Response to Referee Comments on “Heavy Precipitation Events of Various Durations Across Germany: A Station-Based Assessment of Spatial and Temporal Variability Using the Block Maxima Method” (egusphere-2026-1067)

- **Referee:**

  Overall recommendation: Major revision

  This manuscript presents a valuable and timely contribution to the analysis of heavy precipitation extremes in Germany. Its main strength is the use of recently digitised and quality-controlled 5-minute precipitation data from German rain-gauge stations to compare annual maximum precipitation totals across durations from 5 minutes to 7 days. The manuscript is especially useful because it treats short-, medium-, and long-duration precipitation extremes within one coherent framework. It also explicitly recognises measurement-system inhomogeneities, which can strongly affect the interpretation of short-duration extremes.

  However, I recommend major revision because several methodological and interpretive issues need to be addressed more carefully, especially regarding trend detection, multiple testing, field significance, and the interpretation of physical processes.

  Response:

  We thank the referee for this constructive assessment of our manuscript and sincerely appreciate the time and effort invested in providing detailed and helpful comments. We also appreciate the referee’s recognition of the relevance of the topic, in particular the comprehensive investigation of short-, medium-, and long-duration precipitation extremes within a coherent framework, as well as the value of the underlying recently digitised and quality-controlled high-resolution dataset. At the same time, we acknowledge the concerns raised regarding methodological and interpretative aspects, including the interpretation of trend analysis results, the treatment of multiple testing and field significance, as well as the physical interpretation of the observed patterns.

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

- **Referee:**

  Measurement-system changes and trend detection
  The mixing of analogue and digital measurements, as well as sensor changes over time, makes it highly challenging to detect robust long-term changes in precipitation extremes. This issue is especially relevant for short-duration precipitation totals. The manuscript discusses this problem, but the implications for trend interpretation should be made even clearer.

  I recommend that the authors more explicitly discuss how much confidence can be placed in the detected trends given these measurement inhomogeneities. Where possible, sensitivity analyses should be added, for example by excluding stations or periods affected by known instrument changes.

  Response:

  We fully agree that measurement-system changes, including the transition from analogue to digital measurements as well as subsequent sensor replacements, make the reliable detection and interpretation of long-term trends in precipitation extremes particularly challenging, especially for the shortest durations. Therefore, and also following similar comments from the other referees, we propose to relocate and expand the description of the underlying 5-minute precipitation data and the associated challenges (e.g. jump detection and limitations of their correction) to Section 2.1, while omitting the interpretation of trend-analysis results for very short durations from Section 3.2 and to remove all panels representing very short durations from Figs. 6-11, as these durations are particularly prone to misinterpretation due to potential measurement inhomogeneities.

We also agree that additional sensitivity analyses, such as splitting the records before and after major measurement-system changes or excluding stations and periods affected by known inhomogeneities, would in principle be valuable. We carefully considered such analyses. However, for most stations, the transition from analogue to digital measurements occurred during the 1990s. Splitting the records at this point would substantially shorten the resulting subperiods, while the post-transition period would generally comprise fewer than 30 years of observations up to the fixed end year 2020. Such record lengths are generally considered insufficient for robust estimation of long-term trends in precipitation extremes and would considerably reduce the statistical power of the analysis. Nevertheless, acknowledging the importance of this aspect, we propose to include additional analyses in Section 2.1 that emphasize the potential influence of measurement-system changes on annual maximum precipitation totals (AMPTs). As an example, Figure 1 compares observed AMPTs, represented by the median and the 1^{st}-99^{th} percentile range, with the effective measurement ranges of historical and modern precipitation instruments used within the German observation network, including early- and modern-generation pluviometers as well as rain recorders with daily and weekly chart replacement. The figure illustrates that measurement limitations become increasingly relevant for very short durations, where observed extremes may approach or exceed the capabilities of some historical instruments. In turn, Figure 2 presents example time series of 5-min, 1-h and 1-day AMPTs from the station Erfurt–Weimar, which is affected by known measurement inhomogeneities associated with the transition from analogue to digital precipitation measurements on 1 January 1993. The figure highlights a pronounced shift in the 5-min AMPT record around the documented transition from analogue to digital precipitation measurements, with the median 5-min AMPTs increasing by approximately 83% (from 3.6 to 6.6 mm) after the measurement-system change. In contrast, the corresponding increases are considerably smaller for the 1-h (34%, from 15.6 to 21.0 mm) and 1-day (16%, from 38.0 to 44.0 mm) AMPTs. Together, these analyses indicate that measurement-system changes may substantially influence AMPTs of very short durations and thereby bias the estimation of long-term trends. Consequently, trend analyses for very short durations should be interpreted with particular caution, whereas corresponding analyses for longer durations are likely to be more robust.

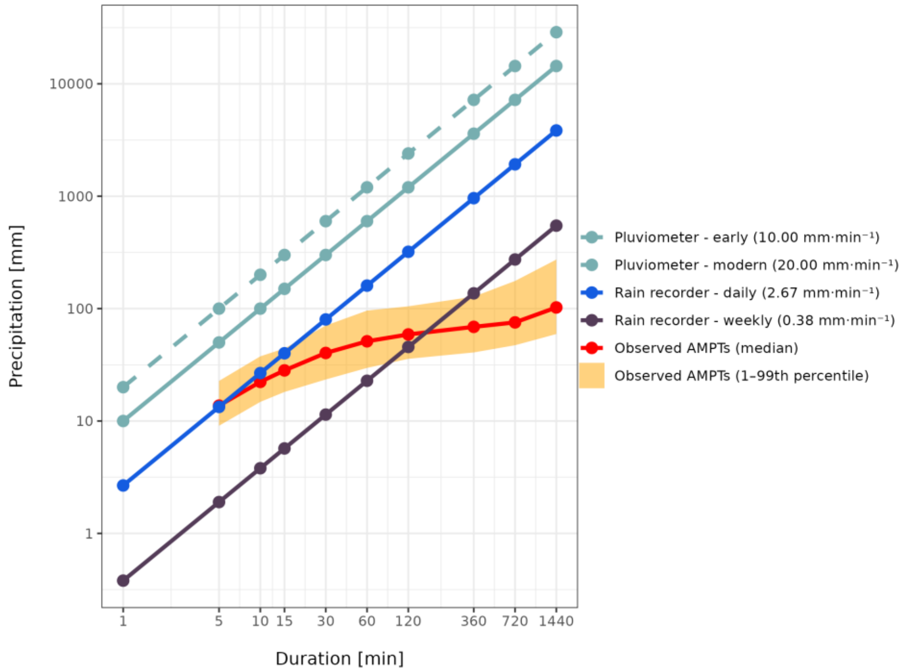

Figure 1: AMPTs represented by the median and the 1st–99th percentile range of the station-specific maximum AMPTs, together with the effective measurement ranges of historical and modern precipitation instruments, including early- and modern-generation pluviometers as well as rain recorders with daily and weekly chart replacement.

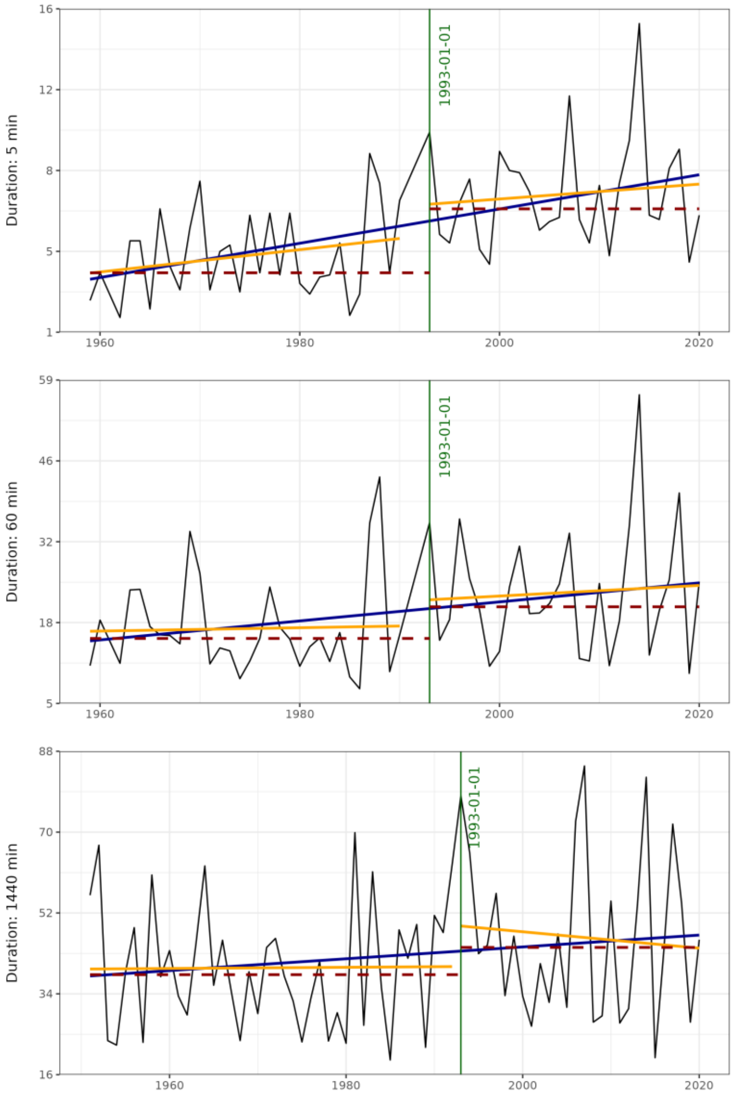

Figure 2: AMPTs for durations of 5 min, 1 h, and 1 day at station Erfurt–Weimar. Black lines show the observed AMPTs, while blue lines represent linear trends fitted to the complete 1951-2020 period and orange lines indicate linear trends fitted separately to the periods before and after the measurement-system change on 1 January 1993 (green vertical line). The visual comparison of these trends illustrates how measurement inhomogeneities can influence the magnitude and even the direction of estimated long-term trends. Dark red dashed horizontal lines denote the median AMPTs before and after the measurement-system change, highlighting a pronounced shift in the 5-min series, whereas corresponding changes are considerably smaller for the 1-h and 1-day durations.

- **Referee:**

Multiple testing and field significance
The manuscript performs many station-based trend tests across different durations and time periods. This raises the issue of multiple testing. Fractions of statistically significant stations close to 5% can arise by chance, particularly when spatial dependence between stations is present.

I recommend adding a false discovery rate correction, field-significance test, or spatial bootstrap/permutation approach. This would make the interpretation of significant trends more robust and would help distinguish spatially coherent signals from results that may occur by chance.

Response:

We fully agree that the issue of multiple testing should be addressed to avoid overestimating the statistical significance of the station-based trend results. Therefore, and also following similar comments from the other referees, we propose to apply a false discovery rate (FDR) correction based on the Benjamini–Hochberg procedure to the set of p-values obtained from the Mann–Kendall trend tests. Following Wilks (2016 – https://doi.org/10.1175/BAMS-D-15-00267.1), 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.

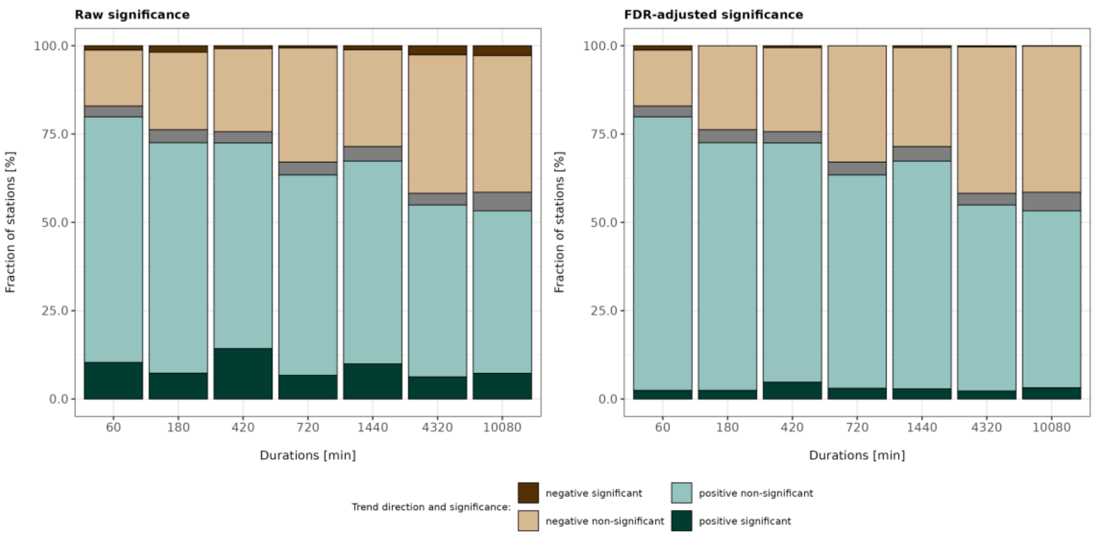

Figure 3: Fraction of stations [%] exhibiting different trend directions and significance in AMPTs (same as Fig. 6 in the manuscript), shown here for both raw (unadjusted) and false discovery rate (FDR)-adjusted results. Very short durations (5 and 30 minutes) are omitted following the referees' comments due to the increased uncertainty associated with data inhomogeneities.

Moreover, we agree that field-significance tests or spatial bootstrap/permutation approaches could provide a further, more in-depth perspective, particularly with respect to spatial dependence between stations, and we will explore whether these approaches can be implemented in the revised manuscript. However, given that the FDR adjustment already leads to a substantial reduction in the number of statistically significant trends, we expect that the application of these more advanced approaches would not materially alter the main conclusions of the manuscript, namely that statistically significant trends are generally very rare.

- **Referee:**

Interpretation of physical processes
The process interpretation is plausible but sometimes appears stronger than what is directly demonstrated by the analysis. The manuscript interprets short-duration events as mainly convective and long-duration events as being shaped by large-scale processes such as frontal systems, slow-moving lows, and Vb cyclones.

This interpretation is physically reasonable and supported by previous work, but the present analysis mainly shows spatial distributions, coefficients of variation, and trends. To strengthen the process-based interpretation, the authors could add a seasonal decomposition, event-month statistics, or a brief circulation-regime analysis. Alternatively, the wording should be softened so that these explanations are clearly presented as plausible interpretations rather than direct results of the analysis.

Response:

We agree that the current manuscript may, in some places, present the interpretation of the underlying physical processes more strongly than is directly supported by the analyses presented. 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 some preliminary results is provided in the figure below – 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 the interpretation remains appropriately cautious and does not extend beyond what is directly supported by the analyses presented.

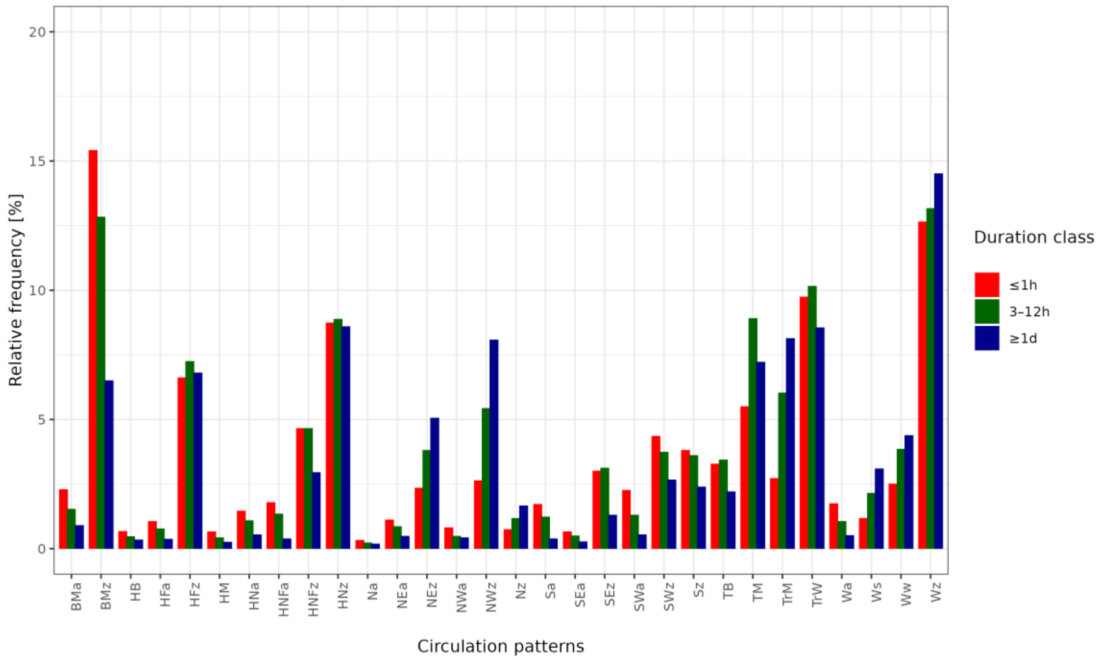

Figure 4: Relative frequency of circulation patterns [%] associated with AMPTs of different durations. For brevity, the original duration classes are grouped into three categories, i.e. ≤ 1 h, 3–12 h, and ≥ 1 day. Circulation patterns are defined using the operational Grosswetterlagen (GWL) classification 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 & Ostermöller (2026, under review).

- **Referee:**

Study period and exclusion of recent years
At line 114, please clarify why data from 2021 to the present are excluded. Since the paper is based on a recently digitised data product, readers will likely wonder why the analysis ends in 2020 and whether more recent years were unavailable, not yet quality controlled, or intentionally excluded for methodological reasons.

Response:

We thank the referee for highlighting this point. The analysis is restricted to the period ending in 2020 because, at the time the dataset for this study was compiled, the corresponding 5-minute precipitation data from the cooperating data providers were only available up to this year. The collection and integration of data for the period from 2021 onward are still in progress and these data were therefore not yet available for the analyses presented in this manuscript. We will clarify this explicitly in Section 2.1 of the revised manuscript.

- **Referee:**

  Title
  The authors may consider simplifying the title by omitting "using the block maxima method." The current title is informative, but somewhat long. Since the methodological approach is explained in the manuscript, the title could focus more on the main scientific contribution: the spatial and temporal variability of heavy precipitation events across durations in Germany.

  Response:

  We agree that the reference to the block maxima method is not essential in the title and that a shorter title places greater emphasis on the main scientific contribution of the study. Therefore, and also following a similar suggestion from another referee, we will revise the title accordingly: "Heavy Precipitation Events of Various Durations Across Germany: A Station-Based Assessment of Spatial and Temporal Variability."

- **Referee:**

  Lines 74 and 77: "Fowel et al. (2021)" should likely be corrected to "Fowler et al. (2021)."

  Response:

  We thank the referee for pointing out these typographical errors. The spelling will be corrected to "Fowler" throughout the revised manuscript.

- **Referee:**

  Line 114: Please explain why the period from 2021 onward is excluded.

  Response:

  As indicated above, the analysis is restricted to the period ending in 2020 because the collection and integration of 5-minute precipitation data for the period from 2021 onward are still in progress and these data were therefore not yet available for the analyses presented in this manuscript. We will clarify this explicitly in Section 2.1 of the revised manuscript.

- **Referee:**

  Figure 7: The duration of 420 minutes appears unusual before approximately 1950. Please check whether this reflects the data, the station network, or a plotting/processing artefact.

  Response:

  We thank the referee for this insightful comment. We have investigated the behaviour observed for the duration of 420 min and confirmed that it results from the very small number of stations available for the pre-1951 time periods, which leads to discrete values (e.g. 0%, 50%, or 100%). To avoid such misleading effects, we will revise Fig. 7 by introducing a minimum threshold for the number of stations required for plotting and will exclude periods with insufficient station coverage.

- **Referee:**

  Line 445: Consider changing "early warning system" to the plural form "early warning systems."

  Response:

  We thank the referee for this suggestion. The wording will be corrected accordingly.

- **Referee:**

Please carefully proofread the manuscript for minor typographical and grammatical issues.

Response:

We will carefully proofread the revised manuscript and correct any remaining typographical and grammatical inconsistencies.