Reviewer comments, response and actions for manuscript (updated with finalised responses column)

A systematic evaluation of 15 actual evapotranspiration formulations within conceptual hydrological models

Co-authors: Gabrielle Burns, Keirnan Fowler, Murray Peel, Clare Stephens

Reviewer 1	Response	Action	Finalised response
Major comments			
I missed any discussion on the spatial transferability of the calibrated parameters. I understand that this methodology is presented as site-specific. IN the line of Klemes et al., 1986 (doi 10.1080/02626668609491024), to assess robustness of the most suitable AET formula, I would also definitely welcome an evaluation of the selected parameters in ungauges locations (apply the 1calibrated parameter set to the 6 remaining basins) and objectively assess, whether the selected AET formula also holds across other locations as well. This could enhance the temporal split-sample test, already presented.	We thank the reviewer for raising the important issue of parameter transferability and for referring to the validation framework of Klemeš (1986). Following the logic of Klemeš, the proposed proxy-basin test is valuable when the objective is prediction in ungauged basins using a fixed model structure and transferable catchment-scale parameters. We would first like to clarify that the split-sample testing presented in this study refers to temporal split-sample validation, rather than a spatial or proxy-basin test. We will revise the manuscript to explicitly state this (i.e. temporal split-sample testing) to avoid any ambiguity and to make clear that spatial transferability is not assessed here. While we agree that a proxy-basin experiment can provide useful insights, the spatial transferability of conceptual models is not our	We consider spatial transferability and proxy-basin testing to be a valuable but distinct line of inquiry, particularly relevant for studies explicitly targeting regionalisation or prediction in ungauged basins. We will acknowledge this more clearly in the discussion as an important avenue for future work, but that it lies beyond the intended scope of the present manuscript.	Added information at lines Methods: 287 291 Discussion: 511 557

Reviewer 1	Response	Action	Finalised response
	focus. As per our existing abstract, our focus is temporal transferability ("models [may] incorrectly simulate long-term changes in partitioning between AET and streamflow"). Spatial and temporal transferability are distinct challenges due to the known site-dependence of evaporative and vegetative processes. This means that poor spatial transferability of calibrated parameters would not necessarily indicate a lack of robustness of a given AET formulation but rather reflect differences in vegetation across space. Thus, there is little value in testing this transferability without an explicit framework to handle these spatial differences. Such a framework would bloat the paper, which is already long (30 pages, 9 figures) and is quite complex already (3x model structures, 15x equations and 7 catchments).		
When focusing on AET, the current formulation does not consider any uncertainty in potential evapotranspiration estimates. The authors should consider their in the discussion, see e.g. studies like Oudin et al, 2005 (JoH), Pimentel et al, 2023 (WRR); Thakur et al,	We agree that PET formulation and precipitation uncertainty are important contributors to uncertainty in hydrological simulations. In this study, the decision to adopt a single PET formulation was intentional, as our aim was to isolate the effects of alternative AET formulations rather	Add in to discussion PET uncertainties & omission of testing different PETs. Example: "An important source of uncertainty not explicitly explored in this study is the estimation of potential evapotranspiration	Added information at lines Discussion: 535

Reviewer 1	Response	Action	Finalised response
2025 (HESS); which presents sometime large differences of employed PET methods on hydrological components from different perspectives, besides uncertainty in the precipitation estimates, which is also a crucial element for the water balance closure.	than to explore the full range of forcing uncertainty. Nevertheless, we agree that the implications of PET formulation uncertainty should be acknowledged more explicitly in the manuscript, and we will expand the discussion accordingly. In addition, to directly address this comment, we have included a targeted sensitivity analysis examining the effect of an alternative PET formulation on one of the key diagnostic results. Specifically, we reproduced Figure 8 using an alternative PET forcing (Morton's Point Potential), with all three models re-calibrated across the seven catchments. This additional figure will be provided in the Supplementary Materials and referenced in the revised discussion.	(PET). Numerous studies have demonstrated that different PET formulations can produce different estimates of atmospheric evaporative demand, with implications for simulated evapotranspiration, soil moisture, and runoff (e.g. Oudin et al., 2005; Pimentel et al., 2023; Thakur et al., 2025). In the present analysis, PET formulation was held constant across all experiments in order to isolate the effects of model AET formulation specifically. One additional analysis in reference to this issue was investigated by not reproducing all the results for a different PET formulation, but rather we reproduced Figure 8, which highlights the systematic tendency for simulated AET to peak earlier in the season than observed, using an alternative PET input (Morton's Point Potential). For this experiment, all three models were re-calibrated across the seven catchments using the alternative PET forcing. While some catchment-specific differences in the magnitude of	

Reviewer 1	Response	Action	Finalised response
		the response were observed (with slightly stronger or weaker signals), the overall patterns and conclusions remained unchanged. The timing bias in simulated AET persisted across PET formulations, indicating that the early seasonal peak identified in the original analysis is not a consequence of the choice of PET input. Given the overall outcomes were relatively robust despite the very different PET formulation, we feel this is a sufficient response to this valid concern.	
It is not very clear, at this temporal resolution the study was done. Daily? I missed information about the length of the calibration periods in the Methods. This should be provided around Lines 241-247, to provide more details on the split-sample testing.	We thank the reviewer for noting this lack of clarity. The simulations were performed at a daily time step, which was not stated explicitly in the Methods and will be added for clarity. We also agree that additional detail on the calibration and split-sample periods should be provided in the Methods section around Lines 241–247. The length of the calibration periods differs among sites, reflecting the varying availability of flux tower observations. While this information is currently described in the Supplementary Material, we	Add detail about flux tower time lengths for split sample test around lines 241-247. Also add in that it was a daily time step in the introduction.	Line 83 Line 291

Reviewer 1	Response	Action	Finalised response
	will add a brief description in the Methods and clearly reference the Supplementary Table where site-specific calibration periods are reported. As noted in the Results (Lines 328–329), the relatively short duration of flux tower records at some sites may limit the ability of the split-sample test to fully capture long-term model behaviour. We will ensure this limitation is clearly introduced earlier in the Methods when describing the split-sample testing procedure.		
Minor comments			
Lines 28-29: be careful, soil moisture is not internal model's flux, please, reformulate sentence accordingly, soil moisture is a state!	We thank the reviewer for this clarification and agree. The sentence will be revised to distinguish between simulated state variables (e.g. soil moisture) and fluxes (e.g. groundwater recharge, AET).	Revise wording accordingly.	Line 27
First paragraph of Introduction: please, say more clearly, which kind of hydrological models you have in mind (conceptual, black box, spatially distributed, etc, there are many subcategories).	We agree that additional clarity is helpful. While this study focuses on conceptual rainfall–runoff models, the statements in the opening paragraph are intended to apply more broadly to hydrological models in general. We will revise the text to make this distinction explicit.	Clarify model classes in Introduction.	Line 37

Reviewer 1	Response	Action	Finalised response
Equation (4) should be written in full in the Methods, providing a reference to original work is not enough here.	Yes agreed, the full function is: E = E_of - 5 * abs(ln(1 + B)) ^ 2.5 where B = mean(sim - obs) / mean(obs) We agree and will include the full formulation of the objective function in the Methods.	Add full equation and definitions.	Adjusted, section 2.6 Calibration
Line 30: remove "perhaps" and rephrase.	We agree and will rephrase the sentence to avoid speculative language while retaining the intended meaning.	Reword sentence.	Line 31
Line 31: instead of "culmination", consider using "integral variable"	We appreciate the suggestion and will revise the wording to better reflect that streamflow integrates multiple catchment processes.	Revise wording.	Line 31
Consider merging the last two paragraphs of Intro, and move the detailed MARRMOT descriptions to the Methods, where it fits better.	We agree that this would improve structure and readability and will revise accordingly. We will keep the aim in the introductory section, and move the remaining material in these paragraphs to the Methods, as suggested.	Reorganise Introduction and Methods.	Moved to methods – line 107
Line 140: you don't say, which methods was used to calculate PET. This is important. Also, in section 2.3, information on the discharge is missing. Additionally, here is the right place to introduce length of the timeseries, number of years, and say, how did you treat	PET is calculated using Morton's wet-environment formulation as provided in the CAMELS dataset; we will make this explicit. Streamflow data are also sourced from CAMELS and will be clearly stated. We will add details on the temporal coverage of the datasets	Include sentence explicitly stating that streamflow data is from CAMELS and PET is Morton's Wet environment. Add details that say the CAMELS dataset is from 01-Jan-1950 to 26-May-2022, and that missing values were ignored. The dates of	Added line 168

Reviewer 1	Response	Action	Finalised response
missing values in your analysis.	and note that missing values were excluded from the analysis.	the flux tower timeseries can be seen in the supplementary materials table S5.	
Line 268: equation 19 should be connected to the AET definition, Current statements looks like it refers to eq. 19, which is not defined in the manuscript.	Noted. References to AET equations will be updated to use consistent naming aligned with the tables (e.g. evap_19).	Update equation references throughout	Referenced throughout as a variant of (dependent on context): • AET equation evap_19 • Equation evap_19 • Evap_19
Table 4: instead of having only categorical values for which AET method was best, it would be also valuable to have some qualitative category.	We note this suggestion. Quantitative performance metrics are already provided in the Supplementary Material. Can the reviewer please provide feedback as to whether this is what is meant?	We will add text to make it clearer that this quantitative information is in the Supplementary Material. This will be added in the caption of Table 4 and associated text.	Updated table 4 caption (line 323)
Figure 4: Consider different colors for native equation: now, every model's native equation is shown by black star, with overlaying colored circle. Considering coloring the star by model's color. Also, I recommend not to use 4 columns, rather 3 columns and then 3 rows. This would allow your figures to become bigger	We agree and will revise the figure accordingly, including exploring an alternative subplot layout to improve readability.	Update figure to have the stars as the model colours Adjust composition of subplots to better show as per recommendation.	Updated figure 4.

Reviewer 1	Response	Action	Finalised response
Figure 6: I would recommend to keep same model's colors as used earlier in Figure 4.	We thank the reviewer for highlighting this inconsistency. In Figure 6, the Simhyd model should indeed be shown in green, consistent with the colour scheme used elsewhere in the manuscript (e.g. Figure 4). We have corrected this in the revised figure and updated the caption accordingly.	Update Figure 6 to ensure consistent colouring with the rest of the manuscript.	Update all of the signature figures

Reviewer 2	Response	Action	Finalised response
Major comments			
Inclusion of AET and catchment correction factors pr. catchment Table 1.	We agree and thank the reviewer for this suggestion. Including catchment-level AET and correction factors will improve transparency regarding the flux-tower adjustments applied prior to calibration.	Update Table 1 to include catchment correction factors	Table 1 updated (see more detail below)
More details on the objective function for discharge and weighting is needed	We agree that additional detail is required. As also noted by Reviewer 1, we will include the full formulation of the discharge objective function in the Methods.	Expand Methods section describing objective functions and weighting.	See more details below
More discussions and potentially analysis of spatial parameter transferability and regionalization	We acknowledge the importance of spatial transferability and regionalisation. However, as clarified in response to Reviewer 1, the objective of this study is not prediction in ungauged basins, but	Add clarification in Discussion.	See details below.

Reviewer 2	Response	Action	Finalised response
	rather the diagnostic evaluation of model AET formulations using site-specific calibration. We believe this comment is expanded on in the final comment below, where we have addressed this in more detail. In summary, we will strengthen the discussion to explicitly acknowledge regionalisation and spatial transferability as important directions for future work, particularly in the context of explicit vegetation representation.		
Tables are not well organized	We thank the reviewer for the feedback. The tables currently follow HESS formatting guidelines, but we will revisit their organisation and presentation to improve readability where possible.	Review table layout and clarity.	See details below.
Minor comments			
Line 9: correct spelling ”partitioniapendng”	Noted. The typo will be corrected.	Correct typo in line 9.	Corrected
Lines 37-49: This section talks about dynamic vegetation modelling, but this is not covered in the manuscript, which is based on simple	We agree that this section places disproportionate emphasis on dynamic vegetation modelling relative to the scope of the paper. We will revise this section to better	Revise Introduction to better align scope.	Adjusted introduction to reduce emphasis, lines 46-54

Reviewer 2	Response	Action	Finalised response
conceptual hydrological models. It seems disproportional and a bit misguiding to dedicate such attention to a topic, which is not part of the paper.	align with the use of conceptual hydrological models and refocus the motivation on evapotranspiration process representation rather than explicit vegetation dynamics.		
Lines 118-127: These selection criteria seem a bit subjective. Here I had written a long section about the need to adjust the site flux to catchment water balances. However, you actually do that in section 2.4, which is great. I strongly suggest that you mention that catchment water balance adjustment already here in lines 118-127. In that way you can mention both careful site selection and the need for adjustment.	We thank the reviewer for this helpful suggestion. We will add a sentence in Section 2.1 noting that, in addition to the site selection criteria, a catchment-scale water balance adjustment was applied, with details provided in Section 2.4.	Add in to section 2.1 a sentence stating that in addition to these criteria, we also applied a catchment water balance adjustment, which is discussed in detail in section 2.4.	Added line 105
Table 1: I strongly suggest that Table 1. Includes catchment mean daily Q and site mean daily AET from the flux-tower and the adjustment factors to illustrate the adjustments made. If the table gets large, I think you can skip the catchment code, PET, the	We agree. Including mean daily streamflow, mean daily flux-tower AET, and adjustment factors will better illustrate the applied corrections. We will revise Table 1 accordingly, simplifying other columns where necessary and reducing numerical precision.	Update table 1 to include: - Mean daily AET (from flux tower) - Mean daily streamflow - Catchment adjustment	Table 1 updated to include - Mean daily AET (from flux tower) - Mean daily streamflow - Catchment adjustment Removed (and moved details to supp materials) - Mean daily PET (mm/d) - Australian State or territory

Reviewer 2	Response	Action	Finalised response
state and High P Low P columns. Here I believe the AET and correction is more interesting than the PET. Also number of decimals are excessive, and flux tower sit description could be simplified.			- Season high P / low P Removed decimal places, simplified site descriptions and improved table spacing
Line 179: Why not substitute the AET formulation across all storages? Why only for the storage that a priori is assumed to contribute the most?	We thank the reviewer for this question. Applying distinct AET formulations to all storages would require the introduction of additional parameters for each storage. For example, in the VIC framework this would increase the number of calibrated parameters from 10 (or 12 in the current framework) to approximately 16. Given these extra dimensions, calibration would become more difficult, increasing the likelihood of a poor result due to failure to converge on the "optimum" solution. In other words, the higher dimensionality could cause problems with attribution of results. Further issues may arise with intercomparability issues. The	Expand methods to explicitly justify why alternative AET formulations were applied only to the dominant soil moisture storage. Specifically, explain (i) the calibration challenges and increased parameter dimensionality that would arise from modifying AET formulations across all storages, (ii) the importance of maintaining intercomparability across different model structures, and (iii) process-based reasons for not altering AET formulations for storages representing distinct physical processes (e.g. interception).	Added into lines 204-211

Reviewer 2	Response	Action	Finalised response
	benefit of the current design is that the consistency of changes across model structure allows us to generalise, thus leading to recommendations likely to apply across all model structures. If we introduce the changes to a different number of storages in each model structure, we lose the intercomparability. Lastly, there are process considerations. Some storages, such as interception, represent fundamentally different physical processes compared to soil moisture stores. Modifying the AET formulation for these storages, as suggested, is inconsistent with their role to represent direct evaporation from wetted canopy surfaces. The chosen approach represents a pragmatic compromise that allows differences attributable to AET formulation to be isolated while maintaining model interpretability and conceptual consistency. We		

Reviewer 2	Response	Action	Finalised response
	will clarify this rationale in the Methods.		
Line 219: I think we need some more details on this bias_penalising_log OF. I have not heard of it before, please write the equation. Also, it cannot be found in Trotter et al. (2022) (Eqn. 4), as stated.	We thank the reviewer for pointing out the incorrect citation. We will include the full mathematical formulation of the bias-penalising log objective function in the Methods and correct the citation accordingly, consistent with the response to Reviewer 1.	Add equation and correct citation (Trotter et al., 2023, Eqn 1). https://doi.org/ 10.1029/2021WR031845	Adjusted, line 240 and 245
Table 2 is hard to read, please try to reorganize.	We will revisit the formatting of Table 2 to improve readability while remaining consistent with journal guidelines.	Reformat Table.	Reformatted.
Table 3 is not important, consider moving to supplement.	We agree and will move Table 3 to the Supplementary Material.	Move Table 3 to supplement.	Moved to supplement.
Table 4, Why not in a similar format as S2 in the supplement, much easier to read.	We thank the reviewer for the suggestion. We will revise Table 4 to more closely match the clearer structure of Table S2, while remaining compliant with journal formatting constraints (no colour).	Revise Table 4 layout.	We agree that the format used in Table S2 is effective for visual comparison. However, replicating this format in the main manuscript is constrained by journal formatting limitations (e.g., use of colour and table structure). Additionally, Table 4 presents objective function values

Reviewer 2	Response	Action	Finalised response
			(OFVs) across multiple catchments, which are not directly comparable due to substantial differences in magnitude between sites. As such, adopting a ranking- or colour-based format similar to Table S2 could be misleading in the main text. For these reasons, we have retained the current format for Table 4, while keeping the alternative presentation in the supplementary material for clarity and reference.
Lines: 225-227: The two OF's are weighted equally, but how does that ensure that they contribute equally to the combined OF? What if the two OF's have different magnitudes, e.g. the KGE OF will be below 1, what is the typical range of bias_penalising_log? From later plots e.g. Fig 4 and 7 it seems as if the OF's are scaled to similar sizes, but please elaborate a bit on that.	We thank the reviewer for raising this important point regarding the combination of the two objective functions. We agree that equal weighting alone does not guarantee equal contribution unless the objective functions are formulated on comparable numerical scales, and we acknowledge that this was not explained sufficiently clearly in the original manuscript. In the current framework, both objective functions are KGE-	Revise the Methods section to clarify that both the streamflow and AET objective functions are KGE-derived metrics with comparable numerical ranges. This now explicitly describes the formulation and typical magnitude of each objective function and explains why equal weighting leads to approximately equal contribution to the combined objective function. Example: "The calibration approach was based on single-objective optimisation using a composite	Adjusted section 2.6 Calibration to reflect these changes

Reviewer 2	Response	Action	Finalised response
	derived metrics. The streamflow objective (referred to as the bias-penalised log objective) is itself a modified KGE formulation that penalises bias separately for high- and low-flow regimes (Trotter et al., 2023), while the AET objective is the KGE applied to square-root–transformed data. As a result, both objective functions are bounded and exhibit similar numerical ranges during calibration. As illustrated in Figures 4 and 7, the two objective functions attain comparable magnitudes across catchments and models, such that equal weighting results in approximately equal influence on the combined objective function. We have revised the Methods section to explicitly clarify the KGE-based nature, typical ranges, and scaling of both objective functions, and to explain why equal weighting is appropriate in this context.	objective function that equally weighted model performance against streamflow observations from CAMELS-AUS and AET observations from flux towers. For streamflow, a Kling–Gupta Efficiency (KGE)-derived objective function was used that incorporates separate penalties for bias in high- and low-flow regimes (hereafter referred to as the bias-penalised KGE; Trotter et al., 2023). For AET, the KGE was applied to square-root–transformed data following Gardiya Weligamage et al. (2025). Both objective functions are bounded KGE-based metrics and therefore exhibit comparable numerical ranges during calibration. As a result, equal weighting of the two objective function values leads to approximately equal contribution to the combined objective function value (OFV), which was computed as the mean of the	

Reviewer 2	Response	Action	Finalised response
		streamflow and AET objective values (Eqn. 6)."	
Lines 243-244: Great with such a split sample test, which I believe is important for this kind of exercise. However, why only do it for one AET equation? Would there be value in adding the split sample test for all equation in the supplement? I would be curious to see if there is a systematic difference between models and equations regarding overfitting and tradeoffs between Q and AET?	We agree that applying the split-sample test to all AET equations could provide additional insight. However, the purpose of the split-sample test here is to provide a final robustness check for the best-performing formulation, rather than a systematic assessment of overfitting across all equations. Extending this analysis to all formulations would substantially increase complexity without directly serving the main study objective. We will clarify this rationale in the text.	Clarify rationale.	Added sentence to lines 280-282
Line 251: "In this figure, Simhyd was run with each of the 15 evapotranspiration (AET)". Please make it clear if this is the result after recalibration to each AET equation.	Yes, all results shown correspond to independent recalibration for each AET equation. We will state this explicitly in the figure description.	Update line 251: In this figure, Simhyd was run calibrated with each of the 15 evapotranspiration (AET) equations consecutively, and all other model components were held constant.	Adjusted line 294
Lines 252: I assume it is the adjusted not the observed flux tower AET.	We agree and will specify that the adjusted flux-tower AET is shown.	Update line 252: the observed flux tower AET (catchment adjusted) is shown as a thick blue line,	Adjusted line 296

Reviewer 2	Response	Action	Finalised response
Line 255: Also here, that depends on whether this is prior or post recalibration. I think that has to be very clear.	Noted. We will ensure this is stated explicitly.	Clarify text.	Adjusted first paragraph in section 3.1 to make this clear
Line 303: I think it is wise to place the other signature results in the supplement. However, I notice one thing for the Interannual variability signature Figure S4.2. It seems as if the pattern here is almost identical across all models indicating that this signature is largely a result of the precipitation data and not controlled by the models or equations.	We thank the reviewer for this insightful observation. We will explicitly note in the Results that the similarity across models suggests this signature is largely driven by precipitation rather than model structure or AET formulation.	Add note to Results.	Added to line 347-349
Line 335: I really appreciate the split-sample test and the split between Q and AET OF's. However, I also feel that with this calibration to a single optimal parameterset, it can be difficult to interpret the possible overfitting. It seems as if Eq. 19 fits very well to AET, but performance drops for Q, could that indicate an overfitting to AET? Perhaps in the discussion or under a	We agree and thank the reviewer for this valuable suggestion. We will expand the Discussion to acknowledge that results are based on a single optimal parameter set and that alternative near-optimal parameter sets may exhibit different trade-offs between Q and AET performance and transferability.	Add limitation discussion on parameter uncertainty.	Added to discussion at lines 565-572

Reviewer 2	Response	Action	Finalised response
description of limitations also mention that all results are based on one unique optimal parameter set and there will be a range of alternative parameter sets that would perform almost equally well, but with different transferability or tradeoffs. In principle the whole analysis could have been presented as results of an ensemble of parameter sets. I am not suggesting to change that, but perhaps worth noting something on this in the discussion.			
Line 337: for the analysis under "3.4 Seasonal timing of AET" you recalibrate to AET alone to test if the model has the capacity to match better. But would a recalibration including this seasonal timing OF (e.g. in combination with the original AET OF) not be a better test of that?	The recalibration to AET alone was designed to isolate the influence of streamflow constraints on seasonal AET timing, while retaining comparability with the original objective function components. Introducing an additional seasonal timing objective would complicate interpretation and comparability. We will clarify this rationale in the Methods or Discussion.	Clarify rationale in text.	Added explanation at line 399

Reviewer 2	Response	Action	Finalised response
Figure 8: I would suggest having the same y-axis scale on all three plots.	We agree and will apply a consistent y-axis scale across all panels.	Change figure 8 to have same y-axis scale on the three sub plots.	
4.3 Implications: Overall, a very nice analysis and great that you take the time in the end to analyze eq. 19 in more depth. One remaining question, that I feel is not addressed, is the possibility to regionalize and thereby parametrize the two extra parameters in eq 19 beyond the current catchments. There is a large spread in those parameters (Table S6). So, there is not a unique parametrization across catchments. I think it would be great to address this, and in the discussion point in some direction regarding this, perhaps under future work. E.g. would the optimal values of p1 and p2 be spatially related to any mappable information? Also, I assume the PET used is a purely meteorologically driven PET, would a spatial scaling, using something like a high-	We thank the reviewer for this insightful comment and for highlighting the broader implications of introducing additional parameters in equation 19. We agree that the observed spread in the calibrated values of p1 and p2 (Table S6) indicates that a single, catchment-invariant parametrisation is unlikely, and that questions of regionalisation and transferability are important considerations. In the present study, p1 and p2 are treated as effective parameters that compensate for unresolved controls on evapotranspiration which are not explicitly represented in the conceptual model framework. As such, their variability across catchments is expected and not interpreted here as a limitation of equation 19, but rather as an indication of site-specific behaviour.	Update limitations section to reflect this: "The additional parameters introduced in equation 19 exhibit substantial variability across catchments (Table S6), suggesting that a single, spatially uniform parametrisation is unlikely. In this study, these parameters are interpreted as effective representations of unresolved controls on evapotranspiration. Their site-specific nature therefore reflects differences in local evaporative regimes rather than parameter uncertainty alone. An important direction for future research is to assess whether these parameters can be regionalised using mappable catchment attributes, such as aridity indices, vegetation functional types, or remotely	Added into lines 499-509

Reviewer 2	Response	Action	Finalised response
resolution crop coefficient approach level out some of the variability between catchments, e.g. the need to reduce AET below PET also under high moisture availability? And what is the impact of PET method used? Could some model or equations struggle with scaling a poor PET input?	We agree that exploring whether these parameters could be related to mappable catchment characteristics (e.g. climate regime, aridity, vegetation type or phenology) is an important avenue for future research. Similarly, incorporating spatially varying PET scaling approaches (e.g. crop coefficients or vegetation-adjusted PET) may help reduce the need for empirical suppression of AET under high moisture availability, although this would require explicit representation of vegetation dynamics beyond the scope of the current study. We are happy to expand the Implications section to more clearly discuss these points and to outline directions for future work.	sensed indicators of canopy dynamics.	

Editor	Finalised response
Line 220: Could you please add the original reference to the KGE here?	Added at line 254.

In addition to the reviewers' comments, the symbol "of_KGE_sqrt" in Eq. (5) does not represent an established mathematical operator. Could you please rephrase it based on traditional mathematical notation?	Rephrased and adjusted equation in section 2.6 Calibration
I may have overlooked this, but could you please ensure that information on the length of the time series is available for each location, and indicate which years are covered?	Made clearer line 163, added to table S1.