Geoscientific
Model Development

Methods for assessment of models

Scalable Feature Extraction and Tracking (SCAFET): a general framework for feature extraction from large climate data sets

Arjun Babu Nellikkattil[1,2]**, Danielle Lemmon**[1,3,4]**, Travis Allen O'Brien**[5,6]**, June-Yi Lee**[1,2,7]**, and Jung-Eun Chu**[8]

[1]Center for Climate Physics, Institute for Basic Science (IBS), Busan 46241, Republic of Korea

[2]Department of Climate System, Pusan National University, Busan 46241, Republic of Korea

[3]Pusan National University, Busan 46241, Republic of Korea

[4]Science and Technology Policy Fellowship Program, The American Association for the Advancement of Science,
Washington, D.C., 20005, USA

[5]Department of Earth and Atmospheric Sciences, Indiana University, Bloomington, IN 47403, USA

[6]Climate and Ecosystem Sciences Division, Lawrence Berkeley National Lab, Berkeley, CA 95720, USA

[7]Research Center for Climate Sciences, Pusan National University, Busan 46241, Republic of Korea

[8]Low-Carbon and Climate Impact Research Centre, School of Energy and Environment,
City University of Hong Kong, Hong Kong SAR, China

Correspondence: Arjun Babu Nellikkattil (arjunbabun@pusan.ac.kr)

Abstract. This study describes a generalized computational mathematical framework, Scalable Feature Extraction and Tracking (SCAFET), that extracts and tracks features from large climate data sets. SCAFET utilizes novel shape-based metrics that can identify and compare features from different mean states, data sets, and between distinct regions. Features of interest such as atmospheric rivers, tropical and extratropical cyclones, and jet streams are extracted by segmenting the data based on a scale-independent bounded variable called the shape index (SI). The SI gives a quantitative measurement of the local geometric shape of the field with respect to its surroundings. Compared to other widely used frameworks in feature detection, SCAFET does not use a posteriori assumptions about the climate model or mean state to extract features of interest and levelize the comparison between different models and scenarios. To demonstrate the capabilities of the method, we illustrate the detection of atmospheric rivers, tropical and extratropical cyclones, sea surface temperature fronts, and jet streams. Cyclones and atmospheric rivers are extracted to show how the algorithm identifies and tracks both the nodes and areas from climate data sets. The extraction of sea surface temperature fronts exemplifies how SCAFET effectively handles curvilinear grids. Last, jet streams are extracted to demonstrate how the algorithm can also detect three-dimensional features. As a generalized framework, SCAFET can be implemented to extract and track many weather and climate features across scales, grids, and dimensions.

1 Introduction

The volume of climate data is growing exponentially, owing to rapid expansions in both observational capabilities and computational power, which are driven in particular by the precision and insights offered by higher-resolution models (Overpeck et al., 2011; Balaji et al., 2018). Frontier research like global cloud-resolving and large-ensemble simulations leads not only to increased volume but also to inflated velocity, variety, veracity, and value (5Vs) (Marr, 2015; Guo, 2017; van Genderen et al., 2019) of climate data. This makes the detection and comparative analysis of important atmospheric and oceanic features, such as atmospheric rivers (ARs), tropical and extratropical cyclones, sea surface temperature fronts (SSTFs), and jet streams, an onerous task. Although these climate phenomena influence regional and global weather and climate with immense societal, economic, and ecological impacts, the volume of data represent-

ing these events and features is a small percentage of the whole simulation. Feature extraction considerably reduces the volume of data that needs to be stored, thus improving computational efficiency in analyzing these features (Yang et al., 2016). Moreover, the mean, variability, and characteristics of features can be compared to observational data sets as a measure of bias within model simulations, thereby improving our understanding of the causal differences between observations and models (Sellars et al., 2013). Thus, efficient and reliable feature extraction is vital to climate data processing, analysis, and model development.

Despite the importance of feature extraction in climate data analysis and model development, there is little consensus on standard best practices for feature extraction. The simplest method for extracting a feature is to use a physical threshold or its derivative for some climate variable (SST, precipitation, wind speed, humidity, etc.), or a combination thereof, to identify ARs, fronts, jet streams, or tropical and extratropical cyclones (Bengtsson et al., 1982, 1995; Vitart et al., 1997; Hewson, 1998; Koch et al., 2006; Strong and Davis, 2007; Rutz et al., 2014; Guan and Waliser, 2015). The limitations and discrepancies in these methods arise from the somewhat arbitrary choice of physical thresholds in relation to the spatiotemporal distributions of the climate variables. In other words, many studies choose a physical threshold that is not theoretically defined but rather a function of the location, time span, and data set used. Validation can then unfortunately come down to the intelligent but subjective human eye or, in other words, tuning an absolute or relative threshold until it appears to have captured all the features of interest, while leaving out the background noise (Zarzycki and Ullrich, 2017; Vishnu et al., 2020).

Choosing an absolute threshold from climate variables for feature extraction that applies to different climate models and spans multiple mean states and model scenarios is not straightforward. Thresholds are often applied to climate variables or derivatives in which the features are most visible, such as relative vorticity (RV) and sea level pressure anomalies for tropical cyclones (e.g., Vitart et al., 1997), integrated water vapor transport (IVT) for ARs (e.g., Guan and Waliser, 2015), or the first derivative of sea surface temperature (SST) for SST fronts (Castelao et al., 2006). Thresholds are often either empirically derived from observational studies or calculated from a model-specific distribution; though, even within the same data set, a particular choice of threshold may be suitable for one region but not for another, given varying regional characteristics and topography. In the case where the feature extraction threshold is an a posteriori assumption of the data set used, one must pre-process large, representative data sets just to calculate reasonable thresholds. While some detection methods have done well to streamline their algorithms to reduce total runtime, the process of posterior threshold calculation for higher-resolution and large-ensemble data sets inherently becomes increasingly less ef-

ficient, highlighting the need to develop feature extraction methods that do not use posterior assumptions.

Aside from the sensitivity of feature detection to intermodel and inter-simulation differences, feature detection is further complicated when trying to detect and compare features between present and future climate change scenarios, as the underlying spatiotemporal climate variable distributions change under global warming. Feature detection must be reconsidered when applied to variables with significant and/or non-linear changes in their means and extremes in response to external forcings such as doubling or quadrupling carbon dioxide concentrations. It should be emphasized that applying different arbitrary thresholds can and does lead to contradictory conclusions regarding the response of these features to greenhouse gas warming (Horn et al., 2014; Zhao, 2020; O'Brien et al., 2022; Nellikkattil et al., 2023). To counter these uncertainties, methods based on topology, machine learning, ridge extraction, edge detection, and various other image-processing techniques have been proposed over the years (Dixon and Wiener, 1993; Post et al., 2003; Molnos et al., 2017; Biard and Kunkel, 2019; Xu et al., 2020). While these methods offer an alternative for the extraction of features in data sets spanning different mean states, many of these methods were developed for detecting specific rather than general features.

The need for a general framework for extracting and tracking features from large climate data sets has been raised in various climate science communities for the last several decades. In a pioneer study, Hodges (1994) developed a general framework for extracting and tracking features from meteorological data sets in the following three steps: segmentation, filtering, and tracking. In the segmentation step, the field is split into distinct regions by applying a threshold and defining each of the connected regions as an object. Segmented regions are then filtered based on the characteristics of each object, and feature nodes are defined for the remaining objects. Finally, the feature nodes are tracked over time to produce the final output for further analysis. This framework was further developed for cyclones, storm tracks, convective systems, ocean eddies, monsoon depressions, etc. (Hodges, 1995; Hogg et al., 2005; Hodges et al., 2011; Burston et al., 2014; Hurley and Boos, 2014; Pinheiro et al., 2016; Priestley et al., 2020; Torres-Alavez et al., 2021; Karmakar et al., 2021). However, it is limited to the detection of points of local maxima in two-dimensional scalar fields, which do not always fully characterize various features.

In 2012, a team from the Lawrence Berkeley National Laboratory developed the Toolkit for Extreme Climate Analysis (TECA), integrating pre-existing, physical thresholddependent detection methods and algorithms into a comprehensive software package that was parallelized to make the algorithms more suitable for large data sets (Prabhat et al., 2012). In a more recent effort, a team led by Paul Ullrich at the University of California, Davis, created TempestExtremes (Ullrich and Zarzycki, 2017; Ullrich et al., 2021),

which is another computationally efficient algorithm package that uses C ++ and several core functions to detect a variety of features. These functions are being actively developed for extraction, characterization, and uncertainty quantification of weather extremes. Both TECA and TempestExtremes have been widely implemented by the climate community and have been monumental in advancing scientific understanding of meso- and synoptic-scale processes and their connections to long-term climate variability.

In this study, we present a novel method called Scalable Feature Extraction and Tracking (SCAFET), which serves as a versatile and general framework for detecting and tracking features of various shapes and intensities across scales, grid types, and dimensions. Simply put, SCAFET uses the curvature measurements of a given scalar field to identify distinct emergent shapes corresponding to features of interest. The local shape calculation is finite, bounded, and scale-independent, and it can be tuned, depending on the specified feature of interest. Unlike traditional methods that rely on physical thresholds often derived from data-specific, posterior conditions, this method relies on shape-based thresholds. As such, it separates the feature detection process from inter- and intra-model variation, making it less sensitive to these differences. Furthermore, this approach allows for the complete parallelization of feature extraction along the time dimension, since the detection operates independently of time. Time-independent feature extraction offers two key advantages. First, it has the potential to boost computational efficiency by enabling data pre-processing such as smoothing to occur in parallel, rather than requiring a single pre-processing step before feature extraction. Second, it holds the promise of being developed and implemented for real-time feature extraction during critical events like hurricanes and tornadoes. Importantly, the code for this framework is fully open-source and written in Python in an easy-to-use package so that even individuals with beginner-level Python skills can readily implement the algorithm (see https://github.com/nbarjun/SCAFET/blob/master/scafet_demo.ipynb, last access: 17 October 2023, for a simple working example).

The novelty of SCAFET compared to pre-existing methods lies in feature detection that does not use a posteriori assumptions and is based on the overall "shape" of a climate variable field, rather than arbitrary thresholding of that field or derivative. The core methodology for the detection of any feature is the same and can be tuned using just two variables: one for the spatial scale and the other for the shape of the features one is looking for. For example, between the two variables, one can tune the difference between a long filament-shaped atmospheric river and a shorter, round-shaped cyclone. The algorithm applies to both rectilinear and curvilinear grids and can also be extended to detect three-dimensional (3D) features. Even in the context of recent advancements in feature extraction such as TempestExtremes and TECA, SCAFET is a comprehensive, efficient, and eas-

ily implementable framework that aims to upgrade the feature extraction process with a novel shape-based approach that does not rely on iterative posterior conditions and could prove to be a robust method for detecting a diverse set of features under different mean climate states. Further discussion on the differences between SCAFET and other detection algorithms can be found in Sect. A.

The paper is organized as follows, Sect. 2 introduces the fundamentals of SCAFET and how it is implemented in a two-dimensional (2D) field. Section 3 presents three specific use cases of SCAFET, demonstrating its capabilities in detecting various climate features across different grid types. Extraction of 3D features using jet steams as an example will be discussed in Sect. 4. Though the application of SCAFET is not limited to the features described here, this study focuses on atmospheric rivers, cyclones, SST fronts, and jet streams, as these examples cover a broad range of phenomena, providing users with insights on how to adapt SCAFET to their specific use cases and requirements.

2 Description of Scalable Feature Extraction and Tracking

SCAFET adopts the same three-step approach as outlined by Hodges (1994): segmentation (yellow boxes in Fig. 1), filtering (orange boxes in Fig. 1), and tracking (green boxes in Fig. 1). However, before commencing these steps, SCAFET requires initialization with essential information describing the data sets and the specific feature to be extracted (as indicated by blue boxes in Fig. 1). The key inputs for this initialization include the following:

– *Primary field* (ϕ_p). This is a gridded data set in which the target feature is most easily distinguishable. For instance, cyclones are readily identified using the RV field, ARs emerge from IVT, and SSTFs are distinguished using the SST gradient. Optionally, one or more secondary fields can be used to further constrain the detected features.

– *Grid properties*. Information on the primary field's grid, including grid cell area/volume, grid distance, and coastlines, is required for calculating derivatives of the basic field and identifying locations of landfall.

– *Feature properties*. The algorithm requires information on the properties of the target feature. This includes estimated spatial scale, shape, eccentricity (for 2D features only), minimum length, minimum area, minimum volume (for 3D feature only), minimum duration, and maximum distance per time step.

In the SCAFET scheme, segmentation, filtering, and tracking are developed and coded as separate Python libraries. This design allows users to substitute any of these components
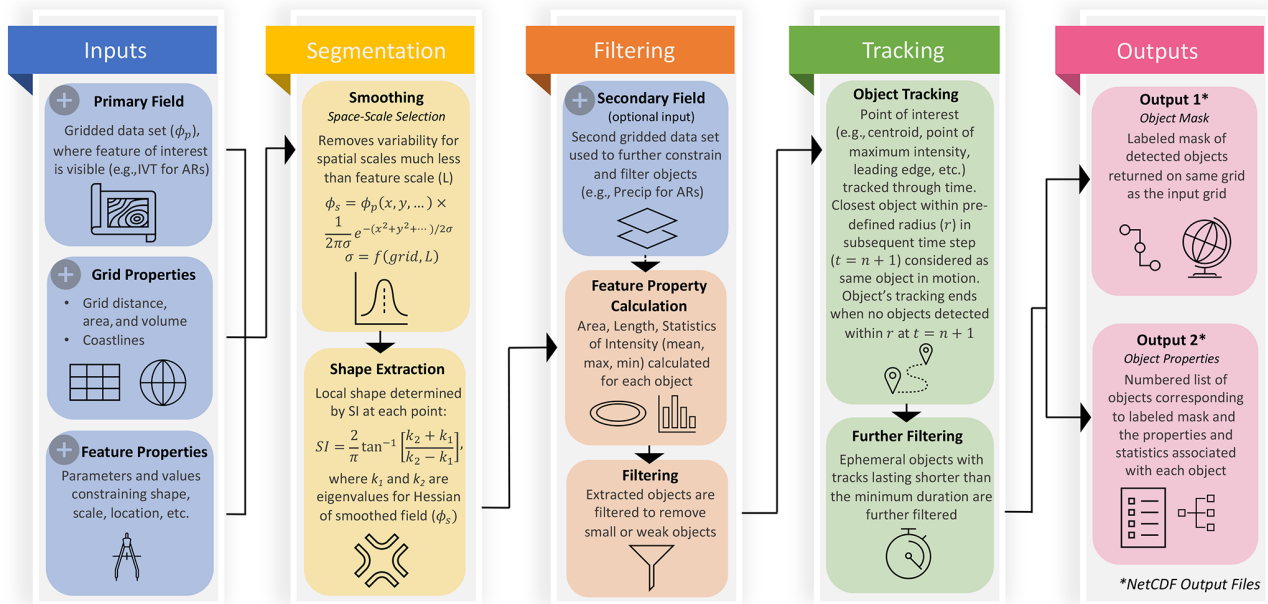
Figure 1. Overall schematic of SCAFET workflow and components. Inputs to the algorithm are depicted in blue, while the algorithm's outputs are shown in pink boxes. Processes related to the segmentation step are highlighted in yellow boxes, whereas the orange boxes represent the filtering processes. The tracking step is denoted by green boxes. Arrows on the periphery of the boxes illustrate the flow of the algorithm. Each section is elaborated upon in detail within the text.

with their own methods, while still being able to execute the algorithm. Once all three steps have been executed, the algorithm yields two outputs: one provides information about the properties of the detected objects and the other produces a labeled mask highlighting the feature of interest on the input grid (pink boxes in Fig. 1).

2.1 Segmentation

The core operation for the feature extraction involves categorizing points within a scalar field into one of five shapes. This categorization is achieved using curvature measurements obtained from the eigenvalues of the Hessian of the basic field. These five selected shapes (see Fig. 2) are an abridged version of the shapes described in previous studies (Koenderink and van Doorn, 1992). Depending on the specific feature of interest, one or more shapes are extracted from the primary field. The segmentation process starts with scale–space selection of the field to remove smaller scales of variability that are background noise compared to the feature of interest. Last, the algorithm calculates SI to estimate the local geometric shape at each point.

2.1.1 Scale–space selection

Scale–space selection is a widely used technique in image processing, signal processing, and computer vision (Lindeberg, 2014). In our current study, scale–space selection involves applying a Gaussian smoothing kernel to sup-

press variability smaller than the chosen smoothing scale (σ) (see https://unidata.github.io/MetPy/latest/api/generated/metpy.calc.smooth_gaussian.html, last access: 20 December 2023, for the implementation of Gaussian smoothing). Mathematically, scale–space selection is performed by convolving the primary field (ϕ_p) with a Gaussian function, which is expressed as follows:

$$\phi_s(x, y, \ldots) = \phi_p(x, y, \ldots) \cdot \frac{1}{2\pi\sigma} e^{-(x^2+y^2+\ldots)/2\sigma^2}. \tag{1}$$

In the context of the meso–synoptic-scale processes examined in this study, the scale–space selection filters out smaller microscale features to isolate features like cyclonic vortexes or atmospheric rivers. Notably, this function can be adjusted to the spatial scale of interest and could also be used to filter out synoptic-scale features in isolating micro- and mesoscale processes. In climate data sets, grid spacing is not always uniform. To account for that, we adapt the above equation to be "grid-aware". The input for the smoothing scale is provided in kilometers, and based on this input, we calculate the value of σ, while considering the grid size. Notably, the value of σ remains constant when smoothing is applied along each longitude, but it varies along each circle of latitude. For future studies, researchers may explore other, more advanced scale–space selection methods to further refine their analyses.

2.1.2 Local shape extraction

The local geometric shape of the field, ϕ_s is calculated as a function of the eigenvalues (k_1 and k_2) of the Hessian of the magnitude of the field ($|\phi_s|$), where the Hessian is given by

$$\mathcal{H}\left(|\phi_s|\right) = \begin{bmatrix} \frac{\partial^2 |\phi_s|}{\partial x^2} & \frac{\partial^2 |\phi_s|}{\partial x \partial y} \\ \frac{\partial^2 |\phi_s|}{\partial y \partial x} & \frac{\partial^2 |\phi_s|}{\partial y^2} . \end{bmatrix} \quad (2)$$

In the context of simple differential geometry, we can determine whether a point is a local maximum or a local minimum based on the eigenvalues k_1 and k_2. Specifically, if $k_2 \leq k_1 < 0$, then the point under consideration is a local maximum, whereas if $k_1 \geq k_2 > 0$, then point is a local minimum. The criterion is primarily applicable to nodal features such as tropical cyclones or monsoon depressions. To expand our ability to identify a range of features, we use the shape index (SI) (Koenderink and van Doorn, 1992), a quantitative measure of the local shape of the field defined as

$$\mathrm{SI}(k_1, k_2) = \frac{2}{\pi} \tan^{-1} \left[\frac{k_2 + k_1}{k_2 - k_1} \right], \quad (3)$$

where k_1 and k_2 are the two eigenvalues, satisfying $k_1 \geq k_2$, for the Hessian matrix. It is important to clarify that in the original work by Koenderink and van Doorn (1992), the principal curvatures, not the eigenvalues of the field, are utilized to calculate the SI. However, the disparity between the SI calculated using principal curvatures and the SI derived from eigenvalues is exceedingly minimal in climate data analysis. The SI is used to categorize the primary field into distinct shapes (see Fig. 2). The choice of SI values is contingent upon the specific type of feature to be extracted. For example, we select caps and domes when extracting features such as atmospheric depressions or cyclones, whereas ridges, caps, and domes are chosen when targeting features like ARs and fronts.

SI is designed to be a bounded value (range -1 to 1) independent of the magnitude of the field (Fig. 3). In simple terms, SI provides a continuous and quantitative measurement of the geometric shape of the field with respect to its immediate background field. This concept is similar to how a climate scientist's trained eye identifies features based on differences in color or value contrast, though the SI is arguably a more objective and precise measure of geometric shape. These characteristics make the SI particularly well-suited for feature extraction from data sets with varying mean states, which is in contrast to traditional physical threshold-based methods. In addition to the two eigenvalues, the shape extraction provides us with corresponding eigenvectors. The eigenvector for k_1 points perpendicular to the local ridge direction, while that of k_2 is parallel to it. This allows us to impose further constraints, such as the coherence of transport or flow with respect to the local ridge when ϕ_s is a vector field. This capability is aptly demonstrated in the context of AR detection, as discussed in Sect. 3.1.

2.2 Filtering

Once the target features are extracted, properties like area, location, mean, minimum, and maximum values of different properties are calculated for each of the objects. A series of filtering is carried out to remove objects that do not satisfy certain conditions regarding (a) grid properties like area, length, and region masks, (b) primary field properties like magnitude and direction, and (c) constraints from the secondary field(s). The primary aim of the filtering process is to remove small, weak, or ephemeral objects.

2.3 Tracking

The properties extracted for each object include key positional details, such as its centroid, and endpoints, as well as the locations of maximum and minimum intensity of input field within each object. To follow objects through time, one of these positional attributes is tracked. In the present study, we employ a straightforward tracking method. For each object at time step n, we identify the closest object to it at time $n + 1$. If this identified object is closer than a pre-defined radius r, we consider it to be the same object in motion. The radius r is defined in kilometers based on the maximum translation speed of the object and the temporal frequency of the input data. At this stage, it is possible to filter out short-lived features as needed. While this uncomplicated tracking approach may not be suitable for microscale processes, it can be adapted to incorporate greater complexity if necessary.

3 Application to 2D features

In this section, we showcase how SCAFET is employed to detect cyclonic vortices, ARs, and SSTFs from various climate data sets. These examples serve to illustrate the versatility of SCAFET as applied to different types of features and grids, though all the examples in this study follow the same general process shown in Fig. 1. Each subsection has a table of parameters detailing the properties of the desired feature. The properties include the feature's typical spatial scales, shape index (SI) regime, minimum length, minimum area, object eccentricity, and minimum duration of its track. To determine the quantitative values for these properties, we refer to a consensus among previous studies, which are cited within each section. A detailed examination of the sensitivity of these parameters in relation to the detected features, using AR detection as an example, can be found in Sect. S1 in the Supplement. In addition to the results discussed in the following sections, videos are also included in the Supplement for each of the features. The primary objective of this work is to demonstrate SCAFET's capability to detect a variety of features. Consequently, we present results for the long-term climatology of each of the features, enabling a comparison with other published detection algorithms.
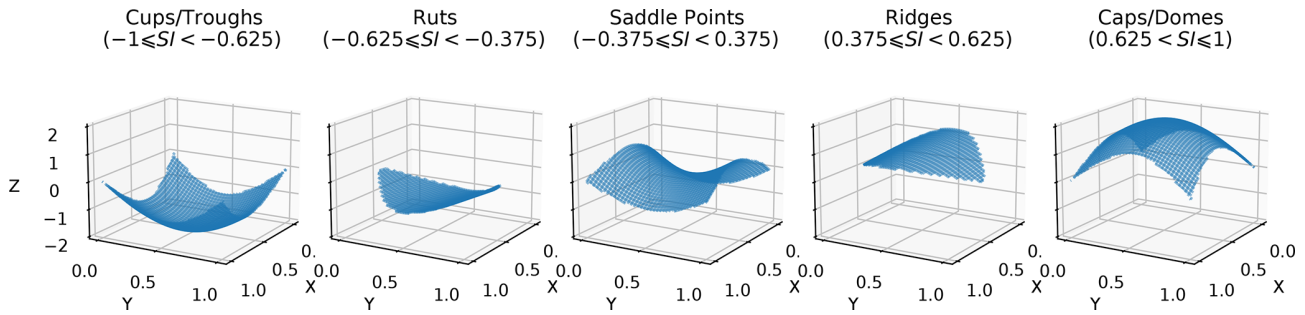
Figure 2. Selected shapes used in this study and the values of the shape index associated with each of them. The x and y axes are a set of general axes, while $Z(X,Y) = \sin(2X) + \cos(2Y)$. Regions within $Z(X,Y)$ satisfying conditions for different shapes are isolated to show the geometry associated with them.
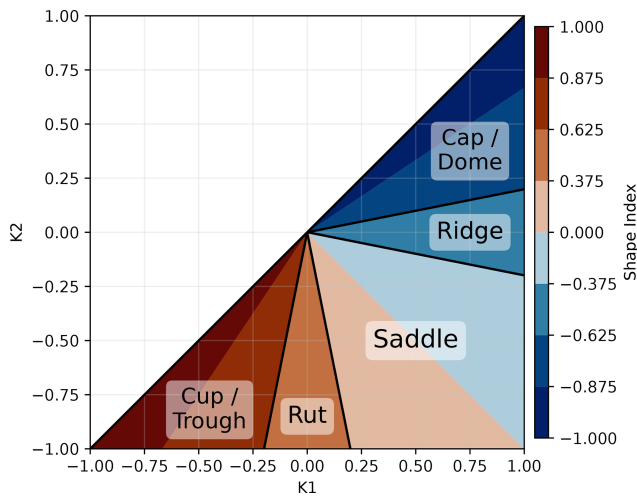

Figure 3. Sensitivity of the shape index (SI) to eigenvalues k_1 and k_2. The x and y axes represent values of the two eigenvalues used for calculating the shape index, while the color indicates the value of the shape index. Shapes corresponding to the SI regimes are labeled. Shapes corresponding to SI regimes are labeled.

3.1 Atmospheric rivers

According to the American Meteorology Society's glossary of meteorology, ARs are "long, narrow, and transient corridors of strong horizontal water vapor transport that are typically associated with a low-level jet stream ahead of the cold front of an extratropical cyclone" (Ralph et al., 2018). A substantial portion of the precipitation and water vapor transport in midlatitude regions is concentrated within ARs (Guan and Waliser, 2015). These atmospheric phenomena play a significant role in midlatitude hydrology, contributing to more than 50 % of the extreme precipitation and wind events in the region (Waliser and Guan, 2017; Nash et al., 2018). The ability to accurately detect, forecast, and project future ARs is of utmost importance for both extreme weather preparedness, as well as for water resource management in basins worldwide.

The ambiguity in AR projections and AR detection tools (ARDTs) stems from the lack of a clear quantitative definition of ARs in strength, length, narrowness, and other such parameters used in detection. In comparison with other criteria, the choice of threshold for AR strength has a significant effect on the inferences drawn between the detection schemes (Zhao, 2020; O'Brien et al., 2022; Nellikkattil et al., 2023). Many ARDTs determine this threshold empirically from the data set itself, which renders them sensitive to spatiotemporal variations and changes in mean-state conditions (Shields et al., 2018). SCAFET defines ARs as long (length $> 2000\,\text{km}$), narrow (eccentricity > 0.75) regions of strong water vapor transport (SI > 0.375) and significant precipitation (minimum AR precipitation $> 1\,\text{mm}\,\text{d}^{-1}$) (see Table 1 for complete details). The sensitivity of these parameters in AR detection to the characteristics of detected ARs is discussed in Sect. S1. This approach reduces the sensitivity of AR characteristics to arbitrary strength thresholds, making it easier to compare ARs across different mean state conditions.

To illustrate how SCAFET identifies ARs, we utilized daily mean data from the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis Version 5 (ERA5; Hersbach et al., 2020) for the period 2000 to 2019. The key fields of interest included the daily mean integrated water vapor transport (IVT) as the primary field and the daily mean total precipitation as the secondary variable. All the data sets employed share a spatial resolution of $0.25° \times 0.25°$. The vector field, IVT, is calculated as

$$\text{IVT}x = -\frac{1}{g}\int_{1000\,\text{hPa}}^{300\,\text{hPa}} q\,u\,\mathrm{d}p \tag{4}$$

$$\text{IVT}y = -\frac{1}{g}\int_{1000\,\text{hPa}}^{300\,\text{hPa}} q\,v\,\mathrm{d}p \tag{5}$$

$$|\text{IVT}| = \sqrt{\text{IVT}x^2 + \text{IVT}y^2}. \tag{6}$$

where u, v, and q are the zonal wind speed, meridional wind speed, and specific humidity, respectively.

To detect AR-like structures, SCAFET employs a search for specific shapes, such as ridges, caps, and domes (see Fig. 2). Following the process outlined in Fig. 1, the SI is calculated after applying a grid-aware smoothing technique that suppresses variability smaller than 1000 km (Fig. 4a). Once the SI is calculated for |IVT| (Fig. 4b), regions where the SI > 0.375 are passed on to the next stage for filtering. To maximally utilize the vector qualities of the primary field, we ensure that the local transport direction (arrows in Fig. 4a) and local ridge direction (arrows in Fig. 4b) do not deviate by more than 45°. The local ridge direction is identified as the eigenvector corresponding to the smallest eigenvalue (k_2). Filtering based on the grid properties removes candidates that are too small (length < 2000 km and area < 2×10^6 km^2) or too wide (eccentricity ≤ 0.75). To eliminate AR-like objects with low strength (precipitation < 1 mm d^{-1}), we constrain our results with the secondary field, total precipitation, to within the object's area. The use of precipitation as a strength indicator is relevant, given its significant socioeconomic impact. In line with other ARDTs, we impose a regional mask to filter out AR-like structures along the equatorial belt. All these steps can be applied in parallel along the time axis, and at each time step, AR-like structures similar to those shown in Fig. 4c are identified. Once all ARs are detected, the tracking algorithm is applied to the daily data to filter out ARs that last shorter than 1 d. Tracking is performed based on the centroid of each identified object. The closest objects within a distance of 4000 km between two consecutive time steps are considered to be the same object evolving over time (Fig. 4d). The annual mean frequency of the detected AR objects and their seasonality are shown in Fig. 4e, f, and g. SCAFET's identification of ARs is consistent with other ARDTs, both in terms of detecting single events and determining their mean climatology, as further detailed in the Sect. S2 (see also Lora et al., 2020).

3.2 Tropical and extratropical cyclones

In the scientific literature, cyclones are generally described as large weather systems ranging from 500–4000 km in size, characterized by strong cyclonic circulation, low pressure at their center, and exceptionally high winds around them (Emanuel, 2003; Schultz et al., 2019; Encyclopaedia, 2022). The dynamics and characteristics of cyclones can vary, depending on factors such as their genesis location and translation speeds. For instance, cyclones generated near the Equator, commonly referred to as tropical cyclones, are typically smaller in size compared to those formed in midlatitudes, known as extratropical cyclones. Regardless of their origin, cyclones have the potential to unleash intense rainfall and powerful winds along their path and can lead to flooding, landslides, and severe damage to coastal infrastructure when they make landfall (Knutson et al., 2010; Mendelsohn et al., 2012; Ranson et al., 2014). Moreover, the impact of cyclones is becoming a subject of heightened public concern, due to

rising sea levels and the potential for increased cyclone intensity in response to global warming. Thus, the identification and future projection of cyclones are a subject of growing attention and importance for the climate community (Woodruff et al., 2013).

Once again, discrepancies among different detection algorithms can be attributed to varying choices of physical thresholds or constraints related to factors such as size, wind speeds, vorticity, or surface pressure anomalies. While most studies generally agree on the present and future characteristics of cyclones, resolving details such as the changes in genesis rate and durations is complicated by the uncertainties in the detection methods (Ulbrich et al., 2009; Neu et al., 2013; Horn et al., 2014; Walsh et al., 2015). In this study, SCAFET identifies cyclones as regions of strong local maxima of cyclonic circulation (SI > 0.625) and maximum wind speeds exceeding 10 m s^{-1}. This definition enables the detection of robust cyclonic vorticities worldwide, including but not limited to tropical and extratropical cyclones. The primary field used for cyclone detection is the absolute value of cyclonic relative vorticity (ζ) defined as

$$\zeta = \nabla \times \boldsymbol{U}, \tag{7}$$

where $\boldsymbol{U}$ is the 6 h wind speeds at 10 m above the surface obtained from the ERA5 reanalysis data set, with a spatial resolution of $0.25° \times 0.25°$ (Hersbach et al., 2020). The magnitude of the wind speed at 10 m is utilized as the secondary field to constrain detection. Additional cyclone-related variables such as surface pressure anomaly and potential temperature can also serve as secondary fields for the identification and classification of cyclones.

In contrast with ARs, the detection of cyclones relies on a scalar field, specifically in this case the cyclonic relative vorticity $|\zeta|$. First, the data are pre-processed with grid-aware Gaussian smoothing to suppress spatial variability smaller than 750 km (Fig. 5a). The chosen smoothing scale allows us to identify both tropical and extratropical cyclones. Caps and dome shapes (SI > 0.625) are then identified within the smoothed $|\zeta|$ field as potential cyclones (Fig. 5b). Subsequently, objects with an area smaller than 10^5 km^2 and a diameter shorter than 20 km are filtered out. Once these spatial criteria are met, we can further refine our selection by excluding weak cyclonic vorticities $|\zeta| < 10^{-6}$ s^{-1} and slow maximum wind speed < 10 m s^{-1}, resulting in the identification of robust cyclonic systems for a given time step (Fig. 5c). Similar to the AR example, all the described steps can be parallelized along the time dimension. Once potential cyclones are identified, they are tracked using a methodology similar to the AR tracking algorithm. However, in this case, the radius for search is limited to 1000 km, since we are using 6 h data, and the translation speeds of cyclones are notably slower than 150 km h^{-1}. A minimum duration of 48 h and a minimum total displacement of 500 km is applied to distinguish moving cyclonic circulations from stationary ones. An example of a tracked cyclone, commonly known as "Dorian"

Table 1. The table presents the values for various parameters used in the detection of ARs using SCAFET. The rows for each step, including segmentation, filtering, and tracking are grouped together and labeled.

	No.	Property	Value	Unit
	1	Smoothing scale	2000	km
Segmentation	2	Angle coherence	45	degrees
	3	Selected shape	(0.375,1.0]	–
	1	Minimum length	2000	km
	2	Minimum area	2×10^6	km^2
Filtering	3	Eccentricity	[0.75, 1.0]	–
	4	Minimum precipitation	1	mm d^{-1}
	5	Latitude mask	(−20, 20)	degrees
Tracking	1	Minimum duration	24	hours
	2	Maximum distance per time step	4000	km

Table 2. Same as in Table 1 but for parameters and values relevant to detecting tropical and extratropical cyclones.

	No.	Property	Value	Unit
Segmentation	1	Smoothing scale	1500	km
	2	Selected shape	(0.625,1.0]	–
	1	Minimum length	20	km
	2	Minimum area	10^5	km^2
Filtering	3	Eccentricity	[0.0,1.0]	–
	4	Minimum vorticity	10^{-6}	s^{-1}
	5	Minimum maximum wind speed	10	m s^{-1}
	1	Minimum duration	48	hours
Tracking	2	Maximum distance per time step	500	km
	3	Net minimum displacement	1000	km

(Avila et al., 2020) is compared with the observed track from IBTrACS (Knapp et al., 2010, 2018) data set (Fig. 5d). In comparison to the observed track, SCAFET's track is much longer, due to the more relaxed conditions applied to cyclonic vorticity and wind speed. Additionally, SCAFET does not differentiate between tropical and extratropical cyclones, which can result in tracking the object throughout its transition from a tropical cyclone to a midlatitude storm. Despite this difference, the long-term averages for cyclone frequency and its seasonal variability calculated using SCAFET are comparable with other studies (e.g., Ullrich and Zarzycki, 2017). What sets SCAFET apart from other conventional cyclone detection algorithms is its approach to identifying cyclones not as point objects but as encompassing surfaces around the point of maximum $|\zeta|$. This enables a more comprehensive analysis of cyclone properties, including maximum and minimum values of wind speed and precipitation within the entire cyclone structure.

3.3 Sea surface temperature fronts

SST fronts are regions where different water masses come together. They are typically characterized by strong horizontal gradients in temperature, salinity, density, and other properties (Bowman, 1978; Legeckis, 1978; Fedorov, 1986; Yoder et al., 1994). Unlike the larger meso- to synoptic-scale features discussed in this study, frontal structures are often observed in much smaller spatiotemporal scales. Accurate identification of SSTFs is essential because these features are frequently associated with strong upwelling and high levels of biogeochemical productivity (Clayton et al., 2014, 2021; Nagai and Clayton, 2017). Additionally, the detection of SSTFs serves as an example of how SCAFET can be applied to identify features in curvilinear grids.

Many prior SSTF detection algorithms rely on edge detection techniques and the gradient of sea surface temperature and/or height to identify these structures (Canny, 1986; Castelao et al., 2006). In our approach, we utilize the magnitude of the daily mean SST horizontal gradient as the primary field for detecting SST fronts. The SST data are obtained from a fully coupled, ultrahigh-resolution (≈ 10 km) Community Earth System Model (CESM) v1.2.2 simulation of present-day mean climate (Small et al., 2014; Chu et al., 2020; Nellikkattil et al., 2023). The data are processed by SCAFET in the tripolar (POP) grid. To illustrate the detec-
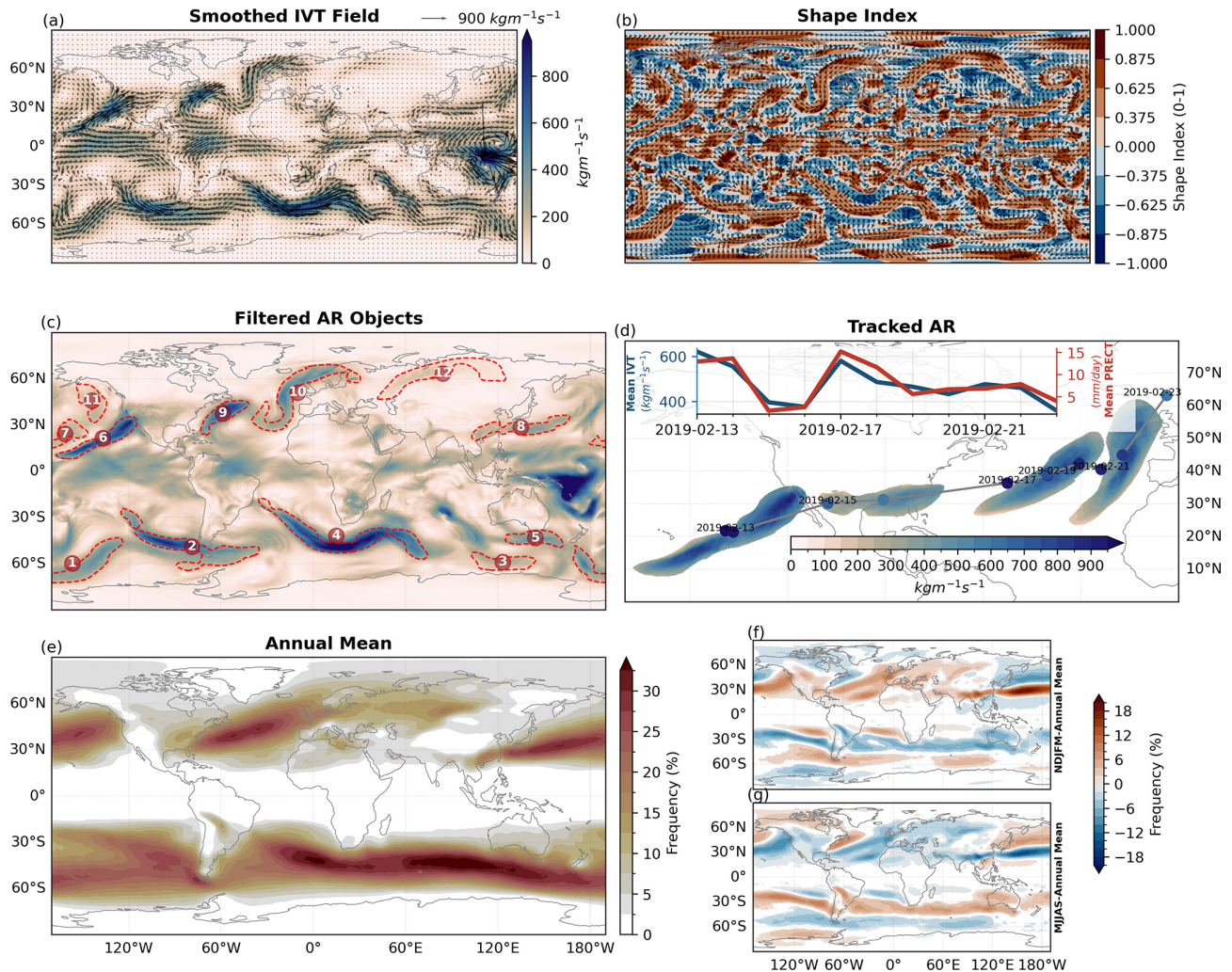
Figure 4. Major steps in the detection and tracking of atmospheric rivers. **(a)** Smoothed primary field of vertically integrated water vapor transport (IVT). Smoothing removes variability smaller than 1000 km from the IVT. The arrows in panel **(a)** represent the direction of unsmoothed IVT. **(b)** Magnitude (shading) of the shape index (SI) and direction of the local ridge (arrows) direction calculated from smoothed IVT. **(c)** Labeled AR objects after filtering out weak, small, and ephemeral candidates. **(d)** Example of tracked AR centroids and marked time; the inlay shows the object's area mean IVT over time. **(e)** AR annual mean frequency for the period 2000 to 2019. **(f, g)** AR frequency anomaly relative to the annual mean for **(f)** November to March and **(g)** May to September.

tion process, the analysis focuses on the Kuroshio frontal and extension region for the last 10 years of the simulation.

The extraction of frontal structures using the selected shapes of ridges, caps, and domes is similar to the method for the detection of ARs. Prior to extraction, a spatial smoothing of approximately 30 km is applied. From the extracted SSTF candidates, objects with a mean SST gradient lower than 10^{-4} K m^{-1} are removed. Circular (eccentricity < 0.5) and small (area < 1000 km^2) objects are also filtered out. It is worth noting that, in contrast to AR detection, frontal structures are not tracked. The detected frontal frequency exhibits general patterns and seasonality consistent with findings in previous studies (Xi et al., 2022).

4 Application to 3D features

This section introduces the extension of SCAFET to detect features within three-dimensional (3D) primary fields. The process of scale–space selection involves applying Gaussian smoothing independently along each of the three dimensions. Notably, a 3D field yields three eigenvalues ($k_1 \geq k_2 \geq k_3$) instead of the usual two. In this context, the SI can be calculated in three different ways by combining these eigenvalues.
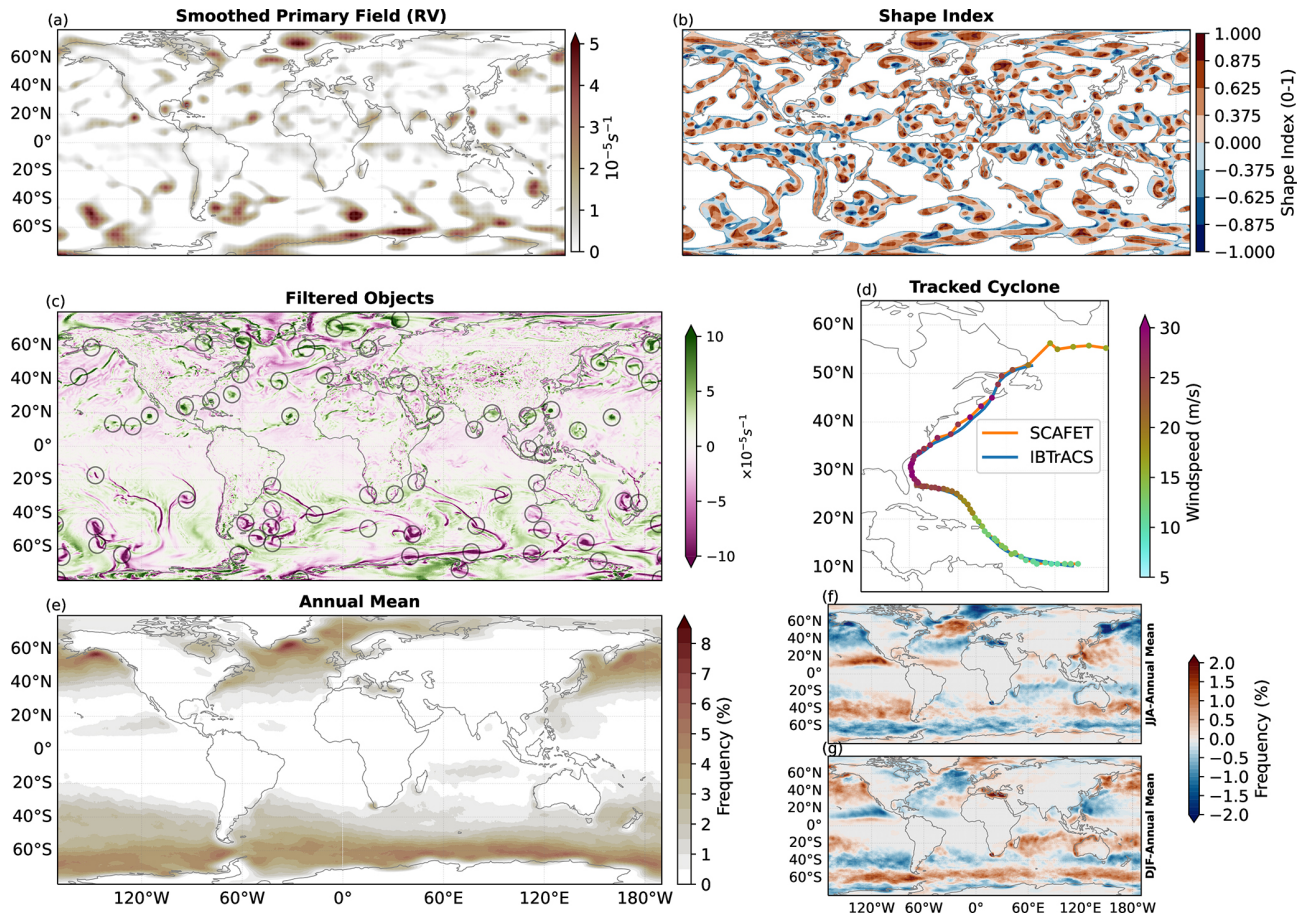
Figure 5. Major steps in the detection and tracking of cyclones. **(a)** Smoothed primary field of cyclonic relative vorticity ($|\zeta|$). The smoothing removes variability smaller than 750 km from $|\zeta|$. **(b)** Magnitude of the SI for the primary field. **(c)** Filtered cyclonic objects, with the background color representing unsmoothed values of ζ. **(d)** Track obtained for cyclone "Dorian" from SCAFET compared with the track from the IBTrACS data set. **(e)** Annual mean frequency of cyclone occurrence for the period 2000 to 2020. **(f, g)** Anomalous cyclone frequencies relative to the annual mean for **(f)** JJA (June–August) and **(g)** DJF (December–February).

Table 3. Same as in Table 1 but for parameters and values relevant to detecting sea surface temperature fronts (SSTFs).

	No.	Property	Value	Unit
Segmentation	1	Smoothing scale	30	km
	2	Selected shape	(0.375, 1.0]	–
Filtering	1	Minimum length	500	km
	2	Minimum area	10^3	km^2
	3	Eccentricity	(0.5, 1.0]	–
	4	Minimum SST gradient	10^{-4}	$K\,m^{-1}$

For the extraction of jet streams, the SI calculated using k_1 and k_2 (the two largest eigenvalues) is used, as it provides a more conservative estimate for the jet-like structure (see Appendix A3 and Fig. S7). The decision to exclude the smallest eigenvalue, denoted as k_3, is based on empirical observations. Empirical evidence suggests that when dealing with regions exhibiting positive maxima (convex curvature), both SI(k_1, k_2) and $SI(k_1, k_3)$ effectively capture the

shape. Meanwhile, SI(k_2, k_3) has a trivial application (refer to Fig. A4). Conversely, for concave shapes, both SI(k_1, k_3) and SI(k_2, k_3) represent the shape, while the conditions for SI(k_1, k_2) become redundant, given that they are satisfied by SI(k_1, k_3) and SI(k_2, k_3).
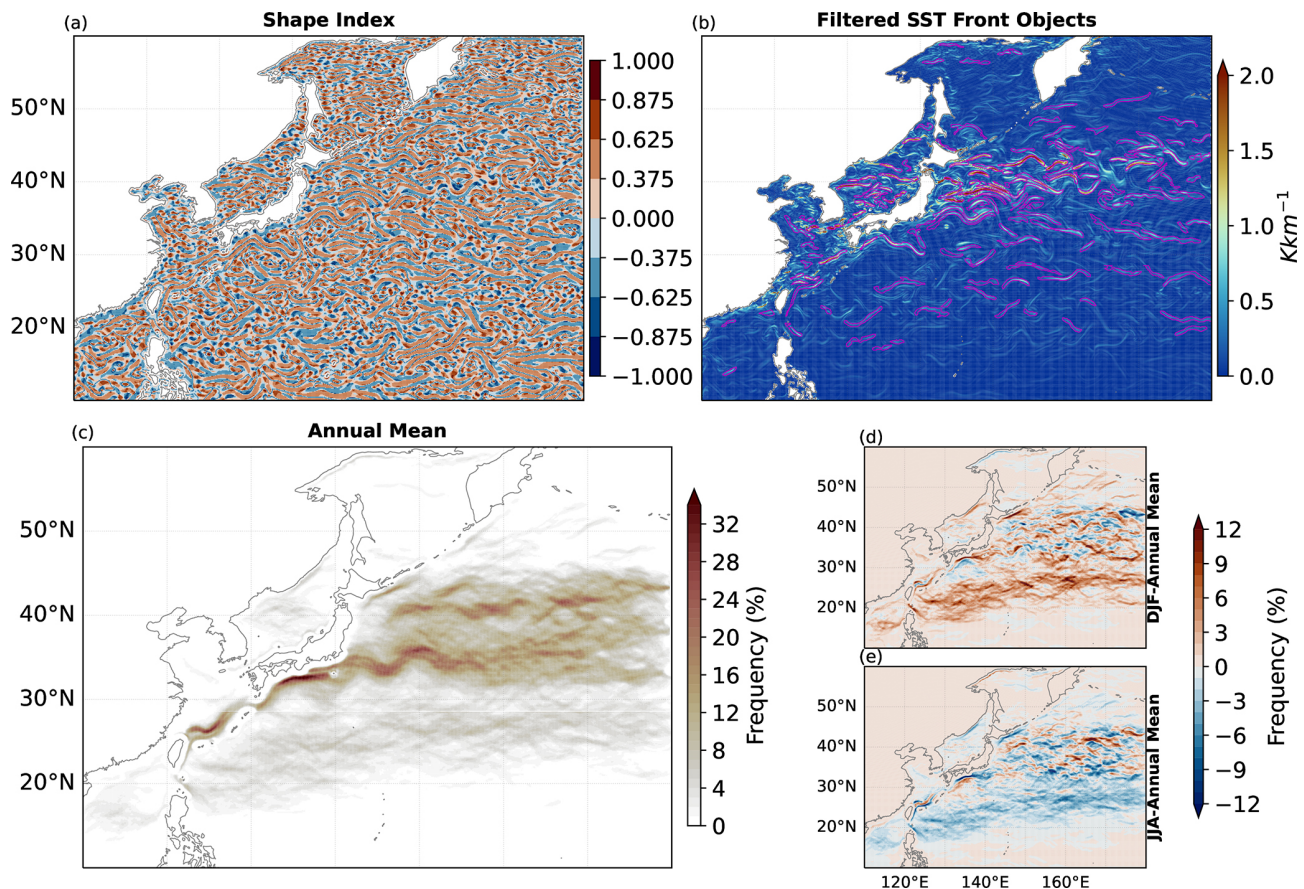
Figure 6. Major steps in the detection of sea surface temperature fronts (SSTFs). **(a)** Magnitude of the SI as calculated from the smoothed primary field of the horizontal gradient of sea surface temperature (∇SST). Smoothing removes variability smaller than 15 km from ∇SST. **(b)** Filtered SSTF objects, in units of Kelvins per kilometer ($\mathrm{K\,km^{-1}}$), where the background color represents unsmoothed values of ∇SST. **(c)** Annual mean frequency of SSTF occurrence across a 10-year period in the present climate simulation. **(d, e)** Anomalous frontal frequencies relative to the annual mean for **(d)** JJA and **(e)** DJF.

4.1 Jet streams

Jet streams, regardless of the underlying dynamics, are narrow, high-wind-speed regions in the upper atmosphere, with faster wind speeds compared to their surroundings (Koch et al., 2006). These jet streams have a significant impact on aviation and strongly influence surface weather conditions. For example, a persistent jet stream in boreal summer can result in extreme heat and flooding events, while a meandering jet stream in winter leads to severe cold spells in the midlatitudes (Petoukhov et al., 2013; Coumou et al., 2014; Kretschmer et al., 2016). Additionally, the northward movement of jet streams due to greenhouse warming contributes to the poleward propagation of tropical cyclones (Studholme et al., 2021). Thus, accurately detecting and characterizing jet streams is crucial for predicting and projecting both climatology and extreme weather systems.

Much like the detection of other weather phenomena discussed in this study, previous research typically employs a physical threshold to identify jet streams. Furthermore, with

the exceptions of Limbach et al. (2012) and Kern et al. (2018), most studies identify jet streams as either one- or two-dimensional features. However, it is important to emphasize that this section's focus is primarily on illustrating the method for detecting jet streams rather than validation of any analysis with published work. There is currently limited analysis available for comparing with a 3D perspective of jet streams, highlighting the need for such an approach. As a result, we present examples of jet stream detection in three selected time steps. A more comprehensive analysis and discussion regarding of the long-term characteristics of jet streams will be a topic for future research. For those interested, a video showcasing the results over an extended period can be found in the Supplement.

The primary field used in the extraction of jet streams is the 6 h, three-dimensional wind speeds obtained from ERA5 reanalysis data set, with a spatial resolution of 1° with 37 vertical levels (Hersbach et al., 2020). The magnitude of wind

Table 4. Same as in Table 1 but for parameters and values relevant for detecting jet streams.

	No.	Property	Value	Unit
Segmentation	1	Smoothing scale	6000	km
	2	Selected shape	(0.375,1.0]	–
Filtering	1	Minimum length	5000	km
	2	Minimum height	5	km
	3	Minimum volume	10^6	km^3
	4	Minimum maximum wind speed	50	$m\,s^{-1}$

speed is calculated as

$$W = \sqrt{u^2 + v^2}, \tag{8}$$

where u and v are the zonal and meridional wind velocities.

The detection process for jet streams begins similarly to the detection of 2D features. Gaussian smoothing is used to remove variability shorter than 3000 km in the horizontal dimensions. No smoothing is applied along the vertical dimension. Next, the SI is calculated using the two largest eigenvalues, k_1 and k_2. The vertical dimension for the three-dimensional wind speed is given in pressure coordinates. To calculate the gradient as change in wind speeds per kilometer, a rudimentary conversion from pressure to height coordinates is used (refer to Wallace and Hobbs, 1977, p. 60–61, and https://unidata.github.io/MetPy/latest/api/generated/metpy.calc.pressure_to_height_std.html, last access: 20 December 2023, for further details).

Similar to the detection of ARs, regions characterized by the selected shapes of ridges, caps, and domes (SI > 0.375) are isolated for filtering. Filtering is then applied to remove objects with a volume less than $10^6\,km^3$, a horizontal length less than 5000 km, and a maximum wind speed within each object less than $50\,m\,s^{-1}$. In the current version of SCAFET, the tracking algorithm is not applied to jet detection (see Fig. 7). The detailed list of parameters used in the detection of jet streams is given in Table 4.

5   Conclusions

In this study, we introduced a novel computational mathematical framework and an open-source Python package for extracting and tracking features from large climate data sets, called Scalable Feature Extraction and Tracking (SCAFET). The purpose of SCAFET is to tackle the challenges posed by the increasing volume and diversity of climate data by providing an alternative to traditional physical threshold-based feature detection methods. It enables the comparison of features between observational and model data with different mean states by attempting to remove the need for posterior data-specific assumptions. Furthermore, SCAFET introduces a novel shape-based approach to feature extraction, which helps uncover discrepancies in climate projections due

to differences in detection methods and aims to help the community in building scientific consensus. To demonstrate SCAFET's capabilities and its potential in advancing these goals, we showcased its ability to detect various features, including two-dimensional features such as atmospheric rivers (ARs), tropical and extratropical cyclones, sea surface temperature fronts, and the detection of three-dimensional jet streams. Each application serves as an illustrative example from which users can customize SCAFET for their specific research needs.

SCAFET offers several significant advantages, including a more comprehensive framework and parallel computing implementation for efficiency. However, its most noteworthy contribution lies in offering a novel perspective on how we can relatively define various features within climate data sets that span extensive periods marked by significant changes in mean climate. Rather than relying on empirically derived, data-specific physical thresholds for feature extraction, SCAFET identifies features using shape-based absolute thresholds and the locally estimated shape within the field. This methodology offers a unique viewpoint, enabling us to observe the continuous changes in feature properties while accounting for shifts in the mean climate state. This approach is particularly valuable, as meso–synoptic-scale studies are highly sensitive to thresholds in a dynamically changing mean climate state. Consequently, the conclusions drawn from such studies can vary significantly, as demonstrated in research examining the response of ARs to greenhouse warming (Zhao, 2020; O'Brien et al., 2022; Nellikkattil et al., 2023). Thus, algorithms like SCAFET which are not influenced by data-specific conditions of various climate models play a crucial role in advancing scientific understanding and facilitating climate model development.

In conclusion, delving deeper into the principles of differential geometry to elucidate the physical interpretation of the relationship between the SI and local geometric shape has the potential to revolutionize our approach to feature extraction from large data sets. This avenue of research has the promise of significantly enhancing the algorithm's robustness and reliability. It is worth noting that, at present, SCAFET may not surpass the computational efficiency of other well-established feature extraction methods discussed above (see Sect. S2.2). However, ongoing efforts to optimize
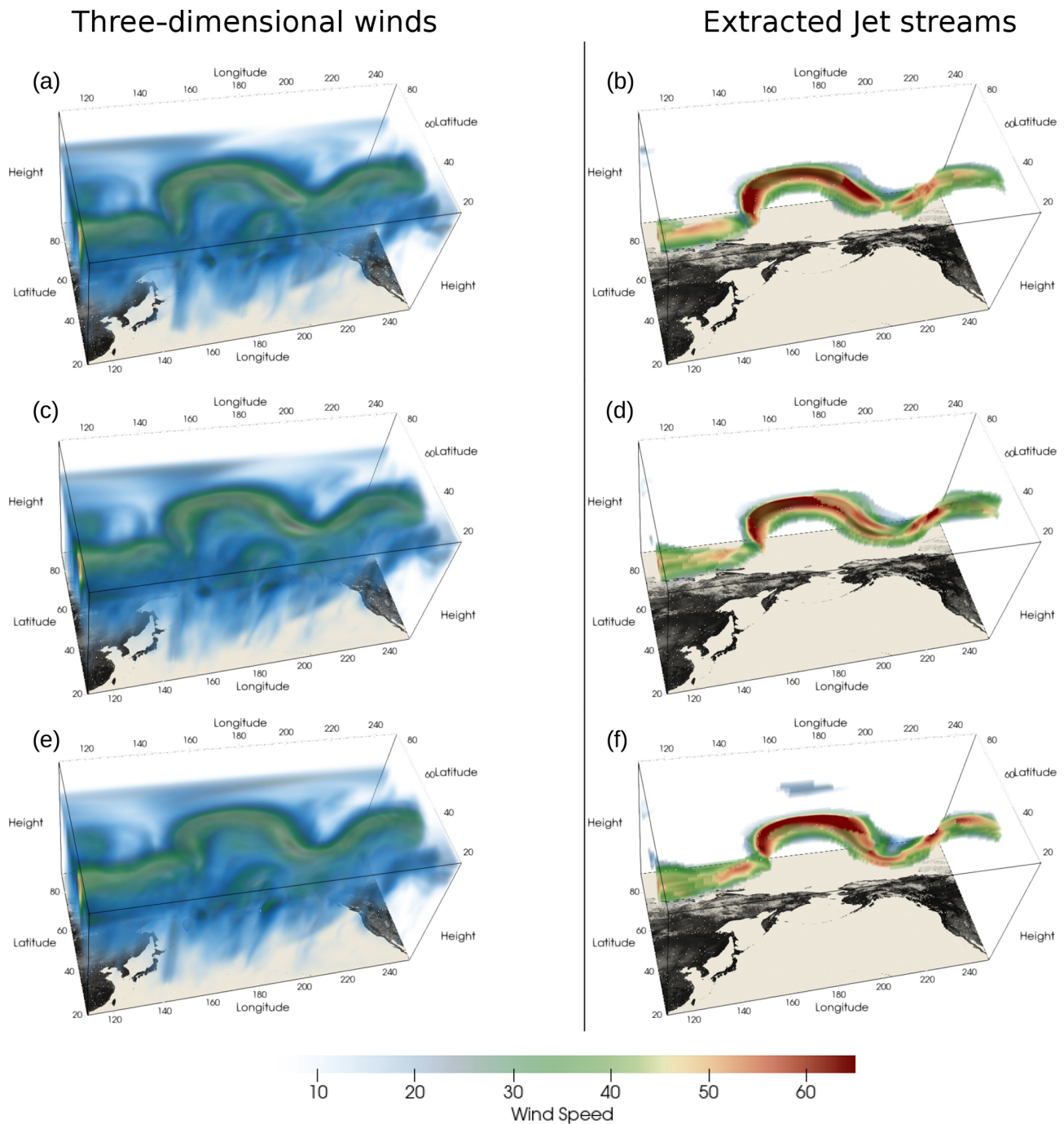
Three-dimensional winds

(a)

(c)

(e)

Extracted Jet streams

(b)

(d)

(f)

Figure 7. The 3D jet streams extracted using SCAFET. Magnitude of 3D wind speed for **(a)** 28 August 2022 at 12:00 UTC, **(c)** 28 August 2022 at 18:00, and **(e)** 29 August 2022 at 00:00 UTC. Extracted 3D jet streams for corresponding periods are shown in panels **(b)**, **(d)**, and **(f)**, respectively. The reader is encouraged to view the full video of these snapshots in the Supplement.

and streamline the algorithm for improved computational efficiency continue. One notable strength of SCAFET's design is its independence from data-set-specific posterior information when identifying features. Moreover, the shape-based thresholds used for detecting specific features remain consistent across various grids, data sets, and climatologies. Be-

tween these strengths and the full parallelization of the feature detection method, there are exciting possibilities for further development. This may eventually enable the algorithm to be used in operational feature identification and early-warning systems for extreme weather events.

Appendix A: Shape-based feature extraction on simple data sets

This section demonstrates how shape-based feature extraction can be performed on scalar fields represented by simple, idealized mathematical functions. It is intended to provide readers with more insights into the basic principles behind shape-based feature extraction and how it differs from other conventional methods. We have also tried to showcase some properties of shape-based feature extraction methods like its insensitivity to linear mean state trends.

A1 Application to 1D data sets

In this section, we draw an analogy between the use of SCAFET on a two-dimensional (2D) data set and shape-based feature extraction from a one-dimensional (1D) data set. Our intention is not to promote the use of shape-based extraction of features from 1D data sets but rather to provide readers with a fundamental understanding of this approach, along with its strengths and limitations.

For any differentiable curve C, the curvature is measured as the instantaneous rate of change in direction along the curve. Simply put, the curvature is measured as the rate of change in the unit tangent to the curve at any given point. An osculating circle can be used to intuitively represent the curvature of a surface or a curve (see Fig. A1). At any point P, the curvature, k is the reciprocal of the radius (R) of the circle. The sign of k determines if the curve has a concave or a convex curvature. More information and mathematical proof for these concepts can be found in any standard differential geometry textbook.

Following the derivation of the shape index (SI) for 2D data sets, we calculate the local shape of a function f using the shape parameter, defined as

$$K = \frac{2}{\pi}\tan^{-1}(f''). \tag{A1}$$

Values of K closer to 1 are identified as regions of local minima, while K values closer to -1 are regions of local maxima (black curve in Fig. A2). Depending on the magnitude of the function, one could adjust the value of K to obtain regions of local maxima (red caps in Fig. A2) and local minima (green caps in Fig. A2). The curvature of the function is insensitive to linear trends and mean state changes. This is evident, as the application of identical shape thresholds identifies the same regions of the curves as local maxima and minima, whether on the base curve (blue curve in Fig. A2) or on the same curve with an added linear trend (orange curve in Fig. A2). The values of K for both curves are represented by the black line in Fig. A2. Thus, the shape parameter can be used to identify the local minima and maxima from a 1D data set, despite background state changes.
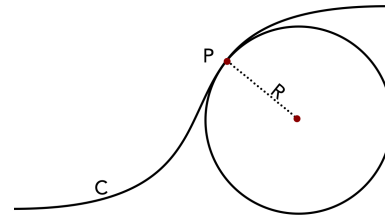

Figure A1. Schematic representation of curvature measurement of a curve C at point P. At P, the curvature is the reciprocal of the radius R of the osculating circle. In differential geometry, an osculating circle is defined as the circle passing through the point P and a pair of additional points infinitesimally close to P.

A2 Application of SCAFET to simple geostrophic motion

In this section, we apply SCAFET to a basic geostrophic rotational motion. The goal of this discussion is to illustrate how the shape-based extraction of 2D features differs from conventional methods. The calculation of the SI involves the computation of the two eigenvalues, k_1 and k_2, of the Hessian matrix for any gridded data set. As discussed in the previous section, the curvature measurement provided by k_1 and k_2 can be visualized as the reciprocal of the radius of two osculating circles that intersect orthogonally at a point on the surface. Large negative eigenvalues signify surfaces with strong convex curvature, while positive values correspond to troughs or cups.

To demonstrate the characteristics and advantages of feature detection based on the SI, let us consider a simple rotational wind field (see Fig. A3a vectors) given by

$$u_g = -\Omega y \tag{A2}$$
$$v_g = \Omega x, \tag{A3}$$

where Ω is a constant ($\Omega = 10^5 \, \text{rad s}^{-1}$), and x, y represents the grid. The geopotential height (h) of the field (see Fig. A3a shading) is used as our primary field in calculating the SI, which is computed as

$$h = \frac{\Omega f}{2g}(x^2 + y^2), \tag{A4}$$

where f and g are the Coriolis parameter and the acceleration due to gravity, respectively. The SI is calculated from the eigenvalues of the Hessian of h, using the formula

$$\text{SI}(k_1, k_2) = \frac{2}{\pi}\tan^{-1}\left[\frac{k_2 + k_1}{k_2 - k_1}\right], \tag{A5}$$

where the eigenvalues k_1 and k_2 are given by

$$k_{12} = \frac{f\zeta_g}{2g} \pm \sqrt{\left(\frac{f}{2g}\right)^2 - \left(\frac{f}{g}\right)^2 \frac{\partial v_g}{\partial x}\frac{\partial u_g}{\partial y} + \frac{\partial u_g}{\partial x}\frac{\partial v_g}{\partial y}}, \tag{A6}$$
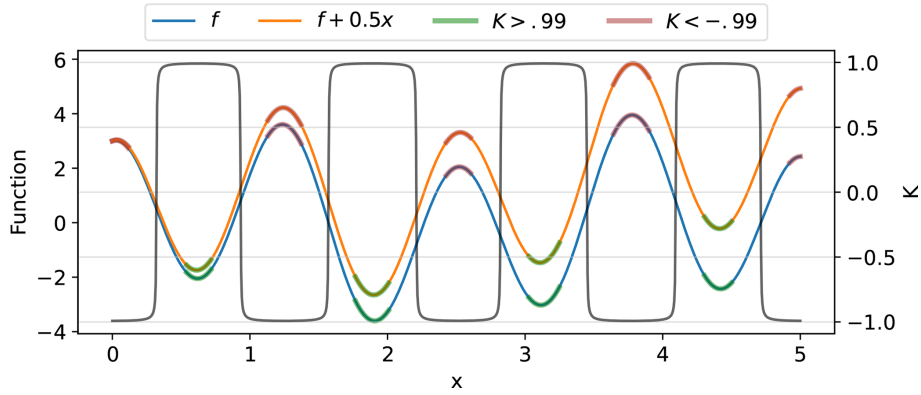
Figure A2. Comparison of shape extraction between a simple one-dimensional curve, given by $f = \sin 2x + 3\cos 5x$ (blue curve) and $f + 0.5x = \sin 2x + 3\cos 5x + 0.5x$ (orange; the blue curve with a linear trend). The left y axis shows the magnitude of both functions, and the right y axis indicates the values of the shape parameter (K). Note that value of K is the same for both functions. The green and red highlighting on the curves shows regions where $K > 0.99$ and $K < -0.99$, corresponding to regions of local maxima and minima, respectively.

where ζ_{g} is the geostrophic vorticity. This then gives SI as

$$\mathrm{SI}(k_1, k_2) = \frac{2}{\pi}\tan^{-1}\left[\frac{\zeta_{\mathrm{g}}}{-2\sqrt{\left(\frac{\zeta_{\mathrm{g}}}{2}\right)^2 - \frac{\partial v_{\mathrm{g}}}{\partial x}\frac{\partial u_{\mathrm{g}}}{\partial y} + \frac{\partial u_{\mathrm{g}}}{\partial x}\frac{\partial v_{\mathrm{g}}}{\partial y}}}\right]. \quad (A7)$$

A detailed derivation of the above equation can be found in Sect. B. By plugging in the values for the rotational motion, we get

$$\zeta_{\mathrm{g}} = \nabla^2 h = \Omega f/g \qquad (A8)$$

$$\frac{\partial u_{\mathrm{g}}}{\partial x} = \frac{\partial v_{\mathrm{g}}}{\partial y} = 0 \qquad (A9)$$

$$\frac{\partial v_{\mathrm{g}}}{\partial x}\frac{\partial u_{\mathrm{g}}}{\partial y} = \Omega^2. \qquad (A10)$$

Therefore,

$$\mathrm{SI} = \frac{2}{\pi}\tan^{-1}\left[\frac{\Omega^2}{-\sqrt{\Omega^2 - \Omega^2}}\right] = -1. \qquad (A11)$$

Thus, SCAFET classifies the whole domain with counterclockwise rotational motion as a trough with the SI ≈ -1, regardless of the absolute value of the field or Ω. In contrast, traditional methods that rely on thresholding the geopotential height would identify regions based on the chosen threshold of h, which would need to be adjusted depending on the mean (time) and background (space) state. Another common approach is to establish a threshold on the smallest eigenvalue, thus aiming to identify extreme features based on the curvature strength rather than the field's actual value. TempestExtremes (Ullrich and Zarzycki, 2017), a feature extraction framework discussed in the main text, follows this method to detect atmospheric rivers from gridded data sets. In the current example, this approach would correspond to setting a threshold on $f\Omega/g$. In other words, TempestExtremes would only identify a trough if the value of Ω exceeds the predetermined threshold. SCAFET, on the other hand, identifies the trough region as a trough regardless of the specific value of the field or Ω. This illustrates how feature extraction, using SI and other published methods, can yield different results, depending on the input data, as they focus on distinct properties of the field.

A3 Application of SCAFET to 3D fields

This section aims to demonstrate the detection of a cylindrical volume within a three-dimensional scalar field. To illustrate the effectiveness of the SI in identifying 3D structures embedded within scalar fields, we offer a straightforward example of how the SI can be used to isolate a cylinder embedded in a scalar field defined by $f = \sin(3X) + \cos(4Y)\cos(Z)$. It is worth mentioning that this specific problem bears significant similarities to the task of identifying 3D jet cores.

As explained in Sect. 4, a three-dimensional field provides us with three eigenvalues, thus satisfying the condition $k_1 \geq k_2 \geq k_3$. The SI can be computed using $\mathrm{SI}(k_1, k_2)$, $\mathrm{SI}(k_1, k_3)$, or $\mathrm{SI}(k_2, k_3)$. Setting a threshold of the SI > 0.375 effectively isolates the cylinder when using either $\mathrm{SI}(k_1, k_2)$ or $\mathrm{SI}(k_1, k_3)$ (see Fig. A4b–d). Between these two options, $\mathrm{SI}(k_1, k_2)$, which utilizes the two largest eigenvalues, imposes a more conservative criterion for identifying the embedded cylinder. The percentage of data identified as the cylinder is provided in the title of each plot in Fig. A4. Notably, employing $\mathrm{SI}(k_2, k_3)$ is not suitable, as it fails to isolate the desired cylinder shape effectively. The choice of using $\mathrm{SI}(k_1, k_2)$ is specifically tailored for extracting convex shapes or local maxima. Interestingly, to identify concave shapes or
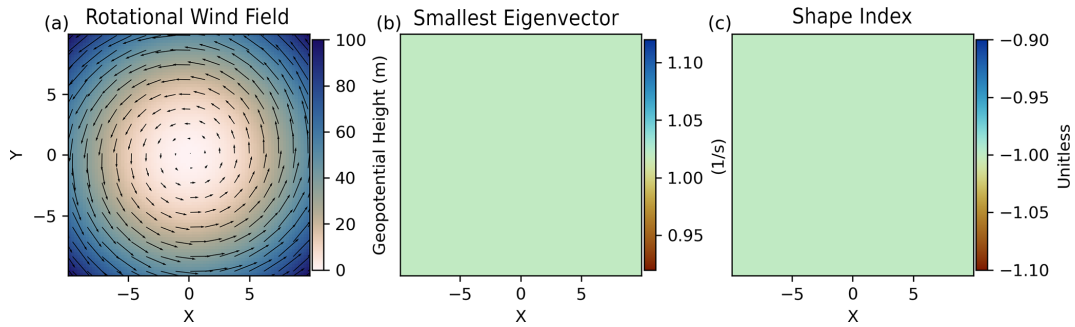
Figure A3. Comparison between two feature extraction techniques on an idealized example of rotational wind field. **(a)** The geopotential height (h) (shading) of the rotational wind field (arrows). h is defined as $\Omega f (x^2 + y^2)/2g$, where $f = 10^{-4} \, \text{s}^{-1}$, $g = 9.805 \, \text{m s}^{-1}$, and $\Omega = 10^5 \, \text{rad s}^{-1}$. **(b)** Magnitude of the smallest eigenvalue, derived from the equation as $f\Omega/g = 1.0199$, thus illustrating a uniform field as expected. **(c)** Value of the SI where SI $= -1$ throughout the domain, as expected.
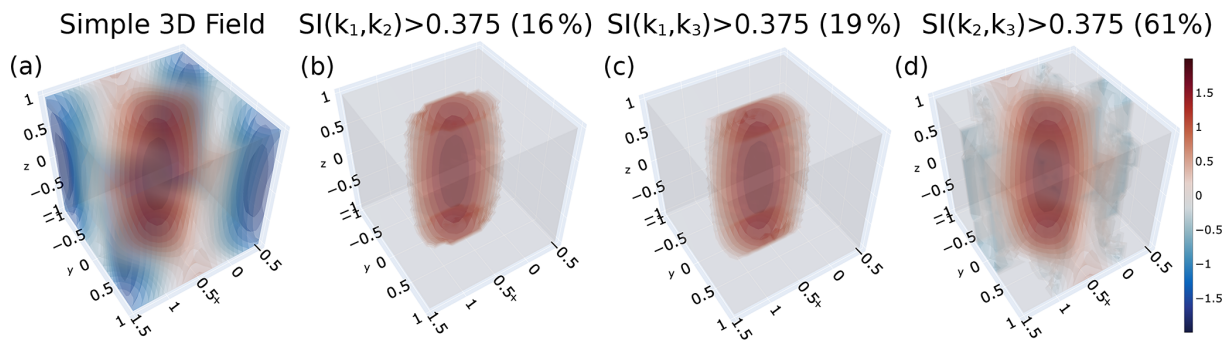
Figure A4. Various approaches for extraction of a 3D cylinder from a scalar field. **(a)** A simple scalar field represented by $\sin(3X) + \cos(4Y) * \cos(Z)$ is shown. **(b–d)** The extracted cylinders by applying the conditions in panel **(b)** SI$(k_1, k_2) > 0.375$, while in panel **(c)** SI$(k_1, k_3) > 0.375$ and panel **(d)** SI$(k_2, k_3) > 0.375$ are shown. The values enclosed in parentheses within the figure titles indicate the percentage of data that satisfies the respective conditions applied in each case.

local minima, one should utilize the SI derived from the two smallest eigenvalues, namely SI(k_2, k_3).

While the simple example presented here may not provide a comprehensive illustration of 3D feature detection, we hope that it encourages further fundamental research into 3D feature extraction to expand the capabilities of analysis and increase precision.

Appendix B: Derivation of shape index for geostrophic motion

The complete derivation of the SI for geostrophic wind fields is shown in this section. The result from the derivation is used in Sect. A.

Let h be the geopotential height at a certain level. The Hessian of h is given by

$$\mathcal{H}(h) = \begin{pmatrix} \frac{\partial^2 h}{\partial x^2} & \frac{\partial^2 h}{\partial x \partial y} \\ \frac{\partial^2 h}{\partial y \partial x} & \frac{\partial^2 h}{\partial y^2} \end{pmatrix}. \tag{B1}$$

The eigenvalues of the symmetric matrix $\mathcal{H}$ are calculated by solving the quadratic equation.

$$\left(\frac{\partial^2 h}{\partial x^2} - \lambda \right) \left(\frac{\partial^2 h}{\partial y^2} - \lambda \right) - \left(\frac{\partial^2 h}{\partial x \partial y} \right)^2 = 0, \tag{B2}$$

which can be expanded as

$$\lambda^2 - \lambda \left(\frac{\partial^2 h}{\partial x^2} + \frac{\partial^2 h}{\partial y^2} \right) + \frac{\partial^2 h}{\partial x^2} \cdot \frac{\partial^2 h}{\partial y^2} - \left(\frac{\partial^2 h}{\partial x \partial y} \right)^2 = 0 \tag{B3}$$

$$\lambda^2 - \lambda \nabla^2 h + \frac{\partial^2 h}{\partial x^2} \cdot \frac{\partial^2 h}{\partial y^2} - \left(\frac{\partial^2 h}{\partial x \partial y} \right)^2 = 0 \,. \tag{B4}$$

Note that the geostrophic vorticity (ζ_g) is defined as

$$\zeta_g = \frac{g}{f} \nabla^2 h. \tag{B5}$$

The geostrophic velocities are defined as

$$u_g = -\frac{g}{f} \frac{\partial h}{\partial y} = -\frac{\partial \psi}{\partial y} \tag{B6}$$

$$v_g = \frac{g}{f} \frac{\partial h}{\partial x} = \frac{\partial \psi}{\partial x} \,, \tag{B7}$$

where ψ is the geostrophic stream function. This implies

$$\frac{\partial^2 h}{\partial x^2} = \frac{f}{g}\frac{\partial^2 \psi}{\partial x^2} = \frac{f}{g}\frac{\partial v_g}{\partial x} \tag{B8}$$

$$\frac{\partial^2 h}{\partial y^2} = \frac{f}{g}\frac{\partial^2 \psi}{\partial y^2} = -\frac{f}{g}\frac{\partial u_g}{\partial y}. \tag{B9}$$

Adding the abovementioned relationships to Eq. (3) leads to the following:

$$\lambda^2 - \frac{\lambda f}{g}\zeta_g - \frac{f^2}{g^2}\frac{\partial v_g}{\partial x}\frac{\partial u_g}{\partial y} + \frac{f^2}{g^2}\frac{\partial u_g}{\partial x}\frac{\partial v_g}{\partial y}. \tag{B10}$$

By solving for λ, we get

$$\lambda_{12} = \frac{f\zeta_g}{2g} \pm \sqrt{\left(\frac{f}{2g}\right)^2 - \left(\frac{f}{g}\right)^2 \frac{\partial v_g}{\partial x}\frac{\partial u_g}{\partial y} + \frac{\partial u_g}{\partial x}\frac{\partial v_g}{\partial y}} \tag{B11}$$

$$\lambda_{12} = \frac{f}{g}\left[\frac{\zeta_g}{2} \pm \sqrt{\left(\frac{\zeta_g}{2}\right)^2 - \frac{\partial v_g}{\partial x}\frac{\partial u_g}{\partial y} + \frac{\partial u_g}{\partial x}\frac{\partial v_g}{\partial y}}\right]. \tag{B12}$$

Thus, the shape index for h is

$$SI = \frac{2}{\pi}\tan\left[\frac{\zeta_g}{-2\sqrt{\left(\frac{\zeta_g}{2}\right)^2 - \frac{\partial v_g}{\partial x}\frac{\partial u_g}{\partial y} + \frac{\partial u_g}{\partial x}\frac{\partial v_g}{\partial y}}}\right]. \tag{B13}$$

Code and data availability. The latest version of the Scalable Feature Extraction and Tracking (SCAFET) algorithm can be downloaded from https://github.com/nbarjun/SCAFET (last access: 20 December 2023). The version of the codes used for feature extraction and creating relevant figures in this work can be downloaded from https://doi.org/10.5281/zenodo.7767301 (Nellikkattil, 2023). A sample data set for the curvilinear SST data is also included in the repository. The directory also includes sample outputs for various features discussed in the article. The single-level ERA5 reanalysis data such as 10 m wind speed are obtained from https://doi.org/10.24381/cds.adbb2d47 (Hersbach et al., 2023a), while three-dimensional variables can be extracted from https://doi.org/10.24381/cds.bd0915c6 (Hersbach et al., 2023b). To see the exact codes used for downloading ERA5 data, readers could refer to the ERA5Data folder in the Zenodo repository. For any further details on code and data, feel free to contact the corresponding author.

Supplement. The supplement related to this article is available online at: https://doi.org/10.5194/gmd-17-301-2024-supplement.

Author contributions. ABN wrote and developed the software package. ABN and DL prepared the draft, with inputs from JYL. TAO'B and DL were involved in developing a mathematical framework for the algorithm. JEC provided input and guidance on the detection and tracking of tropical and extratropical cyclones. JYL, TAO'B, and JEC contributed equally to revising the paper.

References

Avila, L. A., Stewart, S. R., Berg, R., and Hagen, A. B.: Hurricane Dorian, Tech. rep., National Hurricane Center, https://www.nhc.noaa.gov/data/tcr/AL052019_Dorian.pdf (last access: 20 December 2023), 2020.

Balaji, V., Taylor, K. E., Juckes, M., Lawrence, B. N., Durack, P. J., Lautenschlager, M., Blanton, C., Cinquini, L., Denvil, S., Elkington, M., Guglielmo, F., Guilyardi, E., Hassell, D., Kharin, S.,

Kindermann, S., Nikonov, S., Radhakrishnan, A., Stockhause, M., Weigel, T., and Williams, D.: Requirements for a global data infrastructure in support of CMIP6, Geosci. Model Dev., 11, 3659–3680, https://doi.org/10.5194/gmd-11-3659-2018, 2018.

Bengtsson, L., Kanamitsu, M., Kållberg, P., and Uppala, S.: FGGE Research Activities at ECMWF, B. Am. Meteorol. Soc., 63, 277–303, https://doi.org/10.1175/1520-0477-63.3.277, 1982.

Bengtsson, L., Botzet, M., and Esch, M.: Hurricane-type vortices in a general circulation model, Tellus A, 47, 175–196, https://doi.org/10.1034/j.1600-0870.1995.t01-1-00003.x, 1995.

Biard, J. C. and Kunkel, K. E.: Automated detection of weather fronts using a deep learning neural network, Adv. Stat. Clim. Meteorol. Oceanogr., 5, 147–160, https://doi.org/10.5194/ascmo-5-147-2019, 2019.

Bowman, M. J.: Introduction and Historical Perspective, in: Oceanic Fronts in Coastal Processes, 2–5, Springer Berlin Heidelberg, https://doi.org/10.1007/978-3-642-66987-3_1, 1978.

Burston, R., Hodges, K., Astin, I., and Jayachandran, P. T.: Automated identification and tracking of polar-cap plasma patches at solar minimum, Ann. Geophys., 32, 197–206, https://doi.org/10.5194/angeo-32-197-2014, 2014.

Canny, J.: A Computational Approach to Edge Detection, IEEE T. Pattern Anal. Mach. Int., PAMI-8, 679–698, https://doi.org/10.1109/tpami.1986.4767851, 1986.

Castelao, R. M., Mavor, T. P., Barth, J. A., and Breaker, L. C.: Sea surface temperature fronts in the California Current System from geostationary satellite observations, J. Geophys. Res., 111, https://doi.org/10.1029/2006jc003541, 2006.

Chu, J.-E., Lee, S.-S., Timmermann, A., Wengel, C., Stuecker, M. F., and Yamaguchi, R.: Reduced tropical cyclone densities and ocean effects due to anthropogenic greenhouse warming, Sci. Adv., 6, eabd5109, https://doi.org/10.1126/sciadv.abd5109, 2020.

Clayton, S., Nagai, T., and Follows, M. J.: Fine scale phytoplankton community structure across the Kuroshio Front, J. Plankton Res., 36, 1017–1030, https://doi.org/10.1093/plankt/fbu020, 2014.

Clayton, S., Palevsky, H. I., Thompson, L., and Quay, P. D.: Synoptic Mesoscale to Basin Scale Variability in Biological Productivity and Chlorophyll in the Kuroshio Extension Region, J. Geophys. Res.-Oceans, 126, e2021JC017782, https://doi.org/10.1029/2021jc017782, 2021.

Coumou, D., Petoukhov, V., Rahmstorf, S., Petri, S., and Schellnhuber, H. J.: Quasi-resonant circulation regimes and hemispheric synchronization of extreme weather in boreal summer, P. Natl. Acad. Sci. USA, 111, 12331–12336, https://doi.org/10.1073/pnas.1412797111, 2014.

Dixon, M. and Wiener, G.: TITAN: Thunderstorm Identification, Tracking, Analysis, and Nowcasting – A Radar-based Methodology, J. Atmos. Ocean. Tech., 10, 785–797, https://doi.org/10.1175/1520-0426(1993)010<0785:ttitaa>2.0.co;2, 1993.

Emanuel, K.: Tropical Cyclones, Annu. Rev. Earth Planet. Sci., 31, 75–104, https://doi.org/10.1146/annurev.earth.31.100901.141259, 2003.

Encyclopaedia, B.: Cyclone, https://www.britannica.com/science/cyclone-meteorology (last access: 20 December 2023), 2022.

Fedorov, K. N.: The physical nature and structure of oceanic fronts, Coastal and Estuarine Studies, Springer, New York, NY, 1986.

Guan, B. and Waliser, D. E.: Detection of atmospheric rivers: Evaluation and application of an algorithm for global studies, J. Geophys. Res.-Atmos., 120, 12514–12535, https://doi.org/10.1002/2015jd024257, 2015.

Guo, H.: Big Earth data: A new frontier in Earth and information sciences, Big Earth Data, 1, 4–20, https://doi.org/10.1080/20964471.2017.1403062, 2017.

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

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J.-N.: ERA5 hourly data on single levels from 1940 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS) [data set], https://doi.org/10.24381/cds.adbb2d47 (last access: 10 January 2023), 2023a.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J.-N.: ERA5 hourly data on pressure levels from 1940 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS), https://doi.org/10.24381/cds.bd0915c6, 2023b.

Hewson, T. D.: Objective fronts, Meteorol. Appl., 5, 37–65, https://doi.org/10.1017/s1350482798000553, 1998.

Hodges, K. I.: A General Method for Tracking Analysis and Its Application to Meteorological Data, Mon. Weather Rev., 122, 2573–2586, https://doi.org/10.1175/1520-0493(1994)122<2573:agmfta>2.0.co;2, 1994.

Hodges, K. I.: Feature Tracking on the Unit Sphere, Mon. Weather Rev., 123, 3458–3465, https://doi.org/10.1175/1520-0493(1995)123<3458:ftotus>2.0.co;2, 1995.

Hodges, K. I., Lee, R. W., and Bengtsson, L.: A Comparison of Extratropical Cyclones in Recent Reanalyses ERA-Interim, NASA MERRA, NCEP CFSR, and JRA-25, J. Climate, 24, 4888–4906, https://doi.org/10.1175/2011jcli4097.1, 2011.

Hogg, A. M. C., Killworth, P. D., Blundell, J. R., and Dewar, W. K.: Mechanisms of Decadal Variability of the Wind-Driven Ocean Circulation, J. Phys. Oceanogr., 35, 512–531, https://doi.org/10.1175/jpo2687.1, 2005.

Horn, M., Walsh, K., Zhao, M., Camargo, S. J., Scoccimarro, E., Murakami, H., Wang, H., Ballinger, A., Kumar, A., Shaevitz, D. A., Jonas, J. A., and Oouchi, K.: Tracking Scheme Dependence of Simulated Tropical Cyclone Response to Idealized Climate Simulations, J. Climate, 27, 9197–9213, https://doi.org/10.1175/jcli-d-14-00200.1, 2014.

Hurley, J. V. and Boos, W. R.: A global climatology of monsoon low-pressure systems, Q. J. Roy. Meteor. Soc., 141, 1049–1064, https://doi.org/10.1002/qj.2447, 2014.

Karmakar, N., Boos, W. R., and Misra, V.: Influence of Intraseasonal Variability on the Development of Monsoon

Depressions, Geophys. Res. Lett., 48, e2020GL090425, https://doi.org/10.1029/2020gl090425, 2021.

Kern, M., Hewson, T., Sadlo, F., Westermann, R., and Rautenhaus, M.: Robust Detection and Visualization of Jet-Stream Core Lines in Atmospheric Flow, IEEE T. Vis. Comput. Gr., 24, 893–902, https://doi.org/10.1109/tvcg.2017.2743989, 2018.

Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., and Neumann, C. J.: The International Best Track Archive for Climate Stewardship (IBTrACS), B. Am. Meteorol. Soc., 91, 363–376, https://doi.org/10.1175/2009bams2755.1, 2010.

Knapp, K. R., Diamond, H. J., Kossin, J. P., Kruk, M. C., and Schreck, C. J.: International Best Track Archive for Climate Stewardship (IBTrACS) Project, Version 4, https://doi.org/10.25921/82TY-9E16, 2018.

Knutson, T. R., McBride, J. L., Chan, J., Emanuel, K., Holland, G., Landsea, C., Held, I., Kossin, J. P., Srivastava, A. K., and Sugi, M.: Tropical cyclones and climate change, Nat. Geosci., 3, 157–163, https://doi.org/10.1038/ngeo779, 2010.

Koch, P., Wernli, H., and Davies, H. C.: An event-based jet-stream climatology and typology, Int. J. Climatol., 26, 283–301, https://doi.org/10.1002/joc.1255, 2006.

Koenderink, J. J. and van Doorn, A. J.: Surface shape and curvature scales, Image Vision Comput., 10, 557–564, https://doi.org/10.1016/0262-8856(92)90076-f, 1992.

Kretschmer, M., Coumou, D., Donges, J. F., and Runge, J.: Using Causal Effect Networks to Analyze Different Arctic Drivers of Midlatitude Winter Circulation, J. Climate, 29, 4069–4081, https://doi.org/10.1175/jcli-d-15-0654.1, 2016.

Legeckis, R.: A survey of worldwide sea surface temperature fronts detected by environmental satellites, J. Geophys. Res., 83, 4501, https://doi.org/10.1029/jc083ic09p04501, 1978.

Limbach, S., Schömer, E., and Wernli, H.: Detection, tracking and event localization of jet stream features in 4-D atmospheric data, Geosci. Model Dev., 5, 457–470, https://doi.org/10.5194/gmd-5-457-2012, 2012.

Lindeberg, T.: Scale Selection, in: Computer Vision, 701–713, Springer US, https://doi.org/10.1007/978-0-387-31439-6_242, 2014.

Lora, J. M., Shields, C. A., and Rutz, J. J.: Consensus and Disagreement in Atmospheric River Detection: ARTMIP Global Catalogues, Geophys. Res. Lett., 47, e2020GL089302, https://doi.org/10.1029/2020gl089302, 2020.

Marr, B.: Big data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance, JohnWiley & Sons, Nashville, TN, 256 pp., ISBN 978-1-118-96583-2, 2015.

Mendelsohn, R., Emanuel, K., Chonabayashi, S., and Bakkensen, L.: The impact of climate change on global tropical cyclone damage, Nat. Clim. Change, 2, 205–209, https://doi.org/10.1038/nclimate1357, 2012.

Molnos, S., Mamdouh, T., Petri, S., Nocke, T., Weinkauf, T., and Coumou, D.: A network-based detection scheme for the jet stream core, Earth Syst. Dynam., 8, 75–89, https://doi.org/10.5194/esd-8-75-2017, 2017.

Nagai, T. and Clayton, S.: Nutrient interleaving below the mixed layer of the Kuroshio Extension Front, Ocean Dynam., 67, 1027–1046, https://doi.org/10.1007/s10236-017-1070-3, 2017.

Nash, D., Waliser, D., Guan, B., Ye, H., and Ralph, F. M.: The Role of Atmospheric Rivers in Extratropical and Po-

lar Hydroclimate, J. Geophys. Res.-Atmos., 123, 6804–6821, https://doi.org/10.1029/2017jd028130, 2018.

Nellikkattil, A. B.: Scalable Feature Extraction and Tracking (SCAFET): A general framework for feature extraction from large climate datasets, Zenodo [code and data set], https://doi.org/10.5281/zenodo.7767301, 2023.

Nellikkattil, A. B., Lee, J.-Y., Guan, B., Timmermann, A., Lee, S.-S., Chu, J.-E., and Lemmon, D.: Increased amplitude of atmospheric rivers and associated extreme precipitation in ultra-high-resolution greenhouse warming simulations, Commun. Earth Environ., 4, 1–11, https://doi.org/10.1038/s43247-023-00963-7, 2023.

Neu, U., Akperov, M. G., Bellenbaum, N., Benestad, R., Blender, R., Caballero, R., Cocozza, A., Dacre, H. F., Feng, Y., Fraedrich, K., Grieger, J., Gulev, S., Hanley, J., Hewson, T., Inatsu, M., Keay, K., Kew, S. F., Kindem, I., Leckebusch, G. C., Liberato, M. L. R., Lionello, P., Mokhov, I. I., Pinto, J. G., Raible, C. C., Reale, M., Rudeva, I., Schuster, M., Simmonds, I., Sinclair, M., Sprenger, M., Tilinina, N. D., Trigo, I. F., Ulbrich, S., Ulbrich, U., Wang, X. L., and Wernli, H.: IMILAST: A Community Effort to Intercompare Extratropical Cyclone Detection and Tracking Algorithms, B. Am. Meteorol. Soc., 94, 529–547, https://doi.org/10.1175/bams-d-11-00154.1, 2013.

O'Brien, T. A., Wehner, M. F., Payne, A. E., Shields, C. A., Rutz, J. J., Leung, L.-R., Ralph, F. M., Collow, A., Gorodetskaya, I., Guan, B., Lora, J. M., McClenny, E., Nardi, K. M., Ramos, A. M., Tomé, R., Sarangi, C., Shearer, E. J., Ullrich, P. A., Zarzycki, C., Loring, B., Huang, H., Inda-Díaz, H. A., Rhoades, A. M., and Zhou, Y.: Increases in Future AR Count and Size: Overview of the ARTMIP Tier 2 CMIP5/6 Experiment, J. Geophys. Res.-Atmos., 127, e2021JD036013, https://doi.org/10.1029/2021jd036013, 2022.

Overpeck, J. T., Meehl, G. A., Bony, S., and Easterling, D. R.: Climate Data Challenges in the 21st Century, Science, 331, 700–702, https://doi.org/10.1126/science.1197869, 2011.

Petoukhov, V., Rahmstorf, S., Petri, S., and Schellnhuber, H. J.: Quasiresonant amplification of planetary waves and recent Northern Hemisphere weather extremes, P. Natl. Acad. Sci. USA, 110, 5336–5341, https://doi.org/10.1073/pnas.1222000110, 2013.

Pinheiro, H. R., Hodges, K. I., Gan, M. A., and Ferreira, N. J.: A new perspective of the climatological features of upper-level cut-off lows in the Southern Hemisphere, Clim. Dynam., 48, 541–559, https://doi.org/10.1007/s00382-016-3093-8, 2016.

Post, F. H., Vrolijk, B., Hauser, H., Laramee, R. S., and Doleisch, H.: The State of the Art in Flow Visualisation: Feature Extraction and Tracking, Computer Graphics Forum, 22, 775–792, https://doi.org/10.1111/j.1467-8659.2003.00723.x, 2003.

Prabhat, Rübel, O., Byna, S., Wu, K., Li, F., Wehner, M., and Bethel, W.: TECA: A Parallel Toolkit for Extreme Climate Analysis, Proced. Comput. Sci., 9, 866–876, https://doi.org/10.1016/j.procs.2012.04.093, 2012.

Priestley, M. D. K., Ackerley, D., Catto, J. L., Hodges, K. I., McDonald, R. E., and Lee, R. W.: An Overview of the Extratropical Storm Tracks in CMIP6 Historical Simulations, J. Climate, 33, 6315–6343, https://doi.org/10.1175/jcli-d-19-0928.1, 2020.

Ralph, F. M., Dettinger, M. D., Cairns, M. M., Galarneau, T. J., and Eylander, J.: Defining "Atmospheric River": How the Glossary of

Meteorology Helped Resolve a Debate, B. Am. Meteorol. Soc., 99, 837–839, https://doi.org/10.1175/bams-d-17-0157.1, 2018.

Ranson, M., Kousky, C., Ruth, M., Jantarasami, L., Crimmins, A., and Tarquinio, L.: Tropical and extratropical cyclone damages under climate change, Clim. Change, 127, 227–241, https://doi.org/10.1007/s10584-014-1255-4, 2014.

Rutz, J. J., Steenburgh, W. J., and Ralph, F. M.: Climatological Characteristics of Atmospheric Rivers and Their Inland Penetration over the Western United States, Mon. Weather Rev., 142, 905–921, https://doi.org/10.1175/mwr-d-13-00168.1, 2014.

Schultz, D. M., Bosart, L. F., Colle, B. A., Davies, H. C., Dearden, C., Keyser, D., Martius, O., Roebber, P. J., Steenburgh, W. J., Volkert, H., and Winters, A. C.: Extratropical Cyclones: A Century of Research on Meteorology's Centerpiece, Meteorol. Monogr., 59, 16.1–16.56, https://doi.org/10.1175/amsmonographs-d-18-0015.1, 2019.

Sellars, S., Nguyen, P., Chu, W., Gao, X., lin Hsu, K., and Sorooshian, S.: Computational Earth Science: Big Data Transformed Into Insight, Eos, Transactions American Geophysical Union, 94, 277–278, https://doi.org/10.1002/2013eo320001, 2013.

Shields, C. A., Rutz, J. J., Leung, L.-Y., Ralph, F. M., Wehner, M., Kawzenuk, B., Lora, J. M., McClenny, E., Osborne, T., Payne, A. E., Ullrich, P., Gershunov, A., Goldenson, N., Guan, B., Qian, Y., Ramos, A. M., Sarangi, C., Sellars, S., Gorodetskaya, I., Kashinath, K., Kurlin, V., Mahoney, K., Muszynski, G., Pierce, R., Subramanian, A. C., Tome, R., Waliser, D., Walton, D., Wick, G., Wilson, A., Lavers, D., Prabhat, Collow, A., Krishnan, H., Magnusdottir, G., and Nguyen, P.: Atmospheric River Tracking Method Intercomparison Project (ARTMIP): project goals and experimental design, Geosci. Model Dev., 11, 2455–2474, https://doi.org/10.5194/gmd-11-2455-2018, 2018.

Small, R. J., Bacmeister, J., Bailey, D., Baker, A., Bishop, S., Bryan, F., Caron, J., Dennis, J., Gent, P., ming Hsu, H., Jochum, M., Lawrence, D., Muñoz, E., diNezio, P., Scheitlin, T., Tomas, R., Tribbia, J., heng Tseng, Y., and Vertenstein, M.: A new synoptic scale resolving global climate simulation using the Community Earth System Model, J. Adv. Model. Earth Sy., 6, 1065–1094, https://doi.org/10.1002/2014ms000363, 2014.

Strong, C. and Davis, R. E.: Winter jet stream trends over the Northern Hemisphere, Q. J. Roy. Meteor. Soc., 133, 2109–2115, https://doi.org/10.1002/qj.171, 2007.

Studholme, J., Fedorov, A. V., Gulev, S. K., Emanuel, K., and Hodges, K.: Poleward expansion of tropical cyclone latitudes in warming climates, Nat. Geosci., 15, 14–28, https://doi.org/10.1038/s41561-021-00859-1, 2021.

Torres-Alavez, J. A., Glazer, R., Giorgi, F., Coppola, E., Gao, X., Hodges, K. I., Das, S., Ashfaq, M., Reale, M., and Sines, T.: Future projections in tropical cyclone activity over multiple CORDEX domains from RegCM4 CORDEX-CORE simulations, Clim. Dynam., 57, 1507–1531, https://doi.org/10.1007/s00382-021-05728-6, 2021.

Ulbrich, U., Leckebusch, G. C., and Pinto, J. G.: Extra-tropical cyclones in the present and future climate: a review, Theor. Appl. Climatol., 96, 117–131, https://doi.org/10.1007/s00704-008-0083-8, 2009.

Ullrich, P. A. and Zarzycki, C. M.: TempestExtremes: a framework for scale-insensitive pointwise feature tracking on unstructured grids, Geosci. Model Dev., 10, 1069–1090, https://doi.org/10.5194/gmd-10-1069-2017, 2017.

Ullrich, P. A., Zarzycki, C. M., McClenny, E. E., Pinheiro, M. C., Stansfield, A. M., and Reed, K. A.: TempestExtremes v2.1: a community framework for feature detection, tracking, and analysis in large datasets, Geosci. Model Dev., 14, 5023–5048, https://doi.org/10.5194/gmd-14-5023-2021, 2021.

van Genderen, J., Goodchild, M. F., Guo, H., Yang, C., Nativi, S., Wang, L., and Wang, C.: Digital Earth Challenges and Future Trends, in: Manual of Digital Earth, 811–827, Springer Singapore, https://doi.org/10.1007/978-981-32-9915-3_26, 2019.

Vishnu, S., Boos, W. R., Ullrich, P. A., and O'Brien, T. A.: Assessing Historical Variability of South Asian Monsoon Lows and Depressions With an Optimized Tracking Algorithm, J. Geophys. Res.-Atmos., 125, e2020JD032977, https://doi.org/10.1029/2020jd032977, 2020.

Vitart, F., Anderson, J. L., and Stern, W. F.: Simulation of Interannual Variability of Tropical Storm Frequency in an Ensemble of GCM Integrations, J. Climate, 10, 745–760, https://doi.org/10.1175/1520-0442(1997)010<0745:soivot>2.0.co;2, 1997.

Waliser, D. and Guan, B.: Extreme winds and precipitation during landfall of atmospheric rivers, Nat. Geosci., 10, 179–183, https://doi.org/10.1038/ngeo2894, 2017.

Wallace, J. M. and Hobbs, P. V.: Atmospheric science: An Introductory Survey, Academic Press, San Diego, CA, 500 pp.,ISBN 978-0127329505, 1977.

Walsh, K. J., McBride, J. L., Klotzbach, P. J., Balachandran, S., Camargo, S. J., Holland, G., Knutson, T. R., Kossin, J. P., cheung Lee, T., Sobel, A., and Sugi, M.: Tropical cyclones and climate change, WIREs Clim. Change, 7, 65–89, https://doi.org/10.1002/wcc.371, 2015.

Woodruff, J. D., Irish, J. L., and Camargo, S. J.: Coastal flooding by tropical cyclones and sea-level rise, Nature, 504, 44–52, https://doi.org/10.1038/nature12855, 2013.

Xi, J., Wang, Y., Feng, Z., Liu, Y., and Guo, X.: Variability and Intensity of the Sea Surface Temperature Front Associated With the Kuroshio Extension, Front. Marine Sci., 9, 836469, https://doi.org/10.3389/fmars.2022.836469, 2022.

Xu, G., Ma, X., Chang, P., and Wang, L.: Image-processing-based atmospheric river tracking method version 1 (IPART-1), Geosci. Model Dev., 13, 4639–4662, https://doi.org/10.5194/gmd-13-4639-2020, 2020.

Yang, C., Huang, Q., Li, Z., Liu, K., and Hu, F.: Big Data and cloud computing: innovation opportunities and challenges, Int. J. Dig. Earth, 10, 13–53, https://doi.org/10.1080/17538947.2016.1239771, 2016.

Yoder, J. A., Ackleson, S. G., Barber, R. T., Flament, P., and Balch, W. M.: A line in the sea, Nature, 371, 689–692, https://doi.org/10.1038/371689a0, 1994.

Zarzycki, C. M. and Ullrich, P. A.: Assessing sensitivities in algorithmic detection of tropical cyclones in climate data, Geophys. Res. Lett., 44, 1141–1149, https://doi.org/10.1002/2016gl071606, 2017.

Zhao, M.: Simulations of Atmospheric Rivers, Their Variability, and Response to Global Warming Using GFDL's New High-Resolution General Circulation Model, J. Climate, 33, 10287–10303, https://doi.org/10.1175/jcli-d-20-0241.1, 2020.