

1 Mapping Australia's precipitation: harnessing 2 the synergies of multi-satellite remote sensing 3 and gauge network data

4 Benjamin Hines^{1†}, Guoqi Qian^{1†} and Antoinette Tordesillas^{1*}

5 ¹School of Mathematics and Statistics, The University of
6 Melbourne, Parkville, 3010, VIC, Australia.

7 *Corresponding author(s). E-mail(s): atordesil@unimelb.edu.au;

8 Contributing authors: benjamin.hines@unimelb.edu.au;
qguoqi@unimelb.edu.au;

9 [†]These authors contributed equally to this work.

11 Abstract

12 Satellites provide a method for estimating precipitation, which is not
13 spatially restricted by the same constraints as traditional rain gauges.
14 However, satellite precipitation data while very useful, are not always
15 accurate enough to replace rain gauge data. Blending satellite and rain
16 gauge data can provide more accurate precipitation estimates than satel-
17 lites, while still maintaining the high spatial coverage. In this study we
18 propose to blend the Australian Bureau of Meteorology (BOM) rain
19 gauge network data with the satellite precipitation data — obtained
20 from Japan Aerospace Exploration Agency's (JAXA) Global Satellite
21 Mapping of Precipitation (GSMaP) and National Oceanic and Atmo-
22 spheric Administration's (NOAA) Climate Prediction Center Morphing
23 technique (CMORPH). Our data blending or fusion method consists
24 of two steps. In step 1, the relationship among the three sources of
25 data is modelled by multiple linear regression at each rain gauge loca-
26 tion, returning the least squares estimates for the associated regression
27 coefficient vector. In step 2, such regression coefficient vectors esti-
28 mates for all rain gauge locations are fitted by a spatial autoregression
29 model, whereafter the multiple linear regression coefficient vectors for
30 those locations void of rain gauges are predicted by spatial interpola-
31 tion. Subsequently, the gauge-satellite blended precipitation estimates
32 at those locations having no gauges are obtained as predictions from
33 those multiple linear regression models with the associated regression

2 Multi-platform remote sensing and in situ precipitation data

coefficients being the predictions from the aforementioned spatial interpolation. There is no methodological difficulty to extend our data fusion method to beyond Australian scope. In our Australian precipitation blending, NOAA is shown to be a more accurate source in most regions of Australia, and is thus more influential in the blending process. JAXA, while accurate in northern Australia, fails to consistently estimate the amount of precipitation in other regions at high accuracy. A clustering analysis of the blended precipitation over the last twenty years suggests two key trends on Australia's changing climate relative to the Bureau of Meteorology's six major climate zones from the previous century (1900-1999): (a) an increased spatial variability to the north, consistent with meteorological expectations, amid a southwards expansion of the wet summer dominant zones across the continent; while (b) the edge of the arid region shifts southwards and pushes out winter dominant rainfall zones across southern Australia. We name the developed platform of precipitation data fusion and interpolation as Precipitation Profiler-Observation Fusion and Estimation (PPrOFusE).

Keywords: data fusion, spatial modelling, precipitation, remote sensing

1 Introduction

Precipitation is one of the essential environmental processes having significantly controlled the way humans and communities have developed [1, 2]. Understanding the precipitation process is vital yet difficult. A changing climate brings about a complex dynamical precipitation system [3, 4]. Emergent extreme precipitation events and related hazards like floods, wildfire, and drought are occurring at higher frequency [5, 6] and intensity [7, 8] in many parts of the world. This puts significant pressure on communities and ecosystems which rely on fresh water supply and resource distribution [9]. The World bank estimates two-thirds of the world's population is likely to face water shortages by 2025 [10]. To fully understand and model the precipitation system, high quality precipitation data is required. However, the two methods for measuring precipitation data (in situ gauge measurements and remote sensing satellite estimates) have significant shortfalls [11–15]. In this paper, we aim to give a simple, yet efficient and effective, two-step method for fusing the multiplatform precipitation data to overcome the limitations which the raw data presents.

The process of data fusion involves integrating multiple data sources into a single variable, in a way that harnesses the best aspects of the individual data platforms, while minimising their weaknesses. Many studies on climate have used remote sensing data [16, 17] due to its ability to give spatial data at a high resolution and broad spatial coverage. However, this data is not necessarily an accurate representative of the ground truth [18], and as a result may negatively affect studies and policies in which it is used [19, 20]. To fully

76 exploit the advantages of remote sensing data, it is desirable to perform data
77 fusion. Numerous studies have chosen this strategy [21, 22] in order to have a
78 more accurate representation of precipitation at high spatial coverage.

79 Data fusion is a very useful technique for mapping global precipitation.
80 The different platforms for obtaining precipitation data, namely, ground rain-
81 gauge network, airborne and space satellites, each bear their own strengths
82 and weaknesses. Rain gauges, while giving us in situ measurements, are highly
83 limited in their spatial coverage capacity [23]. Each rain gauge has an asso-
84 ciated cost, installation and maintenance requirement, and is also subject to
85 terrain restrictions (e.g., cannot be used at inaccessible locations). Thus, a
86 large optimally placed rain gauge network with high spatial coverage cannot be
87 accomplished in many situations. There have been many studies done on opti-
88 mal rain gauge network placement [24, 25]. While measured at the source, rain
89 gauges are still subject to error. There are numerous reasons for this. Factors
90 such as blockages, wind direction and evaporation can all reduce a rain gauge's
91 ability to properly measure precipitation [26]. Satellite estimation of precipi-
92 tation can provide data at a higher spatial coverage than rain gauges, which
93 is also not restricted by the nature of the terrain. However, satellites can only
94 provide estimates (i.e., derived measurements) for the amount of precipitation
95 which has occurred. These estimates are subject to error due to affectedness
96 from several factors such as terrain and cloud composition [27, 28]. Detailed,
97 complete and reliable mappings of precipitation are of high importance for not
98 only precipitation modelling, but also for enhancing the early warnings sys-
99 tems for climate related risks [29–33]. To obtain the most useful and complete
100 precipitation data, fusion is a necessary approach.

101 The Australian continent presents a unique challenge when it comes to
102 measuring precipitation by rain gauge. Due to its size, and non-uniform con-
103 centration of the relatively small population, operating and maintaining a rain
104 gauge network with high spatial coverage is not feasible [34]. As a result,
105 there are very large gaps in the rain gauge network, where precipitation is
106 not measured. However, as Australia is the driest inhabited continent [35],
107 understanding of precipitation is vitally important. Thus remote sensing pre-
108 cipitation estimates are of a high necessity to fill in the gaps left from the rain
109 gauge network.

110 Our aim in this case study on Australia is twofold. The first and main aim is
111 to develop, Precipitation Profiler-Observation Fusion and Estimation (PPrO-
112 FusE), a tool to deliver high quality gauge and satellite fused precipitation
113 data. The second is to compare the satellite precipitation platforms around
114 Australia and assess their ability in accurately representing the amount of pre-
115 cipitation which has occurred. We aim to fuse Australian monthly rain gauge
116 data (provided by the Bureau of Meteorology), with monthly satellite data
117 from the Japanese Aerospace Exploration Agency and the National Oceanic
118 and Atmospheric Administration in a way which exploits the strengths of the
119 individual data set, while minimising their relative weaknesses. To fuse these
120 three data sets, we employ a two-step approach. Firstly, the relationship among

4 Multi-platform remote sensing and in situ precipitation data

the three sources of data is modelled by multiple linear regression at each rain gauge location, returning the least squares estimates for the associated regression coefficient vector. Secondly, such regression coefficient vectors estimates for all rain gauge locations are fitted by a spatial autoregression model, whereafter the multiple linear regression coefficient vectors for those locations void of rain gauges are predicted by spatial interpolation. Subsequently, the gauge-satellite blended precipitation estimates at those locations having no gauges are obtained as predictions from those multiple linear regression models with the associated regression coefficients being the predictions from the aforementioned spatial interpolation.

The paper is arranged as follows. In Section 2, we review the relevant literature about the current methods regarding the fusion of gauge and satellite precipitation data in Australia. In Section 3, we present the data for this study and discuss the two-step methodology we propose for fusing gauge and satellite precipitation data. The developed platform of precipitation data fusion and interpolation is named Precipitation Profiler-Observation Fusion and Estimation (PPrOFusE). Results and their significance are presented in Section 4. Concluding remarks are given in Section 5.

2 Current Australian Precipitation Data Fusion Methods

To put this work in better context, we will briefly discuss the literature on gauge-satellite data fusion for Australia. In the area of hydrology and climatology, data fusion is referred to as data blending. To remain consistent with the literature, we will henceforth refer to the methodology as data blending. We will not review the entire literature on gauge and satellite data blending as there have been two recent excellent review studies on this topic [36, 37]. Instead, we summarise the salient state-of-the-art approaches relevant to this work for Australia.

An approach for blending daily gauge and the Tropical Rainfall Measuring Mission satellite data involved spatially modelling the difference between the two variables at each location [38]. Several spatial models were used and their performances were compared. These models included different kriging models and also non-parametric kernel smoothing models. While the kernel smoothing method proved to be the optimal model, there were several limitations in the study. Firstly, there is no clear approach for adding explanatory variables into the models. Secondly, only 120 days of data from the year 2001 were used. Thirdly, the estimation of the uncertainty is influenced by the correlation structure.

Another study looking at gauge and Tropical Rainfall Measuring Mission satellite daily data blending in Australia was presented in [34]. A kriging based approach was taken to spatially interpolate daily rainfall to all of Australia. However, their work showed blended satellite and rain gauge data performed worse than other techniques mentioned in their paper. It was also showed

164 blending of satellite and rain gauge data significantly reduced the estimation
165 variance.

166 There have been studies more relevant to the work presented in this paper
167 by using monthly precipitation data. The work presented in [39] provided
168 a two-step method where again spatial kriging models were used. This was
169 applied to the Australian Gridded Climate Data set (a gauge substitute) and
170 the Japanese Aerospace Exploration Agency data. Twenty years of monthly
171 data was used, and their results showed their methods were successful in reduc-
172 ing the amount of bias in their estimates over region with high gauge density.
173 However, while their method can be transferred to other regions, a limitation
174 is the regional effectiveness cannot be generalized. Thus its performance in
175 regions where the rain gauge density is lower cannot be easily quantified.

176 A comparison of various correction and blending techniques for monthly
177 precipitation data for Australia was studied in [40]. The Japanese Aerospace
178 and Exploration Agency data was compared to the Australian Gridded Cli-
179 mate Data set. Again, kriging methods were used for spatial interpolation.
180 However, the results showed the choice of which kriging technique being used
181 did not have a significant impact of the performance. It was also found the best
182 performing blending technique was an inverse error variance blending method.

183 There are common themes amongst these studies of blending gauge and
184 satellite data over Australia. Firstly, kriging is used as the preferred spatial
185 interpolation method. Kriging is a very popular spatial model, due to its ease to
186 be implemented, but it does have its limitations [41]. Secondly, only one satel-
187 lite data set was ever used in each of the blending processes. This prevents the
188 ability to compare different satellite data sets and their relative performances
189 in the blending process for Australian precipitation.

190 3 Materials and Methods

191 3.1 Australian Precipitation Data

192 In this study, we will be using three data sets giving monthly precipitation
193 measurements/estimates. The ground based rain gauge measurement data set
194 provided by the Bureau of Meteorology, and two satellite precipitation estima-
195 tion data sets provided by the Japanese Aerospace Exploration Agency and
196 the National Oceanic and Atmospheric Administration. Figure 1 shows the
197 square root precipitation measurement/estimate for June of 2021, for the three
198 data sets (square root is used to better see the small scale difference). While
199 there are more than two satellite precipitation data sets available, for several
200 important reasons, we have chosen to only use the two named above. Firstly,
201 the data needs to be free and easily accessible. Secondly, the data needs to
202 have full coverage over Australia. Thirdly, the data needs to display a degree
203 of accuracy in representing the true amount of precipitation which has been
204 measured by a rain gauge.

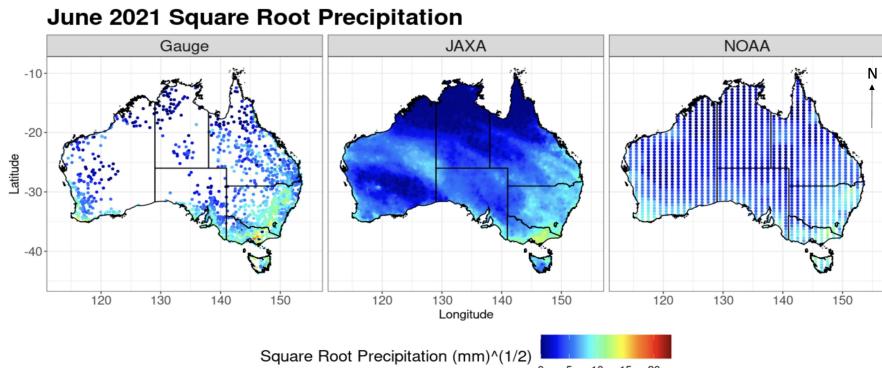
6 *Multi-platform remote sensing and in situ precipitation data*

Fig. 1 June 2021 square root precipitation Bureau of Meteorology rain gauge measurement (left), Japanese Aerospace Exploration Agency (middle) and National Oceanic and Atmospheric Administration (right).

205 3.1.1 Bureau of Meteorology Rain Gauge Data

206 The Bureau of Meteorology (BOM) is an Australian government agency which
 207 has the duty of providing weather research, analysis and services for Australia
 208 and the Australian regions. The data we are using from the Bureau of Meteorology
 209 comes from government rain gauge stations. A rain gauge station measures
 210 precipitation for an exact location at predetermined intervals. The time period
 211 for observations from the Bureau of Meteorology rain gauge stations are not
 212 consistent across Australia. The earliest rain gauge observation comes from
 213 January of 1841. The constantly developing nature of Australian society results
 214 in new rain gauges being introduced over the years, while redundant stations
 215 are removed. Nearly 6000 rain gauge stations have been operational in Aus-
 216 tralia at some point since January of 1841. Figure 2 shows where these rain
 217 gauges have been located, and the number of observed months which they
 218 were/have been in operation. Early established population centers have more
 219 densely population rain gauge networks, with many of the stations having a
 220 relative higher number of observed months.

221 In this study, we will only consider 3098 stations which have been opera-
 222 tional since April 2000 (earliest satellite estimate for the Japanese Aerospace
 223 Exploration Agency), and with at least three years of monthly data. Some of
 224 these gauge stations are only operational for the part of the total data his-
 225 tory. The rain gauge stations (1707 in total) displayed in Figure 1 are fully in
 226 operation as of March 2022.

227 3.1.2 Japanese Aerospace Exploration Agency Satellite Data

228 The Japanese Aerospace Exploration Agency (JAXA) is a Japanese govern-
 229 ment agency that focuses on research, technology development and launching
 230 of satellites into orbit. JAXA's Global Satellite Mapping of Precipitation
 231 (GSMap) estimates can provide measurements up to an hourly frequency.

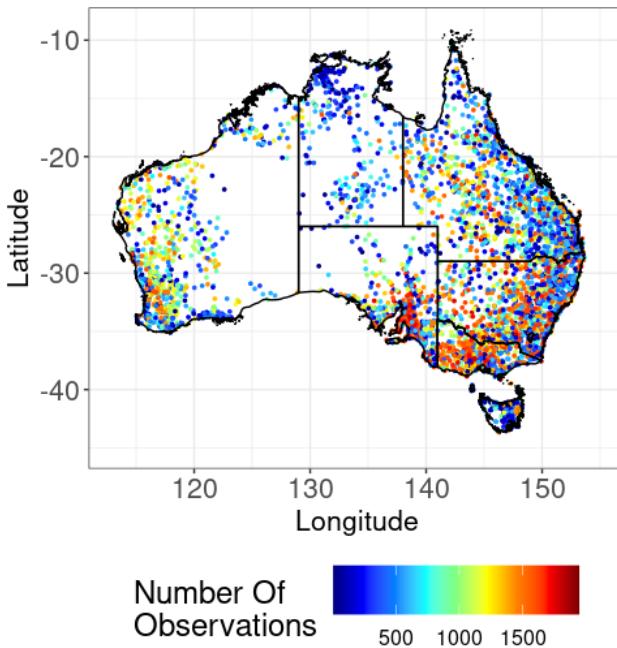


Fig. 2 Locations of all Bureau of Meteorology Rain Gauge stations which have been in operation at some point since January 1841. Colour represents how many months the stations was operational for.

GSMaP's estimates combine precipitation retrievals from the Tropical Rainfall Measuring Mission satellite (TRMMSS), as well as other orbiting satellites, and spatially interpolate them with cloud motion vectors derived from infrared images produced from geostationary satellites [42]. This data source allows for exact location estimation for a given longitude and latitude, where estimates are available from April 2000 to the present. We can obtain gridded measurements for the JAXA precipitation data at a spatial resolution of $0.1^\circ \times 0.1^\circ$. Figure 1 shows the locations where JAXA satellite estimates are available.

Figure 3 shows the correlation between the rain gauge measurements and the JAXA satellite estimates at each of the 3098 rain gauge locations. That is, the gauge precipitation measurements are compared to the satellite estimates for all observed time periods at each location. There is a strong degree of spatial dependency, with locations in northern Australia having higher correlation, which decreases the more south the location is. This is consistent with a previous study in Japan which showed JAXA to be more accurate at higher rainfall regions [43]. Figure 4 shows the percentage of observations, at each location in which the JAXA satellite overestimates the rain gauge measurements. This is very spatially dependent, with locations in central and northern Australia typically underestimating the amount of precipitation which has occurred. Some locations in southern regions do overestimate majority of the

8 Multi-platform remote sensing and in situ precipitation data

252 observations. However, the satellite is more likely to underestimate the amount
 253 of precipitation that a rain gauge has measured overall.

254 3.1.3 National Oceanic and Atmospheric Administration 255 Satellite Data

256 The National Oceanic and Atmospheric Administration (NOAA) is an Amer-
 257 ican government agency that is responsible for weather forecasts, and oceanic
 258 and atmospheric condition monitoring. NOAA's Climate Prediction Center
 259 Morphing technique (CMORPH) data is in the form of monthly satellite esti-
 260 mates on a $1^\circ \times 1^\circ$ grid over Australia. CMORPH's estimates are produced
 261 from passive microwave satellite scans, which are propagated by motion vec-
 262 tors from geostationary satellite infrared data [44]. The locations of the rain
 263 gauges do not coincide with the locations of available NOAA observations. To
 264 overcome this, we use a simple kriging model to interpolate the observations
 265 from the NOAA grid, to the gauge locations for each time period [45]. This data
 266 source has a longer history than JAXA and most of BOM with observations
 267 starting in January of 1979 for all of Australia. Figure 1 shows the locations
 268 which NOAA satellite estimates are available. Figure 3 shows the correlation
 269 between the rain gauge measurements and the NOAA satellite estimates at
 270 each of the gauge locations. Similarly to the JAXA-Gauge correlation, there
 271 is a strong degree of spatial dependency. However, unlike the JAXA data set,
 272 the high correlation between the variables is not limited to just northern Aus-
 273 tralia. Most of the country has high correlation. South Australia and parts
 274 of western Australia present the only locations where the correlation coeffi-
 275 cient is low. This is also the case with the JAXA data set. Figure 4 shows
 276 the percentage of observations in which the NOAA satellite overestimates the
 277 rain gauge measurement for each location. Again like JAXA, this is very spa-
 278 tially dependent. However, with NOAA, the vast majority of locations typically
 279 underestimate the amount of precipitation which has occurred. It is only in
 280 the east and south eastern regions where the satellite is not underestimating
 281 the gauge measurement the vast majority of the time.

282 Table 1 gives a summary of the different characteristics for the three
 283 different data sets.

Source	BOM	JAXA	NOAA
Method	Gauge Measurement	Satellite Estimate	Satellite Estimate
Start Date	Location Dependent	Apr 2000	Jan 1979
Temporal Resolution	Daily/Monthly	Hourly	Monthly
Spatial Resolution	Non-gridded	$0.1^\circ \times 0.1^\circ$	$1^\circ \times 1^\circ$

284 **Table 1** Characteristics of the three precipitation data sources.

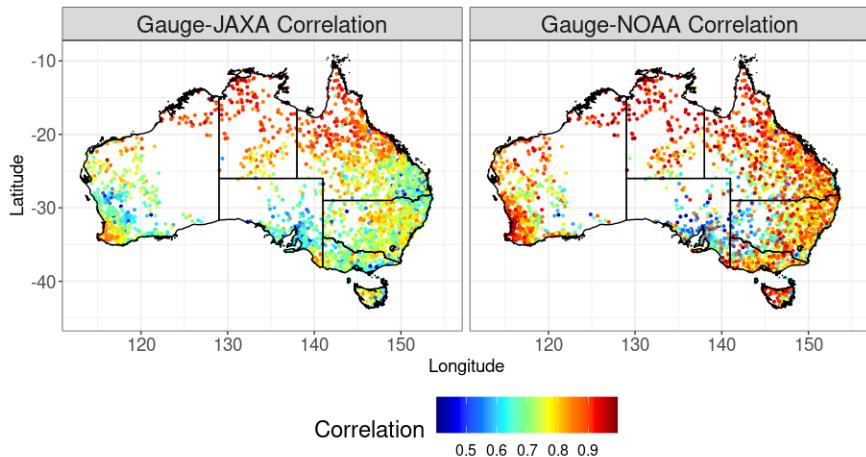


Fig. 3 Sample Pearson correlation between the gauge measurements, and the JAXA (left) and NOAA (right) satellite estimate of the available 3098 Bureau of Meteorology rain gauge stations.

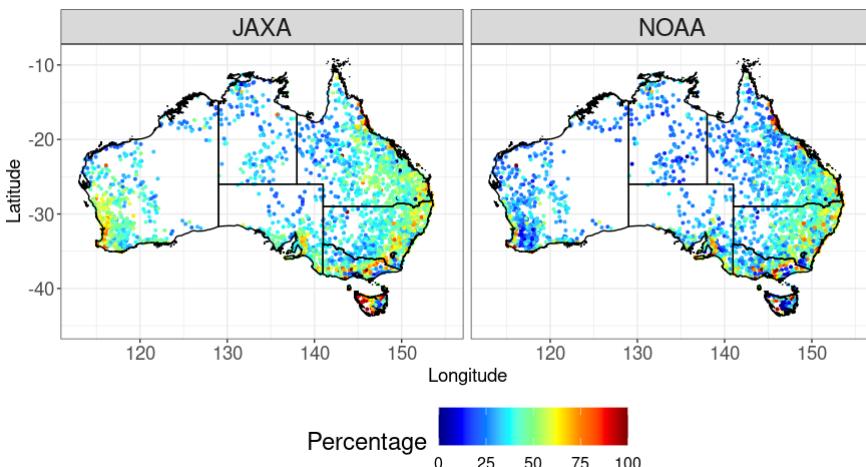


Fig. 4 Mappings showing percentage of observations in which the JAXA (left) and NOAA (right) satellites overestimate the gauge measurement.

3.2 Methods

A summary of the two-step process for blending rain gauge measurements with satellite precipitation estimates is shown in Figure 5. In the first step, the primary input data (gauge and satellite data) are modelled through a multiple linear regression relationship. In the second step, we spatially model the regression coefficient vectors from the first step, where the secondary input data, if tested significant, are used as explanatory variables to enhance the results.

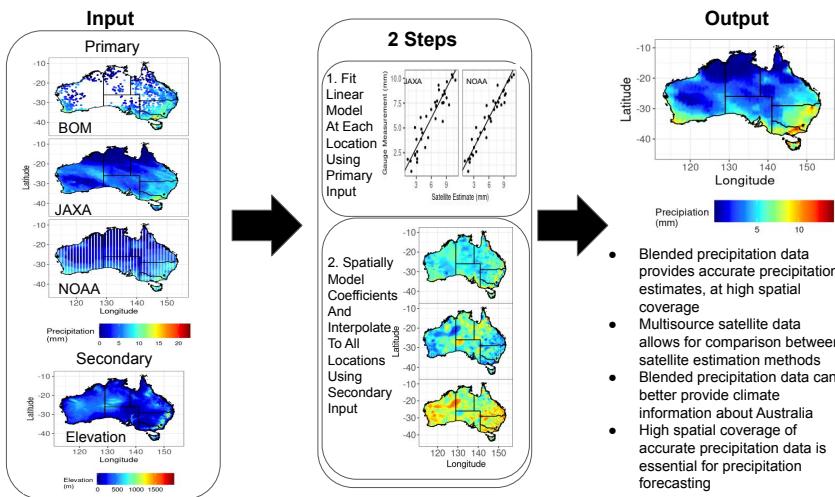


Fig. 5 Flowchart demonstrating the 2-step process for blending BOM rain gauge data, with JAXA Satellite and NOAA Satellite Data. The first step involves linearly modelling (i.e., multiple linear regression) the primary input data (gauge and satellite precipitation data). The second step involves spatially modelling the relationship (i.e., the multiple regression coefficient vectors) from the first in-between all locations, using the secondary input data to enhance the results.

3.2.1 Step 1: Linear Modelling (multiple linear regression)

We first model the relationship between the rain gauge measurements and the satellite estimates for each rain gauge location (primary input). This is achieved through assuming a linear relationship between the variables (gauge measurements and satellite estimates), then applying a multiple linear regression model at each location for synergistically combining the strengths of the different platforms. Previous studies have considered a linear relationship between gauge and satellite variables [46]. Precipitation is a non-negative variable, however linear models map to the entire real line. To overcome the possibility of a domain violation, we will be considering a square root transformation of all variables. This is a very common transformation to make for non-negative variables in linear regression [47], and is often used for precipitation data [48].

Let $s_i \in \mathbb{R}^2$ for $i = 1, \dots, m$ be the two-dimensional coordinates description of the i^{th} rain gauge location where s_{i1} and s_{i2} are the longitude and latitude for location i respectively. We then define $Y_{ij}^{[G]}$ to be the square root gauge measurement for the i^{th} location respectively, for the j^{th} time period, where $j = 1, \dots, n_i$ (as not all rain gauge locations have the same observed periods). Similarly, define $Y_{ij}^{[J]}$ and $Y_{ij}^{[N]}$ to be the square root JAXA and NOAA satellite estimates respectively for the i^{th} location respectively for the j^{th} time period where $j = 1, \dots, n_i$. Consider $Y_{ij}^{[G]}, Y_{ij}^{[J]}$ and $Y_{ij}^{[N]}$ to have a linear relationship

313 subject to random error $\varepsilon_{i,j}$, i.e.

314

$$Y_{i,j}^{[G]} = \beta_0^{[i,j]} + \beta_1^{[i,j]} Y_{i,j}^{[J]} + \beta_2^{[i,j]} Y_{i,j}^{[N]} + \varepsilon_{i,j}.$$

315 If we assume the relationship between the gauge measurements and satellite
 316 estimates to be temporally invariant (coefficients are the same regardless of
 317 time) which is true when the data are temporally stationary [49], we can
 318 express this relationship for location i across all observed time periods as the
 319 linear equation

320

$$\mathbf{Y}_i^{[G]} = \beta_0^{[i]} \mathbf{1}_{n_i} + \beta_1^{[i]} \mathbf{Y}_i^{[J]} + \beta_2^{[i]} \mathbf{Y}_i^{[N]} + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, m. \quad (1)$$

321 Here $\mathbf{Y}_i^{[G]} = (Y_{i,1}^{[G]}, \dots, Y_{i,n_i}^{[G]})'$, $\mathbf{Y}_i^{[J]} = (Y_{i,1}^{[J]}, \dots, Y_{i,n_i}^{[J]})'$, $\mathbf{Y}_i^{[N]} =$
 322 $(Y_{i,1}^{[N]}, \dots, Y_{i,n_i}^{[N]})'$, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, \dots, \varepsilon_{i,n_i})' \sim \mathcal{N}(\mathbf{0}, \sigma_i^2 I_{n_i})$ is the noise term
 323 with $\mathbf{1}_{n_i}$ being an $n_i \times 1$ vector with every entry being 1. The conventional
 324 method for estimating the coefficients from equation (1) is by minimising the
 325 ordinary least squares function so that the least squares estimates

326

$$\hat{\boldsymbol{\beta}}^{[i]} = \arg \min_{\beta_0^{[i]}, \beta_1^{[i]}, \beta_2^{[i]}} \|\mathbf{Y}_i^{[G]} - \beta_0^{[i]} \mathbf{1}_{n_i} - \beta_1^{[i]} \mathbf{Y}_i^{[J]} - \beta_2^{[i]} \mathbf{Y}_i^{[N]}\|_2^2 \quad (2)$$

327 where $\hat{\boldsymbol{\beta}}^{[i]} = (\hat{\beta}_0^{[i]}, \hat{\beta}_1^{[i]}, \hat{\beta}_2^{[i]})'$ and $\|\cdot\|_2$ is the Euclidean distance. The solution
 328 to equation (2) can be shown to be

329

$$\hat{\boldsymbol{\beta}}^{[i]} = (V_i' V_i)^{-1} V_i' \mathbf{Y}_i^{[G]}$$

330 with $V_i = (\mathbf{1}_{n_i}, \mathbf{Y}_i^{[J]}, \mathbf{Y}_i^{[N]})$ an $n_i \times 3$ matrix [50].

331 3.2.2 Step 2: Spatial Modelling of Regression Coefficients

332 From step 1, we will obtain spatially dependent coefficient vector estimates
 333 $\hat{\boldsymbol{\beta}}^{[i]}$ for the linear model representing the relationship between the square root
 334 of the BOM rain gauge measurements, and the square root of the JAXA and
 335 NOAA satellite estimates. Let

336

$$\boldsymbol{\beta}_\ell = [\hat{\beta}_\ell^{[1]} \ \hat{\beta}_\ell^{[2]} \ \dots \ \hat{\beta}_\ell^{[m]}]', \quad \ell = 0, 1, 2, \quad (3)$$

337 be an $m \times 1$ vector, so that $(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ may be regarded as a spatial process
 338 of the linear relationship between the BOM rain gauge measurements
 339 and the JAXA and NOAA satellite estimates over all locations [51]. Figure 3
 340 demonstrates strong spatial dependency of the correlations between the gauge
 341 measurements and satellite estimates, implying spatial dependency in the
 342 regression coefficient process $(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$. Such spatial dependency can be

12 *Multi-platform remote sensing and in situ precipitation data*

343 accounted for by the following spatial linear model (also known as a spatial
 344 autoregressive model or SAR):

345

$$\boldsymbol{\beta}_\ell = X\boldsymbol{\alpha}_\ell + \lambda_\ell W_{m,\ell}\boldsymbol{\beta}_\ell + \boldsymbol{\varepsilon}(\boldsymbol{\beta}_\ell) \quad \text{for } \ell = 0, 1, 2, \quad (4)$$

346 where λ_ℓ is the autocorrelation scalar parameter, $\boldsymbol{\alpha}_\ell$ is a coefficient vector
 347 associated with the data matrix X of covariates (secondary input), $W_{m,\ell}$ is a
 348 given $m \times m$ weight matrix giving the influences of the spatial distances of the
 349 observed locations, and random error $\boldsymbol{\varepsilon}(\boldsymbol{\beta}_\ell) \sim \mathcal{N}(\mathbf{0}, \sigma^2(\boldsymbol{\beta}_\ell)I_m)$ for $\ell = 0, 1, 2$
 350 [52]. Furthermore, $W_{m,\ell} = \{w_{ij}^{[\ell]}\}$, where $w_{ij}^{[\ell]}$ is given by some function of
 351 a known distance metric $d(\mathbf{s}_i, \mathbf{s}_j)$ where locations \mathbf{s}_i and \mathbf{s}_j are the locations
 352 associated with $\hat{\beta}_\ell^{[i]}$ and $\hat{\beta}_\ell^{[j]}$ respectively. Note $w_{ii}^{[\ell]} = 0$ for $i = 1, \dots, m$ as the
 353 observation cannot depend on itself. The choice of which metric and function
 354 we use to define the weight matrix $W_{m,\ell}$ is vital to the modelling process. A
 355 weight matrix not properly representing the true nature of the spatial domain
 356 of interest is likely to result in biased or inconsistent estimates of the model's
 357 parameters [53, 54]. Selection of the weight matrix is essential for providing
 358 unbiased estimates of parameters, and should be done in a way such that every
 359 entry is consistent to some rule. There are multiple methods for defining a
 360 weight matrix that could be used with different distance metrics. In this paper
 361 a mixture of k -nearest neighbours and inverse distance weighting (IDW) [55]
 362 with an addition of the location's distance to the nearest coast is used. That is,

363

$$w_{i,j} = \begin{cases} \frac{1/(d(\mathbf{s}_i, \mathbf{s}_j)^\gamma + dc(\mathbf{s}_i) + dc(\mathbf{s}_j))}{\sum_{\mathbf{s}_t \in \text{ne}(\mathbf{s}_i)} 1/(d(\mathbf{s}_i, \mathbf{s}_t)^\gamma + dc(\mathbf{s}_i) + dc(\mathbf{s}_t))} & , \text{ if } \mathbf{s}_j \in \text{ne}(\mathbf{s}_i) \\ 0 & , \text{ otherwise} \end{cases} \quad (5)$$

364 where $\text{ne}(\mathbf{s}_i)$ is the neighbourhood of the location \mathbf{s}_i which is determined by
 365 which other observed locations are the k closest to it, $dc(\mathbf{s}_i)$ is the distance from
 366 the location \mathbf{s}_i to the nearest coastal point, and γ is a non-negative value that
 367 represents the smoothness of the weight matrix. The distance to the coast is
 368 used as a factor in creating the weight matrices. As we are working with rainfall
 369 data, the nature of the spatial relationships may change when a location is
 370 further away from the coast [56], where there is significantly less rain and also
 371 for Australia, significantly fewer gauge stations. As the land of Australia lays
 372 on the surface of an approximate sphere, we should use a distance metric that
 373 considers the spherical nature of the domain. Thus we use a cosine distance
 374 metric, given by

375

$$d(\mathbf{s}_i, \mathbf{s}_j) = r\Delta\sigma \quad (6)$$

376 for $\Delta\sigma = \arccos(\sin s_{i,1} \sin s_{j,1} + \cos s_{i,1} \cos s_{j,1} \cos |s_{i,2} - s_{j,2}|)$ with $r \approx$
 377 6371 km (radius of Earth). Whilst using the cosine distance metric doesn't
 378 significantly change the results compared to Euclidean distance, both in terms
 379 of the tuning parameters (k and γ) and also the end result, it is better to use
 380 a more accurate representation of the distance between locations. The tuning
 381 parameters are selected via 10-fold cross validation [57]. This is done by first

randomly dividing the data (estimated linear model coefficients β_ℓ) into ten mutually exclusive groups, then fitting the spatial linear model on nine of the groups with weight matrices defined for different values of k and γ for all variable. Secondly, the regression coefficients are interpolated to the locations of the tenth group and compared with the actual values. This is then repeated for each of the ten groups, and the k and γ values which give the smallest average error are selected.

It can easily be shown that

$$\beta_\ell \sim \mathcal{N}((I_m - \lambda_\ell W_{m,\ell})^{-1} X \alpha_\ell, (I_m - \lambda_\ell W_{m,\ell})^{-1} (I_m - \lambda_\ell W'_{m,\ell})^{-1} \sigma^2(\beta_\ell)) \quad (7)$$

for $\ell = 0, 1$. As β_ℓ has normally distributed residuals, estimation of λ_ℓ and α_ℓ would be traditionally done by a least squares approach. The ordinary least squares estimates of λ_ℓ and α_ℓ for given β_ℓ satisfy

$$\left\{ \hat{\lambda}_\ell, \hat{\alpha}_\ell \right\} = \arg \min_{\alpha_\ell, \lambda_\ell} \|\beta_\ell - X \alpha_\ell - \lambda_\ell W_{m,\ell} \beta_\ell\|_2^2 \quad (8)$$

Directly solving (8) has been shown through testing to result in biased and inconsistent estimates, especially for λ_ℓ when $\|\alpha_\ell\|_2$ is large [58–61]. Another issue in solving (8) is it does not take into account the spatial dependence in β_ℓ . To overcome these obstacles, we propose a new loss function based on the squared fitting errors from the multiple regression models (1), and to fit the spatial autoregression models (4) simultaneously by minimising this loss function for the given gauge and satellite precipitation data. This new loss function is given as

$$L = \sum_{i=1}^m \xi_i \left\| \mathbf{Y}_i^{[G]} - \mathbf{1}_{n_i} (\mathbf{x}'_i \alpha_0 + \lambda_0 \mathbf{w}'_{0i} \beta_0) - \mathbf{Y}_i^{[J]} (\mathbf{x}'_i \alpha_1 + \lambda_1 \mathbf{w}'_{1i} \beta_1) - \mathbf{Y}_i^{[N]} (\mathbf{x}'_i \alpha_2 + \lambda_2 \mathbf{w}'_{2i} \beta_1) \right\|_2^2 \quad (9)$$

where $\mathbf{x}'_i \alpha_\ell + \lambda_0 \mathbf{w}'_{\ell i} \beta_\ell$ is the ℓ^{th} coefficient estimate from equation (4) for the i^{th} location, and ξ_i is the i^{th} location's weighting. The weighting ξ_i can be chosen in a number of ways. It may be dependent on the number of observations at the location, or potentially its spatial isolation. An iterative algorithm is used to find the estimates of the parameters $(\alpha_0, \alpha_1, \alpha_2)$ and $(\lambda_0, \lambda_1, \lambda_2)$ from minimising (9) (details of calculations are given in Appendix A). Then the blended gauge-satellite precipitation estimates for those locations void rain gauges can be calculated by spatial interpolations from (4) and (1). While there have been other studies which use multivariate spatial and spatio-temporal models to blend gauge and satellite precipitation data [46], the loss function in equation (9) has the added benefit of using the gauge and satellite precipitation data to help estimate the spatial linear model parameters.

For simplicity, let $\mathbf{Z}_{0i} = \mathbf{1}_{n_i}$, $\mathbf{Z}_{1i} = \mathbf{Y}_i^{[J]}$, $\mathbf{Z}_{2i} = \mathbf{Y}_i^{[N]}$, and $\delta_{ji} = \mathbf{x}'_i \boldsymbol{\alpha}_j + \lambda_j \mathbf{w}'_{ji} \boldsymbol{\beta}_j$ for $j = 0, 1, 2$. The algorithm to estimate the parameters by minimising the loss function (9) is given in Algorithm 1 below.

Algorithm 1 Generalised Combined Loss Parameter Estimation (GCLPE) Algorithm

```

1: procedure GCLPE( $\boldsymbol{\beta}_\ell, W_{m,\ell}, X$ )
2:   Initialise  $\lambda_\ell^{[0]} = 0$  and  $\boldsymbol{\alpha}_\ell^{[0]} = \mathbf{0}'$  for  $\ell = 0, 1, 2$ . For a pre-specified  $K$ ,
3:   while  $|\lambda_\ell^{[k]} - \lambda_\ell^{[k-1]}| > \epsilon$  for  $k = 1, 2, \dots, K$  do
4:     for  $\ell = 0, 1, 2$ 
5:        $\lambda_\ell^{[k]} = [\sum_{i=1}^m \xi_i (\mathbf{w}'_{\ell i} \boldsymbol{\beta}_\ell) (\mathbf{Z}'_{\ell i} \mathbf{Z}_{\ell i}) (\boldsymbol{\beta}'_\ell \mathbf{w}_{\ell i})]^{-1} \left( \sum_{i=1}^m \xi_i \left\{ (\mathbf{Z}'_{\ell i} \mathbf{Z}_{\ell i}) (\mathbf{Z}'_{\ell i} \mathbf{Y}_i^{[G]}) \right. \right.$ 
6:            $\left. \left. - (\mathbf{w}'_{\ell i} \boldsymbol{\beta}_\ell) (\mathbf{Z}'_{\ell i} \mathbf{Z}_{\ell i}) (\mathbf{x}'_i \boldsymbol{\alpha}_\ell^{[k-1]}) - (\mathbf{w}'_{\ell i} \boldsymbol{\beta}_\ell) \sum_{j \neq \ell} \delta_{ji} \mathbf{Z}'_{\ell i} \mathbf{Z}_{ji} \right\} \right)$ 
7:        $\boldsymbol{\alpha}_\ell^{[k]} = [\sum_{i=1}^m \xi_i \mathbf{x}_i (\mathbf{Z}'_{\ell i} \mathbf{Z}_{\ell i}) \mathbf{x}'_i]^{-1} \left( \sum_{i=1}^m \xi_i \left\{ \mathbf{x}_i (\mathbf{Z}'_{\ell i} \mathbf{Y}_i^{[G]}) - \mathbf{x}_i \sum_{j \neq \ell} \delta_{ji} \mathbf{Z}'_{\ell i} \mathbf{Z}_{ji} \right. \right.$ 
8:            $\left. \left. - \mathbf{x}_i (\mathbf{Z}'_{\ell i} \mathbf{Z}_{\ell i}) (\lambda_\ell^{[k]} \mathbf{w}'_{\ell i} \boldsymbol{\beta}_\ell) \right\} \right)$ 
9:        $k = k + 1$ 
10:      end while
11:      return the last  $\lambda_\ell^{[k]}$  and  $\boldsymbol{\alpha}_\ell^{[k]}$  as the estimates of  $\lambda_\ell$  and  $\boldsymbol{\alpha}$ , respectively, with  $\ell = 0, 1, 2$ .
12: end procedure

```

With estimates for λ_ℓ and $\boldsymbol{\alpha}_\ell$ for $\ell = 0, 1, 2$, we proceed to predict the value of coefficient vector $(\hat{\beta}_0^{[h]}, \hat{\beta}_1^{[h]}, \hat{\beta}_2^{[h]})$ for any location h having no in situ rainfall data by spatial interpolation. Namely, let \mathbf{s}_h be the coordinates of location h having in situ rainfall data, and let $\mathbf{w}_{h,\ell} = (w_{h,1}^{[\ell]}, \dots, w_{h,m}^{[\ell]})^T$ be the spatial weight vector with each entry being the function of the distance metric from \mathbf{s}_h to the m locations specified by (5) and (1). Then the prediction of $(\hat{\beta}_0^{[h]}, \hat{\beta}_1^{[h]}, \hat{\beta}_2^{[h]})$ at location h can be computed from the fitted model (4) and is of the following form

$$\tilde{\beta}_\ell^{[h]} = \mathbf{x}'_h \hat{\boldsymbol{\alpha}}_\ell + \hat{\lambda}_\ell \mathbf{w}_{h,\ell} \boldsymbol{\beta}_\ell, \quad \ell = 0, 1, 2. \quad (10)$$

Prediction of what the gauge measurement of location h for the j^{th} time interval would be can be accordingly computed from the corresponding satellite estimates by

$$\tilde{Y}_{hj}^{[G]} = \hat{\beta}_0^{[h]} + \hat{\beta}_1^{[h]} Y_{hj}^{[J]} + \hat{\beta}_2^{[h]} Y_{hj}^{[N]}. \quad (11)$$

Using equation (11), we are able to provide blended nowcasts and hindcasts of precipitation for any locations in a region. We would name the whole procedure of getting to (11) for blended nowcasts and hindcasts as Precipitation Profiler-Observation Fusion and Estimation (PPrOFusE), which will be further discussed in Sections 4.3 and 4.4.

437 **3.2.3 Clustering Analysis**

438 Through the steps above, we can obtain blended precipitation data, by spatial
 439 interpolation and temporal prediction, for months from April 2000 to March
 440 2022 ($n = 264$) and at a grid of locations uniformly distributed over Australia.
 441 For both simplicity and computational feasibility, data fusion is performed
 442 by substituting equation (10) and the JAXA and NOAA data into equation
 443 (11) to obtain the blended precipitation estimates at a uniform grid of 1391
 444 locations where in situ rain data are not necessarily available or complete.

445 Let $\mathbf{y}(i) = (y_{i1}, \dots, y_{in})'$ be the blended precipitation vector at location
 446 i , $i = 1, \dots, m$. Note $m = 1391$ for the blended data obtained above. The k -
 447 means method for partitioning the m data vectors $\mathbf{y}(1), \dots, \mathbf{y}(m)$ into k clusters
 448 is summarized as following.

1. Arbitrarily partition $\{1, \dots, m\}$ into k subsets (named step-0 partition).
 Note in this paper each location i is specified by its latitude and longitude
 value $(\text{lat}_i, \text{lon}_i)$. The step-0 partition can be achieved by a uniform partition
 of $\{(\text{lat}_i, \text{lon}_i) : i = 1, \dots, m\}$. Denote the subset j in the step-0 partition as

$$\mathcal{C}_j^{(0)} = \left\{ C_j^{(0)}(1), \dots, C_j^{(0)}(m_j^{(0)}) \right\}, \quad j = 1, \dots, k;$$

where $m_j^{(0)}$ is the size of subset j . Denote the cluster j of the blended
 precipitation vectors as

$$\mathbf{y}(\mathcal{C}_j^{(0)}) = \left\{ \mathbf{y}(C_j^{(0)}(1)), \dots, \mathbf{y}(C_j^{(0)}(m_j^{(0)})) \right\}, \quad j = 1, \dots, k.$$

Then compute the center of cluster $\mathbf{y}(\mathcal{C}_j^{(0)})$ by its sample mean:

$$\overline{\mathbf{y}(\mathcal{C}_j^{(0)})} = \left(m_j^{(0)} \right)^{-1} \sum_{l=1}^{m_j^{(0)}} \mathbf{y}(C_j^{(0)}(l)), \quad j = 1, \dots, k.$$

2. Compute the Euclidean distance $\|\cdot\|$ between each $\mathbf{y}(i)$ and each cluster

$$d \left(\mathbf{y}(i), \mathbf{y}(\mathcal{C}_j^{(0)}) \right) = \min_{1 \leq l \leq m_j^{(0)}} \| \mathbf{y}(i) - \overline{\mathbf{y}(\mathcal{C}_j^{(0)})} \|, \quad i = 1, \dots, m; \quad j = 1, \dots, k.$$

Then assign each $\mathbf{y}(i)$ to cluster $j^{(1)}(\mathbf{y}(i))$ such that

$$j^{(1)}(\mathbf{y}(i)) = \operatorname{argmin}_{1 \leq j \leq k} d \left(\mathbf{y}(i), \mathbf{y}(\mathcal{C}_j^{(0)}) \right), \quad i = 1, \dots, m.$$

This results in a new partition (named step-1 partition) of $\{\mathbf{y}(1), \dots, \mathbf{y}(m)\}$ into k clusters: $\mathbf{y}(\mathcal{C}_1^{(1)}), \dots, \mathbf{y}(\mathcal{C}_k^{(1)})$, with

$$\mathcal{C}_j^{(1)} \equiv \left\{ C_j^{(1)}(1), \dots, C_j^{(1)}(m_j^{(1)}) \right\} = \left\{ i I(j^{(1)}(\mathbf{y}(i)) = j) : i = 1, \dots, m \right\}$$

where $I(\cdot)$ is the indicator function equaling either 1 or empty. In addition, compute the center of $\mathbf{y}(\mathcal{C}_j^{(1)})$ by its sample mean:

$$\overline{\mathbf{y}(\mathcal{C}_j^{(1)})} = \left(m_j^{(1)} \right)^{-1} \sum_{l=1}^{m_j^{(1)}} \mathbf{y}(C_j^{(1)}(l)), \quad j = 1, \dots, k.$$

3. Using the step-1 partition as the starting point, repeat the process for obtaining the step-1 partition to get step-2 partition. Denote the resultant step-2 clusters as $\mathbf{y}(\mathcal{C}_1^{(2)}), \dots, \mathbf{y}(\mathcal{C}_k^{(2)})$, with

$$\mathcal{C}_j^{(2)} \equiv \left\{ C_j^{(2)}(1), \dots, C_j^{(2)}(m_j^{(2)}) \right\} = \left\{ i I(j^{(2)}(\mathbf{y}(i)) = j) : i = 1, \dots, m \right\}$$

and the cluster center of $\mathbf{y}(\mathcal{C}_j^{(2)})$

$$\overline{\mathbf{y}(\mathcal{C}_j^{(2)})} = \left(m_j^{(2)} \right)^{-1} \sum_{l=1}^{m_j^{(2)}} \mathbf{y}(C_j^{(2)}(l)), \quad j = 1, \dots, k.$$

This process is further repeated for either a pre-specified number of times, or until the sum of squared within-cluster differences (SSWD) reaches the minimum (assume this is achieved after q steps) with the step- q SSWD being

$$\text{SSWD} \left(\mathbf{y}(\mathcal{C}_1^{(q)}), \dots, \mathbf{y}(\mathcal{C}_k^{(q)}) \right) = \sum_{j=1}^k \sum_{i=1}^m d^2 \left(\mathbf{y}(i), \overline{\mathbf{y}(\mathcal{C}_j^{(1)})} \right).$$

In practice, it is often computationally infeasible to reach the global minimum SSWD. To tackle this difficulty, a small number s (say $s \leq 5$) of step-0 partitions are performed to obtain s step- q partition results (say $q \leq 20$). Then the s step- q partition (i.e., clustering) results are compared with each other to determine the best one.

3.2.4 Relevant Statistical Tests

Standard statistical tests are used in this paper to validate various postulated hypotheses underpinning data fusion. These tests and their references are listed below:

- 458 1. Correlation coefficient test [62]. Used for
 459 2. Regression coefficient test [50]. Used to test coefficients of the linear
 460 regression model in equation (1).
 461 3. Augmented Dickey-Fuller test [63]. Used for testing the temporal station-
 462 arity assumption.

4 Results and Discussion

4.1 Assumption Justifications

463 There are two direct assumptions that have been made in the methodology
 464 section. Firstly, the relationship between the square root rain gauge measure-
 465 ment and the square root satellite estimates is linear. Secondly, the relationship
 466 between the rain gauge measurement and a satellite estimate is temporally
 467 invariant. Both of these assumptions have been used to blend gauge and
 468 satellite data previously [46].

469 The assumption for the referenced linear relationship is natural. If satellite
 470 estimates are accurate, a linear relationship should be observed. Figure 3 vali-
 471 dates the linear relationship assumption, as the vast majority of locations have
 472 high correlation values, especially for the NOAA data set. Also four out of the
 473 3098 gauge locations did not have significant correlation between their mea-
 474 surements and the JAXA satellite estimates (at significance level $\alpha = 0.05$),
 475 whereas only two out of the 3098 locations did not have significant correla-
 476 tion between the gauge and NOAA data. In Figure 6, the total amount of
 477

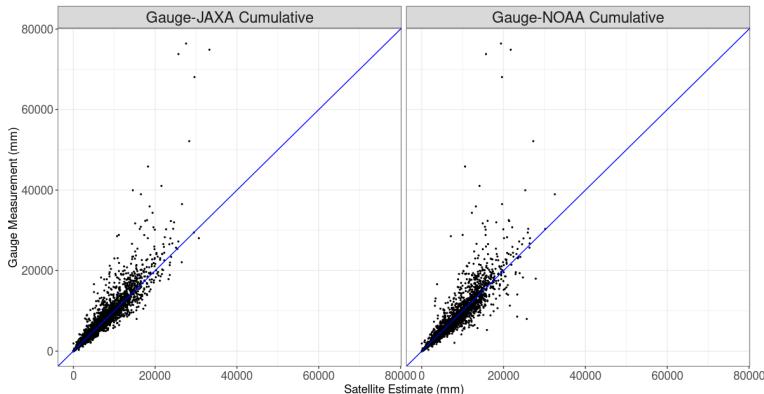


Fig. 6 Comparisons of the cumulative gauge measurements against the cumulative JAXA (left) and NOAA (right) satellite estimates of precipitation for each location. The blue line represents $y = x$.

478 precipitation recorded at each rain gauge location for the entire time it was
 479 operational is compare to the total satellite estimates for that location over
 480 the same period (see supplementary figures for seasonal comparison). Whilst
 481

the trend follows the blue $y = x$ line quite closely, giving more evidence to the assumption of a linear relationship, the variance in the difference between the total gauge measurements and the total satellite estimates increases as the measurement increases. If the satellite consistently overestimates or underestimates the gauge measurement in locations where there is a lot of precipitation, this problem will only be exacerbated when summing up the entire series. Four examples of linearity between the variables are shown in Figure 7, where we have scatter plots showing the data for four spatially diverse locations (locations shown in Figure 11). There is a clear trend of linearity between the gauge measurements and satellite estimates at each of these locations subject to minor noise. While there are some points where the gauge and satellite values are significantly different, most of the points follow a linear trend. We also looked at including further terms in the linear model, such as quadratic and cubic transformations of the satellite estimate. No additional terms were found significant at level $\alpha = 0.05$. With only the regression coefficients associated with the JAXA and NOAA data and the intercept were statistically significant at any location, we conclude the aforementioned relationship is linear.

To justify the temporal stationarity assumption we can consider the difference between the gauge and satellite observations at each location ($Y_{i,j}^{[G]} - Y_{i,j}^{[J]}$ and $Y_{i,j}^{[G]} - Y_{i,j}^{[N]}$) as a time series. We perform the augmented Dickey-Fuller test to check for temporal stationarity. We are testing the null hypothesis that there is a unit root present in the time series against the alternative that the time series is stationary [63]. The result of this test gives no locations with a p -value above 0.05, meaning the time series at each location is stationary. Thus the mean, variance and autocorrelation of the time series do not change with time [49], and we can justify using one multiple regression model for each location with one set of coefficients which are temporally invariant.

4.2 Linear Modelling Results

Figure 8 shows the results for the estimated coefficients for the linear model of the square root gauge measurements against the square root JAXA and NOAA satellite estimates. There is a clear spatial trend for the intercept coefficients. In southern Australia and south west Australia, the estimates are very close to zero. Interestingly, along the great dividing range in eastern Australia (a large mountain range, elevation shown in Figure 11), the intercept coefficients are also very close to zero. Higher values of the NOAA coefficient are associated with lower values of the JAXA coefficient. These estimates are shown on the same colour scale for easier comparison. JAXA has uniformly lower estimates than NOAA. JAXA's relative highest estimates occur in the northern regions of Australia, where its correlation coefficient in Figure 3 was at its highest. The NOAA estimates however have less spatial variability than JAXA, with the vast majority of estimates being between 0.75 and 1.3. There are locations where the coefficients estimated values are moderately different to those around it. There are numerous possible reasons for this. Firstly, there is no

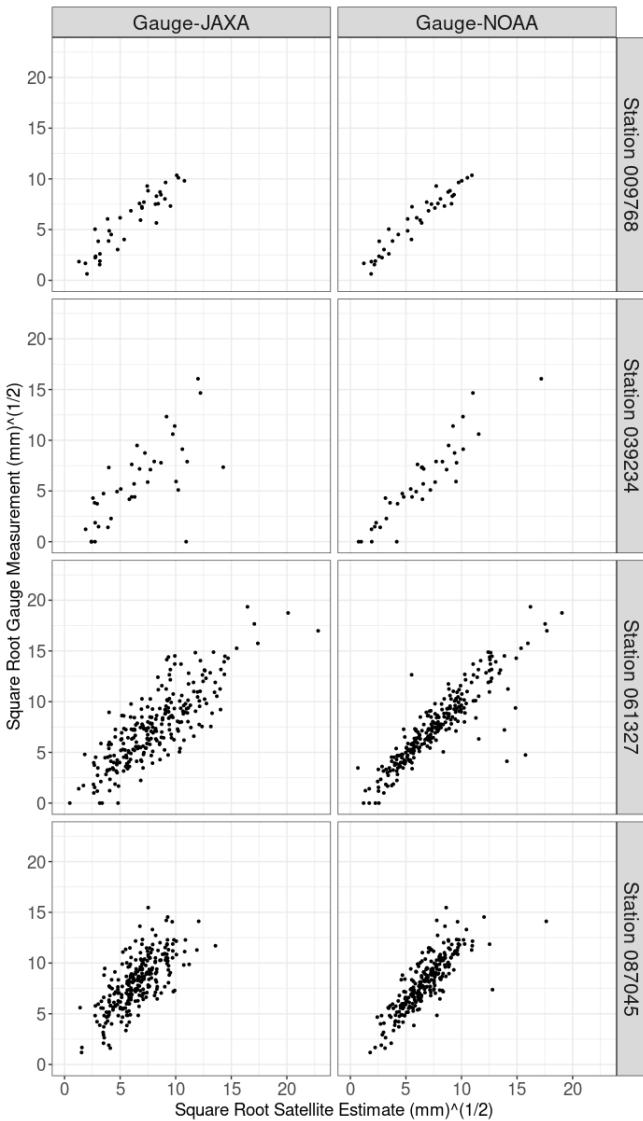


Fig. 7 Monthly square root gauge precipitation measurements against their corresponding monthly square root satellite estimates for JAXA (left) and NOAA (right). Shown for station 009768 (top, Mount Barker WA), 015657 (second from the top, Canteen Creek NT), 039234 (second from the bottom, Pine Creek QLD) and 048018 (bottom, Collarenebri NSW) demonstrating linear relationship between the variables.

guarantee rain gauges give perfectly accurate measurements for the amount of precipitation which has occurred in a given time interval. There are several factors which can lead to rain gauge inaccuracy, such as blockages, wind direction and evaporation [26]. These could all potentially affect certain rain gauges to varying degrees. Secondly, it is possible for the satellites' behaviour

530 to vary moderately over short distances due to factors such as complex topog-
 531 raphy [64] and proximity to the coastline [56]. Thus resulting in very different
 coefficient estimates.

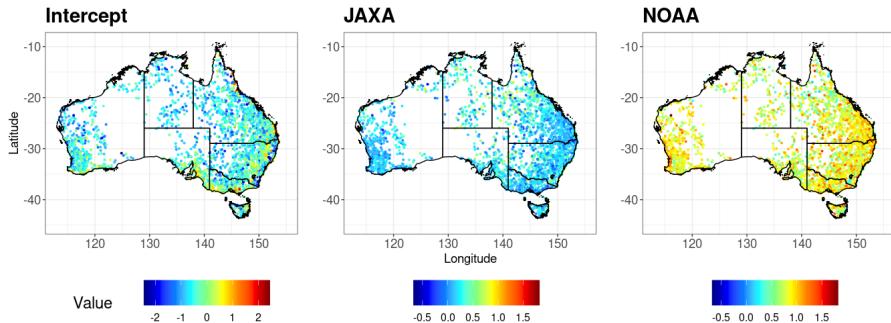


Fig. 8 Mappings of the linear regression coefficient estimates (intercept:left, JAXA:middle, NOAA:right) for the linear relationship between the gauge measurement and the JAXA and NOAA satellite estimates for every location. The JAXA and NOAA estimates have the same colour for easier comparison.

532
 533 It must be noted the estimates for the coefficients are all relative to the
 534 scale of the satellite estimate at a given location. A coefficient's estimate will
 535 depend on the relative domain of the variable. For example, if a satellite source
 536 consistently overestimates the true gauge measurement amount at some loca-
 537 tion, its coefficient will likely be small. This does not necessarily indicate the
 538 satellite is not influential in the regression process at the given location. It may
 539 indicate the need to calibrate the satellite significantly, to properly model the
 540 gauge measurements. Figure 4 showed the percentage of observations at each
 541 location which the satellite would overestimate the gauge measurement. Loca-
 542 tions where the satellite often underestimated the true gauge measurement,
 543 would more likely have a larger coefficient estimate. JAXA and NOAA satel-
 544 lite have been shown to overestimate when the true precipitation is low, but
 545 underestimate when the true precipitation is high [15]. As Australia is a very
 546 dry continent, there will be more opportunities for the satellites to overesti-
 547 mate the low-level true precipitation, compared to underestimating high-level
 548 true precipitation. We can test the significance of each of the coefficient esti-
 549 mates at each location through a simple *t*-test. Figure 9 shows the locations
 550 where the coefficients are statistically significant at the $\alpha = 0.05$ level. The
 551 red colour indicates the locations with statistically insignificant coefficient esti-
 552 mates. In other words, can be considered to be zero, and have no influence in
 553 the blending process. Approximately 63% of locations have a non significant
 554 intercept coefficient. The intercept parameter is in an essence, a measure of
 555 the bias of the satellite estimate. The notion of the satellites being accurate
 556 would lead to the expectation of most locations having very little, to no bias.
 557 Figure 7 showed at some locations, when the gauge measurement is zero, the
 558 satellites may estimate significantly higher. The locations which this occurs

are more likely to have a significant intercept parameter. Approximately 33% of locations have a non significant JAXA coefficient. Compared to the NOAA coefficient, where only about 1% of locations have a non significant estimate. A satellite coefficient is non significant when in the presence of the other variables (other satellite and intercept) provides no increased accuracy for modelling the gauge measurements. The NOAA data being significant in nearly all location indicates it is almost uniformly, the better satellite precipitation estimate platform. There are no locations where both the JAXA and NOAA variables are not significant. Thus there is no gauge location which does not have a significant relationship with at least one of the satellite data sets.

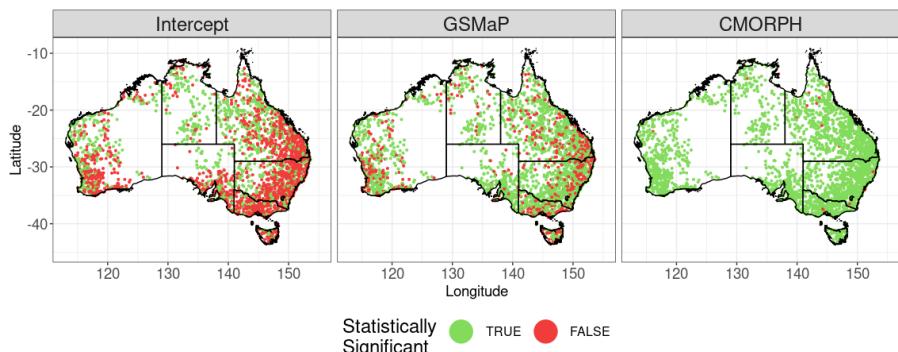


Fig. 9 Mappings showing the locations of the statistically significant (green) and statistically non significant (red) regression coefficient estimates (intercept:left, JAXA:middle, NOAA:right) for the linear relationship between the gauge measurement and the JAXA and NOAA satellite estimates for every location.

The residual standard error for the model at each location can be easily quantified. Figure 10 shows the residual standard error for the linear model at each gauge location. Most locations have a low residual standard error. The relatively higher residual standard error locations are on the east and north eastern coasts of Australia. These locations are characterised by having very high seasonal precipitation [65]. It has been shown that both JAXA and NOAA exhibit less accuracy for higher precipitation amounts [15]. At lower precipitation climate regions, the linear model gives a lot less error. This is partly due to the relatively reduced magnitude of total precipitation, making errors less significant. Shown also in Figure 10 is a density plot for the residual standard error. The vast majority of locations have a residual standard error less than 2, with an average residual standard error of about 1.6. There are some locations with a residual standard error of up to 4. However, these are in locations on the coastline where precipitation is typically higher, and the both JAXA and NOAA are less accurate, thus as a relative percentage, the error is not so large.

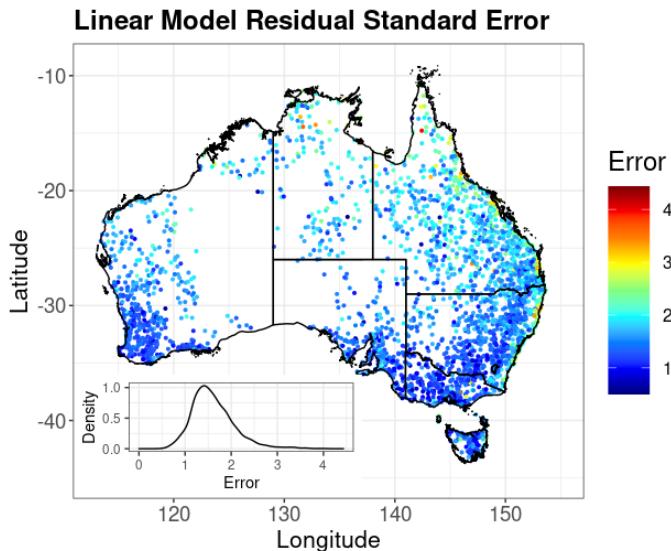


Fig. 10 Main: Mapping of the residual standard error for each gauge station when modelled using equation (1). Minor: Density plot for the residual standard error.

585 4.3 Spatial Modelling Results

586 To use the spatial linear model in equation (4), a data matrix X is required,
 587 which will have corresponding coefficients to be estimated in α . This data
 588 matrix will contain an intercept parameter, but can also contain spatially
 589 specific characteristics for each location. The intercept coefficient results in
 590 Figure 8 indicated there is a relationship between the estimated linear regres-
 591 sion coefficients $\hat{\beta}_\ell$ and the elevation of the location. Thus the elevation of a
 592 location may be a useful parameter to use in the spatial model. The satellite's
 593 images, in which the precipitation is estimated from, are taken from above.
 594 Rain gauge stations which are at a higher elevation will be closer to both the
 595 cloud and the satellite. Whilst the difference in elevation is marginal compared
 596 to the altitude of the satellite, it is not marginal compared to the altitude of
 597 the clouds. The elevation range of the rain gauge stations is 1,868 metres, with
 598 multiple at sea level and the highest located in Kosciuszko National Park, New
 599 South Wales. Figure 11 gives an elevation mapping of Australia with locations
 600 of interest labelled. Precipitation causing clouds can occur anywhere between
 601 ground level and 6,000 metres [66]. Therefore the altitude of the rain gauge
 602 station may help us better model the coefficients' behaviour. Previous gauge
 603 satellite precipitation blending studies have used elevation as an explanatory
 604 variable to improve the blending process [40, 67]. We will use both a lin-
 605 ear and quadratic elevation parameter in this model (the significance of these
 606 parameters can be tested).

607 Table 2 shows the parameter estimates of the model in (4) when using
 608 algorithm 1. The high values of the spatial dependency parameters (λ_ℓ),

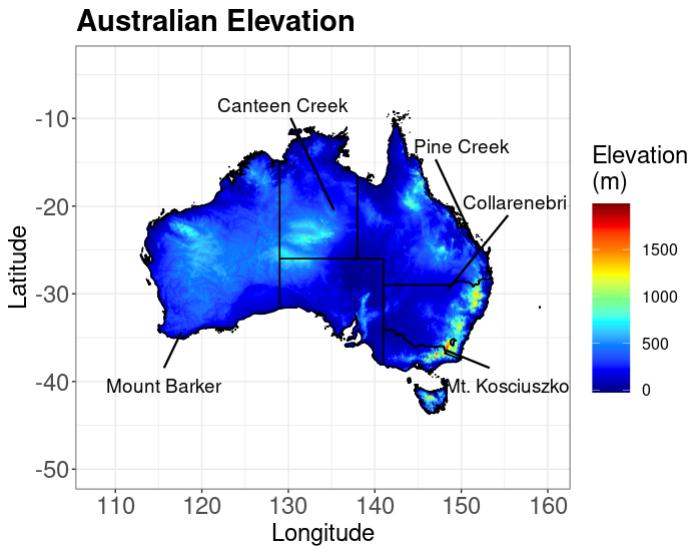


Fig. 11 Elevation map of Australia with locations of interest shown.

demonstrate the highly spatially dependent nature of the relationship between the gauge measurements and the satellite estimates. We also see the estimate for the spatial model intercept ($\alpha_{\ell 0}$), the elevation parameter ($\alpha_{\ell 1}$), the quadratic elevation parameter ($\alpha_{\ell 2}$), and the parameters used to give the optimal weight matrices (k_ℓ and γ_ℓ). It is clear each spatial model parameter behaves differently for linear model coefficient $\hat{\beta}_\ell$. The low value for the estimated $\hat{\alpha}_{00}$ indicates with the inclusion of the JAXA and NOAA variables, the bias becomes almost negligible (as seen in Figure 9). However, for the NOAA coefficient it is relatively large, highlighting the importance of this variable. There is more similarity in the linear and quadratic elevation parameters. While the linear elevation parameter is small for all three variables, it is statistically significant (at $\alpha = 0.05$) for the intercept and NOAA variables. The quadratic elevation parameter is not statistically significant for any variable. Previous studies have also indicated there is a significant decrease in the performance of satellites over mountainous areas, with satellites underestimating rainfall over higher elevations [68, 69].

Figure 12 shows the results when algorithm 1 is used to estimate the parameters from equation (4) and spatially interpolated to all of Australia in equation (10). Similarly to Figure 8, the JAXA and NOAA values maintain their negatively correlated relationship. In northern parts of Australia, the JAXA estimate is relatively high, whereas it's at its relative lowest in central and southern Australia. On the eastern and south western coasts the NOAA estimate is also relatively high whereas the JAXA estimate is at its lowest, and in some cases it is even estimated to be negative. Whilst adding another explanatory variable to the model may allow for better results, it may make it more difficult to interpret the relationship between the variables.

Variable	Intercept	JAXA	NOAA
$\hat{\lambda}_\ell$	0.9024	0.8104	0.7901
$\hat{\alpha}_{\ell 0}$	-8.84×10^{-3}	5.054×10^{-2}	1.517×10^{-1}
$\hat{\alpha}_{\ell 1}$	-3.015×10^{-4}	-6.813×10^{-5}	1.223×10^{-4}
$\hat{\alpha}_{\ell 2}$	4.557×10^{-7}	6.547×10^{-8}	-1.074×10^{-7}
k_ℓ	7	7	6
γ_ℓ	1.31	0.64	0.58

Table 2 Spatial linear model parameters and weight matrix tuning parameter estimates for the square root relationship for monthly gauge measurements against JAXA satellite and NOAA satellite estimates using algorithm 1.

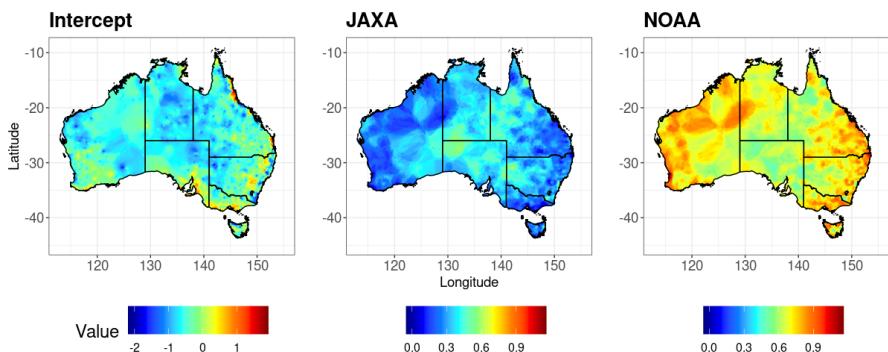


Fig. 12 Mappings of the linear regression intercept (top), JAXA estimate (middle) and NOAA estimate (bottom) from equation (1) using the model in equation (4) then interpolated to all locations using equation (10). Estimation of spatial model parameters done using algorithm 1.

We now turn to the final results of this blending process, PPrOFusE. Figure 13, highlights two recent significant events. The first being the 2021 eastern Australian floods which began in March of 2021 (first column). The second being the 2022 eastern Australian floods which began in February of 2022 (second column). Other months are shown in the supplementary materials. Figure 13 shows the PPrOFusE blended values (first row), the BOM gauge measurements (second row), the JAXA satellite estimates (third row, and the NOAA satellite estimates (fourth row) for the respective months. It should be noted the BOM gauge measurements of these two months were not used to calibrate the parameters of the model as to have a proper comparison. In March of 2021 heavy rainfall occurred between the 16th and 23rd on the New South Wales coast [70]. Some affected regions experienced between 400mm and 600mm during the time stated above, where the mean annual rainfall is between 1000mm and 15000mm [71]. In February of 2022 heavy rainfall occurred from the 23rd of February to the 7th of March in southern Queensland and northern New South Wales. Some locations in the flood affected areas recorded nearly 700 mm of rain in a three day period [72]. We can see the JAXA and NOAA maps exhibit very similar behaviour for the respective months, which is also very

similar to the gauge measurements. The square root of the precipitation is used here such that the difference in values are more easily observed. Based on the results in Figure 12, the PPrOFusE values in eastern Australia, where the floods were located, are more influenced by NOAA estimates as the coefficient is significantly higher than the JAXA coefficient. However, in central and north eastern regions of Australia, where the influence of the two satellite sources are about the same, the PPrOFusE values are a more even mixture of the two mappings.

While unfortunately we cannot properly quantify the accuracy of the model in locations where there are no rain gauges, we will do this for the locations where all three precipitation variables are available. We quantify the accuracy at the rain gauge locations by the following steps:

1. Use equation (10) to fit the regression coefficients $\hat{\beta}_\ell^{[h]}$ at the gauge locations (the regression coefficients estimates $\hat{\beta}_\ell^{[h]}$ at the gauge locations from equation (1) are not used for to estimate itself)
2. Apply equation (11) to estimate the amount of precipitation $\tilde{Y}_{i,j}^{[G]}$ which has occurred over the observed time periods, using the JAXA and NOAA estimates
3. Compare the estimates in $\tilde{Y}_{i,j}^{[G]}$ to the actual observed observation $Y_{i,j}^{[G]}$.

Figure 14 shows the residual errors for PPrOFusE at each of the gauge locations available for March 2021 (left) and February 2022 (right). There is less spatial dependency than what was seen in Figure 10. This is due to some locations' gauge measurements having a moderately different relationship with the satellite estimates, than the locations around it. This results in the spatially estimated coefficients having more error which similarly affects the fitted precipitation values the same way. Majority of the locations have low residual standard error with respective means of 1.69 and 1.92. This indicates the modelling process we have used gives useful results and gives an alternative approach for blending rain gauge measurements with multiple satellite estimates.

4.4 Clustering Results

To show the ability of the PPrOFusE data to give insight into the precipitation system of Australia, we apply the k -means clustering algorithm as described in section 3.2.3 using all the time periods available (March 2000 to 2022 February). That is, we obtain the PPrOFusE values for each month at the 1391 locations from the NOAA dataset. Figure 15 shows the results of this clustering with $k = 6$ centres. We use $k = 6$ centres to have our results be comparable to the Australian major rainfall zones and the Köppen climate regions, as defined by the Bureau of Meteorology. BOM defines there to be six major rainfall zones; summer dominant (wet summer and dry winter), summer (wet summer and low winter rainfall), uniform (even rainfall year round), winter (wet winter and low summer rainfall), winter dominant (wet winter

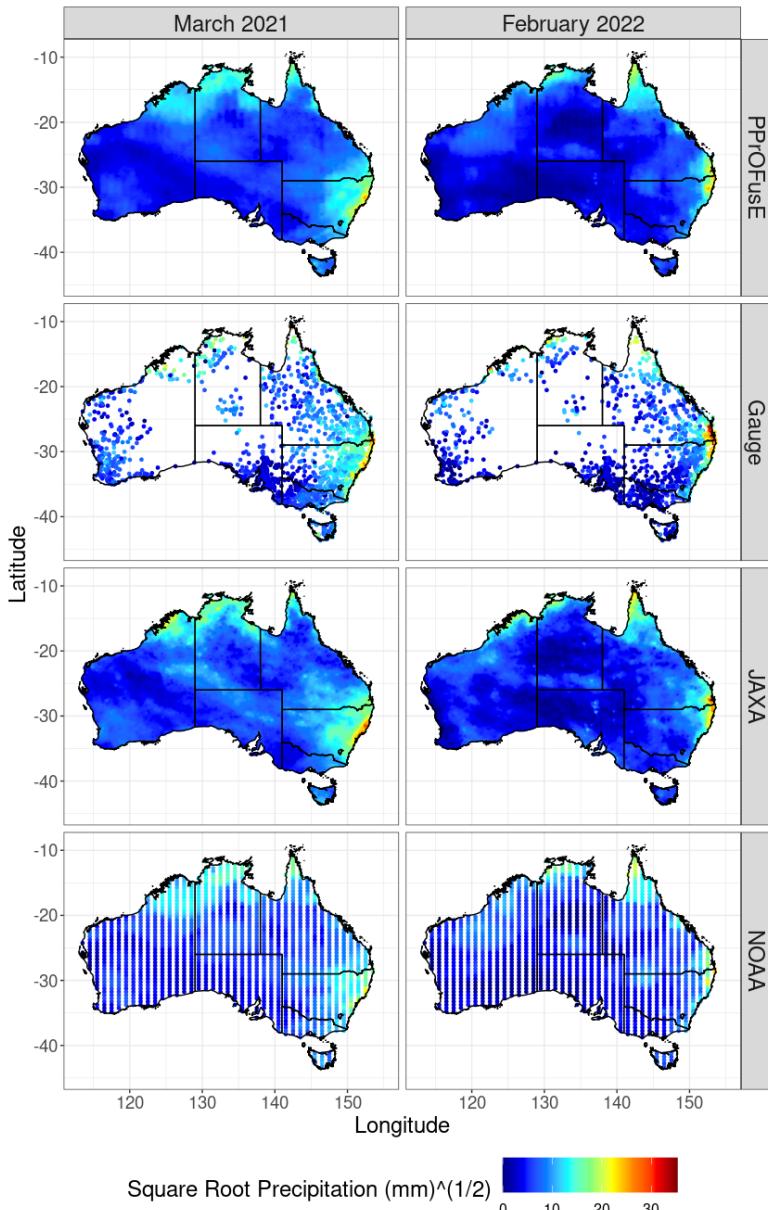


Fig. 13 PPrOFusE results for March 2021 (first column) and February 2022 (second column). Square root blended values (first row), square root BOM rain gauge measurements (second row), square root JAXA satellite estimates (third row), and the square root NOAA satellite estimates (fourth row) using the model in equation (11). Extreme rainfall across south-eastern Queensland and eastern New South Wales, resulting in the 2021 and 2022 Eastern Australian Floods.

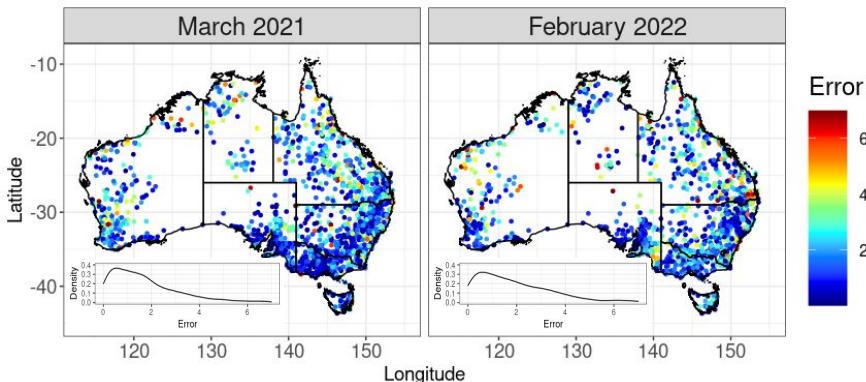


Fig. 14 Main: Mapping of the error for each gauge station when modelled using equations (10) and (11) for March 2021 (left) and February 2022 (right). Minor: Density plot for the residual standard error.

and dry summer), and arid (low rainfall) as shown in the bottom left plot in Figure 15. These major rainfall zones are defined by the differences between summer and winter rainfall across Australia using median annual rainfall and seasonal incidence (the ratio of median rainfall over the period November to April to the period May to October), using data from 1900 to 1999. Previous studies have used the major rainfall zones as factors in their precipitation models [73]. The PPrOFusE data gives slightly different results to the Bureau classification with the clusters now being summer dominant (high, k), summer dominant (medium, k), summer dominant (low, k), summer (k), winter dominant (k), Arid (k). The k is to distinguish between the clustering classifications, and the BOM classifications. For the same number of zones, the PPrOFusE data clustering gives a greater distinction for the precipitation in northern Australia. There are now four categories for summer precipitation and only one for winter, where we can see the summer dominant regions are expanding southwards more toward the centre of Australia. The distinction the k means clustering makes for the northern regions of Australia is more similar to what is seen in all the rainfall zones plot, where there are 16 categories (bottom centre plot of Figure 15). For the southern regions of Australia, the clustering of the blended data does not give as great of a distinction as what it does for northern Australia. A possible reason for this is that the spread of precipitation is far greater than for northern Australia and much more dependent on the season. The PPrOFusE data offers richer information about the spatial and temporal variation across Australia which can capture the diverse climate nature of northern Australia [74]. Another possible reason is the shift of precipitation systems due to climate change, where the major rainfall zones defined with data from 1900 to 1999 cannot capture this. Thus the k means clustering and the major rainfall zones are not entirely comparable however.

There is also a difference in the data capturing itself. Past measurement capabilities where both precision and sparsity of the gauge network are inferior to those achieved today. Similarly, the Köppen climate classification, which takes into account both precipitation and temperature [75], would be affected by the different precipitation data. Australia has six major Köppen climate regions (bottom right plot of Figure 15), where the zones take the climatic limits of native vegetation into consideration. Studies have used the Köppen climate classifications for both a reference [76] and as an additional input to enhance modelling outcomes [77, 78]. A locations classification can change over time, and this has been used to help quantify climate change [79, 80]. With new blended precipitation data, the Köppen climate classifications would change. In turn this change would affect any model that use the Köppen climate classifications. Of course, we can change the number of clusters k , if we want to further refine the classifications. Analysis of the quality of clustering through the silhouette method [81] and the elbow method [82] suggest having 5 and 6 clusters to be comparable. This clearly merits further work about what would be the optimal number of cluster/zones to define the Australian rainfall and climate systems.

To give more credibility to the k -means results in Figure 15, we look at histograms of the precipitation for locations in each of the clusters. Figure 16 compares the distributions of precipitation in each of the given clusters. There are clearly great distinctions between some of the different clusters in the distributions. The arid regions are concentrated on the lower end of the scale. Compare this to the summer dominant (high) cluster, where there is a large concentrated amount of precipitation near zero, but there is a lot of a reasonably uniform spread up to 600 mm. The summer dominant high, medium and low, all have similarities in their histograms. However, while for all three regions there is a high concentration at the lower end of the domain, the summer dominant (low) region has fewer observations above 200 mm, but has had observations reach as high as 550 mm. The summer dominant (high) and summer dominant (medium) have many observations above 200 mm, with summer dominant (high) going significantly higher. It should be noted that when the precipitation at each locations in each zone is aggregated into a single plot, we do lose a sense of seasonality which may cause there not to be as great distinction visually between the clusters. Histograms for the precipitation of the major rainfall zones and the Köppen climate zones are shown in the supplementary material. The histograms for the major rainfall zones and the Köppen climate zones show less distinction between zones compared to the k means clustering. This indicates the major rainfall zones and the Köppen climate zones may not truly be capturing the current dynamics and making distinctions of the Australian precipitation system.

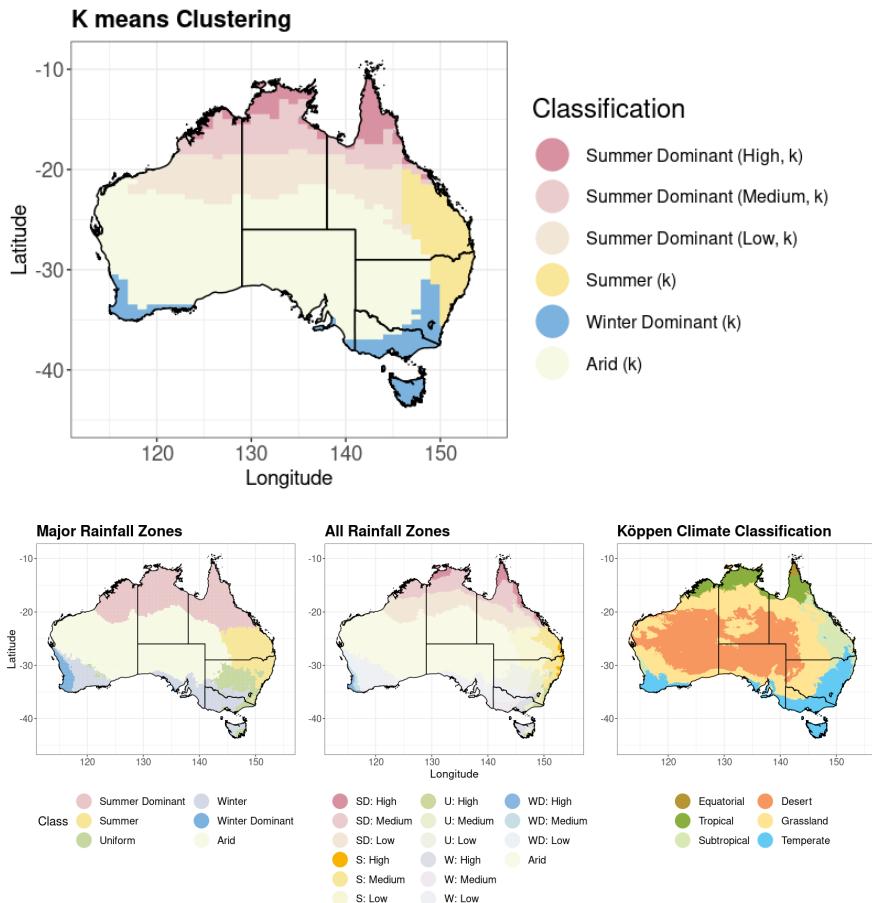


Fig. 15 Top: k means clustering and classification for blended precipitation data for $k = 6$ across all time periods. Bottom left: Bureau of Meteorology's six major rainfall zones. Bottom centre: Bureau of Meteorology's 16 rainfall zones ($SD =$ Summer Dominant, $S =$ Summer, $U =$ Uniform, $W =$ Winter, $WD =$ Winter Dominant). Bottom right: Bureau of Meteorology's six major Köppen climate regions.

5 Conclusion

In this paper, we developed PProFusE, a method for blending ground based rain gauge data with multi-satellite platform precipitation data. This was achieved by first finding the relationship between the gauge measurements and satellite estimates at each gauge location, then spatially interpolating the relationship to locations where there are no rain gauges. We developed a new loss function with an accompanying algorithm for a model which has not been used for gauge satellite precipitation data blending before. This new loss function enabled us to deliver more accurate spatial regression coefficient for multiple spatial autoregressive models, and thus more higher-quality blended gauge-satellite estimates in the form of PPrOFusE for all of Australia.

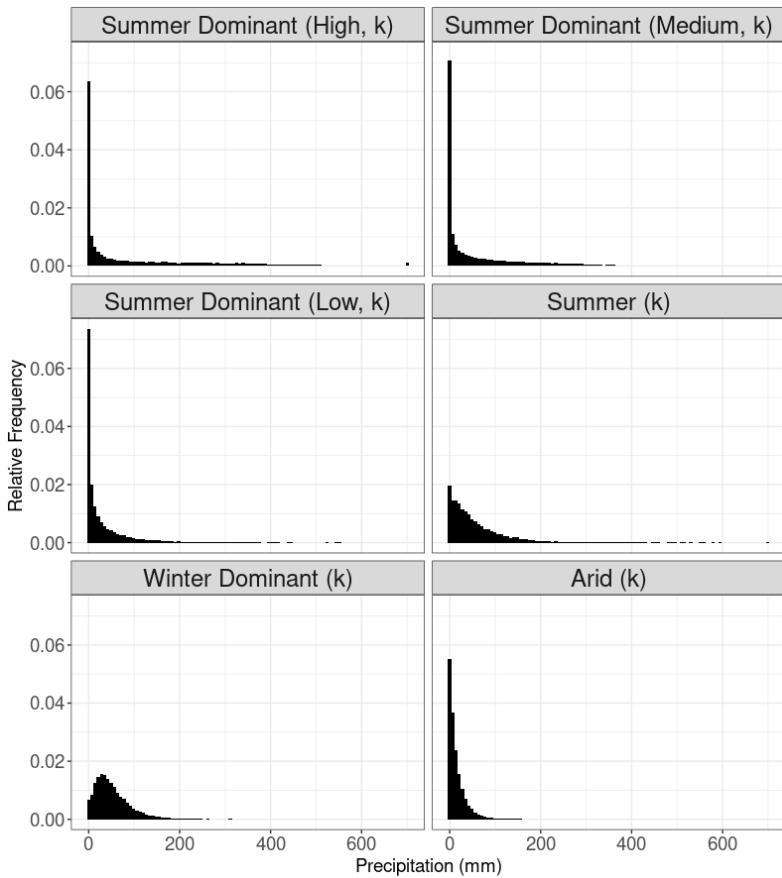


Fig. 16 Histograms for the amount of precipitation across each cluster defined through the k means clustering.

The results in this paper suggest there are underlying differences in the dynamics of the JAXA and NOAA satellite data sets. On balance, NOAA data can better replicate gauge measurements across Australia. The JAXA satellite estimates for one-third of Australian rain gauge locations were not of a significant influence in the presence of the NOAA estimates. Nevertheless, JAXA still has advantages to offer in the blending process and we have exploited this in locations where applicable (greater spatial resolution). However it needs to be noted that satellite precipitation estimation is a constantly improving method [83, 84]. Thus we would expect in the future, these satellite estimates will become more and more accurate, the correlation coefficients will increase, and the need for rain gauges will decrease.

A clustering analysis of the blended precipitation over the last two decades suggests two main trends on Australia's changing climate relative to the Bureau of Meteorology's six major rainfall zones from the previous century (1900-1999). First there is an increased spatial variability to the north, which

is consistent with meteorological expectations, while the edge of the wet summer dominant zone is creeping towards the middle of the continent, stretching across Western Australia, Northern Territory and west of New South Wales. Second the edge of the arid region is shifting southwards, pushing out the winter dominant rainfall zone across southern Australia.

There are limitations with the methods we have proposed which do merit further work. Firstly, no correction was made to account for the potentially decreasing bias in the satellite estimates. To achieve this, we would need to fully understand what changes have been made to the satellite measuring method and equipment. Secondly, we have only used elevation as an explanatory variable to explain the relationship between the gauge and satellite observations. There are many spatially specific characteristic which could be used such as, urban classification, tree coverage average wind etc. This data is not easily obtained in a large, mostly uninhabited continent like Australia. Thirdly, the selection of the tuning parameters k_ℓ and γ_ℓ is a computationally intensive task. Using cross validation for six parameters simultaneously can limit the ability of this methodology to be reproduced and applied to new data.

Much future work remains with gauge-satellite blending as more heterogeneous observation data becomes available from different platforms and as blending methods continue to improve and yield more accurate estimates of precipitation across the globe. In a changing climate, a more accurate and broader spatial coverage of precipitation is crucial, and until satellites can perfectly replicate direct gauge measurement, blending the data will be necessary.

Finally, while this methodology was developed in the context of blending gauge and satellite precipitation data, it is not limited to this setting. Algorithm 1 can be adapted to allow for many data sets, which can provide an alternative multivariate spatial model to the most common cokriging model [85].

818 Acknowledgement

819 The authors thank the U.S. Army International Technology Center Pacific
820 (ITC-PAC) and US DoD High Performance Computing Modernization Pro-
821 gram (HPCMP) under Contract No. FA5209-18-C-0002 for financial support.
822 The authors would also like to thank Reviewers for taking the time and effort
823 necessary to review the manuscript.

824 Data availability

825 The Bureau of meteorology rain gauge data is openly accessed
826 at <http://www.bom.gov.au/climate/data/>. The Japanese Aerospace
827 and Exploration Agency satellite data is openly accessed at
828 <https://sharaku.eorc.jaxa.jp/GSMaP/>. The National Oceanic and
829 Atmospheric Administration satellite data is openly accessed at
830 <https://www.ncei.noaa.gov/data/>.

831 References

- [1] Scott, M.: American City Planning Since 1890: A History Commemorating the Fiftieth Anniversary of the American Institute of Planners. 1, vol. 3. Univ of California Press, Berkeley, California (1969)
- [2] Weltzin, J.F., Loik, M.E., Schwinnning, S., Williams, D.G., Fay, P.A., Haddad, B.M., Harte, J., Huxman, T.E., Knapp, A.K., Lin, G., *et al.*: Assessing the response of terrestrial ecosystems to potential changes in precipitation. *Bioscience* **53**(10), 941–952 (2003)
- [3] Dore, M.H.: Climate change and changes in global precipitation patterns: what do we know? *Environment International* **31**(8), 1167–1181 (2005)
- [4] Trenberth, K.E.: Changes in precipitation with climate change. *Climate Research* **47**(1-2), 123–138 (2011)
- [5] Mukherjee, S., Aadhar, S., Stone, D., Mishra, V.: Increase in extreme precipitation events under anthropogenic warming in India. *Weather and Climate Extremes* **20**, 45–53 (2018)
- [6] Kunkel, K.E.: North American trends in extreme precipitation. *Natural Hazards* **29**(2), 291–305 (2003)
- [7] Jung, I.-W., Bae, D.-H., Kim, G.: Recent trends of mean and extreme precipitation in Korea. *International Journal of Climatology* **31**(3), 359–370 (2011)
- [8] Gummadi, S., Rao, K., Seid, J., Legesse, G., Kadiyala, M., Takele, R., Amede, T., Whitbread, A.: Spatio-temporal variability and trends of precipitation and extreme rainfall events in Ethiopia in 1980–2010. *Theoretical and Applied Climatology* **134**(3), 1315–1328 (2018)
- [9] Keune, J., Miralles, D.: A precipitation recycling network to assess freshwater vulnerability: Challenging the watershed convention. *Water Resources Research* **55**(11), 9947–9961 (2019)
- [10] Alcamo, J., Henrichs, T., Rosch, T.: World water in 2025. *World Water Series Report* **2** (2000)
- [11] Michaelides, S., Levizzani, V., Anagnostou, E., Bauer, P., Kasparis, T., Lane, J.: Precipitation: Measurement, remote sensing, climatology and modeling. *Atmospheric Research* **94**(4), 512–533 (2009)
- [12] Michelson, D.B.: Systematic correction of precipitation gauge observations using analyzed meteorological variables. *Journal of Hydrology* **290**(3-4), 161–177 (2004)

- 866 [13] Peterson, T.C., Easterling, D.R., Karl, T.R., Groisman, P., Nicholls, N.,
867 Plummer, N., Torok, S., Auer, I., Boehm, R., Gullett, D., *et al.*: Homogenity
868 adjustments of in situ atmospheric climate data: a review. International Journal of
869 Climatology: A Journal of the Royal Meteorological Society **18**(13), 1493–1517 (1998)
- 871 [14] Groisman, P.Y., Legates, D.R.: The accuracy of United States precipitation
872 data. Bulletin of the American Meteorological Society **75**(2), 215–228
873 (1994)
- 874 [15] Chua, Z.-W., Kuleshov, Y., Watkins, A.: Evaluation of satellite precipita-
875 tion estimates over Australia. Remote Sensing **12**(4), 678 (2020)
- 876 [16] Wu, W., Hall, C.A., Scatena, F.N., Quackenbush, L.J.: Spatial mod-
877 ellings of evapotranspiration in the luquillo experimental forest of Puerto
878 Rico using remotely-sensed data. Journal of Hydrology **328**(3-4), 733–752
879 (2006)
- 880 [17] Wu, W., Zhang, L.: Comparison of spatial and non-spatial logistic regres-
881 sion models for modeling the occurrence of cloud cover in north-eastern
882 Puerto Rico. Applied Geography **37**, 52–62 (2013)
- 883 [18] Yan, J., Gebremichael, M.: Estimating actual rainfall from satellite
884 rainfall products. Atmospheric Research **92**(4), 481–488 (2009)
- 885 [19] Ramarohetra, J., Sultan, B., Baron, C., Gaiser, T., Gosset, M.: How satel-
886 lite rainfall estimate errors may impact rainfed cereal yield simulation in
887 West Africa. Agricultural and Forest Meteorology **180**, 118–131 (2013)
- 888 [20] Lashkari, A., Salehnia, N., Asadi, S., Paymand, P., Zare, H., Ban-
889 nayan, M.: Evaluation of different gridded rainfall datasets for rainfed
890 wheat yield prediction in an arid environment. International Journal of
891 Biometeorology **62**(8), 1543–1556 (2018)
- 892 [21] Fiddes, S., Pepler, A., Saunders, K., Hope, P.: Redefining southern Aus-
893 tralia's climatic regions and seasons. Journal of Southern Hemisphere
894 Earth Systems Science **71**(1), 92–109 (2021)
- 895 [22] Gemedo, D.O., Korecha, D., Garedew, W.: Monitoring climate extremes
896 using standardized evapotranspiration index and future projection of
897 rainfall and temperature in the wettest parts of southwest Ethiopia.
898 Environmental Challenges, 100517 (2022)
- 899 [23] Villarini, G., Mandapaka, P.V., Krajewski, W.F., Moore, R.J.: Rain-
900 fall and sampling uncertainties: A rain gauge perspective. Journal of
901 Geophysical Research: Atmospheres **113**(D11) (2008)

34 *Multi-platform remote sensing and in situ precipitation data*

- 902 [24] Cheng, K.-S., Lin, Y.-C., Liou, J.-J.: Rain-gauge network evaluation and
903 augmentation using geostatistics. *Hydrological Processes: An Interna-*
904 *tional Journal* **22**(14), 2554–2564 (2008)
- 905 [25] Adhikary, S.K., Yilmaz, A.G., Muttill, N.: Optimal design of rain gauge
906 network in the Middle Yarra River catchment, Australia. *Hydrological*
907 *Processes* **29**(11), 2582–2599 (2015)
- 908 [26] Devine, K.A., Mekis, E.: Field accuracy of Canadian rain measurements.
909 *Atmosphere-ocean* **46**(2), 213–227 (2008)
- 910 [27] Krajewski, W., Smith, J.A.: Radar hydrology: Rainfall estimation.
911 *Advances in Water Resources* **25**(8-12), 1387–1394 (2002)
- 912 [28] Dinku, T., Ceccato, P., Connor, S.J.: Challenges of satellite rainfall esti-
913 mation over mountainous and arid parts of east Africa. *International*
914 *Journal of Remote Sensing* **32**(21), 5965–5979 (2011)
- 915 [29] Zhou, S., Tordesillas, A., Intrieri, E., Di Traglia, F., Qian, G., Catani, F.:
916 Pinpointing early signs of impending slope failures from space. *Journal of*
917 *Geophysical Research: Solid Earth*, 2021–022957 (2021)
- 918 [30] Tordesillas, A., Kahagalage, S., Campbell, L., Bellett, P., Intrieri, E., Batterham,
919 R.: Spatiotemporal slope stability analytics for failure estimation
920 (SSSAFE): linking radar data to the fundamental dynamics of granular
921 failure. *Scientific Reports* **11**(1), 1–18 (2021)
- 922 [31] Tordesillas, A., Kahagalage, S., Campbell, L., Bellett, P., Batterham, R.:
923 Introducing a data-driven framework for spatiotemporal slope stability
924 analytics for failure estimation. *Technical report, Australian Centre for*
925 *Geomechanics* (2021)
- 926 [32] Tordesillas, A., Zhou, S., Traglia, F.D., Intrieri, E.: New insights into the
927 spatiotemporal precursory failure dynamics of the 2017 Xinmo landslide
928 and its surrounds. In: *Workshop on World Landslide Forum*, pp. 331–338
929 (2020). Springer
- 930 [33] Wang, H., Qian, G., Tordesillas, A.: Modeling big spatio-temporal
931 geo-hazards data for forecasting by error-correction cointegration and
932 dimension-reduction. *Spatial Statistics* **36**, 100432 (2020)
- 933 [34] Chappell, A., Renzullo, L.J., Raupach, T.H., Haylock, M.: Evaluating
934 geostatistical methods of blending satellite and gauge data to estimate
935 near real-time daily rainfall for Australia. *Journal of Hydrology* **493**, 105–
936 114 (2013)
- 937 [35] Pilz, R.D.: Lessons in Water Policy Innovation from the World's Driest

- 938 Inhabited Continent: Using Water Allocation Plans and Water Markets
939 to Manage Water Scarcity. *U. Denv. Water L. Rev.* **14**, 97 (2010)
- 940 [36] Bui, H.T., Ishidaira, H., Shaowei, N.: Evaluation of the use of global
941 satellite-gauge and satellite-only precipitation products in stream flow
942 simulations. *Applied Water Science* **9**(3), 1–15 (2019)
- 943 [37] Duan, Z., Ren, Y., Liu, X., Lei, H., Hua, X., Shu, X., Zhou, L.: A
944 comprehensive comparison of data fusion approaches to multi-source
945 precipitation observations: A case study in Sichuan Province, China.
946 *Environmental Monitoring and Assessment* **194**(6), 1–17 (2022)
- 947 [38] Li, M., Shao, Q.: An improved statistical approach to merge satellite
948 rainfall estimates and raingauge data. *Journal of Hydrology* **385**(1–4),
949 51–64 (2010)
- 950 [39] Chua, Z.-W., Kuleshov, Y., Watkins, A.B., Choy, S., Sun, C.: A two-
951 step approach to blending GSMAp satellite rainfall estimates with gauge
952 observations over Australia. *Remote Sensing* **14**(8), 1903 (2022)
- 953 [40] Chua, Z.-W., Kuleshov, Y., Watkins, A.B., Choy, S., Sun, C.: A compari-
954 son of various correction and blending techniques for creating an improved
955 satellite-gauge rainfall dataset over Australia. *Remote Sensing* **14**(2), 261
956 (2022)
- 957 [41] Armstrong, M.: Problems with universal kriging. *Journal of the Interna-
958 tional Association for Mathematical Geology* **16**(1), 101–108 (1984)
- 959 [42] Tian, Y., Peters-Lidard, C.D., Adler, R.F., Kubota, T., Ushio, T.: Evalua-
960 tion of GSMAp precipitation estimates over the contiguous United States.
961 *Journal of Hydrometeorology* **11**(2), 566–574 (2010)
- 962 [43] Kachi, M., Kubota, T., Ushio, T., Shige, S., Kida, S., Aonashi, K.,
963 Okamoto, K., Oki, R.: Development and utilization of “jaxa global rainfall
964 watch” system based on combined microwave and infrared radiometers
965 aboard satellites. *IEEJ Transactions on Fundamentals and Materials*
966 **131**(9), 729–737 (2011)
- 967 [44] Joyce, R.J., Janowiak, J.E., Arkin, P.A., Xie, P.: CMORPH: A method
968 that produces global precipitation estimates from passive microwave
969 and infrared data at high spatial and temporal resolution. *Journal of*
970 *Hydrometeorology* **5**(3), 487–503 (2004)
- 971 [45] Cressie, N.: Spatial prediction and ordinary kriging. *Mathematical Geol-*
972 *ogy* **20**(4), 405–421 (1988)
- 973 [46] Jin, B., Wu, Y., Miao, B., Wang, X.L., Guo, P.: Bayesian spatiotemporal

36 *Multi-platform remote sensing and in situ precipitation data*

- 974 modeling for blending in situ observations with satellite precipitation esti-
975 mates. *Journal of Geophysical Research: Atmospheres* **119**(4), 1806–1819
976 (2014)
- 977 [47] Scealy, J., Welsh, A., *et al.*: Properties of a square root trans-
978 formation regression model. In: *Proceedings of CoDaWork'11: 4th*
979 *International Workshop on Compositional Data Analysis* (2011). Universitat de Girona. Departament d'Informàtica Matemàtica Aplicada.
980 <http://hdl.handle.net/2117/366730>
- 981
- 982 [48] Fu, G., Viney, N.R., Charles, S.P.: Evaluation of various root transforma-
983 tions of daily precipitation amounts fitted with a normal distribution for
984 Australia. *Theoretical and Applied Climatology* **99**(1), 229–238 (2010)
- 985 [49] Hamilton, J.D.: *Time Series Analysis*. Princeton University Press, Prince-
986 ton, New Jersey (2020)
- 987 [50] Freedman, D.A.: *Statistical Models: Theory and Practice*. Cambridge
988 University Press, Cambridge, United Kingdom (2009)
- 989 [51] Schabenberger, O., Gotway, C.A.: *Statistical Methods for Spatial Data
990 Analysis*. Chapman and Hall/CRC, London, United Kingdom (2017)
- 991 [52] Ord, K.: Estimation methods for models of spatial interaction. *Journal of
992 the American Statistical Association* **70**(349), 120–126 (1975)
- 993 [53] Herrera, M., Mur, J., Ruiz Marin, M.: Selecting the Most Adequate Spat-
994 ial Weighting Matrix: A Study on Criteria. Technical report, University
995 Library of Munich, Germany (2012)
- 996 [54] Beenstock, M., Felsenstein, D., *et al.*: *The Econometric Analysis of Non-
997 Stationary Spatial Panel Data*. Springer, Gewerbestrasse 11, 6330 Cham,
998 Switzerland (2019)
- 999 [55] Wikle, C.K., Zammit-Mangion, A., Cressie, N.: *Spatio-temporal Statistics
1000 With R*. Chapman and Hall/CRC, Wollongong, New South Wales (2019)
- 1001 [56] Tian, Y., Peters-Lidard, C.D.: A global map of uncertainties in satellite-
1002 based precipitation measurements. *Geophysical Research Letters* **37**(24)
1003 (2010)
- 1004 [57] Refaeilzadeh, P., Tang, L., Liu, H.: Cross-validation. *Encyclopedia of
1005 Database Systems* **5**, 532–538 (2009)
- 1006 [58] Lee, L.-f., Yu, J.: Estimation of spatial autoregressive panel data models
1007 with fixed effects. *Journal of Econometrics* **154**(2), 165–185 (2010)

- 1008 [59] Lee, L.-F.: Asymptotic distributions of quasi-maximum likelihood esti-
1009 mators for spatial autoregressive models. *Econometrica* **72**(6), 1899–1925
1010 (2004)
- 1011 [60] Li, H., Calder, C.A., Cressie, N.: Beyond Moran's I: testing for spa-
1012 tial dependence based on the spatial autoregressive model. *Geographical*
1013 *Analysis* **39**(4), 357–375 (2007)
- 1014 [61] Kelejian, H.H., Prucha, I.R.: A generalized spatial two-stage least squares
1015 procedure for estimating a spatial autoregressive model with autoregres-
1016 sive disturbances. *The Journal of Real Estate Finance and Economics*
1017 **17**(1), 99–121 (1998)
- 1018 [62] Asuero, A.G., Sayago, A., González, A.: The correlation coefficient: An
1019 overview. *Critical Reviews in Analytical Chemistry* **36**(1), 41–59 (2006)
- 1020 [63] Fuller, W.A.: *Introduction to Statistical Time Series* vol. 428. John Wiley
1021 & Sons, Ames, Iowa (2009)
- 1022 [64] Milewski, A., Elkadiri, R., Durham, M.: Assessment and comparison
1023 of TMPA satellite precipitation products in varying climatic and topo-
1024 graphic regimes in Morocco. *Remote Sensing* **7**(5), 5697–5717 (2015)
- 1025 [65] Nicholls, N., Drosdowsky, W., Lavery, B.: Australian rainfall variability
1026 and change. *Weather* **52**(3), 66–72 (1997)
- 1027 [66] Houze Jr, R.A.: *Cloud Dynamics* vol. 104. Academic press, Cambridge,
1028 Massachusetts (2014)
- 1029 [67] Ma, Y., Sun, X., Chen, H., Hong, Y., Zhang, Y.: A two-stage blending
1030 approach for merging multiple satellite precipitation estimates and rain
1031 gauge observations: An experiment in the northeastern Tibetan Plateau.
1032 *Hydrology and Earth System Sciences* **25**(1), 359–374 (2021)
- 1033 [68] Dinku, T., Ceccato, P., Grover-Kopec, E., Lemma, M., Connor, S.,
1034 Ropelewski, C.: Validation of satellite rainfall products over East Africa's
1035 complex topography. *International Journal of Remote Sensing* **28**(7),
1036 1503–1526 (2007)
- 1037 [69] Habib, E., Haile, A.T., Tian, Y., Joyce, R.J.: Evaluation of the high-
1038 resolution cmorph satellite rainfall product using dense rain gauge obser-
1039 vations and radar-based estimates. *Journal of Hydrometeorology* **13**(6),
1040 1784–1798 (2012)
- 1041 [70] Yague-Martinez, N., Leach, N.R., Dasgupta, A., Tellman, E., Brown, J.S.:
1042 Towards Frequent Flood Mapping with the Capella Sar System. The 2021
1043 Eastern Australia Floods case. In: 2021 IEEE International Geoscience

38 *Multi-platform remote sensing and in situ precipitation data*

- 1044 and Remote Sensing Symposium IGARSS, pp. 6174–6177 (2021). IEEE
- 1045 [71] Australian Government, B.o.M.: Special Climate Statement 74 –
1046 extreme rainfall and flooding in eastern and central Australia in
1047 March 2021. <http://www.bom.gov.au/climate/current/statements/scs74.pdf>. Accessed: 2022-08-11
- 1049 [72] Australian Government, B.o.M.: Special Climate Statement 76 – Extreme
1050 rainfall and flooding in south-eastern Queensland and eastern New South
1051 Wales. <http://www.bom.gov.au/climate/current/statements/scs76.pdf?20220525>. Accessed: 2022-08-11
- 1053 [73] Feng, P., Wang, B., Li Liu, D., Ji, F., Niu, X., Ruan, H., Shi, L.,
1054 Yu, Q.: Machine learning-based integration of large-scale climate drivers
1055 can improve the forecast of seasonal rainfall probability in Australia.
1056 Environmental Research Letters **15**(8), 084051 (2020)
- 1057 [74] Lee, D., Neal, A.: The climate in northern australia. Northern Australia:
1058 The arenas of life and ecosystems on half a continent, 29–55 (1984)
- 1059 [75] Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg,
1060 A., Wood, E.F.: Present and future Köppen-geiger climate classification
1061 maps at 1-km resolution. Scientific Data **5**(1), 1–12 (2018)
- 1062 [76] d'Andrimont, R., Defourny, P.: Monitoring African water bodies from
1063 twice-daily MODIS observation. GIScience & Remote Sensing **55**(1), 130–
1064 153 (2018)
- 1065 [77] Mohammad, P., Goswami, A.: Quantifying diurnal and seasonal variation
1066 of surface urban heat island intensity and its associated determinants
1067 across different climatic zones over Indian cities. GIScience & Remote
1068 Sensing **58**(7), 955–981 (2021)
- 1069 [78] Lohmann, U., Sausen, R., Bengtsson, L., Cubasch, U., Perlitz, J., Roeckner,
1070 E.: The Köppen climate classification as a diagnostic tool for general
1071 circulation models. Climate Research, 177–193 (1993)
- 1072 [79] Beck, C., Grieser, J., Kottek, M., Rubel, F., Rudolf, B.: Characterizing
1073 global climate change by means of Köppen climate classification.
1074 Klimastatusbericht **51**, 139–149 (2005)
- 1075 [80] Cui, D., Liang, S., Wang, D.: Observed and projected changes in global
1076 climate zones based on Köppen climate classification. Wiley Interdisci-
1077 plinary Reviews: Climate Change **12**(3), 701 (2021)
- 1078 [81] Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and
1079 validation of cluster analysis. Journal of Computational and Applied

1080 Mathematics **20**, 53–65 (1987)

- 1081 [82] Syakur, M., Khotimah, B., Rochman, E., Satoto, B.D.: Integration k-
1082 means clustering method and elbow method for identification of the best
1083 customer profile cluster. In: IOP Conference Series: Materials Science and
1084 Engineering, vol. 336, p. 012017 (2018). IOP Publishing
- 1085 [83] Jiang, S., Ren, L., Hong, Y., Yang, X., Ma, M., Zhang, Y., Yuan, F.: Improvement of multi-satellite real-time precipitation products for ensemble
1086 streamflow simulation in a middle latitude basin in South China.
1087 Water Resources Management **28**(8), 2259–2278 (2014)
- 1088 [84] Tang, G., Clark, M.P., Papalexiou, S.M., Ma, Z., Hong, Y.: Have satellite
1089 precipitation products improved over last two decades? A comprehensive
1090 comparison of GPM IMERG with nine satellite and reanalysis datasets.
1091 Remote Sensing of Environment **240**, 111697 (2020)
- 1092 [85] Nerini, D., Monestiez, P., Manté, C.: Cokriging for spatial functional data.
1093 Journal of Multivariate Analysis **101**(2), 409–418 (2010)

1095 A Derivation of equations in Algorithm 1

1096 We obtain the equations in Algorithm 1, by differentiating the loss function

$$L = \sum_{i=1}^m \xi_i \left\| \mathbf{Y}_i^{[G]} - \sum_{j=0}^2 \mathbf{Z}_{ji} \delta_{ji} \right\|_2^2$$

$$\begin{aligned} \frac{\partial L}{\partial \lambda_\ell} &= \sum_{i=1}^m \xi_i \left(-2 (\mathbf{w}'_{\ell i} \boldsymbol{\beta}_\ell) (\mathbf{Z}'_{\ell i} \mathbf{Y}_i^{[G]}) + 2 \delta_{\ell i} (\mathbf{w}'_{\ell i} \boldsymbol{\beta}_\ell) (\mathbf{Z}'_{\ell i} \mathbf{Z}_{\ell i}) + 2 (\mathbf{w}'_{\ell i} \boldsymbol{\beta}_\ell) \sum_{j \neq \ell} \delta_{ji} \mathbf{Z}'_{\ell i} \mathbf{Z}_{ji} \right) \\ \frac{\partial L}{\partial \boldsymbol{\alpha}_\ell} &= \sum_{i=1}^m \xi_i \left(-2 \mathbf{x}_i (\mathbf{Z}'_{\ell i} \mathbf{Y}_i^{[G]}) + 2 \delta_{\ell i} \mathbf{x}_i (\mathbf{Z}'_{\ell i} \mathbf{Z}_{\ell i}) + 2 \mathbf{x}_i \sum_{j \neq \ell} \delta_{ji} \mathbf{Z}'_{\ell i} \mathbf{Z}_{ji} \right). \end{aligned}$$

1098 with respect to $\boldsymbol{\lambda}_\ell$ and $\boldsymbol{\alpha}_\ell$, where $\mathbf{Z}_{0i} = \mathbf{1}_{n_i}$, $\mathbf{Z}_{1i} = \mathbf{Y}_i^{[J]}$, $\mathbf{Z}_{2i} = \mathbf{Y}_i^{[N]}$, and
1099 $\delta_{ji} = \mathbf{x}'_i \boldsymbol{\alpha}_j + \lambda_j \mathbf{w}'_{ji} \boldsymbol{\beta}_j$ for $j = 0, 1, 2$. We get

$$\begin{aligned} \lambda_\ell &= \left[\sum_{i=1}^m \xi_i (\mathbf{w}'_{\ell i} \boldsymbol{\beta}_\ell) (\mathbf{Z}'_{\ell i} \mathbf{Z}_{\ell i}) (\boldsymbol{\beta}'_\ell \mathbf{w}_{\ell i}) \right]^{-1} \left(\sum_{i=1}^m \xi_i \left\{ (\mathbf{Z}'_{\ell i} \mathbf{Z}_{\ell i}) (\mathbf{Z}'_{\ell i} \mathbf{Y}_i^{[G]}) \right. \right. \\ &\quad \left. \left. - (\mathbf{w}'_{\ell i} \boldsymbol{\beta}_\ell) (\mathbf{Z}'_{\ell i} \mathbf{Z}_{\ell i}) (\mathbf{x}'_i \boldsymbol{\alpha}_\ell) - (\mathbf{w}'_{\ell i} \boldsymbol{\beta}_\ell) \sum_{j \neq \ell} \delta_{ji} \mathbf{Z}'_{\ell i} \mathbf{Z}_{ji} \right\} \right) \end{aligned}$$

1101

$$\alpha_\ell = \left[\sum_{i=1}^m \xi_i \mathbf{x}_i (\mathbf{Z}'_{\ell i} \mathbf{Z}_{\ell i}) \mathbf{x}'_i \right]^{-1} \left(\sum_{i=1}^m \xi_i \left\{ \mathbf{x}_i \left(\mathbf{Z}'_{\ell i} \mathbf{Y}^{[G]}_i \right) - \mathbf{x}_i \sum_{j \neq \ell} \delta_{ji} \mathbf{Z}'_{\ell i} \mathbf{Z}_{ji} \right. \right.$$

1102

$$\left. \left. - \mathbf{x}_i (\mathbf{Z}'_{\ell i} \mathbf{Z}_{\ell i}) (\lambda_\ell \mathbf{w}'_{\ell i} \beta_\ell) \right\} \right).$$

1103

B Supplementary Material

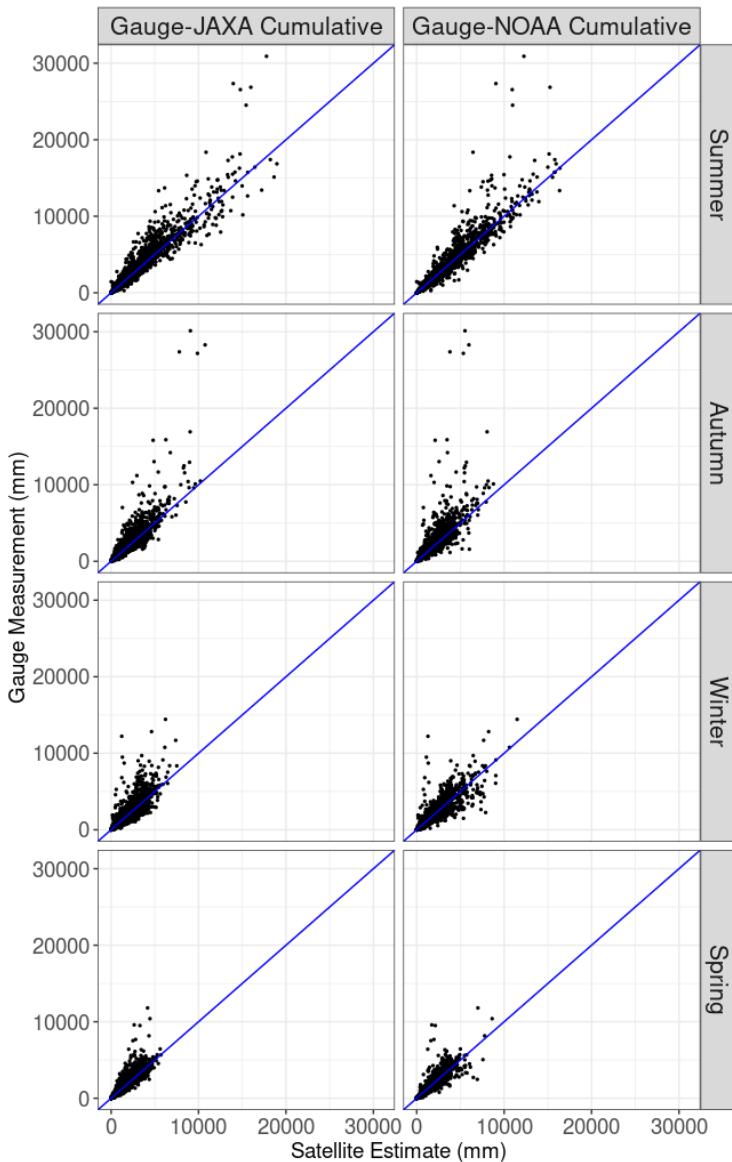


Fig. 17 Comparisons of the cumulative gauge measurements against the cumulative JAXA (left) and NOAA (right) satellite estimates of precipitation for each location, in summer (first row), autumn (second row), winter (third row), and spring (fourth row). The blue line represents $y = x$.

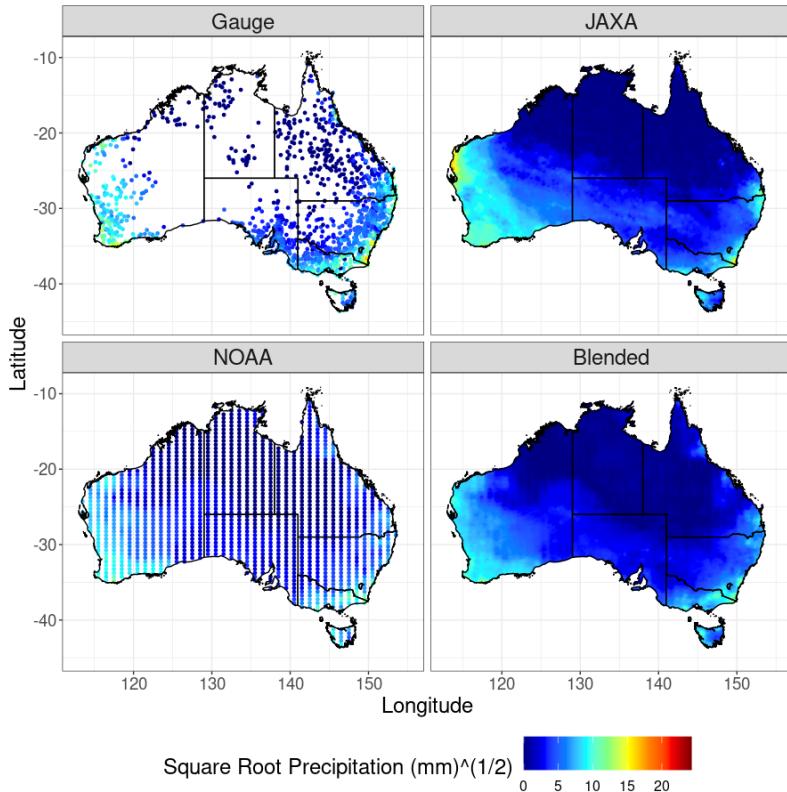


Fig. 18 May 2021 square root BOM rain gauge measurements (top left), square root satellite precipitation estimates for the JAXA (top right) and NOAA (bottom left), and the square root blended values (bottom right) using the model in equation (11).

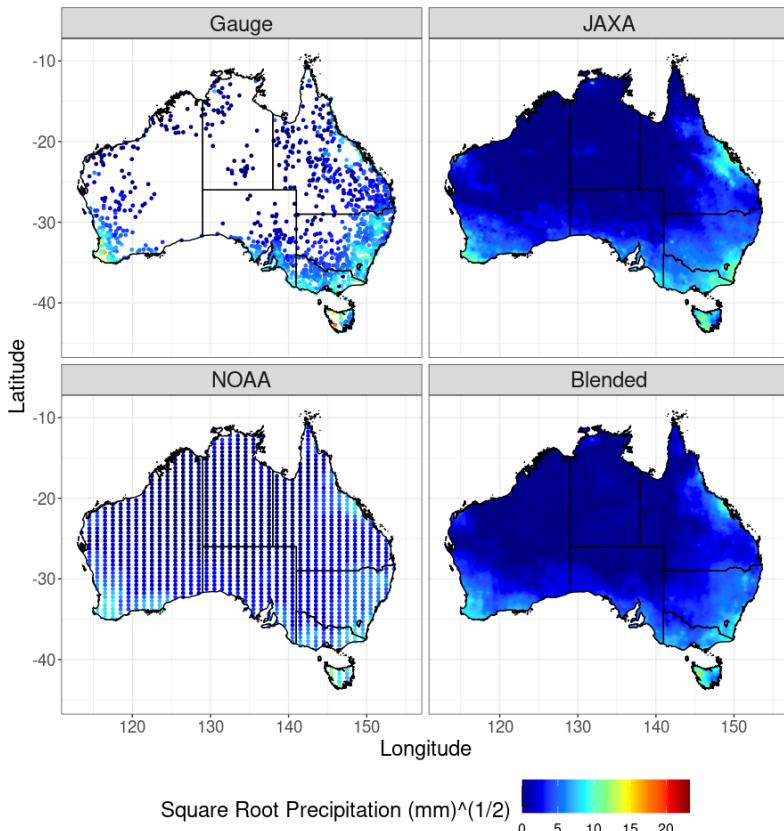


Fig. 19 August 2021 square root BOM rain gauge measurements (top left), square root satellite precipitation estimates for the JAXA (top right) and NOAA (bottom left), and the square root blended values (bottom right) using the model in equation (11).

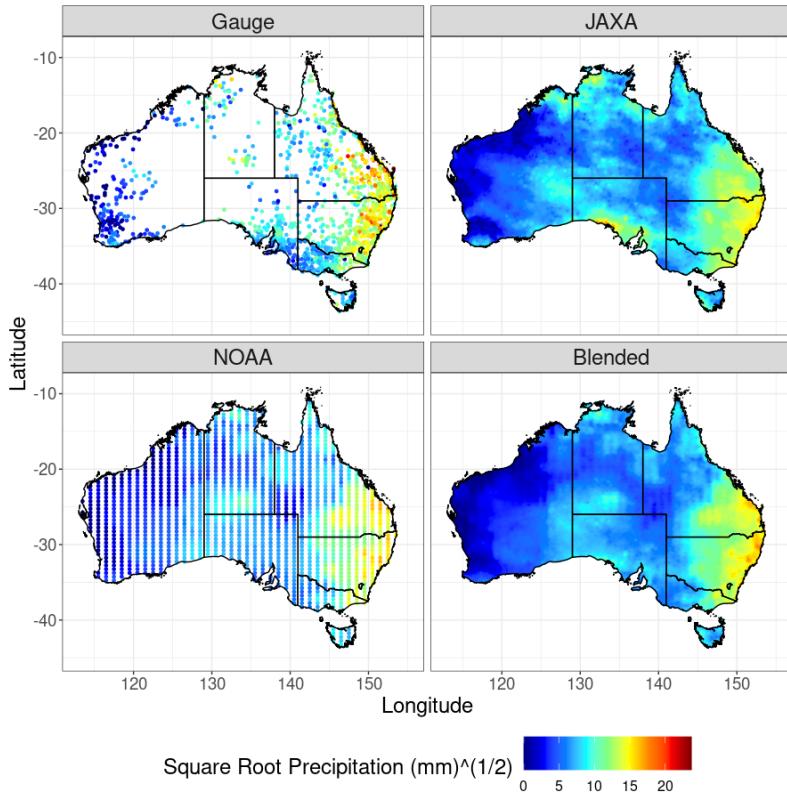


Fig. 20 November 2021 square root BOM rain gauge measurements (top left), square root satellite precipitation estimates for the JAXA (top right) and NOAA (bottom left), and the square root blended values (bottom right) using the model in equation (11).

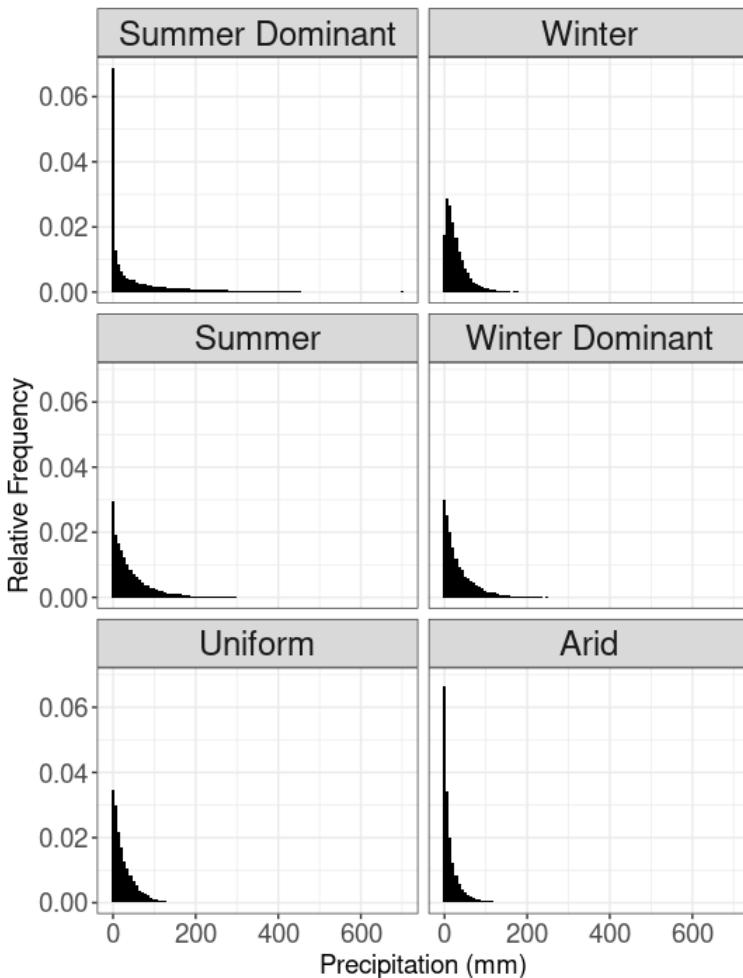


Fig. 21 Histograms for the amount of precipitation across each cluster defined through the Bureau of Meteorology's six major rainfall zones from April 2000 to March 2022.

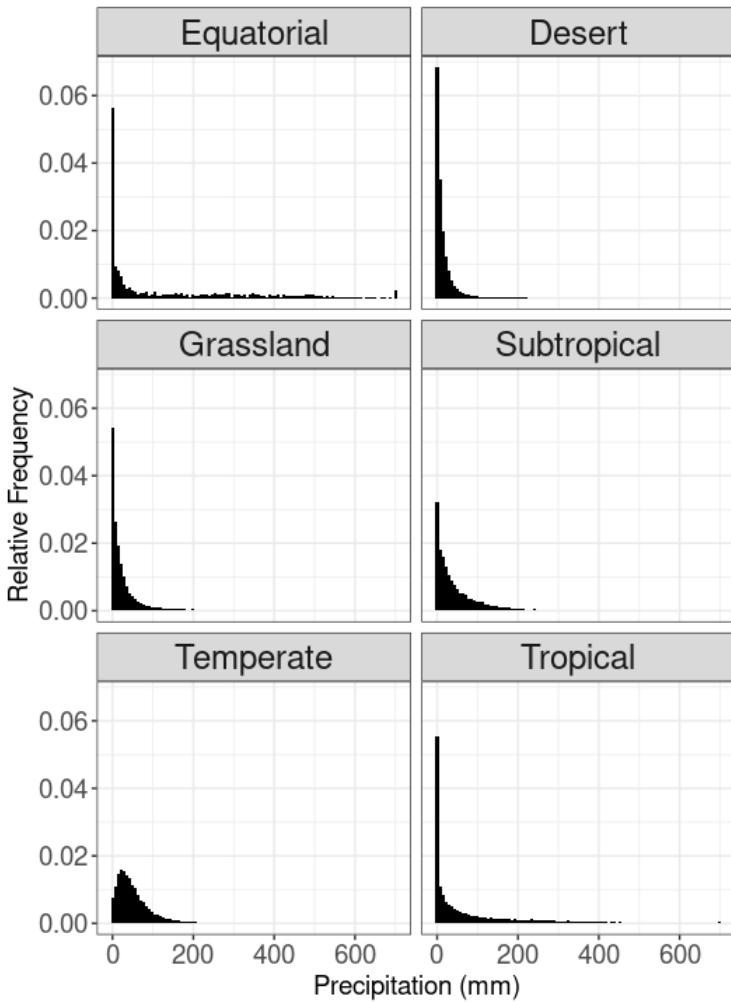


Fig. 22 Histograms for the amount of precipitation across each cluster defined through the Bureau of Meteorology's six major Köppen climate regions from April 2000 to March 2022.