An Exploratory Framework for Cyclone Identification and Tracking

A THESIS

SUBMITTED FOR THE DEGREE OF

Master of Science (Engineering)

IN THE

Faculty of Engineering

BY

Akash Anil Valsangkar

Computer Science and Automation
Indian Institute of Science
Bangalore – 560 012 (INDIA)

January, 2018

Declaration of Originality

I, **Akash Anil Valsangkar**, with SR No. **04-04-00-10-21-15-1-12597** hereby declare that the material presented in the thesis titled

An Exploratory Framework for Cyclone Identification and Tracking

represents original work carried out by me in the **Department of Computer Science and Automation** at **Indian Institute of Science** during the years **2015-2017**.
With my signature, I certify that:

- I have not manipulated any of the data or results.
- I have not committed any plagiarism of intellectual property. I have clearly indicated and referenced the contributions of others.
- I have explicitly acknowledged all collaborative research and discusions.
- I have understood that any false claim will result in severe disciplinary action.
- I have understood that the work may be screened for any form of academic misconduct.

Date: Student Signature

In my capacity as supervisor of the above-mentioned work, I certify that the above statements are true to the best of my knowledge, and I have carried out due diligence to ensure the originality of the report.

Advisor Name: Advisor Signature

DEDICATED TO

Aai, Baba and Kja (sis)

Acknowledgements

I would like to express my deep gratitude towards Prof. Vijay Natarajan for his help in my work. My heartfelt appreciation for his deep insights and the patient encouragement in the work. I am thankful to our collaborator, Joy Mervin Monterio for his constant help in domain specific knowledge. I am also thankful to Vidya Narayanan and Prof. Ingrid Hotz for helping us in developing the idea. I am grateful to the ex-chairman, Prof. Jayant Haritsa and the current chairman Prof. Shalabh Bhatnagar for the excellent research environment at CSA. I am thankful to all the faculty members for providing excellent courses in the department.

I would also like to express gratitude towards my parents, friends and the almighty for constant encouragement. I am grateful to all my labmates for all the exciting activities and awesome trips.

Lastly, my research was supported by funding from Department of Science and Technology, India (DST/SJF/ETA-02/2015-16) and from the Joint Advanced Technology Programme, Indian Institute of Science (JATP/RG/PROJ/2015/16). I am thankful to these organizations for their support.

Abstract

Analyzing depressions plays an important role in meteorology, especially in the study of cyclones. In particular, the study of the temporal evolution of cyclones requires a robust depression tracking framework. To cope with this demand we propose a pipeline for the exploration of cyclones and their temporal evolution. This entails a generic framework for their identification and tracking. The fact that depressions and cyclones are not well-defined objects and their shape and size characteristics change over time makes this task especially challenging. Our method combines the robustness of topological approaches and the detailed tracking information from optical flow analysis. At first cyclones are identified within each time step based on well-established topological concepts. Then candidate tracks are computed from an optical flow field. These tracks are clustered within a moving time window to distill dominant coherent cyclone movements, which are then forwarded to a final tracking step. In contrast to previous methods our method requires only a few intuitive parameters. An integration into an exploratory framework helps in the study of cyclone movement by identifying smooth, representative tracks. Multiple case studies demonstrate the effectiveness of the method in tracking cyclones, both in the northern and southern hemisphere.

Publications based on this Thesis

1. Akash Anil Valsangkar, Joy Merwin Monteiro, Vidya Narayanan, Ingrid Hotz, and Vijay Natarajan, An Exploratory Framework for Cyclone Identification and Tracking *TVCG: IEEE Transactions on Visualization and Computer Graphics 2017* under second round review.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Cyclones are large-scale, nonlinear, coherent, and long-lived structures that exist in planetary atmospheres. They are characterized by strong vertical winds and a strong nonlinear momentum balance in the horizontal winds. A cyclone derives energy when water vapor condenses within it releasing the latent heat of condensation. Extratropical cyclones play an important role in the general circulation of the earth's atmosphere by transporting energy, angular momentum, and water vapor towards the poles from the subtropics [21].

Extratropical cyclones are typically generated in the winter in subtropical baroclinic zones (or "storm tracks") and move towards the poles [11]. Quite often, these cyclones make landfall and are commonly known as "winter storms" in the mid-latitudes. Due to their economic and social impact, prediction and tracking of extratropical cyclones remains an important aspect of mid-latitude meteorology. With the projected changes in storm tracks due to climate change [26], understanding the processes governing the genesis and life-cycle of extratropical cyclones is a priority. The first step towards understanding these processes is to be able to identify and track individual cyclones across time and space, and this forms an important part of cyclone visualization. In this thesis we propose a cyclone exploration framework based on a novel extraction and tracking algorithm.

1.2 Related work

Cyclone tracking is a well studied problem in meteorology, with many available algorithms for objective (i.e, automated) identification of cyclones from observational data [17, 12, 27, 13, 10] (See [18] for a comprehensive list). Traditional methods used for cyclone tracking are usually dependent on many parameters requiring specific domain knowledge. Parameters include spatial extent of the cyclone, strength of the cyclone and gradient based thresholds. The diversity of algorithms available reflects the diversity of requirements of users and the diversity of operational definitions of a cyclone used in the literature [18]. While a variety of methods exist for cyclone centre identification, cyclone tracking in most algorithms is performed by some variant of a "nearest neighbour" heuristic. The method begins with a first guess for the future location of a cyclone. A cyclone centre at the next time step is considered to be part of a cyclone if its distance, in a chosen metric, from the first guess location is lower than a threshold value. A notable exception is the method proposed by Inatsu [13], which tracks connected regions rather than single points, and considers two connected regions from different time steps to belong to the same cyclone if the regions overlap. A method to track multicentre cyclones was introduced by Hanley and Caballero [10]. It uses conventional identification and tracking techniques but includes two additional steps. First, cyclone centres are merged to form cyclone systems and second, a 'track surgery' is performed to extract significant tracks. It is capable of identifying a multicentred cyclone as a single cyclone. A multicentered cyclone may contain splits and mergers during its life cycle but is represented as a single track. More closely related to our work is the work by Doraiswamy et al. [5] for tracking clouds. They compute and display global and local tracks of clouds and cloud systems. However, their approach does not scale well for cyclones because it results in a large number of tracks with limited lifetime, resulting in a cluttered visualization.

Most of these methods work well within their restricted setting. However they also leave many challenging problems unanswered. Cyclones are typically defined via extremal values of sea level pressure or relative vorticity. Each extremum of sea level pressure or relative vorticity is considered as a depression. Cyclones can contain multiple depressions, each having it's own fuzzy border or structure. Their shape and size changes strongly over time. The lifespan of cyclones can vary from a day to multiple weeks. This makes them extremely difficult to identify and track. Naïve Gaussian smoothing can ease these effects. However, it might also remove small but still important cyclones. Further, the interaction and exploration possibilities are so far largely limited.

Several approaches to tracking features have been proposed within the visualization literature. Most relevant here are the isocontour and topology-based approaches, where the regions enclosed by isosurface or isocontour (in general, level set) components correspond to the features of interest [16]. The level set is computed either for the input scalar field or for a scalar field derived from the input data. The Reeb graph and its loop-free version, the contour tree, are topological structures that provide an abstract representation of the connectivity of level sets for the scalar field restricted to a single time step [6]. The evolution of the level set connectivity over time can be computed efficiently and stored in a time-varying Reeb graph [7] or a time-varying contour tree [29]. Most methods that are based on a level set based approach to feature identification and tracking build a directed acyclic graph (DAG) representing all feature tracks. Correspondence between features in consecutive time steps is identified either via spatial overlap [29], critical point tracking [28], or via an analysis of join and split trees [3, 19]. The concept of combinatorial feature flow fields [23] provides a purely topological tracking method for two-dimensional scalar fields. It also introduces a spatio-temporal importance measure to the feature tracks. Our method also constructs a DAG to represent all correspondences between the cyclone centres. However, different from the above methods, we compute feature tracks via a spatio-temporal analysis. A novel track clustering step helps identify unique and significant cyclone tracks.

Noise in the data results in a large number of features being reported. The above feature tracking methods use topological simplification to reduce the number of critical points and hence remove the insignificant features [3, 5]. In contrast to the above methods, our spatio-temporal approach ensures that the temporal component plays an important role in determining the significance of a cyclone. Independent noise removal within each time step may result in the removal of cyclone centres in the early and late stages of a cyclone's lifetime when the spatial component is not significant.

1.3 Contributions

We propose an exploratory framework using an extraction and tracking algorithm. In contrast to conventional methods it is not a two-step approach that identifies and tracks cyclone centres separately. Instead, it utilises the temporal correlation between time steps to identify cyclones. Our method first combines, topological feature extraction with optical flow tracking to generate short candidate tracks. In a second step, the short candidate tracks are clustered and assembled to build representative cyclone tracks. Using topological concepts like the join and split trees results in a stable identification of the initial candidates. The method can robustly deal with multicentre cyclones without cluttering the visualization, see Figure 2.6(c). Furthermore, it supports the identification of merge and split events. Our integrated approach is stable with respect to spatial and temporal noise resulting in smooth representative cyclone tracks. A query-driven visualization framework with a few easy to understand parameters enables a flexible means of visualizing features of interest. It also enables climate scientists to explore the sensitivity of the obtained results to both parameter choices and the scalar field being used for identification of cyclones. This motivates the formulation and implementation of the framework described in this thesis.

The main contributions of the thesis are:

1. Robust identification of cyclonic regions using a spatio-temporal approach combining topology, optical flow, and clustering.

2. Computing representative tracks per cyclone resulting in a clutter free visualization.

3. Supporting query based user interaction enabled by a cyclone motion graph and a cyclone track graph.

4. A generic framework for tracking extrema of mean sea level pressure and relative vorticity based on a few easy-to-understand parameters.

Case studies from IMILAST [18], a recent intercomparison study, show that the proposed exploratory framework computes cyclone tracks that are comparable to results from existing methods while being generic and flexible enough to allow the use of either scalar field (mean sea level pressure or relative vorticity). Statistics on the spatial and temporal variability of cyclones are compared with previously known results to validate the method.

Chapter 2

Cyclone Identification and Tracking

A cyclone centre is often defined as a local minimum of mean sea level pressure or an extremum of cyclonic vorticity [18]. In this chapter, we describe methods for identifying and tracking cyclones that are based on a novel combination of techniques such as topological analysis and clustering applied to time-varying data. The method consists of a pipeline of four different steps: (i) the extraction of cyclone centres, (ii) the computation of the raw *cyclone motion graph* based on the optical flow, (iii) the distillation of the final *cyclone track graph* summarising the coherent movements of dominant cyclonic regions, and (iv) the visualization and computation of *representative tracks*. Figure 2.1 presents an overview of the algorithm and summarizes these steps. After introducing the cyclone data the individual steps will be discussed in more detail in the following.

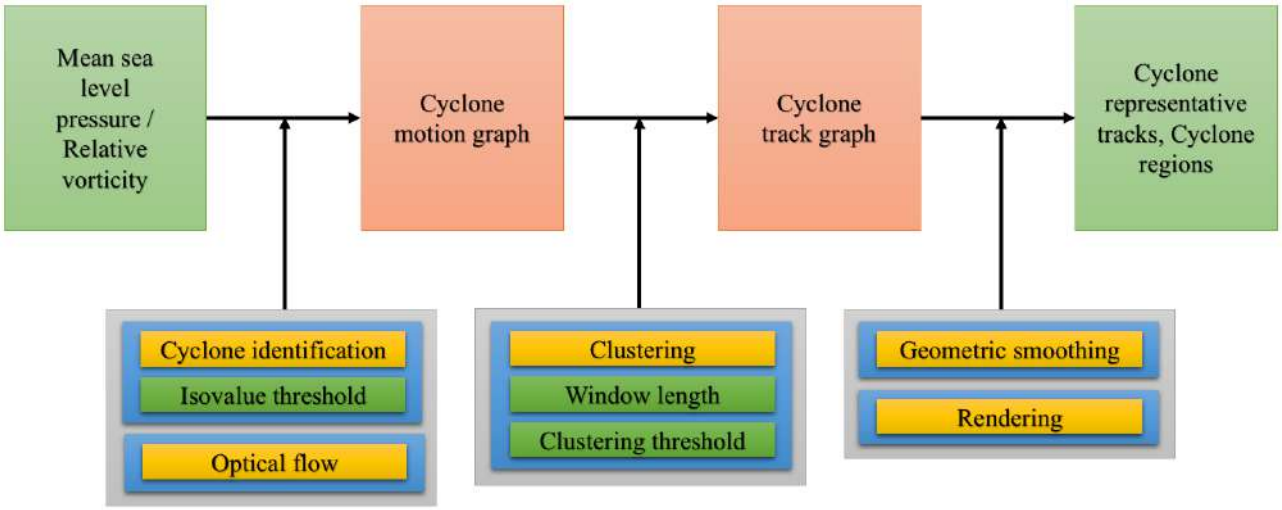


Figure 2.1: Cyclone identification and tracking workflow. Blocks in yellow indicate computations required to construct the intermediate graphs and tracks. The associated parameters are listed below.

2.1 Data

The data used in this study is part of the ERA-Interim reanalysis dataset [4], which is generated by a numerical model constrained by observational data. The data includes a large variety of surface parameters, describing weather as well as ocean-wave and land-surface conditions. In this project we are especially interested in the part of the data related to cyclone formation. Cyclone centres are characterized by the balance in forces due to the pressure gradient and the Coriolis forces. So, they tend to have a pressure minimum accompanied by strong rotational wind near the centre. Due to this reason, the quantities typically used to track cyclones are the vorticity (defined as the vertical component of the curl of the horizontal winds) and mean sea level pressure (MSLP). We use both MSLP and relative velocity in our analysis. Relative vorticity is defined as the vorticity of the wind relative to the earth's planetary vorticity, which is induced due to rotation. Clockwise movements (negative relative vorticity) in the southern hemisphere and counter clockwise movements (positive relative vorticity) in the northern hemisphere are cyclonic. The relative vorticity field is defined at a pressure level of 850 millibars, which represents the free atmosphere above any boundary layer effects. Both fields are available in the ERA-Interim dataset on an equiangular grid with a spatial resolution of 0.75° in both latitude and longitude and a temporal resolution of 6 hours. For each time step, the field is available on a 240×480 grid. Since reanalysis data is typically provided on a latitude-longitude grid, the effective resolution of the data increases closer to the poles. This increase in resolution increases the number of cyclones resolved closer to the poles, resulting in a positive bias in the number of cyclones identified in the polar regions. To overcome this issue, the data is first interpolated onto an equal-area projection (the Lambert equal area projection for each hemisphere, centered on the north/south pole) to equally weight all regions, before interpolating back onto a regular latitude-longitude grid (following [10]). For a given study, the number of time steps is determined by the time period under consideration. Figure 2.2 shows the relative vorticity scalar field over the entire globe on 1st May 1994 using NCVIEW software [22]. The data is available in netCDF format [24]. Similar visualization for mean sea level pressure data is shown in Figure 2.3.

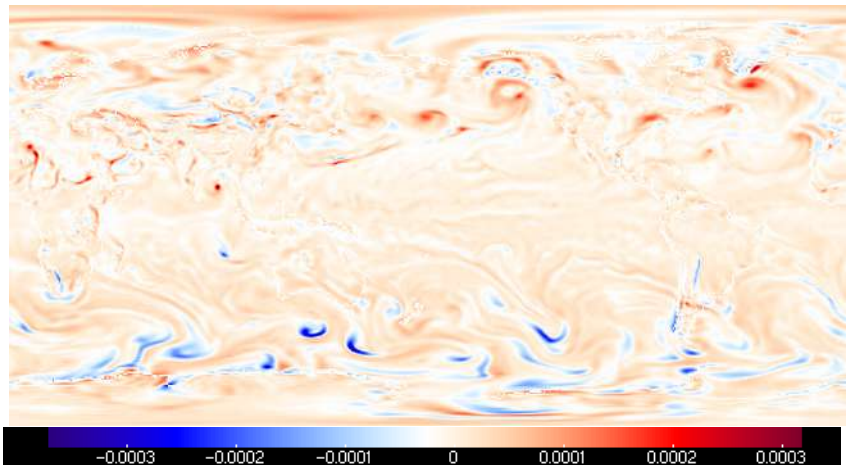


Figure 2.2: Relative vorticity scalar field (s^{-1}) on 1st May 1994. Typically, extreme relative vorticity values indicate cyclones. Clockwise movements(negative values) in the southern hemisphere and counter clockwise movements(positive values) in the northern hemisphere indicate cyclonic regions. This image has been obtained using NCVIEW software [22].

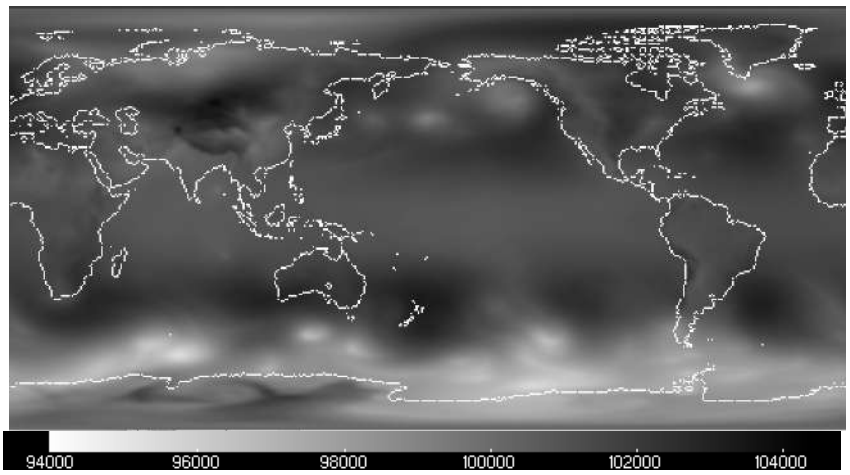


Figure 2.3: Mean sea level scalar field (Pa) on 1st May 1994. Minima seen as hotspots indicates possible existence of cyclones. This image has been obtained using NCVIEW software [22].

2.2 Cyclone Centres

Potential cyclone centres are identified by computing extrema of the two fields. In particular, local minima of MSLP, local minima of relative vorticity in the southern hemisphere, and local maxima of relative vorticity in the northern hemisphere correspond to potential cyclone centres. These definitions are sensitive to noise and the resulting list of centres needs to be filtered to remove noise and weak depressions, and hence to potential cyclone centres. Instead of applying a typical Gaussian smoothing operator, we propose to use a level-set driven simplification for this task. Specifically, for the MSLP scalar field, we compute a topological structure called the join tree and employ a level-set driven simplification to remove or merge minima corresponding to weak depressions. Given a real value c, a *level set* of MSLP is defined as the preimage of the isovalue c. A *sub-level set* of MSLP is the preimage of the interval $(-\infty, c]$. The *join tree* is an abstract representation of the connectivity of sub-level sets of a scalar field, see Figure 2.5(a). Leaf nodes of the join tree correspond to local minima of the scalar field and the root node of the tree corresponds to the global maximum. Given an isovalue threshold, the corresponding level set decomposes the join tree into multiple subtrees, see Figure 2.5(b). Each subtree corresponds to a single cyclone and its deepest minimum is assigned as the cyclone centre. We refer to the subdomain that maps to this subtree as the *cyclonic region*. Local minima with MSLP higher than the threshold belong to the subtree containing the global maximum and are discarded. A remaining challenge is to find an appropriate isovalue to be used as a threshold. We propose an automatic selection where the join tree plays a crucial role in determining an appropriate isovalue threshold as shown in the following section. Further, the join tree data structure enables efficient processing of the minima to identify the cyclone centres. Alternatively, local extrema of relative vorticity may also be used to identify cyclone centres. The same method as above is applied on the relative vorticity data with a focus on maxima for the northern hemisphere and minima for southern hemisphere. Figure 2.4 shows cyclones identified over a region near Australia using the above method applied on relative vorticity data. The *split tree*, which represents connectivity of super-level sets, is computed for identifying cyclone centres in the northern hemisphere. Local maxima of relative vorticity correspond to leaf nodes of the split tree and represent the cyclone centres. Again, the level set corresponding to an input isovalue threshold decomposes the tree into subtrees representing cyclonic regions.

2.3 Cyclone Motion Graph

Studying the temporal behavior of cyclones requires the computation of cyclone tracks. Computing correspondences between cyclone centres in successive time steps using a simple distance-based neighborhood search may not produce expected results, particularly for extra tropical cyclones, because the mean flow advecting the cyclones can vary significantly. Instead, we utilize the optical flow between the MSLP distribution (or relative vorticity) for two consecutive time steps to establish the correspondences. This flow field is used to construct a *cyclone motion graph*, whose nodes are cyclone centres and arcs indicate correspondences between two cyclone centres. An arc is inserted between cyclone centres from two consecutive time steps if the optical flow connects the corresponding cyclonic regions. Details about the generation of these correspondences can be found in Doraiswamy et al. [5], who employ a similar approach to find correspondences between clouds. The optical flow field is computed using OpenCV [2, 8]. A key advantage of using the optical flow field is that the tracking results depend on the entire cyclonic region as opposed to just the cyclone centre. Figure 2.6 shows all the extracted paths generated by traversing the cyclone motion graph. We refer to these paths as the raw tracks. The algorithm identifies 1092 raw tracks during the month of May 1994. This visualization gives a first impression of the cyclones in the region during the month of May but the image is cluttered and contains several redundant and short tracks. The figure shows multiple tracks from different days of the month overlaid on the same region, which evidently make it hard to visually recognize individual tracks or trends. It is also not possible to see clusters of cyclones moving together or separating.

2.4 Cyclone Track Graph

The goal of the final cyclone track graph is to summarize the dominant motion of the cyclone and identify trends of coherent motions of cyclone clusters. It serves as the major input for the visual exploration framework. The input for the generation of this graph are the raw tracks from the cyclone motion graph. Direct clustering of the raw tracks, however, would not give the desired result since it is possible that two tracks are moving in a coherent fashion for some small time window but are not globally similar. We capture these coherent motions by clustering tracks within a moving time window of length w. The parameter w corresponds to the minimum track length that we are interested in. Consider the raw tracks shown in Figure 2.7. A track whose length is less than w is discarded. The red and purple tracks are similar within the $[t, t + w]$ time window. As a distance measure we use the sum of square distances between corresponding cyclone centres of two red (or purple) tracks. For each time window similar tracks are clustered into a group and represented by a single node in the *cyclone track graph*. A directed arc is inserted between two nodes that belong to consecutive time windows if at least a single track from the first node flows into the second node, see Figure 2.8. A node that belongs to a time window $[t, t + w]$ may be represented by the raw track in the cluster that passes through the lowest minimum (MSLP) or highest maximum (relative vorticity) at time t. *Representative tracks* corresponding to all paths in the cyclone track graph are computed next. Figure 2.6(b) shows all representative tracks for $w = 9$ (54 hours, 10 timesteps), with a total of 1479 cluster centres and 137 connected components of cyclone track graph. The highlighted connected component is selected for further study. The tracks are cohesive when compared to the raw tracks and are fewer in number.

Current approaches to cyclone tracking identify significant cyclone centres within each time step and track only these significant cyclone centres. In contrast, our approach to tracking considers both the spatial and temporal distribution of cyclone centres to identify representative cyclone tracks. A natural advantage of our approach is that it reports a single track even when the cyclone contains multiple minima (for MSLP) or maxima (for relative vorticity). Further, the former approaches are sensitive to noise because each time step is processed independent of the next. In particular, cyclonic regions in the early and later stages of a cyclone's lifetime may have a small spatial extent and may hence get filtered away by the former approaches.

2.5 Representative Tracks

An exploratory framework supports the selection of individual tracks for further study. Representative tracks are processed to generate smoother geometry. Specifically, all vertices on a representative track are considered as control points to compute a B-spline curve [31]. Figure 2.6(c) shows the selected representative track with smooth geometry. The cyclonic regions for this track are also computed on demand as the connected components of the sub-level set / super-level set that contain the cyclone centres. The video [1] demonstrates functionalities of the framework. Two cyclonic regions are assigned the same color if they belong to a common region in the past and split, or will merge in future. This merge/split information is readily available from the cyclone motion graph. The video shows a single representative track selected using the exploratory framework. A single smooth track is extracted despite the presence of multiple merges and splits. Other algorithms report multiple tracks near New Zealand and employ specialized postprocessing techniques such as track surgery to remove clutter. In contrast, the track clustering based method generates a single track without the need for additional postprocessing.

[1] https://youtu.be/aEfY7Mo8WuY

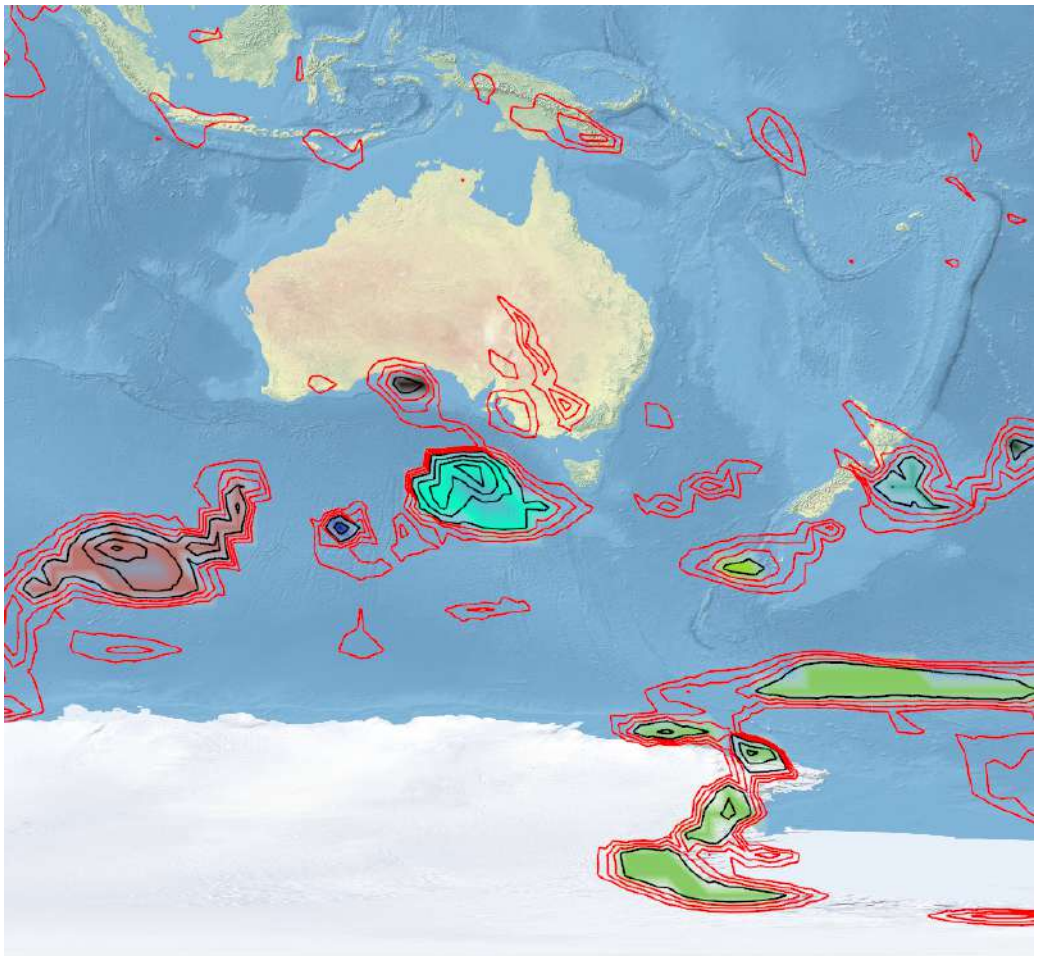
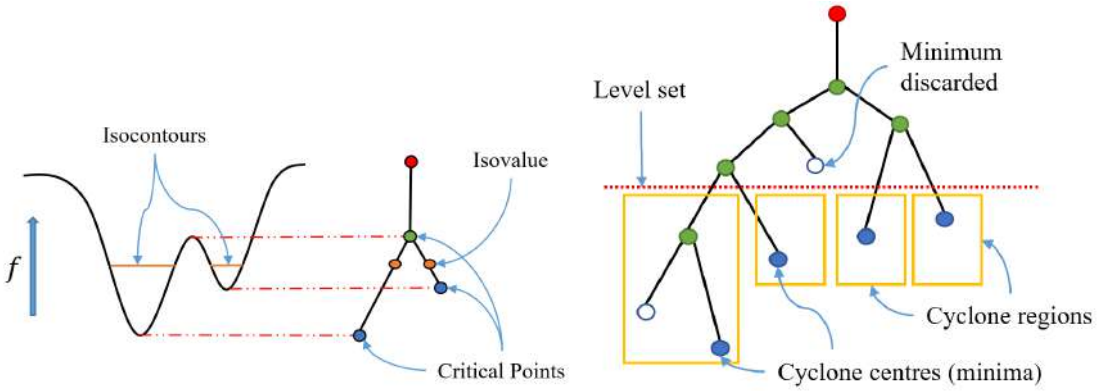
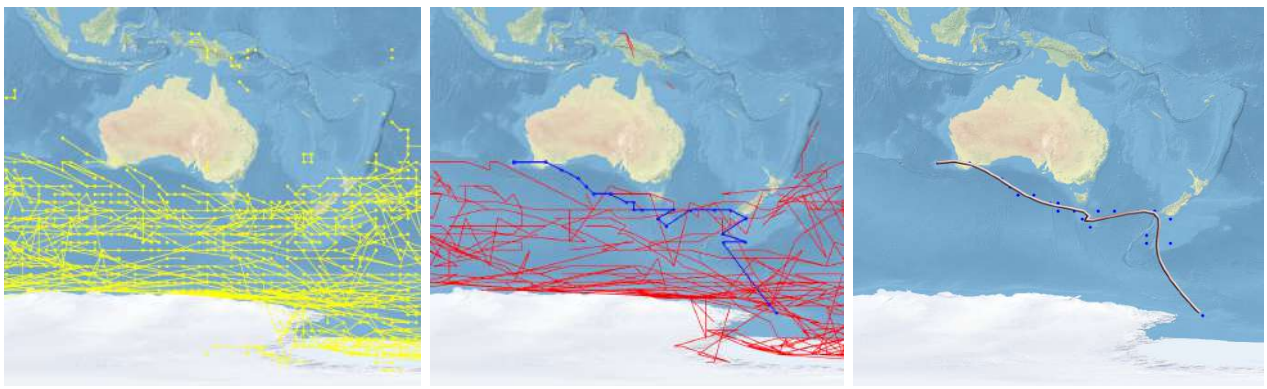


Figure 2.4: Cyclones detected in the southern hemisphere during May 1994 for a vorticity isovalue threshold of $-6 \times 10^{-5}\ s^{-1}$ The black contours have isovalue less than the threshold and belong to the cyclonic regions reported by the algorithm. Red contours have isovalue greater than the threshold. Centres surrounded only by red contours are discarded as shallow cyclones. Cyclonic regions that share a common color have either split in the past or will merge in future.



(a) (b)

Figure 2.5: (a) Join tree is an abstract representation of the connectivity of sub-level sets of a scalar field. Height function defined on a 2D domain (left) and the corresponding join tree (right). Leaf nodes of the tree are minima (blue) of the scalar field or its global maximum (red). (b) Each subtree below a chosen level set corresponds to a cyclone centre.



(a) (b) (c)

Figure 2.6: (a) Raw tracks obtained in the southern hemisphere during May 1994 using a join tree based cyclone centre identification step followed by optical flow field based tracking. Raw tracks from different days of the month may pass through the same region. (b) Track clusters containing tracks within a time window of 54 hours (10 time steps). A connected component of the cyclone track graph is highlighted. This component may be selected for further analysis. (c) One representative track is selected using an exploratory framework and rendered after smoothing. The cyclonic regions may be viewed on demand, see accompanying video.

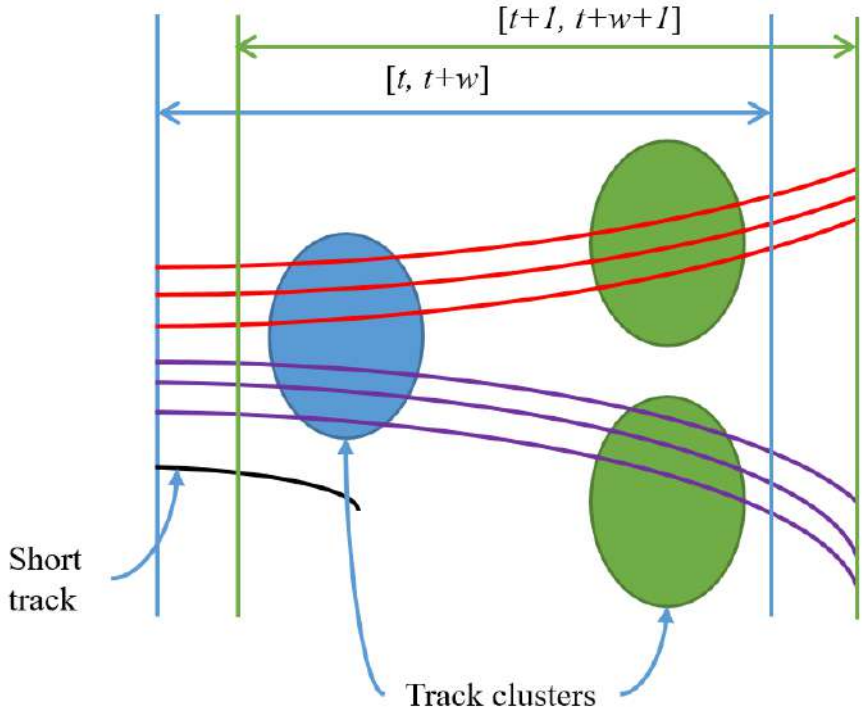


Figure 2.7: Clustering tracks within a moving time window. All tracks that do not span the entire window are ignored. A clustering algorithm is applied for each window. Each cluster is represented as a node in the cyclone track graph.

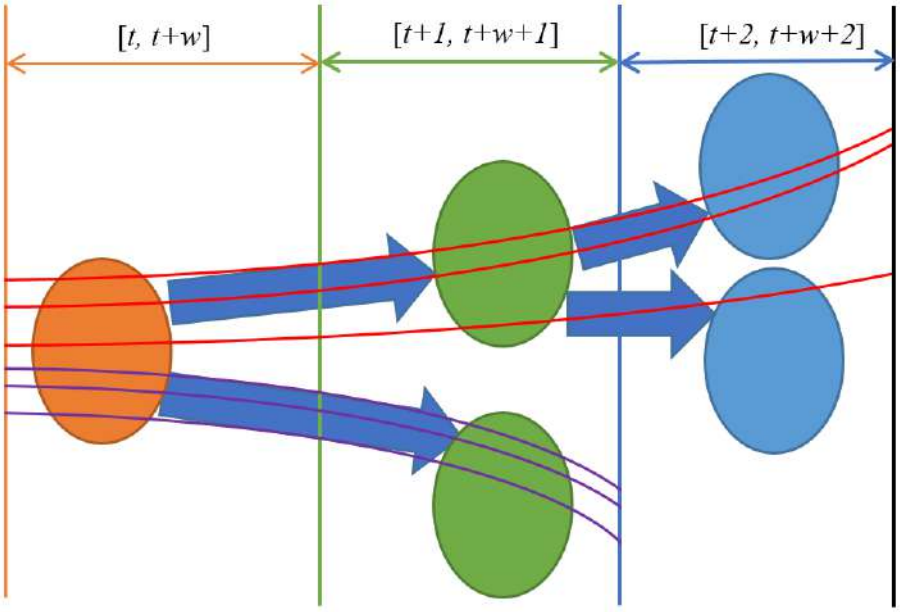


Figure 2.8: An edge is inserted into the cyclone track graph if at least one track passes from one node to another in the next time window. The time windows are shown adjacent to each other for clarity.

Chapter 3

Parameters and Thresholds

We now discuss three different parameters that determine the output (Figure 2.1) and choice of thresholds for each parameter. We aim to develop a generic framework and hence attempt to reduce the number of parameters that need to be tuned by the user. In our experiments, the values for two of the parameters are fixed to constants and only one parameter is tuned for different data sets.

3.1 Isovalue threshold

Cyclone centres are identified from the join/split tree based on an isovalue parameter. A plot of the number of level set components accumulated over all time steps for increasing isovalues helps identify an appropriate threshold. The number of level set components equals the number of subtrees of the join/split tree and the number of cyclone centres. Consider the plot in Figure 3.1 for relative vorticity computed during May 1-30, 1994. The field values are normalized to lie within the interval [0,1] so that the parameter value is chosen as a fraction of the range of field values. Local minima of relative vorticity correspond to cyclone centres in the southern hemisphere. So, the join tree is computed and analyzed here. A low isovalue selects only the most intense cyclone centres and hence the number of components is small. The number of level set components increases as the threshold increases. This is followed by a rapid increase due to the presence of noise and eventually the number of level set components decreases because the different components merge. A point in the neighborhood of the knee of the curve (high curvature region) in this plot is chosen as the threshold. The plot is similar for other time periods considered in our experiments. In this case, we chose a threshold value of 0.45.

Sensitivity: We now study the effect of the choice of isovalue threshold on the results. Consider extreme values (both low and high) and values close to the chosen threshold of 0.45. For each threshold value, we compute and count the raw tracks from the cyclone motion graph and representative tracks from the cyclone track graph. First, we discuss observations for extreme isovalue thresholds. Figure 3.2 shows the raw tracks over the Australian bay area during the month of May 1994 for different isovalue thresholds. Figure 3.2(b) shows the raw tracks computed for an isovalue threshold of 0.45 and may be considered as a reference. A low value of 0.3 results in a small cyclone motion graph and hence significantly small number of raw tracks. At a high value of 0.6, level set components merge into large components and hence raw tracks are longer but small in number. The total number of raw tracks over the southern hemisphere and the symmetric difference with the set of raw tracks obtained for a threshold of 0.45 are shown in Table 3.1. Next, we study how the results are affected by small changes to the isovalue threshold. The symmetric difference with the set of raw and representative tracks for the reference threshold of 0.45 is computed for thresholds in the range [0.44, 0.46]. We note that the cardinality of the symmetric difference increases gradually as expected. A visual inspection reveals the possible reason for the significant size of the symmetric difference. Figure 3.3 shows a few of representative tracks together with the symmetric difference for neighboring threshold values. We observe minor variations in the genesis and lysis of the cyclone tracks and no

changes otherwise.

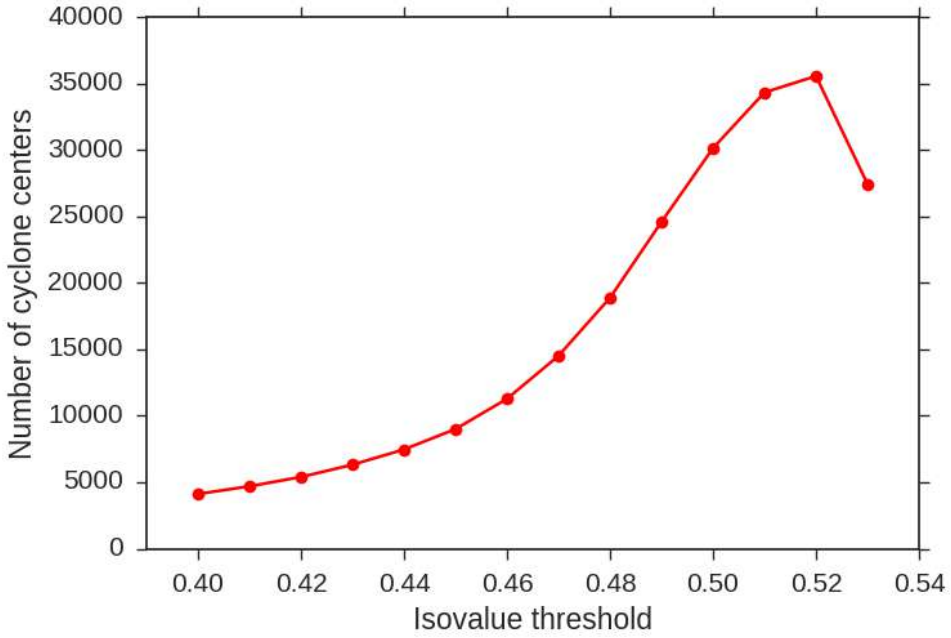


Figure 3.1: Plot of number of isocontour components for various isovalues of relative vorticity during May 1-30, 1994. The relative vorticity values are normalized to lie within [0,1] and lower values at cyclone centres correpond to deeper cyclones. Fewer cyclones are reported for low isovalues. The number of cyclones increases with increasing isovalues. The increase is rapid due to noise or presence of weak depressions. The number of cyclones eventually decreases. We select the knee of the curve, 0.45, as the threshold.

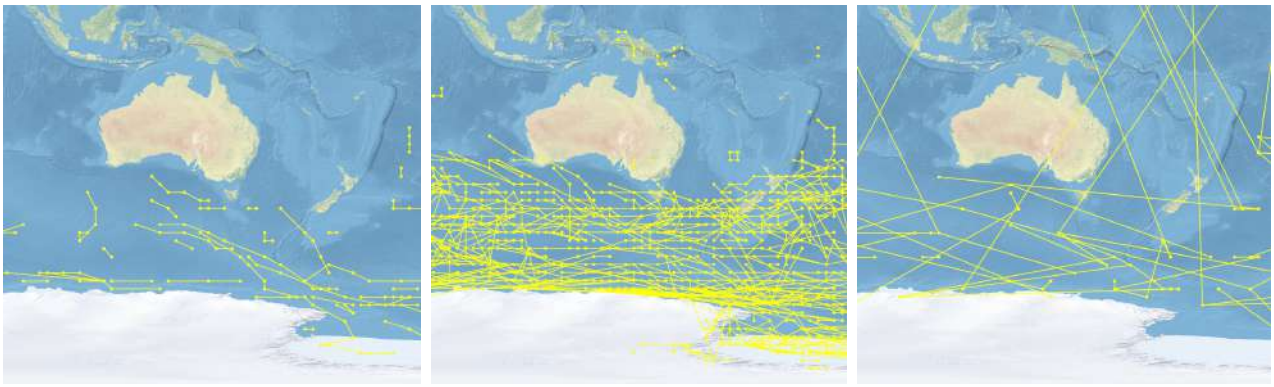


Figure 3.2: Analyzing the effect of choosing extreme values for isovalue threshold on the raw tracks. Tracks are computed over the Australian bay area during May 1994. (a) Raw tracks obtained using an isovalue threshold of 0.3. Only deep cyclone centres and a small number of raw tracks are identified for low thresholds values. (b) Raw tracks obtained for the reference isovalue threshold of 0.45, which was selected using the plot in Figure 3.1. This threshold captures most of the cyclonic activities in the region. (c) Isovalue threshold of 0.60 gives rise to large isocontours resulting in longer raw tracks. Multiple cyclone centres combine into a few large isocontours thereby reducing the number of cyclonic regions and raw tracks. The cyclone motion graph changes significantly with large changes in isovalue threshold.

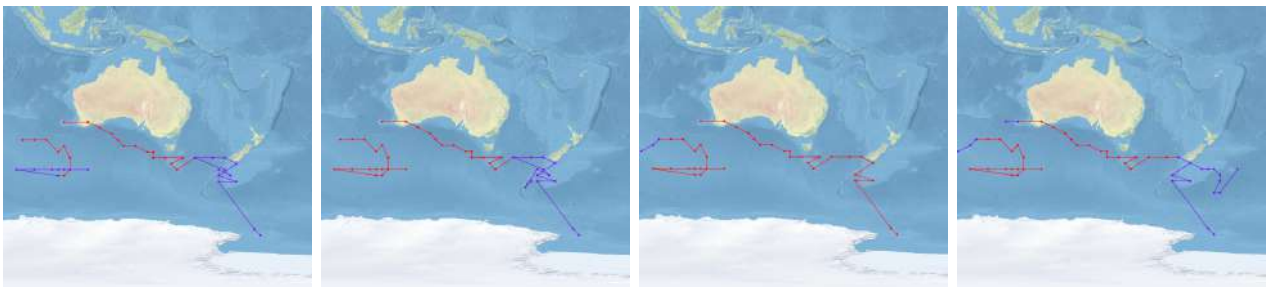


Figure 3.3: Analyzing the effect of small changes to isovalue threshold on the representative tracks computed over the Australian Bay area during May 1994. The clustering threshold is kept constant at 44. (a) The representative tracks (red + purple) computed using an isovalue threshold of 0.45. Purple sections of the tracks correspond to the symmetric set difference between the tracks obtained using isovalue thresholds 0.45 and 0.445. (b,c,d) Symmetric set difference of representative tracks obtained using isovalue threshold of 0.45 against thresholds 0.447, 0.453, and 0.455. Minor variations in the isovalue threshold may affect the genesis and lysis of the cyclone track reported by the algorithm. However, a significant fraction of the track remains unaffected.

3.2 Window length

A good choice of window length w helps eliminate temporal noise. We set $w = 9$ (10 timesteps) as the default window length with an aim to capture longer cyclone tracks in our experiments. This also ensures that the minimum cyclone track length is 54 hours.

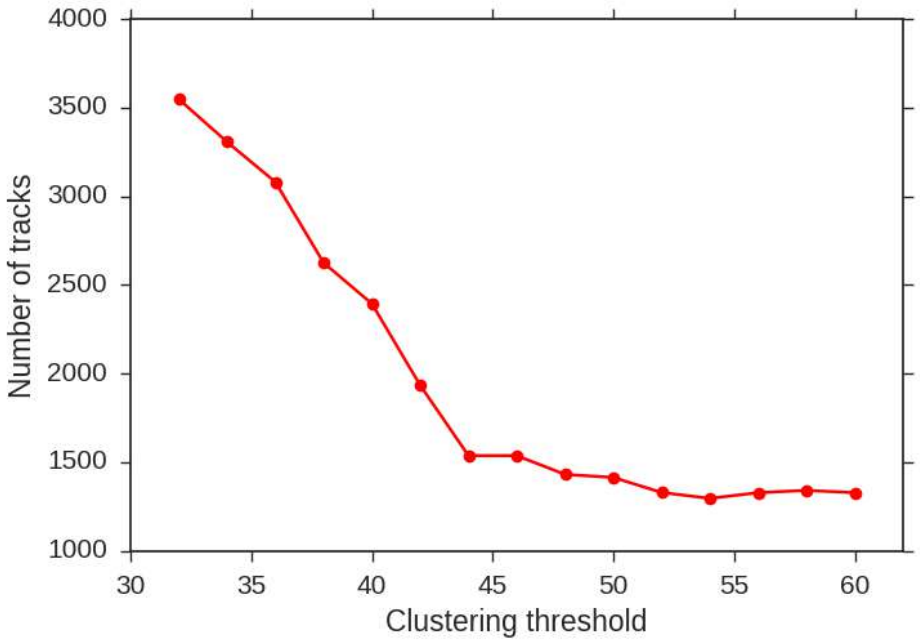


Figure 3.4: Plot of number of tracks computed from relative vorticity during May 1-30, 1994 for varying clustering threshold values. Each path in the cyclone track graph corresponds to a distinct track. Large number of tracks are detected for small clustering thresholds. As the threshold increases, tracks merge together. When the threshold increases further, far away tracks are clustered together. A clustering threshold of 44 being the knee of the curve is selected.

3.3 Clustering threshold

Once the window length is determined, tracks are clustered within each time window. We observed that a simple distance threshold based clustering method performs well. The current implementation of the framework uses single link hierarchical clustering [20]. Each track is represented as a point in a $2(w+1)$-dimensional space with coordinates specified by appending the 2D coordinates of cyclone centres. The distance between two tracks is computed as the euclidean distance in this $2(w+1)$-dimensional space. The result of the hierarchical clustering algorithm is represented as a dendrogram. A clustering threshold helps identify the appropriate level of the hierarchy in the dendrogram and hence the clusters. We use the cophenetic distance as the clustering threshold [30, 14]. For a given window length w, the clustering threshold is equal to the cophenetic distance divided by $\sqrt{2(w+1)}$. For clarity, we ignore the term in the denominator in our discussion. The cophenetic distance between two tracks is defined as the height of the dendrogram at which the tracks merge into a single cluster. We report two tracks as belonging to the same cluster if their cophenetic distance is smaller than the clustering threshold.

Figure 3.4 shows a plot of the number of tracks reported for different values of clustering threshold. Multiple tracks are detected for small distance thresholds. We observe an initial steep decrease in the number of tracks because the proximal raw tracks that correspond to a single cyclone are clustered together. Further increase in the threshold bring spatially remote raw tracks together. We expect the raw tracks corresponding to different cyclones to be well separated in space-time. In all experiments, we observe a similar steep decrease in the graph plot followed by a relatively flat section. We set the clustering threshold equal to the knee of the curve, which effectively separates spatially remote tracks. The plot is similar for other time steps as well. So, we set a uniform clustering threshold of 44 when $w = 9$ (54 hours, 10 timesteps). The knee of the curve, and the relatively stable number of cyclone tracks identified above this value, probably reflect the fact that cyclones tend to have an effective radius of around 650-700 km [25]. The steep increase above the knee of the curve suggests that multiple tracks belonging to a single cyclone (i.e, within 700 km of the cyclone center) are getting clustered into one track. Below the knee of the curve, tracks from different cyclones might be clustered. However, since cyclones are usually separated in both space and time, clustering is no longer as effective, and the number of cyclone tracks tends to stabilise.

Sensitivity: We study the sensitivity of the results to the choice of clustering threshold by comparing the representative tracks. The isovalue threshold and window length parameters are fixed. First, we study the behaviour at extreme values. Figure 3.5 shows a representative track computed for varying clustering thresholds. The tracks for the extreme values of 20 and 60 have minor to significant differences from the one computed using the chosen threshold of 44. The tracks in Figure 3.5(a) contains a small split before lysis and hence result in two representative tracks that differ only at the lysis. Remote tracks are clustered together when the clustering threshold is increased to 60, again resulting in multiple representative tracks that are significantly different from each other. Next, we study the behaviour for neighboring values in the range [40,50]. We observe that the same representative track is obtained for all values in this range.

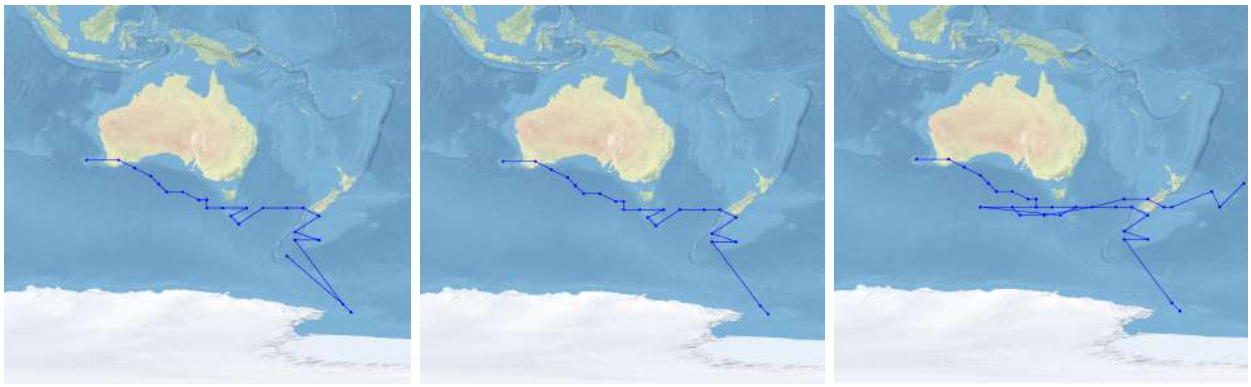


Figure 3.5: Sensitivity analysis of the clustering threshold. Representative tracks over the Australian bay area during May 1994 computed with constant isovalue threshold of 0.45. (a) Clustering threshold is set to 20. A small split is visible close to lysis. The two proximal sections of the tracks are not clustered together. (b) Representative track for clustering threshold equal to 44, as identified from the plot in Figure 3.4. This track, and its lysis, is physically more plausible, because large jumps of the cyclone track towards the west are highly unusual since the prevailing winds in this region are towards the east. Varying the clustering threshold between 40-50 has no effect on the representative track. This indicates stability of the selected clustering threshold. (c) Increasing the clustering threshold to 60 results in remote tracks to be clustered together.

Analysis	**Isovalue**	**Raw tracks**		**Representative tracks**	
		Total number of edges	**Symmetric difference (Reference: 0.45)**	**Total number of edges**	**Symmetric difference (Reference: 0.45)**
Extreme Values	0.2	60	5464	0	2121
	0.3	414	5292	51	2138
	0.45	5516	0	2121	0
	0.5	21549	21347	7234	8111
	0.6	442	5842	355	2386
Sensitivity	0.440	4627	2157	1759	1468
	0.445	5038	1288	1957	864
	0.447	5215	857	2020	563
	0.450	5516	0	2121	0
	0.453	5901	887	2253	688
	0.455	6146	1486	2369	998
	0.460	6879	2913	2531	1696

Table 3.1: Effect of the choice of isovalue threshold (relative vorticity) on cyclone tracks that are identified. A track, raw or representative, is a sequence of cyclone centres connected by edges. The number of edges in the tracks are significantly lower for extreme values of the isovalue threshold. A threshold of 0.45 is set as reference value. The symmetric set difference between tracks obtained for the reference value and extreme isovalues is large as expected. Cardinality of the symmetric difference is low for values in the neighborhood of 0.45 implying that a majority of tracks are captured by both isovalues.

Chapter 4

Exploratory Framework

Interactive study of cyclones within a region of interest requires an exploratory framework. The join/split tree, cyclone motion graph, and cyclone track graph are data structures that form the basis of this framework. The user is typically presented with all cyclone tracks within a single time step. The user may choose to visualize individual cyclone tracks via a query interface. The cyclonic regions can be optionally displayed over the track to understand the cyclone movement. All queries are transformed into simple path traversals within the corresponding graph and hence efficiently processed. The tracks are overlaid on a 1:50m raster map of the earth provided by Natural Earth [15], which has shaded relief and water and coloring based on land cover.

The framework consists of 4 important phases:

4.1 Preprocessing

The data is available in the form of satellite imagery and represented as a time-varying scalar field. For each time step, function values are available at vertices of a structured grid. For convenience, all the function values are normalized in the range of [0,100] where 0 is minimum and 100 is maximum across all the time steps. We want to compute a vector field between two time steps using the optical flow algorithm. So, every time step is stored individually as an `.off` file. The first line of this file indicates the number of vertices, number of faces, and number of edges as comma separated values. The following lines are the coordinates of vertices together with their function values. This is followed by a list of faces specified by vertex indices.

A python script is invoked to generate a vector field, the optical flow field, between every

pair of consecutive time steps using the Farneback algorithm [8]. The algorithm is run twice to obtain forward and backward going vectors. The output vectors are written to binary files.

4.2 Computing Cyclone Motion Graph

This stage begins with reading `.off` files from the previous stage. For every `.off` file, a join/split tree is computed. The minima/maxima and other topological features identified by join/split tree are stored for each time step. The minima are represented as nodes of the cyclone motion graph. Next, the binary files created using optical flow algorithm are read to fetch vector field per timestep. If there exists a vector going from one node to another node in adjacent timestep, an edge is inserted in the graph. A path extracted from cyclone motion graph is called as a raw track.

4.3 Computing Track Graph

As explained in second chapter, clustering algorithm is used to simplify tracks and remove clutter. However, not entire time length can be used for clustering. This is because not every track would span the entire interval. Hence, we introduce the notion of a time window. We iteratively compute all raw tracks within each time window and write it to a file. We use SCIPY library [14] in python to perform clustering on the tracks from the file. The output is written to another file and read back by the framework. The time window is then shifted by one and clustering is performed. At the end, track graph is constructed. Every cluster per time window becomes a node in the track graph. Edges are inserted between the nodes if at least one raw track passes from one node to another. A path extracted from cyclone track graph is called as a representative track. The extracted representative tracks are smoothened using B-spline curves [31].

4.4 Visualization

The framework uses Java Swings and OpenGl to generate the desired output. Figure 4.2 shows one of the instances for February 1992 mean sea level pressure data. The right top display shows the date and time of the time-frame. The user selected isovalue threshold is displayed next to it. The user can walk through the next or previous time steps using right and left keys on the keyboard. The cyclonic regions with the same color have either split in the past or will merge in future. This information is available from the cyclone motion graph. Next, the user can select interesting cyclonic regions using `ctrl` with the left click button. Latitude-longitude and deepest normalized minimum value of the selected cyclonic region are displayed on the right panel. This helps the user to quickly identify the interesting cyclonic area and visually track them. Zooming, panning features are supported using standard mouse operations.

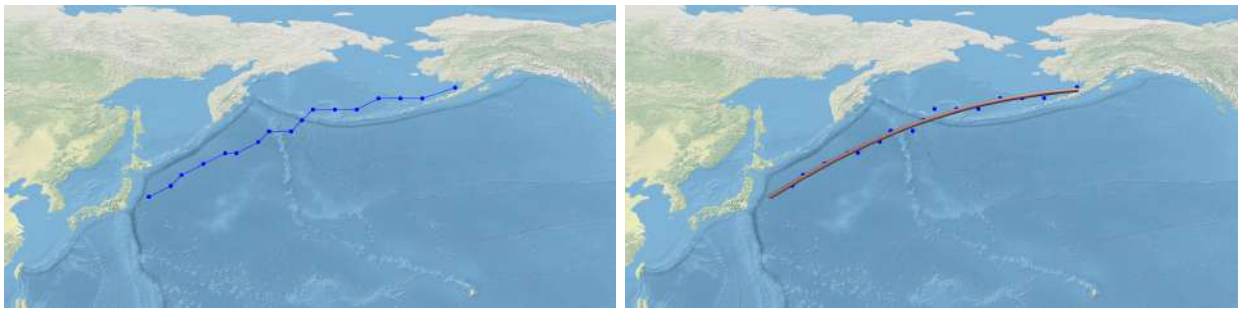
4.4.1 Visualization of cyclonic regions

The framework supports the study of the spatial extent of cyclonic regions within a time step. The isovalue corresponding to the boundary of the cyclonic region may be tuned to explore the distribution of mean sea level pressure or relative vorticity in the neighborhood. The cyclonic region may also be interactively tracked over time. The color of the cyclonic region indicated splits in the past and future merge events. Isocontours of the field within a neighborhood of the cyclone are also displayed. Cyclones often contain multiple minima/maxima. The isocontours help the climate scientist understand the multicentered nature of the cyclone, its intensity, and spatial extent. The join/split tree enables efficient extraction of the relevant isocontours components corresponding to a cyclone centre.

4.4.2 Visualization of representative tracks

Each representative track is displayed as a B-spline curve. Optionally, the cyclone centres of the representative track and the corresponding raw tracks may also be displayed. This view may help understand any unexpected track geometry. Vertices of the representative track may also be located at the weighted average of all cyclone centres at time t in a cluster instead of the cyclone centre with the deepest field value. This avoids zig-zag between different raw tracks and leads to smoother tracks. Figure 4.1 shows the representative track obtained for one of the cyclones in February 1992. Figure 4.1 (a) shows the skeleton of the cyclone representative track. This representation provides deeper understanding of geographical movement of the cyclone. Figure 4.1 (b) shows the same track after geometric smoothing. This is obtained

using smooth curved pipe based rendering. This shows how easily one can switch between two representations to gain interesting insights. Figure 4.3 shows cyclone evolution along its track. This can be manually visualized by the user by moving forward/backward in time using keyboard keys. This helps meteorologist to manually study effect of neighboring weather system on cyclone evolution. Another feature enables visualization of cyclone evolution in a single frame as shown in Figure 4.4. This helps in quick visual analysis of cyclone evolution along the track. The cyclonic regions in Figure 4.4 are stretched/elongated due to underlying relative vorticity scalar field. Additional key controls are provided so as to toggle the display of tracks, cyclone regions, contours, etc. for better visualization. The framework also allows the user to change thresholds and visualize their effect. Thus, apart from the suggested threshold using the plots, a meteorologist can manually change any threshold as per the application and find the interesting features. If an interesting area is already known, such as cyclone in Arabian Sea or user is interested in studying a specific cyclone track obtained using the framework, the respective latitude-longitude configuration can be used to restrict the area under consideration. This captures genesis and lysis of the cyclones well.



(a) (b)

Figure 4.1: One of the cyclone track obtained in February 1992. (a) Representative track obtained by connecting cyclone centers. (b) Same track shown after geometric smoothing. Circular dots represents cyclone centers.

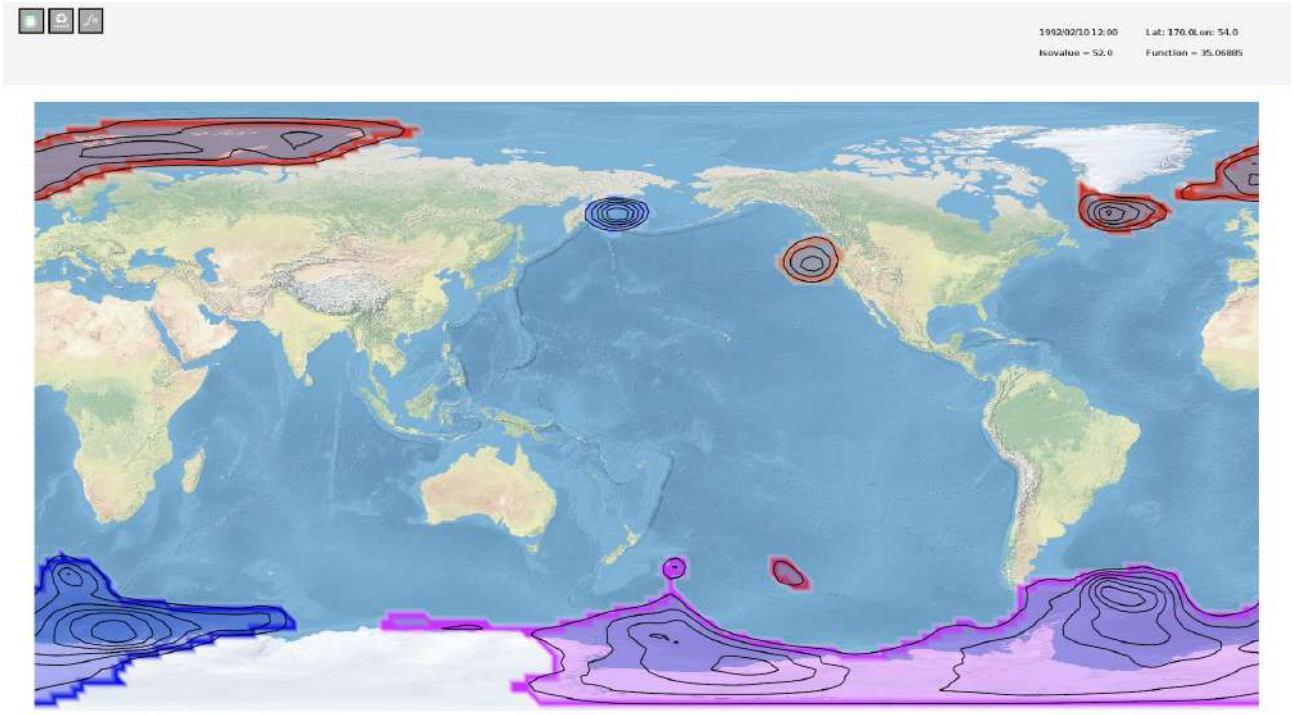


Figure 4.2: Screenshot of framework in action. Canvas area displays identified cyclonic regions. Regions with same color have either split in past or will merge in future. Isocontours within each cyclonic regions help in understanding the location of depression in the cyclone. Top right panel shows the details regarding the timesteps. Details of individually selected cyclone using mouse click are also shown alongside. Other graphical operations like panning and zooming are also implemented.

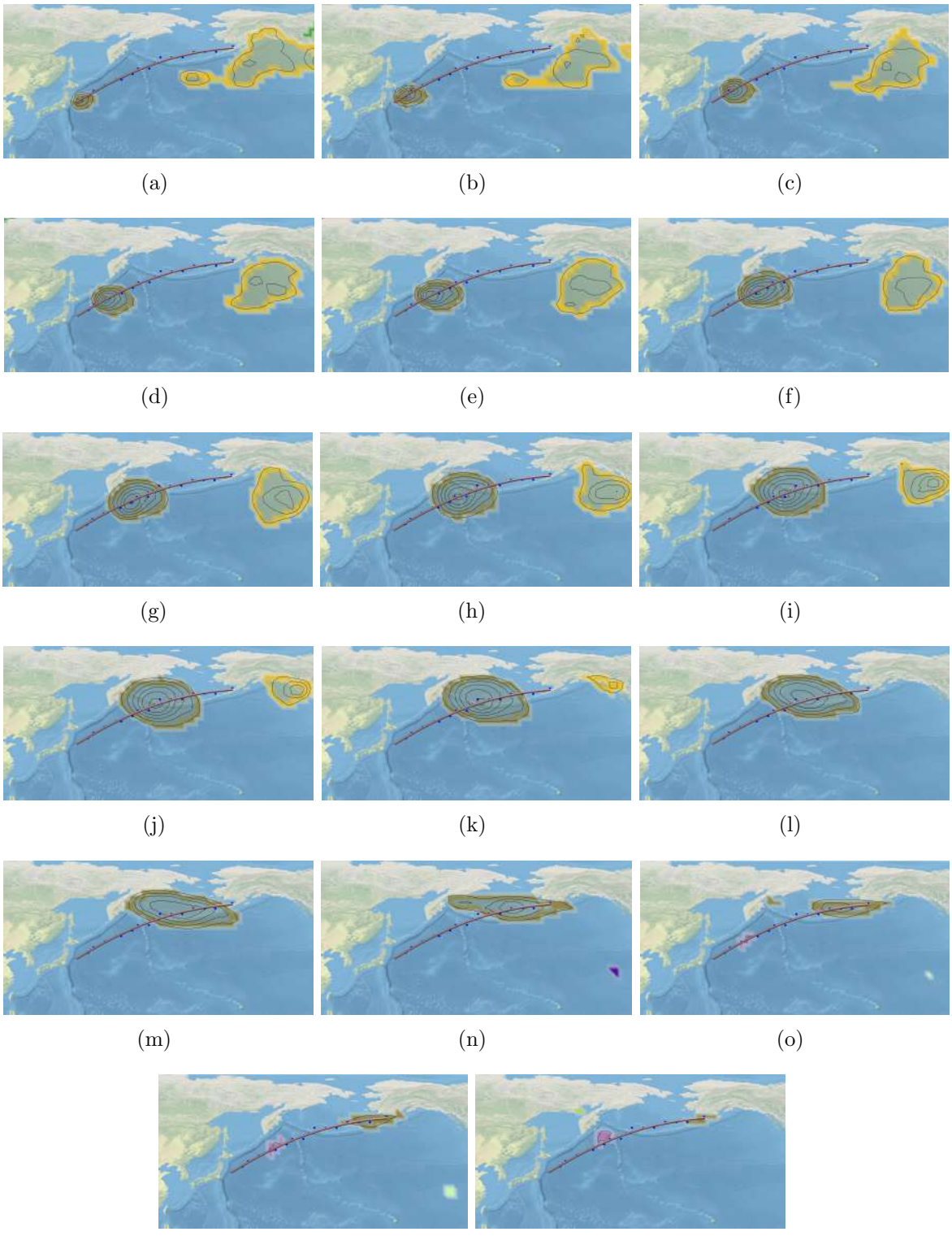


Figure 4.3: Cyclone evolution along the identified track in the month of February 1992. The scalar field is mean sea level pressure. User can manually animate cyclone evolution along the track using keyboard. This helps in comparing the overall cyclone evolution with its surroundings.

Figure 4.4: Figure shows cyclone isocontour evolution in time along the representative track on May 1994. Cyclone isocontours are sampled along the track. The cyclone area extent appears to be elongated due to relative vorticity field. This feature in the framework helps in understanding cyclone evolution over time.

Chapter 5

Results

We perform experiments to study extratropical cyclones over different periods of time.

5.1 Experimental setup

All experiments were performed on an Intel® Xeon(R) CPU E5-2670 v3 @ 2.30GHz x 24. Data is obtained from the ERA-Interim [4] reanalysis data repository. The scalar field under consideration is 6 hourly data of relative vorticity or mean sea level pressure. We first preprocess the data by reprojecting it onto an equal area grid and then normalizing the values to lie within [0,1]. This corrects for bias due to the larger area of a latitude-longitude grid cell near the equator as compared to one near the pole [27]. Next, we employ the workflow described in Figure 2.1. The exploratory framework only loads data pertaining to cyclone centres from the current time step. This process is quick and also reduces the memory footprint. However, all tracks are stored in a single file and loaded initially for visualization. For a 1-month long data set, the preprocessing takes approximately 35 seconds, the optical flow computation takes 2 seconds, and the motion graph and track graph computation takes 40-50 seconds.

5.1.1 Cyclone centre density

Since extratropical cyclones are the main mechanism by which the climate system transports heat from the edge of the tropical regions to the poles, understanding the geographical variability of cyclone activity provides some insight into the "hotspots" of energy transfer. Since these hotspots can be also calculated by alternate means (by calculating the eddy kinetic energy, for example), such a calculation also provides a means to verify the performance of the tracking algorithm.

We compute the cyclone centre density, defined as the number of cyclones centres at a given grid location over 100 time steps. The cyclone centre density distribution over the northern and southern hemisphere for the period 1979-2012 is shown in Figure 5.1 and 5.2. The cyclone centre density is calculated over a grid for an isovalue threshold of 0.45 for southern hemisphere and 0.55 for northern hemisphere. For each identified cyclone centre, we increment the count of cyclones for all grid cells that lie within a 1000 km^2 box neighborhood and normalize by the number of time steps. The cyclone centre density captures the major centres of cyclone activity (or storm tracks) in both hemispheres. The heatmap corresponds well to the density computed using other algorithms [18], see Figure 5.3 and 5.4 for heatmaps from the earlier study. Changing the isovalue threshold from 0.45 to 0.4 for southern hemisphere does not change the overall statistical picture, which suggests that an isovalue in this range would be suitable for scientific applications. Moreover, we use all cyclones identified without filtering based on area or lifetime (as opposed to the IMILAST results which only use cyclones that live for longer than 24 hours).

5.1.2 Temporal variability

Extratropical cyclones are formed mainly in the winter in each hemisphere, and the total number of cyclones formed per season is sensitive to the state of the climate system in that season. For example, cyclone activity is known to be sensitive to the North Atlantic Oscillation and the Pacific North American mode of variability [9]. Therefore, the intrinsic inter-annual and inter-decadal variability of cyclones must first be accounted for before any attempt to understand the impacts of anthropogenic climate change on cyclones.

While a complete characterisation of cyclone statistics is out of the scope of this thesis, we present results for the cyclone counts per season obtained by the algorithm in Figures 5.5 and 5.6. The inter-annual variability of cyclone center counts in Figure 5.5 remains quite similar regardless of the isovalue chosen, which implies that the largest variability is due to changes in the deepest cyclones (which are captured by all thresholds). Thus, natural climate variability seems to influence the genesis of deep cyclones more than shallow cyclones. This inference is inline with the picture presented in the IMILAST comparison as well. Figure 5.6 shows the total number of cyclones identified per year in the northern hemisphere. The number of cyclones is approximately 120 on average, which is half the number identified by Gulev et al. [9]. This suggests that the current implementation captures deeper cyclones more effectively than weaker ones.

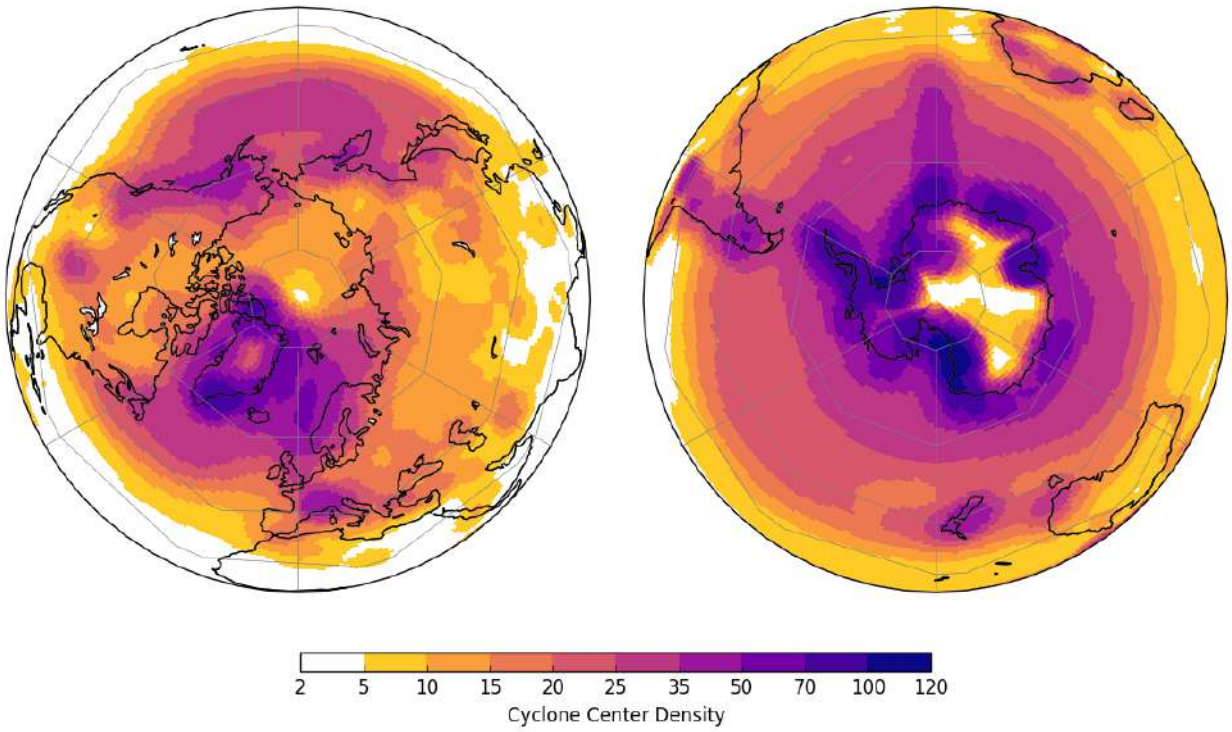


Figure 5.1: Heatmap of cyclone centre density in the winter season for northern (left) and southern hemispheres (right) during 1979-2012. The cyclone centres are identified using relative vorticity.

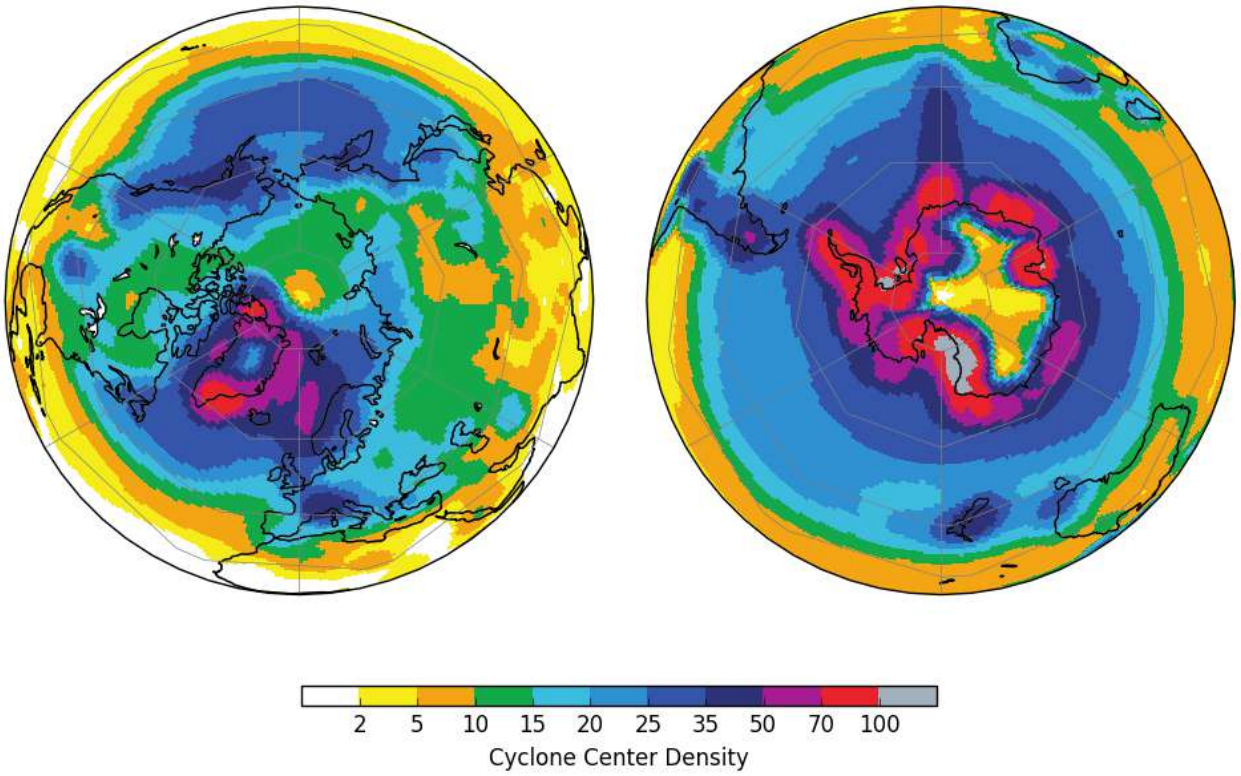


Figure 5.2: Heatmap of cyclone centre density in the winter season for northern (left) and southern hemispheres (right) during 1979-2012. The scalar field used is relative vorticity. The color map used here is similar to the one used by IMILAST.

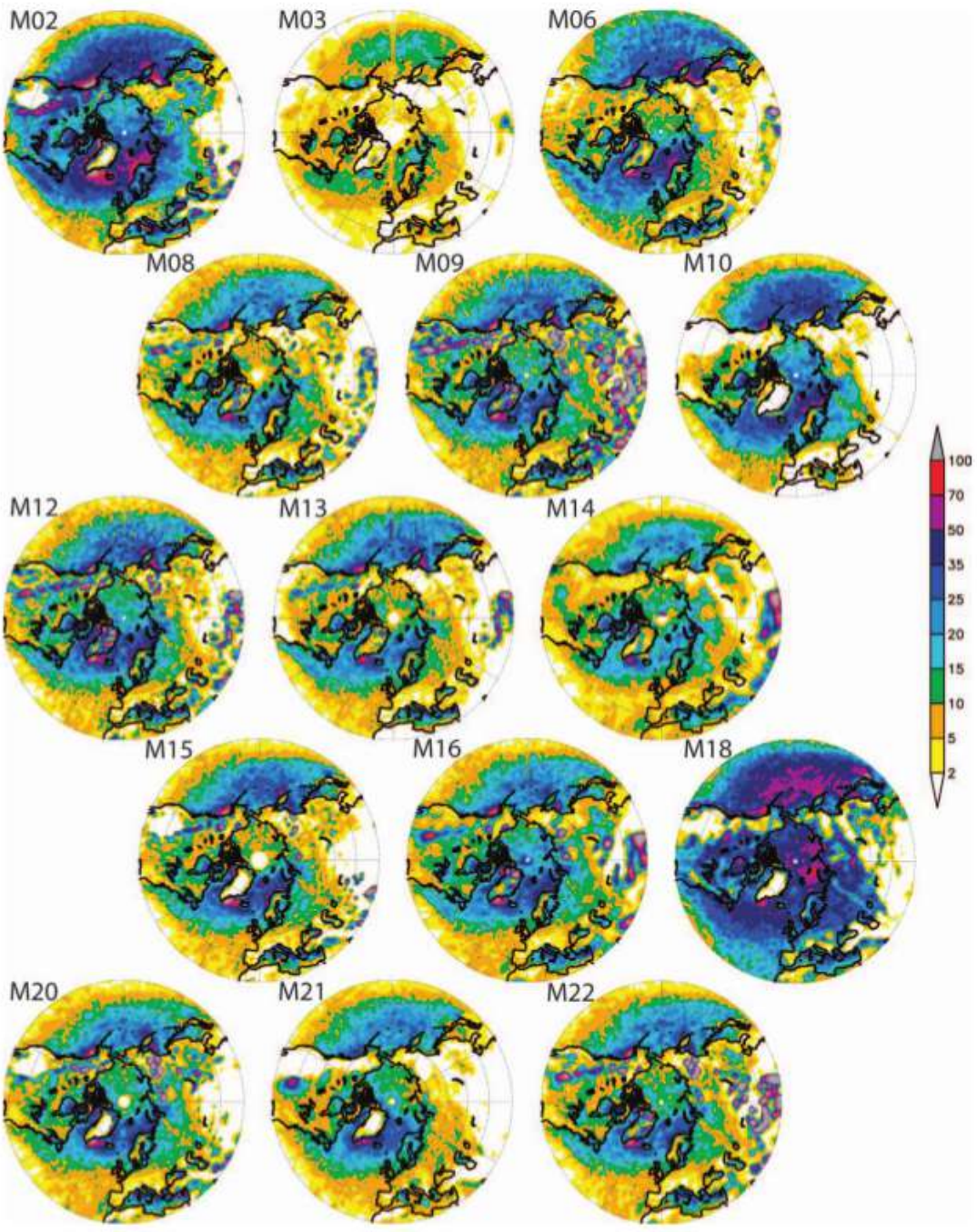


Figure 5.3: Total cyclone center density in the northern hemisphere for cyclones lasting 24 h or more in the winter season for various detection and tracking methods [18].

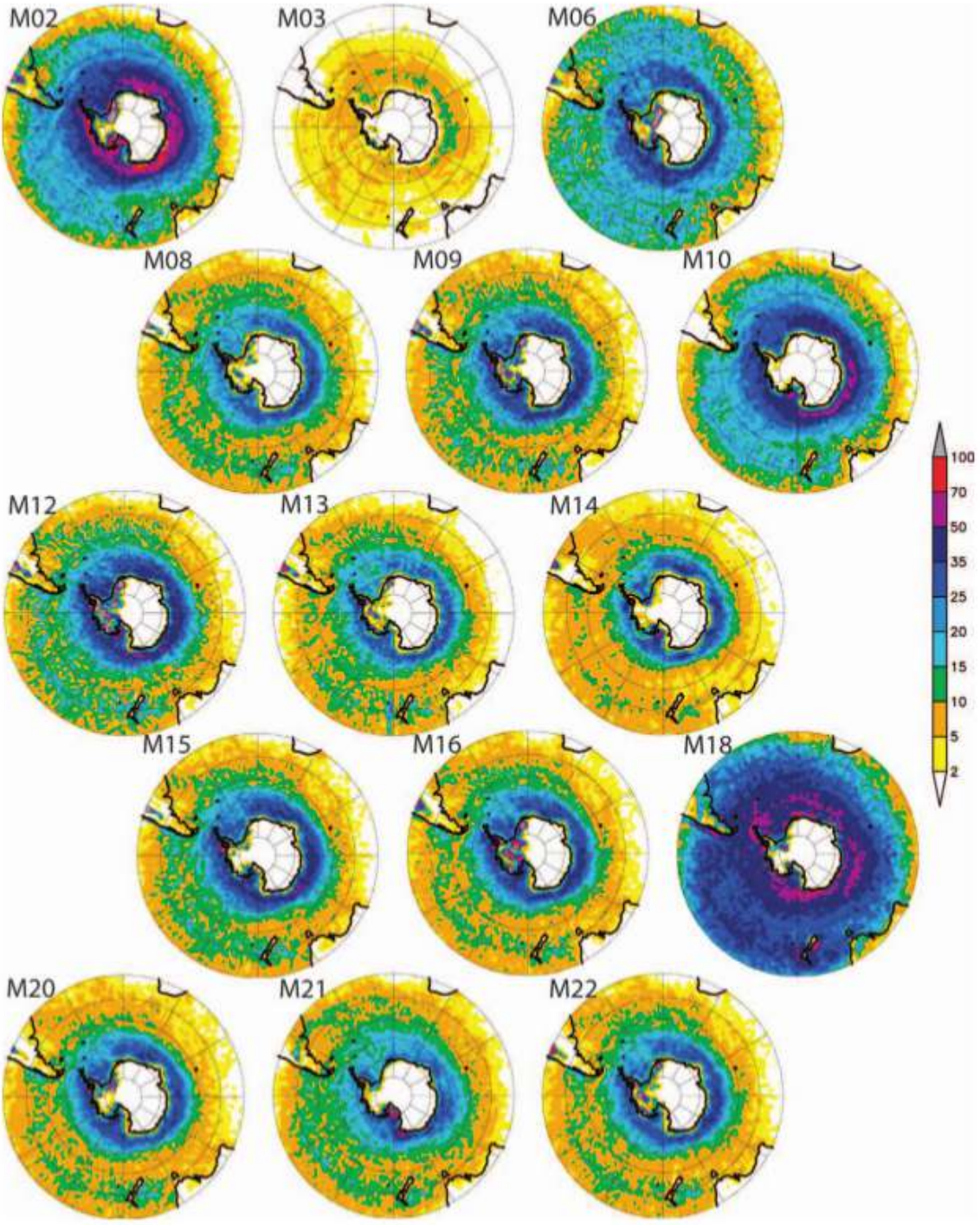


Figure 5.4: Total cyclone center density in the souther hemisphere for cyclones lasting 24 h or more in the winter season for various detection and tracking methods [18]

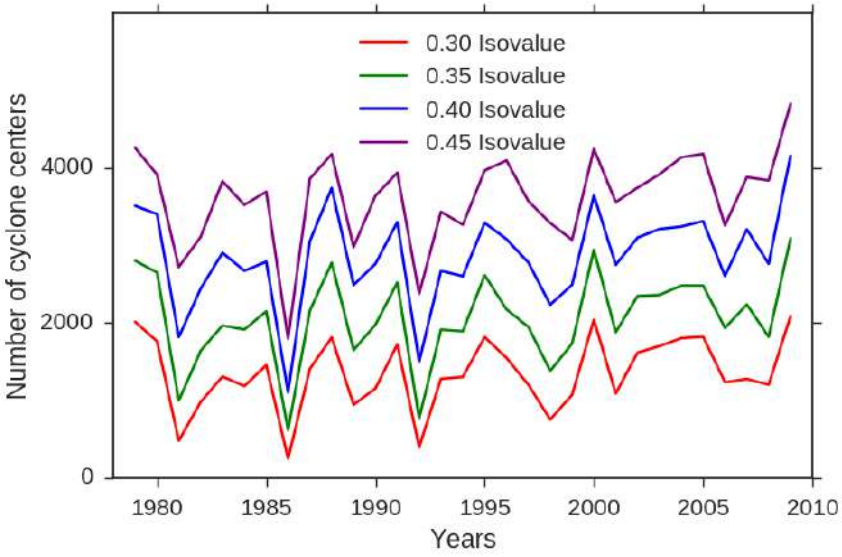


Figure 5.5: Number of cyclone centers identified per boreal winter season (December-February) in the northern hemisphere for different isovalues of mean sea level pressure from 1979 to 2009.

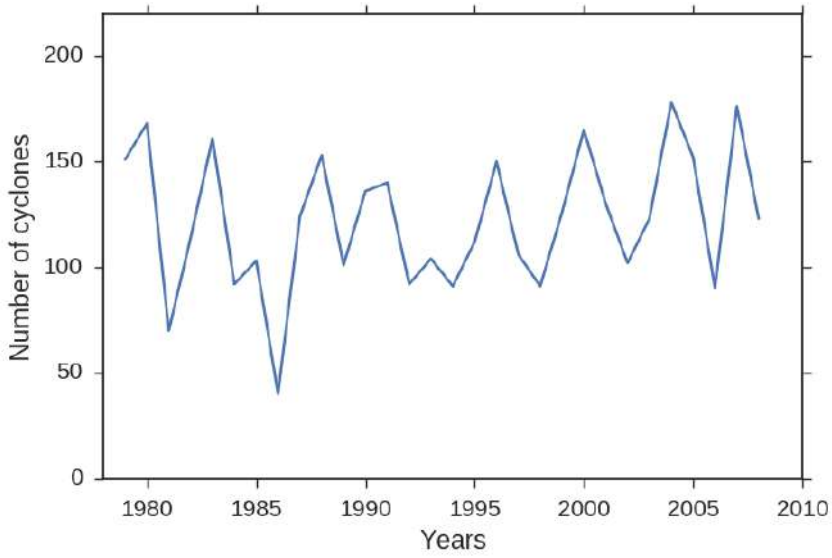


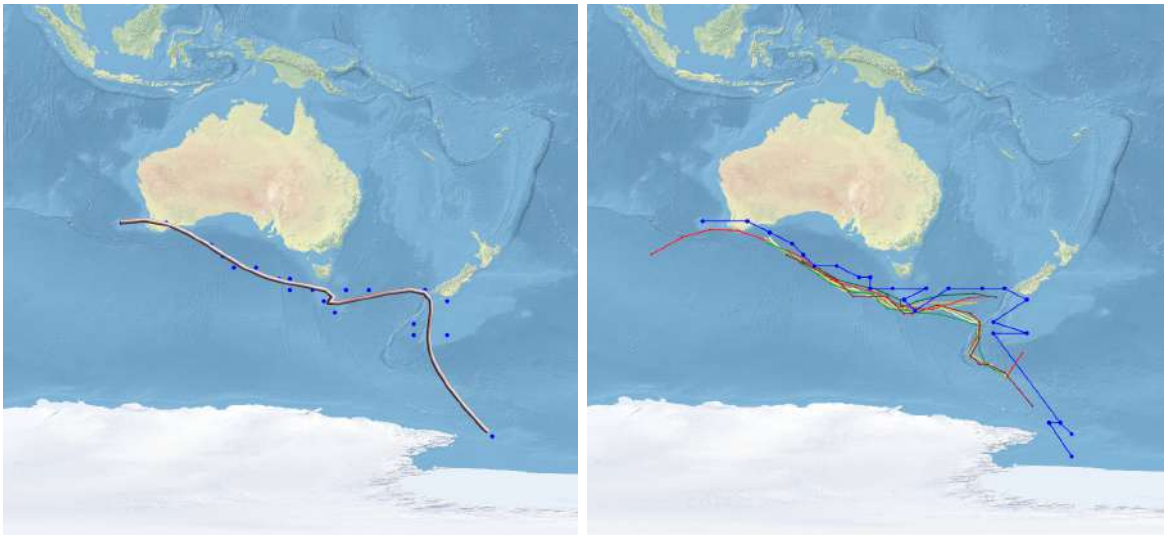
Figure 5.6: Number of cyclones per year in the northern hemisphere for the period 1979 to 2009. The mean sea level pressure isovalue threshold is 0.45.

5.2 IMILAST case studies

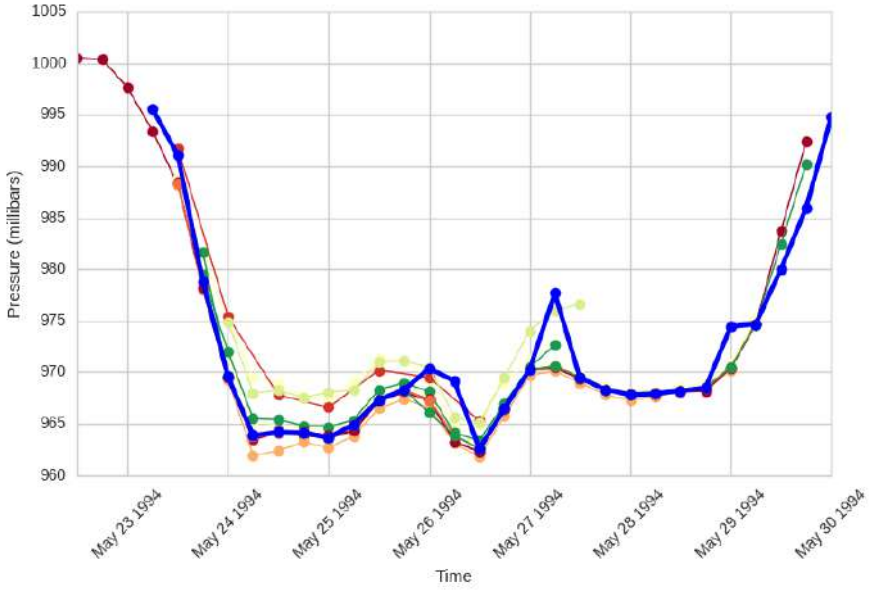
IMILAST [18] was a community effort for the intercomparison of extratropical cyclone tracking algorithms, directed at studying the differences between various tracking algorithms available and the reason for these differences. For purposes of comparison, IMILAST selected two high impact winter cyclones which were previously analysed in the literature and were characterised by complicated life cycle characteristics such as cyclone splitting and re-intensification. They were chosen in different hemispheres to test the tracking algorithm's sensitivity to a change in hemisphere. The detailed synoptic situation and impacts of the two cyclones are described in [18] and references therein.

5.2.1 IMILAST storm 1: 22-29 May 1994

The first case study is an unnamed cyclone that occurred to the south of Australia during the austral winter in 1994. Cyclones are identified using an isovalue threshold of 0.45 and other parameter set to default values. The path traced by different tracking algorithms in the IMILAST comparison is shown in Figure 5.7. The IMILAST study suggests that there was a large disagreement between the methods. Some tracks follow a path towards New Zealand while others turn to the south depending on the sensitivity of the algorithm towards handling of the splitting of the cyclone. Our method is able to capture the southern track as well as the extension of the track towards New Zealand. We do not explicitly show the split in the track because it is small in length and gets clustered into a single track. The evolution of the minimum pressure over the life cycle of the cyclone is shown in Figure 5.7(c). The evolution captured by our algorithm is again similar to earlier algorithms. In particular, the re-intensification of the cyclone on May 26 is captured.



(a) (b)



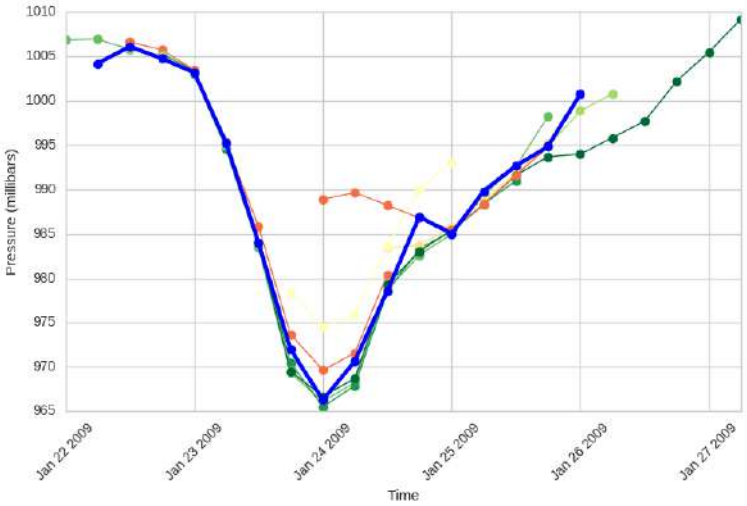
(c)

Figure 5.7: A track obtained near the Australian Bight due to a southern hemisphere storm (22-29 May, 1994). (a) A representative track obtained using our framework. (b) Comparison with tracks obtained from different algorithms reported in the IMILAST study [18]. The track obtained using our framework is shown in blue. The slight upward shift in the track is due to difference in the resolution of the data used. Most tracks either capture eastward or southward movement of the cyclone, not both. (c) Core pressure evolution over the representative track (blue) is similar to the results from IMILAST.

5.2.2 IMILAST storm 2: 22-27 Jan 2009

The next case study is of the cyclone Klaus that originated in the Atlantic and traced a path through the Mediterranean region. It affected northern Spain and southwestern France between 22-27 January, 2009. Maxima of relative vorticity are computed to locate the cyclone centres, with an isovalue threshold of 0.59. The isovalue threshold is used to identify subtrees of the split tree. As shown in Figure 5.8, we compare the representative track obtained from the framework with the tracks reported in the IMILAST study. The genesis of the cyclone is captured by the representative spline track. The representative track is not able to capture the lysis of the cyclone to the east of Italy. We note that earlier algorithms that track cyclones using relative vorticity as an input also fail to capture lysis of the cyclone. The evolution of the minimum pressure over the life cycle of cyclone Klaus is shown in Figure 5.8(c). The life cycle of the cyclone as computed by our algorithm is similar to the results from other algorithms described in the IMILAST study.

(a) (b)



(c)

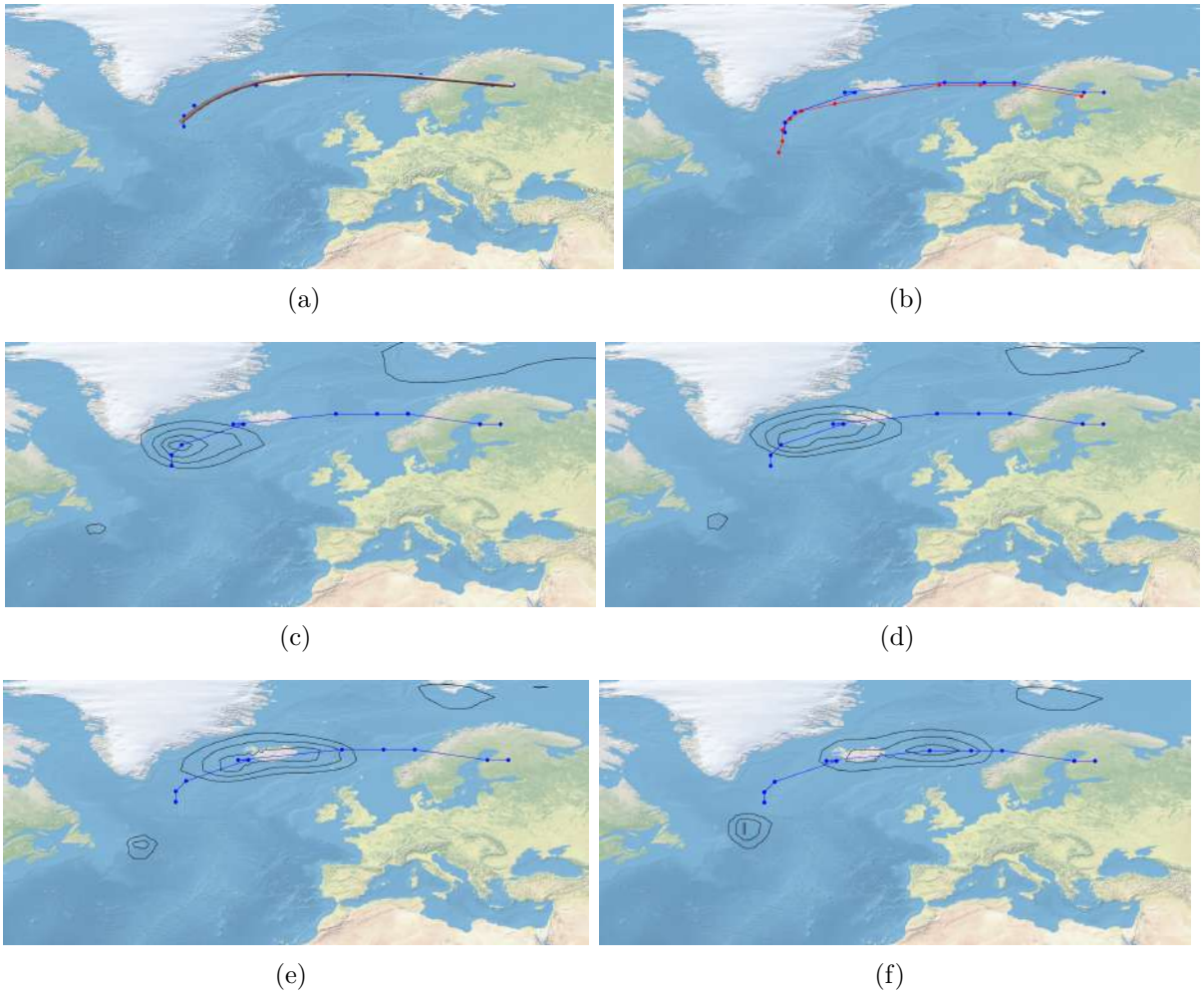
Figure 5.8: A track obtained due to storm Klaus in the northern hemisphere (22-27 January, 2009). (a) A representative track obtained using our framework. (b) Comparison with tracks obtained from different algorithms reported in the IMILAST study [18]. The blue track obtained using our framework traces a similar path as others. (c) Core pressure evolution over the representative track (blue) is similar to the results from IMILAST.

5.3 North Atlantic cyclone

Tracking of cyclones with multiple centres has always been a challenge. Most algorithms consider cyclones with multiple centres as two separate cyclones, leading to double counting and erroneous tracks. This case study illustrates the capability of the proposed algorithm and framework to track multicentre cyclones. This particular case study was used by Hanley and Caballero [10] to illustrate their track surgery algorithm and contrast it with algorithms that are unable to distinguish between single and multicentre cyclones. Figures in their paper are from comparative study, which show that naïve approaches that track individual depressions will generate two separate tracks, one for each cyclone minimum, and not a single longer multicentred cyclone track. Their track surgery operation computes a single track where the cyclonic region at each time step contains mutiple centres. Our framework reports similar results without resorting to the detailed case analysis required by the track surgery operation. Figure 5.9 shows the single representative track obtained by our framework using an isovalue threshold equal to 0.56 and $w = 9$ (10 timesteps). A visual comparison shows that the result is similar to that of Hanley and Caballero.

5.4 Recent cyclone activity: December 2011

December 2011 was marked by multiple high impact winter storms affecting Europe in succession: cyclones Friedhelm (December 7-12) and Joachim (December 15-21) as shown in Figure 5.10. Due to the fact that these cyclones overlapped in space and time, this period was chosen to test the robustness of the tracking algorithm to a complex and dynamic situation. For this case study, we use the mean sea level pressure instead of the vorticity field to track the cyclones. Such an option is not easily available in current tracking algorithms. The generated tracks were compared with weather charts archived at the Institute of Meteorology, FU Berlin [1]. In the case of cyclone Friedhelm, the algorithm captures the life cycle well, apart from the lysis on December 12, when the sea level pressure signature was relatively weak. This is also the case for cyclone Joachim, with the final day of the life cycle being missed. Upon inspection of the weather charts, the reason for both appears to be that the cyclone centre is a low pressure relative to its environment, but not in a global sense. This suggests that using a preprocessing step (such as subtraction of a running mean value) to enhance the cyclone signature or using regional isovalue thresholds might help enhance the ability of the algorithm to capture the full life cycle of the cyclone.



(a) (b)

(c) (d)

(e) (f)

Figure 5.9: Tracking a multicentred north Atlantic cyclone (Dec 2008). (a) A single representative track produced by our framework. (b) Red track computed using a specialized track surgery method for tracking multicentred cyclones [10]. The blue track obtained using our framwork is similar to the previous result. (c-f) Contour plots indicate the existence of two minima within a single cyclone. Tracing the evolution of individual minima will result in two tracks as shown by Hanley and Caballero [10].

(a) (b)

Figure 5.10: Tracks obtained for cyclone Friedhelm (a) and cyclone Joachim (b) in December 2011 using mean sea level pressure data. An isovalue threshold of 0.37 was uniformly used in both cases.

Chapter 6

Conclusions

The identification and tracking of extratropical cyclones remains an operational challenge for meteorologists due to their varied sizes, strengths, and lifecycle characteristics. In this thesis, we have provided an exploratory framework that facilitates identification of cyclones and generates simplified smooth tracks. Our method is generic and requires only a few parameters to be tuned. This allows the use of different scalar fields such as vorticity and sea level pressure for tracking cyclones, a flexibility that is not present in existing methods. We have demonstrated the use of topological methods in simplifying and storing only the important features of the scalar field. Next, we employ clustering of tracks within a fixed time window, which has multiple benefits. It removes temporal noise, identifies redundant tracks, and also merge / split behavior. This approach contrasts with currently used tracking methods that typically use only two time steps, the current and previous ones, to obtain the cyclone tracks. Query based exploratory methods help the user to select and obtain detailed visualization of interesting features.

While the exploratory framework is useful for studying individual cyclones, there are multiple directions for future work. Our current tracking implementation uses a latitude-longitude map projection. This projection increases the area of identified cyclonic regions close to the poles. Further work is required to systematically remove such bias. The algorithm may be further improved to obtain correct density distributions over the polar regions by using physically motivated filtering methods. Incorporating these improvements without introducing additional parameters is a challenge. The query based framework is interactive for time periods of 1-2 months. Improved data structures are required for query processing to scale to data that spans multiple years.

Bibliography

[1] Weather charts. http://www.met.fu-berlin.de/de/wetter/maps/. [Online, accessed 2017-05-19]. 44

[2] Gary Bradski et al. The OpenCV library. *Doctor Dobbs Journal*, 25(11):120–126, 2000. 10

[3] Peer-Timo Bremer, Gunther Weber, Julien Tierny, Valerio Pascucci, Marc Day, and John Bell. Interactive exploration and analysis of large-scale simulations using topology-based data segmentation. *IEEE Transactions on Visualization and Computer Graphics*, 17(9): 1307–1324, 2011. 3

[4] D. P. Dee et al. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597, 2011. 7, 32

[5] Harish Doraiswamy, Vijay Natarajan, and Ravi S Nanjundiah. An exploration framework to identify and track movement of cloud systems. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2896–2905, 2013. 2, 3, 10

[6] Herbert Edelsbrunner and John Harer. *Computational Topology - an Introduction*. American Mathematical Society, 2010. 3

[7] Herbert Edelsbrunner, John Harer, Ajith Mascarenhas, and Valerio Pascucci. Time-varying Reeb graphs for continuous space-time data. In *Proceedings of the twentieth annual symposium on Computational geometry*, pages 366–372. ACM, 2004. 3

[8] Gunnar Farnebäck. Two-frame motion estimation based on polynomial expansion. In *Scandinavian Conference on Image Analysis*, pages 363–370. Springer, 2003. 10, 26

[9] S. K. Gulev, O. Zolina, and S. Grigoriev. Extratropical cyclone variability in the Northern Hemisphere winter from the NCEP/NCAR reanalysis data. *Climate Dynamics*, 17(10): 795–809, July 2001. ISSN 0930-7575, 1432-0894. doi: 10.1007/s003820000145. 34

[10] John Hanley and Rodrigo Caballero. Objective identification and tracking of multicentre cyclones in the era-interim reanalysis dataset. *Quarterly Journal of the Royal Meteorological Society*, 138(664):612–625, 2012. x, 2, 7, 44, 45

[11] Isaac M. Held. The General Circulation of the Atmosphere. http://www.gfdl.noaa.gov/cms-filesystem-action/user_files/ih/lectures/woods_hole.pdf, 2000. [Online, accessed 2015-08-20]. 1

[12] K. I. Hodges. A general method for tracking analysis and its application to meteorological data. *Monthly Weather Review*, 122(11):2573–2586, November 1994. doi: 10.1175/1520-0493(1994)122⟨2573:AGMFTA⟩2.0.CO;2. 2

[13] Masaru Inatsu. The neighbor enclosed area tracking algorithm for extratropical wintertime cyclones. *Atmospheric Science Letters*, 10(4):267–272, 2009. 2

[14] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: open source scientific tools for Python. http://www.scipy.org/, 2001–. [Online, accessed 2017-05-19]. 22, 26

[15] Nathaniel Vaughn Kelso and Tom Patterson. Introducing natural earth data-naturalearthdata. com. *Geographia Technica*, 5:82–89, 2010. 25

[16] Ajith Mascarenhas and Jack Snoeyink. Isocontour based visualization of time-varying scalar fields. In *Mathematical Foundations of Scientific Visualization, Computer Graphics, and Massive Data Exploration*, pages 41–68. Springer, 2009. 3

[17] Ross J. Murray and Ian Simmonds. A numerical scheme for tracking cyclone centres from digital data. *Australian Meteorological Magazine*, 39(3), 1991. 2

[18] Urs Neu, Mirseid G Akperov, Nina Bellenbaum, Rasmus Benestad, Richard Blender, Rodrigo Caballero, Angela Cocozza, Helen F Dacre, Yang Feng, Klaus Fraedrich, et al. IMILAST: A community effort to intercompare extratropical cyclone detection and tracking algorithms. *Bulletin of the American Meteorological Society*, 94(4):529–547, 2013. ix, x, 2, 4, 5, 33, 37, 38, 40, 41, 43

[19] P Oesterling, C Heine, GH Weber, D Morozov, and G Scheuermann. Computing and visualizing time-varying merge trees for high-dimensional data. In Hamish Carr, Christoph Garth, and Tino Weinkauf, editors, *Topological Methods in Data Analysis and Visualization IV: Theory, Algorithms, and Applications*. Springer International Publishing, 2017. 3

[20] Michael Steinbach Pang-Ning Tan and Vipin Kumar. *Introduction to Data Mining*. Addison Wesley, 2005. 22

[21] J. P. Peixoto and Abraham H. Oort. *Physics of Climate*. American Institute of Physics, New York, 1992. 1

[22] DW Pierce. Ncview. *Climate Res. Div., Scripps Inst. of Oceanogr., Univ. of California, San Diego. http://meteora. ucsd. edu*, 80, 2009. vi, 7, 8

[23] Jan Reininghaus, Jens Kasten, Tino Weinkauf, and Ingrid Hotz. Efficient computation of combinatorial feature flow fields. *IEEE Transactions on Visualization and Computer Graphics*, 18(9):1563–1573, 2012. 3

[24] Russ Rew and Glenn Davis. Netcdf: an interface for scientific data access. *IEEE computer graphics and applications*, 10(4):76–82, 1990. 7

[25] Irina Rudeva and Sergey K. Gulev. Climatology of cyclone size characteristics and their changes during the cyclone life cycle. *Monthly Weather Review*, 135(7):2568–2587, July 2007. ISSN 0027-0644. doi: 10.1175/MWR3420.1. 22

[26] T. A. Shaw, M. Baldwin, E. A. Barnes, R. Caballero, C. I. Garfinkel, Y.-T. Hwang, C. Li, P. A. O'Gorman, G. Rivire, I. R. Simpson, and A. Voigt. Storm track processes and the opposing influences of climate change. *Nature Geoscience*, 9(9):656–664, September 2016. doi: 10.1038/ngeo2783. 1

[27] Mark R. Sinclair. Objective identification of cyclones and their circulation intensity, and climatology. *Weather and Forecasting*, 12(3):595–612, September 1997. doi: 10.1175/1520-0434(1997)012⟨0595:OIOCAT⟩2.0.CO;2. 2, 32

[28] Primoz Skraba and Bei Wang. Interpreting feature tracking through the lens of robustness. In *Topological Methods in Data Analysis and Visualization III*, pages 19–37. Springer, 2014. 3

[29] B-S Sohn and Chandrajit Bajaj. Time-varying contour topology. *IEEE Transactions on Visualization and Computer Graphics*, 12(1):14–25, 2006. 3

[30] Robert R Sokal and F James Rohlf. The comparison of dendrograms by objective methods. *Taxon*, pages 33–40, 1962. 22

[31] Fujio Yamaguchi. *Curves and Surfaces in Computer Aided Geometric Design*. Springer Science & Business Media, 2012. 12, 26