Institute of
Computer Science

JG|U
JOHANNES GUTENBERG
UNIVERSITÄT MAINZ

Identification, Tracking and Analysis of 3-D Meteorological Features

Dissertation

submitted for the award of the title

“Doctor of Natural Sciences”

to the Faculty of Physics, Mathematics, and Computer Science
of Johannes Gutenberg University Mainz
in Mainz

Christoph Fischer

Born in Oberkirch

Mainz, June 18, 2024

Christoph Fischer
Identification, Tracking and Analysis of 3-D Meteorological Features
Dissertation, June 18, 2024
Date of oral examination: November 22, 2024
Reviewers: [name removed] and [name removed]

Johannes Gutenberg University Mainz
Computational Geometry
Institute of Computer Science
FB08
Staudingerweg 9
55128 Mainz

Abstract

Meteorological data sets, which detail the state of the Earth's atmosphere, have become increasingly complex and diverse over the past decades. The growing volume and quality of these data have created a demand for advanced tools capable of efficiently handling and analyzing large, multidimensional atmospheric data sets. This dissertation meets this demand by combining computer science techniques with meteorological research to develop practical software tools and novel algorithms. These tools aim to help atmospheric scientists extract valuable insights from vast data sets through efficient feature identification, tracking, and analysis.

The main focus of this work is the development of new algorithms for identifying and tracking meteorological features in three dimensions. By utilizing algorithmic geometry, this dissertation introduces low-dimensional, geometry-based descriptors that simplify the analysis of wave-related features such as tropical waves and anomalies along the dynamical tropopause. These descriptors significantly reduce the complexity of the data while preserving important information, making the analysis more efficient and potentially improving the predictability of weather events linked to these features.

A key aspect of this dissertation is the application of the Potential Vorticity (PV) framework, a well-established method for studying atmospheric dynamics. We introduce a novel algorithm for identifying and tracking 3-D PV anomalies along the tropopause, offering new insights that traditional 2-D methods might overlook and addressing the associated challenges. Another application involves identifying and tracking PV features associated with African easterly waves (AEWs). These tropical waves are known to be linked to tropical cyclone development and contribute to high weather prediction uncertainty in tropical West Africa.

To support these advancements, a dedicated software framework has been developed, tailored specifically to the needs of meteorologists. This framework allows for the easy implementation of the proposed algorithms, with configurable options and templates for straightforward feature identification and tracking. Additionally, a near-real-time web display has been created to

visualize identified and tracked AEWs, which evolved into a valuable tool for both researchers and operational forecasters.

Summarized, this dissertation bridges the gap between computer science and meteorology, showing how advanced computational techniques can improve the understanding and predictability of complex atmospheric phenomena. The developed tools and algorithms provide a solid foundation for future research and practical applications. They pave the way for more accurate weather predictions and a deeper understanding of meteorological processes.

Zusammenfassung

Meteorologische Datensätze, die den Zustand der Erdatmosphäre beschreiben, sind in den letzten Jahrzehnten zunehmend komplexer und vielfältiger geworden. Das wachsende Volumen und die steigende Qualität dieser Daten haben die Nachfrage nach fortschrittlichen Werkzeugen erhöht, die in der Lage sind, große, mehrdimensionale atmosphärische Datensätze effizient zu verarbeiten und zu analysieren. Diese Dissertation erfüllt diese Anforderungen, indem sie Techniken der Informatik mit meteorologischer Forschung kombiniert, um praktische Softwarewerkzeuge und neuartige Algorithmen zu entwickeln. Diese Werkzeuge sollen Meteorologen helfen, wertvolle Erkenntnisse aus umfangreichen Datensätzen durch effiziente Merkmalserkennung, -verfolgung und -analyse zu gewinnen.

Der Schwerpunkt dieser Arbeit liegt auf der Entwicklung neuer Algorithmen zur Identifizierung und Verfolgung meteorologischer Merkmale in drei Dimensionen. Durch die Nutzung algorithmischer Geometrie führt diese Dissertation niedrigdimensionale, geometriebasierte Beschreibungen ein, die die Analyse von wellenbezogenen Merkmalen wie tropischen Wellen und Anomalien entlang der dynamischen Tropopause vereinfachen. Diese Beschreibungen reduzieren die Komplexität der Daten erheblich, während wichtige Informationen erhalten bleiben, was die Analyse effizienter macht und potenziell die Vorhersagbarkeit wetterbedingter Ereignisse, die mit diesen Merkmalen verbunden sind, verbessert.

Ein zentraler Aspekt dieser Dissertation ist die Anwendung des Potential Vorticity (PV)-Frameworks, einer bewährten Methode zur Untersuchung atmosphärischer Dynamiken. Wir stellen einen neuartigen Algorithmus zur Identifizierung und Verfolgung dreidimensionaler PV-Anomalien entlang der Tropopause vor, der neue Einblicke bietet, die traditionelle zweidimensionale Methoden möglicherweise übersehen, und die damit verbundenen Herausforderungen in Angriff nimmt. Eine weitere Anwendung umfasst die Identifizierung und Verfolgung von PV-Merkmalen, die mit afrikanischen Ostwellen (African easterly waves, AEWs) verbunden sind. Diese tropischen Wellen sind bekanntermaßen mit der Entwicklung tropischer Wirbelstürme

verbunden und tragen zu hoher Unsicherheit bei Wettervorhersagen in Westafrika bei.

Um diese Fortschritte zu unterstützen, wurde ein spezielles Framework entwickelt, das auf die Bedürfnisse von Meteorologen zugeschnitten ist. Dieses Framework ermöglicht die einfache Implementierung der vorgeschlagenen Algorithmen mit konfigurierbaren Optionen und Vorlagen für eine unkomplizierte Merkmalserkennung und -verfolgung. Zusätzlich wurde eine nahezu in Echtzeit arbeitende Webanzeige erstellt, um identifizierte und verfolgte AEWs zu visualisieren, die sich zu einem wertvollen Werkzeug sowohl für Forscher als auch für operative Wettervorhersagedienste entwickelt hat.

Zusammenfassend überbrückt diese Dissertation die Lücke zwischen Informatik und Meteorologie und zeigt, wie fortschrittliche Rechentechniken das Verständnis und die Vorhersagbarkeit komplexer atmosphärischer Phänomene verbessern können. Die entwickelten Werkzeuge und Algorithmen bieten eine solide Grundlage für zukünftige Forschung und praktische Anwendungen. Sie ebnen den Weg für genauere Wettervorhersagen und ein tieferes Verständnis meteorologischer Prozesse.

Contents

1 Introduction

> *The most fruitful areas for the growth of the sciences were those which had been neglected as a no-man's land between the various established fields. [...] Science has been increasingly the task of specialists, in fields which show a tendency to grow progressively narrower. [...] It is these boundary regions which offer the richest opportunities to the qualified investigator.*
>
> — **Norbert Wiener, 1961**
> Cybernetics: Or Control and Communication
> in the Animal and the Machine

The landscape of scientific exploration has transformed, no longer confined by the boundaries of individual disciplines. As Norbert Wiener noted, the most fruitful areas for scientific growth lie at the crossroads of established fields. This is particularly evident in the intersection of computer science and meteorology, a realm where the complexities of atmospheric phenomena intersect with the capabilities of cutting-edge computational tools, adding both opportunities but also arising challenges.

In this context, the collaboration between computer science and meteorology has become increasingly crucial. Meteorology, a multidisciplinary domain, requires a profound understanding of physics, mathematics, and atmospheric science. However, it is through collaboration with computer science that meteorology gains the ability to reveal hidden patterns, predict weather events with precision, and derive actions to safeguard communities and resources based on statistical analyses.

This dissertation explores the interplay between computer science and meteorology, addressing challenges in identifying and understanding meteorological phenomena, referred to as features. The focus is on the identification, analysis, and tracking of these features, developing novel solutions that meet the demands of both fields. Going beyond traditional approaches, the investigation extends into 3-D space, providing valuable insights while introducing additional challenges. The dissertation centers around two peer-reviewed scientific publications published in the journal *Geoscientific Model Development* (GMD), showcasing distinct strategies for identifying atmospheric features.

1.1 Motivation and Problem Statement

This dissertation is motivated by the close collaboration between computer science and meteorology, with the goal of efficiently identifying and understanding meteorological features. The goal is to leverage the potential arising from the convergence of these two fields and develop innovative solutions for the dynamic and complex atmospheric realm.

At the heart of our motivation is the aspiration to enhance the efficiency of identifying and tracking meteorological features. By exploring algorithmic geometry, we aim to create streamlined descriptions of predictors, with a specific focus on developing low-dimensional, geometry-based descriptors for wave-related features. These descriptors significantly reduce dimensionality, streamlining the recognition process. The methodology involves defining the locations and extents of features using simple two- or three-dimensional geometric shapes. Additionally, our approach includes the tracking of connected components, contributing to the establishment of climatologies of features. The dissertation specifically investigates features influencing tropical cyclone development, including tropical waves and anomalies along the dynamical tropopause.

1.2 Thesis Structure

The subsequent chapter explores foundational aspects of computer science, establishing the groundwork for feature identification and tracking. It explores the conceptualization of features, drawing from classical computer science notions and tailoring them to the unique characteristics of meteorological phenomena. Various approaches to identifying and tracking these features are discussed, along with methods for their representation in a low-dimensional manner to facilitate efficient analyses.

Moving to meteorology in Chapter 3, we introduce relevant atmospheric processes, specifically exploring the vorticity measure and the Potential Vorticity framework. Furthermore, we will introduce the dynamics in the lower troposphere in tropical West Africa, specifically the formation of the African easterly jet (AEJ) and the associated African easterly waves (AEWs). These concepts are crucial for the design decisions made in the upcoming feature identification algorithms.

Recognizing the need for a dedicated software framework for efficient feature identification and tracking, Chapter 4 outlines the design decisions behind the software developed for this purpose. The chapter emphasizes tailoring the software to meet the specific requirements of meteorologists. We will showcase different configuration options and use cases.

Chapters 5 and 6 demonstrate the practical applications of the interdisciplinary framework. Chapter 5 draws from the peer-reviewed publication Fischer et al. [49], identifying 3-D Potential Vorticity features along the tropopause. Chapter 6 focuses on the identification, tracking, and analysis of 3-D PV features associated with AEWs, published in Fischer et al. [48].

Chapter 7 concludes the thesis, summarizing contributions and reflecting on advancements in the identification, analysis, and tracking of meteorological phenomena. The chapter also provides an outlook on future avenues in this dynamic field.

Atmospheric features

2

> *A picture is worth a thousand words. An interface is worth a thousand pictures.*
>
> — **Ben Shneiderman**
> (Professor for Computer Science)

The world of meteorology is not just about forecasting weather; it is also a treasure trove of data, ripe for exploration and analysis from a computer science perspective. In this chapter, we will address feature identification and tracking within meteorological data sets in depth: a vital intersection between computer science and atmospheric sciences.

Meteorological data are no longer confined to hand-drawn weather maps and manual observations. They are complex and multidimensional digital entities closely linked to physical principles and real-world processes. To extract valuable insights and make informed decisions, we need to identify and track specific phenomena within these data. These phenomena, often referred to as *atmospheric features*, can range from cyclones and high-pressure systems to more abstract atmospheric structures based on derived fields or wave patterns.

We begin with the fundamental aspects crucial to this thesis, gradually bridging the gap between computer science and atmospheric sciences.

2.1 Meteorological data sets

Meteorological data encompass a wide range of information, collected from various heterogeneous data sources, for example observations from weather stations, satellites and radars, and outputs from numerical weather Prediction (NWP) models. These sources differ strongly in resolution, accuracy, availability, and parameters. Observations, often recorded as scattered data points, contrast with NWP models, which operate on diverse grids, each with unique characteristics. For example, the German Weather Service (DWD) employs the ICON grid, based on an Icosahedral structure [207], while the Integrated Forecasting System (IFS) at the European Centre for Medium-Range Weather Forecasts (ECMWF) deploys various grids, such as triangular and Gaussian grids [197]. Post-simulation, model outputs are typically resampled onto more intuitive grids, frequently in Geographic Coordinate Systems (GCS) based on longitude and latitude in the horizontal domain.

Regarding the vertical domain (along the z direction), meteorological data sets can have vastly different representations. The default output of weather models is based on model levels, which mimic the surface orography at lower levels (the ones near the surface) and become more horizontally aligned at higher levels. Subsequent resampling often occurs onto levels of constant atmospheric pressure (isobars) or levels of constant potential temperature (isentropes). Beyond these spatial dimensions, data may include additional dimensions. Both observations and NWP outputs incorporate a temporal component, introducing time as an extra dimension. Leading-edge NWP models also operate with ensemble data, generating multiple forecasts with diverse initial conditions, offering insights into predictability and atmospheric state uncertainty. And, of course, atmospheric data is multivariate: both measurements and output of numeric models contain a wide array of variables, including temperature, wind, humidity, aerosols, and many more. Sophisticated data processing techniques need to adapt to all these dimensions.

It is worth highlighting the recent strides made by weather services in introducing open data policies. These policies have significantly enhanced the transparency and accountability of meteorological research, enabling easier access to observations and forecast data for researchers and the general public. Furthermore, remarkable progress in standardization, driven by organizations like the World Meteorological Organization (WMO) and global climate and weather laboratories, led to more reliable and consistent measurements. The establishment of the Climate and Forecast Metadata Conventions (CF conventions) [71] has resulted in a comprehensive catalog of data types, formats, and associated metadata. Coupled with the standardization of specific file formats for storing and exchanging data (mostly netCDF [@128] and GRIB [@203]), this greatly facilitates global collaborations and data accessibility. This evolution over the past few decades stands as a remarkable success story from the perspective of data analysts and meteorological researchers.

Fig. 2.1: A projection of the Earth by Sanson and Jaillot's in 1691, a decorative masterpiece of both hemispheres, embodies the 17th-century worldview. The aesthetics have been prioritized over geographical accuracy. (From Jaillot, H., Atlas Nouveau, 1691 edition)

2.1.1 Map projections

Snyder [171] defined a map projection as *a systematic representation of all or part of the surface of a round body, especially the Earth, on a plane*. The earliest theories on projections date back to the second century when the Greek mathematician, astronomer, and geographer Claudius Ptolemy worked on the *Geographike Hyphegesis* ("Guide to Drawing the Earth"), discussing the drawing of maps in geographical coordinates for parts of the known Roman World. Many of the projections we use today started to gain attention in the 16th and 17th centuries. Figure 2.1 shows one such projection, which would be today referred to as a *Nicolosi globular projection*. By well-defined mappings, data can be transformed back and forth between projections. Meteorological data sets are always provided using a defined map projection. A map projection can be thought of a function T that maps the naturally parametrized variables of the globe, namely longitude λ and latitude ϕ, to a 2-D xy-plane:

$$T(\lambda, \phi) = (x(\lambda, \phi), y(\lambda, \phi)). \tag{2.1}$$

Here, the question arises: which projection is the *best*, especially to bridge the gap between computer science and meteorology, and to perform operations on gridded data sets? As Snyder [171] pointed out, there is no best projection due to the inherent distortions that occur when projecting a sphere onto a plane. Therefore, scientists must prioritize certain characteristics relevant to the application at hand and choose a projection that favors these characteristics at the expense of others, leading to compromises. Figure 2.2 shows a small selection of regularly used projections in science and visualization. Often, so-called circular *Tissot indicatrices* are drawn on maps (yellow shapes in Fig. 2.2) to illustrate the spatial change in distortion. The most important characteristics of projections, which might be more or less crucial depending on the application, are:

- *Preserving areas.* Equal-area projections maintain the property that any shape defined by a boundary at any point on the projection covers the same area as it would on the actual sphere. However, these projections distort shapes, angles, and scales. The Albers projection illustrated in Fig.2.2d exemplifies this characteristic, where the Tissot indicatrices exhibit varying shapes while maintaining their area.

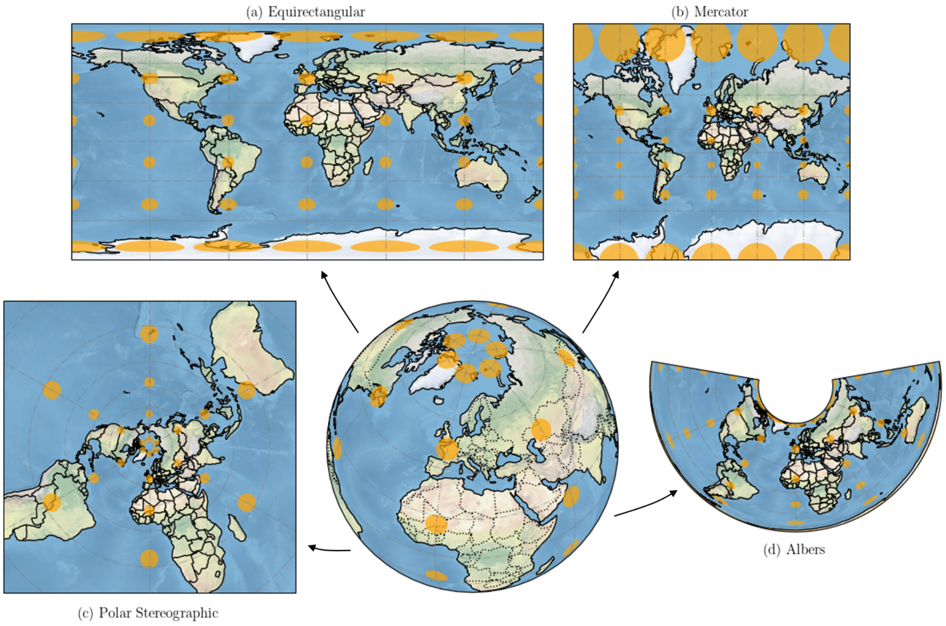
- *Preserving shapes.* Conformal projections preserve local angles at any point. Thus, two lines intersecting on the sphere with an angle α will also do so in the projection. Furthermore, the local scale in every direction around any given point remains constant, and circles on the sphere project to circles in the projection, but of different areas. The stereographic projection depicted in Fig.2.2c has this property.

- *Preserving scales.* These projections maintain correct scale throughout the map, although this is only feasible on regional maps or along certain lines, not on a global scale.

It is impossible to construct a map projection that is both equal-area and conformal, because the sphere is not a developable surface [171].

In meteorology, the predominant standard for providing data sets is the equirectangular projection (Fig. 2.2a), which is equivalent to a grid consisting of longitude λ and latitude ϕ dimensions with a constant grid spacing $\Delta\lambda$ and $\Delta\phi$. The projection function T_{Eq} mapping λ, ϕ to x, y can quite simply be expressed by

$$T_{Eq}(\lambda, \phi) = (x(\lambda, \phi), y(\lambda, \phi)) = (\lambda, \phi), \tag{2.2}$$


(a) Equirectangular

(b) Mercator

(c) Polar Stereographic

(d) Albers

Fig. 2.2: Different map projection of the earth, including Tissot indicatrices as yellow shapes. Shown is (a) the neither equal area nor conformal equirectangular projection; (b) the conformal Mercator projection; (c) the conformal stereographic projection; and (d) the equal-area Albers projection. We refer to Snyder [171] for a mathematical definition of all projections.

if we choose the point defined by 0°N, 0°E as projection center. Despite lacking any of the above-mentioned characteristics, the equirectangular projection's simplicity in format and storage, intuitiveness, and compatibility with legacy and diverse geographic information systems (GIS) make it the predominant data format for global meteorological data. However, for data analysis tasks, it may be advisable to transform the data onto a different projection more suited to the specific tasks and domains of interest. In chapters 5 and 6 we use the following projections:

- A stereographic projection is used in Chapter 5 and depicted in Fig. 2.2c. This projection maps circles on the sphere to circles in the projection and is primarily used to display polar maps with either the North Pole or South Pole in the projection center. More details *why* this projection is vital for the identification method can be found in Sect. 5.5.1. Conceptually, the projection is constructed by projecting points from a sphere onto a plane, typically done from one of the poles, where the projection point is directly opposite to the point of tangency on the sphere. In polar coordinates, the projection $T_{Eq}(\lambda, \phi)$ can be computed by [171]

$$T_{Eq}(\lambda, \phi) = 2\tan(\frac{\pi}{4} - \frac{\phi}{2})(\cos\lambda, \sin\lambda). \tag{2.3}$$

Drawbacks are singularities at the opposing pole and extreme distortions far from the projection center, but these areas will never be considered and analyzed due to physical constraints in this specific application.

- An equirectangular projection is used in Chapter 6. The features identified in that chapter are located over tropical West Africa and the tropical Atlantic basin, mainly between 5° and 20°N (see Sect. 3.4 and Burpee [21]). Despite major distortions and singularities near the poles, this projection reflects areas near the equator reasonably precise. A degree of latitude and longitude at the equator corresponds to around 111 km, and at 20°N, a degree in longitude is still 104.5 km, resulting in a distortion error of not more than 6%. For tropical West Africa, this projection strikes a reasonable compromise between the simplicity and availability of the data fields and the induced distortion error.

2.1.2 Meteorological features

Meteorological features refer to various observable and distinct patterns or phenomena in the Earth's atmosphere, occurring across a range of spatial and temporal scales. These features can be as large as large-scale atmospheric systems like Rossby waves [159] or as small as short-lived convective bursts. Additionally, they can span from multi-year global oscillating patterns like *El Niño* to localized short-lived wave anomalies. If the features are defined by specific regions with a well-defined boundary, we also call them *objects* in the scope of this work. These objects are often defined based on case-specific criteria and thresholds, and can be considered as a (not necessarily coherent) set of voxels in 3-D space.

To analyze and compare meteorological features, it is essential to describe their properties in a structured manner. This is achieved through a *feature description*, also referred to as a *feature vector* in this work. Feature vectors are crucial in representing complex meteorological patterns in a low-dimensional format, enabling statistical analysis and feature comparisons. Feature vectors aim to possess properties such as completeness, congruence, invariance, and compactness, as highlighted by Nixon and Aguado [126]. However, in meteorology, these principles may be selectively applied to achieve a concise yet meaningful definition tailored to specific applications. While achieving congruence and having a complete set are feasible, maintaining invariances in feature vectors (ensuring similar descriptors for features undergoing affine transformations) proves challenging, as discussed in Sect. 2.4.1.

Conceptually, feature vectors for objects consist of two descriptions: the *region* description and the *boundary* description. The region description describes the arrangement of pixels within the area, and includes descriptors such as moments (see Sect. 2.4.1), perimeter, and shape. Basic regional descriptors focus on the geometric properties of the region, while moments concentrate on the density of the region. On the other hand, the boundary description defines the contour or boundary of the region. The contour's shape can be described by a mesh connecting all contour voxels, or a more condensed representation can be achieved using its convex hull, a Fourier description, or other distinct geometrical representations. Some of the boundary representations are compared and evaluated in Sect. 2.4.4, for an extensive analysis we refer to Nixon and Aguado [126], Chapter 7.

2.2 Identification of atmospheric features

In atmospheric sciences, identifying and tracking specific meteorological features plays a pivotal role in the process of analyzing complex structures and understanding the dynamics of the atmosphere. With the growing volume of atmospheric data generated, processing entire data sets for various computational tasks becomes increasingly challenging. To address this challenge, extracting key features from atmospheric data sets offers a potent solution. This not only reduces the data load but also simplifies the complexity of the information that needs to be explored and analyzed. Moreover, it enables interactive visualization of these features.

Meteorological features have been identified and documented in a structured way since the existence of weather maps and charts in the mid-19th century. Early pioneers like Francis Galton plotted data on maps using symbols that depicted what could be regarded as features, including high and low-pressure systems [58]. In the early 1910s, a group of researchers around Vilhelm Bjerknes in Norway introduced the theory that weather activity is concentrated in relatively narrow zones, which form the boundaries between warm and cold air masses, so-called fronts, which have been added since as features to weather maps [17]. The identification of meteorological features has since become more sophisticated, evolving from manual identification to automated methods. Today, feature identification is not only performed manually but also automatically, and extended to higher dimensions: Atmospheric fronts can now be identified in 3-D fields [10], and a lot of identification results are publicly available to researchers for statistical analyses (e.g., [51]).

Features could not only be defined on 3-D spatial information, but also based on higher dimensional information. Outputs from ensemble NWP models can serve as source for features which are also characterized by derived fields, like ensemble variance, or through clustering of the members. Furthermore, features can be classified based on the descriptors, allowing to better understand classifiable phenomena like weather regimes or types of precipitation.

2.2.1 Feature identification in meteorology

Post et al. [142] classified feature extraction approaches into three categories, all of which are utilized in this thesis: image processing, topological analysis, and physical properties.

- *Image Processing Techniques*. Originally developed for 2-D or 3-D digital images, these techniques can be adapted to various use cases in meteorology. The scientific fields of medical imaging and image enhancement are two of the primary fields that benefit greatly from these techniques. These techniques use basic operations on image data, such as segmentation (thresholding, edge detection) or quantitative descriptions, such as principal component analysis (PCA). In Chapter 5, we use an image processing technique based on morphological operators that has been adapted to the specific requirements of the input data.

- *Topological Analysis*. This analysis is performed on vector fields and involves the detection of critical points such as sources, sinks, and saddle points. These critical points are identified using eigenanalysis theory. We used topological analysis to find vortex centers in the African easterly waves. These vortex centers have been investigated in a related master thesis [110].

- *Physical Features*. Here, features are detected by physical properties of the field, an obvious approach in atmospheric sciences. Physical values, often defined by domain experts, are used to extract features in the fields, such as thresholds of pressure for tropical storms or thresholds of temperature over a certain time span to identify heat waves. For our potential vorticity analyses in Chapters 5 and 6, some thresholds are needed to separate regions of interest from background flow and noise.

When bridging the gap from algorithms to process discrete image data to extracting features from atmospheric data sets, we get confronted with several problems and challenges:

Meteorological domains can be complex. While most operations in digital image processing are defined and performed on discrete multidimensional regular grids (e.g., pixel space in 2-D images), we are confronted in atmospheric sciences with continuous temporal multidimensional data. This data can be present in memory in a variety of projections, as introduced in 2.1. Furthermore, the vertical domain of the data can pose different challenges: Pressure levels are typically represented in a logarithmic fashion with a singularity at 0 hPa, following the distribution of air pressure in the atmosphere, with the z-axis pointing down; model levels are highly irregular near the surface, setting up a complex problem for rendering and processing these fields; and data on isentropic levels could be non-differentiable since the change in potential temperature with height can change its sign. Image processing techniques have to be adapted to suit the different restrictions and limitations of these domains. Due to these complexities, algorithms in literature are restricted to data based on a predefined vertical domain, which fits best the problem statement. For example, potential vorticity feature analysis is performed on isentropes due to their nice meteorological properties (see Sect. 3.3), while isobaric charts are often used for analyzing weather patterns.

Lack of precise definitions. In atmospheric science, accurately defining complex atmospheric processes and their associated features can be a tough nut to crack. Some features, like the strength of a tropical storm measured by the Saffir-Simpson Hurricane Wind Scale (SSHWS), have clear and established classifications. However, many other features are more tricky and lack precise definitions that can be easily translated into code. Often, features are still being labeled manually by domain experts. This means that when designing objective identification methods, collaboration with domain experts is crucial. Regarding the methods presented in this thesis, it took multiple iterations, adjusting configurations doing compromises, to fine-tune these algorithms and achieve the desired results.

Interpretability. For meteorologists, connecting observations and statistical analyses with meaningful interpretations is essential for gaining a deeper understanding of the atmosphere and its workings. Often, this understanding is best achieved when these interpretations are linked to mathematical equations that describe the state of the atmosphere. Furthermore, it's crucial to explain *why* algorithms classify structures in a certain way. Meeting these requirements can be quite challenging, especially when dealing with machine learning methods or image processing techniques applied to gridded data. It is vital to convert the data and processing steps into the atmospheric context.

2.2.2 Morphological image processing

Image processing is a field of information processing that focuses on the manipulation and analysis of digital images. It involves techniques to both transform and extract valuable information from images. While its applications extend to various domains such as computer vision, remote sensing, and medical imaging, these techniques are also highly relevant in the context of meteorology. Input data can be as straightforward as satellite data images, but in general all gridded data sets can be thought of as digital images. *Mathematical morphology* is a branch of this field coined and firstly introduced by Matheron [115] and Serra [166], and focuses on the analysis and extraction of geometric information and structures using set-theoretic operators and operations. These morphological operators have diverse effects, some of which hold significant value in meteorological applications. One crucial aspect to consider when employing morphological operators is that they are non-conservative. This means data can be altered or lost during the process, potentially leading to erroneous conclusions when analyzed by domain experts. Therefore, close collaboration with domain experts is imperative when utilizing these techniques.

Among the many applications of morphological operators in meteorology, image segmentation stands out as a prominent use case. Image segmentation involves partitioning an image into distinct subgroups, known as image segments. This is achieved by classifying or labeling each point in space to a specific group based on predefined criteria. The segmentation process relies on various methods such as thresholds, clustering, edge detection, region-growing algorithms, and more. Thresholding, in particular, is widely utilized in meteorology due to established threshold values, for example, for distinguishing between tropospheric and stratospheric air masses (see Sect. 3.3). This technique finds application in Chapters 5 and 6 of this thesis.

Structuring elements are key components in morphological image processing. They are small geometric patterns or kernels that define the neighborhood of each pixel during morphological operations and are used to probe or interact with a given data. Consider a binary image A and a structuring element B, which is also a binary image, usually much smaller than A. The structuring element is positioned at each possible location in the image, and a predefined operation is applied to the image under the structuring element. *Dilation* and *erosion* are two key concepts that alter the shape and size of objects in an image using structural elements, and are the motivation of the feature identification process outlined in Chapter 5.

A morphological erosion, denoted by the $\ominus$ operator, shrinks or thins objects in an image. Mathematically, the erosion of set A by structuring element B is defined as:

$$A \ominus B = \{z \in \mathbb{Z}^2 \mid B_z \subseteq A\}. \tag{2.4}$$

This means that the erosion of A by B is the set of all points z such that when B is centered at z, B is completely contained within A.

Morphological dilation, denoted by the $\oplus$ operator, involves adding pixels to object boundaries. Given two sets A and B, where A again represents the image and B the structuring element, the dilation is defined as:

$$A \oplus B = \{z \in \mathbb{Z}^2 \mid (z - B) \in A \text{ for some } b \in B\}. \tag{2.5}$$

This definition implies that the set $A \oplus B$ consists of all points z such that when B is translated by z, at least one element of B overlaps with an element of A.

Dilation expands the boundaries of a feature based on the structuring element, while erosion shrinks them. Combined, the effects of the dilation and erosion operators unfold their potential by enabling image manipulation and feature extraction capabilities. The combination of erosion followed by dilation $(A \ominus B) \oplus B$ is known as morphological opening. Conversely, dilation followed by erosion $(A \oplus B) \ominus B$ is a morphological closing. These operations are showcased in Fig. 2.3. Here, an exemplary potential vorticity feature identified

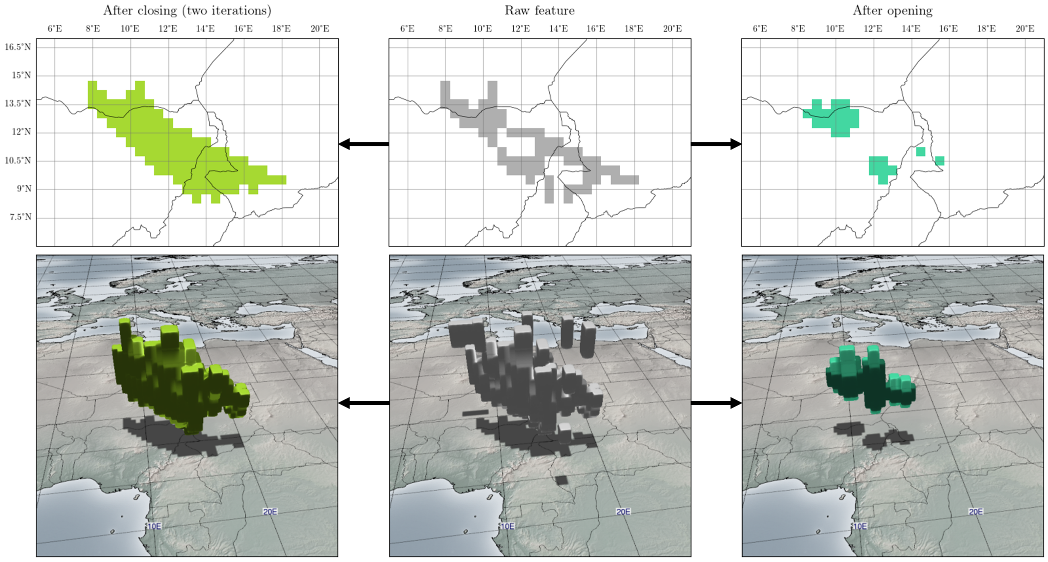


Fig. 2.3: A 2-D and 3-D view of an exemplary potential vorticity feature being identified in Chapter 6 to showcase the morphological operations. The raw identified feature in the middle column is processed using the closing (left) and opening (right) operators on the raw structure.

in Chapter 6 is displayed. The raw feature in the middle looks quite disrupted, but the human eye can clearly identify the area the feature encompasses. Generally, a morphological closing fills small holes and gaps within objects, fuses minor breaks, and smoothens the boundaries. This operation has been applied to the structure, the result is visible in the left column. Both, the 2-D and 3-D view of the feature exhibit a more cohesive structure. The output of the closing process can be used to compute shape parameters of the cohesive structure. On the other hand, a morphological opening removes noise, small objects and thin protrusions, keeping only a smoothed inner core of the feature. This task proves to be vital in many application as a preprocessing step, however, in our use case, it would remove too much information of the feature.

An obvious problem in meteorological applications is here again the interpretability (see Sect. 2.2.1). While all morphological operators can fundamentally be applied to gridded data sets, they operate in pixel space. When a meteorologist is interested in features with a spatial extent of at least 100 km, static structuring elements are not usable due to distortions in the projection, and a coarse resolution might propagate and increase these errors. In Sect. 5.4 we propose a novel strategy to perform these morphological operators on conformal projections, working in real-world units instead of pixel space, by using a crossover with the Fast Marching Method to compute distances in the projection.

2.3 Tracking of atmospheric features

Before analyzing identified features, *feature tracking* can be employed if the input data has a temporal dimension which makes it possible to meaningfully connect the identified features in time. Along a track, growth and decay of these features, and their spatial distribution can be analyzed and compared across various data sets.

Long before objective tracking has been employed, many researchers tracked their features manually, due to the lack of computing power, low data resolution and noisy input data (e.g., Streten and Troup [176] for Southern Hemisphere vortices, Akyildiz [3] for winter cyclones over the Northern Atlantic). They used these tracks to explore variability of the atmosphere and to generate statistics on various parameters of the tracked features. With increasing data and computing power however, automated tracking based on objective criteria became the more advantageous approach. In meteorology, the first predominant use case for objective tracking was by tracking clouds from satellite data via brightness centers [42, 108]. Williamson [202] tracks cyclones via anomalies in geopotential height on the 500 hPa isosurface, more sophisticated techniques based on surface pressure data have been developed by Le Treut and Kalnay [107] and Murray and Simmonds [123]. One of the pioneers in tracking features in the tropics was Reed et al. [155], who implemented the first objective AEW tracking algorithm. Similar to identified features drastically reducing the amount of data to analyze, the tracks connecting the features is a valuable addition to this low-memory description. These tracks facilitate the investigation of feature evolution in individual case studies, to correlation of tracks within ensemble forecasts, to statistics on entire climatological timescales.

Regarding tracking motions in image sequences, Aggarwal and Nandhakumar [2] published a review on tracking techniques, distinguishing between two paradigms: feature-based approaches and optical flow-based approaches. The latter is motivated by the fact that spatial and temporal variations in image brightness indicate motion in the sequence. Although technically applicable to some atmospheric science use cases, optical flow is predominantly used for stereo correspondence and is shown to be less robust. This leads both the authors and us to prefer the feature-based approach. In this process, identified features at different points in time must be compared over multiple time steps to determine whether they represent the same entity. This is known as the *correspondence problem*.

2.3.1 Correspondence problem

The correspondence problem poses a challenge as features may, depending on the use case and methodology, undergo variations such as disappearing, reappearing, being occluded by other features or the environment, or they might experience rapid changes in shape. Aggarwal et al. [1] categorized solutions to this problem into two main types: iconic models and structural models.

Iconic models rely on *templates*, which are features detected and extracted from the initial frame (in our case the first time step), and then sought in subsequent frames by pattern matching. For meteorological applications, this is problematic because the features might not be present at the first time step, and drastically change their shape over their entire life cycle.

Contrary, when using structural models, *tokens* are extracted from each picture, containing a set of attributes and constraints. These tokens are then matched with each other. According to Aggarwal et al. [1], this proves to be way more robust, yet more computational expensive then iconic models. In meteorology, these tokens align closely with what we defined as feature vectors in Sect. 2.1.2. When matching tokens, constraints are often expert-defined, incorporating assumptions about the smoothness of motion [32, 77], the expected velocity and direction of advection of air masses [11, 138], or masking a certain sub-space where features are expected to occur [76, 172].

Given N identified and characterized feature vectors $\{f_1, \ldots, f_N\}$ at times $\{t_{f_1}, \ldots, t_{f_N}\}$, we can define the correspondence function C as:

$$C(f_i, f_j, S) = \begin{cases} 1 & \text{if } f_i \text{ and } f_j \text{ represent the same entity and } t_{f_i} < t_{f_j} \\ 0 & \text{otherwise} \end{cases} \tag{2.6}$$

Here, S represents the complete state of the system from which the constraints are derived to solve the correspondence problem. The matrix $E \in \mathbb{Z}_2^{N \times N}$ with

$E_{ij} = C(f_i, f_j, S)$ is a binary matrix, where ones indicate that the correspondence is fulfilled and zeros indicate otherwise. Figure 2.4 shows an example: ten features f_0 to f_9 with their respective timestamps t and connected features in the table on the top left, where features are connected if the correspondence function C equals 1. The associated matrix E is on the top right.

Feature	Time	Connections
f_0	t_0	$\{f_1, f_2\}$
f_1	t_1	$\{f_3\}$
f_2	t_1	$\{f_3\}$
f_3	t_2	$\{f_6\}$
f_4	t_2	$\{f_5\}$
f_5	t_3	$\{\}$
f_6	t_4	$\{f_8\}$
f_7	t_4	$\{f_8\}$
f_8	t_5	$\{\}$
f_9	t_5	$\{\}$

$$E = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ & & & & 0 & 1 & 0 & 0 & 0 & 0 \\ & & & & & 0 & 0 & 0 & 0 & 0 \\ & & & & & & 0 & 0 & 1 & 0 \\ & & \text{\huge 0} & & & & & 0 & 1 & 0 \\ & & & & & & & & 0 & 0 \\ & & & & & & & & & 0 \end{bmatrix}$$

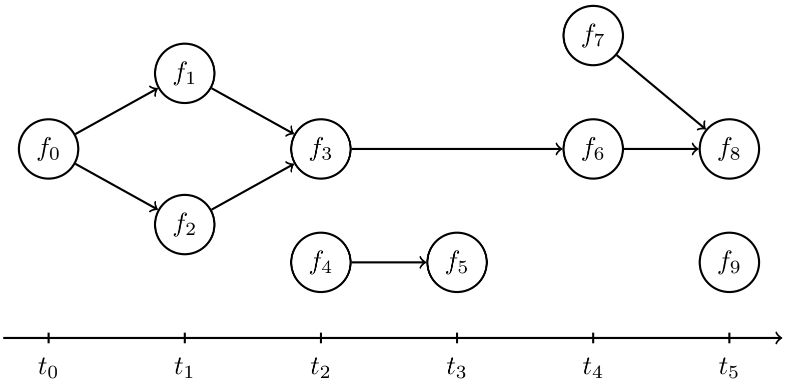


Fig. 2.4: Three representations to show correspondences between features: Table view on the top left, adjacency matrix on the top right, and the tracking graph on the bottom.

2.3.2 Tracking graph

When considering the matrix E as adjacency matrix, as on the top right of Fig. 2.4, we define $G = (F, E)$ as the tracking graph of the features $F = \{f_1, \ldots, f_N\}$, which are represented by nodes. Similar to the definition of the Event Graph introduced by Limbach et al. [109], this graph encapsulates both the identified features and their correspondence relations to other features in an abstract fashion. While Limbach et al. [109] restricts edges to connections between features from consecutive time steps only, the more general matrix definition also permits connections over longer intervals.

In a tracking graph, the adjacency matrix has the following properties when features are ordered by time ($i < j \Leftrightarrow t_{f_i} \leq t_{f_j}$):

- the elements along the diagonal are zero $(E_{ii} = 0)$,
- the matrix is triangular since $t_{f_i} \leq t_{f_j}$ for $i < j$,
- if $\forall j, E_{ij} = 0$ and $\forall k, E_{ki} = 0$, the feature f_i is *isolated*; it is the only manifestation of this entity in the data set,
- if $\exists j, E_{ij} = 1$ and $\forall k, E_{ki} = 0$, there are only outgoing edges from f_i, representing the *genesis* of a new track,
- if $\forall j, E_{ij} = 0$ and $\exists k, E_{ki} = 1$, there are only incoming edges to f_i, representing the *disintegration* of a track,
- one incoming edge to and multiple outgoing edges from feature f_i signifies the *splitting* of the track into multiple continuations,
- multiple incoming edges to and one outgoing edge from feature f_i signifies the *merging* of multiple tracks into one track.

2.3.3 Extracting tracks

While the tracking graph contains all information about the correspondences, it is required for most applications to extract single tracks from this graph structure. A track, which represents the full evolution of one entity over its life cycle, can be represented as a subgraph $T^i = (F^i, E^i)$ induced by all features $F^i \subseteq F$ making up the i-th track and their correspondences $E^i \subseteq E$. Note that for easier notation, we refer to the correspondences E as a list of edges. Given the sparsity of the adjacency matrix, our implementation, provided with this thesis, uses edge lists for memory efficiency. Additionally, our graph model requires tracks to be disjoint, thus, $F^i \cap F^j = \emptyset$ and $E^i \cap E^j = \emptyset$ for $i \neq j$. The union of all track nodes forms exactly the set of the original nodes $\bigcup_{i=1}^{n} F^i = F$, and the correspondences form a subset $\bigcup_{i=1}^{n} E^i \subseteq E$.

From a practical perspective, extracting tracks is challenging. Features can merge with other features, split up again, or in extreme cases, features might be stationary and interact multiple times with other features passing by. Even experts in the field may struggle to define a track when complex interactions with the atmosphere or other tracks occur. Here, we present two practical methods of track extraction. It is important to mention that the graphs we are working with are acyclic, as the graph connections represent temporal progressions of features.

The first method, as depicted in Algorithm 1, is based on the invariant that in the case of a split event, the track continues on the path that yields the longest possible track in total, while the other nodes in the split (merge) event form (end) a different track. Extracting the longest paths in general graphs is, in fact, $\mathcal{NP}$-hard: The well-known Hamiltonian Path Problem, which is $\mathcal{NP}$-complete, can be reduced to the longest path problem [90]. However, this problem becomes linear for acyclic graphs. We realize the linear time complexity through two depth-first search (DFS) setups. In the first DFS, we compute for each node its maximum time until dissipation, and which child node is connected to this exact path. Since the tracking graph is not fully connected, we have to start this DFS multiple times, marking nodes as discovered. By processing the starting nodes for the DFS pre-sorted by time (as they are output from the identification process), we ensure that we do not traverse parts of the graph multiple times or override already existing paths. This results in a computational complexity of $\mathcal{O}(|F|+|E|)$. In the second step, another DFS is performed with starting nodes sorted by time, where we traverse along the saved longest paths. Each time a new root node is processed, a new identifier is generated for the new track starting with this node. If we reach a node that is already labeled, the encountered node is already part of a track that started earlier in time, thus, a longer track. The currently traversed track ends there. This method also operates in $\mathcal{O}(|F|+|E|)$ complexity. Figure 2.5a shows an exemplary result of this procedure.

Although our approach resembles topological sorting, it extends beyond merely ordering vertices. Topological sorting in a directed acyclic graph (DAG) orders vertices such that for every directed edge uv, vertex u comes before vertex v. While this sorting ensures linear processing order, our method further incorporates path length determination and handling of merge and split events.

The second method defines a track as a connected subgraph in the tracking graph. Therefore, all split and merge events continue the same tracked entity. This can be realized using a union-find procedure to find connected subgraphs.

We iterate over the edges, connect features by having them represented by a designated parent node, and update the parent node while performing path compression to increase performance. Excluding the $\mathcal{O}(|F|)$ space allocation, this leads to a quasi-linear runtime of $\mathcal{O}(|E|\alpha(|F|))$ (see [178]), where α denotes the inverse Ackermann function, which is less than three in our applications. Alternatively, a DFS or a BFS can also be performed to extract connected subgraphs. In this method, we start at unvisited nodes, marking all reachable nodes visited and as part of the same track. The time complexity is, as in the first method, $\mathcal{O}(|F| + |E|)$. Figure 2.5b shows the identified connected subgraphs in this example.

Both approaches have moments where they excel and cases where the other algorithm would provide a more reasonable result. Ultimately, it comes down to the question: What defines a track? In complex scenarios, ten different domain experts may suggest ten different sets of tracks. Occasionally, a feature splits and merges with itself a short time later (see Fig. 2.5). In such cases, the second algorithm would still consider the whole evolution as one track, while the first algorithm would only continue one of the subtracks. This leads to more tracks in total, as illustrated. The first algorithm will always continue the longest path to assemble a track.

A potential third algorithm could generate new tracks at every split and merge event. However, this would fragment tracks as soon as split events occur. Both labeling tracks by connecting subgraphs and breaking tracks at every merge and split event are less robust than the first option. In noisy data sets and with poorly optimized algorithms, many false positive edges can connect feature nodes incorrectly, either merging subgraphs that should be separate or inflating the number of split and merge events. Different track generation heuristics were employed in Chapter 6, with the first method yielding the most robust results.

2.3.4 Example: Point feature tracking

One of the simplest features to track are point features. These are defined primarily by their spatial location in 2-D or 3-D space, for example the point of lowest pressure for storm tracking algorithms [22], or structural centroids of air parcels. Then, a feature vector $f = (\lambda, \phi, p, \dots)$ contains the location

Algorithm 1 Track extraction with longest path invariant

```
Input: $G = (F, E) = (\{f_1, \dots, f_N\}, E)$: Tracking graph with features sorted by corresponding time step
Output: Track identifier $f.\mathrm{id}$ for each feature

procedure DFS($f$)
    ▷ This DFS assigns each node its time until dissipation along the longest track, as well as the the continuation node along that track
    label node $f$ as discovered
    if $\nexists g \in F \mid (f, g) \in E$ then
        return $f, 0$
    else
        for $i \in \{0, \dots, k\} | (f, g_i) \in E$ do
            $c_i, t_i \leftarrow$DFS($g_i$)                    ▷ Recursive call to all child nodes
        end for
        $i_{\max} \leftarrow \arg\max_{i=\{0,\dots,k\}} t_i$          ▷ Child index of longest path
        $f.\mathrm{next} \leftarrow c_{i_{\max}}$                    ▷ Keep track of path
        $f.t \leftarrow t_{i_{\max}} + \Delta t$                  ▷ Time until dissipation
        return $c_{i_{\max}}, t_{i_{\max}} + \Delta t$
    end if
end procedure

procedure INDEXING($f$, id)                          ▷ Assign track IDs
    if $f.\mathrm{id} \geq 0$ then
        return                                        ▷ Already considered
    end if
    $f.\mathrm{id} = \mathrm{id}$
    if $f.\mathrm{next} \neq f$ then
        INDEXING($f.\mathrm{next}, \mathrm{id}$)
    end if
end procedure

for $f \in F$ do                                      ▷ First loop: DFS
    if $f$ is not discovered then
        DFS($f$)
    end if
end for
$\mathrm{id} \leftarrow 0$
for $f \in F$ do                                      ▷ Second loop: Indexing
    if $f.\mathrm{id} < 0$ then            ▷ Not considered yet → new track identifier
        INDEXING($f, \mathrm{id}$)
        $\mathrm{id} = \mathrm{id} + 1$
    end if
end for
```

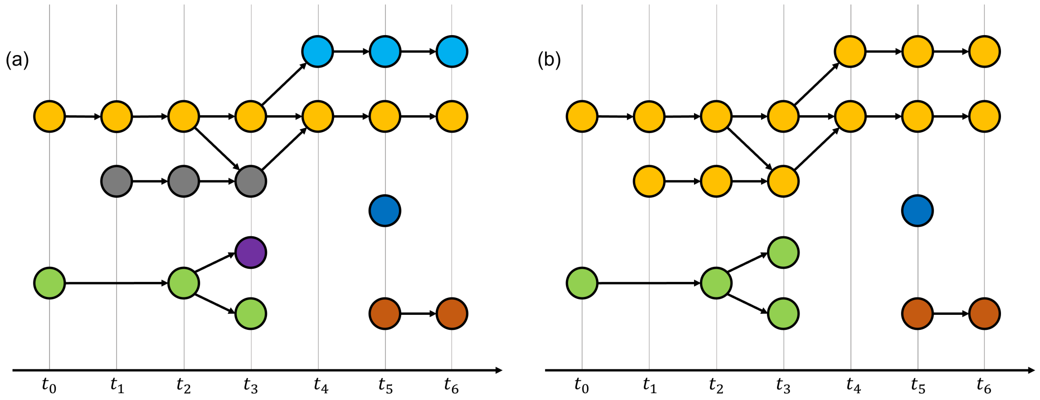


Fig. 2.5: An exemplary tracking graph where the nodes are colored by the track they belong to. Possible results for (a) using Algorithm 1, and (b) with union-find to identify connected subgraphs.

$f_{pos} = (\lambda, \phi, p)$ of the feature. The correspondence function can be expressed in the most simple fashion as

$$C(f_i, f_j, S) = \begin{cases} 1 & \text{if } ||f_{i,pos} - f_{j,pos}|| < v_{max} \cdot (t_{f_j} - t_{f_i}) \\ 0 & \text{otherwise} \end{cases} \tag{2.7}$$

with an expert-defined physically motivated maximum velocity threshold v_{max} of the underlying dynamical process, and an application specific distance norm $|| \cdot ||$. In most cases, the applied norm is simply the spherical distance of the latitude-longitude point pairs. A minimum threshold for the velocity can be set similarly.

2.3.5 Example: Area feature tracking

For many atmospheric phenomena, representing a feature as a single point is insufficient, as these features often influence and impact entire areas. These areas are crucial for producing and presenting forecasts, conducting predictability studies, and managing disasters. Instead, area features are used, which are defined as coherent finite-size spatial regions in two or three dimensions. Examples include potential vorticity structures [174, 49, 48], areas of convection [69, 68, 113], and high wind features [85, 37].

In literature, by far the most common strategy to track area features is *overlap tracking*. Here, the designated area of different features at (mostly consecutive) time steps are compared, and they represent the same entity if there is an overlap between both features. Mostly prominent this is done on binary fields, and image processing techniques are applied to improve robustness of the strategy [109, 68]. For overlap tracking, the feature vector contains specific information about the covered area, typically a list of grid points or a binary mask. A feature vector could then be represented as $f = (P, \dots)$, where P is the set of gridded data points making up the feature area $f_P = P$, and

$$C(f_i, f_j, S) = \begin{cases} 1 & \text{if } f_{i,P} \cap f_{j,P} \neq \emptyset \\ 0 & \text{otherwise} \end{cases} \tag{2.8}$$

would represent the most simple tracking overlap strategy.

Besides point (0-D) and area (2-D, 3-D) feature tracking, other structures can also be tracked with suitable definitions. For example, features can consist of 1-D line strings [11, 206, 48] or other higher-dimensional structures. Moreover, when features are defined by geometric descriptions, overlaps can be computed analytically, rather than merely intersecting binary masks.

2.4 Feature Analysis

2.4.1 Moment analysis

Moment analysis in statistics provides a means to describe the arrangement of pixels within a shape, offering valuable insights into its quantitative measures. Drawing inspiration from mechanical moments in engineering, these statistical moments play a crucial role in understanding shape characteristics.

Hu [84] laid the foundation for moment-based feature extraction in image analysis, and proposed the 2-D geometric moments of an image as a structured element of what he named "moment invariants". These mathematical moments have gained wide acceptance in many image analysis tasks and applications, including shape recognition and object classification. They showed up properties and computation instructions for specific sets of moments, which are invariant under translation, scaling and rotation: a crucial criterion for many applications.

Although moments are defined for arbitrary many dimensions, much of the existing literature and applications deal with 2-D features and moments. Our work extends this framework to 3-D tasks. Generalizing to three dimensions, the Cartesian moments m_{pqr} of order $p+q+r$ of a function F are defined as

$$m_{pqr} = \iiint x^p y^q z^r F(x,y,z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z. \tag{2.9}$$

Conversely, the moments $m_{pqr}\,\forall p,q,r \in \mathbb{N}_0$ uniquely determine the original function, similar to the reconstruction via an inverse Fourier transform. In practical applications, where data is given on discrete grids and represented as images $I(x,y,z)$, moments are computed in the discrete equivalent of Eq. 2.9:

$$\bar{m}_{ijk} = \sum_x \sum_y \sum_z x^i y^j z^k I(x,y,z) V(x,y,z), \tag{2.10}$$

where V defines the volume of each voxel.

In the context of this thesis, moments of order 0 to 2 are particularly important. In this section, we provide a concise overview of these moments from both theoretical and practical perspectives. These moments have clear statistical and visual interpretations.

When computing the 0-th order moment, Eq. 2.10 simply collapses to the sum of all points in the image weighted by their volume. The first order moments compute the mean along each dimension. Therefore, the vector $\{m_{100}, m_{010}, m_{001}\}$ represents the center of mass relative to the input function. The second order moments have a more intricate definition with very useful properties, and are explored more in detail in Sect. 2.4.3.

While the moments m and $\bar{m}$ respectively are not invariant under any affine transformations, alternative sets of moments can be computed. Central moments, defined as

$$\mu_{pqr} = \iiint (x-\bar{x})^p (y-\bar{y})^q (z-\bar{z})^r F(x,y,z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \tag{2.11}$$

are moments centered around the center of mass ($\bar{x} = \frac{m_{100}}{m_{000}}, \bar{y} = \frac{m_{010}}{m_{000}}, \bar{z} = \frac{m_{001}}{m_{000}}$). These are trivially invariant under translation. Moreover, Hue moments [84], which can be computed from the central moments, are also invariant under scale and rotation. Generally, the computation of discrete moments from gridded data can only give an approximation of the true moments, therefore the precision of the invariance becomes more unclear.

2.4.2 Moment analysis for meteorological data

Applying moment analysis to meteorological data involves treating gridded output from Numerical Weather Prediction (NWP) systems as images. Here, for introductory purposes, we assume the data to be present on an equirectangular projection (see Sect. 2.1.1) and consider 3-D data with pressure levels in the vertical domain. Later, we explore the possibility of substituting different projections or grids. Adapting Eq. 2.10 to meteorological dimensions—longitude (λ), latitude (ϕ), and atmospheric pressure level (p)—we have:

$$M_{ijk} = \sum_{\lambda} \sum_{\phi} \sum_{p} \lambda^i \phi^j p^k I(\lambda, \phi, p) V(\lambda, \phi, p). \tag{2.12}$$

Here, the function I could represent any field, be it continuous (like a wind field) or discrete (such as a mask of an identified structure). Crucially, compared to most use cases in image analysis, the volume function V is not constant due to the used data projection and vertical resolution, and therefore an integral part of building the bridge from pixel space to physical units. Typically, applications use a static grid, therefore, V can be precomputed. While equal-area projections offer a solution to this issue, they come with drawbacks (see Sect. 2.1.1). In spherical space, more of the desired feature properties introduced by Nixon and Aguado [126] hold, however, different grids have to be used to account for the curvature (e.g., geodesic grids), which are more challenging to compute and more challenging to interpret. Expert-driven knowledge is here vital to find a good compromise between the meteorological and mathematical interpretability.

Unfortunately, moment invariants are lost under projection. Even two features with identical central moments in the projection space can represent vastly different entities. For example, a latitudinal displacement of a feature in an equirectangular projection changes the structure's extent in spherical space due to varying distortions based on latitude. Additionally, the geometric distance between two pressure levels varies over location and time, depending on the atmospheric dynamics in place. However, in regions where the distortion is minimal (e.g., near the equator in an equirectangular projection) and with certain assumptions about the atmospheric state, moment similarity can still be used to compare features. In such instances, approximations of scale and rotational invariants prove valuable. Bridging mathematical techniques with meteorological understanding requires careful consideration of the atmospheric processes' characteristics.

2.4.3 Eigenanalysis

The second order moments can give information about the shape's variance and orientation. Especially the orientation is a measure of interest in the field of meteorology. Structures in the atmosphere exhibit dynamical behaviors along all spatial dimensions. For example, the tilt of a feature can give insight into its life cycle, the intensity of the feature, or even help experts in doing forecasts (e.g., tilt of upper-level troughs [181] or structure of simulated thunderstorms [99]).

We define the main orientation of a feature as the axis, along which the variance of the structure is maximized. To compute this axis, we compute the variance and covariance along all dimensions, for example:

$$\begin{aligned} \mathrm{Var}[\lambda] = \mathrm{E}[(\lambda - \bar{\lambda})^2] &= \sum_{\lambda}\sum_{\phi}\sum_{p}(\lambda - \bar{\lambda})^2 I(\lambda, \phi, p) V(\lambda, \phi, p) = \tilde{\mu}_{200}, \\ \mathrm{Cov}[\lambda, \phi] &= \sum_{\lambda}\sum_{\phi}\sum_{p}(\lambda - \bar{\lambda})(\phi - \bar{\phi}) I(\lambda, \phi, p) V(\lambda, \phi, p) = \tilde{\mu}_{110}, \end{aligned} \tag{2.13}$$

where $\tilde{\mu}$ also represent the discrete central moments falling in line with the variances and covariances. The other variances and covariances between the dimensions can be computed similarly. The connection between second order moments and covariance becomes clear. Therefore, we can express the covariance matrix of a structure using central moments:

$$\mathbf{\Sigma} = \begin{pmatrix} \mathrm{Var}[\lambda] & \mathrm{Cov}[\lambda, \phi] & \mathrm{Cov}[\lambda, p] \\ \mathrm{Cov}[\phi, \lambda] & \mathrm{Var}[\phi] & \mathrm{Cov}[\phi, p] \\ \mathrm{Cov}[p, \lambda] & \mathrm{Cov}[p, \phi] & \mathrm{Var}[p] \end{pmatrix} = \begin{pmatrix} \tilde{\mu}_{200} & \tilde{\mu}_{110} & \tilde{\mu}_{101} \\ \tilde{\mu}_{110} & \tilde{\mu}_{020} & \tilde{\mu}_{011} \\ \tilde{\mu}_{101} & \tilde{\mu}_{011} & \tilde{\mu}_{002} \end{pmatrix}. \tag{2.14}$$

Now, to compute the main orientation of a feature, we can look at the problem from an algebraic point: Let c be the axis, along which the variance is maximized if the data is projected onto c. Therefore, we maximize $c^T\mathbf{\Sigma}c$ with $|c| = 1$ to find the axis corresponding to the main orientation. Using the eigendecomposition of the symmetric matrix $\mathbf{\Sigma} = \mathbf{PDP}^T$, where $\mathbf{P}$ is a 3x3 matrix consisting of the eigenvectors of $\mathbf{\Sigma}$, and $\mathbf{D}$ being a diagonal matrix with associated eigenvalues on the diagonal (see for example [26]), it becomes clear that the variance is maximized exactly for the biggest eigenvector of $\mathbf{\Sigma}$:

$$c^T\mathbf{\Sigma}c = c^T\mathbf{PDP}^Tc = v^T\mathbf{D}v = \sum_i v_i^2 d_i \leq \max_i d_i = d_1, \tag{2.15}$$

Boundary Description	Memory	Accuracy	Comp. Speed
Volume Bounding Box	Low	Low	$\mathcal{O}(n)$
Quadric Approximation	Low	Moderate	$\mathcal{O}(n)$
Compressed Mesh	High	High	varies
3-D Fourier Transform [193]	Moderate	High	$\mathcal{O}(n^3 \log n^3)$
Raw Object Mesh	Highest	Highest	-

Tab. 2.1: Different boundary description types with a rough order of magnitude for their memory, accuracy, and computational speed for n 3-D points, given a raw object mesh.

with $v = \mathbf{P}^T c$ and without loss of generality $d_1 \geq d_i$. Thus, the largest eigenvector of $\mathbf{\Sigma}$ corresponds to the primary orientation of the feature. We then project the feature onto the plane orthogonal to this main axis, resulting in a 2-D projection of the 3-D feature. For this 2-D field, we again compute the axis of maximum variance, ultimately providing a set of three semi-principal axes representing the feature's orientation. Note that the eigenvectors of a symmetric matrix are orthogonal.

Numerous algorithms have been developed and optimized to find the eigenvalues and corresponding eigenvectors of a matrix (see Parlett [135] for an overview). For the real symmetric covariance matrices we encounter, we use the Jacobi eigenvalue algorithm, as it converges quickly (quadratic convergence, see Golub and Van Loan [60]) to the desired eigenvalues and eigenvectors. This method allows us to compute three orthogonal axes of the given 3-D structure, providing a representation that describes the feature's orientation.

2.4.4 Best-fit geometry

Boundary representations provide a means to represent objects with varying levels of accuracy. Table 2.1 offers an overview of selected boundary representation approaches, considering factors like memory efficiency, accuracy, and computational speed. Clearly, established representations using additional memory are also generally capable of reducing the approximation error.

Each technique offers its own set of advantages and challenges. The most straightforward approach is the volume bounding box. It is defined by two points, designated as (λ_1, ϕ_1, p_1) and (λ_2, ϕ_2, p_2), which define two opposing corners of the bounding box. Consequently, this method utilizes six scalar values to describe the state. Following its computational simplicity, it lacks detailed shape information. In contrast, the utilization of quadric surfaces presents a more sophisticated approach. These are three-dimensional surfaces that can be expressed through second-degree polynomial equations. In general, a quadric surface is defined by the equation:

$$A\lambda^2 + B\phi^2 + Cp^2 + D\lambda\phi + E\lambda p + F\phi p + G\lambda + H\phi + Ip + J = 0, \tag{2.16}$$

and can also be expressed in matrix form:

$$x^T \mathbf{A} x + bx + c = 0, \mathbf{A} = \begin{pmatrix} A & D & E \\ D & B & F \\ E & F & C \end{pmatrix}, b = \begin{pmatrix} G \\ H \\ I \end{pmatrix}, c = J, x = \begin{pmatrix} \lambda \\ \phi \\ p \end{pmatrix}, \tag{2.17}$$

where $A, \ldots, J$ are the real coefficients characterizing the specific quadric surface, and at least one of A, B, C must be nonzero. Depending on the choice of parameters, as nicely illustrated by Salmon [162], this surface takes the form of an ellipsoid, a paraboloid or a hyperboloid.

Ellipsoids are a specific type of quadric where the matrix $\mathbf{A}$ is symmetric and positive definite, and $J < 0$. According to the principal axis theorem, and as pointed out by Strang [175], the eigenvectors of $\mathbf{A}$, with a vector scaling of $\frac{1}{\sqrt{\lambda_i}}$ related to their corresponding eigenvalues λ_i, form the semi-principal axes of the ellipsoid. This falls in line with the covariance matrix introduced in Sect. 2.4.3, where axes along maximum variance were identified. Therefore, setting $\mathbf{\Sigma} = \mathbf{A}$, an ellipsoid fitting to the data in the sense of a Principal Component Analysis can be computed. The computation complexity is linear with the amount of points making up the feature assuming the points are always 3-D. Ellipsoids offer a compact and robust representation, characterized by straightforward operations and clear interpretation. Smith and Woolf [170] termed this geometrical representation as the *most economical representation* for meteorological observations. However, it's important to note that they have limited accuracy, particularly in capturing complex or non-coherent features.

Other approaches non-related to the covariance matrix can be considered to compute a best-fitting ellipsoid. An adaption of Khachiyan's algorithm [94] is a rather complex approach based on linear programming to compute the minimum-volume enclosing ellipsoid of a given set of points. This way, the ellipsoid can be done in $\mathcal{O}(n^{3.5} \ln(n\epsilon^{-1}))$ time for n points and a relative accuracy of ϵ. Optimizers can also be used to minimize the squared distance between the boundary and an ellipsoid. In practice, we discovered that these two strategies are sensitive to outliers and require the computation of a boundary of the feature (e.g., its convex hull) beforehand. Generally, the PCA approach scales much better for the input data at hand.

Compressed mesh representations are flexible in accuracy but memory-intensive. Fourier descriptors capture frequency information, suitable for shapes with periodic features but lack meteorological interpretability. Representing objects with raw meshes offers high accuracy but is memory-expensive, limiting its use for statistical analyses.

While there is generally no best approximation, the choice of representation depends on the specific application, characteristics, and constraints of the objects to analyze. Especially in meteorology, interpretability and statistical conciseness are key for efficient data analysis and conclusive evaluations. While ellipsoids take center stage in upcoming chapters due to their interpretability and statistical efficiency, some complex meteorological features may require hybrid approaches. Such hybrids could involve computing an ellipsoid to encapsulate an object's general shape for tracking and statistical evaluations, while retaining a more detailed description for further in-depth analyses.

2.4.5 Shape parameters

Table 2.2 presents key shape parameters computed in Chapters 5 and 6 for the identified features, respectively. These descriptors, calculated for the respective data sets mentioned in the publications, are available in the code and data availability section in the appendix of this work. The computational complexity for all 3-D descriptors is linear with respect to the number of voxels defining the feature.

Parameter	**Representation**	Chap. 5	Chap. 6
Bounding Box	2 vectors	✓	✓
Centroid	1 vector	✓	✓
Semi-principal axes	3 vectors	✓	✓
Volume	scalar	✓	✓
Intensity (average, peak)	scalar	✓	✓
Eccentricity	scalar	✓	✓
2-D projected semi-principal axes	2 vectors		✓

Tab. 2.2: A selection of feature descriptors computed in the upcoming chapters.

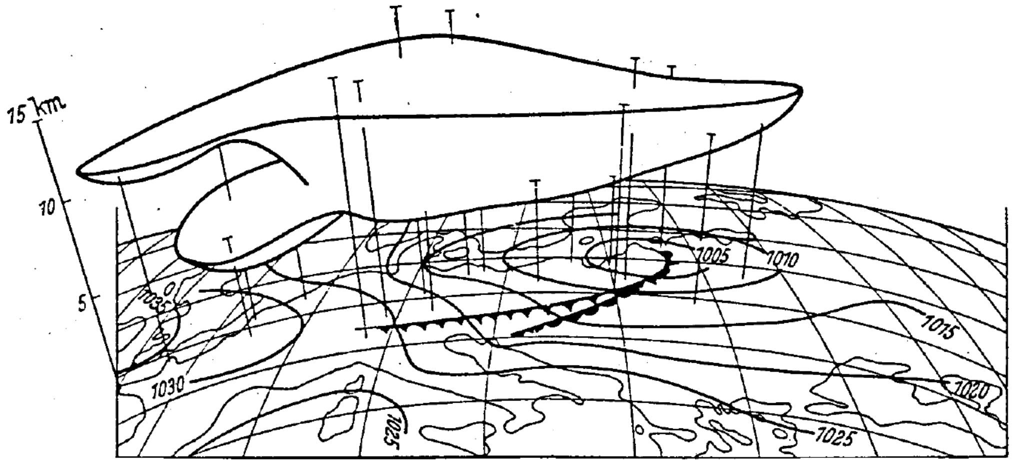


Fig. 2.6: A hand-drawn 3-D visualization of an upper-level PV anomaly by Ernst Kleinschmidt [98], resembling one of the first of its kind.

2.5 Data visualization in meteorology

> *Visualization transforms the symbolic into the geometric, offering a method for seeing the unseen. It enriches the process of scientific discovery and fosters profound and unexpected insights.*
>
> — **Bruce H. McCormick**
> Visualization in Scientific Computing

The foundation of visualizing meteorological data can be traced back to the dissertation of German physicist Heinrich Wilhelm Brandes in 1826 [20]. He created the first weather maps by reconstructing and displaying observations from 1783. Since then, synoptic meteorology has evolved, and maps have become a standard tool for analyzing the state of the atmosphere. Advancements in technology have significantly increased the volume and availability of meteorological data, leading to more frequent and sophisticated map visualizations. Weather maps started to utilize a wide range of visualization techniques: surface pressure is typically represented using iso-contours, temperature with different color shadings, wind with streamlines or wind barbs, and observational data with specific synoptic symbols and syntax.

With the advent of radiosondes and aircraft measurements since the early and mid-20th century, an additional dimension of observational data, representing different heights in the atmosphere, became available. To visualize the state of the troposphere, maps are drawn which display data mainly from single pressure levels. Still, visualization tasks were performed by hand, by putting station data on a map, connecting the lines, and identify structures using the human eye. During this time, Kleinschmidt [98] published one of the first 3-D visualizations of potential vorticity structures, drawn by hand (Fig. 2.6). This work pioneered the 3-D view of atmospheric processes, demonstrating how anomalies aloft might be connected to weather phenomena at the surface. In the 1960s, the advent of NWP models further increased the volume and granularity of data to be processed and analyzed. The introduction of ensemble forecasts in the late 20th century added another dimension to the output data of NWP models, while the resolution of other dimensions continued to improve, resulting in a massive increase in data load. Thanks to the rapid development of personal and institutional computers, today's researchers can process and visualize data sets of previously unimaginable size. Manual visualizations are now rare, making way for computer-aided visualization.

Effective visualization is vital to communicate information from the increasingly large and complex data sets to scientists and the general public. However, as data sets become larger, conventional visualizations can become more cluttered and less informative. To address such challenges, interactive visualization techniques have emerged, allowing users to selectively explore and analyze data of interest, enabling the exploration of much larger data sets. A 3-D visualization of meteorological features fully relies on such digital displays and interactive capabilities. Rautenhaus et al. [149] provide a comprehensive and extensive state-of-the-art overview of visualization techniques and software solutions applied to a wide range of meteorological applications, highlighting the continued importance and advancements in data visualization for meteorology.

2.5.1 Feature visualization

A user-friendly toolkit that fits the needs of many different meteorological applications and allows interactive exploration of 3-D data helps developing

Parameter	Visualization technique
Bounding Box	Wireframe geometry
Centroid	Point sphere geometry
Intensity (average, peak)	Color coded
Semi-principal axes	Ellipsoid geometry
2-D projected semi-principal axes	Line geometry
Line string	Line geometry
Wind field	Streamlines, barbs
Identified object	Iso-surface
Other 3-D fields	Direct Volume Rendering, iso-surfaces

Tab. 2.3: Selected feature descriptors presented in the subsequent chapters, along with corresponding visualization techniques.

a better understanding of the data sets and the atmospheric processes. Visualization of identified atmospheric features help to focus on the regions and structures of interest. This facilitates the step forward from an exploratory to an explanatory analysis of meteorological features. Visualization techniques can be manifold in this context. Table 2.3 showcases several feature descriptors and data field types computed, and how these are visualized in the chapters of this thesis.

2.5.2 Met.3D

Met.3D [148] is an open-source tool designed for interactive, three-dimensional visualization of gridded atmospheric data sets. Utilizing state-of-the-art computer graphics technology, it is a powerful visualization tool in this thesis, offering a feature-rich and modular design optimized for meteorological data. The software is actively developed and maintained by the Visual Data Analysis Group at the House of Computing and Data Science at the University of Hamburg. It is an integral part of the Transregional Collaborative Research Center *Waves to Weather*, providing both support and computational resources to members of the Waves to Weather project.

Met.3D is compatible with Linux machines equipped with GPUs capable of OpenGL 4. It supports various vertical level types, projected grids, ensemble data, and multiple input formats. Unlike general-purpose visualization software such as ParaView [@96], which may lack meteorology-specific modules, or domain-specific tools like Vapor [@125] and IDV [@127], which occasionally fall short in features and support, Met.3D is custom-crafted for meteorological data sets.

One of the notable advantages of Met.3D is its focus on creating a transition or "building the bridge" [148] from 2-D to 3-D visualization. Meteorologists mainly focus their trained eye on visualizations of 2-D maps to understand processes in the atmosphere; therefore, setting up a bridge from a novel 3-D visualization to a known 2-D display is crucial. The software architecture of Met.3D follows a data processing pipeline scheme, akin to VTK [@97], and incorporates a memory manager to support caching of already processed data, optimizing performance and interactivity during exploratory analyses.

Details on the design and extent of the software can be obtained from Rautenhaus et al. [148] and the software documentation [@118]. For this thesis, several new additions to the open-source software have been contributed. A geometry reader has been implemented, which is able to read geometrical feature descriptors (i.e., principal axes of an ellipsoid) from CF compliant netCDF files. Furthermore, a geometry renderer has been implemented, which is capable of rendering both line and ellipsoid data to the scene. Existing visualization techniques based on line data (e.g., jet-stream visualization) have been incorporated into the abstract geometry interface. Instanced rendering is used to efficiently render these features to the display.

Meteorological background

3

> *The real voyage of discovery consists not in seeking new landscapes but in having new eyes.*
>
> — **Marcel Proust**

If you do not have a background or strong interest in meteorology, you might be wondering, "Do I need to be a meteorology expert to follow along?" The answer is a resounding no! Coming from a computer science background, I only encountered meteorology later during my master's studies and was fascinated by it. This perspective has helped me appreciate the importance of making complex meteorological concepts accessible and understandable to all. While a basic understanding of the atmospheric processes discussed in the upcoming chapters will be helpful, these processes are governed by physical laws that I will strive to explain clearly. Throughout this background chapter, I will simplify equations and focus on the essential aspects relevant to this thesis. For those interested in diving deeper, I encourage you to explore the cited books and papers in the appropriate sections to expand your knowledge.

3.1 3-D structure of the atmosphere

The Earth's atmosphere is a remarkable envelope surrounding our planet. It primarily consists of nitrogen and oxygen, with trace amounts of other gases like carbon dioxide and water vapor. Despite extending for hundreds of kilometers above Earth's surface, nearly 99% of the atmosphere's mass is concentrated within the lowest 30 kilometers.

The atmosphere's vertical structure is often conceptualized by dividing it into distinct layers. The closest layer to the Earth's surface is the troposphere, which extends from the surface up to approximately 6 km (around 400 hPa atmospheric pressure) at the poles and up to about 18 km (around 70 hPa) at the equator. This layer is where most of our daily weather phenomena occur. Temperature decreases with increasing altitude in the troposphere, as shown by the dashed contours in Fig. 3.1, primarily due to the expansion of air in lower-pressure conditions.

Just above the troposphere lies the stratosphere, which plays a significant role in shaping larger-scale weather patterns. Here, we'll find features such as jet streams and the ozone layer. The stratosphere sometimes noticeably influences the weather at the surface but with a time delay, acting as a sort of long-range forecast system [9]. In the stratosphere, temperatures rise with increasing altitude (see Fig. 3.1), primarily due to the presence of ozone. This temperature inversion creates a distinct boundary known as the tropopause, separating the stratosphere from the troposphere. Beyond the stratosphere, the atmosphere extends into the mesosphere and the thermosphere, where temperature variations are influenced by the vertical pressure gradient and solar radiation.

On the other hand, the horizontal variation of the atmosphere is mainly driven by the Earth's inclination and the distribution of solar radiation. Near the equator, where solar radiation is most intense, the atmosphere warms, making the air hotter and less dense. This leads to a pronounced height of the tropopause. In contrast, at higher latitudes, where solar radiation is less direct, the atmosphere remains cooler, and the tropopause is situated at a lower altitude. This interplay between air masses of different origin due to differing solar radiation is the basis that shapes the global circulation patterns.

For the purposes of this thesis, our primary focus will be on the troposphere and the tropopause. Anomalies along the tropopause have been identified as relevant for extratropical weather patterns [82, 169].

3.2 Concept of Vorticity

Rotation is a fundamental force shaping the dynamics of the atmosphere, from mesoscale systems and tornadoes to large-scale tropical and extra-tropical cyclones. It is defined based on the circular movement around a central line, the axis of rotation. In meteorology, we often simplify this vector quantity to

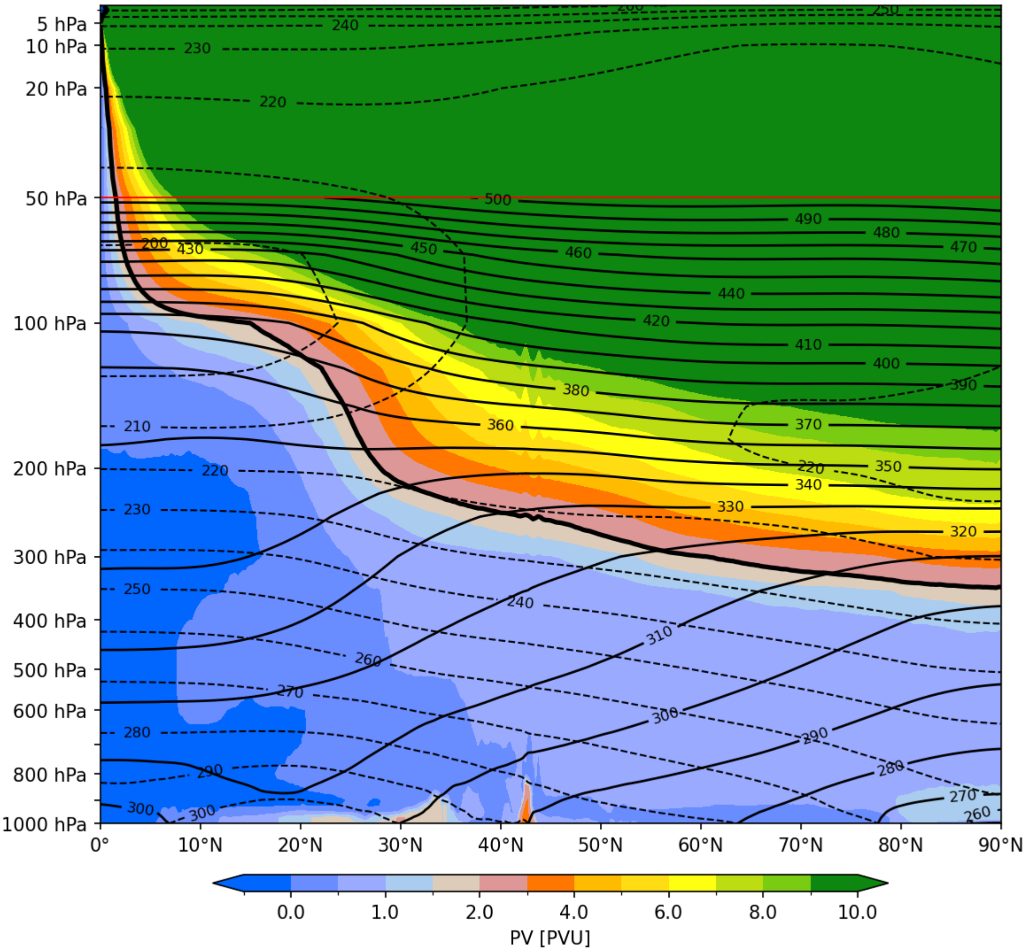


Fig. 3.1: A cross section of the mean state of the atmosphere along the prime meridian. Visualized is the temperature (dashed contours), potential temperature (solid contours), and potential vorticity (shading) along pressure levels from 0°N to 90°N. The thick solid contour depicts the dynamical tropopause, defined by the 2 PVU threshold. To keep the increase in temperature in the stratosphere visible, isentropes are only drawn until the 50 hPa line and a linear vertical scale is chosen in the upper-levels. The mean is calculated over all seasons from 2002 to 2022 and is based on the ERA5 reanalysis.

its vertical component, representing the horizontal component of rotation. This simplification leads to the definition of relative vorticity ζ_r as

$$\zeta_r = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}, \tag{3.1}$$

where u and v denote the zonal and meridional wind component on the xy plane. Earth as constantly rotating body exert also an additional fictitious force on masses in the atmosphere, the Coriolis force. In the northern hemisphere, this force deflects masses to the right, and ultimately governs the equilibrium in rotation around high and low pressure systems. In the reference frame of the earth, this Coriolis force can be expressed as $f = 2\Omega sin(\phi)$, where ϕ is the latitude at a given point and Ω the angular velocity of Earth. When considering this Coriolis effect, the absolute vorticity (ζ_a) is expressed as:

$$\zeta_a = \zeta_r + f. \tag{3.2}$$

3.2.1 Shear and Curvature Vorticity

While vorticity can originate from various sources, certain applications may be particularly interested in specific components of the vorticity field. Unwanted contributions can obscure the data, complicating the analysis. To address this issue, we can decompose the relative vorticity into different terms by expressing it in natural coordinates. Such a coordinate system is not spatially uniform, but always follows the motion of the air parcels, and is spanned up by a unit vector $\mathbf{s}$ parallel to the air parcel (parallel to the stream function) and a unit vector $\mathbf{n}$ orthogonal to it. As further elaborated by Holton [81], this approach yields an expression for ζ_r as follows:

$$\zeta_r = -\frac{\partial}{\partial n}V + \frac{V}{r}. \tag{3.3}$$

Here, n represents the direction orthogonal to the stream function, V denotes the horizontal wind speed, and r is the radius of the curvature of the wind field at a particular point. The sign of r can be positive or negative depending on whether the field exhibits a cyclonic or anticyclonic behavior at that particular point.

Both terms on the right hand side of Eq. 3.3 have nice interpretations. The first term, known as *shear vorticity*, resembles vorticity by the rate of change

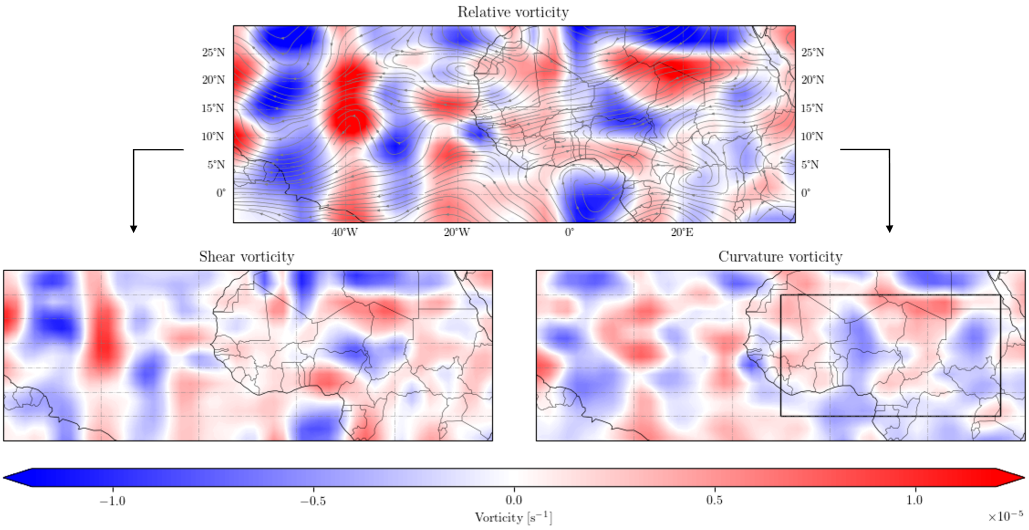


Fig. 3.2: Relative vorticity ζ_r (top), shear vorticity $-\frac{\partial}{\partial n}V$ (left), and curvature vorticity $\frac{V}{r}$ (right) at 700 hPa over southern West Africa and parts of the tropical Atlantic basin on 10 August 2021, derived from the ERA5 reanalysis. The wind field has been band-pass filtered (2–8 days) and smoothed. The black box highlights a region with a strong African easterly jet, where curvature vorticity more clearly depicts the flow's undulations, indicated by the alternating red and blue areas.

of wind speed normal to the direction flow. Therefore, just a sheared environment already induces vorticity. The second term, referred to as *curvature vorticity*, reflects the turning of the wind along a streamline. Figure 3.2 shows an example of splitting the relative vorticity in these two terms.

This formulation allows us to isolate the vorticity resulting from the curvature of the wind field. The scenario depicted in Fig. 3.2 is one such example: In Chapter 6, we are interested in the undulations of the easterly flow over northern West Africa, as visible by the streamlines over this region. The relative vorticity, however, originates not only from the actual advection of curvature vorticity along the jet (African easterly jet, see Sect. 3.4), but also from the temporal and spatial variability in shear introduced by the jet itself. Just by analyzing the curvature vorticity in Fig. 3.2, the red-blue alternating vorticity signature over Africa can clearly be identified contrary to the relative vorticity view.

3.2.2 Vorticity Equation

To investigate vorticity features in the atmosphere, we need to explore how the vorticity evolves over time in its atmospheric surroundings. This is described by the vorticity equation. This equation takes on slightly different forms depending on the vertical coordinate in which the equation is expressed. For instance, following Holton [81], if we examine relative vorticity on an isobaric surface (ζ_p), often referred to as isobaric vorticity, the equation unfolds as follows:

$$\frac{D\zeta_p}{Dt} = -(\zeta_p + f)(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}) - \frac{2\Omega \cos\phi}{a} v + (\frac{\partial \omega}{\partial y}\frac{\partial u}{\partial p} - \frac{\partial \omega}{\partial x}\frac{\partial v}{\partial p}), \tag{3.4}$$

where p is the atmospheric pressure, a the Earth's radius, and $\omega = \frac{Dp}{Dt}$ the change in pressure.

Every term in this equation has a practical interpretation. The left side reflects the change of vorticity over time. The first term on the right represents the effect of horizontal divergence (compression or stretching), akin to a figure skater extending or pulling in their arms to spin faster or slower. The second term represents is the northward variation of the Coriolis parameter times meridional wind v, and describes the change in vorticity resulting from latitudinal displacement of the air parcel: An increase in planetary vorticity along y has to come along with a decrease of isobaric vorticity ζ_p. The third term, often called the twisting or tilting term, captures a change of the vorticity direction. Horizontal gradients in vertical velocity tilts the vorticity vector towards the p axis, increasing the vertical component of the vorticity.

When examining isentropic levels under adiabatic conditions, air parcels move along isentropic surfaces, causing the vertical component ω to become zero. By expressing the vorticity along isentropes as ζ_θ, following Holton [81], the twisting term vanishes, leading to the simplified equation:

$$\frac{D\zeta_\theta}{Dt} = -(\zeta_\theta + f)(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}) - \frac{2\Omega \cos\phi}{a} v. \tag{3.5}$$

These simplified properties of vorticity along isentropic surfaces motivates a special framework for investigating upper-level flow in the atmosphere based on conserved vorticity budgets: The Potential Vorticity framework.

3.3 Potential Vorticity

Many different frameworks can be utilized to study atmospheric dynamics and to understand dynamical processes involved. One framework that became widely established in the late 20th century is the Potential Vorticity (PV) framework. Since the features being identified and analyzed in this thesis involve an investigation of their PV structures, an introduction to this concept is necessary.

PV provides a conceptually appealing perspective on atmospheric dynamics, particularly when analyzing upper-level dynamics. It can be considered a measure of the ratio of absolute vorticity ζ_a (Eq. 3.2) to the effective depth of the layer containing the vortex [81]. Hoskins et al. [83] offers a comprehensive review of the usefulness of PV as a diagnostic tool on isentropic surfaces, highlighting its role in understanding the connections between upper tropospheric PV dynamics and surface weather systems.

Following Ertel [44] and Hoskins et al. [83], PV is defined as the product of absolute vorticity and the gradient of potential temperature. In a framework where horizontal motion predominates, this relationship is expressed as:

$$PV = \frac{1}{\rho}\zeta_a \frac{\partial \theta}{\partial z}, \tag{3.6}$$

where ρ denotes the density and θ the potential temperature. The term $\frac{\partial \theta}{\partial z}$ represents the gradient of potential temperature with height and is commonly referred to as the static stability of the air parcel. When this gradient is positive, the atmosphere at this point is stably stratified, and vertical motions are suppressed. If the potential temperature decreases with height, hence $\frac{\partial \theta}{\partial z} < 0$, the atmosphere is unstable under vertical motion.

Coming back to the structure of the atmosphere, the stratosphere is characterized by its high static stability due to an increase in potential temperature with height, as illustrated in Fig. 3.1: isentropic levels are much denser in the stratosphere. Therefore, higher values for PV can be found in that layer. Moreover, an inclination towards higher PV values is observed towards polar regions due to differences in heating and a stronger planetary vorticity f. As a result, isentropic surfaces are generally tilted upward towards the pole, and can also intersect levels of constant PV. Relative maxima in PV correspond to cyclonic curvature (positive ζ_a) and high stability (high θ), minima correspond to anticyclonic features and warmer air, typically associated with ridges and high pressure systems.

The PV framework proves particularly valuable for analyzing upper-tropospheric flow due to two key aspects:

- *Conservation in adiabatic processes*: Following Rossby [159] and Ertel [44], the frictionless evolution of PV can be expressed as

  $$\frac{D}{Dt}PV = \frac{1}{\rho}\zeta_a \frac{\partial}{\partial z}\left(\frac{D}{Dt}\theta\right). \tag{3.7}$$

  Thus, considering adiabatic processes where the diabatic heating rate ($\frac{D}{Dt}\theta$) is zero, PV is conserved under idealized conditions. This conservation implies that an air parcel maintaining its potential temperature also conserves its PV. Therefore, air parcels conserve PV along isentropic surfaces, resulting in quasi-horizontal motion becoming the sole dynamics involved when both friction and diabatic heating are absent. Haynes and McIntyre [73] underline this conservation of the PV by showing that within a layer bounded by two isentropic surfaces, PV can neither be created or destroyed. This allows us to analyze air mass movements exclusively on isentropic levels, making these maps *physically meaningful and conceptually beautiful* [141].

- *Invertibility Principle*: PV follows the invertibility principle, which allows for the determination of the stream function of the field solely from the PV distribution, assuming balance constraints and background conditions [83]. This principle is based on the relationship between vorticity ζ and stream function Ψ through the Laplace Operator $\zeta = \nabla^2\Psi$. By knowing the stream function, one can gain insights into the field dynamics and infer other quantities, such as temperature and density.

3.3.1 The significance of PV in the extratropics

In the extratropics, PV plays a significant role in understanding the behavior of the upper-level troposphere and the stratosphere. As illustrated in the vertical cross-section in Fig. 3.1, tropospheric air has typical values around or below 1 PVU, and stratospheric air exceeds 2 PVU ($1\,\mathrm{PVU} = 10^{-6}\mathrm{K\,kg^{-1}m^2s^{-1}}$). This motivates the definition of the "dynamical tropopause." Due to the classic definition of the tropopause being based on the reversal of the vertical

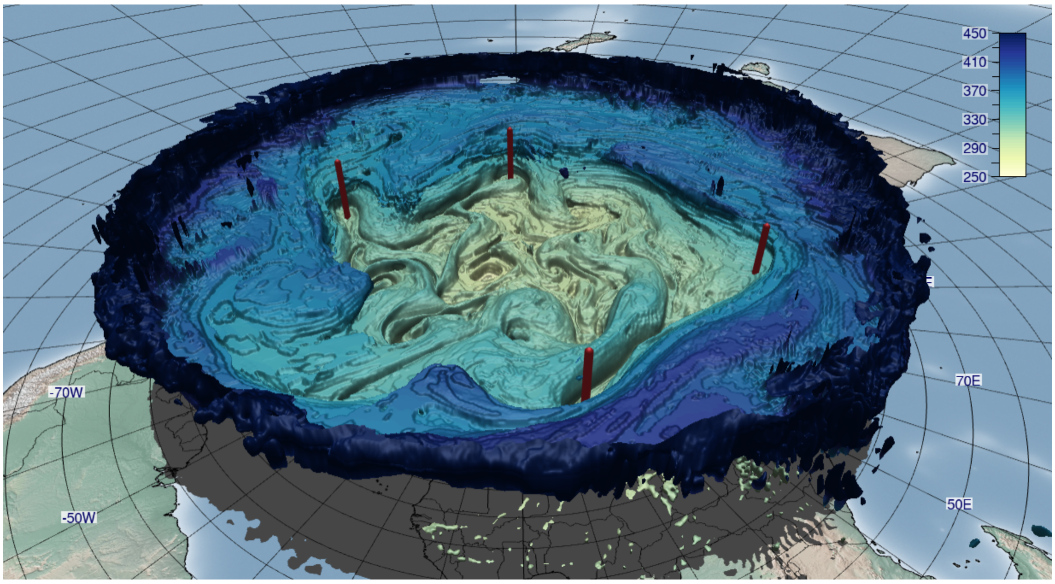


Fig. 3.3: A 3-D visualization of the northern hemisphere depicting the dynamical tropopause defined by the 2 PVU isosurface shaded by potential temperature. The complex dynamics in place are well visible. The red poles denote locations of anomalies in the field, where stratospheric air penetrates into the troposphere.

temperature gradient and the close relationship between temperature and potential temperature, the manifestations of both boundaries fall closely in line with each other. Empirically, a threshold value of 2 PVU is often employed to identify the dynamical tropopause in data fields [18]. This is visually represented by the black thick contour closely following the PV field in Fig. 3.1. This definition is based on the idea that the region of the strongest PV gradient is a suitable indicator of the tropopause [154]. In 3-D data fields, the tropopause is not anymore represented by one or multiple iso-lines, but instead by isosurfaces. Figure 3.3 shows an exemplary state of the dynamical tropopause in all its complexity.

Inspecting this 3-D visualization, contrary to the mean state of the atmosphere depicted in the cross-section in Fig. 3.1, wave-like anomalies can be observed in many places. Particularly, prominent undulations appear in regions marked by the vertical poles. Already the regularity of these undulations in this specific case point to the presence of large-scale upper-level waves. These are a result of baroclinic instabilities and their manifestation was theoretically introduced by Carl-Gustav Rossby [159]. Subsequently, these waves, commonly referred to as Rossby waves, led to the pioneering work by Rossby and Eliassen [39] to introduce the concept of PV for describing their behavior. Rossby waves manifest as meridional undulations of the frontal zone, separating cooler air on the northern side from the warmer air on the southern side. These undulations materialize as anomalies within the PV field, where high-PV air masses from the stratosphere, often termed the "stratospheric reservoir," penetrate into the troposphere. Consequently, an anomalous low tropopause height can be observed, which takes the shape of "valleys" in the 3-D view.

Around the time when Rossby proposed his framework, Sutcliffe [177] introduced a theory that explained surface convergence and divergence as consequences of upper-tropospheric geostrophic flow. This theory provided a tangible link between surface weather phenomena, specifically extratropical cyclones, and upper-tropospheric flow. Subsequent work by Charney [23] and Eady [36] showed that kinetic energy within extratropical cyclones can actually arise from these undulations. Consequently, Rossby waves exhibit tight coupling with cyclones and anticyclones observed at the Earth's surface.

Instabilities in Rossby Waves can lead to a phenomenon known as Rossby-Wave Breaking (RWB) events, which involve irreversible deformations of PV iso-surfaces. Typically, RWB events are characterized by a meridional overturning of Ertel's potential vorticity along an isentropic level [116]. Figure 3.4 depicts a visual representation of this process and its manifestations in both 2-D and 3-D perspectives. Note that the left column illustrates the 2-D view, while the right column provides a 3-D view along the blue line depicted on the left; this region corresponds to the RWB event.

In panels (a)-(b) and (c)-(d), an elongated filament along the 180°E meridian is prominently visible, extending far from the stratospheric PV reservoir. This filament is already laterally deformed, leading to a tilting of the trough over the Pacific Ocean near Hawaii. Commonly referred to as a PV streamer, these elongated high-PV filaments exhibit a 3-D appearance resembling 'valleys,' indicative of regions with anomalously low tropopause heights. However, it's crucial to note that these valleys do not signify a local minimum of the tropopause height at this stage.

Moving on to panels (e)-(f), the wave in the 2-D view seems to detach from the primary reservoir, creating so-called 'drops' or PV 'cut-offs'. Stratospheric air masses, forming these drops, become separated from the reservoir on the corresponding 2-D view. In the 3-D representation, most of these drops maintain vertical connectivity to the PV reservoir, now manifesting as local minima in the tropopause height.

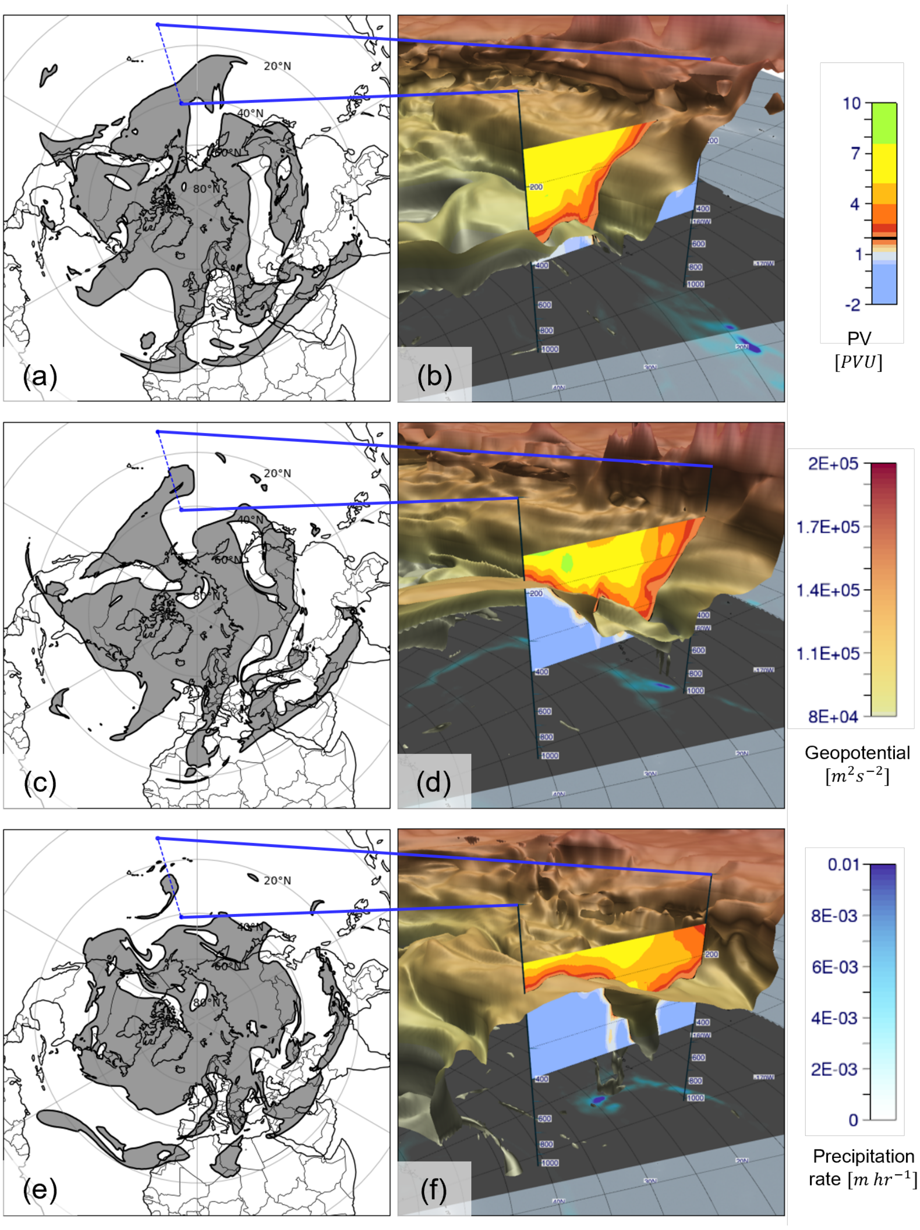


Fig. 3.4: Three phases of a Rossby-wave breaking event are illustrated in the three rows. The left side displays the 330 K isentrope, with white regions representing tropospheric air (<2 PVU) and gray regions indicating stratospheric air (>2 PVU). The boundary marks the dynamical tropopause. On the right, a vertical cross-section cuts through the wave anomaly along the blue line. The iso-surface is the dynamical tropopause, shaded by geopotential. At the surface, the precipitation rate is depicted in blue. Panels (a)-(b) display the state on March 18, (c)-(d) on March 20, and (e)-(f) on March 22, 2023 at 00 UTC. Data taken from the ERA5 reanalysis.

As evident in panel (f), these anomalies can extend deeply into the troposphere, significantly influencing surface weather. Porcù et al. [139] demonstrated that more than a third of such cut-offs over the Mediterranean are associated with surface cyclones. In the scenario depicted in Fig. 3.4, heavy rainfall was observed in parts of the central Pacific Ocean. Remarkably, this PV anomaly prompted the National Weather Service in Honolulu to issue a flood watch for the island of Oahu due to the anticipated heavy rain [@31], even though the island is located approximately 1,000 km away from the center of the PV anomaly.

The significance of these PV anomalies along the tropopause have gotten more attention over the last decades, mostly by investigating streamers and cut-offs on isentropic layers. In Chapter 5, we will take the identification of such anomalies one step further: by looking at them from a 3-D perspective.

3.3.2 The significance of PV in the tropics

In tropical regions, the conceptual elegance and analytical behavior of the PV framework proves to be more difficult due to different environmental conditions. Near the equator, the horizontal component of the Coriolis force becomes small, contributing less to the vorticity analysis. Nevertheless, from Eq. 3.6 it becomes evident that the presence of vorticity is essential for the existence of PV. Furthermore, convective events involve diabatic heating, leading to non-conservativeness of PV, and an examination of the non-conservative processes influencing PV becomes vital. Müller et al. [122] investigated the PV structure in proximity to convective systems, emphasizing both a horizontal and vertical PV dipole configuration. The horizontal dipole structure has also been identified by other researchers within convective updrafts, and has been linked to dynamic processes [151, 199, 198, 132]. These structures only become evident through analysis of diabatic PV sources and sinks, rendering the adiabatic view (Eq. 3.5) unfeasible.

In convective systems, air parcels rise, cool adiabatically until condensation takes place. The condensation process releases latent heat, increasing the potential temperature of the air parcel over time ($\dot{\theta} > 0$). Therefore, the change of heating with height is positive as well ($\frac{\partial \dot{\theta}}{\partial z} > 0$). Consequently, the PV tendency in these lower tropospheric regions also becomes positive. This anomaly grows until it reaches the peak of diabatic heating. Above this point, the potential temperature decreases with height again ($\frac{\partial \dot{\theta}}{\partial z} < 0$), eventually returning to the surrounding state, thereby producing a negative PV anomaly and a resulting vertical dipole. Müller et al. [122] underline that while these anomalies might be more pronounced in the extratropics due to the Coriolis parameter f, they also manifest in the tropics as long as a background vorticity is present. This background vorticity can arise from various sources, including large-scale wave patterns like African easterly waves (Sect. 3.4).

To manifest a horizontal dipole of positive and negative PV anomalies, a potent horizontal gradient of vertical velocity coupled with significant vertical wind shear of horizontal wind is required [89, 122, 132]. Within idealized convective updrafts, where the maximum vertical velocity resides in the center and decreases with distance from it, coupled with an increase of horizontal wind with height (baroclinicity), the twisting term now contributes significantly to the vorticity equation (Eq. 3.4). This leads to a negative PV pole to the left, and a positive PV pole to the right of the wind shear vector. A high Coriolis parameter isn't essential for this analysis. Importantly for this work, Müller et al. [122] analyzed the precipitation associated with the PV dipoles and conclude that the magnitude and vertical extent of the dipole structure reach their zenith during moments of maximal precipitation intensity and are collocated with them as well. Therefore, they identify "the relative potential vorticity as a useful quantity to diagnose and study atmospheric phenomena and processes on the convective scale", suggesting PV anomalies may serve as a proxy for evaluating intense rainfall events. In general, the 3-D structure of PV on the convective scale has been explored to a limited extent.

3.4 African easterly jet and African easterly waves

During the summer months in the Northern Hemisphere, the southern West African land mass undergoes significant heating due to seasonal insolation. This results in pronounced temperature and moisture gradients between the Gulf of Guinea and the Sahara Desert, leading to complex atmospheric circulations. These circulations manifest in specific meteorological phenomena: the African easterly jet (AEJ) and African easterly waves (AEWs) [46].

The AEJ is a low-altitude easterly jet stream that forms in response to the strong surface heating over North Africa. When the surface heats up intensely during the day, it creates a positive temperature gradient from south to north.

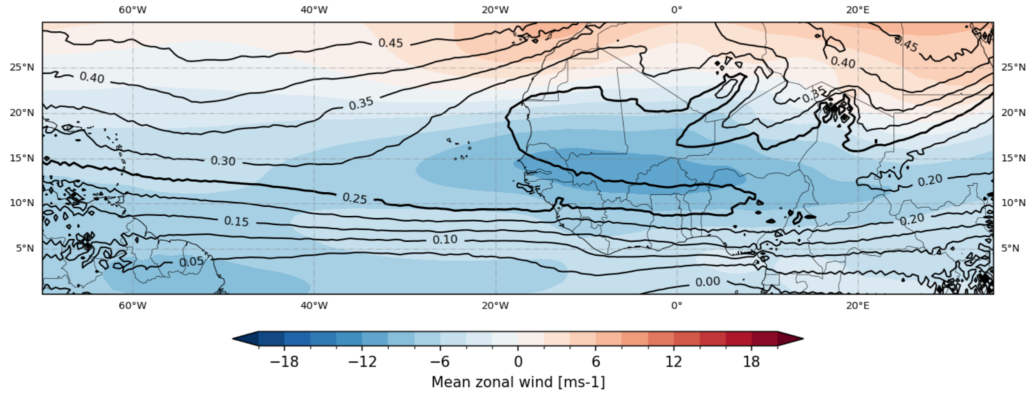


Fig. 3.5: Depicted is the mean zonal wind (shading) and mean potential vorticity (contours in PVU) on the 320 K isentropic level for the months spanning from June to October in the years 2002 to 2022. Data is taken from the ERA5 reanalysis.

This creates a baroclinic instability, where the warmer air expands and therefore levels of constant pressure have a meridional tilt. As a result, winds develop increasing with altitude, this vertical wind shear is known as thermal wind. The wind itself reaches its maximum around 600–700 hPa. The height of the AEJ can be traced back to the level at which the surface-induced positive temperature gradient transitions into the free atmosphere's negative temperature gradient [27]. While the winds in an inertial system would be northerlies, due to the Coriolis effect deflecting the winds to the right, a balance establishes. This balance of easterly winds is known as the AEJ. The shading in Fig. 3.5 shows the mean zonal wind in the summer months, when the AEJ is located roughly at 15°N. The AEJ is clearly visible by the maximum in blue shading. This zonal flow is rather unique because contrary to the planetary-scale air masses in the northern hemisphere, around the AEJ, the warm air is located to the north, and the colder air to the south. The AEJ is characterized by a peak in the zonal easterly wind of about 15 ms^{-1} and has a substantial impact on the dynamics of weather systems, including the formation of AEWs.

AEWs are disturbances in the AEJ that propagate westward across North Africa and the tropical Atlantic Ocean, primarily during the summer months, and are often associated with convective systems, and therefore PV anomalies as introduced in Sect. 3.3.2. The first researchers to observe and describe these waves were probably Regula [157] and Piersig [137] independently. Since then, countless studies investigated the properties of AEWs (e.g., [47, 21, 95, 117]).

Synoptically, two tracks of waves can be observed, which are located mainly between 5-20°N. The southern track, located around 10°N and peaking at around 700 hPa, is characterized by relatively moist waves, drawing moisture from the Gulf of Guinea. Conversely, the northern track, situated approximately at 20°N, features drier waves with their maximum intensity around 850 hPa [156, 47]. Typically, the periodicity of these waves ranges from 3 to 6 days, but can extend up to 8 days during the early monsoon season [62].

From a dynamical perspective, AEWs can be seen as a result of hydrodynamic instability along with a meridional gradient reversal of the PV [54, 24]. The PV along the 320 K isentropic level is depicted by the black contours in Fig. 3.5. The emphasized 0.25 PVU contour has a reversal over Western Africa, in the region where the AEJ is located. This reversal has the implication, that the PV increases, then decreases, and then increases again along a meridional line from south to north. This observed gradient reversal in PV satisfies the Charney–Stern necessary condition for mixed barotropic-baroclinic instability of internal jets [21, 38]. Then, as hypothesized by Berry and Thorncroft [14], convection over the Darfur mountain region and large convective outbursts can perturb the state of the AEJ, and these perturbations can grow downstream due to instabilities. These perturbations manifest as waves along the jet, as shown in Fig. 3.2.

However, it should be noted that this argumentation is only part of the complex dynamical setup still being researched, because easterly waves are often embedded and interacting with strong moist convection. Actually, Thorncroft et al. [183] showed that the large-scale baroclinic and barotropic instabilities are too weak to explain the fast intensification of these waves. Furthermore, Holton [80] suggested that latent heat released in convection is of primary importance for these waves. Thorncroft and Hoskins [184] confirm that AEWs grow by combined barotropic and baroclinic instabilities associated with the jet environment, but see the role of latent heating as vital as well. While AEWs are important in initiating organised convective systems and modulating summer rainfall over tropical West Africa during their westward propagation, the convection also could directly drive AEWs growth and propagation as well as modulate the strength and position of the AEJ [65, 161].

According to Russell et al. [160], over half of the tropical cyclones forming in the North Atlantic Basin have a direct origin in these AEW disturbances. Schlueter et al. [163] showed that up to 30% of rainfall variability in the Sahel Zone can be attributed to AEWs. These waves play a pivotal role in modulating rainfall patterns, which, in turn, have a direct bearing on food security, water resources, and the economic well-being of the millions of people inhabiting these areas [64].

Therefore, accurate rainfall forecasting in these areas is vital. Nevertheless, NWP models often exhibit very limited skill in tropical regions. This limitation is so significant that purely statistical forecasts, which rely solely on climatological information or meteorological variables from previous days, can surpass their performance [191, 146, 194]. To build on these statistical models, in Chapter 6, we introduce an identification and tracking algorithm for PV anomalies in AEWs. These features can serve as input predictors for statistical models or hybrid statistical-dynamical models.

4

A framework for identification and tracking of meteorological features

4.1 Motivation

The detection and tracking of features and their analysis are fundamental steps in the methodology of various geoscientific domains. Some recent examples include the objective identification of features in winter storms [37], jet streams [109, 91], convective systems [69], or the tracking of tropical cyclones [77] and Rossby wave packets [55]. While all these algorithms operate on specific subsets of data structures, require specific spatial and temporal resolution in certain cases, operate on analysis or ensemble forecast data, have different computational requirements, and generate output in specific formats suited for the different use cases, they connect the desire to generate a general representation of features and tracks in a format useful for and shareable to all researchers in an efficient parallelized environment.

A unified framework for identification and tracking enables efficient exchange of methods and results within and in between research projects, as well as facilitates broader exchange with the scientific community. In collaborative meetings within the *Waves to Weather* project, various similarities between different projects have been discovered, and a framework has been found to have a high potential impact and achieve mutual advantage in terms of productivity and exchange of methods. While the projects involved are very diverse in their nature (3-D analysis of frontal zones [10], evolution of PV uncertainties in the extratropics [164], identification of tropical wave features [48]), there was the strong need for a unification of the code bases.

While not all projects utilized the framework for identifying their respective meteorological features — often because methods had already been implemented in different programming languages before the framework became usable — these projects provided the initial motivation to develop this framework specifically for identifying and tracking features within meteorological data sets. Within the Waves to Weather project, the *enstools* framework has been developed in Phase 1 [152], containing modules and functionality for various data processing steps (e.g., input/output functionality and data compression). Custom modules can easily be added by researchers, contributing to enstools. Therefore, it has been decided to make the planned feature identification framework a part of enstools, which can then be easily deployed together with the other modules, as *enstools-feature*.

4.2 Goals

Several goals for the framework have been discussed and defined. These goals were shaped by the collective experiences and necessities of the researchers. Throughout the software development process, the requirements underwent iterative refinement. Some initially considered features were deemed non-essential as time progressed, conversely, new features emerged as a result of this ongoing feedback process. These features align with the overarching guiding principles established below.

4.2.1 Functional scope and features

Specific aspects have to be considered in particular when designing software for researchers. The framework should

- provide adequate **IO functionality**. Specific data formats are used in geoscientific research, that are suited to reflect the different types of grids and data types that exist. While clearly no software can support all data formats in this scientific domain, the goal is to support the majority of widely used formats, which are also used by most data distributors (weather services like DWD and ECMWF), including netCDF [@128], and GRIB [@203]. The data can be present in different projections.
- provide a **robust**, **efficient** and **automated** environment. With the high granularity of meteorological data and the unimaginable amount of generated and observed data, the framework should be well scalable to these demands. Computations are in research environments often

done on computing clusters, where the framework should be capable of exploiting the parallel setup of the hardware.

- be **easy** to install and learn. It should be in a programming language researchers are already familiar with. For non-expert programmers, learning a new programming language to use a framework might be a hurdle that is not worth it to overcome. Extensive documentation and templates as starting point can be a way to motivate researchers to use the framework.

- provide an environment for easy **prototyping**. Researchers often need to rapidly prototype and test new ideas. The framework should support easy implementation of new prototypes of identification and tracking strategies quickly, apply these to big data sets in a scalable fashion, and provide an environment to compare the prototype to existing algorithms.

- be **transparent** and follow open source philosophies. This allows the researchers to access, review, and modify the framework and the implemented features. Publications in geoscientific journals more and more set high standards for open source policies. The publication guidelines employed by the big geoscientific publishers *Copernicus* [@28] and *AMS* [@4] clearly state that code published along an article should be archived in *FAIR*[1]-aligned repositories. The software should be accessible and preserved in a trusted repository and contain appropriate licensing statements.

The framework's boundaries have also been carefully defined to maintain a concise and focused functionality, avoiding unnecessary complexity. To create visualizations of features, other software solutions should be considered, e.g. Met.3D (see Sect. 2.5.2). The output IO structures of the framework and added geometry functionality in Met.3D allow for straightforward visualization. Furthermore, while the framework strives for efficiency, it does not automatically optimize the user's algorithms for identification and tracking. The responsibility of code optimization remains with the user. The same applies to analysis tasks.

A few guiding terms have emerged to achieve the aforementioned goals in diverse ways.

---

[1] **F**indability, **A**ccessibility, **I**nteroperability, and **R**euse

Unification

In the current state of academic research in meteorology, researchers typically maintain their individual code bases and scripts for feature extraction and data analysis. This code heterogeneity poses challenges. This framework seeks to unify these approaches, offering a cohesive platform that harmonizes various identification and tracking algorithms. It facilitates code sharing, collaboration, and offers a wide range of meteorological feature extraction methods. The framework should find a balance of being broad enough to accommodate various needs and constraints of feature detection and tracking algorithms, while being specific enough to the meteorology domain to offer tailored functionality. In this framework, unification will mostly be achieved by having clearly documented interfaces and useful output types, as well as by providing an optimized backbone tailored to the needs of geoscientific data analysis.

Modularity

Researchers should be able to seamlessly interchange individual identification and tracking strategies, and integrate them effortlessly. For instance, two cyclone identification algorithms should be seamlessly interchangeable, and without further changes the same tracking strategy should be applicable, as identification or tracking modules remain highly independent. This facilitates comparative studies where researchers evaluate the performance of distinct algorithms within the framework both qualitatively and quantitatively. Researchers can work on improving individual modules without affecting the overall framework. Some identification strategies with straightforward heuristics (e.g., proximity of features, overlap of area features) should be applicable easily if the data descriptions provide these parameters. In this framework, modularity is achieved by the architecture that will be outlined in Sect. 4.4. Abstract classes with simple interfaces between the data processing steps ensures that modules can easily be interchanged and decoupled.

Usability and Documentation

In interdisciplinary research, particularly where users might lack either extensive programming or meteorological expertise, usability is crucial. An excessively high learning curve can discourage researchers from adapting their existing algorithms to fit the framework or using it at all. Therefore, the framework should provide an intuitive, user-friendly, and efficient experience regardless of the researcher's background. Documentation should be accessible in web-like formats, including setup instructions and module details, as well as templates and sample data. Further enhancing usability, the framework should seamlessly integrate with systems and environments already familiar to researchers, including third-party Python libraries.

4.3 Core dependencies

The framework's design revolves around carefully selected libraries and standards for meteorological research. Python was the natural choice for the framework's implementation, given its popularity in the scientific community, ease of use, and wealth of libraries, especially for atmospheric data analysis. The language's rapid prototyping capabilities further enhance algorithm implementation in the framework.

Anaconda

Anaconda [@5] is a distribution of the Python programming language suited for scientific computing, with a key focus on package management and deployment. It enables users to create isolated virtual environments for their Python code, ensuring deterministic and replicable deployment across various systems. It can be seen as a Docker environment tailored to Python programming. Environments can be defined and shared using generated *.yml* files. This framework also provides such a file so users can directly create a functional environment to run the framework in.

enstools

The feature identification framework has been realized as a module within the *enstools*[2] Python package. This package, developed and maintained within the Waves to Weather project, provides a modular environment for processing meteorological data, with a special focus on ensemble data. This package

---

[2] Ensemble Tools

encompasses multiple modules for tasks ranging from clustering and interpolation to IO and data validation. It is available on GitHub [@195] and archived at Redl et al. [153]. A web-based documentation is available [@196].

xarray

xarray [@204] is an excellent choice for working with meteorological data within our framework due to its versatility and efficiency in handling multidimensional labeled data. Its powerful features align well with the complexities of meteorological data sets. This Python library allows for intuitive and, thanks to its C back end, very efficient computations. The open-source library is well integrated within the meteorological community, supporting various data formats and coordinate-aware operations. An extensive documentation [@205] is available.

JSON

JSON (JavaScript Object Notation) is a lightweight and human-readable data interchange format that has gained widespread adoption due to its simplicity and versatility. The seamless integration of JSON in Python streamlines data parsing and generation, offering a familiar structure compatible with Python's data handling. This simplicity and versatility make JSON a natural fit for the framework, enabling easy data exchange.

Protobuf

In tandem with JSON, *protobuf* (Protocol Buffers) [@61] serve as an additional data format within the framework. Protobuf is essentially JSON in binary format, offering reduced memory consumption and efficient data exchange capabilities. Therefore, it is particularly beneficial for serializing boundary descriptions of area features or other memory intensive data. Additionally, the required schema definition, so called *.proto* files, provide explicit data structure and aiding in feature documentation. Such a file therefore explicitly contains the definition of the features. Protocol Buffers are compatible with many programming languages, and parsing these files is straightforward with the existing schema definition file.

4.4 Implementation

Figure 4.1 shows the main classes involved in the data processing pipeline. The `Pipeline` class encapsulates the program flow and has the currently set identification algorithm, tracking algorithm and data set as attributes to operate on. To achieve the goals of unification and modularity, the interfaces have a clear definition and the implementations of identification and tracking algorithms have to inherit from the according abstract classes to be able to be used in the framework.

4.4.1 Interfaces and data processing pipeline

Through dedicated setter functions, identification and tracking algorithms can be configured within the processing pipeline. Input data files should adhere to the CF Metadata Conventions [71] introduced in Section 2.1. Following these conventions, crucial metadata for the algorithms present in the input files (e.g., dimensions and units) can be detected automatically. Supported data dimensions in our presented framework encompass the three spatial dimensions, time, initialization time of a forecast, and ensemble member. Users can specify the `processing_mode` when initializing the pipeline, choosing between the 2-D and 3-D modes, based on their specific needs. In the 2-D mode, if a vertical dimension is present, the framework processes levels separated in 2-D chunks. The 3-D mode processes data in 3-D chunks, typically suited for spatial 3-D features. The classes relevant for the identification and tracking interface are shown in the class diagram in Figure 4.1. An exemplary implementation of this interface is shown in Sect. 4.5.

The identification interface provides a streamlined method for implementing identification strategies. It consists of three key methods:

- Precomputation `precompute(dataset)`: Prior to executing the identification algorithm, the precomputation step prepares the input data set. Users can implement custom actions, such as regridding and computation of derived fields, to ensure the data is optimally configured for the algorithm.
- Identification `identify(chunk, index)`: This method operates on individual 2-D or 3-D chunks and is called in parallel. Here, the actual

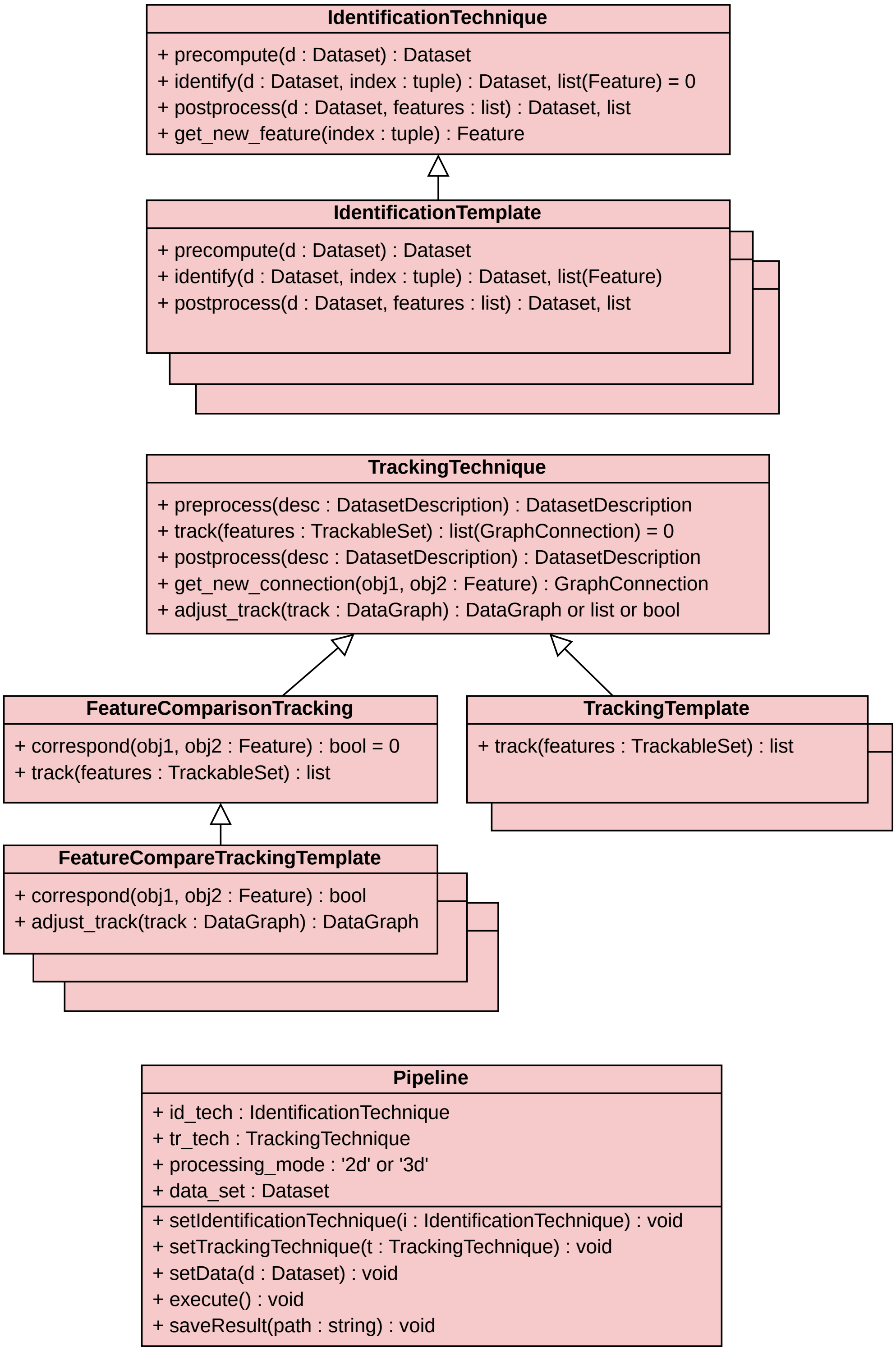


Fig. 4.1: Classes involved in the data processing pipeline.

identification of features takes place. The method yields a list of identified features for the specified block, as specified by the feature definition. The data block itself can also be altered and returned, e.g. by writing a bit mask. The given index contains relevant meta data of the subset relative to the full data set.

- Postprocessing `postprocess(dataset, features)`: After all data blocks have undergone identification, the postprocessing step can be used for a review and modification of both the data set and identified features. Users can perform sanity checks, analyses, and potentially filter the features before returning and serializing the result.

The tracking module interface also consists of several functions:

- Precomputation `precompute(features)`: Similar to the identification interface, here, preprocessing can be performed on the identified features, which was either loaded from file or directly forwarded from the identification algorithm.
- Tracking `track(track_set)`: This method facilitates feature tracking parallelized over the available *trackable sets* in the feature description. Trackable sets encapsulate disjoint subsets of data and feature descriptions, which can be tracked individually. One trackable set thus is defined by the spatial dimensions and a time dimension. The method processes the data from the set, returning a list of feature connections which represents the solution to the correspondence problem (see Sect. 2.3.1) that subsequently form the basis of the tracking graph.
- Postprocessing `postprocess(feature_desc)`: Users can implement this function to refine the tracking results. It allows for targeted filtering of tracks to enhance the robustness of tracking outcomes.
- `adjust_track(track)` can be implemented to return an altered track for each identified track, or discard it entirely, for example if it does not reach a specified life time.

Furthermore, as a realization of the tracking interface, the abstract tracking strategy `FeatureCompareTracking` is provided. This strategy simplifies the correspondence problem by making it only dependent on the two features in question. This is realized by requiring subclasses to implement the `bool correspond(`f_1`,`f_2`)` instead of the generic `track()` function. Executed for all pairs of features within a defined time delta in the set, this function determines if the features represent the same entity. Most general state-of-the-art tracking algorithms implemented in the framework fit into this sub scheme. Users can adjust the maximum time difference between two features to consider, with the default setting evaluating only consecutive time steps. More advanced tracking algorithms may require information from multiple time frames or a feature's historical context to decide on the correspondence problem. That is where the more general tracking interface becomes handy. A practical illustration of utilizing the `FeatureCompareTracking` technique is provided in Sect. 4.5.

The data processing pipeline, illustrated in Figure 4.2, shows how the interfaces come together. When the user initiates the pipeline with an `execute()` call, initial data preparation takes place. This includes analyzing input data sets, merging when possible, and extracting dimensions following the CF conventions. The pipeline splits input data into chunks based on dimensions and processing mode (2-D or 3-D) and runs the `identify()` method in parallel. Each chunk returns an optionally altered data chunk and a list of identified features present in the chunk. After processing, the identified features are labeled with meta-data and collected into the output structure.

If a tracking strategy is provided, there is an optional preprocessing step to filter the output of the identification for relevant data and features. The `track()` function processes each trackable set in parallel, and the results form the basis of the tracking graph. After processing all trackable sets, the resulting tracks are collected and serialized. From the tracking graph, tracks can be extracted based on the algorithms outlined in Sect. 2.3.3.

4.4.2 Data output

The manner in which data is presented, as noted by Hunt and Thomas [87, p. 75], has lasting significance: *"Human-readable forms of data, and self-describing data, will outlive all other forms of data and the applications that created them."* While human-readable output introduces overhead, it yields substantial advantages: It ensures platform-independent usability and rapid prototyping for instant data validation. However, for extensive feature descriptions, the additional overhead of human-readable formats may become problematic. In such cases, binary file formats that prioritize memory efficiency offer a suitable alternative.

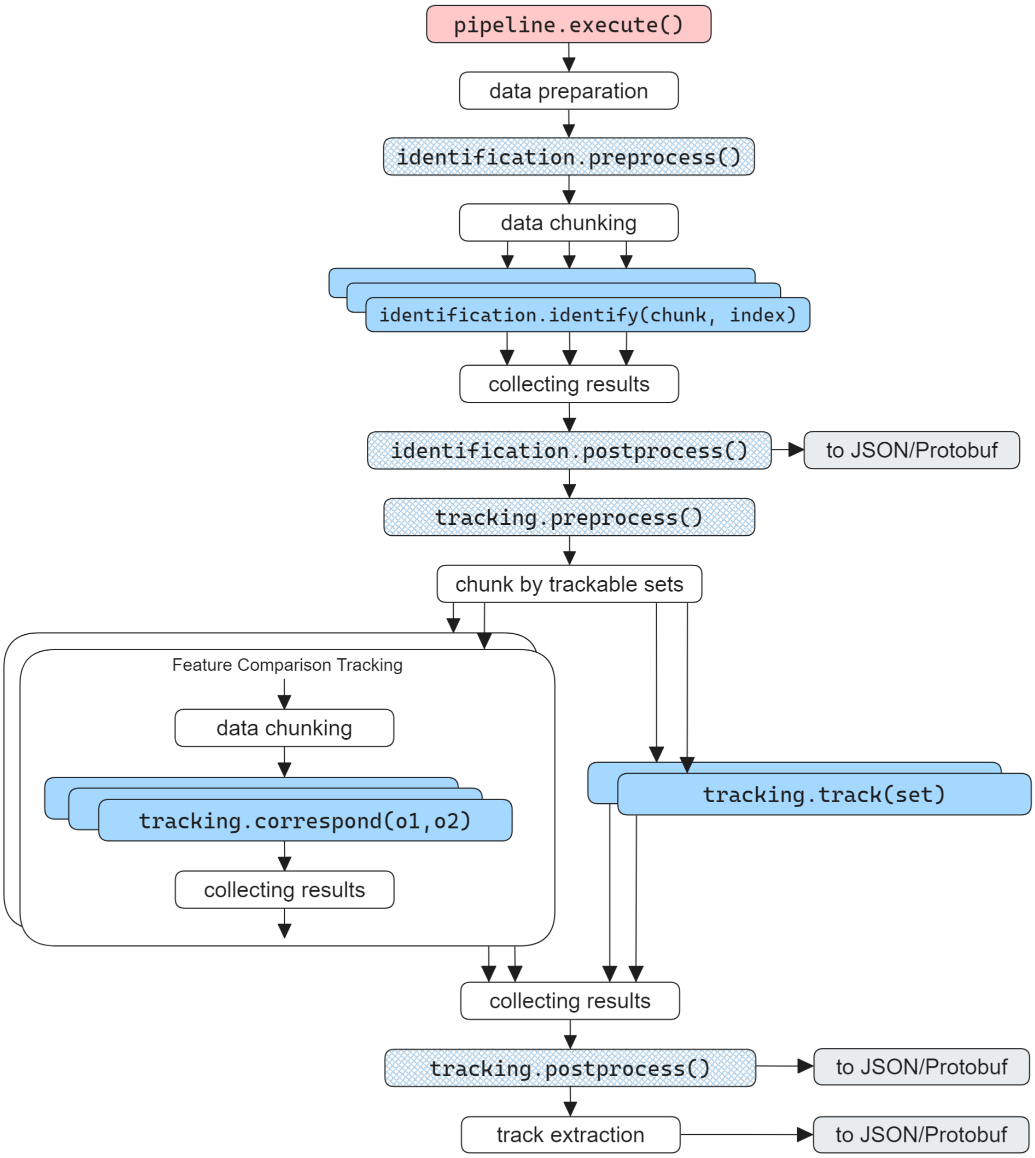


Fig. 4.2: A schematic of the data processing pipeline. The dark blue methods require an implementation by the researcher, the hatched ones are optional. The red methods are called by the user from the run-file.

Within this framework, users have the flexibility to employ both output formats: the human-readable JSON and additionally, the binary Protobuf format (as introduced in Section 4.3). Saving processing results is realized through a call to `save_result()`, where the type of output can be easily specified. Both output options necessitate a Protobuf description, which also serves as the foundation for defining feature descriptions.

Feature output

The default Protobuf description for a feature is represented in our framework by:

```
syntax = "proto2";
message Properties {
    // ...
}
```

The Properties message can be populated with data types like `float`, `string` and `boolean`, and more complex types in the form of nested messages, or lists of both types and messages. Section 4.5 illustrates a straightforward use case. This structure is intentionally designed to have a canonical encoding in JSON.

Identification output

After the identification part of the pipeline is done, the results are bundled into trackable sets. The results can be saved into a Protobuf or JSON file following the structure prescribed by the framework. The full code of the Protobuf description is available in the code repository of the framework as stated in the code and data availability appendix. Below is a concise summary of the structure.

The primary structure, `DatasetDescription`, encapsulates the process's metadata along with a collection of `TrackableSets`. Each `TrackableSet` contains a sequence of `Timesteps`. Each `Timestep` is characterized by its timestamp and the identified features. Each feature consists of an identifier unique to the trackable set, and the encapsulated attributes introduced earlier. This structure, while concise, is highly versatile in terms of handling various input data sets and feature descriptions.

Tracking output

When tracking is activated and executed, the results are saved to a JSON or Protobuf file containing additional fields. We again refer the reader to the source code for the complete Protobuf description files.

In this description, each tracked set now includes a graph representation, which resembles the tracking graph discussed abstractly in Sect. 2.3.2, as well as a list of disjoint graphs constituting the extracted tracks from this tracking graph. Each graph is constructed as an edge list, with the nodes being identified by their respective identifiers assigned during the identification process. Notably, the format does not constrain the edge to connect solely with features from consecutive time steps. In a verbose mode, not just the identifiers, but the entire features are copied to the tracks instead of just referencing the identifiers.

4.5 The users perspective: cyclone tracking

To illustrate the simplicity of the framework from the user's perspective, let's consider a simple task: implementing an identification and tracking strategy for tropical storms based on surface pressure data. Despite the existence of sophisticated algorithms [78, 188] and already available storm track data bases [100], we'll keep it straightforward and focus on identifying and tracking positions of surface low-pressure systems.

First, we create a `.proto` file to define the structure of the storm features. This resembles the feature vector f introduced in Sect. 2.1.2. In our case, a storm at a given point in time is described by its position (longitude and latitude) and the minimum surface pressure at that point. Here's the `.proto` file:

```
syntax = "proto2";

message Position {
  required float latitude = 1;
  required float longitude = 2;
}

message Properties {
  required Position position = 1;
```

```
  required float min_pressure = 2;
}
```

For the identification step, we need to derive from the abstract identification class and implement the identification algorithm. If no preprocessing or postprocessing is required, these methods can be omitted. Parameters, such as the maximum pressure threshold, can be passed as arguments. Here's an example in Python:

```python

class StormIdentification(IdentificationStrategy):

    def __init__(self, pressure_threshold_hpa=1000, **kwargs):
        self.threshold = pressure_threshold_hpa

    def identify(data_chunk: xarray.Dataset, index):
        # Looking for minima, e.g. by gradient analysis
        minima = find_local_minima(data_chunk.pressure,
            threshold=self.threshold)

        objs = []
        for minimum in minima:
            # Get an instance of a new feature
            obj = self.get_new_feature(index)
            obj.position.latitude = minimum.latitude
            obj.position.longitude = minimum.longitude
            obj.min_pressure = minimum.pressure
            objs.append(obj)

        return data_chunk, objs


```

For tracking, we'll use the `FeatureCompareTracking` strategy. Therefore, the correspondence method needs to be implemented to determine if two storms represent the same entity. We consider two storms as the same entity if their movement speed does not exceed 25 ms^{-1}. Below is the Python code for implementing such proximity tracking:

```python

class StormTracking(ProximityTracking):

    def __init__(self, speed_range_ms=[0,25], *args, **kwargs):
        super(ProximityTracking, self).__init__(*args, **kwargs)
```

```python
        self.min_speed = speed_range_ms[0]
        self.max_speed = speed_range_ms[1]

    def correspond(obj1, obj2):
        t1, t2 = obj1.time, obj2.time
        pos1 = obj1.properties.position
        pos2 = obj2.properties.position

        # compute spherical distance
        distance_m = distance(pos1, pos2)
        speed_ms = distance_m / (t2 - t1).seconds()

        return (self.min_speed < speed_ms < self.max_speed)
```

And that’s it. We have an identification and tracking algorithm for storms, based on a pressure threshold for the identification and speed threshold for the tracking. To execute the identification and tracking algorithm, we create a run file in Python. This file sets up the pipeline, initializes the algorithms, specifies the data path, and executes the pipeline:

```python

from identification._proto_gen import storm_pb2
from identification.storm import StormIdentification
from tracking.storm import StormTracking

# initialize the pipeline with the processing mode
pipeline = FeaturePipeline(storm_pb2)

# initialize algorithms
i_strat = StormIdentification()
# With optional max time interval for correspondences
t_strat = StormTracking(max_compare_dt_hours=12)

# set the algorithms and data path
pipeline.set_data_path('*.nc')
pipeline.set_identification_strategy(i_strat)
pipeline.set_tracking_strategy(t_strat)

# execute pipeline
pipeline.execute()
pipeline.save_result(description_type='json')

```

Provided a data set with only a single vertical level, the above pipeline returns storm tracks on that specific 2-D field. When given data with multiple vertical levels, it returns multiple tracked sets, one for each level, with storms tracked with respect to the pressure data on the specific levels. With this setup, the generated JSON file contains all identified features, their connections in the tracking graph, and the extracted tracks.

4.6 Current and future use

The algorithms introduced in chapters 5 and 6 to identify PV structures and anomalies in the tropics and extratropics are available to use in the framework. Additionally, besides these quite specific algorithms, version 1.0 of the framework contains the following identification techniques:

- `TemplateIdentification`: A template with method stubs serves as a starting point for new strategies.
- `DoubleThresholdIdentification`: This technique identifies features based on specified threshold values of a given field. It supports both double thresholding and single thresholding (as a special case). The output consists of object features, each represented by a unique identifier. This identifier corresponds to a labeled data field, where identifier i represents the area where the data field equals i.
- `StormIdentification`: This technique offers a simple identification of storm centers based on local minima in the pressure field. The algorithm includes two key parameters: a pressure threshold and a prominence distance threshold. By setting these parameters, the algorithm assigns a point feature to each identified storm, denoting its center. The prominence distance threshold ensures that nearby local pressure minima are grouped together into a single feature, enhancing the stability of storm tracking.
- `VortexIdentification`: Identifies vortex centers by analyzing critical points in the streamlines of the wind field. Similar to the storm identification outlined above, a prominence distance threshold is applied to the resulting point features as well.

For feature tracking, on top of the strategies employed in chapters 5 and 6, the framework offers the following techniques:

- `TemplateTracking`: Similar to the template identification, this method provides a starting point with method stubs for customization.
- `ProximityTracking`: This tracking strategy solves the correspondence problem by pairwise comparison of features defined by points. It determines correspondence based on the distance between points, with input parameters specifying the minimum and maximum speed of the features. The storm tracking introduced above is a specific proximity tracking.
- `OverlapTracking`: Designed for area features, this tracking strategy checks for overlaps between features to establish correspondence.

The framework has already been used in various theses [102, 110]. It is our hope that in the future, researchers will utilize this framework as a starting point for developing their own feature extraction algorithms. They can leverage the framework's functionality and easily pass on their work to future researchers. The framework is continuously evolving and open to contributions. A documentation is available and hosted at *ReadTheDocs*[3].

---

[3] `https://enstools-feature.readthedocs.io/`

5

A novel method for objective identification of 3-D potential vorticity anomalies

5.1 Preface

This chapter covers the theoretical framework, the methodology, and a detailed case study of a novel method for identifying 3-D PV anomalies along the tropopause. A substantial portion of this chapter has been drawn from an article published in GMD[1] [49] and therefore undergone an anonymous peer-review process. While the core ideas originate from my work and the writing was mainly done by myself, the article has been bolstered by the invaluable insights, feedback, and revisions provided by my co-authors. The meteorological expertise for the case study has been provided by Roderick van der Linden. To emphasize the computer science facets inherent in this research, certain sections have been restructured, cross-referencing to the background chapters 2 and 3. Although the article was officially published in 2022, it's worth noting that the algorithm underwent subsequent minor adaptations and refinements. Notably, since the publication, the algorithm has been successfully applied in a master's thesis [102], with a summary of their findings found in Sect. 5.10. Moreover, substantial modifications have been introduced to the framework itself, to accommodate newly implemented algorithms that have been added since the publication. References regarding code and data availability can be found in the appendix.

5.2 Abstract

Potential vorticity (PV) analysis plays a central role in studying atmospheric dynamics and in particular in studying the life cycle of weather systems. The

---

[1] `https://www.geoscientific-model-development.net/`

three-dimensional (3-D) structure and temporal evolution of the associated PV features, however, are not yet fully understood. An automated technique to objectively identify 3-D PV features can help to shed light on 3-D atmospheric dynamics in specific case studies, as well as facilitate statistical evaluations within climatological studies. Such a technique to identify PV features fully in 3-D, however, does not yet exist. This study presents a novel algorithm for the objective identification of PV anomalies along the dynamical tropopause in gridded data, as commonly output by numerical simulation models. The algorithm is inspired by morphological image processing techniques and can be applied to both two-dimensional (2-D) and 3-D fields on vertically isentropic levels. The method maps input data to a horizontally stereographic projection and relies on an efficient computation of horizontal distances within the projected field. Candidates for PV anomaly features are filtered according to heuristic criteria, and feature description vectors are obtained for further analysis. The generated feature descriptions are well suited for subsequent case studies of 3-D atmospheric dynamics as represented by the underlying numerical simulation. We evaluate our approach by comparison with an existing 2-D technique, and demonstrate the full 3-D perspective by means of a case study of an extreme precipitation event that was dynamically linked to a prominent subtropical PV anomaly. The case study demonstrates variations in the 3-D structure of the detected PV anomalies that would not have been captured by a 2-D method. We discuss further advantages of using a 3-D approach, including elimination of temporal inconsistencies in the detected features due to 3-D structural variation, and elimination of the need to manually select a specific isentropic level on which the anomalies are assumed to be best captured. These advantages, as well as the suitability of the implementation to process big data sets, also open applications for climatological analyses. The method is made available as open-source for straightforward use by the atmospheric community.

5.3 Introduction

Weather systems and extreme weather events result from nontrivial three-dimensional (3-D) interactions in the atmosphere. The potential vorticity (PV) perspective on atmospheric dynamics provides an often-used conceptual framework to understand such interactions [as reviewed by 83]. A key component of this framework is that PV is materially conserved in the absence of nonconservative processes such as, e.g., latent heat release in clouds, longwave radiative cooling, or turbulent mixing. In the absence of strong latent heat release, it turns out that nonconservative PV modification is comparatively small and that material conservation of PV provides a very good first approximation. When considered on isentropic levels, the temporal evolution of PV at a given location is hence governed by quasi-horizontal advection only, which typically yields a rather smooth PV evolution.

Of particular importance to the PV perspective is the tropopause, as it separates air masses with typically low PV values in the troposphere from typically high-PV air in the stratosphere[2]. A strong PV gradient across the tropopause has motivated the definition of the so-called dynamical tropopause as an upper-level PV isosurface in the literature. Typically used PV values range between 1.5 – 4 PVU, with PV = 2 PVU being an established standard [e.g., 120].

The tropopause slopes from higher isentropic levels in the tropics towards lower isentropic levels at the poles. On a given isentropic level, poleward excursions of the tropopause denote an anomaly of low PV, whereas equatorward excursions denote an anomaly of high PV. A relatively smooth meridional undulation of the tropopause signifies quasi-linear Rossby wave dynamics. In the context of extreme weather, however, much attention has been given to highly non-linear PV features, in particular zonally narrow and meridionally extended ”tongues” of high PV that intrude equatorwards [114], known in the literature as PV streamers (PVSs). These elongated anomalies can have complex life cycles and behavior: They are often linked to Rossby wave breaking (RWB) events [e.g., 116, 181], interact with surface cyclones [e.g., 57, 13, 112], and may detach from the main stratospheric high-PV reservoir on a given isentropic level, forming so-called PV cutoffs. These cutoffs can diabatically decay, be reabsorbed by the main reservoir, undergo further dynamical processes, and can be associated with extreme weather events [141].

In the present study, we consider the objective identification of PV structures. Objective identification of atmospheric features from numerical simulation output has proven beneficial for a number of applications in both atmospheric research and operational meteorology. Typical uses include statistical analysis such as the computation of climatologies of feature occurrence [e.g., 109, 30] and operational weather forecasting purposes [e.g., 76] including ensemble

---

[2] PV is typically positive on the northern hemisphere and negative on the southern hemisphere. Low PV air here refers to low absolute values and high PV air to high absolute values. For notational convenience we consider the northern hemispheric situation only.

forecast analysis [e.g., 76, 148]. Recently, combination of state-of-the-art interactive 3-D visualization techniques (a comprehensive survey can be found at [149]) with 3-D detection of atmospheric features opened the door for comprehensive case studies of 3-D atmospheric dynamics [148, 91, 92, 7]. With respect to objective identification of PV structures, Papin et al. [134] recently provided an overview of identification techniques for PVSs. They classified the methods into techniques based on the reversal of the meridional PV gradient, and techniques based on distance thresholds along a 2-PVU contour. For example, a 2-D identification technique that has found wide acceptance and use, and may thus effectively serve as a state-of-the-art benchmark technique, has been introduced by Wernli and Sprenger [200]. This technique is based on a scale analysis, according to which the 2-PVU contour of a PVS contains points that are far apart in the direction along the contour, yet relatively close together regarding their great circle distance. Wernli and Sprenger [200] applied this technique to generate a climatology from ERA-15 reanalysis data, which Kunz et al. [104] later extended to a 33 year time period.

All techniques discussed by Papin et al. [134] operate on 2-D data only. Furthermore, gradient-reversal techniques only consider a subset of PVSs (often specifically focusing on RWB events), which are sensitive to small changes in the input data. To comprehensively study the atmospheric dynamics associated with a weather event of interest, however, it is often important to consider the full 3-D structure of the atmosphere, including related PV features. For instance, Bithell et al. [16] put effort into analyzing and visualizing the 3-D structure of atmospheric PV. They showed a wide spectrum of 3-D PV anomalies (PVAs, see below) that are difficult to analyze (or even to track over time) in 2-D vertical or horizontal (isentropic) cross-sections alone. They further demonstrated that the evolution of wave-related features along the tropopause may cover an extensive vertical range that is difficult to describe in the framework of a 2-D analysis. Therefore, Bithell et al. [16] concluded that the 3-D perspective is a useful tool for the comprehensive study of the evolution of the dynamical tropopause and related weather systems.

We note that in 2-D analysis, the concepts of both PVSs and PV cutoffs have been widely established. They are clearly defined and widely used. However, a 3-D anomaly stretching over multiple isentropes can exhibit both types, PVS and PV cutoff, depending on the considered level. Therefore, this 2-D classification is challenging for evaluations, and especially does not hold for an analysis of 3-D features. In this study and for 3-D analysis, we hence name these features PV anomalies (PVAs). These anomalies can exhibit characteristics of both streamers and cutoffs when considered on a single isentropic level only. 3-D cutoffs, i.e. fully isolated areas of stratospheric air surrounded by tropospheric air, are only rarely a result of detachment from the main reservoir and mostly appear in the lower troposphere due to latent heat release [e.g., 12].

An identification of 3-D PV structures has previously been performed for subsets of anomalies only, although the importance of their vertical structure has been demonstrated in the literature [e.g., 141, 105]. For example, Škerlak et al. [169] investigated the 3-D structure of one subset of PVAs, namely tropopause folds and their link to extreme weather events. These folds are defined as structures with multiple tropopause crossings in the vertical direction. Portmann et al. [140, 141] focused on cutoffs that are identified independently on multiple isentropic levels and are subsequently stacked vertically using an overlap and proximity heuristic. Regarding a 3-D field, these can be thought of as stalactites attached to the stratospheric PV reservoir. Portmann et al. [141] found that cutoffs can be potentially complex 3-D features, which can intensify on higher levels while decaying on lower levels, resulting in a vertical (cross-isentropic) displacement of these features. Therefore, tracking PVSs on a chosen 2-D isentropic level is not always sufficient to analyze an event during its entire life cycle. As investigated by Bithell et al. [16], stalactites and tropopause folds are only a subset of possible anomaly types in a 3-D PV field. They classified the complex structures they encountered as *tubes*, *spirals*, *stalactites*, *folds*, among others, demonstrating complex scenarios that are not discernible within a 2-D perspective. Existing approaches used in 2-D identification lack the concepts of 3-D cohesiveness and information regarding extent, genesis and evolution of these structures. For instance, using the complete 3-D field offers the possibility to consider long time periods without the need to adjust a selected isentropic level to changes occurring between different seasons.

In summary, while many previous studies contributed to identifying different aspects of PV features on single isentropic levels, the objective identification of fully 3-D PV structures is still an open challenge and motivates our work. We expect that a 3-D identification method will yield the opportunity for novel analyses, including new aspects within climatological studies and comprehensive case studies of 3-D atmospheric dynamics. In this chapter, we hence present a novel algorithm for the identification of PV anomalies, which can be applied not only to 2-D, but also to 3-D fields. In the latter case, the algorithm operates on the complete 3-D fields instead of on individual isentropic levels.

To the best of our knowledge, this is the first approach aiming for a strategy that identifies the full spectrum of PVAs in 3-D. We use image processing techniques for the identification, more specifically numerical solutions that are based on morphological operators. These operators, which originate from analytical geometry, have been adapted to suit the requirements of a meteorological application. This adaptation mainly revolves around using physical units to ensure interpretability, while considering the properties of the used projection. Parameters of the identification are adjustable and the algorithm is usable on different resolutions. A low-dimensional and human-readable feature vector is computed for each identified PVA consisting of quantifiable measures, e.g., centroid, intensity or best-fit geometry. This memory-efficient feature representation can serve as a basis for statistical analyses or as potentially relevant predictors. The identified 3-D PVA features are also well suited for studying atmospheric dynamics by means of interactive 3-D visual analysis [IVA; cf. 149]. For example, reducing atmospheric processes of importance to concise visual depictions facilitates combination of multiple aspects of atmospheric dynamics in a comprehensive 3-D display well suited for rapid exploration of the considered numerical simulation data [e.g., 148, 92]. Here, we demonstrate such an analysis by incorporating the identified PVA features into the 3-D meteorological visualization framework “Met.3D” [150], which we use to shed light on the 3-D structure of PVAs encountered during an extreme precipitation event previously investigated by Van der Linden et al. [189].

This chapter is structured based on the structure of the publication as follows. Section 5.4 introduces the basic principle of the algorithm. In Sect. 5.5, a distance measure required for the method is introduced. The identification algorithm is described for 2-D data in Sect. 5.6, and compared to the method of Wernli and Sprenger [200] in Sect. 5.7. Section 5.8 introduces the generalized 3-D algorithm, which we apply in Sect. 5.9 for a case study, discussing the potential of 3-D PVA analysis. Results of applying this method in a master thesis are summarized in Sect. 5.10. Conclusions from this chapter are summarized in Sect. 5.11.

5.4 Algorithmic motivation

The identification algorithm is motivated by so-called morphological image processing techniques, which have been introduced in Sect. 2.2.2. In the following, we first consider the 2-D case to illustrate our approach. Afterwards

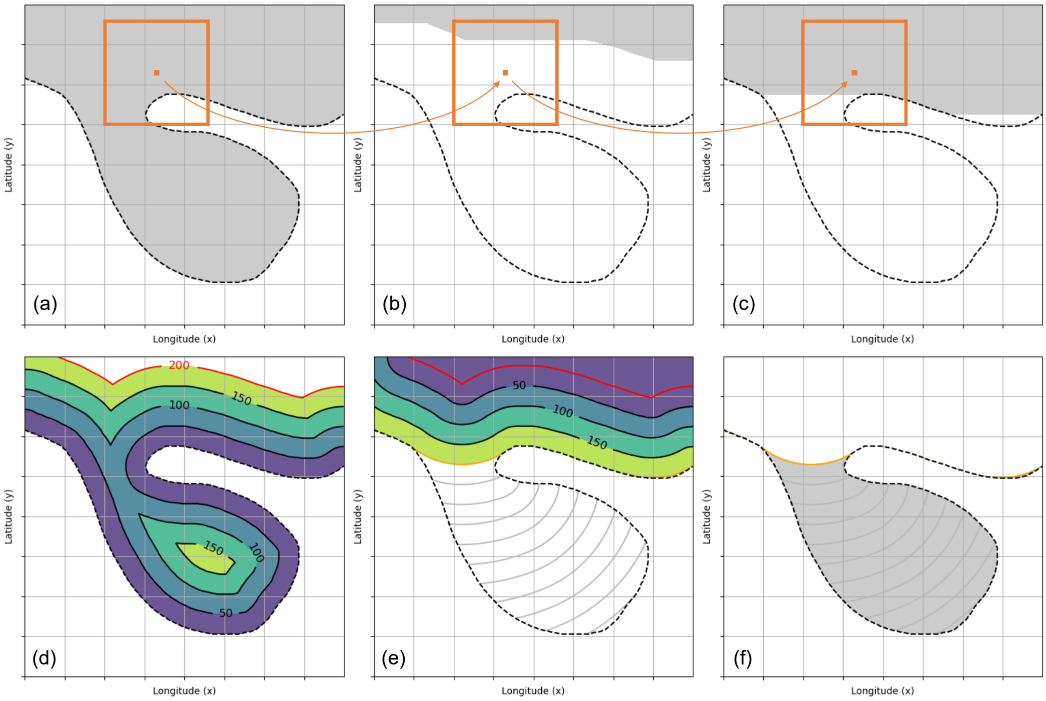


Fig. 5.1: A sketch of our identification strategy applied to a PVS-like structure. (a-c): The process of a morphological opening (erosion followed by dilation) is applied to the gray structure using a square mask (orange). This removes thin structures depending on the mask's size. (d-f): Our adapted strategy, which uses distances instead of a binary mask. First, the distance emerging from the outer boundary is computed. Then, starting at a set distance (red line), the same distance is added back, respecting the boundaries of the domain. Areas not reconstructed are identified as anomalies.

the method will be expanded to 3-D. To make use of morphological techniques, anomalies in the PV field (e.g., intrusions, cutoffs etc.) can be thought of as morphological "noise". A "noise-free" PV field then resembles an idealized state of the atmosphere, where the height of the tropopause is symmetrical around the poles.

Fig. 5.1a-c shows the process of a morphological opening applied to a binary image, which could resemble a PVS. For the erosion (Fig. 5.1a-b), a structuring element (orange) is used as mask. This mask is shifted over the binary input image. Each pixel (x, y) is kept as part of the domain only if all pixels in the structuring element surrounding (x, y) are also part of this domain. Thus, the erosion removes all areas along edges depending on the size of the mask. As seen in the figure, it especially destroys elongated and filament-like structures.

Following the erosion, the dilation step (Fig. 5.1b-c) reconstructs areas of the domain. Using the same structuring element, a pixel (x, y) is added back to the domain if any of the surrounding pixels defined by the structuring element is part of the domain. As seen in Fig. 5.1c, this reconstructs the general shape of the original object, but filament-like structures and other forms of noise are not reconstructed. These non-reconstructed areas can be extracted and interpreted as PVSs.

This approach uses an abstract concept of binary masks in discrete environments, and care is required to maintain interpretability and to design the algorithm with meaningful parameters to apply it in meteorological applications (see Sect. 2.2.1). Fig. 5.1d-e shows an adaptation using Euclidean distances instead of a mask. Masks are typically square, and results based on these masks would give a distorted distance measure. The use of Euclidean distances in this Cartesian field results in a uniform effect along all directions. The adapted strategy works as followed: First, we measure distances emerging from the boundary (e.g., tropopause) into the domain, as seen in Fig. 5.1d. Then, all points with a specific distance from the boundary based on a given parameter are extracted (e.g., the red line in Fig. 5.1d with a distance of 200). A comparable parameter for the upper part of the figure would be the size of the structuring element. Then, starting from the newly defined boundary (red line), distances are measured outwards. Points with the previously specified distance from the red boundary form a new boundary (orange line in Fig. 5.1e-f). Areas with a higher distance are identified as anomalies.

However, the distance measure required must follow the domain instead of using a direct spherical distance, as seen in Fig. 5.2. Most algorithms that satisfy this requirement suffer from metrical errors induced by a discrete representation of the projected grid. Instead, Añel et al. [6] achieve significant improvements in area computations using a higher-order numerical scheme based on a region-of-interest approach. On the other hand, the present study requires an approach to compute distances (as shown in Fig. 5.2b), but similarly must compensate for distortions in the field. For more precise results we choose a strategy based on a higher-order numerical scheme also, as outlined in the next section.

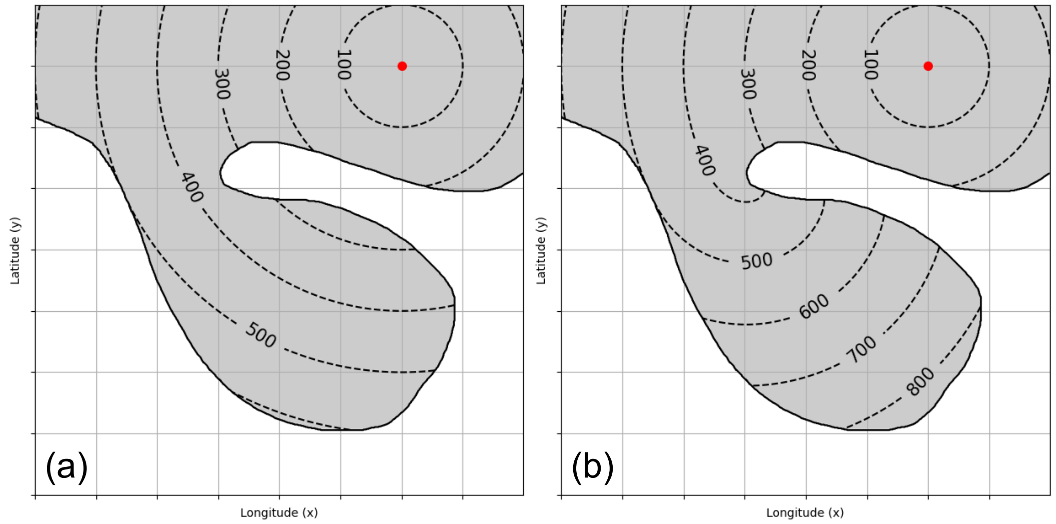


Fig. 5.2: A visual example (same PVS-like structure as in Fig. 5.1) showing the shortest distance measure required for the algorithm. (a): Starting at the red dot, the direct spherical distance is shown, which can be calculated easily. (b): The distance following a well defined domain (e.g., PV field) is computed using a Fast Marching Method (FMM) as required by our strategy, see Sect. 5.5.2 for details.

5.5 Solving the distance measure problem

5.5.1 Stereographic projection

For the identification algorithm described in this study we choose a pole-centered stereographic projection (Fig. 5.3). Its point of projection is on the surface of the sphere opposite to the tangent plane where the projection is created. Thus, we can use the South Pole as point of projection to have the view centered around the North Pole. This creates a singularity at the South Pole while minimizing the distortions in the northern hemisphere [171]. A mathematical formulation for the transformation T from spherical coordinates (θ, λ) to the projection plane (x, y) has been introduced in Sect. 2.1.1.

We choose this type of projection due to several advantages:

1. The PV field encircles the earth and is centered around the poles. Using the pole as center of projection leads to a clear depiction of the features to be detected.
2. Contrary to equirectangular projections, this stereographic projection avoids singularities at the North Pole. PV intrusions over poles can occur,

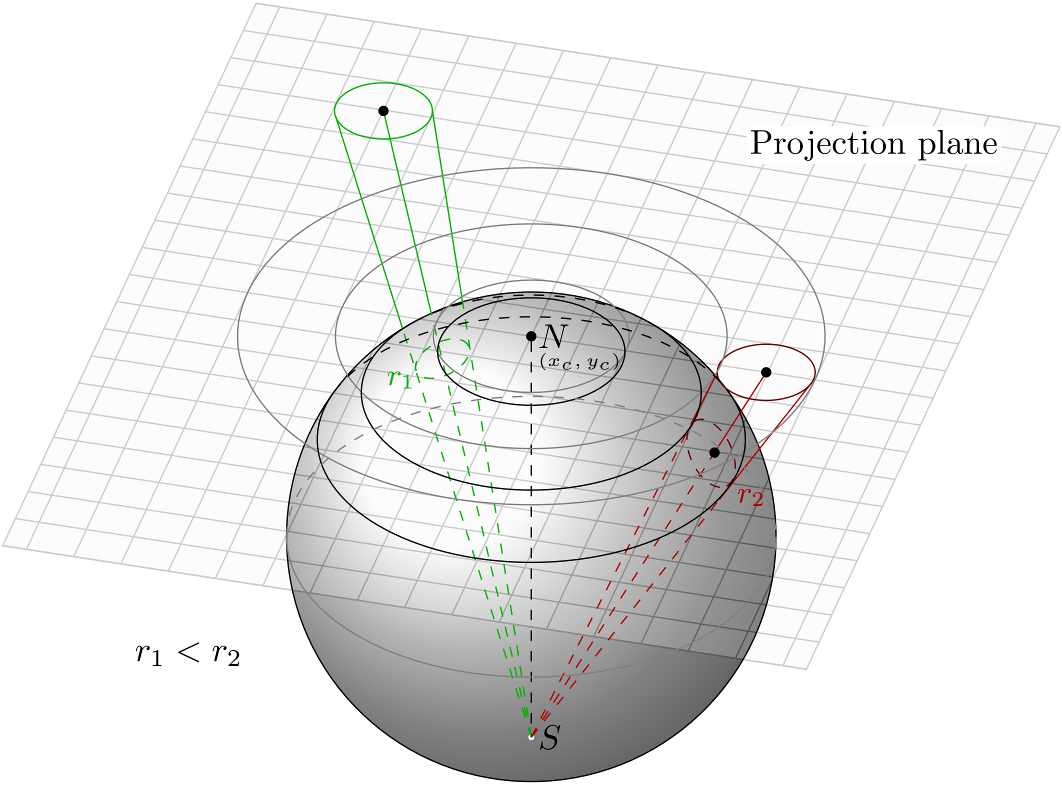


Fig. 5.3: Depiction of the stereographic projection used in our identification strategy. This figure shows that circles on the sphere are mapped to circles on the projection plane. Therefore, this projection is conformal, a vital property for this strategy. Note that despite having the same shape, the sizes of the circles on the sphere vary by distance from the projection center.

and their analysis is highly difficult in projections where a singularity is involved. Additionally, the stereographic projection solves issues handling the antimeridian boundary.

3. A stereographic projection is conformal, and the only known true perspective projection with this property [171]. A conformal projection preserves angles, thus, two lines on a sphere will intersect in the same angle as in their projection. As a result, infinitesimal shapes on a sphere are conserved in this projection. All Tissot's indicatrices are circles, therefore perfect circles on the sphere will always map to perfect circles in the projection, generally of a different area (see Fig. 5.3). This property will be useful to efficiently compute distances along the earth's surface regarding this projection later on.

The main drawback of this strategy is the introduction of a singularity at the projection center, i.e. the opposing pole. To avoid this problem while mapping global data, we create one stereographic projection centered around the North Pole and one centered around the South Pole, both extending towards the equator. The singularity at the pole shifts to a boundary at the equator. Since the definition of the dynamical tropopause changes its sign at the equator and generally lacks proper definition at lower latitudes, this area is not considered as a region of interest for this present study.

5.5.2 Computing distances within the projection

As introduced in the previous section, our novel algorithm requires a way to calculate distances within the projection following a given field, as illustrated in Fig. 5.2. Here, we can use the conformal property of the projection. Since circles on the surface of a sphere are projected to circles in the stereographic projection, we can define a distance map $\Delta s(x, y)$ for all points in the projected field, or for all pixels (x, y) respectively. Each pixel in the map Δs contains the radius of a circle on the sphere in km that projects onto a circle of radius of 1 pixel around (x, y) in the projection. For example in Fig. 5.3, the centers of the green and red unit circle on the projection plane are assigned values $\Delta s(x, y) = r_i$, where r_i is the radius of the corresponding circle on the sphere. Thus, based on the conformal properties, $\Delta s(x, y)$ can be expressed as a scalar map and be calculated using just the distortion factor at every point. For a sphere, according to Snyder [171], the distortion factor h at a point (x, y) in a stereographic projection can be calculated by

$$h = \cos(\frac{\theta(x, y)}{2})^{-2}, \tag{5.1}$$

where $\theta(x, y)$ denotes the angle from the center of the projection at (x, y). h denotes the scale factor in all directions. For example, projecting the northern hemisphere with the North Pole as projection center yields a factor of 1 at the pole ($\theta = 0$) as expected, and a factor of 2 at the equator ($\theta = \frac{\pi}{2}$). Towards the South Pole, h becomes infinite. The distance map $\Delta s(x, y)$ can simply be calculated by the pixel radius in km at the projection center (x_c, y_c) (e.g., the pole, see Fig. 5.3), divided by the distortion factor:

$$\Delta s(x, y) = \frac{\Delta s(x_c, y_c)}{h}. \tag{5.2}$$

$\Delta s(x_c, y_c)$ denotes the radius on the sphere that projects onto a circle on the projection plane centered around (x_c, y_c) with a radius of one pixel. Since the pole is the projection center, the pole of the sphere touches the projection plane in (x_c, y_c). By mapping one neighboring pixel of the pole to their regarding latitude and longitude coordinates, $\Delta s(x_c, y_c)$ can be obtained by computing the spherical distance between the pole and this neighboring pixel. Due to the conformal property of the projection, any of the neighboring pixels yield the same distance.

As next step, we define the distance map $u(z)$ containing the minimal distance from a well defined boundary Γ to all $z = (x, y)$, where the distance $u(\Gamma)$ from the boundary to the boundary is zero. This problem is similar to the computation of signed distance functions or solving path planning problems. Popular algorithms to generate this map u starting at Γ contain variations of the Dijkstra algorithm [34]. Starting at the boundary, this construction is performed by evolving a level set to the destination based on a distance map [136].

However, most of these approaches suffer from metrication errors. They construct the path on segments parallel to the grid dimensions. Directions invariant to the grid orientation will always lead to errors induced by the L_1 norm [25]. To get consistent results for the identification independent of the location of a PVS, these inaccuracies should be avoided.

This shortest path problem can also be formulated as a line integral. In this context, we are interested in finding the shortest distance u_γ from a point γ

in the boundary Γ along a smooth and differentiable curve $\mathcal{C}$ to a point p in the field, considering a given cost function Δs. The integral form of this can be expressed as:

$$u_\gamma(p) = \min_{\mathcal{C}} \int_{\mathcal{C}} \Delta s(z) dz. \tag{5.3}$$

Here, the integral $\int_{\mathcal{C}} \Delta s(z)dz$ represents the cumulative cost along the path $\mathcal{C}$, where Δs is the cost function evaluated at any point $z = (x, y)$. This cost function encapsulates the "weight" or "distortion" at each point int the 2-D field.

To better understand this, imagine a landscape, where the cost function Δs could represent the difficulty of walking through different terrains. More difficult terrains equal to higher cost. The goal is to find the path from point γ to point p that accumulates the least total cost, effectively the easiest path to traverse, considering all possible paths.

In our specific application, the distance map Δs is used to represent distortions due to projections. For uniform Δs, i.e., a field without distortions, the formulation simplifies to the Euclidean distance. However, when distortions are present, this path integral formulation ensures that we account for these variations in the projection.

Then, by a well defined boundary Γ, we can express the shortest distance from the boundary Γ to any point p in the field as:

$$u(p) = \min_{\gamma \in \Gamma} u_\gamma(p). \tag{5.4}$$

Cohen and Kimmel [25] explored that this path formulation satisfies the Eikonal equation

$$||\nabla u(p)|| = \Delta s(p). \tag{5.5}$$

This first-order nonlinear differential equation is used widely in scientific applications, mainly in wave propagation problems as a front propagation approach. In these cases, the cost function on the right-hand side is typically expressed as a propagation velocity at every point in the domain. To our knowledge, we are the first to formulate this equation in a sense to compensate for distorted distances in a projection instead, leading to a formulation for distances from a given boundary for conformal projections. The map of shortest distances u along the projection plane contains the distances from a given boundary to points in this map. Intuitively, the gradient of u can be thought of as a measure that is anti-proportional to the distortions defined by the distance map Δs.

Solving the above differential equation yields a far more accurate result than graph based approaches on a discrete field, resolving metrication errors. There are approaches solving the Eikonal equation, mainly the *Fast Marching Method* (FMM) introduced by Sethian [167]. Starting at known distances in the distance map, i.e. at the boundary where the distance is zero, the algorithm takes advantage of the fact that this map can iteratively be built from the boundary outward since the cost function Δs is strictly positive. We refer the interested reader to Petres et al. [136] for a more in-depth description and evaluation about this algorithm.

5.6 Identification technique in 2-D

Based on the above introduced strategies, we outline our identification technique for 2-D PVSs on a given isentropic level. Figure 5.4 illustrates the individual steps; pseudo-code of the algorithm is shown in Algorithm 2.

Algorithm 2 Pseudocode of the identification process, both for 2-D and 3-D application. 2-D and 3-D identification differ with respect to distance measure (scale analysis for 3-D) and filtering strategy (additional area based filtering for 3-D); cf. Sect. 5.8.

```
Input: $d_{PV}$: PV field, $w$: width threshold
$d_{PV}^{proj} \leftarrow$ PROJECT($d_{PV}$)
        ▷ Project data to stereographic grid (Fig. 5.4b)
$\Omega \leftarrow$ GETDOMAIN($d_{PV}^{proj} > 2$ PVU)
        ▷ Set stratospheric air mass as domain (Fig. 5.4c)
$\Gamma_1 \leftarrow$ GETBOUNDARY($d_{PV}^{proj} = 2$ PVU)
        ▷ Determine dynamical tropopause (Fig. 5.4c)
$d_{inner} \leftarrow$ DISTANCES($\Gamma_1, \Omega, \Delta s$)
        ▷ Compute distances from boundary into domain (Fig. 5.4d)
$\Gamma_2 \leftarrow$ GETBOUNDARY($d_{inner} = \frac{w}{2}$)
        ▷ Extract boundary where distance equals half the width threshold $\frac{w}{2}$ (Fig. 5.4d, e)
$d_{outer} \leftarrow$ DISTANCES($\Gamma_2, \Omega, \Delta s$)
        ▷ Compute distances from $\Gamma_2$ out of domain (Fig. 5.4f)
objects $\leftarrow$ LABEL($d_{outer} > \frac{w}{2}$)
        ▷ Extract and label areas with distance greater than $\frac{w}{2}$ (Fig. 5.4g)
return FILTERED(objects)
        ▷ Filter anomalies by length or other heuristics (Fig. 5.4h)
```

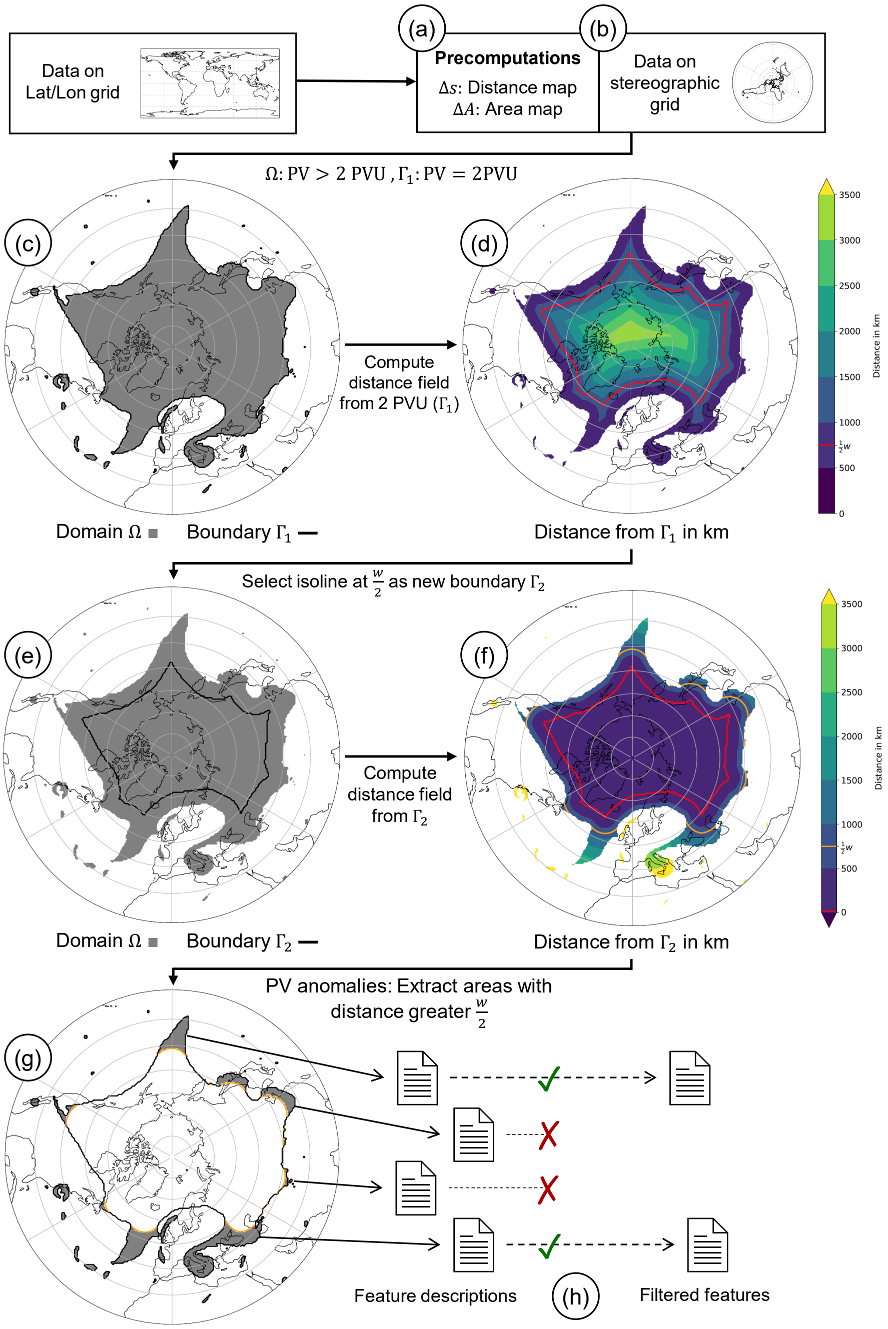


Fig. 5.4: Illustration of the algorithm to identify PVSs. (a, b): Input data is mapped to a stereographic grid; precomputations are performed. (c, d): The distance map from the 2-PVU boundary is computed (cf. Sect. 5.5). Points at a distance equal to half the width threshold $\frac{w}{2}$ are denoted by the red isoline. (e, f): From this isoline, distances outwards are computed. Again, the isoline corresponding to $\frac{w}{2}$ is drawn with respect to the new distance map (orange). (g): Regions with a distance greater than $\frac{w}{2}$ from (f) are extracted. They form anomalies in the field. (h): The anomalies are labeled and filtered, yielding a set of identified PVSs. Visualized is the 335 K isentrope from the ERA5 reanalysis at 7 September 2016, 00 UTC.

5.6.1 Data

For the 2-D analysis, we use isentropic levels from both the ERA-5 reanalysis [75] data interpolated to 0.5° x 0.5° in both latitude and longitude, and data from the sub-seasonal to seasonal (S2S) prediction database [190], as indicated in the corresponding figures. The latter contains forecasts out to 45 days and is available daily on a regular latitude-longitude grid with a grid spacing of 1.5° x 1.5°. The data is remapped to a stereographic grid.

5.6.2 Computation of the distance map

Before executing the identification strategy, the distance map Δs (as introduced in Sect. 5.5) has to be computed (see Fig. 5.4a). This map is of importance for the core functionality of the identification, as well as to produce metrics for each PVS. Since the values of the distance map do not depend on the input data itself, this can be done as a precomputation step, and cached for future use.

In this study, we use a polar centered stereographic grid with a projection pixel spacing of 75 km and a total size of 340×340 pixels. This configuration covers one hemisphere completely. A finer spacing (e.g., 50 km with a size of 510×510 pixels) increases memory consumption and run time significantly. The resolution of the projection should be chosen to match the resolution of the input data, i.e. the pixel footprints should cover similar areas. For our ERA5 analyses, the mentioned resolution is in that regard sufficient.

Furthermore, as additional metric, the *area map* ΔA is computed in this step as well. This map contains for each pixel in the projection the spherical area it projects. It can be approximated by the square of the distance map $\Delta A = (\Delta s)^2$. The area map can be used after the identification to calculate the centroid of a given PVS more accurately by taking area distortions into account.

5.6.3 Identification algorithm

Inputs for the algorithm are the field of PV data, and the width threshold w. It will become clear that this parameter leads to an identification of PVS up to a maximum width of w km. In our study, we use a value of w =1500 km, a width threshold used widely in existing techniques like the one by Wernli and Sprenger [200]. More input parameters are not required for the 2-D identification. In a first step, the stratospheric body of air is extracted from the projected data set using the previously introduced threshold of 2 PVU. This results in a binary field as seen in Fig. 5.4c.

The next step is called *erosion*, which is based on the respective morphological operation and encapsulates the main idea of the strategy, as introduced in Chapter 2 and Sect. 5.4. The domain Ω is defined as the stratospheric air mass, and the boundary Γ_1 as the dynamical tropopause, as shown in Fig. 5.4c. Starting from the boundary, the distance from this boundary to every point in the stratospheric air mass is computed by solving Eq. 5.5 using the FMM as outlined in Sect. 5.5. Figure 5.4d illustrates this step and shows the color coded distance map. These are the minimum distances from the boundary Γ_1 to each grid point, respecting the boundaries of the PV field.

After that, areas that are $\geq \frac{w}{2}$ km away from the boundary are extracted, these areas form the “inner core”. Higher values for the width threshold w would remove more areas along the tropopause and thus smoothing the inner core further. Generally, this inner core may contain multiple disjunct areas. In these cases we pick the biggest one as the core to proceed with, defined by its area. The outer contour of the previously defined inner core (red/black contour in Fig. 5.4d/e) is defined as boundary Γ_2. This boundary is the starting point for the next operation.

The *dilation* operation resembles in morphology the second basic operator, as introduced in Sect. 5.4. The contour of the previously defined inner core is used as input boundary, and distances outward with respect to the domain Ω are computed. Keeping this domain is necessary to measure distances following the stratospheric domain with respect to the distance map (as illustrated in Fig. 5.2). The distances can be seen in Fig. 5.4f. While filament-like structures are assigned meaningful distances, note that cutoffs cannot be reached by the algorithm in this step because they are not spatially connected to the stratospheric reservoir. Their distance is set to infinity, making them easily recognizable later. Feature descriptions for these objects can be computed, nonetheless.

These extracted areas resemble the identified PVS (shaded areas in Fig. 5.4g). Since $\frac{w}{2}$ km have been eliminated along the contour (from all sides) during the erosion step, anomalies with a maximum width of w km are identified.

Figure 5.5 shows outputs of the algorithm for different values of w. Higher values for this parameter lead to the identification of wider anomalies, as well to further extension of anomalies towards the main reservoir. The identification of cutoffs is not sensitive to parameter changes.

5.6.4 Feature Vectors

The identified PV anomalies are labeled by cohesiveness and features as defined in Chapter 2. Feature vectors are computed for each PV streamer, containing various characteristics. We compute the moments up to order two for each feature, which includes its centroid and orientation, as introduced in Sect. 2.4.1. Furthermore, the average and maximum PV of a feature, and its length are computed. The length measure is used to filter the features: We only keep PVSs that satisfy a length threshold of 1000 km. This threshold is motivated by previous work from Wernli and Sprenger [200], who use a threshold of 2000 km regarding the perimeter of the feature defined by the 2-PVU contour. Given that the perimeter is approximately twice the length for thin and elongated structures, this threshold provides a practical approximation.

5.7 Evaluation and comparison in 2-D

Before extending the idea to 3-D data sets, the introduced algorithm will be evaluated. For this purpose, the identification strategy by Wernli and Sprenger [200] and Sprenger et al. [174, 173] (hereafter abbreviated by WS07) will serve as a benchmark. Their strategy is well established, used in various follow-up work, and considered as state-of-the-art.

The PVS identification strategy by WS07 is motivated by the thin and elongated nature of PVSs. They identify the outermost 2-PVU contour, and connect points along it with a spacing of about 30-50 km. This step returns a polygon resembling the outermost dynamical tropopause for the selected isentropic level. Then, the vertices of this polygon are pairwise compared regarding their direct spherical distance d and length along the contour l. For small d and big l ($d <$1500 km and $l >$2000 km according to Sprenger et al. [174]), these points are regarded as streamer base points. Then, these base points

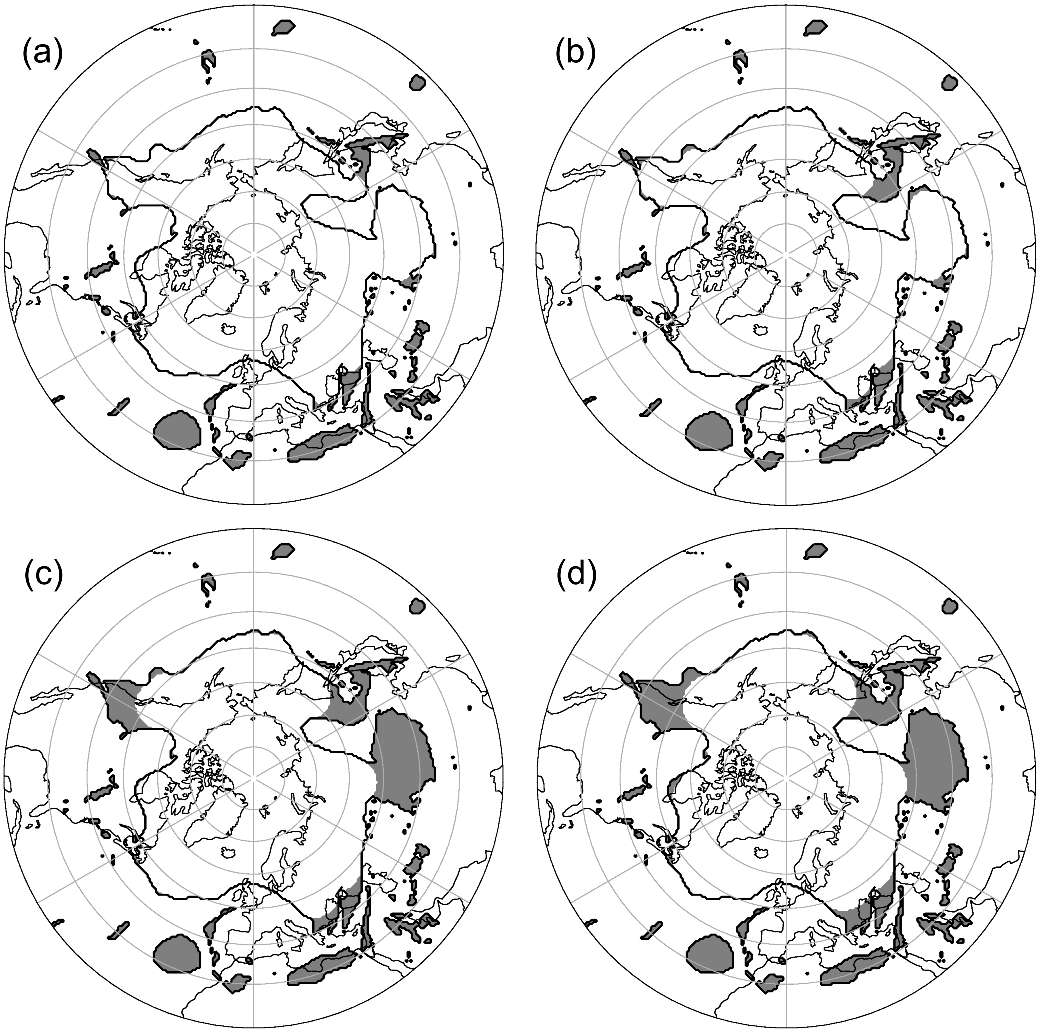


Fig. 5.5: Sensitivity of the identification process to the width threshold w. Here, w equals (a) 500 km; (b) 1000 km; (c) 1500 km; (d) 2000 km; using the example of the 330 K isentrope from the ERA5 reanalysis at 14 July 2015, 00 UTC. Results are not filtered to highlight the effect of the parameter. Our default configuration is (c).

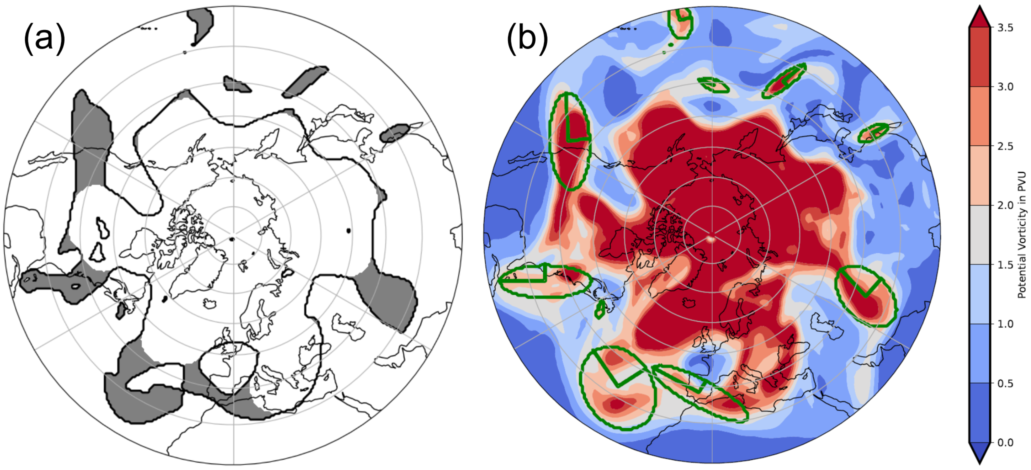


Fig. 5.6: 200–500 hPa layer-averaged PV based on the S2S prediction database [190] initialized and valid at 7 May 1998, 00 UTC. (a) PVSs identified using the default configuration without filtering the results; in (b) filtered PVSs using a length threshold of 1000 km, approximated by best-fit ellipses, with raw PV values shaded in the background. Ellipses are defined by their computed main axes.

are connected in a straight line and, together with the 2-PVU contour, this forms a closed polygon which resembles the PV streamer.

Comparing the algorithm parameters, our width threshold w has a similar interpretation to the d in the work by WS07, so we set $w = d$ for evaluation purposes. This choice is also confirmed by climatological analyses we conducted, showing an inordinate increase in detected structures for smaller values of w. While their strategy has the additional l parameter (length along contour), we do not set length restrictions in the first step, but filter the identified PVSs later based on a length threshold. This filtering parameter can be refined and changed depending on the use case.

Figure 5.7 shows the identification results in a 2-D PV field for both techniques. This example represents one of the predominant instances where we expect similar results. As shown in the figure, both techniques indeed identify the same structures as streamers. These also resemble the structures an educated user would identify as such. However, the shapes slightly differ: While the method by WS07 (Fig. 5.7b) identifies streamer base points and connects them in a straight line, our approach naturally creates roundings at these spots. These roundings reveal the boundary on what is considered as an anomaly in the field by our algorithm. The amount of detected anomalies is similar to

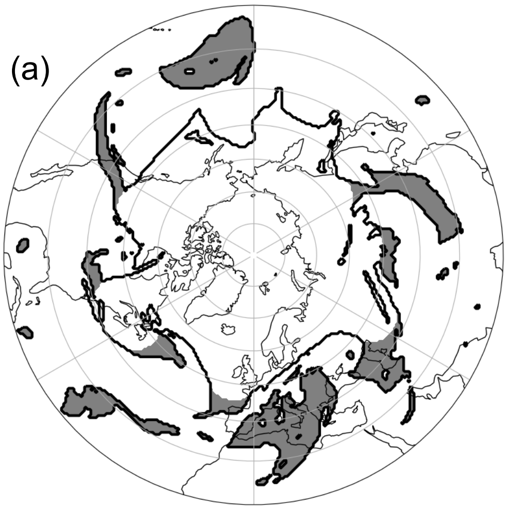


Fig. 5.7: Comparison of the identified PVSs based on the 340 K isentrope from the ERA5 data set at 24 July 2015, 00 UTC. (a) Results of our strategy using the default configuration; and (b) by the algorithm from Wernli and Sprenger [200]. The anomaly over East Asia will be investigated in the case study later on.

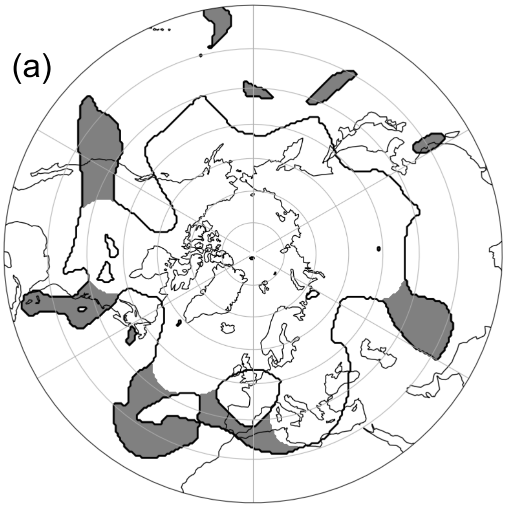


Fig. 5.8: Same as Fig. 5.7, but with the data set used in Fig. 5.6. The red arrow indicates the outermost 2-PVU contour, which is used to identify anomalies in the work by Wernli and Sprenger [200].

the climatological results presented in their study, as supported by analyses we conducted.

Figure 5.8 on the other hand shows a more complex example demonstrating the improvements gained by considering the full PV field compared to only the outermost 2-PVU contour. Over Europe and the Eastern Atlantic, two PV intrusions (one cyclonic and one anti-cyclonic) are being observed at this specific time step.

These structures are not identified by the approach from WS07. By only tracing the outermost 2-PVU boundary (red arrow in Fig. 5.8b), the inner state of the PV domain is not considered. However, this inner state might reveal important information on the large-scale flow in this area. By just following this red arrow, this results in a broad PV trough being spotted from the Central Atlantic all the way to the Black Sea. This wide trough does not fulfill the required thresholds for a streamer. The strategy we employ also considers the inner state of the reservoir, including all troposphere-stratosphere boundaries in the domain. The erosion step starts at these inner boundaries as well. Therefore, these elongated structures are not reconstructed in the following dilation step, identifying them thereafter. In this specific case, the two anomalies are spatially connected. Our strategy therefore considers them as one merged object. Further heuristics are required to distinguish the origin of these merged anomalies in order to separate them, but this issue is beyond the scope of this work.

As introduced, we filter PVSs that do not satisfy a length threshold of 1000 km, which is comparable with the 2000 km contour length used in WS07. Note that our identification strategy does not have to be re-executed when changing the length threshold, instead the filtering step returns a different subset of identified streamers. Furthermore, our length calculation yields a precise measure for the streamer, from the reservoir base to its tip. Contour length techniques can suffer from inconsistencies introduced by noise. All in all, our strategy generally leads to additional identified streamers, especially in complex environments.

For a set test case, the computation of anomalies takes a similar amount of time. Regarding our strategy, this time also includes the computation of feature vectors (e.g., best fit geometry and centroids) for all detected structures. Nevertheless, due to different approaches in preprocessing, different programming languages and different use of parallelism, a quantifiable comparison of the run times proves to be difficult. For our algorithm, 85% of the computational time is required to project the data onto a stereographic grid. Analyzing bigger data sets, data level parallelism (e.g., Single Instruction Multiple Data (SIMD) processing) can highly improve the efficiency of projecting data from multiple isentropic levels and multiple time steps. Projecting a 3-D data set with 50 isentropic levels takes only twice the time compared to a single isentropic level.

Similar to WS07, we currently do not consider the origin of a PVA as a filtering strategy. For example high PV of non-stratospheric origin (e.g., in tropical cyclones [119] or related to deep moist convection [198]) could also be identified as a PVS or cutoff. More sophisticated time series and algorithms are needed to distinguish the origin of a PV filament, e.g., using specific humidity thresholds as suggested by Škerlak et al. [169]. Our concept of feature descriptions can give some insight and provide quantified values for each identified anomaly, but this problem goes beyond the scope of this work.

5.8 Extending the principle to three dimensions

5.8.1 Data

For the 3-D identification, ERA-5 data on model levels (here we use levels 40-137) are interpolated to isentropes with a vertical spacing of 2 K. We consider the region between 290 K and 380 K. The vertical spacing of 2 K proved to be a reasonable trade-off between computation time and disk usage on one hand, and representing the vertical structure of the PV field on the other hand. Our experiments showed that coarser spacings (e.g., 5 K) hide dynamically relevant and fine features in the vertical structure, while a finer spacing would increase memory and computational cost significantly without adding much more relevant structures. Using pressure or model levels would be an alternative worth considerating, since it is natively available in higher vertical resolution as model output. Therefore, it would not require the interpolation on isentropic levels, but would reduce the interpretability of the PV structures, especially when evaluating them using cross-sections or development over time.

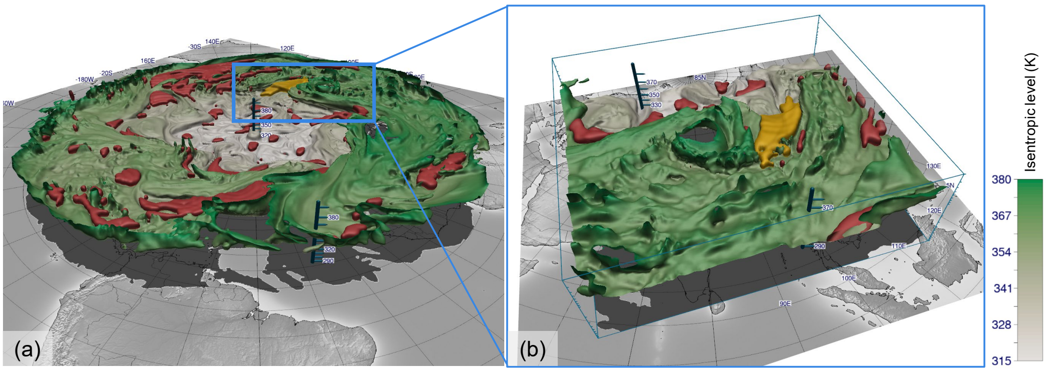


Fig. 5.9: 3-D visualizations of the dynamical tropopause (defined by the 2-PVU threshold), which is shaded by isentropic height. Red and yellow are the identified anomalies. The yellow one is emphasized for the case study later on. (a) The entire northern hemisphere is visualized; in (b) focused view of southeastern Asia, which is investigated in the case study in Sect. 5.9. Displayed on both panels is the ERA5 reanalysis at 24 July 2015, 00 UTC.

5.8.2 Strategy

There are several possibilities to extend the algorithm introduced in Sect. 5.6 to 3-D. A straightforward idea consists of applying the 2-D identification to each isentropic layer individually and stacking the results on top of each other to create 3-D objects (e.g., Portmann et al. [141] for cutoffs). However, applying our algorithm in such fashion would not include all information available in the full 3-D environment at a given point. For example, the use of thresholds (e.g., minimum contour length or aspect-ratio) in a 2-D identification strategy can lead to artifacts when applied to multiple vertical levels: A PVS might be above a given identification threshold on two specific isentropic levels, but below this threshold on a level in between these two, therefore not identifying this structure on this intermediate level. This leads to unwanted vertical gaps in the detected structure.

Figure 5.9a shows a 3-D visualization using the open-source visualization framework Met.3D [150] of the tropopause (defined by the 2-PVU isosurface) and the identified anomalies. Like for the previous 2-D visualizations, we use a stereographic map projection to emphasize the circulating nature of the flow. The tropopause is shaded by isentropic height. The air mass below (above) this isosurface is classified as tropospheric (stratospheric) air. As expected, the dynamical tropopause has a lower height towards the North Pole (beige shaded area), and a sharp gradient is situated in the mid-latitudes where the jet stream is located [101]. Visible holes in the isosurface resemble tropopause levels beyond our visualization domain of 290 K to 380 K. The red and yellow regions (latter emphasized for the case study) are the identified PVAs. As visible in Fig. 5.9 and supported by the video supplement, identified anomalies can manifest in various shapes. Anomalies are typically located in valleys, folds, or generally along depressions of the tropopause boundary. Because the algorithm works in a broader sense to identify irregularities along the boundary, it does not distinguish between the type of anomaly. Especially for our case study, we use the East Asian region shown in Fig. 5.9b to analyze the event.

Our algorithm is designed in a manner that it can be executed on 2-D and 3-D data in a similar way. Note that the basic morphological operations (erosion and dilation, see Sect. 5.4) can be applied independently of the number of dimensions. This requires the computation of distances in 3-D space. Therefore, we need to revise the distance measure due to mismatching units in the horizontal domain (km) and vertical domain (K). An exact distance approach would compute the height of every grid point in the atmosphere given its current state or based on a standard atmosphere model. Due to too high computational effort, we keep isentropic levels as vertical distance measure, and combine the mismatching units in the horizontal and vertical dimension motivated by a scale analysis. One challenge in computing meaningful distance is posed by the different horizontal and vertical scales, and in our case also different units, of the PVAs in the atmosphere. Typically we find the horizontal scale of PVAs to be on the order of 1000 km, and their vertical scale on the order of 10 K. To achieve meaningful 3-D visualization, Bithell et al. [16] used a vertical scaling factor of about 100 along the z-axis to achieve distinctly visible 3-D structures along the tropopause. To achieve similar order of magnitude in all three spatial dimensions, we follow Bithell et al. [16] and scale the vertical dimension by a factor of $100\,\mathrm{km}\,\mathrm{K}^{-1}$, then use a Euclidean distance measure. With respect to the width threshold w, the same value as in 2-D is used.

Figure 5.10 shows the main steps of the 3-D algorithm. The 3-D erosion step is illustrated in Fig. 5.10a, analogous to the 2-D case in Fig. 5.4d. As in 2-D, the domain Ω is the stratospheric air mass, and the boundary Γ_1 is the dynamical tropopause defined by the 2-PVU threshold. Thus, adding the vertical dimension, the domain Ω becomes the three-dimensional stratospheric air mass, and the boundary Γ_1 becomes the 2 PVU isosurface, separating stratospheric and tropospheric air. Distances are measured emerging from the boundary Γ_1 into the stratosphere using the same algorithm as in 2-D but applied along all

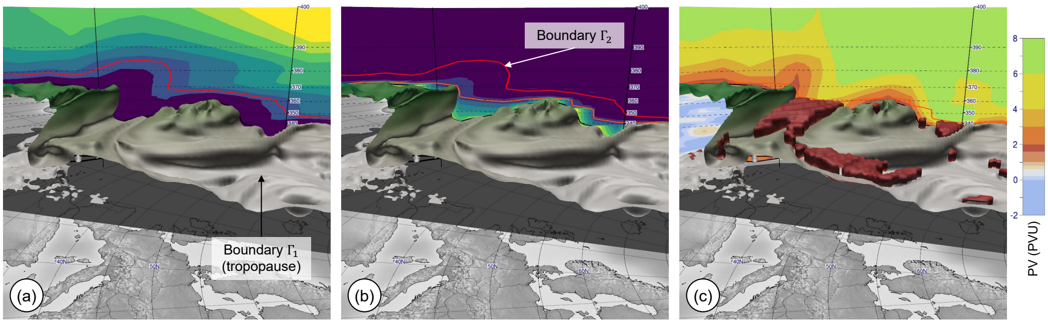


Fig. 5.10: The identification method introduced in Fig. 5.4 for 2-D extended to 3-D. The tropopause over the North Atlantic (2 PVU isosurface) is displayed along with a vertical cross section showing: (a) distances from the dynamical tropopause (boundary Γ_1) towards the stratosphere (see Fig. 5.4d for the 2-D equivalent and the color map); (b) distances measured outwards from the red boundary Γ_2 into the stratospheric domain (see Fig. 5.4f for the 2-D equivalent and the color map); (c) the potential vorticity. Additionally shown are the identified anomalies on top of the tropopause (red objects). Note that all distances are measured in 3-D, while the visualizations only show a 2-D cross section. Displayed on all panels is the ERA5 reanalysis at 24 July 2015, 00 UTC.

dimensions. This results in a 3-D distance map (same color mapping as in 2-D, c.f. Fig. 5.4). Figure 5.10a shows the distance along a vertical cross section. As in 2-D, we define the new boundary Γ_2 as all points that have a distance of $\frac{w}{2}$ km from the boundary (red contour in Fig. 5.10a, note that in 3-D space this becomes a 3-D isosurface). The isosurface Γ_2 resembles the inner core and lies above the dynamical tropopause, bounded by a smoother boundary than the tropopause itself. Distances outwards are computed starting at Γ_2 (Fig. 5.10b, cf. Fig. 5.4f for 2-D), using the same distance measure. Distances here are constrained by the domain Ω and grow into the anomalies. Then, as in 2-D, another boundary is defined at a distance of $\frac{w}{2}$ km (orange contour in Fig. 5.10b, cf. Fig. 5.4f). Areas farther away than $\frac{w}{2}$ km from Γ_2 are extracted. These are the areas identified as anomalies. As shown in Fig. 5.10c, these anomalies are volumetric objects. Stratospheric air masses isolated from the stratospheric reservoir (3-D cutoffs) are identified similar to the 2-D process. These areas cannot be reached by the distance computation starting at Γ_2.

The same filtering process as in 2-D is applied: The goal is primarily to keep as many detected features as possible. The user can then decide which of these features are of importance for a specific use case and filter these by their feature descriptions accordingly. The centroid of an anomaly is computed in 3-D, while we replace the area measure with a volume measure (in km^2K, or approximated in km^3 using a standard atmosphere model). Using the centroid

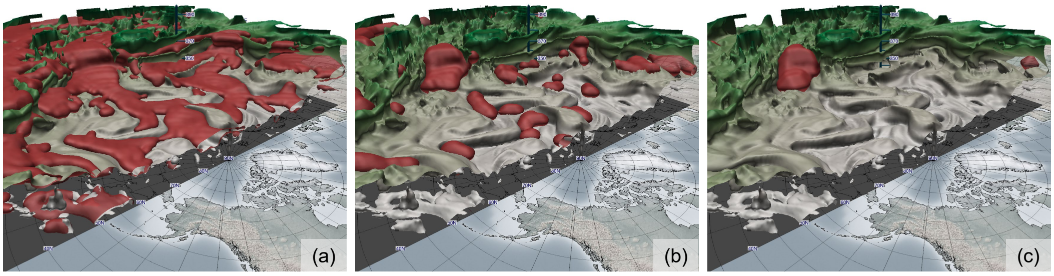


Fig. 5.11: Effect of filtering anomaly candidates with respect to the extent threshold b. Areas of anomalies with a vertical extent smaller than b are omitted from the identification, this leads to clear and separated features. (a) $b = 0\,\mathrm{K}$ (no filtering); (b) $b = 6\,\mathrm{K}$; (c) $b = 12\,\mathrm{K}$. As specified in Sect. 5.8, our default configuration is 6 K.

in 3-D, the vertical evolution of an anomaly can be analyzed. The case study in Sect. 5.9 investigates an anomaly with a clear shift in vertical position over time. Some of the feature descriptions introduced in Sect. 5.6.4 are not computed in 3-D due to misleading interpretations introduced by the scale analysis. For example, the length measure implicitly contains the applied scale factor of $100\,\mathrm{km\,K^{-1}}$. As stated above, this scaling is required for the 3-D functionality. Instead, for analyses we suggest to use measures that are easy to interpret, like the object's bounding box or volume. Typical use cases for filtering involve setting thresholds for position, size, or values of other fields in the same area (e.g., humidity).

However, in 3-D, we have to introduce an area based filtering strategy (cf. Sect. 5.6.3). Fig. 5.11a shows the raw and unfiltered results of our identification applying the strategy described in this section. Despite the distinct nature of isentropes in 2-D, the 3-D display reveals that these are actually connected dynamically to each other in 3-D. This behavior is expected from a globally coherent flow structure. However, it is desired to have clear and distinct anomalies in order to describe these individually for statistical analyses. Therefore, we filter the identified anomaly areas to get a set of detached and separated objects. We discard identified areas with a vertical extension of smaller than b =6 K. This specific value has been chosen as a good compromise of grasping clear features, while separating them at areas of weak vertical extent. Illustrated in Fig. 5.11b-c are identification results using different values of the extent threshold b. Figure 5.11b shows our default configuration.

Existing identification techniques (see an overview of them in Papin et al. [134]) rely on following the 2-PVU boundary, and searching points along this boundary meeting certain criteria (e.g., Wernli and Sprenger [200]). However, in 3-D, there is no unique direction to follow the boundary (the dynamical tropopause), making such techniques not applicable. Our strategy does not rely on a spatial tracking of a contour, but works in a broader sense, where the introduced domain Ω and boundaries Γ_1, Γ_2 can be extended to a 3-D perspective.

5.8.3 Implementation details

Our feature identification strategy is implemented in the Python framework for general identification and tracking of meteorological structures, introduced in Chapter 4. The algorithm itself relies on the CDO Python wrapper [165] for projections. The *scikit-fmm* toolkit [56] provides an implementation for the FMM algorithm, which has been adapted to fit our requirements. The algorithm is part of version 1.0 of the framework and is available as stated in the code and data availability appendix.

On a test system (AMD Ryzen 5 2600X), the identification process for a 3-D data set with 46 isentropes (290 K to 380 K) and a horizontal grid spacing of 0.5° x 0.5° for the northern hemisphere takes 5.9 s, where projecting the data takes about 23% of the total time. The computation time heavily depends on what features are being computed for the individual objects (e.g., main axes, volume). Compared to the 2-D identification, where one isentropic level took about 1 s to process, this 3-D identification reveals a big speedup regarding the amount of data being processed.

5.9 Case Study: Precursors of an extreme precipitation event affecting northeastern Vietnam

We evaluate our identification method along a case study. The selected extreme precipitation event affected northeastern Vietnam in late July and early August 2015. This event has been analyzed in detail by Van der Linden et al. [189], particularly focusing on the predictability of the event in ensemble forecasts. According to Van der Linden et al. [189], predictability of the event was strongly related to the presence and location of an upper-level trough in ensemble forecasts. The trough was diagnosed using 200 hPa geopotential height and upper-tropospheric PV fields. Here, we want to take a closer look into the precursors of this trough, and to analyze whether our novel technique can help to improve the analysis of the dynamics leading to the event. Different configurations of the algorithm are considered and our intent is to shed light on the 3-D structures. The visualized data is based on the ERA5 reanalysis as introduced in Sect. 5.8.1, using a 6-hourly temporal resolution. To assess the rainfall associated with the event, the 6-hourly rainfall totals centered around the respective time steps and based on the satellite-based NASA Global Precipitation Measurement (GPM) Integrated Multisatellite Retrievals for GPM [86] product were computed. This data is available at a horizontal grid spacing of 0.1° x 0.1°.

Between 25 July and 3 August 2015, record-breaking precipitation amounts of more than 1000 mm were observed at the coast of northeastern Vietnam. According to the analysis by Van der Linden et al. [189], intense and long-lasting rainfall was mainly related to a surface low over the Gulf of Tonkin, which slowly moved westwards. Van der Linden et al. [189] provided evidence that the formation and westward movement of the surface low was related to a subtropical trough that was located over southern China slightly to the northwest of Vietnam. Figure 5.12a, b shows the 200 hPa isosurface in an isentropic field. The isosurface is shaded by geopotential height. It also shows the location of the identified trough axis by Van der Linden et al. [189], using a 2-D analysis. In Fig. 5.12b, the identified PVA has been added. Clearly, the geopotential anomaly coincides with the position of the trough axis and the PVA. Fig. 5.12c shows for the same point in time a corresponding vertical cross-section, which will be used for the PV analysis in this case study. The location of the PVA matches the PV intrusion over southeastern Asia.

5.9.1 PV analysis

To examine the precursors of the extreme event, we visualize the 3-D PV environment over eastern Asia for five different time steps in Fig. 5.13. These panels provide snap-shots, starting four days prior to the extensive rainfall. For an additional depiction of the event based on an extended time frame, also including intermediate time steps, we provide an animation of the synoptic evolution of the PVA over time in a supplementary video at Fischer et al. [53].

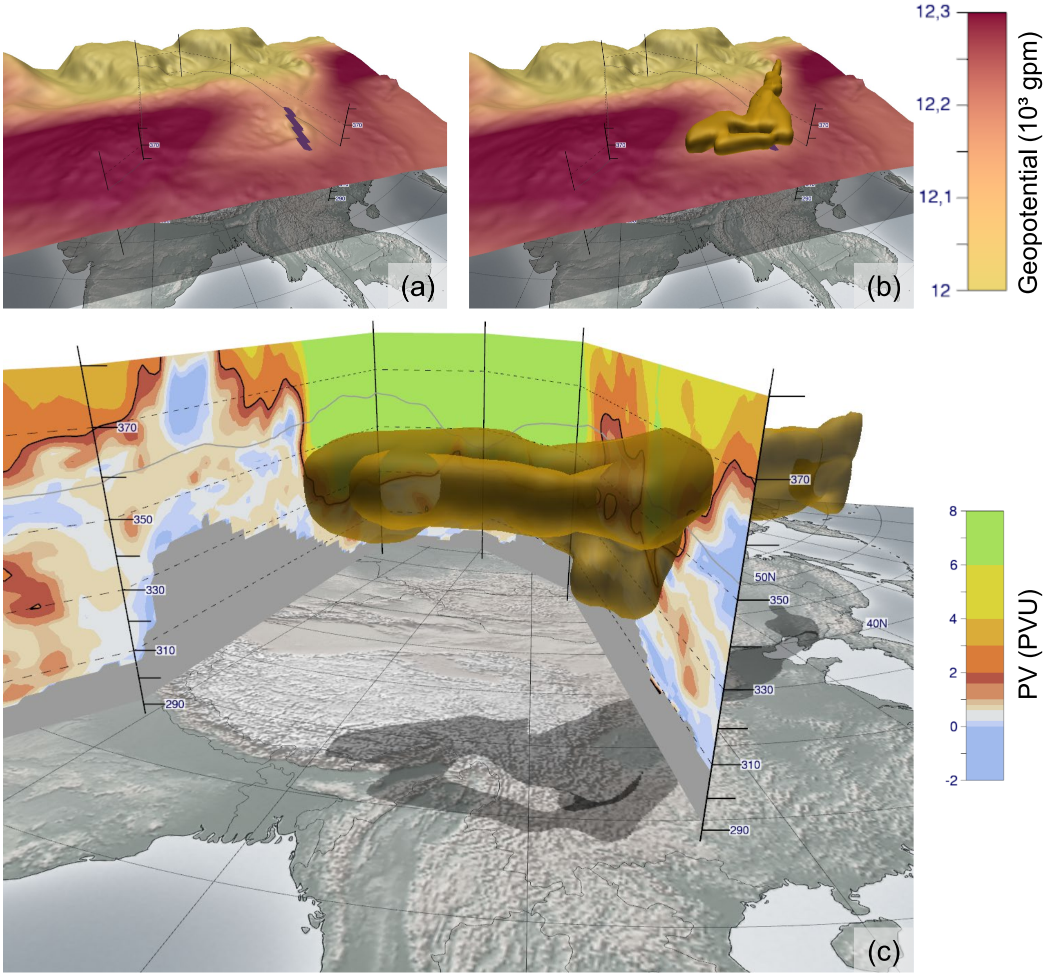


Fig. 5.12: Comparison of the location of the anomalies defined by the 200-hPa geopotential height and by the dynamical tropopause, at 25 July 2015, 12 UTC. (a) Shown is the 200 hPa isosurface on isentropic levels, shaded by geopotential height. The purple line is the trough identified by Van der Linden et al. [189] in a 2-D environment, cf. their Fig. 7. (b) Same as (a), but the identified PVA has been added. Its location clearly coincides with the geopotential trough and the identified trough axis. (c) A vertical cross-section of the area along the section indicated in (a, b), and the identified PVA. Shaded on the cross-section is the PV, where the black contour highlights the 2-PVU boundary. The gray line is the 200-hPa contour, thus a cut of panel (a).

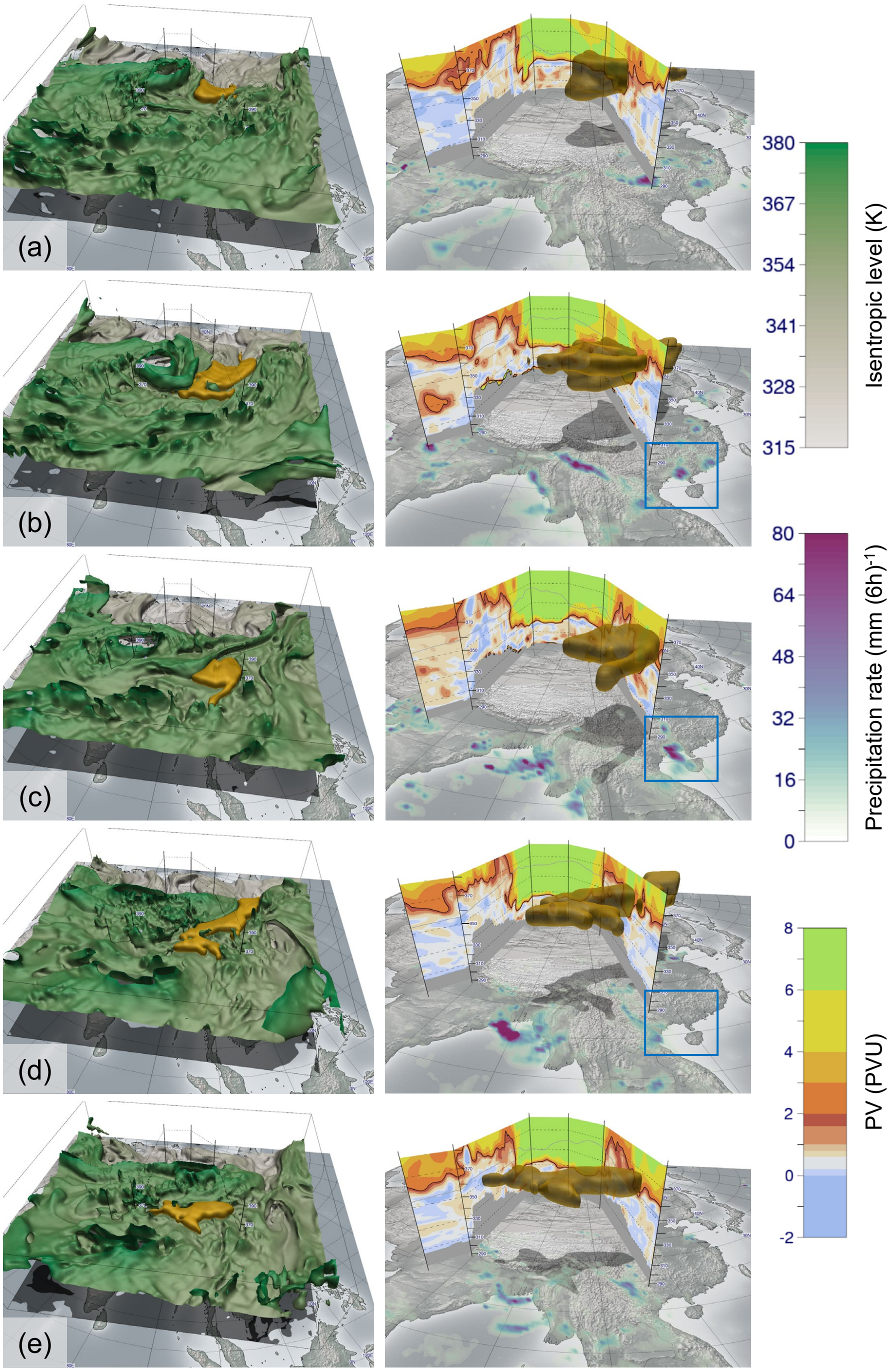


Fig. 5.13: The synoptic-dynamic development associated with the event. Here, five time steps are visualized: (a) 22 July, 12 UTC; (b) 24 July, 00 UTC; (c) 27 July, 00 UTC; (d) 30 July, 12 UTC; (e) 31 July, 12 UTC. Respectively on the left, a 3-D visualization of the dynamical tropopause shaded by isentropic height, and the identified anomaly of interest is shown. On the right, the cross-section as displayed in Fig. 5.12c is shown. The blue box in (b-d) emphasizes the torrential rainfall triggering the flood. An animation with an extended time frame is provided at Fischer et al. [53].

At 21 July 2015, a wide trough is located over eastern China, which is below the thresholds for our identification strategy in 2-D and 3-D to be detected as a PVA. Over the following days, a high tropopause area is moving from the Arabian Peninsula towards the Tibetan Plateau, further illustrated in the supplementary animation [53]. In the visualization, this anomaly is characterized by a “hole” in the 2-PVU isosurface, and a high tropopause on the western side of the vertical cross-sections in Fig. 5.13a-c. Holes in the tropopause are heights beyond our domain that has an upper limit of 380 K. From 22 July 2015, 00 UTC onward, our algorithm identifies a PVA related to the event over Vietnam (shaded in yellow in Fig. 5.13). The high tropopause over the Tibetan Plateau shows to be stationary over multiple days, leading to intense advection of high PV air southwards along its eastern flank. This leads to an intensification of the identified anomaly until 24 July (Fig. 5.13a-c). After this, the influence of the Tibetan high becomes weaker, cutting the intrusion of high PV air from the north. The identified structure partially cuts off horizontally from the main reservoir (Fig. 5.13c). During this process, the centroid of the PVA moves southward from around 30°N over mainland China to around 20°N just west of Vietnam. This is clearly visible by the PVA moving southwards along the cross-section from Fig. 5.13a-c. The period afterwards coincides with the most intense rainfall over the coast of northeastern Vietnam as indicated by the 6-hourly precipitation rate (cp. Fig. 5 in Van der Linden et al. [189], and emphasized blue boxes in Fig. 5.13b-d). Over the following days some dynamical coherence to the main tropopause fold can still be observed (Fig. 5.13d, e). Finally, on the first days of August the PVA dissipates [see 53].

Comparison with the results of Van der Linden et al. [189], who used 200-hPa geopotential height and 500–200-hPa vertically averaged PV to detect the subtropical trough (their Fig. 7), clearly illustrates the advantage of our novel 3-D approach. Using vertically averaged PV, and single-level geopotential height, the trough could not be identified after 25 July, and 29 July, respectively, by Van der Linden et al. [189]. Since high precipitation amounts were also observed after 29 July and in more western parts of the country (not shown), a continuing influence of the PVA seems reasonable.

5.9.2 Algorithm evaluation

As outlined in Sect. 5.8, two parameters in this algorithm for 3-D cases are used, the width threshold w (see Sect. 5.6) and the extent threshold b to separate PV

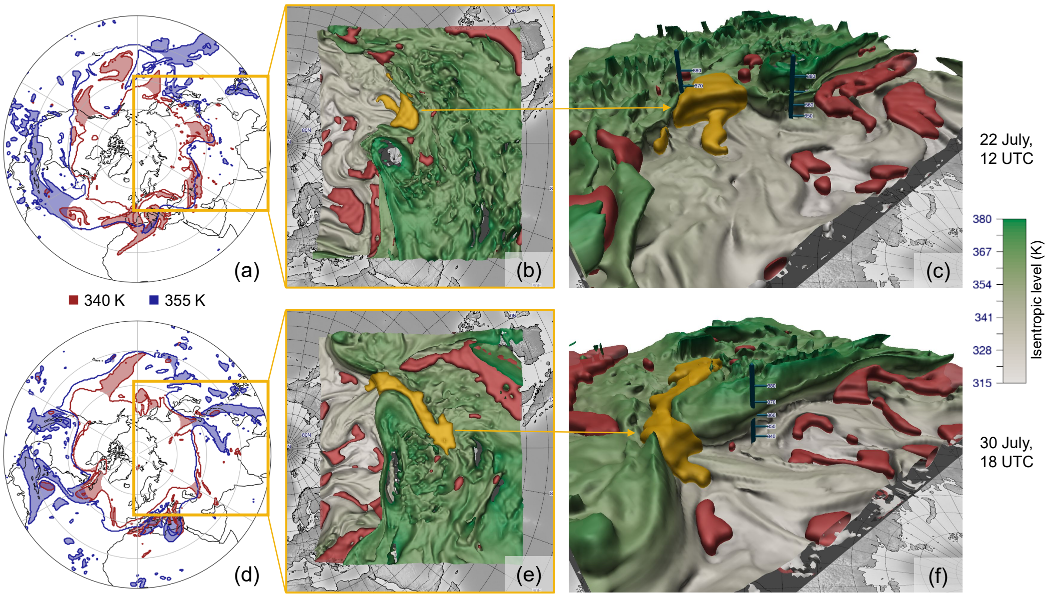


Fig. 5.14: Comparison of the 340 K and 355 K isentrope, and the corresponding 3-D analysis for two time steps throughout the case study. (a): The 2-D analysis at 22 July, 12 UTC, where the anomaly is identified on 340 K (red), but not on 355 K (blue); (b): top-down view of the 3-D analysis; and (c): side view of the 3-D anomaly. (d-f): Same as (a-c), but 8 days later. The anomaly is now identifiable on 355 K, but not anymore on 340 K. Therefore, a 2-D analysis of the event considering a single isentrope is insufficient.

anomalies into clear disjunct features. For this case study, we use the default configuration of w =1500 km and b =6 K.

In order to demonstrate the distinction between identifying 3-D structures compared to their 2-D counterparts, Fig. 5.14 presents snapshots from two different time steps of the case study. In Fig. 5.14a-c, a time step early in the development of the event is shown. The yellow anomaly can be identified as an equatorward intrusion of stratospheric air in the 3-D view. Considering identification on 2-D levels (340 K and 355 K isentropes in Fig. 5.14a), the structure is not identified on the 355 K isentrope due to its broad shape, but the anomaly is clearly distinct on the 340 K level. Because our 3-D technique (Fig. 5.14b, c) takes the whole 3-D neighborhood into account, the identified object covers both selected levels. Just looking at single 2-D isentropes individually is not sufficient to grasp the vertical structure of the PV object. Later in the event (Fig. 5.14d-f) the same anomaly that is clearly identifiable in 3-D can be detected on the 355 K level, but not anymore on the 340 K level using the 2-D identification. This also falls into place when analyzing the height of the 3-D PVA's centroid (position weighted by volume and PV, see Sect. 5.6.4). During genesis, the centroid of the PVA has a height of 345 K, and in later stages of the event, it is situated at over 360 K. Therefore, the PVA moved vertically upward throughout the event, making it impossible to automatically analyze and track when considering a single isentropic level.

The case study shows as well that distinguishing between the concepts of a PVS and PV cutoff in 2-D is not applicable to the 3-D view. The vast majority of cutoffs are attached vertically to the main stratospheric reservoir, while PVS are connected typically both horizontally and vertically. Furthermore, a structure could often be interpreted as a streamer or a cutoff depending on the isentropic cross-section one is investigating. This supports our suggested nomenclature to take a more abstract approach and name these intrusions more generally PV anomalies (PVAs).

5.10 Investigating main drivers of Tropical Transition Events

The algorithm introduced in this chapter has been applied by Kriening [102] in a master thesis to investigate main drivers of Tropical Transition (TT) Events. Understanding and predicting tropical storms are crucial, yet the predictability of these events, particularly during the TT process, remains a challenge. TT involves the transformation of an extratropical storm into a tropical one, occurring in an environment conducive to this transition. This process often encounters a predictability barrier, introducing high uncertainties in forecasting due to the complexity of dynamical processes.

The master thesis by Kriening [102] focused on exploring the main drivers of TTs using forecast ensemble data in selected case studies. The algorithm presented in this chapter served as a tool for identification and visualization, aiding in the exploratory analysis of ensemble data.

The findings by Kriening suggested that factors such as the storm's location relative to the 3-D PV anomaly, the penetration depth of the anomaly, its intensity, and shape significantly contribute to the predictability of the transition event. Analyzing ensemble members revealed distinct structures in the two clusters of transition and non-transition group members. The evolution of transition group members was characterized by the formation of a hook-shaped upper-level PV anomaly, indicating a pronounced warm seclusion and leading to increased relative humidity at lower levels. The 3-D visualization highlighted that the storm's location relative to the trough structure, along with the penetration depth and intensity of the PV tower, were crucial factors for a successful TT event in the studied case.

It's important to note that Kriening's work did not include statistical or climatological analyses. Therefore, further examination of additional datasets is required to draw significant conclusions regarding the relationship between the 3-D PV structure and tropical transition predictability.

5.11 Conclusion

In this study, a novel automated identification strategy for PVAs along the dynamical tropopause has been presented. It is based on sequential distance measures in the PV field to extract anomaly-like structures. The method is capable of identifying these structures in both 2-D and 3-D data sets robustly. It is based on image processing operations, which are able to detect disturbances in a multidimensional field, and has been adapted to suit the needs and requirements of the present application. Specifically, the adaptation keeps the interpretability of the process by using physical units, and compensates the distortions introduced by the particular flattened projections used. Here, we use the Eikonal equation and reinterpret it in a novel fashion to efficiently compute accurate distances in the distorted field, taking advantage of the properties of the projection. The strategy has proven well scalable for processing big data sets. The presented algorithm takes a width parameter w as input, which controls the maximum width of anomalies being identified. Identified anomalies are assigned feature vectors containing metrics for the specific object, e.g., centroid, bounding box, intensity or best-fit ellipsoid. The objects are then filtered based on these metrics contained in the feature vectors.

A drawback of existing 2-D identification techniques is that their decisions are solely based on the outermost 2-PVU contour for a given isentropic level. It is shown that this hides important information about the structure of the field in some cases. By applying our algorithm, which considers the whole field, we therefore identify additional features not being easily captured by most algorithms described in the literature. The algorithm can be executed in the same fashion for 2-D and 3-D data sets since the individual image processing operations are independent of the number of dimensions. One main issue of the 3-D identification consists of clearly identifying separate anomalies in the 3-D PV field. We therefore added a heuristic that separates anomalies by detaching them at points of weakest vertical extent for a clear separation. This is important for meaningful feature descriptions afterwards.

In 3-D analysis, the algorithm provides an independent configuration for all seasons and use cases, and is therefore suitable for climatologies. In 2-D analysis on the other hand, care has to be taken of choosing a suitable isentrope for analysis during a season or for a specific event, or depending on which regimes one is interested in. Also more interesting features can be detected by considering the full 3-D neighborhood instead of focusing on a single 2-D level. The 3-D structures are more cohesive and consistent in space and time compared to their 2-D counterparts. Furthermore, vertical gaps introduced by stacking results of the 2-D identification are prevented.

A case study has been investigated to show up advantages of the 3-D analysis. A trough detected using single-level geopotential height or vertically averaged PV as in Van der Linden et al. [189], which was instrumental in the evolution of the presented extreme event, would dissipate too early when compared with the results using the novel approach. Comparisons of 2-D and 3-D identification results reveal that the 3-D object is present both over a broader vertical range and longer time period. Especially the longer period coincides more closely with the observed rainfall than the 2D analysis by Van der Linden et al. [189]. Therefore, the novel method provides a more detailed view of the dynamical forcings of the extreme event by taking into account the vertical evolution and movement. This supports conclusions by Bithell et al. [16]. They emphasized that without a full 3-D view of the developing system the extent of features, such as tropopause folds, and their depth of penetration into the troposphere, are difficult to follow.

We put effort in describing the identified object in an abstract geometric fashion, similar to ellipses in 2-D. There, most structures can be well approximated by one (or in degenerative cases a few) simple geometric shapes. These are also clearly interpretable, e.g., the main axis yields the tilt of the streamer. For 3-D objects, we considered using subsets of quadrics (second order surfaces, see Krivoshapko and Ivanov [103]) but recognized that the extension of 2-D fitting to 3-D data is a highly non-trivial task. To explore the geometrical structure, but also to exploratorily analyze the presented algorithm and the data set itself, interactive 3-D visualizations proved to be an essential tool for comprehensive depictions. Derived displays, such as cross-sections, are vital to highlight dynamical processes in complex environments and to perform coherent conclusions.

In the future, this technique can be applied both to more individual use cases and to big data sets. Analyses and climatologies help to exploratorily examine the properties and behavior of PVAs in certain 3-D environments. Computed feature descriptions could be a base to find correlations and clusters in data, classifying anomalies into distinct categories. Since the algorithm is well scalable, it can be applied to ensemble forecasts as well. However, the performance bottleneck for big data analysis is the computation of high resolution isentropic data out of model levels. Therefore, processing large data sets might require a trade-off by using a different vertical dimension even though this leads to a more challenging interpretation, especially when evaluating horizontal cross-sections. Although this study focuses on the identification on anomalies along the dynamical tropopause, the algorithm in a broader sense can be applied to identify anomalies along any 2-D or 3-D spatial boundary. Therefore, other identification tasks in meteorology can also benefit from these ideas and strategies.

Further research is required on certain aspects of this identification, which might lead to improved results. For individual use cases, the tracking of these anomalies could be done by hand, whereas for climatologies automated tracking techniques are required. Ideas for further work include simple overlap tracking [e.g., 109], or a Lagrangian view [e.g., 141]. The feature descriptions, which have been computed for each PVA, might serve as base for similarity measures and be used for tracking purposes.

The automated 3-D PV identification, tracking, and its application to many meteorological cases opens an avenue to study, which characteristics of the PV objects are related to the genesis and improved forecast using statistical and machine learning approaches. For example, Maier-Gerber et al. [112] successfully used ensemble mean and standard deviation of vertically averaged upper-level PV forecasts in the Gulf of Mexico as a predictor for subseasonal statistical-dynamical forecasting of tropical cyclone occurrence. Here, a 3-D PV object related approach holds the promise of further improvements. These abstracted objects can also facilitate combinations with other displays, e.g., other identified atmospheric features, to explore further connections.

6

An objective identification technique for potential vorticity structures associated with African Easterly Waves

6.1 Preface

This chapter covers the identification and tracking of 3-D PV structures within AEWs, and the climatological analysis of these identified features. A substantial portion of this chapter has been published in GMD[1] [48] and therefore undergone an anonymous peer-review process. While the design and implementation of the processes, the computation of the climatological analyses, and the writing was mainly done by myself, the article received substantial improvements by the co-authors and the anonymous peer-reviewers. Especially thanks to Andreas Fink for including his meteorological expertise to analyze and interpret the climatological results to give the work an applied component. To emphasize the computer science facets, certain sections have been expanded, cross-referencing to the background chapters 2 and 3. The generated data set of 2-D wave troughs based on the ERA5 reanalysis has been used by multiple researchers and forecasters. Code and data is available as stated in the appendix at the end of this thesis.

6.2 Abstract

Tropical Africa and the North Atlantic Ocean are significantly influenced by African easterly waves (AEWs), which play a fundamental role in tropical rainfall and cyclogenesis in that region. The dynamics of AEWs can be described

---

[1] `https://www.geoscientific-model-development.net/`

in a potential vorticity (PV) framework. The important impact of latent heat release by cloud processes is captured in this framework by the diabatic generation of PV anomalies. This paper introduces an innovative approach for the identification and tracking of PV structures within AEWs. By employing AEW tracking and computing the wave phase of each point within the AEW domain using a Hilbert transform, we are able to effectively identify and collect 3-D PV structures associated with specific AEWs. To facilitate a climatological analysis, performed here over the months of June to October from 2002 to 2022, these structures are subsequently characterized by low-dimensional descriptors, including their location, intensity, and orientation.

Our climatological analysis reveals the seasonal evolution and the structural attributes of PV anomalies within AEWs over the study domain. PV feature locations closely align with the African easterly jet's latitudinal shift during the summer season. Analysis of the mean pressure level of the 3-D PV structures shows a remarkable shift during their life cycle, indicating deep moist convection characteristics over land, and more shallow convection characteristics over the ocean. On average, PV features identified within AEW troughs tilt downshear over land and equatorward over the ocean. The trough-centered analysis reveals distinct differences between satellite-estimated and model-predicted rainfall. Agreement between the results of a more traditional composite analysis and our new feature analysis provides confidence in our feature approach as a novel diagnostic tool. The feature framework provides a low-dimensional representation of the PV structure of AEWs, which facilitates future statistical analyses of the relation of this structure to, for example, tropical cyclogenesis or the predictability of tropical rainfall.

6.3 Introduction

African easterly waves (AEWs) are synoptic-scale disturbances that play a crucial role in the weather and climate of tropical West Africa and the tropical Atlantic region. AEWs are quasiperiodic perturbations, typically originating over the broader Lake Chad region in central North Africa or being triggered by the high topography over the Ethiopian Highlands [117, 66]. These waves propagate westward across West Africa and the North Atlantic Ocean and as far as the eastern Pacific. They have drawn considerable attention due to their substantial impact on Atlantic tropical cyclone (TC) genesis [e.g., 160, 129, 145], the rainfall variability over the West African monsoon region, their relation to the West African offshore rainfall maximum [e.g., 67], and their role in extreme precipitation events over tropical West Africa [e.g., 47, 29, 43]. Understanding the formation, propagation, and interaction of AEWs with the ambient (thermo-)dynamical state of the troposphere is essential for improving the prediction of North Atlantic TCs and rainfall patterns in West Africa.

The potential vorticity (PV) framework is a fundamental fluid-dynamical conceptual model widely utilized in extratropical meteorology, including the understanding of barotropic and baroclinic instabilities, Rossby wave propagation and amplification [e.g., 83], Rossby wave breaking [116, 181], and the importance of latent heat release and other diabatic processes to Rossby wave dynamics [e.g., 180]. Furthermore, under a balance assumption, the wind, temperature and density fields can be derived solely from the PV field. Strong latent heat release associated with convection or intense rainfall [198, 122] leads to horizontal and vertical dipoles of PV and, thus, rich small-scale structures. Müller et al. [122] highlighted the correlation of strong PV anomalies and intense precipitation and suggested that PV anomalies may serve as a proxy for evaluating intense rainfall. In the absence of nonconservative processes, PV is materially conserved, which in combination with large PV gradients in the midlatitudes makes it relatively straightforward to identify and track PV features associated with Rossby waves and other large-scale flow features [e.g., 180, 49, 72]. In the tropics, however, much smaller PV gradients and more prominent contributions of convective-scale latent heat release imply a more complex and smaller-scale nature of the PV distribution, which is in particular true for AEWs. A feature-based PV perspective of AEWs thus faces challenges that demand a more detailed investigation.

There exists a substantial body of research on AEWs, especially encompassing their dynamical interaction with the environment and their relationship with TCs. The dry dynamics of AEWs can be understood in terms of downstream propagation along the African easterly jet (AEJ) from an upstream wave source [183], with (small) amplification by baroclinic and barotropic growth [65]. As in the midlatitudes, these processes can be described from the PV perspective. More important for AEW amplification is latent heat release associated with embedded convection [14, 183]. From a PV perspective, this amplification is seen as the diabatic generation of PV anomalies. Besides amplitude, the diabatically generated PV anomalies signify modification of AEW structure. Tomassini et al. [185] investigated in detail the contributions of different parameterization schemes to diabatically modified PV in an operational numerical model. Russell et al. [160] and Russell et al. [161] give further insight into the structure and sources of PV in AEWs, including the role of moist convection and its coupling with the background wave environment. Essentially, the diabatically generated PV anomalies encapsulate the impact of moist processes on AEW intensity and structure that outlasts a period of active convection.

In terms of AEW predictability, there is a strong indication that uncertainty in latent heat release can be linked to larger forecast errors in AEW characteristics [40, 41]. Recently, Núñez Ocasio et al. [130] suggested that the type of convective organization and the location of the convection relative to the AEW trough may play a discriminating role in terms of AEW-related TC genesis in the North Atlantic Ocean. Dunkerton et al. [35] provide a conceptual framework that links AEWs and TC genesis. In this framework, the nonlinear critical layer of AEWs provides a region of recirculating air masses, in which moisture and PV may accumulate with time in relative isolation from the drier and thus more hostile environment. Despite these potential applications in understanding AEW growth, predictability and their role in TC genesis, the application of a PV-centric view, especially in three dimensions, in studying AEWs has been relatively limited.

In this study, we address this gap by providing a comprehensive tool for the PV-centric analysis of AEWs. Specifically, we address the following objectives:

- We propose a novel identification and tracking technique for 3-D PV features associated with AEWs. It facilitates the quantification of feature properties in case studies and climatological analyses. By describing features by a low-dimensional vector, statistical analyses including feature climatologies [e.g., 109], ensemble forecast analysis [e.g., 148], and statistical postprocessing [e.g., 147] can be performed.

- We perform a climatological analysis of these identified PV features to explore the properties of these features over their life cycle. Furthermore, it provides confidence in the previously introduced method by comparing the identified features to climatological PV composites.

Multiple approaches exist for objectively identifying AEWs. Early approaches by Burpee [21], Reed et al. [155], and Diedhiou et al. [33] focus on the identification of mean tracks rather than individual wave anomalies. In order to be able to compute climatologies and to perform statistical analyses of the data, an objective identification is absolutely vital. Objective identification techniques for atmospheric features have proven beneficial in both atmospheric research for statistical analyses and verification tasks and operational meteorology [e.g. 74]. Fink and Reiner [47] summarize the strength and weaknesses of automatic and manual approaches. Thorncroft and Hodges [182] were the first ones to exploit the usefulness of the vorticity measure to objectively track individual AEWs in the tropics. Other approaches include the identification using Hovmöller diagrams [8]. Berry et al. [15] and Belanger et al. [11] most recently used the advection of curvature vorticity as the primary measure of AEW activity and to identify wave tracks. This measure proved to be very robust after applying downscaling and smoothing operators.

To enhance our analysis of areas with high PV associated with AEWs, we incorporate phase filtering through the Hilbert transform. This technique is essential because the geographic location of an AEW trough often does not match the regions with the highest PV anomalies, due to the complex interactions between wave movements and diabatic PV influenced by convection. According to Shapiro [168] and Tomassini et al. [185], convection within AEWs typically begins ahead, or west, of the trough over Africa, where atmospheric conditions are most conducive to convection [156, 47]. Conversely, over the ocean, convection generally occurs closer to, or slightly east of, the trough [158]. Since convection can potentially take place across the entire trough area, accurately assigning the current phase to every point in the domain becomes crucial for filtering PV signals in AEWs.

To reach the aforementioned goals, we employ a technique similar to the one by Belanger et al. [11] to get a robust position of AEWs on 700 hPa, this is outlined in Sect. 6.4.2 and 6.4.3. Then, in Sect. 6.4.4, we compute the wave phase at every point in the domain by performing a Hilbert transform, which is required to narrow down the regions around AEWs where high PV associated with these AEWs is expected [47]. Then finally, in Sect. 6.4.5, we identify and extract the PV structures. The identified 3-D features are assigned geometric descriptions as low-dimensional representation, including a best-fitting ellipsoid. Section 6.5 shows results of a climatological analysis of this feature, shedding light on the distinct characteristics of the waves. Additionally, the approach is verified through comparison with PV composites. A trough-centered analysis highlights differences between satellite-estimated and model-predicted rainfall. The chapter is concluded in Sect. 7 where we discuss the implications of our findings, their relevance to previous research, and the potential avenues for further investigation.

6.4 Strategy

6.4.1 Data

For the identification of AEWs and the corresponding PV features, as well as for the analyses in this study, we utilize data from the global ERA5 reanalysis [75]. Our analysis focuses on the period from June to October to align with the West African monsoon season, as detailed by Fink et al. [46]. The selected data are provided on a regular grid with a grid-point spacing of 0.5° in both latitude and longitude and with a temporal resolution of 6 hours. This timescale is sufficient since AEW characteristics do not vary on hourly timescales. To identify the wave trough of AEWs, we use the zonal (u) and meridional (v) wind components at 700 hPa, where the AEJ has its maximum speed [46]. The study domain covers the region of AEW activity over tropical West Africa and parts of the North Atlantic Ocean, specifically from 75°W to 45°E and 0°N to 40°N. The PV analysis is conducted on 16 pressure levels between 200 and 900 hPa with intervals of 50 hPa. The PV data on pressure levels originate from the ERA5 archive. We exclude the pressure levels beneath 900 hPa where surface-induced effects come into play. For rainfall analyses, the satellite-gauge-based GPM IMERG V06B data set [86] and the twice-daily ERA5 short-range forecasts will be used to quantify rainfall in relation to AEWs and PV occurrence.

6.4.2 Identification of AEW wave troughs

To robustly identify PV features within AEWs, we build on the method by Belanger et al. [11] to firstly identify AEW troughs on 700 hPa. Their method is an improvement over previous work [e.g., 182] by applying curvature vorticity (CV) anomalies instead of absolute values of relative vorticity and by ensuring that waves are westward propagating, which more closely aligns with the characteristics of AEWs. We have improved the tracking of the waves to achieve more robust and consistent tracks. Additionally, we introduce a data structure to facilitate the analysis of the tracks, including the identification of split events and merge events, the computation of average wave speeds, and other parameters.

A climatology of CV is computed for each month and each time of day to take into account both seasonal and diurnal effects. Then, the CV itself is band-pass-filtered to only retain frequencies that are linked to AEW disturbances. We use a filter to keep frequencies between 2–8 days, following Russell et al. [161]. To identify wave troughs at a given point in time, we compute the anomalies of the CV (CVA) by subtracting this band-pass-filtered CV from the climatology. In the idealized case of westward-propagating waves, positive CVAs are linked to cyclonic rotation present in the wave trough, while negative CVAs are present in the ridge regions. Figure 6.1a shows an example of CV anomalies. The band-pass-filtered wind leads to distinct alternating positive and negative poles of CVA. Following Belanger et al. [11], the data are smoothed twice using a nine-point local smoother, which retains only the synoptic-scale easterly wave structure. Then, the zeros of the CVA advection are determined, which resemble the lines where the sign of the advection changes. These are the troughs and ridges in the CVA field. Given the grid's discrete nature, a cell-wise computation would almost never find zero values. Hence, we employ the Marching Squares algorithm to interpolate line segments between grid cells that approximate these zero lines. These segments are then merged to form continuous lines. In Fig. 6.1a, the grey lines indicate zeros of CVA advection, which are collocated with troughs and ridges.

For filtering purposes, following Belanger et al. [11], two masks are applied to the identified lines:

- The zonal wind must be $u < +2.5\mathrm{ms}^{-1}$.
- The CVA must be over the 66-th percentile of the entire reanalysis.

Furthermore, we consult the second derivative of the CVA to extract only troughs in the data set, masking out the ridges. The red lines in Fig. 6.1a are the result of applying the other masks and filters to the data set. This results in the identified wave troughs.

6.4.3 Tracking of AEW wave troughs

Research on tracking AEWs has produced various methodologies, where the most established ones are based on analyzing the CV field. Hollis et al. [79], for example, utilize an approach based on the well-known TRACK algorithm, originally proposed by Hodges [78]. This method primarily focuses on linking point features across successive time frames based on a predefined, physically reasonable propagation speed. Lawton et al. [106] adopt a different approach by tracking AEWs through meridional averages of CV and velocity. Bain et al. [8] and Brammer and Thorncroft [19] employ Hovmöller plots for their

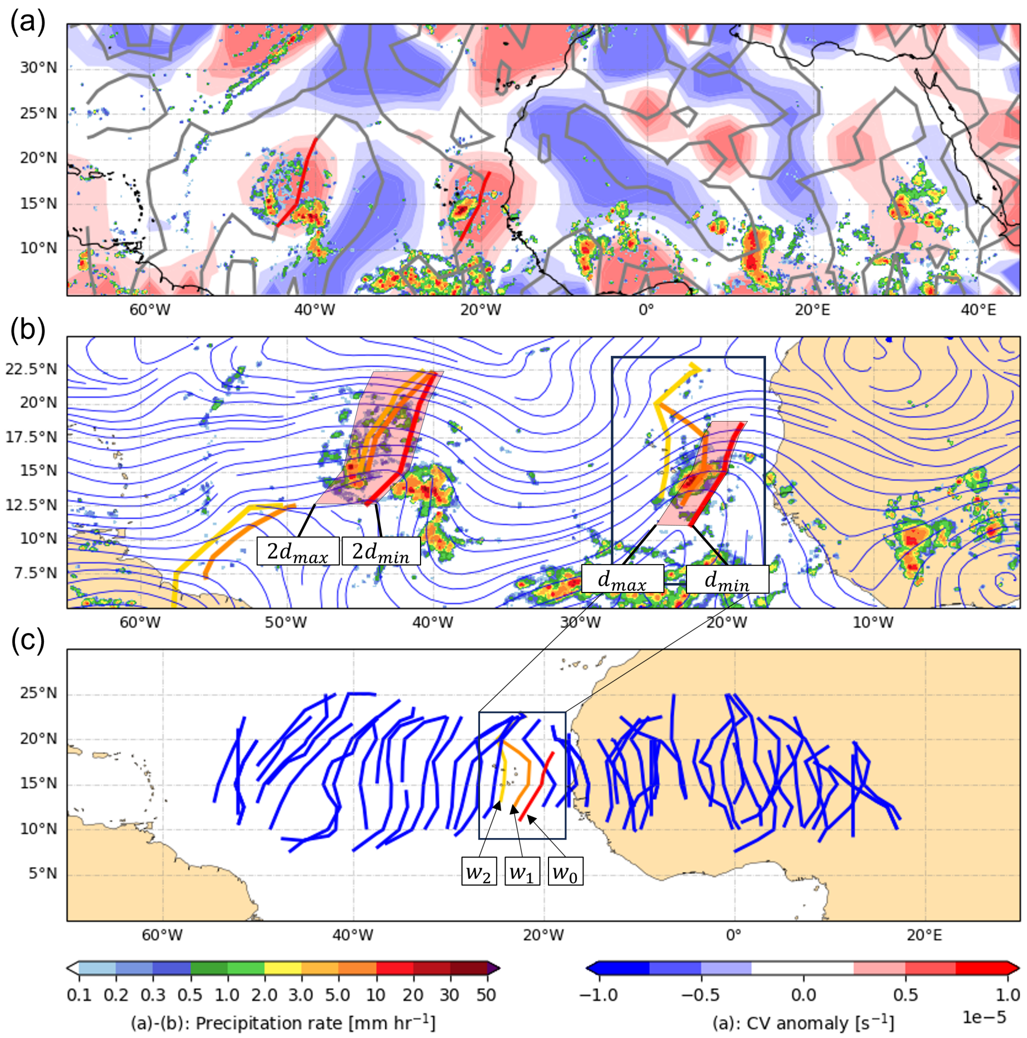


Fig. 6.1: Illustration of the identification and tracking process for AEWs using a polygonal search approach. Panel (a) shows the CV anomalies and the precipitation rate according to the GPM IMERG data set from 13 September 2022 at 00 UTC. The grey lines indicate the zeros of CVA advection, and the red lines show the extracted wave troughs based on the filters introduced in Sect. 6.4.2. In panel (b), two identified wave troughs are shown at 00 UTC (red lines, as in (a)), 06 UTC (orange lines), and 12 UTC (yellow lines). The red boxes sketch the polygonal search areas computed for both wave troughs, initiated at 00 UTC. The left wave trough is compared with the +12 h wave trough, while the right wave trough is compared with the +6 h wave trough. The 700 hPa streamlines of the band-pass-filtered wind at 00 UTC are depicted as blue lines. In panel (c), an entire AEW track is depicted, demonstrating the continuity of the tracking process. The three wave troughs from panel (b) are highlighted.

tracking, analyzing the longitudinal movement of waves. While each of these methodologies has proven effective in their respective applications and has gained popularity in the field, they predominantly focus on point features. In contrast, our work, along with that of Belanger et al. [11], leverages additional information provided by the identified wave trough features.

To form tracks from the identified individual wave troughs, we employ an overlap approach. Overlap tracking has proven to be a robust tracking technique in meteorological applications, such as the tracking of mesoscale convective systems [MCS, e.g., 130, 45, 143] and general purpose feature extraction [e.g., 187]. However, since our identified wave troughs are represented as line strings, they do not directly lend themselves to traditional overlap tracking methods.

To address this, we create area features by predicting the future positions of each trough for upcoming time steps, $t + \Delta t$ and $t + 2\Delta t$, with $\Delta t = 6\,\mathrm{h}$. This prediction uses an anticipated propagation speed to define a polygonal area that represents where the trough is expected to be. The presence of overlap between these predicted polygonal areas and the actual locations of wave troughs at future time steps facilitates the tracking. The range of expected propagation speeds, which we set from $u_{max} = -15\,\mathrm{ms}^{-1}$ to $u_{min} = +2\,\mathrm{ms}^{-1}$, defines the size and shape of these polygons, as illustrated in Fig. 6.1b. While AEWs typically propagate westward with a speed between $5\,\mathrm{ms}^{-1}$ and $10\,\mathrm{ms}^{-1}$ [63], we use a broader range to account for sampling artifacts due to the data resolution. Nonetheless, based on the expected wavelength of AEWs [e.g., 46], this range still ensures that we do not mix up different waves into the same track. Given a time granularity Δt, the bounds of expected traveled distance d_{min} and d_{max} can be computed by multiplying u_{min} and u_{max} with the elapsed time.

With this approach, wave tracks can persist even when a trough is detected at t and $t + 2\Delta t$ but not at $t + \Delta t$. This happens when one of the abovementioned identification masks is not fulfilled for this time step, but for the ones before and after. Consequently, maintaining longer continuous wave tracks, rather than multiple short tracks, enhances the quality of the tracking results.

Using the identified connections, a connection graph $G = (W, E)$ is formed, akin to the approach outlined by Limbach et al. [109] and introduced in detail in Section 2.3.2. This structure enables the application of graph theory concepts and algorithms to our identified features, as also employed by Whitehall et al. [201] in their analysis of MCS. W represents the set of all wave troughs and serves as the nodes in the graph, while E denotes the edges connecting elements of W. A pair of wave troughs $w_1, w_2 \in W$ form an edge $(w_1, w_2) \in E$ if they represent the same entity, thus fulfilling the overlap strategy outlined above. This graph provides a concise representation of the union of all wave troughs and their evolution over time. For instance, if there exist two wave troughs w_1, w_2 following wave trough w, such as $(w, w_1) \in E$ and $(w, w_2) \in E$, then the wave trough splits in two, indicating a split event. The connection graph is being simplified by removing transitive edges that are primarily introduced through the comparison of wave troughs from nonconsecutive time steps. Thus, if all $\{(w_1, w_2), (w_2, w_3), (w_1, w_3)\} \subseteq E$, then (w_1, w_3) is eliminated from E.

After the connection graph has been simplified, tracks are extracted from the graph. Each track T^i itself is a graph $T^i = (W^i, E^i)$, based on a subset of wave troughs $W^i \subseteq W$ and connections $E^i \subseteq E$. The tracks T^i are disjoint subsets of the overall connection graph G. Following the challenges and solutions for extracting tracks from the connection graph outlined in Section 2.3.3, we adhere to the rule to generate tracks with the longest possible lifespan. This makes it easier to investigate the life cycle of these waves. Alternative heuristics, such as extracting connected subgraphs or initiating new tracks at every split and merge event, were considered. Extracting connected subgraphs could reduce the number of tracks and inherently capture split and merge events within a single track. However, this approach is sensitive to inaccuracies in feature identification and overlooks the origins of merged features, connecting too many features into one. On the other hand, creating new tracks at each split or merge event significantly increases the number of tracks, making life span analyses unfeasible.

For each identified wave trough being part of a track, we compute its current speed as the average speed over a 2-day window centered around the trough's time step along this track. For this purpose, the position of a wave trough is defined by the center of its bounding box. We discard parts of tracks with an average speed of less than 3 ms^{-1} at any given point in time. This removes stationary features based on orographic effects. Furthermore following Belanger et al. [11], we also remove tracks with a lifetime of less than 2 days from the track set.

Figure 6.1c shows the full life cycle of one tracked AEW from panel (b). Each AEW trough represents a node in the graph. The troughs of the depicted track are linked using edges: the w_0 manifestation of the wave is connected to w_1;

thus $(w_0, w_1) \in E$. w_1 itself is connected downstream to the next trough w_2 and so on. When the track splits, both ways can be continued in this way. We refer to Limbach et al. [109] for a detailed introduction on this type of data structure.

The output of the tracking algorithm consists of a list of AEW troughs with unique identifiers and their descriptions and a list of tracks where each track is defined by a list of edges defining that track. The location of the AEW is defined as a line string, thus, a list of latitude-longitude points. The computed AEW trough data set, spanning the entire ERA5 reanalysis period from 1940 to 2022, is available as detailed in the *Code and data availability* section of this chapter. Also, a near-real-time implementation displaying identified waves from the ECMWF, GFS, and ICON deterministic forecasts is available[2].

6.4.4 Phase computation

Although an AEW trough defines the line of maximum CV, its location does not necessarily coincide with the expected region of maximum PV anomalies, which could occur in the entire trough phase. This discrepancy underscores the need to compute the wave phase for each point across the domain. Looking at the meridional wind component v in the background flow, the trough area can be assigned a phase between $-\pi$ and 0, and the ridge can be assigned a phase between 0 and π. However, it is not straightforward to extract the phase information from a real-valued field, since all frequencies in the band-pass-filtered range of 2–8 days can contribute to it.

Zimin et al. [208] faced a similar problem and utilized digital signal processing methods to extract amplitude and phase information of Rossby wave packets. Using the so-called Hilbert transform (see a practical introduction in [144]), a complex (consisting of a real and imaginary) field can be computed from a real-valued field, which provides additional information such as amplitude and phase. This method involves performing a Fourier transform on the signal and then applying a back transform only to the positive part of the frequency spectrum. This approach is suitable for wave patterns that encompass a range of wavenumbers, like tropical waves do. Following Zimin et al. [208], we look

---

[2] `www.kit-weather.de/aew_deterministic_maps.php` (last accessed at 15 May 2024)

at the meridional wind component v along a latitude circle and perform a 2-D discrete Fourier Transform over both the wavenumber and time domains:

$$V_{f,k} = \sum_{t=0}^{T-1} \sum_{n=0}^{N-1} v(t,n) e^{-2\pi j (f \frac{t}{T} + k \frac{n}{N})}, \tag{6.1}$$

where t denotes the time step in the data set consisting of T time steps, and n denotes the longitude index out of N indices, while f and k denote the frequency and wavenumber coefficients of the transformed signal. Following Reed et al. [156] and Russell et al. [161], we choose a wavelength of 2000–6500 km and a wave frequency of 2–8 days to identify perturbances that can be assigned to AEWs. Therefore, the back transformation is only applied for the subset of the positive wavenumbers $k_{min} \leq k \leq k_{max}$ and positive frequencies $f_{min} \leq f \leq f_{max}$ corresponding to the chosen wavelength and frequency criteria:

$$v_H(t,n) = \frac{2}{TN} \sum_{f=f_{min}}^{f_{max}} \sum_{k=k_{min}}^{k_{max}} V_{f,k} e^{2\pi j (f \frac{t}{T} + l \frac{n}{N})}. \tag{6.2}$$

This process can be described as a band-pass-filter that generates complex wave information. The real part of the restored signal represents the band-pass-filtered data. Additionally, the imaginary part allows for the computation of the phase at any point in the domain using the following equation:

$$\theta(v_H) = arctan(\frac{Im[v_H]}{Re[v_H]}), \tag{6.3}$$

where $Im[v_H]$ and $Re[v_H]$ denote the imaginary and real components of v_H, respectively.

Figure 6.2 provides an illustration of a wind field at 700 hPa, along with the identified wave troughs highlighted in red. To remove phase signals in the trough area $(-\pi, \ldots, 0)$ that are not near any AEW trough at all (e.g., vorticity anomalies from other tropical and extratropical waves), we restrict the validity of the phase to a radius of 1000 km around the trough line, depicted by the red contour in Fig. 6.2. Since the transform is performed for wavelengths greater than 2000 km, this ensures we consider a substantial area that does not interfere with signals of other waves. The shading represents the phase of the wave at each point in the domain periodically from $-\pi$ to π. The wave troughs clearly align with a phase of $-\frac{\pi}{2}$ (white areas). The hatched area in Fig. 6.2 represents the designated search area for PV, called "trough area"

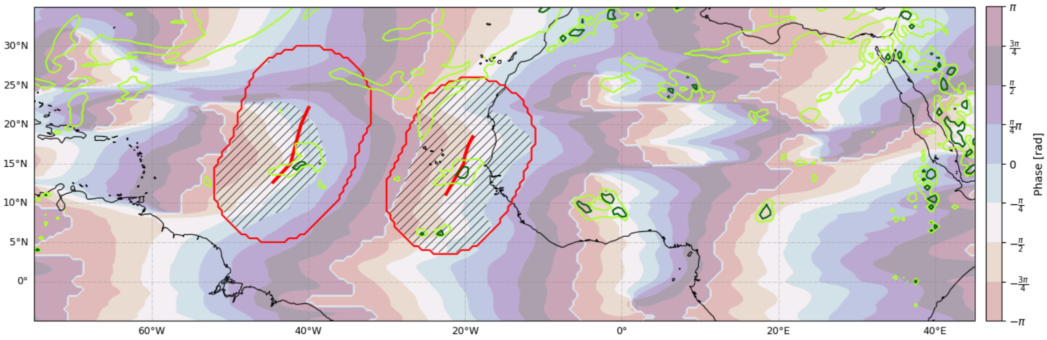


Fig. 6.2: Visualization of the computed wave phase using a Hilbert transform (shading), and the identified wave troughs with a 1000 km radius in red from 13 September 2022 at 00 UTC. The hatched region represents the intersection of the expected phase and the vicinity area of the wave trough, delineating the search area for high PV. High-PV regions are indicated by the green iso-contours (light green represents 0.4 PVU and dark green represents 0.8 PVU), which are based on the averaged PV values between 900 and 300 hPa.

from hereon, which is determined by the areas in the trough phase and near a wave trough (within the red radius).

This highlights that the Hilbert transform produces correct results with regard to the expected wave signal. In areas less conducive to AEW activity, the Hilbert transform's reliability diminishes, leading to increased noise and less accurate results in the phase field. This outcome is anticipated, as different types of atmospheric waves and other atmospheric disturbances overshadow the AEW signal in these locations. In addition, the green iso-contours in Fig. 6.2 indicate the vertically averaged PV in the lower and mid troposphere, highlighting the regions of high PV. Notably, these high-PV areas align well with the phase of and proximity to the wave troughs.

6.4.5 Identification of PV features

These previously identified trough areas in Sect. 6.4.4 delineate the search area for PV. Analyses of the PV field (e.g., Fig. 6.3), supported by 3-D analyses using Met.3D [150] have indicated that a threshold of 0.7 PVU retains the strong signal associated with the waves while effectively separating the clusters. Therefore, we identify all grid points within trough areas that have a PV value exceeding 0.7 PVU. Figure 6.3a displays a raw 3-D PV field in the tropics, visualized using the 0.7 PVU iso-contour. The noisy PV field highlights the need for a procedure to extract PV that is linked to AEWs. In Fig. 6.3b, the PV areas over this threshold have been confined based on the trough areas defined in Sect. 6.4.4 and displayed as yellow volumetric objects. Additionally, a so-called morphological opening filter was applied to these objects to refine the features. A comprehensive overview of these morphological filters is provided by Najman and Talbot [124], with a recent implementation for extratropical PV features applied in Fischer et al. [49]. These filters process the 3-D volumetric features by eliminating small, isolated outliers and smoothing out noisy areas, while preserving their general structure. Based on the proximity to the 2-D wave trough lines, each PV feature can be linked to such a wave trough. Furthermore, the tracks of the 3-D PV anomalies can be inferred from the tracks of the 2-D wave troughs, as introduced in Sect. 6.4.3. A set of feature descriptions is computed for each PV feature and consists of a bounding box,, maximum and average PV values, volume, and, as outlined next, a set of image moments and a geometric representation.

6.4.6 Ellipsoid computation

To facilitate statistical analyses of the identified features, it is desirable to represent them using a set of intuitive and well-suited parameters. Previous studies have revealed that the structure and orientation of PV features in AEWs depend on different factors, for example their location and the (thermo-)dynamic constraints of the environment [185, 161, 131]. Ellipsoids are deformed spheres and can be defined by their center and their three main axes. This simple representation encapsulates information about the shape's position, size, orientation, and elongation, while being defined by only four vectors. It has been termed as the "most economical representation" for meteorological observations by Smith and Woolf [170] and is also employed in various other scientific fields, including robotics and metrology.

Various approaches have been considered for computing a best-fit ellipsoid. Minimizing the squared error between the surface of the feature and the ellipsoid is computationally expensive and problematic on discrete grids (see an overview in [186]). An ellipsoid with the minimum volume that fully encloses the PV feature can be computed using Khachiyan's algorithm [93], but this approach is not robust against noise. We have determined that the most suitable method for computing ellipsoids is to calculate so-called image moments of the feature. Image moments have been widely used in image analysis and pattern recognition [e.g., 179, 133] as they compactly describe

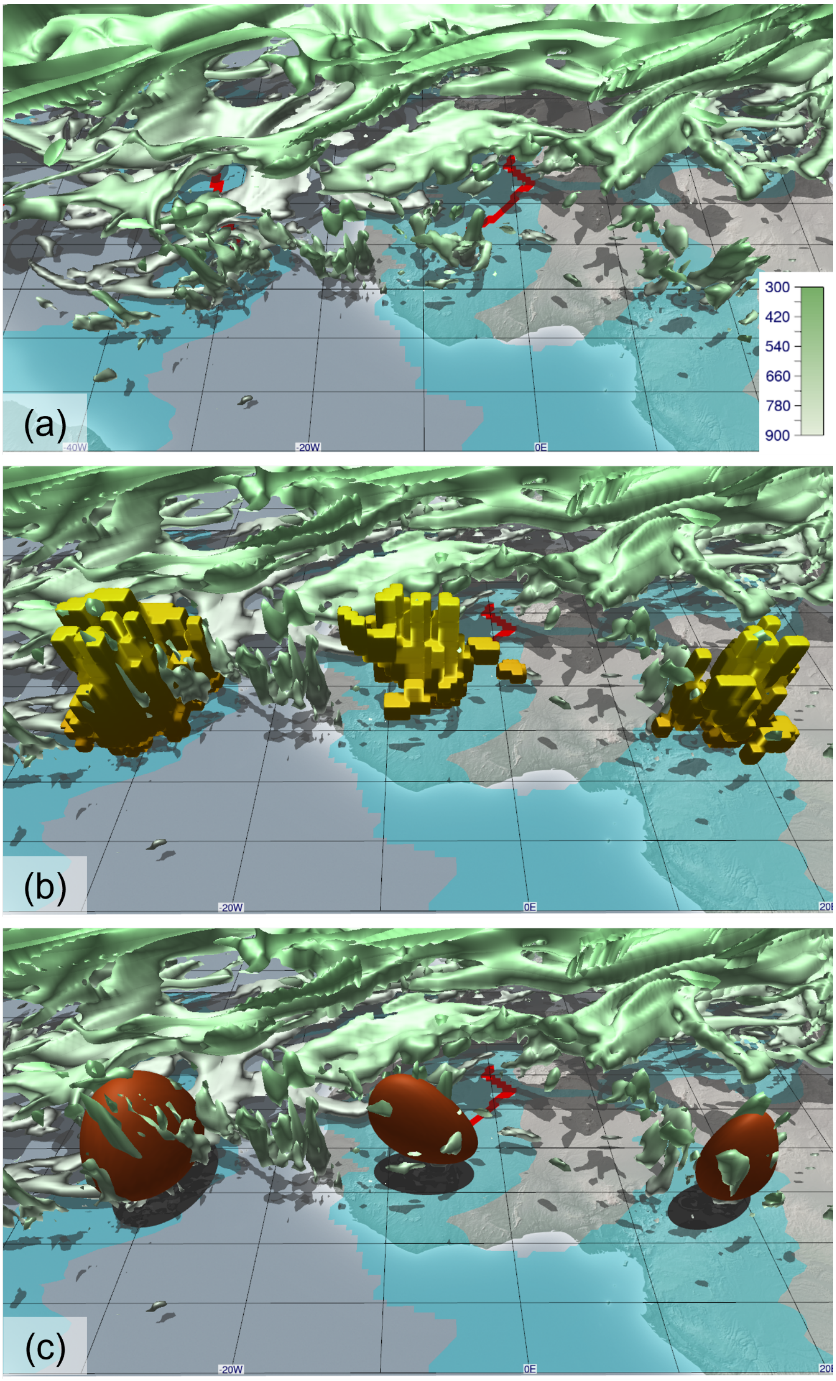


Fig. 6.3: Depicted is a 3-D visualization, created using Met.3D, of the PV across the North Atlantic Ocean and West Africa from 13 September 2022 at 00 UTC. In panel (a), the 0.7 PVU iso-contour is displayed with the shading corresponding to atmospheric pressure and the identified wave troughs in red; the trough phase ranging from $-\pi$ to 0 is highlighted in blue. Panel (b) showcases the identified PV objects in yellow as outlined in Sect. 6.4.5, while panel (c) illustrates the ellipsoids representing the identified features from panel (b).

the spatial characteristics and geometric properties of a feature. We refer to Mukundan and Ramakrishnan [121] for a detailed introduction of this concept. Importantly, we will exploit the second-order moments to compute the main orientation of a feature.

For a discrete setting, as for our grid, the 3-D moments M_{ijk} of order $i+j+k$ for the function PV are defined as

$$M_{ijk} = \sum_x \sum_y \sum_z x^i y^j z^k PV(x,y,z) V(x,y,z), \tag{6.4}$$

where PV represents the PV restricted to the identified volumetric feature (e.g., yellow structures in Fig. 6.3b); V represents the volume of the grid cell; and x, y, and z are the grid dimensions (longitude, latitude, pressure levels). Calculating the moments up to the second order provides multiple interesting characteristics of the shape. The 0th moment M_{000} represents the PV-weighted volume of the object, while the vector

$$(x_c, y_c, z_c) = (\frac{M_{100}}{M_{000}}, \frac{M_{010}}{M_{000}}, \frac{M_{001}}{M_{000}}) \tag{6.5}$$

represents the centroid of the feature weighted by PV. The second-order moments represent the variance and covariance between each pair of dimensions. Therefore, we can construct the covariance matrix $\mathbf{\Sigma}$ of the feature as

$$\mathbf{\Sigma} = \begin{bmatrix} \mu'_{200} & \mu'_{110} & \mu'_{101} \\ \mu'_{110} & \mu'_{020} & \mu'_{011} \\ \mu'_{101} & \mu'_{011} & \mu'_{002} \end{bmatrix}, \tag{6.6}$$

where $\mu'_{ijk} = \frac{M_{ijk}}{M_{000}} - x_c^i y_c^j z_c^k$ are the second-order central moments. As demonstrated in Jackson [88], the eigenvectors of the covariance matrix $\mathbf{\Sigma}$ correspond to the main axes of the feature, while the length of the main axes s_i can be computed from the eigenvalues λ_i as $s_i = 2\sqrt{\lambda_i}$. These principal axes are determined for each PV feature, defining the best-fit ellipsoid by spanning these three axes. An example is illustrated in Fig. 6.3c, where the ellipsoids are the ones computed for the yellow objects in panel (b). Furthermore, ellipsoids are invariant under projections: projecting ellipsoids on any 2-D plane yields ellipses [see 70], which in turn can be characterized by two main axes. We compute the projected ellipses on the three axis-aligned planes (xy, xz, and yz plane) and define these ellipses by their main axes.

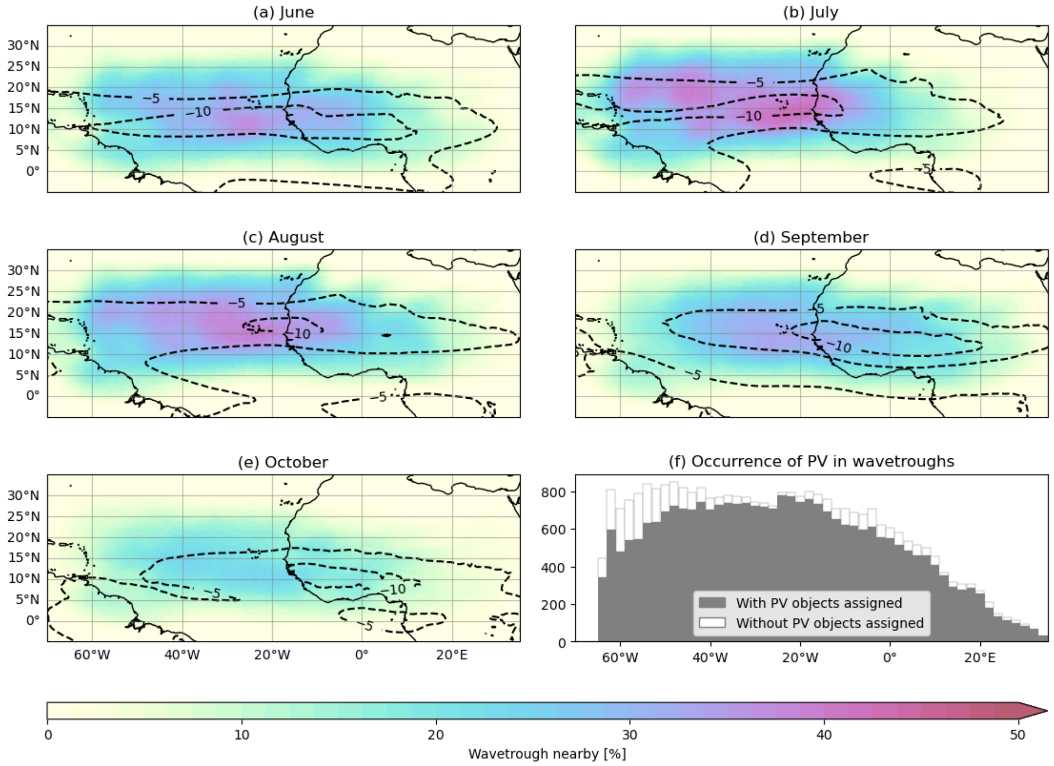


Fig. 6.4: Occurrence percentages of trough areas for the months of (a) June to (e) October at a specific point, along with iso-lines depicting mean zonal wind at 700 hPa (-5 and -10 ms^{-1}). Panel (f) shows the number of identified wave troughs at various longitudes based on their bounding box center and indicates the ratio of wave troughs which have been assigned a geometric 3-D PV representation. The ERA5 reanalysis of 2002 to 2022 has been used.

6.5 Climatology

To assess the data set generated by the identification algorithm and provide an overview of the features, a climatological analysis is conducted. Figure 6.4 presents a climatology depicting the occurrence of wave troughs and geometrical representations. In this context, occurrence represents the fraction of time steps within a given month where a grid point resides within the trough area introduced in Sect. 6.4.4, indicated by the colored shading. The climatological mean zonal wind at 700 hPa, representing the AEJ, exhibits a latitudinal movement and coincides with regions of higher wave trough occurrence. Additionally, for specific longitudes, the histogram in Fig. 6.4f displays the number of identified wave troughs centered at each longitude. It indicates whether a geometric representation, as defined in Sect. 6.4.6, has been assigned to the respective feature. Errors during the computation of the

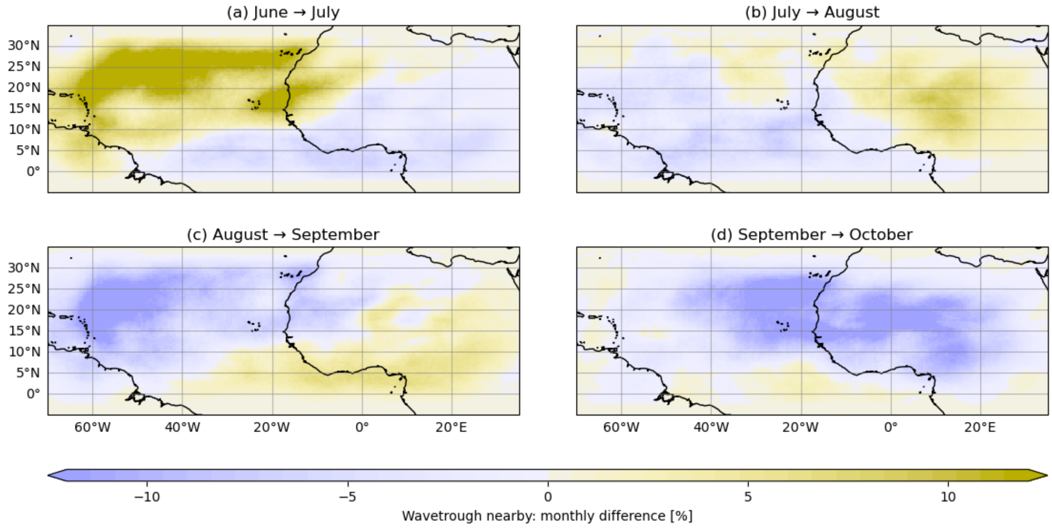


Fig. 6.5: Differences in occurrence percentage of consecutive months in the June–October season, based on the monthly data from Fig. 6.4.

ellipsoid (e.g., due to an insufficient number of data points exceeding the PV threshold) may prevent the generation of a geometric representation for a PV feature. The plot reveals that in regions where well-developed AEWs are expected, the majority of these waves exhibit a distinct PV signal. However, the PV signal weakens in the genesis region of northern East Africa due to the lack of deep moist convection and the western end of the study domain, where almost half of the waves moving towards northern South America are accompanied with inhibition of precipitation [59].

Additionally, Fig. 6.5 illustrates the changes in PV feature occurrence between the months shown in Fig. 6.4. The number of PV features increases from June to July, corresponding to the West African monsoon season's onset. Warmer and moister atmospheric conditions favor PV anomaly formation. Then, going into August, the PV activity increases further over tropical West Africa, aligning with peak convective activity during the monsoon season and the peak of wave activity in general [46]. Inversely to panel (a), September shows a decline in PV activity, mainly over the Atlantic Ocean, as the Intertropical Convergence Zone (ITCZ) shifts southward. Towards October, PV activity over West Africa retreats southward as well in tandem with the monsoon's withdrawal.

As a 3-D reference, Figure 6.6 shows a composite of the PV anomalies relative to the identified wave troughs categorized in *ocean* and *land*. The categorization is based on whether the midpoint of the wave trough's bounding box is situated over tropical West Africa (land) or the Atlantic Ocean (ocean). For each grid point along each identified wave trough being part of a track, the data are split into bins based on the longitude offset ahead and behind the trough. This leads to a zonal cross section of the climatologically average AEW, looking from the south into the wave. Additionally, below the 3-D composites, the estimated rainfall (based on GPM IMERG satellite retrievals) and the ERA5-modeled rainfall are shown. The latter one is an 18-hour forecast initialized twice a day (06 UTC, 18 UTC) with hourly data available. The lead times from +6 h to +18 h are concatenated to avoid model spin-up effects. Then, both the forecast's data and satellite retrieval's data are centered to a 6-hourly time window around the given wave trough time to get an estimate of the rainfall rates at a given location by both data sources.

Over land, the PV column exhibits a noticeable downshear tilt (the lower tropospheric shear vector points westward due to the AEJ), whereas over the ocean, the PV column appears upright in relation to this cross section. Centered around the wave trough, a clear dipole structure in the meridional wind can be observed, with a maximum around 650-700 hPa, falling in line with the maximum intensity of the AEJ [e.g., 21]. Moreover, over land, high-PV values extend higher into the troposphere than over the ocean. The composite structure over land is reminiscent of the deep convection that occurs ahead of the trough [47]. This is corroborated by the ERA5 model rainfall that peaks slightly ahead of the trough. Over land, clear differences occur, with higher estimated rainfall intensities well ahead of the trough line. As discussed in Fink and Reiner [47], squall lines are initiated ahead of the trough but move about twice as fast as the trough into the preceding ridge where they tend to dissipate. This is one potential explanation of the rainfall intensities increasing west of the trough in the composite. Over the ocean, the PV tower is shallower and centered on the trough. The shallower PV object over the ocean is related to less-deep convection over the ocean due to lower CAPE values. The largest model rainfall is found at and to the east of the trough axis. This is consistent with an AEW evolution described in Riehl [158] and Russell et al. [161]. The satellite-based rainfall estimations match the model rainfall in phase and amplitude quite well over the ocean.

In Figure 6.7, a similar composite is presented from another viewing angle. It shows the latitude-pressure cross section perspective, viewing from east to west into the PV anomalies. To get a composite of PV structures independent of their current latitude, we use here the computed center of the ellipsoid as reference latitude in the center (red line). Areas to the left indicate locations south of the center of the PV anomaly, and areas to the right denote areas north. This view validates the PV feature locations, aligning well with the red reference line. Over land, the PV column is upright, extending higher in the vertical, signifying intense deep convection. Over the ocean, anomalous PV to the north of the features can be observed, especially in the 600–700 hPa range, which coincides with the peak intensity zone of the AEJ. The contours in the figure distinctly mark the core of the AEJ just north of the wave centers. This lets us suggest that this PV anomaly in the composite can be traced back to PV advection taking place from tropical West Africa to the Atlantic Ocean.

The ERA5-modeled precipitation and rainfall estimated by satellite retrievals further match with the interpretation from the first composite. Both data sets reveal good agreement over the ocean, while significant differences appear over land. Centered around the trough, much less precipitation is observed. Following the argument by Fink and Reiner [47], squall lines tend to move faster than the trough and, therefore, move out of the displayed latitude-pressure cross section.

Figure 6.8a shows boxplots depicting the orientation of PV features along the longitude-pressure plane. As described in Sect. 6.4.6, projecting a 3-D ellipsoid results in an ellipse, which is defined by two main axes. The more vertically oriented axis is used to calculate the angle of the PV feature relative to the pressure axis. The plot clearly illustrates distinct orientation patterns of PV features depending on their location. Over land, the features exhibit a downshear tilt, while over the ocean, the PV column appears upright, aligning with the composite shown in Fig. 6.6, validating the structure of the identified ellipsoids. Figure 6.8b showcases similar boxplots for the latitude-pressure plane, which can be visualized by observing the PV features looking from east to west. While the longitude-pressure cross section is visible in Fig. 6.6 and has been studied in the literature [e.g., 161], the latitude-pressure cross section (composite in Fig. 6.7) remains relatively unexplored. Here, a clear tilt pattern is evident as well, which differs between ocean and land. Over the Atlantic Ocean, the PV column tilts to the south, while it appears more upright over land, as also evident in the composite. As explained earlier, we propose that this southern tilt results from PV advection along the AEJ, inducing PV anomalies on the northern flank of the waves at lower levels and creating this asymmetry. Further research could explore the decoupling of PV sources to validate this advection process, distinguishing advection from diabatically generated PV, similar to Tomassini et al. [185].

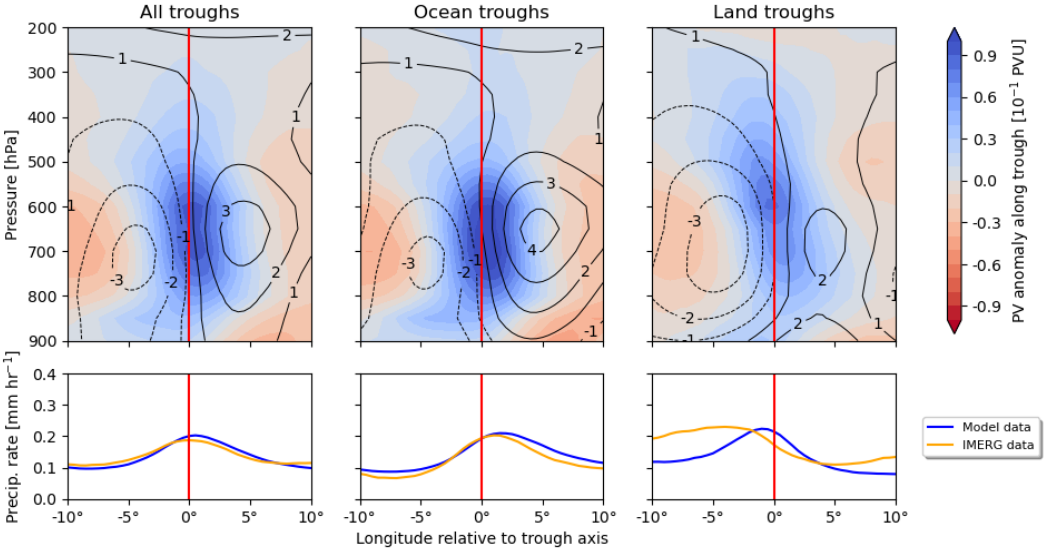


Fig. 6.6: Composite of PV anomalies relative to the identified wave troughs along a longitude-pressure cross section from June–October 2002–2022 based on the ERA5 reanalysis. The red line indicates the longitude of the wave trough, with regions to the left denoting positions ahead (west) of the trough and those to the right indicating positions behind (east) the trough. Line contours denote the mean meridional wind around the identified troughs in ms^{-1}. What is shown are the following: panel (a) focuses on all identified wave troughs which are part of the tracks, panel (b) focuses on the subset of troughs over the North Atlantic Ocean, and panel (c) focuses only on wave troughs over West Africa. In the lower panels, 6-hourly centered satellite-estimated (GPM IMERG) and model-predicted (ERA5 short-range forecasts) rainfall along this cross section is visualized.

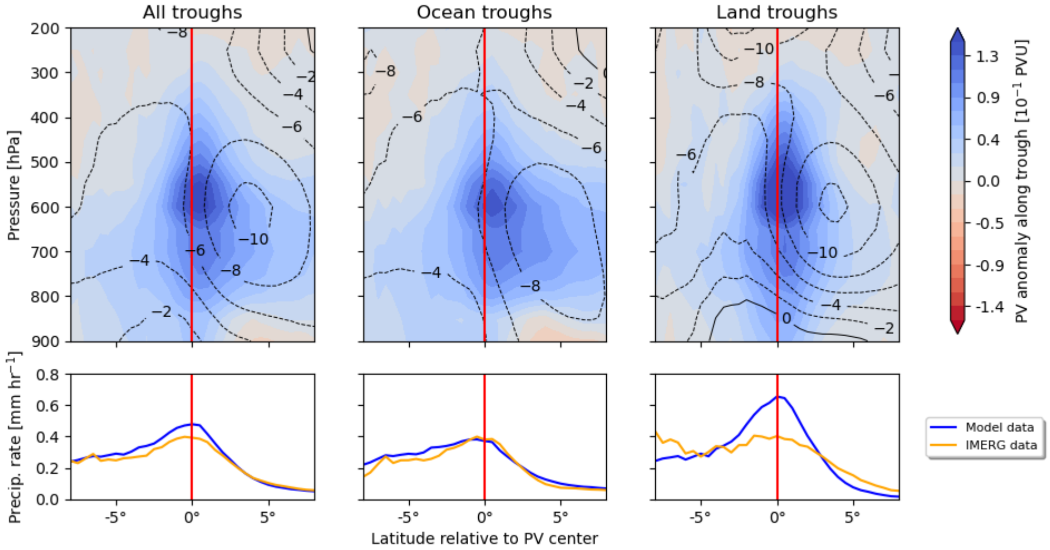


Fig. 6.7: Similar to Fig. 6.6, this composite displays the latitude-pressure cross section for the same data set, broken down in the same fashion with the same precipitation data. The contours indicate the mean zonal wind in ms^{-1}. The red line denotes the image-moment-based latitude center of the identified PV anomalies, as outlined in Sect. 6.4.6.

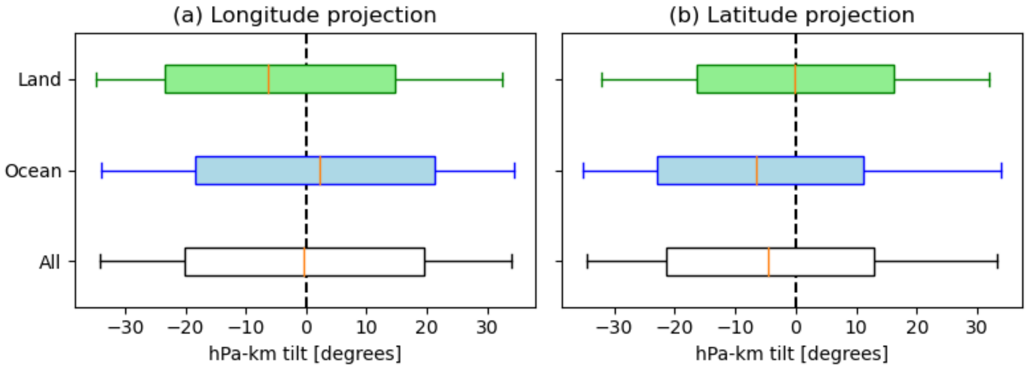


Fig. 6.8: Orientation of projected ellipsoids on the (a) longitude-pressure plane and the (b) latitude-pressure plane. Box plots depict the orientation split by location: *land* contains ellipsoids centered over West Africa, *ocean* contains ellipsoids over the North Atlantic Ocean, and *all* indicates all identified features. Positive angles indicate a tilt towards higher longitude (latitude) values with height.

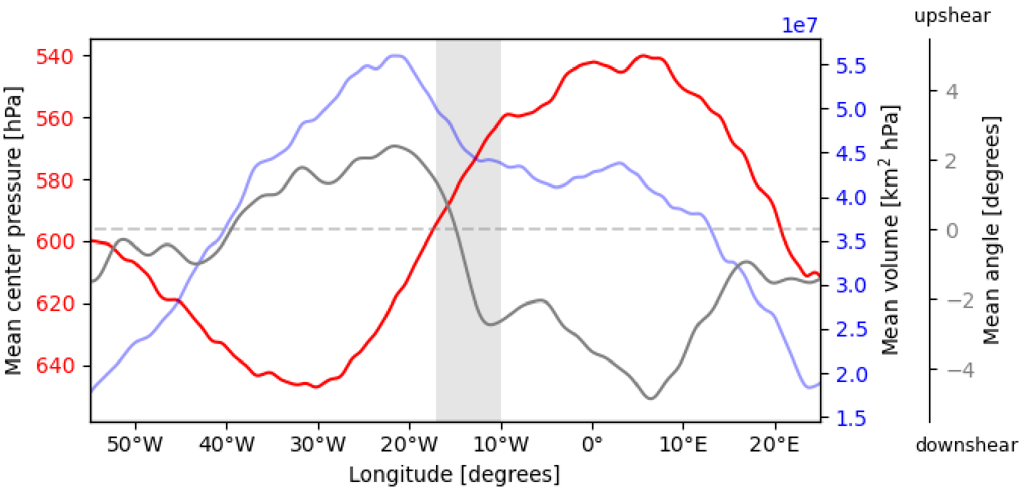
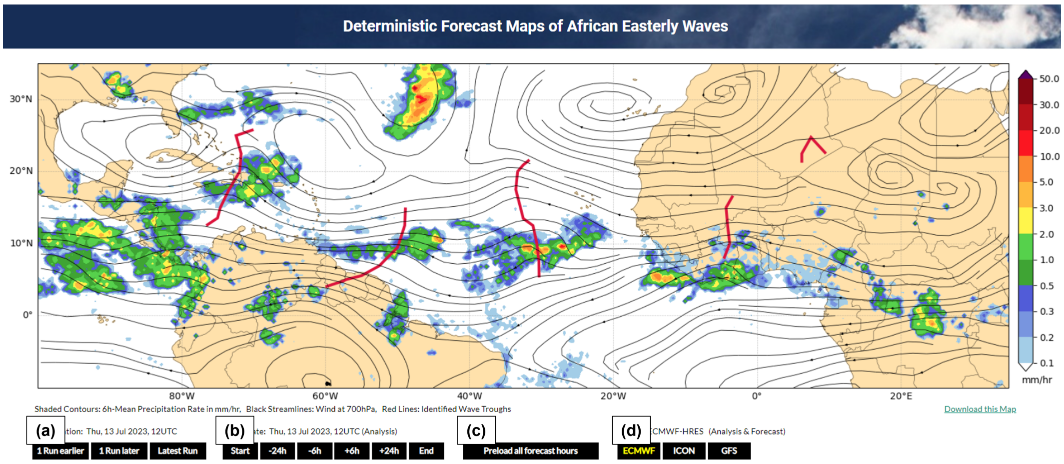


Fig. 6.9: Graph showing the mean pressure level (in red), volume (in blue), and angle along the longitude-pressure plane relative to the vertical (in grey) for all identified PV features from June to October, spanning 2002 to 2022, plotted against longitude. The horizontal grey line marks the zero angle, indicating a perfectly vertical orientation. The grey area highlights the region where features transition from land to ocean within our study domain.

Figure 6.9 shows key characteristics of identified PV features within the study area split by longitude, more specifically their mean level and volume. As the AEWs progress from their genesis region towards 0°W, there is a noticeable upward trend in the mean level of the PV features. This upward trend in altitude suggests the evolution of the convective structure associated with the wave, evolving into deep convective structures. Subsequently, beyond 0°W, the mean level gradually decreases until approximately 40°W, indicating a descent of the diabatic heating centers in the AEWs and a shift towards lower atmospheric levels. This is consistent with the composite figures and reflects less-deep convection over the ocean. Changes in PV feature volume mirror the evolution of convection and diabatic processes. Peak volumes correspond to the mature phase of convective activity, aligning with the deep convection phase in these regions [113]. The orientation of the features depicted by the grey line also confirms the findings from the boxplots in Fig. 6.8. An abrupt change in orientation is visible across the land–ocean transition zone, which shows the different dynamical behaviors in these contrasting environments.



Shaded Contours: 6h-Mean Precipitation Rate in mm/hr, Black Streamlines: Wind at 700hPa, Red Lines: Identified Wave Troughs

tion: Thu, 13 Jul 2023, 12UTC

te: Thu, 13 Jul 2023, 12UTC (Analysis)

CMWF-HRES (Analysis & Forecast)

Fig. 6.10: Screenshot of the Near-Real Time web page accessible at `www.kit-weather.de`, taken at 14 Jul 2023, 00 UTC. The user can switch between (a) different forecast initialization times and (b) different time steps, (c) download all time steps in advance, and (d) switch between different forecasting models.

6.6 Near-Real Time presentation

The website `https://www.kit-weather.de`[3] hosts various visualization products developed by members of the Waves to Weather consortium. Within the scope of this thesis, a near-real-time product has been implemented to display identified and tracked AEWs. This product is updated twice daily using data from ECMWF, GFS, and ICON forecasts. The generated plots provide a current snapshot of AEWs over southern West Africa and the Atlantic basin. This web display serves as a valuable tool for meteorologists and enthusiasts, offering insights for educational and forecasting purposes. Notably, it is actively utilized by African forecasting agencies, including the African Center of Meteorological Applications for Development (ACMAD).

6.7 Summary & Conclusion

In this study, we have introduced a novel identification and tracking strategy for 3-D PV features within AEWs, allowing for in-depth analyses of their characteristics and statistical properties. Our algorithm builds on a robust identification and tracking of AEW troughs in 2-D at 700 hPa. Identified wave

---

[3] last accessed 22 May 2024

troughs and tracks for the entire ERA5 reanalysis from 1940–2022 are provided, along with the framework for identifying and tracking meteorological features (see the Code and data availability section in the appendix). The framework offers support for parallel execution and supports output in JSON or Protobuf formats. Furthermore, a near-real-time web page displays wave troughs for ECMWF, GFS, and ICON forecasts. These 2-D wave troughs, combined with the computed wave phase at every point in the study domain, serve as input for the 3-D extension of our strategy where we extract PV features associated with specific AEWs. Within the identified trough area, we collect and process PV in the column above to create cohesive structures. These PV features are further represented using best-fitting ellipsoids, enabling statistical analyses.

An evaluation of the feature climatology demonstrates the robustness and relevance of our approach. The occurrence, extent, and orientation of the identified PV features closely align with established expert knowledge. By focusing our composite analysis on PV across longitude-pressure and latitude-pressure cross sections from a trough-centric perspective, we achieved a consistent analysis framework that is not influenced by the seasonal north-south migration of the waves. This climatological review, spanning June to October for the years 2002 to 2022, validates our feature identification approach, particularly highlighting the concurrence between the orientation of PV features in our analysis and existing climatological data.

Additionally, we undertook a comparative analysis of these PV features with estimated (GPM IMERG) and NWP-simulated (ERA5 short-range forecast) rainfall data. Notable differences in the spatial distribution and intensity of rainfall between the simulated and the estimated data can be observed. These disparities are primarily attributed to variations in the movement of squall lines over tropical West Africa, where satellite estimates show a much faster progression of these lines. A prominent tilt of PV features over tropical West Africa and over the North Atlantic Ocean can be observed both in the composites and in the statistical evaluation of the main axes. An identified southern tilt of these features over the Atlantic suggests PV advection along the AEJ originating from the convective activity over northern Africa. Along with additional feature descriptors, like mean level and volume of the PV feature, the life cycle of these waves can be investigated, originating as small disturbance to deep moist convection over Africa and towards less-deep convection over the Atlantic Ocean.

The data generated through our strategy have diverse applications. Besides the generation of feature climatologies, they can be utilized as input for statistical forecasting techniques. Links between PV features and TC activity can be explored using the features as input to statistical or mixed statistical-dynamical models [111], as well as for forecasts of tropical rainfall over southern West Africa [192]. Furthermore, our strategy enables detailed case studies and facilitates in-depth investigations of AEWs. One particular focus here could be on the relation between the 3-D PV feature evolution and the propensity of an AEW to undergo TC genesis [e.g., 35, 130]. In all these applications, an important aspect is the sensitivity of the gross characteristics of PV features to different reanalyses or models because the PV features can be expected to be strongly influenced by the representation of convection. In this context, differences in models with parameterized and explicit convection are considered most worthy of further study. In combination with state-of-the-art interactive 3-D visualization techniques [149], objective features also open the door for comprehensive case studies.

In conclusion, this study significantly contributes to the field of AEW research by providing a tool to comprehensively analyze PV features within AEWs and, thus, to make the PV perspective more easily applicable to AEWs. The identification and tracking strategy, along with the availability of comprehensive 3-D data, opens up new avenues for studying individual cases, understanding AEW dynamics, improving forecasts, and gaining a deeper understanding of the role of AEWs in the weather and climate of tropical West Africa and the North Atlantic Ocean.

Conclusion

7

> *Visualization gives you answers to questions you didn't know you had.*
>
> — **Ben Shneiderman**
> (Computer Scientist)

In this dissertation, we explored the convergence of computer science and meteorology. Our aim was to create innovative tools for examining atmospheric waves, driven by the belief that integrating different scientific disciplines can lead to novel and promising developments. We devised methods that enhance the understanding of atmospheric wave phenomena through sophisticated identification, tracking, and climatological analysis.

At the heart of our work was the establishment and implementation of a versatile framework for the identification and tracking of features within meteorological data. This framework has been tailored to the needs of meteorologists to analyze complex atmospheric phenomena. It enables the transformation of complex atmospheric patterns into simpler, low-dimensional feature vectors, making these dynamics easier to study. Additionally, we developed innovative approaches for identifying potential vorticity anomalies at the tropopause and within African Easterly Waves, creating a vital link between theoretical atmospheric science and practical data analysis.

Achieved Goals

Framework Development

A notable success of this research was the creation of a comprehensive framework that provides a universal method for identifying and tracking atmospheric features. Its flexibility allows it to be applied to a variety of atmospheric phenomena beyond the specific case studies addressed. The framework excels in transforming complex atmospheric data into clear and manageable low-dimensional representations. Developed with an emphasis on user-friendliness, modularity, and ease of prototyping, the framework proved effective in two primary studies, which have been published in peer-reviewed journals.

Bridging the gap from computer science to meteorology

In Chapter 2, we laid the groundwork for feature identification and tracking from a conceptual perspective. We introduced the tracking graph, which forms the backbone of our tracking methodology. This chapter bridges the gap from the theoretical definitions in computer science, such as morphological operations and eigenanalysis of features, to practical applications in meteorology. Challenges such as distortions in projections and interpretability in a meteorological context, which are not addressed by purely theoretical approaches in computer science, are tackled to ensure the applicability of these methods to atmospheric data. Subsequently, Chapter 3 introduces the meteorological concepts used in the presented applications of the framework.

Application to PV Streamers

The first study introduced a novel approach to identify PV anomalies along the tropopause. Utilizing image processing techniques adapted to the requirements of meteorological data, the method successfully interprets disturbances in multi-dimensional fields. The unique contribution of this study is its strategy for calculating distances within the chosen conformal projection, which is particularly effective in examining the elongated nature of 2-D PV streamers and their 3-D counterparts. This concept can also be applied to other anomaly-type structures in the atmosphere. This study provided a new look at the tropopause and its complexity in 3-D. It enabled us to generalize the previously established 2-D view of the anomalies to a 3-D view, allowing the identification and analysis of these features over their entire life-time.

Application to AEWs

The second study focused on African Easterly Waves and involved the identification and tracking of these complex atmospheric entities along with their associated 3-D PV features. By building on a robust foundation of 2-D AEW trough identification and tracking, this study both adds new value to existing 2-D analyses in literature, and extending the analysis of 2-D wave troughs in AEWs to a 3-D perspective of PV features.

A Hilbert transform was used to locate PV features within AEWs, serving as proxies for convective activity. To ensure more accurate and reliable detection of these features, morphological operators were applied, enhancing the robustness of the identification process. The research confirmed the effectiveness of low-dimensional geometric descriptions for comprehensive analysis, with feature vector statistics aligning well with previous studies. This validation supports the use of low-dimensional descriptors as statistical inputs for various prediction models. Results from the 2-D identification and tracking are publicly available for the ERA5 data set. Additionally, analysis of the most recent forecasts are accessible on an automatically updated near-real-time web page, offering practical tools for both research and operational forecasting.

Outlook

Practical Applications and Future Extensions

The framework has already seen application in numerous research projects and shows great potential for future use. Its straightforward interface and capability for parallelization set the stage for future advancements, such as incorporating enhanced visualization modules like feature graph representation and expanding testing and analysis functionalities. Beyond feature detection, the framework offers a robust environment for swift prototyping and processing of meteorological data, which could greatly benefit a diverse range of atmospheric researchers.

Impact on Meteorological Research

The peer-reviewed studies in Chapters 5 and 6 offer novel insights into the 3-D structure of PV anomaly features along the tropopause and within AEWs. For a detailed summary of the findings, please refer to the conclusions of the respective chapters. The low-dimensional descriptions of these features have demonstrated robustness and reliability in statistical analyses, providing a strong foundation for future research.

These low-dimensional descriptors can be utilized in statistical analyses to correlate the identified features with other atmospheric phenomena or weather events. For example, the AEW data can be used to investigate the relationships between AEWs, tropical rainfall, and tropical cyclone genesis. This can lead to more accurate and reliable weather predictions. PV anomalies along the tropopause are known to be linked to pressure systems near the surface, shaping our weather. Using the novel feature descriptors, this connection be evaluated statistically, too.

Conclusion

In conclusion, this dissertation has successfully bridged the gap between computer science and meteorology by proposing a versatile methodology for analyzing atmospheric wave features. By overcoming the challenges inherent in integrating these disciplines, this research has not only opened up new avenues for further study but has also underscored the immense potential that lies at their intersection.

Through collaborative efforts across disciplines, this work aligns with the increasing recognition within the scientific community of the value of innovative opportunities. The introduction of 3-D visualization techniques has not only offered fresh insights into complex atmospheric data but has also underscored the essential of visualization in uncovering new perspectives on longstanding questions. By incorporating image processing methods and pixel-based interpretations in feature detection and analysis, this study has demonstrated a robust, effective, and elegant approach to building automatic detection and analysis systems.

In essence, this interdisciplinary endeavor paves the way for future scientific exploration by showcasing the benefits of venturing into uncharted territory on the interface between multiple sciences, as Norbert Wiener pointed out in his quote that started this thesis. It serves as a reminder to upcoming researchers to explore the vast possibilities at the crossroads of diverse fields and to continue the quest for knowledge in these promising realms.

Bibliography

[1]J. K. Aggarwal, L. S. Davis, and W. N. Martin. “Correspondence processes in dynamic scene analysis”. In: *Proceedings of the IEEE* 69.5 (1981), pp. 562–572. DOI: `10.1109/PROC.1981.12025` (cit. on p. 18).

[2]J. K. Aggarwal and N. Nandhakumar. “On the computation of motion from sequences of images-A review”. In: *Proceedings of the IEEE* 76.8 (1988), pp. 917–935. DOI: `10.1109/5.5965` (cit. on p. 17).

[3]V. Akyildiz. “Systematic errors in the behaviour of cyclones in the ECMWF operational models”. In: *Tellus A* 37A.4 (1985), pp. 297–308. DOI: `10.1111/j.1600-0870.1985.tb00429.x`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1600-0870.1985.tb00429.x` (cit. on p. 17).

[6]J. A. Añel, D. R. Allen, G. Sáenz, L. Gimeno, and L. de la Torre. “Equivalent latitude computation using regions of interest (ROI)”. In: *PLOS ONE* 8.9 (Sept. 2013), pp. 1–8. DOI: `10.1371/journal.pone.0072970` (cit. on p. 78).

[7]R. Bader, M. Sprenger, N. Ban, et al. “Extraction and Visual Analysis of Potential Vorticity Banners around the Alps”. In: *IEEE T. Vis. Comput. Gr.* (2019), pp. 1–1. DOI: `10.1109/TVCG.2019.2934310` (cit. on p. 74).

[8]C. L. Bain, K. D. Williams, S. F. Milton, and J. T. Heming. “Objective tracking of African Easterly Waves in Met Office models”. In: *Q. J. R. Meteorolog. Soc.* 140.678 (2014), pp. 47–57. DOI: `10.1002/qj.2110`. eprint: `https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.2110` (cit. on pp. 113, 115).

[9]M. P. Baldwin and T. J. Dunkerton. “Stratospheric Harbingers of Anomalous Weather Regimes”. In: *Science* 294.5542 (2001), pp. 581–584. DOI: `10.1126/science.1063315`. eprint: `https://www.science.org/doi/pdf/10.1126/science.1063315` (cit. on p. 38).

[10]A. A. Beckert, L. Eisenstein, A. Oertel, et al. “The three-dimensional structure of fronts in mid-latitude weather systems in numerical weather prediction models”. In: *Geosci. Model Dev.* 16.15 (2023), pp. 4427–4450. DOI: `10.5194/gmd-16-4427-2023` (cit. on pp. 12, 53).

[11]J. I. Belanger, M. T. Jelinek, and J. A. Curry. “A climatology of easterly waves in the tropical Western Hemisphere”. In: *Geosci. Data J.* 3.2 (2016), pp. 40–49. DOI: `10.1002/gdj3.40` (cit. on pp. 18, 25, 113–115, 117, 118).

[12]D. A. Bennetts and B. J. Hoskins. “Conditional symmetric instability - a possible explanation for frontal rainbands”. In: *Q. J. R. Meteorolog. Soc.* 105.446 (1979), pp. 945–962. DOI: `https://doi.org/10.1002/qj.49710544615`. eprint: `https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.49710544615` (cit. on p. 75).

[13]A. Bentley, L. Bosart, and D. Keyser. “Upper-Tropospheric Precursors to the Formation of Subtropical Cyclones that Undergo Tropical Transition in the North Atlantic Basin”. In: *Mon. Weather Rev.* 145 (2017), pp. 503–520. DOI: `10.1175/MWR-D-16-0263.1` (cit. on p. 73).

[14]G. J. Berry and C. Thorncroft. “Case Study of an Intense African Easterly Wave”. In: *Mon. Weather Rev.* 133.4 (2005), pp. 752–766. DOI: `10.1175/MWR2884.1` (cit. on pp. 51, 111).

[15]G. J. Berry, C. Thorncroft, and T. Hewson. “African Easterly Waves during 2004—Analysis Using Objective Techniques”. In: *Mon. Weather Rev.* 135.4 (2007), pp. 1251–1267. DOI: `10.1175/MWR3343.1` (cit. on p. 113).

[16]M. Bithell, L. J. Gray, and B. D. Cox. “A Three-Dimensional View of the Evolution of Midlatitude Stratospheric Intrusions”. In: *J. Atmos. Sci.* 56.5 (1999), pp. 673–688. DOI: `10.1175/1520-0469(1999)056<0673:ATDVOT>2.0.CO;2` (cit. on pp. 74, 75, 95, 106).

[17]V. Bjerknes. *On the dynamics of the circular vortex: with applications to the atmosphere and atmospheric vortex and wave motions*. Vol. 2. 4. I kommission hos Cammermeyers bokhandel, 1921 (cit. on p. 12).

[18]H. B. Bluestein. *Synoptic-Dynamic Meteorology in Midlatitudes: Vol II Observations and Theory of Weather Systems*. New York: Oxford University Press, 1993 (cit. on p. 45).

[19]A. Brammer and C. D. Thorncroft. “Variability and Evolution of African Easterly Wave Structures and Their Relationship with Tropical Cyclogenesis over the Eastern Atlantic”. In: *Mon. Weather Rev.* 143.12 (2015), pp. 4975–4995. DOI: `10.1175/MWR-D-15-0106.1` (cit. on p. 115).

[20]H. W. Brandes. *Dissertatio physica de repentinis variationibus in pressione atmosphaerae observatis*. Vol. 1. Staritzii, 1826 (cit. on p. 32).

[21]R. W. Burpee. “The Origin and Structure of Easterly Waves in the Lower Troposphere of North Africa”. In: *J. Atmos. Sci.* 29.1 (1972), pp. 77–90. DOI: `10.1175/1520-0469(1972)029<0077:TOASOE>2.0.CO;2` (cit. on pp. 10, 50, 51, 112, 127).

[22]S. J. Camargo and S. E. Zebiak. “Improving the Detection and Tracking of Tropical Cyclones in Atmospheric General Circulation Models”. In: *Weather Forecasting* 17.6 (2002), pp. 1152–1162. DOI: `10.1175/1520-0434(2002)0172.0.CO;2` (cit. on p. 22).

[23]J. G. Charney. “The dynamics of long waves in a baroclinic weserly current”. In: *J. Atmos. Sci.* 4.5 (1947), pp. 136–162. DOI: `10.1175/1520-0469(1947)0042.0.CO;2` (cit. on p. 46).

[24]J. G. Charney and M. E. Stern. “On the Stability of Internal Baroclinic Jets in a Rotating Atmosphere”. In: *J. Atmos. Sci.* 19.2 (1962), pp. 159–172. DOI: `10.1175/1520-0469(1962)0192.0.CO;2` (cit. on p. 51).

[25]L. D. Cohen and R. Kimmel. “Global minimum for active contour models: A minimal path approach”. In: *Int. J. Comput. Vision* 24.1 (1997), pp. 57–78. DOI: `10.1023/A:1007922224810` (cit. on pp. 82, 83).

[26]M. X. Cohen. *Practical Linear Algebra for Data Science*. O’Reilly Media, 2022 (cit. on p. 28).

[27]K. H. Cook. “Generation of the African Easterly Jet and Its Role in Determining West African Precipitation”. In: *J. Clim.* 12.5 (1999), pp. 1165–1184. DOI: `10.1175/1520-0442(1999)0122.0.CO;2` (cit. on p. 50).

[29]J. Crétat, E. Vizy, and K. Cook. “The relationship between African easterly waves and daily rainfall over West Africa: observations and regional climate simulations”. In: *Clim. Dyn.* 44 (2015), pp. 385–404. DOI: `10.1007/s00382-014-2120-x` (cit. on p. 111).

[30]J. T. Dawe and P. H. Austin. “Statistical Analysis of an LES Shallow Cumulus Cloud Ensemble Using a Cloud Tracking Algorithm”. In: *Atmos. Chem. Phys.* 12.2 (2012), pp. 1101–1119. DOI: `10.5194/acp-12-1101-2012` (cit. on p. 73).

[32]“Feature Point Matching Using Temporal Smoothness in Velocity”. In: *Pattern Recognition Theory and Applications*. Ed. by P. A. Devijver and J. Kittler. Berlin, Heidelberg: Springer Berlin Heidelberg, 1987, pp. 119–131 (cit. on p. 18).

[33]A. Diedhiou, S. Janicot, A. Viltard, P. De Felice, and H. Laurent. “Easterly wave regimes and associated convection over West Africa and tropical Atlantic: Results from the NCEP/NCAR and ECMWF reanalyses”. In: *Clim. Dyn.* 15 (1999), pp. 795–822. DOI: `10.1007/s003820050316` (cit. on p. 112).

[34]E. W. Dijkstra et al. “A note on two problems in connexion with graphs”. In: *Numer. Math.* 1.1 (1959), pp. 269–271. DOI: `10.1007/BF01386390` (cit. on p. 82).

[35]T. J. Dunkerton, M. T. Montgomery, and Z. Wang. “Tropical cyclogenesis in a tropical wave critical layer: easterly waves”. In: *Atmos. Chem. Phys.* 9.15 (2009), pp. 5587–5646. DOI: `10.5194/acp-9-5587-2009` (cit. on pp. 112, 134).

[36]E. T. Eady. “Long Waves and Cyclone Waves”. In: *Tellus* 1.3 (1949), pp. 33–52. DOI: `10.3402/tellusa.v1i3.8507`. eprint: `https://doi.org/10.3402/tellusa.v1i3.8507` (cit. on p. 46).

[37]L. Eisenstein, B. Schulz, G. A. Qadir, J. G. Pinto, and P. Knippertz. “Identification of high-wind features within extratropical cyclones using a probabilistic random forest – Part 1: Method and case studies”. In: *Weather Clim. Dyn.* 3.4 (2022), pp. 1157–1182. DOI: `10.5194/wcd-3-1157-2022` (cit. on pp. 24, 53).

[38]A. Eliassen. “The Charney—Stern Theorem on Barotropic—Baroclinic Instability”. In: *Instabilities in Continuous Media*. Ed. by L. Knopoff, V. I. Keilis-Borok, and G. Puppi. Basel: Birkhäuser Basel, 1985, pp. 563–572. DOI: `10.1007/978-3-0348-6608-8_11` (cit. on p. 51).

[39]A. Eliassen and E. Kleinschmidt. “Dynamic Meteorology”. In: *Geophysik II / Geophysics II*. Ed. by J. Bartels. Berlin, Heidelberg: Springer Berlin Heidelberg, 1957, pp. 1–154. DOI: `10.1007/978-3-642-45881-1_1` (cit. on p. 45).

[40]T. J. Elless and R. D. Torn. “African Easterly Wave Forecast Verification and Its Relation to Convective Errors within the ECMWF Ensemble Prediction System”. In: *Weather Forecasting* 33.2 (2018), pp. 461–477. DOI: `10.1175/WAF-D-17-0130.1` (cit. on p. 112).

[41]T. J. Elless and R. D. Torn. “Investigating the Factors That Contribute to African Easterly Wave Intensity Forecast Uncertainty in the ECMWF Ensemble Prediction System”. In: *Mon. Weather Rev.* 147.5 (2019), pp. 1679–1698. DOI: `10.1175/MWR-D-18-0071.1` (cit. on p. 112).

[42]R. M. Endlich, D. E. Wolf, D. J. Hall, and A. E. Brain. “Use of a Pattern Recognition Technique for Determining Cloud Motions from Sequences of Satellite Photographs”. In: *J. Appl. Meteorol. Climatol.* 10.1 (1971), pp. 105–117. DOI: `10.1175/1520-0450(1971)0102.0.CO;2` (cit. on p. 17).

[43]T. Engel, A. H. Fink, P. Knippertz, G. Pante, and J. Bliefernicht. “Extreme Precipitation in the West African Cities of Dakar and Ouagadougou: Atmospheric Dynamics and Implications for Flood Risk Assessments”. In: *J. Hydrometeorol.* 18.11 (2017), pp. 2937–2957. DOI: `10.1175/JHM-D-16-0218.1` (cit. on p. 111).

[44]H. Ertel. “Ein neuer hydrodynamischer Erhaltungssatz”. In: *Naturwissenschaften* 30.36 (1942), pp. 543–544. DOI: `10.1007/BF01475602` (cit. on pp. 43, 44).

[45]Z. Feng, J. Hardin, H. C. Barnes, et al. “PyFLEXTRKR: a flexible feature tracking Python software for convective cloud analysis”. In: *Geosci. Model Dev.* 16.10 (2023), pp. 2753–2776. DOI: `10.5194/gmd-16-2753-2023` (cit. on p. 117).

[46]A. H. Fink, T. Engel, V. Ermert, et al. “Mean Climate and Seasonal Cycle”. In: *Meteorology of Tropical West Africa*. John Wiley & Sons, Ltd, 2017. Chap. 1, pp. 1–39. DOI: `10.1002/9781118391297.ch1`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118391297.ch1` (cit. on pp. 49, 114, 117, 126).

[47]A. H. Fink and A. Reiner. “Spatiotemporal variability of the relation between African Easterly Waves and West African Squall Lines in 1998 and 1999”. In: *J. Geophys. Res.: Atmos.* 108.D11 (2003). DOI: `10.1029/2002JD002816` (cit. on pp. 50, 51, 111, 113, 127, 128).

[48]C. Fischer, A. H. Fink, E. Schömer, M. Rautenhaus, and M. Riemer. “An objective identification technique for potential vorticity structures associated with African easterly waves”. In: *Geosci. Model Dev.* 17.10 (2024), pp. 4213–4228. DOI: `10.5194/gmd-17-4213-2024` (cit. on pp. 3, 24, 25, 53, 109).

[49] C. Fischer, A. H. Fink, E. Schömer, et al. “A novel method for objective identification of 3-D potential vorticity anomalies”. In: *Geosci. Model Dev.* 15.11 (2022), pp. 4447–4468. DOI: `10.5194/gmd-15-4447-2022` (cit. on pp. 3, 24, 71, 111, 122).

[50] C. Fischer. *enstools-feature: A framework for identification and tracking of meteorological features*. Nov. 2023. DOI: `10.5281/zenodo.10062273` (cit. on p. 169).

[51] C. Fischer, A. H. Fink, E. Schömer, M. Rautenhaus, and M. Riemer. *AEWDAT - African Easterly Waves: Identified and Tracked Wave Troughs*. Zenodo, Oct. 2023. DOI: `10.5281/zenodo.8403744` (cit. on pp. 12, 169).

[52] C. Fischer, A. H. Fink, E. Schömer, M. Rautenhaus, and M. Riemer. *Identified Potential Vorticity Structures associated with African Easterly Waves*. Zenodo, Nov. 2023. DOI: `10.5281/zenodo.10061471` (cit. on p. 169).

[53] C. Fischer, M. Rautenhaus, A. H. Fink, et al. *A novel method for objective identification of 3-D potential vorticity anomalies - Visualizations using Met.3D*. Nov. 2021. DOI: `10.5281/zenodo.5639001` (cit. on pp. 99, 101, 102, 165).

[54] R. Fjørtoft. *Application of integral theorems in deriving criteria of stability for laminar flows and for the baroclinic circular vortex*. Vol. 17. Grøndahl & søns boktr., I kommisjon hos Cammermeyers boghandel Oslo, 1950 (cit. on p. 51).

[55] G. Fragkoulidis and V. Wirth. “Local Rossby Wave Packet Amplitude, Phase Speed, and Group Velocity: Seasonal Variability and Their Role in Temperature Extremes”. In: *J- Clim.* 33.20 (2020), pp. 8767–8787. DOI: `10.1175/JCLI-D-19-0377.1` (cit. on p. 53).

[56] J. Furtney. *scikit-fmm*. 2019 (cit. on p. 98).

[57] T. J. Galarneau Jr., R. McTaggart-Cowan, L. F. Bosart, and C. A. Davis. “Development of North Atlantic tropical disturbances near upper-level potential vorticity streamers”. In: *J. Atmos. Sci.* 72.2 (2015), pp. 572–597. DOI: `10.1175/JAS-D-14-0106.1` (cit. on p. 73).

[58] F. Galton. *Meteorographica, or, methods of mapping the weather: Illustrated by upwards of 600 printed and lithographed diagrams referring to the weather of a large part of Europe, during the month of December 1861*. Macmillan, 1863 (cit. on p. 11).

[59] S. Giraldo-Cardenas, P. A. Arias, S. C. Vieira, and M. D. Zuluaga. “Easterly waves and precipitation over northern South America and the Caribbean”. In: *Int. J. Climatol.* 42.3 (2022), pp. 1483–1499. DOI: `10.1002/joc.7315` (cit. on p. 126).

[60] G. H. Golub and C. F. Van Loan. *Matrix computations*. JHU press, 2013 (cit. on p. 29).

[62] J. P. Grist. “Easterly Waves over Africa. Part I: The Seasonal Cycle and Contrasts between Wet and Dry Years”. In: *Mon. Weather Rev.* 130.2 (2002), pp. 197–211. DOI: `10.1175/1520-0493(2002)130<0197:EWOAPI>2.0.CO;2` (cit. on p. 51).

[63] G. Gu, R. F. Adler, G. J. Huffman, and S. Curtis. “African easterly waves and their association with precipitation”. In: *J. Geophys. Res.: Atmos.* 109.D4 (2004). DOI: `https://doi.org/10.1029/2003JD003967`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2003JD003967` (cit. on p. 117).

[64] M. Haile. “Weather patterns, food security and humanitarian response in sub-Saharan Africa”. In: *Philos. Trans. R. Soc. London, Ser. B* 360.1463 (2005), pp. 2169–2182. DOI: `10.1098/rstb.2005.1746`. eprint: `https://royalsocietypublishing.org/doi/pdf/10.1098/rstb.2005.1746` (cit. on p. 52).

[65] N. M. J. Hall, G. N. Kiladis, and C. D. Thorncroft. “Three-Dimensional Structure and Dynamics of African Easterly Waves. Part II: Dynamical Modes”. In: *J. Atmos. Sci.* 63.9 (2006), pp. 2231–2245. DOI: `10.1175/JAS3742.1` (cit. on pp. 51, 111).

[66] H. L. Hamilton, K. M. Núñez Ocasio, J. L. Evans, G. S. Young, and J. D. Fuentes. “Topographic Influence on the African Easterly Jet and African Easterly Wave Energetics”. In: *J. Geophys. Res.: Atmos.* 125.8 (2020). e2019JD032138 2019JD032138, e2019JD032138. DOI: `10.1029/2019JD032138`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019JD032138` (cit. on p. 110).

[67] H. L. Hamilton, G. S. Young, J. L. Evans, J. D. Fuentes, and K. M. Núñez Ocasio. “The relationship between the Guinea Highlands and the West African offshore rainfall maximum”. In: *Geophys. Res. Lett.* 44.2 (2017), pp. 1158–1166. DOI: `10.1002/2016GL071170`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2016GL071170` (cit. on p. 111).

[68] L. Han, S. Fu, L. Zhao, et al. “3D Convective Storm Identification, Tracking, and Forecasting—An Enhanced TITAN Algorithm”. In: *J. Atmos. Oceanic Technol.* 26.4 (2009), pp. 719–732. DOI: `10.1175/2008JTECHA1084.1` (cit. on pp. 24, 25).

[69] J. Handwerker. “Cell tracking with TRACE3D—a new algorithm”. In: *Atmos. Res.* 61.1 (2002), pp. 15–34. DOI: `10.1016/S0169-8095(01)00100-4` (cit. on pp. 24, 53).

[70] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003 (cit. on p. 124).

[71] D. Hassell, J. Gregory, J. Blower, B. N. Lawrence, and K. E. Taylor. “A data model of the Climate and Forecast metadata conventions (CF-1.6) with a software implementation (cf-python v2.1)”. In: *Geosci. Model Dev.* 10.12 (2017), pp. 4619–4646. DOI: `10.5194/gmd-10-4619-2017` (cit. on pp. 6, 59).

[72] S. Hauser, F. Teubler, M. Riemer, P. Knippertz, and C. M. Grams. “Towards a holistic understanding of blocked regime dynamics through a combination of complementary diagnostic perspectives”. In: *Weather Clim. Dyn.* 4.2 (2023), pp. 399–425. DOI: `10.5194/wcd-4-399-2023` (cit. on p. 111).

[73]P. H. Haynes and M. E. McIntyre. “On the Evolution of Vorticity and Potential Vorticity in the Presence of Diabatic Heating and Frictional or Other Forces”. In: *J. Atmos. Sci.* 44.5 (1987), pp. 828–841. DOI: `10.1175/1520-0469(1987)0442.0.CO;2` (cit. on p. 44).

[74]T. Hengstebeck, D. Heizenreder, P. Joe, and P. Lang. “The mesocyclone detection algorithm of DWD”. In: *6th European Conference on Severe Storms, ECSS*. 2011, pp. 3–7 (cit. on p. 113).

[75]H. Hersbach, B. Bell, P. Berrisford, et al. “The ERA5 global reanalysis”. In: *Q. J. Roy. Meteor. Soc.* 146.730 (2020), pp. 1999–2049. DOI: `10.1002/qj.3803`. eprint: `https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3803` (cit. on pp. 86, 114).

[76]T. D. Hewson and H. A. Titley. “Objective Identification, Typing and Tracking of the Complete Life-Cycles of Cyclonic Features at High Spatial Resolution”. In: *Meteorol. Appl.* 17.3 (2010), pp. 355–381. DOI: `10.1002/met.204` (cit. on pp. 18, 73, 74).

[77]K. I. Hodges. “A General Method for Tracking Analysis and Its Application to Meteorological Data”. In: *Mon. Weather Rev.* 122.11 (1994), pp. 2573–2586. DOI: `10.1175/1520-0493(1994)1222.0.CO;2` (cit. on pp. 18, 53).

[78]K. I. Hodges. “Feature Tracking on the Unit Sphere”. In: *Mon. Weather Rev.* 123.12 (1995), pp. 3458–3465. DOI: `10.1175/1520-0493(1995)1232.0.CO;2` (cit. on pp. 65, 115).

[79]M. A. Hollis, R. R. McCrary, J. P. Stachnik, C. Lewis-Merritt, and E. R. Martin. “A global climatology of tropical easterly waves”. In: *Clim. Dyn.* 62 (2024), pp. 2317–2332. DOI: `10.1007/s00382-023-07025-w` (cit. on p. 115).

[80]J. R. Holton. “A Diagnostic Model for Equatorial Wave Disturbances: The Role of Vertical Shear of the Mean Zonal Wind”. In: *J. Atmos. Sci.* 28.1 (1971), pp. 55–64. DOI: `10.1175/1520-0469(1971)0282.0.CO;2` (cit. on p. 51).

[81]J. R. Holton. “An introduction to dynamic meteorology”. In: *International geophysics series* 48 (1992), pp. 1–497 (cit. on pp. 40, 42, 43).

[82]J. R. Holton, P. H. Haynes, M. E. McIntyre, et al. “Stratosphere-troposphere exchange”. In: *Rev. Geophys.* 33.4 (1995), pp. 403–439. DOI: `10.1029/95RG02097` (cit. on p. 38).

[83]B. J. Hoskins, M. E. McIntyre, and A. W. Robertson. “On the use and significance of isentropic potential vorticity maps”. In: *Q. J. R. Meteorolog. Soc.* 111.470 (1985), pp. 877–946. DOI: `10.1002/qj.49711147002`. eprint: `https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.49711147002` (cit. on pp. 43, 44, 72, 111).

[84]M.-K. Hu. “Visual pattern recognition by moment invariants”. In: *IRE Transactions on Information Theory* 8.2 (1962), pp. 179–187. DOI: `10.1109/TIT.1962.1057692` (cit. on pp. 25, 26).

[85]X. Huang, C. Hu, X. Huang, et al. “A long-term tropical mesoscale convective systems dataset based on a novel objective automatic tracking algorithm”. In: *Clim. Dyn.* 51.7 (2018), pp. 3145–3159. DOI: `10.1007/s00382-018-4071-0` (cit. on p. 24).

[86]G. J. Huffman, D. T. Bolvin, E. J. Nelkin, and J. Tan. “Integrated Multi-satellitE Retrievals for GPM (IMERG) technical documentation”. In: *NASA/GSFC Code* 612.47 (2015), p. 2019 (cit. on pp. 99, 114).

[87]A. Hunt and D. Thomas. *The Pragmatic programmer : from journeyman to master*. Boston [etc.]: Addison-Wesley, 2000 (cit. on p. 62).

[88]J. E. Jackson. *A user’s guide to principal components*. John Wiley & Sons, 2005 (cit. on p. 124).

[89]S. C. Jones. “The evolution of vortices in vertical shear. I: Initially barotropic vortices”. In: *Q. J. R. Meteorolog. Soc.* 121.524 (1995), pp. 821–851. DOI: `https://doi.org/10.1002/qj.49712152406`. eprint: `https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.49712152406` (cit. on p. 49).

[90]D. Karger, R. Motwani, and G. D. Ramkumar. “On approximating the longest path in a graph”. In: *Algorithmica* 18.1 (1997), pp. 82–98. DOI: `10.1007/BF02523689` (cit. on p. 21).

[91]M. Kern, T. Hewson, F. Sadlo, R. Westermann, and M. Rautenhaus. “Robust Detection and Visualization of Jet-stream Core Lines in Atmospheric Flow”. In: *IEEE T. Vis. Comput. Gr.* 24.1 (2018), pp. 893–902. DOI: `10.1109/tvcg.2017.2743989` (cit. on pp. 53, 74).

[92]M. Kern, T. Hewson, A. Schatler, R. Westermann, and M. Rautenhaus. “Interactive 3D Visual Analysis of Atmospheric Fronts”. In: *IEEE T. Vis. Comput. Gr.* 25.1 (Jan. 2019), pp. 1080–1090. DOI: `10.1109/TVCG.2018.2864806` (cit. on pp. 74, 76).

[93]L. G. Khachiyan. “Rounding of Polytopes in the Real Number Model of Computation”. In: *Math. Oper. Res.* 21.2 (1996), pp. 307–320. DOI: `10.1287/moor.21.2.307`. eprint: `https://doi.org/10.1287/moor.21.2.307` (cit. on p. 122).

[94]L. G. Khachiyan. “A polynomial algorithm in linear programming”. In: *Doklady Akademii Nauk*. Vol. 244. 5. Russian Academy of Sciences. 1979, pp. 1093–1096 (cit. on p. 30).

[95]G. N. Kiladis, C. D. Thorncroft, and N. M. J. Hall. “Three-Dimensional Structure and Dynamics of African Easterly Waves. Part I: Observations”. In: *J. Atmos. Sci.* 63.9 (2006), pp. 2212–2230. DOI: `10.1175/JAS3741.1` (cit. on p. 50).

[98]E. Kleinschmidt. “What is COST?” In: *Meteorologische Rundschau* 3.1 (1950) (cit. on pp. 32, 33).

[99] J. B. Klemp and R. Rotunno. “A Study of the Tornadic Region within a Supercell Thunderstorm”. In: *J. Atmos. Sci.* 40.2 (1983), pp. 359–377. DOI: https://doi.org/10.1175/1520-0469(1983)040<0359:ASOTTR>2.0.CO;2 (cit. on p. 28).

[100] K. R. Knapp, M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann. “The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying Tropical Cyclone Data”. In: *Bull. Am. Meteorol. Soc.* 91.3 (2010), pp. 363–376. DOI: 10.1175/2009BAMS2755.1 (cit. on p. 65).

[101] P. Koch, H. Wernli, and H. C. Davies. “An event-based jet-stream climatology and typology”. In: *Int. J. Climatol.* 26.3 (2006), pp. 283–301. DOI: 10.1002/joc.1255. eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.1255 (cit. on p. 94).

[102] M. Kriening. “The 3-D Potential Vorticity Development during the Tropical Transition of Hurricane Leslie (2018) and Paulette (2020) over the North Atlantic Ocean”. MA thesis. University of Heidelberg, 2022 (cit. on pp. 69, 71, 104).

[103] S. N. Krivoshapko and V. N. Ivanov. “The Second Order Surfaces”. In: *Encyclopedia of Analytical Surfaces*. Cham: Springer International Publishing, 2015, pp. 613–626. DOI: 10.1007/978-3-319-11773-7_35 (cit. on p. 106).

[104] A. Kunz, M. Sprenger, and H. Wernli. “Climatology of potential vorticity streamers and associated isentropic transport pathways across PV gradient barriers”. In: *J. Geophys. Res.-Atmos.* 120.9 (2015), pp. 3802–3821. DOI: 10.1002/2014JD022615. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2014JD022615 (cit. on p. 74).

[105] J.-F. Lamarque and P. G. Hess. “Cross-tropopause mass exchange and potential vorticity budget in a simulated tropopause folding”. In: *J. Atmos. Sci.* 51.15 (1994), pp. 2246–2269. DOI: 10.1175/1520-0469(1994)051<2246:CTMEAP>2.0.CO;2 (cit. on p. 75).

[106] Q. A. Lawton, S. J. Majumdar, K. Dotterer, C. Thorncroft, and C. J. Schreck. “The Influence of Convectively Coupled Kelvin Waves on African Easterly Waves in a Wave-Following Framework”. In: *Mon. Weather Rev.* 150.8 (2022), pp. 2055–2072. DOI: 10.1175/MWR-D-21-0321.1 (cit. on p. 115).

[107] H. Le Treut and E. Kalnay. “Comparison of observed and simulated cyclone frequency distribution as determined by an objetive method”. In: *Atmosfera* 3.1 (1990), pp. 57–71 (cit. on p. 17).

[108] J. A. Leese, C. S. Novak, and B. B. Clark. “An Automated Technique for Obtaining Cloud Motion from Geosynchronous Satellite Data Using Cross Correlation”. In: *J. Appl. Meteorol. Climatol.* 10.1 (1971), pp. 118–132. DOI: 10.1175/1520-0450(1971)0102.0.CO;2 (cit. on p. 17).

[109] S. Limbach, E. Schömer, and H. Wernli. “Detection, tracking and event localization of jet stream features in 4-D atmospheric data”. In: *Geosci. Model Dev.* 5.2 (2012), pp. 457–470. DOI: `10.5194/gmd-5-457-2012` (cit. on pp. 19, 20, 25, 53, 73, 107, 112, 117, 119).

[110] C. C. N. Lo. “Representation of African Easterly Waves and associated rainfall in two versions of ECMWF Integrated Forecasting System”. MA thesis. Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, 2024 (cit. on pp. 12, 69).

[111] M. Maier-Gerber, A. H. Fink, M. Riemer, et al. “Statistical–Dynamical Forecasting of Subseasonal North Atlantic Tropical Cyclone Occurrence”. In: *Weather Forecasting* 36.6 (2021), pp. 2127–2142. DOI: `10.1175/WAF-D-21-0020.1` (cit. on p. 134).

[112] M. Maier-Gerber, M. Riemer, A. H. Fink, et al. “Tropical Transition of Hurricane Chris (2012) over the North Atlantic Ocean: A Multiscale Investigation of Predictability”. In: *Mon. Weather Rev.* 147.3 (2019), pp. 951–970. DOI: `10.1175/MWR-D-18-0188.1` (cit. on pp. 73, 107).

[113] M. Maranan, A. H. Fink, and P. Knippertz. “Rainfall types over southern West Africa: Objective identification, climatology and synoptic environment”. In: *Q. J. R. Meteorolog. Soc.* 144.714 (2018), pp. 1628–1648. DOI: `10.1002/qj.3345`. eprint: `https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3345` (cit. on pp. 24, 131).

[114] A. C. Massacand, H. Wernli, and H. C. Davies. “Heavy precipitation on the Alpine southside: An upper-level precursor”. In: *Geophys. Res. Lett.* 25.9 (1998), pp. 1435–1438. DOI: `10.1029/98GL50869` (cit. on p. 73).

[115] G. Matheron. *Éléments pour une théorie des milieux poreux*. Masson et Cie, 1967 (cit. on p. 14).

[116] M. E. McIntyre and T. Palmer. “Breaking planetary waves in the stratosphere”. In: *Nature* 305.5935 (1983), pp. 593–600. DOI: `10.1038/305593a0` (cit. on pp. 46, 73, 111).

[117] A. Mekonnen, C. D. Thorncroft, and A. R. Aiyyer. “Analysis of Convection and Its Association with African Easterly Waves”. In: *J. Clim.* 19.20 (2006), pp. 5405–5421. DOI: `10.1175/JCLI3920.1` (cit. on pp. 50, 110).

[119] J. Molinari, S. Skubis, D. Vollaro, F. Alsheimer, and H. E. Willoughby. “Potential vorticity analysis of tropical cyclone intensification”. In: *J. Atmos. Sci.* 55.16 (1998), pp. 2632–2644. DOI: `10.1175/1520-0469(1998)055<2632:PVAOTC>2.0.CO;2` (cit. on p. 93).

[120] M. C. Morgan and J. W. Nielsen-Gammon. “Using Tropopause Maps to Diagnose Midlatitude Weather Systems”. In: *Mon. Weather Rev.* 126.10 (1998), pp. 2555–2579. DOI: `10.1175/1520-0493(1998)126<2555:UTMTDM>2.0.CO;2` (cit. on p. 73).

[121]R. Mukundan and K. R. Ramakrishnan. *Moment Functions in Image Analysis — Theory and Applications*. World scientific, 1998. DOI: `10.1142/3838`. eprint: `https://www.worldscientific.com/doi/pdf/10.1142/3838` (cit. on p. 124).

[122]A. Müller, B. Niedrich, and P. Névir. “Three-dimensional potential vorticity structures for extreme precipitation events on the convective scale”. In: *Tellus A: Dyn. Meteorol. Oceanogr.* 72.1 (2020), pp. 1–20. DOI: `10.1080/16000870.2020.1811535`. eprint: `https://doi.org/10.1080/16000870.2020.1811535` (cit. on pp. 48, 49, 111).

[123]R. J. Murray and I. Simmonds. “A numerical scheme for tracking cyclone centres from digital data. Part II: Application to January and July general circulation model simulations”. In: *Aust. Meteorol. Mag.* 39.3 (1991), pp. 167–180 (cit. on p. 17).

[124]L. Najman and H. Talbot. *Mathematical morphology: from theory to applications*. John Wiley & Sons, 2013 (cit. on p. 122).

[126]M. Nixon and A. Aguado. *Feature extraction and image processing for computer vision*. Academic press, 2019 (cit. on pp. 11, 27).

[129]K. M. Núñez Ocasio. “Tropical cyclogenesis and its relation to interactions between African easterly waves and mesoscale convective systems”. In: (2021) (cit. on p. 110).

[130]K. M. Núñez Ocasio, J. L. Evans, and G. S. Young. “Tracking Mesoscale Convective Systems that are Potential Candidates for Tropical Cyclogenesis”. In: *Mon. Weather Rev.* 148 (2020), pp. 655–669. DOI: `10.1175/MWR-D-19-0070.1` (cit. on pp. 112, 117, 134).

[131]K. M. Núñez Ocasio and R. Rios-Berrios. “African Easterly Wave Evolution and Tropical Cyclogenesis in a Pre-Helene (2006) Hindcast Using the Model for Prediction Across Scales-Atmosphere (MPAS-A)”. In: *J. Adv. Model. Earth Syst.* 15.2 (2023). e2022MS003181 2022MS003181, e2022MS003181. DOI: `10.1029/2022MS003181`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2022MS003181` (cit. on p. 122).

[132]A. Oertel, M. Boettcher, H. Joos, M. Sprenger, and H. Wernli. “Potential vorticity structure of embedded convection in a warm conveyor belt and its relevance for large-scale dynamics”. In: *Weather Clim. Dyn.* 1.1 (2020), pp. 127–153. DOI: `10.5194/wcd-1-127-2020` (cit. on pp. 48, 49).

[133]G. A. Papakostas, D. E. Koulouriotis, and E. G. Karakasis. “Computation strategies of orthogonal image moments: A comparative study”. In: *Appl. Math. Comput.* 216.1 (2010), pp. 1–17. DOI: `10.1016/j.amc.2010.01.051` (cit. on p. 122).

[134]P. P. Papin, L. F. Bosart, and R. D. Torn. “A Feature-Based Approach to Classifying Summertime Potential Vorticity Streamers Linked to Rossby Wave Breaking in the North Atlantic Basin”. In: *J. Climate* 33.14 (2020), pp. 5953–5969. DOI: `10.1175/JCLI-D-19-0812.1` (cit. on pp. 74, 97).

[135] B. N. Parlett. *The symmetric eigenvalue problem*. SIAM, 1998 (cit. on p. 29).

[136] C. Petres, Y. Pailhas, Y. Petillot, and D. Lane. “Underwater path planing using fast marching algorithms”. In: *Europe Oceans 2005*. Vol. 2. 2005, 814–819 Vol. 2. DOI: `10.1109/OCEANSE.2005.1513161` (cit. on pp. 82, 84).

[137] W. Piersig. “Schwankungen von Luftdruck und Luftbewegung: sowie ein Beitrag zum Wettergeschehen im Passatgebiet des östlichen Nordatlantischen Ozeans”. In: *Arch. Deut. Seewarte* 54.6 (1936) (cit. on p. 50).

[138] J. G. Pinto, T. Spangehl, U. Ulbrich, and P. Speth. “Sensitivities of a cyclone detection and tracking algorithm: individual tracks and climatology”. In: *Meteorol. Z.* 14.6 (2005), pp. 823–838 (cit. on p. 18).

[139] F. Porcù, A. Carrassi, C. M. Medaglia, F. Prodi, and A. Mugnai. “A study on cut-off low vertical structure and precipitation in the Mediterranean region”. In: *Meteorol. Atmos. Phys.* 96 (2007), pp. 121–140 (cit. on p. 48).

[140] R. Portmann, B. Crezee, J. Quinting, and H. Wernli. “The complex life cycles of two long-lived potential vorticity cut-offs over Europe”. In: *Q. J. Roy. Meteor. Soc.* 144.712 (2018), pp. 701–719. DOI: `10.1002/qj.3239`. eprint: `https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3239` (cit. on p. 75).

[141] R. Portmann, M. Sprenger, and H. Wernli. “The three-dimensional life cycles of potential vorticity cutoffs: a global and selected regional climatologies in ERA-Interim (1979–2018)”. In: *Weather Clim. Dynam.* 2.2 (2021), pp. 507–534. DOI: `10.5194/wcd-2-507-2021` (cit. on pp. 44, 73, 75, 94, 107).

[142] F. H. Post, B. Vrolijk, H. Hauser, R. S. Laramee, and H. Doleisch. “The State of the Art in Flow Visualisation: Feature Extraction and Tracking”. In: *Computer Graphics Forum* 22.4 (2003), pp. 775–792. DOI: `https://doi.org/10.1111/j.1467-8659.2003.00723.x`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8659.2003.00723.x` (cit. on p. 12).

[143] A. F. Prein, P. A. Mooney, and J. M. Done. “The Multi-Scale Interactions of Atmospheric Phenomenon in Mean and Extreme Precipitation”. In: *Earth’s Future* 11.11 (2023). e2023EF003534 2023EF003534, e2023EF003534. DOI: `https://doi.org/10.1029/2023EF003534`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2023EF003534` (cit. on p. 117).

[144] S. Purves. “Phase and the Hilbert transform”. In: *The Leading Edge* 33.10 (2014), pp. 1164–1166. DOI: `10.1190/tle33101164.1`. eprint: `https://doi.org/10.1190/tle33101164.1` (cit. on p. 119).

[145] V. Rajasree, X. Cao, H. Ramsay, et al. “Tropical cyclogenesis: Controlling factors and physical mechanisms”. In: *Trop. Cyclone Res. Rev.* 12.3 (2023), pp. 165–181. DOI: `10.1016/j.tcrr.2023.09.004` (cit. on p. 110).

[146]A. Rasheeda Satheesh, P. Knippertz, A. H. Fink, E.-M. Walz, and T. Gneiting. "Sources of predictability of synoptic-scale rainfall during the West African summer monsoon". In: *Q. J. R. Meteorolog. Soc.* 149.757 (2023), pp. 3721–3737. DOI: `https://doi.org/10.1002/qj.4581`. eprint: `https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.4581` (cit. on p. 52).

[147]S. Rasp and S. Lerch. "Neural Networks for Postprocessing Ensemble Weather Forecasts". In: *Mon. Weather Rev.* 146.11 (2018), pp. 3885–3900. DOI: `https://doi.org/10.1175/MWR-D-18-0187.1` (cit. on p. 112).

[148]M. Rautenhaus, C. M. Grams, A. Schäfler, and R. Westermann. "Three-dimensional visualization of ensemble weather forecasts – Part 2: Forecasting warm conveyor belt situations for aircraft-based field campaigns". In: *Geosci. Model Dev.* 8.7 (2015), pp. 2355–2377. DOI: `10.5194/gmd-8-2355-2015` (cit. on pp. 34, 35, 74, 76, 112).

[149]M. Rautenhaus, M. Böttinger, S. Siemen, et al. "Visualization in meteorology—a survey of techniques and tools for data analysis tasks". In: *IEEE T. Vis. Comput. Gr.* 24.12 (2018), pp. 3268–3296. DOI: `10.1109/TVCG.2017.2779501` (cit. on pp. 33, 74, 76, 134).

[150]M. Rautenhaus, M. Kern, A. Schäfler, and R. Westermann. "Three-dimensional visualization of ensemble weather forecasts – Part 1: The visualization tool Met.3D (version 1.0)". In: *Geosci. Model Dev.* 8.7 (2015), pp. 2329–2353. DOI: `10.5194/gmd-8-2329-2015` (cit. on pp. 76, 94, 121).

[151]D. J. Raymond and H. Jiang. "A Theory for Long-Lived Mesoscale Convective Systems". In: *J. Atmos. Sci.* 47.24 (1990), pp. 3067–3077. DOI: `10.1175/1520-0469(1990)0472.0.CO;2` (cit. on p. 48).

[152]R. Redl, C. Keil, G. Craig, S. Lerch, and J. Eichhorn. "Towards a Framework for Parallelized Post-Processing and Evaluation of Ensemble Forecasts". In: *EGU General Assembly Conference Abstracts*. EGU General Assembly Conference Abstracts. Apr. 2018, p. 12322 (cit. on pp. 54, 169).

[153]R. Redl, O. T. Prims, baurflorian, and cagau. "wavestoweather/enstools: Release v2022.11.1". In: (Nov. 2022). DOI: `10.5281/zenodo.7331674` (cit. on p. 58).

[154]R. J. Reed. "A study of a characteristic type of upper-level frontogenesis". In: *J. Atmos. Sci.* 12.3 (1955), pp. 226–237. DOI: `10.1175/1520-0469(1955)0122.0.CO;2` (cit. on p. 45).

[155]R. J. Reed, E. Klinker, and A. Hollingsworth. "The structure and characteristics of African easterly wave disturbances as determined from the ECMWF operational analysis/forecast system". In: *Meteorol. Atmos. Phys.* 38 (1988), pp. 22–33 (cit. on pp. 17, 112).

[156]R. J. Reed, D. C. Norquist, and E. E. Recker. "The Structure and Properties of African Wave Disturbances as Observed During Phase III of GATE". In: *Mon. Weather Rev.* 105.3 (1977), pp. 317–333. DOI: `10.1175/1520-0493(1977)1052.0.CO;2` (cit. on pp. 51, 113, 120).

[157]H. Regula. “Druckschwankungen und Tornados an der Westküste von Afrika”. In: *Ann. Hydrogr. Maritimen Meteor.* 64 (1936), pp. 107–111 (cit. on p. 50).

[158]H. Riehl. “Tropical Meteorology”. In: McGraw-Hill Book Co., 1954. Chap. 9, pp. 210–234 (cit. on pp. 113, 127).

[159]C.-G. A. Rossby. “Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semi-permanent centers of action”. In: *J. Mar. Res.* 2 (1939), pp. 38–55 (cit. on pp. 10, 44, 45).

[160]J. O. Russell, A. Aiyyer, J. D. White, and W. Hannah. “Revisiting the connection between African Easterly Waves and Atlantic tropical cyclogenesis”. In: *Geophys. Res. Lett.* 44.1 (2017), pp. 587–595. DOI: `10.1002/2016GL071236` (cit. on pp. 51, 110, 112).

[161]J. O. H. Russell, A. Aiyyer, and J. Dylan White. “African Easterly Wave Dynamics in Convection-Permitting Simulations: Rotational Stratiform Instability as a Conceptual Model”. In: *J. Adv. Model. Earth Syst.* 12.1 (2020), e2019MS001706. DOI: `10.1029/2019MS001706` (cit. on pp. 51, 112, 115, 120, 122, 127, 128).

[162]G. Salmon. *A treatise on the analytic geometry of three dimensions*. Vol. 1. Longmans, Green and Company, 1912 (cit. on p. 30).

[163]A. Schlueter, A. H. Fink, P. Knippertz, and P. Vogel. “A Systematic Comparison of Tropical Waves over Northern Africa. Part I: Influence on Rainfall”. In: *J. Clim.* 32.5 (2019), pp. 1501–1523. DOI: `10.1175/JCLI-D-18-0173.1` (cit. on p. 51).

[164]S. Schmidt, M. Riemer, and T. Selz. *A feature based perspective on upscale error growth.* Tech. rep. Copernicus Meetings, 2023 (cit. on p. 53).

[165]U. Schulzweida. *CDO User Guide*. Oct. 2019. DOI: `10.5281/zenodo.3539275` (cit. on p. 98).

[166]J. Serra. *Image Analysis and Mathematical Morphology*. USA: Academic Press, Inc., 1983 (cit. on p. 14).

[167]J. A. Sethian. “A fast marching level set method for monotonically advancing fronts”. In: *P. Natl. Acad. Sci. USA* 93.4 (1996), pp. 1591–1595. DOI: `10.1073/pnas.93.4.1591` (cit. on p. 84).

[168]L. J. Shapiro. “The Vorticity Budget of a Composite African Tropical Wave Disturbance”. In: *Mon. Weather Rev.* 106.6 (1978), pp. 806–817. DOI: `10.1175/1520-0493(1978)106<0806:TVBOAC>2.0.CO;2` (cit. on p. 113).

[169]B. Škerlak, M. Sprenger, S. Pfahl, E. Tyrlis, and H. Wernli. “Tropopause folds in ERA-Interim: Global climatology and relation to extreme weather events”. In: *J. Geophys. Res.-Atmos.* 120.10 (2015), pp. 4860–4877. DOI: `10.1002/2014JD022787`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2014JD022787` (cit. on pp. 38, 75, 93).

[170]W. L. Smith and H. M. Woolf. “The Use of Eigenvectors of Statistical Covariance Matrices for Interpreting Satellite Sounding Radiometer Observations”. In: *J. Atmos. Sci.* 33.7 (1976), pp. 1127–1140. DOI: `10.1175/1520-0469(1976)0332.0.CO;2` (cit. on pp. 30, 122).

[171]J. P. Snyder. *Map projections–A working manual.* Vol. 1395. USGS, 1987 (cit. on pp. 7–10, 79, 81, 82).

[172]M. B. Souders, B. A. Colle, and E. K. M. Chang. “The Climatology and Characteristics of Rossby Wave Packets Using a Feature-Based Tracking Technique”. In: *Mon. Weather Rev.* 142.10 (2014), pp. 3528–3548. DOI: `10.1175/MWR-D-13-00371.1` (cit. on p. 18).

[173]M. Sprenger, G. Fragkoulidis, H. Binder, et al. “Global Climatologies of Eulerian and Lagrangian Flow Features based on ERA-Interim”. In: *B. Am. Meteorol. Soc.* 98.8 (2017), pp. 1739–1748. DOI: `10.1175/BAMS-D-15-00299.1` (cit. on p. 88).

[174]M. Sprenger, O. Martius, and J. Arnold. “Cold surge episodes over southeastern Brazil – a potential vorticity perspective”. In: *Int. J. Climatol.* 33.12 (2013), pp. 2758–2767. DOI: `10.1002/joc.3618`. eprint: `https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.3618` (cit. on pp. 24, 88).

[175]G. Strang. *Introduction to linear algebra.* SIAM, 2022 (cit. on p. 30).

[176]N. A. Streten and A. J. Troup. “A synoptic climatology of satellite observed cloud vortices over the Southern Hemisphere”. In: *Quarterly Journal of the Royal Meteorological Society* 99.419 (1973), pp. 56–72. DOI: `10.1002/qj.49709941906`. eprint: `https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.49709941906` (cit. on p. 17).

[177]R. C. Sutcliffe. “Cyclonic and anticyclonic development”. In: *Q. J. R. Meteorolog. Soc.* 65.282 (1939), pp. 518–524. DOI: `https://doi.org/10.1002/qj.49706528208`. eprint: `https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.49706528208` (cit. on p. 46).

[178]R. E. Tarjan and J. van Leeuwen. “Worst-Case Analysis of Set Union Algorithms”. In: *J. ACM* 31.2 (Mar. 1984), pp. 245–281. DOI: `10.1145/62.2160` (cit. on p. 22).

[179]M. R. Teague. “Image analysis via the general theory of moments”. In: *J. Opt. Soc. Am.* 70.8 (Aug. 1980), pp. 920–930. DOI: `10.1364/JOSA.70.000920` (cit. on p. 122).

[180]F. Teubler and M. Riemer. “Potential-vorticity dynamics of troughs and ridges within Rossby wave packets during a 40-year reanalysis period”. In: *Weather Clim. Dyn.* 2.3 (2021), pp. 535–559. DOI: `10.5194/wcd-2-535-2021` (cit. on p. 111).

[181] C. D. Thorncroft, B. J. Hoskins, and M. E. McIntyre. “Two paradigms of baroclinic-wave life-cycle behaviour”. In: *Q. J. Roy. Meteor. Soc.* 119.509 (1993), pp. 17–55. DOI: `10.1002/qj.49711950903`. eprint: `https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.49711950903` (cit. on pp. 28, 73, 111).

[182] C. Thorncroft and K. Hodges. “African Easterly Wave Variability and Its Relationship to Atlantic Tropical Cyclone Activity”. In: *J. Clim.* 14.6 (2001), pp. 1166–1179. DOI: `10.1175/1520-0442(2001)014<1166:AEWVAI>2.0.CO;2` (cit. on pp. 113, 114).

[183] C. D. Thorncroft, N. M. J. Hall, and G. N. Kiladis. “Three-Dimensional Structure and Dynamics of African Easterly Waves. Part III: Genesis”. In: *J. Atmos. Sci.* 65.11 (2008), pp. 3596–3607. DOI: `10.1175/2008JAS2575.1` (cit. on pp. 51, 111).

[184] C. D. Thorncroft and B. J. Hoskins. “An idealized study of African easterly waves. II: A nonlinear view”. In: *Q. J. R. Meteorolog. Soc.* 120.518 (1994), pp. 983–1015. DOI: `https://doi.org/10.1002/qj.49712051810`. eprint: `https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.49712051810` (cit. on p. 51).

[185] L. Tomassini, D. J. Parker, A. Stirling, et al. “The interaction between moist diabatic processes and the atmospheric circulation in African Easterly Wave propagation”. In: *Q. J. R. Meteorolog. Soc.* 143.709 (2017), pp. 3207–3227. DOI: `10.1002/qj.3173`. eprint: `https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3173` (cit. on pp. 111, 113, 122, 128).

[186] D. A. Turner, I. J. Anderson, J. C. Mason, and M. G. Cox. *An algorithm for fitting and ellipsoid to data, Technical Report*. Tech. rep. RR9803. School of Computing and Mathematics, University of Huddersfield, 1999 (cit. on p. 122).

[187] P. A. Ullrich, C. M. Zarzycki, E. E. McClenny, et al. “TempestExtremes v2.1: a community framework for feature detection, tracking, and analysis in large datasets”. In: *Geosci. Model Dev.* 14.8 (2021), pp. 5023–5048. DOI: `10.5194/gmd-14-5023-2021` (cit. on p. 117).

[188] P. A. Ullrich and C. M. Zarzycki. “TempestExtremes: a framework for scale-insensitive pointwise feature tracking on unstructured grids”. In: *Geosci. Model Dev.* 10.3 (2017), pp. 1069–1090. DOI: `10.5194/gmd-10-1069-2017` (cit. on p. 65).

[189] R. Van der Linden, A. H. Fink, J. G. Pinto, and T. Phan-Van. “The dynamics of an extreme precipitation event in northeastern Vietnam in 2015 and its predictability in the ECMWF ensemble prediction system”. In: *Weather Forecasting* 32.3 (2017), pp. 1041–1056. DOI: `10.1175/WAF-D-16-0142.1` (cit. on pp. 76, 98–100, 102, 106, 165).

[190] F. Vitart et al. “The subseasonal to seasonal (S2S) prediction project database”. In: *B. Am. Meteorol. Soc.* 98.1 (2017), pp. 163–173. DOI: `10.1175/BAMS-D-16-0017.1` (cit. on pp. 86, 90, 164).

[191]P. Vogel, P. Knippertz, A. H. Fink, A. Schlueter, and T. Gneiting. “Skill of Global Raw and Postprocessed Ensemble Predictions of Rainfall over Northern Tropical Africa”. In: *Weather Forecasting* 33.2 (2018), pp. 369–388. DOI: `10.1175/WAF-D-17-0127.1` (cit. on p. 52).

[192]P. Vogel, P. Knippertz, T. Gneiting, et al. “Statistical Forecasts for the Occurrence of Precipitation Outperform Global Models over Northern Tropical Africa”. In: *Geophys. Res. Lett.* 48.3 (2021), e2020GL091022. DOI: `10.1029/2020GL091022`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020GL091022` (cit. on p. 134).

[193]D. Vranic and D. Saupe. “3D Shape Descriptor Based on 3D Fourier Transform”. In: *Proceedings of ECMCS- 2001, the 3 rd EURASIP Conference on Digital Signal Processing for Multimedia Communications and Services, 11 - 13 September 2001, Budapest, Hungary*. Ed. by K. Fazekas. Budapest: Scientific Assoc. of Infocommunications, 2001, pp. 271–274 (cit. on p. 29).

[194]E.-M. Walz, P. Knippertz, A. H. Fink, G. Köhler, and T. Gneiting. *Physics-based vs. data-driven 24-hour probabilistic forecasts of precipitation for northern tropical Africa.* 2024. arXiv: `2401.03746 [physics.ao-ph]` (cit. on p. 52).

[197]N. P. Wedi, P. Bauer, W. Deconinck, et al. *The modelling infrastructure of the Integrated Forecasting System: Recent advances and future challenges*. Technical Memorandum 760. ECMWF: Reading, UK, 2015 (cit. on p. 6).

[198]C. Weijenborg, J. Chagnon, P. Friederichs, S. Gray, and A. Hense. “Coherent evolution of potential vorticity anomalies associated with deep moist convection”. In: *Q. J. Roy. Meteor. Soc.* 143.704 (2017), pp. 1254–1267. DOI: `10.1002/qj.3000` (cit. on pp. 48, 93, 111).

[199]C. Weijenborg, P. Friederichs, and A. Hense. “Organisation of potential vorticity on the mesoscale during deep moist convection”. In: *Tellus A: Dyn. Meteorol. Oceanogr.* 67.1 (2015), p. 25705. DOI: `10.3402/tellusa.v67.25705`. eprint: `https://doi.org/10.3402/tellusa.v67.25705` (cit. on p. 48).

[200]H. Wernli and M. Sprenger. “Identification and ERA-15 Climatology of Potential Vorticity Streamers and Cutoffs near the Extratropical Tropopause”. In: *J. Atmos. Sci.* 64.5 (2007), pp. 1569–1586. DOI: `10.1175/JAS3912.1` (cit. on pp. 74, 76, 87, 88, 91, 97, 164).

[201]K. Whitehall, C. A. Mattmann, G. Jenkins, et al. “Exploring a graph theory based algorithm for automated identification and characterization of large mesoscale convective systems in satellite datasets”. In: *Earth Sci. Inf.* 8 (2015), pp. 663–675. DOI: `10.1007/s12145-014-0181-3` (cit. on p. 117).

[202]D. L. Williamson. “Storm track representation and verification”. In: *Tellus* 33.6 (1981), pp. 513–530. DOI: `10.1111/j.2153-3490.1981.tb01778.x`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.2153-3490.1981.tb01778.x` (cit. on p. 17).

[206] G. Xu, X. Ma, P. Chang, and L. Wang. “Image-processing-based atmospheric river tracking method version 1 (IPART-1)”. In: *Geosci. Model Dev.* 13.10 (2020), pp. 4639–4662. DOI: `10.5194/gmd-13-4639-2020` (cit. on p. 25).

[207] G. Zängl, D. Reinert, P. Rípodas, and M. Baldauf. “The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core”. In: *Q. J. R. Meteorolog. Soc.* 141.687 (2015), pp. 563–579. DOI: `https://doi.org/10.1002/qj.2378`. eprint: `https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.2378` (cit. on p. 6).

[208] A. V. Zimin, I. Szunyogh, D. J. Patil, B. R. Hunt, and E. Ott. “Extracting Envelopes of Rossby Wave Packets”. In: *Mon. Weather Rev.* 131.5 (2003), pp. 1011–1017. DOI: `10.1175/1520-0493(2003)131<1011:EEORWP>2.0.CO;2` (cit. on p. 119).

Webpages

[@4] American Meteorological Society. *Data and Software Policy Guidelines for AMS Publications*. URL: `https://www.ametsoc.org/index.cfm/ams/publications/ethical-guidelines-and-ams-policies/data-and-software-policy-guidelines-for-ams-publications/` (visited on June 10, 2024) (cit. on p. 55).

[@5] Anaconda Software Distribution. *Conda*. URL: `https://www.anaconda.com/` (visited on June 10, 2024) (cit. on p. 57).

[@28] Copernicus Programme. *Copernicus Publications: Manuscript preparation*. URL: `https://publications.copernicus.org/for_authors/manuscript_preparation.html` (visited on June 10, 2024) (cit. on p. 55).

[@31] Department of Agronomy, Iowa State University. *Iowa Environmental Mesonet*. URL: `https://mesonet.agron.iastate.edu/vtec/#2023-O-NEW-KHFO-FA-A-0004/USCOMP-N0Q-202303231140` (visited on June 10, 2024) (cit. on p. 48).

[@61] Google LLC. *Protocol Buffers*. URL: `https://protobuf.dev/` (visited on June 10, 2024) (cit. on p. 58).

[@96] Kitware Inc. *ParaView*. URL: `https://www.paraview.org/` (visited on June 10, 2024) (cit. on p. 34).

[@97] Kitware Inc. *The Visualization Toolkit*. URL: `https://vtk.org/` (visited on June 10, 2024) (cit. on p. 35).

[@118] Met.3D contributors. *Met.3D*. URL: `https://met3d.readthedocs.io/` (visited on June 10, 2024) (cit. on p. 35).

[@125]National Center for Atmospheric Research's Computational and Information Systems Lab. *Vapor*. URL: `https://www.vapor.ucar.edu/` (visited on June 10, 2024) (cit. on p. 34).

[@127]NSF Unidata, University Corporation for Atmospheric Research. *Integrated Data Viewer (IDV)*. URL: `https://www.unidata.ucar.edu/software/idv/` (visited on June 10, 2024) (cit. on p. 34).

[@128]NSF Unidata, University Corporation for Atmospheric Research. *Network Common Data Form (NetCDF)*. URL: `https://www.unidata.ucar.edu/software/netcdf/` (visited on June 10, 2024) (cit. on pp. 6, 54).

[@195]Waves to Weather - Transregional Collaborative Research Project. *Ensemble Tools*. URL: `https://github.com/wavestoweather/enstools/` (visited on June 10, 2024) (cit. on p. 58).

[@196]Waves to Weather - Transregional Collaborative Research Project. *Ensemble Tools Documentation*. URL: `https://enstools.readthedocs.io/en/latest/` (visited on June 10, 2024) (cit. on p. 58).

[@203]World Meteorological Organization. *General Regularly-distributed Information in Binary form (GRIB)*. URL: `https://www.nco.ncep.noaa.gov/pmb/docs/grib2/grib2_doc/` (visited on June 10, 2024) (cit. on pp. 6, 54).

[@204]Xarray Developers. *xarray*. URL: `https://xarray.dev/` (visited on June 10, 2024) (cit. on p. 58).

[@205]Xarray Developers. *xarray documentation*. URL: `https://docs.xarray.dev/en/stable/` (visited on June 10, 2024) (cit. on p. 58).

List of Figures

Code and data availability

The implementation of the feature identification and tracking framework *enstools-feature* is available at `https://gitlab.com/Christoph.Fischer/enstools-feature` (last accessed 16 June 2024) and archived at Zenodo [50]. It is realized as part of the open-source framework *enstools*, available at `https://github.com/wavestoweather/enstools` (last accessed 16 June 2024) [152].

The data sets used in this thesis (ERA5 reanalysis and forecasts, S2S database, GPM-IMERG) are publicly available at the relevant cited sources. Furthermore, tracked AEW troughs based on the method introduced in Chapter 6, and the corresponding 3-D PV features for the ERA5 reanalysis are available at Zenodo [51], [52].

Declaration

I hereby declare that I have written the present thesis independently and without use of other than the indicated means. I also declare that to the best of my knowledge all passages taken from published and unpublished sources have been referenced. The paper has not been submitted for evaluation to any other examining authority nor has it been published in any form whatsoever.

Mainz, June 18, 2024

Christoph Fischer