Detection of coherent airstreams using cluster analysis: application to an extratropical cyclone

Detection of coherent airstreams using cluster analysis: application to an extratropical cyclone

Neil C. G. Hart * and Suzanne L. Gray and Peter A. Clark

Department of Meteorology, University of Reading

* *Corresponding author address:* Department of Meteorology, University of Reading, Earley Gate, Reading, United Kingdom

E-mail: n.c.hart@reading.ac.uk

ABSTRACT

Flow in geophysical fluids is commonly summarized by coherent streams, for example conveyor belt flows in extratropical cyclones or jet streaks in the upper troposphere. Typically, parcel trajectories are calculated from the flow field and subjective thresholds are used to distinguish coherent streams of interest. This methodology contribution develops a more objective approach to distinguish coherent airstreams within extratropical cyclones. Agglomerative clustering is applied to trajectories along with a method to identify the optimal number of cluster classes. The methodology is applied to trajectories associated with the low-level jets of a well-studied extratropical cyclone. For computational efficiency, a constraint that trajectories must pass through these jet regions is applied prior to clustering; the partitioning into different airstreams is then performed by the agglomerative clustering. It is demonstrated that the methodology can identify the salient flow structures of cyclones: the warm and cold conveyor belts. A test focusing on the airstreams terminating at the tip of the bent-back front further demonstrates the success of the method in that it can distinguish fine-scale flow structure such as descending sting jet airstreams.

Keywords: Airstreams, Extratropical Cyclones, Warm Conveyor Belt, Cold Conveyor Belt, Ward's method, Trajectory Analysis

1. Introduction

The structure of a wide range of geophysical flows has often been analyzed in terms of distinct, coherent airstreams, such as jet streams, jet streaks and conveyor belts. For example, a conveyor belt view of flow within extratropical cyclones is widely accepted (Harrold 1973; Carlson 1980; Browning and Roberts 1994; Wernli and Davies 1997; Schultz 2001). Though clearly defined features in satellite imagery or synoptic-scale analyses, the precise definition of these airstreams often relies on a relatively arbitrary choice of threshold. The goal of this contribution is to demonstrate that cluster analysis of flow trajectories is a successful method to automate the identification of coherent airstreams in a more objective way.

Cluster analysis has gained wide use in geophysical sciences, particularly in applications where identifying archetypes is useful. Cheng and Wallace (1993) identified large-scale atmospheric flow regimes by applying a hierarchical clustering approach to 500 hPa geopotential height fields. Hierarchical clustering has also been used to categorize synoptic-scale rainfall patterns from a high-density rain-gauge network (e.g. Tennant and Hewitson 2002; Crétat et al. 2012). Fuzzy clustering approaches such as the K-means algorithm have become the favored methods for determining weather regimes due to the variety of advanced statistical tests that can be used to test the robustness of the regimes that are determined (e.g. Michelangeli et al. 1995; Fauchereau et al. 2009).

Application of cluster analysis to two-dimensional and later three-dimensional airflow trajectories was first carried out in research focusing on understanding variability in atmospheric composition at observations sites (eg. Moody and Galloway 1988; Harris and Kahl 1990). Both hierar-

chical and fuzzy clustering approaches have been used to characterize the trajectories computed backwards from atmospheric composition observatories (eg. Dorling et al. 1992; Moody et al. 1995; Cape et al. 2000). Stohl (1998) summarised the strengths and shortcomings of these trajectory computations and classification techniques. Despite this wide use of cluster analysis, we are unaware of an application to the flow in extratropical cyclones. There is an important difference between the application we present and the literature noted above: most applications have little a priori knowledge of the classifications cluster analysis may produce whereas, extratropical cyclones have a wealth of literature describing airflow features.

The warm conveyor belt (WCB) is a warm moist (high-valued equivalent potential temperature) rain-producing ascending airstream advancing polewards ahead of the cold front (Harrold 1973). The cold conveyor belt (CCB) is a cool low-level airstream that forms on the cool side of the warm front flowing rearwards in relation to cyclone motion (Carlson 1980; Schultz 2001). In extratropical cyclones where the warm front bends cyclonically around behind the low pressure center of the system, the CCB flow can wrap around to produce very strong earth-relative winds immediately south of the cyclone center. This is common in Shapiro-Keyser type cyclones (Shapiro and Keyser 1990), producing a "poisonous-tail" of damaging winds (Grønås 1995). In such cyclones, finer-scale flow structures have sometimes been found associated with damaging winds ahead of the CCB (Browning 2004). Termed "sting jets", since they occur near the tip of the poisonous tail, these airstreams develop less frequently than the CCB and WCB (Martínez-Alvarado et al. 2012) and can have a more transient nature when they do develop, persisting for periods of only several hours (Clark et al. 2005; Baker 2009; Martínez-Alvarado et al. 2010).

Conveyor belt and sting jet airstreams are often identified subjectively or with simple thresholding techniques applied to Lagrangian trajectories describing these flows (Wernli and Davies 1997; Schultz 2001; Clark et al. 2005). This approach has proved particularly effective for automating

the identification of the WCBs in climatology studies of extratropical cyclones (Stohl 2001; Eckhardt et al. 2004; Madonna et al. 2014; Pfahl et al. 2014). Trajectories that start near the surface, and exceed a total ascent threshold (e.g. $\triangle p > 600$ hPa where p is pressure) can be retained as the coherent ensembles of trajectories describing the WCB, as demonstrated by Wernli and Davies (1997). Madonna et al. (2014) notes that a 600 hPa ascent criterion (within a two-day time period) is "fairly strong" and found some changes in the spatial distribution of WCB starting points and average evolution of parameter values (such as specific humidity and potential vorticity) along the trajectories in sensitivity tests in which the ascent pressure change or time criterion were relaxed.

Applying thresholds to identify airstreams works well where the airstream is largely similar across cyclone populations and has an easily identifiable characteristic such as strong ascent or descent. Nevertheless, calculating statistics across the airstreams has the caveat that the thresholding may admit more or fewer trajectories in the WCB of each cyclone, resulting in statistical artifacts. If a study focused on fine differences between WCB flow in, for example an ensemble simulation of one storm, this caveat could become a serious issue. Furthermore, if the airstream of interest exhibits wide variability in total ascent or descent, such as a sting jet, threshold criteria may well prevent detection of valid coherent ensembles of trajectories. This contribution proposes cluster analysis as an appropriate tool in such a situation. Clustering trajectories which pass near the frontal structures of extratropical cyclones should naturally result in coherent ensembles of trajectories, based on similarities in their dynamical histories. These could then be classified based on the a priori knowledge of conveyor belt flow; e.g. a coherent ensemble of trajectories would be identified as a WCB if it was a near-saturated ascending airstream ahead of the cold front. The WCB could thus be selected without the need to choose a threshold criterion other than a test of ascent.

The clustering method is demonstrated here in application to the well-observed and well-studied extratropical cyclone Friedhelm (2011). Fig. 1 provides schematic representation of where the airstreams formed during the development of cyclone Friedhelm, which developed explosively and produced very strong and damaging surface winds over Scotland on 8 December 2011 (Baker et al. 2013; Martínez-Alvarado et al. 2014, (MA14 hereafter)). During the early stages of development the WCB was the primary coherent ensemble of trajectories associated with this cyclone (Fig. 1a). The CCB became associated with strong Earth-relative winds as the warm front was cyclonically bent-back around the low pressure center (Fig. 1b). By this stage the low-level jet associated with the WCB was starting to weaken. A sting jet descended on the southern flank of the CCB as the cloud head of the cyclone continued to wrap around with the bent-back warm front (Fig. 1c). While the schematic shows the evolution of these flows for cyclone Friedhelm, this evolution generally occurs for developing cyclones with diagnosed sting jets (Clark et al. 2005, e.g. the Great October storm of 1997).

A general description of the method, independent of application to extratropical cyclones, is given in section 2. Section 3 describes the model simulation of Friedhelm, the cyclone used to demonstrate and test the methodology. The results of applying this method to the dominant low-level conveyor belt airstreams and the mesoscale jet structure near the bent-back front are described in sections 3a and 3b respectively. Section 4 provides a summary of these results and concludes this study.

2. Methodology: Agglomerative Clustering

Agglomerative (also termed hierarchical) clustering depends on assessing the similarity between many individual instances which are commonly referred to as observations in clustering algorithms. In this application each trajectory is one observation. These observations are stored in

a matrix $\mathbf{W}$ containing J observations each described by N dimensions. A common measure of similarity is Euclidean distance, d, computed as the l^2-norm. Between two observation vectors in $\mathbf{W}$ this would be

$$d_{jk} = \sqrt{\sum_{n=1}^{N} (\mathbf{W}_j(n) - \mathbf{W}_k(n))^2}, \tag{1}$$

where j and k index the two observations under comparison and n indexes the dimension of each observation. Clustering starts by agglomerating the most similar observations into new cluster classes. The algorithm used here is Ward's variance minimization (Ward 1963); this has a straightforward implementation , which lends itself to automation, a key goal of this study. Its primary drawback is the tendency to produce cluster classes containing similar numbers of observations. This produces the caveat that coherent streams containing many trajectories may be described by more than one cluster class. The SciPy Hierarchical Clustering module for Python (Jones et al. 2001), used here, implements Ward's method with the widely-used approach of updating a matrix storing the Euclidean distances between each cluster centroid (Wishart 1969). The algorithm ensures that, at each iteration, new classes are created such that variance between members within a cluster class is minimized across all possible combinations of members of a class, at that step. Iteration continues until all observations are agglomerated into a single class. Cheng and Wallace (1993) provide a detailed account of the algorithm with additional descriptions available in Crétat et al. (2012) and Ramos (2001). The succession of agglomerations are represented graphically by dendrograms, as shown in Fig. 2 (a and c).

Any distinguishing variable can be used in an observation vector. In the context of classifying atmospheric flows, building the observation vectors from positional information $(x(t), y(t), z(t)$, where t is time) is appropriate (e.g. Dorling et al. 1992). Inclusion of an airmass tracer variable $(\gamma(t))$ can add further distinguishing information to the flows under consideration. Having computed trajectories for a given flow field (from start points at time t_0) a time period of interest can

be chosen, e.g. $t = [t_0 - 3hrs, t_0 + 3hrs]$, and the observation vector for one of these trajectories can be specified:

$$\mathbf{W}_j = [\mathbf{x}_j, \mathbf{y}_j, \mathbf{z}_j, \gamma_j], \tag{2}$$

where bold variables denote rows of values for this time period. This observation vector would have dimension $N = 28$, if the time period was over seven hourly positions.

Scaling of these variables is necessary due to the different units of the horizontal, vertical and airmass variables. We choose to scale the values of each variable by their respective standard deviations at time $t = t_0$, giving $\hat{\mathbf{x}}_j = \mathbf{x}_j / \sigma[\mathbf{x}_{all}(t_0)]$ and similarly for $\mathbf{y}, \mathbf{z}, \gamma$. The matrix $\mathbf{x}_{all}$ represents a matrix of $\mathbf{x}$ positional vectors for all trajectories and σ represents the standard deviation. Thus, $\hat{\mathbf{x}}_j$ is the j-th $\mathbf{x}$-coordinate vector scaled with the standard deviation of the x position of all trajectories at $t = t_0$. Scaling in such a manner for all variables in each observation produces the data matrix that is passed to the clustering algorithm:

$$\hat{\mathbf{W}}_j = [\hat{\mathbf{x}}_j, \hat{\mathbf{y}}_j, \hat{\mathbf{y}}_j, \hat{\gamma}_j]; j = 1 : J. \tag{3}$$

Two subtleties to these choices bear mention. First, the mean is not removed (if so this would be a normalization) as this would remove the ability to distinguish the geographic locality of the trajectories, leaving only the shape of each trajectory as the distinction. Second, scaling by the standard deviation calculated only at the trajectory start time ensures that the relative time evolution of each variable is unmodified, i.e. a trajectory with substantial three-dimensional curvature retains its character in relation to more linear trajectory paths.

The nature of the task requires an approach to automatically decide on the number of clusters to retain. This decision would ideally result in a classification with coherent, dissimilar airstreams assigned to separate classes, and similar airstreams grouped in one class. This can be achieved by exploiting a feature of the clustering algorithm itself. As agglomerative clustering proceeds,

d between successive clusters joined at each iteration increases gradually. This distance d can be averaged over all the clusters joined at one iteration step to give $\overline{d}$. Fig. 2b shows this increase in $\overline{d}$ as a function of agglomeration step. When notably distinct clusters start to be co-joined $\overline{d}$ rapidly increases. The last set of cluster classes present before this sudden jump represents the classification of trajectories into classes most distinct from one another, as shown by curvature in Fig. 2b. These classes are the classification retained as the salient airstreams in this methodology. Dorling et al. (1992) used a similar technique in application of a fuzzy clustering method for 2-30 clusters. The number of cluster classes was chosen as that just before a sudden increase in intra-cluster variance, as expected when classes contain very different class members. This was decided by visual inspection (Dorling et al. 1992). In our method, the decision of cluster numbers is automated which leads to the method admitting a caveat as follows. If a maximum in curvature occurs two agglomerations or more before only one class is left, the method works as described; if not, the last value of the curvature is the highest and the number of classes chosen is forced to be the number of classes present in the third-last iteration. This is borne out in comparison of Fig. 2 (b) and (d). At worst, this caveat results in more cluster classes being selected than would have been selected by visual selection. Therefore, no attempt is made to draw conclusions from the number of classes.

Each of these automatically chosen classes contains a population of trajectories from which class-median trajectories are calculated by taking the median trajectory properties at each time in the coherent ensemble of trajectories. Trajectory classification can be repeated for trajectory populations calculated from start locations defined at consecutive times in the evolution of a weather system. Class-median trajectories from each consecutive time can be calculated relative to the position of the center of the weather system at the given time. This results in a population of system-relative class-median trajectories that summarize the weather system development. This

summary class-median-trajectory population can then be classified with same clustering approach described above to obtain a "super" classification of airstreams that form as a weather system evolves. In a Shapiro-Keyser type extratropical cyclone, this super-classification should show the WCB flow during early cyclone development and the CCB in later stages (Shapiro and Keyser 1990). This is demonstrated in section 3a.

3. Coherent ensembles of trajectories in a test case cyclone

In this section we describe cyclone Friedhelm and demonstrate that the method of clustering trajectories can identify both synoptic-scale and mesoscale structures in its flow field. Cyclone Friedhelm was observed in-situ using a research aircraft during intensive observing period 8 of the Diabatic Influences on Mesoscale Structures in Extratropical Storms (DIAMET) field campaign (Vaughan et al. 2014). It was the subject of a detailed case study into the airstreams that that constituted the low-level jet in the bent-back frontal region (MA14). MA14 analyzed a numerical simulation which was compared to flight data. They defined the low-level jet in this region by the 45 $\mathrm{m\,s^{-1}}$ isotach and found three constituent airstreams with distinct equivalent potential temperatures and airflow histories: a primary CCB flow, a secondary CCB flow, and a sting jet descent. Comparison with MA14 provides a stringent test for this clustering methodology.

The cyclone was simulated using the operational numerical weather prediction model used by the Met Office, the Unified Model (MetUM). Version 8.2 of the MetUM was used with the (until recently operational) North Atlantic and European domain configuration which extends from approximately 30° to 70°N in latitude and from 60°W to 40°E in longitude (figures in this paper show a subregion of this domain). This configuration has 0.11° ($\sim$ 12 km) grid spacing in the horizontal in both latitude and longitude on a rotated grid. The model lid is at $\sim$80km with the 70 stretched vertical levels spaced such that slantwise circulations of slope 1:40, with absolute

vertical level spacing of 300m, are resolvable near 600 hPa and slopes shallower than 1:50 are resolvable below 750hPa. This compares well with the 1 in 50 slope suggested as necessary to simulate the release of conditional symmetric instability by slantwise circulations (Persson and Warner 1993) and similarly slantwise sting-jet descents (Clark et al. 2005; Gray et al. 2011).

The simulation was initialized at 0000 UTC 8 December from the operational North Atlantic and European analysis, with boundary conditions provided from the operational global MetUM. This modeling setup is nearly identical to that used in MA14; the only differences are the updates due to the change in model version from 7.3 to 8.2. Simulations of Friedhelm from the two model versions compare very closely (not shown). Model data from the version 8.2 simulation were interpolated onto pressure levels ($\triangle p = 25$ hPa) before calculating the diagnostics used in this study.

The aim is to demonstrate that the clustering methodology can characterize airstreams that flow through the low-level jet regions of this cyclone. In principle, the methodology could identify these low-level jets from trajectories that describe the full flow around the cyclone. Jet regions of strong wind speed would be distinguished by coherent classes of trajectories that trace greater distances than those not associated with jet regions. To reduce the computational resources needed, including the computational challenge of clustering $\sim 10^5$ trajectories, we speed up this identification process by pre-selecting gridpoints within strong wind regions. However, this pre-selection is a practical step, not a necessary one. To this end we choose where to seed trajectories by identifying grid points that lie within a threshold isotach in the lower troposphere (950-650 hPa). Considering 25 hPa pressure increments, trajectories identifying the conveyor belts at each level are started within any 40 m s^{-1} isotach that is vertically contiguous with a 40 m s^{-1} isotach at 850 hPa, an arbitrary but reasonable wind speed for low-level jets (results shown in section 3a). For comparison with the bent-back front jets studied in MA14, their threshold of 45 m s^{-1} is chosen (results shown

in section 3b). Start points were thus selected hourly from the model output and used to initialize both forward and backward trajectories. Trajectories were calculated with the Lagrangian Analysis Tool LAGRANTO (Wernli and Davies 1997; Sprenger and Wernli 2015) using the iterative Euler scheme applied to hourly model output, with an iteration timestep of 5 minutes.

The observation vector ($\mathbf{W}_j$) for each trajectory is described by latitude, longitude, and pressure coordinates with equivalent potential temperature (θ_e) providing the airmass characteristic. For moist flows θ_e is a conserved variable; however, θ_e can evolve in time along the trajectories and trajectories with similar θ_e will be preferentially clustered. While other airmass tracers could be chosen, a priori knowledge of flow around extratropical cyclones suggests inclusion of a measure of the moist entropy of air parcels is worthwhile: Browning and Roberts (1994) describe how the warm and cold conveyor belts can be distinguished by their high and low θ_e values respectively (after Carlson 1980); and Clark et al. (2005) show that θ_e is also approximately conserved during the descent of air in sting jets.

a. Identification of conveyor belts

To identify the conveyor belts, each trajectory was calculated over the time period $[t_0 - 3hrs, t_0 + 3hrs]$ where t_0 is the initialization time from when both forward and backward trajectories were calculated. This six hour period was chosen as a minimum time span in which to capture the key features of conveyor belt flows, i.e. location relative to the storm center, curvature, and ascending or descending character. Tests with a longer time period $[t_0 - 6hrs, t_0 + 6hrs]$ produced very similar results (not shown). As Wernli and Davies (1997) noted, these airstreams are coherent for the duration of storm intensification, longer than 12 hours. However, these key features are also present on shorter timescales (e.g. Fig. 10 in Schultz 2001).

The resulting classification for cyclone Friedhelm is shown for trajectories passing through low-level jet regions at 0600 UTC 8 December 2011 (Fig. 3). The full population of trajectories is shown in Fig. 3a. Classes with both CCB (class #1) and WCB (class #5) characteristics are identified (Fig. 3 b and c, respectively). At this time the cyclone structure was identified as corresponding to stage three of the Shapiro-Keyser conceptual model (Shapiro and Keyser 1990) by MA14: the fronts had formed a T-bone structure with the WCB still present and the CCB starting to wrap around the cyclone center. Fig. 3d presents a WCB trajectory population obtained by thresholding for saturated ascent. Due to the short 21 hr duration of these trajectories, a 600 hPa ascent criterion (as applied in previous studies such as Wernli and Davies (1997) for a 48 hr period) only admits 183 trajectories. So to provide a comparable population size for comparison to the WCB cluster class, a 400 hPa ascent criterion is chosen. The WCBs obtained by the agglomerative clustering and thresholding methods match closely (compare Fig. 3 c and d). Obtaining the WCB by thresholding was almost instantaneous, whereas cluster analysis of the full 9738 trajectories passing through the low-level jet region took $\sim$ 10min. However, the clustering of the 887 trajectories passing through the strong wind region at the tip of the bent-back front to produce Fig. 7 completes in order seconds; this issue is discussed in section 4. Median trajectories for the CCB and WCB class populations are overlaid in Fig. 3 (b and c). These median trajectories, along with the others resulting from the clustering for this start time, are mapped in Fig. 4a with Fig. 4b, c and d displaying their evolution in pressure, relative humidity, and θ_e respectively. These are shown for the full time period of the simulation but, following the previous discussion, classification was only performed on trajectory histories from 0300 UTC to 0900 UTC. The dendrogram associated with these clusters classes is found in Fig. 2a and shows the distance cutoff used to obtain these six classes.

The sensitivity of the final classification to variables chosen for clustering was tested (not shown) by considering both the number of classes produced and comparison of resulting class-median trajectories. This showed that pressure evolution was a strongly distinguishing variable, with latitude and longitude less so. Inclusion of θ_e in the observation vector tended to increase the number of final clusters. This implies it assists in distinguishing different air masses constituting the airstreams.

The wind maximum directly south of and closest to the cyclone center is typical of the wrap around of the CCB and the associated development of strong surface winds (Fig. 4a). The near-surface airstream describing the CCB flow is captured in class #1. This median trajectory remains saturated and below 800 hPa while a drier airstream, class #2, follows a more zonal path above 700 hPa (Fig. 4b and c). Together with class #4, these three classes represent flow in the cool sector of the cyclone, northwest of the cold front, as demonstrated in Fig. 4d. The wind maximum some distance southeast of the cyclone center in Fig. 4a is associated with the cold front and attendant WCB. The classification in Fig. 4 captures the WCB with class #5, as identified by a median trajectory that starts near the surface and rises 100 hPa between 0300 UTC and 0900 UTC before ascending more rapidly to about 350 hPa (Fig. 4b). Class #6 describes a similar airstream at an elevated altitude. These median trajectories remain saturated, along with class #3 which captures flow through a lower extension of the upper-level jet. Classes #3, #5, and #6 summarize flow in the warm sector of this cyclone (Fig. 4d).

How coherent are the airstreams summarized by these class medians? Following from Wernli and Davies (1997) who used variance, we quantify coherency by the standard deviations of the trajectory observation variables at each time and then average this value for all times (Table 1). Considering θ_e, trajectory class #5 (the WCB airstream) is the most coherent relative to other classes. Likewise class #1, the CCB has small variance in θ_e. However, small standard deviations

in pressure ($\sim$50 hPa, about 4 times the model vertical level spacing) and θ_e ($\sim$1 K) in all the classes support the assertion that this clustering methodology is capturing coherent ensembles of trajectories. These results provide a characterization of airstreams constituting the low-level jets at 0600 UTC in the evolution of Friedhelm. We now consider the later hours of development.

Agglomerative clustering of trajectories was applied separately to the model output for initial times set to every hour of the simulation from 0500 UTC–1700 UTC. This resulted in a population of class-median trajectories, i.e. the information shown in Fig. 4 for each time. The entire set (for all initial times) of class-median trajectories can be classified as follows. Fig. 5a shows an illustrative selection those class-median trajectories with WCB characteristics. Before classification the system-relative (relative to the storm center) class-median trajectories were calculated (Fig. 5b). Classification was then performed on all system-relative class-median trajectories, which in this case clustered these six trajectories and another seven (not shown for clarity) into a class capturing WCB flow during the period 0500 UTC to 1200 UTC of this cyclone. This cluster class can be summarized by the mean of all system-relative class-median trajectories. This "super-class" mean is shown in Fig. 5b as the dashed black line. Note that for the super-classification the class median of the few trajectories is noisy, hence the use of the class mean.

All the super-class means for this storm are presented in Fig. 6. Analysis of the times of the trajectories in the classes (not shown) associated with each super-class mean in Fig. 6 enables characterization of the evolution of the conveyor belt airstreams in cyclone Friedhelm. Super-class mean #1 contains a population of class-median trajectories present from 0500 UTC until 1200 UTC. This was the WCB while it was still a part of the low-level jet regions of this cyclone. The elevation of this super-class mean (Fig. 6b) indicates that while containing obvious WCB class-medians such as class #5 in Fig. 4, it also contains class-median trajectories of elevated WCB-like flows such as class #6 in Fig. 4. Super-class mean #4 summarizes the population of

CCB class-medians present throughout the period analyzed (0500 UTC until 1800 UTC). The weak, late ascent of this class and mean warming (Fig. 6b) of 5K (Fig. 6d) contrast with the cooling ($\sim$2K) and sharper ascent of class-mean #1. Classes #2 and #3 capture the low-level extensions of upper-level jets above the cold front and bent-back warm front.

b. Identification of airstreams terminating at the tip of the bent-back front

For this comparison with airstreams classified in MA14, the focus is on the airstreams that enter an intense low-level jet region near the tip of the bent-back front and positioned south of the cyclone center (in the frontal fracture zone as defined in MA14). Start points for trajectories in this low-level jet were only retained if they were within the isotach that was contiguous with the near-surface jet, defined by wind speed exceeding 45 $\mathrm{m\,s^{-1}}$ at 850 hPa. The trajectories were classified based on their positional and thermodynamic histories in the time period $[t_0 - 5hrs, t_0]$.

MA14 describe descending sting jets arriving in this low-level jet at 0900 UTC, 1300 UTC and 1600 UTC on 8 December and undercut by two distinct CCB flows. Further analysis of these sting jets revealed mesoscale dynamical instabilities (conditional, conditional symmetric, and inertial instabilities) associated with the sting-jet descent which were absent in the CCB flows. Using convection-permitting (2.2 km horizontal grid spacing) ensemble simulations of cyclone Friedhelm, Vaughan et al. (2014) reveal fine-scale banding in the wind and precipitation structure in the region where the sting jet airstream emerges from the cloud head.

Having demonstrated that the clustering algorithm can distinguish conveyor belts in section 3a, the following question is now addressed: can the clustering methodology capture a mesoscale feature such as the sting jet with minimal use of threshold criteria? Figure 7 presents the results of the cluster analysis for trajectories arriving in the low-level jet in the frontal fracture zone of cyclone Friedhelm at 1600 UTC. For comparison, Fig. 8 (adapted from MA14) shows the

pressure, relative humidity, and θ_e evolutions of the sting jet (S2) and CCB (S1 and S3) airstreams identified in MA14. In MA14, the trajectories were split into coherent ensembles of trajectories using subjectively chosen threshold values for θ_e (Fig. 8a shows the pressure evolutions for four arrival times; Fig. 8b shows the relative humidity evolution for arrival at 1600 UTC only). Cluster classes #2-5 distinguish nuances of the CCB (Fig. 7a). Together these classes describe similar evolutions to those of S1 and S3 in MA14. They remain saturated at low altitude while rising weakly and warming $\sim$3K, characteristic of CCB flow.

In contrast to the other classes, the median trajectory in class #1 descended more than 150 hPa during this period (Fig. 7b). This class-median is most similar to the median trajectory labeled S2@16 by MA14 due to its descent from $\sim$500 hPa to $\sim$700 hPa, drying to about 50% relative humidity and small change in θ_e ($\sim$1K). These features are characteristic of sting jet descents and MA14 demonstrated the presence of mesoscale instability associated with the descent of the trajectory population that this class summarizes. Although sting jets are often associated with strong surface winds (e.g. in the Great October storm of 1987, Browning 2004), trajectory analysis which uses the resolved model wind-field, cannot show the interaction of the sting jet with the boundary layer.

Some differences in evolution of the ensemble median CCB and sting jet trajectories calculated here and in MA14 are to be expected as the different clustering methods result in slightly different apportioning of individual trajectories to each ensemble. For example, the inclusion into S2 of some trajectories assigned to S3 would result in a moister S2 class median. Differences in the ensemble-median trajectories would also be expected from slight differences in forecast evolution due to the different model version used, the rejection of some trajectories in MA14 using a conservation of potential temperature criterion (but not in this study), and differences in the start points of the trajectories (the windspeed threshold used to identify the start points is the same in this study

and MA14, but MA14 used all points exceeding this threshold within a specified box whereas in this study points contiguous to the jet at 850 hPa were used). However, the differences are minor; the overall resemblance between the full trajectory populations of class #1 and S2 (Fig. 9 a and b respectively) is striking. Both contain trajectories that rose from near the surface up to $\sim$600 hPa before joining air parcels with a more westerly and elevated source. These parcels then descended together as an airstream into the strong wind region above Scotland. This similarity demonstrates that the agglomerative clustering methodology can distinguish mesoscale flow structures such as sting-jet descents even in cases such as Cyclone Friedhelm in which the CCB undercuts the sting jet.

4. Summary and conclusions

This study has demonstrated the ability of cluster analysis to identify the salient airstreams of an extratropical cyclone in an automated way. This was possible by making specific choices with regards to the time period, distinguishing variables, and cluster algorithm. These choices benefit from a priori knowledge of extratropical cyclone structure. Focus on strong wind regions introduced a wind magnitude threshold which was applied to limit the number of start points of trajectories to be passed through cluster analysis.

The first test was to identify the primary low-level flows of extratropical cyclone Friedhelm. With focus on the low-level jets, the CCB and WCB were identified at a specific time. The cluster-determined WCB matched closely to a WCB obtained by thresholding for saturated ascent. The automated method was applied to the classification of trajectories passing through the low-level wind maxima during a period in which the cyclone continued to develop. Cluster analysis was performed on the population of class-medians from classifications at each hour in this period.

This produced a "super-classification" which summarised the Lagrangian flow of this storm in a single system-relative figure.

Following the identification of conveyor belts the second test was to identify flows arriving in the strong (>45 m s^{-1}) low-level jet region in the frontal fracture zone, just south of the center of cyclone Friedhelm. We demonstrated that the clustering methodology was capable of distinguishing mesoscale flow structure comparable to that identified by careful case study analysis of this cyclone and clustering of trajectories using subjectively chosen θ_e thresholds (MA14). The success in passing this second test motivates the use of this clustering methodology in the study of mesoscale features of the flow such as sting jets in extratropical cyclones.

The cluster analysis method successfully passed both these tests and thus provides a more objective way of identifying airstreams in extratropical cyclones than the use of threshold criteria. We noted that identifying the WCB by an ascent threshold is substantially more computationally efficient, but the cluster analysis will be computationally acceptable for many applications. The main computational drawback of Ward's method is due to the implementation using a stored Euclidean matrix. This contains the distance between each possible pair of trajectories. Therefore, the matrix size increases as the square of number of trajectories to be classified, which can lead to large memory requirements.

The caveat of the presented clustering algorithm is that both the choice of number of cluster classes and Ward's clustering algorithm can result in more clusters than the data natural describes with large clusters classified into a few clusters of similar size. This may have occurred with the multiple CCB clusters in Fig. 7. Further work could experiment with alternative clustering methodologies such as, K-means (MacQueen 1967; Hartigan and Wong 1979) or Affinity Propagation (Frey and Dueck 2007), in order to address these issues. Some of these other clustering algorithms also have less computational memory demand, addressing the problem of clustering

large numbers ($> 10^4$) of trajectories (see Dorling et al. 1992). They can however, demand more computational processing time.

In conclusion, an extratropical cyclone has been used to demonstrate this methodological approach to airstream classification. This method has application in studies where the caveats of thresholding are unacceptable. For example, sting jets may have strong variability in descent rates, so analysis of ensemble simulations of sting jet storms would be a natural application of the approach presented here. Furthermore, the method is sufficiently general to be used in other contexts where such distinguishing of flows is needed.

Acknowledgments. This work made use of the facilities of HECToR, the UK's national high-performance computing service. The Met Office is thanked for making the MetUM and associated operational start dump and boundary condition files available as is the NERC-funded National Centre for Atmospheric Sciences (NCAS) Computational Modelling Services (CMS) for providing computing and technical support. Two anonymous reviewers and D. Schultz and H. Sodemann are thanked for their thorough reviews and suggestions. We especially appreciate the care taken by one anonymous reviewer which identified a bug in the code. This work was supported by AXA-Research Fund under the project 'Sting Jet Windstorms in Current and Future Climates'.

References

Baker, L. H., 2009: Sting jets in severe Northern European wind storms. *Weather*, **64**, 143–148, doi:10.1002/wea.397.

Baker, L. H., O. Martínez-Alvarado, J. Methven, and P. Knippertz, 2013: Flying through extratropical cyclone Friedhelm. *Weather*, **68**, 9–13, doi:10.1002/wea.2047.

Browning, K. A., 2004: The sting at the end of the tail: Damaging winds associated with extratropical cyclones. *Quart. J. Roy. Meteor. Soc.*, **130**, 375–399, doi:10.1256/qj.02.143.

Browning, K. A., and N. M. Roberts, 1994: Structure of a frontal cyclone. *Quart. J. Roy. Meteor. Soc.*, **120**, 1535–1557.

Cape, J. N., J. Methven, and L. E. Hudson, 2000: The use of trajectory cluster analysis to interpret trace gas measurements at Mace Head, Ireland. *Atmos. Environ.*, **34**, 3651–3663, doi:10.1016/S1352-2310(00)00098-4.

Carlson, T. N., 1980: Airflow through midlatitude cyclones and the comma cloud pattern. *Mon. Wea. Rev.*, **108**, 1498–1509.

Cheng, X., and J. M. Wallace, 1993: Cluster analysis of Northern Hemisphere wintertime 500-hPa height field: spatial patterns. *J. Atmos. Sci.*, **50 (6)**, 2674–2692.

Clark, P. A., K. A. Browning, and C. Wang, 2005: The sting at the end of the tail: Model diagnostics of the fine-scale three-dimensional structure of the cloud band. *Quart. J. Roy. Meteor. Soc.*, **131**, 2263–2292, doi:10.1256/qj.04.36.

Crétat, J., Y. Richard, B. Pohl, M. Rouault, C. J. C. Reason, and N. Fauchereau, 2012: Recurrent daily rainfall patterns over South Africa and associated dynamics during the core of the austral summer. *Int. J. Climatol.*, **32**, 261–273, doi:10.1002/joc.2266.

Dorling, S., T. Davies, and C. Pierce, 1992: Cluster analysis: A technique for estimating the synoptic meteorological controls on air and precipitation chemistry - method and applications. *Atmos. Environ. A*, **26 (14)**, 2575–2581, doi:10.1016/0960-1686(92)90110-7.

Eckhardt, S., A. Stohl, H. Wernli, P. James, C. Forster, and N. Spichtinger, 2004: A 15-year climatology of warm conveyor belts. *J. Clim.*, **17**, 218–237.

Fauchereau, N., B. Pohl, C. Reason, M. Rouault, and Y. Richard, 2009: Recurrent daily OLR patterns in the Southern Africa / Southwest Indian Ocean region, implications for South African rainfall and teleconnections. *Clim. Dyn.*, **32**, 575–591.

Frey, B. J., and D. Dueck, 2007: Clustering by passing messages between data points. *Science*, **315**, 972–976, doi:10.1126/science.1136800.

Gray, S. L., O. Martínez-Alvarado, L. H. Baker, and P. A. Clark, 2011: Conditional symmetric instability in sting jet storms. *Quart. J. Roy. Meteor. Soc.*, **137**, 1482–1500, doi:10.1002/qj.859.

Grønås, S., 1995: The seclusion intensification of the New Year's day storm 1992. *Tellus A*, **47**, 733–746, doi:10.1034/j.1600-0870.1995.00116.x.

Harris, J. M., and J. D. Kahl, 1990: A descriptive atmospheric transport climatology for the Mauna Loa Observatory, using clustered trajectories. *J. Geophys. Res.: Atmos.*, **95 (D9)**, 13 651–13 667, doi:10.1029/JD095iD09p13651.

Harrold, T. W., 1973: Mechanisms influencing the distribution of precipitation within baroclinic disturbances. *Quart. J. Roy. Meteor. Soc.*, **99**, 232–251.

Hartigan, J. A., and M. A. Wong, 1979: Algorithm as 136: A k-means clustering algorithm. *J. Roy. Stat. Soc. C*, **28**, 100–108.

Jones, E., T. E. Oliphant, P. Peterson, and Coauthors, 2001: SciPy: Open source scientific tools for Python. URL http://www.scipy.org/, [Online; accessed 2015-02-21].

MacQueen, J., 1967: Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, University of California Press, Berkeley, Calif., 281–297, URL http://projecteuclid.org/euclid.bsmsp/1200512992.

Madonna, E., H. Wernli, H. Joos, and O. Martius, 2014: Warm conveyor belts in the ERA-Interim dataset (1979-2010) Part I: Climatology and potential vorticity evolution. *J. Climate*, **27**, 3–26, doi:10.1175/JCLI-D-12-00720.1.

Martínez-Alvarado, O., L. H. Baker, S. L. Gray, J. Methven, and R. S. Plant, 2014: Distinguishing the cold conveyor belt and sting jet air streams in an intense extratropical cyclone. *Mon. Wea. Rev.*, **142**, doi:10.1002/met.297.

Martínez-Alvarado, O., S. L. Gray, J. L. Catto, and P. A. Clark, 2012: Sting jets in intense North-Atlantic windstorms. *Environ. Res. Lett.*, **7**, 024 014, doi:10.1088/1748-9326/7/2/024014.

Martínez-Alvarado, O., F. Weidle, and S. L. Gray, 2010: Sting jets in simulations of a real cyclone by two mesoscale models. *Mon. Wea. Rev.*, **138**, 4054–407, doi:10.1175/2010MWR3290.1.

Michelangeli, P., R. Vautard, and B. Legras, 1995: Weather regimes: Recurrence and quasi stationarity. *J. Atmos. Sci.*, **52**, 1237–1256, doi:10.1175/1520-0469(1995)052⟨1237:WRRAQS⟩2.0.CO;2.

Moody, J. L., and J. N. Galloway, 1988: Quantifying the relationship between atmospheric transport and the chemical composition of precipitation on Bermuda. *Tellus B*, **40B (5)**, 463–479, doi:10.1111/j.1600-0889.1988.tb00117.x.

Moody, J. L., S. J. Oltmans, H. Levy, and J. T. Merrill, 1995: Transport climatology of tropospheric ozone: Bermuda, 1988-1991. *J. Geophys. Res. Atmos.*, **100**, 7179–7194, doi:10.1029/94JD02830.

Persson, P. O. G., and T. T. Warner, 1993: Nonlinear hydrostatic conditional symmetric instability: Implications for numerical weather prediction. *Mon. Wea. Rev.*, **121**, 1821–1833, doi:10.1175/1520-0493(1993)121⟨1821:NHCSII⟩2.0.CO;2.

Pfahl, S., E. Madonna, M. Boettcher, H. Joos, and H. Wernli, 2014: Warm conveyor belts in the ERA-Interim dataset (1979-2010) Part II: Moisture origin and relevance for precipitation. *J. Climate*, **27**, 27–40, doi:10.1175/JCLI-D-13-00223.1.

Ramos, M. C., 2001: Divisive and hierarchical clustering techniques to analyse variability of rainfall distribution patterns in a Mediterranean Region. *Atmos. Res.*, **57**, 123–138.

Schultz, D. M., 2001: Reexamining the cold conveyor belt. *Mon. Wea. Rev.*, **129**, 2205–2225, doi:10.1175/1520-0493(2001)129⟨2205:RTCCB⟩2.0.CO;2.

Shapiro, M. A., and D. A. Keyser, 1990: Fronts, jet streams, and the tropopause. *Extratropical Cyclones: The Erik Palmén Memorial Volume*, C. W. Newton, and E. O. Holopainen, Eds., American Meteorological Society, 167–191.

Sprenger, M., and H. Wernli, 2015: The Lagrangian analysis tool LAGRANTO - version 2.0. *Geosci. Mod. Dev. Discuss.*, **8**, 1893–1943, doi:10.5194/gmdd-8-1893-2015.

Stohl, A., 1998: Computation, accuracy and applications of trajectories - A review and bibliography. *Atmos. Environ.*, **32 (6)**, 947–966, doi:10.1016/S1352-2310(97)00457-3.

Stohl, A., 2001: A 1-year Lagrangian "climatology" of airstreams in the northern hemisphere troposphere and lowermost stratosphere. *J. Geophys. Res.*, **106**, 30 445–30 462.

Tennant, W. J., and B. C. Hewitson, 2002: Intra-seasonal rainfall characteristics and their importance to the seasonal prediction problem. *Int. J. Climatol.*, **22**, 1033–1048.

Vaughan, G., and Coauthors, 2014: Cloud banding and winds in intense European cyclones: Results from the DIAMET project. *Bull. Amer. Meteor. Soc.*, doi:10.1175/BAMS-D-13-00238.1, in Press.

506 Ward, J. H., 1963: Hierarchical grouping to optimize an objective function. *J. Amer. Stat. Assoc.*,
507 **58**, 236–244, doi:10.1080/01621459.1963.10500845.

508 Wernli, H., and H. C. Davies, 1997: A Lagrangian-based analysis of extratropical cyclones. I:
509 The method and some applications. *Quart. J. Roy. Meteor. Soc.*, **123**, 467–489, doi:10.1002/qj.
510 49712353811.

511 Wishart, D., 1969: An algorithm for hierarchical classification. *Biometrics*, **25**, 165–170.

LIST OF TABLES

TABLE 1. Average standard deviation (σ) of each variable for the ensemble of trajectories in each class obtained from clustering trajectories passing through low-level wind maxima at 0600 UTC (shown in Fig. 4) and 1600 UTC (shown in Fig. 7). Average σ for each variable is computed as the mean of σ's at each hour for the classification period chosen: $[t_0 - 3hrs, t_0 + 3hrs]$ at 0600 UTC and $[t_0 - 5hrs, t_0 hrs]$ at 1600 UTC.

	0600 UTC Classes						**1600 UTC Classes**				
	#1	**#2**	**#3**	**#4**	**#5**	**#6**	**#1**	**#2**	**#3**	**#4**	**#5**
σ(**Lon**.)(deg)	0.53	0.62	0.98	1.43	0.52	0.93	0.35	0.22	0.24	0.31	0.27
σ(**Lat**.)(deg)	0.70	0.80	0.34	0.63	0.71	0.80	0.27	0.16	0.25	0.22	0.21
σ(**Pres**.)(hPa)	52.99	45.21	39.78	32.20	43.14	46.40	33.66	39.88	49.09	35.17	47.11
$\sigma(\theta_e)$(K)	0.50	1.12	0.90	1.13	0.40	1.18	0.31	0.24	0.30	0.21	0.39
no. trajs	1192	1847	1294	1742	1324	2339	249	89	188	133	228

LIST OF FIGURES

564 **Fig. 8.** Evolution of (a) pressure and (b) relative humidity and (c) θ_e with respect to ice along the
565 ensemble median trajectories of the subjectively clustered trajectories arriving in the low-
566 level jet region in the frontal fracture zone at 1600 UTC 8 December 2011 in MA14. The
567 ensemble median trajectories represent airstreams which are labeled as "SX@HH", where
568 X indicates the airstream number and HH indicates the arrival hour. Pressure evolutions
569 are shown for airstreams arriving at four times whereas relative humidity evolutions are
570 only shown for those airstreams arriving at 1600 UTC. Figure adapted from MA14, Fig. 8
571 (©American Meteorological Society. Used with permission.). 37

572 **Fig. 9.** Positional evolution of the trajectories constituting (a) agglomeratively clustered class #1
573 and (b) the subjectively clustered S2 airstream (from MA14, Fig. 5 ©American Meteoro-
574 logical Society. Used with permission.). The back trajectories are colored by pressure. (a)
575 also shows the position of the minimum pressure marked by **X** at 1600 UTC on the smoothed
576 cyclone track (black line) and (b) also shows mean sea level pressure at 1600 UTC (con-
577 tours). Black dots in both panels represent the positions of the trajectories at 1500 UTC.
578 The trajectories in both panels extend backwards to 0100 UTC. 38

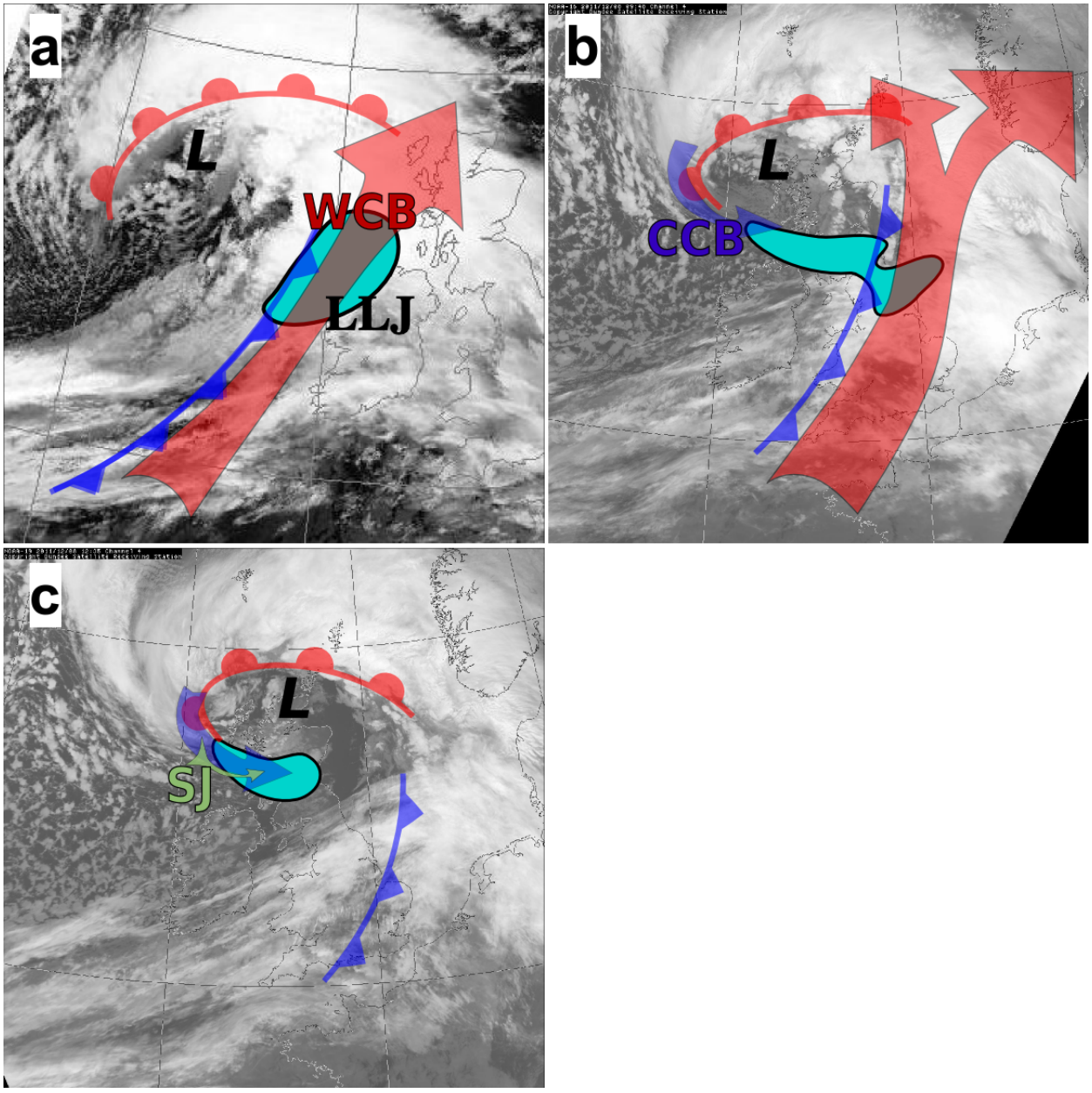


FIG. 1. Schematic of the development of airstreams and front locations in extratropical cyclone Friedhelm during 8 December 2011 overlaid onto infrared imagery at (a) 0300 UTC (MODIS); (b) 1000 UTC (AVHRR); and (c) 1300 UTC (AVHRR). Light blue region denotes approximate region of strongest low-level (eg. 850hPa) winds. (Courtesy: Dundee Satellite Receiving Station).

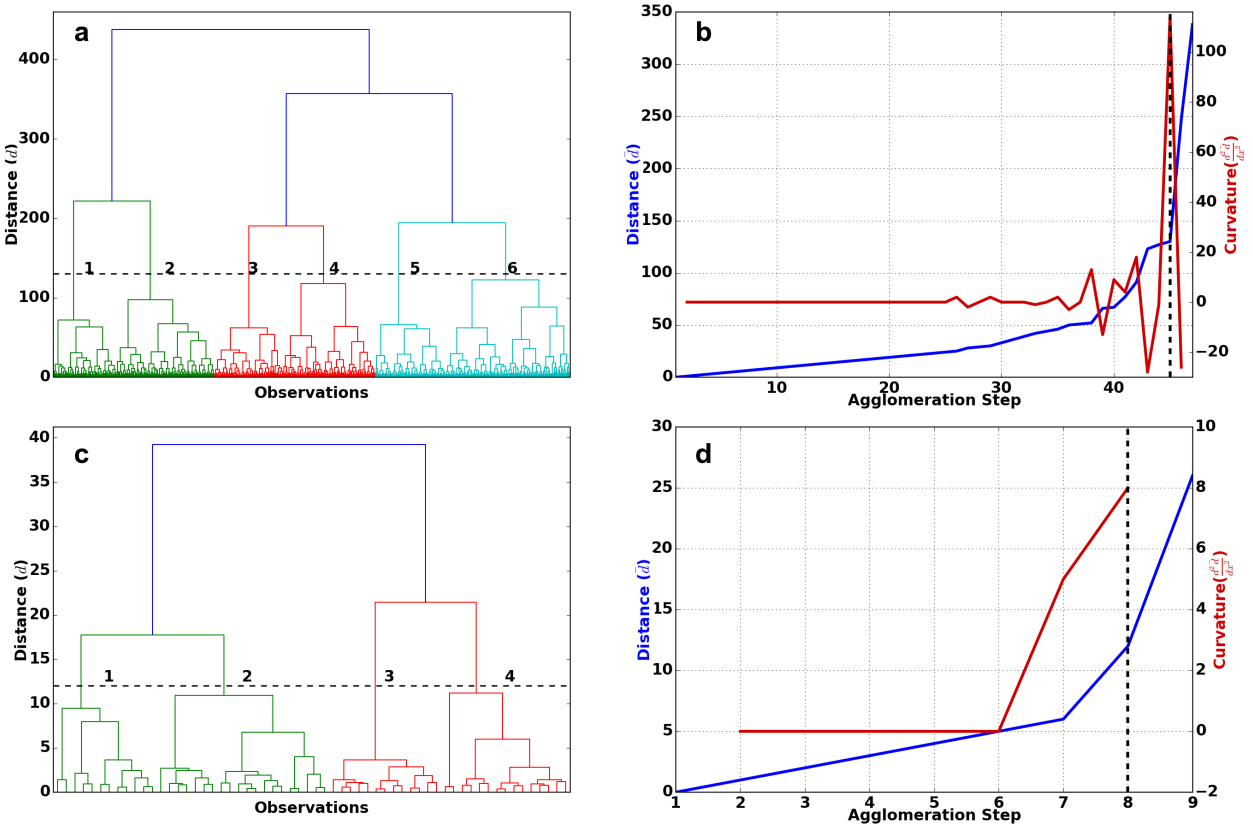


FIG. 2. (a) Dendrogram indicating the successive agglomeration of clusters from observations (x-axis) with increasing distance between clusters (y-axis). (b) $\overline{d}$ as a function of agglomeration step (blue) with the peak value of curvature ($\frac{d^2\overline{d}}{dx^2}$, red) indicated (dashed). See text for definition of $\overline{d}$. The mean distance at the agglomeration step when the peak curvature occurs is the classification cutoff distance as shown by horizontal dashed line in (a). The resulting 6 clusters are labeled in (a). (a and b) correspond to clustering of trajectories initialized at 06UTC (no. obs. = 9738) shown in Fig. 4. (c and d) correspond to super-clustering (no. obs. = 60) shown in Fig. 6.

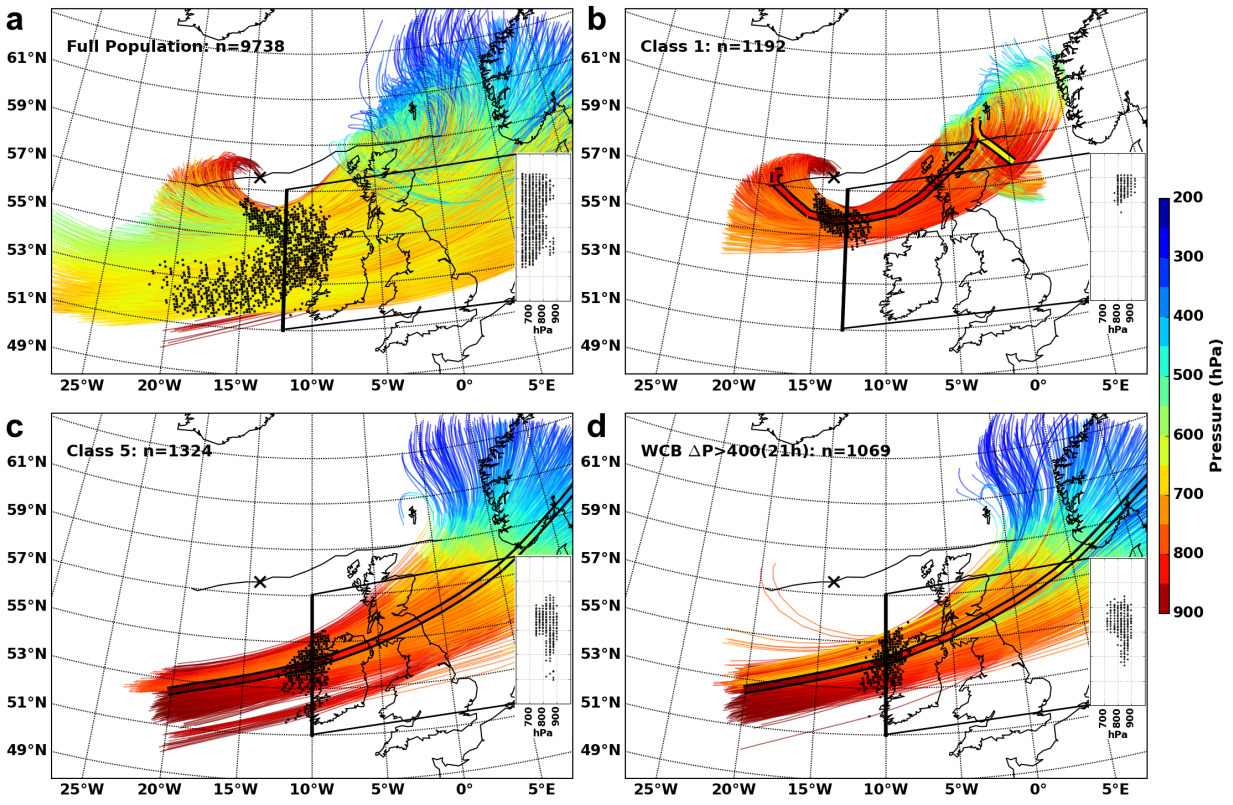


FIG. 3. (a) Full population of trajectories passing through low-level jet (windspeed exceeding 40 $\mathrm{m\,s^{-1}}$) at 0600 UTC 8 December 2011. (b) Trajectories assigned to cluster class 1 (CCB) with the class median overlaid (shaded and black-edged thick line). (c) As in (b) for cluster class 5 (WCB). (d) A WCB trajectory population obtained from thresholding for saturated ($RH > 90\%$) ascent ($\triangle p > 400$ hPa in 21 hours). Every 10^{th} trajectory is plotted in (a) with every 3^{rd} shown in (b-d). Dots indicate 0600 UTC initial locations of displayed trajectories, with insets showing vertical cross-sections of these start locations between 600hPa and 1000hPa through longitude drawn in bold. Trajectories are clustered over the period 0300-0900 UTC but trajectory histories are shown for the extended period 0100-2200 UTC.

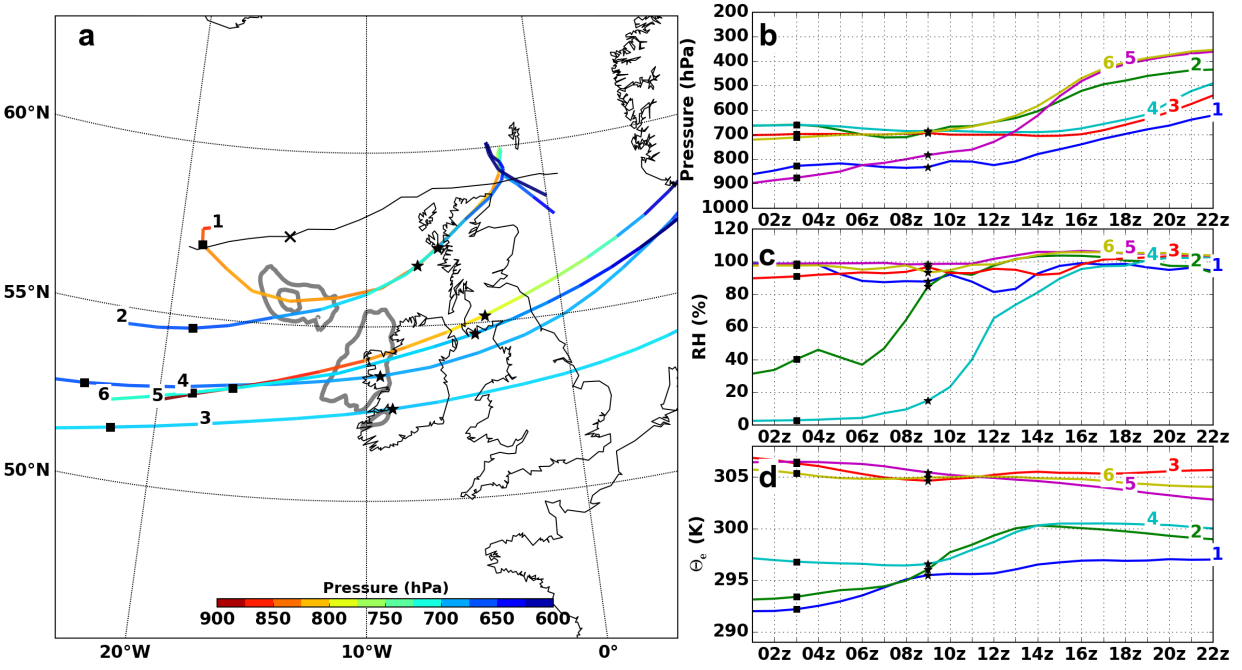
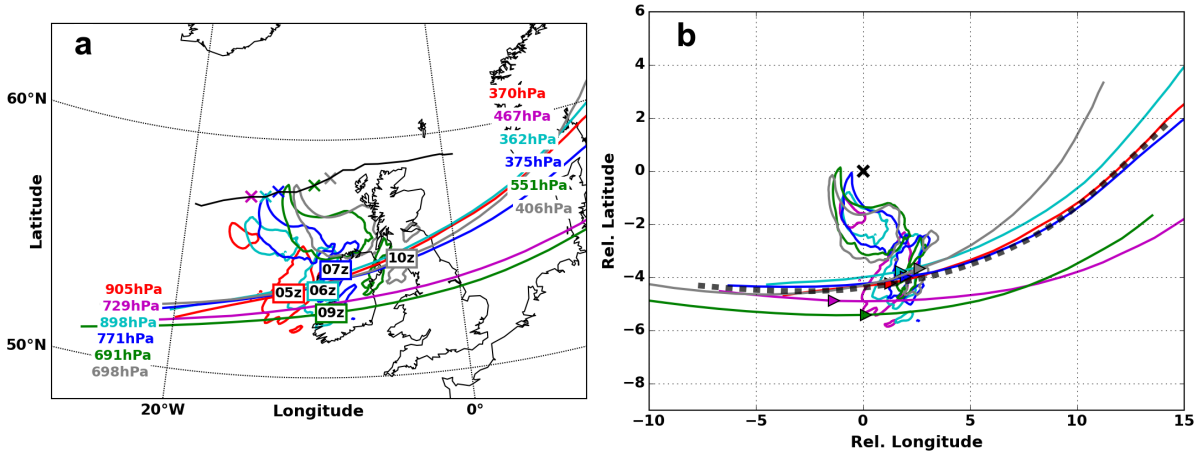
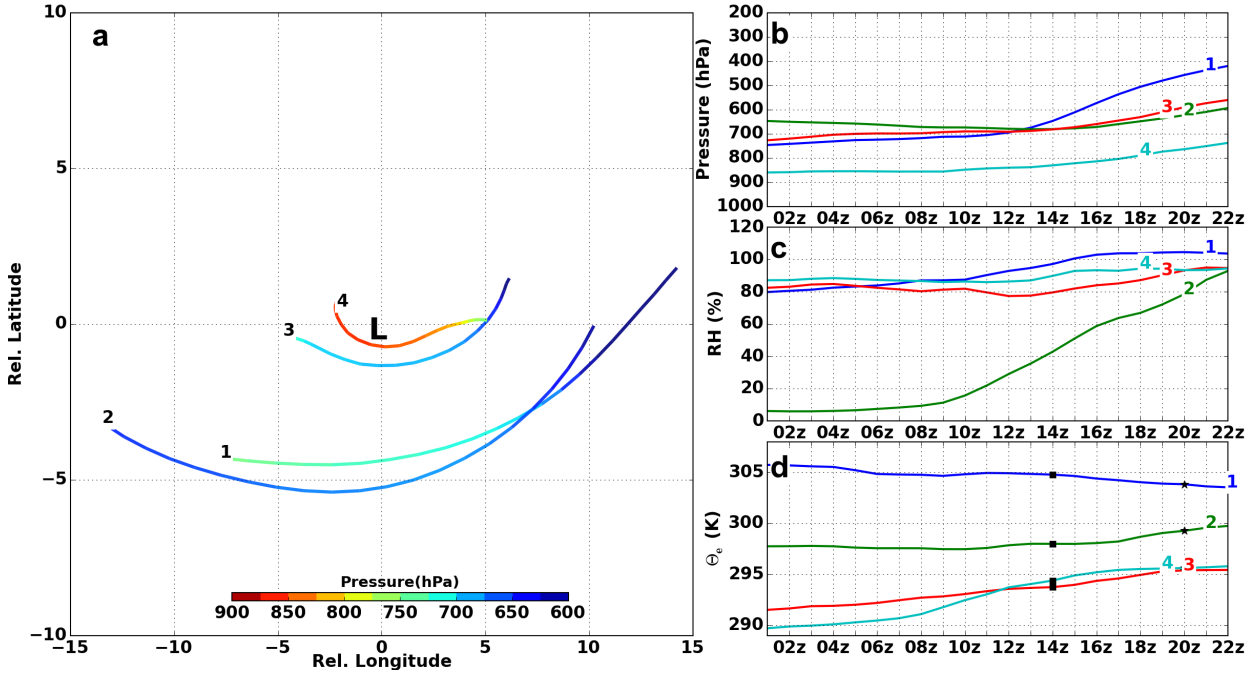


FIG. 4. Classification of trajectories arriving in low-level jets (windspeed exceeding 40 m s^{-1}) at 0600 UTC 8 December 2011 labeled in each panel by class number. (a) Class-median trajectories colored by pressure with the position of the minimum cyclone pressure marked by **X** at 0600 UTC on the smoothed cyclone track (black line). The direction of the trajectories can be determined from the numerical trajectory labels which are positioned near the beginning of the trajectories (at $t_0 - 3hrs$). Contours of 40 and 45 m s^{-1} windspeed are marked in gray. (b, c and d) Class-median evolutions of pressure (b) and relative humidity (c) and θ_e with the start and end times of the $[t_0 - 3hrs, t_0 + 3hrs]$ classification period denoted by squares and stars respectively.



579 FIG. 5. Illustration of the computation of a WCB super-class mean. (a) class-median trajectories with WCB
580 characteristics, as in Fig. 4a, but only showing every second class-median trajectory to avoid clutter. Contours
581 denote 40 m s^{-1} isotach at 850 hPa at each time, with the corresponding storm center marked on the storm track.
582 (b) system-relative class-median trajectories and isotachs (as for (a) but in system-relative coordinates) and the
583 super-class mean (dashed black line in (b)) computed from all system-relative class-median trajectories with
584 WCB characteristics in the period 0500–1200 UTC.



579 FIG. 6. Super-classification of all trajectories arriving in the low-level jet between 0500 UTC and 1700 UTC
580 8 December 2011. The longitude and latitude coordinates in (a) are relative to the center of the cyclone, marked
581 by 'L'. Otherwise the figure format is as in Fig. 4

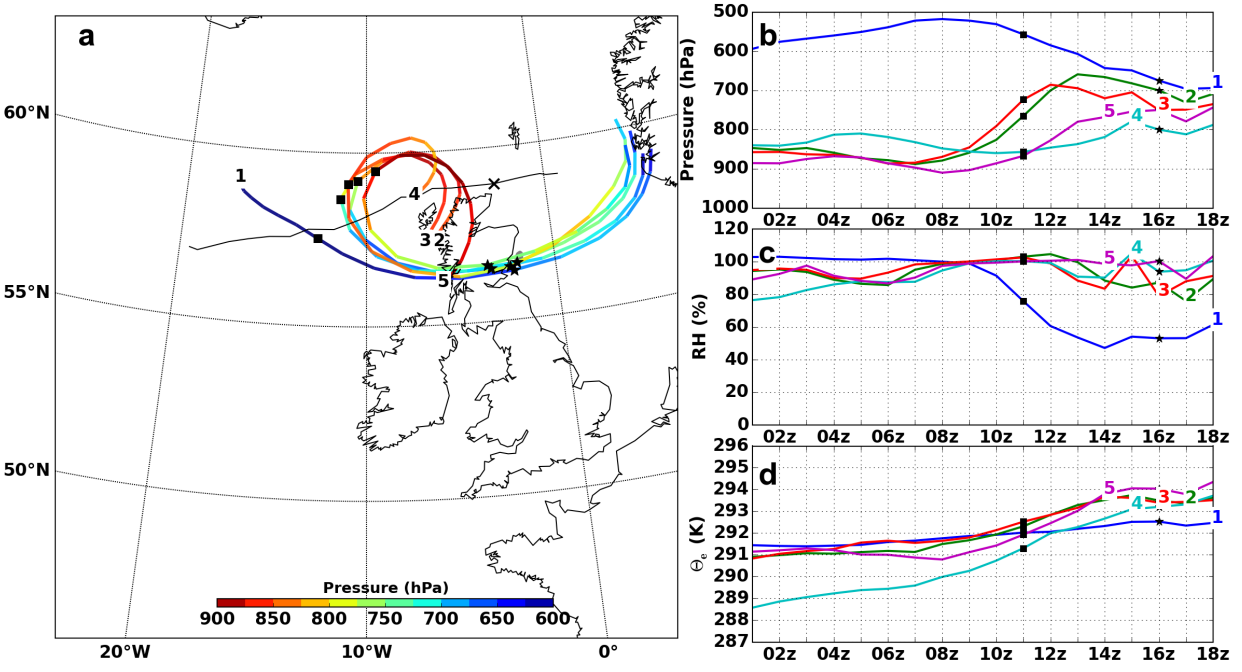


FIG. 7. Classification of trajectories arriving in the low-level jet region (windspeed exceeding 45 $\mathrm{m\,s^{-1}}$) in the frontal fracture zone at 1600 UTC 8 December 2011 labeled in each panel by class number. The format of the figure is as for Fig. 4.

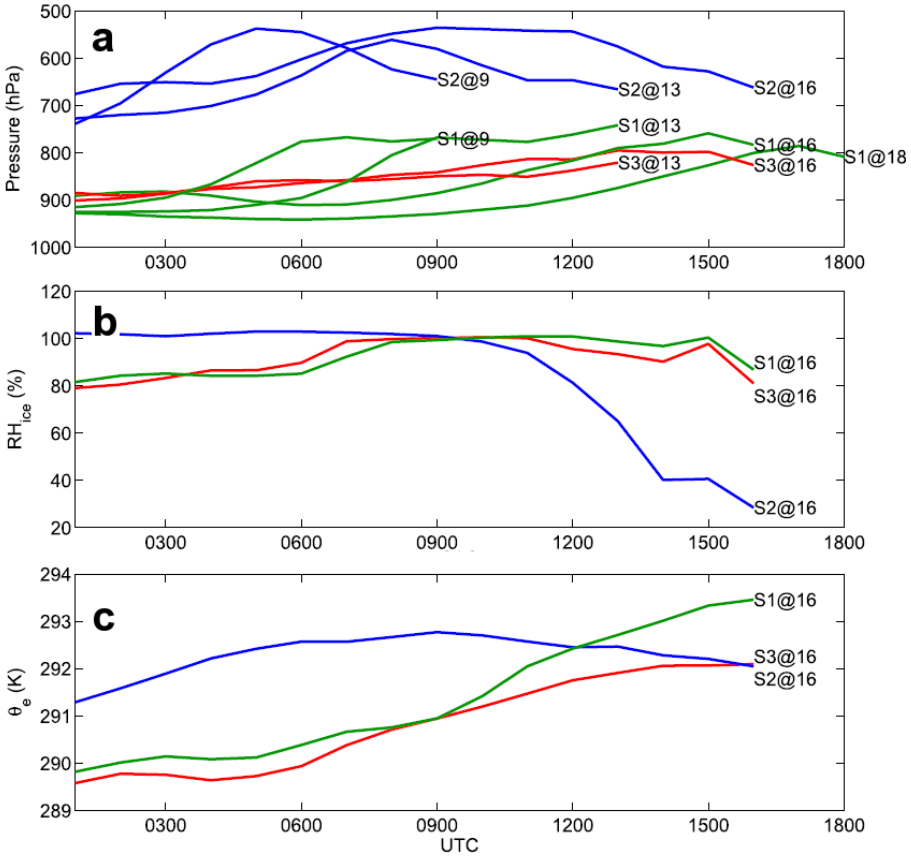
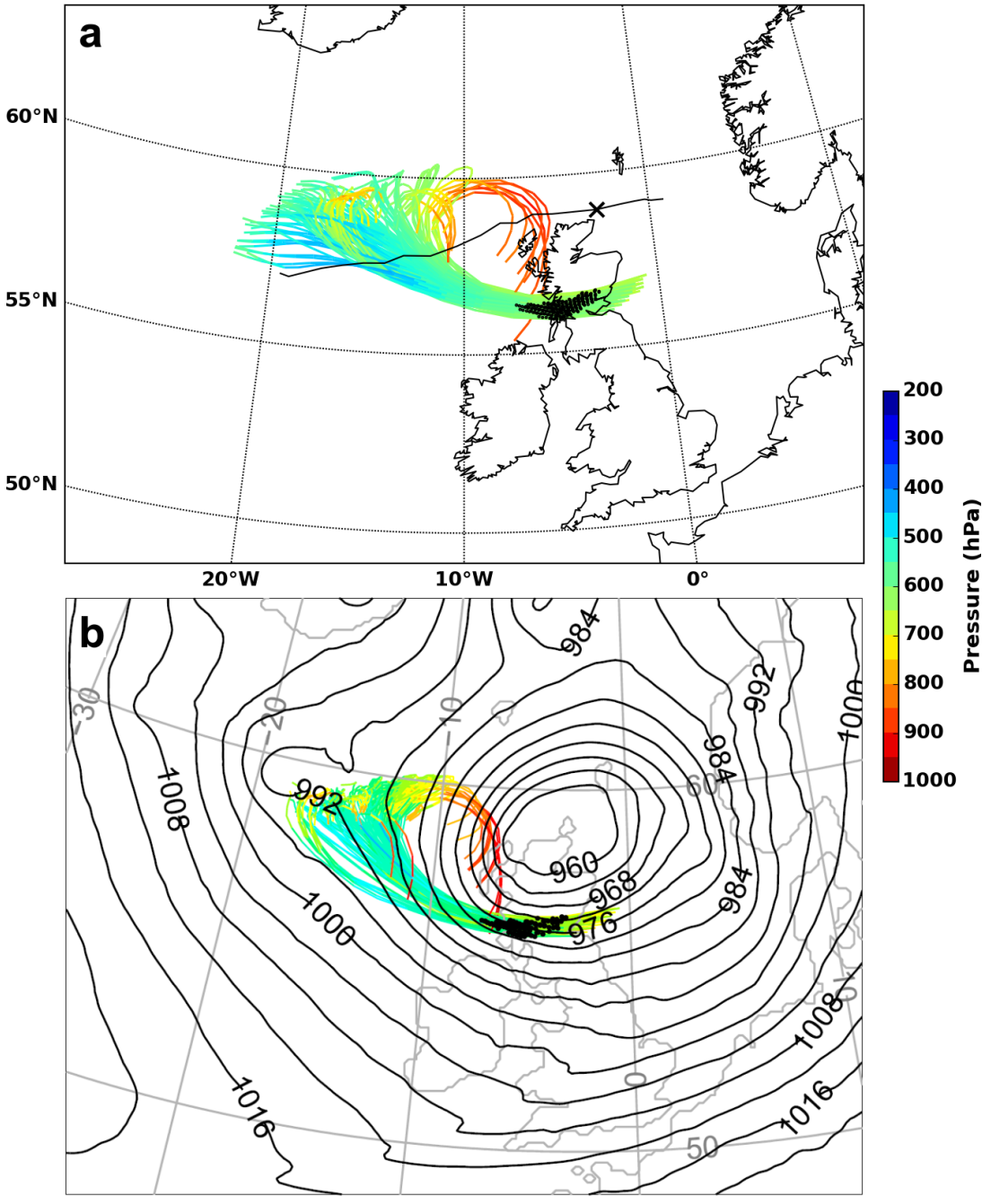


FIG. 8. Evolution of (a) pressure and (b) relative humidity and (c) θ_e with respect to ice along the ensemble median trajectories of the subjectively clustered trajectories arriving in the low-level jet region in the frontal fracture zone at 1600 UTC 8 December 2011 in MA14. The ensemble median trajectories represent airstreams which are labeled as "SX@HH", where X indicates the airstream number and HH indicates the arrival hour. Pressure evolutions are shown for airstreams arriving at four times whereas relative humidity evolutions are only shown for those airstreams arriving at 1600 UTC. Figure adapted from MA14, Fig. 8 (©American Meteorological Society. Used with permission.).



579 FIG. 9. Positional evolution of the trajectories constituting (a) agglomeratively clustered class #1 and (b)
580 the subjectively clustered S2 airstream (from MA14, Fig. 5 ©American Meteorological Society. Used with
581 permission.). The back trajectories are colored by pressure. (a) also shows the position of the minimum pressure
582 marked by **X** at 1600 UTC on the smoothed cyclone track (black line) and (b) also shows mean sea level pressure
583 at 1600 UTC (contours). Black dots in both panels represent the positions of the trajectories at 1500 UTC. The
584 trajectories in both panels extend backwards to 0100 UTC.