Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-607
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 19 February 2019

Hydrology and Earth System Sciences
Discussions

An evaluation of daily precipitation from atmospheric reanalyses over Australia

Suwash Chandra Acharya[1], Rory Nathan[1], Quan J Wang[1], Chun-Hsu Su[2], Nathan Eizenberg[2]

[1] Department of Infrastructure Engineering, The University of Melbourne, Melbourne, Australia
[2] Bureau of Meteorology, Melbourne, Australia

Correspondence to: S.C. Acharya (suwasha@student.unimelb.edu.au)

Abstract.

An accurate representation of spatio-temporal characteristics of precipitation fields is fundamental for many hydro-meteorological analyses but is often limited by the paucity of gauges. Reanalysis models provide systematic methods of representing atmospheric processes to produce datasets of spatio-temporal precipitation estimates. The precipitation from the reanalysis datasets should, however, be evaluated thoroughly before use because it is inferred from physical parameterization. In this paper, we evaluated the precipitation dataset from the Bureau of Meteorology Atmospheric high-resolution Regional Reanalysis for Australia (BARRA) and compared it against (a) gauged point observations, (b) an interpolated gridded dataset based on gauged point observations (AWAP), and (c) a global reanalysis dataset (ERA-Interim). We utilized a range of evaluation metrics such as continuous metrics (correlation, bias, variability, modified Kling-Gupta efficiency), categorical metrics, and other statistics (wet day frequency, transition probabilities and quantiles) to ascertain the quality of the dataset. BARRA, in comparison with ERA-Interim, shows a better representation of rainfall of larger magnitude at both point and grid scale of 5 km. BARRA also consistently reproduces the distribution of wet days and transition probabilities. The performance of BARRA varies spatially, with better performance in the temperate zone than in the arid and tropical zones. A point-to-grid evaluation based on correlation, bias and modified Kling-Gupta efficiency (KGE') indicates that ERA-Interim performs on par or better than BARRA. However, on a spatial scale, BARRA outperforms AWAP in terms of KGE' score and the components of the KGE' score. Our evaluation illustrates that BARRA, with richer spatial variations in climatology of daily precipitation, provides an improved representation of precipitation compared with the coarser ERA-Interim. It is a useful complement to existing precipitation datasets for Australia, especially in sparsely gauged regions.

1 Introduction

Availability of accurate precipitation datasets is an essential requirement for the modelling of natural processes, hydro-meteorological analyses and forecasting, monitoring climatic variations and changes (Kirschbaum et al., 2017; Kucera et al.,

2013; Robertson et al., 2013). A comprehensive knowledge of occurrence and distribution of precipitation is however hindered by the sparseness of the gauging network. Variations in the density and coverage of the gauging network make it difficult to capture information on the spatial and temporal variability of rainfall. This is particularly the case in areas covered by deserts, mountains, and oceans and in large areas with low population densities (Salio et al., 2015; Thiemig et al., 2012). This presents a challenge for the Australian continent where the gauges are mostly located along the densely populated coastal regions. The station network is less dense in the central region which represents the more arid part of the continent (Johnson et al., 2016).

The difficulties inherent in existing observation networks have prompted the development of various gridded datasets with consistent spatial and temporal scale. One such precipitation dataset for Australia is the interpolated precipitation product from the Australian Water Availability Project (AWAP). Because of the uneven gauge network distribution, it is not consistent in terms of accuracy (Jones et al., 2009). Global and regional reanalysis datasets provide another source of precipitation data at a consistent spatial and temporal resolution. Such reanalysis datasets are generated using a numerical weather prediction (NWP) model and a data assimilation scheme to incorporate the available observations thereby providing a consistent method of representation of the atmosphere at a regular interval over larger spatial and temporal domain (Parker, 2016). The available global reanalysis datasets such as NCEP-CFSR (Saha et al., 2010), ERA-Interim (Dee et al., 2011), JRA-55 (Kobayashi et al., 2015) cover the Australian region, but their horizontal resolutions are relatively coarse (≥ 80 km) and unsuitable for fine-scale application in hydro-meteorological analysis. The resolution of the global reanalysis can be enhanced by downscaling approaches such as dynamic downscaling (Soares et al., 2012) or statistical analysis using high-resolution surface observations (Vidal et al., 2010). Alternatively, the application of a regional model-based data assimilation can provide a better representation of local climate features and extreme events (Bollmeyer et al., 2015). The regional reanalysis, unlike global reanalysis, allows the integration of abundant local surface observations at a finer scale (Bollmeyer et al., 2015; Isotta et al., 2014; Jakob et al., 2017).

The Bureau of Meteorology Atmospheric high-resolution Regional Reanalysis for Australia (BARRA) is the first 12 km regional reanalysis conducted over Australia, New Zealand and southeast Asia (Su et al., 2018). It provides a greater understanding of the past weather, particularly extreme events. It will, therefore, support better planning and management to reduce risks for the future. BARRA makes use of local surface observations and locally derived wind vectors which are not available to global reanalysis models. However, precipitation observations are not assimilated in BARRA. Precipitation is modelled using a microphysics parameterization and a mass flux convection scheme. The modelled precipitation can be erroneous (Bukovsky and Karoly, 2007; Parker, 2016). Therefore, it is necessary to understand the nature of the uncertainty and inaccuracies involved and to use this understanding to correct for any systematic errors.

The criteria for the evaluation of the reanalysis precipitation need to be rigorous and relevant to the intended purpose of use (Parker, 2016). The type of reference datasets (point or grid), the spatial and temporal scale, and the metrics used vary widely

among the studies. In general, gridded precipitation products including reanalysis datasets, satellite precipitation estimates, and/or interpolated datasets are evaluated against benchmark datasets (Baez-Villanueva et al., 2018; Beck et al., 2017; Gebremichael, 2010; Isotta et al., 2014; de Leeuw et al., 2015; Peña-Arancibia et al., 2013; Zambrano-Bigiarini et al., 2017). The suitable benchmark datasets are comprised of point measurements (Baez-Villanueva et al., 2018; Chiaravalloti et al., 2018; Salio et al., 2015; Thiemig et al., 2012; Zambrano-Bigiarini et al., 2017) and/or high-quality gridded datasets (Chiaravalloti et al., 2018; Isotta et al., 2014; Peña-Arancibia et al., 2013). Catchment-scale representativeness of precipitation datasets is assessed by evaluating catchment average precipitation (e.g. Thiemig et al., 2012). Evaluation studies for hydrological applications focus on a daily timescale analysis (Baez-Villanueva et al., 2018; Thiemig et al., 2012; Zambrano-Bigiarini et al., 2017), though this has been extended to a sub-daily time interval for a purpose of hydrologic risk assessment (Chiaravalloti et al., 2018).

Given the complex statistical behaviour of precipitation, a range of evaluation metrics are available which have their own assumptions and limitations. Typical metrics are unconditional scores (e.g. correlation, bias, root mean square error, mean absolute error), categorical metrics (e.g. probability of detection, false alarm ratio, skill scores) and distributional statistics (Gebremichael, 2010). In addition, statistical properties like quantiles, wet day frequency, and transition probabilities help to ascertain if the dataset under evaluation preserves the statistical properties related to sequencing. Any single metric cannot adequately represent the nature of all errors in the precipitation products. Therefore, it is essential to evaluate multiple metrics which describes different aspects of precipitation to identify the possible sources of mismatch between datasets (Baez-Villanueva et al., 2018).

This study is one of the first comprehensive explorations of the BARRA precipitation estimates. It aims to identify its strengths and limitations relative to other available precipitation datasets (specifically, to AWAP and ERA-Interim, as described below). The precipitation estimates in the BARRA reanalysis are available at hourly timescale. While we are interested in the accuracy of the sub-daily estimates as well, this study focuses only on a daily scale. The accuracy at a daily scale provides us with an important benchmark as it is applicable to many hydrological applications and also forms the basis for further examination at finer timescales. We explore a range of metrics to evaluate the depth and statistical distribution of precipitation at a daily level. In addition, we examine the wet day frequency and transition probabilities to assess the suitability of the datasets for climate studies. The evaluation is carried out on both point and areal precipitation estimates. This evaluation sheds light into the potential hydro-meteorological applications of the dataset such as climate change studies, water resource management, floods, and droughts.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-607
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 19 February 2019

2 Datasets

The sub-hourly time series of recorded rainfall from continuous rainfall stations are obtained from the Bureau of Meteorology. Daily rainfall is generated by aggregating the sub-daily rainfall observations. The period of analysis was determined by the duration at which all the datasets used in the study are available i.e. January 2010 to December 2015. The gauge stations used in the study are chosen based on the availability of information for the entire period over which BARRA data estimates are available. A total of 441 stations are selected, and their spatial distribution is shown in Figure 1(a).

The Australian Water Availability Project (AWAP) provides a daily high-quality 0.05 ° × 0.05 ° (around 5 km) gridded rainfall product, dating back to 1900, based on an extensive network of rain gauges, as described by (Jones et al., 2009). The AWAP product makes use of daily (9 am to 9 am) data from all available gauges across the whole of Australia. The gridded estimates are obtained using a weighting scheme that incorporates an optimized Barnes successive-correction algorithm and orographic corrections (Jones et al., 2009). AWAP uses land-based observations only and therefore does not provide information over the ocean.

The global reanalysis ERA-Interim of the European Centre for Medium-Range Weather Forecasts (ECMWF, www.ecmwf.int) covers the period from 1 January 1979 to present (Dee et al., 2011). The core component of the ERA-Interim data assimilation system is the 12h 4D-variational (4DVar) analysis scheme of the upper-air atmospheric state, which is on a spectral grid with triangular truncation of 255 waves (corresponding to approximately 80 km) and a hybrid vertical co-ordinate system with 60 vertical levels. The precipitation is estimated by the numerical model based on temperature and humidity information derived from assimilated observations originating from Passive Microwave and in-situ measurements. In this study, the sub-daily precipitation from ERA-Interim is converted to daily by accumulation to 24 hours to be consistent with the gauged and AWAP data.

BARRA also uses 4DVar but at the 36 km resolution, and uses the Unified Model as its forecast model (Su et al., 2018). BARRA extends spatially over 65.0° to 196.9° east, -65.0° to 19.4° north at a spatial resolution of 0.11° (approximately 12 km) and with 70 levels up to 80 km into the atmosphere. The project aims to produce the dataset dating back to 1990, however, at present, only a six-year period (2010-2015) is available. The analysis for BARRA is conducted 4 times a day with a 6-hour analysis window centred at time t_0 = 0, 6, 12 and 18 UTC. The forecast cycle is of 12 hours which ranges from t_0-3h to t_0+9h (Su et al., 2018). The BARRA precipitation forecasts from t_0+4 hour to t_0+9 hour are used to generate a daily time series for the study.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-607
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 19 February 2019

3 Methodology

3.1 Methodology for data comparison

The study is conducted over a period of 6 years, from January 2010 to December 2015. We have performed a point-to-grid and grid-to-grid evaluation of the reanalysis datasets against suitable benchmark datasets. The gauged data represents the best estimate of rainfall at a point, and AWAP data provides the best estimate of rainfall over a grid cell. Both estimates are, however, limited by the available gauging density. AWAP is based on all available gauges. The point data considered here are from selected sub-daily rainfall stations with maximum availability of data over the study period. The gauges are distributed all over Australia and are spatially representative of the available gauging network. As gauges measure point rainfall and gridded datasets represent an areal average, it may be expected that interpolated areal rainfalls differ from point estimates. The AWAP estimates will contain inaccuracies due to the interpolation method, but the point estimates are also an imperfect estimate of rainfall over a grid cell area. Consequently, we compared the BARRA and ERA-Interim estimates to both point gauged and AWAP areal data.

In order to compare gridded rainfalls with point rainfalls, a nearest neighbour method is employed. At each location of point rainfall, the nearest neighbour method assigns the precipitation amount from the nearest grid point. In grid-to-grid evaluation, both reanalysis datasets are interpolated to a common AWAP spatial scale using nearest neighbour method. That means, for each AWAP grid, the precipitation variable is obtained from the nearest grid of reanalysis datasets. Given the spatial inconsistency in the quality of AWAP dataset, the grid-to-grid evaluation is limited to grid points that are nearest to the gauge stations used in this study.

The performance of daily precipitation from the reanalysis data is assessed considering all days of the year and by seasons (Summer: DJF, Autumn: MAM, Winter: JJA, and Spring: SON). In addition, evaluation is stratified across three broad climatic zones (arid, tropical and temperate) as defined by the Köppen-Geiger classification (Peel et al., 2007). Most comparisons are undertaken using estimates from every day in the 6-year period, and thus the number of wet days varies with the different datasets and locations considered. However, it is recognized that reanalysis products tend to produce a high number of days with light drizzle, and therefore they over-estimate the frequency of wet days. Thus, following Baez-Villanueva et al. (2018), Ebert et al. (2007) and Zambrano-Bigiarini et al. (2017), a threshold of 1 mm/day is used to classify a day as ‘wet’ or ‘dry’.

3.2 Performance indices

The evaluation of precipitation data generally involves an assessment of detection capabilities and biases in the form of continuous and categorical metrics. The continuous metrics used in this study are Modified Kling-Gupta efficiency (KGE’) (Gupta et al., 2009; Kling et al., 2012) along with its three individual components: correlation (r), bias ratio (β), and variability ratio (γ). Baez-Villanueva et al. (2018) and Zambrano-Bigiarini et al. (2017) also suggested the use of KGE’ and its

components for evaluating precipitation datasets as it provides an overall assessment along with an error in the representation of magnitude and variability of the reference precipitation.

The four categorical indices adopted are the probability of detection (POD) or hit rate, false alarm ratio (FAR), critical success (or threat) index (CSI), and frequency bias (fBias). These are evaluated over five different rainfall intensity classes, namely: no rain (<1 mm), light rain (≥1 mm and <5 mm), moderate rain (≥5 mm and <20 mm), heavy rain (≥20 mm and <40 mm) and violent rain (≥40 mm) (Baez-Villanueva et al., 2018; Zambrano-Bigiarini et al., 2017). The continuous and categorical metrics used in the study are described in Table 1.

3.3 Precipitation statistics

In this study, we evaluate the capacity of the reanalysis datasets to reproduce a range of precipitation statistics related to the frequency of wet days, transition probabilities between wet and dry days, and the distribution of rainfall amounts based on the 90%, 95%, and 99% daily exceedance values.

The transition probabilities considered here are p_{01} and p_{11}, which denote the probabilities of a dry day followed by a wet day, and a wet day followed by a wet day, respectively. Consider X_0 and X_1 are two consecutive days, and i, j are two states (0: dry, 1: wet), then the transition probability is expressed as

$$p_{ij} = Pr(X_1 = j | X_0 = i)$$

For dry-wet day: $p_{01} = Pr(X_1 = 1 | X_0 = 0)$

For wet-wet day: $p_{11} = Pr(X_1 = 1 | X_0 = 1)$

4 Results

4.1 Mean precipitation

Figure 1 shows the average annual precipitation over Australia for the period of six years (2010-2015) as estimated using the different data sets (gauged point rainfalls, AWAP, BARRA, and ERA-Interim). It is seen that there is a high spatial variability in the average precipitation over Australia. Regions of high rainfall (northern and eastern coasts along with western Tasmania) are similarly represented by all four datasets. ERA-Interim provides a coarser representation of the precipitation field, which expectedly fails to capture the higher spatial variability in the coastal regions and orographic precipitations in the Great Dividing ranges and western Tasmania. By comparison, BARRA precipitation captures this variability in the AWAP precipitation. It should be noted, however, that the AWAP data provides a poor estimate of precipitation over central Australia

where there is a paucity of gauging information. BARRA, on the other hand, provides high-resolution precipitation pattern over the ocean as well as the central Australian region where gauges are sparse.

4.2 Point-to-grid assessment

KGE' and its components

The variation of KGE' at the gauge locations is shown in Figure 2. As expected, the higher performance of KGE' is observed in the AWAP dataset as this is derived by interpolating gauged data, which includes the selected high-quality point rainfall gauges. The pattern of KGE' score looks similar among both reanalysis datasets. The performance, however, varies spatially with a better score for the gauges located in the southern region in comparison to the central and northern region. A difference in KGE' score between BARRA and ERA-Interim (Figure 2d) shows the difference between the performance of reanalysis datasets in various regions. In most of the locations, the difference is minimal indicating the similar performance by both reanalysis datasets. ERA-Interim generally performs better in the central arid region, whereas BARRA exhibits better scores in the temperate region.

Figure 3 presents the summary of KGE' and its components at a daily scale calculated with reference to gauge data for overall, summer, and winter seasons. In using all the data at a daily scale, the AWAP precipitation proves to be the best estimate at gauge location with the highest value of KGE'. In decomposing the overall performance of KGE' into its components (linear correlation (r), bias ratio (β), and variability ratio (γ)), AWAP achieves the best score for the correlation and variability terms. The AWAP estimates of point rainfall are higher than the gauged observations, and a similar degree of overestimation is exhibited by the BARRA dataset. Despite having a slightly lower correlation compared to ERA-Interim, the variability of the rainfall is better captured by the BARRA dataset. The overall KGE' score is slightly better for ERA-Interim compared to BARRA dataset.

Evaluation of daily rainfall in different seasons reveals a mixed performance between the different reanalysis products. The performance during winter is better than in summer for both reanalysis datasets. However, this difference is larger for BARRA than ERA-Interim. There is less discrepancy in the correlation between BARRA and ERA-Interim. However, the variation of correlation across stations is higher for the BARRA dataset. BARRA tends to overestimate the depth of rainfall and the ERA-Interim under-estimates variability component across all the seasons. The mean and variability for BARRA during winter (JJA) are closer to gauge estimates than in summer (DJF) season. Based on KGE' score, the performance of BARRA is lower than ERA-Interim during summer whereas similar during the winter.

The summary of the KGE' scores based on the climatic zone is presented in Table 2. As expected, the KGE' scores are better for AWAP than for BARRA and ERA-Interim for all three climatic zones. The reanalyses match AWAP most closely for the

Temperate zone, followed by the Arid zone and performance is worst for the Tropical zone. The KGE' metric shows that BARRA performs better at the temperate zone while ERA-Interim at the tropical zone.

Wet day frequency and transition probabilities

Figure 4 shows the comparison of the wet day frequency and transition probabilities (p_{01}, dry day followed by a wet day and p_{11}, wet day followed by a wet day) at different gauge locations. Overall, BARRA shows a better correlation to the gauge estimates than ERA-Interim and is less biased except for dry-wet transition probability. As expected, the frequency of wet days and transition probabilities recorded at gauge are lower compared to all the gridded datasets The wet day frequency in all gridded dataset exhibit positive bias with coarser dataset showing higher bias. The difference is higher for the gauge stations in the northern region. From the colour of scatter, it can be observed that the frequency of wet days is lower over central Australia and higher over the south-east coast and Tasmania. The level of agreement between ERA-Interim and gauge estimates of wet day frequencies varies with the proportion of wet days. At the gauges where the frequency of the wet days is very low (<0.2), the estimates from the ERA-Interim are closer to the gauge. However, for the remaining gauges with higher wet day frequencies, the estimation is positively biased and highly scattered. The wet day frequency varies spatially with the location of the station. The spatial pattern, however, is similar for all datasets. Transition probabilities are also dependent on the location of stations with a similar overall trend for all datasets. The model estimates of transition probability p_{11} are greater than p_{01}. The variation of p_{01} and p_{11} in the northern region is greater than the southern region. BARRA and AWAP both exhibit higher p_{01} values than gauges for all locations. The p_{01} for the northern region is estimated correctly by ERA-Interim, however, it exhibits more bias and spread for the southern region. In comparison to p_{01}, p_{11} shows more scatter and bias. The general trend of overestimation of p_{11} is similar across all datasets and the estimates are higher for the gauges located in northern region compared to southern. The estimates of p_{11} are highly correlated for AWAP. The spatial pattern and correlation are similar for both reanalysis datasets while BARRA exhibits less bias than ERA-Interim.

Quantiles

Figure 5 shows the comparison of quantiles obtained from the gridded datasets to the corresponding quantiles derived for the observed point rainfalls. Since all days were considered in computing the quantile (i.e. both wet and dry days were considered), the 99% value represents "large" rainfalls that are exceeded on average only three or four times per year. On the other hand, the rainfall represented by the 90% quantile (exceeded on average 36 times per year) corresponds to the more frequent wet day rainfalls which depend upon the climatology of the station considered. The AWAP rainfall corresponding to these quantiles are all higher than the point rainfall estimates. The quantile estimates in BARRA and ERA-Interim shows very different patterns to those of the AWAP data set. BARRA estimates for all quantiles (90%, 95%, and 99%) exhibit no consistent bias with location. In contrast, ERA-Interim greatly underestimates the larger precipitation, where the degree of underestimation

increases with increasing quantile. The nature of the differences varies with location and quantile: for example, the ERA-Interim 90% quantile estimates are biased low compared to the point rainfalls in the northern region, yet there is a tendency for all higher quantiles to be biased high. The ERA-Interim reanalysis appears unable to represent higher precipitation magnitudes, and the largest value plateaus at about 40mm for the rarer quantiles.

Categorical Metrics

Figure 6 shows the boxplot of four categorical performance indices computed between three gridded datasets with reference to gauged point rainfalls at a daily scale. The probability of detection (POD) is highest for the AWAP dataset, and the POD values for the two reanalysis data sets are consistently lower and somewhat similar. For rainfall intensities greater than 20 mm and lower than 5mm, BARRA shows higher detection of rainfalls, whereas, for the rainfall intensity 5-20 mm, ERA-Interim shows slightly better detection skill. In contrast, the false alarm ratio (FAR) is higher for BARRA for the heavy rainfall classes (> 20 mm). For the low rainfall classes (<20 mm), FAR is lower for BARRA. The Critical Success Index (CSI) shows that the reanalysis data are not able to adequately represent the distribution of rainfalls greater than 1 mm/day. AWAP, as expected, has higher CSI for each of the rainfall classes. BARRA has higher CSI for all rainfall classes than ERA-Interim except the moderate rainfall (5-20 mm/day) where both reanalysis datasets have similar CSI. The frequency bias shows the positive bias for AWAP and BARRA for all the rain days. ERA-Interim shows a negative bias for higher rainfall intensities (>20 mm).

4.3 Grid-to-grid analysis

KGE' and its components

The boxplot of KGE' and its components shown in Figure 7 is computed between reanalysis datasets with reference to AWAP dataset at a daily scale. Among the components, the correlation is similar for both reanalysis datasets, though BARRA exhibits a slightly wider range of correlations than ERA-Interim. However, the bias ratio (β) and variability ratio (γ) obtained from BARRA is closer to 1, which is appreciably better than the corresponding statistics for ERA-Interim. The difference in the overall KGE' measure for the two data sets is not as pronounced compared to its components.

In Table 2, a summary of the KGE' score is also presented with reference to the AWAP dataset. The reanalysis datasets show mixed performance depending on the climate zone and the metrics used. BARRA shows improved performance over ERA-Interim in the temperate zone with better correlation, bias ratio, variability ratio, and KGE' statistics. The difference in correlation coefficients between the reanalysis datasets is higher for the tropical zone but similar for the arid and temperate zones. Further, the BARRA data set exhibits smaller bias than ERA-Interim across all climatic zones. The variability ratio, however, shows contrasting patterns for the two reanalysis datasets. ERA-Interim shows a marked underestimation in

variability across all zones. On the other hand, BARRA closely represents the variability in arid and temperate zones with overestimation in the tropics.

Wet day frequency and transition probabilities

The frequency of wet day and transition probabilities obtained from reanalysis datasets at AWAP grid locations are shown in Figure 8. Overall, the representation of wet day frequency and transition probabilities follow a similar pattern to the point-to-grid assessment (Figure 4). In comparing the gridded datasets with AWAP data as the benchmark, the estimates of wet day frequency and transition probabilities from BARRA are closer than those of ERA-Interim. However, both reanalysis datasets show an improved correlation and reduced bias in grid-to-grid over point-to-grid evaluation. This is expected as this bias is attributed to the spatial resolution of the data.

Quantiles

The comparison of quantiles in gridded datasets in Figure 9 shows a similar pattern as observed in Figure 5 for the gauged point rainfalls. Both reanalysis datasets represent the 90% quantile reasonably well. However, the difference between the reanalysis products increases at higher quantiles. The degree of bias in ERA-Interim is considerably more pronounced compared to BARRA.

Categorical Metrics

Figure 10 shows the boxplot of four categorical performance indices computed between reanalysis datasets with reference to AWAP dataset at a daily scale. In general, the POD for both datasets is similar to the comparison against gauge. For the larger precipitation intensities, the detection capacity of BARRA is higher whereas, for the smaller rainfall intensities, the detection from both datasets is close to each other. The FAR scores for BARRA estimates of higher rainfalls are slightly greater, indicating the larger occurrence of false alarms while detecting higher rainfall. For the low rainfall classes, BARRA exhibits slightly better FAR than ERA-Interim. The CSI score is higher for BARRA across all rainfall classes and indicates that BARRA is slightly more skilful than ERA-Interim. At larger rainfall class, the CSI is notably higher for BARRA indicating its better performance in representing larger rainfall. The frequency bias shows that the BARRA precipitation tends to produce more events of light rainfall while missing out on some larger rainfalls. The bias in BARRA is, however, smaller in comparison with ERA-Interim which shows a marked underestimation of larger rainfall events (>20 mm).

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-607
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 19 February 2019

5 Discussion

The evaluation of daily precipitation from BARRA and its comparison against existing datasets reveals a mixed performance under varying benchmark dataset and the metrics. The key insights on the BARRA reanalysis precipitation obtained from the results are discussed below.

5.1 Representation of spatial precipitation structure

The BARRA dataset exhibits similar spatial patterns of mean annual rainfall as the AWAP dataset in regions where gauge density is highest (Figure 1). However, the reliability in the evaluation of BARRA in the central regions of Australia is confounded by the lack of gauges. The coarser resolution of ERA-Interim misses the small-scale variability. In contrast, a more realistic representation of complex topography in BARRA helps to capture fine-scale features such as orographic precipitation (Su et al., 2018). A good level of agreement in spatial patterns of rainfalls with AWAP and similarity with ERA-Interim estimates (over both land and ocean) at a coarser scale suggests that the BARRA dataset might provide useful information on the distribution of rainfall in regions where direct measurement of precipitation is not available.

5.2 Representation of average precipitation and variability

In general, a slight overestimation is observed in mean rainfall from high-resolution gridded data when compared to gauged point rainfall estimates (bias in mm/day for AWAP = 0.48, BARRA = 0.43). In AWAP, the wet bias observed at low rainfall is due to the Barnes analysis (Jones et al., 2009). In addition, the weight functions tend to spread rainfalls across the grid making coastal stations wetter since the density of stations near the coast is greater. When average AWAP precipitation is considered a benchmark for evaluation, BARRA (R=0.93, bias = -0.05mm/day) shows better agreement than ERA-Interim (R=0.84, bias = -0.49 mm/day). In the locations with higher average precipitation, the ERA-Interim underestimates gauged point rainfalls. This can be attributed to smoothing out of the large rainfall due to coarser resolution of ERA-Interim.

In addition, the temporal variability is represented well by BARRA while this is underestimated by ERA-Interim (Figure 3 and Figure 7). As with the assessment of spatial variability, the reason for this underestimation is the coarser resolution of ERA-Interim. The variability ratio varies across climatic zones with largest difference in the tropical zone. BARRA greatly overestimates variability in the tropical region, whereas the opposite is true for ERA-Interim (Table 2). The discrepancy in the performance across climatic zones is further discussed in section 5.5.

5.3 Representation of wet day frequency and transition probabilities

All the gridded datasets exhibit higher frequencies of wet day occurrence compared to gauge point rainfalls. This difference is consistent with the physical reasoning that the likelihood of rainfall occurring over an area is always higher than that over a point location. Moreover, an increase in the area produces a higher difference in the wet days and this is consistent with the

increasing levels of bias with grid cell size between AWAP, BARRA, and ERA-Interim (Figure 4). In comparison with AWAP as a benchmark, BARRA closely represents the wet day frequency and exhibits less uncertainty at higher estimates compared to ERA-Interim (Figure 8).

The variation of transition probabilities p_{01} and p_{11} at a location can be attributed to the seasonal distribution of precipitation at that location. Most rainy days in the northern region occur during October-March with very low or no rainfall for the rest of the season. Due to distinct dry (and wet) season, the likelihood of a dry day following a dry day (and a wet day following a wet day) is higher. This results in larger p_{11} and smaller p_{01} estimates. In contrast, there is not a distinct wet and dry season in the south, and p_{01} is close to p_{11} for the southern region, mostly due to the light rainfall distributed over the longer period. The ability of the reanalysis datasets to represent transition probabilities also varies with location due to the climatology and varying precipitation mechanisms across locations. The transition probabilities p_{01} spread is estimated well by all gridded datasets, whereas the p_{11} estimates are more varied. Due to the likelihood of higher frequency of wet days in gridded datasets, the transition probability p_{11} is higher in gridded datasets and exhibits higher variability for the tropical zone.

5.4 Over/Under-estimation of large rainfall events

The comparison of quantiles (Figures 5 and 9) demonstrates that the BARRA dataset is able to represent the higher quantiles more accurately than ERA-Interim. The BARRA estimates of quantiles show less bias, but variance increases with the magnitude of precipitation amount. This uncertainty could be due to displacement error as the precipitation becomes more localized with an increase in magnitude. ERA-Interim, on the other hand, greatly underestimates the higher quantiles. This could be attributed to its coarser resolution resulting in the averaging of rainfall over a larger area. These results are consistent with published findings (Isotta et al., 2014; Jermey and Renshaw, 2016) that high-resolution regional reanalysis improves over ERA-Interim in the representation of large rainfall.

The categorical evaluation (Figure 6 and Figure 10) also illustrates the improved performance of the BARRA dataset over ERA-Interim during larger rainfall events. Because of the higher spatial resolution, BARRA provides a more accurate representation of larger rainfalls. Therefore, the probability of detection is higher for the larger rainfall classes. Despite exhibiting a higher hit rate, BARRA also shows higher false alarm ratios. The increased false alarm in BARRA is due to the higher number of large rainfall events reported by the BARRA dataset. With larger rainfall events, the false alarm is also likely to increase. Such a trend is usually observed in the assessment of reanalysis and satellite rainfall estimates (Zambrano-Bigiarini et al., 2017). Critical Success Index, the metric which penalizes both misses and false alarms, represents the overall skill of the data. The performance of BARRA lies between the skilful AWAP and less skilful ERA-Interim. The greater CSI for BARRA compared to ERA-Interim suggests that its improved hit rate outweighs the frequency of false alarms. The

difference in CSI between BARRA and ERA-Interim is greater for larger rainfall classes, with the former yielding a better score.

5.5 Superior performance over temperate than over tropical and arid

Most metrics suggest that the performance of BARRA is superior in the southern (temperate) region compared to the northern (tropical) region. The summary of KGE' and its component across climatic zones (Table 1) clearly shows the variation in performance over different climatic zones. The possible explanation for this may be the difference in the climatic system driving the precipitation in those regions and the scheme used to generate precipitation (Su et al., 2018). Ebert et al. (2007) and de Leeuw et al. (2015) also observed less accurate performance of NWP models at the regime of convective precipitation. The rainfall in the Northern region of Australia is largely convective in nature during summer which is not well captured by the reanalysis model. In addition, the network of surface observations used to develop BARRA is denser in the southern region compared to the northern. This difference in the information content used to derive the estimates may have caused discrepancies in performance in those regions.

5.6 Varying performance in point-to-grid and grid-to-grid evaluation:

AWAP is the best estimate of gridded rainfalls over Australian and not surprisingly it shows the best agreement with point rainfall. Despite its limitation in some areas due to poor gauging density, the gridded nature of the estimates provides the best basis for comparison with reanalysis datasets at the area with dense gauging network.

This study shows that the general pattern of performance between the reanalysis datasets is similar in the point-to-grid evaluation. However, the evaluation against the gridded AWAP estimates showed a markedly better performance by BARRA in terms of overall bias, variability, wet day frequencies, transition probabilities, quantiles and categorical metrics. As expected, BARRA better represents areal rather than point rainfall as it represents the average precipitation field over a grid cell.

6 Conclusions

The purpose of the current study is to document the performance of the BARRA dataset at a daily scale and to provide a comparative analysis of its strengths and limitations relative to other available datasets. The analysis includes point-to-grid and grid-to-grid evaluations at the gauge locations. A range of metrics representing correlation, daily precipitation statistics, and categorical performance are explored and compared on an annual as well as a seasonal basis.

The high-resolution nature of the BARRA dataset provides more detailed and accurate estimates of rainfall across the Australian region than the coarser ERA-Interim. BARRA precipitation exhibits good agreement with the average annual rainfall from AWAP as well as to the gauge dataset. The correlation statistics of the BARRA estimates are slightly lower than

for the global reanalysis (ERA-Interim). However, the depth and variability of daily precipitation from AWAP are better reproduced by BARRA than by ERA-Interim. We can conclude that BARRA precipitation is representative at the spatial scale of AWAP considering that AWAP data provides the best basis for comparison with the reanalysis datasets.

BARRA provides largely unbiased estimates of larger rainfall quantiles whereas ERA-Interim are clearly underestimated. Categorical evaluation also shows a better correspondence of the larger events in BARRA compared to ERA-Interim. This has important implications for hydrological modelling as simulations of runoff processes are heavily dependent on how realistically the precipitation field is represented both spatially and temporally. Information on large rainfalls also important for many engineering investigations where the design loading conditions of interest are dependent on the characteristics of daily rainfall. The superior performance of BARRA in representing large rainfall strengthen a case for its use over ERA-Interim where information about extremes is required.

BARRA closely reproduces the frequency of wet days and dry-wet transition probabilities. This evaluation broadly supports BARRA precipitation for its ability to reproduce precipitation statistics at a daily scale. BARRA precipitation could be useful for assessing variation in the spatial and temporal characteristics of precipitation in a consistent manner that is not influenced by differences in gauge density. In addition, it could potentially be used as a source of data in the central arid zone where AWAP estimates are poor or not available. Considering the limitations in the availability of gauged datasets, BARRA could be considered as a valuable reference dataset for hydro-climatic analysis across the whole of Australia, particularly where gauging density is low.

The core attraction of the BARRA dataset is the availability of sub-daily precipitation estimates. Such information is not available in the AWAP data set, and the spatial resolution of the estimates is higher than the currently available global reanalysis and satellite datasets. Accordingly, future work will be directed towards an analysis of BARRA estimates of sub-daily precipitation to assess its ability to represent precipitation at finer time scales.

Acknowledgments

The authors gratefully acknowledge the financial support provided by Seqwater and the Bureau of Meteorology to partially fund the first author's PhD scholarship. We would like to thank colleagues at the Bureau of Meteorology (Peter Steinle, Robert Smalley and Alex Evans) for discussions at various stages of research. We are also grateful to Dörte Jakob for commenting on early results and for providing feedback on drafts of the manuscript.

References

Baez-Villanueva, O. M., Zambrano-Bigiarini, M., Ribbe, L., Nauditt, A., Giraldo-Osorio, J. D. and Thinh, N. X.: Temporal

and spatial evaluation of satellite rainfall estimates over different regions in Latin-America, Atmos. Res., 213(December 2017), 34–50, doi:10.1016/J.ATMOSRES.2018.05.011, 2018.

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M. J. M., Weedon, G. P., Brocca, L., Pappenberger, F., Huffman, G. J. and Wood, E. F.: Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling, Hydrol. Earth Syst. Sci., 21(12), 6201–6217, doi:10.5194/hess-21-6201-2017, 2017.

Bollmeyer, C., Keller, J. D., Ohlwein, C., Wahl, S., Crewell, S., Friederichs, P., Hense, A., Keune, J., Kneifel, S., Pscheidt, I., Redl, S. and Steinke, S.: Towards a high-resolution regional reanalysis for the european CORDEX domain, Q. J. R. Meteorol. Soc., 141(686), 1–15, doi:10.1002/qj.2486, 2015.

Bukovsky, M. S. and Karoly, D. J.: A Brief Evaluation of Precipitation from the North American Regional Reanalysis, J. Hydrometeorol., 8(4), 837–846, doi:10.1175/JHM595.1, 2007.

Chiaravalloti, F., Brocca, L., Procopio, A., Massari, C. and Gabriele, S.: Assessment of GPM and SM2RAIN-ASCAT rainfall products over complex terrain in southern Italy, Atmos. Res., 206(December 2017), 64–74, doi:10.1016/j.atmosres.2018.02.019, 2018.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., Mcnally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J. N. and Vitart, F.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc., 137(656), 553–597, doi:10.1002/qj.828, 2011.

Ebert, E. E., Janowiak, J. E. and Kidd, C.: Comparison of near-real-time precipitation estimates from satellite observations and numerical models, Bull. Am. Meteorol. Soc., 88(1), 47–64, doi:10.1175/BAMS-88-1-47, 2007.

Gebremichael, M.: Framework for satellite rainfall product evaluation, pp. 265–275., 2010.

Gupta, H. V., Kling, H., Yilmaz, K. K. and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377(1–2), 80–91, doi:10.1016/j.jhydrol.2009.08.003, 2009.

Isotta, F. A., Vogel, R. and Frei, C.: Evaluation of European regional reanalyses and downscalings for precipitation in the Alpine region, Meteorol. Zeitschrift, 24(1), 15–37, doi:10.1127/metz/2014/0584, 2014.

Jakob, D., Su, C.-H., Eizenberg, N., Kociuba, G., Steinle, P. and Al., E.: An atmospheric high-resolution regional reanalysis for Australia, Bull. Aust. Meteorol. Oceanogr. Soc., 30(3), 16–23, 2017.

Jermey, P. M. and Renshaw, R. J.: Precipitation representation over a two-year period in regional reanalysis, Q. J. R. Meteorol. Soc., 142(696), 1300–1310, doi:10.1002/qj.2733, 2016.

Johnson, F., Hutchinson, M. F., The, C., Beesley, C. and Green, J.: Topographic relationships for design rainfalls over Australia, J. Hydrol., 533, 439–451, doi:10.1016/j.jhydrol.2015.12.035, 2016.

Jones, D., Wang, W. and Fawcett, R.: High-quality spatial climate data-sets for Australia, Aust. Meteorol. Oceanogr. J., 58(04), 233–248, doi:10.22499/2.5804.003, 2009.

Kirschbaum, D. B., Huffman, G. J., Adler, R. F., Braun, S., Garrett, K., Jones, E., McNally, A., Skofronick-Jackson, G., Stocker, E., Wu, H. and Zaitchik, B. F.: NASA's Remotely Sensed Precipitation: A Reservoir for Applications Users, Bull. Am. Meteorol. Soc., 98(6), 1169–1184, doi:10.1175/BAMS-D-15-00296.1, 2017.

Kling, H., Fuchs, M. and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, J. Hydrol., 424–425, 264–277, doi:10.1016/j.jhydrol.2012.01.011, 2012.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., ONODA, H., ONOGI, K., KAMAHORI, H., KOBAYASHI, C., ENDO, H., MIYAOKA, K. and TAKAHASHI, K.: The JRA-55 Reanalysis: General Specifications and Basic Characteristics, J. Meteorol. Soc. Japan. Ser. II, 93(1), 5–48, doi:10.2151/jmsj.2015-001, 2015.

Kucera, P. A., Ebert, E. E., Turk, F. J., Levizzani, V., Kirschbaum, D., Tapiador, F. J., Loew, A. and Borsche, M.: Precipitation from space: Advancing earth system science, Bull. Am. Meteorol. Soc., 94(3), 365–375, doi:10.1175/BAMS-D-11-00171.1, 2013.

de Leeuw, J., Methven, J. and Blackburn, M.: Evaluation of ERA-Interim reanalysis precipitation products using England and Wales observations, Q. J. R. Meteorol. Soc., 141(688), 798–806, doi:10.1002/qj.2395, 2015.

Parker, W. S.: Reanalyses and Observations: What's the Difference?, Bull. Am. Meteorol. Soc., 97(9), 1565–1572, doi:10.1175/BAMS-D-14-00226.1, 2016.

Peel, M. C., Finlayson, B. L. and McMahon, T. A.: Updated world map of the Köppen-Geiger climate classification, Hydrol. Earth Syst. Sci., 11(5), 1633–1644, doi:10.5194/hess-11-1633-2007, 2007.

Peña-Arancibia, J. L., van Dijk, A. I. J. M., Renzullo, L. J. and Mulligan, M.: Evaluation of Precipitation Estimation Accuracy in Reanalyses, Satellite Products, and an Ensemble Method for Regions in Australia and South and East Asia, J. Hydrometeorol., 14(4), 1323–1333, doi:10.1175/JHM-D-12-0132.1, 2013.

Robertson, D. E., Shrestha, D. L. and Wang, Q. J.: Post-processing rainfall forecasts from numerical weather prediction models for short-term streamflow forecasting, Hydrol. Earth Syst. Sci., 17(9), 3587–3603, doi:10.5194/hess-17-3587-2013, 2013.

Saha, S., Moorthi, S., Pan, H. L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y. T., Chuang, H. Y., Juang, H. M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., Van Den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J. K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C. Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R. W., Rutledge, G. and Goldberg, M.: The NCEP climate forecast system reanalysis, Bull. Am. Meteorol. Soc., 91(8), 1015–1057, doi:10.1175/2010BAMS3001.1, 2010.

Salio, P., Hobouchian, M. P., García Skabar, Y. and Vila, D.: Evaluation of high-resolution satellite precipitation estimates over southern South America using a dense rain gauge network, Atmos. Res., 163, 146–161, doi:10.1016/j.atmosres.2014.11.017, 2015.

Soares, P. M. M., Cardoso, R. M., Miranda, P. M. A., de Medeiros, J., Belo-Pereira, M. and Espirito-Santo, F.: WRF high resolution dynamical downscaling of ERA-Interim for Portugal, Clim. Dyn., 39(9), 2497–2522, doi:10.1007/s00382-012-1315-2, 2012.

Su, C.-H., Eizenberg, N., Steinle, P., Jakob, D., Fox-Hughes, P., Rennie, S., Franklin, C., Dharssi, D. and Zhu, H.: BARRA v1.0: The Bureau of Meteorology Atmospheric high-resolution Regional Reanalysis for Australia, submitted to Geoscientific Model Development, 2018.

Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., Levizzani, V. and De Roo, A.: Validation of Satellite-Based Precipitation Products over Sparsely Gauged African River Basins, J. Hydrometeorol., 13(6), 1760–1783, doi:10.1175/JHM-D-12-032.1, 2012.

Vidal, J. P., Martin, E., Franchistéguy, L., Baillon, M. and Soubeyroux, J. M.: A 50-year high-resolution atmospheric reanalysis over France with the Safran system, Int. J. Climatol., 30(11), 1627–1644, doi:10.1002/joc.2003, 2010.

Zambrano-Bigiarini, M., Nauditt, A., Birkel, C., Verbist, K. and Ribbe, L.: Temporal and spatial evaluation of satellite-based rainfall estimates across the complex topographical and climatic gradients of Chile, Hydrol. Earth Syst. Sci., 21(2), 1295–

1320, doi:10.5194/hess-21-1295-2017, 2017.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-607
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 19 February 2019

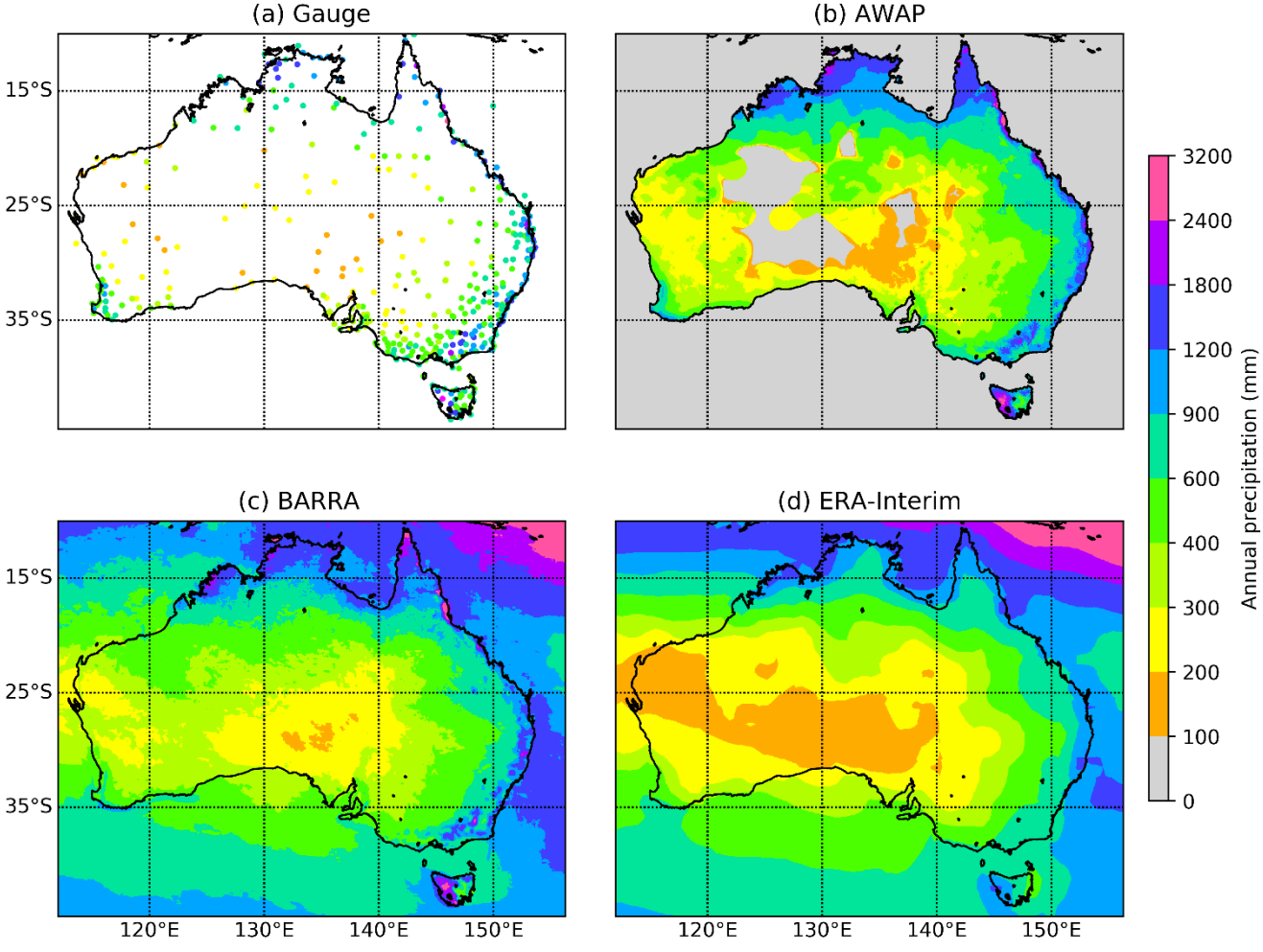

Figure 1 Average annual precipitation over Australian region (a)Gauge data, (b) AWAP, (c) BARRA, and (d) ERA-Interim. Missing values and ocean are masked in AWAP and represented by gray colour.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-607
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 19 February 2019

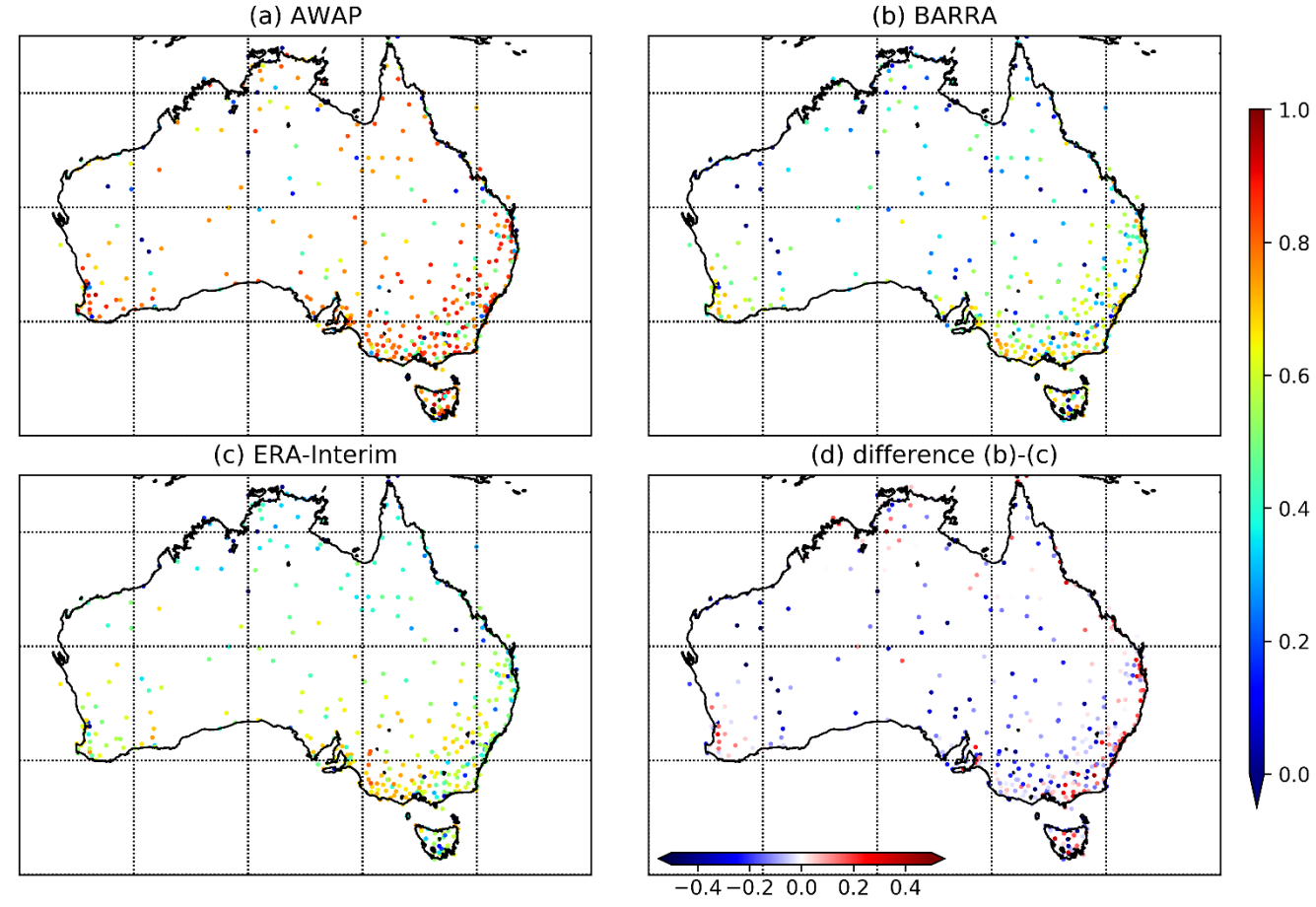

Figure 2 Modified Kling-Gupta efficiency (KGE') against gauge dataset at daily scale for (a) AWAP, (b) BARRA, (c) ERA-Interim, and (d) difference of KGE' between BARRA and ERA-Interim.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-607
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 19 February 2019

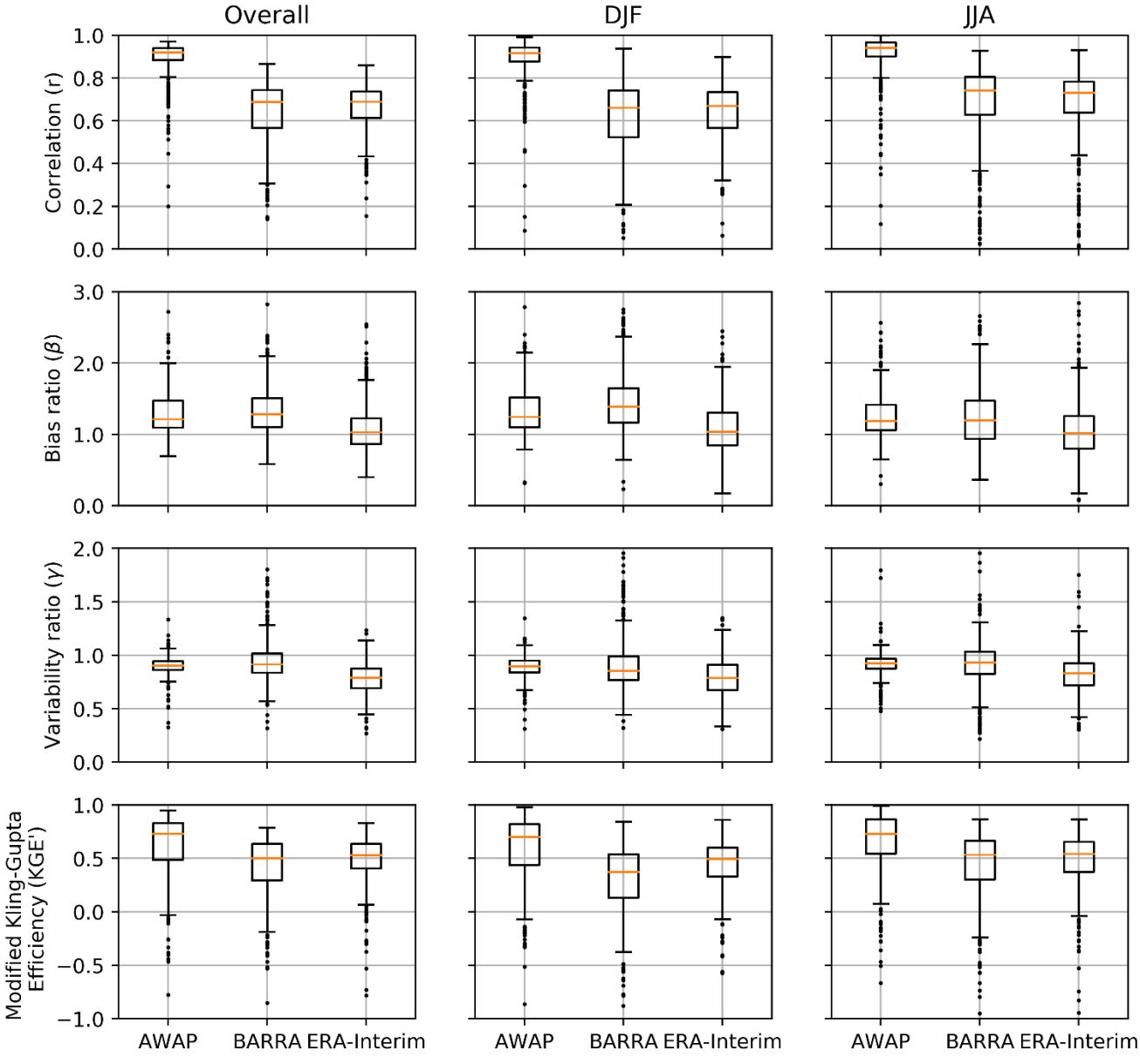

Figure 3 Boxplot of correlation (r), bias ratio (β), variability ratio (γ), and modified Kling-Gupta Efficiency (KGE') between daily precipitation estimates from gridded datasets and observations at gauge locations. Columns represent overall data, summer (DJF), and winter (JJA) respectively. Each box extends from first to third quartile, medians are marked in each box, and whisker extends to furthest point or 1.5 times the interquartile range whichever is closer.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-607
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 19 February 2019

Hydrology and Earth System Sciences
Discussions

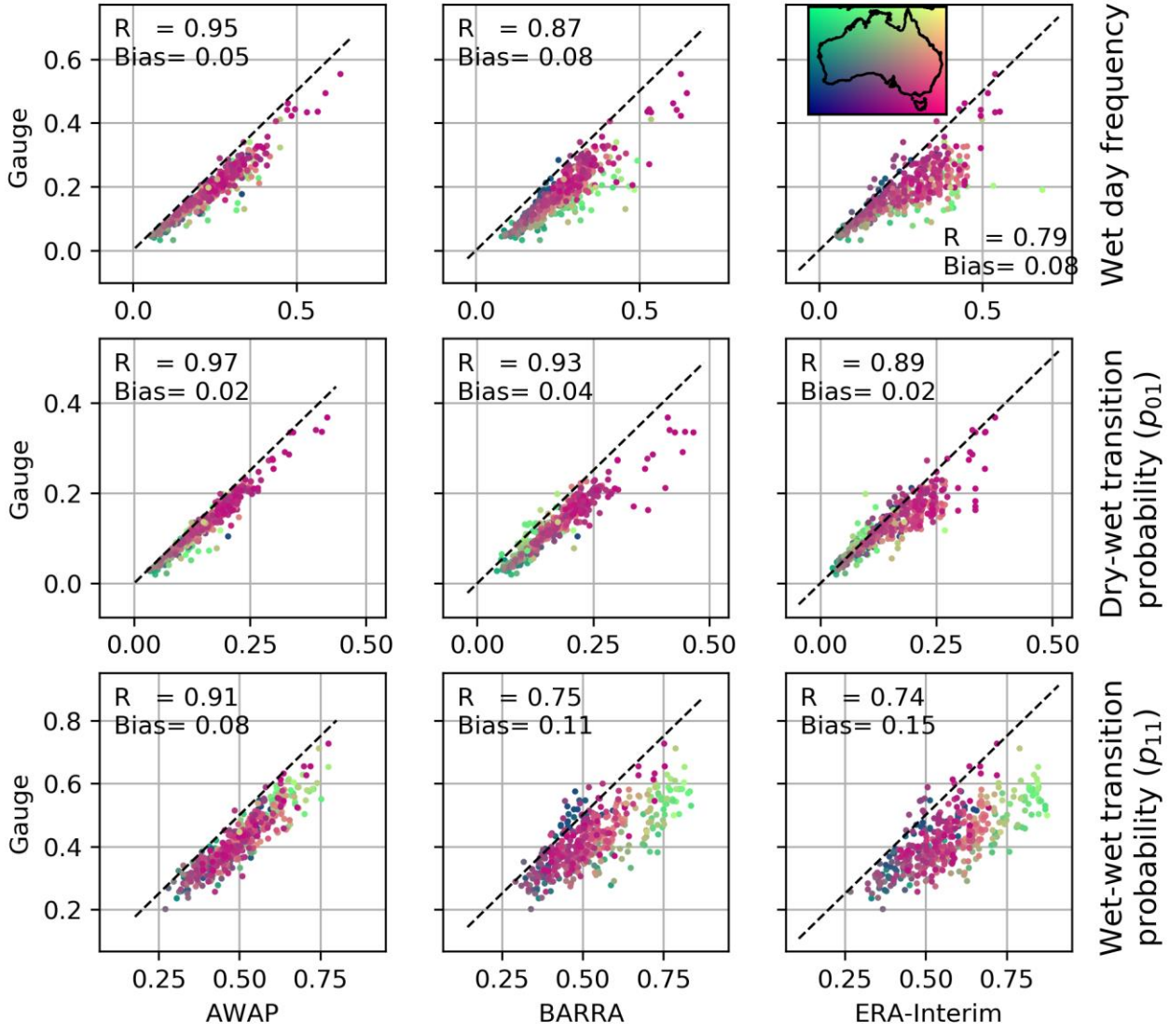

Figure 4 Wet day frequency estimates from gridded datasets (AWAP, BARRA, and ERA-Interim) and observations at gauge locations (first row). Transition probabilities: dry-wet (p_{01}, second row), and wet-wet (p_{11}, third row). Colour of scatter indicates the location of the station.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-607
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 19 February 2019

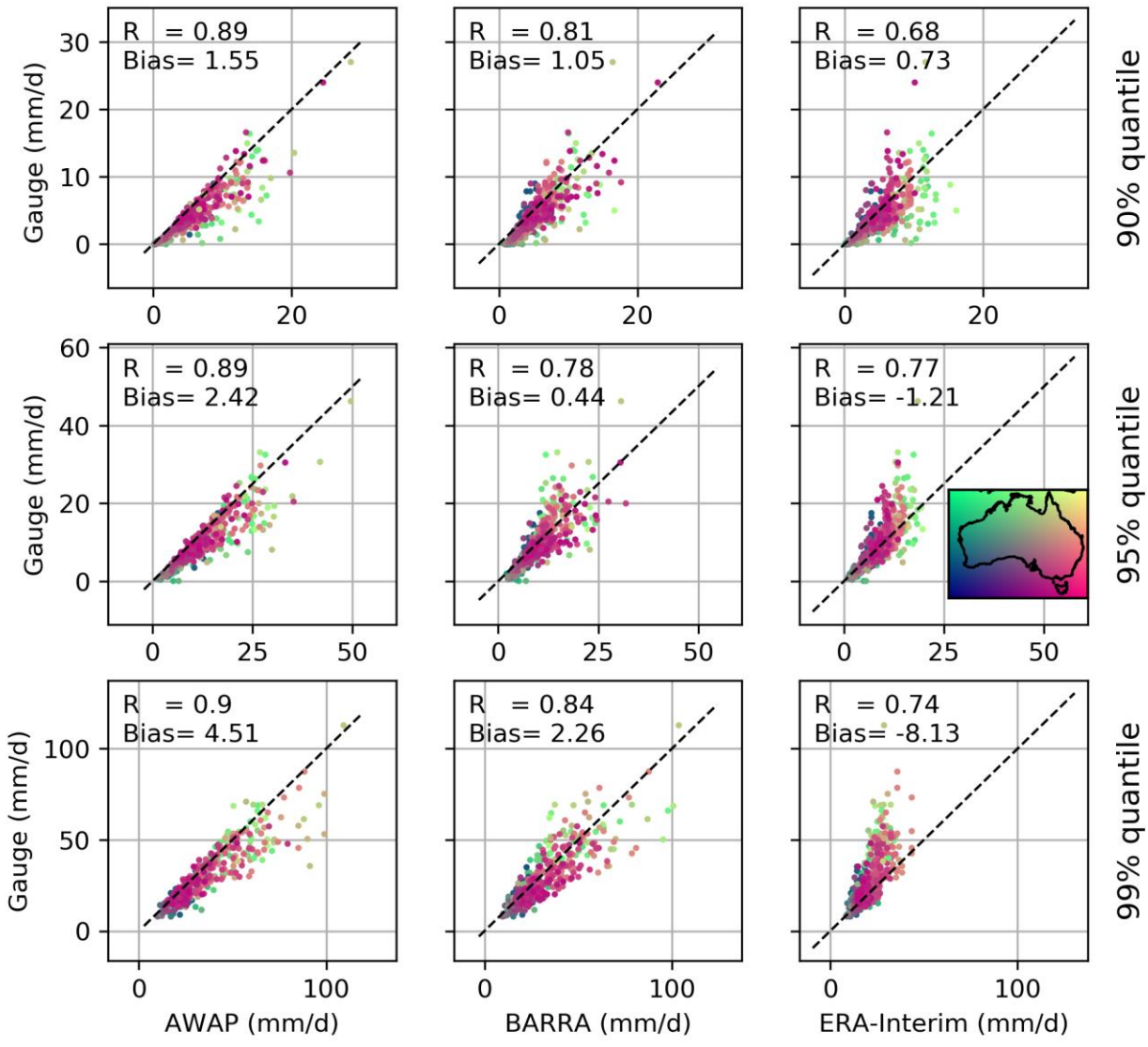

Figure 5 Quantile estimates (mm/day; 90, 95 and 99% in the top, middle and bottom row respectively) from gridded datasets (AWAP, BARRA, and ERA-Interim) and observations at gauge locations. The colour of the scatter indicates the location of the stations.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-607
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 19 February 2019

Hydrology and Earth System Sciences
Discussions

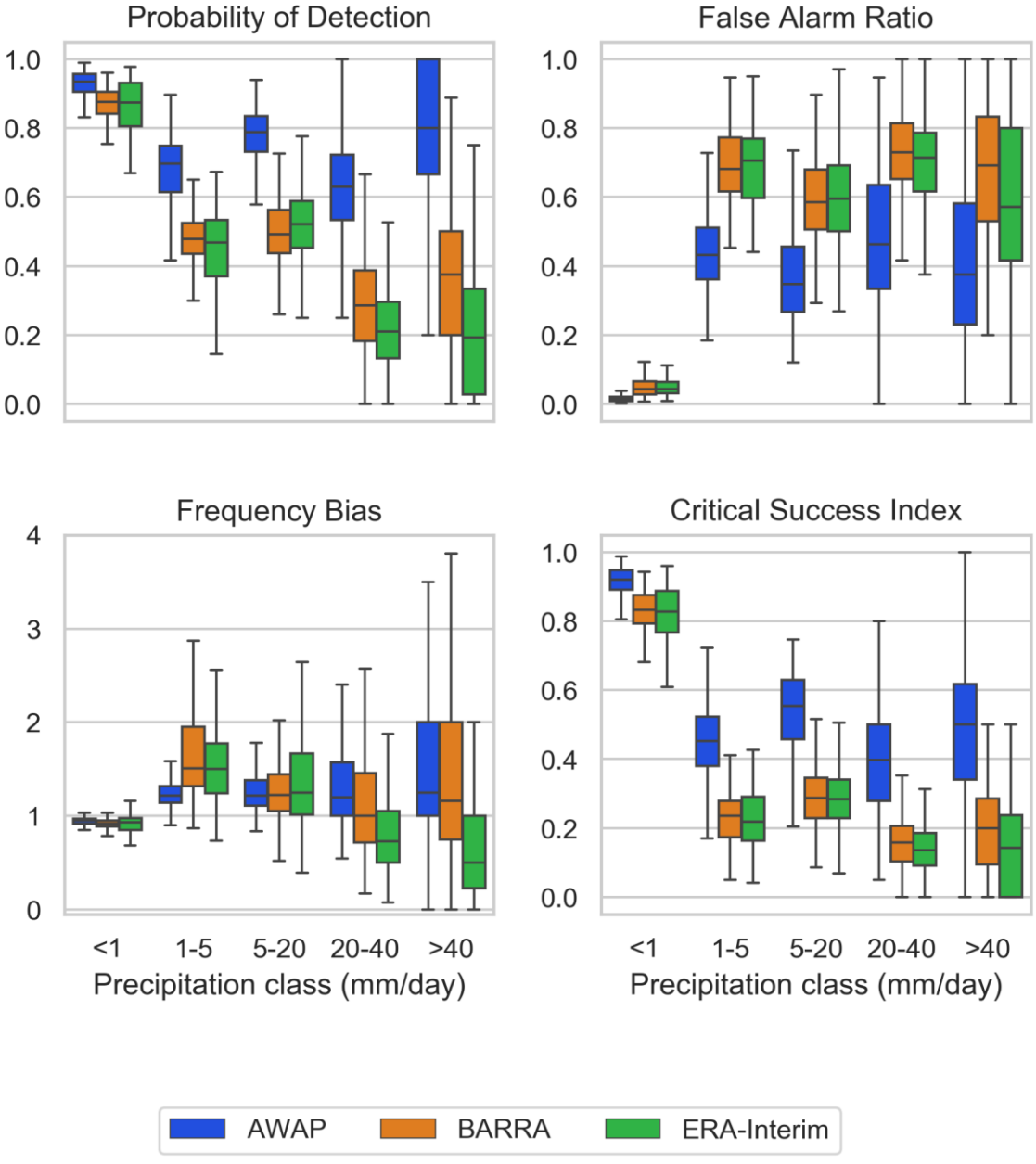

Figure 6 Boxplot of categorical performance indices (Probability of detection, false alarm ratio, frequency bias , and critical success index) calculated against gauge data for five classes of rainfall intensity. Each box extends from first to third quartile, medians are marked in each box, and whisker extends to furthest point or 1.5 times the interquartile range whichever is closer.

5

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-607
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 19 February 2019

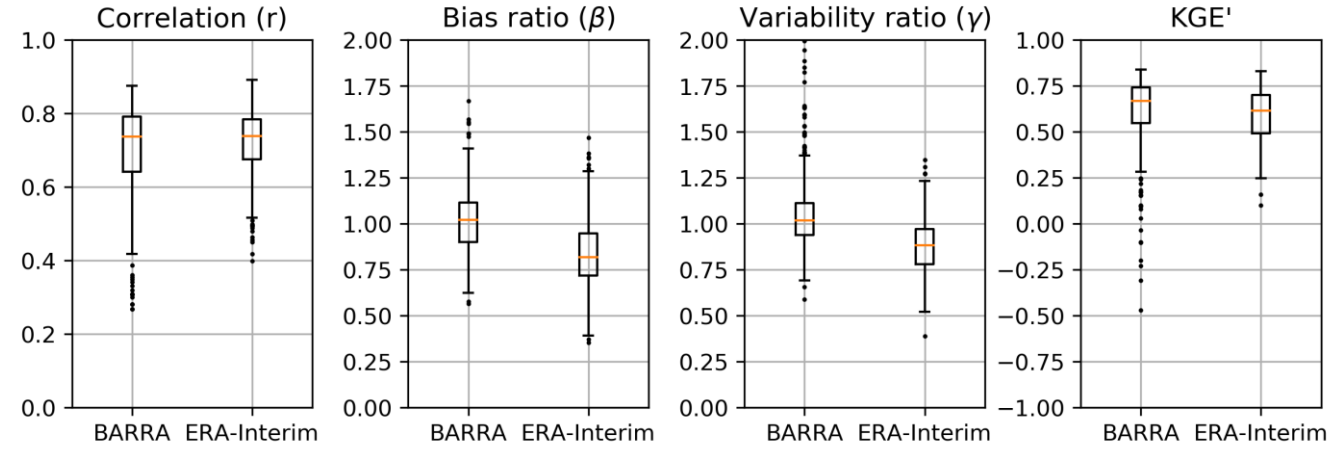

Figure 7 Boxplot of correlation (r), bias ratio (β), variability ratio (γ), and modified Kling-Gupta Efficiency (KGE') against AWAP dataset. Each box extends from first to third quartile, medians are marked in each box, and whisker extends to furthest point or 1.5 times the interquartile range whichever is closer.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-607
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 19 February 2019

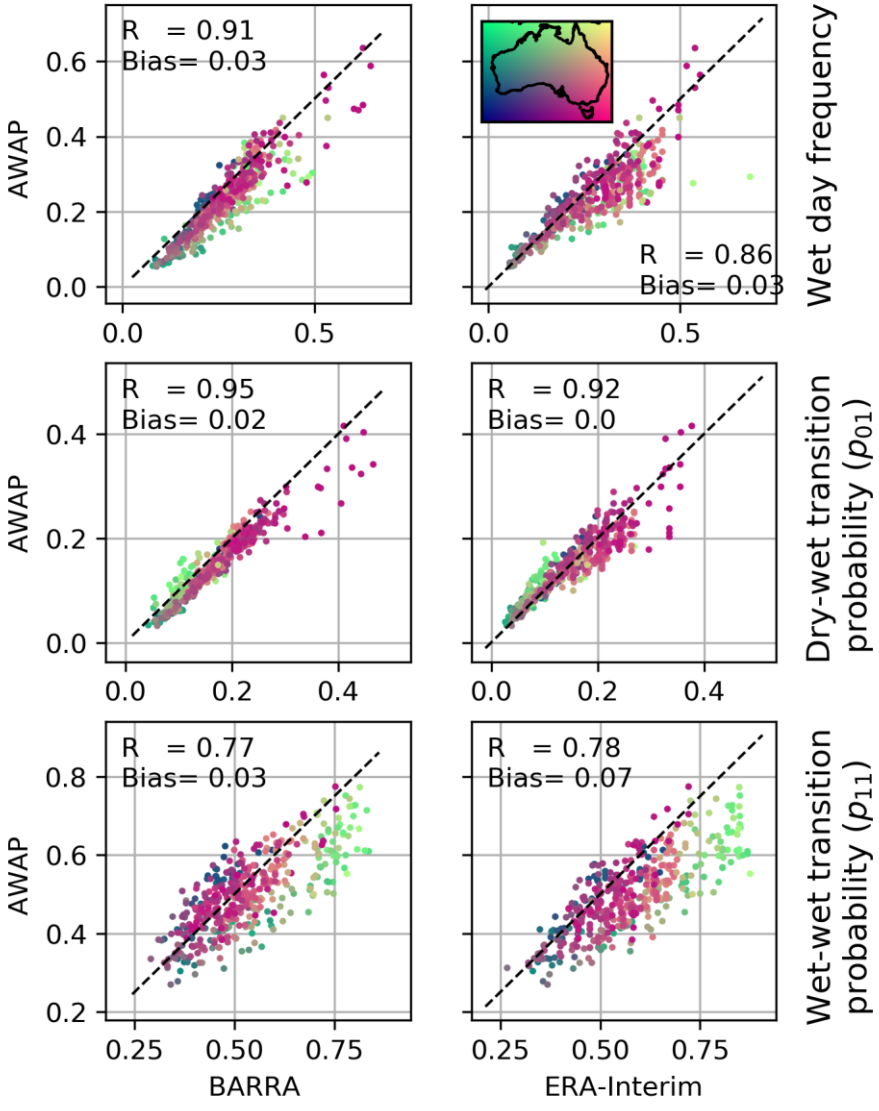

Figure 8 Wet day frequency estimates from reanalysis datasets (BARRA, and ERA-Interim) and AWAP at gauge locations (first row). Transition probabilities: dry-wet (p_{01}, second row), and wet-wet (p_{11}, third row). Colour of scatter indicates the location of the AWAP grid.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-607
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 19 February 2019

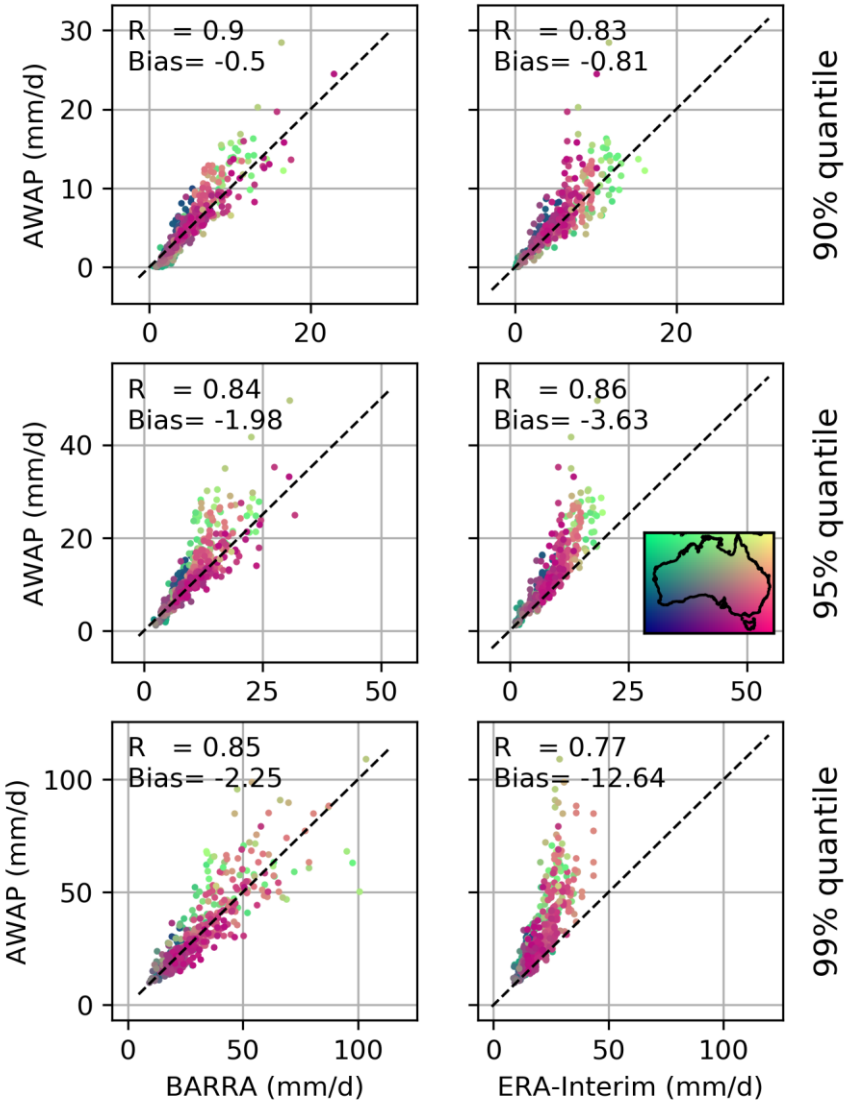

Figure 9 Quantile estimates (mm/day; 90, 95 and 99%, in the top, middle and bottom row respectively) from reanalysis datasets (BARRA, and ERA-Interim) and AWAP at gauge locations. The colour of the scatter indicates the location of the AWAP grid.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-607
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 19 February 2019

Hydrology and Earth System Sciences
Discussions

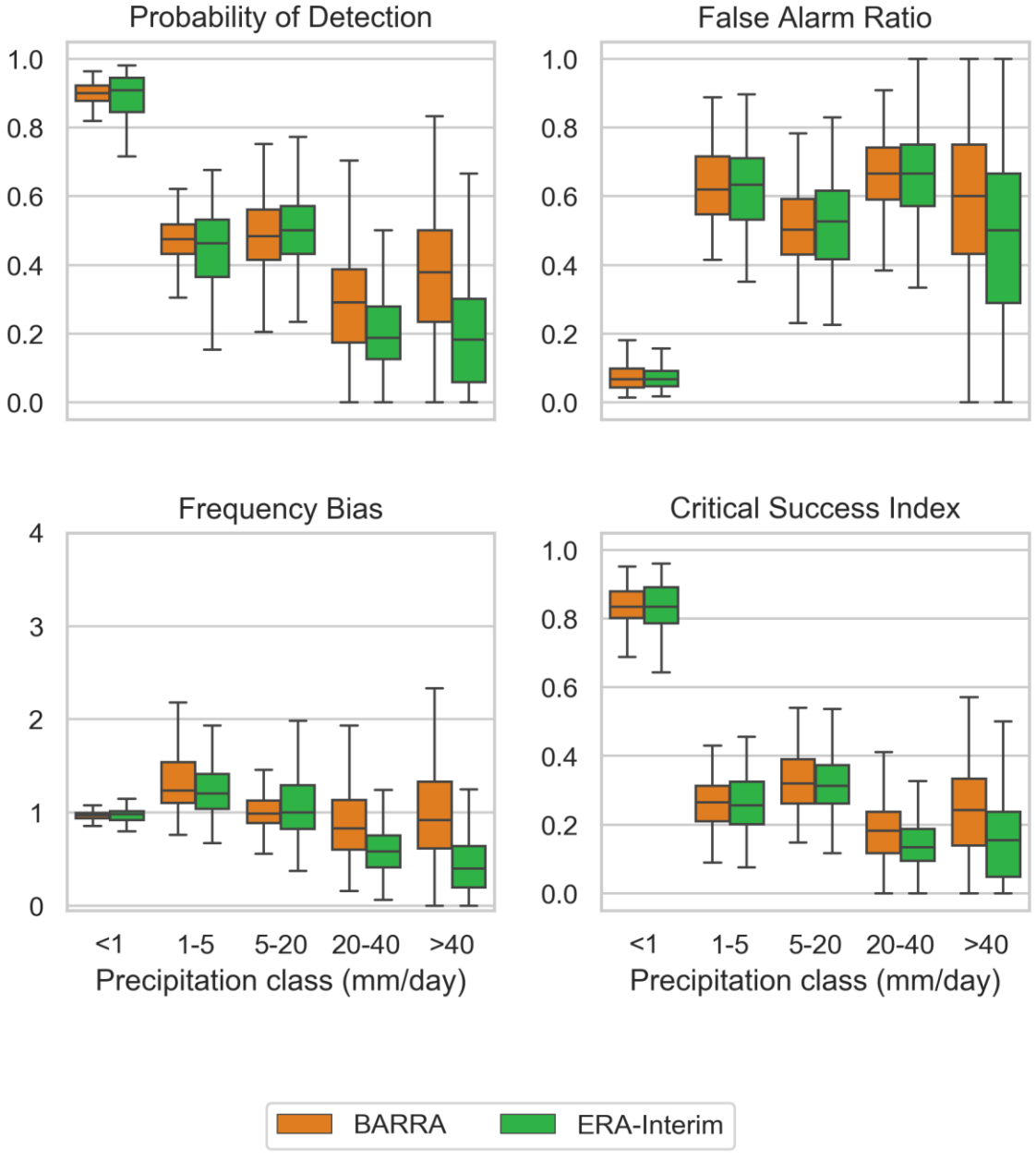

Figure 10 Boxplot of categorical performance indices (Probability of detection, false alarm ratio, frequency bias , and critical success index) calculated against AWAP data for five classes of rainfall intensity. Each box extends from first to third quartile, medians are marked in each box, and whisker extends to furthest point or 1.5 times the interquartile range whichever is closer.

5

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-607
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 19 February 2019

Hydrology and Earth System Sciences
Discussions

Table 1 Metrics used in the evaluation of precipitation data.

	Measures	Equation	Description
Continuous metrics	Pearson correlation coefficient (r)	$\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$	Pearson product moment correlation measures the linear correlation of the observed and modelled values. Range [-1,1] Best value =1
	Bias ratio (β)	$\frac{\bar{y}}{\bar{x}}$	Bias ratio measures the agreement between means of modelled and observed values. Evaluates bias in total volume over-estimation (β>1) under-estimation (β<1) Range [0, +∞] Best value =1
	Variability ratio (γ)	$\frac{CV_S}{CV_O}=\frac{s_y/\bar{y}}{s_x/\bar{x}}$	Variability ratio measures the agreement between Coefficient of Variation (CV) of modelled and observed values. The use of CV in the calculation of γ ensures that the bias and variability ratio are not cross-correlated (Kling et al., 2012) Evaluates spread . over-dispersed (γ>1) or under-dispersed (γ<1) Range [0, +∞] Best value =1
	Modified Kling-Gupta	$1-\sqrt{(r-1)^2+(\beta-1)^2+(\gamma-1)^2}$	Summarises correlation, bias ratio and variability ratio

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-607
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 19 February 2019

	Efficiency (KGE')		Range [-∞,1] Best value =1
Categorical metrics	Probability of Detection (POD) or Hit Rate	$\frac{H}{H+M}$	POD measures the fraction of number of events that were correctly reported by the model. Sensitive to hits but ignores false alarms. Suitable for rare events Range [0,1] Best value =1
	False Alarm Ratio (FAR)	$\frac{FA}{H+FA}$	FAR measures the fraction of number of non-events that were incorrectly reported as events by the model. Sensitive to false alarm but ignores misses Range [0,1] Best value =0
	Threat Score (TS) or Critical Success Index	$\frac{H}{H+FA+M}$	Measures the fraction of observed events that were correctly modelled. It penalizes both misses and false alarms Range [0,1] Best value =1
	Frequency Bias (fBias)	$\frac{H+FA}{H+M}$	Measures the ratio of number of events modelled and observed. Range [0, +∞] Best value =1

Where,

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-607
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 19 February 2019

x is gauge precipitation and *y* is gridded precipitation being evaluated.

H =Hits, FA=False alarm, and M = Miss for categories under consideration

Table 2 Median values of KGE' and its components grouped for broad Koppen-Geiger climate categories for Australia. The number in parenthesis indicates the number of stations in the climatic zones. Value in bold represent the best score in each group

		Tropical (50)			Arid (125)			Temperate (266)		
		AWAP	BARRA	ERA Interim	AWAP	BARRA	ERA Interim	AWAP	BARRA	ERA Interim
Against Gauge	Correlation (R)	**0.88**	0.42	0.55	**0.91**	0.60	0.65	**0.93**	0.72	0.71
	Bias ratio (β)	1.24	1.28	**0.97**	1.22	1.35	**1.01**	1.20	1.22	**1.06**
	Variability ratio (γ)	**0.87**	1.15	0.61	**0.90**	0.85	0.86	0.91	**0.93**	0.79
	KGE'	**0.67**	0.27	0.35	**0.72**	0.41	0.53	**0.74**	0.58	0.56
Against AWAP	Correlation (R)		0.49	**0.63**		0.66	**0.67**		**0.77**	0.76
	Bias ratio (β)		**0.99**	0.76		**1.08**	0.83		**0.99**	0.82
	Variability ratio (γ)		1.35	**0.72**		**0.95**	0.94		**1.02**	0.88

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-607
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 19 February 2019

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

KGE'	0.38	**0.43**	**0.60**	**0.60**	**0.71**	0.64