This supplement compiles the quality-control and robustness diagnostics referenced in the main text. Section S1 covers the calibration and time-basis checks underlying the correction, and Section S2 presents the barometer test for all regions; Fig. 4 of the main paper shows two representative examples. All figures in this supplement are based on the released one-degree tiles and corrected product and were not used in developing the correction.

S1 Calibration and time-basis diagnostics

The modern reference can be sampled as either a single noon snapshot or an eight-point daily mean. These two methods have very similar distribution shapes: their quantile–quantile relation is close to a unit-slope line with a small offset (Fig. S1), so choosing one or the other affects only the level of the wind, not its variability. Because every drought statistic in the paper is defined by a threshold relative to each series' own distribution, this choice has no effect on them; this is verified directly in main-text §6, so we adopt the eight-point daily mean.

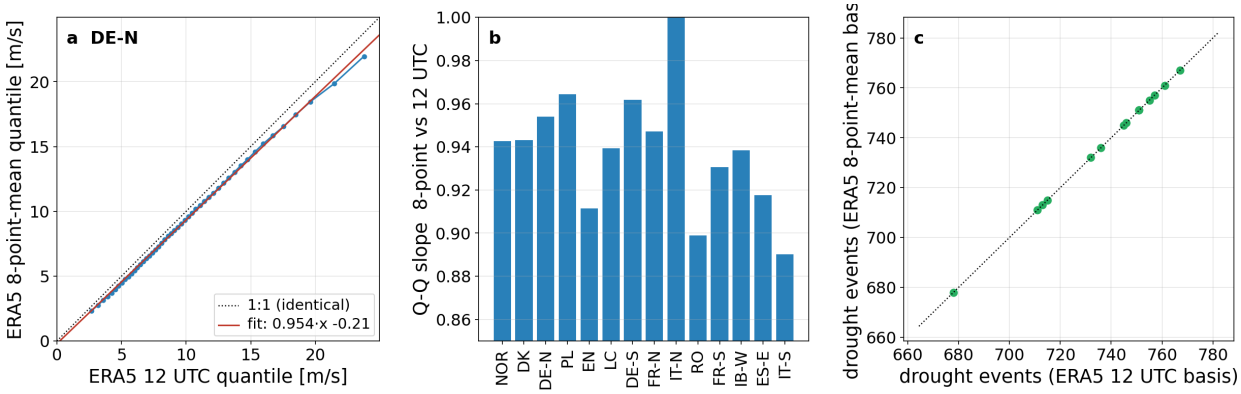


Figure S1: The two ways of sampling the modern ERA5 reference, a single 12 UTC snapshot and an eight-point daily mean, differ only in level, not in distribution shape. **(a)** for one region (N. Germany), the quantile–quantile relation of the eight-point-daily-mean quantiles (y) against the 12 UTC quantiles (x): a near-unit-slope line (fit $0.954\,x - 0.21$) lying just below the 1:1 diagonal. **(b)** the fitted quantile–quantile slope (eight-point versus 12 UTC) for all fourteen regions, all close to unity. **(c)** the number of NDJF wind-drought events detected under the two samplings, region by region, on the 1:1 line: the drought count is invariant to the choice of basis.

The ensemble-mean smoothing of the sparse early pressure field (main-text §4.3) dampens the sharp weather, so it is necessary to verify that the amplitude rescaling in §6 restores day-to-day variability rather than leaving a residual over-smoothing. Figure S2 shows how much variability remains in the corrected record compared with the modern reference, broken down by region; it confirms that the rescaling restores the amplitude of day-to-day variability without leaving evidence of residual oversmoothing.

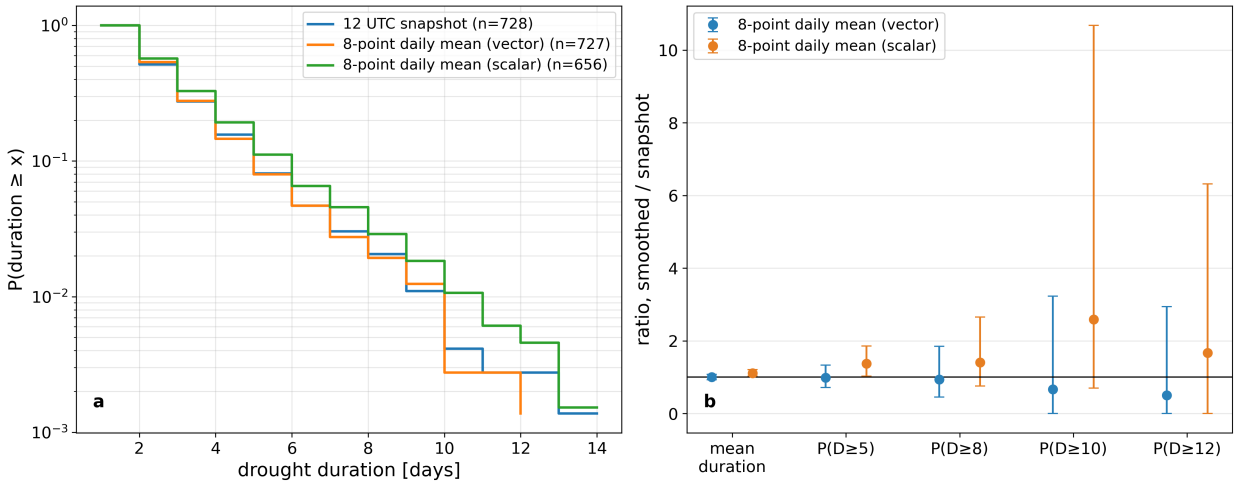


Figure S2: Day-to-day wind variability retained by the corrected record relative to the modern reference, region by region. The rescaling restores the amplitude of daily variability without evidence of residual oversmoothing.

S2 The barometer test, resolved by region

Main-text Fig. 4 shows the barometer test — the decomposition of the early-epoch wind bias into an additive (instrument-offset) component b and a multiplicative (ensemble-smoothing) component $(a-1)\mu$, in bins of daily temperature — for two representative regions. The temperature axis is a *test* axis, not a correction axis: the released correction carries no temperature term (§4.2). The figures below provide the full set of data that supports this conclusion.

Across all fourteen regions, the total bias follows the multiplicative (smoothing) component while the additive (barometer) component remains close to zero (Fig. S3). The uncertainties confirm this: the multiplicative component is bounded well away from zero in every region and temperature bin, whereas the additive component is consistent with zero in the majority of bins. Where the additive component's temperature slope differs from zero, it is *negative*, the opposite of the sign an under-corrected mercury column would produce (Fig. S4). The bias is therefore not an instrument error.

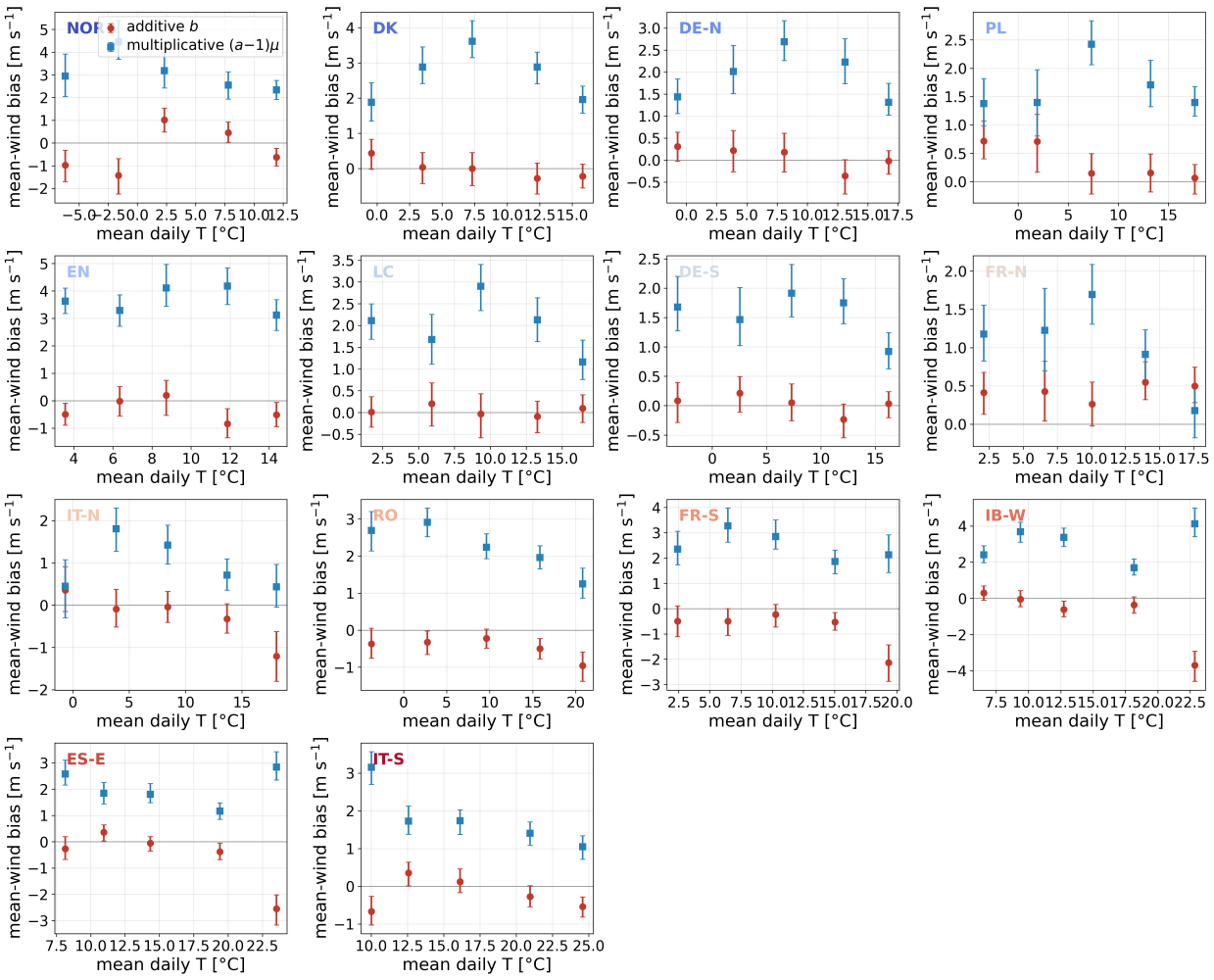


Figure S3: The barometer test performed on all fourteen released regions. For each region, the mean-wind bias between the sparsely observed early epoch (1806–1849) and the well-observed decades (1930–1980) is split, temperature bin by temperature bin, into an additive component (b, a fixed shift, the barometer signature) and a multiplicative component $(a-1)\mu$ (the smoothing signature). The two regions shown in main-text Fig. 4 (northern Germany and southern Italy) are drawn from this set. The per-region additive fraction and its 95 % confidence interval are listed in Table S1.

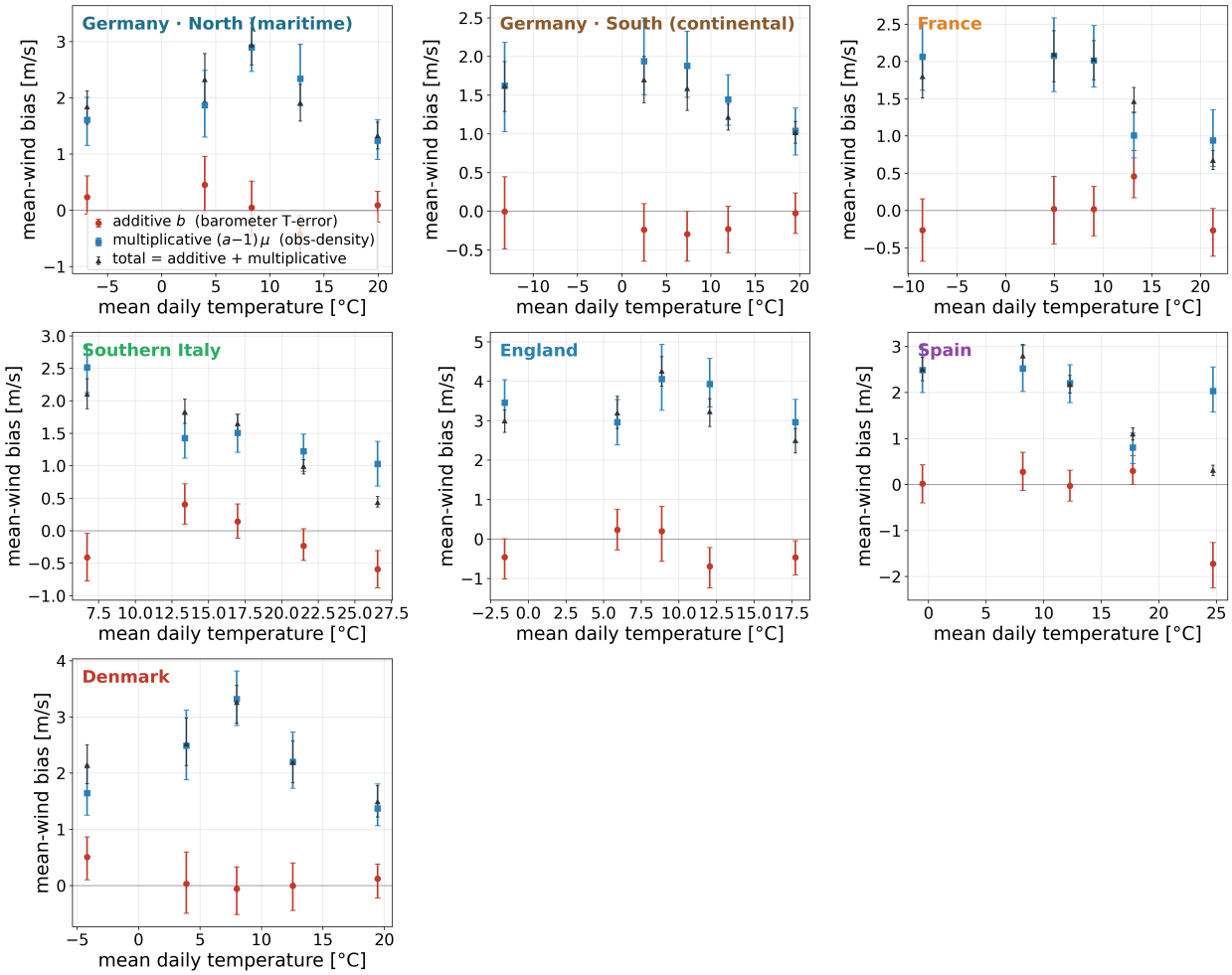


Figure S4: The additive component b and the multiplicative component $(a-1)\mu$ of Fig. S3, now with fit uncertainties: each point carries a 95 % confidence interval obtained by year-block bootstrapping (whole calendar years resampled within each epoch). The resulting per-region additive fraction is given in Table S1.

Table S1: Additive fraction of the barometer temperature error for all fourteen regions, the quantity shown per panel in Figs. S3 and S4: the additive component b as a fraction of the total temperature-dependent bias, $b/(b+(a-1)\mu)$, with a 95 % confidence interval from a year-block bootstrap. The character applies the rule additive if the interval lies wholly above 0.6, multiplicative if wholly below 0.4, and mixed otherwise.

Region	additive fraction [95 % CI]	character
NOR	0.22 [0.16, 0.27]	multiplicative
DK	0.08 [0.05, 0.15]	multiplicative
DE-N	0.10 [0.06, 0.17]	multiplicative
PL	0.17 [0.11, 0.26]	multiplicative
EN	0.10 [0.07, 0.15]	multiplicative
LC	0.05 [0.04, 0.15]	multiplicative
DE-S	0.07 [0.05, 0.16]	multiplicative
FR-N	0.35 [0.25, 0.45]	mixed
IT-N	0.31 [0.23, 0.45]	mixed
RO	0.19 [0.15, 0.22]	multiplicative
FR-S	0.22 [0.17, 0.26]	multiplicative
IB-W	0.18 [0.14, 0.23]	multiplicative
ES-E	0.20 [0.16, 0.25]	multiplicative
IT-S	0.18 [0.14, 0.23]	multiplicative

The temperature axis on which the barometer test is stratified must itself be reliable. It is: the daily-mean near-surface temperature drawn from the 20CRv3 reanalysis (the same historical source as the wind) is consistent with the modern reference (ERA5) in each region over their overlap (Fig. S5), and the year-to-year variability is consistent as well (Fig. S6), so the stratification is not distorted by a bias or a variance mismatch in the temperature field. The fitted warming trends for both sources and all regions are listed in Table S2.

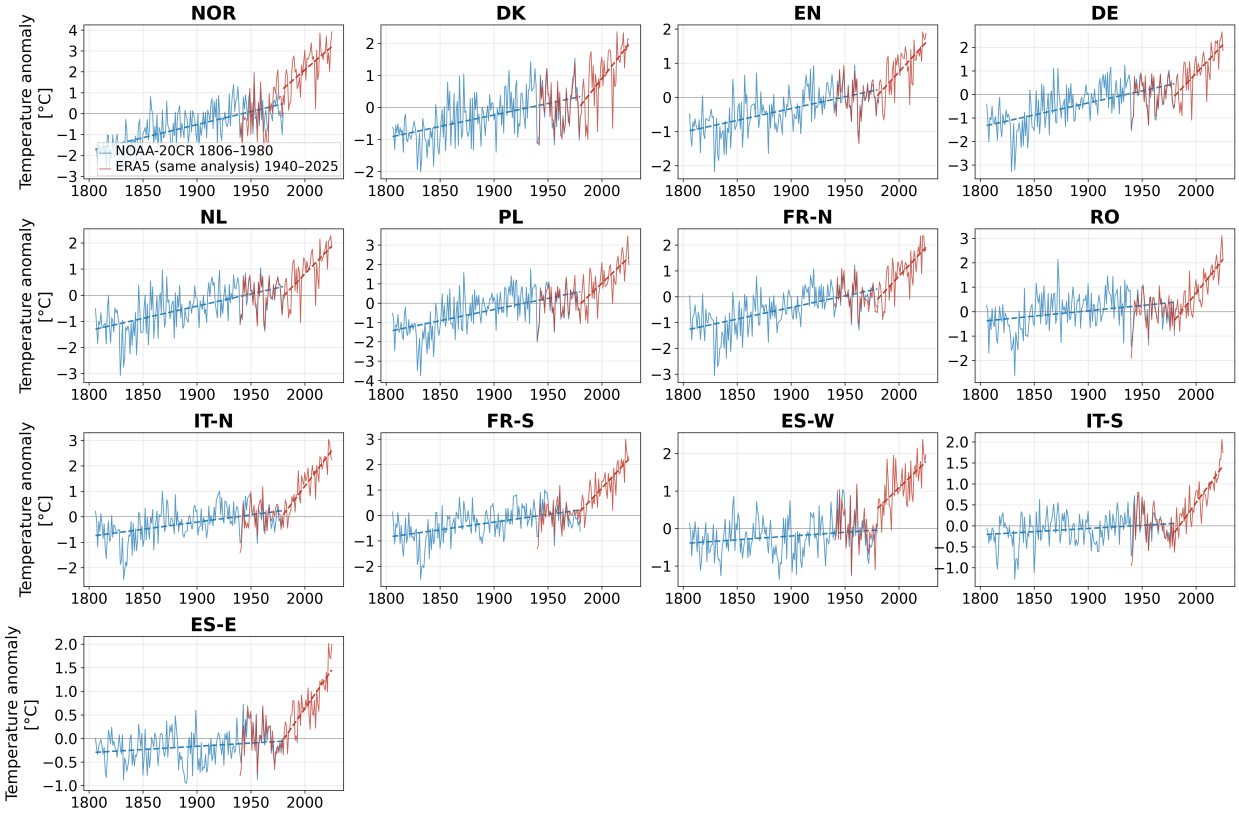


Figure S5: Daily-mean near-surface 2 m temperature, two independent sources analysed identically: 20CRv3 (blue, 1806–1980) continued by ERA5 (red, the modern reference, 1940–2025), with the same box and processing per region. Each panel is shown as an anomaly relative to the common 1940–1980 overlap mean, with a constant level offset (given per panel) removed so that the trends compare directly; dashed lines are linear fits, the ERA5 fit taken from 1980 on. The fitted trends for both periods and all regions are listed in Table S2.

Table S2: Fitted near-surface temperature trends by region for the two periods shown in Fig. S5: the 20CRv3 trend over 1806–1980 and the ERA5 trend from 1980 onward. These are the temperature-test boxes and do not coincide exactly with the fourteen released wind regions.

Region	20CRv3 1806–1980 (°C century^{-1})	ERA5 ≥1980 (°C century^{-1})
NOR	+1.25	+4.47
DK	+0.72	+4.27
EN	+0.69	+3.45
DE	+1.02	+4.87
NL	+0.94	+4.27
PL	+1.16	+5.35
FR-N	+0.90	+4.47
RO	+0.44	+5.54
IT-N	+0.56	+5.60
FR-S	+0.60	+4.55
ES-W	+0.20	+2.77
IT-S	+0.15	+3.52
ES-E	+0.14	+3.28

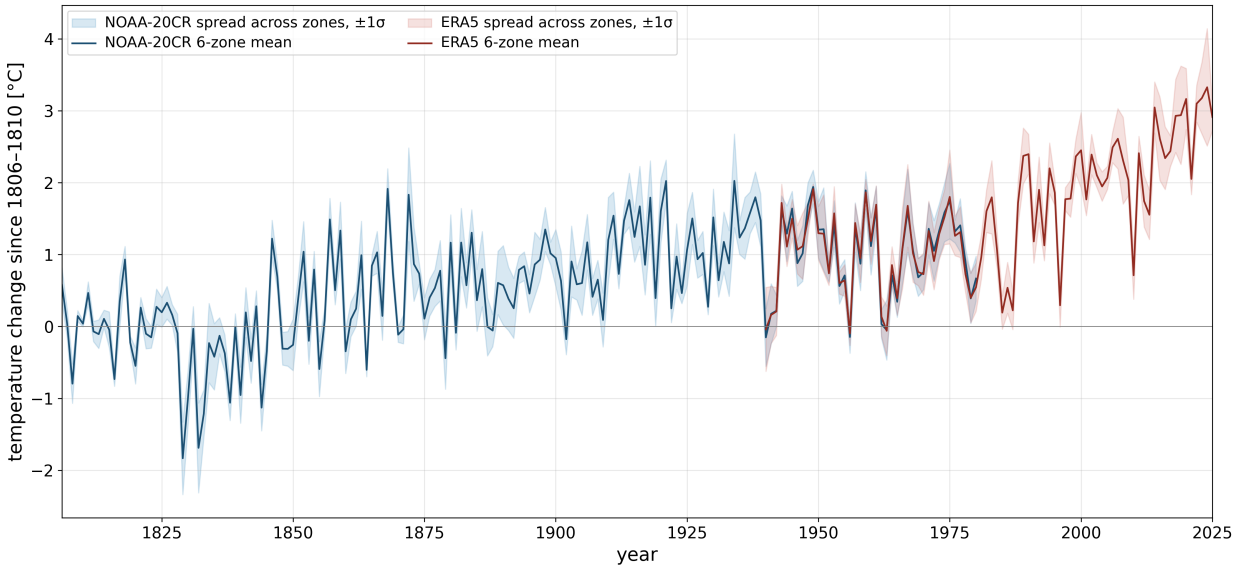


Figure S6: Interannual evolution of the near-surface temperature that stratifies the barometer test, as the anomaly since 1806–1810: the multi-region mean with its $\pm 1\sigma$ across-region spread, for the 20CRv3 reanalysis (blue) and ERA5 (red). The two agree over their overlap.

S3 Residual correction and stream continuity

Within each region, the relationship between the raw 20CR wind and the modern reference is approximately a straight line lying above the 1:1 diagonal, so the old wind is too weak by the same fraction at every wind speed. A single per-region multiplicative factor (the tilt of that line, quoted as the residual correction $100\,(1 - 1/a)$) therefore describes the residual completely, and it ranges from the flat, barely-biased regions (Germany S, +2.2 %) to the strongly channelled ones (Rhône corridor, +16.8 %). Figure S7 resolves this region by region; it is the per-region breakdown behind the summary bars of main-text Fig. 6.

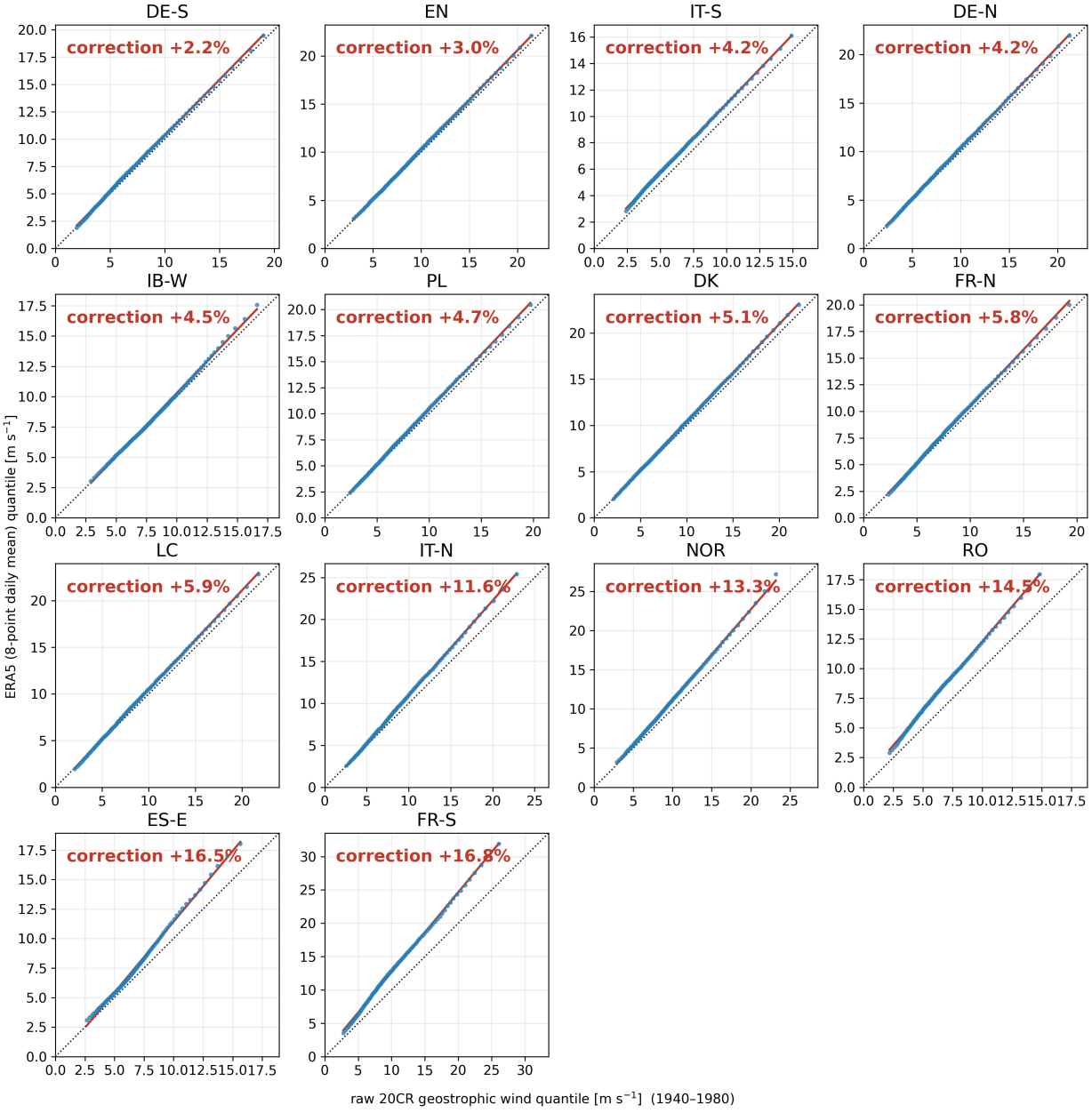


Figure S7: The residual correction resolved region by region: for each of the fourteen released regions, the quantiles of the raw 20CR wind (x) against those of the matched-basis modern reference (ERA5 as an eight-point daily mean, y) over their 1940–1980 overlap, with the 1:1 diagonal dotted. Panels are ordered by the residual correction $100\,(1 - 1/a)$.

The 20CRv3 pressure field is provided in two streams that meet at 1980/81: the sparser-input stream before and the modern fuller-input stream after. A reconstruction combining the two could in principle contain a discontinuity at this boundary. Before correction, however, the decadal ra-

tio of the modern reference to the raw reconstruction varies smoothly across 1980/81 (Fig. S8a), indicating that the two streams are mutually consistent for the present application. After correction, all regions remain continuous across the boundary, with a median ratio step of only +0.008 (Fig. S8b).

The correction is calibrated on the modern decades but applied across the full record, including both 20CR streams and the ERA5 extension. Continuity across the 1980/81 boundary therefore provides an out-of-sample check: the correction maps data outside the calibration period onto the modern reference without introducing a discontinuity.

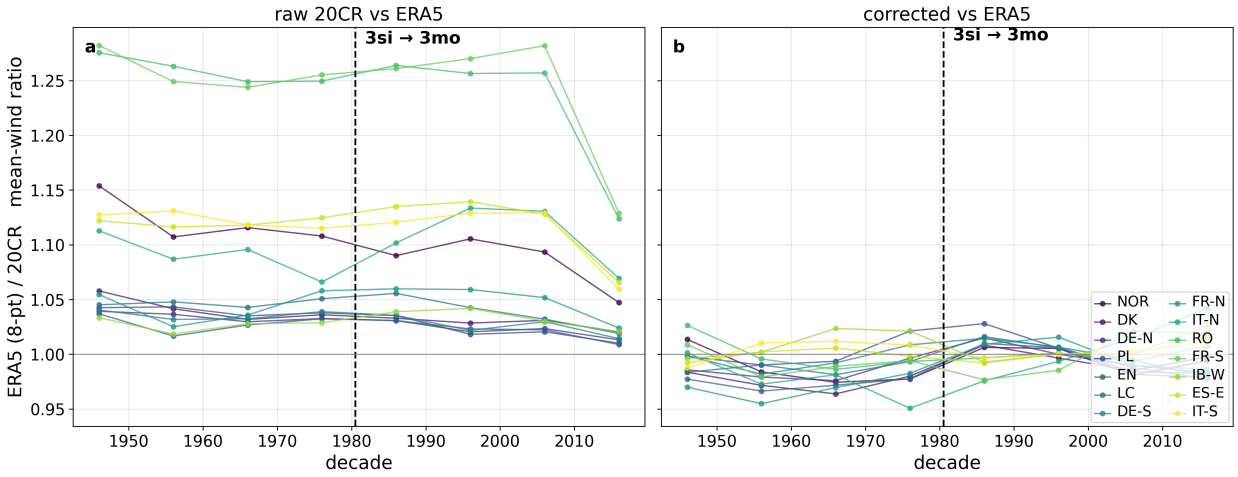


Figure S8: The decadal ratio of the modern reference based on the ERA5 eight-point daily mean to the reconstruction is shown for each region for the period 1941–2015, across the 20CRv3 production-stream change at 1980/81. **(a)** Raw 20CR: the ratio ranges from a few per cent in flat regions to around 25 per cent in channelled regions. The dashed line marks the 1980/81 boundary, where the source switches between the two 20CRv3 production streams. **(b)** After correction, the ratio is approximately 1 in every region and the median ratio step at the boundary is +0.008.

S4 Robustness of the barometer test: the intercept

The barometer test of §4.2 fits, within each region–temperature bin, the quantiles of the early reconstructed wind, Q_{early}, against those of the modern reference, Q_{ref}, using

$$Q_{\mathrm{ref}} = a\,Q_{\mathrm{early}} + b.$$

The slope a measures the multiplicative adjustment required to match the early and modern winds and is the only fitted quantity used in the correction. The quoted bias, $(a-1)\,\mu$, is evaluated at the regional mean wind μ, within the observed range. By contrast, the intercept b corresponds to the extrapolated value of Q_{ref} at zero wind and is neither used nor interpreted.

Figure S9 illustrates this distinction. Allowing a quadratic term lowers the residual, but makes the intercept larger and reverses its temperature trend in several regions, showing that this extrapolated parameter is not robust. The multiplicative bias $(a-1)\,\mu$, on which the conclusion of §4.2 rests, remains constrained by the bulk of the data and is essentially unchanged. The additional curvature therefore redistributes variance with the extrapolated intercept without altering the physical conclusion.

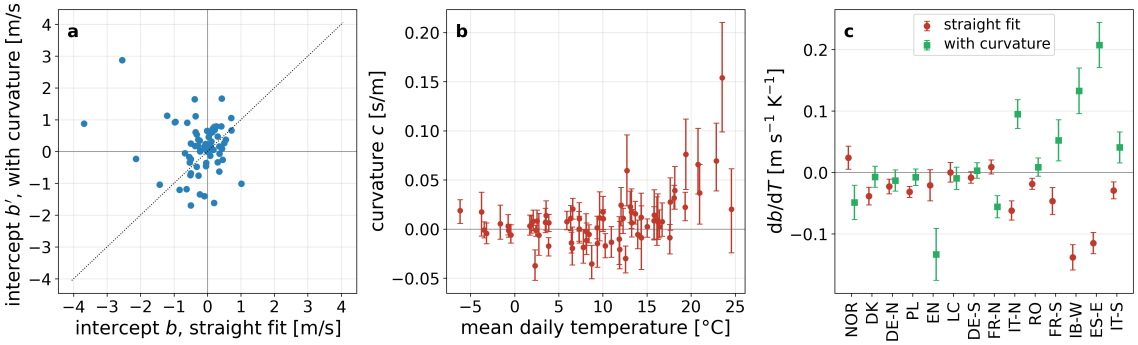


Figure S9: Curvature test of the barometer-test fit (§4.2) over the seventy region–temperature bins. The straight-line fit $Q_{\mathrm{ref}} = a\,Q_{\mathrm{early}} + b$ (slope a, intercept b) is compared with a version carrying an added quadratic term, $Q_{\mathrm{ref}} = a'Q_{\mathrm{early}} + b' + c\,Q_{\mathrm{early}}^2$ (curvature c); Q_{early} and Q_{ref} are the wind quantiles of the early reconstruction and of the modern reference. **(a)** root-mean-square residual of the two fits, 0.146 versus 0.085 $\mathrm{m\,s^{-1}}$. **(b)** the fitted curvature c per bin, differing from zero in 23 of the 70 bins, with no consistent sign. **(c)** the fitted intercept (b and b') against bin temperature, by region, for both fits.

S5 Effect of the observing network on drought persistence

The released reconstruction uses the mean of the eighty members that 20CRv3 provides. Each member is a separate reconstruction of the atmosphere that is consistent with the available pressure observations and their error. Where observations are sparse they pin down the position and movement of pressure systems only loosely, so the members place a given system in slightly different places. Averaging the members then blurs the pressure field: its gradients weaken, and with them the geostrophic wind. The blurring grows further back in time and with distance from the nearest observation (Fig. S10), and it can also shift the timing of a drought when members disagree on the day it ends.

To compare persistence between periods we borrow the drought-duration law of the companion paper. The chance that a drought lasts at least D days is

$$P(\geq D) = \left[1 + (q-1)\,\frac{D}{\tau}\right]^{-1/(q-1)}, \tag{1}$$

where τ (in days) sets the typical length and q sets how heavy the tail is: $q \to 1$ is the memoryless case $e^{-D/\tau}$, and $q > 1$ means droughts persist longer than that. We use q and τ here only as compact summaries; the companion paper (Paul, submitted to *Nature Communications*) develops the law, how it is fitted, and the return levels it implies.

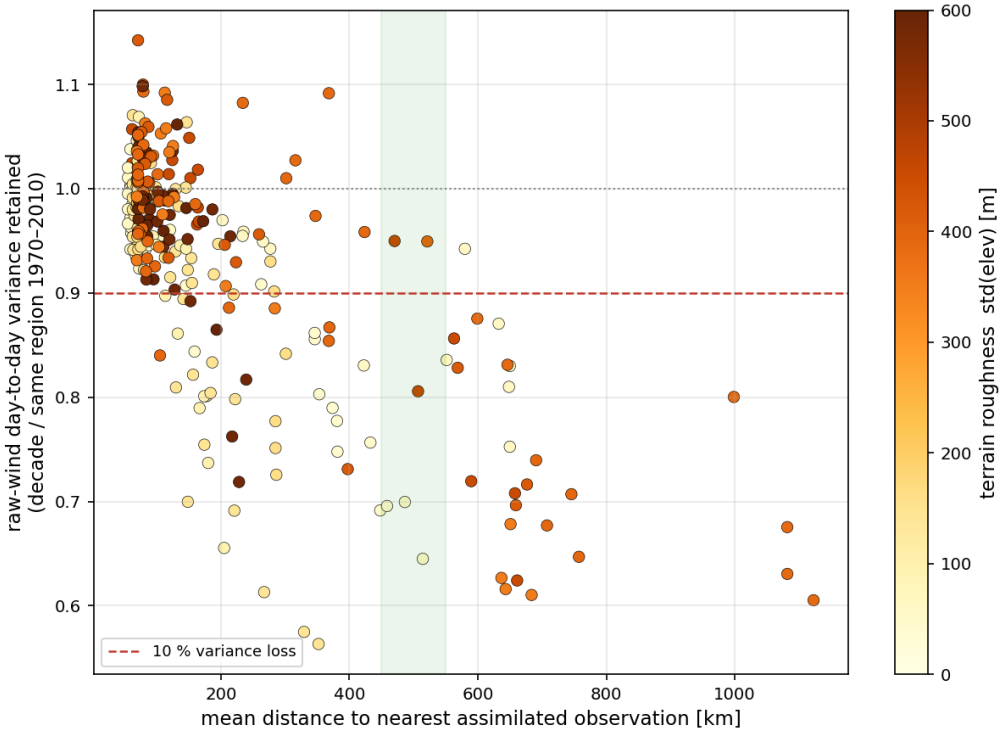


Figure S10: Artificial smoothing increases with distance to the nearest assimilated observation. Day-to-day winter variance retained by the raw reconstruction, normalised to the 1970–2010 reference of each region, versus mean distance to the nearest assimilated pressure observation. Points show region–decade values for 1810–2010; colour indicates terrain roughness (standard deviation of elevation), and the green band marks the ∼500 km coverage scale. Spearman $\rho = -0.63$ ($p < 10^{-30}$, $n = 294$).

An alternative combination order would derive the geostrophic wind for each member first and then average the eighty winds. Because wind speed is a nonlinear function of the gradient, this per-member average is systematically higher than the wind from the mean field, a Jensen inequality, with the gap largest in the sparse early record and shrinking as the network densifies. The difference is one of absolute level only. On a pooled sample of eleven early years (1811–1896) that retain

most of the eighty members, the two orders give the same wind-drought duration law within the fit scatter (pooled shape $q = 1.00$ versus 1.01 and scale $\tau = 3.2$ versus 3.1 d for the per-member and mean-field winds; the τ elevated in both reflects the raw early-record smoothing that the released detector and quantile map remove), and the per-member ensemble uncertainty is likewise unchanged. The released product therefore uses the mean-field wind, whose level the quantile map anchors to ERA5, with no consequence for the drought statistics or their uncertainties.

If this ensemble-mean smoothing is responsible for the early record's behaviour, applying comparable temporal smoothing to the well-observed modern record should reproduce its statistical signature. We therefore smooth the dense modern epoch (1980–2024, all regions pooled) with running-mean windows of one to seven days and repeat the identical drought detection and duration-law fit (Fig. S11). The fitted parameters move systematically along a single locus: q falls from 1.11 to 0.99 while τ lengthens from 1.8 to 5.6 d. The early record (1806–1849; $q = 1.05$, $\tau = 2.1$ d) lies along the same direction from the full-record pooled fit and is statistically consistent with the approximately one-day temporal smoothing inherent in the reconstruction.

Adding first-order autoregressive (AR(1)) measurement noise, in which each day retains a fixed fraction of the previous day's error so that the errors are correlated from one day to the next rather than independent, to the same modern record instead lowers q while shortening τ ($q = 1.06$, $\tau = 1.6$ d at forty per cent injected error; §S6.4). The opposite response of τ distinguishes smoothing from measurement noise, and the early record follows the smoothing signature. This experiment identifies a mechanism consistent with the early behaviour; it does not attribute the effect to a particular source. The region-dependent recommended start years (§6.5) incorporate this additional uncertainty.

The quantile correction of §6 restores the wind-amplitude distribution but, because it preserves rank, cannot undo temporal smoothing that has already merged separate calm spells. Persistence statistics in the earliest decades therefore carry reduced confidence even where the wind amplitude is fully corrected, motivating the region-specific recommended start years provided with the released product (§6.5).

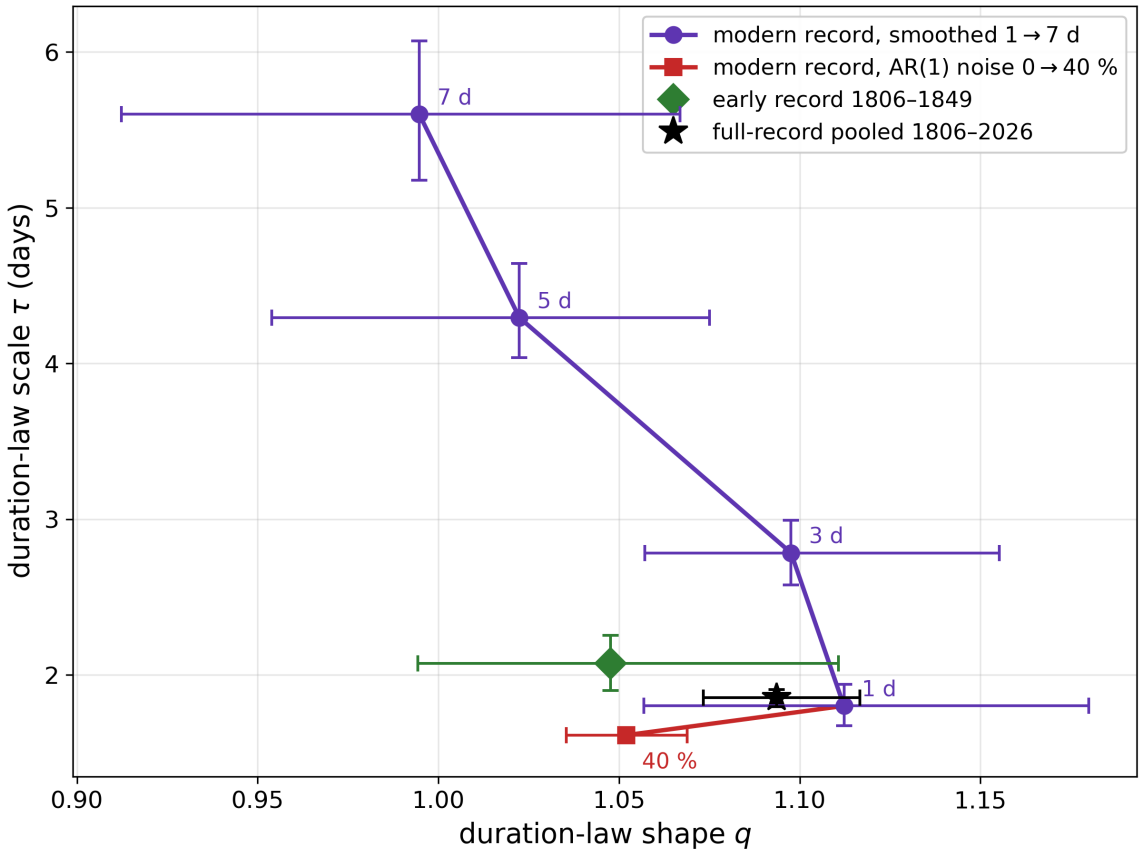


Figure S11: Duration-law parameters recovered from the modern, well-observed record (1980–2024, all regions pooled) as it is artificially degraded, shown in the shape–scale (q, τ) plane. Purple: progressive smoothing of the daily wind with running-mean windows of one to seven days. Red: addition of AR(1) measurement noise with lag-1 correlation $\rho = 0.36$, from 0 to 40 % (see text). Green diamond: the fit to the early record (1806–1849); black star: the pooled fit to the full record. Error bars show 95 % confidence intervals from a winter-year block bootstrap. For the full-record and 40 %-noise points, the large sample constrains τ to about ± 0.05 d, so the vertical error bar is shorter than the symbol size.

S6 Uncertainty of the reconstructed wind, and the reliability horizon

This section supports the uncertainty and reliability statements in the main-text §8. The wind uncertainty is estimated directly from the spread of the eighty alternative atmospheric reconstructions provided by the 20CRv3 ensemble. We quantify its magnitude and temporal persistence, test its effect on the fitted duration law, examine its relation to the observing network, and compare the resulting reliability with the region-specific recommended start years.

S6.1 Size of the uncertainty

For each 20CRv3 member we calculate the geostrophic wind independently from the 500 hPa geopotential-height field, $v_g = (g/f)\,|\nabla Z|$. The ensemble provides $M = 80$ such members, indexed $m = 1, \ldots, M$; we write $v_g^{(m)}(t)$ for the wind of member m at time step t. At each t the member winds have an ensemble mean $\overline{v_g}(t)$ and an across-member standard deviation $\Delta v_g(t)$,

$$\overline{v_g}(t) = \frac{1}{M}\sum_{m=1}^{M} v_g^{(m)}(t), \qquad \Delta v_g(t) = \sqrt{\frac{1}{M-1}\sum_{m=1}^{M}\left(v_g^{(m)}(t) - \overline{v_g}(t)\right)^2}. \tag{2}$$

Here Δv_g is the across-member spread and provides the ensemble-based estimate of wind uncertainty in physical units ($\mathrm{m\,s^{-1}}$). Throughout this section we express it as the fractional uncertainty

$$u(t) = \frac{\Delta v_g(t)}{\overline{v_g}(t)}, \tag{3}$$

a dimensionless number; the entries in Table S3 and in Figs. S13 and S14 are its NDJF mean. Because geostrophic wind depends on the pressure gradient, spatially coherent differences in geopotential height between ensemble members largely cancel. The resulting wind uncertainty is therefore much smaller than one would infer from the pointwise height spread alone.

Table S3: Fractional uncertainty $u(t)$ (Eq. 3) of the reconstructed geostrophic wind at representative epochs, by averaging time scale.

Time scale	1836	1930	2000
instantaneous (3 h)	∼33 %	∼18 %	∼11 %
daily mean	∼33 %	∼16 %	∼10 %
seasonal (NDJF) mean	∼5 %	∼4 %	∼2 %

The uncertainty does not average away substantially within a day. Ensemble differences remain correlated on synoptic timescales: the autocorrelation of the wind deviation is 0.96 at 3 h and 0.66 at 12 h. The eight sub-daily samples therefore correspond to only about 1.3 effectively independent samples per day. Seasonal averaging, by contrast, reduces the uncertainty to a few per cent.

The resulting fractional uncertainty is about 30–33 % in the sparse early record and about 10 % in the modern period.

S6.2 Effect on drought statistics

A large uncertainty in wind speed on an individual day does not imply an equally large uncertainty in drought duration. We test this directly by adding correlated noise to the well-observed modern record (§S6.4).

The fitted duration statistics remain within their baseline uncertainty up to about 30 % relative daily error. At larger errors the inferred persistence weakens rather than strengthens. The level of wind uncertainty found in the early record cannot create the persistence seen in the reconstruction.

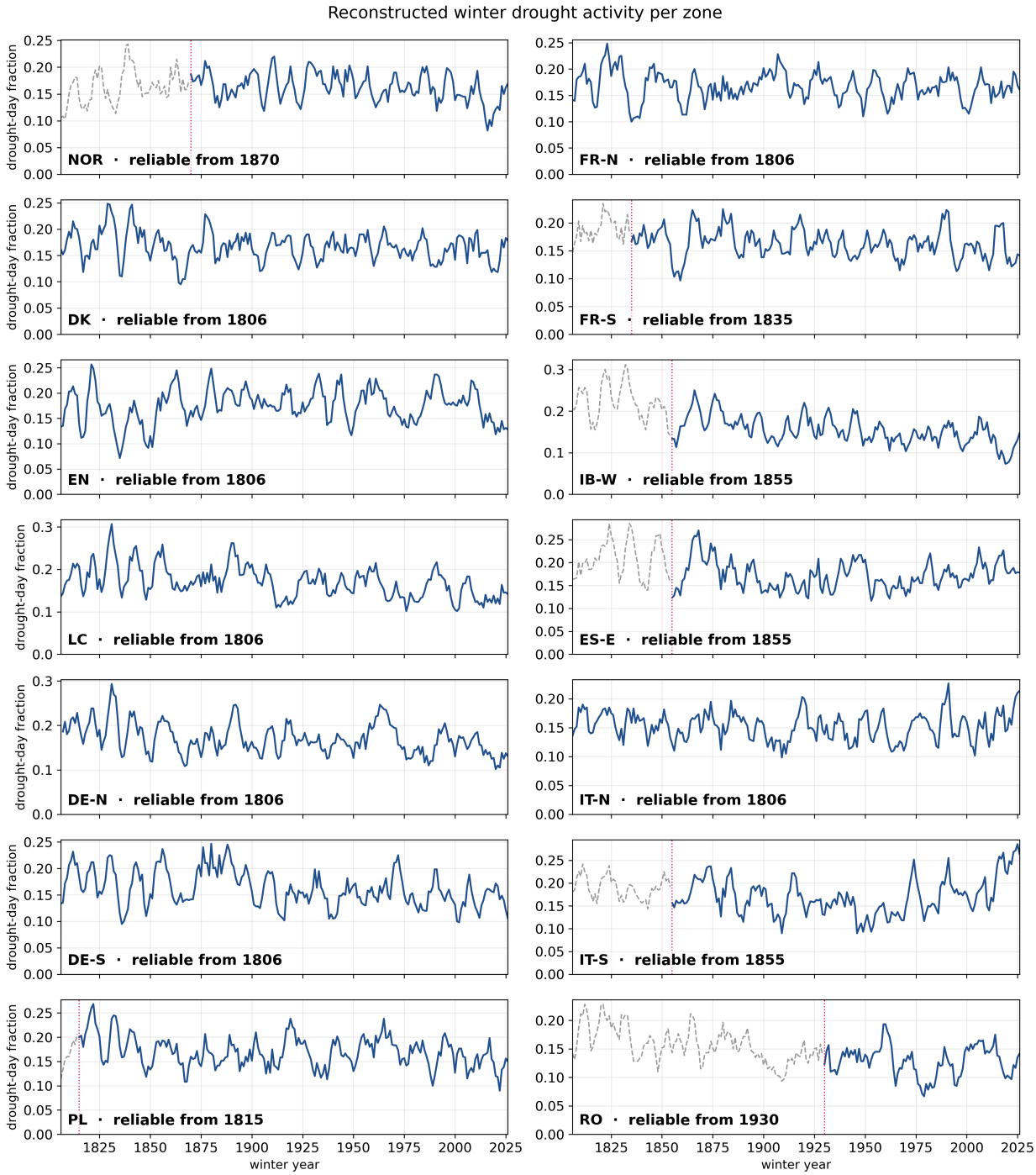


Figure S12: Per-region NDJF wind-drought-day fraction, 1806–2026 (five-winter running mean). The red dotted line marks each region's recommended start year (main-text §6.5); the curve is drawn solid from that year on and dashed before it, where the reconstruction is not used for quantitative event statistics.

The earliest decades are nevertheless less reliable for individual events. Uncertainty near the threshold can change short droughts through marginal crossings, while temporal smoothing can merge neighbouring calm intervals and shift the start or end of longer droughts. The region-specific recommended start years mark the period from which event-level statistics are used quantitatively. Earlier values remain in the released reconstruction and are shown dashed in Fig. S12, but are not included in those statistics.

S6.3 Dependence on the observing network

The reconstruction becomes more certain as more historical surface-pressure observations are available to constrain it. Across seven anchor years and fourteen regions, daily fractional uncertainty decreases from about 34 % to 10 % as the number of surface-pressure reports assimilated into 20CRv3 rises from approximately 10^2 to 10^5 per NDJF season (Fig. S13; Spearman $\rho = -0.83$). The dependence on observation spacing is weaker (Fig. S14; $\rho = +0.75$).

The number of assimilated observations predicts regional uncertainty better than their geographical spacing.

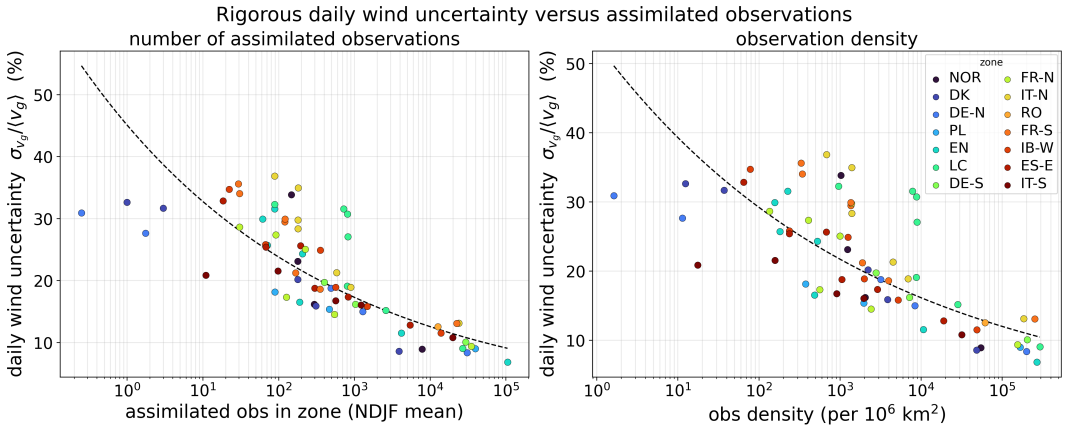


Figure S13: Daily fractional wind uncertainty versus the number of assimilated surface-pressure observations, for seven anchor years and fourteen regions. Left: number of assimilated observations; right: observation density (per 10^6 km^2). The uncertainty falls as observations increase; the dashed lines are power-law fits with slope -0.14 and Spearman $r = -0.83$ for the number of observations and slope -0.13 and $r = -0.75$ for the density, so the number of observations organises the regions marginally better.

S6.4 Noise-injection test

We test the sensitivity of the fitted duration statistics to observational error by adding correlated noise to the modern corrected record (the released, quantile-mapped wind; 1980–2024, fourteen regions pooled). The added noise is AR(1) with the observed day-to-day correlation ($\rho = 0.36$); we then repeat the entire drought detection and fit (Fig. S15).

Without added noise the fit gives $q = 1.110\,[1.074, 1.174]$ ($n = 3443$). The fitted value remains within this baseline interval up to about 30 % added daily error and departs from it only near 35 %. Even at 33 %, the level of the sparse early era, q moves only to about 1.08, and always towards smaller q and shorter τ, that is, towards *weaker* persistence. This is a conservative robustness test: observational error of the size estimated for the early record cannot produce the persistence found in the reconstruction.

S6.5 Uncertainty at the recommended start years

Figure S16 compares the ensemble-derived fractional wind uncertainty with the region-specific recommended start years. At those start years, the mean uncertainty across the fourteen regions is 30.2 %. Northern regions generally lie near this value, with some at 30–35 %, whereas the quieter southern regions are lower, at 19–26 %.

The two quantities were determined independently. Recommended start years are based on observation coverage and terrain (main-text §6.5), whereas the approximately 30 % robustness level comes from the noise-injection test (§S6.4). Their agreement is therefore a consistency check: the

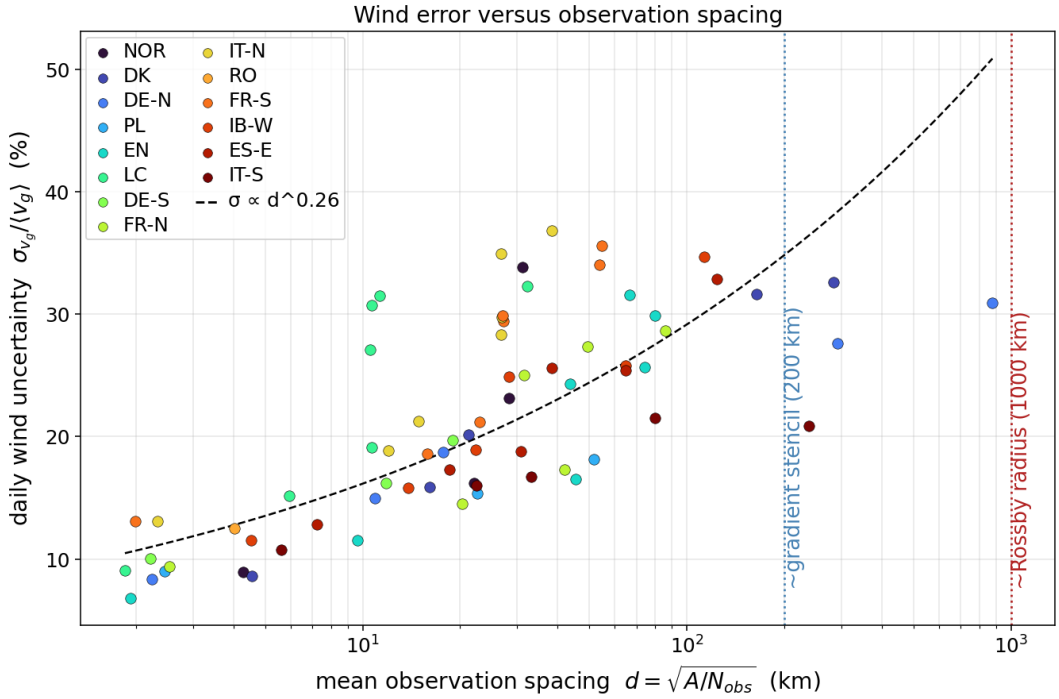


Figure S14: Daily fractional wind uncertainty versus mean inter-observation spacing, $d = \sqrt{A/N_{\mathrm{obs}}}$, for seven anchor years and fourteen regions. The uncertainty stays low while the spacing is below the finite-difference gradient stencil (∼200 km, blue) and rises once observations thin past it toward the Rossby radius (∼1000 km, red); the dashed power-law fit is $\sigma \propto d^{0.26}$ with Spearman $r = 0.75$.

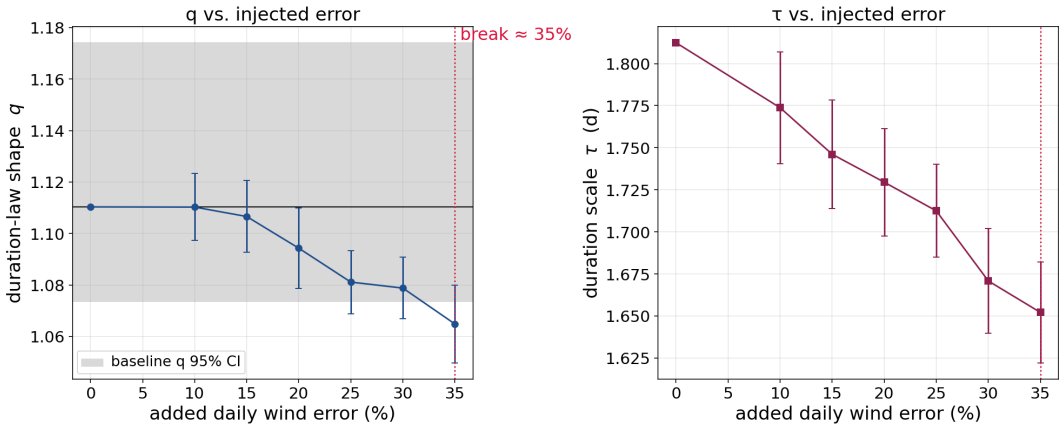


Figure S15: Reliability benchmark for the duration-law fit: parameters recovered after adding AR(1) noise of increasing relative amplitude to the modern corrected record, shape q (left) and scale τ (right). The shaded band is the baseline 95 % confidence interval and the dashed vertical marks where q first leaves it (about 30–35 % added daily error), the tolerance quoted in §S6.4.

observation-based start years generally begin where the ensemble-derived wind uncertainty has fallen to a level at which the duration statistics remain robust.

Data support. The full 80-member reconstruction is available at 31 years spanning 1806–2015 (about one every five years through the nineteenth century, then 1930 and 2000–2015). The uncertainty-versus-coverage regressions above use the seven of these for which the assimilated-observation counts were also compiled (1806, 1836, 1855, 1870, 1900, 1930 and 2000); the remaining years carry the same per-member winds and bearings, so the wind can be reconstructed for each member and averaged afterwards (§S5) and the directional uncertainty evaluated (§S7) at each of

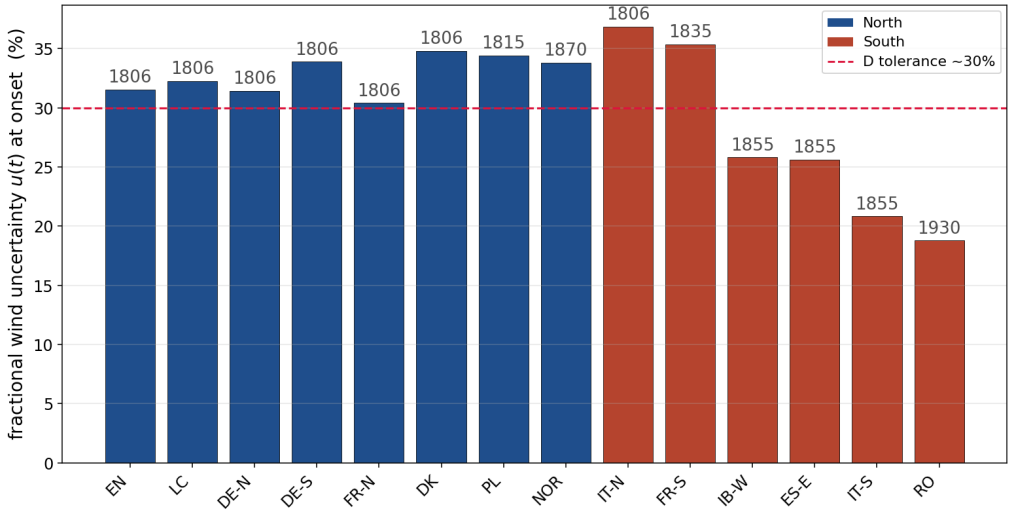


Figure S16: Fractional wind uncertainty $u(t)$ at each region's recommended start year (blue: northern regions; red: southern), interpolated where the start year falls between anchor years. The dashed line is the noise-injection tolerance of §S6.4 (≈30 %); the comparison asks whether the observation-based onset corresponds to a common wind-error level.

them. These years densely cover 1806–1886, where most regions cross their reliability horizon.

S7 Reconstructed wind direction and its uncertainty

Directional uncertainty. The pressure-based reconstruction provides wind direction as well as wind speed. For each winter (NDJF) time step, we calculate the geostrophic bearing of all 80 ensemble members relative to the ensemble-mean bearing and pool these deviations over the analysis period. Their circular standard deviation,

$$\sigma_{\mathrm{dir}} = \sqrt{-2 \ln \bar{R}}, \qquad \bar{R} = \left| \left\langle e^{i\delta} \right\rangle \right|,$$

measures the spread among ensemble members and hence the uncertainty in the reconstructed wind direction. Figure S17 compares this quantity between a sparse early period (1806–1835) and a dense modern period (2000–2015) for four representative regions.

The mean wind direction remains strongly westerly in both periods (274–292°), whereas the across-member directional uncertainty falls from 45–56° in the early period to 14–23° in the modern period. Table S4 gives the corresponding values for all fourteen regions, where the directional uncertainty likewise decreases by roughly a factor of two to five while the mean direction changes little.

Table S4: Across-member (reconstruction) uncertainty of the daily geostrophic wind direction, and the mean bearing the wind blows from, for all fourteen regions in a sparse early period (1806–1835) and a dense modern one (2000–2015). The uncertainty is the circular standard deviation of the 80 ensemble members about the ensemble-mean bearing of the same time step, pooled over all NDJF time steps.

		mean bearing (°)		member spread (°)	
Code	Region	early	modern	early	modern
NOR	Southern Norway	291	274	56	17
DK	Denmark	291	274	51	15
DE-N	North German Plain	284	274	45	14
PL	Poland	277	270	48	14
EN	England	280	283	49	10
LC	Low Countries	287	280	45	14
DE-S	Southern Germany	280	278	46	15
FR-N	Northern France	290	285	44	13
IT-N	Po Basin	281	284	52	20
RO	Romania	268	272	49	18
FR-S	Rhône Corridor	293	294	50	20
IB-W	Atlantic Iberia	291	292	55	20
ES-E	Spanish Mediterranean	292	292	50	23
IT-S	Southern Italy	278	286	45	18

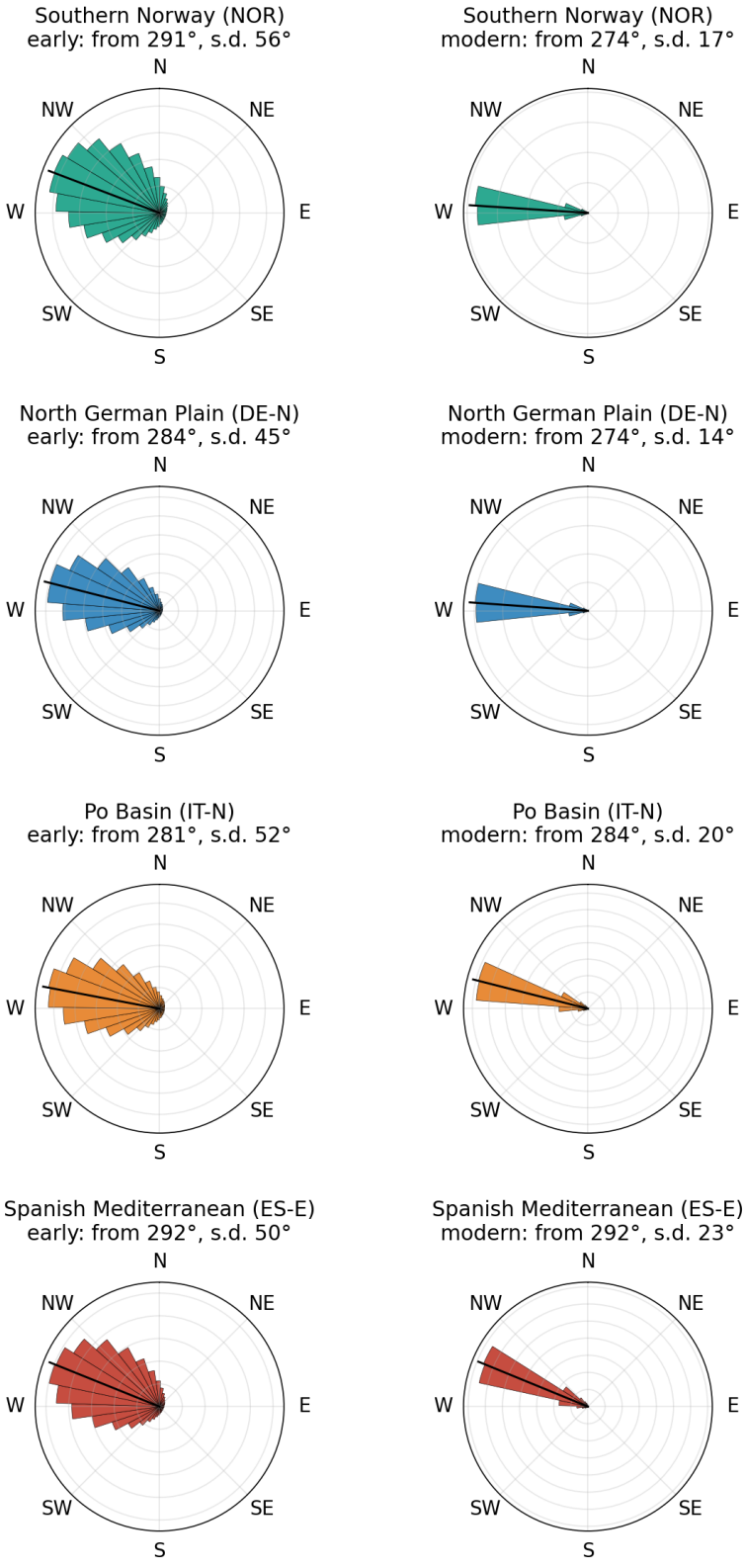


Figure S17: Reconstructed geostrophic wind-direction uncertainty for four regions (Southern Norway, the North German Plain, the Po Basin and the Spanish Mediterranean), shown for a sparse early period (1806–1835, left) and a modern period with dense observations (2000–2015, right). Each circular distribution shows the NDJF bearings of the 80 ensemble members relative to the ensemble-mean bearing at each time step, pooled over the period and centred on the climatological wind direction (black line; direction from which the wind blows). Bearings are measured clockwise from north. Labels give the mean bearing and the circular standard deviation across ensemble members.