



# 1 CAMELS-AUS: Hydrometeorological time series and landscape 2 attributes for 222 catchments in Australia

3 Keirnan J. A. Fowler<sup>1</sup>, Suwash Chandra Acharya<sup>1</sup>, Nans Addor<sup>2</sup>, Chihchung Chou<sup>1\*</sup> and Murray C. Peel<sup>1</sup>

4 <sup>1</sup>Department of Infrastructure Engineering, University of Melbourne, Parkville, Victoria, Australia

5 <sup>2</sup>Department of Geography, University of Exeter, Exeter, UK

6 \*now at: Department of Earth Sciences, Barcelona Supercomputing Centre, Barcelona, Spain

7 *Correspondence to:* Keirnan Fowler (fowler.k@unimelb.edu.au)

## 8 **Abstract.**

9 This paper presents the Australian edition of the Catchment Attributes and Meteorology for Large-sample Studies (CAMELS)  
10 series of datasets. CAMELS-AUS comprises data for 222 unregulated catchments, combining hydrometeorological timeseries  
11 (streamflow and 18 climatic variables) with 134 attributes related to geology, soil, topography, land cover, anthropogenic  
12 influence, and hydroclimatology. The CAMELS-AUS catchments have been monitored for decades (more than 85% have  
13 streamflow records longer than 40 years) and are relatively free of large scale changes, such as significant changes in landuse.  
14 Rating curve uncertainty estimates are provided for most (75%) of the catchments and multiple atmospheric datasets are  
15 included, offering insights into forcing uncertainty. This dataset, the first of its kind in Australia, allows users globally to  
16 freely access catchment data drawn from Australia's unique hydroclimatology, particularly notable for its large interannual  
17 variability. Combined with arid catchment data from the CAMELS datasets for the USA and Chile, CAMELS-AUS constitutes  
18 an unprecedented resource for the study of arid-zone hydrology. CAMELS-AUS is freely downloadable from  
19 <https://doi.pangaea.de/10.1594/PANGAEA.921850> (Fowler et al., 2020a).

## 20 **1 Introduction**

21 For some time, the ideals of 'comparative hydrology' and 'large-sample hydrology' have been advanced as complementary  
22 and necessary components of hydrology (eg. Falkenmark and Chapman 1989; Andréassian et al, 2006; Gupta et al. 2014).  
23 Alongside traditional hydrological studies, which may focus on a single catchment, or possibly compare results among several  
24 catchments within a region, large-sample studies aim to establish the generality of results and to test paradigms applicable on  
25 regional-to-global scales (eg. McMahon et al 1992; Peel et al., 2004; Kuenst et al, 2018; Ghiggi et al., 2019; Mathevet et al.,  
26 2020). Large samples of catchments are also insightful for certain tasks, such as prediction in ungauged basins (eg. Pool et  
27 al., 2019) or training and evaluation of machine learning algorithms (eg. Kratzert et al, 2018; Shen, 2018; Kratzert et al., 2019).  
28 Thus, large sample studies are a growing component of recent hydrological research (see review by Addor et al., 2019).



29

30 However, issues of data availability and commensurability, which are endemic to environmental sciences, are exacerbated for  
31 large sample hydrology. Large samples may cross jurisdictions or data providers, require harmonisation across different data  
32 formats or nomenclatures (eg. quality codes) and are more likely to suffer from spatial gaps due to different data sharing  
33 policies of water agencies (Viglione et al., 2010; Addor et al., 2019). Thus, the importance of FAIR data (Findable, Accessible,  
34 Interoperable and Reusable, see Wilkinson et al. 2016 and the Open Data Charter 2015) in hydrology is amplified in large  
35 sample hydrology and there is a clear need for open publication of datasets wherever possible to allow equal access. Such  
36 policies also encourage hydrologists to work across boundaries – an important ideal since the spatial distribution of  
37 hydrologists globally does not reflect the spread of interesting hydrological environs, nor the pressing need for hydrological  
38 insights to inform policy.

39

40 Responding to these needs, multiple recent projects have publicly released large sample hydrological datasets (e.g., Arsenault  
41 et al., 2016; Do et al., 2018; Lin et al., 2019; Linke et al., 2019; Olarinoye et al., 2020). Here we contribute to one such ongoing  
42 project – the Catchment Attributes and Meteorology for Large-sample Studies, or CAMELS, project. Originally launched for  
43 the United States (Newman et al., 2015; Addor et al., 2017), CAMELS datasets now exist for Chile (Alvarez Garreton et al.,  
44 2018), Great Britain (Coxon et al., 2020) and Brazil (Chagas et al., 2020). The defining features of a CAMELS dataset are  
45 that they complement data on streamflow (which is often publicly available) with other data types: (i) pre-processed climatic  
46 data for each catchment, such as would be required to run a hydrological model; and (ii) catchment attributes which characterise  
47 various aspects of the catchment without the need for field visitation (impractical for large samples). They also support  
48 download of the entire dataset in contrast to agency websites which may only support one-at-a-time download (if at all).

49

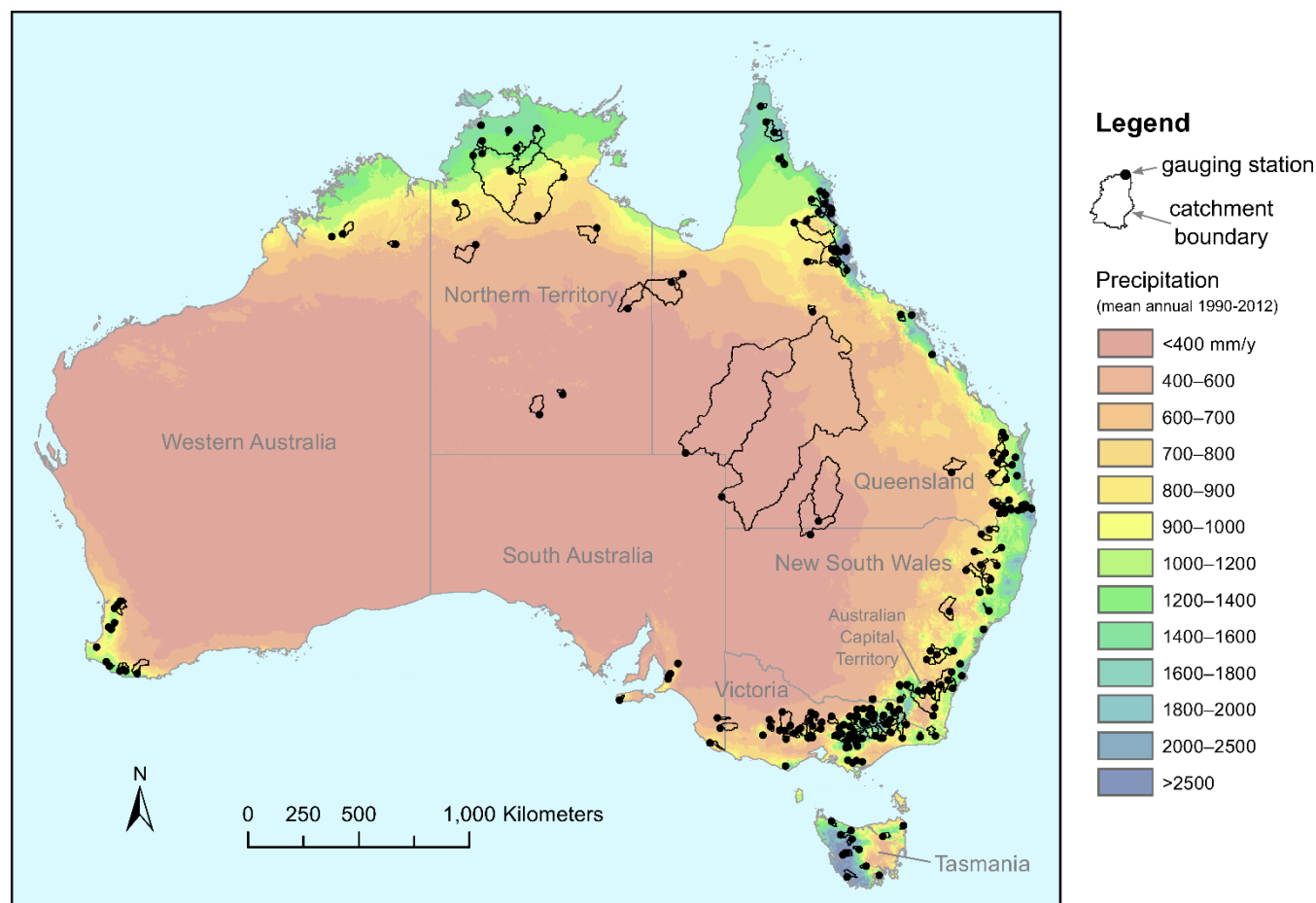
50 The present dataset focusses on the continent of Australia, including the southern state of Tasmania, but excluding other  
51 territories held by the Australian nation. Australia is the world's sixth largest country (approximately  $7.7 \times 10^6$  km<sup>2</sup>) and is  
52 comparable in size to the conterminous USA or Europe, but the hydrologically active parts of the country tend to be limited to  
53 coastal regions, with the interior being semi-arid or arid (Figure 1; see also Knoben et al., 2018). Thus, dense gauging of  
54 streamflow covers only a small proportion of the total area, with the remaining areas providing few gauged locations. While  
55 sparsely gauged, the dry parts of Australia provide interesting arid-zone catchment examples, many of which are included in  
56 the CAMELS-AUS dataset. In addition to arid regions, Australia includes northern areas with tropical climate and southern  
57 areas with temperate climate.

58

59 This paper is structured as follows. In Section 2 we describe the rationale for the dataset, including considerations of why  
60 Australian hydroclimate is interesting and relevant to hydrologists globally; and factors shaping the dataset, including local  
61 data availability. Section 3 provides a technical description of the dataset and forms the bulk of the paper. Sections 4 and 5  
62 explain CAMELS-AUS data availability and conclude the manuscript, respectively.



63  
64



65

66

67

**Figure 1: Location of the 222 CAMELS-AUS flow gauging stations and catchments, along with mean annual precipitation (from Jones et al., 2009) and Australian states and territories.**

68

## 2 Rationale

69

This section lays out the motivations underpinning the release of this dataset for Australia. It focusses on why CAMELS-AUS takes its present form, including two chief aspects: catchment selection; and inclusion of local versus global datasets.

70

71

### 2.1 Motivation: Australian hydroclimate and its place in the study of arid-zone hydrology and hydrology under climatic change

72

73

Every region on earth is unique and has characteristics of interest for hydrological study. Within Australia and for CAMELS-AUS, two characteristics are noted here. Firstly, Australia contains many arid landscapes, and considerable advances in arid-

74



75 zone hydrology have been made there (eg. Western et al., 2020). CAMELS-AUS contains more than twenty arid-zone rivers  
76 (depending on definition but see Figure 1), so the publication of the dataset opens the study of these rivers to a global pool of  
77 scientists. Added together with included arid-zone rivers in the USA and Chile (Addor et al., 2017; Alvarez-Garreton et al.,  
78 2018), the CAMELS datasets together provide a significant sample for the study of arid-zone hydrology.

79  
80 Another notable characteristic of Australian hydroclimatology is its tendency for multi-year spells of relatively deep climatic  
81 anomalies compared to most other regions of the world (Peel et al., 2005), due partly to the strong influence of climate  
82 teleconnections such as the El Nino Southern Oscillation (ENSO, eg. Peel et al., 2002; Verdon-Kidd and Kiem, 2009). Recent  
83 severe droughts have affected south-east Australia, including the 13-year Millennium Drought (Van Dijk et al., 2014) which  
84 provided the opportunity for knowledge sharing with other drought-prone regions (Aghakouchak et al., 2014) and supplied  
85 many case studies of model failure (eg. Saft et al., 2016), which are under ongoing investigation (eg. Fowler et al., 2020b). In  
86 the context of providing credible runoff projections in regions subject to drying climate, it is hoped that the public release of  
87 datasets such as CAMELS-AUS may hasten scientific progress towards more defensible and robust hydrological models.

## 88 **2.2 Context: hydrometeorological monitoring in Australia**

89 Systematic climatic measurement in Australia extends back to the late 1800s (eg. Ashcroft et al., 2014), with widespread  
90 streamflow gauging of headwater catchments commencing from the 1950s and 60s. Meteorological monitoring is the  
91 responsibility of a federal Bureau of Meteorology (BOM), but streamflow monitoring falls to the states and territories of  
92 Australia, rather than the federal government (Skinner and Langford, 2013). Thus, Australian streamflow data has historically  
93 been dispersed between its six states and two territories (Figure 1) and, while quality control is relatively well-established,  
94 methods and formats (eg. quality codes) are not consistent between states and territories. Since the 2000s this situation has  
95 partially been rectified after federal legislation required the BOM to collate data from the states under new ‘Water Information’  
96 powers (Vertessy, 2015).

## 97 **2.3 Catchment choice: the Hydrologic Reference Stations dataset**

98 Under its new responsibilities the BOM initiated several national hydrological projects, one of which is called the *Hydrologic*  
99 *Reference Stations* project (Turner et al., 2012). This project selected a large set of gauging stations, each on unregulated  
100 streams, to serve as a “platform to investigate long-term trends in water resource availability” (Turner et al., 2012, p. 1555).  
101 The project has a website for provision of streamflow data to the public ([www.bom.gov.au/water/hrs/](http://www.bom.gov.au/water/hrs/)).

102

103 We adopted the Hydrologic Reference Stations as the basis for CAMELS-AUS, for three reasons:

- 104 • The selection criteria used by the BOM, including record length, lack of regulation, and stationarity of anthropogenic  
105 influence (see Section 3.2) are consistent with the aim of the CAMELS project to provide high-quality scientific data;



- 106 • Considerable effort has already been expended by the BOM to standardise and quality check the streamflow data,  
107 which was only possible via contacts with state agencies that are not necessarily available to academic authors (for  
108 an example, see BOM, 2020). It is logical to take advantage of this prior effort; and
- 109 • The Hydrologic Reference Stations have attained a degree of acceptance within the Australian hydrological  
110 community, partly due to extensive consultation with stakeholders during development (see Section 3.2). Also, they  
111 have been adopted by numerous academic studies (eg. Zhang et al., 2014; 2016; Wright et al., 2018; McInerney et  
112 al., 2017; Fowler et al., 2016, 2018, 2020b).

113  
114 It is noted that this choice is not intended to limit future inclusion of a wider range of stations/catchments. We envisage that  
115 the Hydrologic Reference Stations may provide the nucleus for future versions of the CAMELS-AUS dataset, while the current  
116 selection provides a sensible and pragmatic starting point. The Hydrological Reference Station dataset itself may be subject  
117 to future expansion, which would inform future CAMELS-AUS versions.

## 118 **2.4 Local versus global datasets**

119 A key choice in developing CAMELS-AUS was whether to use local or global datasets (or both) when extracting  
120 hydrometeorology time series and catchment attributes. On the one hand, global datasets are important to facilitate  
121 intercontinental comparisons. On the other hand, when local datasets are available, they are generally the highest quality  
122 information that exists for a given region (eg. Acharya et al., 2019). With the advent of large-sample hydrology, it is now  
123 possible to conduct near-global studies using very large samples of catchments (eg. over two thousand in Mathevet et al., 2020)  
124 and future studies might compose such large samples by combining continental-scale datasets like the various CAMELS.  
125 However, the lack of standardised approaches and sources between national large sample datasets remains a key limitation of  
126 large-sample studies (Addor et al., 2019).

127  
128 The approach followed by the CAMELS datasets so far is to use the best possible data available for each country, so national  
129 datasets have been prioritised over global datasets. In some cases, global datasets have been employed, for instance the Global  
130 Lithological Map (Hartmann and Moosdorf, 2012) in CAMELS and CAMELS-CL or the Multi-Source Weighted-Ensemble  
131 Precipitation (Beck et al., 2017) in CAMELS-CL. But overall, the best national data products were selected for each country,  
132 leveraging the knowledge of CAMELS creators. This enables global users, who may not be familiar with these national  
133 products, to benefit from this local knowledge. It also gives direct access to the best available data to users whose study  
134 focusses on catchments from a single country (see, eg., intercomparisons in Acharya et al., 2019). In keeping with this  
135 approach, the priority was given to national data products to produce CAMELS-AUS.

136  
137 In parallel, efforts are ongoing to increase the consistency among the CAMELS datasets (in terms of data products used to  
138 derive the time series and catchment attributes, and also naming conventions and data format, see Addor et al., 2019), in order



139 to create a dataset that is globally consistent. This is part of a second phase, which will build upon the current phase which is  
140 focussed on the release of national products, such as CAMELS-AUS. To contribute to this effort, we have supplied the  
141 CAMELS-AUS catchment boundaries and gauge locations. Because of these ongoing efforts, our expectation is that the data  
142 introduced here, derived from Australian sources, will in time be complimented by data derived from global datasets.

### 143 3 CAMELS-AUS dataset technical description

144 The previous section outlined key decisions made for CAMELS-AUS, ie. it is based on the Hydrological Reference Stations,  
145 and its data are derived from Australian rather than global sources. This section provides more detail and presents each aspect  
146 of the dataset in turn. Work not undertaken by the present authors (eg. earlier efforts by the BOM for the Hydrological  
147 Reference Stations project) is clearly marked. In many cases, subsections end with a “*Included in dataset:*” sentence to clearly  
148 outline items in the online repository related to the sub-section text.

149  
150 Before presenting the detail, we note that the online repository of the dataset (Fowler et al., 2020a) includes the following:

- 151 • A file containing the overall attribute table, containing all non-timeseries data (see Tables 1, 3 and 4); and
- 152 • 27 timeseries files, each containing data for all catchments for a given hydroclimatic variable (see Table 2).
- 153 • Extra files such as shapefiles and readme files as noted below.

#### 154 3.1 Catchment selection rules

155 Given the decision (Section 2.2 above) to base the CAMELS-AUS dataset on the BOM’s Hydrologic Reference Stations, this  
156 subsection summarises the process of catchment selection undertaken earlier by the BOM. As described in Turner et al. (2012):

- 157 • **Initial selection:** 246 potential stations were initially selected based on three criteria: (i) record length (minimum of  
158 1975 onwards); (ii) availability of data including historic rating curve information; and (iii) lack of regulation by large  
159 dams.
- 160 • **Invitation for stakeholders to suggest additional stations:** BOM consulted with seventy stakeholders from federal,  
161 state and territory agencies and water authorities, who were given the opportunity to add new stations to the list. This  
162 enlarged the list to 362 stations.
- 163 • **Targeted fact-finding:** To elicit information about each candidate station/catchment, the relevant agencies were  
164 asked a series of questions about the catchments in their jurisdiction relating both to past and present practices. Topics  
165 included diversions, irrigation structures, upstream point source discharge, land clearing, forestry, urbanisation, fire,  
166 and farm dams.
- 167 • **Final selection:** the final selection process considered all the above information. A good coverage of Australia’s  
168 various hydroclimatic regions was desired, although this is inherently limited by the coverage of the gauging network.  
169 Where possible, only stations with < 5% missing data and < 10% change in forest cover were selected.



170

171 The above process provided the first version of the Hydrologic Reference Stations, with a total of 221 catchments.  
172 A subsequent update in 2015, which included a detailed review and update of streamflow data up to 2014 (BOM, 2020),  
173 resolved to retain all existing stations and add one more (ID 215207). Thus, the final number of stations is 222 (Figure 1).

174

175 *Included in dataset:* The following variables are provided in the CAMELS-AUS attribute table (see Table 1): Station ID,  
176 station name (including river name and station name), drainage division and river region (out of 13 drainage divisions and 218  
177 river regions across Australia). Unfortunately, information is not available about which catchments were included or excluded  
178 under the above rules.

### 179 3.2 Catchment boundaries

180 For all but ten of the catchments, catchment boundaries were derived via flow path analysis (using Esri's Arc Hydro) of  
181 topographic data undertaken by the authors. The input data was: (i) the post-processed and hydrologically enforced DEM of  
182 Gallant et al. (2012) which is derived from the 1-second (approximately 30 m) grid Shuttle Radar Topography Mission (SRTM)  
183 dataset; and (ii) the location of the streamflow gauges as provided by the BOM. The Arc Hydro analysis determines the  
184 apparent position of streams from the DEM data, and it was found that the published locations rarely fall precisely on these  
185 digital streamlines. The mismatch is unsurprising given location data may be decades old and significant figures may have  
186 been truncated with the passage of data between databases (or never reported in the first place). Also, the position of the digital  
187 streamline may or may not match reality, particularly in flat landscapes. To derive catchment areas, the BOM-published gauge  
188 locations were shifted to the nearest streamline with expected catchment area. This movement was generally less than 200 m.  
189 As noted, this method was used for most catchments, with the following exceptions:

- 190 • For the six largest catchments (A0030501, A0020101, G8140040, G9030250, 424002 and 424201A), this process  
191 was not undertaken due to excessive computational requirements. For context, the largest catchment is approximately  
192 the size of the United Kingdom (see Figure 1);
- 193 • For a further four catchments (A2390519, A2390523, 307473 and 606185), the Arc Hydro process resulted in a  
194 catchment boundary that was inconsistent with the boundaries displayed on the Hydrologic Reference Station website.  
195 Although severely degraded for fast mapping, the website boundaries show the approximate position of the boundary  
196 as agreed with stakeholders / agencies who have local knowledge. Therefore, in cases of obvious mismatch, the Arc  
197 Hydro-derived boundaries were assumed to be in error. Despite the 'blockiness' of the website boundaries, they were  
198 considered to be a better option for these four catchments.

199

200 For these ten catchments a protocol was developed to read the website's .json file to extract the boundary vertices. The website  
201 boundaries were then adopted. Note, more detail on the above considerations, including a selection of figures, is given in the  
202 dataset within the readme file README\_CAMELS\_AUS\_Boundaries.pdf.



203

204 *Included in dataset:* The main inclusions are a point shapefile of adopted gauge locations and a polygon shapefile of adopted  
205 catchment areas. Further information included are: point shapefile of BOM published gauge locations; polygon shapefile of  
206 website mapped boundaries; readme file explaining the above logic but in more detail and with figures. As listed in Table 1,  
207 the CAMELS-AUS attribute table lists the coordinates of the catchment outlet and centroid, along with notes which expand  
208 on issues listed above, on a catchment-by-catchment basis.

### 209 **3.3 Catchment area and nestedness**

210 To calculate catchment areas, the catchment boundaries were first projected into the appropriate local coordinate system under  
211 the Map Grid of Australia (MGA). Due to Australia's size, the MGA defines different coordinate systems based on  
212 longitude. Using the catchment centroid, each catchment was placed within a zone and this zone was used to calculate area  
213 using the standard tool within Esri's ArcMap. Inspection of catchment boundaries revealed that some of the catchments are  
214 'nested' (ie. entirely contained) within others, for example, when two gauges lie on the same stream (one downstream of the  
215 other) and both have been included in the dataset. The upstream (ie. entirely contained) catchments are clearly marked in the  
216 CAMELS-AUS attribute table (see Table 1). Catchments containing nested catchments are also marked.

217

218 Before moving on from considerations of spatial data, it is noted that: (i) CAMELS-AUS does not come with a spatial layer  
219 for the river network; (ii) users may find the 15s Hydrosheds River Network ([www.hydrosheds.org/downloads](http://www.hydrosheds.org/downloads)) or the BoM  
220 Geofabric v2 SH network ([www.bom.gov.au/water/geofabric/download.shtml](http://www.bom.gov.au/water/geofabric/download.shtml)) useful; and (iii) the reason these are not  
221 included in CAMELS-AUS is because of licencing concerns (for Hydrosheds) and file size concerns (for the Geofabric).

222

223 *Included in dataset:* The following variables are provided in the CAMELS-AUS attribute table (see Table 1): catchment area,  
224 map zone, and three indicators related to nestedness (NestedStatus, NextStationDS, NumNestedWithin).

### 225 **3.4 Streamflow data and uncertainty**

226 Streamflow timeseries data are provided by the BOM in two variants: non gap filled, and gap filled. The gap filled variant is  
227 filled using the daily rainfall-runoff model GR4J (Perrin et al., 2001) but no details are provided about calibration method,  
228 validation procedures, or the specifics of the interpolation method. The BOM also provide quality codes. As mentioned in  
229 Section 2.1, the quality codes of each state of Australia are different but the BOM has harmonised these to a common set  
230 ([www.bom.gov.au/water/hrs/qc\\_doc.shtml](http://www.bom.gov.au/water/hrs/qc_doc.shtml)). For CAMELS-AUS, these data are supplied as follows. Firstly, summary  
231 statistics about period of record (start date, end date and proportion of missing data) are provided in the attribute table, as listed  
232 in Table 1. Regarding timeseries data (Table 2), each of the above three data types (gap filled, non-gap-filled, and quality  
233 codes), are provided within CAMELS-AUS exactly as supplied by the BOM, except that they are presented as a single file

234





235

236

**Table 1: Basic catchment information provided in the attribute table of CAMELS-AUS**

| Short name               | Description  | Data source / notes  |
|--------------------------|--|--|
| <i>station_id</i>        | Station ID used by the Australian Water Resources Council.   | Hydrologic Reference Stations (HRS) project, Bureau of Meteorology (BOM)<br><a href="http://www.bom.gov.au/water/hrs">www.bom.gov.au/water/hrs</a> |
| <i>station_name</i>      | River name and station name  |  |
| <i>drainage_division</i> | Drainage division, of the 13 defined by the BOM.   |  |
| <i>river_region</i>      | River region, of the 218 defined by the BOM.   |  |
| <i>notes</i>             | General notes about data issues and/or catchment area calculations   | This study   |
| <i>lat_outlet</i>        | Latitude and longitude at outlet. Note, in most cases this will be slightly different to the BoM published value because most outlets needed to be moved onto a digital streamline in order to facilitate flow path analysis.  |  |
| <i>long_outlet</i>       |  |  |
| <i>lat_centroid</i>      | Latitude and longitude at centroid of the catchment.   |  |
| <i>long_centroid</i>     |  |  |
| <i>map_zone</i>          | Map zone used to calculate catchment area (function of longitude)  |  |
| <i>catchment_area</i>    | Area of upstream catchment in km <sup>2</sup>  |  |
| <i>nested_status</i>     | "Not nested" indicates the catchment is not contained within any other. "Level1" means it is contained within another, except in cases where it is contained in another "Level1" catchment in which case it is marked "Level2". There are no "Level3" catchments in the present dataset. |  |
| <i>next_station_ds</i>   | For "Level1" and "Level2" nested catchments, <i>NextStationDS</i> ('DS' meaning downstream) indicates the catchment they are contained within.   |  |
| <i>num_nested_within</i> | Indicates how many catchments are nested within this catchment.  |  |
| <i>start_date</i>        | Streamflow gauging start date (yyyymmdd)   | HRS (see above)  |
| <i>end_date</i>          | Streamflow gauging end date (yyyymmdd)   |  |
| <i>prop_missing_data</i> | Proportion of data missing between startdate and enddate   |  |

237

238

239

240

241

242

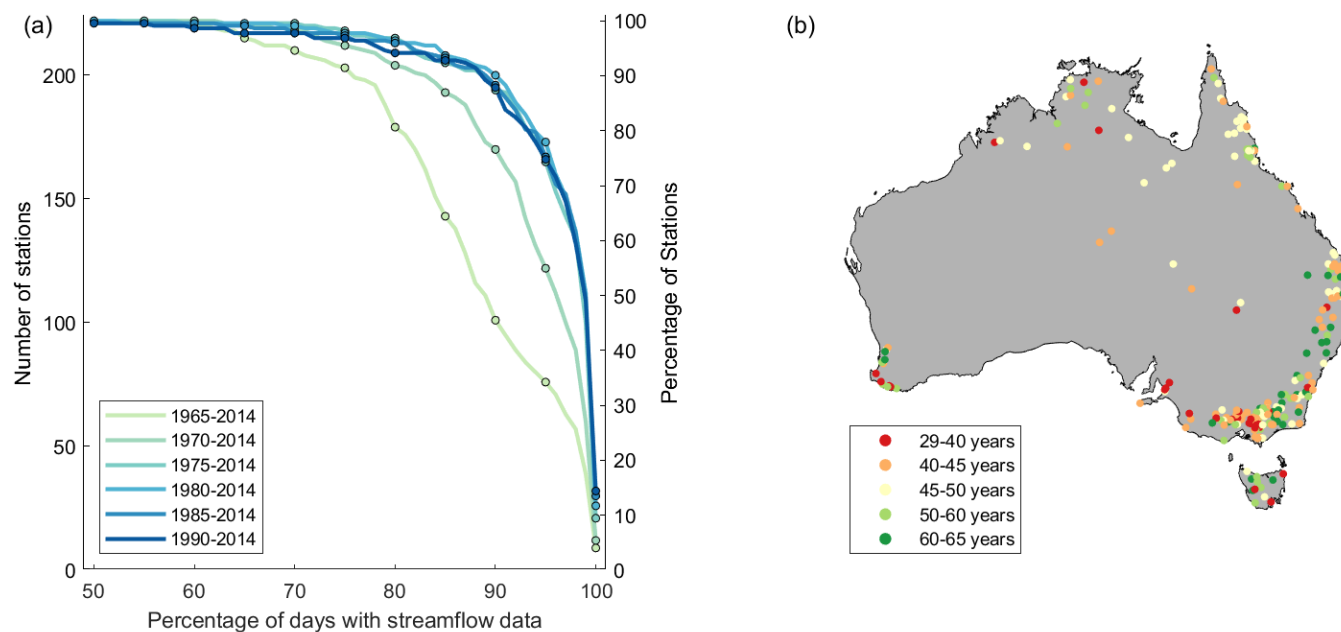
243

244

245

across all catchments. In addition, since the units of the streamflow files are ML d<sup>-1</sup> whereas modelling studies typically use mm d<sup>-1</sup>, CAMELS-AUS provides an additional streamflow timeseries file in mm d<sup>-1</sup>.

Figure 2 shows that CAMELS-AUS stations are typically long term gauges, with the shortest record being 29 years. All but 17 gauges commence by 1975 (in line with the selection rules in Section 3.1) and all but 22 of the records contain data up until the cut-off date for this dataset which is 31<sup>st</sup> December 2014. Thus, records longer than 40 years are typical (Figure 2b). Figure 2a considers both the record extent and missing data to determine the overall data availability for different overlapping periods. The data availability for the periods starting in 1965 and 1970 are lower than the others, as expected given the remarks



246

247 **Figure 2: Plot after Coxon et al. (2020) showing (a) Number of stations with percentage of available streamflow data for different**  
248 **periods, b) Length of the flow time series for each gauge.**

249

250 about record length. An increase in missing data post-1990 means that the data availability curves decrease slightly for the  
251 most recent period (dark blue).

252

253 Information about streamflow uncertainty is provided with CAMELS-AUS (Table 1) from an earlier study by McMahon and  
254 Peel (2019). McMahon and Peel (2019) examined available rating curve data for 166 of the 222 stations, developed rating  
255 curves based on Chebyshev polynomials and estimated uncertainties using an approach which considered regression error and  
256 uncertainty in water level. The original authors post-processed their data to provide the following statistics (Table 3) for  
257 CAMELS-AUS: (i) Number of separate rating curves considered for a given station (median value across all stations was 3);  
258 (ii) Number of days considered across all curves (median value was ~14,700 or ~40 years); (iii) low, medium and high flow  
259 rates in  $\text{mm d}^{-1}$  (discharge exceeded 90%, 50% and 10% of the time over days considered by the curves); (iv) 95% confidence  
260 intervals around the low, medium and high flow estimates, expressed in percentage terms. However, for some stations  
261 considered by McMahon and Peel (2019) the above data are not supplied in full, for the following reasons: (a) the percentile  
262 flow is zero (cease to flow), leading to undefined uncertainty estimates; (b) the percentile flow is outside the rated range, in  
263 which case neither upper or lower bounds are reported for that flow; and (c) the lower uncertainty bound goes below zero, in  
264 which case it is missing (but the upper bound is not). In a small number of cases the uncertainty bound numbers are very high,  
265 and these cases are generally associated with near-cease-to-flow conditions. For example, the highest value of



266 *Q\_uncert\_Q10\_upper* (refer Table 3 for naming conventions) occurs for catchment 919309A, for which Q10 is 0.000023 mm  
267 d<sup>-1</sup> but the upper bound is 0.05 mm d<sup>-1</sup>, which is >2000 times higher. Thus, *Q\_uncert\_Q10\_upper* for this catchment is  
268 201,400%.

269  
270 *Included in dataset:* Three streamflow timeseries files for flow, as explained above and listed in Table 2. One timeseries file  
271 for streamflow quality codes. In the CAMELS-AUS attribute table: three attributes related to record extent and availability  
272 (startdate, enddate, prop\_missing\_data; see Table 1) plus eleven attributes related to streamflow uncertainty  
273 (*Q\_uncert\_NumCurves*, *Q\_uncert\_N*, *Q\_uncert\_Q10*, *Q\_uncert\_Q10\_upper*, *Q\_uncert\_Q10\_lower*, *Q\_uncert\_Q50*,  
274 *Q\_uncert\_Q50\_upper*, *Q\_uncert\_Q50\_lower*, *Q\_uncert\_Q90*, *Q\_uncert\_Q90\_upper*, *Q\_uncert\_Q90\_lower*; see Table 3).

### 275 3.5 Hydrometeorological timeseries

#### 276 3.5.1 Availability of gridded hydrometeorological data in Australia

277 It is common practice in large sample hydrology studies to derive climate timeseries inputs by processing gridded data rather  
278 than directly using gauged point information (as is still common in industry). The first Australia-wide gridded climate product  
279 was the Scientific Information For Land Owners (SILO) project of the government of the State of Queensland (Jeffrey et al.,  
280 2001). Later, the BOM developed a separate set of climate grids under the Australian Water Availability Project (AWAP;  
281 Jones et al., 2009). SILO and AWAP are similar: they are both interpolated products based purely on the BoM's climate  
282 monitoring sites and (where relevant) incorporating topography as a co-variate. They both output grids on a resolution of  
283 0.05° × 0.05° (approximately 5 km). However, the datasets differ in the variables they provide: AWAP provides precipitation,  
284 temperature, vapour pressure and radiation, all of which SILO also provides in addition to vapour pressure deficit, and,  
285 importantly for modelling studies, various formulations of potential evapotranspiration (PET). They also differ in spatial  
286 interpolation method: the SILO method forces an exact match to measured values, whereas AWAP does not (Tozer et al.,  
287 2012). Both AWAP and SILO are commonly used in Australia. Rather than select one dataset over another, CAMELS-AUS  
288 includes both datasets and leaves the choice to users. When possible, users are encouraged to compare the datasets to obtain  
289 insights into interpolation uncertainty for the forcing data. For all AWAP and SILO variables, timeseries for each catchment  
290 were compiled by the CAMELS-AUS project by calculating the catchment spatial average separately for each day. The full  
291 available period was extracted which for most variables is 1900-2018 (SILO) and 1911-2017 (AWAP). Exceptions to these  
292 record extents are noted in the text below.

#### 293 3.5.2 Conventions for definition of daily timesteps

294 Variables such as precipitation and streamflow are continuous variables, and formatting into a daily timestep requires arbitrary  
295 conventions to split continuous time into 24 hour periods. For example, the BOM convention is that precipitation is split at  
296 9am each day, and a daily value refers to the precipitation that occurred over the preceding 24 hours. Thus, if the BOM reports



297 18 mm precipitation for 14th March, this means that 18 mm was recorded between 9am 13<sup>th</sup> March and 9am 14<sup>th</sup> March. For  
298 streamflow, the conventions may vary depending on state or territory, but in collating the HRS data the BOM claims that  
299 conventions have been standardised to 9am to 9am (ie. the same as precipitation). However, an audit of HRS data conducted  
300 by Jian et al. (2017) investigated this standardisation. They report that data from the states of Victoria, New South Wales,  
301 Queensland and the Australian Capital Territory (which together account for 168 of 222 stations) were consistent with the 9am  
302 to 9am claim. In contrast, they report that Western Australia (16 stations) data appear to be subject to a 01:00 split (ie. 8 hours  
303 earlier than expected) and South Australia and Northern Territory data (25 stations) appear to be subject to a 23:30 split (ie.  
304 9.5 hours earlier than expected). Modellers should be mindful of these points when designing studies and interpreting results.  
305 Regarding PET, the key variables (eg. temperature) are aligned directly with the day they are reported. This creates a time  
306 offset between PET and precipitation. In the experience of the CAMELS-AUS authors, this offset will typically make little  
307 difference to the results of (eg.) a rainfall runoff modelling study, since PET typically influences streamflow via seasonal, not  
308 daily, dynamics, in most CAMELS-AUS catchments. In the interests of providing CAMELS-AUS data subject to minimal  
309 manipulation, we do not apply a time shift to PET (or any other data), but users may wish to manually shift PET earlier by one  
310 day to minimise the time offset between precipitation and streamflow.

### 311 3.5.2 Precipitation

312 AWAP and SILO precipitation are provided in the files *precipitation\_awap.csv* and *precipitation\_silo.csv*, respectively (Table  
313 2). Users interested in a comparison of AWAP and SILO precipitation are referred to Tozer et al. (2012) who note that the  
314 two products vary due to differences in interpolation methods, as noted above. They also assess the impact of adopting these  
315 gridded products on rainfall runoff modelling outcomes, which may be of interest to CAMELS-AUS users.

316  
317 One further rainfall-related timeseries file is *precipitation\_var\_awap.csv* which provides, for each day, the spatial variance  
318 due to differences between grid cell values within a given catchment. This analysis was conducted using the tool *AWAPer*  
319 (Peterson et al., 2020) and the outputs can be used to understand how representative areal averages are across a given  
320 catchment, and how this varies with time.

### 321 3.5.3 Evaporative demand

322 As noted, evaporation and evapotranspiration variables are provided by SILO only (Table 2). SILO provides PET estimates  
323 for the FAO56 short crop (FAO, 1956) and ASCE tall crop (ASCE, 2000) methodologies, in addition to three  
324 evapotranspiration formulations from Morton (1983), namely point potential, areal wet environment potential, and areal actual.  
325 Three additional evaporation products are also provided, namely Morton (1983) shallow lake, interpolated Class A pan  
326 evaporation (which only covers the measured period, 1970 onwards), and synthetic Class A pan evaporation extended to the  
327 full SILO period using the method of Rayner (2005). See Table 2 for adopted file names.



328  
 329

**Table 2: Hydrometeorological time series data supplied with CAMELS-AUS. All timesteps are daily. All non-streamflow data were processed as part of the CAMELS-AUS project to extract catchment averages from Australia-wide AWAP/SILO grids.**

| Category   | File name                          | Source data  | Description / comments   | Unit                            |
|--|------------------------------------|--|--|---------------------------------|
| streamflow   | <i>streamflow_MLd.csv</i>          | Hydrologic Reference Stations (HRS) project, Bureau of Meteorology (BOM)<br><a href="http://www.bom.gov.au/water/hrs">www.bom.gov.au/water/hrs</a>   | Streamflow (not gap filled)  | ML d <sup>-1</sup>              |
|  | <i>streamflow_MLd_infilled.csv</i> |  | Streamflow gap filled by the BOM using GR4J (Perrin et al, 2003)   | ML d <sup>-1</sup>              |
|  | <i>streamflow_mmd.csv</i>          |  | Streamflow (not gap filled) expressed as depths relative to CAMELS-AUS adopted catchment areas (Table 1).  | mm d <sup>-1</sup>              |
|  | <i>streamflow_QualityCodes.csv</i> |  | Quality codes as supplied by the HRS website, with meanings listed at <a href="http://www.bom.gov.au/water/hrs/qc_doc.shtml">www.bom.gov.au/water/hrs/qc_doc.shtml</a> | -                               |
| precipitation  | <i>precipitation_awap.csv</i>      | BOM's Australian Water Availability Project (AWAP), (Jones et al., 2009)<br><a href="http://www.bom.gov.au/climate/maps/AWAP">www.bom.gov.au/climate/maps/AWAP</a> provides 0.05° grids.                           | catchment average precipitation  | mm d <sup>-1</sup>              |
|  | <i>precipitation_var_awap.csv</i>  |  | Spatial internal variance in precipitation as calculated by the 'AWAPer' tool (Peterson et al. 2020).  | mm <sup>2</sup> d <sup>-2</sup> |
|  | <i>precipitation_silo.csv</i>      |  | catchment average precipitation  | mm d <sup>-1</sup>              |
| Actual and potential evapo-transpiration (AET and PET) | <i>et_short_crop_silo.csv</i>      | Scientific Information for Land Owners (SILO) project, Government of Queensland (Jeffrey et al., 2001)<br><a href="http://www.longpaddock.qld.gov.au">www.longpaddock.qld.gov.au</a><br>SILO provides 0.05° grids. | FAO56 short crop PET (see FAO, 1998)   |                                 |
|  | <i>et_tall_crop_silo.csv</i>       |  | ASCE tall crop PET (see ASCE, 2000)  |                                 |
|  | <i>et_morton_wet_silo.csv</i>      |  | Morton (1983) wet-environment areal PET over land  |                                 |
|  | <i>et_morton_point_silo.csv</i>    |  | Morton (1983) point PET  |                                 |
|  | <i>et_morton_actual_silo.csv</i>   |  | Morton (1983) areal AET  |                                 |
| evaporation  | <i>evap_morton_lake_silo.csv</i>   |  | Morton (1983) shallow lake evaporation   |                                 |
|  | <i>evap_pan_silo.csv</i>           |  | Interpolated Class A pan evaporation   |                                 |
|  | <i>evap_syn_silo.csv</i>           |  | Interpolated synthetic extended Class A pan evaporation (Rayner, 2005)   |                                 |
| temperature  | <i>tmax_awap.csv</i>               | AWAP (see above)   | Daily maximum temperature  | °C                              |
|  | <i>tmax_silo.csv</i>               | SILO (see above)   |  |                                 |
|  | <i>tmin_awap.csv</i>               | AWAP (see above)   | Daily minimum temperature  |                                 |
|  | <i>tmin_silo.csv</i>               | SILO (see above)   |  |                                 |
| Other variables  | <i>solarrad_awap.csv</i>           | AWAP (see above)   | Solar radiation  | MJ m <sup>-2</sup>              |
|  | <i>radiation_silo.csv</i>          | SILO (see above)   |  |                                 |
|  | <i>vprp_awap.csv</i>               | AWAP (see above)   | Vapour pressure  | hPa                             |
|  | <i>vp_silo.csv</i>                 | SILO (see above)   |  |                                 |
|  | <i>vp_deficit_silo.csv</i>         |  | Vapour pressure deficit  | hPa                             |
|  | <i>rh_tmax_silo.csv</i>            |  | Relative humidity at the time of maximum temperature   | %                               |
|  | <i>rh_tmin_silo.csv</i>            |  | Relative humidity at the time of minimum temperature   | %                               |
| <i>mslp_silo.csv</i>                                   | Mean sea level pressure            |  | hPa  |                                 |



330

331

**Table 3: Flow uncertainty information, climatic indices and streamflow signatures provided in the attribute table of CAMELS-AUS**

| Short Name                | Description   | Units              | Data source / notes  |   |
|---------------------------|---|--------------------|--|---|
| <i>q_uncert_NumCurves</i> | Flow uncertainty: number of rating curves considered in analysis by McMahon and Peel (2019), and total number ( <i>Q_uncert_N</i> ) of days the curves apply to | -                  | McMahon and Peel (2019)  |   |
| <i>q_uncert_N</i>         |   |                    |  |   |
| <i>q_uncert_q10</i>       | Q10 (ie. flow exceeded 90% of the time) flow value with 95% confidence limits. Note, only calculated considering days for which rating curves are available.    | mm d <sup>-1</sup> |  |   |
| <i>q_uncert_q10_upper</i> |   | %                  |  |   |
| <i>q_uncert_q10_lower</i> |   | %                  |  |   |
| <i>q_uncert_q50</i>       | As above but for the median flow  | mm d <sup>-1</sup> |  |   |
| <i>q_uncert_q50_upper</i> |   | %                  |  |   |
| <i>q_uncert_q50_lower</i> |   | %                  |  |   |
| <i>q_uncert_q90</i>       | As above but for Q90 (flow exceeded 10% of the time)  | mm d <sup>-1</sup> |  |   |
| <i>q_uncert_q90_upper</i> |   | %                  |  |   |
| <i>q_uncert_q90_lower</i> |   | %                  |  |   |
| <i>p_mean</i>             | mean daily precipitation  | mm d <sup>-1</sup> | Climatic signatures are calculated using code from Addor et al. (2017), using the following datasets (cf. Table 1)<br>- Precipitation is based on AWAP rainfall.<br>- PET is based on SILO Morton Wet Env. PET<br>- temperature data is based on AWAP temperature<br><br>For <i>p_seasonality</i> see Eq. 14 in Woods (2009) |   |
| <i>pet_mean</i>           | mean daily potential evapotranspiration (PET) (Morton's Wet Environment)  | mm d <sup>-1</sup> |  |   |
| <i>aridity</i>            | aridity ( <i>pet_mean</i> / <i>p_mean</i> )   | -                  |  |   |
| <i>p_seasonality</i>      | precipitation seasonality (0: uniform; +ve: Dec/Jan peak; -ve: Jun/Jul peak)  | -                  |  |   |
| <i>frac_snow</i>          | fraction of precipitation on days colder than 0° C  | -                  |  |   |
| <i>high_prec_freq</i>     | frequency of high precipitation days, ≥5 times <i>p_mean</i>  | d y <sup>-1</sup>  |  |   |
| <i>high_prec_dur</i>      | average duration of high precipitation events   | days               |  |   |
| <i>high_prec_timing</i>   | season during which most high precip. days occur (djf, mam, jja, or son)  | season             |  |   |
| <i>low_prec_freq</i>      | frequency of dry days (≤ 1 mm/d)  | d y <sup>-1</sup>  |  |   |
| <i>low_prec_dur</i>       | average duration of low precipitation periods (days ≤ 1 mm/d)   | days               |  |   |
| <i>low_prec_timing</i>    | season during which most dry days occur (djf, mam, jja, or son)   | season             |  |   |
| <i>q_mean</i>             | mean daily streamflow   | mm d <sup>-1</sup> |  | Hydrologic signatures are calculated using code from Addor et al. (2017). Where required, climate datasets are the same as above. |
| <i>runoff_ratio</i>       | ratio of mean daily streamflow to mean daily precipitation  | -                  |  |   |
| <i>stream_elas</i>        | sensitivity of annual streamflow to annual rainfall changes   | -                  |  |   |
| <i>slope_fdc</i>          | slope of flow duration curve (log transformed) from percentiles 33 to 66  | -                  |  |   |
| <i>baseflow_index</i>     | baseflow as a proportion of total streamflow, calculated by recursive filter  | -                  | Original sources of signature formulations:<br>- <i>stream_elas</i> - Sankarasubramanian et al. (2001);<br>- <i>slope_fdc</i> - Sawicz et al. (2011);<br>- <i>baseflow_index</i> - Ladson et al. (2013); and<br>- <i>hdf_mean</i> - Court (1962).  |   |
| <i>hdf_mean</i>           | mean half flow date (date marking the passage of half the year's flow). Calculated according to April-March water years.  | day of year        |  |   |
| <i>Q5</i>                 | 5% flow quantile (low flow – flow exceeded 95% of the time)   | mm d <sup>-1</sup> |  |   |
| <i>Q95</i>                | 95% flow quantile (high flow – flow exceeded 5% of the time)  | mm d <sup>-1</sup> |  |   |
| <i>high_q_freq</i>        | frequency of high flow days (≥9 times mean daily flow)  | d y <sup>-1</sup>  |  |   |
| <i>high_q_dur</i>         | average duration of high flow events  | days               |  |   |
| <i>low_q_freq</i>         | frequency of low flow days (< 0.2 times mean daily flow)  | d y <sup>-1</sup>  |  |   |
| <i>low_q_dur</i>          | average duration of low flow periods  | days               |  |   |
| <i>zero_q_freq</i>        | frequency of days with Q = 0  | d y <sup>-1</sup>  |  |   |



### 332 3.5.4 Other timeseries

333 AWAP timeseries are provided for a further four variables: daily maximum temperature, daily minimum temperature, vapour  
334 pressure (1950 onwards), and solar radiation (1990 onwards). Solar radiation AWAP data has numerous gaps which have  
335 been filled by the average Julian Day value: for example, if the 5<sup>th</sup> March is missing, we adopt the average value over all non-  
336 missing instances of the 5<sup>th</sup> March. SILO timeseries are provided for the following variables: daily maximum temperature,  
337 daily minimum temperature, vapour pressure, vapour pressure deficit, solar radiation, mean sea level pressure (1957 onwards),  
338 relative humidity at time of maximum temperature and relative humidity at time of minimum temperature. See Table 2 for  
339 adopted file names.

### 340 3.6 Catchment attributes

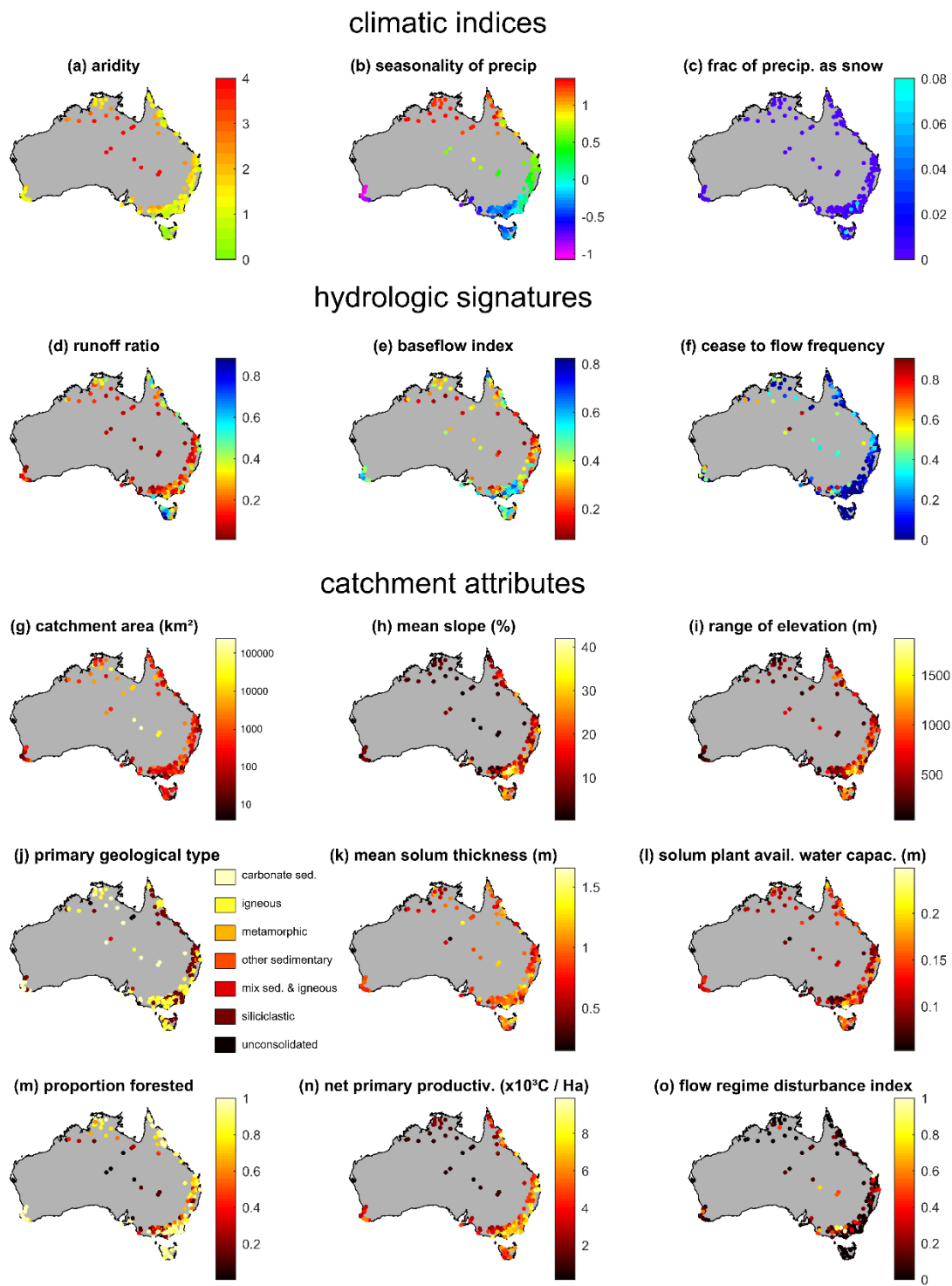
341 The following subsections, along with Tables 3 and 4, summarise the set of CAMELS-AUS catchment attributes. Spatial  
342 distributions of selected attributes are mapped in Figure 3.

343  
344 We note that the CAMELS-AUS dataset owes much to the earlier work of Stein et al. (2011), whose *National Environmental*  
345 *Stream Attributes* project calculated a broad variety of catchment attributes spatially across Australia, 74 of which are included  
346 in CAMELS-AUS dataset. Stein et al. (2011) calculated these for the upstream area of each stream segment in Australia based  
347 on a 250k scale stream and catchment dataset (the BOM Geospatial Fabric v2.1, [www.bom.gov.au/water/geofabric/](http://www.bom.gov.au/water/geofabric/)), and the  
348 contribution of the CAMELS-AUS project for the 74 indices is limited to (i) spatially matching each outlet to the appropriate  
349 segment (of which there are 1.4 million to choose from); and (ii) sorting through the attributes to identify those relevant to  
350 CAMELS-AUS (eg., not all Stein et al. 2011 attributes relate to the upstream catchment area; others relate to the local area  
351 immediately around the stream segment and are thus irrelevant as CAMELS-AUS attributes in nearly all cases).

### 352 3.7 Climatic indices and streamflow signatures

353 Eleven climatic indices are provided, as listed in Table 3, calculated using the same code used in the original CAMELS (Addor  
354 et al., 2017). The code requires input timeseries of precipitation, temperature and PET, and for this purpose AWAP was used  
355 where available (precipitation, temperature) and for PET, SILO Morton Areal Wet Environment PET was used (this  
356 combination of inputs is consistent with past modelling studies such as Fowler 2016; 2018; 2020b). Likewise, thirteen  
357 streamflow signature indices are provided, as listed in Table 3, also calculated using code from Addor et al. (2017). Together,  
358 the climatic and streamflow indices cover a wide range of statistics commonly used to characterise hydroclimate in modelling  
359 and regionalisation studies, and their common formulation with Addor et al. (2017) aids intercontinental comparison.

360



361

362

363

**Figure 3: Maps of selected climatic indices (a – c), hydrologic signatures (d – f) and other catchment attributes (g – o). For definitions, see Tables 3 and 4.**





### 364 3.7.1 Geology and soils

365 Geology data are taken from Stein et al. (2011) which in turn is from the 1:1,000,000 scale *Surface Geology of Australia*.  
366 In Table 4 this dataset is cited for brevity as *Geoscience Australia (2008)* but here we acknowledge the detailed state-by-state  
367 work of Liu et al. (2006), Raymond et al. (2007a, 2007b, 2007c), Stewart et al. (2008), and Whitaker et al. (2007, 2008). For  
368 each catchment the proportion taken up by each of the seven geological types is provided as separate attributes. Additionally,  
369 we follow Alvarez-Garreton et al. (2018) in defining separate categorical attributes for the primary and secondary geological  
370 units (see Figure 3j for a map of the primary types) with their respective areas defined as separate numerical attributes.

371  
372 Soils data are taken from a variety of sources. The soil depth attribute (*SolumThickness*) is based on the Atlas of Australian  
373 Soils (Isbell, 2002), which divides Australia into soil ‘map units’, each with associated ‘principle profile forms’ (ppfs) in order  
374 of dominance. In turn, the dataset provides estimates (McKenzie et al., 2000) of the distribution of solum thicknesses (as 5<sup>th</sup>,  
375 50<sup>th</sup> and 95<sup>th</sup> percentiles) associated with each ppf. The CAMELS-AUS *SolumThickness* is defined as a spatial average across  
376 the map units that occur in the catchment, where the depth assumed for a given map unit is the median value for its dominant  
377 ppf. Soil saturated hydraulic conductivity (*ksat*) and water holding capacity (*solpawhc*) are taken from Stein et al. (2011)  
378 which in turn is from Soil Hydrologic Properties of Australia (Western and McKenzie, 2004).

### 379 3.7.2 Topography and geometry

380 Maximum elevation and average elevation are each taken from Stein et al. (2011), but because the gauging stations themselves  
381 are not features in the Stein et al. dataset, we calculate the elevation at the outlet separately. Catchment slope is calculated as  
382 the spatial average of the slope product of Gallant et al. (2012), which is itself based on the 1 second SRTM DEM.

383  
384 Stein et al. (2011) provide a variety of attributes related to the geometry of the catchment and/or stream network. Each of  
385 these are based on the geometry of the streams and catchments defined in the BOM’s Geospatial Fabric v2.1  
386 ([www.bom.gov.au/water/geofabric/download.shtml](http://www.bom.gov.au/water/geofabric/download.shtml)), which itself is based on the 9 second (approximately 270 m) DEM of  
387 Hutchinson et al. (2008). The attributes are: (i) maximum flow path length *upsdist* upstream from the outlet; (ii) stream density;  
388 (iii) Strahler (1957) stream order at outlet; (iv) elongation ratio; (v) relief, here defined as ratio of the mean and maximum  
389 elevations above the outlet; and (vi) relief ratio, here defined as elevation range divided by flow path distance.

390  
391 Further attributes are defined based on the *Multi-Resolution Valley Bottom Flatness* (MRVBF) index of Gallant et al. (2012).  
392 As the name indicates, the index relates to the shape of the landscape and the degree of deposited sediment. As explained in  
393 Table 4, the index values contrast erosional (MRVBF=0) locations with depositional (MRVBF>0) locations ranging from  
394 ‘small hillside deposits’ (MRVBF=1) through to ‘extensive depositional basins’ (MRVBF=9). Ten separate attributes are  
395 defined based on each integer value (0, 1 ... 9) that MRVBF can take, indicating the proportion of the catchment in the given



396 class. Lastly, using an earlier MRVBF version, Stein et al. (2011) analysed how common it is for a stream to pass through  
397 erosional landscapes (MRVBF=0), and defined this as an additional attribute, ‘confinement’.

### 398 **3.7.3 Land cover and vegetation**

399 Land cover and vegetation attributes are primarily based on the Dynamic Land Cover Dataset (DLCD), v2 of Lymburner et  
400 al. (2015). Across Australia, the DLCD maps 22 land cover classes using MODIS satellite data over rolling two year windows,  
401 providing 13 separate time-slices (Jan. 02 – Dec. 2003, Jan. 03 – Dec. 04 ... Jan. 2014 – Dec. 2015). The CAMELS-AUS  
402 dataset incorporates this data in three ways:

- 403 1. A separate attribute for each land cover class, where the attribute value indicates the temporal average proportion of  
404 the catchment taken up by the class over the 13 time-slices;
- 405 2. Since ‘proportion forested’ is an oft-used catchment attribute, a separate attribute is defined as the sum of the four  
406 DLCD classes which mention trees (‘trees – closed’, ‘trees – open’, ‘trees – scattered’ and ‘trees – sparse’); and
- 407 3. The timeseries data itself is provided in full for each catchment, in a separate spreadsheet *Landcover\_timeseries.xlsx*

408  
409 The DLCD dataset is complemented by data from Stein et al. (2011), in turn sourced from the National Vegetation Information  
410 System (NVIS; DEWR, 2006). Stein et al. (2011) report the proportion of the catchment occupied by NVIS “major vegetation  
411 subgroups” (categories are grasses, forests, shrubs, woodlands and bare). This has considerable overlap with the DLCD, and  
412 the reason it is included is because the NVIS also estimates the proportion of these vegetation types that existed in the  
413 catchment’s ‘natural’ state (pre-1750; note this is pre-European but not pre-Indigenous settlement). For each of the 5  
414 categories, the NVIS provides natural pre-1750 (‘\_n’) and ‘extant’ (meaning current, ‘\_e’) statistics.

### 415 **3.7.4 Anthropogenic influences**

416 Anthropogenic influences are taken from Stein et al. (2011) based on earlier work with the same lead author (Stein et al.,  
417 2002). The earlier study aimed to identify the ‘wild’ rivers of Australia by quantifying human impacts on two broad categories:  
418 the flow regime (sub categories: impoundments, flow diversions and levee banks) and the catchment (sub categories:  
419 infrastructure, settlements, extractive industries and landuse). Following the same method, Stein et al. (2011) provide a unitless  
420 index varying between zero and one to quantify human effects in each of these categories and subcategories, and these are all  
421 included in CAMELS-AUS.

422  
423 In addition to the Stein et al. (2002) indices, one further attribute from the Stein et al. (2011) dataset is included in CAMELS-  
424 AUS: the length of river upstream before encountering a dam. Although most of the current catchments lack large dams (and  
425 thus this will be the same as *upsdist*, see Section 3.7.2), it is possible that future releases may include catchments that are  
426 marginally regulated and the index might be relevant in these cases.

427



428  
429

**Table 4: Catchment attributes included in the attributes table of CAMELS-AUS (apart from climatic and hydrologic indices)**

|                           | Short name  | Description   | Unit                        | Data source                         | Notes/references   |  |                                     |
|---------------------------|---|---|-----------------------------|-------------------------------------|--|--|-------------------------------------|
| Geology and Soils         | <i>geol_prim</i>  | Two most common geologies (see list in cell below) with corresponding proportions.  | -                           | Geoscience Australia (2008)         | Preprocessed by Stein et al. (2011)  |  |                                     |
|                           | <i>geol_prim_prop</i>   |   |                             |                                     |  |  |                                     |
|                           | <i>geol_sec</i>   |   |                             |                                     |  |  |                                     |
|                           | <i>geol_sec_prop</i>  |   |                             |                                     |  |  |                                     |
|                           | <i>unconsoldted</i>   | Proportion of catchment taken up by individual geological types, specifically: unconsolidated rocks; igneous rocks, siliciclastic/undifferentiated sedimentary rocks; carbonate sedimentary rocks; other sedimentary rocks; metamorphic rocks; and mixed sedimentary/igneous rocks.   | -                           |                                     |  |  |                                     |
|                           | <i>igneous</i>  |   |                             |                                     |  |  |                                     |
|                           | <i>silicised</i>  |   |                             |                                     |  |  |                                     |
|                           | <i>carbatesed</i>   |   |                             |                                     |  |  |                                     |
|                           | <i>othersed</i>   |   |                             |                                     |  |  |                                     |
|                           | <i>metamorph</i>  |   |                             |                                     |  |  |                                     |
|                           | <i>sedvolc</i>  |   |                             |                                     |  |  |                                     |
|                           | <i>oldrock</i>  | Catchment proportion old bedrock  | -                           |                                     |  |  |                                     |
|                           | <i>claya</i>  | Percent clay in the soil A & B horizons, for the stream valley in the reach containing gauging station.   | %                           |                                     |  | National Land and Water Resources Audit (2001) | Preprocessed by Stein et al. (2011) |
|                           | <i>clayb</i>  |   |                             |                                     |  |  |                                     |
| <i>sanda</i>              | As above, but % sand in the soil A horizon  | %   |                             |                                     |  |  |                                     |
| <i>solum_thickness</i>    | Mean soil depth considering all principle profile forms   | m   | McKenzie et al. (2000)      | -                                   |  |  |                                     |
| <i>ksat</i>               | Saturated hydraulic conductivity (areal mean)   | mm h <sup>-1</sup>  | Western and McKenzie (2004) | Preprocessed by Stein et al. (2011) |  |  |                                     |
| <i>solpawhc</i>           | Solum plant available water holding capacity (areal mean)   | mm  |                             |                                     |  |  |                                     |
| Topography and geometry   | <i>elev_min</i>   | Elevation above sea level at gauging station  | m                           | Gallant et al. (2009)               | -  |  |                                     |
|                           | <i>elev_max</i>   | Catchment maximum and mean elevation above sea level  | m                           | Hutchinson et al. (2008)            | Preprocessed by Stein et al. (2011)  |  |                                     |
|                           | <i>elev_mean</i>  |   |                             |                                     |  |  |                                     |
|                           | <i>elev_range</i>   | Range of elevation within catchment: <i>elev_max</i> - <i>elev_min</i>  | m                           |                                     | -  |  |                                     |
|                           | <i>mean_slope_pct</i>   | Mean slope, calculated on a grid-cell-by-grid-cell basis  | %                           | Gallant et al. (2012)               | -  |  |                                     |
|                           | <i>upsdist</i>  | Maximum flow path length upstream   | km                          | Hutchinson et al. (2008)            | Preprocessed by Stein et al. (2011). For <i>strahler</i> , see Strahler (1957) For <i>elongratio</i> , see Gordon et al. (1992).                   |  |                                     |
|                           | <i>strdensity</i>   | Ratio: (total length of streams) / (catchment area)   | km <sup>-1</sup>            |                                     |  |  |                                     |
|                           | <i>strahler</i>   | Strahler stream order at gauging station  | -                           |                                     |  |  |                                     |
|                           | <i>elongratio</i>   | Factor of elongation as defined in Gordon et al. (1992)   | -                           |                                     |  |  |                                     |
|                           | <i>relief</i>   | Ratio: (mean elev. above outlet)/(max elev. above outlet)   | -                           |                                     |  |  |                                     |
|                           | <i>reliefratio</i>  | Ratio: (elevation range)/(flow path distance)   | -                           |                                     |  |  |                                     |
|                           | <i>mrvbf_prop_0 through to mrvbf_prop_9</i>   | Proportion of catchment occupied by classes of Multi-Resolution Valley Bottom Flatness (MRVBF). These indicate areas subject to deposition. Broad interpretations are: 0 – erosional; 1 – small hillside deposit; 2-3 – narrow valley floor; 4 – valley floor; 5-6 – extensive valley floor; 7-8 – depositional basin; 9 – extensive depositional basin   | -                           | CSIRO (2016)                        | Gallant and Dowling (2003)   |  |                                     |
| <i>confinement</i>        | Proportion of stream segment cells & neighbouring cells that are not valley bottoms (as defined by MRVBF) | -   | Hutchinson et al. (2008)    | Preprocessed by Stein et al. (2011) |  |  |                                     |
| Land Cover and Vegetation | <i>lc01_extracti</i>  | Proportion of catchment occupied by land cover categories within the <i>Dynamic Land Cover Dataset</i> (DLCD): <i>mines and quarries</i> (ISO name: <i>extraction sites</i> ) <i>lakes and dams</i> ( <i>inland water bodies</i> ) <i>salt lakes</i> ( <i>salt lakes</i> ) <i>irrigated cropping</i> ( <i>irrigated cropping</i> ) <i>irrigated pasture</i> ( <i>irrigated pasture</i> ) <i>irrigated sugar</i> ( <i>irrigated sugar</i> ) <i>rain fed cropping</i> ( <i>rainfed cropping</i> ) <i>rain fed pasture</i> ( <i>rainfed pasture</i> ) <i>rain fed sugar</i> ( <i>rainfed sugar</i> ) <i>wetlands</i> ( <i>wetlands</i> ) <i>closed tussock grassland</i> ( <i>tussock grasses - closed</i> ) | -                           | Lymburner et al. (2015)             | Note, the source dataset has 13 timeslices; these attributes indicate the temporal average. The timeslices are separately supplied with CAMELS-AUS |  |                                     |
|                           | <i>lc03_waterbo</i>   |   |                             |                                     |  |  |                                     |
|                           | <i>lc04_saltilak</i>  |   |                             |                                     |  |  |                                     |
|                           | <i>lc05_irrcrop</i>   |   |                             |                                     |  |  |                                     |
|                           | <i>lc06_irrpast</i>   |   |                             |                                     |  |  |                                     |
|                           | <i>lc07_irrsuga</i>   |   |                             |                                     |  |  |                                     |
|                           | <i>lc08_rfcropp</i>   |   |                             |                                     |  |  |                                     |
|                           | <i>lc09_rfpastu</i>   |   |                             |                                     |  |  |                                     |
|                           | <i>lc10_rfsugar</i>   |   |                             |                                     |  |  |                                     |
|                           | <i>lc11_wetlands</i>  |   |                             |                                     |  |  |                                     |
|                           | <i>lc14_tussclo</i>   |   |                             |                                     |  |  |                                     |
|                           | <i>lc15_alpineg</i>   |   |                             |                                     |  |  |                                     |



|                          | Short name   | Description   | Unit                                   | Data source   | Notes/references                    |
|--------------------------|--|---|--|---|-------------------------------------|
|                          | <i>lc16_openhum</i>  | <i>alpine meadows (alpine grasses - open)</i>   |  |   |                                     |
|                          | <i>lc18_opentus</i>  | <i>open hummock grassland (hummock grasses - open)</i>  |  |   |                                     |
|                          | <i>lc19_shrbsca</i>  | <i>open tussock grasslands (tussock grasses - open)</i>   |  |   |                                     |
|                          | <i>lc24_shrbden</i>  | <i>scattered shrubs and grasses (shrubs and grasses - sparse - scattered)</i>   |  |   |                                     |
|                          | <i>lc25_shrbope</i>  | <i>dense shrubland (shrubs - closed)</i>  |  |   |                                     |
|                          | <i>lc31_forclos</i>  | <i>open shrubland (shrubs - open)</i>   |  |   |                                     |
|                          | <i>lc32_foropen</i>  | <i>closed forest (trees - closed)</i>   |  |   |                                     |
|                          | <i>lc33_woodope</i>  | <i>open forest (trees - open)</i>   |  |   |                                     |
|                          | <i>lc34_woodspa</i>  | <i>open woodland (trees - scattered)</i>  |  |   |                                     |
|                          | <i>lc35_urbanar</i>  | <i>woodland (trees - sparse)</i>  |  |   |                                     |
|                          | <i>lc35_urbanar</i>  | <i>urban areas (urban areas)</i>  |  |   |                                     |
|                          | <i>prop_forested</i>   | sum(LC_31, LC_32, LC_33, LC_34)   |  |   |                                     |
|                          | <i>nv_grasses_n</i>  | Major vegetation sub-groups within the <i>National Vegetation Information System</i> (NVIS). Despite redundancy with the DLCD attributes (see above), these are included because NVIS quantifies alteration from ‘natural’ by differentiating between ‘pre-1750’ (‘_n’) and ‘extant’ (‘_e’). Subgroups:   | -                                      | DEWR (2006)   | Preprocessed by Stein et al. (2011) |
|                          | <i>nv_grasses_e</i>  |   |  |   |                                     |
|                          | <i>nv_forests_n</i>  |   |  |   |                                     |
|                          | <i>nv_forests_e</i>  |   |  |   |                                     |
|                          | <i>nv_shrubs_n</i>   |   |  |   |                                     |
|                          | <i>nv_shrubs_e</i>   |   |  |   |                                     |
|                          | <i>nv_woodl_n</i>  |   |  |   |                                     |
|                          | <i>nv_woodl_e</i>  |   |  |   |                                     |
|                          | <i>nv_bare_n</i>   |   |  |   |                                     |
|                          | <i>nv_bare_e</i>   |   |  |   |                                     |
|                          | <i>nv_nodata_n</i>   |   |  |   |                                     |
| <i>nv_nodata_e</i>       |  |   |  |   |                                     |
| Anthropogenic Influences | <i>distupdamw</i>  | maximum distance upstream before encountering a dam or water storage  | km                                     | Geoscience Australia (2004)                         | Preprocessed by Stein et al. (2011) |
|                          | <i>impound_fac</i>   | Dimensionless factors quantifying human impacts on catchment hydrology, in two broad categories:<br>- Flow regime factors: impoundments ( <i>ImpoundmF</i> ), flow diversions ( <i>FlowDivF</i> ), and levee banks ( <i>LeveebankF</i> ). The combined effect is disturbance index <i>FlowRegimeDI</i> ;<br>- Catchment factors: infrastructure ( <i>InfrastrucF</i> ), settlements ( <i>SettlementF</i> ), extractive industries ( <i>ExtractiveIndF</i> ) and landuse ( <i>LanduseF</i> ). The combined effect is captured in <i>CatchmentDI</i> .<br><i>FlowRegimeDI</i> and <i>CatchmentDI</i> are combined in <i>RiverDI</i> | -                                      | Stein et al. (2002), updated by Stein et al. (2011) |                                     |
|                          | <i>flow_div_fac</i>  |   |  |   |                                     |
|                          | <i>leveebank_fac</i>   |   |  |   |                                     |
|                          | <i>infrastruc_fac</i>  |   |  |   |                                     |
|                          | <i>settlement_fac</i>  |   |  |   |                                     |
|                          | <i>extract_inf_fac</i>   |   |  |   |                                     |
|                          | <i>landuse_fac</i>   |   |  |   |                                     |
|                          | <i>catchment_di</i>  |   |  |   |                                     |
| <i>flow_regime_di</i>    |  |   |  |   |                                     |
| <i>river_di</i>          |  |   |  |   |                                     |
| Other                    | <i>pop_mean</i>  | Average and maximum human population density in catchment across 3" grid squares.   | km <sup>2</sup>                        | ABS (2006)  | Preprocessed by Stein et al. (2011) |
|                          | <i>pop_max</i>   |   |  |   |                                     |
|                          | <i>pop_gt_1</i>  | Proportion of catchment with population density exceeding 1 person / km <sup>2</sup> and 10 people / km <sup>2</sup>  | -                                      |   |                                     |
|                          | <i>pop_gt_10</i>   |   |  |   |                                     |
|                          | <i>erosivity</i>   | Rainfall erosivity (spatial average across catchment)   | MJ mm ha <sup>-1</sup> h <sup>-1</sup> | NLWRA (2001)  |                                     |
|                          | <i>anngro_mega</i>   | Average annual growth index value for megatherm, mesotherm and microtherm plants, respectively  | -                                      | Xu and Hutchinson (2011)                            |                                     |
|                          | <i>anngro_meso</i>   |   |  |   |                                     |
|                          | <i>anngro_micro</i>  |   |  |   |                                     |
|                          | <i>gromega_seas</i>  | Seasonality of growth index value for megatherm, mesotherm and microtherm plants, respectively  | -                                      |   |                                     |
|                          | <i>gromeso_seas</i>  |   |  |   |                                     |
| <i>gromicro_seas</i>     |  |   |  |   |                                     |
| <i>npp_ann</i>           | Net Primary Productivity estimated by Raupach et al. (2002) for pre-European settlement conditions:<br>- annually; and<br>- for the twelve calendar months of the year | tC Ha <sup>-1</sup>   | Raupach et al. (2002)                  | Preprocessed by Stein et al. (2011)                 |                                     |
| <i>npp_1</i>             |  |   |  |   |                                     |
| <i>through to</i>        |  |   |  |   |                                     |
| <i>npp_12</i>            |  |   |  |   |                                     |



431

### 432 **3.7.5 Other catchment attributes**

433 This final category contains indices that do not easily fit in one category, or that fit into more than one. The attributes  
434 quantifying human population are included here as they are relevant to both the land cover category and the anthropogenic  
435 influences, but fit neatly into neither. These population attributes, taken from Stein et al. (2011), are based on aggregation of  
436 census population to 9 second grid squares, and quantify the spatial average, the maximum grid value present in the catchment,  
437 and the proportion of grid squares exceeding 1 and 10 people km<sup>-2</sup>. A further inclusion is the erosivity which is primarily a  
438 climatic attribute but is often used by studies associated with the soil category. The erosivity is taken from Stein et al. (2011)  
439 and in turn from the National Land and Water Resources Audit (NLWRA, 2001).

440

441 Finally, there are two further subcategories of attributes: growth indices of plants, and net primary productivity statistics. The  
442 growth indices of plants, compiled by Stein et al. (2011) and calculated using the Australian National University's ANUCLIM  
443 program (Xu and Hutchinson, 2011), quantify the suitability of growing conditions (and the seasonality thereof) for three types  
444 of plants: megatherm (plants living in relatively high temperatures year round), mesotherm (plants living in seasonally high  
445 temperatures) and microtherm (plants living in low temperatures). Net primary productivity (NPP) statistics are provided from  
446 Stein et al. (2011) based on Raupach et al. (2002). NPP is defined by Raupach et al. (2002) as "*plant photosynthesis less plant*  
447 *respiration ... the carbon or biomass yield of the landscape*" and "*the most important driver of the coupled balances of water,*  
448 *C, N and P*". Although Raupach et al. (2002) quantified both baseline (pre-agricultural) and current NPP, only the baseline  
449 figures were processed by Stein et al. (2011). The attributes include the annual average NPP in addition to averages for each  
450 calendar month separately.

## 451 **4 Data availability**

452 The CAMELS-AUS dataset is freely available for download from the Pangaea online repository at  
453 <https://doi.pangaea.de/10.1594/PANGAEA.921850> (Fowler et al., 2020a). The dataset can only be downloaded via Pangaea's  
454 'view dataset as html' option, not 'download dataset as tab-delimited text'. The dataset (along with datasets on which it is  
455 based) is subject to a Creative Commons BY (attribution) licence agreement (<https://creativecommons.org/licenses/>).

## 456 **5 Conclusions**

457 This paper introduced a new freely available dataset for Australia, CAMELS-AUS. It is the first large sample hydrology  
458 dataset for Australia and the fifth CAMELS dataset worldwide. CAMELS-AUS provides timeseries data (streamflow and 18  
459 climatic variables) and a broad set of 134 attributes, for 222 unregulated catchments from across Australia. Given the unique



460 hydroclimate of Australia, with high hydroclimatic variability and many case studies of multi-year drought, it is hoped that the  
461 release of this dataset will accelerate progress in such fields as arid zone hydrology and the study of hydrology under a changing  
462 climate.

## 463 **6 Author contribution**

464 KF and NA conceived the dataset with the support of MP. KF, NA and MP designed the dataset. KF, CC and SCA analysed  
465 and compiled the hydrometeorological timeseries and catchment attribute data. MP analysed earlier work (McMahon and  
466 Peel, 2019) to provide the uncertainty estimates included in the dataset. KF wrote the initial draft of the manuscript and all  
467 co-authors edited and amended it to provide the final manuscript.

## 468 **7 Competing interests**

469 The authors declare they have no conflict of interest.

## 470 **8 Acknowledgements**

471 The authors acknowledge the Bureau of Meteorology, Australia for their support and permission to undertake this project on  
472 the Hydrologic Reference Stations. The authors acknowledge the excellent work of Janet Stein and coauthors (Stein et al.,  
473 2011) to whom this dataset owes more than half the listed catchment attributes. KF acknowledges funding from the Australian  
474 Research Council (LP170100598) and the Bureau of Meteorology (TP705654) during the period of preparation of this  
475 manuscript. KF appreciates the support of Gemma Coxon (University of Bristol) who encouraged his foray into large sample  
476 hydrology. NA acknowledges support from the Swiss National Science Foundation (fellowship P400P2\_180791).

## 477 **9 References**

478 ABS (Australian Bureau of Statistics): Australian Census 2006 Population Statistics. Raw data available at  
479 <https://www.abs.gov.au/websitedbs/censushome.nsf/home/historicaldata2006?opendocument&navpos=280>, 2006

480 Acharya, S. C., Nathan, R., Wang, Q. J., Su, C.-H., and Eizenberg, N.: An evaluation of daily precipitation from a regional  
481 atmospheric reanalysis over Australia, *Hydrol. Earth Syst. Sci.*, 23, 3387–3403, <https://doi.org/10.5194/hess-23-3387-2019>,  
482 2019

483 Addor, N., Do, H. X., Alvarez-Garreton, C., Coxon, G., Fowler, K. and Mendoza, P. A.: Large sample hydrology: recent  
484 progress, guidelines for new datasets and grand challenges, *Hydrol. Sci. J.*, <https://doi.org/10.1080/02626667.2019.1683182>,  
485 2019.



- 486 Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N. and Clark, M. P.: A Ranking of Hydrological Signatures Based  
487 on Their Predictability in Space, *Water Resour. Res.*, 54(11), 8792– 8812, <https://doi.org/10.1029/2018WR022606>, 2018.
- 488 Addor, N., Newman, A. J., Mizukami, N. and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for  
489 large-sample studies, *Hydrol Earth Syst Sci*, 21(10), 5293–5313, 655 <https://doi.org/10.5194/hess-21-5293-2017>, 2017.
- 490 Aghakouchak, A., D. Feldman, M. Stewardson, J. Saphores, S. Grant, and B. Sanders: Australia's Drought: Lessons for  
491 California, *Science*, 343 (6178), 1430-1431, <https://doi.org/10.1126/science.343.6178.1430>, 2014
- 492 Alvarez-Garreton, C., Mendoza, P. A., Boisier, J. P., Addor, N., Galleguillos, M., ZambranoBigiarini, M., Lara, A., Puelma,  
493 C., Cortes, G., Garreaud, R., McPhee, J. and Ayala, A.: The CAMELS-CL dataset: catchment attributes and meteorology for  
494 large sample studies – Chile dataset, *Hydrol. Earth Syst. Sci.*, 22(11), 5817–5846, <https://doi.org/10.5194/hess-22-5817-2018>  
495 , 2018.
- 496 Andréassian, V., et al., 2006. Introduction and synthesis : why should hydrologists work on a large number of basin data sets  
497 ? Large sample basin experiments for hydrological model parameterization: results of the Model Parameter Experiment–  
498 MOPEX. Vol. 307. CEH Wallingford, UK: IAHS Publ, 1–5.
- 499 Arsenault, R., Bazile, R., Ouellet Dallaire, C. and Brissette, F.: CANOPEX: A Canadian hydrometeorological watershed  
500 database, *Hydrol. Process.*, 30(15), 2734–2736, <https://doi.org/10.1002/hyp.10880>, 2016.
- 501 ASCE (American Society for Civil Engineering): ASCE's Standardized Reference Evapotranspiration Equation, proceedings  
502 of the National Irrigation Symposium, Phoenix, Arizona, 2000
- 503 Ashcroft, L., Karoly, D. J., & Gergis, J.: Southeastern Australian climate variability 1860–2009: a multivariate analysis.  
504 *International Journal of Climatology*, 34(6), 1928-1944. <https://doi.org/10.1002/joc.3812>, 2014
- 505 BOM (Bureau of Meteorology, Australia): Hydrologic Reference Stations data update 2015. Website.  
506 [www.bom.gov.au/water/hrs/update\\_2015.shtml](http://www.bom.gov.au/water/hrs/update_2015.shtml), accessed June 1<sup>st</sup> 2020.
- 507 Chagas, V. B. P., Chaffe, P. L. B., Addor, N., Fan, F. M., Fleischmann, A. S., Paiva, R. C. D. and Siqueira, V. A.: CAMELS-  
508 BR: Hydrometeorological time series and landscape attributes for 897 catchments in Brazil, *Earth Syst. Sci. Data Discuss.*,  
509 <https://doi.org/10.5194/essd-2020-67>, 2020.
- 510 Court, A.: Measures of streamflow timing, *J. Geophys. Res.*, 67, 4335–4339, <https://doi.org/10.1029/JZ067i011p04335>, 1962.



- 511 Coxon, G., Addor, N., Bloomfield, J. P., Freer, J., Fry, M., Hannaford, J., Howden, N. J. K., Lane, R., Lewis, M., Robinson,  
512 E. L., Wagener, T., and Woods, R.: CAMELS-GB: Hydrometeorological time series and landscape attributes for 671  
513 catchments in Great Britain, *Earth Syst. Sci. Data Discuss.*, <https://doi.org/10.5194/essd-2020-49>, in review, 2020.
- 514 CSIRO: AUS SRTM 1sec MRVBF mosaic v01. Bioregional Assessment Source Dataset. Viewed 18 June 2018,  
515 <http://data.bioregionalassessments.gov.au/dataset/79975b4a-1204-4ab1-b02b-0c6fbbbbbcb5>, 2016
- 516 DEWR (Department of the Environment and Water Resources, Australia): Estimated Pre-1750 Major Vegetation Subgroups  
517 - NVIS Stage 1, Version 3.1. [https://www.environment.gov.au/land/native-vegetation/national-vegetation-information-](https://www.environment.gov.au/land/native-vegetation/national-vegetation-information-system)  
518 [system](https://www.environment.gov.au/land/native-vegetation/national-vegetation-information-system), 2008
- 519 Do, H. X., Gudmundsson, L., Leonard, M. and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM)-Part  
520 1: The production of a daily streamflow archive and metadata, *Earth Syst. Sci. Data*, 10(2), 765–785,  
521 <https://doi.org/10.5194/essd-10-765-2018>, 2018.
- 522 Falkenmark, M. and Chapman, T.: *Comparative hydrology: An ecological approach to land and water resources*. Unesco,  
523 Paris, 1989.
- 524 FAO (Food and Agriculture Organization of the United Nations): Irrigation and drainage paper 56: Crop evapotranspiration -  
525 Guidelines for computing crop water requirements, 1998.
- 526 Fowler, K., Peel, M., Western, A., & Zhang, L. Improved rainfall-runoff calibration for drying climate: Choice of objective  
527 function. *Water Resources Research*, 54. <https://doi.org/10.1029/2017WR022466>, 2018
- 528 Fowler, K., Peel, M., Western, A., Zhang, L., and Peterson, T. J.: Simulating runoff under changing climatic conditions:  
529 Revisiting an apparent deficiency of conceptual rainfall-runoff models, *Water Resources Research*, 52,  
530 <https://doi.org/10.1002/2015WR018068>, 2016
- 531 Fowler, K.; Acharya, S. C.; Addor, N.; Chou, C.; Peel, M.: CAMELS-AUS v1: Hydrometeorological time series and landscape  
532 attributes for 222 catchments in Australia. Dataset archived with online repository PANGAEA,  
533 <https://doi.pangaea.de/10.1594/PANGAEA.921850>, (dataset in review at time of article), 2020a
- 534 Fowler, K., Knoben, W., Peel, M., Peterson, T., Ryu, D., Saft, M., Seo, K., Western, A. Many commonly used rainfall-runoff  
535 models lack long, slow dynamics: implications for runoff projections. *Water Resources Research*,  
536 <https://doi.org/10.1029/2019WR025286>, 2020b
- 537 Gallant, J. C., and T. I. Dowling: A multiresolution index of valley bottom flatness for mapping depositional areas,  
538 *WaterResour. Res.*, 39(12), 1347, <https://doi.org/10.1029/2002WR001426>, 2003





- 539 Gallant, J., Wilson, N., Tickle, P.K., Dowling, T., Read, A.: 3 second SRTM Derived Digital Elevation Model (DEM) Version  
540 1.0. Record 1.0. Geoscience Australia, Canberra. <http://pid.geoscience.gov.au/dataset/ga/69888>, 2009.
- 541 Gallant, John; Austin, Jenet: Slope derived from 1" SRTM DEM-S. v4. CSIRO. Data Collection.  
542 <https://doi.org/10.4225/08/5689DA774564A>, 2012
- 543 Geoscience Australia: Dams and Water Storages 1990, Geoscience Australia, Canberra. Later versions at  
544 <https://data.gov.au/data/dataset/ce5b77bf-5a02-4cf8-9cf2-be4a2cee2677>, 2004
- 545 Geoscience Australia: Surface Geology of Australia 1:1 million scale dataset. Latest version is available at  
546 <https://data.gov.au/dataset/ds-dga-48fe9c9d-2f10-49d2-bd24-ac546662c4ec/details>, 2008
- 547 Ghiggi, G., Humphrey, V., Seneviratne, S. I. and Gudmundsson, L.: GRUN: An observations-based global gridded runoff  
548 dataset from 1902 to 2014, Earth Syst. Sci. Data Discuss., 1–32, <https://doi.org/10.5194/essd-2019-32>, 2019.
- 549 Gordon, N. D., McMahon, T. A., Finlayson, B. L., & Christopher, J.: Stream Hydrology: an Introduction for Ecologists. John  
550 Wiley & Sons, Ltd. 1992
- 551 Gupta, H.V., et al. Large-sample hydrology: a need to balance depth with breadth. Hydrology and Earth System Sciences, 18,  
552 463–477. <https://doi.org/10.5194/hess-18-463-2014>, 2014
- 553 Hutchinson, M.F., Stein, J.L., Stein, J.A., Anderson, H., Tickle, P.K.: GEODATA 9 second DEM and D8: Digital Elevation  
554 Model Version 3 and Flow Direction Grid 2008. Record DEM-9S.v3. Geoscience Australia, Canberra.  
555 <http://pid.geoscience.gov.au/dataset/ga/66006>, 2008
- 556 Isbell, R. F. The Australian Soil Classification. Revised Edition. CSIRO Publishing, Melbourne. See  
557 [https://www.asris.csiro.au/themes/Atlas.html#Atlas\\_Digital](https://www.asris.csiro.au/themes/Atlas.html#Atlas_Digital), 2002
- 558 Jeffrey, S. J., Carter, J. O., Moodie, K. B., & Beswick, A. R.: Using spatial interpolation to construct a comprehensive archive  
559 of Australian climate data. Environmental Modelling and Software, 16(4), 309–330. [https://doi.org/10.1016/S1364-8152\(01\)00008-1](https://doi.org/10.1016/S1364-8152(01)00008-1), 2001
- 561 Jones, D. A., Wang, W., & Fawcett, R.: High-quality spatial climate data-sets for Australia. Australian Meteorological and  
562 Oceanographic Journal, 58(4), 233–248. <https://doi.org/10.22499/2.5804.003>, 2009
- 563 Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S. and Nearing, G. S.: Toward Improved Predictions in  
564 Ungauged Basins: Exploiting the Power of Machine Learning, Water Resour. Res., 55(12), 11344–11354,  
565 <https://doi.org/10.1029/2019WR026065>, 2019.



- 566 Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S. and Nearing, G.: Benchmarking a Catchment-Aware Long  
567 Short-Term Memory Network (LSTM) for Large-Scale Hydrological Modeling, *Hydrol. Earth Syst. Sci. Discuss.*,  
568 <https://doi.org/10.5194/hess-2019-368>, 2019.
- 569 Kuentz, A., Arheimer, B., Hundecha, Y. and Wagener, T.: Understanding hydrologic variability across Europe through  
570 catchment classification, *Hydrol. Earth Syst. Sci.*, 21, 2863–2879, <https://doi.org/10.5194/hess-21-2863-2017>, 2017.
- 571 Ladson, A., Brown, R., Neal, B., and Nathan, R.: A standard approach to baseflow separation using the Lyne and Hollick filter,  
572 *Aust. J. Water Resour.*, 17, 25–34, <https://doi.org/10.7158/W12-028.2013.17.1>, 2013.
- 573 Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., David, C. H., Durand, M., Pavelsky, T. M., Allen, G. H.,  
574 Gleason, C. J. and Wood, E. F.: Global reconstruction of naturalized river flows at 2.94 million reaches, *Water Resour. Res.*,  
575 2019WR025287, <https://doi.org/10.1029/2019WR025287>, 2019.
- 576 Linke, S., Lehner, B., Dallaire, C. O., Ariwi, J., Grill, G., Anand, M., Beames, P., Burchard-levine, V., Moidu, H., Tan, F. and  
577 Thieme, M.: HydroATLAS : global hydro-environmental sub-basin and river reach characteristics at high spatial resolution,  
578 *Sci. Data*, 0–25, <https://doi.org/10.1038/s41597-019-0300-6>, 2019.
- 579 Liu, S. F., Raymond, O. L., Stewart, A. J., Sweet, I. P., Duggan, M., Charlick, C., Phillips, D. and Retter, A. J.: Surface geology  
580 of Australia 1:1,000,000 scale, Northern Territory [Digital Dataset]. The Commonwealth of Australia, Geoscience Australia.,  
581 Canberra Retrieved from: <http://www.ga.gov.au>, 2006
- 582 Lymburner, L., Tan, P., McIntyre, A., Thankappan, M., Sixsmith, J.: Dynamic Land Cover Dataset Version 2.1. Geoscience  
583 Australia, Canberra. <http://pid.geoscience.gov.au/dataset/ga/83868>, 2015
- 584 Mathevet, T., Gupta, H., Perrin, C., Andréassian, V., & Le Moine, N.: Assessing the performance and robustness of two  
585 conceptual rainfall-runoff models on a worldwide sample of watersheds. *Journal of Hydrology*,  
586 <https://doi.org/10.1016/j.jhydrol.2020.124698>, 2020.
- 587 McKenzie, N.J., Jacquier, D.W., Ashton L.J. and Cresswell, H.P.: Estimation of Soil Properties Using the Atlas of Australian  
588 Soils. CSIRO Land and Water Technical Report 11/00. see [https://www.asris.csiro.au/themes/Atlas.html#Atlas\\_Digital](https://www.asris.csiro.au/themes/Atlas.html#Atlas_Digital), 2000
- 589 McInerney, D., Thyer, M., Kavetski, D., Lerat, J., & Kuczera, G. (2017). Improving probabilistic prediction of daily  
590 streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors. *Water Resources Research*,  
591 53(3), 2199–2239. <https://doi.org/10.1002/2016WR019168>. 2017
- 592 McMahon, T.A., Finlayson, B.L., Haines, A.T. and Srikanthan, R.: *Global runoff: continental comparisons of annual flows  
593 and peak discharges*. Catena Verlag, 1992.



- 594 McMahon T and Peel M.: Uncertainty in stage–discharge rating curves: application to Australian Hydrologic Reference  
595 Stations data, *Hydrological Sciences Journal*, 64:3, 255-275, <https://doi.org/10.1080/02626667.2019.1577555>, 2019
- 596 Morton, F. I.: Operational estimates of areal evapotranspiration and their significance to the science and practice of hydrology.  
597 *Journal of Hydrology*, 66, 1–76, 1983
- 598 National Land and Water Resources Audit: Gridded soil information layers. Canberra. View at  
599 [www.asris.csiro.au/mapping/viewer.htm](http://www.asris.csiro.au/mapping/viewer.htm), 2001
- 600 Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R., Blodgett, D., Brekke, L., Arnold, J. R.,  
601 Hopson, T. and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological dataset for the contiguous  
602 USA: dataset characteristics and assessment of regional variability in hydrologic model performance, *Hydrol. Earth Syst. Sci.*,  
603 19, 209–223, <https://www.hydrol-earth-syst-sci.net/19/209/2015/>, 2015.
- 604 Olarinoye, T., Gleeson, T., Marx, V., Seeger, S., Adinehvand, R., et al. Global karst springs hydrograph dataset for research  
605 and management of the world’s fastest-flowing groundwater, *Sci. Data*, 7(1), 1–9, [https://doi.org/10.1038/s41597-019-0346-](https://doi.org/10.1038/s41597-019-0346-5)  
606 5, 2020.
- 607 Peel, M.C., McMahon, T.A. and Finlayson, B.L.: Variability of annual precipitation and its relationship to the El Niño–  
608 Southern Oscillation. *Journal of Climate*, 15(5), 545-551, 2002.
- 609 Peel, M.C., McMahon, T.A. and Finlayson, B.L.: Continental differences in the variability of annual runoff-update and  
610 reassessment. *Journal of Hydrology*, 295(1-4), 185-197, 2004.
- 611 Peel, M. C., G. G. S. Pegram, and T. A. McMahon: Global analysis of runs of annual precipitation and runoff equal to or below  
612 the median: Run magnitude and severity, *International Journal of Climatology*, 24 (7), 549-568,  
613 <https://doi.org/10.1002/joc.1147>, 2005
- 614 Peterson, T. J., Wasko, C., Saft, M., & Peel, M. C.: AWAPer: An R package for area weighted catchment daily meteorological  
615 data anywhere within Australia. *Hydrological Processes*, 34(5), 1301-1306. <https://doi.org/10.1002/hyp.13637>, 2020
- 616 Pool, S., Viviroli, D., & Seibert, J. Value of a limited number of discharge observations for improving regionalization: A  
617 large-sample study across the United States. *Water Resources Research*, 55, 363–377.  
618 <https://doi.org/10.1029/2018WR023855>, 2019
- 619 Raupach MR, Kirby JM, Barrett DJ, and Briggs PR: Balances of Water, Carbon, Nitrogen and Phosphorus in Australian  
620 Landscapes version 2.04, CSIRO Land and Water, Canberra, <http://www.clw.csiro.au/publications/technical2001/tr40-01.pdf>,  
621 2002



- 622 Raymond, O. L., Liu, S. F. and Kilgour, P.: Surface geology of Australia 1:1,000,000 scale, Tasmania - 3rd edition [Digital  
623 Dataset]. The Commonwealth of Australia, Geoscience Australia., Canberra Retrieved from: <http://www.ga.gov.au>, 2007a
- 624 Raymond, O. L., Liu, S. F., Kilgour, P. L., Retter, A. J., Stewart, A. J. and Stewart, G.: Surface geology of Australia 1:1,000,000  
625 scale, New South Wales - 2nd edition [Digital Dataset]. The Commonwealth of Australia, Geoscience Australia., Canberra  
626 Retrieved from: <http://www.ga.gov.au>, 2007b
- 627 Raymond, O. L., Liu, S. F., Kilgour, P., Retter, A. J. and Connolly, D. P.: Surface geology of Australia 1:1,000,000 scale,  
628 Victoria - 3rd edition [Digital Dataset]. The Commonwealth of Australia, Geoscience Australia., Canberra Retrieved from:  
629 <http://www.ga.gov.au>, 2007c
- 630 Rayner, D. Australian synthetic daily Class A pan evaporation. Technical Report December, Queensland Department of  
631 Natural Resources and Mines, Indooroopilly, Qld., Australia, 40 pp.,  
632 <https://data.longpaddock.qld.gov.au/static/silo/pdf/AustralianSyntheticDailyClassAPanEvaporation.pdf>, 2005
- 633 Sankarasubramanian, A., Vogel, R. M., and Limbrunner, J. F.: Climate elasticity of streamflow in the United States, Water  
634 Resour. Res., 37, 1771–1781, <https://doi.org/10.1029/2000WR900330>, 2001.
- 635 Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G.: Catchment classification: empirical analysis of  
636 hydrologic similarity based on catchment function in the eastern USA, Hydrol. Earth Syst. Sci., 15, 2895–2911,  
637 <https://doi.org/10.5194/hess-15-2895-2011>, 2011.
- 638 Shen, C.: A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists, Water  
639 Resour. Res., 54, 8558–8593, <https://doi.org/10.1029/2018WR022643>, 2018.
- 640 Skinner, D., & Langford, J. Legislating for sustainable basin management: the story of Australia's Water Act (2007). Water  
641 Policy, 15(6), 871-894. 2013
- 642 Stein, J. L., Stein, J. A. and Nix, H. A.: Spatial analysis of anthropogenic river disturbance at regional and continental scales:  
643 identifying the wild rivers of Australia. Landscape and Urban Planning, 60, 1-25, [https://doi.org/10.1016/S0169-  
644 2046\(02\)00048-8](https://doi.org/10.1016/S0169-2046(02)00048-8), 2002
- 645 Stein, J. L., Hutchinson, M. F. and Stein, J. A.: National Catchment and Stream Environment Database version 1.1.4. Available  
646 at <http://pid.geoscience.gov.au/dataset/ga/73045>, 2011
- 647 Stewart, A. J., Sweet, I. P., Needham, R. S., Raymond, O. L., Whitaker, A. J., Liu, S. F., Phillips, D., Retter, A. J., Connolly,  
648 D. P. and Stewart, G.: Surface geology of Australia 1:1,000,000 scale, Western Australia [Digital Dataset]. The  
649 Commonwealth of Australia, Geoscience Australia., Canberra Retrieved from: <http://www.ga.gov.au>, 2008



- 650 Strahler, A. N. Quantitative analysis of watershed geomorphology. *Eos, Transactions American Geophysical Union*, 38(6),  
651 913-920. <https://doi.org/10.1029/TR038i006p00913>, 1957
- 652 Tozer, C. R., Kiem, A. S., & Verdon-Kidd, D. C. On the uncertainties associated with using gridded rainfall data as a proxy  
653 for observed. *Hydrology and Earth System Sciences*, 16(5), 1481. <https://doi.org/10.5194/hess-16-1481-2012>, 2012
- 654 Turner, M., Bari, M., Amirthanathan, G., & Ahmad, Z.: Australian network of hydrologic reference stations-advances in  
655 design, development and implementation. In *Hydrology and Water Resources Symposium 2012* (p. 1555). Engineers Australia,  
656 <http://www.bom.gov.au/water/hrs/media/static/papers/Turner2012.pdf>, 2012
- 657 Verdon-Kidd, D. C., and A. S. Kiem: Nature and causes of protracted droughts in southeast Australia: Comparison between  
658 the Federation, WWII, and Big Dry droughts, *Geophysical Research Letters*, 36 (22), 1-6,  
659 <https://doi.org/10.1029/2009GL041067>, 2009.
- 660 Vertessy, R. A.: Water information services for Australians. *Australasian Journal of Water Resources*, 16(2), 91-105.  
661 <https://www.tandfonline.com/doi/abs/10.7158/13241583.2013.11465407>, 2013
- 662 Viglione, A., Borga, M., Balabanis, P. and Blöschl, G.: Barriers to the exchange of hydrometeorological data in Europe: Results  
663 from a survey and implications for data policy. *Journal of Hydrology*, 394(1-2), 63-77, 2010.
- 664 Western, A. and McKenzie, N.: *Soil hydrological properties of Australia Version 1.0.1*, CRC for Catchment Hydrology,  
665 Melbourne, 2004
- 666 Western, A. W., Matic, V., & Peel, M. C.: Justin Costelloe: a champion of arid-zone water research. *Hydrogeology Journal*,  
667 28(1), 37-41, <https://doi.org/10.1007/s10040-019-02051-7>, 2020
- 668 Whitaker, A. J., Champion, D. C., Sweet, I. P., Kilgour, P. and Connolly, D. P.: Surface geology of Australia 1:1,000,000  
669 scale, Queensland 2nd edition [Digital Dataset]. The Commonwealth of Australia, Geoscience Australia., Canberra Retrieved  
670 from: <http://www.ga.gov.au>, 2007
- 671 Whitaker, A. J., Glanville, D. H., English, P. M., Stewart, A. J., Retter, A. J., Connolly, D. P., Stewart, G. A. and Fisher, C.  
672 L.: Surface geology of Australia 1:1,000,000 scale, South Australia [Digital Dataset]. The Commonwealth of Australia,  
673 Geoscience Australia., Canberra Retrieved from: <http://www.ga.gov.au>, 2008
- 674 Woods, R. A.: Analytical model of seasonal climate impacts on snow hydrology: Continuous snowpacks, *Adv. Water Resour.*,  
675 32, 1465–1481, <https://doi.org/10.1016/j.advwatres.2009.06.011>, 2009.



- 676 Wright, D. P., Thyer, M., & Westra, S.: Influential point detection diagnostics in the context of hydrological model calibration.  
677 Journal of Hydrology, 527, 1161-1172. <https://doi.org/10.1016/j.jhydrol.2018.01.036>, 2018
- 678 Xu, T., & Hutchinson, M.: ANUCLIM version 6.1 user guide. The Australian National University, Fenner School of  
679 Environment and Society, Canberra. <https://fennerschool.anu.edu.au/files/anuclim61.pdf>, 2011
- 680 Zhang, S. X., Bari, M., Amirthanathan, G., Kent, D., MacDonald, A., & Shin, D. Hydrologic reference stations to monitor  
681 climate-driven streamflow variability and trends. In Hydrology and Water Resources Symposium 2014 (p. 1048), Engineers  
682 Australia, 2014.
- 683 Zhang, X. S., Amirthanathan, G. E., Bari, M. A., Laugesen, R. M., Shin, D., Kent, D. M., et al.. How streamflow has changed  
684 across Australia since the 1950s: evidence from the network of hydrologic reference stations. Hydrology and Earth System  
685 Sciences, 20(9), 3947. <https://doi.org/10.5194/hess-20-3947-2016>, 2016