

**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:**

*This manuscript addresses an important and timely topic: the spatial and temporal variability of heavy precipitation events across Germany using recently digitized and quality-controlled high-resolution rain-gauge observations. The main strength of the study is the use of 5-minute station data across various accumulation durations, which allows a systematic comparison from short-duration to multi-day events. The descriptive results are generally plausible, especially the contrast between relatively uniform short-duration maxima and increasingly topography-influenced long-duration maxima. The manuscript also offers an overview of how trend behaviour changes with duration.*

*At the same time, the paper is not yet ready for publication in its present form. The most important weaknesses concern the treatment of short-duration inhomogeneities, the lack of a clearly separated discussion section, and several methodological and presentation issues that currently limit the robustness of the conclusions.*

*[...] This is a relevant manuscript built on a valuable dataset, and it has the potential to become a useful contribution to the literature on heavy precipitation in Germany. However, the current version requires a very substantial revision before publication. I therefore recommend rejection of the current manuscript, but due to the potential of the new data set, I would like to encourage the authors for resubmission after substantial revisions.*

**Response:**

We thank the referee for the careful assessment of our manuscript and sincerely appreciate the time and effort invested in providing detailed comments. We also appreciate the referee’s recognition of the relevance of the topic and the value of the underlying high-resolution dataset. At the same time, we acknowledge the concerns raised regarding the treatment of short-duration inhomogeneities, the lack of a clearly separated discussion section, as well as some methodological and presentation issues.

In the revised manuscript, the former issue will be addressed by relocating and expanding 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, given their limited reliability. In addition, we will provide a new clearly separated Discussion section, which 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. Following the detailed comments provided, we will also carefully revise all methodological and presentation concerns raised by the referee to improve the overall clarity and robustness of the manuscript.

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

- **Referee:**

*Treatment of short-duration trends*

*The manuscript shows that very short durations, especially 5-minute annual maxima, are strongly affected by inhomogeneities associated with the transition from analogue to digital instruments. Although the authors state that formal inference should therefore be restricted to durations of one hour or longer, the 5- and 30-minute results remain visible in the main trend figures and continue to shape the overall impression of the results. This should be handled more consistently. Either these durations should be removed from the main trend analysis and moved to supplementary material, or their status should be stated much more clearly as diagnostic rather than inferential.*

**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 very short durations, particularly 5-minute AMPTs, are strongly affected by inhomogeneities related to changes in the measurement system

(i.e. sensor transitions), and should therefore 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 a consistent treatment of very short durations and allows the manuscript to focus on robust and interpretable results, while transparently documenting the limitations of the underlying data.

- **Referee:**

*Missing discussion section & other structural comments*

*The final section currently reads more like a combined discussion and conclusion than a true conclusion. A separate discussion section is needed. This would allow the authors to distinguish more clearly between results directly supported by the analyses and broader physical interpretation, including the proposed contrast between convective controls for short-duration events and large-scale circulation controls for long-duration events.*

**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. We will also expand the discussion by drawing on a broader set of relevant studies analysing the contrast between convective and large-scale controls on heavy precipitation events of different durations (e.g. Lengfeld et al. 2021 – <https://doi.org/10.1127/metz/2021/1088> or Rulfová & Kyselý 2014 – <https://doi.org/10.1155/2014/647938>), which will allow us to better frame and support the physical interpretation of our results. The Conclusions section will be shortened to provide a clear and concise summary of the key results.

- **Referee:**

*Figures: To my opinion, Figures 1, 2, and 3 can be removed or put to the supplementary information material.*

**Response:**

We agree that the number and selection of figures can be carefully revised and reduced to ensure a clear and focused presentation of the most important results. In the revised manuscript, we will reassess the necessity and placement of Figs. 1-3 and evaluate whether parts of these figures can be moved to the Appendix or Supplementary Material, where appropriate.

- **Referee:**

*Methodological issues*

*AMAX: The paper correctly describes the pros and cons of AMAX. Due to limited data availability (which is highlighted also as a major shortcoming), the choice of AMAX compared to other approaches such as the POT is not fully understandable. This might heavily impact on the robustness of the extreme value estimation. This requires at least some discussion.*

*The Mann-Kendall test and Sen's slope estimator are appropriate and widely used, but the manuscript should discuss the implications of performing many station-wise significance tests across multiple durations and periods. The issue of multiple testing is currently not addressed. Even if no formal correction is applied, this limitation should be acknowledged explicitly so that local significance patterns are not overinterpreted.*

**Response:**

We thank the referee for this important comment regarding the methodological aspects of the study.

- **AMAX vs. POT:** We agree that the choice between the annual maxima (AMAX) and peaks-over-threshold (POT) approaches requires clearer justification and will expand this discussion in Section 2.2. In the present study, the use of AMAX is primarily motivated by the structure and availability of the underlying dataset. In particular, the 5-minute precipitation data exhibit substantial gaps during the winter season, either because certain measurement systems were not operated year-round or because historical data have not yet been fully digitised. While both approaches have limitations under incomplete data conditions, we argue that, in the presence of substantial temporal gaps – particularly when entire seasons are missing – a POT approach is more strongly affected, as, firstly, the threshold value itself may be biased due to incomplete sampling and, secondly, threshold exceedances may be systematically missed. In contrast, the AMAX approach, which relies on identifying a single maximum per year, is more likely to capture the true annual maximum when data coverage is reasonably complete (i.e. meets the applied data availability criteria), although it may still underestimate extremes in years with substantial missing data. Furthermore, the use of AMAX ensures consistency with established approaches applied in operational contexts such as KOSTRA-DWD-2020 and enables a systematic comparison across a wide range of durations. Nevertheless, we acknowledge that POT approaches may offer advantages for more complete and homogeneous datasets and will therefore consider their application in future work as data availability improves.
- **Mann-Kendall and multiple testing:** We agree that the implications of performing many station-wise significance tests across multiple durations and periods should be addressed explicitly. 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 – <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. 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.

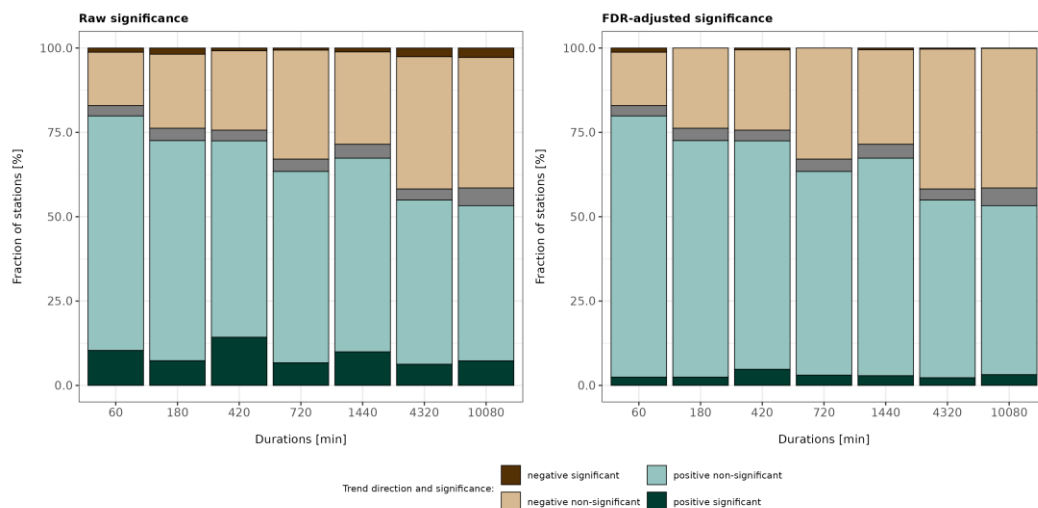


Figure 1: Fraction of stations [%] exhibiting different trend directions and significance in AMPTs (same as Fig. 6 in the manuscript), but shown here for both raw (unadjusted) and false discovery rate (FDR)-adjusted results. Very short durations (5 and 30 minutes) are omitted following reviewer comments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

- **Referee:**

*Interpretation of long-duration trends*

*The more balanced or partly negative trends for 3- to 7-day events are potentially interesting, but the interpretation remains limited. Annual maxima may merge physically different events occurring in different seasons and regions. Opposing seasonal tendencies could therefore mask clearer signals in annual statistics. This point should at least be discussed, and ideally explored more directly.*

**Response:**

We thank the referee for this insightful comment. We agree that the interpretation of trends based on annual maxima is limited by the fact that physically different types of events occurring in different seasons may be combined, potentially masking clearer seasonal signals. A more detailed seasonal analysis would therefore be valuable. 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. At the same time, we are currently assessing whether, at least for longer durations, such an analysis can be performed using daily-resolution data, where data availability is more complete. 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:**

*Level of scientific novelty*

*The strongest potential of the paper lies in the data basis and the systematic duration-dependent comparison across Germany. That is valuable! However, beyond the descriptive results, the scientific novelty remains moderate unless the interpretation is sharpened or additional analyses are included. The main message that short-duration precipitation extremes do not follow the topography is not new and already published by colleagues from DWD. Thus, the paper would benefit from either a more explicit positioning as a benchmark descriptive study or a deeper analysis of the proposed physical drivers.*

**Response:**

We acknowledge this comment regarding the level of scientific novelty and agree that the manuscript would benefit from a stronger treatment of the underlying physical drivers. 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.

At the same time, we will more clearly position the study within the existing literature and emphasise that our results provide a station-based confirmation of previously reported radar-based findings, particularly the reduced influence of topography on short-duration precipitation extremes. To our knowledge, such a systematic, duration-dependent assessment based on long-term high-resolution station data has not been available previously due to data limitations.

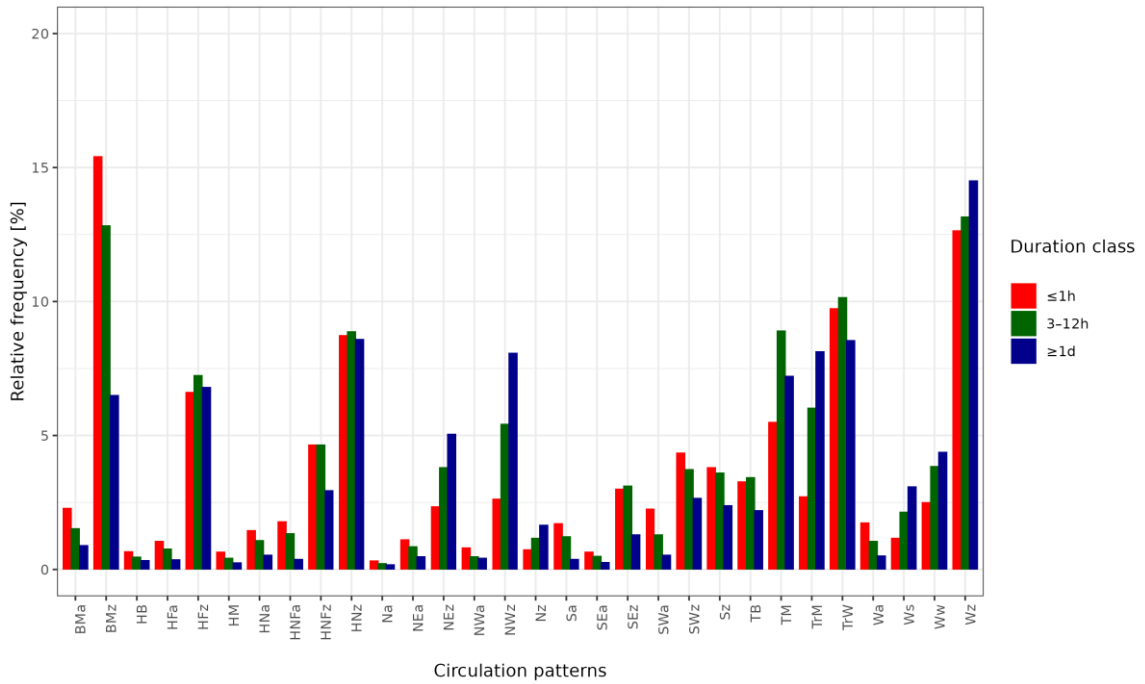


Figure 2: Relative frequency of circulation patterns [%] associated with AMPTs of different durations. For brevity, the original duration classes are grouped into three categories, i.e.  $\leq 1$  h, 3–12 h, and  $\geq 1$  day. Circulation patterns are defined using the operational Grosswetterlagen (GWL) 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:**

The abstract should state the trend-analysis periods more clearly and directly acknowledge the limitations of very short durations.

**Response:**

We agree that the temporal framework of the trend analyses should be stated more clearly in the abstract and that the limitations associated with very short durations should be explicitly acknowledged. In the revised manuscript, we will therefore revise the corresponding sentences to clearly specify the analysis period (e.g. “The duration-dependent nature of AMPTs is further reflected in their long-term variability, which is assessed using station records of at least 30 years, with the end year fixed at 2020 and start years varying according to data availability at individual stations”). In addition, we will explicitly note that results for very short durations are subject to higher uncertainty due to data inhomogeneities and should therefore be interpreted with caution.

- **Referee:**

Several figures need improvement in readability, especially Figure 3 and some of the multi-panel maps.

**Response:**

We agree that the readability of several figures, particularly Fig. 3 and some of the multi-panel maps, can be improved. In the revised manuscript, we will carefully revise their design, improve their readability, and adjust the figure captions to enhance clarity.

- **Referee:**

*The manuscript should also report the effective station sample size more clearly for the varying start-year analyses, and the reference list should be checked carefully for completeness and consistency.*

**Response:**

We agree that the effective station sample size should be reported more clearly for the varying start-year analyses, in particular in Fig. 7. In the revised manuscript, we will therefore provide additional information on the temporal evolution of the station sample size, both in the text and in the corresponding figures. To avoid misleading interpretations associated with small sample sizes, we will also revise the figures by introducing a minimum threshold for the number of stations required for plotting and exclude periods with insufficient station coverage. In addition, we will carefully review the reference list to ensure completeness and consistency.

- **Referee:**

*Introduction: few times “Fowel et al.F” is cited, but should be “Fowler”*

**Response:**

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