The relationship between climate and mechanisms of tropical cyclone formation

Pavan Harika Raavi
ORCID: 0000-0002-6981-1157

Doctor of Philosophy

November 2020

School of Earth Sciences
The University of Melbourne
Australia

Submitted in total fulfilment of the requirements of the degree of Doctor of Philosophy

Abstract

A major uncertainty in future projections of tropical cyclone (TC) frequency is due to inadequate understanding of the atmospheric mechanisms leading to a reduction in the TC formation in warmer climates. Although the recently proposed "marsupial pouch theory," of TC formation indicates that a semi-enclosed recirculating region known as a "pouch" within large-scale disturbances provides necessary conditions for TC formation, it is important to link the frequency of the pouch environments to the evolving climate conditions. Therefore, this thesis examines the changes in the marsupial pouch TC formation environments and their relationship to the large-scale environmental conditions in the current climate and idealized warmer climates.

Here we use ERA-interim reanalysis data and high-resolution Australian Community Climate and Earth-System Simulator (ACCESS) climate model simulations of the current climate and idealized warmer climates (aqua-planet simulations) with two fundamentally different TC tracking schemes. The first scheme is the Okubo-Weiss Zeta Parameter (OWZP), a phenomenon-based tracking scheme that detects TC-favorable locations within marsupial pouches using resolution-independent thresholds. The other scheme is the Commonwealth Scientific and Industrial Research Organization (CSIRO), a traditional TC tracking scheme that uses resolution-dependent thresholds.

Environmental and structural composite analysis of developing and non-developing tropical depressions (TDs) 48 hours before TD formation, which are detected using the OWZP scheme in the reanalysis, showed that developing circulations have a strong protective layer around the core from the lower to mid troposphere that protects it from external disruptive influences. The relative importance of the environmental variables influencing tropical storm (TS) formation varies across ocean basins due to differences in the large-scale disturbances and surrounding environmental conditions. Statistical TS prediction schemes are also developed using the environmental conditions of developing and non-developing TDs. This analysis notes that random forests and support vector machine algorithms have higher accuracy than decision trees (DTs). Additionally, the hybrid approach of DTs and Markov decision process is proposed which indicates that developing TDs have a higher likelihood of remaining in more favorable environmental conditions than non-developing TDs.

In current climate simulation using the climate model, the TCs detected within marsupial pouches are a subset of the traditionally detected TCs. Also, the OWZP scheme performs better in representing the TC frequency statistics and has stronger relationships with the surrounding environmental conditions in the current climate than the CSIRO scheme. In idealized aquaplanet model simulations, we observe both reduced marsupial pouch environments and traditionally detected TCs with increasing sea surface temperatures (SSTs). The increased saturation deficit and increased stability explain the reduction in the frequency of formations with increased SST. The present study also notes a decrease in the frequency of low-intensity storms and an increase in the frequency of intense storms with increasing SSTs. In a drier and more stable atmosphere, the initial vortices need to be stronger for TC formation to occur, with higher low-deformation vorticity and higher upward mass flux. This research indicates that the marsupial pouch theory may be a fundamental paradigm of TC formation due to its better relationships with the large-scale environmental conditions in different climates.

Declaration

I. This thesis is my original work towards the completion of my Ph.D. except where indicated in the preface;

II. Due acknowledgment has been made in the text to all other material used;

III. The thesis is fewer than 100 000 words in length, exclusive of tables, maps, bibliographies and appendices or that the thesis is 44000 words as approved by the Research Higher Degrees Committee.

Raavi Pavan Harika

November 2020

Preface

This thesis is submitted as 'thesis with publications'. As such, the thesis is carried out under the supervision of Prof. Kevin Walsh. The nature and the contribution are described as follows:

- Prof. Kevin Walsh: Supervision and reviewing of the manuscripts of published/unpublished thesis chapters with about 20% of the contribution in each of the published/unpublished thesis chapters.
- The remainder of the work was carried out by the author and consists of her original work.

No work towards the thesis has been submitted for other qualifications.

No work towards the thesis has been carried out prior to the enrolment in the degree.

No third party editorial assistance was provided in the preparation of the thesis.

The publication status of all chapters presented in the article format is as follows:

- **Chapter 4:** Published by Journal of Geophysical research Atmospheres on 26/04/2020.
  Raavi, P.H. and Walsh, K.J.E. (2020b), Basinwise statistical analysis of factors limiting tropical storm formation from an initial tropical circulation. *Journal of Geophysical Research: Atmospheres*. (https://doi.org/10.1029/2019JD032006).

- **Chapter 5:** Unpublished material not submitted for publication.

- **Chapter 6:** Published by Earth and Space Science on 25/02/2020.

  Raavi, P.H., Walsh, K. J. (2020a), Sensitivity of tropical cyclone formation to resolution-dependent and independent tracking schemes in high-resolution climate model simulations. *Earth and Space Science*. (https://doi.org/10.1029/2019EA000906).

- **Chapter 7:** Unpublished material not submitted for publication.

The two published chapters (chapter 4 & 6) have been included as the corresponding chapters without any major changes to the main published text.

The author is grateful for financial support from Melbourne Research Scholarship for conducting Ph.D. research. The author is thankful for partial funding from the Australian Research Council Center of Excellence for Climate Extremes and the National Environmental Science Program.

Acknowledgements

First, I would like to show my sincere gratitude to my supervisor, Prof. Kevin Walsh, for his valuable guidance, encouragement, and support. My Ph.D. journey would have been impossible without the constant support and motivation of my supervisor. I am thankful for allowing me to research under his supervision with valuable advice on how to become an independent researcher.

I am extremely thankful to Dr. Kevin Tory, Bureau of Meteorology, Melbourne, for the valuable discussions on the current project. I am grateful for my co-supervisor, Prof. Prasad Kumar Bhaskaran, from the Indian Institute of Technology, Kharagpur, India, for supporting every step of the Ph.D. research. I am thankful to my other co-supervisors, Prof. Ian Simmonds and Dr. Claire Vincent, for their suggestions, comments, and encouragement during the Ph.D. journey. I thank my Ph.D. committee Chair, Prof. Peter Rayner, and Prof. Todd Lane, for their advice and support in these Ph.D. years. I am thankful to the Head of the School, Prof. David Phillips, for his help and for providing academic facilities.

I sincerely acknowledge the University of Melbourne for the funding of the Melbourne Research scholarship through the Melbourne-India Postgraduate Program to pursue my Ph.D. research. I am thankful to the Australian Research Council Center of Excellence for Climate Extremes and the National Environmental Science Program for climate change for their partial research funding.

My sincere thanks for the technical support of Scott Wales, Australian Research Council Center of Excellence for Climate Extremes team, and National Computational Infrastructure facilities team.

I want to thank Charlotte, Toni, Jamena, Katrina, and Bill for their support and discussions other than research, which bring a smile on my face during this research journey.

Heartfelt gratitude to Chirantan Bhagawati for his invaluable and constant support throughout my research work. I am also thankful for the support of my friends and colleagues, in particular, Manasa, Sushma, Suchita, Meghalim, Karishma, Sopia, Yohanna, Sandra, Apoorv, Dinesh, and Achyut.

I want to thank all other friends and lab mates, Roberta, Rob, Mahesh, Clemente, Eric, Ewan, Nasimah, Jaweria, Manohar, Aswini, Sindhu, Soumya, Priyanka, Sudhakar, Praveen, Radhika, Bhavana, and Neelima for helping me during this period.

I am thankful to my teacher Siva Rama Krishna for his constant encouragement from high school to pursue research career. I always remember nothing would have been accomplished without the care, constant support, and encouragement from my parents Linga Murthy Raavi, Sesharatnam Raavi, and my sister Deepika. They have been by the side throughout my research and encouraged me to pursue my interest through constant reassurance, love, and calming presence when I needed it most. I again express my heartfelt gratitude to my parents to whom I dedicate this thesis.

Table of contents

List of Figures

List of Tables

List of figures and tables in Appendices

Chapter 1. Introduction

1.1 Motivation

Tropical cyclones (TCs) are one of the most severe natural disasters that influence coastal communities across the world. TCs form under certain environmental conditions within different synoptic-scale disturbances across the global ocean basins. The changing climate conditions can further increase the severity of the socio-economic losses due to the enhanced TC intensity in a future climate (Holland & Bruyere, 2014; Knutson et al., 2019b). One primary concern in TC research is the ongoing debate on the mechanisms leading to changes in TC frequency in the future climate (Sugi et al., 2012; Emanuel, 2013). Recent studies on TC characteristics using advanced climate models reveal an increase in the TC intensity from the future global TC projections, whereas a decrease occurs in the frequency of formation, with uncertainties at a regional scale (Knutson et al., 2010; Walsh et al., 2016; Walsh et al., 2019; Knutson et al., 2019b). In contrast, the genesis potential indices derived based on the best statistical fit between the frequency of formation and the environmental variables do not reveal a decrease in the TC frequency (Camargo et al., 2014). This disagreement between the statistically derived indices and the numerical models shows a lack of quantitative understanding of the relationship between large-scale environmental variables and TC formation.

However, a consensus exists between the model projections of the increased TC intensity in the future climate based on the maximum potential intensity theory (Emanuel, 1986, 1988). In contrast, although most of the modeling studies project reduced TC frequency in a future climate, there is no climate theory of TC formation, which could quantify the TC frequency under different climates (Walsh et al., 2016). Nevertheless, there are two hypotheses explaining the reduction in TC frequency: increased stability of the atmosphere leading to reduced upward mass flux (Sugi et al., 2002; Held & Zhao, 2011; Sugi et al., 2012) and increased saturation deficit of the lower troposphere, which is a variation between saturation-related moisture and actual moisture (Emanuel et al., 2008; Emanuel, 2010). However, there is no precise mechanism that can quantitatively explain the reduction in the TC frequency under warming conditions. The increased intensity of tropical storms and associated coastal risks in the future conveys the need for a better understanding of the changes in the TC formation and its relationship to large-scale climate. Therefore, in the current research, we focus on understanding the relationship between TC formation and environmental conditions using observation and high-resolution climate model simulations in the current climate and future idealized warmer climates.

Significant advances have been previously achieved in understanding the TC formation mechanism. For instance, conceptual models of observed TCs have recently been aided by the marsupial pouch theory of TC formation (Dunkterton et al., 2009). This theory states that a closed recirculating region called a pouch within the large-scale disturbances (tropical waves) in the Atlantic and the Pacific Ocean basins shelters the incipient circulation from disruptive influences and provide necessary conditions for its development. A specific location within this protective pouch region of tropical waves acts as the sweet spot for cyclone formation. Another study by Wang et al. (2010) hypothesized that a deep pouch from the lower troposphere to the middle troposphere is important for TC formation. Based on this marsupial pouch paradigm, Tory et al. (2013a, b) developed a phenomenon-based TC detection and tracking scheme which detects the locations within the pouch that have potential for TC formation. They also noted that there exists a close resemblance between the observed TCs and the detected pouch locations. Therefore, any climate theory of TC formation likely needs to take into account the frequency of the pouch formation environments. Also, the relationship between the large-scale environmental variables and the pouch formations using the state-of-art global circulation models may improve our current understanding of the relationships between TCs and environmental conditions.

1.2 Aim

Taking note of the discrepancies in future projections of TC frequency, the present study examines the relationship between environmental variables and marsupial pouch formations in reanalysis data and state-of-art high-resolution global circulation model simulations run with different model climates. To achieve this, we use two fundamentally different TC detection and tracking schemes, that is, phenomenon-based (resolution-independent) and traditional (resolution-dependent) schemes. Also, the study attempts to understand the atmospheric conditions leading to a reduction in the TC frequency with increasing sea surface temperatures in aqua-planet simulations. Furthermore, the study examines whether marsupial pouch theory acts as a fundamental paradigm for TC formation.

1.2.1 Research Questions

The primary research questions addressed in this study are as follows:

1. There is a good relationship between TCs detected within marsupial pouches and observed tropical cyclone formation in the current climate (Tory et al., 2013b). Do all the vortices within pouches develop into tropical cyclones? What are the significant influencing environmental variables for developing cases?

2. Do the frequency statistics of TCs detected within marsupial pouches resembles traditionally detected TCs in a high-resolution climate model simulation in the current climate?

3. Do the detected TCs within marsupial pouches have significant relationships with large-scale environmental conditions in the climate model?

4. Do the marsupial pouches have the same relationship to large-scale environmental variables in other model experiments run with different climates?

5. Do the pouches have a better relationship with environmental variables than tropical cyclones detected by traditional tracking schemes in other model climates?

1.2.2 Objectives

The main objectives of the project are as follows:

1. To identify the environmental factors limiting tropical storm formation from an initial tropical depression stage vortex.

2. Application of different machine learning algorithms to predict developing tropical depressions with background environmental conditions as the predictors.

3. Compare the frequency statistics of TCs identified within marsupial pouch environments using a resolution-independent tracking scheme (phenomenon-based) with the frequency of the TCs detected by a resolution-dependent tracking scheme (traditional) in a high-resolution GCM simulation of the current climate. Also, to identify the relationships of the detected TCs with the large-scale environmental variables.

4. To examine the relationship of marsupial pouch environments with traditionally detected TCs and environmental variables using idealized GCM simulations run with different values of sea surface temperatures.

1.3 Thesis Outline

In the present thesis, the research work is presented in different Chapters as follows:

Chapter 1: Introduction

This chapter gives a general introduction and motivation for the Ph.D. project, followed by the aim of the study, which includes the research questions and objectives, the thesis outline describing each of the thesis chapters briefly, and finally the contribution of the thesis.

Chapter 2: Literature review

This chapter gives a general review of tropical cyclones (TCs) and favorable environment, TC formation mechanisms, and the relationship between the climate and TCs from both observational and numerical modeling studies. Additionally, a separate and more specific introduction is given at the start of each chapter.

Chapter 3: Data and Methods

This chapter gives a general overview of the general circulation models, description of climate model specifications, different observational datasets, and tracking schemes used in the current thesis.

Chapter 4: Basinwise statistical analysis of factors limiting tropical storm formation from an initial tropical circulation

This chapter uses a phenomenon-based tracking scheme to detect the developing and non-developing circulations. A storm-centered statistical analysis identifies the limiting factors for TC formation in each ocean basin as well as the structural differences between developing and non-developing circulations. This chapter has been published in the Journal of Geophysical Research: Atmospheres.

Chapter 5: Prediction of tropical storm formation from initial tropical depressions using machine learning algorithms

This chapter describes the results obtained by comparing three machine learning techniques, decision trees, support vector machines and random forests, to classify the developing and non-developing tropical circulations and predict the developing cases based on the background environmental variables as the predictors.

Chapter 6: Sensitivity of tropical cyclone formation to resolution-dependent and independent tracking schemes in high-resolution climate model simulations

This chapter compares the performance of a traditional TC detection and tracking scheme with a phenomenon-based tracking scheme based on TC frequency statistics and their relationships to large-scale environmental variables. The traditional scheme detects the circulations that resemble the observed TCs using resolution-dependent thresholds of vorticity and 10m wind speed. In contrast, the phenomenon-based tracking scheme detects the locations within the marsupial pouches that have TC formation potential using resolution-independent thresholds of different large-scale environmental variables. This chapter has been published in Earth and Space Science.

Chapter 7: Response of tropical cyclone frequency to sea surface temperatures in aqua-planet simulations

This chapter describes the response of TCs to increased sea surface temperatures detected using both traditional and phenomenon-based tracking schemes in different idealized model experiments. It also examines different environmental conditions to identify the contributing variables leading to the changes in the TC frequency.

Chapter 8: Concluding remarks and future work

This chapter gives the overall concluding remarks from all the chapters and summarize the results and gives the future directions of the current research.

1.4 Scientific contribution of the thesis

The increased intensities of the TCs in the future climate impose a further increase in the risk associated with storm surge in coastal cities and ports on the background of rising sea levels due to global warming. Some potential implications due to changes in TC activity include changes in the design standards of the coastal infrastructure, ports, and harbors to account for the incidence of storm surges and damaging winds. Thus studies on understanding the relationship between environmental conditions and TC formation improve our confidence in the future TC frequency projections, which are useful for better climate risk and adaptation strategies. The present thesis depicts the environmental factors limiting TC formation across different ocean basins by examining the developing and non-developing tropical circulations identified within the marsupial pouch environments. The research work also provides a forecasting approach for the prediction of developing tropical depressions from pre-genesis environmental conditions 48 hours before the tropical depression declaration. The study compares the performance of a phenomenon-based TC detection and tracking scheme and a traditional TC tracking scheme in diagnosing the TC frequency characteristics simulated using high-resolution climate model simulations in the current climate. The study notes that the phenomenon-based TC tracking scheme is a better performing scheme (a resolution-independent scheme) that can be applied to different resolution models compared to the traditional TC detection scheme (a resolution-dependent scheme). The TCs have close relationships to the large-scale environmental variables in the current climate using the global climate model simulation. Finally, the study explores the response to warmer sea surface temperatures of global TC genesis frequency and the contributing environmental conditions using aqua-planet experiments. The study notes that increased stability and saturation deficit leads to reduced formation rates of both initial tropical circulations and TCs. This project, therefore, improves our understanding of the relationships between climate and tropical cyclone formation. Based on the present findings, the marsupial pouch theory acts as a fundamental paradigm of TC formation in other climates. In summary, the study paves the way for developing more efficient TC forecast schemes in the present as well as future climate through a better understanding of the role of environmental variables on TC formation within the marsupial pouch environments

Chapter 2. Literature review

2.1 Introduction

This chapter gives a brief introduction of the tropical cyclones and its impacts. The literature describing different theories of the cyclone formation, favorable conditions for the formation, different global climate modeling studies on the TC frequency in the current and future climate and different hypothesis explaining the changes in the frequency in the future climate using idealized climate modeling experiments.

2.1.1 Tropical cyclones

Tropical cyclones (TC) are intense vortices with rotating, organized cloud clusters that originate over the warm tropical oceans or in the subtropical regions across the globe. TCs are characterized by low atmospheric pressure, strong winds, and heavy rain. The life cycle of a TC involves different stages based on the surface wind speeds (a measure of intensity) as shown in Figure 2.1: (i) a tropical disturbance stage where there is slight circulation in the lower levels and wind speeds are less than 20 knots; (ii) a tropical depression stage where the surface wind speeds are between 20-33 knots with at least one closed isobar; (iii) a tropical storm stage when the wind speeds are between 33-64 knots, a further drop in the central pressure and an organized vertical structure; and (iv) a hurricane/severe tropical cyclone stage where the surface wind speeds are higher than 64 knots with a further organized vertical structure, as shown in the schematic below (Plutchak et al., 2010). In tropical cyclone forecasting, either the tropical depression stage or a tropical storm stage may be considered as the initial stage (i.e., genesis) (Tory & Montgomery, 2006). In different global ocean basins, different terminologies are used for the mature storms. They are called hurricanes in the Atlantic and eastern North Pacific basins, typhoons in the western North Pacific, and severe tropical cyclones in the rest of the ocean basins.

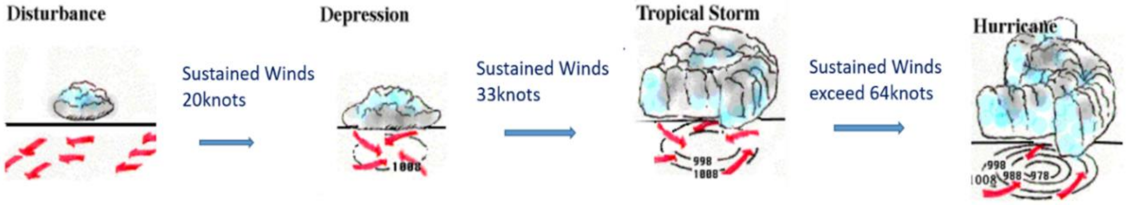

Figure 2. 1. Stages of tropical cyclone formation based on the measure of surface wind speeds (modified after Plutchak et al., 2010)

2.1.2 Impacts of tropical cyclones

TCs are natural catastrophic events causing tremendous losses to both life and property across the globe. Every year around 80-90 cyclones form affecting different countries across the world through high impact winds, extreme precipitation, tornadoes, and inundation due to storm surges. Of those TCs, Hurricane Katrina in August 2005 in the North Atlantic (NA) basin was one of the costliest storms causing losses to infrastructure and killing at least 1833 people (Sampe & Xie, 2007). The recent Hurricane Harvey in August 2017 in the NA basin also caused severe socio-economic losses. Furthermore, Cyclone Nargis (2008) that formed in the Bay of Bengal caused high fatalities of at least 138,000 in Myanmar. The most significant number of deaths associated with tropical cyclones are with Cyclone Bhola of 1970 that killed at least 300,000 people in the low lying areas of Bangladesh due to the extreme storm surge (Holland, 1993).

In recent times, there is evidence of a global increase in the intense storms (Holland and Bruyere, 2014) and the exceedance probability of intense tropical cyclones (TCs) has been considerably increasing in the observations for the period 1979-2017, with a statistically significant increase of 2-15% per decade (Kossin et al., 2020). This phenomenon is strongly associated with the sea surface temperature suggesting that climate change and TCs are strongly linked. The intrinsic links of TCs with the global heat budget (Emanuel, 2001; Sriver & Huber, 2007; Sriver et al., 2010; Scoccimarro et al., 2011; Manucharyan et al., 2011) as well as their disastrous impacts in coastal regions, suggest a better theoretical understanding of the relationship between climate and TC formation contributes to the efforts for effective disaster management and climate adaptation policies. Additionally, the changes to the TC activity in the future climate, both the intensity and the frequency of formations are therefore important for the insurance industry when estimating the future risk of changing climate conditions (Dailey et al., 2009; Bakkensen & Mendelsohn, 2016; Klettner et al., 2019).

2.2 Tropical cyclone formation

Tropical cyclone formation is a multiscale interaction between organized clusters of deep convection within the trough region of large scale disturbances and individual mesoscale convective systems within these clusters. The genesis stage of a TC is the least understood part of the tropical cyclogenesis process due to fewer in-situ observations and its multiscale nature (Emanuel, 2005). The formation of a tropical cyclone within large-scale tropical disturbances involves two stages. The first stage is the formation of the favorable environment surrounding the "large scale disturbance," and the second stage, "inner-core formation" or "intensification stage," describes the transition of a disturbance to a low-intense depression stage vortex or a tropical storm (Ooyama, 1982; Tory & Frank, 2010). The following sections briefly elaborate on these stages.

2.2.1 Large Scale disturbances

The first stage of TC formation involves preconditioning of the environment surrounding large scale disturbances so that it becomes conductive for TCs. Earlier studies noted that around 80-90% of the storms form within 20 degrees of the equator near the region of the monsoon trough at the Inter-Tropical Convergence Zone (ITCZ) and in regions with an appropriately placed upper-level trough (Gray, 1968; Briegel and Frank, 1997; Ferrieria and Schubert, 1997; Wang & Magnusdottir, 2006). Many individual wave-troughs occur in Eastern and Central Atlantic. Although, Frank (1970) notes that around 20% of the storms originate from African easterly waves (AEW) in the Atlantic basin, Landsea (1993) suggests that more than 20% of the storms originate from AEWs. Other types of large-scale disturbances favorable to TC formation have been established over the global ocean basins in recent decades, as shown in Figure 2.2.

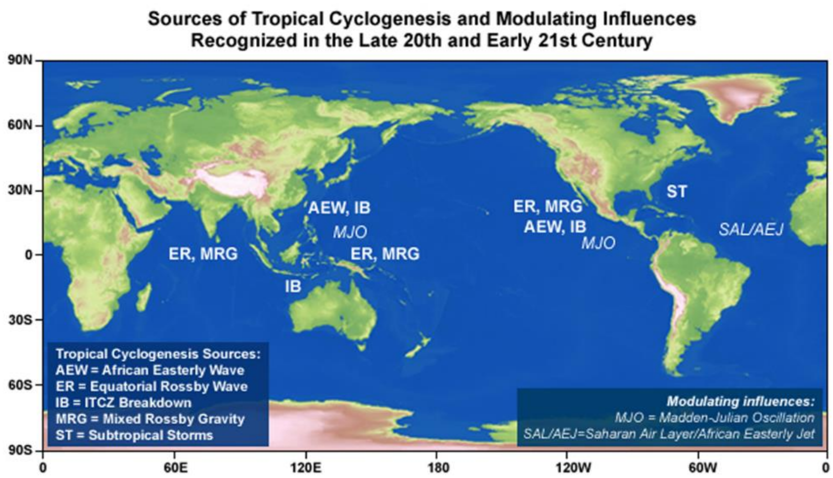

Figure 2. 2. Different sources of tropical cyclogenesis and their modulating influences across global ocean basins (modified after Laing & Evans, 2011)

Large scale disturbances include AEW in the North Atlantic region (Frank, 1970; Landsea, 1993; Thorncroft & Hodges, 2001), mixed Rossby gravity waves (MRG) in the North Pacific region and the equatorial Rossby (ER) waves and MRG waves in the south Indian ocean and Australian region (Dickinson and Molinari, 2002; Bessafi & Wheeler, 2006; Frank & Roundy, 2006). A recent study by Russell et al. (2017) note that around 61% of the TCs in the NA basin originate from AEWs by using eddy kinetic energy associated with the AEWs to account for their activity. These large-scale environmental disturbances provide an initial favorable environment for TC formation in the form of enhanced values of lower-tropospheric cyclonic vorticity, which favors deep convection from which TCs form. The next stage is forming a core region within these large-scale disturbances, a favorable region for the development of a cyclonic strength vortex.

2.2.2 TC core Formation Theories

The development of a cyclone strength vortex on the ocean surface within the large scale (synoptic-scale) disturbances takes place with cooperative interactions between large-scale disturbances (synoptic) and mesoscale convective systems (Figure 2.3). A mesoscale convective system (MCS) is an organized multicellular convective system with a convective and a large stratiform cloud shield from the mid to upper troposphere with a horizontal scale of 20-1000 km (Houze, 2004). Tropical convection is often organized into an MCS and is important for transporting heat, moisture, and momentum. It has been theorized that when successive cycles of convection occur over warm ocean waters, MCSs play a crucial role in the transition of weakly organized convection into tropical depressions under favorable large scale environmental conditions (Houze, 2004). These MCSs develop into cyclonic circulations when they move over warm tropical waters. Although many MCSs develop in a hurricane season, only a few develop into tropical storms due to variations in environmental conditions. From an observational study using satellite images of cloud clusters, McBride and Zehr (1981) have suggested that the cloud clusters develop into tropical storms under certain conditions: (a) a warm core in the upper level with more warming in the developing circulations than the non-developing circulations and (b) higher values of low-level vorticity in developing cases than non-developing cases. These convective systems produce mesoscale vortices in the mid-level without a surface circulation, which may be the reason for the small fraction of these disturbances that become depressions (Harr and Elsbery, 1996).

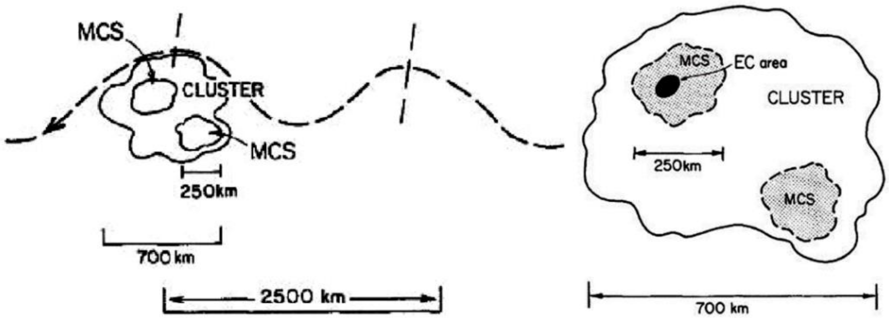

Figure 2. 3. Graphical representation of synoptic-scale flow in the form of a large-scale disturbance with a cluster of convection within the trough of the wave. The cluster contains mesoscale convective systems (MCSs) and extreme convection (EC) (modified after Gray, 1998)

The core formation of a TC occurs at the mesoscale (i.e., with a horizontal scale of 20-1000 km) within the initial large scale disturbance. Different theories explain this core formation from an MCS based on observational datasets obtained from a few field campaigns and idealized numerical experiments with hypotheses focussing on the organization of a mid-level vortex into a surface-based TC like circulation. The following section describes the gradually improving understanding of the TC core formation mechanisms starting from the top-down theory (with two categories viz. top-down merger and top-down showerhead), followed by the bottom-up theory to the recently developed new paradigm, the marsupial pouch theory.

(a) Top-down Merger

The first theory, "top-down merger," emphasizes the importance of two mid-level vortices within a stratiform region of an MCS for TC formation. This idea is based on both the observational data from the Tropical Cyclone Motion (TCM-92) field experiment (Dunnavan et al. 1992) and idealized modeling experiments on the mesoscale vortex interactions within the large scale environment (Ritchie and Holland, 1997). This theory hypothesizes that the strong interaction between two or more mesoscale convective vortices (MCVs) on scales of 100-200 km at mid-levels results in a merger process. They create a more intense and larger MCV with a higher penetration depth to the top of a boundary layer, forming a cyclone scale low-level vortex. The numerical experiments indicate that the background rotation makes the merger more efficient. The vortex merger with a background rotation results in a significant increase of vortex depth in vertical layers, consistent with the Rossby penetration depth (Ritchie and Holland, 1997).

(b) Top-down shower head

The second type of the first theory of TC core formation within an MCS is called "top-down showerhead." It emphasizes thermodynamical processes with a single mid-level vortex in spinning up the low-level vortex (Figure 2.4). This theory hypothesizes that the sustained precipitation in the stratiform region of an MCS or MCV and evaporation of the rain droplets will saturate the lower troposphere to create an environment favorable for TC formation. This idea has been tested using a non-hydrostatic model and from observational data of a field experiment, the Tropical Experiment in Mexico (TEXMEX) (Bister & Emanuel, 1997). The slantwise flow within the stratiform region of an MCS advects the cyclonic vorticity from mid-levels to the ocean's surface, subsequently increasing surface winds and the moisture fluxes. Moistening and cooling due to stratiform precipitation destabilize the boundary layer which supports deep convection and creates low-level convergence and stretching of vorticity, thus raising low-level winds (Bister and Emanuel, 1997). A criticism of this theory was made by Tory and Montgomery (2006). They have pointed out that absolute circulation cannot spin up the low-level circulation due to the impermeability of the vorticity between isobaric surfaces (Haynes and McIntrye, 1987).

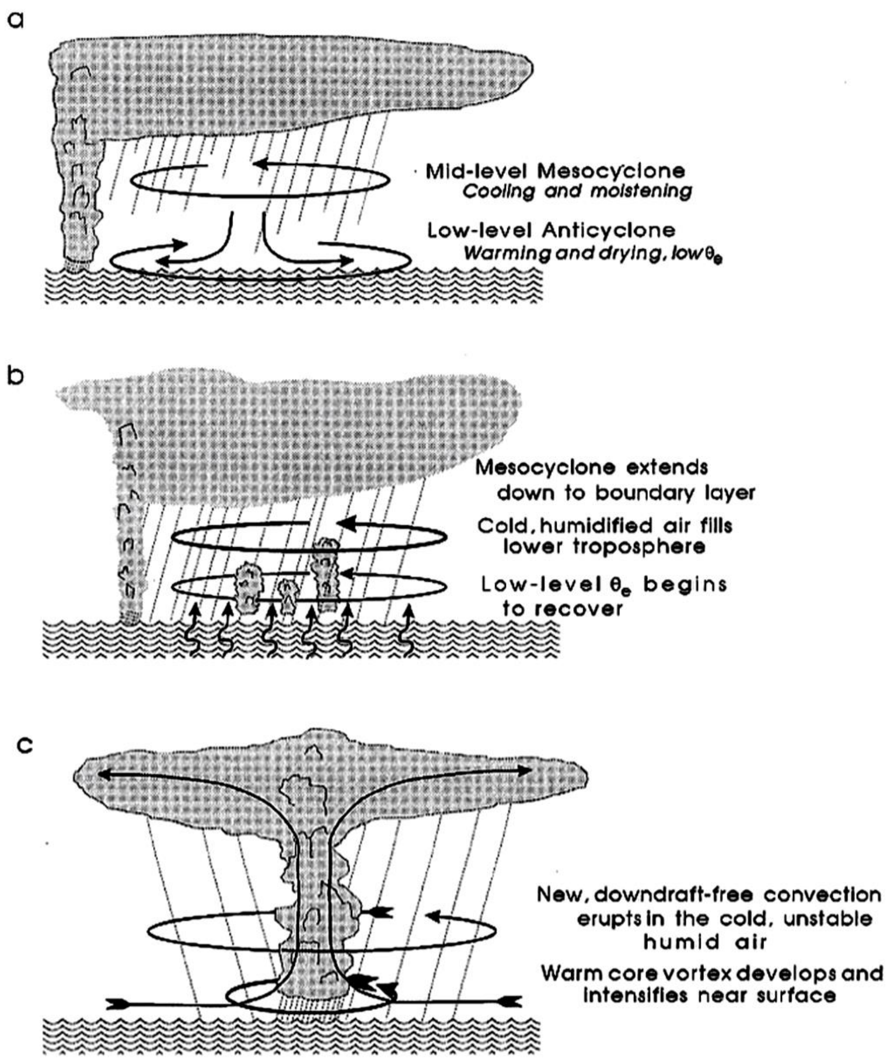

Figure 2. 4. Schematic of top-down showerhead mechanism of tropical cyclone formation from Bister and Emanuel (1987).

(c) Bottom-up Theory

The second theory of a TC formation within an MCS is the “bottom-up theory," which emphasizes the role of the near-surface vorticity. In this theory, the vorticity enhancement at the low levels occurs through the formation of multiple small scale vertical hot towers (VHT) with a heating source within the convective region leading to the formation of vortices, which finally merge to form a cyclone scale vortex (Hendricks et al. 2004). Some idealized experiments on the VHTs within an MCS using numerical models by Montgomery et al. (2006) suggest that merging of the VHTs leads to the formation of a low-level vortex with increased size and intensity. This merged vortex acts as the seed for TC genesis. The important point to note in this theory is that the initial non-alignment of the lower level vortex with the mid-level vortex later becomes vertically aligned to form a deep convective vortex. A more recent mechanism explaining the formation of a low-level vortex within a large scale disturbance, thereby developing into a TC under favorable conditions, is explained by the "marsupial pouch theory."

(d) Marsupial pouch Theory

Dunkerton et al. (2009) have proposed a new theory of TC formation, the "marsupial pouch theory." This theory hypothesizes that a closed recirculating region, a “pouch” (or Kelvin cat’s eye) within an easterly wave, provides favorable conditions for TC formation. They note that TC genesis initiates in a region near the intersection of the trough of a westward-moving large scale disturbance (like tropical waves) and the critical latitude of the wave within some of the easterly waves over the Atlantic and Pacific basins, based on the survey of the storms in 1998-2001 using observational datasets. The critical latitude is the location in the large-scale disturbance at which the wave phase speed matches the speed of the background flow. The intersecting point is the sweet spot for TC formation. A semi-enclosed recirculating region around it known as a pouch, as shown in Figure 2.5, provides sufficient low-level cyclonic vorticity with weaker deformation forces through breaking of waves or rolling up of the vorticity and surrounding moisture. It also provides a recirculating region with a continuously moist environment where the lateral intrusion of dry air is inhibited, thereby providing necessary conditions for deep convection, which enhances the cyclonic vorticity. Also, the mesoscale vortices inside the Kelvin cat's eye maintain or enhance the parent wave disturbance. Therefore, sustained convection within the trough of a wave and vortex interaction within a critical layer of the wave subsequently leads to the formation of a self-sustaining vortex with deep convection. This marsupial paradigm represents the bottom-up development of TC formation. Dunkerton et al. (2009) also note that the top-down and bottom-up development of a TC formation is not contrasting pathways; instead, they co-exist on different scales.

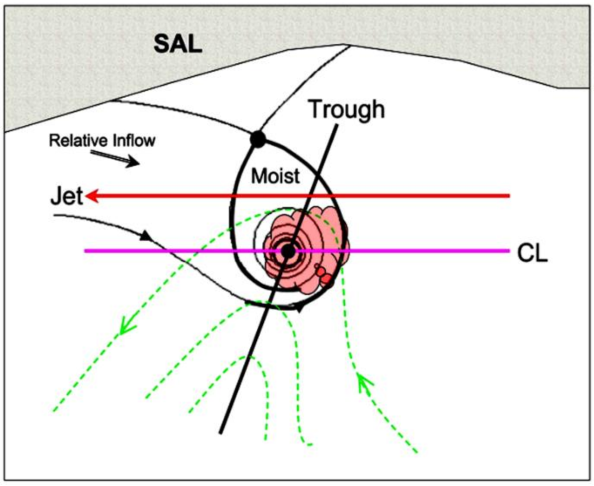

Figure 2. 5. Schematic of the marsupial pouch (solid black curve) of easterly waves. The green dotted lines indicate the streamlines of an easterly wave in an Earth frame of reference. Here the purple line CL denotes the critical latitude, and the intersecting point of the CL with the trough axis (black line) acts as a sweet spot for TC formation. The pink shading indicates persistent convection (modified after Wang et al. (2010a)).

A further examination of the pouch's structure to test the marsupial pouch hypothesis is carried out using the observations from the field experiments of Tropical Cyclone Structure 2008 (TCS-08; Montgomery et al., 2010; Raymond & Lopez Carrillo, 2010) and PRE-Depression Investigation of Cloud Systems in the Tropics (PREDICT; Montgomery et al., 2012). The marsupial pouch theory is also tested for a real storm case using high-resolution idealized simulations (Wang et al., 2010 a, b; Montgomery et al., 2010). They have noted that rotating deep convection within the pouch region enhances the low-level vorticity, and the convergence in the middle troposphere increases the vorticity in the mid-levels, thereby creating a deep vortex. These numerical studies with different horizontal scales (meso-beta scale: bottom-up; meso-alpha scale: top-down) have analyzed the marsupial paradigm also note that both TC development pathways co-exist. The later studies have developed a pouch identification technique by tracking the initial tropical disturbances (Wang et al., 2009; Montgomery et al., 2010). Wang et al. (2012) have proposed that TC formation from initial disturbances requires a deep pouch from the mid-troposphere (600-700 hPa) to the top of the boundary layer. Tory et al. 2013(a) have developed a TC detection scheme that identifies the locations within pouch that seem to have potential for TC formation using a storm characteristic parameter, the product of standardised Okubo-Weiss parameter and absolute vorticity (denoted OWZP), with certain thermodynamic and dynamic conditions. They have noted that

95% of the observed TCs have increased values of OWZ at lower and mid-troposphere in the ERA-Interim reanalysis data. The marsupial pouch theory better explains TC formation mechanism within a large-scale disturbance. As the marsupial pouches accurately represent the observed TCs (Tory et al., 2013a, b, c), it is important to understand the relationship between marsupial pouches and different large-scale environmental conditions using both statistical and different idealized model experiments. Also, application of the OWZP tracking scheme which identifies potential locations for TCs within marsupial pouches not only helps to detect the developed TCs accurately but also detect the initial tropical depressions (Tory et al. 2017). The study of the environmental conditions of the initial tropical depressions helps us to identify important predictors and limiting factors for TC formation. Therefore, the current Ph.D. thesis focuses on this marsupial pouch theory of formation to understand the relationship between large-scale climate and the TC formation.

2.3 TC formation and Favourable Environment using observational datasets

TCs form under specific favorable large-scale thermodynamical and dynamical conditions, which are observed to be necessary but not sufficient (Gray, 1968; 1975; Tuleya & Kurihara, 1981; Tory & Frank, 2010; Camargo et al., 2010). These conditions are: (a) sea surface temperature (SST) values above 26.5℃ with a mixed layer depth of 50m, (b) larger values of low-level relative vorticity, (c) a deep tropospheric layer of conditional instability, (d) weak to moderate vertical wind shear (in the easterly direction), (e) a moist mid-troposphere and deep organized convection and (f) a non-zero value of Coriolis parameter (i.e., at least 5 degrees away from the equator).

Recent studies reveal that TC potential intensity is an important parameter which takes larger values in relatively warmer tropics and that there is no absolute SST threshold for TC formation. This has been shown in climate modeling simulations of much colder climates (Last Glacial Maximum; Korty et al. 2012) and of much warmer climates (much much higher CO_2; Korty et al. 2017), as well as in f-plane simulations (Cronin and Chavas 2019) and single column models imposing Weak Temperature Gradient approximation (Ramsay and Sobel 2011). The aforementioned 26.5°C threshold of SST is simply an approximate one that applies for the current climate. The TC potential intensity may be influenced by the sensible and latent heat flux which lead to changes in the static stability and vertical wind shear of the atmosphere. These changes are due to variations in the vertical potential temperature and wind profile by turbulent heat, momentum fluxes and heat release due to phase changes between solids, liquids, and gases. The stability of the atmosphere affects vertical transport, thereby influencing the upward mass flux. A highly stable atmosphere has lower upward mass flux values and negative sensible heat flux unfavorable for TC formation. The changes in atmospheric moisture content through processes of evaporation and precipitation contribute to changes in the saturation deficit. A higher saturation deficit is an unfavorable condition for TC formation.

Globally around 80-90 cyclones form every year over the tropical oceans, but there is no significant trend in the annual numbers (Gray, 1968; Emanuel and Nolan, 2004; Klotzbach, 2006). The number of TCs in individual basins exhibits variability at different time scales, such as seasonal, interannual, and decadal scales. These variations in the TC frequency are often associated with the changes in the large-scale thermodynamical and dynamical environmental conditions due to natural climate variability and possible anthropogenic effects. Natural causes may be due to significant dominating phenomena such as the El Niño-Southern Oscillation (ENSO) phenomenon and the Madden-Julian Oscillation (MJO) (Camargo et al., 2007; Frank & Young, 2007; Camargo et al., 2010; Bell et al., 2014; Patricola et al., 2018; Wu et al., 2018). At the same time, anthropogenic influences lead to an increase in global surface temperatures, influencing the sea surface temperatures, atmospheric circulations, and other large-scale variables which further influences the TC formation (Vecchi & Soden, 2007a; Johnson & Xie, 2010; Held & Zhao, 2011; Sugi et al., 2012; Murakami et al., 2013b; Patricola et al., 2014, 2016; Walsh et al., 2015; Knutson et al., 2019b). The future changes in the ENSO phenomenon due to changing temperatures may influence the convection and the TC activity (Chand et al., 2017; Williams & Patricola, 2018). Recent studies have observed a poleward migration in the TC genesis locations (Daloz & Camargo, 2018) due to changes in the atmospheric circulations and the climate variables (Studholme & Gulev, 2018; Sharmila & Walsh, 2018).

In the NA basin, a large number of tropical disturbances emerge each year, but only a few develop into tropical cyclones due to variations in the surrounding conditions (Avila, 1992; Avila et al., 2000). Thorncroft & Hodges (2001) noted a correlation of TC frequency with AEW number. Hopsch et al. (2007) developed the work of the Thorncroft & Hodges (2001) and noted that the interannual variation of the AEWs is not correlated with the TC frequency changes. Another study by Agudelo et al. (2011) noted that around 14% of AEWs develop into TCs. The regional differences in the environmental conditions surrounding large-scale disturbances may lead to differences in the TC frequency statistics. Therefore, the current study investigates the environmental factors influencing the TC formation from initial circulations across the global ocean basins, as discussed in detail in chapter 4. Various statistical schemes have developed to differentiate developing and non-developing cases using different observational and forecast analysis datasets. These studies have identified the prominent environmental variables leading to TC formation in the Pacific and Atlantic basins (Perrone & Lowe, 1986; Hennon & Hobgood, 2003; Schumacher et al., 2009; Hopsch et al., 2010; Chand & Walsh, 2011; Fu et al., 2012). Chapter 5 estimates the relative importance of different environmental variables for TC formation in global ocean basins using advanced data mining techniques.

2.4 TC characteristics using General Circulation Models (GCMs)

Climate models are important for understanding the relationship between large-scale climate and various TC features. The three main approaches to infer TC activity by using GCM are as follows: (i) direct simulations of the TCs, here models simulate the large-scale environmental variables, and TCs are identified with the help of different TC detection and tracking schemes; (ii) The dynamical downscaling approach involves embedding a high-resolution model within a global model to simulate the TC characteristics; (iii) the statistically derived potential genesis indices measure the frequency of storms based on the background environmental conditions. Numerous studies have analyzed TC activity using different high-resolution climate models both globally and regionally. Several models can simulate TC formation rate, geographical distribution and seasonal variability at close to the observations in most of the basins (Zhao et al., 2009; Knutson et al., 2010; Walsh et al., 2013; Bell et al., 2013; Shaevitz et al., 2014; Walsh et al., 2015; Camargo & Wing, 2016; Walsh et al., 2019; Sharmila et al., 2019; Raavi & Walsh 2020a). Some of the models reasonably simulate the other TC characteristics such as landfall activity, ENSO induced changes, TC size, intensity, and the rainfall rates (Zarzycki & Jablonowski 2014; Knutson et al., 2015; Krishnamurthy et al., 2016; Lok & Chan, 2017; Bhatia et al., 2018; Liu et al., 2018; Schenkel et al., 2018). Murakami et al. (2012) have examined the role of model physics and sea surface temperature (SST) variations in global and regional variations of TC annual numbers using 60km horizontal resolution atmospheric GCM simulations. They have suggested that there are three significant sources of uncertainty in the projection of TC frequency: (1) model horizontal resolution, (2) model physics, and (3) changes in SST patterns. They have also noted that projections of future SST changes are causing significant future changes in TC frequency rather than the choice of convective parameterization. A recent study by Camargo et al. (2020) also notes that the climatological TC activity in the global climate models largely depends on model physics, dynamical core and resolution.

Most of the recent climate models continue to predict the reduced global number of TCs in a warmer climate, especially in the Southern Hemisphere (Sugi et al., 2012; Tory et al., 2013c, d; Walsh et al., 2016; Knutson et al., 2019b). However, regional uncertainties exist in the frequency of the formations (Knutson et al., 2010; Park et al., 2017; Knutson et al., 2019b). The TC detections using climate model projections with different TC detection and tracking schemes produce specific variations in the TC frequency due to a different choice of thresholds for wind speeds, formation latitudes, and storm duration (Horn et al. 2014). However, most of the Coupled Model Intercomparison Project Phase 3 (CMIP3; Meehl et al., 2007) and Phase 5 (CMIP5; Taylor et al., 2012) models project this reduction in TC frequency (Camargo, 2013; Tory et al., 2013c, d; Knutson et al., 2013; Murakami et al., 2014a). An exception to these studies of reduced TC frequency projections is the downscaling study by Emanuel (2013) in which the TC numbers are found to be increasing in a warmer climate when the downscaling technique is applied to CMIP5 model output, compared to results from a similar downscaling technique applied to CMIP3 models resulting in a reduction in TC frequency.

However, there have been uncertainties in both numbers and intensity of TC projections at regional scales using the CMIP5 models (Kossin et al., 2016; Nakamura et al., 2017; Knutson et al., 2019b; Bell et al., 2019a, b, c). From the above studies, we can infer that there is some inconsistency in TC frequency projections from modeling studies due to variations in downscaling methodologies, model physics, inconsistent projections of large-scale conditions, and TC tracking algorithms (Walsh et al., 2016). Therefore, examining the variations in TC frequency statistics and their relationships to the large-scale environment using fundamentally different tracking schemes in the current state-of-the-art high-resolution climate models helps to identify a more reliable tracking scheme. Therefore, Chapter 6 estimates the TC frequency statistics and their relationship to environmental conditions using two fundamentally different tracking schemes in a high-resolution atmospheric climate model simulation.

2.5 TC characteristics using idealized GCM simulations

Some recent studies using GCM simulations predict a decrease in the global frequency of TC formations in a future climate, with regional variations (Knutson et al., 2019b). In contrast, Vecchi et al. (2019) found an increase in global TC frequency with increasing resolution under isolated SST warming experiments and decrease in the frequency with CO2 doubling using coupled GCM simulations. They also note that there is an increase in the frequency of formations with increasing model resolution. The decrease in TC frequency is linked to changes in the mean vertical circulation (Held & Soden, 2006; Held and Zhao, 2011; Zhao and Held, 2012). The increased mass flux favors sustained deep convection favoring TC genesis. Therefore reduced upward mass flux is an unfavorable condition for TC formation, which decreases the frequency of formations with increasing temperatures. However, the quantitative link between the mean vertical circulations and TC formation is not well understood. There are various studies on different sets of numerical experiments using global climate models to understand further the mechanisms leading to a reduction in TC frequency. Some of these studies include idealized experiments such as an SST increase by 2K and a doubling of CO_2 concentration (Sugi et al. 2002; Held and Zhao, 2011; Sugi et al. 2012; Sugi et al., 2015; Walsh et al., 2015). These studies suggest that an upward mass flux hypothesis that is the decrease in the upward mass flux at mid-troposphere leads to a reduction in TC frequency. Sugi et al. (2012) have proposed two mechanisms explaining the reduced mass flux in future climate: an increase in the dry static stability of the atmosphere in increased SST experiments and reduced precipitation in doubling CO_2 experiments. Satoh and Yamada (2015) have proposed that the increase in TC strength combined with the constraint for total upward mass flux leads to a decrease in TC frequency.

In other studies, increased tropospheric moisture deficit explains the reduction in the frequency of TC formation (Emanuel et al., 2008; Emanuel, 2010; Rappin et al., 2010). The increased moisture content in the troposphere favors deep convection and the increase of strength of the initial low-level vortex to a tropical storm strength. The increased saturation deficit leads to the intrusion of the dry unfavorable for the development of the initial vortex, thereby reducing TC frequency in a future warmer climate. Sugi et al. (2015) notes that increased saturation deficit leads to reduced upward mass flux due to the intrusion of the dry air which disrupts the deep convection, suggesting that saturation deficit and the upward mass flux hypotheses maybe not be independent. Therefore, further studies are required to understand the mechanisms leading to a reduction in the TC frequency with increasing temperatures using the current state-of-the-art climate models to increase the confidence in the TC frequency projections.

Additionally, the regions of TC occurrence and their resulting impacts may change in a warmer climate. Johnson and Xie (2010) have examined the observations of SST threshold and precipitation. They have found that in a warmer world, the SST threshold needed for the deep convection regions liable to TC genesis is likely to rise. Evans and Waters (2012) have noted similar result using five different atmosphere-ocean coupled climate models. One of the implications of this increase in the SST threshold is that the general increase in SST accompanies a change in the future regions of TC genesis due to changes in the atmospheric circulation. Chavas et al. (2017) explains the relationship between central pressure, maximum wind speed, storm size and rotation rate using idealized, earth like simulations and observations. Since cyclones form over oceans, aqua-planet simulations represent simplified versions of the global climate models with lower boundary conditions covered with oceans and no land. These experiments act as an emerging modeling framework for a detailed understanding of the climate controls on TC activity (Hayashi & Sumi 1986; Merlis & Held, 2019). The study by Chavas and Reed (2019) using dynamical aquaplanet simulations with uniform thermal forcing isolates the dependence of TC genesis rate on Coriolis parameter. Their study identifies minimum genesis distance from the Equator as the equatorial deformation radius, and shows that the Rhines scale is an important length scale for TC dynamics particularly as a limit on storm size. A diagnostic study of Hsieh et al. (2020) using idealized simulations with AGCMs notes that ventilation index acts as an important parameter which explains the differences in the TC frequency projections of the global models.

As the dominant variations in climate conditions are in the meridional direction, these aqua-planet experiments with zonally symmetric SSTs as the lower boundary conditions help us better understand the influence of the variations in SSTs and associated changes in the atmospheric variables on TC formation. The aqua-planet simulations with different SST magnitudes can help us to identify the influence of SSTs on the TC formations quantitatively. Therefore, idealized experiments using global climate models are required to assess the relative importance of large-scale environmental variables to the TC formation. Chapter 7 investigates the effects of SSTs on marsupial pouch formation environments, and TCs identified using a resolution-dependent TC tracking scheme and the associated atmospheric conditions leading to the changes in the frequency of formations.

2.6 Summary

From these studies, we can infer that observational and modeling studies have predicted an increase in TC intensity with certain uncertainties in the TC frequency that makes it crucial to thoroughly explore the linkage between the mean climate and TC formation. The recent theory of the TC formation marsupial pouch theory advances our current understanding of the formation of an initial vortex within large-scale disturbances. It is, therefore, necessary to understand the changes in the TC formation within marsupial pouch environments and the associated environmental conditions in both current and future climate. Both observations and different idealized numerical experiments using a high-resolution atmospheric climate model, different TC detection and tracking schemes are necessary to understand the marsupial pouch formations. The essential aspects identified from the above literature, which are further explored in the current thesis are:

(i) Regional variations in environmental conditions and structural differences in large-scale disturbances across the NA and WNP basins result in differences in the relative importance of environmental variables favoring TC formation. A basinwide comparative study of the environmental differences surrounding the developing and non-developing circulations improves our current understanding of significant influencing environmental variables for TC formation in the current climate. These differences have been addressed in Chapter 4.

(ii) The recent advancements in the machine learning algorithms that involve complex algorithms to delineate the combined influences of different parameters within the given dataset can advance our understanding on the relationship between environmental variables and the TC formation. The use of these machine learning algorithms across different ocean basins further helps develop more robust TC formation forecast schemes. The comparison of the three machine learning algorithms in predicting the TC formation is carried out in Chapter 5.

(iii) The frequency of the TCs detected within the marsupial pouches using a phenomenon-based tracking scheme that uses the resolution-independent thresholds in the reanalysis datasets is close to the observations (IBTrACS). Therefore, it is important to observe the TC detections within pouches in climate models, which are important tools to understand future changes. The comparison of this pouch detection scheme with a traditional TC detection scheme that uses resolution-dependent thresholds identifies a better performing TC tracking scheme that could be applied to different climate models. This comparison of the tracking schemes is carried out in Chapter 6.

(iv) The differences in the environmental variables due to both natural and anthropogenic influences lead to differences in the TC frequency. These background environmental changes also influence the formation of the marsupial pouch environments. Therefore, it is important to understand the relationships of the large-scale environment variables with the TCs identified within marsupial pouches in the climate model simulations. A better performing model in the current climate that has significant relationships to the large-scale climate variables can be used to study the future changes in the TCs using different idealized configurations. These relationships are studied in Chapter 6.

(v) Further, idealized climate model simulations to detect the TCs within the marsupial pouches and their fundamental relationships to sea surface temperatures and associated changes in the atmosphere improve our understanding of the relationship between TC formation and climate. This aspect has been addressed in Chapter 7.

Chapter 3. Data and Methods

3.1 Introduction

In this chapter, the data and the methods required to achieve the objectives of section 1.2.2 are described. The chapter gives a broad overview of general circulation models (GCMs) and the technical description of the model used in the study. The details of the observational and reanalysis data used in the study are also described. The description of two different tropical cyclones (TC) tracking schemes employed in the current study is given in detail.

3.1.1 General circulation models (GCMs)

The global climate system involves complex interactions between different physical processes of the Earth system (i.e., atmosphere, ocean, land surface, and cryosphere) at distinct spatial and temporal scales. Gettleman & Rood (2016) have provided a detailed description of the different components of climate system models. GCMs are numerical models which solve the mathematical equations representing the thermodynamical and dynamical processes of the Earth system on a three-dimensional grid at various spatial and temporal scale. The GCMs are primarily employed to understand the mechanisms leading to different global atmospheric and oceanic processes. Although both theory and observation provide an understanding of our climate system, GCMs are essential tools to understand the climate in the present and the future. Held (2005) provides a detailed review on the application of the GCMs to link observations and theory, to understand the underlying phenomenon/mechanism of the complex climate system in more detail by using idealized model conditions and finally to study the changes associated with global warming and provide forecasts for the future climate. Slingo et al. (2009) describes the advances and future challenges in developing global climate models.

GCMs apply time and space discretization to solve the governing equations representing the climate system. They use either grid-point formulation where its values represent a field at discrete grid points as shown in Figure 3.1. In spectral formulation, the field is represented using a discrete set of coefficients of known harmonic functions (Durran, 1999). As the GCMs consider different physical processes at distinct spatial scales, some of these processes occur at smaller spatial scales (sub-grid scale processes) than the scale that the GCM can solve explicitly. Parameterization schemes are used to represent these processes within the model in terms of the properties of the grid boxes. Atmosphere-only GCMs (AGCMs) use observed SSTs as the lower boundary condition, and these models typically focus on a specific time period, using time-slice experiments to achieve higher resolutions (Bengtsson et al., 1995). In time-slice experiments the SSTs are produced from a coarser-resolution coupled ocean atmosphere model for a specific future time period, and then these SSTs are used for a much finer-resolution AGCM simulation. Although low-resolution models are able to generate vortices that resemble TC observations, high-resolution AGCMs have improved geographical distributions of TCs, their annual frequency, seasonal and interannual variability (Walsh et al., 2016; Camargo et al., 2016). In the current study, we employ the ACCESS GCM model for simulating the TC activity in the current and idealized climates.

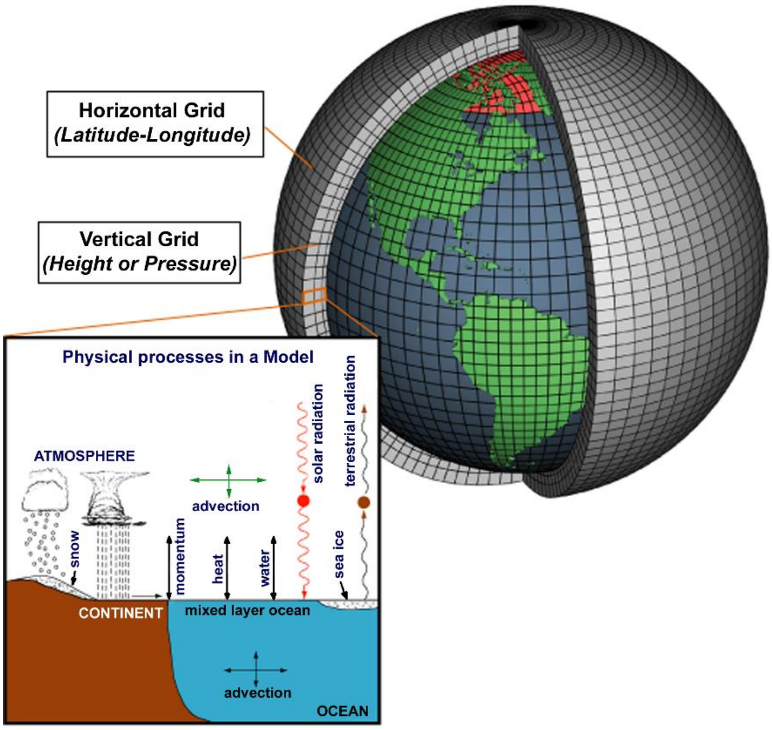

Figure 3. 1. Schematic of the three-dimensional grid representing the earth's atmosphere where the climate models estimate different atmospheric variables represented in mathematical equations (source: https://en.wikipedia.org/wiki/General_circulation_model).

Atmosphere-ocean GCMs (AOGCMs) involve the coupling of the ocean models with the atmosphere models, and they usually involve sea-ice and land surface models. The AOGCMs require higher computational resources compared to the AGCMs as they have to solve the dynamical and physical equations of both oceans and atmospheres. Also, AOGCMs require more than 1000 years spin-up period to attain an equilibrium model climate. Therefore, they typically have low resolutions as described in the Intergovernmental Panel on Climate change (IPCC) Fourth Assessment Report (IPCC, 2007); the AOGCMs have a resolution between 1.5° and 3° in the atmosphere and 1° in the oceans. The models have slightly increased resolutions in the IPCC Fifth Assessment Report (IPCC, 2013). Tropical cyclone studies using AOGCMs also require high resolution for effectively capturing the TCs and the relevant feedbacks.

Recent Earth system models involve further complexity of the climate system as they incorporate interactive biogeochemistry that includes the carbon cycle. The IPCC Fifth Assessment Report describes the results of the recent Earth system models representing different climate features. These models are of increasing interest in understanding the future TC activity and their associated impacts and mitigation strategies. A study by Dunstone et al. (2013) has employed the Earth system models and suggested a mitigation pathway via aerosols for the Atlantic TC activity. Also, the study by Evan et al. (2011) argues that the decrease in the atmospheric aerosols near the Indian Ocean leads to a decrease in the vertical wind shear and thereby increasing the TC activity in the pre-monsoon season in the Arabian Sea. The Earth system models further improve our understanding of the implications of TCs on the large-scale climate.

3.1.2 AGCMs horizontal resolution and model physics

Earlier studies using AGCMs have noted that low-resolution models could also generate storms that resemble observations (Bengtsson et al., 1982; Camargo et al., 2005). A study by Bengtsson et al. (2007b) using two different resolutions AGCMs have noted that higher resolution simulations are necessary for better representation of the future frequency changes of intense TCs. The horizontal resolution of the GCMs plays an important role in reliable TC climatology, frequency, and intensity estimates in the current and future climates (Camargo & Wing, 2016). An intercomparison of climate models with different resolutions and TC activity show that coarser-resolution models have specific difficulties in simulating TC climatology, particularly in the North Atlantic basin and other TC characteristics (Walsh et al., 2010; Shaevtiz et al., 2014). They also have noted that fine resolution models with improved convective parametrization schemes improve the TC activity simulated by the climate models. Recent studies are moving towards high-resolution simulations due to increased computational resources and advances in modeling capabilities (Bengtsoon et al., 2007a; Zhao et al., 2009;

Murakami & Sugi, 2010; Murakami et al., 2012a; Walsh et al., 2013; Camargo & Wing, 2016).

Also, higher resolution AGCMs are required for a better geographical representation of TC genesis frequency and better TC formation statistics across the global ocean basins (Walsh et al., 2013). Murakami et al. (2010) have showed that high-resolution AGCMs simulations are preferable for representing realistic future changes in the TC frequency and intensity characteristics. They also have noted that models with finer than 60km could simulate the future changes in the TC activity compared to the coarser resolution model simulations. The TC activity's sensitivity also depends on the choice of the cumulus convection scheme, which is shown by Murakami et al. (2012) using multi-model and multi-physics experiments. Therefore, a high-resolution AGCM with improved model physics and parametrization schemes improves different TC characteristics and gives more realistic projections for future climate (Camargo & Wing, 2016). The current study uses a higher resolution version of the ACCESS GCM having around 40km horizontal resolution with improved model physics for studying the tropical cyclone frequency characteristics.

3.2 ACCESS model description

In the current study, we employ the atmospheric component of the Australian Community Climate and Earth-System Simulator GCM (ACCESS, Bi et al., 2013). Previous climatological simulations with observed SSTs to study TC climatology are from the earlier version unified model GA3.0 having a horizontal resolution of 25km (Roberts et al., 2015). They have noted improved TC numbers by increasing the horizontal resolution. Further updates to the previous version of the unified model are carried out to improve the TC characteristics as detailed in a recent study by Walters et al. (2017). One such update includes a change in the entrainment rate, which improved the representation of the deep convection, thereby improving the TC intensities. Another improvement is the reduced TC intensity bias when compared with the observations. The advanced dynamical core of the Unified Model, “Even Newer Dynamics for General Atmospheric Modelling of the Environment” (ENDGame) (Wood et al., 2014) improved the simulation of large-scale fields, which then had a positive impact on the TC simulations (Reed et al., 2015). These improvements in the dynamical core and the entrainment rate of the Unified Model improved the short-range tropical cyclone tracks forecasts by reducing the tracking error by 7%, as described in a study by Heming

(2016). There are further improvements in the representation of the tropical wave activity, the eastward propagating Kelvin waves, and a better Madden Julian Oscillation (MJO) signal (Walters et al., 2017).

The description of the current ACCESS modeling system which has the above updates and improvements mostly follows that contained in the previous work using a similar model (Sharmila et al., 2020). This ACCESS model uses the atmospheric part of the UK Met Office Unified Model (UM v8.5) with improved parameterization schemes of the Global Atmosphere (GA 6.0) configuration. It includes the JULES (Joint UK Land Environment Simulator) land surface model (Best et al., 2011; Walters et al., 2017). The dynamical core of the atmosphere within the model uses the semi-implicit semi-Lagrangian approach in a longitude-latitude grid, solving both the large-scale processes (solar and terrestrial radiation, precipitation and gravity wave-drag) and sub-grid scale processes (boundary layer processes, convection, atmospheric aerosols, and chemistry). A prognostic cloud fraction and prognostic condensate (PC2) scheme are employed (Wilson et al., 2008). The model employs the radiation scheme of Edwards and Slingo (1996) for representing the cloud droplets, the boundary layer scheme of Lock et al. (2000) with additional modifications for representing the boundary layer, as described in Lock (2001) and Brown et al. (2008).

The model is run in an atmospheric mode with prescribed monthly and interannually varying Atmospheric Model Intercomparison Project (AMIP) SSTs (Kanamitsu et al. 2002) as a lower boundary condition, and has a 40 km horizontal resolution with 17 vertical levels. The model is initialized with the atmospheric conditions corresponding to SSTs of 1 September 1988, which themselves have been produced from a spinup run starting from 1 September 1950 forced with AMIP-II SSTs and seasonally varying radiative forcing. The model is then run from 1st September 1988 until 31st December 2010 with the model outputs saved for every 6 hour time interval. For the current climate simulation we have used 20 years of data from 1990-2009 for the analysis of objective 3 in chapter 6. For the idealized climate model simulations to achieve objective 4 we have carried out three fixed SST experiments as described in chapter 7 using the same initialized ACCESS model.

3.3 Observational data

The climate model simulations are evaluated by comparing with the observations for both the simulated TCs and the large-scale environmental variables. This section describes the sources of different observational datasets used in the current thesis to evaluate the performance of the ACCESS model.

3.3.1 Tropical cyclones (IBTrACS)

The global observations of Tropical cyclones are from the International Best Track Archive for Climate Stewardship (IBTrACS) TC data for 1990-2018. This IBTrACS data provides the best information about the observed TCs from the different regional TC observing centers synthesized and merged into a single dataset (Knapp et al., 2010). This data includes the TC latitude-longitude position, maximum sustained wind speeds (3min or 10 min), and the TC's central sea level pressure at six hourly intervals covering the whole duration of the storm. The TS genesis location of IBTrACS is the first position at which the system satisfies a maximum sustained low-level 10 min wind speed higher than 33 knots (17 ms^{-1}) during its lifetime. For the years analyzed here, the IBTrACS data should have an almost complete representation of TC formation. Knapp et al. (2010) note small differences between the operational centers that compile the best track data in the total number of declared TCs (see Ren et al., 2011; Hodges and Emerton, 2015). Here, IBTrACS serves as ground truth for TC formation. An issue in all best track data sets is the lack of a consistent definition of extratropical transition between operational centers, which leads to biases in the length of the observed TC tracks (Schreck et al. 2014).

3.3.2 Reanalysis data

Hodges et al. (2017) have used objective detection schemes having vorticity or pressure extrema thresholds for TC detection to assess the representations of different TC characteristics in six different reanalysis data (viz. ERAI, JRA-25, JRA-55, NCEP-CFSR, and MERRA). Their study shows that all the reanalysis datasets shows good representation (~98%) of TCs observed in best track data. With a motivation of applying an objective based TC detection and tracking scheme to different resolution climate models, Tory et al. (2013a, b) formulated the Okubo-Weiss Zeta Parameter (OWZP) TC detection (resolution-independent) scheme by tuning suitable thresholds of large-scale environmental variables in the ERA-interim reanalysis data at 1°×1° resolution. Also, as the TC detection threshold criteria using objective schemes are tailored to a particular reanalysis (Murakami, 2014), the present study uses the OWZP detection scheme proposed by Tory et al. (2013a, b) in the ERA Interim reanalysis dataset.

Reanalysis data is based on both the short-range forecast model data and diverse observational data using data assimilation techniques. The forecast model involves a cyclic process with a forecast made based on the previously assimilated input data. The reanalysis data provides the best 4D view of the atmosphere across the globe, constrained by observations but more homogenous compared to the inhomogeneous observations. The reanalysis data depends on the forecast model, data assimilation techniques, observations, and different satellite retrieval algorithms. Therefore, a better assimilation technique, improved model physics, and advanced retrieval algorithms to extract the features from different sets of observations improve the quality of the reanalysis data. The current study uses the European Centre of Medium-range Weather Forecasts (ECMWF) reanalysis ERA-interim. The ERA-interim reanalysis is an updated version of the reanalysis compared to the earlier version ERA-40. This version involves an integrated forecasting system that uses complete 4D VAR data assimilation techniques within a spectral model. It also includes improved and additional observations both satellite and buoy (Compo et al., 2011; Dee et al., 2011). The current study uses the ERA-interim reanalysis data having a horizontal resolution of around 80km with six-hourly data for a period of 30 years from 1989-2018.

3.4 Tropical cyclone tracking algorithms

Different TC tracking schemes have been developed. These are often basin specific and model resolution-dependent, making it more difficult to compare the different studies (Camargo & Zebiak, 2002; Walsh et al., 2007; Zhao et al., 2009; Strachan et al., 2013). These traditional TC tracking schemes employ certain thresholds to identify the vortices which resemble observed TCs (as taken from IBTrACS). As the GCMs employ coarser resolution due to computational constraints compared to the fully resolvable GCM resolution required to represent TC processes, which is about 1 km, they use some resolution-dependent thresholds to identify TC-like vortices in the global models (Walsh et al., 2007). Horn et al. (2014) carried out an intercomparison of the different models that employ different tracking schemes in diagnosing the TCs. They noted specific differences in the future TC activity due to differences in the thresholds of wind speeds and duration thresholds. A TC detection and tracking scheme (OWZP) that employs a different method to other existing schemes is by Tory et al. (2013a), which detects the locations favorable for TC formation by using resolution-independent thresholds of large-scale dynamical and thermodynamical conditions. In the current study, we employ both a traditional tracking scheme CSIRO (Walsh et al., 2007; Horn et al., 2014) and a phenomenon-based TC tracking scheme OWZP (Tory et al., 2013a, b, d).

3.4.1 CSIRO TC detection and tracking scheme

The modified CSIRO scheme, a traditional TC tracking algorithm that detects TC-like vortices using certain resolution-dependent objective thresholds, as shown in the Figure 3.2. The following description follows that of Horn et al. (2014). The thresholds to be satisfied for declaring a low-pressure system as a TC are:

(i) An absolute estimate of 850 hPa cyclonic vorticity which is calculated from the model simulated wind fields using centered difference scheme should be higher than $1 \times 10^{-5}\ s^{-1}$ to avoid grid points having weaker vorticity;

(ii) A closed pressure minimum with a 350 km distance in both x and y directions from the location satisfying the first condition is taken as the center of the storm;

(iii) Mean wind speed within a 700 km x 700 km box region around the storm center at 850 hPa pressure level is higher than wind speed at the 300 hPa;

(iv) The maximum 10-meter wind speed should exceed the resolution-dependent threshold, as proposed by Walsh et al. (2007). The wind speed threshold is 16.5 m/s for the 40 km horizontal resolution ACCESS model data and 15.5 m/s for the lower-resolution ERA-Interim reanalysis;

(v) Finally, the above conditions are to be satisfied for 24 consecutive hours to be declared as a TC.

Here using the CSIRO scheme, the TCs are detected only over oceans, based on the degraded topography data to the model resolution. The detections are tracked to create storm tracks by linking the consecutive detections within a 6° region for six-hourly temporal resolution data. The tracks which are lasting less than 24 hours are discarded. Here no additional checks are carried out to separate the extratropical storms. The TCs are separated from the extratropical storms using a latitudinal distribution of the genesis locations. The genesis locations within the 30N-30S latitude band are considered as the tropical cyclones.

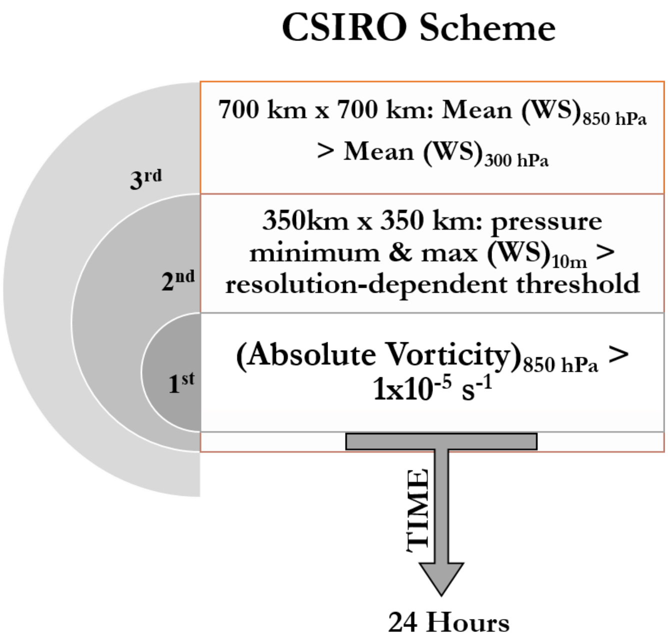

Figure 3. 2. Schematic of the steps involved in the CSIRO TC detection and tracking scheme. Here WS represents wind speed.

3.4.2 OWZP TC detection and tracking scheme

Tory et al. (2013a) found that 95% of observed TSs (as defined in IBTrACS) have enhanced values of OWZ at both 850 hPa and 500 hPa levels relative to their surroundings, prompting them to base their OWZP TC tracking scheme on these two quantities along with other environmental variables. The Okubo-Weiss Zeta (OWZ) parameter detects the low-deformation vorticity regions within the large-scale disturbances (marsupial pouch regions) that have the potential for TC formation. Therefore, the phenomenon-based TC detection and tracking scheme (OWZP) uses the diagnostic variable OWZ, which replaces the absolute vorticity variable of other traditional TC tracking schemes along with thresholds of other environmental variables. The OWZ calculation involves the normalization of the OW parameter by the square of the relative vorticity. The Okubo Weiss (OW) parameter is a relative measure of vorticity to deformation forces that identifies the dominant rotation regions and strain-dominated areas such as the shear sheath.

$$OW = \zeta^2 - (E^2 + F^2) \qquad (1)$$

where $\zeta = \left[\frac{\partial v}{\partial x}\right] - \left[\frac{\partial u}{\partial y}\right]$; $E = \left[\frac{\partial u}{\partial x}\right] - \left[\frac{\partial v}{\partial y}\right]$; $F = \left[\frac{\partial v}{\partial x}\right] + \left[\frac{\partial u}{\partial y}\right]$

Here, ζ is the relative vorticity, E and F represent the stretching and shearing deformation, respectively.

The OWZ uses only positive values, including a range of flows, from deformation-dominated flows (OW_{norm} = 0) to rotation dominant regions (OW_{norm} = 1). The phenomenon-based TC tracking scheme uses the thresholds of the OWZ variable instead of the OW variable. The equation representing the OWZ is:

$$OWZ = \max(OW_{norm}, 0) \times (\zeta + f) \times sign(f) \qquad (2)$$

$$OW_{norm} = \frac{OW}{\zeta^2} \qquad (3)$$

Here OW_{norm} is the normalized OW parameter, and f is the Coriolis parameter.

The other atmospheric variables used in the tracking scheme are relative humidity (RH) at 700 and 950 hPa, specific humidity at 950 hPa (SH_{950}), and vertical wind shear (VWS), which is the vector wind difference between 850 and 200 hPa pressure levels as shown in Table 3.1. This scheme is designed to be applied to data on a $1° \times 1°$ grid. Therefore, we interpolate the ERA-interim data in the current study to $1° \times 1°$ grid resolution.

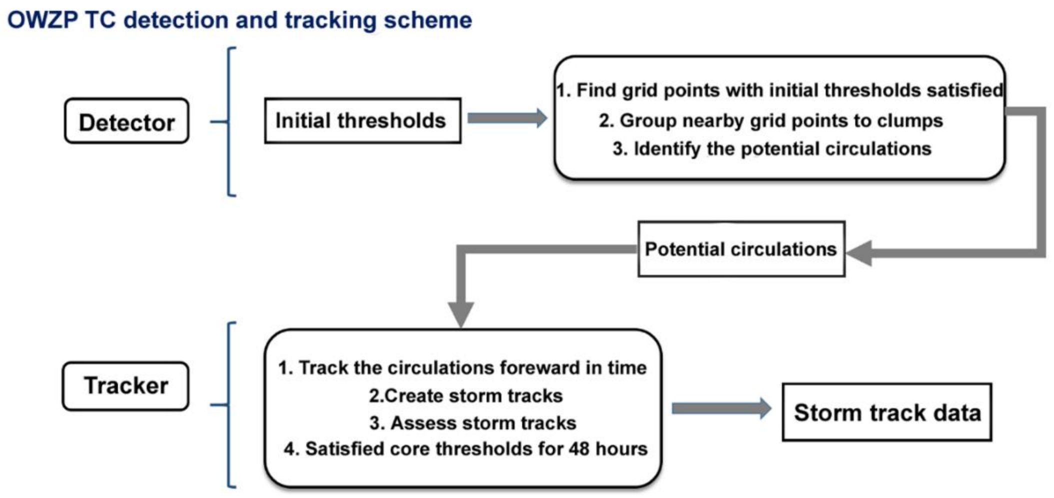

Figure 3. 3. The steps involved in the OWZP TC detection and tracking scheme for either climate model or reanalysis gridded data.

Tory et al. (2013a, 2013b) give a detailed explanation of the detection scheme, which is a multi-step process, as shown in Figure 3.3:

(i) Detect storms: Identify circulations that have the potential for TC formation by testing every grid point with an initial set of threshold conditions (Table 3.1 gives the "initial" thresholds). Neighboring grid points satisfying the initial thresholds are clustered into individual clumps that represent the storm at specific times.

(ii) Create storm tracks: Individual storm tracks are constructed by linking the above instantaneous clumps forward in time based on the circulation's expected position from the 700 hPa steering flow (the average wind at 700hPa within a $4° \times 4°$ box region around the grid point center). The tracks cease when there is no future circulation in the vicinity of the expected storm position.

(iii) Detect TCs: Figure 3.4 gives a schematic diagram of this process. At each point along the storm track, clump averages of different large-scale variables are assigned to each clump (storm) position. The core thresholds of Table 3.1 are checked with each clump position and labeled as "True" if the clump satisfies these core thresholds or "False" if not. Each storm track is assessed to see if it satisfies tropical depression (TD) status, and then tropical storm (TS) status. A particular track is declared a TS (TD) if it satisfies the core thresholds for at least 48 (24) hours consecutively, which amounts to five (three) consecutive true values for 12 hourly data.

Table 3. 1. Threshold values of parameters considered in the OWZP detection and tracking scheme

OWZP detection Scheme threshold values of parameters						
Criterion	**OWZ_{850}**	**OWZ_{500}**	**RH_{950}**	**RH_{700}**	**$VWS_{850\text{-}200}$**	**SH_{950}**
Initial	$50 \times 10^{-6} s^{-1}$	$40 \times 10^{-6} s^{-1}$	70%	50%	$25 ms^{-1}$	$10\ g\ kg^{-1}$
Core	$60 \times 10^{-6} s^{-1}$	$50 \times 10^{-6} s^{-1}$	85%	70%	$12.5 ms^{-1}$	$14\ g\ kg^{-1}$

This condition of five consecutive true clumps is critical for a storm track in coastal regions as they may be unsuccessful temporarily in satisfying the above additional core thresholds. Therefore, each track is assessed for an additional set of conditions: (i) if less than two clumps are over water then it must be affected by land, and the clump is assigned false; and (ii) there should be at least four consecutive true clumps irrespective of the length of the storm track. The Coriolis contribution to the OWZ calculation was capped to latitude 20° to avoid detecting any anticyclonic circulations that satisfy the OWZ thresholds at 500 hPa at latitudes greater than 20°. This additional criteria is imposed to account for the very rare occasion in which the tracking scheme detects an anticyclonic circulation as a TC (Tory et al., 2013d).

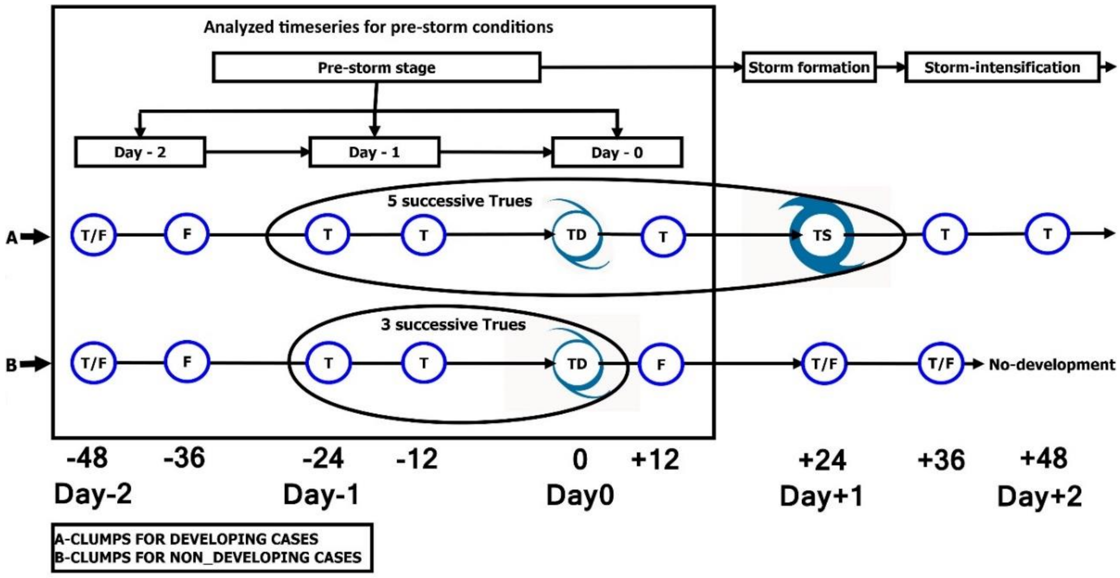

Figure 3. 4. Diagram showing the tracks of developing and non-developing tropical depressions with F(False) indicating that the core thresholds are not satisfied and T(True) where the thresholds are satisfied. The 3rd consecutive true is the TD declaration time for both developing and non-developing TDs. Path A shows that all TSs were once TDs, and path B is a TD that did not develop into TS.

A dynamical technique is imposed to separate the sub-tropical storms after detecting the TCs using the OWZP scheme. Here we define a sub-tropical jet as the latitude at which the 200 hPa wind speeds exceed 25 m/s, and the zonal winds exceed 15 m/s. Here we estimate the subtropical jet position at every longitude and then check these jet positions with the detected genesis locations of the tracking scheme and then identify the sub-tropical storms (Tory et al., 2013d). The genesis locations poleward of this jet position are considered as the sub-tropical storms and are excluded for further analysis. For aquaplanet analysis in the current study (chapter 7), which have uniform SSTs across the globe, we do not use this dynamical technique to separate the subtropical storms due to the absence of a defined subtropical jet.

Chapter 4. Basinwise statistical analysis of factors limiting tropical storm formation from an initial tropical circulation

This chapter addresses the first research objective (section 1.2.2) and is reproduced from the published paper of Raavi and Walsh (2020b): Basinwise statistical analysis of factors limiting tropical storm formation from an initial tropical circulation. *Journal of Geophysical Research: Atmospheres*, (https://doi.org/10.1029/2019JD032006).

Abstract

Tropical cyclones (TC) form within a closed circulation region of large-scale disturbances under certain dynamical and thermodynamical conditions. This study investigates differences in the large-scale environmental conditions of a TC 48 hours before the declaration of tropical depression (TDs), comparing those that developed into tropical storms (TSs) with those depressions that did not develop. Here, we apply the Okubo-Weiss Zeta Parameter (OWZP) detection and tracking scheme to ERA-interim reanalysis from 1989-2018, across different ocean basins. The method detects storm-system-scale environmental conditions that favor TC formation. We construct spatial composites of thermodynamical and dynamical quantities for both developing and non-developing depressions, as well as storm-relative streamlines. A statistical index (the Box Difference Index) is used to quantitatively estimate the dominant limiting factors of TS formation from the area-averaged quantities of large-scale variables. The relative contribution of large-scale environmental variables impeding the development of an initial TD to TS differs between ocean basins perhaps due to regional variations in the characteristics of large-scale disturbances and the surrounding environmental conditions. A streamline analysis shows the developing storms have a more pronounced cyclonic core and are protected from external influences by a stronger shear-sheath layer surrounding the core extending from the lower- to mid-troposphere. This study, therefore, identifies the potential limiting variables and structural differences in an initial circulation that impede TS formation across different ocean basins.

4.1. Introduction

Tropical cyclones (TCs) are highly destructive extreme events that lead to substantial socio-economic losses, primarily in coastal regions. TCs form under certain favorable thermodynamical and dynamical conditions within closed recirculating regions (Gray, 1968, 1975; Dunkerton et al., 2009). Favorable thermodynamical conditions include sea surface temperatures (SSTs) higher than 26.5^{o}C, a deep layer of a conditionally unstable environment, and moist mid-troposphere (Gray, 1968). Dynamical conditions include weak to moderate vertical wind shear (vector wind difference magnitude between the 850 and 200 hPa), deep organized convection, and locally enhanced values of low-level cyclonic relative vorticity (Gray, 1968, 1975; McBride & Zehr, 1981). Additionally, Gray (1968) suggested that TCs generally do not form within 5 degrees of the equator. Although these conditions are necessary for TC formation, it is still not clear what are the sufficient conditions, due to multi-scale interactions in the TC formation activity.

TC formation undergoes different stages of development. The first stage involves pre-conditioning of the large-scale environment along with the formation of minimum atmospheric pressure in the lower levels. The second stage involves the formation of a self-sustaining mesoscale vortex and its aggregation, with modest stretching and shearing deformations enabling a near-solid body rotation (Lee et al., 1989; Briegel & Frank, 1997; Ritchie & Holland, 1999; Montgomery et al., 2006b; Dunkerton et al., 2009). Some large-scale disturbances provide suitable conditions such as increasing low-level relative vorticity, convective activity associated with increased updraft strength, and altering the vertical wind shear distribution to conditions favorable for TC formation. These disturbances include: 1) tropical depression type waves also called African Easterly Waves (AEWs) in the North Atlantic (NA) region (Landsea, 1993; Thorncroft & Hodges, 2001); 2) Rossby wave trains induced by a preexisting TC, Pacific easterly waves and synoptic wave trains, or tropical depression type waves in the Western North Pacific (WNP) region (Li & Fu, 2006; Li et al., 2006; Fu et al., 2007; Xu et al., 2013); and 3) Equatorial Rossby waves and mixed Rossby-Gravity waves in the South Indian Ocean (SI), South Pacific (SP), and Australian (AUS) regions (Hall et al., 2001; Bessafi & Wheeler, 2006; Frank & Roundy, 2006). Other forms of large-scale disturbances favorable for TC formation are the monsoon trough, regions with upper-level troughs and low-level wind surges (Ferreira & Schubert, 1997; Briegel & Frank, 1997; Ritchie & Holland, 1999; Wang & Mognusdottir, 2005, 2006; Zong & Wu, 2015).

Every year a large number of tropical depressions form in global ocean basins from various types of tropical disturbances. However, despite apparently suitable conditions provided by these disturbances, not all the depressions develop into tropical storms. For instance, in the NA, a large number of AEWs form every year, but only around 30% of them develop into tropical storms (Avila, 1991). Understanding of the TC formation pathway has recently been advanced by the "marsupial paradigm" defined within tropical easterly waves over the Atlantic and Pacific regions (Dunkerton et al., 2009). This innovative work recognizes that TCs form in a relatively rare recirculating region called a pouch within a critical layer of the tropical wave under favorable surrounding conditions. A deep pouch from the low to mid-troposphere within an easterly wave maintains an area of closed circulation, where the entry of outside air is restricted, shelters the storm from disruptive influences and enables it to intensify under a favorable environment (Wang et al., 2010a, 2010b; Montgomery et al., 2010; Wang et al., 2012). Asaadi et al. (2016, 2017) have found that developing easterly waves appear to propagate as a packet of four to five waves, tilted somewhat northwest-southeast from zonal. Also, of these waves, only one wave is located at the critical latitude (where the mean flow speed equals phase speed of the wave) and in a region of weak potential vorticity gradient, where TC formation can occur. These studies highlighted the role of a deep pouch and the critical latitudes of the easterly waves favorable for TC formation.

Peng et al. (2012) & Fu et al. (2012) have examined the environmental differences between developing and non-developing tropical disturbances across the WNP and NA basins detected by time filtering technique to extract synoptic-scale disturbances (tropical waves). They have noted that thermodynamical conditions are more important across the NA basin, and dynamical conditions are more critical across the WNP basin. These studies showed that the relative contribution of large-scale variables to TC formation varies between these two basins due to distinctive large-scale disturbances and varying regional environmental conditions. It is, therefore, essential to study differences between developing and non-developing cases across global ocean basins that have diverse environmental conditions and distinct large-scale disturbances to identify the important predictors for TC formation.

There are many other studies on the environmental differences between the developing and non-developing AEWs across the NA basin (Hopsch et al., 2010; Montgomery & Smith, 2010; Davis & Ahijevych, 2012; Komaromi, 2013; Brammer & Throncroft, 2015; Fowler & Galarneau, 2017). These studies highlight the importance of relative flow within AEWs, increased low-level vorticity, vertical shear, and sufficient tropospheric moisture for the development of an initial disturbance. A moist mid-troposphere increases the strength of convective updrafts by decreasing the effect of dry air, which disrupts the deep convection (Nolan et al., 2007; Smith & Montgomery, 2012; Wang, 2012). Persistent broad-scale, deep convection can then spin up the primary vortex via a system-scale convergence by the in-up-out secondary circulation (Hendricks et al. 2004; Montgomery et al., 2006; Tory et al., 2006a, 2006b, 2007; Raymond & Sessions, 2007).

Some studies have found that moderate shear or easterly shear is favorable for developing cases (Tuleya & Kurihara, 1981, 1982; Zehr, 1992; Zeng et al., 2010). In contrast, enhanced vertical wind shear misaligns the storm which sometimes leads to ventilation, thereby disrupting the intensification process (Riemer et al., 2010; Tang & Emanuel, 2010; Davis & Ahijevych, 2012; Wang et al., 2012; Alland et al., 2017). Other studies have showed that abundant low-level convergence and enhanced mass flux through organized outflow is associated with sustained deep convection in developing cases, with an anticyclonic circulation in the upper levels (McBride & Zehr, 1981; Lee, 1989a, 1989b; Zehr, 1992; Tuleya, 1991; Zhang & Zhu, 2012; Sears & Velden, 2014).

In addition to the above studies on environmental differences, from particle trajectory analysis around the developing storms in the NA basin, Rutherford et al. (2015) identify a protective layer termed the "shear sheath" surrounding the core of the storm that can inhibit shear-induced storm weakening. This shear sheath is a layer of high deformation (that is, a strain-dominated region) that serves to isolate the core from its surroundings. These strain-dominated regions are termed as the rapid filamentation zones, where convection is suppressed, and subsidence is more pronounced (Rozoff et al., 2006). The shear sheath acts as a barrier layer that protects the initial vortex from external influences by obstructing the horizontal inflow of dry air or opposite signed vorticity fields into the core region of developing storms. Rutherford et al. (2015, 2018) have found that the development of a tropical depression to a tropical storm involves a notable shear-sheath region around the vortex core region.

From the studies above, we note that there are no global studies on the environmental and structural differences between the developing and non-developing circulations. In the current study, we use a phenomenon-based TC detection and tracking scheme to detect the initial circulations and identify the environmental and structural differences across different ocean basins. This TC detection scheme is developed based on the marsupial paradigm of TC formation. It detects the low deformation vorticity regions within the pouch that have the potential for tropical storm (TS) formation by using a diagnostic (OWZ) along with other environmental variables. This diagnostic is a product of the normalized Okubo-Weiss (OW) parameter and absolute vorticity (Zeta) parameter (Tory et al., 2013a, b; see section 2 for a more detailed description). The Okubo-Weiss Zeta Parameter (OWZP) tracking scheme detects both a TS that is any TC with low-level winds higher than 17 m /s and a tropical depression (TD) that is any TC weaker than a TS. We consider those TDs that become TSs as developing TCs and the remaining TDs non-developers (Tory et al., 2018). This tracking scheme is particularly helpful in the basins where there are no observational aircraft missions for identifying and tracking the initial depressions. Therefore, we perform a basinwide composite analysis on ERA-Interim reanalysis data using these detected circulations to identify the limiting factors for the development of an initial TD.

From the above literature, TC formation from an initial circulation requires a favorable "external environment," such as enhanced vorticity, low-level cyclonic circulation, a moist atmospheric column, and weak to moderate vertical wind shear. As the circulation intensifies, the "internal environment" (the core of the circulation) becomes more favorable: it becomes moister, vorticity amplifies, and the shear-sheath layer around the core strengthens. It is, therefore, difficult to separate the internal and external environment, so a more developed TC looks like it is in a more favorable environment than a weaker TC. By focusing on the differences between developing and non-developing TCs 48 hours before the TD declaration, we have some confidence that most of the differences we identify are likely to be external rather than internal. This composite analysis helps us to find out the external environmental conditions and structural variations that assist or hinder TS formation, and whether these conditions are similar or differ between ocean basins.

4.2. Data and Methodology

4.2.1 TS observations and ERA-Interim data

The observed TS information for 1989 to 2018 period is from the International Best Track Archive for Climate Stewardship (IBTrACS) database (Knapp et al., 2010) as described in section 3.3.1. The reanalysis fields for analyzing the environmental conditions of developing and non-developing circulations are taken from the ERA-Interim data (Compo et al., 2011; Dee et al., 2011) as described in section 3.3.2. These data have a horizontal resolution of $0.75° \times 0.75°$. We extract the data for a period of 30 years, from 1989 to 2018, at 12 hourly intervals (00 and 12 UTC) to create variables used by (i) the TC detector and (ii) the composite analysis of the environmental conditions.

The variables used by the detector at 12 hourly intervals (00 and 12 UTC) include the zonal (u) and meridional (v) wind fields, to compute 1) the OWZ at both 850 and 500-hPa pressure levels (Tory et al., 2013a; see also chapter 3); 2) vertical wind shear (vector wind difference magnitude) between 850 and 200 hPa pressure levels; and 3) the TC steering velocity (the average wind at 700hPa within a $4° \times 4°$ box region around the grid point center) (Tory et al., 2013a). Additional parameters include relative humidity at 700 and 500 hPa and 950-hPa specific humidity (SH_{950}).

The variables used to examine the composite storm environment are those already found to be important for TC formation (Gray, 1968; McBride & Zehr, 1981; Emanuel & Nolan, 2004; Camargo et al., 2007; Tippet et al., 2011; Tory et al., 2013a; Ditchek et al., 2017). These variables include the total column water (TCW; in units of kg/m^2), 700 hPa relative humidity (RH; %), maximum potential intensity (MPI; m/s), 850 hPa Relative vorticity (RVor; s$^{-1}\times10^{-6}$), 500 hPa Omega (Omega; Pa/s), 200 hPa divergence (DIV; s$^{-1}\times10^{-6}$), vertical shear of zonal wind ($U_{200} - U_{850}$) (Ushear; m/s), and OW parameter (s$^{-2}\times10^{-10}$) (see chapter 3 for the definition). We consider the positive values of Ushear as westerly shear and negative values as easterly shear. Here, the signs of Omega globally and RVor in the Southern Hemisphere are reversed, meaning that positive Omega is now upward motion, and positive RVor in the Southern Hemisphere is now cyclonic flow. We calculate the MPI using the equation formulated by Bister and Emanuel (1998):

$$MPI = \frac{T_s}{T_o}\frac{C_k}{C_D}\left[CAPE_e - CAPE_b\right] \qquad (4)$$

where T_s is the sea surface temperature, T_o is the outflow temperature, C_k denotes the enthalpy exchange coefficient, C_D denotes the drag coefficient, and $CAPE$ is the convectively available potential energy with the subscripts e and b denoting the environmental and boundary layer atmospheric sounding respectively. In addition to the above large-scale environmental variables, we use a meridional gradient of absolute vorticity (Beta*) variable:

$$Beta^* = \frac{\partial \eta}{\partial y} = \frac{\partial f}{\partial y} + \frac{\partial \zeta}{\partial y} \qquad (5)$$

where η denotes the absolute vorticity. Tory et al. (2018) found that TC formation is favored in regions with lower values of the Beta* and is extremely rare in regions of enhanced values.

4.2.2 Tracking of developing and non-developing tropical depressions

Here we use OWZP TC detection and tracking scheme to identify both the TDs and the TSs across the global ocean basins with similar procedure described in section 3.4.2 with detailed steps given in the Figure 4.1. A particular track is declared a TS (TD) if it satisfies the core thresholds for at least 48 (24) hours consecutively, which amounts to five (three) consecutive true values for 12 hourly data shown in Figure 3.4. Application of the TD and TS detection routine to 30 years of ERA-Interim data yielded 2300 TSs and 3432 TDs. The TD numbers include the 2300 TSs, as all TSs were once TDs. This shows that the TD to TS ratio is about 3:2, implying globally there are 50% more lows that satisfy the TD than the TS thresholds, which is consistent with the 1979-2013 period considered by Tory et al. (2018). To obtain the non-developing depressions, we subtract the developed depressions (TSs) from the total tropical depression (TD) dataset. A total of 1132 non-developing storms are detected, located across all the ocean basins. The fifth True position along the track of the TS (Figure 3.4) is considered as the TS genesis which matches with the IBTrACS genesis position in most of the ocean basins (Bell et al., 2018) whereas third true position is considered as the TD genesis. Tory et al. (2018) tuned the OWZP scheme to detect the TDs in the tropical disturbance dataset across the Australian region. Also, Bell et al. (2018) assessed the pre-genesis tracks of the OWZP and IBTrACS TSs and noted that OWZP detected TSs have earlier tracks with longer duration because the scheme detects the locations that have potential for TS formation rather than the matured TSs.

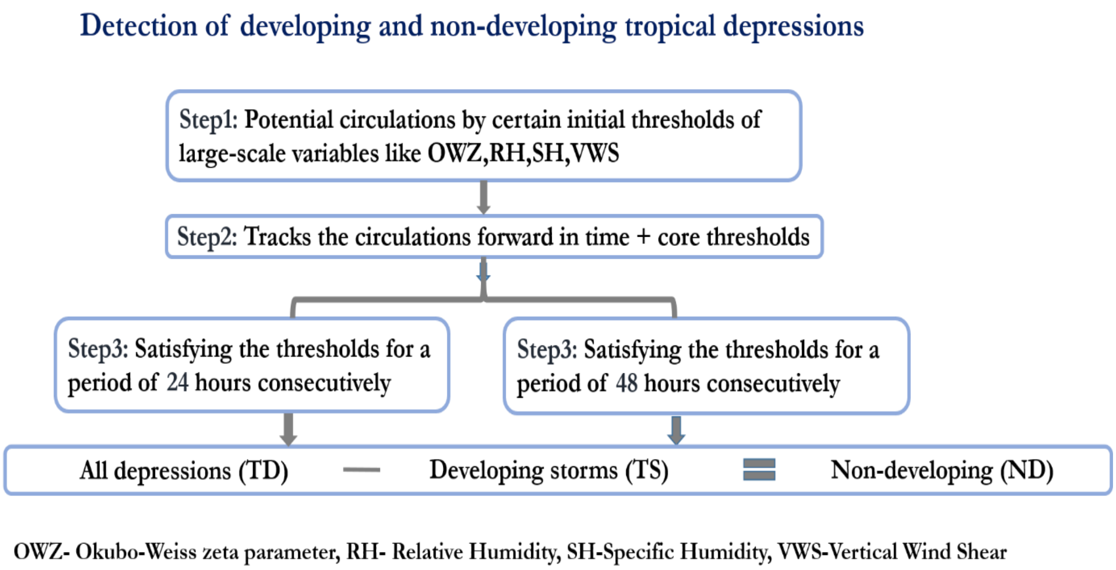

Figure 4. 1. Steps involved in the detection of the developing and non-developing tropical depressions using the OWZP detection scheme

4.2.3 Storm-centered spatial composites of developing and non-developing cases

To understand the environmental differences between developing and non-developing TDs, storm centered composites of the environmental variables are created within a 24° ×24° region around the depression center. Also, we calculated the streamlines of the composite storm overlaid with the OW parameter and RH contours to understand the structural differences and the influence of the surrounding environment on the initial circulation. We include only the storms with a detected storm track that existed 48 hours before TD declaration. This group contains developing (PDV48) and non-developing storms (PND48) across each ocean basin (Table 4.1 shows basin boundaries). The analysis considers three periods: 48 and 24 hours before the time of TD formation, and the day of TD declaration, labeled Day-2, Day-1, and Day0, respectively as shown in Figure 3.4.

Table 4. 1. Latitude and longitude of defined ocean basins. A tilde (~) indicates an irregular land boundary

Basin	Latitude; Longitude
North Atlantic - main development region (MDR_NA)	10°-20°N, ~100°W- 0
Western North Pacific(WNP)	0-15°N, 100°-180°E
Western North Pacific(poleward)(WNP_P)	15°-30°N, 100°-180°E
Eastern North Pacific(ENP)	0-30°N, 180°- ~100°W
North Indian Ocean(NI)	0-30°N, 40°-100°E
South Indian Ocean(SI)	0-30°S, 20°-135°E
Australian region(AUS)	0-30°S;100-170°E
South Pacific(SP)	0-30°S;135°E-120°W

4.2.4 Statistical Index

To quantify the differences in the environmental conditions between the developing and non-developing TDs, we calculate the Box Difference Index (BDI) (Peng et al., 2012) for PDV48 and PND48 cases across all the ocean basins within a 12°×12° box region around the depression center (blue boxes in Figure 4.2). This domain size appears to show the maximum differences between developing (DV) and non-developing (ND) compared to the 24°× 24° region which covers a larger spatial extent and may disrupt the box averages. The OW variable which includes the vorticity dominated and shear-sheath regions is divided into two variables. For this we use a 18°×18° box region (green boxes in Figure 4.2), the "OW center" variable uses the centered box showing the low-deformation vorticity strength and the "OW surrounding" variable determines the strength of the shear sheath uses the remaining boxes. The BDI index is defined as:

$$BDI = \frac{mean(DV)-mean(ND)}{std(DV)+std(ND)} \qquad (6)$$

where *mean(DV)* and *mean(ND)* are the mean of the environmental variables for DV and ND cases and *std(DV)* and *std(ND)* are the standard deviation of the DV and ND environmental variables respectively. The higher the magnitude of the BDI (which ranges from 0 to 1), the better the variable can differentiate the two cases. The sign of the BDI index tells us the environmental behavior of the variable. In general, positive values indicate that the quantity favors development. However, the opposite is true for quantities in which small values favor TC formation, such as Ushear, Beta* and OW surrounding variables. The negative values of Ushear, Beta* and OW surrounding variables indicates favorable shear, barotropic stability regions and stronger shear sheath regions respectively that support the TC formation. We use a two-sample t-test to determine the statistical significance of the differences between DV and ND cases.

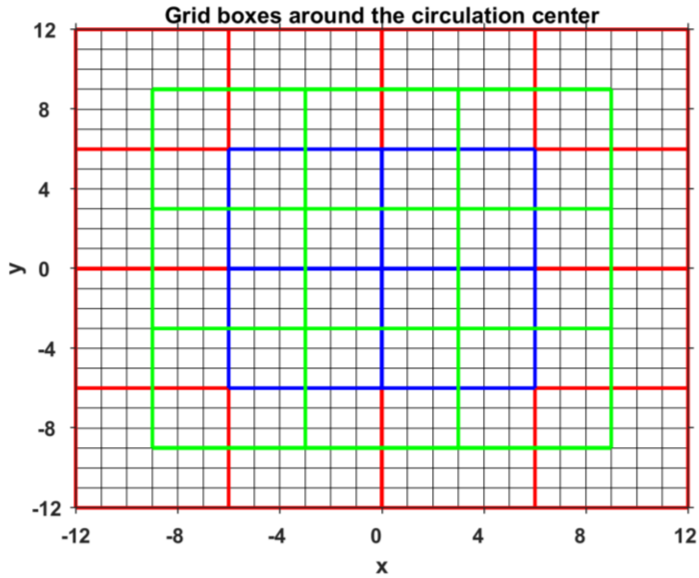

Figure 4. 2. The 24°×24° box region (in red grid boxes) around the TC center used for spatial composites, and the 12°×12° region (in blue grid boxes) is used for BDI analysis. Also, we use the 18°×18° box regions (in green grid boxes) for the OW parameter BDI analysis with the centered 6°×6° representing the regions for calculating the OW center and the remaining green boxes represent the OW surrounding variables.

4.3. Results

This section describes the storm-centered composite analysis of initial circulations that exists two days before TD declaration to identify structural differences and key limiting variables across each ocean basin that lead to the non-development of an initial circulation.

4.3.1 Geographical distribution of TS and TD genesis locations

Figure 4.3 shows the geographical distribution of genesis locations of IBTrACS and OWZP detected DV and ND depressions using the ERA-interim reanalysis. The genesis dots of the IBTrACS and the OWZP detected TSs have similar geographical locations in most of the ocean basins. Tory et al. (2013b) statistically tested the performance of this TC detection scheme when compared with the IBTrACS for 20 years from 1989-

2008. They showed globally 78% of IBTrACS storms are correctly detected using this scheme with a 25% false alarm rate. Also, of all the ocean basins the NI basin has poor performance, perhaps due to the NI being a unique land-enclosed basin and mis-identification of the monsoon lows as TCs. We perform a basin-wide analysis of the mean annual frequency bias of the OWZP detected storms and IBTrACS for the 30 years 1989-2018. Table 4.2 shows that the NA, ENP, and SP basins have underpredicted genesis compared to observations. These negative biases across the ENP, NA, and SP basins may be due to more short-lived and land influenced storms compared to the WNP and SI basins that have more long-lived storms over the open ocean waters, which are usually detected well by the OWZP scheme (Tory et al., 2013b).

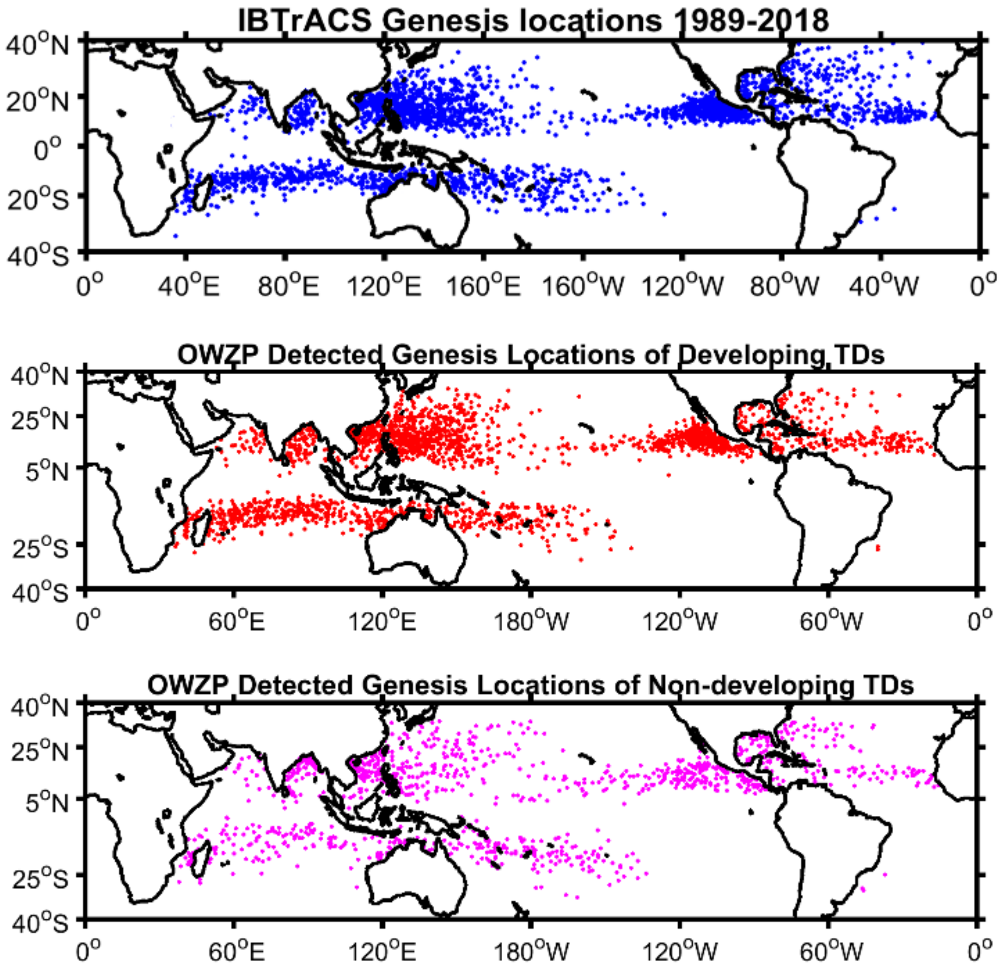

Figure 4. 3. Locations of IBTrACS genesis locations (top panel) and OWZP detected genesis locations of developing (middle panel) and non-developing TDs (bottom panel) in the ERA-Interim reanalysis data from 1989-2018.

Table 4. 2. Basinwise mean annual genesis frequency and percentage bias (per year) of the storms detected by OWZP scheme in the reanalysis data when compared with IBTrACS and * indicates 95% significance level. (% bias = (ERA-Interim minus observations/observations)*100)

BASIN	IBTrACS (TCs/year)	ERA-Interim (OWZP) (TCs/year)	% bias of OWZP storms
MDR_NA	6.8	5.6	(-) 17%*
ENP	15.5	14.2	(-) 8%
WNP	24.4	25.6	(+) 5%
NI	3.8	4.8	(+) 26%*
SI	15.4	14.9	(+) 5%
SP	8.7	6.6	(-) 24%*
AUS	9.3	7.5	(-) 18%*

4.3.2 Structural differences between developing and non-developing tropical depressions

Typically, a TS has a developed vortex core region from low to mid-level with cyclonic circulation along with anticyclonic outflow at upper levels associated with a secondary circulation (McBride and Zehr, 1981; Dunkerton et al., 2009; Wang et at., 2012). Addtionally, developing cases have a stronger shear-sheath layer from the low to mid-troposphere around the core of the storm. This shear sheath protects the vortex core region from external influences like the horizontal intrusions of drier air or lower values of opposite sign vorticity (Rutherford at al., 2015, 2018). Therefore structural variations, along with spatial distribution of environmental variables, are explored in detail 48 hours before TD declaration using spatial composites within a 24° ×24° box region around the depression center across all the ocean basins. This box region covers a larger spatial extent including the centered cyclonic circulation and the surrounding region which helps us to understand the influence of external environment on the circulation.

4.3.2.1 Northern Hemisphere (NH) basins

We examine the spatial composites of the atmospheric variables evolving from Day-2 to Day0 to differentiate the DV and ND cases across the NA basin. The differences in the spatial composites of relative humidity (RH) between the DV and ND cases in Figure 4.4a illustrate that developing cases have higher moisture at the center and surrounding regions compared to the non-developing cases from lower (850 hPa) to mid-troposphere (500 hPa). Also, the DV cases have fully enclosed streamlines around the center at 500 hPa shown in Figure 4.6, whereas in ND cases streamlines gradually diverge to the northwest of the center on Day0 with dry air surrounding the core region in the northwest side.

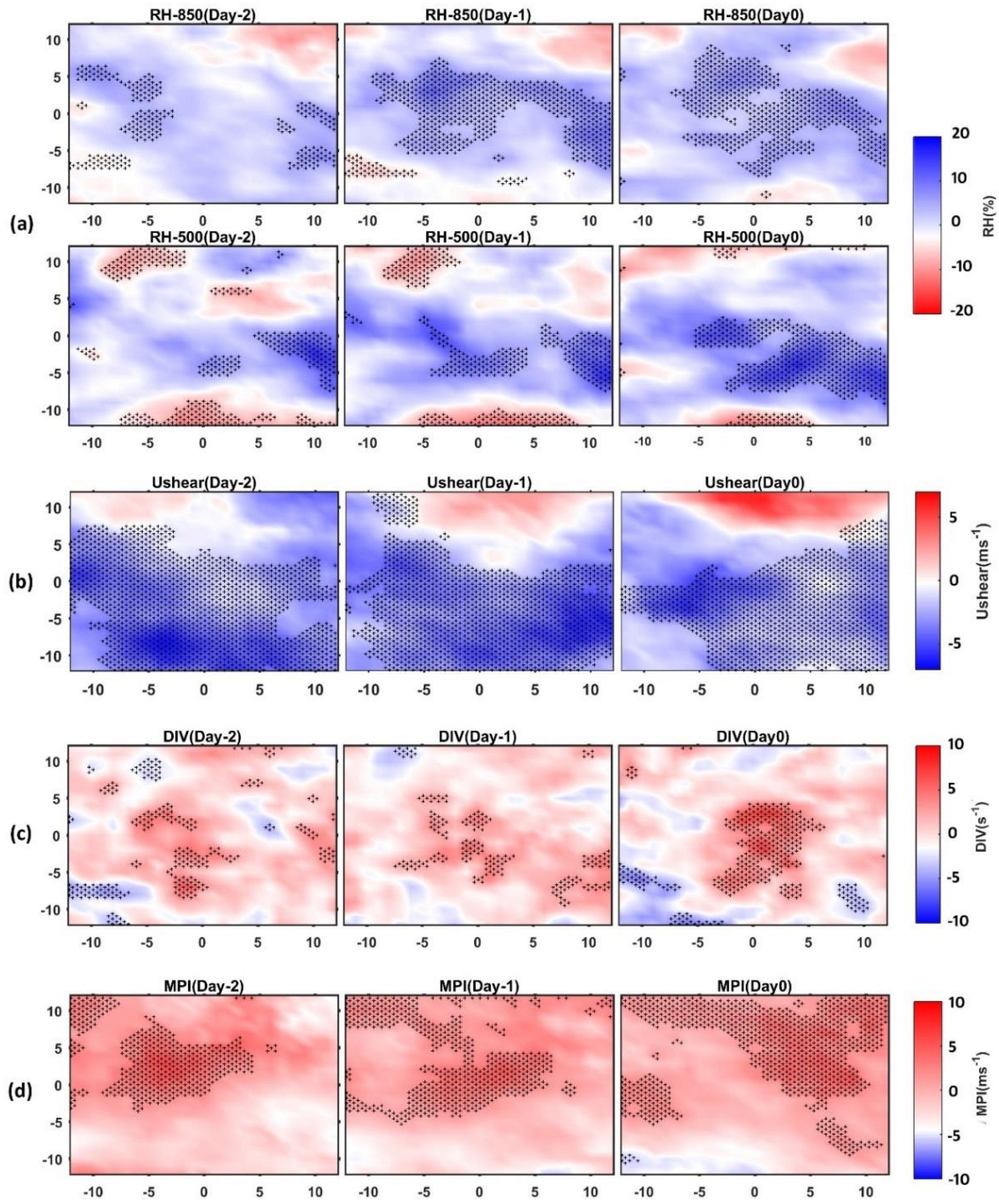

Figure 4. 4. Differences in spatial composites of developing and non-developing cases (DV minus ND) within the $24° \times 24°$ box region region around the circulation center over the North Atlantic basin, (a) 850 and 500 hPa RH, (b) Ushear, (c) DIV and (d) MPI. The stippling indicates the 95% significance level differences.

Figure 4.4b indicates that ND cases have unfavorable shear at the center and surrounding areas compared to the DV cases as it propagates from Day-2 to Day0 (Figure 4.7a shows averaged values of Ushear). The negative values of Uhsear indicate that ND cases have either stronger westerly shear (positive values) or weaker easterly shear (lower negative values). Also, Figure 4.4c shows that DV cases have increasing values of DIV from Day-2 to Day-0 at the center and surrounding regions compared to the ND cases indicating the development of stronger outflow jet associated with sustained convection in the DV cases. Similarly, Figure 4.4d shows that DV cases have higher values of MPI at the center and surrounding regions compared to the ND cases indicating greater thermodynamic instability in DV cases.

We further examine the spatial composites of the OW parameter overlaid with streamlines in the co-moving frame of the composite storm for both PDV48 and PND48 cases over the NA basin at 850, 700, 500, and 200 hPa levels. Figure 4.5 and Figure 4.6 shows that closed or nearly closed streamlines surround the cyclonic cores from low to mid-levels of the atmosphere. Each cyclonic circulation has a core of positive OW (strain-free rotation) surrounded by a band of negative OW (higher lateral strain, indicating the shear sheath), which extends from the lower (850 hPa) to the middle troposphere (500 hPa). A general decrease in the size and strength of the cyclonic circulation with height is evident, culminating in stronger anticyclonic outflow streamlines to the east of the center at 200 hPa for developing cases (Figure 4.6). In order to understand quantitatively the strength of centered circulation and the surrounding shear-sheath in the DV and ND cases we use box averages of the OW center and OW surrounding parameters (Table 4.3 and 4.4) as described in section 4.2.4. The OW center values of non-developing cases show that the cyclonic circulation is significantly weaker than in the developing cases (Table 4.3). The OW surrounding values shows a development of a more pronounced shear-sheath (higher negative OW values) in DV cases on Day0 at both lower and mid-levels (Table 4.4). The stronger the circulation, the stronger the OW core, and the stronger the shear sheath develops in DV cases.

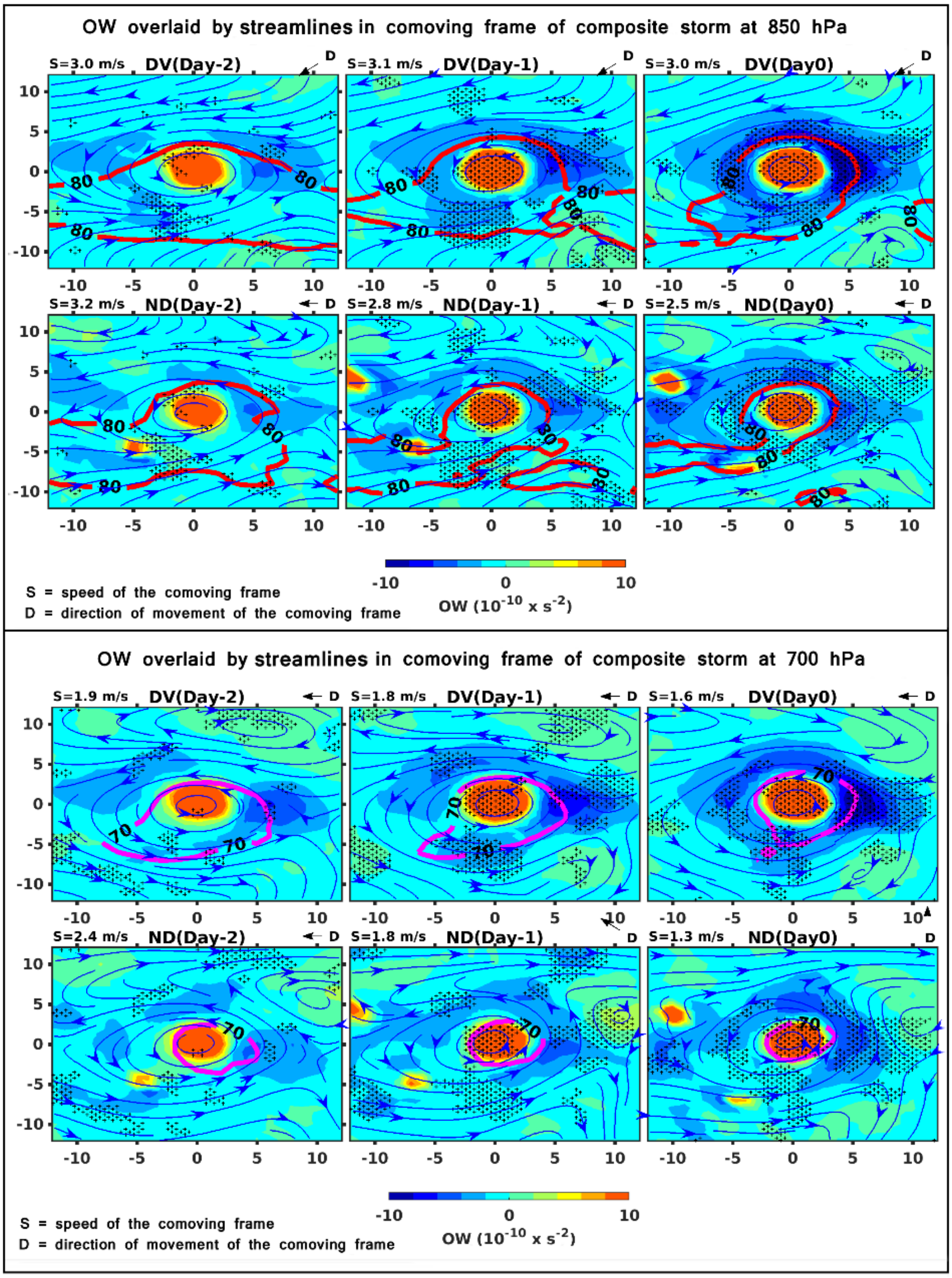

Figure 4. 5. OW overlaid by streamlines computed in a co-moving frame for DV and ND cases at 850 and 700hPa over the North Atlantic basin. The x and y-axis indicating the distance from the center in degrees, the red and magenta contours indicate 80% and 70% RH, respectively and stippling indicates the 95% significance level differences between DV and ND cases.

Table 4. 3. Basinwise bias of OW center mean values in a $6° \times 6°$ box region around the circulation center (developing minus non-developing) at both 850 and 500 hPa levels (* indicates that the differences are statistically significant at the 95% significance level).

BASIN	Day-2 (850 hPa)	Day-1 (850 hPa)	Day0 (850 hPa)	Day-2 (500 hPa)	Day-1 (500 hPa)	Day0 (500 hPa)
MDR_NA	1.5*	6.2*	12.7*	1.1*	1.6*	7.1*
ENP	1.8*	2.5*	7.4*	0.2	2.9*	7.0*
WNP	2.4*	7.0*	15.5*	2.3*	5.6*	14.2*
WNP_P	3.4*	3.8*	9.7*	0.46	2.1*	7.6*
NI	0.34	0.32	5.9*	0.36	1.1	5.6*
SI	3.0*	5.4*	13.1*	0.5	2.1*	8.5*
SP	-4.1	-0.1	12.5*	-1.3	1.3	7.6*
AUS	4.9*	9.4*	17.4*	-1	0.3	5.4*

Table 4. 4. Basinwise bias of OW surrounding mean values with a $18° \times 18°$ box region by excluding the center $6° \times 6°$ box region around the circulation center (developing minus non-developing) at both 850 and 500 hPa levels (* indicates the differences are statistically significant at 95% significance level).

BASIN	Day-2 (850 hPa)	Day-1 (850 hPa)	Day0 (850 hPa)	Day-2 (500 hPa)	Day-1 (500 hPa)	Day0 (500 hPa)
MDR_NA	-0.5*	-0.5*	-1.2*	-0.16	-0.13	-0.85*
ENP	-0.47*	-0.63*	-1.4*	-0.17	-0.46*	-1.2*
WNP	-0.26*	-0.63*	-1.2*	-0.3*	-0.7*	-1.4*
WNP_P	-0.36*	-0.23	-0.8*	-0.36*	-0.07	0.36
NI	0.003	0.8	-0.4	0.47	0.2	-0.6*
SI	-0.15	-0.46*	-1.3*	-0.06	-0.44*	-1.1*
SP	1.0	-0.2	-1.5*	0.7	-0.3	-0.7*
AUS	-0.1	-1.2*	-1.8*	-0.02	-0.56*	-0.8*

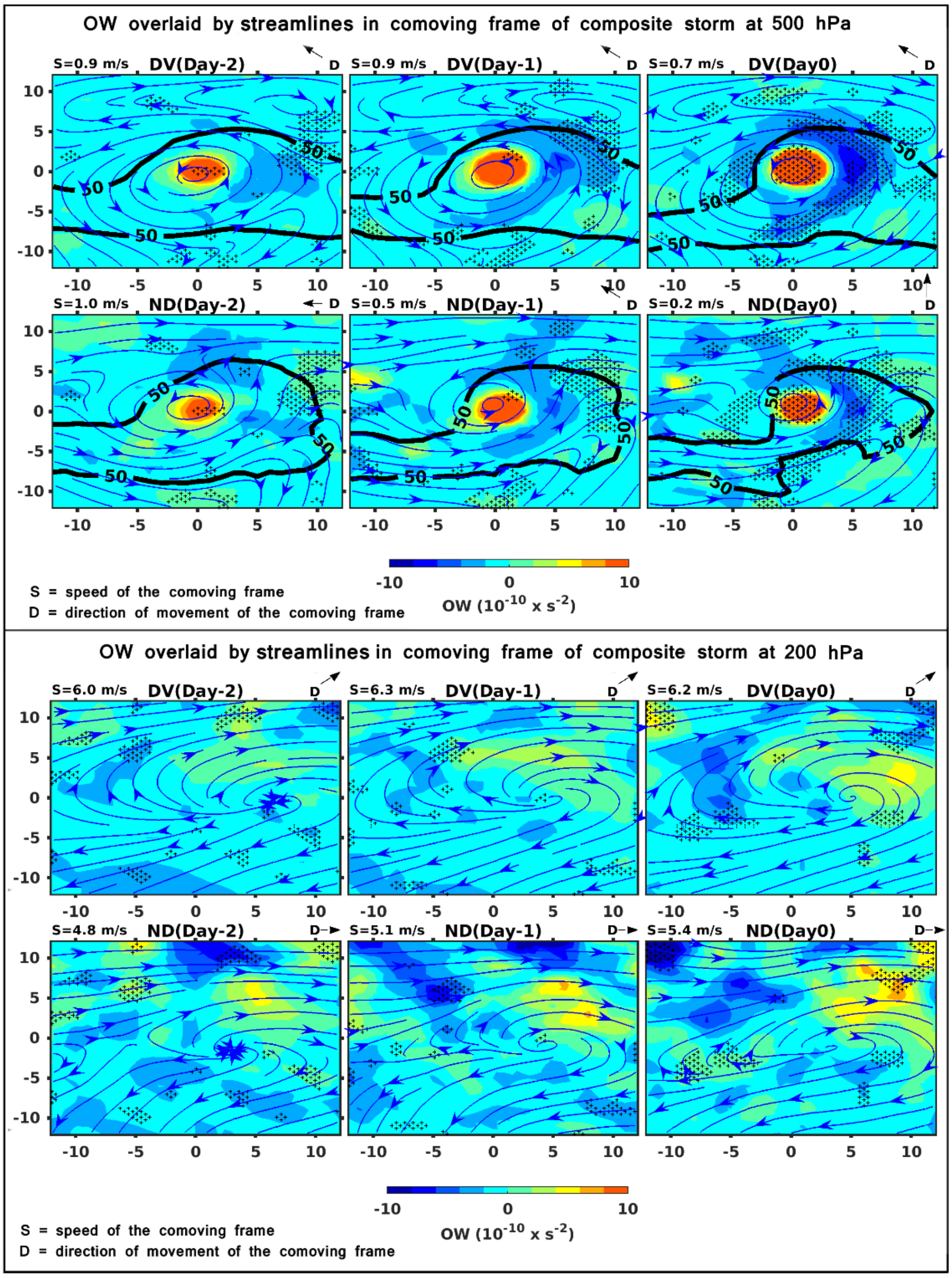

Figure 4. 6. The same as Figure 3 at 500 and 200 hPa; the black contour indicates the 50% RH.

As the depression develops, a stronger shear sheath layer forms in the north, east, and west side of the storm center enclosing the core region in the developing cases. This layer offers higher resistance for the surrounding dry air or opposite signed vorticity to penetrate towards the center. In contrast, non-developing cases have a weaker shear sheath layer enclosing a smaller area around the vortex center at 700 and 500 hPa levels (Figure 4.5 & 4.6), with a relatively weaker layer on the west side of the storm. The non-developing cases in Figure 4.6 show an intrusion of dry air at the 500 hPa level (as indicated by the 50% RH values in black contour) into the vortex core region on day0 from the northwest side of the depression may be due to weaker shear sheath layer.

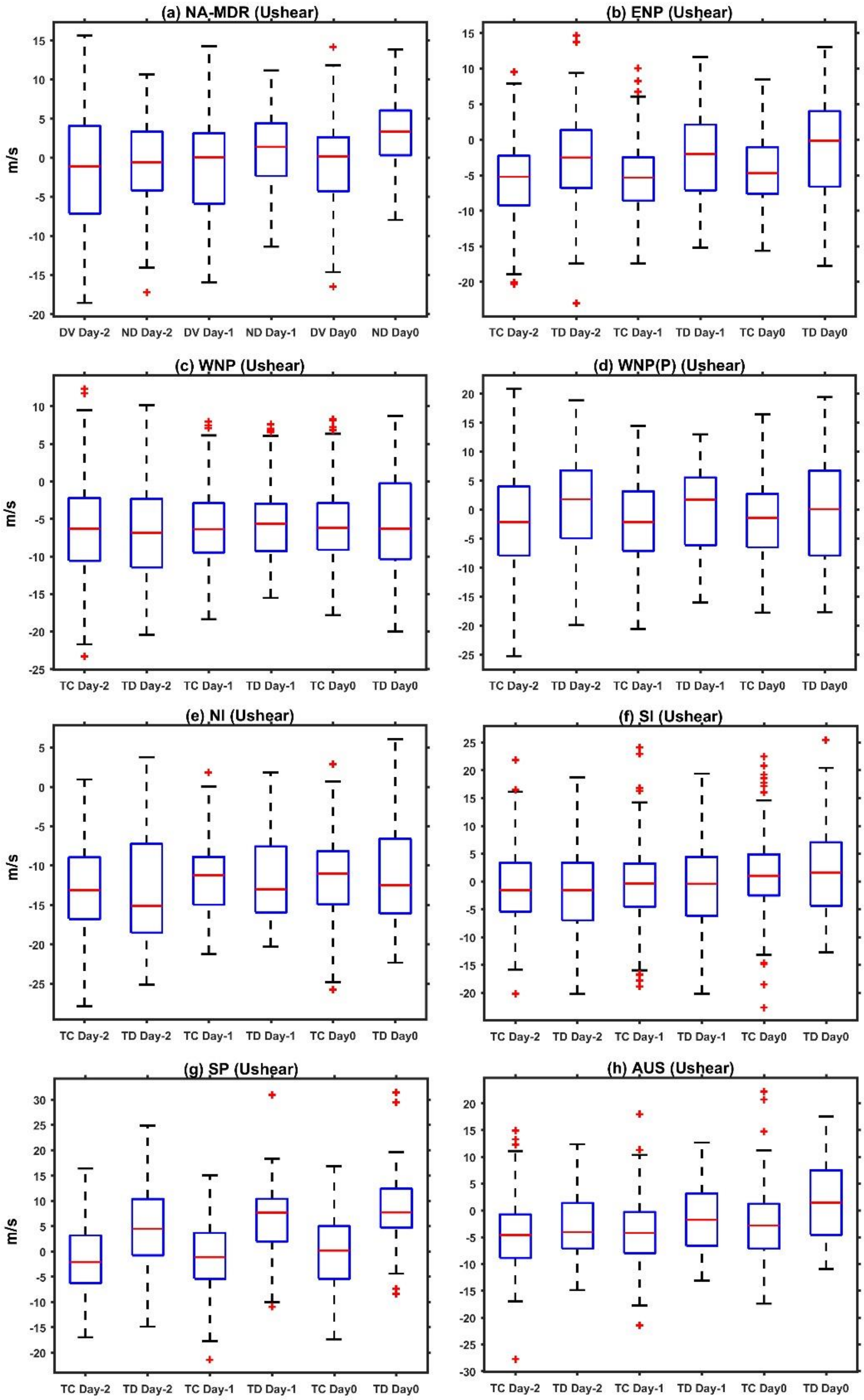

Figure 4. 7. Box-plots of averaged Ushear for DV and ND cases within a 12°×12° box region. The horizontal line of the box indicates median of the distribution, the left, and right box edges show 75th and 25th percentiles, the whiskers indicate maximum and minimum values and the red crosses indicate outliers. The statistical significance of Ushear is given in Figure 4.9.

Additionally, Figure 4.7a shows that there is an increased mean westerly shear in the ND cases as the circulation approaches Day0. Shu et al. (2014) noted that westerly shear is detrimental to the development of a TC as it draws in cold, dry air from the poleward side of the storm. Brammer and Thorncroft (2015) noted that moisture in the north and northwest side is an important factor for developing storms from AEWs. Therefore a less protective vortex having a weaker shear-sheath layer particularly in the northwest side of the center, westerly shear, and lower atmospheric moisture content leads to the non-development of initial TDs in the NA basin.

A similar spatial composite analysis is performed in other NH basins (ENP, WNP, WNP_P), as shown in Figure A.1, Figure A.2 and Figure A.3 (in appendix A). We separate the WNP basin into equatorward WNP (0-15N) and poleward WNP_P (15-30N) regions due to different TD failure rates (Tory et al., 2018). These basins also show that ND cases have weaker shear sheath layers leading to dry air intrusion due to weak mean easterly shear (except in WNP) shown in Figure 4.7 (b, c, d), thereby disrupts the deep convection suppressing the development of initial depression. Also, the box-averaged values of OW center and OW surrounding variables show that DV cases have higher positive centered core region with a stronger negative surrounding shear sheath (Table 4.3 and 4.4). In the NI basin (Figure A.4), the non-developing cases have disorganized circular structure of the centered positive OW core region at middle levels of the troposphere with significantly lower magnitude on Day0 (Table 4.4). Also, the ND cases across the NI basin have weaker shear sheath layer at mid-levels (500hPa) on Day0 but do not indicate the mid-level dry air intrusion into the center due to favorable mean easterly shear (Figure 4.7e). Therefore, a lack of organized structure of centered circulation in the ND cases leads to the non-development of the circulation across the NI basin.

4.3.2.2 Southern Hemisphere (SH) basins

The spatial composites of the atmospheric variables in the SH indicate a larger spatial extent with a higher magnitude of these variables in DV cases compared to the ND cases (not shown). Figures 4.8, A.5 A.6 show the OW composites overlaid with the streamlines over the SP, SI, and AUS basins, respectively. These basins also depict similar structural features as observed in the NH basins with a stronger cyclonic circulation at low-levels and anti-cyclonic at upper levels. The DV cases have significantly higher OW center values and a stronger shear sheath with stronger negative OW surrounding values from the low (850 hPa) to mid-troposphere (500 hPa) (Table 4.3 and 4.4). The ND cases have weaker mean easterly shear or increased westerly shear (Figure 4.7g, h) in SP and AUS basins compared to the DV cases. In contrast, the SI basin has similar mean shear for both DV and ND cases (Figure 4.7f). Across these ocean basins, the non-developing cases have a weaker shear-sheath layer on the west side of the depression indicating mid-level dry air intrusion (with 55-60 % RH values) to the center of the depression impeding the development of TD. Therefore a less protected vortex having a weaker shear-sheath layer with unfavorable shear disrupts the intensification of the initial circulation to a TS.

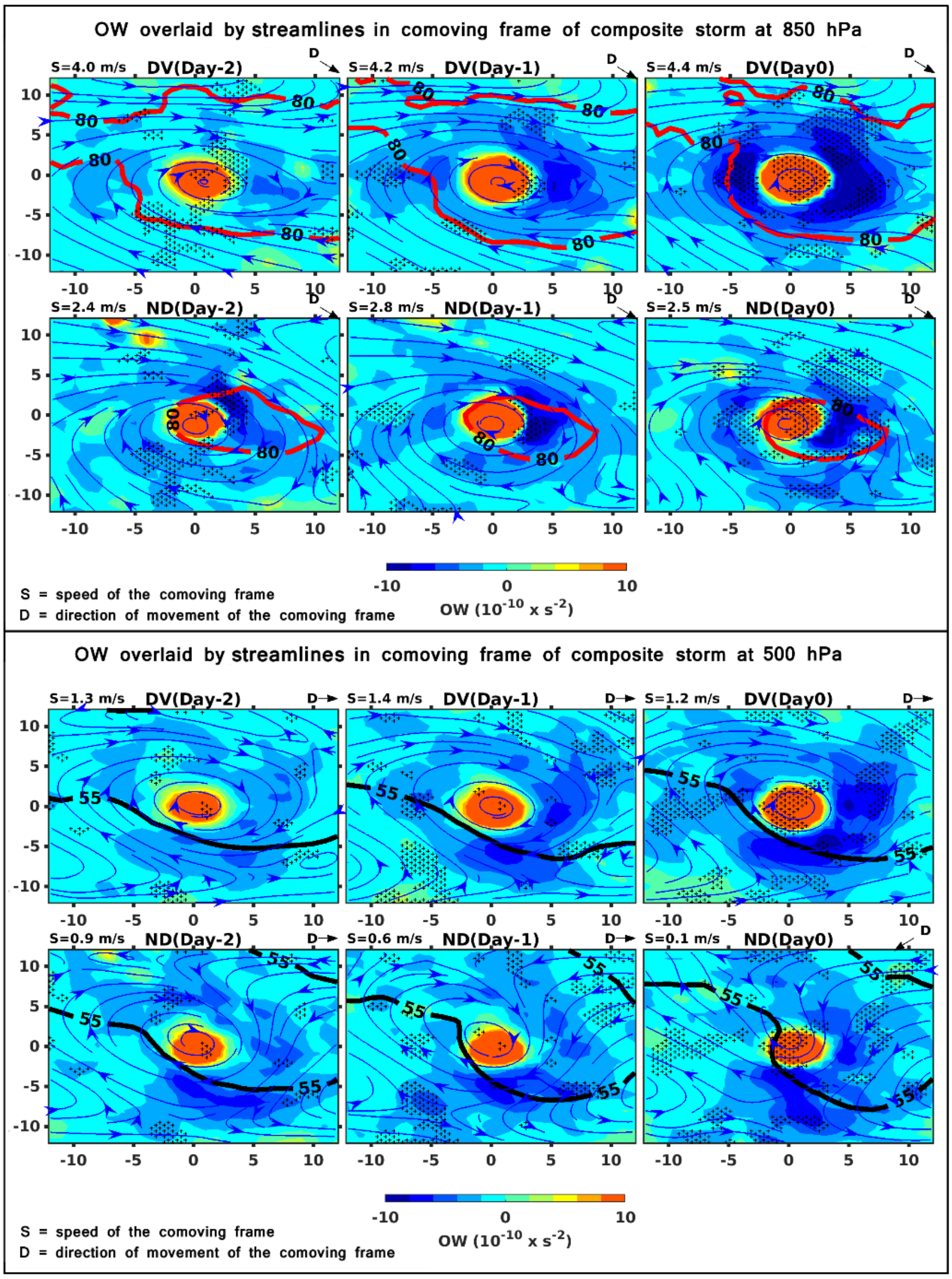

Figure 4. 8. The same as Figure 4.5 at 850 and 550 hPa across the SP basin, the red and black contour indicate the 80% and 55% RH.

4.3.3 Quantification of differences from statistical distributions using the BDI

We computed the statistical distributions of the environmental variables averaged within a $12^{o}\times 12^{o}$ box region on Day-2, Day-1, and Day0 to reveal the variables that show significant differences between the DV and ND cases (A.7 to A.10). Although statistical distributions identify the variables showing significant differences between DV and ND cases, they do not give the order of the important variables. The BDI index, which uses the mean and standard deviation of the samples, describes the relative importance of variables in distinguishing DV from ND and is calculated for each large-scale variable on Day-2, Day-1, and Day0. In particular, this index identifies which quantities/processes are most detrimental to TS formation. Here, we use both Day-2 and Day0 BDI values to rank the order of prominent variables (Table 4.5), including their statistically significance, that hinder/favor the intensification of an initial circulation. The BDI values of 700 hPa RH correlate with TCW, and BDI values of Omega correlate with DIV across all the ocean basins, therefore we include only TCW and Omega in the BDI discussion.

4.3.3.1 NH basins

Across the NA basin, Figure 4.9a shows that as the DV circulation approaches Day0, we observe increased updrafts (Omega) and low-level vorticity (RVor) suggesting a developing circulation due to availability of sufficient tropospheric moisture (TCW) from Day-2. This higher moisture favors sustained convection by increasing the efficiency of upward mass flux and thereby increases the strength of the circulation. Also, as the circulation approaches Day0, we observe that DV cases have favorable mean shear compared to the ND cases that have mean westerly shear (Figure 4.7a). The lower values of moisture and unfavorable shear leads to disruption of deep convection that suppresses the development in ND cases. Also, the BDI values show that DV cases have higher thermodynamical instability measured in terms of MPI.

This analysis shows that tropospheric moisture is an important limiting factor across this basin for an initial circulation to develop into a TS. A few studies noted that the Saharan Air Layer (SAL) acts as a modulating influence for the TC activity across this basin, as it suppresses the convection, entrains dry air to the center of the circulation, and increases vertical wind shear leading to non-development of the circulation (Dunion & Veldon, 2004). Other studies noted that pre-existence of a dry troposphere or the advection of the dry air by the AEWs, large-scale subsidence, TC-TC interactions or TC interaction with environment creates a dry atmosphere and higher vertical wind shear that disrupts storm formation (Hopsch et al., 2010; Braun, 2010; Brammer & Thorncroft, 2015; Fowler & Galarneau, 2017). Therefore, the availability of sufficient tropospheric moisture around the disturbance with favorable shear supports the development of the incipient circulation.

From the BDI values across the ENP basin (Figure 4.9b), as the circulation progresses from Day-2 to Day0, DV cases have increased TCW at the center and surrounding of the circulation along with favorable mean easterly shear (Ushear) compared to the ND cases (Figure 4.7b). The lower BDI values of RVor indicate that both DV and ND cases have similar low-level vorticity. Therefore, the lack of sufficient tropospheric moisture and unfavorable shear suppresses the development of an ND case across the ENP basin. The limiting variables for TS formation identified from this analysis also agree with other studies at different timescales. For instance, at interannual timescales across the ENP basin, Zhao and Raga (2015) noted that mid-level relative humidity acts as a significant contributor to TS formation in inactive years (i.e., years observed with low TC activity) and vertical wind shear is the most important parameter in the active years (high TC activity).

The WNP basin is the most active basin for TC formation across the global ocean basins due to favorable regional environmental conditions. There is a strong monsoon circulation that provides sufficient low-level convergence and higher SSTs due to the warm pool region favor TC formation. Nevertheless, there exist environmental differences between DV and ND cases due to the prevalence of different synoptic-scale disturbances across the basin (Li, 2006; Li & Fu, 2006; Fu et al., 2012). The higher BDI values from Figure 4.9c and 4.9d shows that DV cases have stronger vorticity (RVor) and increased updrafts (Omega) compared to the ND cases. These higher values of vorticity in DV cases from Day-2 indicate that the initial circulation originated in a strongly convective region, which favors further intensification by deep convection. Also, the thermodynamic favorability measured in terms of MPI and TCW is higher for DV cases compared to the ND cases across the entire basin.

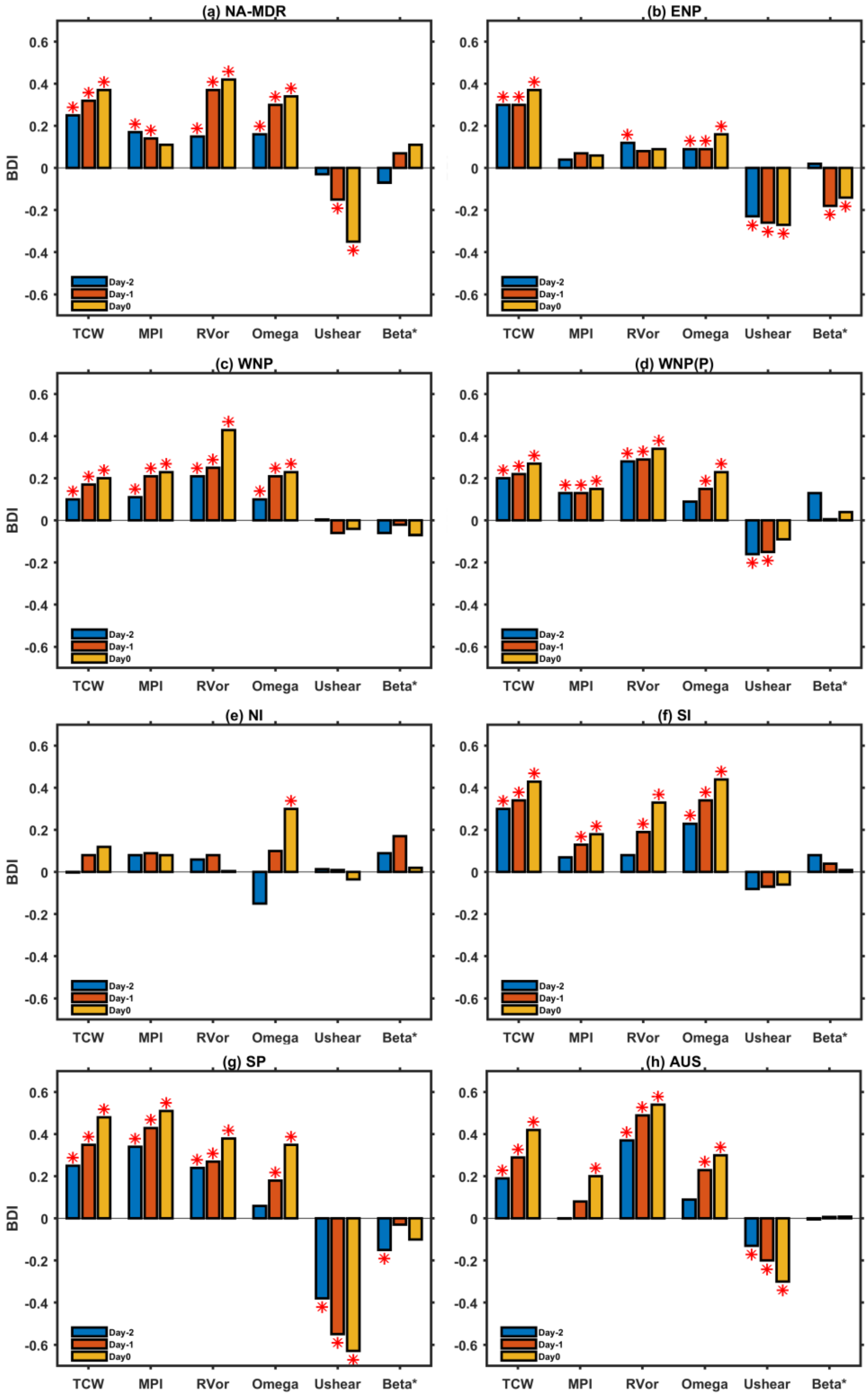

Figure 4. 9. Basinwide BDI index of environmental variables for PDV48 and PND48 cases on Day-2, Day-1, and Day0. Positive values of the BDI index for TCW, MPI, RVor, and Omega indicate more favorable conditions for development, while negative values of Ushear, Beta* are more favorable (* indicates the 95% significance level).

In addition, across the WNP_P basin, DV cases have favorable mean easterly shear (Ushear) compared to the ND cases (Figure 4.7d). The initial circulation originating in a weakly convective region with a less thermodynamically unstable environment leads to the non-development across the basin. Therefore, the dynamical condition is more important for the development of an initial circulation across the WNP basin (McBride& Zehr, 1981; Lee, 1989 a, b; Fu et al., 2012). Across the NI Ocean, from the BDI values of Figure 4.9e, we observe that DV cases have stronger updrafts (Omega) on Day0 compared to the ND cases. A study by Tsuboi and Takemi (2014) have noted that Omega acts as a major contributor the changes in the TC frequency during the active phase of the Madden-Julian Oscillation (MJO) across this basin. The NI Ocean is a unique basin compared to other basins as it doesn't observe significant differences for most of the large-scale variables perhaps due to poor performance of TC frequency characteristics using the OWZP scheme (Tory et al., 2013b).

Table 4. 5. The order of importance of variables in each ocean basin based on the statistically significant values at the 95% significance level of the BDI index at Day-2 and Day0 (the day of TD declaration).

BASIN	Order of variables (Day-2)	Order of variables (Day0)
NA-MDR	TCW, MPI, Omega, RVor	RVor, TCW, Ushear,Omega
ENP	TCW, Ushear, RVor, Omega	TCW, Ushear, Omega, Beta*
WNP	RVor, Omega, MPI, TCW	RVor, Omega, MPI,TCW
WNP-P	RVor, TCW, Ushear, MPI	RVor, TCW, Omega, MPI
NI	--	Omega
SI	TCW, Omega	TCW, Omega, RVor, MPI
SP	Ushear, MPI,TCW, RVor, Beta*	Ushear, MPI,TCW, RVor, Omega
AUS	RVor, TCW, Ushear	RVor, TCW, Ushear, Omega, MPI

4.3.3.2 SH basins

The BDI analysis over the SH basins show some similarities with the NH basins in the significant limiting variables with a difference in the order of the variables (Table 4.5).

Figure 4.9f shows that the DV cases in SI Ocean has higher TCW, Omega and RVor compared to the ND cases. Over the SP basin, the BDI values in Figure 4.9g show that DV cases have favorable thermodynamic conditions and mean easterly shear compared to the ND cases (Figure 4.7g). Therefore both thermodynamical (MPI & TCW) and dynamical (Ushear) conditions act as the significant influencing variables for the development of a TD across the SP basin. From the BDI values across the AUS basin (Figure 4.9h), we observe that DV cases have higher RVor and TCW compared to the ND cases. Also, the DV cases have favorable mean easterly shear compared to ND cases (Figure 4.7h). Earlier studies at different time scales show that sea surface temperatures, vertical wind shear, RH, and RVor are significant influencing variables across SH basins (Camargo et al., 2007; Ramsay et al., 2008; Kuleshov et al., 2009; Chand & Walsh, 2011; Dowdy et al., 2012; Liu & Chan, 2012).

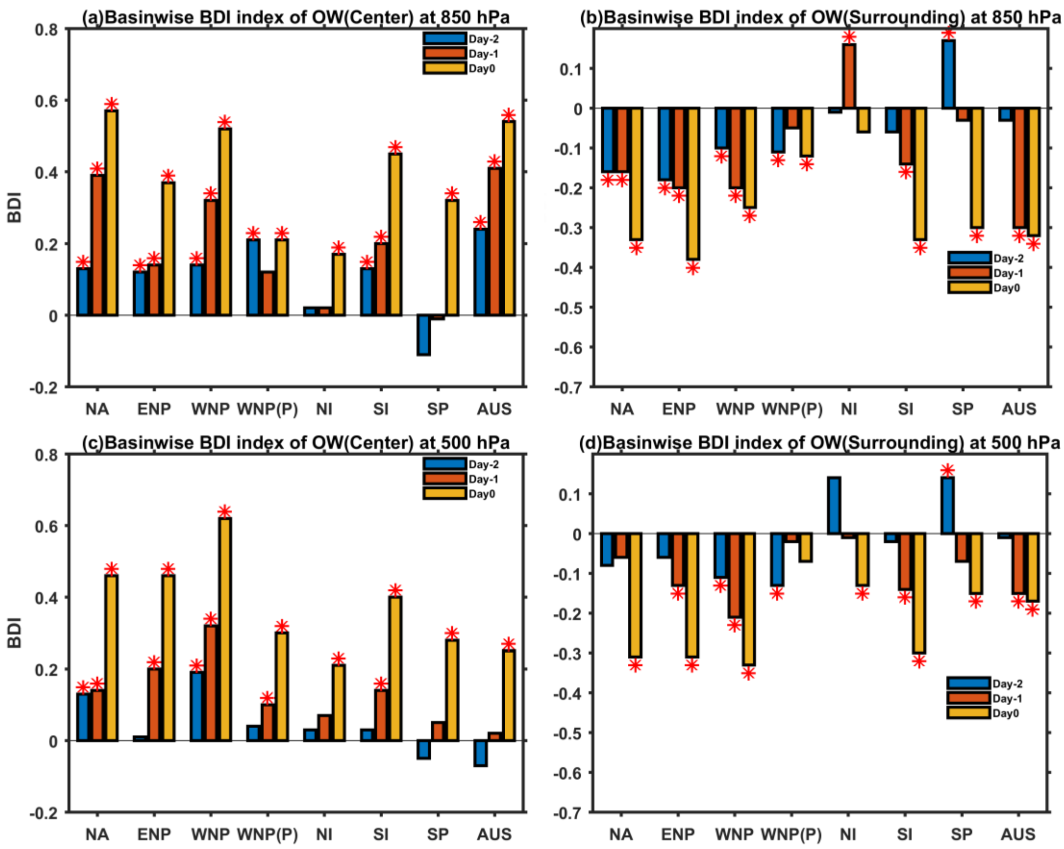

Figure 4. 10. Basinwide BDI index of OW center and OW surrounding for DV and ND cases at 850 and 500 hPa levels on Day-2, Day-1, and Day0. Positive values of OW center variable and negative values of OW surrounding variable are more favorable for TS formation (* indicates a 95% significance level).

In addition to the above large-scale variables, we estimate the BDI index of the Beta*, OW center, and OW surrounding variables across these ocean basins. The Beta* variable which shows the barotropic instability regions is found to differentiate the DV and ND cases across the ENP and SP basins (Figure 4.9b, g). Also, from the BDI index values of the OW center variable (Figure 4.10), we observe that developing cases have higher magnitudes of OW center across all the ocean basins, indicating strong low-deformation vorticity regions. Also, the BDI values of the OW surrounding variable (Figure 4.10) show that DV cases have stronger negative values compared to the ND cases in most of the ocean basins indicating a stronger shear-sheath that protects the vortex from external negative influences. Therefore, in addition to the above large-scale variables, the Beta*, OW center, and OW surrounding parameters can be used as predictors across the ocean basins to discriminate the DV and ND cases.

4.4. Discussion

The circulation centered spatial composites of the RH, DIV, and MPI variables evolving from Day-2 to Day0 show that developing cases have more considerable spatial extent surrounding the vortex core of a region of the higher magnitude of these variables (Figure 4.4). In contrast, the non-developing cases have less spatial extent and lower values of the environmental conditions from Day-2. The Ushear spatial composites show that non-developing cases have either strong mean westerly shear or weak easterly shear unfavorable for TS formation. The composites of the streamlines overlaid with shaded OW parameter and RH contours help us to identify the structural variations between DV and ND cases (Figure 4.5 & 4.6). These composites show that non-developing cases have lower values of centered OW along with a weaker band of negative OW values (the shear sheath) surrounding the core region from lower-levels to mid-levels compared to the DV cases across most of the basins (Table 4.3 & 4.4). In contrast, developing cases have stronger OW core regions and a stronger surrounding shear sheath layer from lower to mid-levels. Also, DV cases have mean easterly wind shear, which draws the warm moist air from the equatorward side of the storm (Shu et al., 2014). In ND cases, the increased westerly shear or weak easterly shear, as well as a weaker shear sheath in the west or northwest side of the vortex center on day0, may lead to horizontal entrainment of surrounding drier air at mid-troposphere (500 hPa) in some of the basins. Drawing the cold, dry air from the side with a weaker shear-sheath layer may disrupt the storm formation in ND cases by reducing the overall moisture content and affecting the sustained convection. In contrast to the above results, across the NI basin, both DV and ND cases have easterly shear and weaker shear sheath region. However, an unorganized structure of the OW centered values leads to the non-development of the initial circulation in this basin. Therefore, a more compact vertical structure from lower levels to mid-levels with a stronger centered vortex along with a stronger shear-sheath layer protects the incipient circulation.

The BDI index illustrates the relative importance of the thermodynamical and dynamical conditions on the development of TS from an initial circulation. The BDI magnitude (Figure 4.9) from day-2 to day0 shows that environmental differences between DV and ND gradually increase from Day-2 to Day0. These differences indicate that either DV cases are in favorable environmental conditions or there is the increased influence of destructive forces in ND cases. Table 4.5, which shows the order of important variables, notes slight differences in the order of these influencing variables on day-2 and day0, with similar environmental variables appearing to be prominent on both days but with a larger magnitude on day0. Figure 4.9 and Figure A.11 gives an overview of the role of environmental variables on the development of an initial circulation from the BDI values. It shows some agreement between the global ocean basins. For instance, TCW, RVor, and Omega variables show significant differences between DV and ND cases with a higher magnitude of BDI index in almost all the ocean basins. In contrast, MPI, Ushear, and Beta* show differences and seem to play a role in fewer ocean basins.

Although the current analysis does not illustrate the precise convective processes leading to the development of a TS, Figure 4.9 shows that the DV cases have higher moisture over different ocean basins indicated by positive BDI values of TCW. The higher tropospheric moisture favors sustained deep convection that leads to a strong low to mid-level vorticity concentration due to increased convective heating (Wang et al., 2012; Wang, 2014). In contrast, the ND cases that has drier troposphere lead to weaker deep convection due to weaker updrafts and downdrafts (James & Markowski, 2009; Kilroy & Smith, 2013). Some earlier studies noted that the thermodynamical variable MPI is a vital predictor to determine the TS formation using statistically derived genesis potential indices (Emanuel & Nolan, 2004; Bruyère et al., 2012). This MPI acts as one of the significant influencing factors across the WNP, WNP_P, NA, SP, and AUS basins indicating favorable thermodynamic instability in developing cases.

In the examination of the dynamical variables, Figure 4.9 shows that the ND cases have lower values of RVor and Omega across all the ocean basins except over the ENP and the NI. The higher magnitudes of RVor and Omega indicates sustained convection and abundant low-level convergence, favorable for TS formation (Tory & Frank, 2010; Montgomery & Smith, 2010). Also, ND cases have weaker mean easterly shear or increased westerly shear across the ocean basins except in the WNP (0-15N), NI, and the SI basins (Figure 4.7), indicating that weak mean easterly shear or increased westerly shear may be detrimental to TS formation. Ritchie and Frank (2007) have shown that easterly shear could offset the northwesterly shear induced by the vortex due to the beta effect as the possible reason for westerly shear being detrimental to development. In contrast to the above differences, there are no significant environmental differences between DV and ND cases across the NI basin except for Omega. The BDI analysis shows some similarities between the ocean basins with similar prominent variables. However, the order of these variables varies from basin to basin, perhaps due to distinct large-scale disturbances, different modulating influences and differences in the background environmental conditions. Also, DV cases have higher positive BDI values of OW center variable and higher negative BDI values of OW surrounding variable favorable for TS formation in most of the basins (Figure 4.10). Therefore, observing the OW values at the center and surrounding the circulation center from low to mid-troposphere in numerical weather prediction models or including the OW center and OW surrounding variables in statistical models may improve the forecast accuracy of the TS formation from a depression stage vortex.

4.5. Conclusion

The differences in the large-scale disturbances and the mean flow conditions across the global ocean basins may lead to differences in the relative contribution of large-scale variables to TS formation. Therefore, in this study, we perform a basinwide statistical analysis on the environmental differences between developing and non-developing depressions by following the depression two days before its declaration in a Lagrangian manner. We use a previously developed TC detection technique to detect both developing and non-developing tropical circulations in the ERA-Interim reanalysis. The spatial composites of atmospheric variables and streamlines from Day-2 to Day0 help us to monitor the evolving flow features around the depression in each ocean basin. We employ a statistical BDI index to quantitatively identify the order of important atmospheric variables that distinguish the developing and non-developing depressions across each ocean basin. We find the following main conclusions:

- The spatial composites of the environmental variables show that DV cases have more favorable conditions at the center and surrounding compared to the non-developing cases from two days before TD declaration (Figure 4.4).
- The DV cases have a protective vortex with a strong positive OW core region and the development of a surrounding region of stronger negative OW, the "shear sheath," from the low to mid-troposphere as it approaches TD declaration stage (day0). In contrast, non-developing cases have less protective vortex with a weaker shear sheath in the northwest/west side of the depression leading to the intrusion of cold, dry air to the center of the circulation at mid-levels (500 hPa) on day0 (Figure 4.5 & 4.6).
- The ND cases have either mean westerly shear or weaker mean easterly shear in some ocean basins (Figure 4.7), which leads to dry air intrusion disrupting the intensification of the vortex.
- Based on the BDI values on day-2 and day0 (Table 4.5), TCW and RVor act as significant limiting factors for TS formation over the NA; RVOR over the WNP basins; TCW, and Ushear over the ENP basin; and Omega over the NI basin. In SH, TCW, and Omega act as main limiting factors for TS formation over the SI; Ushear, and MPI over the SP; and RVor and TCW over the AUS basin. This BDI analysis shows that TCW and RVor are important in most of the ocean basins and Ushear, which shows the large-scale influence is prominent in a few ocean basins (Figure 4.9).
- The BDI values of the OW center and OW surrounding variables show that DV cases have stronger core and stronger shear sheath from Day-2 compared to the ND cases in most of the ocean basins (Figure 4.10).

This study provides a set of large-scale environmental conditions that act as limiting factors for the development of an initial TD to TS across each ocean basin. We observe similar structural differences between developing and non-developing TDs across most of the ocean basins. In contrast, there are varying contributions of environmental variables that may be due to the distinct large-scale disturbances and their interactions with the basin's mean flow conditions. This study identifies additional predictors for the TS formation forecast in addition to earlier large-scale variables. Therefore, the inclusion of both OW center and OW surrounding variables in building statistical forecast models of TS formation from the initial depressions along with other atmospheric variables such as TCW, MPI, RVor, Omega, Ushear may improve TS forecasts at the synoptic scale. Future work focuses on the application of this methodology to forecast analysis fields to observe whether similar large-scale variables act as impediments for the TS formation from an initial circulation and establish a statistical TS formation forecast model. Also, more detailed analysis will be necessary to understand the variations in the order of important variables across the global ocean basins which improves our understanding of the relationship between large-scale environmental variables and TS formation.

Chapter 5. Prediction of tropical storm formation from initial tropical depressions using machine learning algorithms

This chapter addresses the research objective two of the thesis: "Application of different machine learning algorithms to predict developing tropical depressions with background environmental conditions as the predictors" to be submitted to Natural Hazards.

Abstract

The current study aims to develop a statistical prediction scheme for tropical storm (TS) formation across the global ocean basins from the pre-genesis environmental conditions of developing and non-developing tropical depressions (TDs). Here, we use the Okubo-Weiss Zeta Parameter (OWZP) detection and tracking scheme to detect the initial TDs in the ERA-interim reanalysis from 1989-2018. Three machine learning (ML) algorithms, namely Decision Trees (DT), Random Forests (RF), and Support Vector Machines (SVM) are employed to predict the developing TDs with the environmental predictors. The RF and SVM algorithms performed better with an F_score of about 0.8, an 80% hit rate, and around a 20% false alarm rate across global ocean basins compared to the DTs. The ML models built with equivalent variables of ventilation index shows comparable results with the actual model that have other additional variables indicating that ventilation theory is an important theory for TC formation from initial tropical depressions. The order of important variables for developing cases from the algorithms varies across basins, perhaps due to differences in the sources of tropical disturbances and their modulating influences. The DTs gives the quantitative values of co-influencing variables favorable for TS formation in the form of decision rules. Assuming the evolution of TS formation from an initial depression as a Markov decision process (MDP), a discrete Markov model is used to estimate the certainty of maintaining the favorable conditions obtained from the DT rules. This discrete dynamical Markov model shows that developing tropical depressions have a maximum likelihood of maintaining favorable conditions from 48 hours before TD formation compared to non-developing cases that involve atmospheric state transitions into less favorable environmental

conditions. Therefore, the DTs and the MDP's hybrid approach improve the certainty of the TS formation forecast from DTs.

5.1. Introduction

Tropical cyclones (TCs) are natural atmospheric phenomena that cause socio-economic losses, mainly in coastal regions. TC formation involves complex spatial and temporal scale interactions between the large-scale and mesoscale features and surrounding environmental conditions (Gray, 1998; Hendricks et al., 2004; Montgomery et al., 2006; Dunkerton et al., 2009; Bell & Montgomery, 2019). TCs form in closed recirculating regions of large-scale disturbances (i.e. tropical waves) under certain favorable thermodynamical and dynamical conditions (Gray, 1968, 1975; Dunkerton et al., 2009). Although the tropical disturbances provide some necessary conditions such as vorticity and favor convection, some of the tropical disturbances lead to the development of tropical cyclones (Avila, 1992; Avila et al., 2000; Russell et al., 2017). The non-development of disturbances indicates that TCs are mainly dependent on the surrounding environment, paving the path for predicting the formation based on the background environmental conditions. For instance Dunion and Veldon (2004) note that the Saharan Air Layer acts as a major regulator for TC activity leading to the increase in the destructive forces such as increased wind shear and dry air intrusion that lead to the non-development of the initial circulation. The formation of an initial vortex (i.e. genesis) from a synoptic-scale disturbance is not fully understood due to generally sparse observations and multi-scale interactions (Rogers et al., 2006). Forecasters use different approaches to predict TC formation, such as dynamical and numerical weather prediction (NWP) models, statistical-dynamical approaches and statistical models built based on the surrounding environmental conditions and storm-representative features (Roy & Kovordanyi, 2012).

Recent advances in the parametrization schemes, data assimilation techniques, and availability of various advanced satellite data with better retrieval algorithms provide better initialization conditions to the numerical weather prediction (NWP) models (Hendricks & Peng, 2012). The improved initial conditions in the model lead to a better TC track and intensity forecast well in advance (Rogers et al., 2006; Gall et al., 2013). TC genesis is difficult to predict one to several days in advance (Hennon & Hobgood, 2003; Kerns et al., 2008). Therefore, different ensemble approaches are considered for TC genesis forecasts to improve the certainty of the numerical forecasts (Cheung &

Elsberry, 2002; Wei et al., 2008; Reynolds et al., 2008; Snyder et al., 2010; Elsberry et al., 2011; Tsai et al., 2011;Thatcher & Pu, 2013; Halperin et al., 2013). A study by Halperin et al. (2013) on the TC genesis forecasts in the North Atlantic (NA) basin illustrates that the performance of five individual models decreases with increasing lead time from 6 to 96 hours. Also, the performance varies between seasons and locations; therefore, it is very difficult to identify the best performing NWP model. This analysis shows that it remains challenging for the models to predict the genesis across the ocean basins. Nevertheless, Halperin et al. (2016) have noted an improvement in the predictions of the TC formations using numerical models in recent years.

Some observational studies have identified certain thermodynamical and dynamical conditions necessary for TC formation (Gray, 1968; McBride & Zehr, 1981; Fu et al., 2012; Peng et al., 2012; Davis & Ahijevych, 2012; Komaromi, 2013; Rutherford et al., 2018). Researchers have developed different linear statistical models with large-scale environmental variables as the predictors that are found to be necessary for TC formation (Henon & Hobgood, 2003; Schumacher et al., 2009; Fan, 2010). These earlier studies mostly employ either the Linear Discriminant Analysis (LDA) or multi-linear regression statistical models using reanalysis datasets for large-scale environmental predictors. These two models consider linear relationships between the predictors (large-scale environmental variables) and the predictand (TC formation likelihood) and have achieved some skill in predicting TC formation. As the TC formation involves multi-scale interactions between the synoptic-scale and mesoscale features, TC statistical forecasts are advancing towards using machine learning (ML) algorithms. These algorithms consider the non-linear interactions and give a better understanding of the interactions between different environmental variables within the available dataset.

The different ML approaches in TC research include clustering analysis, binary classification, feature selection, pattern matching, association rules, and decision rules (Camargo et al., 2007b; Li et al., 2009; Jaiswal et al., 2013; Yang, 2016). Some of these approaches, such as classification, decision rules, and association rules, help to determine the correlations between environmental variables and their influence on TC formation and intensity characteristics. The decision rules establish quantitative thresholds for the environmental variables and establish precise criteria for future statistical predictions. A study by Camargo et al. (2007b) has implemented the clustering approach to identify the differences between types of tracks, such as straight moving and recurving. They have also noted that both intensity and seasonality are stratified within clusters as also the frequency of occurrence and regions of impact are largely cluster dependent. Therefore, clustering techniques discover differences across various clustered track features in each of the ocean basins and improves our current understanding of TC activity. The association rules technique helps to identify the favorable conditions for TC intensification, and rapid intensification (RI) processes through different combinations of environmental conditions from a set of environmental predictors with both balanced and unbalanced datasets (if the number of samples of one class in a binary classifier is significantly less than the other class then it is considered as an unbalanced dataset) (Yang, 2007; Yang et al., 2011 Yang, 2016).

There are studies on the application of ML techniques to convective processes and TC formation. The most widely used ML algorithms are decision trees (DTs), random forests (RF) and support vector machines (SVM) for TC formation characteristics with a set of environmental predictors from forecast analysis, reanalysis and different satellite image data (Li et al., 2009; Zhang et al., 2013a, b, c; Zhang et al. 2015; Han et al. 2015; Park et al. 2016; Kim et al., 2019). The DTs iteratively classify the data into separate groups using an information gain ratio measure that quantifies the goodness of split and build a tree-like structure with specific thresholds for the environmental variables (Quinlan, 1987, 1993). The information gain ratio is a probabilistic measure that measures the quality of the split between developing and non-developing cases, in other words, how good the predictor or predictors are at distinguishing between developing and non-developing cases. The RFs are a more robust tree-based technique that uses random data samples to build an ensemble of DTs with identical distribution for all the trees (Brieman, 2001). The SVM is a non-parametric technique that does not consider any distribution for the underlying data and constructs a hyperplane that separates the two different groups (i.e. it can be either developing or non-developing cases). The SVM algorithm transforms the data into higher dimensional space if the underlying data is not linearly separable using different kernel functions. The miss-classification rate is minimal in the SVM as it chooses an optimal margin around the separating hyperplane (Mountrakis et al., 2011; Park et al., 2016). Li et al. (2009) have used DTs to identify the relationships between environmental variables and TC formation processes. This study suggests that DTs improve our current understanding of the TC formation as they reveal the co-effects of the variables, and also identify significant predictors for predictions of TC formation. Zhang et al. (2015) have employed DT analysis to discriminate the developing and non-developing tropical disturbances across the Western North Pacific (WNP) basin based on environmental predictors from the forecast analysis datasets.

Park et al. (2016) have applied DTs to discriminate developing and non-developing tropical disturbances using the remote sensing data of the storm-representative parameters employing cyclonic circulation symmetry, the intensity of the circulation, and organizational indices. Han et al. (2014) have employed the three ML techniques DT, RF, and SVM with remote sensing data for detecting convective initiation. This study notes that the RF algorithm performs better than the other two algorithms. A recent study by Kim et al. (2019) compares the performance of three ML techniques (DT, RF, SVM) and LDA across the WNP basin using the satellite measurements and noted that ML techniques have a high hit rate of predicting TC formation compared to the linear classifier LDA. The above studies highlight the importance of the ML techniques in advancing our knowledge of TC formation and building robust statistical models. These studies employ statistical schemes for a better understanding of different TC characteristics with either environmental or storm-representative predictors across the NA and WNP ocean basins. Some of the recent studies using statistical analysis have noted that dynamical variables are more important across the WNP basin and thermodynamical variables are important across the NA basin, perhaps due to different modulating influences such as MJO and ENSO across global ocean basins (Fu et al., 2012; Peng et al., 2012; Raavi & Walsh, 2020b (chapter 4)). Therefore, it is essential to understand the prominent environmental variables influencing the TC formation across different global ocean basins using advanced data mining techniques.

In the current study, we classify the developing and non-developing tropical depressions across the global ocean basins using the environmental predictors from the ERA-Interim reanalysis with three ML techniques, the DTs, RFs, and SVM. Here, we also test the ventilation theory and estimate the order of the important predictors for TC formation across each ocean basin. A recent theory of TC formation, the "marsupial pouch theory", as described in chapter 2 states that there exists a closed circulation region within large-scale disturbances that favors TSs by providing necessary conditions for their formation (Dunkerton et al., 2009). A phenomenon-based TC detection tracking scheme, the Okubo-Weiss Zeta Parameter (OWZP) given in chapter 3, (Tory et al., 2013a, b) identifies the locations within the marsupial pouches that have the potential for TS formation. This OWZP algorithm is further tuned with the tropical disturbance dataset of the Australian region to detect the tropical depressions (Tory et al., 2018) as described in chapter 4. Here, we consider any TC with wind speeds greater than 17 m/s as a TS and weaker storms are TDs. This detection scheme is used to detect both developing (TS) and non-developing tropical depressions (ND) across global ocean basins. Therefore, the current study gives a global view of the performance of different ML algorithms in classifying the developing and non-developing cases using a similar set of environmental predictors.

5.2. Data and Methodology

5.2.1 ERA-interim data

The environmental variables are from the ERA-interim data (see section 3.3.2 for more details) with a horizontal resolution of $0.75° \times 0.75°$ at six hourly intervals for a period of 30 years from 1989-2018. The environmental variables which are input to the machine learning algorithms are total column water (WV; in units of kg/m^2), 700 hPa relative humidity (RH; %), maximum potential intensity (MPI; m/s), 850 hPa Relative vorticity (RVor; s$^{-1}\times10^{-6}$), 500 hPa Omega (Omega; Pa/s), 200 hPa divergence (DIV; s$^{-1}\times10^{-6}$), vertical shear of zonal wind ($U_{200} - U_{850}$) (Ushear; m/s). With a suspicion that non-developing cases may have stronger deformation forces unfavorable for the development of the initial vortex compared to the developing cases the current study uses two additional variables. These variables include OW center (OWC) that determines the strength of the circulation by considering the deformation forces relative to vorticity; and the OW surrounding (OWS), which identifies the strength of the surrounding protective layer called the shear-sheath (Rutherford et al., 2018) at 850 and 500 hPa levels as described in section 4.2.4. A stronger shear sheath layer does not allow the intrusion of dry air or opposite signed vorticities to the center of the storm. The OW center (OWC) and the OW surrounding (OWS) variables are correlated across the ocean basins also the values of these variables at 850 hPa level correlates with the values at 500 hPa level; therefore, we retain one (OWC850) of them in building the ML model. The variables of the current study are also identified as prominent variables for TC formation from the initial depression stage vortex based on the statistical analysis (Raavi & Walsh, 2020b; chapter 4). Also, a subset of the variables used in the current study comprises the equivalent terms of ventilation index (combination of vertical wind shear, potential intensity and column water vapor) which is an excellent predictor of TC genesis from a precursor disturbances in the historical data in both theoretical and empirical way (Tang and Emanuel, 2012). Vecchi et al. (2019) has applied this index in global climate models and noted that increased values of the index in the CO2 doubling experiment suggesting unfavorable environment for TC formations in the tropical regions. Here, we also built machine learning models with the equivalent variables of ventilation index (Ushear, MPI, and WV parameters) in the current climate to validate the physical theory of Tang and Emanuel (2012).

5.2.2 Tropical Cyclone detection and tracking

Here we use the OWZP TC detection and tracking scheme as described in section 3.4.2 and section 4.2.2 for detecting both developing and non-developing tropical depressions. Globally, around 2300 developed TSs, and 3432 initial TDs are detected for 30 years 1989-2018. The TD numbers include the 2300 TSs, as all TSs are once TDs. There are 50% more low-pressure systems that satisfy the TD threshold conditions than TSs. In order to obtain the non-developing depressions, we subtract the developed TSs from the TDs, and 1132 cases are identified as non-developing depressions across the global ocean basins. The global ocean basin boundaries are given in table 4.1.

5.2.3 SMOTE Resampling

A dataset is considered as imbalanced when the first class sample numbers are significantly larger than the second class in a binary classification problem. The classification methods tend to bias the results towards the higher sampled class. In order to avoid this bias in the classifier, the lower sampled class, that is the non-developing cases, in the current study are oversampled to the size comparable to the developing cases using a synthetic minority oversampling technique SMOTE (Chawla et al. 2002). This SMOTE method produces new samples for the minority class between the existing samples using a specified number of nearer neighbor samples. A recent study by Zhang et al. (2015) on the discrimination of developing and non-developing tropical disturbances employs this SMOTE algorithm for balancing both the classes.

5.2.4 Input data for ML algorithms

The input data of the environmental variables to the ML algorithms include the storm-centered mean values of the variables (section 5.2.1) within a 12° × 12° box region around the center of both developing and non-developing cases. This box region is found to show maximum environmental differences between developing and non-developing cases compared to a bigger region. The increase in the box region includes the surrounding values of the environmental variables which disrupts the mean values of the storm generated environmental conditions of the box region. In the current study, we include the tropical depressions that exist two days before the TD declaration stage (as discussed in Section 4.2.3) across each ocean basin. The input data to the algorithm include three categories representing 48 and 24 hours before the time of TD declaration and the day of TD declaration, labeled Day-2, Day-1, and Day0, respectively, as shown in Figure 3.4. Each of these categories has two values (i.e., at 12 hourly time steps) for each of the environmental variables along the tracks. Therefore, each of the developing and non-developing case have six values along the track for each of the environmental variables. We use 70% of the developing and non-developing data after applying the SMOTE algorithm for training and 30 % of the data for testing the performance of the classifier on the new data. We will be discussing the performance of each ML algorithms for the test data in the results section. The collinearity of the variables is checked in each of the ocean basins by observing the partial correlations that are greater than 0.5 before starting the classifier's training. If the partial correlation between two variables is greater than 0.5 then one of the variable is excluded for further analysis. The final environmental variables OWC850, Beta*, MPI, WV, and Ushear are included in building the classifiers. Figure 5.1 shows the schematic of the machine learning algorithms framework. This framework mainly involves three steps: (1) Pre-processing of the data which involves collinearity check, removing outliers, and balancing both the developing and non-developing cases; (2) Learning of the algorithms through training data followed by testing the learned model with new data; (3) Calculation of different forecast accuracy metrics from the confusion matrix (final output) of the algorithms to evaluate the performance of the models.

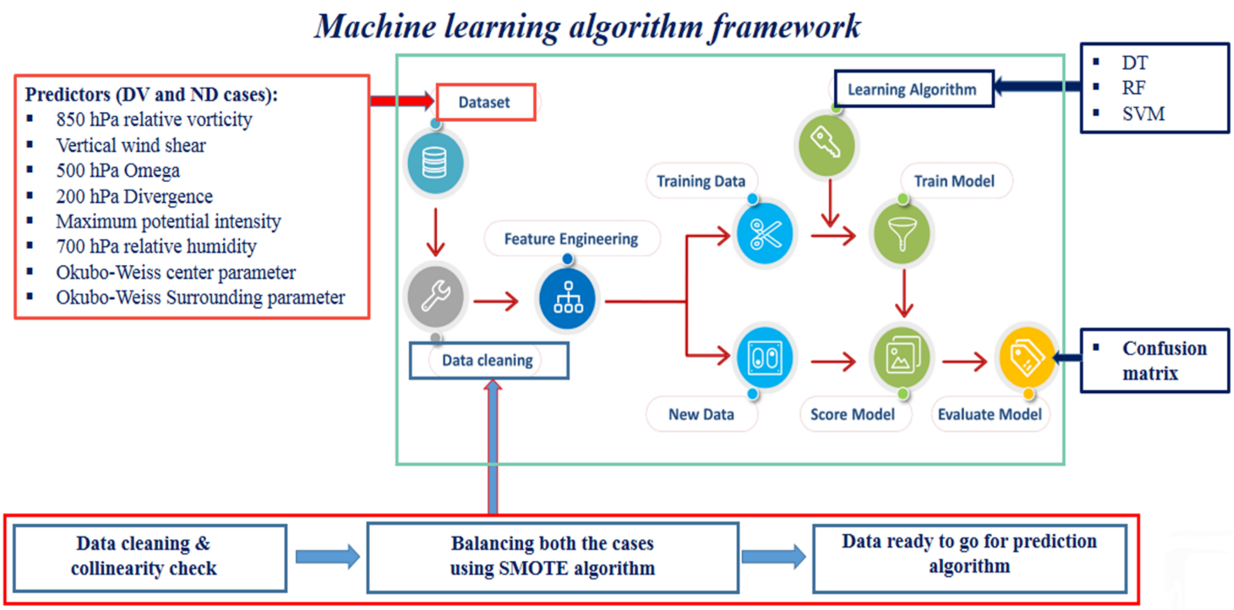

Figure 5. 1 Steps involved in running different machine learning algorithms

5.2.5 C4.5 classification algorithm - Decision trees (DTs)

Classification is a data mining approach in gaining information for easy interpretation and forming rules to make decisions from historical datasets (Quinlan, 1987). A binary classification algorithm's main objective is to predict the categorical variable assigned as 0 to non-developing cases and 1 for developing cases based on the values of certain predictive variables by constructing a decision tree-like structure with various combinations of predictor variables within the dataset. Here, we use the C4.5 algorithm, a classical decision tree algorithm that uses a specified measurement called "information gain ratio," (as described in section C.1 of appendix C) which divides the samples into two or more classes, i.e., two classes for binary classification (Quinlan, 1987, 1993). This algorithm divides the data consecutively into subsets and assigns class labels (i.e., 0 or 1 for binary classification) for each observation. In binary classification, any node within a decision tree has a primary node and two successor nodes. The different tests at each node determine the structure of the decision tree. An environmental variable that produces a high degree of segregation between the two classes acts as the root of the tree. Quinlan (1993) states that available information of any problem is dependent on its probability which is quantified as the logarithmic function of that probability known as "information gain ratio". The variable having a higher information gain ratio acts as the more important variable (refer section C.1 of Appendix C for more details). Here, the variable that has maximum information gain partitions the data at each node of the decision tree.

The decision tree's construction involves selecting a variable with maximum information gain criterion as the "root node." The subsequent splitting of the decision tree stops when the minimum leaf size threshold is reached. A minimum leaf size is chosen to stop the successive splitting of the decision tree. This minimum leaf size parameter indicates the number of samples that a leaf node can hold and represents the size of the decision tree. A lower threshold leads to a larger tree (with the risk of overfitting), whereas a higher value leads to a smaller tree and may not represent the entire information. A decision tree with overfitting does not perform better for a new test dataset, as it is not generalized well for the overall training dataset. Therefore, an optimum value for the “minimum leaf size” is gained through cross-validation with reduced error pruning option that gives the highest classification accuracy by pruning the leaf nodes and branches with reduced accuracy. In the current study, the Weka 3.6.2 software package is used to perform the C4.5 algorithm. The source link for the open-source software is https://www.cs.waikato.ac.nz/ml/weka/.

5.2.6 Markov decision process (MDP)

Tropical storm formation from an initial tropical disturbance is a stochastic process that incorporates the random variation due to different dynamical features across the global ocean basins. Regnier and Harr (2006) have considered atmospheric dynamics as a Markovian process. A Markov model is a stochastic model used to model time-varying systems with an assumption that future evolving states are dependent on the current state. After estimating the thresholds of significant environmental variables for developing depressions and non-developing depressions from the decision trees, we use a Markov model to address the certainty in the decision-making process by considering the atmospheric state transitions within the 48 hours before the TD declaration dataset. The certainty in each decision rule is achieved through the Discounted Markov Decision Process (MDP) with a policy iteration algorithm (Puterman, 1994), as described in the section C.2 of appendix C. MDP is a sequential decision model in which an action is chosen at each step, which transforms the state of the system into a new state and final decision is based on the maximum cumulative reward. Here, the state of a dynamic system represents its physical properties at a given time. Based on the current environmental state, the decision-maker chooses an action and receives a reward at each step of the iterative process. A reward is a real value which is assigned by the decision maker to each of the environmental states based on the transition probabilities and the favorability for TS formation (refer to section C.2.3 for more details on reward). For instance, if the environmental state having lower atmospheric moisture moves to a state having higher moisture due to chosen action during the iterative process, then the decision maker gives higher reward for this transition as higher moisture is favorable for TS formation. In contrast, the decision maker gives lower reward when the higher moisture state moves to lower moisture state. The reward value therefore relies on the current environmental state and the selected action. The rewards are necessary to obtain a discrete solution (maximum expected reward) to the MDP problem at the end of the iterative process.

The actions and associated rewards for each state are prescribed in policy, and there exists an optimum policy in MDP. The optimum policy is defined for every environmental state in MDP there exists an action (either developing (TS) or non-developing (ND)) that gives the highest sum of rewards. The optimum policy depends on the transitional probabilities and the rewards. Therefore, MDP estimates the optimum policy by maximizing the rewards in an iterative decision process. To be precise, the main goal of an MDP is to find the best actions (TS or ND) for each of the environmental states by estimating the optimum policy that gives maximum expected cumulative reward. The two methods to solve an MDP (i.e., to find an optimum policy) are value iteration and policy iteration. The policy iteration algorithm (refer appendix C for more details) adopted here using a MATLAB functions (Chades et al., 2005) which computes an optimum policy (which environmental states favors TS formation) at the end of the iterative process to maximize the expected total reward for the prevalent environmental states in the entire dataset of developing and non-developing cases (Puterman, 1994). The MDP helps to ascertain the probabilistic evolution of the TS suitable states under the observed uncertainties with state transitions two days before the TD formation. The detailed description of the MDP for establishing different states and finding an optimal policy is given in section C.2.3 of appendix C.

5.2.7 Random Forests (RF)

RF is an ensemble tree-based method using the Classification and Regression Trees (CART) algorithm, which is a rule-based classification scheme (Breiman, 2001) using the measure of information gain. It is well known that the decision trees are highly sensitive to the training data, and we get entirely different decision trees when we change the input training data (Lawrence & Wright, 2001). The RF algorithm constructs an ensemble of individual DTs by randomly selecting the predictors and the training samples. While building each tree, the RF considers a random subsample of total observational data and the left out data is known as the "out-of-bag" observations. The final decision of the algorithm in classifying the developing and non-developing cases is based on the majority or weighted voting of the total individual trees (i.e., by measuring maximum number of times the prediction for the out-of-bag observation from different decision trees). For instance, if the developing case is the output from the majority of the decision trees for that particular observation then we assign that atmospheric state as a developing case based on majority voting. We also use weighted voting as the number of random samples used to build the decision trees vary from one tree to another tree. The relative importance of the predictor variables used in an RF model is determined using the strategy of Breiman (2001), that is the variable having the highest mean decrease in the accuracy of the RF model when that variable is randomly permuted is considered as the important predictor variable. Similar strategies are used in the current study to build the RF model and determine the order of the predictors in each of the global ocean basins. The current study uses MATLAB to build the RF model by randomly varying the training samples.

5.2.8 Support Vector Machines (SVM)

SVM is a supervised machine learning algorithm that efficiently classifies the linearly separable data into the target classes by constructing optimal hyperplanes that best separates the binary classes. For two-dimensional data this is a straight line (the decision boundary) and for three dimensional data it is a plane separating the two classifying groups (refer to section C.3 of appendix C for more details). In higher dimension, the separation is a hyperplane of the appropriate dimension. The SVM choses the optimal hyperplane that has maximal margin between the training observations and the dividing hyperplane. If the data is not linearly separable then the SVM uses the Kernel trick (Murty & Raghava, 2016) to separate the two underlying groups. SVM presumes that the samples in the multidimensional space can be separable linearly in the given input feature dimension. Nevertheless, there is overlapping between the target class data points, making it difficult to separate the classes linearly. Therefore, SVM uses specific kernel functions for transforming the input data into higher dimensional space that can be separable. There are three different types of kernel functions: linear, non-linear, polynomial, and radial basis function (RBF). The RBF is a widely used kernel for various Earth Sciences applications using remote sensing data (Gleason & Im, 2012; Guneralp et al., 2013; Kim et al., 2014). In the current study, we implemented the SVM using the RBF kernel for separating the developing and non-developing cases using the functions in MATLAB. The relative importance of the predictors using the SVM classifier is determined by using the measure of F-Score (refer appendix C for more details). Here each of the variables is removed from the SVM model one at a time and the F-score is estimated. On excluding a particular large-scale variable, if the SVM model gives the lowest F-score then the variable is considered as the most important predictor. Table 5.1 compares the three different machine learning algorithms.

Table 5. 1 Comparison of the machine learning algorithms

Machine learning schemes	Working Principle	Predictive power & interpretability
C4.5 Algorithm - (decision trees)	Information gain ratio (a quantitative measure based on the probability)	Less predictive power; easy to interpret
Random Forests - (a bag of decision trees)	Information gain ratio; uses bootstrapped samples of data	High predictive power; less interpretable
Support vector Machines - (constructs hyperplanes to classify the data)	Maximal margin between hyperplanes and the training observations; uses kernel functions and transforms the data to higher dimensional space	High predictive power; less interpretable

5.2.9 Forecast Skill Scores

The confusion matrix is the output of the classifiers/ ML algorithms and is similar to the contingency table which gives the hits (true positives), false alarms (true negatives), misses (false negatives) and true negatives (false positives) as defined by Wilks (2006).

In order to measure the forecast skill score of the classification schemes, we use the F-score, which is a weighted harmonic mean of precision and recall (Powers, 2011). The precision identifies what proportions of positive instances are correct (ratio of true positives to the sum of true positives and false positives). In contrast, the recall is a measure of correctly identified positive instances (the ratio of true positives to the sum of the true positive and false negative). The performance of the models build on the training datasets are verified with the test data using the metrics hit rate (HT) and false alarm rate (FAR) using the values of the contingency table. The HT rate is the number of classified developing cases to the total number of observed developing cases (the ratio of true positive to the sum of true positive and false positive). The FAR represents the number of misclassified developing cases as non-developing cases to the total number of observed non-developing cases (the ratio of false-negative to the sum of a true negative and false negative). The section C.4 in appendix C gives the formulae for these different forecast skill scores.

5.3. Results

The current section describes the performance of the DT, RF, and SVM algorithms in classifying the developing and non-developing cases and predicting the developing cases across global ocean basins using environmental variables as the predictors.

5.3.1 Classification using DTs

DTs are built across each global ocean basin using the C4.5 algorithm to classify tropical depressions into the developing and non-developing cases with the large-scale environmental variables. The classification accuracy varies from 67% to 72% across the global ocean basins (Table 5.2). The WNP (0-15N) has lower accuracy of 67% compared to other basins which have moderate accuracy in classifying the developing and non-developing cases. Also, the equal values of F-score and accuracy indicates that each test case is correctly classified to the respective classes that is the number of correctly detected developing cases are equal to number of correctly detected non-developing cases. The F-score of the reduced model which includes Ushear, MPI, and WV parameters to validate the ventilation theory of Tang and Emanuel (2012) is almost similar to the F-score of full model that has other additional parameters. Similar to the study of the Tang and Emanuel (2012), this analysis with the DT machine learning algorithm also indicates that ventilation index is an important predictor for the TC formation from the initial tropical depressions across global ocean basins. The information gain ratio metric identifies the significant predictors in each of the ocean basins. The OWC parameter is found to be prominent in all of the ocean basins, whereas the moisture parameter WV is found to be important in most of the ocean basins except the WNP basins where the dynamical conditions are more important compared to the thermodynamical conditions. There are some differences in the order of the important predictors across different ocean basins (Table 5.2). The weaker performance of the DTs in classifying the data to developing and non-developing cases is perhaps due to the smaller sample size of the input training data and may require additional predictors. Therefore, it is essential to test the performance of other ML algorithms in classifying the developing and non-developing cases with a similar set of prominent predictors.

Table 5. 2. Accuracy, F-score (All Variables), F-score of reduced model (RM) with USHEAR, MPI, and WV variables, and the order of important variables obtained from the DTs using the C4.5 algorithm in different ocean basins

BASIN	Accuracy	F-Score	F-Score (RM)	Order of Variables
MDR_NA	72%	0.72	0.73	OWC850, MPI, WV, Ushear
ENP	69%	0.69	0.65	Ushear, WV, OWC850
WNP	67%	0.67	0.7	OWC850, MPI
WNP_P	68%	0.68	0.65	OWC850, MPI, Ushear
SI	68%	0.69	0.67	OWC850, WV, MPI
SP	71%	0.71	0.72	WV, Ushear, OWC850, MPI
AUS	70%	0.70	0.65	OWC850, WV

5.3.2 Classification using RF

The RFs results are an ensemble of 1000 iterations. They have 400 trees in each iteration with bootstrapped samples of data that uses the Classification and Regression Trees (CART) algorithm with a set of environmental predictors. The RFs have a classification accuracy between 79% and 84%, with F-scores of 0.79 to 0.84 (Table 5.3). The Eastern North Pacific (ENP) basin has a lower accuracy of 79%, and North Atlantic main development region (NA-MDR) has higher accuracy of 84% in classifying the developing and non-developing tropical depressions. The F-score of reduced model with RF algorithm using Ushear, MPI, and WV variables is slightly less than (i.e. around 0.05-0.12) the full model with other additional variables. Similar to the DTs, the RF algorithm also validates that equivalent variables of ventilation index (Tang and Emanuel (2012)) acts as important predictors of TC formation. As the ensemble methods are more robust than single DT, the performance of the RFs is superior to the DTs, which are less sensitive to changes in the input training data. Aggregating the multiple DTs built on the random sets of subsamples of data reduces the overfitting error compared to the single DTs, which are more prone to overfitting. The averaged value of the prediction error obtained by randomly permuting the each predictor for the out-of-bag observations (as defined in section 5.2.7) in each tree, is used to estimate the relative importance of the predictors. The higher the value of the error, the better it differentiates the developing and non-developing cases. The order of the variables is defined by the descending order of the prediction error. The important variables identified from the RFs agree (Table 5.3) with the variables identified by DTs built using the C4.5 algorithm (Table 5.2) in most of the ocean basins.

Table 5. 3. The same as Table 5.2 but using the RF classifier.

BASIN	Accuracy	F-Score	F-Score (RM)	Order of Variables
MDR_NA	84%	0.83	0.76	OWC850, MPI, WV, Ushear, Beta*
ENP	79%	0.79	0.72	Ushear, WV, OWC850, MPI, Beta*
WNP	82%	0.82	0.71	OWC850, MPI, WV, Ushear, Beta*
WNP_P	82%	0.82	0.7	OWC850, MPI, WV, Ushear
SI	83%	0.83	0.75	WV, MPI, OWC850, Beta*, Ushear
SP	81%	0.81	0.76	Ushear, MPI, OWC850, WV, Beta*
AUS	82%	0.82	0.74	OWC850, WV, MPI, Ushear, Beta*

5.3.3 Classification using SVM

The SVM results are an ensemble of 1000 iterations with random samples of data for training and testing in each iteration. The classification accuracy using the SVM classifier with RBF kernel varies from 80% to 87% and the F-scores from 0.8 to 0.87

across global ocean basins (Table 5.4). The South Pacific basin has the lowest accuracy of 80% of all the ocean basins, and the South Indian basin has an accuracy of 87% in classifying developing and non-developing cases. The F-score values of the reduced model using SVM algorithm with Ushear, MPI, and WV variables and full model with additional variables Beta* and OWC850 is similar to the results of RF algorithm suggesting that ventilation index is an important predictor for prediction of TCs from the initial tropical depressions. The relative importance of predictors in the SVM classifier is measured based on the F-score when a target predictor is not included in building the SVM model. The order of importance of predictors is derived based on the ascending order of the F-scores (Table 5.4). The order of the important variables for classifying the tropical depressions differs compared to the DTs and RFs as the SVM algorithm uses a kernel technique to classify the data compared to the other two algorithms which uses a probabilistic measure based on the information entropy.

Table 5. 4. The same as Table 5.2 but using the SVM classifier.

BASIN	Accuracy	F-Score	F-Score (RM)	Order of Variables
MDR_NA	85%	0.85	0.75	MPI, OWC850, WV, Ushear, Beta*
ENP	81%	0.81	0.72	Ushear, MPI, WV, OWC850, Beta*
WNP	82%	0.82	0.70	MPI, Ushear, Beta*, OWC850, WV
WNP_P	83%	0.83	0.69	MPI, OWC850, Ushear, WV, Beta*
SI	87%	0.87	0.74	MPI, Beta*, OWC850, WV, Ushear
SP	80%	0.80	0.73	Ushear, MPI, WV, OC850, Beta*
AUS	81%	0.81	0.74	MPI, OWC850, WV, Ushear, Beta*

5.3.4 Forecast verification of the ML techniques

The performance of the ML algorithms to forecast the developing depressions (TSs) is estimated using two accuracy metrics, namely HR and FAR as defined in section 5.2.9. Figure 5.2 shows the performance of all the three ML techniques across different ocean basins. The RF and SVM have better accuracy metrics than the DTs, with higher HR in prediction of the developing cases with background environmental conditions. When comparing the FAR, both RF and SVM have a low FAR compared to the DTs. The DTs have slightly better performance in HR in comparison to the FAR. The HR of reduced model with equivalent variables of ventilation index is comparable to the full model with additional parameters. In contrast, the FAR is higher in the reduced model compared to the actual model. Based on this analysis, we can suggest the RF and SVM are the better performing techniques compared to the DTs. Although the DTs have a weaker performance with this set of data, they have better performance in earlier studies, with an accuracy of up to 91% (Kim et al., 2019). The inclusion of additional predictors and increasing the sample size may improve the performance of the DTs.

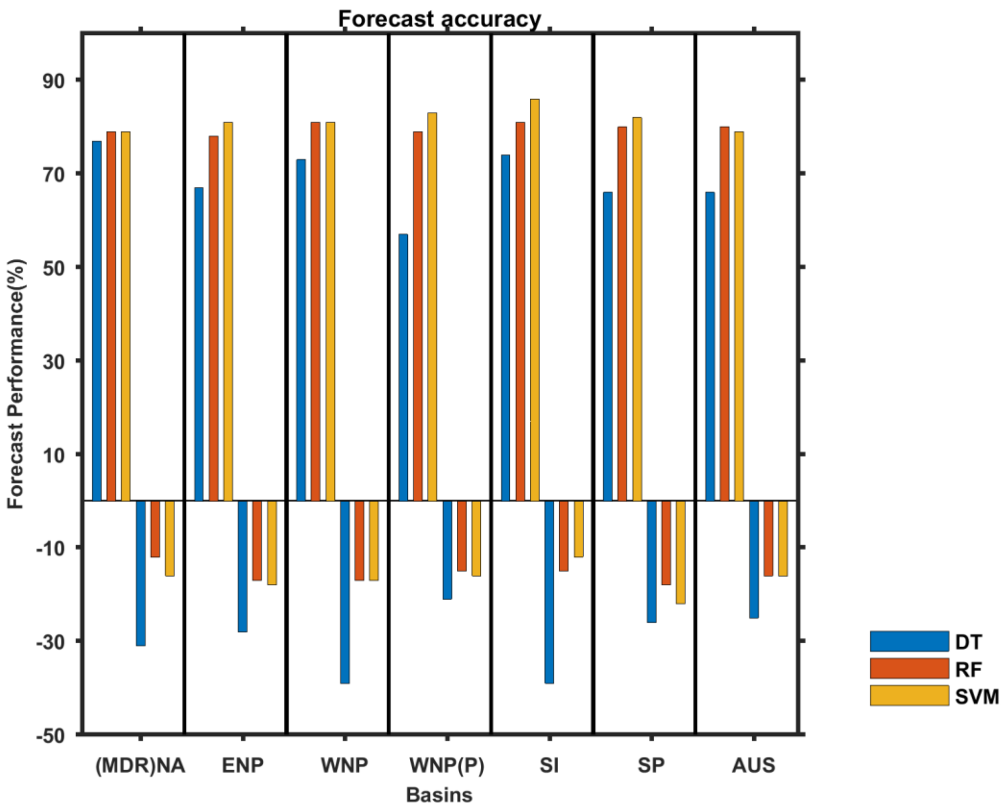

Figure 5. 2 Forecast accuracy metrics for all the three ML algorithms across global ocean basins. Positive values indicates the Hit Rate (correctly classifying the developing cases) and negative values indicates the False Alarm Rate.

5.3.5 Decision tree rules using C4.5 algorithm

The forecast results using RF and SVM algorithms have good skill in identifying developing cases with background environmental conditions. However, these schemes do not explicitly provide the thresholds for the environmental predictors for classifying the data. Therefore, we use a binary classification algorithm C4.5 to obtain the combinations of different variables with quantification of their respective thresholds favoring the development of an initial tropical depression into a TS across global ocean basins. An optimum set of conditions is obtained with the classification algorithm by giving a minimum leaf size and reduced error pruning. The methodology section 5.2.5 includes a detailed explanation of the algorithm and its optimized settings. The summary of the major decision rules from the decision trees built across different ocean basins (Table 5.5) is discussed below.

5.3.5.1 Northern Hemisphere Basins

The decision tree rules across the North Atlantic (NA) basin include the values attributes of the OW center parameter at 850 hPa (OWC850), MPI, and Ushear from the tropical depression dataset initiating 48 hours before its genesis (Table 5.5). The decision tree rules indicate that OWC850 acts as the most significant variable compared to other environmental variables. As shown in Table 5.5, the first decision rule is when the OWC850 is higher than $14.5 \times 10^{-10}\ s^{-2}$, a tropical depression develops into a TS. The second rule is when the OWC850 is less than or equal to $14.5 \times 10^{-10}\ s^{-2}$, then MPI is higher than 45.47 m/s, and Ushear should be less than or equal to -5.84 m/s for a tropical depression to develop. The summary of the decision tree algorithm with combinations of the conditions is that a higher value of low-deformation vorticity (OWC850), favorable thermodynamical instability (MPI), and lower values of vertical shear of horizontal wind is necessary for tropical cyclone formation across the NA basin.

Over the equatorward portion of the Western North Pacific (WNP) basin (0-15N), the final decision tree includes the variables OWC850 and MPI, as shown in Table 5.5. The significant attribute for classifying the developing and non-developing cases over this basin is the OWC850 parameter. The first decision rule is that when the OWC850 is greater than $25.49 \times 10^{-10} s^{-2}$, the initial depression develops into a TS. The second rule is when the OWC850 ranges between 9.17 and $25.49 \times 10^{-10}\ s^{-2}$, and MPI is less than 64.17m/s, an initial depression to develop into a TS. The decision tree rules summarize that higher values of OWC850 or moderate OWC850 with an adequate amount of thermodynamic instability (MPI) are necessary for TS formation.

Table 5. 5. Decision tree rules governing the development of an initial depression (48 hrs. prior to TC genesis) through quantitative relations and co-influences of variables with diverse combinations in different ocean basins

MDR_NA Basin		
Rules	**Attributes**	**Accuracy**
1: If OWC850 > 14.5×10^{-10} s^{-2} then TS forms	OWC	(120) 100%
2: If OWC850 <= 14.5×10^{-10} s^{-2}, MPI > 45.47 m/s and USHEAR <= -5.84 m/s then TS forms	OWC, MPI, USHEAR	(97/20) 79%
WNP Basin		
Rules	**Attributes**	**Accuracy**
1: If OWC850 > 25.49×10^{-10} s^{-2} then TS forms	OWC	(119/7) 94%
2: If 9.2×10^{-10} s^{-2} < OWC850 <= 25.49×10^{-10} s^{-2} and MPI < 64.17 m/s then TS forms	OWC, MPI	(541/195) 63%
WNP_P Basin		
1: If OWC850 > 8.07×10^{-10} s^{-2} then TS forms	OWC	(414/118) 72%
1: If 3.3×10^{-10} s^{-2} < OWC850 < 8.07×10^{-10} s^{-2}, MPI > 62.64 m/s and USHEAR > -5.2 m/s then TS forms	OWC, MPI, USHEAR	(118/41) 65%
ENP Basin		
Rules	**Attributes**	**Accuracy**
1: If USHEAR <= 0.34 m/s and WV > 56.25 kg/m^2 then TS forms	USHEAR, WV	(299/61) 80%
SI Basin		
Rules	**Attributes**	**Accuracy**
1: If OWC850 > 23.35×10^{-10} s^{-2} then TS forms	OWC	(243/41) 83%
2: If OWC850 <= 23.35×10^{-10} s^{-2} and WV > 57.25 kg/m^2 then TS forms	OWC, WV	(136/23) 83%
3: If OWC850 <= 23.35×10^{-10} s^{-2}, 53.23 kg/m^2 < WV <= 57.25 kg/m^2 and MPI <= 68.9 m/s then TS forms	OWC, WV, MPI	(253/96) 62%
Australian Region		
Rules	**Attributes**	**Accuracy**
1: If OWC850 > 24.08×10^{-10} s^{-2} then TS forms	OWC	(76/1) 99%
2: If OWC850 <= 24.08×10^{-10} s^{-2} and WV > 57.1 kg/m^2 then TS forms	OWC, WV	(184/60) 67%
SP Basin		
Rules	**Attributes**	**Accuracy**
1: If WV > 45.56 kg/m^2, USHEAR <= 4.2 m/s and OWC850 > 18.69 then TS forms	WV, USHEAR, OWC	(81/8) 90%
2: If WV > 45.56 kg/m^2, USHEAR <= 4.2 m/s, OWC850 <= 18.69×10^{-10} s^{-2} and MPI > 64.23 m/s then TS forms	WV, USHEAR, OWC, MPI	(171/52) 70%

Similarly, across the poleward region of the WNP basin from 15-30N (WNP_P), the OWC850, MPI, and USHEAR (Table 5.5) variables act as influencing variables for discriminating the developing and non-developing tropical depressions. The significant variable over this region is OWC850, suggesting that higher values of low-deformation vorticity are favorable for TS formation. These decision tree rules across WNP_P basin indicate that sufficient low-deformation vorticity with higher values of MPI and moderate Ushear favors TS formation from a depression stage vortex. The earlier studies discriminating developing and non-developing cases using environmental conditions from different datasets and approaches identify that higher values of vorticity and MPI are favorable for TC formation across the WNP basin (McBride and Zehr, 1981, Fu et al., 2012, Zhang et al., 2015; Tory et al., 2018).

In the Eastern North Pacific (ENP) basin, the USHEAR and WV act as significant influencing parameters for TS formation. The prominent variable in classifying the developing versus non-developing instances is USHEAR. The decision tree rule of the classification algorithm given in Table 5.5 is that USHEAR should be less than 0.34 m/s and WV is greater than 56.25 kg/m2 for an initial depression to develop. These classification rules reveal that lower vertical wind shear or adequate moisture throughout the atmospheric column supports TS formation. These results agree with a recent study by Raavi & Walsh (2020b) given in chapter 4 that WV and Ushear are the important predictors for TS formation across the ENP basin. In another study, Zhao and Raga (2015) have noted that mid-level relative humidity and low-level relative vorticity are the significant contributors to TS genesis in inactive years across the ENP basin.

5.3.5.2 Southern Hemisphere (SH) Basins

The decision tree rules over the South Indian (SI) Ocean and the Australian region indicate that the OWC850 and WV variables are important in classifying developing and non-developing cases. Over the South Pacific (SP) basin, we observe that in addition to the OWC850 and WV, lower values of Ushear and sufficient thermodynamical instability (MPI) lead to TS formation. Therefore, in the SH basins, sufficient low-deformation vorticity, higher atmospheric moisture, and lower values of shear favor TS formation. Other studies at different time scales showed that mid-level relative humidity, vertical wind shear and low-level relative vorticity (850hPa) play a vital role across the SH basins (Camargo et al., 2007; Ramsay et al., 2008; Chand & Walsh, 2010; Kuleshov et al., 2009; Liu and Chan 2012).

5.3.6 Markov Decision Process for developing and non-developing classification rules

The prominent large-scale environmental conditions with their quantitative values for TS formation across different ocean basins are obtained from decision trees with data from the 48 hours before TD declaration (Table 5.5). There are different dynamical features across the global oceans which sometimes can increase in the destructive environment such as high vertical wind shear or intrusion of dry air which lead to the non-development of the depression within 48 hours. It is therefore essential to observe the probability of maintaining these favorable environmental states, to make the TS genesis prediction more accurate. The main aim of using the MDP is that there may be three probable prestorm environmental states: least favorable, moderately favorable, and most favorable state. Each state has a probability of transitioning to the other two states as well as to a TS (genesis) favorable state. These transition probabilities across the three states and most favorable environmental states for tropical storm formation are estimated from the data. Keeping this in view, the MDP is used to address the uncertainty in the decision tree rules by considering the transition probabilities between atmospheric states. Here the atmospheric conditions observed from the decision rules are categorized into three different states. One state is upper bounded by the mean of the variable, a second state is bounded between the mean and thresholds of the variable obtained from decision tree rules and the third state is by thresholds of the decision tree rule (see also appendix C.2.3). The possible action to be taken from the MDP process is either a developing (TS) or a non-developing depression (ND). Transition probabilities represent the probability of moving into any of the three states given the current state is estimated for both developing and non-developing cases using the entire input dataset. Each of the three states is associated with an action, and higher rewards are given to the TS suitable states in developing cases obtained from the decision tree analysis (refer C.2.3 of appendix C for more details on rewards).

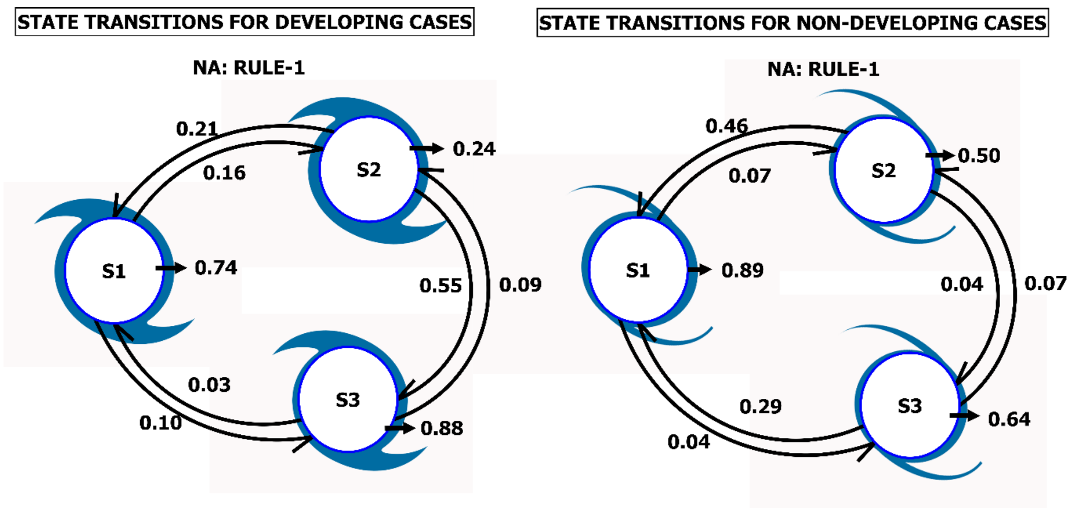

Figure 5. 3. Transition probabilities of decision rule1 (Table 5.5) of North Atlantic (NA) basin with S1 as State1, S2 as State2 and S3 as State3. Here S3 is the favorable state for TC formation across the basin.

The MDP for rule1 (Table 5.5) over the North Atlantic basin starts by calculating the transition probabilities for the different states: State 1 (OWC850<=10.47), State 2 (10.47< OWC850<14.5) and State-3 (OWC850>14.5), for both developing and non-developing cases. It is noticed that State 3 (which is suitable for TS formation from the decision tree rules) has a higher probability (0.88) of retaining its state in developing cases. In non-developing cases, it remains in state 3 with a moderate probability (0.64) or weakens to state 2 (having probability 0.29). Based on the favorability of TS environment from the decision tree rules we declare a reward matrix (refer section C.2.3) to the three respective states of developing cases for Action 1 (i.e., a developing TD) with the discounted reward of 0.95 to converge the solution in an infinite horizon domain. Similarly, a reward matrix assigned to each of the three states of non-developing cases for Action 2 (non-developing TD). This MDP problem is solved using the policy iteration, which gives an optimum policy (probable action to be taken for a given state). It shows a higher likelihood of TS formation in state 3 (higher atmospheric moisture content) in developing cases, while the other two states, i.e., states 1 and 2, are found to be favorable for non-developing cases. A schematic diagram in Figure 5.3 shows the transitions between state1 (S1), state2 (S2) and state 3 (S3) over the North Atlantic basin for developing and non-developing cases for rule1 from Table 5.5 with state3 as the favorable state for TS formation (i.e., higher atmospheric moisture content).

Table 5. 6. Steady-state probabilities and optimal action for the decision tree rules (Table 5.5) using Markov decision process

BASIN	Suitable state from the decision tree rules (Rules from table4)	Transition probabilities of this state		Optimum action from MDP using policy iteration
		Developing	Non-developing	
MDR_NA	OWC850 > 14.5×10^{-10} s^{-2}	0.88	0.64	TS
MDR_NA	OWC850 <= 14.5×10^{-10} s^{-2}, MPI > 45.47 m/s, USHEAR <= -5.84 m/s	0.76	0.56	TS
WNP	OWC850 > 25.49×10^{-10} s^{-2}	0.85	0.65	TS
WNP	9.2×10^{-10} s^{-2} < OWC850 <= 25.49 $\times10^{-10}$ s^{-2}, MPI < 64.17 m/s	0.60	0.40	TS
WNP_P	OWC850 > 8.07×10^{-10} s^{-2}	0.92	0.84	TS
WNP_P	3.3×10^{-10} s^{-2} < OWC850 < 8.07×10^{-10} s^{-2}, MPI > 62.64 m/s, USHEAR > -5.2 m/s	0.83	0.87	TS
ENP	USHEAR <= 0.34 m/s, WV > 56.25 kg/m^2	0.88	0.78	TS
ENP	USHEAR > 0.34 m/s, WV <= 56.25 kg/m^2, OWC850 > 13.89×10^{-10} s^{-2}	0.77	0.54	TS
SI	OWC850 > 23.35×10^{-10} s^{-2}	0.86	0.79	TS
SI	OWC850 <= 23.35×10^{-10} s^{-2}, WV > 57.25 kg/m^2, MPI <= 68.9 m/s	0.80	0.52	TS
AUS	OWC850 > 24.08×10^{-10} s^{-2}	0.84	0.55	TS
AUS	OWC850 <= 24.08×10^{-10} s^{-2}, WV > 57.1 kg/m^2	0.90	0.71	TS
SP	WV > 45.56 kg/m2, USHEAR <= 4.2 m/s, OWC850 > 18.69	0.91	0.40	TS
SP	WV > 45.56 kg/m^2, USHEAR <= 4.2 m/s, OWC850 <= 18.69×10^{-10} s^{-2}, MPI > 64.23 m/s	0.93	0.83	TS

Similarly, for rule 2, the division of the states (as described in section C.2.3) is as follows in order to calculate the transition probabilities: State 1 (OWC850<14.5, MPI<=45.47, USHEAR<=-5.84), State 2 (OWC850<14.5, MPI<=45.47, -5.84<=USHEAR<3) and State-3 (OWC850<14.5, MPI<=45.47, USHEAR>=3). Note that transition probabilities for developing cases show that there is less chance for the TS favorable state 1 to go to the other two states. The probability of remaining in the same state is 0.76. Based on the favorability of the TS environment, we choose a reward matrix as described in section C.2.3 of appendix C for the three states in developing cases and a reward matrix for non-developing cases with a discounted reward of 0.95. Similarly, the policy iteration shows a higher likelihood of TS formation in state 1 (under the specific values of OWC850 and MPI values the TC formation requires a low-wind shear region) with a lower likelihood of TS formation in states 2 and 3. In summary, for the two rules over the North Atlantic basin, the Markov decision process shows that higher values of OWC850 or moderate OWC850, high MPI, and moderate shear have the optimum chances of TS development under the observed state transition probabilities and TS suitable conditions.

Likewise, the Markov decision process is carried out in other ocean basins for the rules of Table 5.5 to determine the certainty of maintaining the favorable state for TS formation and are summarized in Table 5.6. Table 5.6 shows the respective transition probabilities for different classification rules and the optimum action for that rules from the iterative process using the policy iteration to solve the MDP problem. The MDP shows that in developing cases, the favorable states for TS formation have a higher probability of retaining their state without transitioning into less favorable states. In contrast, non-developing cases have a lesser probability of maintaining the TS favorable states which are transitioned to less favorable states. For example, the transition probabilities in non-developing cases are higher for specific rules in Table 5.6, which shows there are a few cases with those favorable states for TS formation. However, developing tropical depressions (viz. TS) have a maximum number of samples having a particular favorable atmospheric state; this leads to TS as the optimum action for that environmental condition in the MDP iterative process. Additionally, the action (i.e. TS or ND) which is chosen based on the maximum value of the expected total reward from the MDP, which is the optimum policy for TS suitable environmental states, also indicates a higher likelihood of TS formation (last column of Table 5.6). Therefore, MDP increases the certainty of each of the decision rules whether that particular rule leads to either a developing or non-developing case.

5.4. Discussion

The current study provides an overview of the ML algorithms' (DT, RF, SVM) performance in predicting the developing tropical depressions with a similar environmental predictor set across the global ocean basins. Of all the three algorithms, the RF and SVM algorithms have comparable performance in forecasting the developing cases (TSs) with slightly better performance by the SVM algorithm in all ocean basins, as shown in Figure 5.2. The performance of the DT algorithm is weaker compared to other schemes, with high FAR and low HR in all ocean basins. The DTs use an optimal choice at each stage of building the decision trees to select the predictors that better differentiate the two groups and are sensitive to the input training data. The RF (ensemble of DTs) and SVM (optimal hyperplanes with kernel functions) are robust algorithms that best predict the developing cases using environmental predictors. Also, all the three ML algorithms built with reduced number of predictors that are equivalent to ventilation index (Ushear, MPI, and WV) suggests that ventilation index is important for the prediction of TC formation from initial tropical depressions. Although the DTs are highly sensitive to the training data, they give quantitative limits to the co-influencing environmental predictors (Table 5.5). The prominent predictors identified from the different ML algorithms have some agreement, with WV (atmospheric moisture content) and OWC850 (low-deformation vorticity) being important predictors in most of the ocean basins. This identification of OWC850 as a prominent variable suggests that deformation forces might be stronger in non-developing cases compared to the developing cases. The order of predictors identified by all the three algorithms shows variations between ocean basins (Table 5.2, 5.3, 5.4), perhaps due to differences in basin environmental conditions due to different modes of variability and the variations in the sources of the TS formation. Also, the Beta* parameter is found to be least important in almost all the ocean basins using RF and SVM algorithms. Whereas the reduced error pruning option in the DT algorithm removes the least important parameters which excludes Beta* from the list of predictors. A recent study by Raavi & Walsh (2020b) given in chapter 4 notes the variations in the limiting factors for TS formation among the global ocean basins. Their study also showed that increased vertical shear leads to the intrusion of dry air at the mid-levels of the atmosphere leading to the non-development of the tropical depression. The thresholds of the environmental variables obtained from the decision tree rules shown in Table 5.5 vary from basin to basin due to differences in the modulating large-scale phenomenon such as El-Nino and Southern Oscillation (ENSO), Indian Ocean Dipole (IOD), North Atlantic Oscillation (NAO) and differences in the large-scale disturbances such as AEWs, monsoon troughs, ITCZ breakdown and Rossby wave trains. The hybrid approach of the DTs and MDP provides additional information to understand the TS formation process from an initial depression stage vortex. The MDP identifies that the developing cases have a high probability of remaining in the favorable atmospheric states compared to the non-developing cases, as shown in Table 5.6. The increase in the destructive forces such as dry air in the non-developing cases perhaps leads to a lower probability of maintaining favorable conditions for TC formation. Therefore, the MDP improves the certainty in the DTs' predictions when using the environmental conditions two days before formation.

5.5. Conclusion

In the current study, the pre-genesis environmental conditions, i.e., 48 hr before the genesis of both developing and non-developing tropical depressions detected from the OWZP detection algorithm, are used to predict the developing cases from three different machine learning (ML) algorithms across the global ocean basins. The ML algorithms include the decision trees (DTs), Random Forests (RFs), and the Support Vector Machines (SVMs). As the DTs are less robust than other algorithms due to their sensitivity to the input training data, a hybrid approach of DTs and Markov Decision Process (MDP) is used to improve the certainty of the TS genesis prediction from DTs. The MDP uses an iterative algorithm by considering the atmospheric state transition probabilities, current atmospheric state, and associated rewards to estimate the most probable action (i.e., either developing or non-developing) for each of the prominent environmental states obtained from the decision trees. Also, we have tested the ventilation theory of Tang and Emanuel (2012) using the three ML algorithms. We find the following main conclusions:

- The forecast skill scores for predicting developing cases from a dataset containing both developing and non-developing cases show superior performance using the RF and SVM algorithms compared to the DTs (Figure 5.2).

- The results of the ML algorithms with equivalent variables of ventilation index show a reasonable skill to predict the TC formation from initial tropical depressions.
- Decision rules formulated from the DTs unveil different combinations of the environmental conditions in quantitative terms favorable for TS formation. In summary, the relative contribution of both thermodynamical and dynamical conditions differs across ocean basins (Table 5.5).
- The MDP shows a higher likelihood of retaining in the suitable environmental states obtained from the decision tree analysis in developing cases. In contrast, non-developing cases involve atmospheric state transitions to unfavorable states (Table 5.6).

Therefore, the tropical depressions detected using the OWZP scheme can be used in the RF and SVM algorithms to identify the developing tropical depressions based on the background environmental conditions. Note that the performance of the DTs in classifying the developing and non-developing cases is moderate compared to the RF and SVM. The DTs provide a simple tree-like representation that reveals the coinfluences of different environmental conditions favoring the TS formation with their quantitative thresholds. Further application of the OWZP scheme to detect the tropical depressions in climate models and then using the ML algorithms to classify the developing and non-developing cases will further enhance our understanding of the relationships between large-scale environment and TS formation. The new hybrid approach of decision trees and MDP can improve the certainty in the prediction of developing cases of the DTs. Further studies with this hybrid approach using different forecast analysis products and with increased data samples may enhance the confidence of the current results obtained using reanalysis products.

Chapter 6. Sensitivity of tropical cyclone formation to resolution-dependent and independent tracking schemes in high-resolution climate model simulations

This chapter addresses the third research objective and is reproduced from the published paper of Raavi and Walsh (2020a): Sensitivity of tropical cyclone formation to resolution-dependent and independent tracking schemes in high-resolution climate model simulations. *Earth and Space Science*.

Abstract

In the present study, global Tropical Cyclone (TC) formation characteristics are estimated using two fundamentally different CSIRO and OWZP tracking schemes in the reanalysis data and in a high-resolution climate model with interannually varying sea surface temperatures (SSTs). Both the schemes have a reasonable global geographical distribution of TC genesis locations with under-simulation in the eastern North Atlantic and north-western Australian regions. The mean annual TC frequency in the model is similar to observations using the CSIRO scheme but higher using the OWZP scheme. Whereas, the annual frequency in reanalysis using the OWZP scheme is similar to observations but halved using the CSIRO scheme. In the CSIRO scheme, both the resolution-dependent thresholds and large-scale climate may play a role for skilful TC formation statistics. In contrast, large-scale climate leads to changes in OWZP TC detections. This highlights the importance of the large-scale environment for TC detections in both the tracking schemes. The OWZP scheme can differentiate the monsoon lows from the actual TCs in the North Indian Ocean compared to the CSIRO scheme, which incorrectly detects them as TCs in the monsoon season. The distribution of TC lifetime in the model using the OWZP scheme is similar to observations. Conversely, the CSIRO scheme detected TCs have shorter lifetimes, perhaps due to intrinsic tracking scheme differences. Although the tracking schemes are fundamentally different, the study shows that there exist some similarities between them and for certain TC formation characteristics the OWZP scheme performs better compared to the CSIRO scheme.

6.1. Introduction

Recent advances in climate models have led to more accurate simulation of the Tropical Cyclone (TC) climatology, geographical distribution, seasonal and interannual variability across various regions of the globe (Zhao et al., 2009; Bell et al., 2013; Tory et al., 2013c; Strachan et al., 2013; Shaevitz et al., 2014; Walsh et al., 2015; Camargo & Wing, 2016; Li et al., 2018). The fine-resolution climate model simulations are particularly crucial for individual ocean basins where the low-resolution models have low skill in simulating TC formation rates comparable to observations. This results in higher variability between models in individual basins, which leads to lower confidence in their future projections (Knutson et al., 2010). A study by Walsh et al. (2013) has noted a specific relationship between model-resolution and TC formation rate after considering resolution-based thresholds. Also, they demonstrated that high-resolution models had better TC formation patterns and rates.

Various observational and modeling studies have investigated the relationship between climate and TC formation. It is observed that different climate models produce distinct frequency rates across different ocean basins, and the precise cause for these variations is not clear. However, these earlier studies have identified some of the prominent factors that lead to changes in TC formation rates. These factors include temporal and geographical variations of Sea Surface Temperature (SST) (Gray, 1968; Vecchi & Soden, 2007a; Murakami et al., 2012); mid-troposphere relative humidity (Bister & Emanuel, 1997; Tang & Emanuel, 2010), vertical wind shear (McBride & Zehr, 1981; Riemer et al., 2010; Davis & Ahijevych, 2012), and a pre-existing convective environment (Hendricks et al., 2004). In addition to the above variables, other factors that lead to changes in TC formation rates are the model convective parameterization schemes (Kim et al., 2012; Zhao et al., 2012), and also different TC tracking schemes with discrete choice of thresholds, storm duration and formation latitudes (Camargo et al., 2005; Walsh et al., 2007; Horn et.al.,2014).

Using analysis from observations, TC formation research has been aided by the recent "marsupial pouch" concept defined within easterly waves of the Atlantic and Pacific regions (Dunkerton et al. 2009; Montgomery et al. 2012). This theory observes that TC formation occurs in semi-enclosed circulation regions (pouches) of easterly waves protected from surrounding influences like the entry of lower entropy drier air and high vertical wind shear. This pouch concept has been further developed to identify favorable regions for TC formation within large-scale disturbances (Wang et al., 2009; Wang et al., 2010 a, b; Wang et al., 2012a; Montgomery et al., 2010a). Based on the pouch concept, in recent studies by Tory et al. (2013a, b, c), a TC tracking scheme is developed using ERA-Interim reanalysis (Dee et al. 2011) which includes a diagnostic term known as the Okubo-Weiss Zeta parameter (OWZP), coexisting with necessary thermodynamic conditions and dynamic conditions. This OWZP represents low deformation vorticity regions favorable for TC formation. The detected pouch locations in reanalysis using this tracking scheme have a close resemblance to observed TC genesis locations (IBTrACS; Knapp et al. 2010) with 95% of the global TCs containing enhanced values of OWZ, not limited to the regions of tropical easterly waves in the Atlantic and Pacific regions. Bell et al. (2018) statistically assessed TC track climatology in ERA-Interim reanalysis using the OWZP scheme and noted good agreement between observed and detected tracks. In addition, this scheme is resolution independent and also observed to have consistent results of TC frequency as well as track changes both in current and future climate using various CMIP5 models (Tory et al., 2013d; Bell et al., 2019a, 2019b).

It would thus be useful to apply these concepts to the output of fine-resolution climate models, as a way of better defining and detecting TCs in model output. Existing TC tracking schemes tend to be resolution-dependent and use wind-speed or vorticity thresholds (Horn et al., 2014; Shaevitz et al., 2014). Studies using climate models are moving to a higher horizontal resolution to better simulate and understand TC characteristics. It has been suggested that 20-100 km horizontal resolutions perhaps sufficient to study different aspects of the TC activity (Oouchi et al. 2006; Bengtsson et al. 2007a, b). In the present study, a high horizontal resolution model of about 40 km, the atmosphere-only ACCESS model (Bi et al., 2013), is used to detect present-day TCs using two different tracking schemes across different ocean basins, with basin boundaries given in Table 4.1. The first scheme is the circulation-based OWZP tracking scheme of Tory et al. (2013a), and the second method is the Commonwealth Scientific and Industrial Research Organisation (CSIRO) detection scheme of Horn et al. (2014), a traditional TC tracking scheme that uses imposed thresholds such as minimum low-level wind speed. The OWZP detection scheme has not been examined for its performance in high-resolution models, whereas the CSIRO scheme has. Therefore, this study compares the performance of the phenomenon based tracking scheme (OWZP) with the traditional scheme (CSIRO) in a fine-resolution climate model simulation and reanalysis when compared against the observations (IBTrACS; Knapp et al. 2010) in simulating various storm characteristics such as climatology, lifetime, seasonal and interannual variability. A particularly challenging region of the globe for TC tracking schemes is the North Indian (NI) Ocean basin, due to the prevalence of monsoon depressions in this basin that can be misidentified as TCs. Accordingly, some of the tracking schemes apply additional thresholds to remove the monsoon depressions falsely detected as TCs to capture the seasonal cycle correctly across this basin (Murakami et al., 2012). Therefore, we also attempted to examine the performance of the OWZP TC detection scheme in the NI Ocean.

6.2. Model, Data and Methods

6.2.1 ACCESS model simulation

In the current study, we employ the atmospheric component of the Australian Community Climate and Earth-System Simulator general circulation model (GCM) (ACCESS, Bi et al., 2013). The description of this model is given in section 3.2. Here we use a model simulation of 20 years from 1990-2009 run with prescribed SSTs from the AMIP-II. Sharmila et al. (2020) evaluated the performance of this model in simulating large-scale climate variables in comparison to the observed climate variables of ERA-Interim reanalysis. They noted certain regional biases in large-scale fields of the model corresponding well with the genesis biases using the CSIRO scheme. Keeping in view the model performance, we compare the model and reanalysis performance in simulating the TC formation characteristics with IBTrACS for 20 years from 1990-2009 using two fundamentally different tracking schemes.

6.2.2 Observational and ERA-Interim data

In the current study, we use the International Best Track Archive for Climate Stewardship (IBTrACS) TC data for the 1990-2009 period (see also section 3.3.1). Here we use the ERA-Interim reanalysis data for comparing with the model simulated large-scale environmental variables and detected TCs using two different tracking schemes. The TCs simulated in the reanalysis data act as a link between GCM simulated TCs and the observed TCs. The reanalysis data are spatially and temporally homogeneous assimilated observation data but have some constraints which are discussed in section 3.3.2. Recent studies have showed that TC intensity is underestimated in the ERA-

Interim reanalysis compared to the observations (Murakami, 2014; Hodges et al., 2017). Two different TC tracking schemes are applied to the ERA-Interim data at a resolution of $0.75° \times 0.75°$ at 12 hourly intervals for 20 years from 1990-2009. The OWZP tracking scheme requires the zonal and meridional wind fields at 850, 500 and 200 hPa pressure levels to estimate the OWZ, the magnitude of vector wind difference between 200 and 850 hPa (the vertical wind shear) and the TC steering velocity (average wind at 700hPa within a box region). In addition to the above variables, it requires relative humidity at 950 and 700 hPa levels, and specific humidity at the 950 hPa level. Similarly, the CSIRO tracking scheme requires the 10m, 850 hPa, and 300 hPa zonal and meridional winds and mean sea level pressure.

6.2.3 Tracking schemes

Here we use a resolution-dependent TC tracking scheme (CSIRO) and a resolution-independent scheme (OWZP) to detect the TCs with the model simulation and the reanalysis. The detailed description of the tracking schemes are given in the section 3.4.

6.3. Results

6.3.1 Simulated TC formation characteristics

The quality of the TC formation climatology in the current climate is essential for any climate model to obtain reliable TC projections under changing climate conditions.

6.3.1.1 Geographical distribution of TC genesis locations

This section illustrates the geographical distribution of the TC genesis locations from IBTrACS and the ACCESS model detected using both CSIRO and OWZP TC tracking schemes (Fig. 6.1).

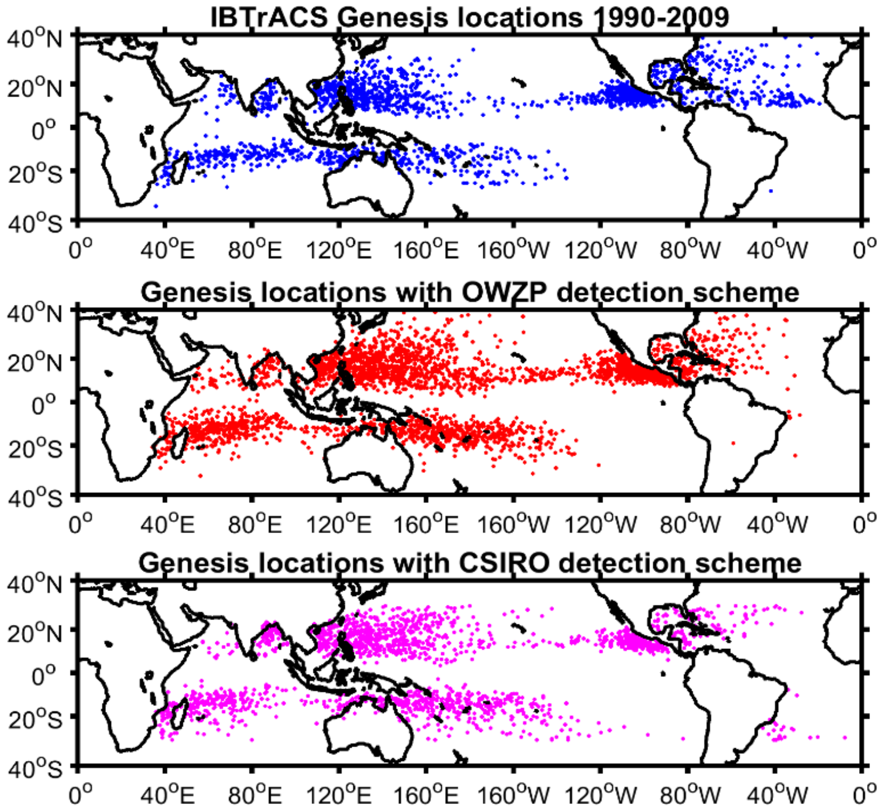

Figure 6. 1. TC genesis locations in observations and the ACCESS model (using both CSIRO and OWZP schemes) for a period of 20 years from 1990 to 2009.

The OWZP detected TCs are more frequent than observations in most of the ocean basins with an under-simulation in the Eastern NA and north-western part of the Australian (AUS) region. In contrast, the CSIRO detection scheme has larger under-simulation in the NA basin. From this, we notice that the ACCESS model simulates reasonably well the TC climatology, and both the tracking schemes identify the TC genesis detections at geographically similar locations to the observations. Similarly, the genesis locations in the reanalysis have a close resemblance to observations using both the schemes (Figure B.1; appendix B).

Figure 6.2 shows the zonal and meridional distribution of TC genesis locations expressed in terms of a mean frequency at each 4-degree grid box across the globe. The zonal mean distribution has two peaks, one in the Northern Hemisphere (NH) and the other in the Southern Hemisphere (SH) at around 15° latitude in the observations (IBTrACS). There is a good agreement in the zonal distribution between the detected TCs of the model and reanalysis using both the schemes and observations (IBTrACS). Both the schemes have peak locations at similar latitudes except with a few more detections in the model at these latitudes using the CSIRO scheme in the SH and significantly more detections using the OWZP scheme in both hemispheres.

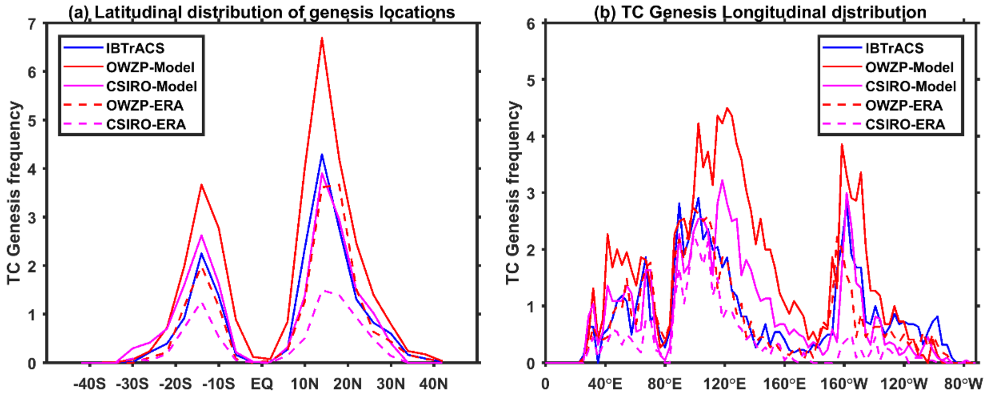

Figure 6. 2. Zonal (a) and meridional (b) mean TC frequency (per 4°×4° box region across the globe) as a function of latitude and longitude in both observations and model (using both the CSIRO and OWZP detection schemes).

Similarly, from the meridional distribution of mean TC genesis frequency in IBTrACS, Figure 6.2 shows two sharp peaks, one at around 90° E in the South Indian (SI) Ocean and the other at 110° W in the Eastern North Pacific (ENP) region. Also, there is a broader peak at around 150° E in the Western North Pacific (WNP) and South Pacific (SP) regions. The meridional distribution in the reanalysis is similar to the observations using both the schemes. The CSIRO detections within the model has peaks at similar longitudes but with an eastward shift of the 150° E peak. In contrast, the OWZP detection scheme in the model has a 150° E (WNP & SP) peak shifted further eastward. There exists a peak at around 85° W and 70° W (ENP & NA) in the model using the OWZP scheme, which is not present using the CSIRO scheme. Therefore, the zonal and meridional mean TC frequency distributions in the model show differences across the ENP, NA, SP, and WNP basins for both the schemes when compared against observations.

6.3.1.2 TC genesis density bias

Here, the global TC genesis density is defined as the number of storms per year and per 4°×4° box region across the globe as shown in Figure B.2 and B.3 for both model and ERA-Interim data, respectively. To understand the quantitative differences between the detection schemes in simulating TCs, we calculate the genesis density bias by subtracting the observations (IBTrACS) from the model and reanalysis detected TCs and

Table 6.1 and Table B.1 shows its basin-averaged values. These basin-averaged values in the reanalysis (Table B.1) show that the CSIRO scheme has a negative bias in almost all the ocean basins compared to the OWZP scheme, which performs better in most of the ocean basins. In the model, basin-averaged values (Table 6.1) show a negative genesis bias in the NA-MDR with both the schemes and in the ENP basin using the CSIRO scheme.

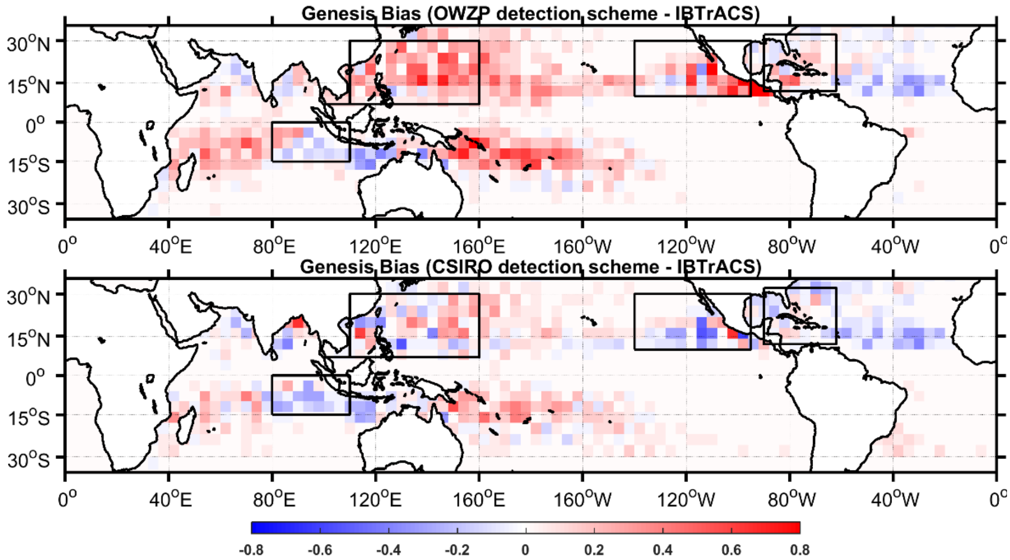

Figure 6. 3. TC genesis density bias in the model using the CSIRO and OWZP tracking schemes when compared against observations (IBTrACS). TC genesis density is defined as the number of storms per 4°×4° box region across the globe for a 20 year period from 1990 - 2009. The black rectangular boxes indicate the major differences in genesis density between the two schemes when compared with observations.

The pattern correlation coefficient of the latitude wise global TC genesis locations between the OWZP detected cases and IBTrACS is 0.79 and the correlation coefficient between the CSIRO detected cases and IBTrACS is 0.77. The spatial genesis density bias (Fig. 6.3) of the model has similarities between both the tracking schemes in the model in the NI, SI, SP and the AUS basins with a similar sign of the bias in the two schemes, but with different magnitudes. There exist two basins with significant differences between both the schemes in the model: one being the NA-MDR basin, having negative bias using both the schemes with significantly higher negative bias using the CSIRO scheme (Table 6.1). The OWZP scheme is able to capture the NA-MDR genesis frequency within the model to some extent which is difficult to capture in most climate models (Walsh et al., 2016). The other basin showing substantial differences between the schemes is the ENP with strong positive bias throughout the basin using the OWZP detection, while the CSIRO scheme shows overall negative bias across the basin. In addition, the WNP basin within the model has patches of negative bias using the CSIRO scheme. From these differences, we can infer that the OWZP scheme has better performance in the NA-MDR, WNP and ENP basins within the model compared to the CSIRO scheme, which has negative genesis density bias in these regions. To understand genesis biases (positive/negative) across the ocean basins, we perform a climatological analysis of large-scale climate variables.

Table 6. 1. Basinwise mean genesis density bias (Model minus observations) per year using both the detection schemes in ACCESS model when compared against IBTrACS and percentage bias of the storms when compared to observations (* indicates 95% significance level)

BASIN	OWZP bias (bias/year)	CSIRO bias (bias/year)	IBTrACS (TCs/year)	% bias of OWZP storms	% bias of CSIRO storms
MDR_NA	-2.35	-4.8*	6.6	(-) 36%	(-) 73%
ENP	15.6*	-1.7*	15.7	(+)100%	(-) 11%
WNP	21.1*	3.55*	24	(+) 88%	(+) 15%
NI	1.15*	0.4*	4.05	(+) 28%	(+) 10%
SI	6.55*	-0.4*	15.3	(+) 43%	(-) 2.6%
SP	16.6*	9.15*	8.9	(+) 186%	(+) 102%
AUS	5.0*	1.4*	9.0	(+) 55%	(+) 15.5%
Global	61	5.6	82	+74.4%	+5.6%

Previously performed basin wise analysis of this model noted a relationship between TC genesis locations detected using the CSIRO scheme and the large-scale climate of the model (Sharmila et al., 2020). Following the analysis of Sharmila and Walsh (2017), the current analysis includes both dynamical variables such as Omega at 500 hPa, vertical wind shear between 850 and 200 hPa (VWS), 850 hPa relative vorticity ($Rvor_{850}$), along with thermodynamic variables including the maximum potential intensity (MPI) and 700 hPa relative humidity (RH_{700}). The climatological bias of the large-scale atmospheric fields for TC peak seasons as shown in Figure 6.4 and Figure 6.5 (July, August, September (JAS) for NH; October, November, December over NI; January, February, March for SH) between the ACCESS model and ERA-Interim reanalysis reveals some of the possible causes behind the TC genesis bias.

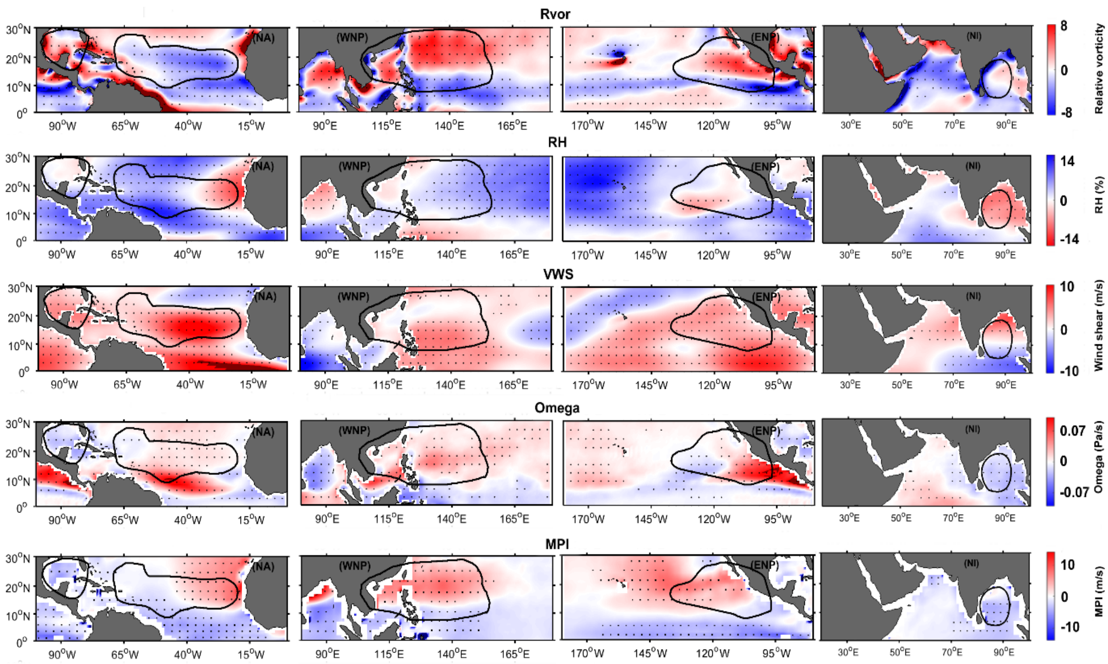

Figure 6. 4. Spatial bias of the large-scale environmental variables between ERA-Interim and ACCESS model. The stippling indicates the 95% significant values and the black contour 2 x 10^{-2} indicates the observed (IBTrACS) main TC genesis development regions in each of the ocean basins in the Northern Hemisphere.

Over the NA basin, the unfavorable environmental conditions in the model (Figure 6.4) such as lower values of $Rvor_{850}$ and MPI along with higher values of VWS and Omega (lesser upward mass flux) may lead to the negative genesis density bias. Over the WNP basin the positive bias in the Rvor850, MPI and RH_{700} variables (Figure 6.4) in the model lead to positive genesis bias. In the ENP basin, the higher values of $Rvor_{850}$ and MPI lead to positive genesis bias in the model using the OWZP scheme. The lower values of MPI and RH_{700} lead to negative genesis bias across the NI basin. Similarly, the northwestern side of the AUS region have higher bias in the $Rvor_{850}$, VWS and RH_{700} environmental variables (Figure 6.5) leading to the negative genesis bias in the model. The favorable values of Rvor850, RH_{700} and MPI across the SP basin lead to positive genesis bias in the model. From this analysis, we notice that higher values of VWS and mid-level Omega (implying descent), lower values of RH_{700}, MPI, and $Rvor_{850}$ lead to negative genesis density bias using both the schemes across different ocean basins (Figure 6.4 & Figure 6.5). Conversely, higher values of RH_{700}, MPI, $Rvor_{850}$, and lower values of mid-level Omega lead to positive bias with both the schemes. The more positive TC genesis density bias in the model using the OWZP detection scheme (Figure 6.3) compared to observations may be due to the more favorable large-scale climate within the model. Therefore, both the schemes show a relationship between large-scale environmental conditions and the TC genesis.

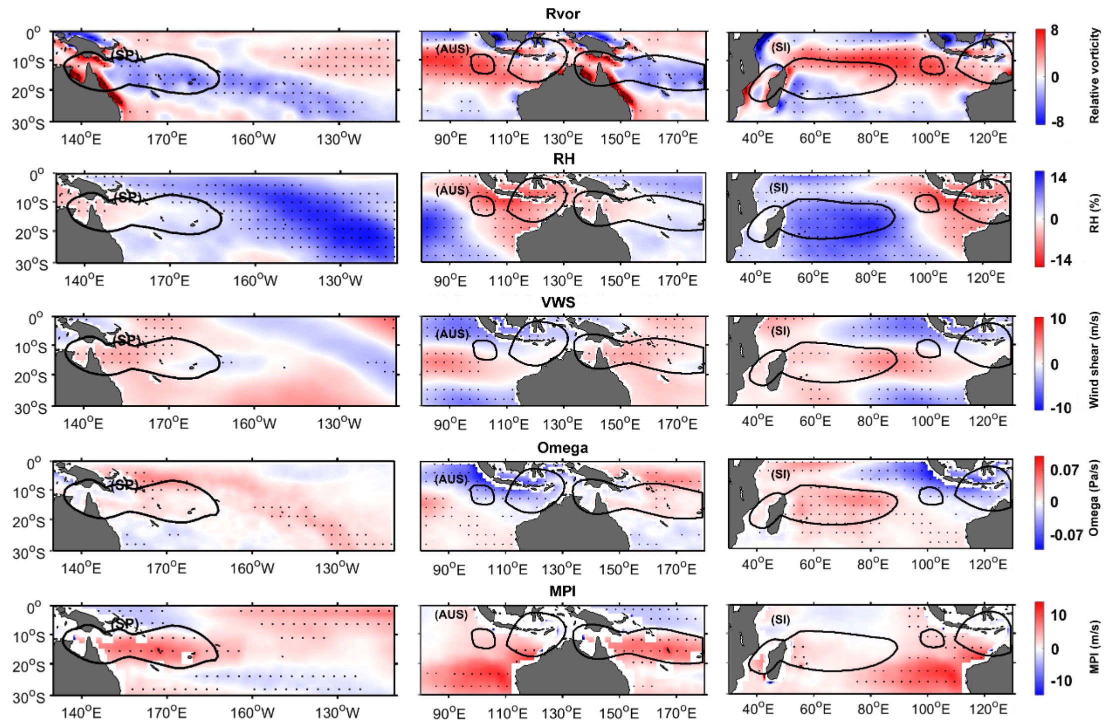

Figure 6. 5. The same as Figure 6.4 but in the Southern Hemisphere.

6.3.1.3 Mean annual number of TCs

Every year around 80 cyclone form across the globe under certain dynamical and thermodynamical conditions (Gray, 1975; Emanuel, 2003). The asymmetry of global annual mean TC formations in the NH and the SH is comparable to the observations using both the schemes (Figure 6.6). In addition, the SH is found to be more active in the model compared to observations. The global annual mean TC detections identified in the model output with the OWZP scheme is 143, whereas IBTrACS observations have 82.4 detections. The OWZP detection scheme shows 50% more TCs than the observed annual TC numbers, a statistically significant difference. In contrast, the mean annual TC number in the model with the CSIRO detection scheme is 86. We have performed a similar analysis with ERA-Interim data, also for 20 years from 1990 to 2009. The mean annual number of detections using the OWZP scheme in the ERA-Interim is 75, which is close to the observations. In contrast, with the CSIRO detection scheme, the mean annual TC frequency is 36.5, which is 50% less than observations, as shown in Figure 6.6.

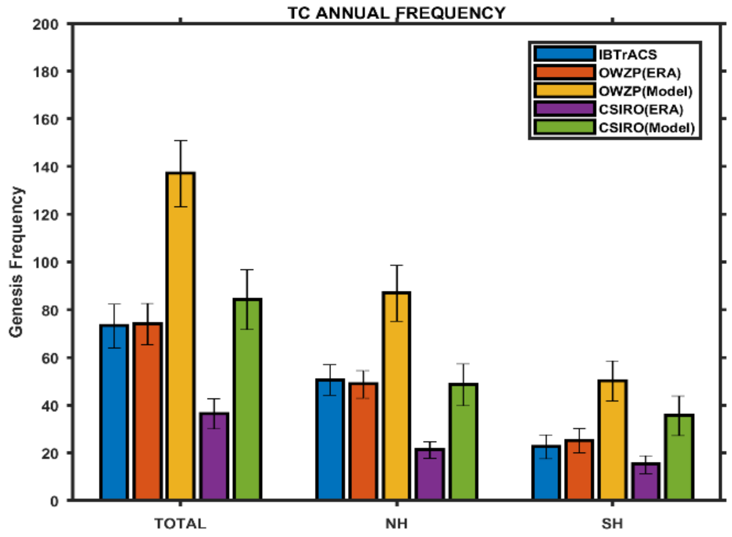

Figure 6. 6. Mean annual TC frequency globally and in Northern and Southern Hemispheres in observations (IBTrACS), ACCESS model and ERA-Interim reanalysis using both OWZP and CSIRO detection schemes for a period of 20 years from 1990-2009

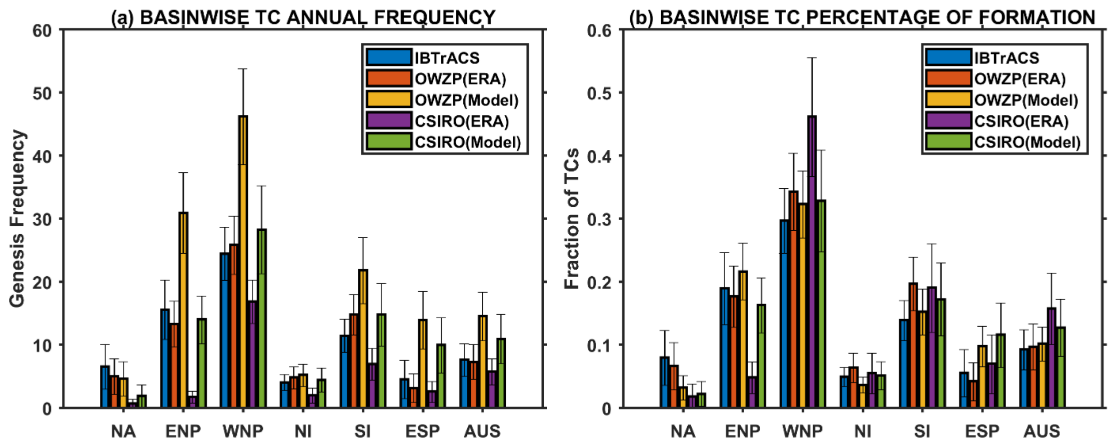

Figure 6. 7. Basinwise mean annual number of TC genesis detections in the ACCESS model and ERA-Interim using both CSIRO and OWZP detection schemes and IBTrACS. (a) The total number of TCs, (b) the fraction of TCs in each basin compared to the global number of TCs.

The finer resolution in the ACCESS model data (0.36°) compared to the reanalysis (0.75°), leads to an increased number of detections in the model with both the tracking schemes. The increased TC detections with increasing resolution using the CSIRO detection scheme may be due to a more favorable large-scale climate and better simulation of TC intensity in high-resolution models. The resolution-dependent 10m wind speed threshold is derived from the azimuthal mean TC intensity profile (Walsh et al. 2007). Murakami (2014) also noted that the CSIRO detection scheme underestimates annual TC numbers in low-resolution reanalysis indicating that the TC intensity is too weak in reanalysis datasets for its resolution if the only effect of the resolution is horizontal smoothing. In contrast, the more favorable large-scale climate within the model leads to more detections using the OWZP scheme compared to the reanalysis.

Figure 6.7a summarizes the basinwide mean annual frequency of the observed TC genesis along with the TCs simulated in the model and reanalysis using both the detection schemes over the 20 year period with higher detections in the WNP basin. Figure 6.7b shows that the relative percentage of the storms detected by both the tracking schemes in the model and reanalysis across each of the ocean basins; for instance, the highest number of detections in the WNP basin is reproduced similar to the IBTrACS. The fraction of TCs in reanalysis using the CSIRO scheme shows that they have more TCs in the WNP and very few TCs in the NA. Therefore, the resolution-independent scheme (OWZP) is observed to have superior performance in simulating TC mean annual numbers and fraction of TCs in both the ACCESS model and ERA-Interim reanalysis data across the global ocean basins compared to the resolution-dependent CSIRO scheme.

6.3.1.4 Seasonal cycle of TC genesis frequency

The seasonal cycle of TC formation across different ocean basins using both CSIRO and OWZP detection schemes with ACCESS model output is compared with IBTrACS, as shown in Figure 6.8. It is observed that both the model and reanalysis data can capture the seasonal cycle in all the ocean basins except in the NI Ocean, with good similarity to observations using both the schemes with the statistical correlations given in Table 6.2. The seasonal cycle in the NI Ocean has a bimodal distribution with two observed peaks in the pre- and post-monsoon seasons (observations) and is not captured well with CSIRO detection scheme in both the model and reanalysis but captured reasonably well using the OWZP detection scheme. The NI basin is where general circulation models with some tracking schemes have difficulty in capturing the seasonal signal due to false alarms of monsoon depressions being identified as TCs during the monsoon season (Murakami et al., 2012, 2013a; Manganello et al., 2012; Camargo et al., 2016). The reason could be that monsoon depressions have certain similar characteristics with actual TCs (Murakami et al., 2011). The CSIRO detection scheme has higher TC detections in the monsoon season (July and August) in the ACCESS model, which degrades the simulated seasonal cycle across this basin.

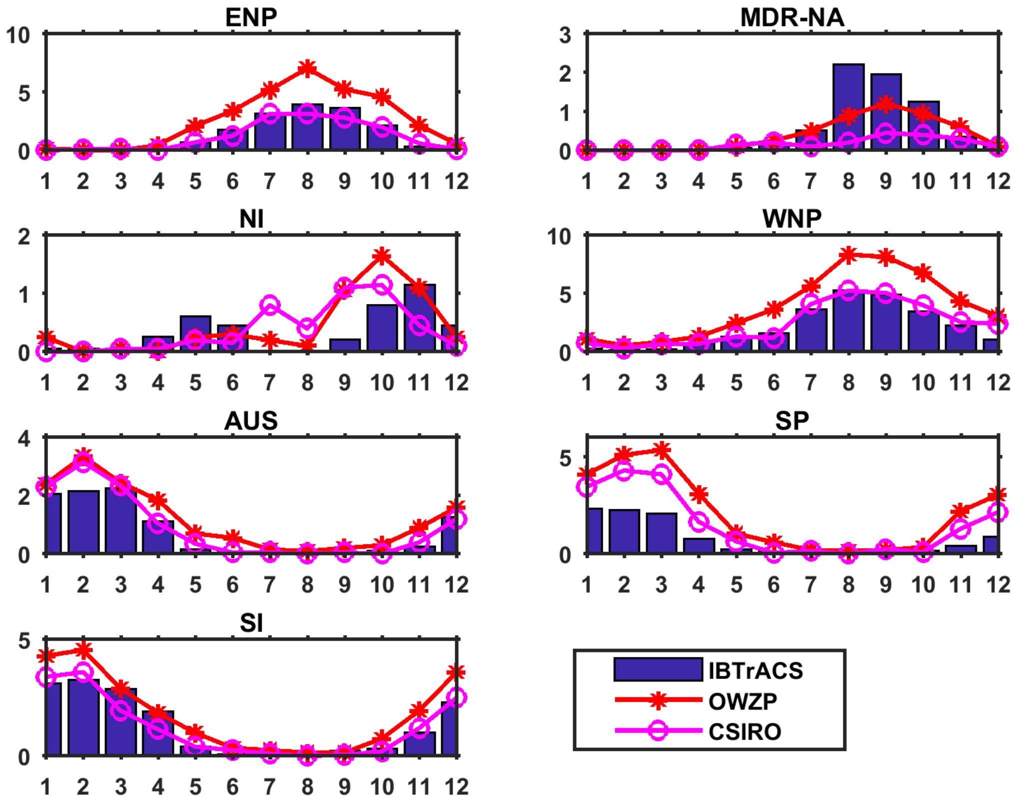

Figure 6. 8. TC seasonal cycle in each basin in both observations (IBTrACS) and ACCESS model detected storms using both CSIRO and OWZP schemes.

Table 6. 2. Basinwise seasonal TC frequency correlations between IBTrACS observations and the two detection schemes applied to ACCESS model output (* indicates above 95% significance level)

BASIN	OWZP (Model)	CSIRO (Model)	OWZP (ERA)	CSIRO (ERA)
MDR_NA	0.92*	0.71*	0.97*	0.94*
ENP	0.97*	0.99*	0.99*	0.85*
WNP	0.99*	0.97*	0.99*	0.99*
NI	0.67*	**0.24**	0.86*	**0.27**
SI	0.96*	0.95*	0.88*	0.97*
SP	0.94*	0.98*	0.89*	0.95*
AUS	0.96*	0.97*	0.88*	0.97*

6.3.2 Interannual variability

The most dominant natural mode of climate variability in almost all the ocean basins, the El-Nino and Southern Oscillation (ENSO) phenomenon, is observed to influence the interannual variability of TC genesis frequency and its characteristics (Camargo et al., 2010; Bell et al., 2013). Therefore, in the current section, we have investigated the model's performance in simulating the basin-wide interannual variability of the mean annual TC frequency in the ACCESS model using both the detection schemes when compared with the observed annual frequency. Table 6.3 shows the interannual correlations between IBTrACS and detected annual TCs using both the detection schemes applied both to the model output and to reanalysis. Overall, the detection schemes give better correlations when applied to reanalysis than to the model output. More so, the OWZP scheme appears to give better correlations than the CSIRO scheme, particularly in the reanalysis. We note that the correlations in Table 6.3 are from a single simulation. The interannual correlations of TC annual count with observations in atmospheric general circulation models tend to improve in ensemble model simulations (Vitart et al. 1997; Vitart et al. 1999; Strachan et al. 2013; Shaevitz et al. 2014).

Table 6. 3. Correlations of the interannual variability of annual TC genesis detections from the model using two different tracking schemes compared with IBTrACS observations (bold letters indicate statistically significant values; * indicates 95% significance level)

BASIN	OWZP (Model)	CSIRO (Model)	OWZP (ERA)	CSIRO (ERA)
MDR_NA	**0.59***	**0.52***	**0.88***	**0.68***
ENP	0.3	**0.62***	**0.79***	0.2
WNP	**0.58***	0.22	**0.7***	**0.53***
NI	-0.12	0.18	0.38	0.03
SI	0.06	0.26	**0.51***	0.29
SP	0.07	0.14	**0.8***	**0.63***
ESP	0.24	0.15	**0.79***	**0.73***
AUS	**0.48***	-0.09	**0.55***	0.12

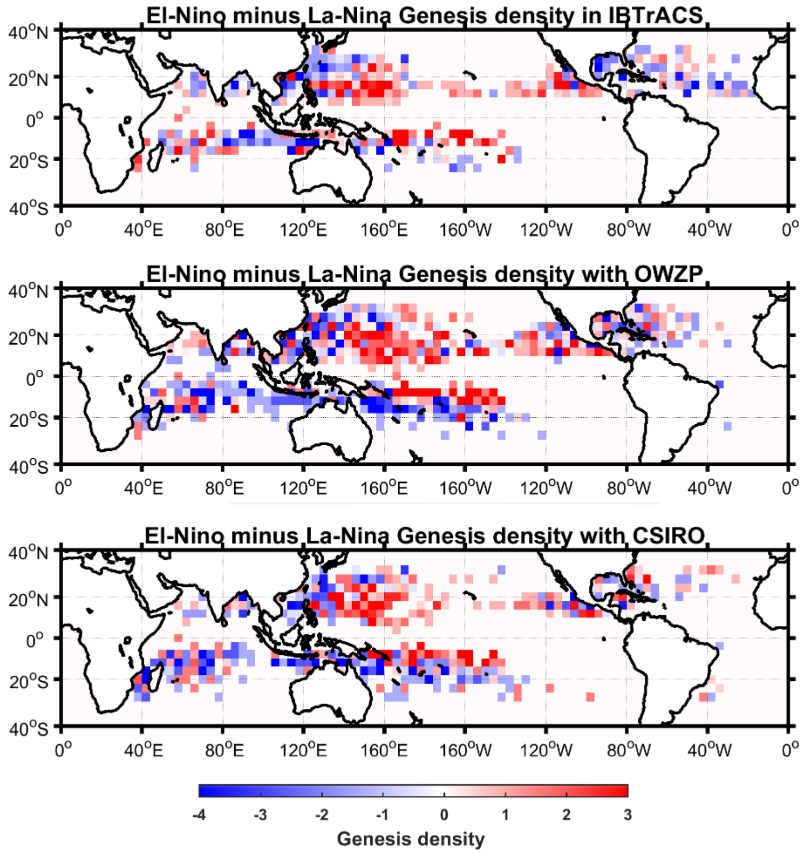

Figure 6. 9. Geographical distribution of the difference in the TC genesis density of El-Nino and La-Nina years, with genesis density being the number of TCs per 4°×4° box region per year, for (top) IBTrACS; (center) ACCESS model with the OWZP detection scheme; and (bottom) ACCESS model with the CSIRO detection scheme.

We further examine the model's ability to capture the variations in TC genesis locations during different phases of ENSO due to changes in atmospheric circulation patterns for a period of 20 years from 1990 - 2009. The model is run with interannually varying SSTs, and this partially transmits the observed ENSO signal, therefore the NH (SH) El-Nino and La-Nina years within the model are considered similar to observations. The composites of TC genesis frequency per 4-degree box region across the global ocean basins are computed for El-Nino and La-Nina years given in Table 6.4 separately for the NH and SH (Sharmila et al., 2020). We use the definition of the ENSO index (the Oceanic Niño Index (ONI) from the Climate Prediction Centre of the National Oceanic and Atmospheric Administration (NOAA)) to determine the warm and cold phases. Figure 6.9 shows the genesis density bias between El Nino and La-Nina phases for both observations and model-simulated TCs using both the schemes. The changes in TC genesis frequency due to ENSO in observations (IBTrACS) show suppressed genesis during El Nino in the Bay of Bengal (Felton et al. 2013), increased genesis with a shift to the southeast region of the WNP (Wang & Chan 2002; and Camargo et al. 2007), increased genesis with a shift towards the southwest in the ENP (Chia & Ropelewski, 2002) and suppressed genesis in the NA basin due to increased vertical wind shear (Shaman et al., 2009). In the SH basins, there is increased genesis in the west of the SI ocean (i.e., between 75°E and 135°E), suppressed genesis in the west of Australian basin (i.e. between 90°E and 135°E), and increased genesis in the SP close to the east coast of Australia (i.e. between 140°E and 170°E) (Kuleshov et al., 2008).

Table 6. 4. El-Nino and La-Nina years considered for the analysis based on the ENSO states from the climate prediction center in both Northern (ASO) season and Southern Hemispheres (JFM) season. In SH basins TC seasons are considered from July to June enclosing two years.

Northern Hemisphere		Southern Hemisphere	
El Nino	**La Nina**	**El Nino**	**La Nina**
1991 1994 1997 2002 2004 2006 2009	1995 1998 1999 2000 2007	1991/1992 1994/1995 1997/1998 2002/2003	1995/1996 1998/1999 1999/2000 2000/2001 2005/2006 2007/2008

The model-simulated genesis frequency in the NH basins captures the increase and southeast shift in WNP, increase and a southwest shift in ENP using both tracking schemes, and the suppressed TC genesis frequency in the NA-MDR basin is captured to a lesser extent by the CSIRO detection scheme compared to the OWZP scheme. In the SH basins, the increased genesis in the SI ocean (west of 135°E) and the east coast of Australia is captured moderately by both the schemes. Additionally, the suppressed genesis on the west coast of Australia basin is captured to a greater extent by the OWZP scheme compared to the CSIRO scheme. From the above analysis, we observe that both the tracking schemes can capture the major differences due to the ENSO phenomenon in most ocean basins. A similar analysis is performed in the reanalysis data with the OWZP scheme showing better performance compared to the CSIRO scheme (Figure B.4).

6.3.3 Geographical distribution of climatological TC track densities

The track density distribution is used to investigate the differences in the spatial distributions of TC tracks in the model using both the tracking algorithms. The tracks of each TC for the 20 years (i.e., 1990-2009) obtained from both the tracking schemes are composited per 4-degree box region and year across the globe, as shown in Figure 6.10. The OWZP scheme detected tracks have a wider spatial extent of track densities, indicating longer tracks and higher genesis frequency compared to the CSIRO detection scheme. The simulated track densities of both the schemes are well comparable with observations (IBTrACS) in most of the basins. The OWZP scheme better captures the track density across the Atlantic basin compared to the CSIRO scheme. In addition, the track density bias of Figure B.5 shows strong negative bias in the north of the WNP basin using both the schemes and negative bias on the east coast of the ENP basin using the CSIRO detection scheme. Furthermore, the OWZP scheme has a more positive track bias in agreement with the positive bias in the genesis plots (Fig. 6.3).

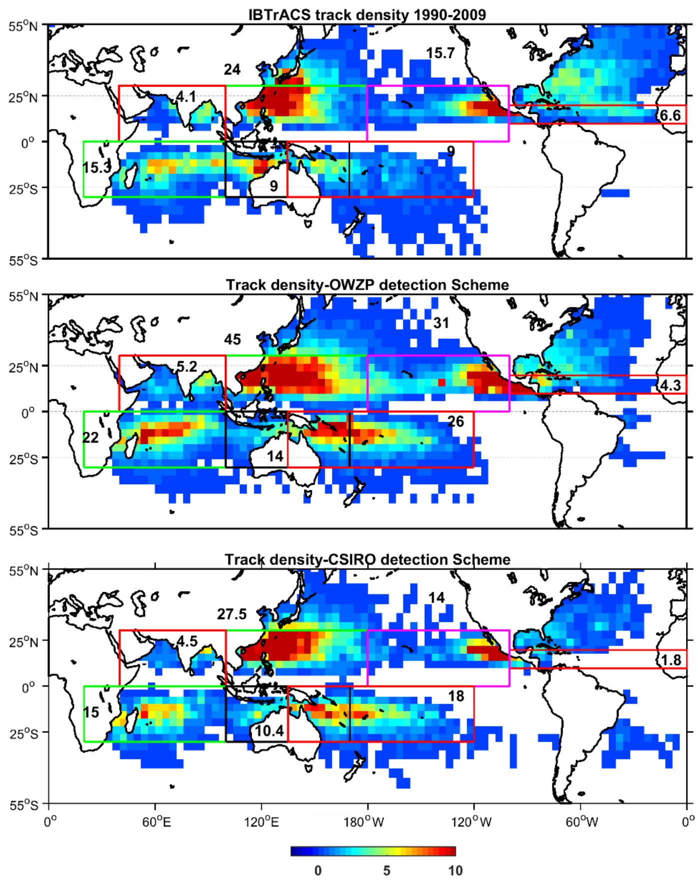

Figure 6. 10. TC track density (the number of tracks per 4°×4° box region and per month) across the globe in both model and observations. The values in the box regions indicate the average annual storms in each ocean basin.

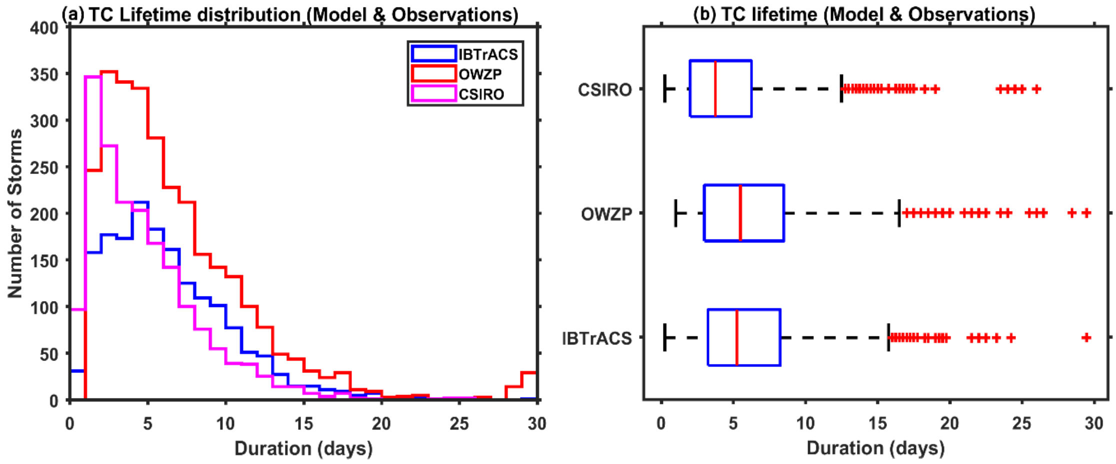

Figure 6. 11. (a) Histograms of TC lifetime in both observations and model using both detection schemes (b) Distributions of TC lifetime with the vertical line of the box indicate the median of the distribution, the left and right box edges are the 75th and 25th percentiles, the whiskers indicate the maximum and minimum values and the red crosses indicate the outliers.

Figure 6.11 shows the TC lifetime histograms of observations and storms detected in the model using both the schemes and their TC lifetime distributions. The CSIRO detected storms have shorter duration compared to the OWZP detected storms that have a longer duration almost similar to the observations. Similar results are observed in the reanalysis data with OWZP detected storms having longer lifetimes (Figure B.6). These differences in the storm's lifetime may be partly due to differences in the duration thresholds, with 24 hours in the CSIRO scheme and 48 hours in the OWZP scheme. A recent study by Horn et al. (2014) also noted that TC lifetime is highly sensitive to the duration threshold of the tracking schemes.

6.3.4 Case studies of TCs detected by both the tracking schemes within the model in the Australian region

The CSIRO detected storms are found to be a subset of the OWZP detected storms in the ACCESS model data in all the basins, as shown in Figure 6.12. To understand the differences and similarities between the TCs detected using both the tracking schemes, the particular year 1990 is selected as a test case to see in detail the behavior of these storms. Figure 6.13a shows that CSIRO tracks are shorter and are a subset of OWZP tracks in the model, except one case of the CSIRO scheme, which has a different formation latitude and track compared to the OWZP scheme.

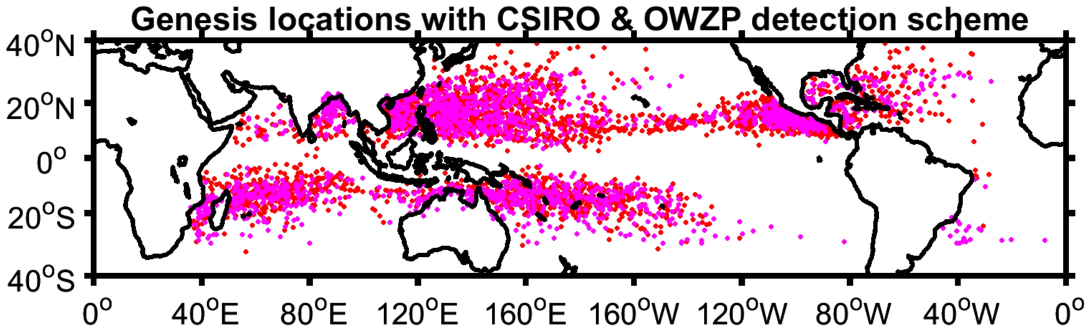

Figure 6. 12. TC genesis locations in the ACCESS model using OWZP scheme (red dots) and CSIRO scheme (magenta dots) for a period of 20 years from 1990-2009. The genesis locations of CSIRO scheme are found to be a subset of OWZP scheme in all the ocean basins.

A similar track analysis is performed using ERA-Interim reanalysis over the Australian region, and we observe that the CSIRO tracks are a subset of the OWZP detected tracks with lengthier tracks in the OWZP detected storms (Fig. 6.13b). Also, the OWZP and the CSIRO detected tracks in the reanalysis closely resemble the IBTrACS tracks but are shorter. This may be partially due to an issue in IBTrACS rather than an inherent failing in the TC detection schemes. The lack of a consistent criterion for extratropical transition may be contributing to an artificial extension of IBTrACs well into the Southern Hemisphere mid-latitudes (Sinclair 2004). This can also be seen in Figure 6.10 south of New Zealand. The CSIRO scheme has fewer tracks across the basin aligned with fewer detections in the reanalysis. These results indicate that although both the tracking schemes are fundamentally different they show similarities in their storm tracks with longer tracks using the phenomenon based tracking scheme (OWZP).

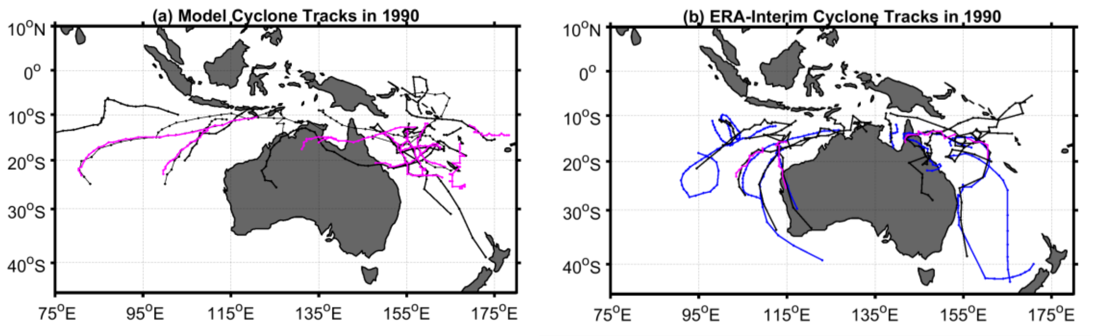

Figure 6. 13. (a) TC tracks in ACCESS using both OWZP(Black) and CSIRO (Magenta) schemes (b) TC tracks of the storms detected in the ERA-Interim using both CSIRO (Magenta), OWZP (Black) detection schemes and IBTrACS (blue) for the year 1990.

Table 6. 5. Comparison of tracking schemes for different TC characteristics in the ACCESS model and reanalysis data (* indicates statistical correlations of mean TC frequency; -ve indicates negative). Bold text shows better performing tracking scheme

Characteristics	**OWZP (ERA-Interim)**	**CSIRO (ERA-interim)**	**OWZP (Model)**	**CSIRO (Model)**	**Remarks**
Mean annual frequency	Similar to IBTrACS (75)	50% < IBTrACS (36.5)	50% > IBTrACS (143)	Similar to IBTrACS (86)	Better performance in ERA-interim using the **OWZP scheme**
Spatial Genesis density bias	(-ve) bias in the ENP, SP and NA-MDR regions	(-ve) bias in almost all the ocean basins	(-ve) in the NA and AUS basins	Higher (-ve) bias in the NA-MDR; (-ve) bias in ENP, WNP, SI and AUS basins	Better performance using the **OWZP** scheme in the model across the NA-MDR, ENP and WNP
Mean Genesis density bias	(-ve) bias in the ENP, SP and NA-MDR regions	(-ve) bias in almost all the ocean basins	(-ve) bias in the NA-MDR region	Higher (-ve) bias in the SI, ENP and NA-MDR region	Better performance of the **OWZP scheme** in the model across SI, ENP and NA-MDR basins
Seasonal variability*	Similar to IBTrACS	Not captured in NI basin	Similar to IBTrACS	Not captured in NI basin	Captured the NI seasonal cycle by the **OWZP scheme**
Interannual variability*	Significant correlations in most of the ocean basins	Significant correlations in a few ocean basins	Significant correlations in only three ocean basins	Significant correlations in only two ocean basins	Performs better in the reanalysis compared to model using the **OWZP scheme**
Lifetime of storms	Similar to IBTrACS	Shorter than IBTrACS	Similar to IBTrACS	Shorter than IBTrACS	Better lifetime of tracks using the **OWZP** scheme in both reanalysis and the model
ENSO induced changes in genesis locations	Similar to IBTrACS	Captured the ENSO changes in a few ocean basins	Captured signal in most of the ocean basins	Similar to the OWZP scheme except in NA and west coast of AUS	Better ENSO signal using the **OWZP** scheme in reanalysis and the model

6.4. Discussion and Conclusion

In the current study, we examine differences between a phenomenon-based tracking scheme (OWZP) and a more traditional TC tracking scheme (CSIRO) in estimating the TC formation characteristics in the reanalysis and in a high-resolution ACCESS climate model simulations run with interannually varying SSTs when compared against observations. A comparison of the sensitivity of TC formation to these two tracking schemes both in the model and reanalysis data is illustrated in Table 6.5.

Although not many variations are observed in geographical TC distribution and interhemispheric asymmetry of global TC frequency, nevertheless noticeable differences can be observed in spatial TC genesis density bias regionally (Figure 6.3). The NA-MDR and the ENP regions in the model show a higher underestimation by the CSIRO scheme compared to the OWZP scheme (Table 6.1). The climatological biases in the large-scale climate variables of the model (Figure 6.4 & 6.5) correspond to the biases of the genesis of both the tracking schemes (Sharmila et al., 2020). Different studies have noted that the climate models that simulate the TCs are sensitive to the convective parameterization schemes for producing deep convection (Kim et al., 2012; Zhao et al., 2012; Camargo & Wing, 2016). Therefore, either running the global models at high horizontal resolution (around 1 km) without the requirement of parameterization schemes or improving convection schemes may improve the model simulated large-scale variables that reduce the genesis biases.

Similarly, variations are observed in the global mean annual TC frequency of the model in both the schemes (Figure 6.6). Here, the OWZP scheme shows higher frequency, while the CSIRO scheme is more comparable to the observations. The mean annual number in the reanalysis using the OWZP scheme is similar to observations, while the CSIRO scheme has fewer detections compared to observations. The increased TC numbers in the ACCESS model using the OWZP scheme may be due to grid resolution bias of the detection scheme but is smaller compared to the effect of the model physics, as stated by Tory et al. (2013d). Therefore, the increased number of detections using the OWZP detection scheme may be due to the more favorable large-scale climate of the ACCESS model. The reduced number of detections in the reanalysis using the CSIRO scheme is due to a resolution-dependent 10 m wind speed threshold, which causes an underestimation of TC numbers in the low-resolution reanalysis datasets (Murakami, 2014; Hodges et al., 2017). The resolution-dependent (CSIRO) TC tracking scheme has better performance in the high-resolution ACCESS model compared to the lower resolution reanalysis, we speculate that the model has improved the simulation of TC intensity compared with the reanalysis. ERA-Interim reanalyses are also not a perfect representation of TC characteristics, and the consistent underrepresentation of TC intensity in these reanalyses (e.g. Murakami 2014) would lead to less reliable detection and (arguably) shorter TC tracks. Accordingly, a high-resolution model with improved model physics gives a better simulation of TC frequency and intensity characteristics, an improvement that is greater than merely the effect of less smoothing due to the improved model horizontal resolution (Walsh et al., 2013).

Also, the TC seasonal cycle of the model and reanalysis using both the tracking schemes has good statistical correlations when compared with observations in all the ocean basins except the NI Ocean (Table 6.2), where the CSIRO tracking scheme has a weaker correlation due to higher TC detections in the monsoon season. In comparison, the OWZP scheme eliminates monsoon low detections in the NI Ocean. The model has statistically significant interannual correlations in the NA and ENP basins. In contrast, using the OWZP scheme, significant correlations are note in the NA, WNP, and AUS basins (Table 6.3). The OWZP scheme has better interannual correlations in the reanalysis compared to the CSIRO scheme. The weaker interannual correlations in the model using the OWZP scheme compared to reanalysis may be due to the model's performance in simulating the large-scale fields and their temporal variation. It is also evident from the study that the CSIRO scheme detected storms are a subset of the OWZP detected storms, with shorter TC lifetime.

Although there are specific differences, the model’s performance in simulating the geographical distribution, the relative percentage of TCs in each basin, seasonal variability (except NI), and ENSO induced changes in TC genesis locations using both schemes are comparable to the observations (Figure 6.9). From this analysis we see that the resolution-independent scheme has better performance in both the ACCESS model and reanalysis. In contrast, the resolution-dependent scheme has better performance in simulating TC frequency characteristics in the ACCESS model compared to reanalysis. Though model resolution plays a role in simulated TCs, it is not the only factor for better TC simulation in climate models. Therefore, this improved TC simulation in the high-resolution model using the CSIRO scheme may be due to increased horizontal resolution as well as improved model physics.

Future studies focus on applying this phenomenon based tracking scheme (OWZP) and resolution-dependent tracking scheme (CSIRO) to other current state-of-the-art high-resolution climate models. Comparisons across models are essential in climate projection studies as there exist differences in projections with different tracking schemes due to the fundamental differences between tracking schemes and also to threshold differences (Horn et al., 2014). Models with increased resolutions might reduce the differences between the tracking schemes as they better simulate the TC intensities. However, there are differences in detecting weaker storms with different schemes (Camargo & Wing, 2016). Therefore, there is a requirement for a tracking scheme that gives consistent results with different models at different resolutions. Also, applying these two schemes in idealized simulations (e.g., aqua-planet configurations) may improve our understanding of the large-scale climate controls on TC formation, as well as better understanding the influence of detection schemes on these relationships.

Chapter 7. Response of tropical cyclone frequency to sea surface temperatures using aqua-planet simulations

This chapter discusses the objective four: "To examine the relationship of marsupial pouch environments with traditionally detected TCs and environmental variables using idealized GCM simulations run with different values of sea surface temperatures" to be submitted to Climate Dynamics.

Abstract

A comparison is undertaken between the results of two different types of tropical cyclone (TC) detection schemes when applied to aquaplanet global climate model simulations having uniform sea surface temperatures (SSTs). The Commonwealth Scientific and Industrial Research Organization (CSIRO) scheme (a resolution-dependent scheme) detects TCs that resemble observed storms, while the Okubo-Weiss zeta parameter (OWZP) tracking scheme (a resolution-independent scheme) detects the locations within marsupial pouches favorable for TC formation. Here, we use these two schemes to observe relative changes in the TC frequency and their relationships to large-scale environmental conditions with increasing SSTs in a high-resolution atmospheric climate model (ACCESS). Both the schemes indicate a decrease in the global mean TC frequency with increasing temperatures due to increased saturation deficit and stability of the atmosphere. There is an increase in the frequency of intense storms using both the tracking schemes. Also, both the schemes show a reduced frequency with increasing temperatures in the tropics, due to reduced values of moisture, upward mass flux, and higher stability of the atmosphere. The OWZP scheme shows a poleward shift in the genesis locations with rising temperatures due to lower vertical wind shear values. We also observe an overall decrease in the formation of tropical depressions (TDs), both for those that develop into TCs and non-developing cases. The environmental variations between the developing and the non-developing tropical depressions identify the Okubo-Weiss (OW) parameter and Omega as significant influencing variables. Initial vortices with lower vorticity or with weaker upward mass flux do not develop into TCs due to lower moisture and higher stability of the atmosphere. The latitudinal variations in the large-scale environmental conditions account for the latitudinal differences in the TC frequency in the OWZP scheme. Therefore, the OWZP scheme appears to be a more generalized tracking scheme applicable to different model resolutions in different

climates. Also, the marsupial pouch theory acts as a fundamental paradigm for TC formation in different climates.

7.1. Introduction

Tropical cyclones (TCs) cause severe socio-economic losses, mainly in coastal communities. TCs are controlled mainly by surrounding environmental conditions and the large-scale dynamical features of the tropics. Both observational and modeling studies have been used to understand the relevant environmental influences and formation mechanisms. A recent generation of climate models has produced simulations with a reasonable geographical distribution and seasonal variability of TCs but have uncertainties in the simulated intensity due to its sensitivity to the representation of sub-grid scale processes and horizontal resolution (Zhao et al., 2009; Bell et al., 2013; Tory et al., 2013c, d; Strachan et al., 2013; Shaevitz et al., 2014; Walsh et al., 2015; Camargo & Wing, 2016; Walsh et al., 2019; Sharmila et al., 2020; Raavi & Walsh, 2020a). From these studies, it can be noted that there are differences in the processes governing TC formation versus intensification. For climate change predictions, there is an agreement between the theory and modeling of TC intensity, both predicting an increase with increasing temperatures (Emanuel, 1986, 1988). In contrast, although most of the modeling studies show a decrease in TC frequency, the precise mechanism leading to these changes is not known. Also, the statistically derived TC genesis potential indices perform well in the current climate but they do not consistently show a decrease in the frequency with increasing temperatures (Camargo et al., 2014). These indices are derived based on the best statistical fit of TC genesis and climate variables in the current climate and are not necessarily a representation of the future climate. Also, there is no climate theory for TC formation that quantifies the frequency based on the background climate conditions (Walsh et al. 2015). These uncertainties in TC frequency projections require us to understand the quantitative values of prominent environmental variables leading to the changes in the TC frequency.

One method for addressing this issue is to use idealized numerical model experiments that remove some of the co-influences of different variables. Many idealized studies on the environmental conditions leading to the changes in TC characteristics such as formation, size, and intensity due to changing climate conditions are carried out using a high-resolution cloud-resolving model (a horizontal resolution of ~1 km) with a simpler TC permitting framework than would generally be included in a typical climate model.

The radiative-convective equilibrium (RCE) model framework has been employed in such studies, and those models use idealized boundary conditions that have a range of uniform sea surface temperatures (SSTs) either in an f-plane or non-rotating geometry, with domains of various sizes (Knutson and Tuleya, 2004; Nolan, 2007, 2008; Held and Zhao, 2008; Khairoutdinov and Emanuel, 2013; Davis, 2015; Wing et al. 2016; Murthy & Boos, 2018). These studies highlight the use of simplified boundary conditions to understand the role of the planetary vorticity, surface fluxes, radiative feedbacks, and mean thermodynamic environmental controls on different TC characteristics. Studies using the large-domain RCE experiments indicate that TC size increases with increasing temperatures, and with an increase in Coriolis parameter *f*, there is a decrease in size and an increase in the intensity. Some studies noted that TC size is scaled using both the potential intensity to the Coriolis parameter ratio and the radius of deformation (Held and Zhao, 2008; Zhou et al., 2014). Using idealized RCE experiments, a study by Chavas and Emnauel (2014) altered radiative cooling to distinguish the TC size variations with potential intensity scale and radius of deformation. They show consistent results to the TC size with the potential intensity to Coriolis parameter than radius of deformation. This simplification of the bottom boundary conditions with uniform SSTs for a specific range of values has been implemented in general circulation models (GCMs) to study the changes in the TC characteristics and the circulation changes induced by numerous vortices generated in such idealized simulations. These types of simulations are known as aqua-planet simulations (Hayashi & Sumi, 1986; Merlis & Held, 2019). Often, such experiments have no zonal SST asymmetries with the surface covered with water having either a slab ocean or constant SSTs.

Merlis et al. (2013) employ an aqua-planet simulation using a GCM and note that TC frequency increases with increasing radiative forcing with no change in the ocean heat flux, due to a shift of the Inter-Tropical Convergence Zone (ITCZ) in the poleward direction. Ballinger et al. (2015), uses a GCM with zonally symmetric and inter-hemispherically asymmetric surface forcing that involves both fixed SSTs and slab ocean boundary conditions. They have noted that the TC frequency response is sensitive to the meridional SST gradients and is directly proportional to the distance of the ITCZ from the equator. These two studies focus on the ITCZ position and its relationship to the genesis frequency, associated with the changes in the atmospheric circulations driven by the underlying asymmetric thermal forcing. Merlis et al. (2016) have compared the TC frequency response to SST in both GCM simulations with spherical geometry (variation of Coriolis parameter) and f-plane RCE simulations using constant SSTs as lower boundary conditions. They have noted an inhomogeneous distribution of TCs with an increasing trend from low to high latitudes, with the subtropics as the preferred region. Moreover, in both RCE and GCM simulations, there is a decrease in TC frequency and a poleward shift in the genesis with increasing temperatures. A study by Shi and Bretherton (2014) using a GCM with spherical geometry, constant SST and no insolation shows the formation of a multiple TC like vortices poleward of 10 degrees latitude, with durations higher than two months. They have noted a weak Hadley circulation due to the transport of momentum by the poleward movement of these vortices. In addition to the above studies, Reed and Chavas (2015) have used uniform aquaplanet experiments with uniform rotation to show that the f-plane RCE simulations can be reproduced on the sphere by setting f constant everywhere on the sphere in the model.

Chavas and Reed (2019) provides a new theory which bridges the gap between the f-plane RCE simulations and the rotating sphere GCM simulations, and then tests it using uniform aqua-planet experiments with varying planetary rotation rate and size. They have noted that on the f-plane there is only one length scale, which scales with 1/f. On the sphere, there is a second length scale associated with beta called the Rhines scale. At high latitudes, the Rhines scale becomes very large and so only the 1/f scale remains relevant and hence this region is dynamically equivalent to an f-plane supporting lots of TCs. A recent review by Merlis and Held (2019) on the various aqua-planet studies of TCs from single TC permitting models to GCMs highlights the importance of the idealized model configurations to understand the mechanisms that control TC activity due to changes in climate conditions. Walsh et al. (2020) have employed different aqua-planet simulations with both constant and meridionally varying SSTs to study the quantitative influence of large-scale climate on the TC frequency changes. This study note that the vertical static stability parameter has a significant influence on the TC frequency and acts as a likely essential parameter in any climate theory of TC formation.

A recent TC formation theory, the marsupial pouch theory, states that a semi-enclosed recirculating region exists within the trough region of tropical waves called a pouch that protects the vortex from external disruptive influences (Dunkerton et al. 2009). Another study by Wang et al. (2012) hypothesizes that a deep pouch spanning from the lower to the middle troposphere could be sufficient for the TC formation. The above studies using idealized model simulations motivate us to analyze the changes in the frequency of the

TC favorable marsupial pouch formation environments and the associated large-scale environmental influences on their formation. The OWZP TC detection scheme (Tory et al., 2013a, d) detects the locations within marsupial pouches favorable for TC formation (see also chapter 3 and 4). This Okubo-Weiss Zeta Parameter (OWZP) detection scheme uses a diagnostic parameter, the OWZ, and other large-scale environmental variables to identify stronger vorticity regions with weaker deformation forces favorable for TC formation. This OWZP tracking scheme is further developed to detect the initial TDs (Tory et al., 2018) given in chapter 4. Although most of the studies project a decrease in the TC frequency in a future climate, a few studies predict an increase in the formation rate (Emanuel, 2013; Bhatia et al., 2018). Also, there is inconsistency in future projections of whether there is a decrease in the frequency of low-intensity TCs (Emanuel, 2013). Therefore, recent studies are focussing on the changes in the incipient vortices of TC formation known as TC seeds (i.e., weaker vortices) to better understand the differences in the future projections of the frequency of storms. One of two recent studies on the future changes in the TC seeds in a high-resolution coupled ocean-atmosphere general circulation model have noted an increase in the frequency of TC seeds but with a reduction in the frequency of development of these seeds to developed storms (Vecchi et al., 2019). Another study by Sugi et al. (2020) using two different high-resolution atmospheric GCMs have noted a decrease in the future TC seed frequency. A recent study by Hsieh et al. (2020) notes varied response in the frequency of TCs and seeds in the idealized AGCM experiments of different SST gradients. These differences in the AGCMs are explained by the changes in the ventilation index suggesting that the index plays a role in explaining the varied responses of TC frequency projections across the models. These discrepancies in the TC seed frequency changes motivate us to use a phenomenon-based TC tracking scheme to observe the changes in tropical depression (TD) (i.e., low strength vortices) frequency in different idealized simulations. Observing the response of incipient vortices to increased SSTs using idealized simulations aims to improve our understanding of the relationships between climate and initial vortices.

Also, previous aqua-planet studies using GCMs have employed resolution-dependent TC tracking schemes that detect the features resembling the observed TCs using windspeed-vorticity thresholds (Kirtman & Schneider, 2000; Merlis et al., 2013; Shi & Bretherton, 2014; Ballinger et al., 2015; Merlis et al., 2016; Walsh et al., 2020). In contrast to these traditional tracking schemes, TempestExtremes scheme (Ullrich & Zarzycki, 2017) use thresholds of surface pressure and warm core to detect TCs. In the current study, we employ the OWZP scheme to detect locations within marsupial pouches that have the potential for TC formation using thresholds of the large-scale environmental conditions. For this, we analyze aqua-planet simulations using a high-resolution atmospheric climate model with three different values of SSTs. We aim to see whether there is a similar response of TCs detected within pouch locations using a resolution-independent scheme to that of the TCs identified using the resolution-dependent detection schemes in these idealized simulations with increased SSTs. This study helps us to understand whether the marsupial pouch paradigm is fundamental for TC formation in various warmer climates. We also examine whether TCs detected within marsupial pouches using the phenomenon-based scheme have better relationships to large-scale environmental conditions than the TCs detected using the traditional TC tracking scheme. The comparison of the OWZP scheme with the CSIRO TC detection scheme in identifying the TSs in idealized simulations also helps us to identify a more generalized TC detection and tracking scheme that could apply to different climates for different model resolutions.

7.2. Model data and methods

In this study we use the idealized simulations of the ACCESS model as described in section 3.2. These simulations include spatially homogenous SSTs as the lower-boundary conditions with seasonally varying radiative forcing, which creates moderate meridional temperature gradients. The different constant SST values are 25℃, 27.5℃, and 30℃. Also, these simulations have a varying Coriolis parameter with latitude the same as the terrestrial climate. We analyze the idealized simulations for five years from 1989-1993 in each of the constant SST experiments. We use the Commonwealth Scientific, and Industrial Research Organization **(**CSIRO) scheme which detects the circulations that resemble the observed TCs. In addition, a phenomenon-based TC detection and tracking scheme is employed using the Okubo-Weiss zeta parameter **(**OWZP) as described in section 3.5.

We consider different environmental variables that act as influencing variables in a recent study with the same set of aqua-planet simulations (Walsh et al., 2020). We also consider saturation deficit one of the equivalent variable of the ventilation index which is found to be an important predictor for TC formation from the ventilation theory of Tang & Emanuel (2010, 2012) which scales with the changing climatic conditions. These variables include the (i) 700 hPa relative humidity (RH; %), (ii) potential temperature difference between the 925 and 200 hPa level (stability; K/km), (iii) 700 hPa saturation deficit that is the difference between saturation specific humidity and specific humidity, (iv) magnitude of vector wind difference between the 850 and 200 hPa levels (VWS; m/s), (v) 500 hPa Omega (Omega; Pa/s), (vi) 10m wind speed (wind speed), and (vii) 850 hPa Okubo-Weiss (OW; $s^{-2}\times10^{-10}$) parameter (see chapter 3 for definition). We take 12-degree and 4-degree box averages around the storm center of the large-scale environmental variables at the time of the genesis for both developing and non-developing cases. We use the box-averaged values of the variables at the time of genesis for individual cyclones to plot the statistical distributions and perform a binary classification for the TCs in the SST25 and SST30 experiments (i.e. SST25 = constant SST of 25℃ GCM simulation and SST30 = constant SST of 30℃ simulation) to identify the contributing variables to TC formation with increasing SSTs. A two sampled t-test is used to estimate the statistical significance of the distributions.

Here, we use a C4.5 algorithm for the binary classification, which uses a specified metric called “information gain ratio” that divides the TCs of the SST25 and SST30 experiments into either of the two classes (TCs of SST25 or TCs of SST30), and constructs a decision tree like structure with the background environmental conditions as the predictors (Quinlan, 1987, 1993). This decision tree gives us the different combination of environmental conditions leading to the TC formation in the SST25 and SST30 experiments. The decision tree algorithm uses the information gain ratio criteria which is a probabilistic measure that describes the quality of split after it divides the dataset into a “TCs (SST25) or TCs (SST30)” class using each of the large-scale environmental variables (refer section C.1 of appendix C for more details). Here the environmental variable that has maximum information gain ratio partitions the data at each node of the tree. The construction of the decision tree involves selecting an environmental variable with maximum information gain ratio criterion as the "root node." The subsequent splitting of the decision tree stops when it reaches the minimum leaf size threshold. The minimum leaf size represents the size of the decision tree; a smaller tree better generalizes the data. Further details of the decision trees using the C4.5 algorithm are described in section 5.2.5 of Chapter 5.

7.3. Results

7.3.1 Mean annual frequency

Figure 7.1 shows that the OWZP scheme detects some storms with lower 10m wind speeds (i.e., lower than 10m/s) compared to the CSIRO scheme. The resolution-dependent 10m wind speed criteria in the CSIRO scheme eliminate the TC detections below this specified threshold of wind speed. In contrast, the OWZP scheme uses the thresholds of large-scale environmental conditions and does not use any thresholds of the wind speeds. With a suspicion of those OWZP detected storms having wind speeds less than 10m/s are not tropical storms, so we remove these storms for further analysis. In addition, Figure 7.1 shows an increase in the number of intense storms with increasing SSTs using both OWZP and CSIRO schemes. We further estimate the mean annual number of OWZP detected storms for each of the constant SST experiments (Table 7.1).

Table 7. 1. The global mean annual number of TCs and TDs for each of the aqua-planet simulations using both the OWZP and CSIRO detection schemes

Experiment	OWZP Scheme (TC)	CSIRO Scheme (TC)	OWZP (greater than 10 m/s)	OWZP Scheme (TDs)	Non-developing cases (ND)
SST25	3218	2277	2583	5007	1789
SST27.5	2929	1808	2286	4449	1520
SST30	2923	1586	2197	4463	1540

Here, we consider the OWZP storms with wind speeds greater than 10m/s for the further analysis. The SST25 experiment is considered as the control experiment, for which we found a global mean number of 2583 storms. By increasing the temperatures, for the SST27.5 and SST30 experiments, we observe a decrease in the global mean annual detections with 2286 and 2197, respectively. Similarly, using the CSIRO scheme, we found more detections, 2277 TCs for the SST25 experiment and decreasing numbers for the SST27.5 and SST30 experiments, with 1808 and 1586 respectively. The increased detections in all the three constant SST experiments using the OWZP scheme compared to the CSIRO scheme is similar to that seen in the current climate simulation using the ACCESS model, which gives higher detections using the OWZP scheme (chapter 6; Raavi & Walsh, 2020a). Using the OWZP scheme, we observe a 12% decrease in the frequency when SST increases from 25℃ to 27.5℃ and a 15 % reduction when the SST increases from 25℃ to 30℃. In contrast, the CSIRO scheme has a substantial decrease of 21% and 30% by increasing the SSTs from 25℃ to 27.5℃ and from 25℃ to 30℃, respectively. Overall, we observe a reduction in the genesis frequency by increasing SSTs using both the tracking schemes.

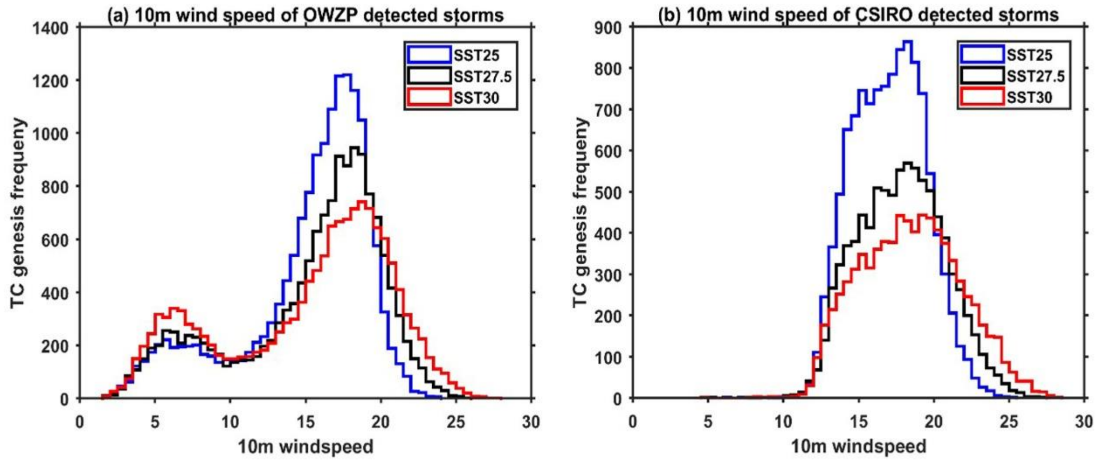

Figure 7. 1 The distribution of maximum 10m wind speed attained during the lifetime of each of the storms detected by both (a) OWZP and (b) CSIRO tracking schemes.

7.3.2 Geographical distribution of TC genesis Frequency

Figure 7.2 shows the geographical distribution of area-weighted TC genesis density across the globe using both CSIRO and OWZP schemes in the three different SST experiments. The aqua-planet geographical distribution of TC genesis differs from the current climate simulation (Figure 6.1 of chapter 6). The differences in geographical distribution are due to the absence of land with favorable SSTs everywhere and weaker vertical wind shear regions due to a lack of stronger meridional temperature gradients. In the OWZP detection scheme, we observe fewer detections between 0 to 15° latitudes with increased detections poleward of 15° in all the three experiments. In contrast, the CSIRO detection scheme has fewer detections between and 0 to 25° latitudes and increased detections poleward of 25°. The higher detections in the mid-latitudes using both the schemes are perhaps due to the absence of baroclinicity (no meridional temperature gradients) and the consequent lack of strong vertical wind shear (Walsh et al. 2020), in contrast to current climate simulations. Also, the symmetric SST specification in both the hemispheres produced a similar number of detections in both hemispheres, in contrast to the current Earth climate.

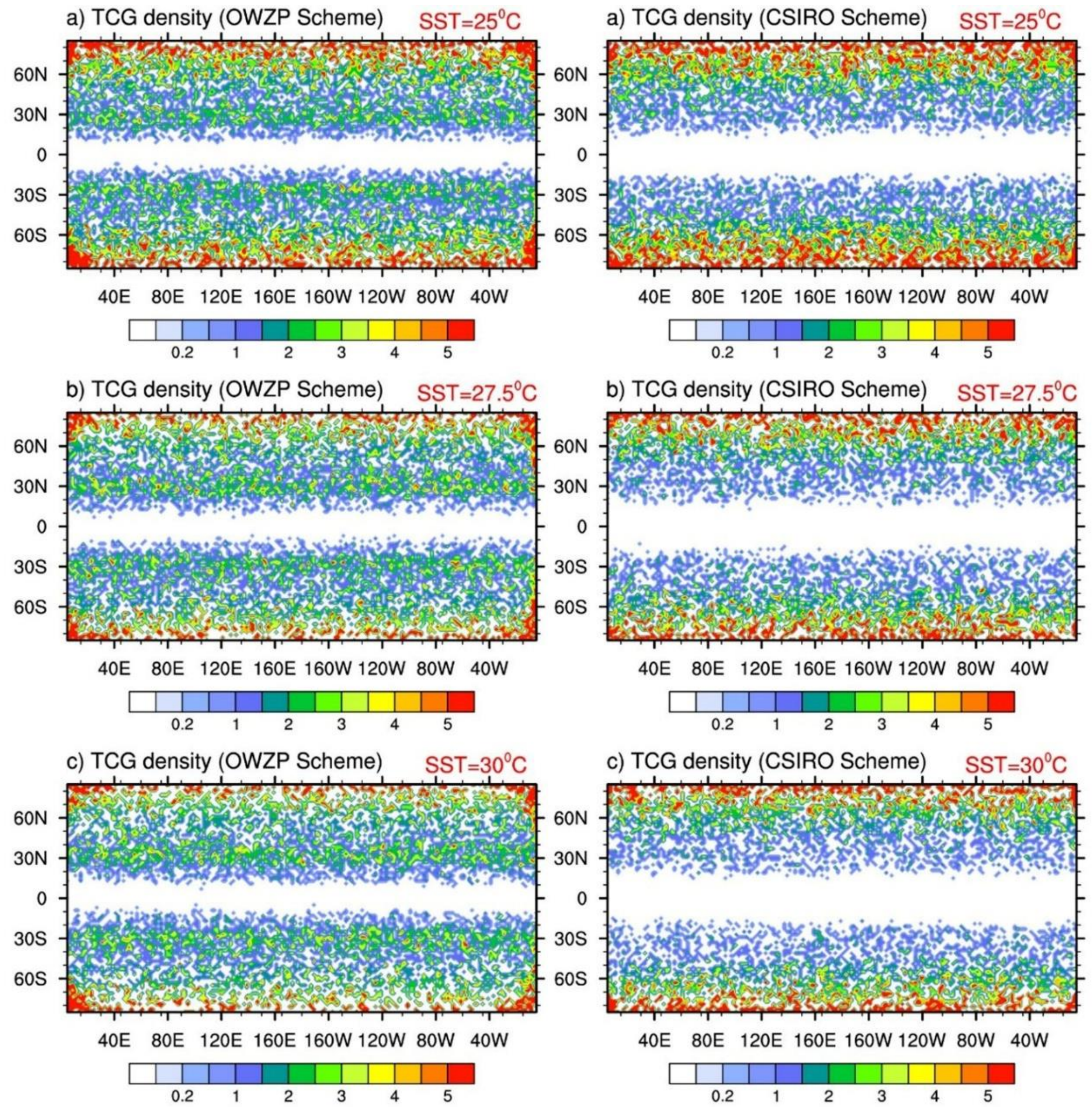

Figure 7. 2. Area-weighted (cosine of latitudes; gradually increasing weightage while moving from equator to poles) spatial distribution of TC genesis (TCG) density given as number of TCs per 2° × 2° region per year for the three constant SST experiments using both OWZP and CSIRO schemes.

We observe individual differences between the two schemes in specific latitude bands from the spatial distribution of TC formation. Figure 7.3 shows the latitudinal distribution of the TC genesis frequency using both the schemes, indicating differences between them in the latitudes 10 to 45°. There is an increased frequency between 10 to 30° latitudes and a decrease in the 30 to 45° latitude band using the OWZP scheme, whereas using the CSIRO scheme, detections gradually increase poleward of 20° in all the three experiments. The difference between the CSIRO and the OWZP scheme is mainly due to a peak at latitude 30° in the OWZP scheme. While observing the 10m wind speeds for different latitude bands, we notice that the OWZP detected storms in the 0-30° latitude band have a lower 10m wind speed magnitude compared to the CSIRO storms (Figure 7.4), which might be due to the wind speed threshold in the CSIRO scheme removing detections in that scheme. Figure 7.5 shows the mean annual TC frequency across different latitude bands for all the three experiments using both the schemes. The mean number of detections is largest in the latitudes poleward of 30° using the CSIRO scheme, whereas the OWZP scheme has higher TC detections in the 0-30° and 30-45° latitude bands.

These results using OWZP scheme agree with a recent study Chavas & Reed (2019) with a gradually increasing genesis to a maximum near 45° latitude and a decreasing genesis towards poles which is filled with long-living storms. The tracking scheme used in the study of Chavas & Reed (2019) and the OWZP scheme doesn't use any threshold for wind speed and therefore perform in a similar manner. In contrast, the wind speed threshold in the CSIRO scheme leads to the difference in results with increasing detections when moving towards the poles. Also, the CSIRO scheme shows a decrease in TC frequency with increasing SSTs across all the latitudes, whereas the OWZP scheme indicates a decrease in TC frequency in 0-30°, 45-60° and 60-90° latitude bands but an increase in the 30-45° latitude band showing a poleward shift in the genesis with increasing temperatures. These differences in the detections between latitude bands with OWZP and CSIRO schemes are perhaps due to fundamental differences in the tracking schemes.

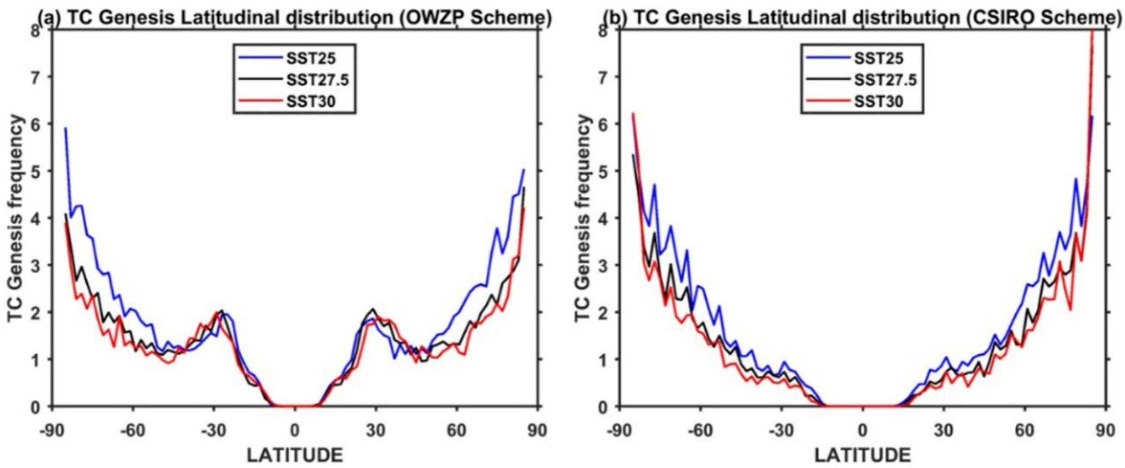

Figure 7. 3. Latitudinal distribution of genesis frequency for the three constant SST experiments using both the (a) OWZP and (b) CSIRO detection schemes.

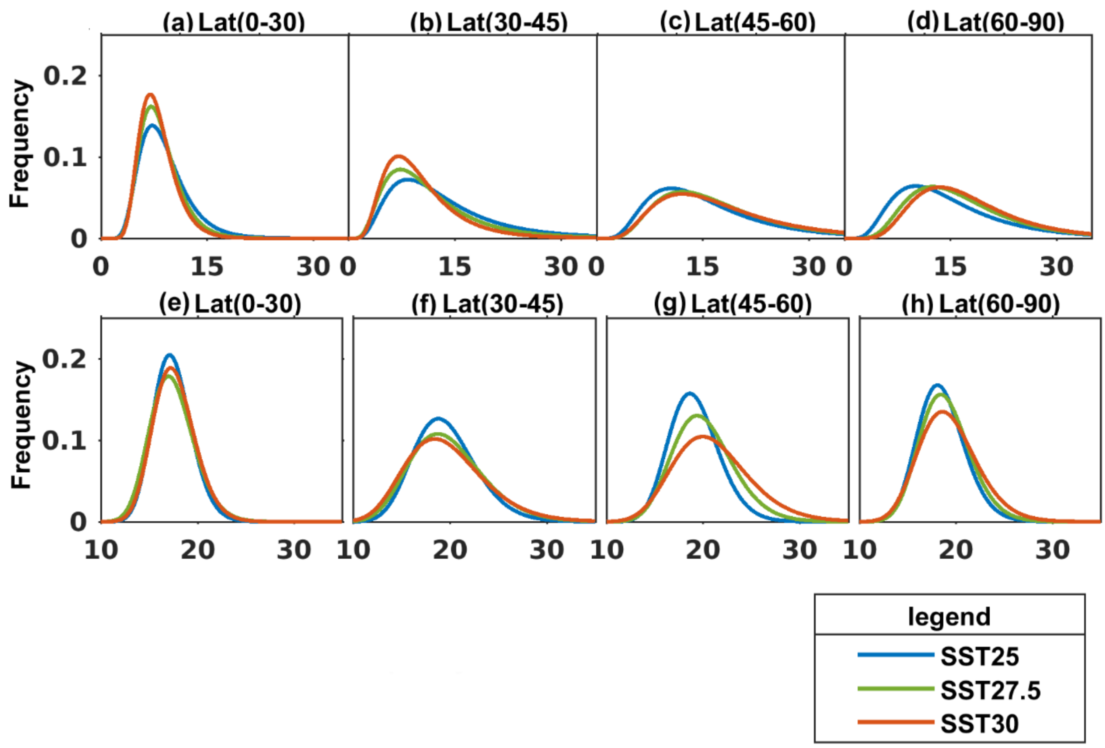

Figure 7. 4. 10m wind speeds (ms^{-1}) of OWZP detected storms (a to d) and CSIRO detected storms (e to h) in each of the latitude bands across the globe

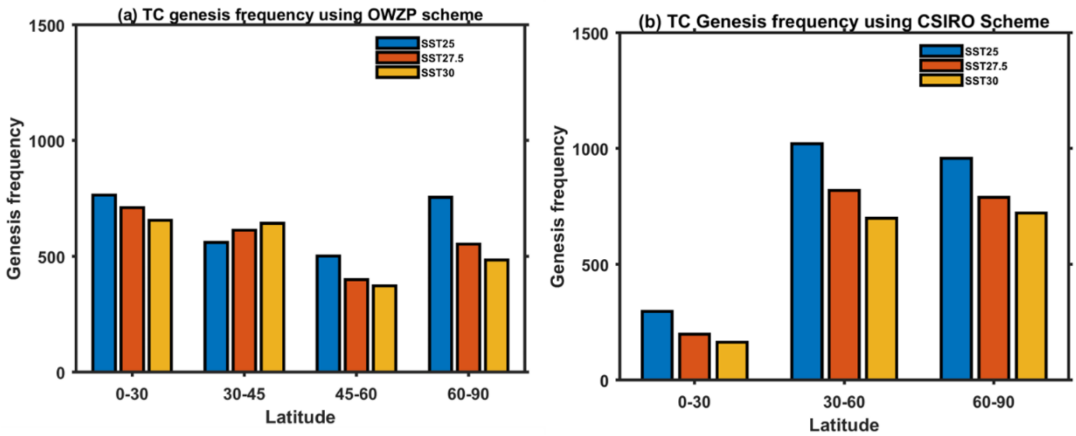

Figure 7. 5. Mean annual frequency of detected storms using both the (a) OWZP and (b) CSIRO detection schemes for each latitude band

7.3.3 TC lifetime and tracks

Figure 7.6 shows that both the OWZP and CSIRO schemes have an almost similar maximum lifetime of storms (30-35 days) in the SST25 simulation, and similar results are seen in the other two simulations (not shown). The OWZP scheme has a higher number of more prolonged duration storms, whereas the CSIRO scheme has shorter duration storms. The lifetime of TCs in using the same model forced with interannually varying AMIPII SSTs (Kanamitsu et al., 2002) for the period 1990-2009 shows that the storms detected by the CSIRO scheme have a shorter duration than storms detected by the OWZP (chapter 6; Raavi & Walsh, 2020a). Figure 7.7 shows the lifetimes of OWZP storms in all three simulations for different latitude bands. The storms in the 0-30° and 30-45° latitude bands have a longer mean duration in the SST25 simulation compared to SST30. This may perhaps be due to their ability to survive for a longer time in a region that has favorable environmental conditions such as stronger vorticity, and higher moisture promoting sustained deep convection (Figure 7.9 & 7.10). In contrast, for higher latitudes, TCs in the SST30 simulation have a slightly longer duration compared to the SST25 experiment. A recent study by Raavi & Walsh (2020a) given in chapter 6 noted that the detected TC tracks from the CSIRO scheme are a subset of the detected tracks from the OWZP scheme.

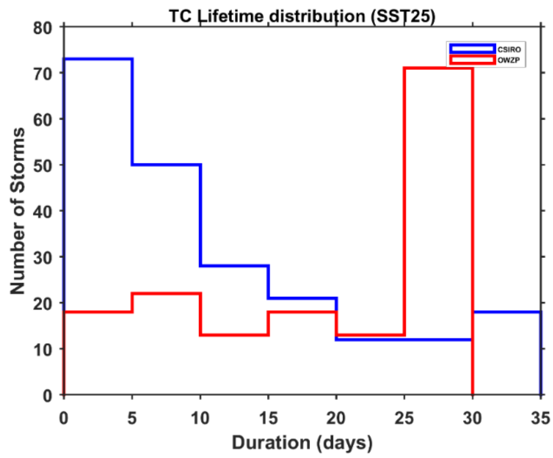

Figure 7. 6. The distribution of TC lifetime (in days) for the detected storms using the OWZP and CSIRO schemes for August 1990 in the SST-25 experiment.

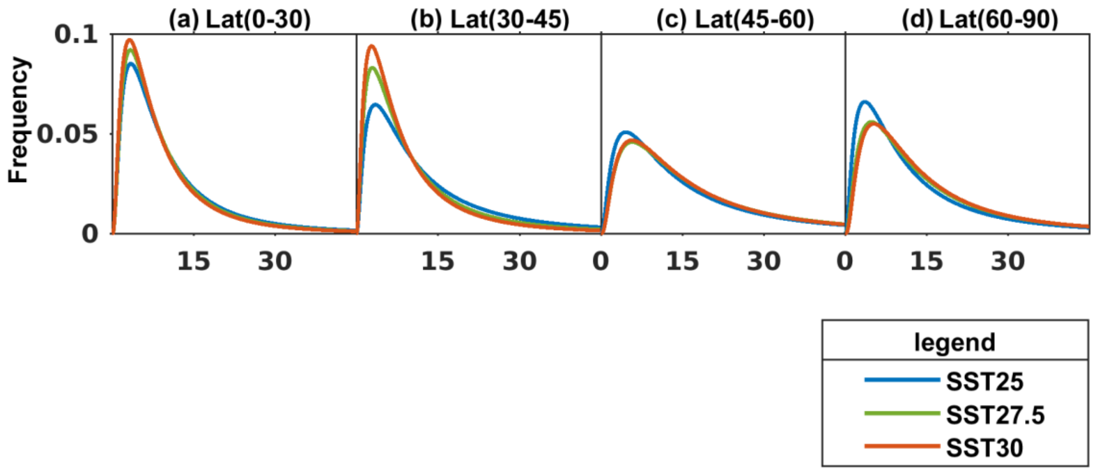

Figure 7. 7. Lifetime of OWZP detected storms (in days) for different latitude bands for all three constant SST experiments

To identify whether the CSIRO tracks in the aqua-planet simulations are a subset of the OWZP tracks, we have considered a particular month (August) in the year 1990 for the SST25 experiment and plotted individual TC tracks detected by both the schemes. Figure 7.8 shows that the CSIRO storm tracks for the SST25 simulation are a subset of the OWZP tracks. Similar results are observed in the other two constant SST experiments (not shown). The TC tracks in these simulations show that genesis occurs mostly in the subtropics, and the TCs move poleward and westward using both the tracking schemes. This poleward and westward movement of the TCs is mainly due to the beta-drift mechanism due to the presence of the latitudinally varying Coriolis force in these simulations (Holland, 1983, 1986; Li and Wang, 1994; Wang and Holland, 1996).

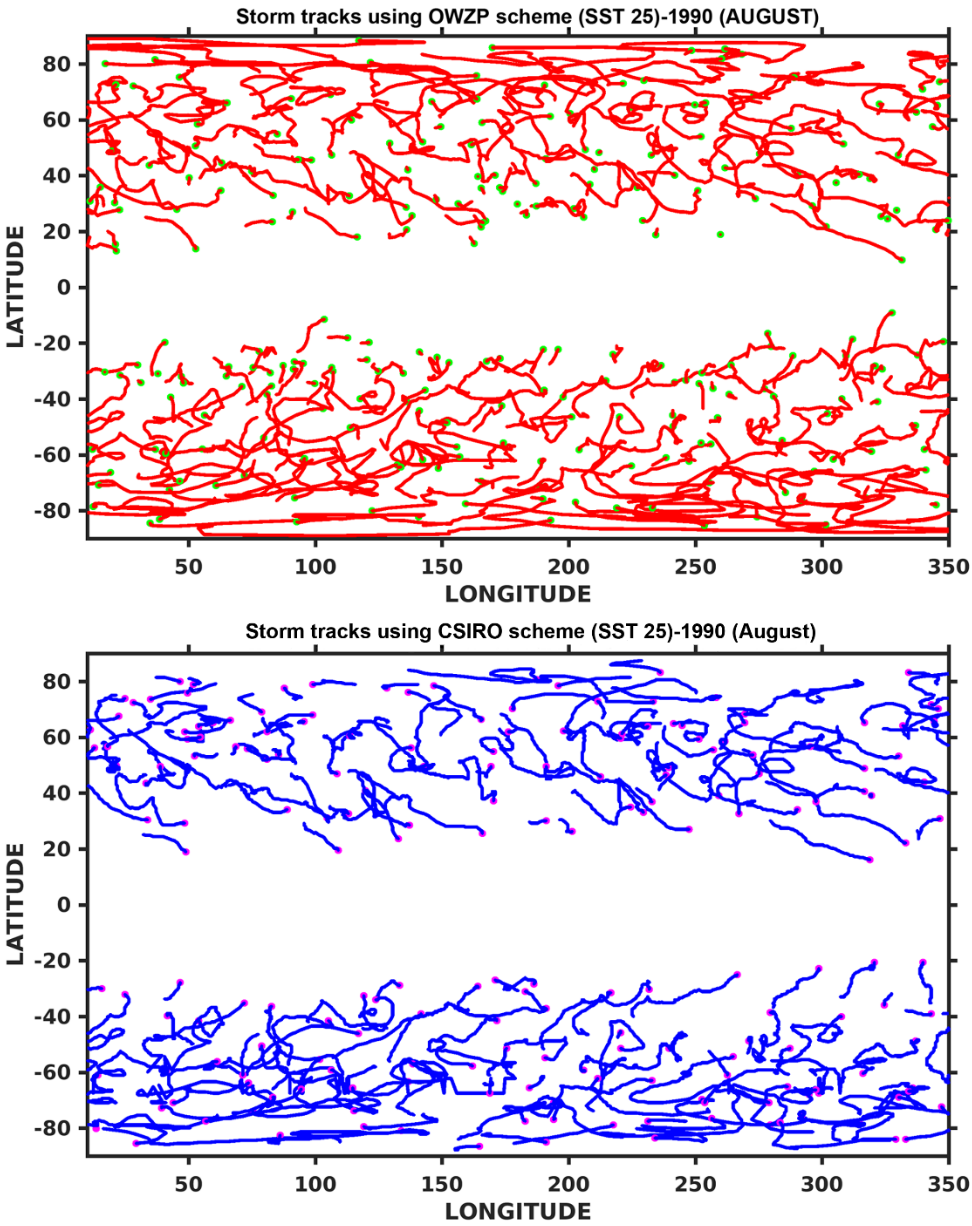

Figure 7. 8. Tracks of the 280 storms detected using the OWZP scheme (red lines; genesis locations in green dots) and 214 storms detected by the CSIRO scheme (blue lines; genesis locations in magenta dots) for August 1990 of the SST 25 experiment.

7.3.4 Large-scale environmental conditions across different SST experiments

To understand the differences in the genesis frequency between different latitude bands, we have calculated the large-scale environmental conditions at the time of the genesis for the OWZP storms for each of the latitude bands. Figure 7.9 and Figure 7.10 show the distribution of the environmental conditions for the storms in all the three SST experiments. Of all the variables, we observe that the saturation deficit and stability parameters show the most significant differences (greater than 95% significance level) between the three simulations. With an increase in SST, we observe an increase in the stability of the atmosphere (Figure 7.9). The stability values also show an increasing trend with latitude, with the polar regions having higher values than the other latitude bands in all the three experiments. This rising trend of stability with latitudes matches the increasing precipitation towards poles in these sets of simulations, as shown in Figure 3 of Walsh et al. (2020). We speculate that the higher values of the precipitation lead to the release of the latent heat flux during the condensation process and increases the temperatures of the upper atmosphere, thereby stabilizing the atmospheric column. This increased stability with increased upper tropospheric temperature leads to a decrease in the frequency of storms across the globe. Similarly, the saturation deficit increases with increasing temperatures across all the latitude bands. In contrast, Figure 7.9 shows prominent differences in RH across the 0-30° and 30-45° latitude bands, with the SST25 experiment having higher values compared to the SST27.5 and SST30 experiments. Therefore, with increasing temperatures, there is a global increase in the stability and saturation deficit of the atmosphere, both more unfavorable for TC formation.

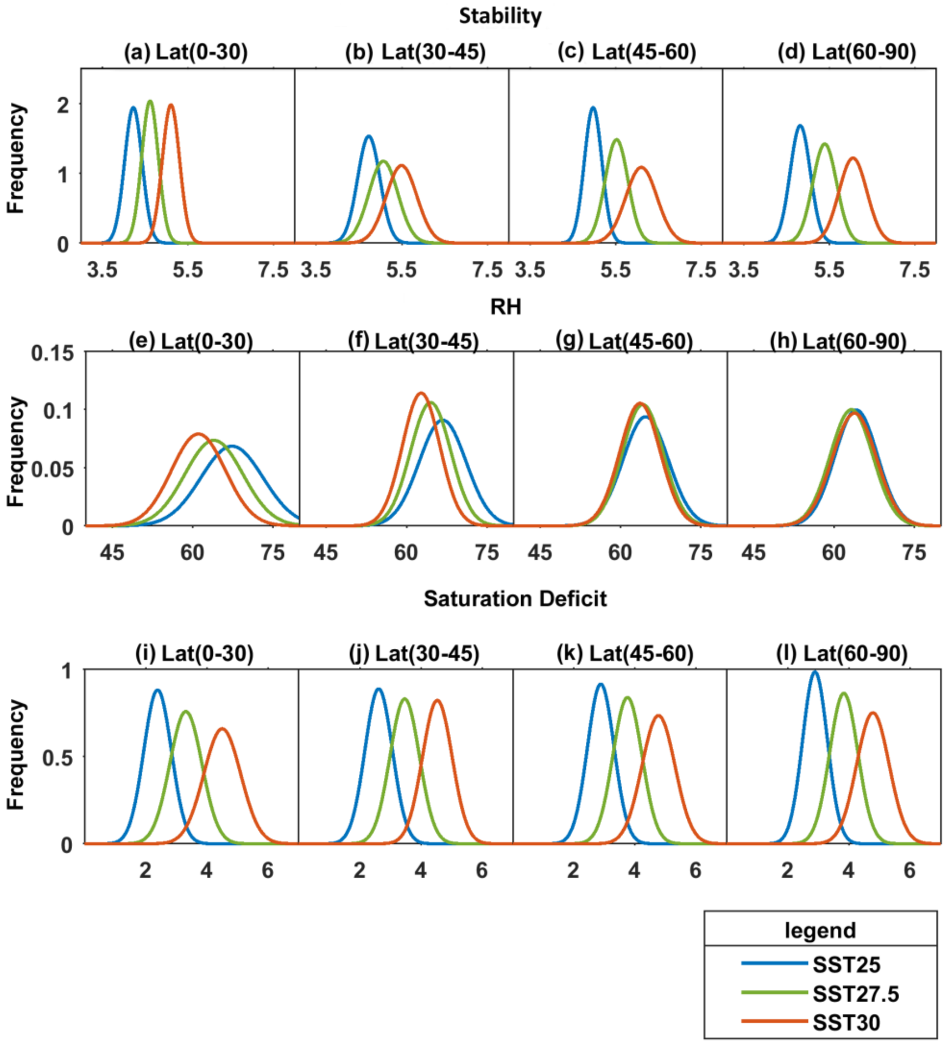

Figure 7. 9. Statistical distributions of stability parameter (a to d; K/Km), RH (e to h; %), and saturation deficit (i to l; kg/kg) for OWZP detected developing storms for different latitude bands in all three constant SST experiments

The distribution of Omega (Figure 7.10) shows slight differences between different SST experiments. For the latitude bands, 0-30° and 30-45°, the storms in the SST25 experiment have more negative values of Omega (thus more favorable for TC formation) compared to the other two experiments. In contrast, for the latitude bands 45-60° and 60-90°, the storms in the SST27.5 and SST30 experiments have more negative values of Omega compared to the storms in the SST25 experiment. The 850 hPa OW parameter (Figure 7.10) also shows similar distribution to Omega but has variations in the magnitudes at different latitude bands. The OW parameter shows a gradually increasing trend in the magnitude with increasing latitudes. The TCs in the low latitudes have lower values of the OW parameter compared to the storms in higher latitudes. The stability parameter also shows an increasing trend in the values with increasing latitudes. The higher values of the OW at the higher latitudes (poleward of 45 degrees latitude) have enhanced stability of the atmosphere compared to the lower latitudes, indicating that as stability increases, the vortices need to have stronger vorticity for genesis to occur. This also means that the vortices having weaker strength indicated by weaker values of the low-deformation vorticity do not develop into tropical storms and may become non-developing cases. Also, there may be a reduction in the formation of the weaker strength vortices in the higher stable atmospheric conditions.

The other variables considered are the VWS and wind speed, which have similar distributions (Figure 7.10): for the four latitude bands, we notice a difference in the latitude band 30-45°, with SST30 and SST27.5 having lower values of shear compared to SST25. These lower values of vertical wind shear lead to increased detections in SST27.5 and SST30 experiments compared to the SST25 (Figure 7.5). Also, 0-30° latitudes, the values of the VWS are lower compared to the other latitude bands, therefore leading to higher detections in the lower latitude bands compared to the other bands (Figure 7.5). Similar to the study of Walsh et al. (2020), we observe that the VWS increases with a decrease in the SSTs in lower latitudes under constant SST forcing. This is due to the presence of the meridional temperature gradients aloft generated by retaining Earth-like radiative forcing.

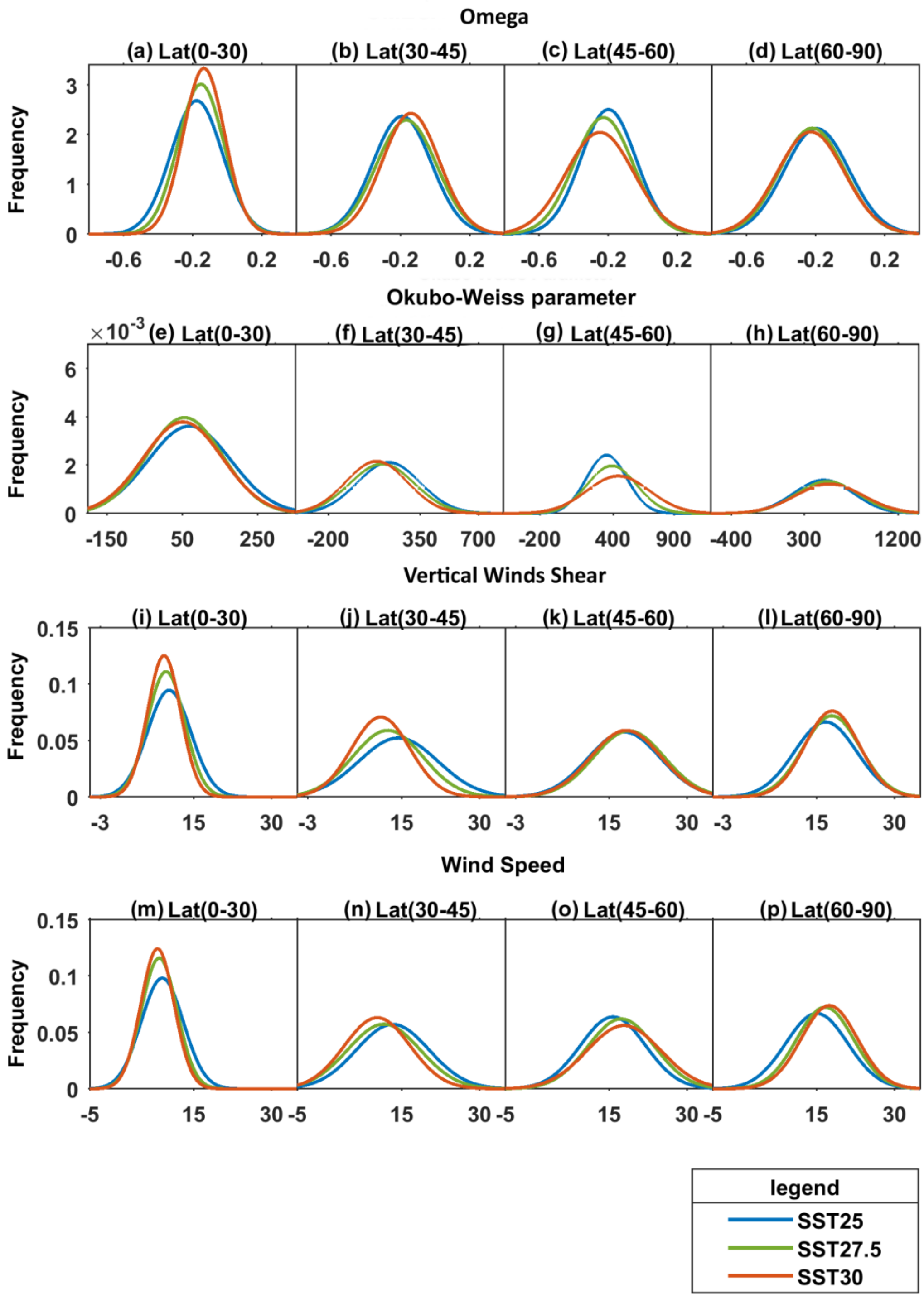

Figure 7. 10. Statistical distributions of Omega (a to d; Pa/s), OW parameter (e to h; $s^{-2}\times10^{-10}$), Vertical wind shear (i to l; m/s), Wind speed (m to p; m/s) for OWZP detected storms across different latitude bands in all the three constant SST experiments

Table 7. 2. Decision tree rules obtained after applying C4.5 classification algorithm to the SST25 and SST30 experiments for different latitude bands (OW parameter correlates with the Omega)

SST25 and SST30 (0-30°)
1. if RH > 65.43 then TC forms in SST25 (1075/210) 2. if RH <=65.43 and VWS <=13.25 then TC forms in SST 30(609/232)
SST25 and SST30 (30-45°)
1. if RH > 67 then TC forms in SST 25 (1216/236) 2. RH<=67 and OW <=31 then TC forms in SST30(398/77) 3. if RH <=67 and OW>502 and then TC forms in SST30 (127/26) 4. RH <=67 and OW<=502 and VWS <20 then TC forms in SST30(361/170)
SST25 and SST30 (45-60°)
1. if OW > 678 then TC forms in SST30 (170/19) 2. if OW < 678, RH > 67 then TC forms in SST25(517/143) 3. if RH <=67 and OW<=554 then TC forms in SST25(1395/575) 4. if RH <=67 and 554<OW<678 then TC forms in SST30(176/63)
SST25 and SST30 (60-90°)
1. if WIND > 18 then TC forms in SST30 (173/32) 2. if WIND <=18 then TC forms in SST25(2622/974) 3. if WIND <=18 and VWS <= 20 then TC forms in SST30(117/20) 4. if WIND <=18 and VWS>20 and RH >64 then TC forms in SST25 (327/148)

7.3.5 Classification of the TCs between SST25 and SST30 experiments

Although the statistical distributions indicate that as the saturation deficit and stability increase, the TC frequency decreases, and it is still unclear what other environmental conditions lead to TC formation in different latitude bands with reduced moisture and increased stability of the atmosphere. Therefore, we have performed a binary classification analysis between the SST25 and SST30 experiments using a C4.5 algorithm to find out the combination of the environmental conditions favoring TC formation. This algorithm identifies the prominent parameters needed to classify the

SST25 and SST30 TCs given different background environmental parameters as the predictors. Table 7.2 shows the different decision rules obtained from the classification algorithm. Over the latitude band 0-30°, higher values of RH lead to TC formation in the SST25 experiment. Although SST30 has lower amounts of relative humidity compared to SST25, the lower values of vertical wind shear lead to TC formation in the SST30 experiment at these latitudes. In the latitude band 30-45°, the higher values of the relative humidity lead to TC formation in the SST25, and higher values of the OW parameter and lower values of vertical wind shear lead to TC formation in SST30 experiment. In the latitude bands 45-60° and 60-90°, the higher values of the OW parameter, upward mass flux (more negative values of Omega) and wind speed lead to TC formation in the SST30 even though there are less favorable thermodynamical conditions such as higher stability and saturation deficit of the atmosphere. This analysis also indicates that under highly stable atmospheric conditions, incipient vortices with higher low-deformation vorticity and with higher values of upward mass flux lead to TC formation in the higher latitudes. In the lower latitudes the atmosphere is less stable compared to the higher latitudes, the lower values of vertical wind shear favor TC formation in the SST30 experiment.

7.3.6 Developing versus non-developing tropical depressions

To understand whether there is an overall decrease in the formation activity or an increase in the non-developing cases compared to the developing cases with increasing global temperatures, we use the OWZP scheme to detect both developing and non-developing tropical depressions (Tory et al. 2018). Table 7.1 shows the global annual frequency of the total tropical depressions, developing, and non-developing depressions. We notice that there is a decrease in the global mean annual frequency of tropical depressions and non-developing tropical depressions with increasing SST, similar to the trend for developing cases. This reduction in the initial tropical depressions may imply that there will be a reduction in the number of weaker storms in the future climate. Although the intensity of the storms (as measured by their 10m wind speeds) is weaker in the ACCESS model compared to the observations, there is an increase in the frequency of intense storms with increasing temperatures using both OWZP and CSIRO storms (Figure 7.1). Using two high-resolution atmospheric models, Sugi et al. (2020) also noted that the frequency of TC seeds (i.e., weaker pre-storm vortices) decreases with increasing temperatures, with the seeds detected based on the minimum sea-level pressure and a warm core criterion (the geopotential height difference between 200-500 hPa). Figure 7.11 shows that in the latitude bands 0-30° and 30-45°, there is an increase in the mean annual non-developing cases with increasing temperatures. In contrast, at latitudes above 45°, we observe a decrease in the non-developing cases with an increase in the SSTs. To further understand the reasons for the changes in the TC frequency for both developing and non-developing cases in all three experiments across different latitude bands, we evaluate the large-scale variables changes due to different underlying SSTs and seasonally varying radiative forcing which may change the atmospheric circulations. In the next section, we discuss the changes in the environmental conditions in different latitude bands.

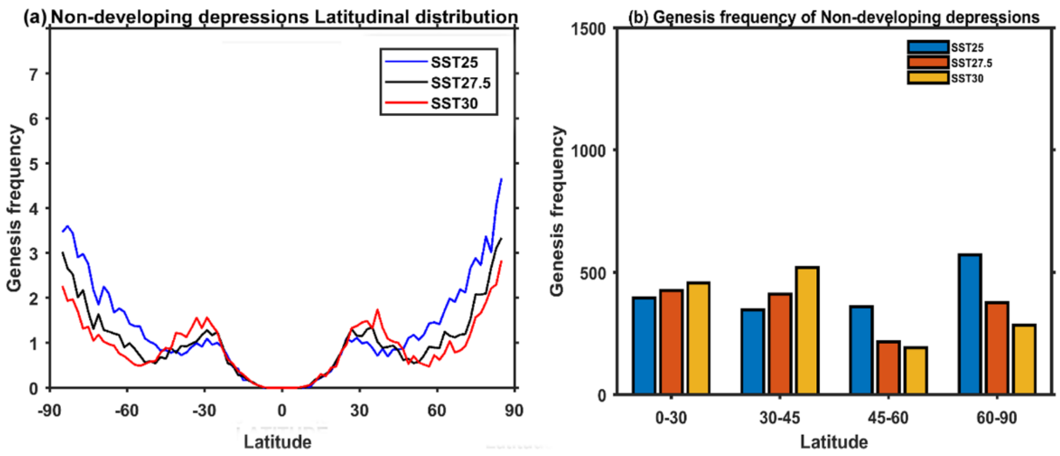

Figure 7. 11. Latitudinal distribution of non-developing depressions genesis frequency per year detected using the OWZP scheme for the three constant SST experiments.

7.3.7 Large-scale environmental conditions for both developing and non-developing circulations

Figure 7.12 shows the variations in the stability of the atmosphere for three different experiments for both developing and non-developing cases. A similar pattern exists for both developing and non-developing cases: increasing stability with increasing SSTs. The increasing stability leads to a net decrease in the formation of the tropical depressions for both developing and non-developing cases. The saturation deficit parameter also shows similar differences between developing and non-developing cases in that there is an increase in the saturation deficit with increasing temperatures. The other thermodynamical variable is RH (Figure 7.12), which shows that there is a decrease in the RH with increasing temperatures in both developing and non-developing cases. The non-developing cases have lower moisture compared to the developing cases in all the three constant SST experiments. The increasing number of non-developing cases with increasing temperatures in the lower latitudes 0-30° and 30-45° as shown in Figure 7.11 is perhaps due to reduced moisture and increased stability of the atmosphere with rising temperatures.

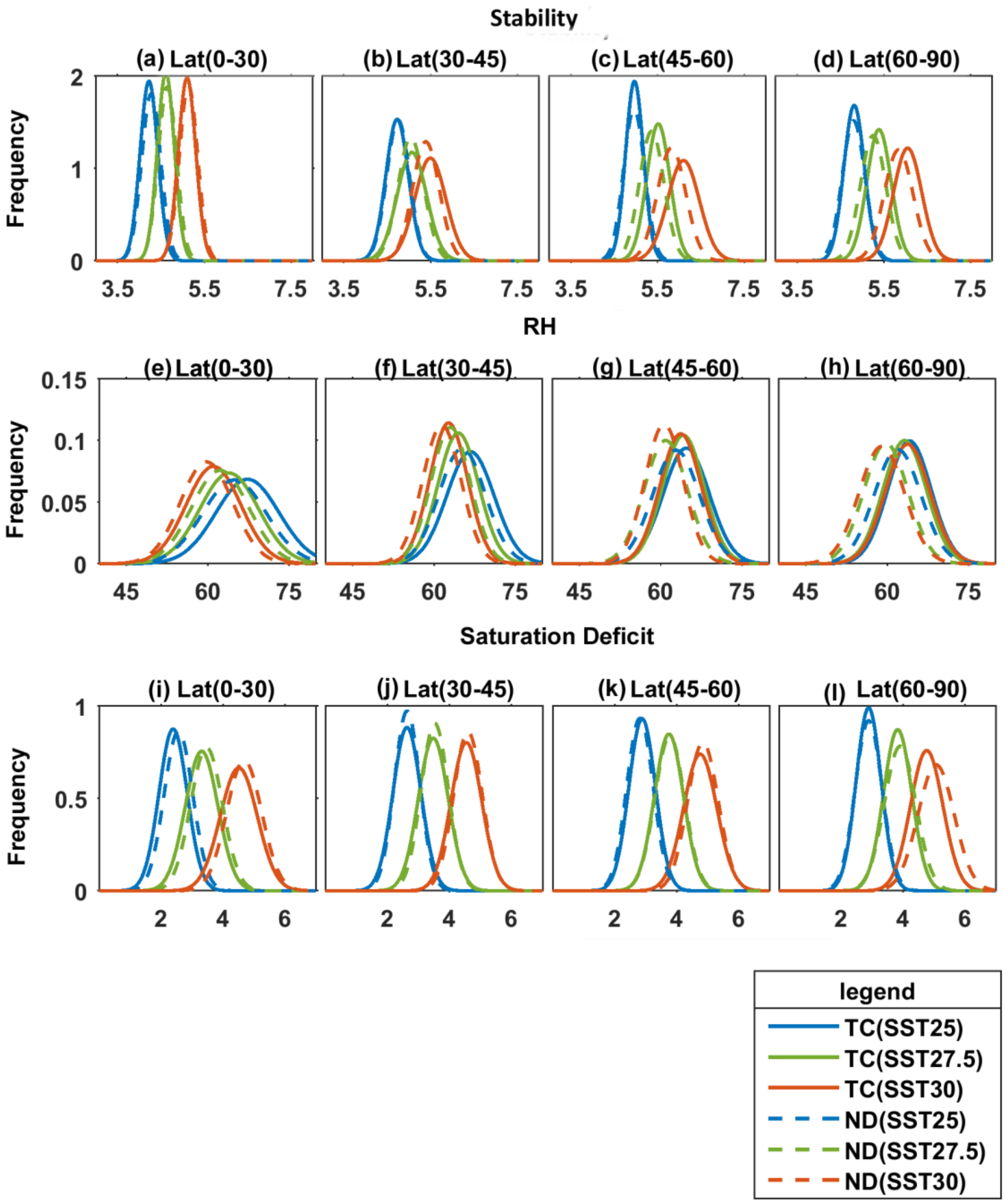

Figure 7. 12. Statistical distributions of Stability parameter (a to d; K/km), RH (e to h; %), and Saturation deficit (i to l; kg/kg) for both developing (TC) and non-developing tropical depressions (ND) across different latitude bands in all the three constant SST experiments

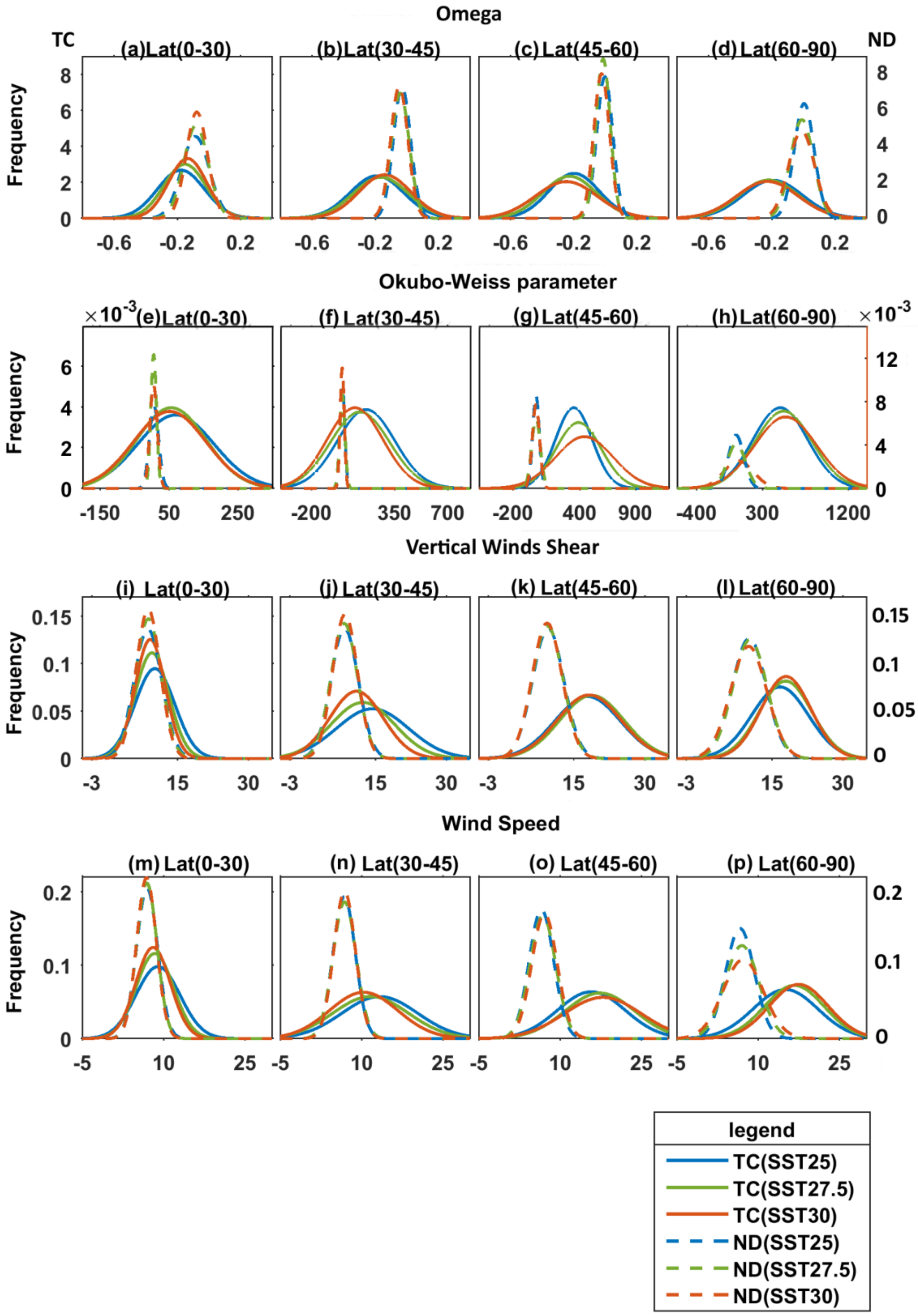

Figure 7. 13. Statistical distributions of Omega (a to d; Pa/s), OW parameter (e to h; $s^{-2}\times10^{-10}$), Vertical wind shear (i to l; m/s), Wind speed (m to p; m/s) across for both developing and non-developing tropical depressions across different latitude bands in all the three constant SST experiments.

The increased values of Omega (i.e., reduced upward mass flux) lead to decreased TC frequency in future projections using GCMs due to weaker atmospheric vertical circulation caused by the decreasing lapse rate with increasing temperatures (Sugi et al. 2002, 2012). These studies noted that the increase in the precipitation is lower than the increase in the atmospheric moisture with increasing SST, thereby indicating a decrease in the upward mass flux. Here, developing cases have reduced values of Omega (Figure 7.13), favoring formation compared to the non-developing cases. Also, the variation of Omega across each of the latitude bands shows that developing cases have favorable upward mass flux compared to the non-developing cases. The statistical distributions of OW and 10m wind speed variables (Figure 7.13) have a similar distribution as Omega, with developing cases having more favorable values compared to the non-developing cases. These results suggest that initial vortices with stronger low-deformation vorticity and higher upward mass flux and 10m wind speeds lead to the formation of an intense TC.

The statistical distributions of the vertical wind shear (Figure 7.13) show that non-developing cases have lower values than the developing cases, perhaps due to the influence of the existing developed circulation. The increased low-level winds with slight upper-level winds in a developing circulation lead to increased VWS compared to the non-developing circulation that has weaker low-level winds. This analysis of the environmental differences between developing and non-developing tropical depressions shows that the RH, OW, and Omega explain the differences in the developing and non-developing cases under the increased saturation deficit and stability of the atmosphere with increasing SSTs. These results agree with studies of both Sugi et al. (2005) and Emanuel (2013), showing that the decrease in TC frequency is due to the decrease of the upward mass flux and increased saturation deficit of the atmosphere. Therefore, unfavorable environmental conditions such as lower moisture content, lower values of upward mass flux, weaker circulation (lower OW values) lead to the non-development of the initial circulation.

7.4. Discussion

To understand the environmental conditions and the associated mechanisms leading to the changes in TC genesis frequency within marsupial pouches in warmer climates, we analyze idealized climate model simulations with different values of globally uniform SSTs in a high-resolution atmospheric model (ACCESS). Here we compare the TC frequency changes within the model obtained from two fundamentally different TC tracking schemes. Both schemes show a decrease in the global TC genesis frequency (Table 7.1) and an increase in the frequency of the intense storms with increasing SSTs (Figure 7.1). Also, the area-weighted global TC genesis frequency (Figure 7.2 & 7.3) gradually increases from the subtropics to poles using both the tracking schemes, with some differences in specific latitudes. Merlis et al. (2016) noted an increase in the TC frequency towards the poles in a constant SST GCM simulation.

When comparing the overall TC detections in different latitude bands, the OWZP scheme has a higher number of TCs in the tropics (Figure 7.5) compared to the CSIRO scheme, due to lower values of the wind speed and VWS (Figure7. 4 & 7.10). The reduced number of TCs using the CSIRO scheme across this 0-30° latitude band is due to the higher 10m wind speed threshold (Figure 7.4), whose value is set to be consistent with the resolution of the model. There are a higher number of detections in mid-latitudes and poles compared to the terrestrial climate simulations using both the detection schemes, perhaps due to the more favorable conditions for the TC formation within the aquaplanet simulations. On examining the changes in the genesis frequency between different SST experiments, both the OWZP and the CSIRO show a decrease in genesis at 0-30° latitudes with increasing SSTs. Also, the OWZP scheme detects an increase in genesis at 30-45° latitude, showing a shift in genesis towards the poles with increasing temperatures (Figure 7.5). These results agree with the earlier study by Merlis et al. (2016) using uniform SSTs, who showed a decrease in frequency and a shift in the genesis locations towards poles with increasing SSTs. At higher latitude bands (45-60° & 60-90°), the tracking schemes show a decrease in the frequency with increasing temperatures (Figure 7.5).

The OWZP and CSIRO scheme detected TCs have a similar maximum lifetime with a higher number of more prolonged duration storms in the OWZP scheme (Figure 7.6). The CSIRO TC tracks are a subset of the OWZP tracks for all the three constant SST aqua-planet simulations, similar to the results of the current climate simulation (chapter 6; Raavi & Walsh, 2020a). The TC tracks move westward and poleward due to the beta-drift mechanism (Figure 7.8). The TCs in the higher latitudes are noted to have longer lifetimes than the lower latitudes in the SST27.5 and SST30 experiments, perhaps due to stronger vorticity and higher upward mass flux, enabling the TCs to survive for a longer period.

We use storm-centered instantaneous values of large-scale environmental variables at the time of the genesis of individual cyclones to understand the changes in the TC genesis frequency across different latitudes. Of all the variables, we observe that the saturation deficit and stability parameters have the most significant differences between different SST simulations (Figure 7.9). There is an increase in the saturation deficit and stability of the atmosphere with increasing temperatures. A consistent result was observed in the recent study of Walsh et al. (2020) using the climatological averages of the large-scale variables for similar model experiments. Also, there is an increase in stability with latitude in all three constant SST simulations. The increasing precipitation toward the poles heats the upper troposphere at a higher rate, making the atmosphere more stable. Therefore, the atmospheric column is more strongly convectively heated in the poles, thereby increasing the stability compared to the tropics.

One of the hypotheses that explain the reduced frequency of TC formation is the reduced upward mass flux hypothesis (Sugi et al., 2002; Held & Zhao, 2011; Sugi et al., 2012). The decrease in the upward mass flux with increasing surface temperatures is found to be a robust result in climate models (Vecchi and Soden, 2007b; Held and Soden, 2006). The decrease in upward mass flux in the mid-troposphere of the tropics with increasing temperatures is due to increased stability of the atmosphere (Sugi et al., 2002, 2012). In the current study, in the lower latitude bands, 0-30°, and 30-45°, we observe the detected individual storms in the SST25 experiment have more upward mass flux compared to the SST30 experiment (Figure 7.10). That means there is a reduced upward mass flux with increasing temperatures in the lower latitudes in the aqua-planet simulations. In contrast, at higher latitudes, the storms in the SST30 experiment have higher upward mass flux. Similar trends are observed for the OW parameter and the 10m wind speed variable, with higher values for SST25 storms at lower latitudes and enhanced values for storms in the SST30 experiment at higher latitudes.

Another hypothesis explaining the reduction in the TC frequency is the saturation deficit hypothesis (Emanuel et al., 2008; Emanuel, 2010). These studies note that increased values of low-level saturation deficit lead to reduced TC formations. An idealized study by Rappin et al. (2010) noted that larger values of saturation deficit lead to a slower rate of vortex development, thereby leading to reduced cyclone genesis. In the current study, we observe with increasing temperatures that there is an increase in the saturation deficit across all the latitude bands (Figure 7.9). In contrast, the RH shows prominent differences between different experiments at the lower latitudes, that there is a decrease in the humidity with increasing temperature. Also, Sugi et al. (2015) noted that both these hypotheses explaining the reduction in the TC frequency might be not independent. With increased saturation, the atmosphere is drier and leads to reduced upward mass flux and thereby reduces the potential for TC formation from an initial vortex. In the current study, for all the latitude bands across the globe, we observe a stronger relationship to the stability and the saturation deficit of the atmosphere compared to the reduction in the low-level relative humidity or the reduced upward mass flux. This weaker relationship with the Omega maybe perhaps due to complex interactions between different variables but is not investigated here.

In order to further understand the co-influences of the different environmental variables on TC formation with increasing temperatures, we have performed a binary classification for the storms in the SST25 and SST30 experiments. The decision rules obtained from the C4.5 decision tree classification algorithm for the SST25 and SST30 storms (Table 7.2) show that 700 hPa RH acts as a vital discriminator for the storms at the lower latitudes (0-30° and 30-45°), whereas the OW parameter (which correlates with Omega) acts as the most significant influencing variable at higher latitudes. Therefore, RH and Omega or the OW parameter act as the most significant influencing variables for TC formation in the SST25 and SST30 experiments. Moreover, TC formation at the higher latitudes, where there is stronger stability and increased saturation deficit of the atmosphere, requires higher values of the low deformation vorticity (OW) or higher values of upward mass flux at the time of genesis. Furthermore, perhaps those vortices having lower values of the OW or lower values of upward mass flux, thus indicating a weaker circulation, do not develop under strong stable and reduced moisture conditions. These unfavorable environmental conditions may, therefore, lead to a reduction in the TC frequency at those latitudes as not all the TCs may have the required higher values of the upward mass flux or vorticity.

Also, with increasing SSTs, we observe a decrease in the number of non-developing tropical depressions, similar to the decrease in developing cases. This decrease in tropical depressions frequency agrees with the decrease in the frequency of TC seeds in a warmer climate shown by Sugi et al. (2020). Therefore, within increasing temperatures, the reduction in the overall formation rate may be due to reduced moisture (higher saturation deficit) and the reduced upward mass flux or the non-development of the weaker vortices under more stable atmospheric conditions. Also, from the environmental differences, we observe that the OW parameter and Omega act as the most significant discriminating variables for developing versus non-developing depressions under reduced moisture and strong, stable atmospheric conditions.

7.5. Conclusion

In this study, we have shown the response of the global tropical cyclone (TC) and tropical depression (TD) frequency to increased sea surface temperatures (SST) and the associated environmental conditions. For this, we use aqua-planet simulations using a global high-resolution climate model (ACCESS) with three different values of uniform SSTs and two fundamentally different TC detection and tracking schemes. A new pathway of the TC formation, the 'marsupial pouch theory,' explains the formation of an initial vortex in a closed pouch region within large-scale disturbances. We use the OWZP scheme (a resolution-independent method) to detect the locations within the pouches favorable for TC formation. Another TC detection scheme, the CSIRO scheme (a resolution-dependent method), detects the vortices that resemble observed TCs using wind speed and vorticity thresholds. In the current study, we have observed the following results:

(i) Globally, we find an increase in the saturation deficit and stability of the atmosphere with increasing temperatures (Figure 7.9) and a decrease in the TC frequency using both the resolution-dependent and independent TC tracking schemes (Table 7.1).

(ii) Across the lower latitudes (0-30°), we observed a reduced TC frequency (Figure 7.5) accompanied by lower values of low-level RH, higher Omega (lower values of upward mass flux), higher saturation deficit, and higher atmospheric stability with increasing temperatures Figure (7.9 & 7.10). This result is consistent with both the earlier hypotheses of reduced upward mass flux and increased saturation deficit. This indicates that in the lower latitudes under increased SSTs, the increased stability of the atmosphere with increasing saturation deficit and reduced upward mass flux leads to a decrease in the TC frequency.

(iii) The phenomenon-based tracking scheme (OWZP) identifies a poleward shift in the genesis locations in low latitudes (30-45°) with increasing SSTs (Figure 7.5) due to lower values of vertical wind shear (Figure 7.10).

(iv) Although the TCs within ACCESS model have weaker intensity than in observations, we observe an increase in the frequency of the intense storms with increasing SSTs using both the tracking schemes (Figure 7.1).

(v) At higher latitudes, which have higher stability than the lower latitudes in the SST30 experiment (Figure 7.9), we notice that the individual TC circulations have higher values of OW and lower Omega (Figure 7.10). This result suggests that under more stable atmospheric conditions with increased values of saturation deficit, those initial vortices with higher strength are more likely to develop into TCs compared to the low strength vortices even with the similar background conditions of other environmental variables.

(vi) Also, we observe a reduction in the formation of both developing and non-developing tropical depressions, suggesting that there may be a reduction in the frequency of the weaker vortices in a future warmer climate (Table 7.1).

(vii) The low-level OW parameter and the Omega act as significant influencing variables leading to the development of an initial depression stage vortex to an intense TC. The non-developing tropical depressions have significantly lower values of low-deformation vorticity (OW) and upward mass flux (increased Omega) compared to the developing cases (Figure 7.13).

This study observes a reduction in tropical depressions (i.e., weaker vortices) and an increase in the number of intense storms with increasing temperatures. Under more stable atmospheric conditions with lower values of moisture (higher saturation deficit), only sufficiently strong vortices at the time of the genesis develop into tropical storms. It is also important to know the influence of vertical wind shear on the TC formation and its development with varying meridional SST gradients due to the expansion of tropics in future climates (Held & Hou, 1980; Lu et al., 2007). Future studies will focus on other aqua-planet experiments with different meridional SST gradients that lead to specific changes in TC frequency due to the changes in the Hadley circulation and the development of vertical wind shear caused by an altered thermal wind balance. Also, the increased values of shear may influence the formation of the shear sheath (as described in chapter 4), which is a protective layer surrounding a vortex that restricts the inflow of dry air or opposite vorticity supporting the vortex for the formation and development (Rutherford et al., 2015, 2018). In addition, it is essential to compare the current results with simulations that have constant insolation and constant radiative forcing (Shi & Bretherton, 2014; Merlis et al., 2016; Chavas & Reed, 2019) instead of seasonally varying values in similar uniform SST experiments.

Chapter 8. Conclusion

8.1. Summary

Direct climate model simulations with the help of different detection and tracking schemes are increasingly used to understand the changes to different tropical cyclone (TC) characteristics in different climates. Climate models show a decrease in the frequency of the formation in a future climate, but the precise mechanism leading to the reduced TC frequency is not clear (Walsh et al., 2016). The discrepancies in the associated mechanisms suggest a lack of proper understanding of the environmental factors influencing TC formation. Therefore, it is vital to understand the relationship between large-scale climate and TC formation for increasing the confidence of the future TC frequency projections. The marsupial pouch theory improves our current understanding of the processes involved in the TC formation within large-scale disturbances. The OWZP TC tracking scheme, which was designed to detect the sweet spots for TC formation within the marsupial pouches of large-scale disturbances, is used in the current study in different model climates (i.e., current and future warmer idealized climates). The current thesis examines the changes in the marsupial pouch environments and the associated environmental conditions in different climates using statistical analysis and numerical modeling techniques.

Moreover, the fundamental differences in the tracking schemes can also lead to specific variations in the frequency statistics in addition to other uncertainties such as model resolution and model physics (Horn et al., 2014; Shaevitz et al., 2014; Walsh et al., 2016; Camargo & Wing, 2016). Hence, there is a requirement of a generalized tracking scheme that can be applied to different models having a different resolution. The current study compares TCs detected within the marsupial pouches by the OWZP scheme (a phenomenon-based method) with the TCs detected using the CSIRO tracking scheme (a traditional method) in different climate model simulations. Accordingly, the present thesis has been compiled from four research objectives to understand the relationship between TC formation and climate, described as follows:

8.1.1 Summary of Objective 1

Earlier studies have noted that only some tropical disturbances lead to TC formation due to differences in the characteristics of the tropical disturbances and the changes in the surrounding environmental conditions (Mc Bride & Zehr, 1981; Dunkerton et al., 2009;

Montgomery & Smith, 2010; Fu et al., 2012; Rutherford et al., 2018). Therefore, only a few of the marsupial pouches may develop into TSs, those pouches that are under favorable thermodynamical and dynamical conditions. Due to differences in the sources of the tropical disturbances and the environmental differences, the relative importance of the environmental variables vary across ocean basins (Fu et al., 2012; Peng et al., 2012). It is essential to understand the basin wise differences in the limiting factors for tropical storm (TS) formation for developing better TS prediction schemes in the global ocean basins. Therefore, objective 1 (Chapter 4 of the current thesis) focuses on the statistical analysis of pre-genesis environmental conditions, i.e., 48 hours before tropical depression (TD) declaration across each of the ocean basins. Here, both developing and non-developing TDs are detected within marsupial pouches using the OWZP scheme in the reanalysis data for the 30 years form 1989-2018. This study gave us a better understanding of why only a few depressions develop into TSs. Here we have observed that developing cases have a stronger protective layer called the shear sheath around the cyclonic core that protects the initial vortex from external influences such as dry air intrusion. In the non-developing cases, the weaker shear sheath layer to the west side of the depression in the mid troposphere along with higher vertical wind shear leads to the intrusion of dry air to the center of the storm. Therefore, the initial circulations with a more compact vertical structure having a stronger protective layer from the low troposphere to the mid troposphere with a strong anti-cyclonic circulation in the upper levels under favorable environmental conditions lead to the development of the initial depression. The contribution of the thermodynamical and dynamical conditions varies from basin to basin, perhaps due to differences in the sources for the tropical cyclone formation (large-scale disturbances) and the modulating influences (e.g. ENSO, MJO). In addition to these environmental variables, the study shows that an additional variable, the OW surrounding parameter, which gives the strength of the surrounding shear sheath, can be used to predict the formation of a TS in addition to other environmental predictors. This study gives an overview of the structural variations and environmental factors influencing an initial tropical depression to develop into a TS across the global ocean basins.

8.1.2 Summary of Objective 2

The forecasts of TC genesis, which involves complex multi-scale interactions, are improving using numerical weather prediction models (Halperin et al., 2016). Also, different statistical prediction schemes for TC formation have been formulated with different sets of predictors (i.e., environmental or storm-representative fields) across global ocean basins. Earlier prediction schemes consider only the linear relationships between the predictor and the predictand (Henon & Hobgood, 2003; Schumacher et al., 2009) and do not consider the non-linear interactions between different environmental predictors. Recent advances in the data mining techniques, which consider the non-linear interactions between different predictors, improve the TC formation prediction statistics (Park et al. 2016; Kim et al., 2019). Also, they indicate a different combination of the predictors, which improves our current understanding of the relationship between the environment variables and TC formation (Li et al., 2009; Zhang et al., 2015). Therefore, objective 2 (chapter 5 of the current thesis) employs three different machine learning algorithms to predict TS formation from an initial tropical depression dataset in each of the ocean basins with the same set of environmental predictors. This analysis indicates that random forests (ensemble decision trees) and support vector machines (optimal hyperplanes and kernel function) have better skill in predicting the developing tropical depressions using background environmental conditions as predictors compared to individual decision trees. Also, the relative importance of the variables influencing the TS formation is estimated using the three machine learning techniques. A variation in the importance of the thermodynamical and the dynamical variables is observed in each ocean basins. Although a difference in the order of importance for the predictors is observed from the statistical analysis for objective 1, this machine learning analysis additionally gave us the threshold values of the important environmental predictors. The decision trees, which have less prediction accuracy than the other two techniques, are combined with a discrete Markov model to estimate the certainty in the prediction statistics. The hybrid approach of decision trees and Markov decision processes shows that developing tropical depressions have a high probability of retaining the more favorable environmental states than the non-developing cases from two days before the formation of a tropical depression. This hybrid approach also improves the confidence in the prediction of TS formation using decision trees. Therefore, this study shows that with the help of the OWZP TC tracking scheme, environmental predictors, and machine learning algorithms, we can predict the developing tropical depressions across the global ocean basins.

8.1.3 Summary of Objective 3

Most of the climate models can generate a reasonable TC climatology, geographical distribution, seasonal climatology, and TC intensities. Certain inconsistencies exist in the prediction of TCs in future climates due to differences in the downscaling methodologies, tracking schemes, warming scenarios, projected large-scale conditions, and model physics (Walsh et al., 2016; Camargo & Wing, 2016). The current climate models are advancing towards high resolution, which improves the geographical distributions, TC intensities, and their detections using resolution-dependent tracking schemes. Still, specific issues remain using different resolution-dependent tracking algorithms in detecting the TCs that have weaker intensities which have an influence on the global TC frequency statistics (Camargo & Wing, 2016). Also, not all the circulations identified using the 10m wind speed thresholds, a typical criterion in such schemes, are actually TCs as they could be monsoon gyres (Lander, 1991). Therefore, a generalized tracking scheme is required to apply to different resolution models, and in all the ocean basins, that better differentiates the TC circulations from the monsoon gyres. Also, the confidence in future projections depends on the quality of the climate model simulations in the current climate. It is, therefore, important to understand the geographical distribution of the TC formation and their relationships to different large-scale climate variables using different tracking schemes. Thus, objective 3 (chapter 6 of the current thesis) compares two fundamentally different tracking schemes in a high-resolution climate model simulation of the current climate in simulating different TC frequency characteristics across the global ocean basins. The study also estimates the relationship of the TC formation to different large-scale environmental variables for both the TCs detected within marsupial pouches using the OWZP scheme and the traditionally detected TCs using the CSIRO scheme. The TCs detected using a traditional TC tracking scheme, the CSIRO scheme (a resolution-dependent method) are a subset of marsupial pouch environments detected using a phenomenon-based tracking scheme OWZP (a resolution-independent method) in a high-resolution ACCESS model simulation. The phenomenon-based tracking scheme (OWZP) has comparatively better performance for different TC frequency characteristics. The OWZP scheme performs better than the CSIRO scheme in the North Indian Ocean in separating the monsoon gyres from the TCs. This study also shows that the large-scale model climate is vital for the reasonable estimation of the TC frequency statistics irrespective of the tracking schemes. This sensitivity study using two fundamentally different tracking schemes is necessary to understand the global TC climatology and environmental controls on TC formation for improving confidence in the projections of future TC activity.

8.1.4 Summary of Objective 4

Even though there is an increased certainty in the projections of the mean large-scale environmental conditions in the future climate, there exist uncertainties in the simulated TC frequency activity in the climate models (Knutson et al., 2019b). A better theoretical understanding of the large-scale climate controls on TC formation is necessary for reducing the existing uncertainties in the climate model simulations (Walsh et al., 2015; Camargo & Wing, 2016). Aqua-planet simulations, which reduce the model complexities by removing some of the co-influencing factors of TC formation, help us to better understand the climate controls on TC formation (Merlis & Held, 2019). For instance, the aqua-planet simulations remove the orography and cover the globe with either SSTs or slab ocean as TCs form over oceans, creating a tropical cyclone world. Therefore, chapter 7 of the current thesis focusses on fundamental controls of TC formation, e.g., the influence of sea surface temperatures (SSTs) using aqua-planet simulations with two fundamentally different TC tracking schemes. This study shows that the phenomenon-based TC tracking scheme (OWZP) has better relationships with large-scale environmental conditions than the traditional TC tracking scheme (CSIRO). Therefore, the OWZP scheme acts as a more generalized tracking scheme for different resolution models in different climates.

Furthermore, these idealized climate model simulations using different values of the sea surface temperatures show that with increasing SSTs, there is a decrease in the frequency of the formation with both the tracking schemes. This decrease in frequency is due to increased saturation deficit of the lower troposphere and increased stability of the atmosphere (Sugi et al., 2012; Emanuel, 2013; Sugi et al., 2015). There is an increase in the frequency of the intense storms with increasing temperatures. The study also shows that there is a decrease in the frequency of the tropical depressions with increasing temperatures suggesting a decrease in the number of weaker storms with increasing temperatures (Sugi et al., 2020). The statistical analysis of environmental variables shows that if the atmosphere is drier and more stable, an initial vortex needs to have more low-deformation vorticity and more upward mass flux for TC formation to occur. This study further emphasized that there is a decrease in the frequency of low intense storms and an increase in the frequency of intense storms with increasing temperatures. This study also improved our current understanding of the mechanisms leading to the decreased TC frequency in future climate.

8.2. Major findings

The research aimed to understand the changes in the marsupial pouch environments and the associated environmental conditions for a better understanding of the relationship between climate and TC formation. The developing tropical depressions detected within marsupial pouches have a strong protective layer surrounding the cortex core region from lower to mid troposphere. The important environmental and storm-representative predictors two days before tropical depression formation are also useful in developing stochastic TS forecast schemes using new machine learning techniques, namely hybrid DT with MDP, RF, and SVM. The marsupial pouch environments have better relationships with the large-scale thermodynamical and dynamical conditions than the TCs detected using the resolution-dependent thresholds in the current and idealized climates. The increased saturation deficit and stability of the atmosphere lead to a reduction in the TC frequency with increasing sea surface temperatures. Based on quantitative analysis using observations and modeling, it can be concluded that marsupial pouch theory acts as a fundamental paradigm of TC formation in different climates. The results also indicate that the OWZP TC detection and tracking scheme is a more robust TC detection scheme that can be applied to different climates and model resolutions.

8.3. Recommendations for Future work

The following section gives a few recommendations for future work, which can improve the results from the objectives carried out in the current thesis work.

1. The detection of the initial tropical depressions in the global forecast analysis data using the OWZP tracking scheme and identifying the environmental and structural differences using similar environmental variables to observe whether the same order of important variables exists as in the reanalysis data. This could help us better understand the environmental controls on TS formation to make better statistical predictions and improve our confidence in the future TC activity.

2. The application of different machine learning algorithms to the forecast analysis data to predict TS formation from the initial tropical depression dataset may be advantageous for real-time forecasts. The inclusion of additional environmental predictors and selecting the best predictors using the advanced lasso and elastic net regularization techniques (Bieli et al., 2020) may be an advance in statistical forecasting of TS formation.

3. The diagnosis of the TC frequency characteristics and their relationships to environmental conditions using the phenomenon-based OWZP tracking scheme and CSIRO schemes could be applied to other atmosphere-only or a coupled ocean-atmosphere climate models having different resolutions. This intercomparison between different models and resolutions helps us to observe further the sensitivity of TC detection to the tracking schemes and the model resolution. The inter-model comparison among a range of advanced models would be advantageous to identify a more generalized tracking scheme that could apply to different model climates.

4. Different sensitivity experiments using the climate models with a varied set of stability profiles to observe the changes in the frequency of formation in the idealized experiments will strengthen the confidence in the stability and the upward mass flux hypothesis explaining the reduction in the TC frequency.

5. Understanding the TC frequency response to different meridional SST gradients and the associated changes in the circulation and the atmospheric conditions using the aqua-planet simulations advances our knowledge of the environmental controls for TC formation. The different meridional SST gradients impose different values of vertical wind shear which influences the TC formation (Walsh et al., 2020). It would be interesting to see the influence of the vertical wind shear on the formation of the shear sheath in the initial tropical depressions and TCs (Rutherford et al., 2018) in the meridional SST gradients experiments (Walsh et al., 2020).

6. In the current study, the idealized simulations include seasonally varying radiative forcing to compare the results to the terrestrial climate. Instead, a more simplified model by keeping the radiative forcing constant and observing the changes to the frequency helps us further isolate the role of the sea surface temperatures (SSTs) in TC formation. Merlis et al. (2016) suggest further investigation of the SST sensitivity of TC formation using more simplified aqua-planet simulations using different models. Also, an intercomparison of the models with aqua-planet configurations will advance the understanding of the TC formation's environmental factors.

7. Analyzing the TC frequency changes and their relationship to the environmental conditions using Earth system models can further advance our understanding on the relationships between climate and TC formation. This analysis using advanced Earth system models also decipher the role of ocean mixing due to TC activity and its influence on biogeochemical variability.

References

Agudelo, P., C. Hoyos, J. Curry, and P. J. Webster (2011), Probabilistic discrimination between large-scale environments of intensifying and decaying African Easterly Waves, *Climate Dyn.*, *36*(7–8), 1379–1401.

Alland, J.J., B.H. Tang, and K.L. Corbosiero (2017), Effects of midlevel dry air on development of the axisymmetric tropical cyclone secondary circulation. *J. Atmos. Sci.*, *74*(5), 1455-1470.

Alpaydin, E. (2020), *Introduction to machine learning*. MIT press.

Alzantot, M. (2017), Deep reinforcement learning demysitifed (Episode 2)-Policy iteration, value iteration and Q-learning. (https://medium.com/@m.alzantot/deep-reinforcement-learning-demysitifed-episode-2-policy-iteration-value-iteration-and-q-978f9e89ddaa).

Asaadi, A., G. Brunet, and M.K. Yau (2016), On the dynamics of the formation of the Kelvin cat's-eye in tropical cyclogenesis. Part I: Climatological investigation. *J. Atmos. Sci.*, *73*(6), 2317-2338.

Asaadi, A., G. Brunet and M.K. Yau (2017), The Importance of Critical Layer in Differentiating Developing from Nondeveloping Easterly Waves. *J. Atmos. Sci.*, *74*(2), 409-417.

Avila, L.A. (1991), Atlantic tropical systems of 1990. *Mon. Weather Rev.*, *119*(8), 2027-2033.

Avila, L.A., and R.J. Pasch (1992), Atlantic tropical systems of 1991, *Mon. Weather Rev.*, *120*(11), 2688–2696.

Avila, L.A., R.J. Pasch, and J.G. Jiing (2000), Atlantic tropical systems of 1996 and 1997: Years of contrasts, *Mon. Weather Rev.*, *128*(10), 3695–3706.

Bakkensen, L.A. and R.O. Mendelsohn (2016), Risk and adaptation: evidence from global hurricane damages and fatalities. *J. Asso. Environ. Res. Econ.*, *3*(3), 555-587.

Ballinger, A.P., T.M. Merlis, I.M. Held, and M. Zhao (2015), The sensitivity of tropical cyclone activity to off-equatorial thermal forcing in aquaplanet simulations. *J. Atmos. Sci.*, *72*, 2286– 2302.

Barto, A.G. (1995), Reinforcement learning and dynamic programming. In *Analysis, Design and Evaluation of Man–Machine Systems*, 407-412. Pergamon.

Bell, R., J. Strachan, P.J. Vidale, K. Hodges, and M. Roberts (2013), Response of tropical cyclones to idealized climate change experiments in a global high-resolution coupled general circulation model. *J. Climate*, *26*, 7966-7980.

Bell, R., K. Hodges, P.J. Vidale, J. Strachan, and M. Roberts (2014), Simulation of the global ENSO–tropical cyclone teleconnection by a high-resolution coupled general circulation model. *J. climate*, *27*(17), 6404-6422.

Bell, M.M., and M.T. Montgomery (2019), Mesoscale Processes during the Genesis of Hurricane Karl (2010). *J. Atmos. Sci.*, *76*(8), 2235-2255.

Bell, S.S., S.S. Chand, K.J. Tory, and C. Turville (2018), Statistical assessment of the OWZ tropical cyclone tracking scheme in ERA-Interim, *J. Climate*, *31*, 2217-2232.

Bell S.S., S.S. Chand, K.J. Tory, A.J. Dowdy, C. Turville, and H. Ye (2019a), Projections of Southern hemisphere tropical cyclone track density using CMIP5 models. *Climate Dyn*., *52*, 6065-6079.

Bell, S.S., S.S. Chand, K.J. Tory, C. Turville, and H. Ye (2019b), Eastern North Pacific tropical cyclone activity in historical and future CMIP5 experiments: assessment with a model-independent tracking scheme. *Climate Dyn.*, *53*, 4841-4855.

Bell, S.S., S.S. Chand, S.J. Camargo, K.J. Tory, C. Turville, and H. Ye (2019c), Western North Pacific tropical cyclone tracks in CMIP5 models: Statistical assessment using a model-independent detection and tracking scheme. *J. Climate*, *32*(21), 7191-7208.

Bengtsson, L., H. Böttger, and M. Kanamitsu (1982), Simulation of hurricane-type vortices in a general circulation model. *Tellus*, 34, 440–457.

Bengtsson, L., M. Botzet, and M. Esch (1995), Hurricane-type vortices in a general circulation model. *Tellus A*, 47, 175–196.

Bengtsson, L., K.I. Hodges, and M. Esch (2007a), Tropical cyclones in a T159 resolution global climate model: Comparison with observations and re-analyses. *Tellus A: Dynam. Meteorol. Oceanogr.*, *59*, 396-416.

Bengtsson, L., K.I. Hodges, M. Esch, N. Keenlyside, L. Kornblueh, J.J. Luo, and T. Yamagata (2007b), How may tropical cyclones change in a warmer climate?. *Tellus A: Dynam. Meteorol. Oceanogr.*, *59*(4), 539-561.

Berwick, R. (2003), An Idiot's guide to Support vector machines (SVMs). *Retrieved on October*, *21*, 2011. (http://web.mit.edu/6.034/wwwbob/svm-notes-long-08.pdf).

Bessafi, M., and M.C. Wheeler (2006), Modulation of south Indian Ocean tropical cyclones by the Madden–Julian oscillation and convectively coupled equatorial waves. *Mon. Weather Rev.*, *134*(2), 638-656.

Best, M.J., M. Pryor, D.B. Clark, G.G Rooney, R. Essery, C.B. Ménard, J.M. Edwards, M.A. Hendry, A. Porson, N. Gedney, and L.M. Mercado (2011), The Joint UK Land Environment Simulator (JULES), model description–Part 1: energy and water fluxes. *Geosci. Model Dev.*, *4*, 677-699.

Bhatia, K., G. Vecchi, H. Murakami, S. Underwood, and J. Kossin (2018), Projected response of tropical cyclone intensity and intensification in a global climate model. *J. Climate*, 31, 8281–8303.

Bi, D., M. Dix, S.J. Marsland, S. O'Farrell, H. Rashid, P. Uotila, A.C. Hirst, E. Kowalczyk, M. Golebiewski, A. Sullivan, and H. Yan (2013), The ACCESS coupled model: description, control climate and evaluation. *Aust. Meteorol. Oceanogr. J*, *63*, 41-64.

Bieli, M., A.H. Sobel, S.J. Camargo, and M.K. Tippett (2020), A Statistical Model to Predict the Extratropical Transition of Tropical Cyclones. *Weather Forecast.*, *35*(2), 51-466.

Bister, M., and K.A. Emanuel (1997), The genesis of Hurricane Guillermo: TEXMEX analyses and a modeling study. *Mon. Weather Rev.*, *125*(10), 2662-2682.

Bister, M., and K.A. Emanuel (1998), Dissipative heating and hurricane intensity. *Meteor. Atmos. Phys.*, 50, 233–240.

Brammer, A., and C.D. Thorncroft (2015), Variability and evolution of African easterly wave structures and their relationship with tropical cyclogenesis over the eastern Atlantic. *Mon. Weather Rev.*, *143*(12), 4975-4995.

Braun, S.A. (2010), Reevaluating the role of the Saharan air layer in Atlantic tropical cyclogenesis and evolution. *Mon. Weather Rev., 138(6),* 2007–2037.

Briegel, L.M., and W.M. Frank (1997), Large-scale influences on tropical cyclogenesis in the western North Pacific. *Mon. Weather Rev.*, *125*(7), 1397-1413.

Breiman, L. (2001), Random Forests. *Machine Learning*, 45**,** 5–32.

Brown, A.R., R.J. Beare, J.M. Edwards, A.P. Lock, S.J. Keogh, S.F. Milton, and D.N. Walters (2008), Upgrades to the boundary-layer scheme in the Met Office numerical weather prediction model. *Bound.-Lay. Meteorol.*, *128*, 117-132.

Bruyere, C.L., G.J. Holland, and E. Towler (2012), Investigating the use of a genesis potential index for tropical cyclones in the North Atlantic basin. *J. Climate*, *25*(24), 8611-8626.

Busoniu, L., B. De Schutter, and R. Babuška (2010), Approximate dynamic programming and reinforcement learning. In *Interactive collaborative information systems*, 3-44, Springer, Berlin, Heidelberg.

Camargo, S.J., and S.E. Zebiak (2002), Improving the Detection and Tracking of Tropical Cyclones in Atmospheric General Circulation Models. *Weather Forecast.*, 17(6), 1152–1162.

Camargo S.J., A.G. Barnston, and S.E. Zebiak (2005), A statistical assessment of tropical cyclones in atmospheric general circulation models. *Tellus*, *57,* 589–604.

Camargo, S.J., A.W. Robertson, S.J. Gaffney, P. Smyth, and M. Ghil (2007b), Cluster analysis of typhoon tracks. Part I: General properties. *J. Climate*, *20*, 3635-3653.

Camargo, S.J., K.A. Emanuel, and A.H. Sobel (2007), Use of a genesis potential index to diagnose ENSO effects on tropical cyclone genesis. *J. Climate*, *20*(19), 4819-4834.

Camargo, S.J., A. Sobel, A. Barnston, and P. Klotzbach (2010), The influence of natural climate variability, and seasonal forecasts of tropical cyclone activity, in *Global Perspectives on Tropical Cyclones*, 2nd ed., edited by J. C. L. Chan and J. D. Kepert, 325– 360, World Scientific Publishing Co., Singapore.

Camargo, S.J. (2013), Global and regional aspects of tropical cyclone activity in the CMIP5 models. *J. Climate,* 26, 9880–9902.

Camargo, S.J., M.K. Tippett, A.H. Sobel, G.A. Vecchi, and M. Zhao (2014), Testing the performance of tropical cyclone genesis indices in future climates using the HiRAM model. *J. Climate*, 27, 9171–9196.

Camargo, S.J., and A.A. Wing (2016), Tropical cyclones in climate models. *Wiley Interdiscip. Rev.: Climate Change*, 7, 211–237.

Camargo, S.J., A.H. Sobel, A.D. Delgenio, J.A. Jonas, M. Kelley, Y. Lu, D.A. Shaevitz, and N. Henderson (2016), Tropical cyclones in the GISS ModelE2. *Tellus A: Dyn. Meteorol. Oceanogr. 68(1)*, 31494.

Camargo, S. J., C. Giulivi, A. H. Sobel, A. A. Wing, D. Kim, Y. Moon, J. D. Strong, A. D. Del Genio, M. Kelley, H. Murakami, K. A. Reed, E. Scoccimarro, G. A. Vecchi, M. F. Wehner, C. M. Zarzycki and M. Zhao (2020), Characteristics of model tropical cyclone climatology and the large-scale environment, *J. Climate*, 33, 4463-4487.

Chadés, I., M. Cros, F. Garcia, and R. Sabbadin (2005), Markov decision process (mdp) toolbox v2. 0 for matlab. *INRA Toulouse, INRA, France, http://www. inra. fr/internet/Departements/MIA/T/MDPtoolbox*.

Chand, S.S., K.J. Walsh, and J.C. Chan (2010), A Bayesian regression approach to seasonal prediction of tropical cyclones affecting the Fiji region. *J. climate*, *23*(13), 3425-3445.

Chand, S.S., and K.J. Walsh (2011), Forecasting tropical cyclone formation in the Fiji region: A probit regression approach using a Bayesian fitting. *Weather Forecast.*, *26*(2), 150-165.

Chand, S.S., K.J. Tory, H. Ye, and K.J.E. Walsh (2017), Projected increase in El Niño-driven tropical cyclone frequency in the Pacific. *Nat. Climate Change*, 7, 123–127.

Chavas, D.R., and K. Emanuel (2014), Equilibrium tropical cyclone size in an idealized state of axisymmetric radiative–convective equilibrium. *J. Atmos. Sci.*, *71*, 1663-1680.

Chavas, D.R., K.A. Reed, and J.A. Knaff (2017), Physical understanding of the tropical cyclone wind-pressure relationship. *Nature communications*, *8*(1), 1-11.

Chavas, D.R. and K.A. Reed (2019), Dynamical aquaplanet experiments with uniform thermal forcing: System dynamics and implications for tropical cyclone genesis and size. *J. Atmos. Sci.*, *76*(8), 2257-2274.

Chawla, N.V., KW. Bowyer, L.O. Hall, and W.P. Kegelmeyer (2002), SMOTE: synthetic minority over-sampling technique. *J. art. Intel. Res.*, *16*, 321-357.

Cheung, K.K., and R.L. Elsberry (2002), Tropical cyclone formations over the western North Pacific in the navy operational global atmospheric prediction system forecasts. *Weather forecast.*, *17*(4), 800-820.

Chia, H.H., and C.F. Ropelewski (2002), The interannual variability in the genesis location of tropical cyclones in the northwest Pacific. *J. Climate*, *15*, 2934-2944.

Christensen, J., T. Carter, M. Rummukainen, and G. Amanatidis (2007), Evaluating the performance and utility of regional climate models: the PRUDENCE project. *Climatic Change*, 81(0), 1–6.

Collobert, R., and S. Bengio (2004), Links between perceptrons, MLPs and SVMs. In *Proceedings of the twenty-first international conference on Machine learning*, 23.

Compo, G.P., J.S. Whitaker, P.D. Sardeshmukh, N. Matsui, R.J. Allan, X. Yin, B.E. Gleason, R.S. Vose, G. Rutledge, P. Bessemoulin, and S. Brönnimann (2011), The twentieth century reanalysis project. *Q. J. Roy. Meteorol. Soc.*, *137*, 1-28.

Cronin, T.W. and D.R. Chavas, (2019), Dry and semidry tropical cyclones. *J. Atmos. Sci.*, *2019*(1), 1-20.

Dailey, P., M. Huddleston, S. Brown, and D. Fasking (2009), The financial risks of climate change. *ABI Research Paper*, (19). (https://www.weadapt.org/sites/weadapt.org/files/legacy-new/knowledge-base/files/1040/504a19b1e3d0efinancial-risks-of-climate-change-pdf.pdf)

Daloz, A.S., and S.J. Camargo (2018), Is the poleward migration of tropical cyclone maximum intensity associated with a poleward migration of tropical cyclone genesis? *Climate Dyn.*, 50, 705-715.

Davis C.A. (2015), The formation of moist vortices and tropical cyclones in idealized simulations. *J. Atmos. Sci.*, 72(9), 3499-3516.

Davis, C.A., and D.A. Ahijevych (2012), Mesoscale structural evolution of three tropical weather systems observed during PREDICT. *J. Atmos. Sci.*, *69*, 1284-1305.

Dee, D.P., S.M. Uppala, A.J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M.A. Balmaseda, G. Balsamo, D.P. Bauer, and P. Bechtold (2011), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. Roy. Meteorol. Soc.*, *137*, 553-597.

Dickinson, M., and J. Molinari (2002), Mixed Rossby–gravity waves and western Pacific tropical cyclogenesis. Part I: Synoptic evolution. *J. Atmos. Sci.*, *59*(14), 2183-2196.

Ditchek, S.D., T.C. Nelson, M. Rosenmayer, and K.L. Corbosiero (2017), The relationship between tropical cyclones at genesis and their maximum attained intensity. *J. Climate*, *30*(13), 4897-4913.

Dowdy, A.J., L. Qi, D. Jones, H. Ramsay, R. Fawcett, and Y. Kuleshov (2012), Tropical cyclone climatology of the South Pacific Ocean and its relationship to El Niño–Southern Oscillation. *J. Climate*, *25*(18), 6108-6122.

Dunion, J.P., and C. S. Velden (2004), The impact of the Saharan Air Layer on Atlantic tropical cyclone activity. *Bull. Amer. Meteor. Soc.*, *85(3)*, 353-365.

Dunkerton, T.J., M.T. Montgomery, and Z. Wang (2009), Tropical cyclogenesis in a tropical wave critical layer: Easterly waves. *Atmos. Chem. Phys.*, *9*(15).

Dunnavan, G.M., E.J. McKinley, P.A. Harr, E. A. Ritchie, M.A. Boothe, M. Lander, and R. L. Elsberry (1992), Tropical Cyclone Motion (TCM-92) mini-field experiment summary. *Naval Postgraduate School Tech. Rep. NPS-MR-93-001*, *98.*

Dunstone, N. J., D.M. Smith, B.B.B. Booth, L. Hermanson, and R. Eade (2013), Anthropogenic aerosol forcing of Atlantic tropical storms. *Nat. Geosci.*, 6(7), 534–539.

Durran, D. (1999), *Numerical methods for wave equations in geophysical fluid dynamics* (Vol. 32). Springer Science & Business Media.

Edwards, J.M., and A. Slingo (1996), Studies with a flexible new radiation code. I: Choosing a configuration for a large-scale model. *Q. J. Roy. Meteorol. Soc.*, *122*, 689-719.

Elsberry, R.L., M.S. Jordan, and F. Vitart (2011), Evaluation of the ECMWF 32-day ensemble predictions during 2009 season of western North Pacific tropical cyclone events on intraseasonal timescales. *Asia-Pacific J. Atmos. Sci.*, *47*(4), 305.

Emanuel K.A. (1986), An air–sea interaction theory for tropical cyclones, part I: steady-state maintenance. *J. Atmos. Sci.,* 43, 585–605.

Emanuel, K.A. (1988), The maximum intensity of hurricanes. *J. Atmos. Sci.*, *45*(7), 1143-1155.

Emanuel, K. (2001), Contribution of tropical cyclones to meridional heat transport by the oceans. *J. Geophy. Res. Atmos.*, *106*(D14), 14771-14781.

Emanuel, K. (2003), Tropical cyclones. *Annu. Rev. Earth Planet. Sci.*, *31*, 75-104.

Emanuel, K., and D.S. Nolan (2004), Tropical cyclone activity and the global climate system. In *26th Conference on Hurricanes and Tropical Meteorology*, 240-241.

Emanuel, K. (2005), *Divine wind: the history and science of hurricanes*. Oxford University Press.

Emanuel, K., R. Sundararajan, and J. Williams (2008), Hurricanes and global warming: Results from downscaling IPCC AR4 simulations, *Bull. Amer. Meteor. Soc.*, 89, 347–367.

Emanuel, K. A. (2010), Tropical cyclone activity downscaled from NOAA-CIRES Reanalysis, 1908–1958, *J. Adv. Model. Earth Syst.*, 2(1).

Emanuel, K. A. (2013), Downscaling CMIP5 climate models shows increased tropical cyclone activity over the 21st century. *Proc. Natl. Acad. Sci. USA*, 110, 12219–12224.

Evan, A.T., J.P. Kossin, and V. Ramanathan (2011), Arabian Sea tropical cyclones intensified by emissions of black carbon and other aerosols. *Nature*, *479*(7371), 94-97.

Evans, J.L., and J.J. Waters (2012), Simulated relationships between sea surface temperatures and tropical convection in climate models and their implications for tropical cyclone activity. *J. Climate*, *25*(22), 7884-7895.

Fan, K. (2010), A prediction model for Atlantic named storm frequency using a year-by-year increment approach. *Weather Forecast.*, *25*(6), 1842-1851.

Felton, C.S., B. Subrahmanyam, and V.S.N. Murty (2013), ENSO-modulated cyclogenesis over the Bay of Bengal. *J. Climate*, *26*, 9806-9818.

Ferreira, R.N., and W.H. Schubert (1997), Barotropic aspects of ITCZ breakdown. *J. Atmos. Sci.*, *54*(2), 261-285.

Fowler, J.P., and T.J. Galarneau Jr (2017), Influence of storm–storm and storm–environment interactions on tropical cyclone formation and evolution. *Mon. Weather Rev.*, *145*(12), 855-4875.

Frank, N.L. (1970), Atlantic tropical systems of 1969. *Mon. Weather Rev*, *98*(4), 307-314.

Frank, W.M., and G.S. Young (2007), The interannual variability of tropical cyclones. *Mon. Weather Rev.*, *135*, 3587-3598.

Frank, W.M., and P.E. Roundy (2006), The role of tropical waves in tropical cyclogenesis. *Mon. Weather Rev.*, *134*(9), 2397-2417.

Fu, B., Li, T., M.S. Peng, and F. Weng (2007), Analysis of tropical cyclogenesis in the western North Pacific for 2000 and 2001. *Weather Forecast.*, *22*(4), 763-780.

Fu, B., M.S. Peng, T. Li, and D.E. Stevens (2012), Developing versus nondeveloping disturbances for tropical cyclone formation. Part II: Western North Pacific. *Mon. Weather Rev.*, *140*(4), 1067-1080.

Gall, R., J. Franklin, F. Marks, E, N. Rappaport, and F. Toepfer (2013), The hurricane forecast improvement project. *Bull. Amer. Meteor. Soc., 94*(3), 329-343.

Gandhi, R. (2018), Support vector machine-introduction to machine learning algorithms. (*URl:https://towardsdatascience.com/supportvector-machine-introduction-to-machine-learning-algorithms-934a444fca47*).

Gettelman A., and R.B. Rood (2016), Essence of a Climate Model. *In Demystifying Climate Models* (pp.37-58). Springer, Berlin, Heidelberg.

Gleason, C.J., and J. Im (2012), Forest biomass estimation from airborne LiDAR data using machine learning approaches. *Rem. Sen. Environ.*, *125*, 80-91

Gray, W.M. (1975), Tropical cyclone genesis. *Atmos. Sci.,* 234.

Gray, W.M. (1968), Global view of the origin of tropical disturbances and storms. *Mon. Weather Rev*., *110*, 572-586.

Gray, W.M. (1998), The formation of tropical cyclones. *Meteor. Atmos. Phys.,* 67, 37–69.

Guneralp, İ., A.M. Filippi, and B.U. Hales (2013), River-flow boundary delineation from digital aerial photography and ancillary images using support vector machines. *GIScience & remote sensing*, *50*(1), 1-25.

Hall, J.D., A.J. Matthews, and D.J. Karoly (2001), The modulation of tropical cyclone activity in the Australian region by the Madden–Julian oscillation. *Mon. Weather Rev.*, *129*(12), 2970-2982.

Halperin, D.J., H.E. Fuelberg, R.E. Hart, and J.H. Cossuth (2016), Verification of tropical cyclone genesis forecasts from global numerical models: Comparisons between the North Atlantic and eastern North Pacific basins. *Weather Forecast., 31*(3), 947-955.

Halperin, D.J., H.E. Fuelberg, R.E. Hart, J.H. Cossuth, P. Sura, and R.J. Pasch (2013), An evaluation of tropical cyclone genesis forecasts from global numerical models. *Weather Forecast.*, *28*(6), 1423-1445.

Hamill, T.M., and J. Juras (2006), Measuring forecast skill: is it real skill or is it the varying climatology?. *Q. J. R. Meteorol. Soc., 132*, 2905-2923.

Han, H., S. Lee, J. Im, M. Kim, M.I. Lee, M.H. Ahn, and S.R. Chung (2015), Detection of convective initiation using Meteorological Imager onboard Communication, Ocean, and Meteorological Satellite based on machine learning approaches. *Rem. Sens.*, *7*(7), 9184-9204.

Harr, P.A., M.S. Kalafsky, and R.L. Elsberry (1996), Environmental conditions prior to formation of a midget tropical cyclone during TCM-93. *Mon. Weather Rev.*, *124*(8), 693-1710.

Hastie, T., R. Tibshirani, and J.J.H. Friedman (2001), The Elements of Statistical Learning: Data Mining, Inference, and Prediction. *Springer Series in Statistics, Springer, 552*.

Hayashi, Y.Y., and A. Sumi (1986), The 30–40 day oscillations simulated in an ‘‘aqua planet’’ model. *J. Meteor. Soc. Japan*, 64, 451–467.

Haynes, P.H., and M.E. McIntyre (1987), On the evolution of vorticity and potential vorticity in the presence of diabatic heating and frictional or other forces. *J. Atmos. Sci.*, 44, 828-841.

Held, I.M., and A.Y. Hou (1980), Nonlinear axially symmetric circulations in a nearly inviscid atmosphere. *J. Atmos. Sci.*, *37*(3), 515-533.

Held, I.M. (2005), The gap between simulation and understanding in climate modeling. *Bull. Amer. Meteor. Soc.,* 86(11), 1609–1614.

Held, I.M., and B. J. Soden (2006), Robust responses of the hydrologic cycle to global warming. *J. Climate,* 19, 5686–5699, https://doi.org/10.1175/ JCLI3990.1.

Held, I.M., and M. Zhao (2008), Horizontally homogeneous rotating radiative–convective equilibria at GCM resolution. *J. Atmos. Sci., 65,* 2003–2013.

Held, I.M., and M. Zhao (2011), The response of tropical cyclone statistics to an increase in CO2 with fixed sea surface temperatures. *J. Climate*, *24*(20), 5353-5364.

Heming, J.T. (2016), Met Office Unified Model tropical cyclone performance following major changes to the initialization scheme and a model upgrade. *Weather Forecast.*, *31*(5), 1433-1449.

Hendricks, E.A., M.T. Montgomery, and C.A. Davis (2004), The role of “vortical” hot towers in the formation of tropical cyclone Diana (1984). *J. Atmos. Sci.*, *61*(11), 1209-1232.

Hendricks, E.A. and M.S. Peng (2012), Initialization of Tropical Cyclones in Numerical Prediction Systems. *Advances in Hurricane Research: Modelling, Meteorology, Preparedness and Impacts*, p.1.

Hennon, C.C., and J.S. Hobgood (2003), Forecasting tropical cyclogenesis over the Atlantic basin using large-scale data. *Mon. Weather Rev.*, *131*(12), 2927-2940.

Hodges K.I., and R. Emerton (2015), The prediction of Northern Hemisphere tropical cyclone extended life cycles by the ECMWF ensemble and deterministic prediction systems. Part I: Tropical cyclone stage .*Mon. Weather Rev.*, *143*, 5091–5114.

Hodges, K., A. Cobb, and P.L. Vidale (2017), How well are tropical cyclones represented in reanalysis datasets?. *J. Climate*, *30*, 5243-5264.

Holland, G. J. (1983), Tropical cyclone motion: Environmental interaction plus a beta effect. *J. Atmos. Sci.*, 40, 328–342.

Holland, G. (1993), Ready Reckoner. Technical report, World Meteorological Organization; Geneva, Switzerland.

Holland, G., and C.L. Bruyère (2014), Recent intense hurricane response to global climate change. *Climate Dyn.*, *42*(3-4), 617-627.

Hopsch, S. B., C.D. Thorncroft, K. Hodges, and A. Aiyyer (2007), West African storm tracks and their relationship to Atlantic tropical cyclones, *J. Climate, 20*, 2468–2483.

Hopsch, S.B., C.D. Thorncroft, and K.R. Tyle (2010), Analysis of African easterly wave structures and their role in influencing tropical cyclogenesis. *Mon. Weather Rev.*, *138*(4), 1399-1419.

Horn, M., K. Walsh, M. Zhao, S.J. Camargo, E. Scoccimarro, H. Murakami, H. Wang, A. Ballinger, A. Kumar, D.A. Shaevitz, and J.A. Jonas (2014), Tracking scheme dependence of simulated tropical cyclone response to idealized climate simulations. *J. Clim.*, *27*, 9197-9213.

Hsieh, T.L., G.A. Vecchi, W. Yang, I.M. Held, and S.T. Garner (2020), Large-scale control on the frequency of tropical cyclones and seeds: a consistent relationship across a hierarchy of global atmospheric models. *Climate Dyn.*, *55*(11), 3177-3196.

Houze, R.A. (2004), Mesoscale convective systems. *Rev. Geophys.*, *42*(4).

IPCC (2007), Climate Change 2007: The physical science basis. Contribution of Working Group 1 to the Fourth Assessment Report of the intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge.

IPCC, (2013), Working Group I Contribution to the IPCC Fifth Assessment Report: Climate Change 2013: The Physical Science Basis, Summary for Policymakers. *IPCC, UN*.

Jaiswal, N., C.M. Kishtawal, and P.K. Pal (2013), Prediction of tropical cyclogenesis in North Indian Ocean using Oceansat-2 scatterometer (OSCAT) winds. *Meteorol. Atmos. Phys.*, *119*(3-4), 137-149.

James, R. P., and P.M. Markowski (2009), A numerical investigation of the effects of dry air aloft on deep convection. *Mon. Weather Rev.*,138, 140-161.

Johnson, N.C., and S.P. Xie (2010), Changes in the sea surface temperature threshold for tropical convection. *Nat. Geosci.*, *3*(12), 842-845.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter (2002), NCEP–DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, 83, 1631–1643.

Kecman, V. (2005), Support vector machines–an introduction. In *Support vector machines: theory and applications*, Springer, Berlin, Heidelberg, 1-47.

Kerns, B., K. Greene, and E. Zipser (2008), Four years of tropical ERA-40 vorticity maxima tracks. Part I: Climatology and vertical vorticity structure. *Mon. weather Rev.*, *136*(11), 4301-4319.

Khairoutdinov, M., and K.A. Emanuel (2013), Rotating radiative-convective equilibrium simulated by a cloud-resolving model. *J. Adv. Model. Earth Syst.*, 5, 816–825.

Kilroy, G., and R.K. Smith (2013), A numerical study of rotating convection during tropical cyclogenesis. *Q. J. Roy. Meteorol. Soc.*, 139, 1255-1269.

Kim, D., A.H. Sobel, A.D. Del Genio, Y. Chen, S.J. Camargo, M.S. Yao, M. Kelley, and L. Nazarenko (2012), The tropical subseasonal variability simulated in the NASA GISS general circulation model. *J. Climate*, *25*, 4641-4659.

Kim, S., and J. Choi (2014), An SVM-based high-quality article classifier for systematic reviews. *J. bio. Inform.*, *47*, 153-159.

Kim, M., M.S. Park, J. Im, S. Park, and M.I. Lee (2019), Machine learning approaches for detecting tropical cyclone formation using satellite data. *Rem. Sens.*, *11*(10), 1195.

Kirtman B.P., and E.K. Schneider (2000), A spontaneously generated tropical atmospheric general circulation. *J. Atmos. Sci.*, 57, 2080–2093.

Klettner, A., S. Kelly, M. Edwards, and P. Brown, (2019), Risk management: Australia's sustainable finance agenda: Implications for corporate governance. *Governance Directions*, *71*(10), 551. (https://www.bis.org/fsi/publ/insights20.htm).

Klotzbach, P.J. (2006), Trends in global tropical cyclone activity over the past twenty years (1986–2005). *Geophys. Res. Lett.*, *33*(10).

Knapp, K.R., M.C. Kruk, D.H. Levinson, H.J. Diamond, and C.J. Neumann (2010), The international best track archive for climate stewardship (IBTrACS) unifying tropical cyclone data. *Bull. Amer. Meteor. Soc.*, *91*, 363-376.

Knutson, T.R., and R.E. Tuleya (2004), Impact of CO2-induced warming on simulated hurricane intensity and precipitation: Sensitivity to the choice of climate model and convective parameterization. *J. Climate*, *17*, 3477-3495.

Knutson, T.R., J.L. McBride, J. Chan, K. Emanuel, G. Holland, C. Landsea, I. Held, J.P. Kossin, A.K. Srivastava, and M. Sugi (2010), Tropical cyclones and climate change. *Nat. Geosci.*, *3*(3), 157-163.

Knutson, T.R., J.J. Sirutis, G.A. Vecchi, S. Garner, M. Zhao, H.S. Kim, M. Bender, R.E. Tuleya, I.M. Held, and G. Villarini (2013), Dynamical downscaling projections of twenty-first-century Atlantic hurricane activity: CMIP3 and CMIP5 model-based scenarios. *J. Climate*, *26*(17), 6591-6617.

Knutson, T. R., J. J. Sirutis, M. Zhao, R. E. Tuleya, M. Bender, G. A. Vecchi, G. Villarini, and D. Chavas (2015), Global projections of intense tropical cyclone activity for the late twenty-first century from dynamical downscaling of CMIP5/ RCP4.5 scenarios. *J. Climate*, 28, 7203–7224.

Knutson, T., S.J. Camargo, J.C. Chan, K. Emanuel, C.H. Ho, J. Kossin, M. Mohapatra, M. Satoh, M. Sugi, K. Walsh, and L. Wu (2019a), Tropical Cyclones and Climate Change Assessment: Part I: Detection and Attribution. *Bull. Amer. Meteor. Soc.*, *100*(10), 1987-2007.

Knutson, T., S.J. Camargo, J.C. Chan, K. Emanuel, C.H. Ho, J. Kossin, M. Mohapatra, M. Satoh, M. Sugi, K. Walsh, and L. Wu (2019b), Tropical cyclones and climate change assessment: Part II. Projected response to anthropogenic warming. *Bull. Amer. Meteor. Soc.*

Komaromi, W.A. (2013), An investigation of composite dropsonde profiles for developing and nondeveloping tropical waves during the 2010 PREDICT field campaign. *J. Atmos. Sci.*, *70*(2), 542-558.

Korty, R.L., S.J. Camargo, and J. Galewsky (2012), Tropical cyclone genesis factors in simulations of the Last Glacial Maximum. *J. Climate*, *25*(12), 4348-4365.

Korty, R.L., K.A. Emanuel, M. Huber, and R.A. Zamora (2017), Tropical cyclones downscaled from simulations with very high carbon dioxide levels. *J. Climate*, *30*(2), 649-667.

Kossin, J.P., K.A. Emanuel, and S.J. Camargo (2016), Past and projected changes in western North Pacific tropical cyclone exposure. *J. Climate*, *29*(16), 5725-5739.

Kossin, J.P., K.R. Knapp, T.L. Olander, and C.S. Velden (2020), Global increase in major tropical cyclone exceedance probability over the past four decades. *Proc. Natl. Acad. Sci.*, 117(22), 11975-11980.

Krishnamurthy, L., G. A. Vecchi, R. Msadek, H. Murakami, A. Wittenberg, and F. Zeng (2016), Impact of strong ENSO on regional tropical cyclone activity in a high-resolution climate model. *J. Climate*, 29, 2375–2394.

Kuleshov, Y., L. Qi, R. Fawcett, and D. Jones (2008), On tropical cyclone activity in the Southern Hemisphere: Trends and the ENSO connection. *Geophys. Res. Lett.*, *35*.

Kuleshov, Y., F. Chane-Ming, L. Qi, I. Chouaibou, C. Hoareau, and F. Roux (2009), Tropical cyclone genesis in the Southern Hemisphere and its relationship with the ENSO. *Ann. Geophys.*, 6, 2523-2538.

Laing, A., and J.L. Evans (2011), Introduction to tropical meteorology. *Educational material from The COMET Program*. (https://www.meted.ucar.edu/tropical/textbook_2nd_edition).

Landsea, C. W. (1993), A climatology of intense (or major) Atlantic hurricanes. *Mon. Weather Rev.*, 121, 1703–1712.

Lawrence, R.L., and A. Wright (2001), Rule-based classification systems using classification and regression tree (CART) analysis. *Photogrammetric engineering and remote sensing*, *67*(10), 1137-1142.

Lee, C.S. (1989a), Observational analysis of tropical cyclogenesis in the western North Pacific. Part I: Structural evolution of cloud clusters. *J. Atmos. Sci.*, *46*(16), 2580-2598.

Lee, C.S. (1989b), Observational analysis of tropical cyclogenesis in the western North Pacific. Part II: Budget analysis. *J. Atmos. Sci.*, *46*(16), 2599-2616.

Lee, C.S., R. Edson, and W.M. Gray (1989), Some large-scale characteristics associated with tropical cyclone development in the North Indian Ocean during FGGE. *Mon. Weather Rev.*, *117*(2), 407-426.

Li, T. (2006), Origin of the summertime synoptic-scale wave train in the western North Pacific. *J. Atmos. Sci.*, *63(3)*, 1093-1102.

Li, T., and B. Fu (2006), Tropical cyclogenesis associated with Rossby wave energy dispersion of a preexisting typhoon. Part I: Satellite data analyses. *J. Atmos. Sci.*, *63*(5), 1377-1389.

Li, T., X. Ge, B. Wang, and Y. Zhu (2006), Tropical cyclogenesis associated with Rossby wave energy dispersion of a preexisting typhoon. Part II: Numerical simulations. *J. Atmos. Sci.*, *63*(5), 1390-1409.

Li, W., C. Yang, and D. Sun (2009), Mining geophysical parameters through decision-tree analysis to determine correlation with tropical cyclone development. *Comp. & Geosci.*, *35*(2), 309-316.

Li, H., and R.L. Sriver (2018), Tropical Cyclone Activity in the High-Resolution Community Earth System Model and the Impact of Ocean Coupling. *J. Adv. Model. Earth Syst.*, *10*, 165-186.

Liu, K.S., and J.C. Chan (2012), Interannual variation of Southern Hemisphere tropical cyclone activity and seasonal forecast of tropical cyclone number in the Australian region. *Int. J. Climatol.*, *32*(2), 190-202.

Lock, A.P. (2001), The numerical representation of entrainment in parameterizations of boundary layer turbulent mixing. *Mon. Weather Rev.*, *129*, 1148-1163.

Lock, A.P., A.R. Brown, M.R. Bush, G.M. Martin, and R.N.B. Smith (2000), A new boundary layer mixing scheme. Part I: Scheme description and single-column model tests. *Mon. Weather Rev.*, *128*, 3187-3199.

Lok, C. C. F., and J. C. L. Chan (2017), Changes of tropical cyclone landfalls in South China throughout the twenty-first century. *Climate Dyn.*, 51, 2467– 2483.

Lu, J., G.A. Vecchi, and T. Reichler (2007), Expansion of the Hadley cell under global warming. *Geophys. Res. Lett.*, *34*(6).

Manganello, J.V., K.I. Hodges, J.L. Kinter III, B.A. Cash, L. Marx, T. Jung, D. Achuthavarier, J.M. Adams, E.L. Altshuler, B. Huang, and E.K. Jin (2012), Tropical cyclone climatology in a 10-km global atmospheric GCM: toward weather-resolving climate modeling. *J. Clim.*, *25*, 3867-3893.

Manucharyan, G.E., C.M. Brierley, and A.V. Fedorov (2011), Climate impacts of intermittent upper ocean mixing induced by tropical cyclones. *J. Geophys. Res. Ocean*, *116*(C11).

McBride, J.L., and R. Zehr (1981), Observational analysis of tropical cyclone formation. Part II: Comparison of non-developing versus developing systems. *J. Atmos. Sci.*, *38*(6), 1132-1151.

Meehl, G. A., C. Covey, T. Delworth, M. Latif, B. McAvaney, J. F. B. Mitchell, R. J. Stouffer, and K. R. Taylor (2007), THE WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bull. Amer. Meteor. Soc.*, 88, 1383–1394.

Merlis, T. M., M. Zhao, and I. M. Held (2013), The sensitivity of hurricane frequency to ITCZ changes and radiatively forced warming in aquaplanet simulations. *Geophys. Res. Lett.*, 40, 4109–4114.

Merlis, T. M., W. Zhou, I. M. Held, and M. Zhao (2016), Surface temperature dependence of tropical cyclone-permitting simulations in a spherical model with uniform thermal forcing. *Geophys. Res. Lett.*, 43, 2859–2865.

Merlis, T.M., and I.M. Held (2019), Aquaplanet simulations of tropical cyclones. *Curr. Climate Change Rep.*, 5, 185–195.

Montgomery, M.T., M.E. Nicholls, T.A. Cram, and A.B. Saunders (2006), A vortical hot tower route to tropical cyclogenesis. *J. Atmos. Sci.*, *63*(1), 355-386.

Montgomery, M.T., and R.K. Smith (2010), *Tropical-cyclone formation: Theory and idealized modelling*. Naval postgraduate school Monterey CA. (https://apps.dtic.mil/dtic/tr/fulltext/u2/a566569.pdf).

Montgomery, M.T., L.L. Lussier III, R.W. Moore, and Z. Wang (2010), The genesis of Typhoon Nuri as observed during the Tropical Cyclone Structure 2008 (TCS-08) field experiment–Part 1: The role of the easterly wave critical layer. *Atmos. Chem. Phys.*, *10*(20), 9879-9900.

Montgomery, M.T., and R.K. Smith (2012), The genesis of Typhoon Nuri as observed during the Tropical Cyclone Structure 2008 (TCS08) field experiment–Part 2: Observations of the convective environment. *Atmos. Chem. Phys.*, *12*(9), 4001-4009.

Montgomery, M.T., C. Davis, T. Dunkerton, Z. Wang, C. Velden, R. Torn, S.J. Majumdar, F. Zhang, R.K. Smith, L. Bosart, and M.M. Bell (2012), The Pre-Depression Investigation of Cloud-Systems in the Tropics (PREDICT) experiment: Scientific basis, new analysis tools, and some first results. *Bull. Am. Meteorol. Soc.*, *93*, 153-172.

Mountrakis, G., J. Im, and C. Ogole (2011), Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(3), 247-259.

Murakami, H., and M. Sugi, (2010), Effect of model resolution on tropical cyclone climate projections. *Sola*, *6*, 73-76.

Murakami, H., B. Wang, and A. Kitoh (2011), Future change of western North Pacific typhoons: Projections by a 20-km-mesh global atmospheric model. *J. Climate*, *24*, 1154-1169.

Murakami, H., R. Mizuta, and E. Shindo (2012), Future changes in tropical cyclone activity projected by multi-physics and multi-SST ensemble experiments using the 60-km-mesh MRI-AGCM. *Climate Dyn.*, *39*(9-10), 2569-2584.

Murakami, H., Y. Wang, H. Yoshimura, R. Mizuta, M. Sugi, E. Shindo, Y. Adachi, S. Yukimoto, M. Hosaka, S. Kusunoki, and T. Ose (2012a), Future changes in tropical cyclone activity projected by the new high-resolution MRI-AGCM. *J. Climate*, *25*, 3237-3260.

Murakami, H., M. Sugi, and A. Kitoh (2013a), Future changes in tropical cyclone activity in the North Indian Ocean projected by high-resolution MRI-AGCMs. *Climate Dyn.*, *40*, 1949-1968.

Murakami, H., T. Li, and M. Peng (2013b), Changes to environmental parameters that control tropical cyclone genesis under global warming. *Geophys. Res. Lett.*, *40*(10), 2265-2270.

Murakami, H., P.C. Hsu, O. Arakawa, and Y. Li (2014a), Influence of model biases on projected future changes in tropical cyclone frequency of occurrence. *J. Climate*, 27(5), 2159-2181.

Murakami, H. (2014), Tropical cyclones in reanalysis data sets, *Geophys. Res. Lett.*, 41, 2133–2141.

Murthy, V.S., and W. R. Boos (2018), Role of surface enthalpy fluxes in idealized simulations of tropical depression spinup. *J. Atmos. Sci.*, *75*, 1811-1831.

Murty, M.N., and R. Raghava (2016), *Support Vector Machines and Perceptrons: Learning, Optimization, Classification, and Application to Social Networks*. Springer.

Nakamura, J., and Coauthors, (2017), Western North Pacific tropical cyclone model tracks in present and future climates. *J. Geophys. Res. Atmos.*, 122, 9721–9744.

Nolan, D. S. (2007), What is the trigger for tropical cyclogenesis? *Aust. Meteorol. Mag.*, *56*, 241-266.

Nolan, D.S., E.D. Rappin, and K.A. Emanuel (2007), Tropical cyclogenesis sensitivity to environmental parameters in radiative–convective equilibrium. *Q. J. R. Meteorol. Soc.*, *133*, 2085-2107.

Nolan, D.S., and E.D. Rappin (2008), Increased sensitivity of tropical cyclogenesis to wind shear in higher SST environments *Geophys. Res. Lett., 35*.

Oouchi, K., J. Yoshimura, H. Yoshimura, R. Mizuta, S. Kusunoki, and A. Noda (2006), Tropical cyclone climatology in a global-warming climate as simulated in a 20 km-mesh global atmospheric model: Frequency and wind intensity analyses. *J. Meteorol. Soc. Jap. Ser. II*, *84*, 259-276.

Ooyama, K.V. (1982), Conceptual evolution of the theory and modeling of the tropical cyclone. *J. Meteorol. Soc. Jpn., II, 60*(1), 369-380.

Park, D.S.R., C.H. Ho, J.C. Chan, K.J. Ha, H.S. Kim, J. Kim, and J.H. Kim (2017), Asymmetric response of tropical cyclone activity to global warming over the North Atlantic and western North Pacific from CMIP5 model projections. *Sci. Rep.*, *7*, 41354.

Park, M.S., M. Kim, M.I. Lee, J. Im, and S. Park, (2016), Detection of tropical cyclone genesis via quantitative satellite ocean surface wind pattern and intensity analyses using decision trees. *Rem. Sens. Environ.*, *183*, 205-214.

Patricola, C.M., R. Saravanan, and P. Chang (2014), The impact of the El Niño-Southern Oscillation and Atlantic Meridional Mode on seasonal Atlantic tropical cyclone activity. *J. Climate*, 27, 5311–5328.

Patricola, C.M., P. Chang, and R. Saravanan (2016), Degree of simulated suppression of Atlantic tropical cyclones modulated by flavour of El Niño. *Nat. Geosci.*, 9, 155–160.

Patricola, C.M., and M.F. Wehner (2018), Anthropogenic influences on major tropical cyclone events. *Nature*, 563, 339–346.

Patricola, C.M., S.J. Camargo, P. Klotzbach, R. Saravanan, and P. Chang (2018), The influence of ENSO flavors on western North Pacific tropical cyclones. *J. Climate*, 31, 5395-5416.

Peng, M.S., B. Fu, T. Li, and D.E. Stevens (2012), Developing versus nondeveloping disturbances for tropical cyclone formation. Part I: North Atlantic. *Mon. Weather Rev.*, *140*(4), 1047-1066.

Perrone, T.J., and P.R. Lowe (1986), A statistically derived prediction procedure for tropical storm formation. *Mon. Weather Rev.*, *114*(1), 165-177.

Plutchak, J., S.E. Hall, D. Wojtowicz, M. Sridhar, M. Ramamurthy, and R.B. Wilhelmson (2010), P1. 2 Weather World, A Customizable, User-Oriented WWW Site. (http://ww2010.atmos.uiuc.edu/(Gh)/guides/mtr/hurr/stages/home.rxml).

Powers, D.M. (2011), Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *J. Mach. Learn. Technol.*, 2, 37–63.

Puterman, M.L. (1994), Markov decision processes: discrete stochastic dynamic programming. *John Wiley & Sons*.

Quinlan J. (1987), Decision trees as probabilistic classifiers. Proc. Fourth Int. Workshop on Machine Learning, Irvine, CA, International Machine Learning Society, 31–37.

Quinlan, J. (1993), *C4.5: Programs for Machine Learning*. Morgan Kaufmann.

Raavi, P.H., and K.J.E. Walsh (2020a), Sensitivity of tropical cyclone formation to resolution-dependent and independent tracking schemes in high-resolution climate model simulations. *Earth and Space Science*, 7(3). (https://doi.org/10.1029/2019EA000906).

Raavi, P.H., and K.J.E. Walsh (2020b), Basinwise statistical analysis of factors limiting tropical storm formation from an initial tropical circulation. *Journal of Geophysical Research: Atmospheres*. (https://doi.org/10.1029/2019JD032006).

Ramsay, H.A., L.M. Leslie, P.J. Lamb, M.B. Richman, and M. Leplastrier (2008), Interannual variability of tropical cyclones in the Australian region: Role of large-scale environment. *J. Climate*, *21*(5), 1083-1103.

Ramsay, H.A. and A.H. Sobel (2011), Effects of relative and absolute sea surface temperature on tropical cyclone potential intensity using a single-column model. *J. Climate*, *24*(1), 183-193.

Rappin, E.D., D.S. Nolan, and K.A. Emanuel (2010), Thermodynamic control of tropical cyclogenesis in environments of radiative-convective equilibrium with shear. *Q. J. Roy. Meteorol. Soc.*, *136*(653), 1954-1971.

Raymond, D.J., and S.L. Sessions (2007), Evolution of convection during tropical cyclogenesis. *Geophys. Res. Lett.*, *34*(6).

Raymond, D.J., S.L. Sessions, and C. López Carrillo (2011), Thermodynamics of tropical cyclogenesis in the northwest Pacific. *J. Geophys. Res. Atmos*, *116*(D18).

Reed, K. A., J. T. Bacmeister, N. A. Rosenbloom, M. F. Wehner, S. C. Bates, P. H. Lauritzen, J. E. Truesdale, and C. Hannay (2015), Impact of the dynamical core on the direct simulation of tropical cyclones in a high-resolution global model. *Geophys. Res. Lett.*, 42, 3603–3608.

Reed, K.A. and D.R. Chavas (2015), Uniformly rotating global radiative-convective equilibrium in the Community Atmosphere Model, version 5. *J. Adv. Model. Earth Syst.*, *7*(4), 1938-1955.

Regnier, E., and P.A. Harr (2006), A dynamic decision model applied to hurricane landfall. *Weather Forecast.*, *21*(5), 764-780.

Ren, F., J. Liang, G. Wu, W. Dong, and X. Yang (2011), Reliability analysis of climate change of tropical cyclone activity over the western North Pacific. *J. Climate*, *24*, 5887–5898.

Reyes, A., L.E. Sucar, E.F. Morales, and P.H. Ibargüengoytia (2006), Solving hybrid Markov decision processes. In *Mexican International Conference on Artificial Intelligence*, 227-236.

Reynolds, C.A., J. Teixeira, and J.G. McLay (2008), Impact of stochastic convection on the ensemble transform. *Mon. weather Rev.*, *136*(11), 4517-4526.

Riemer, M., M.T. Montgomery, and M.E. Nicholls (2010), A new paradigm for intensity modification of tropical cyclones: thermodynamic impact of vertical wind shear on the inflow layer. *Atmos. Chem. Phys.*, *10*(7).

Ritchie, E.A., and G.J. Holland (1997), Scale interactions during the formation of Typhoon Irving. *Mon. Weather Rev.*, *125*(7), 1377-1396.

Ritchie, E.A., and G.J. Holland (1999), Large-scale patterns associated with tropical cyclogenesis in the western Pacific. *Mon. Weather Rev.*, *127*(9), 2027-2043.

Ritchie, E.A., and W.M. Frank (2007), Interactions between simulated tropical cyclones and an environment with a variable Coriolis parameter. *Mon. Weather Rev.*, *135*(5), 1889-1905.

Roberts, M. J., and Coauthors (2015), Tropical cyclones in the UPSCALE ensemble of high-resolution global climate models. *J. Climate*, 28, 574–596.

Rogers, R., S. Aberson, M. Black, P. Black, J. Cione, P. Dodge, J. Dunion, J. Gamache, J. Kaplan, M. Powell, and N. Shay (2006), The intensity forecasting experiment: A NOAA multiyear field program for improving tropical cyclone intensity forecasts. *Bull. Am. Meteorol. Soc.*, *87*(11), 1523-1538.

Roy, C., and R. Kovordányi (2012), Tropical cyclone track forecasting techniques―A review. *Atmos. Res.*, *104*, 40-69.

Rozoff, C.M., W.H. Schubert, B.D. McNoldy, and J.P. Kossin (2006), Rapid filamentation zones in intense tropical cyclones. *J. Atmos. Sci.*, *63*(1), 325-340.

Russell, J.O., A. Aiyyer, J.D. White, and W. Hannah (2017), Revisiting the connection between African Easterly Waves and Atlantic tropical cyclogenesis. *Geophy. Res. Lett.*, *44*(1), 587-595.

Rutherford, B., T.J. Dunkerton, and M.T. Montgomery (2015), Lagrangian vortices in developing tropical cyclones. *Q. J. Roy. Meteorol. Soc.*, *141*(693), 3344-3354.

Rutherford, B., M.A. Boothe, T.J. Dunkerton, and M.T. Montgomery (2018), Dynamical properties of developing tropical cyclones using Lagrangian flow topology. *Q. J. Roy. Meteorol. Soc.*, *144*(710), 218-230.

Sampe, T., and S.P. Xie (2007), Mapping high sea winds from space. *Bull. Amer. Meteorol. Soc.*, 88, 1965–1978.

Satoh, M., Y. Yamada, M. Sugi, C. Kodama, and A.T. Noda (2015), Constraint on future change in global frequency of tropical cyclones due to global warming. *J. Meteorol. Soc. Jpn., II*, *93*(4), 489-500.

Schenkel, B. A., N. Lin, D. Chavas, G. A. Vecchi, M. Oppenheimer, and A. Brammer (2018), Lifetime evolution of outer tropical cyclone size and structures as diagnosed from reanalysis and climate model data. *J. Climate*, 31(19), 7985-8004.

Schreck III, C.J., K.R. Knapp, and J.P. Kossin (2014), The impact of best track discrepancies on global tropical cyclone climatologies using IBTrACS. *Mon. Wea. Rev.*, *142*, 3881-3899.

Schumacher, A.B., M. DeMaria, and J.A. Knaff (2009), Objective estimation of the 24-h probability of tropical cyclone formation. *Weather Forecast.*, *24*(2), 456-471.

Scoccimarro, E., S. Gualdi, A. Bellucci, A. Sanna, P. Giuseppe Fogli, E. Manzini, M. Vichi, P. Oddo, and A. Navarra (2011), Effects of tropical cyclones on ocean heat transport in a high-resolution coupled general circulation model. *J. Climate* 24(16), 4368–4384.

Sears, J., and C.S. Velden (2014), Investigating the role of the upper-levels in tropical cyclone genesis. *Tropical Cyclone Res. & Rev.*, *3*(2), 91-110.

Shaevitz, D.A., S.J. Camargo, A.H. Sobel, J.A. Jonas, D. Kim, A. Kumar, T.E. LaRow, Y.K. Lim, H. Murakami, K.A. Reed, and M.J. Roberts (2014), Characteristics of tropical cyclones in high-resolution models in the present climate. *J. Adv. Model. Earth Syst.*, *6*, 1154-1172.

Shaman, J., S.K. Esbensen, and E.D. Maloney (2009), The dynamics of the ENSO–Atlantic hurricane teleconnection: ENSO-related changes to the North African–Asian jet affect Atlantic basin tropical cyclogenesis. *J. Climate*, *22*, 2458-2482.

Sharmila, S., and K.J.E. Walsh (2017), Impact of large-scale dynamic versus thermodynamic climate conditions on contrasting tropical cyclone genesis frequency. *J. Climate*, *30*, 8865-8883.

Sharmila, S., and K.J.E. Walsh (2018), Recent poleward shift of tropical cyclone formation linked to Hadley cell expansion. *Nat. Clim. Change,* 8, 730-736.

Sharmila, S., K.J.E. Walsh, M. Thatcher, S. Wales, and S. Utembe (2020), Real world and tropical cyclone world. Part I: high-resolution climate model verification. *J. Clim.*, *33*, 1455-1472.

Shi, X., and C. S. Bretherton (2014), Large-scale character of an atmosphere in rotating radiative-convective equilibrium. *J. Adv. Model. Earth Syst.*, 6, 616–629.

Shu, S., F. Zhang, J. Ming, and Y. Wang (2014), Environmental influences on the intensity changes of tropical cyclones over the western North Pacific. *Atmos. Chem. Phys.*, *14*(12), 6329-6342.

Sinclair, M.R. (2004), Extratropical transition of southwest Pacific tropical cyclones. Part II: Midlatitude circulation characteristics. *Mon. Weather Rev, 132,* 2145-2168.

Slingo, J., K. Bates, N. Nikiforakis, M. Piggott, M. Roberts, L. Shaffrey, I. Stevens, P.L. Vidale, and H. Weller (2009), Developing the next-generation climate system models: challenges and achievements. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* 367(1890), 815–831.

Smith, R.K., and M.T. Montgomery (2012), Observations of the convective environment in developing and non-developing tropical disturbances. *Q. J. Roy. Meteorol. Soc., 138*(668), 1481-1739.

Snyder, A.D., Z. Pu, and Y. Zhu, (2010), Tracking and verification of east Atlantic tropical cyclone genesis in the NCEP global ensemble: Case studies during the NASA African monsoon multidisciplinary analyses. *Weather Forecast.*, 25(5), 1397-1411.

Sriver, R. L. (2010). Tropical cyclones in the mix. *Nature*, 463(25), 1032–1033.

Sriver, R.L., and M. Huber (2007), Observational evidence for an ocean heat pump induced by tropical cyclones. *Nature*, *447*(7144), 577-580.

Strachan, J., P.L. Vidale, K. Hodges, M. Roberts, and M.E. Demory (2013), Investigating global tropical cyclone activity with a hierarchy of AGCMs: The role of model resolution. *J. Clim*., *26*, 133-152.

Studholme, J., and S. Gulev (2018), Concurrent changes to Hadley circulation and the meridional distribution of tropical cyclones. *J. Climate*, 31, 4367-4389.

Sugi, M., A. Noda, and N. Sato (2002), Influence of the global warming on tropical cyclone climatology: An experiment with the JMA global model. *J. Meteorol. Soc. Jpn., II*, *80*(2), 249-272.

Sugi, M., H. Murakami, and J. Yoshimura (2012), On the mechanism of tropical cyclone frequency changes due to global warming. *J. Meteorol. Soc. Jpn., II, 90*, 397-408.

Sugi, M., Y. Yoshida, and H. Murakami, (2015), More tropical cyclones in a cooler climate? *Geophys. Res. Lett.*, 42, 6780–6784.

Sugi, M., Y. Yamada, K. Yoshida, R. Mizuta, M. Nakano, C. Kodama, and M. Satoh (2020), Future changes in the global frequency of tropical cyclone seeds. *SOLA*. (https://doi.org/10.2151/sola.2020-012).

Tang, B., and K. Emanuel (2010), Midlevel ventilation’s constraint on tropical cyclone intensity. *J. Atmos. Sci.*, *67*(6), 1817-1830.

Tang, B. and K. Emanuel (2012), A ventilation index for tropical cyclones. *Bull. Amer. Meteor. Soc.*, *93*(12), 1901-1912.

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.

Thatcher, L., and Z. Pu (2013), Evaluation of Tropical Cyclone Genesis Precursors with Relative Operating Characteristics (ROC) in High resolution Ensemble Forecasts: Hurricane Ernesto. *Tropical Cyclone Research and Review*, *2*(3), 131-148.

Thorncroft, C. D., and K. Hodges (2001), African easterly wave variability and its relationship to Atlantic tropical cyclone activity. *J Climate, 14*, 1166–1179.

Tippett, M.K., S.J. Camargo, and A.H. Sobel (2011), A Poisson regression index for tropical cyclone genesis and the role of large-scale vorticity in genesis. *J Climate*, *24*(9), 2335-2357.

Tory, K.J., and M.T. Montgomery (2006), Internal influences on tropical cyclone formation. In: Sixth International Workshop on Tropical Cyclones, San Jose, Costa Rica. World Meteorological Organization.

Tory, K.J., M.T. Montgomery, and N.E. Davidson (2006a), Prediction and diagnosis of tropical cyclone formation in an NWP system. Part I: The critical role of vortex enhancement in deep convection. *J. Atmos. Sci.*, *63*(12), 3077-3090.

Tory, K.J., M.T. Montgomery, N.E. Davidson, and J.D. Kepert (2006b), Prediction and diagnosis of tropical cyclone formation in an NWP system. Part II: A diagnosis of Tropical Cyclone Chris formation. *J. Atmos. Sci.*, *63*(12), 3091-3113.

Tory, K.J., N.E. Davidson, and M.T. Montgomery (2007), Prediction and diagnosis of tropical cyclone formation in an NWP system. Part III: Diagnosis of developing and nondeveloping storms. *J. Atmos. Sci.*, *64*(9), 3195-3213.

Tory K.J., and W.M. Frank (2010), Tropical cyclone formation. Chapter 2: Global perspectives on tropical cyclones. In: Chan J, Kepert JD (eds) World Scientific, Second Edition

Tory, K.J., R.A Dare, N.E. Davidson, J.L. McBride, and S.S. Chand (2013a), The importance of low-deformation vorticity in tropical cyclone formation. *Atmos. Chem. Phys.*, *13*(4), 2115-2132.

Tory, K.J., S.S. Chand, R.A. Dare, and J.L. McBride (2013b), The development and assessment of a model-, grid-, and basin-independent tropical cyclone detection scheme. *J. Climate, 26*, 5493-5507.

Tory, K.J., S.S. Chand, R.A. Dare, and J.L. McBride (2013c), An assessment of a model-, grid-, and basin-independent tropical cyclone detection scheme in selected CMIP3 global climate models. *J. Climate*, *26*, 5508-5522.

Tory, K.J., S.S. Chand, J.L. McBride, H. Ye, and R.A. Dare (2013d), Projected changes in late-twenty-first-century tropical cyclone frequency in 13 coupled climate models from phase 5 of the Coupled Model Intercomparison Project. *J. Climate*, *26*, 9946-9959.

Tory, K.J., H. Ye, and R.A. Dare (2018), Understanding the geographic distribution of tropical cyclone formation for applications in climate models. *Climate Dynam.*, *50*(7-8), 2489-2512.

Tsai, H.C., K.C. Lu, R.L. Elsberry, M.M. Lu, and C.H. Sui (2011), Tropical cyclone–like vortices detection in the NCEP 16-day ensemble system over the western North Pacific in 2008: Application and forecast evaluation. *Weather Forecast.*, *26*(1), 77-93.

Tsuboi, A., and T. Takemi (2014), The interannual relationship between MJO activity and tropical cyclone genesis in the Indian Ocean. *Geosci. Lett.*, *1*(1), 9.

Tuleya, R.E., and Y. Kurihara (1981), A numerical study on the effects of environmental flow on tropical storm genesis. *Mon. Weather Rev.*, *109*(12), 2487-2506.

Tuleya, R.E., and Y. Kurihara (1982), A note on the sea surface temperature sensitivity of a numerical model of tropical storm genesis. *Mon. Weather Rev.*, *110*(12), 2063-2069.

Tuleya, R.E. (1991), Sensitivity studies of tropical storm genesis using a numerical model. *Mon. Weather Rev.*, *119*(3), 481-733.

Ullrich, P.A. and C.M. Zarzycki (2017), TempestExtremes: A framework for scale-insensitive pointwise feature tracking on unstructured grids. *Geosci. Model Dev.*, *10*(3), 1069.

Vecchi, G. A., and B. J. Soden (2007a), Global warming and the weakening of the tropical circulation, *J. Climate*, 20, 4316–4340.

Vecchi, G. A., and B. J. Soden (2007b), Effect of remote sea surface temperature change on tropical cyclone potential intensity. *Nature*, 450(7172), 1066-1070.

Vecchi, G.A., T. L. Delworth, H. Murakami, S. D. Underwood, A.T. Wittenberg, F. Zeng, W. Zhang, J.W. Baldwin, K.T. Bhatia, W. Cooke, and J. He (2019), Tropical cyclone sensitivities to CO 2 doubling: roles of atmospheric resolution, synoptic variability and background climate changes. *Climate Dynam.*, *53*, 5999-6033.

Vitart, F., J.L. Anderson, and W.F. Stern (1997), Simulation of interannual variability of tropical storm frequency in an ensemble of GCM integrations. *J. Climate*, *10*, 745-760.

Vitart, F., J.L. Anderson, and W.F. Stern (1999), Impact of large-scale circulation on tropical storm frequency, intensity, and location, simulated by an ensemble of GCM integrations. *J. Climate*, *12*, 3237-3254.

Walsh, K.J.E., M. Fiorino, C.W. Landsea, and K.L. McInnes (2007), Objectively determined resolution-dependent threshold criteria for the detection of tropical cyclones in climate models and reanalysis. *J. Climate*, *20*, 2307-2314.

Walsh, K., S. Lavender, H. Murakami, E. Scoccimarro, L.P. Caron, and M. Ghantous (2010), The tropical cyclone climate model intercomparison project. In *Hurricanes and climate change*, Springer, Dordrecht, 1-24.

Walsh, K., S. Lavender, E. Scoccimarro, and H. Murakami (2013), Resolution dependence of tropical cyclone formation in CMIP3 and finer resolution models. *Climate. Dyn.*, *40*, 585-599.

Walsh, K.J., S.J. Camargo, G.A. Vecchi, A.S. Daloz, J. Elsner, K. Emanuel, M. Horn, Y.K. Lim, M. Roberts, C. Patricola, and E. Scoccimarro (2015), Hurricanes and climate: the US CLIVAR working group on hurricanes. *Bull. Am. Meteorol. Soc.*, *96*(6), 997-1017.

Walsh, K.J., J.L. McBride, P.J. Klotzbach, S. Balachandran, S.J. Camargo, G. Holland, T.R. Knutson, J.P. Kossin, T.C. Lee, A. Sobel, and M. Sugi (2016), Tropical cyclones and climate change. *Wiley Interdisciplinary Reviews: Climate Change*, *7*(1), 65-89.

Walsh, K.J., S.J. Camargo, T.R. Knutson, J. Kossin, T.C. Lee, H. Murakami, and C. Patricola (2019), Tropical cyclones and climate change. *Tropical Cyclone Research and Review*, *8*(4), 240-250.

Walsh, K.J.E., S. Sharmila, M. Thatcher, S. Wales, S. Utembe, and A. Vaughan (2020), Real World and tropical cyclone world. Part II: sensitivity of tropical cyclone formation to uniform and meridionally varying sea surface temperatures under aquaplanet conditions. *J. Climate*, 33(4).

Walters, D., M. Brooks, I. Boutle, T. Melvin, R. Stratton, S. Vosper, H. Wells, K. Williams, N. Wood, T. Allen, and A. Bushell (2017), The Met Office unified model global atmosphere 6.0/6.1 and JULES global land 6.0/6.1 configurations. *Geosci. Model Dev.*, *10*, 1487-1520.

Wang, B., and J.C. Chan (2002), How strong ENSO events affect tropical storm activity over the western North Pacific. *J. Clim.*, *15*, 1643-1658.

Wang, C.C., and G. Magnusdottir (2006), The ITCZ in the central and eastern Pacific on synoptic time scales. *Mon. Weather Rev.*, *134*(5), 1405-1421.

Wang, Z., M.T. Montgomery, and T.J. Dunkerton (2009), A dynamically-based method for forecasting tropical cyclogenesis location in the Atlantic sector using global model products. *Geophys. Res. Lett.*, *36(3)*.

Wang, Z., M.T. Montgomery, and T.J. Dunkerton (2010a), Genesis of pre–Hurricane Felix (2007). Part I: The role of the easterly wave critical layer. *J. Atmos. Sci.*, *67*(6), 1711-1489.

Wang, Z., M.T. Montgomery, and T.J. Dunkerton (2010b), Genesis of pre–Hurricane Felix (2007). Part II: Warm core formation, precipitation evolution, and predictability. *J. Atmos. Sci.*, *67*, 1730-1744.

Wang, Z., M.T. Montgomery, and C. Fritz (2012), A first look at the structure of the wave pouch during the 2009 PREDICT–GRIP dry runs over the Atlantic. *Mon. Weather Rev.*, *140*(4), 1144-1163.

Wang, Z. (2014), Role of cumulus congestus in tropical cyclone formation in a high-resolution numerical model simulation. *J. Atmos. Sci.*, *71*(5), 1681-1700.

Wei, M., Z. Toth, R. Wobus, and Y. Zhu (2008), Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. *Tellus A: Dynam. Meteorol. Oceanogr.*, *60*(1), 62-79.

Wilks, D.S. (2006), comparison of ensemble-MOS methods in the Lorenz'96 setting. *Meteorol. Appl.*, *13*(3), 243-256.

Williams, I.N., and C.M. Patricola (2018), Diversity of ENSO events unified by convective threshold sea surface temperature: a nonlinear ENSO index. *Geophys. Res. Letters,* 45, 9236-9244.

Wilson, D.R., A.C. Bushell, A.M. Kerr-Munslow, J.D. Price, and C.J. Morcrette (2008), PC2: A prognostic cloud fraction and condensation scheme. I: Scheme description. *Q. J. Roy. Meteorol. Soc.*, *134*, 2093-2107.

Wing, A.A., S.J. Camargo, and A. H. Sobel (2016), Role of radiative–convective feedbacks in spontaneous tropical cyclogenesis in idealized numerical simulations. *J. Atmos. Sci.*, *73*, 2633-2642.

Wood, N., and Coauthors (2014), An inherently mass-conserving semi-implicit semi-Lagrangian discretization of the deep atmosphere global non-hydrostatic equations. *Quart. J. Roy. Meteor. Soc.,* 140, 1505–1520.

Wu, M. C., H. Zhang, J.M. Chen, and T. Feng (2018), Impact of two types of El Niño on tropical cyclones over the western North Pacific: Sensitivity to location and intensity of Pacific warming. *J. Climate,* 31, 1725–1742.

Xu, Y., T. Li, and M. Peng (2013), Tropical cyclogenesis in the western North Pacific as revealed by the 2008–09 YOTC data. *Weather Forecast.*, *28*(4), 1038-1056.

Yang, R., J. Tang, and M. Kafatos (2007), Improved associated conditions in rapid intensifications of tropical cyclones. *Geophys. Res. Lett.*, *34*(20).

Yang, R., J. Tang, and D. Sun (2011), Association rule data mining applications for Atlantic tropical cyclone intensity changes. *Weather Forecast.*, *26*(3), 337-353.

Yang, R. (2016), A systematic classification investigation of rapid intensification of atlantic tropical cyclones with the ships database. *Weather Forecast.*, *31*(2), 495-513.

Zarzycki, C. M., and C. Jablonowski (2014), A multidecadal simulation of Atlantic tropical cyclones using a variable-resolution global atmospheric general circulation model. *J. Adv. Model. Earth Syst.*, 6, 805–828.

Zehr, R. (1992), Tropical cyclogenesis in the western North Pacific. NOAA Tech. Rep. 61,181.(https://repository.library.noaa.gov/gsearch?terms=NOAA%20Technical%20Report%20NESDIS%2061%20&ss_publishyear=1992).

Zeng, Z., Y. Wang, and L. Chen (2010), A statistical analysis of vertical shear effect on tropical cyclone intensity change in the North Atlantic. *Geophys. Res. Lett.*, *37*(2).

Zhang, D.L., and L. Zhu (2012), Roles of upper-level processes in tropical cyclogenesis. *Geophys. Res. Lett.*, *39*(17).

Zhang, W., S. Gao, S., B. Chen, and K. Cao (2013a), The application of decision tree to intensity change classification of tropical cyclones in western North Pacific. *Geophys. Res. Lett.*, *40*(9), 1883-1887.

Zhang, W., Y. Leung, and J.C. Chan (2013b), The analysis of tropical cyclone tracks in the western North Pacific through data mining. Part I: Tropical cyclone recurvature *J. Appl. Meteorol. Climatol.*, *52*(6), 1394-1416.

Zhang, W., Y. Leung, and J.C. Chan (2013c), The analysis of tropical cyclone tracks in the western North Pacific through data mining. Part II: Tropical cyclone landfall. *J. Appl. Meteorol. Climatol.*, *52*(6), 1417-1432.

Zhang, W., B. Fu, M.S. Peng, and T. Li (2015), Discriminating developing versus nondeveloping tropical disturbances in the western North Pacific through decision tree analysis. *Weather Forecast.*, *30*(2), 446-454.

Zhao, M., I.M. Held, S.J. Lin, and G.A. Vecchi (2009), Simulations of global hurricane climatology, interannual variability, and response to global warming using a 50-km resolution GCM. *J. Climate*, *22*, 6653-667.

Zhao, M., and I.M. Held (2012), TC-permitting GCM simulations of hurricane frequency response to sea surface temperature anomalies projected for the late-twenty-first century. *J. Climate*, *25*(8), 2995-3009.

Zhao, M., I.M. Held, and S.J. Lin (2012), Some counterintuitive dependencies of tropical cyclone frequency on parameters in a GCM. *J. Atmos. Sci.*, *69*, 2272-2283.

Zhao, H., and G.B. Raga (2015), On the distinct interannual variability of tropical cyclone activity over the eastern North Pacific. *Atmósfera*, *28*, 161-178.

Zhou, W., I.M. Held, and S.T. Garner (2014), Parameter study of tropical cyclones in rotating radiative–convective equilibrium with column physics and resolution of a 25-km GCM. *J. Atmos. Sci.*, 71, 1058–1069.

Zong, H., and L. Wu (2015), Re-examination of tropical cyclone formation in monsoon troughs over the western North Pacific. *Adv. Atmos. Sci.*, *32*(7), 924-934.

Appendix A

This appendix gives additional figures to support the results of the chapter 4 titled "Basinwise statistical analysis of factors limiting tropical storm formation from an initial tropical circulation."

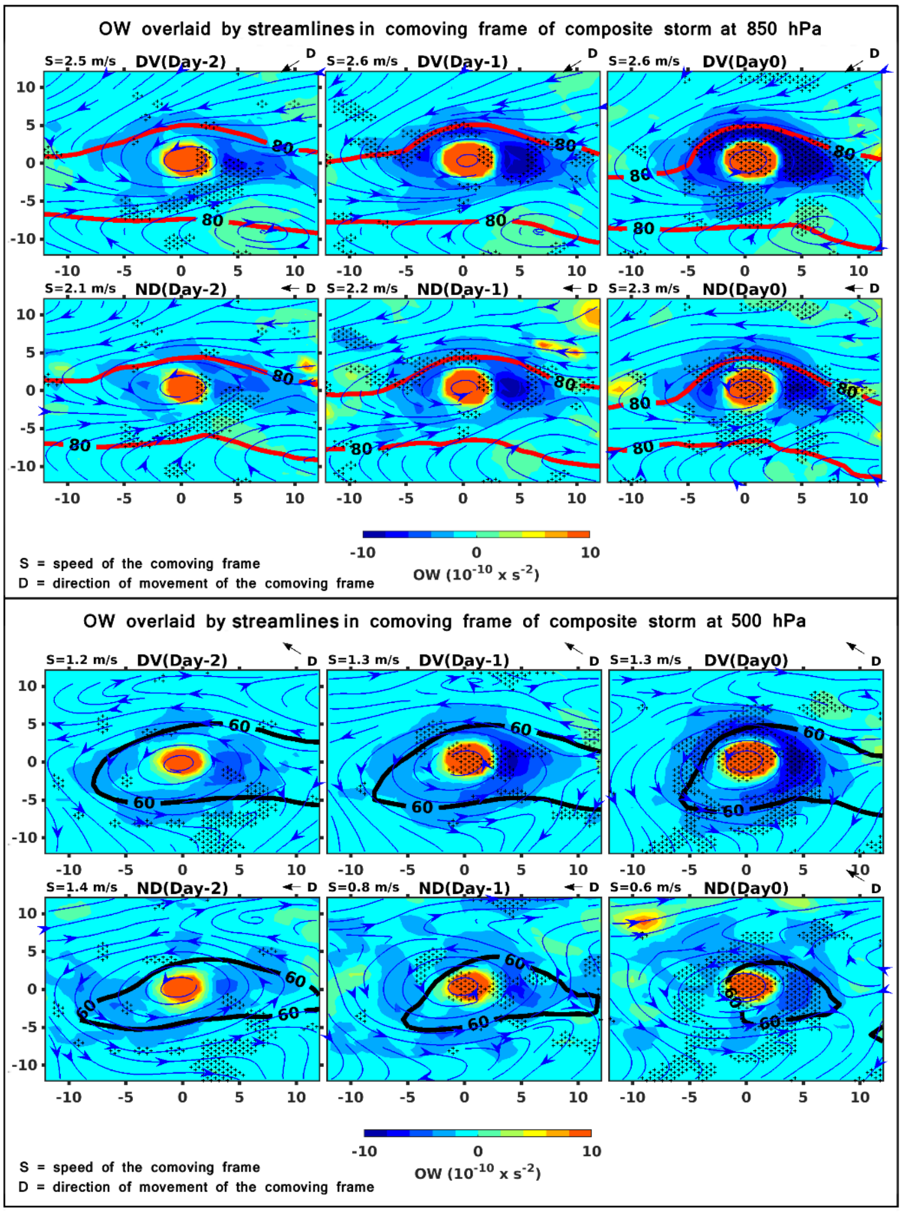

Figure A. 1. OW overlaid by streamlines computed in a co-moving frame for DV and ND cases at 850 and 500hPa over the Eastern North Pacific basin. The x and y-axis indicating the distance from the center in degrees, the red and black contours indicate the 80% and 60% RH, respectively and stippling indicates the 95% significance level differences between DV and ND cases.

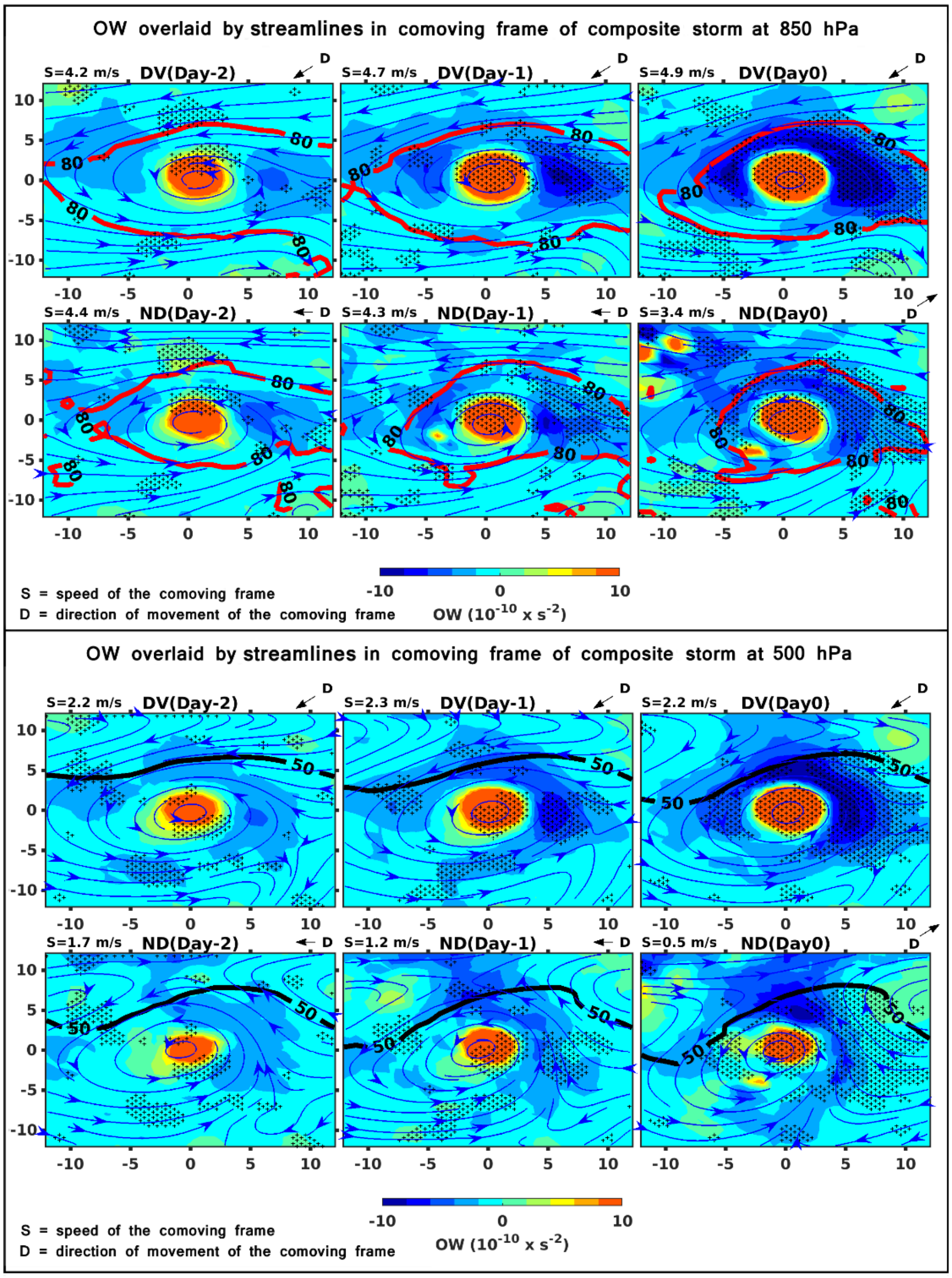

Figure A. 2. The same as Figure A.1 over the Western North Pacific (0-15N).

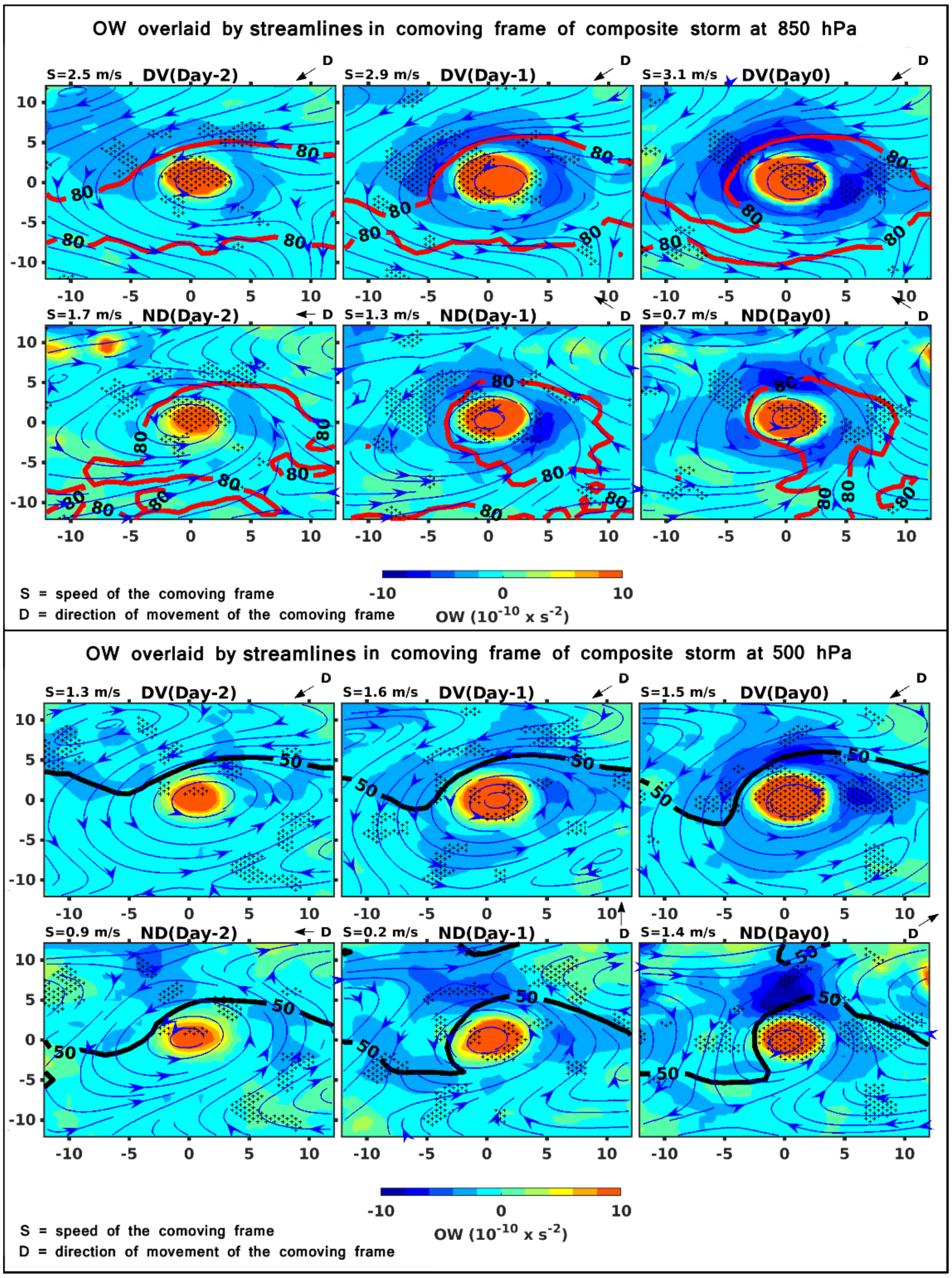

Figure A. 3. The same as Figure A.1 over the Western North Pacific poleward region (15-30N).

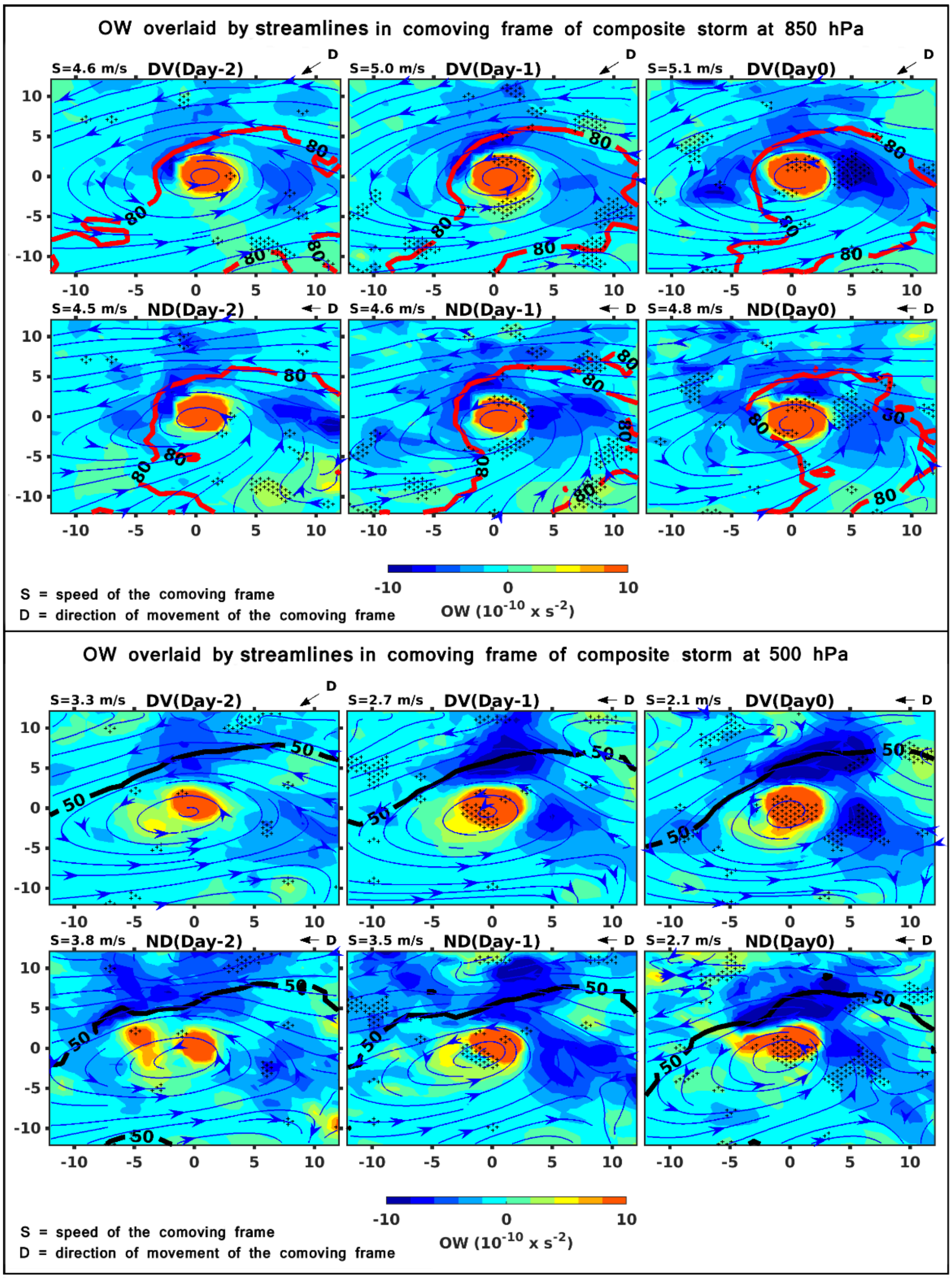

Figure A. 4. The same as Figure A.1 over the North Indian region.

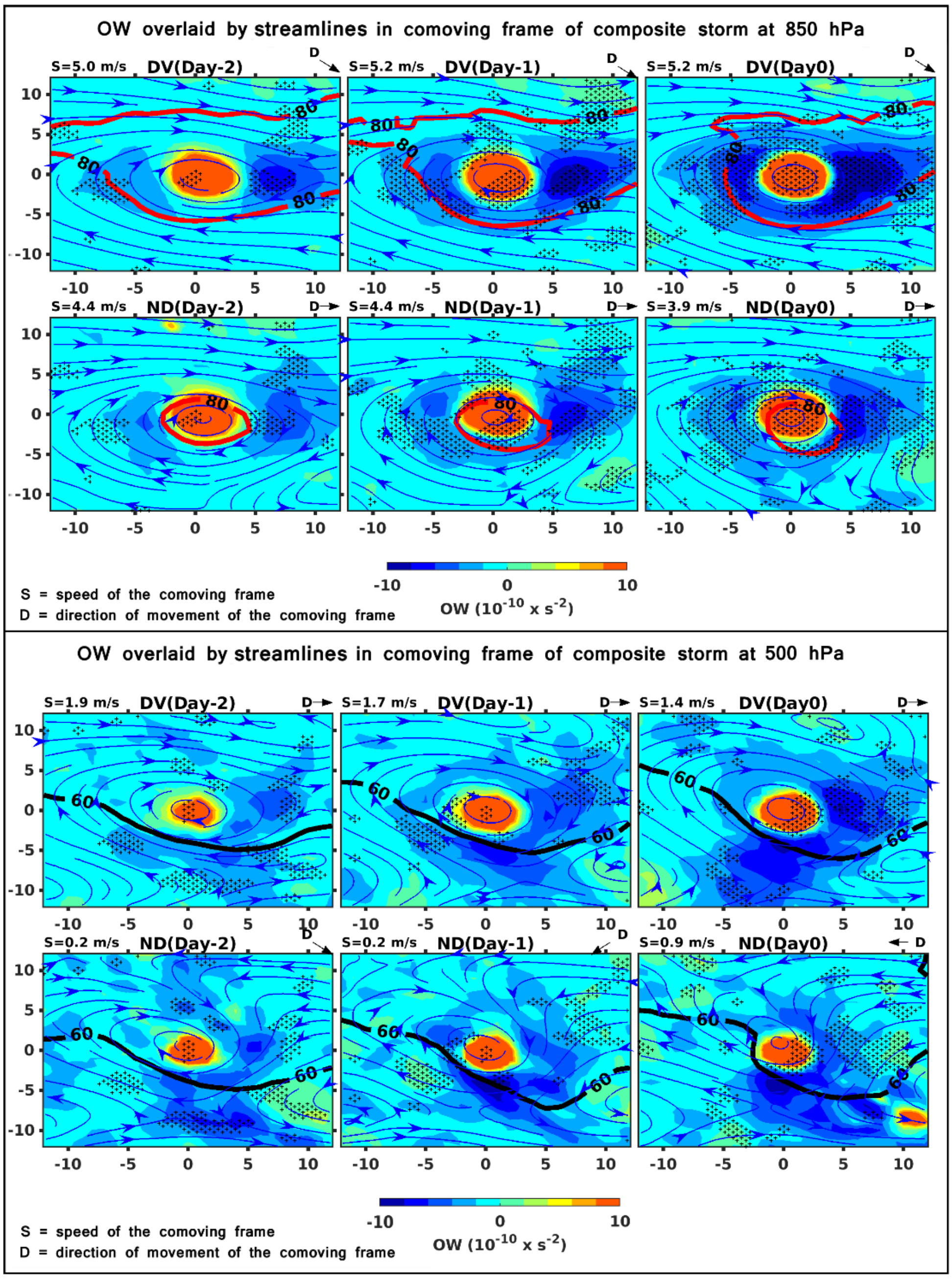

Figure A. 5. The same as Figure A.1 over the South Indian basin.

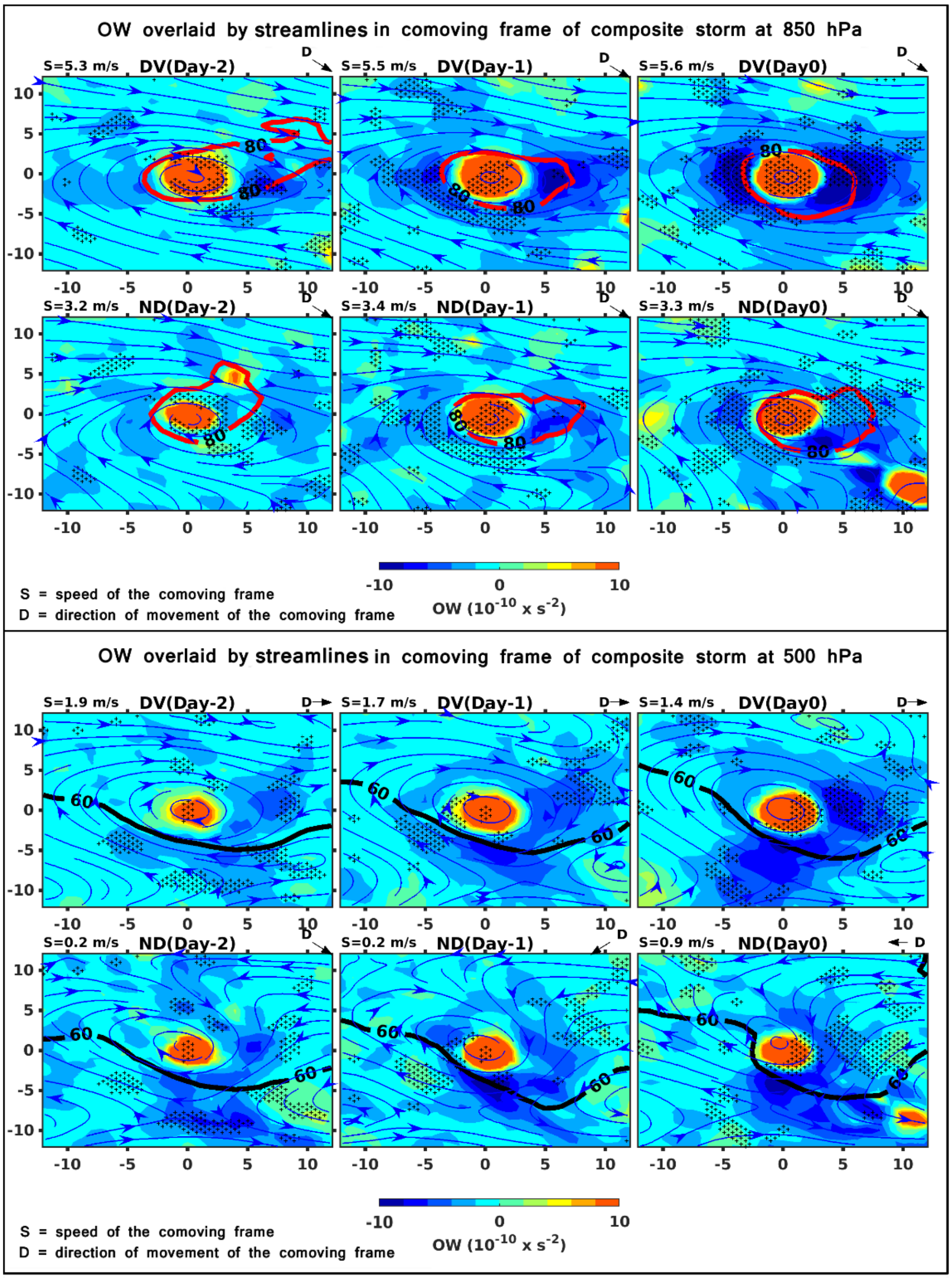

Figure A. 6. The same as Figure A.1 over the Australian basin.

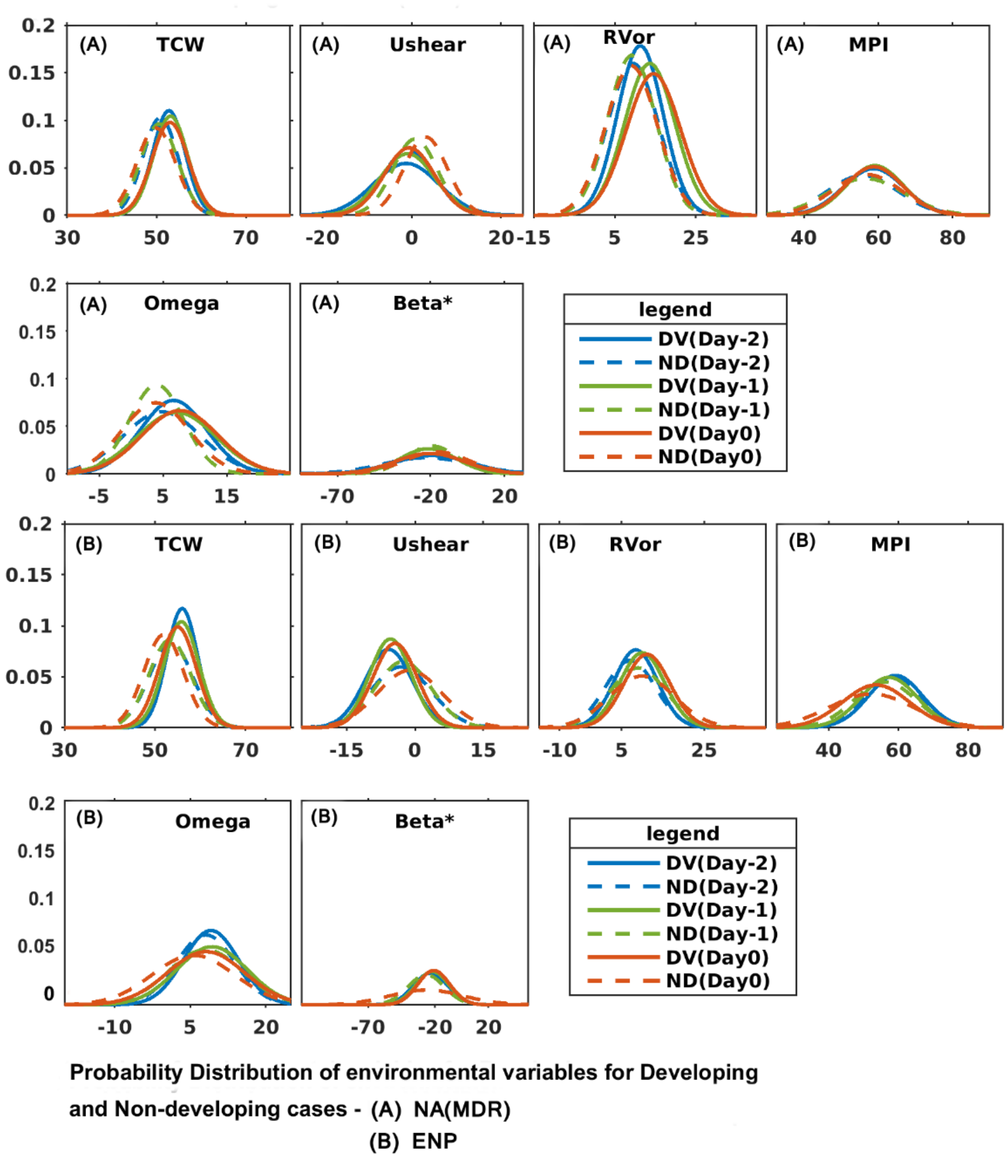

Figure A. 7. Statistical distributions of large-scale variables in the North Atlantic (A) and Eastern North Pacific (B) basins, on (Day-2), (Day-1) and (Day0) days before TD declaration. DV indicates developing tropical depressions, and ND indicates non-developing tropical depressions.

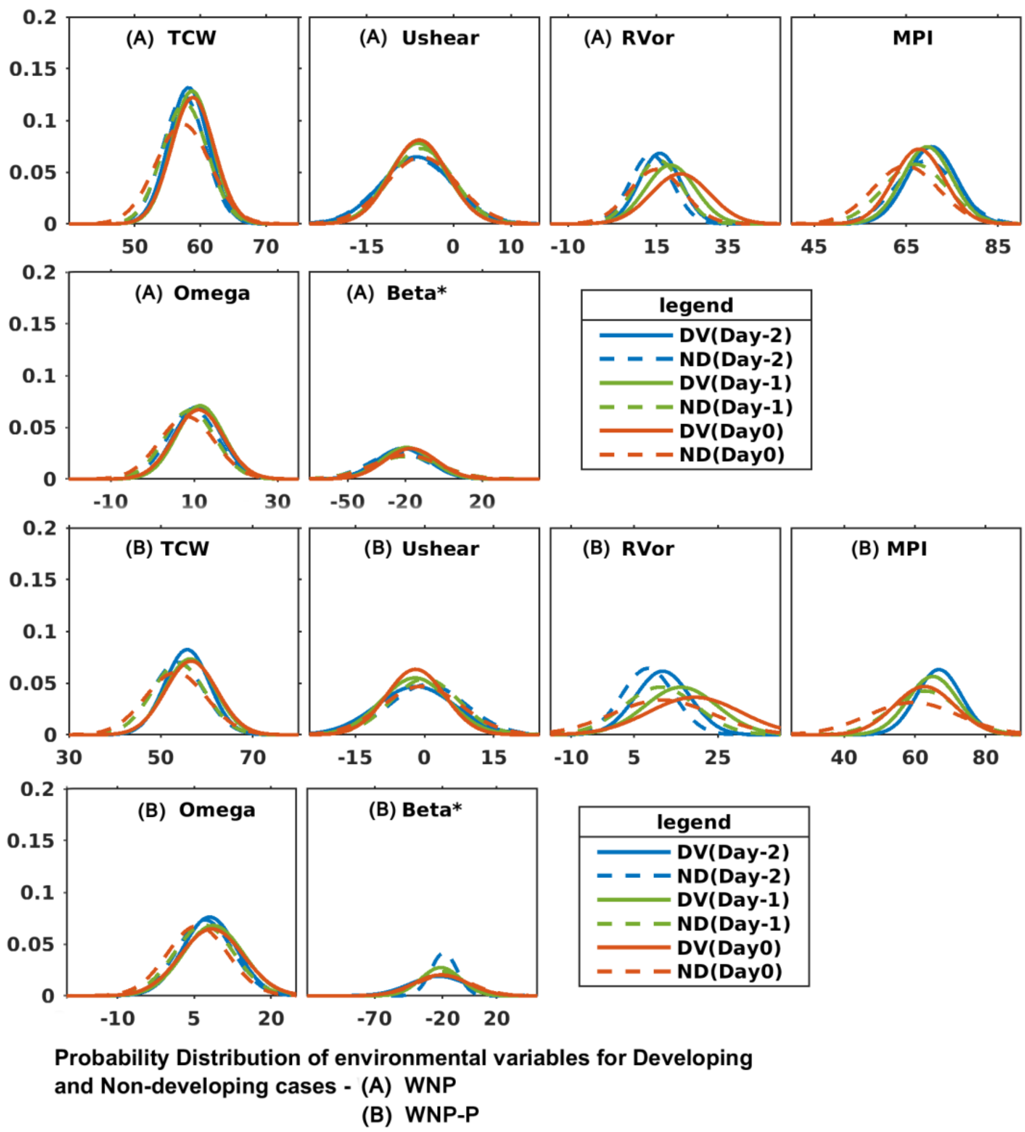

Figure A. 8. The same as Fig. A.7 but for (A) the Western North Pacific equatorward region (0-15N) and (B) Western North Pacific poleward region (15-30N).

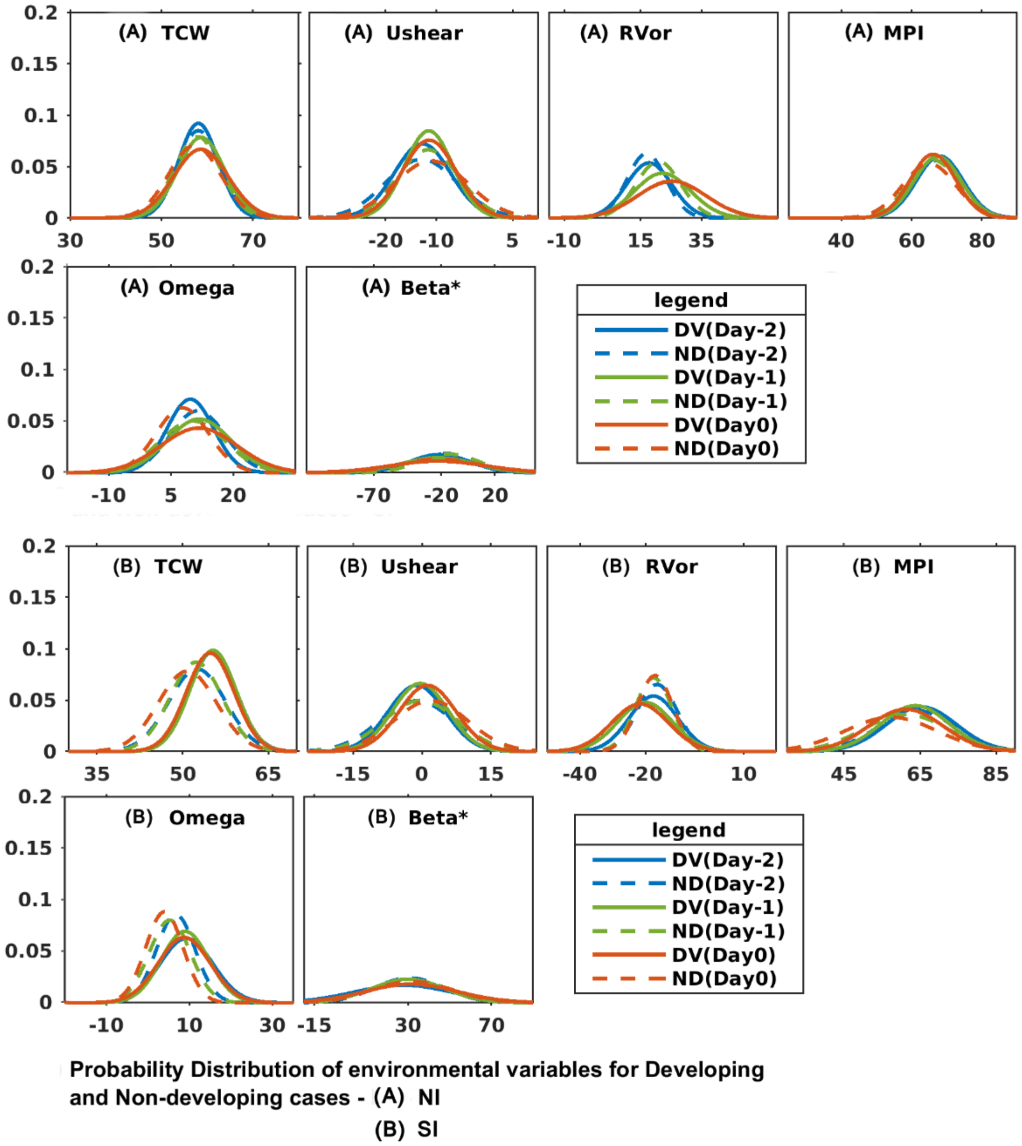

Figure A. 9. The same as Figure A.7 but for (A) the North Indian and (B) South Indian basins.

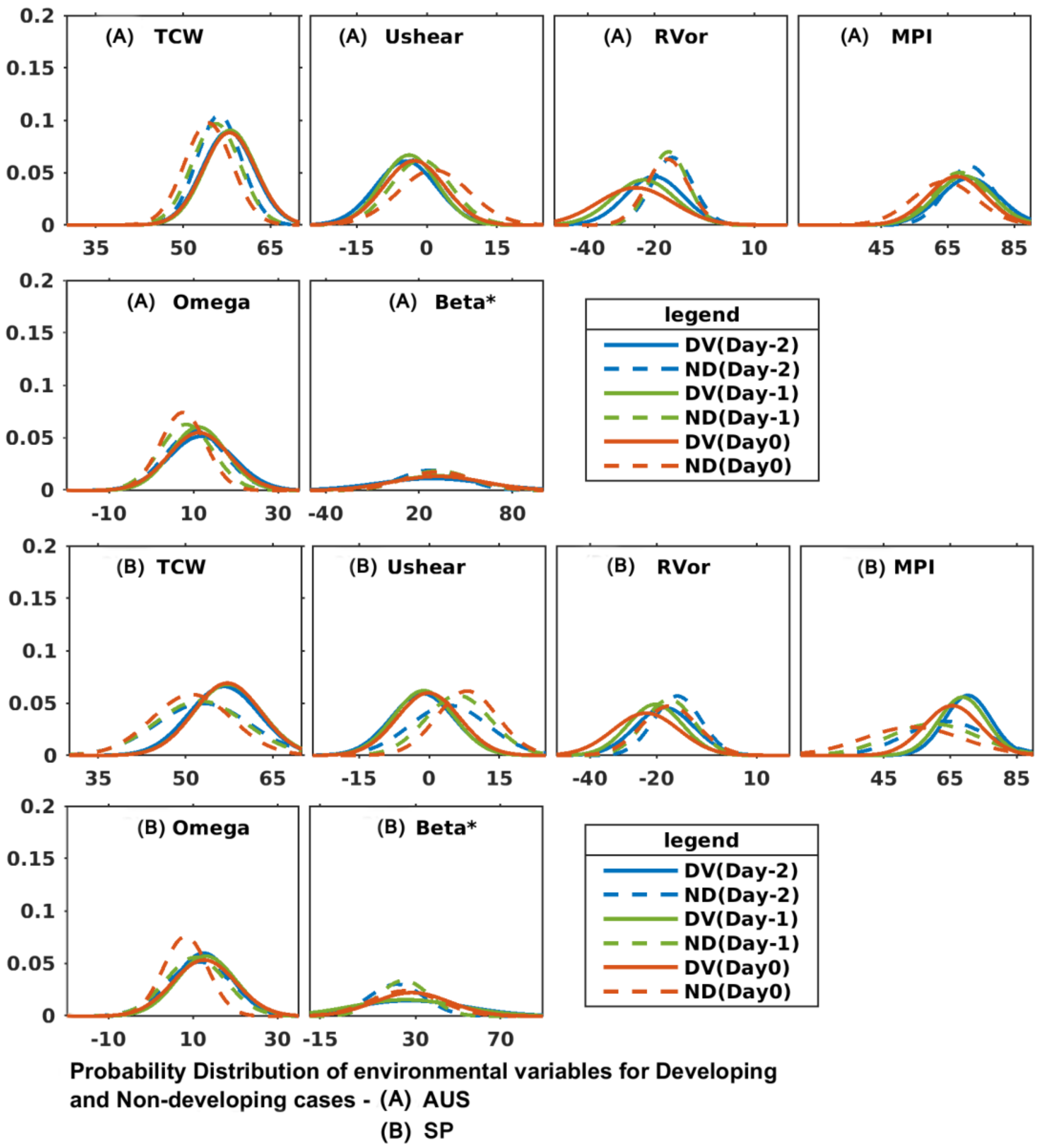

Figure A. 10. The same as Figure A.7 but for (A) the Australian and (B) South Pacific basins.

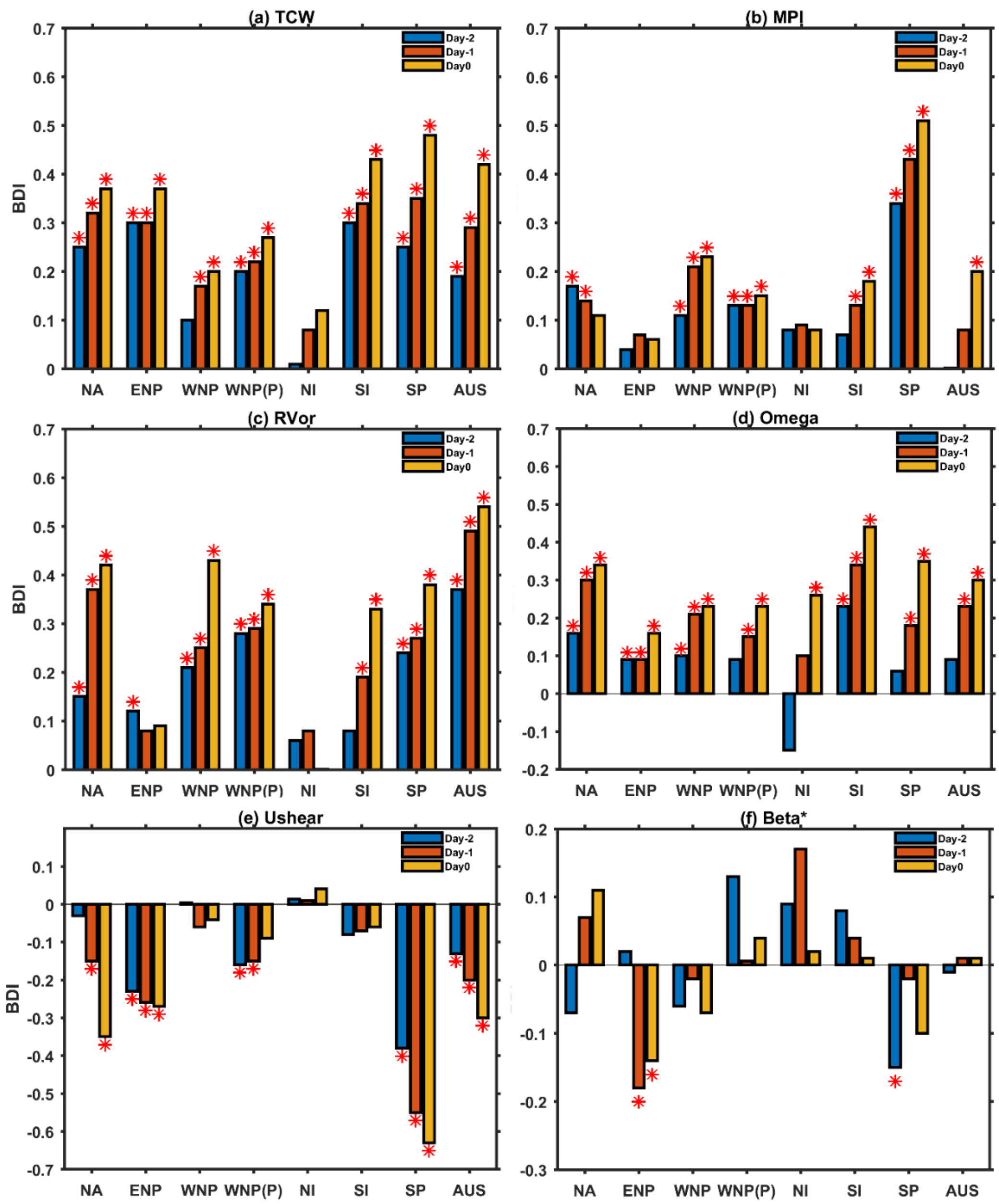

Figure A. 11. The same as Figure 4.9 (chapter 4) except that the differences between developing and non-developing systems are arranged by variable instead of the basin.

Appendix B

This appendix gives additional figures and tables to support the results of the chapter 6 titled "Sensitivity of tropical cyclone formation to resolution-dependent and independent tracking schemes in high-resolution climate model simulations."

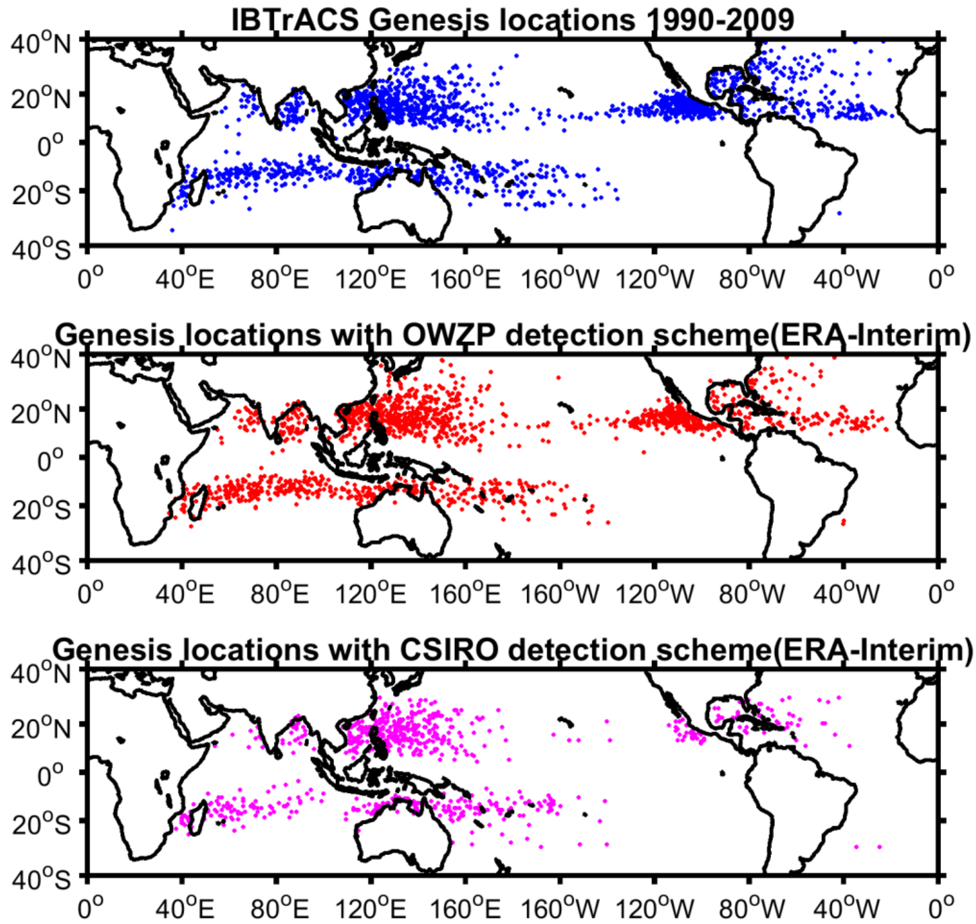

Figure B. 1. TC genesis locations in observations and the ERA-Interim reanalysis (using both CSIRO and OWZP schemes) for a period of 20 years from 1990 to 2009.

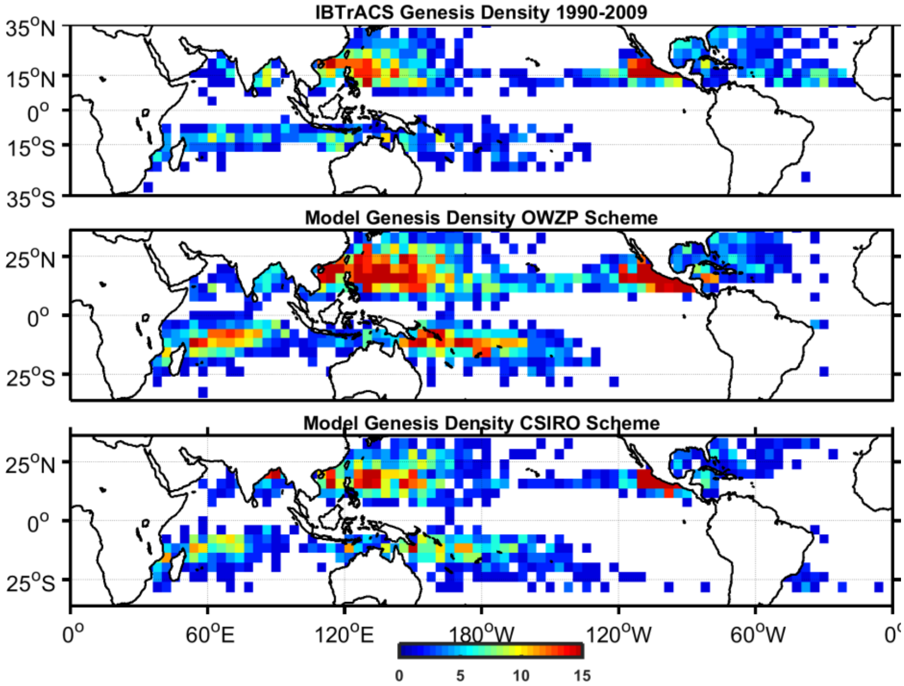

Figure B. 2. Genesis density per 4-degree box region across the globe in observations (IBTrACS) and in ACCESS model using both the CSIRO and OWZP detection schemes for a period of 20 years from 1990-2009.

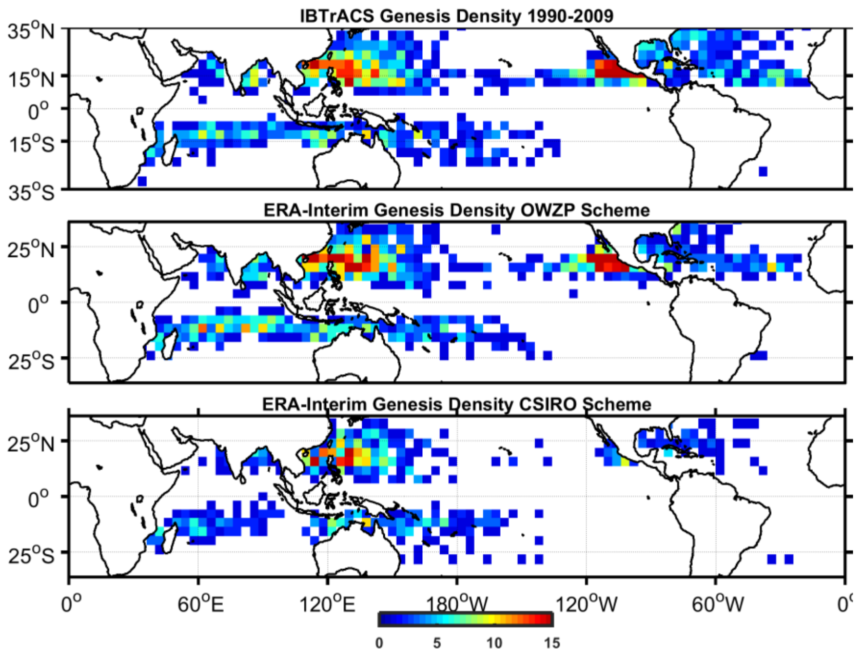

Figure B. 3. Genesis density per 4-degree box region across the globe using both the CSIRO and OWZP detection schemes with ERA-Interim reanalysis data for a period of 20 years from 1990-2009.

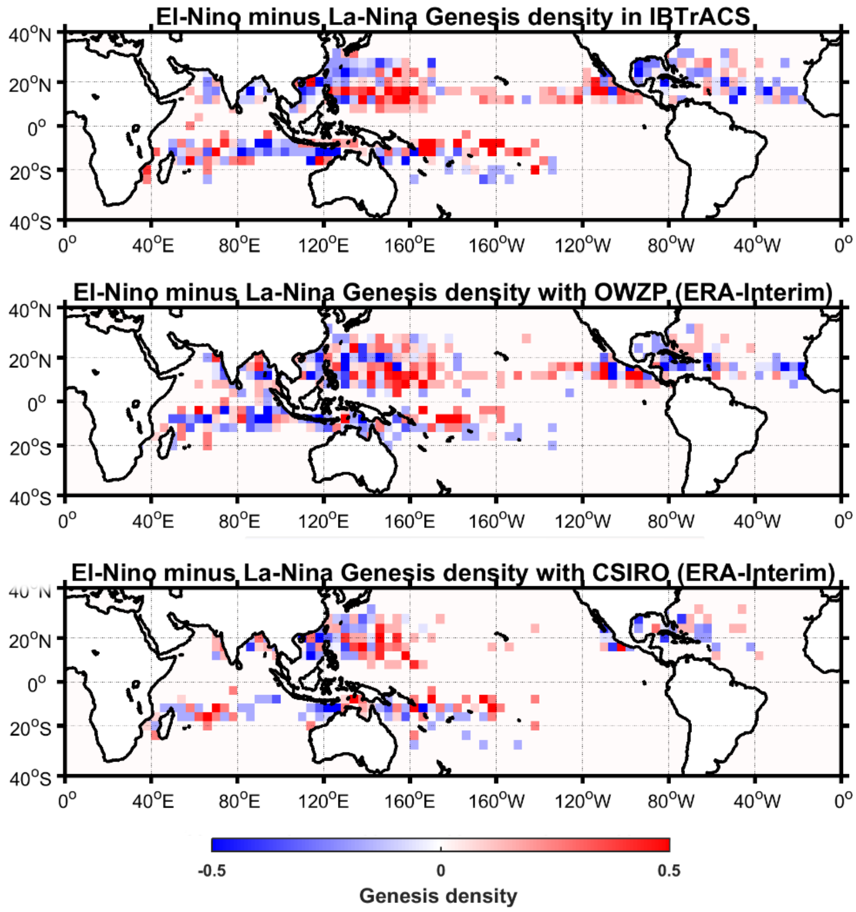

Figure B. 4. Geographical distribution of the difference in the TC genesis density of El-Nino and La-Nina years, with genesis density being the number of TCs per $4^0 \times 4^0$ box region across the globe, for (top) IBTrACS; (center) ERA-Interim reanalysis with OWZP detection scheme; and (bottom) ERA-Interim reanalysis with CSIRO detection scheme.

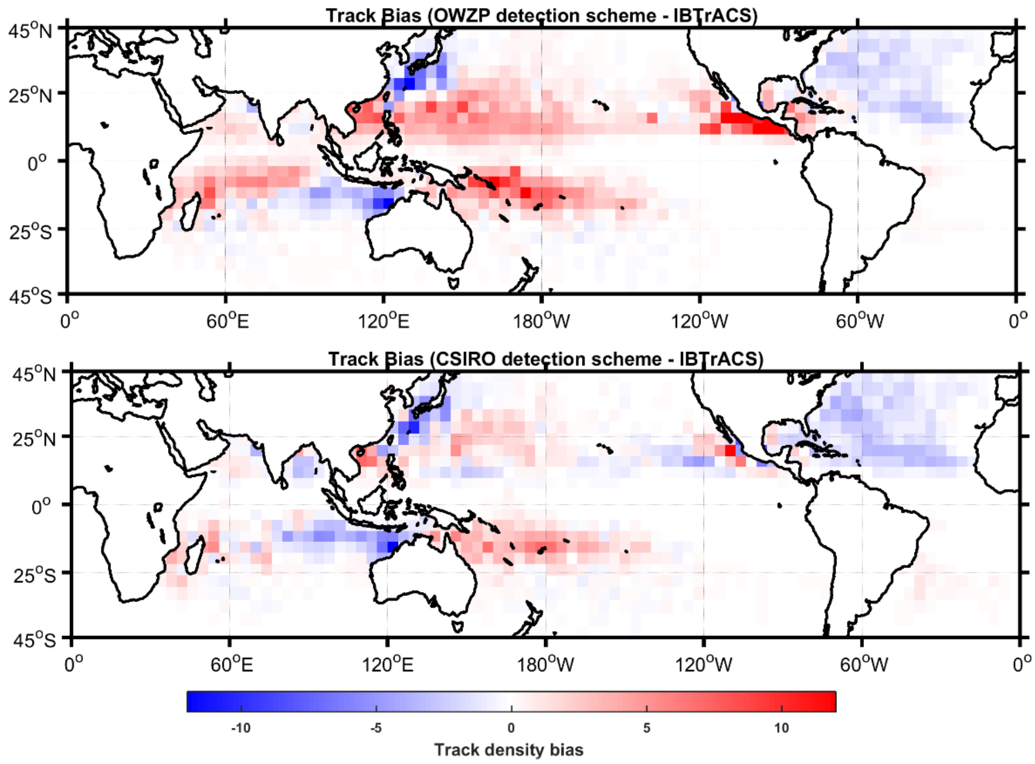

Figure B. 5. TC track density (number of tracks per $4^0 \times 4^0$ box region across the globe) bias obtained by subtracting the observed track density (IBTrACS) from model simulated track density (OWZP and CSIRO detected tracks).

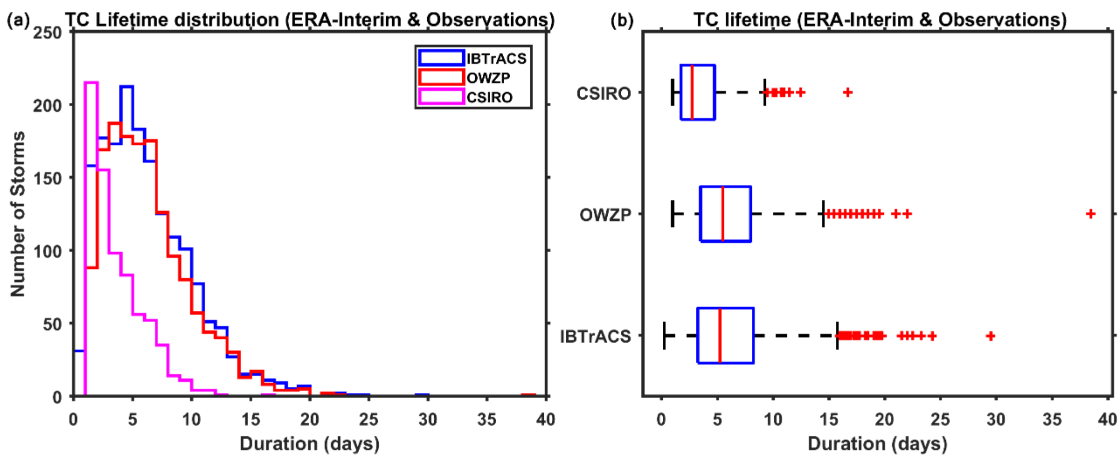

Figure B. 6. (a) Histograms of TC lifetime in both observations and reanalysis using both detection schemes (b) Distributions of TC lifetime with the vertical line of the box indicate the median of the distribution, the left, and right box edges show the 75th and 25th percentiles and the whiskers indicate the maximum and minimum values and plus signs indicate the outliers.

Table B. 1. Basinwise mean genesis density bias (ERA-Interim minus observations) per year using both the detection schemes in reanalysis when compared against IBTrACS and percentage bias of the storms when compared to observations (* indicates 95% significance level)

BASIN	OWZP-ERA (TCs/year)	CSIRO-ERA (TCs/year)	% bias of OWZP storms(ERA)	% bias of CSIRO storms(ERA)
MDR_NA	-1.55*	-5.9*	(-) 23%	(-) 89%
ENP	-2.3*	-13.85*	(-)15%	(-) 89%
WNP	1.35*	-7.6*	(+) 5.6%	(-) 32%
NI	0.75*	-2.05*	(+) 19%	(-) 50.6%
SI	3.35*	-4.5*	(+) 22%	(-) 29%
SP	-2.4*	-2.8*	(-) 26%	(-) 31%
AUS	-0.3*	-1.85*	(-) 3.1%	(-) 19%

Appendix C

This appendix section gives additional technical details for the machine learning algorithms of chapter 5.

Decision trees use an information gain ratio measure to select an attribute that best separates the two different groups (i.e., either developing tropical depression or non-developing tropical depression) using the surrounding environmental conditions.

C.1 Information gain ratio

Information gain ratio is a quantitative measure giving the performance of the split in separating the developing from the non-developing tropical depressions using a particular environmental variable. It is estimated using a parameter termed information entropy. The information entropy is a measure of variance in the data: a lower variance dataset will have lower entropy, and higher variance data will have higher entropy. The information gain is calculated by subtracting the weighted information entropy after splitting the data from the information entropy before the split. The information gain ratio is the ratio of information gain and the information entropy of a particular variable. Therefore, information gain ratio measures how much variance in the dataset is reduced when a particular environmental variable separates the classes (i.e., either developing or non-developing). The formula to measure information entropy (E) is:

$$E = -\sum_{i}^{C} p_i \log p_i$$

where p_i is the probability of class i (for i = 1, 2). For instance, K is a training data set that has k samples of data with two classes Ci (for i = 1, 2), and then k ($C1$) and k ($C2$) represents the number of samples for each class. Then the information entropy ($E(K)$) required for classifying the training set K is

$$E\ (K) = -\sum_{i=1}^{2} \frac{k(Ci)}{k} \log\left[\frac{k(Ci)}{k}\right]$$

Now we have an environmental variable Z, and it has some P number of distinct values, and information entropy ($E\ (K|Z)$) required to separate the training set K to P different sets {K1,K2,K3,...,KP}

$$E\ (K|Z) = -\sum_{j=1}^{P} \frac{Kj}{k} E(K_i)$$

The information gain is measured as:

$$Information\ Gain\ = E(K) - E(K|Z)$$

Eventually, the information gain ratio can computed as:

$$Gain\ ratio\ (Gain) = \frac{Information\ Gain}{E(K|Z)}$$

In the current study we have five different environmental variables as described in section 5.2.4 of chapter 5.

C.2 Reinforcement learning and dynamic programming

Reinforcement learning (RL) and dynamic programming (DP) are the fields in machine learning that make a sequence of decisions to achieve a goal in an uncertain and continuously evolving complex environment. These methods are therefore suitable for modeling dynamical processes (Barto, 1995; Busoniu et al., 2010). Here, a Markov model is used which involves a sequence of steps where, at each step the state (that is, the physical property of the circulation) of the system transforms to a new state due to an assigned action (e.g. developing or non-developing tropical depression) and receives a reward each time for the chosen action. A reward is a real value which depends on the environmental state and the action which is required by the Markov model to obtain a discrete solution (maximum expected reward). In this model, a decision-maker can assign rewards for different actions based on its suitability to different states as per the previous knowledge of the tropical storms (TS). The decision maker or the agent receives the reward for the chosen action after observing the environmental state at each step of the iterative process. The iterative process ceases when no further increase in the cumulative reward of the Markov model. The final decision of the Markov model is based on the maximum of the cumulative rewards received through multiple sequential decisions. In contrast, decision trees give a single decision for each environmental state without considering other possible decisions due to the dynamical state transitions before the TS formation. Therefore, RL and DP involve a multiple-decision process with a reward system in a time varying dynamic environment. Unlike supervised learning (like decision trees), in RL and DP, we do not get an exact result; instead, it obtains a reward for its preformed action and the final result is based on the maximum value of cumulative rewards.

C.2.1 Markov Decision Process

Both RL and DP methods involve a chain of decisions that require a specific method to finalize the discrete decision after including the dynamical transitions between the environmental states. The Markov decision process (MDP) is a method that can formulate the RL and DP problems for obtaining the discrete solution (Puterman, 1994). The MDP has five different elements to solve a problem that involves sequential decisions:

S: A set of states. In the present work the state represents the environmental conditions during the initial development of tropical depressions derived from the set of atmospheric conditions obtained through the decision tree rules

A: A set of actions. Here there are two actions, i.e., either developing or non-developing tropical depression. At each time step of the iterative process the learning agent/decision maker selects an action.

P: State transition probabilities. These are the probabilities obtained when the state transition occurs with the agent choosing an action based on the current environmental state and rewards.

R: Rewards are real values which are assigned by policy maker based on the prior knowledge of the suitability of environmental state to an action. At each time when the state changes, the agent receives the reward for assigned action. The maximum cumulative reward values are used to find the discrete solution to the MDP problem.

γ: a discount factor (varies from 0 to 1) or discounted reward which gives higher importance to immediate rewards than future rewards and helps in converging the solution.

The way the learning agent decides which action to perform for a specific state is called a policy. The policy (**π**) is a functional representation of the current state that returns the action.

$$\boldsymbol{\pi}(\mathbf{s}): \mathbf{S} \rightarrow \mathbf{A}$$

where '*s*' is the current state chosen from the set of states 'S' which returns a set of actions 'A'.

The MDP shown in Figure C.1 aims to train the agent to find that optimal action for each of the atmospheric states. This combination of the optimal actions to each environmental states is termed as optimal policy (environmental states + optimal actions (i.e. either developing or non-developing tropical depression) = optimal policy). The optimal policy is based on the maximum cumulative rewards obtained over time.

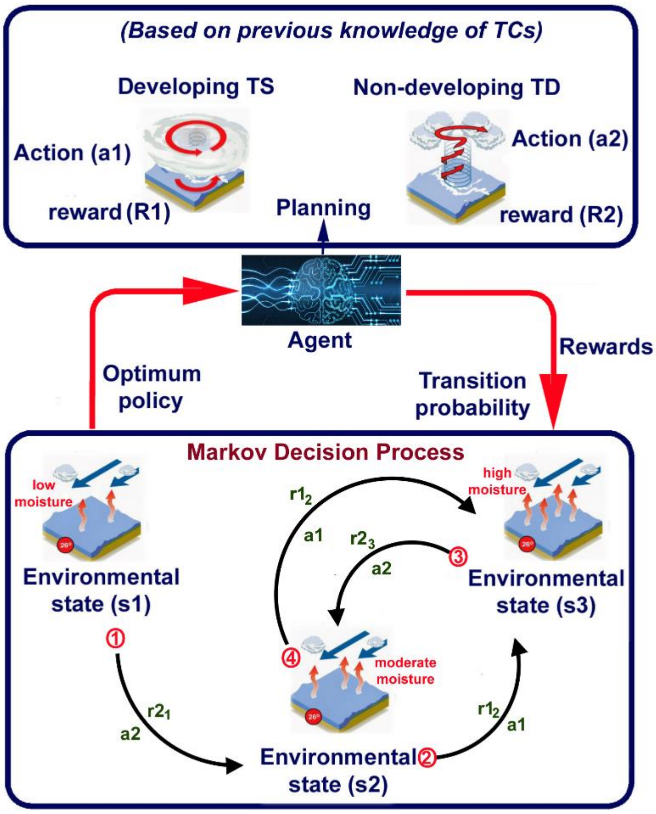

Figure C. 1. Schematic of the MDP depicting the assignment of reward matrix R1 and R2 (having rewards $r1_i$, and $r2_i$ respectively, where i indicates the environmental states) for action a1 and a2 are based on knowledge of TS favorable environmental conditions. The rewards and state transition probabilities are the input to the MDP which results in a four step Markov chain (from step 1 to step 4 shown in red fonts) having state transitions $s1 \rightarrow s2 \rightarrow s3 \rightarrow s2 \rightarrow s1$ with rewards $r2_1 \rightarrow r3_2 \rightarrow r2_3 \rightarrow r1_2$ and actions $a2 \rightarrow a1 \rightarrow a2 \rightarrow a1$ thereby conversing at action $a1$ (development of TS) in the environmental state 3. The MDP chooses the actions for respective environmental states (optimum policy) at the end of the iterative process.

The MDP problems are usually formulated for two different problems (Barto, 1995; Busoniu et al., 2010): (i) dynamic programming (DP) requires the knowledge of the state transitions and the rewards, (ii) RL does not require prior knowledge of the state transitions and rewards, and it learns about them by using a trial and error method. In the current study we employ the DP to obtain the solution (i.e. favorable environmental states for tropical cyclone formation) after considering the dynamical state transitions of developing and non-developing tropical depressions atmospheric conditions. There are two methods to solve the MDP problem by dynamic programming (Barto, 1995): (i) value-iteration and (ii) policy iteration. In both methods, the agents know information about the state-transition probabilities and the rewards. Therefore, an agent in MDP plans actions (i.e. either developing or non-developing case) for each atmosphere state with the given knowledge on the environment through transition probabilities and rewards. In the current study, we use the policy iteration algorithm using MATLAB functions (Chades et al., 2005) for estimating the actions to the atmospheric states obtained from the decision trees using the data 48 hours before the declaration of a tropical depression stage vortex. The following section explains the mathematical equations involved in solving an MDP problem.

C.2.2 Value function

The value function V(s) determines goodness of a starting from a specific state 's' to the agent. It is represented as the total expected reward value for initiating from a state ‘s.’ Therefore, the V(s) depends on the policy as the agent assigns an action for a specific state (Barto, 1995; Alzantot, 2017; Alpaydin, 2020). For a given policy π the value function is given as

$$V^{\pi}(s) = E\left[\sum_{i=1}^{T} \gamma^{i-1} r_i\right] \quad \forall s \in S$$

where γ and r denote the discount factor and the reward respectively at the i^{th} time step, and s denotes a state ‘s’ in a set of states ‘S.’

Among the different value functions there exists an optimal value function ($V^*(s)$) which is given as:

$$V^*(s) = max\, V^{\pi}(s) \quad \forall s \in S$$

The optimal policy ($\pi^*(s)$) for this corresponding optimal value function is:

$$\pi^*(s) = arg\, max\, V^{\pi}(s) \quad \forall s \in S$$

Also, to this state-value function, there exists another Q-function of state-action (state (S) and action (A)) that gives a real-value (R) necessary for solving the RL algorithms.

$$Q: S \times A \rightarrow R$$

The Q*(s, a) (the optimal Q-function) is the expected total reward received by the agent while beginning from 's' state by picking an 'a' action. This optimal Q* function indicates the goodness of choosing an action 'a' for a state 's', while V*(s) denotes the maximum expected total reward for initializing from a state 's,' therefore it is the maximal of the Q* function across all feasible actions.

$$V^{*}(s) = maxQ^{*}(s, a) \qquad \forall s \in S$$

From the optimal Q-function Q*(s, a), we can get the optimal policy, which is the action that gives the maximum of function Q*.

$$\pi^{*}(s) = \arg maxQ^{*}(s, a) \qquad \forall s \in S$$

For these RL problems, the Bellman equation defines optimal Q*(s, a) recursively (Barto, 1995). That is, the Q*(s, a) is combination of the immediate reward (R(s, a)) and the discounted expected future reward for transitioning into a new state s` after choosing an action.

$$Q^{*}(s, a) = R(s, a) + \gamma E_{s`}[V^{*}(s`)]$$

The above equations of value and Q functions are for a deterministic environment solved by finite MDP. For a stochastic environment which is a time varying requires a dynamical equation that gives time varying changes in terms of probabilities. The equations for the stochastic environment are described below.

$$Q^{*}(s, a) = R(s, a) + \gamma \sum_{s` \in S} p(s`|s, a) V^{*}(s`) \text{ ; for 'm' different states}$$

$$\text{Since } V^{*}(s) = maxQ^{*}(s, a)$$

$$V^{*}(s, a) = max[R(s, a) + \gamma \sum_{s` \in S} p(s`|s, a) V^{*}(s`)$$

Here p(s`|s, a) is the probability for moving into state s` from state s after taking an action a in the iterative process. If there are three states then the above equation will be a sum of three probabilities multiplied by the value function.

The final equation given above is the Bellman equation used to solve by dynamic programming. In DP, we will break the complex problem into sub-problems and then estimate the solutions and store them. If the same subproblem occurs in the iterative process, we use the stored solution rather than recalculating the solution.

We solve a Bellman equation using two powerful algorithms: (i) Value Iteration, (ii) Policy Iteration (Figure C.2). The steps involved in solving using the policy iteration employed in the current study are given below (Barto, 1995):

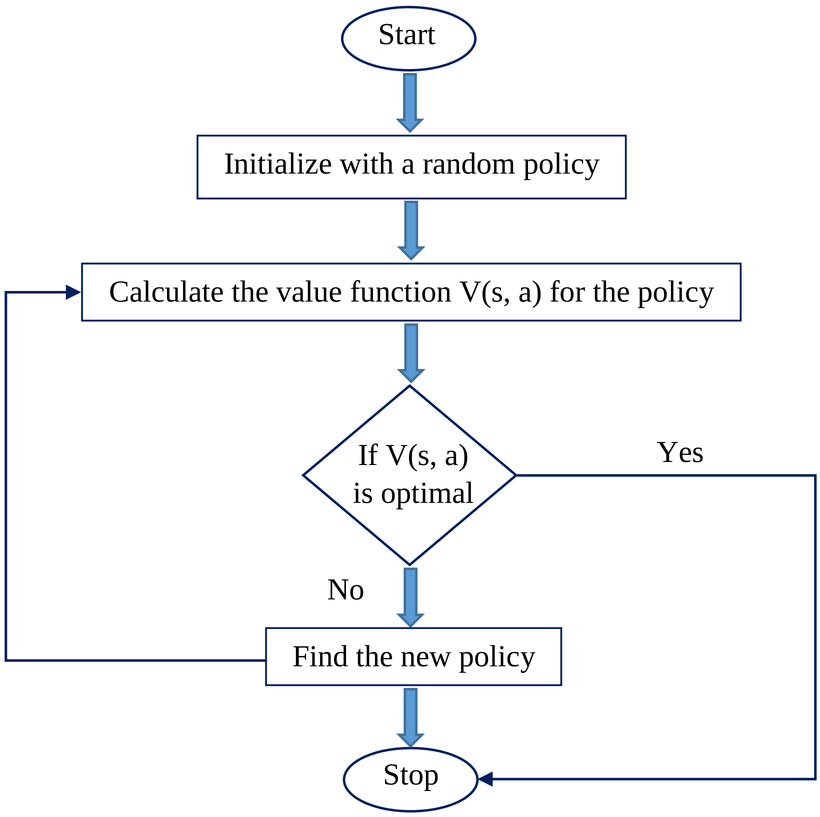

Figure C. 2. Flowchart for policy iteration showing the calculation of value function in a loop until an optimum V policy is achieved (courtesy: https://medium.com/@sanchittanwar75/bellman-equation-and-dynamic-programming-773ce67fc6a7).

C.2.3 Steps followed in the MDP problem for the current study

In order to model the uncertainty in the TS suitable states predicted (48 hrs before genesis) from decision trees, we use a Discounted Markov Decision Process (MDP) with a policy iteration algorithm, suitable for stationary infinite horizon problems. The infinite horizon problems consider infinite time steps (i.e., Markovian state transitions termed epochs) where a discount factor is used for convergence of the iterative state transitions (Puterman, 1994). In the present dynamic programming, a specific time horizon is difficult to assign. We do not know how much time instances (i.e., epochs) it might take for the development of a tropical depression to a TS. In order to find an optimum policy, an infinite time horizon is considered by assigning a discount factor that converges the mathematical solution, which is otherwise challenging to converge in an infinite horizon. Here we also consider stationarity at each time step for simplicity, so the action is not changed for a state at a new time during the iterative process.

Reyes et al. (2006) used a C4.5 decision tree to build some abstract states and rewards for the MDP and suggested a similar hybrid approach of C4.5 and MDP. Therefore, here we use the decision tree rules obtained from C4.5 algorithm to create the states for the MDP problem.

The major steps to solve the MDP problem using the policy algorithm are finding the set of environmental states, state transition probability matrix and the reward matrix which depends on both the states and the transition probabilities. The following section describes the each of these steps carried out to estimate the optimal action for each of the environmental states using policy iteration algorithm.

(i) The first step in formulating the MDP problem is defining the Markovian states. In the current study, the states represent the initial environmental conditions through which the initial depression gradually develops or disrupts to non-developing cases. Therefore, we categorize the important predictors observed from each of the decision rules into three different classes. One class upper bounded by the mean of the variable, a second class bounded between the mean and thresholds of the variable obtained from decision tree rules and the third class by thresholds of the decision tree rule. If the mean of the large-scale variable is very close to a threshold of the decision tree rule variable, then the value at one standard deviation away is considered as the boundary. For example, in South Pacific basin the Rule of decision tree shows (Table 5.5) that if WV > 45.56 kg/m^2, USHEAR <= 4.2 m/s, OWC850 <= 18.69×10^{-10} s^{-2} and MPI > 64.23 m/s then a TC forms. Here, the mean of MPI is 65.188 m/s, which is close to the threshold. The standard deviation of the MPI distribution is 11.5523 m/s. Therefore, the first boundary is obtained one standard deviation away from the mean (i.e., 65.188 -11.5523 = 53.6357 m/s). This division of decision rule is carried out separately for developing and non-developing depressions for obtaining their respective three Markovian states. The environmental conditions of a tropical depression in each of the respective input dataset depicts one of these three classes representing the states possible before the TS formation.

(ii) Next, we estimated transition probabilities from the time series records of the developing and non-developing cases, which give the chances of the favorable state (as per prior knowledge on TS favorable environmental conditions) to move to any of the two unfavorable states for the developing and non-developing cases separately. The process of developing into a TS represents action one, while the non-developing process is another action.

(iii) After that, we need to assign the initial rewards to the Markovian states and the respective actions (developing or non-developing) based on previous knowledge of the favourability of these states to TS development or non-development. Therefore, in the case of the action one (favoring TS formation), we assign higher rewards to the states that are favorable for TS formation, whereas lower rewards to the unfavorable conditions and vice versa. To estimate the favourability of the Markovian states, we consider the decision tree threshold and the state transition probability, which depicts the uncertainty of retaining that state. Henceforth, we estimate the rewards for each action by multiplying the normalized transition probabilities by either 2 for favorable states or 1 for unfavorable states. For instance, the state showing higher tropospheric moisture content will be given higher rewards, and the state showing increased vertical wind shear will be given lower rewards for action one. The reward is assigned as follows:

$$r_{s|a} = \begin{cases} 2 \times \dfrac{TP_{s|a}}{total\ state\ transitions_a} & if\ \ state\ s\ is\ suitable\ \ for\ action\ a \\ 1 \times \dfrac{TP_{s|a}}{total\ state\ transitions_a} & if\ state\ s\ is\ not\ suitable\ \ for\ action\ a \end{cases}$$

where $r_{s|a}$, $TP_{s|a}$ denotes the reward and transition probability respectively for state *s* given the action *a*, whereas *total state transitions* $_a$ is the summation of state probabilities of all three states (here *total state transitions* $_a$ =3). For example let us consider the NA basin, where the decision Rule 1 states that OWC850 > 14.5×10^{-10} is suitable for TS formation (as illustrated in Table 5.5 of section 5.3.5). This environmental condition is designated as state S3 in the MDP formulation (as illustrated in Figure 5.1 of section 5.3.6) and can transit to a less favourable state S1 or S2 with a probability of 0.03 or 0.09 respectively when TS develops (say action 1). In contrast, the state S3 is unfavourable for non-development (say action 2) and can transit to state S1 or S2 with a probability of 0.29 and 0.07 respectively. So the rewards assigned to state S3 for action 1 (developing TS) are 1/3×0.03 and 1/3×0.09 while transiting to state S1 and S2 respectively, and similarly for action 2 (non-developing TD) they are 2/3×0.29 and 2/3×0.07 respectively.

(iv) Therefore, solving an MDP involves mapping an action (either developing or non-developing) to each given state (environmental conditions). It considers probabilistic transitions of the dynamic states, thereby accounting for uncertainty in the evolution of a state under different specified actions. A discount factor of 0.95 is chosen such that rewards are reduced by 5% at each iteration in estimating the future sum of rewards to attain the mathematical convergence for the iterative algorithm.

(v) Finally, we apply the discrete Markov Model to find the optimum chances of TS formation based on the probability of suitable states in developing cases to grow to the TSs. The MDP helps to ascertain the probabilistic evolution of the TS suitable states under the observed uncertainties with state transitions.

C.3 Support Vector Machines (SVM)

The main aim of the SVM algorithm is to identify the hyperplane in an N-dimensional space that best separates the two underlying groups developing and non-developing tropical depressions (Murty & Raghava, 2016). The following Figure C.3 shows the hyperplanes in two-dimensional and three-dimensional space. These hyperplanes are the decision boundary separating the classifying groups. There are many possible hyperplanes, but the SVM selects the optimum hyperplane.

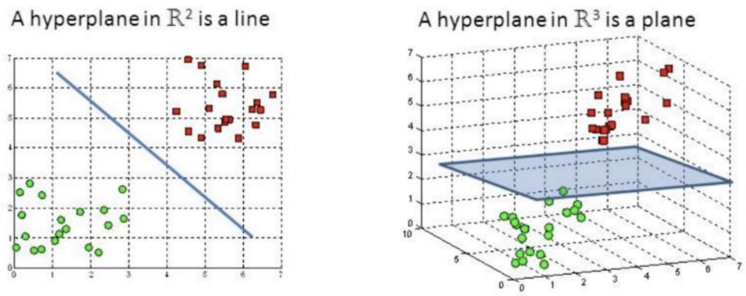

Figure C. 3. A schematic showing one of the possible hyperplanes in two and three-dimensional space separating the classifying groups (modified after Gandhi, 2018).

The hyperplane, with a maximal margin between the decision boundaries and the training data points (with least misclassification error), as shown in Figure C.4 acts as an optimal hyperplane (Kecman, 2005; Murty & Raghava, 2016).

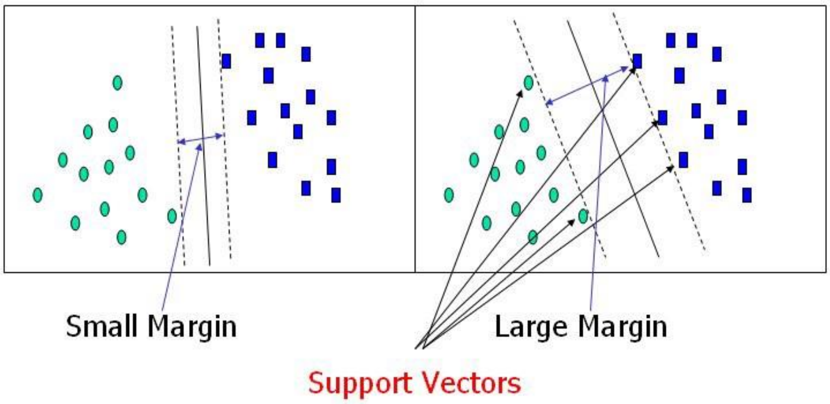

Figure C. 4. Support vectors and the margins considered by the SVM algorithm to identify the optimal hyperplane (modified after Gandhi, 2018).

Non-linear SVM: For the data that is not linearly separable (Figure C.5), the SVM algorithm uses kernel functions (generalized dot products) that transform the data into higher dimensional space, which is then separable (Murty & Raghava, 2016).

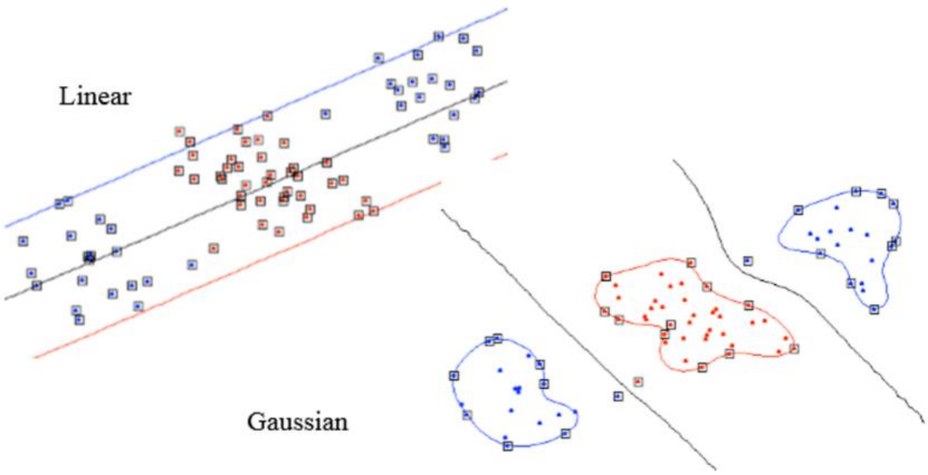

Figure C. 5. The data on top (left) is not separable using a linear function. Therefore, the data is transformed into a higher-dimensional space with a Gaussian kernel function in the bottom (right) panel (modified after Berwick, 2003).

Perceptron (Linear classifier):

A perceptron is a binary classifier that uses linear discriminant functions to classify the two underlying groups. The mathematical framework of the perceptron learning algorithm is described below.

Defining input (*x*) and target (*y*) data: Let's assume a training data (x_i, y_i) for i = 1 … N. Here $x_i \in \Re^d$ (real-valued samples of environmental variables having d dimensions) and $y_i \in \{-1, \; 1\}$ represents their respective classes. The goal of SVM is to define a classifier $f(x)$ such that $f(x_i) \begin{cases} \geq 0 & y_i = +1 \\ < 0 & y_i = -1 \end{cases}$.

Finding the separating hyper-plane *f* (x_i): In case of linearly separable data, the classifier *f* (x_i) can be a linear function representing a straight line for 2d data, a plane for 3d data, or a hyper-plane for N (d) data. This process can be represented as $f(x) = w^T x + b$, where *w* is the weight vector (normal), and *b* is the bias. The simplest way to understand a linear classifier is given by the Perceptron Learning Algorithm. In this algorithm, the classifier *f* (x_i) can be estimated using which maps the input *x* and target *y*, as described below.

At first, we define a separation plane $\vec{a} \cdot \vec{x} = 0 \quad \left| \begin{array}{l} \vec{a} \cdot \vec{x} > 0 \\ \vec{a} \cdot \vec{x} < 0 \end{array} \right. \quad if \quad \begin{array}{l} y_i = +1 \\ y_i = -1 \end{array}$ as shown in Figure C.6.

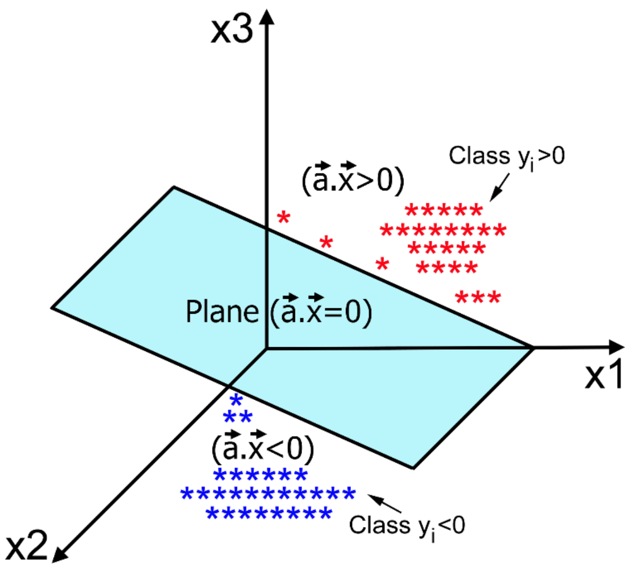

Figure C. 6. Separation plane

Here $\vec{a} \cdot \vec{x} = 0$ is separation plane such that $\vec{a} \cdot \vec{x} > 0$ if $y_i = +1$ and $\vec{a} \cdot \vec{x} < 0$ if $y_i = -1$ in the x1, x2, x3 coordinates. If the negative samples are transformed to positive, (i.e. if $y_i = -1$, set $x_i \rightarrow -x_i$ and $y_i \rightarrow -y_i$), then we get $sign\ (\vec{a} \cdot \vec{x}_i) = y_i$. So the perceptron algorithm now simplifies to find the plane $\vec{a} \cdot \vec{x}_i \geq 0$ for i = {1, …, N}. Thereafter, we initialise $\vec{a} \cdot (0) = 0$ and loop over {i = 1, …, N}. Following this we calculate the error of misclassification by finding the total distance of all the misclassified samples (i.e. $\vec{a} \cdot \vec{x}_i < 0$) from the classifier plane. Then update $\vec{a}$ as $\vec{a} \rightarrow \vec{a} + \vec{x}_i$. The above step is repeated in a loop until the error is minimized through gradient descent.

SVM algorithm:

The approach of SVM is similar to the Perceptron Learning Algorithm as discussed above. However it has some privileges over the Perceptron algorithm. A noteworthy advantage of SVM is that it gives one optimised separation plane whereas Perceptron algorithm can give more than one solution (theoretically infinite solutions). Additionally, if the input data is not linearly separable, the SVM uses kernel functions which are capable of bringing a linear separation in higher dimensional space (Figure C.5).

The framework of the SVM (Figure C.7) can be understood as follows (Collobert & Bengio, 2004; Murty & Raghava, 2016):

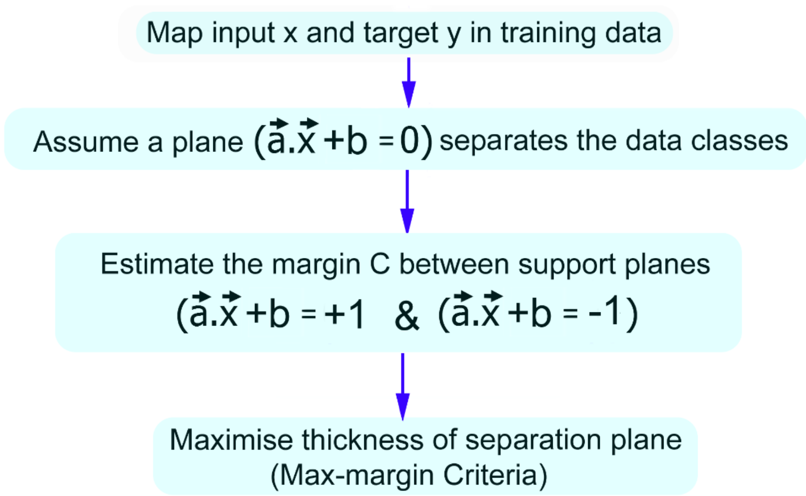

Figure C. 7. The framework of SVM

The linear SVM considers a hyperpalne $\vec{a}\cdot\vec{x}+b=0$. Here input data having y_i =1 can be considered to lie on plane $\vec{a}\cdot\vec{x}+b=1$ or above (viz. $\vec{a}\cdot\vec{x}+b>1$) as shown in Figure C.8. Similarly, the input data having y_i =-1 can be considered to lie on plane $\vec{a}\cdot\vec{x}+b=-1$ or below (viz. $\vec{a}\cdot\vec{x}+b<-1$). The planes $\vec{a}\cdot\vec{x}+b=1$ and $\vec{a}\cdot\vec{x}+b=-1$ are termed as support planes. The points lying on the support plane are called support vectors and the distance between the support planes is called the margin (Figure C.8). The SVM maximizes the margin (C) of the hyper-plane ($\vec{a},b$) so that it has maximum thickness touching the nearest support vectors on either side as:

$$\max_{\vec{a},b,|\vec{a}|=1} C \quad \Big| \; y_i(x_i\cdot\vec{a}+b)\geq C, \qquad \forall i\geq 1 \quad to \quad n$$

Since perfect classification is often impossible to achieve, SVM allows some misclassification. For this, some slack variables $\{z_1, z_2, \ldots z_n\}$ are included using which samples can move in direction of $\vec{a}$ (viz. right side as shown in Figure C.8). However, a penalty is required for using the slack variables which can be given as $\sum_{i=1}^{n} z_i$. That means that if classification is perfectly accurate (without any misclassification) then z_i = 0, else z_i > 0. Now, the maximisation of the margin C can be done using the Max-Margin criteria where we maximise the margin and minimise the amount of slack variables, which can be represented as:

$$\max_{\vec{a},b,|\vec{a}|=1} C \quad \Big| \; y_i(x_i\cdot\vec{a}+b)\geq C(1-z_i), \qquad \forall i\geq 1 \quad to \quad n \big| Z_i\geq 0 \qquad \forall i$$

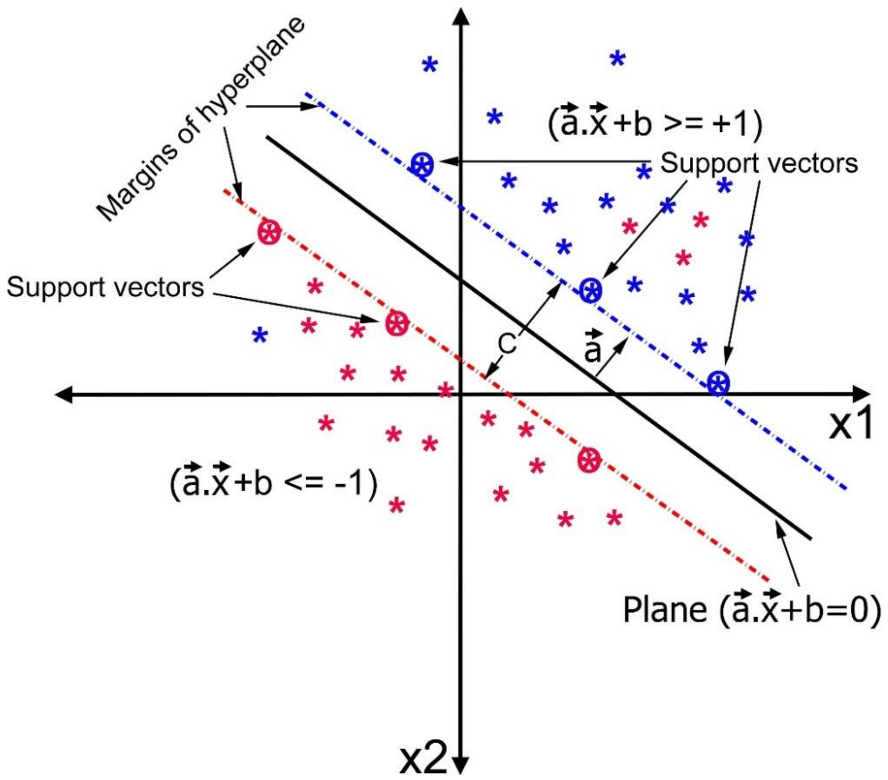

Figure C. 8. Maximisation of the separation hyper-plane showing the margins having thickness C. Here the vector $\vec{a}$ directing rightward is normal to the hyper-plane. Note that some samples are misclassified which need to be moved across the classification plane using slack samples.

C.4 Formulae of skill scores

The output of machine learning algorithms to predict the developing and non-developing cases is a confusion matrix or contingency table, as shown in the Figure C.9 Here 1 denotes the developing cases and 0 as the non-developing cases.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Figure C. 9. Schematic of the confusion matrix or the contingency table, which is the output of the machine learning (ML) algorithms. Here 1 indicates the developing cases, and 0 indicates the non-developing cases.

There are different scores to get the performance of the machine learning algorithms (Hamill & Juras, 2006). The hit rate (HR) or the probability of detection gives us the measure for what the algorithm correctly detects a fraction of the developing cases. The false alarm rate (FAR) gives what fraction of the non-developing cases are incorrectly forecast as developing cases. The overall accuracy in classifying the developing and non-developing case is:

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

$$Hit\ Rate\ (HR) = \frac{TP}{(TP + FN)}$$

$$False\ Alarm\ Rate\ (FAR) = \frac{FP}{(FP + TN)}$$

The other forecast skill scores are precision, recall, and F-score. The precision is the ratio of the correctly predicted cases as developing to total labeled as developing cases by the model.

$$precsion = \frac{TP}{(TP + FP)}$$

The recall is the ratio of the number of correctly predicted developing cases to the number of developing cases in the dataset.

$$recall = \frac{TP}{(TP + FN)}$$

The F-score is the harmonic mean of both precision and recall.

$$Fscore = \frac{2 * (recall * precision)}{recall + precision}$$