Author response and manuscript changes for ESSD manuscript:

Fowler et al., (2021): CAMELS-AUS: Hydrometeorological time series and landscape attributes for 222 catchments in Australia

Green: comments from referees

Blue: author's response (identical to Author Comments submission)

Purple: author's changes in manuscript

Note: line numbers for changes refer to the revised manuscript (non-track changes version)

Anonymous Reviewer 1

I only see praises to formulate concerning this paper. Congratulations.

Author response: Thank you very much for your effort to review the paper, and for these kind words.

A few very minor propositions:

1. in the abstract, you write "This dataset, the first of its kind in Australia, allows users…" : the "first of its kind in Australia" could be contested, it may not be useful here.

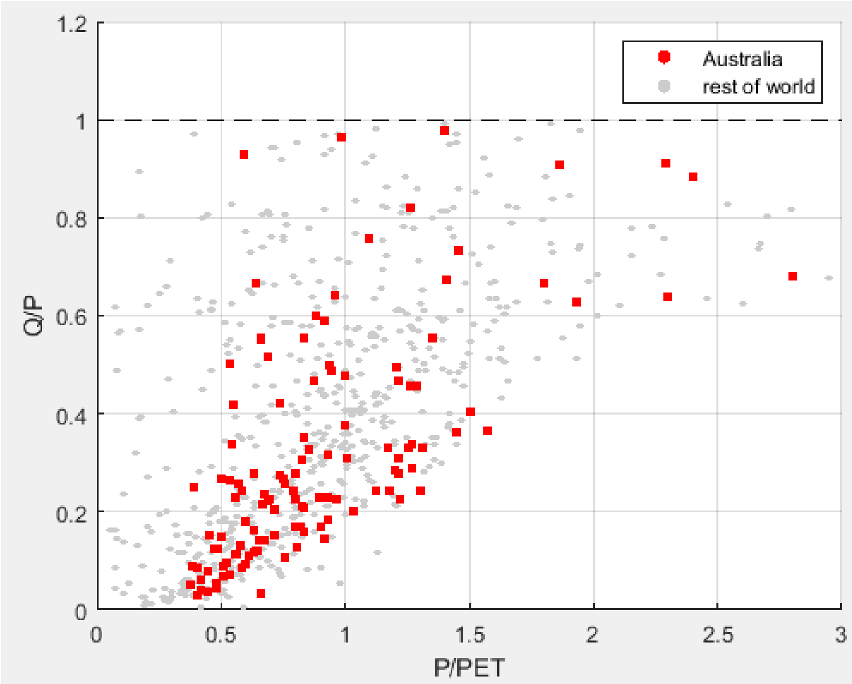
   Author response: we have removed these words, thank you. ***Line 15 now reads "This dataset allows users …"***

2. l. 51 "other territories held by the Australian nation.": this seems almost a poetical-political statement… why not writing "other Australian territories"

   Author response: accepted, thank you ***(see Line 53-54)***

3. section 2.1 : I wonder whether it would be appropriate to also mention one originality of Australian hydrology: the presence of a large number of catchments with low average rainfall yield (am I right? That's the opinion I have at least, but I may be wrong). And there would be in my opinion a great graph to add: plot the 222 catchments in the hydrological ("pseudo-Budyko") space (Q/P vs P/E is my favorite), with the "physical" limits and add the catchments from the existing CAMELS datasets.

   Author response: We have considered this carefully and decided to ***include a plot on a similar theme [Figure 2]****. As you can see in the revised manuscript* ***[Figure 2]****, the plot compares distributions of values for Australia and the rest of the world, for four variables (P, PET, Q and rainfall runoff ratio). As you say, Australian catchments tend to have lower rainfall runoff ratios, linked to higher PET* ***[and we have added text to this effect at lines 80-83]****. We considered compiling data across CAMELS datasets as suggested, but decided against it due to the dual problems of (i) the datasets adopt different PET formulations, making comparison difficult; and (ii) the CAMELS datasets are not yet representative of the globe. In response to these issues, we adopted the global dataset of Peel et al. (2010, https://doi.org/10.1029/2009WR008233) and trialed the plot you suggested (see below), but* *there's some values in the plot that seem quite unrealistic (eg. Q/P = 97% when P/PET = 0.4) which arise from some methods used in dataset compilation that were the best available at the time but would now be considered out of date (eg. Theissen polygons to calculate P for a mountainous catchment). Seeking to avoid the reader being distracted by these anomalies, and reasoning that the distribution of values would be relatively reliable in spite of them, we settled on showing the differences in distributions of values. We hope this serves the intention of the reviewer's comment.*



4. section 2.1: insist on the fact that only the occurrence of long-lasting droughts can allow to test hypotheses on the memory of hydrological systems.

   Author response: good point. ***We have added the following text [line 94-97]: "In the context of providing credible runoff projections, 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)."***

5. section 2.3: may be mention how the issue of time zones has been dealt with?

   Author response: Thanks for the prompt. It is easy to forget about timezones because, unfortunately, it is outside of our control to harmonise them. At Section 3.5.1 we have added the following text: "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 … 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 … but unfortunately most observations are only taken once per day rather than continuously. Thus, there is little choice but to accept the use of this data despite these limitations."*

6. l. 261: I don't understand the problem with defining the uncertainty of zero flows… it seems to me that it is clearly 0. Even if there is a discontinuity when defining the uncertainty in relative terms (but not in absolute terms)

   Author response: Firstly, recall that the way the uncertainty information is provided is: we first provide low, medium and high flow rates in mm/day (ie. discharge exceeded 90%, 50% and 10% of the time); and then relative uncertainties for each of these three discharges are provided, in the form of upper and lower bounds, provided in relative (percentage) terms. Now, to respond to your comment.

   i) If the comment refers to cases where low flow discharge itself is zero (cease to flow): as you've noted, we can't define uncertainty in relative terms if the denominator is zero. In such cases, both lower and upper bounds are marked as "n/a". You are right that the lower bound is clearly zero from a physical perspective (since flows can't be negative), but this refers to an absolute number and it is unclear how to express this as a relative change (since 0/0 is undefined).

   ii) If the comment refers to cases where low flow discharge is non-zero but the lower bound is marked as "n/a", we feel you make a good point. As per the original manuscript, if the mathematical scheme returns a negative number for the lower bound then we marked this as "n/a", but on reflection it is, in your words, "clearly zero", in which case the lower bound should be "-100%". ***In response to your comment, we have made this change [see Table 1 row 12-16, in addition to updated tables in dataset].***

7. list of attributes: I know that no list can be exhaustive. But there is a long-term seasonality index that I like very much, that addresses the dephasing between Potential Evaporation and Precipitation (the lambda in the paper of de Lavenne and Andréassian in 2018). But I can understand that you may not have any energy left after so much computation…

   Author response: Thank you very much for the suggestion. We downloaded the paper with the intention of adding the required code, but on closer inspection it seems that the suggested index suffers from some limitations when applied to catchments with high aridity. To recap, the paper (https://doi.org/10.1016/j.jhydrol.2018.01.009) suggests a seasonality index that is simply the ratio [(P∩PET)/(P∪PET)], with numerator and denominator each calculated on a monthly timestep and then summed over the temporal domain, prior to calculation of the ratio. Consider a hypothetical case where PET > P and there is exact synchronization between P and PET. In this case, numerator = P∩PET = Σ P, and denominator = P∪PET = Σ PET. Thus, the result will be Σ P/ Σ PET. It was felt that, in the case of exact synchronization, a seasonality index should have a value of 1.0, but in this hypothetical case the value is not 1.0, it is Σ P/ Σ PET. So if the aridity is, say, 2.3, the value of the seasonality index (in case of exact synchronization) will be 1/2.3 = 0.43. An aridity of 2.3 may seem high but 10% of CAMELS-AUS catchments are more arid than this. When the aridity is so high, even cases of significant dephasing between P and PET will be unduly influenced by aridity value. To conclude: in cases where the formulation is being used to *add value to equations that already consider aridity (such as the above paper), these issues may not matter. But in cases where the index may be adopted as a standalone value, the unwelcome interaction with the aridity means that the index cannot be used as a standalone measure of seasonality, and thus, we decided not to include it in the dataset. I (Keirnan) would be interested to build upon the concepts in the above paper to arrive at a measure more suited to arid environments, ideally for inclusion in later versions of CAMELS-AUS – perhaps a future discussion?* ***[note: no changes to manuscript in response to this comment]***

Anonymous Reviewer 2

This paper and associated dataset sets out to provide a large sample dataset for Austrlia, specifically one in the CAMELS series, joining others on three other continents.

I think the authors have fully delivered on this aim. CAMELS-AUS is an impressive dataset, rigorously derived as with other CAMELS initiatives and clearly articulated in the paper, which is very well-written. I believe it will be a major asset to the research community in Australia and, as the others point out, for researchers globally given the possibility of analyses across the CAMELS domains.

I like how the paper describes the unique Australian context and rationale, as well as the typical explanation of the dataset. The list of spatial/contextual metadata available across CAMELS-AUS is particularly impressive, and I think the authors have made many pragmatic decisions which will help the user community (e.g. using both major climatic datasets as both are well established).

I have only very minor comments, which will be easy to resolve, I imagine.

Finally, I apologise for the slowness of this review.

Author response: Thank you for your positive feedback and effort reviewing the article. Many of the 'minor' comments are in fact quite insightful and the feedback has improved the article.

Minor Comments:

L32. I think I know what you mean here by 'quality codes' but on its own this expression quite open to interpretation. You might want to avoid ambiguity, e.g. ('hydrometric data quality codes/flags'). By the way, the second mention at L94 is OK to me as it is clear of the context.

Author response: amended with your suggested wording, thank you ***[see line 32]****. For consistency, we have used the phrase "quality codes/flags" at later instances also* ***[lines 106, 248, 252, 284, Table 2].***

L48. For me another key benefit is that CAMELS are 'frozen', citable open datasets that are not subject to change, ensuring repeatability. Whereas going to agencies means data are always likely to be updated (e.g. reprocessed due to rating changes). There's no guarantee it will be the same data used between different studies. Of course, some users demand the most up-to-date data ad hence those sources are the most suitable in such cases, whereas for large-sample science studies the consistency and repeatability is key. Maybe just worth a line or two on this benefit?

Author response: very pertinent point, thank you. ***We have added the following text [line 48-51]: "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)."***

L80. Minor point, but I'm not sure what constitutes 'relatively deep'. Rephrase? Is it not the persistence that is so characteristic, compared to most regions?

Author response: interestingly, if you question Dr Peel on this matter, he insists that it is not the persistence/length of Australian spells but their magnitude that is different, which may be surprising (see Peel et al., 2004 https://doi.org/10.1002/joc.1041). Still, a change in wording seems reasonable. ***The sentence [lines 88-90] now reads as follows (with the new words underlined): "Another notable characteristic of Australian hydroclimatology is its tendency for multi-year spells of climatic anomalies <u>of larger magnitude</u> 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)."***

L85. Again, super minor, but what do you mean by 'model failure'? I know it's probably described in the reference but I'm interested, as many readers will be – worth a very short extra line/parenthesis about what you mean here?

Author response: no problem, ***the sentence [lines 94-97] now reads (with new words underlined): "[The Millennium Drought has] supplied many case studies of hydrological model failure (ie. <u>the high bias and low model performance in differential split sample testing reported by,</u> eg., Saft et al., 2016), which are under ongoing investigation (eg. Fowler et al., 2020b)".*** *We hope this is enough detail for the present context.*

L103 onwards. This is all very sensible, the decision to use the HRS. However, it's not universal that these CAMELS datsets (or any large sample) should be undisturbed sites. In fact, many researchers are interested in the mix of influences, and in other cases networks are deliberately chosen to represent a range of degrees of influence. E.g. in the UK where the CAMELS dataset includes many disturbed sites as they are a subject of study. There is then a separate 'benchmark' set of undisturbed sites (analogous to the HRS) which can be used and compared against the wider set with many and diverse influences. In short: it's not a necessary condition for a Large Sample datset to be free from influences – that was a decision taken in Australia but not UK, Chile etc. You could note this somewhere. This really isn't a criticism, just a clarification.

Author response: Again, an excellent point. ***We have added the following text [lines 129-134] at the end of an existing paragraph (section 2.3) that already discussed the possibilities for 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)."***

L192. Excellent line here, to humble British hydrologists like me…..

Author response: glad to hear this line was taken in the spirit in which it was written!

L200. I think I follow the process here. To an (admittedly ill-informed) outsider like me this seems like desperate measures to have to scrape this info from the website. I'm not challenging the end result, I am sure its fine, but was there no other way to obtain this from BOM (and if so possibly benefit from possibly more 'official' handover of spatial data that may even be higher res, given likely downsizing for the website?).

Author response: This would seem to be a reasonable comment, and we must be careful of our answers given the public nature of this forum. Suffice to say that there is a large difference between giving in-principle support for a project to proceed (on the one hand) and providing active and engaged assistance to said project (on the other). On a more positive note, most of the 10 catchments mentioned are quite large, so the impact of the 'blocky' boundaries is relatively small. Thus, it's unlikely that the end result would be much different if we had access to the higher resolution boundaries.

L227. Do you mean this information is not provided in CAMELS-AUS or just not provided anywhere? I appreciate it may just be how it is, but it would be good for users to find out more (somewhere) on the gap filling if they want to, so if it is available somewhere a signposting may help.

Author response: Sorry this wasn't clear. ***We have changed [line 243-245] the wording to (with new words underlined): "The gap filled variant is filled using the daily rainfall-runoff model GR4J (Perrin et al., 2001) but <u>the BOM have not published further methodological details</u> about calibration method, validation procedures, or the specifics of the interpolation method."***

L267 I imagine this is just propagating the approach in the original paper and I respect that decision if so, but this seems like a potentially misleading approach to me (thinking of how users take datasets and use them at face value). I would have thought it may have made sense to remove these cases.

Author response. We are a little unsure how to respond, and perhaps there has been a misunderstanding. We presume you're referring to the example provided, namely where Q10 is 201,400%. The main thing to note with this example is that there is no error, mathematical artifact, or anything untoward happening. The seemingly large number (201,400%) arises because a fraction is calculated with the Q10 as the denominator and, in this case, Q10 is very small (0.000023). The numerator, while also small (0.05), is three orders of magnitude larger, or five orders when expressed as a percentage. We feel it is appropriate to retain such examples since the validity of the uncertainty analysis in these cases is not compromised by the seemingly over-large numbers, and it would be a pity to lose this information. ***[note: no changes to manuscript in response to this comment. The relevant lines are 277-281]***

L288. Agreed, this seems like a sensible choice, to put this decision to users, given how established both datasets are.

Author response: Thanks for this feedback, we are glad you agree.

L304 – 310. It does seem like there is some uncertainty surrounding timing conventions and you have done the right thing in signposting this. I imagine it will not make too much difference (although I note the Jan Seibert reference) and I agree on putting this to the user rather than processing the data further. However, I just wonder whether this is also clearly highlighted within the dataset, i.e. in the metadata for these fields?

Author response: We have ***added three fields outlining the start-time of the day, in UTC, for (i) the precipitation data; (ii) the temperature data; and (iii) the streamflow data. In addition, we precede these fields with an additional field listing the state of Australia each catchment is within (since this relates to these issues). Table 3 [note, this was an error in the Author Comments, it should have said Table 1] in the manuscript is updated to reflect this.*** *We hope this is sufficient documentation to be useful to users.*

L340. Is there any issue with numbering/sub-section names? Here, in 3.6 there is a reference to the ‘following subsections’ but 3.6 has none. 3.7 does so I assume just numbering errors.

Author response: There was indeed an error (3.7 should have been 3.6.1), which has now been fixed. Thank you for bringing this to our attention ***[Amended at line 381]***

L416. Given the catchments are all HRS, at face value there’s a tension here (given that Fig 3 shows some catchments score highly on this attribute with some yellowish dots implying high degrees of disturbance). I guess these are just two different definitions of ‘anthropogenic influence’, and this distinction is fine. However it may be worth being really explicit about that at the start of this section to avoid ambiguity and incorrect interpretations (L424 goes on to note the lack of regulation in the dataset but you could be clearer early on here).

Author response: Happy to ***include an extra sentence at the start of this sub-section [line 443-445] to try to avoid confusion: “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).”***

L457. Given CAMELS inherits from HRS, this 'first' is not strictly true in a wide sense, even if true in tehe narrower sense of as a dataset specifically for large-sample analysis with associated metadata etc.

Author response: This is true, so we have ***changed this sentence [line 484-485] to now read: “[CAMELS-AUS is] the first large sample hydrology dataset for Australia to include data on climatic forcing, catchment attributes, and gauging uncertainty”.***

Comment from Jan Seibert

First of all, thanks to the authors for making this valuable data set available!

Author response: Thanks Jan, our pleasure!

Secondly, I want to make two points related to a specific aspect of the data set, namely the definition of hydrological and climatological days. I appreciate that the authors raise this issue (section 3.5.2, P11, L293):

1) Honestly, I never thought that this could have much of an effect until our former PhD student Eduardo Reynolds demonstrated convincingly that I was wrong – see https://doi.org/10.1080/02626667.2018.1451646

Author response: thanks for this reference – very interesting, and ***now referenced in the manuscript [at line 318].***

2) As I understand, for some catchments the hydrological day is defined differently from 9am-9am. However, I am a bit unsure whether you in these cases adapted the climatological day? I fully agree that modellers should be aware of this as you write; to make this easier perhaps the definition of the hydro-day could be given as info in the attribute table (Table 3)

Author response: Regarding the first part of your point, sorry that this was unclear. We have ***considerably expanded the relevant section, due in part to Reviewer 1's question about timezones, and it now contains [at line 335-339] the text: "Unfortunately, these limitations … 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 this data despite these limitations."*** *Regarding the second part of your point, as noted above* ***we have added three fields outlining the start-time of the day, in UTC, for (i) the precipitation data; (ii) the temperature data; and (iii) the streamflow data. In addition, we precede these fields with an additional field listing the state of Australia each catchment is within (since this relates to these issues). Table 3 [note, this was an error in the Author Comments, it should have said Table 1] in the manuscript is updated to reflect this.*** *We hope this is sufficient documentation to be useful to users.*
