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

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

[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]Department of Earth and Atmospheric Sciences, Indiana University Bloomington, Indiana, USA 47403
[4]Climate and Ecosystem Sciences Division, Lawrence Berkeley National Lab, Berkeley, USA 95720
[5]Office of Clean Energy Demonstration, U.S. Department of Energy, Washington, DC 20585
[6]Research Center for Climate Sciences, Pusan National University, Busan, Republic of Korea, 46241
[7]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 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 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. 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 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. SCAFET can be implemented to extract and track most weather and climate features.

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 (Overpeck et al., 2011; Balaji et al., 2018). Frontier research like global 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 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 task. Although these features of interest 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 a small percentage of the whole simulation. Thus, extraction

20   of features not only enables us to focus our analysis on high-impact rare events but can also considerably reduce the amount of data that needs to be stored, improving computational efficiency in analysing 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 vital to climate data processing, analysis, and model development.

25   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 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

30   methods are linked to the somewhat arbitrary choice of physical thresholds in relation to the underlying spatio-temporal 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, and dataset used. Validation then unfortunately comes down to the intelligent but subjective human eye, or in other words tuning a 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).

35   Choosing an absolute threshold from climate variables for feature extraction that is applicable 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 in which the features are most visible, such as relative vorticity and sea

40   level pressure anomalies for tropical cyclones (e.g., Vitart et al., 1997) or integrated water vapor transport for ARs (e.g., Guan and Waliser, 2015). Thus, before the actual detection process is applied, one must pre-process entire datasets just to calculate reasonable thresholds that will allow for comparison within and between models. This process becomes increasingly infeasible for higher resolutions and large ensemble data sets, highlighting the need for a method of feature extraction that is not empirically derived and thus less sensitive to the climate mean state.

45   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 future climate change scenarios is further untenable as the underlying spatio-temporal climate variable distributions change under global warming. Feature detection must be reconsidered when applied to variables with significant 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

50   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). To counter these uncertainties, methods based on topology, machine learning, ridge extraction, edge detection, and various other image processing techniques have been proposed over the years (Dixon and Wiener, 1993; Post et al., 2003; Molnos et al., 2017; Biard and Kunkel, 2019; Xu et al., 2020). While these meth-

ods offer an alternative for the extraction of features in datasets spanning different mean states, many of these methods were
55 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
60 shape-based rather than physical thresholds (see Appendix A), 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
65 can easily implement the algorithm.

The need for a general framework in 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 general framework for extracting and tracking features from meteorological datasets. The first step is segmentation, where the field is split into distinct regions by applying a threshold and then labelling each of the connected regions as an object. Later, feature nodes are
70 extracted by filtering out regions outlined in the first step based on the characteristics of each object. Finally, the feature nodes are tracked over time to produce the final output for dynamical analysis. This framework was developed further for cyclones, storm tracks, convective systems, ocean eddies, monsoon depressions, and more (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,
75 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,
80 2017; Ullrich et al., 2021), another algorithm package that uses several core functions to detect a variety of features. These functions are being actively developed for extraction, characterization and uncertainty quantification of weather extremes. Both TECA and TempestExtremes have been widely implemented by the climate community and have been monumental in advancing scientific understanding of meso and synoptic scale processes and their contributions to long-term climate trends. Further discussion on the differences between SCAFET and other detection algorithms can be found in Appendix A. Even in
85 the context of these recent advancements, SCAFET aims to upgrade the detection process to be less sensitive to the physical thresholds used while presenting a novel shape-based approach to feature extraction.

The novelty of SCAFET compared to pre-existing methods lies in the use of a comprehensive mathematical framework for extracting different features based on the overall "shape" of a climate variable field, rather than thresholding of that field. 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
90   and the other for the shape of 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 to both
rectilinear and curvilinear grids and can also be extended to detect three-dimensional (3D) features. In a nutshell, SCAFET is
mathematically comprehensive, computationally efficient, and easily implementable, and could prove to be a robust method
for detecting a diverse set of features under different mean climate states.

95   The paper is organized as follows, section 2 describes the basics of SCAFET and how it is implemented in a two-dimensional
(2D) field. Section 3 provides three SCAFET use cases for the detection of various climate features and different grid types.
Extraction of 3D features will be discussed in section 4. Though the application of SCAFET is not limited to the features
described here, this study showcases atmospheric rivers, cyclones, SST fronts, and jet streams to cover a broad range of
phenomena through which users could learn to adapt SCAFET to their needs.

2   Description of Scalable Feature Extraction and Tracking

100

SCAFET follows the same three processes as discussed by Hodges -Segmentation, Filtering and Tracking (yellow boxes in
Figure 1). Before starting the this process, SCAFET is initialized with information describing the datasets and the type of
feature to be extracted (blue boxes inFigure 1). Primary inputs includes the following.

- **A primary field** (ϕ_p): This is a gridded dataset in which the feature to be extracted is most clearly visible. Optionally,
105   one or more secondary fields can be used to further constrain the detected features.

- **Grid Properties**: Information on the primary field including grid cell area/volume, grid distance, and coastlines are
    required for calculating derivatives of the basic field and identifying landfalling locations.

- **Object properties**: The algorithm requires information on the feature properties. This includes approximate spatial
    scale, shape, eccentricity (for 2D features only), minimum length, minimum area, minimum volume (for 3D feature
110   only), minimum duration, and maximum distance per time step. (see section 3 and section 4 for examples)

As the segmentation, filtering, and tracking are mutually independent in the SCAFET scheme, users can replace any of them
and still run the algorithm. After implementing all three steps, two outputs are obtained: one describing the properties of the
detected objects and the other containing the labelled mask of the feature of interest in the input grid (pink boxes in Figure 1).

2.1   Segmentation

115   The core operation behind the extraction of features is to classify points in a scalar field into one of five shapes using the
two principal curvature measurements derived as eigenvalues of the Hessian of the basic field. The five chosen shapes (see
Figure 2) are an abridged version of the shapes described in previous studies (Koenderink and van Doorn, 1992). Depending on
the feature of interest, one or more shapes are extracted from the primary field. Segmentation starts with scale-space selection
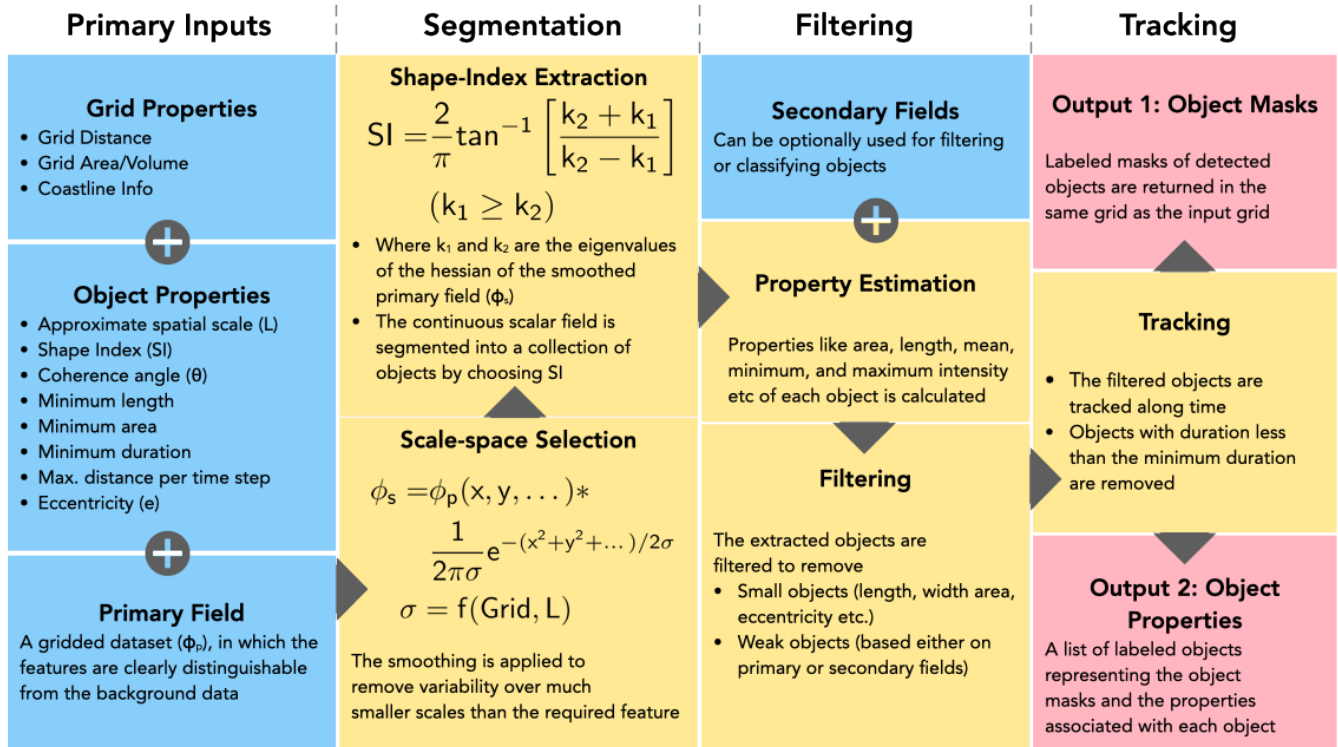
Figure 1. The overall schematic, workflow, and components of SCAFET. The inputs, processes/calculations, and outputs of the algorithm are shown in blue, yellow, and pink boxes respectively. The arrows in the periphery of the boxes represent the workflow of the algorithm. Each section is explained in detail in text.

of the field to remove smaller scales of variability that are background noise compared to the feature of interest. Next, the local

120 geometric shape is calculated.

2.1.1 Scale-space Selection

Scale-space selection is a very common tool used in image and signal processing, as well as computer vision (Lindeberg, 2014). In this current work, the scale-space selection is limited to the application of a gaussian smoothing kernel to suppress variability less than the smooth scale (σ). Scale-space selection is mathematically implemented as a convolution of the primary

125 field (ϕ_p) with a gaussian function given as,

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

In the context of the meso-synoptic scale processes explored in this study, scale-space selection will filter out smaller micro-scale features to isolate features like cyclonic vortexes or atmospheric rivers. However, this function can be adjusted to the

spatial scale of interest and could even theoretically be used to filter out synoptic scale features in isolating micro-meso scale
130 processes. In climate datasets, the grids are not always uniformly spaced. To account for that, we adapt the above equation to
be "grid-aware". This is implemented by calculating the value of σ along each circle of latitude. In future studies, one could
experiment with other, more sophisticated scale-space selection methods.

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
135 magnitude of the field ($\|\phi_s\|$), where the Hessian is given by,

$$\mathcal{H}\left(\|\phi_s\|\right) = \begin{bmatrix} \frac{\partial\|\phi_s\|}{\partial x^2} & \frac{\partial\|\phi_s\|}{\partial xy} \\ \frac{\partial\|\phi_s\|}{\partial yx} & \frac{\partial\|\phi_s\|}{\partial y^2} \end{bmatrix} \qquad (2)$$
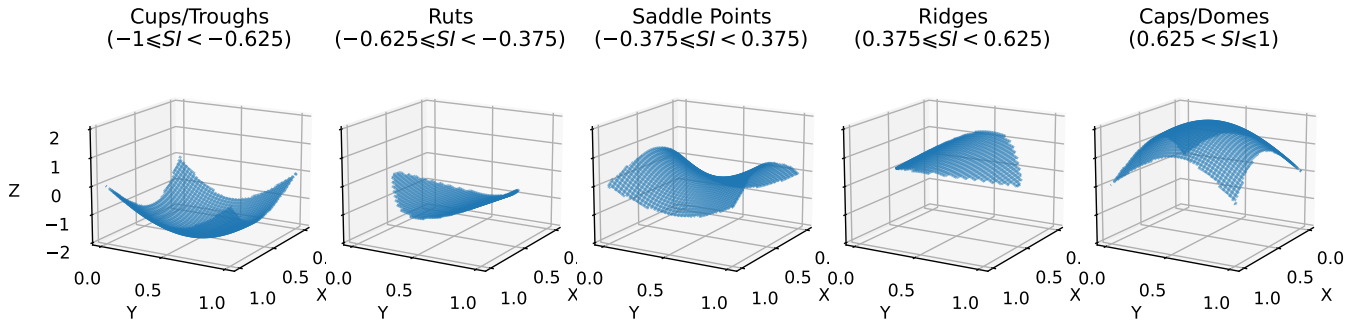
Figure 2. The abridged version of the shapes used in this study and the values of the shape index associated with each of them. The X and Y
axis are a set of general axis while $Z(X,Y) = sin(2X) + cos(2Y)$. Regions within $Z(X,Y)$ satisfying conditions for different shapes are
isolated to show the geometry associated with them.

From simple differential geometry, we know that if $k_2 \leq k_1 < 0$, 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 to nodal
features like tropical cyclones or monsoon depressions. To induce continuity in the shape extraction, we use shape index (SI)
140 (Koenderink and van Doorn, 1992), a quantitative measure of the local shape of the field defined as,

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

Where k_1 and k_2 are the two eigenvalues, satisfying $k_1 \geq k_2$, for the Hessian matrix. The SI is used to classify the primary
field into distinct shapes (see Figure 2). The values of the SI selected are set depending on the type of feature to be extracted.

For instance, caps and domes are selected to extract features such as atmospheric depressions or cyclones. Ridges, caps, and
145 domes are selected to extract ARs and fronts.

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 quantitative measurement of geometric shape of the field with respect to its immediate background field.
This might be similar to how the trained eye of a climate scientist detects features from color/value contrast, though SI is
arguably a more objective measure of geometric shape. These characteristics make it more suitable for feature extraction from
150 datasets with varying mean states compared to traditional physical threshold-based methods. In addition to the two eigenvalues,
the shape extraction provides us with corresponding eigenvectors. The eigenvector for k_2 points perpendicular to the local ridge
direction while that of k_1 is parallel to it. This allows us to set further constrains on the local shape extraction if ϕ_s is a vector
field, as demonstrated in the detection of ARs in subsection 3.1.
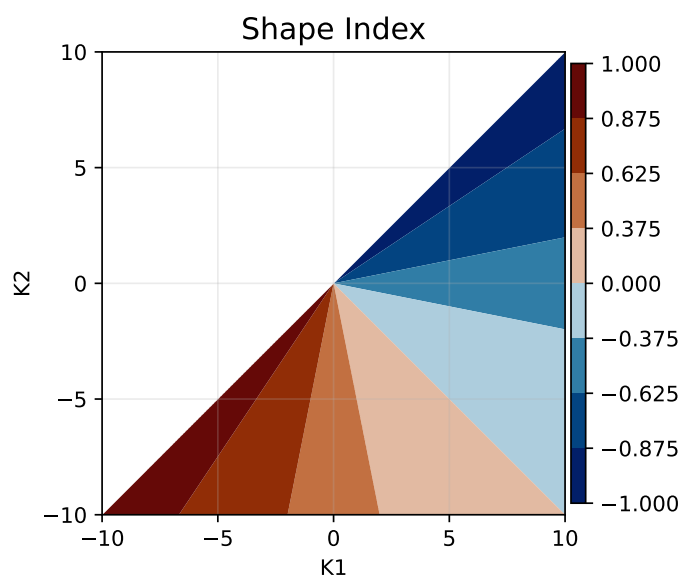
Figure 3. Sensitivity of shape index to eigenvalues k_1 and k_2. The X and Y axis represent values of the two eigenvalues used for calculating
shape index while the color indicates the value of shape index.

2.2 Filtering

155 Once the right shapes 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 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 secondary field(s). The aim of the filtering process is mainly to remove small or weak objects. Since
filtering is applied to the extracted objects rather than an entire field, computational cost is decreased relative to other methods.

2.3 Tracking

The extracted properties of each object include positional information for the center, maximum, and minimum values. To track objects through time, one of the pieces of the positional information is used. In the current study, a simple radius is defined and the closest object within the given radius to each object at time n is clustered and identified from time n+1 as the same object in motion. While this simple tracking method may not translate to micro-scale processes, it could be modified with more complexity if necessary.

3 Application to 2D Features

In this section, we exemplify SCAFET to detect cyclonic vortices, ARs, and SSTFs from various climate datasets. While the highlighted examples demonstrate SCAFET's broader capabilities as applied to diverse features and grid types, all the examples go through 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 typical spatial scales of the feature, shape index (SI), minimum length, minimum area, object eccentricity, and minimum duration of the track. The quantitative values for the properties are obtained from a consensus of previous studies referenced within each section. Apart from the results discussed in the sections below, videos for each feature are also attached in the supplementary section. As the aim of this work is to demonstrate the ability of SCAFET to detect various features, the results for long term climatology for each of the features presented for comparison with other published detection algorithms.

3.1 Atmospheric Rivers

According to the American Meteorology Society's glossary of meteorology, atmospheric rivers (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 of the precipitation and water vapor transport in midlatitudes occur within AR structures (Guan and Waliser, 2015). They are also responsible for over 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 extreme weather preparedness as well as for water resource management in basins across the globe.

The ambiguity in the AR detection schemes stems from the quantitative definition of strength, length, narrowness, and other parameters used in detection. In comparison with other criterion, choosing how to fix the threshold for AR strength changes 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 mean-state variances (Shields et al., 2018). SCAFET defines ARs as long (length $> 2000km$) narrow (eccentricity > 0.85) regions of strong water vapor transport (SI > 0.375) and precipitation (minimum AR precipitation $> 1mm/day$) (see Table 1 for complete details). This makes the comparison of AR characteristics between different mean states less sensitive to arbitrary strength thresholds.

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

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

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

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

To detect AR-like structures, SCAFET looks for shapes such as ridges, caps, and domes (see Figure 2). Following Figure 1, shape index (SI) is calculated after applying a grid aware smoothing that suppresses variability smaller than $1000km$ (Figure 4(a)). Once SI is calculated for $\|IVT\|$ (Figure 4(b)), regions where SI > 0.375 is 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}. The local ridge direction is defined as the eigenvector corresponding to the larger eigenvalue (k_1). Filtering based on the grid properties removes small (length $< 2000km$ and area $< 1e^{12}km^2$), and wide (eccentricity < 0.75) candidates. The secondary field, total precipitation within each object is 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 a regional mask to get rid of AR-like structures along the equatorial belt. All the previously mentioned steps can be applied in parallel along the time axis, making it computationally fast. Each time step will identify AR-like structures similar those shown in Figure 4(c). Once all ARs are identified, a simple tracking algorithm is implemented on 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 centroid of each detected object to track it. Closest objects within a distance of $4000km$ between two time steps are considered the same object progressing in time (Figure 4(d)). The annual mean frequency of the detected AR objects and their seasonality is 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).
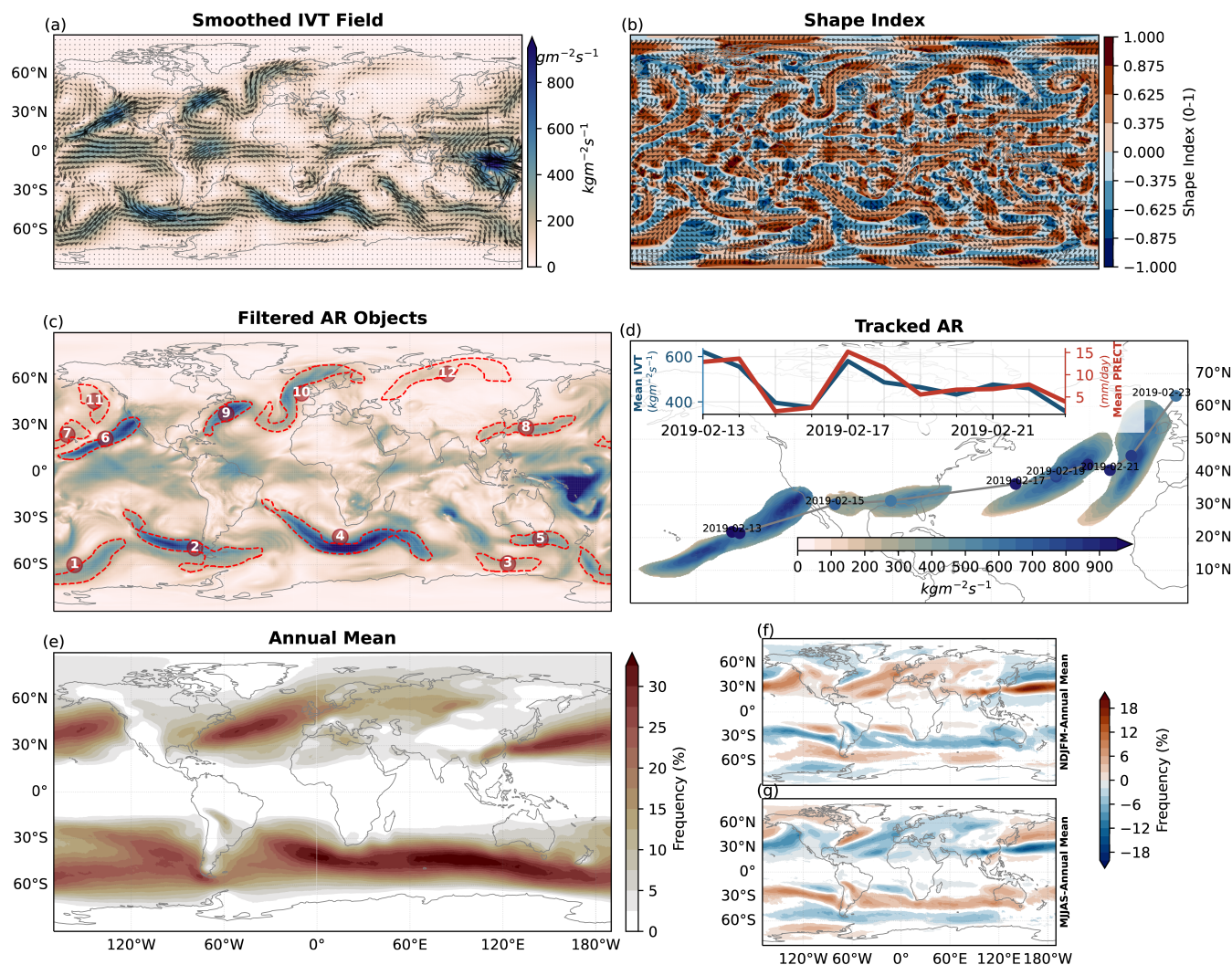
Figure 4. Major steps in the detection of Atmospheric Rivers. (a) is the smoothed primary field, which is the vertically integrated water vapor transport (IVT). The smoothing removes variability smaller than 1000 km from the IVT. The arrows in (a) represent the direction of unsmoothed IVT. (b) shows the magnitude (shading) of shape index (SI) and direction of the local ridge (arrows) calculated from (a). In the next step ridges, caps, and domes are extracted from (b) and weak and small AR candidates are filtered out. The AR objects after this filtering is shown in (c). Finally all the objects in (c) are tracked as shown in (d) to obtain tracks as well as other properties like mean IVT and precipitation. Objects that does not last more than one day is removed in this step. AR annual mean frequency for the period 2000 to 2019 is shown in (e), the anomaly for November to March (f), and May to September (g) relative to the annual mean are also plotted.

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

Table 1. The table shows the values of all the different parameters used in the detection of ARs. Rows for each step −segmentation, filtering, and tracking− are grouped together and labeled.

3.2 Tropical and Extratropical Cyclones

Cyclones are defined in the literature as large (500–4000 km) regions of strong cyclonic circulation with low pressure at the center and extremely 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. For instance,
220 cyclones formed near the equator (tropical cyclones) are in general smaller in area compared to that of the cyclones formed in midlatitudes (extratropical cyclones). Independent of this, they produce extremely high rainfall, and winds along the track and flooding, landslides, and severe damage to infrastructure along the coastlines where it makes landfall (Knutson et al., 2010; Mendelsohn et al., 2012; Ranson et al., 2014). With the rise in sea level and enhanced intensity of cyclones in response to warming, identification and future projection of cyclones is gaining a lot of attention from the climate community (Woodruff
225 et al., 2013).

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

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

Once again, discrepancies between the different detection algorithms can be traced back different choices of physical thresholds or functions on size, wind speeds, vorticity or surface pressure anomaly. Even though most studies converge on the conclusions for present and future characteristics of cyclones, ironing out the details such as the changes in genesis rate, dura-

tion are hindered by the uncertainties in the detection methods (Ulbrich et al., 2009; Neu et al., 2013; Horn et al., 2014; Walsh
230 et al., 2015). In this study, SCAFET identifies cyclones as regions of strong local maxima of cyclonic circulation (SI > 0.625)
with maximum wind speeds greater than $10 m/s$. This definition is able to identify strong cyclonic vorticities all over the globe
including, but not limited to, tropical and extratropical cyclones. The basic field used for cyclone detection is the absolute value
of cyclonic relative vorticity (ζ) defined as,

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

235 Where U is the 6 hourly wind speeds at 10 meters from surface obtained from ERA5 reanalysis with a spatial resolution
of $0.25° \times 0.25^{circ}$ (Hersbach et al., 2020). The magnitude of wind speed at 10 meters is used as the secondary field. Other
cyclone related variables like the surface pressure anomaly, and potential temperature can also be used as secondary field to
identify/classify cyclones.

In contrast with ARs, cyclones are detected using a scalar field, and in this case cyclonic relative vorticity. A grid-aware
240 gaussian smoothing is applied to suppress spatial variability smaller than 750 km (Figure 5(a)). The smoothing scale was
chosen so that we can identify both tropical and extratropical cyclones. Caps and dome shapes (SI > 0.625) from the smoothed
relative vorticity is identified as potential cyclones (Figure 5(b)). The next step is to filter out objects with area less than
$10^4 km^2$, and diameter less than $20 km$. Once the aforementioned spatial characteristics are fulfilled, we can further filter out
weak cyclonic vorticities (relative vorticity $< 10^{-6} s^{-1}$), and maximum wind speed $< 10 m s^{-1}$) giving us all the strong cyclonic
245 systems identified for a given time step (Figure 5(c)). All the processes described can be parallelized along the time dimension
as in the AR example. Once all potential cyclones are identified, they are tracked like AR tracking algorithm. However, the
radius for search is limited to $1000 km$ as we are using 6 hourly data and translations speeds of cyclones are much lower than
$150 km/h$. A minimum duration of 48 hours and a minimum total displacement of $500 km$ is applied to isolate propagating
cyclonic circulations from stationary ones. An example of a tracked cyclone, commonly known as cyclone Dorian (Lixion
250 A. Avila and Hagen, 2020) is compared with the observed track from IBTrACS dataset (Knapp et al., 2010, 2018) 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
speed thresholds. Another reason for the longer track is that, SCAFET does not distinguish between tropical and extratropical
cyclones and would end up following the vorticity while it transitions from tropical to midlatitude storms. The long-term
averages for cyclone frequency and its seasonal variability are comparable with other studies (e.g., Ullrich and Zarzycki, 2017).
255 Unlike other cyclone detection algorithms, SCAFET does not identify cyclones as a point object, but as a surface encompassing
the point of maximum ζ. This will help us study the nuanced properties of cyclones like the maximum and minimum values of
wind speed, precipitation within the whole cyclone structure.
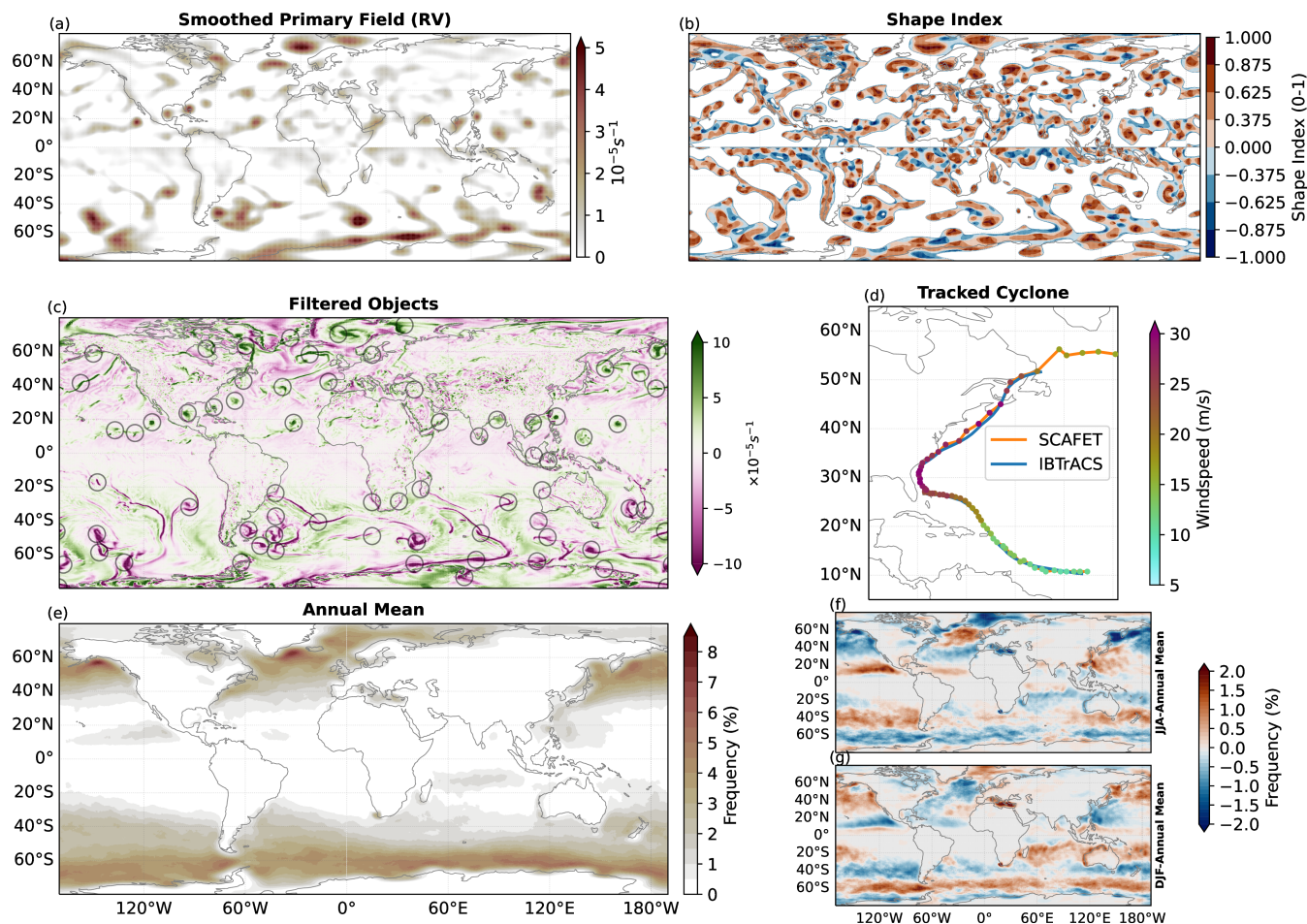
Figure 5. Major steps in the detection of cyclones. (a) Smoothed primary field, which in this case is the absolute value of cyclonic relative vorticity (ζ). The smoothing removes variability smaller than 750 kms from ζ. (b) Magnitude of SI calculated from (a). In the next step, caps, and domes are extracted from (b) and weak and small cyclone candidates are filtered out. Cyclonic vorticities after this filtering are shown in (c) 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 IBTrACS dataset (d). Objects that do not last more than 48 hours are removed in this step. The annual mean frequency cyclone occurrence for the period 2000 to 2020 is shown in (e), the anomalous cyclone frequencies for JJA (f), and DJF (g) relative to the annual mean are also plotted.

3.3 Sea Surface Temperature Fronts

SST fronts are the confluence regions of different water masses. They are often manifested as having strong horizontal gradi-
ents in temperature, salinity, density, and other characteristics (Bowman, 1978; Legeckis, 1978; Fedorov, 1986; Yoder et al.,
1994). Frontal structures are often observed in much smaller spatio-temporal scales than the other features described in this
study. Accurate identification of SSTFs are important as these features are often associated with strong upwelling, and high
biogeochemical productivity (Clayton et al., 2014, 2021; Nagai and Clayton, 2017). Identification of SSTFs also demonstrates
how SCAFET can be used to detect features in curvilinear grids.

Most previous frontal detection algorithms use edge detection algorithms and the gradient of sea surface temperature and/or
height and to identify fronts (Canny, 1986; Castelao et al., 2006). We use the magnitude of daily mean SST horizontal gradient
as the primary field for the detection of SST fronts. The SSTs were obtained from a fully coupled, ultra-high-resolution
($\approx 10km$) CESM v1.2.2 simulation of present day mean climate (Small et al., 2014; Chu et al., 2020). The data is fed into
SCAFET in the tripolar POP grid. To demonstrate the detection process, the analysis is confined to the Kuroshio frontal and
extension domain for the last 10 years of the simulation.

Frontal structures, ridges, caps, and domes are extracted very similarly as in the detection of ARs. A spatial smoothing of
approximately $30km$ is applied before extracting the frontal structures. From the extracted SSTF candidates, objects with a
mean SST gradient lower than $10^{-4} K/m$ is removed. So are circular (eccentricity < 0.5) and small (area $< 1000 km^2$) objects.
Unlike AR detection, SSTFs are not tracked as ocean fronts are stationary rather than transported into other regions. The
detected frontal frequency shows familiar patterns and seasonality as in previous studies (Xi et al., 2022).

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

Table 3. Same as in table 1 but for detecting Sea Surface Temperature Fronts (SSTFs)
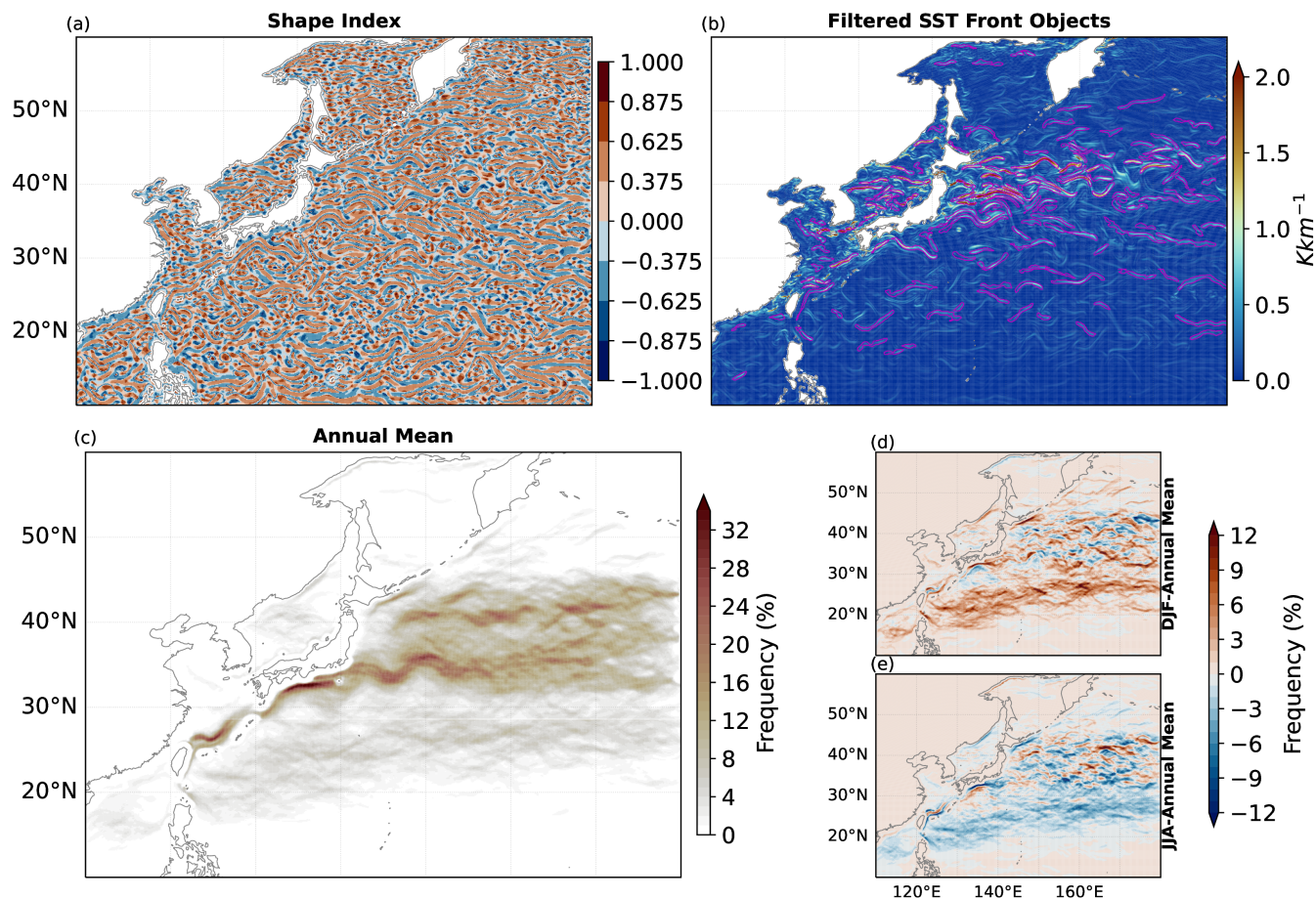
Figure 6. Major steps in the detection of sea surface temperature fronts (SSTFs). (a) Magnitude of shape index (SI) calculated from the smoothed primary field, which is the horizontal gradient of sea surface temperature (∇SST). The smoothing removes variability smaller than 15 km from ∇SST. In the next step, ridges, caps, and domes are extracted from (b) and weak and small SSTF candidates are filtered out. SSTFs after this filtering are shown in (b) with the background color representing the unsmoothed values of ∇SST. The annual mean frequency SSTF occurrence for the 10 years of the present climate simulation is shown in (c), the anomalous frontal frequencies for JJA (d), and DJF (e) relative to the annual mean are also plotted.

4 Application to 3D Features

In this section, we show how to extend SCAFET to detect features from three-dimensional (3D) primary fields. The scale-space selection is carried out by applying gaussian smoothing along the three dimensions separately. Also, a 3D basic field would yield three eigenvalues ($k_1 \geq k_2 \geq k_3$) rather than two. Here, SI can be calculated by combining the eigenvalues in three different ways. The SI calculated using k_1 and k_2 is used for the extraction of jet streams as it provides a more conservative estimate for the jet domain (see Supplementary Figure S1).

4.1 Jet Streams

Independent of the dynamics, jet streams are manifested as narrow regions in the upper atmosphere with relatively high wind speeds compared to its surroundings (Koch et al., 2006). Apart from the obvious direct impact on aviation, the location and characteristics of jet streams strongly influence surface weather conditions. For instance, a persistent jet in boreal summer can lead to extreme heat and flooding events while a meandering jet in the winter induces extreme cold spells in the midlatitudes (Petoukhov et al., 2013; Coumou et al., 2014; Kretschmer et al., 2016). Further, the northward movement of jet streams in response to greenhouse warming leads 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 and extreme weather systems. Similar to the detection of other weather phenomenon discussed in this study, most previous studies use a physical threshold in identifying jet locations. Moreover, most of these studies except Limbach et al. and Kern et al. identifies jet streams as either a 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 detection in three selected time steps. A video showing the results for a longer period can be seen in the supplementary section.

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

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

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

Comparable to detection of 2D features, the detection process starts by applying a gaussian smoothing to remove wavelengths less than $6000km$ in the horizontal dimensions. No smoothing is applied along the vertical dimension. Next, SI is calculated using the two largest eigenvalues k_2 and k_3. Regions corresponding to ridges, caps, and domes (SI > 0.375) are isolated for filtering. Filtering removes objects with volume less than $1000km^3$, horizontal length less than $5000km$, and maximum wind speed within each object less than $50m/s$. In the current version of SCAFET, the tracking algorithm is not applied on jet detection (see Figure 7).
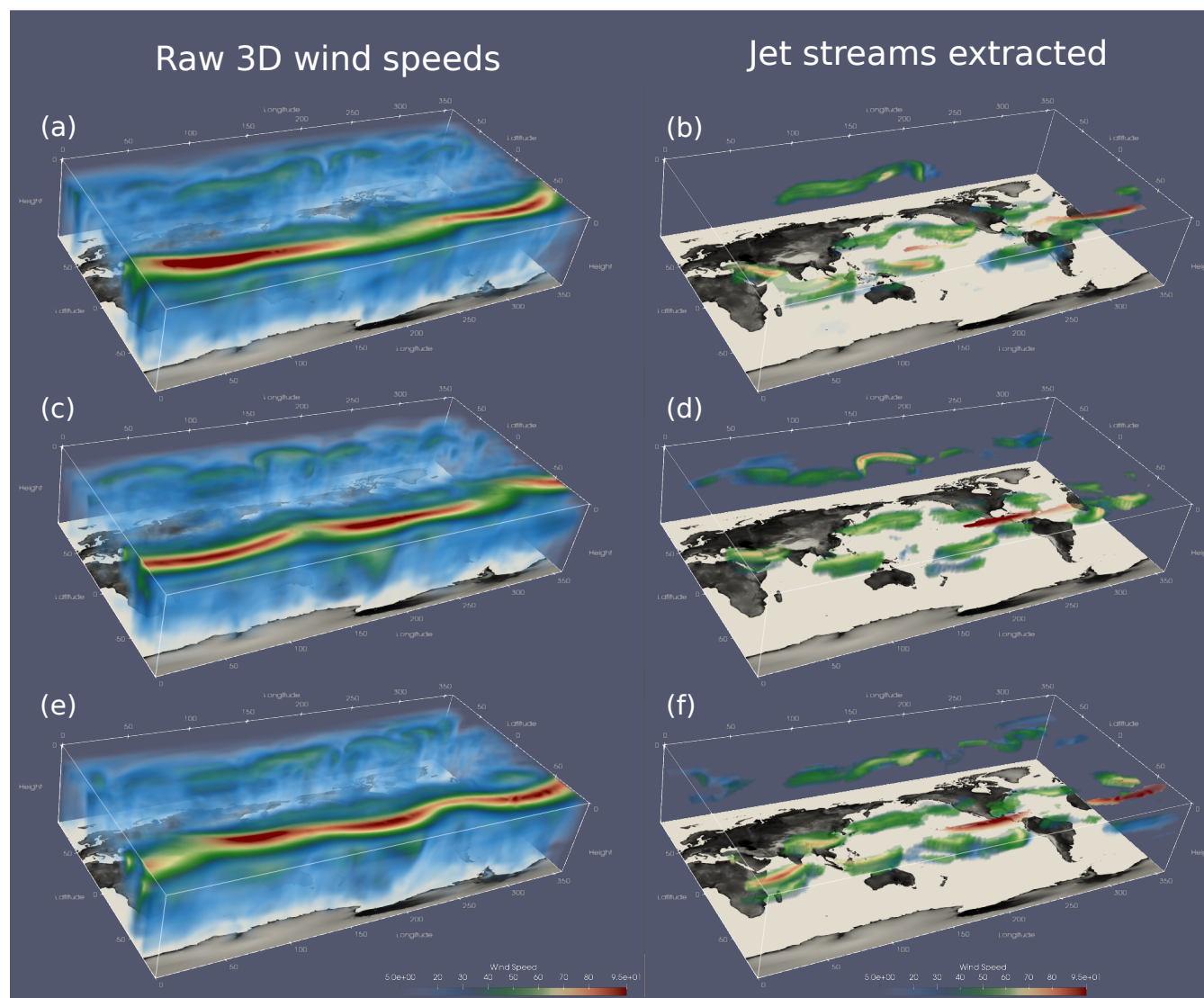
Figure 7. 3D jet streams extracted using SCAFET. (a), (c), (e) shows the magnitude 3D wind speed for 2022-08-25 00:00, 2022-08-28 06:00, and 2022-08-31 18:00 respectively. The 3D jet streams extracted for the corresponding time period is show in (b), (d), and (e) respectively.

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

Table 4. Same as in Table 1 but for Jet Streams.

5 Conclusions

In this study we introduced a new framework and algorithm package for extracting and tracking meso-synoptic scale features from large climate datasets, called Scalable Feature Extraction and Tracking (SCAFET). As the volume and diversity of ob-
servational and model climate data grow, an alternative method to physical threshold-dependent feature detection is necessary to compare features within and between observational and model data sets. Furthermore, a novel shape-based approach for feature extraction will give us further insights into detection method discrepancies in projections and aid the community in building scientific consensus. To demonstrate the ability of SCAFET in advancing these goals, we illustrated 2D detection of atmospheric rivers (ARs), tropical and extratropical cyclones, sea surface temperature fronts, and 3D detection of jet streams. Each application was intended to give characteristic examples from which users can customize SCAFET for their own research purposes.

Apart from the obvious benefits like a more generalized framework and parallelized implementation, SCAFET, more importantly, provides a novel perspective on how we could define various features in climate datasets covering large periods of time, in which the mean climate varies significantly. Instead of extracting features from physical thresholds, 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 the continuous changes in feature properties that account for mean state changes. Since results of meso-synoptic scale studies are sensitive to thresholds in a varying mean state, conclusions inherently depend on the feature extraction method. Such varying conclusions are noticed while studying the response of ARs to greenhouse warming Zhao (2020); O'Brien et al. (2022). Thus physical threshold-independent algorithms such as SCAFET may be crucial in furthering scientific understanding and climate model development.

Undertaking a more fundamental level research into differential geometry and mathematical derivation of the relationship between SI and local geometric shape could transform the way we identify extreme events 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.

Author contributions. ABN wrote the software package and prepared the manuscript draft with inputs from co-authors. TAO and DL was 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, DL, and JEC contributed equally on 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
340   downloading ERA5 data, readers could refer to *ERA5Data* folder in the zenodo repository. For any further details on code and data, feel free to contact the first 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. O'Brien's
345   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, and Dr. Pavan Harika Ravi from IBS Center for Climate Physics for their comments on shape based feature extraction. We would also like to thank the ARTMIP community for their feedbacks on SCAFET. Part of the analysis of the detected features were conducted on the IBS/ICCP
350   supercomputer "Aleph" 1.43 peta flops 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. 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 how shape-based matrices can be used to extract features from simple mathematical
355   functions. This section is indented to provide readers with more insights into the basic principle behind shape-based feature extraction and how it differs from other published methods. We have also tried to demonstrate some properties of shape-based feature extraction methods like its insensitivity to mean state changes and linear trends.

A1   Application to 1D datasets

In this section, we are constructing an analogy between application of SCAFET on a 2D dataset and shape-based feature
360   extraction from a one-dimensional dataset. The purpose of this discussion is not to advocate for a shape-based extraction of features from 1D datasets but to help the readers understand its strengths and weaknesses.

   Conventionally, 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
365   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
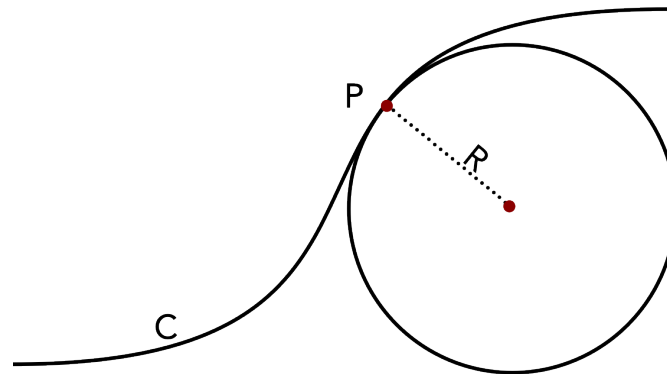
Figure A1. A schematic representation of measuring curvature of a curve C at point P. At P the curvature is the reciprocal of the radius R of the osculating circle. In differential geometry, an osculating circle is defined as the circle passing through the point P and a pair of additional points infinitesimally close to P.

or a convex curvature. More information and mathematical proof for these concepts can be found in any standard differential geometry textbooks.

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

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

370

Values of K closer to 1 can be 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, one could choose a value for K to get 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 as the application of the same shape thresholds identifies identical regions as local maxima

375 and minima in a simple trigonometric curve (blue curve in Figure A2) and the same curve with an added linear trend (orange curve in Figure A2). The values of K for both curves are represented by the black line inFigure A2. Thus, it can be used to identify extreme events from datasets without being affected by the background state changes.

A2   Application of SCAFET to simple Geostrophic Motion

In this section we demonstrate the application of SCAFET to a simple geostrophic rotational motion. The goal of this discussion
380 is to see how shape-based extraction of 2D features differ from other conventional methods used. The calculation of SI involves the computation of the two eigenvalues, k_1 and k_2 of the hessian of any gridded dataset. As discussed in the previous section, the measurement of curvature, k_1 and k_2 can be visualized as the reciprocal of the radius of two osculating circles orthogonally intersecting at a point in the surface. Large negative eigenvalues represent surfaces with strong convex curvature while positive
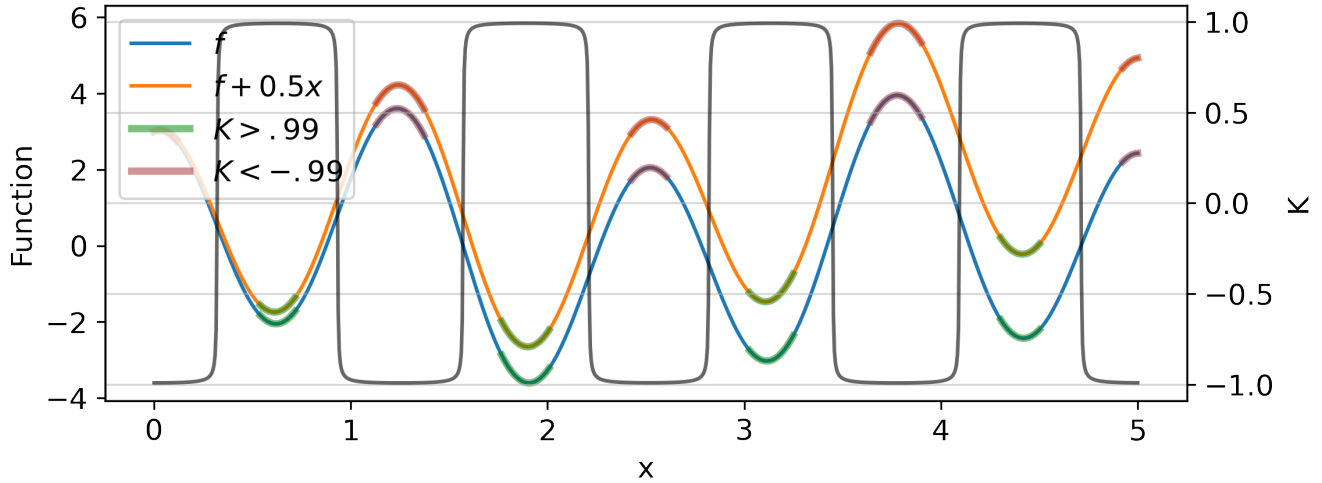
Figure A2. Shape extraction for a simple one-dimensional curves given by $f = sin2x + 3cos5x$ and $f + 0.5x = sin2x + 3cos5x + 0.5x$. The first one is a simple trigonometric curve while the second curve includes a linear trend as is evident from their functional forms. The magnitudes of both the functions are shown in the left Y axis while the right Y axis indicates the values of the shape parameter (K). Note that K is same for both the functions. The green and red highlighting on the curves indicates the regions where $K > 0.99$ and $K < -.99$. The red highlighted are tagged as extreme maxima events and those highlighted with green are be characterized as extreme minima events.

values are identified as troughs or cups. To demonstrate the characteristics and strengths of feature detection based on Shape
385   Index (SI) let's consider a simple rotation 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 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
390   estimated as,

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

f and g are the Coriolis parameter and the gravitational constant respectively. SI is calculated from the eigenvalues of the hessian of h using the formula,

$$SI = \frac{2}{\pi}\arctan\left[\frac{k_2 + k_1}{k_2 - k_1}\right] \tag{A5}$$

Where the eigen values 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,

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

A detailed derivation of the above equation can be found in Appendix B. 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,

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

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 the geopotential height would identify regions depending on the value of the threshold on h. However, the value of the threshold must be adjusted depending on the mean (time) and background (space) state. Another widely used practice is to define a threshold on the smallest eigenvalue. The intention of such methods is to identify extreme features based on the strength of the curvature rather than the actual value of the field. TempestExtremes, a feature extraction framework previously discussed, follow this method to identify Atmospheric Rivers from gridded datasets. In the current example, this would correspond to thresholding on $f\Omega/g$. In other words, TempestExtremes would only detect the trough if the value of Ω is greater than the predetermined threshold. Contrastingly, SCAFET would identify the trough as a trough regardless of the actual value of the field or Ω. Hence, we see that feature extraction using SI and other published methods can give us different results depending on the input data as they are looking at different properties of the field.
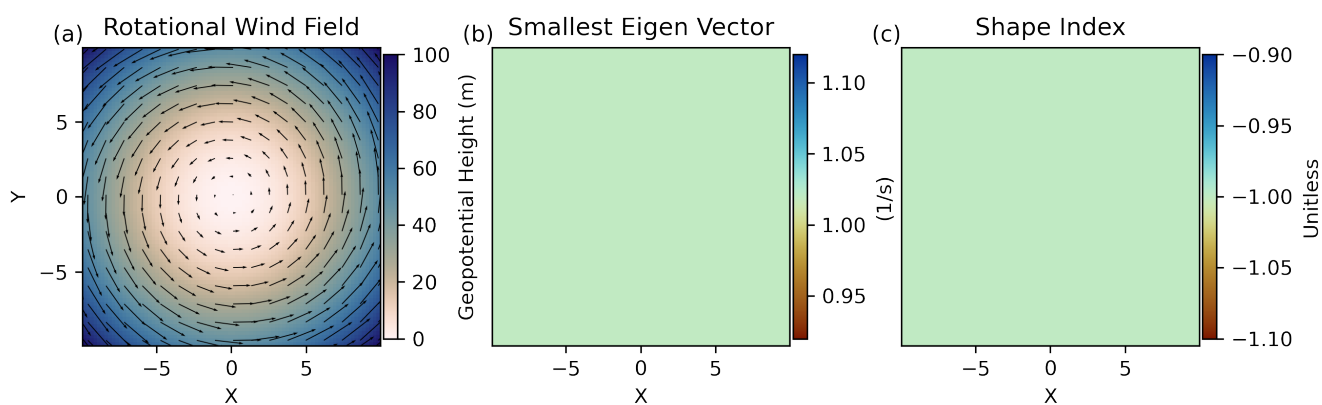
Figure A3. Comparing different feature extraction techniques on synthetic rotational wind field. (a) The geopotential height (h) (shading) of the rotational wind field (arrows). h is defined as $\Omega f(x^2 + y^2)/2g$, where $f = 10^{(}-4)1/s$, $g = 9.805m/s$ and $\Omega = 10^5 rad/s$. (b) The magnitude of the smallest eigen value. From the equation, the smallest eigen value can be derived as $f\Omega/g = 1.0199$. (c) Shows the value of the shape index (SI). From the equations and the plot, we can see that the value of $SI = -1$ throughout the domain.

Appendix B: Derivation of Shape Index for Geostrophic Motion

The complete derivation of the SI for geostrophic wind fields are shown in this section. The result from the derivation is used in the previous section. Let h be the geopotential height at a certain level. The hessian of h is given by.

$$420 \quad \mathcal{H}(h) = \begin{pmatrix} \frac{\partial^2 h}{\partial x^2} & \frac{\partial^2 h}{\partial x \partial y} \\ \\ \frac{\partial^2 h}{\partial y \partial x} & \frac{\partial^2 h}{\partial y^2} \end{pmatrix} \tag{B1}$$

The eigen values 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 can be expanded as;

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

$$425 \qquad\qquad \lambda^2 - \lambda \nabla^2 h + \frac{\partial^2 h}{\partial x^2} \cdot \frac{\partial^2 h}{\partial y^2} - \left(\frac{\partial^2 h}{\partial x \partial y} \right)^2 = 0 \tag{B4}$$

NOTE: The geostrophic vorticity (ζ_g) is defined as

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

The geostrophic velocities are defined as

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

$$430 \quad v_g = \frac{g}{f} \frac{\partial h}{\partial x} = \frac{\partial \psi}{\partial x} \tag{B7}$$

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)

$$435 \quad \lambda^2 - \frac{\lambda f}{g} \zeta_g - \frac{f^2}{g^2} \frac{\partial v_g}{\partial x} \frac{\partial u_g}{\partial y} + \frac{f^2}{g^2} \frac{\partial u_g}{\partial x} \frac{\partial v_g}{\partial y} \tag{B10}$$

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

440 $$SI = \frac{2}{\pi}arctan\left[\frac{\zeta_g}{-2\sqrt{\left(\frac{\zeta_g}{2}\right)^2 - \frac{\partial v_g}{\partial x}\frac{\partial u_g}{\partial y} + \frac{\partial u_g}{\partial x}\frac{\partial v_g}{\partial y}}}\right] \tag{B13}$$

References

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, 12331–12336, https://doi.org/10.1073/pnas.1412797111, 2014.

Dixon, M. and Wiener, G.: TITAN: Thunderstorm Identification, Tracking, Analysis, and Nowcasting—A Radar-based Methodology, 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, 12514–12535, https://doi.org/10.1002/2015jd024257, 2015.

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

480 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, 485 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/10.1175/1520-490 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.

495 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.

500 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.

505 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.: 510 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.

515   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, Geosci-
520   entific 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.

Lixion A. Avila, Stacy R. Stewart, R. B. and Hagen, A. B.: Hurricane Dorian, Tech. rep., National Hurricane Center, https://www.nhc.noaa
.gov/data/tcr/AL052019_Dorian.pdf, 2020.

Lora, J. M., Shields, C. A., and Rutz, J. J.: Consensus and Disagreement in Atmospheric River Detection: ARTMIP Global Catalogues,
525   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.

530   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
535   of Geophysical Research: Atmospheres, 123, 6804–6821, https://doi.org/10.1029/2017jd028130, 2018.

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
540   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
545   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 Hemi-
sphere weather extremes, Proceedings of the National Academy of Sciences, 110, 5336–5341, https://doi.org/10.1073/pnas.1222000110,
550   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.

555 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
560 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
565 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,
570 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):
575 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,
580 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.

585 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.

590 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.

595 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.

600 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.

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.,
605 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.

610 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,
615 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.