Reply document

This document includes our reply to the comments of the first Reviewer. In the following, the original comments are shown in Italics and our reply in blue-colored text.

Overview

This manuscript constitutes a case study for cyclone/medicane "Daniel" which, at different stages in its lifecycle, delivered devastating weather and impacts in parts of both Greece and Libya, in early September 2023. Four aspects are examined, as listed at the start of the title.

The most novel and publication-worthy features of the study are the moisture source analyses, for both the Greek and Libyan floods, the sea wave analysis and the use of cyclone and precipitation objects.

The remainder of the study does not add much to previous published literature on this case (notably Hewson et al, 2024) which is cited and latterly Couto et al (2024), which focusses on broadscale aspects. Admittedly the Couto paper, available here: https://www.mdpi.com/2073-4433/15/10/1205, has only just appeared so was probably unavailable to the authors pre-submission. In some respects these two papers go much further than the one under review, particularly with regard broadscale patterns, local details of the extreme weather and considerations with regard to high impact warnings. Given that a standard requirement for publication, in any journal, is that one adds to previously established knowledge (rather than detracting from it) it is clear, in the opinion of this reviewer, that a very substantial reworking of the paper's content would be required for acceptance.

There is a clear reluctance to include observations in this paper – notably rainfall measurements. If numerical model analyses were perfect, this might be acceptable, but given that they are not, particularly with regard to rainfall, which is the impact centrepiece of this study, this is a major omission.

Furthermore, reviewing the paper has been a frustrating process due to the many inconsistencies in different segments of the text, inconsistencies between what the figures show and what the text says, simple errors, poorly explained figures, and unsubstantiated conclusions. Rather than go through absolutely everything which is of concern, which would take a very long time and replicate the checks the authors themselves should have carried out before submission, I will instead go through the figures, which are the bedrock of the paper, and highlight the key issues via those.

We thank the reviewer for taking the time to comment on our manuscript in detail. We answered all the comments and suggestions below.

Main points

Figure 1a: Some of the red spots are missing (assuming the time interval is 6 hours, which should anyway be stated); some of the labels show the wrong time, and the size-for-mslp legend is hard to interpret. At one stage in the text it is stated that the cyclone was intense early in its lifecycle – by normal measures 1004mb is not intense – and indeed elsewhere in the text this statement is contradicted. Somewhere else in the text it says that the cyclone intensified on 6th and 7th, as can be seen on this Figure; this was not the case and nor does the figure show it, even allowing for mislabelling errors. Somewhere else the text says the minimum pressure of 997mb was reached on 9th September. This is not correct either, nor does the figure show this.

Thank you for this comment. Indeed, the original plot contained mistakes related to dates and time. We have corrected these, and added the minimum SLP values in the labels next to the red dots to help the reader follow the evolution of the cyclone's SLP. In addition, we corrected all the wrong statements mentioned in the revised version of the manuscript according to the reviewer's suggestion.

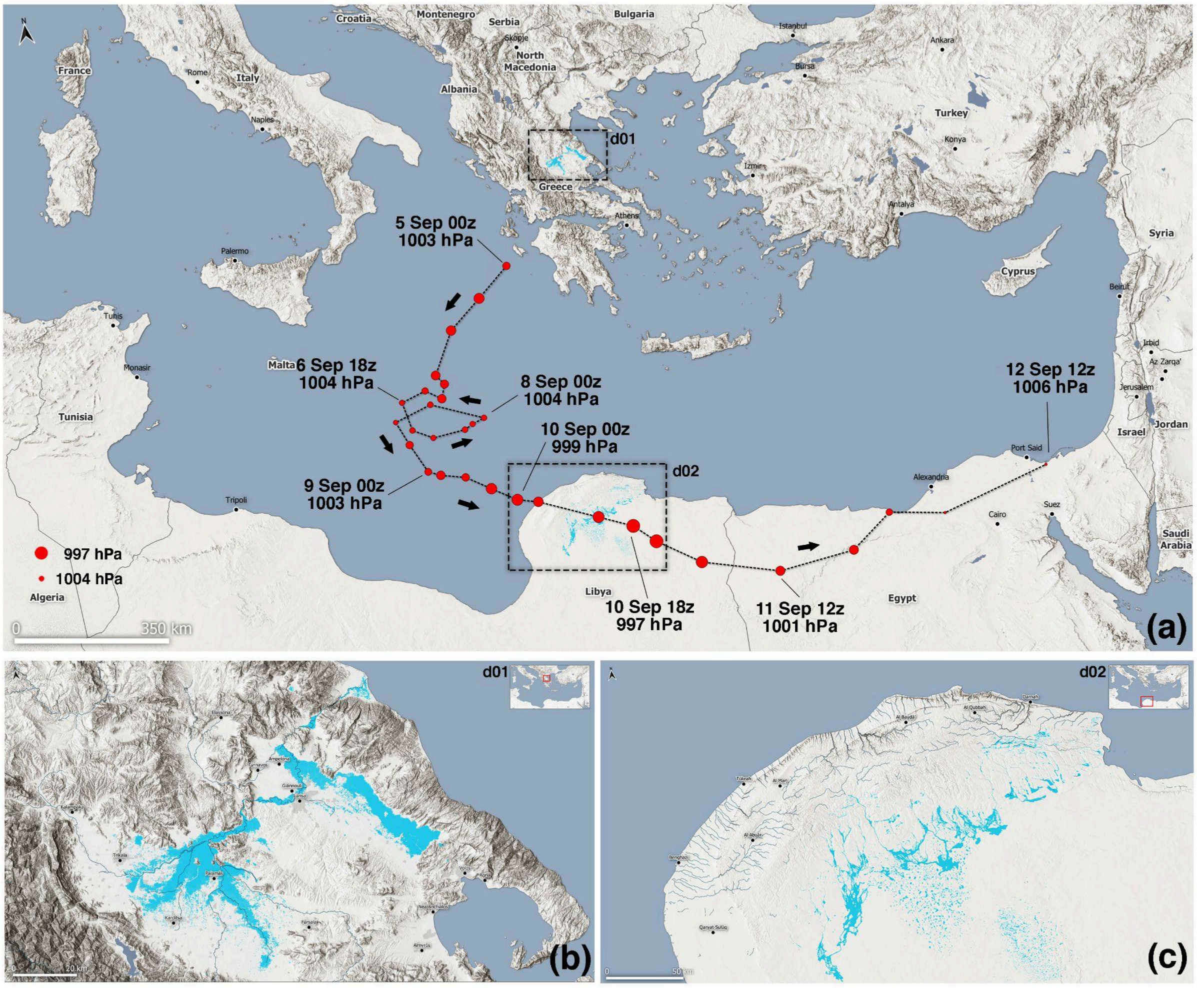

Figure 1 (a) Track of Storm Daniel at six-hour intervals based on ECMWF analysis, where the size of red dots is proportional to cyclone depth in terms of minimum mean sea level pressure. Flooded areas are shown in cyan and blue tones (acquired by one of the Copernicus Sentinel-2 satellites on 10 and 12 September 2023). Panels **(b)** and **(c)** zoom over central Greece and Libya (square boxes in panel **a**).

Figure 2a: The rainfall area is too small to see properly (even when zooming on the pdf), as are the wind barbs. Where does the data come from – ERA5 or ECMWF analyses/short range forecasts? The latter is much higher resolution (9km versus 31km) and so would likely show much more useful rainfall detail (if one could see it). I can quite imagine that the PV of 2 PVU, in the caption, is actually 2. In the text it is stated that "there is a high wind speed pattern aligned with the PV streamers' orientation". I do not know what aligned with means here. The text cites 750mm rainfall in 24 hours on 5th – why not say where! This actually occurred east of Volos, at 3 close sites (see Table 2 in Dimitriou et al, 2024), and is where the model rainfall pattern, when zoomed in maximally, shows about 110mm. So the reference to a 50% shortfall in the model should be 85%. The authors actually refer directly to purple colours (which are seen elsewhere), which represent 200mm, so quite clearly even those don't represent 50% of 750mm. This all then makes the statement that (ECMWF) models provided good guidance somewhat incorrect.

Thank you for the careful estimations and the suggestions.

The IFS dataset used in this study has 9 km spatial resolution. Two additional panels have been included to increase clarity in the areas affected by heavy precipitation.

We have added some additional information about the differences between the IFS and observation maximum 24-hr accumulated precipitation:

"Notably these peak values are underestimated by about 40% in the ECMWF analysis (max IFS 24-h accumulated rainfall equal to 434 mm on 6 September 2023 00 UTC)."

Figure 2b: The model rainfall, up to 00UTC on 11th, is not much in the Derna catchment (location unfortunately not shown but included in Hewson et al, 2024) despite the fact that the dams broke an hour or two later. Again this is very concerning with regard to model validation and impact predictability. These aspects are not discussed at all. Text says the PV streamer was "much weaker" at this time. I do not know what this means. It gives the wrong impression too as this streamer is, on the contrary, probably a marker for a substantial lobe of upper level forcing that helped trigger the main intensification of Daniel.

We have added the following sentence:

"It is worth mentioning that simulated 24-hr total accumulated precipitation on 11 September 2025 in Libya, up to 382 mm, was not located within the Derna catchment, as it has been discussed in Hewson et al. (2024), which was the most impacted area."

Regarding the role of the PV streamer in cyclone intensification, we have revised the text according to the reviewer's suggestion:

"A comparison of Figs. 2a and 2b shows that, at the time of maturity, the area covered by at least 2 PVUs at 300 hPa is significantly smaller than during cyclogenesis. Nevertheless, Fig. 2b shows that the 2-PVU patch is collocated with the cyclone center, advected from the west. Hewson et al. (2024) proposed that this collocation is responsible for the cyclone's intensification just before landfall. In fact, the intensification of a Mediterranean cyclone due to the synergy of upper-level baroclinic forcing and deep convection is a common characteristic of intense Mediterranean cyclones, including medicanes (Flaounas et al., 2021). A previous case of a medicane intensifying due to the collocation of a PV streamer with the cyclone center was documented by Chaboureau et al. (2012). This phenomenon reflects, on the one hand, the anomalous nature of this medicane (as medicanes generally intensify over the sea and weaken inland), on the other hand, the critical role of upper-level features in the evolution of Mediterranean cyclones."

Figure 3: It was nice to see the moisture sources, even if the propensity to uptake most moisture just upwind of the heaviest rain for Daniel, in strong wind areas, was not hugely surprising. The uptake in the composited cases is much harder to second guess, so this is a nice result. I am not sure why 10 day trajectories were used. That seems quite long? Also I am not sure what "30km grid" means, in the main text. Coastlines and sea areas are impossible to see on the figure in this form, so that aspect has to be improved. The main discussion of the moisture uptake elects to ignore any sources over land, yet clearly they are relevant – more so than the Atlantic Ocean which is mentioned. In the conclusions uptake over landmasses is mentioned for the first time.

Thank you for these questions that allow us to better clarify this important part of the paper.

- We use 10-day backward trajectories because the explained fraction of the moisture sources decreases strongly with shorter trajectories (Fig. A1). With 10-day backward trajectories the sources of around 90% of the precipitation can be explained by the moisture source diagnostic. For Libya on 11 Sep, the length of the trajectories also affects the land fraction of the moisture sources. For 10-day trajectories, one third of the moisture originates from land areas, while for 5-day trajectories, this fraction decreases to one quarter.

- Concerning the 30 km grid, it refers to the spatial position of the trajectory starting locations. The trajectory starting locations are positioned in a regular meridional-zonal grid with 30 km grid spacing. We adjusted the text to make this clearer:

  "Ten-day air parcel backward trajectories are calculated every 20 hPa between 1000 and 300 hPa from starting locations on a regular latitude-longitude grid with a 30 km grid spacing within boxes over Greece and Libya..."

- Fig. 3 has been adjusted following suggestions from both reviewers.
- Thank you for pointing out the missing information on regional attribution of moisture sources. We've calculated land/ocean fraction and added these values to Fig. 3. Further, we now discuss the source regions in more detail in the main discussion and also compare our source regions to a recently published study (Argüeso et al., 2024), which calculated moisture source regions for storm Daniel using an Eulerian moisture source diagnostic.

  We added the following discussion on regional moisture source attribution:

  "These source regions are in general agreement with a recent study (Argüeso et al., 2024), which investigated moisture sources of rainfall over Greece from 3 to 9 Sep 2023 using a Eulerian moisture source diagnostic. Our moisture source analysis shows larger contributions from land (54.7%) than in Argüeso et al. (2024) (27%). The Lagrangian method used in our study shows relatively large moisture contributions from north of the Black Sea because most of the air parcels arriving on 5 Sep 2023 descended and took up moisture in this region before moving southwestward along the western flank of the PV streamer. The differences in the land fraction between the two methods might originate from different periods used for the moisture source calculations, different handling of moisture uptakes above the boundary layer, a lower explained fraction of the total moisture sources (84%) with the Eulerian compared to the Lagrangian diagnostic (explained fraction of 90%), and general differences in Eulerian versus Lagrangian approaches. An ongoing comparison study of moisture source diagnostics is investigating differences in these methods in detail and will shed more light on disagreements between various moisture source diagnostics."

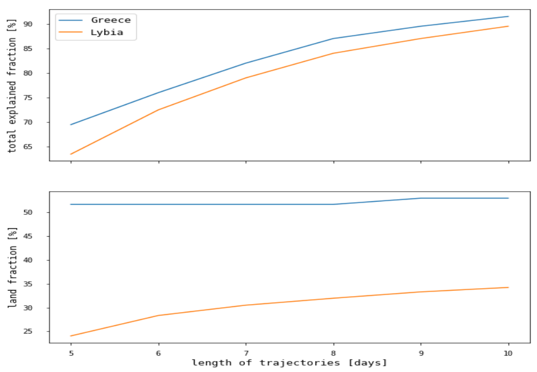

Figure A1: (Top) Explained fraction of the total precipitation during 5 Sep 2023 in Greece (blue line) and 10 Sep 2023 in the Derna region in Libya (orange line) for 5-, 6-, 7-, 8-, 9- and 10-days backward trajectories. (Bottom) Similar to the top figure but for the land fraction of the moisture sources.

References:

Argüeso, D., Marcos, M. & Amores, A. Storm Daniel fueled by anomalously high sea surface temperatures in the Mediterranean. *npj Clim Atmos Sci* 7, 307 (2024). https://doi.org/10.1038/s41612-024-00872-2

Figure 4: Although this looks initially quite convincing on closer inspection one sees that there is virtually no signal in (b) of a particularly high discharge near to where the heaviest rainfall was in Greece (its all time 24h record), east of Volos, nor in Derna in Libya, or its catchment. These aspects should have been extensively discussed. Maybe this relates to the rainfall errors on Figures 2a and 2b that I reference above, which were also not discussed.

We appreciate the reviewer's detailed observations in Figure 4 and acknowledge the importance of addressing the discrepancies between the discharge signals and the rainfall patterns highlighted in Figures 2 and 4. Upon revisiting the data and methodology, we provide the following explanation and revisions to the manuscript.

First, it is essential to clarify the fundamental difference in the temporal scope of Figures 2 and 4, which likely contributed to the perceived inconsistencies. Figure 2 represents 24-hour total accumulated precipitation for specific time frames during the storm: from 00 UTC on September 5 to 00 UTC on September 6, 2023 for Thessaly (west of Volos), and for September 11, 2023, at 00 UTC for Derna. These snapshots capture rainfall over single days and focus on localized phases of the event. In contrast, Figure 4 shows the maximum simulated peak river discharge for September 2023, integrating hydrological impacts over the entire lifecycle of Storm Daniel. This integration of effects means that discharge signals in Figure 4 reflect cumulative responses to rainfall over time rather than the specific short-term intensities depicted in Figure 2.

The absence of strong discharge signals in Figure 4(b) near the east of Volos and Derna can be attributed to several factors. First, the limitations of the GloFAS v4.0 model, particularly its spatial resolution and calibration scope, play a significant role in the observed discrepancies. The model operates at a spatial resolution of approximately 5 km (0.05°), sufficient for global-scale flood awareness but inadequate for resolving fine-scale hydrological processes in regions with complex topography and small catchments. For instance, the catchments east of Volos, including the wider Pelion area, are approximately 30 km^2, while the Derna basin spans around 575 km^2. In both cases, localized rainfall-runoff dynamics play a critical role. Moreover, Greece and Libya were not included in the GloFAS calibration dataset due to the limited availability of in-situ discharge measurements (see here). As a result, discharge predictions for these regions rely on generalized parameter regionalization rather than site-specific calibration, introducing further uncertainties.

Furthermore, inaccuracies in the rainfall inputs depicted in Figure 2 propagate into the river discharge simulations shown in Figure 4. For the Thessaly region and within the Peneus catchment, the maximum recorded 24-hour total accumulated precipitation from 5 to 6 September 00 UTC was 274 mm at Zappeio and 226mm at Neraida stations, as reported by Dimitriou et al. (2024, Table 2). However, accumulations of up to 750 mm, referenced in the discussion, correspond to stations outside the Peneus catchment, specifically over the Pelion area, east of Volos. This distinction is important as the GloFAS model simulates river discharge for the Peneus catchment, and the underrepresentation of precipitation within the catchment impacts the accuracy of discharge predictions. In the case of Derna, torrential rainfall of 150–240 mm was recorded in several cities, with Al-Bayda experiencing the highest daily total of 414.1 mm, as reported by the World Meteorological Organization (WMO). These extreme rainfall events, critical in triggering catastrophic flash floods and dam failures, were underrepresented in the GloFAS input data.

To address these issues, we revised the manuscript to highlight the temporal distinction between Figures 2 and 4. In the results section, we also discussed the limitations of the GloFAS v4.0 model, particularly its resolution and lack of calibration, which contributed to discrepancies. Additionally, we discussed the influence of rainfall input inaccuracies on the discharge signals and their implications for interpreting Figure 4. The revised results section reads as follows:

"The hydrological impacts of Storm Daniel were profound and unprecedented. Figure 4 compares the peak mean daily river discharge during Daniel with the historical records over three decades, integrating the cumulative hydrological impacts over the entire event. Figure 4a shows the spatial distribution of the maximum simulated peak discharge from January 1993 to August 2023 (i.e., before Daniel), demonstrating typical peak discharge patterns in the Eastern Mediterranean. On the other hand, Fig. 4b compares the event-wide mean daily peak discharge during September 2023, when Daniel occurred, against the historical peak discharges of the last 30 years in Fig. 4a. Results reveal an unprecedented magnitude of Daniel's impacts, with several areas experiencing discharges that exceeded the historical maximums by 300 to 500%. The darkest shades in Fig. 4b indicate the most heavily affected regions, where the river discharge during Daniel exceeded previous records by at least a factor of five, highlighting that Daniel was an unprecedented event of increased river discharge levels (further discussed in section 5). At this cyclone stage, 17 human casualties were registered in Thessaly, along with a profound hydrological aftermath. The extreme rainfall from 3 to 8 September 2023 led to widespread flooding across 1,150 km² in the Thessalian plain, 70% of which constituted agricultural land. The inundation severely affected the cotton crops, with floodwaters covering more than 282 km², roughly 30% of the region's total cotton fields. Over 35,000 farm animals were also affected (He et al., 2023).

…..

Figure 4 highlights the exceptional river discharges in the region, as in the case of Greece. However, the absence of similarly strong discharge signals in several severely impacted regions, such as the wider Pelion area in Greece and Derna (Libya), is notable and can be attributed to several factors. First, the GloFAS model has limitations in spatial resolution and calibration. The model operates at a resolution of approximately 5 km (0.05°), which, while adequate for global-scale flood awareness, is insufficient for resolving localized hydrological dynamics. For instance, the catchments east of Volos, including the wider Pelion area, are approximately 30 km2, while the Derna basin spans around 575 km2. In both cases, localized rainfall-runoff dynamics are critical in shaping discharge patterns, particularly during extreme events. Due to insufficient in-situ discharge data, the absence of Greece and Libya in the GloFAS calibration dataset further exacerbates these limitations since the model relies on generalized parameter regionalization rather than site-specific calibration, introducing significant uncertainties into discharge predictions. Furthermore, inaccuracies in the rainfall inputs depicted in Figure 2 propagate into the discharge simulations shown in Figure 4. For instance, within the Peneus catchment, the maximum recorded 24-hour accumulated precipitation was 274 mm at Zappeio and 226 mm at Neraida stations, as Dimitriou et al. (2024) reported, while accumulations of up to 750 mm were recorded outside the catchment, specifically over the Pelion area, east of Volos. In the Wadi Derna catchment, extreme rainfall exceeded 400 mm day-1, with torrential rainfall ranging between 150 and 240 mm across several locations and Al-Bayda recording a maximum of 414.1 mm (WMO, 2023). These rainfall extremes were underrepresented in the GloFAS rainfall inputs, propagating into the discharge simulations and contributing to the muted signals observed in Figure 4(b)."

Figure 5a,b: The colour scheme used is poorly chosen as it does not allow for accurate values to be read off. However to me it looks like the value of a +2C anomaly quoted in the text should actually be +1C (save perhaps for the area N of Derna on (b) where it may be +2C). This is especially true if one references both 5a and 5b instead of just 5a, which would *be justified as the lifecycle is then better covered. This would be a bit of a counter argument against the misleading statements regarding climate change influence made late in the manuscript. Furthermore the blue patch of negative SST anomalies on 5b, which may be a legacy of Daniel's upward fluxes, is not discussed; indeed the manuscript contains no reference to 5b at all, so far as I can see.*

We appreciate the reviewers' comments on Figure 5 and acknowledge the importance of providing precise and clear figures for the readers. In response, we have enhanced the figure by adding additional isolines to eliminate any ambiguity regarding the SST distribution during the selected days. Additionally, we have revised the corresponding text in the manuscript as follows:

"Figure 5a, b shows the SST anomaly in the area affected by Storm Daniel on 3 and 9 September, respectively. Before the passage of Storm Daniel, positive SST anomalies dominated the study area, with values exceeding 1°C between the Libyan coast and Greece, and lower anomalies (0 to 0.5°C) observed east of Sicily. Following the storm's passage, a significant drop in SST resulted in an extensive area of negative anomalies greater than 1°C between Libya and Greece. A colder SST core with a decrease of less than 1.5°C was observed east of Sicily, while the northern Aegean Sea experienced an even more pronounced decline. Such SST cooling after the passage of medicanes has been previously diagnosed using explicitly resolved air-sea interactions in coupled atmosphere-ocean models (Ricchi et al., 2017; Bouin and Lebeaupin Brossier, 2020; Varlas et al., 2020) and SST observations (Avolio et al., 2024). Nevertheless, the feedback mechanism between cyclones intensity and SST cooling is expected to be less important than the one typically observed in tropical cyclones."

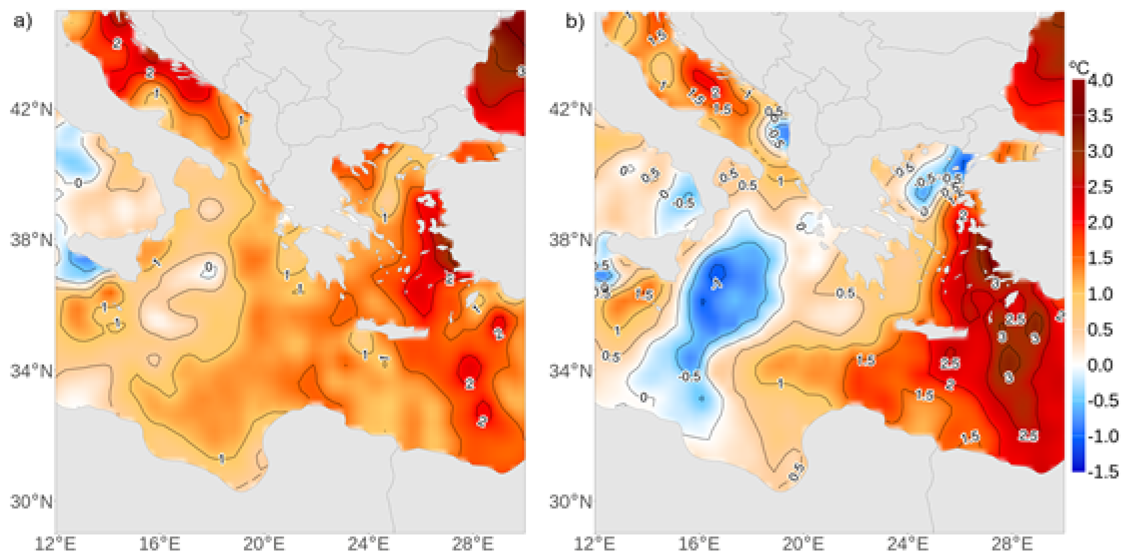

Figure 5 (a) Daily SST anomaly from ERA5, for 3 September 2023, and **(b)** 9 September 2023. The reference climatology for anomaly determination is 1982-2011.

Figure 6: This is a nice figure. However a related statement in the text that "it is impossible to evaluate the relative socio-economic impact of each threat (storm surge, waves, rain, river flood)" seems rather preposterous when we know that >5000 people lost their lives in Derna as a result of a dam burst (due to rain and river flooding causing overtopping).

We have removed this phrase according to the reviewer's suggestion.

Figure 7: Ok but the right-hand panels are not valid for 10 September at 12UTC. I also have doubts about the valid time of the left hand panels given that the cyclone centre seems to have a rather different position to that shown on Figure 2a. Or maybe Fig 2a is the one that's wrong? The statement in the text that 7a shows a much larger area of high PV than Fig. 2 is not correct. It is the other way round (note we only have 2PVU on Fig 2). And for PV averaging it might anyway be better to take the log first, given PV structure/ranges? An analogy is that one cannot meaningfully average visibility (across several orders of magnitude). Whilst this figure and the next one highlight clear convergence in the EPS solutions, which is OK, the text fails to acknowledge that relative to what came beforehand, the forecasts from 12UTC 1st (the first one included) actually represented a big positive step in skill – at least they had cyclones – due to much better handling of the mid-Atlantic Rossby wave train, due in turn to better handling of a tropical cyclone (as in Hewson et al, 2024). This is an example where one sees that the manuscript is not adding much to previous work, and indeed is contradicting it somewhat. These two results would need to be placed alongside each other in this paper to give the full context of cyclogenesis predictability for this case, and thereby advance the science as is required for a publishable standard. For Fig 7h the discrete 300hPa high PV blob west of Daniel is not mentioned. This very likely links to the upper level low moving in from the west from Hewson et al (2024), that is referenced, so a useful connection could be made here, pointing out also the increased specificity of this feature as lead times reduce, as shown by 7b,d,f,h.

Thank you for spotting this inconsistency. We corrected the valid time for consistency with Fig. 2. The text has been modified accordingly, and now we show the median of ensemble members. We understand the concern about averaging PV, but this is practically the case for any meteorological variable that does not follow a Gaussian distribution. Here, it is already addressed with the colour crosses showing percentiles of members exceeding different thresholds. We prefer not to take the log of PV values, as it is neither physically found nor applicable to zero or negative values.

Regarding comparison with previous work, we extended the lead time to one week ahead to include the jump in predictability between 5 and 7 days for the cyclogenesis stage. We now discuss the link with the upstream ET of Hurricane Franklin referred to in Hewson et al. 2024. We also clarify that the PV blob west of Daniel during the mature stage, discussed at the end of Section 4.2, is marked by high PV values in Fig. 7h. However, we would like to stress that Hewson et al. 2024 suggest but do not demonstrate the impact of the upstream hurricane and PV blob.

Figure 8: The left hand panels do not appear to be valid for 5 Sep 12UTC. Judging from the cyclone spot cluster they may be for 5 Sep 18UTC. Similarly the spots on (g) do not seem to correspond with the mslp minimum on Fig 7g, suggesting these panels are not for the same time. This is all rather confusing. The valid time for b,d,f,h looks to be correct.

Thank you for this remark. The left-hand panels are now valid for 6 Sep 00 UTC. Please also note that the shown lead times have been changed to be consistent with Fig. 7. We thus now include the predictability "jump" in the ensemble system.

Figure 9: The reader is left to guess what the valid time range is for the precipitation objects. It may be that it is the 24h periods ending at the stated valid times, yet if that is the case why

use 5 Sep 12UTC as an end time when the main 24h rainfall period was 00-24UTC on 5 Sep or a bit later (again reference Table 2 in Dimitriou et al, 2024)?

Indeed, we show 24h accumulated precipitation ending at the stated valid times. We clarified the caption of Fig. 9 in the revised manuscript. As we now show 6 Sep at 00 UTC as valid time, we include the main rainfall period from 00-24 UTC on 5 Sep. Please also note that the shown lead times have been changed to be consistent with Figs 7 & 8.

Figure 10: This figure is fine but I do not understand what it intends to show – the text "This shift is plausibly relevant.." I have not managed to decipher. Adding spots on the grey track lines, to show cyclone centres at a particular valid time, could help.

This part of the text has been revised and the figure has been updated.

"When Daniel made landfall and produced impacts on the Libyan coasts, the EPS showed higher predictability, with cyclone objects and associated extreme precipitation being predicted at least five days in advance by several EPS members (Fig. 10d), albeit the location of both cyclone and precipitation objects are still displaced to the southwest compared to the analysis (Figs. 8d and 10d).This comes in accordance with the southern displacement of several ensemble member tracks in Fig. 9a. The probability strongly increases at shorter lead times (Figs 10f and 10h) mostly and all EPS members tend to converge to similar cyclone locations when reaching a lead time of one day (Fig. 8h)."

Figure 11: This figure looks potentially informative but the elements of it are not explained, and furthermore some elements are barely visible (grey tick marks overlapping the box and whiskers). First readers should be pointed to where the Pinios river outlet is, and what its catchment is. According to Wikipedia the spelling should be Pineios (though I concede that could be "wrong"). It would also help to see the Wadi Derna catchment – the relative size of this, versus the Pineios catchment, is very important for predictability and impact prediction and this is not discussed. Then where does the "perfect forecast" benchmark come from. Is it related to the rainfall in Figure 2b, which as stated above looks wrong (hardly perfect!) in the critical area? Then what do the box and whiskers relate to, and why do they have a strange shape? What are all the percentiles represented? It is fairly clear to me that the forecasts for Greece converge onto the "right" solution (if the red curve can be trusted), whilst the forecasts for Libya, though overall they get a bit better with lead time, basically do not converge. The forecasts from 9th for Derna, which might be at the most critical for triggering preventative measures, step back from those of the previous day, and then even from 10th we still have huge spread and a big shortfall in the box and whisker median (if that's what the middle black line is). Yet all the text says about the Derna forecast is that it follows a "similar pattern" to the one for Greece. This is an incorrect and unhelpful sweeping statement. Furthermore, the following paragraph goes on to say that Fig 4b highlights the unprecedented nature of the event, when for Derna and its catchment the signal is rather weak. The much stronger signal is well to the west (also discussed above).

We appreciate the reviewer's detailed feedback on Figure 11 and acknowledge the need to provide additional clarity and context for this figure.

First, in the revised Figure 4, the Peneus and Wadi Derna catchments are delineated, and the Peneus River outlet is marked. The addition provides context for understanding the relative sizes, critical for understanding the contrasting hydrological behaviours and predictability challenges of the two basins.

New Fig. 4:

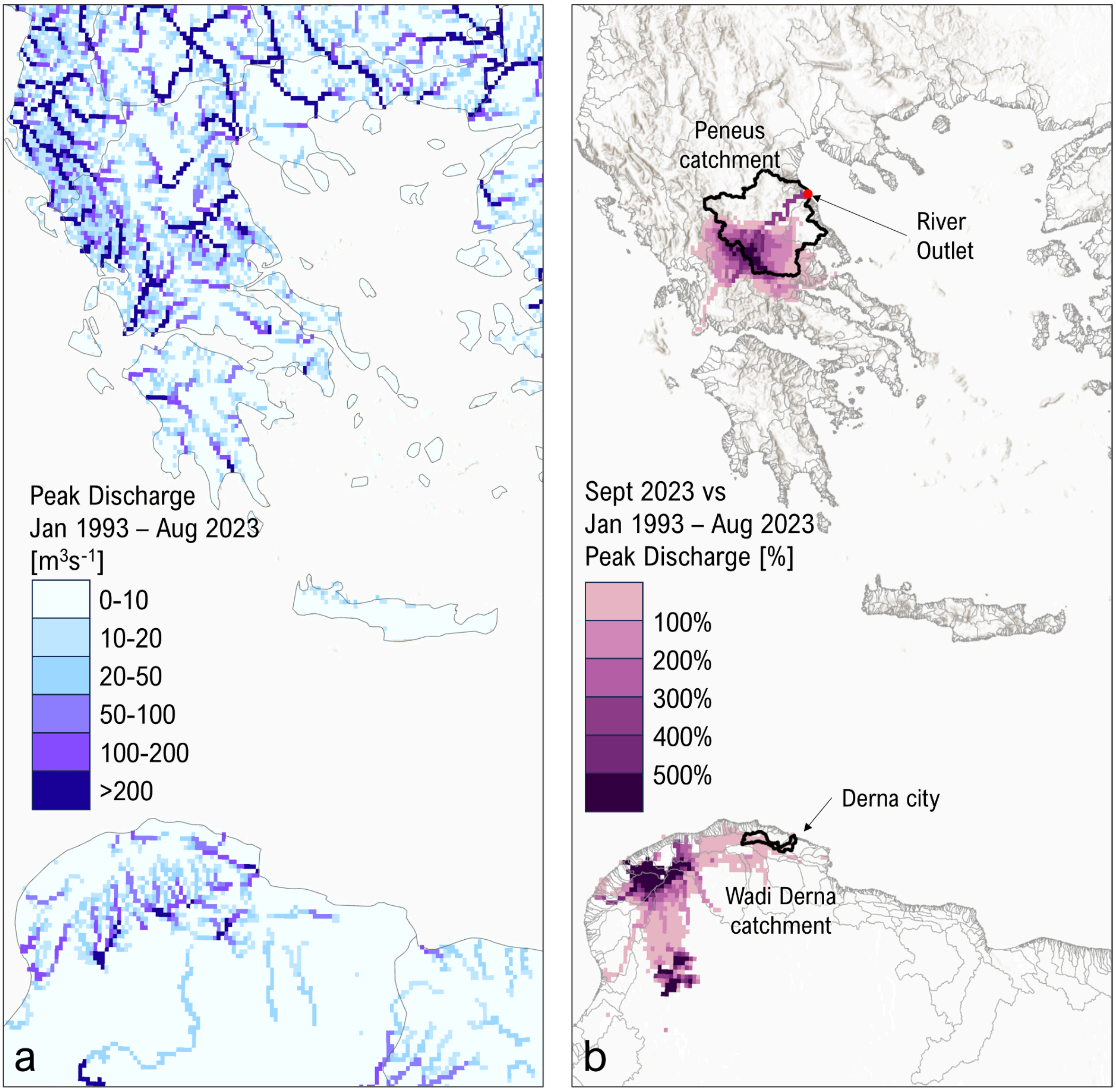

Figure 4 *Peak discharge over three recent decades (Jan 1993 – Aug 2023) versus Daniel storm as represented by the Global Flood Awareness System* ***(a)*** *spatial distribution of the maximum peak river discharge from January 1993 to August 2023,* ***(b)*** *comparison map for September 2023 illustrating the event-wide peak discharges as a percentage increase over the maximum peak discharges during the 30 years January in* ***(a)****.*

Regarding the spelling, we acknowledge the reviewer's observation that the standard spelling is not "Pinios" (nor "Pineios" as listed in Wikipedia) but "Peneus", and we revised the manuscript to use this spelling throughout consistently.

Next, Figure 11 has been revised to ensure clarity and provide a more detailed explanation of its elements. The updated caption now explicitly defines all the features of the figure, addressing the concerns raised regarding the visualization and interpretation of the data.

The grey tick marks represent individual ensemble members from the EFAS model, driven by the 51 ensemble members of the ECMWF EPS. Overlapping tick marks darken, visually highlighting areas of ensemble member agreement (convergence). This approach intentionally uses light grey to ensure that convergence areas stand out, helping readers intuitively grasp the degree of forecast agreement. Forecast summary data are displayed as boxplots, with the box representing the interquartile range (IQR), the whiskers showing the range of values within 1.5 times the IQR, and the horizontal black line inside the box indicating the median forecast. Additionally, the notches around the median depict the 95% confidence interval, providing a measure of uncertainty around the median forecast. The "perfect forecast" benchmark (red line) represents the initialization of each forecast for all time steps across the event, serving as a reference for evaluating forecast accuracy. The figure demonstrates the contrast between forecast performance in the Peneus and Wadi Derna catchments. For the Peneus catchment (~11,062.2 km^2), the forecasts converge well onto the observed discharge as lead time decreases, reflecting greater predictability for larger basins with distributed hydrological processes. In contrast, forecasts for the Wadi Derna catchment (~575 km^2) exhibit significant variability and lack convergence, even at shorter lead times. The boxplots in Figure 11 have been explained in detail to clarify the elements. The box represents the IQR, the whiskers show the range of values within 1.5 times the IQR, and the horizontal black line indicates the median. The notches around the median provide a 95% confidence interval, highlighting forecast uncertainty. This explanation is now explicitly included in the updated caption to ensure readers fully understand the visual representation. The revised caption reads as follows:

“Figure 11 Six-hourly ensemble river discharge forecasts for the Peneus and Wadi Derna catchments compared to the “perfect forecast” benchmark (red line). The "perfect forecast" represents the initialization of each forecast for all time steps across the event, taken as a reference for evaluating forecast accuracy. With the observed timing of rising hydrograph limbs marked on 5 September, noon local time (09 UTC) for the Peneus River in Thessaly, and 10 September, 18:00 local time (16 UTC) for the Wadi Derna River. Grey stripes (tick marks) represent individual ensemble members from the EFAS model, driven by the 51 ensemble members of the ECMWF EPS. Overlapping tick marks darken, visually highlighting areas of member agreement (convergence). Forecast summary data are displayed as boxplots, where the box represents the interquartile range (IQR), the whiskers show the range of values within 1.5 times the IQR, and the horizontal black line inside the box indicates the median. The notches around the median show the 95% confidence interval.”

Furthermore, we have revised the discussion to emphasize the distinct differences in forecast performance between the two catchments. The earlier text inaccurately stated that the Wadi Derna forecasts followed a "similar pattern" to those for the Peneus catchment, which oversimplified the differences. This has been corrected to highlight the lack of forecast convergence in the Wadi Derna catchment and its implications for predictability and response planning. We acknowledge the reviewer’s observation that Figure 4b shows a weaker discharge signal for the Wadi Derna catchment compared to areas further west. As already stated in our reply to comments regarding Figure 4, this reflects limitations in the GloFAS and EFAS models, including their resolution and rainfall input accuracy. This has been addressed in the main text to provide context for the variability and underrepresentation in forecasts. The revised text reads as follows:

“The potential of extreme precipitation leading to substantial socio-economic impacts has also been transferred to hydrologic discharge forecasts. The hydrographs presented in Fig. 11 examine river discharge predictability as forecast by the operational European Flood Awareness System (EFAS) during Daniel. For the Peneus River outlet in Thessaly, the forecast initiated on 1 September underpredicted the peak discharge on 5 September. Nevertheless, extreme discharges were evident for several members five days in advance. The forecast accuracy improved getting closer to the event, with ensemble members (grey stripes) converging towards the peak discharge (“perfect forecast” - red line). This trend indicates an increasing reliability of the forecast as the lead time decreases, particularly within 48 hours of the event. The skill in discharge predictability for the Peneus River can be

attributed, in part, to the large size of the basin (11.063 km2), which aligns relatively well with the spatial resolution of the EFAS model, enabling an accurate representation of distributed hydrological processes and moderating runoff variability.

The forecasts for the Wadi Derna River outlet (Fig. 11, right panels) exhibit significant variability and fail to converge during the earlier forecast initialization dates as well as at shorter lead times. This persistent lack of convergence can be attributed to distinct challenges of both temporal scales. For earlier forecast initialization dates, the primary source of variability lies in the westward displacement of extreme precipitation predicted by the EPS (Figs. 10b and 10d). For example, forecasts initialized on 9 September, during a critical period for implementing preventative measures, display a wide spread and a shortfall in the median forecast compared to the benchmark (red line). This variability persists even for forecasts initialized on 10 September. The failure to converge at shorter lead times is compounded by challenges inherent to the Wadi Derna catchment. The resolution of the precipitation forcings used in the forecasts combined with the relatively small size (575 km2) and flash-flood-prone nature of this basin amplify the uncertainties in predicting discharge, particularly in response to localized extreme rainfall.

Figure 4 provides critical context by comparing the peak mean daily river discharge during Storm Daniel with the historical baseline. The unprecedented magnitude of the event is evident in Fig. 4b, where discharges exceeded the historical reanalysis by at least fivefold in certain regions. However, the relatively weak signal for the Wadi Derna catchment underscores the limitations of the GloFAS and EFAS systems in accurately resolving runoff dynamics in smaller basins. This discrepancy is primarily attributed to insufficient model resolution, inaccuracies in rainfall inputs, and the lack of detailed hydrological calibration for these catchments. In contrast, the much stronger signal observed in the Peneus catchment aligns with larger basin sizes and better-resolved hydrological processes, where models more effectively captured the extreme nature of the event.

The ability of EFAS to predict extreme events, as shown in Fig. 11, highlights its value in forecasting severe hydrological impacts. However, discrepancies in simulated peak discharge remain apparent, such as the overestimation of runoff for the Peneus River outlet. EFAS simulated peak discharge at approximately 5000 m^3 s-1, whereas observed values, based on station-level data and H-Q curve estimates, were less than 2000 m^3 s-1 (Dimitriou et al, 2024). This overestimation reflects inherent limitations in the model's spatial resolution and hydrological representation. Furthermore, the absence of flood protection infrastructure, such as levees or dams that attenuate runoff and peak flows is not accounted for in the EFAS and GloFAS systems, contributing to these discrepancies. Additionally, the simplified representation of retention processes, including floodplain storage and wetland buffering, further amplifies discharge estimates in some regions. For smaller basins such as Wadi Derna, the rapid hydrological response to localized extreme rainfall presents additional challenges. The variability in rainfall distribution, coupled with the model's limited ability to capture localized hydrological dynamics, results in a weaker signal for the catchment, even during an event as extreme as Storm Daniel. These limitations emphasize the need for improved model resolution, enhanced precipitation forcings, and better calibration tailored to local catchment characteristics.

Nonetheless, the ability of EFAS to predict extreme discharges, particularly within short lead times, demonstrates the potential of operational forecast systems in capturing the extreme values of such events. Supported by EFAS and GloFAS, the Copernicus Emergency Management Service (CEMS) provides critical insights into the timing and magnitude of extreme hydrological events. These forecasts are vital for enhancing preparedness and response strategies in the face of escalating climate extremes, offering essential tools for civil protection efforts and mitigating the socio-economic impacts of such disasters."

Figures 12 and 13: On many of the panel legends the numbers do not align with the colour bars. So the reader does not know what the colour bars mean. This is obviously important when one tries to cross-reference with the text – e.g. on Fig 12h it is stated that temperatures have gone up by 2C in the Ionian Sea when it looks like rather less than that. Then why are there contours as well as shading on panels d,h,l and p? The fact that the rainfall amounts for the 2023 case in this depiction under-represent reality by a large margin is not mentioned, when clearly this has relevance (panels i). The worst part about this part of the study is that the conclusions in the text do not reflect what the figures show. For example, the authors state "we conclude that Mediterranean depressions like Daniel hitting Greece and Libya show lower MSLP and higher precipitation in the present climate than in the past". The evidence for this is supposed to be panels d which show basically no mslp change at all; and panels l which show drier over Greece and slightly wetter over the seas around Libya. And maybe +2mm or so per day over northern Libya itself, but when >400mm/24h was recorded at one site for Daniel this seems irrelevant. The text of Section 5.2 contains many other errors and inconsistencies, too numerous to go into here. In my opinion the vast majority of Section 5.2, for which these Figures are the "evidence" should be removed from the paper, as it shows very little of substance. One could much more usefully and honestly say, in brief, that "an in-depth study using standard methods indicates that in the ERA5 dataset there is no evidence of climate change influencing features like Daniel in the 1980-2020 period". The only non-neutral "result" I can see on these figures is a signal for an increased frequency for cyclones, in the SOND period, in the SE Mediterranean near the N African coast (Fig 13x). So that could be referenced too. Furthermore, it seems to me that trying to link El Nino, the PDO and the AMO to Daniel-like cyclones over just a 40-year period is stretching physical credibility beyond its natural limit.

We appreciate the reviewer's detailed comments, particularly regarding the interpretation of Figures 12 and 13. The different color scales used in panels d, h, l, and p were chosen deliberately to ensure that the changes in variables remain visible. If these panels were placed on the same color scale as the others, the magnitude of changes would be difficult to discern, making it harder to interpret the results. However, we have carefully rechecked the data and confirmed that the value of +2°C in the Ionian Sea, as indicated in Figure 12h, is correct. Regarding the contours in these panels, we clarify that they represent areas where the changes are not statistically significant. Only shaded regions indicate significant differences, ensuring the reader can easily distinguish between robust trends and regions where changes may occur due to natural variability. To enhance clarity, we will explicitly state this in the figure captions and corresponding text.

We also acknowledged the limitation of the MSWX dataset in capturing extreme precipitation values. The reviewer rightly pointed out that our figures under-represent the actual rainfall totals observed during Daniel, particularly in Libya. To address this, we have now explicitly stated in the text that our analysis focuses on large-scale climatological trends rather than station-level extremes. While MSWX provides valuable insights into broad atmospheric patterns, it does not fully capture the localized intensities that contributed to the flooding disaster in Derna. Recognizing this, we revised our discussion to ensure that our conclusions remain in line with the limitations of our dataset.

The most significant revision involved refining our conclusions regarding changes in MSLP and precipitation. Initially, our text suggested that Mediterranean depressions like Daniel show lower MSLP and higher precipitation in the present climate. However, as the reviewer correctly noted, the evidence for significant MSLP changes is weak. After carefully reassessing our figures, we adjusted our conclusions to clarify that there is no strong trend in MSLP over Greece or Libya. Instead, we emphasize that the observed increases in precipitation are likely driven by rising sea surface temperatures, which enhance atmospheric moisture availability. By making this distinction clearer, we ensure that our findings are scientifically sound and accurately represent the data.

We also addressed the contrast between the two phases of Daniel's evolution. On September 5, when the storm affected Greece, we found suitable past analogues, indicating that this type of event had occurred before. However, by September 10, when Daniel reached Libya, our analog search identified no comparable historical events, underscoring the exceptional nature of the atmospheric conditions at this stage. This revision strengthens our argument that while the storm's early track and intensification over the Mediterranean were within expected climatological behavior, its final phase was highly unusual. Notably, we also revised our discussion on the disaster in Libya to highlight that while Daniel's rainfall was intense, the catastrophic flooding in Derna resulted primarily from infrastructure failure, particularly the collapse of the dams. We referenced recent research, including Dente et al. (EGU 2024) and Shirzaei et al. (2025), which showed that regional precipitation levels were not exceptionally high but that vulnerabilities in urban planning and emergency response significantly worsened the disaster.

The reviewer also raised concerns about our large-scale climate variability modes analysis, particularly ENSO, AMO, and PDO. We recognized that attempting to link these modes to Daniel-like cyclones over a 40-year period is speculative and complex to establish confidently. In response, we revised our text to clarify that this part of the analysis is exploratory rather than conclusive. We now explicitly state that while these modes may influence atmospheric conditions, our findings do not establish a causal link between them and Daniel's development.

Finally, we revised Section 5.2 to focus on the most robust findings and ensure our conclusions align with the evidence. Rather than removing the section entirely, we streamlined it to emphasize three key points: (1) no significant changes in MSLP were detected, (2) increased precipitation is most likely due to higher sea surface temperatures rather than shifts in atmospheric dynamics, and (3) there is a clear increase in the frequency of similar Mediterranean depressions in the southeastern Mediterranean near the North African coast. These revisions ensure the section remains informative and scientifically rigorous while avoiding overstated conclusions.

We appreciate the reviewer's constructive feedback, which has substantially improved our study. These changes have already been implemented in the revised manuscript, ensuring that the data clearly presents and accurately supports our findings.