Response to Reviewer 1

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 would like to express our gratitude for your constructive suggestions and critical analysis in this second round of review. We appreciated the time you dedicate to the paper to ensure the clarity our research. We applied different changes to incorporate the reviewer’s suggestions and adequately present the results of our study. 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.

General Comments.

1. In my opinion, the definition of “drought severity” and the role of “multivariate” is still unclear for most of the paper. Even after reading the paper multiple times (in two rounds of reviews), the fact that “multivariate” refers to 3 “drought severity” values for each basin representing the total number of months in each drought class becomes clear to me only halfway through the text. Since modelling these quantities is the key goal of the study, I strongly recommend the author to clarify this definition early in the text. As an example, in the abstract “…drought events… and their severity”. The use of classes and duration is never mentioned, and the term “events” may be misleading in conveying the message that severity of multiple events are characterized instead of multiple categories. Also, the authors use “moderate”, “severe” and “extreme” as label for these classes, but the term “severe” severity is quite confusing, and it make the results and discussion sections difficult to read.

We thank the reviewer for this comment because it gave us the opportunity to clarify our manuscript. The definitions of drought severity and drought severity categories are now introduced early in the paper, and in consequence the abstract and the introduction have been modified. In the revised manuscript the abstract reads now as follows:

The Soil Water Assessment Tool (SWAT) is used for hydrological modelling. Model outputs, soil moisture and streamflow, are used to calculate the drought indices, namely the Soil Moisture Deficit Index and the Standardized Stream

*Flow Index. 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. From the identified droughts, the number of months for each drought severity category are sum up. Lastly, the Multivariate regression tree technique is applied to assess the relationship between hydroclimatic characteristics (represented by different simulated hydroclimatic parameters) and the severity of agricultural and hydrological droughts (represented by the number months in different drought severity categories).*RM Ln 17 to Ln 24.

Drought planning also uses research progress on drought characterisation. Using drought indices is a widespread methodology used for drought characterisation. (Zargar et al., 2011). Drought indices are computed numerical representations of drought severity (Hao & Singh, 2015; Keyantash & Dracup, 2002). Severity refers to the drought strength, also described as the deficit degree (Cavus & Aksoy, 2019), soil moisture deficit in the case of agricultural droughts and streamflow deficit in the case of hydrological droughts. Generally, severity is divided into different categories (e.g. moderate, severe, extreme), providing a qualitative assessment of the drought state in a region during a given period. Drought indices (and their categories) are crucial for tracking or anticipating drought-related damage and impacts (WMO & GWP, 2016). RM Ln 61 to Ln 69.

Regarding the comment on the severity categories, the authors agree with the reviewer that referring to severe severity may be confusing; however, this terminology comes from the indices, SMDI and SSI proposed by Narasimhan & Srinivasan (2005) and Modarres (2007), respectively. The authors named one of the indices' categories "severe", even when the indices measure "drought severity". We do not like this, but these categories are consistently used in drought studies, and we prefer not to modify the terminology.

2. The results and discussion sections are still lacking in the "explainable ML" component. The results section is mostly a mere description of Figs. 7-10, without any added values beside what can be inferred directly by the readers by looking at the plots. As an example, no connections are made between the leaves extracted by the MVRT and the obtained spatial patter. Is there any explanation on the grouping? Are all the

separations meaningful? Why some quantities are relevant for some basins on not for others? These are some of the insights that the authors should provide, in my opinion, to the readers to support the value of "explainable ML".

We would like to emphasise that Explainable ML refers to data-driven techniques that provide information on the model's decision-making process. We can say that a model is explainable if a human can tell how a model comes to a decision. MVRT is considered an explainable clustering technique since the method's outcome includes the clusters and the parameters used in the decision-making process of creating them. In this study, MVRT explainability presented in Figures 7 and 9 refers to the fact that the parameters (and the values) to create each one of the clusters are explicit. The authors consider that the explainable component of the MVRT was fundamental for establishing the relationship between the areas experiencing severe (and mild) drought and the parameters influencing that condition.

Regarding the description of Figures 7-10 in the results section, we agree with the reviewer that the description of the MVRT branches can be enhanced. Following the positive remark by the reviewer in specific comment 22, the description of Figures 7-10 is improved in the RM (Please see the answer to specific comment 22).

We address the specific comments on the "explainable ML component" in the following.

No connections are made between the leaves extracted by the MVRT and the obtained spatial patter

Assuming that by "connections", the reviewer refers to potential similarities or differences between the basins clustered at the same leaf, from the MVRT results, we can determine two main similarities: 1) it is possible to conclude that subbasins clustered at the same group have a comparable susceptibility to droughts (agricultural or hydrological), despite these subbasins are distant from each other, 2) it is possible to conclude that the average values of the parameters influencing droughts are similar in the subbasins clustered at the same group, although these subbasins are distant from each other.

Determine other potential similarities or differences that may connect distant subbasins clustered at the same leaf cannot be obtained from MVRT results since the technique is applied explicitly to cluster in the same leaf subbasins with similar susceptibility to droughts.

Is there any explanation on the grouping?

The grouping is the numerical result of the MVRT computation (see section Building the MVRT). Please note that MVRT is a data-driven technique, and building the tree relies entirely on generating multiple partitions of the data, calculating the standard deviation for each group created, and iteratively retaining the groups with the lowest standard deviation. The groups with low standard deviation are those that cluster subbasins with similar susceptibility to droughts (agricultural or hydrological).

Are all the separations meaningful?

From a numerical point of view, all the separations are meaningful since each retained leaf contributes to reducing the tree's relative error; in other words, the variance of the data explained by the tree is influenced by the number of separations. From the physical point of view, all the separations provide information on the drought generation process in the subbasins groped at each leaf. All the generated separations allow the discovery of possible unknown links between hydroclimatic parameters and drought severity.

Why some quantities are relevant for some basins on not for others?

From a physical point of view some parameter are relevant for some subbasins and not for others because the parameters influencing drought severity at each group of clusters are different. From a numerical point of view, some parameters are relevant for some subbasins and not for others because the explanatory variables that produce the partitions (groups) with the low standard deviation are different for each branch and leaf.

Specific Comments

1. L21. Clarify here what is actually modelled (number of months in multiple drought severity classes).

We like the reviewer suggestion. 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. From the identified droughts, the number of months for each drought severity category are summed up. Lastly, the Multivariate regression tree technique is applied to assess the relationship between hydroclimatic characteristics (represented by different simulated hydroclimatic parameters) and the severity of agricultural and hydrological droughts (represented by the number of months in different drought severity categories)* .RM Ln 19 to Ln 24.

2. L31. The interplay.

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

> *Results show that the presented methodology, combining hydrological modelling and a machine learning tool, provides valuable information about the interplay between the hydroclimatic factors that influence drought severity in the Cesar River basin.* RM Ln 33 to Ln 35.

3. L36. United States department of Agriculture?

The authors 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 38 to Ln 39.

4. L37. Drought severity, which is expected to…

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

> *A similar trend is predicted for hydrological drought severity, which is expected to increase by the end of the twenty-first century, with regional hotspots in central and western Europe and South America, where the frequency of hydrological drought may increase by more than 20 % (Prudhomme et al., 2014).* RM Ln 39 to Ln 42.

5. L75-77. Here you should clarify why drought severity is different and why a multi variable approach is needed.

The authors thank the reviewer for suggesting that the importance of the methodology should be presented in this paragraph. For this purpose, we updated the following paragraphs in the RM.

> *We have found two studies employing machine learning to assess the non-linear relationship between climate and basin processes and droughts (Konapala & Mishra, 2020; Valiya Veettil & Mishra, 2020). The studies reported relevant findings on the parameters driving droughts; however, the selected techniques showed a limitation for the drought analysis since they allow only one output variable. In both cases, it was necessary to apply the chosen technique multiple times to find the relationships between hydroclimatic parameters and the different categories of the evaluated drought characteristics. For example,*

Valiya Veettil et al. (2020) used a classification and regression tree (CART) to identify the variables influencing drought duration. CART allows one output variable; then, the authors applied the approach three times to evaluate the variables influencing short-term, medium-term and long-term drought events. Meanwhile, Konapala et al. (2020) used a random forest (RF) algorithm to identify the climate and basin parameters influencing the characteristics (duration, frequency and intensity) of three different drought regimes (long duration and mild intensity, moderate duration and intensity, short duration and high intensity). As the core of RF is a decision tree that allows one output variable (in this case, each characteristic of each drought regime), the authors repeated the procedure nine times, one for each drought regime and characteristic.

The aforementioned research shows the potential of machine learning techniques for drought-related analysis; nevertheless, it also suggests that assessing the parameters driving drought characteristics requires techniques capable of simultaneously handling the different categories of drought characteristics. Commonly used in ecology to relate independent environmental conditions to populations of multiple species, the Multivariate Regression Tree (MVRT) arises as a suitable technique for this purpose. .RM Ln 76 to Ln 98.

6. L85. You introduce here "individual categories of drought severity" but never clarify the concept.

The definition of drought severity is presented in the RM in Ln 61 to Ln 67. To address the reviewer's comment, we enhanced the definition of severity in the RM.

Drought planning also uses research progress on drought characterisation. Using drought indices is a widespread methodology used for drought characterisation. (Zargar et al., 2011). Drought indices are computed numerical representations of drought severity (Hao & Singh, 2015; Keyantash & Dracup, 2002). Severity refers to the drought strength, also described as the deficit degree (Cavus & Aksoy, 2019), soil moisture deficit in the case of agricultural droughts and streamflow deficit in the case of hydrological droughts. Generally, severity is divided into different categories (e.g. moderate, severe, extreme), providing a *qualitative assessment of the drought state in a region during a given period. Drought indices (and their categories) are crucial for tracking or anticipating drought-related damage and impacts (WMO & GWP, 2016).* RM Ln 61 to Ln 67.

7. L87. Why is it important to highlight "supervised" What is the role of supervision here?

It is important to highlight that MVRT is a *supervised clustering* technique because this type of technique uses explanatory variables and response variables to create clusters. That is to say, it uses two datasets. The dataset of response variables is recursively divided using the set of explanatory variables to build the tree and create the clusters. Another technique, *Unsupervised clustering*, only uses one dataset to create the clusters. That is the reason why is important to highlight that this is a supervised technique.

8. L99. "drought events… describe their severity". This gives the impression that the severity of each drought event is modelled, and that the multiple variables are the severity of multiple events. Please reword.

We improved the description of the response variable 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). Drought indices are utilised to identify the agricultural and hydrological drought events in the analysis period. Then, we calculate the months for each drought severity category during the observed droughts. Finally, the MVRT approach is applied to assess the relationship between hydroclimatic characteristics (represented by the simulated parameters in each subbasin) and drought severity categories (represented by the total number of months for each drought severity category in each subbasin). RM Ln 103 to Ln 109.

9. Table 1. Discharge is not an input data, but rather used for calibration. As it is listed here it can be misleading.

The authors agree with the reviewer that the information on the discharge should not be included in Table 1. In the RM, details on the discharge data are presented in the "Model calibration and validation section".

Table 1. SWAT model input data

Data type	Details	Source
Digital elevation model	25 × 25 m	Dataset ALOS PALSAR L1.0, Cartography 1:25000 Geographic Institute Agustín Codazzi (IGAC), Colombia

Data type	Details	Source
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
Rainfall and temperature daily data	Period 1985–2018 (34 years)	Institute of Hydrology, Meteorology and Environmental Studies (IDEAM), Colombia

The model was calibrated from 1985 to 2002 and validated from 2003 to 2018 using the streamflow series from four stream gauges (Figure 1). The source of the discharge data is the Institute of Hydrology, Meteorology and Environmental Studies (IDEAM), Colombia. RM Ln 177 to Ln 178.

10. L163-164. Please correct the reference.

The authors apologize for the mistake the reference is updated in the RM.

Based on expert judgment and the available literature (Arnold et al., 2012; ASABE, 2017), the following SWAT parameters were used in the calibration and validation process. RM Ln 168.

11. L182. Please report there the reasoning behind the use of SMDI rather than soil moisture anomalies (as stated in the replies to the reviewers).

Agree. The criteria to choose SMDI is included in the RM.

The present study used the soil moisture deficit index (SMDI) to analyse agricultural droughts. We chose this index since it was developed to use simulated soil moisture as input parameter, particularly the SWAT simulated soil moisture in the soil profile at each subbasin (Narasimhan & Srinivasan, 2005). RM Ln 189 to Ln 190.

12. L216-217. The values adopted to define a drought event seem rather arbitrary. Please provide some support to these assumptions.

Agreed. Information about the criteria to define the spatial and temporal minimum thresholds is included in the RM.

It is worth highlighting that the minimal extension of a drought is not defined, but it is accepted that droughts typically occur on a large scale (Sheffield & Wood, 2011b). Setting a spatial threshold is a common practice to maintain a minimum drought-affected and prevent identifying isolated areas experiencing *dry spells as drought events (Brunner et al., 2021). Regarding the temporal threshold, it was used to avoid including sort periods of water shortage and minor and flash droughts in the analysis. These are events that occur within days or weeks (J. Shah et al., 2022).* RM Ln 225 to Ln 230.

13. L231. The response multiple variables are the number of months observed in the three drought severity categories.

Indeed, the multivariate response is the number of months in the three drought severity categories. We update the sentence in the RM.

> *The multivariate response is the number of months observed in the three drought severity categories (moderate, severe and extreme) at each subbasin.* RM Ln 241 to Ln 243.

14. L254. Outputs or inputs (e.g. precipitation is an input).

The sentence is updated using the word "results" to prevent readers from confusion between the outputs of the hydrological model and the inputs of the MVRT technique.

> *The averages were computed using the SWAT model results at each subbasin.* RM Ln 263 to Ln 264.

15. L310. Low flow conditions using SSI are not limited to the dry season (see table 6, with some events during winter months). Why not reporting correlation on SSI values?

We agree with the first part of the statement; drought events occur in the dry and rainy seasons. Regarding the correlation on SSI values, assuming the reviewer refers to the correlation between the streamflow (in the dry and rainy season) and the SSI values, the correlation is high since streamflow is the input variable used to calculate the index. The authors consider that presenting the correlation between these two variables does not provide new insights into the objective of this study.

16. L321. Or inputs. Again, precipitation and potential ET are forcing.

The sentence is updated using the word "results" to prevent readers from confusion between the outputs of the hydrological model and the inputs of the MVRT technique.

> *Figure 4a to h presents the multi-annual average of the numerical hydroclimatic drivers of droughts at each subbasin. The average was calculated using the hydrological model's results during the simulation period (1987 to 2018).* RM Ln 332 to Ln 334.

17. L330. It would be good to have some more details on the events reported. Is there any info on the severity? Do they align well with your modelled classes? What about spatial patters?

The authors agree with the reviewer that it would be interesting to contrast the results; nevertheless, the authors did not find information about drought severity in the National Study of Water. This study is conducted nationally and covers diverse topics associated with water availability in the country. Regarding droughts, the study focuses on the chronology and the duration of drought events in Colombia. Accordingly, we used the available information to compare the drought periods and their duration.

18. L335. Each subbasin, as represented in Figures 5 and 6 for agricultural and hydrological droughts, respectively.

We like the suggestion. In the RM, we applied the reviewer's suggestion.

> *After identifying the agricultural and hydrological drought events, it was possible to determine the number of months for each drought category in each subbasin, as represented in Figures 5 and 6 for agricultural and hydrological droughts, respectively.* RM Ln 348 to Ln 349.

19. Figure 5. The colour scale is difficult to read, especially for the extreme category.

The color scale of Figures 5 and 6 is improved in the RM.

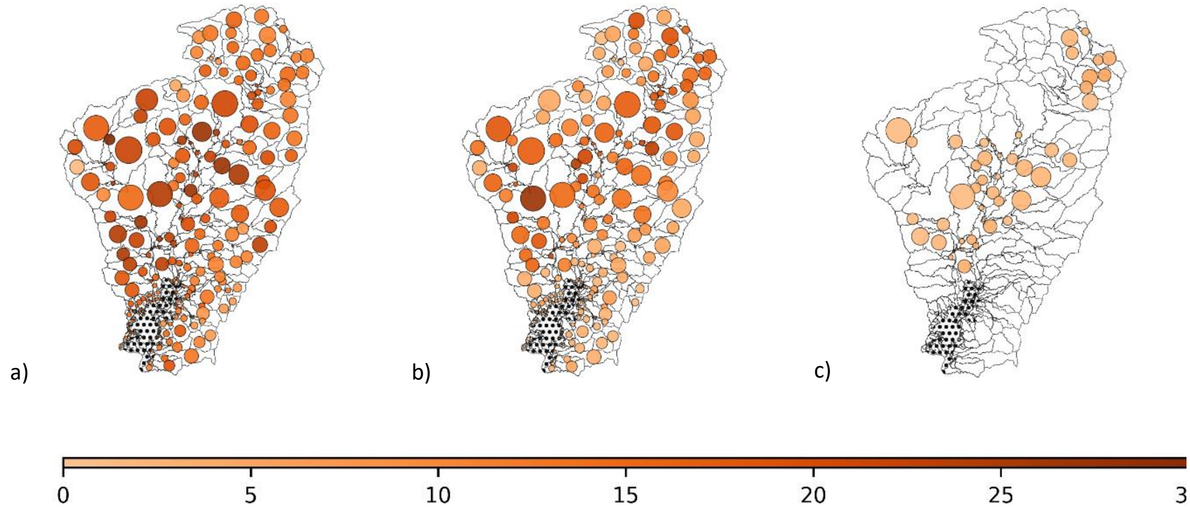

Figure 1 Months counted in each agricultural droughts category: a) moderate, b) severe and c) extreme. SMDI was not calculated in the wetland subbasins (i.e. hatched area).

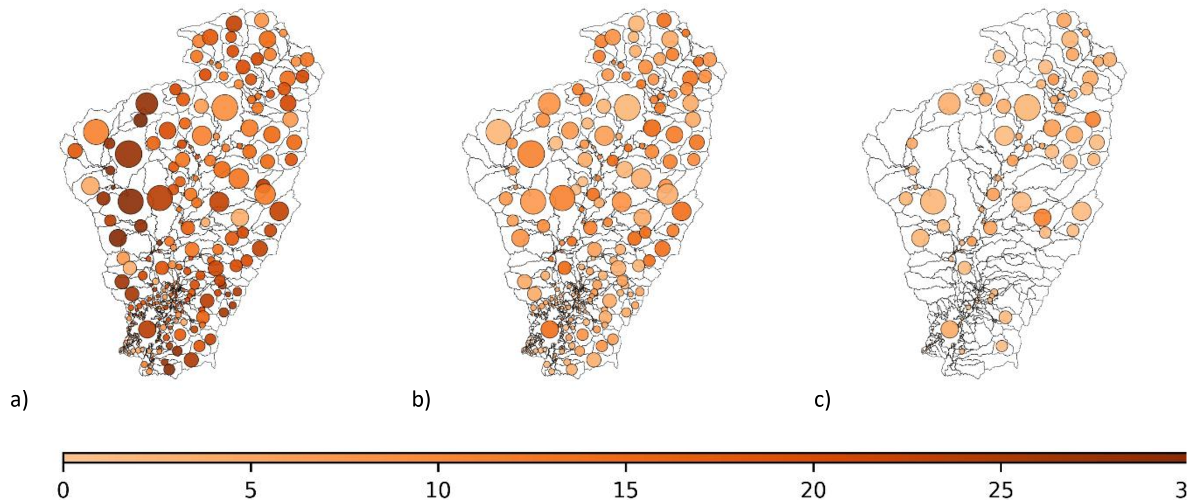

Figure 2 Months counted in the hydrological droughts category: a) moderate, b) severe and c) extreme.

20. L351. Actual evapotranspiration.

The line is updated in the RM.

> *The MVRT indicated that actual evapotranspiration was a strong driver of agricultural droughts; it appeared three times at different tree levels in the splitting rules.* RM Ln 365 to Ln 366.

21. L360. More can be said on the differences between some classes. As an examples, h, I and j looks very similar in terms of frequencies in the three classes. Are the differences modelled by MVRT justified?

The authors agree with the reviewer that the severity in the subbasins clustered at groups *h, i* and *j* is similar, and the variables influencing the severity in clusters *h* and *i* are the same. Considering the severity is alike for clusters *h* and *i,* it can be asserted that these subbasins could be clustered into the same leaf. The method to merge leaves is "pruning" the tree. Tree pruning combines the leaves starting from the leaves produced at the lower levels of split. If we had pruned the tree to merge leaves *h* and *i*, the leaves at the fifth level of split (*e* and *f*) would have become one; then the leaves at the fourth level of split *c, d, h, i, k* and *l* would have been merged. The criteria to decide what leaves to merge is determining the combination of leaves that produces the group with the lowest standard deviation. The authors highlight that pruning the tree increases the CVRE and compromises the explained variance (EV); that is to say, the tree reduces its explanatory power.

The authors opted for retaining the 12 leaves of the tree because the clusters produced at the fourth and fifth level of split provide relevant information on the links between hydroclimatic parameters and the drought severity categories (e.g. leaves *e* and *f*). We consider that the similarities between the leaves *h* and *i* do not compromise the results of this study. On the contrary, retaining these two clusters allowed the authors to extract relevant information from other clusters that would have disappeared when combining leaves *h* and *i* into one cluster.

We included an observation about this two clusters in the section *4.4 Accuracy of the MVRTs*

> *The high EV (0.81) value reflects the good explanatory power of the tree built for agricultural drought. This confirms that the selected explanatory variables significantly influence the severity of agricultural drought. Nevertheless, two potential disadvantages of the tree are identified. First, clusters h and i are very similar. Drought severity is alike in these leaves, and the parameters influencing droughts are the same. This suggests that these two clusters can be merged into one cluster. Second, leaves b and k cluster two subbasins. Accordingly, the distribution presented in the boxplots must be interpreted cautiously. Neither of these disadvantages compromises the study's main findings; however, further analysis is recommended to determine the size of the tree (number of clusters) that better fits the objective of this study.* RM Ln 585 to Ln 591.

22. L361-367. This description of classes a and b does a fir job at introducing the main drivers. The results for the other classes (as well as the ones for hydrological drought) should follow a similar structure.

The authors agree with the reviewer that there is still room to improve the description of the built trees for agricultural and hydrological droughts. The description is improved in the RM.

For agricultural droughts:

> *Leaves c and d cluster twenty-four and nineteen subbasins, respectively. Leaf c groups subbasins located in the upper part of the river course and the basin east. Precipitation was slightly below the basin average in the subbasins located in the north and was close to the average in subbasins in the east (Figure 4a). Leaf d groups subbasins located in the upper course of the river and in the basin's western part. The actual evapotranspiration threshold to split leaves c and d is 1,064 mm, value above the basin average (Figure 4c). For subbasins with actual evapotranspiration below the threshold, leaf c, the median of months in the severe drought category is below ten (Figure 8c). For subbasins with actual evapotranspiration above the threshold, leaf d, the median of months in the severe drought category is sixteen, one of the highest among the terminal groups (Figure 8d).*

Leaves e, f and g cluster twenty-four, six and twelve subbasins, respectively. Subbasins are located in the river valley and the basin's western part. In these subbasins, precipitation was below the basin average (Figure 4a), and actual evapotranspiration was above the average (Figure 4c). The percolation threshold to split leaves e and f from leaf g is 111 mm, a value considerably below the basin average (Figure 4d). At the fifth level of split, the sediment yield threshold to split leaves e and f is 101 metric tons/ha, a value close to the average in the basin (Figure 4h). Figure 8e, f and g show that subbasins clustered in these leaves are prone to agricultural droughts. The median of months in the moderate drought category was above twenty months; the severe category was above ten months; and the three leaves exhibited months in the extreme drought category.

Leaves h, i and j cluster twenty-six, fifty-two and fifty-six subbasins, respectively. Subbasins are mainly located in the wetland surroundings, La Serranía (leaf i), and some outliers are located in the basin's north (leaves h and j). Percolation in leaves h, i and j was close to the basin average (Figure 4d). Actual evapotranspiration in terminal groups h and i was relatively close to the basin average (Figure 4c). The water yield threshold to split clusters h and i is 352 mm. Overall, subbasins clustered at leaves h, i and j presented low susceptibility to severe and extreme agricultural drought conditions. The median of months in the moderate drought category was slightly higher than ten; the median for months in the severe category was the lowest for the study area and showed no months in the extreme drought category (Figure 8h, i, j).

Leaves k and l cluster two and six subbasins, respectively. Subbasins are located towards the basin's north, and one outlier is observed in the subbasin east (leaf l). In these subbasins, percolation was lower than 271 mm, value relatively low compared to other basin areas (Figure 4d). In leaf k, the curve number was lower than sixty-seven, while in leaf l, it was higher. In leaf k, the median of months for the moderate category is ten, and for the severe category, *it is above 10. In leaf l, the median of months in the moderate category is above ten, and the subbasins experienced some months in severe drought. Leaves k and l show no months in the extreme drought category (Figure 8k and l).*RM Ln 380 to Ln 409.

For hydrological droughts:

Leaf a clusters twenty-eight subbasins in the upper basin and one outlier located in the western part of the subbasin (Figure 9a). In these subbasins, precipitation was considerably below the basin average (Figure 4a). Figure 10a shows that the subbasins in this terminal group repeatedly experienced moderate, severe and extreme hydrological drought.

Leaves b and c cluster thirty-seven and thirteen subbasins, respectively. Subbasins clustered at leaf b are relatively distant; most are towards the eastern part of the basin, and the rest are in the north and west of the basin. Subbasins in leaf c are located in the river's middle course towards the western part of the basin and some outliers in the north. Precipitation and percolation were slightly above the basin average in subbasins clustered at leaves b and c (Figure 4a and d). The curve number threshold to split leaves c and d is 51. Subbasins with a curve number above the threshold, leaf b, experience months in extreme drought and present one of the highest median of months for severe drought (Figure 10b). For subbasins with curve number below the threshold, leaf c, the median of months at moderate drought is almost 20 and experience months at severe and extreme category (Figure 10c).

Leaf d clusters twenty-nine subbasins in the river's middle course and the basin's eastern part. Figure 10d indicates that in this terminal group, the subbasins experienced fewer months in the severe and extreme drought categories than the other clusters in the tree's left branch; however, subbasins experienced one of the highest median of months at moderate drought.

In leaves e (n = 72) and f (n = 18), precipitation and water yield exceeded the basin average (Figure 4a and g). The actual evapotranspiration threshold to split leaves e and f is 833 mm, value below the basin average (Figure 4c). Both *terminal groups describe moderate exposure to hydrological drought. At leaf e, the median of months in the severe and extreme drought categories is below ten, while the median of months in the moderate drought category is twenty (Figure 10e). The hydrological drought exposure of the subbasins clustered at leaf f is also mild. In these subbasins, actual evapotranspiration is above the threshold and close to the basin average. These subbasins present the lowest median of months for all drought categories (Figure 10f). Notably, the Zapatosa marsh and upstream subbasins are clustered in this terminal group (Figure 9f).*

Leaves g and h cluster seventy-one and forty subbasins, respectively. Subbasins clustered at these leaves are located upstream of the Zapatosa marsh. The surface runoff threshold to split the leaves g and h is 0.5 mm. Figure 10g shows that the subbasins grouped at leaf g present the low suceptibility to hydrological drought. The median of months for all categories is the lowest in the basin. In leaf h, the surface runoff was lower than 0.5 mm. In these subbasins, the medians of months in the severe and extreme categories are relatively low, while the median of months in the moderate category is eighteen (Figure 10h).

RM Ln 430 to Ln 459.

23. Figure 8. The use of box plots for some leaves is questionable when very few subbasins are included (< 6). Please revise this figure, or highlight the limitations is these cases.

The authors agree that using box plots for groups clustering a small number of subbasins may be debatable. We included an observation about these two clusters in section 4.4 Accuracy of the MVRTs (Please see the answer to specific comment 21).

24. Figures 8 and 10. Differences in term of extreme severity are difficult to judge, due to the low magnitude compared to the other classes. Please consider representing these data in a different way or to use a secondary axis.

We thank the reviewer for pointing this out since it allowed us to notice that for some clusters (7b, h, i, j, k, l) the blox-plot presented “zero” months in the extreme as a significant value. Accordingly, we corrected the error. We did not apply additional changes to the Figures. We consider that the Figures adequately present the necessary information to discuss our results.

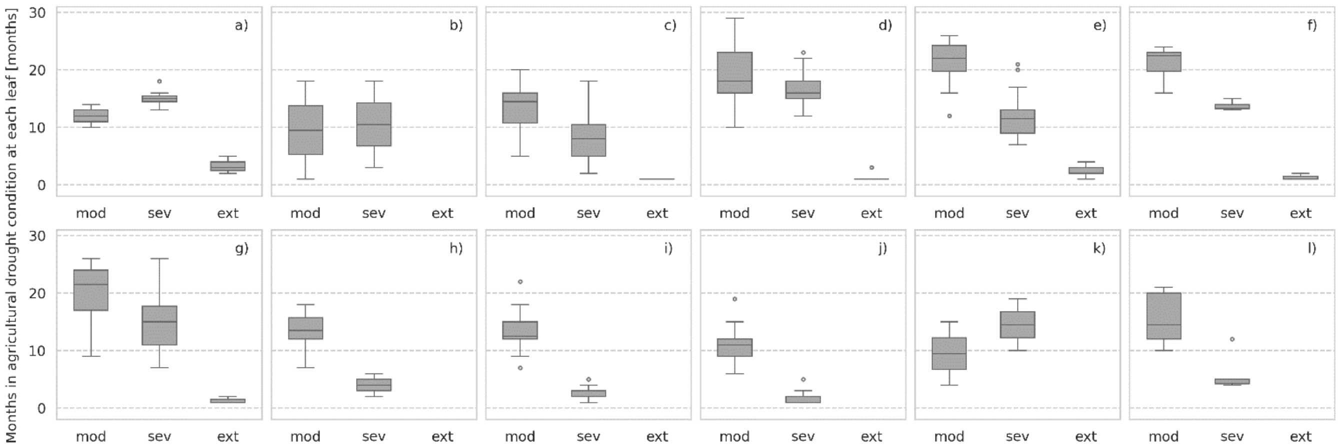

Figure 3 Number of months in agricultural drought categories (moderate, severe, extreme) at each leaf. Tree leaves are named from a to l.

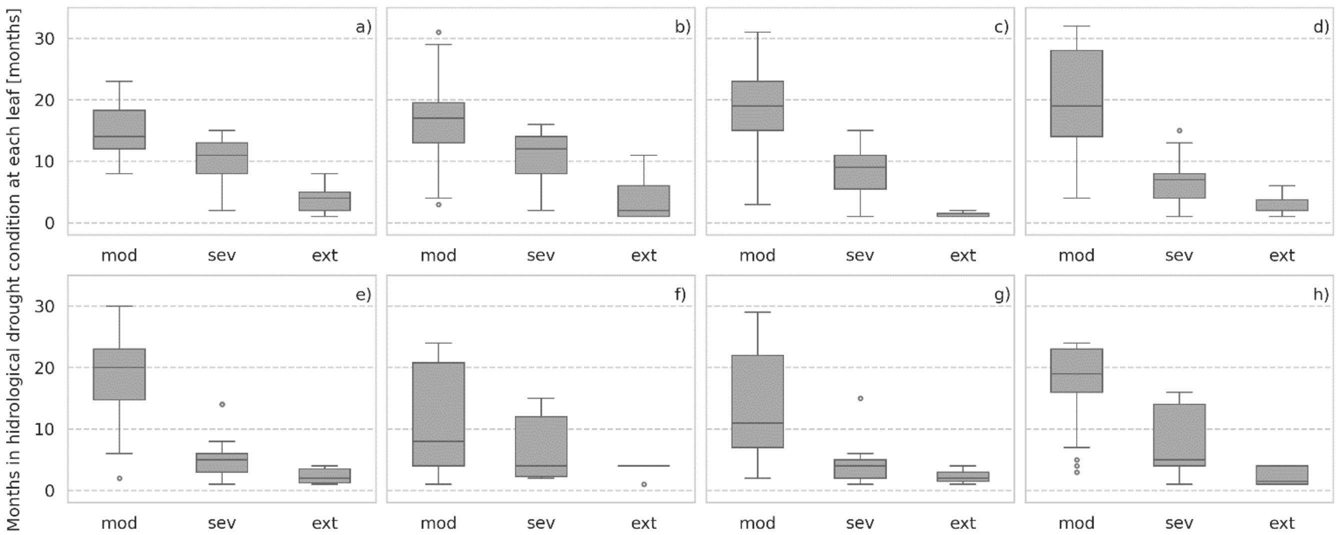

Figure 4 Months in hydrological drought categories (moderate, severe, extreme) at each leaf. Tree leaves are named from a to h.

25. Figures 7 and 9. Please add some comments on the spatial patters and also on outliers. Some examples from Fig. 9: what is the difference between basins in class a (in the north) and class e (in the south) in terms of hydrology? Why is an isolated basin included in class a (is it just a numeric problem or is there any hydrological explanation)?

Comments on the spatial allocation of the subbasins grouped at each leaf were included in sections *3.4.1 Drivers of agricultural drought and 3.4.2 Drivers of hydrological drought* (Please see answer to Specific comment 22). The comments on the subbasins hydrology focus on the parameters influencing droughts at each leaf.

It is important to reiterate that according to the MVRT results, subbasins clustered in the same group have a comparable susceptibility to droughts (agricultural or hydrological) despite these subbasins being distant. In addition, the average values of the parameters influencing droughts are similar in the subbasins clustered in the same group, although these subbasins are distant. Thus, the clusters' outliers do not result from a numerical error.

26. Figure 10. The caption refers to agricultural drought.

The authors apologize for the mistake. The typo error is corrected in the RM. Please see answer to specific comment 24.

27. Discussion: classes b, c, and d are never mentioned.

Agreed. Clusters b, c and d are mentioned in the RM.

> *Leaves b, c, and d corroborate the significant influence of evapotranspiration on agricultural drought severity. A comparison of clusters a and b, and c and d indicates that the leaves with higher evapotranspiration are more prone to experience severe drought. It is interesting to notice that in clusters c and d, the actual evapotranspiration threshold causes a notable difference in drought severity. While the leaf c, clustering subbasins with actual evapotranspiration below 1064 mm, presents the lowest median of months at severe category at the left branch of the tree, leaf d shows the highest median of months at the same category in the tree.* RM Ln 478 to Ln 483.

28. L446. "…Severity in leaves a, e, f and g was comparable…". Is this true? Are they more similar then the other classes?

Indeed, the agricultural drought severity in leaves a, e, f and g is comparable (more similar if compared with the classes). The authors highlight that the similarities specifically refer to the agricultural drought susceptibility observed in the subbasins clustered at these leaves.

From Figure 8, it is possible to see that subbasins clustered in leaves *a*, *e*, *f* and *g* experienced months in extreme drought category; the median of months for severe drought category is above ten months, and for moderate category, the median is above 20 months in leaves *e*, *f* and *g*. Although the values are not exactly the same, a similar trend is observed. Although the drought situation is comparable, the drivers are different, as presented in tree built for agricultural droughts (Figure 7). Our results show how the influence of different hydroclimatic parameters can lead to comparable drought susceptibility like leaves *a*, *e*, *f* and *g*. This is relevant information for understanding the drought-generating process in the Cesar River Basin.

29. L457. Is this true only for leaf e?

The authors agree with the reviewer that the statement is not only valid for leaf *e*. It is also true for leaves *f* and *g*. The paragraph is updated 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. Particularly, leaf e grouped the subbasins that experienced the most severe agricultural drought in the analysis period.* RM Ln 490 to Ln 492.

30. L457. "the most severe…" This is one example where the term severe (referring to severity) may be confused with severe as category.

Please see answer to general comment 1.

31. L485. I would avoid the use of "expose" here, as it refers to something different in risk analysis.

To prevent inadequate use of the terminology, the sentence t is updated in the RM

> *The subbasins clustered on the left branch of the tree were prone to hydrological drought (Figure 10a, b, c, d).* RM Ln 519.

Other paragraphs of the paper were the word was used are also corrected.

> *The scattering of the outputs in each leaf allows us to identify the susceptible subbasins to agricultural droughts.* RM Ln 363
>
> *This information allowed us to identify the clusters of subbasins prone to hydrological droughts.* RM Ln 423 to Ln 424.
>
> *Figure 10a shows that the subbasins in this terminal group repeatedly experienced to severe and extreme hydrological drought.* RM Ln 434.
>
> *The left branch of the MVRT clusters the subbasins susceptible to severe agricultural drought (Figure 8a, e, f and g).* RM Ln 470.

32. L563. Conclusions.

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

33. L572. "MVRTs indicate… course of the river". This can be inferred from the severity data (see Fig. 6) even without the need of MVRT. Please rephrase.

The authors agree with the reviewer. We updated the conclusion as follows:

> *The outcomes of the MVRT provide valuable information on the hydroclimatic parameters influencing the drought-generating process in the Cesar River basin. MVRTs indicate that severe agricultural and hydrological*

drought conditions observed in the upper and middle course of the river are influenced by different hydroclimatic factors. The interaction between precipitation shortfalls and high potential evapotranspiration drives severe agricultural drought in the headwater. Conversely, severe hydrological drought condition is mostly caused by limited precipitation. In subbasins in the middle course, droughts' severity is linked to inadequate rainfall partitioning and an unbalanced water cycle favouring water loss through evapotranspiration and low percolation values. Notably, results suggest that poor soil structure enhances severe agricultural drought conditions, and high curve numbers seem to increase hydrological drought severity. RM Ln 610 to Ln 617.

34. L574. Solely mostly.

We like the suggestion. In the RM, we use "mostly". Please see answer to the minor comment 33.

35. L588. Vulnerable is not the right term here. Also, as mentioned above, the area with most frequent drought can be inferred even without MVRT. Please rephase and emphasise the added value of MVRT is detecting the drivers.

The authors agree with the reviewer. The conclusion is updated as follows:

It can also be concluded that the MVRT (and other machine learning techniques that generate 'explainable AI' models based on progressive tree-like data partitioning and simplified models in leaves) is a relevant tool for defining drought management strategies. The tool helps to identify drought-prone areas and design management strategies that contribute to maintaining the hydrological parameters influencing droughts above (or below) the thresholds that trigger severe and extreme drought conditions. RM Ln 625 to Ln 629.

Response to Reviewer 2 Comments

Manuscript title: Multivariate regression trees as an 'explainable machine learning' approach to exploring relationships between hydroclimatic characteristics and agricultural and hydrological drought severity

Author's general response:

The authors would like to thank the reviewer for thoroughly reviewing the manuscript. We are particularly grateful for the time dedicated to reviewing the Results section and finding many typo errors that were compromising the description of the outcomes of our analysis. We have considerably improved the trees' description and corrected the errors identified by the reviewer.

General Comments.

1. Thanks for addressing my comments. However, I would like to raise a concern regarding the explanation provided on the use of the Standardized Soil Moisture Index (SSMI or SSI and ESMI in some papers) that is not correct. The authors stated that 1) SSMI is an agricultural drought index derived from daily satellite data, 2) the index is developed for short-term drought monitoring, and 3) there is no previous assessment of the index performance using simulated SM as the input parameter. All these three arguments are misleading. First, the SSMI is a standardized drought index like SPI (for precipitation), SSI (for streamflow), SGI (for groundwater). It uses monthly data instead of daily data (e.g., Ndehedehe et al., 2016; Carrão et al., 2016; Das et al., 2022). These publications used monthly soil moisture data to derive SSMI. Second, SSMI has been employed not only for drought monitoring but also for drought forecasting in some studies (e.g., AghaKouchak, 2014; Xu et al., 2018). Last, some of the aforementioned publications have indeed utilized models to simulate soil moisture variable used in drought identification (SSMI). The second paragraph is accepted.

The authors thank the reviewer for providing additional insights about the SSMI. We reviewed the references and realised some studies had used simulated soil moisture to compute the SSMI. We apologise for not conducting more exhaustive research on the SSMI applications and computation principle. However, we consider that SMDI is still a good choice since the authors of SMDI developed the index to use simulated soil moisture as input parameter. The good results for analysing the hydroclimatic drivers of agricultural droughts confirm that the index was appropriate for representing the severity of agricultural droughts.

2. Thanks for your explanation. The authors may consider to move the PCA analysis in the appendix or supplementary material instead of delete it. PCA analysis is still useful to indicate the variance.

The authors appreciate the reviewer's suggestion. However, we prefer to leave the results of the PCA analysis out. The authors consider that for this study, the most significant variance is the proportion of explained variance (EV) calculated for each tree. This parameter defines the explanatory power of the tree and is presented in the *Results* section and discussed in the section *Accuracy of the MVRT.*

New line by line comments.

1. P1L1: Suggestion for 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".

We like the suggestion. The title is updated accordingly in the RM.

2. P1L17: The authors may write "(SWAT)" here.

The abbreviation is included in the RM.

> *The Soil Water Assessment Tool (SWAT) is used for hydrological modelling.* RM Ln 17.

3. P1L18: Suggested text revision: ….the drought indices namely Soil Moisture….

We like the suggestion. The abstract is updated accordingly in the RM.

> *Model outputs, soil moisture and streamflow are used to calculate the drought indices, namely Soil Moisture Deficit Index and the Standardized Stream Flow Index.* RM Ln 17 to Ln 19.

4. P2L59-60: Rephrase this sentence. It is unclear.

The sentence is rephrased in the RM.

> *Using drought indices is a widespread methodology for drought characterisation.* RM Ln 61 to Ln 62.

5. P3L83: Mentioned -> The aforementioned research

The sentence is updated in the RM.

> *The aforementioned research shows the potential of machine learning techniques for drought-related analysis; nevertheless, it also suggests that assessing the parameters driving drought characteristics requires techniques*

capable of simultaneously handling the different categories of drought characteristics. RM Ln 89 to Ln 91.

6. P4L98: Between indices and Soil Moisture Deficit Index, the author my write either "which are" or "i.e.,"

We like the suggestion. The sentence is improved in the RM.

> *The Second is the analysis of droughts. SWAT outputs, soil moisture and streamflow are used to calculate the drought indices, i.e., Soil Moisture Deficit Index (SMDI) and the Standardized Stream Flow Index (SSI).* RM Ln 103 to Ln 104.

7. P10L231: Instead of "(the drought indices give categories)" -> "(moderate, severe, and extreme)"

We like the suggestion. The sentence is updated in the RM.

> *The multivariate response is the number of months observed in the three drought severity categories (moderate, severe and extreme) at each subbasin.* RM Ln 241 to Ln 243.

8. P10L234: What do the authors mean with four technique attributes are relevant to this study?

The authors thank the reviewer for the question since it allows us to notice that referring to "four technique attributes" can be misleading. In the sentence, we want to highlight the technique attributes that make it suitable for this study. The sentence is rephrased in the RM.

> *The following MVRT attributes are relevant for this study.* RM Ln 245.

9. P10L238: The authors may remove "The drought indicators give these three categories to represent the drought severity". It is redundant.

Agreed. The sentence is not included in the RM.

> *MVRT's capability to handle multiple outputs allowed us to evaluate the influence of the hydroclimatic parameters on moderate, severe and extreme drought conditions simultaneously (response variables).* RM Ln 247 to Ln 249.

10. P12L274: What is SS?

SS is the within-group sums of squared distances to the group means. To prevent readers confusion the paragraph is updated in the RM.

> *Second, for each partition, it was calculated the resulting sum of within-group sums of squared distances to the group means for the response data*

(within-group SS). Within-group SS is equivalent to standard deviation. RM Ln 282 to Ln 283.

11. P13L310: Please give a low flow definition here. How do the authors identify low flow? Is it using a threshold method?

The authors thank the reviewer for the question. We realised that using the term low flow is inaccurate. We did not develop an analysis to determine low flows. The results in Table 5 correspond to the model performance simulating the discharge in the dry season. To adequately describe the analysis developed, the terminology in the paragraph and the caption of Table 5 are updated in the RM.

Since the study focuses on droughts, the model performance simulating streamflow in the dry season was analysed separately. Performance indicators were calculated for the period corresponding to the basin's dry season (December to March). The intermediate period of precipitation decrease from June to July was also included in this analysis. Table 5 summarises the calibration and validation performance indicators in the dry season. According to the rating guidelines, the model performance simulating streamflow in the dry season is satisfactory (ASABE, 2017). RM Ln 323 to Ln 327.

Table 1. *SWAT model performance simulating flows in the dry season.*

12. P15L324: Figure 4. It is annual average right? Also, readers need an explanation about soil type A, B, C, and D. What are those?

Figure 4a to h presents the multi-annual average of the numerical hydroclimatic drivers of droughts at each subbasin. In the RM, we indicate that Figure 4 presents the multi-annual average.

Regarding soil types in the Table 2 (last row), we introduce the hydrologic soil groups and we clarify that they refer to soil's infiltration characteristics.

Figure 4 presents the numerical and categorical hydroclimatic parameters used as potential drivers of droughts. Figure 4a to h presents the multi-annual average of the numerical hydroclimatic drivers of droughts at each subbasin. RM Ln 332 to Ln 333.

Table 2. *Explanatory variables used in MVRT*

Hydroclimatic parameter	*Abbreviation*	*Unit*	*Definition*
Precipitation	*PRECP*	*mm*	*Average precipitation at each subbasin*
Potential evapotranspiration	*PET*	*mm*	*Average potential evapotranspiration at each subbasin*
Evapotranspiration	*ET*	*mm*	*Average actual evapotranspiration at each subbasin*
Percolation	*PERC*	*mm*	*Average percolation past the root zone*
Surface runoff	*SURFQ*	*mm*	*Average surface contribution to the streamflow at each subbasin*
Groundwater	*GRWQ*	*mm*	*Average groundwater contribution to the streamflow at each subbasin*
Water yield	*WYLD*	*mm*	*Average amount of water that leaves the subbasin and contributes to the streamflow at each subbasin*
Sediment yield	*SYLD*	*metric tons/ha*	*Average sediment from the subbasin transported into the reach*
Curve number	*CN*	–	*Dominant curve number at each subbasin*
Slope	*SLP*	–	*Dominant slope at each subbasin*
Hydrologic soil group	*STY*	–	*Dominant hydrologic soil group (A, B, C, and D) at each subbasin. The soil hydrologic groups refer to the soil's infiltration characteristics. Properties of each soil type can be found in USDA (2007)*

13. P15L328: What drought category is represented in Table 6? Is it severe drought, extreme, or moderate drought?

Table 6 does not present information on drought severity. Severity changes every month in each subbasin, and the table cannot show such variability. Figures 4 and 5 are used to present the severity variability. These figures present the total number of months observed for each drought category. Using that information, we build the dataset of explanatory variables to apply the MVRT technique.

Starting from here, please read carefully and do comprehensive check.

14. P17L357: Here the authors say: potential evapotranspiration (1,679 mm). However, I cannot see this number in Figure 7.

The authors regret for this mistake, potential evapotranspiration was not used at the fifth level of split. The sentence was *corrected* in the RM.

> *Then, the left branch was recursively split as follows: at the third level, according to potential evapotranspiration (1,888 mm) and evapotranspiration (1,191 mm); at the fourth level, according to evapotranspiration (1,064 mm) and percolation (111 mm); and at the fifth level, according to sediment yield (101 tons/ha).* RM Ln 367 to Ln 370.

15. P18L369: I think it is lower and not above.

The authors regret for this mistake, as the reviewer indicates the correct word is below not above. The paragraph is updated in the RM.

For subbasins with actual evapotranspiration below the threshold, leaf c, the median of months in the severe drought category is below ten (Figure 8c).. RM Ln 386 to Ln 387.

16. P18L372-373: It is not Figure 8b and also please check your statement about "highest median of months in the severe drought category"

The authors regret for this mistake, as the reviewer indicates referring to Figure 8b is incorrect. The statement is also corrected in the RM.

For subbasins with actual evapotranspiration below the threshold, leaf c, the median of months in the severe drought category is below ten (Figure 8c). For subbasins with actual evapotranspiration above the threshold, leaf d, the median of months in the severe drought category is sixteen, one of the highest among the terminal groups (Figure 8d). RM Ln 386 to Ln 387.

P18L375,376: It is not above but lower.

The authors thank the reviewer for this comment since we realized that there is room to improve the description of the leaves *e*, *f* and *g*. The paragraph is updated in the RM.

Leaves e, f and g cluster twenty-four, six and twelve subbasins, respectively. Subbasins are located in the river valley and the basin's western part. In these subbasins, precipitation was below the basin average (1318 mm), and actual evapotranspiration was above the average (1191 mm). The percolation threshold to split leaves e and f from leaf g is 111 mm, a value considerably below the basin average. At the fifth level of split, the sediment yield threshold to split leaves e and f is 101 metric tons/ha, a value close to the average value in the basin. Figures 8e, f and g show that subbasins clustered in these leaves are prone to agricultural droughts. The median of months in the moderate drought category was above twenty months; the severe category was above ten months; and the three leaves exhibited months in the extreme drought category. RM Ln 382 to Ln 388.

17. P18377: Here the authors stated that moderate drought category was above 20 months and severe category was above 10 motnhs -> but this is not for Figure 8h.

The authors regret for this mistake, in this paragraph we are describing leaves *e*, *f* and *g*. Referring to Figure 8h was a mistake. The paragraph is updated in the RM (Please see answer to specific comment 17).

18. P20L404: Check the number 1362 mm, cannot see this number.

The authors regret for this typo error. The correct number is 1632 mm. The error is corrected in the RM.

> *The subbasins were separated at the first split level according to precipitation (1632 mm).* RM Ln 426.

19. P20L421: I think it is higher and not lower.

The authors regret for this typo error. As the reviewer indicates, the correct word is higher. The paragraph is updated in the RM.

> *In leaves e (n = 72) and f (n = 18), precipitation and water yield exceeded the basin average. The actual evapotranspiration threshold to split leaves e and f is 833 mm, value below the basin average. Both terminal groups describe moderate exposure to hydrological drought. At leaf e, the median of months in the severe and extreme drought categories is below five, while the median of months in the moderate drought category is twenty (Figure 10e). The hydrological drought exposure of the subbasins clustered at leaf f is also mild. In these subbasins, actual evapotranspiration is above the threshold and close to the basin average. These subbasins present the lowest median of months for all drought categories (Figure 10f). Notably, the Zapatosa marsh and upstream subbasins are clustered in this terminal group (Figure 9f).*RM Ln XX to Ln XX.

20. P20L426: Check if it is Figure 9f?

The authors double-check and the Zapatosa marsh is clustered at leaf *f*; then referring to Figure 9f is correct. Please see answer to specific comment 19).

21. P20L428: Check if it is 37?

The authors regret for this typo error. As the reviewer indicates, the correct number is forty. The paragraph is updated in the RM.

> *Leaves g and h cluster seventy-one and forty subbasins, respectively*.RM Ln 456.

22. P21L442: Why figure 8e? 8d is higher

Reviewer's comment allowed us to notice that group *d* also experience a considerable number of months in severe drought category, accordingly the leaf *d* was included in the statement. Leaf *e* is also included in the sentence since subbasins clustered at that group experience a higher number of months in extreme drought category compared with leaf *d*.

23. P21L444: Clusters h, i, and j are seen in Figure 8i, j, and k?

The authors regret for this typo error. As the reviewer noticed, there is an error in the figure we are referring to. The error is corrected in the RM.

> *The subbasins in leaves h, i and j predominately experienced months in the moderate drought category (Figure 8h, i, and j).*RM Ln 471 to Ln 472.

24. P22L475: It think it is not groups I and j but h and i.

The authors regret for this typo error. As the reviewer indicates, the correct leaf is "*h*".

> *At terminal groups h and i, water yield was found to influence the severity of agricultural drought.* RM Ln 508 to Ln 509.

25. P23L479: This sentence is confusing. Subbasins grouped at leaf i showed in Figure 10b and c?

We double-checked the statement, and it is correct. The sentence indicates that subbasins in leaf i show low susceptibility to agricultural droughts but high exposure to hydrological droughts. We mention Figures 10b and c because they present the subbasins experiencing hydrological drought susceptibility.

26. P25L544: were located

The authors apologize for this mistake, the paragraph is updated in the RM.

> *Overall, these subbasins were located in the southern part of the basin*. RM Ln 578 to Ln 579.

27. P26L591: software -> model

The paragraph is updated in the RM.

> *The study's limitations include its simplified approach to modelling a complex phenomenon using SWAT model (e.g. representing the groundwater components that impact hydrological drought conditions) and using only a single ML technique to build explainable models.* RM Ln 631 to Ln 633.