Australian Meteorological and Oceanographic Journal 63 (2013) 315–324

An evaluation of the precipitation forecasts of the Poor Man's Ensemble for wintertime rainfall across the southern portion of Australia

Christopher D. Johnson[1], Steven T. Siems[1], Michael J. Manton[1] and Elizabeth E. Ebert[2]
[1]School of Mathematical Sciences, Monash University, Monash, Australia
[2]Centre for Australian Weather and Climate Research,
Bureau of Meteorology, Melbourne, Australia

(Manuscript received December 2011; revised February 2013)

The Poor Man's Ensemble (PME) forecasts of daily rainfall have been investigated for the eight-year period of 2001–2008 with a focus on wintertime precipitation (May–October) across the southern portion of Australia (south of 30°S). Such rainfall is commonly the product of storm systems that come from the Southern Ocean (SO). The skill of the forecasts is assessed against the surface precipitation observations of the Australian Water Availability Project (AWAP). Moving beyond the basic scores such as the Frequency Bias and the Equitable Threat Score, as well as the Symmetric Extremal Dependency Index, the aim was to explore the sensitivity of the forecasts to geographic factors, specifically orographic enhancement and coastal boundaries. Eight different 1° × 1° grid boxes were chosen across the southern portion of Australia including three in Tasmania. The sensitivity to forecast length was also examined. The PME precipitation product shows considerable skill but is found to fare relatively poorly in the mountain and coastal regions. The PME precipitation for western Tasmania, which can be classified as both orographic and coastal, is roughly half of what is observed. The rainfall was broken down into light (≥1 mm per day) and heavy (≥10 mm per day) days. An analysis of the frequency bias score and equitable threat score finds the PME has particular trouble with the heavy precipitation days.

Introduction

The simulation of precipitation is wrought with challenges across scales ranging from the microphysics to synoptic scale dynamics. Even in more simplified conditions, specifically precipitation over the ocean, Stephens et al. (2010) concluded that the present generation of global weather prediction models consistently overestimates the frequency at which rain occurs but underestimates the intensity. These simulations were evaluated against the precipitation retrieved from CloudSat (Haynes et al. 2009) with the largest discrepancies occurring over the Southern Ocean (SO), predominantly between 45°S and 60°S. This difficulty in simulating the meteorology over the SO is similarly identified in Trenberth and Fasullo (2010) in an evaluation of the energy budget of both the reanalyses and coupled global climate simulations. It was concluded that too few clouds were being generated over the SO leading to relatively large biases in the absorbed short-wave radiation and outgoing long wave radiation.

This difficulty in simulating the meteorology over the SO is of direct consequence to Australia as these systems contribute the wintertime precipitation across the southern portion of the continent and Tasmania. 'Cut-off lows' originating in the SO storm track accounted for over 50 per cent of all wintertime precipitation and 80 per cent of the heavy rainfall days in the Mallee of the Murray-Darling Basin (Pook et al. 2006). Risbey et al. (2009) extended this climatology to explore connections to climate indices, such as El Niño-Southern Oscillation (ENSO), the Southern Annular Mode (SAM) and the Indian Ocean Dipole (IOD). In particular, positive SAM years were linked to 'wet' years across southeast Australia. Chubb et al. (2011) studied

Corresponding author address: Christopher Johnson, email: thisform2002@gmail.com

precipitation for the Snowy Mountains over a 20-year period (1990–2009) and, similar to Pook et al. (2006), found that 'cut-off lows' and 'embedded lows' accounted for the vast majority (96 per cent) of the wintertime precipitation. At these high elevations the precipitation was roughly a factor of four greater than over the upwind Mallee due to the orographic enhancement. In contrast to Risbey et al. (2009), however, Chubb et al. (2011) found that 'wet' years in the Snowy Mountains were linked to a negative SAM index. The argument presented was that during such years, the winds across the Snowy Mountains were more zonal leading to a greater production of orographic precipitation.

The difficulty in simulating precipitation over the southern portion of Australia is evident in Ebert et al. (2003), which evaluated the quantitative precipitation forecasts (QPFs) of six global and one regional weather prediction model across all of southeast Australia (east of 135°E and south of 25°S). In general, the prediction of rainfall across southeast Australian was more skilled than that across tropical Australia (north of 20°S), which is understandable; tropical rainfall is intense and highly variable. When compared with QPFs over the United States or Germany, however, those over southeast Australia fared poorly. Ebert et al. (2003) went on to highlight that QPFs made from an ensemble of forecasts offered a 'promising and practical' way to improve the overall skill.

The primary aim of this paper is to evaluate the Bureau of Meteorology's Poor Man's Ensemble (PME) precipitation forecasts (Ebert 2001) against observations from the Australian Water Availability Project (AWAP) (Raupach et al. 2008). Rather than looking at a broad region of analysis, such as all of southeast Australia, the focus is on looking at specific regions such as the Mallee, as investigated in Pook et al. (2006) or the Snowy Mountains, as investigated in Chubb et al. (2011). In all, eight 1° × 1° blocks (the native resolution of the PME) are explored spanning an eight-year period (2001–2008). First, the data products are more fully discussed, which is followed by the methodology. Next, the analysis is undertaken and expanded to examine the potential effect of orographic enhancement on the QPFs. Thresholds of 1 and 10 mm day^{-1} are tested to explore the assertion of Stephens et al. (2010) that the frequency and intensity of simulated precipitation is biased against observations.

Precipitation records

The QPFs evaluated herein are a product of the Bureau of Meteorology's PME (Ebert 2001), which is an ensemble composed of seven independent numerical weather prediction forecasts from various meteorological centres around the globe. Only the ensemble mean product was considered out of the PME; the probability of precipitation was not taken into consideration. This economical means of developing an ensemble forecast has the strength of spanning a variety of input data, physical parameterisations and forecast methodologies in spite of its limited size. Ebert (2001) demonstrated that the skill of the ensemble forecast exceeded that of any of the individual ensemble members and demonstrated skill in heavy precipitation (50 mm day^{-1}) at 24 hours and lighter precipitation at 48 hours. The seven individual members of the ensemble will have been improved over the eight-year period (2001–2008) meaning that it is expected that the PME forecasts will have improved, too (WMO 2008).

These QPFs are evaluated against the precipitation analyses developed for the AWAP (Raupach et al. 2008). AWAP precipitation analyses are, in general, more accurate than the Bureau of Meteorology's operational analyses (Weymouth et al. 1999) because they use data from a much larger number of rain gauges, and the interpolation scheme takes topographical effects into account. Daily precipitation is produced on a 0.05° × 0.05° grid across all of Australia. To make the resolution between the two datasets comparable, the 400 AWAP grid boxes within a single PME grid are simply averaged together.

The average annual precipitation (2001–2008) over the southern portion of Australia (south of 30°S) of the AWAP dataset and the PME 24-hour forecast (PME24) (Fig. 1) are in broad agreement with the greatest annual precipitation located along the west coast of Tasmania and the least over central Australia. Overall, however, the magnitude of the PME24 precipitation is moderate in comparison to the AWAP dataset. The inland dry regions of Western and South Australia are slightly drier than forecast by the PME24. Conversely the PME24 is found to under predict the magnitude of the precipitation in coastal and orographic regions. Much of the western portion of Tasmania is observed to be more than a factor of two greater than the PME24 average.

It is not surprising that the PME apparently has less skill in orographic regions due to its limited resolution. The PME resolution, 1° × 1°, is simply too coarse to resolve the mountains in any detail; the highest peaks that force the greatest precipitation are smoothed out. Poor resolution may also affect coastal precipitation, such as in southwest Western Australia, where sea breeze circulations will not be resolved.

In an effort to assess the relative strengths and weaknesses of the PME forecasts across these varying conditions, a set of eight PME grid boxes have been chosen. In Fig. 1, from left to right across the mainland the five of the eight grid boxes chosen are southwest Western Australia (SWWA), the Mallee (MAL), the Thomson Dam – Victorian Alps (TDAL), the Snowy Mountains (SNOW) and the grid to the northwest of Sydney (SYD) which extends into the Blue Mountains. As the primary focus is on weather systems originating from the SO, three grids are chosen across Tasmania: western Tasmania (WTAS), northern Tasmania (NTAS) and eastern Tasmania (ETAS). The average altitude of these regions is shown in Table 1. Shepherd (1995) investigated the rainfall across Tasmania and roughly divided the climatology into three such regions, where WTAS is the wettest region and dominated by SO weather systems. ETAS is the driest being

Fig. 1. The average annual rainfall from AWAP (top) and the 24-hour forecast from the Bureau of Meteorology's PME24 (bottom). Eight test regions are identified. Across mainland Australia (from west to east) they are southwest Western Australia (SWWA), the Mallee (MAL), Thomson Dam – Victorian Alps (TDAL), the Snowy Mountains (SNOW) and Sydney (SYD). The three sites across Tasmania are western, northern and eastern Tasmania (WTAS, NTAS and ETAS, respectively).

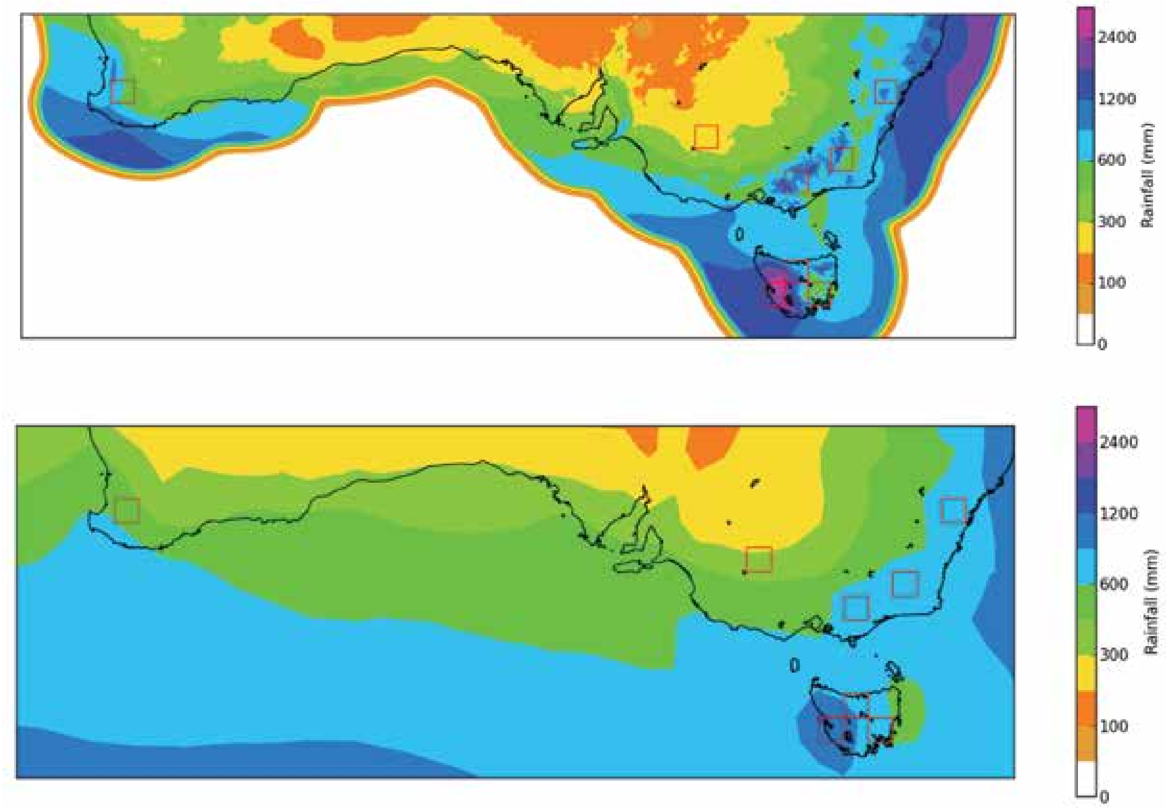


Table 1. Identification and characterisation of the eight test regions employed.

Location	*Abbreviation*	*Gridbox average altitude (m)*	*Coastal (Yes/No)*
Thomson Dam – Victorian Alps	TDAL	474	No
Snowy Mountains	SNOW	890	No
Northwestern Sydney	SYD	352	Yes
Western Tasmania	WTAS	193	Yes
Eastern Tasmania	ETAS	219	Yes
Northern Tasmania	NTAS	585	Yes
Mallee	MAL	90	No
Southwestern Western Australia	SWWA	237	No

located in the lee of the mountains, and NTAS is somewhat in between; it experiences heavy rainfall during frontal passages, but is in the lee of the mountains during post-frontal periods when the winds come off the SO from the southwest.

Figure 2 shows the average monthly rainfall (2001–2008) for these eight regions and highlights the seasonal cycle through the southern portion of Australia for TDAL, WTAS, NTAS, SNOW and SWWA. No significant seasonal cycle is evident for ETAS and only a weak cycle is evident for MAL. SYD offers a point of contrast as the region actually gets more precipitation during the summer months from sub-tropical systems, which are typically relatively intense and more difficult to forecast. It is worth noting that the peak in precipitation observed in the month of February is primarily in response to two large rainfall events in 2002 and 2008. Along the New South Wales coast, Speer et al. (2009) undertook a synoptic classification of precipitation events. They found that during the winter months, 'wave on a front lows' were the most common conditions to bring precipitation, but in general were associated with weaker precipitation events. This synoptic type is defined by the development of cyclonic rotation along a cold front that is grounded to a synoptic low located to the south, either a cut-off low or an embedded low in the terms of Chubb et al. (2011).

These annual cycles also more clearly illustrate the nature of the bias in the PME. For the majority of these regions, the overall bias/error in the PME24 forecast is greatest during the winter months. For example in SWWA, which has a relatively modest elevation, there is little difference between AWAP and PME24 during the summer and early autumn, but during the winter season the error is pronounced. In WTAS, by contrast, a mean bias exists year round although it is stronger in magnitude during the winter months.

Analysis

The accuracy of the PME forecasts will be assessed through the lens of the contingency table:

		Observed	
		Yes	*No*
Forecast	*Yes*	Hit	False alarm
	No	Miss	Correct negative

Contingency tables for the six 'winter' months (May–October) for the 8-year period (2001–2008) are calculated for each of the eight regions (Table 2). Heavy precipitation (10 mm day^{-1}) and light precipitation (1 mm day^{-1}) thresholds are considered for both the 24-hour and 72-hour (PME72) PME forecasts. Consistent with the climatology, WTAS is seen to have the greatest number of rainy days, both heavy and light, and MAL has the fewest. It is noted that the PME forecast was not available for 174 days of this period due to product development. Such days were removed from any further analysis. There are also very few heavy precipitation days recorded for MAL and ETAS meaning that it is difficult to draw robust conclusions about such events at these locations.

Based on the contingency table three metrics have been defined to measure various aspects of a forecast. They are: the Frequency Bias Score (FBS), Equitable Threat Score (ETS) and the Symmetric Extremal Dependency Index (SEDI). The frequency bias score,

$$FBS = \frac{hits + false\ alarms}{hits + misses}$$

simply compares the frequency at which events are forecast to the frequency at which they are observed. The FBS ranges from a value of zero (no forecast events) to infinity (no observed events) with an ideal score of one occurring when the forecasts are predicting events at the same frequency as observed, although the timing of such events is not considered.

Consistent with the assertion of Stephens et al (2010), both the PME24 and PME72 forecasts are found to over predict the occurrence of light precipitation at all regions except WTAS (Fig. 3). Conversely, both the PME24 and PME72 are found to consistently under predict the number of heavy rainfall days, especially across Tasmania. By this measure the under prediction of the heavy precipitation events is more frequent than the over prediction of the light events. It is further noted that the FBS slightly degrades when moving from the PME24 to the PME72.

Focusing on light precipitation at 24 hours, a weak relationship can be drawn to elevation. The lower elevation regions (MAL, ETAS and NTAS) perform worse than the high elevation regions (WTAS, SNOW and TDAL). SYD marginally scores the poorest of all regions, but these heavy rainfall events commonly arise from subtropical systems, even during winter (Speer et al. 2009). SWWA is most interesting as this region is actually predominantly at low elevation (<300 m), yet the FBS scores are close to one for light rainfall at both 24 and 72 hours.

When considering heavy rainfall, it is difficult to make out much of a geographic signal. Here SYD stands out as having considerably more skill. It is difficult to place much confidence in the results for MAL and ETAS considering how few heavy precipitation days are recorded.

The equitable threat score (ETS) seeks to quantify the skill in forecasting the timing of events. First a measure of the unskilled or random chance of correctly forecasting an event is quantified,

$$random\ hits = \frac{(hits + misses)(hits + false\ alarms)}{(hits + misses + false\ alarms + correct\ negatives)}$$

and then the ETS is defined as,

$$ETS = \frac{(hits - random\ hits)}{(hits + misses + false\ alarms - random\ hits)}$$

An ideal score of one would reflect that there were no missed events or false alarms. The ETS ranges from –1/3 to 1 with a value of zero suggesting no skill.

Focusing first on light precipitation for the PME24 forecast (Fig. 3), the best ETS are recorded for SWWA, TDAL and SNOW. Values of 0.5 and greater compare favourably with those recorded in Ebert et al. (2003) for southeast Australia, tropical Australia, the US and Germany, which likely reflects the increasing skill of the individual components and thus the enhanced skill of an ensemble forecast (WMO 2008). The scores are observed to decrease quickly when moving from the PME24 to PME72 datasets in all areas except SYD. The ETS scores drop even more severely when moving to heavy rainfall days, which suggests that there is relatively little skill in the timing of heavy precipitation forecasts. When examining the ETS scores for elevation or geography, there is little evidence of a consistent pattern.

The Symmetric Extremal Dependency Index (SEDI) (Ferro and Stephenson 2011) is a measure of the skill when it comes to extreme events, in this case exceeding the precipitation thresholds (1 mm and 10 mm). To calculate this index, the hit rate (H):

$$H = \frac{hits}{hits + misses}$$

Table 2. Contingency tables for the eight test regions for the months May–October. Rainfall thresholds of 1 and 10 mm day^{-1} are employed for both the PME24 and PME72 forecasts.

		>1 mm		*>10 mm*	
		Observed Y	*Observed N*	*Observed Y*	*Observed N*
TDAL					
PME24	*Forecast Y*	429	109	49	8
	Forecast N	46	714	40	1201
PME72	*Forecast Y*	355	251	29	13
	Forecast N	120	572	60	1196
SNOW					
PME24	*Forecast Y*	387	95	49	6
	Forecast N	53	763	59	1184
PME72	*Forecast Y*	307	236	30	18
	Forecast N	133	622	78	1172
SYD					
PME24	*Forecast Y*	279	127	55	16
	Forecast N	42	850	29	1198
PME72	*Forecast Y*	228	202	23	38
	Forecast N	93	775	61	1176
WTAS					
PME24	*Forecast Y*	790	47	143	6
	Forecast N	122	339	213	936
PME72	*Forecast Y*	766	131	111	45
	Forecast N	146	255	245	897
ETAS					
PME24	*Forecast Y*	368	169	19	5
	Forecast N	53	708	32	1242
PME72	*Forecast Y*	278	261	9	13
	Forecast N	143	616	42	1234
NTAS					
PME24	*Forecast Y*	445	133	41	13
	Forecast N	40	680	65	1179
PME72	*Forecast Y*	382	236	22	27
	Forecast N	103	577	84	1165
MAL					
PME24	*Forecast Y*	231	97	7	4
	Forecast N	36	934	11	1276
PME72	*Forecast Y*	159	173	4	8
	Forecast N	108	858	14	1272
SWWA					
PME24	*Forecast Y*	544	82	89	8
	Forecast N	65	607	78	1123
PME72	*Forecast Y*	442	176	55	36
	Forecast N	167	513	112	1095

and the false alarm rate (F):

$$F = \frac{false\ alarms}{false\ alarms + correct\ negatives}$$

are calculated and from these two indices, the SEDI is then defined as:

$$SEDI = \frac{log(F) - log(H) - log(1 - F) + log(1 - H)}{log(F) + log(H) + log(1 - F) + log(1 - H)}$$

The SEDI ranges from –1 to 1, with one indicating perfect performance with events that exceed the precipitation upper bound.

Focusing on light precipitation at 24 hours (Fig. 3), all the regions recorded an SEDI score of at least 0.8. Moving towards the heavier rainfall thresholds and longer forecast time, SYD and all regions in TAS performed worse than the other areas with a SEDI score close to 0.5. As discussed in Speer et al. (2009), precipitation in the SYD region is commonly dominated by sub-tropical systems rather than mid-latitude systems. The poor forecasting of such events could readily lead to the poorer SEDI scores.

Ebert et al. (2003) demonstrated a seasonal cycle in the skill scores over southeast Australia: the FBS of the seven models were, on average, closer to one during the winter (JJA) and spring (SON) seasons and more variable during summer (DJF) and autumn (MAM). Similarly the ETS of the models peaked, on average, in winter or spring. Repeating this analysis for the eight individual regions for the PME24 dataset, a similar behaviour is observed. The time series of the FBS averaged into 'high elevation' and 'low elevation' regions (Fig. 4) show increasing skill during winter and spring but also illustrates the large grid-to-grid variability within the PME. This observation is not immediately apparent when examining the time series of the ETS. While the average ETS does marginally improve during the winter, the time series is dominated by the large variability between regions.

Finally the effect of any systematic error/bias in the forecasts is considered. For example in regions like WTAS or SWWA, the PME rainfall accumulation forecasts are, on average, considerably smaller than observed, which will readily affect the skill scores especially at the heavy precipitation threshold. An immediate means of addressing such shortcomings is to simply scale the PME forecasts. Such a re-scaling can be undertaken on a month-by-month basis for each of the eight regions. For example, the average rainfall for August over the WTAS domain was observed to be ~200 mm while the PME24 forecast it to be ~100 mm (Fig. 2). For this month all the daily PME24 forecasts were according doubled (approximately). In general this re-scaling will lead to more forecasts of both heavy and light precipitation days, especially in the mountain regions.

The bias-adjusted FBS and ETS are presented in Fig. 5 and compared against those in Fig. 3. Looking first at the FBS, it is observed that the skill decreases at the light rainfall threshold (e.g. increasing from 1.2 to 1.25 for the PME24 at TDAL) but increases significantly at the high precipitation threshold (e.g. increasing from 0.65 to 0.88 for the PME24 at TDAL). Such changes are largely observed for all regions except MAL, where the FBS actually decreases for both the heavy and light rain days. This is consistent as the PME forecasts were larger than the observations in most months in the region (Fig. 2). Overall, the re-scaling most dramatically improved the forecasting of heavy precipitation days, especially in the high elevation regions (WTAS, TDAL and SNOW).

Turning to the ETS, rescaling will not affect the timing of events, but certainly some events will be reclassified. In

Fig. 2. The average monthly rainfall (AWAP – blue; PME24 – green; PME72 – red) for the eight test regions for 2001–2008.

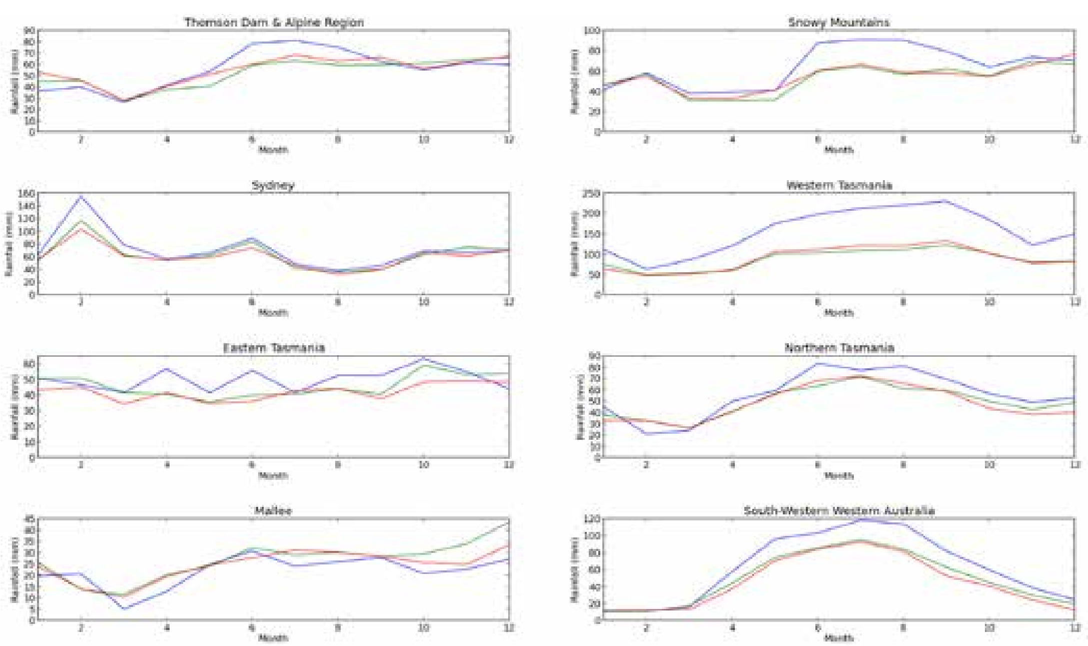


Fig. 3. Frequency Bias Score (FBS) (top), Equitable Threat Score (ETS) (middle) and Symmetric Extreme Dependency Index (SEDI) (bottom) for both the 24 and 72-hour PME forecasts for the six winter months (May–October) from 2001–2008.

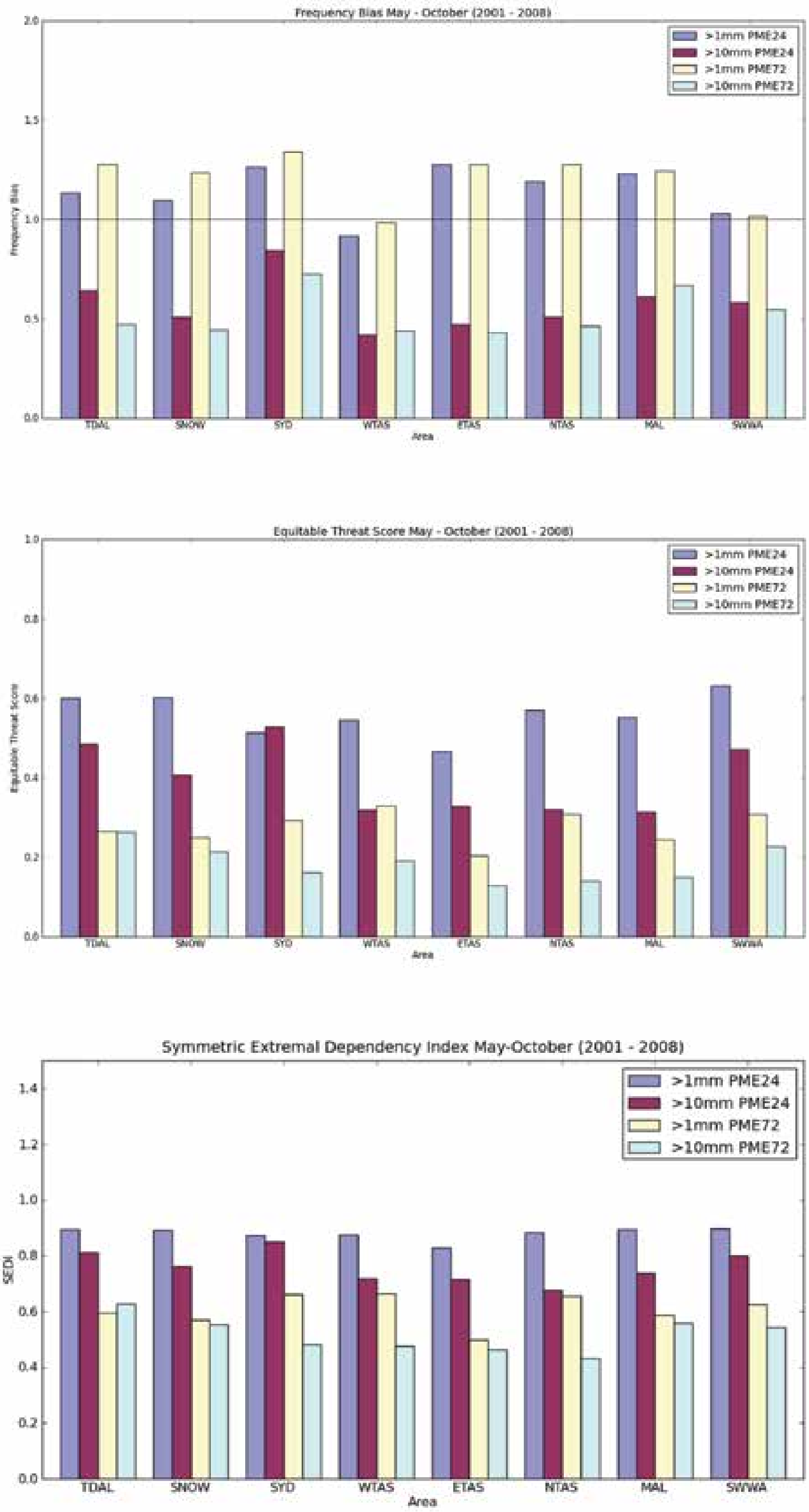
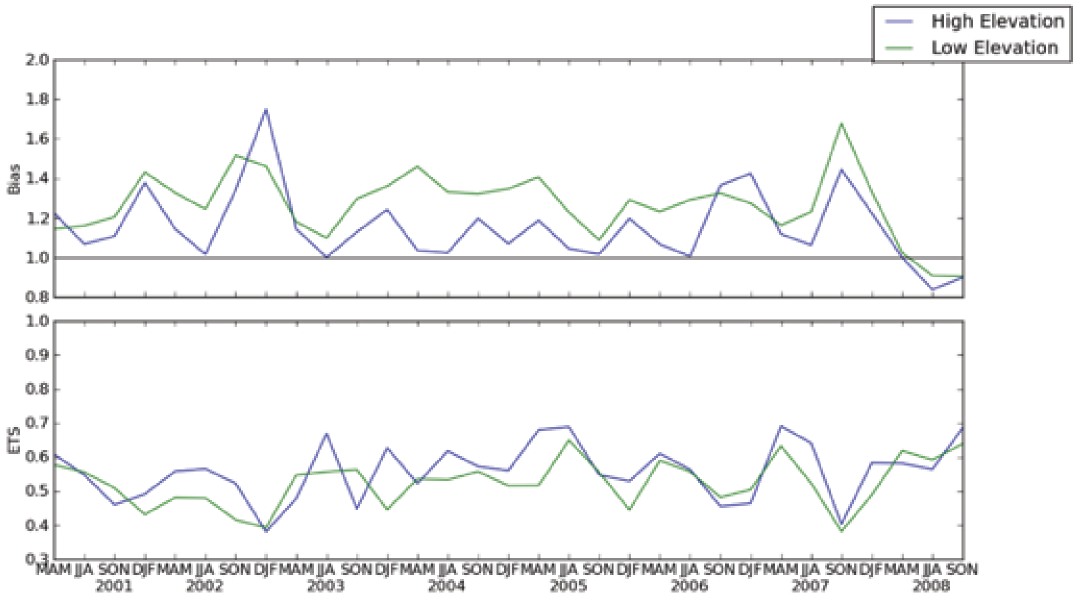


Fig. 4. The seasonal FBS and ETS of the high elevation regions (TDAL, SNOW and WTAS) and the low elevation regions (SWWA, MAL, SYD, NTAS and ETAS).



general, the re-scaling does not significantly change the ETS for light rainfall days but does improve the ETS for heavy rainfall days in all regions except MAL (Fig. 5).

The SEDI index improves across all the regions for rainfall above 10 mm when monthly biases are removed, outperforming that of the 1 mm threshold for most of the regions, except for MAL where extreme events are harder to predict.

It is interesting to revisit the question about the sensitivity of the FBS and ETS to geographic locations with the rescaled forecasts. For the unscaled data little signal could be attached to the effect of elevation or geographic location (Fig. 3), although WTAS and SWWA did display the least bias for light rainfall. These are the regions that are most immediately exposed to storm systems off the SO. No pattern was evident in the ETS of the unscaled PME forecasts. When the forecasts are scaled, however, the high elevation regions (TDAL, SNOW and WTAS) perform relatively well for both the ETS and FBS, especially for heavy precipitation. These enhanced scores in mountain regions suggest that, in general, errors in the timing of precipitation events are not as significant as errors in the magnitude. This is likely to simply be a reflection that rainfall is more common in mountain regions and timing is less of an issue. More specifically, it is likely that relatively more precipitation in mountainous areas is forced by the local orography, and thus there is greater predictability in such regions.

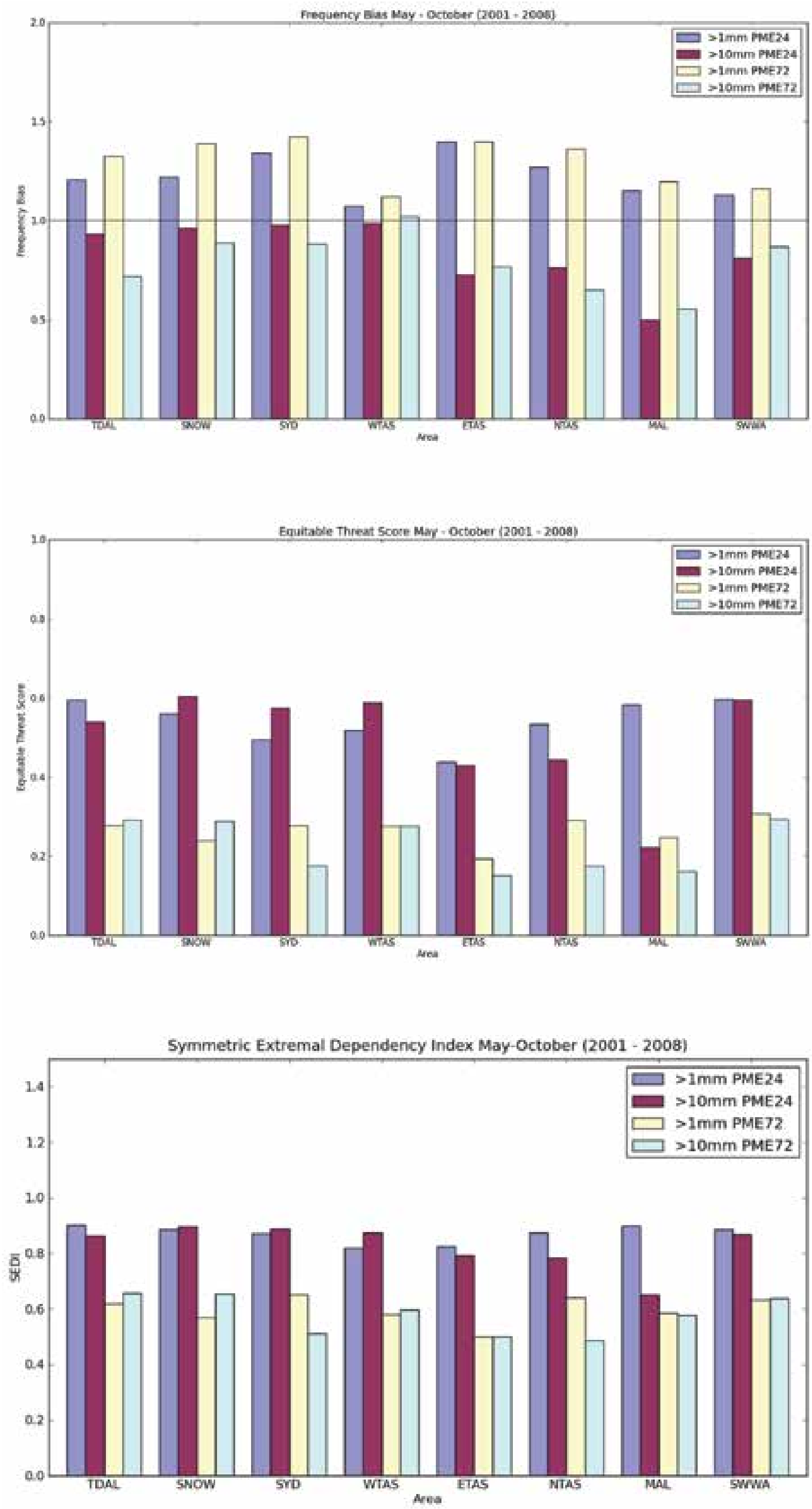
Discussion and conclusions

The PME forecasts of daily rainfall have been investigated for the eight-year period of 2001–2008 with a focus on wintertime precipitation (May–October) across the southern portion of Australia (south of 30°S). Such rainfall is commonly the product of storm systems off the SO. The skill of the forecasts is assessed against the surface precipitation observations of the Australian Water Availability Project (AWAP). Moving beyond the basic scores, the aim has been to explore the sensitivity of this skill to matters of forecast length, orographic enhancement and basic geographic location by looking at individual regions. Eight different (1° × 1°) domains were chosen across the southern portion of Australia including three in Tasmania.

A direct comparison of the annual precipitation from the 24-hour PME forecasts against the AWAP observations reveals a few immediate biases in the forecasts. Overall the PME precipitation is moderated in comparison to the observations: central Australia is slightly too wet, while the coastal regions are not wet enough. The PME has particular difficulty in the high elevation regions along the Great Dividing Range and Tasmania, which is to be expected; the PME and its ensemble members are generally at poor resolution and simply cannot resolve the orography and thus the orographic enhancement. Figure 1 also suggests that the PME has difficulty with coastal precipitation, especially along southwest Western Australia, southern South Australia/southwest Victoria and western Tasmania. These are regions most immediately exposed to wintertime storm systems off the Southern Ocean. Western Tasmania suffers from both of these conditions and suffers poorest forecasts; the annual precipitation from the 24-hour PME forecasts is less than half of that observed. The eight grids across the southern portion of Australia and Tasmania were chosen to sample varying geographic conditions.

These shortcomings were quantified through three common skill metrics, the frequency bias score (FBS), the

Fig. 5. The same as Fig. 3 only the PME24 and PME72 forecasts have been scaled to remove mean monthly biases for each of the eight test regions, individually. .



equitable threat score (ETS) and the Symmetric Extreme Dependency Index (SEDI). The FBS measures the skill in frequency at which events are forecast, and the ETS measures the skill at the timing at which events are forecast. Both light rainfall days (>1 mm day^{-1}) and heavy rainfall days (>10 mm day^{-1}) are defined as events. The 72-hour PME forecasts were also considered. The SEDI is a measure of the skill when it comes to extreme events.

In general it was difficult to make conclusive statements based on the FBS and ETS. Perhaps surprisingly, the FBS for the high elevation regions (TDAL, SNOW, WTAS) and the coastal region (SWWA) for light precipitation fared better (closer to a value of one) than for the other regions. During the winter these regions experience rainfall quite frequently. For heavy rainfall, however, these regions fared no better or worse than the others. As for the timing of these rainfall events (both light and heavy), a marginally higher ETS is observed for TDAL, SNOW and SWWA, but not WTAS.

When the PME rainfall was scaled to remove systematic error/bias, the skill in the mountain regions improved most noticeably for heavy precipitation events. Little change was noted for light precipitation. While scaling the PME forecasts was simple and effective, it provided no insight into the mechanism by which orography or a coastal circulation modifies rainfall. For example, Morrison et al. (2011) recently noted that the wintertime precipitation across the mountains of southeast Australia and Tasmania contains relatively large amounts of supercooled liquid water, which may directly effect the simulation of precipitation through the microphysics scheme. A simple scaling of the PME forecasts simply merges an adjustment to the precipitation due to orography with any potential adjustment due to microphysics.

Some broader conclusions have been reached by this exercise, as well, when considering the earlier analysis of Ebert et al. 2003. While this earlier work covered only a 39-month period (1997–2000) for a broad region of all of southeast Australia, it illustrated a seasonal cycle in the FBS and ETS with increasing skill in the winter months. Such a cycle is evident in the PME forecasts, too, although there is considerable noise on the scale of an individual grid box. Further the FBS and ETS skill level has risen substantially since this earlier analysis. Independent tests by one of the authors suggest that this gain is primarily attributable to the ensemble approach of the PME.

Returning to the conclusion of Stephens et al. (2010) that global weather prediction models tend to over estimate the frequency of precipitation but under estimate its intensity, this work supports the conclusion in southern Australia. The FBS score suggested that light rainfall was over predicted but heavy rainfall was under predicted and is attributed to the ensemble framework, where a single heavy precipitation forecast may readily be moderated by milder forecasts of other ensemble members.

Acknowledgments

The authors thank Ms. Jingru Dai for her assistance with the AWAP dataset.

References

Chubb, T.H., Siems, S.T. and Manton, M.J. 2011. On the decline of wintertime precipitation in the Snowy Mountains of South-Eastern Australia. *J. Hydrometeor., 12,* 1483–97, DOI: 10.1175/JHM-D-10-05021.1.

Ebert, E.E. 2001. Ability of a Poor Man's Ensemble to Predict the Probability and Distribution of Precipitation. *Mon. Weather Rev., 129,* 2461–80.

Ebert E.E., Damrath, U., Wergen, W. and Baldwin M.E. 2003. The WGNE assessment of short-term quantitative precipitation forecasts. *Bull. Am. Meteorol. Soc., 84,* 481–92.

Ferro, C.A.T. and Stephenson, D.B. 2011. External Dependence Indices: Inproved Verification Measures for Deterministic Forecasts of Rare Binary Events. *Weather and Forecasting, 26,* 699–713.

Haynes, J.M., L'Ecuyer, T.S., Stephens, G.L., Miller, S.D., Mitrescu, C., Wood, N.B. and Tanelli, S. 2009. Rainfall retrieval over the ocean with spaceborne W-band radar. *J. Geophys. Res., 114,* D00A22, DOI: 10.1029/2008JD009973.

Morrison, A.E., Siems, S.T. and Manton, M.J. 2011. A Cloud-top Phase Climatology of Southern Ocean Clouds. *J. Clim., 24,* 2405–18, DOI10.1175/2010JCLI3842.1.

Pook, M.J., McIntosh, P.C. and Meyers, G. 2006. The Synoptic Decomposition of Cool-Season Rainfall in the Southeastern Cropping Region. *J. Appl. Meteorol. Climatol., 45,* 1156–70.

Raupach, M.R., Briggs, P.R., Haverd, V., King, E.A., Paget, M. and Trudinger, C.M. 2008. Australian Water Availability Project (AWAP). CSIRO Marine and Atmospheric Research Component: Final Report for Phase 3. http://www.csiro.au/awap/doc/AWAP3.FinalReport.20080601.V10.pdf.

Risbey, J.S., Pook, M.J., McIntosh, P.C., Ummenhofer, C.C. and Meyers G. 2009. Characteristics and variability of synoptic features associated with cool season rainfall in southeastern Australia. *Int. J. Climatol., 29,* 1595–613.

Shepherd, D. 1995. Some characteristics of Tasmanian rainfall. *Aust. Meteorol. Mag., 44,* 261–74.

Speer, M.S., Wiles, P. and Pepler, A. 2009. Low pressure systems off the New South Wales coast and associated hazardous weather: establishment of database. *Aust. Met. Oceanogr. J., 58,* 29–39.

Stephens, G.L., L'Ecuyer, T., Forbes, R., Gettlemen, A., Golaz, J.-C., Bodas-Salcedo, A., Suzuki, K., Gabriel, P. and Haynes, J.M. 2010. The dreary state of precipitation in global models. *J. Geophys. Res., 115,* D24211, DOI:10.1029/2010JD014532.

Trenberth, K.E. and Fasullo, J.T. 2010. Simulation of Present-Day and Twenty-First-Century Energy Budgets of the Southern Oceans. *J. Clim., 23,* 440–53.

Weymouth, G., Mills, G.A., Jones, D., Ebert, E.E., and Manton, M.J. 1999. A continental-scale daily rainfall analysis system. *Aust. Meteorol. Mag., 48,* 169–79.

World Meteorological Organization. 2008. Report of the 23rd Session of the Working Group on Numerical Experimentation (WGNE), Shanghai, China, 22–26 October 2007. WMO TD No. 1429. World Meteorological Organization, Geneva.