

Power spectral estimate for discrete data

Norbert Marwan^{1,2, a)} and Tobias Braun¹

¹⁾*Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Telegrafenberg A31, 14473 Potsdam, Germany*

²⁾*University of Potsdam, Institute of Geoscience, Karl-Liebknecht-Straße 32, 14476 Potsdam, Germany*

(Dated: 23 April 2023)

The identification of cycles in periodic signals is a ubiquitous problem in time series analysis. Many real-world data sets only record a signal as a series of discrete events or symbols. In some cases, only a sequence of (non-equidistant) times can be assessed. Many of these signals are furthermore corrupted by noise and offer a limited number of samples, e.g., cardiac signals, astronomical light curves, stock market data, or extreme weather events.

We propose a novel method that provides a power spectral estimate for discrete data. The edit distance is a distance measure that allows to quantify similarities between non-equidistant event sequences of unequal lengths. However, its potential to quantify the frequency content of discrete signals has so far remained unexplored. We define a measure of serial dependence based on the edit distance which can be transformed into a power spectral estimate (EDSPEC), analogously to the Wiener-Khinchin theorem for continuous signals.

The proposed method is applied to a variety of discrete paradigmatic signals representing random, correlated, chaotic, and periodic occurrences of events. It is effective at detecting periodic cycles even in the presence of noise and for short event series. Finally, we apply the EDSPEC method to a novel catalogue of European atmospheric rivers (ARs). ARs are narrow filaments of extensive water vapor transport in the lower troposphere and can cause hazardous extreme precipitation events. Using the EDSPEC method, we conduct the first spectral analysis of European ARs, uncovering seasonal and multi-annual cycles along different spatial domains. The proposed method opens new research avenues in studying of periodic discrete signals in complex real-world systems.

Many dynamical processes exhibit characteristic periodic time scales, which can be assessed by power spectral density estimation. This is the usual and standard procedure of time series analysis, where the data is, in general, equally sampled and follows a more or less continuous nature. However, in specific applications, only events are observable, which could be irregularly occurring activities or extreme events. Identification of cyclical behaviour in such kind of data is difficult. We propose a novel but simple tool to estimate the power spectral density in such discrete data. This method can be applied on symbolic sequences, discrete data, or (extreme) event series.

I. INTRODUCTION

The estimation of the power spectral density (PSD) of cyclical signals is an important field in signal and data analysis¹⁻⁴. However, the estimation of PSD from discrete data, such as symbolic time series or event series, is difficult. A time series can generally be denoted by a set of ordered pairs $\{(t_i, x_i)\}$ of time t_i with $t_{i+1} > t_i$

and corresponding data value x_i ; and with sampling index i . PSD estimation becomes even more challenging, if instead only the time points of the events are available (i.e., in contrast to the time series definition above, we have only a set of time points $\{t_i\}$ which indicate the presence of an event; time series analysis tools usually handle only data in the form $\{(t_i, x_i)\}$).

Analysing discrete or event series is of general interest in many scientific fields. For example, heart beats are discrete events, described as the R-waves in an electrocardiogram (ECG). Pathologic states cause significant changes in the beat-to-beat intervals, which can be used to diagnose or even predict life-threatening cardiac events^{5,6}. Specific methods designed for discrete time series analysis are available for classifying cardiological signals, based on symbolic dynamics and ordinal patterns^{7,8}. Brain activity is controlled by the firing of neurons, occurring as sequences of events, also called “spike trains”⁹. The synchronization between firing neurons is the base of transmitting information through the nervous system, but can also cause serious pathological states, e.g., seizures^{10,11}. Symbolic dynamics is a promising tool to study seizure activity using electroencephalographic measurement¹². In climatology, a profound understanding of extreme weather events such as droughts, floods or heavy precipitation is becoming more and more crucial^{13,14}. Detrimental societal and technological impacts of single events and cascades of compound weather extremes are projected to increase on a warming planet. In this context, atmospheric rivers (ARs) receive increasing interest as they can trigger extreme

^{a)}marwan@pik-potsdam.de

rainfall¹⁵. Frequency and intensity of AR events are controlled by the Madden-Julian Oscillation, the Quasi-Biannual Oscillation, the El Niño/ Southern Oscillation (ENSO), and the Pacific Decadal Oscillation^{16,17}. Extreme rainfall and river flood event series are studied with event synchronization, edit distance, and complex networks to shed light on their serial dependency and spatial propagation^{18,19}. Discrete data also appears in engineering, e.g., in the investigation of how weather conditions affect car crashes, using integer-valued autoregressive models²⁰. In the analysis of stock markets, identification of regularities in the occurrence of trades in the order book or (ultrafast) extreme events is a common research objective^{21,22}. The temporal structure, heterogeneity, and serial dependencies in financial data are studied using response functions, spectral analysis, and recurrence plots^{23–25}. Communication between human individuals via social media platforms is intricately linked to recurrence of periodic social events and is, for instance, studied by means of periodic patterns in tweets on Twitter²⁶. Finally, detection of periodicities in irregular and event-like time series is also one of the main challenges faced in computational astrophysics^{27,28}.

Although discrete or (extreme) event time series are ubiquitous in many scientific disciplines, the study of cyclicities within these events is, to our best knowledge, still in its infancy. Potential methods would be Walsh transform^{29–31} and spectral analysis using the Haar wavelet^{32–34}. The Walsh transform provides a specific set of discrete basis functions consisting of binary sequences of different shape and length. Therefore, the interpretation of a plain Walsh spectrum is not straightforward. Nevertheless, Walsh transforms are recently receiving more interest, e.g., in combination with machine learning tools for classification purposes³⁵. Applications of the Haar wavelet to non-equidistant and event-like data have so far remained sparse³⁶. While not providing a PSD-like framework, another family of methods based on line-search algorithms, phase folding and maximum-likelihood estimation aims at finding periods in event series³⁷. Finally, some methods characterise event time series based on inter-event times^{7,38}.

Haar wavelet and Walsh power spectrum require time series in the form of ordered pairs $\{(t_i, x_i)\}$. Walsh power spectrum even needs equidistant time series (i.e., $t_{i+1} - t_i = \text{const.}$). However, series of (extreme) events, e.g., extracted from data by thresholding and further considering only the times of these events t_i , consist only of the set of these time stamps $\{t_i\}$. Alternatively, if assigning corresponding values to get an (extreme) event time series $\{(t_i, x_i)\}$, it would be one which is strongly non-equidistant (i.e., $t_{i+1} - t_i \neq \text{const.}$) or, as a further alternative, to ensure equidistant sampling, it would consist mainly of zero values. Most methods of time series analysis are limited for such kind of data. Specific methods considering the extreme nature of discrete data points have their roots in neuroscience^{39–41}. The main purpose of these methods is testing coincidence or syn-

chronization between event series. Among these methods are event synchronization⁴⁰, edit distance³⁹, and coincidence analysis⁴². Event synchronization has been applied in climate science to construct complex networks to investigate the spatio-temporal dynamics of extreme rainfall^{18,43}. Edit distance has been applied in different applications and in combination with recurrence analysis, to study serial dependence^{19,24,44} and to allow analysing non-equidistantly (palaeoclimate) time series^{45,46}.

We propose a novel power spectral analysis for discrete and event series $\{t_i\}$ based on the edit distance metric (modified Levenshtein metric^{39,47}). This method allows to estimate a power spectrum directly from the event sequence without interpolation.

In Sect. II we introduce the edit distance and the power spectral estimation based on this metric. We illustrate this approach in Sect. III on selected paradigmatic examples of random, periodic, and chaotic event sequences. Finally we apply it on atmospheric rivers in Europe to find typical recurring cycles in Sect. IV.

II. METHOD

For the estimation of a power spectrum we propose a combination of a similarity measure for discrete time series (applied as an auto-correlation measure of this discrete time series) and the Wiener-Khinchin theorem. We use the edit distance metric as a similarity measure. Instead of the standard definition of a time series as ordered pairs of time and value $\{(t_i, x_i)\}$, we consider to have only the time stamps of the discrete or extreme events $\{t_i\}$, finally forming a sequence $\mathcal{S} = \{t_1, t_2, \dots, t_N\}$ of N events.

A. Edit distance

The edit distance quantifies the dissimilarity between two sequences \mathcal{S}_a and \mathcal{S}_b of (discrete) symbols by means of how much it would minimally cost to transform \mathcal{S}_a into \mathcal{S}_b . It was initially introduced in computer science (Levenshtein metric) where the two sequences are usually identified with two strings or words⁴⁷. The operations that are typically allowed include inserting a symbol, deleting a symbol, and shifting two symbols (Levenshtein metric would consider addition, deletion, and replacement). The edit distance offers the advantage that it neither requires sequences of equal lengths nor equally displaced symbols. In practise, measuring similarities with the edit distance is, thus, not limited to strings but it has been applied to marked point processes^{24,39}, extreme events¹⁹, spatial trajectories,⁴⁸ and irregularly sampled time series^{46,49}. Recent modifications to the edit distance include accounting for a saturation of shifting costs given a large temporal displacement between events¹⁹ and a correction method for non-stationary sampling rates⁵⁰. Another recent study already hints at the

potential of the edit distance to reveal scale-dependent variability in the context of recurrence analysis⁵¹.

Here, we consider sequences of binary symbols with no continuous amplitude. These could, e.g., represent spike trains or series of time stamps of the occurrence of extreme events. Hence, we refer to the symbols of a sequence as events in the following, with sequence \mathcal{S} of N events.

The edit distance between two sequences \mathcal{S}_a and \mathcal{S}_b of N_a and N_b events is computed by identifying the transformation path with minimum costs from all possible transformations between \mathcal{S}_a and \mathcal{S}_b :

$$D(\mathcal{S}_a, \mathcal{S}_b) = \min \left\{ \underbrace{\Lambda_S (N_a + N_b - 2|\mathcal{C}|)}_{\text{adding and deleting}} + \underbrace{\sum_{\alpha, \beta \in \mathcal{C}} \Lambda_0 \left\| t_\alpha^{(a)} - t_\beta^{(b)} \right\|}_{\text{shifting}} \right\} \quad (1)$$

with cost parameters Λ_S and Λ_0 ; N_a and N_b the number of events in \mathcal{S}_a and \mathcal{S}_b ; \mathcal{C} the set of events that are shifted, i.e., $\alpha, \beta \in \mathcal{C}$; and $|\mathcal{C}|$ the cardinality of the set \mathcal{C} (i.e., the number of events which are shifted). The cost parameters for deletion/insertion Λ_S and the cost parameter for shifting Λ_0 have to be selected to balance between deletion/insertion and shifting of events. For the sake of simplicity, we can set $\Lambda_S = 1$ and adjust Λ_0 . In previous work we found an appropriate choice for Λ_0 by fixing it to the inverse average time interval between all events⁴⁹

$$\Lambda_0 = \frac{M}{T} \quad (2)$$

with the total number of events $M = N_a + N_b$. As mentioned before, the choice of cost parameters determines whether cost computation is carried out in a balanced way. If the ratio Λ_0/Λ_S is chosen too high, deletion/insertion will always be preferred against shifting (and *vice versa*), yielding inadequate values of the edit distance.

B. Power spectral estimate

Given a finite length, discrete-time (but *continuous*-amplitude) time series $\{(t_i, x_i)\}$, the power spectrum is given by the squared absolute value of the signal's Fourier transform

$$S_x(f) = \frac{1}{N} \left| \sum_{i=0}^{N-1} x_i e^{-j2\pi f i} \right|^2 \quad (3)$$

with the discrete time index i , frequency f , length of the signal N , and the imaginary unit j . Provided the time series is wide-sense stationary and the autocorrelation

$$R_x(\tau) = \sum_{i=0}^{N-1-\tau} x_i x_{i+\tau}^* \quad (4)$$

of the time series is integrable ($\sum_{\tau=-\infty}^{\infty} |R_x(\tau)| < \infty$), the Wiener-Khinchin theorem states that the time series's autocorrelation and power spectral density are directly linked

$$S_x(f) = \sum_{\tau=-\infty}^{\infty} R_x(\tau) e^{-j2\pi f \tau}. \quad (5)$$

In practise, only a limited number of time delays τ is assessable due to the finite length of the time series, granting validity of the integrability assumption.

We use this relationship for a heuristic definition of a novel power spectral estimate for discrete signals, e.g., event series as discussed in Sect. II A. Using the edit distance, Eq. (1), we define the edit distance-based autocorrelation function (EDACF) for event series

$$R_S^{\text{edit}}(\tau) = 1 - \tilde{D}(\mathcal{S}, \mathcal{S}(\tau)) \quad (6)$$

with $\tilde{D}(\mathcal{S}, \mathcal{S}(\tau))$ the normalized edit distance between the event series \mathcal{S} and its shifted copy $\mathcal{S}(\tau)$ shifted by a time delay τ . The normalisation of D limits R_S^{edit} into the range $[0, 1]$ and is performed by dividing D by the maximum transformation cost D_{max} . D_{max} is the larger value of the edit distance between \mathcal{S} and a completely empty sequence and the edit distance between $\mathcal{S}(\tau)$ and a completely empty sequence. This is equivalent to the Hamming distance of the event series and an empty event series, multiplied by the deletion cost parameter Λ_S .

Applying the Wiener-Khinchin theorem to the (z -normalized) EDACF, Eq. (6), we obtain a power spectral estimate for discrete signals (EDSPEC)

$$S_S^{\text{edit}}(f) = \sum_{\tau=-\infty}^{\infty} \frac{R_S(\tau) - \langle R_S(\tau) \rangle}{\text{std}(R_S(\tau))} e^{-j2\pi f \tau} \quad (7)$$

(where $\langle \cdot \rangle$ is the arithmetic mean [but median would also work] and $\text{std}(\cdot)$ is the standard deviation).

A technical note needs to be added to the computation of this power spectral estimate: in the analysis of binary event series, events are usually recorded just by their time instance, time stamp $\{t_i\}$. It is therefore sufficient to add the time delay τ to all time instances in order to shift the event series against itself. This, however, results in non-overlapping segments at the beginning and end of the event series when computing $R_S^{\text{edit}}(\tau)$ for some specific time delay τ . In order to not include the transformation costs that solely result from the transformation of the non-overlapping events in the numerical computation, we prune the shifted event series to their largest intersection. The selection of the considered maximum time delay will change the lowest assessable frequency resolved by the power spectral estimate. While a large maximum value of τ allows for studying longer cycles, the length of the overlap shrinks and introduces undesired effects such as spectral leakage. We thus recommend to carefully select this parameter based on the number of observed events and research question at hand.

III. PARADIGMATIC EXAMPLES

We illustrate the proposed EDSPEC by power spectral analysis of selected paradigmatic example systems (Fig. 1). In particular, we consider sequences with events of random (Fig. 1A), periodic (Fig. 1B), stochastic with serial dependence (Fig. 1C), and chaotic (Fig. 1D) occurrences. For each example, we generate $N = 1,000$ events over a time span of $T = 10,000$ arbitrary time units (a.u.). For the examples studied here, the considered maximum time delay τ is chosen as $\tau = T/10$. For evaluating the significance of the potentially detected periodicities, we perform a surrogate based statistical test. We generate $n = 100$ surrogate event series by drawing N time instances between 0 and T with equal probability. The corresponding null-hypothesis is the absence of any non-stochastic/periodic pattern that controls the sequence of events. The 95%-quantile of spectral power for each frequency from the surrogate ensemble is then used as the respective 95%-confidence level.

A. Random sequence

First, we generate a random sequence of events by drawing exponentially distributed waiting times between consecutive events (Fig. 1A). The resulting process is a Poisson process and is, e.g., used as a Markovian model for arrival of costumers in a store⁵². The EDACF rapidly declines to a value close to zero, reflecting vanishing serial dependence as it is expected for a Poisson process (Fig. 1A). The EDSPEC follows a power law over two orders of magnitude. No significant frequencies exceed the 95%-confidence interval for the studied Poisson process, confirming the absence of any periodicities in this event series.

B. Periodic sequence

The opposite is expected for a periodic event series. We generate an event series of 1,000 equally spaced events with a frequency of $\omega = 1/10$. The resulting EDACF is strictly periodic (Fig. 1B). Consequently, the EDSPEC gives the expected result of a sharp peak at a frequency of $f = \omega$, exceeding the 95%-confidence interval. The observed spectral leakage is similar to what is found with standard power spectral density when it is computed for a monochromatic sinusoidal with a rectangular window function. Aliases of the original frequency can be identified at (positive) integer multiples of ω . Spectral power below ω is constant.

C. Stochastic sequence with serial dependence

We next study a stochastic sequence of events with non-zero serial dependence, based on an autoregressive

process of first order (AR(1)-process, Fig. 1C). First, a sample of a regular AR(1)-process is generated:

$$x_i = ax_{i-1} + \xi_i, x_0 = 0 \quad (8)$$

with standard-normally distributed random numbers ξ_i and an autoregressive coefficient of $a = 0.9$. Next, we shift all x_i -values to the positive domain ($x_i \rightarrow \tilde{x}_i : x_i - \min\{x_i\} > 0$) to finally use them as serially correlated inter-event times:

$$t_i = t_{i-1} + kx_{i-1}, t_0 = \tilde{x}_1 \quad (9)$$

The constant k is chosen such that, like for the other studied systems, we obtain $N = 1,000$ events over a time span of $T = 10,000$ a.u. Such a process could model a scenario where events occur in a clustered or cascade-like fashion.

For the first 100 delays, the EDACF declines to a value slightly below 0.5. On top of this decline, we observe small oscillations. The corresponding EDSPEC exhibits higher spectral power than expected from a random null-model for the lowest frequencies and for a band of higher frequencies around $f = 1/10$. In the scaling region up to $f = 1/20$, scaling is more steep than for the random event series. The serial dependence in the serially correlated event series does consequently not result in sharp significant frequencies (as expected), but gives rise to significantly high spectral power in a broader, but limited band around the average time interval between consecutive events. For event series with a unimodal distribution of inter-event times, this average sampling interval could effectively be understood as a cycle. This is due to the fact that inter-events times center around this preferential interval, in contrast to the exponentially distributed inter-event times in Fig. 1A.

D. Chaotic sequence

Finally, we generate an event series from a chaotic time series. To this extent, we compute the x -component of the Lorenz system⁵³

$$\begin{aligned} \frac{dx}{dt} &= \sigma(y - x) \\ \frac{dy}{dt} &= x(\rho - z) - y \\ \frac{dz}{dt} &= xy - \beta z \end{aligned} \quad (10)$$

with $\sigma = 10$, $\rho = 28$, $\beta = 8/3$ (chaotic dynamics). We fix an integration time of $\Delta t = 0.1$ and generate 10,000 values (after discarding transients) with initial conditions $x(t = 0) = 0.0$, $y(t = 0) = 1.0$, $z(t = 0) = 1.05$. The x -component is transformed into an event series by identifying each local maximum with an event (Fig. 1D). We only extract 1,000 local maxima and rescale the respective time axis of these events such that it covers a time

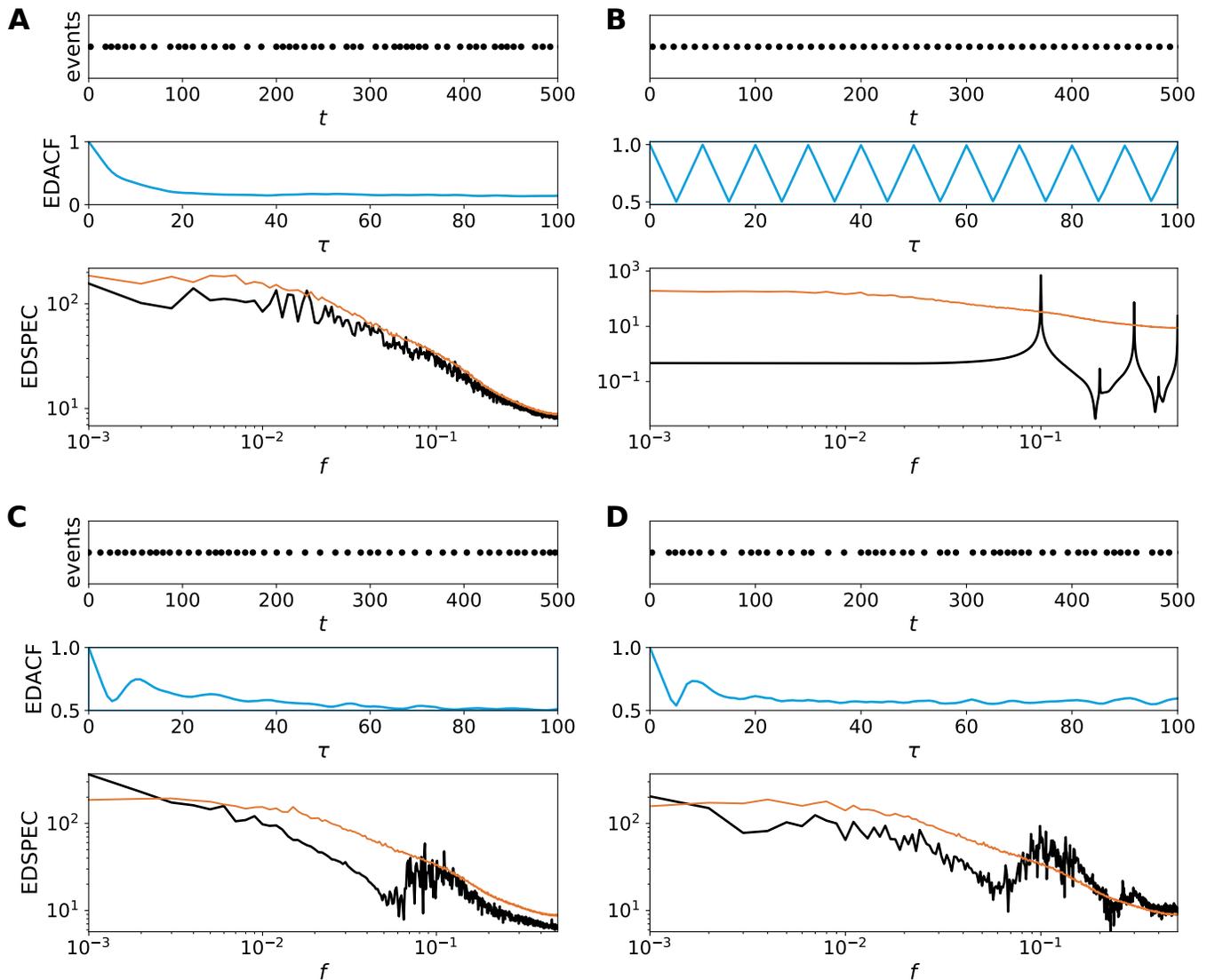


FIG. 1. Paradigmatic examples of event series: (A) random, (B) periodic, (C) autoregressive, and (D) chaotic events. For each example, three panels show an excerpt of the event series (up to $t = 500$ a.u., black dots), an excerpt of the EDACF (blue) and the EDSPEC (black) (from top to bottom). The respective 95%-confidence levels are shown in orange.

span of $T = 10,000$ such that it aligns with the other examples. As expected, the EDACF exhibits non-zero serial correlation which is even slightly higher than for the autoregressive event series. We find a broad band of significant frequencies above $f = 1/10$, likely reflecting the presence of many unstable periodic orbits. In the scaling region up to $f = 1/20$, the slope of the EDSPEC is less steep than for the random event series.

E. Influence of noise and number of events

In the analysis of real-world event series, it is common to find considerable uncertainty in the timing of events. This uncertainty can, for example, be represented by jit-

ter around the true time instance of an event. The impact of this effect on the estimation of the proposed edit distance-based power spectrum can be investigated by superimposing observational noise onto an event series. As the proposed method is predominantly designed to identify periodicities in event series, we generate a periodic event series with two distinct frequencies $\omega_1 = \frac{1}{20}$ and $\omega_2 = \frac{1}{12.7}$ (i.e., frequencies are not integer multiples of each other). Another challenge in the study of observational event series is that only a low number of events can be recorded, e.g., due to constraints in the measurement procedures. In order to test the sensitivity of our method against the presence of observational noise and shortness of event series, we generate two samples of this periodic event series, one with 130 and the other with

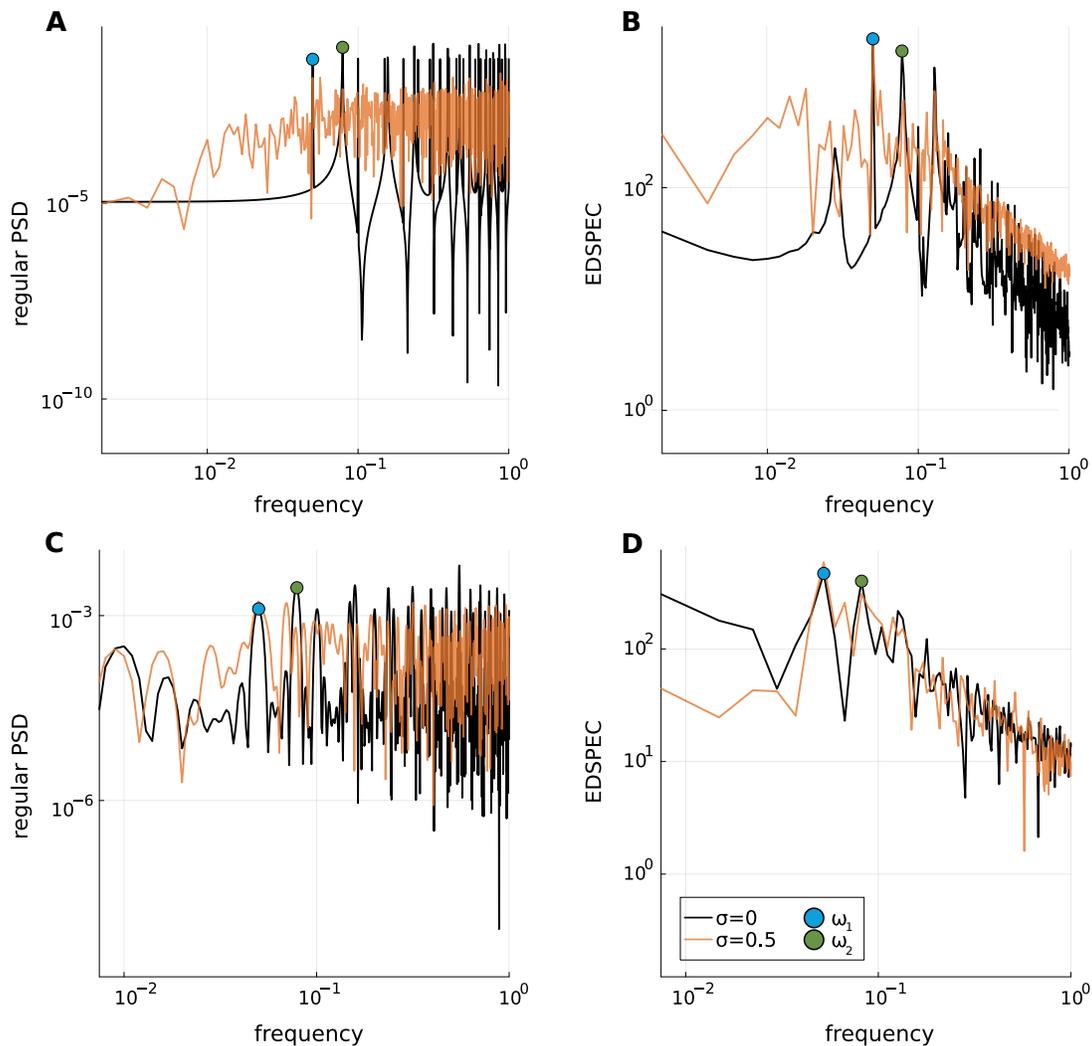


FIG. 2. (A, C) Regular PSD and (B, D) EDSPEC for a periodic event series with two frequencies ω_1 and ω_2 (cyan and purple markers). Black lines show the noise-free case, orange lines show a single realization with superimposed noise of relatively high strength ($\sigma = 0.5$). Peaks can be more clearly distinguished for the longer event series ($N = 130$, A/B) than for the very short event series ($N = 20$, C/D).

only 20 events between 0 and $T = 1,000$ a.u. Due to the short event series lengths, we set a maximum time delay of $\tau = T/2$ in order to still assess lower frequencies. To establish a comparison with a typical alternative method, we also compute the regular power spectral density (PSD) of the event series using the Fourier periodogram of the binary event series (Fig. 2). Overall, the proposed EDSPEC generates more distinct peaks in presence of noise and deteriorates less rapidly when the number of events is reduced.

In absence of noise, both the regular PSD and the EDSPEC effectively identify both frequencies as distinct peaks (Fig. 2A/B, black curves). Additional peaks can be found at integer multiples of each frequency as well as at integer multiples of the sum and difference of them. If we crop the event series such that there are only $N = 20$

events, both peaks are still visible but become harder to delineate (Fig. 2C/D, black curves).

Next, we superimpose normally distributed noise $\mathcal{N}(0, \sigma\Delta)$ onto the event series with noise strength σ the average time interval Δ between consecutive events. It appears that the EDSPEC gives visually more convincing peaks at the expected frequencies for relatively high noise strength $\sigma = 0.5$ (Fig. 2, orange curves). To systematically study the noise sensitivity of our method, we increase $\sigma \in [0, 1]$ and obtain the regular PSD and EDSPEC for 500 distinct noise realizations that are superimposed onto the periodic event series. As a quantitative measure of peak quality, we compute the prominence of each of the two peaks by computing the vertical distance between the peak height and the height of its baseline. The height of each baseline is estimated by identifying

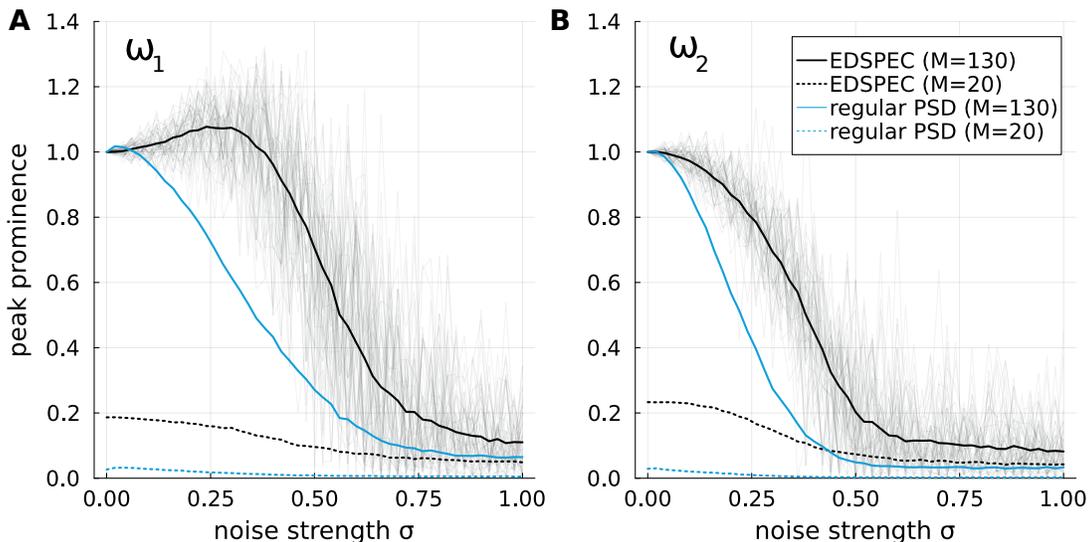


FIG. 3. Sensitivity of the EDSPEC (black) against noise and shortness of event series, as compared to the regular PSD (blue). A single periodic event series with two frequencies ω_1 (A) and ω_2 (B) is generated and superimposed with noise of varying strength σ . For each computed regular PSD and EDSPEC, prominence of peaks is estimated. Averages over all realizations are shown as thick lines, EDSPECs for single realizations are shown by shaded gray lines. Solid (dotted) lines show the results of this analysis for $N = 130$ ($N = 20$).

the closest (local) minimum value of spectral power in proximity of the peak. Consequently, we obtain two values of peak prominence for each noise strength (ω_1 and ω_2). We normalize both peak prominence curves separately for the two methods to their respective value in the noise-free case ($\sigma = 0$). We do the same for the shorter event series.

The identification of both frequencies ω_1, ω_2 deteriorates considerably more slowly for the proposed EDSPEC than for the standard periodogram (Fig. 3). The prominence of the peak that is associated with the lower frequency ω_1 is on average more easy to detect than the one for ω_2 . Interestingly, small amounts of noise (up to 30%) seem to enhance prominence of the peak associated with ω_1 when we use the EDSPEC method. The validity of this finding will, however, most likely depend on the selected measure of peak prominence. In the study of very short event series ($N = 20$, black dotted lines), peak prominence for detecting both frequencies is reduced by about 1/5 for the EDSPEC method, i.e., roughly proportional to the reduction in length as compared to $N = 130$ (black solid line). While this reduction is certainly significant and also visually hampers detection of peaks especially in the presence of noise (Fig. 2D), the traditional PSD reveals considerably stronger deterioration in peak prominence for very short event series (Fig. 3, blue dotted line). Finally, distinct noise realizations can allow considerably better or worse peak detection, as indicated by the shown peak prominence curves for single realizations (gray lines).

Overall, we find that the proposed EDSPEC is less sensitive to timing jitter/observational noise than the regu-

lar PSD when applied to a simple periodic event series with two frequencies. The prominence of peaks deteriorates less strongly when the number of events is reduced. This should be of high relevance for sparse and noisy event series obtained from observational data. The robustness of other important properties beyond peak detection, such as spectral scaling, should be investigated in future studies.

IV. APPLICATION TO EUROPEAN ATMOSPHERIC RIVERS

Atmospheric rivers (ARs) are narrow filaments of extensive water vapor transport in the lower troposphere. They can transport moisture for thousands of kilometers, predominately from the tropics towards the mid-latitudes¹⁵. Their life times can range between a few hours and several days. Contrasting common beliefs, ARs represent the largest “rivers” on Earth as they on average transport more than twice the amount of freshwater of the Amazon River⁵⁴. While some recent studies hint at alternative possible origins of ARs, most ARs are associated with the front of an extratropical cyclone⁵⁵. Low-intensity ARs provide vital amounts of freshwater and rarely result in heavy precipitation events. On the other hand, high-level ARs can cause detrimental impacts, including floods and high economic damage, when they land-fall or ascent along an orographic barrier.

In Europe, links between ARs and heavy precipitation events have recently started to attract attention^{56,57} (similar to research of ARs at the American west coast⁵⁸).

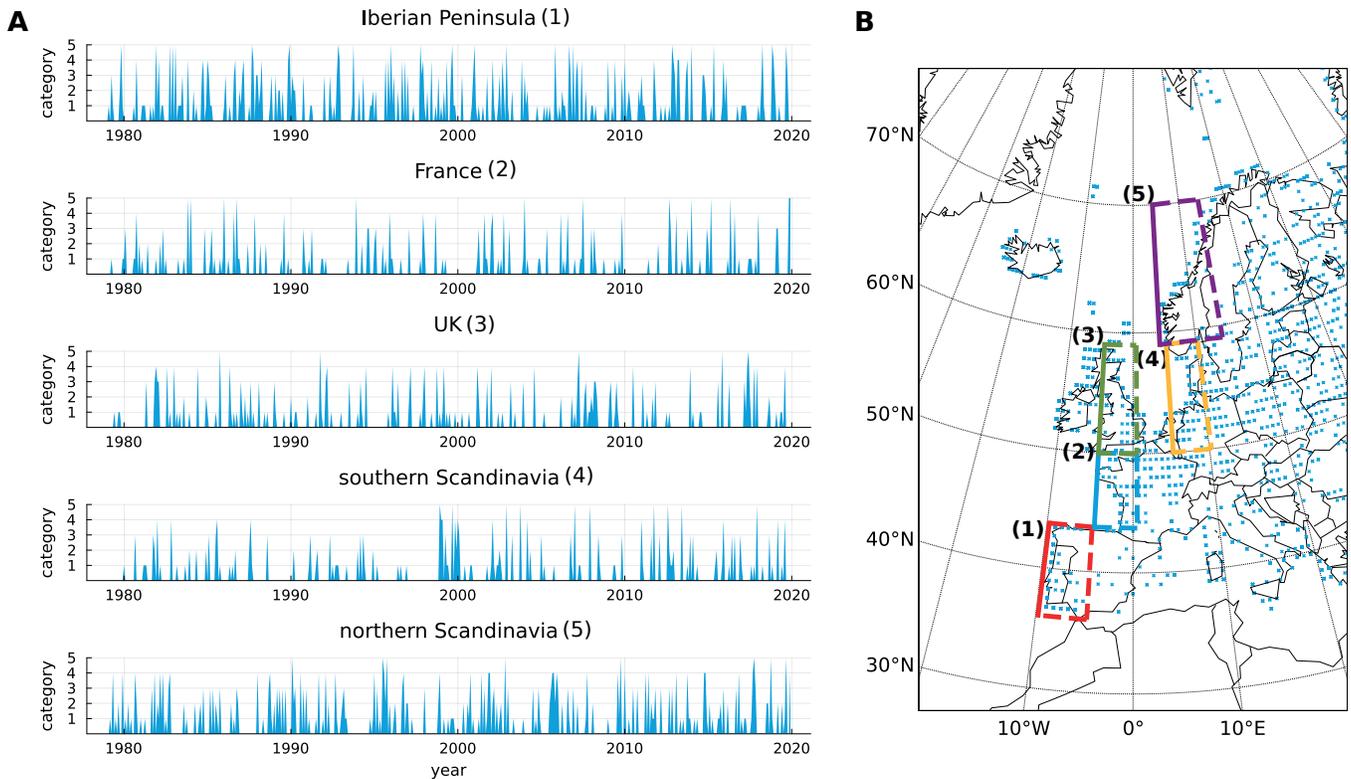


FIG. 4. (A) Monthly-resolved time series of AR categories in five distinct European domains between 1979-2019. (B) Map of Europe, showing the five spatial domains: (1) Iberian Peninsula, (2) France, (3) the UK, (4) southern Scandinavia and the Netherlands, (5) northern Scandinavia. Blue markers indicate landfalling ARs in the studied AR catalogue (from which we focus on the ones that fall into the marked domains only). Some ARs seem to landfall over the ocean which is due to smaller islands that are not indicated here.

The frequency of ARs and related heavy precipitation events in Europe is projected to increase due to enhanced atmospheric moisture caused by recent climate change, whereas considerable uncertainties exist⁵⁹. Impact by high-category ARs is regionally heterogeneous over Europe and likely linked to regional presence of orographic barriers. Fall and winter have been identified as the seasons where most strong ARs occur, indicating the presence of a seasonal cycle in AR occurrence⁵⁶. Furthermore, AR frequency in northern and southern Europe has been found to correlate with the phase of the North Atlantic Oscillation (NAO)⁵⁶.

Here, we study periodic cycles in the occurrence of European ARs between 1979 and 2019 by means of the proposed power spectral measure. To this extent, we track European ARs and regard the final (continental) position along their track as an event, as the termination of an AR potentially triggers rainfall (“landfalling” AR). Detection of ARs is based on localizing anomalous atmospheric transport of moisture. Many approaches track ARs by defining a threshold that identifies local anomalies in integrated vapor transport (IVT). Such approaches effectively presume stationary atmospheric moisture levels and often exclude low-level ARs. Here, we instead

employ an AR-detection framework (*ARtracks*⁶⁰) based on global ERA5 reanalysis data that utilizes image processing techniques (using the IPART algorithm⁶¹). We implement the most widely used scale that characterizes European ARs based on their strength and persistence⁵⁷ to distinguish five categories of ARs. While ARs of categories 1 and 2 are predominately beneficial, ARs of categories 4 and 5 often entail hazardous heavy precipitation events; category-3 ARs rank in between.

First, we generate the European AR catalogue using the *ARtracks* algorithm. The used parameters can be found in the documentation of *ARtracks* and align with standard calibration of the IPART algorithm⁶¹. As an intermediate result, we obtain a catalogue of (uncategorized) ARs along with detailed AR properties, including unique AR identifiers and AR strength. To generate the AR catalogue, ERA5 reanalysis data set is re-gridded to a temporal resolution of six hours and a spatial resolution of $0.75^\circ \times 0.75^\circ$. Next, we apply a set of additional selection criteria to this AR catalogue and categorize every landfalling AR. A more detailed description can be found in App.A. Based on the resulting filtered and categorized AR catalogue, 65.5% of days between 1979 and 2019 have at least one active AR in the covered spa-

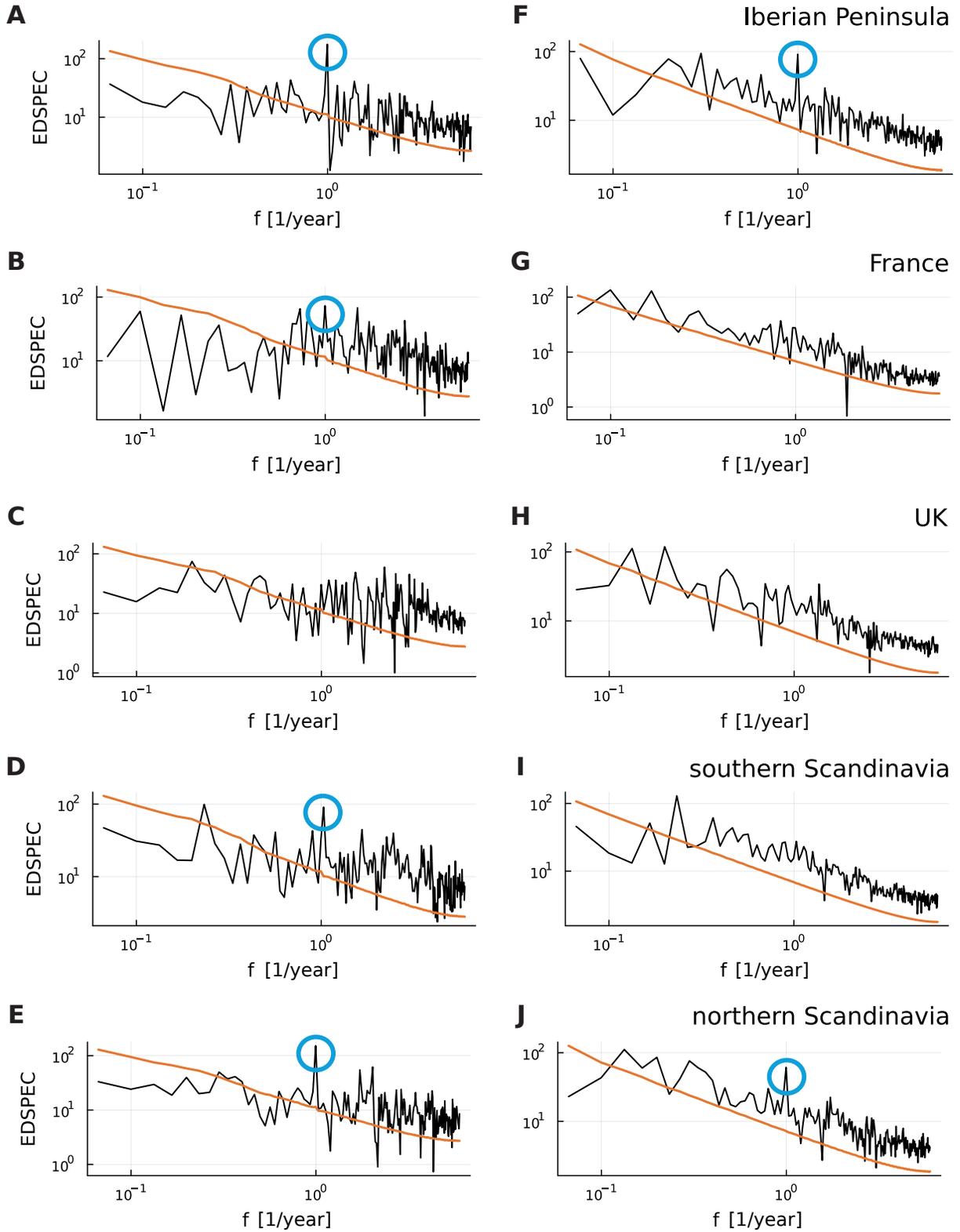


FIG. 5. Spectral power estimates for European ARs in the five spatial domains (top: south, bottom: north), computed with the proposed EDSPEC method (black). (A–E) EDSPEC for low-category (≤ 2) ARs and (F–J) EDSPEC for (moderate to) high-category (≥ 3) ARs. 99.99% confidence levels based on 1000 random surrogates for each event series are shown in orange. Pronounced seasonal cycles that can clearly be identified visually are marked by blue circles.

tial region (Fig. A1 and Fig. 4B). We identify various hot spots of landfalling ARs along the European west coast, such as the Iberian Peninsula, the UK, and Scandinavia (Fig. A1). In the Atlantic, Greenland as well as some islands attract landfalling ARs. On the other hand, we also identify landfalling ARs along the Mediterranean coast of Italy. For 7.1% of all days, there is a landfalling AR which we identify with an AR event. For by far the most ARs that are studied here, the final landfall location is mostly covering the European continent (Fig. A2, gray bars). Only few AR tracks terminate predominately over Africa, Asia, or North America (but still exhibit a European contribution). Most European ARs are potentially beneficial as they are assigned to category 1 or 2 (Fig. A2, colored bars).

We generate event series from the filtered and categorized AR catalogue by both temporally and spatially integrating ARs. ARs are integrated onto a monthly time axis by selecting the highest category AR for each month. For the spatial integration, we largely follow the spatial domains oriented along three reference meridians as defined in⁶²: we distinguish between the (1) Iberian Peninsula, (2) west coast of France, (3) the UK, (4) southern Scandinavia and the Netherlands, and (5) northern Scandinavia (Fig. 4B). We only extend the latter domain (northern Scandinavia, longitudes: 4.25° to 14.25°) to include more AR events for sufficient spectral estimates. Each monthly and domain-specific time series of AR categories can be regarded as a symbolic sequence with integer values (symbols) between 1 and 5 (Fig. 4A). Most (integrated) landfalling AR events are identified along the Iberian Peninsula and northern Scandinavia (Fig. A3). Especially the studied domains of France, UK, and southern Scandinavia exhibit an overproportionally high number of category-5 ARs whereas northern Scandinavia exhibits the most regular distribution of AR categories.

For the spectral analysis, we only distinguish between low-category (category < 3) and (moderate to) high-category (category ≥ 3) ARs. For each of these two groups, we obtain a single event series, indicating the occurrence of a low- or high-category AR at times $\{t_i\}$. We chose the considered maximum time delay as $\tau = 30$ years to include decadal frequencies of AR occurrence.

We find evidence for a clear seasonal cycle in the occurrence of low-category ARs in all domains except the UK, corroborating findings from other studies on European ARs^{56,57} (Fig. 5A–E). This showcases the ability of the proposed method to identify known cycles in complex observational discrete data. While we do not analyze the phases of the identified cycles, it is likely that they align with the known enhanced frequency of AR occurrence in boreal autumn/winter. In the UK domain, the (significant) seasonal cycle is still noticeable, but it is suppressed compared to the other domains. This suggests a reduced seasonal predictability of ARs along parts of the British coast. Tested against random event series (99.99% confidence level from 1,000 surrogates), broad bands of frequencies stand out as significant. This hints

at a certain degree of predictability of ARs based on non-random cyclicities that could be exploited in forecasting approaches.

Even for high-category ARs, a broad range of significant cycles exist (Fig. 5F–J). High-category ARs only exhibit a pronounced seasonal cycle along the Iberian peninsula and northern Scandinavia. Considering both low- and high-category ARs, seasonality is expressed least between 0 and 5° W longitude (France and UK). For high-category ARs, spectral power at lower frequencies is enhanced as compared to low-category ARs. These multi-annual/decadal cycles could be linked to the NAO: A negative (positive) NAO is likely linked to enhanced AR frequency in southern (northern) Europe⁵⁶.

V. CONCLUSION

In many disciplines, research focuses on the identification of cycles in periodic discrete data (e.g., symbolic time series). In particularly challenging cases, only a sequence of (non-equidistant) times is recorded. Standard techniques, such as the Fourier PSD are not designed to quantify the spectral content of such signals. Noise contamination and limited signal length are additional obstacles in the spectral analysis of such systems.

In this work, we proposed a novel power spectral estimate for discrete data. It is based on the edit distance, a powerful distance measure that quantifies the similarity of two non-equidistant event sequences of unequal lengths by means of their transformation costs. The fundamental idea proposed here is to, firstly, normalize the edit distance as an auto-correlation measure of a discrete time series and, secondly, transform this measure into a power spectral estimate using the Fourier transform (analogously to the Wiener-Khinchin theorem). The proposed spectrum is referred to as EDSPEC. It enables us to identify cycles in non-equidistant discrete ordered pairs of time and values $\{(t_i, x_i)\}$ as well as for the more challenging case when only time stamps of the discrete events $\{t_i\}$ can be assessed. Here we focus on the latter case.

We applied the proposed EDSPEC to a range of discrete paradigmatic signals, both of stochastic and deterministic origin. A simple hypothesis test based on random event surrogates is provided to test for the presence of significant cycles. The EDSPEC proves effective in identifying known cycles and broad bands of significant frequencies, also in presence of relatively high noise intensity (or timing jitter). Even for a very low number of events, the EDSPEC performs comparably well.

In a challenging real-world application, we used the EDSPEC to investigate potential cycles in a novel catalogue of European ARs. With the increasing moisture capacity of a warming atmosphere, ARs are expected to become a more and more vital source of atmospheric water supply, including heavy rainfall events. While seasonality in European ARs is undisputed and has been

inferred from other AR catalogues (e.g., based on composites), our study confirms the presence of a seasonal cycle in AR occurrence based on our more suitable spectral analysis framework. This warrants further credibility to (spatially heterogeneous) seasonality in the landfall of European ARs. Furthermore, the proposed framework allows to separately study the presence of seasonality as well as multi-annual cycles for different AR categories. We found that hazardous high-category ARs show less seasonal regularity but offer predictability based on a broad range of significant cycles, potentially linked to the phase of the NAO.

In future studies, additional features of the proposed EDSPEC should be studied, e.g., spectral scaling and leakage effects. A multitude of complex discrete signals from real-world systems calls for application of the proposed measure, potentially revealing unknown periodicities.

ACKNOWLEDGMENTS

This study was supported by BMBF project climX-treme (No. 01LP1902J) “Spatial synchronization patterns of heavy precipitation events”. We thank Sara M. Vallejo-Bernal for fruitful discussions.

DATA AND CODE AVAILABILITY

Code used for the creation of event series and EDSPEC calculation of the paradigmatic examples, as well as code for the EDSPEC calculation of the ARs is available at Zenodo [10.5281/zenodo.7555049](https://zenodo.org/doi/10.5281/zenodo.7555049). AR events are extracted from the ARtracks Atmospheric River Catalogue available at Zenodo <https://doi.org/10.5281/zenodo.7018725>.

Appendix A: Spectral analysis of AR catalogue

We filter the AR tracks as returned by *ARtracks* with respect to several criteria in order to ensure the obtained European ARs are physically meaningful with respect to our research objective. Firstly, the information on which proportion of the AR is located over Europe is used to ensure that only ARs are considered that are located over Europe for most track instances (Fig. A1A). This still preserves ARs that overall have mixed continental contributions, e.g., over Africa. Next, an AR is fully discarded if (a) it exhibits IVT less than 250 kg/ms for any instance along its track, (b) does not persist longer than 6 h, (c) terminates over the ocean, (d) is solely continental. We further only consider AR tracks that can be identified with a continuous track, i.e., no time 6-hourly instance is missing. Track instances where an AR is fully located over a different continent than Europe are discarded (while not discarding the full track). Finally, we

categorize ARs adhering to the scale proposed in⁵⁷. If an AR exceeds the scale, either in terms of persistence or strength, we assign the most suitable category at the respective edge of the scale (e.g., an AR with maximum IVT magnitude between 750 and 1000 kg/ms and persistence of more than 72 h is still ranked as category 4). Categories are computed exclusively from all instances of a track where the AR landfalls, i.e., some fraction of its area is located over Europe. For the strength-based classification of ARs, we choose an AR’s category based on which category would be correct for its strength along most instances of its track. The resulting spatial distribution of ARs (based on the final instances of their centroids) is strongly oriented along the European west coast (Fig. A1). There are various hot spots of landfalling ARs along the European west coast, such as the Iberian Peninsula, the UK, and Scandinavia. In the Atlantic, Greenland as well as some islands attract landfalling ARs. On the other hand, we also identify landfalling ARs along the Mediterranean coast of Italy. For by far the most

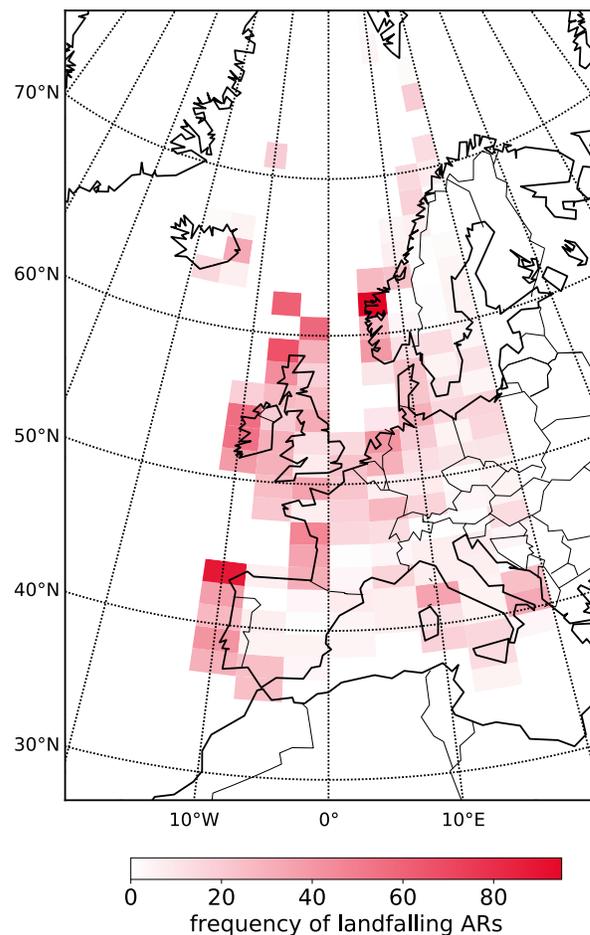


FIG. A1. Spatial distribution of European landfalling ARs. Frequencies are estimated using a spatial histogram of AR landfall coordinates with 60 longitudinal and 40 latitudinal bins.

ARs that are studied here, the final landfall location is mostly covering the European continent (Fig. A2, gray bars). Only in few AR tracks terminate predominately over Africa, Asia, or North America (but still exhibit a European contribution). Most European ARs are potentially beneficial as they are assigned to category 1 or 2 (Fig. A2, colored bars).

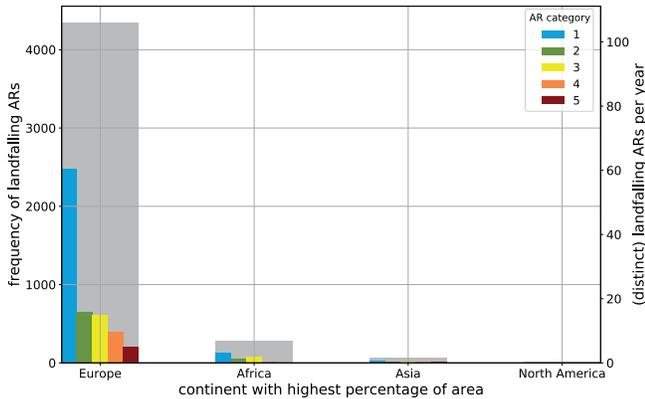


FIG. A2. Continental contributions of European landfalling ARs (gray bars) and respective distribution of AR categories (colored bars). The displayed frequencies refer to the categorized *ARtracks* catalogue of ARs without spatial or temporal integration. The second y -axis gives average numbers of distinct landfalling ARs per year (not to be mistaken with the yearly time at least one AR is active).

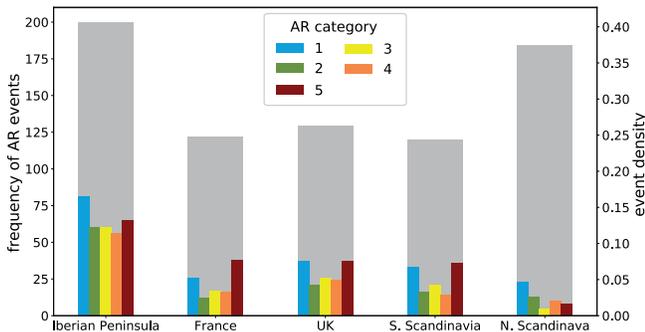


FIG. A3. Distribution of AR events in the different five spatial domains. Gray bars indicate the total number of AR events, regardless of their category, within each domain. The second y -axis shows the density of these events, i.e. fraction of months that exhibit an AR event. Frequencies of different AR categories are shown by colored bars.

A more detailed assessment of the robustness of the obtained spatial distribution and statistics of AR properties beyond the findings mentioned here is investigated in a study that is currently prepared.

REFERENCES

- 1S. Makeig, S. Debener, J. Onton, and A. Delorme, “Mining event-related brain dynamics,” *Trends in Cognitive Sciences* **8**, 204–210 (2004).
- 2S. Cole, W. J. Percival, J. A. Peacock, P. Norberg, C. M. Baugh, C. S. Frenk, I. Baldry, J. Bland-Hawthorn, T. Bridges, R. Cannon, M. Colless, C. Collins, W. Couch, N. J. G. Cross, G. Dalton, V. R. Eke, R. De Propriis, S. P. Driver, G. Efstathiou, R. S. Ellis, K. Glazebrook, C. Jackson, A. Jenkins, O. Lahav, I. Lewis, S. Lumsden, S. Maddox, D. Madgwick, B. A. Peterson, W. Sutherland, and K. Taylor, “The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications,” *Monthly Notices of the Royal Astronomical Society* **362**, 505–534 (2005).
- 3R. B. Randall and J. Antoni, “Rolling element bearing diagnostics—A tutorial,” *Mechanical Systems and Signal Processing* **25**, 485–520 (2011).
- 4M. H. Trauth, “Spectral analysis in Quaternary sciences,” *Quaternary Science Reviews* **270**, 107157 (2021).
- 5M. K. Das, B. Khan, S. Jacob, A. Kumar, and J. Mahenthiran, “Significance of a Fragmented QRS Complex Versus a Q Wave in Patients With Coronary Artery Disease,” *Circulation* **113**, 2495–2501 (2006).
- 6N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, and J. Kurths, “Recurrence Plot Based Measures of Complexity and its Application to Heart Rate Variability Data,” *Physical Review E* **66**, 026702 (2002).
- 7N. Wessel, C. Ziehmann, J. Kurths, U. Meyerfeldt, A. Schirdewan, and A. Voss, “Short-term forecasting of life-threatening cardiac arrhythmias based on symbolic dynamics and finite-time growth rates,” *Physical Review E* **61**, 733–739 (2000).
- 8U. Parlitz, S. Berg, S. Luther, A. Schirdewan, J. Kurths, and N. Wessel, “Classifying cardiac biosignals using ordinal pattern statistics and symbolic dynamics,” *Computers in Biology and Medicine* **42**, 319–327 (2012).
- 9K. D. Harris, D. A. Henze, H. Hirase, X. Leinekugel, G. Dragoi, A. Czurkó, and G. Buzsáki, “Spike train dynamics predicts theta-related phase precession in hippocampal pyramidal cells,” *Nature* **417**, 738–741 (2002).
- 10M. Steriade, “Corticothalamic resonance, states of vigilance and mentation,” *Neuroscience* **101**, 243–276 (2000).
- 11E. M. Izhikevich, “Synchronization of elliptic bursters,” *SIAM review* **43**, 315–344 (2001).
- 12M. Staniek and K. Lehnertz, “Symbolic Transfer Entropy,” *Physical Review Letters* **100**, 158101 (2008).
- 13A. F. Prein, R. M. Rasmussen, K. Ikeda, C. Liu, M. P. Clark, and G. J. Holland, “The future intensification of hourly precipitation extremes,” *Nature Climate Change* **7**, 48–52 (2017).
- 14M. Kemter, B. Merz, N. Marwan, S. Vorogushyn, and G. Blöschl, “Joint Trends in Flood Magnitudes and Spatial Extents Across Europe,” *Geophysical Research Letters* **47**, e2020GL087464 (2020).
- 15A. Gershunov, T. Shulgina, R. E. S. Clemesha, K. Guirguis, D. W. Pierce, M. D. Dettinger, D. A. Lavers, D. R. Cayan, S. D. Polade, J. Kalansky, and F. M. Ralph, “Precipitation regime change in Western North America: The role of Atmospheric Rivers,” *Scientific Reports* **9**, 9944 (2019).
- 16C. F. Baggett, E. A. Barnes, E. D. Maloney, and B. D. Mundhenk, “Advancing atmospheric river forecasts into subseasonal-to-seasonal time scales,” *Geophysical Research Letters* **44**, 7528–7536 (2017).
- 17B. D. Mundhenk, E. A. Barnes, E. D. Maloney, and C. F. Baggett, “Skillful empirical subseasonal prediction of landfalling atmospheric river activity using the Madden–Julian oscillation and quasi-biennial oscillation,” *npj Climate and Atmospheric Science* **1**, 19–21 (2018).
- 18N. Boers, B. Bookhagen, N. Marwan, and J. Kurths, “Spatiotemporal characteristics and synchronization of extreme rainfall in

- South America with focus on the Andes Mountain range,” *Climate Dynamics* **46**, 601–617 (2016).
- ¹⁹A. Banerjee, B. Goswami, Y. Hirata, D. Eroglu, B. Merz, J. Kurths, and N. Marwan, “Recurrence analysis of extreme event-like data,” *Nonlinear Processes in Geophysics* **28**, 213–229 (2021).
- ²⁰T. Brijs, D. Karlis, and G. Wets, “Studying the effect of weather conditions on daily crash counts using a discrete time-series model,” *Accident Analysis & Prevention* **40**, 1180–1190 (2008).
- ²¹Z. Eisler, J.-P. Bouchaud, and J. Kockelkoren, “The price impact of order book events: market orders, limit orders and cancellations,” *Quantitative Finance* **12**, 1395–1419 (2012).
- ²²T. Braun, J. A. Fiegen, D. C. Wagner, S. M. Krause, and T. Guhr, “Impact and recovery process of mini flash crashes: An empirical study,” *Plos one* **13**, e0196920 (2018).
- ²³S. Wang, R. Schäfer, and T. Guhr, “Cross-response in correlated financial markets: individual stocks,” *The European Physical Journal B* **89**, 1–16 (2016).
- ²⁴S. Suzuki, Y. Hirata, and K. Aihara, “Definition of distance for marked point process data and its application to recurrence plot-based analysis of exchange tick data of foreign currencies,” *International Journal of Bifurcation and Chaos* **20**, 3699–3708 (2010).
- ²⁵J. Barunik and T. Křehlík, “Measuring the Frequency Dynamics of Financial Connectedness and Systemic Risk,” *Journal of Financial Econometrics* **16**, 271–296 (2018).
- ²⁶D. Preoțiuc-Pietro and T. Cohn, “A temporal model of text periodicities using Gaussian Processes,” in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (2013) pp. 977–988.
- ²⁷E. D. Feigelson, R. S. De Souza, E. E. Ishida, and G. J. Babu, “Twenty-first-century statistical and computational challenges in astrophysics,” *Annual Review of Statistics and Its Application* **8**, 493–517 (2021).
- ²⁸C. Koen, “Identifying multiple periodicities in sparse photon event time series,” *Monthly Notices of the Royal Astronomical Society* **459**, 3012–3024 (2016).
- ²⁹P. A. Morettin, “Walsh Spectral Analysis,” *SIAM Review* **23**, 279–291 (1981).
- ³⁰I. Dinstein and T. Silberberg, “Average Walsh Power Spectrum for Periodic Signals,” *IEEE Transactions on Electromagnetic Compatibility* **EMC-23**, 407–412 (1981).
- ³¹W. D. Smith, “Walsh Versus Fourier Estimators of the EEG Power Spectrum,” *IEEE Transactions on Biomedical Engineering* **BME-28**, 790–793 (1981).
- ³²C. Torrence and G. P. Compo, “A Practical Guide to Wavelet Analysis,” *Bulletin of the American Meteorological Society* **79**, 61–78 (1998).
- ³³R. S. Stanković and B. J. Falkowski, “The Haar wavelet transform: its status and achievements,” *Computers & Electrical Engineering* **29**, 25–44 (2003).
- ³⁴C. Capilla, “Application of the Haar wavelet transform to detect microseismic signal arrivals,” *Journal of Applied Geophysics* **59**, 36–46 (2006).
- ³⁵X. Xiang, J. Zhou, X. An, B. Peng, and J. Yang, “Fault diagnosis based on Walsh transform and support vector machine,” *Mechanical Systems and Signal Processing* **22**, 1685–1693 (2008).
- ³⁶R. Hébert, K. Rehfeld, and T. Laepple, “Comparing estimation techniques for temporal scaling in palaeoclimate time series,” *Nonlinear Processes in Geophysics* **28**, 311–328 (2021).
- ³⁷I. V. L. Clarkson, “Approximate maximum-likelihood period estimation from sparse, noisy timing data,” *IEEE transactions on signal processing* **56**, 1779–1787 (2008).
- ³⁸T. Sauer, “Interspike interval embedding of chaotic signals,” *Chaos* **5**, 127–132 (1995).
- ³⁹J. D. Victor and K. P. Purpura, “Metric-space analysis of spike trains: theory, algorithms and application,” *Network: Computation in Neural Systems* **8**, 127–164 (1997).
- ⁴⁰R. Quian Quiroga, T. Kreuz, and P. Grassberger, “Event synchronization: A simple and fast method to measure synchronicity and time delay patterns,” *Physical Review E* **66**, 41904 (2002).
- ⁴¹M. Ciba, R. Bestel, C. Nick, G. F. de Arruda, T. Peron, C. C. Henrique, L. d. F. Costa, F. A. Rodrigues, and C. Thielemann, “Comparison of Different Spike Train Synchrony Measures Regarding Their Robustness to Erroneous Data From Bicuculline-Induced Epileptiform Activity,” *Neural Computation* **32**, 887–911 (2020).
- ⁴²J. Donges, C.-F. Schleussner, J. Siegmund, and R. Donner, “Event coincidence analysis for quantifying statistical interrelationships between event time series,” *The European Physical Journal Special Topics* **225**, 471–487 (2016).
- ⁴³N. Malik, B. Bookhagen, N. Marwan, and J. Kurths, “Analysis of spatial and temporal extreme monsoonal rainfall over south asia using complex networks,” *Climate Dynamics* **39**, 971–987 (2012).
- ⁴⁴N. Marwan, “Challenges and perspectives in recurrence analyses of event time series,” *Frontiers in Applied Mathematics and Statistics* **9**, 1129105 (2023).
- ⁴⁵I. Ozken, D. Eroglu, S. F. M. Breitenbach, N. Marwan, L. Tan, U. Tirnakli, and J. Kurths, “Recurrence plot analysis of irregularly sampled data,” *Physical Review E* **98**, 052215 (2018).
- ⁴⁶D. Eroglu, F. H. McRobie, I. Ozken, T. Stemler, K.-H. Wyrwoll, S. F. M. Breitenbach, N. Marwan, and J. Kurths, “Saw relationship of the Holocene East Asian-Australian summer monsoon,” *Nature Communications* **7**, 12929 (2016).
- ⁴⁷W. J. Masek and M. S. Paterson, “A faster algorithm computing string edit distances,” *Journal of Computer and System sciences* **20**, 18–31 (1980).
- ⁴⁸I. R. Cleasby, E. D. Wakefield, B. J. Morrissey, T. W. Bodey, S. C. Votier, S. Bearhop, and K. C. Hamer, “Using time-series similarity measures to compare animal movement trajectories in ecology,” *Behavioral Ecology and Sociobiology* **73**, 1–19 (2019).
- ⁴⁹I. Ozken, D. Eroglu, T. Stemler, N. Marwan, G. B. Bagci, and J. Kurths, “Transformation-cost time-series method for analyzing irregularly sampled data,” *Physical Review E* **91**, 062911 (2015).
- ⁵⁰T. Braun, C. N. Fernandez, D. Eroglu, A. Hartland, S. F. M. Breitenbach, and N. Marwan, “Sampling rate-corrected analysis of irregularly sampled time series,” *Physical Review E* **105**, 024206 (2022).
- ⁵¹C. Ozdes and D. Eroglu, “Transformation Cost Spectrum for Irregularly Sampled Time Series,” *European Physical Journal – Special Topics* (in press), 10.1140/epjs/s11734-022-00512-x.
- ⁵²W. Fischer and K. Meier-Hellstern, “The Markov-modulated Poisson process (MMPP) cookbook,” *Performance evaluation* **18**, 149–171 (1993).
- ⁵³E. N. Lorenz, “Deterministic Nonperiodic Flow,” *Journal of the Atmospheric Sciences* **20**, 130–141 (1963).
- ⁵⁴F. M. Ralph, J. J. Rutz, J. M. Cordeira, M. Dettinger, M. Anderson, D. Reynolds, L. J. Schick, and C. Smallcomb, “A scale to characterize the strength and impacts of atmospheric rivers,” *Bulletin of the American Meteorological Society* **100**, 269–289 (2019).
- ⁵⁵Z. Zhang, F. M. Ralph, and M. Zheng, “The relationship between extratropical cyclone strength and atmospheric river intensity and position,” *Geophysical Research Letters* **46**, 1814–1823 (2019).
- ⁵⁶D. A. Lavers and G. Villarini, “The nexus between atmospheric rivers and extreme precipitation across Europe,” *Geophysical Research Letters* **40**, 3259–3264 (2013).
- ⁵⁷J. Eiras-Barca, A. M. Ramos, I. Algarra, M. Vázquez, F. Dominguez, G. Miguez-Macho, R. Nieto, L. Gimeno, J. Taboada, and F. M. Ralph, “European West Coast atmospheric rivers: A scale to characterize strength and impacts,” *Weather and Climate Extremes* **31**, 100305 (2021).
- ⁵⁸A. Gershunov, T. Shulgina, F. M. Ralph, D. A. Lavers, and J. J. Rutz, “Assessing the climate-scale variability of atmospheric rivers affecting western North America,” *Geophysical Research Letters* **44**, 7900–7908 (2017).
- ⁵⁹A. E. Payne, M.-E. Demory, L. R. Leung, A. M. Ramos, C. A. Shields, J. J. Rutz, N. Siler, G. Villarini, A. Hall, and F. M.

- Ralph, “Responses and impacts of atmospheric rivers to climate change,” *Nature Reviews Earth & Environment* **1**, 143–157 (2020).
- ⁶⁰D. Traxl, “ARtracks – A Global Atmospheric River Catalogue Based on ERA5 and IPART (1.0.0),” Zenodo (2022), <https://doi.org/10.5281/zenodo.7018725>.
- ⁶¹G. Xu, X. Ma, and P. Chang, “IPART: A Python Package for Image-Processing based Atmospheric River Tracking,” *Journal of Open Source Software* **5**, 2407 (2020).
- ⁶²A. M. Ramos, R. Nieto, R. Tomé, L. Gimeno, R. M. Trigo, M. L. Liberato, and D. A. Lavers, “Atmospheric rivers moisture sources from a Lagrangian perspective,” *Earth System Dynamics* **7**, 371–384 (2016).