Response to reviewer 01 – Manuscript: *"Combining uncertainty quantification and entropy-inspired concepts into a single objective function for rainfall-runoff model calibration"* by Pizarro, Montanari & Koutsoyiannis

Below, we list reviewer's comments verbatim in **bold**, followed by responses to these comments in blue. We *italicise* the revised additions to the manuscript.

Thank you again for your helpful feedback!

Reviewer 01: Keith Beven

Obs. 1: This is a really nice paper but needs a few issues of clarification in both the introduction and in the presentation of the methodology before publication in line with the comments in the MSS.
Ans. 1: Thank you for your comments and suggestions. Introduction and Methods sections were revised and improved accordingly.

Obs. 2: Line 24, "…treatment of error sources (see, e.g., Blazkova and Beven 2002; 2004; Krzysztofowicz 2002)…", highlighting "Krzysztofowicz 2002": Beven 2009 Environmental Modelling - An Uncertain Future provides a more general review
Ans. 2: Cited works in Line 24 were updated, incorporating Beven (2018).

References:
Beven, K. (2018). *"Environmental modelling: an uncertain future?"*. CRC press.

Obs. 3: Line 31, "…understandable to end users (Beven, 2024)…", highlighting "Beven, 2024": perhaps mention recent Vrugt attempts to provide more flexible likelihood functions (albeit with more parameters)
Ans. 3: Thank you for the suggestion. We incorporated Vrugt et al. (2022) alongside Beven (2024).

References:
Beven, K. (2024). *"A brief history of information and disinformation in hydrological data and the impact on the evaluation of hydrological models"*. Hydrological Sciences Journal, 69(5), 519-527.

Vrugt, J. A., de Oliveira, D. Y., Schoups, G., & Diks, C. G. (2022). *"On the use of distribution-adaptive likelihood functions: Generalized and universal likelihood functions, scoring rules and multi-criteria ranking"*. Journal of Hydrology, 615, 128542.

Obs. 4: Line 51, "…possibly statistically incoherent and potentially unreliable parameter and predictive distributions…": This is misleading - you know very well that incoherence argument could only be made in respect of ideal cases when the likelihood is known - and that GLUE could also be used with the correct likleihood to give the same result if that information is known (Beven et al. 2008). But real cases are non-ideal in that respect - as the nonsense that can come from applying formal likelihoods shows (Beven and Smith 2015).
Ans. 4: We shared the reviewer's vision in terms of real cases are non-ideal and, therefore, uncertainty quantification is a must being the core of RUMI formulation. The sentence in question was modified and reads now:

"However, GLUE has faced criticism in terms of the subjective decisions required in its application and how these affect prediction limits (informal likelihood function, lacks of maximum likelihood parameter estimation, and omission of explicit model error consideration). This subjectivity might lead to not being formally Bayesian (for that reason, GLUE includes the term "generalized" in its name)"

Obs. 5: Line 56, "…with formal Bayesian approaches…": But GLUE can also be used with MCMC/DREAM and has been - please recognise that defining a likelihood and searching the space are two seperate issues.

In that respect it might also be inclusive to mention the use of limits of acceptability within GLUE to get around these problems (going back at least to Freer et al 2004, also Vrugt and Beven, 2018 using DREAM_LOA; Beven and Lane, 2022 and in CURE). Still some subjectivity but involving more thoughtful decisions about the data. There is also a UPH20 paper (Beven et al., https://doi.org/10.5194/piahs-385-129-2024)
Ans. 5: We recognised that defining likelihood functions and searching the solution space in calibration are two different and separate issues. We wrote it explicitly in the main text which reads now:

"Building on previous work (see, e.g., Blasone et al. 2008), researchers have compared GLUE with formal Bayesian approaches. At this regard, both formal Bayesian approaches as well as GLUE can be used with advanced Monte Carlo Markov Chain (MCMC) schemes such as the Differential Evolution Adaptive Metropolis (DREAM, Vrugt et al. 2008). Important to note is that defining likelihood functions and searching the solution space at calibration are two independent issues. One way to get around these problems relies on the limits of acceptability which are typically used (but no mandatory) with GLUE (see, e.g., Freer et al, 2004; Vrugt and Beven, 2018; Beven and Lane, 2022; Page et al., 2023; Beven et al., 2024), involving more thoughtful decisions about the data (even though still with subjectivity)."

References:
Beven, K., & Lane, S. (2022). On (in) validating environmental models. 1. Principles for formulating a Turing-like Test for determining when a model is fit-for purpose. Hydrological Processes, 36(10), e14704.

Beven, K., Page, T., Smith, P., Kretzschmar, A., Hankin, B., & Chappell, N. (2024). UPH Problem 20–reducing uncertainty in model prediction: a model invalidation approach based on a Turing-like test. Proceedings of IAHS, 385, 129-134.

Freer, J. E., McMillan, H., McDonnell, J. J., & Beven, K. J. (2004). Constraining dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures. Journal of Hydrology, 291(3-4), 254-277.

Page, T., Smith, P., Beven, K., Pianosi, F., Sarrazin, F., Almeida, S., ... & Wagener, T. (2023). The CREDIBLE Uncertainty Estimation (CURE) toolbox: facilitating the communication of epistemic uncertainty. Hydrology and Earth System Sciences, 27(13), 2523-2534.

Vrugt, J. A., & Beven, K. J. (2018). Embracing equifinality with efficiency: Limits of Acceptability sampling using the DREAM (LOA) algorithm. Journal of Hydrology, 559, 954-971.

Obs. 6: Line 139-140, "For each point, a sample is established comprising neighbouring simulated river flows, defined by m_1 Flows smaller and m_2 flows larger than the point's discharge, both with the smallest differences.": Not very clear - what is meant by neighbouring? Illustrate with a simple figure... It is also then not clear how the m1 and m2 samples are related to u in RUMI

Ans. 6: Thank you for rising this issue. As in the original BLUECAT paper, neighbouring was taken in terms of the magnitude of stream flows. Figure R01.1 was generated and incorporated into the manuscript. The figure in question reads now:

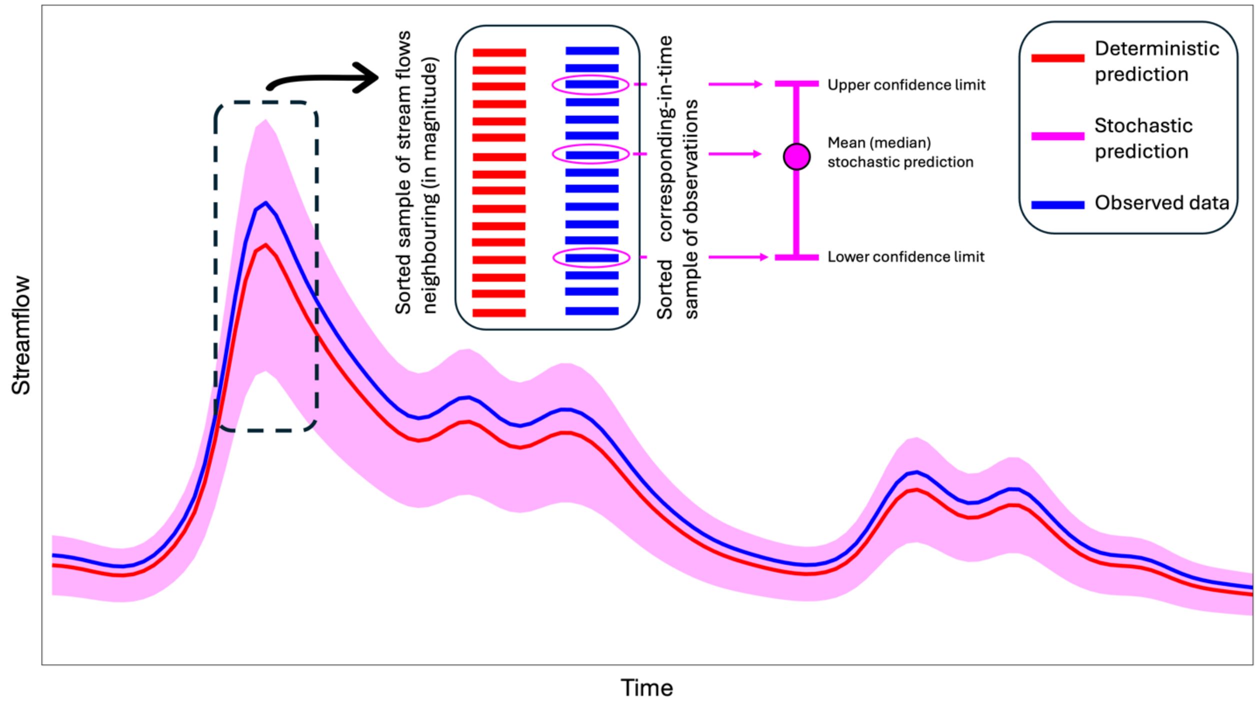

Figure R01.1. Conceptual illustration of BLUECAT methodology. Blue colour represents observed (streamflow) data, whereas red and pink colours are deterministic and stochastic predictions respectively.

Obs. 7: Line 145, "…Worth mentioning is…": It is worth mentioning ... is better English (also below)

Ans. 7: The sentence in question (and others similar) was (were) updated.

<u>Obs. 8</u>: Line 164, "…normalised amount of information…": Critical here is how H(Y|X) is estimated - needs to be more explicit. And for the limits in u (see comment above)
<u>Ans. 8:</u> Caption in Figure 2 (original submission) gave relevant information in this regard that is reported again as follows (see also Fig. 2d summarising RUMI computation methods):

"… Marginal and conditional entropies are computed empirically with bins. The filled cyan band is the area between the 97.5 and 2.5 percentiles of simulation estimated by BLUECAT."

We acknowledge that writing explicitly (and again) this information in the main text might be insightful for the reader, we added the following sentence to the main text:

"Additionally, and with the intention to avoid any additional assumption, marginal and conditional entropies are computed empirically with bins."

<u>Obs. 9</u>: Line 167, "…proper…": Is this proper in the sense of Vrugt?
<u>Ans. 9:</u> "Proper" was in the sense of Jorquera & Pizarro (2023) and Montanari & Koutsoyiannis (2025). Considering that this word does not add any additional information, we decided to remove it from the main text. The sentence in question reads now:

"… Furthermore, an uncertainty measure (in line with Jorquera and Pizarro (2023) and Montanari and Koutsoyiannis (2024) uncertainty quantification proposal) of the stochastic model computed with BLUECAT…"

<u>References:</u>
Jorquera, J., & Pizarro, A. (2023). *"Unlocking the potential of stochastic simulation through Bluecat: Enhancing runoff predictions in arid and high-altitude regions"*. Hydrological Processes, 37(12), e15046.

Montanari, A., & Koutsoyiannis, D. (2025). *"Uncertainty estimation for environmental multimodel predictions: The BLUECAT approach and software"*. Environmental Modelling & Software, 106419.

<u>Obs. 10</u>: Line 176, "…it is possible…": It is not possible????????? You can draw it (as in Fig 2a) but how can that actually arise? Also Figure 2c implies perfect data (that unrealistic ideal case again) - but you certainly do not have that here, if only because of daily discretisation and input interpolation errors..
<u>Ans. 10:</u> We thank reviewer's comment and share his view in terms of ideal and non-ideal available data. However, the intention behind the Figure 2 was to illustrate conceptually the reasoning behind RUMI formulation. We strongly believe this figure helps to clarify RUMI computation steps and line of thought (even though they can be extremely conceptual).

<u>Obs. 11</u>: Line 292 – 293, "…The latter with the intention to quantify the metric uncertainty…": Not clear what you mean here
<u>Ans. 11:</u> The sentence in question was rewritten with the intention to avoid confusion. It reads now:

"The latter with the intention to quantify the sensibility of RUMI as a function of those additional methodologies."

<u>Obs. 12</u>: Line 294, "high-quality input data and length of the time series": But that was surely not the case for some of these catchments? Using daily data itself introduces discretisation issues that will add to simulation errors.
<u>Ans. 12:</u> Used data were taken from the CAMELS-CL database (which, indeed, followed a quality control analysis), ensuring high-quality data. Additionally, in section "2.2 Data", an additional filter was applied with the intention to consider: a) near-natural hydrological regimes catchments; b) catchments with no more than 25% of missing data; and, c) standardisation of the time series length, finally being from 1990 to 2018.

Time discretisation was out of the scope of the presented work, even though it was highlighted as possible future research at the end of the conclusion section (see Ans. 15).

<u>Obs. 13</u>: Line 295, "…upgrade from the deterministic to the stochastic model)…": Do you mean in terms of sampling the m1 and m2 values to get at the local distribution of errors?
<u>Ans. 13:</u> Not the local distribution of errors but the local distribution of the variable under analysis (at each time step). See Ans. 6 and Figure R01.1.

<u>Obs. 14</u>: Line 297, "…might be…": might be? you have not said how it was estimated?
<u>Ans. 14:</u> See Ans. 8. Additionally, codes for RUMI computation in R and Matlab were provided with this submission and, therefore, any additional metric computation doubt can be resolved looking at the codes. The latter gives transparency and foster reproducibility of results.

<u>Obs. 15</u>: Line 305, "…additional research…": add testing with higher resolution data to reduce discretisation issues in smaller catchments
<u>Ans. 15:</u> Thank you for this suggestion. We added it to the conclusion section, reading now:

"Possible additional research is mentioned as follows: (a) Testing the RUMI-based approach with other rainfall-runoff models (lumped, semi-distributed, and distributed hydrological models); (b) Testing the RUMI-based approach under other hydroclimatological catchment characteristics and in a higher number of catchments; (c) Testing alternative uncertainty quantification methods; (d) Exploring the impact of varying data quality on RUMI performance to establish guidelines for data requirements; (e) Testing with higher resolution data to reduce discretisation issues; and, (f) Exploring the applicability of the RUMI in other disciplines such as meteorology, environmental science, and ecology where modelling and uncertainty quantification are critical."