Response to Reviewer 1 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 the time given to this manuscript and for providing insightful and detailed comments to help us to improve this manuscript's overall scientific quality and readability. Your attention to detail has undoubtedly enhanced the overall strength of our study. Notably, we appreciate the comments about the missing definition of drought severity and the lack of information on the application of MVRT to this particular study. We will apply multiple changes to incorporate the reviewer's suggestions and clearly define the study's objective. 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 will apply in the Revised Manuscript (RM) are in italics.

General Comments.

1. My main concern regards the lack of clarity in the objective of the study. The title fails to mention that it is an application to a specific case study. In addition, it is not clear until deep in the "results section" what exactly the authors mean with drought severity. For almost the entire paper the readers are left wondering what exactly is modelled with the MVRT. Is it the severity of a series of events on the entire basin? Is it the spatial distribution of the severity? This should be made clear already in the objective described in the introduction, and then detailed in the methodology.

The authors thank the reviewer for pointing out that the title does not mention that the study is an application to a case study. Accordingly, we will update the RM title, including the case of study:

> *Multivariate regression trees as an 'explainable machine learning' approach to exploring relationships between hydroclimatic characteristics and agricultural and hydrological drought severity. Case of study Cesar River basin.*

Regarding the second part of the comment, we agree that the introduction needs to include the definition of drought severity and how it is represented in this study. In addition, it fails to describe what is modelled by applying the MVRT technique. Since both concepts are crucial elements of this study, we will apply two changes in the RM. In the introduction, we will include a paragraph presenting the definition of drought's severity and how it is represented using drought indices. In addition, we will update the introduction indicating that drought severity categories (moderate, severe and extreme) are the three response variables modeled with the MVRT. We present the updated version of the introduction (There are no changes in the first two paragraphs of the introduction).

Projections indicate that drought frequency, severity and duration are expected to increase globally in the twenty-first century (UNDRR, 2021). 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 United States Department of Agriculture (Lu et al., 2019). A similar trend is predicted for hydrological drought severity. This 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). The intensification of drought characteristics (in combination with other factors) could force the migration of up to 216 million people by 2050 (The World Bank, 2021), increase wildfire risk and tree mortality, and negatively affect regional air quality, among other ecosystem impacts (Vicente-Serrano et al., 2020).

It is essential that we better understand drought drivers if we are to foster preparedness and resilience to projected drought events. Remarkable progress has been achieved in understanding drought propagation through the hydrological cycle (Van Loon et al., 2012). Drought occurs due to climatic extremes, which may be enhanced or alleviated by region characteristics and anthropogenic influence (Hao et al., 2022; Seneviratne et al., 2012; Tijdeman et al., 2018). Typically, droughts are triggered by atmospheric circulation and weather systems that combine to cause lower precipitation and/or higher than normal evaporation in a region (Destouni & Verrot, 2014; Sheffield & Wood, 2011a). Reduced precipitation leads to a decrease in soil moisture, causing agricultural drought. When soil moisture depletion is high, it is restored in the wet season, thus reducing subsurface flow and groundwater recharge and giving rise to hydrological drought (Iglesias et al., 2018). Regional characteristics such *as soil type, stratigraphy, elevation, slope, vegetation cover, drainage networks, water bodies and groundwater systems play a relevant role in response to the climate anomalies that affect drought propagation and contribute to different levels of agricultural and hydrological drought (Sheffield & Wood, 2011a; Zhang et al., 2022). Equally important, human interventions in the hydrological cycle (e.g. reservoirs, water diversion, deforestation, over-pumping groundwater, overgrazing, urbanisation) can reduce water supplies, triggering a drought situation or exacerbating a climate-driven drought (Rangecroft et al., 2019; Wang et al., 2021).*

Drought planning also uses research progress on drought characterisation. Various methodologies for drought characterisation exist, using drought indices is widespread (Zargar et al., 2011). Drought indices are computed numerical representations of drought severity (Hao & Singh, 2015; Keyantash & Dracup, 2002). Severity refers to the departure from the normal of an index. 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).

Despite remarkable progress achieved in understanding the drought-generating process and drought characterisation, there is still a need for studies that assess the complex interplay between the different drivers of droughts and how their combined effect influences drought characteristics (e.g. duration, severity, intensity) (Valiya Veettil & Mishra, 2020). Previous studies focus on the influence of one driver (Margariti et al., 2019; Mastrotheodoros et al., 2020; Shah et al., 2021; Xu et al., 2019), and some of the methodologies applied cannot adequately address the non-linear relationship between climate, basin processes and droughts (Peña-Gallardo et al., 2019; Saft et al., 2016; Van Loon, 2015).

We have found two studies that employ machine learning to analyse the non-linear relationship between climate and basin processes and droughts (Konapala & Mishra, 2020; Valiya Veettil & Mishra, 2020). Valiya Veettil et al.

(2020) used a classification and regression tree (CART) to identify the variables influencing drought duration. Since CART allows one output variable (drought duration), the authors applied the technique 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 for each drought regime and characteristic. Both studies focused on drivers of hydrological drought and were developed in the continental United States.

Mentioned research shows the potential of machine learning techniques for drought-related analysis; nevertheless, there is still a need for testing a technique capable of simultaneously assessing the influence of different drought drivers on the individual categories of drought severity. Commonly used in the field of ecology to relate independent environmental conditions to populations of multiple species, Multivariate Regression Tree (MVRT) arises as a suitable technique for this purpose. MVRT is a supervised clustering technique that links explanatory variables to multiple response variables while maintaining the individual characteristics of the responses. Significantly, the technique does not assume a linear relationship between explanatory and response variables. Furthermore, it allows for the so-called "interpretable machine learning" algorithms that make decisions and predictions understandable to humans (Molnar, 2022). MVRT interpretably is a relevant attribute for drought researchers and planners since the method allows them to identify the parameters influencing severe (or mild) drought conditions.

To understand the relationship between the drivers of droughts and the individual categories of agricultural and hydrological droughts severity, this *study employs a methodology that consists of three steps. The first is hydrological modelling. We used Soil Water Assessment Tool (SWAT) to simulate the hydroclimatic parameters required for analysing droughts and applying the MVRT approach. The Second is the analysis of droughts. SWAT outputs, soil moisture and streamflow, are used to calculate the drought indices Soil Moisture Deficit Index (SMDI) and the Standardized Stream Flow Index (SSI). Drought indices are utilised to identify the agricultural and hydrological drought events during the period of analysis and describe their severity. Finally, the MVRT approach is applied to assess the relationship between hydroclimatic characteristics (represented by the simulated parameters at each subbasin, see Table 2) and droughts severity categories (represented by the observed number of months for each drought severity category at each subbasin, see Table 3). The analyses for agricultural and hydrological droughts were conducted separately; thus, two MVRTs were obtained. A concrete application of this methodology is developed in the Cesar River basin (Colombia, South America).*

2. Another related issue of the paper is the lack of specific details of the application of MVRT to the given study case. Most of the description is rather generic, and do not answer key questions about the specific application. The authors state that one of the advantages of MVRT is the capability to output multiple variables, but it is never clarified why this is needed here and how this is exploited.

The authors agree with the reviewer that the introduction and methodology do not explicitly present the reasons for choosing the MVRT approach and how the technique capabilities are exploited in this study. Accordingly, two changes will be included in the RM. First, we will update the introduction presenting the MVRT capabilities relevant to the study.

Despite remarkable progress achieved in understanding the drought-generating process and drought characterisation, there is still a need for studies that assess the complex interplay between the different drivers of droughts and how their combined effect influences drought characteristics (e.g. duration, severity, intensity) (Valiya Veettil & Mishra, 2020). Previous studies focus on the influence of one driver (Margariti et al., 2019; Mastrotheodoros et al., 2020; Shah et al., 2021; Xu et al., 2019), and some of the methodologies applied cannot *adequately address the non-linear relationship between climate, basin processes and droughts (Peña-Gallardo et al., 2019; Saft et al., 2016; van Loon, 2015).*

We have found two studies that employ machine learning to analyse the non-linear relationship between climate and basin processes and droughts. Valiya Veettil et al. (2020) used a classification and regression tree (CART) to identify the variables influencing drought duration. Since CART allows one output variable (drought duration), the authors applied the technique 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 for each drought regime and characteristic. Both studies focused on drivers of hydrological drought and were developed in the continental United States.

Mentioned research shows the potential of machine learning techniques for drought-related analysis; nevertheless, there is still a need for testing a technique capable of simultaneously assessing the influence of drought drivers on the individual categories of drought severity. Commonly used in the field of ecology to relate independent environmental conditions to populations of multiple species, Multivariate Regression Tree (MVRT) arises as a suitable technique for this purpose. MVRT is a supervised clustering technique that links explanatory variables to multiple response variables while maintaining the individual characteristics of the responses. Significantly, the technique does not assume a linear relationship between explanatory and response variables. Furthermore, it allows for the so-called "interpretable machine learning" algorithms that make decisions and predictions understandable to humans (Molnar, 2022). MVRT interpretably is a relevant attribute for drought researchers and planners since *the method allows them to identify the parameters influencing severe (or mild) drought conditions.*

Second, in the methodology, we will update the introductory paragraph of Section ***Multivariate regression tree approach for evaluating the relationships between hydroclimatic characteristics and droughts severity*** and include a paragraph describing the reasons for selecting the technique.

MVRT is an extension of a regression tree (Breiman, 2001), but it differs in that it allows for multiple outputs (see De'ath, 2002). It allows the recursive split of a quantitative response variable (predictand, output) controlled by a set of numerical or categorical explanatory variables (predictors, input). The technique approach yields a set of non-linear models, each a piece-wise linear regression model (of zero order). An MVRT result is a tree whose terminal groups (leaves) of instances (input-output vectors) comprise subsets of instances selected to minimise the within-group sums of squares. Each successive split is given by a threshold value of the explanatory variables (Borcard et al., 2018). MVRT applies to dataset exploration, description and prediction (De'ath, 2002). In this study, explanatory variables are the hydroclimatic parameters at each subbasin, represented by the average value of each parameter during the analysis period (1987 to 2018). The number of months observed at each drought severity category (Categories are given by the drought indices) are the response variables. The analyses for agricultural and hydrological droughts were conducted separately; thus, two MVRTs were obtained.

Four technique attributes are relevant to this study. First, MVRT can capture the non-linear interactions between the parameters influencing droughts and their severity. Second, the technique can handle numerical and categorical hydroclimatic parameters influencing drought severity (explanatory variables). Third, 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). The drought indicators give these three categories to represent the drought severity. Simultaneous analysis of different drought categories provides a comprehensive understanding *of the drought-generating process and the factors influencing severe (or mild) drought conditions. Fourth, MVRT results can be easily visualised and interpreted. The resulting tree structure provides a clear representation of the relationship between the drivers of droughts and the severity of agricultural and hydrological droughts.*

3. A lot more can be said on the "explainable" portion of the study. The authors provide some comments on the outcomes of the two MVRTs, but the link between these outputs and a physical interpretation is lacking. In both the discussion and the conclusion sections (as well as in the abstract), the authors stress how a main finding is the division of the domain in 3 macro regions. However, it is not clear how this conclusion is drawn from the outputs of MVRT, and how MVRT are "explained" to derive this conclusion. At the moment, it is seems that this conclusion is derived from previous knowledge of the area rather than the actual outcomes of the study.

Regarding the reviewer's concern about dividing the basin into three regions, the authors realized that the analysis results should be summarised differently. It is more precise to say that we identify different sets of parameters that govern drought severity in the basin. First, severe agricultural and hydrological drought conditions are driven by precipitation shortfalls and high potential evapotranspiration. This interaction is observed in the upper part of the river valley. Second, severe agricultural and hydrological drought conditions are caused by inadequate rainfall partitioning and an unbalanced water cycle favouring water loss through percolation and evapotranspiration. According to the results, the middle part of the river valley is affected by the interplay of these parameters. Finally, moderate exposure to agricultural and hydrological droughts is related to the capacity of the subbasins to retain water, which lowers evapotranspiration losses and promotes percolation. Moderate drought severity is observed in the Zapatosa marsh and the Serrania del Perijá foothills.

To improve the description of our results and ensure readers' clarity, we will not include the reference to the three regions in the RM. Following Reviwer's General Comment 4, we will compare the results from the two MVRT trees (See answer to General Comment 4). We agree that this is a better way to describe differences and similarities between the parameters influencing the severity of agricultural and hydrological droughts and present the spatial distribution of the areas experiencing severe and mild drought conditions. In the RM, the abstract and the conclusion will be updated accordingly.

5. Conclusion (Second paragraph)

Our results 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 occurred in the upper and middle course of the river. Nevertheless, the severe droughts were influenced by different hydroclimatic factors. The interaction between precipitation shortfalls and high potential evapotranspiration drove severe agricultural drought in the headwater. Conversely, severe hydrological drought condition was solely caused by limited precipitation. In subbasins in the middle course, droughts' severity was 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. Subbasins in the basin's southern part experienced moderate agricultural and hydrological drought severity. Mild agricultural drought was linked to low evapotranspiration losses and basin capacity to retain water in the soil profile, improving percolation. In turn, moderate hydrological drought severity relates to the subbasins' proximity to the marsh (which acted as a natural control reducing the water yield) and surface runoff contributions to the streamflow. The outcomes of this study demonstrate that the combined effect of parameters with low impact can trigger a drought situation as severe as the one produced by one or two of the most influential hydroclimatic parameters.

4. In addition, the outcomes of the two MVRTs are rather different, and it would be interesting to discuss the analogies and differences between the two (in spatial patterns, explanatory variables, etc.). In the current version, the two analyses are almost independent from each other. Is the division in 3 macro regions valid for both agricultural and hydrological droughts? Is yes, how it is so given the differences in the trees?

To improve the description of our results and ensure readers' clarity, we will not include the reference to the three regions in the RM. In the RM, we will have a section highlighting similarities and differences between the MVRTs.

4.3 Comparison of the hydroclimatic parameters influencing the severity of agricultural and hydrological droughts

Crucial similarities and differences emerge from contrasting the parameters influencing the severity of droughts and the spatial distribution of the subbasins experiencing severe and mild drought conditions. MVRTs indicate that severe agricultural and hydrological drought conditions occurred in the upper and middle course of the river. Nevertheless, the severe droughts were influenced by different hydroclimatic factors. Severe agricultural drought in the headwater was driven by the interaction between precipitation shortfalls and high potential evapotranspiration (Figure 7a). Conversely, severe hydrological drought condition was solely driven by limited precipitation. It is worth highlighting that the severe hydrological situation extends from the headwater to some subbasins in the middle course (Figure 9a).

Downstream, in subbasins located in the middle course, the agricultural and hydrological drought situation was also severe. In this area, droughts' severity was linked to inadequate rainfall partitioning and an unbalanced water cycle that favours water loss through evapotranspiration and low percolation values (Figure 7d, e, f and g, and Figure 9b, c and d). Significantly, agricultural and hydrological droughts in these leaves were more severe than in leaves experiencing precipitation deficits (Figure 7a and Figure 9a). Results suggest that poor soil structure enhances severe agricultural drought conditions (Figure 7e), and high curve numbers seem to increase hydrological drought severity (Figure 9b).

MVRTs also showed subbasins experiencing mild agricultural and hydrological drought severity. Overall, these subbasins were in the southern part of the basin. However, for agricultural drought, a few cases were observed in the north of the basin (Figure 7h, i and j). Subbasins presenting mild hydrological drought severity allocate upstream of the Zapatosa marsh (Figure 9g). Moderate agricultural drought severity was linked to low evapotranspiration losses and basin capacity to retain water in the soil profile, improving percolation (Figure

7j). In turn, moderate hydrological drought severity relates to the subbasins' proximity to the marsh (which acted as a natural control reducing the water yield) and surface runoff contributions to the streamflow (Figure 9g). Remarkably, some of these subbasins also showed mild agricultural drought conditions (Figure 7i).

5. Finally, given the focus on drought, I would have expected a validation of the model also in term of drought quantities, especially low-flow conditions. The validation of the SWAT model should be expanded to highlight reasonable performances during drought conditions, and possibly expanded to soil moisture as well.

The authors agree with the referee that given the focus of the study on droughts, it is appropriate to evaluate the model performance simulating low-flows. In the RM manuscript, we will include the model performance indicators for the dry season.

Considering the study focus is on droughts, the model performance simulating low flows was analysed separately. Performance indicators were calculated for the dry season, which lasts from 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 low flows is satisfactory (Transactions of the ASABE (American Society of Agricultural and Biological Engineers), 2018).

Table 1. *SWAT model performance simulating low flows.*

Gauging station	*Calibration*		*Validation*	
	NSE	*PBIAS [%]*	*NSE*	*PBIAS [%]*
Puente Salguero	*0.65*	*-19.4*	*0.53*	*-21.3*
Puente Carretera	*0.67*	*-15.3*	*0.53*	*17.2*
Cantaclaro	*0.67*	*-3.6*	*0.58*	*16.3*
Puente Canoas	*0.55*	*-15.7*	*0.60*	*-13.5*

Regarding the comment about expanding the validation to soil moisture, the authors agree with the reviewer that calibration and validation of the model using soil moisture may contribute to reducing the uncertainty for the drought analysis; nevertheless, monthly soil moisture data is needed for calibrating and validating the model, either in-situ measurements, satellite-derived soil moisture, or reanalysis soil moisture, at subbasin level. There are no in-situ soil moisture measurements in the study area, and the spatial resolution of the available datasets of satellite-derived soil moisture or reanalysis soil moisture is coarse (0.25°×0.25°). Accordingly, data availability constraints that analysis. In the absence of data to conduct that calibration, good performance simulating streamflow indicates that the model adequately reproduces the land phase of the water cycle in the basin.

Specific Comments

1. L12-13. You mention anthropogenic interventions and region's characteristics, but those are factors that are barely included in your analysis. If this is a key point of your study, it should be better reflected in the analysis.
2. L51. "MAY play…" Actually, I have the impression from your results that some of these quantities do not play a major role, at least in your study region.
3. L53-55. Again, you stress the role of human interventions but only marginally included them in the study.

The authors highlight that "the region's characteristics" refer to hydroclimatic parameters recognised as potential drought drivers. We consider that the region's characteristics are adequately reflected in the analysis. The manuscript's introduction presents different hydroclimatic parameters that influence the drought-generating process and the characteristics of droughts. These parameters include soil type, stratigraphy, elevation, slope, vegetation cover, drainage networks, water bodies and groundwater systems. In the methodology section, Table 2 presents the hydroclimatic parameters used in this study as potential drivers of droughts (percolation, surface runoff, groundwater, water yield, sediment yield, curve number, slope and soil type). Comparing the parameters presented in the introduction with the parameters in the methodology confirms that both are in good agreement. Furthermore, the results and discussion section show that most of the parameters included in the analysis influence the drought's severity.

The authors agree with the reviewer that hydroclimatic parameters selected at the first split levels have more influence on droughts than those at lower levels. However, a relevant outcome of this study is that the combined effect of parameters with low impact can trigger a drought situation as severe as the one produced by one or two of the most influential parameters.

Regarding comments 1 and 3, the authors agree that the representation of anthropogenic interventions is limited. Land use change (represented by the CN2) is the only anthropogenic intervention included in the analysis. At the initial stage of the study, the authors asked local and regional authorities about the available information on irrigation systems and groundwater pumping in the area. The authorities confirm that the activities are developed in the region, but there was no consolidated information on these systems' location and operation characteristics. Accordingly, it was not possible to represent these interventions in the study and evaluate the impact on drought severity.

Although the influence of anthropogenic activities is not widely analysed in this study (due to the lack of data), it is relevent to mention them in the introduction. In that section, we aim to provide an overall picture of all the potential drivers of droughts and various studies have demonstrated that human activities can enhance a drought situation.

4. L76. This is the right place to highlight why a multivariate approach may be needed here.

Indeed, the authors agree with the reviewer that the introduction needs to indicate why a multivariate approach is relevant to this study. As shown in answer to General Comment 2, we will improve the introduction to explicitly present why we opted for this technique and how its capabilities are used and relevant for this work.

5. L87. Please better link this line and figure to the rest of the text reported later (description of the methodology).

To improve the structure of the section and better link Figure 1 to the description of the methodology, we will apply the following changes in the RM. The section title and subtitles will be updated: Section 2 is *Study location and methods*, and the subsections are: *2.1 Case of study and 2.2 Methods. Section 2.2 includes 2.2.1 Hydrological modelling, 2.2.3 Agricultural and hydrological drought analysis and 2.2.3 Multivariate regression tree approach for evaluating the relationships between hydroclimatic characteristics and droughts severity.* Figures 1 and 2 are swapped according to the new section's order.

2 Study location and methods

2.1 Case study

Figure 1 presents the Cesar River basin's location, topography and land use. The basin is located between 72º53'W 74º04'W and 10º52'00'N 7º41'00''N latitude (Colombia). It extends for an area of 22,312 km2. The basin's topography is defined in three distinct climatic regions (Universidad del Atlantico, 2014). In the north is La Sierra Nevada de Santa Marta. This sector is characterised by steeply sloped mountains reaching up to 5,700 meters above sea level (masl). The temperature ranges from 3°C to 6°C, and the mean annual precipitation is 1,000 mm. In the east is La Serranía del Perijá. This mountainous area is an extension of the eastern branch of the Andes range. In this sector, the altitude ranges from 1,000 to 2,000 masl. The average temperature is 24°C, and *the average annual precipitation varies from 1,000 mm to 2,000 mm. Lastly, the valley of the Cesar River and the Zapatosa marsh are in the west and south of the basin, respectively. The valley is characterised by flat topography and a complex system of marshes formed by the Cesar River floodplains and its confluence with the Magdalena River. The average temperature is 28°C, and the mean annual precipitation is 1,500 mm. At the basin, the annual rainfall pattern presents a dry season from December to April, followed by a rainy season from April to May. In the intermediate period from June to July, precipitation decreases. The main rainfall events occur between August and November.*

The predominant land use is pasture, followed by agriculture (Universidad del Atlantico, 2014). The primary land use in La Sierra Nevada foothills is pastures for cattle farming. In La Serranía del Perijá, the high altitude areas are covered by forests in very good condition; at the lower altitudes, the principal land use is agriculture, particularly subsistence crops. The Cesar River valley's soils are rich in nutrients, providing favourable conditions for agriculture. The riverbanks are covered by forest with low tree density.

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. Nevertheless, the region is threatened by the overexploitation of its forest resources and overfishing. In addition, climate change projections indicate that the basin's temperature may increase by 2.7°C, and precipitation may reduce by ten percent by 2070 (Universidad del Magdalena et al., 2017). Accordingly, multiple initiatives are oriented to improve water management and create resilience to hydroclimatic extremes (Ministerio de Ambiente y Desarrollo Sostenible (Colombia), 2015).

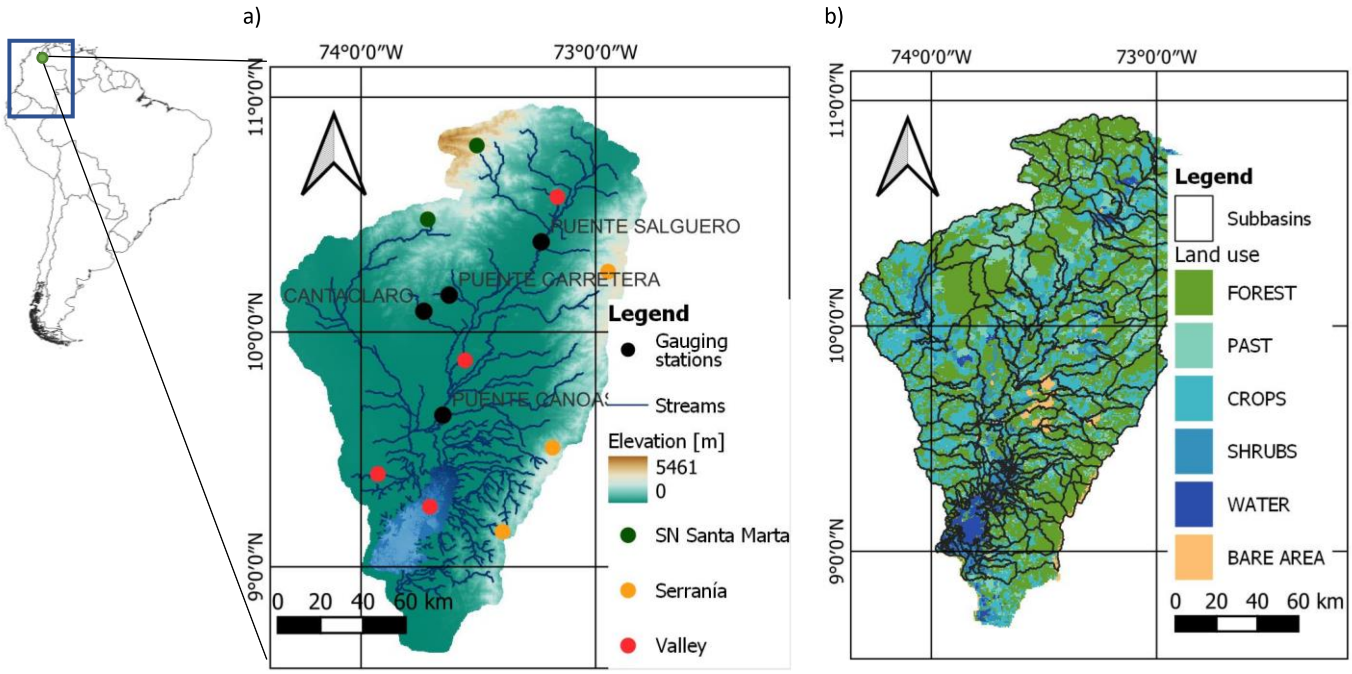

Figure 1 *Cesar River basin: a) topography and b) land use.*

2.2 Methods

Figure 2 illustrates the three steps methodology applied in this study. Section 2.2.1 describes the hydrological modelling, and 2.2.2 the drought analysis. Section 2.2.3 presents the application of the MVRT technique.

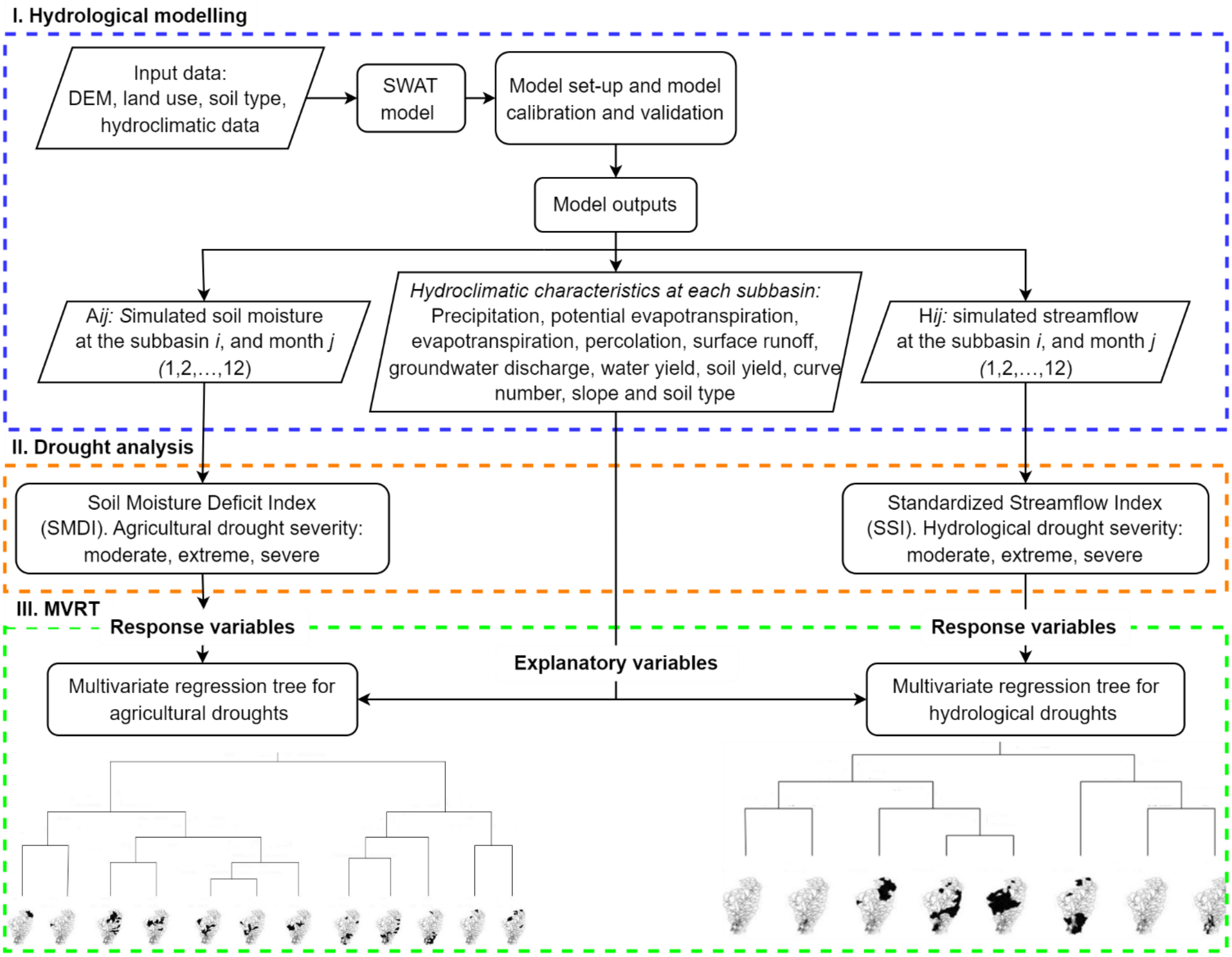

Figure 2 Flow chart of the methodology

6. L91. Please mark these three sub-regions in the map for the people not familiar with the region.

See answer to Specific Comment 5, Figure 1a

7. L105. You mention pasture here, but no "pasture" class is reported in the Figure. Please align the text with the figure.

There was an error in the figure. The category "GRASS" is actually "PASTURE". The authors apologize for the mistake. The figure is corrected in the RM. See answer to Specific Comment 5, Figure 1b.

8. L115. Reference?

The reference will be included in the RM.

> *Accordingly, multiple initiatives are oriented to improve water management and create resilience to hydroclimatic extremes (Ministerio de Ambiente y Desarrollo Sostenible (Colombia), 2015).*

9. L121. I would link this sentence to the next.

The sentence will be link to next in the RM.

A SWAT model with an ArcSWAT extension was used to develop the Cesar River basin model used in this research. SWAT is a continuous-time, semi-distributed, process-based river watershed-scale model developed by The Agricultural Research Service of the United States Department of Agriculture (ARS-USDA). The model is designed to simulate the quality and quantity of surface and groundwater and predict the environmental impacts of land use, land management and climate change (Neitsch et al., 2011).

10. L142. I assume that CN2 is the initial CN for soil moisture condition 2, since the actual CN is a variable. Please clarify.

The authors apologize for the mistake. Indeed, CN2 is the initial SCS runoff curve number for moisture condition II. The CN2 definition will be corrected in the RM.

CN2 (initial SCS runoff curve number for moisture condition II).

11. L143. No calibration on the Manning factor?

The manning factor was used in the calibration of the model. It was not included in the *Section Model Calibration and Validation* by mistake. The parameter will be included in the RM.

Based on expert judgment and the available literature (Arnold et al., 2012; Transactions of the ASABE (American Society of Agricultural and Biological Engineers), 2018), the following SWAT parameters were used in the calibration and validation process: baseflow alpha factor (ALPHA_BF), effective hydraulic conductivity in main channel alluvium (CH_K), Manning's value for the main channel (CH_N2), SCS runoff curve number for moisture condition II (CN2), soil evaporation compensation factor (ESCO), groundwater delay (GW_DELAY), threshold depth of water in the shallow aquifer required for return flow to occur (GWQMN), deep aquifer percolation fraction (RCHRG_DP), threshold depth of water in the shallow aquifer for percolation to the deep aquifer to occur (REVAPMN) and available water capacity of the soil layer (SOL_AWC).

12. L152. Since your focus is on hydrological drought, I suggest adding some evaluation metrics focused specifically on low flow. It is a well-known issue that NSE may return high values even when low flow conditions are not well represented due to a good matching of flood values. Also, given the relevance of soil

moisture in your study, some kind of validation/evaluation of the performances in terms of soil moisture is needed.

See the answer to General Comment 5.

13. L162. No details are provided on the soil profile. Is it a single soil layer? How depth? Please clarify.

More details on the soil profile will be provided in the RM.

> *According to the soil profiles and the secondary information used to elaborate the soil map, three soil layers were identified in the Cesar River basin. The soil layers' thickness (vertical distance from the surface) varies. The first layer reaches up to 350 mm, the second 1000 mm, and the third 1500 mm.*

14. L190. What is the reference period? 1987-2018? Clarify.

The reference period will be included in the RM

> *To this aim, the monthly simulated streamflow at each subbasin in the analysis period (1987 to 2018) was fitted to the gamma probability distribution function.*

15. L194. This sentence is not clear to me. Does the 30% refer to the total area of the basin, meaning that a minimum number of sub-basins (covering at least 30% of the total area) need to be in moderate drought?

Indeed, the reviewer's description of the sentence is correct. The sentence will be updated in the RM to prevent the reader's confusion,

> *SMDI and SSI were calculated monthly using the simulated soil water and streamflow values at each subbasin. The drought events during the period of analysis were then identified. A drought (agricultural or hydrological) event was assumed to occur in the basin when a number of subbasins (covering at least 30 % of the basin's total area) were in a moderate drought state for at least two consecutive time steps (i.e. in this study month). According to the spatial and temporal thresholds, a drought event began when both conditions were met and continued until one of them failed to be met. We set a minimum spatial extension threshold because droughts typically extend regionally (Sheffield & Wood, 2011b). By setting the temporal threshold, we avoided identifying periods of water shortage or scarcity as drought events.*

16. L196. You mention short periods, but I do not see any constrains on the duration of an event. Please better clarify the definition of drought event used here (i.e. starts when at least 30%...., and end when…). Also, if any kind of spatial or temporal pooling is performed please clarify.

The authors agree with the reviewer that the paragraph fails to adequately describe the temporal threshold used to identify droughts (agricultural and hydrological). As indicated in the answer to Comment 15, this paragraph will be improved in the RM.

17. L198. The PCA has a very limited role in this study. I suggest reevaluating the need to include this section and this analysis in the study.

The authors thank the reviewer for questioning the relevance of the PCA results. Before applying the MVRT, we used PCA to explore the dataset of explanatory variables. Our goal was to identify the most influential parameters of the dataset and discard non-influential parameters. The PCA results showed that all the parameters considerably influenced at least one of the PC retained; thus, for the MVRT technique, we used all the parameters initially selected. Reviewer's Specific Comments 17 and 31 make us reevaluate the relevance of the PCA results since the method did not produce changes in the set of explanatory variables. We concluded that using PCA was a good strategy for explanatory variables exploration, but the outcome of the analysis is not relevant to the objective of this study. Accordingly, we will remove the PCA analysis in the RM.

18. L216-221. This a rather generic description of the methodology. Please contextualize the method to your study. This section should answer the questions: What is a predictand (see comment below)? Why are they multiple? Why do you need MVRT instead of simple RT?

In the RM, we will update the paragraph to indicate the response variables explicitly and include a new paragraph to contextualize the technique in the study.

> *MVRT is an extension of a regression tree (Breiman, 2001), but it differs in that it allows for multiple outputs (see De'ath, 2002). It recursively splits a quantitative response variable (predictand, output) controlled by a set of numerical or categorical explanatory variables (predictors, input). The technique approach yields a set of non-linear models, each a piece-wise linear regression model (of zero order). An MVRT result is a tree whose terminal groups (leaves) of instances (input-output vectors) comprise subsets of samples selected to minimise the within-group sums of squares. Each successive split is given by a threshold value of the explanatory variables (Borcard et al., 2018).*

MVRT is applied to dataset exploration, description and prediction (De'ath, 2002). In this study, the explanatory variables are the hydroclimatic parameters at each subbasin, represented by the average value of each parameter during the analysis period (1987 to 2018). The response variables are the number of months observed at each drought severity category (the drought indices give categories). The analyses for agricultural and hydrological droughts were conducted separately; thus, two MVRTs were obtained.

Four technique attributes are relevant to this study. First, MVRT can capture the non-linear interactions between the parameters influencing droughts and their severity. Second, the technique can handle numerical and categorical hydroclimatic parameters influencing drought severity (explanatory variables). Third, 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). The drought indicators give these three categories to represent the drought severity. Simultaneous analysis of different drought categories provides a comprehensive understanding of the drought-generating process and the factors influencing severe (or mild) drought conditions. Fourth, MVRT results can be easily visualised and interpreted. The resulting tree structure provides a clear representation of the relationship between the drivers of droughts and the severity of agricultural and hydrological droughts.

19. L223. The response variables need to be better identified here. The generic "drought severity" used here leaves a lot of questions to the readers. Is it a time series of event severity for each sub-basin? A time series over the entire basin? Just a single value (average or similar)? This need to be clarified here (and eventually detailed later) in order to justify the multivariate dimension of the problem.

In the RM, we will update the paragraph to indicate the response variables explicitly (Se answer to Specific Comment 18). In addition, we will improve the description of the set of response variables.

Set of response variables

We used the drought analysis outcomes to define the response variables (Table 3). Following the methodology presented in 2.3, we identified the *agricultural and hydrological drought events during the analysed period. After identifying the drought events, we counted the months for each drought severity category at each subbasin. The observed months for each one of the three drought categories were used as response variables. The analyses for agricultural and hydrological droughts were conducted separately; thus, two sets of response variables were obtained.*

20. L223-229. Related to the previous point. Here you first give the impression that agrological and hydrological drought severities are the two “multivariate” variables. Then, you clarify that the two are studied separately, leaving the question on what is the “multivariate” variable then. This can be only indirectly inferred from the results section, but it must be clearly stated already here.

We will update the introduction and methodology in the RM to define the response variables clearly. See the answer to General Comment 1 and Specific Comments 18 and 19.

21. Since section 2.5 is supposed to be the main methodology section, you need to significantly extend this section and add all the needed clarifications. Also link to the flow chart should me reported here.

The following we summarize the changes we will apply to the Section 2.5 (Section 2.2.3 in the RM).

- We will define the sets of explanatory and response variables. See answer to Specific Comment 18.
- We will include a new paragraph to properly contextualize the MVRT technique in the study and highlight the attributes relevant for this study. See answer to Specific Comment 18.
- We will improve the description of the set of explanatory variables. See answer to Specific Comment 26.
- We will improve the description of the set of response variables. See answer to Specific Comment 19.

22. L234. Again, similarly to the previous section, it is not clear what average means here. Is it a spatial average? A temporal average? Do you use time series of spatial-average values for each sub-basin or just a single value. This can be indirectly inferred from the results, but it should be made clear here.

In this study, the explanatory variables are the hydroclimatic parameters at each subbasin, represented by the average value of each parameter during the analysis period (1987 to 2018). The introduction and the methodology will be updated in the RM to improve the description of the explanatory variables. See answer to General Comment 1, and Specific Comments 18 and 26.

23. L240. Following the previous comment: so, do you have 3 values for each sub-basin as response variables? Are then the frequency in the 3 categories the “multivariate”?

Indeed, the drought severity categories were the multivariate response. We agree with the reviewer that the manuscript needed more clarity about the application of the MVRT technique and why the drought severity was considered a multivariate output. To improve the description of the methodology, we will apply the changes presented in the answers to General Comment 1 and Specific Comments 18 and 19.

24. L251.Which two groups?

The sentence will be rewritten in the RM.

> *The data partitioning consisted of three steps. First, for each explanatory variable were generated all possible partitions of the sites (subbasins) into two groups.*

25. L276. This sentence seems to imply that two methods are used to choose the size, which is in contrast with the next sentence. Please clarify.

The authors apologize for the mistake. In the RM, we will update the paragraph to indicate the approach we used to choose the tree size. In the RM, we will not include information on the method we did not use.

> *To choose the tree size that retained the most descriptive partition, we used the approach suggested by De'ath (2002). According to the author, a tree with the smallest CVRE offers the best explanatory power and interpretability combination. Once the tree was built, the proportion of explained variance (EV) was calculated as 1-〖RE〗_tree (tree relative error) (Cannon, 2012).*

26. L294-295. This should be clarified in the methodology and not here.

In the RM, we will update the methodology description to indicate the explanatory variables explicitly; see the answer to Specific Comment 18. In addition, we will improve the description of the set of explanatory variables.

> ***Set of explanatory variables***
>
> *To select the set of explanatory variables, we used the outcomes of previous studies on governing drivers of droughts (Sheffield & Wood, 2011a; Zhang et al., 2022). Table 2 describes the eleven parameters selected as the potential drivers of droughts. The used values correspond to the parameters' average in the analysis period (1987 to 2018). The averages were computed using the SWAT model outputs at each subbasin. We used the dominant category at each subbasin for the curve number, the slope, and the soil type (categorical variables).*

27. I am not 100% sure that the data reported in sections 3.1 and 3.2 are results of the study. They may fit better in the “Data and method section”, since they do not bring much to the discussion on the use of MVRT.

We thank the reviewer for the suggestion but prefer to maintain Sections 3.1 and 3.2 in the results. We consider that model calibration results and simulated hydroclimatic parameters are results of this study and fit best the in that section.

28. Section 3.3. It is not clear how these 6 events are derived from the methodology described in section 2.3. There, only a minimum fraction of the area in the sub-basin is defined, and nothing is said on duration/continuity of an event. Is there any constrain on duration? Did you remove the minor events? Please clarify.

Indeed, the minor events were not included in this analysis. The authors agree with the reviewer that the methodology fails to provide details on how the drought events identified during the simulation period were derived from the methodology. In the RM, we will adequately describe the temporal threshold used to identify droughts (agricultural and hydrological). See answer to Specific Comment 15.

29. Table 5. There is a typo on event 4 (IV).

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

Table 2. *Agricultural and hydrological droughts during the period of analysis*

Event	***Agricultural droughts***		***Hydrological droughts***	
	Date	***Duration [months]***	***Date***	***Duration [months]***
I	*May 1991 – Jun 1992*	*13*	*Apr 1991 – May 1992*	*14*
II	*Jun 1997 – April 1998*	*11*	*Apr 1997 – Feb 1998*	*11*
III	*Jun 2001 – Aug 2001*	*3*	*May 2001 – Jun 2001*	*2*
IV	*Oct 2009 – Jan 2010*	*4*	*Sep 2009 – Nov 2009*	*3*
V	*Jun 2014 – Aug 2014*	*3*	*Jun 2014 – Jul 2014*	*2*
VI	*May 2015 – Jul 2016*	*15*	*Apr 2015 – Apr 2016*	*13*

30. L310. This should be made clear much sooner in the text, and clearly highlight that the multivariate of the MVRT is referring to the 3 categories.

We improved the description of the response variables in the introduction and the methodology. See answers to General Comment 1 and Specific Comments 18 and 19.

31. 3.4 As a said before, this has very marginal impacts on the analysis. At the end, you included all the variables in the MVRT analysis, but some of them where not actually used in the final trees (and some very marginally). What does this say on the usefulness of the PCA in this case? I suggest removing this part and

focus more on analyzing the variables used in the two final MVRTs and the differences between the two trees.

We will not include the PCA analysis in the RM. See answer to Specific Comment 17.

32. L334-342. Was an analysis on a limited number of explanatory variables also performed? As an example: how different are the results if only ET and PREC are used? Are some leaves really necessary? As an example, h) and i) are separated only at the end and based on WYLD, but the plots in Fig. 9 are quite similar. Are all 12 leaves relevant, considering that you then discuss only 3 macro regions? Some leaves are also quite small (just 2 basins for b) and k) for instance); if these are relevant, then they shouldn't be grouped in the 3 macro regions in the discussion and conclusion sections.

33. The same considerations are true for the results on hydrological drought.

Answer to Specific Comments 32 and 33

To build the MVRT, All the explanatory variables are used to recursively generate the partitions resulting in the three's final leaves. We did not perform the analysis using fewer explanatory variables because it may result in MVRTs with lower explanatory power. Including multiple explanatory allows the technique to produce the partitions that maximize the explanatory power of the three (maximize the proportion of the explained variance). In addition, before applying the MVRT technique, we used PCA to explore the dataset of explanatory variables (As explained in answer to the Specific Comment 17). Our goal was to identify the most influential parameters of the dataset and discard non-influential parameters. The PCA results showed that all the parameters considerably influenced at least one PC retained. It indicates that all the parameters included in the set of explanatory variables are relevant to this study. Accordingly, for the MVRT technique, we used all the parameters initially selected. It is worth mentioning that we chose the threes with the lowest CVRE. According to De'ath (2002), these trees offer the best explanatory power and interpretability combination.

Regarding the importance of all the leaves retained, we consider that all leaves provide relevant information on the different hydroclimatic parameters influencing droughts' severity. Figures 7, 8, 9 and 10 show that the severity of doughs (agricultural and hydrological) is different in each leaf and influenced by different parameters.

Regarding the three regions mentioned in the abstract and the conclusion, the authors realized that the statement does not properly summarize the study results. It is more precise to say that we identify different sets of parameters that govern drought severity in the basin (See answer to General Comment 3). The RM will not include the paragraphs referring to these three regions.

34. L424-426. This should be better supported by some synthetic results, rather than leaving the extraction of meaningful information to the readers.

The reviewer refers to the first paragraph of Section 4.1. In that paragraph, we summarize the information presented in Section 3.5.1 and link the tree description (results) with the discussion. In addition, in the following paragraphs of Section 4.1, we provide a detailed discussion about the parameters influencing the droughts and the severity in each leaf. We consider that the reviewer's comment may arise from the expression "subbasins most exposed to agricultural droughts". We will update the sentence in the RM to ensure readers' clarity.

> *The left branch of the MVRT clusters the subbasins exposed to severe agricultural drought (Figure 8a, e, f, g). Conversely, the right branch of the MVRT clusters the subbasins experiencing moderate agricultural drought severity. The subbasins in leaves h, i and j predominately experienced months in the moderate drought category (Figure 8i, j, k).*

35. L514-521. This explanation is a little lacking, since the explanatory variables and the targets are both derived from the same modelling framework. I am wondering if some variables that are relevant for the hydrological drought were not included in the analysis.

The authors agree with the reviewer that the three's explanatory power may also be linked to relevant parameters for the hydrological drought not included in the analysis. In the last part of this paragraph, we refer to this limitation.

> *Conversely, the explanatory power of the tree built for hydrological drought is not very high (EV = 0.48). This may be related to the inaccurate representation of groundwater contribution to the streamflow. Streams depend significantly on groundwater during droughts to maintain flow; nevertheless, groundwater contribution to the streamflow was not included as a key drought driver in the MVRT, although it was in the list of explanatory variables. It is possible that the model's simplifications for the simulation of groundwater flow and storage did not adequately represent the groundwater contribution to the streamflow (Molina-Navarro et al., 2019). The lack of adequate information about this relevant factor hydrological drought may have compromised the MVRT's accuracy. Unexplained variability may also link to factors that influence hydrological drought but were not considered in the dataset of explanatory*

variables (e.g. abstractions such as water for irrigation, industry or human consumption).

In addition, in the limitations of the study we mentioned that parameters influencing droughts were not included in this analysis.

> *Additionally, there is still a need to better represent anthropogenic interventions (and other relevant parameters influencing droughts) in the set of explanatory variables (e.g. abstractions such as water for irrigation, industry or human consumption, groundwater pumping).*

36. L523-529. Even if 9/11 were included, some have a very limited role and appears only in hydrological drought. This discussion needs to be expanded, and a more in-depth comparisons of the two trees need to be added.

Regarding the first part of the comment, the authors considered that the relevance of parameters is not given by the number of times it was selected at different split levels in one or both threes. We evaluated a parameter's relevance by contrasting the drought's severity in the different leaves. For instance, in the MVRT for hydrological droughts, precipitation and water yield are alike for leaves g and h. Surface runoff is selected at the third split level, dividing the subbasins into two groups. Figure 10 shows that in the leave g, the median of months in the moderate drought category is ten, while at h is eighteen. Furthermore, each leave shows different number of months in severe and extreme drought categories. Although surface runoff was used at the third split level (and not included in the MVRT for agricultural droughts), results show that the parameter is utilized to divide subbasins presenting different agricultural drought severity. Similar analysis can be developed for sediment yield (Figure 8 leaves e and f) and curve number (Figure 8 leaves k and l, and Figure 10 leaves b and c).

About the second part of the comment, the comparison of the two trees was included in the RM. See answer to General Comment 4.

37. L542. Is this true also for hydrological drought?

In the line indicated by the reviewer both types of droughts are mentioned.

> *This study applied the MVRT technique, which served as an explanatory approach (in the line of 'explanatory AI') to assess the relationship between a subbasin's hydroclimatic characteristics (i.e. explanatory variables) and the severity categories of agricultural and hydrological drought (i.e. response*

variables). The results show that the machine learning technique successfully identified drought severity's primary drivers and critical thresholds. The MVRT also provided valuable information on which parameters can contribute to reducing agricultural and hydrological drought severity.

38. L546-547. This subdivision in three sub-areas is never highlighted in the results, and it is not evident how and why these three sub-areas are the same for agricultural and hydrological droughts, given that different trees and explanatory variables are identified.

Agreed. See answer to General Comments 3 and 4.