Australian Meteorological and Oceanographic Journal 61 (2011) 171–184

Evaluating the TRMM 3B43 monthly precipitation product using gridded raingauge data over Australia

K. Fleming[1,2], J.L. Awange[1,3], M. Kuhn[1] and W.E. Featherstone[1]

[1]Western Australian Centre for Geodesy and The Institute for Geoscience Research, Curtin University, GPO Box U1987, Perth, Western Australia, 6845, Australia.

[2]Earthquake Risk and Early Warning, Helmholtz Centre Potsdam German Research Centre for Geosciences, D-14467, Potsdam, Germany.

[3]Geodetic Institute, Karlsruhe Institute of Technology, Engler-Strasse 7, D-76131, Karlsruhe, Germany.

(Manuscript received June 2010; revised October 2011)

This work compares two monthly rainfall datasets for Australia for the period between January 1998 and December 2010: The *TRMM (Tropical Rainfall Measuring Mission) and Other Sources Rainfall Product (3B43)* and the Australian Bureau of Meteorology (BoM) monthly gridded values based on raingauge data. The spatial variability in the cross correlation between each dataset over time reveals values > 0.90 for areas with denser ground-based networks of raingauges and higher rainfall, falling to around 0.20 for more arid areas with fewer raingauges. There appears to be some variation in the cross correlation over time, with values on average decreasing slightly after 2004, accompanied by increased scatter. The application of principal component analysis confirms the high correlation between these rainfall products, demonstrated by very similar spatial and temporal variability patterns. Histograms reveal that both rainfall products follow an almost identical probability distribution. The austral summer appears to show the strongest correlation between datasets, with the other seasons being slightly less correlated, but very similar to each other, although the differences between all seasons are minor with cross correlation values > 0.80 for Australia as a whole. Regarding the correlation between datasets as a function of climate, the steppe and temperate climate zones show a higher correlation than the tropical and desert zones. It is therefore proposed that the TRMM 3B43 monthly product would be appropriate for use in climate studies in Australia, keeping in mind that some discrepancies between the datasets for individual months arise. It is also potentially of value for the areas of Australia with fewer raingauge stations.

Introduction

Changing rainfall patterns and the effect this has on water resources is one of the possible consequences of climate change (Solomon et al. 2007). Therefore, it is essential that an accurate picture of precipitation and its changing nature over adequate spatial and temporal resolutions be available (e.g. Meinke et al. 2005). Considering the case of Australia, this is problematic given that a significant proportion of the continent has very few, if any, weather recording stations (e.g. Jones et al. 2009). While the sparsely instrumented areas are generally those with an arid climate (and as a result, little or no population or agricultural activity), there are others, for example in the north of the country, where rainfall is very high yet there is a dearth of ground-based observations owing again to the low population and limited agricultural or other activities.

This is a prime example of where satellite-derived rainfall estimates may be of benefit for filling in gaps in a ground-based data-acquisition network. Satellite- and ground-based observations complement each other; while satellite

Corresponding author address: K. Fleming, Earthquake Risk and Early Warning, Helmholtz Centre Potsdam German Research Centre for Geosciences, D-14467, Potsdam, Germany
email: kevin@gfz-potsdam.de

observations potentially provide global coverage, including less accessible areas (in particular the oceans), ground-based networks have the advantage of providing—for their point locations—better temporal resolution. Nonetheless, there must be confidence in the estimates provided by satellites, and this can only be gained when they are validated by ground-truth data, e.g. raingauges (e.g. Debo Adeyewa and Nakamura 2003; Nicholson et al. 2003) or ground-based radar (e.g. Schumacher & Houze 2000; Liao & Meneghini 2009).

An important satellite mission for this purpose is the Tropical Rainfall Measuring Mission (TRMM), a joint project between the United States (National Aeronautics and Space Administration, NASA) and Japan (Japan Aerospace Exploration Agency, JAXA). TRMM products have already been used for Australian meteorological studies, although more for gaining insight into daily rainfall patterns. For example, Ebert et al. (2007) included TRMM products in their comparison of near-real-time satellite observations with numerical models. Oke et al. (2009) used TRMM data to assess the interpolation of daily rainfall around Australia, while Huang & Bringi (2009) compared TRMM products with data from a rainfall radar near Brisbane, eastern Australia. Other uses of TRMM products have involved the correlation of soil moisture and vegetation density with El Niño events (e.g. Liu et al. 2007), and its use in conjunction with results from the Gravity Recovery and Climate Experiment (GRACE, e.g. Awange et al. 2009, 2011; Rieser et al. 2010).

The primary aim of this paper deals with the validity of the *TRMM and Other Sources Rainfall Product* (denoted as 3B43) monthly dataset when compared to a gauge-only analysis provided by the Australian Bureau of Meteorology (BoM) for the period between January 1998 to December 2010. While there are fine-temporal resolution products provided by TRMM (e.g. a multiple satellite analysis product, 3B42, with a three-hour temporal resolution) that have the potential to assist in detailed hydrological modelling (e.g. Su et al. 2008), our interest is more concerned with general trends in climate change, hence not requiring highly-temporally resolved data. The TRMM 3B43 product has been assessed in a number of studies, with those of particular relevance to this work being concerned with its validity to Africa (Debo Adeyewa & Nakamura 2003; Nicholson et al. 2003). However, to our knowledge, such an assessment has not been carried out for Australia.

The cross correlation (over space and time) between the TRMM and BoM datasets is determined to see how well they match. This is done with respect to (i) Australia as a whole, (ii) seasonal variations, and (iii) the various climatic zones that Australia is divided into, which also generally represent areas of different raingauge coverages. Principal component analysis (PCA), a statistical method for identifying the most dominant patterns of spatial and temporal variability within a dataset (e.g. Preisendorfer 1988), is also carried out. The use of PCA is motivated by its successful application in the analysis of rainfall over Australia as observed by TRMM (e.g. Awange et al. 2011; Rieser et al. 2010). In essence, this paper sets out to answer two questions: (1) does the TRMM 3B43 product produce an adequate representation of monthly rainfall in Australia, as verified by areas with dense ground-truth data; and (2) would the TRMM product be a more appropriate dataset to use for regions of Australia with sparse raingauge coverage?

Datasets

The *TRMM and Other Sources Rainfall Product* (3B43)

TRMM was launched on 27 November 1997 from the Tanegashima Space Center, Japan. It orbits at an altitude of around 400 km with an inclination of 35° and an orbital period of around 92.5 minutes, completing approximately sixteen orbits each day. The primary objective of TRMM is to measure rainfall and energy exchange over the tropical and subtropical regions of the world, in particular covering the oceans and unsampled land areas (Kummerow et al. 1998, 2000). TRMM provides a variety of products to meet the needs of different users. For this work, we employ one of the Level 3 products, which are space-time grids representing average rates of rainfall over different time scales (e.g. Kummerow et al. 2000).

The product used in this paper is the *TRMM and Other Sources Rainfall Product* or 3B43, (e.g. Huffman et al. 1995; Kummerow et al. 2000; Huffman et al. 2007). It is a time-series of monthly average rainfall (in mm hr^{-1}), inferred from data provided by multiple satellites in addition to TRMM, as well as raingauge data provided by the Global Precipitation Climatological Center (GPCC, produced by the German Weather Service) and the Climate Assessment and Monitoring System (CAMS), produced by the National Oceanic Atmospheric Administration (NOAA). Additional details may be found on the TRMM 3B43 webpage[1].

The satellite data is provided by two types of sensors: (1) passive microwave data collected by various low-Earth orbiting (LEO) satellites, including TRMM, the US Defence Meteorological Satellite Program satellites, and the NOAA satellite series; and (2) infrared data collected by an international constellation of geosynchronous satellites (Kummerow et al. 2000). The TRMM satellite itself has multiple sensors, the most important being the TRMM precipitation radar (TRMM-PR), the passive TRMM microwave imager (TMI), and the visible and infrared scanner (VIRS). Each of these provides products in its own right for the scientific community. However, the interest of this work is only in the 3B43 product because it has been found to be more representative (e.g. Nicholson et al. 2003).

The 3B43 data is provided in the Hierarchical Data Format (HDF) on a 0.25° × 0.25° grid that covers the globe between latitudes 50°N to 50°S. Improving the TRMM products is an ongoing process with superior versions of the

[1]http://disc.sci.gsfc.nasa.gov/precipitation/documentation/TRMM_README/TRMM_3B43 readme.shtml

various products released when appropriate (e.g. Shige et al. 2006). The version used in this work is number 6, although the months released from May 2007 onwards are referred to as version 6A[2]. For this study, areas outside of Australia have been masked out (i.e. the grid cell values have been set to zero) to enable a more direct comparison with the BoM rainfall product (described below) on land.

Bureau of Meteorology (BoM) rainfall product

The BoM rainfall product of interest is provided in the form of 0.05° × 0.05° total monthly rainfall grids. These grids are derived from raingauge data and are produced on the last day of each calendar month. They are then updated at later dates when more data arrives since many observation records are still provided in the form of paper-based documentation. It is also during this process that suspect values are confirmed and removed if found to be erroneous (see the BoM website[3]).

The BoM analyses use a hybrid analysis technique (Jones et al. 2009), comprising (i) a Barnes successive correction technique analysis of the monthly fraction of the climatological monthly average and (ii) a three-dimensional (topography resolving) thin-plate smoothing-spline analysis of the climatological monthly average, which are multiplied together to form the analysis which is in effect a weighted average of the raingauge data. For the purpose of this work, these grids are interpolated to the same 0.25° × 0.25° resolution as the TRMM time-series by fitting a minimum curvature spline (e.g. Smith & Wessel 1990) using the surface routine of the Generic Mapping Tools suite of programs (Wessel & Smith 1998), and then masked in the same manner as the TRMM grids to allow direct comparisons to be made between the TRMM and BoM datasets.

The information provided by BoM with regards to the rainfall observation stations[4] includes their locations and the period over which they operated, as well as the completeness of the record for each station, expressed as a percentage. Figure 1 presents various details about the BoM rainfall measuring network (see also Oke et al. 2009; Jones et al. 2009). Over 17 000 stations, at some time or another, have provided rainfall data over the Australian continent, with the earliest raingauge station dating from 1834. Of these, ~ 4900 were continuously operational during the full period covered by this study (January 1998 to December 2010), the distribution of which is shown in Fig. 1(a). The raingauge stations are concentrated in the more populated regions, in particular along the east coast and the southwest of the continent. Figure 1(b) shows how the numbers of stations has varied over time. Note that while the number of stations operating at any one time may exceed the number of stations that continuously operated during the study period, it is the latter that is of interest here, as they will define which grid cells will be used in the analysis outlined in the following section.

Figure 1(c) outlines how the number of stations with a given percentage of completeness varies, as well as the cumulative proportion. The majority of raingauges provide records more than 80 per cent complete. For this work, a cutoff of 90 per cent completeness was chosen as the basis to select a raingauge, leading to the use of ~70 per cent (~3300) of the locations of raingauges that operated continuously during the study period. Based on these locations, the 0.25° × 0.25° grid cells (i.e. the grid corresponding to the TRMM and regridded BoM rainfall products) that had at least one raingauge were identified. These are shown in Fig. 1(d) (~ 2000 cells), and it is these that form the basis of much of the following analysis.

Our reason for limiting the number of raingauge stations is to reduce the influence of the changing numbers of raingauges over time, which would have had some influence on the interpolation schemes used to infer the BoM grids. For the remainder of this paper, the use of the complete Australian grid will be referred to as Case A, while the use of only those grid cells that had at least one raingauge with records at least 90 per cent complete (Fig. 1(d)) will be termed Case B.

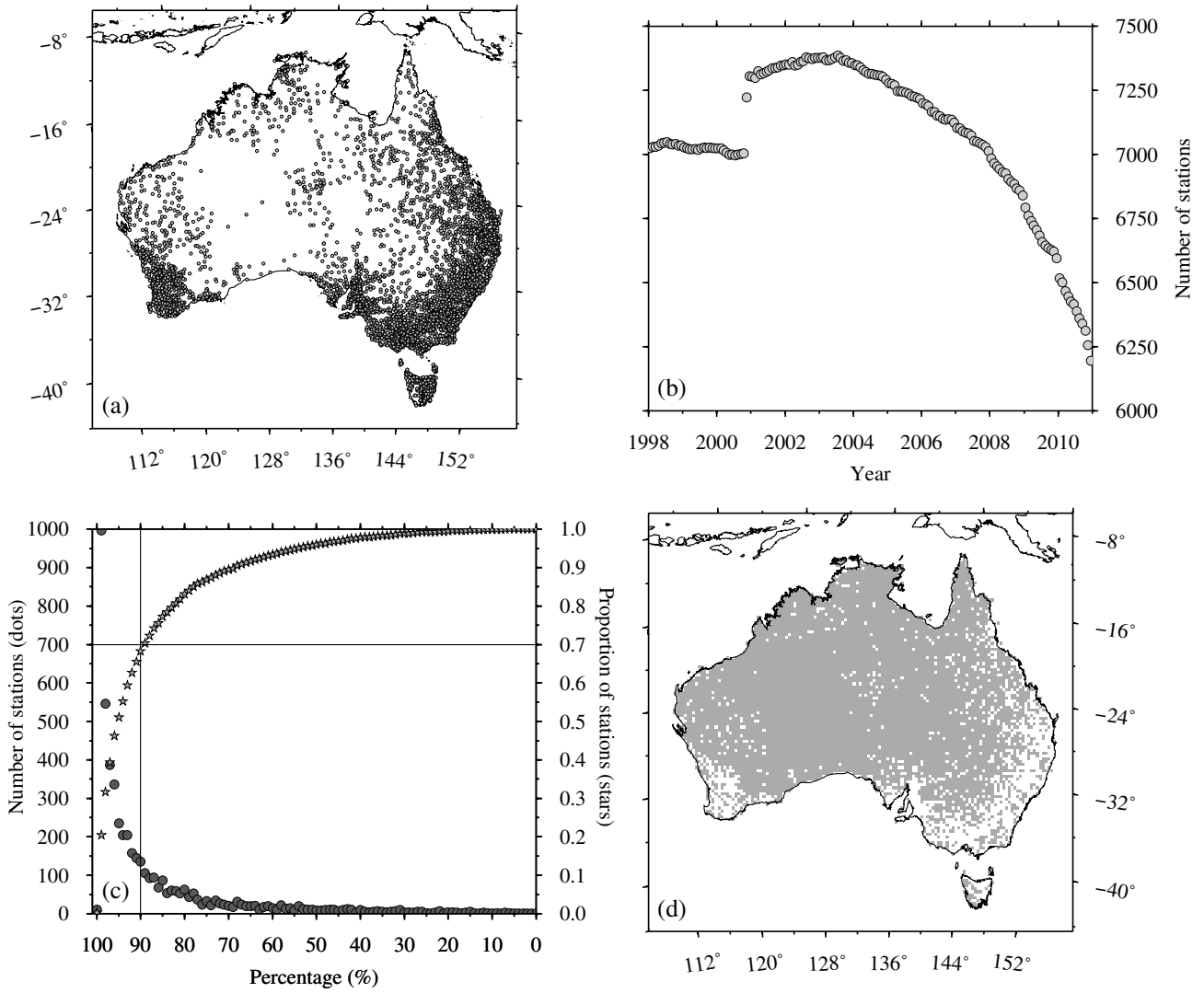
Australia's climate

As part of the analyses (see following section), we will examine how the two datasets compare as a function of Australia's climate, which also reflects the variation of raingauge coverage. BoM makes use of a modified version of the Köppen-Geiger climate classification scheme (Stern et al. 2000), which was based on the assumption that native vegetation is the best indicator of an area's climate. Stern et al. (2000) applied this to Australia by modifying the existing climate subdivisions, and by incorporating BoM mean monthly rainfall, annual rainfall, and mean maximum and minimum temperature gridded datasets. In this work, we use an updated version of this scheme, developed by Peel et al. (2007), who used long-term global records of temperature and precipitation to redraft the Köppen-Geiger world climate map[5]. A simplified version of this climate scheme was used, with each of the climate zones shown in Fig. 2 being the sum of several other subdivisions (for more details, see Peel et al. 2007). This was simply to have the continent divided into similar proportions of its total area. Table 1 lists the number of raingauges that are located in each climate zone, along with the number of associated grid cells and the parameters that describe each zone. The reason for this was to see if there was some climatic dependence in the correlation between the datasets that could be related to the TRMM 3B43 processing.

[2]ftp://precip.gsfc.nasa.gov/pub/trmmdocs/3B42_3B43_doc.pdf
[3]http://www.bom.gov.au/jsp/awap/rain/index.jsp
[4]ftp://ftp.bom.gov.au/anon2/home/ncc/metadata/lists_by_element/alpha/alphaAUS 139.txt

[5]http://people.eng.unimelb.edu.au/mpeel/koppen.html

Fig. 1 (a) The location of the raingauge stations that operated continuously during the period of interest (January 1998 to December 2010). (b) How the number of stations varied over time. (c) Number of stations (dots) as a function of the completeness of their records and a cumulative proportion (stars) over the period of interest. The vertical and horizontal lines mark the 90 per cent completeness and ~0.70 proportion of raingauges, respectively. (d) The areas in Australia where at least one raingauge station operated continuously (white areas) during the study period and were ≥ 90% complete, plotted using the same cell size (0.25° ×0.25°) as used in the analysis. The maps in this and forthcoming figures are in Lambert conformal conic projection.

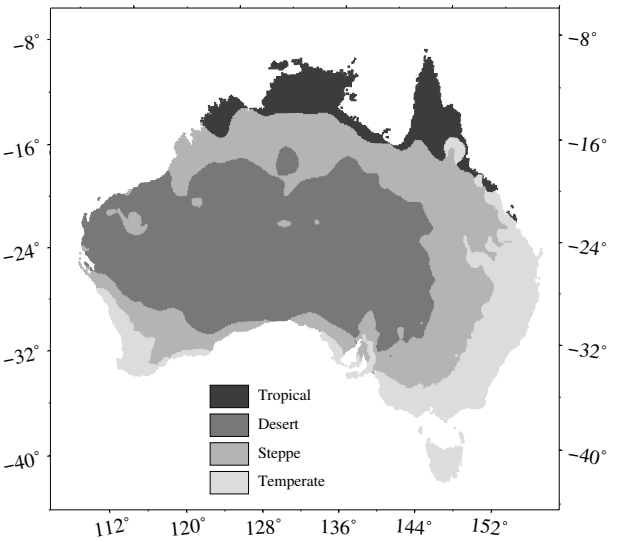


Analysis

The analysis focuses on three situations: (1) comparing TRMM and BoM Australia-wide for Cases A and B (cf. section 2.2); (2) comparing TRMM and BoM with respect to the total rainfall that occurred during the austral seasons (i.e. three month accumulations), divided into December to February for summer (DJF), March to May for autumn (MAM), June to August for winter (JJA), and September to November for spring (SON); and (3) comparing TRMM and BoM with respect to the four broad climatic zones of Australia defined earlier (cf. Fig. 2). The primary means of analysis is to determine the cross-correlation between the datasets, given by:

$$R = \frac{\sum_{i=1}^{n}[(x(i)-\bar{x})(y(i)-\bar{y})]}{\sqrt{\sum_{i=1}^{n}(x(i)-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y(i)-\bar{y})^2}}, \quad \ldots 1$$

Fig. 2 The climate zones of Australia, as described by a simplified version of an updated Köppen-Geiger climate classification scheme (Peel et al. 2007). Some details of how the climate zones are defined are included in Table 1.



where x and y are the TRMM and BoM rainfall products with averages of $\bar{x}$ and $\bar{y}$, respectively, and n is the number of data. R will be determined for two situations: 1) considering the time-series of a given grid point, resulting in location-dependent cross correlation values ($R(x, y)$, where x and y represent longitude and latitude), with n being the number of months (n = 156, i.e., the number of months between January 1998 to December 2010); and 2) a correlation between the grid cells of each dataset for a given month, giving a cross correlation for each month ($R(t)$, where t represents the epoch), with n being the number of grid cells considered (n = 11 359 for case A and n = 2086 for Case B, see Fig. 1(d)). This allows us to assess how the datasets compare over both time and space.

In addition to the cross correlation, principal component analysis (PCA, see e.g. Preisendorfer 1988) is used. This may be described by the following (e.g. Kuhn et al. 2005):

$$H(\Omega, t) = \sum_{j=1}^{p} e_j(\Omega) a^T{}_j(t) = EA^T, \quad \ldots 2$$

where $H(\mathbf{\Omega},t)$ is the scatter matrix (in our case, rainfall minus the long-term average) that varies over time t and space $\mathbf{\Omega}$, $e_j(\mathbf{\Omega})$ represents a set of eigenvectors or the empirical orthogonal function (EOF) that describe the spatial distribution of a mode's signal, $a^T{}_j(t)$ denotes the the principal component (PC) which describes the temporal change in the signal, while p is the total number of modes possible, defined by the number of epochs considered. Hence, a specific mode's signal at a given time is found by multiplying that time's EOF by its associated PC value.

Table 1. The numbers of rainfall-reporting weather stations that operated during the study period (January 1998 to December 2010) with records ≥ 90 per cent complete and the numbers of 0.25°×0.25° grid cells that contain at least one station (Case B, see Figs. 1(d) and 2) for the main climatic zones of Australia, following the scheme of Peel et al. (2007).

Zone	*No. stations*	*No. grid cells*
Tropical	152	90
Desert	261	235
Steppe	862	625
Temperate	2545	1136

Australia-wide analysis

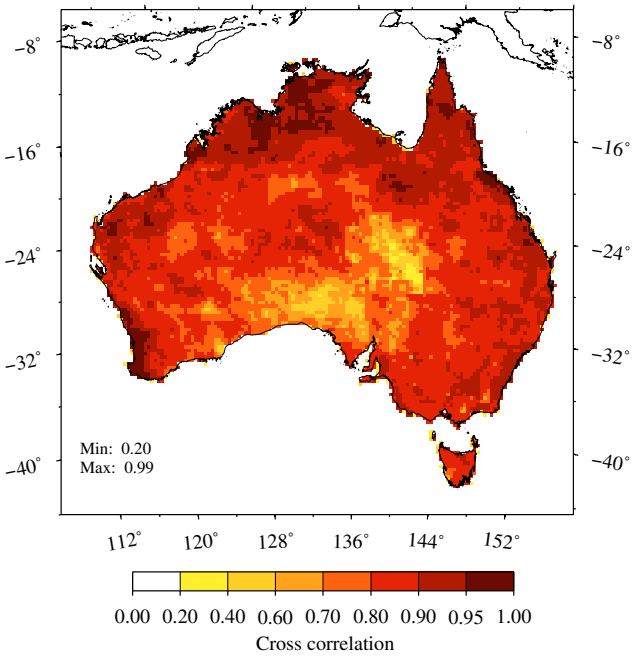
Considering Australia as a whole, Fig. 3 shows the cross correlation between the time-series of the datasets at each grid cell, giving a spatial distribution of R. In general, high correlations are observed, with values > 0.90 found for the better-instrumented areas (cf. Fig. 3 with Fig. 1(a)) that also receive higher rainfall (see Fig. 3.2 in Australian Bureau of Meteorology, 2008), in particular the southwest and along the east coast. High correlations also arise in the north which experiences the highest rainfalls (> 3, 000 mm yr^{-1} in some areas) although being sampled by relatively few raingauges (Fig. 1(a)). However, much lower correlations occur in the sparsely instrumented parts of Australia, falling to values as low as ~0.20. Considering the entire dataset (all months over the entire continent), we obtain a R value of 0.93.

How the correlation between the datasets varies over time is presented in Fig. 4(a) for Case A (i.e. when all grid cells are used). In general, the correlation is > 0.90, although there appears to be a slight decrease with an accompanying increase in the scatter after 2004. However, as commented upon before, not all grid cells actually had a raingauge station that operated over the entire study period. Hence, in Fig. 4(b) this exercise is repeated for Case B (i.e. only those cells that had an operational raingauge whose record is 90 per cent complete, Fig. 1(d)) and for the cells that had no raingauge station during the study period (i.e. their values for BoM are wholely dependent upon the gridding methodology followed).

Considering Case B, we note there is generally an improvement in the cross correlation, with the scatter in the results from 2004 lessened compared to those in Fig. 4(a), but still apparent. Considering for a moment the three months that display the worse correlation values for Case A, namely October 2006, January 2007 and October 2008 (marked by circles in Figs. 4(a) and 4(b)), the results for Case B showed two of them to provide significantly improved correlations

Fig. 3 The variation in the cross correlation (Eqn. 1) between the TRMM and BoM rainfall products over Australia on a 0.25° × 0.25° grid for the time period of interest (January 1998 to December 2010). The values plotted are found over the study period for each cell.

FIGURE 3



compared to Case A (October 2006, from 0.71 to 0.80, and October 2008, from 0.47 to 0.73), while the third showed very little change (January 2007, from 0.78 to 0.79). These months will be dealt with in greater detail in the discussion section. Considering the case of the cells without raingauges, we see that there is in general a noticeable decrease in the R values, although they are still mostly above 0.8.

The results outlined in Figs 4(a) and (b) are shown again in Fig. 4(c), which presents the variation in R averaged for each year with their respective standard deviations of the mean. As could be expected from Figs 4(a) and (b), the standard deviation increases over time with slightly decreasing R, although over the study period, the values generally agree within their standard deviations of each other. Considering for example the first (1998) and last (2010) years for both Case A and B, the R values for these years are 0.93 ± 0.01 and 0.90 ± 0.03 for 1998 and 2010, respectively, for Case A, and 0.95 ± 0.01 and 0.93 ± 0.03 for 1998 and 2010, respectively, for Case B. If the two datasets are considered over their full time-series, effectively combining the cross correlation over space and time, values of $R = 0.93$ for Case A, and $R = 0.94$ for Case B are obtained.

To further examine the correlation between datasets for Australia as a whole, Fig. 5 presents results from the PCA (Eqn. 2). The percentage values given in the EOF maps represent the proportion of the total signal variability described by that mode, hence the first mode for both datasets explains over 60 per cent of their respective total variability. A visual

Fig. 4 The variation over time of the cross correlation (Eqn. 1) between the TRMM and BoM rainfall products over Australia. (a) The results when considering all of Australia (Case A); (b) when considering only the cells with a rain-gauge whose record is ≥ 90% complete (Case B, see Fig. 1(d)) or cells that lacked any rain-gauges during the study period; and (c) the annual averages and standard deviations of the mean for the above cases. The circled months in (a) and (b) are dealt with in greater detail in the 'Discussion'. The values in (c) have a slight temporal offset and the same vertical scale is used for ease of comparison.

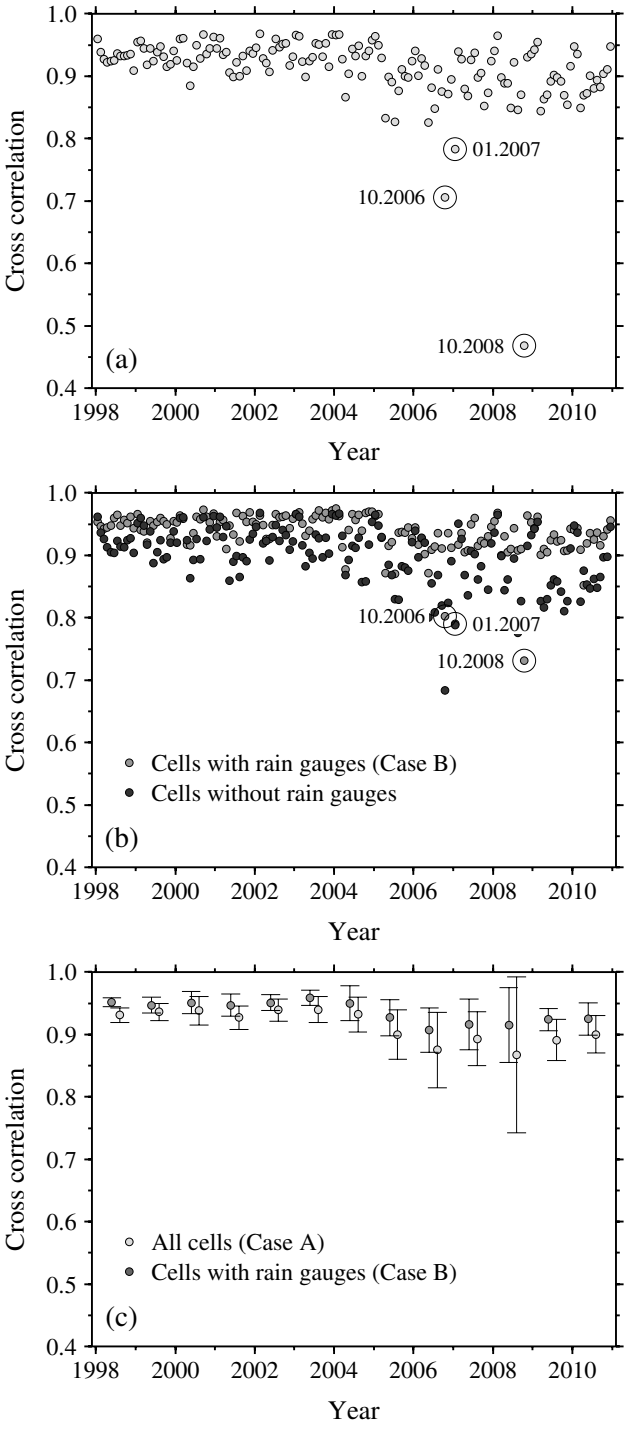
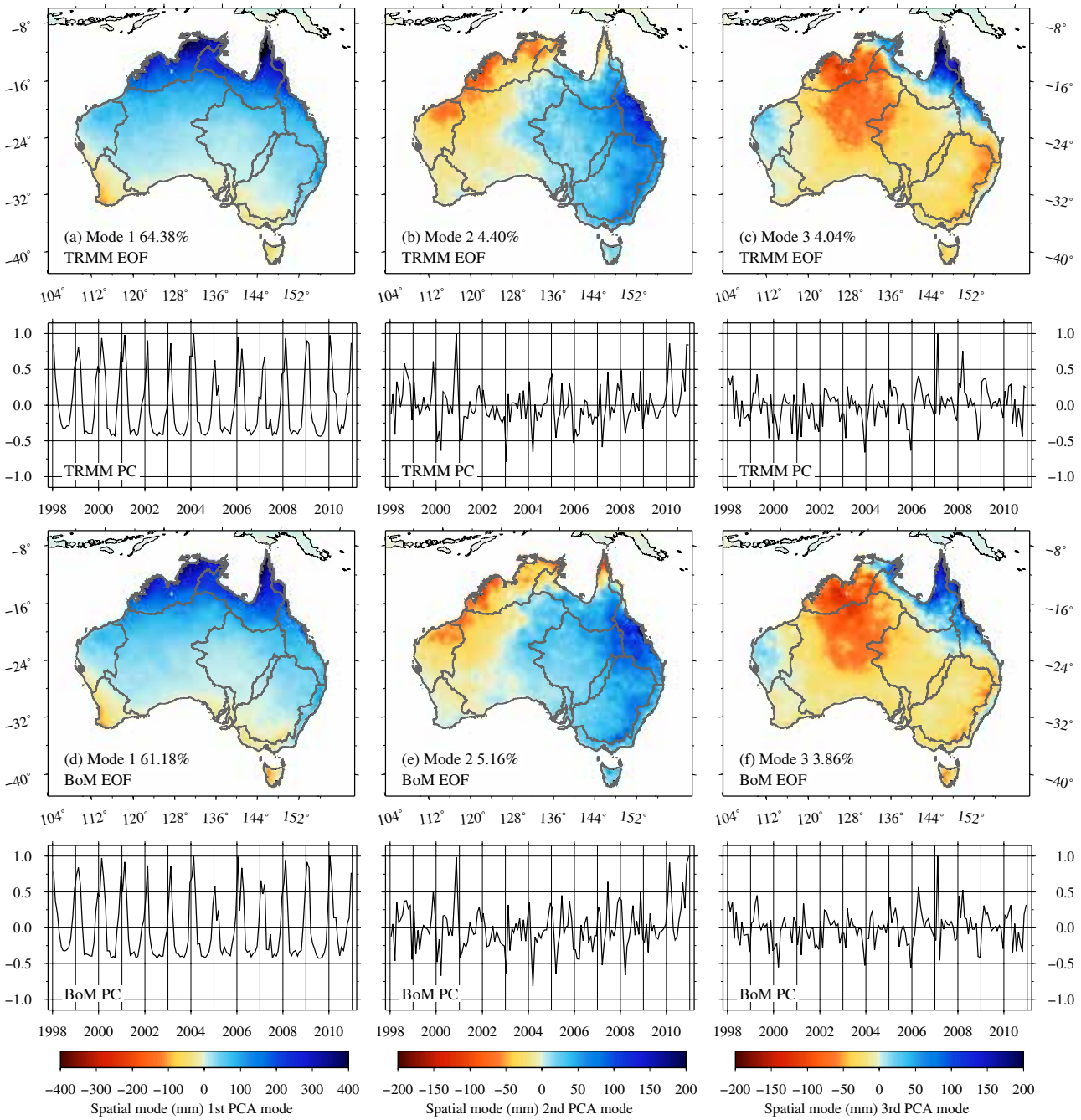


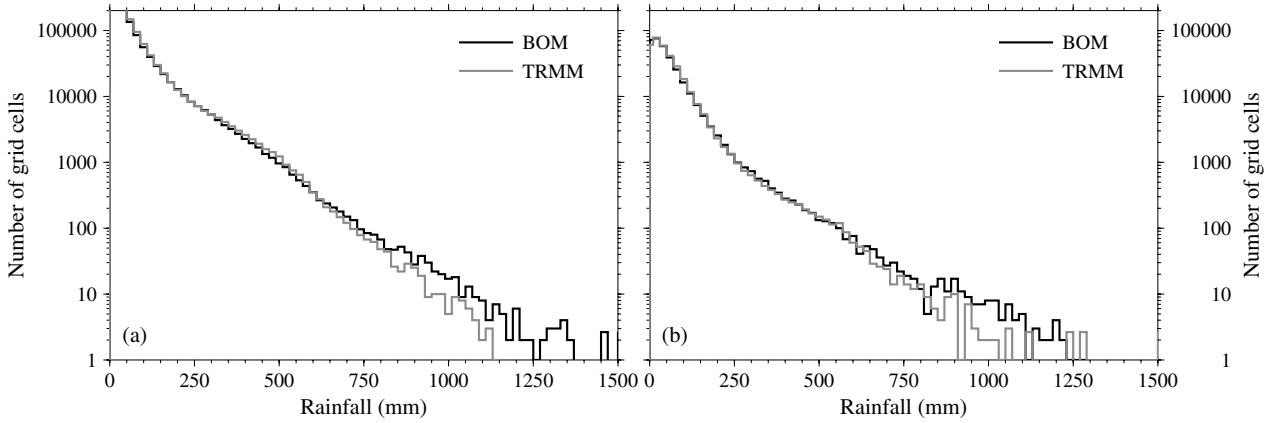
Fig. 5 Results of the principal component analysis of the TRMM (a, b, c) and BoM (d, e, f) datasets. (a, d) mode 1; (b, e) mode 2; and (c, f) mode 3. The percentages in the EOF maps represent the contribution of that mode to the total signal variability.



inspection of the first three modes reveals strong similarities between the datasets in both the EOF and PCA. Determining a cross correlation between the TRMM and BoM of the first three modes of the EOF and PC values shows 0.99, 0.95 and 0.93 for the EOF and 0.99, 0.95 and 0.94 for the PC for for modes 1, 2 and 3, respectively. As PCA provides information about the most dominant temporal/spatial variability (in our case ~70 per cent of the total signal is described by the first three modes), it confirms the high correlation over space and time between these datasets.

As a final comparison for the datasets considering Australia as a whole, Fig. 6 shows histograms of the rainfall distribution for all grid cells over the entire study period for Case A (Fig. 6(a)) and Case B (Fig. 6(b)). These results show that both datasets effectively follow the same probability distribution, with the largest differences in the distribution of the higher values (i.e. in the north of the continent, where there is also sometimes a lack of raingauges). Furthermore, the large number of low values reflect the arid nature of much of Australia.

Fig. 6 Histograms of the monthly rainfall distribution as given by the BOM (black lines) and TRMM (grey lines) datasets. (a) Case A (all grid cells); and (b) Case B (considering only the cells with a raingauge whose record is ≥ 90% complete).



Seasons

The next tests involve examining how the cross correlation between the datasets varies with the seasons (Fig. 7). Considering the full time-series, some variability in the spatial pattern of R over Australia between seasons is observed, with summer showing the lowest mean value (0.81) and autumn the highest (0.86). Figure 8 shows the cross correlation over time for each season determined for Case A (Fig. 8(a)) and Case B (Fig. 8(b)). Considering Case A, the summer R values are consistently higher than for the other seasons, with the exception of 2007, which is largely a function of the poor correlation in January 2007, as shown earlier in Fig. 4. Likewise for the spring of 2006 and 2008, the lower R is again a result of the months that displayed poor cross correlations between datasets (namely October 2006 and October 2008, Fig. 4). Moreover, the autumn, winter and spring appear to display a more obvious decrease in their R values after 2004. Considering the R values over the full time-series for each season, a value of $R = 0.97$ for summer is gained, which is consistant with the plot in Fig. 8(a), and $R = 0.94$ for autumn, winter, and spring.

Examining now Case B, and while summer still generally displays better correlations, it is not as obvious as for Case A. Autumn appears to show the lowest correlation between datasets, however, what must be emphasised is how minor these differences are. This is shown again by considering the R values over the full time-series for each season, where high values of R are gained for each season, namely $R = 0.97$ for summer, $R = 0.94$ for autumn, $R = 0.96$ for winter, and $R = 0.96$ for spring.

Fig. 7 Variation in the cross correlation (Eqn. 1) between the TRMM and BoM rainfall products over Australia for each season for the time period from January 1998 to December 2010. (a) Austral summer (DJF); (b) autumn (MAM); (c) winter (JJA); and (d) spring (SON).

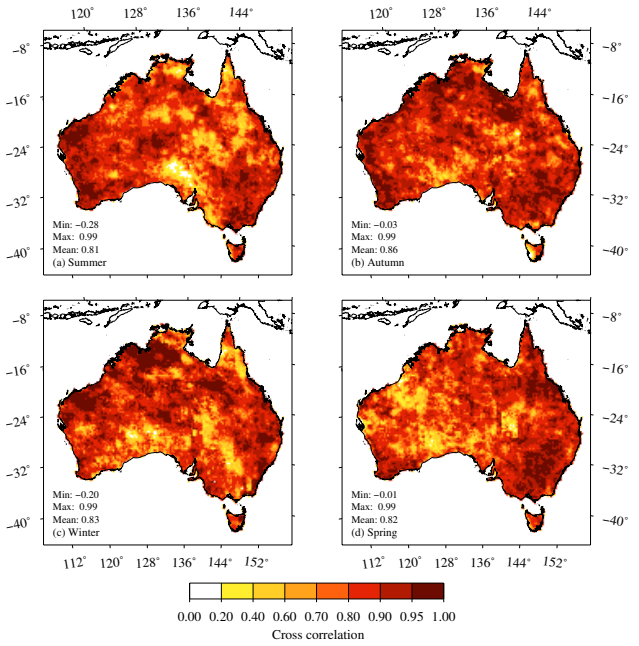
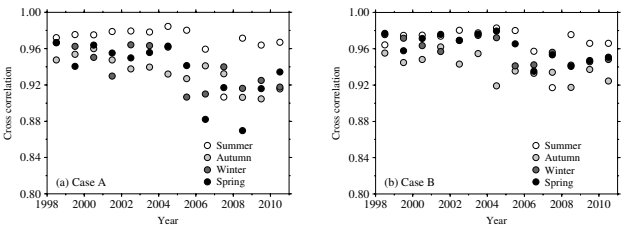


Fig. 8 Variation over time of the cross correlation (Eqn. 1) for each season between the TRMM and BoM rainfall products for (a) Case A and (b) Case B.



Climate

The next study involves the same analysis as above, but after subdividing the cells into their respective climatic zones.

Figure 9 presents the temporal variations in the cross correlation for each of the four climate zones for Case B. The TRMM and BoM datasets appear much better correlated for the steppe and temperate climate zones than for the desert and tropical zones. However, it is still clear that there is generally a high correlation for all zones, with $R > 0.80$

most of the time. This is verified when one examines the correlation over the full time-series, with R = 0.95 for the tropical zone, R = 0.88 for the desert zone, and R = 0.94 for the steppe and temperate zones. Again, the increased scatter in the R values after 2004 is apparent in all cases.

This test is repeated in Fig. 10 to show the temporal variations in the cross correlation for each of the climate zones as a function of season. Winter and spring show the greater scatter in the cross correlations arising for each climate zone, with the temperate zone generally showing the best correlations. The increase in the scatter of the cross-correlation values seen after 2004 in Figs 4 and 8 is also apparent here, especially in spring for the tropical, desert and steppe climates, but slightly less so for the temperate climate (Fig. 10(d)).

Discussion

Figure 11 shows the spatial distribution of rainfall for some representative months: namely February 2002, which provided the highest correlation when considering all grid cells (R = 0.97), and two of the months discussed in the previous section that gave the lowest R values, January 2007 and October 2008. From the general form of the rainfall distribution—as observed by the TRMM and BoM products—a high R value is achieved for February 2002 (Figs 11(a) and (b)). By contrast, for the case of January 2007, very significant differences between the datasets are seen. However, they arise in cells both with and without a raingauge, hence the relatively small change when only considering those cells that have a raingauge. Meanwhile, for October 2008, most of the differences arise in those areas without raingauges, hence the improvement in R when considering only those cells with a raingauge. A similar result is obtained for October 2006 (not shown).

The causes for these differences are not immediately clear. One reason may be that, since TRMM is a major component of the 3B43 product, one would expect it to not record some rainfall owing to unfavorable timing of its overflight of Australia, which may be the case for January 2007 (Fig. 11(c) and (d)), the TRMM 3B43 product used in this work incorporates other satellite data (e.g. geostationary infrared observations). However, in October 2008, there is an area of higher rainfall evident in the TRMM product in central Australia that is not apparent in the BoM product, although it is an area of very few raingauges. This may be evidence of a localised rain event detected by TRMM, yet not able to be accommodated by the sparse-area interpolation procedures followed by BoM, or the precipitation does not reach the ground, and is lost by evaporation before it accumulates in a raingauge.

Another way of displaying how the two data products correlate is shown in Fig. 12, where each dataset's rainfall values for the months shown in Fig. 11 are plotted against each other in terms of their climatic zones. A straight line, $y = Ax$, is fitted to the data, where y represents the BoM values, x the TRMM, with the coefficient A approaching 1 as the two datasets are more highly correlated, and R^2 is the coefficient of determination (i.e., if $R^2 = 1$, then the line perfectly predicts the relationship between data). The resulting straight line is plotted for each month/climate zone case, as is a 1:1 line, along which the points would plot if the datasets were perfectly correlated.

The rainfall values for February 2002 (Figs 11(a) and (b)) are more closely aligned with the 1:1 line than for the other two months (Figs 12(a)–(d)), evident also by the A coefficient and R^2 values being closer to 1 compared to the other months. This is to be expected as this month gave the highest cross-correlation value. For this month, distinct patterns are difficult to recognise, although there seems to be a trend for the TRMM steppe values to be overestimated relative to the BoM values. In the case of January 2007 (Figs 12(e)–(h)), as

Fig. 9 Variation over time of the cross correlation (Eqn. 1) between the TRMM and BoM rainfall products for each climate zone (cf. Fig. 2). (a) Tropical; (b) desert; (c) steppe; and (d) temperate. These values are for Case B (Fig. 1(d)).

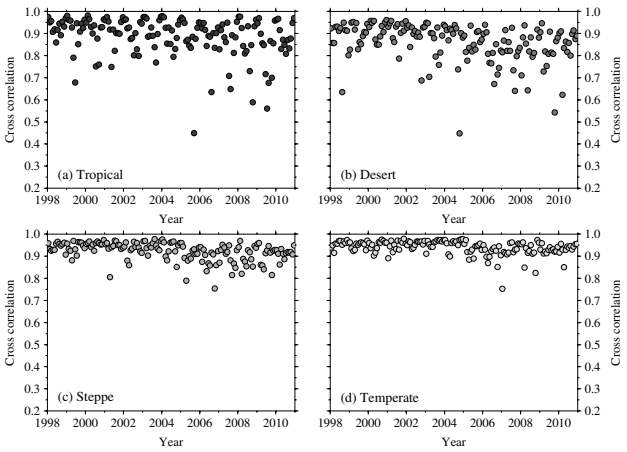


Fig. 10 Variation over time of the cross correlation (Eqn. 1) between the TRMM and BoM rainfall products for each season and climate zone (cf. Fig. 2). (a) Tropical; (b) desert; (c) steppe; and (d) temperate. These values are for Case B (cf. Fig. 1(d)).

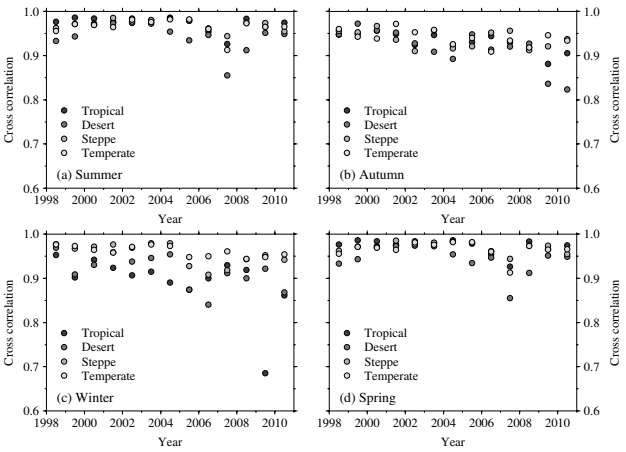
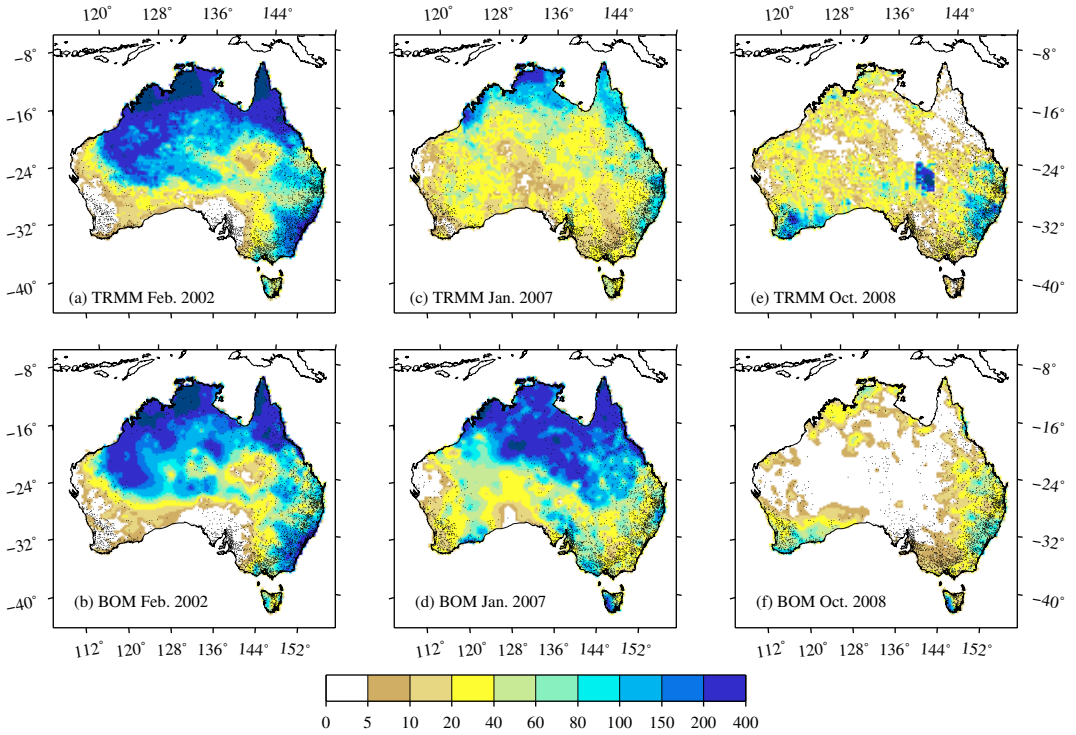


Fig. 11 Comparison of theTRMM (a, c, e) and BoM (b, d, f) rainfall values for the specific months marked out in Fig. 4 (a, b) February 2002; (c, d) January 2007; and (e, f) October 2008.

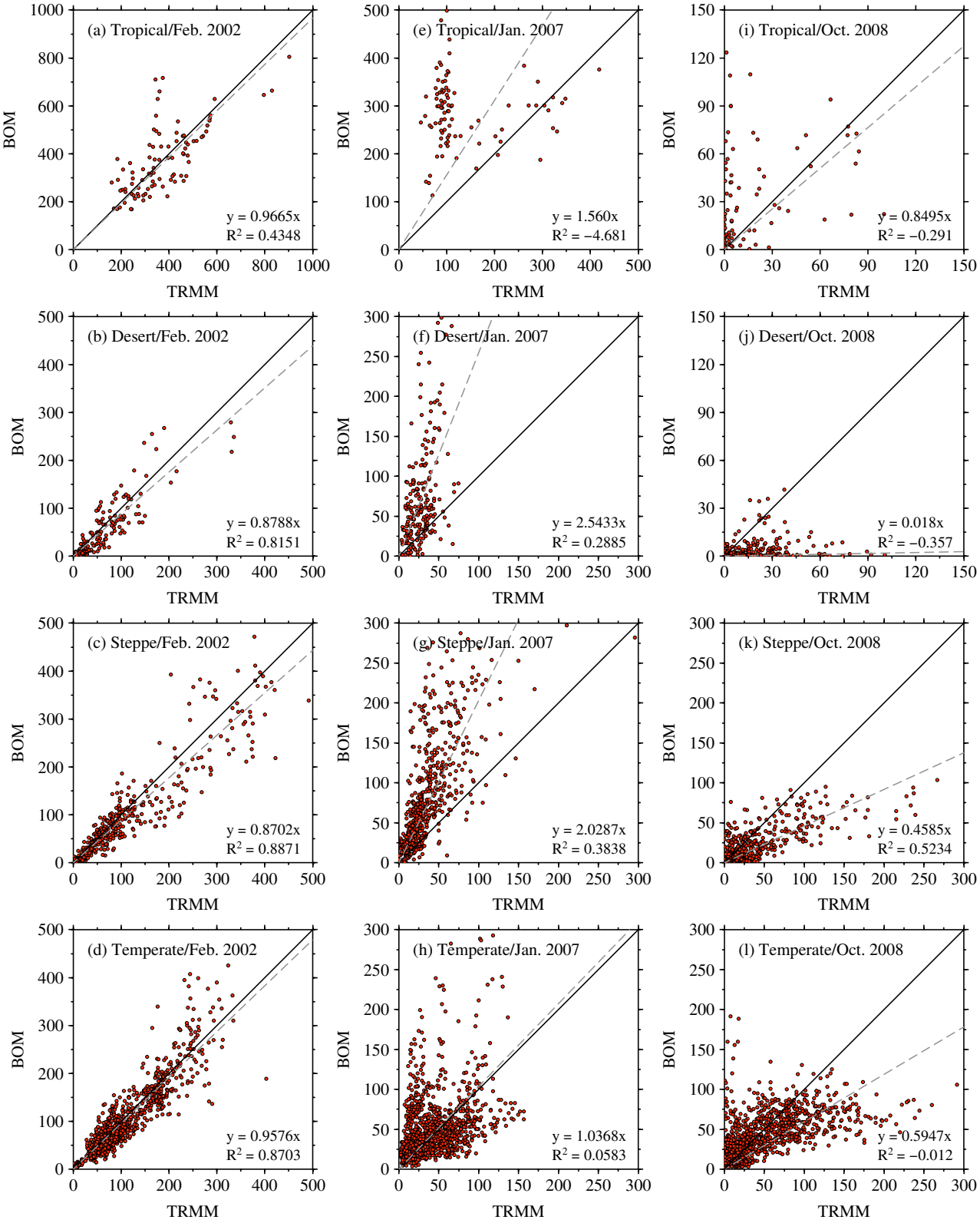


expected from Figs 11(c) and (d), the BoM values are often greater than their TRMM counterparts, while the opposite is the case for October 2008 (Figs. 11(e) and (f) and 12(i)–(l)). For these two months, the poor correlation between the datasets is seen by the A values being quite different from 1 (with the exception of the temperate zone for January 2007) and the very poor R^2 values. For the cases where a negative R^2 value has arisen, this is a result of our forcing a zero-intercept in the linear regression, i.e., excluding a constant term. This leads to the linear regression model being forced to produce a worse data fit than fitting the data to their simple mean value. This situation is clearly the case in Fig. 12(e) leading to a large negative R^2 where, for this situation the R^2 value is largely meaningless.

Examining the full time-series in terms of the total volumes of water associated with rainfall for each of the four climatic zones (Fig. 13), found by simply multiplying a grid cell's rainfall by its area, then summing these volumes, the BoM rainfall volumes (black solid line) are consistently within the range of the TRMM estimates (grey shading), with the exception of January 2007 where the BoM volume exceeds the TRMM maximum for the desert and steppe climates (Figs 13(b) and (c)). To continue this line of discussion, Table 2 presents the average differences in monthly rainfall volumes and the associated standard deviations between the TRMM and BoM products (a positive value indicates that TRMM is larger than BoM) for Australia as a whole and for each of the climate zones for three periods: the full study period (January 1998 to December 2010), and periods divided between January 1998 to December 2003 and January 2004 to December 2010, approximately corresponding to when the scatter in the cross correlations increases (cf. Fig. 4). Also shown in Table 2 are the associated RMS values, found by considering the difference between products for each grid cell volume for all months considered in each time period. The tropical values are much larger than the others, reflecting the much greater rainfall. These values are relatively small when compared to the standard deviations, adding further confidence in the use of either dataset.

In all cases (except for the second period for the tropical zone) the TRMM values are on average larger than the BoM, and the standard deviation exceeds this difference, showing that the biases are insignificant at the 1-sigma level. This is further shown in Fig. 14 which presents the spatial variability in these differences and the associated standard deviations. While the larger standard deviations appear to correspond to those areas that receive higher rainfall, for example the north of the country, a spatial pattern in the differences is not clear. Moreover, an increase in the associated standard deviations is noted when the January 2004 – December 2010 period (Figs 14(d), (e) and (f)) is compared with January 1998 – December 2003 (Figs 14(a), (b) and (c)). Nonetheless, as stated before, the differences between datasets are still less than their associated standard deviations, further providing confidence in the use of either database. Considering again the RMS values, only the tropics appear to show much difference when examining each time period. Hence, we cannot identify any climatic dependency that may influence the TRMM 3B43 product.

Fig. 12 Regression of the TRMM (x-axis) and BoM (y-axis) rainfall values (in mm) for several months of interest discussed in Fig. 11 for each climate zone (cf. Fig. 2). Only the cells that had a raingauge ≥ 90% complete for the study period are considered (Case B). Included are the 1:1 line (black line) and a fitted straight line (grey dashed lines), with the associated linear coefficient and R^2 values.



Conclusions

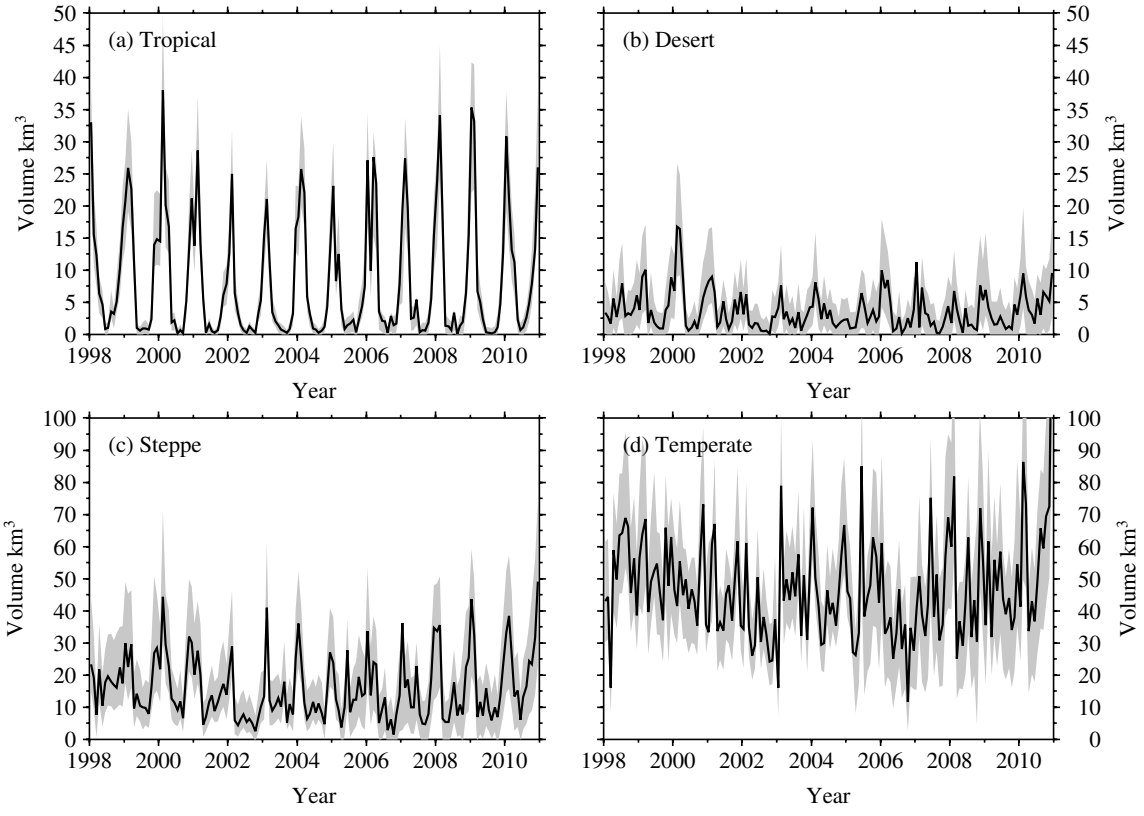
The aim of this work was to compare the *TRMM and Other Sources Rainfall Product* 3B43 with an equivalent, after re-gridding, dataset provided by the Australian Bureau of Meteorology (BoM), which is based solely on raingauge data, for the period between January 1998 to December 2010. In general, the two datasets show a strong agreement with a correlation coefficient of $R = 0.93$ between them when considering the full datasets' time-series over Australia as a whole (Figs (3) and (4)). This is confirmed by results from a PCA analysis (Fig. 5), where correlations of > 0.90 were obtained for the mode 1 EOF and PC results, which contributes over 60 per cent of the total signal, and 70 per cent when considering the first three modes.

Comparing the datasets with respect to the seasons, the austral summer gave the best correlation values over time, followed by the others, all of which displayed very similar correlations with $R > 0.90$ (Figs 7 and 8). It is therefore reasonable to treat the two datasets as being equivalent if one was interested in seasonal rainfall variability over the time scale covered by the TRMM 3B43 product (from January 1998 to the present). When considering the TRMM and BoM datasets with respect to the four climate zones that Australia can be divided into (cf. Figs 2, 9 and 10), much greater scatter was observed for the tropical and desert regions as opposed to the steppe and temperate zones. Although the less-instrumented regions are arid areas with little in the way of population or agricultural activity, accurately understanding

Table 2. The average difference in volumes of rainfall (km³) over Australia per month between theTRMM and BoM rainfall products for the cells that included a rain-gauge ≥ 90 per cent complete (Case B). The values in brackets represents the associated RMS values (considers differences for all gridcells over the time period of interest). A positive value indicates that the TRMM value is larger than the BoM. See also Fig 14.

	January 1998 – December 2010	*January 1998 – December 2003*	*January 2004 – December 2010*
All of Australia	2.43 ± 6.86 (0.11)	4.52 ± 4.69 (0.10)	0.49 ± 8.02 (0.13)
Tropical	0.08 ± 1.33 (0.76)	0.28 ± 0.79 (0.56)	–0.17 ± 1.66 (0.90)
Desert	0.41 ± 0.93 (0.13)	0.50 ± 0.55 (0.12)	0.34 ± 1.19 (0.14)
Steppe	1.21 ± 2.65 (0.12)	2.28 ± 1.59 (0.11)	0.23 ± 3.16 (0.12)
Temperate	0.73 ± 4.00 (0.16)	1.46 ± 3.60 (0.16)	0.09 ± 4.11 (0.15)

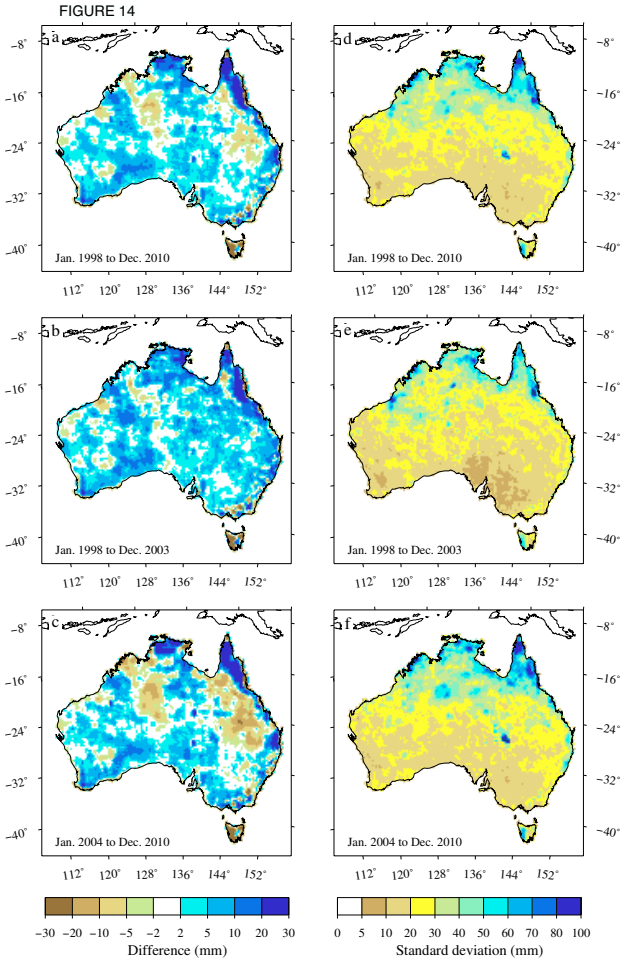
Fig. 13 Comparison of the total volumes of water associated with the TRMM (grey shaded) and BoM (solid black lines) rainfall values for each climate zone (cf. Fig. 2). (a)Tropical; (b) desert; (c) steppe; and (d) temperate. These values are for Case B (cf. Fig. 1(d)).



rainfall in these regions is crucial, given the danger of expanding deserts (e.g. Zeng & Yoon 2009). Nonetheless, in terms of volumes of rainfall water, the BoM dataset is generally within the range defined by the TRMM product. Hence, as commented before for the seasonal analysis, either of the two datasets may be used if one were assessing rainfall behaviour as a function of climatic zone.

It was also found that there were individual months where the correlation between the datasets was much less than for the overall time-series (cf. Figs 4, 8, 9, and 10). Examining the average differences between datasets (Fig. 14), we can divide the time-series, based on the temporal variability in the cross correlation as shown in Figure 4, into two periods, January 1998 – December 2003 and January 2004 – December 2010, where the second period shows slightly lower cross correlation values and greater scatter. There is also an increase in the differences between datasets (Figs 14(a), (b) and (c)) during the second period, although they are still within the standard deviations of the mean (Figs 14(d), (e) and (f)). The reasons for this decrease in the cross correlation after 2004, and for the poorer correlations for individual months, remain to this date unknown.

Fig. 14 Comparison of the differences in the averages over three time periods ((a, d) January 1998 to December 2010; (b, e) January 1998 to December 2003; and (c, f) January 2004 to December 2010) between the TRMM and BoM time-series (positive values indicate TRMM is greater than BoM) (a–c) and the associated standard deviations of the mean (d–f).



Finally, to return to the questions raised in the 'Introduction': (1) is this TRMM product an adequate representation of monthly rainfall in Australia, and (2) would the TRMM product be appropriate for regions of Australia without raingauges? The answer to the first question is yes with some statistical confidence given the high level of correlation between datasets when considering regions with adequate instrumentation. While there are individual months where the two rainfall products are in less agreement, for longer term studies, it is believed that they certainly exhibit adequate correlation, provided care is taken when employing the less correlated months. Such a result is promising, given the potential for future satellite missions that will deal with rainfall measurements, for example the NASA Global Precipitation Measurement mission, which is planned for launch in 2013[6].

As for the second question, the areas where the cross correlation values indicate a poorer relationship between the datasets are those with a much lower density of raingauges, although, when these regions are examined as a whole (cf. Fig. 4(b)), the R values are still generally > 0.80. Nonetheless, based on the results for areas with denser raingauge coverage, there is a strong likelihood that the TRMM 3B43 product would provide accurate rainfall estimates for less densely instrumented areas. To pursue this issue further, it may be necessary to incorporate other datasets, such as those resulting from the Gravity Recovery and Climate Experiment (GRACE) space mission (cf. Leblanc et al. 2009).

Regarding the decrease in the cross correlations from 2004, as stated above, no explanation at the moment is offered. What must be kept in mind is that TRMM is past its original design lifespan, which was to be until March 2004 (Kummerow et al. 2000), hence some degradation of the spacecraft's systems is only to be expected. However, as TRMM is still operational, the question of whether this correlation continues to decrease will be the subject of future investigation.

Acknowledgments

The authors would first like to express their graditude to the two reviewers and the associate editor, whose comments have greatly improved this work. K. Fleming acknowledges the support of this research under the Australian Research Council's Discovery Projects funding scheme (project DP087738). J.L. Awange acknowledges the support of a Curtin Research Fellowship and the Alexander von Humboldt (Ludwig Leichhardt Memorial Fellowship) Foundation that support his time at Curtin University and

[6]http://pmm.nasa.gov/

Karlsruhe Institute of Technology, respectively. M. Kuhn acknowledges the support of a Curtin Research Fellowship. W. Featherstone was the recipient of an Australian Research Council Professorial Fellowship (project number DP0663020).

The authors are further grateful for the datasets used in this work that were obtained from the following sources.

TRMM 3B43 product:

Goddard Earth Sciences Data and Information Services Center (http://disc.sci.gsfc.nasa.gov/precipitation).

BoM monthly gridded data:

Bureau of Meteorology, Climate data online (http://www.BoM.gov.au/jsp/ncc/climate averages/rainfall).

Modified Köppen-Geiger scheme:

Updated Köppen-Geiger climatemap of the world (http://people.eng.unimelb.edu.au/mpeel/koppen.html).

This work is a TIGeR (The Institute of Geoscience Research) publication (no. 386).

References

Australian Bureau of Meteorology, 2008. *Climate of Australia*. Commonwealth of Australia.

Awange, J., Sharifi M., Baur, O., Keller, W., Featherstone, W. and Kuhn, M. 2009. GRACE hydrological monitoring of Australia: current limitations and future prospects. *Journal of Spatial Science, 54(1),* 23–36.

Awange, J., Fleming, K., Kuhn, M., Featherstone, W., Anjasmara, I. and Heck, B. 2011. On the suitability of the 4° × 4° GRACE mascon solutions for remote sensing Australian hydrology. *Remote Sensing of Environment, 115(3),* 864–875, DOI: 10.1016/j.rse.2010.11.014.

Debo Adeyewa, Z. and Nakamura, K. 2003. Validation of TRMM radar rainfall data over major climatic regions in major climatic regions in Africa. *Journal of Applied Meteorology, 42(2),* 331–347, DOI:10.1175/1520–0450(2003)042<0331:VOTRRD>2.0.CO;2.

Ebert, E., Janowiak, J. and Kidd, C. 2007. Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bulletin of the American Meteorological Society, 88(1),* 47–64, DOI:10.1175/BAMS–88–1–47.

Huang, G.-J. and Bringi, V. 2009. A new approach for analyzing and comparing TRMM PR and ground CP2 radar data for two convective storms near Brisbane, Australia, in *Proceedings of the 34th Conference on Radar Meteorology,* American Meteorological Society, Williamsburg, Virginia, USA.

Huffman, G., Adler, R., Rudolf, B., Schneider, U. and Keehn, P. 1995. Global precipitation estimates based on a technique for combining satellite-based estimates, rain gauge analysis, and NWP model precipitation information. *Journal of Climate, 8(5),* 1284–1295, DOI:10.1175/1520–0442(1995)008<1284:GPEBOA>2.0.CO;2.

Huffman, G., Adler, R., Bolvin, D., Gu, G., Nelkin, E., Bowman, K., Hong, Y., Stocker, E. and Wolff, D. 2007. The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology, 8(1),* 38–55, DOI:10.1175/JHM560.1.

Jones, D., Wang, W. and Fawcett, R. 2009. High-quality spatial climate data-sets for Australia. *Australian Meteorological and Oceanographic Journal, 58(4),* 233–248.

Kuhn, M., Bosch, W. and Kaniuth, R. 2005. Low-frequency variations of the North Atlantic sea level measured by TOPEX/Poseidon altimetry. *Marine Geodesy, 28(1),* 19–37, DOI:10.1080/01490410590884449.

Kummerow, C., Barnes, W., Kozu, T., Shiue, J. and Simpson, J. 1998. The Tropical Rainfall Measuring Mission (TRMM) sensor package. *Journal of Atmospheric and Oceanic Technology, 15(3),* 809–817, DOI:10.1175/1520–0426(1998)015<0809:TTRMMT>2.0.CO;2.

Kummerow, C., Simpson, J., Thiele, O., Barnes, W., Chang, A., Stocker, E., Adler, R., Hou, A., Kakar, R., Wntz, F., Aschroft, P., Kozu, T., Hing, Y., Okamoto, K., Iguchi, T., Kuroiwa, H., Im, E., Haddad, Z., Huffman, G., Ferrier, B., Olson, W., Zipser, E., Smith, E., Wilheit, T., North, G., Krishnamurti, T. and Nakamura, K. 2000. The status of the Tropical Rainfall Measuring Mission (TRMM) after two years in orbit. *Journal of Applied Meteorology, 39(12),* 1965–1982, DOI:10.1175/1520–0450(2001)040<1965:TSOTTR>2.0.CO;2.

Leblanc, M., Tregoning, P., Ramillien, G., Tweed, S. and Fakes, A. 2009. Basin-scale, integrated observations of the early 21st century multiyear drought in southeast Australia. *Water Resources Research, 45(W04408),* DOI:10.1029/2008WR007333.

Liao, L. and Meneghini, R. 2009. Validation of TRMM precipitation radar through comparison of its multiyear measurements with ground-based radar. *Journal of Applied Meteorology and Climatology, 48(4),* 804–817, DOI:10.1175/2008JAMC1974.1.

Liu, Y., de Jeu, R., van Dijk, A. and Owe, M. 2007. TRM-TMI satellite observed soil moisture and vegetation density (1998–2005) show strong connection in El Niño in eastern Australia, *Geophysical Research Letters, 34(L15401),* DOI:10.1029/2007GL030311.

Meinke, H., DeVoil, P., Hammer, G., Power, S., Allan, R., Stone, R., Folland, C. and Potgieter, A. 2005. Rainfall variability at decadal and longer time scales: Signal or noise? *Journal of Climate, 18(1),* 89–96, DOI:10.1175/JCLI–3263.1.

Nicholson, S., Some, B., McCollum, J., Nelkin, E., Klotter, D., Berte, Y., Diallo, B., Gaye, I., Kpabeba, G., Ndiaye, O., Noukpozounkou, J., Tanu, A., Thiam, A., Toure, A.A. and Traore, A. 2003. Validation of TRMM and other rainfall estimates with a high-density gauge dataset for West Africa: Part II: Validation of TRMM rainfall products, *Journal of Applied Meteorology, 42(10),* 1355–1368, DOI:10.1175/1520–0450(2003)042<1355:VOTAOR>2.0.CO;2.

Oke, A., Frost, A. and Beesley, C. 2009. The use of TRMM satellite data as a predictor in the spatial interpolation of daily precipitation over Australia, in *Proceedings of the 18th World IMACS/MODSIM Congress,* edited by R. Anderssen, R. Braddock, & L. Newham, Cairns, Australia.

Peel, M., Finlayson, B. and McMahon, T. 2007. Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Discussions, 4,* 439–473, DOI:10.5194/hessd–4– 439–2007.

Preisendorfer, R. 1988. *Principal component analysis in meteorology and oceanography*. Elsevier, Amsterdam, New York.

Rieser, D., Kuhn, M., Pail, R., Anjasmara, I. and Awange, J. 2010. Relation between GRACE derived surface mass variations and precipitation over Australia. *Australian Journal of Earth Sciences, 57(7),* 887–900, DOI: 10.1080/08120099.2010.512645.

Schumacher, C. and Houze, R. J. 2000. Comparison of radar data from the TRMM satellite and Kwajalein oceanic validation site. *Journal of Applied Meteorology, 39(12),* 2151–2164,DOI:10.1175/1520–0450(2001)040<2151:CORDFT>2.0.CO;2.

Shige, S., Sasaki, H., Okamoto, K. and Iguchi, T. 2006. Validation of rainfall estimates from the TRMM precipitation radar and microwave imager using a radiative transfer model: 1. Comparison of the version-5 and -6 products. *Geophysical Research Letters, 33(L13803),* DOI:10.1029/2006GL026350.

Smith, W. and Wessel, P. 1990. Gridding with continuous curvature splines in tension. *Geophysics, 55(3),* 293–305, DOI: 10.1190/1.1442837.

Solomon, S., Qin, D., Manning, M., Alley, R., Berntsen, T., Bindoff, N., Chen, Z., Chidthaisong, A., Gregory, J., Hegerl, G., Heimann, M., Hewitson, B., Hoskins, B., Joos, F., Jouzel, J., Kattsov, V., Lohmann, U., Matsuno, T., Molina, M., Nicholls, N., Overpeck, J., Raga, G., Ramaswamy, V., Ren, J., Rusticucci, M., Somerville, R., Socker, T., Whetton, P., Wood, R. and Wratt, D. 2007. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Chap. Technical Summary, Cambridge University Press, Cambridge.

Stern, H., de Hoedt, G. and Ernst, J. 2000. Objective classificatio of Australian climates. *Australian Meteorological Magazine, 49(2),* 87–96.

Su, F., Hong, Y. and Lettenmaier, D. 2008. Evaluation of TRMM multisatellite precipitation analysis (TMPA) and its utility in hydrological prediction in the La Plata Basin. *Journal of Hydrometeorology, 9(4),* 622–640, DOI:10.1175/2007JHM944.1.

Wessel, P. and Smith, W. 1998. New, improved version of Generic Mapping Tools released. *EOS Transactions of the American Geophysical Union, 79(47),* 579, DOI:10.1029/98EO00426.

Zeng, N. and Yoon, J. 2009. Expansion of the world's deserts due to vegetation-albedo feedback under global warming. *Geophysical Research Letters, 36(L17401),* DOI:10.1029/2009GL039699.