Response to Editor

The authors thank the Editor for the suggestions and sincerely appreciate the Reviewers' insightful and constructive comments, which helped to clarify several important aspects and significantly improve the overall quality of the manuscript.

Below, we provide our responses to the general comments and detailed point-by-point remarks.

We believe that some of the remaining points of discussion reflect differing perspectives, particularly between an experimental and a modeling-oriented approach. We have attempted to clarify these aspects in the revised manuscript to improve transparency and facilitate dialogue across these complementary viewpoints.

General comments: *Thank you for submitting your manuscript to HESS. The reviewers acknowledge the value of the dataset and the ambition of developing an integrated, field-based characterization of groundwater–surface water (GW–SW) interactions in a complex, data-scarce mountain catchment. However, both reviews raise substantive concerns regarding (i) whether the monitoring design and the spatial and temporal resolution of the data are sufficient to support the strength of the conclusions, and (ii) whether key methodological components, particularly baseflow separation, storage change estimation from recession analysis, and the water-budget framework, are presented and interpreted with adequate rigor and explicit consideration of uncertainty. I therefore invite you to prepare and submit a major revision.*

Response: We thank the Editor and the two Reviewers for their careful evaluation of the manuscript and for highlighting key issues related to monitoring design, methodological rigor, and uncertainty. We have substantially revised the manuscript by adding new text, revising figures, updating analyses to clarify the rationale, strengths, and limitations of the monitoring strategy, as well as the interpretation of the results.

All Reviewer comments have been addressed through targeted revisions, which are detailed below.

Comment 1. *Monitoring design, representativeness, and limits of inference.*

- *Strengthen and clarify the justification for discrete discharge campaigns and thermal imagery timing, and explicitly discuss what the sampling frequency/resolution can and cannot resolve.*
- *Expand the description of spring identification, sampling, and gauging procedures, and how these data support the GW–SW interaction interpretation.*

Response: We expanded the manuscript to clarify the conceptual basis of the monitoring design, the rationale for discrete discharge campaigns, and the interpretive limits associated with the adopted spatial and temporal resolution.

Specifically, as reported in our response to Reviewer #1's general comments, we added new text at ***lines 155-166*** to strengthen and expand the description of the monitoring and measurement concepts and to explain why discrete discharge monitoring in S1, S3, and S4 secions, coupled with continuous discharge monitoring in two stream sections (S2 and S5 sections), was performed on those specific dates. Moreover, we presented the results of a new drone flight conducted in summer 2025 (***lines 422-428*** and Fig. 5b), which explains why the winter thermal imagery allowed us to achieve the best possible result, considering the multiple factors involved in a complex mountainous region. In detail, although groundwater inflow points and stream stretches were identified by a single drone survey in January 2024, the analysis is considered reliable because it was conducted during a no-recharge period (e.g., Fig. 4a) and under hydrometeorological conditions favorable for highlighting GW–SW water temperature contrasts. On the contrary, the summer vegetation did not provide further information about the GW inflow (see discussion at ***lines 543-547***). Thanks to the winter drone survey, we identified some punctual springs (I1 and I2 in Figure 1b) between sections S3 and S5, whose water was sampled along with stream water during the geochemical measurement campaigns. The discharge of these springs was monitored using spot measurements with the OTT MF Pro flow meter, as we mainly focused on the total increase in discharge between sections S3-S4 and S4-S5. These increases in discharge are exclusively linked to GW's inflows into the stream. In these conditions, stream discharge data, coupled with thermal drone imagery, enabled robust conclusions about the locations of stretches with groundwater inflow, thereby defining geochemical sampling points. In the discussion section, we added new text to emphasize that, in mountain regions, the best practice is to obtain basic, insightful information to understand the hydrogeological system (***lines 531-540***). Overall, we believe the data analysis is well-suited to addressing research question (a) and yields robust conclusions that can be easily incorporated into water management systems.

Comment 2. *Uncertainty and robustness of derived conclusions*

- *Provide a more explicit, quantitative treatment of uncertainty where feasible (measurement uncertainty, rating-curve uncertainty, uncertainty across climate products, and propagation to recharge/water-balance terms). If a full propagation analysis is beyond scope, the authors should provide a transparent "minimum viable" uncertainty evaluation (e.g., bounding/ensemble estimates and sensitivity).*
- *Ensure that conclusions are phrased in a way that matches the resolution and uncertainty of the underlying data (avoid over-claiming robustness).*

Response: We agree that an explicit treatment of uncertainty is essential for interpreting the robustness of the derived conclusions. However, the primary objective of this manuscript is to provide an integrated, preliminary quantitative framework for investigating GW–SW interactions in a complex, data-scarce mountain catchment, rather than a comprehensive uncertainty propagation analysis, which would constitute a substantial methodological study on its own.

Within this scope, we implemented a transparent, ensemble-based "minimum viable" uncertainty assessment, as suggested by the Reviewers. We added metrics to evaluate rating-curve uncertainty (see ***lines 155-158***), we reported, when available, the uncertainty across climate products from the literature (see Supplement material), and included an ensemble of water budget estimates, resulting in a more robust assessment of the variance in recharge-area extent, including the impact of snowmelt. We added two additional ET datasets, GLEAM and ECOSTRESS (Table 1). The latter has a high spatial resolution (70 m), increasing the number of water budget evaluations from 19 (initial version) to 35 (revised version). Although more sophisticated methods could further improve uncertainty quantification, maintaining this level of simplicity is appropriate given the scope and length of the manuscript. In our opinion, the water budget ensemble we provided is a transparent "minimum viable" uncertainty assessment of the recharge area extent for the Ussita stream. According to both Reviewers, the manuscript has been strengthened by adding a new section (4.4) that discusses perspectives on the use of uncertainty quantification and propagation for the water budget components, and provides possible directions for implementing such analyses in future studies.

Conclusions are rephrased to avoid overclaiming robustness (see ***lines 669-671***).

Comment 3. *Methods requiring clearer validation/defense*

- *Baseflow separation and recession analysis: The reviewers question the robustness of simplified/top-down approaches in complex systems. If the authors retain these methods, they must (a) more clearly document validation steps (e.g., performance during no-recharge periods), (b) expand discussion of known limitations and conditions for applicability, and (c) clarify what the method is used for (diagnostic interpretation vs. precise flux quantification).*
- *Water-budget closure: Make the treatment of leakage/inflow/outflow explicit, including how residual terms are interpreted and what alternative explanations remain plausible.*

Response: We acknowledge the Reviewers' concerns regarding the use of simplified, top-down approaches for baseflow separation and water-budget closure in a complex mountain catchment. In response to Comment 7 from Reviewer #2, we explained that the approach we used to estimate the BF (the LH digital filter method) is one of the possible techniques. We clarified that BF estimates are not treated as exact fluxes, but as diagnostic indicators of groundwater contribution to streamflow (***lines 206-208***). According to Xie et al. (2020), the digital filter methods perform well in many catchments, including those with high infiltration rates, such as the Ussita catchment. Moreover, a recent paper by Mo et al. (2025) examined a karst catchment and found that the LH method performed well. However, we also believe that evaluating the accuracy of baseflow estimation methods remains an open question in the literature, because baseflow cannot be directly measured except during periods of no recharging or prolonged dry spells (i.e., in these periods, stream discharge corresponds to BF). Thanks to the Master Recession Curves, we found that during no-recharge periods, stream discharge is described by the Maillet equation (i.e., linear reservoir depletion). Using the recession constant, we computed the k-filter and verified that the derived BF were appropriate for the available recession periods. In other words, the computed BF data were validated against continuous streamflow data during no-recharge periods (S2 and S5 sections), once again highlighting the importance of acquiring data in complex mountain environments to constrain BF estimates as effectively as possible.

Regarding the water budget closure, the issue of leakage/inflow/outflow (i.e., groundwater exchange) is explained in the response to Comment 7 from Reviewer #2. In complex mountain hydrogeological systems, water budget closure is strictly dependent on groundwater outflow to neighboring hydrogeological systems and by groundwater inflow into the catchment (it is illustrated graphically in the schematic representation of Fig. 4). In this way, we computed in sections 3.3.1 and 3.3.2 the difference of these two GW components as residuals, which has helped us understand that very few external GW inflows come into the catchment. We discussed this point in the new section 4.4, introducing a new paper by Bouaziz et al. (2018), who examining the water budget in 58 small limestone fractured/karst catchments along the Meuse River at the border between France and Belgium, concluded that, due to the nature and complexity of the catchments, net groundwater exchanges are the primary cause of water inbalances, which are most significant in small catchments, consistent with the earlier findings of Schaller and Fan (2009). Anyway, in the Supplement we reported, when available, the uncertainty across climate products from the literature, which can be useful for considering how uncertainty propagates in future analyses (***lines 629-632***).

Comment 4. *Scope and framing*

- *Reviewers asked for model-based interpretation and broader generalization across catchments. Given the focus of the study on integrated field characterization, the authors are not required to add a full numerical modeling study or multiple catchments, but they should sharpen the framing.*

The paper should clearly present a transferable approach while avoiding language that could be interpreted as claiming broadly transferable hydrological behavior.

- *Add a short, balanced discussion of how modeling could build on your dataset (conceptualization needs, what is constrained vs. not constrained by your observations).*

Response: We agree that model-based interpretation can provide additional insight into the processes investigated in this study. However, the primary objective of the manuscript is to develop an integrated, field-based characterization of groundwater–surface water interactions, and including a full numerical modeling experiment would substantially extend the manuscript beyond its intended scope. The results highlight the challenges of model development in such a complex mountain catchment, including strong spatial heterogeneity, scale mismatches among datasets, and contrasting hydrological conditions. In the revised version, we added a new subsection (4.5) to the discussion that emphasizes how evidence from our integrated multidisciplinary approach can support future modelling.

As reported in the response to Reviewers' comments, we further clarified that the manuscript presents a transferable approach, adding new text to prevent any misunderstanding and explicitly stating that our study does not aim to generalize catchment behavior (***lines 85-87***; ***lines 99-100***; ***lines 669-671***).

Comment 5. *Structure and presentation*

- *Shorten/streamline the Introduction and Methods by moving background material to Supplement.*
- *Improve readability by simplifying subsection nesting, and ensure equations and formatting meet journal standards.*

Response: Following the suggestions of both Reviewers, we added a new table (Table 1) that synthesizes the basic characteristics of the P, SWE, and ET datasets at the regional and global scales. We moved the product details to the Supplement for less experienced readers, along with descriptions of the Thornthwaite-Mather and BIGBANG methods and details on the acquisition and treatment of hydrochemical and isotopic data. In this way, the materials and methods section has been shortened, even though additional text has been added to clarify points raised by the Reviewers. We simplified the nesting of subsections and ensured that equations and formatting meet journal standards.

Response to Reviewer#1

The authors would like to thank the Reviewer for their time in reviewing the manuscript. We appreciate the Reviewer's valuable feedback and insights, which improved some aspects of the manuscript. The lines where changes and integrations are made in the revised version are reported after the response to each comment.

General comments: *The work of Ortenzi et al. presents an interesting dataset for studying mountain catchments in the Mediterranean area. I personally found the introduction and methodology sections rather long, and some basic concepts could easily be moved to the Supporting Information for less experienced readers. Conversely, the description of the monitoring and measurement concepts should be expanded, focusing on important aspects such as the use of the data in the study and their associated uncertainty. In particular, it should be explained why discrete monitoring was performed on those specific dates, why only one thermal image was taken, and whether the temporal and spatial resolution are sufficient to derive robust conclusions. This should be a key point in explaining why other researchers should follow the proposed approach when conducting similar investigations in other catchments. The presentation of the results and the discussion could benefit from more detailed analyses and a more critical assessment of the authors' conclusions. In my view, the data analysis is somewhat questionable: on one hand, I appreciate its simplicity, but on the other, it should not be oversimplified.*

Answer: We thank the reviewer for feedback and suggestions. As also required by reviewer #2, section 2 has been shortened to improve readability (we also simplified the nesting of subsections). In detail, we added a new table in section 2.2.2 (Table 1) that synthesizes the basic characteristics of the P, SWE, and ET datasets at the regional and global scales. We moved the products' description, along with descriptions of the Thornthwaite-Mather and BIGBANG methods, and details on the acquisition and treatment of hydrochemical and isotopic data, to the Supplement. In this way, the materials and methods section has been shortened.

Following the new Table (***Line 196***):

Table 1: Basic characteristics of the datasets.

Sources	**Datasets**	**Reference**	**Spatial-temporal resolution**
The Modified Conditional Merging (MCM) algorithm	P	Pignone et al. (2015) https://www.cimafoundation.org/news/tag/dati/	about 1 km Daily data
The Meteorological Reanalysis Italian Dataset (MERIDA)	P	Bonanno et al. (2019) https://merida.rse-web.it/#download1	about 7 km 3-hours data
European Reanalysis 5th generation (ERA5-Land)	P	Muñoz Sabater et al. (2021); C3S (2022) https://cds.climate.copernicus.eu/datasets/reanalysis-era5-land?tab=overview	0.1°x0.1° (about 7 km) Hourly data
Multi-satellitE Retrievals for Global Precipitation Measurements (GPM IMERG) – Final run	P	Huffman et al. (2023) https://gpm.nasa.gov/data/imerg	0.1°x0.1° (about 10 km) Daily data
IT-SNOW v4.0 product	SWE	Avanzi et al. (2023) https://zenodo.org/records/14093436	about 500 m Daily data
MOD16A2 v061 (MODIS)	ET	Mu et al. (2011); Gallego et al. (2023) https://www.earthdata.nasa.gov/data/catalog/lpcloud-mod16a2-061	about 500 m 8-days data
EUMETSAT LSA SAF (Land Surface Analysis Satellite Application Facility)	ET	Trigo et al. (2011) https://lsa-saf.eumetsat.int/en/	about 5 km Daily data
Global Land Evaporation Amsterdam Model (GLEAM)	ET	Miralles et al. (2025) https://www.gleam.eu/	0.1°x0.1° (about 9 km)

			Daily data
ECO_L3T_JED (ECOSTRESS)	ET	Hook and Halverson (2024) https://www.earthdata.nasa.gov/data/catalog/lpcloud-eco-l3t-jet-002#toc-user-s-guide	about 70 m Daily data

Moreover, as required, we expanded the description of the monitoring and measurement concepts and explained why discrete monitoring was performed on those specific dates. We included some metrics to quantify the uncertainty of the rating curves. Moreover, in the Supplement we reported, when available, the uncertainty across climate products from the literature.

Following the revised text:

Lines 156-166: For both stream gauges, the stream level data (H) were plotted against stream discharge data (Q) taken by the OTT MF Pro flow collected during the 2022-2025 period, yielding curves with the following metrics: S2 section (R^2 = 0.957, RMSE = 0.032 m^3/s); S5 section (R^2 = 0.842, RMSE = 0.054 m^3/s). Overall, the gauging sections exhibit stable channel geometry; thus, the reliability of the rating curves is rated as good (e.g., Tomkins, 2014). Moreover, during the 2022-2025 period, the spot discharge measurements by the OTT MF Pro flow were extended to stream sections S1, S3, and S4 (10 discharge measures in the range 0.70-1.18 m^3/s, Fig. 1b) to investigate GW inflow from limestone aquifers across the larger stream stretch. The spot stream discharge measurements were carried out in periods mostly falling within the recession curve, where baseflow contribution to the stream is dominant: it was possible by checking remotely the stream-level dataloggers (placed in stream sections S2 and S5 in Fig. 1). Combining continuous and discrete monitoring enabled us to develop robust analyses, which help us understand stream segments primarily fed by groundwater moving from the stream headwaters toward section S5.

Lines 193-194: More detailed descriptions of the P, SWE, and ET products, along with uncertainty taken from the literature (when available), are provided in the Supplement.

Regarding a single thermal drone flight, it should not be seen as a lack of information, but rather as an optimization of the cost/benefit ratio to achieve the best possible result, given the multiple factors involved in a complex, mountainous region. We conducted a new drone flight in summer 2025, and the results are now included in the revised manuscript. We added new text to both the methods and results sections regarding this new drone flight. The main groundwater inflow points into the stream were carefully identified using the drone surveys and subsequently confirmed by visual inspections. In these conditions, stream discharge data, coupled with thermal drone imagery, enabled robust conclusions about the locations of stretches with groundwater inflow, thereby defining geochemical sampling points. In the discussion section, we added new text to emphasize that, in mountain regions, the best practice is to obtain basic, insightful information to understand the hydrogeological system. Overall, we believe the data analysis is well-suited to addressing research question (a) and yields robust conclusions that can be easily incorporated into water management systems.

Lines 383-385: To analyze the entire region, the investigated stream stretch was divided into three smaller areas, and two flight campaigns were conducted on January 30, 2025 (take-off at 10:37 local time) and July 31, 2025 (take-off at 12:24 local time).

Lines 398-399: The thermal drone imagery helps identify stream stretches with groundwater inflow, thereby defining geochemical sampling points, as described in section 2.6.

Lines 424-428: From Figure 5a, it can be seen that water 1-2 °C warmer entered the main stream channel from I1 (left bank upstream of S4) and I2 (right bank downstream of S4), with additional diffuse GW inflow along the surveyed stream reach. The survey was repeated in July, but no relevant information was obtained due to dense vegetation along the river path (Figure 5b). Instead, the absence of vegetation allows observation of the temperature contrast between groundwater and the surrounding land (including stream water).

Lines 531-534: In mountain regions, acquiring basic data through complementary techniques provides insights into the hydrogeological system, making it a best practice to optimize the cost-benefit ratio and achieve the best possible results, given the multiple factors involved in complex environmental systems (e.g., Samani and Moghaddam, 2022).

New reference added:

Samani, S., and Kardan Moghaddam, H.: Optimizing groundwater level monitoring networks with hydrogeological complexity and grid-based mapping methods. Environmental Earth Sciences, 81(18), 453, https://doi.org/10.1007/s12665-022-10569-5, 2022.

Comment 1: *a critical uncertainty quantification and propagation to the derived quantities*.

Answer: We thank the Reviewer for this valuable comment. We fully agree that an uncertainty quantification and its propagation to the derived quantities would provide a more robust assessment of the water balance components.

Following the Reviewer's suggestion, we expanded the ensemble of water budget estimates, resulting in a more robust assessment of the variance in the recharge area extent, including the impact of snowmelt. We enhanced the methods section as described in our response to the previous point (see sections 2.2.2). We added two additional ET datasets, GLEAM and ECOSTRESS (Table 1). The latter has a high spatial resolution (70 m), increasing the number of water budget evaluations from 19 to 35. This update is reflected in the water budget synthesis in Tables 2 and 3. Specifically, the inclusion of these new datasets improved quantification of the recharge area, resulting in about 52.00±4.10 when considering only the WS, and about 43.00±4.10 when including the snowmelt contribution. Our enhanced analysis confirms that snowmelt plays a significant role in aquifer recharge, accounting for approximately 18%, consistent with recent studies of other hydrogeological systems in the Central Apennines (Lorenzi et al., 2023; Di Giovanni and Rusi, 2024). Furthermore, when snowmelt is included in the water budget, the recharge area is very close to the catchment area, indicating that GW inflows (Q_{in}^{gw}), net of errors inherent in other components of the water balance, are negligible.

We are aware that the approach used for the water budget may be considered basic, even if innovative in the field of hydrogeology, where the common practice is to evaluate the water budget components' uncertainty (also considering Earth Observation datasets) by assuming no other input and output water balance components, i.e. $Q_{in}^{gw} - Q_{out}^{gw} = 0$ (e.g., Genereux et al., 2005; Yoon et al., 2019; Barbosa et al., 2022). According to Genereux et al. (2002), due to convenience or the lack of available data, simplifying assumptions that eliminate certain water budget components (e.g., assuming negligible net groundwater exchange) can lead to a water budget imbalance. In our study, using different ground-based and EO datasets, we computed the recharge area extent by analyzing $Q_{in}^{gw} - Q_{out}^{gw}$ residuals (specifically, net groundwater exchange is an important component in complex tectonic mountain environments). A comprehensive uncertainty analysis is beyond the scope of the present work, which presents an integrated, preliminary quantitative framework for investigating surface–groundwater interactions in a complex mountain catchment. In this way, although more sophisticated methods could further improve the quantification of the uncertainty in the water budget components, maintaining this level of simplicity is appropriate given the scope and length of the manuscript. The manuscript aims to demonstrate how the use of different datasets can support water budget estimation and what insights can be derived (e.g., the role of snowmelt). Anyway, according to the Reviewer, the manuscript has been completed by adding a new section (4.4) that discusses perspectives on the use of uncertainty quantification and propagation for the water budget components, providing possible directions for implementing such analyses in future studies.

Following the updated Tables and text.

Table 2: Water budget results over the MDU catchment (A of about 44 km^2) for the 2019-2023 period, using different ground-based, satellite, and reanalysis products to compute the water surplus (WS). A* indicates the potential extension of the hydrogeological catchment.

P products	ET products	Q (Mm^3/year)	ΔS (Mm^3/year)	WS = P-ET (Mm^3/year)	$Q_{in}^{gw} - Q_{out}^{gw}$ (Mm^3/year)	Q/WS (-)	A* (km^2)
MCM	MODIS	29.55	-0.964	23.77	4.82	1.24	53.01
	LSAF			25.50	3.09	1.16	49.41
	GLEAM			26.92	1.67	1.10	46.79
	ECOSTRESS			25.26	3.33	1.17	49.88
MERIDA	MODIS			24.78	3.81	1.19	50.84
	LSAF			25.71	2.88	1,15	48.99
	GLEAM			28.56	0.03	1.03	44.12
	ECOSTRESS			26.08	2.51	1.13	48.31
ERA5	MODIS			20.72	7.87	1.43	60.82
	LSAF			23.00	5.59	1.28	54.77
	GLEAM			24.49	4.10	1.21	51.44
	ECOSTRESS			22.70	5.89	1.30	55.51
IMERG	MODIS			20.99	7.60	1.41	60.02
	LSAF			23.89	4.70	1.24	52.74
	GLEAM			24.63	3.96	1.20	51.16
	ECOSTRESS			23.19	5.40	1.27	54.33
USSITA weather station	T–M (FC = 100 mm)			24.70	3.89	1.20	51.01
	T-M (FC = 150 mm)			23.47	5.12	1.26	53.67
BIGBANG	T–M			24.77	3.82	1.19	50.86
			Average value	24.38	**4.22**	**1.22±0.10**	**51.98±4.08**

Table 3: Water budget results over MDU catchments using different ground-based, satellite, and reanalysis products to compute the water surplus (WS), integrated with P_{snow} derived from the IT SNOW dataset. A* indicates the potential extension of the recharge area.

P_{rain} products	ET products	Q (Mm^3/year)	ΔS (Mm^3/year)	WS + P_{snow} = (P_{rain}-ET)+P_{snow} (Mm^3/year)	$Q_{in}^{gw} - Q_{out}^{gw}$ (Mm^3/year)	Q/(WS+P_{snow}) (-)	A* (km^2)
MCM	MODIS	29.55	-0.964	30.60	-2.01	0.97	41.17
	LSAF			30.89	-2.30	0.96	40.78
	GLEAM			32.52	-3.93	0.91	38.74
	ECOSTRESS			34.97	-6.38	0.85	36.03

MERIDA	MODIS		31.40	-2.81	0.94	40.13
	LSAF		30.66	-2.07	0.96	41.10
	GLEAM		33.77	-5.18	0.88	37.31
	ECOSTRESS		27.38	1.21	1.08	46.01
ERA5	MODIS		27.81	0.78	1.06	45.31
	LSAF		28.30	0.29	1.04	44.52
	GLEAM		29.94	-1.35	0.99	42.08
	ECOSTRESS		25.66	2.93	1.15	49.09
IMERG	MODIS		27.18	1.41	1.09	46.35
	LSAF		28.14	0.45	1.05	44.77
	GLEAM		29.12	-0.53	1.01	43.27
	ECOSTRESS		24.74	3.85	1.19	50.93
		Average value	29.57	**-0.98**	**1.01±0.09**	**42.97±4.09**

***Lines (604-632)*: 4.4 Water budget components uncertainty: possible future directions**

Characterizing uncertainty in water budget components, including Earth Observation datasets, is complex, especially in small mountain catchments (Levin et al., 2023; Marti et al., 2023). Generally, the water budget equation includes a residual term (Res), which sums the potential inaccuracies of the datasets used in the calculation (Res_i) and the omitted components in the original equation (Res_o). Most studies simplify water budget uncertainty analyses by neglecting a priori certain components, such as assuming no net groundwater exchange ($Q_{in}^{gw} - Q_{out}^{gw} = 0$); e.g., Yoon et al. (2019) and Barbosa et al. (2022). As reported by Genereux et al. (2002), intercatchment groundwater flows (included in Res_o) cannot be directly measured and are therefore difficult to quantify, which can explain why they are often overlooked in catchment studies. In this way, the entire water imbalance error (Res) is often fully redistributed across budget components, leading to contradictory results such as a decline in the accuracy of corrected hydrological datasets (Luo et al., 2025).
Recently, Zheng et al. (2025) provided a comprehensive list of methods to reduce the impact of data inconsistency for improving water budget closure, such as the constrained ensemble Kalman filter (CEnKF, Pan and Wood, 2006), the multiple collocation (MCL), and proportional redistribution (PR) methods (Abolafia-Rosenzweig et al., 2020; Abhishek et al., 2022; Luo et al., 2023), as well as the post-processing filtering technique (PF) and bias correction method (Munier et al., 2014; Weligamage et al., 2023). It should be noted, as Abolafia-Rosenzweig et al. (2020) pointed out, that the potential incorrect assignment of residuals results from the closure constraints and assumptions imposed by the methods mentioned above. Zheng et al. (2025) proposed a more advanced approach to quantify Res_o by modelling 653 catchments in the USA, which have much larger areas than the Ussita catchment, covering only 44 km². As reported by Muñoz et al. (2024), the use of more sophisticated approaches to evaluate uncertainties in water management in mountain regions requires probability distributions, which can be difficult to obtain in data-scarce regions. Bouaziz et al. (2018) examining the water budget in 58 small limestone fractured/karst catchments along the Meuse River at the border between France and Belgium, concluded that, due to the nature and complexity of the catchments, net groundwater exchanges are the primary cause of water balance discrepancies, which are most significant in small catchments, consistent with the earlier findings of Schaller and Fan (2009). Although evaluating water budget uncertainties is beyond the scope of our study, the multi-source water budget we conducted (35 combinations, 16 of which included the snowmelt contribution) identified groundwater exchanges as the main residual, as reported by Bouaziz et al. (2018). We recognize that more advanced methods for accounting for uncertainty in water budget components could further enhance quantification and refine the assumptions (e.g., by considering the propagation of uncertainty of products reported in the Supplement), even if the results from the multi-source water budget align with those from tracer tests (Nanni et al. 2020; Fronzi et al. 2020; Fronzi et al. 2021; Cambi et al. 2022; Mammoliti et al. 2022).

New added references:

Abhishek, Kinouchi, T., Abolafia-Rosenzweig, R., and Ito, M.: Water Budget Closure in the Upper Chao Phraya River Basin, Thailand Using Multisource Data, Remote Sensing, 14, 173, https://doi.org/10.3390/rs14010173, 2022.

Abolafia-Rosenzweig, R., Pan, M., Zeng, J., and Livneh, B.: Remotely sensed ensembles of the terrestrial water budget over major global river basins: An assessment of three closure techniques, Remote Sens. Environ., 252, 112191, https://doi.org/10.1016/j.rse.2020.112191, 2020.

Barbosa, L. R., Coelho, V.H.R., Gusmão, A.C.V., Fernandes, L.A., da Silva, B. B., Galvão, C.D.O., ... & Almeida, C.D.N.: A satellite-based approach to estimating spatially distributed groundwater recharge rates in a tropical wet sedimentary region despite cloudy conditions. Journal of Hydrology, 607, 127503, https://doi.org/10.1016/j.jhydrol.2022.127503, 2022.

Bouaziz, L., Weerts, A., Schellekens, J., Sprokkereef, E., Stam, J., Savenije, H., & Hrachowitz, M.: Redressing the balance: quantifying net intercatchment groundwater flows. Hydrology and Earth System Sciences, 22(12), https://doi.org/10.5194/hess-22-6415-2018, 6415-6434.

Luo, Z., Li, H., Zhang, S., Wang, L., Wang, S., and Wang, L.: A Novel Two-Step Method for Enforcing Water Budget Closure and an Intercomparison of Budget Closure Correction Methods Based on Satellite Hydrological Products, Water Resour. Res., 59, e2022WR032176, https://doi.org/10.1029/2022WR032176, 2023.

Munier, S., Aires, F., Schlaffer, S., Prigent, C., Papa, F., Maisongrande, P., and Pan, M.: Combining data sets of satellite-retrieved products for basin-scale water balance study: 2. Evaluation on the Mississippi Basin and closure correction model, J. Geophys. Res.-Atmos., 119, 12100–12116, 2014.

Pan, M. and Wood, E.: Data Assimilation for Estimating the Terrestrial Water Budget Using a Constrained Ensemble Kalman Filter, J. Hydrometeorol., 7, 534–547, https://doi.org/10.1175/JHM495.1, 2006.

Weligamage, H., Fowler, K., Peterson, T., Saft, M., Peel, M., and Ryu, D.: Partitioning of Precipitation Into Terrestrial Water Balance Components Under a Drying Climate, Water Resour. Res., 59, e2022WR033538, https://doi.org/10.1029/2022WR033538, 2023.

Yoon, Y., Kumar, S.V., Forman, B.A., Zaitchik, B.F., Kwon, Y., Qian, Y., ... & Mukherjee, A.: Evaluating the uncertainty of terrestrial water budget components over high mountain Asia. Frontiers in Earth Science, 7, 120, https://doi.org/10.3389/feart.2019.00120, 2019.

Comment 2: *Although the authors' conclusions may be correct, the fact that many data are collected at a relatively coarse spatial and temporal scale is an issue that should be discussed more critically and taken into consideration*.

Answer: We thank the Reviewer for this useful comment. We agree that the spatial and temporal resolutions of some datasets could be a limitation that needs further discussion. Note that most of the datasets used in our analysis have relatively fine spatial resolutions of about 1 km (MCM) and 500 m (IT-SNOW and MOD16A2). The other datasets, such as MERIDA, IMERG, ERA5, and EUMETSAT LSA SAF, have resolutions exceeding 5 km, which may introduce additional uncertainty and smoothing of the spatial variability of hydrological processes. As noted in the response to Comment 1, the revised version includes two new ET datasets (GLEAM and ECOSTRESS), the latter with a higher spatial resolution (70 m). Overall, the datasets used are among the most accurate and detailed currently available for the study area, allowing for a multi-source water budget. In the revised version of the manuscript, we will explicitly address this aspect in the Discussion section as follows:

Line 556-559: We carried out a multi-source annual-average water budget (e.g., Vargas Godoy et al., 2024) using the most detailed datasets available for the study area, with the highest spatial and temporal resolutions. Among the products we used, IT-SNOW, MCM, and ECOSTRESS (Table 1) have spatial resolutions (ranging from 70 m to 1 km) that can be suitable for analyzing hydrological processes in mountain watershed areas (e.g., Shuai et al., 2022).

New added references:

Shuai, P., Chen, X., Mital, U., Coon, E. T., & Dwivedi, D.: The effects of spatial and temporal resolution of gridded meteorological forcing on watershed hydrological responses. Hydrol. Earth Syst. Sci., 26, 2245–2276, https://doi.org/10.5194/hess-26-2245-2022, 2022.

Vargas Godoy, M. R., Markonis, Y., Rakovec, O., Jenicek, M., Dutta, R., Pradhan, R. K., ... and Hanel, M. Water cycle changes in Czechia: a multi-source water budget perspective. Hydrology and Earth System Sciences, 28(1), 1-19, doi:10.5194/hess-28-1-2024, 2024.

Comment 3: *It is somewhat disappointing to see a study focusing on groundwater–surface water interactions that does not actually include data (particularly high-resolution time series) from piezometers. It is also unclear how the spring water was sampled and how the springs were gauged. I do not see a dataset that truly captures the spatio-temporal dynamics of groundwater–surface water interactions; rather, it seems to describe the contribution of the springs to the river. But what about the contribution of the river to the groundwater system? Where is the interaction?*.

Answer: We thank the Reviewer for allowing us to better clarify this point. As reported in section 2.4.1, the lack of piezometric observations is common in mountain systems due to the high depths of the groundwater table, which significantly increases drilling costs, making them economically unsustainable. Moreover, the Ussita system is entirely within a National Park, and the environmental constraints limit the feasibility of such drilling.

As stated in the manuscript, the Ussita stream is primarily sustained by groundwater in specific stream stretches, as confirmed by the BFI computed from streamflow data recorded at stream sections S2 and S5 (BFI values of 80% and 90%, respectively). It is a key point because analyzing stream discharge during no-recharge periods provides information about the groundwater system when piezometric data are unavailable. During the no-recharge period, the stream water is spring water; thus, monitoring the stream discharge effectively means monitoring groundwater. The absence of piezometers in the area does not mean groundwater cannot be studied. In section 2.4.1, we explained in greater detail how stream recession analysis can help determine the change in water storage (ΔS).

We explained this point better, as also required by Reviewer#2:

Lines 237-249: Water storage changes (ΔS) must be considered in hydrogeological analysis because the water budget spans four years (2019-2023), making it essential not to overlook this component. As is often the case in mountain systems, piezometric observations were unavailable for calculating ΔS in the Ussita catchment because the groundwater table is deep, and drilling costs and environmental constraints (i.e., the Ussita catchment is entirely within a National Park) limit the feasibility of such measurements. Under these conditions, stream recession analysis can help determine ΔS (e.g., Korkmaz, 1990; Dewandel et al., 2003; Malik and Vojtková, 2012; Płaczkowska et al., 2018). Since the Maillet equation accurately describes stream discharge data of the Ussita stream during no-recharge periods (Di Matteo et al., 2020; Mastrorillo et al., 2020), estimating water storage changes through recession analysis—despite its oversimplification—can still provide valuable insights, especially when included in the mean annual water budget (e.g., Tallaksen, 1995; Raeisi, 2008; Krakauer and Temimi, 2011). As noted by Hameed et al. (2025), baseflow recession analysis serves as a reliable, high-resolution proxy for monitoring groundwater storage. It is particularly useful in regions with limited well-level observation networks or where satellite data are too coarse for small watersheds, such as Ussita. Although the Korkmaz method has been in use for several decades, it remains relevant (e.g., Abirifard et al., 2022).

Thanks to drone flights, we also identified some punctual springs between sections S3 and S5, whose water has been sampled along with stream water during the geochemical measurement campaigns. The discharge of these springs (I1 and I2 in Figure 1b) was monitored using spot measurements with the OTT MF Pro flow meter, as we mainly focused on the total increase in discharge between sections S3-S4 and S4-S5. These increases in discharge are exclusively linked to GW's inflows into the stream.

Lines 405-416: Moving from S1 to S3 (see Fig. 4b for the location of stream sections), the stream discharge during the no-recharge periods (e.g., stream discharge values are equal to baseflow) remains almost stable, indicating that the stream is mainly sustained by the VDP spring; thus, no significant GW inflows are present downstream of the spring up to section S3. It should be noted that downstream of S3 towards S5, no tributaries that feed the Ussita stream are present. Based on mean discharge data from spot OTT MF Pro measurements acquired during the no-recharge periods (2022-2024), an initial increase in stream discharge due to GW inflow was registered between S3 and S4, estimated at about 200 L/s. Downstream of S4, during 2022-2024, a further mean discharge increase of about 450 L/s was measured up to the closing section of the MDU catchment (section S5). Overall, between sections S3 and S5, a mean increase in discharge of about 650 L/s is recorded. This evaluation was also confirmed by the discharge measurements conducted by tracer injection on January 29, 2024 (Fig. S2), at the beginning of a stream recession phase (Fig. 4a). Between S3 and S5, a stream discharge increase of about 695 L/s was registered, a value close to the OTT MF Pro measurements (660 L/s) carried out during the tracer test, confirming the reliability of the streamflow measurement system. Figure 5a also shows the base flow (BF) curves estimated for the S2 and S5 stream sections.

Thanks to the integrated approach, we captured the spatio-temporal dynamics of groundwater–surface water interactions. Gaining stream is one possible scenario of GW-SW interactions (Irvine et al., 2024), which depend strongly on the spatial scale at which they occur. Using streamflow data and other techniques, we demonstrated where and how GW-SW interactions occur. We carefully identified stream reaches where groundwater feeds the stream and highlighted stretches where no interaction occurs (e.g., stream flow remains almost constant between sections S1 and S3). The Ussita stream is a gaining stream, especially in the headwaters (section S2) and at the catchment closure (between sections S3 and S5), where the surface water body increases in volume through seepage, fracture flow, and macropore discharge (e.g., Irvine et al., 2024).

We added the following sentence with a new citation in section 4.1:

Lines 526-528: Thanks to the integrated approach, we captured the spatial-temporal dynamics of groundwater–surface water interactions. Gaining stream is one possible scenario of GW-SW interactions (Irvine et al., 2024), which depend strongly on the spatial scale at which they occur.

Comment 4: *The work would greatly benefit from a model-based interpretation of the results. Would a hydrological model support these conclusions?*.

Answer: We thank the Reviewer for this important comment. We agree that a model-based interpretation could provide additional insight into the processes discussed in this study and how to build a model suitable for reproducing hydrological processes in the area (see, for example, Kavetski and Fenicia, 2011). However, including a full modeling experiment would substantially lengthen the manuscript and fall outside the current scope, which focuses on developing an integrated and preliminary framework based on the available data.

This study, nevertheless, represents a fundamental step toward any future modeling effort (see, for example, Anderson et al., 2015; Enemark et al., 2020). Our results already highlight the challenges of building and calibrating a hydrological model in such a complex catchment, given the heterogeneity of data sources, scales, and hydrological conditions (see, for example, Azimi et al., 2022). In the revised version, we added a new subsection (4.5) to the discussion that emphasizes how evidence from our integrated multidisciplinary approach can support future modelling.

Lines (633-650)**: 4.5 Modelling approach perspectives**

The experimental evidence derived from our integrated, multi-disciplinary approach provides a basis for formulating conceptual hypotheses about the functioning of the groundwater system feeding the Ussita stream. Such hypotheses are a fundamental prerequisite for designing a conceptual model that supports the development of numerical and analytical groundwater models (e.g., Anderson et al., 2015). As reported by Enemark et al. (2020), traditionally, a single conceptual model serves as the basis for model predictions, even if the available data on the groundwater system support more than one conceptualization. Multi-model or flexible-structure approaches offer promising tools for testing an ensemble of conceptual understandings consistent with prior knowledge and observational data (e.g., Rojas et al., 2010; Kavetski and Fenicia, 2011; Mustafa et al., 2020). Refsgaard et al. (2006) reported that the modelling approach may involve a step-by-step procedure that defines a protocol for assessing conceptual model uncertainty. Our data and groundwater hypotheses can help future modelling approaches to define and refine the model structure. In this framework, the analysis of the recharge area extent that we conducted using a multi-source approach can support the definition of boundary conditions and help the model calibration. In detail, the PTV thrust and the outcrop of the Marne a Fucoidi hydrogeological complex (Fig. 1) constitute the no-flow boundaries that support the definition of the initial model domain, while the stream stretches fed by groundwater represent the constant-head boundaries where the baseflow represents the main component of streamflow (e.g., Staudinger et al., 2019). Moreover, streamflow data from the two stream sections (S2 and S5) are invaluable for calibrating and validating numerical models, especially in complex mountain environments; thus, data collection and experimental design represent a substantial field effort and play a pivotal role in reducing uncertainty and constraining future model simulations.

New added References:

Anderson, M.P., Woessner, W.W., & Hunt, R.J.: Modeling Purpose and Conceptual Model, Applied Groundwater Modeling, Elsevier Inc, pp. 27-67, doi: 10.1016/B978-0-08-091638-5.00002-X, 2015.

Enemark, T., Peeters, L. J., Mallants, D., & Batelaan, O.: Hydrogeological conceptual model building and testing: A review. Journal of Hydrology, 569, 310-329, doi: 10.1016/j.jhydrol.2018.12.007, 2019.

Mustafa, S. M. T., Nossent, J., Ghysels, G., and Huysmans, M.: Integrated Bayesian Multi-model approach to quantify input, parameter and conceptual model structure uncertainty in groundwater modeling. Environmental Modelling & Software, 126, 104654, doi: 10.1016/j.envsoft.2020.104654, 2020.

Refsgaard, J. C., Van der Sluijs, J. P., Brown, J., and Van der Keur, P.: A framework for dealing with uncertainty due to model structure error. Advances in water resources, 29(11), 1586-159, doi: 10.1016/j.advwatres.2005.11.013, 2006.

Rojas, R., Kahunde, S., Peeters, L., Batelaan, O., Feyen, L., and Dassargues, A.: Application of a multimodel approach to account for conceptual model and scenario uncertainties in groundwater modelling. Journal of Hydrology, 394(3-4), 416-435, doi: 10.1016/j.jhydrol.2010.09.016, 2010.

Staudinger, M., Stoelzle, M., Cochand, F., Seibert, J., Weiler, M., and Hunkeler, D.: Your work is my boundary condition!: Challenges and approaches for a closer collaboration between hydrologists and hydrogeologists. Journal of Hydrology, 571, 235-243, doi: 10.1016/j.jhydrol.2019.01.058, 2019.

Kavetski, D., & Fenicia, F.: Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights. Water Resources Research, 47(11), doi: 10.1029/2010WR010174, 2011.

Response to Reviewer#2

The authors would like to thank the Reviewer for their time in reviewing the manuscript. We appreciate the Reviewer's valuable feedback and insights, which improved some aspects of the manuscript. The lines where changes and integrations are made in the revised version are reported after the response to each comment.

General comments: *The paper "Exploring groundwater-surface water interactions and recharge in fractured mountain systems: an integrated approach" examines the relationship between surface and groundwater in an Italian mountain headwater catchment using isotope-based analysis. Despite its aim to understand surface and groundwater patterns, there are significant conceptual and methodological weaknesses that affect its suitability for publication in HESS*.

Response: The Reviewer's general comments highlight "significant conceptual and methodological weaknesses" found in the manuscript. Point-by-point responses to the comments aim to clarify some aspects of the manuscript highlighted by the Reviewer. We believe that, thanks to the integrations, the quality of the work has improved.

Comment 1. *The introduction (Lines 1-75) lacks adequate support for an integrated surface-groundwater analysis. The study doesn't employ a robust, quantitative, model-based approach, which is crucial for a comprehensive understanding of the complex interactions between surface and groundwater systems. It would be beneficial to consider relevant literature (e.g., Camporese et al., 2019; Betterle and Bellin, 2024) that provides quantitative assessments of surface and groundwater relationships at plot and hillslope scales. These studies demonstrate the importance of integrating observational data with numerical modeling to capture the full complexity of hydrological processes.*

Response: We thank the reviewer for this comment and agree that numerical modeling can be a powerful tool in many hydrogeological and hydrological studies. However, our manuscript has a different, complementary objective: to provide an integrated experimental characterization of surface–groundwater interactions in a fractured, partially karst mountain catchment in a severely data-scarce environment.

We appreciate the reviewer's suggestion to include quantitative modeling; such an approach is beyond the scope and objectives of the present paper, which focuses on acquiring an experimental baseline of hydrological and hydrogeological data to support future modeling in such complex geological settings. As such, in the revised version of the manuscript, as suggested by the reviewer, we included new text with new literature about the potential support for an integrated surface-groundwater analysis through modelling. This, however, cannot be conducted without comprehensive monitoring through an integrated experimental approach, which is essential before any meaningful quantitative modeling is attempted.

Following the new sentence with the revised text in the introduction section:

Lines 57-61: These features also render quantitative studies through groundwater modeling intricate, as the boundary conditions are difficult to constrain, inter-aquifer exchanges may be significant, and hydrogeological parameters are hard to estimate in heterogeneous carbonate massifs (e.g., Silberstein, 2006; Beven, 2007; Clark et al., 2017; Azimi et al., 2022). As a result, physically based models may remain poorly constrained, or even unfeasible, without a minimum level of monitoring data beyond plot and hillslope scales.

New references added:

Beven, K. (2007). Towards integrated environmental models of everywhere: uncertainty, data and modelling as a learning process. Hydrology and Earth System Sciences, 11(1), 460-467.

Clark, M.P., Bierkens, M.F., Samaniego, L., Woods, R.A., Uijlenhoet, R., Bennett, K.E., Pauwels, V.N.R, Cai, X., Wood, A.W., Peters-Lidard, C.D. (2017). The evolution of process-based hydrologic models: historical challenges and the collective quest for physical realism. Hydrology and Earth System Sciences, 21(7), 3427-3440.

Silberstein, R.P. (2006). Hydrological models are so good, do we still need data? Environmental Modelling & Software, 21(9), 1340-1352.

In the revised version of the manuscript, we also included a new sub-section (4.5 Modelling approach perspectives) to discuss the potential support for an integrated surface-groundwater analysis through modeling, stressing the need for a set of base hydrological, hydrochemical, and hydrogeological information obtainable by an integrated approach, which is essential before any quantitative modeling can be meaningfully attempted.

Subsection 4.5 discusses possible future directions for modelling.

***Lines (633-650)*: 4.5 Modelling approach perspectives**

The experimental evidence derived from our integrated, multi-disciplinary approach provides a basis for formulating conceptual hypotheses about the functioning of the groundwater system feeding the Ussita stream. Such hypotheses are a fundamental prerequisite for designing a conceptual model that supports the development of numerical and analytical groundwater models (e.g., Anderson et al., 2015). As reported by Enemark et al. (2020), traditionally, a single conceptual model serves as the basis for model predictions, even if the available data on the groundwater system support more than one conceptualization. Multi-model or flexible-structure approaches offer promising tools for testing an ensemble of conceptual understandings consistent with prior knowledge and observational data (e.g., Rojas et al., 2010; Kavetski and Fenicia, 2011; Mustafa et al., 2020). Refsgaard et al. (2006) reported that the modelling approach may involve a step-by-step procedure that defines a protocol for assessing conceptual model uncertainty. Our data and groundwater hypotheses can help future modelling approaches to define and refine the model structure. In this framework, the analysis of the recharge area extent that we conducted using a multi-source approach can support the definition of boundary conditions and help the model calibration. In detail, the PTV thrust and the outcrop of the Marne a Fucoidi hydrogeological complex (Fig. 1) constitute the no-flow boundaries that support the definition of the initial model domain, while the stream stretches fed by groundwater represent the constant-head boundaries where the baseflow represents the main component of streamflow (e.g., Staudinger et al., 2019). Moreover, streamflow data from the two stream sections (S2 and S5) are invaluable for calibrating and validating numerical models, especially in complex mountain environments;

thus, data collection and experimental design represent a substantial field effort and play a pivotal role in reducing uncertainty and constraining future model simulations.

New added References:

Anderson, M.P., Woessner, W.W., & Hunt, R.J.: Modeling Purpose and Conceptual Model, Applied Groundwater Modeling, Elsevier Inc, pp. 27-67, doi: 10.1016/B978-0-08-091638-5.00002-X, 2015.

Enemark, T., Peeters, L. J., Mallants, D., & Batelaan, O.: Hydrogeological conceptual model building and testing: A review. Journal of Hydrology, 569, 310-329, doi: 10.1016/j.jhydrol.2018.12.007, 2019.

Mustafa, S. M. T., Nossent, J., Ghysels, G., and Huysmans, M.: Integrated Bayesian Multi-model approach to quantify input, parameter and conceptual model structure uncertainty in groundwater modeling. Environmental Modelling & Software, 126, 104654, doi: 10.1016/j.envsoft.2020.104654, 2020.

Refsgaard, J. C., Van der Sluijs, J. P., Brown, J., and Van der Keur, P.: A framework for dealing with uncertainty due to model structure error. Advances in water resources, 29(11), 1586-159, doi: 10.1016/j.advwatres.2005.11.013, 2006.

Rojas, R., Kahunde, S., Peeters, L., Batelaan, O., Feyen, L., and Dassargues, A.: Application of a multimodel approach to account for conceptual model and scenario uncertainties in groundwater modelling. Journal of Hydrology, 394(3-4), 416-435, doi: 10.1016/j.jhydrol.2010.09.016, 2010.

Staudinger, M., Stoelzle, M., Cochand, F., Seibert, J., Weiler, M., and Hunkeler, D.: Your work is my boundary condition!: Challenges and approaches for a closer collaboration between hydrologists and hydrogeologists. Journal of Hydrology, 571, 235-243, doi: 10.1016/j.jhydrol.2019.01.058, 2019.

Kavetski, D., & Fenicia, F.: Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights. Water Resources Research, 47(11), doi: 10.1029/2010WR010174, 2011.

Comment 2. *The claim of catchment representativeness (Lines 91 and following) is debatable and potentially problematic. Generalizing findings from a single catchment to thousands of others may oversimplify the inherent variability in hydrological systems. The Instantaneous Unit Hydrograph (IUH) varies significantly between cases, as do geomorphological structures and recharge patterns. Rigon et al. (2015) provide valuable insights into this variability, emphasizing the need for caution when extrapolating results from a single catchment study.*

Response: We thank the Reviewer for raising this point. We would like to clarify that we do not extrapolate the hydrological behavior of our study catchment to other basins, nor do we claim that the results observed here are representative of thousands of other catchments. Our work does not aim to generalize catchment responses or to propose a common IUH or GIUH framework, nor does it attempt to model the catchment using these concepts.

On the contrary, the manuscript presents a transferable methodological approach, not transferable hydrological results. This point is now stated in the introduction section:

Lines 85-87: The novelty of this work lies in the integration strategy and workflow, designed to be applicable to other data-scarce mountainous carbonate regions facing increasing drought stress.

Lines (99-100): Although developed for the Ussita catchment, the methodology is designed to be adaptable to other Mediterranean mountain catchments and to fractured systems worldwide with limited high-elevation monitoring.

We intend to show (i) how complex these fractured and partially karst mountain catchments can be, and (ii) how an integrated experimental strategy, combining multiple observational datasets and analytical methods, can be used to gain insight into their functioning. Therefore, what is scalable or applicable to other catchments is the approach itself, not the specific hydrological response of this catchment. In fact, applying this methodology elsewhere would very likely reveal different hydrological behaviors, structures, and recharge patterns, fully consistent with the variability discussed by Rigon et al. (2015). In this regard, the work of Azimi et al. (2022), which also includes Rigon as co-author, clearly demonstrates that, in the Nera catchment, which includes the Ussita catchment, numerical modeling would not have been feasible without prior experimental characterization (i.e., in that case the recharge area resulted higher than the catchment one as delineated by a full experimental approach based on the application of tracer tests). This reinforces our point that field-based data and integrated experimental analyses are indispensable prerequisites for any reliable modeling effort in such fractured and tectonically complex environments.

In the conclusions of the revised manuscript, we further clarified the points above to prevent any misunderstanding and explicitly state that our study does not aim to generalize catchment behavior.

Lines 669-671: Although the hydrogeological functioning of carbonate mountain catchments remains site-specific, the integrated workflow proposed here is transferable and can guide field campaigns in similar data-scarce Mediterranean catchments, supporting the identification of the dominant controls on GW inflows to streams and the delineation of recharge areas.

Comment 3. *The high climatic variability in Mediterranean catchments makes it challenging to prove the general validity of the findings. Recharge patterns can vary significantly across Mediterranean regions, and it's unclear how this specific catchment's hydrological patterns could be considered generally valid for such a diverse area. It would be advisable to examine multiple catchments with various geomorphological and climatic patterns to establish more robust, generalizable conclusions. Jiang et al. (2015) demonstrate the importance of analyzing a large number of catchments (108 in their case) to comprehensively understand climate change and human activity impacts on runoff and water resources. This approach allows for capturing regional variability and identifying common or divergent patterns among different catchments. Therefore, analyzing a single catchment and extending its findings may oversimplify hydrological complexities.*

Response: We agree that Mediterranean catchments exhibit high climatic and geomorphological variability, and that large-sample analyses such as Jiang et al. (2015) are valuable for identifying regional-scale patterns, particularly when based on modeled or readily available hydrological indicators. However, our study has a very different scope and purpose.

As discussed in our response to comment 2, we do not aim to generalize hydrological responses. What our manuscript proposes to scale is the methodology, not the results (see answer to comment 2). The integrated experimental approach we present, combining hydrological measurements, hydrochemistry, and isotope analysis, and thermal drone investigation, is unique in the literature and is intended to serve as a framework that can be replicated in other data-scarce mountainous catchments, each of which would naturally exhibit distinct hydrological behavior.

Large-sample analysis of the type cited by the Reviewer is not directly comparable to the kind of field-intensive experimental characterization conducted here. Detailed field investigations in fractured and tectonically complex mountain environments require significant logistical effort, infrastructure, and long-term monitoring, which are essential to developing a reliable numerical modelling approach.

For these reasons, and we trust the Reviewer is well aware of this, applying such a detailed experimental approach to dozens or even hundreds of catchments is not feasible and is rarely attempted in the hydrological literature.

Comment 4. *The geological introduction (Lines 97-104) provides a classic overview but lacks specific information on how different formations contribute quantitatively to groundwater flow. To strengthen the groundwater investigation aspect, it would be beneficial to address questions about groundwater domain conceptualization, such as: How many layers are present? What is the vertical compartmentalization, or are there semi-confined horizons? What is the connectivity on the z-axis between these formations? How are fractures interpreted - as preferential pathways or double continua? Enemark et al. (2019) could provide useful insights into integrating geological information with hydrological modeling, offering a more comprehensive approach to understanding the physical reality of the system.*

Response: Section 2.1, "The Ussita experimental catchment," details the main hydrogeological complexes and structural features in the study area, and the discussion clearly states how different hydrogeological complexes contribute quantitatively to groundwater flow. Only after an in-depth analysis of all the data was it possible to draw this information. It is reported in section 4.1 as follows:

Lines 534-540: In detail, the main advantage of the integrated approach is the investigation of the spatial distribution of GW inflow along the Ussita stream, revealing that most of the baseflow in the stream from S1 to S3 is sustained by the VDP spring (Q ≃ 220 L/s), with a huge baseflow increase between S3 and S5 sections (Q ≃ 650 L/s) delineating, two different sources of alimentation: i) the Maiolica Complex for the VDP spring (EC ≃ 210 μS/cm; SO_4 ≃ 2.5 mg/l), and ii) the Base Limestone Complex for punctual and linear springs downstream of the S3 section and up to S5 (EC ≃ 310 μS/cm; SO_4 ≃ 18.7 mg/l), with some mixing water with intermediate characteristics in the I1 sampling point related to the Maiolica Complex contribution (Ussita left bank, Fig. 1, EC ≃ 264 μS/cm; SO_4 ≃ 6.9 mg/l).

The Reviewer's questions concern the practical implementation of the numerical modeling, which requires a 3-D schematization of the main aquifers (including their interconnections, if present) and the definition of their hydrogeological properties, which are useful but outside the scope of our study. The objective of our work is not to define a conceptual model for numerical modeling but to acquire information to support hypotheses about the groundwater system through an integrated approach, leaving to future studies the attempt to define a conceptual model for numerical modeling, which also include geological insights or expert interpretation to define the tridimensional system structure, hydrogeological parameters, etc. (Anderson and Woessner, 1992; Enemark et al. 2019). Considering the complexity of the study catchment, which is located in a highly heterogeneous and tectonically fractured/deformed carbonate system, simple hydrogeological assumptions, such as laterally continuous layers, well-defined confined or unconfined aquifers, or vertically structured flow domains, can oversimplify the problem. For this reason, a quantitative definition of discrete layers or vertical compartments is beyond what can be supported by available field information and would risk introducing unwarranted assumptions.

As noted in the response to comment 1, we added a new subsection (4.5) to discuss how evidence from our integrated multidisciplinary approach can support future modelling (***lines 633-650***).

Comment 5. *Figure 1 would benefit from including a map of Italy to frame the catchment's location. This addition would provide important context for readers unfamiliar with the study area and help situate the research within the broader geographical landscape of Italy.*

Response: Figure 1 already includes a map of Italy to frame the catchment's location.

Comment 6. *The use of different climate products (ERA5, GRISO, etc.) requires more explanation regarding their disparate input datasets, spatial resolutions, and interpretations of phenomena. These products have different underlying data sources, spatial resolutions, and focus on different aspects of climate. An analysis of biases and their impacts on basin recharge would be helpful to understand how these differences affect the study's results. Consider discussing the impact of time steps on bias correction and providing relevant plots. It would be beneficial to address questions such as: What are the differences in quantiles and extreme events between these products? Do these products undergo bias correction (except for reanalysis)? What is the impact of these products' heterogeneous nature on basin recharge at this scale? Canon et al. (2015) could provide valuable guidance on addressing these issues.*

Response: We thank the Reviewer for this detailed observation. We agree that climate products differ in terms of input data, spatial resolution, and methodological assumptions, and that these aspects are important in climate-focused studies. However, the scope of our work differs substantially from that of Cannon et al. (2015) or other climatological investigations centered on high-resolution bias correction or extreme-event analysis.

Our analysis is conducted on a monthly timescale to capture broad seasonal climatic patterns rather than short-duration extremes. For this reason, differences in quantiles of sub-monthly extremes, which are highly relevant for climate-model evaluation, are not directly applicable to our experimental hydrogeological framework. We fully acknowledge that all climatic products carry uncertainties, but it is important to note that observational datasets also include errors, and reanalysis products similarly require careful interpretation due to scale mismatch and interpolation biases (e.g., Ebert et al., 2007; Massari et al., 2017; Maggioni et al., 2018). In our work, we used the monthly meteo-climatic data from satellite and reanalysis products to derive quantitative mean annual recharge estimates instead of performing climate impact assessments.

We fully agree that an uncertainty quantification and its propagation to the derived quantities would provide a more robust assessment of the water balance components. In the revised manuscript, as also requested by reviewer 1, to complete the manuscript, a new section (4.4) has been added to discuss perspectives on the use of uncertainty quantification and propagation for the water budget components, providing possible directions for implementing such analyses in future studies.

However, a comprehensive uncertainty analysis is beyond the scope of the present work, which primarily aims at presenting an integrated, preliminary quantitative framework for investigating surface–groundwater interactions in complex catchments. The manuscript focuses on demonstrating which instruments, methods, and datasets can support this type of analysis and what insights can be derived.

We expanded the ensemble of water budget estimates, resulting in a more robust assessment of the variance in the recharge area extent, including the impact of snowmelt. We enhanced the methods section (2.2.2). We added two additional ET datasets, GLEAM and ECOSTRESS (Table 1). The latter has a high spatial resolution (70 m), increasing the number of water budget evaluations from 19 to 35. This update is reflected in the water budget synthesis in Tables 2 and 3. Specifically, the inclusion of these new datasets improved quantification of the recharge area, resulting in about 52.00±4.10 when considering only the Water Surplus (WS), and about 43.00±4.10 when including the snowmelt contribution. Our enhanced analysis confirms that snowmelt plays a significant role in aquifer recharge, accounting for approximately 18%, consistent with recent studies of other hydrogeological systems in the Central Apennines (Lorenzi et al., 2023; Di Giovanni and Rusi, 2024). Furthermore, when snowmelt is included in the water budget, the recharge area is very close to the catchment area, indicating that GW inflows (Q_{in}^{gw}), net of errors inherent in other components of the water balance, are negligible.

We are aware that the approach used for the water budget may be considered basic, even if innovative in the field of hydrogeology, where the common practice is to evaluate the water budget components' uncertainty (also considering Earth Observation datasets) by assuming no other input and output water balance components, i.e. $Q_{in}^{gw} - Q_{out}^{gw} = 0$ (e.g., Genereux et al., 2005; Yoon et al., 2019; Barbosa et al., 2022). According to Genereux et al. (2002), due to convenience or the lack of available data, simplifying assumptions that eliminate certain water budget components (e.g., assuming negligible net groundwater exchange) can lead to a water budget imbalance. In our study, using different ground-based and EO datasets, we computed the recharge area extent by analyzing $Q_{in}^{gw} - Q_{out}^{gw}$ residuals (specifically, net groundwater exchange is an important component in complex tectonic mountain environments). A comprehensive uncertainty analysis is beyond the scope of the present work, which presents an integrated, preliminary quantitative framework for investigating surface–groundwater interactions in a complex mountain catchment. In this way, although more sophisticated methods could further improve the quantification of the uncertainty in the water budget components, maintaining this level of simplicity is appropriate given the scope and length of the manuscript. The manuscript aims to demonstrate how the use of different datasets can support water budget estimation and what insights can be derived (e.g., the role of snowmelt). Anyway, according to the Reviewer, the manuscript has been completed by adding a new section (4.4) that discusses perspectives on the use of uncertainty quantification and propagation for the water budget components, providing possible directions for implementing such analyses in future studies.

***Lines (604-632)*: 4.4 Water budget components uncertainty: possible future directions**

Characterizing uncertainty in water budget components, including Earth Observation datasets, is complex, especially in small mountain catchments (Levin et al., 2023; Marti et al., 2023). Generally, the water budget equation includes a residual term (Res), which sums the potential inaccuracies of the datasets used in the calculation (Res_i) and the omitted components in the original equation (Res_o). Most studies simplify water budget uncertainty analyses by neglecting a priori certain components, such as assuming no net groundwater exchange ($Q_{in}^{gw} - Q_{out}^{gw} = 0$); e.g., Yoon et al. (2019) and Barbosa et al. (2022). As reported by Genereux et al. (2002), intercatchment groundwater flows (included in Res_o) cannot be directly measured and are therefore difficult to quantify, which can explain why they are often overlooked in catchment studies. In this way, the entire water imbalance error (Res) is often fully redistributed across budget components, leading to contradictory results such as a decline in the accuracy of corrected hydrological datasets (Luo et al., 2025).
Recently, Zheng et al. (2025) provided a comprehensive list of methods to reduce the impact of data inconsistency for improving water budget closure, such as the constrained ensemble Kalman filter (CEnKF, Pan and Wood, 2006), the multiple collocation (MCL), and proportional redistribution (PR) methods (Abolafia-Rosenzweig et al., 2020; Abhishek et al., 2022; Luo et al., 2023), as well as the post-processing filtering technique (PF) and bias correction method (Munier et al., 2014; Weligamage et al., 2023). It should be noted, as Abolafia-Rosenzweig et al. (2020) pointed out, that the potential incorrect assignment of residuals results from the closure constraints and assumptions imposed by the methods mentioned above. Zheng et al. (2025) proposed a more advanced approach to quantify Res_o by modelling 653 catchments in the USA, which have much larger areas than the Ussita catchment, covering only 44 km². As reported by Muñoz et al. (2024), the use of more sophisticated approaches to evaluate uncertainties in water management in mountain regions requires probability distributions, which can be difficult to obtain in data-scarce regions. Bouaziz et al. (2018) examining the water budget in 58 small limestone fractured/karst catchments along the Meuse River at the border between France and Belgium, concluded that, due to the nature and complexity of the catchments, net groundwater exchanges are the primary cause of water balance discrepancies, which are most significant in small catchments, consistent with the earlier findings of Schaller and Fan (2009). Although evaluating water budget uncertainties is beyond the scope of our study, the multi-source water budget we conducted (35 combinations, 16 of which included the snowmelt contribution) identified groundwater exchanges as the main residual, as reported by Bouaziz et al. (2018). We recognize that more advanced methods for accounting for uncertainty in water budget components could further enhance quantification and refine the assumptions (e.g., by considering the propagation of uncertainty of products reported in the Supplement), even if the results from the multi-source water budget align with those from tracer tests (Nanni et al. 2020; Fronzi et al. 2020; Fronzi et al. 2021; Cambi et al. 2022; Mammoliti et al. 2022).

New added references:

Abhishek, Kinouchi, T., Abolafia-Rosenzweig, R., and Ito, M.: Water Budget Closure in the Upper Chao Phraya River Basin, Thailand Using Multisource Data, Remote Sensing, 14, 173, https://doi.org/10.3390/rs14010173, 2022.

Abolafia-Rosenzweig, R., Pan, M., Zeng, J., and Livneh, B.: Remotely sensed ensembles of the terrestrial water budget over major global river basins: An assessment of three closure techniques, Remote Sens. Environ., 252, 112191, https://doi.org/10.1016/j.rse.2020.112191, 2020.

Barbosa, L. R., Coelho, V.H.R., Gusmão, A.C.V., Fernandes, L.A., da Silva, B. B., Galvão, C.D.O., ... & Almeida, C.D.N.: A satellite-based approach to estimating spatially distributed groundwater recharge rates in a tropical wet sedimentary region despite cloudy conditions. Journal of Hydrology, 607, 127503, https://doi.org/10.1016/j.jhydrol.2022.127503, 2022.

Bouaziz, L., Weerts, A., Schellekens, J., Sprokkereef, E., Stam, J., Savenije, H., & Hrachowitz, M.: Redressing the balance: quantifying net intercatchment groundwater flows. Hydrology and Earth System Sciences, 22(12), https://doi.org/10.5194/hess-22-6415-2018, 6415-6434.

Luo, Z., Li, H., Zhang, S., Wang, L., Wang, S., and Wang, L.: A Novel Two-Step Method for Enforcing Water Budget Closure and an Intercomparison of Budget Closure Correction Methods Based on Satellite Hydrological Products, Water Resour. Res., 59, e2022WR032176, https://doi.org/10.1029/2022WR032176, 2023.

Munier, S., Aires, F., Schlaffer, S., Prigent, C., Papa, F., Maisongrande, P., and Pan, M.: Combining data sets of satellite-retrieved products for basin-scale water balance study: 2. Evaluation on the Mississippi Basin and closure correction model, J. Geophys. Res.-Atmos., 119, 12100–12116, 2014.

Pan, M. and Wood, E.: Data Assimilation for Estimating the Terrestrial Water Budget Using a Constrained Ensemble Kalman Filter, J. Hydrometeorol., 7, 534–547, https://doi.org/10.1175/JHM495.1, 2006.

Weligamage, H., Fowler, K., Peterson, T., Saft, M., Peel, M., and Ryu, D.: Partitioning of Precipitation Into Terrestrial Water Balance Components Under a Drying Climate, Water Resour. Res., 59, e2022WR033538, https://doi.org/10.1029/2022WR033538, 2023.

Yoon, Y., Kumar, S.V., Forman, B.A., Zaitchik, B.F., Kwon, Y., Qian, Y., ... & Mukherjee, A.: Evaluating the uncertainty of terrestrial water budget components over high mountain Asia. Frontiers in Earth Science, 7, 120, https://doi.org/10.3389/feart.2019.00120, 2019.

Comment 7. *The baseflow quantification method could be more robust. Equation (4) is a top-down approach that may need validation before being employed. The use of simplified analytical equations to estimate baseflow and groundwater recharge from stream discharge records has significant limitations, particularly in complex hydrogeological systems. Such equations often fail to account for the heterogeneity of hydraulic properties, vertical stratification of aquifer layers, and intricate interactions between groundwater and surface water bodies, including leakage phenomena. Consider the limitations highlighted by Halford &*

Mayer (2000), Bresciani et al. (2016), Woessner (2000), and Betterle and Bellin (2024). These studies collectively underscore the inadequacy of simplified baseflow equations and advocate for the use of numerical groundwater flow models to accurately quantify baseflow and aquifer-stream exchanges. Betterle and Bellin (2024) even more underlined the role of groundwater fluxes in complex landscapes. Groundwater-fed surface drainage networks under various morphological and geological settings show a complex behavior, following a Gamma distribution, whose parameters are modulated by recharge, hydraulic conductivity, and topography, to be carefully assessed and validated for each case study.

Response: We acknowledge the Reviewer's concerns regarding the use of simplified analytical equations to estimate baseflow in a complex mountain catchment. The approach we used to estimate the BF (the LH digital filter method) is one of the possible techniques. The results of BF separation should not be considered as exact fluxes, but as diagnostic indicators of groundwater contribution to streamflow (***lines 205-208***). We think that simplified analytical approaches have limitations, particularly in hydrogeologically complex environments. Most of the papers suggested by the reviewer analyzed hypothetical catchments in mountain areas or combined hypothetical and field sites (e.g., Halford and Mayer, 2000). Among the suggested papers, Halford and Mayer (2000) presented the results of BF separation by a modelling approach considering only one karst limestone and dolomite catchment having an extension of more than 20,000 km^2, with a minimum estimated BF of 44 m^3/s, the latter of two orders of magnitude higher than the mean BF of our case study. Moreover, as highlighted by Halford and Mayer (2000), multiple processes (e.g., drainage from bank storage) can affect recession curves in such systems. The approach we used to estimate the BF (LH method) is one of the possible techniques, but in our case, the problems highlighted by Halford and Mayer (2000) are absent. According to Xie et al. (2020), the digital filter methods perform well in many catchments, including those with high infiltration rates, such as the Ussita catchment. Moreover, a recent paper by Mo et al. (2025) examined a karst basin and found that the LH method performed well. However, we also believe that evaluating the accuracy of baseflow estimation methods remains an open question in the literature, because BF cannot be directly measured except during periods of no recharging or prolonged dry spells (i.e., in these periods, stream discharge corresponds to BF). Thanks to the Master Recession Curves, we found that during no-recharge periods, stream discharge is described by the Maillet equation (i.e., linear reservoir depletion). Using the recession constant, we computed the k-filter and verified that the derived BF were appropriate for the available recession periods. In other words, the computed BF data by the LH method were validated against continuous streamflow data during no-recharge periods (S2 and S5 sections), once again highlighting the importance of acquiring data in complex mountain environments to constrain BF estimates as effectively as possible.

As noted in the response to the previous point, numerical groundwater models can be helpful but may introduce significant structural uncertainty and risk, yielding results that are not physically constrained. For this reason, our objective is not to provide a precise quantification of groundwater fluxes but rather to characterize the system's integrated hydrological response. Since we found a linear reservoir depletion, the information on baseflow dynamics we obtained, even if it is not an exact estimator, can be considered as a diagnostic tool to help interpret the system's behavior in combination with hydrochemical and isotopic evidence.

Revised text:

Lines 208-230: Although the evaluation of BF separation methods remains an open question in the literature, Xie et al. (2020) showed that digital-filter methods perform well across 1145 catchments in the USA, including systems characterized by high infiltration rates, such as the Ussita catchment. Therefore, we adopted the one-parameter recursive digital filter proposed by Lyne and Hollick (1979) (Eq. 1), which provides an intuitively satisfactory representation of baseflow (e.g., Duncan, 2019) and has also been successfully applied in karst catchments (Mo et al., 2025). In this method, the baseflow at time t (B_t) is separated from the total stream discharge at time t (Q_t) using a filter parameter k that effectively separates low-frequency signals (associated with baseflow). Selecting an appropriate k-parameter value remains an open question, even though the Lyne and Hollick (1979) method is considered helpful for comparative hydrology and regionalization, as it can be used to consistently characterize differences between catchments (Landson, 2013). According to Nathan and McMahon (1990), applying the Lyne and Hollick filter requires selecting several passes and a single k-parameter value. More specifically, the k-parameter affects the degree of baseflow attenuation, and the number of passes determines the degree of smoothing (Nagy et al., 2024). The k-parameter typically ranges from 0.90 to 0.99 (e.g., Kang et al., 2022). One of the main sources of uncertainty in digital filter methods is the linear reservoir assumption (Xie et al., 2020), which holds for the Ussita catchment because discharge during the recession period is described by the Maillet equation (e.g., Di Matteo et al., 2020; Mastrorillo et al., 2020). In this way, Chapman (1991) and Tan et al. (2009) linked the k value to the recession coefficient α ($k = e^{-\alpha t}$) estimated by the Master Recession Curve (MRC), considering daily or hourly discharge time steps (e.g., Tallaksen, 1995; Posavec et al., 2006; Gregor and Malik, 2012; Di Matteo et al., 2017; Carlotto and Cheffe, 2019; Di Matteo et al., 2020). Following this approach, the k-parameter value for each catchment was estimated, and the recursive filter was applied by setting the number of passes to three (forward–backward-forward) to minimize phase distortion effects on the peak values (Nathan and McMahon, 1990). As reported by Kang et al. (2022), it is necessary to verify that the selected digital filter is appropriate for separating baseflow during the dry season (when baseflow represents stream discharge). In other words, to constrain baseflow estimates as effectively as possible, the BF data computed by the LH method must be validated against continuous streamflow data during no-recharge periods (S2 and S5 sections for the Ussita catchment).

New added references:

Mo, C., Jiang, C., Long, S., and Cen, W.: Comprehensive evaluation and attribution analysis of baseflow variation in a typical karst basin, Southwest China. Journal of Hydrology: Regional Studies, 57, 102185, doi:10.1016/j.ejrh.2025.102185, 2025.

Xie, J., Liu, X., Wang, K., Yang, T., Liang, K., & Liu, C. (2020). Evaluation of typical methods for baseflow separation in the contiguous United States. Journal of Hydrology, 583, doi:10.1016/j.jhydrol.2020.124628,124628.

Comment 8. *Section 2.7: The water budget calculation should consider leakage for closure. Without accounting for leakage, the water budget may not be accurately represented, potentially leading to misinterpretation of the hydrological processes within the catchment.*

Response: We thank the Reviewer. The issue of groundwater exchange is of paramount importance, especially in complex mountain hydrogeological systems where the water budget closure is strictly dependent on groundwater outflow to neighboring hydrogeological systems and groundwater inflow into the catchment (it is illustrated graphically in the schematic representation of Fig. 4). In this way, we computed in sections 3.3.1 and 3.3.2 the difference of these two GW components as residuals in the water budget, which has helped us understand that very few external GW inflows come into the catchment. We discussed this point in the new section 4.4 (see the response to comment 6), introducing a new paper by Bouaziz et al. (2018), who examining the water budget in 58 small limestone fractured/karst catchments along the Meuse River at the border between France and Belgium, concluded that, due to the nature and complexity of the catchments, net groundwater exchanges are the primary cause of water imbalances, which are most significant in small catchments, consistent with the earlier findings of Schaller and Fan (2009). Anyway, in the Supplement we reported, when available, the uncertainty across climate products from the literature, which can be useful for considering how uncertainty propagates in future analyses (***lines 629-632***).

Comment 9. *Section 2.8: The estimation of water storage changes using stream recession analysis and the Maillet exponential relationship may be oversimplified. While this method provides a rough approximation, it fails to capture the inherent complexity and heterogeneity of groundwater systems. A more robust approach could involve developing a three-dimensional geological model of the aquifer system, accounting for the spatial variability of hydraulic properties. Recent groundwater modeling studies, such as Schorpp et al. (2025), have emphasized the importance of incorporating heterogeneity into numerical models to accurately represent the subsurface environment. By incorporating heterogeneity into the geological model and subsequently into the numerical groundwater model, researchers can more accurately simulate the spatial and temporal variations in water storage, accounting for the effects of local-scale variations in hydraulic properties and geological structures.*

Response: Developing a three-dimensional geological model of the hydrogeological system is beyond the scope of this manuscript, although the collected data can support hypotheses about groundwater circulation that may aid further modeling studies. We discussed this point and added a new section 4.5 (see the response to comment 1).

Since the Maillet equation accurately describes stream discharge during no-recharge periods, the estimation of water storage changes by recession analysis, despite its oversimplification, can still offer valuable insights, especially when incorporated into the mean annual water budget (e.g., Tallaksen, 1995; Raeisi, 2008; Krakauer and Temimi, 2011). Recently, a study by Hameed et al. (2025) investigated the reliability of baseflow recession analysis for estimating long-term changes in groundwater storage in minimally disturbed watersheds in the USA (1144 gauged watersheds). Despite differences in magnitude, storage changes derived from baseflow recession analysis show remarkably similar spatial patterns to those derived from GRACE (for 75% of the catchment analyzed). As reported in Hameed et al. (2025), "*that baseflow recession analysis is a reliable, high-resolution proxy for groundwater storage monitoring. It is especially useful in regions with sparse well-level observation networks or where satellite data are too coarse for small watersheds*". The hydrograph recession analysis was also used to estimate streamflow sensitivity to changes in water storage across Europe (725 catchments; Berghuijs et al., 2016). Thus, the method we used (Korkmaz, 1990) provides an estimate of changes in reservoir water storage for the Ussita catchment (e.g., no well-level observation network), which is to be considered satisfactory for computing the mean annual water budget. Although the Korkmaz method has been around for several decades, it is still in use, as shown by recent references (e.g., Abirifard et al., 2022). It addresses two main constraints: the validity of the extrapolated recession curves and the availability of annual recession periods (since, in some years, recharge did not completely stop for a certain period during the observed recession). These issues do not apply to Ussita because the recession curves are clearly defined for each year and are well-fitted by the Maillet equation.

Overall, line pointed out that the average annual change in water storage we estimated is two orders of magnitude smaller than the water surplus.

We revised section 2.4.1 of the manuscript, including references supporting the method used to estimate mean changes in water storage.

Lines 237-249: Water storage changes (ΔS) must be considered in hydrogeological analysis because the water budget spans four years (2019-2023), making it essential not to overlook this component. As is often the case in mountain systems, piezometric observations were unavailable for calculating ΔS in the Ussita catchment because the groundwater table is deep, and drilling costs and environmental constraints (i.e., the Ussita catchment is entirely within a National Park) limit the feasibility of such measurements. Under these conditions, stream recession analysis can help determine ΔS (e.g., Korkmaz, 1990; Dewandel et al., 2003; Malik and Vojtková, 2012; Płaczkowska et al., 2018). Since the Maillet equation accurately describes stream discharge data of the Ussita stream during no-recharge periods (Di Matteo et al., 2020; Mastrorillo et al., 2020), estimating water storage changes through recession analysis—despite its oversimplification—can still provide valuable insights, especially when included in the mean annual water budget (e.g., Tallaksen, 1995; Raeisi, 2008; Krakauer and Temimi, 2011). As noted by Hameed et al. (2025), baseflow recession analysis serves as a reliable, high-resolution proxy for monitoring groundwater storage. It is particularly useful in regions with limited well-level observation networks or where satellite data are too coarse for small watersheds, such as Ussita. Although the Korkmaz method has been in use for several decades, it remains relevant (e.g., Abirifard et al., 2022).

New references added:

Abirifard, M., Birk, S., Raeisi, E., Sauter, M. (2022). Dynamic volume in karst aquifers: Parameters affecting the accuracy of estimates from recession analysis. Journal of Hydrology, 612, 128286.

Dewandel, B., Lachassagne, P., Bakalowicz, M., Weng, Ph., Al-Malki A. (2003). Evaluation of aquifer thickness by analysing recession hydrographs. Application to the Oman ophiolite hard-rock aquifer. Journal of Hydrology 274, 248–269.

Hameed, M., Nayak, M.A., Ahangar, M.A. (2025). Groundwater storage changes in the United States using baseflow recession method: Comparison with GRACE and well observations. Journal of Hydrology: Regional Studies, 62, 102946.

Krakauer, N.Y., Temimi, M. (2011). Stream recession curves and storage variability in small watersheds. Hydrology and Earth System Sciences, 15(7), 2377-2389.

Malik, P., Vojtková, S. (2012). Use of recession-curve analysis for estimation of karstification degree and its application in assessing overflow/underflow conditions in closely spaced karstic springs. Environmental and Earth Sciences 65, 2245–2257.

Płaczkowska, E., Siwek, J., Maciejczyk, K., Mostowik, K., Murawska, M., Rzonca, B. (2018). Groundwater capacity of a flysch-type aquifer feeding springs in the Outer Eastern Carpathians (Poland). Hydrology Research, 49(6), 1946-1959.

Raeisi, E. (2008). Groundwater storage calculation in karst aquifers with alluvium or no-flow boundaries. Journal of Cave and Karst Studies, 70(1), 62–70.

Formal issues:

Comment 10. *Lines 78-81 present direct, open questions in the Introduction. While these questions effectively highlight the study's objectives, they might be more impactful if presented as Highlights. This would allow readers to quickly grasp the key focus areas of the research without disrupting the flow of the introduction.*

Response: We modified them accordingly. We added two highlights.

Comment 11. *Equation 2 would benefit from formatting in a Math editor to improve clarity and professional presentation.*

Response: It has been modified accordingly.

Comment 12. *The sub-paragraph structure (e.g., 2.2.2.2) could be simplified for better readability. While detailed organization can be helpful, excessive subdivision can make the paper structure cumbersome, confusing, and annoying for readers.*

Response: We considered the suggestions. In detail, we added a new table in section 2.2.2 (Table 1) that synthesizes the basic characteristics of the P, SWE, and ET datasets at the regional and global scales. We moved the products' description, along with descriptions of the Thornthwaite-Mather and BIGBANG methods, and details on the acquisition and treatment of hydrochemical and isotopic data, to the Supplement. In this way, the materials and methods section has been shortened, simplifying the nesting of subsections.

Comment 13: *The number of authors (13) seems high, given that only 9 actively participated in the work. Consider acknowledging those involved solely in data collection rather than including them as co-authors. This approach would more accurately reflect the contributions to the research and writing process.*

Response: Authorship decisions follow the journal guidelines and the standard practices of our research institutions. All authors listed made substantial contributions to the study, either through field design, data collection, and curation. Moreover, all the authors contributed to writing the original draft and revising and editing it.

In highly field-intensive experimental studies such as ours, conducted in challenging mountainous terrain and involving long-term monitoring, tracer experiments, and coordinated multidisciplinary work, data collection is a core scientific activity. It requires rigorous planning, specialized technical expertise, continuous field presence, and responsibility for the integrity and interpretation of the resulting dataset. In the hydrological community, such contributions are universally recognized as legitimate intellectual input and are standard criteria for authorship in experimental research.

For this reason, excluding colleagues who played essential roles in the design, execution, and field campaign would misrepresent the true nature of the work. We trust that this clarification makes it clear that all listed co-authors fully meet the authorship criteria.

For these reasons, we maintained the current author list and clarified each author’s role in the Author Contributions section.

Conclusions of the Reviewer: *In conclusion, while the study relies on valuable data, it appears as a preliminary investigation, focusing primarily on isotope data collection and interpretation. The work would benefit from addressing issues of novelty, methodological clarity, conceptual solidity, and case-study representativeness. To strengthen the paper, consider incorporating more robust modeling approaches, expanding the analysis to multiple catchments, and providing more detailed explanations of the methodologies used.*

Response: Although the Reviewer highlighted that the manuscript is based on valuable data, the concluding remarks suggest that he did not fully consider the manuscript's overall assessment, overlooking the significant effort involved in monitoring groundwater processes (not only isotope data collection and interpretation) in a data-scarce, structurally complex mountain setting. High-quality field observations of this kind are extremely rare in the Central Apennines and represent precisely the type of foundational knowledge that modeling studies depend on. In regions where the subsurface architecture, hydraulic properties, and boundary conditions are unknown, such experimental datasets are indispensable, and their scientific value goes well beyond a “preliminary investigation.” It is important to emphasize that our study provides new, hard-to-obtain empirical evidence on groundwater–surface water interactions in a catchment where modelling alone cannot advance understanding without prior field-based constraints. We trust that this clarification and the point-by-point responses highlight the significance of our contribution to improving hydrogeological knowledge in data-scarce mountain regions, which can help future modeling analysis.