACT) governments in Australia. NARClIM (NSW/ACT Regional Climate Modelling project) is a regional climate modelling project that aims to provide a comprehensive and consistent set of climate projections that can be used by all relevant government departments when considering climate change. To maximise end user engagement and ensure outputs are relevant to the planning process, a series of stakeholder workshops were run to define key aspects of the model experiment including spatial resolution, time slices, and output variables. As with all such experiments, practical considerations limit the number of ensemble members that can be simulated such that choices must be made concerning which global climate models (GCMs) to downscale from, and which regional climate models (RCMs) to downscale with. Here a methodology for making these choices is proposed that aims to sample the uncertainty in both GCM and RCM ensembles, as well as spanning the range of future climate projections present in the GCM ensemble. The RCM selection process uses performance evaluation metrics to eliminate poor performing models from consideration, followed by explicit consideration of model independence in order to retain as much information as possible in a small model subset. In addition to these two steps the GCM selection process also considers the future change in temperature and precipitation projected by each GCM. The final GCM selection is based on a subjective consideration of the GCM independence and future change. The

created ensemble provides a more robust view of future regional climate changes. Future research is required to determine objective criteria that could replace the subjective aspects of the selection process.

## 1 Introduction

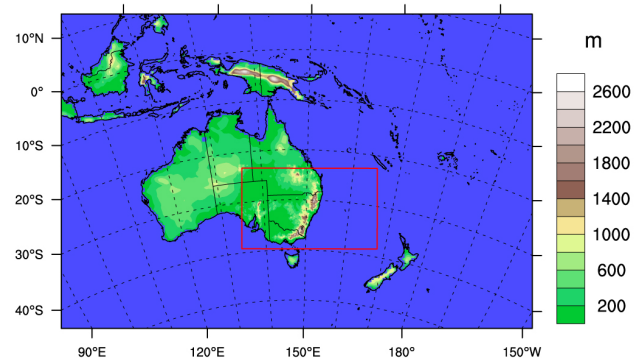
Global warming is a major international concern and requires a global effort to reduce anthropogenic greenhouse gas concentrations. Nevertheless, as global warming continues adaptation to the inevitable changes in climate will have to be done at regional and local scales. This requires climate projection information at a spatial scale relevant to the system of interest, which is frequently significantly smaller than the resolution of global climate models (GCMs). Dynamic downscaling with regional climate models (RCMs) is one method to address this scale gap. A number of previous projects have produced regional climate projections using RCM ensembles including PRUDENCE (Christensen et al., 2007), ENSEMBLES (van der Linden and Mitchell, 2009), RMIP (Fu et al., 2005), NARCCAP (Mearns et al., 2012), CLARIS-LPB (Solman et al., 2013), and now a globally coordinated project in CORDEX (Giorgi et al., 2009). In each case various strategies were used to design the experimental procedure in order to sample the model uncertainties given the practical limitations of computation time and data storage.

While some aspects of the experimental design have developed through successive projects, such as the adoption of

a sparse matrix pairing of GCM and RCM in ENSEMBLES and NARCCAP, other aspects remain to be addressed. The original choice of GCMs and RCMs to include in a project is a primary example, as projects to date have made this choice largely due to convenience. That is, GCMs have generally been chosen based on the ease of access to the data required to create RCM boundary conditions, or due to members of a particular GCM's organisation being involved in the project, and RCMs have been chosen if project members have past experience using them. While such choices were quite pragmatic, advances in computing infrastructure, data sharing and international cooperation through projects such as the 5th Coupled Model Intercomparison Project (CMIP5) and CORDEX, allow more objective choices to be made (McSweeney et al., 2012; Overland et al., 2011). Here we propose a methodology for making these choices, and provide an example of using this methodology within the NSW/ACT Regional Climate Modelling (NARClIM) project. This methodology aims to sample the uncertainty in both GCMs and RCMs, as well as spanning the range of future climate projections present in the full GCM ensemble.

## 2 The NARClIM project design

The express purpose of NARClIM is to deliver robust climate change projections for New South Wales (NSW) and the Australian Capital Territory (ACT) at a scale relevant for use in local-scale decision-making. State governments in Australia have the primary responsibility for natural resource management and the delivery of most community services. This covers many sectors including water resources, biodiversity, infrastructure, health and emergency services. Through a process involving multiple stakeholder workshops, which involved compromise amongst stakeholders from the various sectors, a project design that was achievable within the available computation and data storage resources, was determined. The NARClIM modelling project is unique within Australia as its project design has been a bottom-up approach, heavily involving end users in the conception and design phases, rather than a top-down approach driven mostly by the climate change science community. In the top-down approaches, much of the key questions relating to model epochs and climate variable outputs are decided by the climate modellers and then these are presented to the end user community, including other scientists and modellers working on impact science programs as a *fait accompli*. This leads to a disconnect between the end user or adaptation community and the climate modelling community as the outputs are often not relevant to the needs of the adaptation practitioners or if they are it is by chance rather than design. Involving the adaptation community in the project design maximises the chances of developing model outputs that are readily used by this group. Other benefits of early end user involvement are an improved understanding of the climate modelling process



**Fig. 1.** Topographic map showing the outer and inner (in red) NARClIM model domain and state borders. New South Wales is just to the left of centre of the inner domain.

and its limitations and greater sense of ownership and user uptake of the outputs by the end users. The overall project design includes mechanisms for project governance and data distribution. Information about various aspects of the project can be found at <http://www.crcr.unsw.edu.au/NARClIM/>.

Largely due to the available computing and data storage facilities, the project is limited to a 12-member GCM/RCM ensemble. This will be created by choosing four GCMs and downscaling each of these with three different RCMs. All RCM simulations will be performed at 10 km resolution over NSW/ACT. This high-resolution domain will be embedded within a 50 km resolution domain that covers the CORDEX-AustralAsia region (Fig. 1). Choosing this larger domain ensures that a future stage of the project focused on CMIP5 results can take advantage of simulations performed for the CORDEX initiative. The inner domain and resolution is chosen with a particular focus on simulations of the east-coast climate as this relatively narrow coastal strip, east of the mountains: contains almost half the population of Australia; displays a unique climate response to oceanic modes compared to further inland (Murphy and Timbal, 2008); is generally poorly modelled by GCMs (Suppiah et al., 2007) but is well modelled at 10 km resolution (Evans and McCabe, 2010, 2013); and is strongly influenced by east-coast lows which are often small, rapidly developing storm systems (Speer et al., 2009).

Like previous regional climate projection projects, NARClIM has two main phases.

In phase one, three RCMs are used to downscale the NCEP/NCAR reanalysis (Kalnay et al., 1996) from 1950 to 2010. This reanalysis was chosen to allow a 60-year long historical simulation. Southeast Australia has experienced strong decadal variability in precipitation over the second half of the 20th century with particularly wet decades in the 1950s and 1970s. These reanalysis-driven simulations provide a strong test of the RCMs ability to simulate both these very wet periods and the recent dry period known as the Millennium Drought (Van Dijk et al., 2013). This phase provides

an estimate of the RCM quality including any systematic RCM biases.

In phase two, three RCMs will downscale four GCMs in three 20-year time slices (1990–2010, 2020–2040, 2060–2080). For future projections the SRES A2 emission scenario (IPCC, 2000) will be used. Careful choice of both RCMs and GCMs is required for this small ensemble to adequately sample the model uncertainty – the methodology used to make these decisions is outlined below.

## 2.1 Choosing RCMs

In this experiment we want the small number of RCMs chosen for downscaling to span the range of uncertainty present in the full collection of RCMs that are able to simulate the climate in the area of interest well. Thus a two-step RCM selection process is proposed.

1. The full set of RCMs are evaluated over the domain of interest in order to remove from the set any models that are not able to adequately simulate the climate.
2. From the set of RCMs that perform well a subset is chosen such that each chosen RCM is as independent as possible from the other RCMs.

When evaluating RCMs many subjective choices concerning the variables to be evaluated, the temporal and spatial averaging used, and the statistical measures calculated must be made. Many past studies have evaluated RCM ensembles using many different combinations of the above (e.g. Kjellstrom and Giorgi, 2010; Mearns et al., 2012), generally finding that no model performs best across all variables and metrics (Kjellstrom et al., 2010). Thus, comprehensive evaluation studies are used here to exclude models that perform consistently poorly across a wide range of variables and metrics, rather than trying to identify a set of best models. This approach is consistent with that adopted in McSweeney et al. (2012) and Overland et al. (2011). The large range in possible evaluations that can be performed, along with the many methods to combine evaluation metrics into a final score, makes it difficult to define a priori an acceptable performance level. Here a relative performance level is assessed such that any group of models that are significantly worse than the rest of the models will be excluded.

Now that we have a set of RCMs that perform well over our area of interest, we wish to choose a small subset that spans the uncertainty of this larger set. Given that climate models often share code, there is broad recognition that they do not provide independent samples from the model space (Knutti et al., 2010; Pennell and Reichler, 2011). Hence this choice can be rephrased as one in which the most independent models should be chosen from the larger set. Here, we present a first attempt to consider model independence during the model selection process. Recently Bishop and Abramowitz (2013) proposed a measure that uses the covariance in model errors as the basis for a definition of model

dependence. Here we rank the models based on the magnitude of these independence coefficients and choose the top models from this ranking. It is important to note that these independence coefficients were not designed for this purpose, but rather to provide an optimal linear combination of models from a multi-model ensemble (Potemski and Galmarini, 2009). It is possible to imagine an idealised experiment where they would not lead to selection of the most independent models (see Supplement). One possible situation where the use of the independence weights to select models will be sub-optimal can be identified using the ensemble correlation matrix. If the models separate into groups such that within each group they are extremely highly correlated, while models in different groups have almost no correlation, then this selection method will be sub-optimal. The levels of correlation required within a group are however extremely high (above 0.96), while those between groups are extremely low (below 0.03). However, when tested against actual climate model ensembles the condition described above has not been found and these independence coefficients do perform as desired. They have been shown to select small ensembles with the desired statistical properties (Evans et al., 2013).

## 2.2 Choosing GCMs

Similar to choosing RCMs, the choice of GCMs in this experiment is made in order to sample the range of uncertainty in the ensemble of GCMs that simulate the climate of the target region well. Since a GCM's ability to simulate the current climate has little relationship with the future climate it projects, an additional criterion is introduced. The GCMs chosen should span the range of projected future change, in order to sample this additional source of uncertainty. That is, a three-step GCM selection process is proposed.

1. The full set of GCMs are evaluated over the domain of interest in order to remove from the set any models that are not able to adequately simulate the climate.
2. The set of GCMs that perform well is then ranked based on a measure of independence.
3. The GCMs are then placed within the future change space and the most independent models that span that space are chosen.

While it is possible to perform evaluation of the GCMs in a similar way to that performed for the RCMs, it is also possible to take advantage of the extensive literature in this regard. Given the plethora of evaluation publications based on CMIP3 (and soon CMIP5) data, a metadata analysis of the literature can provide evidence with which to evaluate the models. When this has been done (e.g. Overland et al., 2011; Smith and Chandler, 2010) it is generally found that it is difficult to identify “best” models. Hence, this evaluation is used to identify those models that are consistently poor performers and remove them from consideration.

Several issues must be overcome in order to combine literature studies into one overall score for a GCM: some studies provide a binary pass/fail outcome based on their internal criteria, while others provide continuous measures, and many published studies use only a subset of the full GCM ensemble. Here we address these issues through the introduction of a fractional demerit score, such that the lower the score, the better the performance of the GCM. Demerit points are added to a GCM in two ways. For evaluations which provided a binary pass/fail outcome, any fail equals one demerit point. For evaluations that provide a continuous measure, any GCM that falls in the 25 % worst performing GCMs receives one demerit point. All demerit points across the published studies are totalled for each GCM. Since not every GCM was present in every study this demerit total is then divided by the total number of studies the GCM appeared in to calculate the fractional demerit score. In this way fractional demerit scores of 0.5 or above indicate that the GCM was amongst the 25 % worst GCMs (or failed the test) at least half of the time. These consistently worst performers were then removed from further analysis.

The GCMs that remain are then ranked based on the independence coefficients of Bishop and Abramowitz (2013). Here we rank the models based on the magnitude of these independence coefficients. These rankings are then placed within the GCM's future climate change space, and the highest rankings that span the space are chosen in a subjective manner. The future climate change space can be defined in terms of any climate variables that are deemed appropriate, here temperature and precipitation are used to define this space as they were the variables of most interest to the project stakeholders. It is worth noting that the relatively small sample size of potential GCMs (<20) does not support consideration of more variables and hence a higher-dimensional analysis, though it is possible to do so (e.g. McSweeney et al., 2012). As such, the independence rankings are plotted on an  $x$ - $y$  plot that shows the GCM's projected climate change as given by the change in temperature and precipitation in the area of interest. The most independent models that subjectively best sample the range of future changes are then chosen.

### 3 NARClIM model selection

The model selection criteria above have been applied within the NARClIM project. Given the resources available to the project some further pragmatic choices were made, but within the ongoing international project CORDEX more comprehensive application of the proposed selection criteria could be applied.

#### 3.1 RCM selection

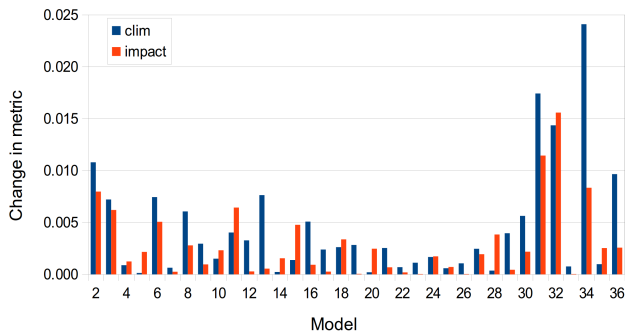
Within a project such as CORDEX, the RCM evaluation could be performed directly on the reanalysis-driven simulations to choose a subset with which to perform the transient GCM-driven simulations. Within NARClIM the available computation resources required the evaluation to be performed using much shorter simulations, and the time constraints limited the number of separate modelling systems that could be implemented. Previous work has shown that the range in the multi-model ensemble can be reproduced within perturbed physics ensembles (Collins et al., 2006). Here the RCM choice is based on a multi-physics ensemble built using the Weather Research and Forecasting modelling system (Skamarock et al., 2008). This system facilitates the use of many RCMs by allowing all model physical parametrisations to be changed and hence many structurally different RCMs can be built. Due to computational limitations, the RCM performance and independence was evaluated based on a series of representative event simulations rather than using multi-year simulations.

By limiting the evaluation period to a series of representative events for the region, a much larger set of RCMs can be tested. In this case an ensemble of 36 RCMs was created by using various parametrisations for the Cumulus convection scheme, the cloud microphysics scheme, the radiation schemes and the Planetary Boundary Layer scheme. Each of these RCMs was used to simulate a set of eight representative storms (Evans et al., 2012; Ji et al., 2014) that cover the various relevant storm types for this region discussed in the literature (Shand et al., 2010; Speer et al., 2009). In each case a 2-week period is simulated centred around the peak of the event. Subsequent analysis then includes pre- and post-event climate as well as the event itself. It should be noted that such an event based evaluation has a number of limitations. During long climate simulations weather periods will arise that were not present in any of the sample events and hence the model performance is untested during these periods, reducing the credibility of the models. Also, by testing a number of relatively short simulations no long-term memory of the system is considered. This may be important if, for example, a model has a strong soil moisture feedback that tends to produce persistent dry states. Ideally, this evaluation would be performed over multiple annual cycles to alleviate these issues, however practical considerations meant that this was not possible.

Evaluation was performed against daily precipitation, minimum and maximum temperature from the Bureau of Meteorology's (BoM) Australian Water Availability Project (BAWAP, Jones et al., 2009). Evaluation was also performed against the mean sea level pressure and the 10 m winds obtained from BoM's MesoLAPS analysis (Puri et al., 1998). The metrics used for the ranking are the bias, root mean square error (RMSE), mean absolute error (MAE) and spatial correlation ( $R$ ) for all variables. The fractional skill score

**Table 1.** The model configuration for the three most independent RCMs.

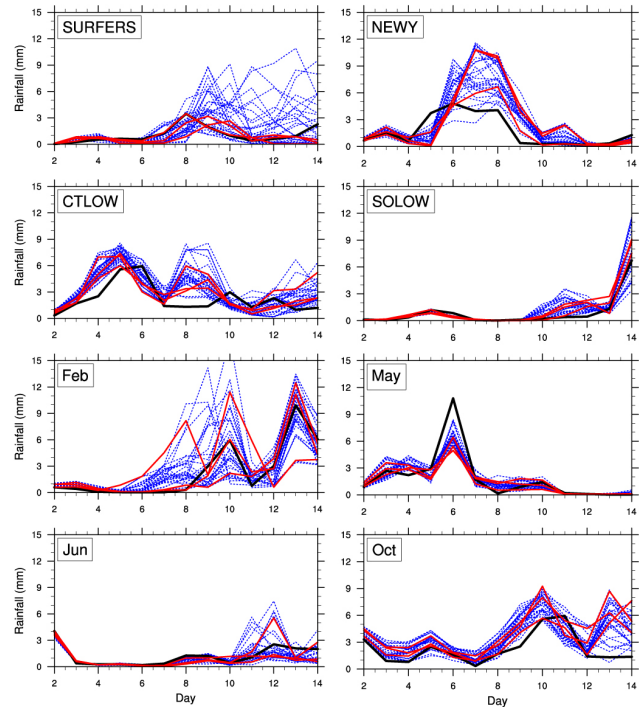
NARcliM ensemble member	Planetary boundary layer physics/ surface layer physics	cumulus physics	Micro-physics	Short-wave/ long-wave radiation physics
R1	MYJ/Eta similarity	KF	WDM 5 class	Dudhia/RRTM
R2	MYJ/Eta similarity	BMJ	WDM 5 class	Dudhia/RRTM
R3	YSU/MM5 similarity	KF	WDM 5 class	CAM/CAM



**Fig. 2.** Change in the overall RCM evaluation metrics between neighbouring models ordered from the best model (left) to the worst model (right).

(FSS) was also used for the rainfall totals. These metrics are calculated for all eight events and combined as described in (Evans et al., 2012). Two overall metrics are calculated such that lower scores indicate better performance (see Tables 1 and 2 of Evans et al., 2012). One metric characterises the climatology (clim) and the other is dominated by the most extreme events (impact). The models are then ordered from the best to the worst model based on the clim metric (the impact metric provides a near-identical ordering), and the differences in the metrics between neighbouring models is shown in Fig. 2. It shows that the overall RCM performance metrics increase gradually from the best to the worst model, with differences between the models of generally less than 0.01. This gradual increase rises sharply at the sixth worst performing model, with differences greater than 0.015 in the clim metric. A similar decrease in performance is seen in the impact metric. Since these six worst performing models show a rapid decrease in performance they are excluded from further analysis.

In the method of Bishop and Abramowitz (2013) the model independence is defined based on the covariance of model errors. For precipitation, minimum and maximum temperature, the daily time series for each event is bias-corrected using the BAWAP observations, to produce an anomaly time series. This anomaly time series for all events is joined together to produce a single long time series for each variable. These time series are then used to create the model error covariance matrix. Bishop and



**Fig. 3.** Daily precipitation time series for each of the eight test periods. Observations are shown in black. All ensemble members retained after the performance evaluation are shown with blue dotted lines. The three members chosen using the independence measure are shown in red.

Abramowitz (2013) are able to show that the coefficients of a linear combination of the models that optimally minimises the mean square error depends on both model performance and model dependence. The solution of this minimisation problem can be written in terms of the covariance matrix already constructed. The size of the coefficients assigned to each model reflects a combination of model performance and independence. That is, the models with the largest coefficients are the best performing/most independent models in the ensemble.

These coefficients are calculated for each variable and then averaged to give the overall performance/independence of each model. The physics parametrisations used in the three most independent/best performing RCMs of the 30-model ensemble are given in Table 1. Figure 3 shows the daily

**Table 2.** Summary of CMIP GCM assessments.

Assessment region		Australia							MDB		SE Australia	
Model	Fractional demerit	A	B	C	D	E	F	G	H	I	J	K
UKMO-HadCM3	0	0	Yes	6	608							179
CSIRO-Mk3.5	0						5	1				207
GFDL-CM2.1	0.111	0	Yes	2	672	Yes			No	Yes	0.72	184
GFDL-CM2.0	0.125	0	Yes	2	671	Yes			No	Yes		252
MIROC3.2 (hires)	0.125	0	Yes	7	608		12	9	Yes			201
CSIRO-Mk3.0	0.182	1	No	7	601	Yes	1	2	Yes	No	0.73	214
UKMO-HadGEM1	0.2	0	No	2	674							163
ECHAM5/MPI	0.222	0	Yes	1	700	Yes			No	No	0.79	173
MIUB-ECHO-G	0.222	0	No	4	632	Yes			Yes	No	0.78	174
INM-CM3.0	0.222	1	No	7	627		9	11		Yes	0.75	192
NCAR CCSM3	0.273	0	No	2	677	No	4	6	No		0.68	245
CNRM-CM3	0.286	0	No	4	542					No	0.73	196
FGOALS-G1.0	0.3	2	No	2	639	No	8	4	Yes		0.66	251
MIROC3.2 (medres)	0.364	2	Yes	7	608	Yes	11	3	Yes	No	0.6	255
CCCM3.1 (T63)	0.375	1		10	478		2	7	No		0.72	241
MRI-CGCM2.3.3	0.455	1	No	3	601	No	10	12	Yes	Yes	0.41	437
CCCM3.1 (T47)	0.455	1	No	8	518	No	3	10	Yes	No	0.77	186
GISS-ER	0.5	0	No	8	515	Yes	6	5	No	No		238
BCCR-BCM2.0	0.5	5		5	590	Yes			No			485
GISS-AOM	0.667	1	No	8	564	No	7	13	Yes		0.6	326
IPSL-CM4	0.8	2	No	14	505	No	13	8	Yes		0.48	394
NCAR PCM	0.833	3	No	11	506						0.64	309
GISS-EH	1	5	No	14	304		14	14				487

A – number of rainfall criteria failed (Smith and Chandler, 2010), B – satisfied ENSO criteria (Min et al., 2005; van Oldenborgh et al., 2005), C – demerit points based on criteria for rainfall, temperature and MSLP (Suppiah et al., 2007), D – M-statistic representing goodness of fit at simulating rainfall, temperature and MSLP over Australia (Watterson, 2008), E – satisfied criteria for daily rainfall over Australia (Perkins et al., 2007), F – order of model based on the total skill scores for each rainfall metric (Kirono et al., 2010), G – order of model based on the total skill scores for each of rainfall and PET metric (Kirono et al., 2010), H – satisfied criteria for daily rainfall over MDB region (Maxino et al., 2008), I – satisfied criteria for MSLP over MDB region (Charles et al., 2013), J – combination of RMSE of mean annual rainfall across south-east Australia and mean NSE (rainfall > 1 mm) comparing GCM-simulated and observed daily rainfall distribution with equal weights (Vaze et al., 2011), K – RMSE of mean annual rainfall over Southeast Australia (Chiew et al., 2009).

precipitation time series for all tested events. The three chosen ensemble members are highlighted in red. Generally the three chosen RCMs display varied simulations of the different events, demonstrating some level of independence between them. The role of performance in the measure can also be seen in the SURFERS case, where none of the models that produced large overestimates of precipitation after the observed peak were chosen. While the models chosen are a compromise across all events, they are still able to sample much of the range of behaviour in the full ensemble for each event.

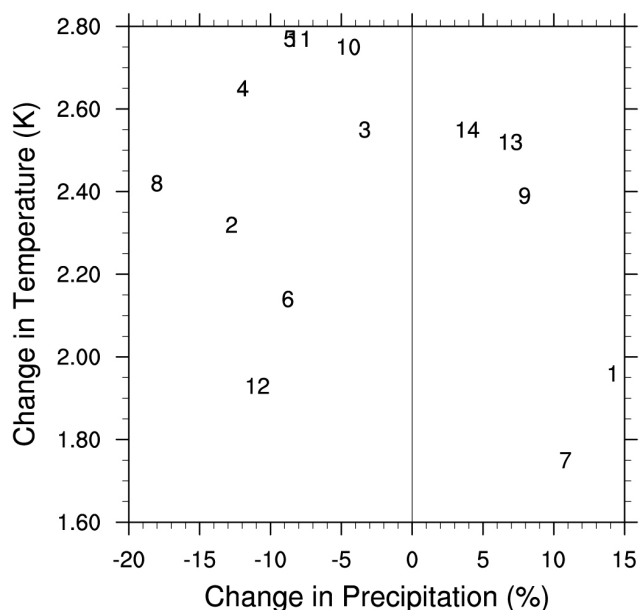
### 3.2 GCM selection

In CORDEX the ensemble from which GCMs are selected is the CMIP5 ensemble. For NARClIM the CMIP3 ensemble is used. Many studies have evaluated the performance of CMIP3 GCMs over south-east Australia using different variables and metrics. Here we build on the meta-analysis of Smith and Chandler (2010). First, more recent evaluations

over Australia, not covered in Smith and Chandler (2010), are added to the analysis for a total of 11 studies (see Table 2). Of these studies four provided a pass/fail assessment of the GCMs, while the rest provided continuous measures. Then a fractional demerit score was calculated to indicate the models overall performance. The lower the fractional demerit the better the performance. Here, six GCMs score 0.5 or higher and are removed from further analysis.

As for the RCMs, the remaining GCMs are then ranked based on their level of model independence using the measure of Bishop and Abramowitz (2013). In this case the independence coefficient is calculated separately for mean temperature and precipitation and then averaged.

The final step requires placing the GCMs within a future climate change space. Such a space could be defined using any combination of climate variables. Here we define the future climate space using the change in mean temperature in Kelvin, and the percent change in mean precipitation. Figure 4 shows the location of the GCMs within this future climate space, numbered by their independence rank order.



**Fig. 4.** Future change space for the CMIP3 GCMs that performed adequately and had the necessary data available, numbered by their independence rank. The change is between the mean of 1990–2009 and the mean of 2060–2079.

Four groupings of GCMs can be seen within this space: top left; top right; centre left; and bottom right. It is desirable then to choose one GCM from each of these groupings that has the highest independence ranking. In this case the models to choose would be the models ranked 3, 9, 2 and 1 respectively. Unfortunately, for various reasons several GCM groups could not supply the required data so alternate GCMs were used. The GCM choice used in practice (and their independence ranking) is MIROC3.2-medres (1), ECHAM5 (5), CCCM3.1 (9), and CSIRO-Mk3.0 (12). Most CMIP5 GCM groups are making available the data required to run RCMs, so within CORDEX the first-choice GCMs should be available.

#### 4 Summary and future work

All regional climate modelling projects require choices to be made concerning the GCMs to downscale from and the RCMs to downscale with. In the past these choices have been largely made based on the convenience of GCM data access and the past modelling experience of project members. Through the greater international cooperation and data access provided by the CMIP5 and CORDEX projects, it is now possible to employ more objective and robust methods for choosing the models to include in regional climate modelling projects.

Here a methodology is proposed to choose models that perform well over the region of interest and that provide as much independent information as possible. This criterion

ensures that the subset of models chosen contains as much of the information available in the full model ensemble as possible. Further, when choosing GCMs, one must also consider their projected future climate change in order to adequately sample all plausible future climates projected by the GCMs that perform adequately over the region.

An application of this methodology within the NARClIM project is presented here. While the method provides a means to objectively select models to use within the project, a number of subjective choices are still required. When evaluating the models a wide range of variables and metrics can be used. How best to combine such measures remains unclear, however the objective here is not to identify the “best” models to use in the ensemble but rather to identify any consistently poor performing models over the area of interest to remove from being considered as possible ensemble members. This identification should be relatively robust to the individual measures used in a comprehensive evaluation as any model whose estimates are far from the observations are likely to perform poorly across a wide range of metrics.

The field of model independence is a relatively new and growing area of research. While the coefficient of Bishop and Abramowitz (2013) is used here as a metric to determine the relative independence of models within an ensemble, it is not an ideal measure and other methods are likely to be developed in the coming years that may also be used within this context.

The future climate change projected by the GCMs is given here by the projected change in temperature and precipitation. This choice was made as these two climate variables were the most sought after by project stakeholders. In practice any climate variables could be used, including the possibility of using a higher-dimensional space (more than two climate variables). Probably the most subjective aspect of the methodology presented here is the choice of models from this future climate change space. Future development of this methodology will include objective methods for making this choice. This may include the application of 2-D clustering techniques to identify clusters from which to choose models, or applying kernel smoothing techniques where the future climate change uncertainty is derived from the inter-annual variability.

Combining the model choice methodology described here with the “sparse matrix” of GCM and RCM combinations used in previous regional climate modelling projects, will result in a climate projection ensemble that more robustly samples the uncertainty space associated with regional climate projections, given limited computational and data storage resources.

**Supplementary material related to this article is available online at <http://www.geosci-model-dev.net/7/621/2014/gmd-7-621-2014-supplement.pdf>.**

*Acknowledgements.* This work was made possible by funding from the NSW Office of Environment and Heritage backed NSW/ACT Regional Climate Modelling (NARCLiM) Project, NSW Environmental Trust for the ESCCI-ECL project, and the Australian Research Council as part of the Future Fellowship FT110100576 and Linkage project LP120200777. This work was supported by an award under the Merit Allocation Scheme on the NCI National Facility at the ANU.

Edited by: J. C. Hargreaves

## References

- Bishop, C. H. and Abramowitz, G.: Climate model dependence and the replicate Earth paradigm, *Clim. Dynam.*, 41, 885–900, doi:10.1007/s00382-012-1610-y, 2013.
- Charles, A., Timbal, B., Fernandez, E., and Hendon, H.: Analog downscaling of seasonal rainfall forecasts in the Murray Darling basin, *Mon. Weather Rev.*, 141, 1099–1117, doi:10.1175/MWR-D-12-00098.1, 2013.
- Chiew, F. H. S., Teng, J., Vaze, J., and Kirono, D. G. C.: Influence of global climate model selection on runoff impact assessment, *J. Hydrol.*, 379, 172–180, 2009.
- Christensen, J., Carter, T., Rummukainen, M., and Amanatidis, G.: Evaluating the performance and utility of regional climate models: the PRUDENCE project, *Climatic Change*, 81, 1–6, doi:10.1007/s10584-006-9211-6, 2007.
- Collins, M., Booth, B., Harris, G., Murphy, J., Sexton, D., and Webb, M.: Towards quantifying uncertainty in transient climate change, *Clim. Dynam.*, 27, 127–147, 2006.
- Evans, J. P. and McCabe, M. F.: Regional climate simulation over Australia's Murray-Darling basin: A multitemporal assessment, *J. Geophys. Res.-Atmos.*, 115, D14114, doi:10.1029/2010JD013816, 2010.
- Evans, J. P. and McCabe, M. F.: Effect of model resolution on a regional climate model simulation over southeast Australia, *Clim. Res.*, 56, 131–145, doi:10.3354/cr01151, 2013.
- Evans, J., Ekström, M., and Ji, F.: Evaluating the performance of a WRF physics ensemble over South-East Australia, *Clim. Dynam.*, 39, 1241–1258, doi:10.1007/s00382-011-1244-5, 2012.
- Evans, J. P., Ji, F., Abramowitz, G., and Ekstrom, M.: Optimally choosing small ensemble members to produce robust climate simulations, *Environ. Res. Lett.*, 8, 044050, doi:10.1088/1748-9326/8/4/044050, 2013.
- Fu, C., Wang, S., Xiong, Z., Gutowski, W. J., Lee, D.-K., McGregor, J. L., Sato, Y., Kato, H., Kim, J.-W., and Suh, M.-S.: Regional Climate Model Intercomparison Project for Asia, *B. Am. Meteorol. Soc.*, 86, 257–266, doi:10.1175/BAMS-86-2-257, 2005.
- Giorgi, F., Jones, C., and Asrar, G. R.: Addressing climate information needs at the regional level: the CORDEX framework, *WMO Bull.*, 58, 175–183, 2009.
- IPCC: IPCC Special Report on Emissions Scenarios, edited by: Nakicenovic, N. and Swart, R., Cambridge University Press, UK, 2000.
- Ji, F., Ekström, M., Evans, J. P., and Teng, J.: Evaluating rainfall patterns using physics scheme ensembles from a regional atmospheric model, *Theor. Appl. Clim.*, 115, 297–304, doi:10.1007/s00704-013-0904-2, 2014.
- Jones, D. A., Wang, W., and Fawcett, R.: High-quality spatial climate data-sets for Australia, *Aust. Meteorol. Mag.*, 58, 233–248, 2009.
- 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.: The NCEP/NCAR 40-year reanalysis project, *B. Am. Meteorol. Soc.*, 77, 437–471, 1996.
- Kirono, D. G. C., Chiew, F. H. S., and Kent, D. M.: Identification of best predictors for forecasting seasonal rainfall and runoff in Australia, *Hydrol. Process.*, 24, 1237–1247, 2010.
- Kjellstrom, E. and Giorgi, F.: Regional Climate Model evaluation and weighting Introduction, *Clim. Res.*, 44, 117–119, doi:10.3354/cr00976, 2010.
- Kjellstrom, E., Boberg, F., Castro, M., Christensen, J., Nikulin, G., and Sanchez, E.: Daily and monthly temperature and precipitation statistics as performance indicators for regional climate models, *Clim. Res.*, 44, 135–150, doi:10.3354/cr00932, 2010.
- Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., and Meehl, G.: Challenges in Combining Projections from Multiple Climate Models, *J. Climate*, 23, 2739–2758, doi:10.1175/2009JCLI3361.1, 2010.
- Maxino, C. C., McAvaney, B. J., Pitman, A. J., and Perkins, S. E.: Ranking the AR4 climate models over the Murray-Darling Basin using simulated maximum temperature, minimum temperature and precipitation, *Int. J. Clim.*, 28, 1097–1112, 2008.
- McSweeney, C. F., Jones, R. G., and Booth, B. B. B.: Selecting Ensemble Members to Provide Regional Climate Change Information, *J. Climate*, 25, 7100–7121, doi:10.1175/JCLI-D-11-00526.1, 2012.
- Mearns, L. O., Arritt, R., Biner, S., Bukovsky, M. S., McGinnis, S., Sain, S., Caya, D., Correia Jr., J., Flory, D., Gutowski, W., Takle, E. S., Jones, R., Leung, R., Moufouma-Okia, W., McDaniel, L., Nunes, A. M. B., Qian, Y., Roads, J., Sloan, L., and Snyder, M.: The north american regional climate change assessment program overview of phase i results, *B. Am. Meteorol. Soc.*, 93, 1337–1362, 2012.
- Min, S.-K., Legutke, S., Hense, A., and Kwon, W.-T.: Internal variability in a 1000-yr control simulation with the coupled climate model ECHO-G – II. El niño Southern Oscillation and North Atlantic Oscillation, *Tellus Ser. Dyn. Meteorol. Ocean.*, 57, 622–640, 2005.
- Murphy, B. and Timbal, B.: A review of recent climate variability and climate change in southeastern Australia, *Int. J. Clim.*, 28, 859–879, doi:10.1002/joc.1627, 2008.
- Overland, J. E., Wang, M., Bond, N. A., Walsh, J. E., Kattsov, V. M., and Chapman, W. L.: Considerations in the Selection of Global Climate Models for Regional Climate Projections: The Arctic as a Case Study\*, *J. Climate*, 24, 1583–1597, doi:10.1175/2010JCLI3462.1, 2011.
- Pennell, C. and Reichler, T.: On the Effective Number of Climate Models, *J. Clim.*, 24, 2358–2367, doi:10.1175/2010JCLI3814.1, 2011.
- Perkins, S. E., Pitman, A. J., Holbrook, N. J., and McAneney, J.: Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions, *J. Climate*, 20, 4356–4376, 2007.



- Potemski, S. and Galmarini, S.: *Est modus in rebus*: analytical properties of multi-model ensembles, *Atmos. Chem. Phys.*, 9, 9471–9489, doi:10.5194/acp-9-9471-2009, 2009.
- Puri, K., Dietachmayer, G., Mills, G., Davidson, N., Bowen, R., and Logan, L.: The new BMRC limited area prediction system, LAPS, *Aust. Meteorol. Mag.*, 47, 203–223, 1998.
- Shand, T. D., Goodwin, I. D., Mole, M. A., Carley, J. T., Coghlan, I. R., Harley, M. D., and Peirson, W. L.: NSW Coastal Inundation Hazard Study: Coastal Storms and Extreme Waves, WRL Technical Report, UNSW Water Research Laboratory, Sydney, Australia, 2010.
- Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-Y., Wang, W., and Powers, J. G.: A Description of the Advanced Research WRF Version 3, NCAR Technical Note, NCAR, Boulder, CO, USA, 2008.
- Smith, I. and Chandler, E.: Refining rainfall projections for the Murray Darling Basin of south-east Australia—the effect of sampling model results based on performance, *Climatic Change*, 102, 377–393, doi:10.1007/s10584-009-9757-1, 2010.
- Solman, S. A., Sanchez, E., Samuelsson, P., da Rocha, R. P., Li, L., Marengo, J., Pessacq, N. L., Remedio, A. R. C., Chou, S. C., Berbery, H., Treut, H. L., de Castro, M., and Jacob, D.: Evaluation of an ensemble of regional climate model simulations over South America driven by the ERA-Interim reanalysis: model performance and uncertainties, *Clim. Dynam.*, 41, 1139–1157, doi:10.1007/s00382-013-1667-2, 2013.
- Speer, M., Wiles, P., and Pepler, A.: Low pressure systems off the New South Wales coast and associated hazardous weather: establishment of a database, *Aust. Meteorol. Ocean. J.*, 58, 29–39, 2009.
- Suppiah, R., Hennessy, K., Whetton, P. H., McInnes, K., Macadam, I., Bathols, J., Ricketts, J., and Page, C. M.: Australian climate change projections derived from simulations performed for the IPCC 4th Assessment Report, *Aust. Meteorol. Mag.*, 56, 131–152, 2007.
- Van der Linden, P. and Mitchell, J. F. B. (Eds.): ENSEMBLES: Climate change and its impacts. Summary of research and results from the ENSEMBLES project, Met Office Hadley Centre, Exeter, UK, 2009.
- Van Dijk, A. I. J. M., Beck, H. E., Crosbie, R. S., De Jeu, R. A. M., Liu, Y. Y., Podger, G. M., Timbal, B., and Viney, N. R.: The Millennium Drought in southeast Australia (2001–2009): Natural and human causes and implications for water resources, ecosystems, economy, and society, *Water Resour. Res.*, 49, 1040–1057, doi:10.1002/wrcr.20123, 2013.
- van Oldenborgh, G. J., Philip, S. Y., and Collins, M.: El Niño in a changing climate: a multi-model study, *Ocean Sci.*, 1, 81–95, doi:10.5194/os-1-81-2005, 2005.
- Vaze, J., Teng, J., and Chiew, F. H. S.: Assessment of GCM simulations of annual and seasonal rainfall and daily rainfall distribution across south-east Australia, *Hydrol. Process.*, 25, 1486–1497, 2011.
- Watterson, I. G.: Calculation of probability density functions for temperature and precipitation change under global warming, *J. Geophys. Res.-Atmos.*, 113, D12106, doi:10.1029/2007JD009254, 2008.