https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

From manual classification to large language models: assessing the quality and consistency of historical convective event records

Franck Schätz[1] and Rüdiger Glaser[1]

[1]University of Freiburg, Institute of Environmental Social Sciences and Geography, Chair of Physical Geography, Stefan-Meier-Strasse 76, 79104 Freiburg, Germany

Correspondence: Franck Schätz (franck.schaetz@geographie.uni-freiburg.de)

Abstract. Historical text sources represent a central, yet methodologically challenging basis for the reconstruction of convective weather events. This study examines the extent to which historical reports on thunderstorms and hailstorms contain reliable climatological information, despite heterogeneous sources, varying degrees of detail and linguistic diversity. Based on a corpus prepared using source criticism, qualitative descriptions are converted into structured evidence levels and intensity classes and analysed using statistical methods and a multilingual BERT language model.

The reconstructed time series show a distinctly stable seasonal signal with a dominant summer maximum that occurs independently of fluctuations in source density and is consistent both in the overall series and in a dense observation window. A comparison with modern observation data from the German Weather Service and with independent historical reconstructions shows a high degree of agreement in seasonal patterns despite different survey methods and time periods. Analysis of the intensity classes also shows that historical sources do not primarily document extreme events, but rather reflect a physically plausible ranking of event strengths.

The results of the automated classification prove that the language model reliably reproduces seasonal and intensity-related patterns and implicitly captures source-specific reporting patterns without levelling them. Overall, the study shows that AI-supported methods can extract robust climatological information from historical texts when processed using rigorous methods, thus opening up new perspectives for quantitative historical climate research.

1 Introduction

Written observations provide a wealth of direct information on weather phenomena such as thunderstorms and hailstorms, with the sequence, intensity, spatial extent and effects of individual events often documented in detail. They are indispensable not only for reconstructing past weather and extreme events, but also for addressing current issues, as modern measurement networks and radar data series do not go back far enough in time (Kahraman et al., 2024; Martius et al., 2018) to reliably capture the variability and long-term trends of convective processes (e.g., Hawkins et al., 2023; Punge and Kunz, 2016; Taszarek et al., 2019; Luterbacher et al., 2024). Historical observations also provide information on the seasonal distribution, frequency and intensity of documented events and their consequences (Brázdil et al., 2016b, a; Diodato et al., 2019; Giordani et al., 2024; Hulton and Schultz, 2024) and thus form a central basis for extreme weather analyses and applications in risk management

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

(Brönnimann et al., 2019; Hawkins et al., 2023). Against this background, texts represent an essential source of data, as they enable consistent classification of mesoscale event dynamics over long periods of time (Brönnimann et al., 2019; Cutter, 2021; Erfurt et al., 2020; Glaser, 2013; Stahl et al., 2016).

Despite this potential, historical texts have only been used to a limited extent so far, as climate-related information must be identified and evaluated manually for the most part. Each event must be recognised individually, contextualised and classified in terms of its meteorological significance. With large text corpora, there is therefore not a lack of sources, but a lack of processing capacity. Modern developments in the field of natural language processing and, in particular, large language models such as BERT (Devlin et al., 2019) are opening up the possibility for the first time of systematic and scalable thematic analysis of historical texts on climate, weather and risk events (e.g., Sakaji and Kaneda, 2023; Webersinke et al., 2022; Zhou et al., 2022).

However, creating such models requires a linguistically and climatologically validated data basis. Historical climate records are linguistically heterogeneous, idiomatic and semantically strongly tied to the time of their origin (Schätz, 2023; Glaser, 2013). They therefore differ fundamentally from modern, scientific weather reports (Grzega, 2022). Furthermore, they are not standardised measurement data, but subjective observational descriptions that vary greatly in terms of detail, temporal and spatial precision, and meteorological interpretation (Brönnimann et al., 2019; Glaser, 2013; Brázdil et al., 2010). Without quality control, there is a risk that a language model will learn and reproduce stylised or traditional forms of description instead of actual weather events. Therefore, a procedure is required that ensures both linguistic and climatological plausibility before such data can be used for fine-tuning a large language model.

There is a particular need for research into thunderstorms and hailstorms: previous work has mainly recorded historical events using simple classifications (such as frequency, intensity or damage), but a formally defined and reproducible classification scheme for text sources is still lacking (cf., Brázdil et al., 2016a, b; Camuffo et al., 2000; Gudd, 2004; Lenke, 1960; Rohr, 2009; Huang et al., 2022). This means that there is no approach that systematically takes into account both the linguistic variability of historical descriptions and meteorological interpretation.

The aim of this study is to develop a scalable method for quality-assured, automated classification of historical thunderstorm and hail events. To this end, a formally defined classification procedure is being developed to serve as the basis for an AI-supported classification procedure. Using the period 1000-1817 as an example, historical event descriptions will first be systematically classified and statistically validated. The text data validated in this way will form the basis for fine-tuning a BERT model, which will then be used for the automated identification and classification of further event descriptions. The study thus demonstrates the potential of modern large language models for the structured evaluation of historical text sources without undermining established scientific standards of climatological data quality. To promote follow-up research, all models are made publicly available via Hugging Face.

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

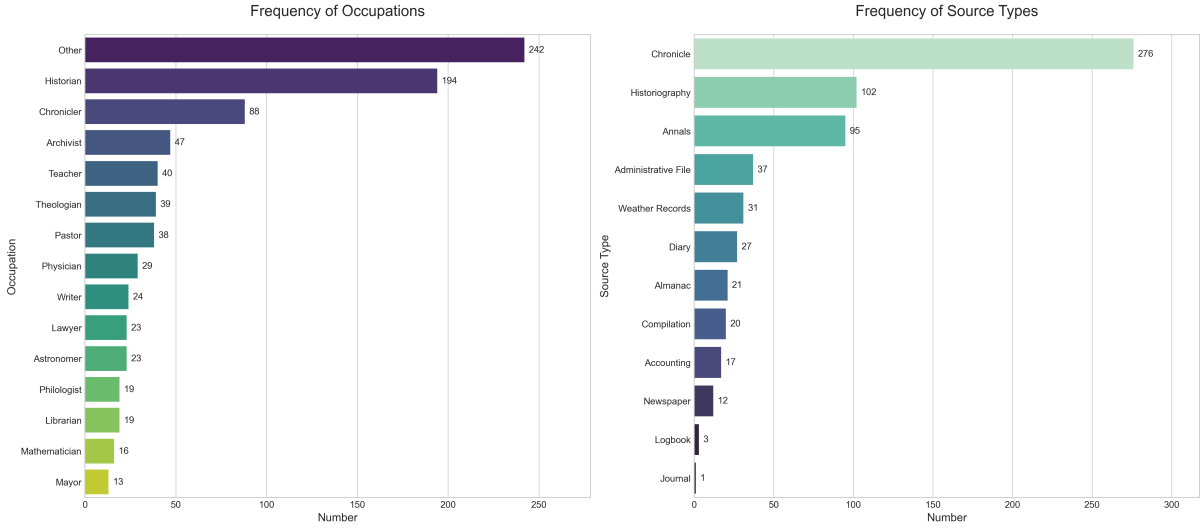

Figure 1. Occupational groups of authors and document types.

2 Data basis

The analyses are based on a systematically structured data set containing metadata on the sources and characteristics from text excerpts (quotes) relating to thunderstorms and hail events. This is based on the HISKLID data collection (Glaser, 2014)), which has been integrated into the virtual research environment tambora.org[1]. The data set comprises 7002 entries with 6678 quotations from 494 sources for Central Europe in the period from 1000 to 1817. The citations are mainly taken from chronicles, historiographies, annals and administrative records (see Figure 1). They document 6155 thunderstorms and 2011 hailstorms and are presented in tabular form. Each row represents a single citation with the associated characteristics and event parameters (cf., Schätz and Glaser, 2025)). The characteristic values are collected and enriched in 21 steps from the source references and citations (Schätz, 2023, cf.,).

In addition to bibliographic details, the source references also contain biographical information about the authors, including their place of work and professional activities. In addition to primary sources, secondary sources are also included, which means that there can sometimes be considerable time gaps between observation and publication (see Figure 2). The increasing number of sources and recorded observations reflects technological and social developments, such as the introduction of printing around 1450, the Reformation from 1517 onwards, the Thirty Years' War from 1618 to 1648, and the university reforms at the beginning of the 19th century (cf., Ernst, 2021).

Since the text sources cover different stages of the German language (Middle High German, Early New High German, New High German, contemporary German), all quotations were manually normalised in terms of language.

[1] https://www.tambora.org

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

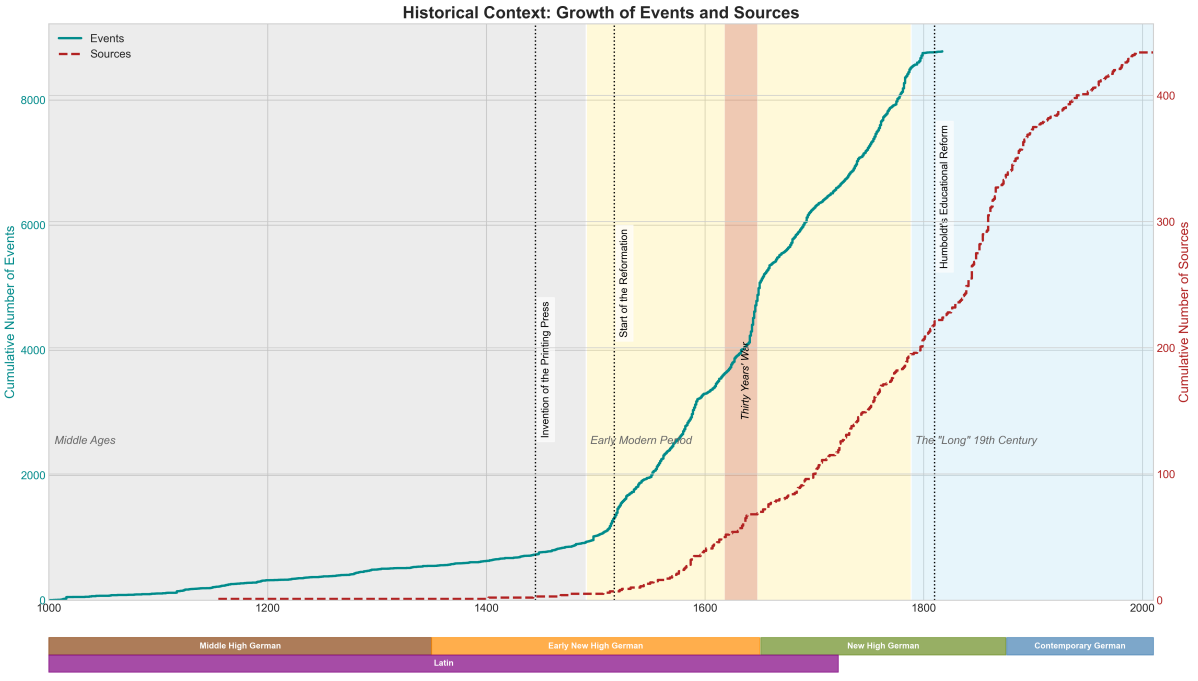

Figure 2. The figure shows the cumulative number of citations relating to thunderstorms and hailstorms, as well as how these are distributed over time according to source. Additionally, historical turning points that influenced the production of text sources are displayed. The stages of the German language – Middle High German, Early New High German, New High German and Contemporary German – are also recorded in accordance with Ernst (2021). Latin texts appear across different eras and end with the 1722 corpus.

This normalisation is a prerequisite for a consistent lexical basis, and therefore for applying modern language models (Ehrmanntraut, 2025). The same applies to the creation of semantic word lists ('Silver Labels') and the coding of linguistic indicators, the quality of which depends directly on a homogeneous linguistic basis.

The spatial distribution of events is concentrated in German-speaking countries, although the location where a source originated does not necessarily correspond to the reported observation location (Figure 4). Newspapers in particular report on events across regions, often beyond territorial and linguistic boundaries. The observations cover the most important landscapes of Central Europe (4). This means that the corresponding climate regions and mesoscale climatic effects such as maritime and continental influences, orography, etc. are represented.

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

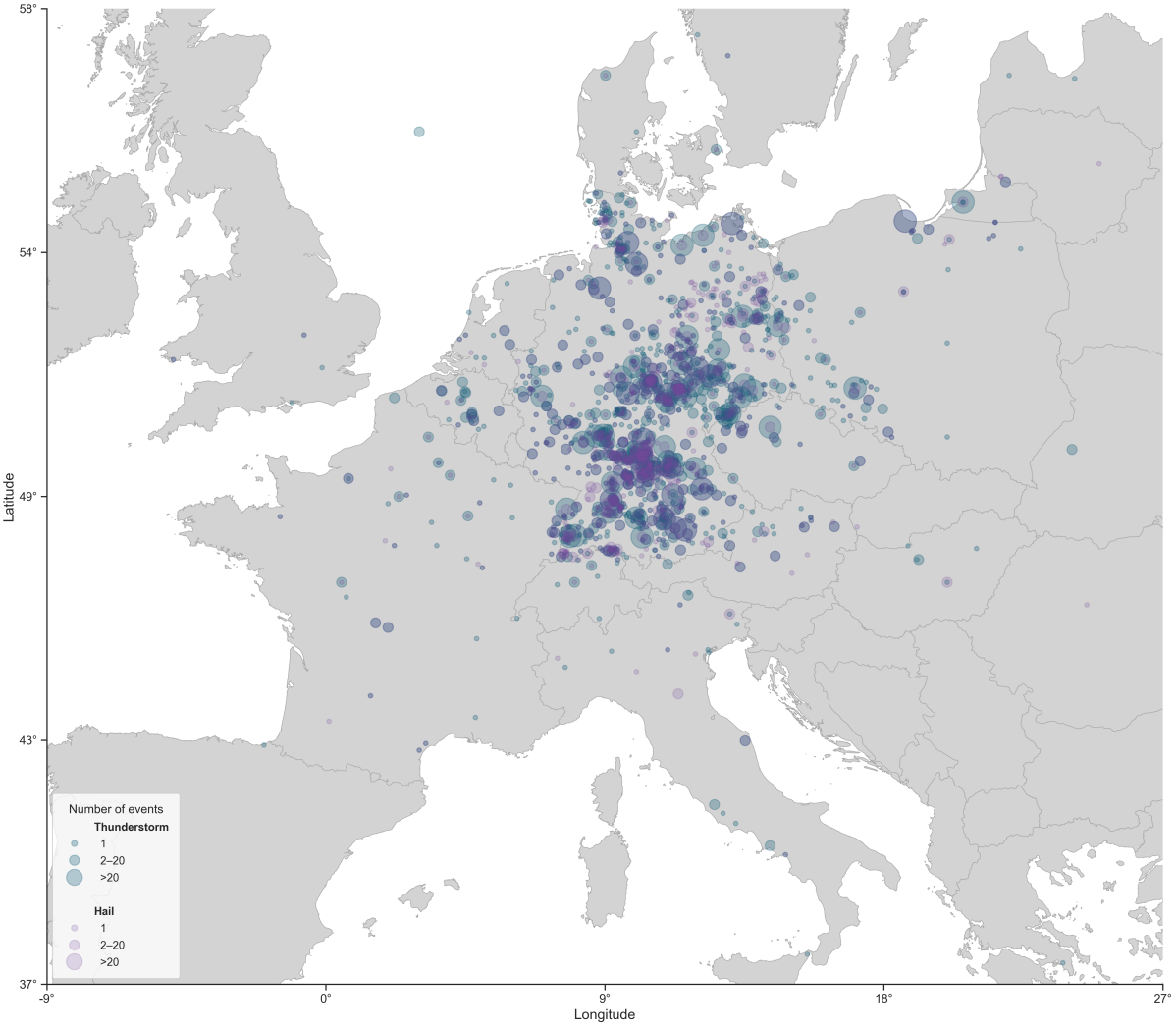

Figure 3. The figure illustrates the spatial distribution of thunderstorms and hailstorms and their number. The events are taken from German-language and Latin-language sources. The place of origin of the source does not always correspond to the place of the reported observation. Newspapers in particular feature supraregional reporting, with events often being communicated across territorial and linguistic boundaries.

3 Methods

This study builds on the approach of Schätz (2023), which describes weather and climatic events using a quadruple consisting of event type (ET), time (T), location (P) and event intensity (EI), where EI is understood as a classification of textual sources. Formally, EI is the result of a mapping f_{EI}, which maps a citation $z \in Z$ from the set of all citations Z to a class $c \in C$ from a finite set of uniquely defined event intensity classes C:

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

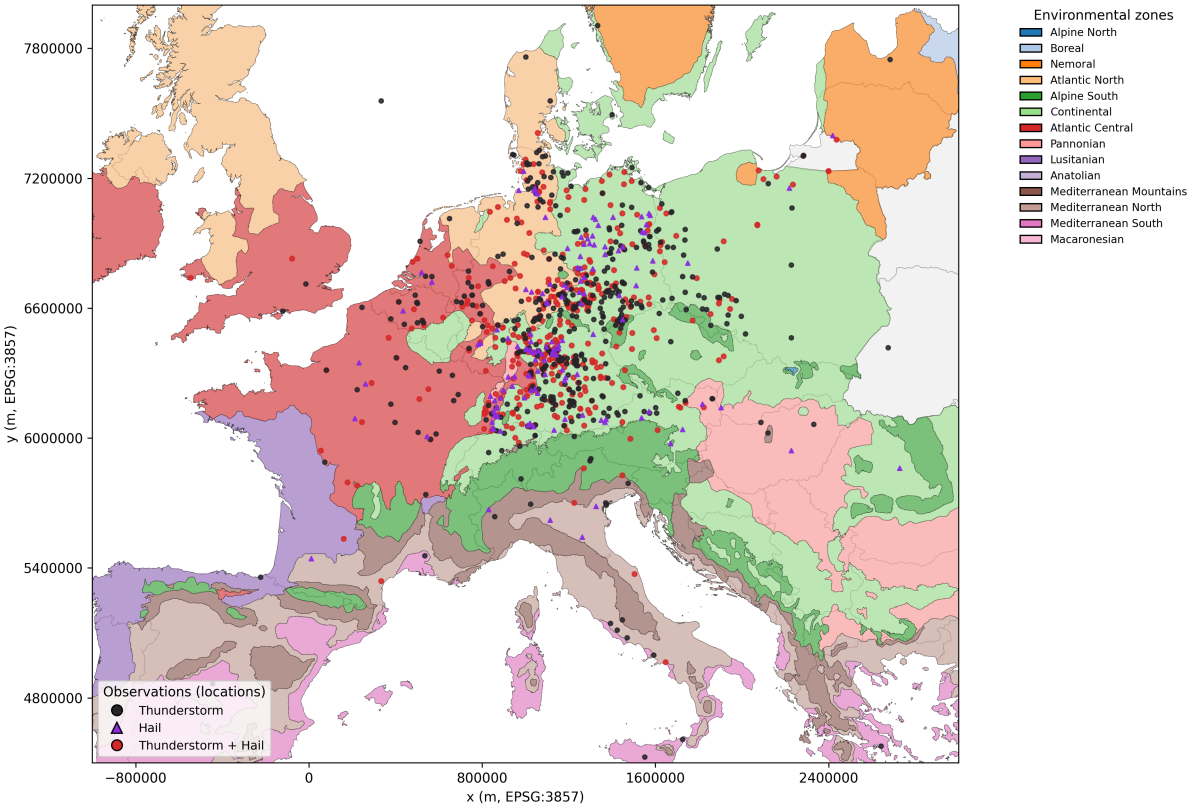

Figure 4. The spatial distribution of thunderstorm and hail frequency in the European context is based on the Environmental Zones of Europe (European Environment Agency, 2020). Fifteen environmental zones are shown (e.g. Alpine North, Continental, Pannonian, Mediterranean North), each representing large-scale ecological and climatic units in Europe. The coloured areas show the extent of these zones, cropped to the coastline of mainland Europe.

$$f_{EI} : Z \rightarrow C, \quad EI = f_{EI}(z) \tag{1}$$

Classification is carried out in a multi-stage process in which linguistic references to accompanying and subsequent phenomena are systematically analysed and recorded at the level of semantically coherent word groups. On this basis, thunderstorm and hail events are divided into intensity classes and then statistically evaluated, including comparison with observation data from the Deutscher Wetterdienst (DWD) for the normal period.

3.1 Classification procedure

The classification scheme consists of five mutually exclusive thunderstorm classes and four hail classes (see Tables A1 and A2). It is based on the classification and warning system of the Deutscher Wetterdienst (DWD) (2025) and Tornado and Storm

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

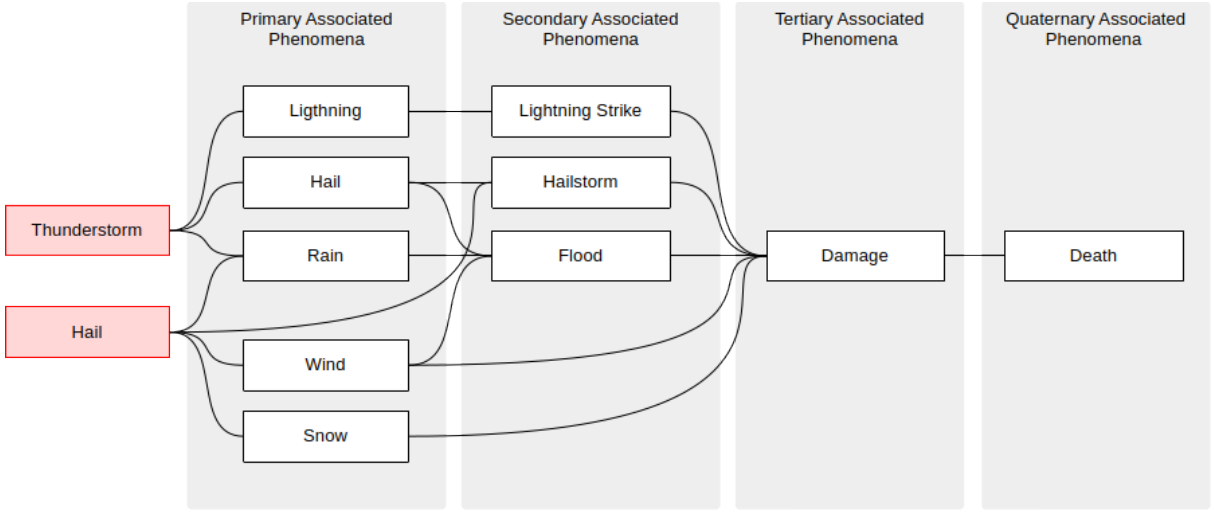

Figure 5. Causal network of thunderstorm-related phenomena. The diagram illustrates the hierarchical relationship between a thunderstorm, hail and its associated phenomena. Primary associated phenomena include rain, hail, lightning, snow, and wind, which represent the direct meteorological manifestations of the storm. These may trigger secondary associated phenomena such as floods, hailstorms, or lightning strikes, leading to tertiary effects in the form of damage. In severe cases, these may escalate to quaternary effects, including fatalities. The classification scheme thus reflects a causal chain from meteorological processes to their environmental and societal impacts.

Research Organisation (TORRO) (2025). The analysis of the citations takes into account the cause-and-effect relationship as explicitly or implicitly formulated in the sources. For the systematic recording of observations, the accompanying phenomena mentioned are classified according to their physical relevance into primary, secondary, tertiary and quaternary characteristics (see Figure 5). This classification serves the purpose of structured source evaluation and forms the basis for the subsequent classification.

Classification is carried out in stages. First, it is checked whether the quotation mentions accompanying phenomena that can be clearly attributed to a thunderstorm or hailstorm. Only if such indications are present is the event considered further. The next step is to classify the intensity of the event based on the criteria defined in tables A3 - A5. The effects described are systematically classified and explicitly included in the assessment.

The classification begins with the highest intensity class (class 2 for hail, class 3 for all other phenomena considered) in order to avoid underestimating particularly severe events. If this class is not met, the lowest intensity class (class 1) is examined and then, if applicable, the middle class. The highest intensity class is assigned exclusively on the basis of explicit damage descriptions. For hail events, qualitative information on grain size is also taken into account (see Table A6).

3.2 Validation through linguistic evidence

To ensure consistent and comprehensible classification of thunderstorm and hail events, class-specific vocabularies are derived from the examined quotations. These contain typical linguistic indicators of the respective accompanying phenomena

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

and are based on frequently used word forms and phrases in the historical sources (e.g. for thunderstorms: "flash'", "thunder"", "storm"). The vocabularies were developed inductively from the source material and checked iteratively. An overview is provided in Table B1 in the appendix.

In addition to the classification of content, the strength of evidence of the linguistic descriptions is recorded in order to transparently evaluate the significance of the underlying text sources (see Table 1). A distinction is made between three classes of evidence. C1 (direct and specific) has the highest reliability, as it provides concrete information, such as hailstone size, wind strength or precipitation levels, which correspond to measurable variables. C2 (indirect) includes descriptions in which damage indicators such as hail, wind or rain damage clearly point to an event. C3 (relative) has the lowest reliability, as purely qualitative adjectives (e.g. 'strong', "violent", 'very large') without reference values allow for a wide range of interpretations.

Table 1. Examples of silver labels for the phenomenon groups rain, hail and wind. The table shows a selection of characteristic linguistic expressions used for the semantic identification and classification of historical storm phenomena. It illustrates how specific word groups and descriptions can typically be assigned to individual phenomena and evidence classes (C1–C3). The complete, rule-based silver label lists, including all recorded variants and regular expression patterns, are documented in the appendix B2, B3 and B4.

Phenomenon	Category	Representative Examples	Class
Hail	Size descriptions	Hazelnut-sized, walnut-sized, hen's egg-sized	C1
	Hail impact	Crops beaten down, vineyards destroyed	C2
	Intensity	Terrible, severe hailstorm	C3
Rain	Precipitation height	Water was foot-deep, one cubit high	C1
	Damage	Bridges broken, cellars flooded	C2
	Intensity	Flood, inundation, deluge, washed away	C3
Wind	Speed	Only indirect descriptions (see Damage/Intensity)	C1
	Damage	Trees uprooted, roofs blown off	C2
	Intensity	Strong wind, hurricane-like, gusts	C3

This evidence-based differentiation supplements the classification with an internal, source-critical assessment of linguistic significance. This increases the traceability of the classifications and explicitly takes into account the interpretative nature of historical text sources.

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

3.3 Validation by statistical-physical evidence

3.3.1 Annual distribution of events

To verify the quality of the historical climatological data sets, we examine whether the historical event corpus exhibits a physically plausible seasonal signal that is consistent with the known seasonality of convective events (summer maximum). This analysis serves to qualitatively validate the corpus with regard to its suitability as a training basis for language model-based classification tasks. Since language models learn statistical regularities from the training data, the presence of a realistic seasonal pattern is a key prerequisite for ensuring that the model reflects real physical relationships and not primarily transmission artefacts.

The analysis is based on monthly aggregated count data of historical thunderstorm and hail events for the period 1000–1817. Since historical sources predominantly document only positive events and since no systematic zero observations are available, a direct reconstruction of the observation density is not possible.

To reduce distortions caused by different source densities, the annual progression is normalised. To do this, the monthly event numbers for each year are normalised to an annual total of 1,

$$p_{m,y} = \frac{E_{m,y}}{\sum_{j=1}^{12} E_{j,y}}, \tag{2}$$

so that each year is described as a distribution over the calendar months.

The final annual curve is derived from the aggregation of the monthly shares, standardised on an annual basis, across all years. In principle, the median provides a robust description of the typical seasonal pattern, as it limits the influence of individual years with a wealth of historical data. However, in the case of highly incomplete time series with predominantly positive event reports, the median leads to months with rarely documented events being systematically underrepresented and the seasonal pattern appearing artificially thinned out.

For this reason, the mean value of the normalised monthly proportions is used as the central estimator for reconstructing the seasonal annual cycle,

$$\bar{p}_m = \mathrm{mean}_y \left(p_{m,y} \right), \tag{3}$$

as it consistently reflects the relative frequency of events across all years and maintains stable seasonal structures even with fragmentary data. Due to the upstream annual normalisation, all years contribute equally to the analysis, so that the mean value is not dominated by years with a high event density, but primarily reflects the shape of the seasonal pattern.

The uncertainty of the reconstructed annual cycle is quantified over the years using bootstrap resampling. 95% confidence intervals are derived from the resulting distributions. Stability is additionally tested for a time window with high source density (1624–1654), which, due to its comparatively complete transmission, allows for independent validation of the seasonal structure.

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

For external validation, the reconstructed historical annual cycle is compared with modern observation data from the DWD for the normal periods 1961–1990 and 1991–2020. The monthly totals of convective events serve as a reference. These are normalised on an annual basis after spatial aggregation, thus ensuring direct comparability of the relative monthly proportions. The agreement of the seasonal patterns is quantified using Spearman's rank correlation (ρ) and the root mean square error (RMSE) of the normalised monthly proportions.

In addition to the standardised annual progression, a summer–winter ratio is calculated for each year. This is based on the monthly aggregated event counts without further normalisation. For each year, the sum of events in the summer months of June to August (JJA) and in the winter months of December to February (DJF) is calculated, with December being assigned to the winter of the following year. The summer–winter ratio is calculated as the quotient of the annual summer and winter event sums. Years without documented winter events are excluded. To improve the comparability of the time series and reduce the influence of extreme quotients, the summer–winter ratio is additionally logarithmised. The resulting logarithmised time series is smoothed by a moving 10-year average.

3.3.2 Class-specific annual distribution

The seasonal distribution of thunderstorm ($TS1$–$TS3$) and hail ($H1$–$H2$) events is analyzed to assess the stability of class-specific seasonal patterns. The aim of this analysis is to examine whether a consistent seasonal hierarchy emerges for the individual classes and whether this order remains stable over the course of the year.

Only events lasting a maximum of 31 days are taken into account. Each event is assigned to all calendar months that it affects ("presence-per-month"). If an event exceeds a monthly limit, it is counted once in each affected month. Events of very long duration that can no longer be clearly assigned to a specific season are excluded.

For each year y, each month m and each class c, the events assigned in this way are aggregated by class. The analysis is based on relative proportions within a month, so that only the class distribution is considered and no assumptions about absolute event frequencies or observation density are required. The monthly class proportion is defined as

$$p_{c,m,y} = \frac{E_{c,m,y}}{\sum_{c'} E_{c',m,y}}, \tag{4}$$

where ($E_{c,m,y}$) denotes the number of events of class c in month m and year y. The summation index c' runs over all classes of the respective event type considered (e.g. $TS1$–$TS3$ or $H1$–$H2$). Thus, $p_{c,m,y}$ describes the proportion of a class in all events reported in a given month of a year.

In order to limit the influence of individual years with high data availability, the monthly class shares are first determined on an annual basis and then aggregated equally across all years. The typical seasonal class structure is estimated as the mean value of the annual class proportions,

$$\bar{p}_{c,m} = \mathrm{mean}_y \left(p_{c,m,y} \right), \tag{5}$$

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

where the upstream normalisation ensures that each year contributes equally to the estimate, regardless of its event density.

The statistical uncertainty of the reconstructed class distributions is quantified over the years using bootstrap resampling. For each bootstrap sample, the monthly class distribution is recalculated; mean values and 95% confidence intervals (2.5th and 97.5th percentiles) are determined from the resulting distributions.

To assess the stability of the seasonal class hierarchy, the proportion of bootstrap samples in which the expected class order is fulfilled is also determined for each month. For thunderstorm events, the strict order

$$\mathrm{TS1} > \mathrm{TS2} > \mathrm{TS3} \tag{6}$$

is used for thunderstorm events and

$$\mathrm{H1} > \mathrm{H2} \tag{7}$$

The resulting proportion (strict support) describes the robustness of the complete class hierarchy with respect to sample variations.

In addition, a weighted support is calculated for thunderstorm events, which distinguishes between fully and partially fulfilled class order. A complete order (TS1 $>$ TS2 $>$ TS3) is weighted higher than a partial order, in which only the dominance of the weakest over the stronger classes is fulfilled ((TS1 $>$ TS2) or (TS2 $>$ TS3)). Weighted support thus allows for a more refined assessment of the seasonal class structure, especially in months with low event density or sparsely populated classes.

3.4 Influence of source type on classification

Potential selection or transmission biases are tested for independence using a chi-square test, which examines the relationship between source type and storm class (TS1–TS3) or hail class (H1–H2). This is based on a contingency table in which the respective number of events is totalled for each combination of source type and class, thereby taking appropriate account of multiple mentions of individual sources.

Under the null hypothesis (H_0), it is assumed that source type and classification are independent of each other. In addition to the chi-square statistic (χ^2), the effect size is calculated according to Cramér's V in order to quantify the strength of the correlation. Standardised residuals are also evaluated in order to identify combinations of source type and class that contribute most strongly to the deviation from independence.

3.5 Fine-tuning the large language model

The validated data sets form the basis for fine-tuning the language model. The pre-trained language model mDeBERTa V3 Base, which is based on multilingual text data and has a high degree of context sensitivity, is used to classify thunderstorm and hail descriptions. The models are adapted to the classification task ($TS0$–$TS3$ and $H0$–$H2$) by means of fine-tuning. The data

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

set is divided into training, validation and test portions in a ratio of 70 : 10 : 20, with the division being stratified in order to maintain the class ratios.

A two-stage hyperparameter search is performed to identify suitable training parameters. In the phase A , an exploratory random search strategy is applied. For this purpose, a combinatorial search space is defined, consisting of learning rates $1 \times 10^{-5}, 2 \times 10^{-5}, 3 \times 10^{-5}$, maximum sequence length $128, 256$, label smoothing factor $0.0, 0.05$, and learning rate scheduler $\mathrm{linear}, \mathrm{cosine}$, resulting in a total of 24 possible hyperparameter combinations.

Twelve combinations are randomly selected from this search space and trained with a fixed seed. This sample corresponds to half of the entire hyperspace and serves as an efficient exploratory coverage without having to fully evaluate all possible configurations. For all training runs, a batch size of 16, a weight decay of 0.01, a warm-up rate of 0.06 and a training duration of five epochs are used. Training is performed with mixed precision (fp16) on GPU hardware. The models are evaluated after each epoch using the validation data, with the best model of a configuration being selected based on the F1 score.

In the phase B, the three hyperparameter configurations with the best ratings in Phase A are selected and retrained with three different random seeds. The aim of this phase is to check the stability and reproducibility of the results against random fluctuations in the training process. For each configuration, the mean value and standard deviation of the F1 score are calculated across the three training runs.

The models from phase B are finally evaluated on the basis of the F1 score and accuracy on the validation and test data set. For qualitative analysis, a confusion matrix is also used to reveal systematic misclassifications between the thunderstorm and hail intensity classes.

4 Results

4.1 Annual distribution

The aggregated historical records of thunderstorms and hail events show a pronounced seasonal cycle in the normalised monthly proportions (Figure 6). For both types of events, the highest proportions occur in the summer months of June to August (JJA). The increase in monthly proportions begins in spring, reaches its maximum in summer and decreases continuously in the autumn months.

During the winter months (DJF), the standardised monthly proportions remain low but consistently show positive values. The transitional seasons are characterised by lower proportions compared to the summer half-year.

The annual cycle calculated for the most densely documented period (1624–1654) follows the same basic structure as the annual cycle for the entire period. The form of the monthly distribution remains the same, while the relative amplitude between the summer and winter months differs from that of the entire period. A comparison with the other years shows a high degree of consistency in the monthly proportions.

For hail events, the highest standardised monthly proportions are also concentrated in the summer months. Compared to thunderstorm events, increased proportions already occur in the late spring months, especially in March and April.

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

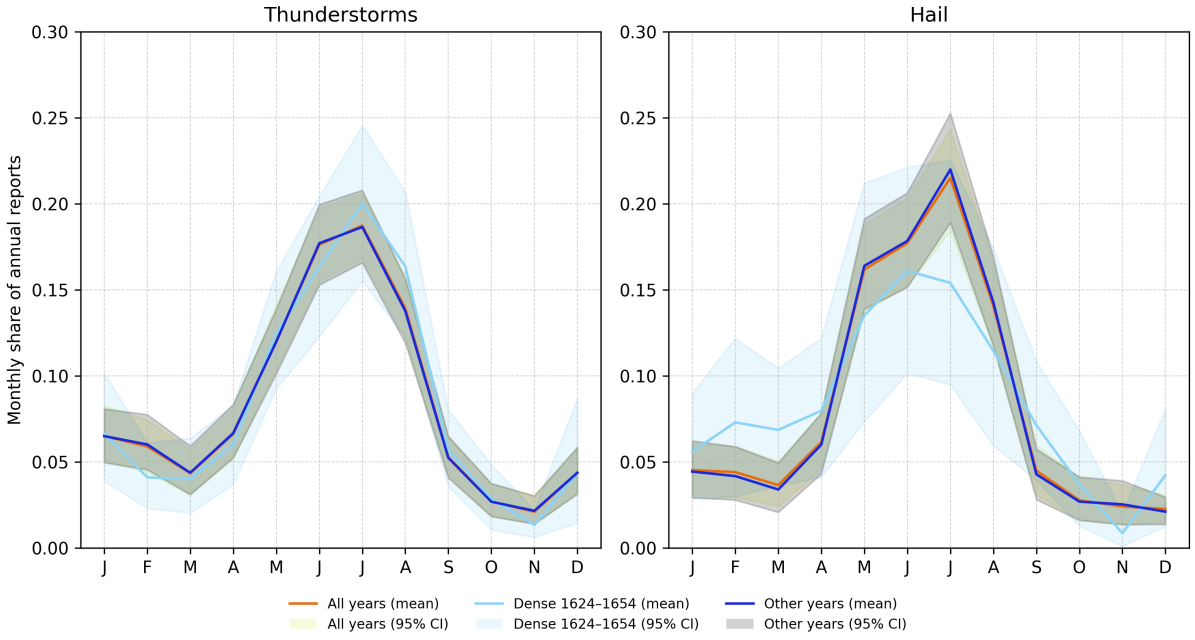

Figure 6. The figure shows the typical seasonal distribution of thunderstorms and hailstorms. The monthly values correspond to the average of the total annual activity (annual total = 1) and thus reflect only the relative frequency within the year.

A comparison with observation data from the DWD for the reference periods 1961–1990 and 1991–2020 (Figure 7) shows a correlation between the seasonal patterns of normalised monthly thunderstorm frequencies. The Spearman rank correlations between the historical data (1624–1654) and the DWD references are between $p = 0.78$ and $p = 0.89$ for the combinations of time period and search radius considered. The mean square errors are $RMSE \approx 0.03$ to 0.04. The results differ only slightly between the selection of DWD stations within a radius of 10 km and 50 km around the historical event locations. The key figures are comparable for both reference periods.

The logarithmic summer–winter ratio (JJA/DJF) shows predominantly positive values for thunderstorm events over the period under investigation (Figure 8), indicating a dominance of the summer months over the winter months. The smoothed time series (10-year average) lies above the zero line for long periods and mostly ranges between approximately 1.0 and 2.0. Elevated values occur particularly in the late 16th and early 17th centuries, while phases with reduced summer dominance are evident in the mid-16th century and early 17th century. Negative individual years occur sporadically, but have only a minor impact on the smoothed curve.

For hail events, the seasonal signal is weaker overall. The logarithmic summer–winter ratio shows greater dispersion and, in the smoothed curve, lies predominantly in the range between approximately 0.3 and 1.0. Values close to or below zero occur at times, especially in the early phase of the time series. From the second half of the 16th century onwards, there is a phase of elevated values, followed by an overall moderate and comparatively stable seasonal ratio until well into the 18th century.

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

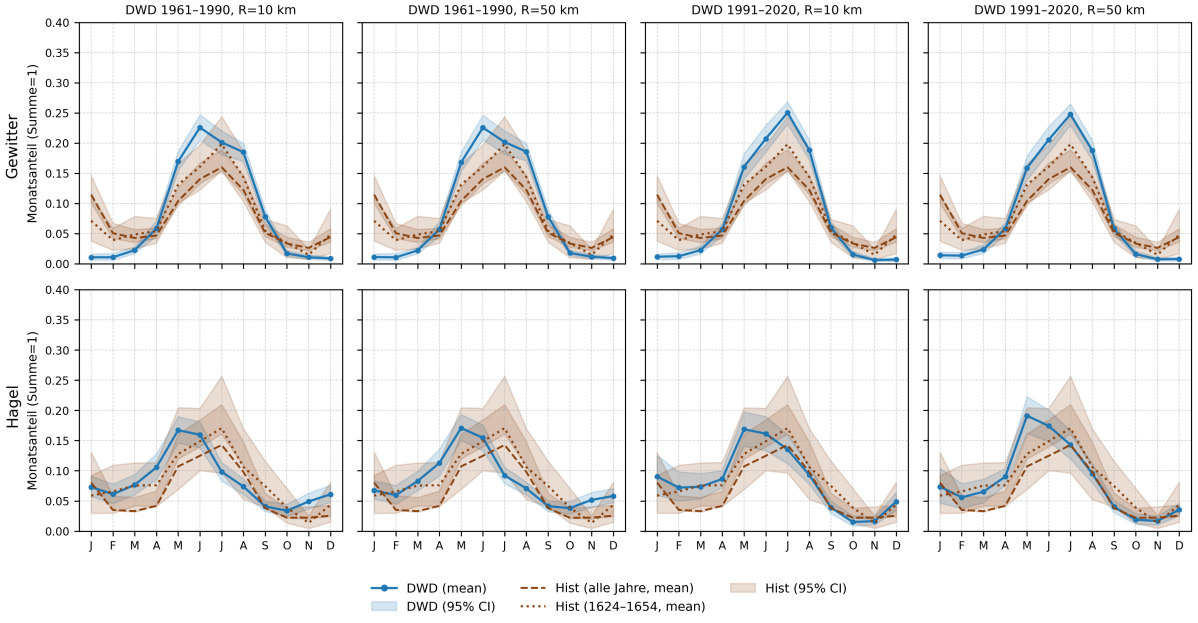

Figure 7. Seasonal trend in normalised monthly frequencies of thunderstorms and hail events from historical observations (1624–1654) compared to observations by the DWD for the periods 1961–1990 and 1991–2020. The DWD stations are selected based on their distance from the historical event locations (circles with a radius of 10 km or 50 km). The values are normalised to the annual total so that the relative proportions of the months are comparable with each other.

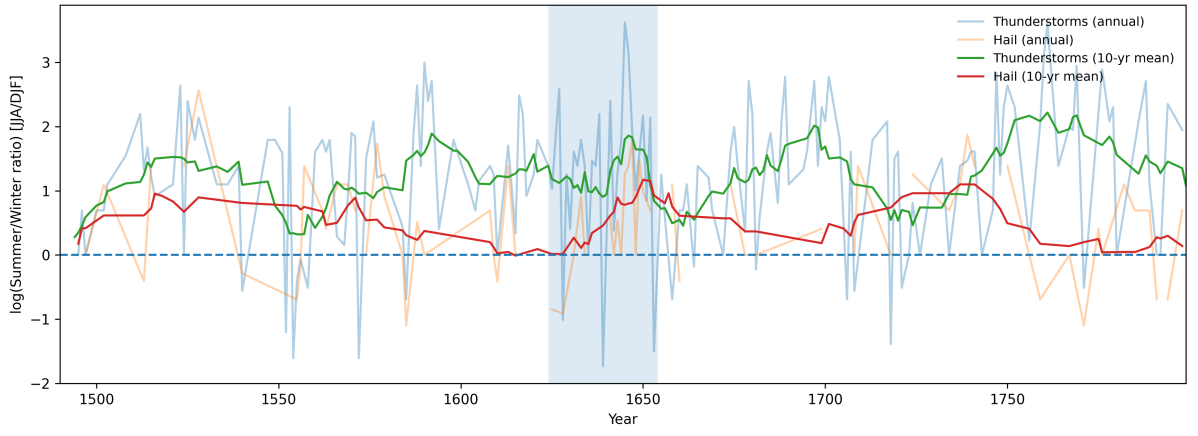

Figure 8. Time series of the logarithmic summer–winter ratio (JJA/DJF) of historical thunderstorm and hail events for the period from 1490 onwards. The dotted line marks a balanced ratio between summer and winter events. The shaded area indicates the source-rich time window 1624–1654.

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

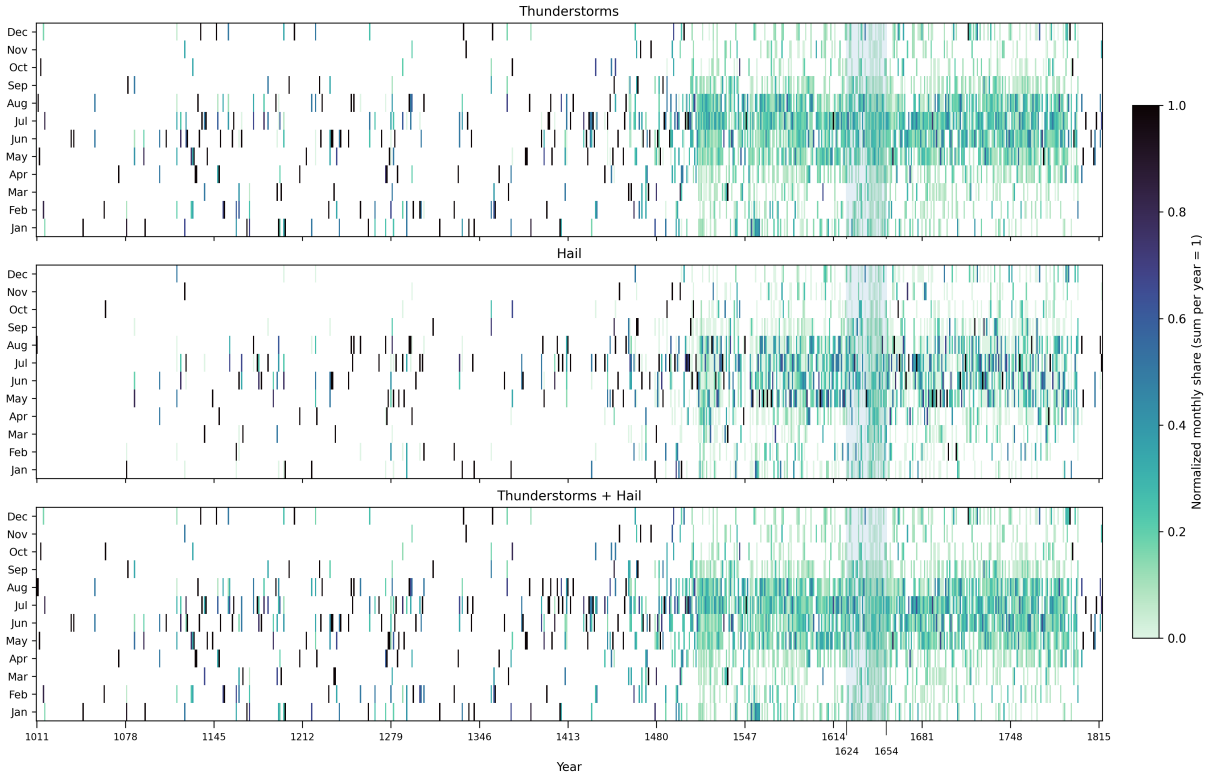

Figure 9. Calendar heat maps of the annualised monthly proportions of thunderstorm and hail events for the entire study period. The relative monthly proportions are shown, with the number of events for each year normalised to an annual total of 1 so that the seasonal distribution can be compared independently of the absolute frequency of events. Separate representations are shown for thunderstorms, hail and the combination of both event types. The colour intensity corresponds to the normalised monthly proportion. The shaded area indicates the period with particularly high source density (1624–1654).

The strength of the seasonal signal varies depending on the number of documented events (Figure 9). When looking at individual event types separately, the monthly time series are characterised by numerous months without documented events. Increased monthly values occur sporadically and are often limited to individual years. In these cases, the seasonal distribution is only recognisable to a limited extent.

When thunderstorms and hail events are evaluated together, the number of events recorded each month increases. The monthly series derived from this shows a clear difference between summer and winter months, with higher proportions in the months of June to August and very low values in the winter half-year. This pattern is consistent throughout much of the period under review.

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

4.2 Annual distribution of thunderstorm and hail classes

The weakest thunderstorm class, $TS1$, dominates in all months throughout the entire study period. The average monthly proportions are predominantly between approximately 55% and 70%. Class $TS2$ contributes approximately 11–34% depending on the month, while the strongest class, $TS3$, reaches proportions between around 10% and just under 30%. Increased proportions of $TS3$ occur particularly in the summer months, with a maximum in the months of June to August. In the winter and transitional months, the ranking $TS1 > TS2 > TS3$ applies predominantly, while in the summer months, constellations in which TS3 achieves higher proportions than $TS2$ occur more frequently.

The bootstrap-based 95% confidence intervals for the monthly class proportions range from approximately $\pm$ 5–10%, with slightly wider intervals in the summer months. Strict rank support ($TS1 > TS2 > TS3$) is very high in the winter months, remaining above 95% almost throughout. In the summer months, this value drops significantly and sometimes reaches very low values, indicating frequent breaches of the strict class hierarchy. The weighted support, which also takes into account cases with $TS1 > TS3 > TS2$, remains significantly higher across all months and is predominantly in the range of approximately 70–85%, even in summer.

For the period 1624–1654, the seasonal structure of the storm classes is similar overall. $TS1$ remains the dominant class in all months, with average shares between approximately 50% and over 75%. Classes $TS2$ and $TS3$ show stronger monthly fluctuations than in the overall data set, especially in the winter months. The 95% bootstrap intervals are partially wider in this period, which is due to the smaller number of years. Strict rank support is significantly reduced in the summer half-year and falls below 20% in individual months, while weighted support is also predominantly in the range of about 75–90% here.

Hail events also show a pronounced seasonal pattern in terms of class distribution. Over the entire period, weaker hail events (H1) dominate in the winter and transitional months, accounting for more than 65–85 per cent of the total. In the summer months, the distribution shifts significantly in favour of stronger events $H2$, whose proportions reach values of around 75–85% in the months of June to August. The 95% bootstrap intervals are mostly in the range of $\pm$ 5–10%.

In the period 1624–1654, the basic ranking $H1 > H2$ remains unchanged in the winter months, while high proportions of class H2 occur in the summer months. The strict ranking support ($H1 > H2$) is very high in the winter months, usually above 95%, while it drops significantly in summer and almost disappears in some months. This reflects the strong seasonal shift in class proportions.

4.3 Influence of source types on classes

The chi-square test for independence shows a highly significant dependence between source type and intensity class for thunderstorm events ($\chi^2 = 529.2$, df $= 22$, $p < 0.001$). The effect size is in the medium range (Cramér's V = 0.21).

The standardised residuals show that weather records in particular are characterised by an overrepresentation of class $TS1$ ($R = +9.0$) and an underrepresentation of class $TS3$ ($R = -11.6$). In chronicles, the strongest class TS3 is overrepresented ($R = +7.6$), while $TS1$ is underrepresented ($R = -5.6$). Almanacs show a clear underrepresentation of $TS3$ ($R = -6.2$).

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

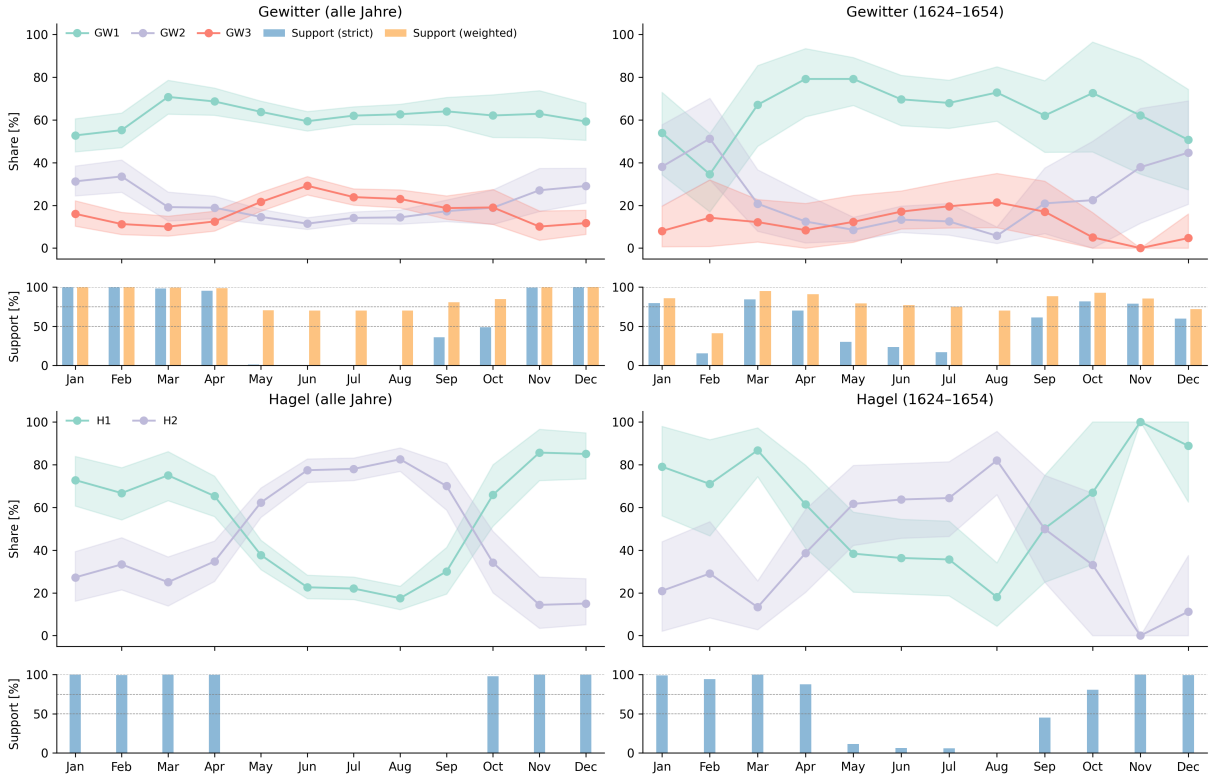

Figure 10. Distribution of class proportions for thunderstorm and hail events for the entire study period and the period 1624–1654. The monthly class proportions (lines) are shown with 95% confidence intervals (shading), calculated from monthly distributions normalised on an annual basis. The lower panels show the proportion of bootstrap samples in which the expected class hierarchy based on the resampled monthly proportions is fulfilled.

Newspapers and historiographical works show increased proportions of class $TS3$ ($R = +4.5$ and $R = +3.6$, respectively). These cells contribute significantly to the total χ^2 value (Figure 11).

For hail events, the chi-square test also shows a highly significant correlation between source type and intensity class ($\chi^2 = 390.6$, $\mathrm{df} = 10$, $p < 0.001$). The effect size is significantly more pronounced than for thunderstorm events (Cramér's V = 0.44).

The standardised residuals show a strong overrepresentation of class $H1$ in weather records ($R = +10.4$) and a simultaneous underrepresentation of class $H2$ ($R = -8.3$). In chronicles, class $H2$ is overrepresented ($R = +4.8$), while $H1$ is underrepresented ($R = -6.0$). Almanacs show increased proportions of $H1$ ($R = +6.6$) and reduced proportions of $H2$ ($R = -5.2$). Administrative files also show an overrepresentation of $H1$ ($R = +4.6$). These deviations explain a large part of the observed χ^2 value (Figure11).

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

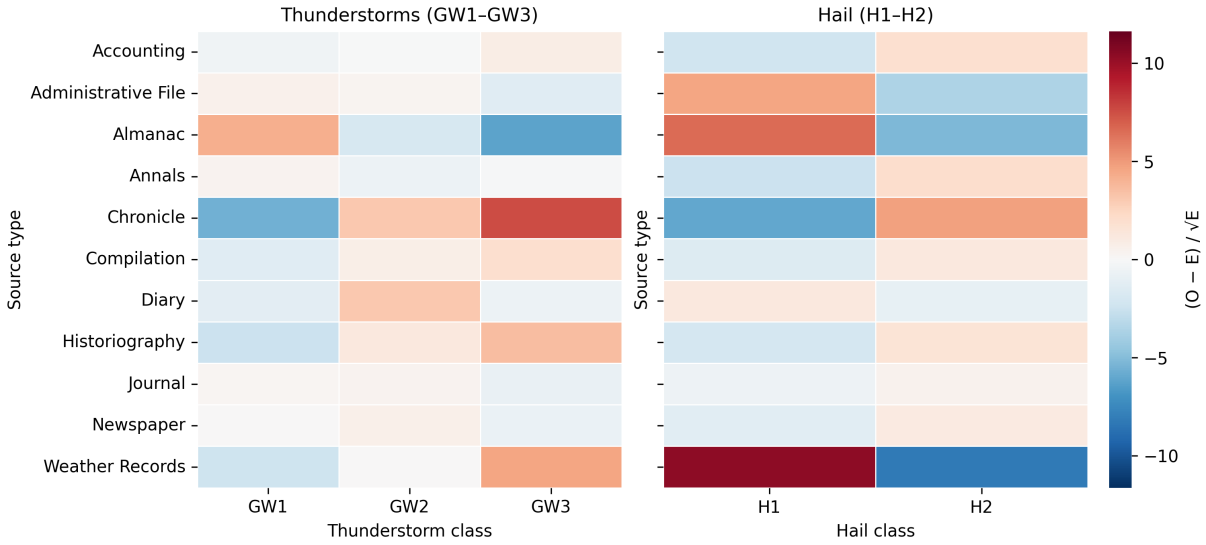

Figure 11. Standardised residuals of the chi-square tests for the dependence between source type and event class for thunderstorms (left, $TS1$–$TS3$) and hail (right, $H1$–$H2$). The residuals $(O-E)/\sqrt{E}$ are shown for each combination of source type and class. Positive values indicate overrepresented combinations, negative values indicate underrepresented combinations relative to the frequency expected under the assumption of independence.

4.4 Large Language Modell

Both classification models (ThunderstormBERT, HailBERT) are based on the same transformer architecture (mDeBERTa-v3-base). The hyperparameters specified in Table 2 correspond to the best settings for the respective classification task determined during model development. Despite identical architecture, there are differences in the training setup, particularly in the learning rate scheduler.

The learning curves of ThunderstormBERT show stable convergence after approximately five epochs with no evidence of relevant overfitting (Figure 12).

ThunderstormBERT achieves a macro-averaged F1 performance of 0.826 on the independent test dataset with an accuracy of 0.871 (corresponding to an error rate of 12.9%). Frequent classes such as *no thunderstorm* and *light thunderstorm* are detected with high precision and high recall (F1 > 0.91), while the rarer intensity classes, as expected, have lower but consistent F1 values (Table 3). Misclassifications occur predominantly between neighbouring intensity levels (Figure 1), suggesting gradual semantic transitions in the historical descriptions, where intensity gradations are often formulated implicitly or contextually.

The learning curves show a continuous increase in the F1 score for HailBERT with a simultaneous decrease in loss (Figure 12). Validation stabilises from the fourth epoch onwards, indicating efficient convergence with no signs of relevant overfitting. The normalised confusion matrix shows very high discrimination between classes, with only a few misclassifications outside neighbouring intensity levels (Figure 13).

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

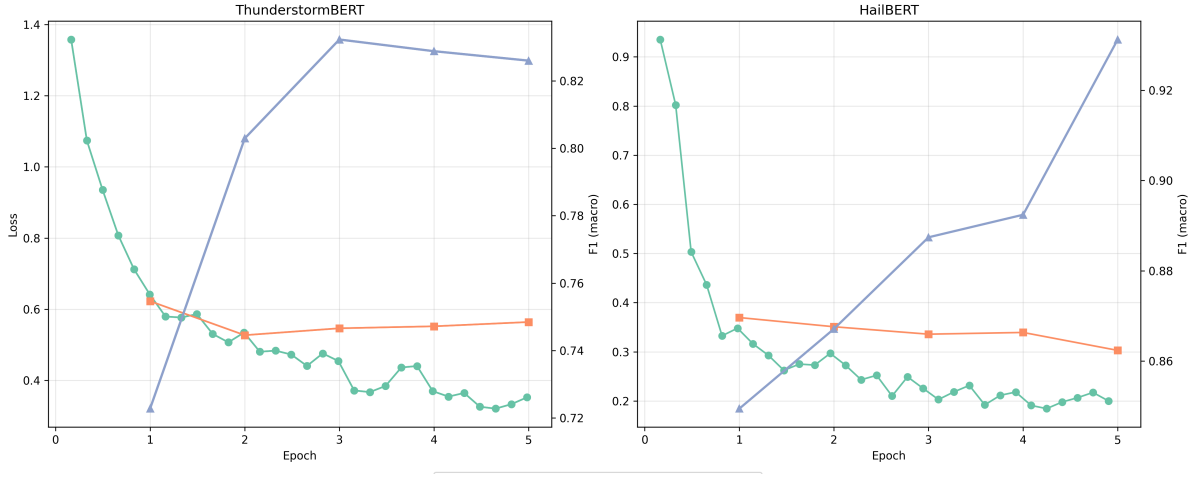

Figure 12. The figure shows the progression of training and validation metrics for ThunderstormBERT and HailBERT over 5 epochs.

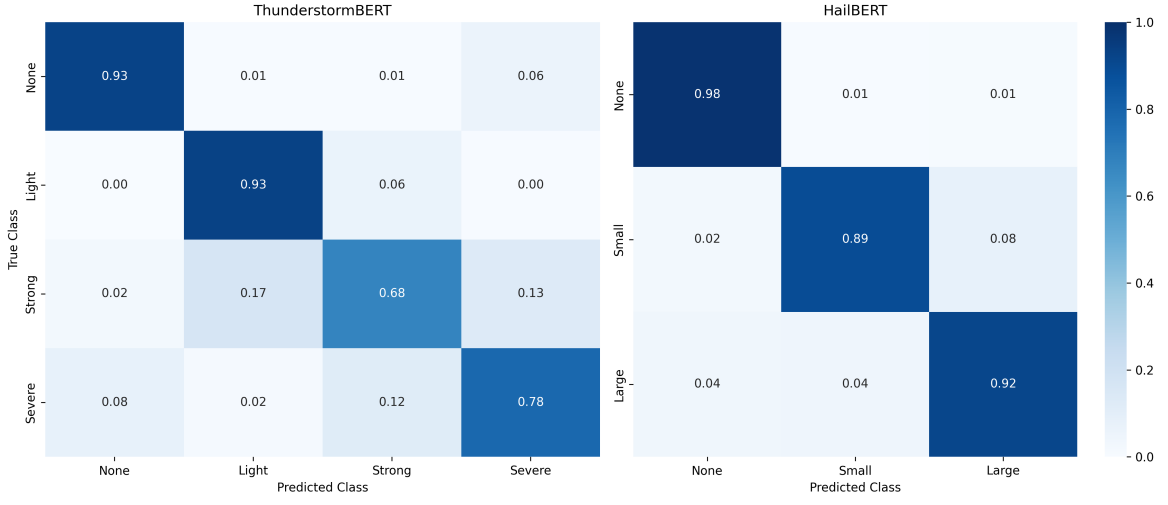

Figure 13. Normalised confusion matrix of ThunderstormBERT and HailBERT on the independent test dataset. For each true class (rows), the probability that the model predicts this class correctly (diagonal) or as another class is shown.

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

Table 2. Key figures for large language models used to classify historical thunderstorm and hail events: ThunderstormBERT and HailBERT.

Model	Thunderstorm (4 Classes)	Hail (3 Classes)
Architecture	mDeBERTa-v3-base	mDeBERTa-v3-base
Max. token length	256	256
Learning rate	3×10^{-5}	3×10^{-5}
Label smoothing	0.05	0.05
Scheduler	linear	cosine
Batch size	16	16
Epochs	5	5
F1 (macro), Validation	0.841	0.907
F1 (macro), Test	0.826	0.931
Accuracy (Test)	0.871	0.961
Test samples (n)	1,377	1,399
Error rate (Test)	12.9%	3.9%

Table 3. Class-specific test performance of the storm model.

Class	Precision	Recall	F1	Support
$TS0$	0.894	0.930	0.911	199
$TS1$	0.948	0.932	0.940	760
$TS2$	0.642	0.677	0.659	201
$TS3$	0.806	0.783	0.794	217

HailBEERT achieves a very high overall classification performance with a macro-averaged F1 score of 0.931 and an accuracy of 0.961 (corresponding to an error rate of 3.9%). All classes are classified with high and balanced scores (Table 4). Here, too, misclassifications are predominantly concentrated in neighbouring intensity classes, which indicates a consistent internal representation of the class order. The higher overall performance of HailBERT must also be interpreted in the context of the smaller number of classes and the clearer semantic distinguishability of hail events.

Table 4. Class-specific test performance of the hail model.

Class	Precision	Recall	F1	Support
$H0$	0.984	0.984	0.984	964
$H1$	0.889	0.894	0.892	161
$H2$	0.919	0.916	0.918	274

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

5 Discussion

The analysis shows that historical thunderstorm and hail reports exhibit a remarkably stable seasonal signal despite highly variable source density and a heterogeneous source base. The reconstructed annual cycle, which is based on monthly proportions normalised on an annual basis, is dominated by a pronounced summer maximum. The maximum determined shows significant agreement with modern observational data in terms of location, shape and relative amplitude. This pattern manifests itself consistently both in the overall series and in the densest observation window (1624–1654). The stability of the seasonal pattern is evident regardless of fluctuations in the density of records and is therefore not attributable to individual phases of increased documentation or the distribution of sources.

Despite a time lag of several centuries and fundamentally different survey methodologies, a comparison with current observation data from the DWD shows a high degree of consistency in seasonal patterns. The high Spearman rank correlations and the low deviations in the normalised monthly proportions indicate that the historical observations consistently refer to real meteorological processes. Furthermore, the seasonal signature proves to be largely independent of the observation density: even pronounced fluctuations in the number of active sources only slightly alter the shape of the annual cycle.

The annual trends presented in this study show significant agreement with the independent reconstructions by Lenke (1960) and Camuffo et al. (2000). This agreement is evident both for the entire series and for the densest window from 1624 to 1654, as shown in Figure 14. The ratio of summer to winter proportions (JJA vs. DJF) in the historical series is approximately $3.0 - 3.5$, while the values according to Camuffo et al. (2000) (6.6) and, in particular, Lenke (1960) (13.2) show a significantly stronger summer-winter contrast. Winter activity is therefore more pronounced in the historical data set than in the comparative series, but remains structurally within the expected seasonal pattern. One potential explanation for the increased observation of thunderstorms in winter could be due to source-typical perceptions of winter thunderstorms and hail events, which were perceived as unusual (eg., Camuffo et al., 2000; Glaser, 2013).

The correlation analysis supports the agreement (Hist vs. Lenke: Spearman 0.86; Hist vs. Camuffo: 0.71), with the period 1624–1654 yielding almost identical results. This indicates a robust seasonal signature of thunderstorm activity across different periods, regions and source types, and underlines the plausibility of the reconstructed annual cycles.

The temporal development of the summer–winter ratio (JJA/DJF) also confirms this robustness in a longitudinal analysis. Over more than four centuries, the ratio has remained predominantly above the unit line and shows no systematic trends that would indicate changes in source density or documentation practices. The consistently pronounced summer maximum suggests that the observed seasonal signature is primarily determined by real convective processes and is not generated by transmission or selection artefacts.

This is further supported by the calendar heat maps. Periods of increased activity are concentrated in the summer months, while the winter months show very little activity. The physically consistent seasonality that can be observed is a key criterion for the plausibility of historical records and speaks for the reliability of the reconstructed thunderstorm and hail time series.

Analysis of the monthly class proportions shows a clear dominance of the weakest storm class (GW1) over the entire study period, with the highest average proportions in all months. Class GW2 occurs with lower but stable proportions over much

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

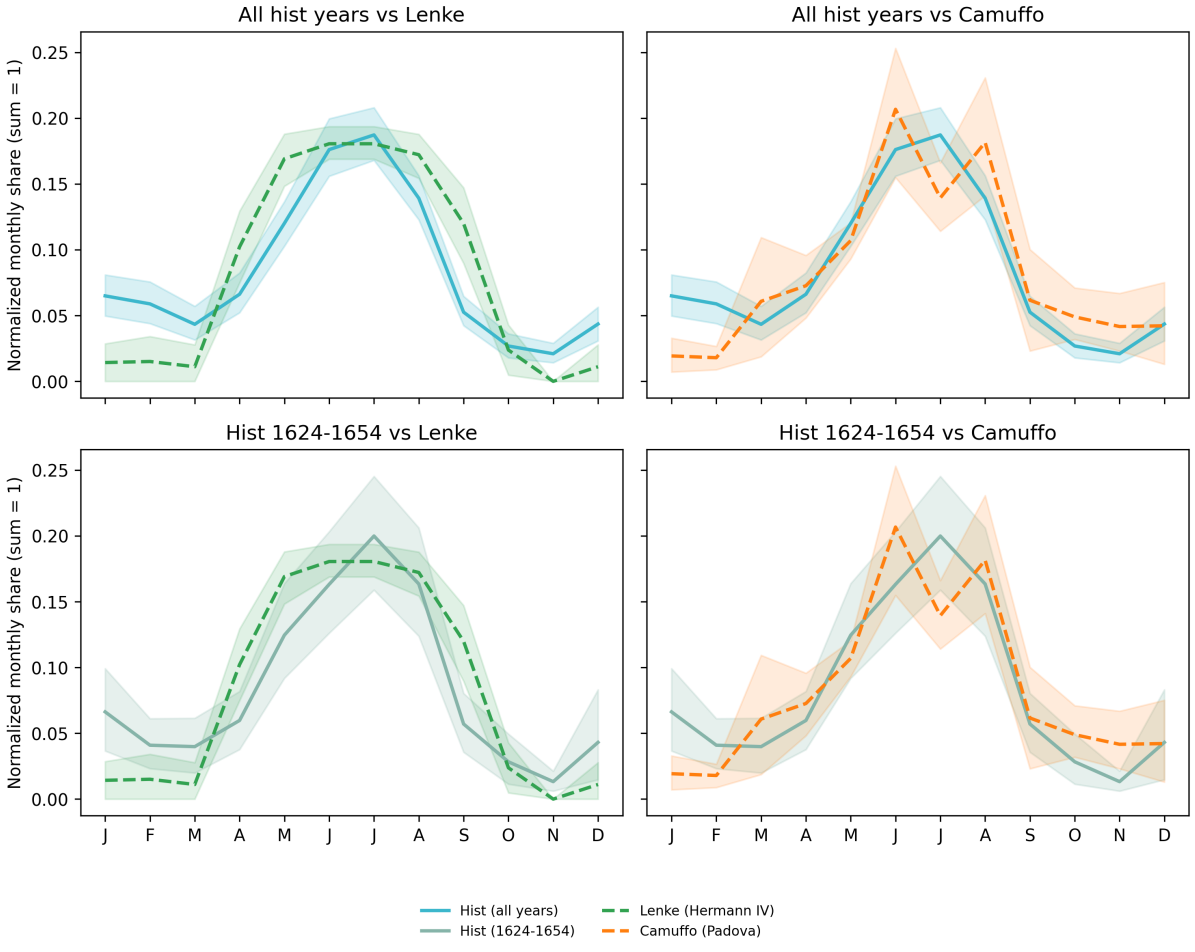

Figure 14. Seasonal progression of historical thunderstorm events compared with two independent reconstruction series. The figures show percentage curves of monthly thunderstorm frequency relative to the annual total, each with 95% bootstrap confidence intervals. Top: Comparison of the entire historical series with Lenke (1960) and Camuffo et al. (2000). Bottom: Identical comparison for the densest study window 1624–1654. The high degree of agreement between the curve shapes underscores the robustness of the seasonal pattern (summer maximum, moderately increased winter activity) across different data sets, regions and methodological approaches.

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

of the year, while the strongest class, GW3, achieves higher relative proportions, especially in the summer months. In the transitional seasons, the ranking is predominantly GW1 > GW2 > GW3. In the summer months, however, the average class proportions repeatedly show a changed secondary ranking of GW1 > GW3 > GW2.

These findings conflict with the assumption frequently expressed in the literature that historical sources primarily record extreme weather events (cf. Lenke, 1960; Camuffo et al., 2000; Brázdil et al., 2016a, b; Burgdorf, 2021). However, such a general dominance of extreme classes cannot be confirmed for thunderstorms and hail events. Under this assumption, a ranking of GW3 > GW2 > GW1 would have been expected. In fact, the observed class proportions in historical records also predominantly follow the physical frequency distribution, with a clear dominance of weaker events (GW1 > GW2 > GW3). This suggests that, at least for thunderstorm events, there is no systematic overrepresentation of extreme intensities.

The analysis of the dependence between source type and intensity classification shows for thunderstorms ($\chi^2 = 479.66$, df = 22, $p < 0.001$, Cramér's V = 0.21) and hail events ($\chi^2 = 376.23$, df = 10, $p < 0.001$, Cramér's V = 0.43). While the correlation for thunderstorm events can be classified as small to moderate (corresponding to an explained variance of about 4%), hail events show a significantly stronger, medium effect (variance explained about 18%). These differences point to source-specific divergent reporting patterns, which are much more pronounced for hail events than for thunderstorm events overall.

The seasonal contrast in the ranking of hail classes can be explained by a combination of physical and source-related effects. In the summer months, high convective energies mean that hail events tend to be more intense when they occur, while weak hail is either meteorologically less frequent or less frequently documented due to its lower visibility. In winter and transitional seasons, on the other hand, hail occurs predominantly in weaker forms, but is recorded even at low intensity due to the overall lower event density and increased perception sensitivity. The transitional months of April and September consistently mark the seasonal change in this dynamic.

These physical differences are further reinforced by source-specific reporting patterns. Narrative sources in particular, such as chronicles, tend to disproportionately document more intense hail events (H2), while weather diaries more frequently record weaker hail events (H1). Hail appears to be particularly newsworthy in narrative sources when it causes damage or takes on unusual forms, while continuous observation formats also systematically record lower intensities.

It is noteworthy that these source-specific distortions are implicitly reflected in the hail classification model, even though the source type was not explicitly integrated as a training feature. The high classification quality of the model with only minor misclassifications suggests that the semantic patterns contained in the historical texts are sufficient to automatically map and take into account such source-related differences.

The results of the language model fit in with the findings previously obtained on the seasonal structure of convective events. They show that the patterns observed in historical sources can also be consistently reproduced at the semantic-linguistic level. Although the model was trained exclusively on texts and does not include any additional meteorological variables, it reproduces the characteristic class hierarchies (e.g. GW1 > GW2 > GW3 over the course of the year) as well as the summer intensification of stronger events. The model thus confirms the coherence of the corpus in terms of content: the historical texts contain sufficient meteorologically relevant information to reliably capture seasonal and intensity-related differences.

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

The class structure used shows stable results across all classes. This illustrates that the model does not primarily memorise formulations, but rather recognises recurring semantic patterns. The few remaining misclassifications occur almost exclusively at transitions between neighbouring intensity levels. This pattern is known from both meteorological practice and source-critical analysis. Serious mix-ups are rare. This indicates a robust separation of the relevant linguistic signals and the resilience of the classification scheme.

This source-specific effect was not explicitly integrated into the language model as an input variable, but is reproducibly reflected in the classification results. This suggests that the model learns implicit structures of historical tradition – including source-specific selection mechanisms – rather than levelling them out. The models thus work consistently with the characteristics of the corpus, which is of central importance for applications in historical climate research.

Both models show stable and reproducible results and are suitable for the automated classification of historical weather descriptions. The robustness of the results is supported by consistent seasonal patterns, narrow confidence intervals and stable rankings of intensity classes. The observed performance differences between thunderstorm and hail classification are significantly greater than differences in the training setup and can be explained primarily by the different class structure and semantic selectivity of the respective target phenomena.

6 Conclusions

The study shows that historical text sources have great potential for the quantitative reconstruction of convective weather events despite linguistic heterogeneity. By combining source-critical analysis, rule-based classification and statistical validation, it was possible to create a consistent database whose event patterns correspond to modern convective climatology. The conversion of qualitative descriptions into structured evidence levels and semantic labels enabled the successful use of a multilingual BERT model that reliably recognises and classifies thunderstorm and hail events. Overall, the work shows that AI-based methods can only extract reliable climatological information from historically data that has been methodically prepared.

Future work should expand the dataset to systematically investigate regional differences between large landscapes.

Code and data availability. Historical Climate Observations of Thunderstorms and Hail in Central Europe (1000–1900): https://doi.org/10.60493/834bd-mww13,
ThunderstormBERT-de-v1: 10.57967/hf/6982 or https://huggingface.co/Stickmu/ThunderstormBERT-de-v1,
HailBERT-de-v1: 10.57967/hf/6989 or https://huggingface.co/Stickmu/HailBERT-de-v1

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

Appendix A: Classification schemes

Table A1. Classification of thunderstorm intensity

Class	Description
-99	The thunderstorm's characteristics cannot be clearly determined.
0	**No thunderstorm:** This class includes events that do not meet the basic criteria for a thunderstorm and serves as a baseline category for differentiation.
1	**Light thunderstorm:** Events that do not meet the criteria for TS3 are classified as light if all accompanying phenomena exhibit only low intensity (rain $\leq$ level 1, wind $\leq$ level 1, hail $\leq$ level 1).
2	**Moderate thunderstorm:** This class includes thunderstorm events that do not reach the thresholds for a severe thunderstorm (TS3) but exceed the intensity of a light thunderstorm (TS1).
3	**Severe thunderstorm:** An event is assigned to this highest level if at least one of the following criteria is met: – Hail diameter > 2.0 cm – Wind speed reaches predefined level 3 (severe storm/hurricane-force gusts) – Rain intensity reaches predefined level 3 (heavy rain)

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

Table A2. Classification of hail intensity

Class	Description
-99	The hail's characteristics cannot be clearly determined.
0	**No hail:** Quote without hail events.
1	**Light hail:** Quote mentioning hail without hail impact. Hail size is less than 2 cm.
2	**Hail impact or large accumulation:** Hail impact occurs at sizes of 2 cm or larger. Frequent comparisons include cherry-, walnut-, pigeon-, or hen's egg-sized hail. Damage may also be described indirectly (e.g., damaged leaves, branches, fruit, window panes, roofs, or metal cladding). Quantity descriptions such as "hail lay knee-deep in the alleys" also fall under this category.

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

Table A3. Classification of rain intensity during thunderstorm and hail events

Class	Description
-99	The rain's characteristics cannot be clearly determined.
0	**No rain:** Thunderstorm or hail without accompanying precipitation.
1	**Rain:** Thunderstorm or hail with ordinary rain.
2	**Heavy rain without flooding:** Thunderstorm or hail with heavy rain but without immediate consequential damage. Typical historical descriptions include *cloudburst* or *torrential rain*.
3	**Heavy rain and flooding:** Thunderstorm or hail with heavy rain and severe impacts such as floods, landslides, damage to mills, or inundated fields. Unlike class 2, this class involves immediate geomorphological and socio-economic consequences.

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

Table A4. Classification of snow intensity during thunderstorm and hail events

Class	Description
-99	The snow's characteristics cannot be clearly determined.
0	**No snow:** Thunderstorm or hail without accompanying snowfall.
1	**Light snowfall:** Thunderstorm or hail with snow. Fresh snow depth below 10 cm in lowland areas or below 20 cm in mountainous regions, without significant consequences.
2	**Heavy snowfall:** Thunderstorm or hail with heavy snowfall. Fresh snow depth of 10 cm or more in lowland areas or 20 cm or more in mountainous regions, without descriptions of significant consequences. Typical historical descriptions include "a lot of snow," "large amounts of snow," or "piling snow."
3	**Heavy snowfall with consequences:** Any snowfall that explicitly causes disruptions or damage, regardless of the amount. Typical indicators include snow breakage in trees, collapsed roofs, impassable roads, snowed-in persons, or avalanches.

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

Table A5. Classification of wind intensity during thunderstorm and hail events

Class	Description
-99	The wind's characteristics cannot be clearly determined.
0	**No wind:** Thunderstorm or hail without significant wind.
1	**Wind:** Thunderstorm or hail with wind; convective gusts up to approx. 7 Beaufort.
2	**Strong wind with minor damage:** Convective storm gusts (8–9 Beaufort). Typical indicators: Large trees sway, shutters open, branches break, significant difficulty walking, minor damage to houses (individual roof tiles or chimney pots lifted).
3	**Strong wind with severe damage:** Convective severe storm gusts, hurricane-force gusts, or tornadoes ($\geq$ 10 Beaufort). Typical indicators: Uprooted trees, snapped trunks, windthrow in forests, severe damage to buildings (roofs blown off, thick walls damaged), walking impossible, widespread devastation.

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

Table A6. Standardized terms and size specifications for historical hail descriptions. The table summarizes typical expressions from historical sources that were used to describe hail events. By assigning them to standardized size ranges and energy levels, it enables consistent and comparable classification of hail intensity within the thunderstorm classification scheme. The definition of the thresholds is based on established meteorological and climatological references to ensure a high degree of comparability of the results. The decisive threshold for hailstones was determined based on the criteria of European Severe Storms Laboratory (ESSL) (2025) and (Tornado and Storm Research Organisation (TORRO) (2025), in which the kinetic energy of the hailstones does not increase linearly, resulting in an increased probability of significant damage to agricultural crops, buildings, and infrastructure.

Term (original)	**Example (database ID)**	**∅ cm**	**Class**
Märmel (marble)	"...teils so groß wie eine Märmel..." (33897)	0.5–0.9	H1;TS2
Hagelstein (hailstone)	"...ungewöhnlichen Größe..." (6)	0.5–5.0	H1-2;TS2-3
Schnellkugel (musket ball)	„...wie große Schnellkugeln..." (5638)	0.9	H1;TS2
Bohne (bean)	"...wie kleine Bohnen..." (479)	1.0–2.0	H1;TS2
Haselnuss (hazelnut)	"...so groß wie Haselnüsse..." (61)	1.2–2.0	H1;TS2
Ital. Nuss (Italian nut)	"...Größe einer ital. Nuss..." (2073)	1.5–2.0	H1-2;TS2-3
Schoßkugel (pistol ball)	"...so groß wie Schoßkeulen..." (496)	1.5–2.0	H2;TS3
Viertelguld. (quarter gulden)	"...¼ Gulden schwer..." (4240)	1.8–2.2	H2;TS3
Daumen (thumb)	"...als ein Daumen dick..." (707)	2.0–2.5	H2;TS3
Baumnuss (walnut)	"...so groß wie Baumnüsse..." (2959)	2.0–3.0	H2;TS3
Kastanie (chestnut)	"...so groß wie Kastanien..." (1010)	2.0–4.0	H2;TS3
Böhm. Grosch. (Bohemian groschen)	"...wie böhmische Groschen..." (928)	2.7–3.0	H2;TS3
Walnuss (walnut)	"...so groß wie Walnüsse..." (354)	2.5–3.5	H2;TS3
Taubenei (pigeon's egg)	"...wie Taubeneier groß..." (86)	3.0	H2;TS3
Taler (thaler coin)	"...so groß wie halbe Taler..." (696)	4.0–4.5	H2;TS3
Halbes Pfund (half pound)	"...anderthalb Pfund..." (361)	≈ 8.0	H2;TS3
Hühnerei (hen's egg)	"...so groß wie ein Hennenei..." (60)	4.5	H2;TS3
Apfel (halb) (half apple)	"...Größe halber Äpfel..." (6676)	6.0–7.5	H2;TS3
Brot (halb) (half loaf)	"...so groß wie ein halbes Brot..." (2787)	7.5–10.0	H2;TS3
Gänseei (goose egg)	"...wie Gänseeier groß..." (281)	8.0	H2;TS3
Faust (fist)	"...wie eine Faust groß..." (1053)	10.0–12.0	H2;TS3
Kürbis (pumpkin)	"...wie Kürbis..." (2994)	20+	H2;TS3
Damage (C2)	"...schlug die Fenster ein..." (385)	n.a.	H2;TS3
Accumulated (C1)	"...fast schuhhoch..." (2140)	n.a.	–

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

Appendix B: Linguistic indicators

Table B1. Linguistic indicators for thunderstorms and hail events in historical sources

Phenomenon	Linguistic indicators (nouns, verbs, adjectives)
Thunderstorm	blitz, blitzeinschlag, blitzen, blitzgewitter, blitzn, blitzschlag, blitzstrahl, blitzte, donner, donnereinschlag, donneren, donnergewitter, donnergrollen, donnerhall, donnerkeil, donnerknall, donnerknalle, donnerkrach, donnern, donnerndes, donnerregenwetter, donnerschlag, donnerstein, donnerstrahl, donnerstreich, donnerwetter, feuerblick, feuerdrach, feuererscheinung, feuerflamme, feuerhagel, feuerklumpe, feuerklumpen, feuerkugel, feuerregen, feuerschlag, feuerstahl, feuerstrahl, feuerwerk, feuerwolke, feuerzeichen, fromalwetter, geblitz, gedonnert, gewitt, gewitter, gewitterheiß, gewitterleuchte, gewitterleuchten, gewitterläuten, gewittern, gewitterregen, gewitterschaden, gewitterschauer, gewittersturm, gewitterwind, gewitterwolke, gewittrig, hagelunwetter, himmelblitzen, hochgewitter, hochwetter, kanonenschuss, knall, krachen, krachend, lichtwolke, niedergestreckt, platzregen, regen-gewitter, regengewitter, regenstrahl, schlagwetter, strichgewitter, sturmdonner, ungewitter, unwetter, wassergewitter, wasserstrahl, wasserstrahle, wasserstreich, wettergeleuchten, wettergeleuchtet, wetterleuchte, wetterleuchten, wetterleuchteten, wetterläuten, wettern, wetterregen, wetterschein, wetterschlag, wetterstrahl, wetterstreich, wetterwolke, wildfeuer, wildstürmend, wirbelwind, wittergeleuchten
Hail	baumnuss, baumnüs, donnerstein, ei, eishagel, eisklumpe, eisschelle, eisstein, eisstück, enteneier, erbse, faust, faustgroß, gehageln, graupel, gänseeier, hagel, hagel-, hagel-regen, hagel-wetter, hagelaufstieg, hagelgewitter, hagelkorn, hageln, hagelregen, hagelregenwetter, hagelschauer, hagelschlag, hagelstein, hagelsturm, hagelsturz, hagelstück, hagelunwetter, hagelwolke, haselnuss, haselnuß, haselnüsse, henneneier, hühnerei, hühnereier, kiesel, kieseln, kieselregen, kieselschlag, kieselschloßen, kieselstein, kieselwetter, nuß, nüsse, schlagwetter, schnellkugel, schoßkeule, steinschlag, streifkiesel, strich, strichgewitter, taubeeier, taubenei, taubeneie, taubeneier, taubeneigroß, verhageln, walnuss, walnussgröße, walnüs, walnüss, welsch, welschen, welschen-nüsse, zerhageln
Wind	braus, brausen, brausend, böe, luftmasse, lüftlein, nord-ost, nord-ost-wind, nord-osten, nord-west, nord-west-wind, nord-westen, nordastwind, norden, nordost, nordosten, nordostwind, nordwesten, nordwestwind, nordwind, nordöstlich, nördlich, orkan, orkansturm, ostwind, schneegestöb, schneegestöber, schneesturm, starkwind, sturm, sturm-balken, sturmbalken, sturmdonner, sturmwetter, sturmwind, sturmzeichen, stürmen, stürmend, stürmisch, süd, süd-ost, süd-ost-süd, süd-ost-wind, süd-süd-ost, süd-west-wind, südwesten, süd-wind, südost, südostwind, südwest, südwesten, südwind, südwärts, unterwind, vorüberziehend, wehen, west, west-nordwest, west-süd-west, west-südwind, west-wind, westlich, westwind, wind, windbrause, windbrausen, windbö, windböe, winde, winden, windfahne, windig, windrauschen, windstill, windstille, windstoß, windsturm, windstöße, windwirbel, wirbel, wirbelsturm, wirbelwind, östlich
Rain	donnerregenwetter, ergießen, ergießung, ergoss, feuerregen, flut, fromalwetter, frühlingswetter, frühlingswitterung, geregnet, gewaltwasser, gewalzt, gewitterregen, gewitterschauer, grobwetter, grundregen, gussregen, hagel-regen, hagelregen, hagelregenwetter, hochwasser, hochwetter, landregen, nass, naß, niederschlag, nieselregen, pegel, platzregen, prassel, regen, regen-gewitter, regenbogen, regenfall, regengewitter, regenguss, regenguß, regengüssen, regenmasse, regenschauer, regenschauerwetter, regenstrahl, regenwasser, regenwetter, regenwolke, regnen, regnerisch, reißend, schauer, schlagregen, schlagwetter, schlamm, schwefel, sindfluth, springflut, sprührege, sprühregen, spülung, starkregen, strich, strichgewitter, strichregen, strom, strömen, strömung, sündflut, tropfen, verschemmung, wasser, wasser-wolke, wasserergüsse, wasserfluss, wasserflut, wassergefahr, wassergewitter, wasserguss, wasserguß, wassergüsse, wassergüssen, wassermasse, wassermenge, wassernebel, wassernot, wasserrinne, wasserspiegelbogen, wasserstrahl, wasserstrahle, wasserstreich, wasserstrom, wasserstürz, wassersäule, wasserwirbel, wegschwemmen, wetterregen, wolkenbruch, überfloss, überflutung, überschwemmen, überschwemmung
Snow	alpenschne, beschneien, eingeschneit, schnee, schneeflocke, schneegestöb, schneegestöber, schneesturm, schneewasser, schneeweiß, schneewetter, schneewolke, schneien, triebschnee

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

Table B2. Vocabulary for semantic identification of hail in the corpus used (Silver Label).

Group	Description	Complete vocabulary
C1	**Direct and specific:** Concrete measurements or physical references (see also table A6).	6 zentnern · apfel · äpfel · baumnuss · baumnüsse · bleikugel · bohne · bohnen · brot · daumen · dotter von eiern · drei männerfäuste · ei · eier · enteneier · erbse · erbsen · faust · fingerlang · gänseei · groschen · groß wie bälle · haselnuss · hennenei · hühnerei · hühnereier · italienischen nuss · kastanien · kieselstein · kleine kürbisse · kugel im feuerrohr · kürbis · lot · märmel · männerfaust · nuss · nüsse · orangen · pfund · pfundstein · schnellkugel · schoßkeule · schoßkugel · sperling · spielkugeln · steingröße · taler · taubenei · taubeneier · walnuss · walnussgröße · walnüsse · welsch nuss · welschen nüsse · werben und berkelinge
C2	**Indirectly:** Indications of damage or effects.	beschädigt · erschlag · erschlagen · feldfrüchte ausgeschlagen · fenster eingeschlagen · flurschaden · früchte beschädigt · große verwüstung · großer hagelschlag · hagel abgeschlagen · hagelschaden · hagelschlag · häuserschaden · eingeschlagen · geschlagen · geschädigt · kieselschlag · mäßiges unheil · niedergeschlagen · ruiniert · schadete · schaden · schädigung durch hagel · schädlich · schädliche hagelwetter · schlug · strich · verdorben · verhagelt · verwüstet · weggeschlagen · wein von den stöcken schlug · wetterschlag · zerknickte · zerstörten · zerschlagen · zerschmettert · zerrieb · zerschlug
C3	**Relative (qualitative):** Linguistic intensifiers or names for strong hail, intensity and severity.	böse hagelwetter · furchtbaren hagel · furchtbares hagelwetter · gewaltigem hagel · gewaltigen schloßen · gewaltiger hagel · grausam hagel · grausam gekieselt · grausamen hagel · grausamer hagel · grausames hagelwetter · große hagel · große kiesel · große schloßen · großem hagel · großen hagels · großen kieselsteinen · großen schloßen · großen steinwürfen · große steine · großes hagelwetter · hagel von erstaunlicher größe · hartes kieselwetter · heftig hagel · heftiger hagelwetter · schloßen so groß · schloßen ungeheuer groß · schreckliche hagelwetter · schrecklichem hagel · schrecklichen hagelwetter · schrecklicher hagel · schreckliches hagelwetter · schwere hagelwetter · schwere kiesel · schweres hagelwetter · schwersten hagelwetter · stark gehagelt · starke schloßen · starkem hagel · starken kieseln · starken schloßen · starker hagel · starkes hagelwetter · übergroßer hagel · ungewöhnliche schloßen · ungewöhnlicher größe · verheerende hagelwetter · verheerender hagel · viele schloßen · große hagelsteine · eisengleichen kugeln · großer menge · ungemeiner größe · große ungeheure schloßen · grausame hagel · großer und schrecklicher vereister hagel · erstaunlichem hagel · solch einer größe · schrecklich zu hageln · heftigen kieselwetter

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

Table B3. Vocabulary for the semantic identification of rain in the corpus used (Silver Label).

Group	Description	Typical word forms and phrases
C1	**Direct and specific:** Direct, measurable, or quantitatively determinable information.	(no direct information on specific measurements available)
C2	**Indirect:** Indications of damage or effects.	weggerissen · wasser riss · erde fortriss · vernichtete · verwüstete · weggeschwemmt · weggespült · ausgewaschen · überschwemmte · überflutete · überspült · verschlemmte · verschemmung · weggeführt · bäume geführt · weggetrieben · wasser riss häuser · keller voll · brücke zerbrochen · schaden mühlen · mühle verdarb · über deiche · nicht ernten · heu verfault · feldfrüchte schaden · land verschoben · schlamm bedeckt · ertranken · ersäufen · versoffen · untergingen · starken regen schaden · wassergefahr · wassernot · wolkenbruch großer gefahr
C3	**Relative (qualitative):** Linguistic descriptions of intensity, amount of water, height, or extent.	hochwasser · überschwemmung · flut · wasserfluten · feldfluten · gewässer · großes wasser · viel wasser · großmächtiges wasser · gewaltige wassermenge · regenmassen · schreckliches wasser · wasser so stark · gewaltigem regen · wilde wasser · wasser grausam wuchsen · hoch stand · sehr hoch angelaufen · hohes wasser · wasser hoch gestanden · steigen des wassers · bäche anschwollen · halben mann hoch · wasser zwei ellen hoch · bäche über ihre ufer gingen · [Fluss/Ort] ausgetreten · Radaune ausbrach · flüsse ergossen · schrecklicher ergießung · übergelaufen · gänzlich überströmt · zusammenfließende wasser · sündflut · unpassierbar · unter wasser · kahn fahren · schwammen · flößte

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

Table B4. Vocabulary for semantic identification of wind in the corpus used (Silver Label).

Group	Description	Typical word forms and phrases
C1	**Direct and specific:** Direct, measurable, or quantitatively determinable information.	(no direct information on specific measured values available)
C2	**Indirect (damage indicators):** Effects on objects, buildings, vegetation, or infrastructure.	gehobelt · bäume gebrochen · bäume riss · bäume entwurzelte · linden zerstörte · schutzmauern umstürzte · türme eingerissen · häuser geworfen · mauern verdarben · häuser abdeckte · kirchenfenster übel zurichtete · dächer abgetragen · mauer eingestürzt · schiffe vergangen · fenster zerschlagen · getreide niedergeschlagen · wurzel gerissen · winde gewütet · großen schaden · schaden häusern
C3	**Relative (qualitative):** Linguistic intensifiers or names for strong winds, intensity and severity.	sturm · sturmwind · sturmwinden · stürmte · sturmdonner · sturmwetter · windsturm · gewittersturm · gewalt des windes · großer stärke · starker wind · sehr starken wind · heftiger wind · böiger wind · heftigen böen · mächtigen wind · schrecklicher wind · ungestümer wind · grausamer wind · stürmischer wind · orkan · tornado · wirbelwind · starker reißender wind

Competing interests. The authors declare that they have no conflict of interest.

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

References

Brázdil, R., Dobrovolný, P., Luterbacher, J., Moberg, A., Pfister, C., Wheeler, D., and Zorita, E.: European climate of the past 500 years: new challenges for historical climatology, Climatic Change, 101, 7–40, https://doi.org/10.1007/s10584-009-9783-z, 2010.

Brázdil, R., Chromá, K., Valášek, H., Dolák, L., and Řezníčková, L.: Damaging hailstorms in South Moravia, Czech Republic, in the seventeenth to twentieth centuries as derived from taxation records, Theoretical and Applied Climatology, 123, 185–198, https://doi.org/10.1007/s00704-014-1338-1, 2016a.

Brázdil, R., Chromá, K., Valášek, H., Dolák, L., Řezníčková, L., Zahradníček, P., and Dobrovolný, P.: A long-term chronology of summer half-year hailstorms for South Moravia, Czech Republic, Climate Research, 71, 91–109, https://www.jstor.org/stable/24897492, publisher: Inter-Research Science Center, 2016b.

Brönnimann, S., Martius, O., Rohr, C., Bresch, D. N., and Lin, K. E.: Historical weather data for climate risk assessment, Annals of the New York Academy of Sciences, 1436, 121–137, https://doi.org/10.1111/nyas.13966, 2019.

Burgdorf, A.-M.: A global inventory of historical documentary evidence related to climate since the 15th century, preprint, Proxy Use-Development-Validation/Historical Records/Decadal-Seasonal, https://doi.org/10.5194/cp-2021-165, 2021.

Camuffo, D., Cocheo, C., and Enzi, S.: Seasonality of instability phenomena (hailstorms and thunderstorms) in Padova, northern Italy, from archive and instrumental sources since AD 1300, The Holocene, 10, 635–642, https://doi.org/10.1191/095968300666845195, 2000.

Cutter, S. L.: The Changing Nature of Hazard and Disaster Risk in the Anthropocene, Annals of the American Association of Geographers, 111, 819–827, https://doi.org/10.1080/24694452.2020.1744423, 2021.

Deutscher Wetterdienst (DWD): Offizielle Webseite des Deutschen Wetterdienstes, https://www.dwd.de/, 2025.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, http://arxiv.org/abs/1810.04805, arXiv:1810.04805 [cs], 2019.

Diodato, N., Ljungqvist, F. C., and Bellocchi, G.: A millennium-long reconstruction of damaging hydrological events across Italy, Scientific Reports, 9, 9963, https://doi.org/10.1038/s41598-019-46207-7, 2019.

Ehrmanntraut, A.: Historical German Text Normalization Using Type- and Token-Based Language Modeling, https://doi.org/10.48550/arXiv.2409.02841, arXiv:2409.02841 [cs], 2025.

Erfurt, M., Skiadaresis, G., Tijdeman, E., Blauhut, V., Bauhus, J., Glaser, R., Schwarz, J., Tegel, W., and Stahl, K.: A multidisciplinary drought catalogue for southwestern Germany dating back to 1801, Natural Hazards and Earth System Sciences, 20, 2979–2995, https://doi.org/10.5194/nhess-20-2979-2020, 2020.

Ernst, P.: Deutsche Sprachgeschichte: eine Einführung in die diachrone Sprachwissenschaft des Deutschen, no. 2583 in utb Sprachwissenschaft, Facultas, Wien, 3. auflage edn., ISBN 978-3-8252-5532-9, 2021.

European Environment Agency: Environmental zones 2018 - version 1.0, June 2020, https://sdi.eea.europa.eu/catalogue/idp/api/records/6ef007ab-1fcd-4c4f-bc96-14e8afbcb688, 2020.

European Severe Storms Laboratory (ESSL): European Severe Storms Laboratory - Official Website, https://www.essl.org/cms/, 2025.

Giordani, A., Kunz, M., Bedka, K. M., Punge, H. J., Paccagnella, T., Pavan, V., Cerenzia, I. M. L., and Di Sabatino, S.: Characterizing hail-prone environments using convection-permitting reanalysis and overshooting top detections over south-central Europe, Natural Hazards and Earth System Sciences, 24, 2331–2357, https://doi.org/10.5194/nhess-24-2331-2024, 2024.

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

Glaser, R.: Klimageschichte Mitteleuropas : 1200 Jahre Wetter, Klima, Katastrophen, Primus, Darmstadt, sonderausg. 2013 3., unveränd. aufl. edn., ISBN 3-86312-350-6, http://deposit.d-nb.de/cgi-bin/dokserv?id=4215166&prov=M&dok_var=1&dok_ext=htmandhttp://www.ub.unibas.ch/tox/IDSBB/006075550/PDF, 2013.

Glaser, R.: HISKLID: Historische Klimadatenbank, https://freidok.uni-freiburg.de/proj/3720, 2014.

Grzega, J.: Climatic Conditions and Lexis: Some Diachronic Notes on Weather-Related Words in English and Other European Languages, Transactions of the Philological Society, 120, 320–331, https://doi.org/10.1111/1467-968X.12243, 2022.

Gudd, M.: Gewitter und Gewitterschäden im südlichen hessischen Berg- und Beckenland und im Rhein-Main-Tiefland 1881 bis 1980, Ph.D. thesis, Johannes Gutenberg-Universität Mainz, https://doi.org/10.25358/OPENSCIENCE-3426, 2004.

Hawkins, E., Brohan, P., Burgess, S. N., Burt, S., Compo, G. P., Gray, S. L., Haigh, I. D., Hersbach, H., Kuijjer, K., Martínez-Alvarado, O., McColl, C., Schurer, A. P., Slivinski, L., and Williams, J.: Rescuing historical weather observations improves quantification of severe windstorm risks, Natural Hazards and Earth System Sciences, 23, 1465–1482, https://doi.org/10.5194/nhess-23-1465-2023, 2023.

Huang, S.-Y., Wu, S.-Y., Chen, Y.-J., Tsai, R. T.-H., and Fan, I.-C.: Climate event classification based on historical meteorological records and its presentation on a Spatio-Temporal research platform, Digital Scholarship in the Humanities, 37, 1022–1032, https://doi.org/10.1093/llc/fqab099, 2022.

Hulton, F. and Schultz, D. M.: Climatology of large hail in Europe: characteristics of the European Severe Weather Database, Natural Hazards and Earth System Sciences, 24, 1079–1098, https://doi.org/10.5194/nhess-24-1079-2024, 2024.

Kahraman, A., Kendon, E. J., and Fowler, H. J.: Climatology of severe hail potential in Europe based on a convection-permitting simulation, Climate Dynamics, https://doi.org/10.1007/s00382-024-07227-w, 2024.

Lenke, W.: Klimadaten von 1621-1650 Nach Beobachtungen des Landgrafen Hermann IV. von Hessen (Uranophilus Cyriandrus), Tech. Rep. 63, Deutscher Wetterdienst, Offenbach am Main, https://www.dwd.de/DE/leistungen/pbfb_verlag_berichte/pdf_einzelbaende/63_pdf.pdf?__blob=publicationFile&v=3, 1960.

Luterbacher, J., Allan, R., Wilkinson, C., Hawkins, E., Teleti, P., Lorrey, A., Brönnimann, S., Hechler, P., Velikou, K., and Xoplaki, E.: The Importance and Scientific Value of Long Weather and Climate Records; Examples of Historical Marine Data Efforts across the Globe, Climate, 12, 39, https://doi.org/10.3390/cli12030039, 2024.

Martius, O., Hering, A., Kunz, M., Manzato, A., Mohr, S., Nisi, L., and Trefalt, S.: Challenges and Recent Advances in Hail Research, Bulletin of the American Meteorological Society, 99, ES51–ES54, https://doi.org/10.1175/BAMS-D-17-0207.1, 2018.

Punge, H. and Kunz, M.: Hail observations and hailstorm characteristics in Europe: A review, Atmospheric Research, 176-177, 159–184, https://doi.org/10.1016/j.atmosres.2016.02.012, 2016.

Rohr, C.: Der Umgang mit Naturkatastrophen im Mittelalter. In: Christian Rohr (Koordinator), Krisen, Kriege, Katastrophen. Zum Umgang mit Angst und Bedrohung im Mittelalter. 9. Interdisziplinäre Ringvorlesung des Interdisziplinären Zentrums für Mittelalter-Studien, Salzburg, Wintersemester 2009/10, type: Other, 2009.

Sakaji, H. and Kaneda, N.: Indexing and Visualization of Climate Change Narratives Using BERT and Causal Extraction, in: 2023 IEEE International Conference on Big Data (BigData), pp. 5674–5683, https://doi.org/10.1109/BigData59044.2023.10386320, arXiv:2408.01745 [cs], 2023.

Schätz, F.: Voraussetzungen und Grenzen der Auswertung klimarelevanter Informationen historischer Textquellen mit Hilfe von Automatisierungsprozessen, Monographie, Universität Freiburg, Freiburg im Breisgau, https://freidok.uni-freiburg.de/data/244079, 2023.

Schätz, F. and Glaser, R.: Historical Climate Observations of Thunderstorms and Hail in Central Europe (1000–1900), https://doi.org/10.60493/834BD-MWW13, 2025.

https://doi.org/10.5194/egusphere-2026-199
Preprint. Discussion started: 6 February 2026
© Author(s) 2026. CC BY 4.0 License.

Stahl, K., Kohn, I., Blauhut, V., Urquijo, J., De Stefano, L., Acácio, V., Dias, S., Stagge, J. H., Tallaksen, L. M., Kampragou, E., Van Loon, A. F., Barker, L. J., Melsen, L. A., Bifulco, C., Musolino, D., de Carli, A., Massarutto, A., Assimacopoulos, D., and Van Lanen, H. A. J.: Impacts of European drought events: insights from an international database of text-based reports, Natural Hazards and Earth System Sciences, 16, 801–819, https://doi.org/10.5194/nhess-16-801-2016, 2016.

Taszarek, M., Allen, J., Púčik, T., Groenemeijer, P., Czernecki, B., Kolendowicz, L., Lagouvardos, K., Kotroni, V., and Schulz, W.: A Climatology of Thunderstorms across Europe from a Synthesis of Multiple Data Sources, Journal of Climate, 32, 1813–1837, https://doi.org/10.1175/JCLI-D-18-0372.1, 2019.

Tornado and Storm Research Organisation (TORRO): The TORRO Hail Intensity Scale (H-Scale), https://www.torro.org.uk/research/hail/hscale, 2025.

Webersinke, N., Kraus, M., Bingler, J. A., and Leippold, M.: ClimateBert: A Pretrained Language Model for Climate-Related Text, https://doi.org/10.48550/arXiv.2110.12010, arXiv:2110.12010 [cs], 2022.

White, S., Pfister, C., and Mauelshagen, F., eds.: The Palgrave handbook of climate history, Palgrave Macmillan, London, United Kingdom, ISBN 978-1-137-43019-9 978-1-349-68260-7, oCLC: ocn934193528, 2018.

Zhou, B., Zou, L., Mostafavi, A., Lin, B., Yang, M., Gharaibeh, N., Cai, H., Abedin, J., and Mandal, D.: VictimFinder: Harvesting rescue requests in disaster response from social media with BERT, Computers, Environment and Urban Systems, 95, 101 824, https://doi.org/10.1016/j.compenvurbsys.2022.101824, 2022.