https://doi.org/10.5194/egusphere-2026-4053
Preprint. Discussion started: 21 July 2026

From climatology to hazard: environmental regimes and tornado wind-speed exceedance in peninsular Spain and the Balearic Islands

Gonzalo Agurto Barragán[1], Javier Aroba Páez[2 3], Isidoro Gutiérrez-Álvarez[1 4], and Enrique Gutiérrez de San Miguel Herrera[1]

[1]Radiation and Environmental Physics Group (FRYMA), Research Centre on Natural Resources, Health and the Environment (RENSMA), University of Huelva, 21007, Huelva, Spain.
[2]Department of Information Technologies, University of Huelva, 21007, Huelva, Spain.
[3]Center for Advanced Studies in Physics, Mathematics and Computing (CEAFMC), University of Huelva, 21007, Huelva, Spain.
[4]Platform for Computational Medicine, Andalusian Public Foundation, Progress and Health-FPS, 41092, Seville, Spain.

Correspondence to: Gonzalo Agurto Barragán (gonzaloagurtob@gmail.com)

Abstract. Tornadoes and waterspouts are infrequent but locally damaging hazards in the Iberian–Mediterranean region, where uneven reporting and intensity uncertainty hamper their assessment. We compiled an updated 1992–2025 catalogue for peninsular Spain and the Balearic Islands by merging the national meteorological agency's tornado records with a pan-European severe-weather database, harmonising event times to UTC and removing duplicates, which yielded 1185 events: 567 tornadoes and 618 waterspouts. The record is temporally heterogeneous, with a late-summer to autumn maximum and a marked diurnal split between morning waterspouts and late-afternoon tornadoes; its rising counts appear to track reporting practices rather than any climatic trend. Using 54 pre-event atmospheric-reanalysis (ERA5) parameters, we tested a six-region geographical framework that proved only partially separable, so a simpler structure emerged: Ward clustering of 50 km hexagonal cells resolves three broad regimes — shear-dominated in the west and southwest, thermodynamically charged in the central-southeast, and weakly forced in the north and east. Adapting a footprint-based occurrence-to-exceedance model, we derived the first regime-dependent wind-speed exceedance estimates for the region. Notably, the least active regime holds four of the five Enhanced Fujita Scale 3 (EF3) tornadoes and dominates exceedance at 180 and 220 km h^{-1}, so where tornadoes are most frequent is not where the strongest winds are most likely. These results provide a first catalogue-based baseline for regime-dependent tornado and waterspout hazard assessment in peninsular Spain and the Balearic Islands.

1 Introduction

Tornadoes and waterspouts are low-frequency but potentially damaging severe convective hazards in Europe. Their climatological assessment is challenging because they are short-lived, spatially narrow phenomena whose detection depends strongly on observation density, post-event documentation, and reporting practices. In recent decades, European tornado climatologies have benefited from the development of severe-weather databases, public reports, and reanalysis-based environmental studies, but reporting heterogeneity and intensity uncertainty remain important limitations, particularly in regions where historical documentation is uneven (Dotzek et al., 2009; Groenemeijer and Kühne, 2014).

Previous work has shown that the Iberian Peninsula and the Balearic Islands host a heterogeneous tornado and waterspout climatology. Early national-scale compilations established this picture: Gayà (2011) characterised the spatial, temporal, and intensity distribution of tornadoes, waterspouts, and convective windstorms across Spain, and the national meteorological agency subsequently produced a systematic climatology for peninsular Spain and the Balearic Islands from a 10-year sample for 2003–2012 (AEMET, 2015). More recent work has focused on the environmental drivers most relevant to the present study. Rodríguez and Bech (2021) analysed 907 events for 1980–2018 and identified a broad northeast–southwest contrast in seasonality and environmental ingredients, with northeastern events generally associated with higher CAPE and lower shear/helicity and a relevant fraction of southwestern events occurring in high-shear, low-CAPE environments. Rodríguez and Lemus-Canovas (2023) further showed that Iberian tornadic storms can be grouped into multiple synoptic patterns, highlighting the diversity of large-scale configurations that may favour tornadogenesis.

Studies in other Mediterranean regions provide a useful analogue. Italian tornado studies have documented the role of low-level shear, deep-layer shear, CAPE, storm-relative helicity, sea-surface temperature anomalies, and coastal/orographic forcing in modulating tornado-prone environments (Ingrosso et al., 2020; Bagaglini et al., 2021; Avolio and Miglietta, 2021). They are a useful reference for Spain, where the Mediterranean coast, the Gulf of Cádiz, the Cantabrian coast, the Ebro valley, and major mountain systems can generate sharp spatial contrasts in moisture, instability, and low-level convergence.

For wind-hazard assessment, climatology alone is insufficient. A catalogue can describe where and when tornadoes and waterspouts have been reported, but exceedance analysis requires translating regional occurrence into the probability that a given wind-speed threshold is exceeded at a representative point or across gridded spatial units. Bridging that gap requires three connected elements: (i) an updated and internally consistent event database; (ii) a spatial framework that groups areas with comparable tornado-prone environments; and (iii) a physically traceable occurrence-to-exceedance model that accounts for tornado frequency, intensity, path length, path width, and the distribution of wind speeds within the damage swath.

A useful framework for this transition was proposed by Balbi and Barbieri (2018), who developed a footprint-based probabilistic approach for tornado hazard assessment in Argentina. The method links tornado occurrence rates with intensity-dependent path geometry and the internal distribution of wind speeds within the tornado footprint, allowing regional tornado climatology to be converted into wind-speed exceedance estimates. Such an approach is particularly useful where direct point-impact observations are scarce, since it yields exceedance probability curves, EP (υ,T), which express the probability of exceeding a wind-speed threshold υ within a given exposure period T. Here, we adapt this methodology to peninsular Spain and the Balearic Islands by combining an updated tornado and waterspout catalogue, environmental regionalization, tornado-path geometry, and a Poisson occurrence model.

https://doi.org/10.5194/egusphere-2026-4053
Preprint. Discussion started: 21 July 2026

The objectives of this study are to (1) present an updated 1992–2025 tornado and waterspout database for peninsular Spain and the Balearic Islands; (2) describe the main temporal, spatial, and intensity-related climatological features of the record; (3) assess whether the environmental information derived from ERA5 supports a meaningful regional organization of tornado-prone environments; and (4) develop a preliminary wind-speed exceedance framework, adapted from footprint-based probabilistic approaches, to estimate EP (υ,T) curves and threshold-based exceedance maps over the study domain.

2 Data and Methods

2.1 Study area and event database

The study domain comprises peninsular Spain and the Balearic Islands. The Canary Islands, Ceuta, and Melilla were excluded to keep a physically coherent Iberian–Mediterranean domain.

The catalogue was built by merging SINOBAS (singular atmospheric observations in Spain, maintained by AEMET) and the European Severe Weather Database (ESWD, maintained by ESSL), a reference source for European severe-weather climatologies (Dotzek et al., 2009). The ESWD contribution includes QC1 and QC2 events. Only tornadoes and waterspouts were retained. Meanwhile, other vortex-like or convective-wind phenomena (dust devils, gust vortices, gust fronts) were excluded.

The 1992 start year satisfies the 30-year minimum recommended by the WMO for climatological analyses (WMO, 2017) — the 1992–2025 record spans 34 years — and coincides with the start of the ESWD data provided by ESSL, giving a common baseline across both sources.

Event times were harmonized to UTC before merging, since SINOBAS/AEMET records are in local time and ESWD records in UTC. Then, local times were converted using the corresponding standard or daylight-saving time, prior to duplicate detection and ERA5 extraction. Following previous Iberian work, reports were treated as duplicate candidates when within 50 km and 2 h of one another (Rodríguez and Bech, 2021).

After harmonization, filtering, and duplicate removal, the final database (updated to 15 December 2025) contains 1185 tornado and waterspout events for 1992–2025, of which 567 are tornadoes and 618 waterspouts. Intensity was harmonized into SD, <EF0, EF0, EF1, EF2, and EF3. The SD class (reports without a determined Fujita/EF intensity) is retained for occurrence climatology. However, its exclusion is tested in the hazard calculations because the regime is dominated by waterspouts, which are rarely rated over water, whereas the rated classes correspond essentially to land-based tornadoes.

Table 1: Number of events in the updated 1992–2025 database by reported Fujita/Enhanced Fujita category. Counts include both tornadoes and waterspouts.

Intensity class	Number of events	Percentage
SD	701	59.2 %
<EF0	20	1.7 %
EF0	200	16.9 %
EF1	216	18.2 %

https://doi.org/10.5194/egusphere-2026-4053
Preprint. Discussion started: 21 July 2026

EF2	43	3.6 %
EF3	5	0.4 %
Total	**1185**	**100.0 %**

2.2 Climatological diagnostics and initial regional framework

The climatological analysis covers annual counts and the monthly and diurnal distributions in local time (LT). Because event times can be uncertain or rounded, the diurnal cycle was examined both with all records and after filtering by temporal uncertainty; the main interpretation uses reports with uncertainty ≤ 1 h, with a stricter ≤ 30 min comparison.

As an initial framework, the domain was divided into six a priori subregions: Mediterranean (MED), Interior Northeast (IntNE), Andalusian/Gulf of Cádiz–Alborán (AND), Interior Southwest (IntSW), Cantabrian (CAN), and Galician (GAL) (Fig. 1). This was not meant to define optimal clusters but to provide a physically motivated baseline, refining the broad NE–SW contrast of Rodríguez and Bech (2021) and the synoptic diversity reported by Rodríguez and Lemus-Canovas (2023), together with first-order controls on convective environments (moisture sources, coastal–inland contrasts, orography, and the spatial density of reports). It is therefore treated as a hypothesis to be tested against ERA5 information, not as an assumed optimum.

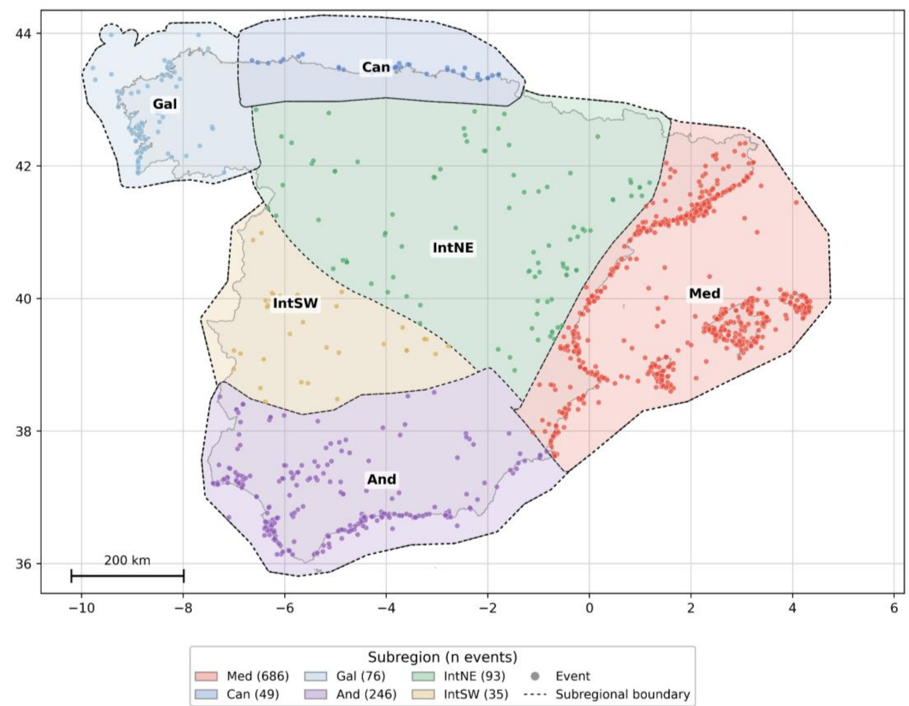

Figure 1: Preliminary subregional framework for peninsular Spain and the Balearic Islands. Colours indicate the six a priori subregions used in the analysis: Mediterranean (MED), Interior Northeast (IntNE), Andalusian/Gulf of Cádiz–Alborán (AND), Interior Southwest (IntSW), Cantabrian (CAN), and Galician (GAL). Points indicate

https://doi.org/10.5194/egusphere-2026-4053
Preprint. Discussion started: 21 July 2026

tornado and waterspout reports, and dashed lines indicate subregional boundaries. The map was generated using GeoPandas–Matplotlib from the subregional GeoJSON polygons, the event database, and official Spain–Balearic Islands geographical base layers.

2.3 ERA5 environmental predictors

Environments were characterized with ERA5 reanalysis (Hersbach et al., 2020). This reanalysis dataset provides hourly estimates of atmospheric variables on a 0.25º x 0.25º horizontal grid. For each event, hourly profiles of temperature, humidity, and wind variables were extracted from 18 h before to 6 h after the reported event time. To reduce sensitivity to location, timing, and single-grid-point representativeness, each profile was defined as the spatial mean of the four ERA5 grid points forming the 2 × 2 grid-point box enclosing the event location.

From these source variables, 54 tornado-relevant parameters were computed using the open-source sounding-analysis package SHARPpy (Blumberg et al., 2017), including thermodynamic, kinematic, and composite indices such as CAPE, lifted indices, storm-relative helicity, effective bulk wind difference, energy-helicity indices, significant tornado parameters, and WMAXSHEAR-type diagnostics (Thompson et al., 2003, 2007; Taszarek et al., 2017). Their full definitions, layers, and primary references are listed in Table S1 in the Supplement.

For the exploratory full-window analysis, hourly series were summarized with event-relative values (t = −6, 0, +6 h) and simple descriptors (mean, maximum, minimum, and net change between −6 and +6 h), following the proximity-environment literature (Púčik et al., 2015; Taszarek et al., 2017) while adding a compact representation of the short-term evolution around each event (Ingrosso et al., 2020; Rodríguez and Bech, 2021).

With respect to the predictive and regionalization stages, only pre-event information was retained to avoid information leakage, i.e. the use of data not legitimately available before the event time (Kaufman et al., 2012), consistent with recent tornado machine-learning work using pretornadic predictors (Sessa and Trapp, 2025). The pre-event summaries were the value at t = −6 h, the pre-event mean, maximum, minimum, last value, first-to-last change, standard deviation, and linear slope.

2.4 Event-based environmental clustering and supervised separability

We first tested whether the six a priori regions emerged as natural groupings in the multivariate space. All predictors were standardized and reduced with PCA, representing each event in an orthogonal space retaining the dominant modes of variance (Jolliffe and Cadima, 2016). Furthermore, the number of components followed a cumulative-variance criterion, keeping those explaining ≈ 90 % of the total variance (up to 10 components).

The machine-learning methods were diagnostic tools, not operational predictors. Three unsupervised approaches were applied to the event-level PCA scores: DBSCAN, which finds dense groups and outliers without a preset number of clusters (Ester et al., 1996); a grid-based diagnostic on the PC1–PC2 plane to detect local density maxima (Wang et al., 1997); and Ward hierarchical clustering, which forms compact groups by minimizing within-cluster variance (Ward, 1963; Murtagh and Legendre, 2014).

https://doi.org/10.5194/egusphere-2026-4053
Preprint. Discussion started: 21 July 2026

Candidate solutions were screened before interpretation: a candidate was usable only if it produced at least two non-noise clusters, retained sufficient coverage (low noise fraction, the share of unassigned events), and did not collapse into one dominant cluster (low dominant-cluster fraction, the share of clustered events in the largest cluster). These criteria were applied ahead of internal scores, which can be misleading when most events become noise or when a partition is technically valid but nearly single-cluster (Hennig, 2015).

A supervised step then quantified whether the six regional labels were distinguishable from pre-event predictors. This tested environmental separability, not forecast skill, since all cases are observed events. Standardized, PCA-reduced predictors were evaluated with a dummy baseline that ignores the ERA5 predictors and assigns the most frequent training-fold class, together with multinomial logistic regression, SVM-RBF (Cortes and Vapnik, 1995), a decision tree (Breiman et al., 1984), and Gaussian naive Bayes (James et al., 2021), under nested spatially grouped cross-validation with non-overlapping spatial blocks to limit spatial leakage (Kohavi, 1995; Roberts et al., 2017). Because the classes are multiclass and imbalanced, performance was summarized with Macro-F1, balanced accuracy and the Matthews correlation coefficient (MCC) (Sokolova and Lapalme, 2009; Brodersen et al., 2010; Matthews, 1975; Gorodkin, 2004; Jurman et al., 2012). To combine metrics defined on different ranges, the MCC was linearly normalized to the unit interval as nMCC = (MCC + 1)/2, following Chicco and Jurman (2020), and the three commensurable scores were aggregated into a single composite score by an unweighted mean. Reducing several normalized performance metrics to a single index for ranking is established practice in Earth-system model evaluation. Because such a single index can mask trade-offs among the individual metrics (Gleckler et al., 2008), the composite is used only as a transparent, parameter-free ranking criterion and the individual metrics are reported alongside it (Sect. 3.2). A dummy-baseline skill score, (composite − dummy) / (1 − dummy), is reported in parallel.

2.5 Objective spatial subregionalization

The second stage shifted from event-level regimes to objective spatial regionalization, delineating spatial units with similar pre-event signatures rather than imposing geographical boundaries (Zhang et al., 2016).

A regular hexagonal grid was generated over the domain; hexagons give a more uniform neighbourhood and more compact geometry than square cells, favourable for spatial aggregation and connectivity analyses (Birch et al., 2007; Majewski et al., 2002). Because ERA5 has a native spacing of 0.25°, the cell size was set comparable to but slightly coarser than the reanalysis scale, with sensitivity tested at approximate hexagon sizes of 35, 50, and 65 km.

Each event was assigned to its hexagon, and each occupied cell was summarized by the median of its events' pre-event predictors, a robust statistic that limits the influence of extreme values where event counts per cell are uneven (Huber and Ronchetti, 2009). Missing values were imputed feature-wise with the median, preserving central tendency without adding extremes (Little and Rubin, 2019); the matrix was then standardized and PCA-reduced before clustering.

Ward hierarchical clustering with a spatial-connectivity constraint was applied to the PCA scores, the constraint preventing geographically fragmented regions. Candidate partitions were generated for k = 2–6 across the tested hexagon sizes, with the six-region framework kept as a baseline.

Rather than selecting k automatically from a single internal criterion such as the gap statistic (Tibshirani et al., 2001), candidates were compared under a common protocol: regional labels were predicted from pre-event predictors and scored against a dummy baseline with a composite of Macro-F1, balanced accuracy, and normalized MCC defined in Sect. 2.4, plus the dummy-relative skill score (model − dummy) / (1 − dummy), which separates absolute performance from improvement over a trivial reference. Spatial cross-validation with 180 km blocks reduced inflation from spatial autocorrelation, and nested validation reduced optimistic bias (Cawley and Talbot, 2010; Roberts et al., 2017).

The final partition was not chosen by a single score but by combining composite performance, skill, temporal robustness, interpretability, parsimony, and suitability for the exceedance analysis.

2.6 Occurrence, return period, and point-scale exceedance framework

The probabilistic analysis was run over the official Spain-Balearic geographic domain clipped to the data. The regionalization was derived from the combined tornado and waterspout sample, but the exceedance calculations were restricted to tornadoes: waterspouts share physical drivers and improve the spatial coverage of ERA5 predictors, yet their overland wind speeds are not measurable from the catalogue. SD events were also excluded from the class-specific footprint and exceedance calculation (Eqs. 2–4), as they cannot be assigned to a wind-speed band without an arbitrary intensity distribution; in practice they correspond to amateur reports, distant waterspout sightings, or cases without a damage survey (Doswell and Burgess, 1988).

Each tornado was assigned to a regime, and regime areas were obtained by dissolving the clipped hexagonal partition in an equal-area projection. For each regime *r* and class *k*, the annual areal occurrence rate is given by Eq. (1):

$$\mu_{r,k} = \frac{N_{r,k}}{A_r \, n_y}, \qquad (1)$$

where $N_{r,k}$ is the count, A_r the regime area, and n_y the number of years. This occurrence model was linked to the footprint-based framework of Balbi and Barbieri (2018), in which climatology is translated into wind-speed exceedance through occurrence rates, path geometry, and the internal distribution of damaging winds.

Path length and mean width were modelled by intensity class with Weibull distributions, fitted by maximum likelihood where the class had at least 7 valid observations. In total, 52 catalogue events had usable path-dimension information, distributed as SD (N = 7), EF0 (N = 17), EF1 (N = 18), and EF2 (N = 7), plus EF3 (N = 3); no <EF0 report had usable dimensions. Local maximum-likelihood estimation (MLE) fits were obtained for the four classes reaching the N ≥ 7 threshold (SD, EF0, EF1, EF2). The SD fit was retained only as a path-

https://doi.org/10.5194/egusphere-2026-4053
Preprint. Discussion started: 21 July 2026

geometry diagnostic and was not propagated into the wind-speed exceedance calculation, because SD events lack an assigned intensity class. Reported path dimensions are strongly rounded — typically given to the nearest kilometre for length and as coarse round figures for width, rather than as precise survey measurements — so the small EF1 and EF2 samples overlap substantially. For EF3, whose three usable reports fell below this threshold, geometry instead followed the intensity–path relationships of Brooks (2004), scaled from the local EF2 fit; <EF0 was scaled from EF0 with fixed ratio factors.

The expected footprint area $E\,[A_k]$ was estimated by Monte Carlo integration over the fitted path-length and width distributions (Robert and Casella, 2010; Metropolis and Ulam, 1949). For each intensity class, N = 200 000 draws of path length and width were generated from the fitted Weibull distributions using a fixed seed for reproducibility. The footprint area for each draw was computed as L × W and $E\,[A_k]$ taken as the sample mean of these areas. The expected area exceeding a threshold υ is given by Eq. (2):

$$E[A_{exc,k}(\upsilon)] = E[A_k]\, f_k(\upsilon)\ , \qquad (2)$$

where $f_k(\upsilon)$ is the footprint fraction with local wind speed above υ, taken from the area-variability matrix of Balbi and Barbieri (2018), itself based on the NUREG/CR-4461 framework of Ramsdell and Rishel (2007); within an EF band, $f_k(\upsilon)$ was linearly interpolated.

The annual exceedance rate and the probability of at least one exceedance in a time horizon T, under a Poisson occurrence model, are given respectively by Eqs. (3) and (4):

$$\lambda_r(\upsilon) = \textstyle\sum_k \mu_{r,k} E[A_{exc,k}(\upsilon)]\ , \qquad (3)$$

$$EP_r(\upsilon, T) = 1 - exp[-\lambda_r(\upsilon)T]\ . \qquad (4)$$

A homogeneous Poisson process was adopted because tornado occurrence within a regime over the 34-year record can reasonably be treated as rare and approximately independent events with a locally constant rate, consistent with standard practice in regional tornado-occurrence modelling (Jagger et al., 2015; Cheng et al., 2015).

Fixed thresholds of 144, 180, and 220 km h^{-1} were used as EF1-, EF2-, and EF3-like references (approximately the lower bounds of each band), with continuous curves over a velocity grid. The exceedance probability is a hazard metric conditional on each regime's observed climatology, not a deterministic forecast.

As a complementary occurrence metric, return periods were computed for a standard neighbourhood of radius 50 km centred on each hexagonal cell ($A_{50} = \pi 50^2$ km²) as $TR_{r,50}(\upsilon) = 1/\lambda_{r,50}(\upsilon)$. The neighbourhood return period and the point-scale EP (υ, T) are complementary but not interchangeable: the point-scale probability is weighted by footprint area and increases with intensity, whereas the return period counts threshold-capable events within A_{50} regardless of footprint. A regime with fewer but more intense tornadoes can therefore show a longer return period yet a higher point-scale exceedance for the same threshold. The neighbourhood

https://doi.org/10.5194/egusphere-2026-4053
Preprint. Discussion started: 21 July 2026

formulation reflects the resolution of the hexagonal regionalization and matches the fixed-radius interpretation of operational severe-convective products.

Uncertainty was quantified by bootstrap resampling of years (1992–2025) with replacement, recomputing the regime rates per replicate (Efron and Tibshirani, 1993) and retaining the median and 5th–95th percentile range. Finally, threshold-based EP (υ, T) maps were produced on the hexagonal grid, with hexagon rates smoothed by a Gamma–Poisson empirical-Bayes formulation whose prior mean equals the regime rate: $\hat{\lambda}_h = (n_h + \lambda_r T_0) / (T_h + T_0)$. A prior exposure $T_0 = 10$ years was adopted as an intermediate regularisation level, giving approximately ≈ 23 % weight to the regime prior and 77 % to the observed count over the 34-year record. This regularisation reduces artificial zero estimates in empty cells where the corresponding regime rate is non-zero, while preserving most of the local spatial signal (Jagger et al., 2015; Cheng et al., 2015).

2.7 Computational implementation

All processing — harmonization, filtering, UTC standardization, duplicate detection, climatological diagnostics, ERA5 extraction, spatial processing, and the machine-learning analyses (PCA, clustering, supervised classification, cross-validation, metrics) — was implemented in Python with Jupyter notebooks (Kluyver et al., 2016), SHARPpy (Blumberg et al., 2017) for sounding-derived parameter computation and scikit-learn (Pedregosa et al., 2011), keeping the workflow reproducible. Ward clustering follows the minimum-variance criterion of Ward (1963) as clarified by Murtagh and Legendre (2014). ERA5 hourly profiles were downloaded and the 54 event-level parameters (Table S1) were computed using the HPC cluster of the "*Centro de Estudios Avanzados en Física, Matemáticas y Computación*" (CEAFMC), University of Huelva.

3 Results

3.1. Updated event catalogue and temporal occurrence patterns

The updated catalogue contains 1185 tornado and waterspout reports over peninsular Spain and the Balearic Islands for 1992–2025 after time harmonisation, filtering, and duplicate removal (Table 1). Annual counts show marked interannual variability rather than a simple monotonic behaviour (Fig. 2a). Specifically, the lowest annual totals are concentrated in the early part of the record, whereas larger totals occur mainly after the early 2000s, with particularly active years such as 2010 and 2018. Such an increase should be read with caution: European severe-weather databases are strongly affected by changes in reporting density, public awareness, documentation practices, and database consolidation (Dotzek et al., 2009; Gayà, 2011; Groenemeijer and Kühne, 2014). We therefore use the annual series to describe the observed record, not to infer a climatic trend in the physical occurrence of tornadoes or waterspouts.

Analysis of the monthly distribution reveals a pronounced late-summer to autumn maximum is revealed, although events are reported throughout the year (Fig. 2b). The combined sample reaches its highest frequency in September and remains high in October and November, with an increase already evident in August. Waterspouts account for a large fraction of the late-summer and early-autumn maximum, especially in September, whereas tornado reports display a broader seasonal distribution, with enhanced activity in spring

https://doi.org/10.5194/egusphere-2026-4053
Preprint. Discussion started: 21 July 2026

and autumn and a marked maximum in October. This seasonal behaviour is consistent with the heterogeneous Iberian and western Mediterranean setting, where coastal moisture sources, sea–land thermal contrasts, orography, and synoptic forcing combine differently across regions (Rodríguez and Bech, 2021; Rodríguez and Lemus-Canovas, 2023). The autumn maximum is also consistent with Mediterranean coastal environments, where warm sea-surface conditions and early-season upper-level disturbances can favour deep moist convection and low-level convergence (Ingrosso et al., 2020; Bagaglini et al., 2021; Avolio and Miglietta, 2021).

The filtered diurnal distribution further indicates that tornadoes and waterspouts have distinct timing characteristics (Fig. 2c). Using only reports with temporal uncertainty equal to or lower than 1 h, the combined sample is concentrated mainly from morning to late afternoon, with relative maxima around mid-morning and late afternoon. Waterspouts show their highest frequencies from mid-morning to early afternoon, whereas tornado reports peak later, mainly during the late afternoon. The contrast is physically plausible: land-based tornado reports are more directly linked to the daytime development of boundary-layer instability, while waterspouts and coastal convective vortices may also occur earlier over marine or near-coastal environments. The stricter 30 min time-filtered sample preserves the same broad contrast between morning–early afternoon waterspout activity and later tornado activity (Fig. S1c in the Supplement). By contrast, the unfiltered distribution shows an anomalously strong noon peak, especially for waterspouts (Fig. S1a in the Supplement), suggesting that records with uncertain or rounded times can distort the apparent diurnal cycle. For this reason, the filtered records provide the basis for the interpretation of the diurnal climatology.

https://doi.org/10.5194/egusphere-2026-4053
Preprint. Discussion started: 21 July 2026

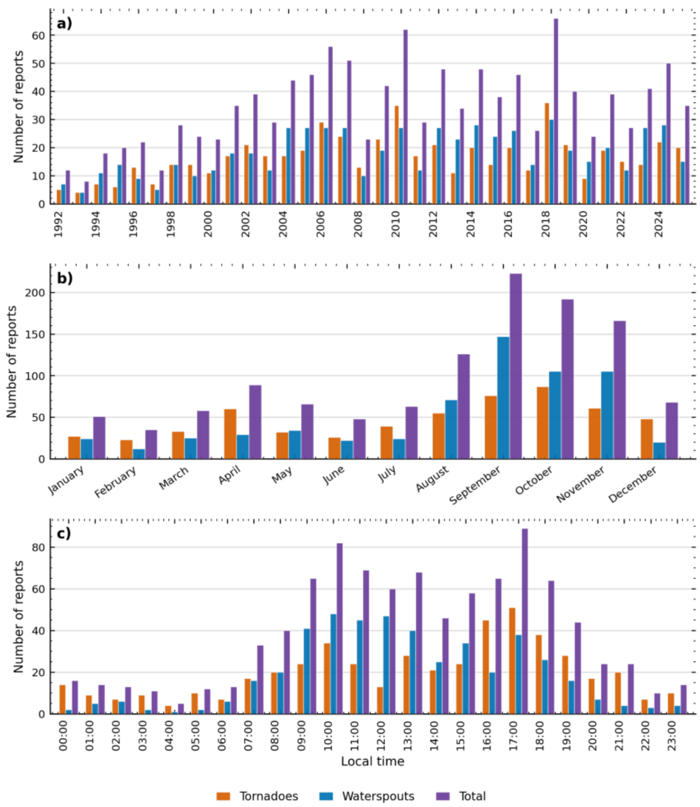

Figure 2. Temporal distribution of tornado and waterspout reports in the updated catalogue for peninsular Spain and the Balearic Islands during 1992–2025. (a) Annual number of tornadoes, waterspouts, and combined events. (b) Monthly distribution aggregated over the full study period. (c) Diurnal distribution in local time using only reports with temporal uncertainty equal to or lower than 1 h. Orange, blue, and purple bars indicate tornadoes, waterspouts, and the combined sample, respectively.

The record is thus strongly heterogeneous in time: interannual swings are large, activity peaks in autumn, and tornadoes and waterspouts differ in their daily timing. A single homogeneous temporal regime cannot represent the whole domain, which motivates the regional and environmental analyses that follow.

3.2. Event-level environmental separability of the initial subregional framework

The ERA5-derived environmental dataset provided a high-dimensional description of the pre-event conditions associated with the tornado and waterspout reports. The 54 tornado-relevant parameters listed in Table S1 include thermodynamic, kinematic, and composite diagnostics, and were summarized using pre-event descriptors in order to avoid information leakage. We first used this predictor set to test whether the six a priori subregions defined in Sect. 2.2 were also distinguishable in the multivariate environmental space.

The projection of the event-level PCA scores onto the first two principal components shows substantial overlap among the six subregions (Fig. 3a). Although some broad spatial–environmental tendencies are apparent, the

https://doi.org/10.5194/egusphere-2026-4053
Preprint. Discussion started: 21 July 2026

regional labels do not form compact or well-separated clusters in the reduced environmental space. Events from the Mediterranean region occupy a wide part of the diagram, while several inland and Atlantic-facing subregions overlap strongly with one another. The overlap indicates that the initial geographical subdivision captures part of the spatial organization of the record but does not correspond to six clearly separated environmental classes.

The supervised separability test confirms this interpretation. The best-performing classifier was the SVM-RBF model, which obtained a composite score of approximately 0.54, compared with 0.26 for the dummy baseline that ignores the ERA5 predictors (Fig. 3b). The model also outperformed the dummy reference in Macro-F1, balanced accuracy, and rescaled MCC, indicating that the ERA5 predictors contain a non-random regional signal. However, the absolute performance remained moderate, with Macro-F1 and balanced accuracy below 0.5. The six a priori regions are only partially recoverable from the pre-event environmental predictors.

The normalized confusion matrix further shows that this environmental separability is uneven across subregions (Fig. 3c). The Mediterranean region was the most clearly identified class, with 79 % of observed Mediterranean events classified correctly. Interior Northeast events also showed moderate separability, with 58 % correct classification. By contrast, Galicia and Interior Southwest were poorly recovered, with correct-class fractions of only about 24 % and 23 %, respectively. Both were frequently assigned to neighbouring or broader environmental categories, especially Mediterranean, Andalusian/Gulf of Cádiz–Alborán, or Cantabrian classes. The Andalusian/Gulf of Cádiz–Alborán and Cantabrian regions showed intermediate behaviour, with correct-class fractions close to 45–47 %.

The comparison among classifiers supports the same conclusion (Fig. S2 in the Supplement). SVM-RBF produced the highest composite score, followed closely by logistic regression, whereas Gaussian naive Bayes and the decision tree performed less well. The relatively small separation between the two best models suggests that the result is not driven by a single algorithmic choice. Instead, it reflects a moderate but incomplete environmental organization of the six-region framework.

https://doi.org/10.5194/egusphere-2026-4053
Preprint. Discussion started: 21 July 2026

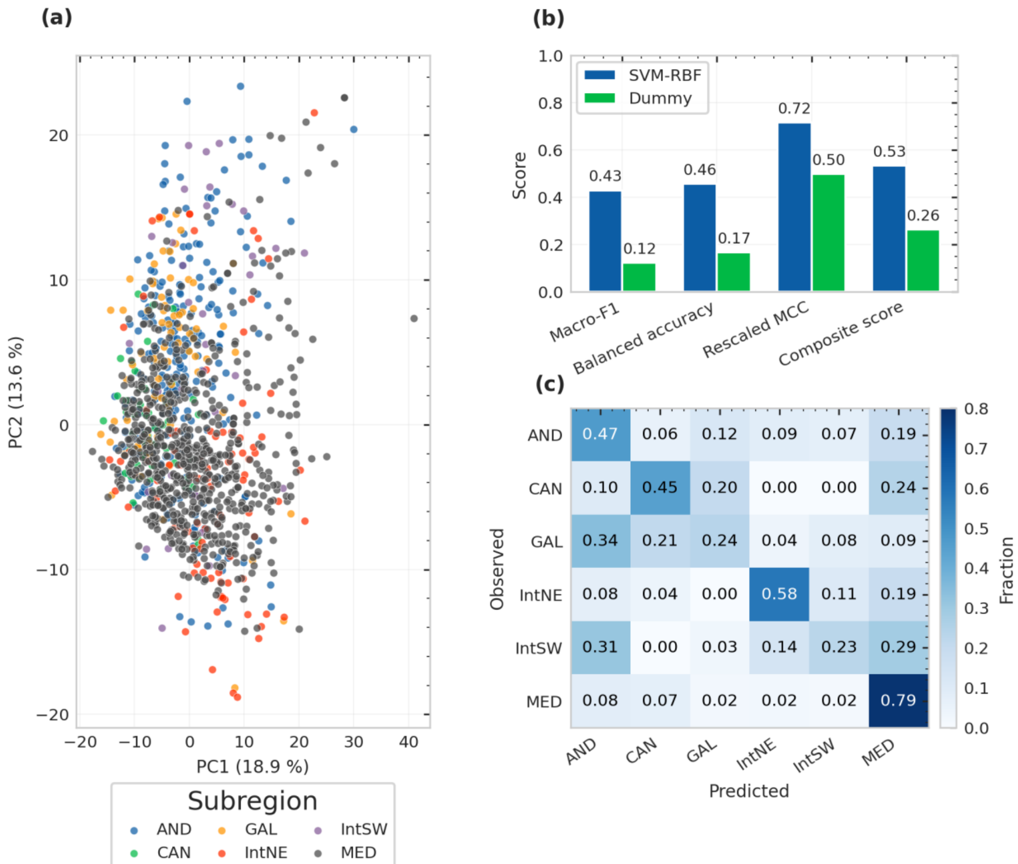

Figure 3. Event-level environmental separability of the six a priori subregions. (a) Projection of tornado and waterspout events onto the first two principal components of the standardized ERA5-derived environmental predictor space, coloured by a priori subregion. (b) Performance of the best supervised classifier compared with a dummy baseline using Macro-F1, balanced accuracy, normalized Matthews correlation coefficient (MCC), and the composite score. The MCC was linearly rescaled to the [0, 1] interval before computing the composite score. (c) Row-normalized confusion matrix for the best classifier, with observed subregions shown on the y axis and predicted subregions on the x axis. Values indicate the fraction of events in each observed subregion assigned to each predicted class.

Unsupervised clustering of event-level PCA scores was then used as an additional diagnostic. DBSCAN, grid-based clustering, and Ward hierarchical clustering were screened using simple validity criteria requiring at least two non-noise clusters, sufficient event coverage, and no collapse into a single dominant cluster (Table S2 and Fig. S3 in the Supplement). DBSCAN did not produce any valid candidate among the 90 tested configurations; its best raw candidate retained most events but was dominated by a single non-noise cluster. The grid-based method also produced no valid candidate among the 20 tested configurations; its best raw candidate identified two dense groups but classified most events as noise, indicating poor event coverage. Ward clustering was the only method that consistently produced valid, complete, and interpretable partitions, with all 30 tested configurations passing the screening criteria. The rejection of DBSCAN and grid-based clustering was based

https://doi.org/10.5194/egusphere-2026-4053
Preprint. Discussion started: 21 July 2026

on basic validity diagnostics rather than on internal clustering scores alone: DBSCAN tended to collapse into one dominant cluster, whereas the grid-based method retained only small dense subsets of the event space.

The four diagnostics agree: the six a priori subregions are a useful geographical hypothesis but too detailed to capture the dominant environmental structure, which is better described by a smaller number of broader regimes. That motivates the objective spatial regionalization developed next, where Ward clustering is applied to hexagonal-cell signatures rather than to the six imposed regions.

3.3. Objective spatial subregionalization

The objective spatial regionalization was derived from Ward clustering applied to hexagon-level pre-event environmental signatures. Among the tested hexagon sizes, the 50 km grid produced the highest composite score, indicating the best compromise between spatial detail, event aggregation, and environmental separability (Fig. 4a). The two-region solution achieved the highest absolute composite score, but it mainly reproduced the first-order NE–SW contrast previously described for Iberian tornadic environments (Rodríguez and Bech, 2021). The three-region solution retained competitive performance while adding a geographically intermediate third regime with its own distinct environmental signature and showed the strongest dummy-relative skill for the tornado-only subset (Fig. 4b, c; the full per-metric breakdown by k is given in Fig. S4). Solutions with more regions increased fragmentation without a clear compensating gain. The Ward 50 km k = 3 partition was retained as the final objective regionalization.

The selected partition defines three spatially coherent regimes over peninsular Spain and the Balearic Islands (Fig. 5). Regime 0 covers most of the northern and eastern sectors, including the Balearic Islands; regime 1 is concentrated over the western and southwestern peninsula; and regime 2 forms a central-to-southeastern transitional belt between them. The raw occupied-hexagon solution already shows this tripartite structure, while the completed and officially clipped maps provide the continuous spatial support used in the subsequent exceedance calculations (Fig. 5). The result refines, rather than replaces, the broad NE–SW environmental contrast reported by Rodríguez and Bech (2021), and is consistent with the diversity of synoptic settings documented for Iberian tornadic storms by Rodríguez and Lemus-Canovas (2023).

https://doi.org/10.5194/egusphere-2026-4053
Preprint. Discussion started: 21 July 2026

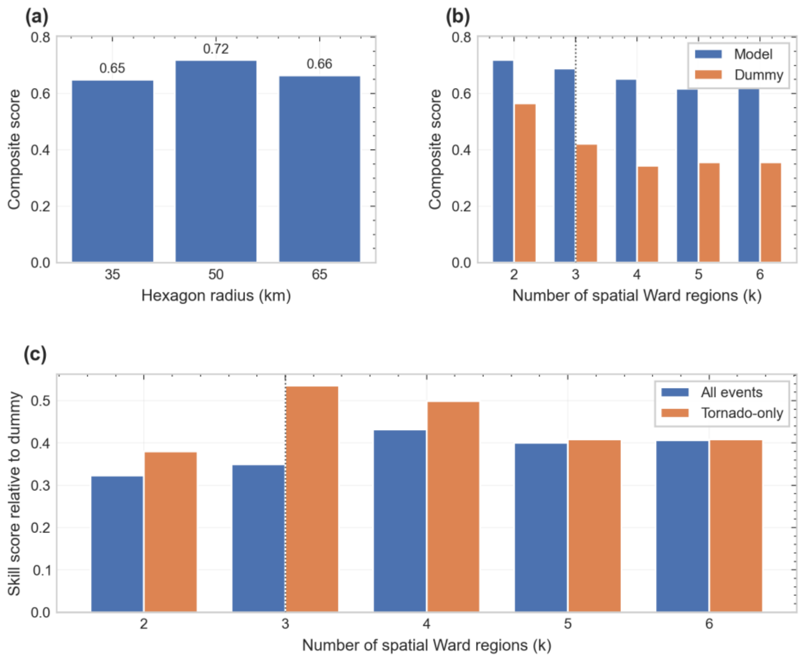

Figure 4. Comparative evaluation of candidate Ward spatial regionalizations. (a) Composite score for the tested hexagon sizes. (b) Composite score of Ward solutions with different numbers of regions (k=2–6), compared with the corresponding dummy baseline. (c) Skill score relative to the dummy baseline for the full event sample and the tornado-only subset. The final regionalization retained in this study corresponds to the 50 km, k=3 solution.

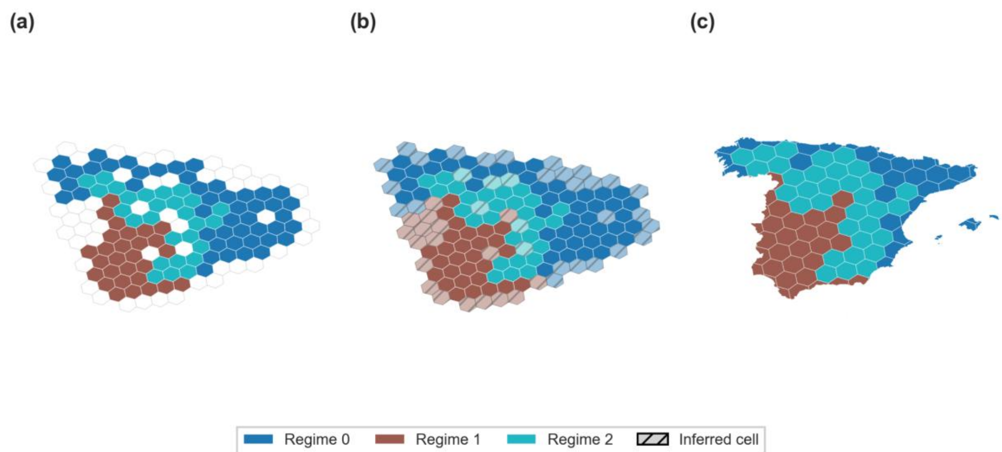

Figure 5. Construction of the final objective spatial regionalization. (a) Raw Ward 50 km k=3 partition for occupied hexagons only. (b) Full hexagonal partition after neighbour-based inference of labels for previously empty cells; inferred cells are indicated by hatching. (c) Final three-regime regionalization clipped to the official domain of peninsular Spain and the Balearic Islands.

https://doi.org/10.5194/egusphere-2026-4053
Preprint. Discussion started: 21 July 2026

The retained partition is also environmentally separable from event-level ERA5 predictors. For the full event sample, the supervised classifier outperformed the dummy baseline in Macro-F1, balanced accuracy, and rescaled MCC (Fig. 6b). The normalized confusion matrix shows that regime 0 is the most clearly recovered, with 84 % of its observed events correctly classified, whereas regimes 1 and 2 are recovered more moderately, at 61 % and 58 %, respectively (Fig. 6a). Critically, the confusion in both cases runs almost entirely toward regime 0 (38 % of regime 1 and 37 % of regime 2 events are misclassified as regime 0) and almost never between regimes 1 and 2 (1 % and 5 %, respectively). Regimes 1 and 2 are therefore well separated from each other, each partially overlapping only with the comparatively weak environmental signal of regime 0, rather than regime 2 sitting on a continuum between regimes 0 and 1.

Bootstrap resampling yielded a median composite score of 0.670, with a 5th–95th percentile range of 0.622–0.714 (Fig. S5 in the Supplement), consistent with the composite score obtained from the full sample and indicating that the retained partition's separability is not an artifact of any particular subset of years. The tornado-only validation reproduces the same pattern: the model again outperformed the dummy baseline (Macro-F1 0.66 vs. 0.17; balanced accuracy 0.67 vs. 0.23; normalized MCC 0.70 vs. 0.50), and the tornado-only confusion matrix closely mirrors the full-sample one, with regime 0 best recovered (75 %) and regimes 1 and 2 (70 % and 59 %) again confused mainly with regime 0 rather than with each other (Fig. S6 in the Supplement).

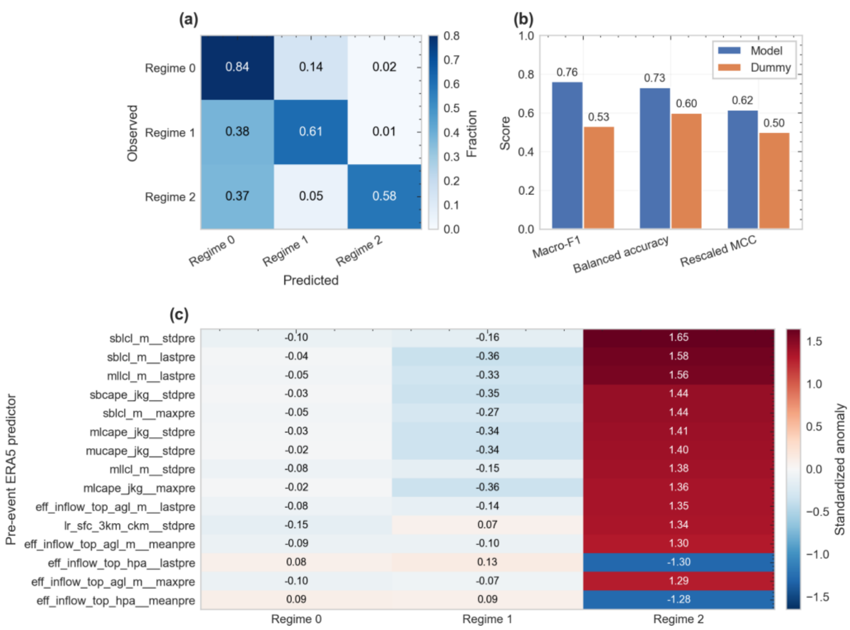

Figure 6. Environmental separability and signature of the retained Ward 50 km k=3 regionalization. (a) Normalized confusion matrix obtained when regional labels are predicted from event-level pre-event ERA5 predictors. (b) Model performance compared with the dummy baseline. (c) Standardized anomalies of the leading

https://doi.org/10.5194/egusphere-2026-4053
Preprint. Discussion started: 21 July 2026

ERA5 predictors for each regime, showing the contrasting environmental signatures that define the final objective regionalization.

The environmental signatures of the three regimes are physically interpretable. The predictors with the largest combined range across the three regimes are dominated by thermodynamic and composite severe-weather diagnostics — CAPE-, LCL-, and effective-inflow-related variables — reflecting regime 2's outsized departure from the full-sample mean (Fig. 6c). Examining the leading predictors regime by regime reveals the full contrast (Fig. S7 in the Supplement): regime 0 shows generally negative standardized anomalies across a range of wind-, shear-, and composite-index predictors relative to the full event sample (Fig. S7a); regime 1 shows the strongest kinematic support of the three regimes, with pronounced positive anomalies in low- to mid-level wind and storm-relative-helicity variables (Fig. S7b); and regime 2 shows by far the largest anomalies of the three regimes, consistent with its dominance of the combined ranking in Fig. 6c (Fig. S7c). These contrasts are consistent with previous Iberian and Mediterranean studies showing that tornado-prone environments may arise from different combinations of instability, vertical shear, moisture, and synoptic forcing (Ingrosso et al., 2020; Bagaglini et al., 2021; Avolio and Miglietta, 2021; Rodríguez and Bech, 2021).

3.4. Occurrence, return period, and point-scale exceedance

By utilizing the Ward 50 km k = 3 environmental regionalization derived from the combined tornado and waterspout sample, the tornado-only exceedance analysis shows that the three regimes differ not only in total tornado occurrence, but also in how occurrence translates into wind-speed exceedance. Within the official clipped domain, regime 0 covers 132 734 km^2, regime 1 covers 152 244 km^2, and regime 2 covers 213 653 km^2. The corresponding domain-assigned tornado counts (all intensity classes, including SD) are 254, 189, and 74 events, respectively, implying total areal occurrence rates of approximately $5.6\ x\ 10^{-5}$, $3.7\ x\ 10^{-5}$, and $1.0\ x\ 10^{-5}$ km^{-2} yr^{-1}. Thus, regime 0 and regime 1 dominate the overall tornado occurrence signal, whereas regime 2 is less frequent but contains a proportionally larger contribution from the highest reported intensities.

Much of this amplification is driven by the fitted footprint geometry, which strongly increases the contribution of the most intense classes to exceedance at high wind-speed thresholds. The expected tornado footprint area increases from 0.09 km^2 for EF0 to 0.92 km^2 for EF1, 1.31 km^2 for EF2, and 11.50 km^2 for EF3. Consequently, the exceedance curves decrease rapidly with increasing wind-speed threshold, but not at the same rate in all regimes (Fig. 7a). For EF1-like winds, represented here by 144 km h^{-1}, the 30-year point-scale exceedance probability is highest in regime 0 ($1.8\ x\ 10^{-4}$), followed by regime 1 ($1.4\ x\ 10^{-4}$) and regime 2 ($1.1\ x\ 10^{-4}$; Fig. 7b). At 180 km h^{-1}, the ranking changes: regime 2 reaches the highest value ($4.1\ x\ 10^{-5}$), followed by regime 0 ($3.2\ x\ 10^{-5}$) and regime 1 ($1.3\ x\ 10^{-5}$). At 220 km h^{-1}, the signal is concentrated in regime 2 ($1.2\ x\ 10^{-5}$) and, to a lesser extent, regime 0 ($5.0\ x\ 10^{-6}$), whereas regime 1 has no non-zero estimate because no EF3 tornadoes were recorded there in the analysed sample. This should not be interpreted as zero physical hazard, but as the direct consequence of the finite observed catalogue and the decision not to impose an arbitrary EF3 rate where none was observed. Accordingly, the 220 km h^{-1} results rest on only five EF3 reports and on

https://doi.org/10.5194/egusphere-2026-4053
Preprint. Discussion started: 21 July 2026

the Brooks (2004) intensity–path scaling, which explains the very wide bootstrap intervals at this threshold and warrants cautious interpretation.

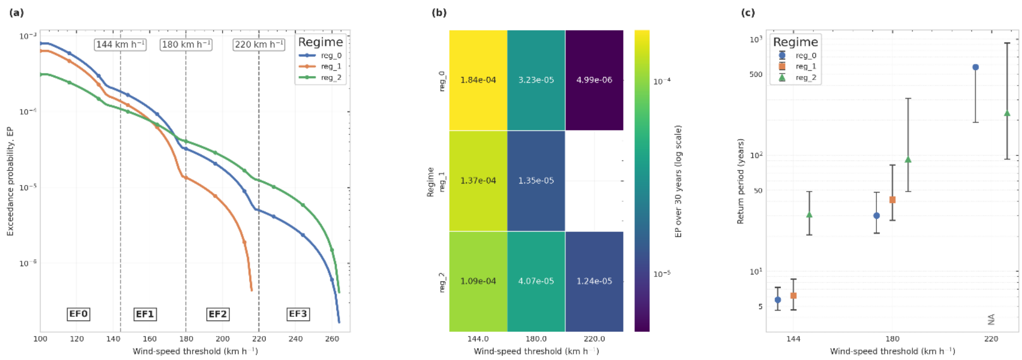

Figure 7. Regime-level tornado exceedance diagnostics computed for the Ward 50 km k = 3 regionalization derived from the combined tornado and waterspout sample. (a) Point-scale exceedance-probability curves EP(v,T) for a 30-year time horizon as a function of wind-speed threshold. Vertical dashed lines mark the reference thresholds of 144, 180, and 220 km h^{-1}, approximately corresponding to EF1-, EF2-, and EF3-like wind speeds. (b) Fixed-threshold 30-year tornado exceedance probabilities by regime. Colours are shown on a logarithmic scale. (c) Bootstrap median return period for threshold-capable tornado occurrence within a 50-km neighbourhood. Points indicate the median and whiskers indicate the 5th–95th percentile range from annual bootstrap resampling.

Complementing this analysis, 50-km neighbourhood return-period estimates provide an occurrence-based view (Fig. 7c and Table S3 in the Supplement). For the 144 km h^{-1} threshold, median return periods are short in regime 0 and regime 1, around 5.7 and 6.2 years, respectively, but increase to 30.8 years in regime 2. For 180 km h^{-1}, the median return periods increase to 30.2 years in regime 0, 41.2 years in regime 1, and 92.5 years in regime 2. For 220 km h^{-1}, only regime 0 and regime 2 produce finite median estimates, with values of 574.6 and 231.2 years, respectively; regime 1 remains undefined because of the absence of EF3 cases. The bootstrap intervals are particularly wide for the highest threshold, reflecting the strong sampling uncertainty associated with very few EF3 events.

To visualize how these regime-level differences project onto the spatial domain, empirical-Bayes hexagonal maps of point-scale EP (υ, T) are presented in Fig. 8. For 144 km h^{-1}, exceedance probabilities are spatially widespread, with local maxima mainly controlled by cells with stronger observed tornado occurrence. At 180 km h^{-1}, the signal becomes more localized and the contrast between regimes increases. At 220 km h^{-1}, exceedance probabilities are sparse and mainly confined to cells associated with the regimes where EF3 events were recorded. The filled-grid maps provide a continuous hazard surface over the official domain, while the observed-cell maps confirm where the estimates are directly supported by observed tornado records (Fig. S8 in the Supplement). Crucially, the most frequent tornado environments are not necessarily those producing the largest high-threshold exceedance probabilities: regime 0 and regime 1 dominate moderate-threshold occurrence, whereas regime 2 becomes disproportionately important for severe wind-speed exceedance.

https://doi.org/10.5194/egusphere-2026-4053
Preprint. Discussion started: 21 July 2026

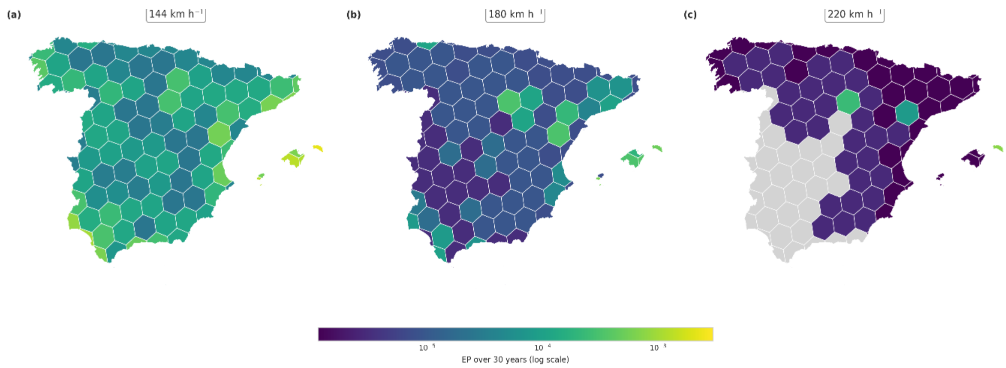

Figure 8. Hexagon-level point-scale 30-year tornado exceedance probability for wind-speed thresholds of (a) 144, (b) 180, and (c) 220 km h^{-1}. Estimates were obtained by applying the footprint-based point-scale exceedance model to empirical-Bayes-smoothed hexagon-level tornado rates, shrunk toward the corresponding Ward-regime prior with an equivalent prior exposure of 10 years. Colours are shown on a logarithmic scale. Grey cells indicate zero estimates that cannot be represented on the logarithmic colour scale. The maps represent the filled official-domain grid after propagating the final Ward-regime labels to previously empty cells.

The resulting probabilities should be interpreted as preliminary climatological hazard estimates conditional on the 1992–2025 tornado catalogue, the adopted EF-scale wind thresholds, and the footprint-based exceedance formulation of Ramsdell and Rishel (2007) and Balbi and Barbieri (2018). They are therefore not deterministic forecasts, but a first spatially explicit translation of the updated Iberian tornado climatology into regime-dependent wind-speed exceedance metrics.

4 Discussion

The updated catalogue reinforces a picture of strong temporal heterogeneity rather than a single dominant regime. The September–November maximum, and in particular the waterspout-driven September peak, matches the seasonal behaviour reported for the western and central Mediterranean, where warm late-summer sea-surface temperatures and the first autumn upper-level troughs favour deep moist convection and low-level convergence near the coast (Ingrosso et al., 2020; Bagaglini et al., 2021; Avolio and Miglietta, 2021). Land tornadoes peak later in the year and later in the day than waterspouts, with a clear late-afternoon mode. The split is physically reasonable: tornadoes over land depend on daytime heating of the boundary layer and the destabilization that follows, whereas waterspouts form over a marine boundary layer that is far less sensitive to the diurnal cycle and can develop under weak synoptic forcing through the morning. Rising counts after the early 2000s are better read as improved reporting than as a physical trend, as has been stressed for European severe-weather databases more generally (Dotzek et al., 2009; Groenemeijer and Kühne, 2014); we therefore treat the annual series as descriptive.

The environmental analysis shows that the six a priori subregions are a useful geographical starting point but not an environmental partition. Only the Mediterranean and, to a lesser degree, the Interior Northeast were well recovered from the pre-event predictors, while Galicia and the Interior Southwest were largely absorbed into

https://doi.org/10.5194/egusphere-2026-4053
Preprint. Discussion started: 21 July 2026

neighbouring classes. The dominant structure is coarser: two regimes reproduce the northeast–southwest contrast already documented for Iberia (Rodríguez and Bech, 2021), and the retained three-regime solution refines it by isolating a third environmentally distinct regime located between the other two. We read the regimes as different ingredient combinations rather than as sharply bounded regions, which is consistent with the range of synoptic settings that Rodríguez and Lemus-Canovas (2023) associated with Iberian tornadic storms.

Each signature maps onto a known Iberian contrast. Regime 1, over the western and southwestern peninsula, is distinguished by enhanced low- to mid-level wind and shear with comparatively modest instability — the high-shear, low-CAPE setting repeatedly linked to Atlantic-facing and Gulf of Cádiz tornadoes (Rodríguez and Bech, 2018, 2021) and, more broadly, to environments that produce frequent but generally weak, short-lived tornadoes (Sherburn and Parker, 2014). Its high tornado count alongside the absence of EF3 cases fits that interpretation. Regime 2, a central-to-southeastern belt, instead shows by far the strongest anomalies of the three regimes, dominated by CAPE-, LCL-, and effective-inflow-related parameters, pointing to more strongly forced, supercell-favourable environments capable of occasional high-end tornadoes. Regime 0, covering the northern and eastern sectors and the Balearic Islands, shows comparatively muted anomalies across the wind- and shear-related predictors that distinguish regime 1, and carries the largest share of waterspouts; its low-shear character agrees with the milder northeastern environments of the binary contrast, although the strongest instability concentrates in regime 2 rather than across the whole northeast, a distinction the simple NE–SW split could not resolve.

A central result is that the most frequent tornado environments are not those that dominate the high-end wind hazard. Regimes 0 and 1 hold most of the tornadoes and control exceedance at 144 km h^{-1}, but at 180 and 220 km h^{-1} the hazard shifts to regime 2, which contains four of the five recorded EF3 tornadoes despite having the fewest events. The footprint-based framework makes this explicit: high-end point-scale exceedance scales with the expected damaged area of the most intense classes, and the EF3 footprint (≈11.5 km^2) is more than an order of magnitude larger than the EF1 footprint (≈0.9 km^2). The neighbourhood return periods and the point-scale exceedance probabilities consequently rank the regimes differently; this is not a contradiction but a direct effect of weighting events by footprint. For hazard purposes, event frequency alone is a poor proxy for severe-wind hazard, and an environment such as regime 2 warrants attention out of proportion to its event count.

Set against the previous national climatology, the results extend rather than overturn the earlier picture. The AEMET (2015) compilation for 2003–2012 recorded no EF3 tornado, whereas the longer 1992–2025 record used here includes five — a reminder of how sensitive the high-intensity tail is to catalogue length, and a reason to treat the 220 km h^{-1} estimates cautiously. This adaptation also represents, to our knowledge, the first spatially explicit translation of the Iberian tornado climatology into wind-speed exceedance, obtained by adapting the footprint approach of Balbi and Barbieri (2018) and the area-partition matrix of Ramsdell and Rishel (2007) to regime-dependent occurrence rates. The framework behaves coherently under the main sensitivity test:

https://doi.org/10.5194/egusphere-2026-4053
Preprint. Discussion started: 21 July 2026

excluding waterspouts shifts the point-scale exceedance probabilities but leaves the neighbourhood return periods almost unchanged, which confirms that the threshold-capable occurrence signal is dominated by tornadoes and is robust to how waterspouts over land are treated.

Several caveats qualify this interpretation. The exceedance estimates inherit the reporting heterogeneity of the source databases and are conditional on the observed catalogue rather than predictive. Estimates for the most intense classes rest on few events — five EF3 reports and only seven EF2 tornadoes with usable path dimensions — so the EF2 and EF3 footprints carry large uncertainty and depend partly on the Brooks (2004) scaling and on coarsely reported path lengths and widths. Recent reassessment work using the International Fujita scale in Catalonia suggests that rating uncertainty is largest for weak tornadoes, whereas EF1 and stronger events show fewer changes after rerating (Rodríguez et al., 2026). This does not remove the sampling uncertainty in our high-threshold estimates, but it suggests that the largest classification ambiguities are concentrated mainly at the weak end of the distribution, with more limited influence on the 180 and 220 km h^{-1} exceedance estimates. Waterspouts were retained for the climatological and environmental analyses but excluded from the hazard calculation, since their overland wind speeds cannot be assigned reliably from the present record. The regionalization was derived from the combined sample and then applied to the tornado-only hazard, an approximation justified by shared environmental drivers but worth revisiting as more rated reports accumulate. Finally, the ERA5 proximity approach describes the pre-event environment at the reanalysis scale and over a 2×2 average, smoothing the sharp coastal and orographic gradients that matter in a domain as varied as the Iberian–Balearic region.

5 Conclusions

We assembled an updated 1992–2025 tornado and waterspout catalogue for peninsular Spain and the Balearic Islands and used it to move from climatology toward a first regime-dependent assessment of wind-speed hazard. The record of 1185 events is temporally heterogeneous, with a late-summer to autumn maximum and a clear split between morning waterspouts and late-afternoon tornadoes; consistent with patterns documented for other European severe-weather databases, its rising counts more plausibly reflect improved reporting practices than a physical trend. The six geographically motivated subregions were only partially separable from the pre-event ERA5 environments, and the dominant structure proved simpler: an objective Ward regionalization of 50 km hexagonal cells resolves three broad regimes that refine, without replacing, the known northeast–southwest contrast.

Each regime carries distinct environmental signatures — shear-dominated in the west and southwest, thermodynamically charged in the central-southeast, and weakly forced in the north and east — and these differences propagate into the hazard. The most informative result is a decoupling between occurrence and severity: the regimes that produce most tornadoes are not those that dominate the high-end wind hazard. The least active regime, the central-southeastern belt, concentrates the strongest instability-related environmental signal and four of the five EF3 tornadoes, and therefore controls exceedance at 180 and 220 km h^{-1}. Occurrence rate alone therefore underestimates where severe winds are most likely.

https://doi.org/10.5194/egusphere-2026-4053
Preprint. Discussion started: 21 July 2026

These estimates are preliminary and conditional on the observed catalogue, yet they offer a first spatially explicit translation of the Iberian tornado climatology into wind-speed exceedance. Further progress will depend mainly on the catalogue. Because tornadoes and their intensity are inherently difficult to document objectively — the events are brief and narrow, many go unsurveyed, and the EF scale was not designed for European structures — more accurate exceedance estimates require denser and more rigorous reporting. Standardized, reproducible damage surveys that record path length, width, and rated intensity for as many events as possible (Rodríguez et al., 2020), together with sustained community contributions to platforms such as SINOBAS, would tighten the EF2–EF3 geometry that drives the severe end and keep both the database and the resulting exceedance maps current. Higher-resolution or convection-permitting environmental information could further sharpen the regional boundaries in a domain with strong coastal and orographic gradients, and applying the approach to other Mediterranean regions would test how transferable the regime-dependent framework is.

Code availability

The code and reproducible workflow used in this study are archived on Zenodo at https://doi.org/10.5281/zenodo.21225731 (Agurto Barragán, 2026). The archived repository includes the Python/Jupyter notebooks used for catalogue harmonization and filtering, ERA5 predictor extraction and summarization, clustering and supervised-separability analyses, and regime-dependent occurrence and exceedance calculations. It also includes the supporting Python source code, ERA5 download and preprocessing scripts, figure-generation routines, documentation, and the computational environment specification. Private/raw datasets and large derived outputs are not redistributed, in accordance with the repository data policy and the access conditions of the original data providers.

Data availability

ERA5 reanalysis data are publicly available from the Copernicus Climate Data Store. SINOBAS tornado and waterspout records are available through AEMET, and ESWD records are available through the European Severe Storms Laboratory under their corresponding access, licensing, and quality-control conditions. The processed datasets used internally for the analyses are not redistributed in the code repository, but the archived workflow documents the processing steps required to reproduce the analyses from the authorized source data.

Author contributions

G.A. curated the database, performed the formal analysis (data harmonization, ERA5 predictor extraction, clustering, classification, and exceedance calculations), and prepared the original draft. J.A.P., I.G.A. and E.G.S.M.H. contributed to conceptualization, supervision, and review and editing. All authors contributed to and approved the final manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

We thank CEAFMC (*Centro de Estudios Avanzados en Física, Matemáticas y Computación*, University of Huelva) for technical support and for access to its HPC cluster, which was used for downloading the ERA5 profiles and computing the 54 event-level parameters described in Sect. 2.7. We also acknowledge support from the Trust-AI project (PID2023-150070NB-I00). Separately, during the preparation of this work, the authors used Claude (Anthropic) to assist with language editing and improve the clarity of the text. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Financial support

This research was partially funded by the Spanish Nuclear Safety Council (CSN) under project SUBV-4/2022 (CLIMATOR).

References

AEMET: Climatología de tornados en España peninsular y Baleares, Agencia Estatal de Meteorología, Madrid, available at: https://www.aemet.es/documentos/es/conocermas/recursos_en_linea/publicaciones_y_estudios/publicaciones/Climatologia_tornados/Climatologia_tornados.pdf (last access: 7 July 2026), 2015.

Agurto Barragán, G.: gagurt01/iberian-tornado-hazard: v1.0.0 – Repository release, Zenodo [code], https://doi.org/10.5281/zenodo.21225731, 2026.

Avolio, E. and Miglietta, M. M.: Multiple tornadoes in the Italian Ionian regions: observations, sensitivity tests and mesoscale analysis of convective storm environmental parameters, Atmos. Res., 263, 105800, https://doi.org/10.1016/j.atmosres.2021.105800, 2021.

Bagaglini, L., Ingrosso, R., and Miglietta, M. M.: Synoptic patterns and mesoscale precursors of Italian tornadoes, Atmos. Res., 253, 105503, https://doi.org/10.1016/j.atmosres.2021.105503, 2021.

Balbi, M. and Barbieri, P.: Enfoque científico del riesgo: evaluación del potencial de tornados en Argentina, Academia Nacional de Ciencias de Buenos Aires, Buenos Aires, ISBN 978-987-537-154-5, available at: https://www.ciencias.org.ar/user/Enfoque%20cientifico%20del%20riesgo%20-%20evaluacion%20del%20potencial%20de%20tornados%20en%20al%20republica%20argentina%202.pdf (last access: 7 July 2026), 2018.

Bidner, A.: The Air Force Global Weather Central Severe Weather Threat (SWEAT) Index, Air Weather Service Aerospace Sciences Review, AWS-RP-105-2, 1970.

Birch, C. P. D., Oom, S. P., and Beecham, J. A.: Rectangular and hexagonal grids used for observation, experiment and simulation in ecology, Ecol. Model., 206, 347–359, https://doi.org/10.1016/j.ecolmodel.2007.03.041, 2007.

https://doi.org/10.5194/egusphere-2026-4053
Preprint. Discussion started: 21 July 2026

Blumberg, W. G., Halbert, K. T., Supinie, T. A., Marsh, P. T., Thompson, R. L., and Hart, J. A.: SHARPpy: An Open-Source Sounding Analysis Toolkit for the Atmospheric Sciences, Bull. Amer. Meteor. Soc., 98, 1625–1636, https://doi.org/10.1175/BAMS-D-15-00309.1, 2017.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J.: Classification and Regression Trees, Wadsworth, Belmont, CA, 1984.

Brodersen, K. H., Ong, C. S., Stephan, K. E., and Buhmann, J. M.: The balanced accuracy and its posterior distribution, in: Proceedings of the 20th International Conference on Pattern Recognition, IEEE, 3121–3124, https://doi.org/10.1109/ICPR.2010.764, 2010.

Brooks, H. E.: On the relationship of tornado path length and width to intensity, Weather Forecast., 19, 310–319, https://doi.org/10.1175/1520-0434(2004)019%3C0310:OTROTP%3E2.0.CO;2, 2004.

Cawley, G. C. and Talbot, N. L. C.: On over-fitting in model selection and subsequent selection bias in performance evaluation, J. Mach. Learn. Res., 11, 2079–2107, 2010.

Cheng, V. Y. S., Arhonditsis, G. B., Sills, D. M. L., Gough, W. A., and Auld, H.: A Bayesian modelling framework for tornado occurrences in North America, Nat. Commun., 6, 6599, https://doi.org/10.1038/ncomms7599, 2015.

Chicco, D. and Jurman, G.: The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation, BMC Genomics, 21, 6, https://doi.org/10.1186/s12864-019-6413-7, 2020.

Cortes, C. and Vapnik, V.: Support-vector networks, Mach. Learn., 20, 273–297, https://doi.org/10.1007/BF00994018, 1995.

Doswell, C. A. III and Burgess, D. W.: On some issues of United States tornado climatology, Mon. Weather Rev., 116, 495–501, https://doi.org/10.1175/1520-0493(1988)116%3C0495:OSIOUS%3E2.0.CO;2, 1988.

Dotzek, N., Groenemeijer, P., Feuerstein, B., and Holzer, A. M.: Overview of ESSL's severe convective storms research using the European Severe Weather Database ESWD, Atmos. Res., 93, 575–586, https://doi.org/10.1016/j.atmosres.2008.10.020, 2009.

Efron, B. and Tibshirani, R. J.: An Introduction to the Bootstrap, Monographs on Statistics and Applied Probability, Vol. 57, Chapman & Hall, New York, 1993.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise, in: Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, Portland, OR, 226–231, 1996.

https://doi.org/10.5194/egusphere-2026-4053
Preprint. Discussion started: 21 July 2026

Galway, J. G.: The lifted index as a predictor of latent instability, Bull. Am. Meteorol. Soc., 37, 528–529, https://doi.org/10.1175/1520-0477-37.10.528, 1956.

Gayà, M.: Tornadoes and severe storms in Spain, Atmos. Res., 100, 334–343, https://doi.org/10.1016/j.atmosres.2010.10.019, 2011.

George, J. J.: Weather Forecasting for Aeronautics, Academic Press, New York, 1960.

Gleckler, P. J., Taylor, K. E., and Doutriaux, C.: Performance metrics for climate models, J. Geophys. Res., 113, D06104, https://doi.org/10.1029/2007JD008972, 2008.

Gorodkin, J.: Comparing two K-category assignments by a K-category correlation coefficient, Comput. Biol. Chem., 28, 367–374, https://doi.org/10.1016/j.compbiolchem.2004.09.006, 2004.

Groenemeijer, P. and Kühne, T.: A climatology of tornadoes in Europe: results from the European Severe Weather Database, Mon. Weather Rev., 142, 4775–4790, https://doi.org/10.1175/MWR-D-14-00107.1, 2014.

Hart, J. A. and Korotky, W.: The SHARP workstation v1.50: a skew-T/hodograph analysis and research program for the IBM and compatible PC, User's Manual, NOAA/NWS Forecast Office, Charleston, WV, 1991.

Hennig, C.: What are the true clusters?, Pattern Recogn. Lett., 64, 53–62, https://doi.org/10.1016/j.patrec.2015.04.009, 2015.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Q. J. Roy. Meteor. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Huber, P. J. and Ronchetti, E. M.: Robust Statistics, 2nd ed., Wiley, Hoboken, https://doi.org/10.1002/9780470434697, 2009.

Ingrosso, R., Lionello, P., Miglietta, M. M., and Salvati, P.: A statistical investigation of mesoscale precursors of significant tornadoes: the Italian case study, Atmosphere, 11, 301, https://doi.org/10.3390/atmos11030301, 2020.

Jagger, T. H., Elsner, J. B., and Widen, H. M.: A statistical model for regional tornado climate studies, PLoS ONE, 10, e0131876, https://doi.org/10.1371/journal.pone.0131876, 2015.

James, G., Witten, D., Hastie, T., and Tibshirani, R.: An Introduction to Statistical Learning: with Applications in R, 2nd ed., Springer, https://doi.org/10.1007/978-1-0716-1418-1, 2021.

https://doi.org/10.5194/egusphere-2026-4053
Preprint. Discussion started: 21 July 2026

Jolliffe, I. T. and Cadima, J.: Principal component analysis: a review and recent developments, Philos. T. R. Soc. A, 374, 20150202, https://doi.org/10.1098/rsta.2015.0202, 2016.

Jurman, G., Riccadonna, S., and Furlanello, C.: A comparison of MCC and CEN error measures in multi-class prediction, PLoS ONE, 7, e41882, https://doi.org/10.1371/journal.pone.0041882, 2012.

Kaufman, S., Rosset, S., Perlich, C., and Stitelman, O.: Leakage in data mining: formulation, detection, and avoidance, ACM T. Knowl. Discov. D., 6, 15, https://doi.org/10.1145/2382577.2382579, 2012.

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., and Willing, C.: Jupyter Notebooks – a publishing format for reproducible computational workflows, in: Positioning and Power in Academic Publishing: Players, Agents and Agendas, IOS Press, 87–90, https://doi.org/10.3233/978-1-61499-649-1-87, 2016.

Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Proceedings of the 14th International Joint Conference on Artificial Intelligence, IJCAI'95, Vol. 2, Morgan Kaufmann, San Francisco, CA, 1137–1143, 1995.

Little, R. J. A. and Rubin, D. B.: Statistical Analysis with Missing Data, 3rd ed., Wiley, Hoboken, https://doi.org/10.1002/9781119013563, 2019.

Majewski, D., Liermann, D., Prohl, P., Ritter, B., Buchhold, M., Hanisch, T., Paul, G., Wergen, W., and Baumgardner, J.: The operational global icosahedral–hexagonal gridpoint model GME: description and high-resolution tests, Mon. Weather Rev., 130, 319–338, https://doi.org/10.1175/1520-0493(2002)130%3C0319:TOGIHG%3E2.0.CO;2, 2002.

Matthews, B. W.: Comparison of the predicted and observed secondary structure of T4 phage lysozyme, Biochim. Biophys. Acta, 405, 442–451, https://doi.org/10.1016/0005-2795(75)90109-9, 1975.

Metropolis, N. and Ulam, S.: The Monte Carlo method, J. Am. Stat. Assoc., 44, 335–341, https://doi.org/10.1080/01621459.1949.10483310, 1949.

Miller, R. C.: Notes on Analysis and Severe-Storm Forecasting Procedures of the Air Force Global Weather Central, Technical Report 200 (Rev.), AWS, Scott AFB, IL, 1972.

Murtagh, F. and Legendre, P.: Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion?, J. Classif., 31, 274–295, https://doi.org/10.1007/s00357-014-9161-z, 2014.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É.: Scikit-learn: machine learning in Python, J. Mach. Learn. Res., 12, 2825–2830, 2011.

https://doi.org/10.5194/egusphere-2026-4053
Preprint. Discussion started: 21 July 2026

Púčik, T., Groenemeijer, P., Rýva, D., and Kolář, M.: Proximity soundings of severe and nonsevere thunderstorms in Central Europe, Mon. Weather Rev., 143, 4805–4821, https://doi.org/10.1175/MWR-D-15-0104.1, 2015.

Ramsdell, J. V. and Rishel, J. P.: Tornado Climatology of the Contiguous United States, NUREG/CR-4461, Rev. 2, PNNL-15112, Rev. 1, U.S. Nuclear Regulatory Commission, Office of Nuclear Regulatory Research, Washington, D.C., available at: https://www.nrc.gov/docs/ML0708/ML070810400.pdf (last access: 7 July 2026), 2007.

Robert, C. P. and Casella, G.: Introducing Monte Carlo Methods with R, Springer, New York, https://doi.org/10.1007/978-1-4419-1576-4, 2010.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., and Dormann, C. F.: Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure, Ecography, 40, 913–929, https://doi.org/10.1111/ecog.02881, 2017.

Rodríguez, O. and Bech, J.: Sounding-derived parameters associated with tornadic storms in Catalonia, Int. J. Climatol., 38, 2400–2414, https://doi.org/10.1002/joc.5343, 2018.

Rodríguez, O. and Bech, J.: Tornadic environments in the Iberian Peninsula and the Balearic Islands based on ERA5 reanalysis, Int. J. Climatol., 41, E1959–E1979, https://doi.org/10.1002/joc.6825, 2021.

Rodríguez, O. and Lemus-Canovas, M.: Synoptic patterns triggering tornadic storms on the Iberian Peninsula and the Balearic Islands, Atmos. Res., 285, 106634, https://doi.org/10.1016/j.atmosres.2023.106634, 2023.

Rodríguez, O., Bech, J., Soriano, J. de D., Gutiérrez, D., and Castán, S.: A methodology to conduct wind damage field surveys for high-impact weather events of convective origin, Nat. Hazards Earth Syst. Sci., 20, 1513–1531, https://doi.org/10.5194/nhess-20-1513-2020, 2020.

Rodríguez, O., Castán, S., and Bech, J.: Revisiting a 25-year database of tornado damage paths using the new International Fujita (IF) scale, Nat. Hazards, 122, 509, https://doi.org/10.1007/s11069-026-08281-9, 2026.

Sessa, M. F. and Trapp, R. J.: The prediction of potential tornado damage intensity using machine learning, Artif. Intell. Earth Syst., 4, e230113, https://doi.org/10.1175/AIES-D-23-0113.1, 2025.

Sherburn, K. D. and Parker, M. D.: Climatology and ingredients of significant severe convection in high-shear, low-CAPE environments, Weather Forecast., 29, 854–877, https://doi.org/10.1175/WAF-D-13-00041.1, 2014.

Sokolova, M. and Lapalme, G.: A systematic analysis of performance measures for classification tasks, Inf. Process. Manage., 45, 427–437, https://doi.org/10.1016/j.ipm.2009.03.002, 2009.

https://doi.org/10.5194/egusphere-2026-4053
Preprint. Discussion started: 21 July 2026

Taszarek, M. and Kolendowicz, L.: Sounding-derived parameters associated with tornado occurrence in Poland and Universal Tornadic Index, Atmos. Res., 134, 186–197, https://doi.org/10.1016/j.atmosres.2013.07.016, 2013.

Taszarek, M., Brooks, H. E., and Czernecki, B.: Sounding-derived parameters associated with convective hazards in Europe, Mon. Weather Rev., 145, 1511–1528, https://doi.org/10.1175/MWR-D-16-0384.1, 2017.

Thompson, R. L., Edwards, R., Hart, J. A., Elmore, K. L., and Markowski, P.: Close proximity soundings within supercell environments obtained from the Rapid Update Cycle, Weather Forecast., 18, 1243–1261, https://doi.org/10.1175/1520-0434(2003)018%3C1243:CPSWSE%3E2.0.CO;2, 2003.

Thompson, R. L., Mead, C. M., and Edwards, R.: Effective storm-relative helicity and bulk shear in supercell thunderstorm environments, Weather Forecast., 22, 102–115, https://doi.org/10.1175/WAF969.1, 2007.

Tibshirani, R., Walther, G., and Hastie, T.: Estimating the number of clusters in a data set via the gap statistic, J. R. Stat. Soc. B, 63, 411–423, https://doi.org/10.1111/1467-9868.00293, 2001.

Wang, W., Yang, J., and Muntz, R. R.: STING: a statistical information grid approach to spatial data mining, in: Proceedings of the 23rd International Conference on Very Large Data Bases, Athens, Greece, 186–195, 1997.

Ward, J. H.: Hierarchical grouping to optimize an objective function, J. Am. Stat. Assoc., 58, 236–244, https://doi.org/10.1080/01621459.1963.10500845, 1963.

WMO: WMO Guidelines on the Calculation of Climate Normals, WMO-No. 1203, World Meteorological Organization, Geneva, ISBN 978-92-63-11203-3, available at: https://library.wmo.int/doc_num.php?explnum_id=4166 (last access: 7 July 2026), 2017.

Zhang, Y., Li, T., and Wang, B.: Optimal cluster analysis for objective regionalization of seasonal precipitation in regions of high spatial-temporal variability, J. Climate, 29, 5497–5513, https://doi.org/10.1175/JCLI-D-15-0582.1, 2016.