Reply to the Comments of Reviewer 1: Second Revision

Rebuttal of the manuscript entitled "On understanding mountainous carbonate basins of the Mediterranean using parsimonious modeling solutions"

We have sincerely appreciate the valuable comments provided by both the Associate Editor and the reviewer. We have carefully considered each of reviewer's comments and, hopefully, have taken the necessary steps to address them. In our revised version, we have provided detailed point-by-point responses and made corresponding modifications to the manuscript accordingly. **Reviewers' comments** are in boldface.

The colour of the text in the revised manuscript is as follows:

1. blue is used to specify the changes related to the **former** reviewer 1's comments.
2. cyan is used to specify the changes related to the **former** reviewer 2's comments.
3. red is showing the changes which refer to the **former**-common comments of both reviewers.
4. brown is used to specify the changes related to the **current** reviewer #1's comments (second round of revision).

Reviewer's Comments:

The authors have made efforts to clarify the novelty of the discussion paper. While I still find that the use of only 3 stations with limited periods of overlapping record is difficult to make general conclusions about mountainous carbonate basins, my area of expertise is more focused on statistical methods in hydrology and I have to evaluate the discussion paper on this aspect. I leave the other aspect - the novelty of the modeling solution for carbonate basins - to hopefully be assessed by other reviewers.

Authors' Reply:

We highly appreciate the reviewer's effort in critically evaluating our approach and providing a valuable feedback. We acknowledge that one of the limitations of our study is the limited number of stations with overlapping records. However, we have to consider that the scarcity of data is a common challenge even in today's data-rich era. Despite this limitation, we believe that by employing a physically based (albeit lumped) modeling approach and robust correlation analysis, the data shortage challenge could be still addressed. In fact, our findings align with the results of other studies which could further support the effectiveness of our proposed approach in mitigating the data limitation problem. In the main text, we have explicitly mentioned this as one of the limitations of our study and highlighted the potential benefits of utilizing a larger and more diverse dataset to gain a deeper understanding of catchment behavior.

Once again, we appreciate the feedback from the reviewer, which has allowed us to address this limitation in our manuscript. We add the following sentences to the main text in lines 409-412:
"Overall, having more data with a longer period of overlapping records would be probably beneficial to improve the simulation of such a complex basin behavior. Although one of the limitations of our study is the limited number of stations with overlapping records, employing a physically based (albeit lumped) modeling approach together with a robust correlation analysis could mitigate the data shortage issue."

Reviewer's Comment 1:

There are still questions about the methodology that need to be clarified further before the discussion paper can be considered for final publication.
L132: Do you actually prove that the mean precipitation is constant in the red-shaded area of Figure 1? I may have missed this, but this assumption appears to be quite critical for the analysis and it is unclear whether this is actually shown in this work.

Authors' Reply:

First, we have to mention that the precipitation data has been spatially interpolated using the kriging method by taking into account the gauges located outside of the basin, in addition to the gauges inside the basin. Figure 1 (in the rebuttal) shows the location of the gauges applied for interpolating the precipitation in the basin and over carbonate areas. This could guarantee a better representation of precipitation over the red-shaded areas which has been applied as the input of the model for these areas. Second, as clarified in the latest version of the manuscript and in our previous response, one of the primary objectives of our study is to develop an efficient modeling approach for simulating the behavior of the carbonate system, particularly in situations where data availability is limited. In Figure 1 (in the manuscript), we have depicted two separate carbonate areas in red (upstream of CSA and Ussita) represented by two lumped models. It is important to note that incorporating more spatial variability for carbonate areas will result in increased model parameters, which could introduce additional uncertainties. Given the limited existing data for calibrating the model parameters, we decided to simulate the groundwater contribution from the carbonate area using only two additional lumped systems. This could satisfy our primary objective which is to investigate the temporal variation of basin storage and its response to the drought at the three main closure points, rather than capturing the overall spatial variability. Also, it should be considered that the random and mixed nature of flow within the carbonate system tends to diminish the influence of individual flow paths, allowing for a more simplified representation. The simulation scores together with the uncertainty analysis show the adequacy of this choice in describing the system. We appreciate the reviewer's attention to this aspect and assure them that our modeling approach, despite its simplifications, effectively captures the temporal variation of basin storage and provides meaningful insights into the behavior of the carbonate system under the given data limitations.

We add more details to clarify the issues mentioned in this comment (Line 184-192 in the text):

"In this study, the primary focus is on the temporal variation of the precipitation-recharge-discharge behavior of the AC water flowing from CSA and Ussita rather than the spatial variability of the carbonate system's behavior. This allows us to specifically investigate the impact of single or multiple-year drought

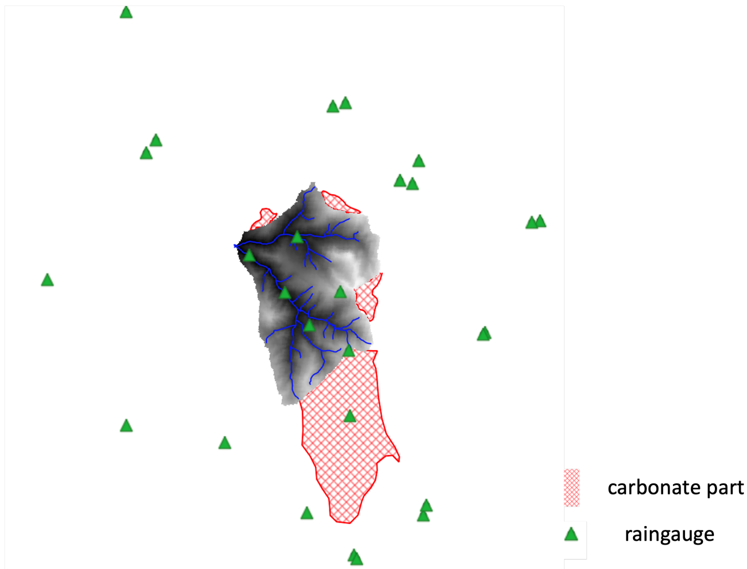

Figure 1. We have taken into account the raingauges outside of the basin and close to the carbonate areas to better representation of precipitation over these areas. (The figure is associated with this rebuttal and it is not added to the main text)

events on the basin storage, as discussed in Section 5.2. So, a lumped modeling approach is more appropriate for providing more insights into the temporal system's response to drought conditions and its implications for basin storage. Furthermore, incorporating more spatial variability for the carbonate areas would result in an increased number of model parameters. This introduces additional uncertainty into the model. Given the limited availability of data for calibrating these parameters, using two separate lumped systems has been considered as an efficient strategy for the modeling. The results will also demonstrate that the temporal behavior of the AC water and its response to drought events could be investigated properly by this modeling approach."

Reviewer's Comment 2:

Because of the limitations of using only 3 stations, it is very difficult to generalize any conclusions from this work. This is especially true when 1 of the 3 stations (the middle station on Figure 3) shows a different relationship between cumulated precipitation and cumulated runoff. It does not appear to be discussed how this affects your assumptions or conclusions about groundwater influxes to the carbonate system. For example in L140, it is stated that "After understanding that the external contribution to the basin is significantly greater than that provided by the terrain analysis..." However, this is not true for this 1 station where cumulated precipitation is slightly less than cumulated runoff in Figure 3. How is this explained? And why is this not mentioned?

Authors' Reply:

Please refer to the answers to the previous comments about using three hydrometric stations. We apologize for any confusion caused and not explicitly addressing the relationship between cumulated precipitation and cumulated runoff in different pannels of Figure 3.
The area of CSA, Ussita, and Visso increases in order, with CSA being the smallest and Visso being the largest which encompasses both CSA and Ussita. Figure 3 illustrates that the contribution from the carbonate (red) catchment decreases with increasing catchment size, which is a reasonable expectation. After Lines 131-138 in the text which is about the variation of runoff coefficient, Lines 138-141 are added as follows:
“The area of CSA, Ussita, and Visso increases in order, with CSA being the smallest and Visso being the largest which encompasses both CSA and Ussita. Fig. 3 illustrates that the contribution from the carbonate (red) catchment decreases with increasing catchment size, which is a reasonable expectation. To further clarification about the lower runoff coefficient of Visso, the readers could refer to Section 5.2 and Fig. 12."

However, it is important to consider this information in conjunction with Figure 12, which presents the water budget components. The water budget analysis in Figure 12(b) allows us to infer that there is a hidden subsurface flow in the catchment fed by the river, which causes sequential positive storage in the basin. This could indicate that the river basin feeds the groundwater system specifically between the CSA and Visso hydrometric stations. In Lines 363-369 in the main text, we have mentioned this feature of the basin as follows:

"For Visso (Fig. 12b) the storage differences remain positive for the period of interest, indicating a potentially infinite stored water accumulation over multiple years. Possible explanations are either that the basin feeds the groundwater system (not simulated by the model) between the CSA and Visso hydrometric stations (see also the observation discharge at CSA and Visso stations) or, considering the long-term memory of the basin, that ten years (2010-2021) is a relatively short period for observing storage changes. While both assumptions remain unanswered, the complexity of the system makes the first assumption plausible, however further investigations are needed to provide compelling evidence of this. Moreover, the first assumption could also justify the lower runoff coefficient at Visso station. "

To avoid any confusion, we also modified Figure 3 and sorted the panels of the figure (top to bottom) according to the order of subbasin area.

Reviewer’s Comment 3:

It is also confusing as a reader where the figures show 3 stations as CSA (which is not spelled out and figure captions should be stand-alone), Ussita, and Visso; however, the text continues to discuss a station MU and does not discuss Ussita. Are these the same station? Or was Ussita dropped for the analysis?

Authors’ Reply:

We apologize for the confusion caused by the inconsistency in the figure captions and the text about different station names.

To clarify, CSA, Madonna dell'Uccelletto (MU), and Visso are the three stations analyzed in our study. In fact, Ussita is the river where the MU station is located on that. So Ussita was not removed from the analysis; rather, it was sometimes mistakenly applied instead of MU in the text. The authors appreciate the reviewer for drawing our attention to this issue. It has been modified in the entire text and MU is applied referring to the station and Ussita is used to call the river basin in the current version of the manuscript. All the station name abbreviations have been already spelt out in Table 1. We also provided more clarification in Lines 101-104:

"The stations used in the study are Visso and Castelsantangelo (CSA) on the Nera River, and Madonna dell'Uccelletto (MU) on the Ussita River. To prevent any confusion, from this point forward in the text, "MU" will be used to refer to the station, while "Ussita" will be used to denote the river basin."

Reviewer's Comment 4:

The new statistical approach that is introduced is difficult to understand.

Authors' Reply:

The proposed method for evaluating the reliability of simulations involves calculating the empirical probability of any measured discharge conditional on the simulated value at the same time step.

To clarify the methodology, we have modified lines 227-239 of the text as follows:

"The process of computing the Empirical Conditional Probability (ECP) involves the following steps:

- Combining the observed discharge and the corresponding simulated values into a single dataset.
- Grouping the dataset into n classes (bins) according to the simulated discharge values. The quantile-based discretization method has been applied for binning data into different classes (see Figure 2 in this rebuttal).
- Computing the Empirical Cumulative Distribution Function (ECDF) for each class j using the formula:

$$ECDF_j(Q) = \frac{1}{m_j} \sum_{i=1}^{m_j} I_{X_i<Q} \qquad (1)$$

Here, $ECDF$ represents the empirical cumulative distribution function of the j-th class, m_j is the number of measures in the group, X_i denotes the i-th measure in the group, and

$$I_{X_i<Q} = \begin{cases} 1 & \text{if } X_i < Q \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

– Computing the features of empirical distribution function (i.e., mode, maximum, minimum, and mean of the discharge) for each class.

From the ECDF, we can derive the empirical distribution functions for different classes (e.g., the one shown in Fig. 6(b) and 6c), which are assigned to each time step. "

and also Lines 248-252 are added as follows:

"The histograms obtained for different bins (e.g., Fig. 6 (b) and (c)) are dedicated to the time steps visualised as green dots and grey-shaded areas in Fig. 7, 8, and 9. The green dots and grey area illustrated in the figures provide an indicator of the reliability of the simulations, according to previous simulated and observed data. In particular, the disparity between the measured and the mode values (green dots) can be considered as a measure of this reliability. The complete estimation procedure is thoroughly documented in a specific Notebook, accessible in the supplementary material. "

Given that you only have 3 stations and very limited data for calibration and validation - in fact, as noted, data were so limited for calibration and validation that it was not always possible to use all stations for these purposes (L186-190) - how can the new statistical method have been robustly validated on this dataset? More justification is needed as to how this small set of data was able to provide reliable empirical PDFs.

Since the available data are hourly, the computed empirical probability distribution function for each class -even for the shortest time series (MU station) with 26281 data points- is based on a meaningful number of data points. For this aim, the number of classes is specified in such a way that a meaningful histogram is obtained for each class. However, we mentioned this point in Lines 252-255 of the text.
"It is important to highlight that the number of classes (bins) has been carefully chosen to ensure a meaningful histogram for each discharge class. Even for the shortest available dataset (at the MU station, which encompasses approximately 26,000 hourly data points), a reasonable number of samples are available for each discharge class"

Minor comments:

There are numerous areas where the referencing has two parentheses. A complete review of the discussion paper needs to be completed to address this aspect. See L108 and L118 for 3 examples of this.

It has been corrected.

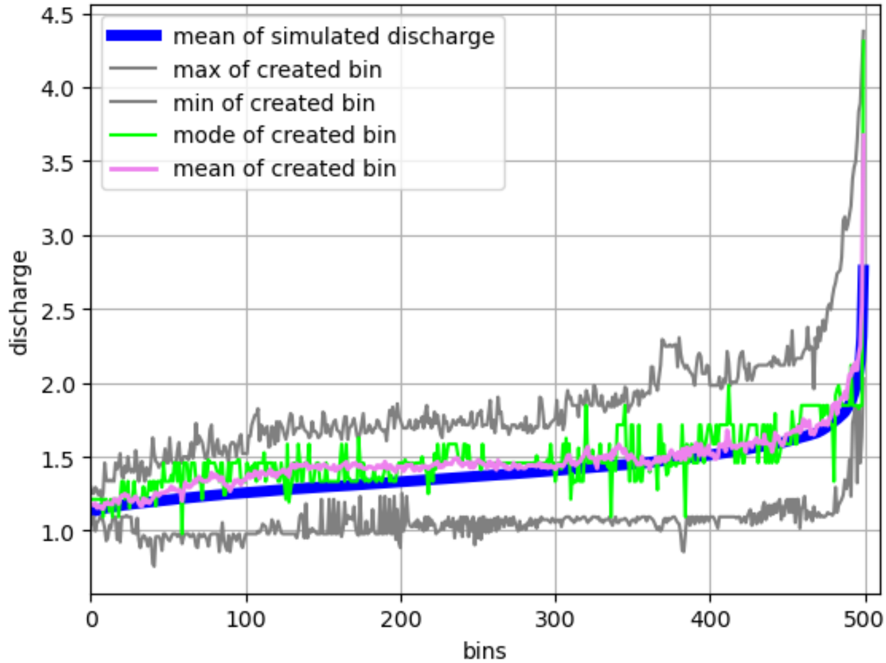

Figure 2. The features of empirical distribution function (i.e., mode, maximum, minimum, and mean of the discharge) for each class (The figure is associated with this rebuttal and it is not added to the main text)

L94-97: The sentence in L94 states the data were "provided by the Marche Region." This is not clear as a Region cannot provide the data. Later in L97, it is stated that the Marche Regional Authority provided some data. Is that what was meant in L94? Also, a reference is needed for the Marche Regional Authority data in L96.

We have written "Marche Region Authority" instead of "Marche Region"

L4: Should read: "climate change conditions"
It has been corrected.

L5-6: Should read "describe behavior in carbonate basins"
It has been modified.

L132: Should read: "the null"
It has been corrected.

Reply to the Comments of Reviewer 2: Second Revision

Rebuttal of the manuscript entitled "On understanding mountainous carbonate basins of the Mediterranean using parsimonious modeling solutions"

Reviewer's Comments:

The revision provided by the Authors accomplished with all comment provided after the first submission phase. I consider the manuscript acceptable for publication in HESS

We sincerely thank the reviewer for dedicating time and effort to thoroughly review our manuscript. We greatly appreciate their valuable comments, which have enhanced the quality of our work. We are grateful for their positive evaluation and consideration of our manuscript for publication in HESS.