Cluster Analysis Tailored to Structure Change of Tropical Cyclones Using a Very Large Number of Trajectories

TOBIAS KREMER AND ELMAR SCHÖMER

Institute of Computer Science, Johannes Gutenberg University Mainz, Mainz, Germany

CHRISTIAN EULER AND MICHAEL RIEMER

Institute for Atmospheric Physics, Johannes Gutenberg University Mainz, Mainz, Germany

(Manuscript received 22 December 2019, in final form 6 August 2020)

ABSTRACT: Major airstreams in tropical cyclones (TCs) are rarely described from a Lagrangian perspective. Such a perspective, however, is required to account for asymmetries and time dependence of the TC circulation. We present a procedure that identifies main airstreams in TCs based on trajectory clustering. The procedure takes into account the TC's large degree of inherent symmetry and is suitable for a very large number of trajectories [$\mathcal{O}(10^6)$]. A large number of trajectories may be needed to resolve both the TC's inner-core convection as well as the larger-scale environment. We define similarity of trajectories based on their shape in a storm-relative reference frame, rather than on proximity in physical space, and use Fréchet distance, which emphasizes differences in trajectory shape, as a similarity metric. To make feasible the use of this elaborate metric, data compression is introduced that approximates the shape of trajectories in an optimal sense. To make clustering of large numbers of trajectories computationally feasible, we reduce dimensionality in distance space by so-called landmark multidimensional scaling. Finally, k -means clustering is performed in this low-dimensional space. We investigate the extratropical transition of Tropical Storm Karl (2016) to demonstrate the applicability of our clustering procedure. All identified clusters prove to be physically meaningful and describe distinct flavors of inflow, ascent, outflow, and quasi-horizontal motion in Karl's vicinity. Importantly, the clusters exhibit gradual temporal evolution, which is most notable because the clustering procedure itself does not impose temporal consistency on the clusters. Finally, TC problems are discussed for which the application of the clustering procedures seems to be most fruitful.

KEYWORDS: Hurricanes; Clustering

1. Introduction

The notion of coherent airstreams plays a long-standing and fundamental role in our understanding of cyclone dynamics (e.g., Hamilton and Archbold 1945; Riehl 1954; Eliassen and Kleinschmidt 1957). In midlatitude cyclones coherent airstreams are referred to as conveyor belts (e.g., Browning 1971; Carlson 1980). These airstreams are distinctly asymmetric with respect to the cyclone's center and have been studied extensively based on trajectory calculations (e.g., Wernli and Davies 1997, and subsequent developments). In tropical cyclones (TCs) the major airstreams are usually separated into the primary and the secondary circulation and are traditionally depicted as streamlines in horizontal and radius–height cross sections, respectively (e.g., Riehl 1954). Such a streamline framework may be employed because mature TCs can be considered, to lowest order, as axisymmetric, steady-state

systems. For a steady state, streamlines and trajectories are equivalent.

A two-dimensional streamline framework, however, is a poor approximation when transient and/or asymmetric motions become important (e.g., in TCs that undergo substantial structure change or TCs embedded in an environmental flow with considerable vertical shear). So far, relatively few comprehensive trajectory analyses of TCs have been performed (e.g., Cram et al. 2007; Stern and Zhang 2013; Riemer and Laliberté 2015). Trajectory analysis of TCs needs to resolve convection in the inner core and thus requires high temporal and spatial resolution of the underlying data when model output is used to calculate trajectories. This requirement for large datasets may partly impede the application of trajectory analysis to TCs. A further and conceptually more severe impediment is the need for a concise interpretation of the wealth of information contained in the computed trajectories. Regarding the thermodynamic interaction with the environment, Riemer and Laliberté (2015) proposed transformation of three-dimensional (3D) trajectories into a two-dimensional thermodynamic (moist entropy–temperature) space as a succinct depiction of a TC's secondary circulation. A succinct

Denotes content that is immediately available upon publication as open access.

Corresponding author: Michael Riemer, mriemer@uni-mainz.de

DOI: 10.1175/MWR-D-19-0408.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) ([www.ametsoc.org/PUBSReuseLicenses](#)).

description of the evolving main airstreams during TC structure change has, to our knowledge, not been proposed yet.

In midlatitude cyclones the conceptual model of conveyor belts is heavily used to organize and diagnose information from trajectory calculations (e.g., Wernli and Davies 1997; Papritz and Schemm 2013; Joos and Forbes 2016). For example, a rather simple set of variables and appropriate thresholds can be used to identify warm conveyor belts as ensembles of trajectories that behave very similarly and that densely populate a compact region in 3D physical space (sometimes referred to as coherent bundles of trajectories; Wernli and Davies 1997). Clustering techniques have been employed more recently in an attempt to identify conveyor-belt trajectories more objectively (e.g., Hart et al. 2015). If trajectory bundles are coherent in 3D physical space, the bundle may be represented simply by its average 3D trajectory or by suitable thresholds of some trajectory-density metric.

In tropical cyclones the large degree of symmetry of the system impedes a definition of coherent airstreams based on similarity in physical space. To illustrate this point, consider an axisymmetric TC, in which some air parcels spiral toward the center in the frictional inflow layer, ascend in the eyewall, and swirl out to larger radii in the outflow layer in the upper troposphere. These air parcels constitute the classical secondary circulation of the TC and thus a physically meaningful class of trajectories. In 3D physical space, however, individual trajectories of this secondary circulation may differ substantially because they may start and end at very different azimuth. In addition, the average 3D trajectory of this class collapses to a line at the TC's center for exact axisymmetry.

The goal of this study is to develop an objective classification of major airstreams in the vicinity of TCs based on cluster analysis. The particular problems to be addressed are (i) to appropriately represent both the convective-scale motion in the TC's inner core and the larger-scale, more quasi-horizontal environmental motions and (ii) to take into account the large degree of symmetry that characterizes a TC. The first problem requires a dense distribution and a high temporal resolution of trajectories in the inner core. To address this problem, downsampling strategies to reduce the density of trajectories outside of the inner core, and thus to reduce the total number of trajectories, could be developed and applied. We here take a different approach and use uniform sampling, which yields order of 10^6 trajectories per considered time period. We will present methods (i) to compress the large amount of trajectory data while preserving the main characteristics of the shape of trajectories and (ii) to reduce the dimensionality of the clustering problem and thus make feasible cluster analysis for a very large amount of trajectories. The second problem, accounting for the large degree of symmetry, will be addressed by a suitable transformation of trajectories before clustering. The basic idea here is to compute similarity of trajectories based on their shape but not on their absolute position in 3D physical space. Addressing these two problems will be the main method developments presented in this study.

The applicability of our approach will be demonstrated for the case of a TC that undergoes substantial structure change and that is strongly affected by vertical shear (i.e., a case in which transient and asymmetric motion can be expected to be of crucial importance). This case will be the extratropical transition

(ET) of Karl (2016), one of the major cases during the North Atlantic Waveguide and Downstream Impact Experiment (NAWDEX; Schäfler et al. 2018). During ET, TCs lose their tropical characteristics and start to transition into midlatitude cyclones (Jones et al. 2003; Evans et al. 2017). An ET case therefore appears to be a well-suited testbed to demonstrate the applicability of our clustering approach. In addition, our clustering results can be compared with results of a previous study that has investigated the ET of Karl focusing on airstreams associated with latent heat release near the inner core (Euler et al. 2019). In the current study, however, we will consider the more general Lagrangian evolution of air masses, including quasi-horizontal environmental flow. Specifically, we will consider all trajectories within a 2.5° radius of Karl's center. This radius is expected to capture the bulk of the air masses above the frictional inflow layer that interact with a TC's inner-core convection within 12 h, which is a typical time scale for TC intensity change due to environmental interaction [as observed in idealized experiments with moderate to strong vertical shear; e.g., Frank and Ritchie (2001); Riemer et al. (2010)].

Section 2 provides an overview of Karl's ET, describes the model simulation that provides the data underlying this study, and gives information on the calculation of trajectories. Our method to account for the large degree of symmetry and the very large number of trajectories are detailed in section 3. The results of the cluster analysis are presented in section 4 where it is demonstrated that all clusters are distinct and represent physically meaningful classes of trajectories. A brief analysis of thermodynamic characteristics along these main airstreams (section 5) provides insight into aspects of Karl's intensification and may illustrate avenues for future analyses based on automatically identified airstreams. Section 6 finally provides a summary, conclusions, and an outlook on potential future work.

2. The ET of Karl: Convection-permitting simulation and trajectory computation

a. Synoptic overview of Karl's evolution

Karl originally developed from a tropical wave off the coast of West Africa, was declared a tropical depression¹ near the Cape Verde Islands on 14 September 2016, and further intensified to tropical-storm strength on 15 September. Tropical Storm Karl moved mainly westward for several days until beginning to interact with an upper-level low on 21 September. During this interaction Karl weakened to a tropical depression and started to move to the northwest. When the upper-level low moved away toward the south, Karl started to move northward and to reintensify to tropical-storm intensity. After recurvature on 23 September, vertical shear increased and Karl accelerated toward the northeast. Karl reached its peak intensity [60 kt ($1 \text{ kt} \approx 0.51 \text{ m s}^{-1}$)] just before the TC became embedded in a midlatitude frontal zone and was declared extratropical on 25 September. A more detailed overview of Tropical Storm Karl is provided by Pasch and Zelinsky (2016).

¹ By the National Hurricane Center of the U.S. National Oceanic and Atmospheric Administration.

FIG. 1. Synoptic overview of Karl's evolution depicting θ_e at 3 km (colors) and wind speed at 10 km height (gray shading) using COSMO data at (a) 2100 UTC 24 Sep (defined as IS stage) and (b) 1200 UTC 25 Sep (defined as XT stage). The official NHC track between 1800 UTC 23 Sep and 1800 UTC 25 Sep is shown in black, with dots every 6 h. The respective track in the COSMO simulation is shown in red. The white squares highlight the center times of the four stages that are investigated in detail in this study. (c) Karl's intensity evolution (black: NHC, red: COSMO simulation, dashed: pressure, solid: wind speed).

This study considers Karl's substantial structure changes that occurred on 24 and 25 September (i.e., the intensification of Tropical Storm Karl toward peak intensity in an environment with increasing vertical wind shear until Karl lost its tropical characteristics and was declared extratropical). An illustration of Karl's track and intensity evolution between 1800 UTC 23 September and 1800 UTC 25 September is provided in Fig. 1, along with the distribution of equivalent potential temperature (θ_e) in the low to midtroposphere and Karl's location relative to the midlatitude jet.

During the extratropical stage of ET (Klein et al. 2000) Karl moved rapidly to the northeast and merged with a preexisting weak cyclone. This merger resulted in an unusually strong midlatitude jet streak. Karl has thus been identified as one of the "triggers" for midlatitude impact during the NAWDEX campaign (Schäfler et al. 2018). This impact manifested itself in a severe precipitation event in Northern Europe. Using ensemble sensitivity analysis, it has been demonstrated that this high-impact event exhibited sensitivity to Karl's evolution during ET (Kumpf et al. 2019).

b. Convection-permitting COSMO simulation

The data underlying this study is the same as in Euler et al. (2019). A convection-permitting simulation of Tropical Storm Karl (2016) has been performed using the Consortium for

Small-scale Modeling (COSMO; Steppeler et al. 2003) model, version 5.04. COSMO is a nonhydrostatic limited-area model for fully compressible flow, which has been designed for both operational numerical weather prediction and scientific applications. Besides in Euler et al. (2019), convection-permitting COSMO simulations have been successfully used before to investigate structure changes of a TC during ET (Lentink et al. 2018).

The horizontal grid spacing of the simulation is 0.025° (~ 2.8 km) with 49 levels between 0 and 21 km height. COSMO default settings have been used for the parameterization schemes: inter alia a turbulent kinetic energy-based turbulence scheme, a shallow-convection scheme with mass-flux closure after Tiedtke (1989), and a single-moment ice microphysics scheme (including graupel). The parameterization schemes are described in Doms et al. (2011) and more details on the setup of the simulation are given in Euler et al. (2019). Our simulation domain spans from 22° to 54° N and from 40° to 70° W. Boundary and initial conditions have been taken from archived operational analysis from the European Centre for Medium-Range Weather Forecasts. The simulation employs rotated spherical coordinates (Doms and Schättler 2002), with the model equator running approximately through the center of the domain at 35° N. This rotation of the coordinate system minimizes the convergence of the meridians such that the grid

FIG. 2. Overview of Karl's structure during the four key stages described in Euler et al. (2019) and further examined in this study: (a) the TS stage, (b) the IS stage, (c) the IM stage, and (d) the XT stage. The individual stages are shown at their respective representative times as defined in the text. Colors and arrows depict the wind speed and wind direction, respectively, at 1.5 km height. Red and blue shading depict up- and downward motion, respectively, averaged between 3 and 7 km. Two contour lines of θ_e at 3 km height (dark purple: 320 K, light purple: 324 K) indicate the midlatitude baroclinic zone. The respective shear vector and magnitude are given in the lower-left corner of each panel.

spacing, specified in degrees latitude and longitude, is approximately constant. The results will be presented in this native rotated coordinate system, denoted by r_{lat} and r_{lon} . Karl's geographical position in longitude and latitude is depicted in Figs. 1a, 1b, and 2.

c. Karl's structure change in the COSMO simulation

The simulation spans the 66-h period from 0000 UTC 23 September to 1800 UTC 25 September. The general track and intensity evolution of Karl is captured well by the simulation (Fig. 1), as compared to the best track analysis by Pasch and Zelinsky (2016). In the simulation, as discussed in some more detail in Euler et al. (2019), Karl moves somewhat faster and reaches somewhat higher intensity, with a position and intensity error at the end of the simulation of approximately

200 km and 7 m s^{-1} , respectively. Despite these differences, we deem the simulation to be sufficiently realistic for the purpose of the current study.

Euler et al. (2019) identified four key stages in the evolution of Karl during the simulated period. Early during the simulation (Fig. 2a), Tropical Storm Karl is situated in a moist tropical environment with low vertical shear from the southwest. Inner-core convection mostly occurs in the downshear-left quadrant and the horizontal winds exhibit a prominent wavenumber-1 pattern with maximum winds to the right of the shear vector. This period has been dubbed tropical storm (TS) stage, represented by time 1200 UTC 24 September. Subsequently, Karl starts to intensify in the COSMO simulation, notwithstanding increasing vertical wind shear, and the inner-core updrafts in the downshear-left quadrant become more vigorous (Fig. 2b).

FIG. 3. Illustration of the transformation of trajectories before clustering. The horizontal projections of 20 randomly selected trajectories are shown. Different colors are used to better distinguish individual trajectories in the same panel and the same trajectory is displayed in the same color in different panels. (a) Horizontal projection on COSMO's native rlon, rlat grid (in degrees). (b) The same trajectories in the storm-relative frame of reference with the origin at the storm center. (c) The same trajectories in the storm-relative frame of reference but now with their seeding points moved to the origin and the trajectories' directions at seeding time aligned along the positive rlon axis.

Karl starts to approach a low-level baroclinic zone to its northwest but interaction with this feature remains weak. Instead, at this stage, Karl exhibits clear characteristics of a TC in moderate shear and the stage has been dubbed intensification (IS) stage. The representative time for this stage is 2100 UTC 24 September.

The remainder of the simulation is characterized by prominent interaction of Karl with the baroclinic zone and thus the midlatitude flow. Vertical wind shear further increases and Karl's convection in the downshear-left quadrant starts to merge with banded convection along the low-level baroclinic zone (Fig. 2c). Prominent descent in the near-core region occurs in the upshear-left quadrant. Overall, Karl starts to lose its tropical characteristics and this stage is thus referred to as intermediate (IM) stage, represented at 0600 UTC 25 September. Finally, Karl starts to substantially distort the baroclinic zone, is affected by very strong vertical shear (above 20 m s^{-1}), and a dipole of vertical motion is evident with ascent in the downshear-left and descent in the upshear-left quadrant (Fig. 2d). This stage is referred to as extratropical (XT) stage, represented at 1200 UTC 25 September. While Euler et al. (2019) focused on trajectories associated with near-core vertical motion our analysis here will consider the behavior of all trajectories in the vicinity of Karl, thus including quasi-horizontal environmental flow also.

d. Trajectory calculation

Data for trajectory calculation are available every 5 min from the simulation. Trajectories are calculated using a fourth-order Runge–Kutta scheme with linear interpolation in space and time and a 2 s time step. Trajectories are seeded with a horizontal spacing equivalent to the grid spacing (i.e., every 0.025° within a radius of 2.5° around the storm center). The center is defined by a low-level centroid of the Okubo–Weiss parameter [Okubo 1970; Weiss 1991; see Euler et al. (2019) for details]. Trajectories are seeded on pressure levels every 20 hPa, with a starting level of 10 hPa above the surface and

extending up to the highest level with pressure greater than 150 hPa. This seeding strategy results in approximately 1.3×10^6 trajectories for each considered seeding time. Trajectories are seeded at the four stages TS, IS, IM, and XT and are computed for 6 h forward and backward in time. The final positions of the forward computation will hereafter be referred to as “end positions” and the final positions of the backward computation as “initial positions.” Trajectory positions are stored at the same times as the underlying model data (i.e., every 5 min). Each trajectory is thus represented by 145 data points.

The presented results are qualitatively robust to small changes of our choice of spatiotemporal scales. Sensitivity tests were performed with a trajectory integration time of 11 and 10 h, respectively, and seeding within a radius of 2.4° and 2.2° , respectively. Only when reducing the radius to 2.2° and only for the XT stage, a qualitative change in the identified clusters occurs. This change can be explained by no longer identifying a cluster that largely comprises trajectories at relatively large radii at seeding time [i.e., outside of the 2.2° radius (cluster ATRC)], and instead identifying an alternative cluster that is a mix of two existing clusters (clusters INUP and OUT; see below for a description of the respective clusters).

3. Clustering under symmetry and large numbers of trajectories

Any cluster analysis crucially depends on how similarity in the underlying data are defined. Our first steps described below serve to adopt the definition of similarity to the problem of trajectories in TCs. We then describe our strategy to handle the very large amount of trajectories that underlies our analysis and our choice of cluster algorithm. Finally, we discuss the representation of the obtained clusters in physical space, which is a nontrivial task in our case because clustering is performed in a transformed, normalized space (see below) and trajectories in the same cluster thus do not necessarily form coherent bundles of trajectories in physical space.

FIG. 4. Schematic illustrating the difference between Euclidean and Fréchet distance of two trajectories T_1 and T_2 (thick solid curves). Euclidean distance is based on the pointwise distance between trajectory positions at each a given time step (dashed). The Fréchet distance is not constrained to comparing trajectory positions at a given time but rather considers the maximum divergence of two trajectory paths (thick solid line, see text for a more formal definition.)

a. Defining similarity of trajectories in the TC context

1) ACCOUNTING FOR TRANSLATION AND SYMMETRIES

As discussed in the introduction, TCs exhibit a relatively large degree of symmetry and we argue that similarity of trajectories within TCs is best measured by similarity of the shape of the trajectory and not primarily by proximity of trajectories in physical space. To account for the underlying approximate symmetries the following steps are performed to preprocess the data before clustering (Fig. 3):

- (i) Taking into account TC motion, trajectories are considered in a storm-relative frame of reference. The storm

center, which constitutes the origin of this frame of reference, is defined by the low-level center as described above (section 2d).

- (ii) Taking translational invariance into account, the seeding locations of all trajectories are moved to the same location in the horizontal plane, arbitrarily defined as the origin.

- (iii) Taking rotational invariance into account, trajectories are rotated in the horizontal plane and the direction of all trajectories is aligned in the same direction at seeding time, arbitrarily defined as the positive rlon axis.

2) NORMALIZATION AND FRÉCHET DISTANCE

Taking into account the vastly different length scales for horizontal and vertical motion, the transformed trajectories are normalized by the standard deviation of the trajectory coordinates in the respective directions (following Hart et al. 2015). Standard deviations in rlat and rlon are reasonably similar (within 5%–10%) so that normalization may be performed after the transformation of trajectories. Specifically, in the following we consider deviations of the trajectories from their seeding location normalized, for the sake of symmetry, by the mean of the standard deviations at the start and the end times of the trajectories. Importantly, considering deviations only imply that the absolute height of the trajectories is not taken into account to define similarity, but only height changes (i.e., the characteristics of ascent and descent).

A widely used metric to measure distance between trajectories is the Fréchet distance, which we employ in this study also. While the Euclidean distance is based on the pointwise distance between air parcel locations at each given time step, the (discrete) Fréchet distance does not consider such pointwise distances at a given time (Fig. 4). Instead, the Fréchet

FIG. 5. Illustration of the the Visvalingam-Whyatt algorithm. (a) The black line shows a discrete trajectory represented by points 1–6. (b) To approximate the trajectory with the least loss of information the Visvalingam-Whyatt algorithm considers the triangles formed by three consecutive points (color-shaded regions) and successively removes the middle point along the trajectory associated with the triangles with the smallest area (triangle A_5 and middle point 5 in this example). (c) Sample application of the Visvalingam-Whyatt algorithm to one of the trajectories from the cluster large CURV . For illustration purpose we have selected a quasi-horizontal trajectory. It is important to note, however, that the algorithm is applied in this study in 3D. The original trajectory is depicted in blue and a representation with 45 instead of 145 data points is depicted by the dashed light-gray line. These two representations are hardly distinguishable. Trajectory representations with 15, 13, and 11 data points are depicted by the dashed medium-gray, dashed dark-gray, and black solid lines, respectively. The approximations that are made when reducing the number of points from 15 to 13 and from 13 to 11 are highlighted by the gray and black arrows, respectively.

FIG. 6. (a) Illustration of MDS. For visual clarity, only two trajectories are shown to illustrate the basic concept. MDS transforms the trajectories from (left) physical space [with coordinates (x, y, z)] to (right) points in a low-dimensional, normalized space [with coordinates (x_1, x_2)]. The transformation is performed under the constraint that the Fréchet distance between trajectories (solid arrow) is optimally approximated by the Euclidean distance between the points in the low-dimensional space (dashed arrow). (b) LMDS uses a subset of trajectories as so-called landmarks and MDS is performed for this subset only [as illustrated in (a), these trajectories are depicted in (b) in black]. Additional trajectories (red, yellow) are embedded as points in the low-dimensional space by approximately preserving their distance *with respect to these landmarks*. For the purpose of clustering, the pairwise distances *between* these additional trajectories do not need to be calculated by their Fréchet distance in physical space but can be approximated with low computational cost by their Euclidean distance in the low-dimensional space (dashed orange arrow).

distance considers reparameterizations (in time) of the trajectories and then seeks the minimum of the maximum distance between trajectories under all reparameterizations (Fréchet 1906; Eiter and Mannila 1994). More illustratively, consider two air parcels that travel along the *path* of their respective trajectory with arbitrary velocities.² Let the two air parcels be connected by an imaginary string. The Fréchet distance equals the length of the shortest string that still allows both air parcels to travel along their respective path when the velocities are optimally matched. Thus, the Fréchet distance emphasizes the shape of trajectories while discounting differences in the trajectories' temporal evolutions. Due to these characteristics, we consider the Fréchet distance to be the preferred, if not the optimal distance metric in the context of this study.

b. Handling very large numbers of trajectories

Investigating a very large number of trajectories (here: order of 10^6) makes the computation of trajectory similarity by pairwise comparison of their distance during clustering computationally

extremely expensive. To make a cluster analysis feasible, we first use a suitable approximate representation of trajectories to reduce the sheer amount of data and then a suitable reduction of dimensionality to make the comparison of the distances between trajectories computationally feasible.

1) DATA COMPRESSION

High-frequency oscillations around the more slowly varying “mean” path of trajectories should not affect the overall similarity of the shape of trajectories. In our case, high-frequency oscillations occur [e.g., as gravity waves in the outflow layer that manifest as vertical fluctuations around an approximately constant mean trajectory height (not shown)]. Filtering out the high-frequency oscillations effectively allows a representation of the trajectories by fewer data points than in the original dataset (i.e., a reduction of the amount of required data).

Filtering out high-frequency oscillations while maintaining sharp but persistent turns of trajectories (i.e., large curvature) can be obtained in an optimal sense by using the Visvalingam–Whyatt algorithm (Visvalingam and Whyatt 1993), a well-known algorithm originally developed in the field of cartography. This algorithm considers the triangles that are formed by three consecutive points along a (discrete) trajectory (see Fig. 5 for illustration). The main idea of the Visvalingam–Whyatt algorithm is to eliminate triangles with small area. For trajectories sampled at constant time intervals, as it is the case here, triangles with small areas are a key property of slow velocities and/or small curvature. Removing all middle points³ along trajectories that are associated with triangles with small areas thus preserves the overall geometry of the trajectory while reducing the amount of data points needed for its representation.

Subjectively, we choose to reduce the number of points representing the trajectory from 145 to 11. This choice is a trade-off between accuracy⁴ and the need for data compression when using the costly (discrete) Fréchet distance, which scales as $\mathcal{O}(M^2 \log M)$, where M is the number of data points representing the trajectory.

2) REDUCTION OF DIMENSIONALITY

Essentially, any cluster algorithm requires information about the pairwise distances between all trajectories to define individual clusters and cluster membership of individual trajectories. The computation of these pairwise distances scales as $\mathcal{O}(cN^2)$, where c is the cost of computing the distance metric and N is the number of trajectories. For large N , this computation is prohibitively expensive; in particular, if an elaborate metric such as the Fréchet distance is used (large c). A main goal here is to make feasible the clustering of a large number of trajectories using an elaborate distance metric. To this end, we dramatically

³ In our case: except for the seeding point.

⁴ On average, the difference between simplified and original trajectories in our case decreases exponentially with increasing number of points. Our choice of 11 points lies in the transition region between the rapid decrease of accuracy for smaller numbers and the high accuracy of the simplification achieved with a higher number (>20) of points.

² The Fréchet distance comprises a monotonicity constraint: The velocities need to be positive into the direction of the trajectory path.

FIG. 7. (a) Illustration of 30 randomly selected landmarks (crosses) and 300 randomly selected embedded trajectories in normalized low-dimensional (3D) space. (b) Illustration of clusters (colors) in normalized low-dimensional (3D) space. Each cluster is represented by 1000 randomly selected members.

reduce the computational cost by approximating these pairwise distances by Euclidean distances in a low-dimensional space and by using the costly Fréchet distance with respect to a subset of the trajectories only.

The first key idea of the approximation is to transform trajectories into a low-dimensional space, in which the trajectories behave with respect to their pairwise distances as similarly as possible as the original trajectories.⁵ This transformation may be performed by a technique that is well-known in the field of information visualization: so-called multidimensional scaling (MDS; Kruskal 1964, here illustrated in Fig. 6a). MDS is conceptually similar to how principal component analysis is usually applied in a meteorological context (e.g., Wilks 2011, chapter 11): MDS, however, performs the reduction of dimensionality by retaining only the leading eigenvalues of a matrix that contains the pairwise distances, whereas principal component analysis considers the covariance matrix. Consequently, MDS requires the knowledge of all pairwise distances, which would render the approach irrelevant for our application. The need to eliminate this prohibitive requirement introduces a second key idea: to apply MDS only for a small subset of l “landmark” trajectories (with $l \ll N$). This approximate version of MDS, so-called landmark MDS (LMDS; De Silva and Tenenbaum 2004, here illustrated in Fig. 6b), requires the calculation of pairwise distance between the l landmarks only. The remaining trajectories are approximately embedded into the low-dimensional space using triangulation with these landmarks as reference points (i.e., the pairwise distances between the remaining individual trajectories and all landmarks are approximately preserved). The latter step

requires the calculation of pairwise distance between all remaining trajectories and the l landmarks. Pairwise distance between all remaining trajectories is then computed with low computational cost in low-dimensional space. A more detailed and technical description of LMDS can be found in De Silva and Tenenbaum (2004).

There are several strategies to define the landmarks. Following De Silva and Tenenbaum (2004), we use a simple and computationally most efficient strategy and sample landmarks uniformly in the low-dimensional space. To yield the best low-dimensional representation, the number of landmarks l should be as high as computationally possible. Here, we choose $l = 1000$. A further choice to be made is the dimension k of the low-dimensional space. In a similar application, albeit for streamlines of steady-state flows, Theisel and Rössl (2012) have successfully applied MDS. Following Theisel and Rössl (2012), we set $k = 3$. Using $k = 15$, which includes all

FIG. 8. Calinski-Harabasz index (y axis, in 10^5) as a function of cluster number (x axis) for each stage as color coded. The optimal number of clusters is indicated by a distinct maximum of the index. The absolute value of the index is not directly meaningful.

⁵ The distance metric in low-dimensional space may be different from the original metric. We here use Euclidean distance in low-dimensional space. The final results of our cluster analysis are essentially insensitive to the choice of this distance metric.

FIG. 9. Representation of clusters smallCURV (purple) and largeCURV (orange) at the TS stage. The shear vector at this stage is given in (a). The spatial distribution of the (a),(d) initial and (c),(f) end positions are depicted by the contour levels of occurrence frequency that encompass 35%, 50%, 65%, and 80% of the data (shaded). The horizontal distribution (a),(c) is shown in a storm-centric view. Radius–height sections are shown in (d) and (f). (b) The shape of the horizontal projection of the cluster mean trajectory is depicted, with the seeding point centered on the origin and the trajectory aligned in the positive rlon axis at seeding time. Crosses mark the starting points of the mean trajectories. The mean trajectory is scaled back from normalized space using the average of the standard deviations of rlon and rlat. (e) The time series of the height of the mean trajectory is depicted relative to the seeding time (0 h).

eigenvalues that are at least 5% of the magnitude of the largest eigenvalue, leads to virtually identical results of the cluster statistics (not shown).

Computation of the pairwise distances is computationally the limiting factor in a cluster analysis. As noted above, without reduction of dimensionality this computation scales as $\mathcal{O}(cN^2)$. Using LMDS, the leading term in the scaling is $\mathcal{O}(cN)$, which implies a reduction of computational cost by a factor of N/l , here 10^3 . Still, on an Intel i7–6800k CPU with 3.4 GHz our cluster analysis for a single stage takes approximately 1 day. A further term in the scaling of landmark multidimensional scaling is $\mathcal{O}(l^3)$. This cubic dependence on l illustrates that the number of landmarks is severely limited and that substantially increasing our choice $l = 10^3$ would be computationally unreasonable.

c. Clustering and representation of clusters in physical space

1) CLUSTERING

Clustering is performed in the low-dimensional space. Nevertheless, the very large amount of trajectories considered in this study leaves only two choices concerning the cluster

algorithm: DBSCAN and k -means. Other clustering algorithms [e.g., hierarchical clustering used for trajectory clustering by Hart et al. (2015)], would be prohibitively expensive. Because the representation of trajectories in low-dimensional space is compact and dense (Fig. 7a), DBSCAN, which is a density-based cluster algorithm, yields only one single cluster (not shown). We thus adopt in this study the only suitable algorithm for our problem: the widely used k -means algorithm. In contrast to DBSCAN, k -means identifies by design a prescribed number of clusters. The seemingly arbitrary partitioning of a dense cloud of data into individual clusters (Fig. 7b) is a well-known feature of k -means. We will demonstrate below, however, that the resulting clusters are indeed physically distinct and meaningful.

To select the number of clusters at each of Karl's stages, we have considered several metrics of cluster quality. It turned out that the so-called Calinski–Harabasz index, which is essentially the ratio of the overall intracluster and the intercluster variances, yields the sharpest signal. With this metric, distinct peaks indicate the preferred number of clusters. Such peaks are found for the TS, IS, and XT stage, indicating 5, 6,

FIG. 10. As in Fig. 9, but for clusters INUP (purple) and OUT (orange).

and 5 clusters,⁶ respectively (Fig. 8). At the IM stage, the Calinski–Harabasz index is basically flat between 2 and 6 clusters. Temporal consistency with the other stages suggests a choice of 5 or 6 clusters. We have chosen 6 clusters at the IM stage. Increasing the number of clusters from 5 to 6 turned out to leave the mean representation of 4 clusters virtually unchanged but splits the fifth cluster into two separate clusters⁷ that can both be identified at the previous IS stage also (not shown).

2) REPRESENTATION IN PHYSICAL SPACE

Representing the clusters in physical space is a nontrivial task. The similarity of trajectories within a cluster pertains to the transformed space (i.e., after translation and rotation of individual trajectories). In physical space, however, individual trajectories within one cluster may differ substantially. As discussed in the introduction, the naive approach of representing the clusters by a mean trajectory in physical space may therefore often be misleading. The results below will thus be presented by showing the

distribution of start and end points of trajectories in physical space, as well as the cluster-mean trajectory in normalized, transformed space. This mean trajectory is well suited to represent the *shape* of trajectories in a given cluster. The physical interpretation is that air masses within a given cluster move from the initial positions to the end positions following a path that is represented by the shape of the cluster-mean trajectory.

4. Clustering results and temporal evolution of main airstreams

We first present the clusters that are identified during the TS stage. Subsequently, we describe how these clusters evolve with time. In addition, a qualitatively new cluster will be described that emerges with the increase of vertical wind shear, and thus storm-relative flow, during the IS stage.

a. Clusters during the TS stage

The clusters identified during the TS stage represent (i) quasi-horizontal, cyclonic swirling around the center, (ii) spiraling inflow, rapid ascent, and anticyclonic outflow, and (iii) slow, corkscrew-like ascent. These clusters meet the expectations that a tropical meteorologist may have for airstreams in the vicinity of a weak TC in a nearly quiescent environment and can thus be considered to be physically meaningful.

1) QUASI-HORIZONTAL, CYCLONIC SWIRLING

Figure 9 illustrates the two clusters that represent quasi-horizontal, cyclonic swirling. The clusters are distinguished by

⁶ When clustering is performed using Euclidean distance, the Calinski–Harabasz index tends to indicate a lower number of optimal clusters (for the IS stage, e.g., four instead of six clusters). Our interpretation of this tendency is that the Fréchet distance does indeed provide a sharper distinction between trajectory shapes, which further supports the use of the Fréchet distance in this application.

⁷ Below, these two clusters will be referred to as clusters INUP and CORK, respectively.

FIG. 11. As in Fig. 9, but for cluster CORK.

the magnitude of the cyclonic curvature of trajectories (Fig. 9b). The cluster with smaller curvature will be referred to as *small*CURV and the cluster with larger curvature as *large*CURV. Trajectories in *small*CURV and *large*CURV hardly undergo, on average, any vertical displacements within the considered time period of 12 h (Fig. 9e). Trajectories in *large*CURV start and end at smaller radii and encircle the center almost completely, whereas trajectories in *small*CURV start and end at larger radii and describe on average a third of a circle (Figs. 9b,d,f).

The spatial distribution of the start and end positions of *large*CURV are located in an annulus around the center, approximately between 1° and 2° radius (Figs. 9a,c). In contrast, the start positions of trajectories in *small*CURV exhibit a bimodal distribution, split into a northern branch that moves west of the center toward the south, and a southern branch that moves east of the center toward the north, mostly outside of 2° radius. Trajectories mostly occur up to 6 and 8 km height in *large*CURV and *small*CURV, respectively (Figs. 9d,f). Trajectories that occur below 1.5 km height tend to move radially inward, consistent with the existence of a frictional inflow layer.

2) INFLOW, RAPID ASCENT, AND ANTICYCLONIC OUTFLOW

Figure 10 illustrates the two clusters that represent inflow, rapid ascent, and anticyclonic outflow. One cluster comprises spiraling inflow and rapid ascent (hereafter referred to as INUP) and one cluster the outflow anticyclone (hereafter

referred to as OUT; Figs. 10b,e). In INUP, the majority of trajectories originate from below 2 km height and from within 4° radius, then rise into the upper troposphere (on average to 12 km height) and remain during the considered time period within 2° radius from the center (Figs. 10a,c,d,f). Trajectories in OUT originate from the upper troposphere, mostly between 9 and 12 km height and predominantly from above and from the southwest of the center. At the end of the TS stage, most of this outflow air is found to the downshear-right between 1° and 4° radius and at heights between 8 and 14 km.

3) SLOW, CORKSCREW-LIKE ASCENT

In the fifth cluster, trajectories ascend slowly while swirling cyclonically around the center (Figs. 11b,e). This movement implies corkscrew-like ascent and thus the cluster will be referred to as CORK. Trajectories in CORK originate from close to the center (within 1°–2° radius) and predominantly from below 8 km height (Figs. 11a,d).

Trajectories remain within this radial range and ascend in the considered 12 h time period on average only by 2 km (Figs. 11d,e,f), despite their location in the inner core. This trajectory behavior is consistent with the observation by Euler et al. (2019) that a substantial part of Karl's inner-core vertical mass flux does not reach the outflow layer at this stage. Air parcels that detrain from and entrain into inner-core updrafts and that spent most of the 12 h period in the eye may contribute to this cluster. It is interesting to note that the characteristics of

FIG. 12. As in Fig. 9, but for cluster $_{\text{largeCURV}}$ only and depicting the temporal evolution from the TS to the XT stage. (a),(c),(d),(f) The gray shading shows the distribution of the initial and end positions at the TS stage. A single contour line representing the distributions at later stages is shown in color (IS, IM, and XT stage in purple, red, and orange, respectively). This representative contour line has been chosen subjectively to represent the main features of the full distribution and here is that encompassing 75% of the data. (b),(e) The mean trajectory at the respective times is shown in as in Fig. 9 and is color coded accordingly. The shear vector, color coded for the respective time, is given in (a).

Karl's inner-core convection manifest themselves as two distinct clusters, INUP and CORK, in the current cluster analysis.

b. Temporal evolution of clusters

The cluster analysis presented herein has been performed independently for each of Karl's stages. Subsequently, clusters are tied together across consecutive stages by using the minimum of the pairwise Fréchet distances between the cluster mean trajectories as an objective criterion.⁸ The analysis thus

does not, per se, require a gradual, coherent temporal evolution of the individual clusters. Therefore, it is most notable that the identified clusters do exhibit such a gradual evolution, strongly supporting the notion that the clusters represent physically meaningful airstreams.

1) CLUSTERS WITH QUASI-HORIZONTAL MOTION

During all of Karl's stages considered herein, the mean trajectories of $_{\text{largeCURV}}$ and $_{\text{smallCURV}}$ remain very similar (Figs. 12b,e and 13b,e). Besides these similarities, small gradual changes with time are evident. One gradual change is a decrease in cyclonic curvature. In addition, for $_{\text{largeCURV}}$, gradual changes comprise an increase in mean height and increasing ascent during the second part of the trajectory. For $_{\text{smallCURV}}$, there is a gradual increase of the length of the mean trajectory, reflecting the increase of vertical shear from the TS to the XT stage. With increasing strength of the environmental flow, trajectories originate and end at larger distance from the center (Figs. 13d,f).

The spatial distribution of the start and end positions exhibit more prominent, but still gradual changes. Most notably, in $_{\text{largeCURV}}$, the distributions of the start and end positions develop with time an increasing degree of asymmetry

⁸ The only exception is cluster OUT and cluster ATRC at the transition from the TS to the IS stage. Cluster ATRC at the IS stage is most similar to cluster OUT at the TS stage. The mean trajectory of cluster OUT, however, clearly exhibits anticyclonic curvature throughout the whole trajectory, whereas the mean trajectory of cluster ATRC first exhibits very little curvature and then exhibits cyclonic curvature during the later part (Fig. 14b). Due to this qualitative difference we introduce cluster ATRC as a new, distinct cluster at the IS stage. For cluster OUT at the IS stage, cluster OUT at the TS stage is the most similar cluster. Therefore, and due to the close Lagrangian connection of the in-up-out circulation, we define cluster OUT at the IS stage as the continuation of cluster OUT at the TS stage.

FIG. 13. As in Fig. 12, but for cluster smallCURV. The subjectively chosen representative contour line has been adapted to best represent the main features of the full distribution and here is that encompassing 67% of the data.

(Figs. 12a,c) and extend to larger radii and heights (Figs. 12d,f). During the IM and the XT stage, the end positions indicate an increasing transport of air masses into the downshear region. The increasing asymmetry and increasing downshear transport are a consequence of the gradual increase of vertical shear from the TS to the XT stage. Increasing vertical shear implies increasing storm-relative flow (e.g., Willoughby et al. 1984), which in turn implies (i) an increasing deformation of storm-relative streamlines and (ii) a decreasing area within a dividing streamline (i.e., less air is being rotationally constrained to circulate around the vortex center) (e.g., Riemer and Montgomery 2011).

In smallCURV, the bimodal distribution of the start and end positions becomes more distinct with time (Figs. 13a,c). In strong vertical shear (IM and XT stage), the start and end positions clearly exhibit an upper- and lower-tropospheric branch (Figs. 13d,f). The occurrence of these two branches (i.e., a lack of midtropospheric trajectories in this cluster), can be explained based on the general kinematic structure of vertical-shear flows (Willoughby et al. 1984; Riemer and Montgomery 2011). Tropospheric-deep vertical shear implies that the storm-relative environmental flow changes sign at some midtropospheric steering level, with downshear flow in the upper troposphere and upshear flow in the lower troposphere. Small storm-relative environmental flow in the vicinity of the steering level, combined with Karl's vortical flow, evidently leads to a trajectory behavior that is distinct from that in

the lower and upper troposphere, where the combination of Karl's vortical flow and prominent storm-relative environmental flow leads to similar curvature of trajectories. The configuration of storm-relative environmental flow further implies that air masses tend to move upshear in the lower troposphere and downshear in the upper troposphere, relative to Karl's center. Consequently, the branch with initial positions downshear (-left) and end positions upshear (Figs. 13a,c) can be identified with the lower-tropospheric branch. Accordingly, air masses in the upper-tropospheric branch move from up- to downshear. Figure 13 further indicates that the lower-tropospheric branch dominates.

Notably, with the advent of considerable vertical shear at the IS stage, a qualitatively new cluster emerges and persists through the IM and the XT stage. At the IS stage, the mean trajectory of this cluster exhibits little vertical motion and curvature, with some cyclonic curvature during the second half of the trajectory (Figs. 14b,e). Trajectories originate from the mid- to upper troposphere (mostly between 6 and 12 km height) and from a widespread region upstream of the center (Figs. 14a,d). Trajectory end positions are mostly right of shear and between 1° and 3° radius (Figs. 14c,f). The mean trajectories at the IM and the XT stage are similar to the IS stage but (i) exhibit more cyclonic curvature during the later part of the trajectory, (ii) start from gradually lower heights, and (iii) indicate weak descent (Figs. 14b,e). The start positions exhibit

FIG. 14. As in Fig. 13, but for the cluster ATRC and the IS to the XT stage only.

more distinct changes and are found at larger radii in the upshear region during the IM and the XT stage (Figs. 14a,d). In contrast, the end positions change relatively little and appear much more localized than the start positions. In reference to this attracting nature of the trajectory behavior, we term this cluster ATRC.

Cluster ATRC represents air masses that are transported toward the storm center from large radii. With increasing storm-relative flow (IM and XT stage) trajectories originate from larger radii upshear. Getting closer to the center, the trajectories get increasingly influenced by the storm's cyclonic flow. In particular, trajectories enter a region of confluent flow (in terms of flow topology: an attracting manifold (Riemer and Montgomery 2011)), which arises from the superposition of vortical and storm-relative flow (cf. e.g., Fig. 18 in Willoughby et al. 1984). Arguably, this confluence organizes the trajectories originating from a widespread region upshear into the more localized region of end positions observed in Fig. 14c.

2) CLUSTERS WITH RAPID ASCENT

As discussed above, the in-up-out circulation of Karl at the TS stage is represented by inflow and rapid ascent in cluster INUP and upper-tropospheric anticyclonic flow in cluster OUT. Cluster INUP persists through the IS and the IM stage, with strongly curved cyclonic inflow and subsequent rapid ascent (Figs. 15b,e), but is no longer identified at the XT stage. With the increasing vertical shear during the IS and the IM

stage, the source region of the inflow becomes increasingly asymmetric and extends out to larger radii (Figs. 15a,d), with inflow from relatively large radii originating predominantly from right of shear. The end positions of the trajectories remain in the vicinity of the center but gradually shift downshear and to larger radii (Figs. 15c,f). These changes are consistent with the increase of shear-induced storm-relative flow in the lower and upper troposphere (i.e., below and above a midlevel steering level). Furthermore, it is interesting to note that trajectories reach lower heights at the IM stage than before (Figs. 15e,f). This change is consistent with the increasingly detrimental thermodynamic environment in which Karl moves during this stage (Euler et al. 2019).

At the IS, the IM, and the XT stage, upper-tropospheric anticyclonic flow in cluster OUT is no longer separated from the rapid ascent below (Fig. 16). In the low-shear environment at the TS stage, upper-tropospheric flow is weak enough such that the outflow anticyclone may remain in the vicinity of the center (Fig. 10). The increased vertical shear during the later stages implies, however, that outflow air is rapidly advected downshear (Figs. 16b,c). Comparing the general characteristics of cluster OUT at these later stages with the general characteristics of cluster INUP at the earlier stages (the TS, the IS, and the IM stage), cluster OUT exhibits a less pronounced cyclonically curved inflow (cf. Figs. 16b and 15b), earlier rapid ascent (cf. Figs. 16d and 15d) and a distinctly larger downshear displacement in the upper troposphere (cf. Figs. 16c,f and

FIG. 15. As in Fig. 12, but for cluster INUP and the TS to the IM stage only.

15c,f). Similar to cluster INUP, the start positions of trajectories become increasingly asymmetric and the outflow gradually reaches lower altitude (Figs. 16a,e).

It is interesting to note that the distinction between clusters INUP and OUT diminishes in the strong-shear environments of the IM and the XT stage: the start height of the mean trajectory of cluster OUT gradually decreases because increasingly more trajectories start at low levels and in the frictional inflow layer, in particular during the XT stage (Figs. 16d,e). In addition, the mean trajectory starts ascending at later times. Arguably, the lack of a distinct INUP cluster at the XT stage is partly due to these diminishing differences.

3) CLUSTER WITH SLOW, CORKSCREW-LIKE INNER-CORE ASCENT

At all stages, cluster CORK exhibits mean trajectories that swirl cyclonically around the center at small radii and that ascend slowly (Figs. 17b,e). The ascent increases from the TS to the XT stage. Start positions of trajectories are mostly from below 6–8 km height and within 2° radius above the frictional inflow layer. Within the inflow layer, trajectories originate also from larger radii. From the TS to the XT stage, trajectories originate from increasingly larger radii within the inflow layer (Figs. 17a,d) and the fraction of inflow-layer trajectories increase (Fig. 17d). Within the inflow layer, trajectories mostly originate from right of shear and exhibit an increasing tendency to end in the downshear region (Figs. 17a,c).

5. Avenues for further analysis: Mean thermodynamic characteristics of clusters

The analysis of Karl's major airstreams provides a kinematic perspective of the evolution. An important further component of a more complete analysis is how thermodynamic properties evolve along these airstreams. To gain more insight into specific aspects of Karl's evolution, and also to illustrate avenues for further analysis based on automatically identified major airstreams, we here consider the evolution of temperature T and moist entropy along trajectories of selected clusters. A more comprehensive thermodynamic analysis is beyond the scope of this study.

The evolution of thermodynamic properties along airstreams can be described succinctly by the associated mass flux in thermodynamic space (Kjellsson et al. 2014; Riemer and Laliberté 2015), here namely temperature (T)–moist-entropy space. For simplicity, we use θ_e as moist-entropy variable [calculated following Bolton (1980)]. The interested reader is referred to Eq. (1) in Riemer and Laliberté (2015) for a formal definition of the thermodynamic mass flux vector. In a nutshell, mass flux directed from low to high T and θ_e manifests an increase of T and θ_e along trajectories, respectively. Accordingly, the mass flux is directed in the opposite direction when T or θ_e decreases along trajectories. Due to a strong dependence on pressure, and thus height, T may serve to a first approximation as vertical coordinate. Two specific examples: (i) mass flux from high to low T at constant θ_e (e.g., around $\theta_e = 350$ K in

FIG. 16. As in Fig. 14, but for cluster OUT.

Fig. 18a) signifies moist-adiabatic ascent and associated cooling by expansion, and (ii) mass flux from low to high θ_e at high T (e.g., for $T > 290$ K and $\theta_e < 350$ K in all panels of Fig. 18) signifies nonconservative increase of θ_e by surface fluxes and turbulent vertical mixing in the boundary layer.

We first consider the thermodynamic characteristics that are associated with the intensification of Karl in increasing vertical shear. Vertical shear increasing from small to moderate values tends to organize convection in weak tropical cyclones (e.g., Molinari et al. 2004) and has been demonstrated to increase inner-core vertical mass flux during the onset of ET (Davis et al. 2008). Euler et al. (2019) have shown that Karl's intensification is associated with such an increased vertical mass flux of relatively high- θ_e air (here illustrated in Figs. 18a,b vs c,d; the increased mass flux is most directly evident in the mid to upper troposphere, i.e., for $T < 270$ K). The new aspect revealed by the current analysis is that not only the rapid, tropospheric-deep convection contributes to this increased mass flux (the combined clusters INUP and OUT; Figs. 18a,c). In addition, the slower corkscrew-like ascent makes a substantial contribution (cluster CORK, Figs. 18b,d). A closer inspection (not shown) reveals that the mass flux in the mid- to upper troposphere increases from the TS to the IS stage by approximately 50% in the combined clusters INUP and OUT and by approximately 100%–150% in cluster CORK. Arguably, in the case of Karl, increasing vertical shear does not only invigorate deep convective ascent but invigorates also the more slowly ascending airstream.

Thermodynamic characteristics of selected clusters elucidate also aspects of Karl's evolving structure during the onset of ET. During the XT stage (Figs. 18e,f), the respective thermodynamic mass flux of the combined clusters INUP and OUT and of the cluster CORK remain qualitatively the same as during the tropical stages of Karl (the TS and IS stages, Figs. 18a–d). From this perspective, Karl's ET at this stage does not imply fundamental changes of the thermodynamic characteristics along these ascending airstreams. Quantitatively, however, ascent during the XT stage occurs with substantially (10–8 K) lower θ_e and ends at substantially higher T (mostly $T > 230$ K instead of $T < 220$ K) than during the tropical stages. We further note that cluster CORK exhibits two ascending branches: one at $\theta_e > 350$ K and one at $\theta_e < 345$ K (Fig. 18f). Euler et al. (2019) has demonstrated that Karl's inner core, notwithstanding Karl's weak intensity, acts as a containment vessel that transports high- θ_e air into the increasingly hostile environment during ET. The current analysis indicates that this high- θ_e air contributes to latent heat release during the XT stage as part of the CORK airstream. Further analysis, however, of such intriguing but more detailed features of Karl's ET are beyond the scope of the current study.

In contrast, the thermodynamic characteristics along the quasi-horizontal airstreams in cluster ATRC and cluster smallCURV exhibit substantial qualitative changes (Fig. 19). During the tropical stages, here exemplified by the IS stage (Figs. 19a,b), the thermodynamic mass flux is generally small and will not be further

FIG. 17. As in Fig. 12, but for cluster CORK.

considered here. During the XT stage, however, cluster ATRC exhibits prominent warming at approximately constant θ_e (for $\theta_e < 340$ K, Fig. 19c) and cluster smallCURV exhibits a prominent increase of θ_e at high T for low- θ_e air (for $\theta_e < 330$ K, Fig. 19d). Arguably, these features are due to adiabatic compression of descending air masses along the baroclinic zone (cluster ATRC) and due to surface fluxes and turbulent mixing in the boundary layer (cluster smallCURV). Further exploration of the significance of these features and comparison with other ET cases are potentially fruitful tasks for future studies.

6. Summary and conclusions

This study presents a trajectory-clustering procedure tailored to TCs and to handle a very large number of trajectories [$\mathcal{O}(10^6)$]. Clustering is performed in a normalized, transformed space to take into account the large degree of axisymmetry and translational invariance of the TC circulation. The Fréchet distance, which emphasizes differences in the shape of trajectories, is used to define similarity between transformed trajectories. To adequately resolve trajectories in a TC's convectively dominated inner core requires high spatiotemporal resolution. Considering the larger-scale flow at the same time may readily yield a very large number of trajectories. A cluster analysis may thus become computationally prohibitive expensive.

To make cluster analysis computationally feasible, we perform data compression by the Visvalingam and Whyatt (1993) algorithm, which reduces the number of data points to represent

a trajectory while approximately preserving the general shape of a trajectory in an optimal sense. A further reduction of dimensionality is achieved by landmark multidimensional scaling (De Silva and Tenenbaum 2004). Exact pairwise distances between trajectories, required for any cluster algorithm, are calculated for a small subset of trajectories only. These so-called landmarks are used to construct a low-dimensional space that optimally approximates the pairwise distances. For the remaining trajectories, pairwise distances are calculated only with respect to the landmarks. The remaining trajectories are then approximately embedded in the low-dimensional space using distance-based triangulation with the landmarks as reference points. Clustering of all trajectories is then performed based on the approximated distances between trajectories in the low-dimensional space using the k -means algorithm. These approximations reduce the computational cost of the clustering procedure by a factor of $\mathcal{O}(10^3)$. Only the clustering procedure is subject to approximations. The actual analyses presented herein show the full unapproximated trajectories.

The ET of Tropical Storm Karl (2016) is investigated to demonstrate the applicability of our clustering procedure. This case has been a focus of a major field campaign and has been linked to a high-impact weather event in the downstream region (Schäfer et al. 2018; Kumpf et al. 2019). In addition, Euler et al. (2019) have investigated inner-core airstreams in Karl that are associated with latent heat release. The focus of the current study is on the more general behavior of airstreams in

FIG. 18. Streamlines of mass flux in thermodynamic space, namely θ_e (horizontal axis, values increasing from right to left) and T (vertical axis, values increasing from top to bottom) space. The magnitude of the mass flux is scaled with the mass m^* of an individual air parcel represented by our trajectories (with $m^* \approx 1.6 \times 10^9$ kg). (a),(c),(e) The mass flux of the combined clusters INUP and OUT and (b),(d),(f) the mass flux in cluster CORK. Note the different scale of the color bars. (a),(b) The TS stage; (c),(d) the IS stage; and (e),(f) the XT stage.

the vicinity of Karl, specifically: within a cylinder of 2.5° radius and for a temporal evolution of 12 h. The choice of these scales is motivated by the typical scales for environmental interaction of TCs and associated intensity change. The presented results are representative for this combination of typical scales. A major change to one of these scales, however, may change the physical problem under consideration and may lead to qualitatively different results.

Our cluster analysis is applied to four key stages of Karl's evolution, as defined by Euler et al. (2019). At all stages, the analysis yields distinct and physically meaningful clusters. The clusters represent (i) quasi-horizontal cyclonic swirling of air

masses around the center, subdivided into clusters with large and small trajectory curvature, (ii) the in-up-out secondary circulation, subdivided into clusters with inflow and rapid ascent, and rapid ascent and outflow, and (iii) slowly ascending air masses in the inner core that describe corkscrew-like ascent. An additional cluster emerges with the advent of substantial vertical wind shear and thus storm-relative flow. In this cluster, air masses pass Karl on quasi-horizontal paths with little curvature. Notably, the identified clusters exhibit a gradual and coherent evolution from stage to stage. Our cluster analysis, however, is performed independently for each stage and thus does not, per se, require such a gradual, coherent evolution.

FIG. 19. As in Fig. 18, but for (a),(c) cluster ATRC and (b),(d) cluster smallCURV. (a),(b) The IS stage and (c),(d) the XT stage. Note that the range of θ_e values has changed with respect to Fig. 18.

We therefore interpret the observed gradual temporal evolution as further evidence that the cluster analysis yields physically meaningful results.

The identified clusters provide an objective and automatable overview of the main airstreams in the vicinity of TCs. Analysis based on such clusters may be particularly useful when intercomparing several different cases or when analyzing many cases (e.g., in an ensemble framework). While the approach proposed herein lacks the details found in studies that focus on more specific processes (e.g., Cram et al. 2007; Stern and Zhang 2013; Euler et al. 2019), such studies usually define airstreams subjectively by introducing threshold criteria for several variables that are tailored to the intensity and structure of a specific TC. In addition, that approach tends to neglect those air parcels that do not fulfill the specific criteria. Taking into account the full Lagrangian solution (i.e., all trajectories within a given region) as in the approach presented herein, may be beneficial for some applications.

A brief analysis of thermodynamic characteristics along main airstreams provides insight into aspects of Karl's intensification in increasing vertical shear and of structure change during Karl's ET. While a more comprehensive analysis is beyond the scope of this study, this brief analysis may illustrate avenues for future analyses based on automatically identified main airstreams. Such Lagrangian analyses seem most fruitful for TC problems, in which flow asymmetries and the interaction with environmental air have been demonstrated to be of crucial importance; for example, the interaction of mature TCs

with vertical shear (e.g., Simpson and Riehl 1958; Frank and Ritchie 2001; Tang and Emanuel 2010; Riemer et al. 2010, 2013), the development and intensification of TCs in shear (e.g., Rios-Berrios et al. 2016a,b, 2018), and ET. Ideally, the objective identification of a TC's main airstreams may provide a framework for future trajectory analyses similarly to the conveyor-belt paradigm for midlatitude cyclones.

Acknowledgments. The research leading to these results has been done within the subproject "A4—Evolution and predictability of storm structure during extratropical transition of tropical cyclones" of the Transregional Collaborative Research Center SFB/TRR 165 "Waves to Weather" funded by the German Research Foundation (DFG). The code to perform the cluster analysis will be available at <https://github.com/wavestoweather>. Parts of this research were conducted using the supercomputer Mogon and/or advisory services offered by Johannes Gutenberg University Mainz (hpc.uni-mainz.de), which is a member of the Alliance for High Performance Computing (AHRP) in Rhineland Palatinate (www.ahrp.info) and the Gauss Alliance e.V.

REFERENCES

- Bolton, D., 1980: The computation of equivalent potential temperature. *Mon. Wea. Rev.*, **108**, 1046–1053, [https://doi.org/10.1175/1520-0493\(1980\)108<1046:TCOEPT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1980)108<1046:TCOEPT>2.0.CO;2).
- Browning, K., 1971: Radar measurements of air motion near fronts. *Weather*, **26**, 320–340, <https://doi.org/10.1002/j.1477-8696.1971.tb04211.x>.

- Carlson, T. N., 1980: Airflow through midlatitude cyclones and the comma cloud pattern. *Mon. Wea. Rev.*, **108**, 1498–1509, [https://doi.org/10.1175/1520-0493\(1980\)108<1498:ATMCAT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1980)108<1498:ATMCAT>2.0.CO;2).
- Cram, T. A., J. Persing, M. T. Montgomery, and S. A. Braun, 2007: A Lagrangian trajectory view on transport and mixing processes between the eye, eyewall, and environment using a high-resolution simulation of Hurricane Bonnie (1998). *J. Atmos. Sci.*, **64**, 1835–1856, <https://doi.org/10.1175/JAS3921.1>.
- Davis, C. A., S. C. Jones, and M. Riemer, 2008: Hurricane vortex dynamics during Atlantic extratropical transition. *J. Atmos. Sci.*, **65**, 714–736, <https://doi.org/10.1175/2007JAS2488.1>.
- De Silva, V., and J. B. Tenenbaum, 2004: Sparse multidimensional scaling using landmark points. Tech. Rep., Stanford University, 41 pp.
- Doms, G., and M. Schättler, 2002: A description of the non-hydrostatic regional model LM. Part I: Dynamics and numerics. Deutscher Wetterdienst, 140 pp., <http://www.cosmo-model.org>.
- , and Coauthors, 2011: A description of the nonhydrostatic regional COSMO model: Part II: Physical parameterization. Consortium for Small-Scale Modeling Doc., Deutscher Wetterdienst, 161 pp.
- Eiter, T., and H. Mannila, 1994: Computing discrete Fréchet distance. Tech. Rep. CD-TR 94/64, Information Systems Dept., Technical University of Vienna, 7 pp.
- Eliassen, A., and E. Kleinschmidt, 1957: Dynamic meteorology. *Handbuch der Physik*, S. Flüge, Ed., Vol. 48, Springer, 1–154.
- Euler, C., M. Riemer, T. Kremer, and E. Schömer, 2019: Lagrangian description of air masses associated with latent heat release in Tropical Storm Karl (2016) during extratropical transition. *Mon. Wea. Rev.*, **147**, 2657–2676, <https://doi.org/10.1175/MWR-D-18-0422.1>.
- Evans, C., and Coauthors, 2017: The extratropical transition of tropical cyclones. Part I: Cyclone evolution and direct impacts. *Mon. Wea. Rev.*, **145**, 4317–4344, <https://doi.org/10.1175/MWR-D-17-0027.1>.
- Frank, W. M., and E. A. Ritchie, 2001: Effects of vertical wind shear on the intensity and structure of numerically simulated hurricanes. *Mon. Wea. Rev.*, **129**, 2249–2269, [https://doi.org/10.1175/1520-0493\(2001\)129<2249:EOVWSO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2249:EOVWSO>2.0.CO;2).
- Fréchet, M. M., 1906: Sur quelques points du calcul fonctionnel. *Rend. Circ. Matematic. Palermo*, **22**, 1–72, <https://doi.org/10.1007/BF03018603>.
- Hamilton, R., and J. Archbold, 1945: Meteorology of Nigeria and adjacent territory. *Quart. J. Roy. Meteor. Soc.*, **71**, 231–264, <https://doi.org/10.1002/qj.49707130905>.
- Hart, N. C. G., S. L. Gray, and P. A. Clark, 2015: Detection of coherent airstreams using cluster analysis: Application to an extratropical cyclone. *Mon. Wea. Rev.*, **143**, 3518–3531, <https://doi.org/10.1175/MWR-D-14-00382.1>.
- Jones, S. C., and Coauthors, 2003: The extratropical transition of tropical cyclones: Forecast challenges, current understanding, and future directions. *Wea. Forecasting*, **18**, 1052–1092, [https://doi.org/10.1175/1520-0434\(2003\)018<1052:TETOTC>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<1052:TETOTC>2.0.CO;2).
- Joos, H., and R. M. Forbes, 2016: Impact of different IFS microphysics on a warm conveyor belt and the downstream flow evolution. *Quart. J. Roy. Meteor. Soc.*, **142**, 2727–2739, <https://doi.org/10.1002/qj.2863>.
- Kjellsson, J., K. Döös, F. Laliberté, and J. D. Zika, 2014: The atmospheric general circulation in thermodynamical coordinates. *J. Atmos. Sci.*, **71**, 916–928, <https://doi.org/10.1175/JAS-D-13-0173.1>.
- Klein, P. M., P. A. Harr, and R. L. Elsberry, 2000: Extratropical transition of western North Pacific tropical cyclones: An overview and conceptual model of the transformation stage. *Wea. Forecasting*, **15**, 373–395, [https://doi.org/10.1175/1520-0434\(2000\)015<0373:ETOWNP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0373:ETOWNP>2.0.CO;2).
- Kruskal, J. B., 1964: Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, **29**, 115–129, <https://doi.org/10.1007/BF02289694>.
- Kumpf, A., M. Rautenhaus, M. Riemer, and R. Westermann, 2019: Visual analysis of the temporal evolution of ensemble forecast sensitivities. *IEEE Trans. Vis. Comput. Graph.*, **25**, 98–108, <https://doi.org/10.1109/TVCG.2018.2864901>.
- Lentink, H. S., C. M. Grams, M. Riemer, and S. C. Jones, 2018: The effects of orography on the extratropical transition of tropical cyclones: A case study of Typhoon Sinlaku (2008). *Mon. Wea. Rev.*, **146**, 4231–4246, <https://doi.org/10.1175/MWR-D-18-0150.1>.
- Molinari, J., D. Vollaro, and K. L. Corbosiero, 2004: Tropical cyclone formation in a sheared environment: A case study. *J. Atmos. Sci.*, **61**, 2493–2509, <https://doi.org/10.1175/JAS3291.1>.
- Okubo, A., 1970: Horizontal dispersion of floatable particles in the vicinity of velocity singularities such as convergences. *Deep-Sea Res.*, **17**, 445–454, [https://doi.org/10.1016/0011-7471\(70\)90059-8](https://doi.org/10.1016/0011-7471(70)90059-8).
- Papritz, L., and S. Schemm, 2013: Development of an idealised downstream cyclone: Eulerian and Lagrangian perspective on the kinetic energy. *Tellus*, **65A**, 19539, <https://doi.org/10.3402/tellusa.v65i0.19539>.
- Pasch, R. J., and D. A. Zelinsky, 2016: Tropical Storm Karl (14–25 September). Tech. Rep. AL122016, NOAA/National Hurricane Center, 17 pp, https://www.nhc.noaa.gov/data/tcr/AL122016_Karl.pdf.
- Riehl, H., 1954: *Tropical Meteorology*. McGraw-Hill, 392 pp.
- Riemer, M., and M. T. Montgomery, 2011: Simple kinematic models for the environmental interaction of tropical cyclones in vertical wind shear. *Atmos. Chem. Phys.*, **11**, 9395–9414, <https://doi.org/10.5194/acp-11-9395-2011>.
- , and F. Laliberté, 2015: Secondary circulation of tropical cyclones in vertical wind shear: Lagrangian diagnostic and pathways of environmental interaction. *J. Atmos. Sci.*, **72**, 3517–3536, <https://doi.org/10.1175/JAS-D-14-0350.1>.
- , M. T. Montgomery, and M. E. Nicholls, 2010: A new paradigm for intensity modification of tropical cyclones: Thermodynamic impact of vertical wind shear on the inflow layer. *Atmos. Chem. Phys.*, **10**, 3163–3188, <https://doi.org/10.5194/acp-10-3163-2010>.
- , —, and —, 2013: Further examination of the thermodynamic modification of the inflow layer of tropical cyclones by vertical wind shear. *Atmos. Chem. Phys.*, **13**, 327–346, <https://doi.org/10.5194/acp-13-327-2013>.
- Rios-Berrios, R., R. D. Torn, and C. A. Davis, 2016a: An ensemble approach to investigate tropical cyclone intensification in sheared environments. Part I: Katia (2011). *J. Atmos. Sci.*, **73**, 71–93, <https://doi.org/10.1175/JAS-D-15-0052.1>.
- , —, and —, 2016b: An ensemble approach to investigate tropical cyclone intensification in sheared environments. Part II: Ophelia (2011). *J. Atmos. Sci.*, **73**, 1555–1575, <https://doi.org/10.1175/JAS-D-15-0245.1>.
- , C. A. Davis, and R. D. Torn, 2018: A hypothesis for the intensification of tropical cyclones under moderate vertical wind shear. *J. Atmos. Sci.*, **75**, 4149–4173, <https://doi.org/10.1175/JAS-D-18-0070.1>.

- Schäfler, A., and Coauthors, 2018: The North Atlantic waveguide and downstream impact experiment. *Bull. Amer. Meteor. Soc.*, **99**, 1607–1637, <https://doi.org/10.1175/BAMS-D-17-0003.1>.
- Simpson, R., and R. Riehl, 1958: Mid-tropospheric ventilation as a constraint on hurricane development and maintenance. Preprints, *Tech. Conf. on Hurricanes*, Miami Beach, FL, Amer. Meteor. Soc., D4-1–D4-10.
- Steppeler, J., G. Doms, U. Schättler, H. W. Bitzer, A. Gassmann, U. Damrath, and G. Gregoric, 2003: Meso-gamma scale forecasts using the nonhydrostatic model LM. *Meteor. Atmos. Phys.*, **82**, 75–96, <https://doi.org/10.1007/s00703-001-0592-9>.
- Stern, D. P., and F. Zhang, 2013: How does the eye warm? Part II: Sensitivity to vertical wind shear and a trajectory analysis. *J. Atmos. Sci.*, **70**, 1849–1873, <https://doi.org/10.1175/JAS-D-12-0258.1>.
- Tang, B., and K. Emanuel, 2010: Midlevel ventilation's constraint on tropical cyclone intensity. *J. Atmos. Sci.*, **67**, 1817–1830, <https://doi.org/10.1175/2010JAS3318.1>.
- Theisel, H., and C. Rössl, 2012: Streamline embedding for 3D vector field exploration. *IEEE Trans. Vis. Comput. Graph.*, **18**, 407–420, <https://doi.org/10.1109/TVCG.2011.78>.
- Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Wea. Rev.*, **117**, 1779–1800, [https://doi.org/10.1175/1520-0493\(1989\)117<1779:ACMFSF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2).
- Visvaligam, M., and J. D. Whyatt, 1993: Line generalisation by repeated elimination of points. *Cartogr. J.*, **30**, 46–51, <https://doi.org/10.1179/caj.1993.30.1.46>.
- Weiss, J., 1991: The dynamics of enstrophy transfer in two-dimensional hydrodynamics. *Physica D*, **48**, 273–294, [https://doi.org/10.1016/0167-2789\(91\)90088-Q](https://doi.org/10.1016/0167-2789(91)90088-Q).
- Wernli, B. H., and H. C. Davies, 1997: A Lagrangian-based analysis of extratropical cyclones. I: The method and some applications. *Quart. J. Roy. Meteor. Soc.*, **123**, 467–489, <https://doi.org/10.1002/qj.49712353811>.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 704 pp.
- Willoughby, H. E., F. D. Marks, and R. J. Feinberg, 1984: Stationary and moving convective bands in hurricanes. *J. Atmos. Sci.*, **41**, 3189–3211, [https://doi.org/10.1175/1520-0469\(1984\)041<3189:SAMCBI>2.0.CO;2](https://doi.org/10.1175/1520-0469(1984)041<3189:SAMCBI>2.0.CO;2).