REVIEWER 2

General comments

The manuscript is interesting and contributes to our knowledge about the hydrometeorological factors of flash flood events, but the presentation is poor. The manuscript structure needs a lot of work; there is no section explicitly for results and discussion. These make the manuscript hard to follow. The study contribution and research gap that aims to be addressed are also unclear. I have recommended a few edits and comments in the PDF.

The authors would like to express their gratitude for reviewing this study and providing exhaustive and constructive comments to enhance the content of the revised manuscript. The titles of Sections 4 and 5, along with their respective subsections, have been made more descriptive of the presented results. The changes are believed to improve the readability. In addition, new discussion and conclusion sections have been included in the revised version of the paper.

Specific comments

Here are additional comments:

Overall, the writing is OK but some improvements should be addressed.

The illustrations need major improvements.

Abstract should be revised to provide information about methods and results. Also, please clarify the unique aspects of this study.

The abstract has been rewritten and the illustrations have been improved to address the concerns raised by the reviewer. The revised abstract now offers insights into the methods employed, present key results and highlights the novel aspects of this study. The new abstract reads:

"On 22 October 2019, the Francolí river basin in Catalonia, north-eastern Spain, experienced a heavy precipitation event that resulted in a catastrophic flash flood, causing six fatalities. This study investigates the hydrometeorological factors that concurred in the unfolding of this event using the high-resolution TRAM mesoscale model, radar-derived precipitation estimates, post-flood field and gauge observations, and the KLEM hydrological model. Results reveal that a persistent south-easterly airflow brought low level moisture and established convective instability in the region, while local orography was instrumental to trigger deep moist convection. A convective train promoted intense, copious and prolonged precipitation over the north-western catchment headwaters. Basin response was significantly modulated by the very dry initial soil moisture conditions. After the long-lasting rainfall, an acute burst of precipitation resulted in extreme flash flooding. Fast and abrupt increases in streamflow leave limited time for the effective implementation of protective measures. This study also

poses special attention on the social dimension by examining the relationship between catchment dynamics and warning response times and by quantifying human behaviour during the course of the flash flood. Few studies comprehensively address both the physical and human dimensions and their interrelations during catastrophic flash flooding. By examining the alignment among all these factors, this research takes a step forward towards filling this gap in knowledge. It also offers insights into the effectiveness of existing social protocols in meeting the requirements of the population at risk and identifies potential areas for improving preparedness for similar natural hazards in the future."

Introduction: Please explicitly discuss the unique aspects and novelty of this paper.

Currently introduction contain some information about flash flood in the case study, but the definition of flash flood is missing. In addition, some examples have been mentioned for small watersheds, but the case study is not as small size as these examples. How is the flash flood dynamic in your case study similar to these cases? Are there other types of floods in these areas?

In some paragraphs of introduction section, several references are presented at the end of a paragraph, but these need to be specifically cited throughout the paragraph.

The introduction has been thoroughly revised in the updated manuscript to address all concerns raised by the reviewer

The term "social response" is too broad and should be more specific. Do you mean management actions?

Effectively, the term "social response" covers a broad spectrum of social actions, encompassing from the warning procedure to responses at individual, group and organizational levels. The warning procedure itself involves several actions such as, monitoring, forecasting strategies, and the planning of management measures. In this study, the authors opted for the more generic term "social response" because the manuscript not only evaluates management activities but also describes human responses and citizen perceptions during the flash flood.

The revised version breaks down the extensive social response into two main components, according to Creutin et al. (2009): management activities and human responses. The former includes three different types of actions: information, organization, and protection. The latter encompasses human responses within three groups: Individual, communal and institutional. Consequently, the title of new section 5.2.1 now incorporates the more precise terms "risk management" and "human response". In addition, a more quantitative analysis has been conducted based on information gathered during the FLOOD-

UP FRANCOLÍ citizen campaign in the Conca de Barberà council. New Table 7 and Figure 12 provide a more detailed account of the types of actions and the timeline followed during the course of the flash flood.

N.	Action	Individual	Community	Institutional
	Warnings by the Catalan Meteorological Service (SMC)			X
	Activation of the INUNCAT plan			X
1	Searching and following information about the event			X
2	Activation of the Municipal Emergency Plans			X
3	Meeting of the INUNCAT Technical Committee at the headquarters of the Ministry of Home Affairs			X
4	Some actions: return home, move to an upper floor,... (Power cuts)	X		
5	First actions of firefighters			X
6	Civil Protection asks the population to stay at home and issues self-protection advice			X
7	The Espluga de Francolí city council ask residents not to go to the affected areas			X
8	Recovery and cleaning tasks. Damages evaluation	X		
9	Collaboration in cleaning tasks, help in recovery of wine cellar bottles, etc.		X	
10	Cleaning tasks and search for missing people. Activation of the Forest Defence Groups (ADF) until October 26			X
11	Activation by the winegrowers' association of a campaign of solidarity through the sale of recovered wine bottles. On October 25, the citizen platform "Riuada Solidària" formed		X	
12	Constitution of Municipal Emergence Command Centre at the Espluga de Francolí			X
13	First visit of the President of the Government of Catalonia to the affected areas			X

Table 7 Types of actions and classification during the course of the flash flood in the Conca de Barberà county from 21 October 2019 08 UTC to 24 October 2019 00 UTC. The colour criteria are the same as those shown in the Figure 12.

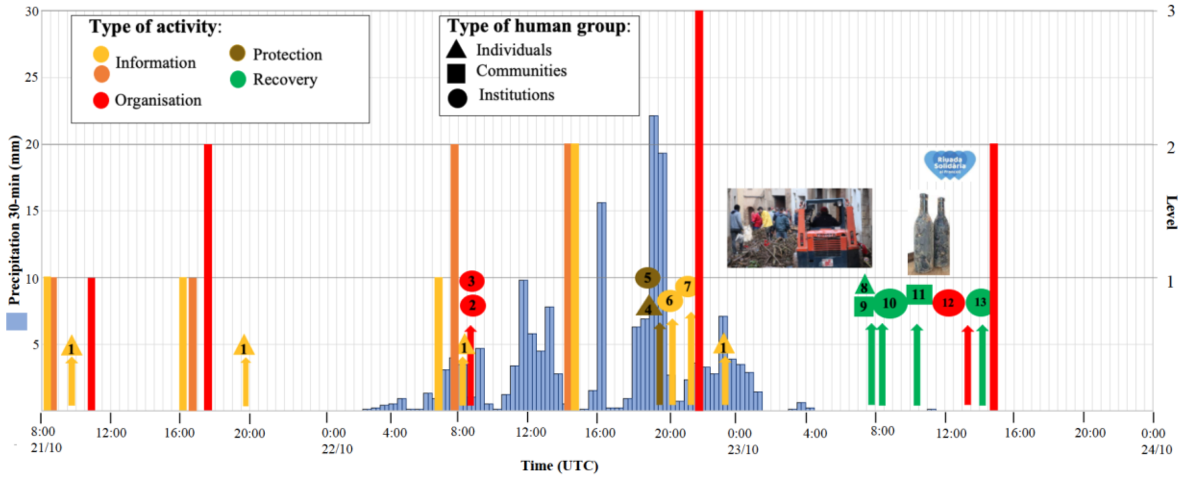

Figure 12: Timeline of warnings issued by the Catalan Meteorological Service (SMC) in the Conca de Barberà county from 21 October 08 UTC to 24 October 00 UTC. Light and dark orange bars denote accumulated precipitation and rainfall rate, respectively. On the right vertical axis, levels 1, 2, and 3 indicate moderate, high, and very high meteorological risk assessments by the SMC for these bars. The progression of activation phases in the INUNCAT plan is illustrated by the red bars. In this case, levels 1, 2 and 3 on the right vertical axis correspond to the pre-alert, alert and emergency stages, respectively. Social actions are also indicated, with colour representing management activities, and shape indicating human responses. The associated numbers align with specific actions detailed in Table 7. Additionally, the background vertical bars in blue showcase the evolution of 30-min rainfall accumulations in Espluga de Francolí.

Case study section needs to discuss the watershed characteristics such as climate, annual precipitation, land cover distribution, topography and other factors related to flash floods.

The authors value the reviewer's comment and acknowledge the importance of better contextualizing the Francolí basin in terms of climate, annual precipitation, etc. Section 2.1.1 of the revised manuscript contains all this pertinent information.

***Please add a schematic view of your methodology as a figure at the beginning of Section 3. Sections 4 and 5 should be renamed as results and discussion*.**

As mentioned earlier, sections 4 and 5 have been renamed to provide clearer indications of the results presented within them. Additionally, a new discussion section has been incorporated in the revised version of the paper.

The control numerical simulation in Section 4.1 should be discussed in detail.

The authors respectfully disagree with the reviewer's comment. This study allocates two pages to discuss the meteorological control simulation. The primary goal of Section 4.1 is to highlight the key physical factors contributing to the development of the HPE. Providing additional details in this section would adversely affect readability and unnecessarily lengthen the study.

The models (TRAM, QPEs KLEM etc.) have inconsistent spatial resolutions. How did you handle this inconsistency?

The TRAM meteorological model is devoted to analyse the physical factors at meso- and synoptic-scales that contributed to this catastrophic flash flooding, with a spatial resolution is 3 km. The quantitative precipitation estimates (QPEs) derived from radar observations have a finer spatial resolution of 1 km and a temporal resolution of 10 minutes. These increased spatio-temporal scales enable a more thorough analysis of the key features of the heavy precipitation event that led to the flash flood in the Francolí basin.

The KLEM model operates at a spatial resolution of 25 meters to examine in detail the interaction between the high spatial and temporal variability in rainfall fields and the geomorphological and hydrological factors influencing basin response to heavy rainfall. In our perspective, there is no inconsistency, as different tools and procedures are employed to investigate distinct physical factors and mechanisms at varying spatial and temporal scales. This chain of models and procedures allows for the description of the cascading succession of physical mechanisms and their interrelations that resulted in this event, spanning from the meso- to the micro-scale.

My understanding is that the automatic gauges record data at sub-daily timescale but the number of these stations are limited, particularly for streamflow. How did you use daily data for a rapid catastrophic flash flood event? What limitations and uncertainties exist here?

There are 59 automatic rain-gauges located inside or very close to the Francolí basin, recording precipitation at temporal resolutions between 5 and 10 minutes and belonging to different regional or state institutions. In addition, the Catalan Water Agency deployed two automatic stream gauges along the Francolí river in Montblanc and Tarragona, two cities crossed by the river. Montblanc encloses the upper Francolí catchment with a drainage area of 339.9 km^2, while Tarragona measures streamflow near the watershed outlet, covering a basin area of 809.1 km^2. Unfortunately, the flood bore destroyed the stream gauge in Montblanc, resulting in the unavailability of a complete time series of data for the 22 October 2019 episode

Daily data are only observed from an independent network of pluviometers deployed by the Spanish Agency of Meteorology. These rainfall data have been used in this work solely for conducting a safety validation test of the QPEs. However, it is acknowledged that the automatic stream gauge is limited to just a river section for this event. Consequently, the hydrological model simulates the entire catchment and is calibrated against streamflow observations at the catchment outlet. Thus, evaluating the performance of the hydrological model at the basin outlet may not be entirely representative of the drainage areas that suffered the most catastrophic impacts.

However, Martín-Vide et al. (2023) conducted a comprehensive post-event field campaign, providing peak flood estimates and timing at various river sections in the upper Francolí catchment. To address the concerns raised by the reviewer, outputs from the hydrological simulation have been compared against these estimates. In this regard, the first two paragraphs in Section 4.3.2 have been modified to address the reviewer's concerns, and a new Table 4 has been added to explicitly compare model results with estimates.

River section	Area (km^2)	Peak discharge (m^3s^{-1})	Simulated peak discharge (m^3s^{-1})	Time of peak discharge (UTC)	Simulated time of peak discharge (UTC)
1-Viern (headwaters)	7.1	40–110*	95.4	–	–
2-Viern	9.5	60–120*	133.7	–	–
3-Milans	26.6	115–360*	286.2	19:30*	19:30
4-Sec	38.8	90-110*	181.4	–	–
5-Espluga	97.3	500–775*	550.1	19:50–20:15*	20:00
6-Montblanc	339.9	610–790*	630.1	20:20–20:45*	20:30
7-Riba	449.0	740–870*	758.6	21:00–21:30*	21:00
8-Tarragona	809.1	*871.0*	798.8	*22:30*	22:40

Table 4 Comparison among data obtained from the hydrological control simulation and estimates based on field observations and hydraulic modelling conducted by Martín-Vide et al. (2023). Estimates are marked with an asterisk (*). Observations have been included for completeness (in italics). Refer to Fig. 1 for locations.

These modelling results confirm the quality of the control simulation not only in reproducing the overall basin response, but also in capturing the hydrological response at smaller drainage areas. In addition, Figure 11 has been modified to include the simulated times at all the aforementioned river sections, allowing for a graphical comparison of model outputs with estimates and observations. To enhance the graphical interpretation, the names of the hydrometric sections where estimates and observations are available have been included.

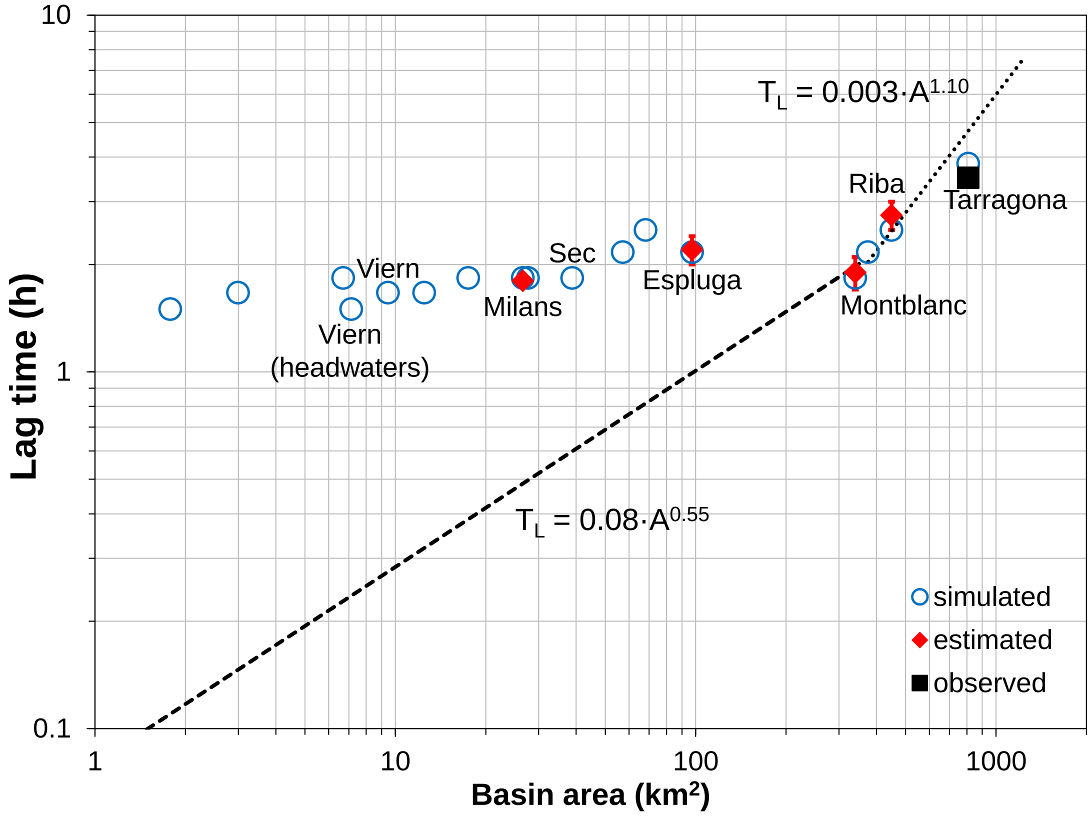

Figure 11. Lag time versus drainage area for the 22 October 2019 flash flood event in the Francolí basin. Uncertainties in the estimated lag times (i.e. derived from the post-event field campaign) are shown as vertical bars. The "observed" lag-time label refers that it has been derived from the stream-gauge measurements at the Tarragona hydrometric section. Also shown the power-law relationships after Marchi et al. (2010). Refer to Table 1 and Fig. 1 for the names and locations of the surveyed river sections.

Section 4.3.3: Add a table and show the sensitivity scenarios.

New Table 5 and Figure 10 present the results obtained from conducting the sensitivity tests

Experiment	**Total rainfall (mm)**	**Flow volume**			**Flow peak**			**Time of peak discharge (UTC)**
		Observed (mm)	**KLEM (mm)**	**Error (%)**	**Observed (m^3s^{-1})**	**KLEM (m^3s^{-1})**	**Error (%)**	**KLEM**
Control	129.9	12.4	15.1	21.5	871.0	798.8	-10.9	22:40
Sensitivity test 1	95.9	12.4	23.8	91.7	871.0	1171.5	32.0	22:40
Sensitivity test 2	95.9	12.4	6.8	-45.4	871.0	315.7	-64.0	22:40
Sensitivity test 3	129.9	12.4	13.4	7.6	871.0	545.5	-37.5	22:50

Table 5 Observed and radar-driven simulated flow volume and peak discharge of Francolí River at the Tarragona hydrometric section for the control and test experiments. Negative values in relative errors denote model underestimation. Total rainfall amount is radar-derived and is expressed as the areal-averaged basin value. Time of peak discharge refers on 22 October 2019.

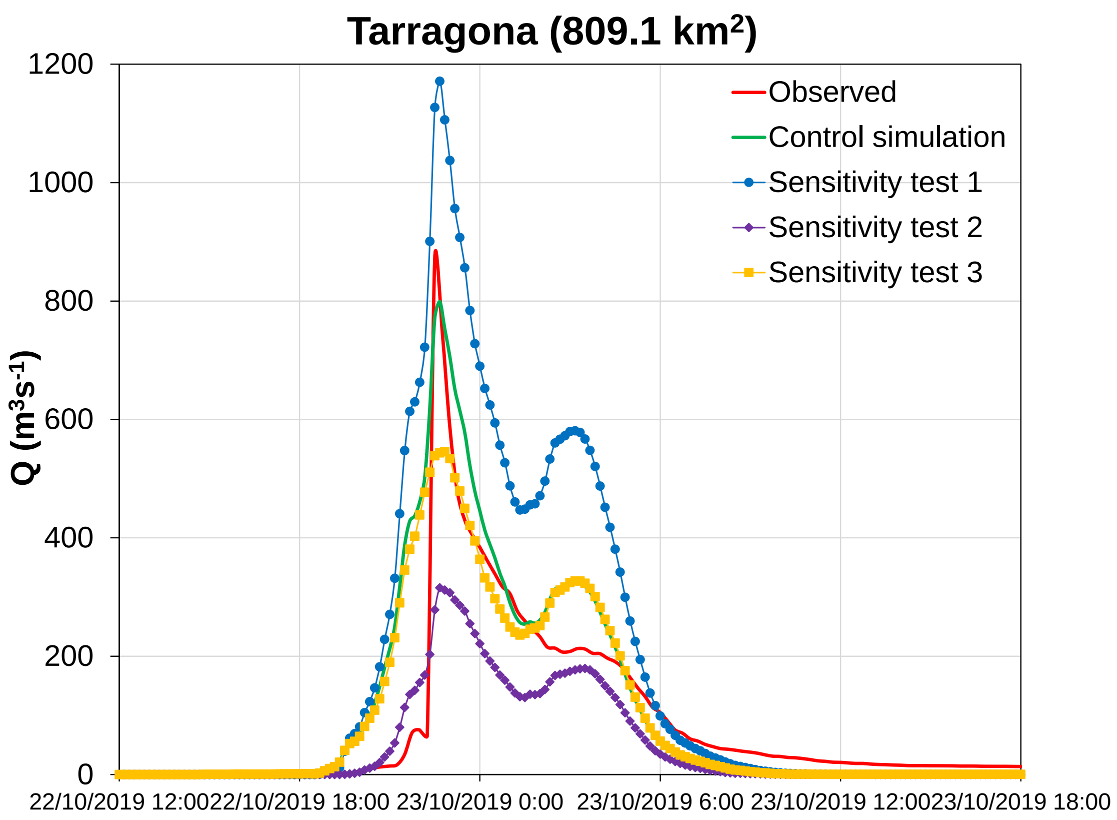

Figure 10. Observed and radar-driven runoff simulations for the control and sensitivity test experiments and the 22 October flash flood at the Tarragona flow-gauge in the Francolí basin.

The initial soil moisture is determined based on the antecedent precipitation, as a standard proxy. Why not using global data like ERA5 and CCI that directly present the soil moisture?

This is a very interesting point. Numerous hydrological models are updated with initial soil moisture estimates derived from ERA5 or the soil moisture project from the ESA Climate Change Initiative. Frequently used for real-time hydrological forecasting, these models build on more complex infiltration equations, often resolving the water balance equation.

In contrast, the soil conservation service curve number relies on antecedent precipitation to evaluate initial moisture conditions, and this approach has been adopted in this study. However, it remains as a future task to incorporate more complex infiltration schemes into KLEM, as well as to start the model by assimilating soil moisture fields coming from these analyses.

I suggest using CN as a commonly used abbreviation for curve number.

Done.

Hydrological model calibration needs details and clarifications. Why CNs were kept invariant? Why did you use an initial abstraction ratio of 0.35 (lambda)? The sensitivity analyses should be extended by evaluating other variables like lambda. What fit metrics (e.g., NSE and PBIAS) were used and how the model performance was judged based on them? Any validation effort on the hydrological model?

The authors appreciate the reviewer's comments and recognize the need for additional details and clarifications regarding the hydrological model calibration. These details and clarifications are now included in the revised Section 4.3.2.

The calibration efforts are focused on reproducing peak discharge, time-to-peak, and runoff volume at the Tarragona hydrometric section where observations are available. In this study, curve numbers represent an input data as they are derived from field measurements. These are set to represent dry antecedent moisture conditions, remaining invariant. However, the initial abstraction ratio is considered a calibration parameter in the infiltration method due to significant soil retention capabilities. The presence of large storativities is associated with exceptionally low initial soil moisture content and the recharge of deep aquifers through infiltration, percolation, and transmission losses along the river beds. This approach allows to correctly simulate the observed water balance.

During the calibration process, the performance of the hydrological model is evaluated against the observed hydrograph using different objective functions, such as the NSE and relative errors in peak discharge and total direct runoff volume. The calibrated value of the initial abstraction ratio is determined to minimize errors in terms of peak discharge and runoff volume, ensuring that KLEM adequately reproduces the overall basin response.

Since calibration is based on the observed flood hydrograph at the catchment outlet, 16 additional river sections have been included in KLEM to explore and validate hydrological response at smaller drainage areas. These 17 hydrometric sections include the 7 river locations surveyed during the post-event field campaign (Table 1). The comparison of numerical simulation results in terms of peak flows and times-to-peak against post-event field estimates serves as a verification test, confirming the quality of the control simulation in reproducing the hydrological response at smaller drainage areas. These information is included in new Table 4 and Figure 11 in the revised version of the manuscript.

Some of the error values on Table 4 are high (<65%). How would you interpret these and the efficiency of your hydrological model?

The hydrological sensitivity tests aim to assess the influence of three specific factors on the development of the flash flood event. These are the roles of the: (i) initial soil moisture content; (ii) early rainfall period preceding to the torrential precipitation rates and amounts, and; (iii) variability of the heaviest rainfall period. Once the overall basin response is adequately reproduced by the control hydrological simulation, the sensitivity tests involve varying one specific ingredient at a time and examining its impact on the basin response for the study case. It is essential to maintain the remaining factors invariant during this procedure to ensure consistency with the hydrological control simulation.

The results of the different sensitivity tests highlight the relative importance of each factor in modulating the overall basin response, quantified by the errors in reproducing the control simulation. The highest deviation in simulated runoff volume occurs when considering normal antecedent conditions in sensitivity test 1, even with a smaller total rainfall. In terms of peak discharge, sensitivity test 2 has the most significant impact by neglecting the effect of the early rainfall period on the overall hydrological response. The variability in rainfall during the heaviest precipitation period plays a crucial role in exacerbating peak discharge.

Can your results be generalized to other flash flood events in the study area or flash events beyond the study area? Please discuss.

The new discussion section compares the results of the present work with previous findings and provides information about elements of the study that are of interest for a wider audience. The conclusions and the abstract have been revised accordingly.

Sources of uncertainty and how they can affect your results should be discussed.

The authors appreciate the reviewer’s comment. In the revised version of the manuscript, the sources of uncertainty related to the quantitative precipitation estimates and the hydrological model simulation have been properly discussed. For further details, the authors refer to rewritten section 4.3.2 and the associated results.

Study limitations and potential areas for future research should be discussed.

Limitations of the study and potential areas for future research have been identified and discussed throughout the revised sections. Specifically, it is important to acknowledge limitations associated with the hydrological model performance at different river sections, and uncertainties in reproducing fine features of the highly variable precipitation pattern due to assumptions made in different correction procedures when estimating rainfall from radar observations. In the revised conclusions, potential avenues for future research have also been discussed.

Table 1: What does "hydrometric section" mean? Please clarify the duration of total rainfall.

In this study, the term "hydrometric section" was initially used as a synonym for the term "river section". To prevent possible confusion, the former term has been replaced for the latter in Table 1 and throughout the manuscript. The duration of the total rainfall is now explicitly included in the first paragraph of section 4.2

Please remove "Color code" column from Table 5.

The authors respectfully disagree with the reviewer's comment. The colour code is aligned with the risk assessment scale used by the Catalan Meteorological Service. This colour scale grading is fundamental for the understanding of section 5.2.1. and Figures 12 and 13. In particular, Figure 13 illustrates the temporal progression of the warning process based on risk assessment and associated colour codes.

Figures 1-3 should be improved by considering the size, alignment etc.

Figures 2 and 3 have been improved in response to the reviewer's concerns. Nevertheless, the authors maintain that Figure 1 adheres to the standard configuration found in scientific literature when introducing a study region. Typically, a top-left figure showcases the main features of the region, while a central figure illustrates the main features of the catchment of interest.

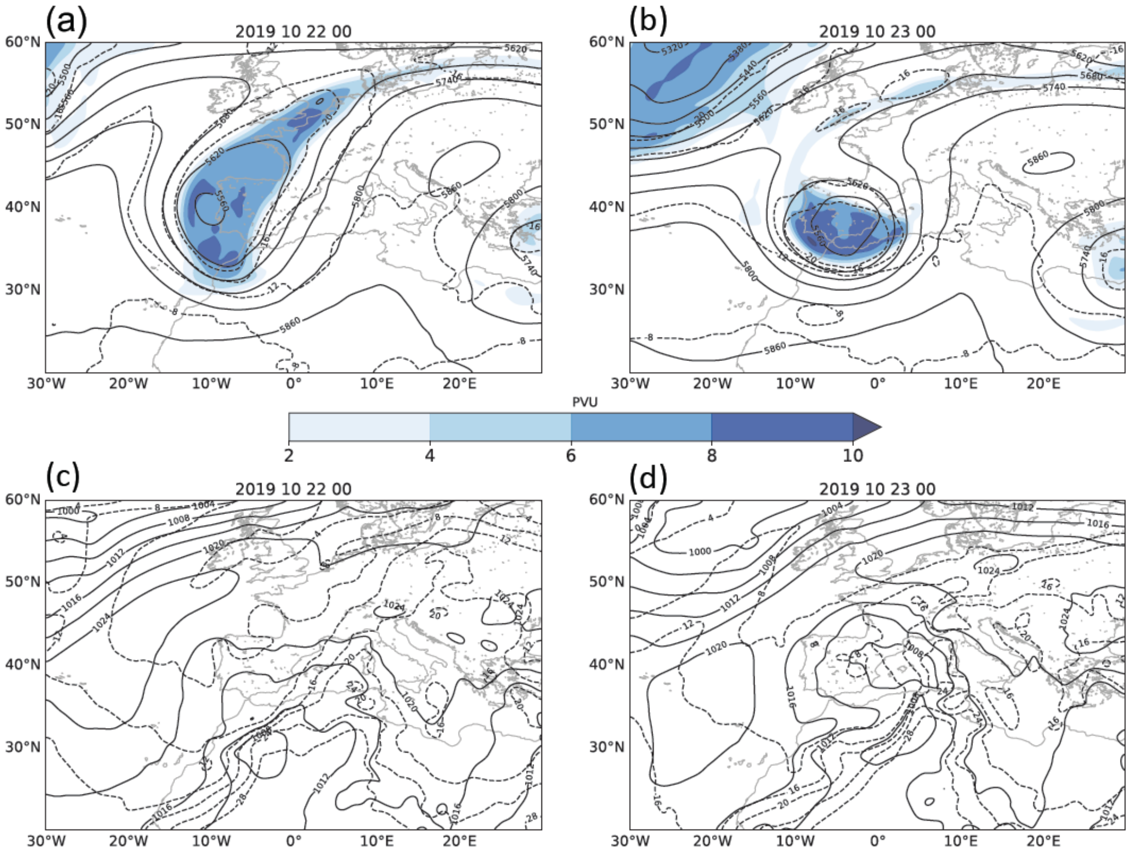

Figure 2. ECMWF analyses for geopotential height (solid lines, in geopotential meters), temperature (dashed, in °C) at 500 hPa, and 250-hPa potential vorticity (PVU, shaded in blue) on: (a) 22 and, (b) 23 October 2019 00:00 UTC. ECMWF analysis for mean sea level pressure (solid lines, hPa) and temperature at 850 hPa (dashed, in °C) on: (c) 22 and, (d) 23 October 2019 00:00 UTC.

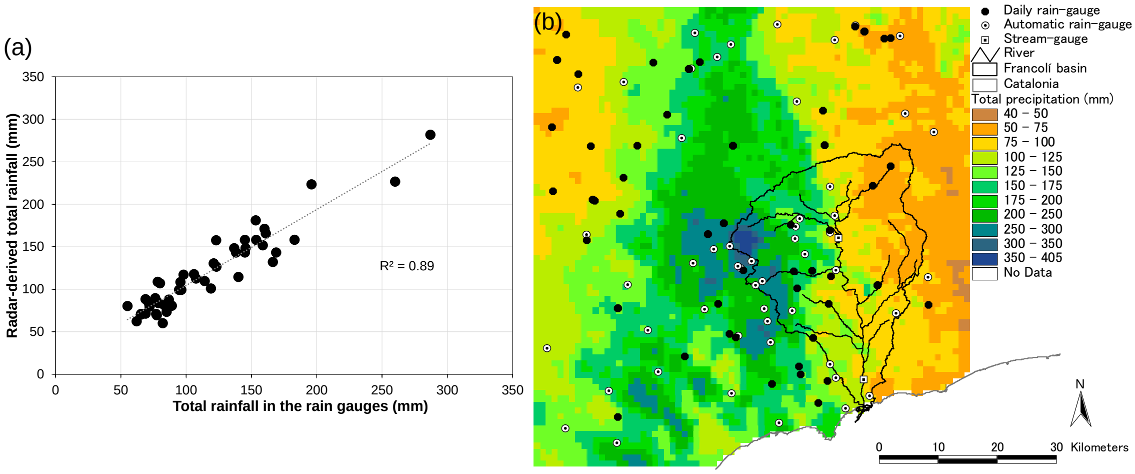

Figure 3. (a) Scatterplot of the 48 h radar-derived rainfall estimates against observed accumulations by the daily AEMET rain-gauge network over the selected region. (b) Spatial distribution of the 48 h accumulated radar-estimated precipitation from 22 to 24 October 2019 at 00:00 UTC. The Francolí river catchment is highlighted with a thin black line. White squares stand for the automatic stream-gauges. White dots show the position of the automatic rain-gauges. Daily pluviometric stations are denoted by black dots.

Section 4.1: Mesoscale processes and role of orography is ambiguous. Please clarify how the simulation works and how Figures 4-5 were produced?

Figures 4-5 can be merged.

The authors refer to Romero (2023) for all the technical details of the control and perturbed simulations of this study. Eliminating orography from a mesoscale simulation is a standard procedure in meteorological research that allows studying the role of this factor. For further technical details, see the reference Romero (2011).

The authors kindly disagree about merging figures 4-5 in just one panel. The authors believe that it would negatively affect the structure and readability of the study.

Romero, R., 2023: TRAM: A new nonhydrostatic fully compressible numerical model suited for all kinds of regional atmospheric predictions. Quart. J. R. Meteorol. Soc., DOI 10.1002/qj.4639.

Romero, R., 2011: Application of factor separation to heavy rainfall and cyclogenesis: Mediterranean examples. Chapter 7 in Factor separation in the

Atmosphere: Applications and Future Prospects, ed. Pinhas Alpert and Tatiana Sholokhman, Cambridge University Press, 87-119.

Figure 8d: What is the main massage of temporal relationship between drainage area and precipitation? Why is the expectation that these two should have a relationship?

Figure 8d explores the basin areas impacted by 10-min rainfall rates exceeding 20 mmh^{-1} and 50 mmh^{-1}, establishing a link between the drainage areas affected by these precipitation rates and their durations. Undoubtedly, these characteristics in rainfall fields are closely connected to runoff generation and subsequent flash flooding. Therefore, it is regarded as an additional metric for describing the spatial and temporal organization in rainfall that led to this flash flooding. This idea has been elaborated further in the revised section 4.2

Figure 11 is odd. Why do you have "estimated" uncertainties only on a few data points? This should be for all simulated values. Why do we have only one "observed" value? How can lag time be even observed?

As mentioned earlier, Figure 11 has undergone modifications to enhance clarity. For ease of graphical interpretation, the names of the river sections with available estimates and observations have been incorporated. In addition, the simulated lag times for the 17 river sections serving as control points in the hydrological simulation have been included for comparison. The uncertainties associated with the estimated lag times stem from the ranges of time of peak discharge estimated at the different hydrometric sections, as shown in Table 1, and derived from the post-event field campaign by Martín-Vide et al. (2023). Times of peak discharge were estimated through interviews with eyewitnesses. Their associated uncertainties are shown as vertical bars

Acknowledging that the lag time cannot be directly observed, the authors recognize that the use of the term "observed" might be misleading in this context. In this study, the lag time is computed as the temporal difference between the centre of mass of the rainfall hyetograph (i.e. the rainfall centroid) and the timing of peak discharge. Therefore, the "observed" lag-time is derived from the automatic stream-gauge data available at the catchment outlet and the rainfall fields obtained from the weather radar, justifying its label as observed. An explanatory sentence has been added to the caption of Figure 11 to prevent confusion.

Please summarize the key findings of your study (e.g., as bullets) in the Conclusions section.

The key findings of the study are now incorporated into the new discussion section. This section also facilitates a comparison of the results with previous findings and offers insights into elements that may be of interest to a broader audience. The conclusions section has been accordingly revised.

Please italicize all parameters and coefficients throughout the text.

Please spell out all abbreviations in the figures, tables and headings; these need to stand alone.

Done

I hope the authors find these comments useful in their research. If the authors decide to submit a revision, both sets of my comments, including the above and in the PDF, have to be addressed.

Certainly. The authors would like to express their gratitude once again to the reviewer for his/her valuable comments, which have improved the revised version of the manuscript.