Australian Government
Bureau of Meteorology

NWP QPFs accuracy analysis through comparative performance assessment of AWAP, AGCD and GPM-IMERG against Gauge Observations

Mohammad Mahadi Hasan, Elizabeth Ebert and Mohammadreza Khanarmuei

Bureau Research Report - 118

NWP QPFs accuracy analysis through comparative performance assessment of AWAP, AGCD and GPM-IMERG against Gauge Observations

Mohammad Mahadi Hasan, Elizabeth Ebert, Mohammadreza Khanarmuei

Bureau of Meteorology

Bureau Research Report No. 118

September 2025

National Library of Australia Cataloguing-in-Publication entry

Authors: Mohammad Mahadi Hasan, Elizabeth Ebert, Mohammadreza Khanarmuei

Title: NWP QPFs accuracy analysis through comparative performance assessment of AWAP, AGCD and GPM-IMERG against Gauge Observations

ISBN: 978-1-923469-09-9

ISSN: 2206-3366

Series: Bureau Research Report – BRR118

Enquiries should be addressed to:

Lead Author: Mohammad Mahadi Hasan

Bureau of Meteorology

GPO Box 1289, Melbourne

Victoria 3001, Australia

mohammad.hasan@bom.gov.au

Copyright and Disclaimer

Contents

List of Figures

List of Tables

Executive Summary

This comprehensive technical report examines the accuracy of Numerical Weather Prediction (NWP) Quantitative Precipitation Forecast (QPF) by comparing it with three observational datasets: Australian Water Availability Project (AWAP), Australian Gridded Climate Data (AGCD), and NASA's Integrated Multi-Satellite Retrievals for Global Precipitation Measurement (GPM-IMERG) rainfall data. Employing various analyses, including categorical assessments, temporal accumulations, spatial analysis and extreme rainfall analyses, the study explores similarities and differences in the AWAP, AGCD, and GPM-IMERG in comparison to gauge rainfall. By investigating multiple rainfall data sources, the report provides an extensive overview of their performance, highlighting strengths, weaknesses, and implications for NWP forecast evaluations.

AWAP and AGCD consistently emerge as reliable sources, closely aligning with ground-based observations. AGCD stands out as the most accurate, making it a preferred choice for various applications. AGCD demonstrates superior accuracy in categorising both no-rain and rain events across various thresholds compared to AWAP and GPM-IMERG. While GPM-IMERG provides global insights, it exhibits a tendency to slightly miscalculate rainfall distribution, especially during severe events. The study identifies regions with unique climate patterns where GPM-IMERG may face challenges, such as high rainfall areas in Eastern Australia and regions with low rainfall, like Western Australia and South Australia. GPM-IMERG, however, shows diminished accuracy in identifying cases with no rainfall, impacting its reliability for low rainfall analysis. GPM-IMERG faces challenges in accurately identifying both higher and lower intensity rainfall, resulting in increased false alarms and misses compared to AWAP/AGCD datasets. Regional variations in rainfall estimates highlight the necessity of accurate observational data for correct model performance evaluation. Temporal variability in rainfall measurements emphasises the need to consider accumulation over time.

The choice of observed data significantly impacts the reported accuracy of ACCESS-G4. When AWAP and AGCD are used instead of GPM-IMERG, there is a noticeable enhancement in ACCESS-G4's accuracy. However, regardless of observed data selection, ACCESS-G4's forecasts exhibit diminishing accuracy as lead days progress.

1. Introduction

The objective of this technical report is to comprehensively evaluate the accuracy of Numerical Weather Prediction (NWP) Quantitative Precipitation Forecast (QPF) in comparison to observational datasets: AWAP (Australian Water Availability Project) and AGCD (Australian Gridded Climate Data) and GPM-IMERG (Global Precipitation Measurement Integrated Multi-Satellite Retrievals) Rainfall data. Applying a range of analytical approaches, including categorical assessments, temporal, spatial and extreme rainfall analyses, this study seeks to find the similarities and disparities among these datasets and extract valuable insights from these comparative assessments. Through a comprehensive analysis of various rainfall data sources, this report aims to offer an extensive overview of their performance. This evaluation aims to find their individual strengths, weaknesses and their potential implications in evaluating NWP forecasts.

Types of Assessment:

Categorical Analyses: Categorical analyses imply the classification of rainfall events into distinct categories based on predefined intensity thresholds. This method enables the comparison of how well rainfall sources capture different intensity levels, aiding in understanding their performance across varied rainfall conditions.

Temporal Accumulation: Temporal accumulation involves aggregating rainfall data over specific time intervals, such as weekly, monthly, seasonal, and yearly accumulations. This evaluation examines how accurately datasets capture rainfall trends over different time scales, providing insights into their consistency and reliability over various temporal periods.

Extreme Rainfall Analyses: Extreme rainfall analyses focus on assessing the performance of rainfall sources in accurately capturing extreme weather events. This evaluation sheds light on the dataset's ability to represent and predict intense precipitation events, crucial for understanding their reliability in extreme weather forecasting.

Statistical Analyses: All data statistics encompass fundamental characteristics of each dataset, offering insights into their accuracy, tendencies, spatial variability, and distribution. This evaluation provides a comprehensive understanding of dataset behaviour, aiding in identifying patterns, biases, and overall performance across diverse geographical locations and conditions.

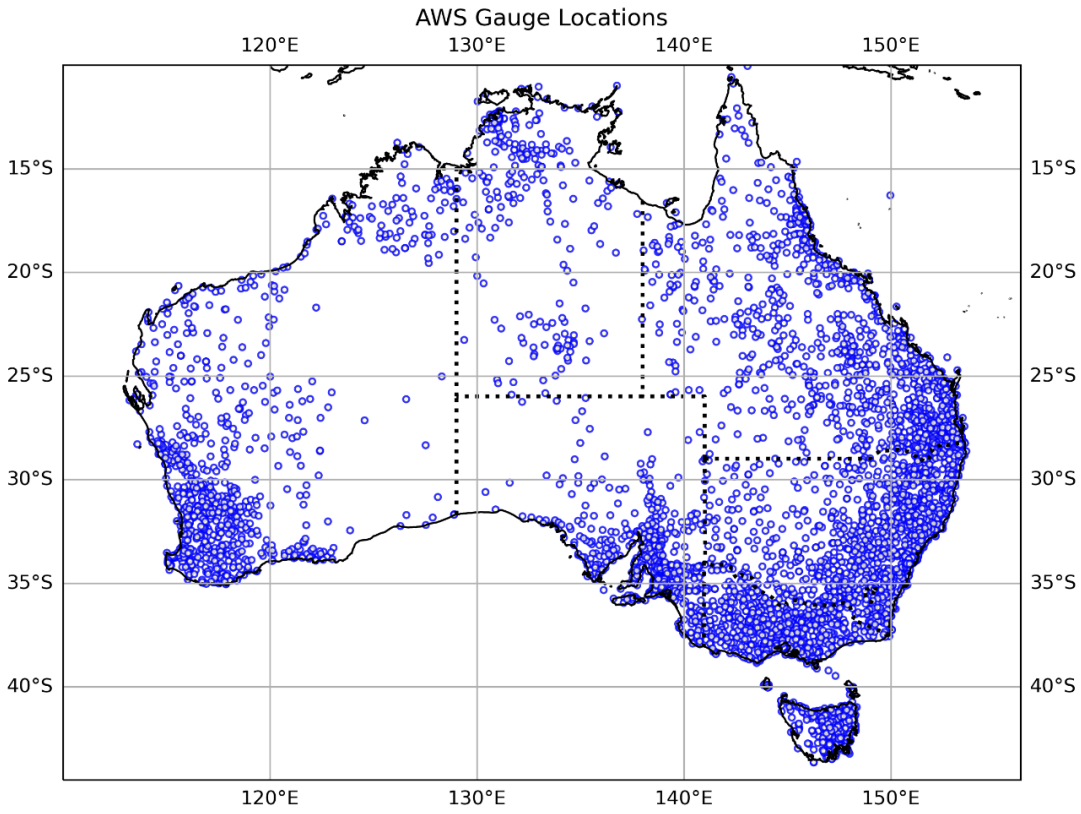


Figure 1: Performance assessment Locations (around 4800 gauges).

Table 1:Data Sources.

Product	Resolution (Lat, Lon)	Data Period
AGCDv2	Spatial resolution 0.01° by 0.01°, daily from 9 am to 9 am.	January, 2006 – June, 2023
AWAP	Spatial resolution 0.05° by 0.05°, daily from 9 am to 9 am.	January, 2006 – June, 2023
GPM-IMERG	Spatial resolution 0.1° by 0.1°, daily from 10 am to 10 am.	January, 2006 – June, 2023
ACCESS-G4	Spatial resolution 0.11719° by 0.175781°, 3-hourly from 10 am to 10 am, accumulated daily, 10 days lead time.	July, 2022 – June, 2023
Gauge Rainfall	Bureau gauges, daily from 9 am to 9 am	January, 2006 – June, 2023

Table 2: Performance evaluation Matrices.

Metrics	Best Value	Comment
Mean Absolute Error (MAE)	0	Lower is better
Root Mean Square Error (RMSE)	0	Lower is better
Normalised Mean Absolute Error (NMAE)	0	Lower is better
Normalised Root Mean Square Error (NRMSE)	0	Lower is better
Nash-Sutcliffe Efficiency (NSE)	1	Higher is better
Kling-Gupta Efficiency (KGE)	1	Higher is better
Relative Bias (RBias)	0	Lower is better
Correlation Coefficient	1	Higher is better

* Equations are included in the appendix

A variety of metrics were utilised to evaluate the accuracy and performance of different datasets in this study. Refer to the following links for detailed information corresponding to different metrics: https://www.cawcr.gov.au/projects/verification/.

2. Gridded Observation Assessment

2.1. Daily Rainfall Accuracy of Gridded Observational Datasets:

- **How does the accuracy of gridded data (AWAP, AGCD) and satellite-based data (GPM-IMERG) compare to gauge observations in daily rainfall measurements?**

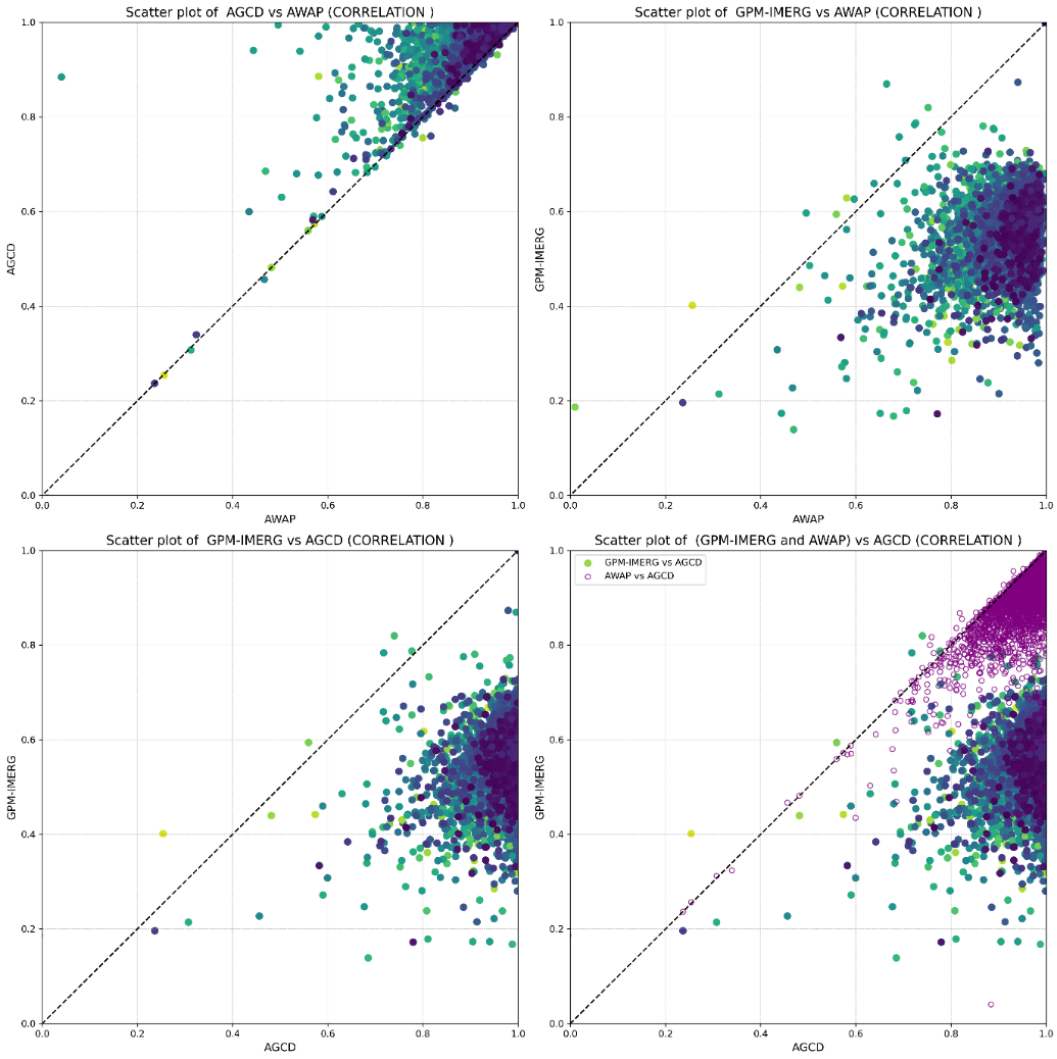


Figure 2:The Correlation coefficient Comparison between AGCD, AWAP, and GPM-IMERG Rainfall against Gauge Rainfall for the data period mentioned in Table 1. Panel Layout: Top Left - Correlation between AGCD and Gauge Rainfall vs. Correlation between AWAP and Gauge Rainfall. Top Right - Correlation between GPM-IMERG and Gauge Rainfall vs. Correlation between AGCD and Gauge Rainfall. Bottom Left - Correlation between GPM-IMERG and AWAP vs. Bottom Right - Correlation Comparison of GPM-IMERG and AWAP with Gauge Rainfall against Correlation of AGCD with Gauge Rainfall.

Table 3: Performance of AWAP, AGCD, and GPM-IMERG against gauge observations.

Metrics name	AWAP	AGCD	GPM-IMERG
Trend of Bias	underestimates	underestimates	overestimates
Relative Bias (%)	6.514	3.394	37.282
MAE (mm/day)	0.541	0.224	2.042
RMSE (mm/day)	2.335	1.454	6.520
NMAE	0.282	0.124	1.143
NRMSE	0.019	0.013	0.058
NSE	0.798	0.863	-1.436
KGE	0.818	0.922	0.344
NNSE	0.891	0.945	0.508
Correlation Coefficient	0.937	0.968	0.636

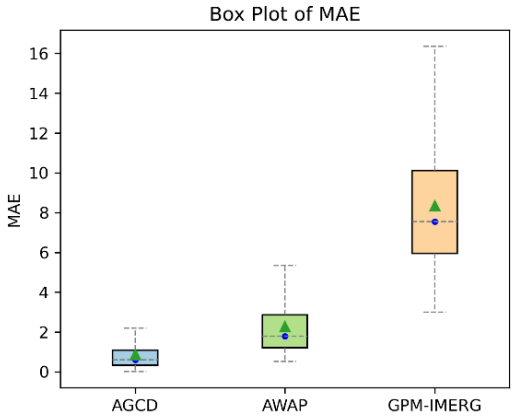


Figure 3: The Mean Absolute Error (MAE) of AGCD, AWAP, and GPM-IMERG Rainfall in comparison to Gauge Rainfall across all locations. (Here green triangle represents the median and the blue dot represents the mean value).

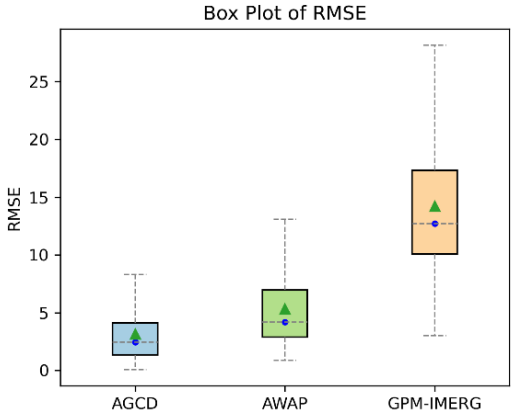


Figure 4: The Root Mean Squared Error (RMSE) of AGCD, AWAP, and GPM-IMERG Rainfall relative to Gauge Rainfall across all locations. (Here green triangle represents the median and the blue dot represents the mean value).

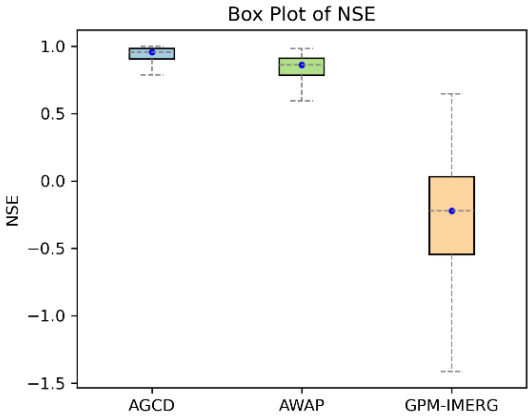


Figure 5: The Nash-Sutcliffe Efficiency (NSE) of AGCD, AWAP, and GPM-IMERG Rainfall concerning Gauge Rainfall across all locations.

AGCD emerges as the most precise source, boasting rainfall measurements that closely mirror ground-based observations. AWAP, while commendably accurate, falls marginally short of AGCD in terms of precision. GPM-IMERG showcases decreased accuracy in comparison to both AGCD and AWAP. While offering broad global coverage and valuable insights, GPM-IMERG might exhibit limitations in capturing daily rainfall events with utmost precision.

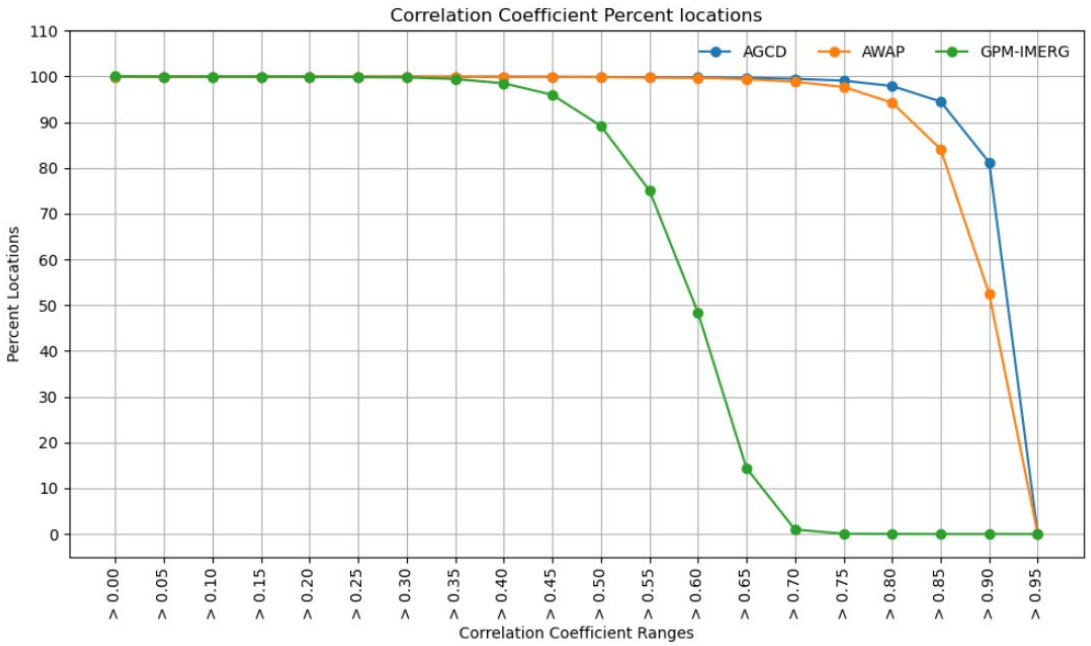


Figure 6: The variability in Correlation between AGCD, AWAP, and GPM-IMERG Rainfall against Gauge Rainfall is presented as a percentage of locations for daily rainfall.

Gridded data like AWAP and AGCD demonstrate strong agreement with gauge observations in daily rainfall, exhibiting correlations above 0.80 across most monitoring stations (more than 90% of the locations). However, for GPM-IMERG rainfall, only approximately 15% of locations display a correlation of 0.65 or higher. AGCD and AWAP, generally more accurate, show slight underestimation compared to gauge measurements. GPM-IMERG tends to notably overestimate rainfall, impacting its reliability in capturing accurate rainfall measurements compared to gauge observations. Comparing 16 years of daily rainfall data, Relative biases are approximately 3.4% (AGCD), 6.5% (AWAP) and over 37.3% (GPM-IMERG). Mean Correlation values for all the locations are 0.96 (AGCD), 0.93 (AWAP), and 0.63 (GPM-IMERG). Mean Absolute Errors range from 0.22 mm/day (AGCD) to notably higher at 2.0 mm/day (GPM-IMERG). Kling-Gupta Efficiency shows scores of 0.92 (AGCD), 0.81 (AWAP), and 0.34 (GPM-IMERG).

- **How consistent are these data sources in capturing daily rainfall patterns across various geographical locations?**

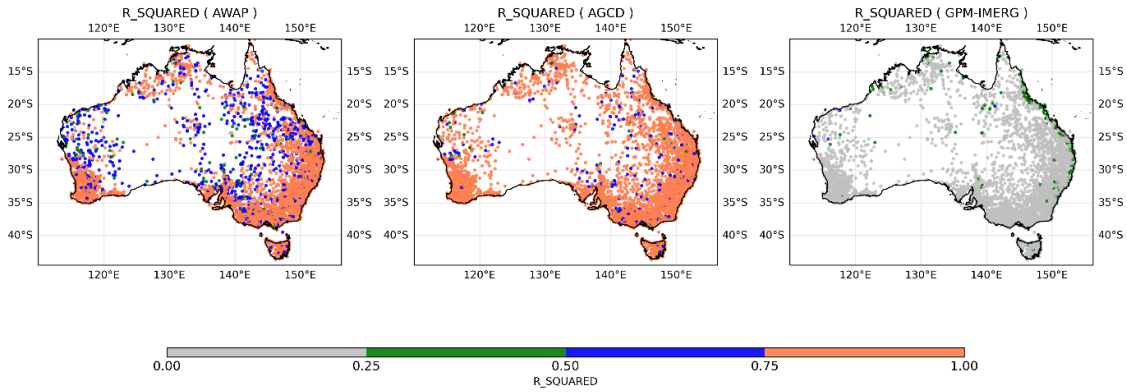


Figure 7: The R-squared Comparison between AGCD, AWAP, and GPM-IMERG Rainfall against Gauge Rainfall.

R-squared quantifies the proportion of variance in the observed data explained by the modelled data, reflecting how well the model aligns with observed data points. Its scale ranges from 0 to 1, where 1 signifies a perfect fit explaining all variance, while lower values denote weaker relationships or less explained variance. Higher R-squared values indicate closer alignment between model predictions and observed data, suggesting a better fit. Conversely, lower values imply the model explains less observed variability.

In the context of AWAP rainfall, R-squared values range from 0.75 to 1.0 along the east and west coasts near Perth and Darwin, indicating high accuracy. Interior locations show values from 0.50 to 0.75, suggesting reasonably good performance. For AGCD rainfall, most locations display R-squared values between 0.75 and 1.0, indicating notably accurate estimations. Interior areas range from 0.50 to 0.75 in AWAP, showcasing reasonably good performance. However, GPM-IMRG exhibits R-squared values mostly between 0 to 0.25 across most locations, with some along the east coast showing values between 0.25 to 0.50, indicating slightly improved accuracy in these regions.

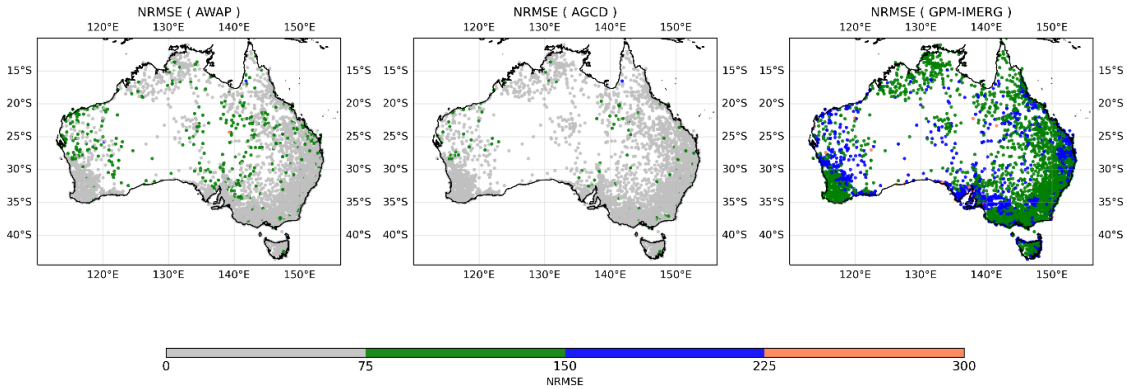


Figure 8: The Normalised Root Mean Square Error (NRMSE) Comparison between AGCD, AWAP, and GPM-IMERG Rainfall against Gauge Rainfall.

Across various locations, the Normalised Root Mean Square Error (NRMSE) for AGCD consistently registers lower values compared to both AWAP and GPM-IMERG rainfall. While the RMSE values vary across locations, the trend indicates that, restricting a few exceptions, AGCD consistently demonstrates lower RMSE compared to AWAP. Additionally, AGCD outperforms GPM-IMERG rainfall in terms of RMSE across most locations. This suggests a more promising accuracy for AGCD in these comparisons.

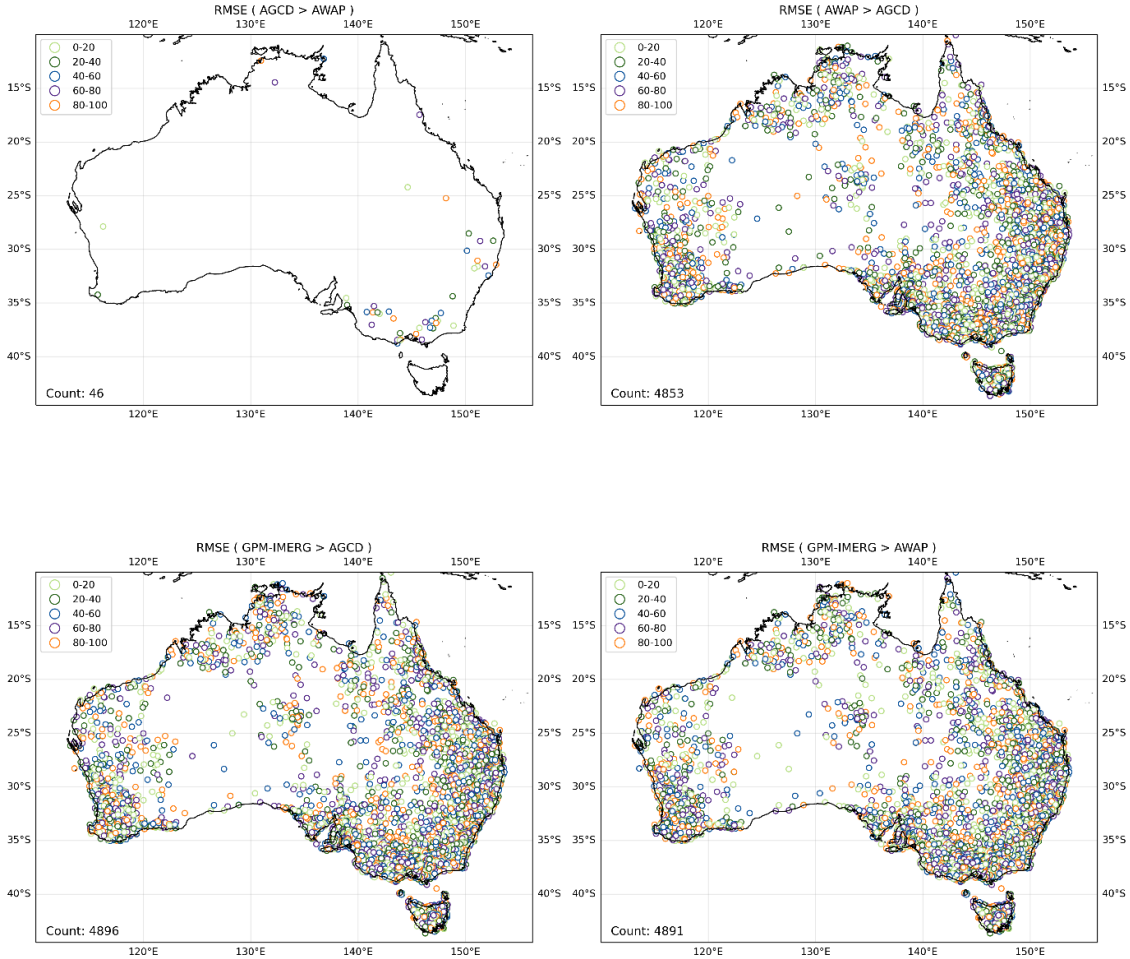


Figure 9: The Root Mean Square Error (RMSE) comparison among AGCD, AWAP, and GPM-IMERG Rainfall against Gauge Rainfall. The panel layout showcases specific locations: the top-left indicates areas where AGCD RMSE surpasses AWAP rainfall, the top-right displays regions with lower AGCD RMSE compared to AWAP rainfall, the bottom-left highlights locations where GPM-IMERG RMSE exceeds AGCD, and the bottom-right illustrates areas where GPM-IMERG RMSE surpasses AWAP.

2.2. Extreme Rainfall Events:

AGCD consistently demonstrates a close alignment with gauge observations across various parameters. Whether examining the 1 Day Maximum Rainfall, Annual Total Rainfall, Count of Wet days, Consecutive Dry days, Maximum Rainfall within 15 or 30 days, Standardised Precipitation Index, Days above Thresholds, or analysing the frequency, intensity, and count of rainfall above the 95th percentile, AGCD showcases a remarkable similarity to gauge observations. This pattern persists even when considering the Annual Consecutive 5 Wet Days of rainfall. AGCD's consistency in aligning with gauge observations across these diverse metrics underscores its reliability and accuracy in capturing various aspects of rainfall data.

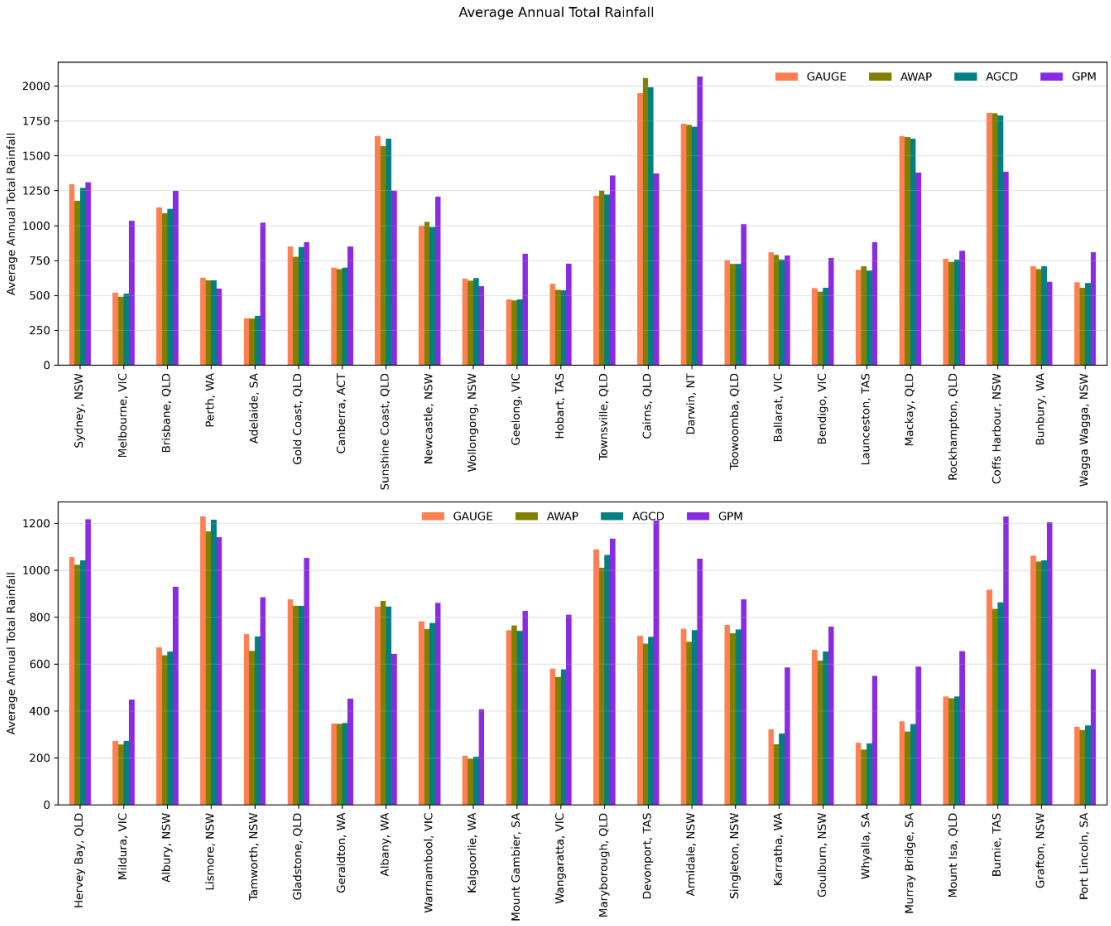


Figure 10: Illustration depicting average annual rainfall from Gauge, AGCD, AWAP, and GPM-IMERG datasets at various chosen towns (using an arbitrarily selected gauge location and the nearest grid values from gridded datasets) across Australia.

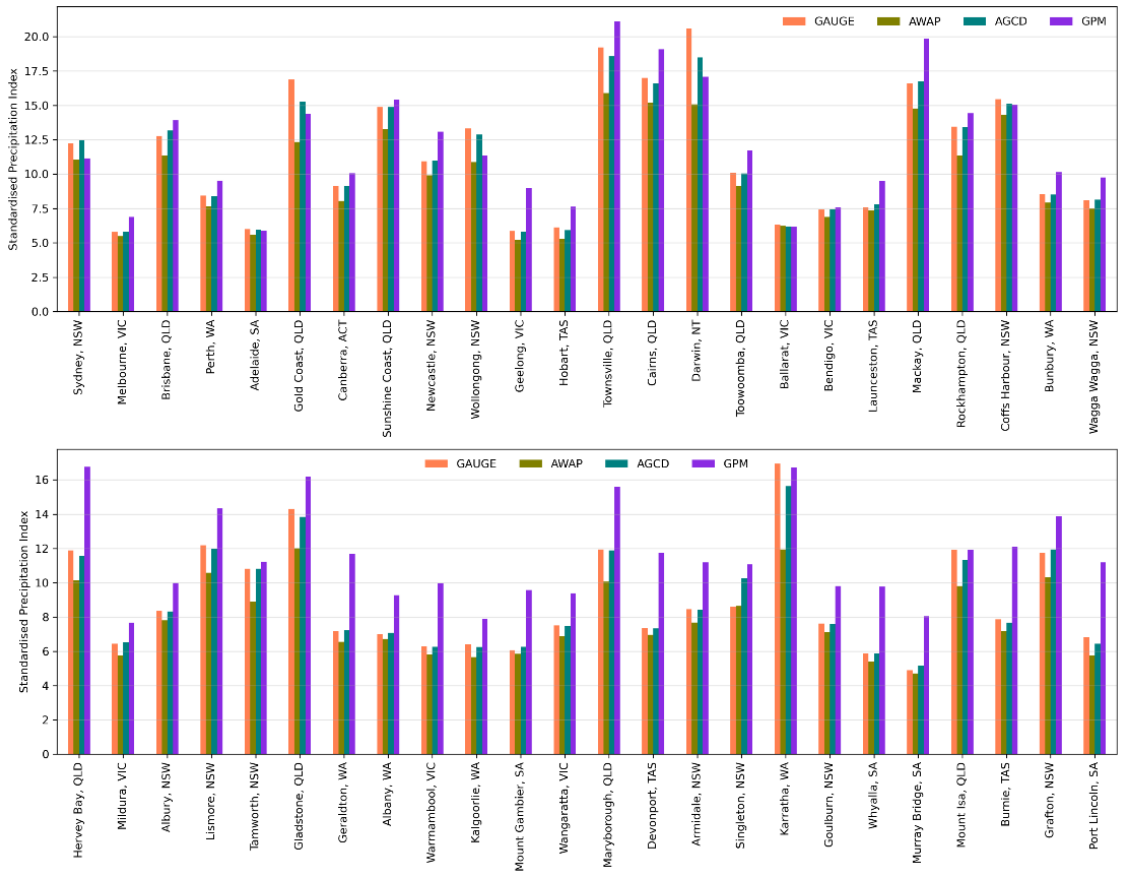


Figure 11: Illustration depicting Standardised Precipitation Index (SPI) from Gauge, AGCD, AWAP, and GPM-IMERG datasets at various chosen towns (using an arbitrarily selected gauge location and the nearest grid values from gridded datasets) across Australia.

GPM-IMERG tends to underestimate average annual rainfall and average 15 days maximum rainfall in Lismore (NSW), Wollongong (NSW), Coffs Harbour, (NSW), Bunbury (WA), Albany(WA), Mackay (QLD), Cairns(QLD) and Sunshine Coast (QLD). However, for the rest of the locations, GPM-IMERG tends to overestimate rainfall.

The average annual count of dry days (<1mm rainfall) is underestimated in Wagga Wagga (NSW), Devonport (TAS), Launceston (TAS), Bendigo (VIC), Ballarat (VIC), Gold Coast (QLD), Toowoomba (QLD), Darwin (NT), Wollongong (NSW), Armidale (NSW), Singleton (NSW), Canberra (ACT), Adelaide (SA), Karratha (WA), Perth (WA), Brisbane (QLD), Melbourne (VIC), Sydney (NSW), Mildura (VIC), Albury (NSW), Tamworth, Whyalla (SA), Port Lincoln (SA). This signifies that the number of days with less than 1mm of rainfall is reported lower than the actual observed count in these locations, especially when considering the GPM-IMERG rainfall data.

The average annual count of wet days (>1mm rainfall) is underestimated in Wollongong (NSW), Sunshine Coast (QLD), Geelong, Hobart, Townsville, Cairns, Bunbury (WA), Geraldton (WA), Albany (WA), Warrnambool (VIC), Mount Gambier (SA), Maryborough (QLD). This indicates that the number of days with more than 1mm of rainfall is reported lower than the actual observed count in these locations, especially when considering

GPM-IMERG rainfall data. The Standardised Precipitation Index (SPI) and 95th Percentile rainfall are underestimated in Sydney (NSW) and Coffs Harbour (QLD), particularly in consideration of the GPM-IMERG rainfall data.

- **To what extent do different rainfall sources capture extreme rainfall events in terms of intensity, frequency, and spatial distribution?**

AWAP and AGCD perform well in capturing extreme rainfall events, closely matching gauge observations in intensity and frequency distribution.

- **Are there discrepancies in identifying and quantifying extreme events between various datasets?**

GPM-IMERG tends to slightly overestimate extreme events' spatial distribution and captures their intensity imprecisely.

2.3. Categorisation of Rainfall Events:

Based on the categorical analysis, gauge-derived sources like AWAP and AGCD exhibit higher overall accuracy, considering their ability to correctly identify True Positives and minimise False Negatives.

Hits: This encompasses both true positives (TP) and true negatives (TN). True positives denote instances correctly identified as positive, while true negatives represent instances correctly identified as negative.

Misses: These are represented by false negatives (FN), indicating instances that were not correctly identified as positive. Misses highlight situations where the model failed to recognise positive occurrences.

False Alarms: This category includes false positives (FP), signifying instances incorrectly identified as positive. False alarms reveal instances where the model incorrectly signals positive occurrences that were, in fact, negative.

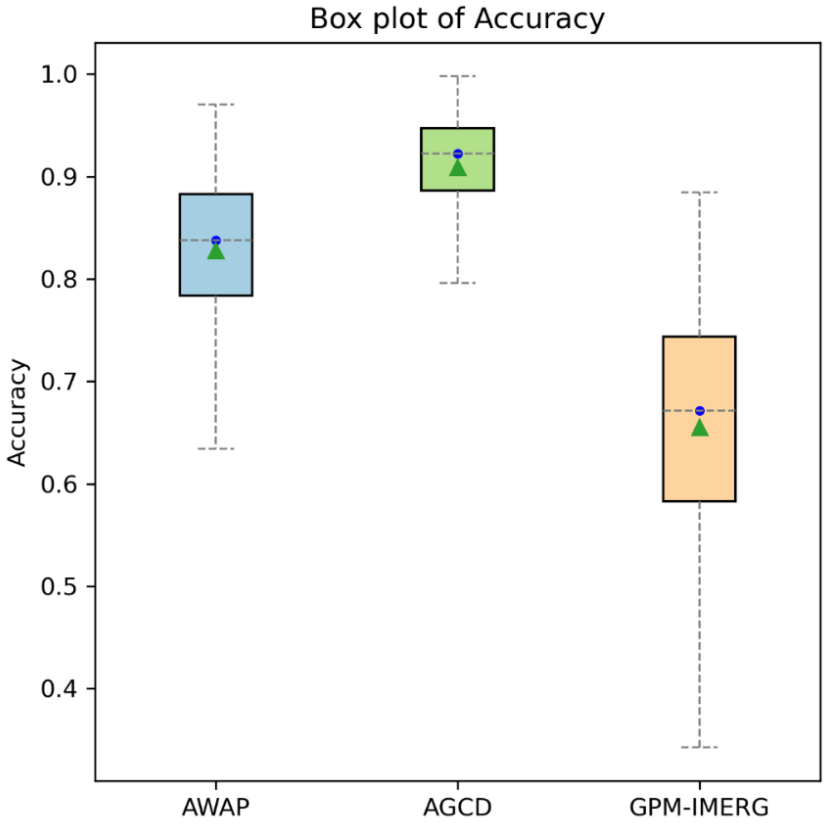


Figure 12: Illustration depicting Categorical Analysis (Accuracy) of AWAP, AGCD, and GPM-IMERG datasets across Australia. Accuracy measures the overall correctness. It is calculated as (TP + TN) / (TP + TN + FP + FN), where TP is the number of true positives (Correctly identified), TN is the number of true negatives (Correct Rejections), FP is the number of false positives (False Alarms), and FN is the number of false negatives (Misses). Accuracy provides a simple and intuitive measure of a model's performance.

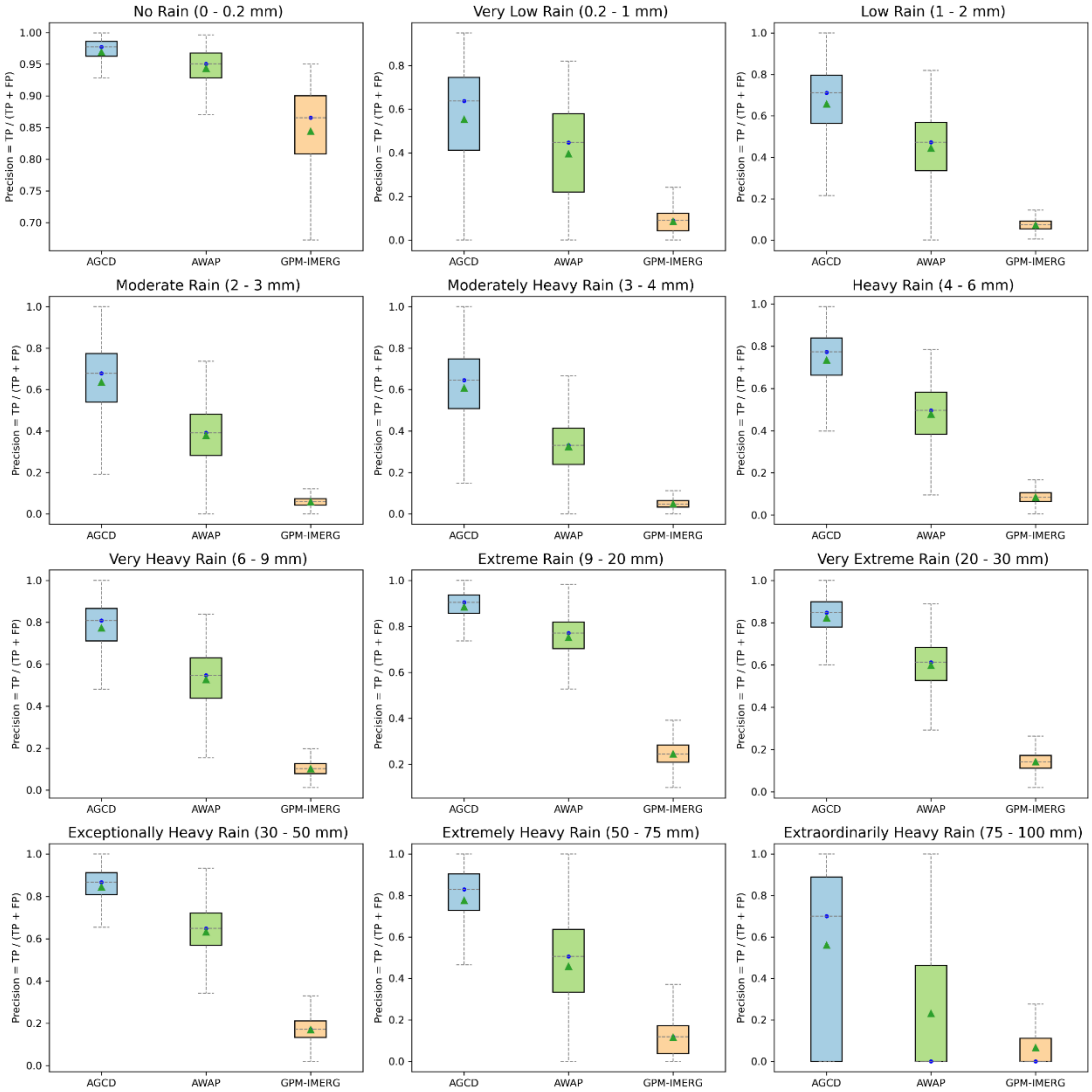


Figure 13: Illustration depicting the Precision Analysis of AWAP, AGCD, and GPM-IMERG datasets across Australia for different rainfall thresholds. Precision measures the accuracy of positive predictions made by the model. It is calculated as TP / (TP + FP), where TP is the number of true positives, and FP is the number of false positives. Precision focuses on the proportion of correctly predicted positive instances among all instances predicted as positive. High precision indicates that the model makes fewer false positive errors, making it useful when minimising false alarms is important.

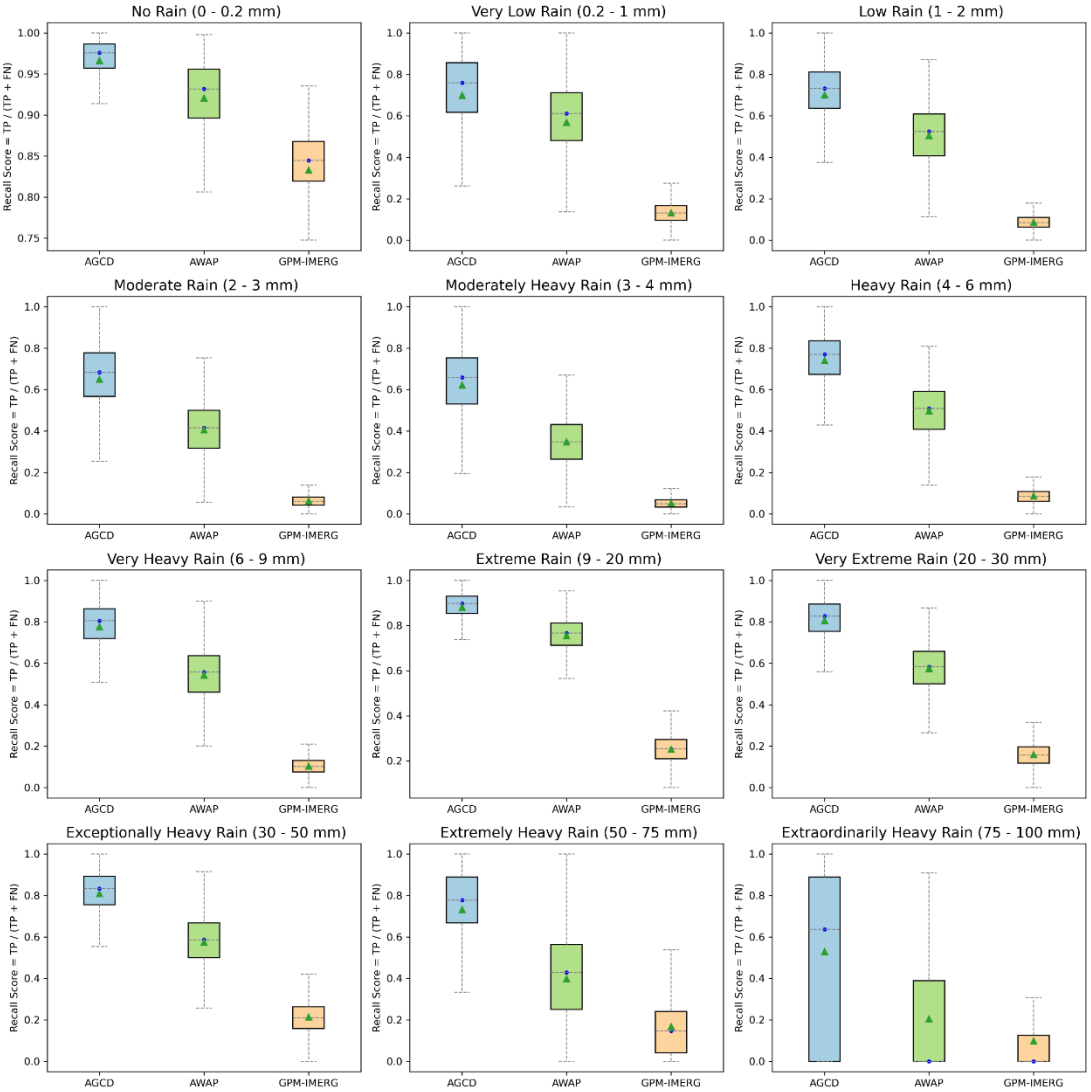


Figure 14: Illustration depicting Recall Analysis of AWAP, AGCD, and GPM-IMERG datasets across Australia for different rainfall thresholds. Recall measures the model's ability to correctly identify all relevant instances (true positives) within a category. It is calculated as TP / (TP + FN), where TP is the number of true positives and FN is the number of false negatives. Recall is useful when the cost of missing positive instances (false negatives) is high.

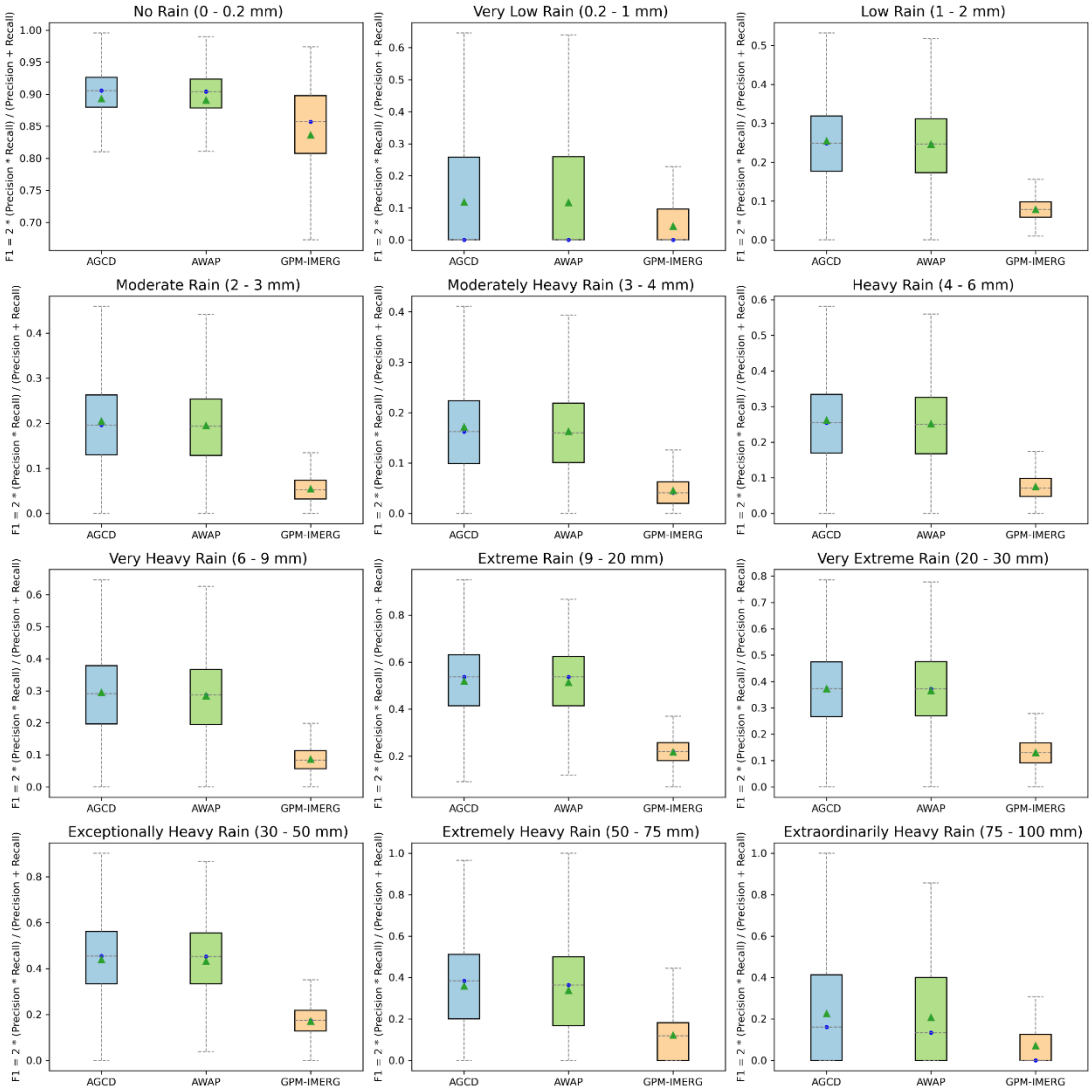


Figure 15: Illustration depicting the F1-score of AWAP, AGCD, and GPM-IMERG datasets across Australia for different rainfall thresholds. The F1-score is the harmonic mean of precision and recall. It balances the trade-off between precision and recall and provides a single metric that considers both false positives and false negatives. F1-score is calculated as 2 * (precision * recall) / (precision + recall), where precision is TP / (TP + FP), and recall is TP / (TP + FN). F1-score is particularly useful when you want to find a balance between minimising false positives and false negatives. Best value at 1 and worst score at 0.

A detailed examination reveals several significant insights. AGCD stands out in precision, accurately categorising both rainy and non-rainy events better than AWAP and GPM-IMERG. Conversely, GPM-IMERG struggles in identifying no-rain situations, indicating reduced reliability for drought analysis.

GPM-IMERG shows weaker precision in detecting rainfall between 0.2 mm to 9 mm and above 75 mm, indicating limitations in spotting smaller and larger rainfall events. AGCD demonstrates higher accuracy than AWAP, especially within the 0.2 mm to 50 mm range. AGCD notably outperforms AWAP in identifying rainfall beyond the 50 mm threshold (Figure 12 and Figure 13).

In recall analysis, GPM-IMERG exhibits notably low recall values, indicating numerous false negatives and missed rainfall events. AGCD outperforms AWAP, showcasing its superior ability to identify rainfall events accurately.

The F1 score analysis reveals smaller variability among gauges for lower rainfall amounts, suggesting consistent performance. However, as rainfall intensity increases, variability in F1 scores also rises, indicating greater performance discrepancies for higher rainfall levels.

- **Do significant differences emerge in the categorisation of rainfall events based on distinct intensity thresholds when utilising different data sources?**

The significant differences exist in categorising rainfall events when using various data sources. AGCD, for instance, showcases better precision in identifying a wide spectrum of rainfall intensities compared to AWAP and GPM-IMERG. It excels in categorising both minimal and extreme rainfall events, demonstrating higher accuracy, especially in distinguishing heavier rainfalls exceeding the 50 mm threshold. Categorisation discrepancies are minimal for AWAP and AGCD, compared to gauge observations, with greater than 90% agreement in classifying rainfall intensity of different thresholds.

GPM-IMERG, on the other hand, exhibits limitations in accurately identifying smaller and larger rainfall events, showcasing weaker precision in the 0.2 mm to 9 mm and above 75 mm categories. This data source struggles in categorising different intensities, which might affect the classification of rainfall events into light, moderate, or heavy categories.

These disparities among data sources in precision and accuracy in identifying varying rainfall intensities contribute to differences in categorising rainfall events, potentially influencing the classification into different intensity levels like light, moderate, or heavy rainfall.

- **How do these discrepancies impact the overall understanding of rainfall patterns?**

GPM-IMERG occasionally misclassifies rainfall due to its satellite-based estimation's limitations. These discrepancies among data sources have a notable impact on the overall comprehension of rainfall patterns. Variations in precision and accuracy in categorising rainfall events influence the representation of rainfall patterns, potentially altering the perceived distribution, frequency, and intensity of rainfall.

For instance, when a data source struggles to accurately identify smaller or larger rainfall events, it might skew the depiction of light, moderate, or heavy rainfall occurrences. This can affect the understanding of regional rainfall variability, misrepresenting the frequency or intensity of rain in certain areas.

Moreover, the discrepancies could lead to biases in assessing drought or flood conditions. If a dataset consistently underestimates or overestimates specific rainfall intensities, it might misguide drought or flood predictions, impacting decision-making processes related to water resource management, agriculture, or disaster preparedness. Overall, these discrepancies in categorising precipitation events impact the overall

understanding of rainfall patterns, potentially introducing inaccuracies in assessing rainfall distribution, intensity, and frequency, consequently influencing various sectors reliant on accurate precipitation data for planning and decision-making.

2.4. Temporal Aggregation Comparisons:

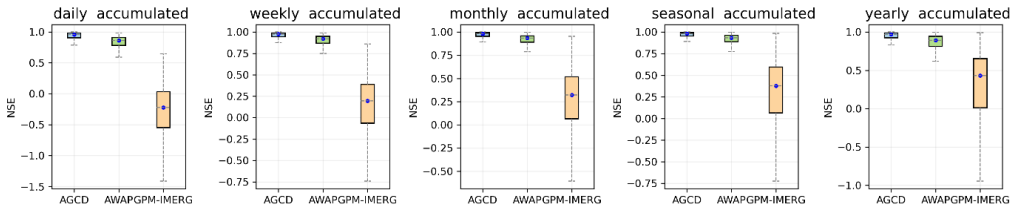


Figure 16: Boxplot showing the Nash-Sutcliffe Efficiency (NSE) Comparison between AGCD, AWAP, and GPM-IMERG Rainfall against Gauge Rainfall for different accumulation periods. Panel Layout: from Left – daily accumulated, weekly accumulated, monthly accumulated, seasonal accumulated and yearly accumulated.

When comparing AGCD, AWAP, and GPM-IMERG Rainfall against Gauge Rainfall across various accumulation periods—daily, weekly, monthly, seasonal, and yearly—it's evident that as the accumulation period extends, NSE values improve compared to daily rainfall. If opting for GPM-IMERG data, it's preferable for monthly, seasonal, or yearly verifications rather than daily, based on the observed trends in accuracy.

- **How do weekly, monthly, seasonal, and yearly accumulations of rainfall compare among different data sources?**

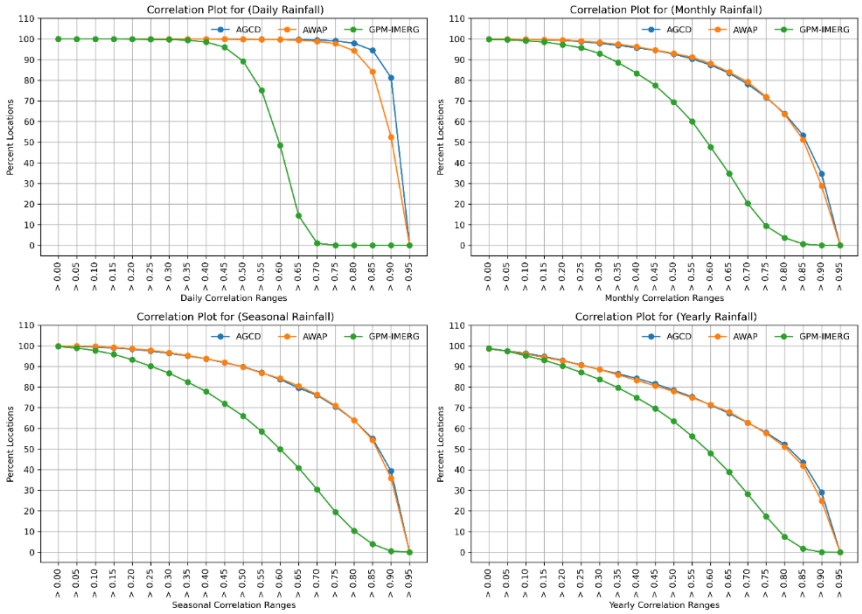


Figure 17: Plot showing the variability in Correlation between AGCD, AWAP, and GPM-IMERG Rainfall against Gauge Rainfall across different accumulation periods, presented as a percentage of locations. Panel Layout: Top Left - Daily Rainfall, Top Right - Monthly Accumulated Rainfall, Bottom Left - Seasonal Accumulated Rainfall, and Bottom Right - Yearly Accumulated Rainfall.

Considering GPM-IMERG rainfall as an example: Merely 10% of locations exhibit a correlation above 0.65 for daily data, whereas approximately 35% of locations display this level of correlation for monthly data. However, for seasonal and yearly accumulations, about 40% of locations reach this correlation threshold.

The reduction in correlation from daily to yearly rainfall accumulations can be attributed to the aggregation of random errors, which tend to average out at the daily scale but accumulate systematically over longer periods. Additionally, discrepancies in the timing and intensity of extreme events, along with potential biases in seasonal rainfall patterns, can disproportionately affect yearly totals. This reduction is further influenced by increased variance in accumulated data, which reduces the signal-to-noise ratio, thereby lowering correlation at coarser temporal scales.

- **Are there consistent variations or biases in temporal aggregation across these datasets?**

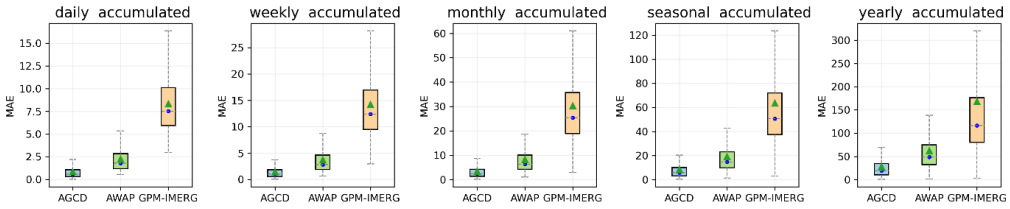


Figure 18: Boxplot showing the Mean Absolute Error (MAE) Comparison between AGCD, AWAP, and GPM-IMERG Rainfall against Gauge Rainfall for different accumulation period. Panel Layout: from Left – daily accumulated, weekly accumulated, monthly accumulated, seasonal accumulated and yearly accumulated.

Monthly accumulations show discrepancies, particularly with GPM-IMERG consistently displaying higher totals compared to gauge-based measurements, in contrast to both AWAP and AGCD rainfall.

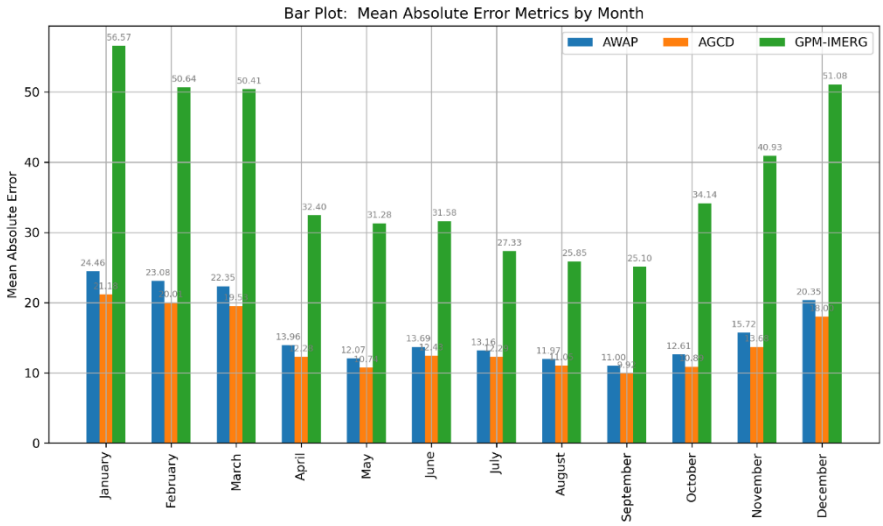


Figure 19: The Mean Absolute Error (MAE) of monthly rainfall for AWAP, AGCD, and GPM-IMERG.

2.5. Spatial Variability Analysis:

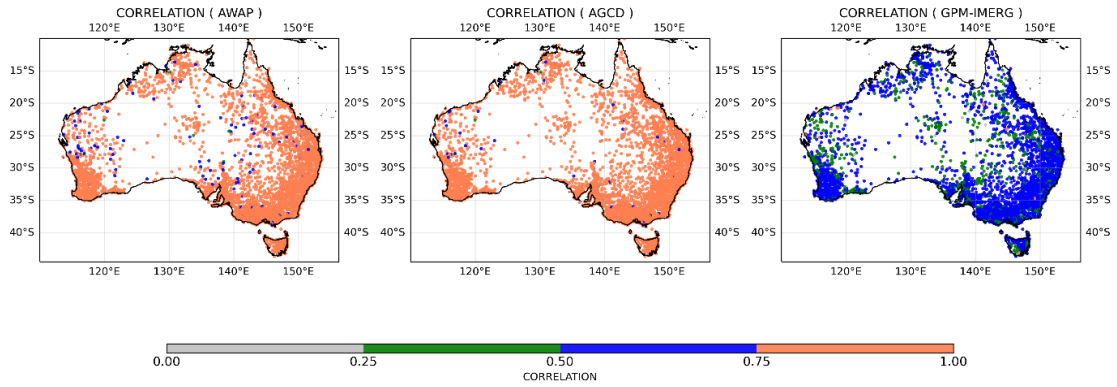


Figure 20: The correlation coefficient comparison between AGCD, AWAP, and GPM-IMERG Rainfall against Gauge Rainfall.

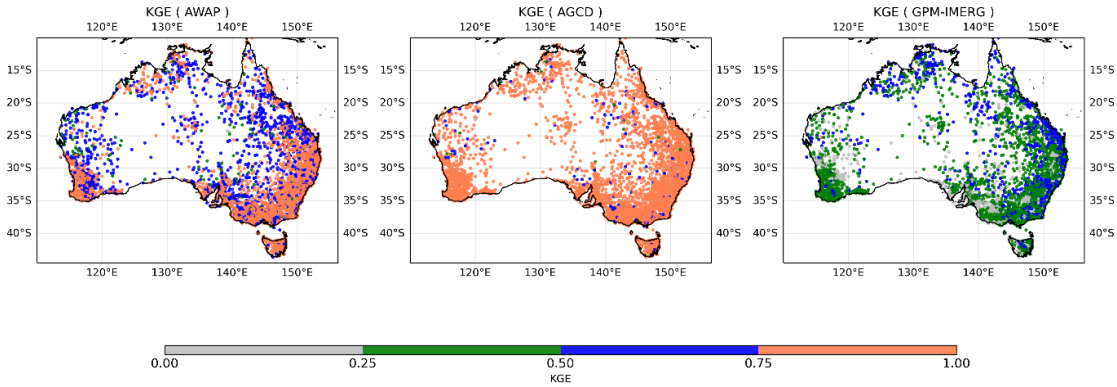


Figure 21: The Kling-Gupta Efficiency (KGE) comparison between AGCD, AWAP, and GPM-IMERG Rainfall against Gauge Rainfall.

The Kling-Gupta Efficiency (KGE) offers a comprehensive evaluation of model performance, considering correlation, variability, and bias within a single metric. Its range spans from -∞ to 1, where values closer to 1 indicate superior performance, mirroring observed data in correlation, variability, and bias. Conversely, values around 0 or negative signify poorer performance. In the case of AWAP rainfall along the east and west coasts near Perth (WA), KGE values fall between 0.75 to 1.0, indicating a high level of accuracy. In interior locations, the values range from 0.50 to 0.75, suggesting reasonably good performance. For AGCD rainfall, most locations exhibit KGE values between 0.75 to 1.0, signifying notably accurate estimations. However, in certain areas along the east coast and Darwin (NT), GPM-IMERG demonstrates KGE values between 0.50 to 0.75, while the rest of the locations indicate values less than 0.50, implying less accuracy in these regions.

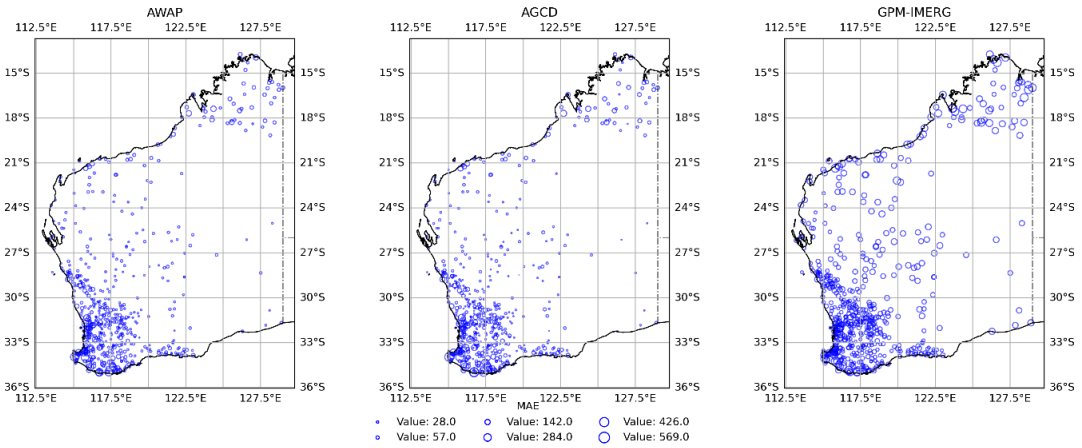


Figure 22: The Mean Absolute error (MAE) comparison between AGCD, AWAP, and GPM-IMERG Rainfall against Gauge Rainfall in Western Australia.

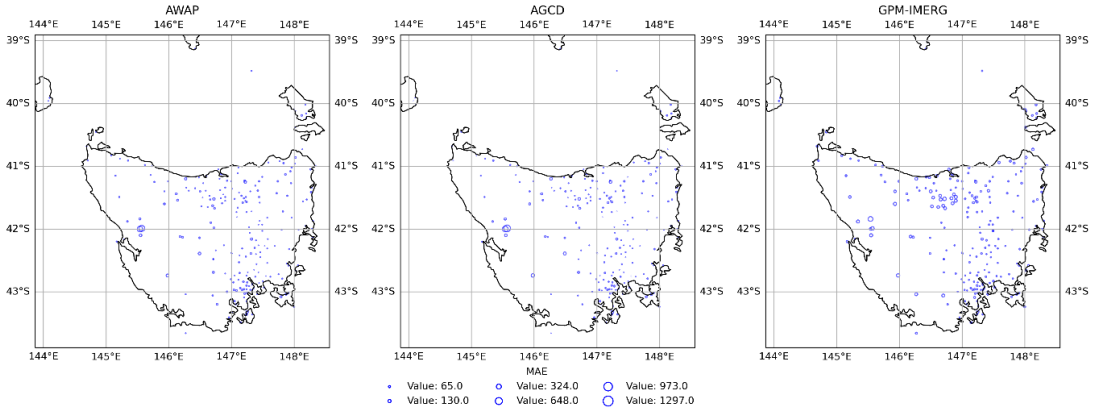


Figure 23: The Mean Absolute Error (MAE) comparison between AGCD, AWAP, and GPM-IMERG Rainfall against Gauge Rainfall in Tasmania.

- **Is there spatial variability in the performance of gridded rainfall products at Bureau and hydro gauge locations?**

AWAP and AGCD consistently demonstrate strong performance across various Bureau gauge locations, showcasing accuracy in most regions assessed.

- **How do the spatial patterns of rainfall discrepancies vary across different regions?**

GPM-IMERG exhibits higher variability in its performance, representing instances of overestimation in specific regions while slightly underestimating rainfall in others.

2.6. Temporal Variability Assessment:

- **What are the temporal variations in the accuracy of rainfall measurements among different sources?**

All datasets portray temporal variability, with fluctuations in accuracy throughout the year. AGCD displays improved accuracy in both the wet season and the dry season.

- **Do these datasets exhibit consistent or contrasting temporal trends in rainfall estimation accuracy?**

Monthly trends in rainfall errors indicate that GPM-IMERG exhibits its highest MAE errors, peaking around 50 mm/month from December to March. During the same period, AWAP and AGCD display errors ranging between 20-24 mm/month. For the other months, GPM-IMERG errors hover around 30 mm, while AWAP and AGCD errors stay between 10-13 mm. These variations suggest error fluctuations tied to monthly accumulations, notably higher during months with increased rainfall and lower during periods of lesser rainfall.

The relative bias plot reveals that GPM-IMERG exhibits elevated bias in December and January, with a subsequent peak in bias during July and August. This pattern indicates that GPM-IMERG displays higher biases during both high rainfall months and low rainfall months.

2.7. Regional Comparisons:

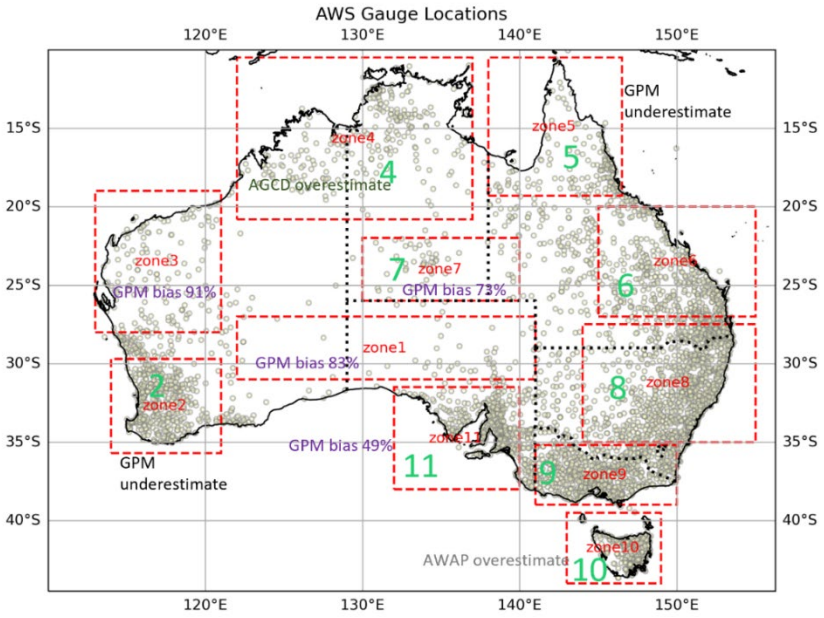


Figure 24: Division of Australia into 11 arbitrarily selected regions labelled as zone 1 to zone 11.

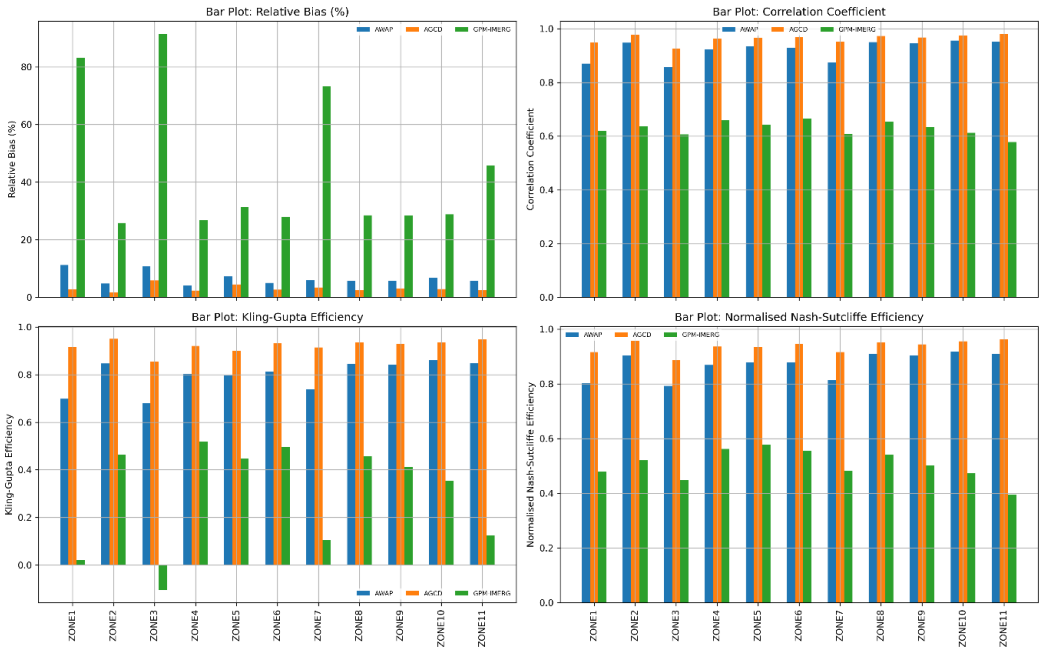


Figure 25: Plot displaying Relative Bias (%), Correlation, KGE (Kling-Gupta Efficiency), and NSE (Nash-Sutcliffe Efficiency) across 11 arbitrarily chosen regions designated as Zone 1 to Zone 11.

- **How does the accuracy of AWAP, AGCD, and GPM-IMERG rainfall vary across different regions?**

Table 4: Accuracy of daily AWAP, AGCD, and GPM-IMERG rainfall across different regions

	Relative Bias (%)			**MAE**			**KGE**		
Zone	AWAP	AGCD	GPM-IMERG	AWAP	AGCD	GPM-IMERG	AWAP	AGCD	GPM-IMERG
zone1	-11.22	-2.77	83.15	0.16	0.07	0.78	0.70	0.92	0.02
zone2	-4.84	-1.74	-25.84	0.29	0.10	1.33	0.85	0.95	0.46
zone3	-10.73	-5.87	91.36	0.31	0.16	1.17	0.68	0.86	-0.11
zone4	-4.23	2.35	26.80	1.13	0.47	2.94	0.80	0.92	0.52
zone5	-7.43	-4.52	-31.33	1.49	0.66	4.18	0.80	0.90	0.45
zone6	-5.06	-2.65	28.03	0.69	0.27	2.26	0.81	0.93	0.49
zone7	-6.00	3.48	73.28	0.30	0.13	1.08	0.74	0.91	0.10
zone8	-5.73	-2.59	28.50	0.58	0.22	2.31	0.84	0.94	0.46
zone9	-5.71	-3.03	28.34	0.46	0.23	1.93	0.84	0.93	0.41
zone10	6.72	-2.91	28.85	0.55	0.22	2.41	0.86	0.93	0.35
zone11	-5.71	-2.57	45.80	0.27	0.10	1.54	0.85	0.95	0.12

Relative bias can significantly vary in regions characterised by either sparse observed rainfall or a limited number of gauges compared to regions with a higher density of gauges. Regions with low observed rainfall or a sparse gauge network often experience evident relative bias.

- **Are there regional biases or discrepancies in the performance of gridded and satellite-based rainfall datasets?**

The regional MAE analysis highlights diverse error patterns across zones concerning monthly errors. For instance, zone 9 (Victoria), zone 10 (Tasmania), and zone 11 (South Australia) exhibit higher errors during mid-year, particularly around July. Conversely, Zone 4 (Darwin), zone 7 (Alice Springs), and zone 5 (Cairns) display lower errors during mid-year months and higher errors from December to March.

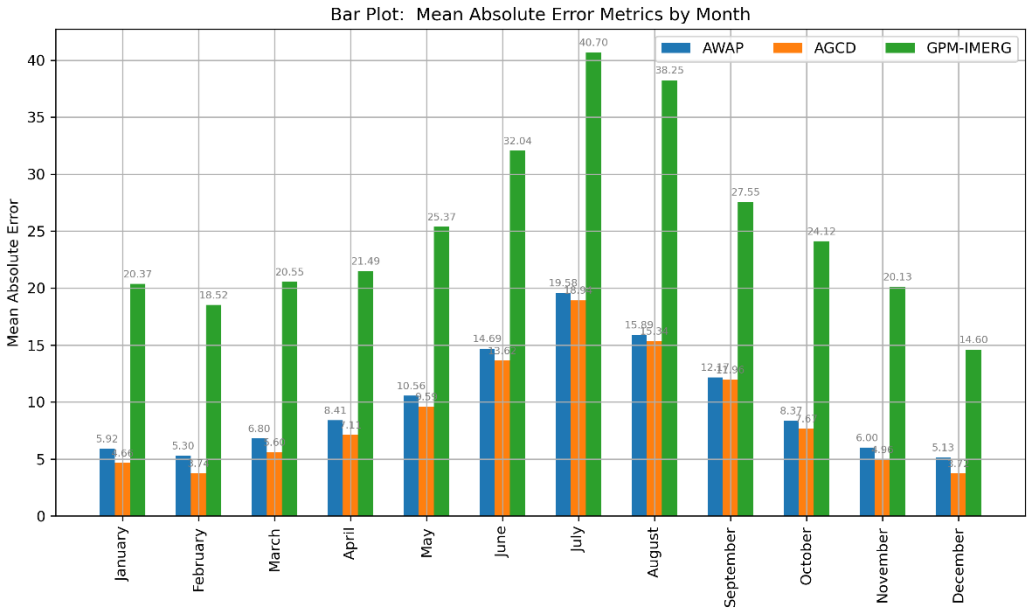


Figure 26: The Mean Absolute Error (MAE) of monthly rainfall for AWAP, AGCD, and GPM-IMERG in zone 2.

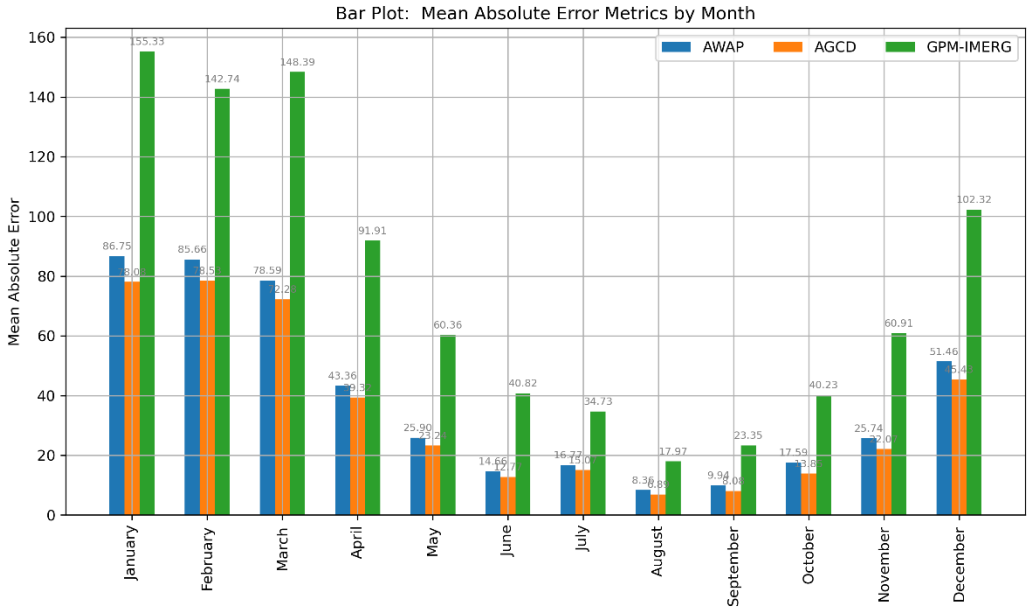


Figure 27: The Mean Absolute Error (MAE) of monthly rainfall for AWAP, AGCD, and GPM-IMERG in zone 5.

2.8. Yearly Rainfall Accumulation:

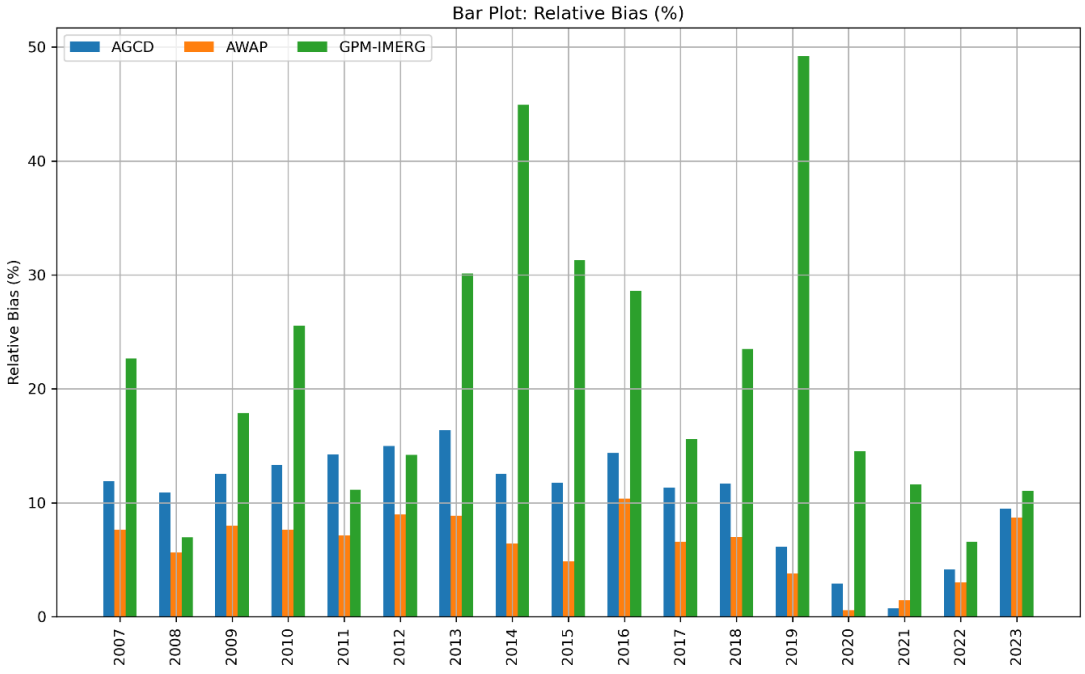


Figure 28: Relative Bias of Annual Rainfall (July -June Year) for AWAP, AGCD, and GPM-IMERG

- **How well do different datasets capture yearly rainfall accumulations, and are there consistent biases or variations?**

AWAP and AGCD consistently demonstrate their effectiveness in capturing yearly rainfall accumulations, showcasing notably lower biases compared to GPM-IMERG rainfall data.

- **Do these datasets accurately represent the long-term yearly rainfall patterns in the study area?**

GPM-IMERG tends to overestimate yearly totals compared to ground-based measurements. The annual financial year accumulation plot illustrates the relative percent bias across different years, particularly for GPM-IMERG over portions of NSW. The percent bias varies, ranging from less than 10% to approximately 48%, peaking during the dry year of 2019, followed by 2014/15 at 13% and 16% respectively. Interestingly, during the higher rainfall year of 2022-23, the bias is notably lower. Additionally, from the annual plot, it's evident that AGCD consistently exhibits a smaller relative bias compared to AWAP. Moreover, AGCD maintains a relative bias of less than 10% consistently from the years 2007 to 2023.

3. Performance Assessment of NWP QPF

This study aims to examine the accuracy of ACCESS-G4 concerning observed rainfall datasets, specifically AWAP, AGCD, and GPM-IMERG. The goal is to find how effectively ACCESS-G4 aligns with the observed rainfall dataset. A key aspect of interest involves investigating the potential impact on the overall reported accuracy of ACCESS-G4 when incorporating observed rainfall data that exhibits varying levels of accuracy. In the subsequent section, the comparative accuracy of ACCESS-G4 is evaluated against diverse observed rainfall datasets, including AWAP, AGCD, and GPM-IMERG, with a focus on a one-year data timeframe (2022 July – June 2023). To ensure a fair comparison, all rainfall datasets were re-gridded onto the ACCESS-G4 model grid, accounting for differences in spatial resolution and alignment. This inquiry seeks to elucidate the potential influence of using less accurate rainfall data, specifically GPM-IMERG, on the reported accuracy of ACCESS-G4.

3.1. Impact of observation dataset errors (biases and random errors) on ACCESS-G's apparent performance:

In the western Tasmania and eastern Victoria regions, along the east coast, and parts of Queensland, particularly from Townsville to Cairns, also the southern part of Western Australia spanning from Perth to Albany, the eastern side of South Australia, significant rainfall was observed in 2022-23. GPM-IMERG recorded lower rainfall amounts compared to AWAP/AGCD in these areas. Conversely, in the rest of the country's interior regions, GPM-IMERG rainfall surpassed AWAP/AGCD rainfall.

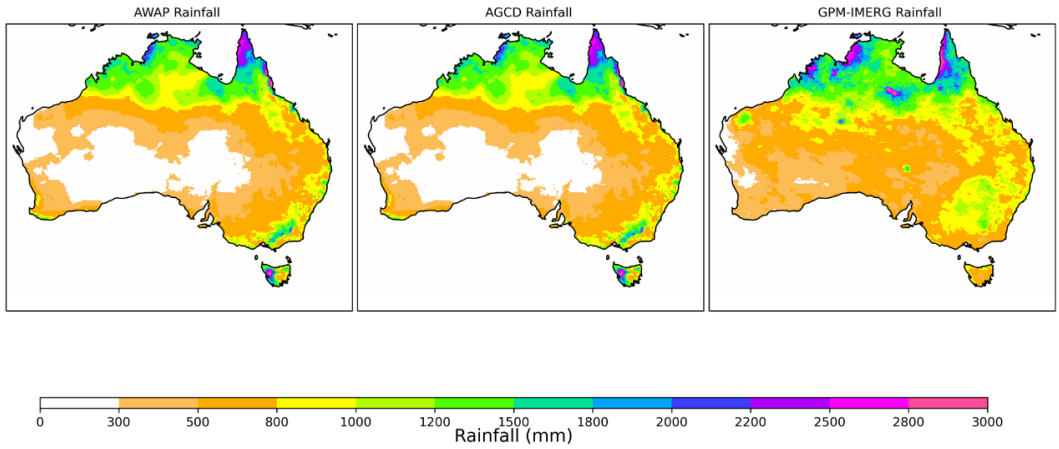


Figure 29: AGCD, AWAP, GPM-IMERG total rainfall (10 July 2022 - 30 June 2023), re-gridded to ACCESS-G4 grid resolution.

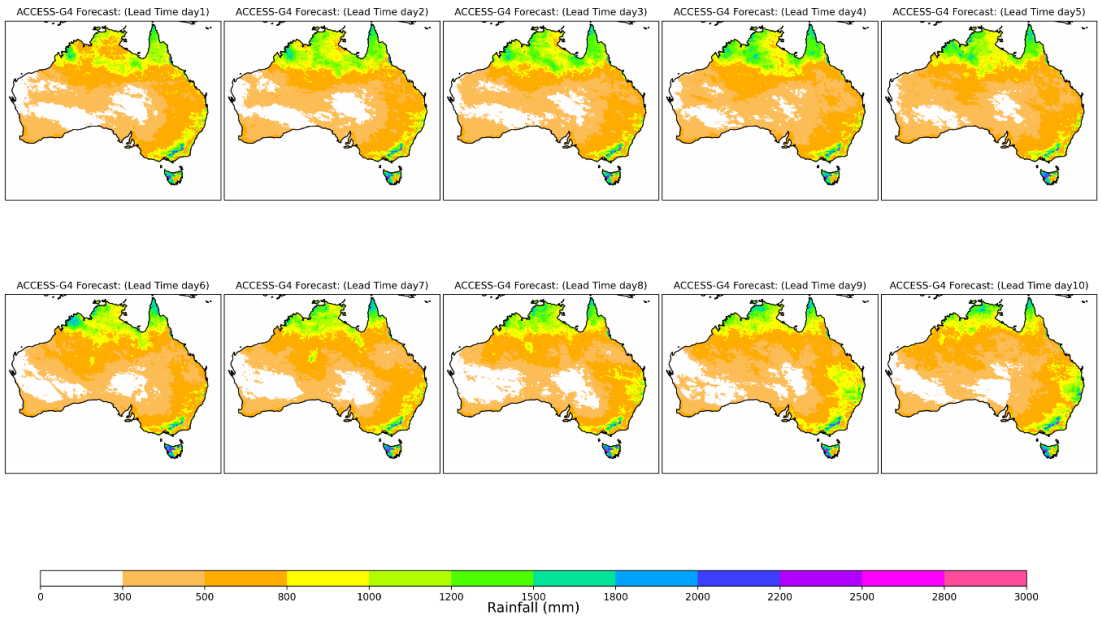


Figure 30: ACCESS-G4 total rainfall (10 July 2022 - 31 June 2023) along the lead times (day 1 to Day 10).

Overall, the ACCESS-G4 forecast indicates reduced rainfall amounts in comparison to AGCD/AWAP, particularly evident in the northern regions near Cairns, specific areas of the Northern Territory, parts of Western Australia around Darwin, segments of the east coast, Victoria, Western Australia, and southern Tasmania. Conversely, in certain interior locations, the total rainfall slightly exceeded both AWAP and AGCD amounts.

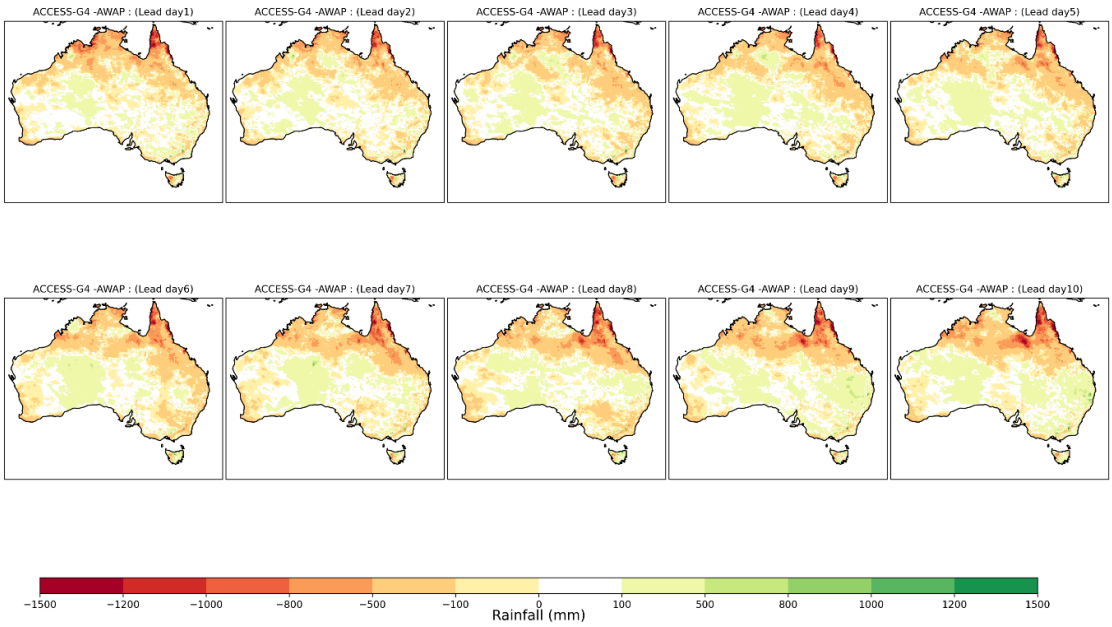


Figure 31: Variation (ACCESS-G4 minus AWAP) in total rainfall (10th July 2022 - 31st June 2023) between ACCESS-G4 and AWAP across different lead times (Day 1 to Day 10).

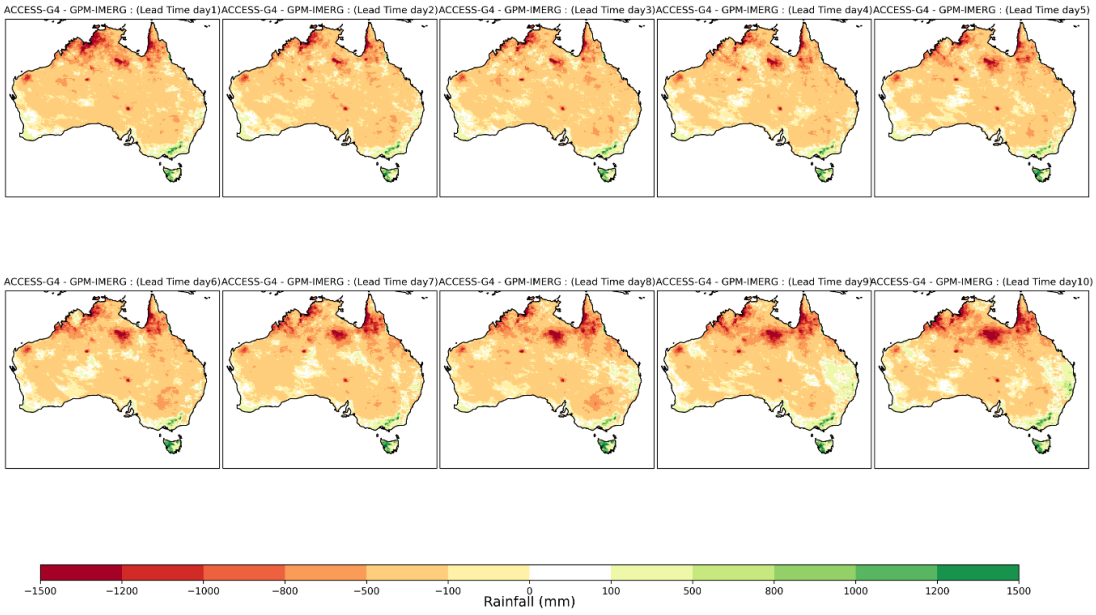


Figure 32: Variation (ACCESS-G4 minus GPM-IMERG) in total rainfall (10th July 2022 - 31st June 2023) between ACCESS-G4 and GPM-IMERG across different lead times (Day 1 to Day 10).

Likewise, in the western region of Tasmania, the ACCESS-G4 forecast predicts increased rainfall in contrast to the GPM-IMERG data. This trend is observed similarly in the southeastern part of Victoria and the southern area of Western Australia, where the ACCESS-G4 forecast predicts higher rainfall compared to GPM-IMERG. On the

contrary, in the northern areas, particularly near Cairns (QLD), Darwin (NT) and specific regions of the Northern Territory, the ACCESS-G4 forecast indicates decreased rainfall compared to the GPM-IMERG data.

- **How does the spatial variability observed in GPM-IMERG, AWAP, and AGCD rainfall data influence the measured accuracy of ACCESS-G4 forecasts across different regions in Australia?**

The spatial variability observed in AWAP, AGCD and GPM-IMERG, rainfall data significantly impacts the accuracy and reliability of ACCESS-G4 forecasts across various regions in Australia. In regions where AWAP, AGCD and GPM-IMERG data exhibit consistent patterns, ACCESS-G4 forecasts tend to align well and demonstrate higher accuracy and reliability. However, discrepancies in rainfall patterns between these datasets, such as when GPM-IMERG records higher rainfall in some regions while AWAP and AGCD indicate lower values, can challenge the reliability of ACCESS-G4 forecasts. These discrepancies might affect the reported verification metrics, especially in areas where the differences between observed data and forecasted values are significant. Therefore, the spatial variability in the observed rainfall among these datasets directly influences the performance metrics of ACCESS-G4 forecasts in different Australian regions.

3.2. Spatial Error when using Different Observational Datasets:

- **How do AWAP, AGCD, and GPM-IMERG data reflect rainfall patterns in different Australian regions and impact the accuracy of ACCESS-G4 forecasts?**

Distinct rainfall patterns are observed across various regions in Australia based on AWAP, AGCD, and GPM-IMERG data:

Northern Australia: Generally, areas near Darwin and certain parts of the Northern Territory exhibit higher rainfall in GPM-IMERG compared to AWAP and AGCD.

Western Australia: The southern part (Perth to Albany) displays lower GPM-IMERG rainfall compared to AWAP and AGCD. Conversely, some inland regions might experience higher GPM-IMERG rainfall than AWAP/AGCD, indicating varied patterns within the state.

Southeastern Australia: Regions like southeastern Victoria and southern Tasmania tend to record lower GPM-IMERG rainfall compared to AWAP and AGCD data.

Eastern Australia: Along the east coast, particularly near Cairns and parts of Queensland, GPM-IMERG records lower rainfall than AWAP and AGCD, indicating a consistent trend in reduced rainfall in these regions.

Additionally, in the country's interior regions, GPM-IMERG rainfall is higher than AWAP/AGCD values, highlighting contrasting patterns within Australia.

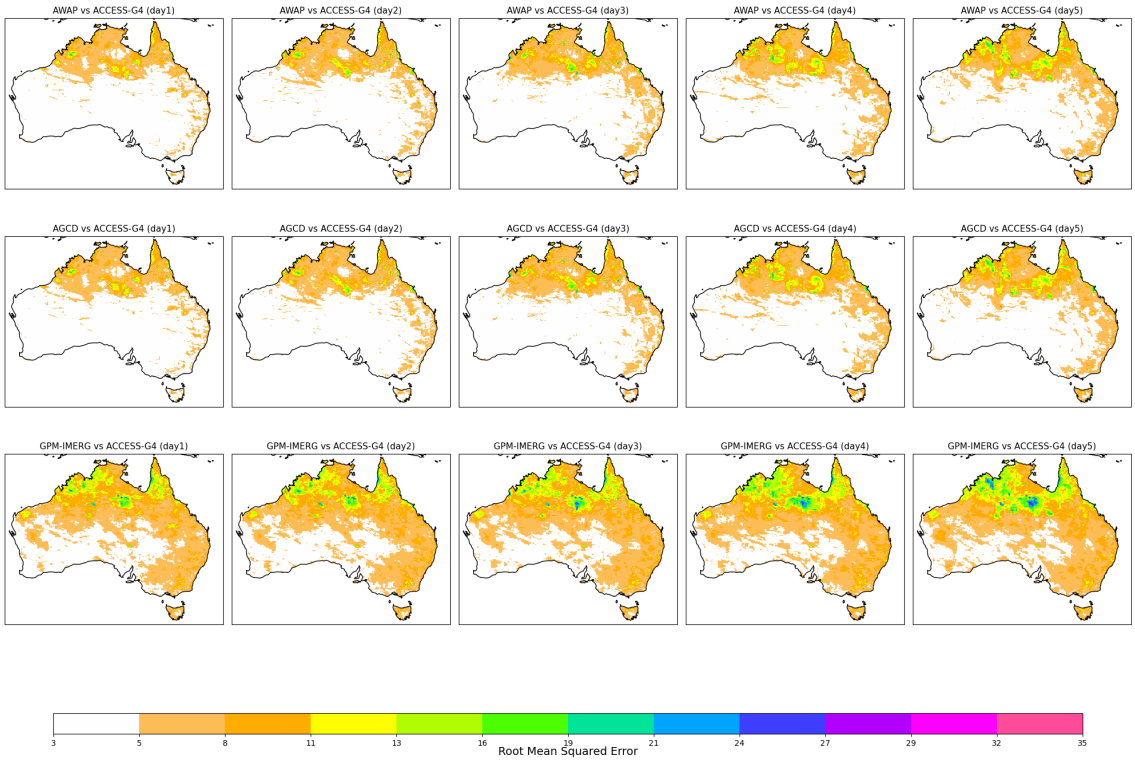


Figure 33: RMSE of ACCESS-G4 rainfall across different lead times (Day 1 to Day 5), top row AWAP vs ACCESS-G4, middle row AGCD vs ACCESS-G4 and bottom row GPM-IMERG vs ACCESS-G4.

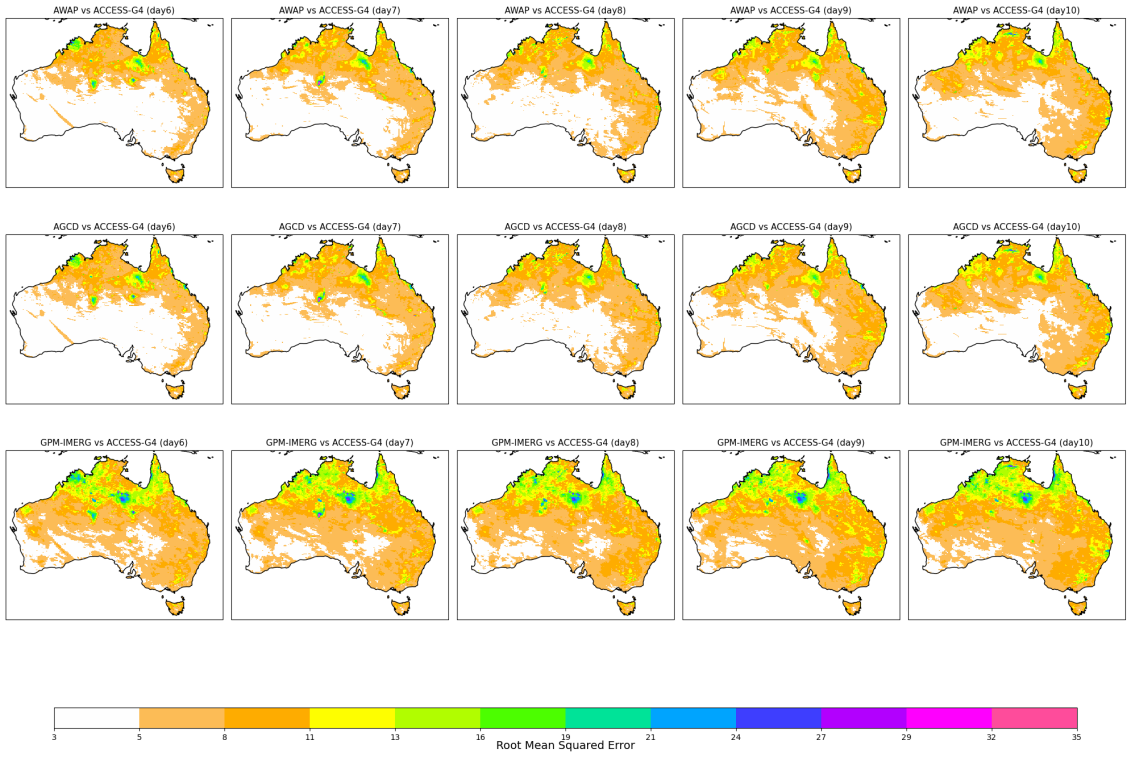


Figure 34: RMSE of ACCESS-G4 rainfall across different lead times (Day 6 to Day 10), top row AWAP vs ACCESS-G4, middle row AGCD vs ACCESS-G4 and bottom row AGCD vs GPM-IMERG vs ACCESS-G4.

As the forecast lead times extend, there is an apparent increase in the Root Mean Square Error (RMSE), signifying a decrease in ACCESS-G4's accuracy over time. This trend of increasing RMSE across all observational datasets indicates that the model becomes less precise with longer lead times. Particularly noteworthy is GPM-IMERG's tendency to exhibit higher RMSE values, especially evident in the northern regions, when compared to the AWAP and AGCD rainfall datasets. This indicates a greater deviation in forecasts from the actual observations, especially evident in the northern areas when using GPM-IMERG data as a reference. This suggests that relying on less accurate rainfall data could potentially lower the reported accuracy of ACCESS-G4 even if the NWP itself is more accurate. When utilising more accurate observed data like AWAP and AGCD, ACCESS-G4's performance appears better compared to using less accurate observed data like GPM-IMERG rainfall. However, regardless of the observed data selected, the forecasts from ACCESS-G4 NWP become less accurate as lead days progress.

3.3. Categorical Performance When Using Different Observational Datasets:

- **How do varying rainfall intensities affect accuracy metrics (CSI, POD, HSS) in ACCESS-G4 across observed datasets (GPM-IMERG, AGCD, AWAP), particularly in false alarm rates during high-intensity events?**

As rainfall intensity rises, ACCESS-G4 displays diminished accuracy compared to lower intensity scenarios. Notably, GPM-IMERG's reduced accuracy in identifying higher intensity rainfall leads to increased false alarms and misses when contrasted with AWAP/AGCD rainfall datasets.

While AGCD and AWAP show similar inclinations towards various rainfall intensities, the CSI values reveal inferior performance for GPM-IMERG compared to AWAP/AGCD. As the forecast lead time extends, ACCESS-G4's CSI performance declines, indicating a reduced capacity to predict events accurately.

At identical lead times and matched rainfall intensities, GPM-IMERG presents a higher false alarm rate than AGCD/AWAP. Specifically, for more intense rainfall, GPM-IMERG's false alarm rate exceeds that of AWAP/AGCD datasets. Consequently, assessing ACCESS-G4's performance against GPM-IMERG rainfall data may result in a higher false alarm rate.

Similar patterns emerged concerning the Probability of Detection (POD) and Heidke Skill Score (HSS) across different intensities. Consistent with higher false alarm rates, GPM-IMERG demonstrates elevated false alarm rates and lower POD and HSS metrics compared to AGCD/AWAP, especially across various intensity levels. This consistent trend suggests that assessing ACCESS-G4 against GPM-IMERG rainfall might yield higher false alarms, reduced POD, and diminished HSS across different intensities.

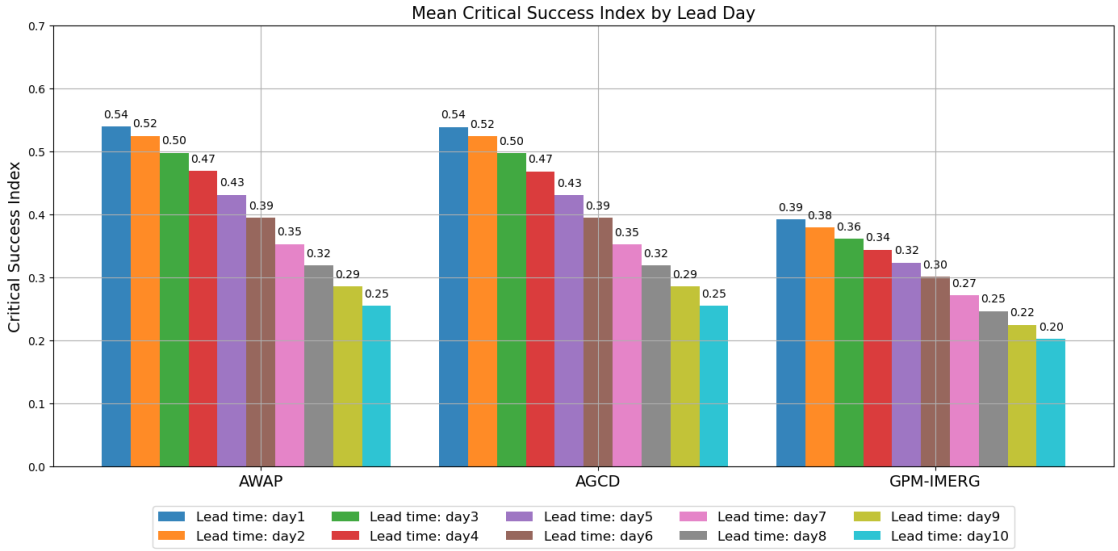


Figure 35: Average Critical Success Index (CSI) of ACCESS-G4 rainfall compared to AWAP, AGCD, and GPM-IMERG rainfall across various lead times (Day 1 to Day 10) for a 1 mm threshold, aggregated across all locations.

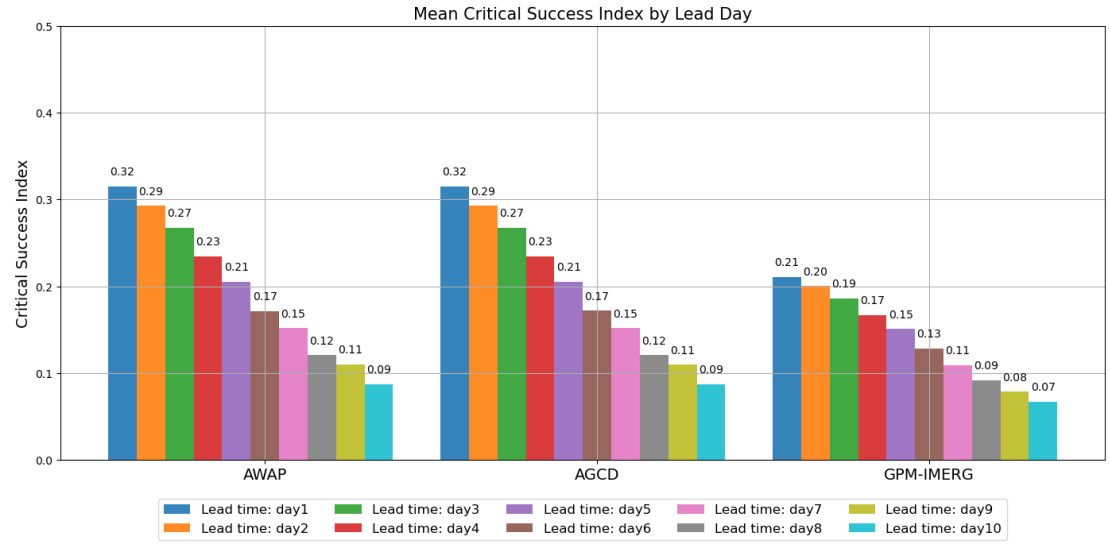


Figure 36: Average Critical Success Index (CSI) of ACCESS-G4 rainfall compared to AWAP, AGCD, and GPM-IMERG rainfall across various lead times (Day 1 to Day 10) for a 10 mm threshold, aggregated across all locations.

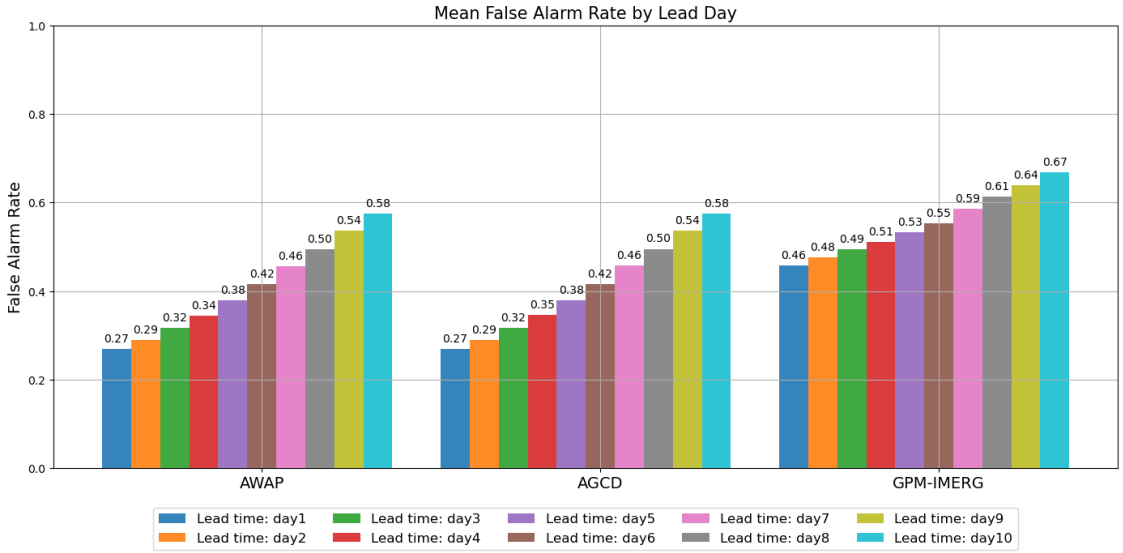


Figure 37: Average False Alarm Rates (FAR) of ACCESS-G4 rainfall compared to AWAP, AGCD, and GPM-IMERG rainfall across various lead times (Day 1 to Day 10) for a 1 mm threshold, aggregated across all locations.

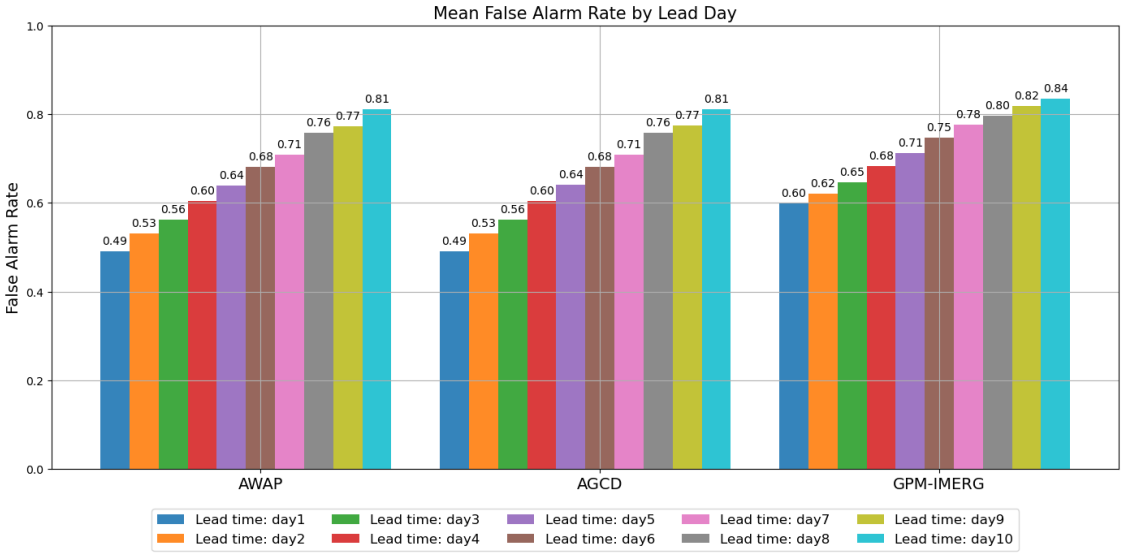


Figure 38: Average False Alarm Rates (FAR) of ACCESS-G4 rainfall compared to AWAP, AGCD, and GPM-IMERG rainfall across various lead times (Day 1 to Day 10) for a 10 mm threshold, aggregated across all locations.

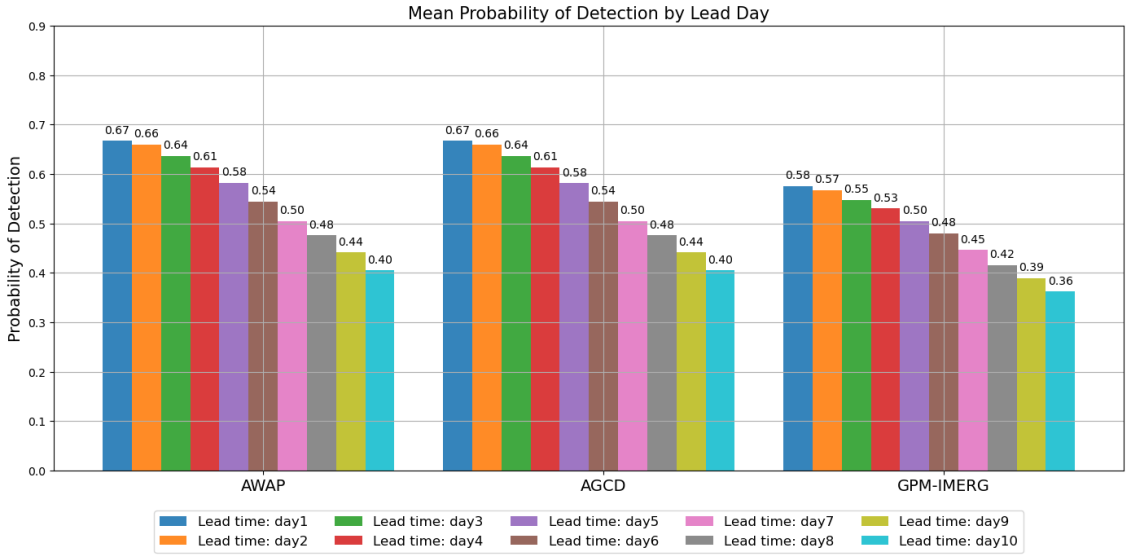


Figure 39: Average Probability of Detection (POD) of ACCESS-G4 rainfall compared to AWAP, AGCD, and GPM-IMERG rainfall across various lead times (Day 1 to Day 10) for a 1 mm threshold, aggregated across all locations.

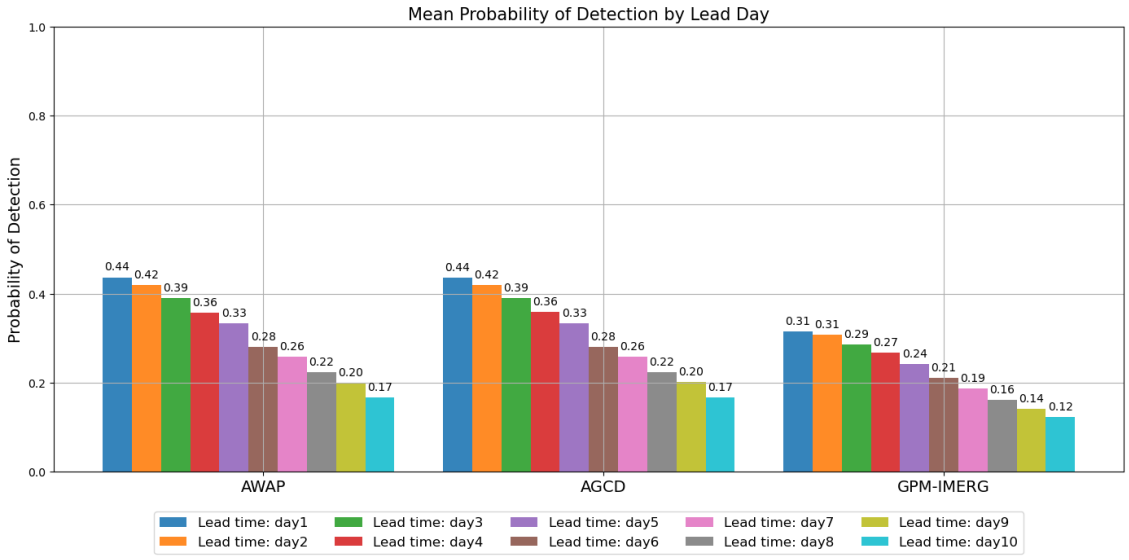


Figure 40: Average Probability of Detection (POD) of ACCESS-G4 rainfall compared to AWAP, AGCD, and GPM-IMERG rainfall across various lead times (Day 1 to Day 10) for a 10 mm threshold, aggregated across all locations.

3.4. GPM-IMERG's Limitations in Identifying Higher Intensity Events:

- **What are the specific limitations or strengths of GPM-IMERG in identifying and quantifying higher intensity rainfall events compared to AGCD/AWAP data? How do these discrepancies impact the accuracy of forecast models like ACCESS-G4, especially in terms of false alarms and missed predictions across different lead times?**

GPM-IMERG exhibits limitations in detecting higher-intensity rainfall when contrasted with AGCD/AWAP data. This deficiency leads to a higher occurrence of false alarms in the forecasts generated by ACCESS-G4. Consequently, ACCESS-G4 tends to inaccurately predict or overestimate instances of heavy rainfall due to GPM-IMERG's reduced capacity to identify such intense precipitation events accurately when compared to the AGCD/AWAP datasets.

3.5. Lead Time Impact on ACCESS-G4 Accuracy:

- **How does the performance of ACCESS-G4 change with increasing lead times, particularly regarding accuracy reduction over longer lead times?**

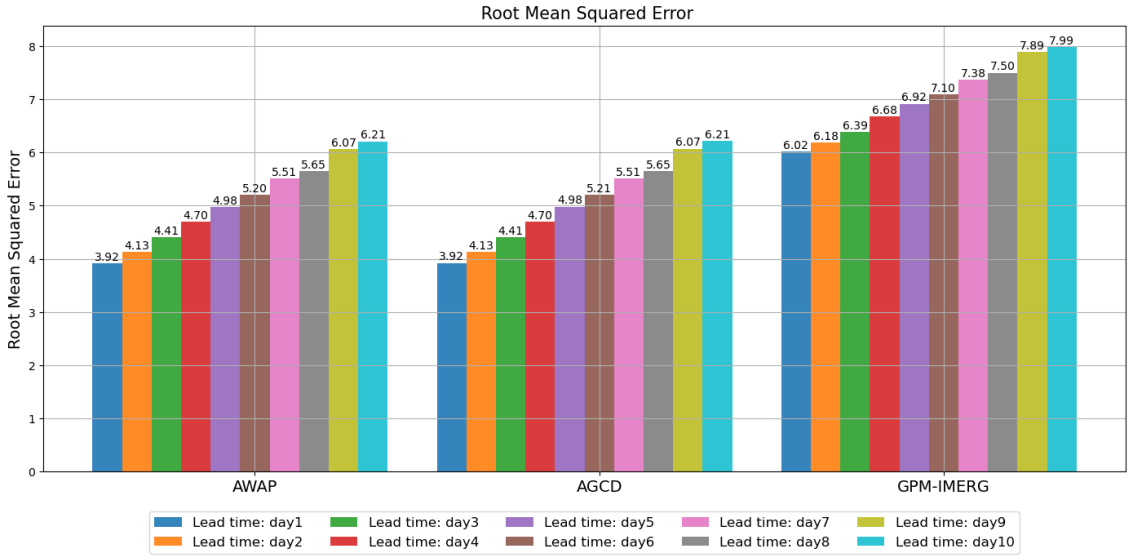


Figure 41: Average Root Mean Squared Error (RMSE) of ACCESS-G4 rainfall compared to AWAP, AGCD, and GPM-IMERG rainfall across various lead times (Day 1 to Day 10), aggregated across all locations.

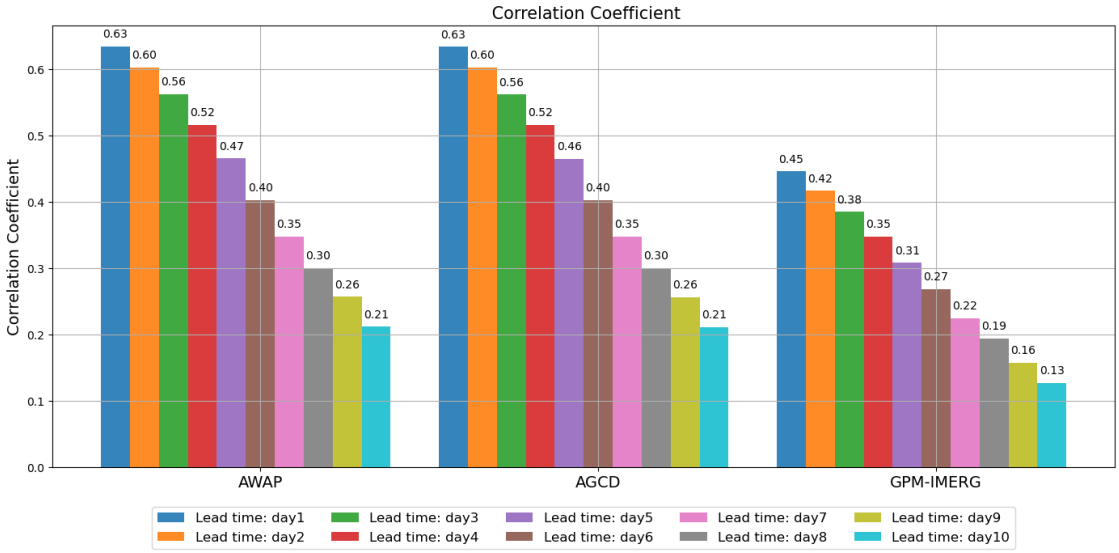


Figure 42: Average Correlation Coefficient of ACCESS-G4 rainfall compared to AWAP, AGCD, and GPM-IMERG rainfall across various lead times (Day 1 to Day 10), aggregated across all locations.

As lead times increase, the performance of ACCESS-G4 tends to reduce in terms of accuracy. Over longer lead times, the model's predictions deviate more from actual observations, leading to reduced accuracy. When compared against AWAP, AGCD, and GPM-IMERG as observed rainfall datasets, ACCESS-G4 demonstrates higher precision when evaluated with AWAP and AGCD. Reliance on GPM-IMERG, despite its wider coverage, seems to compromise the reported accuracy of ACCESS-G4 due to GPM-IMERG's lower accuracy.

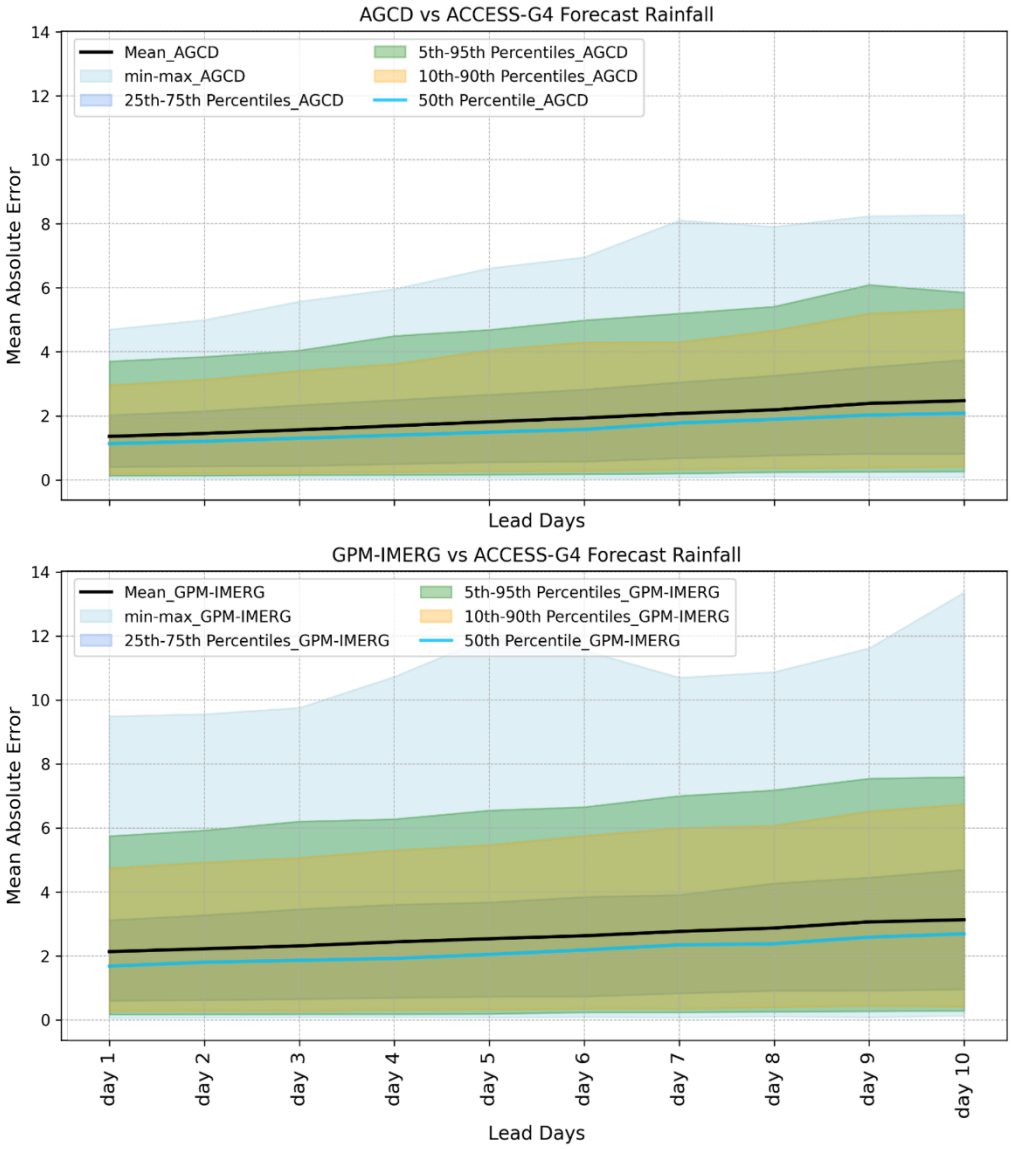


Figure 43: Mean Absolute Error (MAE) of ACCESS-G4 rainfall in comparison to AGCD and GPM-IMERG across various lead times (Day 1 to Day 10). Percentile values for MAE are derived from data encompassing all locations.

3.6. Uncertainty in ACCESS-G4 Performance Across Lead Days:

- **How does the uncertainty in ACCESS-G4's performance vary across different lead days when using AGCD/AWAP versus GPM-IMERG as observation data, and what factors contribute to the larger uncertainties introduced by GPM-IMERG?**

Notably, smaller lead days exhibit lesser uncertainty in ACCESS-G4's performance when using AGCD/AWAP as observed data. Conversely, GPM-IMERG introduces larger uncertainties, particularly in extended lead times, potentially due to its limitations in capturing extreme events accurately.

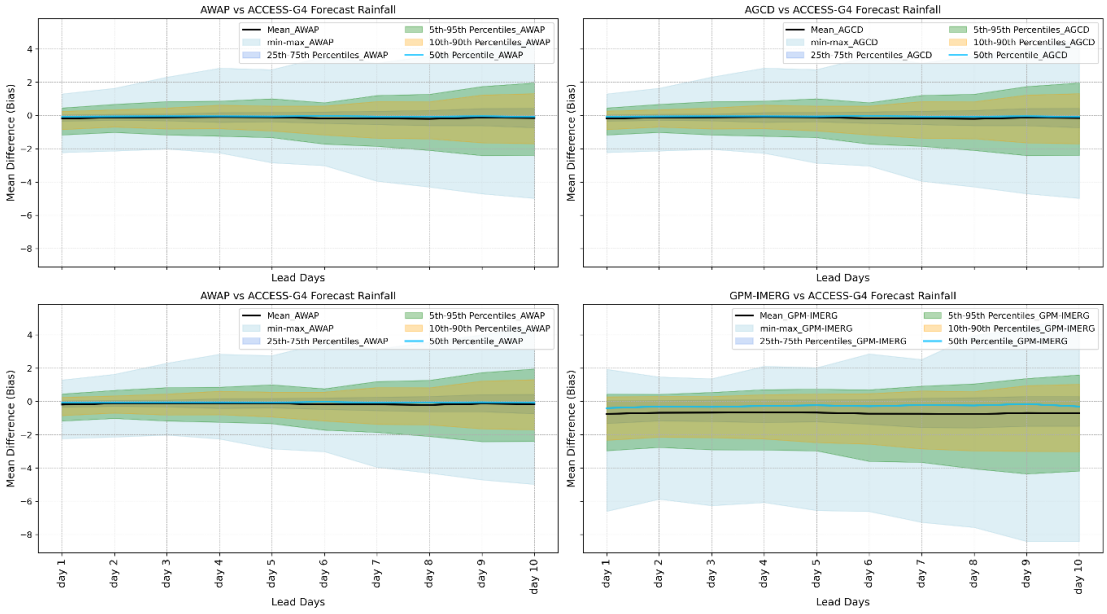


Figure 44: Mean Difference (ME) a long Lead Time for ACCESS-G4 rainfall compared to (AWAP, AGCD and GPM-IMERG rainfall. across various lead times (Day 1 to Day 10). Percentile values for ME are derived from data encompassing all locations.

As the lead time extends in ACCESS-G4 forecasts, its accuracy tends to decrease. When utilising less accurate observation data, such as GPM-IMERG, verification results demonstrate reduced accuracy or larger errors due to the amplification of observational inaccuracies and forecast errors. This amplification arises from the compounding effect of discrepancies in the observational data and inherent forecast errors over extended lead times, leading to heightened discrepancies between the forecasted and observed values. GPM-IMERG often introduces larger uncertainties due to its spatial resolution, which may not align precisely with model grid cells, alongside potential error characteristics inherent in satellite-derived rainfall estimations. These discrepancies contribute to heightened uncertainties in ACCESS-G4's performance, especially across longer lead times.

3.7. Impact of Observed Data Choice on ACCESS-G4's NWP Performance:

- **What variations exist in ACCESS-G4's performance in NWP QPF based on observed data choices (AWAP/AGCD vs. GPM-IMERG), as depicted by mean error, normalised Nash-Sutcliffe Efficiency (NSE), and correlation plots?**

Utilising more accurate observed data like AWAP and AGCD consistently showcases the superior performance of ACCESS-G4 in NWP QPF. The mean error, normalised NSE, and correlation plots consistently highlight this trend.

3.8. Consistency in ACCESS-G4's Forecast Accuracy Over Lead Days:

- **How do trends or differences in ACCESS-G4's forecast accuracy persist as the lead days progress, regardless of the selected observed data (AWAP/AGCD or GPM-IMERG)?**

Irrespective of the observed data selected, ACCESS-G4's forecasts exhibit a common pattern of decreasing accuracy as lead days progress. This trend remains consistent across different observed data sources, emphasising the challenge of longer-lead days forecasting precision.

3.9. Consistency in ACCESS-G4's Forecast Accuracy Over Spatial (Gridded) analysis and Temporal (time series) analysis

- **What insights can be gained regarding the forecast accuracy of ACCESS-G4 through spatial (gridded) analysis and Temporal (series) analysis?**

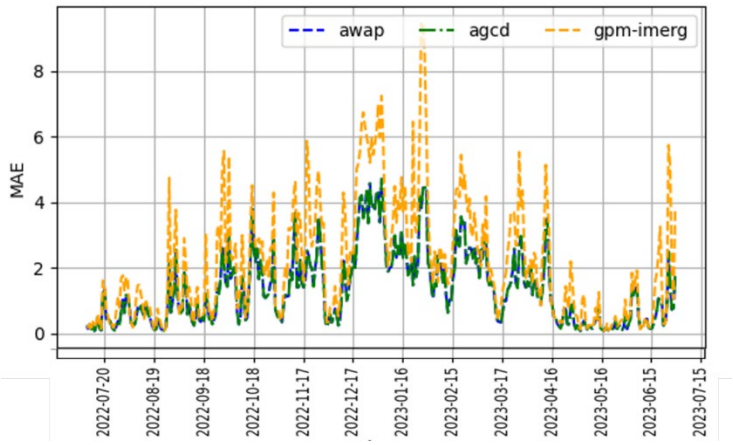


Figure 45: Figure: Mean Absolute Error (MAE) of ACCESS-G4 rainfall compared to AWAP, AGCD and GPM-IMERG for Leadtime (Day 1). Daily MAE values are obtained through spatial analysis encompassing all locations for the specified day.

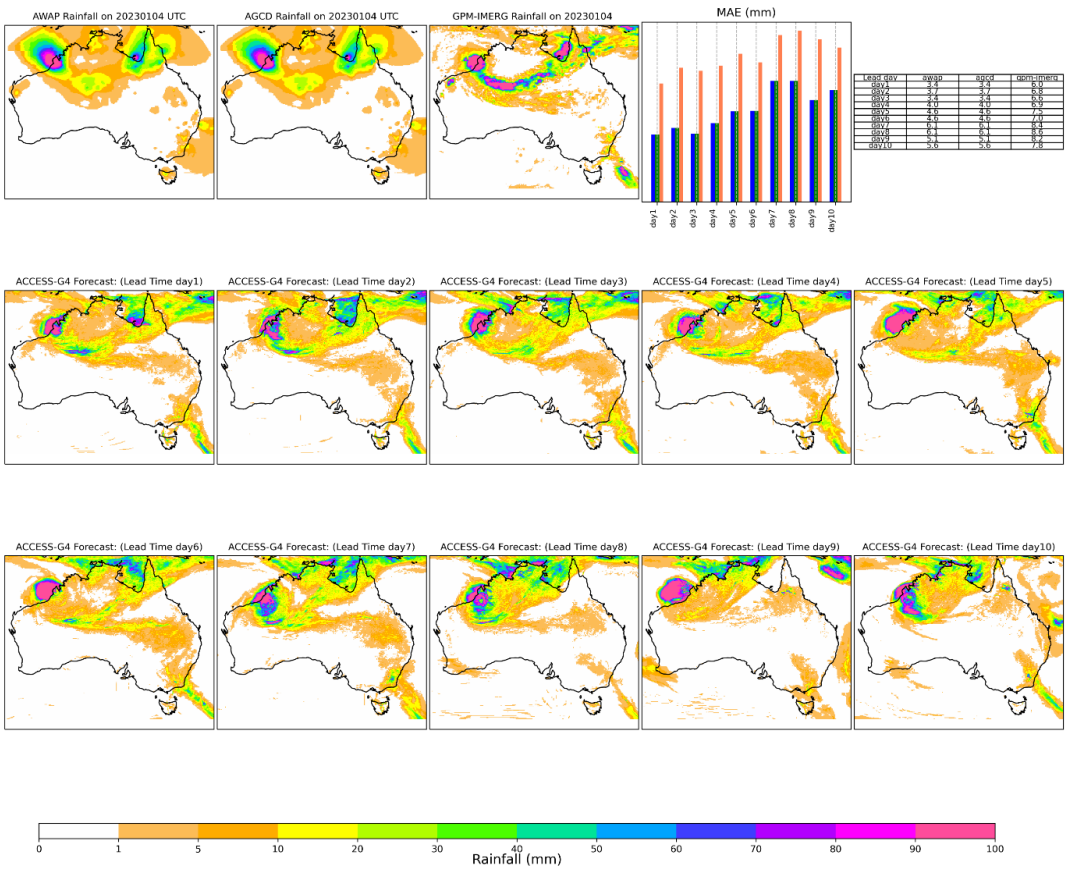


Figure 46: Spatial analysis comparing daily ACCESS-G4 rainfall with daily AWAP, AGCD, and GPM-IMERG rainfall across various lead times (Day 1 to Day 10). Observed daily AWAP, AGCD, and GPM-IMERG rainfall on the top row and daily ACCESS-G4 rainfall for lead times (Day 1 to Day 10) in the middle and last row.

Table 5: The Daily Mean Absolute Error (MAE) values are calculated through spatial analysis, considering all locations within the Australian landmass for the specified day on the above plot.

Lead Times (day)	AWAP	AGCD	GPM-IMERG
day 1	3.4	3.4	6.0
day 2	3.7	3.7	6.8
day 3	3.4	3.4	6.6
day 4	4.0	4.0	6.9
day 5	4.6	4.6	7.5
day 6	4.6	4.6	7.0
day 7	6.1	6.1	8.4
day 8	6.1	6.1	8.6
day 9	5.1	5.1	8.2
day 10	5.6	5.6	7.8

Gridded analysis involves considering all grid points within the analysis domain for the specific day, and this procedure is iterated across all dates in the timeline. Temporal analysis encompasses all data over time for the chosen points, and this process is reiterated across all points within the domain.

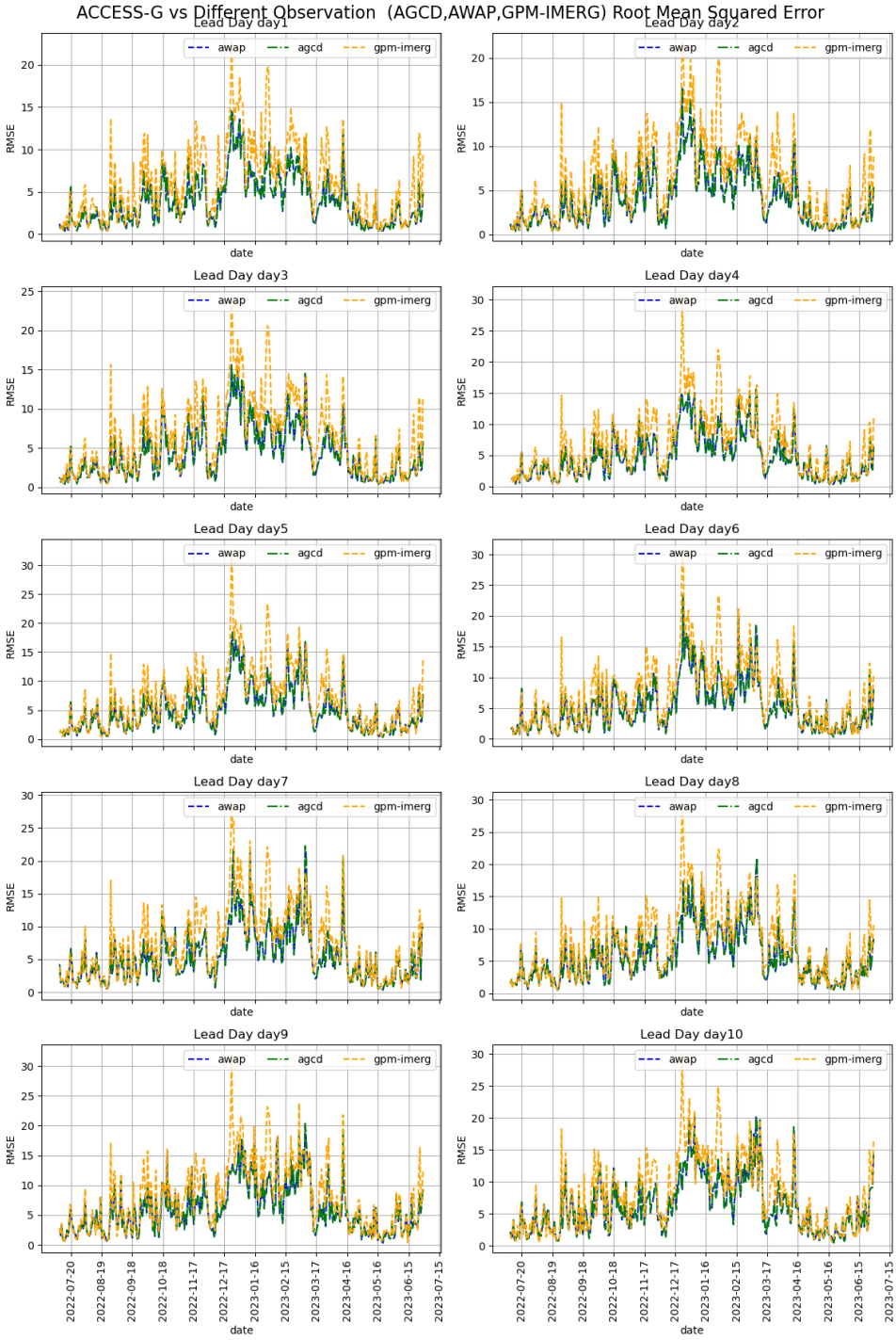


Figure 47: Root Mean Squared Error (RMSE) of ACCESS-G4 rainfall compared to AWAP, AGCD and GPM-IMERG for Leadtime (day 1 to day 10). Daily RMSE values are obtained through spatial analysis encompassing all locations for the specified day.

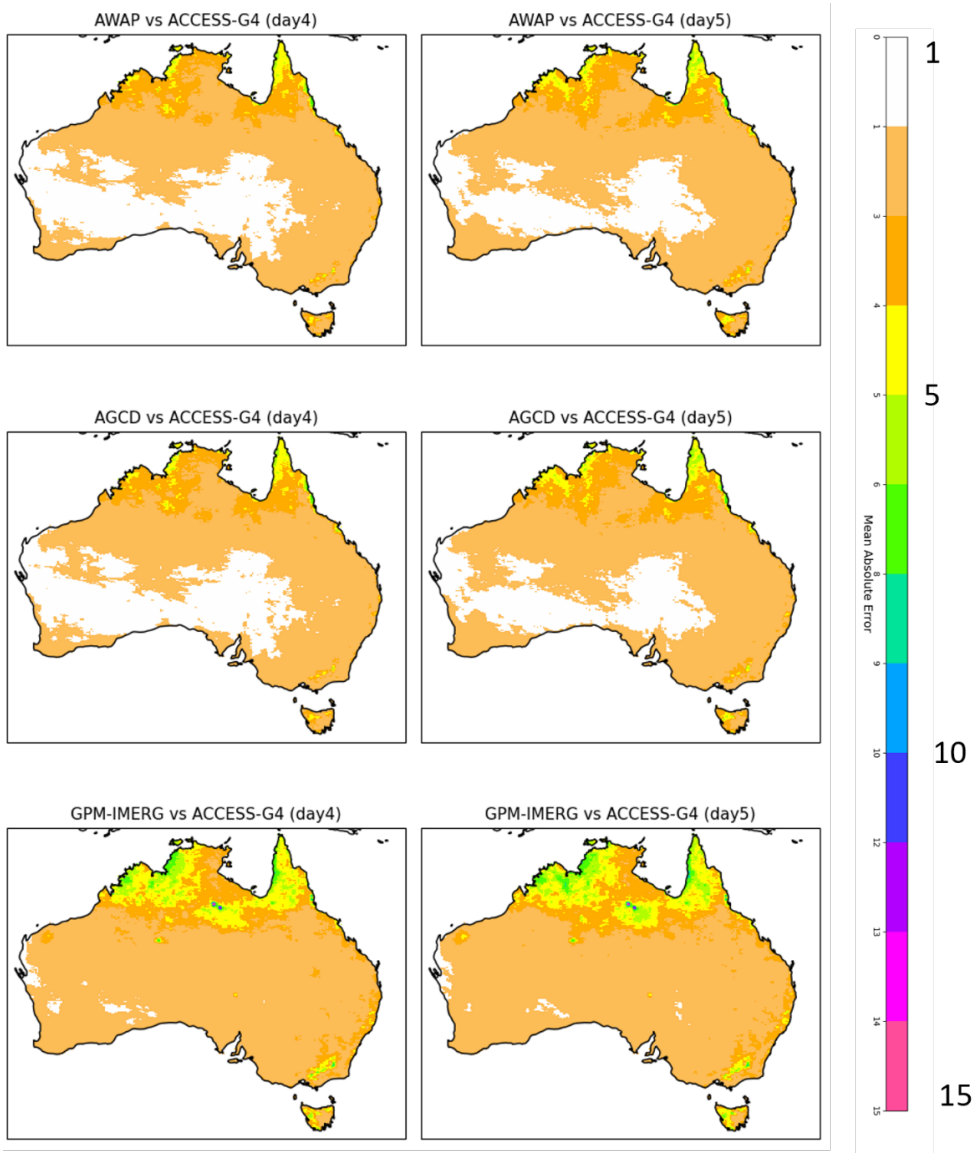


Figure 48: Mean Absolute Error (MAE) of ACCESS-G4 rainfall compared to AWAP, AGCD and GPM-IMERG for Leadtime (Day 4 and Day 5). Spatial MAE values are obtained through temporal analysis encompassing all dates for the specified location.

Spatial (Gridded) Analysis:

Analysing ACCESS-G4's forecast accuracy spatially (using gridded data) and temporally (through time series) provides distinct insights. Spatial analysis illuminates regional variations and discrepancies, while temporal analysis tracks accuracy trends over time.

When evaluating ACCESS-G4's NWP QPFS at identical lead times, it becomes apparent that employing AGCD and AWAP as observed rainfall yields greater precision compared to GPM-IMERG. This implies that utilising less accurate rainfall data for NWP QPF assessments might reduce reported accuracy, even if the NWP model itself is highly accurate. Consequently, relying on less precise rainfall data has the potential to decrease the reported accuracy of ACCESS-G4.

Temporal (Time Series) Analysis:

Through series analysis, it's evident that ACCESS-G4's overall performance depends significantly on both observed and forecasted rainfall across the continent. When GPM-IMERG serves as observed data, notably higher error values emerge, particularly in northern regions, compared to using AWAP and AGCD rainfall datasets as observations. This discrepancy becomes notably prominent in the Relative Bias (%), showcasing GPM-IMERG's higher bias values, especially in western Tasmania and southeastern Victoria, in contrast to AWAP and AGCD datasets. However, in regions with limited gauge coverage, AWAP and AGCD exhibit more bias than GPM-IMERG rainfall.

4. Summary

The analysis presented in this report highlights several key observations that significantly contribute to our understanding of the accuracy and performance of NWP QPFs. AWAP and AGCD emerge as dependable sources, consistently aligning closely with gauge observations, thereby establishing their reliability for rainfall-related applications. In contrast, GPM-IMERG, while offering global insights, tends to slightly miscalculate the distribution of rainfall, particularly during severe events.

AGCD stands out as the most accurate source. It provides rainfall measurements that closely align with ground-based observations, making it a reliable choice for various applications. AWAP, while also demonstrating good accuracy, falls slightly behind AGCD in terms of accuracy. GPM-IMERG may exhibit reduced accuracy when applied to regions with unique climate patterns, such as high rainfall areas like Eastern Australia, or in regions characterised by low rainfall, like Western Australia and South Australia.

In the categorical analysis, AGCD demonstrates a superior ability to accurately categorise both no-rain and rain events of different thresholds compared to AWAP and GPM-IMERG. On the contrary, GPM-IMERG exhibits diminished accuracy in identifying cases with no rainfall, indicating lower reliability for low rainfall analysis.

An interesting aspect revealed in this study is the impact of observed data choice on the reported accuracy of ACCESS-G4. When AWAP and AGCD are applied as observed rainfall instead of GPM-IMERG, there is an evident enhancement in the reported accuracy of ACCESS-G4, highlighting the important role of accurate observed. The temporal analysis of forecast performance shows a critical trend: regardless of the observed data selected, ACCESS-G4's forecasts exhibit diminishing accuracy as lead days progress.

Furthermore, the study shows that the GPM-IMERG's challenges in accurately identifying higher and lower intensity rainfall, resulting in an increased frequency of false alarms and misses when compared to AWAP/AGCD datasets. Consistently measuring higher false alarm rates, lower Probability of Detection (POD), and reduced Critical Success Index (CSI) metrics compared to AGCD/AWAP, GPM-IMERG poses challenges in evaluating ACCESS-G4 against its rainfall data.

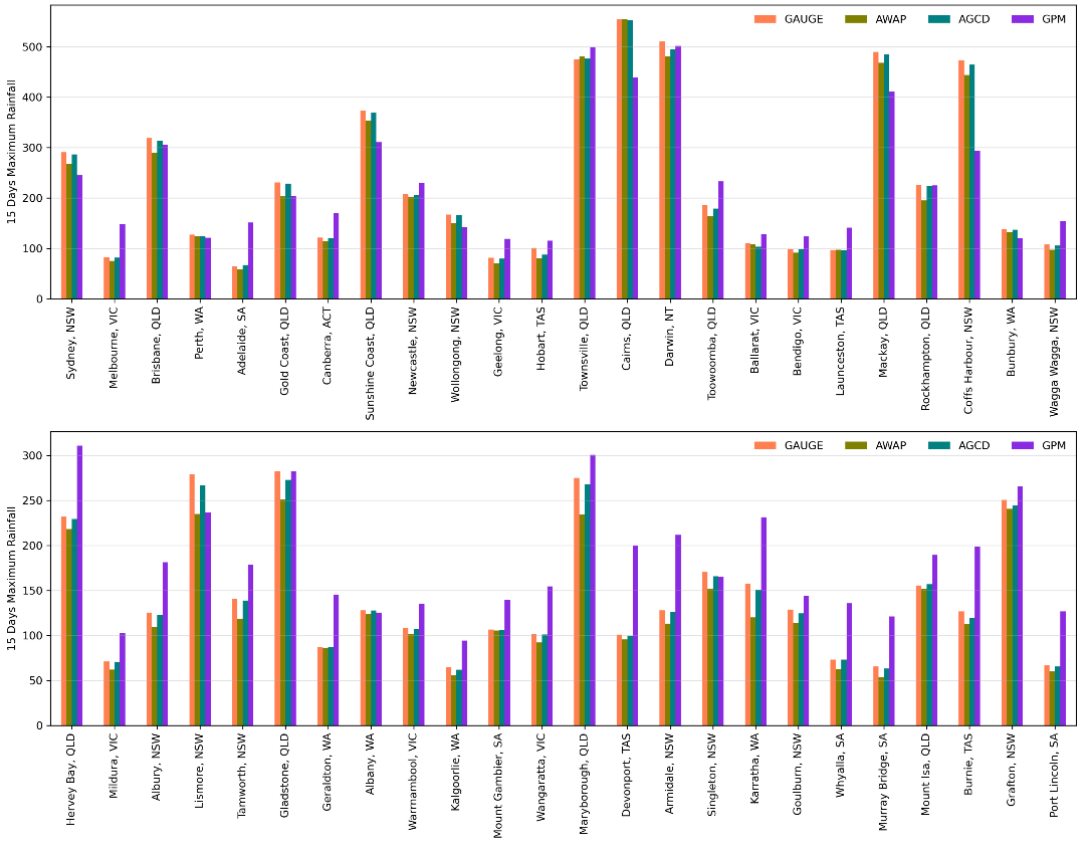
The temporal variability observed in rainfall measurements highlights the need to consider temporal accumulation over time when assessing forecast accuracy. Notably, both AWAP and AGCD sources consistently exhibit a tendency to underestimate accumulated rainfall over time, including 5-day consecutive rainfall, 15-day accumulated rainfall, annual accumulation totals, and even the 95th percentile of rainfall values. In contrast, GPM-IMERG tends to overestimate rainfall.

Regarding the number of days above the 95th percentile rainfall, GPM-IMERG generally underestimates this metric, but there's an interesting exception in Adelaide, SA, where it significantly overestimates these heavy rainfall days. The Standardised Precipitation Index (SPI) mostly leans towards underestimation, highlighting the satellite data's limitations in capturing dry or wet conditions accurately. Furthermore, GPM-IMERG demonstrates inconsistencies in estimating annual wet days, overestimating in some regions while underestimating in others. Dry days are also underestimated, particularly in Adelaide, SA, and Melbourne, VIC.

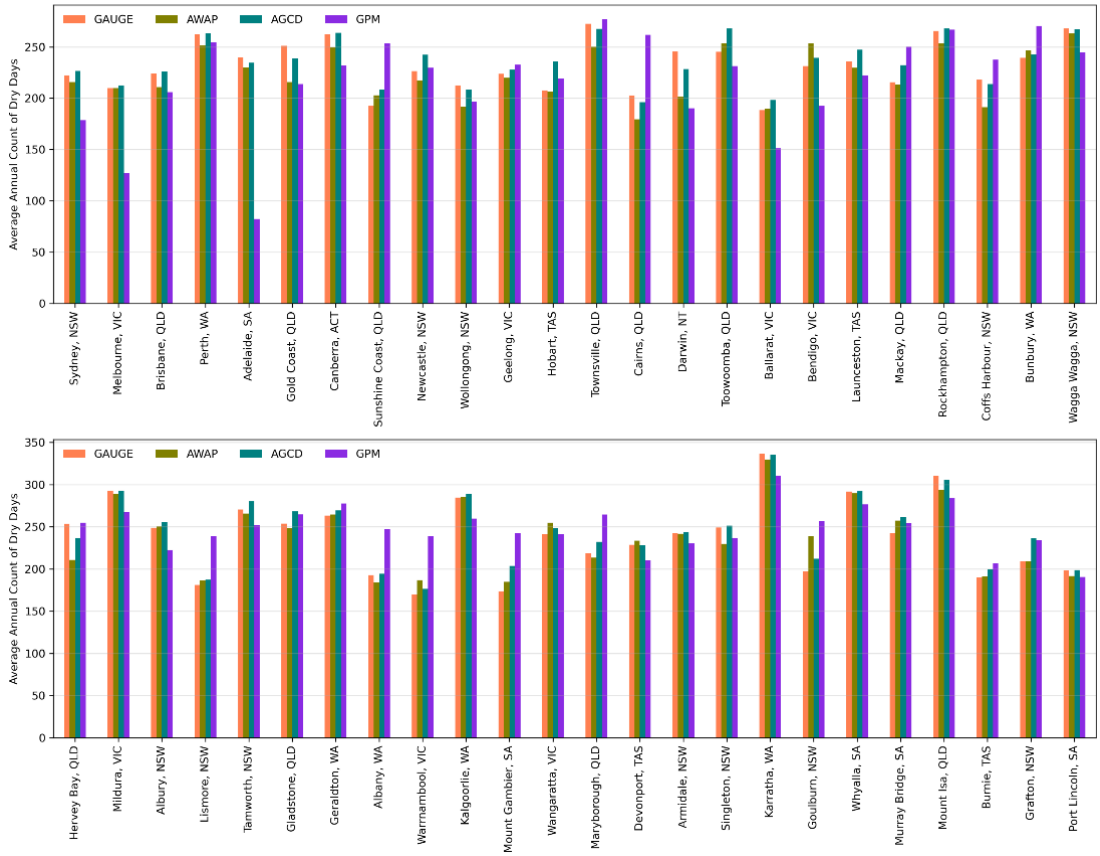
Regional variations in the accuracy of NWP models highlight the necessity of accurate observational data for correct model performance evaluation. In Tasmania's western region, GPM-IMERG rainfall tends to fall below AWAP and AGCD. Similar trends are observed near Perth in Western Australia and the southeastern part of Victoria, where GPM-IMERG rainfall remains lower than AWAP and AGCD. Conversely, in northern areas like Darwin and certain regions of the Northern Territory and central Australia, GPM-IMERG exceeds AWAP and AGCD. Along the east coast near Cairns, GPM-IMERG rainfall tends to be lower than AWAP and AGCD.

The insights provided in this report not only contribute to our understanding of the strengths and limitations of different rainfall datasets but also highlight the factors influencing the performance of NWP QPFs. As meteorological forecasting continues to advance, the findings emphasise the ongoing need for careful consideration of observed data choices, regional variations, and temporal accumulations to correctly report the performance of the NWP models in forecasting rainfall events.

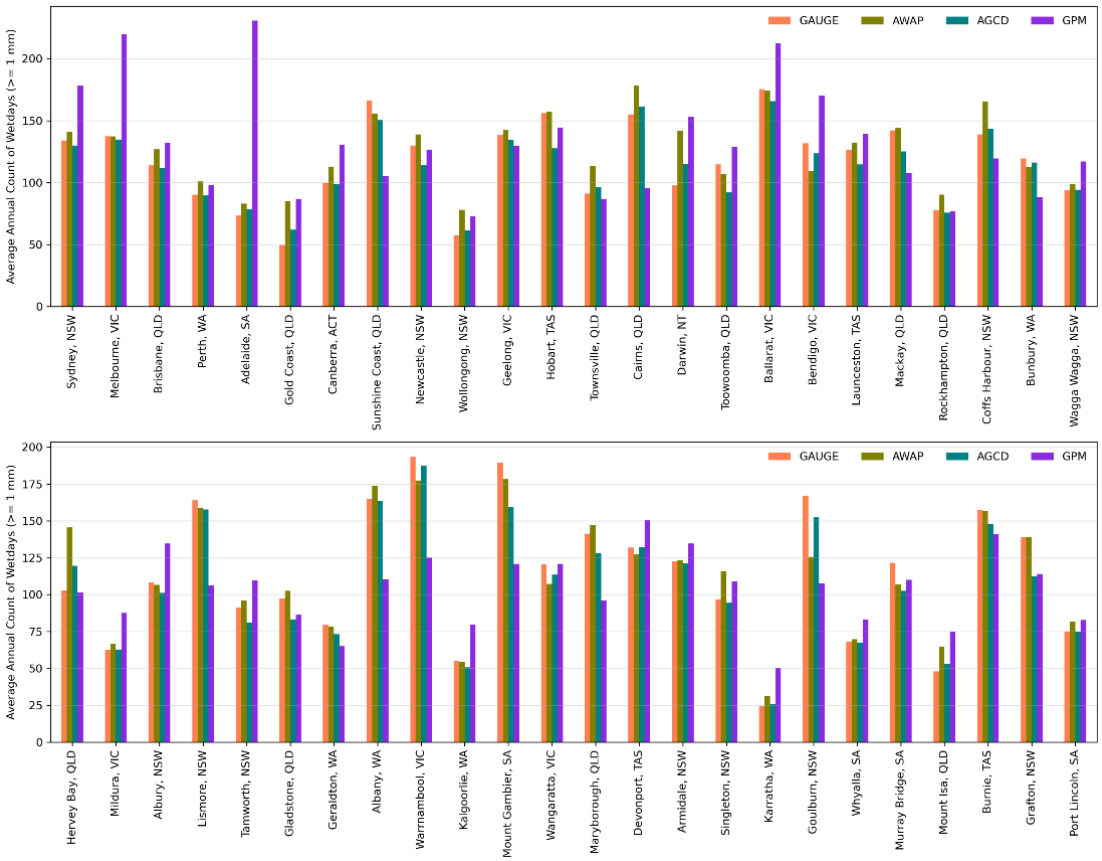
Supplementary Information



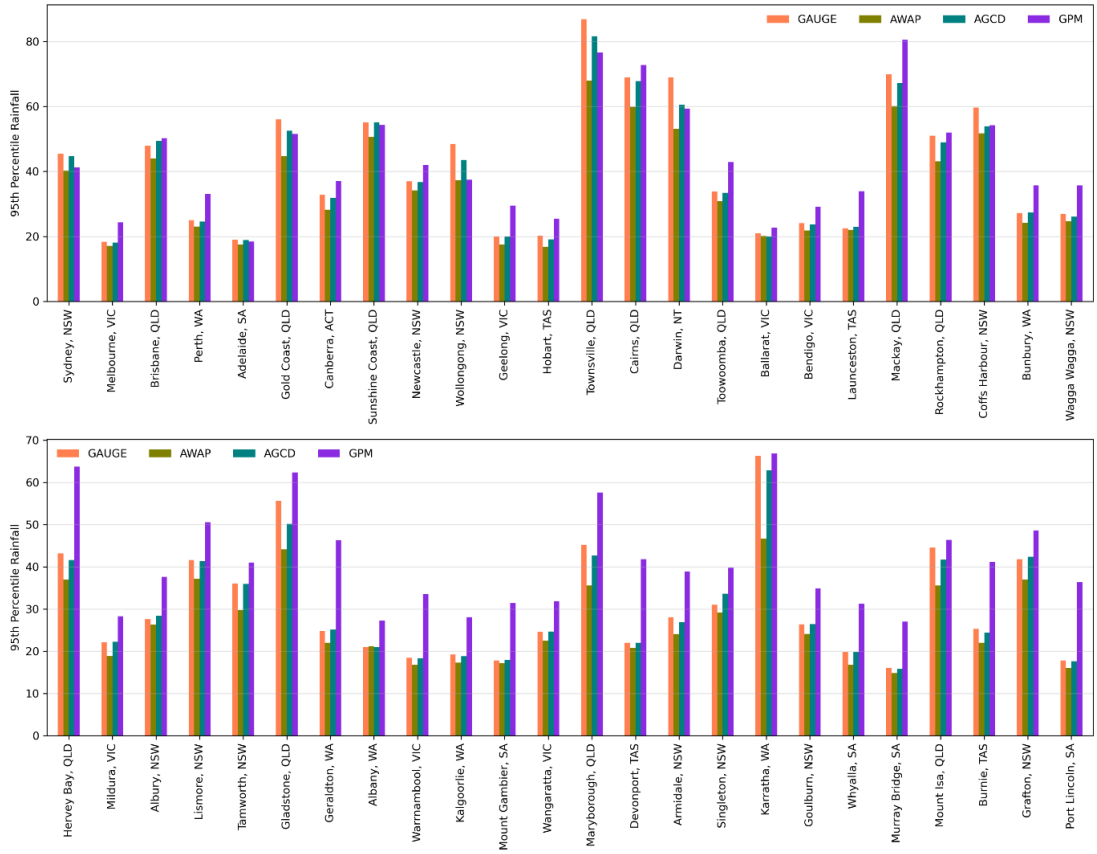
SI Figure 1: Illustration depicting average 15 days accumulated maximum rainfall from the Gauge, AGCD, AWAP, and GPM-IMERG datasets at various chosen towns (using an arbitrarily selected gauge location and the nearest grid values from gridded datasets) across Australia.



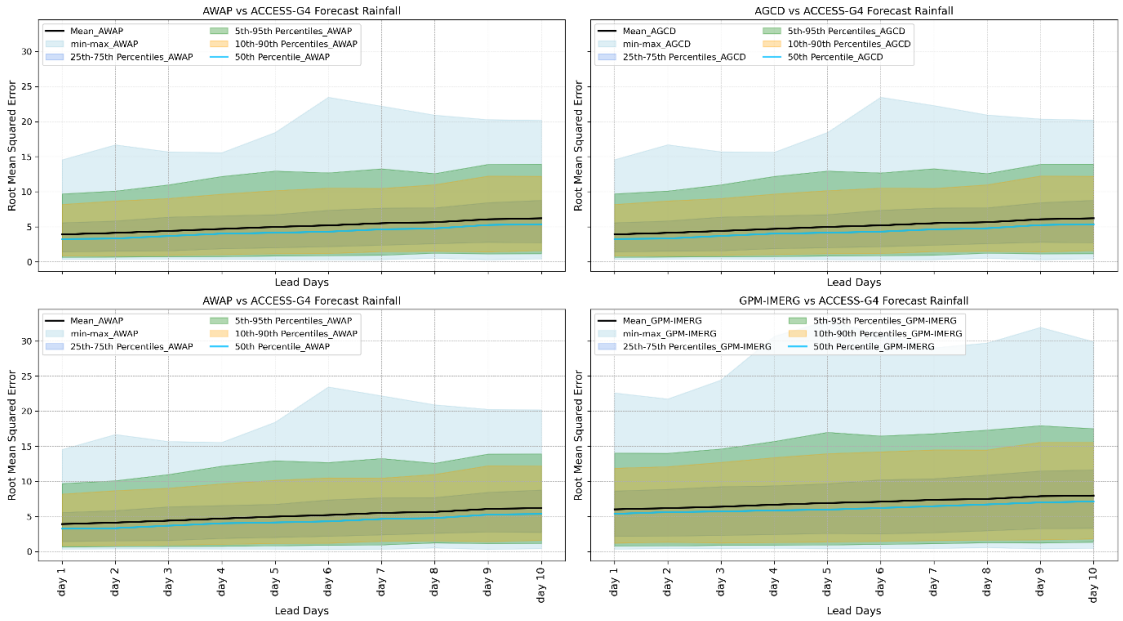
SI Figure 2: Illustration depicting average annual dry days (<1mm) rainfall from Gauge, AGCD, AWAP, and GPM-IMERG datasets at various chosen towns (using an arbitrarily selected gauge location and the nearest grid values from gridded datasets) across Australia.



SI Figure 3: Illustration depicting average annual wet days (>1mm) rainfall from Gauge, AGCD, AWAP, and GPM-IMERG datasets at various chosen towns (using an arbitrarily selected gauge location and the nearest grid values from gridded datasets) across Australia.



SI Figure 4: Illustration depicting 95th Percentile rainfall from Gauge, AGCD, AWAP, and GPM-IMERG datasets at various chosen towns (using an arbitrarily selected gauge location and the nearest grid values from gridded datasets) across Australia.



SI Figure 5:Mean Difference (ME) a long Lead Time for ACCESS-G4 rainfall compared to (AWAP, AGCD and GPM-IMERG rainfall across various lead times (Day 1 to Day 10). Percentile values for ME are derived from data encompassing all locations.

Appendix

$$MAE = \frac{\sum_{i=1}^{n}|O_i - F_i|}{n} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - F_i)^2} \quad (2)$$

$$NMAE = \frac{MAE}{\overline{O_{mean}}} \quad (3)$$

$$NRMSE = \frac{RMSE}{\overline{O_{mean}}} \quad (4)$$

$$RBias = \frac{\sum_{i=1}^{n}(O_i - F_i)}{\sum_{i=1}^{n}(O_i)} \text{x100} \quad (5)$$

$$\text{PearsonCorrelation} = \frac{\sum_{i=1}^{n}(O_i - \overline{F_{mean}})(O_i - \overline{O_{mean}})}{\sqrt{\sum_{i=1}^{n}(F_i - \overline{F_{mean}})^2}\sqrt{\sum_{i=1}^{n}(O_i - \overline{O_{mean}})^2}} \quad (6)$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(O_i - F_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O_{mean}})^2} \quad (7)$$

$$NNSE = \frac{1}{2 - NSE} \quad (8)$$

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (9)$$

Here O_i was observed rainfall, F_i was forecast rainfall $\overline{O_{mean}}$ is the mean observed rainfall, *and* $\overline{F_{mean}}$ is the mean Forecast rainfall, r is the Pearson correlation coefficient between observed and forecast values. It measures the linear relationship between the two datasets, α is a term that represents the variability of the forecast errors and is defined by the ratio of the standard deviation of the observed and forecast data ($\frac{\sigma_f}{\sigma_{obs}}$) and β is the ratio of the mean of the observed and forecast data ($\frac{\mu_{forecast}}{\mu_{obs}}$) respectively.