Response to anonymous referee #2 on HESS egusphere-2025-3373 '*Using evapotranspiration signatures to assess evapotranspiration realism in rainfall-runoff models*'

Citation: https://doi.org/10.5194/egusphere-2025-3373-RC2

#	Remarks to author	Authors' response
General comment		
	This paper by Gardiya Weligamage et al. deals with the issue of evapotranspiration in conceptual rainfall-runoff models. This is a very important topic for the community of hydrological modelling. Evapotranspiration is in general overlooked, although it represents a significant (and sometimes dominant) term of catchment water balance. To this date, only very few studies have proposed a proper evaluation of evapotranspiration in models, so the approach described here is extremely promising. The paper is also well organised and written very clearly. A companion paper, also submitted to HESS, presents the AET signatures in more details, but enough information is given here so that the present paper can be read independently, which is appreciated.	Thank you for your constructive comments on the manuscript. We find your suggestions are valuable for strengthening the manuscript. We are happy to improve the manuscript in response to the comments you made below. Please find our responses below.
Comments		
1	1/ However, given the title of the paper and the main part of the introduction, I was expecting a study on the evaluation of realism of AET in the tested models, containing interpretations and conclusions about these models and their structures. Unfortunately, apart from brief comments here and there, nothing such appears in the paper, that turns out to be a simple (though very carefully thought-out and well executed) comparison study of two calibration methods (with streamflow only of with	Yes, we agree. Thank you for this descriptive comment. We agree that discussing model structures and their representation of AET, beyond comparing the two calibration approaches would strengthen the interpretation of the results in terms of model behaviour and the underlying model structures. We will particularly improve the sections you have highlighted below, as listed in our individual comments as follows.

	streamflow + AET data), which is much more limited in terms of interest and impact. This is really regrettable, and I think it would not be a big amount of work to add these elements. In particular:	
	- Methods section: when describing the models used (p7, l 142) and mentioning their "various complexities with various evapotranspiration representations": describe them + add a table to sum up how the calculations of AET are made in the models.	We will add a summary table describing model structures and AET representation of the selected models. Also, we will guide readers to appendix of the MARRMoT toolbox manual (used in this study for the modelling), where schematic visualization of the mode structures is provided.
	- Results section : develop comments and interpretation according to model structures and AET calculations, for example l 250, l 290, l 319-320	Similarly, as you suggested we will include comments and interpretations related to model structure and AET representation in the selected models in Results section.
	- Add a full paragraph in the Discussion section	Also, we will discuss the results related to model structures, relating this back to the earlier section on structures, and further discuss AET calculations by the equations in different models.
2	2/ Another discussion point that seems overlooked is the interest of the AET signatures for model assessment. Some of the signatures seem to be not discriminant between models for example, which raises questions about their added value (see Fig 11: annual median, water stress in particular). I	Yes, we agree Thank you for this important comment. We agree that the differing diagnostic value of the AET signatures deserves further discussion. We will discuss the suitability of different AET signatures in rainfall-runoff model

	believe this point could be developed in the paper as well.	assessment in the manuscript. We note that other models not tested here may show deficiencies in such signatures, so (to some extent) the signature selection should be guided by the question of which dynamics may be important to common modelling contexts.
3	3/ Figures could also be improved. A lot of them (Figs 3 to 10) are very similar and almost indistinguishable. Some of them don't seem to bring a lot of information and could probably be transferred to supplementary material (Figs 4, 5, 6). For all of them, the symbols for the different models are indistinguishable, which limits the interpretation of AET deficiencies (eg in Fig 6: there are comments about some of the models but they can't really be seen on the graphs). On the other hand, the colours for the catchments don't seem to be very useful, as they are almost never discussed / commented. And finally, as the point of interest is the comparison between two calibration methods, maybe just merging both graphs into one representing the difference between the calibration methods would be more synthetic and straightforward? These remarks apply also to Figure 12.	Thank you for these helpful suggestions regarding the presentation of the figures. Firstly, we apologise that the quality of the figures seems to have degraded in the pre-print process, and we will endeavour to manage this more closely should the manuscript go for final proofing. This drop in quality may be a factor in this comment. For example, we believe that the symbols denoting models will be more distinguishable in this case. Admittedly in some cases the difficulty is due to overlap--a problem that would only be exacerbated if we combined the two into one as suggested--thus, in the interests of clarity, we will not be doing this combination. We take the reviewer's point that we do not sufficiently discuss the colours denoting catchment, and will add extra text throughout to ensure the manuscript leverages this information.
Additional comments		
4	4/ additional remarks - I am surprised by the use of a single flux tower for a 4800 km2 catchment (p3, l90). There are some considerations about the spatial representativeness of the flux tower data (section 2.1 + supplementary material), but they address a slightly	Yes, we agree. Thank you for this insightful comment. We agree that the representativeness of flux tower AET depends not only on the distance between the flux tower and the catchment, but also on the similarity of their climate

	different question as they present a comparison between nearest neighbour flux tower data and remotely sensed data, without consideration for the catchments themselves (no mention of the catchments can be found in Table S1). There is some important information missing, in particular about the climate characteristics of the catchments (climate zones such as Koppen-Geiger as mentioned in Table S1 or any relevant classification), and the dominant vegetation. Representativeness of the flux tower data over such large distances depends on the spatial variability of climate and vegetation in the study area. This is important for the potential generalization of the method. How would it apply to another part of the world?	characteristics and dominant vegetation. These factors were considered when formulating this research and selecting the study sites. Similarly, in the supplementary analysis evaluating nearest-neighbour flux tower data, we deliberately selected comparison sites with broadly similar climate and vegetation characteristics so that spatial separation was the primary factor influencing representativeness. We acknowledge that these considerations were not described clearly in Section 2.1 or the Supplementary Material. We will revise the manuscript to explicitly describe the climate characteristics and dominant vegetation of the selected catchments and flux towers, clarify the criteria used to pair flux towers with catchments, and discuss the implications for the general applicability of the proposed methodology.
	- PET is never properly defined (I don't think that the acronym is even developed once). I don't think that readers can take the definition for granted, as there is a bit of confusion around this term with several definitions. For example Allen et al (1998) in the FAO reference define ET0 (= reference evapotranspiration) rather than PET. Is it the same?	Thank you for highlighting this. We will include the Morton's wet environment PET equation in Section 2.4 Data, where the PET source is described. $M_{wet} = b_1 + b_2 \left(1 + \frac{\gamma_p}{\Delta_p}\right)^{-1} R_{TP}$ where, b1, b2 – Empirical constants, Δ_p – slope of saturation vapour pressure at equilibrium temperature (TP), R_{TP} – net radiation for land surface at TP

	- performance of models 3.1 + Fig 2: Are the improvement/ degradation of performance statistically significant? Or is it only visual?	Thank you for this comment. Since the primary objective of this manuscript is to assess the ability of rainfall–runoff models to reproduce AET dynamics using evapotranspiration signatures, we did not assess the statistical significance of the improvements or degradations in streamflow and AET performance calculated using traditional metrics shown in Fig. 2. This reflects that statistical significance is usually assessed in cases where, for example, a sample is sufficient for an inference, whereas in the case where you have the same forcing data in such a controlled experiment, the concept is perhaps less relevant. So yes, the purpose of this section is to provide a visual comparison of model performance using conventional evaluation metrics, rather than to formally test differences between calibration methods or model performances.