Assessing the credibility of downscaled rainfall extremes for the Greater Sydney region – a novel approach using blended radar/gauge data

Dörte Jakob
Manager Hydrometeorological Advisory Service, Bureau of Meteorology, Melbourne, Australia, E-mail: d.jakob@bom.gov.au

Aurel Griesser
Hydrometeorological Analyst, Bureau of Meteorology, Melbourne, Australia
E-mail: a.griesser@bom.gov.au

Alan Seed
Principal Research Scientist, Bureau of Meteorology, Melbourne, Australia
E-mail: a.seed@bom.gov.au

Abstract

In conjunction with the revision of the ''Australian Rainfall and Runoff', a project was undertaken to a) quantify possible changes and uncertainties in rainfall Intensity-Frequency-Duration (IFD) curves due to anthropogenic climate change and b) provide advice to practitioners on how to include these changes in design and planning. We will present and discuss the key findings from one of the five components of this project.

We evaluated the realism of downscaled rainfall extremes from two Regional Climate Models (Cubic Conformal Atmospheric Model, CCAM; Weather Research and Forecasting Model, WRF) driven by two Global Climate Models (ACCESS 1.0; CSIRO Mk 3.5) and reanalysis-driven (ERA Interim; NCEP/NCAR) for the Greater Sydney region. The evaluation used blended radar/gauge data (Rainfields) and gridded station data (Australian Water Availability Project, AWAP). The use of radar data in this context is a novel approach, motivated by their superior spatial and temporal resolution. Our work identified limitations in both observational datasets. We therefore recommend using more than one observational dataset in similar evaluation studies, analogous to the use of ensembles in climate model simulation.

To produce credible projections of changes in rainfall extremes, models need to be able to realistically simulate key mechanisms and characteristics of rainfall relevant to these extremes. We found that models produce unrealistically high rainfall depths and while bias correction reduced errors for modelled annual accumulations, it did not remove the tendency for overestimation of rainfall extremes and introduced artefacts. Simulated precipitation is less random in space and time than observed precipitation, which implies that not all rainfall-producing mechanisms have been correctly captured. This is important because changes in rainfall extremes are likely to be driven by changes in convective precipitation.

1. INTRODUCTION

The frequency and intensity of rainfall extremes may change as a consequence of climate change (IPCC, 2012). Likely changes in design rainfalls can be assessed on the basis of projections from Global Climate Models (GCMs). These models provide information on fairly coarse spatial scales (~100 km) but for decision-making purposes, information on regional and local scales is required. GCM results are therefore interpreted to provide information on regional and local scales. The relevant technique is referred to as "downscaling" and involves statistical and/or physics-based approaches. Projected changes in downscaled rainfall extremes exhibit a large degree of variability spatially (i.e. projected increases and decreases manifest in close proximity) and between studies (Alexander & Arblaster, 2009).

A particular downscaling technique can be assessed by comparing downscaled model output to observations for the current climate. Observational data sets are typically based on either station data or gridded information derived from station data. The drawback of using station-based data sets is mainly the coarse temporal and/or spatial distribution of the data used. An alternative data set is radar data, which provides consistent spatial and temporal information that cannot be readily gleaned from station information alone.

2. DATA

2.1. Model data

To understand the value of downscaled information for planning purposes, the realism of these projections needs to be assessed and the associated uncertainty needs to be quantified. This paper assesses the realism of downscaled rainfall extremes on the basis of two different dynamical models, also known as Regional Climate Models (RCM): the Weather Research and Forecasting Model (WRF) (Evans & McCabe, 2013) run at a 2 km spatial resolution, and the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Cubic Conformal Atmospheric Model (CCAM) (Thatcher & McGregor, 2009) run at spatial resolutions of both 10 km and 2 km over the region of interest. Downscaled information for the current climate was made available for the periods 1990 – 2009 (WRF) and 1980 – 2010 (CCAM).

2.2. Observational data sets

2.2.1. Radar data

The Australian Bureau of Meteorology's radar network consists of a mixed collection of about 66 radars with an average age of 9 years. Weather radars are operated for real-time applications rather than for climatological studies. In an operational setting, it is vital to process these radar measurements in a very short time, effectively limiting the complexity of processing algorithms that can be applied in real-time. The evaluations presented in this paper are based on spatio-temporal characteristics derived from blended radar and gauge rainfall accumulations and are in the following referred to as "Rainfields" (Seed et al 2007).

Algorithms have been developed to detect and remove artefacts in radar data (Peter et al 2014). Such artefacts include partial beam blocking by local topographic effects, clutter (e.g. dust, insect swarms, reflection on the sea surface) and – in an operational setting – bias correction that requires real-time adjustment. Rain gauge data also need quality control, and information which is obviously wrong - in particular false zeros, where the continuous rain gauge incorrectly records zero during rainfall, – should be filtered out. The conversion of observed radar reflectivity (Z) to estimated rain rate (R) is one of the most critical steps in the estimation of realistic radar rainfall accumulations. There are significant differences in *Z-R* relationships for convective and stratiform rain, which are accounted for using complex algorithms. Instantaneous measurements of rain rate have to be integrated over time to derive rainfall accumulations.

For this project a new version of Rainfields (version 3) was produced using a more sophisticated processing algorithm with the intention of providing improved quantitative precipitation estimates through improved quality control. As part of the new processing algorithm, a spatially varying calibration of radar against gauge data was implemented as well as measures to assess the quality of each 30-minute rainfield. These techniques have been applied to radar data for the Greater Sydney region. The resulting product is a mosaic of radar data from five radar sites (Wollongong, Newcastle, Sydney, Canberra, Kurnell), covering a 500 km by 500 km domain, centred on the Wollongong radar (150.87˚E and 34.26˚S) at a resolution of 0.01˚ (approximately 1 km) and 30 minute rainfall accumulations.

There is a fundamental difference between radar and ground based measurements: radars observe precipitation at height while rain gauges measure at the surface. A number of meteorological processes (e.g. evaporation, wind drift, fall speed) affect raindrops on their way to the surface. Due to these factors, even if the conversion from radar reflectivity to rain rate were reasonably accurate, there are difficulties in comparing rain rate at height with rain rate at the ground because the rainfall is being modified during its descent. While gauge data provide accurate information at the gauge location, the radar estimate is generally more accurate than a gauge-based estimate at distances from a gauge of more than 10 km or so.

2.2.2. AWAP data

In an attempt to define a yard stick for the evaluation of gridded data, gridded analyses of daily rainfall data (24 hour accumulations from 9 am Local Standard Time (LST)) were used. The following section discusses the background to the development of AWAP based on Jones et al (2009) and limitations in AWAP data relevant to rainfall extremes.

The analysis techniques were developed as part of the Australian Water Availability Project (AWAP) and the grids are commonly referred to as AWAP grids. The resolution of these grids is 0.05° by 0.05°, which is approximately 5 km by 5 km. The AWAP grids were developed for the purpose of monitoring of climate change and climate variability for the Australian region and are produced operationally at the Bureau of Meteorology. The spatial analyses are based on in-situ data only. An anomaly-based approach was used, i.e. data were decomposed into the long-term average and anomaly components. A Barnes technique was used to grid the anomalies (Koch et al 1983) while smoothing splines are used for the average grids (Wahba & Wendelberger, 1980). Finally, daily rainfall grids were derived by multiplying climatology and anomaly analyses.

The AWAP grids are an improvement over existing analyses. However, the accuracy of these grids is limited by the density of the station network. Errors tend to increase where rainfall gradients are strong, such as for regions with significant orography. The mean average error is still about 50% of the average daily rainfall. A recent study (King et al 2013) assessed the performance of AWAP grids for daily rainfall extremes. Comparisons of frequency and intensity of rainfall events above the 95th percentile were undertaken against 119 high-quality stations. While the AWAP grid typically underestimates intensity and frequency of extreme rainfall events, the intensity and frequency of low rainfall events is typically overestimated. The underestimate in the magnitude of the average 95th percentile for the study region is of the order of 10%. Problems were identified for remote areas but considering the location of the study region this is not a concern for the investigations presented in this report. The rank correlation between AWAP and station rainfall for the study region is high (0.8 to 0.9). AWAP estimates can therefore be expected to be a good indicator for where and when extremes occurred but they will be less reliable in terms of the magnitude of rainfall accumulations.

3. METHODS AND RESULTS

3.1. Identifying limitations in observational and model data

Preliminary investigations were undertaken to assess and understand the suitability of available observational and climate model data for the planned evaluation. These investigations indicate that biases exist in both observational gridded data sets (AWAP and Rainfields). It is therefore sensible - where possible – to make use of more than one observational data set for the evaluation of downscaled rainfall extremes to gain a better understanding of the uncertainty in observed rainfall extremes. AWAP estimates are less reliable in regions with steep gradients and for sparse networks, and the intensity and frequency of extreme events is generally underestimated. Based on analyses of annual and seasonal totals, Rainfields underestimates rainfall depth. These biases are large for low and light precipitation but smaller for extreme precipitation. Statistical measures for AWAP data for the period 1990-2009 are similar to those for the period 2009-2012 for which Rainfields data are available, indicating that Rainfields data can be used to evaluate the model simulations.

The bias-corrected WRF-GCM simulations for both the current climate and the climate projections show a strong, increasing trend in annual rainfall accumulations over each of the two 20-year simulation periods, whereas no such trend is apparent in the observations for the period of investigation. For details on the bias correction applied refer to Argüeso et al., 2013.
This unrealistic trend makes it difficult to calculate meaningful statistics over the time slice as a whole.

While no such trend was found in CCAM simulations, CCAM model output is available at 10 km and 2 km resolutions, which enabled the effect of model resolution to be investigated. Based on the analysis of daily rainfall it was found that the choice of resolution does not markedly affect the results of the evaluation.

3.2. Analysing spatio-temporal characteristics of observations and simulations

Maps of annual and seasonal rainfall accumulations were used to provide an initial impression of how well observations match the simulations. They are useful in identifying obvious biases in gridded data and to explore how these biases vary across the domain. If there is a consistent bias (over or underestimation) this will become very clear in the annual and seasonal accumulations. While it is informative to assess the simulations in terms of averages, it is also important in the context of this study to evaluate simulations of rainfall extremes; maps of 95th and 99th percentile of daily rainfall accumulations were used for that purpose.

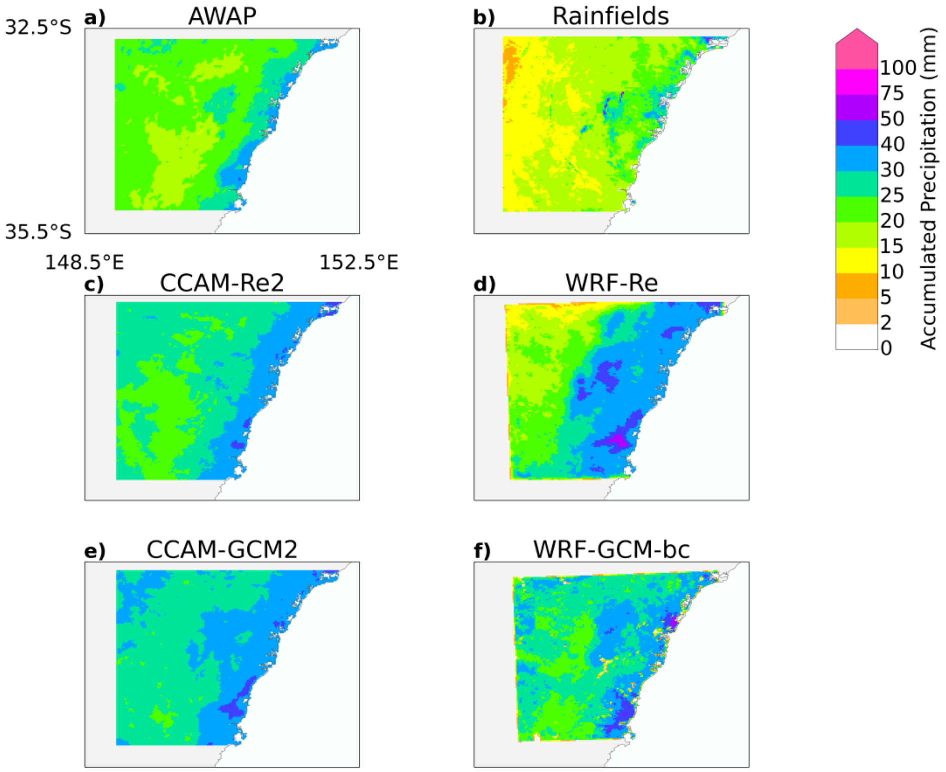

Figure 1 95th percentile of daily rainfall in the Greater Sydney region over 1990-2009 period for (a) AWAP, (b) Rainfields (2009-2012), (c) CCAM-Re2, (d) WRF-Re, (e) CCAM-GCM2 and (f) WRF-GCM-bc.

Figure 1 maps the 95th percentile for daily rainfall. The highest values for the 95th percentile of daily accumulations in AWAP occur in the coastal region. This pattern is matched by Rainfields although the magnitude for Rainfields accumulations is notably lower. Both of the CCAM simulations exhibit consistently higher estimates than the observations and have the same broad spatial pattern although the reanalysis-driven CCAM run at the 2 km resolution (CCAM-Re2) looks more realistic. The reanalysis-driven WRF simulation (WRF-Re) overestimates the magnitudes somewhat relative to Rainfields and does not replicate the location of the highest accumulations well. The magnitudes of the bias corrected GCM-driven WRF run (WRF-GCM-bc) are more realistic although still too high and it appears the bias correction is destroying some of the spatial patterns seen in the raw WRF-GCM output (not shown).
Any biases identified in the average annual rainfall accumulations are not necessarily characteristic of

extreme events. Furthermore, biases for low and high intensity events may partially compensate each other. Histograms of daily and hourly accumulations, boxplots and quantile-quantile plots were used to compare the frequency distributions of observed and simulated data. Where the distribution of simulated rainfall accumulations is distinctly different from the distribution of observed rainfall accumulations, the simulated changes are less reliable and their interpretation becomes more complicated.

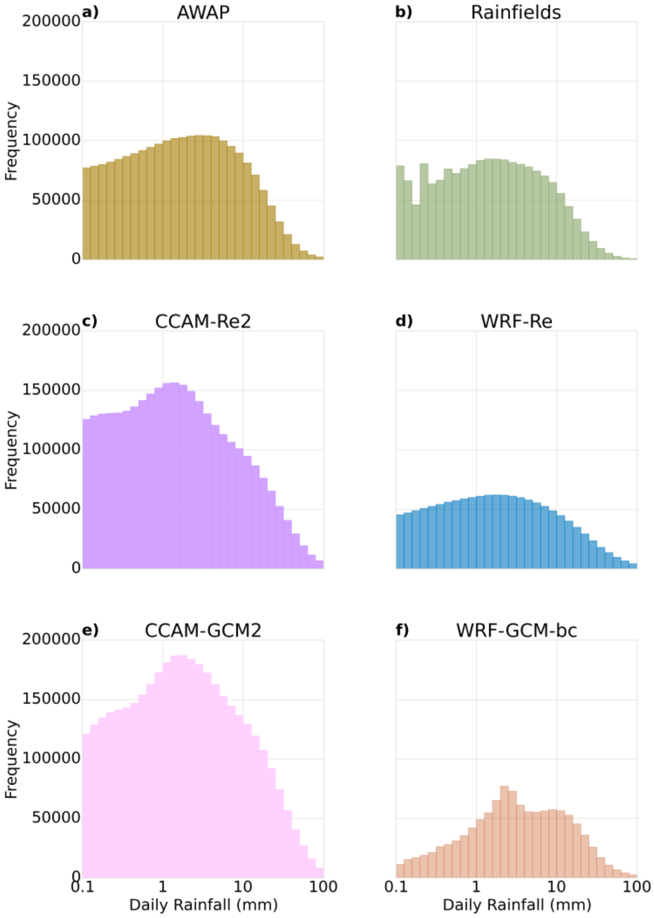

Figure 2 Histograms of daily rainfall accumulations (above 0.1 mm, regridded to 0.02 degree) for (a) AWAP, (b) Rainfields (2009-2012), (c) CCAM-Re2, (d) WRF-Re, (e) CCAM-GCM2 and (f) WRF-GCM-bc.

Histograms of daily rainfall (Figure 2) indicate that the simulations produce a higher number of moderate to heavy rainfall events than is observed in Rainfields. It is likely that the lower tail from AWAP data is too fat because light precipitation is smeared out by the gridding algorithms. Apart from the lower tail there is good agreement between the frequency of events for AWAP and Rainfields. Rainfields data is stored with a resolution of 0.05 mm and this granularity leads to the spikes for the lowest intervals in the histogram plot.

CCAM-GCM2 has an unrealistically heavy tail at the low and at the high end, so that overall CCAM-GCM2 accumulations are more uniformly distributed than suggested by observations. CCAM, like many other models, has a known drizzle or light rainfall problem, i.e. the grid box can “smear out” light precipitation (M. Thatcher, pers. comm.) For WRF-GCM-bc the bias correction removes the heavy lower tail (perhaps too strongly) and the heavy upper tail (perhaps insufficiently), it also introduces bimodality that is not found in the observations.

The set of measures used to characterise the degree of spatio-temporal organisation and lifetime of events in observations and simulations include temporal autocorrelation, wet spell duration (consecutive hours of precipitation) and feature analysis (contiguous areas with rainfall above a given threshold).

Rainfall extremes in the Greater Sydney area occur more frequently during the warmer months and the highest rainfall extremes tend to occur during summer, and changes in rainfall extremes are likely driven by changes in summer precipitation. It was therefore also assessed how well downscaled rainfall extremes replicate the seasonality of observed precipitation.

Daily rainfall totals for five individual rainfall events were analysed to explore how well reanalysis driven model simulations replicate the characteristics of observed events. The selection of these five multi-day events was based on the following criteria: spatially averaged precipitation, rain cell fraction (percentage of the domain that has received precipitation), maximum rainfall intensity and information about the associated synoptic conditions and impact of an event. An important caveat in the interpretation of these comparisons is that ‘the ability to model rain that resembles an observed event depends on whether the large-scale synoptic situation is sufficient to constrain the model outcome without the benefit of high-resolution initial conditions’. (M Thatcher, pers. comm.)

4. KEY FINDINGS AND CONCLUSIONS

The findings for CCAM and WRF show a number of similarities so that it may be reasonable to assume they are more widely applicable. On the other hand, both models show similar deficiencies and may therefore not provide a good indication of the range of projected changes in rainfall extremes. To be able to improve the confidence in likely range of potential changes in rainfall extremes, it will be necessary to assess a larger number of GCMs. The following statements summarise key findings and their implications.

1. There are deficiencies in the observational datasets that are relevant to the evaluation of downscaled rainfall extremes.
    - AWAP grids were specifically developed for the purpose of monitoring climate change and climate variability. These grids are based on sophisticated gridding techniques and a definite improvement over previously available gridded datasets. Errors in average and extreme AWAP precipitation are well documented (Jones et al 2009; King et al 2013). AWAP estimates are known to be less reliable in regions with steep precipitation gradients and sparse networks. The latter is not an issue for the study region but would need to be considered if similar analyses were to be undertaken for other parts of the country. Due to the gridding techniques used (splines and Barnes) the resulting grids are overly smooth.
    - Radar data provide a great opportunity to study the spatio-temporal characteristics of precipitation and it is difficult to ignore their usefulness in the evaluation of downscaled rainfall extremes. Limitations to the accuracy of Rainfields data are associated with the fact that radar measures precipitation at height rather than at the ground. Complex algorithms are required to convert radar reflectivity to rainfall accumulations.
    - Analogous to the use of ensembles in climate model simulations, using more than one observational data set helps give an indication of inherent uncertainty in the observational data set because of different data sources and analysis techniques.

2. The frequency distributions of observed and simulated precipitation exhibit marked differences.
    - The model runs evaluated here (CCAM and WRF) typically simulate too many light rainfall events as well as unrealistically high rainfall extremes. This leads to distributions of daily rainfall accumulations that are flatter than the observed with a median value that is typically higher than for the observations. This affects the assessment of projected changes in design rainfall estimates. The difference in distribution of observed and simulated precipitation also raises questions about whether models correctly replicate rainfall-producing mechanisms.

3. Compared to GCM-driven simulations, reanalysis-driven CCAM simulations are typically a better match with observations.
    - Simulations of future climate cannot be undertaken using reanalysis data. Confidence in the (GCM driven) CCAM runs for the climate projections is therefore lower than in (reanalysis driven) runs for the current climate.
4. Bias-correction of GCM driven WRF output (WRF-GCM-bc) leads to a marked improvement in performance but does not fully address the bias in extremes (e.g. the modelled magnitude of the

95th percentile) and introduces artefacts.

- A histogram-matching approach is used to post-process WRF outputs (Argüeso et al., 2013). While this leads to improvements for average rainfall accumulations, it does not remove the tendency to overestimate extreme precipitation in WRF simulations. The bias-correction also appears to introduce a bimodality to the rainfall distribution which is not representative of observations and results in an unrealistic spatial distribution for the extremes (95th and 99th percentile).
- The post-processing of model output implicitly assumes that the bias correction which is valid under current climate conditions is also applicable for climate projections. These findings limit the degree of certainty in the projected changes in magnitude of rainfall extremes.

5. Model simulations may not adequately replicate extent, location and magnitude of events.
   - Based on the limited number of events assessed, it appears that simulations perform better for large-scale events than for small-scale events. This limits confidence that all rainfall-producing mechanisms have been correctly captured and relevant changes can be correctly projected.

6. Simulated precipitation is typically more organised in space and time than observed precipitation.
   - This finding is based on comparisons of temporal autocorrelation (dropping off more quickly in simulations than observations), wet spell duration (consistently longer for simulations than observations) and visual assessment of spatial patterns for five events. This is important because there are indications that changes in rainfall extremes may be driven to a considerable degree by changes in convective precipitation (Berg, Moseley & Haerter, 2013).

7. The seasonality of downscaled rainfall extremes from WRF simulations agrees well with the seasonality of observed extremes.
   - Analyses of historical observations show that any long-term change in rainfall extremes is likely to be driven by changes in rainfall extremes during summer. Rainfall extremes for the study region occur most frequently during summer and this is replicated in the simulations. The seasonal distributions from WRF simulations are a far better match with the observed than those from CCAM, especially since CCAM strongly overestimates the frequency of moderate events in summer.
8. Based on the comparisons of CCAM runs at two resolutions (10 km and 2 km) there is little indication that increasing the resolution leads to an improvement in the simulation of rainfall extremes.
   - Given that model runs at higher resolution are computationally much more expensive it should be assessed whether the additional effort is justified.

9. While model simulations show some skill in matching the peak intensities of rainfall extremes on an annual scale, this skill is lower when simulated events are compared to observations.
   - It may therefore be preferable to consider providing indicative estimates of projected changes for the study region as a whole without distinguishing by geographical location within the study region.

10. GCM-driven WRF simulations for the current and the future climate (1990 – 2009 and 2040 to 2059) exhibit a marked drift with annual accumulations that increase over time.
    - A comparable increase in annual accumulations is not apparent from the observations or in the CCAM simulations and is probably unrealistic. This model artefact makes it more difficult to interpret the projections derived using WRF.

The very limited number of simulations analysed makes it difficult to assess whether these findings can be generalised. A more robust assessment would require the analysis of a larger number of global climate models and regional climate models.

5. ACKNOWLEDGMENTS

This work was undertaken as part of the revision of Australian Rainfall and Runoff (ARR) with funding from the Australian Government and significant in-kind contributions by Bureau of Meteorology staff.

6. REFERENCES

Alexander L V. and J.M. Arblaster (2009), Assessing trends in observed and modelled climate extremes over Australia in relation to future projections International Journal of Climatology, 29(3), 417–435. doi: 10.1002/joc.1730

Argüeso, D., J. P. Evans, and L. Fita (2013), Precipitation bias correction of very high resolution regional climate models. Hydrology and Earth System Sciences, 17(11), 4379–4388. doi:10.5194/hess-17-4379-2013

Berg, P., C. Moseley, and J. O. Haerter (2013), Strong increase in convective precipitation in response to higher temperatures. Nature Geoscience, 6(3), 181–185. doi:10.1038/ngeo1731

Evans, J., and M. McCabe, M. (2013), Effect of model resolution on a regional climate model simulation over southeast Australia. Climate Research, 56(2), 131–145. doi:10.3354/cr01151

IPCC (2012), Summary for Policymakers. In: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK, and New York, NY, USA, pp. 1-19.

Jones, D. A., W. Wang, and R. Fawcett, R. (2009), High-quality spatial climate data-sets for Australia. Australian Meteorological and Oceanographic Journal, 58, 233–248.

King, A. D., L.V. Alexander, and M.G. Donat (2013), The efficacy of using gridded data to examine extreme rainfall characteristics: a case study for Australia. International Journal of Climatology, 33(10), 2376–2387. doi:10.1002/joc.3588

Koch S. E., M. desJardins, and P.J. Kocin (1983), An interactive Barnes objective map analysis scheme for use with satellite and conventional data J. Clim. Appl. Meteor. 22 1487-1503

Peter, J., A. Seed, P. Steinle, S. Rennie, and M. Curtis, (2014), A Bayesian methodology for detecting anomalous propagation in radar reflectivity observations. CAWCR Technical Report No. 077, available here http://www.cawcr.gov.au/publications/technicalreports/CTR_077.pdf

Seed, A., E. Duthie, and S. Chumchean, S. (2007), Rainfields: The Australian Bureau of Meteorology system for quantitative precipitation estimation. Proceedings of the 33rd AMS Conference on Radar Meteorology, Cairns, Australia.

Thatcher, M., and J.L. McGregor, (2009), Using a Scale-Selective Filter for Dynamical Downscaling with the Conformal Cubic Atmospheric Model. Monthly Weather Review, 137(6), 1742–1752. doi:10.1175/2008MWR2599.1

Wahba, G. and J. Wendelberger (1980), Some new mathematical methods for variational objective analysis using splines and cross validation. Mon. Weath. Rev., 108, 1122-1143