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

Keirnan J. A. Fowler[1], Suwash Chandra Acharya[1], Nans Addor[2], Chihchung Chou[1*] and Murray C. Peel[1]

[1]Department of Infrastructure Engineering, University of Melbourne, Parkville, Victoria, Australia
[2]Department of Geography, University of Exeter, Exeter, UK
[*]now at: Department of Earth Sciences, Barcelona Supercomputing Centre, Barcelona, Spain

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

Abstract.

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

1 Introduction

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

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

Responding to these needs, multiple recent projects have publicly released large sample hydrological datasets (e.g., Arsenault 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 project – the Catchment Attributes and Meteorology for Large-sample Studies, or CAMELS, project. Originally launched for the United States (Newman et al., 2015; Addor et al., 2017), CAMELS datasets now exist for Chile (Alvarez Garreton et al., 2018), Great Britain (Coxon et al., 2020) and Brazil (Chagas et al., 2020). The defining features of a CAMELS dataset are that they complement data on streamflow (which is often publicly available) with other data types: (i) pre-processed climatic data for each catchment, such as would be required to run a hydrological model; and (ii) catchment attributes which characterise various aspects of the catchment without the need for field visitation (impractical for large samples). They also support download of the entire dataset in contrast to agency websites which may only support one-at-a-time download (if at all). Lastly, whereas government agencies reserve the right to retrospectively re-process their streamflow data (eg. due to rating curve changes), CAMELS datasets enable repeatability because a given CAMELS release effectively "freezes" the data, creating a consistent version that is available indefinitely via a persistent digital object identifier (DOI).

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

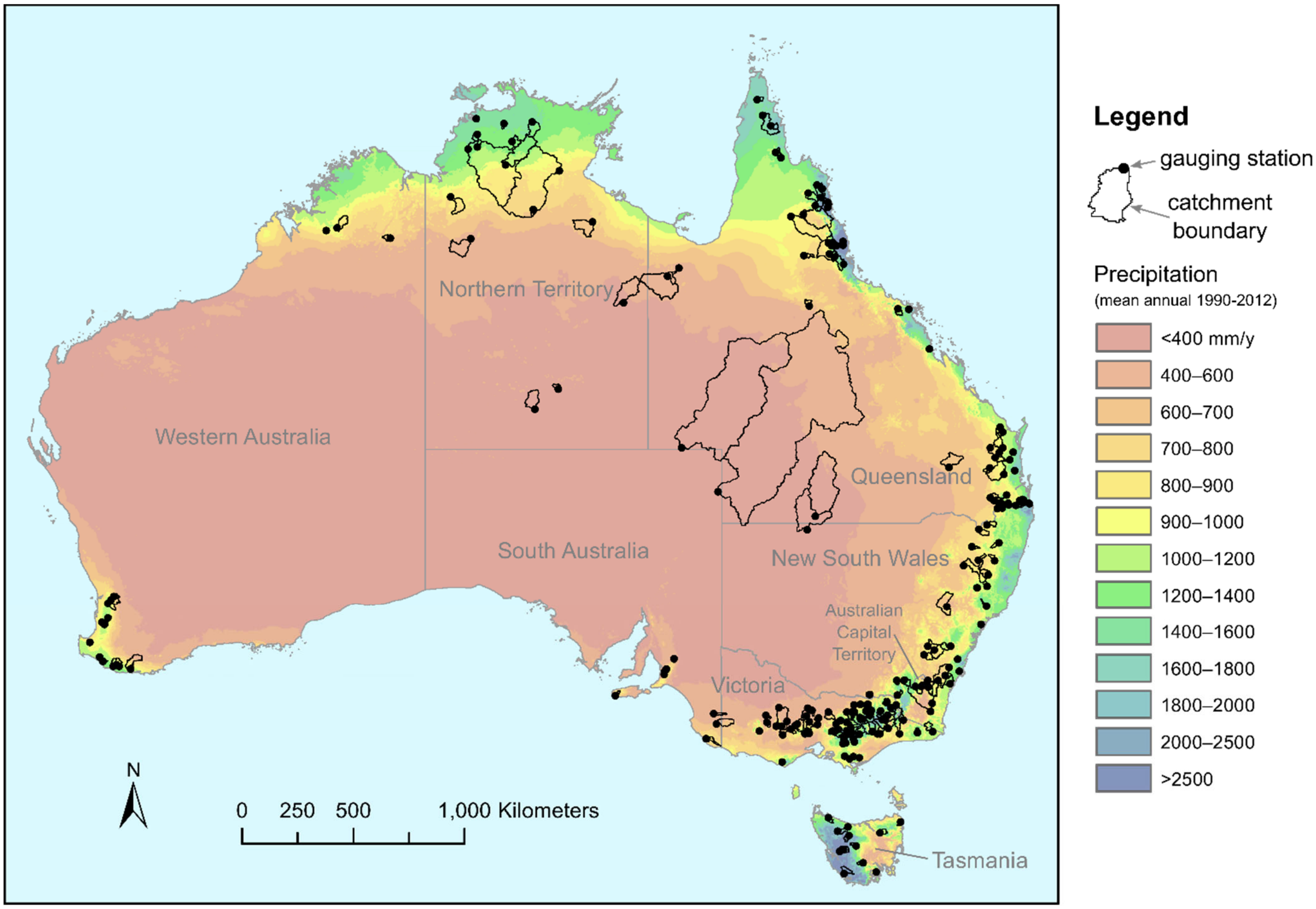

Figure 1: Location of the 222 CAMELS-AUS flow gauging stations and catchments, along with mean annual precipitation (from Jones et al., 2009) and Australian states and territories.

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

2 Rationale

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

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

Every region on earth is unique and has characteristics of interest for hydrological study. Within Australia and for CAMELS-AUS, three~~two~~ characteristics are noted here. Firstly, Australia contains many arid landscapes, and considerable advances in arid-zone hydrology have been made there (eg. Western et al., 2020). CAMELS-AUS contains more than twenty arid-zone rivers (depending on definition but see Figure 1), so the publication of the dataset opens the study of these rivers to a global pool of scientists. Added together with included arid-zone rivers in the USA and Chile (Addor et al., 2017; Alvarez-Garreton et al., 2018), the CAMELS datasets together provide a significant sample for the study of arid-zone hydrology.

Secondly, Australian catchments tend to have lower rainfall-runoff ratios, linked to higher evaporative demand. As shown in Figure 2d, the median rainfall runoff ratio among Australian catchments is approximately 0.25, compared to approximately 0.4 for the rest of the world. Australian catchments are often water-limited (at least on a seasonal basis), providing different modelling challenges to energy limited catchments from higher latitudes.

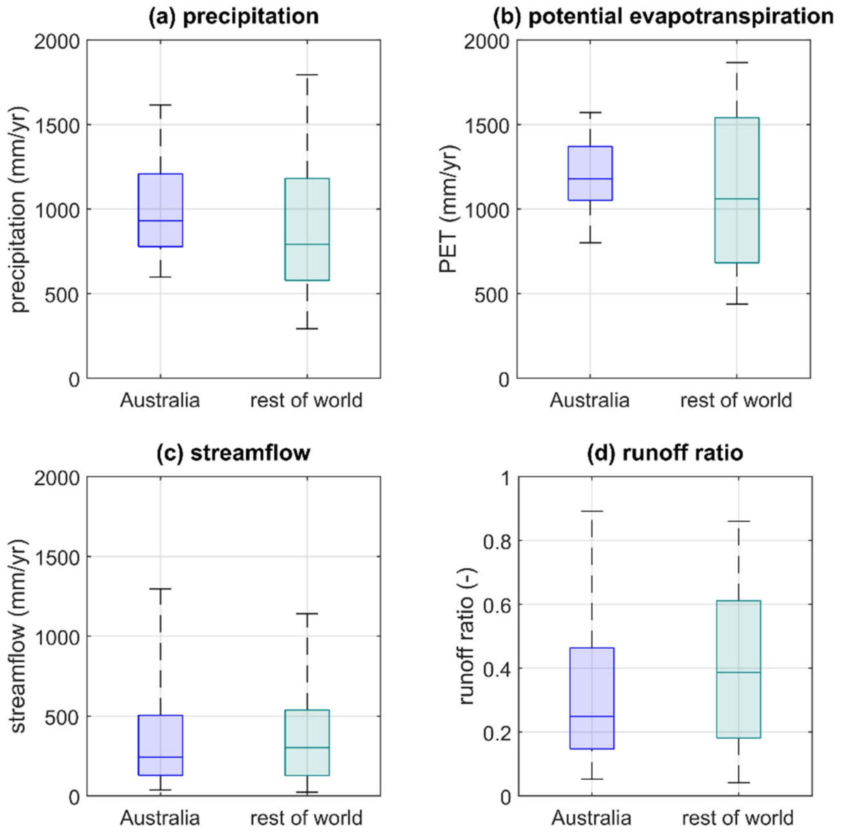

Figure 2: Mean annual values of hydrological variables for the global set of 699 catchments presented by Peel et al. (2010; $n_{\text{Australia}} = 123$; $n_{\text{rest of world}} = 576$). Boxplots show the 5th, 25th, 50th, 75th and 95th percentiles. Potential evapotranspiration in this dataset is a reference crop estimate using a method similar to Hargreaves' method, as outlined in Adam et al. (2006).

~~Another~~ Finally, a notable characteristic of Australian hydroclimatology is its tendency for multi-year spells of ~~relatively deep~~ climatic anomalies of larger magnitude compared to most other regions of the world (Peel et al., 2005), due partly to the strong influence of climate teleconnections such as the El Nino Southern Oscillation (ENSO, eg. Peel et al., 2002; Verdon-Kidd and Kiem, 2009). Recent severe droughts have affected south-east Australia, including the 13-year Millennium Drought (Van Dijk et al., 2014) which provided the opportunity for knowledge sharing with other drought-prone regions (Aghakouchak et al., 2014) and supplied many case studies of hydrological model failure (ie. the high bias and low model performance in differential split sample testing reported by, eg., Saft et al., 2016), which are under ongoing investigation (eg. Fowler et al., 2020b). In the context of providing credible runoff projections ~~in regions subject to drying climate~~, case studies of long droughts are the only means by which hydrologists can test hypotheses regarding how catchments respond physically to the onset of drier conditions, including aspects of long 'memory' (eg. Fowler et al., 2020b) and potential to shift behaviour, possibly in a quasi permanent fashion (eg. Peterson et al., 2021). Thus, it is hoped that the public release of datasets such as CAMELS-AUS may hasten scientific progress towards more defensible and robust hydrological models.

2.2 Context: hydrometeorological monitoring in Australia

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

2.3 Catchment choice: the Hydrologic Reference Stations dataset

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

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

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

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

It is noted that this choice is not intended to limit future inclusion of a wider range of stations/catchments. We envisage that the Hydrologic Reference Stations may provide the nucleus for future versions of the CAMELS-AUS dataset, while the current selection provides a sensible and pragmatic starting point. The Hydrological Reference Station dataset itself may be subject to future expansion, which would inform future CAMELS-AUS versions. Furthermore, whereas the Hydrological Reference Stations project, by definition, sought catchments which are minimally disturbed (or at least having stationarity of anthropogenic influence), future versions could be more inclusive so as to cater for studies examining diverse anthropogenic influences including changes over time – an approach already taken by CAMELS-GB (Coxon et al., 2020) and CAMELS-BR (Chagas et al., 2020). In summary, the current form of CAMELS-AUS should not be interpreted as setting a norm for future versions (or other datasets).

2.4 Local versus global datasets

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

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

In parallel, efforts are ongoing to increase the consistency among the CAMELS datasets (in terms of data products used to derive the time series and catchment attributes, and also naming conventions and data format, see Addor et al., 2019), in order to create a dataset that is globally consistent. This is part of a second phase, which will build upon the current phase which is focussed on the release of national products, such as CAMELS-AUS. To contribute to this effort, we have supplied the CAMELS-AUS catchment boundaries and gauge locations. Because of these ongoing efforts, our expectation is that the data introduced here, derived from Australian sources, will in time be complimented by data derived from global datasets.

3 CAMELS-AUS dataset technical description

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

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

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

3.1 Catchment selection rules

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

- **Initial selection:** 246 potential stations were initially selected based on three criteria: (i) record length (minimum of 1975 onwards); (ii) availability of data including historic rating curve information; and (iii) lack of regulation by large dams.
- **Invitation for stakeholders to suggest additional stations:** BOM consulted with seventy stakeholders from federal, state and territory agencies and water authorities, who were given the opportunity to add new stations to the list. This enlarged the list to 362 stations.
- **Targeted fact-finding:** To elicit information about each candidate station/catchment, the relevant agencies were asked a series of questions about the catchments in their jurisdiction relating both to past and present practices. Topics

included diversions, irrigation structures, upstream point source discharge, land clearing, forestry, urbanisation, fire, and farm dams.

- **Final selection:** the final selection process considered all the above information. A good coverage of Australia's various hydroclimatic regions was desired, although this is inherently limited by the coverage of the gauging network. Where possible, only stations with < 5% missing data and < 10% change in forest cover were selected.

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

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

3.2 Catchment boundaries

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

- For the six largest catchments (A0030501, A0020101, G8140040, G9030250, 424002 and 424201A), this process was not undertaken due to excessive computational requirements. For context, the largest catchment is approximately the size of the United Kingdom (see Figure 1);
- For a further four catchments (A2390519, A2390523, 307473 and 606185), the Arc Hydro process resulted in a catchment boundary that was inconsistent with the boundaries displayed on the Hydrologic Reference Station website. Although ~~severely~~ degraded for fast mapping, the website boundaries show the approximate position of the boundary as agreed with stakeholders / agencies who have local knowledge. Therefore, in cases of obvious mismatch, the Arc

Hydro-derived boundaries were assumed to be in error. Despite the 'blockiness' of the website boundaries, they were considered to be a better option for these four catchments.

For these ten catchments a protocol was developed to read the website's .json file to extract the boundary vertices. The website boundaries were then adopted. Note, more detail on the above considerations, including a selection of figues, is given in the dataset within the readme file README_CAMELS_AUS_Boundaries.pdf.

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

3.3 Catchment area and nestedness

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

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

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

3.4 Streamflow data and uncertainty

Streamflow timeseries data are provided by the BOM in two variants: non gap filled, and gap filled. The gap filled variant is filled using the daily rainfall-runoff model GR4J (Perrin et al., 2001) but the BOM have not published further methodological details about calibration method, validation procedures, or the specifics of the interpolation method. In addition to the

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

Short name	Description	Data source / notes
station_id	Station ID used by the Australian Water Resources Council.	Hydrologic Reference Stations (HRS) project, Bureau of Meteorology (BOM) www.bom.gov.au/water/hrs
station_name	River name and station name	
drainage_division	Drainage division, of the 13 defined by the BOM.	
river_region	River region, of the 218 defined by the BOM.	
notes	General notes about data issues and/or catchment area calculations	This study For daystart_Q, see Jian et al., (2017)
lat_outlet	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.	
long_outlet		
lat_centroid	Latitude and longitude at centroid of the catchment.	
long_centroid		
map_zone	Map zone used to calculate catchment area (function of longitude)	
catchment_area	Area of upstream catchment in km²	
state_outlet	Indicates which state or territory of Australia the outlet is within	
state-alt	If the catchment crosses a state or territory boundary, the alternative state or territory is listed here, otherwise "n/a"	
daystart	Time (UTC) for midnight local standard time (for *state_outlet*). This is the day start time for T_{max} and T_{min} (see Section 3.5.2).	
daystart_P	Time (UTC) for 9am local standard time (for *state_outlet*). 9am is when once-per-day precipitation measurements are reported (see Section 3.5.2).	
daystart_Q	Time (UTC) for streamflow day start time, assuming local standard time for *state_outlet*. This varies by state/territory (Section 3.5.2).	
nested_status	"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.	
next_station_ds	For "Level1" and "Level2" nested catchments, *NextStationDS* ('DS' meaning downstream) indicates the catchment they are contained within.	
num_nested_within	Indicates how many catchments are nested within this catchment.	
start_date	Streamflow gauging start date (yyyymmdd)	HRS (see above)
end_date	Streamflow gauging end date (yyyymmdd)	
prop_missing_data	Proportion of data missing between startdate and enddate	

streamflow data, ~~T~~the BOM also provide quality codes/flags. As mentioned in Section 2.1, the quality codes/flags of each state of Australia are different but the BOM has harmonised these to a common set (www.bom.gov.au/water/hrs/qc_doc.shtml). For CAMELS-AUS, these data are supplied as follows. Firstly, summary statistics about period of record (start date, end date and proportion of missing data) are provided in the attribute table, as listed in Table 1. Regarding timeseries data (Table 2), each of the above three data types (gap filled, non-gap-filled, and quality codes/flags), are provided within CAMELS-AUS exactly as supplied by the BOM, except that they are presented as a single fileacross all catchments. In addition, since the units of the streamflow files are ML d^{-1} whereas modelling studies typically use mm d^{-1}, CAMELS-AUS provides an additional streamflow timeseries file in mm d^{-1}.

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 31st 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 about record length. An increase in missing data post-1990 means that the data availability curves decrease slightly for the most recent period (dark blue).

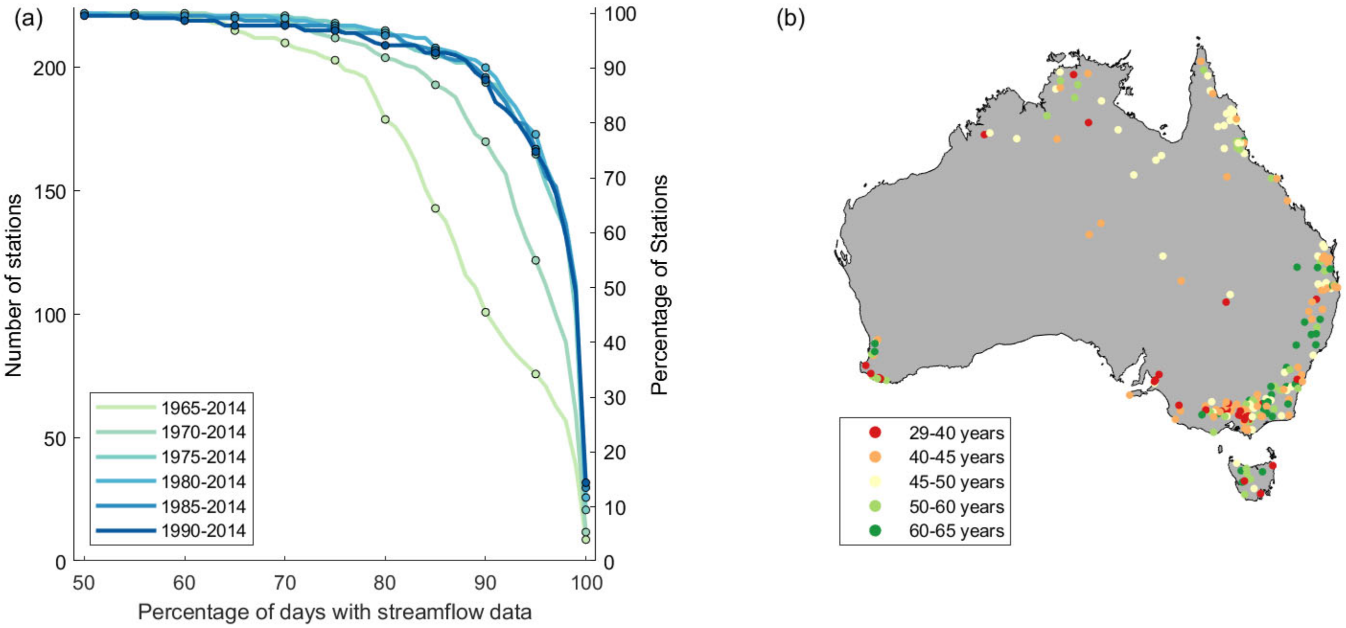

Figure 2: Plot after Coxon et al. (2020) showing (a) Number of stations with percentage of available streamflow data for different periods, b) Length of the flow time series for each gauge.

Information about streamflow uncertainty is provided with CAMELS-AUS (Table 1) from an earlier study by McMahon and Peel (2019). McMahon and Peel (2019) examined available rating curve data for 166 of the 222 stations, developed rating curves based on Chebyshev polynomials and estimated uncertainties using an approach which considered regression error and uncertainty in water level. The original authors post-processed their data to provide the following statistics (Table 3) for CAMELS-AUS: (i) Number of separate rating curves considered for a given station (median value across all stations was 3); (ii) Number of days considered across all curves (median value was ~14,700 or ~40 years); (iii) low, medium and high flow rates in mm d^{-1} (discharge exceeded 90%, 50% and 10% of the time over days considered by the curves); (iv) 95% confidence intervals around the low, medium and high flow estimates, expressed in percentage terms. However, for some stations considered by McMahon and Peel (2019) the above data are not supplied in full, for the following reasons: (a) the percentile flow is zero (cease to flow), leading to undefined relative uncertainty estimates due to the need to divide by zero; or (b) the percentile flow is outside the rated range, in which case neither upper or lower bounds are reported for that flow~~; and (c) the lower uncertainty bound goes below zero, in which case it is missing (but the upper bound is not)~~. In a small number of cases the uncertainty bound numbers are very high, and these cases are generally associated with near-cease-to-flow conditions. For example, the highest value of *Q_uncert_Q10_upper* (refer Table 3 for naming conventions) occurs for catchment 919309A, for which Q10 is 0.000023 mm d^{-1} but the upper bound is 0.05 mm d^{-1}, which is >2000 times higher. Thus, *Q_uncert_Q10_upper* for this catchment is 201,400%.

Included in dataset: Three streamflow timeseries files for flow, as explained above and listed in Table 2. One timeseries file for streamflow quality codes/flags. In the CAMELS-AUS attribute table: three attributes related to record extent and availability (startdate, enddate, prop_missing_data; see Table 1) plus eleven attributes related to streamflow uncertainty (Q_uncert_NumCurves, Q_uncert_N, Q_uncert_Q10, Q_uncert_Q10_upper, Q_uncert_Q10_lower, Q_uncert_Q50, Q_uncert_Q50_upper, Q_uncert_Q50_lower, Q_uncert_Q90, Q_uncert_Q90_upper, Q_uncert_Q90_lower; see Table 3).

3.5 Hydrometeorological timeseries

3.5.1 Availability of gridded hydrometeorological data in Australia

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

3.5.2 Limitations arising from ~~C~~conventions for definition of daily timesteps

Variables such as precipitation and streamflow are continuous ~~variables~~, and formatting into a daily timestep requires arbitrary conventions to split continuous time into 24 hour periods. For example, the BOM convention is that precipitation is split at 9am each day, and a daily value refers to the precipitation that occurred over the preceding 24 hours. Thus, if the BOM reports 18 mm precipitation for 14th March, this means that 18 mm was recorded between 9am 13th March and 9am 14th March. For streamflow, the conventions may vary depending on state or territory, but in collating the HRS data the BOM claims that conventions have been standardised to 9am to 9am (ie. the same as precipitation). However, an audit of HRS data conducted by Jian et al. (2017) investigated this standardisation. They report that data from the states of Victoria, New South Wales, Queensland and the Australian Capital Territory (which together account for 168 of 222 stations) were consistent with the 9am 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 earlier than expected) and South Australia and Northern Territory data (25 stations) appear to be subject to a 23:30 split (ie. 9.5 hours earlier than expected). Modellers should be mindful of these points when designing studies and interpreting results, since modelling results may be sensitive (Reynolds et al., 2018; Jian et al., 2017) to the day definitions for both precipitation and discharge (and, if relevant, the degree to which they are offset from one another). Regarding PET, the key variables (eg. temperature) are aligned directly with the day they are reported. This creates a time offset between PET and precipitation. In the experience of the CAMELS-AUS authors, this offset will typically make little difference to the results of (eg.) a rainfall runoff modelling study, since PET typically influences streamflow via seasonal, not daily, dynamics, in most CAMELS-AUS catchments. In the interests of providing CAMELS-AUS data subject to minimal 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 day to minimise the time offset between precipitation and streamflow.

A further consideration is that, due to Australia's large size, the CAMELS-AUS catchments occupy three different timezones. The majority are in a single zone (UTC+10:00) covering Queensland, New South Wales, the Australian Capital Territory, Victoria and Tasmania. However, South Australia and the Northern Territory are in a separate zone (UTC+09:30); while Western Australia uses UTC+08:00. In addition, daylight saving time is used in South Australia, New South Wales, the

Australian Capital Territory, Victoria, and Tasmania. During the daylight savings period (typically October to April) one hour needs to be added to the UTC times stated above. Given this multiplicity of combinations, measurements taken on either side of a state border that are marked with the same timestamp (eg. 9am) may, in reality, have been taken at different times.

Unfortunately, these limitations (related to timezones and day definitions) are inherent to the observations, and this then carries across into derivative products such as gridded climate data. In principle, if data were measured continuously it would be possible to redefine the day definitions and thus harmonise across timezones and data products, but unfortunately most observations are only taken once per day rather than continuously. Thus, there is little choice but to accept the use of these data despite these limitations.

3.5.2 Precipitation

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

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

3.5.3 Evaporative demand

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

3.5.4 Other timeseries

AWAP timeseries are provided for a further four variables: daily maximum temperature, daily minimum temperature, vapour pressure (1950 onwards), and solar radiation (1990 onwards). Solar radiation AWAP data has numerous gaps which have been filled by the average Julian Day value: for example, if the 5th March is missing, we adopt the average value over all non-

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

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

missing instances of the 5th March. SILO timeseries are provided for the following variables: daily maximum temperature, daily minimum temperature, vapour pressure, vapour pressure deficit, solar radiation, mean sea level pressure (1957 onwards), relative humidity at time of maximum temperature and relative humidity at time of minimum temperature. See Table 2 for adopted file names.

3.6 Catchment attributes

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

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

~~3.7.0~~3.6.1 Climatic indices and streamflow signatures

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

~~3.7.1~~3.6.2 Geology and soils

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

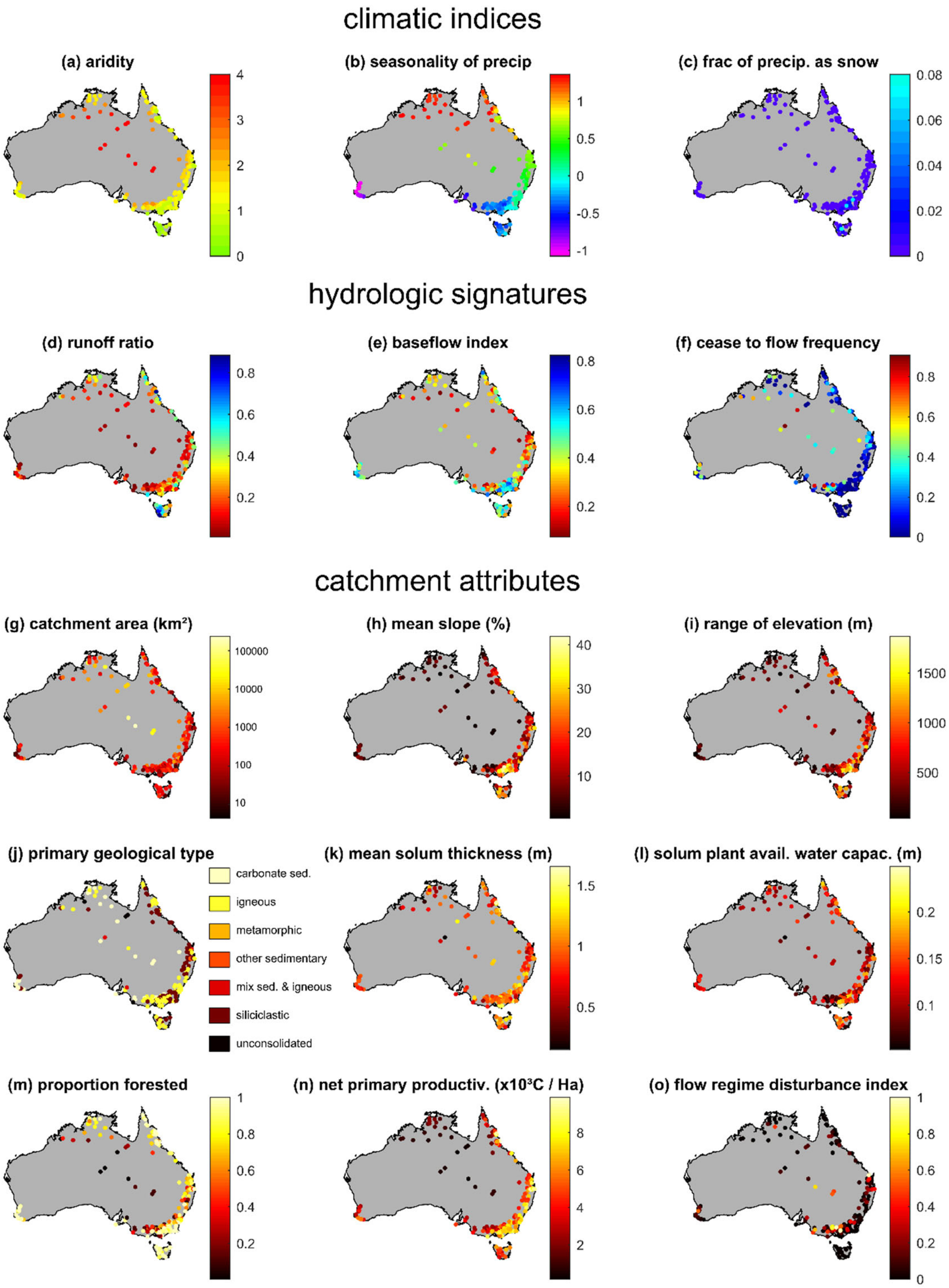

Figure 3: Maps of selected climatic indices (a – c), hydrologic signatures (d – f) and other catchment attributes (g – o). For definitions, see Tables 3 and 4.

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

~~3.7.2~~3.6.3 Topography and geometry

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

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

Further attributes are defined based on the *Multi-Resolution Valley Bottom Flatness* (MRVBF) index of Gallant et al. (2012). As the name indicates, the index relates to the shape of the landscape and the degree of deposited sediment. As explained in Table 4, the index values contrast erosional (MRVBF=0) locations with depositional (MRVBF>0) locations ranging from 'small hillside deposits' (MRVBF=1) through to 'extensive depositional basins' (MRVBF=9). Ten separate attributes are defined based on each integer value (0, 1 … 9) that MRVBF can take, indicating the proportion of the catchment in the given class. Lastly, using an earlier MRVBF version, Stein et al. (2011) analysed how common it is for a stream to pass through erosional landscapes (MRVBF=0), and defined this as an additional attribute, 'confinement'.

~~3.7.3~~3.6.4 Land cover and vegetation

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

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

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

~~3.7.4~~3.6.5 Anthropogenic influences

Anthropogenic influences are relevant to CAMELS-AUS because some catchments are minimally disturbed (eg. pre-European vegetation cover, few roads) while others, although unregulated, are nonetheless significantly changed from their natural state (eg. due to agricultural land use, small private (farm) dams, small towns and/or paved roads). Data on ~~A~~anthropogenic influences are taken from Stein et al. (2011) based on earlier work with the same lead author (Stein et al., 2002). The earlier study aimed to identify the 'wild' rivers of Australia by quantifying human impacts on two broad categories: the flow regime (sub categories: impoundments, flow diversions and levee banks) and the catchment (sub categories: infrastructure, settlements, extractive industries and landuse). Following the same method, Stein et al. (2011) provide a unitless index varying between zero and one to quantify human effects in each of these categories and subcategories, all of which are in CAMELS-AUS.

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

~~3.7.5~~3.6.6 Other catchment attributes

This final category contains indices that do not easily fit in one category, or that fit into more than one. The attributes quantifying human population are included here as they are relevant to both the land cover category and the anthropogenic influences, but fit neatly into neither. These population attributes, taken from Stein et al. (2011), are based on aggregation of census population to 9 second grid squares, and quantify the spatial average, the maximum grid value present in the catchment,

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	*geol_prim*	Two most common geologies (see list in cell below) with corresponding proportions.	-	Geoscience Australia (2008)	Preprocessed by Stein et al. (2011)
	geol_prim_prop				
	geol_sec				
	geol_sec_prop				
	unconsoldted	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.	-		
	igneous				
	silicsed				
	carbnatesed				
	othersed				
	metamorph				
	sedvolc				
	oldrock	Catchment proportion old bedrock	-		
	claya	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)
	clayb				
	sanda	As above, but % sand in the soil A horizon	%		
	solum_thickness	Mean soil depth considering all principle profile forms	m	McKenzie et al. (2000)	-
	ksat	Saturated hydraulic conductivity (areal mean)	$mm\ h^{-1}$	Western and McKenzie (2004)	Preprocessed by Stein et al. (2011)
	solpawhc	Solum plant available water holding capacity (areal mean)	mm		
Topography and geometry	*elev_min*	Elevation above sea level at gauging station	m	Gallant et al. (2009)	-
	elev_max	Catchment maximum and mean elevation above sea level	m	Hutchinson et al. (2008)	Preprocessed by Stein et al. (2011)
	elev_mean				
	elev_range	Range of elevation within catchment: elev_max-elev_min	m		-
	mean_slope_pct	Mean slope, calculated on a grid-cell-by-grid-cell basis	%	Gallant et al. (2012)	-
	upsdist	Maximum flow path length upstream	km	Hutchinson et al. (2008)	Preprocessed by Stein et al. (2011). For *strahler*, see Strahler (1957) For *elongratio*, see Gordon et al. (1992).
	strdensity	Ratio: (total length of streams) / (catchment area)	km^{-1}		
	strahler	Strahler stream order at gauging station	-		
	elongratio	Factor of elongation as defined in Gordon et al. (1992)	-		
	relief	Ratio: (mean elev. above outlet)/(max elev. above outlet)	-		
	reliefratio	Ratio: (elevation range)/(flow path distance)	-		
	mrvbf_prop_0 through to mrvbf_prop_9	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)
	confinement	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	*lc01_extracti*	Proportion of catchment occupied by land cover categories within the *Dynamic Land Cover Dataset* (DLCD): *mines and quarries (*ISO name: *extraction sites)* *lakes and dams (inland water bodies)* *salt lakes (salt lakes)* *irrigated cropping (irrigated cropping)* *irrigated pasture (irrigated pasture)* *irrigated sugar (irrigated sugar)* *rain fed cropping (rainfed cropping)* *rain fed pasture (rainfed pasture)* *rain fed sugar (rainfed sugar)* *wetlands (wetlands)* *closed tussock grassland (tussock grasses - closed)* *alpine meadows (alpine grasses - open)* *open hummock grassland (hummock grasses - open)*	-	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
	lc 03_waterbo				
	lc 04_saltlak				
	lc 05_irrcrop				
	lc06_irrpast				
	lc07_irrsuga				
	lc08_rfcropp				
	lc09_rfpastu				
	lc10_rfsugar				
	lc11_wetlands				
	lc14_tussclo				
	lc15_alpineg				
	lc16_openhum				

	Short name	**Description**	**Unit**	**Data source**	**Notes/references**
	lc18_opentus	*open tussock grasslands (tussock grasses - open)*			
	lc19_shrbsca	*scattered shrubs and grasses (shrubs and grasses - sparse - scattered)*			
	lc24_shrbden	*dense shrubland (shrubs - closed)*			
	lc25_shrbope	*open shrubland (shrubs - open)*			
	lc31_forclos	*closed forest (trees - closed)*			
	lc32_foropen	*open forest (trees - open)*			
	lc33_woodope	*open woodland (trees - scattered)*			
	lc34_woodspa	*woodland (trees - sparse)*			
	lc35_urbanar	*urban areas (urban areas)*			
	prop_forested	sum(LC_31, LC_32, LC_33, LC_34)			
	nv_grasses_n	Major vegetation sub-groups within the *National Vegetation Information System* (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: *grasses* *forests* *shrubs* *woodlands* *bare* *no data*	-	DEWR (2006)	Preprocessed by Stein et al. (2011)
	nv_grasses_e				
	nv_forests_n				
	nv_forests_e				
	nv_shrubs_n				
	nv_shrubs_e				
	nv_woodl_n				
	nv_woodl_e				
	nv_bare_n				
	nv_bare_e				
	nv_nodata_n				
	nv_nodata_e				
Anthropogenic Influences	*distupdamw*	maximum distance upstream before encountering a dam or water storage	km	Geoscience Australia (2004)	Preprocessed by Stein et al. (2011)
	impound_fac	Dimensionless factors quantifying human impacts on catchment hydrology, in two broad categories: - Flow regime factors: impoundments (*ImpoundmF*), flow diversions (*FlowDivF*), and levee banks (*LeveebankF*). The combined effect is disturbance index *FlowRegimeDI*; - Catchment factors: infrastructure (*InfrastrucF*), settlements (*SettlementF*), extractive industries (*ExtractiveIndF*) and landuse (*LanduseF*). The combined effect is captured in *CatchmentDI.* *FlowRegimeDI* and *CatchmentDI* are combined in *RiverDI*	-	Stein et al. (2002), updated by Stein et al. (2011)	
	flow_div_fac				
	leveebank_fac				
	infrastruc_fac				
	settlement_fac				
	extract_inf_fac				
	landuse_fac				
	catchment_di				
	flow_regime_di				
	river_di				
Other	*pop_mean*	Average and maximum human population density in catchment across 3" grid squares.	km^{-2}	ABS (2006)	Preprocessed by Stein et al. (2011)
	pop_max				
	pop_gt_1	Proportion of catchment with population density exceeding 1 person / km² and 10 people / km²	-		
	pop_gt_10				
	erosivity	Rainfall erosivity (spatial average across catchment)	MJ mm ha^{-1} h^{-1}	NLWRA (2001)	
	anngro_mega	Average annual growth index value for megatherm, mesotherm and microtherm plants, respectively	-	Xu and Hutchinson (2011)	
	anngro_meso				
	anngro_micro				
	gromega_seas	Seasonality of growth index value for megatherm, mesotherem and microtherm plants, respectively	-		
	gromeso_seas				
	gromicro_seas				
	npp_ann	Net Primary Productivity estimated by Raupach et al. (2002) for pre-European settlement conditions: - annually; and - for the twelve calendar months of the year	tC Ha^{-1}	Raupach et al. (2002)	Preprocessed by Stein et al. (2011)
	npp_1 through to *npp_12*				

and the proportion of grid squares exceeding 1 and 10 people km^{-2}. A further inclusion is the erosivity which is primarily a climatic attribute but is often used by studies associated with the soil category. The erosivity is taken from Stein et al. (2011) and in turn from the National Land and Water Resources Audit (NLWRA, 2001).

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

4 Data availability

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

5 Conclusions

This paper introduced a new freely available dataset for Australia, CAMELS-AUS. It is ~~the first large sample hydrology dataset for Australia and~~ the fifth CAMELS dataset worldwide, and the first large sample hydrology dataset for Australia to include data on climatic forcing, catchment attributes, and gauging uncertainty. CAMELS-AUS provides timeseries data (streamflow and 18 climatic variables) and a broad set of 134 attributes, for 222 unregulated catchments from across Australia. Given the unique hydroclimate of Australia, with high hydroclimatic variability and many case studies of multi-year drought, it is hoped that the release of this dataset will accelerate progress in such fields as arid zone hydrology and the study of hydrology under a changing climate.

6 Author contribution

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

7 Competing interests

The authors declare they have no conflict of interest.

8 Acknowledgements

The authors acknowledge the Bureau of Meteorology, Australia for their support and permission to undertake this project on the Hydrologic Reference Stations. The authors acknowledge the excellent work of Janet Stein and coauthors (Stein et al., 2011) to whom this dataset owes more than half the listed catchment attributes. KF acknowledges funding from the Australian Research Council (LP170100598) and the Bureau of Meteorology (TP705654) during the period of preparation of this manuscript. KF appreciates the support of Gemma Coxon (University of Bristol) who encouraged his foray into large sample hydrology. NA acknowledges support from the Swiss National Science Foundation (fellowship P400P2_180791).

9 References

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

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

Adam, J. C., Clark, E. A., Lettenmaier, D. P., & Wood, E. F.: Correction of global precipitation products for orographic effects. Journal of Climate, 19(1), 15-38. https://doi.org/10.1175/JCLI3604.1, 2006.

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

Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N. and Clark, M. P.: A Ranking of Hydrological Signatures Based on Their Predictability in Space, Water Resour. Res., 54(11), 8792– 8812, https://doi.org/10.1029/2018WR022606, 2018.

Addor, N., Newman, A. J., Mizukami, N. and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrol Earth Syst Sci, 21(10), 5293–5313, 655 https://doi.org/10.5194/hess-21-5293-2017, 2017.

Aghakouchak, A., D. Feldman, M. Stewardson, J. Saphores, S. Grant, and B. Sanders: Australia's Drought: Lessons for California, Science, 343 (6178), 1430-1431, https://doi.org/10.1126/science.343.6178.1430, 2014

Alvarez-Garreton, C., Mendoza, P. A., Boisier, J. P., Addor, N., Galleguillos, M., ZambranoBigiarini, M., Lara, A., Puelma, C., Cortes, G., Garreaud, R., McPhee, J. and Ayala, A.: The CAMELS-CL dataset: catchment attributes and meteorology for large sample studies – Chile dataset, Hydrol. Earth Syst. Sci., 22(11), 5817–5846, https://doi.org/10.5194/hess-22-5817-2018 , 2018.

Andréassian, V., et al., 2006. Introduction and synthesis : why should hydrologists work on a large number of basin data sets ? Large sample basin experiments for hydrological model parameterization: results of the Model Parameter Experiment–MOPEX. Vol. 307. CEH Wallingford, UK: IAHS Publ, 1–5.

Arsenault, R., Bazile, R., Ouellet Dallaire, C. and Brissette, F.: CANOPEX: A Canadian hydrometeorological watershed database, Hydrol. Process., 30(15), 2734–2736, https://doi.org/10.1002/hyp.10880, 2016.

ASCE (American Society for Civil Engineering): ASCE's Standardized Reference Evapotranspiration Equation, proceedings of the National Irrigation Symposium, Phoenix, Arizona, 2000

Ashcroft, L., Karoly, D. J., & Gergis, J.: Southeastern Australian climate variability 1860–2009: a multivariate analysis. International Journal of Climatology, 34(6), 1928-1944. https://doi.org/10.1002/joc.3812, 2014

BOM (Bureau of Meteorology, Australia): Hydrologic Reference Stations data update 2015. Website. www.bom.gov.au/water/hrs/update_2015.shtml, accessed June 1st 2020.

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-BR: Hydrometeorological time series and landscape attributes for 897 catchments in Brazil, Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2020-67, 2020.

Court, A.: Measures of streamflow timing, J. Geophys. Res., 67, 4335–4339, https://doi.org/10.1029/JZ067i011p04335, 1962.

Coxon, G., Addor, N., Bloomfield, J. P., Freer, J., Fry, M., Hannaford, J., Howden, N. J. K., Lane, R., Lewis, M., Robinson, E. L., Wagener, T., and Woods, R.: CAMELS-GB: Hydrometeorological time series and landscape attributes for 671 catchments in Great Britain, Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2020-49, in review, 2020.

CSIRO: AUS SRTM 1sec MRVBF mosaic v01. Bioregional Assessment Source Dataset. Viewed 18 June 2018, http://data.bioregionalassessments.gov.au/dataset/79975b4a-1204-4ab1-b02b-0c6fbbbbbcb5, 2016

DEWR (Department of the Environment and Water Resources, Australia): Estimated Pre-1750 Major Vegetation Subgroups - NVIS Stage 1, Version 3.1. https://www.environment.gov.au/land/native-vegetation/national-vegetation-information-system, 2008

Do, H. X., Gudmundsson, L., Leonard, M. and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM)-Part 1: The production of a daily streamflow archive and metadata, Earth Syst. Sci. Data, 10(2), 765–785, https://doi.org/10.5194/essd-10-765-2018, 2018.

Falkenmark, M. and Chapman, T.: *Comparative hydrology: An ecological approach to land and water resources*. Unesco, Paris, 1989.

FAO (Food and Agriculture Organization of the United Nations): Irrigation and drainage paper 56: Crop evapotranspiration - Guidelines for computing crop water requirements, 1998.

Fowler, K., Peel, M., Western, A., & Zhang, L. Improved rainfall-runoff calibration for drying climate: Choice of objective function. Water Resources Research, 54. https://doi.org/10.1029/2017WR022466, 2018

Fowler, K., Peel, M., Western, A., Zhang, L., and Peterson, T. J.: Simulating runoff under changing climatic conditions: Revisiting an apparent deficiency of conceptual rainfall-runoff models, Water Resources Research, 52, https://doi.org/10.1002/2015WR018068, 2016

Fowler, K; Acharya, S. C.; Addor, N.; Chou, C.; Peel, M.: CAMELS-AUS v1: Hydrometeorological time series and landscape attributes for 222 catchments in Australia. Dataset archived with online repository PANGAEA, https://doi.pangaea.de/10.1594/PANGAEA.921850, (dataset in review at time of article), 2020a

Fowler, K., Knoben, W., Peel, M., Peterson, T., Ryu, D., Saft, M., Seo, K., Western, A. Many commonly used rainfall-runoff models lack long, slow dynamics: implications for runoff projections. Water Resources Research, https://doi.org/10.1029/2019WR025286, 2020b

Gallant, J. C., and T. I. Dowling: A multiresolution index of valley bottom flatness for mapping depositional areas, WaterResour. Res., 39(12), 1347, https://doi.org/10.1029/2002WR001426, 2003

Gallant, J., Wilson, N., Tickle, P.K., Dowling, T., Read, A.: 3 second SRTM Derived Digital Elevation Model (DEM) Version 1.0. Record 1.0. Geoscience Australia, Canberra. http://pid.geoscience.gov.au/dataset/ga/69888, 2009.

Gallant, John; Austin, Jenet: Slope derived from 1" SRTM DEM-S. v4. CSIRO. Data Collection. https://doi.org/10.4225/08/5689DA774564A, 2012

Geoscience Australia: Dams and Water Storages 1990, Geoscience Australia, Canberrra. Later versions at https://data.gov.au/data/dataset/ce5b77bf-5a02-4cf8-9cf2-be4a2cee2677, 2004

Geoscience Australia: Surface Geology of Australia 1:1 million scale dataset. Latest version is available at https://data.gov.au/dataset/ds-dga-48fe9c9d-2f10-49d2-bd24-ac546662c4ec/details, 2008

Ghiggi, G., Humphrey, V., Seneviratne, S. I. and Gudmundsson, L.: GRUN: An observations-based global gridded runoff dataset from 1902 to 2014, Earth Syst. Sci. Data Discuss., 1–32, https://doi.org/10.5194/essd-2019-32, 2019.

Gordon, N. D., McMahon, T. A., Finlayson, B. L., & Christopher, J.: Stream Hydrology: an Introduction for Ecologists. John Wiley & Sons, Ltd. 1992

Gupta, H.V., et al. Large-sample hydrology: a need to balance depth with breadth. Hydrology and Earth System Sciences, 18, 463–477. https://doi.org/10.5194/hess-18-463-2014, 2014

Hutchinson, M.F., Stein, J.L., Stein, J.A., Anderson, H., Tickle, P.K.: GEODATA 9 second DEM and D8: Digital Elevation Model Version 3 and Flow Direction Grid 2008. Record DEM-9S.v3. Geoscience Australia, Canberra. http://pid.geoscience.gov.au/dataset/ga/66006, 2008

Isbell, R. F. The Australian Soil Classification. Revised Edition. CSIRO Publishing, Melbourne. See https://www.asris.csiro.au/themes/Atlas.html#Atlas_Digital, 2002

Jeffrey, S. J., Carter, J. O., Moodie, K. B., & Beswick, A. R.: Using spatial interpolation to construct a comprehensive archive of Australian climate data. Environmental Modelling and Software, 16(4), 309–330. https://doi.org/10.1016/S1364-8152(01)00008-1, 2001

Jian, J., J. Costelloe, D. Ryu, Q. J. Wang.: Does a fifteen-hour shift make much difference? – Influence of time lag between rainfall and discharge data on model calibration. 22nd International Congress on Modelling and Simulation, Hobart, Tasmania, December 2017. https://www.mssanz.org.au/modsim2017/H3/jian.pdf, 2017

Jones, D. A., Wang, W., & Fawcett, R.: High-quality spatial climate data-sets for Australia. Australian Meteorological and Oceanographic Journal, 58(4), 233–248. https://doi.org/10.22499/2.5804.003, 2009

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S. and Nearing, G. S.: Toward Improved Predictions in Ungauged Basins: Exploiting the Power of Machine Learning, Water Resour. Res., 55(12), 11344–11354, https://doi.org/10.1029/2019WR026065, 2019.

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S. and Nearing, G.: Benchmarking a Catchment-Aware Long Short-Term Memory Network (LSTM) for Large-Scale Hydrological Modeling, Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-368, 2019.

Kuentz, A., Arheimer, B., Hundecha, Y. and Wagener, T.: Understanding hydrologic variability across Europe through catchment classification, Hydrol. Earth Syst. Sci., 21, 2863–2879, https://doi.org/10.5194/hess-21-2863-2017, 2017.

Ladson, A., Brown, R., Neal, B., and Nathan, R.: A standard approach to baseflow separation using the Lyne and Hollick filter, Aust. J. Water Resour., 17, 25–34, https://doi.org/10.7158/W12- 028.2013.17.1, 2013.

Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., David, C. H., Durand, M., Pavelsky, T. M., Allen, G. H., Gleason, C. J. and Wood, E. F.: Global reconstruction of naturalized river flows at 2.94 million reaches, Water Resour. Res., 2019WR025287, https://doi.org/10.1029/2019WR025287, 2019.

Linke, S., Lehner, B., Dallaire, C. O., Ariwi, J., Grill, G., Anand, M., Beames, P., Burchard-levine, V., Moidu, H., Tan, F. and Thieme, M.: HydroATLAS : global hydro-environmental sub-basin and river reach characteristics at high spatial resolution, Sci. Data, 0–25, https://doi.org/10.1038/s41597-019-0300-6, 2019.

Liu, S. F., Raymond, O. L., Stewart, A. J., Sweet, I. P., Duggan, M., Charlick, C., Phillips, D. and Retter, A. J.: Surface geology of Australia 1:1,000,000 scale, Northern Territory [Digital Dataset]. The Commonwealth of Australia, Geoscience Australia., Canberra Retrieved from: http://www.ga.gov.au, 2006

Lymburner, L., Tan, P., McIntyre, A., Thankappan, M., Sixsmith, J.: Dynamic Land Cover Dataset Version 2.1. Geoscience Australia, Canberra. http://pid.geoscience.gov.au/dataset/ga/83868, 2015

Mathevet, T., Gupta, H., Perrin, C., Andréassian, V., & Le Moine, N.: Assessing the performance and robustness of two conceptual rainfall-runoff models on a worldwide sample of watersheds. Journal of Hydrology, https://doi.org/10.1016/j.jhydrol.2020.124698, 2020.

McKenzie, N.J., Jacquier, D.W., Ashton L.J. and Cresswell, H.P.: Estimation of Soil Properties Using the Atlas of Australian Soils. CSIRO Land and Water Technical Report 11/00. see https://www.asris.csiro.au/themes/Atlas.html#Atlas_Digital, 2000

McInerney, D., Thyer, M., Kavetski, D., Lerat, J., & Kuczera, G. (2017). Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors. Water Resources Research, 53(3), 2199-2239. https:/doi.org/10.1002/2016WR019168. 2017

McMahon, T.A., Finlayson, B.L., Haines, A.T. and Srikanthan, R:. *Global runoff: continental comparisons of annual flows and peak discharges*. Catena Verlag, 1992.

McMahon T and Peel M.: Uncertainty in stage–discharge rating curves: application to Australian Hydrologic Reference Stations data, Hydrological Sciences Journal, 64:3, 255-275, https://doi.org/10.1080/02626667.2019.1577555, 2019

Morton, F. I.: Operational estimates of areal evapotranspiration and their significance to the science and practice of hydrology. Journal of Hydrology, 66, 1–76, 1983

National Land and Water Resources Audit: Gridded soil information layers. Canberra. View at www.asris.csiro.au/mapping/viewer.htm, 2001

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T. and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological dataset for the contiguous USA: dataset characteristics and assessment of regional variability in hydrologic model performance, Hydrol. Earth Syst. Sci., 19, 209–223, https://www.hydrol-earth-syst-sci.net/19/209/2015/, 2015.

Olarinoye, T., Gleeson, T., Marx, V., Seeger, S., Adinehvand, R., et al. Global karst springs hydrograph dataset for research and management of the world's fastest-flowing groundwater, Sci. Data, 7(1), 1–9, https://doi.org/10.1038/s41597-019-0346-5, 2020.

Peel, M.C., McMahon, T.A. and Finlayson, B.L.: Variability of annual precipitation and its relationship to the El Niño–Southern Oscillation. Journal of Climate, 15(5), 545-551, 2002.

Peel, M.C., McMahon, T.A. and Finlayson, B.L.: Continental differences in the variability of annual runoff-update and reassessment. Journal of Hydrology, 295(1-4), 185-197, 2004.

Peel, M. C., G. G. S. Pegram, and T. A. McMahon: Global analysis of runs of annual precipitation and runoff equal to or below the median: Run magnitude and severity, International Journal of Climatology, 24 (7), 549-568, https://doi.org/10.1002/joc.1147, 2005

Peel, M. C., McMahon, T. A., & Finlayson, B. L.: Vegetation impact on mean annual evapotranspiration at a global catchment scale. Water Resources Research, 46(9). https://doi.org/10.1029/2009WR008233, 2010

Peterson, T. J., Wasko, C., Saft, M., & Peel, M. C.: AWAPer: An R package for area weighted catchment daily meteorological data anywhere within Australia. Hydrological Processes, 34(5), 1301-1306. https://doi.org/10.1002/hyp.13637, 2020

Peterson, T. J., Saft, M., Peel, M. C., & John, A.: Watersheds may not recover from drought. Science, 372(6543), 745-749. https://doi.org/10.1126/science.abd5085, 2021

Pool, S., Viviroli, D., & Seibert, J. Value of a limited number of discharge observations for improving regionalization: A large-sample study across the United States. Water Resources Research, 55, 363–377. https://doi.org/10.1029/2018WR023855, 2019

Raupach MR, Kirby JM, Barrett DJ, and Briggs PR: Balances of Water, Carbon, Nitrogen and Phosphorus in Australian Landscapes version 2.04, CSIRO Land and Water, Canberra, http://www.clw.csiro.au/publications/technical2001/tr40-01.pdf, 2002

Raymond, O. L., Liu, S. F. and Kilgour, P.: Surface geology of Australia 1:1,000,000 scale, Tasmania - 3rd edition [Digital Dataset]. The Commonwealth of Australia, Geoscience Australia., Canberra Retrieved from: http://www.ga.gov.au, 2007a

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 scale, New South Wales - 2nd edition [Digital Dataset]. The Commonwealth of Australia, Geoscience Australia., Canberra Retrieved from: http://www.ga.gov.au, 2007b

Raymond, O. L., Liu, S. F., Kilgour, P., Retter, A. J. and Connolly, D. P.: Surface geology of Australia 1:1,000,000 scale, Victoria - 3rd edition [Digital Dataset]. The Commonwealth of Australia, Geoscience Australia., Canberra Retrieved from: http://www.ga.gov.au, 2007c

Rayner, D. Australian synthetic daily Class A pan evaporation. Technical Report December, Queensland Department of Natural Resources and Mines, Indooroopilly, Qld., Australia, 40 pp., https://data.longpaddock.qld.gov.au/static/silo/pdf/AustralianSyntheticDailyClassAPanEvaporation.pdf, 2005

Reynolds, J. E., S. Halldin, J. Seibert, & C. Y. Xu: Definitions of climatological and discharge days: do they matter in hydrological modelling?. Hydrological Sciences Journal, 63(5), 836-844. https://doi.org/10.1080/02626667.2018.1451646, 2018

Sankarasubramanian, A., Vogel, R. M., and Limbrunner, J. F.: Climate elasticity of streamflow in the United States, Water Resour. Res., 37, 1771–1781, https://doi.org/10.1029/2000WR900330, 2001.

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G.: Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA, Hydrol. Earth Syst. Sci., 15, 2895–2911, https://doi.org/10.5194/hess-15-2895-2011, 2011.

Shen, C.: A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists, Water Resour. Res., 54, 8558–8593, https://doi.org/10.1029/2018WR022643, 2018.

Skinner, D., & Langford, J. Legislating for sustainable basin management: the story of Australia's Water Act (2007). Water Policy, 15(6), 871-894. 2013

Stein, J. L., Stein, J. A. and Nix, H. A.: Spatial analysis of anthropogenic river disturbance at regional and continental scales: identifying the wild rivers of Australia. Landscape and Urban Planning, 60, 1-25, https://doi.org/10.1016/S0169-2046(02)00048-8, 2002

Stein, J. L., Hutchinson, M. F. and Stein, J. A.: National Catchment and Stream Environment Database version 1.1.4. Available at http://pid.geoscience.gov.au/dataset/ga/73045, 2011

Stewart, A. J., Sweet, I. P., Needham, R. S., Raymond, O. L., Whitaker, A. J., Liu, S. F., Phillips, D., Retter, A. J., Connolly, D. P. and Stewart, G.: Surface geology of Australia 1:1,000,000 scale, Western Australia [Digital Dataset]. The Commonwealth of Australia, Geoscience Australia., Canberra Retrieved from: http://www.ga.gov.au, 2008

Strahler, A. N. Quantitative analysis of watershed geomorphology. Eos, Transactions American Geophysical Union, 38(6), 913-920. https://doi.org/10.1029/TR038i006p00913, 1957

Tozer, C. R., Kiem, A. S., & Verdon-Kidd, D. C. On the uncertainties associated with using gridded rainfall data as a proxy for observed. Hydrology and Earth System Sciences, 16(5), 1481. https://doi.org/10.5194/hess-16-1481-2012, 2012

Turner, M., Bari, M., Amirthanathan, G., & Ahmad, Z.: Australian network of hydrologic reference stations-advances in design, development and implementation. In Hydrology and Water Resources Symposium 2012 (p. 1555). Engineers Australia, http://www.bom.gov.au/water/hrs/media/static/papers/Turner2012.pdf, 2012

Verdon-Kidd, D. C., and A. S. Kiem: Nature and causes of protracted droughts in southeast Australia: Comparison between the Federation, WWII, and Big Dry droughts, Geophysical Research Letters, 36 (22), 1-6, https://doi.org/10.1029/2009GL041067, 2009.

Vertessy, R. A.: Water information services for Australians. Australasian Journal of Water Resources, 16(2), 91-105. https://www.tandfonline.com/doi/abs/10.7158/13241583.2013.11465407, 2013

Viglione, A., Borga, M., Balabanis, P. and Blöschl, G.: Barriers to the exchange of hydrometeorological data in Europe: Results from a survey and implications for data policy. Journal of Hydrology, 394(1-2), 63-77, 2010.

Western, A. and McKenzie, N.: Soil hydrological properties of Australia Version 1.0.1, CRC for Catchment Hydrology, Melbourne, 2004

Western, A. W., Matic, V., & Peel, M. C.: Justin Costelloe: a champion of arid-zone water research. Hydrogeology Journal, 28(1), 37-41, https://doi.org/10.1007/s10040-019-02051-7, 2020

Whitaker, A. J., Champion, D. C., Sweet, I. P., Kilgour, P. and Connolly, D. P.: Surface geology of Australia 1:1,000,000 scale, Queensland 2nd edition [Digital Dataset]. The Commonwealth of Australia, Geoscience Australia., Canberra Retrieved from: http://www.ga.gov.au, 2007

Whitaker, A. J., Glanville, D. H., English, P. M., Stewart, A. J., Retter, A. J., Connolly, D. P., Stewart, G. A. and Fisher, C. L.: Surface geology of Australia 1:1,000,000 scale, South Australia [Digital Dataset]. The Commonwealth of Australia, Geoscience Australia., Canberra Retrieved from: http://www.ga.gov.au, 2008

Woods, R. A.: Analytical model of seasonal climate impacts on snow hydrology: Continuous snowpacks, Adv. Water Resour., 32, 1465–1481, https://doi.org/10.1016/j.advwatres.2009.06.011, 2009.

Wright, D. P., Thyer, M., & Westra, S.: Influential point detection diagnostics in the context of hydrological model calibration. Journal of Hydrology, 527, 1161-1172.https://doi.org/10.1016/j.jhydrol.2018.01.036, 2018

Xu, T., & Hutchinson, M.: ANUCLIM version 6.1 user guide. The Australian National University, Fenner School of Environment and Society, Canberra. https://fennerschool.anu.edu.au/files/anuclim61.pdf, 2011

Zhang, S. X., Bari, M., Amirthanathan, G., Kent, D., MacDonald, A., & Shin, D. Hydrologic reference stations to monitor climate-driven streamflow variability and trends. In Hydrology and Water Resources Symposium 2014 (p. 1048), Engineers Australia, 2014.

Zhang, X. S., Amirthanathan, G. E., Bari, M. A., Laugesen, R. M., Shin, D., Kent, D. M., et al.. How streamflow has changed across Australia since the 1950s: evidence from the network of hydrologic reference stations. Hydrology and Earth System Sciences, 20(9), 3947. https://doi.org/10.5194/hess-20-3947-2016, 2016