A New Method to Objectively Classify Extratropical Cyclones for Climate Studies: Testing in the Southwest Pacific Region

Jennifer L. Catto *

Monash University, Melbourne, Australia; University of Exeter, Exeter, UK

* *Corresponding author address:* Jennifer L. Catto, College of Engineering, Mathematics and Physical Sciences, North Park Road, University of Exeter, UK

E-mail: j.catto@exeter.ac.uk

ABSTRACT

Extratropical cyclones can vary widely in their configuration during cyclogenesis, development mechanisms, spatial and temporal characteristics, and impacts. An automated method to classify extratropical cyclones identified in the ERA-Interim reanalysis data from 1979–2010 in the Australia and New Zealand region has been developed. The technique uses K-means clustering on two upper tropospheric flow fields at the time of cyclogenesis and identifies four distinct clusters. Composites of these clusters are investigated, along with their lifecycles, and their spatial and temporal variability.

The four clusters are similar to a previous manual classification. Cluster 1 develops in the equatorward entrance region of the subtropical jet; clusters 2 and 4 develop in the poleward exit region of the subtropical jet but with different relative positions of the upper level trough and jet streak; and cluster 3 resembles secondary cyclogenesis on a pre-existing front far poleward of the subtropical jet. The clusters have different impacts in terms of their precipitation (cluster 1 has the highest average precipitation), different seasonal cycles, and different preferred genesis locations. Features of the composite cyclones resemble extratropical cyclones from other regions, indicating the utility of the method over larger regions. The method has been developed to be easily applied to climate model output in order to evaluate the ability of models to represent the full range of observed extratropical cyclones.

1. Introduction

Not all extratropical cyclones are created equal. While they are a ubiquitous feature of the midlatitudes, there are many different configurations of the atmospheric circulation conducive to their cyclogenesis. They can also develop and intensify in different ways, giving rise to differing impacts. It is these impacts—the precipitation and strong winds that they bring—that are of socioeconomic importance, and that are one of the reasons for the great interest in how extratropical cyclones will change in the future (e.g., Kirtman et al. 2013).

There have been many techniques and methods used to distinguish and classify different flavors of extratropical cyclone genesis and development (see the review by Catto 2016)). Some of these methods have developed from a weather forecasting point of view (e.g., Young 1993) or focused on small regions (e.g., Evans et al. 1994). Past studies have considered the cloud features using satellite imagery (e.g., Zillman and Price 1972; Evans et al. 1994), with certain cloud features visible in satellite images giving information about how the surface cyclone below is developing. Others have considered atmospheric precursors (e.g., Dacre and Gray 2013; Graf et al. 2016), or considerations of upper versus lower level forcing (e.g., Petterssen and Smebye 1971; Deveson et al. 2002; Graf et al. 2016) to better understand the variability of cyclone development. Sinclair and Revell (2000) (hereafter SR00) performed a manual classification on 40 cyclones (from 1990–1994) identified in the Australia and New Zealand region. Their classification was based on upper tropospheric flow features (300 hPa wind speed and geopotential height) and yielded 4 distinct classes. Despite the use of only 2 variables, these classes exhibited differences in frontal development and cyclone structure through the lifecycle.

In this paper, a relatively simple automated method based on clustering has been developed to classify extratropical cyclones in the Australia and New Zealand region during the period 1979–2010, using an objective feature tracking method applied to the European Centre for Medium Range Weather Forecasting (ECMWF) reanalysis product, ERA-Interim (Dee et al. 2011). The results of the automated classification used here will be compared using composites against the manual classification of SR00 to determine the wider utility of such a method to realistically depict the variation of cyclone types. Composites have been used in a number of previous studies to investigate the most important features of extratropical cyclones, while removing some of the noise associated with individual systems (e.g., SR00; Field and Wood 2007; Catto et al. 2010; Hawcroft et al. 2016). As well as comparing the features studied in SR00, additional fields are investigated here to further understand the different impacts of the cyclones. The average frequency of cold and warm fronts and their associated precipitation are determined for the different cyclone classes.

The spatial and temporal variability of the cyclone clusters are investigated using the long period of reanalysis data. The Australian and New Zealand region are strongly influenced by interannual variability related to the El Niño Southern Oscillation (ENSO), and by the Southern Annular Mode (SAM), so a question arises of how these may affect the types of cyclones that occur in this region.

A further goal of this work is to develop a method that could be easily applied to data from climate models such as the 5th Coupled Model Intercomparison Project (CMIP5; Taylor et al. 2012)). This methodology will provide an objective means to evaluate the realism of climate models against observational data and to intercompare different climate models. In order to understand how extratropical cyclones and their associated impacts may change in the future, we need to first investigate whether climate models—our primary tool in providing projections of future climate—are able to represent the full spectrum of extratropical cyclone behavior.

Many studies have evaluated the mean extratropical storm tracks in climate models (e.g., Ulbrich et al. 2008; Catto et al. 2011; Colle et al. 2013; Zappa et al. 2013) and overall these seem to be improving over time (Flato et al. 2013). The dynamical structure of the most intense extratropical cyclones in the High Resolution Global Environment Model (HiGEM; Shaffrey et al. 2009) in the NH was found to be well represented compared to ERA-40 reanalysis (Catto et al. 2010), however, many studies show an underestimate of average cyclone intensity (Zappa et al. 2013), and a negative bias in the frequency of the most intense extratropical cyclones (e.g., Seiler and Zwiers 2015).

Studies have shown that models can have an incorrect distribution of clouds within extratropical cyclones (Field et al. 2008; Naud et al. 2010; Booth et al. 2013; Govekar et al. 2014; Hawcroft et al. 2017). The relationship between the dynamical and cloud features in models may also be poorly represented (Govekar et al. 2014). In order for models to represent the cyclone intensification associated with diabatic processes, high resolution is required (Willison et al. 2013).

Following these results, it seems possible that extratropical cyclones that are more dependent on latent heating for their development will be less well represented in climate models. This sort of detail would be masked by the consideration of all types of cyclones together. Separating out different classes may lead to added insight into the representation of these features in models.

The objective classification method described here offers a way of grouping cyclones, with atmospheric variables that are widely available from climate model simulations. Manual techniques are impractical for long periods and for multiple models, and they are also not a repeatable methodology—a different synoptic practitioner may place the cyclones into different classes. In order to test the method, the study region has been chosen to match with SR00. However, if it is seen to produce similar groups of cyclones to the manual classification, it gives motivation to further develop the technique to be applied on larger regions (potentially global) in order to be able to compare against climate model output.

Questions that this paper seeks to address are:

i. Can a simple clustering method identify previously defined classes of extratropical cyclones?

ii. What are the distinct characteristics of cyclones in the various clusters?

iii. What is the spatial and temporal variability of the cyclones in the different clusters?

In section 2, a brief description is given of the data used, and the cyclone identification method. A longer explanation of the selection of tracks and the clustering method are also given. Section 3 gives a description of the clusters found using the described method, and how they compare with SR00. Section 4 explores the spatial and temporal variability of the clusters, and section 5 gives a summary and discussion.

2. Data and Methods

a. Data

Data from the European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis product (ERA-Interim), extracted at 1.5° resolution, have been used for the period 1979–2010. Most of the fields have been obtained directly from the reanalysis, except for the 850 hPa relative vorticity, which is required for the objective features tracking, and is calculated from the zonal and meridional wind, and the frontogenesis function (described below). Precipitation is also taken from ERA-Interim forecast fields and represents 6-hourly accumulations (in the 6 hours following the analysis time) from the 0–6hr and 6–12hr forecasts from 0000Z and 1200Z (as used in Catto and Pfahl 2013). Hawcroft et al. (2016) showed that there is a low bias in the precipitation from ERA-Interim in the first 12 hours of the forecast due to model spin-up. In their analysis, Hawcroft et al. (2016) opted to use the 12–24hr forecast lead time precipitation to overcome this bias. The sensitivity of the results to the use of the data from Hawcroft et al. (2016) has been investigated (see Appendix), and found not to alter the conclusions of the study.

The front identification of Berry et al. (2011) based on the work of Hewson (1998) has been used. This algorithm has been described previously in Berry et al. (2011); Hope et al. (2014); Catto and Pfahl (2013); Catto et al. (2015) (among others), so only a very brief description of the method is given here. Frontal points are identified where the gradient of a thermal front parameter (TFP) is zero, where $TFP(\theta_w) = -\nabla|\nabla\theta_w|.(\nabla\theta_w/|\nabla\theta_w|)$ and θ_w

is the wet bulb potential temperature on 850 hPa, after the TFP field has been masked out above a threshold value. The fronts are separated into cold, warm and quasi-stationary, and only the warm and cold fronts are used in this study. The fronts are identified using the ERA-Interim data degraded to a resolution of 2.5° as in Catto et al. (2012).

Frontogenesis has been calculated on the 850 hPa level using the temperature and wind fields from ERA-Interim and with the Petterssen (1936) definition:

$$\frac{d}{dt}|\nabla_p|\theta = -\frac{1}{2}|\nabla\theta|(D - E\cos 2\beta) \tag{1}$$

where D is the divergence ($D = \partial u/\partial x + \partial v/\partial y$), E is the total deformation magnitude, $E = (E_{st}^2 + E_{sh}^2)^{1/2}$, where $E_{st} = \partial u/\partial x - \partial v/\partial y$ (the stretching deformation) and $E_{sh} = \partial v/\partial x + \partial u/\partial y$ (the shearing deformation), and β is the angle between the isentropes and the axis of dilatation.

b. Cyclone Identification and Tracking

Extratropical cyclones are identified using the objective feature identification and tracking methodology of Hodges (1994, 1995, 1999), and demonstrated in Hoskins and Hodges (2002, 2005). This method identifies cyclones in the SH as minima (maxima in the NH) of the 850 hPa relative vorticity, which is first truncated to spectral T42 resolution (approximately 300 km grid-spacing) in order to maintain only the synoptic scales. Cyclone centers are grouped into tracks by first applying a nearest neighbor search, and then by minimizing a cost function to determine the smoothest tracks. In order to keep only mobile cyclones in the dataset, two further criteria are applied; the cyclones must live for at least 2 days (8 time steps), and travel at least 1000 km in order to eliminate quasi-stationary features such as heat lows. Cyclogenesis is defined as the first point that is identified for each track.

For much of the analysis that follows, the spatial fields surrounding the cyclone center are considered. These fields are extracted from the ERA-Interim data by using a radial coordinate system (e.g., Bengtsson et al. 2006; Catto et al. 2010). A spherical cap is centered on the pole, and rotated to the cyclone center. The latitude-longitude gridded field is then regridded on the 20° radial cap and saved for each cyclone and track point.

c. *Cyclone track selection*

In order to be able to sensibly compare the results of the cyclone classification with SR00, a number of steps were performed in the selection of the tracks to analyze.

i. Selection of tracks within a certain region.

The same region of study as used in SR00 was chosen here (25°S–50°S and 150°E–150°W; shown by the black boxes in Figure 1). SR00 selected only cyclones with their entire lifetime within the box. Due to the typical length of the cyclone tracks identified here (mean of 16.9 points — just over 4 days — for all cyclones in the SH), a less restrictive criterion was used, so that cyclones with their genesis (first identified point) and at least 50% of their track points within the box are chosen.

ii. Selection of developing cyclones.

Here the evolution of the T42 resolution central vorticity at 850 hPa is used to identify cyclones that develop after their identified genesis time. The same criteria as SR00 are used; the central cyclone intensity (the T42 vorticity at the track point) must be less than 4 cyclonic vorticity units ($1\,\mathrm{CVU} = -1 \times 10^{-5}\, s^{-1}$ in the SH) at the time of genesis, but increase to greater than 6 CVU at some point later in its lifecycle.

iii. Selection of cyclones of sufficient strength (circulation).

SR00 also used the criterion of the cyclones reaching at least 6 circulation units (CU; $1\,\mathrm{CU} = 1 \times 10^{7}\, m^2 s^{-1}$) using the methods described in Sinclair (1997), and we have applied the same method to our data as follows. First the region where cyclonic vorticity is decreasing with distance from the cyclone center is defined on the cyclone-centered radial grid. Next, the circulation (C) of the cyclones is calculated using the T42 vorticity within this region as; $C = \sum_{n=1}^{P} \xi_{T42,n} A_n$, where P is the number of grid boxes within the specified area, $\xi_{T42,n}$ is the T42 resolution vorticity for each grid box, and A_n is the area of each grid box.

The number of cyclones in the dataset for each season at each of these steps is given in Table 1. Considering all the cyclone tracks identified in the SH, the summer season (December, January and February; DJF) has the fewest, consistent with previous studies (Simmonds and Keay 2000a,b). Spring (September, October and November; SON) and Autumn (March, April and May; MAM) feature the largest number of cyclone tracks, likely associated with the maxima in mid-tropospheric temperature gradients seen at high latitudes in these seasons (van Loon 1967). However, many of these occur around the edge of Antarctica (not shown), and are not relevant for this study. Once the region selection has taken place, the highest frequency of cyclones occurs during winter (June, July and August; JJA) when the subtropical jet is at its strongest, with a similar number in SON, and the lowest during DJF. This is consistent with other climatologies of cyclones in the Australian region (e.g., Speer et al. 2009), and the latitudinal variations shown by Simmonds and Keay (2000a). Once the filtering for strong, developing cyclones has taken place, JJA has by far the largest number of cyclones, indicating the occurrence of the strongest cyclones in winter. The final number of cyclones included in the clustering is 483. These tracks have a mean track length of 24.6 points (just over 6 days), with one track of 78 points (19.5 days).

The genesis density and track density (both with units of cyclones per month per 5° radius circle) are shown in Fig. 1. Note that although no cyclones have their genesis outside of the region of interest, because of the smoothing that occurs with the counting, the genesis density outside the box is non-zero in some places. DJF (Fig. 1a) shows the minimum genesis density in the region, while MAM (Fig. 1b) and JJA (Fig. 1c) show the maximum. In all seasons there is a local maximum of cyclogenesis close to the east coast of Australia, which is most pronounced during winter. This is consistent with the higher frequency of east coast lows (ECLs) identified in these seasons (e.g., Dowdy et al. 2013). During DJF, the highest cyclogenesis occurs to the west of New Zealand. Most of the cyclogenesis in all seasons occurs in the western half of the study region, associated with the criterion of 50% of track points lying within the box, and the typical eastward propagation of the systems. The track density statistics (Fig. 1e-h) further highlight this feature, with large track density values across much of the region. During winter, the maximum track density lies at a lower latitude than the other seasons, but there are more cyclones that track further poleward (indicated by the higher values of track density to the south of the region in the box). This is potentially an artefact of the region selection, with the winter cyclones remaining in the box long enough to fulfill the selection criterion of 50% of points within the box more often than the summer cyclones.

d. Clustering

In order to automatically group the cyclones according to their similarities, a simple clustering method is employed. Graf et al. (2016) showed in their recent classification that cyclone genesis features exist as a continuum. However, the wide variability allows classes with distinct features to be identified. Clustering has been used in the fields of meteorology and climate since the late 1960s, and there are a number of different available methods (see review by Gong and Richman 1995). Ayrault et al. (1995) used clustering on 850 hPa vorticity to separate different subsynoptic cyclone types over a small region of the UK. Here K-means clustering has been used. K-means clustering specifically has recently been used in a number of ways related to synoptic scale meteorology: to identify distinct synoptic scale patterns associated with different vertical profiles from radiosonde data in Northern Australia (Pope et al. 2009b); to identify wind regimes in the Antarctic region (Coggins et al. 2013); on spatial patterns of precipitation to again identify important synoptic scale features (Raut et al. 2014); and to objectively identify extratropical transition of tropical cyclones (Studholme et al. 2015).

1) Fields used

Fields of upper level winds (250 hPa; U_{250}) and anomalies from the zonal mean of potential temperature on the tropopause (where potential vorticity is -2PVU; θ_{PV2}) have been calculated from the ERA-Interim dataset and used in the clustering. The cyclone-centered spatial fields are used as the basis for the clustering. The 20° radial region of these fields around the cyclone center are extracted at the time of genesis for each cyclone. The full spatial patterns are used in the clustering algorithm, rather than any single point or value. The 20° radial fields for each cyclone are normalized using the mean and the standard deviation of the spatial pattern. This normalization is necessary when using two fields (winds and potential temperature) of different magnitude and variability.

2) Clustering method

K-means clustering is an iterative technique that objectively groups objects (in this case cyclones) that are most similar to each other. With this type of clustering method, the number of clusters (often referred to as K) needs to be chosen *a priori*. In order to compare the results with those of SR00 four clusters were chosen. The clustering algorithm was also applied using differing values of K in order to investigate the variability of cluster centroids that would result. With only 3 clusters, one of the classes of SR00 was clearly missing. For 4 clusters and above, the main classes are usually visible in the cluster centroids. The centroids of the clusters for K=3, 5, 6, and 7 are shown in supplementary figures 1-4.

Methods to objectively select the most appropriate number of clusters have been suggested (Gong and Richman 1995; Rossow et al. 2005; Pope et al. 2009a). Many of these are suited to datasets where the clustering is performed on a single variable, thereby allowing the changes in cluster means to be analyzed statistically as the number of clusters is increased. Although these methods have not been used here to choose the number of clusters, some analysis has been done to check that the choice of 4 clusters is statistically sensible. An elbow analysis (investigating the proportion of explained variance as the number of clusters increases), applied to the spatial fields of both wind speed and potential temperature on the tropopause, indicates that the increase in the explained variance slows after 4 clusters (see Supplementary Figure 5), and again after 7 clusters. On inspection of the centroids produced for 7 clusters (see Supplementary Figure 4), it could be seen that some centroids were very similar to each other and so 7 clusters was deemed to be too many. This analysis gives some confidence in the choice of the number of clusters and the choice of 4 classes by SR00. When considering larger regions without a manual classification against which to compare, the use of such objective methods will be more important.

The steps in the clustering are as follows:

i. The two-dimensional spatial patterns of U_{250} and θ_{PV2} are read in and normalized.

ii. A random cluster number is generated from 1 to K for each cyclone, and the initial cluster centroids for U_{250} and θ_{PV2} are calculated by averaging the spatial patterns of these initial clusters.

iii. The Euclidean distance between each cyclone (p) and each cluster centroid (q) is calculated using equation 2 for each field, where there are n points in the normalized 20° radial spatial patterns, and the squared sum of the distances from the two fields are calculated.

$$d_{p,q} = \sqrt{\sum_{i=0}^{n} (p_i - q_i)^2} \tag{2}$$

iv. The cyclones are allocated to the cluster to whose centroid they are closest and the new centroids are recalculated.

v. Since there is some sensitivity to the initial random cluster allocation, for 50 iterations, 30 cyclones are randomly changed to different clusters before the new centroids are calculated. After these 50 iterations, the clusters are allowed to converge. Once there are no more shifts between clusters (convergence has been achieved), the algorithm terminates.

e. Climate indices

Extratropical cyclones in the Australian and New Zealand region are influenced by large-scale climate drivers (e.g., Rudeva and Simmonds 2015) that impact the subtropical jet and hence may determine the variability of the occurrence of the different cyclone clusters. The El Niño-Southern Oscillation (ENSO) index used is based on the Niño3.4 region (5°N-5°S, 120°-170°W) sea surface temperature (SST) anomalies (calculated from the Extended Reconstructed Sea Surface Temperature (ERSST) v4 (Huang et al. 2015)), and obtained from `http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml`). Seasons where the index is less than -0.5°C are classed as La Niña, and seasons where the index is greater than 0.5°C are classed as El Niño.

The SAM index is based on pressure differences between 40°S and 65°S (Marshall 2003), obtained from `https://legacy.bas.ac.uk/met/gjma/sam.html`, and describes the meridional movement of the strongest westerlies. SAM negative seasons are those with SAM index less than -0.5 and SAM positive seasons are those with SAM index greater than 0.5. While an index of greater than 1.0 or less than -1.0 may often be considered, this was found to create sampling issues and was not used here.

3. The clusters and how they compare with manual classification

The composites of cyclone-centered fields at the time of genesis are given for the 4 clusters in Figures 2–5. The number of the cluster is an arbitrary label, but will be used consistently in the remainder of the paper. The number of cyclones represented by clusters 1–4 are 139, 123, 104, and 117 respectively.

Cluster 1 (Fig. 2) shows cyclogenesis on the equatorward side of the main jet streak, downstream of an upper level trough (Fig. 2a), and resembles the equatorward entrance (E) class of SR00 (see their Figure 8). This cluster is one of the clusters with the highest equivalent potential temperature at 850 hPa ($\theta_{e,850}$) at the time of cyclogenesis with values up to 325 K in the northern region of the composite, and a strong temperature gradient in the cold frontal region to the west of the cyclone center (Fig. 2b). At the time of genesis in the north of the composite the surface winds are easterly and the feature resembles an easterly dip (e.g., Holland et al. 1987). This is somewhat different to the surface pattern seen for class E in SR00. At the time of genesis there are already large values of frontogenesis in both the warm frontal and cold frontal regions (Fig. 2b), which occurs in a classic deformation flow field due to the low pressure systems to the north and south, and high pressure to the west and east. As the cyclone develops the frontogenesis in the warm frontal region increases markedly, while that in the cold frontal region and the cyclone center decreases (Fig. 2e,h), consistent with the analysis of the equatorward entrance class of SR00. Surface winds are diffluent in the region of the strongest temperature gradient, which contributes to the strong frontogenesis that occurs between genesis and maximum growth. Highest values of composite precipitation occur where there is maximum frontogenesis (Fig. 2c), with the heaviest precipitation occurring at the time of maximum growth (Fig. 2f). The central pressure decreases from about 1002 hPa at the second stage, to 990 hPa at maximum intensity.

Cluster 2 closely resembles the Class D composite from SR00 (see their Figure 9), with the upper level jet streak downstream of the upper level trough, and the cyclone genesis occurring in the poleward jet exit region (Fig. 3a). There is a fairly deep upper level trough with its axis to the west of the cyclone center. As the composite cyclone develops, the jet streak appears weaker, and the upper level trough starts to tilt to the east (Fig. 3d). By the time of maximum intensity, there is a distinct cut off almost directly above the surface low

(Fig. 3g). At the time of cyclogenesis, there is frontogenesis greater than $4\,K/10^3\,km/day$ in both the cold front and warm front regions (Fig. 3b). Only cold fronts are visible along the baroclinic zone from northwest to southeast (Fig. 3c). As the cyclone undergoes its maximum growth, there is an increase in frontogenesis in the warm front region (Fig. 3e,f), accompanied by an increase in precipitation in this region (Fig. 3f), and the identification of high frequency of warm fronts.

Cluster 3 is most similar to the T Class (trough) of SR00, with the cyclone forming underneath a deep upper level trough, far from the main jet streak (Fig. 4a). This upper level trough appears much like a PV streamer (e.g., Martius et al. 2008), eventually becoming a cut off feature lying almost directly above the surface cyclone center by the time of maximum intensity (Fig. 4g and h). In the SR00 Trough class, there was a single jet streak visible to the north of the cyclone center, whereas in the composite shown for cluster 3 here, there is also a jet streak to the west of the upper level trough (Fig. 4a). At the time of genesis, there is no closed pressure contour in the composite (Fig. 4b), rather a strong indication of genesis occurring on a pre-existing cold front associated with a low pressure system at higher latitudes; so-called "secondary cyclogenesis" (e.g., Parker 1998). At the time of genesis, frontogenesis can be seen mainly in the cold frontal region (Fig. 4b,c), associated with the cold advection from the southwest. As the surface cyclone develops a closed contour of MSLP, frontogenesis can be seen in the warm frontal region and by the final stage, warm fronts can be identified, collocated with the main precipitation region (Fig. 4f, i). This cluster has the weakest warm front region due to relatively weaker warm advection. The maximum in cold front frequency seen at maximum intensity to the west of the cyclone center (Fig. 4i) is more likely a bent-back warm front (which the front identification would pick up as a cold front).

Cluster 4 resembles the Class U (upstream exit) composite from SR00 (see their Figure 7). The cyclone develops in the poleward jet exit region, as with cluster 2, but here the upper level jet streak is upstream of the upper level trough (Fig. 5a). As with the other clusters, there is a deep upper level trough which undergoes cyclonic rotation as the lifecycle progresses (Fig. 5a, d, g). There is no cut off feature at upper levels by the time of maximum intensity, as there is in Cluster 2. At the time of genesis, there is a rather zonally oriented baroclinic zone with temperatures up to about 320 K in the north of the composite (Fig. 5b). There are quite weak surface winds, resulting in weak cold advection, slightly weaker frontogenesis, and fewer cold fronts identified, compared to the other clusters (Fig. 5c). There are already quite large values of precipitation in the warm front region at the time of genesis (Fig. 5c). By the time of maximum growth, the frontogenesis has increased markedly (Fig. 5e), especially in the warm front region, and the precipitation has increased (Fig. 5f). The strong development of the warm fronts can be related to the fairly strong warm advection seen on the east of the cyclone.

a. Cluster life cycles

The clusters clearly have a number of synoptic orientations. An interesting question is whether these characteristics impact on the development and the life cycles of features of the composite cyclones. Figure 6 shows the life cycles of the cluster composites centered on the time of maximum vorticity (at T42 resolution). The life cycles of vorticity (Fig. 6a; this is the full resolution vorticity rather than the T42 resolution; note that the vorticity has been multiplied by -1 as cyclonic vorticity is negative in the SH) show that the average composite maximum vorticity 120 h before the time of maximum intensity is about 10 CVU. About 60 h before maximum intensity the vorticity begins to increase in all clusters, although cluster 3 seems to develop slightly later than the others and has a quicker increase in vorticity (consistent with the rapid development seen in secondary cyclogenesis Parker 1998). Cluster 1 vorticity begins to increase from 72 h prior to maximum intensity, likely associated with the latent heating from the increase in precipitation seen around the same time (Figure 6c), and reaches the highest peak vorticity of all the clusters. The maximum vorticity is reached 6-12 h after the maximum precipitation for all clusters.

The MSLP lifecycles look remarkably similar between the clusters in the period before the maximum intensity (which coincides with the minimum MSLP). All clusters show a slow decrease in pressure until about 60 h before maximum intensity, then an increased rate of deepening until the minimum pressure. After the time of maximum intensity, the MSLP lifecycle curves diverge, likely related to the path the cyclones take. For example, cluster 2 maintains quite a low central MSLP as it moves poleward into a region of climatologically low pressure (see Supplementary Figure 8). Wind speeds in the cyclones also do not vary much between the clusters. Maximum wind speeds are seen at the time of maximum vorticity, and cluster 1 has slightly higher maximum values than the other clusters.

There is a large amount of variability within the clusters, indicated by the dashed lines at 1 standard deviation. Despite the different synoptic orientations, the lifecycles of MSLP are similar. There is a larger difference between the clusters in the lifecycles of impactful variables (precipitation and winds). To further determine the significance of the differences in precipitation between the clusters normal distributions with cluster mean and standard deviation of the maximum precipitation along the tracks in each cluster (averaged within 5° of the cyclone centers) are shown in Figure 7 along with the mean values and the statistical significance. The maximum precipitation in clusters 1 and 4 are not statistically significantly different to each other, which can be seen by how close the density functions are for these clusters. However, the differences between all the other clusters are statistically significantly.

b. Summary of Clusters

Overall the clusters identified using this automated method closely resemble the manually identified classes from SR00. The upper level features, surface and low level temperatures, and the frontogenesis seen in these clusters are mostly consistent with those seen in the classes of SR00. Cluster 3, which looks like class T from SR00, here looks very much like secondary frontal cyclogenesis, and is associated with highly amplified upper level flow in the form of a PV streamer and cut off PV feature (e.g., Wernli and Sprenger 2007). This type of upper tropospheric feature is also seen in the B_{moist} class of Graf et al. (2016), but this occurs predominantly in the lower midlatitudes in contrast to cluster 3. The other clusters are more associated with cyclonic wave breaking at upper levels, which has previously been shown to be associated with a large number of intense extratropical cyclones in the North Atlantic region (Gómara et al. 2014).

The inclusion of the precipitation and objectively identified fronts gives additional information about the potential impacts of the different classes. Cluster 1, similar to the equatorward entrance class of SR00, produces the highest volume of precipitation. This is associated with the advection of warm moist air from the northeast on the equatorward side of the jet. Cluster 1 also reaches the highest central vorticity and 925 hPa wind speeds (Fig. 6), which is consistent with a number of other studies showing the strong relationship between cyclone intensity and precipitation (Chang and Song 2006; Rudeva and Gulev 2011; Pfahl and Sprenger 2016).

While cluster 1 develops in confluent upper level flow on the warm equatorward side of the jet entrance, clusters 2 and 4 develop in diffluent upper level flow on the cold poleward side of the upper level jet. Cluster 2 has a larger number of identified cold fronts throughout the lifetime of the composite cyclone, with high frequency of cold fronts at genesis, while cluster 4 has higher frequency of warm fronts and associated higher precipitation.

4. Spatial and Temporal Variability of the Clusters

a. Genesis locations and tracks

The second goal of the study is to investigate if the identified clusters have distinct characteristics in their preferred locations and their temporal variability. There is some differentiation in the spatial distribution of the cyclogenesis in the identified clusters. Figure 8 shows the genesis locations of each of the cyclones within the clusters. Cluster 1, whose composite shows cyclogenesis on the equatorward side of the jet, has a high density of cyclogenesis in the northern part of the region, with 45% of the cyclogenesis of this cluster occurring at or northward of 30°S. Cluster 4 also has many cyclogenesis events in the north of the region, explaining the warm temperatures in the northern part of the composite cyclone (Fig. 2). Cluster 2 cyclogenesis occurs mostly at higher latitudes than clusters 1 and 4. Cluster 3 has the highest latitude cyclogenesis on average and shows a high density of cyclogenesis around New Zealand, similar to the climatology of cut off lows found by Fuenzalida et al. (2005). This is different to the class T cyclones of SR00 (structurally similar to cluster 3), which predominantly develop off the coast of Australia. In contrast, the cyclogenesis regions of the other clusters are bunched around the east coast of Australia—a region of climatologically high cyclogenesis (Hoskins and Hodges 2005). The seasonal changes in the genesis locations (Fig. 8b-e) mostly reflect differences in the seasonal cyclone numbers (discussed in the next section).

To more clearly show the differences in genesis latitude between the clusters, normal distributions with the mean and standard deviation of the genesis latitudes of each cluster were plotted (right panel of Fig. 8a). Clusters 1 and 4 have similar genesis latitudes (mean of 32.5° and 33.0°S respectively), but develop on different sides of the jet. The mean genesis latitude of cluster 3, the most poleward cluster, is 41.2°S. Clusters 2 and 4, while showing some similarities in their composite structure and development, have different latitudinal distributions. The mean latitude of cluster 2 genesis is 36.5°S, and of cluster 4 is 33.0°S. These are statistically significantly different at the 95% level (as determined by a t-test).

The genesis density and track density of the cluster cyclones in each season are shown in Supplementary Figures 6–9, and Figure 9 shows the annual average track density for each cluster. The tracks of the cluster cyclones are quite closely associated with the orientation of the jet (and the steering flow) of the composite cyclones. Cluster 1 shows tracks with a strong southeastward movement in all seasons. Cluster 3 cyclones track quite zonally and the track density highlights the genesis around New Zealand. Clusters 2 and 4 show some variability in their track directions, but the track density also highlights the longitudinal variations between these two clusters. These figures demonstrate that although the clustering method uses only a single time in the cyclone lifecycle (genesis), the tracks of the groups do show differences in their propagation direction.

b. Temporal variability of the clusters

As well as having distinct spatial characteristics and genesis locations, the clusters show differences in their temporal variability. Figure 10a shows the number of cyclones identified in each cluster per season as well as the total number per season. The total number of cyclones varies strongly by season, with an average of 1.9, 4, 5.9, and 3.3 cyclones per season for DJF, MAM, JJA, and SON respectively. Each cluster displays differences in its seasonal cycle, however all clusters have a minimum during DJF.

Clusters 2 and 4 show the strongest seasonal variability, with the smallest number of cyclones in the summer season (DJF) and the largest in the winter season (JJA) (Fig. 10a). Cluster 2 has a larger number during MAM than cluster 4. Clusters 2 and 4 both have their cyclogenesis on the poleward side of the jet streak, but fairly close to it, so their variability will be similarly linked to the seasonal shifting of the subtropical jet. The subtropical jet is at its strongest, and most zonal during JJA (see e.g., Risbey et al. 2009).

Cluster 1, the equatorward entrance cluster, has a weak seasonal cycle, with 1.3, 1.2, and 1.2 cyclones per season occurring during MAM, JJA, and SON respectively. During the winter, the cyclogenesis region of this cluster is quite well constrained to the north of 35°S, but during the other seasons, the genesis locations are quite spread throughout the region. This may be related to the weaker northern component of the split jet system during the other seasons.

Cluster 3, which forms in a deep trough, also has a fairly weak seasonal cycle with the highest number per season (just over 1) occurring during MAM, and the lowest (0.6) occurring during DJF. For this cluster the strength and orientation of the subtropical jet is not as important as the distance from the main jet streak. This cluster has its cyclogenesis in highly amplified upper level flow, and along a front associated with a more poleward extratropical cyclone. The seasonal cycle of this cluster is likely associated with the seasonal changes in the jet (which is weakest in DJF), as well as the seasonal cycle of the frequency of cold fronts seen in this region (which is higher in MAM and SON than JJA; Catto et al. 2012) .

With the long record from ERA-Interim of 32 years (compared to the 5 years used in SR00), it is possible to investigate the impact of large-scale climate drivers (ENSO and SAM, which influence the subtropical jet) on the occurrence of the different identified clusters. Figure 10b shows the number of cyclones per season identified during La Niña and El Niño seasons. Cluster 1 shows some differences during MAM, JJA, and SON with more cyclones in this cluster identified during La Niña events than El Niño events. During DJF and MAM , cluster 3 cyclones occur more frequently during El Niño conditions than La Niña. Clusters 2 and 4 show opposing signals in most seasons, with the largest differences occurring during JJA. There are more cluster 2 cyclones during La Niña than El Niño, and more cluster 4 cyclones during El Niño than La Niña. None of the differences are statistically significant at the 95% level, despite some of the differences being quite large, because of the small sample size.

A similar comparison between negative and positive SAM seasons is shown in figure 10. Overall there are not many differences, and considering all seasons together, the total numbers of cyclones identified are very similar, with an average of about 3.5 cyclones per season during negative SAM, and 3.8 cyclones per season for positive SAM. During DJF there are slightly more cluster 1 cyclones during negative SAM (the only result that is statistically significant). During JJA and SON cluster 3 shows fewer cyclones in negative SAM and more in positive SAM. This may be related to the poleward contraction of the storm tracks and more pronounced split jet that can be seen during positive SAM. There is more blocking over New Zealand and associated higher amplitude upper level wave activity, which would influence cluster 3. Risbey et al. (2009) showed an increase in cut off lows in the Australian region during positive SAM and overall wetter Australia during the wet season, however, Rudeva and Simmonds (2015) showed very little difference in cyclone numbers in this region between positive and negative SAM during JJA.

As well as large seasonal variability, there is also large interannual variability in the total number of cyclones identified, and the number in each cluster. Figure 11 shows the total numbers of cyclones identified each year, as well as the numbers identified in each cluster. The maximum total of 23 cyclones occurs in 1986, and the minimum of 7 in 1995. There are no apparent trends over this period. The selection of only the strong developing cyclones means that the sample size each year is fairly small. These results may be different for a larger region, or when considering all cyclones, although Pepler et al. (2017) found no significant trends in Australian east coast lows since 1911 using the 20th Century Reanalysis product.

5. Summary and Discussion

Here an automated method using K-means clustering to classify extratropical cyclones from 1979–2010 in the Australia and New Zealand region has been presented. Fields of upper level (250 hPa) wind speed and zonal anomalies of θ_{PV2} from a radial region of 20° were used as the fields to cluster on. The results of this classification have been compared against the manual classification of 5 years of extratropical cyclones by Sinclair and Revell (2000). Some of the features of the four clusters are as follows.

- Cluster 1 has its genesis in the equatorward entrance region of the upper level jet streak in diffluent flow. It is the cluster with the highest 850 hPa θ_w and the highest average rainfall, associated with advection of warm moist air from the tropics, and a high frequency of warm fronts.

- Cluster 2 occurs in the poleward exit region of the upper level jet streak associated with a jet streak that is downstream of the upper level trough. It has quite low rainfall totals and more cold fronts than warm fronts. This cluster has a large seasonal cycle with a maximum in winter (JJA) and minimum in summer (DJF).

- Cluster 3 occurs beneath a deep upper level trough, far polewards of the main jet streak, and clearly resembles secondary frontal cyclogenesis. This cluster has strong cold advection at the time of genesis and the highest number of identified cold fronts. It exhibits rapid development, but has the lowest rainfall totals.

- Cluster 4 has its genesis in the poleward exit region of the upper level jet streak, similar to Cluster 2, but with the jet streak upstream of the upper level trough. This cluster occurs further north that cluster 2 but has very similar seasonal variability. It has the 2^{nd} highest rainfall totals and shows higher frequency of warm fronts.

The clusters identified here closely resemble the classes from SR00 in terms of their structure at the time of genesis, and also how they develop through their lifecycles. The upshot of this result is two-fold. First, we see that the clustering method is clearly good enough to identify the types of cyclones that an experienced synoptician would identify. Second, it provides further evidence of the success of the manual classification technique of SR00. As well as showing that a simple clustering can reproduce classes of cyclones similar to a manual classification, the results presented here add extra credence to such manual classifications.

A next step, having established the utility of the method, will be to expand the region of study to the hemispheric or global scale. A brief investigation into the robustness of the clusters found in this study when considering a larger region has been performed (using a region expanded by 10° to the south and east). The cluster centroids were similar, but not exactly the same (not shown). When considering larger regions it may be that more clusters are required (e.g., Graf et al. 2016, found 5 distinct classes over the Northern Hemisphere). Since there are no global manual classifications against which to compare, a more objective method of selecting the number of clusters would likely need to be employed. It has been suggested that cyclones with different characteristics occur preferentially in different regions (e.g., Dacre and Gray 2013), and such classification may help to further investigate this suggestion. The resemblance of cluster 3 to secondary cyclogenesis could be used to more systematically analyze the occurrence of this type of cyclogenesis globally. The relative importance of upper level versus lower level forcing, as well as diabatic processes at upper and lower levels could be investigated in order to link these clusters to the three-fold classification scheme of Deveson et al. (2002) and Plant et al. (2003), the objective cyclogenesis classification of Graf et al. (2016), and the recent work of Binder et al. (2016) on the role of warm conveyor belts in the intensification of cyclones. While these avenues of research are beyond the scope of the present study, they would add greatly to knowledge on the processes within extratropical cyclones and how they vary spatially and temporally over the globe.

There are clearly some differences in the precipitation associated with the different classes. The composites and the lifecycles indicate that cluster 1 has the highest rainfall and cluster 3 has the lowest. A statistical analysis of the maximum precipitation along the cyclone tracks in each cluster reveals that there are statistically significant differences between most of the clusters (only clusters 1 and 4 are indistinguishable in this measure). These results suggest that the latitude of the cyclones plays an important role in determining the amount of precipitation produced since cluster 1 occurs at the lowest latitudes. Supplementary figure 10 shows the relationship between the maximum precipitation and latitude. The latitude at which the maximum precipitation occurs is a stronger influence than the genesis latitude of the cyclone, but the correlations are still only between 0.44 and 0.53. This indicates that there are other factors responsible for determining the precipitation of the cyclones— the dynamics of the features themselves. Recently Pfahl and Sprenger (2016) showed that stronger cyclones are associated with higher rainfall, with the strongest relationship with the precipitation from before the maximum cyclone intensity. The cyclones with higher rainfall may, therefore, be more dependent on latent heating for their intensification. Such latent heating and other diabatic processes would need to be diagnosed using multiple datasets (such as remotely sensed cloud data) since these processes in ERA-Interim show large biases (Hawcroft et al. 2017). The clustering developed here may help to investigate these aspects in future research. We will use cyclone classification to better understand both the impact of latent heating on the development of different types of extratropical cyclones, and also the impacts they have in terms of the precipitation they produce.

A particular motivation for developing a simple automated method of extratropical cyclone classification is the need to evaluate state-of-the-art climate models for their ability to represent the full spectrum of extratropical cyclone characteristics (Catto 2016). Climate models need to be able to replicate the clusters observed in nature. Since cyclones involve complex non-linear interactions on a variety of space and time scales, their climatologies and classification provide an exacting means of assessing the realism of climate model results. These systems are responsible for bringing a large proportion of the total and extreme rainfall to many regions of the midlatitudes (Pfahl and Wernli 2012; Catto et al. 2012; Catto and Pfahl 2013; Dowdy and Catto 2017) and so we need to have confidence in projections of their future changes. The method that has been presented here is simple enough to apply to climate model output and will help to understand model shortcomings and how extratropical cyclones and their associated impacts may change in the future.

Acknowledgments.

This work was funded by the Australian Research Council (ARC) through a Discovery Early Career Research Award (DE140101305), and spported by the Centre of Excellence for Climate Systems Science (CE110001028). Thanks go to Julian Quinting and Duncan Ackerley for comments on an earlier version of the manuscript. Thanks also to Matt Hawcroft (University of Exeter) for use of his precipitation data. ERA-Interim data are available online (http:// apps.ecmwf.int/datasets/).

The author acknowledges wit thanks the valuable comments and suggestions from a number of anonymous reviewers.

APPENDIX

In order to investigate the impact of using the 12–24hr forecast lead time precipitation (as per Hawcroft et al. 2016, 2017), the composites (Fig.10) and lifecycles (Fig. 11) of precipitation from the clusters have been produced.

The precipitation from Hawcroft et al. (2017) represents the 6-hourly accumulation before the analysis time, whereas from Catto and Pfahl (2013) it represents the 6-hourly accumulation following the analysis time. The effect of this offset can be seen when comparing the composites of precipitation, where the original precipitation is slightly further ahead of the cyclone, and the Hawcroft precipitation is closer to the cyclone center.

At the times of genesis and maximum growth, the composites of precipitation show lower precipitation volumes around the cyclones with more spread out features. Whereas, at the time of maximum intensity the precipitation volumes are higher. Early in the lifecycles of the cyclones, the forecasts may show more uncertainty in the exact location of the cyclone and associated fronts. This would produce a smearing effect on the precipitation composites. At the time of maximum intensity, the forecast errors at the longer lead times would be smaller and so the features would likely line up more coherently, producing higher volume composites consistent with Hawcroft et al. (2016). The lifecycles show higher precipitation in the 5° region surrounding the composite cyclones, consistent with the more centrally located precipitation when using the Hawcroft precipitation data.

REFERENCES

Ayrault, F., F. Lalaurette, A. Joly, and C. Loo, 1995: North Atlantic ultra high frequency variability. *Tellus*, **47A**, 671–696.

Bengtsson, L., K. I. Hodges, and E. Roeckner, 2006: Storm tracks and climate change. *J. Climate*, **19**, 3518–3543.

Berry, G., M. J. Reeder, and C. Jakob, 2011: A global climatology of atmospheric fronts. *Geophys. Res. Lett.*, **38**, doi:10.1029/2010GL046451, L04809.

Binder, H., M. Boettcher, H. Joos, and H. Wernli, 2016: The role of warm conveyor belts for the intensification of extratropical cyclones in Northern Hemisphere winter. *J. Atmos. Sci.*, **73**, 3997–4020, doi:10.1175/JAS-D-15-0302.1.

Booth, J. F., C. M. Naud, and A. D. Del Genio, 2013: Diagnosing warm frontal cloud formation in a GCM: A novel approach using conditional subsetting. *J. Climate*, **26**, 5827–5845, doi:10.1175/JCLI-D-12-00637.1.

Catto, J. L., 2016: Extratropical cyclone classification and its use in climate studies. *Rev. Geophys.*, **54**, doi:10.1002/2016RG000519.

Catto, J. L., C. Jakob, G. Berry, and N. Nicholls, 2012: Relating global precipitation to atmospheric fronts. *Geophys. Res. Lett.*, **39**, doi:10.1029/2012GL051736, L10805.

Catto, J. L., E. Madonna, H. Joos, I. Rudeva, and I. Simmonds, 2015: Global relationship between fronts and warm conveyor belts and the impact on extreme precipitation. *J. Climate*, **28**, 8411–8429, doi:10.1175/JCLI-D-15-0171.1.

Catto, J. L. and S. Pfahl, 2013: The importance of fronts for extreme precipitation. *J. Geophys. Res.*, **118**, 10,791–10,801, doi:10.1002/jgrd.50852.

Catto, J. L., L. C. Shaffrey, and K. I. Hodges, 2010: Can climate models capture the structure of extratropical cyclones? *J. Climate*, **23**, 1621–1635.

Catto, J. L., L. C. Shaffrey, and K. I. Hodges, 2011: Northern Hemisphere extratropical cyclones in a warming climate in the HiGEM high-resolution climate model. *J. Climate*, **24**, 5336–5352, doi:10.1175/2011JCLI4181.1.

Chang, E. K. M. and S. Song, 2006: The Seasonal Cycles in the Distribution of Precipitation around Cyclones in the Western North Pacific and Atlantic. *J. Atmos. Sci.*, **63**, 815–839.

Coggins, J. H. J., A. J. McDonald, and B. Jolly, 2013: Synoptic climatology of the Ross Ice Shelf and Ross Sea region of Antarctica: k-means clustering and validation. *Int. J. Climatol.*, **34**, 2330–2348, doi:10.1002/joc.3842.

Colle, B. A., Z. Zhang, K. A. Lombardo, E. Chang, P. Liu, and M. Zhang, 2013: Historical evaluation and future prediction of Eastern North American and Western Atlantic extratropical cyclones in the CMIP5 models during the cool season. *J. Climate*, **26**, 6882–6903, doi:10.1175/JCLI-D-12-00498.1.

Dacre, H. F. and S. L. Gray, 2013: Quantifying the climatological relationship between extratropical cyclone intensity and atmospheric precursors. *Geophys. Res. Lett.*, **40**, 2322–2327, doi:10.1002/grl.50105.

Dee, D. P., et al., 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597.

Deveson, A. C. L., K. A. Browning, and T. D. Hewson, 2002: A classification of FASTEX cyclones using a height-attributable quasi-geostrophic vertical-motion diagnostic. *Quart. J. Roy. Meteor. Soc.*, **128**, 93–117.

Dowdy, A. and J. L. Catto, 2017: Extreme weather caused by concurrent cyclone, front and thunderstorm occurrences. *Scientific Reports*, **7**, doi:10.1038/srep40359.

Dowdy, A. J., G. A. Mills, and B. Timbal, 2013: Large-scale diagnostics of extratropical cyclogenesis in eastern Australia. *Int. J. Climatol.*, **33**, 2318–2327, doi:10.1002/joc.3599.

Evans, M. S., D. Keyser, L. F. Bosart, and G. M. Lackmann, 1994: A satellite-derived classification scheme for rapid maritime cyclogenesis. *Mon. Wea. Rev.*, **122**, 1381–1416.

Field, P. R., A. Gettelman, R. Neale, R. Wood, P. J. Rasch, and H. Morrison, 2008: Midlatitude cyclone compositing to constrain climate model behaviour using satellite observations. *J. Climate*, **21**, 5887–5903.

Field, P. R. and R. Wood, 2007: Precipitation and Cloud Structure in Midlatitude Cyclones. *J. Climate*, **20**, 233–254.

Flato, G., et al., 2013: Evaluation of climate models. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley, Eds., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 741–866, doi: 10.1017/CBO9781107415324.020, URL `www.climatechange2013.org`.

Fuenzalida, H. A., R. Sánchez, and R. D. Garreaud, 2005: A climatology of cutoff lows in the Southern Hemisphere. *J. Geophys. Res.*, **110**, doi:10.1029/2005JD005934, d18101.

Gómara, I., J. Pinto, G. Masato, P. Zurita-Gotor, and B. Rodríguez-Fonseca, 2014: Rossby wave-breaking analysis of explosive cyclones in the Euro-Atlantic sector. *Quart. J. Roy. Meteor. Soc.*, **140**, 738–753, doi:10.1002/qj.2190.

Gong, X. and M. B. Richman, 1995: On the application of cluster analysis to growing season precipitation data in North America east of the Rockies. *J. Climate*, **8**, 897–931.

Govekar, P., C. Jakob, and J. L. Catto, 2014: The relationship between clouds and dynamics in Southern Hemisphere extratropical cyclones in the real world and a climate model. *J. Geophys. Res.*, **119**, 6609–6628, doi:10.1002/2013JD020699.

Graf, M. A., H. Wernli, and M. Sprenger, 2016: Objective classification of extratropical cyclogenesis. *Quart. J. Roy. Meteor. Soc.*, doi:10.1002/qj.2989.

Hawcroft, M. K., H. F. Dacre, R. Forbes, K. I. Hodges, L. C. Shaffrey, and T. Stein, 2017: Using satellite and reanalysis data to evaluate the representation of latent heating in extratropical cyclones in a climate model. *Clim. Dynam.*, **48**, 2255–2278, doi:10.1007/s00382-016-3204-6.

Hawcroft, M. K., L. C. Shaffrey, K. I. Hodges, and H. F. Dacre, 2016: Can climate models represent the precipitation associated with extratropical cyclones? *Clim. Dynam.*, **47**, 679–695, doi:10.1007/s00382-015-2863-z.

Hewson, T. D., 1998: Objective fronts. *Meteorol. Appl.*, **5**, 37–65.

Hodges, K. I., 1994: A General Method for Tracking Analysis and Its Application to Meteorological Data. *Mon. Wea. Rev.*, **122**, 2573–2586.

Hodges, K. I., 1995: Feature tracking on the unit sphere. *Mon. Wea. Rev.*, **123**, 3458–3465.

Hodges, K. I., 1999: Adaptive Constraints for Feature Tracking. *Mon. Wea. Rev.*, **127**, 1362–1373.

Holland, G. J., A. H. Lynch, and L. M. Leslie, 1987: Australian east-coast cyclones. Part I: Synoptic overview and case study. *Mon. Wea. Rev.*, **115**, 3024–3036.

Hope, P., et al., 2014: A Comparison of automated methods of front recognition for climate studies - a case study in south west Western Australia. *Mon. Wea. Rev.*, **142**, 343363.

Hoskins, B. J. and K. I. Hodges, 2002: New perspectives on the Northern Hemisphere winter storm tracks. *J. Atmos. Sci.*, **59**, 1041–1061.

Hoskins, B. J. and K. I. Hodges, 2005: A new perspective on Southern Hemisphere storm tracks. *J. Climate*, **18**, 41084129, doi:10.1175/jcli3570.1.

Huang, B., et al., 2015: Extended reconstructed sea surface temperature version 4 (ersst.v4). part i: Upgrades and intercomparisons. *Journal of Climate*, **28 (3)**, 911–930, doi:10.1175/JCLI-D-14-00006.1, URL http://dx.doi.org/10.1175/JCLI-D-14-00006.1, http://dx.doi.org/10.1175/JCLI-D-14-00006.1.

Kirtman, B., et al., 2013: Near-term climate change: Projections and predictability. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley, Eds., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, chap. 11, 953–1028, doi:10.1017/CBO9781107415324.023, URL www.climatechange2013.org.

Marshall, G. J., 2003: Trends in the Southern Annual Mode from observations and reanalyses. *J. Climate*, **16**, 4134–4143.

Martius, O., C. Schwierz, and M. Sprenger, 2008: Dynamical tropopause variability and potential vorticity streamers in the Northern Hemisphere – a climatological analysis. *Adv. Atmos. Sci.*, **25**, 367–379.

Naud, C. M., A. D. Del Genio, M. Bauer, and W. Kovari, 2010: Cloud vertical distribution across warm fronts in CloudSat-CALIPSO data and a general circulation model. *J. Climate*, **23**, 3397–3415, doi:10.1175/2010JCLI3282.1.

Parker, D. J., 1998: Secondary frontal waves in the North Atlantic region: A dynamical perspective of current ideas. *Quart. J. Roy. Meteor. Soc.*, **124**, 829–856.

Pepler, A. S., J. Fong, and L. V. Alexander, 2017: Australian east coast mid-latitude cyclones in the 20th Century Reanalysis ensemble. *Int. J. Climatol.*, **37**, 2187–2192, doi:10.1002/joc.4812.

Petterssen, S., 1936: Contribution to the theory of frontogenesis. *Geofys. Publ.*, **11**, 1–27.

Petterssen, S. and S. J. Smebye, 1971: On the development of extratropical cyclones. *Quart. J. Roy. Meteor. Soc.*, **97**, 457–482.

Pfahl, S. and M. Sprenger, 2016: On the relationship between extratropical cyclone precipitation and intensity. *Geophysical Research Letters*, **43 (4)**, 1752–1758, doi:10.1002/2016GL068018, URL `http://dx.doi.org/10.1002/2016GL068018`, 2016GL068018.

Pfahl, S. and H. Wernli, 2012: Quantifying the relevance of cyclones for precipitation extremes. *J. Climate*, **25**, 6770–6780.

Plant, R. S., G. C. Craig, and S. L. Gray, 2003: On a threefold classification of extratropical cyclogenesis. *Quart. J. Roy. Meteor. Soc.*, **129**, 2989–3012.

Pope, M., C. Jakob, and M. J. Reeder, 2009a: Objective classification of tropical mesoscale convective systems. *J. Climate*, **22**, 5797–5808, doi:10.1175/2009JCLI2777.1.

Pope, M., C. Jakob, and M. J. Reeder, 2009b: Regimes of the North Australian wet season. *J. Climate*, **22**, 6699–6715, doi:10.1175/2009JCLI3057.1.

Raut, B. A., C. Jakob, and M. J. Reeder, 2014: Rainfall changes over Southwestern Australia and their relationship to the southern annular mode and ENSO. *J. Climate*, **27**, 5801–5814, doi:10.1175/JCLI-D-13-00773.1.

Risbey, J. S., M. J. Pook, P. C. McIntosh, C. C. Ummenhofer, and G. Meyers, 2009: Characteristics and variability of synoptic features associated with cool season rainfall in southeastern Australia . *Int. J. Climatol.*, **29**, 1595–1613, doi:10.1002/joc.1775.

Rossow, W. B., G. Tselioudis, A. Polak, and C. Jakob, 2005: Tropical climate described as a distribution of weather states indicated by distinct mesoscale cloud property mixtures. *Geophys. Res. Lett.*, **32**, doi:10.1029/2005GL024584.

Rudeva, I. and S. K. Gulev, 2011: Composite analysis of North Atlantic extratropical cyclones in NCEP-NCAR reanalysis data. *Mon. Wea. Rev.*, **139**, 1419–1446.

Rudeva, I. and I. Simmonds, 2015: Variability and trends of global atmospheric frontal activity and links with large-scale modes of variability. *J. Climate*, **28**, 3311–3330, doi:10.1175/JCLI-D-14-00458.1.

Seiler, C. and F. W. Zwiers, 2015: How well do CMIP5 climate models reproduce explosive cyclones in the extra tropics of the Northern Hemisphere. *Clim. Dynam.*, **46**, 1241–1256, doi:10.1007/s00382-015-2642-x.

Shaffrey, L. C., et al., 2009: U.K.-HiGEM: The new U.K. High resolution Global Environment Model - Model description and basic evaluation. *J. Climate*, **22**, 1861–1896.

Simmonds, I. and K. Keay, 2000a: Mean Southern Hemisphere extratropical cyclone behavior in the 40-year NCEP-NCAR Reanalysis. *J. Climate*, **13**, 873–885.

Simmonds, I. and K. Keay, 2000b: Variability of Southern Hemisphere extratropical cyclone behavior, 1958–97. *J. Climate*, **13**, 550–561.

Sinclair, M. R., 1997: Objective identification of cyclones and their circulation intensity and climatology. *Wea. Forecasting*, **12**, 595–612.

Sinclair, M. R. and M. J. Revell, 2000: Classification and composite diagnosis of extratropical cyclogenesis events in the Southwest Pacific. *Mon. Wea. Rev.*, **128**, 1089–1105.

Speer, M. S., P. Wiles, and A. Pepler, 2009: Low pressure systems off the New South Wales coast and associated hazardous weather: Establishment of a database. *Aust. Meteor. Ocean J.*, **58**, 29–39.

Studholme, J., K. I. Hodges, and C. M. Brierley, 2015: Objective determination of the extratropical transition of tropical cyclones in the Northern Hemisphere. *Tellus A*, **67**, doi:10.3402/tellusa.v67.24474, 24474.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, doi:10.1175/BAMS-D-11-00094.1.

Ulbrich, U., J. G. Pinto, H. Kupfer, G. C. Leckebusch, T. Spangehl, and M. Reyers, 2008: Changing Northern Hemisphere Storm Tracks in an Ensemble of IPCC Climate Change Simulations. *J. Climate*, **21**, 1669–1679.

van Loon, H., 1967: The half-yearly oscillations in middle and high southern latitudes and the coreless winter. *J. Atmos. Sci.*, **24**, 472–486.

Wernli, H. and M. Sprenger, 2007: Identification and ERA-15 Climatology of Potential Vorticity Streamers and Cutoffs near the Extratropical Tropopause. *J. Atmos. Sci.*, **64**, 1569–1586.

Willison, J., W. A. Robinson, and G. M. Lackmann, 2013: The importance of resolving mesoscale latent heating in the North Atlantic storm track. *J. Atmos. Sci.*, **70**, 2234–2250, doi:10.1175/JAS-D-12-0226.1.

Young, M. V., 1993: Cyclogenesis: Interpretation of satellite and radar images for the forecaster. Tech. Rep. 73, Forecasting Research Division, United Kingdom Meteorological Office.

Zappa, G., L. C. Shaffrey, and K. I. Hodges, 2013: The ability of CMIP5 models to simulate North Atlantic extratropical cyclones . *J. Climate*, **26**, 5379–5396, doi:10.1175/JCLI-D-12-00501.1.

Zillman, J. W. and P. G. Price, 1972: On the thermal structure of mature Southern Ocean cyclones. *Aust. Meteor. Mag.*, **20**, 34–48.

List of Tables

Table 1. Number of cyclones after each step of the selection process.

Stage of process	DJF	MAM	JJA	SON
Total number	12198	21960	14741	22028
Region selection	310	573	653	646
Developing cyclones	71	188	205	156
Strong cyclones	59	128	190	106

List of Figures

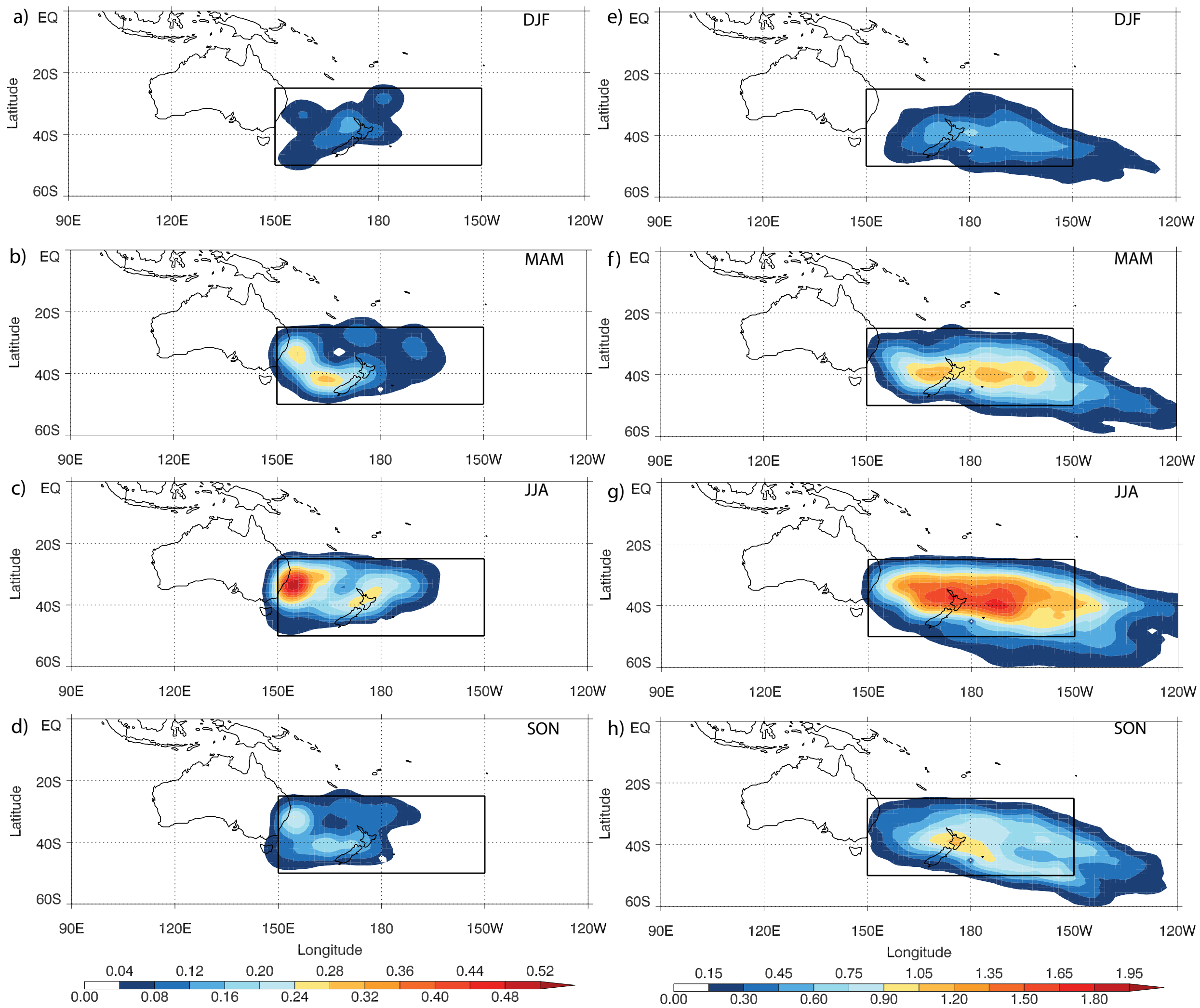

FIG. 1. Genesis density (a-d) and track density (e-h) for the cyclones identified and used in this study for DJF (a,e), MAM (b,f), JJA (c,g), SON (d,h). Units are cyclones per month per 5 degree spherical cap. Note the different color scales for genesis and track density. The black box shows the region of the study.

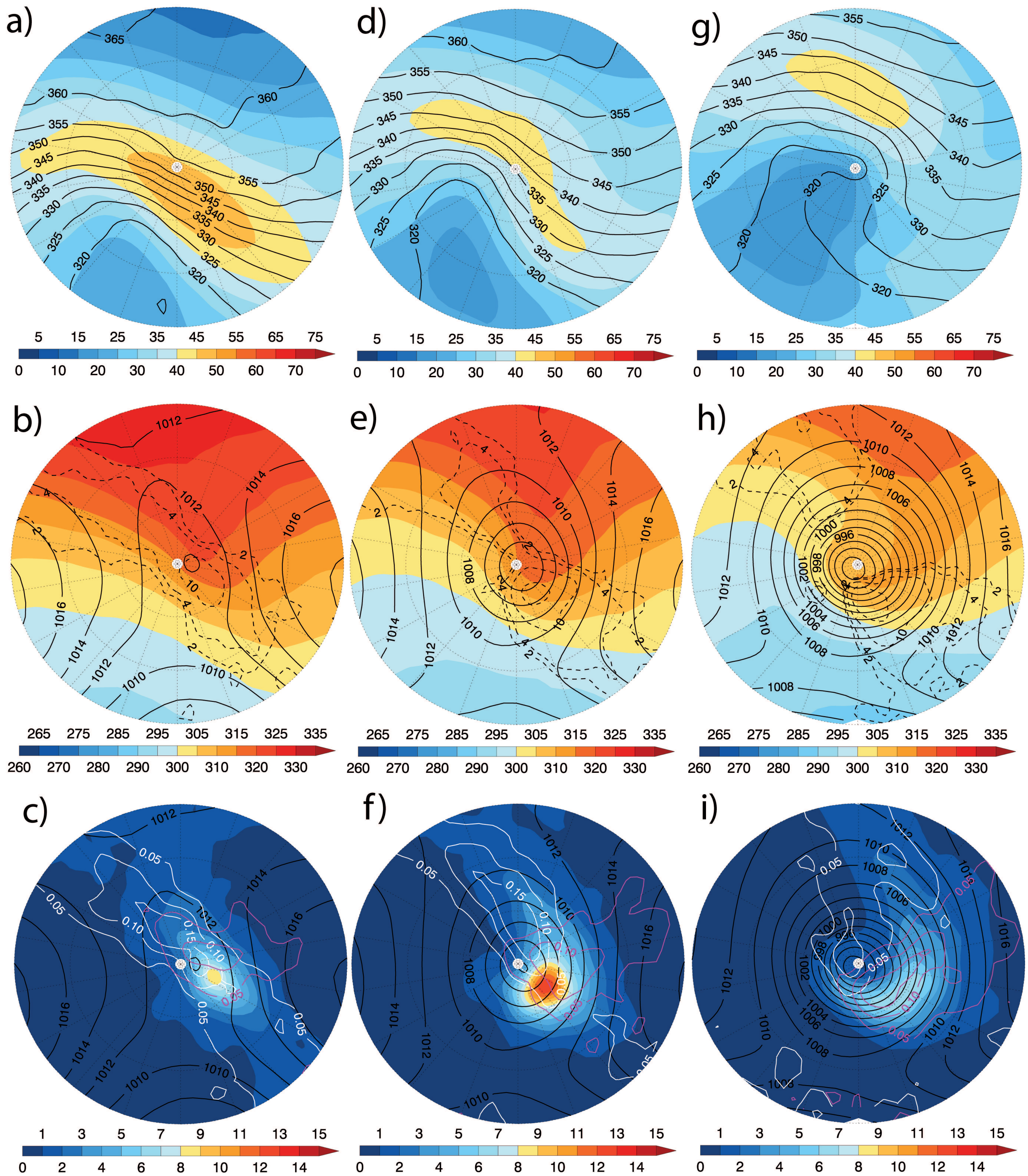

Fig. 2. Composites of cyclones classified as Cluster 1 at the times of genesis (a,b,c), maximum growth (d,e,f), and maximum T42 vorticity (g,h,i). Top row: composites of 250 hPa wind speed (colors; ms^{-1}), θ_{PV2} (black contours; K). Middle row: composites of $\theta_{e,850}$ (colors; K), MSLP (black solid contours; hPa), and frontogenesis (dashed black contours; K/10^3km/day). Bottom row: composites of 6 h accumulated precipitation (colors; mm/6h), MSLP (black contours; hPa), cold front frequency (white contours), and warm front frequency (pink contours).

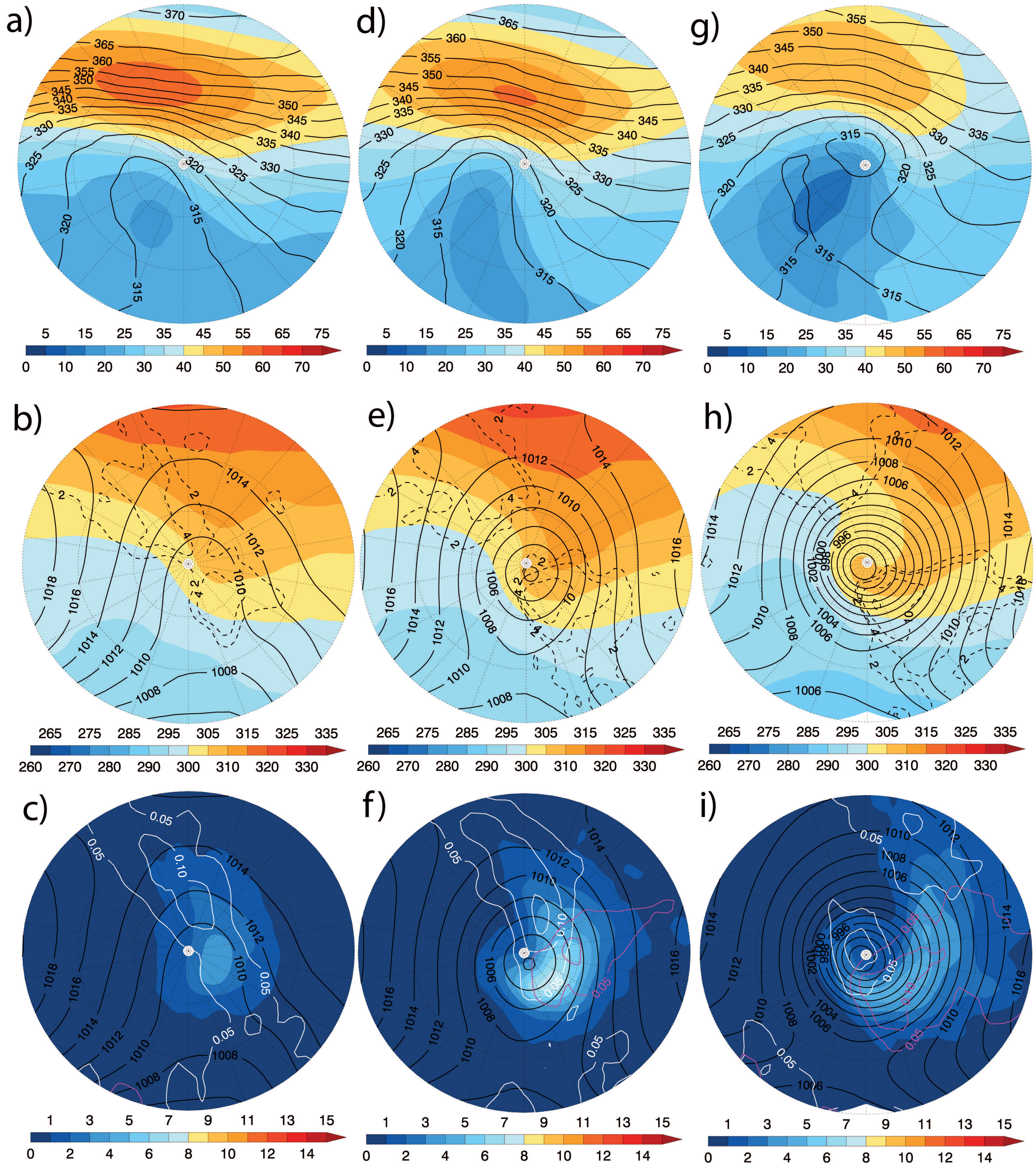

FIG. 3. As Figure 2 but for Cluster 2.

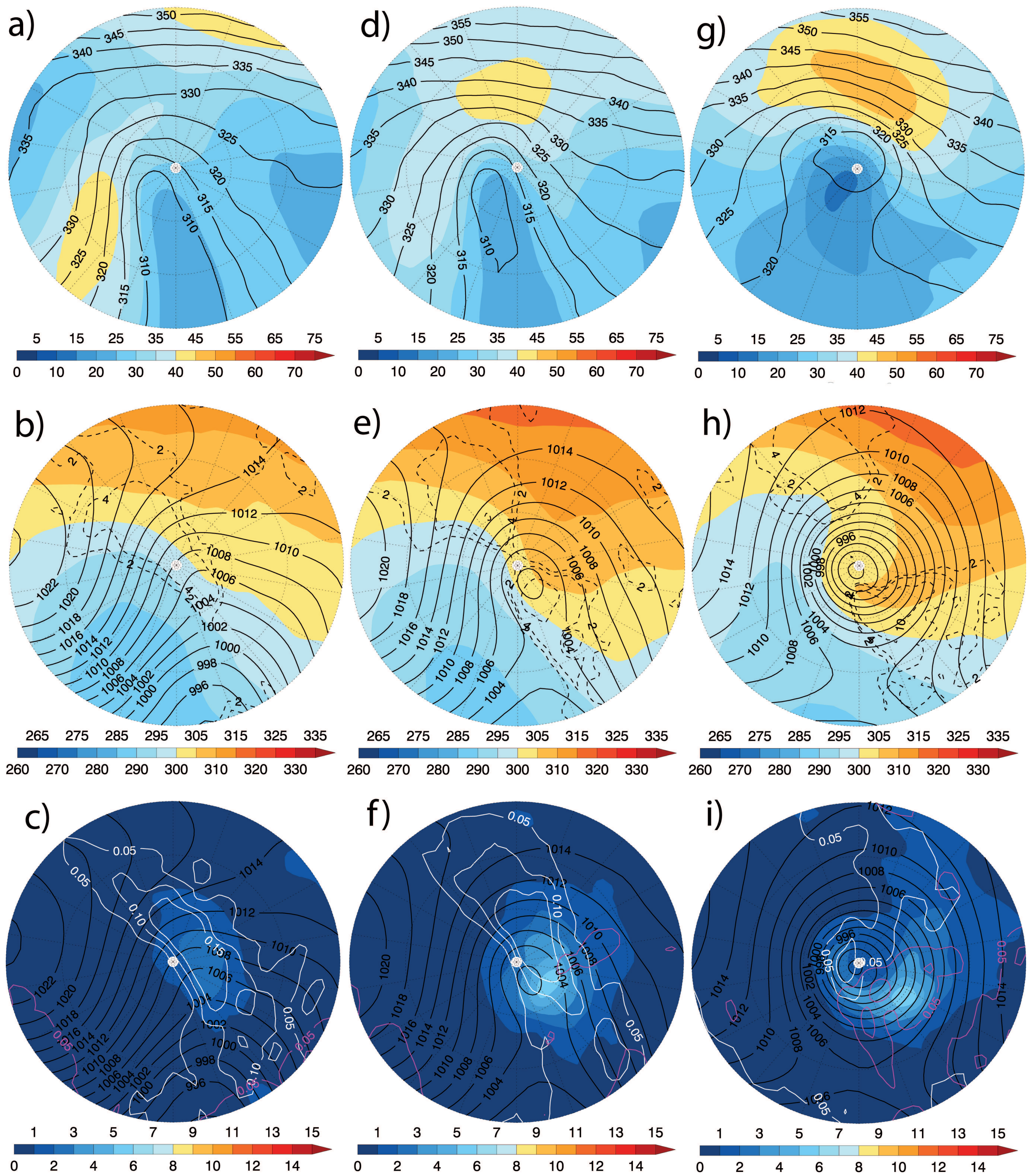

FIG. 4. As Figure 2 but for Cluster 3.

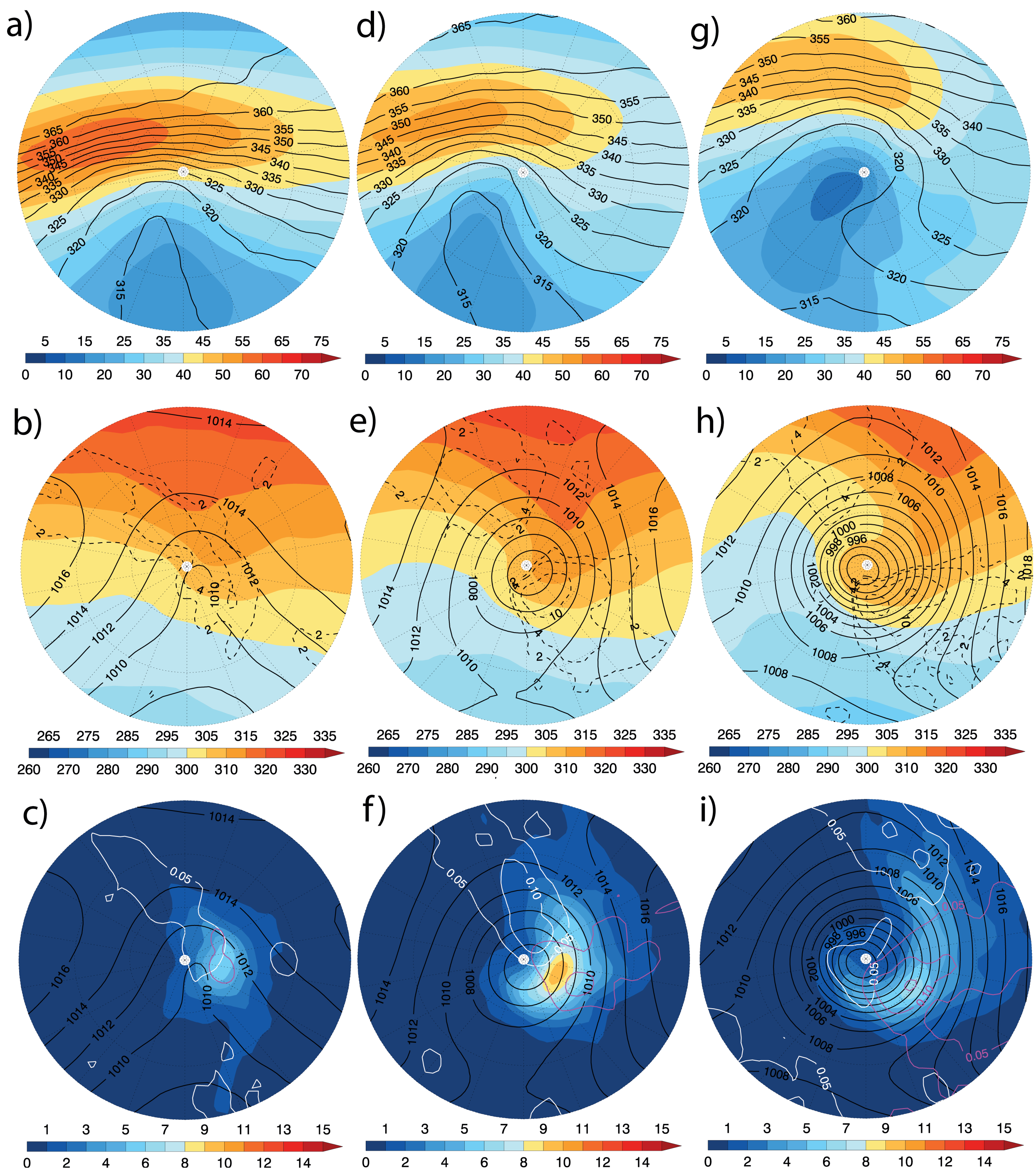

FIG. 5. As Figure 2 but for Cluster 4.

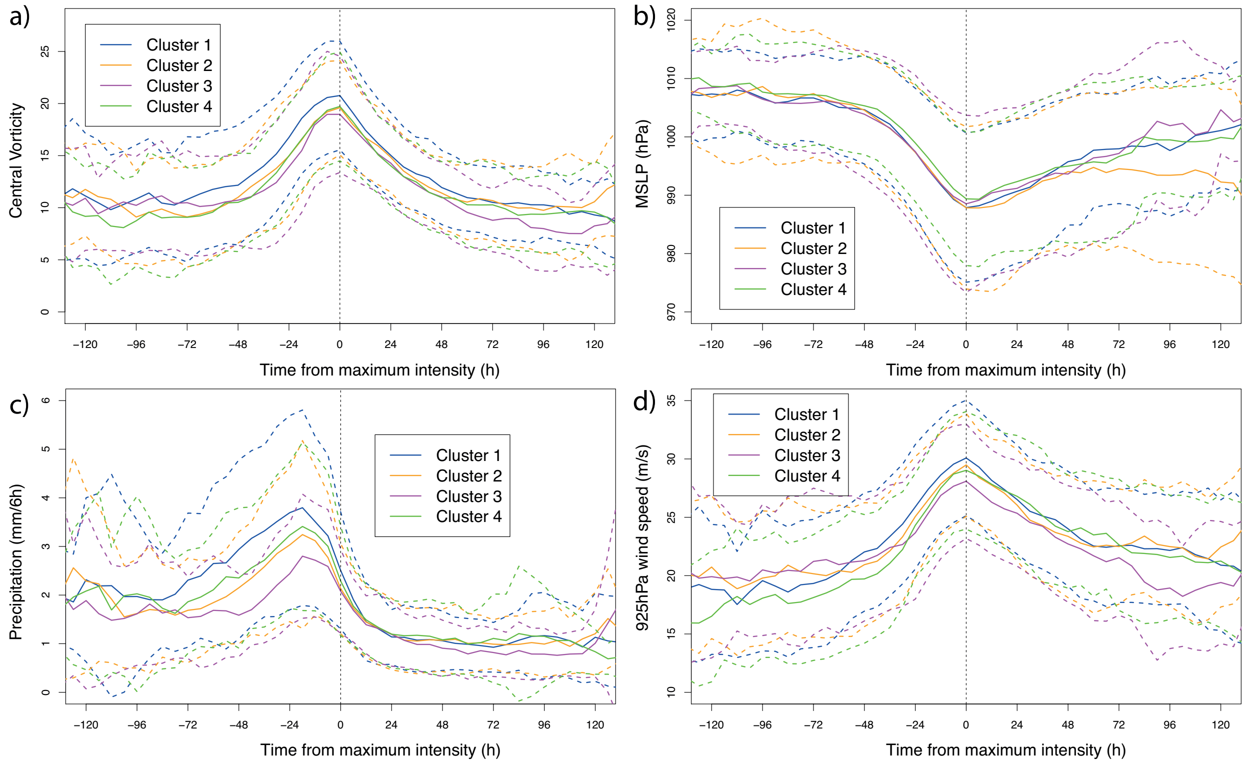

FIG. 6. Composite lifecycles for the 4 clusters centered on the time of maximum T42 vorticity. Solid colored lines show the cluster mean, and the dashed colored lines show plus and minus 1 standard deviation. a) Maximum vorticity (CVU), b) MSLP (hPa), c) precipitation (mm/6h), d) 925 hPa wind speed (ms^{-1}). For the vorticity and MSLP, the maximum (minimum) value of the full resolution vorticity is found within a 5 degree radius of the cyclone center (defined as the location of the maximum T42 vorticity). For the wind speed, the maximum value is found within a 10 degree radius of the cyclone center. For the precipitation, the value is the average precipitation within a 5 degree radius of the cyclone center.

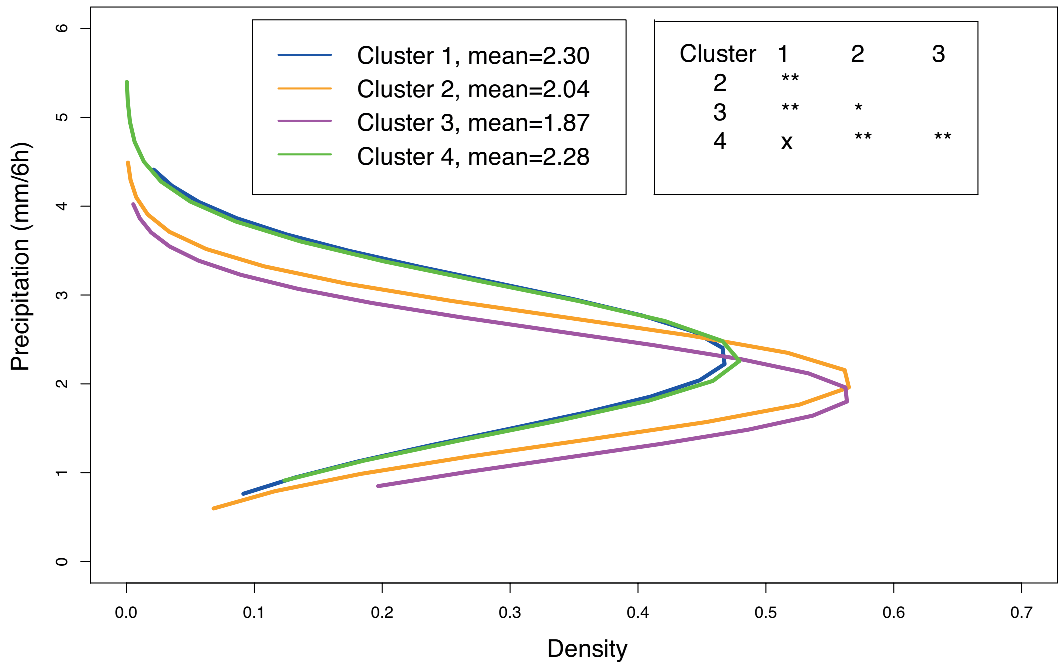

FIG. 7. Maximum 5 degree radius average precipitation for the cyclones in each cluster. (a) Density functions for the clusters with the mean values given in the legend. (b) Scatter plot of maximum precipitation against the latitude at which the maximum precipitation occurs. Colored lines show the linear regressions, and the slope and R squared values are given in the legend. The table shows the significance of the differences between each cluster: ** is significant at the 95% level, * is significant at the 90% level, and x is not significant.

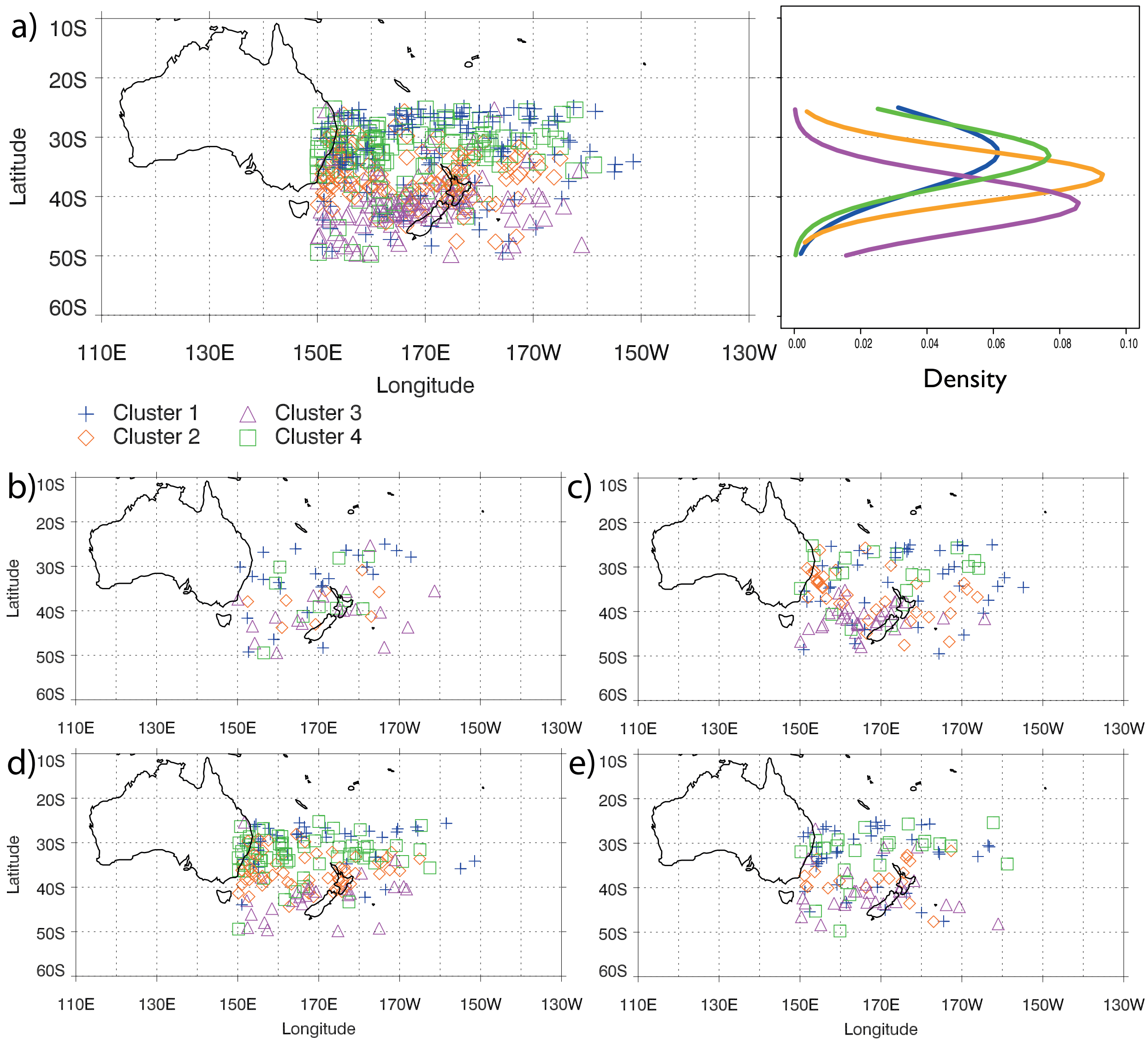

FIG. 8. Genesis locations for all cyclones included in the study sorted by cluster (see legend). (a) Annual, (b) DJF, (c) MAM, (d) JJA, (e) SON. The second panel in (a) shows the density function of genesis latitude for the different clusters.

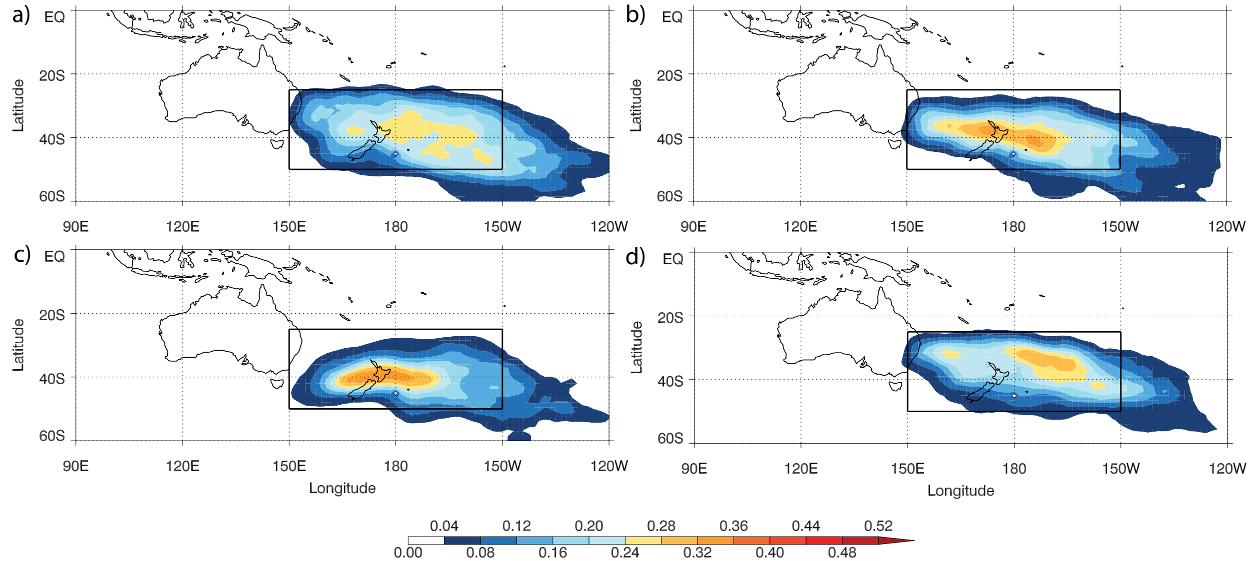

FIG. 9. Track density (annual) for the cyclones identified for (a) cluster 1, (b) cluster 2, (c) cluster 3, and (d) cluster 4. Units are cyclones per month per 5° radius circle. The black box shows the region of the study.

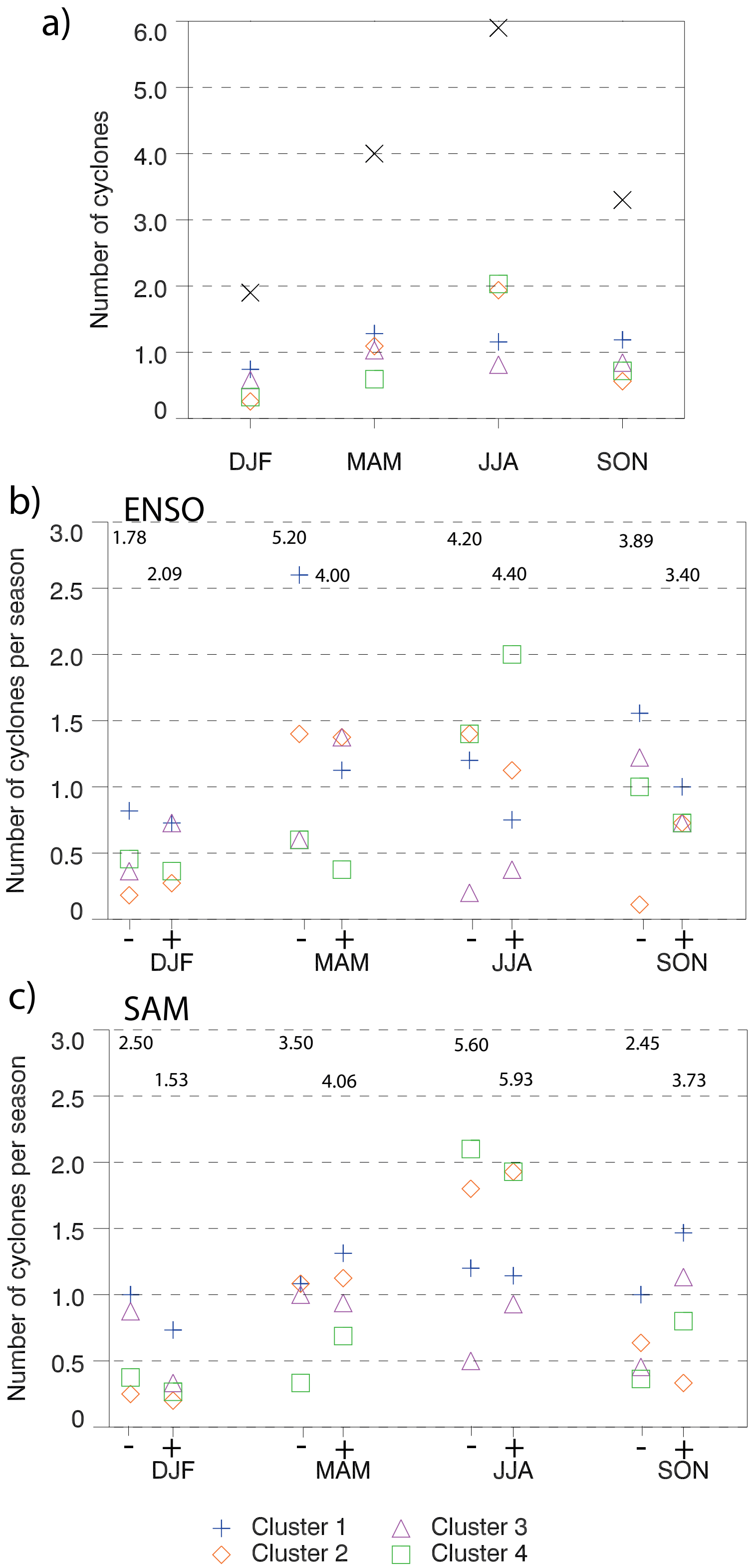

Fig. 10. Seasonal cycle of cyclone occurrences sorted by clusters. (a) All seasons, (b) seasons sorted by ENSO index (see Methods section) with negative ENSO index (La Niña) on the left, and positive ENSO index (El Niño) on the right, (c) seasons sorted by SAM index with negative SAM on the left and positive SAM on the right. Numbers are given as cyclones per season. The total number of cyclones per season is given at the top of each column of data.

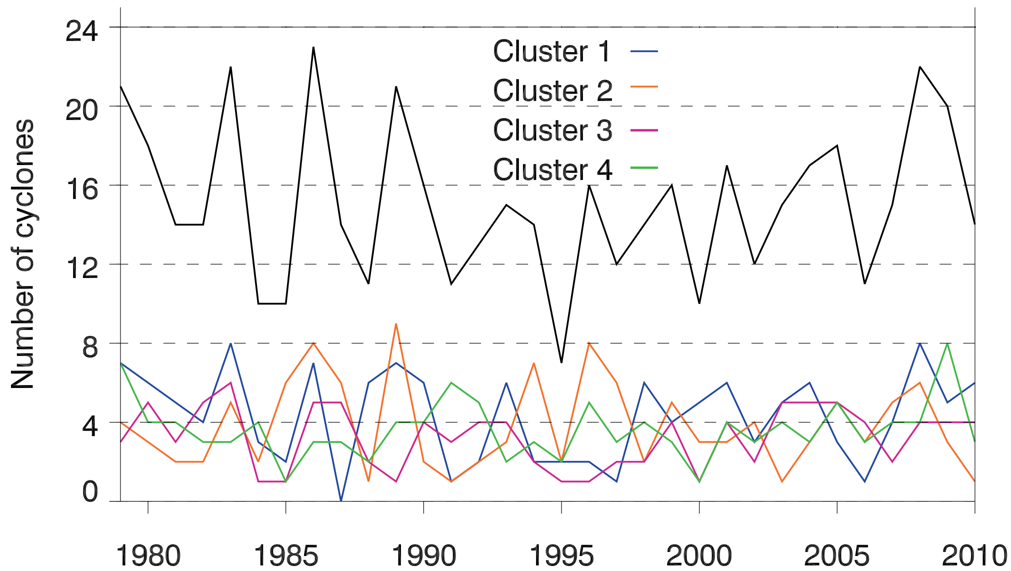

FIG. 11. Number of cyclones per year for all clusters (black line) and each cluster separately (see legend).

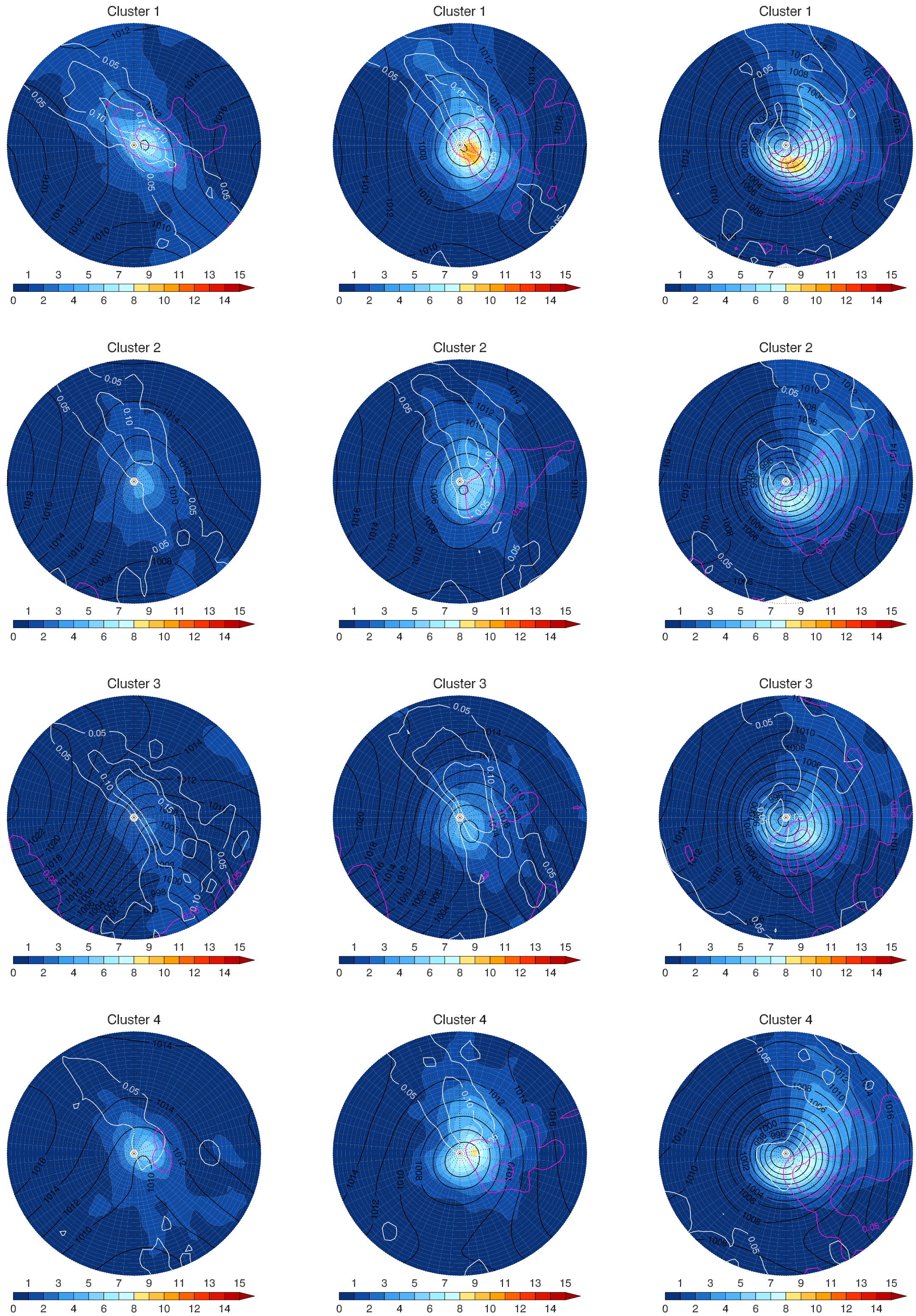

FIG. 12. Composites of different clusters (1-4 from top to bottom) at genesis (left), maximum tendency (middle), and maximum intensity (right). Colors show precipitation (mm/6h) using longer lead time forecasts from ERA-Interim (Hawcroft et al. 2016). Pink contours show frequency of warm fronts, and white contours show frequency of cold fronts. Black contours are MSLP.

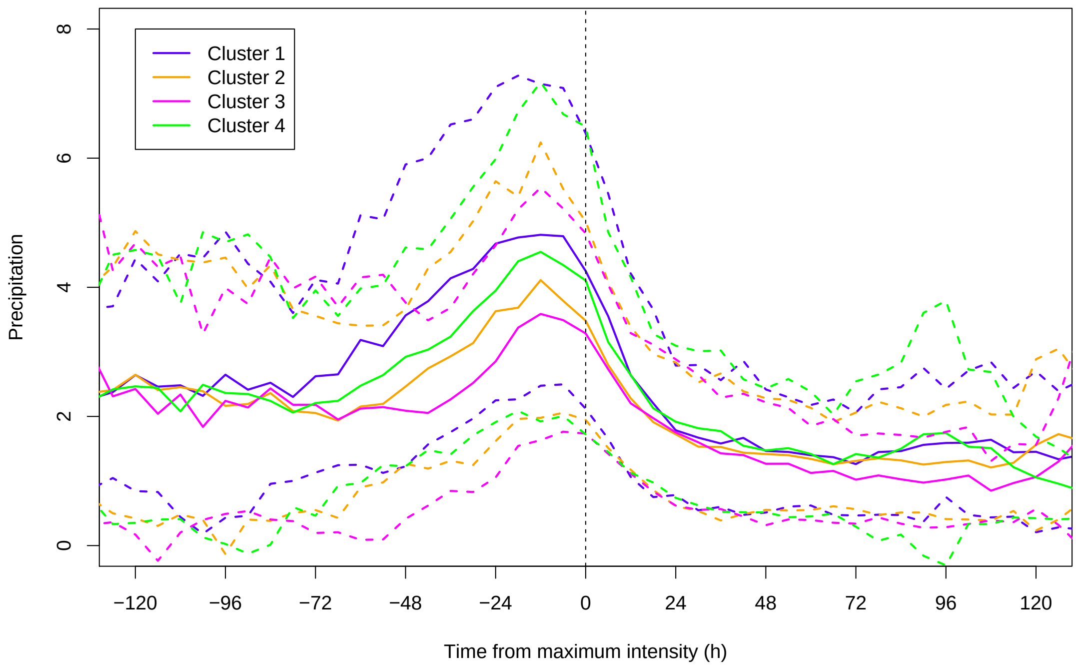

FIG. 13. Lifecycles of precipitation (mm/6h) using longer lead time forecasts from ERA-Interim (Hawcroft et al. 2016). Lifecycles are centered on the time of maximum vorticity. Dashed lines show the range of plus and minus 1 standard deviation.