Manuscript title: Multivariate regression trees as an 'explainable machine learning' approach to explore relationships between hydroclimatic characteristics and agricultural and hydrological drought severity: Case of study Cesar River basin.

Author's general response:

We want to express our gratitude to the reviewers for their contributions to the review process of our paper. Your revision and constructive criticism have significantly improved the quality of our research. We appreciate the time and effort you invested in providing recommendations that have strengthened the clarity of our findings. In the following, you will find the answers to the general and specific comments. Some of them required a particular action or change in the manuscript. The changes we apply in the Revised Manuscript (RM) are in italics.

Response to Reviewer 1

General Comments.

1. I understand that not all the explicit thresholds can be easily explained, and I am not asking for a full explanation of both trees, but an attempt to explains some of the selection is advisable to better clarify the added value of this methodology over other 'not-explainable' ML. Some examples are reported in the detailed comments, but comments on why a certain threshold is meaningful (e.g. what a precipitation of 1600 mm represents for the region compared to the common climatology) may still be added for some of the key thresholds.

The authors thank the reviewer for the remark. We consider that the manuscript provides the necessary information to relate key thresholds to the basin climatology. For example, the study case description indicates the mean annual precipitation in the three climatic regions defined by the basin topography. Additionally, in section *3.2 Hydroclimatic drivers of droughts*, Figure 4 shows the multi-annual average of the numerical hydroclimatic drivers of droughts at each subbasin. Figure 4 presents an overall picture of the basin climatology and allows the reader to identify the areas showing the lowest and the highest values of the hydroclimatic parameters used in the study. In sections *3.4.1 Drivers of agricultural drought* and *3.5.2 Drivers of hydrological drought*, the authors describe the MVRTs and indicate whether the value falls above or below the basin's average for the key thresholds. These two sections were substantially improved in the second round of review by following the reviewer's recommendation on refine the description of the MVRTs and relate the key thresholds to the climatology of the basin. Accordingly, we believe that it is not necessary to include additional values of reference to relate the thresholds to the basin climatology.

2. The second comment is a follow up (clarification) to the previous review on model validation. You are using SSI as a drought stream flow index, so why don't you compare the simulated SSI (from SWAT) with the observed SSI (from stream gauge records)? The comparison between simulated and observed stream flow data during the dry season is not a proxy variable of the performance of your drought index, as many events are outside the dry season (see Table 6). I suggest to revisit this analysis.

Why don't you compare the simulated SSI (from SWAT) with the observed SSI (from stream gauge records)?

The authors clarify that the SWAT model does not simulate the SSI or any other drought indicator. The SSI was calculated by the authors using the results of the SWAT model, particularly the simulated streamflow at each subbasin. Regarding the question, it is not possible to calculate the SSI using gauge records because there is no observed data (gauging stations) at each subbasin. Hence, it is not possible to develop the comparison proposed by the reviewer (comparing the SSI calculated using simulated streamflow with the SSI calculated using gauge records). In the absence of observed streamflow at each subbasin, we consider that using simulated streamflow from the calibrated model is a valid approach to calculate the hydrological drought index at each subbasin.

The comparison between simulated and observed stream flow data during the dry season is not a proxy variable of the performance of your drought index, as many events are outside the dry season

The authors clarify that this comparison does not intend to assess the accuracy of the SSI calculated using simulated data. The comparison between the simulated and the observed streamflow is part of the calibration and validation process to evaluate the model performance for simulating streamflow. Calibration and validation aim to minimise the difference between model simulations and the observed data. This procedure reduces the uncertainty of the model results, and it can be asserted that a model with good performance indicators correctly simulates the basin hydrology.

The performance indicators used in the calibration and validation of the Cesar River Basin indicate that the model adequately simulates streamflow (in the wet and dry seasons, see Table 4 and 5); accordingly, the model results can be utilised to study the basin hydrology or develop further analyses, such as using the simulated streamflow to calculate the SSI.

3. Finally, a careful read of the text is recommended, as there are still many typos in the text (some are highlighted below).

The authors thank the reviewer's suggestion. Accordingly, we conducted an exhaustive revision of the entire manuscript to correct the typo errors in the text.

Minor comments

1. L13. lower than normal precipitation

Agreed. The paragraph is updated in the RM.

> *The typical drivers of drought events are lower than normal precipitation and/or higher than normal evaporation.* RM Ln 13.

2. L14-15. …influencing droughts.. influences droughts… Redundant. Please, reword.

Agreed. The paragraph is updated in the RM.

> *Evaluating the combined effect of the multiple factors influencing droughts requires innovative approaches.* RM Ln 14 to Ln 15

3. L18. Two drought indices.

Agreed. The paragraph is updated in the RM.

> *Model outputs, soil moisture and streamflow, are used to calculate two drought indices, namely the Soil Moisture Deficit Index and the Standardized Stream Flow Index.* RM Ln 17 to Ln 18.

4. L20-21. I suggest removing 'From the… Lastly,'

Agreed. The paragraph is updated in the RM.

> *Then, drought indices are utilised to identify the agricultural and hydrological drought events during the analysis period, and the indices categories are employed to describe their severity. Finally, the Multivariate regression tree technique is applied to assess the relationship between hydroclimatic characteristics and the severity of agricultural and hydrological droughts.* RM Ln 19 to Ln 21.

5. L22. Not only simulated. Precipitation is not simulated by SWAT. I suggest removing the part in parentheses.

Agreed. Please see answer to specific comment 4.

6. L28 …agricultural drought, whereas only limited precipitation…

Agreed. The paragraph is updated in the RM.

> *Precipitation shortfalls and high potential evapotranspiration drive severe agricultural drought, whereas limited precipitation influences severe hydrological drought.* RM Ln 24 to Ln 26.

7. L31. In my opinion, exposure is not the right term here.

Agreed. The paragraph is updated in the RM.

> *Moderate sensitivity to agricultural and hydrological droughts is related to the capacity of the subbasins to retain water, which lowers evapotranspiration losses and promotes percolation.* RM Ln 29 to Ln 30.

8. L51. Lower than normal precipitation

Agreed. The paragraph is updated in the RM.

> *Typically, droughts are triggered by atmospheric circulation and weather systems that combine to cause lower than normal precipitation and/or higher than normal evaporation in a region.* RM Ln 47 to Ln 48.

9. L53-59. I still found quite strange this early emphasis on such factors, which albeit very important, and not accounted (or play a very minor role) in your study.

The authors agree with the reviewer that not all the drivers of droughts listed play a significant role in this study; nevertheless, it is relevant to mention them since these factors influence droughts in other areas. In the introduction, we aim to provide an overall picture of all the potential drivers of droughts. Previous studies have demonstrated that the listed factors can enhance or alleviate a drought.

10. L103. 's'econd

We apologize for the mistake. The typo error is corrected in the RM.

> *The second is the analysis of droughts.* RM Ln 99 to Ln 100.

11. L103. …, soil moisture and stream flow, are…

We apologize for the mistake. The typo error is corrected in the RM.

> *SWAT outputs, soil moisture and streamflow, are used to calculate the drought indices, i.e., the Soil Moisture Deficit Index (SMDI) and the Standardized Stream Flow Index (SSI).* RM Ln 100 to Ln 101.

12. L113. Topography (panel a) and land use (panel b).

Agreed. The paragraph is updated in the RM.

> *Figure 1 presents the Cesar River basin's location, topography (Fig. 1a) and land use (Fig. 1b).* RM Ln 110.

13. L113. 72º53'W 74º04'W longitude and…

We apologize for the mistake. The typo error is corrected in the RM.

The basin is located between 72º53’W 74º04’W longitude and 10º52’00’N 7º41’00’’N latitude (Colombia). RM Ln 110 to Ln 111.

14. L123-124. This description does not match. If the rainy season is April-May, how is it possible that the main rainfall events occur in August November. I guess there is a typo somewhere here.

To prevent readers confusion, we clarity that basin’s annual rainfall pattern is bimodal.

The basin’s annual rainfall pattern is bimodal. The dry season occurs from December to April, followed by a rainy season from April to May. From June to July, precipitation decreases, and the main rainfall events occur between August and November. RM Ln 119 to Ln 121.

15. L132. Maybe a reference to Ramsar sites can be useful here for non-experts.

Agreed. The reference is included in the RM.

The Zapatosa marsh is recognised as one of the most important wetlands in the country, and considering the relevance of this ecosystem, it was declared a Ramsar site in 2018 (Ramsar sites are wetlands of international importance for containing rare or unique wetland types or for their relevance in conserving biological diversity). RM Ln 129 to Ln 130.

16. L134. Compared to… (pre-industrial period 1850-1900 if this is the case). Please clarify.

Agreed. The paragraph is updated in the RM.

In addition, climate change projections indicate that by 2070, the basin’s temperature may increase by 2.7°C, and precipitation may reduce by 10 % compared to the reference period 1971-2000 (Universidad del Magdalena et al., 2017). RM Ln 131 to Ln 133.

17. L146. The SWAT… as it seems to me that you are using the term to refer to one particular SWAT model (the USDA one). If you want to leave this as ‘A SWAT’, then you need to explicitly say that you are using one specific SWAT in the next sentence (i.e. the SWAT model developed by USDA is here used…).

Agreed. The paragraph is updated in the RM.

The SWAT model with an ArcSWAT extension was used to simulate the hydrological balance of the Cesar River. RM Ln 144 to Ln 145.

18. L146. To simulate the hydrological balance of the Cesar River basin.

Agreed. Please see answer to specific comment 17.

19. L149. Groundwater,

Agreed. The paragraph is updated in the RM.

The model is designed to simulate the quality and quantity of surface and groundwater and predict the environmental impacts of land management and climate change (Neitsch et al., 2011). RM Ln 146 to Ln 147.

20. L155. Is this average area correct? It does not seem to add up to 22,500 km2.

We apologize for the mistake. The error is corrected in the RM.

The model was built for the period from 1987 to 2018. The Cesar River basin was divided into 313 subbasins with a median area of 70 km^2. RM Ln 152 to Ln 153.

21. Table 1. since you are using Hargreaves, I am assuming that min and max daily temperature are used. Please clarify.

Agreed. The Table 1 is updated in the RM.

Data type	*Details*	*Source*
Digital elevation model	*25 × 25 m*	*Dataset ALOS PALSAR L1.0, Cartography 1:25000 Geographic Institute Agustín Codazzi (IGAC), Colombia*
Soil map	*300 × 300 m*	*Soil profiles Project GEF Magdalena–Cauca VIVE, GEF, BID, Fundación Natura, Colombia*
Land use map	*25 × 25 m*	*Land use map Geographic Institute Agustin Codazzi (IGAC), Colombia*
Daily precipitation and daily minimum and maximum temperature	*Period 1985–2018 (34 years)*	*Institute of Hydrology, Meteorology and Environmental Studies (IDEAM), Colombia*

22. L169. I am not familiar with this tool, but I am assuming that these parameters are calibrated spatially for each HRU or subbasins. Please add a short sentence to clarify how it works.

Agreed. We include a short comment on the model calibration in the RM.

We used the SWAT-CUP software package with Sequential Uncertainty Fitting version 2 (SUFI-2) for automatic model calibration and validation. SUFI-2 operates by performing several iterations. The calibration parameters are sampled in each iteration using the Latin hypercube technique against the objective function values (Abbaspour et al., 2018). RM Ln 161 to Ln 163.

In the calibration process, a physically meaningful range is set for each parameter in each iteration. Then, a new parameter value (within the range) is selected and applied at each HRU or subbasin. RM Ln 171 to Ln 172.

23. L185. Represents… representing. Please reword.

Agreed. The paragraph is updated in the RM.

> *The NSE is a dimensionless indicator ranging from -∞ to 1, with 1 representing a perfect match between the observed and simulated values (Moriasi et al., 2007).* RM Ln 184 to Ln 185.

24. L187. Low (in absolute value). Or closer to zero.

The following sentence was included in the RM.

> *The ideal PBIAS is 0, with low-magnitude values indicating accurate model simulation (Moriasi et al., 2007).* RM Ln 186 to Ln 187.

25. L200. If the index is scaled between -4 and 4, how is it possible to have a category -4 or less? Please check and clarify.

The authors deeply regret this mistake. As the reviewer points out, the minimum value that the SMDI can take is -4. It occurs when the monthly soil water available in the soil profile (SW_j) corresponds to the long-term minimum soil water available in the soil profile ($minSW_j$). We corrected the mistake in the RM.

> *and extreme drought (SMDI -4).*RM Ln 199.

26. L215. To reiterate on the previous comment. Here you correctly say 'mainly' ranges between -2 and 2. Anyway, I would reword as: zero average and unitary standard deviation.

We appreciate the reviewer's suggestion, but we consider that it is not necessary introduce changes into the sentence since it adequately indicates the SSI range.

27. L229. Short

We apologize for the mistake. The typo error is corrected in the RM.

> *The temporal threshold, it was used to avoid including short-term droughts (i.e., daily or weekly) in the analysis (Li et al., 2020).* RM Ln 229 to Ln 230.

28. L229. And minor and flash drought… This sentence (and also the next) is awkwardly worded. Also, you should check the actual definition of flash drought, which is not fitting what you are saying here (quickly developing not shortly lasting).

The reference to flash droughts is not included in the RM to avoid the readers' confusion, and the sentence is updated as presented in the answer to specific comment 27.

29. L279. As there are already many sub-sub-sub-sections, I suggest merging the next two (constrained…, and cross-validation) with the previous one (building the MVRT) and to reduce the text. You are here describing standard procedures, that are not specific of your study.

We like the reviewer's suggestion. The sections are merged in the RM. Regarding the description of the methodology, the authors consider that the description of the methodology is appropriate and presents the main elements of the technique applied. Accordingly, we do not introduce changes in the text.

> ***Building the MVRT: Constrained partitioning of the data and cross-validation***
>
> *Building the MVRT consisted of two processes: (1) the constrained partitioning of the data, and (2) the cross-validation of the results. The mvpart package run both processes in parallel. The two procedures are briefly explained below, and a more detailed description can be found in Borcard et al. (2018).*
>
> RM Ln 276 to Ln 278.

30. L322. The ability of SSI to correctly capture drought events depends on the accuracy of your SSI compared to the observations, and not the accuracy of modelled stream flow during the dry season (i.e., ability to capture the fluctuations across the entire year vs. ability to reproduce on average the flow during the dry period).

Please see answer to General Comment 2.

31. L358. Figure 7. The sum of the sub-basins classified does not match the total number of sub basins. Is this due to the wetlands? Please clarify. This is not the case for stream flow drought.

The reviewer's observation is correct. The wetland subbasins are not included in the agricultural drought analysis, which is why the sum is not equal to 313 subbasins. In section 2.2.2 Agricultural and hydrological drought analysis, we indicate: "*SMDI was not calculated for the subbasins that correspond to the Zapatosa marsh. In these subbasins, the predominant land cover is water. See Figure 5*" (RM Ln 211 to Ln 212). Additionally, in Figure 5's caption, the authors clarify that: "*SMDI was not calculated in the wetland subbasins (i.e. hatched area)*". We include the following sentence in Figure 7's caption.

> *Figure 7 MVRT of hydroclimatic drivers of agricultural droughts at the Cesar River basin and spatial distribution of the subbasins clustered at each leaf. Tree leaves are named from a to l, and n indicates the number of subbasins clustered at each leaf. The wetland subbasins are not included in the analysis for agricultural drought.*

32. L478. b, c and d should be in italic. Please check the entire text.

We apologize for the mistake. The error is corrected in the RM.

> Leaves *b*, *c*, and *d* corroborate the significant influence of evapotranspiration on agricultural drought severity. RM Ln 484 to Ln 485.

33. L482. How this value compares with the average basin actual ET. Are these subbasins with low or high actual ET?

In the section *3.4.1 Drivers of agricultural drought* we indicate: *The actual evapotranspiration threshold to split leaves c and d is 1,064 mm, value above the basin average (Figure 4c).* RM Ln 388 to Ln 389.

34. L491. Leaves shouldn't be in italic.

We apologize for the mistake. The error is corrected in the RM.

> In contrast, the MVRT outcomes suggest that a lack of precipitation is not a primary driver of agricultural drought in the subbasins clustered at leaves *e*, *f* and *g*. RM Ln 496 to Ln 497.

35. L500. Enhance the characteristics… I suggest rewording this sentence.

The sentence is updated as follows.

> *This agrees with earlier findings concluding that soil degradation enhances agricultural drought characteristics (Masroor et al., 2022; Santra & Santra Mitra, 2020; Trnka et al., 2016).* RM Ln 505 to Ln 506.

36. L546-547. This sentence is unclear to me.

The paragrap is updated in the RM.

> *The subbasins in terminal group g experienced the lowest median number of months for all hydrological drought categories (Fig. 10g). Low water yield indicates good water retention capacity in these subbasins. It can be explained by the proximity of the subbasins to the marsh (which acted as a natural control), the low slope in the area (which reduced streamflow velocity) and the presence of water bodies (which collected and stored runoff during the rainy season). The runoff threshold indicates that part of rainfall water reaches the streamflow; nevertheless, the subbasins in cluster g have one of the lowest runoff potentials in the basin (Fig. 4e). On the contrary, in these subbasins, percolation is considerably high (Fig. 4d). This seems to confirm that low susceptibility to hydrological droughts is linked to subbasins water retention capacity. The present findings suggest that the water storage capacity of the Zapatosa marsh*

can compensate for the increased evaporation that occurs during drought events, thereby alleviating hydrological drought severity upstream. Our results concur well with previous analyses concluding that wetlands (located in different climatic regions) significantly alleviate hydrological drought severity when direct evaporation from the water body does not significantly reduce water storage (Wu et al., 2023). RM Ln 551 to Ln 561.

37. L548. By the proximity of the subbasins to…

The paragraph is updated in the RM. Please see answer to specific comment 36.

38. L551. Is this referring to studies on this basin or studies on wetlands in general? Please clarify.

The paragraph is updated in the RM. Please see answer to specific comment 36.

39. L558. If the runoff is 'negligible', the reliability of SSI as drought index is very limited, as it is a drought index designed for rivers with flow. Please clarify the meaning of 'negligible' and highlight this limitation if needed.

The authors highlight that low runoff values do not necessarily imply streamflow ceasing during dry periods. It is important to keep in mind that runoff and baseflow are the two major components of streamflow. Low runoff values indicate that rainfall water infiltrates, and the baseflow sustains the streamflow between precipitation events. In the area of study, the streams hold water throughout the year; therefore, SSI is an adequate index to analyse hydrological droughts.

40. L571. Drought severity.

We like reviewer's suggestion. The sentence is updated in the RM.

In this area, drought severity was linked to inadequate rainfall partitioning and an unbalanced water cycle that favours water loss through evapotranspiration and low percolation values (Figs.7d, e, f and g, and Figs. 9b, c and d). RM Ln 576 to Ln 578.

41. L575. How high CNs increase hydrological drought? Streamflow should be higher for high CN, is this an effect of increased evaporation? Please argument.

We agree with the reviewer that high CN numbers are associated with high runoff potential. High CN values also indicate impervious surfaces and high runoff potential. This means less time is available for the water to percolate and recharge the groundwater. As a result, less water is available for sustaining baseflow in streams, contributing to lower streamflow during dry periods, when droughts are more severe and more likely to occur.

High CN numbers reduce the basin's water retention capacity and are linked to rapid streamflow increase during wet periods (flash floods) and lower streamflow during dry periods.

42. L604. Please add MVRT.

We like reviewer's suggestion. The sentence is updated in the RM.

> *In this study a machine learning technique, namely multivariate regression tree (MVRT), was applied.* RM Ln 609.

43. L607. I would stress that this is particularly true for agricultural drought.

Agreed.

> *Notably, the MVRT built for agricultural drought shows a good explanatory power.* RM Ln 613.

44. L614. In the subbasins in the middle course, drought severity…

We like reviewer's suggestion. The sentence is updated in the RM.

> *In subbasins in the middle course, drought severity is linked to inadequate rainfall partitioning and an unbalanced water cycle favouring water loss through evapotranspiration and low percolation values.* RM Ln 620 to Ln 622

45. L610-624. I suggest to further summarize this section, as many sentences are 1:1 re-proposition of previous paragraphs.

We appreciate the reviewer's suggestion. We consider that the paragraph adequately presents the main findings of this study, and we prefer not to summarise it.

46. L629. I suggest rewording to something like: …hydrological parameters influencing droughts based on threshold values that discriminate between different drought severity conditions.

We appreciate the reviewer's suggestion. The authors consider the sentence clear and prefer not to introduce the proposed change.

Response to Reviewer 2

General comment.

1. I acknowledge the responses from the authors for line-by-line comments. However, there are still very minor typos found in the revised manuscript. I urge the authors to carefully check it again.

The authors thank the reviewer's suggestion. Following the recommendation, we conducted an exhaustive revision of the entire manuscript to correct the typo errors in the text.

Specific comments.

1. P2L38: The authors may remove comma in between words significant and large-scale.

We apologize for the mistake. The error is corrected in the RM.

> *Upcoming soil moisture drought scenarios predict statically significant large-scale drying, especially in scenarios with strong radiative forcing in Central America and tropical South America (Lu et al., 2019).* RM Ln 35 to Ln 36.

2. P8L191-194: Redundant statement about 3 soil profiles. It was already explained before.

The sentence is reworded as follows.

> *SWAT calculates the soil water content of the entire soil profile. Three soil layers were identified in the Cesar River basin. The first layer thickness (vertical distance from the surface) reaches up to 350 mm, the second 1000 mm, and the third 1500 mm.* RM Ln 190 to Ln 192.

3. P10L230: J. Shah et al., 2022 -> Shah et al., 2022.

The reference is not included in the RM.

4. P10L234: Missing spacing for reference De'ath (2022).

We apologize for the mistake. The error is corrected in the RM.

> *MVRT is an extension of the popular regression tree (Breiman, 2001), but it differs in that it allows for multiple outputs (see De'ath (2002)).* RM Ln 232 to Ln 233.

5. P10L235: The authors may choose either technique or approach

We appreciate the reviewer's suggestion. Since technique and approach are synonyms, we prefer using both to prevent using the same word too much.

6. P13L297: Y here should be not capital.

To prevent readers confusion, we update the equation description in the RM.

> *and the denominator is the overall sum of squares of the response data*. RM Ln 298.

7. P13: Table 4: Move the words calibration and validation to the middle.

Agreed. The Table 4 is updated in the RM.

Gauging station	*Calibration*		*Validation*	
	NSE	*PBIAS [%]*	*NSE*	*PBIAS [%]*
Puente Salguero	*0.61*	*4.28*	*0.52*	*-8.3*
Puente Carretera	*0.50*	*-5.34*	*0.52*	*7.6*
Cantaclaro	*0.58*	*-11.30*	*0.50*	*-11.7*
Puente Canoas	*0.70*	*-1.34*	*0.57*	*10.64*

8. P16: Figure 4. Again please explain in the figure caption, what are soil types A, B, C, and D.

We apologize for the mistake. The caption is updated in the RM.

> *and k) soil type, the soil hydrologic groups A, B, C and D refer to the soil's infiltration characteristics.*RM Ln 340 to Ln 341.

9. P16: Table 6. Please explain if the drought duration here is total from moderate to severe or only for severe?

Agreed.

> *We identified the drought events and estimated their duration following the definition of droughts presented in 2.2.2. A month was summed to the duration of an event when a number of subbasins, covering at least 30 % of the basin's total area, were in a drought state (moderate, severe or extreme).* RM Ln 343 to Ln 345.

10. P18L365: Since it is a new paragraph, please indicate again that you refer to Figure 7.

Agreed.

> *The MVRT indicated that actual evapotranspiration was a strong driver of agricultural droughts; it appeared three times at different tree levels of split (Fig.7).* RM Ln 368 to Ln 369.