Scalable Feature Extraction and Tracking (SCAFET): A general framework for feature extraction from large climate datasets

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, South Korea, 46241
[2]Department of Climate System, Pusan National University, Busan, Republic of Korea, 46241
[3]Pusan National University, Busan, Rep. of Korea, 46241
[4]The American Association for the Advancement of Science, Science and Technology Policy Fellowship Program, Washington D.C., United States
[5]Department of Earth and Atmospheric Sciences, Indiana University Bloomington, Indiana, USA 47403
[6]Climate and Ecosystem Sciences Division, Lawrence Berkeley National Lab, Berkeley, USA 95720
[7]Research Center for Climate Sciences, Pusan National University, Busan, Republic of Korea, 46241
[8]Low-Carbon and Climate Impact Research Centre, School of Energy and Environment, City University of Hong Kong, Hong Kong, China

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

Abstract. This study describes a generalized computational mathematical framework, Scalable Feature Extraction and Tracking (SCAFET) to extract and track features from large climate datasets. SCAFET utilizes novel shape-based metrics that can ~~efficiently~~ identify and compare features from different mean states, datasets, and between distinct regions. Features of interest such as atmospheric rivers, tropical and extratropical cyclones, jet streams, etc. are extracted by segmenting the data based on a scale-independent bounded variable called shape index (SI). SI gives a quantitative measurement of the local geometric shape of the field with respect to its surroundings. Compared with 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 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 ~~from the ERA5 reanalysis dataset~~ to show how the algorithm ~~extracts both locations~~ identifies and tracks both nodes and areas from climate datasets. The extraction of sea surface temperature fronts exemplifies how SCAFET effectively handles curvilinear grids. Lastly, jet streams are extracted to demonstrate how the algorithm can also detect ~~3D features.~~ three-dimensional features. As a generalized framework, SCAFET can be implemented to extract and track ~~most~~ many weather and climate features across scales, grids, and dimensions.

1 Introduction

The amount of climate data is growing exponentially owing to rapid expansions in both observational capabilities and computational power, driven ~~by the need to observe and simulate ever-higher resolutions~~ 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~~

cloud-resolving and large ensemble simulations leads not just to increased volume but also to inflated velocity, variety, veracity, and value (5Vs) ~~of climate data~~ (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, ~~a daunting~~ an onerous task. Although these ~~features of interest~~ climate phenomena influence regional and global weather and climate with immense societal, economic, and ecological impacts, the amount of data representing these events and features ~~would be~~ is a small percentage of the whole simulation. ~~Thus, extraction of features not only enables us to focus our analysis on high-impact rare events but can also considerably reduce~~ Feature extraction considerably reduces the amount of data that needs to be stored, improving computational efficiency in ~~analysing~~ 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~~and various parameterizations (Sellars et al., 2013). Efficient and reliable extraction of these features is thus~~ , improving our understanding about 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 ~~informed~~ 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 ~~of and discrepancies between these methods are linked to~~ and discrepancies in these methods arise from the somewhat arbitrary choice of physical thresholds in relation to the ~~underlying spatio-temporal~~ 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, ~~timespan~~time span, and dataset used. Validation ~~then unfortunately comes~~ can then unfortunately come down to the intelligent but subjective human eye, or in other words tuning ~~a~~ 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 ~~is applicable~~ applies to different climate models and spans multiple mean states and model scenarios is not ~~straight forward. Even within the same model, a particular choice of threshold may be suitable for one region but not for another, given varying regional characteristics and topography. To account for these inter and intra-model discrepancies, thresholds are often calculated from the model- and/or simulation-specific distribution of basic climate variable fields~~ 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)~~or~~ , integrated water vapor transport (IVT) for ARs (e.g., Guan and Waliser, 2015)~~. Thus, before the actual detection process is applied~~, 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 dataset 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 entire~~ preprocess large, representative datasets just to calculate reasonable thresholds~~that will allow for comparison within and between models. This process becomes increasingly infeasible~~. While some detection methods have done well to streamline their algorithms to reduce total runtime, the process of posterior threshold calculation for higher resolutions and large ensemble ~~data sets~~datasets inherently becomes increasingly less efficient, highlighting the need ~~for a method of feature extraction that is not empirically derived and thus less sensitive to the climate mean state~~to develop feature extraction methods that do not use posterior assumptions.

Aside from the ~~inter and intra-model discrepancies that arise from detecting features in present and historical model simulations, applying empirical present thresholds to detect features in~~ sensitivity of feature detection to inter-model and inter-simulation differences, feature detection is further complicated when trying to detect and compare features between present and future climate change scenarios ~~is further untenable~~ as the underlying ~~spatio-temporal~~ 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 ~~(CO_2)~~ 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)~~ (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~~ 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 datasets spanning different mean states, many of these methods were developed for detecting specific rather than general features.

~~In this study we introduce a novel method, Scalable Feature Extraction and Tracking (SCAFET), which is a general framework to detect and track features of various shapes, scales, and intensities. Simply put, SCAFET uses the curvature of a given scalar field to identify emergent shapes that correspond with distinct features of interest. The shape is calculated as a finite, bounded, and scale-independent quantity and can be tuned depending on the desired phenomenon. As this tuning relies on shape-based rather than physical thresholds (see), the characteristics of the detected features are less sensitive to spatio-temporal and mean state variances. This also makes the feature extraction fully parallelizable along the time dimension, as the detection is carried out independent of the time information. SCAFET utilizes some of the modern python packages like *xarray* to handle NetCDF files and *dask* to parallelize the detection process in any machine with multiple cores. The code for this framework is fully open-source and written in Python in an easy-to-use package so that even beginner-level Python users can easily implement the algorithm.~~

The need for a general framework ~~in~~ for extracting and tracking features from large climate datasets has been raised in various climate science communities for the last several decades. In a pioneer study, ~~Hodges developed a three-step~~ Hodges (1994) developed a general framework for extracting and tracking features from meteorological datasets ~~. The first step is segmentation, where~~ in three steps: segmentation, filtering, and tracking. In the segmentation step, the field is split into distinct regions by applying a threshold and ~~then labelling~~ defining each of the connected regions as an object. ~~Later, feature nodes are extracted by filtering out regions outlined in the first step~~ 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 ~~dynamical~~ further analysis. This framework was ~~developed further~~ further developed for cyclones, storm tracks, convective systems, ocean eddies, monsoon depressions, ~~and more~~ 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 threshold-dependent detection methods and algorithms into a comprehensive software package that was parallelized to make the algorithms more suitable for large datasets (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), 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~~ meso- and synoptic-scale processes and their ~~contributions~~ connections to long-term climate ~~trends. Further discussion on the differences between SCAFET~~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 ~~other detection algorithms can be found in . Even in the context of these recent advancements, SCAFET aims to upgrade the detection process to be~~ intra-model variation, making it less sensitive to ~~the physical thresholds used while presenting a novel shape-based approach 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. Firstly, 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. Secondly, 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 for a simple working example).

The novelty of SCAFET compared to pre-existing methods lies in ~~the use of a comprehensive mathematical framework for extracting different features~~ 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 ~~is applicable~~ applies to both rectilinear and curvilinear grids and can also be extended to detect three-dimensional (3D) features. ~~In a nutshell~~Even in the context of recent advancements in feature extraction such as Tempest Extremes and TECA, SCAFET is ~~mathematically comprehensive, computationally~~ a comprehensive, efficient, and easily 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 Appendix A.

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

2 Description of Scalable Feature Extraction and Tracking

SCAFET ~~follows the same three processes as discussed by Hodges~~ adopts the same three-step approach as outlined by Hodges (1994) -Segmentation ~~, Filtering~~ (yellow boxes in Figure 1) , Filtering (orange boxes in Figure 1), and Tracking (~~yellow~~ green boxes in Figure 1). ~~Before starting the this process, SCAFET is initialized with~~ However, before commencing these steps, SCAFET requires initialization with essential information describing the datasets and the ~~type of~~ specific feature to be extracted (as indicated by blue boxes in Figure 1). ~~Primary inputs includes the following.~~ The key inputs for this initialization include the following:

– **~~A primary~~ Primary field** (ϕ_p): This is a gridded dataset in which the ~~feature to be extracted is most clearly visible.~~ 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 are required for calculating derivatives of the basic field and identifying landfalling locations.

– **~~Object properties~~Feature Properties**: The algorithm requires information on the ~~feature properties~~ properties of the target feature. This includes ~~approximate~~ estimated spatial scale, shape, eccentricity (for 2D features only), minimum

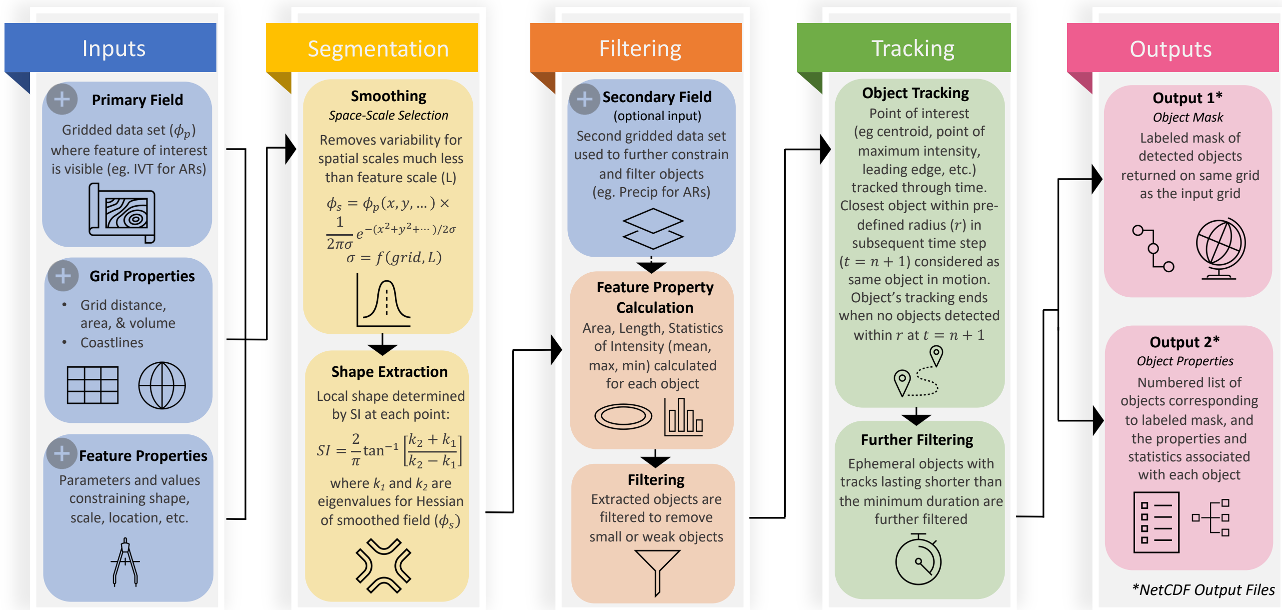

Figure 1. ~~The overall~~ Overall schematic ~~,~~ of SCAFET workflow ~~,~~ and components~~of SCAFET~~. ~~The inputs, processes/calculations, and outputs of~~ Inputs to the algorithm are ~~shown~~ depicted in blue, while the algorithm's outputs are shown in pink boxes. Processes related to the segmentation step are highlighted in yellow boxes, ~~and pink~~ whereas the orange boxes ~~respectively~~represent the filtering processes. The ~~arrows in~~ tracking step is denoted by green boxes. Arrows on the periphery of the boxes ~~represent~~ illustrate the ~~workflow~~ flow of the algorithm. Each section is ~~explained~~ elaborated upon in detail ~~in~~ within the text.

length, minimum area, minimum volume (for 3D feature only), minimum duration, and maximum distance per time step. ~~(see and for examples)~~

~~As the~~ In the SCAFET scheme, segmentation, filtering, and tracking are ~~mutually independent in the SCAFET scheme, users can replace any of them and still run~~ developed and coded as separate Python libraries. This design allows users to substitute any of these components with their own methods while still being able to execute the algorithm. ~~After implementing~~ Once all three steps ~~, two outputs are obtained: one describing~~ have been executed, the algorithm yields two outputs: one provides information about the properties of the detected objects, and the other ~~containing the labelled mask of~~ produces a labelled mask highlighting the feature of interest ~~in~~ on the input grid (pink boxes in Figure 1).

2.1 Segmentation

The core operation ~~behind the extraction of features is to classify points in~~ for the feature extraction involves categorizing points within a scalar field into one of five shapes~~using the two principal curvature measurements derived as~~. This categorization is achieved using curvature measurements obtained from the eigenvalues of the Hessian of the basic field. ~~The five chosen~~ These five selected shapes (see Figure 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. ~~Segmentation~~ 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. ~~Next, the~~ Lastly, the algorithm calculates SI to estimate the local geometric shape ~~is calculated~~at each point.

2.1.1 Scale-space Selection

Scale-space selection is a ~~very common tool used in image and~~ widely used technique in image processing, signal processing, ~~as well as~~ and computer vision (Lindeberg, 2014). In ~~this current work, the~~ our current study, scale-space selection ~~is limited to the application of a gaussian~~ involves applying a Gaussian smoothing kernel to suppress variability ~~less than the smooth~~ smaller than the chosen smoothing scale (σ) ~~. Scale-space selection is mathematically implemented as a convolution of~~ (see https://unidata.github.io/MetPy/latest/api/generated/metpy.calc.smooth_gaussian.html for implementation of Gaussian smoothing). Mathematically, scale-space selection is performed by convolving the primary field (ϕ_p) with a ~~gaussian functiongiven as,~~ Gaussian function, expressed as follows:

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

In the context of the meso-synoptic scale processes ~~explored~~ examined in this study, scale-space selection ~~will filter~~ filters out smaller micro-scale features to isolate features like cyclonic vortexes or atmospheric rivers. ~~However~~Notably, this function can be adjusted to the spatial scale of interest and could ~~even theoretically~~ also be used to filter out ~~synoptic scale~~ synoptic-scale features in isolating ~~micro-meso scale~~ micro- and meso-scale processes. In climate datasets, ~~the grids are not always uniformly spaced~~grid spacing is not always uniform. To account for that, we adapt the above equation to be "grid-aware". ~~This is implemented by calculating~~ The input for the smoothing scale is provided in kilometers, and based on this input, we calculate the value of σ ~~along each~~ 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. ~~In~~ For future studies, ~~one could experiment with~~ researchers may explore other, more ~~sophisticated~~ 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 ($\cancel{\|\phi_s\|}|\phi_s|$), where the Hessian is given by,

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

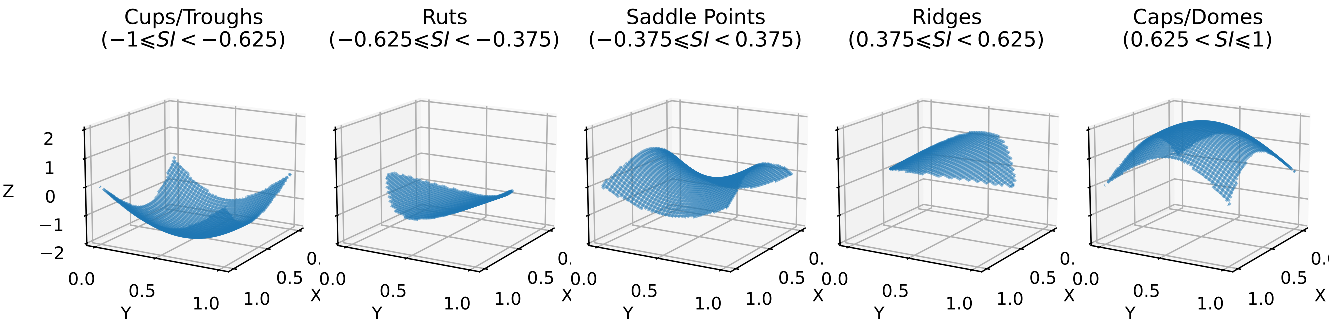

Figure 2. ~~The abridged version of the~~ Selected shapes used in this study and the values of the shape index associated with each of them. The X and Y axis are a set of general axis while $\cancel{Z(X,Y) = sin(2X) + cos(2Y)}$ $Z(X,Y) = \sin(2X) + \cos(2Y)$. Regions within $Z(X,Y)$ satisfying conditions for different shapes are isolated to show the geometry associated with them.

~~From~~ In the context of simple differential geometry, we ~~know that if $k_2 \leq k_1 < 0$,~~ 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 $k_1 \geq k_2 > 0$, the point is a local minimum. The ~~applicability of such a criterion for feature extraction is limited~~ criterion is primarily applicable to nodal features ~~like~~ such as tropical cyclones or monsoon depressions. To ~~induce continuity in the shape extraction~~expand our ability to identify a range of features, we use shape index (SI) (Koenderink and van Doorn, 1992), a quantitative measure of the local shape of the field defined as,

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

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 SI calculated using principal curvatures and SI derived from eigenvalues is exceedingly minimal in climate data analysis. The SI is used to ~~classify~~ categorize the primary field into distinct shapes (see Figure 2). The ~~values of the SI selected are set depending on the~~ choice of SI values is contingent upon the specific type of feature to be extracted. For ~~instance,~~ example, we select caps and domes ~~are selected to extract~~ when extracting features such as atmospheric depressions or cyclones~~. Ridges,~~ , whereas ridges, caps, and domes are ~~selected to extract~~ chosen when targeting features like ARs and fronts.

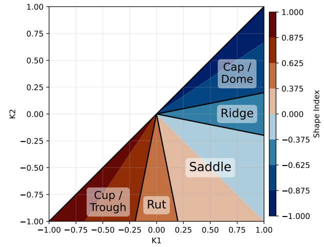

Figure 3. Sensitivity of the shape index (SI) to eigenvalues k_1 and k_2. The X and Y ~~axis~~ axes represent values of the two eigenvalues used for calculating the shape index while the color indicates the value of the shape index. Shapes corresponding to SI regimes are labelled. Shapes corresponding to SI regimes are labelled.

SI is designed to be a bounded value (range -1 to 1) independent of the magnitude of the field (Figure 3). In simple terms, SI ~~gives a continuous~~ provides a continuous and quantitative measurement of the geometric shape of the field with respect to its immediate background field. This ~~might be~~ concept is similar to how ~~the trained eye of~~ a climate scientist~~detects features from color /~~'s trained eye identifies features based on differences in color or value contrast, though SI is arguably a more objective and precise measure of geometric shape. These characteristics make ~~it more suitable~~ SI particularly well-suited for feature extraction from datasets with varying mean states~~compared~~ , 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 $\cancel{k_2}$ k_1 points perpendicular to the local ridge direction while that of $\cancel{k_1}$ k_2 is parallel to it. This allows us to ~~set further constrains on the local shape extraction if~~ impose further constraints, such as the coherence of transport or flow with respect to the local ridge when ϕ_s is a vector field~~, as~~. This capability is aptly demonstrated in the ~~detection of ARs~~ context of AR detection, as discussed in subsection 3.1.

2.2 Filtering

Once the ~~right shapes~~ 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 ~~which~~ that do not satisfy certain conditions regarding (a) grid properties like area, length, region masks, etc. (b) primary field properties like magnitude and direction, and (c) constraints from the secondary field(s). The primary aim of the filtering process is ~~mainly~~ to remove small~~or weakobjects. Since filtering is applied to the extracted objects rather than an entire field, computational cost is decreased relative to other methods.~~ , weak, or ephemeral objects.

2.3 Tracking

The ~~extracted properties of~~ properties extracted for each object include ~~positional information for the center, maximum, and minimum values. To track~~ 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 ~~the pieces of the positional information is used~~these positional attributes is tracked. In the ~~current study, a simple radius is defined and~~ present study, we employ a straightforward tracking method. For each object at time step n, we identify the closest object ~~within the given radius to each object at time n is clustered and identified from time n+1 as~~ to it at time $n+1$. If this identified object is closer than a predefined radius r, we consider it to be the same object in motion. ~~While this simple tracking method may not translate to~~ 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 micro-scale processes, it ~~could be modified with more~~ can be adapted to incorporate greater complexity if necessary.

3 Application to 2D Features

In this section, we ~~exemplify SCAFET~~ showcase how SCAFET is employed to detect cyclonic vortices, ARs, and SSTFs from various climate datasets. ~~While the highlighted examples demonstrate SCAFET 's broader capabilities~~ These examples serve to illustrate the versatility of SCAFET as applied to ~~diverse features and grid types,~~ different types of features and grids, though all the examples ~~go through~~ in this study follow the same general process shown in Figure 1. Each subsection has a table of parameters detailing the properties of the desired feature. The properties include the feature's typical spatial scales~~of the feature~~, shape index (SI) regime, minimum length, minimum area, object eccentricity, and minimum duration of ~~the track. The~~ its track. To determine the quantitative values for ~~the propertiesare obtained from a consensus of previous studiesreferenced~~ these properties, we refer to a consensus among previous studies, which are cited within each section. ~~Apart from the~~ 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 Supplementary Section 1. In addition to the results discussed in the ~~sections below, videos for each feature are also attached in the supplementary section. As the aim~~ following sections, supplementary videos are also included for each of the features. The primary objective of this work is to demonstrate ~~the ability of SCAFETto detect various features, the results for long-term climatology for~~ SCAFET's capability to detect a variety of features. Consequently, we present results for the long-term climatology of each of the features~~presented for~~, enabling a comparison with other published detection algorithms.

3.1 Atmospheric Rivers

According to the American Meteorology Society's glossary of meteorology, ~~atmospheric rivers (ARs) are "~~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). Much~~" (Ralph et al., 2018). A substantial portion of the precipitation and water vapor transport in ~~midlatitudes occur within AR structures (Guan and Waliser, 2015). They are also responsible for over~~ 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 ~~midlatitude~~ region (Waliser and Guan, 2017; Nash et al., 2018). ~~Detection and accurate projection of ARs are crucial for~~

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 ~~across the globe~~worldwide.

The ambiguity in ~~the AR detection schemes~~ 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 ~~criterion, choosing how to fix the~~ criteria, the choice of threshold for AR strength ~~changes~~ has a significant effect on the inferences drawn between the detection schemes ~~. Most of the AR detection algorithms empirically derive this threshold from dataset directly, making it sensitive to spatio-temporal and~~ (Zhao, 2020; O'Brien et al., 2022; Nellikkattil et al., 2023). Many ARDTs determine this threshold empirically from the dataset itself, which renders them sensitive to spatiotemporal variations and changes in mean-state ~~variances~~ conditions (Shields et al., 2018). SCAFET defines ARs as long (length > ~~$2000km$)~~2000 km), narrow (eccentricity > ~~0.85~~0.75) regions of strong water vapor transport (SI > ~~0.375)and~~ 0.375), and significant precipitation (minimum AR precipitation > ~~$1mm/day$~~1 mm day^{-1}) (see Table 1 for complete details). ~~This makes the comparison~~ The sensitivity of these parameters in AR detection to the characteristics of detected ARs is discussed in Supplementary Section 1. This approach reduces the sensitivity of AR characteristics ~~between different mean states less sensitive~~ to arbitrary strength thresholds, making it easier to compare ARs across different mean state conditions.

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

$$IVTx = -\frac{1}{g}\int_{1000hPa}^{300hPa} q.\mathbf{U}dp \tag{4}$$

$$IVTy = -\frac{1}{g}\int_{1000hPa}^{300hPa} q.\mathbf{V}dp \tag{5}$$

$$\|IVT\| = \sqrt{IVTx^2 + IVTy^2} \tag{6}$$

To detect AR-like structures, SCAFET ~~looks for shapes~~ employs a search for specific shapes, such as ridges, caps, and domes (see Figure 2). Following the process outlined in Figure 1, ~~shape index (SI)~~ SI is calculated after applying a ~~grid aware smoothing~~ grid-aware smoothing technique that suppresses variability smaller than ~~$1000km$~~ 1000 km (Figure 4(a)). Once SI is calculated for ~~$\|IVT\|$~~ |IVT| (Figure 4(b)), regions where SI > ~~0.375 is~~ 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 Figure 4(a)) and local ridge direction (arrows in Figure 4(b)) do not deviate by more than ~~45^{circ}~~45°. The local ridge direction is ~~defined~~ identified as the eigenvector corresponding to the ~~larger eigenvalue (k_1~~smallest eigenvalue (k_2). Filtering based on the grid properties removes candidates that are too small (length < ~~$2000km$~~ 2000 km and area < ~~$1e^{12}km^2$), and~~ 2×10^6 km^2), or too wide (eccentricity ≤ 0.75). To eliminate AR-like objects with low strength (precipitation $<$ ~~0.75)candidates. The~~ 1 mm day^{-1}) we constrain our results with the secondary field, total precipitation~~within each objectis used as to filter out weak (precipitation < $1mm/day$) AR-like structures. Precipitation is used to assess the strength of AR as it is the most socio-economically relevant . Following other AR detection methods, we have also imposed~~ , within the object's area. The use of precipitation as a strength indicator is relevant given its significant socio-economic impact. In line with other ARDTs, we impose a regional mask to ~~get rid of~~ filter out AR-like structures along the equatorial belt. All ~~the previously mentioned~~ these steps can be applied in parallel along the time axis, ~~making it computationally fast. Each time step will identify~~ and at each time step AR-like structures similar to those shown in Figure 4(c) are identified. Once all ARs are ~~identified, a simple~~ detected, the tracking algorithm is ~~implemented on~~ applied to the daily data to filter out ARs that last less than ~~two days. Tracking can be implemented based on one of the location parameters, i.e., the center, maximum, or minimum points of each detected object. For ARs, we use the~~ one day. Tracking is performed based on the centroid of each ~~detected objectto track it. Closest~~ identified object. The closest objects within a distance of ~~$4000km$ between two~~ 4000 km between two consecutive time steps are considered the same object ~~progressing in~~ evolving over time (Figure 4(d)). The annual mean frequency of the detected AR objects and their seasonality ~~is~~ are shown in Figure 4(e), (f), and (g). ~~The spatial distribution can be found to be within the uncertainty induced by other detection algorithms from the Atmospheric River Tracking Method Intercomparison Project (ARTMIP)catalog (Lora et al., 2020)~~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 Supplementary Section 2 (see also Lora et al. (2020)).

3.2 Tropical and Extratropical Cyclones

~~Cyclones are defined in the literature as large (~~In the scientific literature, cyclones are generally described as large weather systems ranging from 500–4000 km ~~) regions of~~ in size, characterized by strong cyclonic circulation~~with~~ , low pressure at ~~the centerand extremely~~ their center, and exceptionally high winds around it (Emanuel, 2003; Schultz et al., 2019; Encyclopaedia, 2022). The dynamics and ~~characteristic of cyclones will be slightly different based on the genesis location , translations speed etc~~characteristics of cyclones can vary depending on factors such as their genesis location and translation speeds. For instance, cyclones ~~formed~~ generated near the equator~~(tropical cyclones) are in general smaller in area compared to that of the cyclones~~ , commonly referred to as tropical cyclones, are typically smaller in size compared to those formed in midlatitudes~~(extratropical cyclones). Independent of this, they produce extremely high rainfall, and winds along the track and~~ , known as extratropical cyclones. Regardless of their origin, cyclones have the potential to unleash intense rainfall, powerful winds along their path, and can lead to flooding, landslides, and severe damage to ~~infrastructure along the coastlines where it makes~~ coastal infrastructure when they make landfall (Knutson et al., 2010; Mendelsohn et al., 2012; Ranson et al., 2014). ~~With the rise in sea level and enhanced intensity of cyclones~~ 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 ~~warming,~~ global warming. Thus, the identification and future projection of cyclones ~~is gaining a lot of attention from~~ are a subject of growing attention and importance for the climate community (Woodruff et al., 2013).

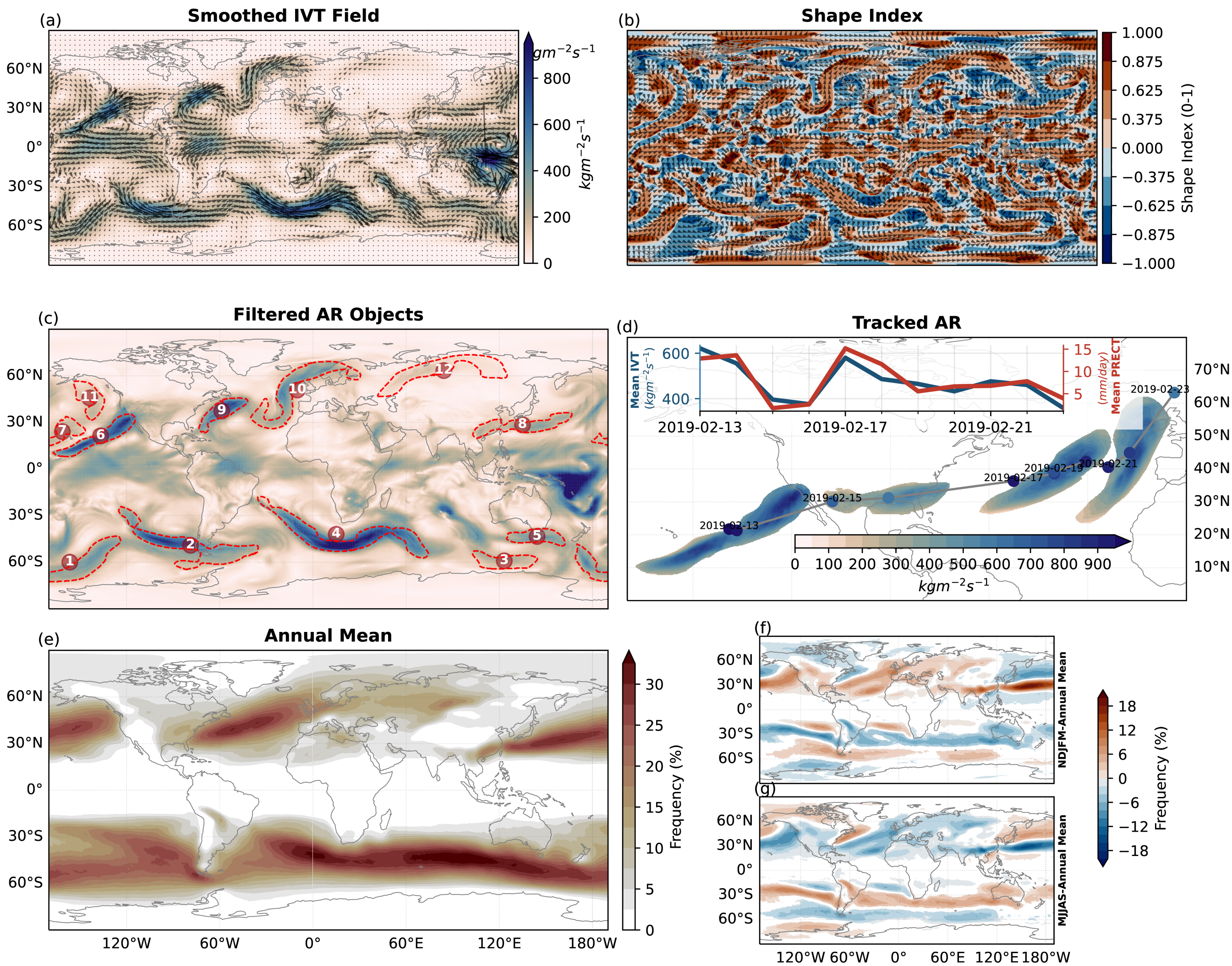

Figure 4. Major steps in the detection and tracking of Atmospheric Rivers. (a) ~~is the smoothed~~ Smoothed primary field ~~, which is the~~ of vertically integrated water vapor transport (IVT). ~~The smoothing~~ Smoothing removes variability smaller than 1000 km from the IVT. The arrows in (a) represent the direction of unsmoothed IVT. (b) ~~shows the magnitude~~ Magnitude (shading) of the shape index (SI) and direction of the local ridge (arrows) direction calculated from ~~(a)~~smoothed IVT. ~~In the next step ridges, caps, and domes are extracted from~~ (~~b~~c) ~~and weak and small~~ Labeled AR ~~candidates are filtered out. The AR~~ objects after ~~this~~ filtering ~~is shown in (c)~~out weak, small, and ephemeral candidates. ~~Finally all the objects in~~ (~~c) are tracked as shown in (~~d) ~~to obtain tracks as well as other properties like~~ Example of tracked AR centroids and marked time, inlay shows object's area mean IVT ~~and precipitation~~over time. ~~Objects that does not last more than one day is removed in this step.~~ (e) AR annual mean frequency for the period 2000 to ~~2019 is shown in~~ 2019. (~~e~~f-g) ~~, the~~ AR Frequency anomaly ~~for November~~ relative to ~~March~~ the annual mean for (f) November to March, and ~~May to September~~ (g) ~~relative~~ May to ~~the annual mean are also plotted~~September.

	No.	Property	Value	Unit
Segmentation	1	Smooth Scale	2000	~~km~~ km
	2	Angle Coherence	45	~~$degrees$~~ degrees
	3	Selected Shape	(0.375,1.0]	-
Filtering	1	Minimum Length	2000	~~km~~ km
	2	Minimum Area	2~~$\times 10^5$~~ $\times 10^6$	~~km^2~~ km^2
	3	Eccentricity	[0.75, 1.0]	-
	4	Minimum Precipitation	1	~~mm/day~~ $\mathrm{mm\,day}^{-1}$
	5	Latitude Mask	(-20, 20)	~~$degrees$~~ degrees
Tracking	1	Minimum Duration	24	~~$hours$~~ hours
	2	Maximum Distance per ~~timestep~~ Timestep	~~1000~~ 4000	~~km~~ km

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

Once again, discrepancies ~~between the~~ among different detection algorithms can be ~~traced back different~~ attributed to varying choices of physical thresholds or ~~functions on~~ constraints related to factors such as size, wind speeds, vorticity, or surface pressure ~~anomaly. Even though most studies converge on the conclusions for~~ anomalies. While most studies generally agree on the present and future characteristics of cyclones, ~~ironing out the~~ resolving details such as the changes in genesis rate ~~, duration are hindered~~ 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) with~~ 0.625) and maximum wind speeds ~~greater than $10m/s$~~exceeding $10\ \mathrm{m\,s^{-1}}$. This definition ~~is able to identify strong cyclonic vorticities all over the globe including,~~ enables the detection of robust cyclonic vorticities worldwide, including but not limited to ~~,~~tropical and extratropical cyclones. The ~~basic~~ primary field used for cyclone detection is the absolute value of cyclonic relative vorticity (ζ) defined as,

$$\zeta = \nabla \times \mathbf{U} \tag{7}$$

	No.	Property	Value	Unit
Segmentation	1	Smooth Scale	1500	~~km~~ km
	2	Selected Shape	(0.625,1.0]	-
Filtering	1	Minimum Length	20	~~km~~ km
	2	Minimum Area	$10^{\cancel{4}\,5}$	~~km^2~~ km^2
	3	Eccentricity	[0.0,1.0]	-
	4	Minimum Vorticity	10^{-6}	~~s^{-1}~~ s^{-1}
	5	Minimum Max. Windspeed	10	~~ms^{-1}~~ ms^{-1}
Tracking	1	Minimum Duration	48	~~$hours$~~ hours
	2	Maximum Distance per ~~timestep~~ Timestep	500	~~km~~ km
	3	Net Minimum Displacement	1000	~~km~~ km

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

Where U is the ~~6 hourly~~ 6-hourly wind speeds at 10 meters ~~from surface obtained from~~ above the surface obtained fro1mmdaym the ERA5 reanalysis dataset with a spatial resolution of ~~$0.25^\circ \times 0.25^{circ}$~~ $0.25^\circ \times 0.25^\circ$ (Hersbach et al., 2020). The magnitude of wind speed at 10 meters is ~~used~~ utilized as the secondary field ~~. Other cyclone related variables like the~~ to constrain detection. Additional cyclone-related variables such as surface pressure anomaly ~~,~~ and potential temperature can also ~~be used as secondary field to identify/classify~~ serve as secondary fields for the identification and classification of cyclones.

In contrast with ARs, ~~cyclones are detected using~~ the detection of cyclones relies on a scalar field, ~~and~~ specifically in this case the cyclonic relative vorticity ~~. A~~ $|\zeta|$. First, the data is pre-processed with grid-aware ~~gaussian smoothing is applied~~ Gaussian smoothing to suppress spatial variability smaller than 750 km (Figure 5(a)). The ~~smoothing scale was chosen so that we can~~ chosen smoothing scale allows us to identify both tropical and extratropical cyclones. Caps and dome shapes (SI > ~~0.625) from the smoothed relative vorticity is identified~~ 0.625) are then identified within the smoothed $|\zeta|$ field as potential cyclones (Figure 5(b)). ~~The next step is to filter out objects with~~ Subsequently, objects with an area less than ~~$10^4 km^2$, and~~ 10^5 km^2 and a diameter less than ~~$20km$. Once the aforementioned spatial characteristics are fulfilled~~ 20 km are filtered out. Once these spatial criteria are met, we can further ~~filter out~~ refine our selection by excluding weak cyclonic vorticities ~~(relative vorticity < $10^{-6}s^{-1}$), and~~ $|\zeta| < 10^{-6}$ s^{-1} and slow maximum wind speed < ~~$10ms^{-1}$) giving us all the strong cyclonic~~

~~systems identified~~ 10 $\mathrm{m\,s^{-1}}$, resulting in the identification of robust cyclonic systems for a given time step (Figure 5(c)). ~~All the processes described~~ Similar to the AR example, all the described steps can be parallelized along the time dimension~~as in the AR example. Once all~~ . Once potential cyclones are identified, they are tracked ~~like~~ using a methodology similar to the AR tracking algorithm. However, in this case, the radius for search is limited to ~~1000km as~~ 1000 km since we are using ~~6 hourly dataand translations~~ 6-hourly data, and the translation speeds of cyclones are ~~much lower than 150km/h~~notably slower than 150 $\mathrm{km\,h^{-1}}$. A minimum duration of 48 hours and a minimum total displacement of ~~500km~~ 500 km is applied to ~~isolate propagating~~ distinguish moving cyclonic circulations from stationary ones. An example of a tracked cyclone, commonly known as ~~cyclone Dorian~~"Dorian" (Avila et al., 2020) is compared with the observed track from IBTrACS ~~dataset (Knapp et al., 2010, 2018)~~ (Knapp et al., 2010, 2018) dataset (Figure 5(d)). In comparison to the observed track, SCAFET's track is much longer ~~as we use a more relaxed condition on ζ and winds speedthresholds. Another reason for the longer track is that~~due to the more relaxed conditions applied to cyclonic vorticity and wind speed. Additionally, SCAFET does not ~~distinguish~~ differentiate between tropical and extratropical cyclones~~and would end up following the vorticity while it transitions from tropical to midlatitude storms. The~~ , 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). ~~Unlike other~~ What sets SCAFET apart from other conventional cyclone detection algorithms ~~, SCAFET does not identify cyclones as a point object~~is its approach to identifying cyclones not as point objects, but as ~~a surface encompassing~~ encompassing surfaces around the point of maximum ~~ζ. This will help us study the nuanced properties of cyclones like the~~ $|\zeta|$. This enables a more comprehensive analysis of cyclone properties, including maximum and minimum values of wind speed ~~,~~ and precipitation within the ~~whole~~ entire cyclone structure.

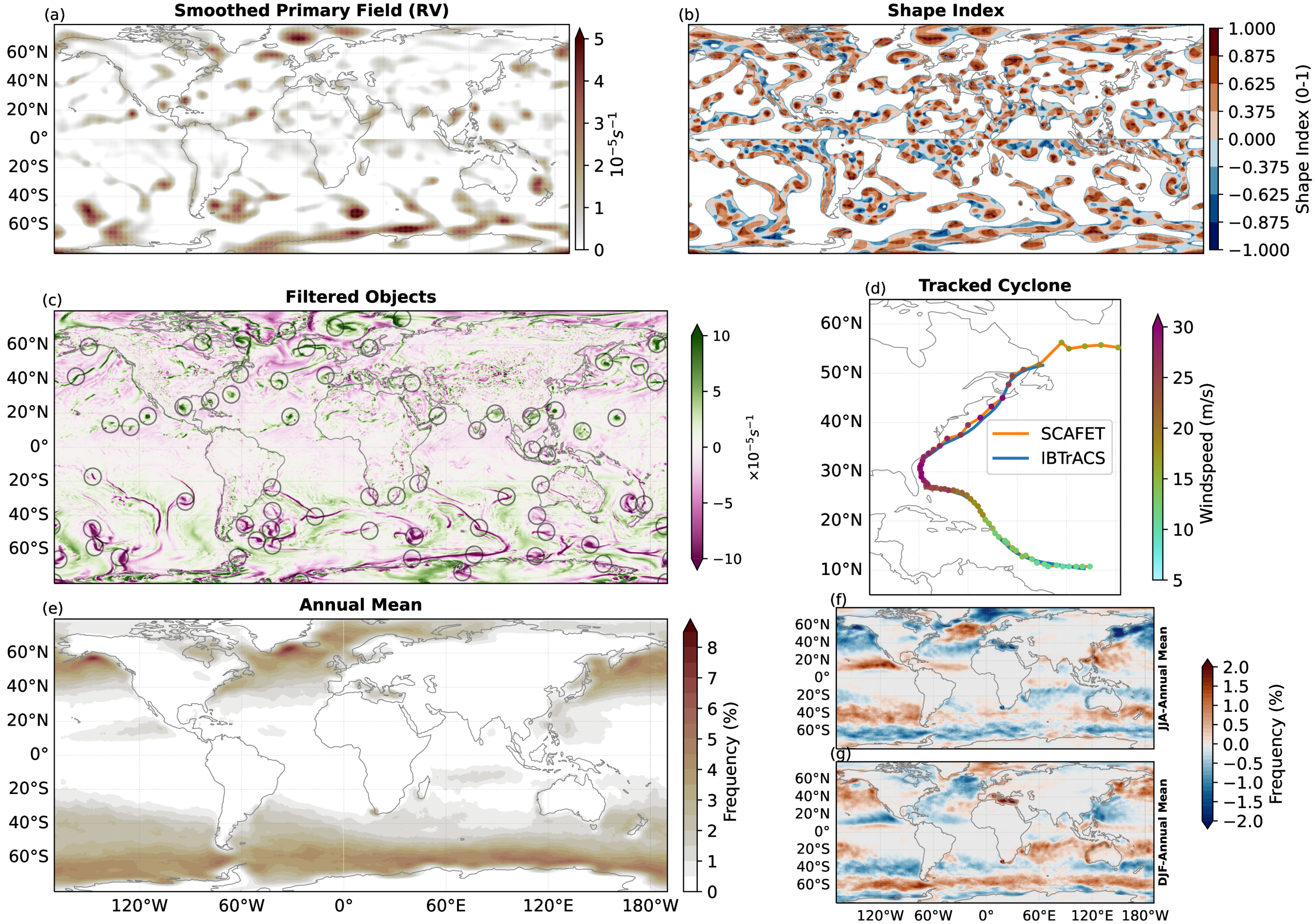

Figure 5. Major steps in the detection and tracking of cyclones. (a) Smoothed primary field ~~, which in this case is the absolute value~~ of cyclonic relative vorticity (~~ξ~~$|\zeta|$). The smoothing removes variability smaller than 750 kms from ~~ξ~~$|\zeta|$. (b) Magnitude of SI ~~calculated from (a). In~~ for the ~~next step, caps, and domes are extracted from (b) and weak and small cyclone candidates are filtered out~~primary field. ~~Cyclonic vorticities after this filtering are shown in~~ (c) Filtered cyclonic objects with the background color representing ~~the~~ unsmoothed values of ζ. ~~Finally, all the objects in (c) are tracked as shown in~~ (d) ~~to obtain tracks as well as other properties like wind speeds and vorticity.~~ Track obtained for cyclone ~~"~~“Dorian~~"~~” from SCAFET ~~is~~ compared with ~~that of~~ the track from the IBTrACS dataset. (~~d~~e) ~~. Objects that do not last more than 48 hours are removed in this step. The annual~~ Annual mean frequency of cyclone occurrence for the period 2000 to ~~2020 is shown in~~ 2020. (~~e~~f-g) ~~, the anomalous~~ Anomalous cyclone frequencies relative to the annual mean for ~~JJA~~ (f) ~~,~~ JJA and ~~DJF~~ (g) ~~relative to the annual mean are also plotted~~DJF.

3.3 Sea Surface Temperature Fronts

SST fronts are ~~the confluence regions of~~ regions where different water masses come together. They are ~~often manifested as having~~ typically characterized by strong horizontal gradients in temperature, salinity, density, and other ~~characteristics~~

properties (Bowman, 1978; Legeckis, 1978; Fedorov, 1986; Yoder et al., 1994). ~~Frontal~~ Unlike the larger meso to synoptic scale features discussed in this study, frontal structures are often observed in much smaller ~~spatio-temporal scalesthan the other features described in this study~~spatiotemporal scales. Accurate identification of SSTFs ~~are important as~~ is essential because these features are ~~often~~ frequently associated with strong upwelling ~~, and high~~ and high levels of biogeochemical productivity (Clayton et al., 2014, 2021; Nagai and Clayton, 2017). ~~Identification of SSTFs also demonstrates~~ Additionally, the detection of SSTFs serves as an example of how SCAFET can be ~~used to detect~~ applied to identify features in curvilinear grids.

~~Most previous frontal detection algorithms use edge detection algorithms~~ Many prior SSTF detection algorithms rely on edge detection techniques and the gradient of sea surface temperature and/or height ~~and to identify fronts (Canny, 1986; Castelao et al., 2006). We use~~ 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 ~~the detection of~~ detecting SST fronts. The ~~SSTs were~~ SST data is obtained from a fully coupled, ultra-high-resolution (~~$\approxeq 10km$~~$\approxeq 10\,km$) CESM v1.2.2 simulation of ~~present day mean climate (Small et al., 2014; Chu et al., 2020)~~present-day mean climate (Small et al., 2014; Chu et al., 2020; Nellikkattil et al., 2023). The data is ~~fed into~~ processed by SCAFET in the tripolar POP grid. To ~~demonstrate~~ illustrate the detection process, the analysis ~~is confined to~~ focuses on the Kuroshio frontal and extension ~~domain~~ region for the last 10 years of the simulation.

~~Frontal structures ,~~ The extraction of frontal structures using the selected shapes of ridges, caps, and domes ~~are extracted very similarly~~ is similar as in the detection of ARs. ~~A~~ Prior to extraction, a spatial smoothing of approximately ~~$30km$ is appliedbefore extracting the frontal structures~~30 km is applied. From the extracted SSTF candidates, objects with a mean SST gradient lower than ~~$10^{-4}K/m$ is removed. So are circular~~ $10^{-4}\,\mathrm{K\,m^{-1}}$ are removed. Circular (eccentricity < ~~0.5~~0.5) and small (area < ~~$1000km^2$) objects . Unlike~~ $1000\,\mathrm{km}^2$) objects are also filtered out. It is worth noting that, in contrast to AR detection, ~~SSTFs~~ frontal structures are not tracked~~as ocean fronts are stationary rather than transported into other regions~~. The detected frontal frequency ~~shows familiar~~ exhibits general patterns and seasonality ~~as~~ consistent with findings in previous studies (Xi et al., 2022).

	No.	Property	Value	Unit
Segmentation	1	Smooth Scale	30	~~km~~ km
	2	Selected Shape	(0.375, 1.0]	-
Filtering	1	Minimum Length	500	~~km~~ km
	2	Minimum Area	~~10^3~~ 10^3	~~km^2~~ km^2
	3	Eccentricity	(0.5,1.0]	-
	4	Minimum SST Gradient	10^{-4}	~~Km^{-1}~~ $\mathrm{K\,m}^{-1}$

Table 3. Same as in ~~table~~ Table 1 but for parameters and values relevant to detecting Sea Surface Temperature Fronts (SSTFs).

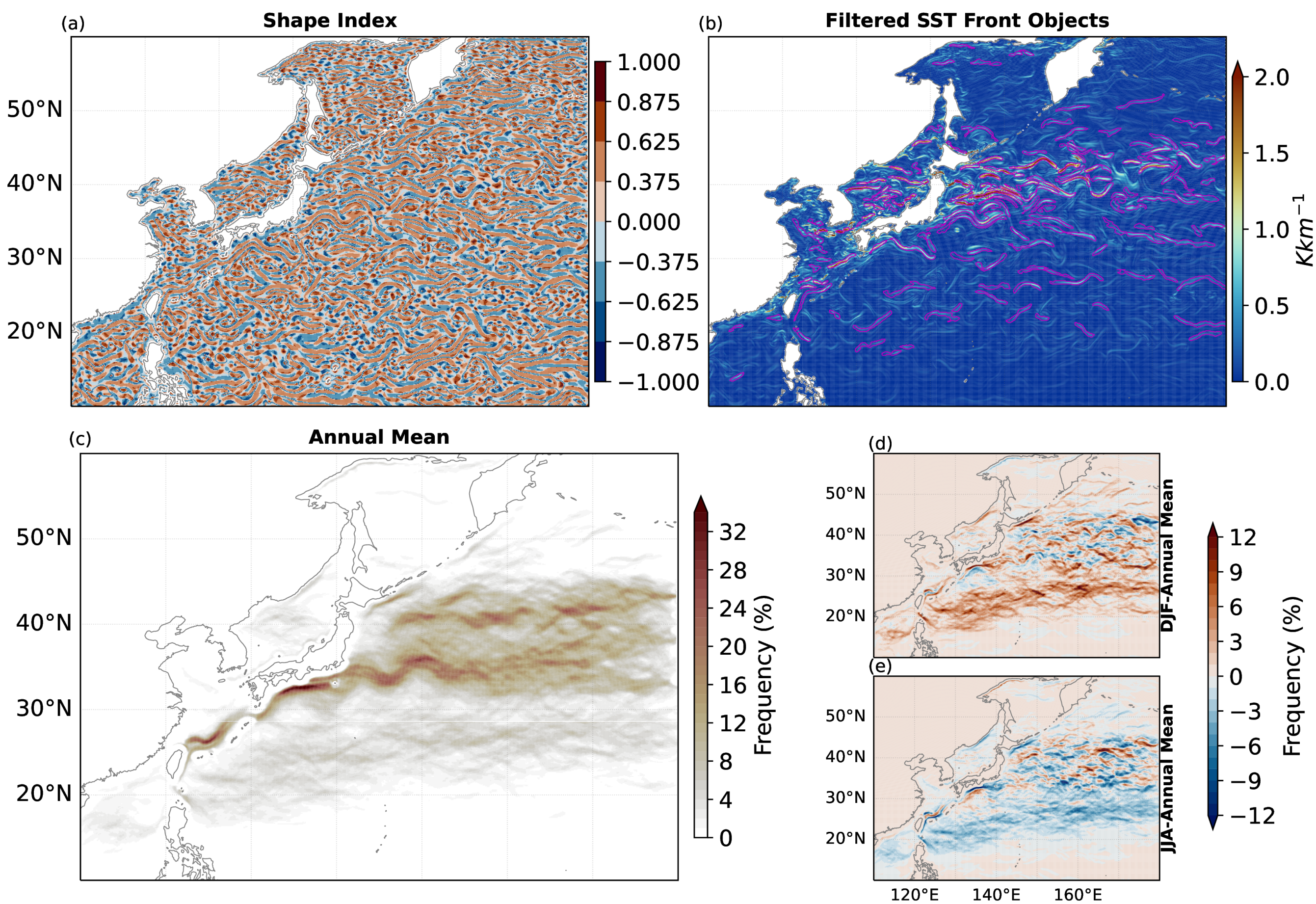

Figure 6. Major steps in the detection of sea surface temperature fronts (SSTFs). (a) Magnitude of ~~shape index (~~SI ~~)~~ as calculated from the smoothed primary field ~~, which is~~ of the horizontal gradient of sea surface temperature (∇SST). ~~The smoothing~~ Smoothing removes variability smaller than 15 km from ∇SST. ~~In the next step, ridges, caps, and domes are extracted from (b) and weak and small~~ Filtered SSTF ~~candidates are filtered out. SSTFs after this filtering are shown~~ objects, in units of Kelvin per kilometer (~~b~~$\mathrm{K\,km}^{-1}$) ~~with the~~ , where

4 Application to 3D Features

~~In this section , we show how to extend~~ This section introduces the extension of SCAFET to detect features ~~from~~ within three-dimensional (3D) primary fields. The process of scale-space selection ~~is carried out by applying gaussian smoothing along~~ involves applying Gaussian smoothing independently along each of the three dimensions~~separately. Also~~. Notably, a 3D ~~basic field would yield~~ field yields three eigenvalues ($k_1 \geq k_2 \geq k_3$) ~~rather than two. Here,~~ instead of the usual two. In this context, the SI can be calculated ~~by combining the eigenvalues~~ in three different ways ~~. The~~ by combining these eigenvalues.

For the extraction of jet streams, the SI calculated using k_1 and k_2 ~~is used for the extraction of jet streams~~ (the two largest eigenvalues) is used as it provides a more conservative estimate for the ~~jet domain (see Supplementary Figure S1).~~ jet-like structure (see Appendix subsection A3 and Supplementary Figure 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 Figure 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 $SI(k_2,k_3)$.

4.1 Jet Streams

~~Independent of the dynamics~~Jet streams, regardless of the underlying dynamics, are narrow, ~~jet streams are manifested as narrow~~high-wind-speed regions in the upper atmosphere with ~~relatively high~~ faster wind speeds compared to ~~its~~ their surroundings (Koch et al., 2006). ~~Apart from the obvious direct~~ These jet streams have a significant impact on aviation ~~, the location and characteristics of jet streams~~ and strongly influence surface weather conditions. For ~~instance~~example, a persistent jet stream in boreal summer can ~~lead to~~ result in extreme heat and flooding events, while a meandering jet ~~in the winter induces extreme~~ stream in winter leads to severe cold spells in the midlatitudes (Petoukhov et al., 2013; Coumou et al., 2014; Kretschmer et al., 2016). ~~Further~~Additionally, the northward movement of jet streams ~~in response~~ due to greenhouse warming ~~leads~~ contributes to the poleward propagation of tropical cyclones (Studholme et al., 2021). Thus ~~accurate and robust detection of jet streams are fundamental in the prediction and projection of mean~~ accurately detecting and characterizing jet streams is crucial for predicting and projecting both climatology and extreme weather systems. ~~Similar to~~

Much like the detection of other weather ~~phenomenon~~ phenomena discussed in this study, ~~most previous studies use~~ previous research typically employs a physical threshold ~~in identifying jet locations. Moreover, most of these studies except Limbach et al. and Kern et al. identifies~~ 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 ~~a one-~~ one or two-dimensional features. ~~Here we intend to demonstrate the capability of SCAFET to detect jet streams as a three-dimensional structure. Since the scope of this section is limited to the validation of the detection method, we have only shown jet~~ 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 ~~video~~

~~showing the results for a longer~~ 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 ~~seen~~ found in the supplementary section.

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

$$W = \sqrt{U^2 + V^2} \tag{8}$$

~~Where, U and V~~ where U and V are the zonal and meridional wind velocities.

~~Comparable to~~ The detection process for jet streams begins similarly to the detection of 2D features~~, the detection process starts by applying a gaussian smoothing to remove wavelengths less than 6000km~~. Gaussian smoothing is used to remove variability less than 3000 km in the horizontal dimensions. No smoothing is applied along the vertical dimension. Next, SI is calculated using the two largest eigenvalues~~k_2~~, k_1 and ~~k_3. Regions corresponding to~~ 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, pg. 60-61), and https://unidata.github.io/MetPy/latest/api/generated/metpy.calc.pressure_to_height_std.html for further details).

Similar to the detection of ARs, regions characterized by the selected shapes of ridges, caps, and domes (SI > ~~0.375~~0.375) are isolated for filtering. Filtering ~~removes objects with~~ is then applied to remove objects with a volume less than ~~1000km^3,~~ 10^6 km^3, a horizontal length less than ~~5000km, and~~ 5000 km, and a maximum wind speed within each object less than ~~50m/s~~50 m s^{-1}. In the current version of SCAFET, the tracking algorithm is not applied ~~on~~ to jet detection (see Figure 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 ~~new framework and algorithm~~ novel computational mathematical framework and an open-source Python package for extracting and tracking ~~meso-synoptic scale~~ features from large climate datasets, called Scalable Feature Extraction and Tracking (SCAFET). ~~As the~~ The purpose of SCAFET is to tackle the challenges posed by the increasing volume and diversity of ~~observational and model climate data grow, an alternative method to physical threshold-dependent feature detection is necessary to compare features within and~~ 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 ~~sets~~with different mean states by attempting to remove the need for posterior data-specific assumptions. Furthermore, SCAFET introduces a novel shape-based approach ~~for feature extractionwill give us further insights into detection method discrepancies in projections and aid~~ 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 ~~the ability of SCAFET~~SCAFET's

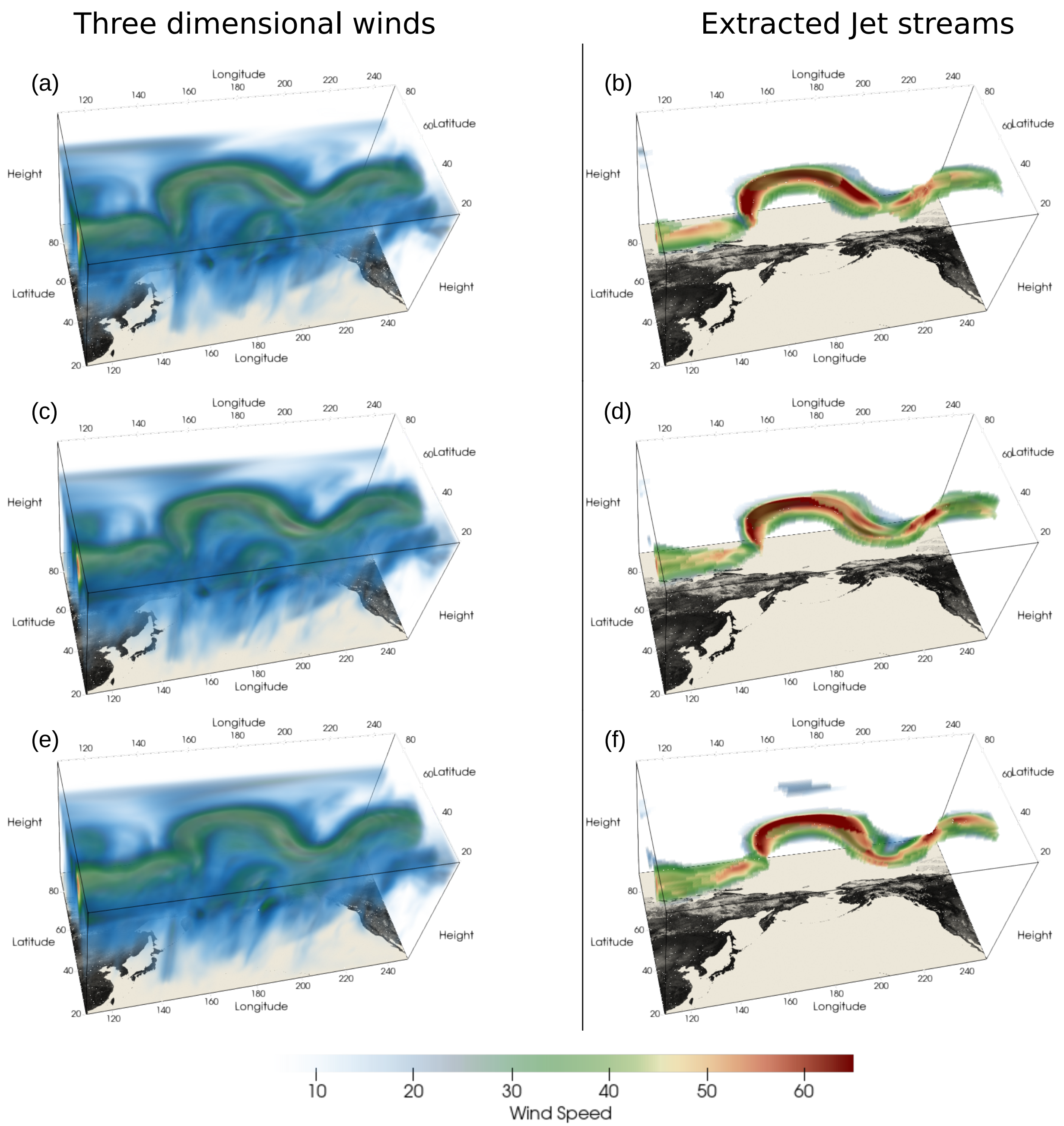

Figure 7. 3D jet streams extracted using SCAFET. ~~(a), (c), (e) shows the magnitude~~ Magnitude of 3D wind speed for ~~2022-08-25 00~~(a) 2022-08-28 12:00, (c) 2022-08-28 ~~06~~18:00, and ~~2022-08-31 18:~~(e) 2022-08-29 00~~respectively. The~~ :00. Extracted 3D jet streams ~~extracted~~ for ~~the~~ corresponding ~~time period is show~~ periods are shown in (b), (d), and (~~e~~f) respectively. The reader is encouraged to view the full video of these snapshots in the supplementary information.

	No.	Property	Value	Unit
Segmentation	1	Smooth Scale	6000	~~km~~ km
	2	Selected Shape	(0.375,1.0]	-
Filtering	1	Minimum Length	5000	~~km~~ km
	2	Minimum Height	5	~~km~~ km
	3	Minimum Volume	~~500~~ 10^6	~~km^3~~ km^3
	4	Minimum ~~Area 10^5 km^2 5 Minimum~~ Max. Wind speed	50	~~m/s~~ $\mathrm{m\,s^{-1}}$

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

capabilities and its potential in advancing these goals, we ~~illustrated 2D detection of~~ 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 3D detection of~~ as well as the detection of three-dimensional jet streams. Each application ~~was intended to give characteristic examples~~ serves as an illustrative example from which users can customize SCAFET for their ~~own research purposes~~specific research needs.

~~Apart from the obvious benefits like a more generalized framework and parallelized implementation , SCAFET, more importantly, provides~~ 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 ~~could~~ can relatively define various features ~~in climate datasets covering large periods of time, in which the mean climatevaries significantly. Instead of extracting features from physical thresholds~~ within climate datasets that span extensive periods marked by significant changes in mean climate. Rather than relying on empirically-derived, ~~we can identify them based on their local shape in the field and refine the analysis by optionally applying a minimum threshold on the extracted objects. This approach provides a view of~~ 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 ~~that account for mean state changes. Since results of~~ 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 ~~varying mean state, conclusions inherently depend on the feature extraction method. Such varying conclusions are noticed while studying~~ 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). Thusphysical threshold-independent algorithms such as SCAFET may be crucial in furthering~~ (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.

~~Undertaking a more fundamental level research into differential geometry and mathematical derivation~~ In conclusion, delving deeper into the principles of differential geometry to elucidate the physical interpretation of the relationship between SI and local geometric shape ~~could transform the way we identify extreme events~~ has the potential to revolutionize our approach to feature extraction from large datasets. ~~Due to its design, SCAFET does not require *a priori* climate information to identify features. This property can be utilized to develop simple web-based solutions for identifying and warning public against presence of extreme weather systems~~This avenue of research has the promise of significantly enhancing the algorithm's robustness and reliability. It's worth noting that, at present, SCAFET may not surpass the computational efficiency of other well-established feature extraction methods discussed above (see Supplementary Section 2.2). However, ongoing efforts to optimize and streamline the algorithm for improved computational efficiency continue. One notable strength of SCAFET's design is its independence from dataset-specific posterior information when identifying features. Moreover, the shape-based thresholds used for detecting specific features remain consistent across various grids, datasets, and climatologies. Between 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.

Author contributions. ABN wrote and developed the software package. ABN and DL prepared the manuscript draft with inputs from JYL. TAO 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, and JEC contributed equally to the manuscript revisions.

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. The version of the codes used for feature extraction and creating relevant figures in this manuscript can be downloaded from https://doi.org/10.5281/zenodo.7767301. A sample dataset for the curvilinear SST data is also included in the repository. The directory also includes sample outputs for various features discussed in the manuscript. The ERA 5 reanalysis data with varying resolutions can be downloaded from https://cds.climate.copernicus.eu/cdsapp. Single-level variables like 10m wind are obtained from https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=form, while three-dimensional variables can be extracted from https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=form. 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.

Competing interests. TAO is a member of the editorial board of the journal Geoscientific Model Development. The peer-review process was guided by an independent editor, and the authors have no other competing interests to declare.

Acknowledgements. The research was supported by the Institute for Basic Science (IBS), Republic of Korea, under IBS-R028-D1. Lemmon's contributions were in part supported by their status as a Science and Technology Policy Fellow with the American Association for the Advancement of Science. O'Brien's contributions were supported by the Director, Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy under Contract No. DE-AC02- 05CH11231 and by the Environmental Resilience Institute, funded by Indiana University's Prepared for Environmental Change Grand Challenge initiative. The authors would like to thank Prof. Axel Timmermann, Dr. Karl Stein, Dr. Ryohei Yamaguchi, and Dr. Pavan Harika Ravi ~~from IBS Center for Climate Physics for~~ for their comments on ~~shape based~~ shape-based feature extraction. We would also like to thank the ARTMIP community for their ~~feedbacks~~ feedback on SCAFET. ~~Special~~ Part of the analysis of the detected features was conducted on the IBS/ICCP supercomputer "Aleph" 1.43 petaflops high-performance Cray XC50-LC Skylake computing system with 18,720 processor cores, 9.59 PB storage, and 43 PB tape archive space. We also acknowledge the support of KREONET for the fast and reliable data transfers. Lastly, we extend special thanks to early users of the algorithm.

Appendix A: Shape Based Feature Extraction on Simple Datasets

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

A1 Application to 1D datasets

In this section, we ~~are constructing~~ draw an analogy between ~~application~~ the use of SCAFET on a two-dimensional (2D) dataset and shape-based feature extraction from a one-dimensional ~~dataset. The purpose of this discussion~~ (1D) dataset. Our intention is not to ~~advocate for a~~ promote the use of shape-based extraction of features from 1D datasets but ~~to help the readers understand~~ rather to provide readers with a fundamental understanding of this approach, along with its strengths and ~~weaknesses~~limitations.

~~Conventionally, for~~ For any differentiable curve C, the curvature is measured as the instantaneous rate of change of direction ~~of a~~ along the curve. Simply put, the curvature is measured as the rate of change of 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 Figure 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 ~~textbooks~~textbook.

Following the derivation of Shape Index (SI) for 2D datasets, we ~~could normalize the curvature for a function f to give~~ calculate the local shape of a function f using the shape parameter~~as,~~ , defined as

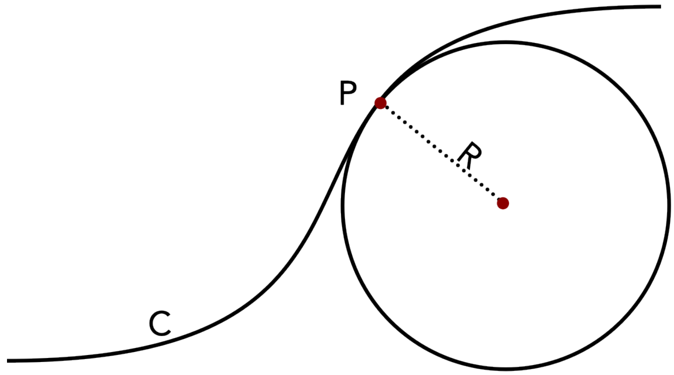

Figure A1. ~~A schematic~~ Schematic representation of ~~measuring~~ curvature measurement of a curve C at point P. At P, the curvature is the reciprocal of the radius R of the osculating circle. In differential geometry, an osculating circle is defined as the circle passing through the point P and a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

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

Values of K closer to 1 ~~can be~~ are identified as regions of local minima while K closer to -1 are regions of local maxima (black curve in Figure A2). Depending on the ~~severity of the extreme event~~magnitude of the function, one could ~~choose a value for K to get~~ adjust the value of K to obtain regions of local maxima (red caps in Figure A2) and local minima (green caps in Figure A2). The curvature of the function is insensitive to linear trends and mean state changes. This is evident as the application of ~~the same~~ identical shape thresholds identifies ~~identical regions~~ the same regions of the curves as local maxima and minima~~in a simple trigonometric~~ , whether on the base curve (blue curve in Figure A2) ~~and~~ or on the same curve with an added linear trend (orange curve in Figure A2). The values of ~~K~~ K for both curves are represented by the black line in Figure A2. Thus, ~~it~~ the shape parameter can be used to identify ~~extreme events from datasets without being affected by the~~ the local minima and maxima from a 1D dataset despite background state changes.

A2 Application of SCAFET to simple Geostrophic Motion

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

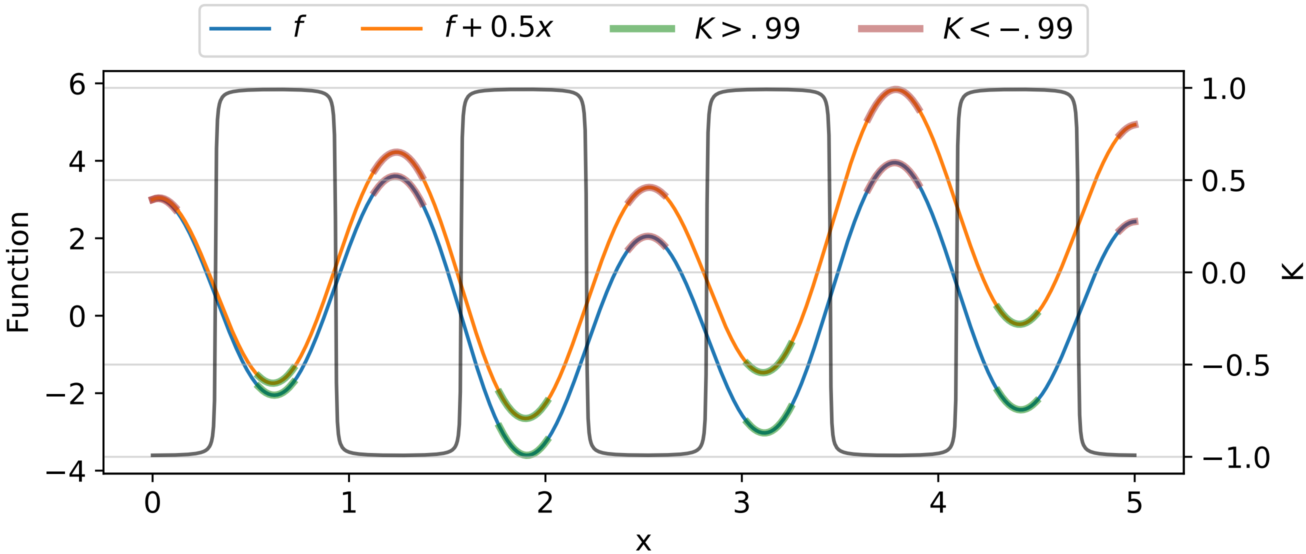

Figure A2. ~~Shape~~ Comparison of shape extraction ~~for~~ between a simple one-dimensional ~~curves~~ curve, given by ~~$f = sin2x + 3cos5x$ and $f + 0.5x = sin2x + 3cos5x + 0.5x$. The first one is a simple trigonometric~~ $f = \sin 2x + 3\cos 5x$ (blue curve~~while~~) and $f + 0.5x = \sin 2x + 3\cos 5x + 0.5x$ (orange; the ~~second~~ blue curve ~~includes~~ with a linear trend~~as is evident from their functional forms~~). ~~The magnitudes~~ Left Y-axis shows magnitude of both ~~the~~ functions~~are shown in the left Y axis while the~~, right ~~Y axis~~ Y-axis indicates ~~the~~ values of the shape parameter (K). Note that value of K is the same for both ~~the~~ functions. The green and red highlighting on the curves ~~indicates the~~ shows regions where $K > 0.99$ and $K < -.99$~~. The red highlighted are tagged as extreme~~, corresponding to regions of local maxima ~~events~~ and ~~those highlighted with green are be characterized as extreme~~ minima~~events~~, respectively.

To demonstrate the characteristics and ~~strengths~~ advantages of feature detection based on ~~Shape Index (SI)~~ SI, let's consider a simple ~~rotation~~ rotational wind field (see Figure A3(a) vectors) given by,

$$u_g = -\Omega y \tag{A2}$$

$$v_g = \Omega x \tag{A3}$$

Where Ω is a constant ($\Omega = 10^5 rad\, s^{-1}$) and x,y represents the grid. ~~In the current example the value of Ω is set as $10^5 rad/s$.~~ The geopotential height (h) of the field (see Figure A3(a) shading) is used as our primary field ~~to identify features using shape index (SI). h is estimated~~ in calculating SI, computed as,

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

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

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

Where the ~~eigen values~~ eigenvalues k_1 and k_2 are given by,

$$k_{12} = \frac{f\zeta_g}{2g} \pm \sqrt{\left(\frac{f}{2g}\right)^2 - \left(\frac{f}{g}\right)^2 \frac{\partial v_g}{\partial x}\frac{\partial u_g}{\partial y} + \frac{\partial u_g}{\partial x}\frac{\partial v_g}{\partial y}} \tag{A6}$$

Where ζ_g is the geostrophic vorticity. Which gives SI as,

$$SI(k_1, k_2) = \frac{2}{\pi}\cancel{tan}\tan^{-1}\left[\frac{\zeta_g}{-2\sqrt{\left(\frac{\zeta_g}{2}\right)^2 - \frac{\partial v_g}{\partial x}\frac{\partial u_g}{\partial y} + \frac{\partial u_g}{\partial x}\frac{\partial v_g}{\partial y}}}\right] \tag{A7}$$

A detailed derivation of the above equation can be found in ~~Appendix B.~~ Appendix B Plugging in the values for the rotational motion, we get

$$\zeta_g = \nabla^2 h = \Omega f/g \tag{A8}$$

$$\frac{\partial u_g}{\partial x} = \frac{\partial v_g}{\partial y} = 0 \tag{A9}$$

$$\frac{\partial v_g}{\partial x}\frac{\partial u_g}{\partial y} = \Omega^2 \tag{A10}$$

Therefore,

$$SI = \frac{2}{\pi}\cancel{tan}\tan^{-1}\left[\frac{\Omega^2}{-\sqrt{\Omega^2 - \Omega^2}}\right] = -1 \tag{A11}$$

Thus, SCAFET classifies the whole domain with ~~the~~ anticlockwise rotational motion as a trough with SI≊-1 regardless of the absolute value of the field or Ω. ~~A traditional method that uses thresholding directly on~~ In contrast, traditional methods that rely on thresholding the geopotential height would identify regions ~~depending on the value of the threshold on h. However, the value of the threshold must~~ based on the chosen threshold of h, which would need to be adjusted depending on the mean (time) and background (space) state. Another ~~widely used practice is to define~~ common approach is to establish a threshold on the smallest eigenvalue~~. The intention of such methods is~~ , aiming to identify extreme features based on the ~~strength of the curvature~~ curvature strength rather than the ~~actual value of the field. TempestExtremes~~ field's actual value. TempestExtremes (Ullrich and Zarzycki, 2017), a feature extraction framework ~~previously discussed , follow~~ discussed in the main text, follows this method to ~~identify~~ detect Atmospheric Rivers from gridded datasets. In the current example, this approach would correspond to ~~thresholding~~ setting a threshold on $f\Omega/g$. In other words, TempestExtremes would only ~~detect the~~ identify a trough if the value of Ω ~~is greater than the predetermined threshold. Contrastingly, SCAFETwould identify the trough~~ exceeds the

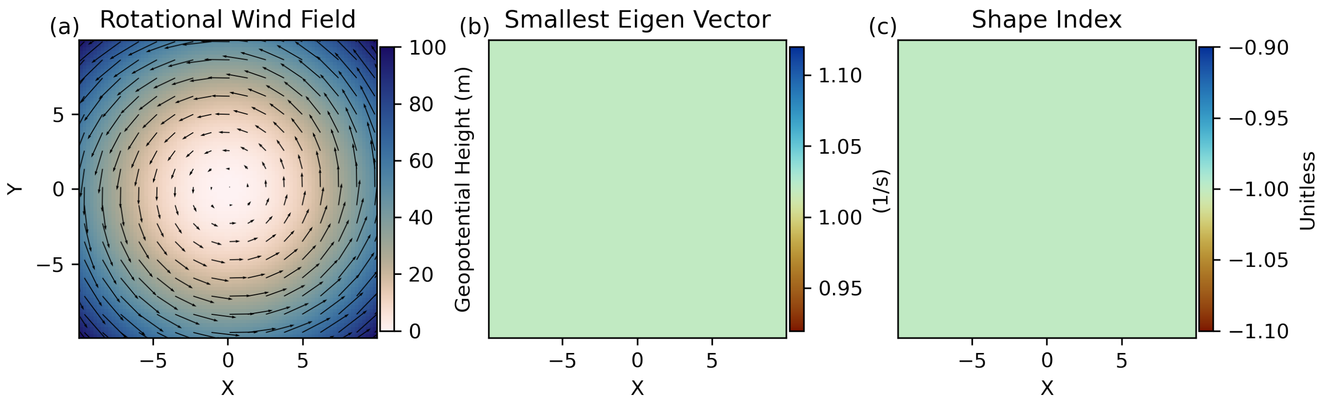

Figure A3. ~~Comparing different~~ Comparison between two feature extraction techniques on ~~synthetic~~ an idealized example of rotational wind field. (a) The geopotential height (h) (shading) of the rotational wind field (arrows). h is defined as $\Omega f(x^2+y^2)/2g$, where ~~$f = 10^(-4)1/s$~~ $f = 10^{-4} s^{-1}$, ~~$g = 9.805 m/s$~~ $g = 9.805\, m\, s^{-1}$ and ~~$\Omega = 10^5 rad/s$~~ $\Omega = 10^5 rad\, s^{-1}$. (b) ~~The magnitude~~ Magnitude of the smallest ~~eigen value. From the equation~~eigenvalue, ~~the smallest eigen value can be~~ derived from the equation as $f\Omega/g = 1.0199$, illustrating a uniform field as expected. (c) ~~Shows the value~~ Value of the ~~shape index (~~SI ~~). From the equations and the plot, we can see that the value of~~ where $SI = -1$ throughout the domain, as expected.

pre-determined threshold. SCAFET, on the other hand, identifies the trough region as a trough regardless of the ~~actual~~ specific value of the field or Ω. ~~Hence, we see that~~ This illustrates how feature extraction using SI and other published methods can ~~give us~~ yield different results depending on the input data~~as they are looking at different~~, 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 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 section 4, a three-dimensional field provides us with three eigenvalues satisfying the condition $k_1 \geq k_2 \geq k_3$. The SI can be computed using $SI(k_1,k_2)$, $SI(k_1,k_3)$, or $SI(k_2,k_3)$. Setting a threshold of SI > 0.375 effectively isolates the cylinder when using either $SI(k_1,k_2)$ or $SI(k_1,k_3)$ (see Figure A4(b-d)). Between these two options, $SI(k_1,k_2)$, 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 Figure A4. Notably, employing $SI(k_2,k_3)$ is not suitable as it fails to isolate the desired cylinder shape effectively. The choice of using $SI(k_1,k_2)$ is specifically tailored for extracting convex shapes or local maxima. Interestingly, to identify concave shapes or 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.

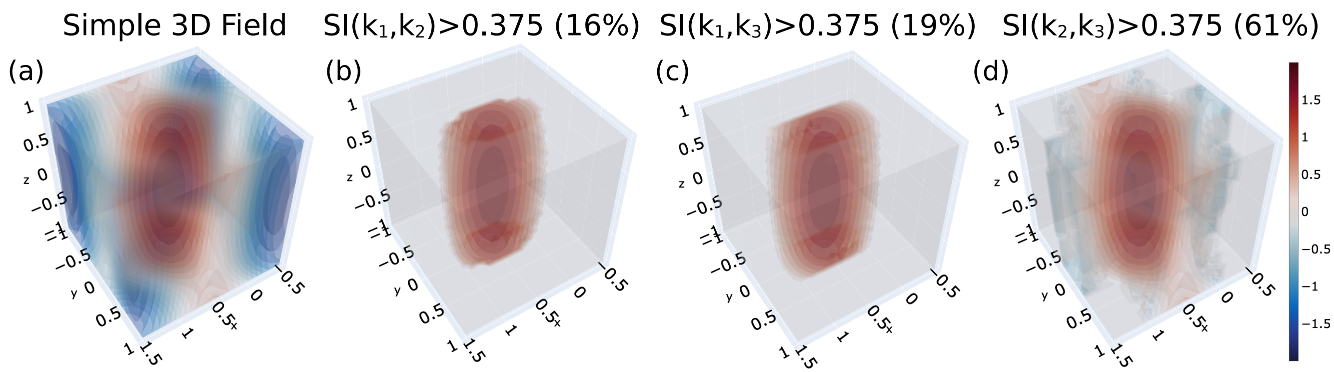

Figure A4. Various approaches for extraction of a 3D cylinder from a scalar field. (a)Simple scalar field represented by $\sin(3X)+\cos(4Y)*\cos(Z)$ is shown. (b-d) The extracted cylinders by applying the conditions (b) $SI(k_1,k_2)>0.375$, while in (c) $SI(k_1,k_3)>0.375$ and (d) $SI(k_2,k_3)>0.375$ 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.

Appendix B: Derivation of Shape Index for Geostrophic Motion

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

Let h be the geopotential height at a certain level. The ~~hessian~~ 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 ~~eigen values~~ eigenvalues of the symmetric matrix $\mathcal{H}$ is 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~~ which can be expanded as;

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

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

NOTE: 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 equation (3)

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

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

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

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, 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, Geoscientific Model Development, 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, Bulletin of the American Meteorological Society, 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, Advances in Statistical Climatology, Meteorology and Oceanography, 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, pp. 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, Annales Geophysicae, 32, 197–206, https://doi.org/10.5194/angeo-32-197-2014, 2014.

Canny, J.: A Computational Approach to Edge Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, 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, Journal of Geophysical Research, 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, Science Advances, 6, 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, Journal of Plankton Research, 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, Journal of Geophysical Research: Oceans, 126, 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, Proceedings of the National Academy of Sciences, 111, 12 331–12 336, https://doi.org/10.1073/pnas.1412797111, 2014.

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

Emanuel, K.: Tropical Cyclones, Annual Review of Earth and Planetary Sciences, 31, 75–104, https://doi.org/10.1146/annurev.earth.31.100901.141259, 2003.

Encyclopaedia, B.: Cyclone, https://www.britannica.com/science/cyclone-meteorology, 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, Journal of Geophysical Research: Atmospheres, 120, 12 514–12 535, 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, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hewson, T. D.: Objective fronts, Meteorological Applications, 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, Monthly Weather Review, 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, Monthly Weather Review, 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, Journal of 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, Journal of Physical Oceanography, 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, Journal of 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, Quarterly Journal of the Royal Meteorological Society, 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, Geophysical Research Letters, 48, 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 Transactions on Visualization and Computer Graphics, 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), Bulletin of the American Meteorological Society, 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, Nature Geoscience, 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, International Journal of Climatology, 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 and Vision Computing, 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, Journal of 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, Journal of Geophysical Research, 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, Geoscientific Model Development, 5, 457–470, https://doi.org/10.5194/gmd-5-457-2012, 2012.

Lindeberg, T.: Scale Selection, in: Computer Vision, pp. 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, Geophysical Research Letters, 47, 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, John Wiley & Sons, Nashville, TN, 2015.

Mendelsohn, R., Emanuel, K., Chonabayashi, S., and Bakkensen, L.: The impact of climate change on global tropical cyclone damage, Nature Climate 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 System Dynamics, 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 Dynamics, 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 Polar Hydroclimate, Journal of Geophysical Research: Atmospheres, 123, 6804–6821, https://doi.org/10.1029/2017jd028130, 2018.

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, Communications Earth & Environment, 4, 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, Bulletin of the American Meteorological Society, 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, Journal of Geophysical Research: Atmospheres, 127, 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, Proceedings of the National Academy of Sciences, 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, Climate Dynamics, 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, Procedia Computer Science, 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, Journal of 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, Bulletin of the American Meteorological Society, 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, Climatic 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, Monthly Weather Review, 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, Meteorological Monographs, 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., Collow, A., Krishnan, H., Magnusdottir, G., and and, P. N.: Atmospheric River Tracking Method Intercomparison Project (ARTMIP): project goals and experimental design, Geoscientific Model Development, 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, Journal of Advances in Modeling Earth Systems, 6, 1065–1094, https://doi.org/10.1002/2014ms000363, 2014.

Strong, C. and Davis, R. E.: Winter jet stream trends over the Northern Hemisphere, Quarterly Journal of the Royal Meteorological Society, 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, Nature Geoscience, 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, Climate Dynamics, 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, Theoretical and Applied Climatology, 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, Geoscientific Model Development, 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, Geoscientific Model Development, https://doi.org/10.5194/gmd-2020-303, 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, pp. 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, Journal of Geophysical Research: Atmospheres, 125, 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, Journal of 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, Nature Geoscience, 10, 179–183, https://doi.org/10.1038/ngeo2894, 2017.

Wallace, J. M. and Hobbs, P. V.: Atmospheric science, Academic Press, San Diego, CA, 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 Climate 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, Frontiers in Marine Science, 9, 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), Geoscientific Model Development, 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, International Journal of Digital 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, Geophysical Research Letters, 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, Journal of Climate, 33, 10 287–10 303, https://doi.org/10.1175/jcli-d-20-0241.1, 2020.