

Response to Referee Comments on “Heavy Precipitation Events of Various Durations Across Germany: A Station-Based Assessment of Spatial and Temporal Variability” (Discussion Phase, egosphere-2026-1067)

- **Referee:**

This manuscript explores the spatial and temporal variability of heavy precipitation events across Germany using high-resolution data from various rain gauges. By applying a block maxima sampling, the authors analyse annual maximum precipitation totals for durations ranging from five minutes to seven days. They apply simple statistical metrics and tests to assess variability and trends. The findings indicate that while short-duration events are scattered and driven by local convective processes, longer-duration events are more influenced by topography and large-scale atmospheric patterns. Although positive trends in intensity are reported in short-duration (e.g. hourly) events, a large fraction of trends are not significant. For long-duration rainfall (e.g. 7 days) also negative trends are reported, where the significance depends on the length of the analysed time series. The study is affected by major drawbacks: [...] Due to these major points I'd suggest to reject the current version of the study but to encourage the authors to re-submit if they are willing to address these major points. At this point, I won't add any minor comments on figures, references, wording, etc.

Response:

We thank the referee for the careful assessment of our manuscript and sincerely appreciate the time and effort invested in providing detailed comments. At the same time, we acknowledge the concerns and major drawbacks raised in the review and take them very seriously.

In the following, we provide a detailed point-to-point response to the referee's comments.

- **Referee:**

The manuscript lacks a scientific discussion section

Response:

We agree that the current manuscript lacks a clearly structured scientific discussion section. As also noted by Referee 1, parts of the current Conclusions contain elements that are more appropriate for a discussion. In the revised manuscript, we will therefore restructure this part by separating it into a dedicated Discussion section and a concise Conclusions section. The Discussion will focus on the interpretation of the main findings, including their relation to previous studies, as well as the limitations and uncertainties of the applied methods. The Conclusions section will be shortened to a clear and concise summary of the key results.

- **Referee:**

5-min rainfall intensities show jumps, which are presented as a scientific finding. In turn, they should be described in the data description, where also the reason (change of sensors) should be reported. Can those jumps be corrected? Otherwise, I'd suggest removing these 5-min-results from the manuscript, especially in the trend analysis.

Response:

We appreciate this important comment regarding the inhomogeneities in 5-minute AMPTs and interpretation of trend results for very short durations. We agree that the observed jumps in 5-minute AMPTs, which are to a large extent related to changes in the measurement system (i.e. sensor transitions), should not be interpreted as a scientific finding. Therefore, in the revised manuscript, we will relocate and expand the description of the 5-minute data and associated challenges (e.g. jump detection and limitations of their correction) to Section 2.1, while omitting the interpretation of the corresponding trend analysis results for very short durations from Section 3.2. In addition, we will remove all panels representing very short durations from Figs. 6-11, as they are prone to misinterpretation. We believe that this approach ensures that the manuscript focuses on robust and interpretable results, while transparently documenting the limitations of the underlying data.

- **Referee:**

Statistical tests seem not to be adjusted for multiple testing correcting the false discovery (FDR; see e.g. Wilks, 2016: <https://doi.org/10.1175/BAMS-D-15-00267.1>).

Response:

We thank the referee for this important suggestion. We agree that the issue of multiple testing should be addressed to avoid overestimation of statistical significance. In the revised manuscript, we will therefore apply a false discovery rate (FDR) correction following the Benjamini–Hochberg procedure to the set of p-values obtained from the Mann–Kendall trend tests. Following Wilks (2016), we will adopt a significance level of $\alpha = 0.10$ for the FDR-adjusted results. In addition, to ensure transparency and comparability with previous studies, which to a large extent do not apply FDR correction, we will present both the raw (unadjusted) and FDR-adjusted results in the revised manuscript. A preview of these results is provided in the figure below – as expected, the comparison indicates that, after accounting for multiple testing, the number of statistically significant trends is further reduced and remains very limited, with only a small fraction of stations showing robust significance.

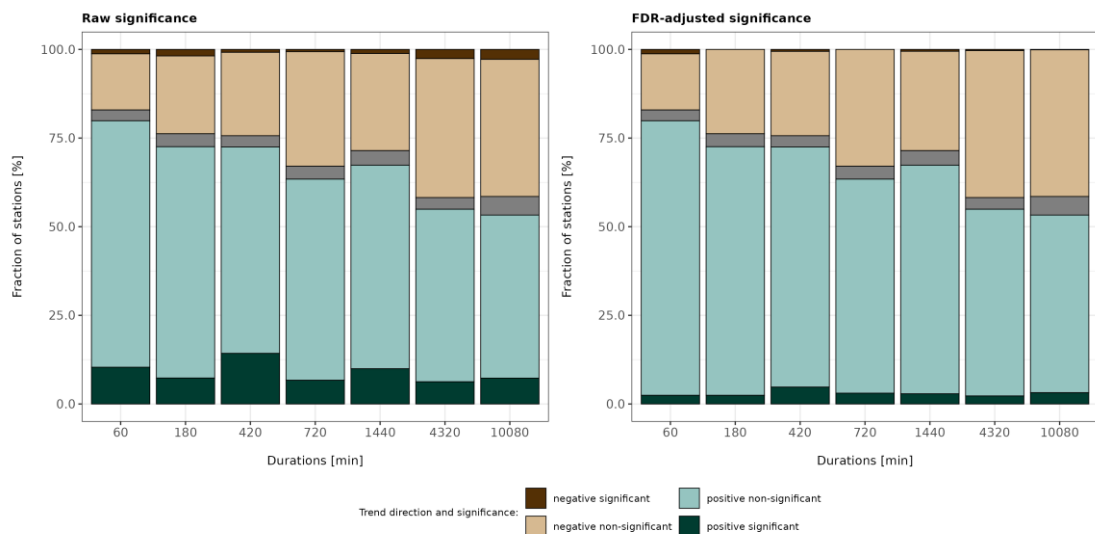


Figure 1: Fraction of stations [%] exhibiting different trend directions and significance in AMPTs (same as Fig. 6 in the manuscript), but shown here for both raw (unadjusted) and false discovery rate (FDR)-adjusted results. Very short durations (5 and 30 minutes) are omitted following reviewer comments due to increased uncertainty associated with data inhomogeneities.

- **Referee:**

Trend analysis for long durations might benefit from seasonal maxima, as the behaviour of decreasing/increasing maxima might be hidden in changes of seasonal maxima. In some parts of Germany, long duration maxima dominate during summer (e.g. southern Germany), while in other parts long duration maxima dominate during winter (e.g. north-western Germany). Both seasons might show different trends (drier summers, wetter winters).

Response:

We thank the referee for this insightful comment. We agree that a seasonal perspective on maximum precipitation totals could provide additional insight into the underlying processes and potentially reveal contrasting trends between seasons. However, such an analysis is currently not feasible within the scope of the present study based on the high-resolution dataset. In particular, the 5-minute precipitation records exhibit substantial gaps during the winter season, either because certain measurement systems were not operated year-round or because historical data for these periods have not yet been fully digitised. The latter is partly related to the digitisation strategy, which prioritised summer periods based on prior analyses at selected stations indicating a higher likelihood of annual maxima occurring during this season. As a result, a seasonally resolved analysis based on the 5-minute precipitation records would not yield sufficiently robust results for all stations currently considered in this study. For example, when requiring at least 30 years of data and a minimum seasonal data availability of 80 % per year, only around half of the stations provide sufficient coverage for DJF, MAM, and SON (49-53 %; 116-125 stations), whereas

JJA retains full coverage (237 stations). In the revised manuscript, we will explicitly discuss this limitation and its implications for the interpretation of the trend analysis results. Nevertheless, we agree that a seasonally resolved perspective represents a promising direction for future work and will highlight this accordingly in the revised manuscript.

- **Referee:**

The manuscript would benefit from a more consistent usage of the term "variability," which currently conflates several distinct concepts. Specific instances of inconsistent usage include:

- *Spatial variability: The term is used to describe the geographic distribution of AMPT medians (Line 191)*
- *CV is used to describe the relative interannual fluctuations of AMPT (Line 197) and "also the event-to-event variability at each station"*
- *Temporal variability: This refers to long-term monotonic trends (Line 161)*
- *Scale-dependent "visual" variability: In lines 175–178, the authors refer to a "visually apparent" greater "overall variability" in long-duration events due to the widening of the interquartile range (box plots), while simultaneously clarifying that this does not represent an "actual increase in their variability". Describing the absolute range of variability for cumulative rainfall is somewhat misleading; if the y-axis of Figure 3 would show intensities instead of cumulative rainfall, the visual notion of variability would probably be inverse.*
- *One important concept of variability (internal climate variability; ICV) is not fully captured in the manuscript. Partly, event-to-event variability is governed by ICV. Later, in lines 366–375, "decadal fluctuations" implicitly describe the realization of ICV on the decadal scale. It is important to understand that the ICV is the driver for year-to-year temporal variability but also spatial variability, especially for smaller-scale convective events governed by the chaotic behaviour of the climate system.*

While these different concepts can and should be applied in the manuscript, they should be clearly distinguished and also understood.

Response:

We thank the referee for this insightful comment regarding the use of the term "variability". We agree that the manuscript currently uses this term to describe several distinct concepts, which may lead to ambiguity. In the revised manuscript, we will carefully revise the terminology throughout the manuscript to ensure a consistent and precise use of the term "variability". In particular, we will consistently differentiate between: (i) spatial variability (i.e. differences in AMPT characteristics across stations), (ii) interannual variability (i.e. year-to-year fluctuations at a given station, quantified, for example, by the coefficient of variation), and (iii) long-term temporal changes (i.e. monotonic trends). Furthermore, we will revise the corresponding passages in the manuscript to avoid ambiguous formulations, especially where "variability" was previously used in a more qualitative or visual sense (e.g. in the interpretation of box plots). In the revised version, we will also clarify that part of the observed interannual and decadal fluctuations, as well as aspects of the spatial heterogeneity of AMPTs, may reflect Internal Climate Variability and the chaotic nature of the climate system, particularly for convectively driven events. This aspect will be addressed more explicitly in the new Discussion section.

- **Referee:**

The authors discuss possible reasons for the reported spatial patterns (convection for short-duration AMPTs, large-scale drivers for long-duration AMPTs, "decadal fluctuations" of short-duration AMPTs "explained by changes in the behaviour of large-scale circulation patterns, potentially reinforced by an intensified convective response to anthropogenic warming". A thorough statistical analysis of these assumptions could add value (and scientific novelty) to the manuscript making use of the high-resolution data set.

Response:

We agree that a more explicit consideration of the underlying physical drivers could further strengthen the interpretation and add scientific value to the manuscript. Therefore, we propose to incorporate an additional analysis of the relationship between AMPTs across durations and large-scale atmospheric circulation patterns using the operational Grosswetterlagen (GWL) classification provided by DWD. This classification is based on and continuously further develops the original Hess-Brezowsky framework (e.g.

James 2007 <https://doi.org/10.1007/s00704-006-0239-3>). Previous versions of the classification have already been applied in studies investigating the relationship between heavy precipitation events and atmospheric circulation patterns (e.g. Palarz et al. 2024, <https://doi.org/10.1002/joc.8323>). This will allow us to relate the observed duration-dependent behaviour of AMPTs to prevailing large-scale circulation patterns and to provide additional context for the proposed distinction between convective and large-scale drivers. A preview of preliminary results is provided in Figure 2. As illustrated, several circulation patterns show some duration-dependent differences. In particular, the cyclonic type of bridge across Central Europe (BMz) tends to be more frequent for short-duration events, whereas cyclonic north-easterly type (NEz), cyclonic north-westerly type (NWz), and trough over Central Europe (TrM) are more commonly associated with long-duration events. The medium-duration events exhibit transitional behaviour between short- and long-duration events. In the revised manuscript, we will further develop and refine this analysis to better account for spatial and temporal variability and the complexity of the underlying processes, while ensuring that it remains within the scope of the present study.

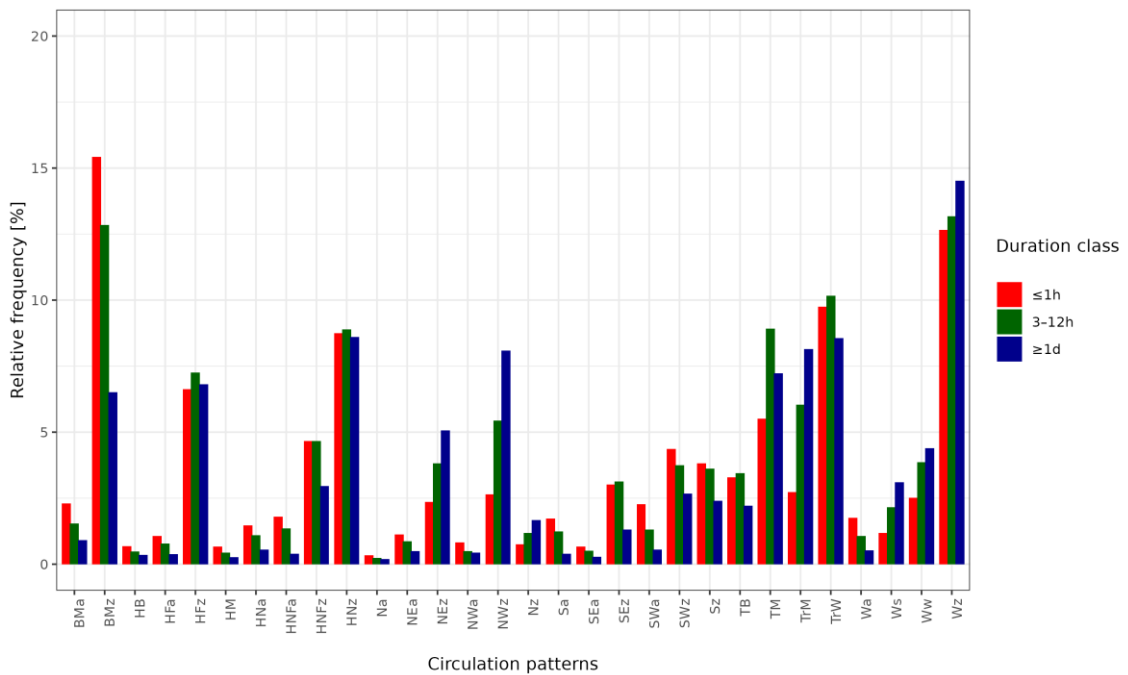


Figure 2: Relative frequency of circulation patterns [%] associated with AMPTs of different durations. For brevity, the original duration classes are grouped into three categories, i.e. ≤ 1 h, 3–12 h, and ≥ 1 day. Circulation patterns are defined using the operational Grosswetterlagen (GWL) classification routinely produced at DWD, which is based on and continuously further develops the original Hess–Brezowsky framework (e.g. James, 2007), with a detailed description of the current operational implementation provided by James & Ostermüller (2026, under review).

- **Referee:**

In sum, the degree of scientific novelty is low to moderate; the data set is valuable, however, this study is not a data description paper and the data cannot be accessed openly. Trends are still not significant (and might be even less after correcting the FDR); the from my perspective biggest novelty would be the finding that short-duration (5 min, 60 min) AMPTs show similar intensities across Germany, i.e. they are not as much governed by topography. In turn, Shehu et al. (2023: <https://doi.org/10.5194/hess-27-1109-2023>) show a considerable topography-influenced gradient for 1-year return levels of 5 min and 60 min rainfall (Fig. 13 in their article). I am not sure if your findings contradict (based on similar data sources; then you have to carefully explain why) or if the visualization of Fig. 4 leads to a biased perception of low differences, as the colorbar extends to 0mm. Hence, all circle markers are in relatively dark greenish-blueish color making them hard to distinguish leading to a low-difference-perception.

Response:

We thank the referee for this thoughtful and detailed assessment and for highlighting several important aspects regarding the interpretation and positioning of our results. We address the points raised below.

- Statistical significance of trends and implications for novelty: While we acknowledge the referee's point regarding the limited number of statistically significant trends, we believe that the scientific value of the study does not rely solely on the detection of significant monotonic trends. As discussed above, applying a correction for multiple testing (False Discovery Rate, FDR) indeed further reduces the fraction of statistically significant trends, leaving only a very limited number of significant results. Nevertheless, this finding is itself informative and of practical relevance. In particular, the general lack of robust and spatially coherent trends provides empirical support for the continued use of stationarity assumptions underlying current design precipitation approaches (e.g. KOSTRA-DWD-2020), at least for the study region and the considered time period. Moreover, our analysis is based on long-term, high-resolution (5-minute) station data, which remain relatively scarce for such duration-dependent assessments. The systematic evaluation of variability and trends across a wide range of durations (5 minutes to 7 days) and investigation periods contributes to an improved understanding of heavy precipitation characteristics in Germany, even in cases where trends are not statistically significant. To further strengthen the robustness and transparency of our analysis, we propose to include both FDR-corrected and non-corrected trend results in the revised manuscript. While the FDR-corrected results provide a more cautious and statistically rigorous assessment, the non-corrected results facilitate comparison with previous studies, where multiple testing corrections are often not applied. This combined presentation would allow for a more comprehensive and methodologically transparent interpretation of the findings, while also illustrating the importance of accounting for multiple testing and the risk of misinterpreting trend results if this aspect is not considered. Both the relevance of our results for the stationarity assumptions used in the derivation of design rainfall and the methodological implications of FDR correction will be discussed more explicitly in the new Discussion section. Furthermore, we will strengthen our analysis by incorporating an assessment of the relationship between AMPTs across durations and their synoptic drivers, using large-scale circulation patterns from the Grosswetterlagen classification routinely produced at DWD. This will allow us to provide additional context for the duration-dependent behaviour of AMPTs and to better link the observed spatial patterns to large-scale atmospheric circulation, thereby enhancing the scientific value of the manuscript.
- Comparison with Shehu et al. (2023): Assuming that the referee's comment refers to the maps of design rainfall (Fig. 14 in Shehu et al. 2023), rather than the maps of GEV parameters (Fig. 13 in Shehu et al. 2023), it should be noted that the results presented by Shehu et al. (2023) and our results cannot be compared on a one-to-one basis. This is due to differences in both the underlying data basis and the applied (geo)statistical methodology. While Shehu et al. (2023) present design rainfall maps derived from fitted extreme value distributions and subsequently interpolated using various geostatistical methods (note that the interpolated variables are the GEV and Koutsoyiannis parameters, not the precipitation sums themselves), our study is based on empirical, station-based statistics of AMPTs without any spatial interpolation. These methodological differences may lead to a smoothing of local-scale variability and the emergence of more coherent large-scale gradients in the approach of Shehu et al. (2023), while preserving the station-based characteristics and local variability of the observations in our analysis. Despite this, our results do not appear to indicate a clear contradiction with those of Shehu et al. (2023). In fact, Shehu et al. (2023) explicitly state that: *“For short durations (i.e. $D = 5\text{min}$), the spatial distribution of rainfall extremes is independent from the elevation and becomes more erratic with higher return periods. This is in accordance with the fact that the convective extreme events can happen anywhere and are very low correlated with the orography”* (Page 1125) and also that *“the elevation itself did not show much effect on the scale (σ) and θ parameter, only to some extent on the location (μ) and η parameter. This means that the extremes of longer duration (affected by the η parameter) and of low return period (affected by the location parameter) will show a pattern resembling the elevation. This is not true for short durations (affected by the θ parameter) and high return periods (affected by the scale parameter)”* (Page 1125). These findings are largely consistent with our results, which show weak topography-related gradients for very short durations (5 and 30 min), although broader regional differences between northern and southern Germany remain visible. At the same time, we agree that the current visualization in Fig. 4 may contribute to a perception of overly weak spatial variability due to the chosen color scale. In the revised manuscript, we will therefore revise the figure design to improve the visual differentiation of spatial patterns and to reduce potential bias in interpretation. Furthermore, we will expand the discussion by relating our findings to those of Shehu et al. (2023), as well as to

Lengfeld et al. (2021, <https://doi.org/10.1127/metz/2021/1088>), who, based on their *Catalogue of Radar-based heavy Rainfall Events (CatRaRE)*, report that: “For W3_ETA [i.e. the CatRaRE version based on DWD warning level 3 precipitation thresholds], *the dependence on orography is much less pronounced for short-term events than for long-lasting rainfall*” (Fig. 9, Page 479) and that “*Short-term events can occur everywhere with more or less the same probability and are less bound to orography*” (Page 489).