Overall Impression:

This manuscript presents a timely, and valuable study that addresses a critical challenge in hydrology: monitoring intermittent rivers and ephemeral streams (IRES). The application of a relatively simple logistic regression model to classify flow states from field camera imagery is both pragmatic and innovative. The methodology is clearly described, the results are robust and convincingly presented, and the discussion thoughtfully places the work in the broader context of IRES monitoring, climate change, and water management. The integration of image classifications for quality control of stage data is a particularly strong and practical contribution. The manuscript is generally well-written and structured. I believe it represents a significant contribution to the field and is a strong candidate for publication after revisions.

Abstract

The abstract effectively summarizes the study's motivation, methods, key findings, and implications. It clearly states the problem (monitoring challenges in IRES), the proposed solution (image classification), the location, and the broader significance of the work. However, I am including some comments and questions that can improve the abstract.

1. The abstract mentions the model was used for quality control of the stage time series. Could you briefly hint at the nature of the discrepancies/uncertainties/errors found (e.g., sensor drift, noise during high flow) to immediately highlight the practical utility of the method?
2. The term "Imagey" in the title appears to be a typo for "Image-based." Was this intentional?
3. The abstract focuses on categorical classification. Did the model's probability output itself provide any additional, continuous-like insight beyond the three discrete categories?

Introduction

The introduction provides a comprehensive and compelling background, effectively building the case for the importance of IRES and the difficulties in monitoring them. The literature review is extensive and covers relevant areas, including remote sensing, citizen science, and various modeling approaches.

1. While you cover technological methods, could you briefly mention the organizational/funding challenges of maintaining sensor networks in remote IRES to further justify the need for low-cost methods?
2. You mention that deep learning has been mainly applied to perennial streams. Could you elaborate on one or two key reasons why these methods are particularly challenging to directly transfer to IRES (e.g., more dynamic channel geometry, greater debris, longer dry periods)?
3. The introduction effectively sets up the use of machine learning with imagery. Would it be valuable to more explicitly state the core hypothesis: that visual features in daytime imagery are sufficient to reliably classify IRES flow states for monitoring purposes?
4. Have you considered citing studies that discuss the hydrological significance of the "pooling" phase in IRES, which your method can detect but stage sensors cannot? For example, Stubbington, R., et al. (2017). The biota of intermittent rivers and ephemeral streams: aquatic and terrestrial assemblages. In *Intermittent Rivers and Ephemeral Streams* (pp. 217-245). Academic Press. could strengthen this point.
5. The transition from the broad introduction to the specific objectives of the study is clear, but could the final paragraph be slightly more structured to explicitly list the primary aims of the paper?

Methods

The methods section is exceptionally detailed and reproducible, a major strength of the manuscript. The description of the study site, data sources, image preparation, model training, and validation is thorough. The handling of unbalanced classes and the development of a confidence metric are particularly sophisticated and commendable.

1. You limited image analysis to 9 am–4 pm PST to avoid low-light issues. Was any consideration given to using the camera's flash-illuminated nighttime images for a simple binary "water"/"no water" classification, given that this is a defining feature of IRES?
2. For the image cropping to 1000x1200 pixels, was this specific size determined empirically? Did you experiment with different crop sizes or aspect ratios to optimize feature recognition?
3. The manual weighting scheme (3.5 for water categories, 3 for obstructed, 1 for 'no water') is interesting. Could you provide a sentence on the rationale behind these specific weight values?
4. The confidence level assignment is well-explained but based on qualitative assessment of probability distributions. Were any quantitative metrics (e.g., maximizing Youden's J index) explored to define the probability thresholds more objectively?
5. You use soil moisture data from a nearby site (DRW) with probably different geology. How might this spatial disconnect influence the interpretation of the relationships between soil moisture and stage at PEC? In addition, a section of uncertainties would be great, since I can see some sources of uncertainties in your work/modelling, for example: image quality and environmental variability (the classification model's performance is inherently tied to the quality and consistency of the input imagery), sensor data reliability and spatial mismatch (the "ground truth" data used for validation and comparison are themselves sources of uncertainty), limited training data and site specificity (the model was trained on a relatively small, manually labeled dataset (537 images) from a single site. this raises uncertainty about its performance when transferred to other IRES with different channel morphology, substrate, vegetation, and water clarity. the model's features (e.g., learned from the specific staff plate and rocks at pec) may be overly tailored to this unique location). Do not get me wrong, I still think there is a lot of value in publishing this paper, however, it is good to show the uncertainties and potential bias of the approach.

Results

The results are clearly presented, with appropriate use of tables, figures, and statistics. The model performance metrics are convincing, and the comparison between image classifications, observed stage, and modeled discharge is effective in demonstrating the value of the approach.

1. The confusion matrix shows that 'obstructed' images were most often misclassified as 'no water'. Given the ephemerality of the stream, do you think this misclassification might be functionally acceptable in many cases, as it likely reflects a true dry state?
2. Figure 7/9 and the text describe how image classifications identified sensor malfunctions. How many erroneous data points would have been missed without this image-based quality control? A rough percentage or count would powerfully quantify this benefit.
3. In Figure 10, the results show a notable discrepancy where the NWM reported zero discharge during periods of observed high water (e.g., Jan-Feb 2018). What is your leading hypothesis for this systematic underestimation by the NWM in this specific catchment?
4. The relationship between stage and soil moisture at 5 cm is stronger than at 100 cm. Does this suggest that flow at PEC is primarily driven by shallow subsurface flow or saturation-excess overland flow rather than deeper groundwater contributions? If so, that should be discussed, showing how the changes over time may impact the local hydrology of the watershed and river.
5. You mention that high flows remain unmeasured due to safety. Could the image classifications be calibrated against the NWM output or other hydraulic models to provide a rough estimate of discharge during these extreme events?

Discussion

The discussion successfully interprets the results, acknowledges limitations, and explores the wider implications of the work. The sections on unique site features and extensibility are particularly thoughtful and elevate the manuscript beyond a simple methods paper.

1. You rightly note that temporal correlation of flow states could be used for further quality control. Could a simple Hidden Markov Model be a natural next step to incorporate this temporal dependency?
2. How does the performance of your logistic regression model (91% accuracy) compare, in your view, to the potential trade-offs of using a more complex but data-hungry model like a CNN for this specific task? That could be a good paragraph in the discussion, showing the drawbacks and positives sides of using a parsimonious but effective model.
3. In the biggening (abstract and objectives) you state the method is "transferable." Could you specify the primary condition for transferability (e.g., the presence of a staff plate or a consistent field of view of the streambed)? Maybe a bit of discussion on the costs of this equipment set up would also help the reader to have an idea of how much it would cost. Perhaps that could be included in the methods?!
4. The discussion on the potential for subsurface flow bypassing the PEC site is fascinating. Could this hypothesis be further supported by comparing the water level in Lake Mendocino with dry/wet periods at PEC?
5. You mention that your code is transferable. What is the minimum number of manually labeled images you would estimate is necessary to achieve reasonable performance at a new, similar site?
6. In the context of climate change, how might your method help in detecting shifts in the timing of flow initiation and cessation in IRES, which is a key impact of warming temperatures?
7. Have you considered referencing studies that have successfully implemented low-cost, image-based methods in data-scarce regions? For instance,
    - Noto, S., Tauro, F., Petroselli, A., Apollonio, C., Botter, G., & Grimaldi, S. (2022). Low-cost stage-camera system for continuous water-level monitoring in ephemeral streams. *Hydrological Sciences Journal*, *67*(9), 1439–1448. https://doi.org/10.1080/02626667.2022.2079415
    - Rodrigues, R. M., Braga, B. B., & Costa, C. A. G. (2025). Efficiency in river discharge measurement: combining Chiu's method with particle image velocimetry techniques. *RBRH*, *30*, e31.

These papers could broaden the perspective on transferability and cost-effectiveness.

Conclusion

The conclusion effectively summarizes the main findings and their significance. It compellingly argues for the role of this low-cost method in improving IRES monitoring and, consequently, water management in a changing climate.

1. Could the conclusion more explicitly state the single most important recommendation for water managers seeking to implement this method?
2. Beyond FIRO, can you speculate on one other specific water management decision (e.g., environmental flow allocations, drought contingency planning) that would benefit from the categorical flow data your method provides?
3. While the method is low-cost, the conclusion could acknowledge the ongoing costs and challenges of maintaining field cameras in harsh environments as a consideration for long-term deployment.
4. What do you see as the next critical technological or methodological advancement needed to make IRES monitoring truly scalable across vast river networks?

I congratulate the authors on an excellent piece of work. I hope that with these revisions, the manuscript will be an even stronger contribution to the literature.