Supplement for

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)

Contents of this file

- Figures S1 - S8
- Table S1 - S3

Introduction

This Supplement complements the main text with the figures and tables supporting the sensitivity tests, diagnostic comparisons, robustness checks, parameter definitions, and exceedance diagnostics used in the analysis. Figure S1 documents the sensitivity of the diurnal distribution to event-time filtering. Figure S2 compares the supervised classifiers used to evaluate the separability of the six a priori subregions. Figure S3 summarizes the screening diagnostics for the event-level clustering methods. Figure S4 provides the per-metric comparison of the candidate Ward spatial regionalizations as a function of the number of regions. Figure S5 assesses the bootstrap robustness of the retained Ward 50 km k = 3 regionalization. Figure S6 tests whether this final regionalization remains environmentally separable when only tornado reports are considered. Figure S7 shows the detailed regime-specific standardized-anomaly profiles supporting the compact environmental-signature heatmap in the main text. Figure S8 maps the observed-cell tornado exceedance probabilities, complementing the filled-grid exceedance maps shown in the main text. Table S1 lists the 54 ERA5-derived parameters with their categories, units, and primary references; Table S2 summarizes the clustering-method screening; and Table S3 gives the bootstrap return-period estimates by regime and wind-speed threshold.

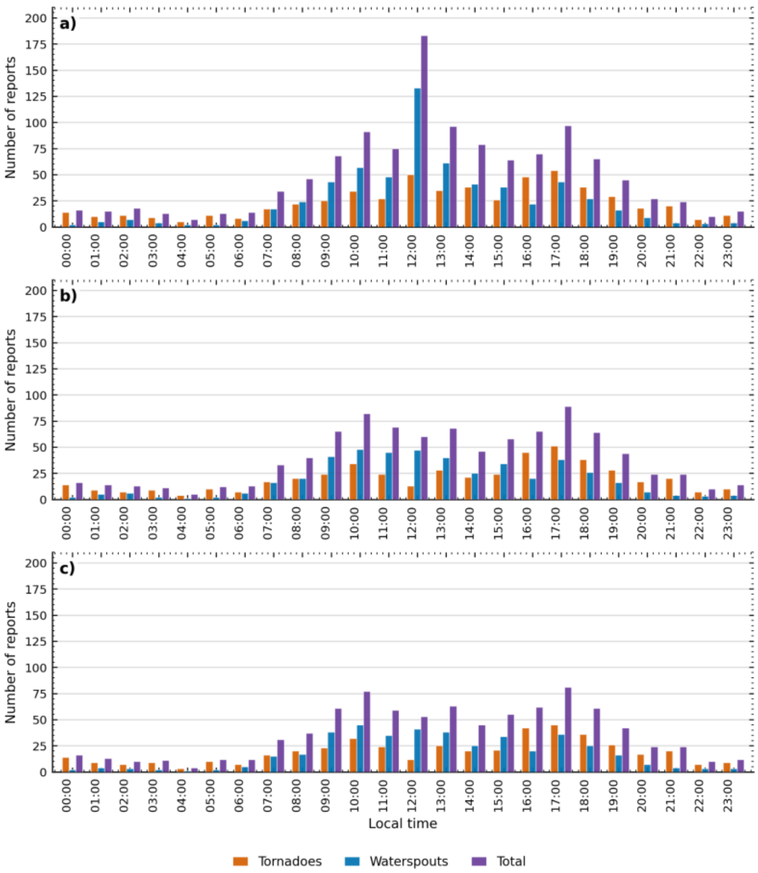

Figure S1. Sensitivity of the diurnal distribution to event-time filtering. (a) Diurnal distribution using all reports without filtering by temporal uncertainty. (b) Diurnal distribution using only reports with temporal uncertainty equal to or lower than 1 h. (c) Diurnal distribution using only reports with temporal uncertainty equal to or lower than 30 min. Colours are as in Fig. 2. The unfiltered distribution is shown as a diagnostic of possible time-rounding or default-time effects and is not used for the main interpretation.

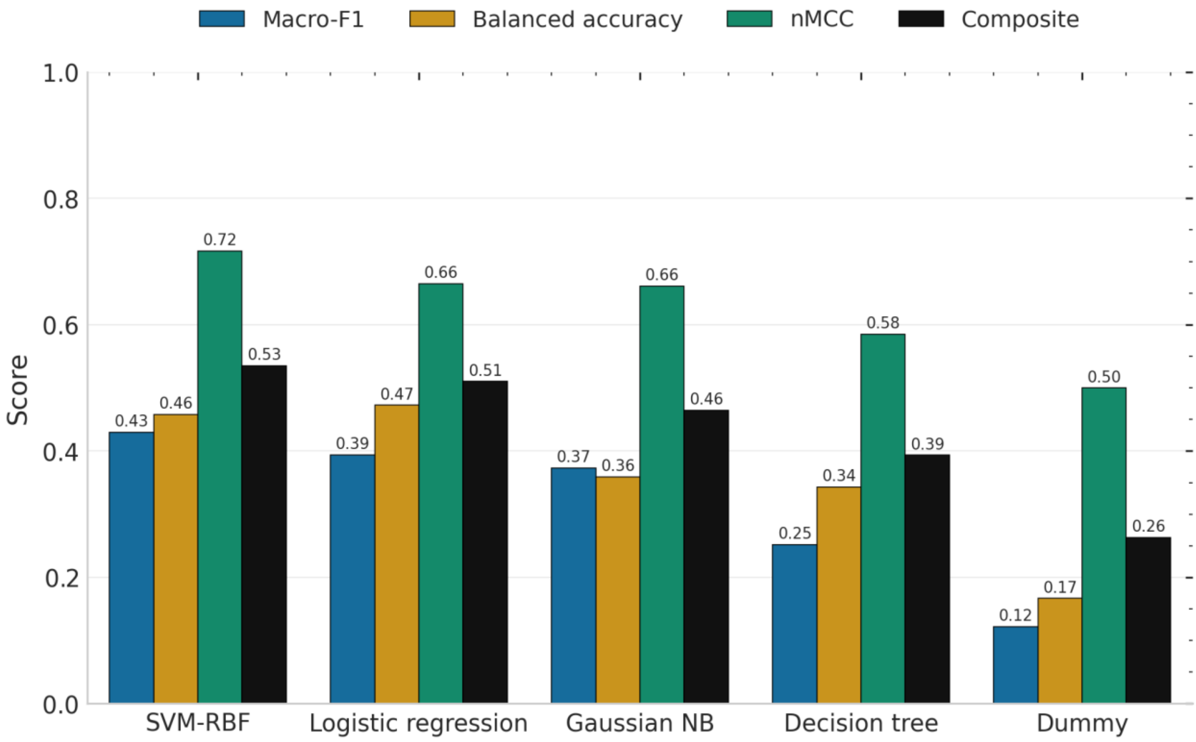

Figure S2. Supervised separability of the six a priori subregions from pre-event ERA5 predictors (full event sample). Bars show Macro-F1, balanced accuracy, normalized MCC (nMCC = (MCC + 1)/2) and the composite score (their unweighted mean) for each classifier and the dummy baseline, on a common 0–1 scale. SVM-RBF scores highest, followed by logistic regression; Gaussian naive Bayes and the decision tree are lower, and all exceed the dummy baseline.

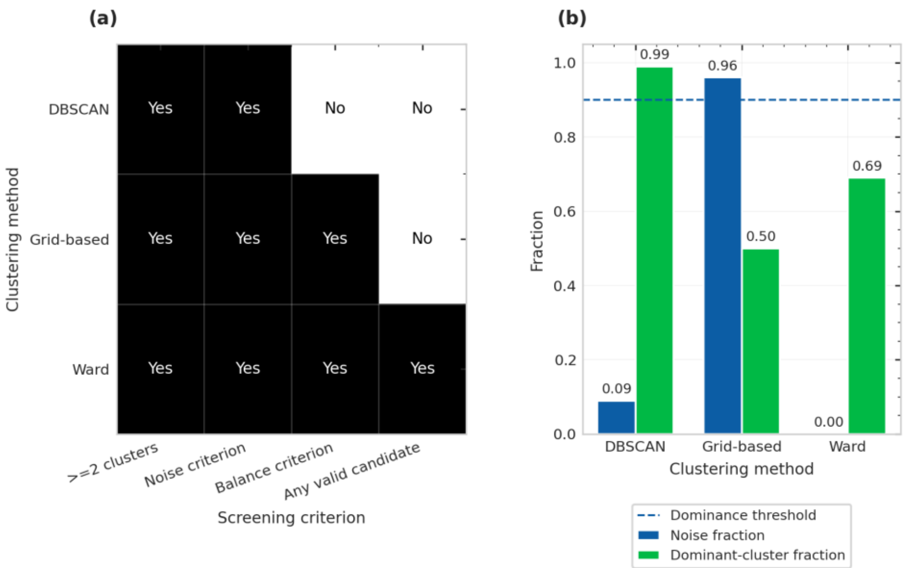

Figure S3. Screening diagnostics for event-level clustering methods applied to the PCA-reduced environmental predictor space. (a) Summary of the predefined screening criteria applied to DBSCAN, grid-based clustering, and Ward hierarchical clustering. A valid method had to produce at least two clusters, satisfy the noise and balance criteria, and yield at least one valid candidate. (b) Diagnostic properties of the best raw candidate generated by each clustering method. High noise fractions indicate poor event coverage, whereas high dominant-cluster fractions indicate collapse into one main cluster. These diagnostics explain why DBSCAN and grid-based clustering were not retained as valid event-level regionalization methods.

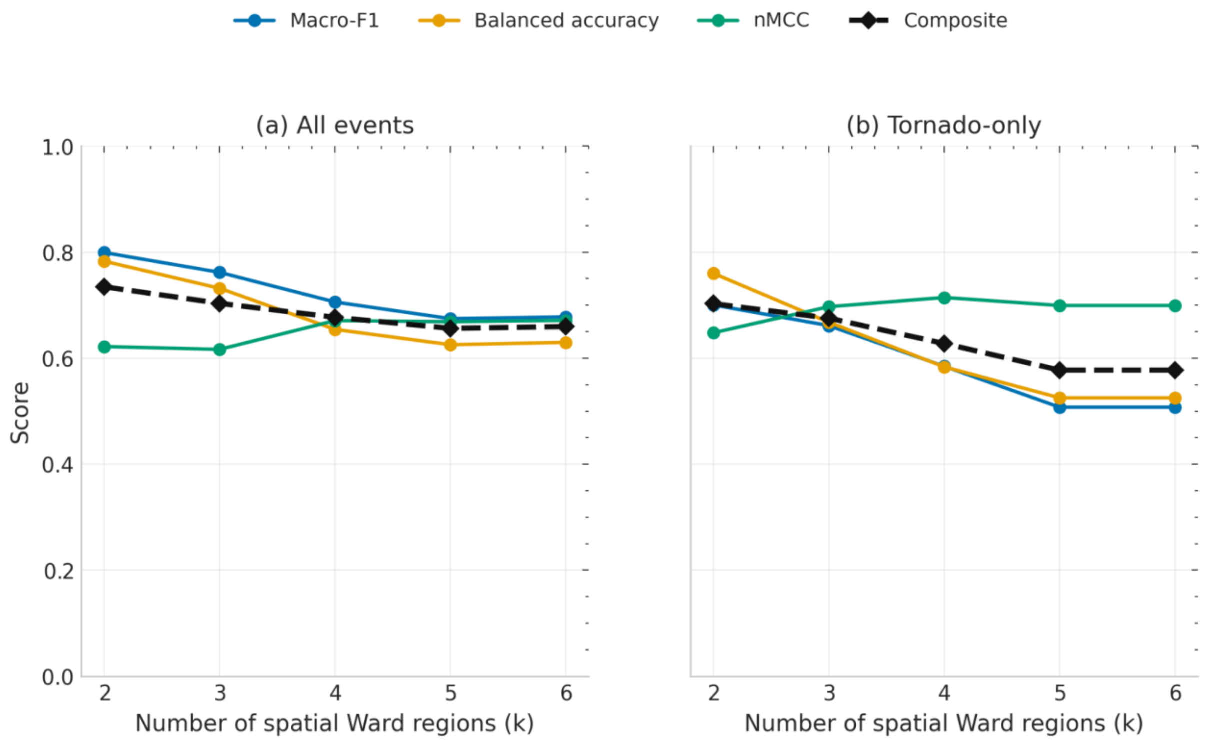

Figure S4. Per-metric comparison of the candidate spatial Ward regionalizations as a function of the number of regions k, at the selected 50 km hexagon radius, for (a) the full event sample and (b) the tornado-only subset. Each line shows a supervised-separability metric obtained when the regional labels are predicted from event-level pre-event ERA5 predictors: Macro-F1, balanced accuracy, the normalized Matthews correlation coefficient (nMCC = (MCC + 1) / 2), and the composite score (their unweighted mean, dashed black line); all scores lie on a common 0–1 scale. The composite score is highest at k = 2 and declines for larger k; the three-region solution (k = 3) was retained as the final regionalization on the basis of its dummy-relative skill, interpretability and parsimony (Sect. 3.3). In the tornado-only panel, the curves for k = 5 and k = 6 coincide because the additional region introduced at k = 6 contains fewer than the minimum of 20 tornadoes required per class and is excluded from the tornado-only evaluation, leaving the same effective classification as for k = 5.

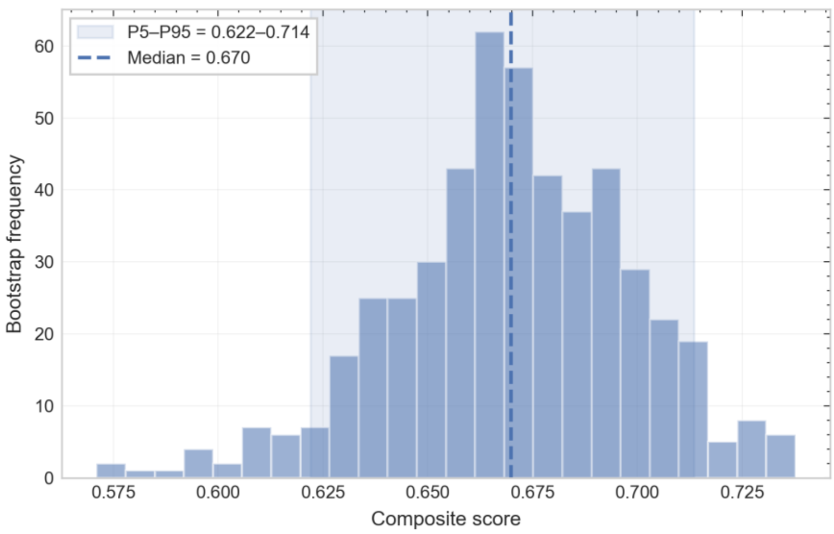

Figure S5. Bootstrap robustness of the retained Ward 50 km k = 3 regionalization. The histogram shows the distribution of the composite score obtained from year-wise bootstrap resampling of the event catalogue. The dashed vertical line indicates the bootstrap median, and the shaded interval indicates the 5th–95th percentile range. This diagnostic quantifies the sensitivity of the selected regionalization to temporal sampling variability.

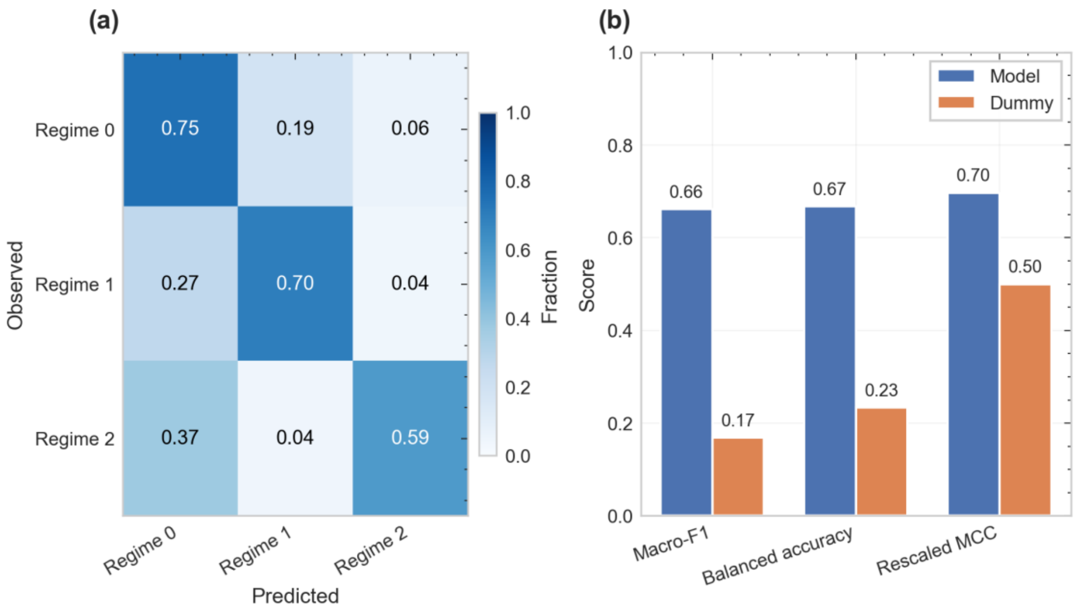

Figure S6. Tornado-only robustness assessment of the retained Ward 50 km k = 3 regionalization. (a) Row-normalized confusion matrix obtained when the final regime labels are predicted from pre-event ERA5 predictors using only tornado reports. Observed regimes are shown on the y axis and predicted regimes on the x axis. (b) Model performance compared with a dummy baseline for the tornado-only subset, using Macro-F1, balanced accuracy, and normalized Matthews correlation coefficient (MCC). This figure tests whether the final regionalization remains environmentally separable when waterspouts are excluded.

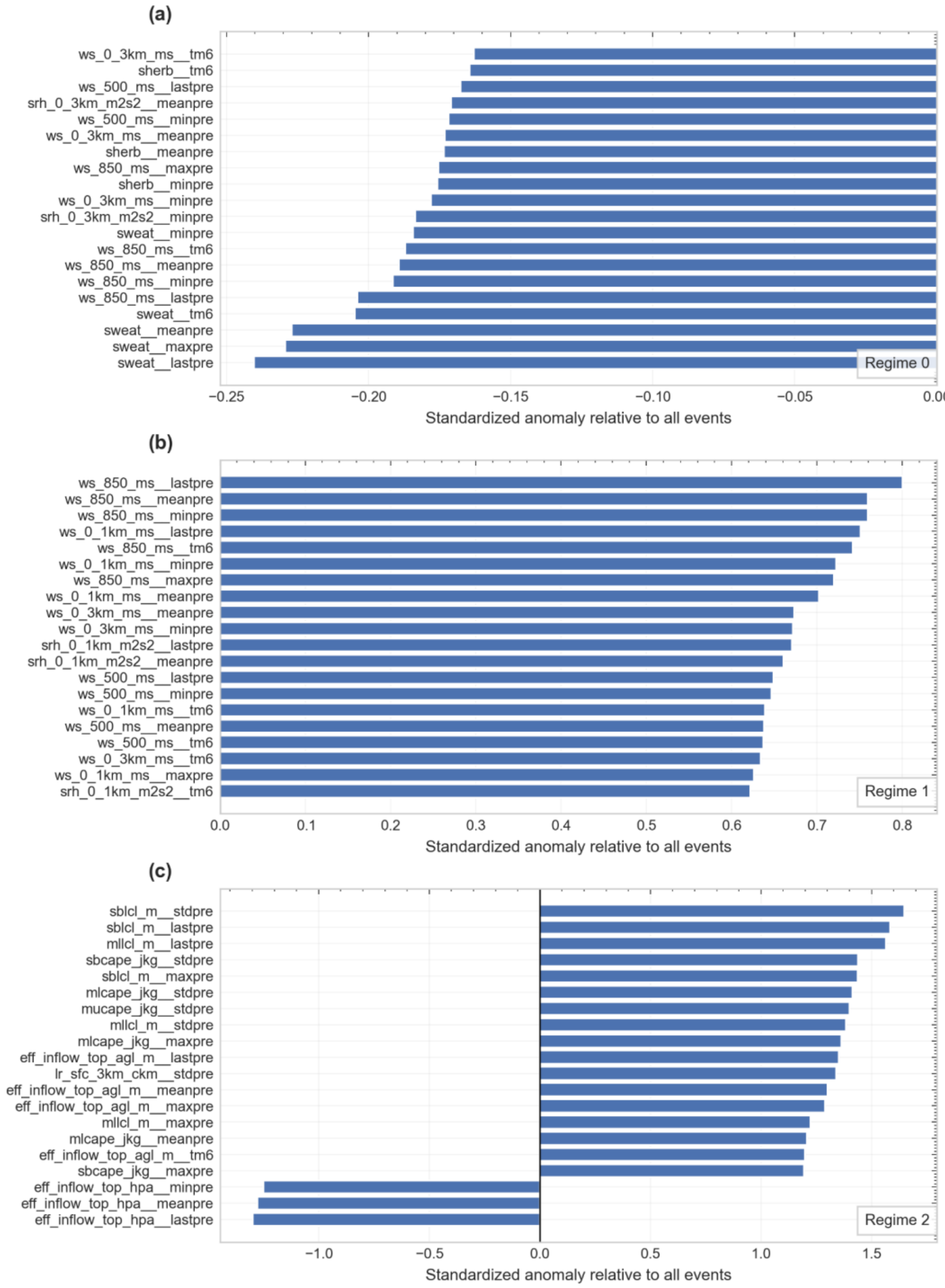

Figure S7. Detailed regime-specific standardized-anomaly profiles for the retained Ward 50 km k = 3 regionalization. Panels show the leading ERA5-derived pre-event predictors for (a) regime 0, (b) regime 1, and (c) regime 2. Bars indicate standardized anomalies relative to the full event sample. The figure provides detailed support for the compact environmental-signature heatmap shown in Fig. 6c.

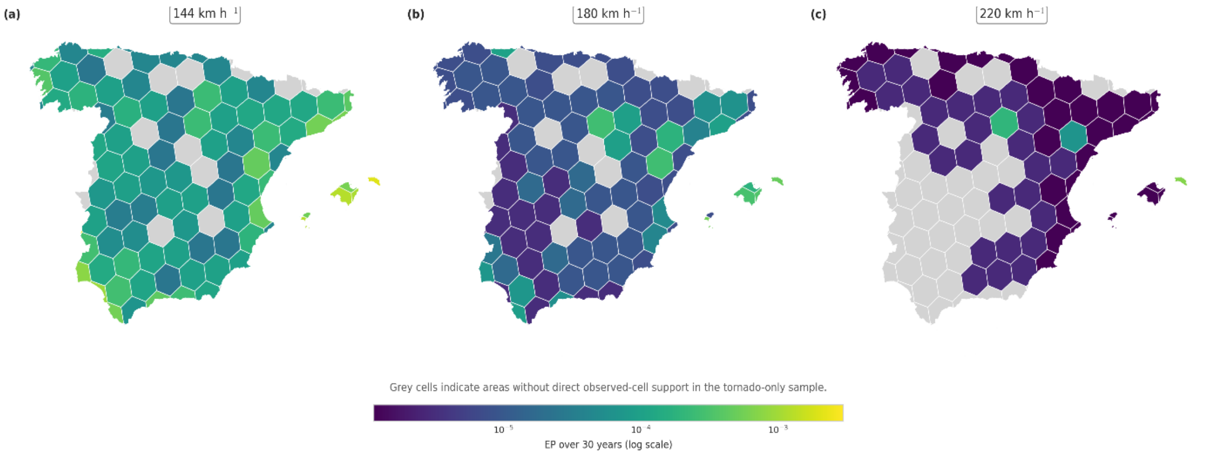

Figure S8. Hexagon-level point-scale 30-year tornado exceedance probability for wind-speed thresholds of (a) 144, (b) 180, and (c) 220 km h^{-1}, shown only for hexagonal cells with direct observed-cell support in the tornado-only sample. Estimates were obtained with empirical-Bayes shrinkage toward the corresponding Ward-regime prior, using an equivalent prior exposure of 10 years. Colours are shown on a logarithmic scale. Grey cells indicate areas without direct observed-cell support in the tornado-only sample.

#	Code	Parameter name	Category	Units	Primary reference(s)
1. Instability, inhibition and parcel levels					
1	*sbcape_jkg*	SB CAPE	Thermodynamic	J kg^{-1}	Thompson et al. (2003)
2	*sbcin_jkg*	SB CIN	Thermodynamic	J kg^{-1}	Thompson et al. (2003)
3	*mucape_jkg*	MU CAPE	Thermodynamic	J kg^{-1}	Thompson et al. (2003)
4	*mucin_jkg*	MU CIN	Thermodynamic	J kg^{-1}	Thompson et al. (2003)
5	*mlcape_jkg*	ML CAPE	Thermodynamic	J kg^{-1}	Thompson et al. (2003)
6	*mlcin_jkg*	ML CIN	Thermodynamic	J kg^{-1}	Thompson et al. (2003)
7	*sblcl_m*	SB LCL height	Thermodynamic	m	Thompson et al. (2003)
8	*mllcl_m*	ML LCL height	Thermodynamic	m	Thompson et al. (2003)
9	*mulcl_m*	MU LCL height	Thermodynamic	m	Thompson et al. (2003)
10	*sblfc_hpa*	SB LFC	Thermodynamic	hPa	Thompson et al. (2003)
11	*mllfc_hpa*	ML LFC	Thermodynamic	hPa	Thompson et al. (2003)
12	*mulfc_hpa*	MU LFC	Thermodynamic	hPa	Thompson et al. (2003)
13	*sbel_hpa*	SB EL	Thermodynamic	hPa	Thompson et al. (2003)
14	*mlel_hpa*	ML EL	Thermodynamic	hPa	Thompson et al. (2003)
15	*muel_hpa*	MU EL	Thermodynamic	hPa	Thompson et al. (2003)
16	*sbli5_c*	SB Lifted Index (500 hPa)	Thermodynamic	°C	Galway (1956)
17	*muli5_c*	MU Lifted Index (500 hPa)	Thermodynamic	°C	Galway (1956)
18	*mlli5_c*	ML Lifted Index (500 hPa)	Thermodynamic	°C	Galway (1956)
19	*sbcape_0_3km_jkg*	SB CAPE 0–3 km	Thermodynamic	J kg^{-1}	Púčik et al. (2015)
20	*mlcape_0_3km_jkg*	ML CAPE 0–3 km	Thermodynamic	J kg^{-1}	Púčik et al. (2015)
2. Shear, helicity and effective inflow layer					
21	*ws_0_1km_ms*	Wind shear 0–1 km	Kinematic	m s^{-1}	Thompson et al. (2003)
22	*ws_0_3km_ms*	Wind shear 0–3 km	Kinematic	m s^{-1}	Thompson et al. (2003)
23	*ws_0_6km_ms*	Wind shear 0–6 km	Kinematic	m s^{-1}	Thompson et al. (2003)
24	*srh_0_1km_m2s2*	SRH 0–1 km	Kinematic	m^2 s^{-2}	Thompson et al. (2003)
25	*srh_0_3km_m2s2*	SRH 0–3 km	Kinematic	m^2 s^{-2}	Thompson et al. (2003)
26	*srh_0_6km_m2s2*	SRH 0–6 km	Kinematic	m^2 s^{-2}	Thompson et al. (2003)
27	*eff_inflow_base_hpa*	Effective inflow base	Kinematic	hPa	Thompson et al. (2007)
28	*eff_inflow_top_hpa*	Effective inflow top	Kinematic	hPa	Thompson et al. (2007)
29	*eff_inflow_base_agl_m*	Effective inflow base (AGL)	Kinematic	m	Thompson et al. (2007)

30	*eff_inflow_top_agl_m*	Effective inflow top (AGL)	Kinematic	m	Thompson et al. (2007)
31	*eff_srh_m2s2*	Effective SRH (ESRH)	Kinematic	$m^2 s^{-2}$	Thompson et al. (2007)
32	*ebwd_ms*	Effective bulk wind diff. (EBWD)	Kinematic	$m s^{-1}$	Thompson et al. (2007)
3. Severe-weather composite indices					
33	*scp*	Supercell Composite Parameter (SCP)	Composite	–	Thompson et al. (2003)
34	*stp_cin*	STP with CIN	Composite	–	Thompson et al. (2007)
35	*stp_fixed*	STP fixed-layer	Composite	–	Thompson et al. (2003)
36	*sbehi_0_1km*	SB EHI 0–1 km	Composite	–	Hart and Korotky (1991)
37	*muehi_0_1km*	MU EHI 0–1 km	Composite	–	Hart and Korotky (1991)
38	*sbehi_0_3km*	SB EHI 0–3 km	Composite	–	Hart and Korotky (1991)
39	*muehi_0_3km*	MU EHI 0–3 km	Composite	–	Hart and Korotky (1991)
40	*sherb*	SHERB	Composite	–	Sherburn and Parker (2014)
41	*total_totals*	Total Totals index	Composite	–	Miller (1972)
42	*k_index_c*	K Index	Composite	°C	George (1960)
43	*sweat*	SWEAT index	Composite	–	Bidner (1970)
44	*wmaxshear_06_m2s2*	WMAXSHEAR 0–6 km	Composite	$m^2 s^{-2}$	Taszarek et al. (2017)
45	*wmaxshear_03_m2s2*	WMAXSHEAR 0–3 km	Composite	$m^2 s^{-2}$	Taszarek et al. (2017)
46	*uti*	Universal Tornadic Index (UTI)	Composite	–	Taszarek and Kolendowicz (2013)
4. Thermal gradients, winds and 0 °C level					
47	*lr_700_500_ckm*	Lapse rate 700–500 hPa	Thermal/ Synoptic	$°C\ km^{-1}$	Púčik et al. (2015)
48	*lr_sfc_3km_ckm*	Lapse rate surface–3 km	Thermal/ Synoptic	$°C\ km^{-1}$	Púčik et al. (2015)
49	*ws_850_ms*	Wind speed at 850 hPa	Thermal/ Synoptic	$m s^{-1}$	Taszarek et al. (2017)
50	*wd_850_deg*	Wind direction at 850 hPa	Thermal/ Synoptic	°	Taszarek et al. (2017)
51	*ws_500_ms*	Wind speed at 500 hPa	Thermal/ Synoptic	$m s^{-1}$	Taszarek et al. (2017)
52	*wd_500_deg*	Wind direction at 500 hPa	Thermal/ Synoptic	°	Taszarek et al. (2017)
53	*isozero_msl_m*	0 °C isotherm height (MSL)	Thermal/ Synoptic	m	Taszarek et al. (2017)
54	*isozero_pres_hpa*	0 °C isotherm pressure level	Thermal/ Synoptic	hPa	Taszarek et al. (2017)

Table S1. ERA5-derived tornado-relevant parameters used in the environmental characterization. Parameters are grouped by physical category. Abbreviations: SB = surface-based; MU = most-unstable; ML = mixed-layer; LCL = lifting condensation level; LFC = level of free convection; EL = equilibrium level; SRH = storm-relative helicity; EBWD = effective bulk wind difference; AGL = above ground level.

Method	Search space	Candidates tested	Valid candidates	Best raw candidate	Noise fraction	Dominant cluster fraction	Decision	Main reason
DBSCAN	PC dimensions, eps, min_samples	90	0	2 clusters; eps=6.12; min_samples=5	0.09	0.99	Rejected	No candidate passed the validity filters; the best raw candidate was dominated by a single non-noise cluster, indicating an outlier/small-cluster split rather than robust regimes.
Grid-based	PC1-PC2 grid resolution and density threshold	20	0	2 clusters; bins=22; density=0.020	0.96	0.5	Rejected	No candidate passed the validity filters; the best raw candidate retained only a small dense subset of events, leaving excessive noise/low coverage.
Ward	PC dimensions and k	30	30	ward_k2_pc5 (2 clusters)	--	0.69	Retained	Generated non-degenerate, interpretable candidate partitions under the predefined validity filters.

Table S2. Screening summary of candidate event-level clustering methods. Candidate solutions were evaluated using predefined validity criteria before any supervised separability analysis. DBSCAN and grid-based candidates were rejected because no candidate simultaneously satisfied the noise, cluster-balance, and coverage requirements, whereas Ward hierarchical clustering produced valid non-degenerate partitions.

Threshold (km h^{-1})	Regime	TR p05 (yr)	TR median (yr)	TR p95 (yr)	(p_0)
144	reg_0	4.60	5.69	7.27	0.000
144	reg_1	4.64	6.16	8.56	0.000
144	reg_2	20.55	30.83	48.68	0.000
180	reg_0	21.28	30.24	47.88	0.000
180	reg_1	27.46	41.19	82.38	0.000
180	reg_2	48.68	92.49	308.30	0.005
220	reg_0	191.54	574.61	574.61	0.361
220	reg_1	∞	∞	∞	1.000
220	reg_2	92.49	231.23	924.90	0.134

Table S3. Bootstrap return-period estimates for threshold-capable tornado occurrence within a 50-km neighbourhood, stratified by Ward regime and wind-speed threshold. Values indicate the 5th percentile, median, and 95th percentile from annual bootstrap resampling over 1992–2025. The p0 column gives the fraction of bootstrap replicates with no threshold-capable event, which produces an undefined or infinite return-period estimate.