Calibration and bias correction of seasonal climate forecasts for use in agricultural models

Peter C. McIntosh and Jaclyn N. Brown

December 2016
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contents</td>
<td>2</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>4</td>
</tr>
<tr>
<td>2 Data and Methods</td>
<td>5</td>
</tr>
<tr>
<td>2.1 Background</td>
<td>5</td>
</tr>
<tr>
<td>2.2 Model Output</td>
<td>8</td>
</tr>
<tr>
<td>2.3 Observations</td>
<td>9</td>
</tr>
<tr>
<td>2.4 Method</td>
<td>10</td>
</tr>
<tr>
<td>3 Results</td>
<td>12</td>
</tr>
<tr>
<td>4 Conclusions</td>
<td>19</td>
</tr>
<tr>
<td>5 Acknowledgements</td>
<td>19</td>
</tr>
<tr>
<td>6 Appendix A</td>
<td>20</td>
</tr>
<tr>
<td>Mean square error and variance plots for all sites</td>
<td>20</td>
</tr>
<tr>
<td>7 References</td>
<td>25</td>
</tr>
</tbody>
</table>
Seasonal climate forecasts from global circulation models have now reached a level of skill where it is becoming appropriate to incorporate them directly into agricultural models and tools. The output from such models arrives on a large spatial grid and can contain biases. Examples of these biases include temperatures being too warm, or rain being the correct amount but falling as drizzle over many days instead of distinct large rain events.

Here we present and evaluate a method to downscale and calibrate the output from a seasonal climate model to locations where weather stations exist. The technique is designed to be effective but with low-computational cost so it can be delivered quickly to websites and Apps.

A frequency correction is first applied to the rainfall in the model output, particularly to adjust for the number of dry days. Weather station records are then used to adjust the statistical properties of the model output through a quantile mapping approach.

In eight case study sites the adjustment is effective with the MSE being close to the ensemble variance.

1 Introduction

Forecasts of seasonal conditions can be of considerable value in the management of agricultural enterprises in a variable climate (e.g. Hammer et al. 2000; McIntosh et al. 2005; McIntosh et al. 2007; Asseng et al. 2012a; Asseng et al. 2012b). Increasingly these forecasts are being produced by computer models of the global coupled ocean-atmosphere climate system. Model forecasts on seasonal timescales generally have moderate skill due to limitations in the grid resolution, the accuracy of the starting conditions, and the chaotic nature of the climate system (Goddard et al. 2001). Despite this, forecasts can provide useful guidance for farmers, particularly when retrospective forecasts (hindcasts) are compared to observations over previous decades to calibrate model behaviour and bias. This article examines techniques for model calibration and bias correction and develops a method for using seasonal model forecasts to drive crop models.

POAMA is the seasonal forecast system run operationally by the Australian Bureau of Meteorology (Hudson et al. 2013). This model contributes to the official seasonal climate outlook provided every month. However, considerably more detailed model output is available for experimental purposes. In particular, forecasts of quantities of relevance to agriculture, such as rainfall, temperature and solar radiation, are made extending out to nine months in the future with a daily time step. This model output could potentially be used to drive crop models such as APSIM (Keating et al. 2003). Furthermore, multiple forecasts are made at any one time in order to simulate the range of possible outcomes. This forecast ensemble could allow the calculation of a range of possible crop outcomes for the current season.

The skill of POAMA seasonal forecasts can be estimated by using a set of hindcasts that extend back to 1981. The model has been shown to have significant skill at forecasting various quantities over a range of different timescales (http://poama.bom.gov.au/publications.shtml). The skill generally varies according to the region and time of year. Some measures of skill, such as whether rainfall will be above or below the median, depend only on the model variation from year to year. This measure of skill is not affected by model bias, such as the model being consistently too wet, since the comparison is made relative to the model’s own climatology. However, when it comes to driving a crop model such as APSIM, it is important that rainfall and other...
variables have realistic values in order to get sensible crop growth. To achieve this, a bias correction can be calculated based on comparing historical model behaviour to observations.

Seasonal forecasts are valuable because they represent the year-to-year variability of the climate, and give an estimate of how the current year differs from the long-term average. It is important, therefore, that the model simulate accurately the amplitude of this interannual variability. Similar to bias correction, calibration can be applied to ensure model interannual variability is scaled to match that observed historically.

The multiple forecasts provided by the model ensemble give a range of possible outcomes, and this range must be realistic in order to be useful. Calibration techniques can also be used to ensure that the forecast spread is representative of the uncertainty in the forecast mean.

The last consideration in the application of POAMA seasonal forecasts is one of downscaling. The POAMA atmospheric model grid for version 2.4 is approximately 250 km on a side, so that a specific location where model output is required could be up to 125 km distant from the nearest model grid point in both the east-west and north-south directions. There are various different methods for downscaling relatively coarse model output to individual locations, ranging from simple interpolation to embedded fine-scale dynamical models. There appears to be no general agreement on the best method for downscaling, and this is an area that would benefit from further research to establish a unified approach. This project takes the view that most of the grain-growing regions are on relatively flat country away from coasts, so that there is less reason to expect large variations in climate over spatial scales shorter than the model scale than there might be elsewhere. In addition, some downscaling methods can be computationally intensive (e.g. Wong et al. 2014), and these would not be suitable for the present application. Therefore we choose to keep the downscaling method as simple as possible consistent with satisfactory results.

2 Data and Methods

2.1 Background

There is a wide literature on the post-processing of model output for use in many applications. The methods cover a range of timescales, from weather to climate change. Model output can consist of daily, monthly,
seasonal or annual averages depending on the end use. The focus here is on daily model output of rainfall, minimum and maximum temperature and radiation as required to drive crop models, with forecasts extending out to nine months.

The simplest method of correcting bias in model forecasts is to apply a shift based on a comparison between the model and observed climatologies (Hansen et al. 2006; Ines and Hansen 2006). In the case of rainfall, a multiplicative adjustment is appropriate to preserve dry days. For temperature and radiation an additive shift is more common. More generally, linear regression can be used to correct the mean and variance of the forecasts, which in turn increases accuracy and statistical reliability (Johnson and Bowler 2009).

However, linear regression cannot correct frequency and intensity errors in model rainfall. The frequency and intensity of model rainfall must first be calibrated by determining a low rainfall threshold below which model rainfall is set to zero so that the number of dry days matches the long-term frequency (Ines and Hansen 2006; Schmidli et al. 2006). The intensity of rainfall can then be adjusted to preserve total model rainfall, or calibrated further to match observed rainfall.

Implicit in the use of linear regression are the assumptions that the model and observed variables are linearly related, and that the residuals after fitting have a normal distribution (Wilks 2007). The desirable properties of forecasts calibrated using linear regression involving only the first and second moments rely on the assumption that the observations and model forecasts are described by normal distributions (Johnson and Bowler 2009). While monthly and longer period averages of rainfall and other variables might have approximately normal distributions, this will not be the case for daily output. Therefore a more general technique for bias correction and calibration of model output is required.

Since there is no guarantee that model and observations will even have similar statistical distributions, the more general technique of quantile mapping is commonly used (Ines and Hansen 2006; Baigorria et al. 2007; Maraun 2013; Wilcke et al. 2013). This method calculates the observed and model cumulative distribution functions (CDFs) and transforms model values so that they have the same distribution as the observations. The CDFs can be determined empirically or by fitting an assumed distribution suited to the particular variable. For example, rainfall is often modelled using a two-parameter gamma distribution, while temperature might use a Gaussian distribution and radiation use a beta distribution (Baigorria et al. 2007). However, it can be
argued that maximum and minimum temperatures more closely follow an extreme value distribution (e.g. Wang et al. 2013). It has also been shown that it might be preferable to use a hyper-exponential distribution for rainfall (Hansen and Mavromatis 2001). To avoid unnecessary assumptions about distributions, and for simplicity, empirical CDFs will be used here.

Quantile mapping is an appropriate technique when there is no scale mismatch between the forecast and observations (Themeßl et al. 2011; Maraun 2013; Wilcke et al. 2013). If the model grid is 250 km on a side (as is POAMA’s grid), then model output ideally would be calibrated to observations averaged over a similar spatial scale. Point observations at stations within the grid box will have greater variability due to sub-grid scale features such as topography and convective storms. The process of taking grid-box scale forecasts and producing forecasts at a point within the box is called downscaling, or sometimes, disaggregation. If model forecasts are downscaled to point observations using a deterministic technique such as quantile mapping, and then aggregated back into area averages, the area average will be over-estimated because the method assumes all local variability is related to the grid-scale variability (von Storch 1999; Maraun 2013). It has also been noted that trends might be affected (Maraun 2013), but we would argue that this is to be expected and not a negative outcome on the seasonal timescale. In any event, this is of most relevance on the longer timescale of climate change.

There are a number of methods used for downscaling: embedded high-resolution dynamical models, statistical relationships with additive stochastic noise (which includes weather generators), and analogue methods based on the similarity of large-scale atmospheric patterns to historical occurrences (Hansen and Ines 2005; Hansen et al. 2006; Timbal et al. 2009). Downscaling is clearly an area of active research, and there is no one solution for all applications. Initial tests developed a simple linear relationship between AWAP averages on the POAMA grid scale and station values (von Storch 1999; Maraun 2013), and this linear correction was applied to POAMA forecasts calibrated to grid-average observations. There were two problems with this approach: the probability distribution and wet day frequencies of the corrected model output no longer matched the point observations, and it was not clear how to develop a suitable noise model for the stochastic component from the linear fit residuals.
In another test, POAMA output was again calibrated to AWAP averages on the POAMA grid scale using quantile mapping, and the process was then repeated to go from the grid scale down to the station scale. This method has the advantage of producing a time series that has the same statistical characteristics as the station data, although without the random noise component. However, when compared to a single step quantile mapping from model to station data, there was very little difference. Therefore we choose to use this simple procedure in the first instance. There is no area averaging of the output to bias results, since we are interested in point time series. Eventually it might be necessary to add stochastic noise, but the use of multiple forecasts from the POAMA ensembles may provide a sufficient random component. It is possible that the downscaling method will evolve throughout the project as we gain experience with the results of running crop models using forecasts.

There are a number of studies that explore the use of daily climate forecasts with crop models (Hansen and Indeje 2004; Hansen et al. 2006; Ines and Hansen 2006; Baigorria et al. 2007; Ines et al. 2011). The importance of bias correction and calibration in growing realistic crops emerges in all these studies. The other important issue to emerge was the length and timing of occurrence of dry spells. Unrealistically long dry spells were found to have a negative impact on crop yields (Ines and Hansen 2006; Ines et al. 2011). In contrast, Baigorria et al. (2007) found that bias correction using quantile mapping improved the number and length of dry spells. However, they also noted the importance of the timing of dry spells in relation to the crucial phases of crop growth. It is unlikely that any calibration technique could correct for this effect. The variation in the timing of rainfall events across ensemble members is simply part of the intrinsic variability associated with seasonal forecasting. Individual weather events have no predictability beyond 10-14 days; it is the average of weather that retains skill many months after the forecast starts.

2.2 Model Output

POAMA is a seasonal forecasting system that consists of a global atmospheric model, a global ocean model, a land surface model, an atmosphere/land initialization scheme, an ocean data assimilation system, and a coupled ensemble generation system (Hudson et al. 2013). Model forecasts are initialized several times every month and run for nine months. Retrospective forecasts (hindcasts) are available from 1981 to the present.
Model output consists of many atmospheric, land surface and ocean variables, including daily rainfall, daily minimum and maximum temperature and daily solar radiation reaching the ground. These are the essential quantities used to drive crop models such as APSIM. These variables are produced on the atmospheric model grid, which is a uniform global grid with side length of approximately 2.5 degrees. This equates to about 250 km north-south everywhere, and 240 km east-west at 30°S.

Each model run is initialized 11 times to produce an ensemble of outcomes. There are three slightly different model versions of POAMA that are each run in this way to provide a total of 33 ensemble members (Hudson et al. 2013). The ensemble spread gives a good indication of the range of possible outcomes given the imperfect model, initial conditions, relatively coarse grid size and chaotic nature of the climate system. We will eventually use ensemble forecasts to produce a spread of possible yield outcomes.

We use the 31 years from 1981 to 2011 as our standard testing and calibration period. Leave-one-out cross-validation will be used to minimise the risk of artificial skill (McIntosh et al. 2005).

### 2.3 Observations

Station observations of daily rainfall, temperature and solar radiation have been obtained from the SILO website (http://www.longpaddock.qld.gov.au/silo/). Any times for which there are missing values have been interpolated from surrounding stations (Jeffrey et al. 2001) so that the time series are complete. All stations used here have adequate surrounding stations for this interpolation to be reasonably accurate.

Initial tests used gridded values of daily rainfall and temperature (solar radiation was not available) calculated to match the POAMA 250 km grid by averaging very fine-scale gridded data from the AWAP project (Jones et al. 2009). The AWAP gridded data is a careful interpolation of in situ observations to a 0.05° (about 5 km) grid incorporating the effects of topography. The AWAP data is now managed and updated by the Australian Bureau of Meteorology.
The simple method eventually used did not require this gridded data. This overcomes the problem of obtaining a suitable gridded data set of solar radiation.

2.4 Method

The bias correction, calibration and downscaling procedure takes place in two steps for rainfall, but only the second step for temperature and radiation:

1. Rainfall frequency correction;
2. Quantile mapping from the model to the station observation;

Each step uses model output and data from all 31 years and within a sliding window of 31 days to ensure enough data is available to calculate accurately the empirical distributions (Themeßl et al. 2012; Wilcke et al. 2013). For forecasts within the first or last 15 days of the 9 month forecast period, the window is maintained at 31 days but is no longer centred. When using a much smaller window, or no window at all, or when using a fixed monthly window, artefacts such as jumps at month boundaries or excessive variability were noticed in the time series. This is attributed to under-sampling of the empirical CDFs for small windows, and time variability of the CDFs in the case of fixed monthly windows.

2.4.1 Rainfall frequency correction

The relative frequency of observed dry days, $f_d$, is determined by simply counting the number of days when observed rainfall is less than 0.1 mm/day and dividing by the total number of days, which is 31x31=961. The forecast rainfall is then sorted, and the smallest $f_d$ values are set to zero (Ines and Hansen 2006). If there are more dry days in the model than the observations, no change is made. Ines et al. (2011) propose adding rainfall days in this circumstance, although they note that this does not happen very often when comparing model gridded rainfall to station observations. Typically $f_d$ is between 0.7 and 0.9, meaning that observed rainfall occurs on 10%-30% of days.
2.4.2 Quantile mapping

The cumulative distribution function (CDF) gives the probability that a variable will exceed a given value. Generally the vertical axis shows probability and the horizontal axis shows the data values (Wilks 2007). The empirical CDF is obtained simply by counting the number of data values exceeding a given value and scaling by the total number of data points to get a relative frequency (probability). In the case of rainfall, only non-zero values are considered after frequency correction.

Quantile mapping involves calculating the model and observed empirical CDFs and mapping the model values at each level of probability to the observed values with the same probability (Themeßl et al. 2011; Maraun 2013; Wilcke et al. 2013). If there are different numbers of observations and model values, the probability values will not line up exactly. In this case, linear interpolation is used. Quantile mapping acts to transform the model values so that they have the same statistical distribution as the observations. Simpler techniques such as linear regression, bias correction (adjusting the mean only) or inflation (adjusting the variance) cannot change the shape of the statistical distribution.

The empirical CDFs are well-resolved because of the number of data points: 961 for each 31 day window of observations, and many more for the model because of the ensemble structure. POAMA consists of three slightly different models which each have 11 ensemble members. Each model is treated separately in case the different models have a different statistical structure. Hence the calculation of the empirical model CDF involves 961x11=10571 values.
3 Results

There are eight sites studied in this project covering the major grain growing regions of NSW, Victoria, South Australia and Western Australia. Site details are given in Table 1, and the locations are plotted in Figure 1. To demonstrate the calibration method here, we focus on Birchip.

Table 1. Details for study sites.

<table>
<thead>
<tr>
<th>Name</th>
<th>Station number</th>
<th>Lon (°E)</th>
<th>Lat (°S)</th>
<th>May-Oct ave rain (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birchip (Vic)</td>
<td>77008</td>
<td>142.85</td>
<td>35.92</td>
<td>239</td>
</tr>
<tr>
<td>Katanning (WA)</td>
<td>10579</td>
<td>117.56</td>
<td>33.69</td>
<td>357</td>
</tr>
<tr>
<td>Longerenong (Vic)</td>
<td>79028</td>
<td>142.30</td>
<td>36.67</td>
<td>284</td>
</tr>
<tr>
<td>Mingenew (WA)</td>
<td>08088</td>
<td>115.44</td>
<td>29.19</td>
<td>304</td>
</tr>
<tr>
<td>Northam (WA)</td>
<td>10111</td>
<td>116.66</td>
<td>31.65</td>
<td>332</td>
</tr>
<tr>
<td>Parkes (NSW)</td>
<td>65026</td>
<td>148.16</td>
<td>33.14</td>
<td>363</td>
</tr>
<tr>
<td>Roseworthy (SA)</td>
<td>23122</td>
<td>138.68</td>
<td>34.51</td>
<td>330</td>
</tr>
<tr>
<td>Temora (NSW)</td>
<td>73038</td>
<td>147.52</td>
<td>34.41</td>
<td>330</td>
</tr>
</tbody>
</table>

The skill of POAMA at predicting rainfall, temperature and radiation varies with season and location. The skill of the current model is available on the POAMA web site (http://poama.bom.gov.au/info/poama_skill.html), and documented in the many publications listed at http://poama.bom.gov.au/info/publications.html. To demonstrate the calibration method, we focus on forecasting conditions from June onwards with forecasts starting on 1 June for the years 1981-2011 at Birchip.
The spatial variation of POAMA correlation skill for June monthly-average rainfall is shown in Figure 2, where the observations from AWAP are averaged to the POAMA grid. The skill at the grid box in NW Victoria containing Birchip is 0.63. The station data and AWAP gridded data are correlated at 0.93, and the correlation
between the Birchip station data and POAMA is 0.53, showing how grid and station data can lead to slightly different results. In any event there is enough skill on the monthly timescale for this to be a suitable location for testing. Maximum and minimum temperatures have a similar level of skill when compared to gridded or station data, while solar radiation also shows a similar level of skill compared to station data. There is no gridded AWAP product for radiation. Skill decreases with lead time, with monthly rainfall skill falling below the 95% significance level of 0.36 in August and beyond for a forecast starting in June.

Recall that each day is calibrated using data from a 31 day moving window that is centred except at the start and end of the nine month forecast period. Furthermore, this calculation is cross-validated so that data from the year being calibrated is not used. This reduces the chance of artificial skill (McIntosh et al. 2005). Note that all ensemble members are combined to give 33 times as many days in June as there are observations.

We consider first the empirical cumulative density functions for the observations, raw model output and calibrated output for the month of June (see Figure 3). This diagram shows the probability that on days when there is rain, daily rainfall lies between 0.1 (the minimum rainfall amount) and the value on the x-axis. The observed and raw model output distributions are quite different, with the model being drier. For example, the probability that the observed daily rainfall (when it does rain) is less than 5mm is about 0.75, whereas the raw model has a probability of 0.87. This is the well-known “drizzle” effect due to the large grid size of the model. The calibrated model distribution, however, lies very close to the observed distribution. This is the effect of quantile mapping. It is not an exact match because of the sliding 31 day window. Each month of the nine month forecast period shows a similar behaviour, with the raw model output being quite different to the observations, and the calibrated output being very close.

The probability of a dry spell being longer than a certain number of days is shown in Figure 4 for June and for the full nine month forecast. The wet bias of the uncorrected model output is clear due to its lower probability of longer dry spells. Again, this is the “drizzle” effect due to the large spatial scale of the model. The calibrated model output has a more realistic dry spell probability, particularly for dry spells up to about a week in duration. There is some sampling error in this calculation in the case of a single month because longer dry spells are likely to overlap with adjacent months.
Calibration and bias correction of seasonal climate forecasts for use in agricultural models

Figure 3. Cumulative distribution functions for observations, raw model output and calibrated model output for June at Birchip. Forecast starts on 1 June.

Figure 4. Probability of a dry spell exceeding a certain number of days for (a) June, and (b) June-February.

The daily amount, intensity and frequency of daily rainfall are all improved considerably by calibration (see Figure 5) at every lead time. Furthermore, an increased ensemble spread is evident. Two desirable properties of an ensemble forecast are that: (i) the ensemble spread is representative of uncertainty in the ensemble mean, and (ii) each ensemble forecast has the same variance as the observations (Johnson and Bowler 2009). Quantile mapping is a more general version of the variance inflation method considered by Johnson and Bowler (2009), and is more appropriate when the underlying distributions are both non-normal and dissimilar between model and observations. Daily rainfall distributions are typically modelled with a gamma distribution (e.g. Ines and Hansen 2006), while daily maximum and minimum temperatures can be modelled...
with a normal distribution (Baigorria et al. 2007) although these data tend to be asymmetric and therefore less suitable for a normal distribution.

Figure 5. The amount (mm/day), intensity (mm/wet day) and frequency (wet days/day) of daily rainfall for the raw and calibrated model forecasts as a function of lead time. Observations (black, square symbols), model ensemble mean (blue, diamond symbols) and individual ensemble members (grey) are shown averaged for each lead month.

The two desirable properties (for normal distributions) can be re-stated as:

(a) The mean-square error (MSE) between the observations and the ensemble mean equals the climatological average of the ensemble variance;

(b) The climatological variance of the observations equals the climatological variance of the ensemble members all binned together.

These properties are closely satisfied by quantile mapping if the underlying distributions are not too different from normal. In the case of rainfall, which does not have a normal distribution, these properties still hold reasonably well (see Figure 6). In the top panel, the calibrated (green) curves are generally closer together.
than the red (raw forecast) curves. In the bottom panel, the observed and calibrated variances are generally very close, particularly at short lead times.

Figure 6. Monthly calibration statistics from Johnson and Bowler (2009) for Birchip rainfall forecasts starting on 1 June; (a) compares the mean square error of the ensemble mean with the climatological average of the ensemble variance; (b) compares the climatological variance of the observations with the average climatological variance of the ensemble members.

The mean square error of the rainfall forecast at Birchip in June happens to be relatively small; the model performs very well in this month, even without calibration, at forecasting the rainfall amount. However, the temperature forecasts are not so good, with the observed mean maximum temperature in June being 14.9°C and the raw forecast more than 3°C cooler. The calibrated forecast maximum temperature, however, is almost exactly right at 14.8°C. Figure 7 shows how calibration has reduced the mean square error for temperature considerably over the first four months of the forecast. Surface radiation shows a similar behaviour to maximum temperature, while minimum temperature shows a substantial reduction in mean square error over almost all forecast months (not shown).
Figure 7. Same as Figure 6 but for daily maximum temperature.

Mean square error and variance plots for all sites are shown in Appendix A. There is some variation between sites, but the general picture is consistent. In the top panel, the calibrated mean square error and the ensemble variance are similar at most lead times, indicating that the ensemble forecast has an appropriate spread. The mean square error is also generally much reduced after calibration. The bottom panel shows that the calibration has ensured that the variance of the ensemble members is consistent with the variance of the observations.
4 Conclusions

There are quite a number of methods for bias adjusting, calibrating and downscaling seasonal climate forecasts. The methods vary in complexity, computational cost and the metrics by which success is assessed. For the present application to incorporating POAMA seasonal forecasts into Yield Prophet, relative simplicity and computational speed are important considerations providing the results are satisfactory. The method should be capable of being explained to agricultural researchers, agronomists and interested farmers, and must run quickly if applications for tablets and smart phones are to be successful.

A method involving adjustment of wet day frequencies followed by quantile mapping directly from model ensemble output to site observations satisfies all the criteria. The results appear to show a reasonable wet day frequency, dry spell statistics, ensemble spread and climatological variance, and to have much-reduced bias as indicated by low mean square errors. Reservations voiced by some authors about the pitfalls of this method apply mainly to area averages and climate trends, neither of which are a concern in this project. It may be that at a later stage a different or improved method could be used or developed, perhaps incorporating the addition of random noise to simulate sub-grid scale processes. However, it would need to be demonstrated that the additional complexity lead to significantly improved results.

5 Acknowledgements

This work forms part of a project funded jointly by GRDC and CSIRO. Model output from POAMA has been made available by the Australian Bureau of Meteorology. The help of Griff Young from BoM is gratefully acknowledged.
6 Appendix A - Mean square error and variance plots for all sites

Figure 8. Monthly calibration statistics from Johnson and Bowler (2009) for Birchip forecasts starting on 1 June. Panel (1) solar radiation, (2) maximum temperature, (3) minimum temperature, and (4) rainfall. Within each panel (a) compares the mean square error of the ensemble mean with the climatological average of the ensemble variance; (b) compares the climatological variance of the observations with the average climatological variance of the ensemble members.
Figure 9. Same as Figure 8 but for Katanning.

Figure 10. Same as Figure 8 but for Longerenong.
Figure 11. Same as Figure 8 but for Mingenew.

Figure 12. Same as Figure 8 but for Northam.
Figure 13. Same as Figure 8 but for Parkes.

Figure 14. Same as Figure 8 but for Roseworthy.
Figure 15. Same as Figure 8 but for Temora.
7 References


