Australian Government
Bureau of Meteorology

An enhanced gridded rainfall analysis scheme for Australia

Alex Evans, David Jones, Rob Smalley, Stephen Lellyett

Updated

June 2020

Bureau Research Report - 41

An enhanced gridded rainfall analysis scheme for Australia

Evans et al

Bureau Research Report No. 041

June 2020

National Library of Australia Cataloguing-in-Publication entry

Author(s): Alex Evans, David Jones, Rob Smalley, Stephen Lellyett

Title: An enhanced gridded rainfall analysis scheme for Australia

ISBN: 978-1-925738-12-4

Series: Bureau Research Report – BRR041

Enquiries should be addressed to:

Climate Services

Bureau of Meteorology
GPO Box 1289, Melbourne
Victoria 3001, Australia

helpdesk.climate@bom.gov.au

Copyright and Disclaimer

Contents

PREFACE **2**

1. INTRODUCTION **3**

2. DATA SOURCES **5**

2.1 Improved quality control of data 6

2.2 Third-party and hydrological network data 8

2.3 Extending the rainfall analyses back to 1880 9

3. GENERATING the SPATIAL ANALYSES **10**

3.1 Defining the station and gridded climate averages 12

3.2 Correlation structure and background error variance 15

3.3 Data-sparse regions and generation of pseudo-observations 17

4. ANALYSIS ACCURACY and ASSESSMENT **20**

4.1 Quality of the analyses 21

4.2 Area-average and spatial comparisons 24

4.3 Pre-1900 analyses 28

4.4 Incorporation of hydrological station network 29

5. SUMMARY and CONCLUSIONS **32**

6. REFERENCES **33**

APPENDIX 1 ..... **37**

List of Figures

Figure 1: Representative example of the mean fraction of stations reporting monthly rainfall data after the last date of the month. The fraction is relative to the number of stations reporting after two years from the observation. This is based on observation reports between June 2015 to March 2017. 5

Figure 2: Annual average count of reporting monthly rainfall stations extracted from ADAM and used here. The grey curve represents the average of all stations, Blue curve representing the operational extraction from ADAM, yellow curve represents stations that have elevation metadata, and orange curve represents the hydrological network which are station reporting for the purpose of (mainly) flood forecasting. 7

Figure 3: Timeseries of the fractional area containing one or more monthly reporting rainfall station per 0.25º grid cell for each State/Territory and the Murray-Darling Basin region (MDB) 9

Figure 4: Map showing the number of years between 1870 and 2019 for which sites report rainfall within a 0.25º grid cell. 10

Figure 5: Cross-validated RMSE (mm) for January (left) and July (right) climatologies (1981-2010) 14

Figure 6: Nationally averaged cross validated RMSE (left) and associated bias (right) (1981-2010), with units in mm, compared to monthly climatological rainfall averages. 14

Figure 7: Observation correlation coefficients (%) as a function of station separation for the month of July (1981-2010). The best fit curve is determined using a convolution filter as described in the text. 16

Figure 8: Annual cycle of the observation decorrelation length scales for monthly rainfall (1981-2010). Correlation values shown are significant at the 95% and 99% level under a two-sided t-test. 16

Figure 9: Representative network station data density grid from a January 2019 analysis. Areas of low station density (nominally less than or equal to 0.65) are used to define data voids. 18

Figure 10: Available station network indicating the count of individual sites located in 25 km × 25 km grid cell that have reported at least once in the decade 2000-2009. Black markers show representative pseudo-observations determined from each monthly analysis in the corresponding decade, as described in text. 19

Figure 11: Zero distance-crossing (r_z) correlation for each climatological month spanning the period 1981-2010. 20

Figure 12: Cross-validated root mean square error (RMSE) for monthly rainfall for the eight years 2011-2018. Left map showing RMSE from SI analysis, right map showing AWAP. Units in mm. 21

Figure 13: National annual average cross-validated RMSE (a), bias (b) and MAE (c) for monthly rainfall for the full analysis period (1900-2018). Blue curves represent results from AWAP, orange curves are results from SI analysis. Units in mm. 22

Figure 14: Cross-validated bias for monthly rainfall for the full analysis period (1900-2019) averaged across all NSW stations. Blue curves represent results from AWAP, orange curves are results from SI analysis. Units in mm. 24

Figure 15: Scatter plots of annual total rainfall from SI against annual total rainfall from AWAP, as averaged across NSW (left) and averaged across Australia (right). 25

Figure 16: Comparison of 16 month (April 2018 to July 2019) realisation of rainfall percentile analysis (Lowest on record = 0th percentile, Severe deficiency = 5th percentile, Serious deficiency = 10th percentile) from the current operational analysis (AWAP on right; 0.05º x 0.05º) and the statistical interpolation (SI on left; 0.01º x 0.01º) procedure, resampled for mapping at 0.05º. 26

Figure 17: Trends in winter rainfall as represented by SI (left) and AWAP (right) over the period 1980-2019. 27

Figure 18: Annual rainfall variability, defined as the 90th rainfall percentile minus the 10th rainfall percentile, with the result divided by the 50th percentile (or median). Data from 1900 to 2018 used in the analysis. SI analysis on left, AWAP analysis on right 27

Figure 19: Rainfall deciles for the 12-month period January-December 1888. The percentile calculation has a reference period of 1885-1965 to enable a direct comparison with Gibbs and Maher (1967). 29

Figure 20: Cross-validated bias for monthly rainfall for the analysis period (January 2000-December 2018) averaged across all stations. Blue curves represent results from AWAP, orange curves are results from SI analysis, grey curve are results from SI with the inclusion of all available hydrological stations. Units in mm. 30

Figure 21: Cross-validated root mean square error (RMSE) for monthly rainfall for the eight years 2011-2018. SI analysis on left (as in figure 12, left) and SI analysis with the inclusion of hydrological locations (right). Units in mm. 31

PREFACE

This report documents a new set of monthly rainfall (climate) analyses for Australia. The analyses are updated in real-time and extend as far back as the late nineteenth century. These analyses improve the spatial representation of Australian rainfall and drought as part of the "Enhanced weather information for farm drought resilience in NSW" initiative. Funding for this work has come from the New South Wales (NSW) government under the 2015 Drought Strategy, administered through the NSW Rural Assistance Authority (RAA).

Gridded rainfall data inform decision-making in response to drought. Specifically, these data are used to highlight regions which are adversely affected by rainfall deficiencies to better target relief measures and plan drought related response. The underlying driver of this project was to improve rainfall analyses across NSW, supporting better targeting of the Australian Government Department of Agriculture's Drought Concessional Loans Scheme (DCLS). Because many farmers and rural groups hold their own rainfall data, it was also agreed to develop third party data specifications and back-end IT infrastructure to enable the inclusion of qualifying third-party data into future analyses.

Although the immediate beneficiary is the funding provider, the NSW RAA, government agencies (Federal, State and Territories) and, more generally, all users of monthly rainfall information across the entire country will benefit from the improved analyses.

Overall, implementation of the new analysis scheme has seen an increase in spatial resolution of monthly rainfall analyses from 5×5 km grids to 1×1 km grids, substantial reduction in interpolation errors and bias, and a new capacity for inclusion of third-party data. Furthermore, early testing has indicated the method is appropriate for extension to other climate variables such as temperature and vapour pressure, and analyses at the daily timescale. These extensions will be pursued when the opportunity arises.

1. INTRODUCTION

Recognising that managing climate variability can pose significant challenges for rural businesses and communities, the Australian Government and States and Territories have developed a range of policies to assist those affected to better manage drought risk and alleviate financial burdens and environmental impacts. With Australian rainfall being highly variable, a question often arising in administering such policies is: "which areas are most affected by poor rainfall conditions causing drought, and how severe are the conditions?" Descriptions like these are often used as input into drought assessment in Australia (Gibbs and Maher 1967). As such, policy outcomes and community response are sensitive to the accuracy of rainfall analyses and their granularity.

In this context, with a policy that used Bureau rainfall data to determine eligibility for financial assistance, the NSW government supported the Bureau to improve rainfall data and analyses. Two pathways were chosen to achieve this:

- The strengthening and upgrading of the weather station network over central and western New South Wales. We have used the data coming from this activity, but do not specifically discuss the details here.
- Providing an enhanced spatial rainfall analysis to better determine the severity of dry periods and the spread of drought across New South Wales and enabling the incorporation of third-party data.

It is important to note that once the physical rainfall network reaches a modest station density there are diminishing returns on investment, at which point investment might be better directed toward improving analyses and leveraging third-party or non-traditional (e.g., satellite) data. As is shown in this report, the length scales for monthly rainfall are such that data can be interpolated over significant distances, but to achieve substantially smaller errors would require unrealistic high station densities. These issues were a key consideration in this work, including ensuring that the systems developed could be applied in the future to alternative observations data, such as those coming from satellites or radar.

This report focuses on the improved analysis scheme and including third-party data. Emphasis is given to providing a detailed description of the scientific basis and methods used for the new analysis scheme which analyse station observations data onto a regular grid ('gridded data').

Gridded data have now been established as the primary means for describing past and current climate over Australia particularly when examining spatial variability (e.g. Jeffrey *et al.*, 2001; Jones *et al.*, 2009). Such gridded products carry many advantages over ‘raw’ station data including continuous values at all points through time, less sensitivity to changes at stations, and ease of use across a wide range of applications because of their consistency and completeness. They are also less influenced by outages or misreporting and have been shown to generally provide a superior estimate of rainfall at the broad scale than single observations alone (Jones *et al.*, 2009). These benefits do come at some costs, including that gridded values may not capture peak values observed at the stations (Jones *et al*. 2009; King *et al.* 2013) and tend to smooth the data.

The application of gridded rainfall data to agricultural decision-making is widespread and varied, such as determining farmer pre-qualification for a range of drought relief schemes, the monitoring of extreme events, generation of web-based maps and a variety of downstream products.

The Bureau's current operational gridded rainfall analyses developed under the Australian Water Availability Project (AWAP), uses a procedure (Jones *et al*. 2009) that employs the Barnes successive-correction gridding method described in Koch *et al*. (1983) and Barnes (1964) combined with a three-dimensional analysis of the climatology of rainfall (Hutchinson 1995). The generation of the climatologies employs a trivariate spline function with the incorporation a continuous spatially varying dependence on elevation, defined with the use of a digital elevation model (DEM) (Hutchinson 1998a). With position coordinates defined in degrees of longitude and latitude, and elevations scaled in units of kilometres above sea-level, this approach has been shown to greatly improve analysis accuracy compared to methods which do not consider altitudinal effect (Hutchinson 1995).

The AWAP approach is based on considering rainfall as a combination of the climatological average and the rainfall ratio (rainfall as a fraction of the climatology). The rainfall ratio for each grid point is calculated from a weighted sum of the rainfall ratio at all rainfall stations within a given search radius of the point, across several passes using different length scales. The weighting given to each station for each pass is a function of only the distance between the station and the grid point (Jones and Trewin, 2000). The weighting applied to stations is optimised in so far as the parameters which define the weighting function are tuned to give the most accurate final analysis. However, this approach may not necessarily reflect the exact characteristics of rainfall events and can fail to account for station covariability by giving too much weight to some stations and too little to others in some cases.

Whilst AWAP remains the Bureau's main operational product for the monitoring of climate across many time scales and several variables, there are known interpolation errors and under-representations of extremes, particularly for rainfall, that account for biases and errors in the product (e.g.: Beesley *et al*. 2008; Tozer *et al*. 2010; King *et al*, 2012; Chubb *et al.* 2016). Indeed, Jones *et al.* (2009) note that analysis errors were only weakly dependent on the optimisation of Barnes parameters, and more likely an indication of the markedly varying length scales for rainfall across time and space, making rainfall analysis difficult (e.g., localised intense precipitation often associated with thunderstorms as compared to wider-spread rainfall from synoptic systems such as cyclones and fronts). Moreover, the product is also no longer state-of-the-art in the sense that that re-calculation of the entire historical dataset to allow refinement given new data and quality control corrections is very computationally expensive, and out of step with the very best practices used in the international meteorological community.

We reviewed gridded dataset generation procedures with the aim of facilitating current and future developments that may include:

- Increased accuracy of national gridded products and support for numerous existing downstream applications;

- Higher spatial resolution, more accurate quantification of errors and the potential to apply different error characteristics to different data sets;

- Inclusion of third-party data sources that comply with the Bureau’s minimum quality standards (see Section 2.2);

- Incorporation of rainfall estimates over data-sparse areas using remotely sensed and earth system model information; and

- Improved presentation of rainfall information and analyses, including greater data accessibility.

Specifically, improvements were sought to the underlying gridded analysis of rainfall, through the use of new and improved statistical analysis techniques. This supports the provision of better information in monitoring and forecasting systems used in agricultural decision-making and across a wide range of other applications. In particular, we demonstrate the applicability of Statistical Interpolation (SI) techniques toward fulfilling the above aims.

2. DATA SOURCES

The data used in this work are obtained from the Australian Data Archive for Meteorology (ADAM) which is Australia's national climate database (Australian Bureau of Meteorology, 2010). The maintenance of climate data held within ADAM is a core activity of the Bureau of Meteorology and underpins many of the services provided by the organisation including analysis, monitoring, forecasting and verification.

The rainfall analyses reported here are generated using observed rainfall data contained in this database. While ADAM is updated in real-time, there is significant non-real-time inputs for some meteorological variables, particularly for total monthly rainfall observations. This is because site-based rainfall data reports have varying delays.

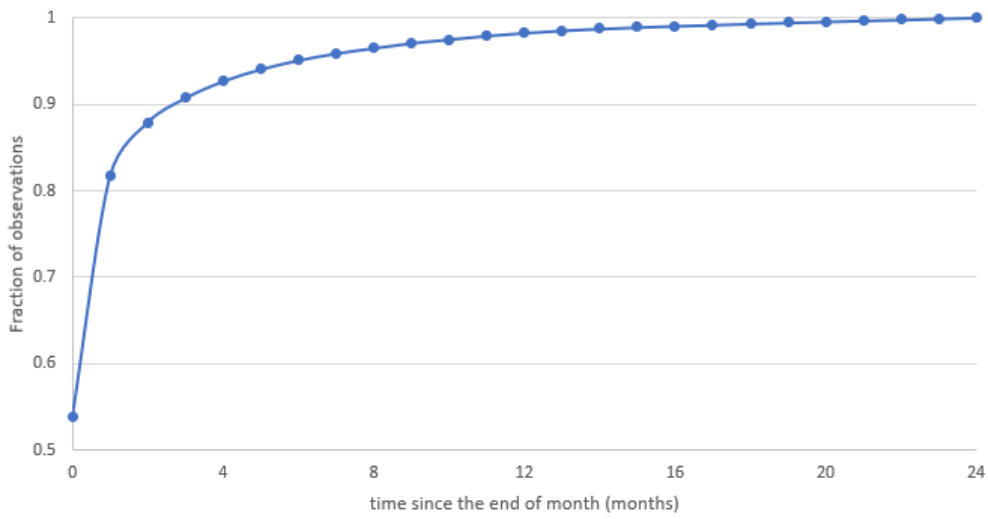

Figure 1: Representative example of the mean fraction of stations reporting monthly rainfall data after the last date of the month. The fraction is relative to the number of stations reporting after two years from the observation. This is based on observation reports between June 2015 to March 2017.

Just over 50% of the current rainfall network in ADAM reports in real-time using electronic means, whilst the remaining data arrive via mail or post with most records added within three months of the end of the month (Figure 1). After 30 days about 80% of the network has

reported, and about 97% of the network is available after one year. As the ADAM database is an evolving resource, with the addition of new and historical data and ongoing quality control, improvements to the station data and any subsequent analyses are frequently made over time. Additional data coming from 8 automatic weather stations (AWSs) and the 20 rainfall gauges newly installed by the NSW government funded project, referred to in the Introduction, have been incorporated into the new analysis.

2.1 Improved quality control of data

While every effort is taken by the Bureau to accurately record rainfall, it is inevitable that some data will have "errors". In this context “error” has a specific meaning, in that the data is not representative of an observable truth. Common factors which contribute to inaccurate rainfall recordings include communication outages or observers being away and hence missing readings, gauge blockages or damage, misreporting either electronically or through hard copy forms, or stations which are not ideally sited (such as too close to trees). Some of the data issues are less about erroneous readings, and more about practices not matching how data should be captured. As an example, it has been quite common for some manual stations not to report rainfall on Sundays (and sometimes Saturdays) which means that the reporting on a Monday is often a two- or three-day rainfall accumulation that may have either been recorded correctly as a multi-day accumulation, or at times, incorrectly as a single day total. This issue is particularly pronounced when the Sunday or Monday occurs in a new month, which may mean the rainfall could be counted against the 'wrong month' or otherwise not able to be used.

Alongside improvements to the rainfall analysis scheme, improvements have also been made to rainfall quality control outside of this project. One significant change has been the systematic application of both manual and automated quality control which is applied to the Bureau's station climate database in ADAM. The approach taken to quality assure data in ADAM has improved through time, with greater automation applied to both historical and new recordings. This quality assurance process is applied to all data prior to use in the analysis process.

Quality control of data in ADAM is multifaceted, with initial gross error checks on incoming data, combined with variable dependent checks looking at data and physical consistency. The methods applied to quality control have tended to improve with time, with growing use of automated methods. Automated data are largely quality controlled in real-time (i.e., with minimal delay) while manually observed data which require digitisation are mainly quality controlled at the time of entry.

In addition to quality assurance on ADAM, the statistical interpolation (SI) procedure supports the detection of likely erroneous data, which are removed through an integrated cross-validation procedure prior to the final rainfall analysis being formed. This is a strength of this new method, as it provides an independent estimate at every station, which includes an interpolated value and the expected station error for each individual monthly value. When a station departs significantly from the interpolated value, it is possible to provide a detailed probabilistic assessment of whether the input station value is likely to be accurate or not. Hence the incorporation of quality control within the scheme is comprehensive, and one of the factors for the SI method producing superior results compared to AWAP.

The implementation of the SI analysis method follows that described in Blomley et al. (1989). This applies a complete data check to ensure that flagged and subsequent removed data are not

negatively influencing the resulting analyses. This is efficiently applied so that when multiple sectors of the full domain are analysed separately, they correctly incorporate the appropriate nearby site data. Multiple (nested) sectors are used, as the spatial domain can only be as large as the dimension of the correlation matrix (set by the maximum computational matrix array size), and the spatial domain (sector) depends on the number of sites in an analysed region.

The removal of suspect data involves a comparison between cross validated interpolated values and observed values, weighted by the observational error variance. The criteria has been set conservatively to ensure only a small percentage of data are rejected (typically around 1% or less of the total network).

We have looked to enhance the use of historical rainfall data which previously were not fully incorporated in the AWAP dataset. One area of significant improvement has been through the updating of stations that had previously not reported a siting altitude. This is achieved by making use of a DEM (BAP 2014) to provide a siting elevation estimate. This is particularly important for defining the climate averages, either for determining or applying the climatology as described in Section 3. Many stations, especially prior to about the 1960s, lack this level of detail (see Figure 2). However, using station position coordinates, we have enabled the inclusion of an additional 1160 stations for NSW, and more than 3800 stations nationally over the full period. Figure 2 shows the average annual count of monthly reporting rainfall stations used in the analysis and highlights the size of the network, throughout time, with and without stations that have reported elevation details.

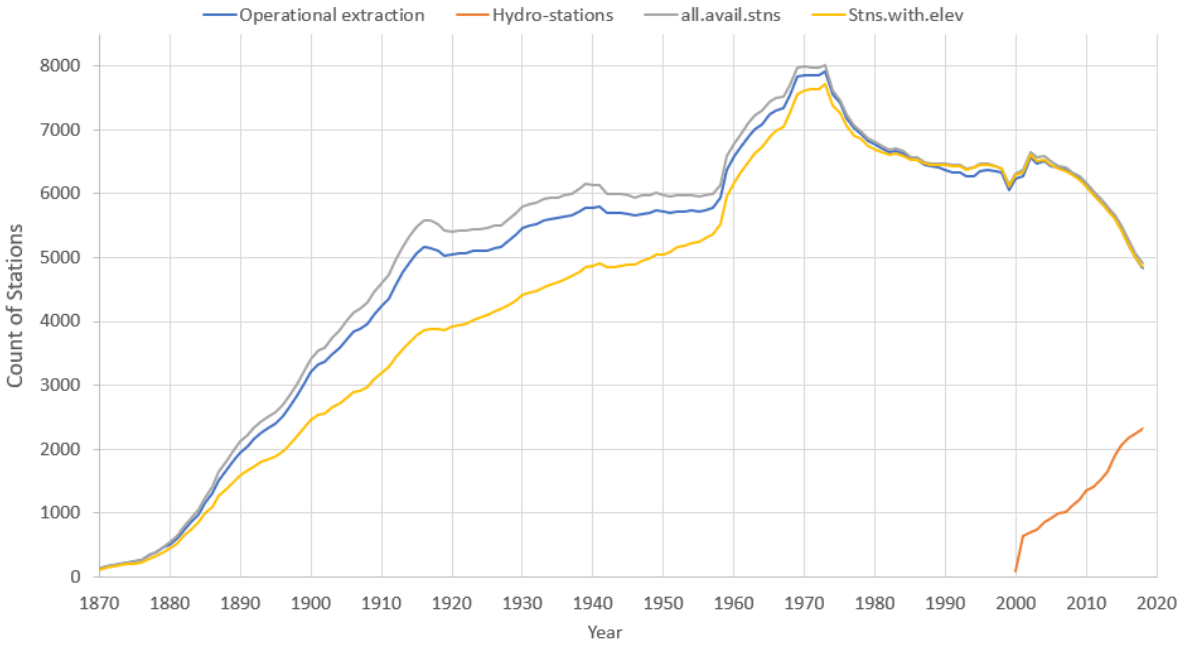

Figure 2: Annual average count of reporting monthly rainfall stations extracted from ADAM and used here. The grey curve represents the average of all stations, Blue curve representing the operational extraction from ADAM, yellow curve represents stations that have elevation metadata, and orange curve represents the hydrological network which are station reporting for the purpose of (mainly) flood forecasting.

2.2 Third-party and hydrological network data

The project has developed new specifications for third-party data to allow the future inclusion of non-Bureau data sources. The specifications correspond to the third tier of the Bureau's observational network, and whilst not appropriate for operational aviation and warning types of applications (the domain of tiers one and two), they are appropriate – subject to compliance and final data quality - for use in a range of other applications and Bureau services including rainfall analyses (e.g., as discussed by King *et al.* 2013). In particular, the developed specifications are for Equipment, Metadata, Siting and Maintenance. The equipment specification covers all standard AWS parameters. The other specifications only cover rainfall, but the intention is to generalise these to all standard AWS parameters post-project.

Whilst the ingestion methodology for third-party rainfall developed under the project has been generalised, as a starting point at the time of writing, approximately 176 stations from the OzForecast network (ozforecast.com.au) were under assessment for potential inclusion in the analysis. A desktop assessment against the Equipment Specification has indicated that if the siting, metadata and maintenance specifications are met, these data could be included in future versions of the SI rainfall analyses.

Given their ready availability and extremely successful operational use in flood warning applications, experiments were undertaken to test the suitability of the hydrology site network, comprising around 3500 stations nationally and 830 stations in NSW. No screening against the Tier 3 Specifications and no additional quality control was undertaken in this process. It was found that overall the use of the data degrades the quality of the rainfall analyses, even though the number of stations was substantially increased. Further investigation is required, but it appears that in part at least this outcome could be the result of non-standard equipment setups relative to the Tier 3 Specifications. An example is rain gauges sited at non-standard heights which tend to give observations which are biased low. To resolve the reasons why, and also extract maximum benefit from hydrology site network, it is intended to undertake a detailed evaluation against the Tier 3 Specifications, from which it is believed will emerge a significant number of suitable, conforming stations for inclusion in the analyses. This brings into sharp focus the importance of conformance of third-party data to the Tier 3 Specifications.

Beyond OzForecast and the hydrology site networks, there are a range of other third-party networks that could potentially total more than 1000 additional stations nation-wide. Incorporation of such networks will be pursued as time and resources allow, and at some point it will be opened up to potential individual suppliers of third-party data, however, in all cases the Tier 3 Specifications will apply.

In a parallel tender process the Bureau approached the equipment market for supply of an all-in-one AWS. The tender has been awarded, subject to extensive testing, which is now underway. This all-in-one unit is expected to conform to the Tier 3 Specifications, and may help in the expansion of networks in the future.

The longer term vision is to have qualifying third-party data across a range of parameters included in Bureau analyses and other products, plus, to have on offer to potential third-party data suppliers for either the Tier 3 Specifications which they could use to go to market, or details of the Bureau approved all-in-one.

2.3 Extending the rainfall analyses back to 1880

The Australian rainfall network from official Bureau sites is particularly sparse prior to 1900, with large data voids across much of the central and western parts of the country. This sparsity is even greater if stations are required to have a recorded elevation as was the case in Jones *et al.* (2009). The number of contributing stations to a given month's national analysis varies from less than 3000 in the last couple of decades of the nineteenth century to a peak of more than 7000 in the late twentieth century. It is for this reason that the AWAP rainfall analyses commenced in January 1900 (Jones and Weymouth 1997; Jones *et al.* 2009).

In theory an analysis can be constructed with very few stations, but this will result in a product which has very large uncertainties and may not be useful or, at worst, misleading. From 1870 to 2010, we analysed the number of stations reporting monthly rainfall per 0.25º (approximately 25km) grid cell. From this, the fractional area of each State and Territory with one or more station per 0.25º grid cell was computed. Results shown in Figure 3 are generally consistent with the analysis of Jones and Weymouth (1997) where they show the change in the *mean* number of stations reporting monthly data by decade.

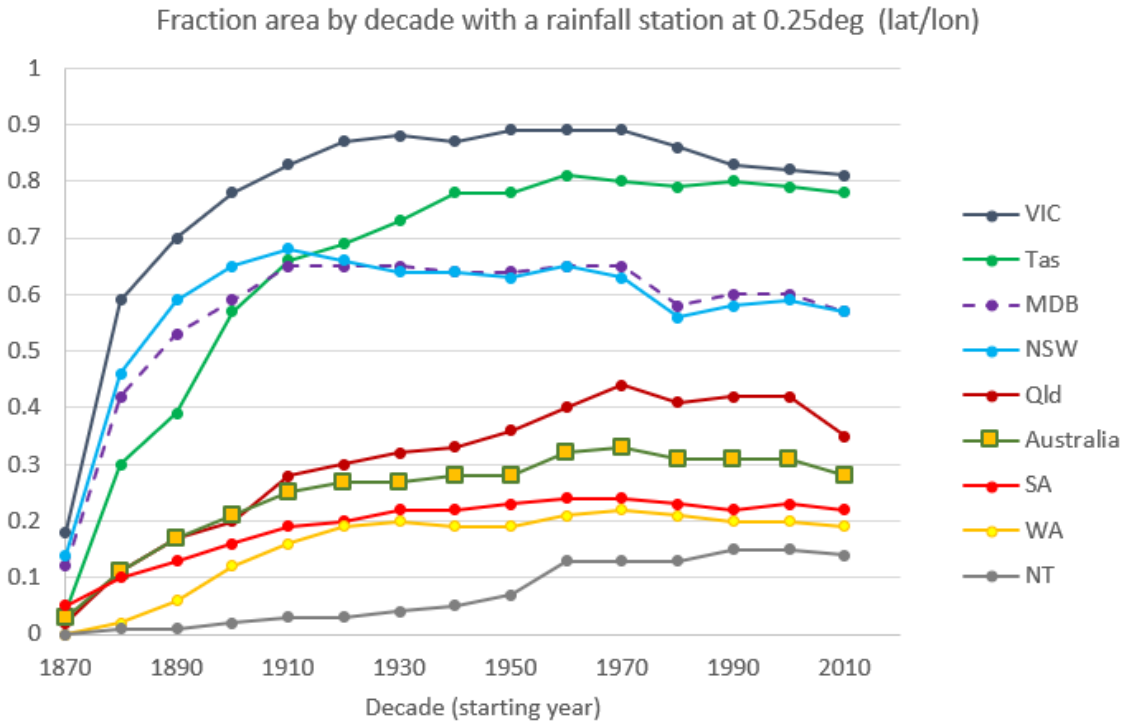

Figure 3: Timeseries of the fractional area containing one or more monthly reporting rainfall station per 0.25º grid cell for each State/Territory and the Murray-Darling Basin region (MDB)

For the Australian region as a whole, the fractional area that contains at least one monthly reporting rainfall station has remained above 20% since 1900 which supports the conclusion that rainfall analyses back to 1900 have a fairly consistent supporting network. Across much of the southeast of Australia there is extensive coverage of sites that have reported rainfall, within 0.25º grid cell, for more than 100 years (Figure 4). It is interesting that the larger states have little change in data coverage from around 1930, highlighting the difficulties of large network expansions. For much of Australia, the ability to observe rainfall is limited by the low

population density and the lack of suitable infrastructure, meaning that future large improvements in networks are unlikely.

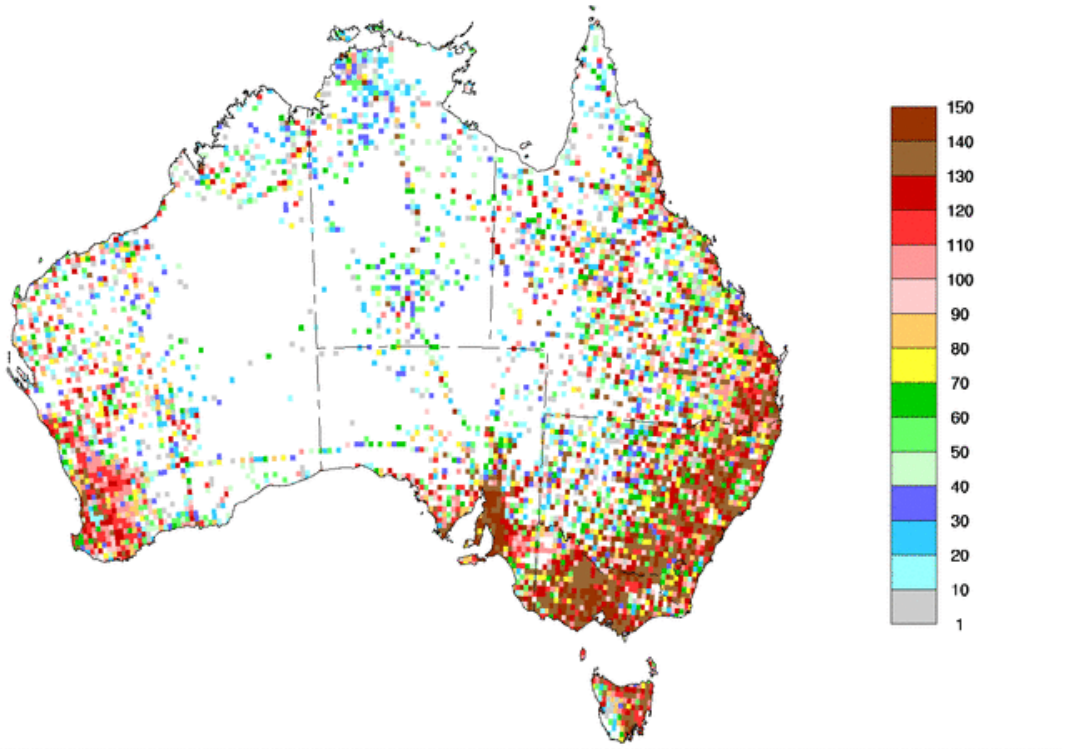

Figure 4: Map showing the number of years between 1870 and 2019 for which sites report rainfall within a 0.25º grid cell.

Across southeast Australia (Victoria, Tasmania and New South Wales) the spatial distribution of stations has extended across non-negligible fractional areas since approximately 1880. This distribution has been relatively stable across time since around 1890 for New South Wales and since 1900 for Victoria. For Australia the fractional area is approximately stable since 1910. With the addition of interpolated station elevations, improved statistical analysis (described in the following section), and use of pseudo-observations (see Section 3.3), it was assessed as feasible to extend the analyses as early as 1880, though noting the network is clearly sparser. The resultant analyses indeed appear realistic and are consistent with other information available from the time.

3. GENERATING the SPATIAL ANALYSES

As already mentioned, AWAP makes use of the Barnes successive-correction procedure. However, in the case of the Australian rainfall network, where large areas are poorly observed, the fixed length-scales underlying the AWAP scheme can often result in under-represented and poorly resolved gridded analyses. We also note that the method applied a weight to stations which does not take into account the common information (covariability) in station data. As an example, two stations sitting at a similar distance from an interpolated grid point will get the same weight whether they are close together (i.e., strongly correlated) or in different directions, which can introduce bias in the end result.

In this work, we utilise a method of gridded analysis known as Statistical Interpolation (Gandin 1963; Lorenc 1981). Statistical Interpolation (SI) is a widely used technique for the objective analysis of meteorological and atmospheric data (Daley 1993; Glowacki *et al.* 2012) and is a simplified form of least squares regression applied in two dimensions. Often referred to as *optimal interpolation*, the technique yields grid-point values from a weighted sum of observation increments to a background field. While for equally weighted stations at similar distances from the interpolated point, SI yields similar results to Barnes, we take advantage of the SI technique to avoid oscillation problems inherent in successive correction methods and make use of more explicit means of using data of differing quality. SI has the additional strength of allowing more precise representation of the rainfall fields correlation structure.

SI is closely related to Kriging, in that it defines a set of *optimal* weights that minimises the estimated analysis error variance (Daley 1993). For the sake of brevity, the reader is referred to existing literature on the full formulation of the SI technique and its various applications (e.g.: Lorenc 1981; Glowacki and Seaman 1987, Blomley *et al.* 1989; Daley 1993; Glowacki *et al.* 2012).

Here we restrict discussion of SI to application of monthly rainfall across Australia. Nevertheless, we present a brief summary on the mathematical basis of SI, closely following that presented in Daley (1993). The estimates of grid-point values (A_i) are determined from the weighted sum of observation increments to background values (B_i), this is written as

$$A_i = B_i + \sum_{k=1}^{K} W_{ik}[O_k - B_k] \quad (1)$$

Here, $[O_k - B_k]$ are the observation increments at observation points (k). W_{ik} is a weight function given to each observation increment at each analysis grid point, obtained from the column vector $\underline{W_i}$ which are called the '*optimum weights*', given by

$$\underline{W_i} = [\underline{\underline{B}} + \underline{\underline{O}}]^{-1} \underline{B_i} \quad (2)$$

$\underline{\underline{B}}$ and $\underline{\underline{O}}$ are the background and observation error covariance matrices, $\underline{B_i}$ is a column vector containing the background error correlation between observations and analysis points. For the most part, the column vector $\underline{B_i}$ and the background error covariance $\underline{\underline{B}}$ form the most important elements in SI and determine the resulting analysis (Daley 1993).

As with the AWAP monthly rainfall analyses, a ratio-based approach is used in the generation of the gridded spatial analyses, except since SI necessarily requires a background estimate B_i, we set $B_i = 0$ and define the anomaly field $R'(t)$ to be zero-mean centred, as

$$R'(t) = \frac{R(t)-[R]}{[R]} = \frac{R(t)}{[R]} - 1 \quad (3)$$

The background estimate, commonly called the "first-guess" background field in meteorological analysis is a field against which observational increments and the associate correlations structures are defined. The closer the first guess field, generally the better the analyses.

Equation (3) expresses the rainfall as an anomaly-ratio, a difference of the rainfall observation R(t) for a given month from the respective monthly climatology average [R]. This representation is applicable for both station observations and the gridded analysis and provides

an unbiased background estimate. The monthly climatological averages are generated using the smoothing thin-plate spline approach based on the work of Hutchinson (1995), details of which are given in the next section. The use of the anomaly ratio ensures that the long-term average of the increments is zero and the values are approximately mean centred, while the variance is also adjusted by the mean rainfall. Noting that when the interpolated anomaly ratios are subsequently back transformed, the rainfall field is obtained.

The approach used here assumes that removal of the climatological signal largely removes the relationship with altitude, and that the climatology is a reasonable first guess of the rainfall in a given month. One reason for applying the SI here is that it is a simple task to improve the first guess field going forwards, for example using a short-term rainfall forecasts and a satellite based gridded product, and this points one way forward for potential further development of the analysis system in future.

3.1 Defining the station and gridded climate averages

Climate averages are calculated for each calendar month for the period 1981-2010. This represents a period when both station density and observation practices are representative of the climate in most recent decades. Figures 2 and 3 also show that the rainfall network has remained more consistent across this period as compared to the earlier period in the record. However, in practice, many stations do not have complete data across this 30-year period due to missed observations and some stations having opened or closed during these 30 years.

Hopkinson et al. 2012 show that, with regression, 30-year mean estimates can be determined for stations with shorter observation records from nearby stations with longer term averages, however for this work, we have only made use of stations that have a minimum of 21 valid months of data for each calendar month during the 1981-2010 period (see section 5.2.3 of WMO-No.1203). In the case for stations at higher elevations, a more relaxed rule with a minimum count of 14 valid months of data was adopted as very few stations are available to the analysis. High elevation stations (at 1000 m or higher) are important in defining the vertical gradient of climate averages and ultimately the final analysis. Jones *et al*. (2009) note that the use of additional high elevation stations in the generation of climate averages made modest but important improvement in the representation of high elevation climate in their final analysis.

The climatology period used in Jones *et al.* (2009) was separated over three different epochs, (1911-1940; 1941-1970; 1971-2000), with the earlier and latter periods applied to data earlier than 1911 and later than 2000 respectively. For analysis of remaining years, the climatology epoch applicable to that period's climatology was used. This contrasts with that used here, which makes use of only the 1981-2010 period. In theory an evolving climatology (1911-1940, 1941-1970, etc) may appear superior as it means that the first guess field is unbiased and closer to the target analysis at each time. However, in practice it was found that changes in stations through time lead to small but detectable drifts in the climatology and resultant analyses and comes at significant computational cost.

The station climate averages are used to generate the gridded climate averages to a resolution of $0.01^{o} \times 0.01^{o}$ (approximately 1 km × 1 km) using the thin-plate spline technique of Hutchinson (1995). Smoothing thin-plate splines are particularly useful for analysing climatological surfaces in three-dimensional space and have been widely used in Australia and elsewhere (e.g. Tait *et al.*, 2006; Jones *et al*. 2009) and can effectively reduce error where stations lack longer term observations (Hutchinson and Bischof 1983). The resulting spline *coefficients* are then used to subsequently provide an estimated climatological average for 'any' station that has, at

any time, reported an observation across the network. Noting substantial changes across the Australian rainfall network, the ability to use estimated climate averages from the spline coefficients across a fluctuating network is practical, preserves continuity across the whole historical record and ultimately results in a more stable analysis system.

Ideally, stations used in creating the climatology must have a recorded elevation value although several studies have shown that making use of generalised interpolated elevations, determined from DEM's, can improve predictive accuracy (e.g.: Hutchinson 1998b, Sharples et al. 2005, Daly *et al.* 2008). In this work we make use of the DEM to determine an elevation, only for where a known elevation estimate is not available. We have chosen not to optimise the smoothing applied to the DEM as previous studies have shown the dependency is not strong, and is also likely to be quite regime, location and seasonally dependent. Retaining the full resolutions has the additional advantage of maintaining consistency with other variables such as temperature for which the relationship with elevation is stronger and more broadly consistent.

The accuracy of the spatial climatologies has been determined using verification against station climate averages. Fully cross-validated estimates have been generated for the 30 years 1981-2010, following the method described in Jones and Trewin (2002). The cross-validation analysis was performed by randomly deleting 10% of the stations in the network, invoking the three-dimensional thin-plate spline analysis using the remaining 90% of station climatologies, and then calculating the climatological analysis errors for the omitted stations. This process was repeated 10 times for each month which then provides independent verification statistics at all stations. We note that this process gives a slightly inflated error measure, because the method itself involves degrading the data network compared to 'reality' (Jones *et al.*, 2009), but is enough for this purpose. The errors at stations are used to map analysis error and to generate all-station average errors which describe our ability to represent climate average rainfall at stations. The determination of the cross-validated analysis errors, root-mean squared error (RMSE), bias and other measures of analysis error are found in Appendix 1.

The cross-validation analyses show that there is a marked north-south gradient in RMSE for monthly climatological rainfall across Australia. In part this reflects the higher rainfall in the tropical regions in the warmer months (Figure 5, left), while the gradient flips in the cooler months (Figure 5, right). Across the tropical regions, the pattern is amplified by the tendency for rainfall to be highly convective, meaning it varies a lot in space and time, whereas in southern Australia rainfall is more often associated with larger-scale weather systems such as cyclones and fronts (Ebert *et al*., 2007; Dowdy and Catto 2017). Such results will lead to better knowledge of analysis errors for a given data smoothness and network density.

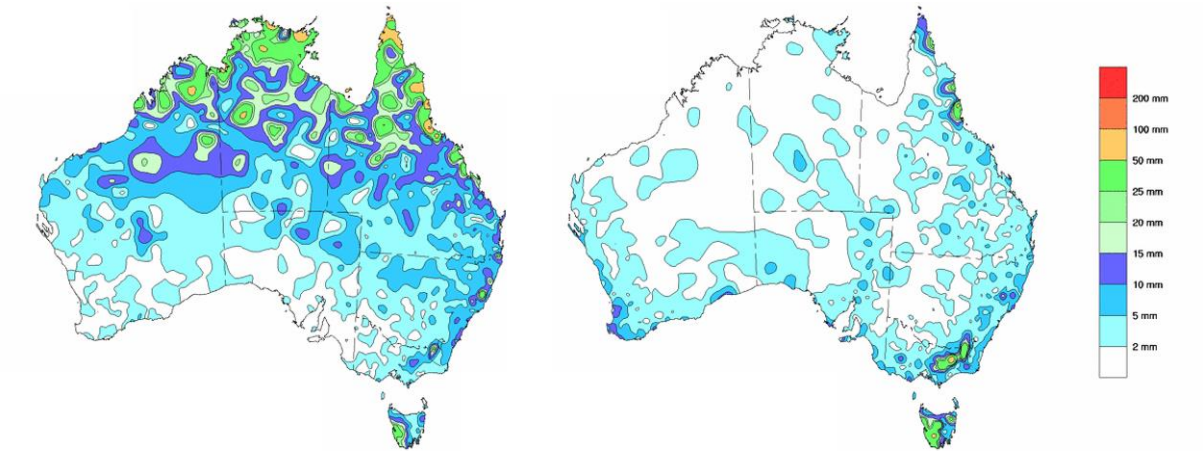

Figure 5: Cross-validated RMSE (mm) for January (left) and July (right) climatologies (1981-2010)

For NSW, climatological rainfall has higher confidence in the cool season, and in central and western areas. This, in part, reflects variations in underlying rainfall (less convection in the west and during winter), as well as the tendency for cool season rainfall to be dominated by larger scale lows and fronts. The annual variation in the long-term (climatological) nationally-averaged RMSE, and associated bias for each calendar month shown in Figure 6 also reflect features of Australian rainfall. Larger values in the transition seasons are indicative of the broadscale variability of rainfall across the country.

Importantly, rapid and easily updateable rainfall climatologies for new stations are available via the use of the spline coefficients. The ability to generate climatologies is necessary to allow stations with short observation records and newly opened stations to be used in the analysis and the future expansion to include third-party data.

A significant departure here compared to Jones *et al.* (2009) is the use of a single fixed climate period for analysis, and the use of fully interpolated climatologies rather than a mix of station and interpolated climatologies. In theory, a time varying climatological period is preferable as it means that the station increments (departures) remain mean centred and that the first guess field is more accurate, but in practice changes in networks introduces drifts in the final dataset such that it is preferable to use a fixed climate period.

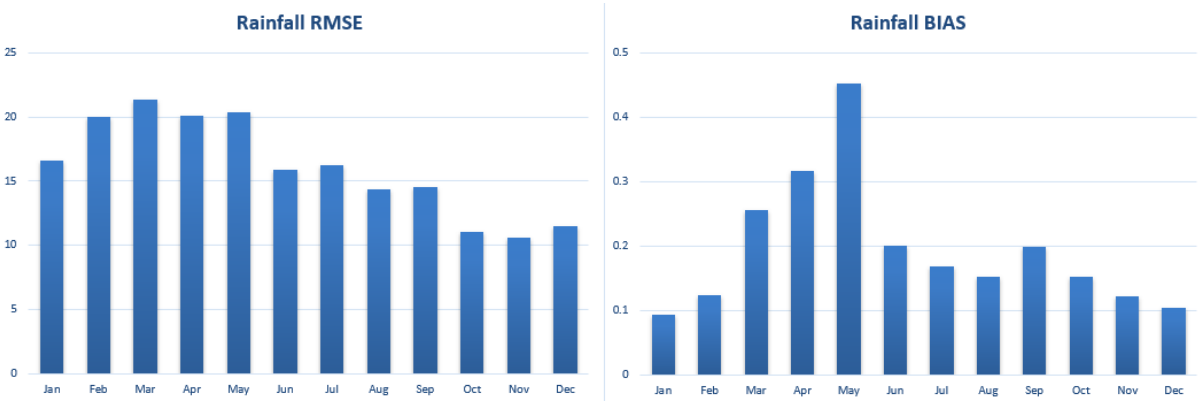

Figure 6: Nationally averaged cross validated RMSE (left) and associated bias (right) (1981-2010), with units in mm, compared to monthly climatological rainfall averages.

3.2 Correlation structure and background error variance

As noted earlier, the most important element of the SI algorithm is appropriate characterisation of the background covariance matrix and the related correlation functions. In practice, these are prespecified and there is a wealth of literature on suggested analytical forms for fitting the best correlation function to meteorological data (e.g. Julian and Theibaux 1975; Theibaux 1975).

Following examples in Jones and Trewin (2002), we examined several fitting functions, including Gaussian, exponential and polynomial damped cosine functions. Each of these correlation models were found to be somewhat unsatisfactory, and poorly represented the observed correlation for monthly rainfall. We found the best performing candidate correlation function by fitting a function using a convolution process as described by Savitzky and Golay (1964). This process takes successive sub-sets of adjacent data points and fits a low-degree polynomial by the method of linear least-squares. The *analytical* solution to the resulting least-squares equations take the form of a set of convolution coefficients that give estimates of a smoothed correlation function and derivatives of the smoothed signal (Savitzky and Golay, 1964; Steinier *et al*. 1972). It is a requirement that the candidate correlation function be positive definite (Gandin 1963, Julian and Thiebaux 1975). The Savitzky-Golay method used here specifies a weighting vector that contains real, positive-valued weights that minimise the least-squares solution which ensures the solution to the system is positive definite (Orfanidis, 1996). Interstation correlations of monthly rainfall anomaly-ratios (R') were calculated for all possible station pairings for each calendar month in the 1981-2010 base period. A representative example, Figure 7, shows the increment correlations for July monthly rainfall spanning 1981-2010, together with the least squares fit provided by the fitting of a low-degree polynomial as described above. Here, we also show a sub-sample of 'all' station pairings (of which there are more than 10 million). Each point in Figure 7 represents the correlation of a single station pair.

Our results from this show the presence of significant correlations (~40%) beyond 500 km for the July monthly rainfall. The scatter of the data points in figure 7 reflects spatial variations and anisotropies in the correlation decay scales, as well as random sampling errors. While we explored using anisotropic correlation functions, we found results to be highly variable in space and time. This presented significant challenges as we noted an improvement in error estimates in one region while leading to a deterioration in another unless multiple functions were used. Based on this work, the pragmatic decision was made to confine the analysis to an isotropic function. In future work, we will look to incorporate anisotropic correlation functions which spread increment values differently in space. Alternatively, a better background field may reduce the anisotropy.

To provide a measure of the characteristic length scales of the climate averaged monthly rainfall data, in Figure 8 we show the distance at which the best fit function falls to 0.463 for each calendar month. This defines the correlation values that are significant at the 99% level under a two-sided *t*-test for 30 years of data. While we note that this measure of length scale is somewhat arbitrary, there is sufficient indication of considerable large-scale structure in the monthly rainfall for length scales of the order of 300-500 km, significantly different to the current fixed-length scale of monthly rainfall generated by AWAP (250 km).

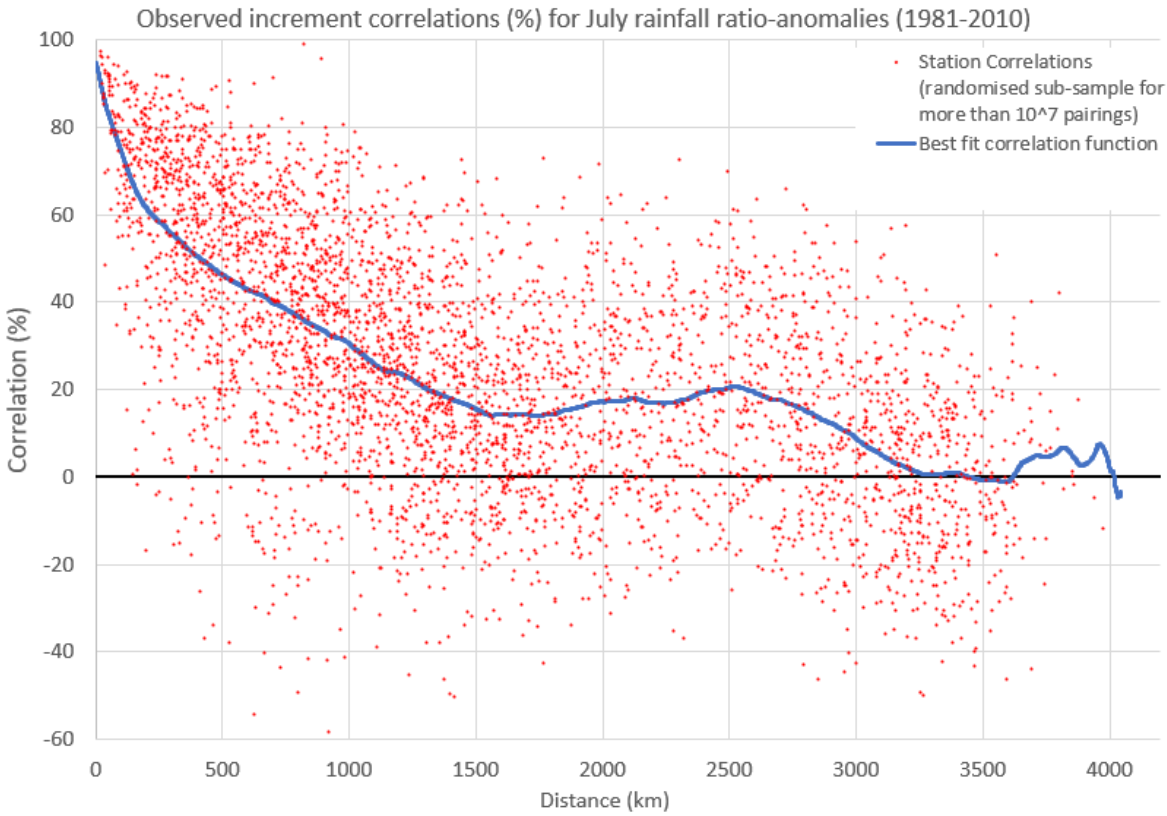

Figure 7: Observation correlation coefficients (%) as a function of station separation for the month of July (1981-2010). The best fit curve is determined using a convolution filter as described in the text.

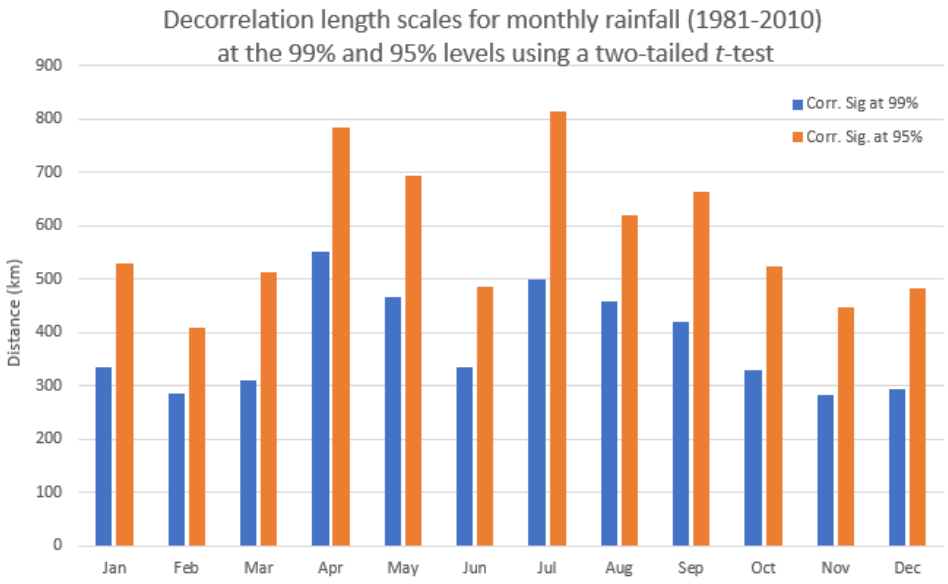

Figure 8: Annual cycle of the observation decorrelation length scales for monthly rainfall (1981-2010). Correlation values shown are significant at the 95% and 99% level under a two-sided t-test.

The decorrelation length scales highlight important and well understood features of Australian rainfall. It is evident that rainfall tends to be broader scale during the cooler months when it is more dominated by frontal systems and mid-latitude lows. There are also spatial structures in the decorrelation length scales (not shown) with smaller values in the northern tropics where convection tends to dominate, and longer values in the south where more rainfall falls from synoptic-scale weather systems. These results are interesting in that the rainfall network tends to be denser in those regions where rainfall analysis is apparently easier, while it is less dense where rainfall analysis is more challenging.

The respective monthly background error variance (BGE) are constructed by running a Barnes analysis on the standard deviation of the observation increments derived from the monthly station climate averages. Under the SI analysis scheme (Eq. 2) the BGE matrix largely determines the resulting analysis (Daley, 1993). Essentially, for a given realisation, the observation increment is spread out according to the spatial structure of the respective BGE for that month. This means that the level of accuracy is expected to be better in data dense areas and *degrade* to those defined by the BGE in areas of data-sparseness.

3.3 Data-sparse regions and generation of pseudo-observations

The general irregularity and sparseness of the observation network presents a significant challenge in generating a consistent spatial analysis of historical data across Australia. It is well known that the SI technique, and indeed other existing spatial analysis approaches, are adversely influenced by network inconsistencies (Barnes, 1964; Bratseth, 1986; Koch *et al*. 2004). This is particularly notable the further back in time we wish to analyse, where the analysis in data-sparse areas can produce spurious noise and larger extrapolation errors. This can be understood in terms of stations effectively coming in and out of the analysis at different locations across a domain (noting that weights approach zero at some distance). When a region is poorly sampled, one station dropping out can effectively lead to a jump in the analysis between neighbouring regions which may be reflected in a step or a wiggle on contours. Since the SI is undertaken in sectors (a division of the whole domain), to keep the size of the covariance matrix manageable, discontinuities between sectors become notable in areas of low network density. While such features have barely any impact on error statistics as they tend to occur in dry and low station density regions, they can leave unrealistic structure in analyses.

We formally addressed this issue by generating a regular spaced coverage of *pseudo*-observations in the very data-sparse areas, following the method of Glowacki *et al.* (2012). The term *pseudo*-observations has a long history in meteorological analyses, and refers to additional data used to infill and stabilise analyses in very poorly observed regions.

This approach is achieved by executing two SI passes, the first pass being a low-resolution analysis from which a gridded data density field is computed, and pseudo-observations determined. The network density grid is computed at a resolution of $0.25^{o}\times0.25^{o}$ for efficiency purposes. The second SI pass is computed at a resolution of $0.01^{o}\times0.01^{o}$, taking in all available observations for that given time stamp with the addition of pseudo-observations. There is little or no evident impact of the two-pass approach where the station network is reasonable (i.e. most of Australia), but it greatly improves the spatial smoothness in very data-sparse parts of the Australian interior. The process of using pseudo-observations has the effect of infilling data sparse regions with a smooth analysis and avoiding data jumps as described previously. It shares similarities with the method of *knots* which may be used in spline analyses.

From its formulation, the SI analysis gives a normalised analysis error estimate, ε_k (e.g. Blomley *et al*. 1989; Daley, 1993; Glowacki *et al*. 2012), which is used to derive the observation network data density:

$$\delta_k = 1 - (\varepsilon_k)^2 \tag{4}$$

Figure 9 shows an example of the gridded data density field. Glowacki *et al*. (2012) describe the suitability of this approach for defining poorly observed, data-sparse areas, since it is based off a measure of analysis error covariances, that will vary according to analysis variable and scales being resolved.

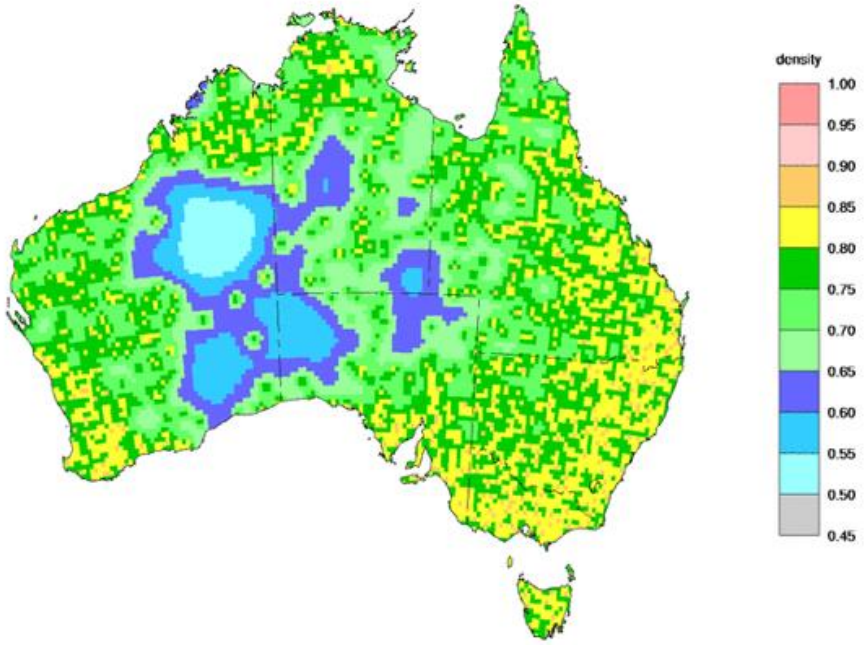

Figure 9: Representative network station data density grid from a January 2019 analysis. Areas of low station density (nominally less than or equal to 0.65) are used to define data voids.

For computational efficiency the first pass, low-resolution analysis also takes a subset of the full network by the method of super-observations, where stations in high data dense regions that are highly correlated are combined into a single observation (Glowacki and Seaman 1987). This reduces the order of the covariance matrix, speeding up the calculations with fewer sectors.

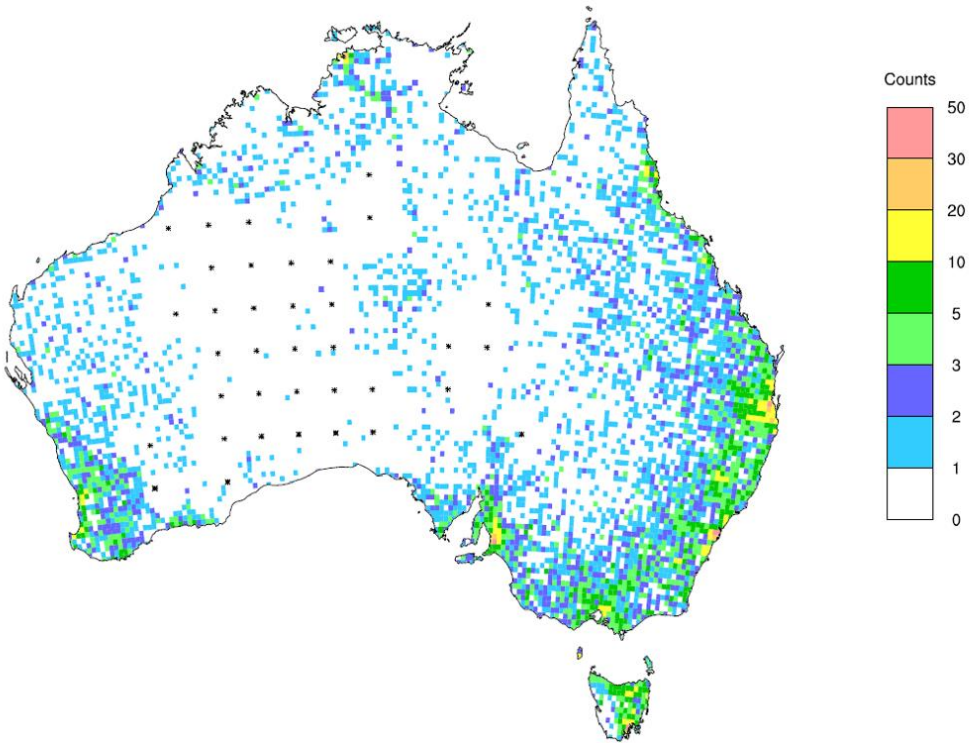

Figure 10: Available station network indicating the count of individual sites located in 25 km × 25 km grid cell that have reported at least once in the decade 2000-2009. Black markers show representative pseudo-observations determined from each monthly analysis in the corresponding decade, as described in text.

Through extensive testing, we implemented this procedure to constrain extrapolation errors and discontinuities in areas of a sparse network. However, we also limit the number of pseudo-observations used in the second SI pass so as to not overstate their influence on the final analysis. We achieve this by selecting grid points from the 0.25º data density field at minimum separation of 2º and take only those grid point values where $\delta_k \leq 0.65$. While the threshold value of network density is set somewhat arbitrarily, extensive testing showed that this threshold does not cut off network density so much so that it introduces too many pseudo-observations. Figure 10 shows an example of the spatial representation of these pseudo-observations (black markers), as determined across a 10-year period 2000-2009.

While the pseudo-observations help to constrain extrapolation errors it should be noted that as late arriving data becomes available to the analysis the network data density increases. Therefore, fewer are incorporated into subsequent re-runs for a given month's realisation. The analysis accuracy is expected to improve on re-runs, however this will mostly depend on appropriate representation of the correlations, and the observational and background variances which can be defined from $O_k - B_k$ as

$$\frac{1}{K}\sum_{k=1}^{K}\overline{(O_k - B_k)^2} = E_O^2 + E_B^2 \qquad (5)$$

Here, E_B^2 is the background error variance and E_O^2 is the observational error variance which sets the floor for analysis accuracy (Daley 1993). In fitting the correlation curve to the inter-station correlation pairs, as discussed in section 3.2 (e.g. Figure 7), it should be noted that only auto-correlations $r_{kk} = 1$, and so it follows that the candidate correlation function must be

extrapolated back to a zero distance (z) crossing. This gives a correlation value referred to as r_z which in itself is the measure of total error, and can be defined by

$$r_z = \frac{E_B^2}{E_B^2 + E_o^2} \tag{6}$$

Figure 11 shows the zero-crossing correlations (r_z) for each climatological month across the 1981-2010 period. This figure is consistent with our earlier point in that it highlights well understood features of Australian rainfall, such as the broader scale rain features during the cooler months (stronger correlations) and more convective, localised rainfall that occurs during warmer months (weaker correlations). The value of (r_z) sets a lower bound on the analysis accuracy, and highlights that even a perfect analysis system will be subject to errors due to imperfect observations.

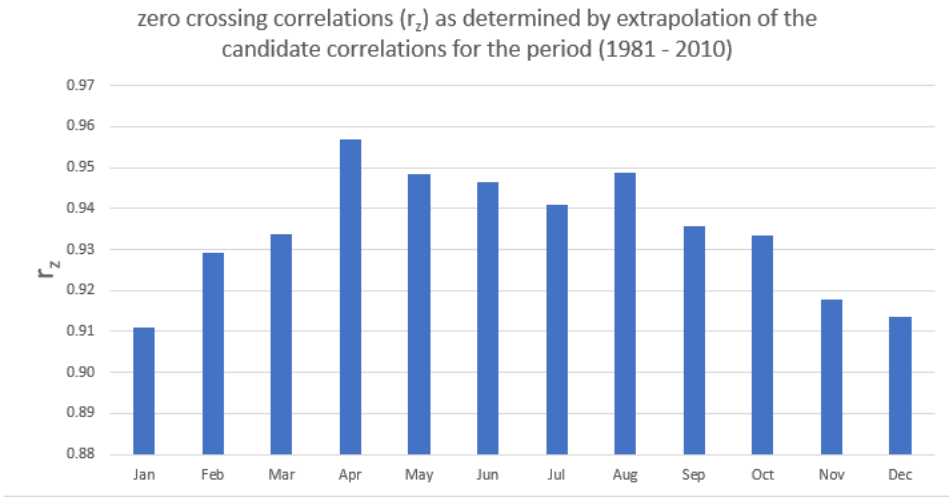

Figure 11: Zero distance-crossing (r_z) correlation for each climatological month spanning the period 1981-2010.

4. ANALYSIS ACCURACY and ASSESSMENT

The accuracy of the spatial analyses for reproducing station data has been determined through generalised cross-validation at stations. One of the benefits of the SI technique is the ability to efficiently use integrated cross-validation, by removing one station from the set of observations for that given month, performing the analysis using the remaining station observations and then calculating the analysis errors for the omitted stations. This process is repeated for each month providing fully independent verification statistics for every station.

We calculate a station root mean square analysis error (RMSE) along with additional measures of bias and mean absolute error (MAE) as defined in Appendix 1. The resulting RMSE is plotted as a function of time and we provide station averaged results, for all stations in NSW as well as all Australian averaged statistics.

4.1 Quality of the analyses

We have generated a full set of cross-validated statistics spanning the entire historical SI analyses (1900-2018). For reference, in the sections that follow, we also provide NSW average and national average statistics for the current operational AWAP product and summary statisticsof running the SI technique with the inclusion of stations from the hydrological station network mentioned in Section 2.2, above. Additional summary statistics shown in section 4.4, represent a fully independent period (2000-2019) which also allows for more direct comparison of the performance of SI with the inclusion of the hydrological network. This period represents the hydrological network at its densest (as shown in figure 2).

Spatial maps of SI analysis error (Figure 12, left) show marked reduction in RMSE across the entire continent compared to AWAP (Figure 12, right). We note that both show a north-south gradient in the rainfall RMSE, as expected given the underlying rainfall climatologies. This was pointed out in Jones *et al.* (2009), reflecting the higher rainfall totals in tropical regions, leading to larger analysis errors for the given station network. Such a pattern also represents the tendency of highly convective rainfall with shorter and more variable length scales (Mills *et al.* 1997; Ebert *et al.* 2007; Jones *et al.* 2009).

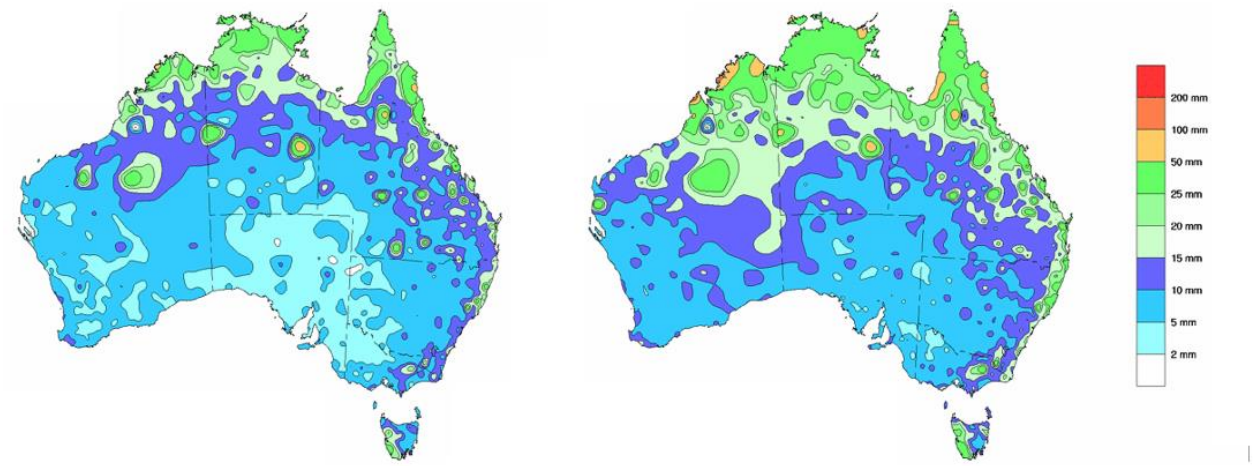

Figure 12: Cross-validated root mean square error (RMSE) for monthly rainfall for the eight years 2011-2018. Left map showing RMSE from SI analysis, right map showing AWAP. Units in mm.

Figure 13 shows a year-by-year comparison of the annual average RMSE (a), bias (b) and MAE (c). These are produced by taking the annual average across each calendar month for all stations across the continent as a whole. It is clear that the SI analysis shows significant improvement across all these measures when compared to AWAP.

The time-series of bias is particularly interesting, with SI bias being small in magnitude, slightly negative and relatively stationary over time, whereas the bias in AWAP exhibits higher magnitude, and is mostly positive, with a large degree of variability. Equations (7) and (13) (Appendix 1) indicate that positive bias at a station results when the analysed value is higher than the observed value. The variation shown in the AWAP bias appears to arise from the fact the AWAP analysis is based upon a 3-epoch, rolling climatology (see Jones *et al*. 2009 for details). This means that while the climatological analysis in each of the epochs are *mean*-centred, they tend to influence analysis outside their respective periods thereby incurring larger bias.

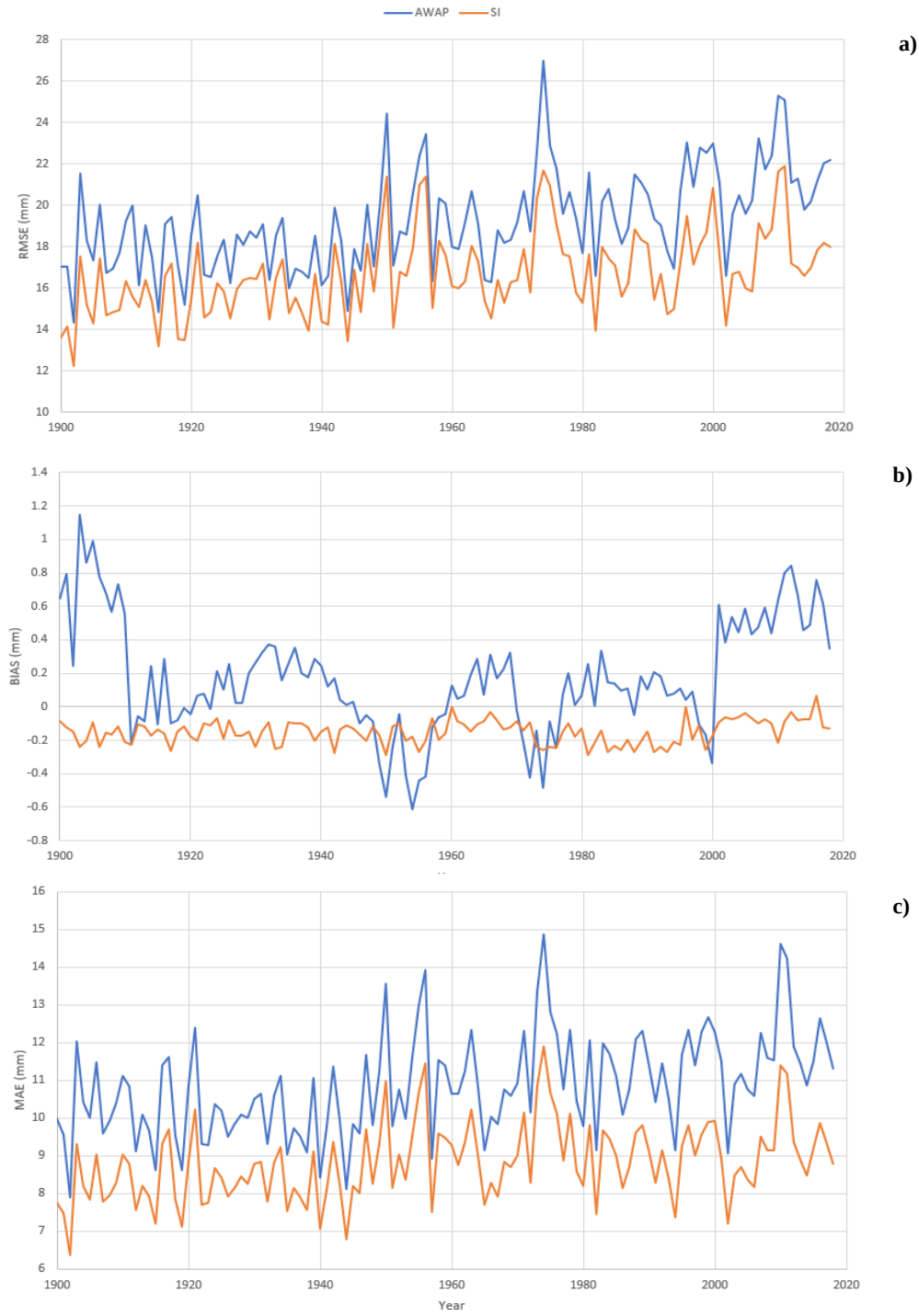

Figure 13: National annual average cross-validated RMSE (a), bias (b) and MAE (c) for monthly rainfall for the full analysis period (1900-2018). Blue curves represent results from AWAP, orange curves are results from SI analysis. Units in mm.

The quality assurance using the SI algorithm is part of the new analysis, so the deletion of erroneous data outliers is part of the analysis improvement. It is worth noting that the AWAP method also deploys a relatively simple quality control process to remove data outliers where those data are more than 7 standard deviations above a local mean, or the estimated percentage

probability of being correct is less than 0.1% and there are no similar outliers. When the rainfall total is < 0.1 mm, erroneous data are removed if there is no supporting nearby data, or are more than 50 mm less than the local mean.

Finally, we note that the bias and MAE which are much less sensitive to outliers show significant improvement, indicating that the improvements are quite general and not confined to large station values.

The RMSE and MAE timeseries comparison in Figure 13 (a and c), shows that the SI is able to reproduce the signal with respect to AWAP, though lower in both measures. Overall, RMSE is reduced by an average of 13.5% and MAE is reduced by an average 18.8%. There is considerable interannual variability in the RMSE and MAE plots, with the peaks being dominated by wet years and as noted in Jones *et al*. 2009, RMSE is substantially larger than the MAE as this metric weights towards large errors. This is owing to significant skewness in the distribution of rainfall errors, with a relatively small number of large errors at the wettest sites. We also assess the analysis (not shown) where rainfall at stations are less than 1 mm. For the period since 1900, the *median* absolute error from the SI is approximately 20% lower compared to AWAP. We report the median absolute error in this case, from each month, due to the skewed distribution at low rainfall totals with stations in climatological dry regions often showing zero. We note for AGCD that the rainfall analysis is not forced to zero in cases of zero or near-zero rainfall, resulting in a slight positive bias for the lowest rainfall totals.

It is interesting that while the rainfall network has improved somewhat through time, the analysis errors tend to slightly increase over the ~120 year period. There are a number of likely reasons for this trend. The expansion of the rainfall network has tended to be greater in the more convective (tropical and inland) parts of Australia where rainfall is harder to analyse and errors are greater. In addition, Australian area-averaged rainfall has shown an increase, particularly since around 1970 driven by positive rainfall trends in inland and northern parts of the country. Finally, there is an indication that large scale rainfall driven out of the middle latitudes, and mainly during the cool season has declined (Bureau of Meteorology and CSIRO 2018). The aggregate of these changes would all be expected to lead to an increase in analysis errors, even those where the input data has improved through time.

Extensive analysis on a month-by-month comparison (not shown here) reveals similar characteristics to the annual averages. For example, at the regional scale, Figure 14 shows the month-to-month bias averaged across all stations in NSW. Again, throughout the whole historical record these results exhibit a striking degree of stationarity and a very large reduction in bias when compared to AWAP.

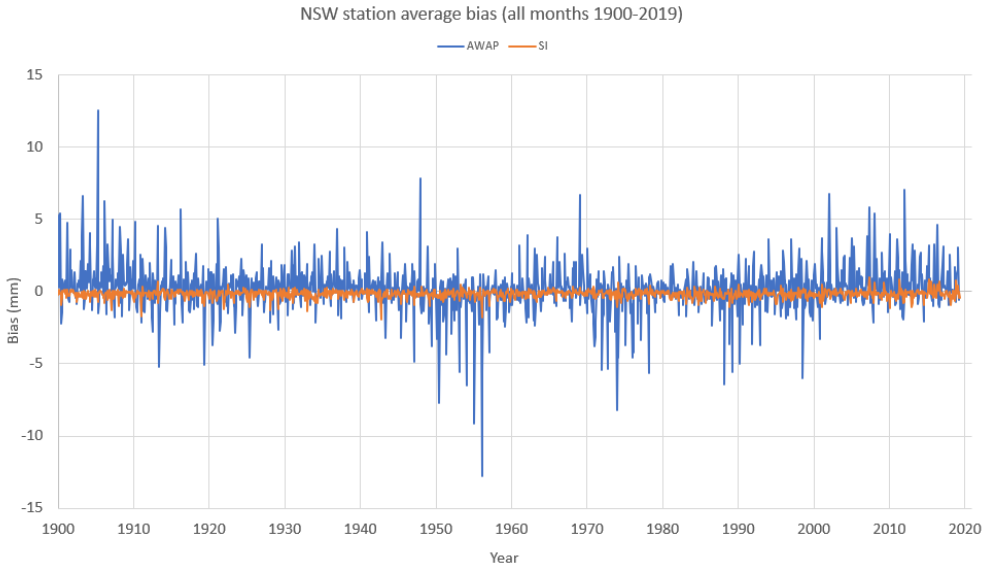

Figure 14: Cross-validated bias for monthly rainfall for the full analysis period (1900-2019) averaged across all NSW stations. Blue curves represent results from AWAP, orange curves are results from SI analysis. Units in mm.

4.2 Area-average and spatial comparisons

The obvious application of the SI analyses is for documenting the temporal and spatial representation of monthly, seasonal and annual rainfall totals in the context of past climatology. Figure 15 shows the comparison of area-averaged annual total rainfall, as generated by both SI and AWAP, for NSW (left) and Australia (right). These figures show that the area average totals generated by both methods show very good agreement. This provides some assurance that, although significant improvements have been achieved with SI, overall at the broadscale, past analyses under AWAP were indeed relatively robust spatially, and that major revisions of decisions or policy with the transition to SI are not warranted. Nevertheless, we do note that the early period analyses (looking at the first few decades in the twentieth century) shows that the Australian annual totals, as generated by SI, are slightly wetter than AWAP, then tend to be slightly drier in the final period of the analyses. In other words. AWAP appears to slightly exaggerate the wetting trend overall when compared to stations, as shown by the biases through time.

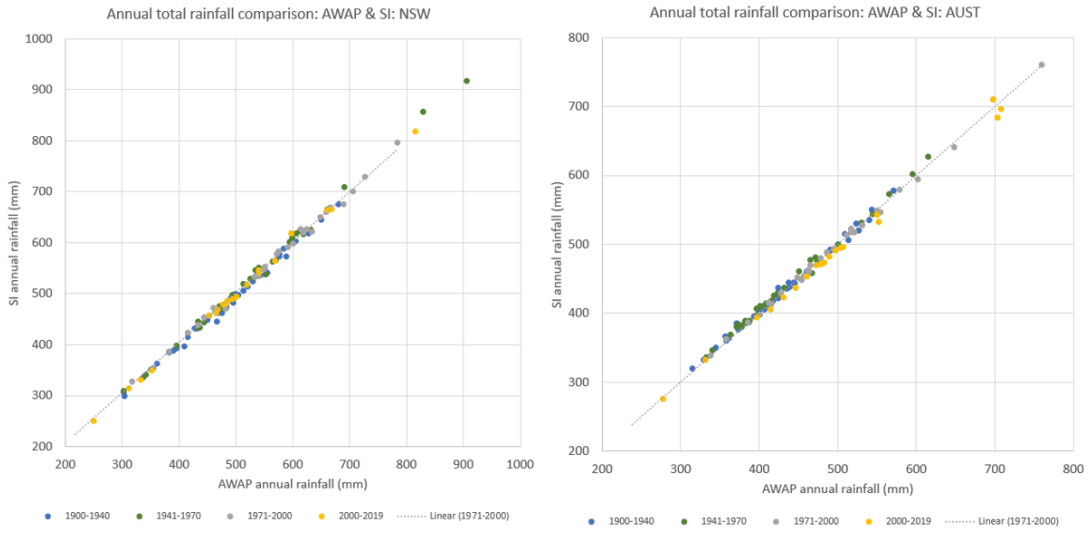

Figure 15: Scatter plots of annual total rainfall from SI against annual total rainfall from AWAP, as averaged across NSW (left) and averaged across Australia (right).

The corresponding scatter plot for NSW area-averaged annual rainfall totals indicates good agreement over time between the two datasets. Table 1 shows the ranked top 10 (wettest) and ranked bottom 10 (driest) years, for both the NSW (left) and national totals (right).

Driest 10 years (NSW)							
YYYY	awap-annual	awap-rank	YYYY	si-annual	si-rank	diff_value	diff_rank
2019	250.15	1	2019	250.19	1	0.04	0
1902	303.54	4	1902	299.4	2	-4.14	2
1940	302.9	3	1940	305.15	3	2.25	0
1944	302.68	2	1944	309.64	4	6.96	-2
2002	312.13	5	2002	313.73	5	1.6	0
1982	318.09	6	1982	327.98	6	9.89	0
2018	332.97	7	2018	331.68	7	-1.29	0
1957	336.38	8	1957	335.43	8	-0.95	0
1965	341.86	9	1965	340.91	9	-0.95	0
2006	352.66	11	2006	348.88	10	-3.78	1

Driest 10 years (AUST)							
YYYY	awap-annual	awap-rank	YYYY	si-annual	si-rank	diff_value	diff_rank
2019	277.59	1	2019	275.19	1	-2.4	0
1902	314.46	2	1902	319.52	2	5.06	0
1905	329.02	3	1905	332.52	3	3.5	0
2002	331.24	4	2002	332.54	4	1.3	0
1961	332.63	5	1961	336.14	5	3.51	0
1994	338.57	6	1994	338.68	6	0.11	0
1965	340.44	7	1965	345.72	7	5.28	0
1935	344.7	8	1935	349.7	8	5	0
1928	357.38	10	1928	360.39	9	3.01	1
1972	357.66	11	1972	362.26	10	4.6	1

Wettest 10 years (NSW)							
YYYY	awap-annual	awap-rank	YYYY	si-annual	si-rank	diff_value	diff_rank
1950	905.64	120	1950	916.51	120	10.87	0
1956	829.52	119	1956	857.36	119	27.84	0
2010	815.1	118	2010	818.33	118	3.23	0
1974	784.05	117	1974	795.58	117	11.53	0
1973	727.25	116	1973	728.71	116	1.46	0
1955	690.62	114	1955	708.73	115	18.11	-1
1978	704.91	115	1978	701.37	114	-3.54	1
1988	689.37	113	1988	676.22	113	-13.15	0
1921	679.69	112	1921	676.18	112	-3.51	0
1976	665.98	110	1976	668.57	111	2.59	-1

Wettest 10 years (AUST)							
YYYY	awap-annual	awap-rank	YYYY	si-annual	si-rank	diff_value	diff_rank
1974	759.65	120	1974	760.57	120	0.92	0
2000	698.3	117	2000	710.57	119	12.27	-2
2011	707.73	119	2011	696.7	118	-11.03	1
2010	703.36	118	2010	683.7	117	-19.66	1
1973	647.79	116	1973	641.11	116	-6.68	0
1950	615.53	115	1950	627.49	115	11.96	0
1956	594.73	113	1956	601.29	114	6.56	-1
1975	602.24	114	1975	593.78	113	-8.46	1
1999	578.27	112	1999	578.77	112	0.5	0
1917	571.04	111	1917	578.01	111	6.97	0

Table 1. Comparison of the rank changes between AWAP and SI for NSW on left and Australia on right, shown for the driest 10 years and wettest 10 years.

There are a number of possible reasons for the change over time in the differences in Australian area averages, between AWAP and SI. The reduction in the station bias using SI compared with

AWAP is one key component. However, the other may relate to the use of an anomaly-ratio for SI, compared with just a ratio (total rainfall/climatology rainfall) for AWAP. In the earlier period of record, when the station network is relatively sparse compared with present time, the background (first-guess) field used in the AWAP analysis is zero. This contrasts with SI where, due to the anomaly-ratio method, the background field, while zero, is actually climatology. In data-sparse regions (i.e. far away from observations) AWAP will have lower rainfall totals than the same region using SI, with the difference forced by the first guess field.

A key output from this work is the application of the analyses to reporting rainfall deficiencies in a drought context. Figure 16 shows a comparison of a 16-month realisation of rainfall percentile analysis from the current operational analysis (AWAP right) to the SI procedure (left) for April 2018 to July 2019. While both methods share similar structure in their spatial distribution of rainfall deficiencies, it is notable that the SI method has resulted in smoother, more contiguous analysis field. This is an interesting result as the SI data is at a higher resolution and suggests that AWAP may have contained rather more excessive local noise.

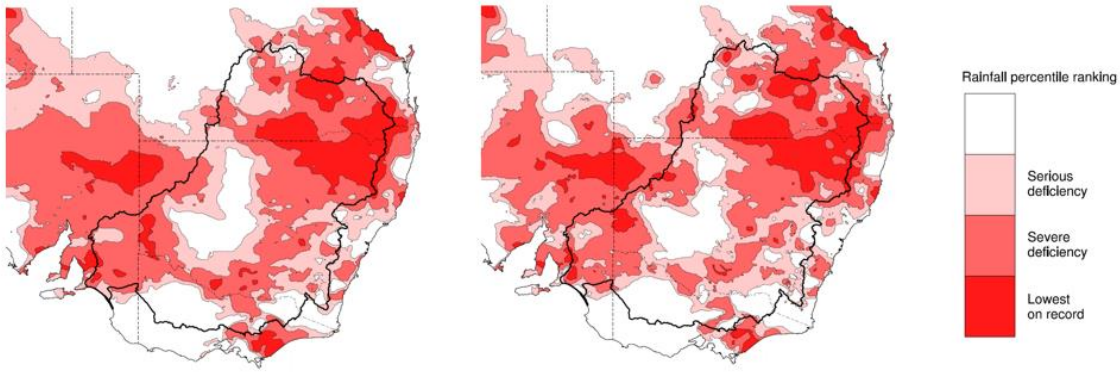

Figure 16: Comparison of 16 month (April 2018 to July 2019) realisation of rainfall percentile analysis (Lowest on record = 0th percentile, Severe deficiency = 5th percentile, Serious deficiency = 10th percentile) from the current operational analysis (AWAP on right; 0.05º x 0.05º) and the statistical interpolation (SI on left; 0.01º x 0.01º) procedure, resampled for mapping at 0.05º.

Again, similar spatial structure but improved contiguity with SI is evident when comparing SI and AWAP rainfall trends and variability. Figure 17 shows the trend in winter rainfall spanning the period 1980 to 2019 (SI on left, AWAP on right). Noting the reduction in overall analysis biases and errors, the data can be expected to more accurately capture climate trends, but as always should be used with caution with reference made back to the contributing stations as applicable.

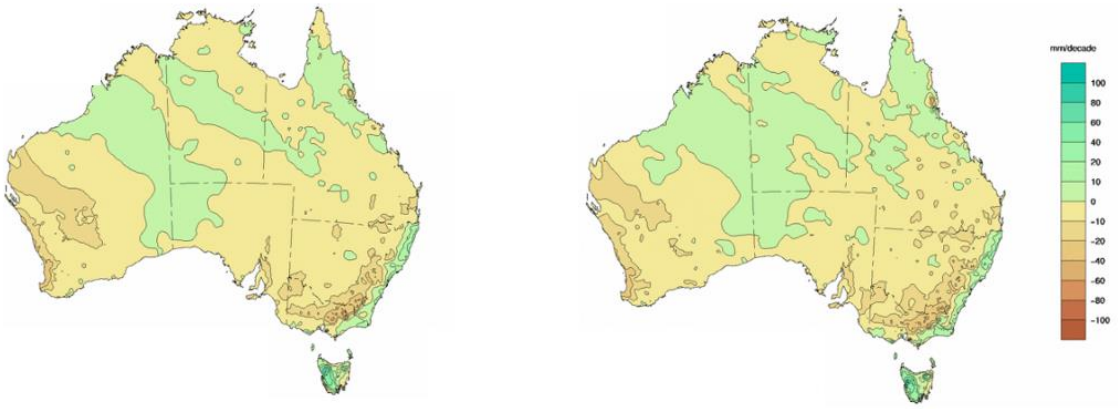

Figure 17: Trends in winter rainfall as represented by SI (left) and AWAP (right) over the period 1980-2019.

We extend the comparison and look at a measure of *variability*, which combines the 10th and 90th percentiles as well as the median (Figure 18). This measure provides an indication of the regularity of rainfall from one year to the next across Australia, but also incorporates within one figure, extreme ends of the rainfall distribution. Overall, where the data density is reasonable, both AWAP and SI have similar contours, showing broad scale agreement. The SI analyses show less noise than AWAP and overall less inter-annual variability in data-sparse regions. This may partly be attributed to AWAP reverting to zero rainfall in data-sparse regions during the early period of record (and so the 10th percentile will be closer to 0 mm), whereas SI has less spread between the 10th and 90th percentiles. It's also noticeable that the structures in the SI data are much less centred on stations over the inland (such as Giles) suggesting they give a better overall analysis consistent with the verification results.

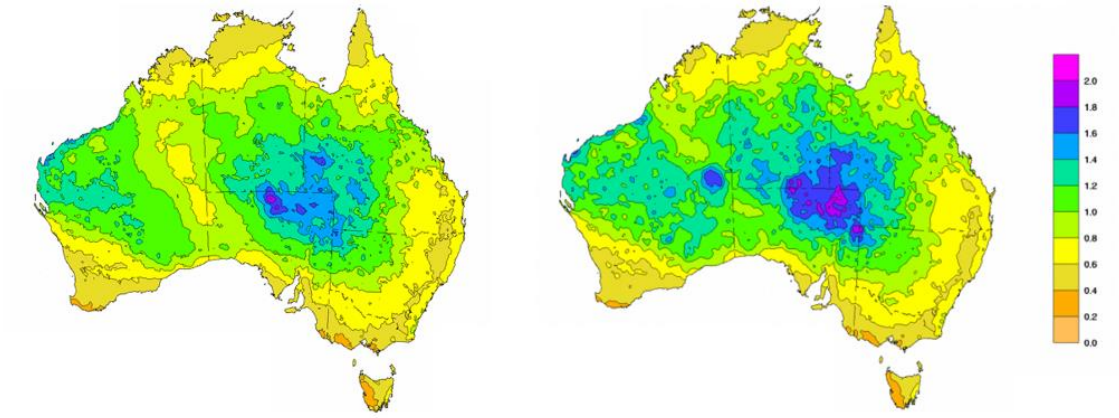

Figure 18: Annual rainfall variability, defined as the 90th rainfall percentile minus the 10th rainfall percentile, with the result divided by the 50th percentile (or median). Data from 1900 to 2018 used in the analysis. SI analysis on left, AWAP analysis on right

4.3 Pre-1900 analyses

In Australia, historical accounts and scientific evidence have revealed the extent and spatial variations of drought, and wet years for that matter, across the continent (Heathcote 1969, Fenby and Gergis 2012, Gergis and Ashcroft 2013). Indeed, the Federation drought that occurred at around the turn of the twentieth century is widely recognised as one of the most notable droughts in Australia's recorded history. This drought became established in the late 1890s and affected large areas across southeast Australia before culminating in severe dry conditions and large stock and agricultural loses in 1901-1902. Despite accounts that the Federation drought was characterised as an extended period of very dry conditions, there were reports that some areas recorded very wet months from unseasonably heavy rainfall (Gibbs and Maher, 1967). Earlier, in the late 1880s, reports described contrasting conditions from one side of the country to the other, hinting to broad scale variability in the climate system. For example, extremely dry conditions occurred in Victoria, Tasmania, New South Wales and Queensland during years 1888-1889 while agricultural areas in Western Australia saw a period of above average rainfall that was then followed by a year of rainfall deficits in 1891 (Foley, 1957).

The presence of drought events and periods of above average rainfall in the early part of the climate record, and their relevance in historical context sets our primary motivation for extending our analyses as far back as feasibly representable.

With consideration of the fractional area that shows reasonable station density coverage across southeast Australia, the decorrelation length-scales and incorporation of pseudo-observations, as discussed in previous sections, we have extended the SI monthly rainfall analyses back to 1880. Despite the sparseness of the rainfall network across inland Australia, previous studies have attempted to define pre-1900 periods of drought through spatial decile analysis (e.g. Gibbs and Maher 1967, Heathcote 1969). We assess the ability for SI to reproduce previously reported percentage area statistics for the 'centennial' drought of 1888 as described in Gibbs and Maher (1967). For the areas defined to be in the decile-1 band in 1888, Table 2 and Figure 19 show a high level of agreement between the values reported in Gibbs and Maher (1967) and SI.

Region	Gibbs and Maher 1967 (percentage area)	SI (percentage area)
Australia	17 %	16.4 %
Qld	17 %	12.4 %
NSW	80 %	71.8 %
Vic	50 %	32.8 %
Tas	28 %	34.4 %
SA	21 %	35.8 %
WA	0 %	0.1%
NT	0 %	0.0 %

Table 2: Comparison of the percentage area in drought (decile 1 rainfall), as defined in Gibbs and Maher (1967) to the analysed percentage area from SI for the year 1888.

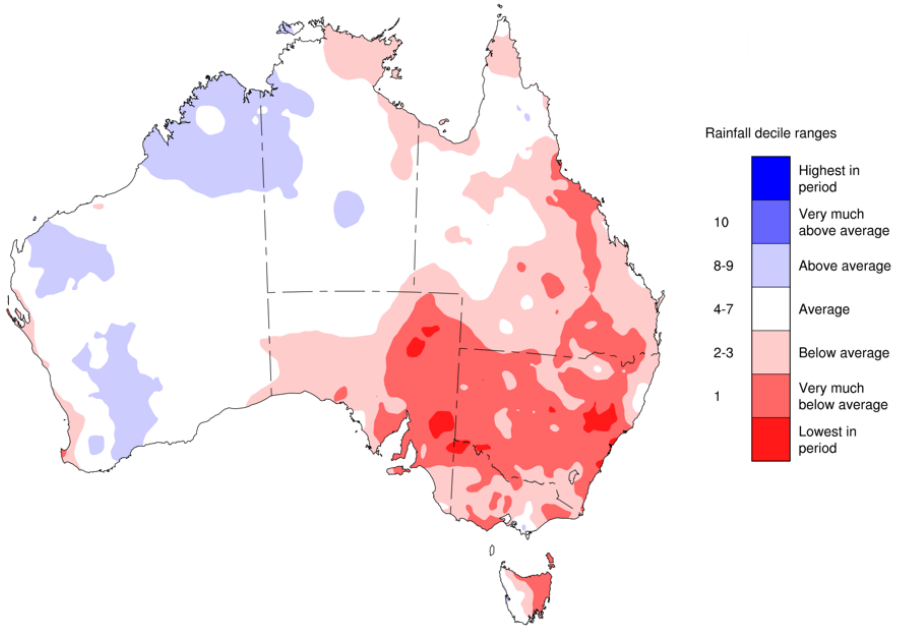

Figure 19: Rainfall deciles for the 12-month period January-December 1888. The percentile calculation has a reference period of 1885-1965 to enable a direct comparison with Gibbs and Maher (1967).

4.4 Incorporation of hydrological station network

In this section we show how the inclusion of the hydrological station network influences the SI analyses. A focus of this study has been to both look to use more rainfall data in the analysis as well as improving the use of those data (i.e. the spatial analysis method).

As shown previously (Figure 2), the hydrological network has grown from a few hundred stations in the early to mid-2000's to more than 1000 from around 2010. In light of this, we show the month-by-month results of analysis bias over the period 2000-2019. The grey curve shown in Figure 20 includes all available hydrological stations included in the SI analysis. As compared to the orange curve, it is evident that the hydrological network recordings are biased low when compared to the conventional network, resulting in an underestimation in the analysed rainfall field when they are used in the SI algorithm. In other words, the rainfall recorded by the hydrological stations appears too low compared to other stations in the rainfall network which are maintained to a higher standard.

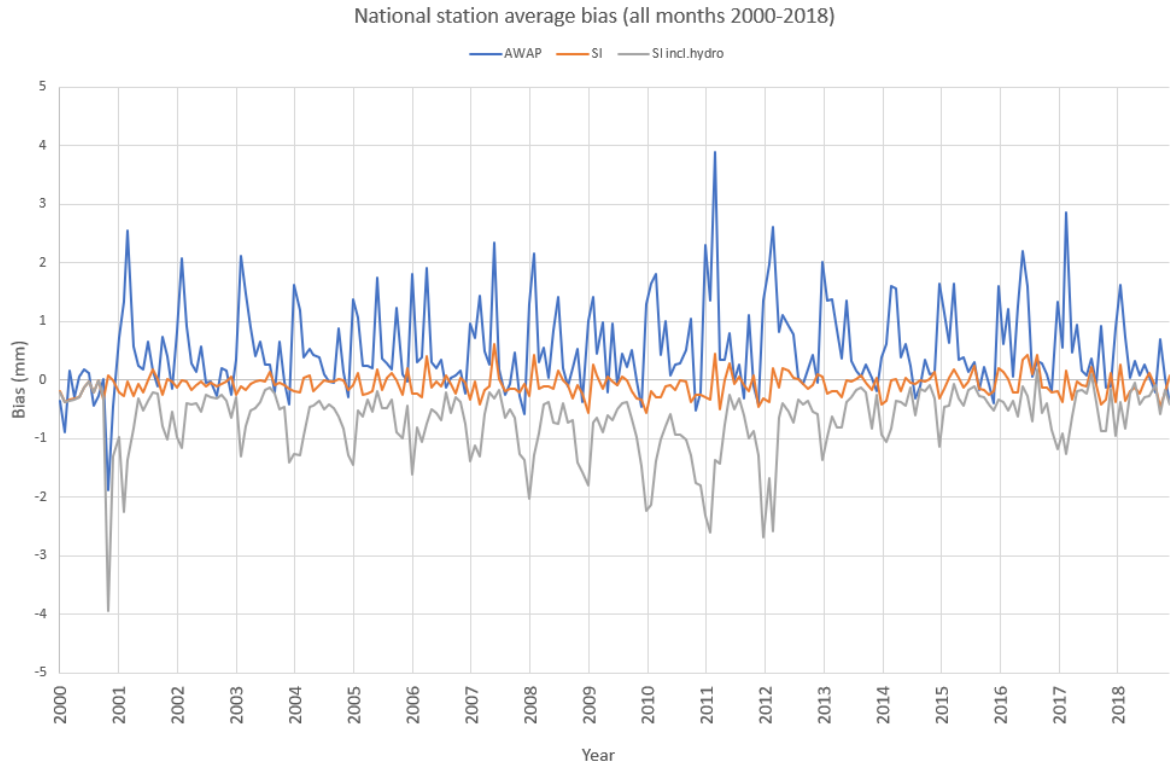

Figure 20: Cross-validated bias for monthly rainfall for the analysis period (January 2000- December 2018) averaged across all stations. Blue curves represent results from AWAP, orange curves are results from SI analysis, grey curve are results from SI with the inclusion of all available hydrological stations. Units in mm.

A possible reason why adding hydrological stations leads to a poorer rainfall analysis could be the use of equipment types and siting characteristics that differ from the Bureau's climate rainfall network specification. Additionally, the data coming from the hydrological sites does not undergo the same level of quality control as Bureau operated stations and tends to have more data outages (i.e., miss reports). The bias does appear to close over the most recent years when a greater effort has gone into the maintenance and support for these networks suggesting that the quality gap between the data has been partially addressed. Further work is required to completely understand why the hydrology stations are biased slightly low compared to the conventional rainfall network. The overall bias is small, but this flows directly through to the analysis errors.

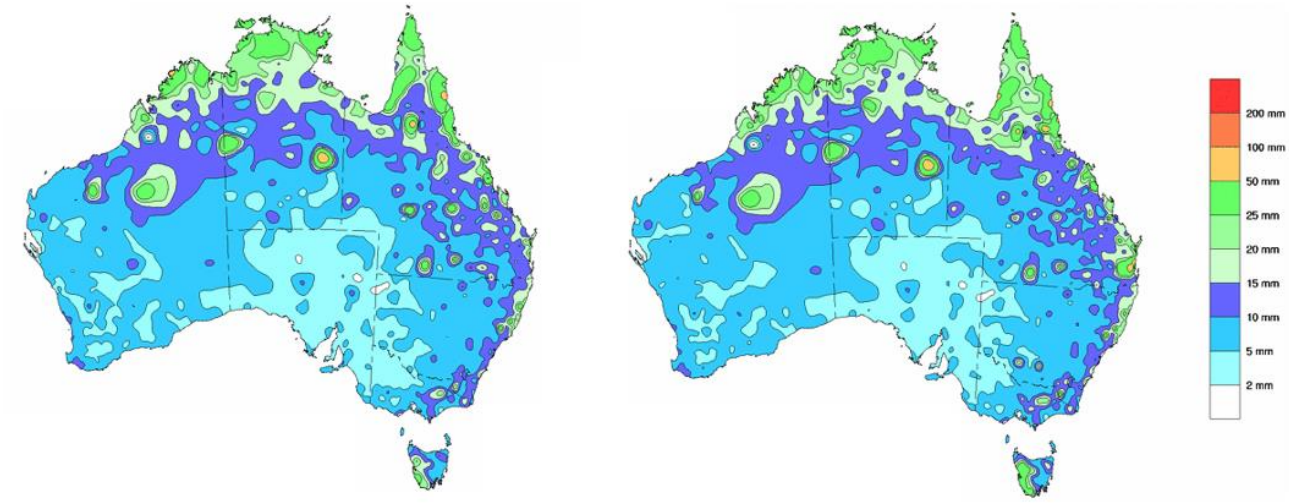

Figure 21: Cross-validated root mean square error (RMSE) for monthly rainfall for the eight years 2011-2018. SI analysis on left (as in figure 12, left) and SI analysis with the inclusion of hydrological locations (right). Units in mm.

The introduction of a particularly dense hydrological network shows a tendency to degrade the bias (Figure 20, Table 3). Indeed, comparison of the mapped RMSE in Figure 21, does indicate that the SI analysis with the inclusion of the hydrological stations tends to *locally* increase the RMSE, particularly throughout southeast and northern Queensland (i.e. Figure 21 right). However, there is negligible difference in the all station average RMSE and MAE across all months for the period 2000-2018, which suggests that if the bias could be removed from the hydrology stations then they could add to an improved rainfall analysis overall..

all months (2000-2018)	AWAP	SI	SI incl.hydro
Absolute BIAS	0.63	0.16	0.70
RMSE	21.38	18.87	18.86
MAE	11.68	9.15	9.11

Table 3: Comparison of verification statistics, bias, RMSE and MAE. All station average, across all months spanning 2000-2018.

The minimal change and local increase in analysis errors when more stations are used can come about for two reasons. The first is when the data have a bias (which we have shown is the case here). The second case is when the stations have notably larger observational errors and so introduce more error into the analysis. Both of these issues can be addressed but require a longer period of data and a better description of the interstation correlations.

The results here do not invalidate the use of hydrological stations for other purposes, but rather highlight that they may not be so easily used for some climate applications. This point is important as it is not unreasonable to expect that other networks such as those using low-cost private weather stations are likely to have similar issues. These do present a potentially rich data source, but their

use will not always be straight-forward or simple and it is possible that they may lead to a somewhat worse analysis outcome on some attributes.

5. SUMMARY and CONCLUSIONS

In this report we have provided a detailed description of a new meteorological analysis product developed by the Australian Bureau of Meteorology as a contribution to the New South Wales (NSW) Rural Assistance Authority (RAA) project "Enhanced weather information for farm drought resilience in NSW". This system is running in real time and is expected to form the basis for ongoing monitoring and mapping of Australia's climate by the Australian Bureau of Meteorology going forwards.

We have shown that, with the use of SI, analysis error of monthly rainfall is greatly improved with a reduction in the annual average RMSE of 13.5% with respect to AWAP. To achieve the same result with AWAP alone, would require up to 2600 additional rainfall stations nationally. Based upon purchase and installation prices plus 5 years of running cost, this equates to $41.6M if the stations are automated rain gauges only. The bias from the implementation of SI, such that its timeseries exhibits near stationarity are also greatly reduced with respect to AWAP. This indicates that the SI tuning coefficients are pretty close to optimal and the choice of a single reference baseline period has significant benefits for analysis accuracy.

Future work will look to extend this analysis technique across multiple variables and multiple time-scales (particularly days). The approach has been found to be robust, to preserve the background climatology in the long term and to be computationally efficient such that the entire historical analyses can be re-run in a short period of time. These systems produce a substantial improvement on existing Bureau practice.

The benefits for adopting the SI methodology also include the ease with which additional datasets can be included into the analysis process. For example, adding radar, satellite rainfall estimates or short-range model forecasts is quite simple. Our results show that the analysis errors are quite insensitive to station number (e.g. they don't change much through time). This means that very large networks will be required if significant further improved analyses are required, and clearly this will not be feasible using traditional rainfall networks. As such, future work should focus on expanding the range of data used for analysis, eventually moving to multivariate analyses where the physical relationship between the variables is preserved. This is done in data assimilation to numerical weather prediction models and is the basis for so-called reanalysis datasets.

6. REFERENCES

Australian Bureau of Meteorology. 2010. ADAM (Australian Data Archive for Meteorology) Manual Version 2.5. Australian Bureau of Meteorology: Melbourne.

Barnes, S. L., 1964. A technique for maximizing details in numerical weather map analysis. J. Appl. Meteor., 3, 396–409.

Beesley, C. and Frost, A. 2009. *A comparison of the BAWAP and SILO spatially interpolated daily rainfall datasets*. Bureau of Meteorology, Melbourne, Australia.

Bioregional Assessment Programme. 2014. GEODATA 9 second Digital Elevation Model Version 3 Hillshade. Bioregional Assessment Derived Dataset. Viewed 10 December 2018, http://data.bioregionalassessments.gov.au/dataset/48f9247c-e15c-4b65-afd9-c8c7e38d5614.

Blomley, J. E., Smith, N. R., and Meyers, G. 1989. An oceanic subsurface thermal analysis scheme: BMRC Research Rep. 18. Bureau of Meteorology, Melbourne, Australia.

Bratseth, A. M., 1986. Statistical interpolation by means of successive corrections. Tellus, 38A, 439–447.

Bureau of Meteorology and CSIRO 2018. State of the Climate 2018. 24pp. Available from http://www.bom.gov.au/state-of-the-climate/State-of-the-Climate-2018.pdf

Chubb, T., Manton, M., Siems, S., & Peace, A. D. 2016. Evaluation of the AWAP daily precipitation spatial analysis with an independent gauge network in the Snowy Mountains. *Journal of Southern Hemisphere Earth Systems Science*, *66*(1), 55-67. https://doi.org/10.22499/3.6601.006

Daley R. 1993. Atmospheric data analysis. Cambridge University Press, New York, U.S.A., 457pp.

Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J. and Pasteris, P.P. (2008), Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. Int. J. Climatol., 28: 2031-2064. doi:10.1002/joc.1688

Dowdy, A., Catto, J. 2017. Extreme weather caused by concurrent cyclone, front and thunderstorm occurrences. *Sci Rep* **7,** 40359 (2017). https://doi.org/10.1038/srep40359

Ebert, E., Janowiak, J.E. and Kidd, C. 2007. Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bull. Am. Met. Soc., 88*, 1-18.

Fenby C, Gergis J. 2012. A rainfall history of south-eastern Australia part 1: comparing evidence from documentary and palaeoclimate records, 1788–1860. *International Journal of Climatology* DOI:10.1002/joc.3640.

Foley, J.C. 1957. Drought in Australia. Bulletin No. 43, Bureau of Meteorology, Melbourne

Gandin, L. S., 1963. Objective Analysis of Meteorological Fields. Gidrometeorologicheskoe Izdaltel'stvo, 286 pp.

Gergis, J., & Ashcroft, L. 2013. Rainfall variations in south-eastern Australia part 2: a comparison of documentary, early instrumental and palaeoclimate records, 1788–2008.

Glowacki, T.J., Xiao, Y. and Steinle, P. 2012. Mesoscale Surface Analysis System for the Australian Domain: Design Issues, Development Status, and System Validation. Weather and Forecasting. 27, 141–157.

Glowacki, T.J. and Seaman, R. 1987. A general purpose package for univariate two-dimensional data checking and analysis. *BMRC Research Report No. 5*, Bur. Met. Australia.

Gibbs, W. J., and J. V. Maher, 1967. *Rainfall Deciles as Drought Indicators.* Australian Bureau of Meteorology, Bull. 48, 37 pp.

Heathcote, R. L., 1969. Drought in Australia: a problem of perception. Geographical Review 59, 175–194

Hopkinson, R.F., Hutchinson, M.F., McKenney, D.W., Milewska, E.J. and Papadopol, P. 2012. Optimizing input data for gridding climate normals for Canada. *Journal of Applied Meteorology and Climatology* 51: 1508-1518. doi: 10.1175/JAMC-D-12-018.1

Hutchinson, M.F. and Bischof, R.J. 1983. A new method for estimating the spatial distribution of mean seasonal and annual rainfall applied to the Hunter Valley, New South Wales, *Australian Meteorological Magazine* 31: 179-184.

Hutchinson M.F. 1995. Interpolating mean rainfall using thin plate smoothing splines. Int. J. Geog. Inf. Systems, **9**, 385-403.

Hutchinson, M.F. 1998a. Interpolation of Rainfall Data with Thin Plate Smoothing Splines - Part I: Two Dimensional Smoothing of Data with Short Range Correlation. Journal of Geographic Information and Decision Analysis. 2. 153-167.

Hutchinson, M.F. 1998b. Interpolation of Rainfall Data with Thin Plate Smoothing Splines: II. Analysis of Topographic Dependence. Journal of Geographic Information and Decision Analysis. 2. 168-185.

Jones D.A. and Weymouth G. 1997. An Australian monthly rainfall data set. Technical Report No. 70, Bureau of Meteorology, Melbourne, Australia. 19pp

Jones D.A. and Trewin B.C. 2000. On the description of monthly temperature anomalies over Australia. Australian Meteorological Magazine, **49**, 261-276. http://www.bom.gov.au/jshess/docs/2000/jones2_hres.pdf

Jones, D.A. and Trewin, B.C. 2002. On the adequacy of historical Australian daily temperature data for climate monitoring. *Aust. Met. Mag., 51*, 237-50.

Jones D.A., Wang W., Fawcett R. and Grant I. 2006. The generation and delivery of Level-1 historical climate data sets. Australian Water Availability Project Milestone Report. 32pp.

Jones, D., W. Wang, and R. Fawcett, 2009. High-quality spatial climate data-sets for Australia. *Aust. Meteorol. Oceanogr. J.*, **58**, 233–248, doi:10.22499/2.5804.003. http://www.bom.gov.au/jshess/docs/2009/jones.pdf

Jeffrey S.J., Carter J.O., Moodie K.B. and Beswick A.R. 2001. Using spatial interpolation to construct a comprehensive archive of Australian climate data. Env. Model. and Software, **66**, 309-330.

Julian P. R., and Thiebaux H. J., 1975. On some properties of correlation functions used in optimum interpolation schemes;*Mon. Weather Rev.* **103** 605–616

King AD, Alexander LV, Donat MG, 2013. The efficacy of using gridded data to examine extreme rainfall characteristics: a case study for Australia. International Journal of Climatology 33(10): 2376-2387.

Koch, S. E., S. C. Albers, J. A. McGinley, P. A. Miller, N. Wang, and Y. F. Xie, 2004. A new approach for mesoscale surface analysis: The space-time mesoscale analysis system (STMAS). Preprints, 11th Conf. on Aviation, Range, and Aerospace Meteorology/22nd Conf. on Severe Local Storms, Hyannis, MA, Amer. Meteor. Soc., J1.5. [Available online at http:// ams.confex.com/ams/pdfpapers/81796.pdf.]

Lorenc, A., 1981. A global three-dimensional multivariate statistical interpolation scheme. Mon. Wea. Rev., 109, 701–721.

Mills, G.A., Weymouth, G., Jones, D., Ebert, E.E., Manton, M., Lorkin, J. and Kelly, J. 1997. A national objective daily rainfall analysis system. *Bureau of Meteorology Development Report No. 1*, Bur. Met., Australia, 30 pp.

Orfanidis, S.J., 1996. *Introduction to Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ.

Savitzky, A. and Golay, M.J.E. 1964. Smoothing and Differentiation of Data by Simplified Least-Squares Procedures. Analytical Chemistry, 36, 1627-1639. http://dx.doi.org/10.1021/ac60214a047

Sharples, J.J. Hutchinson, M.F. and Jellett, D.R. 2005. On the horizontal scale of elevation dependence of Australian monthly precipitation. *Journal of Applied Meteorology* 44: 1850-1865.

Steinier, J.; Termonia, Y.; Deltour, J. 1972. Smoothing and differentiation of data by simplified least square procedure. *Anal. Chem.* 1972, *44*, 1906– 1909, DOI: 10.1021/ac60319a045

Tait, A.B., Henderson, R., Turner, R., & Zheng, X. 2006. Thin plate smoothing spline interpolation of daily rainfall for New Zealand using a climatological rainfall surface.

Tozer, C. R., Kiem, A. S., and Verdon-Kidd, D. C. 2012. On the uncertainties associated with using gridded rainfall data as a proxy for observed, Hydrol. Earth Syst. Sci., 16, 1481–1499, https://doi.org/10.5194/hess-16-1481-2012.

Thiebaux, H. J., 1975. Experiments with correlation representations for objective analysis. Mon. Wea. Rev., 103, 617–627.

World Meteorological Organization, 2017. *WMO Guidelines on the Calculation of Climate Normals* (WMO-No.1203). Geneva.

APPENDIX 1

We calculate the Root Mean Square Error (RMSE) of the analysis at the observation stations as follows:

Consider a cross-validated estimate of a station value T at station k and time t, denoted by $\hat{T}(x_k, y_k, z_k, t)$. The cross-validated analysis error is given by

$$E_k(t) = \hat{T}(x_k, y_k, z_k, t) - T_k(t). \tag{7}$$

Aggregating across time, the RMSE at the station is

$$RMSE_k = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left[E_k(t)\right]^2} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left[\hat{T}(x_k, y_k, z_k, t) - T_k(t)\right]^2}. \tag{8}$$

Noting that N will vary from station to station and according to whether the analysis is for daily or monthly data. The observation $T_k(t)$ can be divided into a "true" component and an "observational error" component $e_k(t)$. The true component is what would be measured if the observation at station k was completely accurate, while the error component is the error introduced due to factors such as instrument miscalibration, misreading by the observer, errors in spatial representativeness arising from specific factors at the observation site and so on. Hence we have

$$T_k(t) = T_k^{True}(t) + e_k(t), \tag{9}$$

giving

$$E_k(t) = \hat{T}(x_k, y_k, z_k, t) - \left(T_k^{True}(t) + e_k(t)\right) \tag{10}$$

and

$$RMSE_k = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left[\hat{T}(x_k, y_k, z_k, t) - \left(T_k^{True}(t) + e_k(t)\right)\right]^2}. \tag{11}$$

On the assumption that the observational errors are statistically independent of the interpolated and true values (Daley 1993), this can be further simplified to

$$\begin{aligned}\left(RMSE_k\right)^2 &= \frac{1}{N}\sum_{t=1}^{N}\left[\hat{T}(x_k, y_k, z_k, t) - T_k^{True}(t)\right]^2 + \frac{1}{N}\sum_{t=1}^{N}\left[e_k(t)\right]^2 \\ &= \left(E_k^{True}\right)^2 + \left(E_k^{Obs}\right)^2.\end{aligned} \tag{12}$$

The first term in eq. 13, $\left(E_k^{True}\right)^2$, is the "true error variance" and is a measure of the accuracy of the analysis in estimating the true field. This value is the true analysis error. The second term in eq. 13, $\left(E_k^{Obs}\right)^2$, is the "observational error variance" and measures the accuracy of the observations. Clearly, even a perfect analysis will have a non-zero cross-validated error because observations have some level of error. To obtain a zero cross-validated error, the

observations need to be “perfect”, also noting that analysis methods are also not perfect. While it is common practice for the cross-validated differences between independent observations and analyses to be treated as “analysis errors”, and they are defined as such here, it is important to keep in mind that they also contain an observation error component. Daley (1993) and Jones and Trewin (2000) describe how the observational errors can be estimated.

In defining the analysis errors averaged across time and stations, additional measures of bias and Mean Absolute Error (MAE) are computed. These are both defined as (e.g., Jones and Weymouth 1997),

$$BIAS_k = \frac{1}{N}\sum_{t=1}^{N}\left[E_k(t)\right] \tag{13}$$

and

$$MAE_k = \frac{1}{N}\sum_{t=1}^{N}\left|E_k(t)\right| \tag{14}$$

The all station average is these terms extended across stations,

$$BIAS = \frac{1}{S}\sum_{k=1}^{S}\left[\frac{1}{N_k}\sum_{t=1}^{N_k}\left[E_k(t)\right]\right] \tag{15}$$

and

$$BIAS = \frac{1}{S}\sum_{k=1}^{S}\left[\frac{1}{N_k}\sum_{t=1}^{N_k}\left|E_k(t)\right|\right]. \tag{16}$$