Comments to RC2

The reviewer's comments are in **black** and our answers in **blue.**

This study addresses a critical gap in intermittent river monitoring in the Umbuzeiro River, Brazil, which I consider significant with implications for other arid areas and for anticipating future climate-change impacts. The integration of UAV surveys, Random Forest classification, and remote sensing–based landscape attributes is a valuable approach. However, the manuscript in its current form suffers from presentation and clarity issues that obscure the scientific contributions. Substantial revision is needed to improve communication, sharpen the articulation of the scientific questions more process-oriented interpretation of results in the context of Earth surface processes.

We sincerely thank the reviewer for the positive evaluation of our study and for highlighting its relevance to dryland hydrology and ecohydrological monitoring. We also appreciate the constructive suggestions aimed at improving the manuscript's clarity and scientific framing. We acknowledge that aspects of the presentation—particularly the articulation of our research questions and the interpretation of our results in light of geomorphological and hydrological processes—can be improved. In the revised manuscript, we:

- **Refine the introduction to clearly articulate the research questions and the knowledge gap addressed;**
- **Improve the flow and clarity of the text, with particular attention to the explanation of methods and model interpretation;**
- **Strengthen the discussion by linking results more explicitly to hydrological and landscape processes, such as flow regulation by dams and topographic accumulation.**

These revisions will help to better communicate the scientific contributions of the study and its applicability to broader contexts, including climate-change-related shifts in river regimes. We thank the reviewer again for the thoughtful and helpful feedback.

Major Comments

1. **Scientific framing**

- While the methodological framework is carefully outlined, the manuscript must articulate the *scientific question* more strongly. The introduction hints at this, but the discussion largely focuses on model outcomes without interpretation in terms of physical or hydrological processes. For example: "distance from the last dam is the most important predictor" does not require a complex Random Forest model to identify. Similarly, Section 4.2 notes that elevation and drainage area are important predictors, but these are already well established. This journal is not a primarily GIS/remote sensing journal and so I encourage the authors to discuss What is the added value of this modeling exercise? Does it allow optimization of resolution? What unexpected findings emerged? What are the implications for other arid environments or climate-driven changes in intermittency?

We thank the reviewer for this thoughtful and constructive comment. We agree that the manuscript would benefit from a clearer articulation of the scientific question and a stronger framing of the results in terms of hydrological and geomorphological processes. In the revised version, we revisit the introduction to explicitly define the knowledge gap we aim to address — namely, the limited understanding of fine-scale spatial patterns of intermittency in intermittent rivers, and the role of multiple landscape controls in shaping these patterns. An excerpt added to the Introduction section is presented below:

> *"This study addresses a critical knowledge gap on spatial patterns of intermittency in a river. We investigate how to effectively map these patterns by incorporating field observations and landscape attributes, along with other variables, into the modeling process. Although the general role of topographic and climatic drivers is well established, little is known about their interaction with human modifications, such as farmer dams, to influence the presence of water on river reaches. Furthermore, methodological approaches capable of integrating multiple data sources (e.g., UAV, landscape attributes, and machine learning) are still limited, particularly in intermittent rivers where wet patches are hard to map."*

> *"Here, we explore how different environmental and anthropogenic variables contribute to the occurrence of water in intermittent reaches. When we combine field-based classifications with Random Forest modelling, we investigate not only the prediction accuracy of different data sources, but also the relative importance of physical attributes and land use drivers on river wetness patterns."*

While predictors such as elevation or drainage area are indeed well established in hydrological modeling, our goal was to explore how these and other variables (e.g., dam proximity, vegetation cover) interact in a high-resolution spatial context, and how machine learning methods like Random Forest can help quantify the relative importance of such predictors across heterogeneous, data-scarce environments. The added value of the model lies in its ability to integrate multi-source data (UAV, satellite, field observations), operate at a fine spatial resolution, and be used as a framework for extrapolation in ungauged systems.

In the discussion section, we also expand the interpretation of our findings with a focus on physical processes — for instance, the role of upstream impoundments in controlling downstream drying patterns, or how topographic accumulation may reflect water retention. Although some key predictors were expected, others offer insights into local regulation mechanisms that might be overlooked in coarser-scale analyses. We also reflect on the potential implications of our approach for studying climate-driven shifts in river intermittency in other basins. To address the reviewer's comment, we will expand the Discussion section with the following excerpts:

> ***4.2 Identification of important predictors***
>
> *"Model performance patterns align with known physical drivers of river intermittency. For example, river reaches with small contributing areas*

were more frequently classified as dry. Conversely, regions with bigger contributing tend to sustain surface water for longer periods."

"Although predictors related to landscape attributes are well-established in hydrology, they gain new value here through their interaction with other spatial variables at finer resolution. For example, areas with high accumulation but located downstream of dams often remain dry, suggesting that topography alone does not explain surface wetness in human-modified catchments."

"This consistent selection of "distance to the next dam" and "distance from the last dam" as top predictors highlights the strong influence of anthropogenic flow regulation on surface water distribution. While the relevance of dams in altering flow regimes is widely acknowledged, our findings provide a fine-scale perspective on how even small impoundments can generate abrupt shifts in flow permanence along the network. This is particularly relevant in semi-arid regions, where it is common to use small-scale storage and retention of surface water. However, as river networks become more intermittent due to climate-driven changes, this can be a reality in all climates."

"Water accumulation in intermittent river seems to be governed by processes captured by landscape attributes (including human-made interferences) and satellite indices. The consistent selection of these predictors across models reinforces them as controls and indicators to water retention."

The revised Conclusion section incorporate the following insights:

"The modelling framework developed in this study contributes to a broader understanding of flow intermittency as a spatially complex and highly dynamic process over time. The integration of high-resolution predictors, especially related to dam presence, landscape attributes and satellite indices, offers a scalable and adaptable approach for mapping wetness conditions in other dryland river systems. These insights are particularly relevant in the context of increasing climate variability and water stress, as they point to key landscape features that can be targeted for monitoring or management. Our results demonstrate that even in the absence of extensive hydrometric data, meaningful patterns can be derived from the careful integration of remote and field-based observations."

2. Presentation quality

- The manuscript follows the standard structure, but the communication of ideas and concepts requires improvement. At present, several sections read like figure captions rather than explanatory text (e.g., line 219). Statements should be fine-tuned for accuracy and consistency (e.g., lines 109–111).

We thank the reviewer for this helpful observation. We agree that certain parts of the manuscript required revisions to improve clarity and narrative flow. In response, we made the following changes:

> **Line 219:**
> **Original:** *"The rightmost column in Table 2 shows which predictor is employed in each model."*
> **Revised:** *"The specific predictors employed in each model are summarized in Table 2."*
>
> **Lines 109–111:**
> **Original:** *" We perform UAV surveys every month during the rainy season (Apr–Jul), and then again in November of 2022 (Fig. 4). Since the rainy season usually occurs during the first months of the year (Soares et al., 2024), satellite images are normally not useful during this period due high cloudiness."*
> **Revised:** *"We perform UAV surveys monthly during the rainy season (Apr to Jul), and once more in November 2022 (Fig. 4). As noted by Soares et al. (2024), the rainy season typically occurs in the first half of the year in this region. The high frequency of cloud cover during the rainy season typically limits the availability of usable satellite imagery, reinforcing the need for UAV-based data collection during this period."*

In addition to the specific revisions mentioned above, we carefully reviewed the manuscript to improve overall clarity and ensure that all sections communicate ideas effectively, avoiding figure-caption-like phrasing and enhancing consistency throughout.

- Section 3.1 Modelling workflow requires rewriting: the repeated use of the generic term “data” is too vague. Please specify clearly whether you are referring to UAV-derived classifications or predictor variables.

We appreciate the reviewer’s observation. In response, we carefully revised Section 3.1 to avoid using the term “data” in a generic way. We now clearly specify the type of data being referenced throughout the section — distinguishing between UAV-derived classifications and predictor variables. The section now reads:

> *“Workflow follows the flowchart shown in Fig. 2. There are three main steps: (1) UAV data collection and mapping, (2) model training, and (3) application to the whole river. Step 1 consists of the river course visual mapping. Then, we conduct UAV surveys multiple times in river reaches, producing high-resolution orthomosaics. We use the images to classify 1 meter reaches in water occurrence classes, which are used as ground-truth data for model training and evaluation. In step 2, input data from predictor variables is used during the training of Random Forest classification models in order to obtain the same class as in the observed data (UAV-derived classification). Training includes the use of static and dynamic landscape attributes as candidate predictors to explain the observed patterns of water occurrence. In a recursive process, we compute*

the importance of the candidate predictors (or features). We test three models based on predictor subgroups and select the most important predictors in each model The Random Forest models are retrained with the selected predictors in order to obtain the best models. In step 3, the models are applied to the entire river."

- Figures require improvement. For example, continuous variables such as monthly precipitation should be plotted as interpolated lines with points not bar plots. Figure captions must be complete and self-contained.

We thank the reviewer for the suggestion. While we understand the recommendation to use line plots for continuous variables, we chose to represent monthly precipitation using bar plots because it is an aggregated variable over discrete time intervals (months), rather than a continuously measured variable. This is a common approach in hydrology and climatology to visually emphasize the magnitude of rainfall in each month.

That said, we acknowledge the importance of clarity and completeness in figure presentation. We will review all figure captions to ensure they are self-contained and provide sufficient information for interpretation without referring back to the main text.

3. Methods

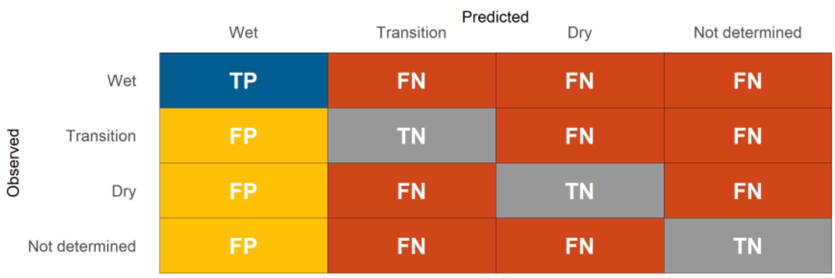
- Although all equations are listed, the conceptual meaning of the metrics (e.g., balanced accuracy) is not unpacked, remember your audience. Please provide intuitive descriptions and, if possible, a conceptual diagram.

Thank you for this valuable comment. We agree that the inclusion of intuitive descriptions enhances the accessibility of the manuscript, particularly for readers who may be less familiar with performance metrics. We have now added brief conceptual explanations for each metric presented — including overall accuracy, balanced accuracy, precision, sensitivity and specificity — highlighting their relevance in the context of imbalanced classes:

"To evaluate performance per class, a one-vs-rest strategy was adopted: each water occurrence class was treated as the positive class in turn, with all remaining classes grouped as negative. This approach allows for consistent metric calculation across imbalanced classes. A conceptual diagram is presented in Fig. 6."

"Overall accuracy (Eq. 1) reflects the proportion of correctly classified instances over the entire dataset and is useful for assessing general performance. However, it may be biased in datasets with imbalanced class distributions. To address this, we also report balanced accuracy (Eq. 4), which equally weights performance across classes by averaging sensitivity (true positive rate) and specificity (true negative rate). This metric provides a more reliable measure of performance when class sizes differ significantly. Together, these metrics allow us to evaluate how well each

model discriminates among the different water occurrence classes, considering both general correctness and class-specific sensitivity."



***Now Figure** 6. Confusion matrix used to compute classification performance metrics. Each color represents each of the possible combinations between predicted and observed classes: TP = true positives, TN = true negatives, FP = false positives, FN = false negatives. In this example, the "positive" class is "Wet". This process was repeated for each class.*

As can be observed, we included a conceptual diagram (now Figure 6) that visually illustrates how the main metrics are derived from the confusion matrix. We also maintain the equations in the text to help the reader understand the calculation process.

- Please clarify why Table 2 lists 25 predictors, but Fig. 8 refers to 40.

Thank you for pointing out this important distinction. Some of the variables presented in Table 2 refer to predictor groups or data sources, each of which may yield multiple individual variables used in the modeling. For example, the land use/land cover predictors is derived from two classification schemes — one with 8 classes and another with 7 — and for each scheme, we compute the class frequency within the surrounding area as separate predictor variables.

As a result, although Table 2 presents 25 predictors, the full set expands to 40 individual predictor variables, which are shown in now Fig. 8. We have clarified this point in the manuscript to avoid confusion:

"While 25 predictors are individually listed, two of them correspond to land use and land cover (LULC) classifications, which include 8 and 7 classes respectively. The frequency of each class is treated as a separate predictor, resulting in a total of 40 variables used in the models."

4. Results and interpretation

- Some results are not fully explained. Why does Model (c) predict >75% dry in March, peak wet season, but not in November, second wet season? Why is Model (a), with 5-day frequency predictors, described as the most successful in simulating intermittency when it has the coarsest temporal resolution?
- Please interpret results in physical terms, not only in terms of model metrics.

Thank you for this important comment. We revised the discussion to incorporate physical interpretations of model results beyond statistical performance.

Specifically, we now relate model outputs to hydrological and ecological processes that govern intermittency. For example, we highlight how predictors such as vegetation indices and topographic convergence correspond to physical controls on water retention. We also discuss how the timing of rainfall may explain temporal discrepancies in model predictions — especially between March and November.

The unexpectedly high proportion of predicted dry conditions in March — typically part of the peak wet season — appears to reflect limitations in the predictor variables used, particularly those derived from simpler datasets (e.g. Model (c) with accumulated precipitation (30 days) as the dynamic variable). While March is climatologically wetter, short-term hydrological conditions (e.g., rainfall deficits or delays in runoff response) can vary substantially. In contrast, the November period, although later in the year, coincided with isolated rainfall events that triggered surface water reappearance — something that may be better captured by the rainfall used in Model (c).

As for Model (a), although it uses predictor variables with a 5-day revisit frequency, it consistently performed better in simulating intermittent dynamics. This is likely because the satellite image captures better the momentary condition, such as recent water presence or vegetation response, which are more stable and informative for distinguishing intermittent patterns than precipitation-only predictors.

We clarified these interpretations in the discussion section to better reflect the physical reasoning behind the model behaviors. Excerpts added to the discussion section are presented below:

> ***4.4 Umbuzeiro River: temporal extrapolation with models***
>
> *"Although March is part of the rainy season, the pattern of drying predicted by Model (c) may reflect lagged hydrological responses not captured by the accumulated precipitation for that month. In contrast, the prediction for November coincided with isolated rainfall events that triggered surface water reappearance — something that may be better captured by the accumulated rainfall."*
>
> *"Since Model (c) only uses rainfall data as dynamic predictor, it was expected to follow the seasonal behavior of water occurrence. However, the poor performance of this model in its application to the whole river indicates that it cannot extrapolate local-trained conditions considering the landscape attributes together with precipitation."*
>
> *"Interestingly, the stronger performance of Model (a) and (b) suggests that satellite-driven variables may better capture ecologically meaningful signals of intermittency, possibly due to their ability to represent spectral landscape responses. It is also possible to observe the complementary convergence of vegetation and water-related satellite indices and landscape attributes, which reinforce each other as physical controls (or responses) related to water retention."*

5. References and context

- The references are strongly weighted toward recent Brazilian studies. Please include some of the classic papers on ephemeral and intermittent streams.

  Thank you for this important comment. We revised the text to incorporate more diverse papers on ephemeral and intermittent streams.

  COSTIGAN, Katie H. et al. Flow regimes in intermittent rivers and ephemeral streams. In: **Intermittent rivers and ephemeral streams**. Academic Press, 2017. p. 51-78.

  ERIS, Ebru et al. Frequency analysis of low flows in intermittent and non-intermittent rivers from hydrological basins in Turkey. **Water Supply**, v. 19, n. 1, p. 30-39, 2019.

  SARREMEJANE, Romain; MESSAGER, Mathis Loïc; DATRY, Thibault. Drought in intermittent river and ephemeral stream networks. **Ecohydrology**, v. 15, n. 5, p. e2390, 2022.

  SEFTON, Catherine EM et al. Visualising and quantifying the variability of hydrological state in intermittent rivers. **Fundamental and Applied Limnology**, v. 193, n. 1, p. 21-38, 2019.

- The paper should also acknowledge interannual variability in flow connectivity, which is critical for interpreting results beyond a single year.

We have added a paragraph to section 4.6 Work limitations acknowledging interannual variability and briefly discussing how stream intermittency may vary between years.

> *"We based our analysis on data from a single hydrological year. While this approach allowed for the capture of short-term hydroclimatic variability, it limits the model's capacity to generalize across years with different rainfall patterns. Future studies including multi-year data and alternative temporal windows could help address this limitation."*

Minor Comments

Lines 109–111: Be consistent in describing rainy season timing. April–July are not "the first months of the year."

Thank you. We revised the sentence to clarify the timing of the rainy season and avoid imprecise phrasing. It now reads: *"We perform UAV surveys monthly during the rainy season (Apr to Jul), and once more in November 2022 (Fig. 4). As noted by Soares et al. (2024), the rainy season typically occurs in the first half of the year in this region."*

Line 134–135: Do you need to direct readers to the Sentinel user guide?

We agree that this reference is not essential and have removed the sentence referring readers to the Sentinel user guide.

Lines 135 & 161: Clarify the role of Google Earth Engine (GEE) for Sentinel and MapBiomas.

We clarified the text to specify that GEE was used as the platform for accessing, processing, and analyzing both Sentinel and MapBiomas data:

Lines 135:
Original: *" Our sentinel images are obtained and processed using Google Earth Engine (GEE)."*
Revised: *"The access, processing and analyzing of sentinel images is performed in Google Earth Engine (GEE). The spectral indices were processed in the platform, and the mean values were obtained per modelling unit."*

"MapBiomas data are freely available, and we use Google Earth Engine (GEE) to access, process and analyze the data. The LULC classes were processed in the platform, and the frequency of each class were obtained per modelling unit."

Line 140: Specify what was done in R.

The sentence was revised to specify that part of image processing was conducted using R: "*If there are missing portions, these gaps are filled with the image of the previous or following day. In those cases, the overlaying of different raster (from subsequent days) is performed in R Statistical Software version 4.1."*

Line 165: Where in Figure 1 is this shown?

We added a reference in the caption of Figure 1 to indicate where the relevant feature appears in the figure: "*Figure 1. Umbuzeiro River in the Benguê catchment with the monitored river reaches used in UAV surveys: upstream, middle, and downstream. Precipitation is measured in the Aiuaba Experimental Basin (AEB) and shown the monthly data. Stream flow is measured in Aroeira section with no streamflow observed during the year of 2022."*

Line 172: Please provide information on data access (can it be downloaded and where?).

We appreciate the comment. The original raw dataset is not publicly available. However, a processed and formatted version of the same dataset, including the relevant variables used in our study, is available through the following peer-reviewed publication:

Fullhart, A., Goodrich, D. C., Meles, M. B., Oliveira, P. T. S., das Neves Almeida, C., de Araújo, J. C., & Burns, S. (2023). Atlas of precipitation extremes for South America and Africa based on depth-duration-frequency relationships in a stochastic weather generator dataset. *International Soil and Water Conservation Research*, *11*(4), 726-742. DOI: https://doi.org/10.1016/j.iswcr.2023.01.004.

We have added this reference to the manuscript and clarified the data availability in the methods section.

Line 203: Suggested rewording: "containing the type of variable…"

Thank you for the suggestion. We revised the sentence to improve clarity and formality. The updated version now reads: *"A summary of all candidate predictors used in our modeling approach is presented in Table 2, containing the type of variable, its spatial distribution, and frequency."*

Line 207: Cite Table 2 after "candidate predictors."

Table 2 is now cited at the appropriate location as suggested.

Line 219: Reword to “The specific predictors employed in each model are summarized in Table 2.”

We have revised the sentence accordingly.

Line 360: Section 4.6 Add acknowledgment of interannual variability and how the intermittency might vary

We have added a paragraph acknowledging interannual variability and briefly discussing how stream intermittency may vary between years.

Line 419: Avoid anthropomorphizing landscapes; specify which aspects of *river metabolism* are relevant.

Thank you for this helpful comment. We revised the sentence to remove anthropomorphic language and to specify which aspects of river metabolism are addressed. The revised version reads: *“From an ecological perspective, mapping the temporal dynamics of wet and dry conditions at the river scale is crucial for assessing species migration and resilience, characterizing habitat availability, and analyzing key components of river metabolism, such as gross primary production and ecosystem respiration.”*

Figures

- Fig. 1: Add discharge/precipitation panel to establish arid context, I know precipitation data is presented in Figure 9

  Thank you for the suggestion. To better contextualize the hydrological regime of the study area, we have added a precipitation panel also to Figure 1, summarizing monthly rainfall averages. This provides a clearer overview of the region’s seasonal aridity, complementing the more detailed data presented later in Figure 9. No discharge occurred during 2022 in the study area.

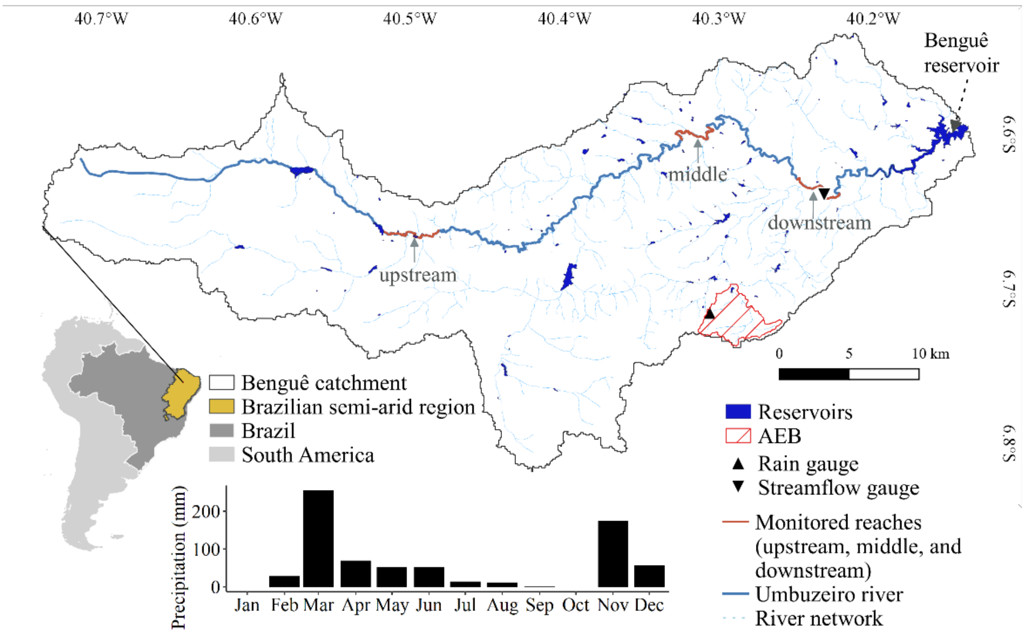


Figure 1. *Umbuzeiro River in the Benguê catchment with the monitored river reaches used in UAV surveys: upstream, middle, and downstream. Precipitation is measured in*

the Aiuaba Experimental Basin (AEB) and it is shown the monthly data. Stream flow is measured in Aroeira section with no streamflow observed during the year of 2022.

- Fig. 2: Show/explain how exactly the UAV surveys are used

Thank you for this observation. We have clarified the role of UAV surveys in both the figure and the main text. In the updated figure caption, we now specify that UAV imagery is used to produce high-resolution (1 meter) water occurrence (Wet/Dry etc) classifications, which serve as ground-truth data for training the random forest models.

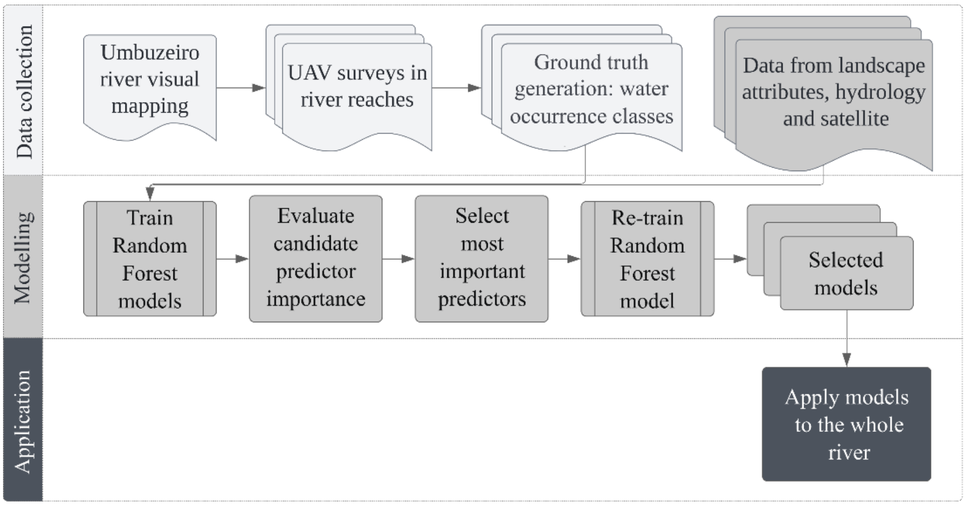


Figure 2. *Flowchart display of the steps taken during the mapping and modelling of water occurrence. UAV surveys were used to generate high-resolution classification of water occurrence (1.0 meter reaches), which served as ground-truth data for training the classification models. These classifications, alongside predictor variables, were used to train and validate random forest models of water occurrence.*

- Fig. 3: Quantify misfit between flow accumulation (DEM) and visual mapping. Adjust blue line color for visibility.

Thank you for this comment. We have changed the color of the flow line to improve visibility and included a quantitative estimate of the misalignment between DEM-derived flow paths and visually mapped river reaches (e.g., average lateral offset in meters): *"The average horizontal misfit between the flow accumulation path and the manually mapped stream was approximately 60 meters."*

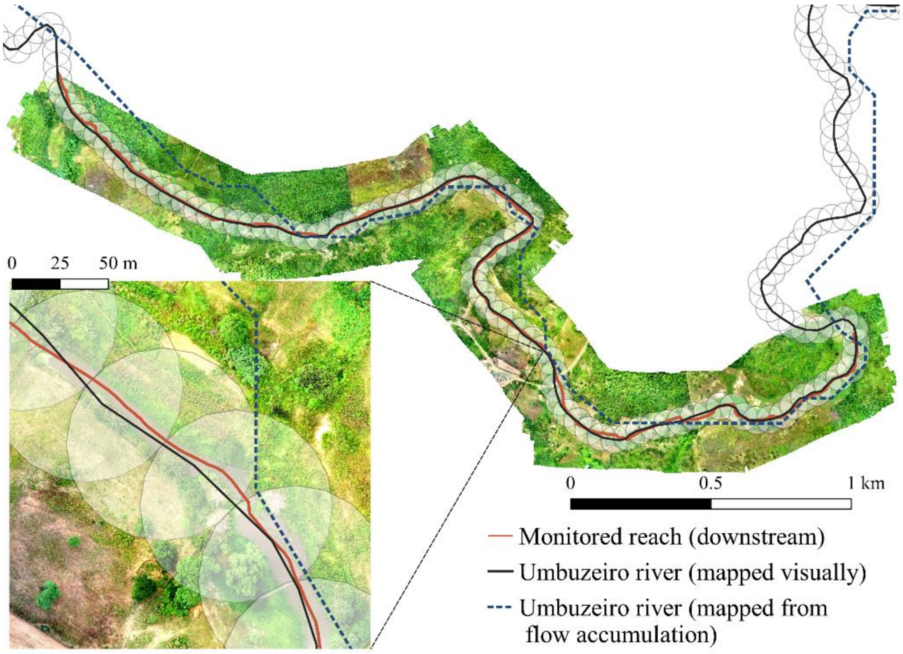


Figure 3. Difference between Umbuzeiro River mappings: the previously available mapping based on flow accumulation (from 30 m resolution DEM) versus visual mapping of the Umbuzeiro River course with satellite imagery. The average horizontal misfit between the flow accumulation path and the manually mapped stream was approximately 60 meters. Image from UAV flight showing the monitored river reach. In highlight the modelling units (diameter of 100 m): river areas whose data is evaluated in water occurrence modelling.

- Fig. 4: Add discharge/precipitation to support text (lines 109–110).

We have updated Figure 4 to include a precipitation panel aligned with UAV survey dates, providing clearer context for interpreting seasonal wetness dynamics.

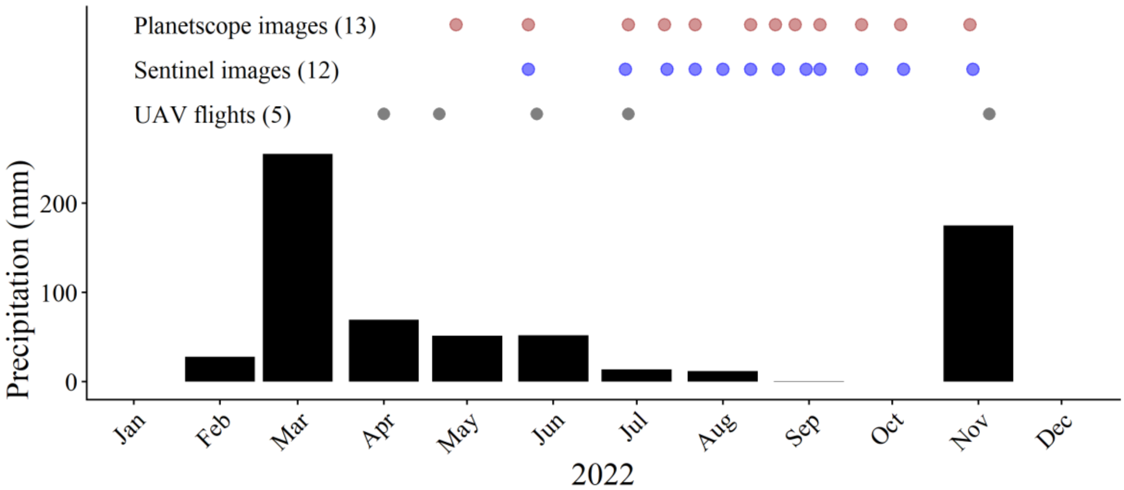


Figure 4. Monthly precipitation and dates of imagery available in 2022, indicating the number of satellite images of each sensor and UAV flights.

- Fig. 5: Missing panel label (a). Clarify "1.0 m reaches," as 1 m seems implausibly small.

Thank you for pointing this out. We have added the missing panel label (a) and clarified in the figure caption and main text that "1.0 m reaches" refers to river

segments of approximately 1 meter length along the mapped river sections. These fine-scale segments were individually classified as Wet, Transition, Dry, Not Determined using UAV imagery collected over three approximately 4-km-long river sections. This high spatial resolution allows detailed mapping of intermittency patterns. We also reviewed the explanation of this classification process in the whole document. The updated figure caption now reads:

Figure 5. *Examples of UAV imagery showing water occurrence classes along the Umbuzeiro River. UAV surveys were used to generate high-resolution classification of water occurrence (1.0 meter reaches). We show in (a) different wet patches and in (b) different examples of damming structures.*

- Fig. 7: Clarify why parameters with high Mean Decrease Accuracy were selected.

  Thank you for raising this point. We have clarified in the caption and text that variables with higher Mean Decrease Accuracy (MDA) values were selected because they contribute more significantly to model performance:

  "Variables with higher Mean Decrease Accuracy (MDA) values were selected because they have a greater impact on the predictive performance of the model. MDA reflects how much the accuracy of the model decreases when a given variable is excluded; thus, a higher MDA indicates that the variable provides important information for distinguishing between classes. By selecting variables with the highest MDA, we focused on those that most strongly influence the model's ability to predict spatial patterns of intermittency. This approach helps identify predictors most relevant to distinguishing our water occurrence classes."

  Now Figure 8. *Random forest model evaluation and feature selection results for three predictor sets: (model a) Sentinel-2 indices, (model b) PlanetScope indices, and (model c) hydrological variables. In (a) recursive feature elimination curves showing accuracy of the models as features are removed. In (b) out-of-bag (OOB) errors comparing models using all candidate predictors versus only the selected ones. BC denotes the benchmark classifier. In (c) mean decrease accuracy for the top 10 most important predictors in their respective model. Highlighted predictors were selected through recursive feature elimination based on their higher Mean Decrease Accuracy.*

- Fig. 8: Please clarify why Table 2 lists 25 predictors, but Fig. 8 refers to 40.

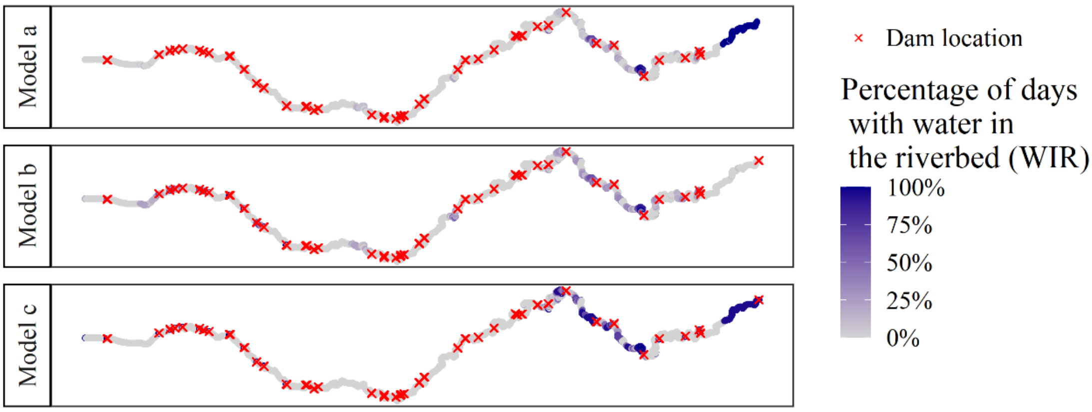
  We have clarified in the text that some predictor groups in Table 2 contribute multiple individual variables. Specifically, land cover predictors are derived from two classifications (8 and 7 classes), and the class frequency of each is included as a separate predictor, resulting in 40 total variables displayed in now Figure 8.

- Fig. 9: Explain Model (c) behavior, see Major comments.

  Please refer to our detailed response under the major comments section. We now discuss in the manuscript how Model (c)'s behavior may reflect limitations in its predictors or seasonal anomalies not captured by the available data. We also contrast its performance with other models in relation to rainfall timing and spatial heterogeneity.

- Fig. 10: Clarify how “High” and “Low” were calculated. Can values be normalized and expressed as percentages?

  Thank you for this insightful comment. To improve interpretability, these values were normalized and expressed as percentages of the total days monitored. We clarified in the text and figure caption that this percentage corresponds to the frequency with which each river reach was classified as “Wet” or “Transition” over the study period.



__Now Figure 11.__ Observations with water in the riverbed (WIR), that is when the reach presents either "Wet" or "Transition" classes. The number of observations is normalized and expressed as percentages. Model (a) uses Sentinel predictors; (b) uses Planetscope indices; and (c) uses hydrological data.

Tables

- Table 2: Organize predictors by model (a–c), consistent with Figs. 7, 9, and 10. Replace “constant” as that is not a Frequency

  Thank you for this comment. We have replaced the term “constant” in the frequency column with “static” to better reflect that these variables do not change during the study period but still serve as important model inputs. Additionally, we agree that consistency across tables and figures is important. However, in this case, we chose to organize Table 2 by predictor type (e.g., vegetation indices, hydrology, landscape), as this structure better reflects the conceptual grouping used in the modeling process. We feel that grouping by variable type improves clarity, especially given that many predictors are used by more than one model. For consistency with Figures 7, 9, and 10, we have ensured that the “Used by model(s)” column clearly indicates the model(s) associated with each predictor.

- Table 3: Define balanced accuracy conceptually, I know the equation is given, and present Train/Test values in a clearer layout.

  Thank you for this valuable suggestion. We have reformatted Table 3 to separate training and testing values into distinct sections, making the results easier to interpret. In addition, we included a conceptual definition of balanced accuracy in the text to complement the equation already provided in the methods section. This

definition highlights its usefulness in evaluating model performance under class imbalance. We added this excerpt to the text:

> *"Balanced accuracy accounts for class imbalance by averaging sensitivity (true positive rate) and specificity (true negative rate), providing a more reliable assessment when classes are unevenly distributed. Values are reported separately for training and testing datasets."*

Table 3*. Evaluation metrics for the three models considering train and test data. Model (a) with Sentinel predictors; (b) with Planetscope indices; and (c) with hydrological data.*

	Models	**Balanced accuracy**				**Overall accuracy**
		Wet	**Transition**	**Dry**	**Not Determined**	
Train	model (a)	0.88	0.50	0.84	0.81	0.80
	model (b)	0.86	0.64	0.84	0.82	0.78
	model (c)	0.87	0.75	0.85	0.85	0.80
Test	model (a)	0.90	0.50	0.80	0.77	0.78
	model (b)	0.93	0.74	0.86	0.89	0.84
	model (c)	0.88	0.76	0.88	0.89	0.83

Data & codes

Please include the data and codes to comply with ESurf's FAIR policies

Thank you for this reminder. In the revised manuscript we will include input data and code used in this study as supplementary material.