A STATISTICAL APPROACH TO UNDERSTANDING AND PREDICTING TROPICAL STORM FORMATION IN THE EAST PACIFIC BASIN

By

Dianna N. Nelson

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

(Atmospheric and Oceanic Sciences)

at the

UNIVERSITY OF WISCONSIN-MADISON

2013

Date of final oral examination: 10 December 2012

The dissertation is approved by the following members of the Final Oral Committee:

Dr. Michael Morgan, Professor, Atmospheric and Oceanic Sciences
Dr. Jonathan Martin, Professor, Atmospheric and Oceanic Sciences
Dr. James Kossin, Adjunct Professor, Atmospheric and Oceanic Sciences
Dr. Larissa Back, Assistant Professor, Atmospheric and Oceanic Sciences
Dr. Grace Wahba, Professor, Statistics

Abstract

Despite its spatial confines, the east Pacific basin is one of the most active basins in the world for tropical cyclone (TC) genesis. While the TCs that form in the basin have important implications for Central America and the southwestern U.S., relatively little research (compared to other tropical basins) has been done on eastern Pacific tropical cyclogenesis. The present study uses two statistical techniques – linear discriminant analysis (LDA) and a Bayesian probabilistic model to identify those variables that are associated with the development of nascent vortices in the east Pacific and uses them to predict tropical storm formation for lead times out to 48 hours. All nascent vortices that last for a minimum of 48 hours and form during the 2001-2009 "peak" hurricane seasons (July-September) are considered in the study. An initial set of 27 spatially averaged variables is considered as potential predictors for the statistical models. Results from both the LDA algorithm and Bayes probabilistic model show that a number of predictors improve the forecast skill of both models. These predictors include the 900hPa relative vorticity, latitude of the vortex, 900hPa deformation fields, 900hPa-500hPa relative humidity, 900hPa zonal wind, and the 900hPa-200hPa equatorward vertical shear of the meridional wind.

Using the aforementioned predictors as a basis, a composite-based conceptual model for environmental elements that favor tropical storm formation is constructed to explain the physical mechanisms of the process. In conjunction with the statistical and composite-based models, Tropical Storm Gil (a 2007 storm that is accurately forecast by the statistical models) and Tropical Storm Enrique (a 2009 case that is poorly forecast) are used to highlight the strengths and weaknesses of each model.

Acknowledgments

First, I would like to thank my advisor, Dr. Michael Morgan, for his guidance and support throughout my graduate career. Michael's genuine and energetic personality makes him an exceptional mentor and friend to all of his graduate students. I sincerely thank him for all of his direction and advice regarding my research over the past several years. I would also like to thank the members of my committee: Dr. Jonathan Martin, Dr. Jim Kossin, Dr. Larissa Back, and Dr. Grace Wahba. It has been an incredible privilege to work with such a talented and enthusiastic group of scientists. I owe a special thanks to Dr. Chris Davis for his programming expertise and Linda Keller for her advise and comments on much of the written work presented in this dissertation.

In addition to working with some brilliant mentors, I have been fortunate enough to work along side an outstanding group of colleagues. Nick Bassill, Brett Hoover, and Ross Lazear have all been amazing to work with, providing me with a myriad of research ideas to build on and creating a friendly and fun work environment (not to mention the number of times they have kept me sane as I suffered setbacks with my research). By extension, I would also like to thank the department's graduate student body that consists of a collaborative and friendly group of graduate students who all have an impeccable sense of camaraderie and friendship.

Technological savvy is just as important as research guidance. Therefore, I am forever indebted to Pete Pokrandt and his fluency in "computer-speak". Pete solved innumerable technical problems that I encountered over both my undergraduate and graduate careers. Perhaps the only thing more impressive than Pete's knowledge of computers is his ability to remain calm and cheerful no matter how hopeless a technological problem may seem.

Finally, I would like to thank my most significant support group, that is, my family and friends. I have been blessed with a wonderful and loving family and devoted friends that have all been there for me throughout my college education. It is truly and incredible feeling to know that there are a number of people who are willing to sacrifice time out of their day to do whatever they can to help you achieve your goals. For that – Mom, Dad, Natalie, Jon, Steph, Grandma, Grandpa, Rachel T., Rachel W., Kate, and a host of other family and friends that I regretfully do not have space to list here (you know who you are though) – I will forever love you and appreciate all you have done to help me throughout the years.

Table of Contents

Chapter 1: The Eastern Pacific Tropical Cyclone Basin

1.1 Introduction

The eastern Pacific is an active basin for tropical cyclone (TC) development. Second only to the west Pacific basin, 19.7% of all tropical cyclones form in the east Pacific basin (Sadler 1964; Landsea 1995). More impressive is that the eastern Pacific basin's main development region is quite small. In fact, 75% of the TCs in the eastern Pacific basin form between 95°W to 115°W and 6°N to 18°N (Davis et al. 2008). While many TCs that form in this basin do not make landfall, those that do have significant impacts on both Central America and southwestern United States. For example, in 2008 Tropical Storm (TS) Alma caused an estimated $33 million in damages to Costa Rica alone (Brown 2008). Two years later, TS Agatha resulted in 190 deaths and over $1 billion in damages for Guatemala, El Salvador and Honduras combined (Beven 2011). However, not all impacts of TCs are negative for this part of the world; indeed, the precipitation associated with eastern Pacific TCs and their remnants is of vital importance to the same regions as it provides a significant portion of the rainfall for Mexico and the southwestern U.S. (Corbosiero et al. 2009). Despite the active nature of this basin, eastern Pacific tropical cyclogenesis has not garnered much interest from the scientific community until recently; therefore, it is imperative that more work be done to understand the genesis mechanisms and improve predictions of TC genesis in this basin.

With the exception of field campaigns, a lack of comprehensive in-situ data over the oceans makes it difficult to determine the interactions – especially on the meso- and smaller scales – which lead to the development of a tropical cyclone from a nascent vortex. The advent of remote sensing from satellites and numerical weather prediction data have allowed

for progress to be made in understanding the process of tropical cyclogenesis. However, the ability to accurately forecast TC genesis is still inconsistent (Pratt and Evans 2004). This study will use a variety of statistical approaches as a means to evaluate the predictive power of several physical mechanisms thought to be important to tropical cyclogenesis in the east Pacific basin. The following is a brief description of these extant genesis mechanisms with a focus on those known broadly to be relevant to the eastern Pacific.

1.2 Climatological features

Observations have shown that TCs tend to occur cluster in both time and space, with the majority of TCs forming in the Northern Hemisphere, specifically in the west and east Pacific and Atlantic basins (Gray 1968, 1979). The spatial confines of TC formation to these basins results from the superposition of features that is conducive to genesis. Based on observations, the necessary, but not sufficient, conditions for tropical cyclogenesis include sufficiently warm (26°C or warmer) sea surface temperatures (SSTs), low-level cyclonic relative vorticity, sufficiently large planetary vorticity, weak vertical shear of the horizontal wind, and high relative humidity in the low to middle troposphere (Gray 1968, 1979). Consideration of each of these features follows.

a. Sea Surface Temperatures

It is clear from observations that the surface of the ocean plays a key role in the formation and maintenance of TCs. Observations have shown that TCs only form over relatively warm SSTs (Gray 1968, 1979), weaken over cooler SSTs (Palmén 1948; Rosendal 1962), and dissipate quickly upon making landfall. The observed minimum SST in TC formation locations is about 26°C (Palmén 1948, 1956; Gray 1968, 1979). The presence of warm SSTs for TC formation is no coincidence. Enhanced fluxes of latent and sensible heat

release from the warm ocean into the atmosphere (Riehl 1948, 1951; Kleinschmidt 1951) in conjunction with thermodynamic disequilibrium between the ocean and the underlying atmosphere (Emanuel 1986, 1991) provide energy for the formation of a TC.

The spatial confines of genesis in the eastern Pacific basin are largely due to the SSTs (Palmén 1948), which are above 26°C from the west coast of Central America and Mexico to approximately 130°W, but drop off significantly further west of this longitude as well as north of 20°N (Figure 1.1). For the present study, the western edge of the basin is considered to be 140°W and the northern edge is 20°N; thus, the vast majority of the basin meets the minimum SST criterion. Therefore, SSTs will not be considered as a parameter for the statistical models (described in Section 1.4) used in this study.

b. Low-level vorticity

The presence of low-level, cyclonic vorticity has been noted as a necessary condition for TC genesis both in the sense of ambient background relative vorticity and the relative vorticity of a pre-existing vortex (Gray 1968, 1979). The main development regions (MDRs) for each basin are located in areas of large-scale, low-level, cyclonic wind shear, which can lead to convection via Ekman frictional pumping of the lower-tropospheric wind (Gray 1968). The likelihood of a nascent vortex developing into a TC depends in part on its relative vorticity, such that larger relative vorticities result in stronger vortices and lead to a greater chance for development (Gray 1968; Davis et al. 2008).

In the east Pacific basin, a low-level westerly jet is present from June through October (Figure 1.2). The presence of this jet, centered on approximately 10°N in the east Pacific basin, gives rise to persistent cyclonic wind shear to the north of the jet, which in conjunction with warm SSTs during the summer months produces favorable conditions for TC

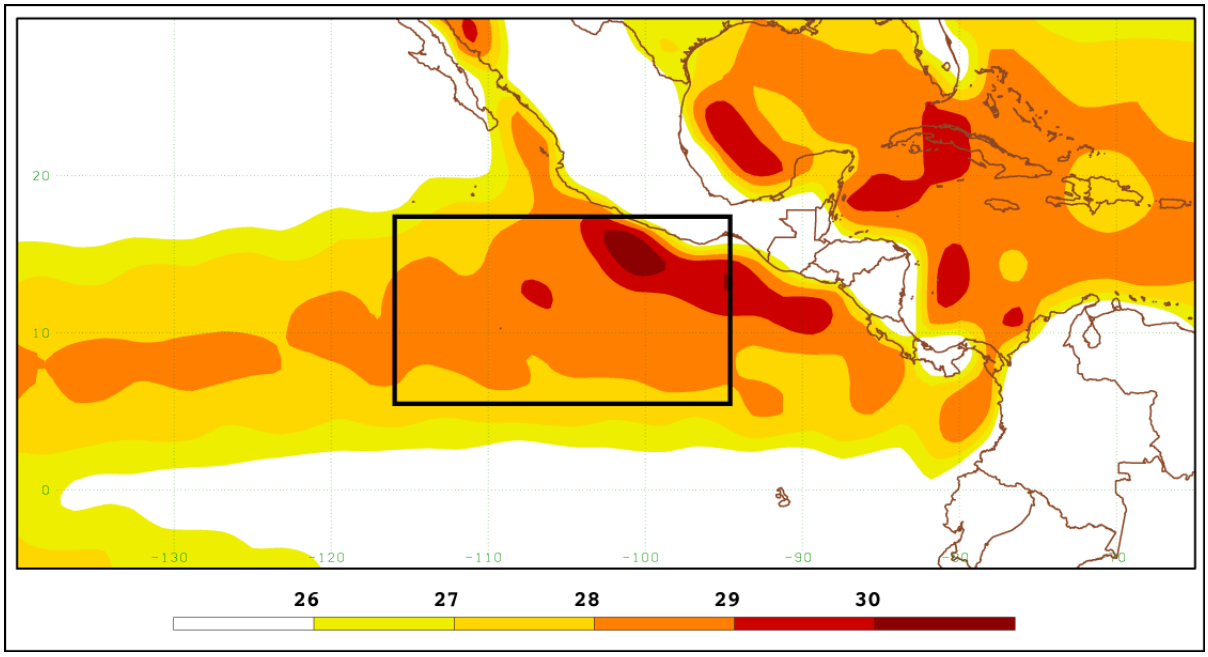

Figure 1.1 The sea surface temperature (in degrees C) climatology (1980-2010) for the peak (July-September) east Pacific hurricane season. The MDR for the east Pacific is denoted by the black box. A sea surface temperature minimum of 26˚C is favorable for TC genesis.

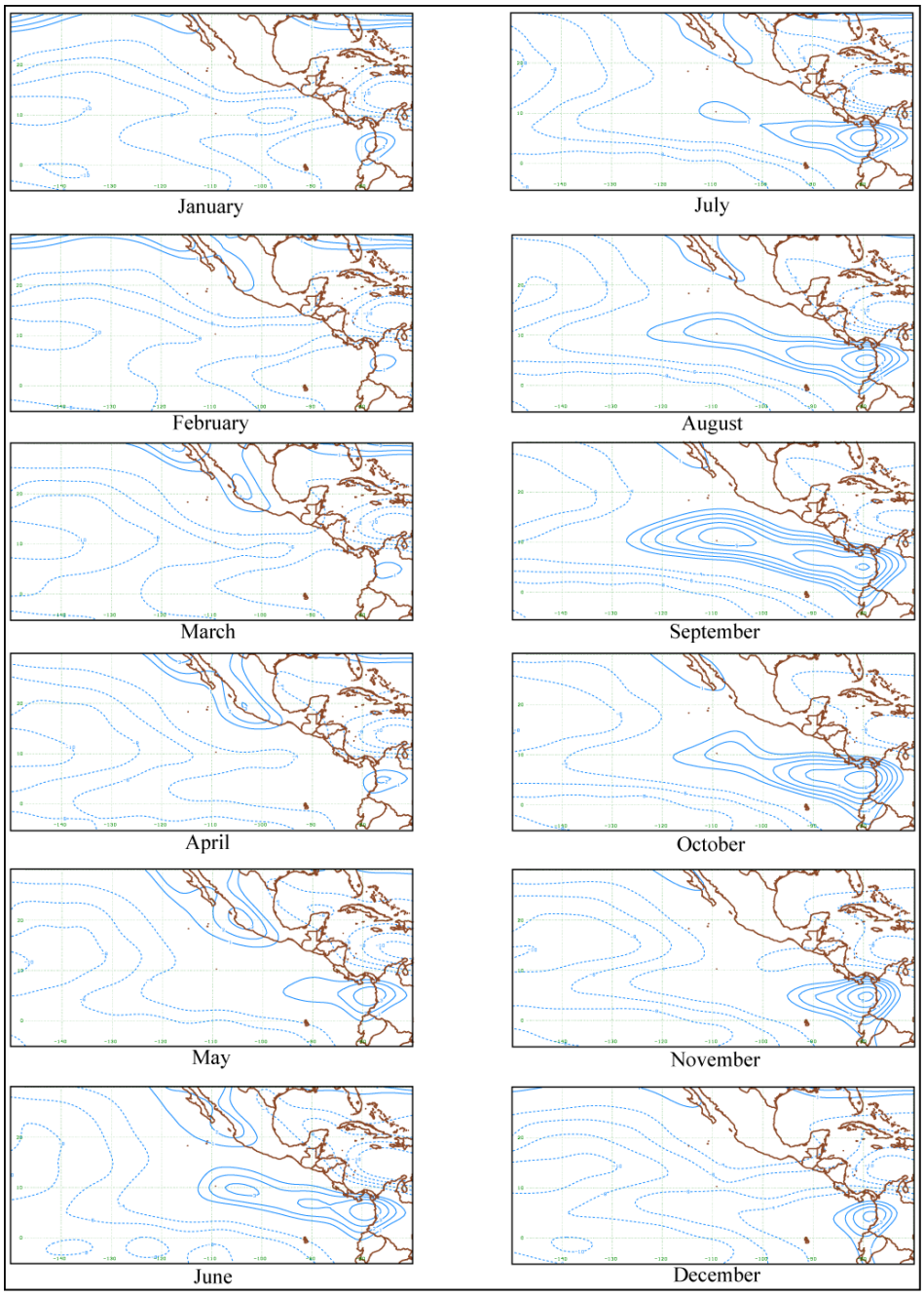

Figure 1.2 The 30-year (1980-2010) monthly climatology of the 900hPa zonal wind plotted in ms^{-1}. Dashed lines indicate easterly winds. The contour interval is $1ms^{-1}$ ($2ms^{-1}$) for westerly (easterly) winds, and the zero contours are omitted.

development in the basin. Since the eastern Pacific basin is located in a positive cyclonic environment, both the background vorticity (vorticity remaining after the vortex relative vorticity has been removed) and the relative vorticity of the vortex will be considered as parameters in the models to test their predictive skill for TC genesis.

c. Vertical wind shear

Vertical wind shear has a significant influence on TC genesis with weak vertical shear favoring development and strong shear inhibiting formation largely due to its effect on the development of the warm core vortex characteristic of a TC. Latent heat release associated with the convection in a low-level, nascent vortex warms the core of the vortex. In a strong vertical shear environment, heat would be advected away from the vortex by the shear (Gray 1968). This loss of heating would inhibit the development of a warm core for TC formation; therefore, it follows that a weak vertical shear environment would be more favorable for development as the heating of the vertical column above the low-level disturbance would be less difficult.

The effect of vertical wind shear on the development of TCs in the east Pacific will be considered from a number of perspectives in this study. The vertical shear over a vortex will be represented by four statistical parameters. These parameters will include the spatially averaged deep (900hPa-200hPa) and shallow (900hPa-500hPa) vertical shear of the zonal and meridional winds. These four measures of vertical shear will be used as a means to identify whether a specific type of shear is more influential on tropical cyclogenesis than others.

There are a number of physical mechanisms that can reduce the vertical shear surrounding a low-level vortex. Some of these mechanisms are on a scale that is too small to be resolved on a 1°x1° grid, and will not be considered in this study. For example, the

vertical momentum transport that occurs in the associated cumulus convection can counteract the increase in vertical shear resulting from heating in the core of the disturbance (Gray 1979). Other, larger-scale mechanisms (e.g. upper-tropospheric patterns and ENSO) can also weaken the environmental vertical wind shear. Some of these features will be considered in this study and will be discussed later.

d. Planetary vorticity

The magnitude of the planetary vorticity is directly related to the vorticity arguments mentioned in section (*b*) previously. In order to achieve the rotation needed for the vortex and surrounding environment, development can only take place at a sufficient distance from the equator (Palmén 1948; Gray 1968). Indeed, in the observations, TCs do not typically form within 5° degrees of the equator.

e. Relative humidity

The last environmental factor for TC genesis is the amount of moisture present in and around the initial low-level disturbance. Having a sufficiently moist atmosphere is necessary to sustain long-lasting convective systems. TCs are no exception as they only form in areas of high mid-tropospheric relative humidity (Gray 1968), and dry environments inhibit TC development. For example, TCs in the Atlantic basin were observed to weaken as they entered into dry atmospheric environments due to downdrafts created by the less buoyant dry air (Rotunno and Emanuel, 1987; Dunion and Velden 2004; Wu et al. 2006). Such downdrafts, accompanied by surface divergence located over a low-level vortex will act to spin-down the vortex vorticity inhibiting tropical cyclogenesis.

1.3 Mechanisms associated with genesis

In addition to the climatological conditions stated above, there are interactions amongst a number of dynamical features that can act to enhance or inhibit the development of TCs to varying degrees. Some of the mechanisms considered here include upper-tropospheric flow patterns, low-level zonal wind fields, barotropic growth, and the Madden-Julian Oscillation (MJO).

a. Upper-tropospheric flow patterns

Riehl (1948) was one of the first to emphasize the importance of upper-tropospheric flow to the formation of TCs, claiming that upper-tropospheric divergence drives the process by enhancing the low-level convergence and convection under it. Since then, a number of upper-tropospheric, large-scale flow patterns have been shown to influence tropical cyclone development. In the Atlantic basin, TC formation was favored in locations where the upper-tropospheric flow was southerly, such as on the eastern side of tropical upper-tropospheric troughs (TUTT), which are also regions of significant upper-level divergence (Colon and Nightingale 1963). TUTTs were also found to play a significant role in the development of a small number of typhoons in the west Pacific basin (Fu et al. 2007).

Building on Riehl's hypothesis that upper-level flow influences TC genesis, studies have found that in addition to enhanced convection, upper-level flows such at TUTTs and upper-tropospheric anti-cyclones can favorably influence genesis via the creation of an outflow channel for the low-level disturbance (Sadler 1976; Frank 1982; Fu et al. 2007) or decrease the vertical shear over the disturbance through its associated ridge (Sadler 1976).

The upper-tropospheric flow will be considered indirectly through deep layer (900hPa-200hPa) shears of the zonal and meridional winds. A surrogate for convection will also be examined through the southern branch of the in-up-out circulation of the vortex (the

900-200hPa equatorward vertical shear of the meridional wind). A more in depth investigation of the upper-level flow will be considered in the composite portion of the study.

b. Low-level zonal winds

In the east Pacific basin, the direction of zonal winds can be useful to differentiate between developing and non-developing vortices with anomalously westerly winds being favorable for development (Davis et al. 2008). In fact, periods of westerly anomalies over the basin produce twice as many vortices and TCs as those times in which the zonal winds are predominantly easterly over the basin (Maloney and Hartmann 2001; Davis et al. 2008). Furthermore, the existence of low-level (900hPa) horizontal stretching deformation was used to distinguish developing from non-developing vortices during the 2005-2006 peak hurricane seasons in the east Pacific basin (Davis et al. 2008). In the case of deformation, vortices embedded in a region of strong horizontal stretching deformation were significantly less likely to develop. A vortex located in an environment characterized by weak to non-existent deformation will be more likely to maintain its structure. However, in the presence of strong deformation, the structure of the vortex (especially during the early stages of its life cycle) is more likely to be destroyed as it is horizontally "stretched out" in the deformation field.

This study will investigate the predictive skill of some low-level (900hPa) wind features on TC genesis. The low-level zonal and meridional wind fields as well as the horizontal shearing, stretching, and total deformation fields will all be considered as parameters in the statistical models employed in the study.

c. Barotropic growth

Many studies have examined the role and influence of barotropic growth on the formation of TCs around the globe. Since the beginning stages of TC formation are typically

barotropic to first order (Shapiro 1978; Davidson and Hendon 1989), it is useful to consider the role of barotropic growth mechanisms in TC genesis. In the east Pacific basin, breakdowns of the inter-tropical convergence zone (ITCZ) may be the result of lower-tropospheric barotropic instability (Ferreira and Schubert 1997) as detailed later in this chapter.

It has been shown that for a low-level jet with intense horizontal shear, the genesis of a modeled TC vortex is strongly sensitive to initial condition perturbations that tilt upshear, taking advantage of the barotropic growth processes (Hoover 2010). Barotropic instability along the low level jet may act to intensify a favorably tilted vortex located on the northern cyclonic shear side by transferring momentum northward in order to conserve angular momentum (Lipps 1970).

While the interaction of the low-level jet and nascent vortices will be investigated in both the composites and case studies, a barotropic conversion term will be evaluated and used as a parameter for barotropic growth in the statistical models. Maloney and Hartmann (2001) considered four barotropic energy conversion terms according to the equation

$$C(\overline{K},K') = -(\frac{\partial \overline{u}}{\partial y}\overline{u'v'} + \frac{\partial \overline{v}}{\partial x}\overline{u'v'} + \frac{\partial \overline{u}}{\partial x}\overline{u'^2} + \frac{\partial \overline{v}}{\partial y}\overline{v'^2}) \quad \textbf{(1)}$$

where bars indicate climatological averages and primes are the deviations from the average. It was found that the first and third terms were important contributors to the total barotropic eddy kinetic energy conversion in the east Pacific basin. Since the current study will use zonally averaged fluxes (as opposed to climatologically averaged fluxes), only the first term will be used as a parameter in the statistical models.

d. Madden-Julian Oscillation (MJO)

Recently, much attention has been focused on the influence of the MJO on the formation and predictability of tropical cyclogenesis. The MJO is a 40 to 50 day oscillation that propagates zonally in the vicinity of the equator (Madden and Julian 1971). Since this oscillation affects many aspects of the atmosphere ranging from angular momentum to upper-tropospheric velocity potential to variations in convection (Madden and Julian 1994), finding a way to effectively measure its progression is no simple task. However, a widely accepted index for identifying the phases of MJO is that developed by Wheeler and Hendon (2004). To develop the MJO index, Wheeler and Hendon used an empirical orthogonal function (EOF) analysis of the outgoing long-wave radiation (OLR) fields and the 850hPa and 250hPa zonal wind anomalies to capture and identify the majority of the MJO's spatial variability. In the end, the first two EOFs of the MJO, called the real-time multivariate MJO (RMM) indices, were used in combination to determine the phase of the oscillation. This index describes the MJO according to eight phases as it propagates around the Earth.

While there have been studies devoted to the effects of the MJO in both the western Pacific and the eastern Pacific basins, much is still unknown about the influence of the MJO on tropical cyclogenesis, particularly in the eastern Pacific. However, the MJO index does shed some light on how the environment over the east Pacific varies according to the phase of the MJO. For instance phases 1, 2, and 8 all tend to place the region in an environment of low-level westerly winds with anomalously low OLR values. This tendency for an increase in convection (indicated by lower OLR), and westerly winds in the lower troposphere increases the favorability of TC formation in this region (Maloney and Hartmann 2000). On the other hand, during the remaining phases, the environment of the east Pacific basin is characterized by anomalous easterly winds and suppressed convection over the majority of the basin

(Wheeler and Hendon 2004) resulting in fewer TC genesis events (Maloney and Hartmann 2000).

The MJO has been shown to have predictive skill for tropical cyclogenesis on a time-scale of one to three weeks (Leroy and Wheeler 2008). However, this study focuses on using statistical models to predict genesis on the scale of one to two days. Favorable phases of the MJO (defined as phases 1, 2, and 8 for this study) have been shown to significantly increase the number of vortices and TCs compared to unfavorable phases (Maloney and Hartmann 2000; Barret and Leslie 2009). Therefore, this study will use the phase and RMM indices as parameters in the statistical models.

1.4 Other mechanisms

Many of the factors that have proven to be influential to TC genesis in the east Pacific basin have been reviewed here, but there are a number of other features that are influential to TC genesis. The following mechanisms will not be considered in the current study as they either do not have direct impacts on the east Pacific basin or are too complex to use as parameters in the statistical models. While the following features are not the focus of this study, they are important to TC genesis in general, and merit some explanation here.

a. El Niño Southern Oscillation (ENSO)

ENSO is the coupled relationship of the sea surface temperature (SST) variability off the west coast of South American, known as El Niño, and the fluctuations of surface pressure in the tropics known as the Southern Oscillation. This phenomenon is responsible for a variety of climate variability across the globe including changes in regional and global precipitation (Walker 1923, 1924; Ropelewski and Halpert 1987) and temperature

(Ropelewski and Halpert 1986) patterns. Considering the impact ENSO has on a global scale, it is no surprise that ENSO can also be connected to TC variability.

ENSO has been shown to have a significant impact on TC activity in the Atlantic basin, with the number of TCs drastically decreased in cold ENSO years versus warm ENSO years (Gray1984) and is thought to have been a leading contributor to the very active 1995 Atlantic hurricane season (Landsea 1998). The increase in TC activity over the Atlantic occurs from teleconnections between the eastern Pacific and Atlantic upper-level flows. As summarized by Gray (1984), during warm ENSO events, convection increases over the warm tongue in the eastern Pacific basin, which leads to an increase in the upper-level outflow over the basin. This increased outflow can act to weaken the vertical shear over the Atlantic basin by decreasing the easterly winds aloft.

While the effect of ENSO on Atlantic TC activity is evident, its effect on the number of TCs in other basins is not as clear (Gray 1984; Whitney and Hobgood 1997). The number of TCs that forms in the western Pacific basin does not have a clear, in-phase relationship to ENSO events (Gray 1984); however, there does appear to by a lagged relationship to the number of TCs and warm ENSO events (Chan 2000). The number of TCs that form in the Southern Hemisphere does not seem to be modulated by ENSO, but the locations for the formation of vortices are affected such that favorable regions are displaced further from the equator during cold ENSO events (Kuleshov et al. 2009).

Despite the effects of ENSO on tropical cyclogenesis in various basins, and its influence on eastern Pacific convection as evidenced by Gray (1984), no strong evidence exists to describe the direct influence of ENSO on eastern Pacific TC genesis. Also, the

influences of ENSO on TC activity is on a much longer time-scale than the 48 hour time-scale considered in this study. Therefore, ENSO is not considered in the present study.

b. Easterly Waves (EWs) and African Easterly Waves (AEWs)

Easterly waves are wave-like disturbances that propagate westward at phase speeds varying from 10-15ms^{-1} (Reed and Recker 1971; Thorncroft and Hoskins 1994) and have two origins in the Northern Hemisphere largely over the African continent and over the west Pacific basin (Gray 1968). Shapiro (1977) hypothesized that small, cyclonic circulations embedded in these waves could develop into TCs when non-linear contributions of vorticity advection into the cyclonic circulation become large. The strengthening of the cyclonic circulation in an environment that is characterized by warm SSTs, weak vertical wind shear, and high relative humidity may then lead to the development of a TC. Indeed, easterly waves have been shown to lead to TC genesis in many basins. Roughly one in five TC cases in the western Pacific 2000-2001 hurricane season were associated with easterly waves (Fu et al. 2007). Many of the TCs that initiate in the Atlantic MDR are associated with AEWs. Some of these AEWs are able to persist or be reinvigorated over the Caribbean and initiate TCs in the east and sometimes even west Pacific basins (Molinari et al. 2000).

While easterly waves can provide nascent vortices, which may develop into TCs, they are not always present during TC genesis events. Therefore, easterly waves are not a focal point of the current work.

c. Monsoon

Local phenomena such as the Indian Monsoon are more likely to maintain an environment that will act to inhibit or promote tropical cyclogenesis for their respective region. For example, there is evidence that years in which the monsoon is weak tend to yield

a 33% increase in the number of tropical cyclones that form in the Indian Ocean compared to years in which the monsoon in strong (Kumar and Krishnan 2005). In the western Pacific, the monsoon provides three favorable environmental flows - 1) the monsoon shear line, 2) the monsoon confluence zone, and 3) the monsoon gyre - in which vortices may develop (Ritchie and Holland 1999). In the Southern Hemisphere, the environment of the monsoon zone is favorable for genesis (Davidson and Hendon 1989), and the North American Monsoon can also lead to TC development through monsoon depressions in the East Pacific basin (Elsberrry and Harr, 2008).

While the monsoon has direct impacts on TC genesis across the globe including the east Pacific basin, it is not a focus of this study, though it may be a topic for future research.

d. ITCZ break down

The ITCZ is characterized by low level convergence and convection that are favorable for TC genesis. Based on observational analyses, 80% of TCs form in or near the ITCZ globally (Gray 1979). Breakdowns of the ITCZ can result in the formation a number of small, separate vortices (Agee, 1972, Thompson and Miller 1976), which – when embedded in a favorable environment – can develop into TCs. Breakdowns of the ITCZ occur from the deep convection that makes up the ITCZ. Intense latent heat release creates a cyclonic potential vorticity (PV) anomaly with a reversal in the gradient on the poleward side, which satisfies the criteria for barotropic and baroclinic instability (Charney and Stern 1962). The instability within the ITCZ can lead to undulations in the ITCZ leading to its breakdown and subsequent vortex formations (Ferreira and Schubert 1997). When this process occurs in the east Pacific basin during boreal summer, many of the environmental features may be favorable for those

vortices to develop into TCs. The breakdown in the ITCZ may also be an explanation for the spatial and temporal clustering observed in TC occurrence (Ferriera and Schubert 1997).

In the east Pacific, ITCZ breakdowns have been shown to result in TC formation within the basin (Gerrish and Mayfield 1989). While breakdowns in the ITCZ may result in nascent vortices that may develop into TCs, it is not within the scope of the study to attempt to predict ITCZ breakdown. Therefore, the breakdown of the ITCZ is not considered as a parameter for the statistical models, though the ITCZ breakdown scenarios may play a tangential role in the barotropic growth process as described previously.

e. Local terrain

One final feature that impacts east Pacific TC development specifically is the local topography, including the Sierra Madre of Mexico. The high topography over the Sierra Madre is directly influential to tropical cyclogenesis in the basin, specifically in its interaction with AEWs. It has been shown that interaction between the topography and AEWs can increase the low-level relative vorticity by 100% as the waves spin-up vorticity on the lee side of the mountains (Zehnder et al. 1999). While this vorticity spin-up may lead to down stream development of TCs, it is outside the scope of this study to include the topography in the statistical models.

1.5 Statistical methods

Recently, statistical methods such as linear discriminant analysis (LDA) and a probabilistic Bayesian adaptation have been successfully implemented to predict various physical aspects of TCs (e.g. Miller 1962; Dewey 1975; Perrone and Lowe 1986; Hennon and Hobgood 2003; Kerns and Zipser 2009; Kossin and Sitkowski 2009). Both of these methods are adapted into statistical models for this study. In addition, composites of developing and

non-developing vortices in the east Pacific basin are made from both vortex-centered and basin-wide perspectives.

a. Linear Discriminant Analysis

LDA is a statistical method used to classify two or more groups of events by finding a linear combination of parameters that demonstrates the greatest skill in separating the groups. This method divides a set of events into two or more predetermined groups by using a combination of variables, or parameters, which characterizes the events of each group. Linear combinations of these parameters are then tested to determine which combination of parameters produces the greatest distinction between the groups.

LDA has been widely used in both science and engineering fields as an efficient and effective tool in data classification. This method was applied in an early meteorological study to predict the occurrence of precipitation (Miller 1962). By the mid to late 1970s, LDA was used in a variety of meteorological studies including the development of a formula that could be used for cumulonimbus forecasts (Randerson 1977), major temperature anomalies in Russia (Bochkova 1978), and lake-effect snow fall probabilities in the Great Lakes region (Dewey 1975). With respect to tropical cyclogenesis, LDA has been used to identify intensifying vs. non-intensifying tropical cloud clusters in the western north Pacific producing promising results in which post agreement scores (percentages of correct forecasts) reached 80% for developing cases (Perrone and Lowe 1986). Another study (Hennon and Hobgood 2003) examined the feasibility of using LDA to forecast the probability of tropical depression formation from individual cloud clusters in the Atlantic basin using large-scale data. While statistical scores used to measure the accuracy of probabilistic forecasts were somewhat lower than the 1986 study, this technique once again proved to be an effective forecasting.

LDA has been used in several studies to examine TC development. Jennrich and Sampson (1977) used LDA to identify the parameters that best distinguished between developing and non-developing vortices. Perrone and Lowe (1986) built an LDA model by using a step-by-step method (known as stepwise linear discriminant analysis) to identify the large-scale variables – including vorticity, Coriolis parameter, potential temperature and divergence – that would yield the most accurate forecasts of tropical cyclogenesis up to three days before storm declaration using data on a 2.5^{o}x2.5^{o} grid. In a more recent study, Kerns and Zipser (2009) used LDA to distinguish between developing and non-developing disturbances in both the eastern Pacific and Atlantic basins. Their findings show that variables such as vorticity, deformation, mixing ratio, and wind shear all helped to distinguish disturbances in the eastern Pacific.

The current study uses a version of LDA that is used specifically to distinguish between two groups of events. This version on LDA is known as Fischer's linear discriminant analysis (FLDA), and it is used in this study to identify those vortex-centered variables that best distinguish developing from non-developing vortices in the eastern North Pacific basin. The parameters used in this method are determined by using a two-sample t-test, which calculates the statistical significance of the difference between the parameter means for each group. The data set in this study has a higher resolution of 1^{o}x1^{o} rather than the resolution used in both the 1986 and 2003 studies. This study will consider some of the variables used in the 2009 study (e.g. vorticity, deformation, and vertical wind shear) and include other features not used in the previous study.

b. Bayes method

The second statistical technique proposed in this study is the implementation of a naïve Bayes model to calculate probabilistic forecasts for TC genesis using a leave-one-season-out cross-validation. As with LDA, this method is used to classify two groups of events, but it does so by finding a *conditional probability* that an event belongs to one group rather than another based on a combination of parameters. While the naïve Bayes method give optimal results when the parameters used in the algorithm are independent of one another, it still gives good results when these conditions are not met (Zhang 2006; Hand and Yu 2001).

As with LDA, the use of a naïve Bayes model and classifier have been successfully employed in the atmospheric sciences. A Bayesian model was found to improve the seasonal forecast skill of the El Nino-3.4 index compared to empirical persistence forecasts and bias-corrected coupled model forecasts (Coelho et al. 2003). In relation to tropical meteorology, it was shown to be skillful as a probabilistic model, specifically for secondary eye-wall formation in eastern Pacific hurricanes (Kossin and Sitkowski 2009).

The current study will implement a naïve Bayes probabilistic model to test its predictive skill for TC genesis in the east Pacific basin using similar parameter sets as used for the LDA algorithm. There are many reasons to consider using a naïve Bayes classifier in addition to the LDA algorithm. One of the most appealing aspects is its probabilistic nature. While, the resultant discriminant function from the LDA algorithm can be converted to a probability in a number of objective ways, the output from Bayes method is already a probability, so no conversion from a classifier to a probabilistic forecast is needed. Second, the Bayes method does not rely on the parameter means from each group to determine whether a given vortex belongs to the DEV or ND group, as does LDA. The reliance on the

means between two groups could be problematic for some features that will be used in the LDA algorithm. For example, the phase of the Madden Julian Oscillation (MJO), which has been shown to have significant impacts on TC formation in the east Pacific basin (Hartmann and Maloney 2001), have similar means between the two groups, which may result in poor predictions from the LDA algorithm.

The results of the Bayesian model will be assessed using various statistical measures including Attributes diagrams, the Brier Skill Score, Threat Score, false alarm rate, and the probability of detection, (Wilks 2006) and the Briggs and Ruppert Skill Score (Briggs and Ruppert 2006).

c. Case Studies and Composites

While testing the predictive skill of these models in relation to TC genesis has its own practical merit, it is also important to understand the underlying physical mechanisms behind the formation of these storms. Therefore, in addition to using the statistical models to predict genesis, they will also be used to identify those variables that have the most influence on tropical cyclogenesis. Once the variables found to distinguish between DEV and ND cases are identified, they will be investigated as a means to better understand the physics behind the genesis of TCs. This study will use two approaches to that end.

Both vortex-centered and basin-wide composites of the strongest 25 DEV (those vortices with the strongest relative vorticity averaged over a 5˚x5˚ box centered on the vortex) and weakest 25 ND (those vortices with the weakest relative vorticity) vortices will be produced to determine which physical mechanisms influence the genesis and non-development of TCs. Vortex-centered composites will include those variables used in the statistical algorithms to gain further insight into their skill, or lack thereof, in predicting TC

genesis. Each variable for the DEV cases will be composited in 12-hour increments from the time of declaration as a tropical storm to 48 hours prior to declaration. The ND cases will be composited at one time, 24 hours after identification.

Each of the aforementioned statistical methods will focus on features that clearly distinguish between developing and non-developing vortices in the east Pacific basin and show promise in skillfully predicting TC genesis. This study will use a selection of physical features and mechanisms that have been shown to have a significant impact on TC genesis and distinguish between developing and non-developing vortices in the east Pacific basin.

This work is organized as follows. Chapter 2 focuses on the use of an LDA algorithm as a statistical forecast model as well as a means to identify features influential to tropical cyclogenesis in the east Pacific basin. Those features are then used in a probabilistic forecast model developed from a Bayes method. The results from this Bayesian model are presented in Chapter 3. Chapter 4 uses the findings of the two statistical models in conjunction with other large-scale features to build a composite-based conceptual framework of the physical characteristics that tend to favor tropical storm formation. (Chapters 2 through 4 are written with the intent of eventually becoming stand-alone published papers; therefore, some of the text in these chapters may seem repetitive). An analysis of the most important parameters, their predictive power in a statistical framework, and their physical relationship to tropical cyclogenesis are discussed in Chapter 5 along with possible directions for future research.

Chapter 2: A statistical approach to predicting tropical cyclogenesis in the eastern Pacific basin: Linear Discriminant Analysis

2.1. Introduction

a. Large-scale dynamical predictors of tropical cyclogenesis in the east Pacific basin

While eastern Pacific tropical cyclones (TCs) have historically received less attention from researchers relative to other basins, they have recently garnered more attention from the scientific community, specifically with regard to the contributing factors of their genesis. One study (Davis et al. 2008) used a vortex-based perspective to identify numerous factors that favorably influence the genesis of these storms. This study used the vortices that developed in the region between approximately 95^{o}W and 115^{o}W, and 5^{o}N and 20^{o}N during the peak seasons (1 July to 30 September) of both the 2005 and 2006 eastern Pacific hurricane seasons to determine the effects of various large scale, environmental factors on their development. Factors such as horizontal stretching deformation, vertical wind shear, vorticity, horizontal low-level winds and a Hadley cell type circulation were all compared for both developing and non-developing cases in a vortex-centered frame-work.

The development of individual cases was found to be greatly influenced by the initial strength of the vortex, with cases exhibiting initial average vorticities greater than $7.0 \times 10^{-5} s^{-1}$ all becoming TCs, and cases with vorticities less than $5 \times 10^{-5} s^{-1}$ rarely developing. Lower-tropospheric horizontal stretching deformation also appeared to be influential to development, with weaker deformation fields being favorable for genesis. Furthermore, it was found that no developing vortices occurred during periods of easterly wind regimes along 10^{o}N, but an increase in both developing and non-developing vortices occurred when a westerly wind

regime was present over the basin. In addition, the presence of a zonally confined Hadley cell-type circulation over the basin also seemed to be favorable for genesis. The existence of this type of circulation was found to be associated with relatively strong northerlies (southerlies) in the upper (lower) troposphere south of the vortex center.

The results of Davis et al. (2008) are compelling in that the consideration of certain variables from a vortex-centered perspective can lead to a distinction between developing and non-developing vortices in the eastern Pacific. This is an important finding that opens new avenues for the exploration of the predictability of eastern Pacific tropical cyclogenesis. Global numerical weather prediction (NWP) models are one of a number of tools used to forecast tropical cyclogenesis, providing timely and widely available data. However, the relatively coarse resolution of global NWP models creates a reliance on large-scale features to predict the genesis of TCs leading to limitations in forecast accuracy. Models such as the National Center of Environmental Prediction's (NCEP) Global Forecast System (GFS) have shown improved accuracy of forecasting tropical storm development from 2001-2006 (Pasch et al. 2006, 2008). However, it may be advantageous to use the storm-centered variables that Davis et al. (2008) found to distinguish between developing and non-developing vortices as a dynamical basis to predict the formation of TCs in the eastern Pacific basin using a statistical method. This study focuses on the use of a statistical technique known as linear discriminant analysis (LDA) to identify those common parameters at vortex-scale, which are most conducive for TC genesis.

b. Linear Discriminant Analysis

LDA is a statistical method used to classify two or more groups of events by empirically identifying a linear combination of parameters that demonstrates the greatest skill

in separating the groups. This method divides a set of events into two or more predetermined groups by using a combination of chosen parameters that characterize the state of members in each group. Linear combinations of these parameters are then tested to determine which combination of parameters produces the greatest distinction between the groups. When the chosen parameters were derived from the state at the time when the members are declared as members of each group, LDA is a classificatory system. When the parameters are derived from the state of members at an earlier time, LDA is a predictive system seeking to classify members into groups before declaration.

LDA has been used as an effective tool in data classification and prediction across a number of disciplines including the atmospheric sciences. One of the first meteorological studies to employ LDA was conducted by Miller (1962) in which this method was used to predict the occurrence of precipitation. In the 1970s, LDA was used in a variety of studies to predict a number of meteorological phenomena including lake-effect snowfall in the Great Lakes region (Dewey 1975), cumulonimbus development (Randerson 1977), and major temperature anomalies in Russia (Bochkova 1978). With respect to tropical meteorology, LDA has been used to identify intensifying vs. non-intensifying tropical cloud clusters in the western north Pacific and produced promising results with the percent of correct forecasts given by the model reaching 80% for developing cases (Perrone and Lowe 1986). Later, another study used large-scale variables as parameters in an LDA model to forecast the development of cloud clusters into tropical depressions in the Atlantic basin (Hennon and Hobgood 2003). The statistical scores used to measure the accuracy of the forecasts were somewhat lower in the 2003 study than the 1986 study, but this technique once again proved to be an effective forecasting tool. More recently, Kerns and Zipser (2009) used LDA to

distinguish between developing and non-developing disturbances in both the eastern Pacific and Atlantic basins. Their findings show that variables such as vorticity, deformation, mixing ratio, and wind shear all helped to distinguish disturbances in the eastern Pacific.

The use of the LDA algorithm in the current study is similar to the methodology used in Perrone and Lowe (1986), which built an LDA model by using a step-by-step method (called stepwise linear discriminant analysis) to identify the large-scale variables that would yield the most accurate forecasts of tropical cyclogenesis. The forecast window in this study went up to three days before storm declaration and used data on a 2.5^{o}x2.5^{o} grid. The current study also uses a version of LDA, known as Fischer's linear discriminant analysis (FLDA), which classifies a set of events into one of two predetermined groups, to identify those vortex-centered variables that are most skillful in distinguishing developing from non-developing vortices in the eastern North Pacific basin. However, rather than using an algorithm to determine which parameters to choose, each parameter is investigated individually by using a two-sample t-test to determine which parameters to test as linear combinations. Furthermore, the data set in this study has a higher resolution of 1^{o}x1^{o} rather than the resolution used in both the 1986 and 2003 studies. While some of the parameters in this study overlap with the earlier studies (e.g. relative vorticity, relative humidity, Coriolis parameter), some additional parameters such as the Okubo-Weiss parameter (defined as $\omega = D^2 - \zeta^2$, where D is the total deformation and ζ is the relative vorticity of the vortex), and MJO indices are also considered.

The data and methodology used in this study is discussed in section 2 of this chapter. The forecast results from the LDA algorithm are covered in section 3, while section 4 details

the physical interpretation of the model findings and conclusions and future work are presented in section 5.

2.2 Data and methodology

a. Data

The data used in this study are obtained from two sources: the National Center of Environmental Prediction (NCEP) six-hourly 1^{o}x1^{o} data,[1] for the period 1 July to 30 September for the 2001-2009 hurricane seasons, is used for the storm-centered set of variables, and archived Madden-Julian Oscillation (MJO) indices[2] (available once daily) are used for the MJO data. The National Hurricane Center's best track database and tropical cyclone reports[3] are used to determine which vortices developed in the east Pacific basin. The vortex-centered variables considered for use in the LDA algorithm included areal averaged vorticity, deformation, relative humidity, Coriolis parameter, and vertical wind shear (Davis et al. 2008). Variables used to characterize the MJO include the phase and amplitude of the oscillation (Wheeler and Hendon 2004).

The preprocessing program for the Fifth-Generation NCAR / Penn State Mesoscale Model (MM5), REGRIDDER, is used to create high resolution, three-dimensional data sets for the vortex-centered variables interpolated from the analysis data sets. Using the output from REGRIDDER, the vortices that formed in the east Pacific basin are identified and tracked. The variables associated with the vortices are averaged over a box centered on the

---

[1] The data for this study are from the Research Data Archive (RDA), maintained by the Computational and Information Systems Laboratory of the National Center for Atmospheric Research (NCAR). The original data are available from the RDA (http://dss.ucar.edu) in dataset number ds083.2.

[2] MJO indices were obtained from the archived text file available online at http://www.bom.gov.au/bmrc/clfor/cfstaff/matw/maproom/RMM/.

[3] NHC website: http://www.nhc.noaa.gov/pastall.shtml.

vortices. The full domain used for this study is centered at 15°N and 100°W with 65 grid points in the meridional direction and 141 grid points in the zonal direction. This results in a domain that encompasses the area from approximately 10°S to 38°N and 44°W to 156°W (Figure 2.1). Vortices considered in the study are confined to a smaller region bounded by the equator, the western coast of Mexico and Central America, 140˚W and 25˚N.

b. Group and parameter selection

An LDA system requires a pre-selected number of classification groups and a collection of classificatory/predictive parameters. Here, two classification groups for the study are selected: those vortices that developed into tropical storms (developing) and those that did not (non-developing). Using findings from previous studies, a set of dynamical variables that is thought to have skill in distinguishing between the chosen groups is selected for testing.

1) GROUP SELECTION

The simplest form of LDA separates data into two groups. For the purposes of this study, the vortices identified are separated into developing or non-developing categories. In order to assign each vortex to a specific category, all vortices that have a maximum vorticity of at least $1.5 \times 10^{-5} s^{-1}$ are tracked and compared against the NHC's best track database. All vortices in the data set that achieved the status of tropical storm are categorized as developing (DEV) cases. Vortices whose relative vorticity remains above $1.5 \times 10^{-5} s^{-1}$ for a minimum of 48 hours, but never reach tropical storm status are grouped as non-developing (ND) cases. Any vortices that fail to meet the above requirements are not used in this study.

2) PARAMETER SELECTION

Two types of variables were chosen for this study. The first type consists of storm-centered variables, whose values are averaged over an 11°x11° box centered on the vortex – with the exception of vorticity, which is averaged over a 5°x5° box (Figure 2.1) – and those variables associated with the MJO. The selection process is carried out in the following manner.

For both sets of parameters, a general list of relevant variables is compiled. While there are many factors that may contribute to the formation of tropical cyclones, the key to this method is to find the features that best separate developing from non-developing vortices. Some factors that have been demonstrated to influence tropical cyclogenesis include latitude (as it relates to the Coriolis parameter), 925hPa vorticity, 900-200hPa vertical wind shear, sea surface temperatures (SSTs), and relative humidity (Gray 1977). In the eastern Pacific, low-level jets, and enhanced Hadley cell type circulations have also been shown to provide favorable conditions for tropical cyclogenesis (Davis et al. 2008). An initial set of 27 environmental variables is constructed as listed along with their abbreviations in (Tables 2.1 and 2.2).

The listed variables are chosen because of their potential to clearly distinguish between the two chosen groups. SSTs are not included because the magnitude with which the SSTs would vary over the time and spatial scales considered in this study are not thought to provide significant distinction between the developing and non-developing events. However, testing the viability of SSTs to distinguish between these groups of events may be considered in future research.

The (equatorward) Hadley cell-type circulation has three branches: lower-tropospheric convergence fed by southerly cross-equatorial flow, divergent equatorward upper-

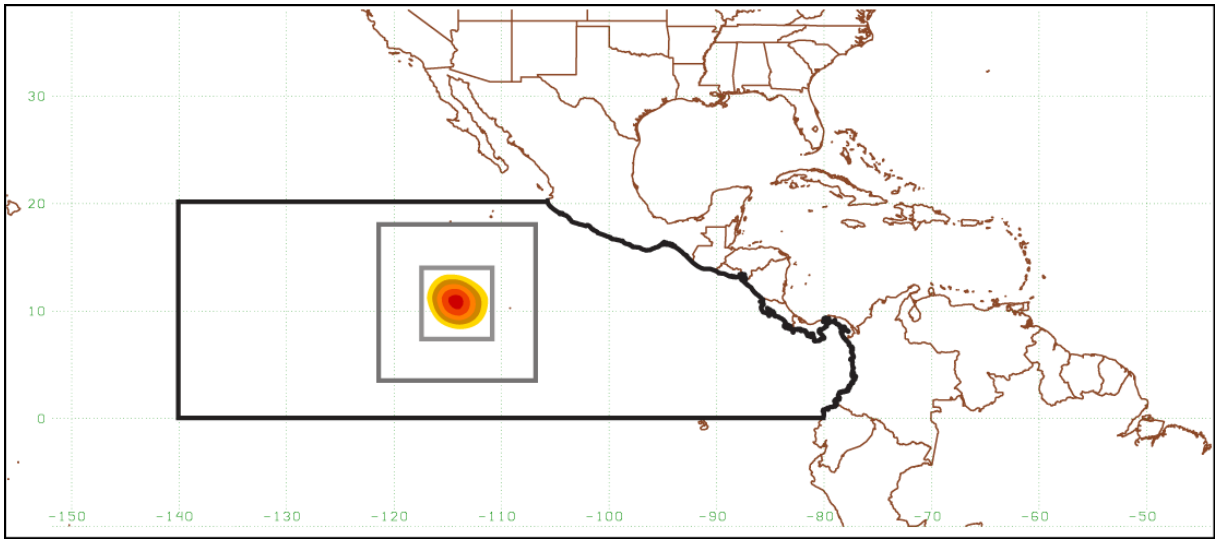

Figure 2.1 The full domain used in the MM5 regridder. The black line indicates the domain in which vortices considered for this study developed. The small and large gray boxes represent the 5˚x5˚ and 11˚x11˚ areal boxes over which vortex-centered variables are averaged.

Parameter Name	Abbreviation	Parameter Name	Abbreviation
latitude	ϕ	900hPa horizontal stretching deformation	$\overline{D}_{str}$
longitude	θ	900hPa horizontal shearing deformation	$\overline{D}_{shr}$
900hPa relative vorticity	$\overline{\zeta}$	900hPa environmental relative vorticity	$\overline{\zeta}_{bk}$
zonal extent of the vortex	$\overline{\xi}\Big\|_W^E$	900hPa environmental total deformation	$\overline{D}_{bk}$
meridional extent of the vortex	$\overline{\xi}\Big\|_S^N$	900hPa environmental horizontal stretching deformation	$\overline{D}_{bkstr}$
900hPa zonal wind	$\overline{u}$	900hPa environmental horizontal shearing deformation	$\overline{D}_{bkshr}$
900hPa meridional wind	$\overline{v}$	Okubo-Weiss parameter	ω
900hPa-500hPa vertical shear of the zonal wind	$\frac{\partial u}{\partial p}\Big\|_{900hPa}^{500hPa}$	energy generation rate of the vortex	$\mathbf{E \bullet D}$
900hPa-500hPa vertical shear of the meridional wind	$\frac{\partial v}{\partial p}\Big\|_{900hPa}^{500hPa}$	realtime multivariate MJO index 1	$RMM1$
900hPa-200hPa vertical shear of the zonal wind	$\frac{\partial u}{\partial p}\Big\|_{900hPa}^{200hPa}$	realtime multivariate MJO index 2	$RMM2$
900hPa-200hPa vertical shear of the meridional wind	$\frac{\partial v}{\partial p}\Big\|_{900hPa}^{200hPa}$	phase of the MJO	φ_{MJO}
900hPa-200hPa equatorward vertical shear of the meridional wind	$\frac{\partial v_{eq}}{\partial p}\Big\|_{900hPa}^{500hPa}$	amplitude of the MJO	ϑ_{MJO}
900hPa-500hPa relative humidity	RH	meridional barotropic conversion term	$-\frac{\partial \overline{U}}{\partial y}\overline{u'v'}$
900hPa total horizontal deformation	$\overline{D}$		

Table 2.1 The full name of each parameter considered for inclusion in the LDA algorithm along with their symbols.

alltime99owp		ω $\overline{\xi}$ $\overline{D}$ $\overline{u}$ $\frac{\partial v_{cg}}{\partial p}\Big\vert_{900hPa}^{500hPa}$ RH ϕ
alltime95owp		ω $\overline{\xi}$ $\overline{D}$ $\overline{u}$ $\frac{\partial v_{cg}}{\partial p}\Big\vert_{900hPa}^{500hPa}$ RH ϕ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $\overline{D}_{str}$
spectime99	m48	alltime99 + $\overline{D}_{str}$ $\overline{D}_{shr}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $\frac{\partial v}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $\overline{D}_{bkstr}$ $-\frac{\partial \overline{U}}{\partial y}\overline{u'v'}$
	m42	alltime99 + $\overline{D}_{str}$ $\overline{D}_{shr}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $\frac{\partial v}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $\overline{D}_{bkstr}$ $-\frac{\partial \overline{U}}{\partial y}\overline{u'v'}$
	m36	alltime99 + $\overline{D}_{str}$ $\overline{v}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $\frac{\partial v}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $\overline{D}_{bkstr}$ $-\frac{\partial \overline{U}}{\partial y}\overline{u'v'}$
	m30	alltime99 + $\overline{D}_{str}$ $\overline{v}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $\frac{\partial v}{\partial p}\Big\vert_{900hPa}^{200hPa}$
	m24	alltime99 + $\overline{D}_{str}$ $\overline{v}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $\frac{\partial v}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $\frac{\partial v}{\partial p}\Big\vert_{900hPa}^{200hPa}$
	m18	alltime99 + $\overline{v}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $\frac{\partial v}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $\frac{\partial v}{\partial p}\Big\vert_{900hPa}^{200hPa}$
	m12	alltime99 + $\overline{v}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $\frac{\partial v}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $\frac{\partial v}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $-\frac{\partial \overline{U}}{\partial y}\overline{u'v'}$
	m06	alltime99 + $\overline{D}_{str}$ $\overline{v}$ $\frac{\partial v}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $\overline{D}_{bk}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{500hPa}$
	m00	alltime99 + $\overline{D}_{str}$ $\overline{v}$ $\frac{\partial v}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $-\frac{\partial \overline{U}}{\partial y}\overline{u'v'}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{500hPa}$
spectime95	m48	alltime95 + $\overline{D}_{str}$ $\overline{D}_{shr}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $\frac{\partial v}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $\overline{D}_{bkstr}$ $-\frac{\partial \overline{U}}{\partial y}\overline{u'v'}$ $\overline{D}_{bk}$
	m42	alltime95 + $\overline{D}_{str}$ $\overline{D}_{shr}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $\frac{\partial v}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $\overline{D}_{bkstr}$ $-\frac{\partial \overline{U}}{\partial y}\overline{u'v'}$ $\overline{v}$ $\overline{D}_{bk}$
	m36	alltime95 + $\overline{D}_{str}$ $\overline{v}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $\frac{\partial v}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $\overline{D}_{bkstr}$ $-\frac{\partial \overline{U}}{\partial y}\overline{u'v'}$
	m30	alltime95 + $\overline{D}_{str}$ $\overline{v}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $\frac{\partial v}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $\overline{D}_{bkstr}$
	m24	alltime95 + $\overline{D}_{str}$ $\overline{v}$ $\frac{\partial v}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $\frac{\partial v}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $RMM1$
	m18	alltime95 + $\overline{v}$ $\frac{\partial v}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $\frac{\partial v}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $\overline{D}_{str}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $-\frac{\partial \overline{U}}{\partial y}\overline{u'v'}$ $\overline{\xi}_{bk}$
	m12	alltime95 + $\overline{v}$ $\overline{D}_{str}$ $\frac{\partial v}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $\frac{\partial v}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $-\frac{\partial \overline{U}}{\partial y}\overline{u'v'}$
	m06	alltime95 + $\overline{v}$ $\overline{D}_{str}$ $\frac{\partial v}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $\overline{D}_{bk}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{200hPa}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $-\frac{\partial \overline{U}}{\partial y}\overline{u'v'}$ $\overline{\xi}_{bk}$ $RMM1$
	m00	alltime99 + $\overline{D}_{str}$ $\overline{v}$ $\frac{\partial v}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $-\frac{\partial \overline{U}}{\partial y}\overline{u'v'}$ $\frac{\partial u}{\partial p}\Big\vert_{900hPa}^{500hPa}$ $\overline{D}_{bk}$

Table 2.2 The four parameter sets tested in the LDA algorithm.

tropospheric flow, and an ascending branch at the latitude of the vortex. Since the parameters used in the LDA method must each be represented by a single number, the three branches of the Hadley cell type circulation need to be summarized in such a fashion. These three branches are closely linked to one another such that a surrogate for the Hadley cell type circulation can be found. Indeed, the presence of strong, deep, northerly vertical wind shear south of developing vortices is indicative of a Hadley cell-like circulation (Davis et al. 2008). Therefore, this variable, which is the average of the 900hPa to 200hPa vertical shear of the meridional wind over an 11^{o}x11^{o} box whose northern boundary crosses through the center of the vortex, is included in the variable list to act as a signal for the circulation.

Recently, more emphasis has been placed on the role of the MJO and its possible influence on creating favorable or unfavorable conditions for TC development (e.g. Maloney and Hartmann, 2001; Camargo et al. 2008). The MJO is an eastward propagating, equatorially confined 40-50 day oscillation of many atmospheric characteristics including zonal 850hPa wind anomalies that propagate zonally about the equator (Madden and Julian, 1971). Wheeler and Hendon (2004) developed an MJO index using empirical orthogonal function (EOF) analysis to track the propagation of the MJO in the tropics. The evolution of the MJO can be described in eight cyclical phases, which affect each tropical basin around the globe; in the eastern Pacific, Phases 1, 2, and 8 of the MJO produce the most favorable conditions for tropical cyclogenesis in the eastern Pacific basin including the presence of westerly zonal winds and enhanced convection (Figure 2.2). The parameters chosen to represent the MJO for this study are obtained from Wheeler's MJO index, specifically the first and second real-time multivariate MJO indices (RMM1 and RMM2) as well the phase and amplitude of the MJO.

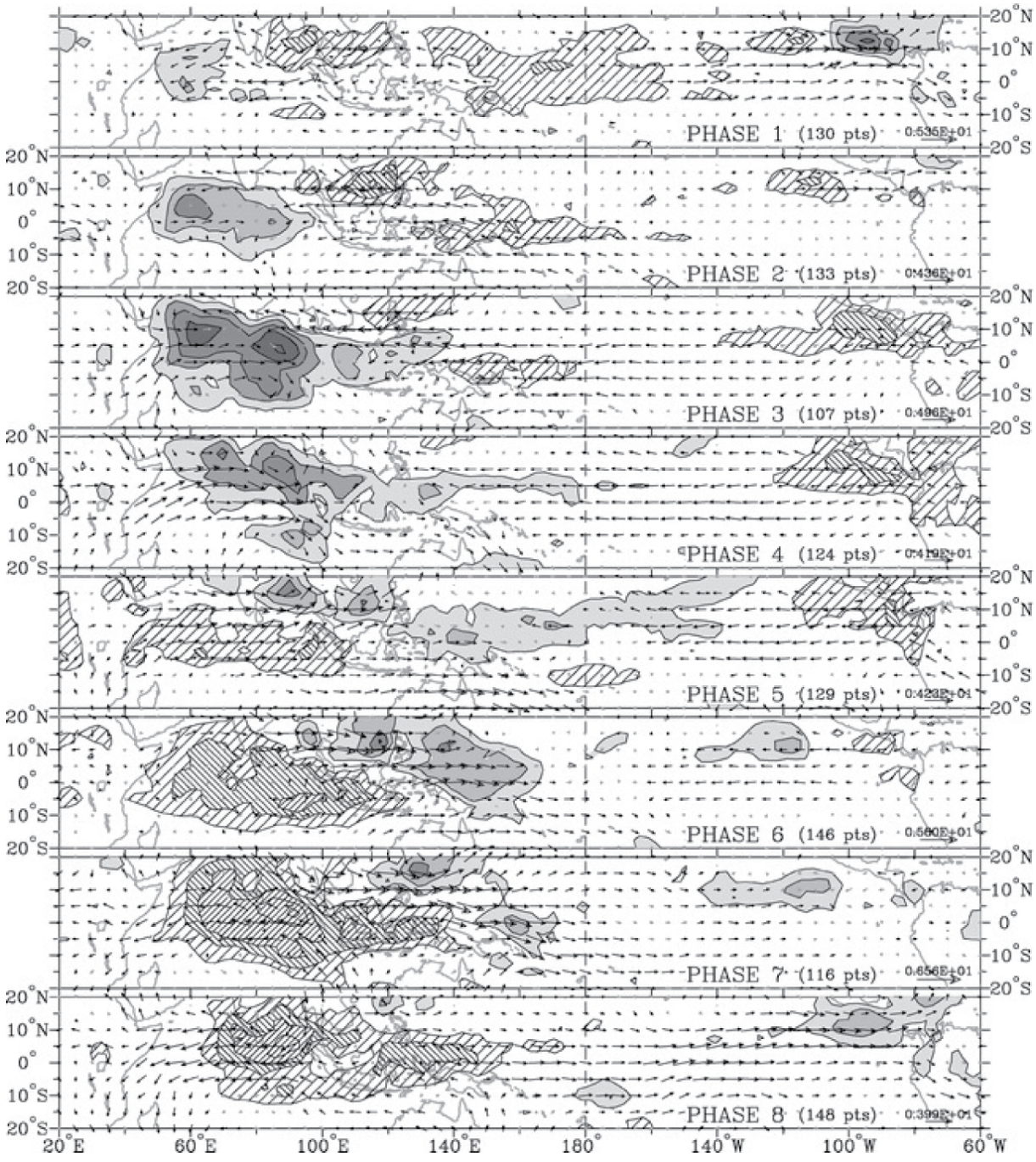

Figure 2.2 (Figure 9 from Wheeler and Hendon 2004) Anomalous 850hPa zonal winds (arrows) and OLR (shaded). Solid shading is anomalously low OLR, while hatching indicates regions of anomalously high values of OLR.

Each of the vortex-centered variables used as parameters for this study represent averaged values for the first 48 hours of their life. Since the MJO variables change slowly, on the order of days, it is not considered necessary to use averaged values for the variables used for the MJO parameters. All of the chosen variables are individually applied to the LDA algorithm to find the combination that resulted in the best classification skill.

c. Fisher's linear discriminant analysis (FLDA)

LDA is a statistical tool that can be used to classify a set of events into one of two or more pre-identified groups (e.g. Kerns and Zipser 2009). First, a calibration algorithm is formulated from events whose classifications are known. A linear combination of parameters associated with these events is then used to produce a discriminant function that yields the greatest separation between the groups of events. The classification skill of this discriminant function can then be tested by using it to classify an independent sample set of events.

This study uses Fisher's LDA, which identifies a linear combination of parameters to divide a set of events into two predetermined groups maximizing the difference between the average values of each group. Two groups labeled as developing (DEV) and non-developing (ND) are used in this study. The general process used is as follows.

All of vortices detected during the peak of the eastern Pacific hurricane season (1 July to 30 September) from 2001 to 2009 are divided into developing (n_{DEV}) and non-developing (n_{ND}) cases. A matrix of environmental variables, $\mathbf{x}_{DEV}$ and $\bar{\mathbf{x}}_{ND}$, were constructed for each group with columns of parameters and rows of events, and their vector averages ($\bar{\mathbf{x}}_{DEV}$ and $\bar{\mathbf{x}}_{ND}$) are calculated separately. The covariance matrices for each group are defined as:

$$\mathbf{C}_{DEV} = \frac{1}{n_{DEV} - 1}(\mathbf{x}_{DEV} - \bar{\mathbf{x}}_{DEV})^{\mathbf{T}}(\mathbf{x}_{DEV} - \bar{\mathbf{x}}_{DEV}) \qquad \textbf{(1)}$$

and

$$\mathbf{C}_{ND} = \frac{1}{n_{ND} - 1}(\mathbf{x}_{ND} - \bar{\mathbf{x}}_{ND})^{\mathbf{T}}(\mathbf{x}_{ND} - \bar{\mathbf{x}}_{ND}). \quad \textbf{(2)}$$

Since $n_D \neq n_{ND}$, it is necessary to calculate a common covariance between the two groups. Following Wilks (2006, eqn. 10.39) the pooled common covariance matrix, $\mathbf{C}_{\text{pool}}$, is calculated as:

$$\mathbf{C}_{pool} = \frac{n_{DEV} - 1}{n_{DEV} + n_{ND} - 2}\mathbf{C}_{DEV} + \frac{n_{ND} - 1}{n_{DEV} + n_{ND} - 2}\mathbf{C}_{ND}. \quad \textbf{(3)}$$

In order to maximize the difference between the means of the two chosen groups, a discriminant vector, **a**, is used to determine the direction of maximum separation between the two means. Using the common covariance matrix, vector **a** can be found using equation 13.3 of Wilks (2006):

$$\mathbf{a} = [\mathbf{C}_{pool}]^{-1}(\bar{\mathbf{x}}_{DEV} - \bar{\mathbf{x}}_{ND})^{\mathbf{T}}. \quad \textbf{(4)}$$

From (4), the discriminant functions (scalar quantities representing the projection of vectors $\mathbf{x}_{DEV}$ and $\mathbf{x}_{ND}$ onto **a**) are calculated for both the DEV and ND groups as:

$$d_{DEV} = \mathbf{a}^{\mathbf{T}}\mathbf{x}_{DEV} \quad \textbf{(5a)}$$

$$d_{ND} = \mathbf{a}^{\mathbf{T}}\mathbf{x}_{ND} \quad \textbf{(5b)}.$$

The discriminant functions can be used to determine into which group an event should be classified by comparing the discriminant function of a single event to the midpoint, which is the projection of the average of $\bar{\mathbf{x}}_{DEV}$ and $\mathbf{x}_{ND}$ onto the vector **a** (Figure 2.3). If the discriminant function for a particular event is greater than or equal to the value of the midpoint, it will be classified into the first group. If the discriminant function is less than the value of the midpoint, it will be classified into the second group.

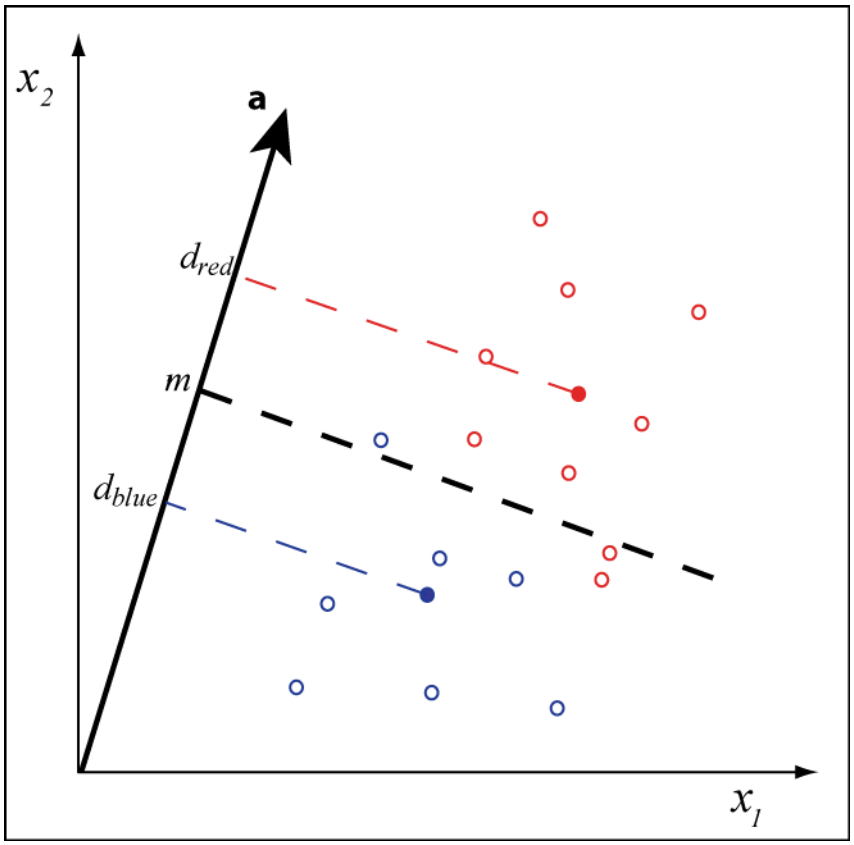

Figure 2.3 A geometric representation (adapted from Wilks 2006) of the calculation of the discriminant function *d* for a simple case that considers only two parameters (x_1 and x_2). The midpoint, *m*, is represented by the thick black dashed line, while the means of two example groups are represented by the blue and red dashed lines. The discriminant function for the red group (d_{red}) is the projection of the red group mean (represented by solid red circle) onto the vector **a**. Likewise, the discriminant function for the blue group (d_{blue}) is the projection of the blue group mean (represented by solid blue circle) onto the vector **a**.

The calculation of the discriminant function as takes into account both the midpoint and the probability that a case would be classified as ND rather than DEV. Therefore, the final discriminant functions are defined as

$$d'_{DEV} = \mathbf{a}^{\mathrm{T}}\mathbf{x}_{DEV} - m - p \quad \textbf{(6a)}$$

$$d'_{ND} = \mathbf{a}^{\mathrm{T}}\mathbf{x}_{ND} - m - p \quad \textbf{(6b)},$$

where m is the midpoint of the averaged data set and p is the ratio of ND to DEV events. Using these relationships the classifications are determined by the sign of the resulting d' such that if d' is positive, the vortex is classified as a DEV case. Likewise, vortices with negative values of d' are categorized as ND cases.

2.3. Forecast Results

a. Parameter sets

The skill of the LDA algorithm is tested at nine times, from the time of tropical storm declaration to 48 hours prior to declaration in increments of six hours. Four parameter sets (Table 2.2) are determined by a two-sample t-test, which tests the statistical significance of the difference between the means of two groups. First, the parameter sets are divided according to those that pass the t-test at the 99% confidence level and those that pass at the 95% confidence level. Parameters that pass at the 95% confidence level are also included because the LDA algorithm tends to perform better when more parameters are used. Since parameters that pass at the 95% confidence level have statistically significant differences between the means, there inclusion is considered to be acceptable.

The second discriminating factor concerns the frequency with which a given parameter passes the required confidence level. Those parameters whose means are statistically significantly different for *all* times are considered as candidates for one parameter

set. For example, any parameter that passes the t-test at the 99% confidence level for all of the nine forecast times is added one at a time to a parameter set. In addition, any parameters that decrease the error percentage of the algorithm results are kept, while those that do not add skill are rejected. Using the 99% confidence level as an example, if a parameter that passes the t-test at the 99% confidence level is added to the set and does not decrease the error percentage of the forecasts produced by the LDA algorithm, it is rejected from the parameter set. However, if a parameter that passes at the required confidence level is added and decreases the error percentage by any amount, it is retained in the parameter set. The parameter sets, which meet these criteria, are labeled as *alltime*. Parameter sets with the *alltime* identifier have a fixed number of parameters for all forecast times.

The other two parameter sets (labeled as *spectime*) are allowed to have the number of parameters change at each forecast time. For these two groups, all parameters are put through a t-test at each time, just as for the other groups. However, if a parameter does not pass the required t-test confidence threshold for all nine times, it is not automatically eliminated. Instead, it is included for anytime in which it passes the required threshold. For example, if a parameter passes the t-test at the 99% confidence level only at the 24 and 36-hour forecasts, it will be included in the parameter set at those two times only. This means that the number of parameters in the *spectime* sets will vary at each forecast time. Table 2.2 lists the parameters included at each time for both the *spectime99* and *spectime95* parameter sets.

b. Skill Scores

All four parameter sets are tested with the LDA algorithm at each of the nine forecast times. All nine seasons of data (2001-2009) are used in the forecast, and a leave-one-season-out cross validation technique is employed. The errors for both the DEV and ND groups are

calculated for each time simply by taking the number of incorrect DEV (ND) forecasts and dividing by the total number of observed DEV (ND) cases. The ND group errors (not shown) show little change for the forecast times, ranging from approximately 0.1 to 0.15.

The DEV group errors (Figure 2.4) prove to be more interesting. From the time of declaration to the 36 hour forecast, the errors are very similar for all parameter sets, ranging from approximately 0.03 to 0.21. However, beyond the 36 hour forecast, there is a noticeable divergence in error from the *alltime99* group and the remaining three. The 48-hour forecast error for *alltime99* is 0.445, while the error for *alltime95* is reduced to 0.284. The difference between these two parameter sets is the addition of the horizontal stretching deformation and the shallow vertical shear of the zonal wind (Table 2.1). This indicates that these features skillfully distinguish between DEV and ND vortices, especially at longer lead times.

There is also a slight decrease in error at the two furthest forecast times for the *spectime99* and *spectime95* compared to *alltime95*, which may again be due to the increase in the number of parameters in the former two sets at those times. The largest improvement occurs at the 42 hour forecast from *spectime99*.

Since the error alone does not give a full representation of the forecast skill of the model, it is important to consider a variety of skill scores to get a complete assessment. Some of the common reliability statistics (Figure 2.5) skill scores (Figure 2.6) used to assess model skill include the Threat Score (ThS), Hit Rate (HR), False Alarm Rate (FR), False Alarm Ratio (FAR), Heidke Skill Score (HSS), Pierce Skill Score (PSS), and Briggs and Ruppert Skill Score (BRSS). All of these skill measures are calculated from 2x2 contingency tables, which display the results of binary forecasts for a dichotomous set of events (Figure 2.7). The table consists of four elements: hits (number of correct Yes forecasts), false alarms (number

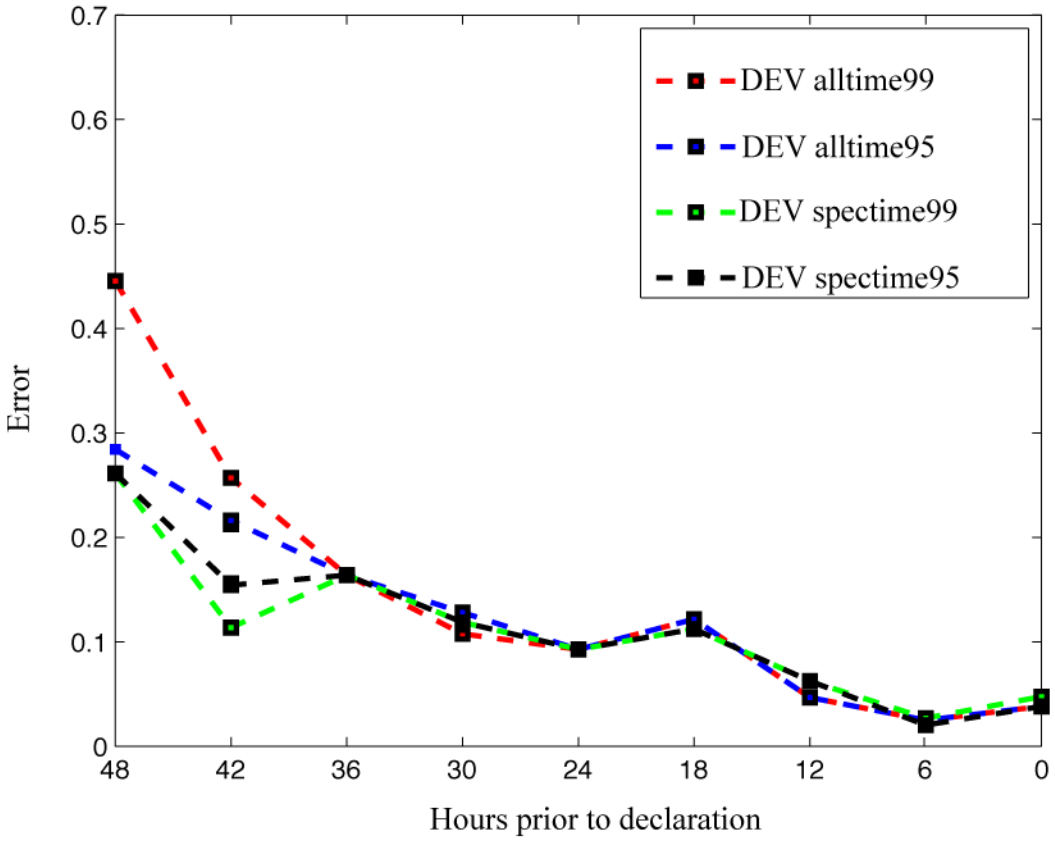

Figure 2.4 The errors (for the DEV cases) of each of the four sets of parameters tested with the LDA algorithm. Errors are plotted for each of the nine forecast times (i.e. 48 hours prior to declaration is the 48 hour forecast time).

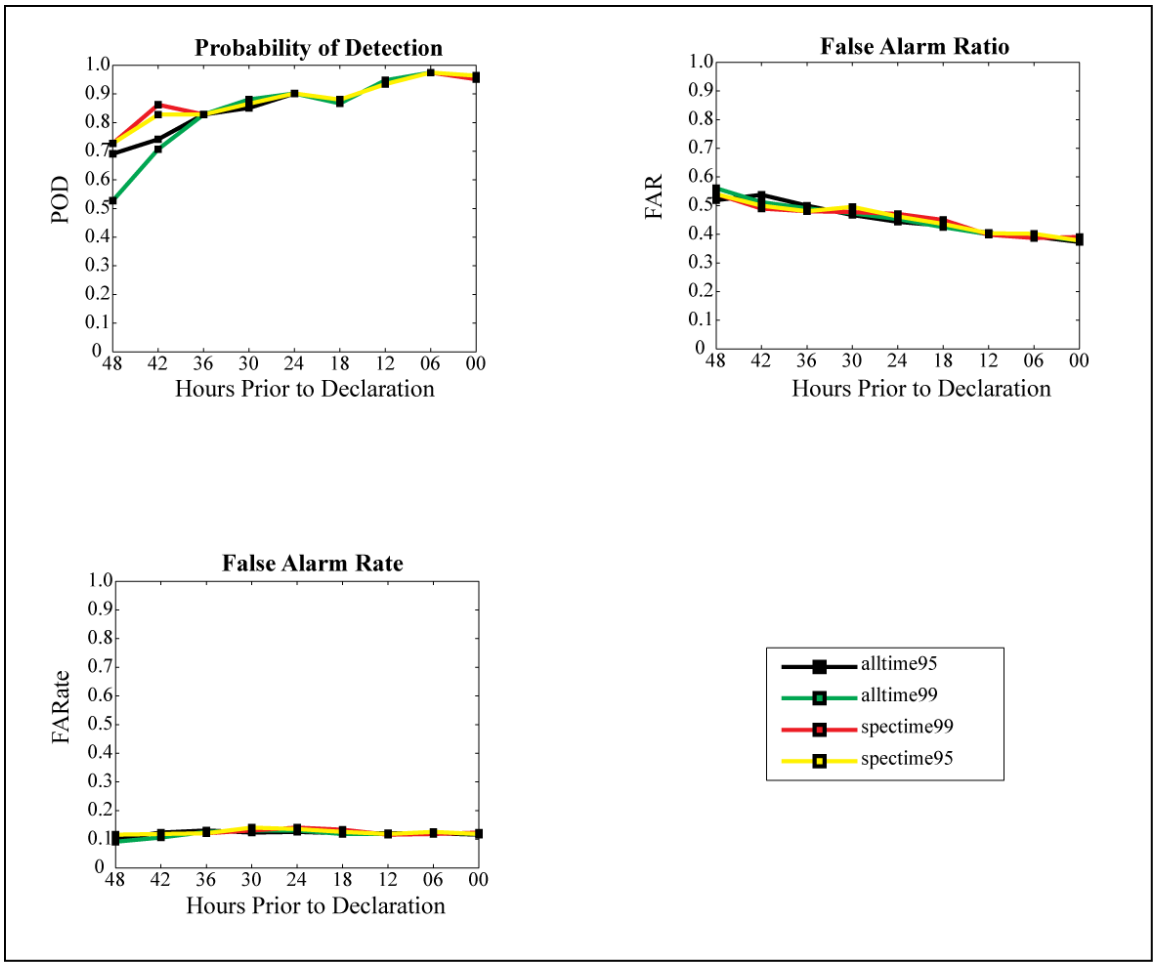

Figure 2.5 The Hit Rate (or POD), False Alarm Rate, and False Alarm Ratio are presented for all four sets of parameters at each of the nine forecast times.

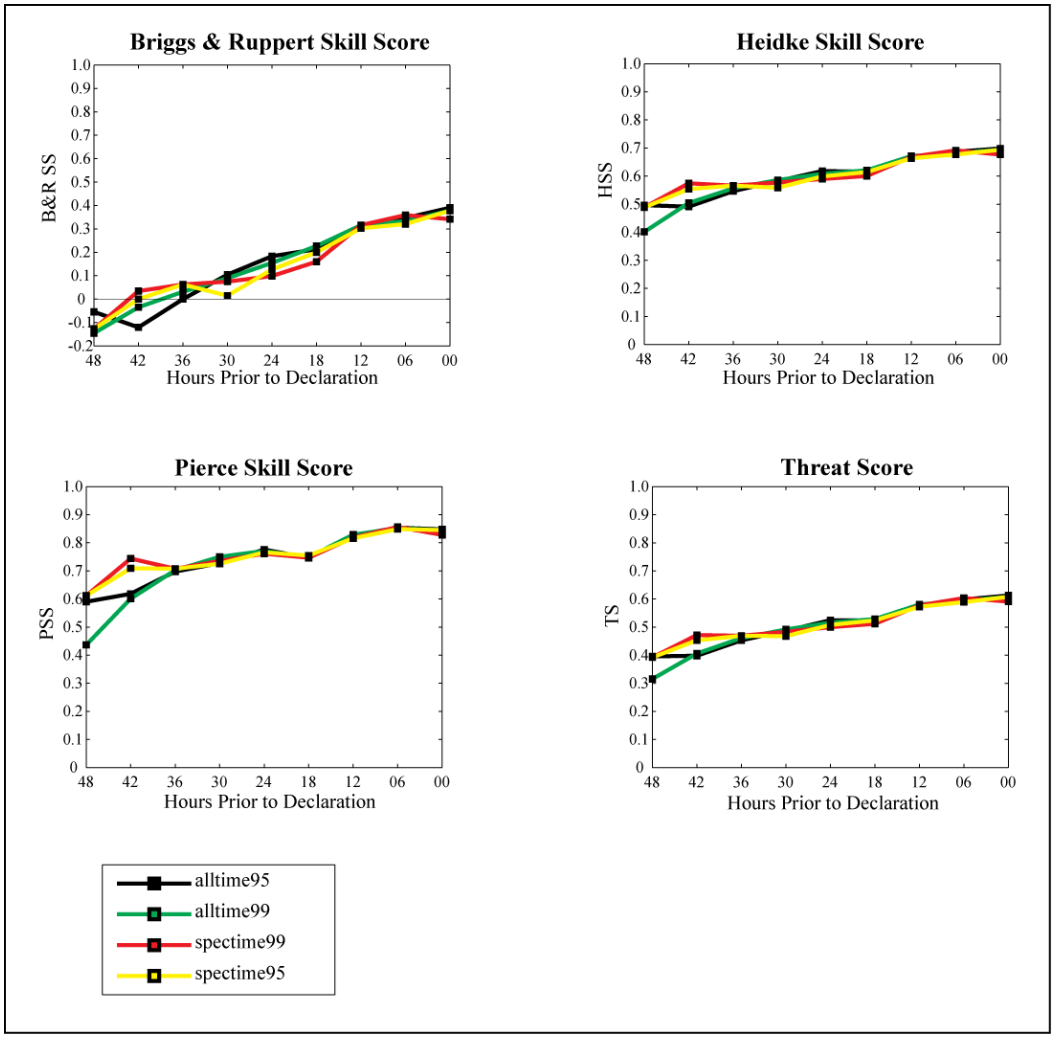

Figure 2.6 The skill scores for each of the four sets of parameters used in the LDA algorithm. Skill scores equal to one indicate perfect model performance, while skill scores of zero demonstrate the skill of random forecasts. Inferior forecasts result in negative skill scores.

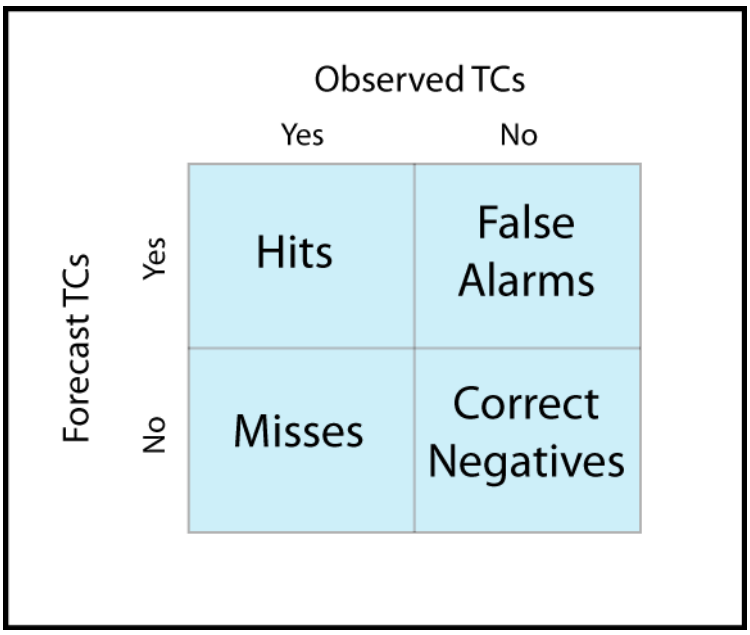

Figure 2.7 A sample contingency table along with its four elements. The four elements that make up this table can be used to calculate a number of reliability statistics and skill scores used to measure the success of a forecast model.

of incorrect Yes forecasts), misses (number of incorrect No forecasts), and correct negatives (number of correct No forecasts).

The HR, FR, and FAR are known as the reliability statistics. They can be calculated from to the 2x2 contingency table, as listed below.

$$HR = \frac{a}{a+c} \quad \textbf{(7)}$$

$$FAR = \frac{b}{a+b} \quad \textbf{(8)}$$

$$FR = \frac{b}{b+d} \quad \textbf{(9)}$$

The HR, also commonly known as the probability of detection, is the ratio of the number of Yes events that are correctly forecast by the model to the total number of Yes events observed. A perfect set of forecasts would result with HR=1, meaning that all of the Yes forecasts were realized. The FR compliments the HR in that it is the ratio of the number of false alarms resulting from the model to the total number of No events observed. A set of forecasts yielding FR=0 would indicated that the model performed perfectly with regards to the No events. The FAR is simply the fraction of the number of Yes forecasts that turn out to be incorrect in which a set of perfect forecasts would result with FAR=0.

The remaining skill scores used to assess model performance are slightly more complicated than the HR and FR ratios, but provide a more comprehensive representation of a model's forecasts. The ThS is an improvement on an accuracy measure known as the Proportion Correct (PC), which is a ratio of the total number of correct forecasts to the total number of forecasts. The PC is a good measure of accuracy for dichotomous events in which the number of Yes and No observations are similar. However, for cases in which one of the

events is rare, the PC can be misleading. To use this study as an example, from 2001 to 2009, there were 406 ND events and 82 DEV events resulting in a disproportionate number of ND to DEV events. If one were to condition a model to make only ND forecasts the resulting PC would be roughly 83%. This would give the model's accuracy a high assessment, when in reality there is no skill in the model. By contrast, the ThS measures the accuracy of a model by ignoring the correct negatives forecast by the model, therefore negating the bias of the disproportionate influence of the number of No events.

The Threat score measures the forecast skill of a model using values that range from zero (no skill) to one (perfect skill) where rare events are typically associated with lower threat scores. Threat scores greater than 0.5 are considered to show significant skill for TCs (Perrone and Lowe 1986). In this study, the threat score is calculated as

$$ThS = \frac{a}{a+b+c}. \tag{10}$$

Note again, that this score ignores the number of correct negative forecasts that are represented by *d* in the contingency table.

Perhaps the most common skill score used to scale the success of a forecast model is the HSS. It compares the accuracy of the forecasts given by a model to the accuracy a set of random forecasts would have. Using a contingency table, the HSS is calculated as below.

$$HSS = \frac{2(ad-bc)}{(a+c)(c+d)+(a+b)(b+d)} \tag{11}$$

The PSS is another common tools to determine the skill of a forecast model, and is calculated as

$$PSS = \frac{ad-bc}{(a+c)(b+d)}. \tag{12}$$

This score is the same as subtracting the FR from the HR so a perfect score would be PPS=1, since HR=1 and FR=0 for a perfect set of forecasts. Likewise, random forecasts would receive a PSS of zero since HR=FR in such a situation, and PSS<0 indicate inferior forecasts.

The final skill score, BRSS, used in this study improves on the PSS in the same way that the ThS improves on the PC accuracy measure (Briggs and Ruppert 2005). The penalty for false alarms in the PSS is disproportionately less than the penalty for misses, while the BRSS penalizes each type of incorrect forecast equally. Using a contingency table, the BRSS is calculated in the formula below.

$$BRSS = \frac{a-b}{a+c} \tag{13}$$

Just as the PSS, the BRSS is equal to one for perfect forecasts, zero when forecasts are random, and less than zero for inferior forecasts.

The skill scores for each of the four parameter sets used in the LDA algorithm are listed for all nine forecast times in Figure 2.6. Just as with the errors in (Figure 2.4), the skill scores are very similar for each group from the 36-hour forecast to the time of declaration. The largest differences are seen during the 42 and 48-hour forecasts between *alltime99* and the other three parameter sets. With the exception of FR and FAR, the scores all show improved skill by the model for the other three sets of parameters with *spectime99* and *spectime95* having slightly better skill than *alltime95*. The HR increases from 0.53 for *alltime99* to 0.71 for *spectime95*, while the PSS, ThS, and HSS, increase between the two groups by 0.16, 0.07, and 0.08, respectively.

All skill scores indicate that the parameter sets have more skill than climatology except for the 48-hour (all parameter sets) and 42-hour (*alltime* parameter sets) forecasts, in

which the BRSS values are negative. The fact that the PSS values are much higher (ranging from 0.44 to 0.73) during the two farthest forecast times than the BRSS indicates that the weaker penalty against false alarms may have masked the skill of the algorithm. However, all other skill scores (including the ThS) show that the algorithm produces skillful results compared to climatology for all parameter sets out to 48 hours. Furthermore, the increases in values for BRSS, HSS, HR, PSS and ThS from the 48-hour forecast to the time of declaration indicate improvement of the model closer to the time of genesis, which is to be expected.

As mentioned earlier, the only assessment scores that do not vary between parameter sets for forecast times are the FAR and FR values. The consistency in these numbers relates to the number of false alarms. In the case of the FAR (FR) statistic, the number of false alarms is divided by the number of Yes forecasts (No events). The fact that neither of these reliability statistics change very much through time for any of the parameter sets (ranging from 0.1 to 0.15) indicates that the number of false alarms forecasts were consistent.

c. Misclassifications

The number of misclassified cases varied for each parameter set and forecast time so, for simplicity's sake, one the misclassifications of the *alltime95* parameter set is investigated here. The total number of DEV (ND) cases over all nine times was 626 (3654), and of these, 79 (439) were misclassified using the *alltime95* parameter set. The *alltime95* set is chosen because it performs well at longer forecast times and its parameters are fixed through time, which makes the identification of parameters that may have caused misclassification more straightforward.

The *alltime95* set included 10 parameters. The parameter value means for the DEV and ND groups at each time are listed in (Table 2.3) along with the difference between the

DEV	$\overline{\phi}$	$\overline{\zeta}$	$\overline{D}_{str}$	$\overline{D}$	$\overline{u}$	$\frac{\partial u}{\partial p}_{900hPa}^{500hPa}$	$\frac{\partial v_{sg}}{\partial p}_{900hPa}^{500hPa}$	RH	ω
total missed = 79									
% > 1 stdv from dev avg	43.0%	50.6%	31.7%	26.6%	27.9%	25.3%	25.3%	16.5%	54.4%
% > 2 stdv from dev avg	3.8%	6.3%	2.5%	1.3%	0.0%	6.3%	3.8%	1.3%	2.5%
			1,1						
ND									
total missed = 439									
% > 1 stdv from nd avg	32.8%	30.8%	15.7%	29.8%	22.6%	12.3%	31.0%	39.4%	18.0%
% > 2 stdv from nd avg	2.5%	4.8%	2.5%	4.6%	1.8%	2.3%	11.2%	16.0%	0.0%
dev group means	lat	vor	strdef	def	mzw	ush5	vsheq	rh	owp
F48	12.43	5.24	0.48	4.10	-1.13	-3.00	-10.63	85.46	-0.10
F42	12.69	5.65	0.52	4.16	-1.34	-2.82	-10.56	85.12	-0.12
F36	12.86	5.86	0.65	4.20	-1.23	-2.89	-10.75	85.87	-0.13
F30	13.19	6.06	0.80	4.38	-1.16	-3.10	-10.49	85.33	-0.13
F24	13.51	6.14	0.87	4.44	-1.14	-3.02	-10.71	85.08	-0.15
F18	13.64	6.32	0.91	4.53	-1.37	-2.58	-10.50	84.75	-0.18
F12	13.89	6.82	0.94	4.82	-1.14	-2.78	-10.75	84.78	-0.21
F06	14.42	6.93	0.85	4.85	-1.23	-3.05	-9.76	84.35	-0.23
F00	14.85	7.15	0.85	4.95	-1.37	-2.83	-9.24	84.06	-0.24
overall	13.50	6.24	0.76	4.49	-1.23	-2.90	-10.38	84.98	-0.17
nd group means									
	10.78	3.84	1.25	3.44	-2.52	-1.80	-7.46	77.35	-0.01
dev - nd group means									
F48	1.65	1.40	-0.77	0.67	1.40	-1.20	-3.17	8.11	-0.09
F42	1.91	1.81	-0.73	0.72	1.19	-1.02	-3.10	7.77	-0.11
F36	2.08	2.02	-0.60	0.76	1.29	-1.09	-3.30	8.52	-0.12
F30	2.41	2.22	-0.45	0.94	1.37	-1.31	-3.03	7.99	-0.13
F24	2.73	2.30	-0.38	1.00	1.39	-1.22	-3.26	7.73	-0.15
F18	2.86	2.47	-0.33	1.09	1.16	-0.78	-3.04	7.40	-0.17
F12	3.11	2.98	-0.30	1.39	1.38	-0.99	-3.29	7.44	-0.20
F06	3.64	3.08	-0.40	1.41	1.29	-1.25	-2.30	7.00	-0.22
F00	4.07	3.31	-0.40	1.51	1.15	-1.03	-1.78	6.72	-0.23
overall	2.72	2.40	-0.48	1.06	1.29	-1.10	-2.92	7.63	-0.16

Table 2.3 The means of the DEV (at all nine forecast times) and ND (24 hours past vortex identification) groups as well as the difference between the two groups for the nine parameters included in *alltime95*. Also listed are the percent of misclassified DEV cases whose parameter values deviated (toward the ND means) from the DEV mean by greater than one or two standard deviations.

two group means. On average, the values of latitude, relative vorticity, total deformation, zonal wind, and relative humidity are greater than those of the ND group for all nine times. The average values of the shallow vertical shear of the zonal wind and equatorward vertical shear of the meridional wind are larger in magnitude for the DEV cases than for the ND cases. The ND group average parameter values are larger for the horizontal stretching deformation and the Okubo-Weiss parameter. Therefore the potential for misclassification exists for any DEV (ND) vortex with a combination of parameter values that are closer to the mean value of the ND (DEV) group.

To get a sense of which parameters have large contributions to misclassification, a comparison of the group means and individual cases is made. First, the means of the DEV group parameters for each of the nine forecast times are calculated. Then, the parameter values for each misclassified DEV cases are subtracted from the DEV group mean. Any parameter value that deviates from the DEV mean (toward the mean of the ND group) by more than one standard deviation is tallied, and the same is done for the parameters that are more than two standard deviations away. Finally, the percentages of all parameters whose values deviate more that one or two standard deviations away from the DEV mean are calculated. This process is repeated for the ND cases.

The percentages of parameters whose values for an individual misclassified DEV (ND) are closer to the means of the ND (DEV) group and whose difference from the DEV (ND) group mean is larger than one or two standard deviations are shown in (Table 2.3). For the DEV cases, the largest contributor to misclassifications seems to be from strong values of Okubo-Weiss parameter as approximately 54% of misclassified DEV cases have an Okubo-Weiss parameter that is over one standard deviation away from the DEV group mean (in a

direction closer to the ND group mean) and roughly 3% of the cases had Okubo-Weiss parameter differences greater than two standard deviations. The next two leading contributors at the one standard deviation level are the relative vorticity and latitude at 51% and 43%, respectively.

For the ND cases, the latitude and relative humidity contribute the most to misclassifications. Interestingly, the relative vorticity contribution is relatively low (fourth) on the contribution list. This could indicate that a vortex that has a relative vorticity characteristic of the ND group may be positioned in a location and environment more typical for a DEV vortex. If these three parameters were similar to the DEV group, the model would classify the ND vortex as DEV even though the vorticity is very weak.

2.4. Physical Interpretation of Results

Based on the error and skill score results of the four parameter sets tested in the algorithm, LDA has proven to be skillful in predicting tropical cyclogenesis improving on climatology out to 48-hours, but what are the underlying reasons for improvement between parameter sets? The largest improvement in forecast skill for the model occurs at the 42 and 48-hour forecast where there is a large divergence in the error and skill scores between the *alltime99* parameter set and the remaining sets. To understand how the parameters may lead to successful or unsuccessful forecasts, two TS cases will be investigated. The first, TS Gil, was declared a tropical storm at 1800 UTC on 29 August 2007. Gil was well forecast by the model as the LDA algorithm correctly forecast it as a DEV case at all forecast times. The second case is TS Enrique, which was declared as a tropical storm at 0000 UTC on 4 August 2009 and was consistently forecast as a ND case at all forecast times.

a. Vorticity and stretching deformation

Based on the confines of the parameter selection, *alltime95* includes all of the parameters included in *alltime99* and two additional parameters (Table 2.2). These parameters are the low-level horizontal stretching deformation and the shallow vertical shear of the zonal wind. The fact that the addition of these two features decrease the error and increase all of the skill score values in comparison to the *alltime99* results at these two times, shows that they are important to the prediction of TCs on a longer time-scale. For the misclassified cases, the stretching deformation and shallow vertical shear of the zonal wind may have contributed to 47% and 36% of the misclassifications, respectively. Furthermore, these findings indicate that both deformation and shallow vertical shear of the zonal wind may play an important role in the development of a nascent vortex into a TC during the early stages of their development.

The vorticity of the vortex is larger for DEV than ND vortices. This suggests that a stronger vortex has a greater chance of developing under favorable environmental conditions than a weaker vortex. The zonal wind is also not a surprising contributor to forecast skill as anomalously westerly winds over the basin (and therefore over the vortex) have been found to favor genesis (Maloney and Hartmann 2001, Davis et al. 2008). A similar situation is found in the means of the vortices here, with weaker *easterly* winds present over the DEV vortex group.

The maximum relative vorticity for TS Gil ranged from $8 \times 10^{-5} s^{-1}$ 48 hours prior to declaration to $12 \times 10^{-5} s^{-1}$ at the time of declaration (Figure 2.8). The stretching deformation over the vortex oscillated from positive to negative over the 48 hour time period. The effects of this oscillation in the stretching deformation field over the vortex can be further understood by considering the orientation of the axes of dilatation. The axes of dilatation present during

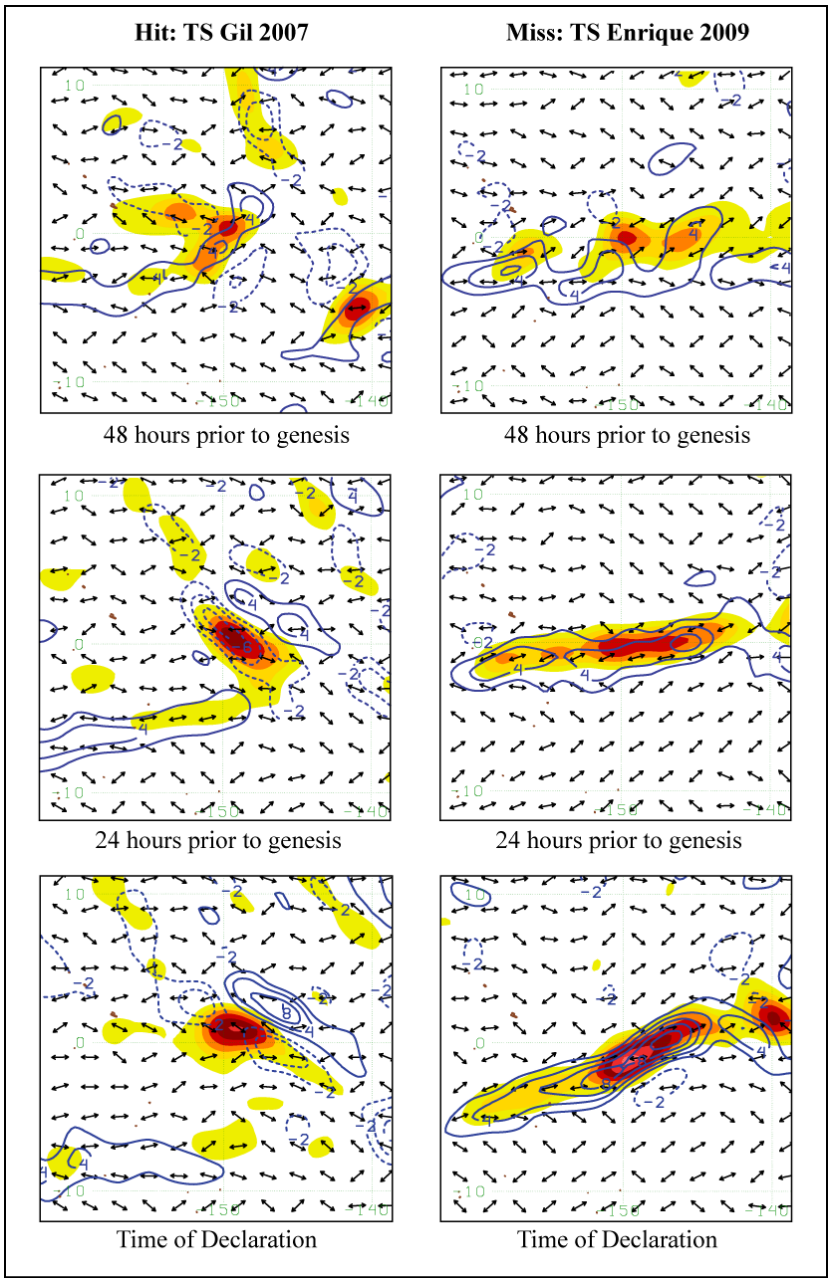

Figure 2.8 The 900hPa relative vorticity (shaded in increments of 2 x$10^{-5}s^{-1}$ beginning with 2x$10^{-5}s^{-1}$) and the 900hPa horizontal stretching deformation contoured every 2x$10^{-5}s^{-1}$ (dashed lines indicate negative values) for TS Gil and TS Enrique. The axes of dilatation associated with the deformation are represented by black arrows.

the development of Gil seem to oppose one another across the center of the vortex. The fact that the dilatation axes are not oriented in a similar fashion across the vortex indicates that the stretching deformation is not acting to "stretch out" the vortex in a particular direction, thus increasing the favorability of Gil's development. The symmetric differences between the sign (orientation) of the stretching deformation (axes of dilatation) over the vortex suggest that the stretching deformation acting the vortex is weak compared to the vortex relative vorticity. Indeed, the magnitude of the total deformation over the Gil's center remained relatively weak compared to the magnitude of the relative vorticity as evidenced by the Okubo-Weiss parameter over the vortex (Figure 2.9). The combination of strong relative vorticity and weak deformation would tilt the forecast of the LDA algorithm to a correct "yes" forecast.

On the other hand, the stretching deformation over TS Enrique was large and strong over the 48-hour period leading up to its declaration, and the orientation of the axes of dilatation over the length of the vortex are more similar than those over Gil (Figure 2.8). However, when comparing the magnitudes between the vorticity and deformation, the value of the Okubo-Weiss parameter is equal to that of TS Gil (Figure 2.8). Therefore, while the vorticity and Okubo-Weiss parameter values would certainly skew the forecast from the algorithm to a "yes" forecast, the strength of the stretching deformation acting over the vortex would bring the forecast closer to "no."

b. Latitude and relative humidity

The parameter set, *alltime99*, includes nine parameters: the latitude of the vortex center, vortex vorticity, total and stretching deformation, low-level zonal wind, relative humidity, equatorward vertical shear of the meridional wind, the 900hPa-500hPa vertical shear of the zonal wind, and the Okubo-Weiss parameter. It is not surprising that a

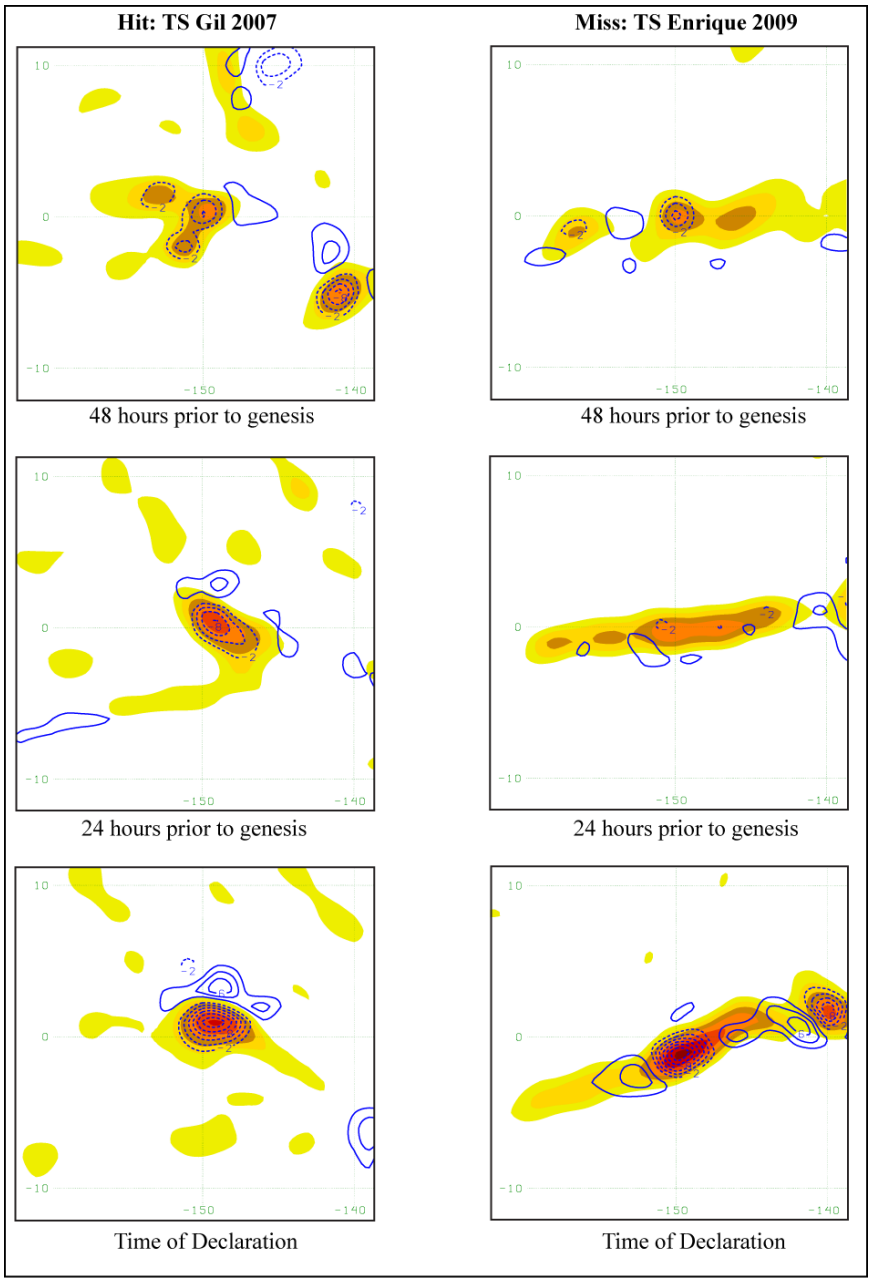

Figure 2.9 The 900hPa relative vorticity (shaded in increments of 2 beginning with $2x10^{-5}s^{-1}$) and 900hPa OWP contoured every $2x10^{-9}s^{-2}$ for TS Gil and TS Enrique.

combination of these parameters produce skillful forecasts, as they are all influential to development.

The location of the vortex is influential to genesis in the east Pacific basin for two reasons. First, the latitude of the vortex impacts genesis as it relates to the planetary vorticity, since planetary vorticity must exist to spin-up a vortex (Gray, 1968). All of the vortices in the data set meet the requirement of being poleward of 6°N, but the mean latitude for the DEV vortices is further north than the ND vortices at all nine forecast times (Table 2.3). At first glance, this may seem to indicate that the larger planetary vorticity gives rise to a greater likelihood that a vortex will develop into a TC. However, a more likely factor may be the location of a vortex in relation to the low level jet. For example, vortices to the north of the jet are located in an environment characterized by cyclonic horizontal shear near the surface. This shear can lead to increased convergence and convection, which are both associated with a favorable environment for TC development (Bister and Emanuel 1997, Davis et al. 2008, etc.).

The second influence of the vortex's location concerns the longitude. The SSTs in the east Pacific basin cool considerably to the west of 130°W, which would inhibit development there. Indeed, the mean longitude of the ND vortices is approximately 120°W, which is roughly 12° further west than DEV cases. However, when the longitude of the vortex center was added to the parameter set, no significant change in the results was seen. This may be due to the fact that the majority of the ND cases remain in the warmest SSTs portion of the basin; therefore, it is possible that the number of ND cases influenced by cooler SSTs was not large enough to impact their predictability.

In addition, the relative humidity can also be expected to distinguish between these two groups (Nolan 2007, Davis et al. 2008). The relative humidity for the DEV group is roughly 7-8% higher in the developing cases.

c. Vertical wind shear

The DEV cases, for an areal average, typically have a larger magnitude of shallow (900hPa-500hPa) vertical shear of the zonal wind than do the ND cases. This indicates that the winds at low-levels are typically more (less) westerly than at mid-levels for the DEV (ND) cases. The patterns of the shallow vertical shear of the zonal wind between Gil and Enrique are not that different (Figure 2.10). However, the areal averaged values are, with Enrique having a positive areal average, while Gil has a value that is closer to the DEV group average.

The final parameter is the equatorward vertical shear of the meridional wind, which can be thought of as a surrogate for convection. Since enhanced convection has been found to be favorable for genesis (Davis et al. 2008), larger negative values of this parameter would be expected for the DEV cases. Indeed, the areal average for this parameter is larger for the DEV than ND cases (Figure 2.11).

The magnitudes of the shear to the south of the vortex are similar between the two cases at 48 hours prior to declaration. However, the areal expanse of the negative shear is larger for Enrique than Gil at both 24 hours prior to declaration and at the time of declaration. Therefore, this parameter would have contributed to a yes forecast from the LDA algorithm.

Overall, a number of parameters including the relative vorticity, latitude, Okubo-Weiss parameter, and equatorward vertical shear would have contributed to a yes forecast for Enrique. However the contributions of the stretching deformation, and shallow vertical shear of the meridional wind must have been sufficiently large in order for this storm to be

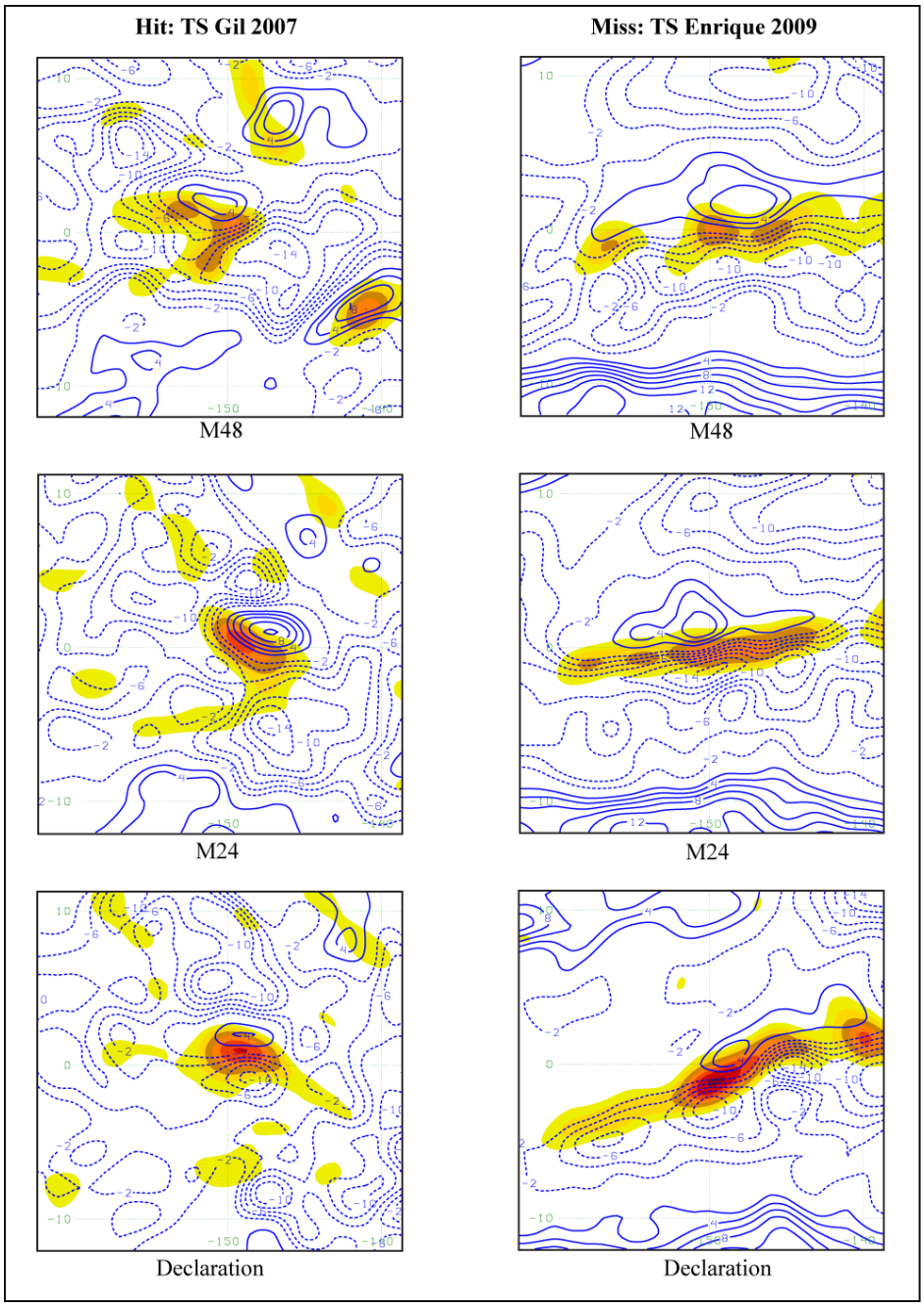

Figure 2.10 The 900hPa relative vorticity (shaded in increments of 2 beginning with $2x10^{-5}s^{-1}$) and shallow (900hPa-500hPa) vertical shear of the zonal wind contoured every $2ms^{-1}$ for TS Gil and TS Enrique.

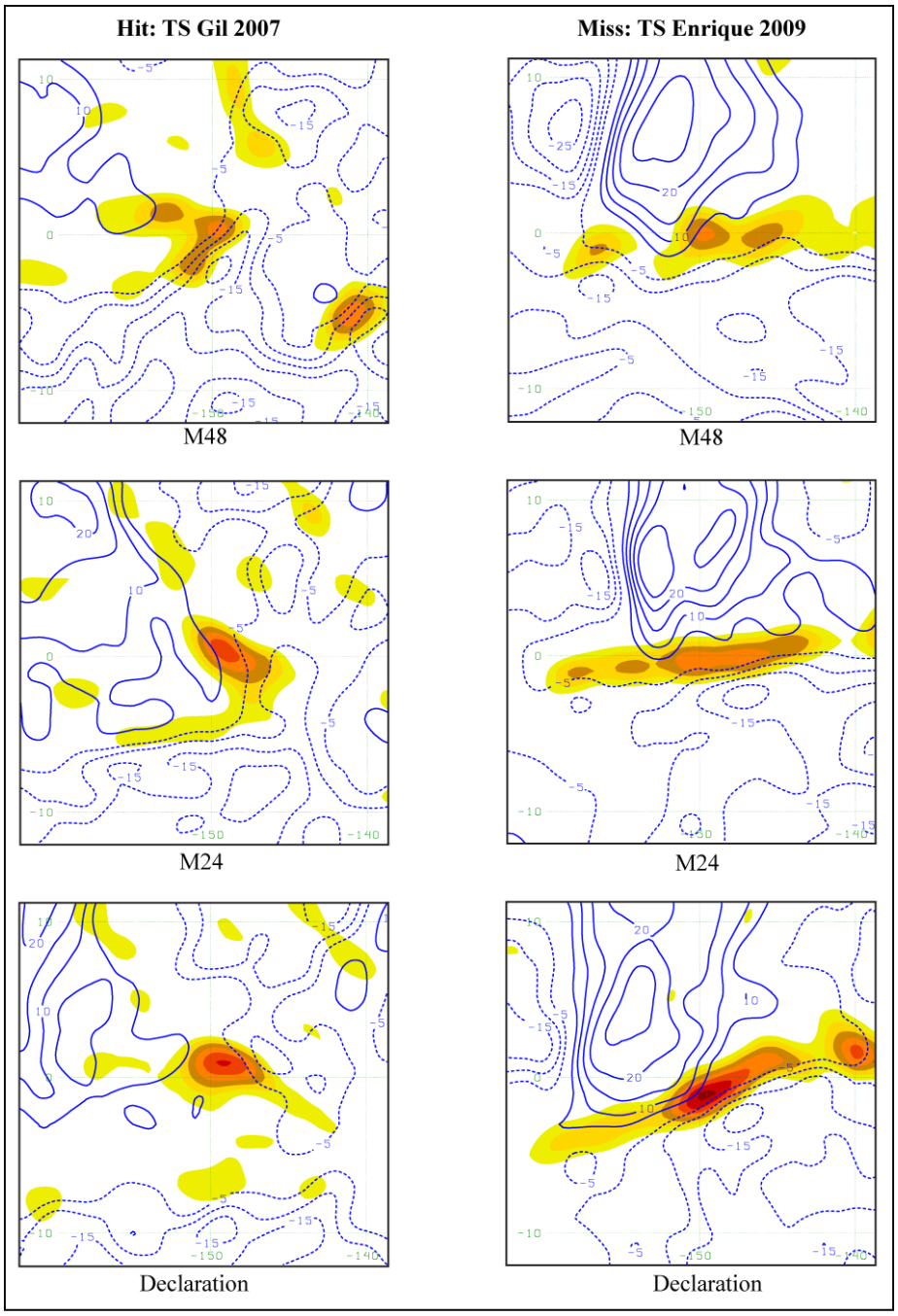

Figure 2.11 The 900hPa relative vorticity (shaded in increments of 2 beginning with $2\text{x}10^{-5}s^{-1}$) and 900hPa-200hPa vertical shear of the meridional wind contoured every $2ms^{-1}$ for TS Gil and TS Enrique.

misclassified by the algorithm. In the case of TS Enrique, the stretching deformation may have ultimately led the storm's misclassification. Considering the synoptic-scale environment present for TS Gil and TS Enrique (Figure 2.12), it is possible that the former case is representative of typical tropical cyclogenesis events (though an investigation of numerous developing cases would be needed to confirm this hypothesis). TS Gil formed to the south of an upper-level anti-cyclone in a relatively low-shear environment. On the other hand TS Enrique formed to the south of an upper-level cyclone in a high shear environment, and was associated with a strong deformation field.

Another interesting difference between these two storms concerns the potential of barotropic growth. A number of studies have shown that tropical vortices can grow barotropically at the expense of the mean flow (Lipps 1970; Maloney and Hartman 2001; Hoover 2010). One of the parameters used in this study, $-\frac{\partial \overline{u}}{\partial y}\overline{u'v'}$, is a zonally averaged variant of a barotropic conversion term that was found to contribute significantly to the barotropic development of vortices in the east Pacific basin by Maloney and Hartmann (2001). While this term did not decrease the forecast error of the LDA algorithm when added to the *alltime* parameter sets, it was included in a number of the *spectime* parameter sets yielding mixed results. Meridional cross-sections of this term are shown for both TS Gil and TS Enrique (Figure 2.12 – bottom row). For TS Enrique, this term is strongly positive indicating that barotropic growth may have played a major role in the storm's development. The same was not true for TS Gil, which shows negative flux values over the vortex. These findings suggest that the LDA model seems to skillfully predict many DEV cases in the east

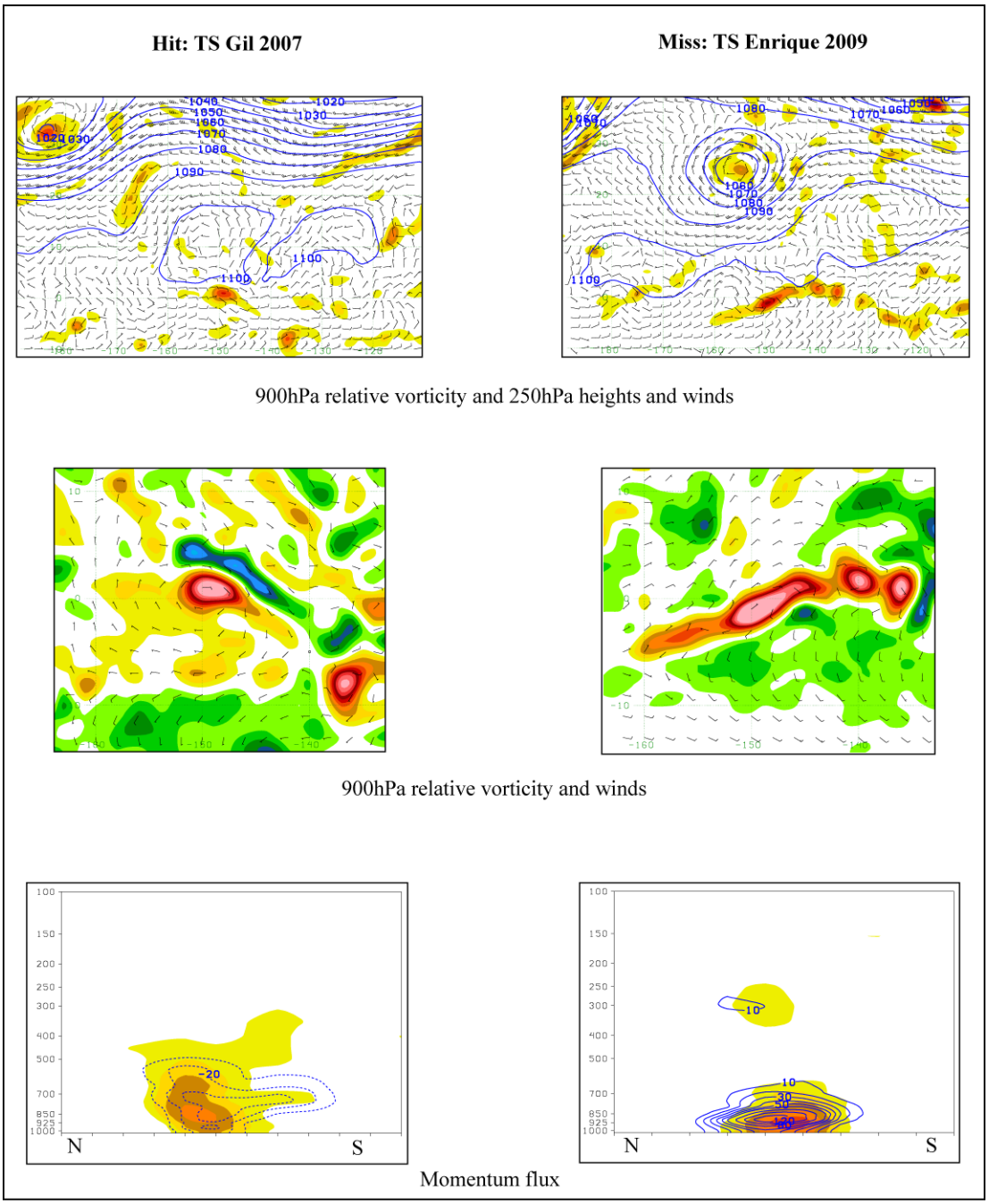

Figure 2.12 The top row is the 900hPa relative vorticity (shaded) and 250hPa heights (contoured in dm) and full winds (represented by barbs) for TS Gil and TS Enrique; the 900hPa relative vorticity (warm colors are positive vorticity and cool colors are negative vorticity) and winds (middle row); and a north-south cross-section through of the vortices (represented in fill pattern by the relative vorticity) and the meridional momentum flux in blue contours (dashed lines represent negative values).

Pacific, but it seems to have little skill in predicting genesis cases associated with a barotropic growth mechanism.

2.5. CONCLUSIONS/FUTURE WORK

This study focused on the feasibility of predicting eastern Pacific tropical cyclogenesis using a linear discriminant analysis technique. Vortices that developed during the peak seasons from 2001 to 2009 east Pacific TC seasons were categorized into developing and non-developing groups and used to calibrate the LDA algorithm. Various sets of parameters were tested but the most skillful set included the latitude of the vortex center, relative vorticity, horizontal stretching and total deformation, zonal wind, vertical shear (900hPa-500hPa zonal and 900hPa-200hPa equatorward meridional) and the Okubo-Weiss parameter. This combination of parameters misclassified 79 (439) of 626 (3654) total DEV (ND) cases.

An inspection of each of the misclassified cases revealed that a number of parameters (including vorticity, latitude, horizontal stretching deformation, and the Okubo-Weiss parameter) seemed to significantly contribute to the misclassifications produced by the FLDA code as it deviated strongly toward values associated with cases that developed. A comparison between TS Gil and TS Enrique revealed that a potential subset of TC formation, which results from a vortex developing in a strong shear flow just north of the jet, may lead to poor forecasts by the LDA algorithm. In the case of TS Enrique, the vortex is misclassified by the strong horizontal shearing deformation surrounding the vortex. However, it seems that the vortex was able to grow at the expense of the jet through barotropic growth. While a momentum flux term was considered as a parameter in the algorithm, it did not improve on the predictive skill of the algorithm and was therefore excluded as a parameter. One potential avenue for future research could be to develop a “secondary” LDA algorithm that would be

calibrated to forecast TC genesis from a barotropic growth mechanism. This algorithm would include the momentum flux term for cases in which a strong vortex is embedded in a highly cyclonically sheared environment.

On a broader scope, in future work, this method could be used to compare the influence that various parameters have on tropical cyclogenesis in different basins. Separate studies have used forms of this method to predict tropical cyclogenesis in the Atlantic and western Pacific basins. However, to the author's knowledge, no study has used LDA as a tool to compare which factors are most influential to tropical storm development in each basin.

A third approach would use LDA to compare the sensitivities to prediction of tropical cyclogenesis for various models. This method could be used to test the predictive skill that a variety of linear combinations of parameters have on classifying whether or not a vortex will develop. Further investigation would compare the skill of these combinations on different models to see if models are more sensitive to certain parameters than others.

Chapter 3: A statistical approach to predicting tropical cyclogenesis in the eastern Pacific basin: Bayes probabilistic model

3.1 Introduction

After decades of study, tropical cyclogenesis remains a still poorly understood and not well-predicted process. Based on observation and analysis, a set of climatological conditions that are necessary, but not sufficient, for formation of TCs has been established. These conditions include sufficiently warm sea surface temperatures (SSTs), a large Coriolis parameter, weak vertical wind shear, high relative humidity in the low- to mid-troposphere, and a pre-existing low-level vortex (Gray 1968, 1979). However, only a minority of the total number of vortices that form in favorable environments actually reach tropical storm (TS) or hurricane strength.

Additional factors that may be influential to TC development include barotropic instability (Lipps 1970; Shapiro 1978; Davidson and Hendon 1989; Maloney and Hartmann 2001), the Madden-Julian Oscillation (MJO) (Maloney and Hartmann 2000), and low-level horizontal stretching deformation (Davis et al. 2008). These factors can lead to a better understanding of why a particular vortex may develop or decay, but they do not necessarily factor into improved predictions. For example, the predictions produced by global spectral numerical weather prediction models like the Global Forecast System (GFS) model are inconsistent at best with relation to the timing and location of TC genesis (Pratt and Evans 2004; Pasch et al. 2006, 2008). One of the main reasons for these inconsistencies is that the coarse resolution of the models does not allow for accurate meso-scale features that may be important to TC formation to be resolved. Therefore, it may be useful to use large-scale

features in a different predictive framework to compliment and perhaps improve predictions of TC formation.

The Bayes probabilistic model is a permutation of Bayes method and can be used to calculate a conditional probability of an event occurring at some point in the future. The probability of an event occurring (such as the genesis of a tropical storm from an nascent vortex) is computed based on features of the event's environment known to have an influence of the event. While Bayes method assumes independence of the features used in the algorithm (Wilks 2006), it can still produce quality results even if the variables are not truly independent (Zhang 2006; Hand and Yu 2001).

A naïve Bayes probabilistic model and classifier have been successfully implemented in a number of meteorological studies in the past decade. Lui et al. (2001) found that a naïve Bayes classifier was more accurate than other statistical forecast methods for forecasting precipitation in Hong Kong with a maximum accuracy rating of 90%. A naïve Bayesian model also improved the seasonal persistence and coupled model forecasts for the El-Niño-3.4 index (Coelho et al. 2003). Most recently in tropical meteorology, a Bayes probabilistic model and classifier showed skill in predicting secondary eye-wall formation in eastern Pacific and Atlantic hurricanes (Kossin and Sitkowski 2009).

This study employs a naïve Bayes probabilistic model and classifier to predict TC genesis in the east Pacific basin using a number of large-scale features shown to distinguish between developing and non-developing vortices. The remainder of this chapter consists of five sections. Section 2 summarizes the data and methodology of the study. Section 3 will detail the results of the Bayes model and classifier, while section 4 will highlight those events

that were misclassified by the model. A discussion of the physical meaning of the results will be contained in section 5 with concluding remarks presented in section 6.

3.2 Data/Methodology

a. Data

The data used to calculate the vortex-centered parameters used for the Bayes model and classifier are the National Centers for Environmental Prediction (NCEP) six hourly, 1°x1° final reanalysis data[4]. Data are formatted using the Penn State University/National Center for Atmospheric Research (PS/NCAR) mesoscale model (MM5) horizontal grid reinterpolation (REGRIDDER) program, converting 1˚x1˚ data to a standard 90km grid. Low-level (900hPa) vortices are identified and tracked using the interpolated data by first smoothing the 900hPa relative vorticity field and locating the maximum vorticity of each individual vortex. An archive of the real-time multivariate Madden Julian Oscillation (MJO) indices and phases (Wheeler and Hendon 2004) is also used to define an MJO parameter for the Bayes probabilistic model.

b. Domain and vortex classification

Using the vortex tracker along with the National Hurricane Center's (NHC) best track database, a set of developing and non-developing vortices is created. Developing vortices are defined as those vortices which are declared as tropical storms at some point in their life cycle, while non-developing vortices never reach TS strength. All vortices considered have a minimum average relative vorticity of $1.5 \times 10^{-5} ms^{-1}$ for at least 48 consecutive hours. During

[4] The data for this study are from the Research Data Archive (RDA), maintained by the Computational and Information Systems Laboratory of the National Center for Atmospheric Research (NCAR). The original data are available from the RDA (http://dss.ucar.edu) in dataset number ds083.2.

the 2001-2009 peak east Pacific hurricane seasons 488 vortices meet these criteria with 82 vortices reaching at least TS strength.

c. Parameter selection

Two types of variables are chosen for this study, those consisting of storm-centered variables, whose values are averaged over a 11°x11° box centered on the vortex – with the exception of vorticity, which is averaged over a 5°x5° box (Figure 2.1) – and those variables associated with the MJO. The selection process is carried out in the following manner.

For both sets of parameters, a general list of relevant variables is compiled, and an initial set of 27 environmental variables is constructed (these variables are listed along with their abbreviations Table 2.1). While there are many factors that may contribute to the formation of tropical cyclones, the key to this method is to find the features that best separate developing from non-developing vortices. Some factors that have been demonstrated to influence tropical cyclogenesis include latitude (as it relates to the Coriolis parameter), 925hPa vorticity, 900-200hPa vertical wind shear, sea surface temperatures (SSTs), and relative humidity (Gray 1977). In the eastern Pacific, low-level jets, and enhanced Hadley cell type circulations have also been shown to provide favorable conditions for tropical cyclogenesis (Davis et al. 2008).

The mean SSTs for the duration of the time considered are above 26°C for the entire area except a small portion of the northwest corner (Figure 1.1). Since SST is geographically dependent in this region, and both the latitude and longitude of the vortex center will be considered as parameters in the model, SSTs will not be considered for the Bayes model. The remaining climatological parameters will be included.

d. Naïve Bayes probabilistic model and classifier

The goal of this study is to use a probabilistic model to skillfully predict TC genesis by using a set of n features (detailed above) that are influential to TC formation. As mentioned above, there are two groups of vortices under consideration: 82 developing cases (V_{dev}) and 406 non-developing (V_{nd}) cases. The values of the n features associated with each vortex are represented by the vector $\mathbf{x} = x_1, x_2, \ldots . x_n$.

The probabilistic model chosen for this study is derived from Bayes Theorem, which uses the Multiplicative Law and the Law of Total Probabilities to "invert" a known conditional probability in order to solve for an unknown conditional probability (Wilks 2006). In other words, if a quantity $P(x_i \mid V_{dev})$ is known, Bayes Theorem can be used to calculate $P(V_{dev} \mid x_i)$ using the formula

$$P(V_{dev} \mid x_i) = \frac{P(V_{dev})P(x_i \mid V_{dev})}{P(x_i)}. \quad \textbf{(1)}$$

The term $P(x_i \mid V_{dev})$ is called the class-conditional probability and describes the likelihood that a specific feature would have the value x_i given the condition that a vortex belongs to the developing group. It should be mentioned here that for continuous functions it is best to use a binning system to determine a value range for x_i instead of using a single scalar value. For example, $.70 < x_i < .75$ would be a binned value range for relative humidity over a vortex. The class conditional probability term for this bin would equal the number of times a developing vortex has a relative humidity between 70% and 75% divided by the total number of developing vortices. (For this study a probability density estimate, consisting of 100 bins, for each parameter is constructed using a normal kernel function.) The other term in the nominator is the prior probability, which is simply the probability that a vortex will belong to the developing group given no other information other than the distribution of the vortices.

For this study $P(V_{dev}) = \frac{V_{dev}}{V_{dev} + V_{nd}}$. The term $P(V_{dev} \mid x_i)$, is known as the posterior probability and gives the probability that a vortex will belong to the DEV group given that a feature has the value x_i. Finally, $P(x_i)$ is strictly the probability that a specific feature will have the value x_i regardless of the class to which a vortex belongs. A more complete expression for the denominator can be written by using of the law of total probability as:

$$P(x_i) = P(V_{dev})P(x_i \mid V_{dev}) + P(V_{nd})P(x_i \mid V_{nd}), \quad \textbf{(2)}$$

where $P(x_i \mid V_{nd})$ is the conditional probability that the value x_i occurs given a non-developing vortex and $P(V_{nd})$ is the "climatology" of the non-developing vortices.

Equation (1) is simplified for a case in which only one feature is used. The model used in this study requires the inclusion of multiple features. Since the naïve Bayes model assumes independence between all of the features in **x**, each of the conditional probabilities for each feature may be multiplied resulting in the equations below.

$$P(\mathbf{x} \mid V_{dev}) = \prod_{i=1}^{n} P(x_i \mid V_{dev}) \quad \textbf{(3a)}$$

and

$$P(\mathbf{x} \mid V_{nd}) = \prod_{i=1}^{n} P(x_i \mid V_{nd}) \quad \textbf{(3b)}$$

Combining Equation (1) with Equations (3a) and (3b) the final equation for the naïve Bayes model becomes:

$$P(V_{dev} \mid \mathbf{x}) = \frac{P(V_{dev}) \prod_{i=1}^{n} P(\mathbf{x}_i \mid V_{dev})}{P(V_{dev}) \prod_{i=1}^{n} P(\mathbf{x}_i \mid V_{dev}) + P(V_{nd}) \prod_{i=1}^{n} P(\mathbf{x}_i \mid V_{nd})}. \quad \textbf{(4)}$$

This study tests the predictive skill of Bayes model for TC genesis at nine forecast times (from the time of declaration as a TS to 48 hours prior to declaration in increments of six hours) using spatially averaged values calculated for each of the variables mentioned in section (2.c) at all nine times for the developing vortices. For comparison, the averaged variables for the non-developing vortices will be calculated for only one time, 24 hours after vortex identification. Various combinations of features that characterize the two groups of vortices are tested with Bayes model to identify the set of features yielding the best forecast skill.

Finally, the most useful parameters are retained and used to construct a naïve Bayes classifier. The classifier is used to convert the probabilistic forecasts produced by Bayes model into a binary set of "yes" and "no" forecasts. Since there is a lot of information lost (concerning the likelihood of development given by a probabilistic forecast) by converting a probabilistic forecast into one that is binary, the classifier is only used in this chapter as a means to compare the results of this model to those produced by the LDA algorithm in Chapter 2.

3.3 Model Results

Five parameter sets are tested with the Bayes model. The same procedure for choosing parameter sets used in Chapter 2 (see Chapter 2.3*a* for complete details on parameter set selection and naming conventions) is employed here as well. A two-sample t-test is used to identify those parameters whose means between the DEV and ND groups are significantly different at the 95% and 99% confidence levels. Those parameters that pass at a particular confidence level for all forecast times are considered for inclusion in the *alltime* parameter sets. The *alltime* parameter sets are finalized by identifying the combination of

parameters, which pass at the required confidence level, that yields the lowest forecast error. The *spectime* parameter sets are determined by including any parameter that passes the t-test at the required confidence level at a specific forecast time. The final five parameter sets that result from this procedure include four parameter sets similar to those used in the linear discriminant analysis model (LDA) discussed in Chapter 2 as well as an additional fixed parameter set that includes the parameters in the *alltime95* set as well as the phase of the MJO. Of the five sets of parameters, *alltime99* proved to produce the most skill based on a variety of skill scores and reliability statistics (Figures 3.1, 3.2 and 3.3). Therefore, the model results for *alltime99* will be presented here.

a. Attributes diagram

The attributes diagram (Figure 3.4) gives a comprehensive view of a probabilistic model's forecast skill by showing the complete joint distribution of forecasts and observations. It is meant specifically for a probabilistic forecast model with a binary predictand. The attributes diagram consists of four line types representing the calibration function, perfect reliability, no skill, and no resolution. The calibration function is represented by a line, which connects each joint distribution pair. One can assess the calibration of a probabilistic model by the proximity of the calibration function to a second line, called the perfect calibration line. Five sample calibration functions (shown in Figure 3.4) demonstrate some potential calibrations.

The first calibration function (represented by the blue line) consist of points which all lie above the perfect calibration line. Such a function indicates a probabilistic model that consistently under-forecasts a particular event. In other words, the average forecast given by the model is smaller than the average observation of the event. The other extreme for this

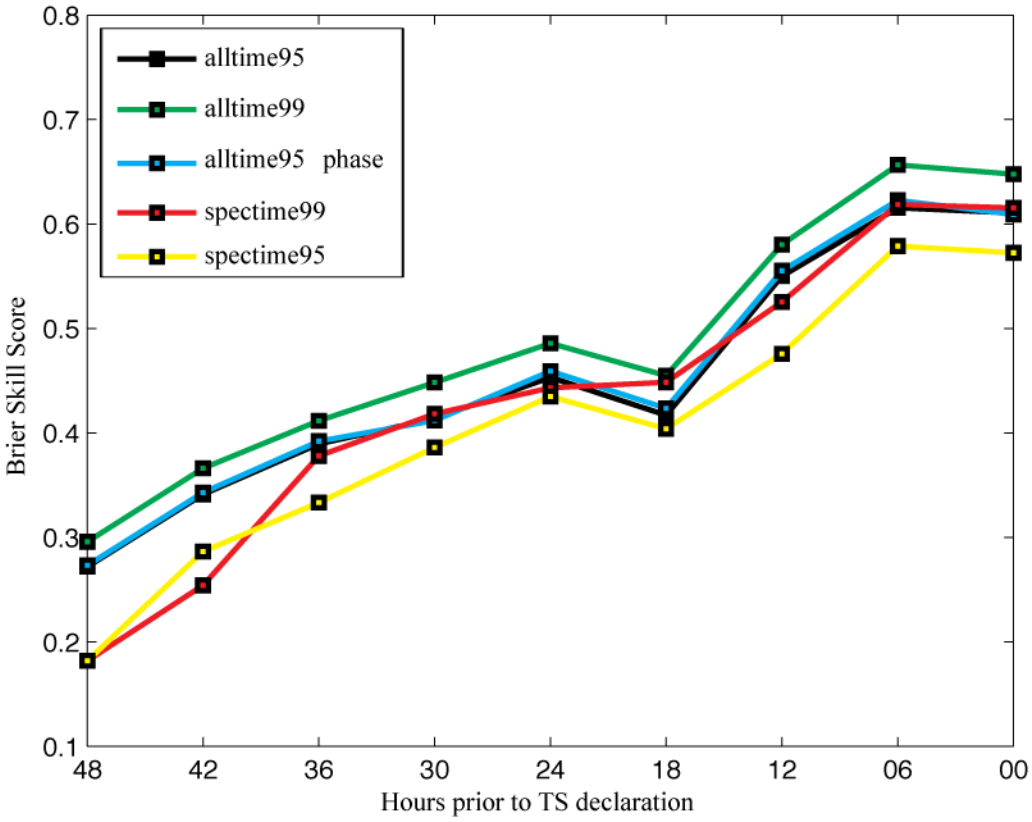

Figure 3.1 The Brier skill scores for each forecast time (48-hour forecast to the zero hour forecast) for the five parameter sets tested.

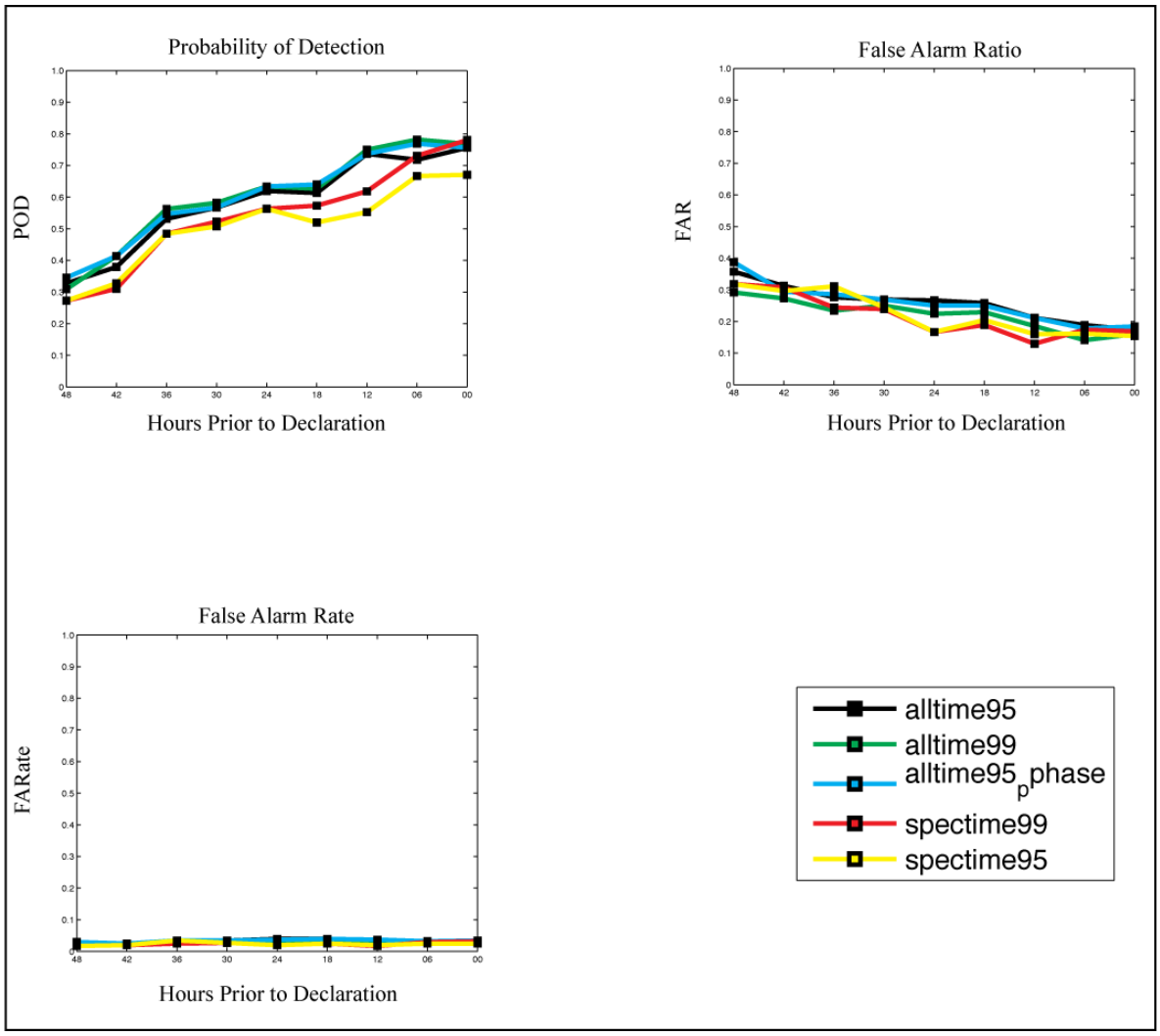

Figure 3.2 The reliability statistics for the five parameter sets. The values for reliability statistics are along the ordinate axis and forecasts are along the abscissa.

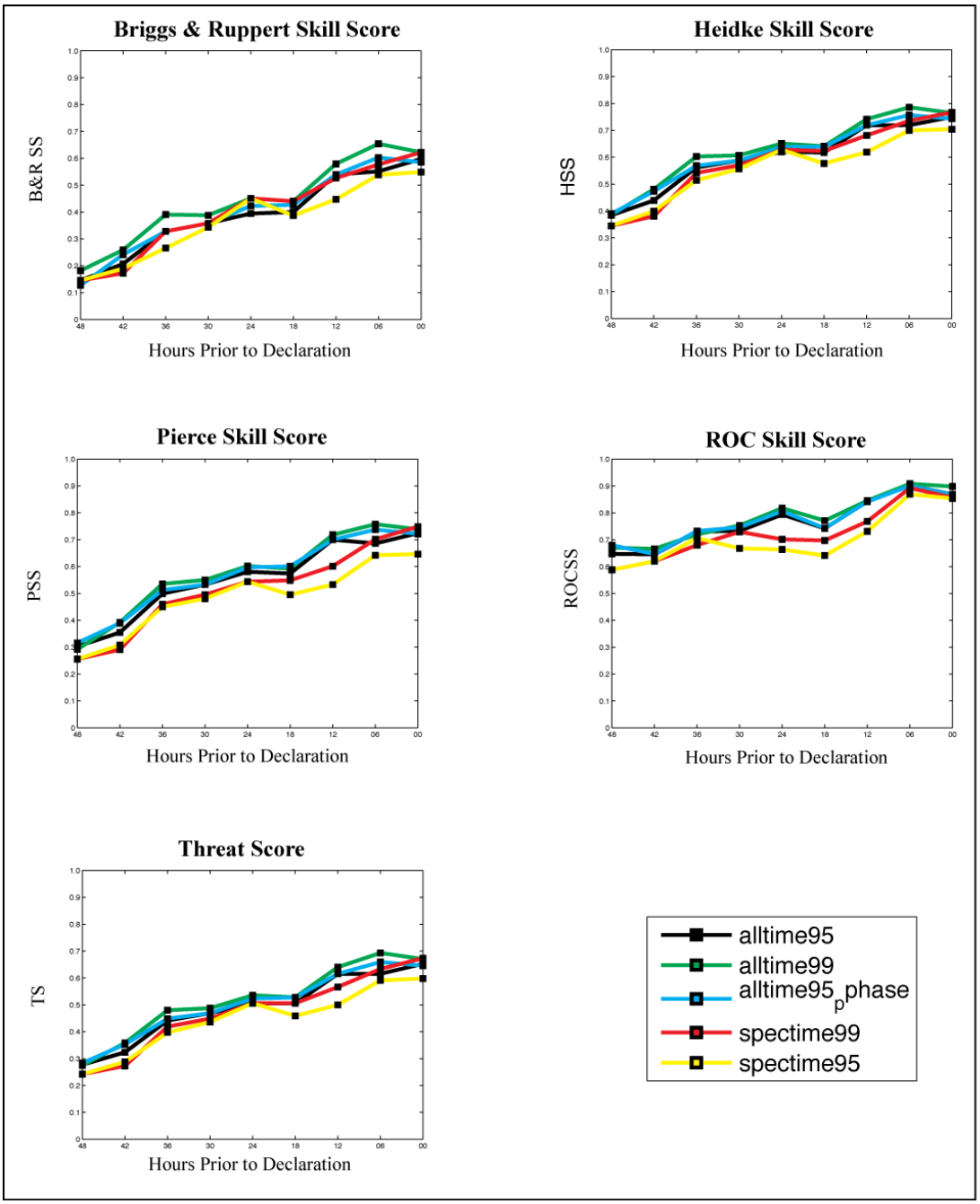

Figure 3.3 The skill scores for the five parameter sets. The values for skill scores are along the ordinate axis and forecasts are along the abscissa.

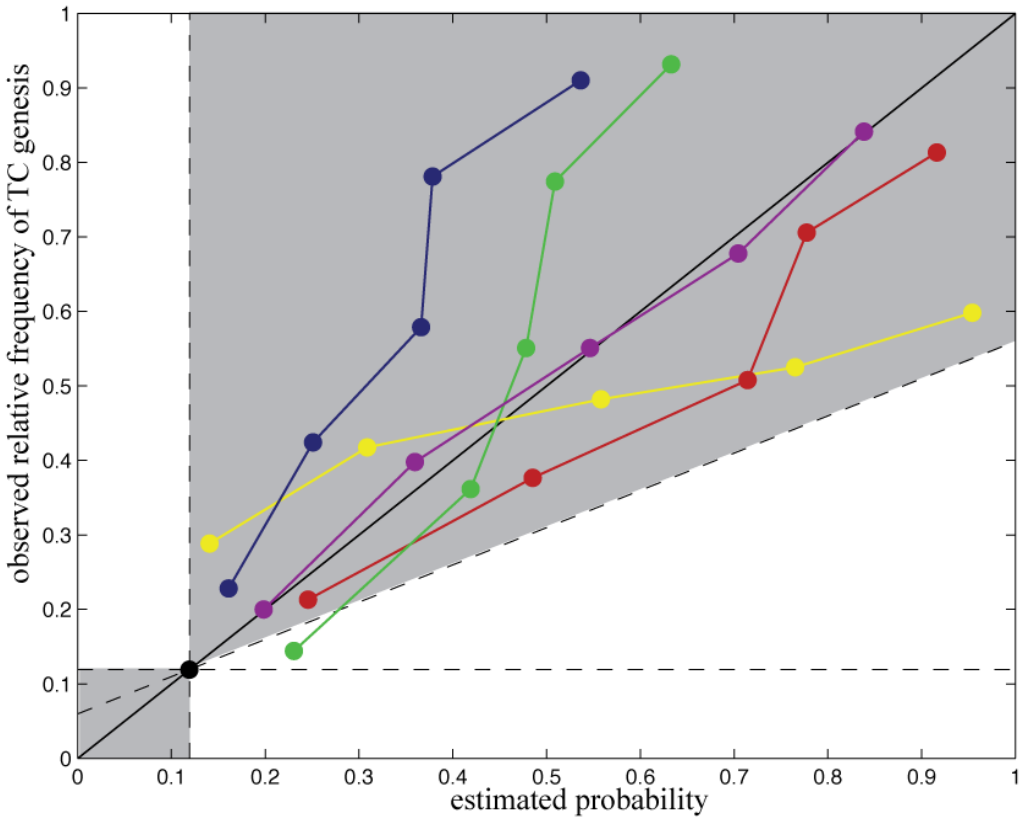

Figure 3.4 Example of an attributes diagram including five sample calibration functions. Calibration functions shown are examples of under-forecasting (blue line), over-forecasting (red line), under confidence (green line), over confidence (yellow line), and good calibration (purple line). Points along a calibration function that lie within the shaded region add forecast skill.

type of miscalibration is a model that consistently over-forecasts an event. For this case, the resulting calibration function would have all points located below the perfect calibration line (e.g. the red line). Both of these calibration functions demonstrate unconditional biases of the model, and the model would be considered unreliable.

The other set of model forecast deficiencies result from conditional biases. Calibration functions from these biases are represented by the green and yellow lines, which show examples of under confidence and over confidence, respectively. In the former case, the model over (under) forecasts the low (high) probability forecasts. In other words, the forecast probability given by the model is larger (smaller) than the observed frequency. The opposite is true in the latter case.

The desired calibration function is represented by the purple line. A well-calibrated model will produce a calibration function whose points all lie along or near the perfect calibration line. Such a function indicates that the forecast probabilities given by the model are very similar to the observed relative frequencies of the predictand.

There are two additional line types on the attributes diagram, which relate to the skill of the model results. The no resolution lines (represented by the horizontal and vertical dashed lines) are determined by the climatology forecast (the black dot). Calibration functions that lie along the no resolution line indicate that the model cannot resolve (better than climatology) whether an event will occur. Finally, the no-skill line (dashed black diagonal line) and the vertical no resolution line act as boundaries for forecast skill. The area between these two lines is shaded for clarity. Any points that lie outside the shaded area are representative of times in which the forecast produced by the model have no skill.

The attributes diagram for the *alltime99* parameter set shows that the model demonstrated skill at all nine forecast times (Figure 3.5). Any dots that lie within the shaded region of the attributes diagram contribute forecast skill to the model, while those that lie outside the shaded region lead to skill degradation. With the exception of two points (one at 6 hours and one at 18 hours prior to declaration), all points are within the shaded region and add to the forecast skill of the model.

The calibration functions for all nine times straddle the perfect skill (dashed) line to some degree and none of them show a consistent trend of being above or below the perfect-skill line so there is no evidence of a persistent over or under forecasting bias. There does appear to be a consistent pattern of slightly under forecasting both the low and high probability events as both the starting and ending points for all (except the endpoint for the 42 hour forecast) the diagrams lie below the perfect-skill line. On the other hand, most of the mid-range probability events are under forecast by the model.

The vertical distance from each point of the calibration function to the perfect skill line decreases as the forecast time approaches the time of declaration indicating an increase in the model's skill at shorter forecast times. For the most part, the high probability forecasts are very close to the perfect-skill line indicating that the model has done well at forecasting high probability events. Overall, according to the attributes diagram, the Bayes probabilistic model produced skillful forecast with the *alltime99* parameter set. It should be noted that the other parameter sets produced somewhat similar attributes diagrams (not shown), but the *alltime99* parameter set outperformed the other sets.

b. Relative Operating Characteristic (ROC) diagram

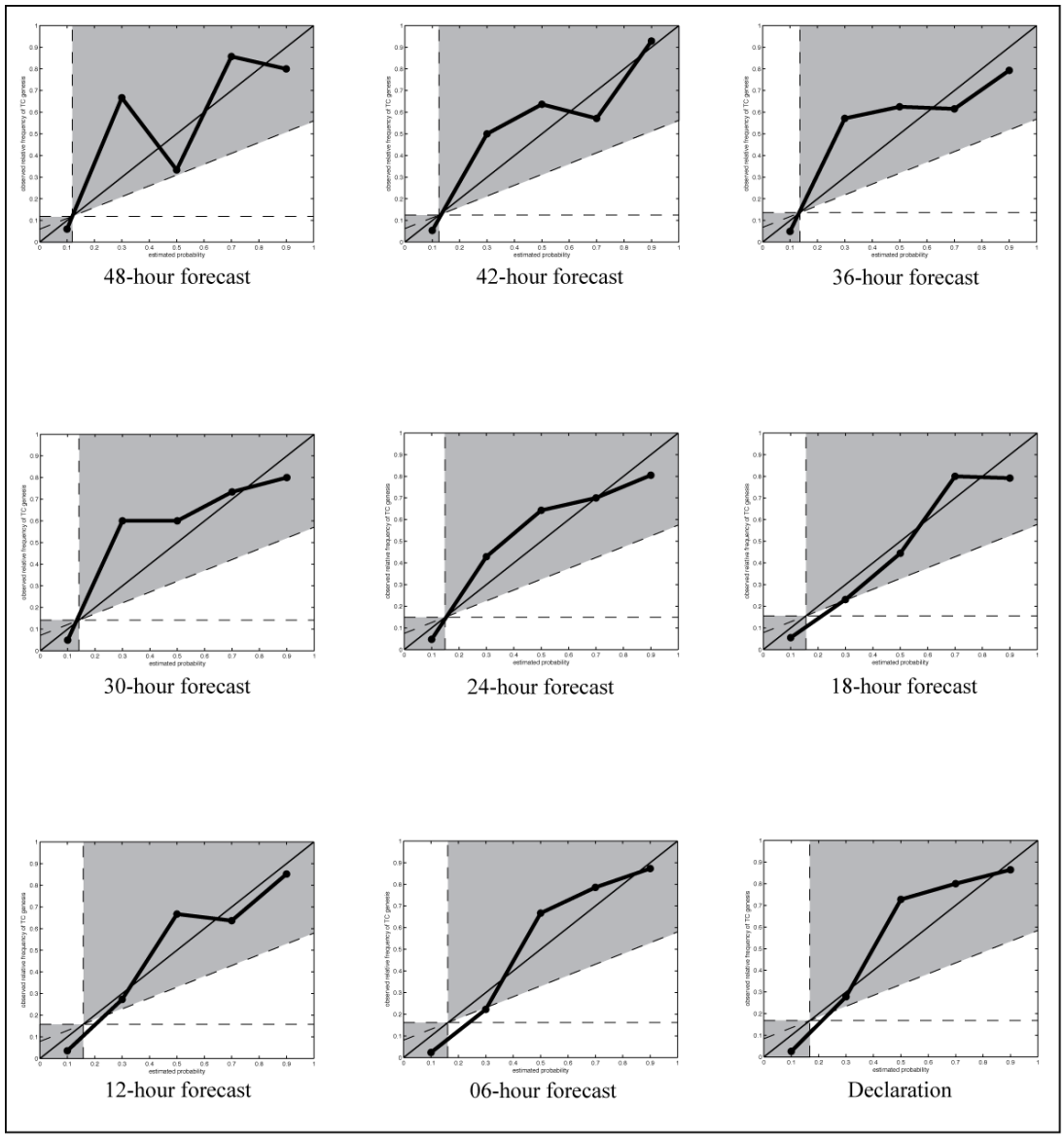

Figure 3.5 Attributes diagrams for each forecast resulting from the *alltime99* parameter set. In each plot, the ordinate axis is the relative frequency of observed TC genesis and the estimated probability is on the abscissa. The shaded region denotes forecast skill.

Although not as comprehensive as the attributes diagram, the ROC diagram offers a graphical summary of the skill of a probabilistic forecast model (Figure 3.6). The ROC diagram is constructed by assigning the forecast probabilities into a set of *n* bins with *n-1* thresholds (in this study 10 bins with thresholds ranging from 0.1 to 1 were used). Contingency tables are then made for each bin such that if the forecast probability is greater than the threshold, the forecast becomes "Yes" while all other forecasts are labeled as "No". The resulting hit rates and false alarm rates calculated from each contingency table are paired and plotted on the diagram.

The hit rate (HR) and false alarm rate (FR) are plotted on the ordinate and abscissa axes, respectively, with the end points defined at (0,0) and (1,1). The point (0,0) corresponds to never forecasting an event, while (1,1) corresponds to always forecasting the event. A perfect set of forecasts yields HR=1, and FR=1. Therefore, the corresponding ROC curve would consist of two line segments: a vertical line segment on the left edge of the graph and a horizontal line across the top. On the other hand, for purely random set of forecasts HR=FR, which produces a diagonal line extending from (0,0) to (1,1).

Keeping in mind that a perfect set of forecasts would result in a two-segmented ROC curve, the *alltime99* parameter set shows skill at all forecast times as indicated by their corresponding ROC curves (Figure 3.6). Once again, an increase in skill for the short-term forecasts is evident from the ROC curves more closely resembling the two-segment curve. Yet, even at the 48-hour forecast, the ROC curve indicates impressive forecast skill with the ROC curve being closer to the two-segmented curve than the diagonal line.

c. Skill scores for probabilistic models

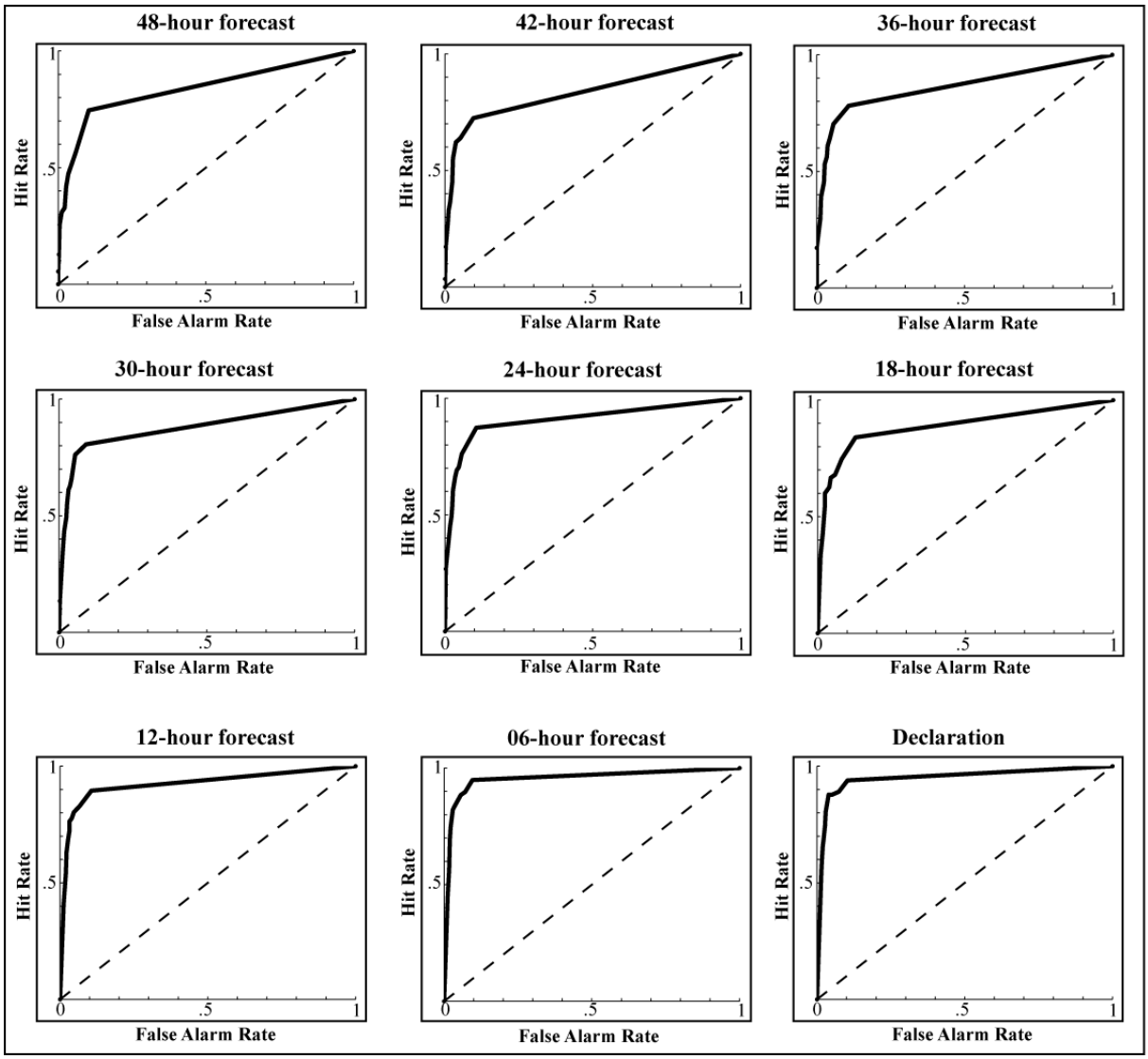

Figure 3.6 The Relative Operating Characteristic (ROC) diagrams for the *alltime99* parameter set forecasts. The ROC curve for each forecast is represented by the solid black line, while the dashed line represents the ROC curve for a set of random forecasts.

In addition to the attributes diagram and ROC curve there are two skill scores that can be used to summarize the forecast skill of a probabilistic model. These scores are known as the ROC skill score and the Brier Skill Score.

The first of the skill scores is a transform the ROC curve into a skill score. This transformation can be performed by using the general skill score formula and considering the area under the ROC curve generated by random forecasts as the underlying skill measure. The resultant equation is

$$ROC_{ss} = \frac{A - A_r}{A_p - A_r}, \quad \textbf{(5)}$$

where A_r is the area under the ROC curve of a set of random forecasts, A_p is the area under a ROC curve resulting from a set of perfect forecasts, and A is the area under the ROC curve of the forecast model.

The ROC skill scores for *alltime99* ranges from 0.67 to 0.91 for the nine forecast times with larger values corresponding to the short-term forecasts (Table 3.1). This indicates that all forecasts show more skill than random forecasts, since all values are greater than 0.5.

The second skill score is derived from the Brier score and is a statistical measure of a forecast skill. The Brier Score is calculated, in this study, according to the equation derived in (Brier 1950):

$$B = \frac{1}{n}\sum_{j=1}^{r}\sum_{i=1}^{n}(f_{i,j} - E_{i,j}), \quad \textbf{(6)}$$

where n is the number of forecast cases, r is the number of classes or categories, and f is the probabilistic forecast that an event will occur. E is set equal to one if an event occurred in class j or zero if the event did not occur in that class. A perfect Brier score is zero, while a

Skill Scores

Alltime99	BRSS	FAR	FR	HSS	HR	PSS	ThS	ROCSS
F48	0.18	0.29	0.02	0.39	0.31	0.29	0.27	0.67
F42	0.26	0.27	0.02	0.48	0.41	0.39	0.36	0.67
F36	0.39	0.23	0.03	0.60	0.56	0.54	0.48	0.72
F30	0.39	0.25	0.03	0.61	0.58	0.55	0.49	0.75
F24	0.45	0.22	0.03	0.65	0.63	0.60	0.54	0.82
F18	0.44	0.23	0.03	0.64	0.63	0.59	0.53	0.77
F12	0.58	0.19	0.03	0.74	0.75	0.72	0.64	0.85
F06	0.65	0.14	0.02	0.79	0.78	0.76	0.69	0.91
F00	0.62	0.16	0.03	0.76	0.77	0.74	0.67	0.90

Table 3.1 The skill scores for each forecast time for the *alltime99* parameter set.

score of one corresponds to no forecast skill. The Brier *skill* score adopts the general skill score formula (as used for the ROC skill score) and uses the Brier score for climatology (B_c) as the underlying accuracy measure. Since a perfect Brier score is equal to zero, the resulting formula for the Brier skill score is

$$BSS = 1 - \frac{B}{B_c}. \tag{7}$$

The Brier skill score ranges from negative infinity to one, where one reflects a perfect forecast and a score of zero indicates that the forecast has the same skill as climatology. A Brier skill score less than zero indicates that the forecast is inferior to climatology. The Brier score for climatology used in this study was calculated by taking the ratio of the number of developing cases to the total number of vortices that occurred during the 2001-2009 peak seasons. The Brier skill scores for the *alltime99* show that the parameter set produced superior forecasts to climatology for all forecast times (Figure 3.1). The lowest score of 0.18 corresponds to the 48 hour forecast, while the six-hour forecast showed the most skill with a Brier skill score of 0.65.

d. Bayes classifier and results

In order to do a relatively direct comparison of the Bayesian model and the LDA algorithm used in Chapter 2, it is useful to convert the probabilistic output of Bayes model into a binary output like that produced by the LDA algorithm. It is possible to perform this conversion by using a decision rule to create a Bayes classifier. A number of methods can be used to determine this classifier. More complex methods choose a decision rule by employing results from the ROC curve mentioned above. Since the only purpose of the classifier in this study is to compare the Bayes model performance to that of the LDA algorithm in Chapter 2,

a simple "largest probability" decision rule is used. Using this rule, any case whose probability of development is greater than or equal to (less than) 0.5 is classified as a DEV (ND) case.

There are a number of reliability statistics and skill scores that can be used to describe the skill of binary forecasts. Some of the more commonly used reliability scores include the Hit Rate (HR), False Alarm Rate (FR), and the False alarm ratio (FAR). The former score is a measure of how well a model forecasts a "Yes" event, while the latter two describe how the model forecasts "No" events. The skill scores are used to describe how well a model performs compared to some underlying accuracy measure. The skill scores used in this study include the Heidke skill score (HSS), Peirce skill score (PSS), Threat score (ThS), and the Briggs and Ruppert skill score (BRSS). (See Chapter 2 for a full description of these reliability statistics and skill scores.)

Recall that, for all skill scores, a value of one reflects a set of perfect forecasts, zeros indicate the skill of randomly generated forecasts, and negative values indicate that the model forecasts are worse than random forecasts. In terms of the reliability statistics, high HR values and low FR and FAR values indicate reliable forecasts. All of the reliability statistics and skill scores reflect less skill at the longer forecast times (Table 3.1) and increase at shorter lead times. When compared to the reliability statistics and skill scores that resulted from the LDA forecasts, the Bayes model does not do as well. This may seem discouraging, but the lower skill is likely due to a non-optimal decision rule. In the end, the classifier produces skillful forecasts at short lead-times.

3.4 Misclassified cases

The parameter set yielding the best results for this model was the *alltime99* parameter set, which includes the vortex center location (latitude and longitude), relative vorticity, north-south extent of the vortex, total deformation over the vortex, the 900hPa zonal wind, and the 900hPa-200hPa vertical shear of the meridional wind. The majority of these variables were also included in the *alltime99* parameter set in Chapter 2. There are some differences between the parameter set in Chapter 2 and the current set. First, the parameter set in Chapter 2 does not include the longitude and north-south extent of the vortex because the addition of those parameters did not influence the LDA algorithm's performance. The kernel density functions, which are probability density estimates based on a normal kernel function using a window parameter, of both features reveal that these features may be more useful in the Bayes model than the LDA algorithm (Figure 3.7). According to the functions, one can infer that the means between the DEV (red line) and ND (blue line) groups would be fairly similar. Since the LDA algorithm relies heavily on the difference between the means of the two groups, it would be more difficult for the algorithm to properly assess whether a vortex should belong to the DEV or ND group given either of these parameters. Bayes model, on the other hand, is able to use the structure of the density curve when assessing its probability. Therefore, a vortex with a longitude of 100˚W would be more easily distinguished as a DEV case in the Bayes model than the LDA algorithm. Similar reasoning would stand for the north-south extent of the vortex, though the differences are subtler than those in the longitude case. The other difference between the sets is the absence of the Okubo-Weiss parameter in the current parameter set, which is also due to the fact that the addition of the parameter did not change the forecast skill of Bayes model.

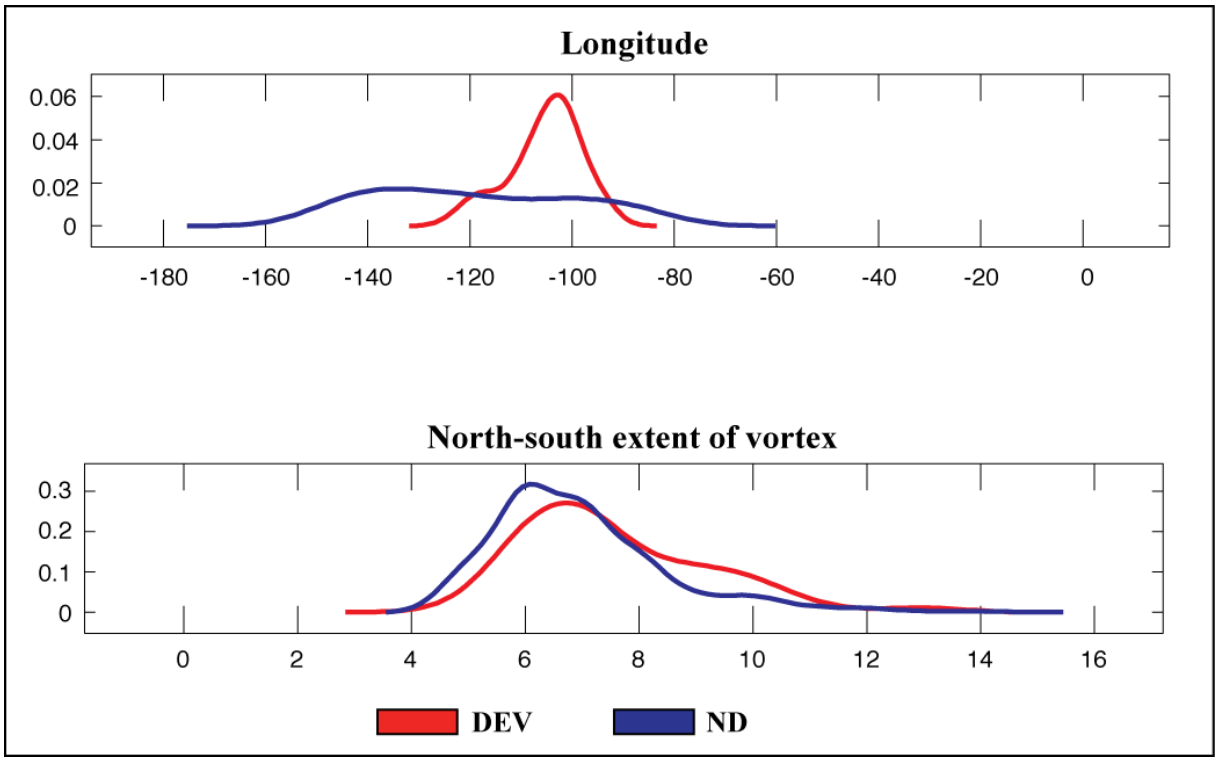

Figure 3.7 The kernel density functions for longitude (in degrees, i.e. -120 is 120˚W) and the meridional extent of the vortex (in grid points where one grid point is 90km).

A list of the misclassified DEV and ND cases for all nine forecast times was compiled. The model had very few false alarms compared to the LDA algorithm with only 92 misclassified ND cases for all nine forecast times (compared to 439 for the LDA algorithm). However, there were many more misclassified DEV cases (237) than there were for the LDA algorithm (79). The increased number of misclassified DEV cases from Bayes model may be a result from using a non-optimal decision rule. Once again, the power of this model is its probabilistic nature, which has been proven to be skillful.

To get a sense of which parameters may contribute to misclassification, a comparison of the group means and individual cases is made. First, the means of the DEV group parameters for each of the nine forecast times are calculated. Then, the parameter values for each misclassified DEV case are subtracted from the DEV group mean. Any parameter value that deviates from the DEV mean (toward the mean of the ND group) by more than one standard deviation is tallied, and the same is done for the parameters that are more than two standard deviations away. Finally, the percentages of all parameters whose values deviate more that one or two standard deviations away from the DEV mean are calculated. This process is repeated for the ND cases.

The percentages of parameters whose values for an individual misclassified DEV (ND) case deviated toward the mean of ND (DEV) group mean by more than one or two standard deviations are shown in (Table 3.2). The leading contributors to the misclassification of DEV cases appear to be the 900hPa total deformation (33.5%) and zonal wind (30%) having the largest percentage of misclassified DEV cases whose values for these parameters are at least one standard deviation closer to the ND values. This indicates that these features more closely resembled those of the ND cases than the DEV cases for a number

DEV	ϕ	θ	$\overline{\zeta}$	$\overline{\zeta}\Big\|_S^N$	$\overline{D}$	$\overline{u}$	$\frac{\partial \overline{v}_{sfc}}{\partial p}\Big\|_{900hPa}^{900hPa}$	RH
total missed = 237								
% > 1 stdv from dev avg	24.9%	26.6%	34.6%	18.6%	33.3%	30.0%	27.8%	24.9%
% > 2 stdv from dev avg	2.5%	6.8%	3.8%	0.0%	0.8%	1.7%	6.3%	7.2%
ND								
total missed = 92								
% > 1 stdv from nd avg	65.2%	48.9%	78.3%	41.3%	67.4%	28.3%	45.7%	60.9%
% > 2 stdv from nd avg	26.1%	34.8%	51.1%	5.4%	32.6%	3.3%	27.2%	45.7%
dev group means	lat	lon	vor	nsext	def	mzw	vsheq	rh
F48	12.4	-106.5	5.2	7.5	4.1	-1.1	-10.6	85.5
F42	12.7	-106.9	5.6	7.5	4.2	-1.3	-10.6	85.1
F36	12.9	-107.9	5.9	7.8	4.2	-1.2	-10.8	85.9
F30	13.2	-108.6	6.1	7.9	4.4	-1.2	-10.5	85.3
F24	13.5	-108.8	6.1	7.9	4.4	-1.1	-10.7	85.1
F18	13.6	-109.3	6.3	7.8	4.5	-1.4	-10.5	84.8
F12	13.9	-110.1	6.8	8.2	4.8	-1.1	-10.8	84.8
F06	14.4	-110.7	6.9	8.2	4.9	-1.2	-9.8	84.4
F00	14.8	-111.4	7.1	8.0	5.0	-1.4	-9.2	84.1
overall	13.5	-108.9	6.2	7.9	4.5	-1.2	-10.4	85.0
nd group means								
p24	10.8	-119.1	3.8	6.9	3.4	-2.5	-7.5	77.4
dev - nd group means								
F48	1.6	12.7	1.4	0.6	0.7	1.4	-3.2	8.1
F42	1.9	12.2	1.8	0.6	0.7	1.2	-3.1	7.8
F36	2.1	11.3	2.0	0.9	0.8	1.3	-3.3	8.5
F30	2.4	10.5	2.2	1.0	0.9	1.4	-3.0	8.0
F24	2.7	10.4	2.3	1.0	1.0	1.4	-3.3	7.7
F18	2.9	9.9	2.5	0.9	1.1	1.2	-3.0	7.4
F12	3.1	9.0	3.0	1.3	1.4	1.4	-3.3	7.4
F06	3.6	8.4	3.1	1.3	1.4	1.3	-2.3	7.0
F00	4.1	7.7	3.3	1.1	1.5	1.2	-1.8	6.7
overall	2.7	10.2	2.4	1.0	1.1	1.3	-2.9	7.6

Table 3.2 For each parameter, the percent of misclassified cases that may have been incorrectly forecast due to the value of the parameter. Also listed are the average values of each parameter for the DEV cases (at each forecast time) and the ND cases (at 24 hours after vortex identification).

of the misclassified DEV cases. Other parameters that may have lead to classification errors for the DEV group include the equatorward vertical shear of the meridional wind (27.8%), and the longitude of the vortex center (26.6%). The largest contributors to the misclassified ND cases were the vortex vorticity (78.3%), total deformation (67.4%), latitude of the vortex center (65.2%), and relative humidity (60.9%).

A better understanding of how these parameters may lead to good and poor forecasts can be obtained by looking at specific cases, whose forecasts were consistently correct or incorrect. To compare model skill, the two cases considered in Chapter 2 are investigated here also. These cases are TS Gil (declared as a tropical storm at 1800 UTC 28 August 2007) and TS Enrique (declared as a tropical storm at 0000 UTC 4 August 2009). Recall that the LDA algorithm consistently (and correctly) forecast that TS Gil would develop at all nine forecast times, while it misclassified TS Enrique as a non-developing case for all forecast times. The same pattern emerges for Bayes model (Table 3.3) where the probabilities for TS Gil's development are above 50% for all forecast times with the exception of the 42-hour forecast. However, the probabilities for development associated with TS Enrique range between zero and 3.2% for all forecast times indicating that Enrique should not have developed given its surrounding environment. The only high probability for development associated with TS Enrique is at the time of declaration where the probability increased drastically to 96.97%! Therefore, both models correctly forecast that TS Gil would develop and yet missed the development of TS Enrique.

The differences between the probabilistic forecasts for the two storms, as well as those that occur over time for each vortex, are compelling and merit further investigation. Comparing how the parameter values vary overtime for each storm (Table 3.4) compared to

Bayes Forecast Probabilities

	TS Gil	TS Enrique
F48	75.73%	0.93%
F42	39.56%	0.00%
F36	53.30%	0.64%
F30	89.27%	0.34%
F24	98.29%	1.23%
F18	98.81%	0.14%
F12	91.52%	1.41%
F06	87.20%	3.20%
F00	97.27%	96.97%

Table 3.3 The Bayes probabilistic forecasts for TS Gil and TS Enrique.

Gil values	$\overline{\phi}$	$\overline{\theta}$	$\overline{\zeta}$	$\overline{\zeta}\Big\|_S^N$	$\overline{D}$	$\overline{u}$	$\frac{\partial \overline{v}_{eq}}{\partial p}\Big\|_{900hPa}^{500hPa}$	$\overline{RH}$
F48	14.6	-102.0	5.2	7.0	4.2	3.1	-14.7	84.5
F42	15.4	-102.0	4.8	9.0	3.6	4.0	-11.4	81.0
F36	16.2	-104.5	4.4	8.0	3.5	2.6	-9.9	85.5
F30	16.9	-105.3	5.7	8.0	4.1	2.7	-8.2	86.1
F24	16.9	-106.1	6.0	9.0	4.8	0.8	-7.7	86.0
F18	17.7	-106.9	6.5	7.0	5.4	2.1	-6.2	87.3
F12	18.5	-107.7	5.7	7.0	5.0	0.4	-2.5	86.5
F06	19.3	-108.5	6.3	7.0	4.5	-0.7	-0.5	82.4
F00	19.3	-109.3	6.7	9.0	4.3	-2.5	-1.0	86.0
Enrique values	lat	lon	vor	nsext	def	mzw	vsheq	rh
F48	9.9	-114.2	4.6	6.0	3.6	-5.8	-9.1	88.1
F42	9.9	-115.8	4.3	6.0	3.3	-5.3	-7.5	89.6
F36	9.9	-118.2	4.9	6.0	3.5	-4.9	-7.3	90.6
F30	9.9	-119.0	4.9	6.0	3.5	-4.5	-9.2	84.0
F24	9.9	-119.8	4.9	5.0	3.8	-3.2	-12.4	90.0
F18	9.9	-123.1	4.8	5.0	3.4	-3.2	-9.1	89.2
F12	9.9	-122.3	5.5	6.0	4.2	-2.5	-11.3	89.9
F06	9.9	-123.1	6.2	6.0	5.0	-2.5	-12.0	90.6
F00	11.5	-123.1	7.9	7.0	6.1	-2.4	-12.4	88.6

Table 3.4 The values for each parameter in *alltime99* for TS Gil and TS Enrique at each forecast time.

the means of the DEV group as a whole (Table 3.2) may shed some light on why the forecasts from Bayes model vary.

The forecasts for TS Gil are relatively consistent for all nine forecast times, except for the 42-hour forecast. Looking at the parameter breakdown of Gil, it is clear that its northwestward propagation keeps the vortex over the Main Development Region (MDR) of the east Pacific basin for the entire 48-hour period, and both the latitude and longitude values are close to the overall mean of the DEV group. The same is true for the north-south extent of the vortex, the total deformation, and zonal wind. Since Bayes model uses the *distribution* of the parameter values rather than the *mean* parameter values to determine whether a vortex will develop or not, comparing the mean values can not definitively identify the cause for misclassifications. However, comparing the parameter values of an individual vortex to the parameter means of the DEV and ND groups can *suggest* which parameters may lead to misclassifications. The fact that a large number of parameter values for TS Gil are close to typical values seen in the DEV group would skew the Bayes model toward higher probabilities for development for TS Gil. The only parameter values that resemble those typified by the ND group are the relative vorticity and relative humidity, which both deviate toward ND values at the 42-hour forecast time. A similar pattern is seen at the 36-hour forecast time, which is the second lowest probability given by the model for this storm.

The model consistently produces low probabilities for the likelihood that TS Enrique would develop. This is not entirely surprising based on the values of the parameters that characterize this vortex. The mean zonal wind surrounding the vortex is easterly as is the case for many of the ND vortices. Likewise, the latitude and longitude values as well as the north-south extent of the vortex and its deformation field all favor that of a ND vortex. The only

parameters that are indicative of a developing vortex are the high relative humidity values (which are often times higher than the values associated with TS Gil), and the equatorward vertical shear of the meridional wind. Only at the time of declaration does the relative vorticity increase to a magnitude typically associated with a developing vortex, while the zonal wind becomes less easterly. The combination of high relative humidity and vorticity along with the large magnitude of the equatorward vertical shear and weakening of the easterly winds must be enough to trigger a high probability forecast by the model.

3.5 Summary/Discussion

a. Vortex-centered variables

Many of the vortex-centered variables that were found to be important for the LDA algorithm proved to be influential to the Bayes model as well. These variables included the latitude of the vortex center, its relative vorticity, total deformation, relative humidity, and the equatorward vertical shear of the meridional wind. Just as for the LDA algorithm, higher vortex relative vorticity, higher relative humidity, larger values of total deformation, larger magnitudes of equatorward vertical shear of the meridional wind, and weaker easterly winds were associated with the DEV cases. (For a more extensive description of how these parameters may affect the development of a vortex, see Chapter 2.)

Two notable differences between the parameter set used in the LDA algorithm and the statistical model used here are the additions of the longitude of the vortex and the north-south extent of the vortex. Both of these variables were important to the Bayes model forecast but not as significant when included in the LDA model. Since Bayes model is able to retain the information given by the shape of the parameters' density functions, while LDA relies only on the means of the two groups, these two parameters added skill to the Bayes model forecasts.

The longitude of the ND vortices varies widely from the west coast of Central America out to 140˚W, while the locations of the DEV vortices are confined between approximately 90˚W and 116˚W. This is not surprising as this area roughly defines the longitudinal boundaries of the east Pacific MDR, where the most favorable conditions (e.g. warmest SSTs, enhanced convection) for genesis occur. The relative significance of the north-south extent of the vortex to the predictive skill of Bayes model was slightly surprising. The fact that larger extents are associated with DEV cases could imply that larger vortices have a higher likelihood of developing, but this is purely speculation. More investigation into this feature is needed in order to make a more accurate assessment.

b. Madden-Julian Oscillation

The MJO has been shown to influence atmospheric conditions including 850hPa zonal wind and convection in tropical basins around the world (Wheeler and Hendon, 2004) including the eastern Pacific basin. During Phases 1, 2, and 8, the eastern Pacific basin tends to have anomalously low outgoing long wave radiation (OLR) embedded in a westerly flow. Both the wind pattern and enhanced convection, as indicative of lower OLR, increase the likelihood for development in the region (Davis et al. 2008; Maloney and Hartmann 2001). The remaining phases of the MJO create less favorable conditions over the basin. For these reasons, it was hypothesized that the phase of the MJO (along with the indices used to determine the phase) would increase the skill of one or both of the statistical models used in Chapter 2 and the current chapter. Therefore, a number of indices related to the MJO were included as parameters in both the LDA algorithm and Bayes probabilistic model. However, the addition of these parameters did not noticeably improve the LDA algorithm's skill in Chapter 2. Likewise, the addition of these parameters into the Bayes model did not improve

its skill (Figures 3.1 and 3.3). To investigate the reasons for this, a composite analysis was performed on some variables used to discern the phase of the MJO. For simplicity, a subset of the nine seasons of data was used, specifically the 2005-2008 seasons, which were relatively active seasons for tropical storms.

Two key variables Wheeler and Hendon (2004) used to determine the phase of the MJO were anomalous outgoing OLR and 850hPa zonal winds. Therefore, these two variables can be used to determine whether the MJO may have influenced the vortices considered in this study. If either composite resembles a phase of the MJO, it could indicate that the MJO helps to grow or inhibit vortex growth in the eastern Pacific depending on its phase. It was hypothesized that if the influence is strong enough, the MJO could be useful in distinguishing between developing and non-developing vortices.

Figure 3.8 shows a comparison of the developing and non-developing anomalous OLR and 850hPa winds for the 48 hours leading up to the time of declaration for the developing cases and the first 48 hours for the non-developing cases. While neither composite is a perfect match to the boreal summer composite phases of the MJO, as shown in WH2004_Figure from Chapter 2 (originally Fig. 9 of Wheeler and Hendon 2004), the developing composite bears some resemblance to both the second and third phases. Anomalously low OLR is embedded in a narrow, westerly flow over the eastern Pacific just as is the case for the second phase of the MJO. The pattern of low OLR over the Indian Ocean is also similar to the pattern found in Phase 2; however the wind pattern is more similar to the eighth phase. The fact that signals from three phases of the MJO are visible in this composite suggests that the majority of developing cases occurred during either Phase 1, 2, or 8 of the MJO. Indeed, according to the histogram (Figure 3.9) a little over half of the

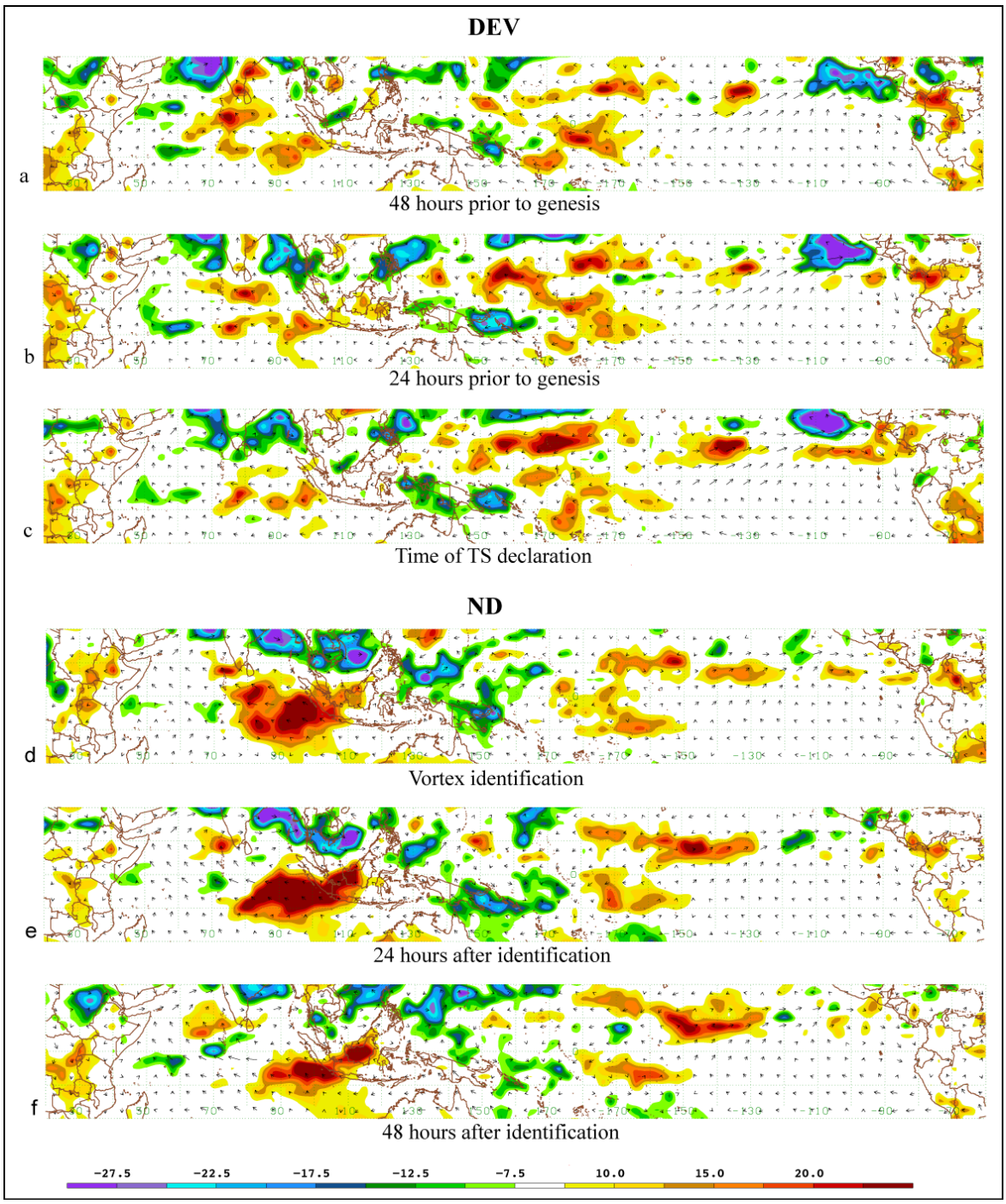

Figure 3.8 Composite of anomalous outgoing long wave radiation (OLR) and 850hPa winds along the equator for 19 developing (a-c) and non-developing (d-f) vortices. Positive OLR anomalies greater than 7.5 Wm^2 are shaded in warm colors, while cool colors indicate anomalous OLR values less than -7.5 Wm^2.

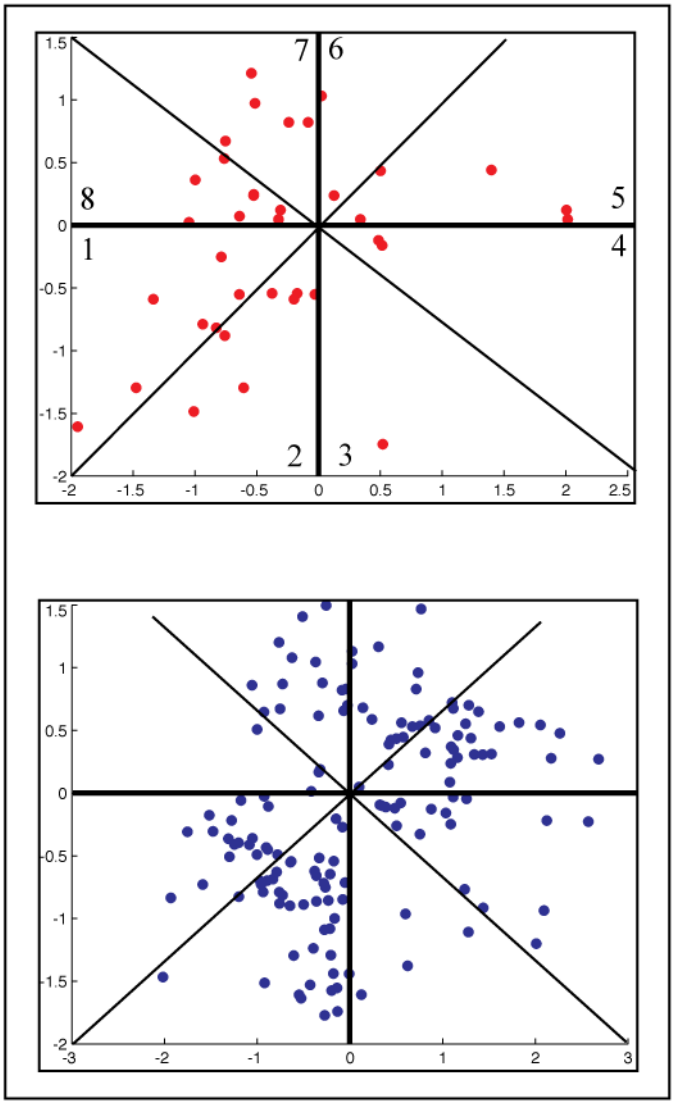

Figure 3.9 The phase space diagrams for the DEV (top) and ND (cases) during the 2005-2008 peak hurricane seasons.

developing vortices in this study did occur during one of the three phases. One other interesting aspect about the DEV composites is the anomalously strong southwesterlies that extend from the equator and 140˚W into the east Pacific. These winds appear to feed into the anomalously low OLR over the basin and do not appear in the ND composite.

The non-developing composite does not show any signal of the MJO at any time. There is no evidence of anomalously low OLR in either the Pacific or Indian basins, nor is there a signature in the wind field. Just as a strong signal in the developing composite was due to many of the developing storms occurring during one of three phases of the MJO, the lack of signal in this composite could stem from vortices occurring during any phases of the MJO. Figure 3.9 also shows this to be the case with at least one of non-developing vortices occurring during each phase.

From the bar graphs shown in Figure 3.9, it is apparent that a 60% majority of the developing vortices that formed during the 2005-2008 seasons originated during Phases 1, 2, or 8 of the MJO. Also, roughly 55% of the non-developing vortices originated during less favorable phases of the MJO. Therefore, it seems as though the MJO does influence the development of vortices in the eastern Pacific, yet including its phase as a parameter in the LDA algorithm did not prove effective. This apparent contradiction may be explained by the cyclical nature the MJO.

Recall that phases 1, 2, and 8 of the MJO all create favorable conditions for development over the basin, while the rest do not. Therefore, even if the occurrences of the developing vortices were equally distributed amongst only favorable phases of the MJO, and the occurrences of the non-developing vortices were distributed equally amongst only non-favorable phases, their averages would be equal to 3.7 and 5.0, respectively. The difference

between these averages did not pass the t-test at the 99% confidence level and only passed at the 95% confidences level at a few of the individual times, which would make distinguishing between the two difficult in terms of the LDA algorithm. However, approximately 40% of developing (45% of non-developing) cases occur during phases that are unfavorable (favorable) to development, which would likely bring the averages of the groups closer. Indeed the phase parameter averages for the 2005-2007 calibration seasons for the developing and non-developing groups are 4.6 and 3.9, respectively. This could explain the lack of improvement provided by adding the phase of the MJO to the LDA algorithm.

While the cyclical nature of the MJO may explain why adding its phase as a parameter in the LDA algorithm did not add skill, it does not explain why it has no impact on the skill of Bayes model. There is another aspect of the MJO indices that can explain why its addition did not add skill for either model. The MJO indices defined by Wheeler and Hendon (2004) result from two variables: the zonal wind and OLR. However, once the MJO propagates outside of the Indian basin the indices largely capture the circulation of the MJO and no longer retain the signal from the OLR (Paul Roundy – personal correspondence). Therefore, over the east Pacific basin, the phase of the MJO is basically another measure for the zonal wind. Since the zonal wind is already a parameter in the model, the MJO offers redundant information, which would explain why its addition does not change the skill of the model. Therefore the MJO (at least its phase as defined by Wheeler and Hendon 2004) may, at times, create favorable conditions for genesis over the east Pacific, but does not seem to have predictive skill for genesis – at least not on a day-to-day forecast timeframe.

3.6 Conclusions/Future Work

a. Bayes model versus the LDA algorithm

This study has shown that Bayes probabilistic model has significant forecast skill for predicting tropical cyclogenesis in the east Pacific basin out to 48 hours using a relatively small set of features. A probabilistic model, like that used in this study, provides more information regarding its forecast than does a binary model like the algorithm in Chapter 2. For example, the probability given by the model describes the confidence of the models prediction more so than a strict "yes" or "no" forecast. In addition, the fact that Bayes model does not rely on the means of the predictands allows for the potential of increased forecast skill for some features. One example resulting from the current research is the longitude of the vortex center. Since the means between the DEV and ND groups are similar, the addition of this parameter in the LDA algorithm did not add forecast skill. However, since the *distribution* of the longitude was different between the DEV and ND groups, Bayes model was able to gain forecast skill from this parameter. One line of future research could explore other features that have similar means, but different distributions to see if they could add forecast skill to this model as well.

b. Bayes classifier

Using a simple decision rule in which a "yes" forecast was assigned to vortices whose probability for development was greater than 50%, the Bayes model performance was compared to that of the LDA algorithm. The Bayes model had fewer false alarms than the LDA algorithm, but it also misclassified more DEV cases. Some of these misclassifications may have been the result of using a sub-optimal decision rule. In the future, a more accurate decision rule could be implemented by using the ROC or lowing the probability threshold for longer-term forecasts. However, if – with a more optimal decision rule – the Bayes model

continues to produce fewer false alarms, it would be interesting to examine how this model improves this statistic compared to the LDA algorithm.

c. Madden-Julian Oscillation

Surprisingly, the MJO indices in this study did not improve the forecast skill for either model, even though it has been shown to modulate TC activity in the basin (Maloney and Hartmann 2001). Due to the cyclical nature of the phase of the MJO, it was determined that the means of the MJO phase between the DEV and ND groups were too similar for the LDA algorithm to extract any forecast skill from the parameter. In addition, the MJO phase largely carries only the zonal wind signal of the oscillation once it leaves the Indian basin. Since the zonal wind is already considered as a parameter in both of the statistical models, the phase of the MJO may have been a redundant parameter and, as a result, did not improve the skill for Bayes model. However, the MJO may be a feature to consider when looking at longer-term (three to four week) forecasts as it does modulate the environment of the east Pacific basin and has been shown to demonstrate predictive skill over this time frame (Leroy and Wheeler 2008).

d. TS Enrique

The complete failure of either statistical model to predict Enrique is interesting. The parameters selected for use in the models are selected to maximize the number of hits. The parameter values for Enrique mostly seemed to indicate that the vortex would not develop. It was in a strongly sheared environment, and yet it did reach tropical storm status. One feature that seemed important to TS Enrique's development was the meridional barotropic conversion term, $-\frac{\partial \overline{u}}{\partial y}\overline{u'v'}$, which was not found to add skill to either model and was subsequently

discarded. This could indicate that a subset of tropical cyclogenesis events in the east Pacific result from barotropic growth. TS Enrique proves to be an interesting case study, which could shed further light onto this type of genesis mechanism. Indeed, TS Enrique raise a number of questions regarding genesis and its predictability in this basin. 1) Can a non-standard model be developed for these cases? 2) Do a subset of the misclassified DEV share characteristics with TS Enrique? 3) Can the meridional barotropic conversion term (or other barotropic conversion terms) be incorporated into either the LDA algorithm or Bayes model for cases like Enrique to improve forecast skill? 4) If Enrique does result from features atypical to development in the east Pacific, what might this non-standard model for development look like? All of these questions could guide future work for tropical cyclogenesis in this basin.

Chapter 4: Constructing a Composite-based Conceptual Framework for Tropical Cyclone Formation

4.1 Introduction

There are a number of favorable features for tropical cyclone (TC) formation. A pre-existing low-level vortex must be located over a region of sea surface temperatures warmer than 26°C that is at least 5° away from the equator (Palmén 1948; Gray 1968,1979). In addition, the environment surrounding the vortex must be characterized by weak vertical wind shear and a moist low- to mid-troposphere (Gray 1968,1979; Bister and Emanuel 1997). There are a number of regions in the tropics that have environments that meet these conditions for substantial portions of the year, yet very few vortices reach tropical storm or hurricane strength – indicating that other physical features must influence the development of TCs.

Perhaps the least studied basin (in terms of tropical cyclogenesis) in the Northern Hemisphere is the east Pacific basin. This basin has perhaps the most spatially confined main development region (MDR) of all other basins with 75% of its TCs forming an area encompassed by 95°W to 115°W, and 6°N to 18°N (Davis et al. 2008). Despite its small size, the basin is second only to the west Pacific basin in terms of TC activity, making it the most concentrated area of TCs in the world. While this region produces many TCs to investigate, few studies have been conducted concerning east Pacific TC genesis. One recent study focused on large-scale features that distinguish between developing and non-developing vortices and found that relative humidity, lower-tropospheric horizontal stretching deformation, and the equatorward vertical shear of the meridional wind all had significant

differences between developing and non-developing vortices in east Pacific basin (Davis et al. 2008).

The current study will build on the findings of the Davis et al. (2008) study as well as two statistical studies (described in Chapters 2 and 3) focused on east Pacific TC genesis predictability in order to construct a composite-based conceptual framework for TC development. Section 2 will describe the data and methodology used in the study. The composite results of the basin-wide composites are summarized in section 3. The results of the vortex-centered composites along with a vortex-centered conceptual model for development are presented in section 4. The time-progression composites of the strongest tropical storms (that also exist 48 hours prior) are discussed in section 5, while the conclusions and future work are summarized in section 6.

4.2 Data and methodology

All variables calculated for the composites used in this study are derived from National Centers for Environmental Prediction (NCEP) six hourly, 1°x1° final reanalysis data[5]. The data fields include zonal and meridional wind components, relative humidity, vertical velocity, temperature, and height on twenty pressure levels from 1000hPa to 150hPa. The dates included in the data set range from 2001 to 2009 for the peak hurricane season (July-September) of the east Pacific.

All vortices considered in this study must have a minimum vortex relative vorticity of $1.5 \times 10^{-5} s^{-1}$ for at least 48 consecutive hours. Data fields are composited for three sets of vortices that meet these criteria. The first set of cases consists of the 25 strongest developing

[5] The data for this study are from the Research Data Archive (RDA), maintained by the Computational and Information Systems Laboratory of the National Center for Atmospheric Research (NCAR). The original data are available from the RDA (http://dss.ucar.edu) in dataset number ds083.2.

vortices 48 hours prior to genesis. The developing cases are composited at a time prior to genesis because the intent of this study is to compare the difference between developing and non-developing nascent vortices; it is not the intent to compare tropical storms to non-developing vortices. The specific time (48 hours prior to declaration) is selected because the vortices should be in an environment conducive to genesis without being on the immediate cusp of being declared as a tropical storm. The remaining two groups consist of non-developing vortices. One is a random set of 25 non-developing vortices, and the second is the weakest 25 non-developing vortices. Both non-developing groups are composited at 24 hours after vortex identification. The strength of the vortices are determined by the magnitude of the relative vorticity averaged over a 5°x5° box centered on each vortex.

A final aspect of this study addresses the question: Do the most intense vortices at 48 hours prior to genesis remain the strongest vortices at their time of declaration? To address this question, the 25 strongest vortices at their time of declaration as a tropical storm (that also exist 48 hours prior) are tracked back 48 hours. The cases included in this data set are compared to those in the initial data set (described above) and reveal a 52% discrepancy between the two data sets. More specifically, 12 vortices were the strongest at both declaration and 48 hours prior, and 13 were weaker at 48 hours prior but grew to become part of the strongest tropical storms. An investigation of this discrepancy is conducted.

4.3 Climatology of the east Pacific basin

The main development region (MDR) of the east Pacific basin is spatially confined between 115°W and 95°W, and 8°N and 16°N and harbors approximately 75% of the 9 to 15 tropical storms and four to eight hurricanes that form annually (NHC) in the basin (Davis et al. 2008). The number of named storms and the most intense hurricanes tend to form during

the months of July, August and September (NHC). A brief investigation of the monthly climatology of the east Pacific environment demonstrates why the "peak season" for hurricanes is July-September.

a. 30-year climatology (1980-2010)

A number of climatological features have been found to be necessary, but not sufficient, conditions for tropical cyclogenesis. These include minimum SSTs of 26°C, a moist low-tropospheric environment, weak vertical wind shear, background vorticity, and the presence of a nascent vortex (Palmén 1948; Rosendal 1962; Gray 1968, 1979). In the east Pacific basin, additional favorable conditions for genesis include enhanced convection and low-level anomalous westerly winds (Maloney and Hartmann 2000, Davis et al. 2008). In order to understand the existence of the peak season of tropical cyclogenesis events in the east Pacific, 30-year (1980-2010) monthly climatologies of 925hPa winds, SSTs, and outgoing long wave radiation (OLR) fields were constructed using data from the National Oceanographic and Atmospheric Administration's (NOAA) Earth System Research Laboratory[6] (ESRL).

Over the course of the year, the SSTs present over the east Pacific MDR remain at or above 26°C (Figure 4.1). The months in which the SSTs remain above 26°C over the entire MDR domain (denoted by the black box) include July through November. Coincident with the warm SSTs is a climatological decrease in OLR. While the decrease in OLR occurs over the MDR from May to October, it is most pronounced during July, August, and September. Since low values of OLR are indicative of enhanced convection, the environment over the

[6] http://www.esrl.noaa.gov/psd/

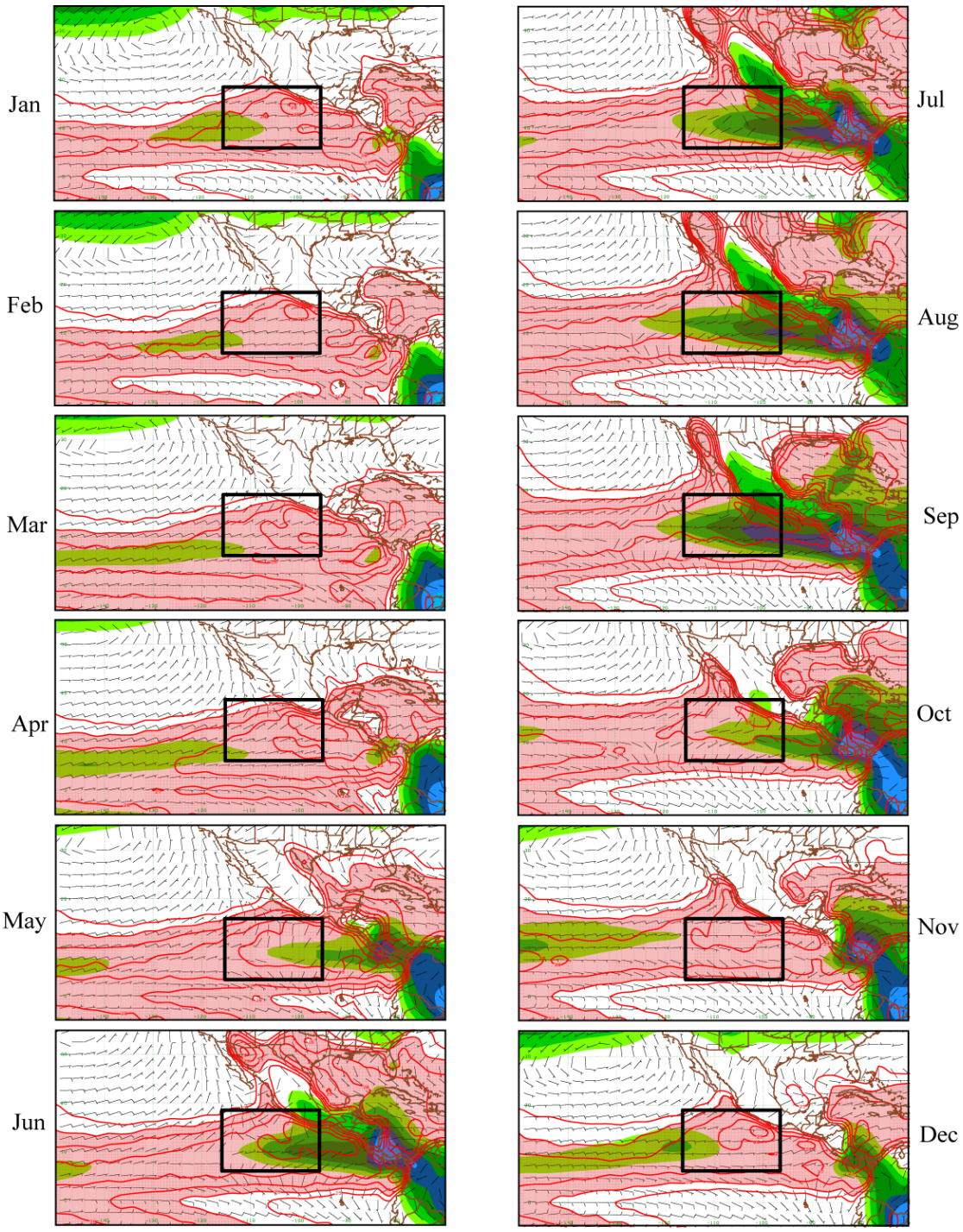

Figure 4.1 The 30-year monthly climatological SST (contoured in red with red shading for values of 26°C and higher) and OLR (shaded in cool colors every $10Wm^{-2}$ with the maximum value at $240Wm^{-2}$).

east Pacific during the summer months is characterized by a warm ocean surface and a moist, convecting atmosphere (Figure 4.2).

The low-level winds also vary from month to month. For example, easterly trade winds exist through most of the year at the 925hPa level over the east Pacific basin. However, a reversal in the zonal winds begins to occur in May extending off of the west coast of Colombia forming a small westerly jet (Figure 4.3). This jet strengthens and extends westward until it reaches a peak in intensity and length in September before retreating in November and becoming virtually non-existent in February. Such a westerly jet creates a favorable environment for tropical cyclogenesis over the basin for two reasons. First, it creates a cyclonic shear along its northern side providing the region with cyclonic background relative vorticity, and second it creates an anomalous westerly wind regime over the MDR, which has been found to be favorable for genesis (Maloney and Hartmann 2000; Davis et al. 2008).

The combination of warm SSTs, a moist environment, positive background vorticity, enhanced convection, and westerly winds create a climatologically favorable environment for the formation of TCs in this basin if an initial disturbance is present. This is especially the case for the three-month period from July to September, when both the low level jet and convection are strong and sufficiently warm SSTs exist over the entire MDR. Given these conditions, one would expect the majority of TCs to occur during this three-month time span. Indeed, this is the case for the 2001-2009 seasons. The number of hurricanes and tropical storms are largest during these three months (Figure 4.4). July has the largest number of tropical storms followed by August and September. All three months have the same number of category 1 and 2 hurricanes (9). Most of the major hurricanes (10) occurred in August,

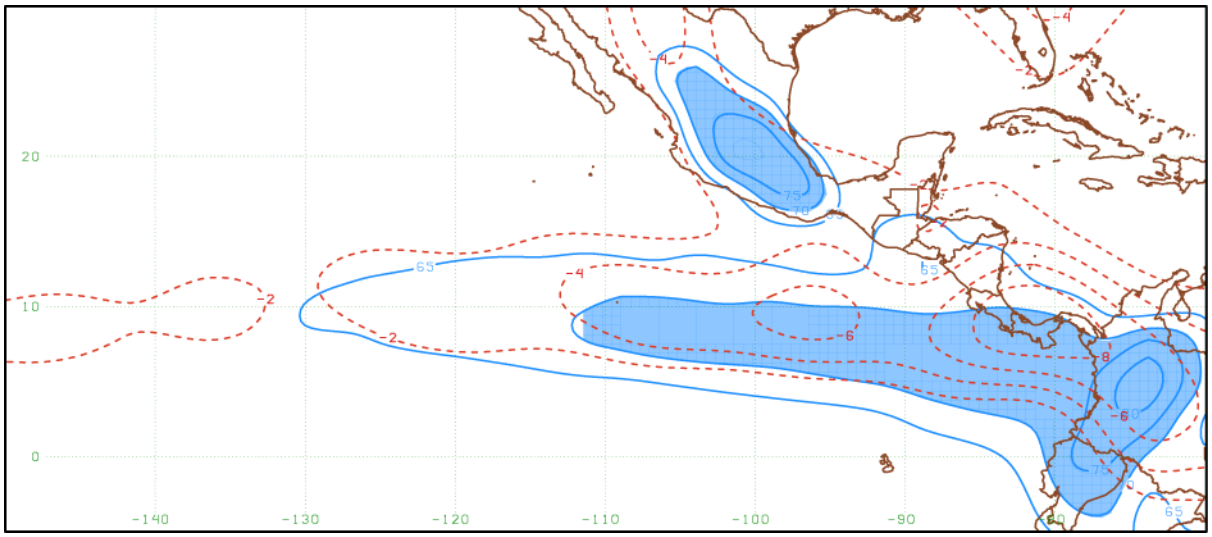

Figure 4.2 The JAS averaged 900hPa-500hPa relative humidity (contoured in blue every 5% with values higher than 70% shaded) and omega (contoured in red every 2Pas^{-1}).

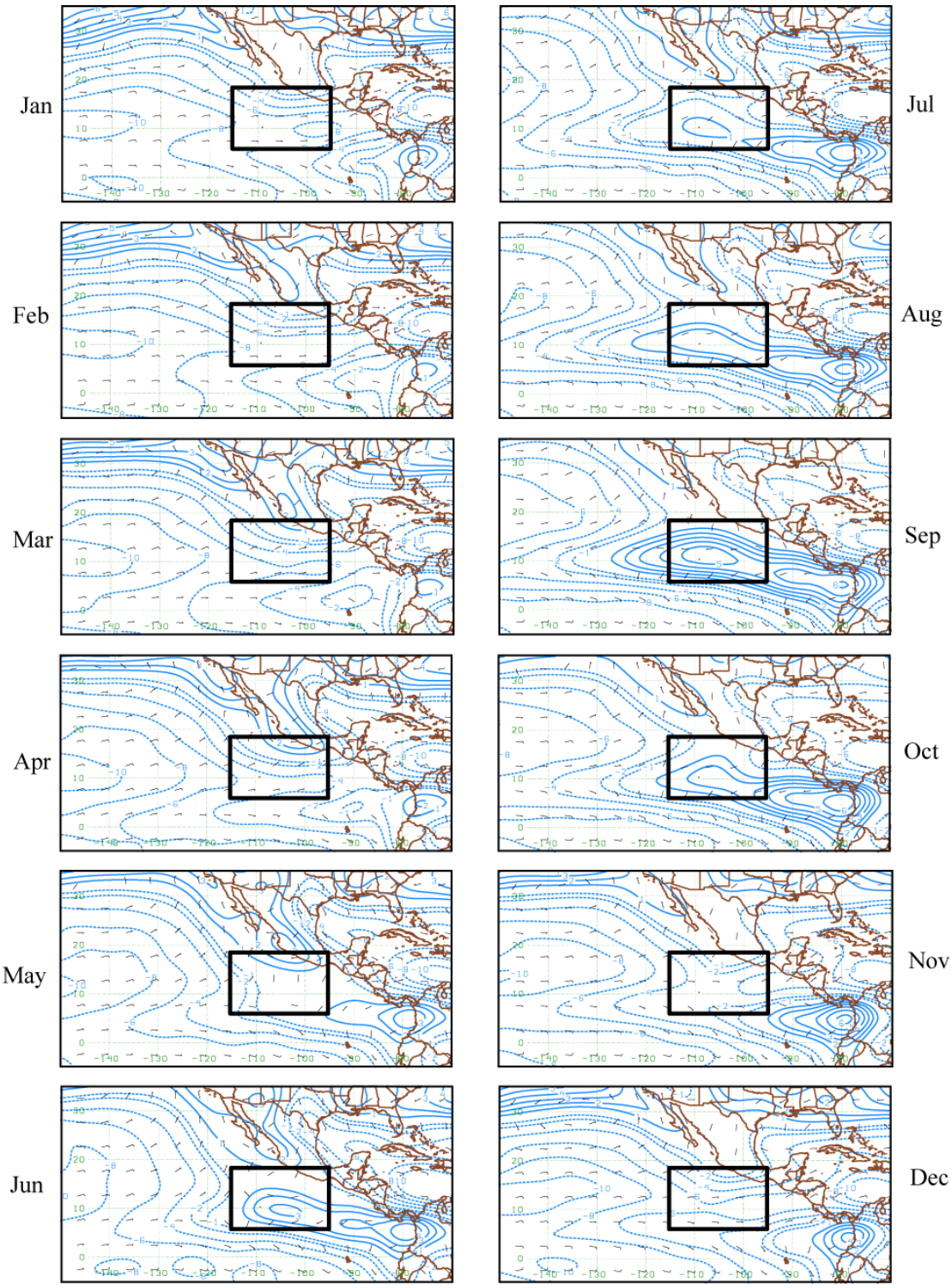

Figure 4.3 The 900hPa zonal wind plotted every 2ms^{-1} for easterly winds (represented by dashed lines) and 1 ms^{-1} for westerly winds (solid contours). The black box indicates the east Pacific MDR.

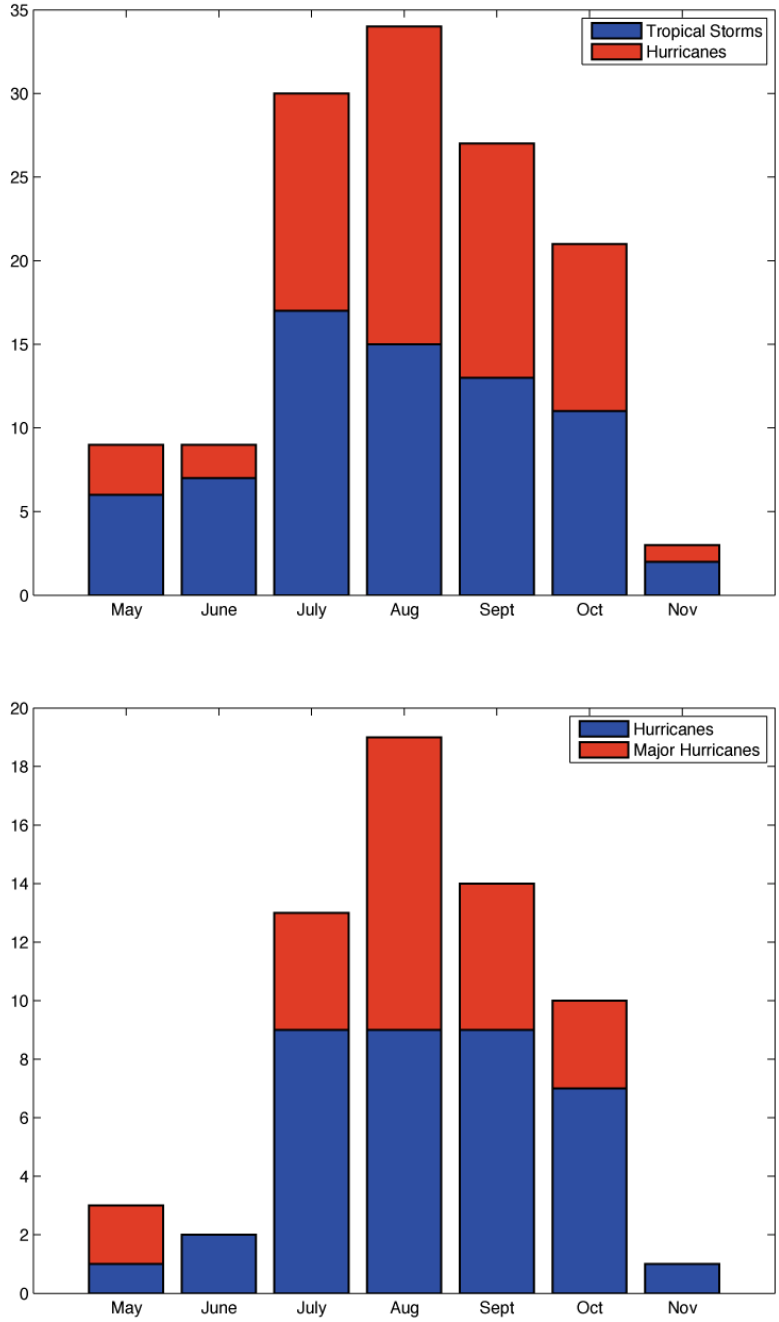

Figure 4.4 Nine season counts of east Pacific tropical storms and hurricanes (top panel), and hurricanes versus major (Category 3 or stronger) hurricanes (bottom panel).

while July and September had fewer with four and five, respectively. Since these three months were the most active for the 2001-2009 hurricane seasons, the vortices considered in this study will be limited to those that developed during the July-September (JAS) months.

b. Nine-season climatology (2001-2009)

As stated above, this study focuses on the nine-season period from 2001-2009. Basin-wide composites of climatological features for these nine seasons are compared to the 30-year climatology to identify any deviations between the two (Figure 4.5). The JAS OLR for the east Pacific is slightly lower over the MDR (4 to 10Wm^{-2}) for the 2001-2009 seasons compared to the climatological values for the same three-month span, while the SSTs over the same area are 0.1°C to 0.3°C warmer than climatology. Furthermore, the low-level zonal jet is slightly stronger (up to 1ms^{-1}) than climatology. While it is not within the scope of this study to determine the variability of tropical cyclogenesis on a decadal time-scale, it should be noted that these differences indicate that the environment may have been slightly more favorable for tropical cyclogenesis than the previous twenty years, and may be an interesting avenue for future research. For the purposes of this study, these results show that the 2001-2009 peak-season climatology is one that is favorable for TC formation in the basin. Indeed, 91 tropical storms and hurricanes were declared during the nine peak-seasons. However, over the same time period approximately 406 other vortices that lasted for at least 48 hours (see Chapters 2 and 3) never reached tropical storm status. The question remains, beyond the necessary climatological conditions, what other features influence the development (or lack thereof) of TCs in the east Pacific? The following sections present composites to identify such features, and develop a conceptual framework of some key elements that favor tropical cyclogenesis in this basin.

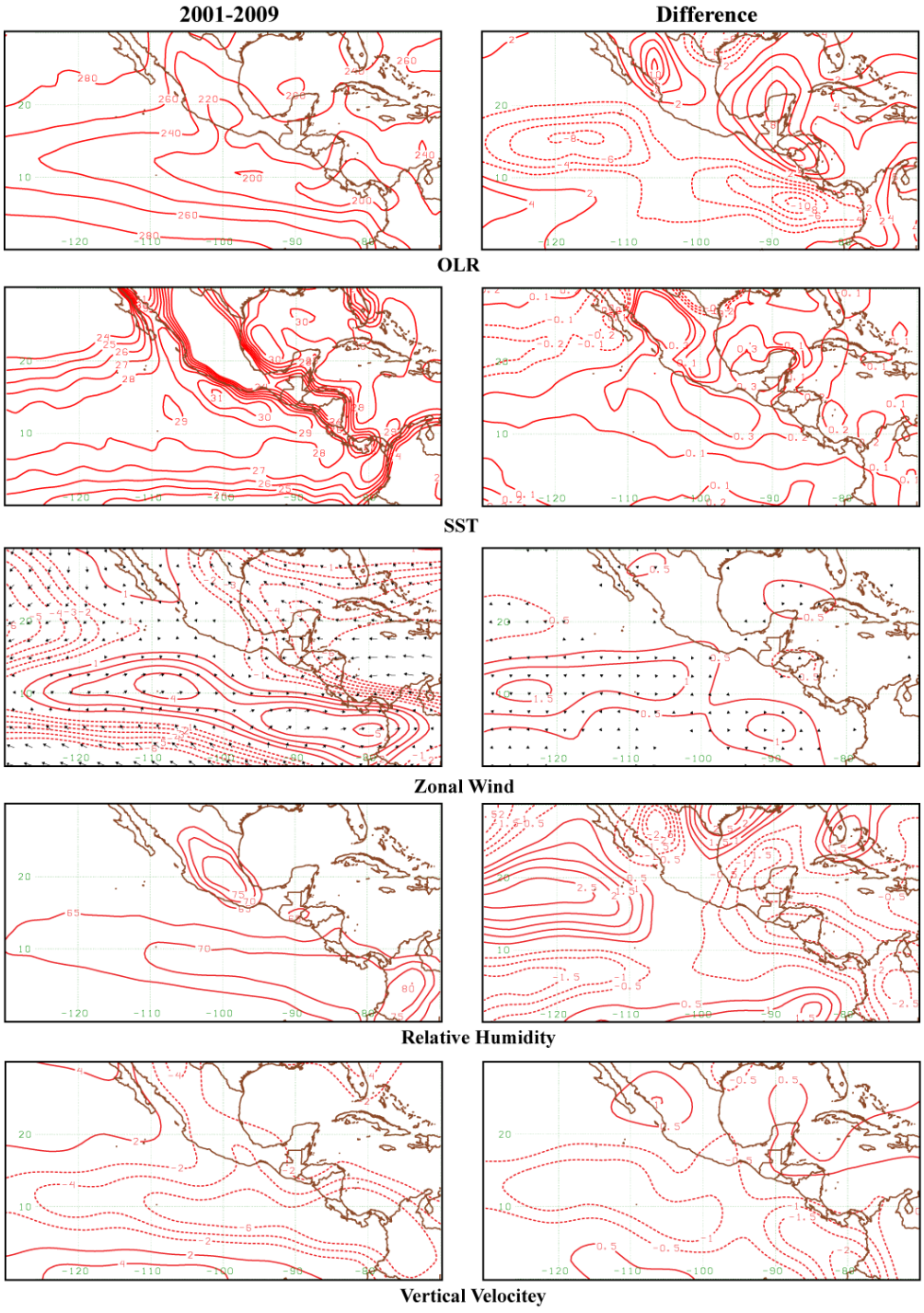

Figure 4.5 The 2001-2009 JAS climatology (left column) and its difference from the 30-year climatology (right column) for the OLR (plotted in Wm^{-2} in the top row), SST (second row in Celsius), 900hPa zonal winds (third row in ms^{-1}), relative humidity (fourth row) and vertical velocity (bottom row in Pas^{-1}).

4.4 Basin-wide composites

The focus is retained on three composite groups: a developing group consisting of 25 vortices 48 hours prior to their declaration as a tropical storm (DEV); a non-developing group consisting of 25 randomly selected non-developing vortices (RND); and another non-developing group consisting of the 25 weakest vortices (BND). Both of the non-developing vortex composites are taken at 24 hours after vortex identification (see section 4.2 for details). The following focuses on the basin-wide composites of these three groups.

(Figure 4.6) shows the 900hPa westerly zonal winds, upper-level (150hPa) full winds, and the location of the vortex centers for the DEV, RND, and BND composites. The 25 strongest vortices during the nine seasons are closely clustered such that the majority is located in the MDR. On the other hand, the locations of the non-developing vortices (both RND and BND groups) are scattered across the basin. Those vortices further west of 130°W are in a climatologically unfavorable environment with respect to SSTs and OLR, which likely contributes, in part, to their lack of development.

The presence of a strong low-level westerly jet is evident in the DEV composite, extending from the west coast of Colombia and Ecuador to approximately 125°W. All of the vortices in this composite are located along the center and poleward of this jet. This is not the case for either ND composite where only a fraction of the vortex centers are located to the north of the jet. The remaining vortices are located to the south of a relatively weak westerly jet or in regions in which no discernable westerly jet exists.

Unlike the lower-tropospheric wind field, the upper-tropospheric flow pattern for all three composites is quite similar. In all three composites, an upper-level anticyclone is present poleward of the lower-tropospheric jet and is centered near the Gulf of California.

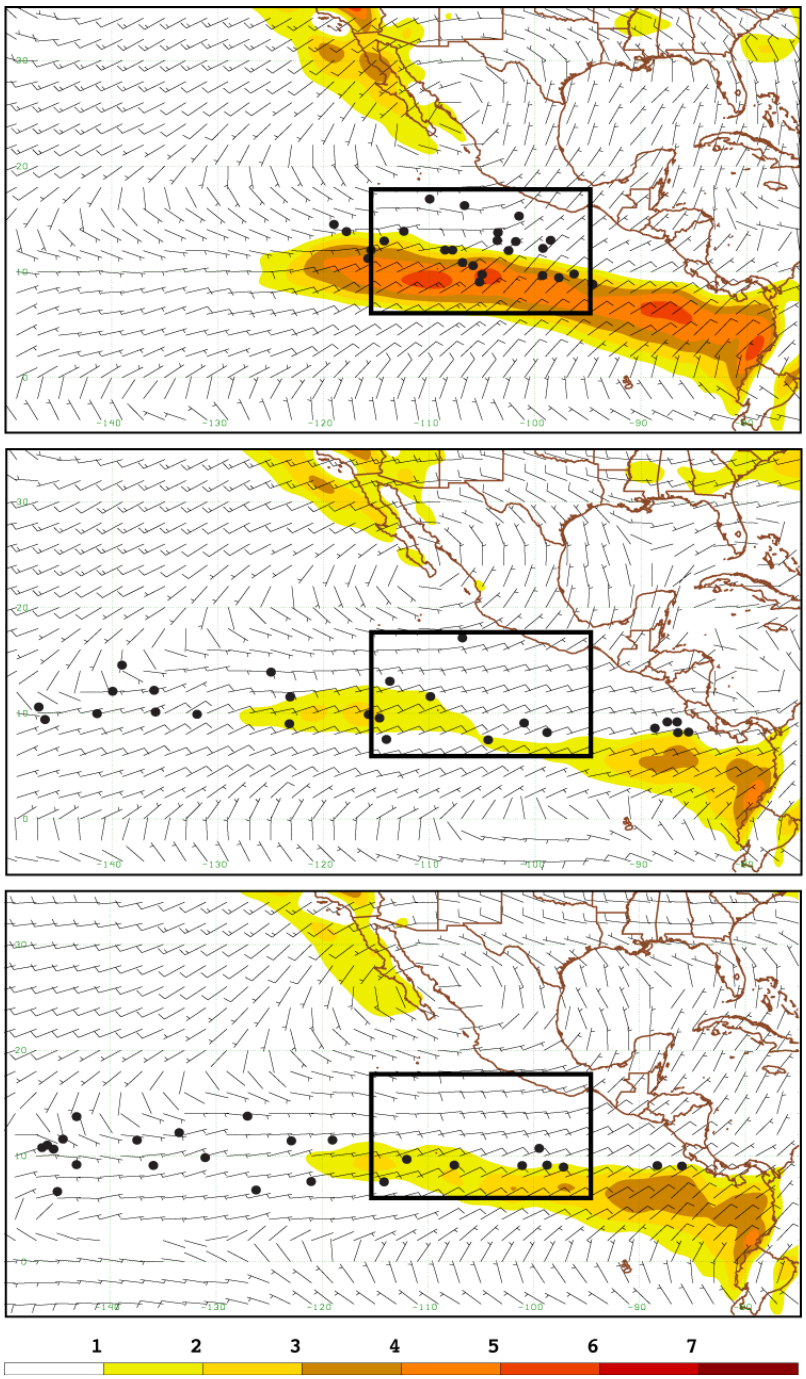

Figure 4.6 The low level (900hPa) zonal winds (shaded) along with the 150hPa wind field (wind barbs) and the vortex centers (represented by dots) for the DEV (top), RND (middle), and BND (bottom) composites. Winds are plotted in ms^{-1}.

The wind direction over the 900hPa jet at upper levels is similar for all three composites as well with northeasterly flow being predominant. The wind speed over the MDR in the DEV composite is slightly stronger than both ND composites. There is also a difference in the wind direction over the far west Caribbean extending over Panama and into the southeastern corner of the east Pacific basin. In this region, strong northeasterly winds are present in the DEV and BND composites, while the RND composite winds are more easterly off the west coast of Colombia and variable over the western Caribbean.

There are also noticeable differences among the three composites (Figure 4.7) in terms of low-level moisture and convection. All three composites have a moist lower-troposphere with minimum relative humidity of 75% over a large fraction of the basin at 850hPa. The relative humidity decreases at higher altitudes for all composites, but the DEV composite has a minimum relative humidity of 80% over the MDR (where the majority of the vortices are located) at 700hPa, and the atmosphere remains moist at 500hPa over the MDR with values ranging from 70% to 85%. The upward vertical motion over the basin is relatively strong between approximately 5°N and 15°N throughout the length of the basin. Velocities ranging from -15 to -30μbs^{-1} are present up to 500hPa across the region and specifically over the MDR. The concentration of upward vertical motion and high relative humidity over the MDR should be expected as the majority of DEV vortices are located here.

Like the DEV composite, both the RND and BND composites have high relative humidity values at 700hPa, though not over the MDR. The BND composite has a maximum in relative humidity to the west of 140°W, which coincides with a cluster of vortices at low levels. Interestingly, there is a local maximum in relative humidity between 120°W and 130°W in the RND composite (a region in which there are three nascent vortices) that does

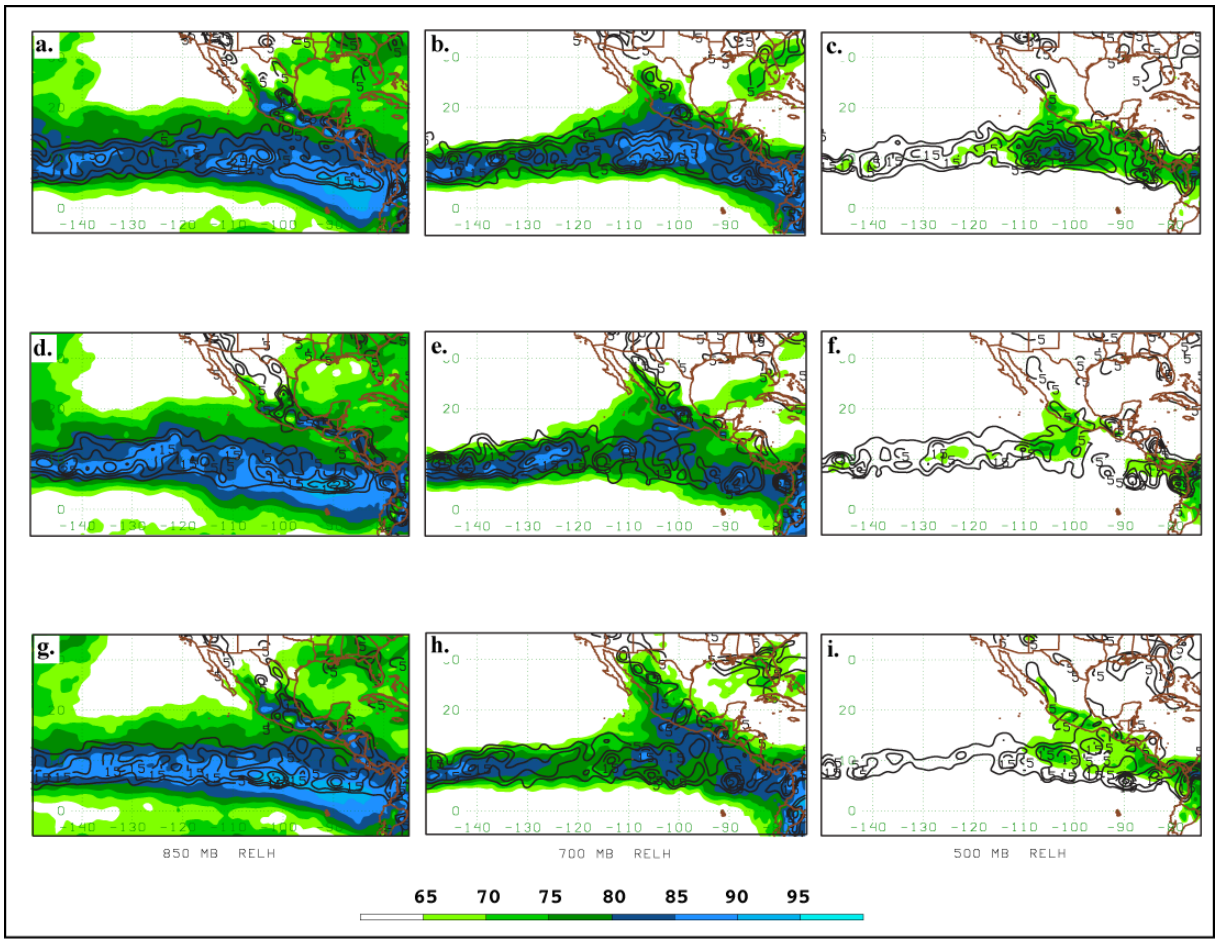

Figure 4.7 The relative humidity (fill pattern) and upward vertical motion (contoured every 5μbs^{-1}) for the DEV (a-c), RND (d-f), and BND (g-i) composites at the 850hPa (left column), 700hPa (middle column), and 500hPa (right column) levels.

not exist in the BND composite (where there are four nascent vortices present). The upward vertical motion (black contours) is also weaker over this longitude band for the BND composite than the RND composite. This could suggest that a lack of enhanced convection over the region may have inhibited the development of one or more of the vortices located in this region in the BND composite. However, the lack of convection could instead be a *result* of the how the composite is constructed as it consists only of weak vortices.

There are a number of distinct differences between the environments present during developing and non-developing vortex formation, especially at low-levels. The DEV composite environment is characterized by the presence of a strong (exceeding 5 ms^{-1}) westerly jet that extends from the coast of Columbia and Ecuador to approximately 125°W. The vortices in this composite tend to be located to the north of the jet on its western side. The low-level relative humidity is high and upward vertical motion is strong over most of the basin along the 5°N to 15°N latitude band at 850hPa for the DEV composite. The moisture and upward vertical motion remain high at 500hPa, though concentrated more over the MDR.

The non-developing environment is somewhat similar to the DEV environment in terms of relative humidity and upward vertical motion. Both composites exhibit relatively large values (though slightly smaller than the DEV composite) of upward vertical motion and relative humidity in the low and mid-troposphere. However, the most noticeable difference is the location of the vortices both in terms of their distribution in the basin and their location relative to the low-level jet. Unlike the DEV vortex locations, both the RND and BND vortices are scattered throughout the basin, and the zonal jet that is predominant in the DEV composite is very weak in the ND composites. The portion of non-developing vortices that lie to the west of 130°W may also be inhibited by the cooler SSTs. The remainder of this

study will focus on the dynamical differences between the DEV and ND composites by investigating them from a vortex-centered perspective.

4.5 Vortex-centered composites

a. low-level jet

As previously mentioned, the low-level flow appears to have an impact on the development of nascent vortices in the basin. As shown in the east Pacific climatology, one of the most prominent features in the lower-troposphere is the low-level westerly jet that forms during boreal summer. The basin-wide composites showed that this jet was stronger for the DEV composite than for the ND composites, but how does this jet compare from a vortex-centered perspective?

Figure 4.8 shows the low-level (925hPa) vortex relative vorticity for the DEV, RND, and BND composites as well as the low-level westerly jet at the 925hPa, 900hPa, and 850hPa levels. It is clear from the DEV composite that a strong westerly jet exists at low levels. While the westerly winds directly to the south of the vortex-center are part of the vortex itself, an obvious westerly jet extends eastward from the vortex with wind speeds ranging from $1ms^{-1}$ to $5ms^{-1}$. This jet persists at the 900hPa level before disappearing at the 850hPa level (above 850hPa, it can usually be inferred that the remaining westerly flow near the vortex is an expression of the vortex circulation alone). One interesting aspect of this composite is that the vortices seem to form on the western side of the low-level jet as evidenced by the absence of westerly winds to the west of the vortex center. In addition, the DEV composite vortex appears to tilt slightly along the shear line at 925hPa, while the RND and BND composite vortices do not. This tilt in the DEV composite vortex suggests that a portion of the vortices in the composite may have developed from barotropic growth from the jet.

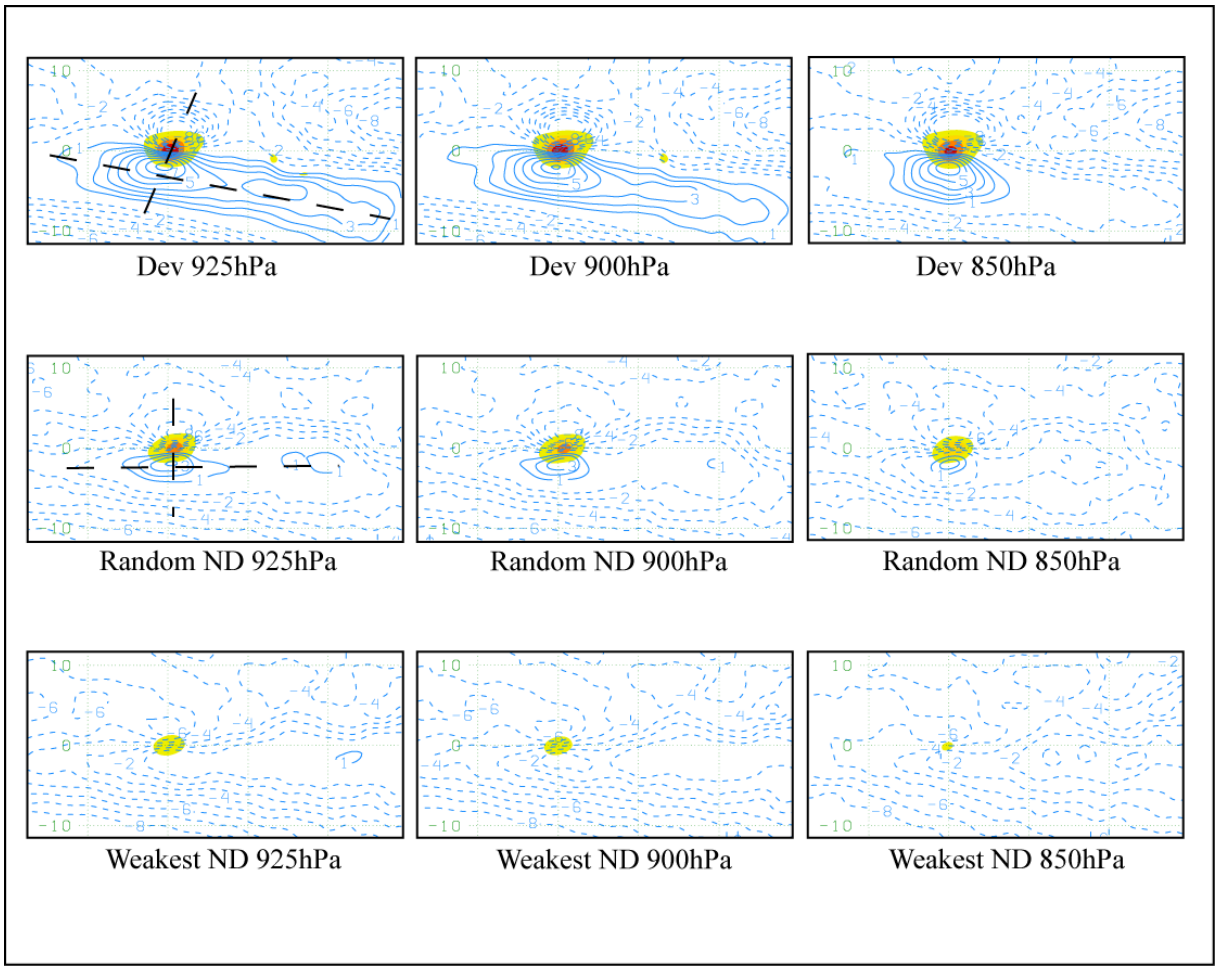

Figure 4.8 The vortex-centered composites of low-level zonal winds with easterly winds contoured every 2ms^{-1} (dashed lines) and westerly winds contoured every 1 ms^{-1}. Also plotted is the 900hPa relative vorticity (shaded) beginning with 2x10^{-5}s^{-1} (interval is also 2x10^{-5}s^{-1}).

The RND composites show little evidence of a lower-tropospheric westerly jet. Again, westerly winds to the south of the vortex are likely associated with the actual vortex winds. While there are small areas of westerly winds to the east of the vortex at the 925hPa and 900hPa levels, there is no evidence of a structured jet and no westerly winds exist outside of the vortex at 850hPa. The pattern continues for the BND composite in which the composite vortex is located in an almost exclusively easterly wind regime, which may be due to the large expanse of the ND vortex locations in the basin. Nonetheless, the difference between these composites indicates that the presence of the low-level jet during TC development.

b. lower-tropospheric deformation

Another feature influential to genesis at low levels is the low-level horizontal stretching deformation associated with the low-level jet, which has been shown to distinguish between developing and non-developing vortices with weaker values favoring genesis (Davis et al. 2008). For this study the stretching deformation as well as the shearing and total deformation fields were calculated for all three composite groups (Figure 4.9). Both the magnitude and distribution of all three variations of deformation differ for the composite groups.

The magnitude of the total deformation over the DEV composite vortex is stronger than the ND composite vortices. The magnitudes of both the stretching and shearing deformation fields are larger in the DEV composite due to the fact that the winds associated with the developing composite vortex are larger than the non-developing composite vortices. Since the total deformation is the sum of the strain rates, $D = \sqrt{D_{str}^2 + D_{shr}^2}$, it follows that the total deformation is larger for the DEV composite than the ND composites. Therefore, it is

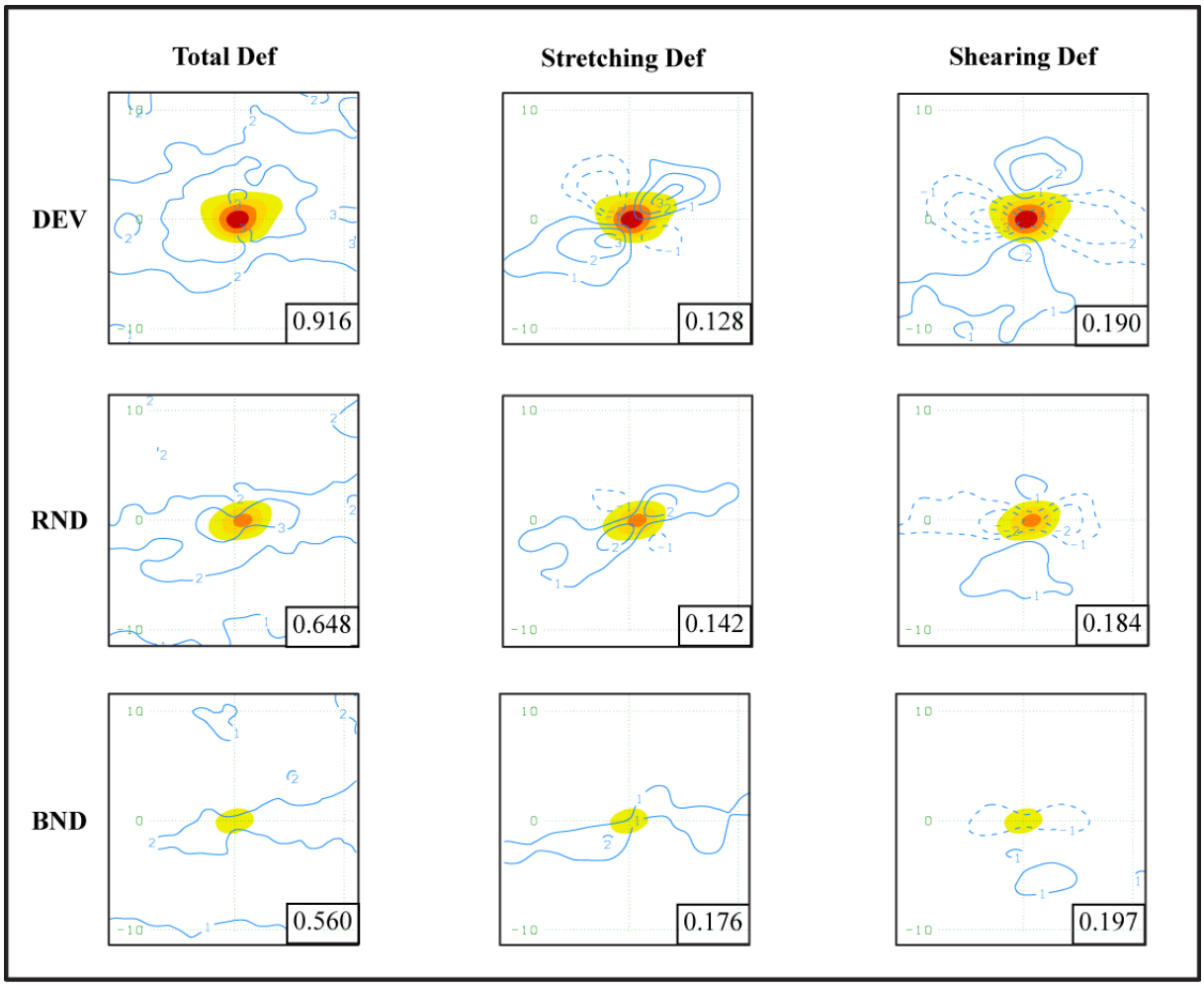

Figure 4.9 Contoured every $1x10^{-5}s^{-1}$ is the total (left column), horizontal stretching (middle column), and shearing (right column) deformation fields for the DEV (top row), RND (middle row), and BND (bottom row) composites. Also plotted in fill pattern is the relative vorticity every $2x10^{-5}s^{-1}$ starting with $2x10^{-5}s^{-1}$. The areal averaged deformation for each panel is plotted in the box in the lower right-hand corner of the panels.

important to look at the components that make up the total deformation to get a better understanding of the interactions between the vorticity and deformation fields.

Looking at the magnitudes of the components of the total deformation, it is clear that both are larger for the DEV composite. At first glance this is counter-intuitive based on results from Davis et al. (2008); however, the difference between the patterns in the deformation field between the composite groups is key. Both the stretching and shearing deformation fields over the DEV composite vortex have positive and negative sectors over the vortex leading to an overall decrease in the areal average of the stretching and shearing deformation. For example, the areal average of stretching deformation for the DEV, RND, and BND composites are 0.128, 0.142, and 0.176, respectively. This indicates that weaker vortices typically tend to be associated with stronger horizontal stretching deformation, which is in agreement with the findings of Davis et al. (2008). Similar results are seen for the shearing deformation in which areal averages of 0.19, 0.184, and 0.197 were found for the DEV, RND, and BND composites, respectively.

Another way to consider the relationship between the vortex vorticity and its surrounding deformation field is through the Okubo-Weiss parameter, $\omega = D^2 - \zeta^2$, where D is the total deformation and ζ is the relative vorticity of the vortex. Large negative values indicate the dominance of rotation over strain and are associated with closed circulations, which have been shown to protect the vortex from dry air intrusion and subsequent downdrafts (Dunkerton et al. 2009; Montgomery et al. 2010). The Okubo-Weiss parameter is thus another indicator of the potential for cyclone intensification (Raymond et al. 1998). The Okubo-Weiss parameter is largely negative ($-4 \times 10^{-9} s^{-1}$ to $-8 \times 10^{-9} s^{-1}$) over the DEV composite vortex and decreases for the weaker composite vortices (Figure 4.10). This gives further

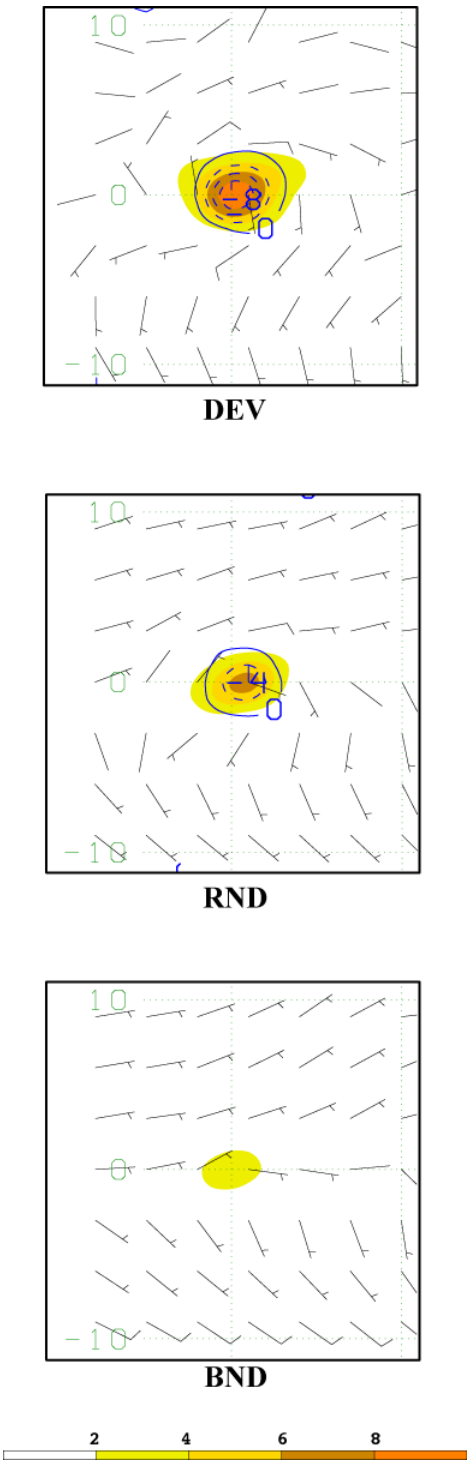

Figure 4.10 The Okubo-Weiss parameter (countered in blue every $2x10^{-9}s^{-1}$) superimposed over the 900hPa vortex relative vorticity (shaded every 2x10-5s^{-1}).

evidence that the vorticity of the vortex must be strong enough to overcome the background deformation acting upon it in order for the vortex to develop into TC.

c. Convergence, divergence, and mid-level relative humidity

The interconnections between the lower and upper troposphere also vary between the developing and non-developing composites. The DEV composite shows strong convergence over the vortex center, with a maximum magnitude of $2 \times 10^{-5} s^{-1}$ to the south of the vortex (Figure 4.11), and the relative humidity surrounding the vortex is at least 85%. Both of the ND composites are associated with similar relative humidity values of at least 85% surrounding the vortex, though the areal coverage is not as expansive as the DEV composite. Furthermore the convergence is half that of the DEV composite. The convergence in the DEV composite is also more concentrated near the vortex than either of the ND composites.

Another favorable environmental feature for tropical cyclogenesis is a weak vertical shear. Figure 4.12 shows the large-scale vertical shear of three wind fields surrounding the composite vortices along with their areal averaged magnitudes (located in the bottom right of each plot). First, the 900hPa-500hPa vertical shear of the zonal wind is calculated because it was identified as a good predictor in the statistical models (see Chapters 2 and 3). The DEV composite shows strong shear in the immediate surroundings of the vortex, but the shear to the south of the vortex is weaker in the DEV composite. Recall that low-level zonal westerly winds are present to the south of the vortex. The fact that wind shear to the south of the vortex is small compared to the shear north of the vortex indicates that westerly winds or weak easterly winds exist at 500hPa. A similar pattern is seen in the RND and BND composites indicating weaker westerly or stronger easterly winds at 500hPa compared to those winds at 900hPa.

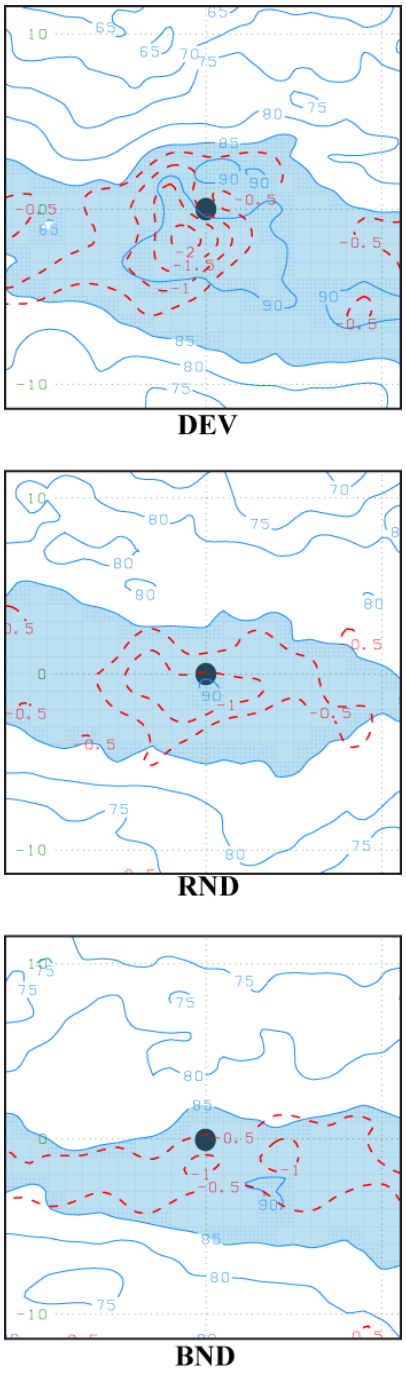

Figure 4.11 The 900hPa relative humidity (blue contours in intervals of 5%; values of 85% and higher are shaded) and divergence (contoured in red every 0.5 x$10^{-5}$$s^{-1}$).

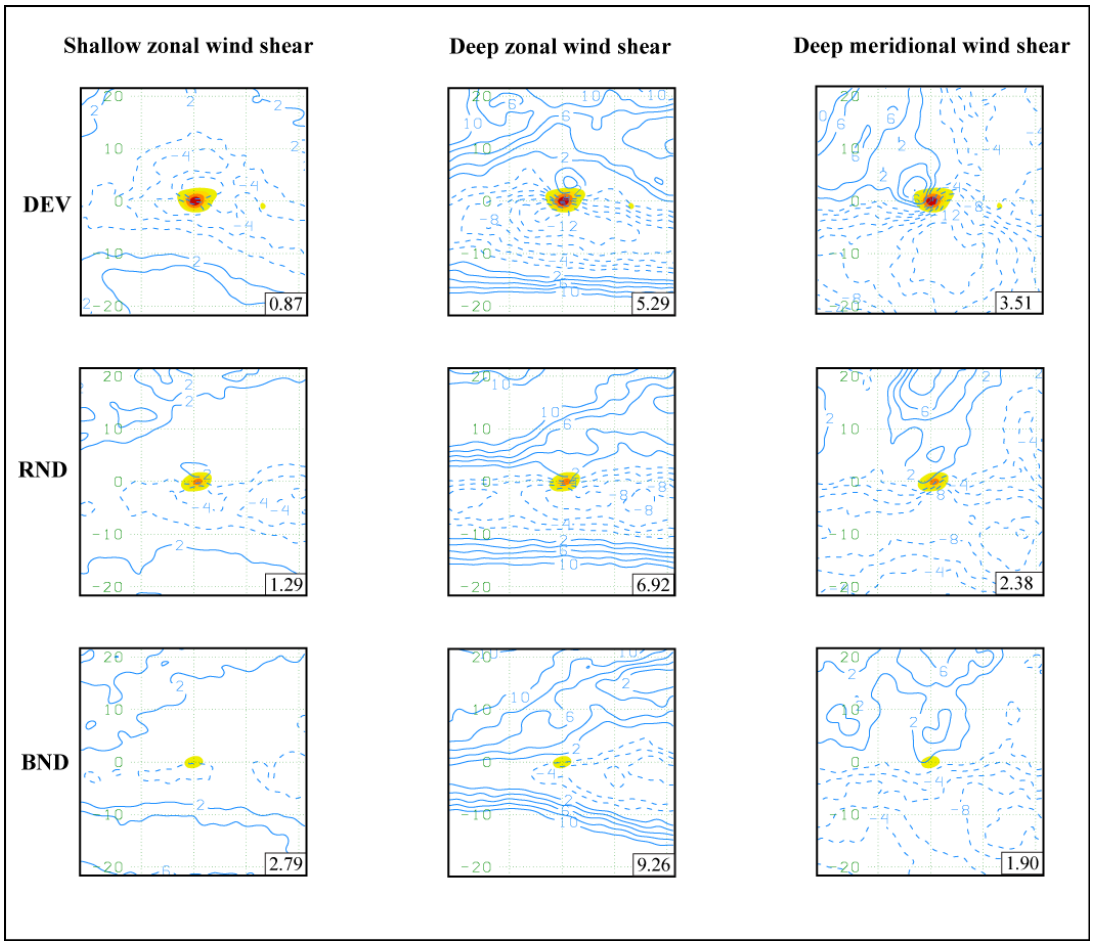

Figure 4.12 The 900-500hPa vertical shear of the zonal wind (left column), 900-200hPa vertical shear of the zonal wind (center column), and 900-200hPa vertical shear of the meridional wind (right column) contoured every 2 ms^{-1} for the DEV (top row), RND (center row), and BND (bottom row) composites. 900hPa relative vorticity is shaded every 2 $x10^{-5}s^{-1}$ for vortex location reference.

The zonal 900hPa-200hPa wind shear is stronger than the 900hPa-500hPa shear for all composites. The most noticeable difference here is the relatively strong shear to the south of the vortex in DEV composite compared to the non-developing composites. The larger difference in the region is likely due to the strong westerly low-level jet that is present in the DEV composite versus the weak low-level jet in the RND and absence of a low-level jet in the BND composite. To the north of the vortex there is positive shear in the DEV composite, but it is relatively weak compared to the RND and BND composites. This pattern is also apparent (though weaker) in the 900hPa-500hPa shear plots. Overall, the areal average for both the shallow and deep vertical shear of the zonal wind is less in the DEV composite and is the largest for the BND composite. This is not surprising because weaker vertical wind shear is favorable for genesis as it reduces the amount of heat (produced by the convection associated with the vortex) advected away by strong winds from the vortex core (Gray 1968).

The final shear considered is the 900hPa-200hPa equatorward vertical shear of the meridional wind, which can be considered as a surrogate of convection (Davis et al. 2008). For all composites, the pattern in the wind shear fields is such that negative (positive) values are to the south and east (north and northwest) of the vortex center. This indicates strong northerly (southerly) winds aloft to the south (north) of the vortex. The magnitude of the areal averaged vertical shear is largest for the DEV composite and weakest for the BND composite. In general, the magnitude of the shear is larger to the south of the vortex than to the north. To get a better understanding of why this might be a vertical north-to-south cross-section was taken through the center of each composite vortex.

Figure 4.13 shows the north to south cross-section through the center of the composite vortices. Plotted are the relative vorticity, winds normal to the cross-section, divergence, and

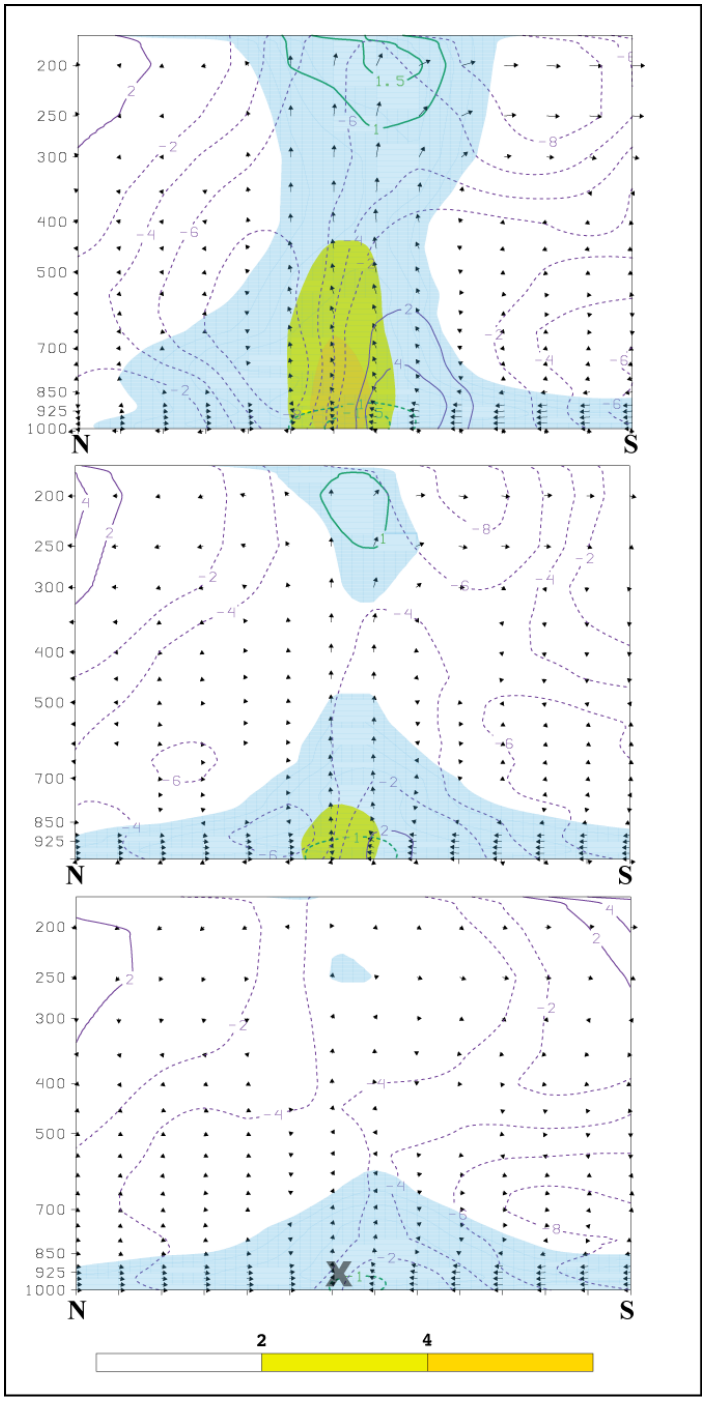

Figure 4.13 North-south cross-section of relative vorticity plotted every $2x10^{-5}s^{-1}$ (fill pattern), normal winds (purple) contoured every $2ms^{-1}$, divergence (green) contoured every $0.5x10^{-5}s^{-1}$, and winds in the plane of the cross-section represented by arrows. Dashed lines indicate negative values and the zero contours are omitted.

the winds in the plane of the cross-section. The DEV composite shows that the relative vorticity associated with the vortex extends to the 450hPa level. Near the surface, strong southerly winds flow to the center of the vortex from the south where they converge with comparatively weak winds from the north. The surface convergence feeds ascent through the moist (not shown) vortex core, and subsequent divergence (green contours) at upper-levels.

There is a marked difference between the divergent winds to the north and south of the vortex with stronger wind speeds occurring to the south. The stronger wind speed to the south of the vortex is likely a result from the fact that the absolute vorticity is weaker closer to the equator. An environment characterized by weak (or near zero values of) absolute vorticity provides conditions in which a northward moving air parcel can travel longer distances before turning westward (in the Northern Hemisphere) from the Coriolis force. To the north of the vortex, the winds traveling southward are in an environment characterized by higher absolute vorticity. These conditions cause the winds to be more affected by the Coriolis force that acts to turn the northerly winds eastward (in the Northern Hemisphere) thereby weakening the southward component of the wind to the north of the vortex in the lower-troposphere.

Winds normal to the cross-section are westerly along the southern edge and to the south of the vortex, though the separation of the two is not clear. These westerlies are likely a combination of the vortex itself and the low-level westerly jet. Further south, the winds are exclusively from the east at approximately $4ms^{-1}$ from the surface to the middle troposphere (where the winds reach a local minimum in the vertical), and then increase to about $8ms^{-1}$ between 300hPa and 150hPa.

Neither of the non-developing composites have a clear relative vorticity structure to them. The RND composite shows a small vortex near the surface, while the vorticity is well

below the threshold used to plot the relative vorticity near the vicinity of the vortex (denoted by the **X**) in the BND composite. Both composites are similar to the DEV composite in terms of the lower-tropospheric portion of in-up-out circulation pattern. Stronger southerly winds converge with weaker northerly winds near the surface at the center of the composite vortices. The surface convergence is slightly weaker in the RND composite than in the DEV composite and weaker still for the BND composite. Subsequently, the upward vertical motion is weak in both composites, though more so for the BND composite.

The differences between the developing and non-developing composites are even more pronounced at upper levels with respect to the divergence of the winds. The RND composite upper-level divergence is weaker than the DEV composite, while it appears to be non-existent the BND composite. The upper-level southerly winds are very weak to the north of the vortex, which is also the case for the DEV composite; however, the northerly winds to the south of the low-level vortex are also weak. This is not the case for the DEV composite that clearly shows strong upper-tropospheric northerly winds to the south of the vortex. The lack of appreciable upper-tropospheric winds to the south of the vortex in the non-developing composites could be due to higher absolute vorticity present at upper levels surrounding the RND and BND composite vortices compared to the DEV composite vortex (for the reasons described earlier). Indeed, the values of absolute vorticity that exist to the north and far south of the vortex in the RND composite are larger than those found in the DEV composite, and even larger values exist the north and south of the vortex in the BND composite (Figure 4.14).

One final note regarding the vertical cross-section composites concerns the environmental zonal winds surrounding the vortices. All composites show the vortices embedded in largely easterly flow, with the exception of westerly winds to the south of the

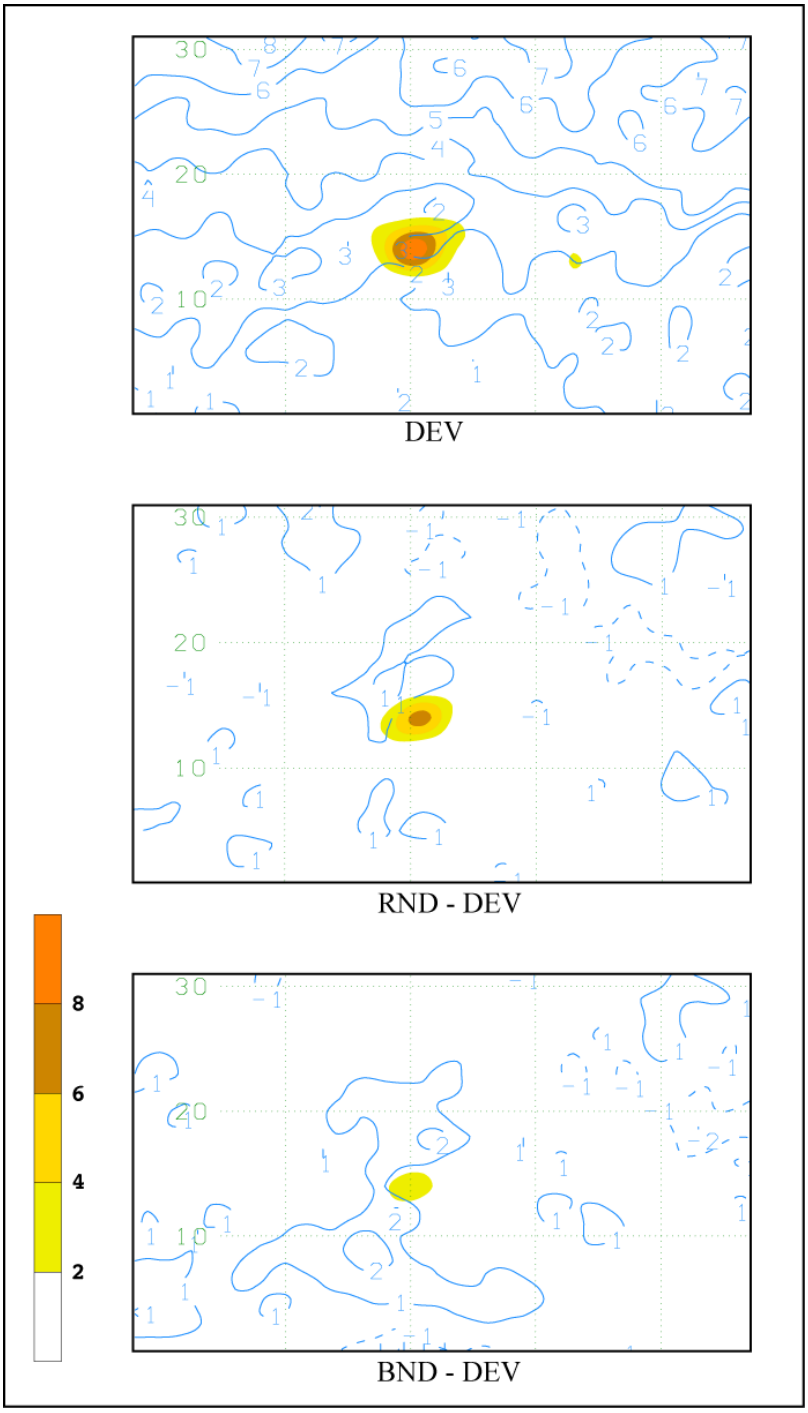

Figure 4.14 The 150hPa absolute vorticity superimposed over the 900hPa vortex relative vorticity. The latitude is the average latitude for the DEV vortices.

DEV composite vortex. While a large contribution of low-level westerly winds are part of the cyclonic rotation of the vortex itself, a partial contribution is likely from a low-level westerly jet (shown in Figure 4.8). The easterly winds further to the south of the vortex in the DEV composite are weaker than those in the non-developing composites with respective low and mid-tropospheric easterly wind speeds of $4ms^{-1}$, $6\ ms^{-1}$, and $8\ ms^{-1}$ for the DEV, RND, and BND composites.

d. Conceptual model of development

The composite analysis results show a number of clear differences between the environmental and vortex-centered features that surround developing and non-developing vortices in the east Pacific basin at all levels in the atmosphere. A graphical representation of the features conducive to genesis is presented in Figure 4.15.

1) LOWER-TROPOSPHERE

The most predominant difference between the DEV and non-developing composites is the presence of a low level jet that is strongest at 925hPa located to the south of the vortex and extending to the east. This jet is also present at 900hPa and has a slight signature at 850hPa. This jet is only present weakly in the RND composite and close to non-existent in the RND composite. The comparison between the total deformation and relative vorticity fields show that developing vortices are associated with deformation fields that are weaker than the relative vorticity of the vortex (as indicated by negative values of the Okubo-Weiss parameter).

In relation to the lower-tropospheric relative vorticity, the low-level jet may also aid in development for a vortex that has a favorable tilt to gain energy at the jet's expense through barotropic energy conversion (Lipps 1977; Ferreira and Schubert 1997; Hoover 2011). The

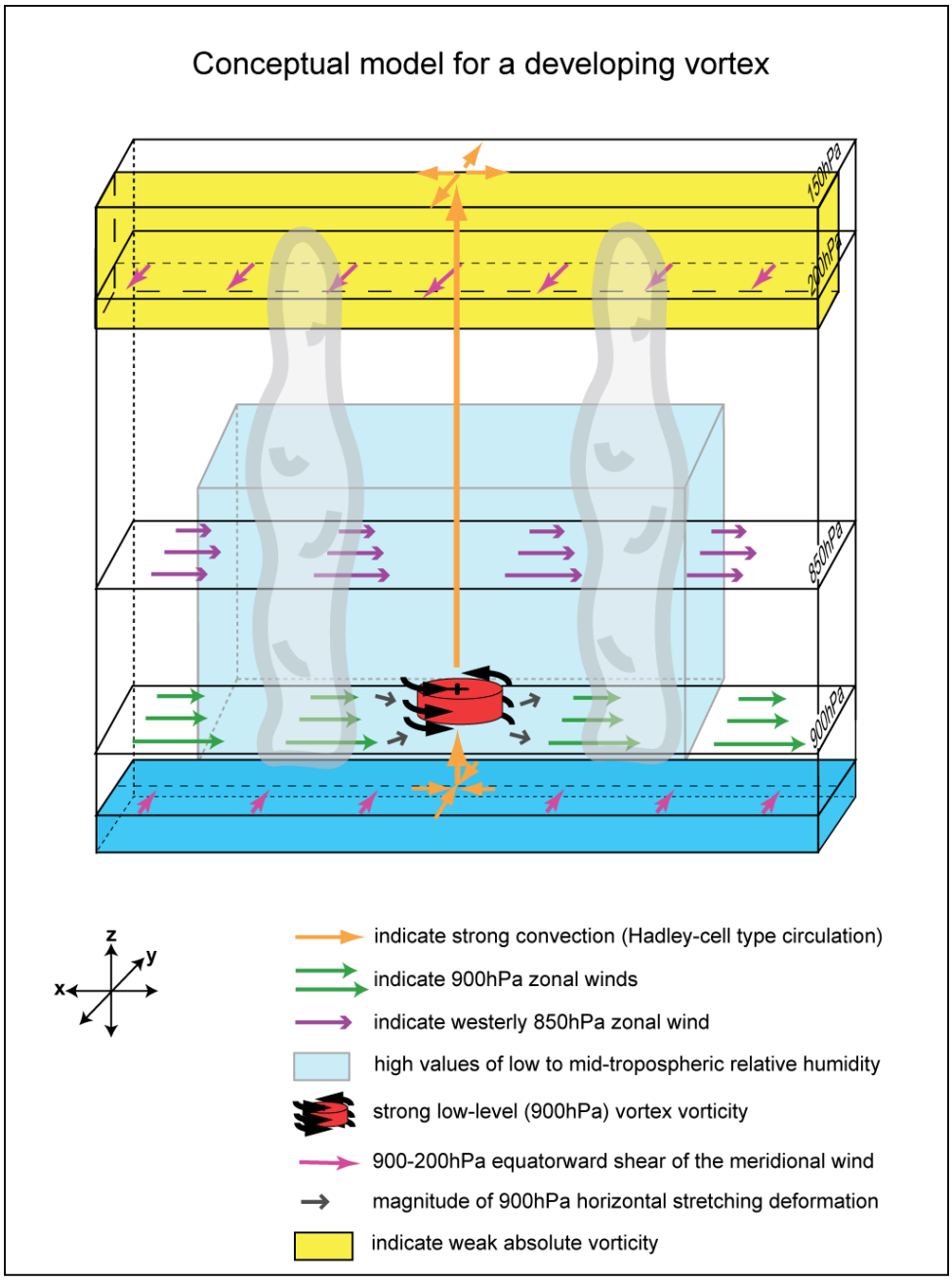

Figure 4.15 A conceptual model of characteristic environmental features favorable for east Pacific tropical cyclogenesis.

fact that the DEV composite vortex showed a slight tilt along the shear line of the jet suggests that a portion of the vortices may have developed via this mechanism. TS Enrique (described in Chapter 2) is one example in which a vortex intensified through barotropic growth. Whether through the process of barotropic growth or the cyclonic shear to the north of the jet, developing vortices tend to be located to the north of the low-level westerly jet (represented by green arrows).

The deformation field at 900hPa was also different between the DEV and non-developing composites, with stronger vortices having a horizontal stretching deformation field that consisted of a quadrupole of positive and negative deformation. Because of this pattern, the areal average of the stretching deformation was the weakest for the DEV composite and increased from the RND to the BND composites. Furthermore, it was shown that for vortex development, it is necessary that the relative vorticity of the vortex be stronger than its surrounding deformation field. The stretching deformation in relation to the vortex is represented in the conceptual model by gray arrows.

The final difference between the composite fields at low levels is the strength of the convergence in the core of the vortex with stronger convergence associated with the DEV composite. The strongest winds associated with the convergence are the northward winds to the south of the vortex center. The near-surface southerly winds were strong in all composites, but the strongest winds were associated with the DEV composite vortex. The upward vertical motion associated with the convergence was also strongest in the DEV composite.

2) THE MIDDLE-TROPOSPHERE

The 850hPa winds are similar to those near the surface with westerly winds to the south of the DEV composite vortex. While the winds associated with the vortex likely contribute to the westerly winds at this level, a portion may also be controlled by the jet, which weakens with height.

The relative humidity was high (above 75%) for all composites at 900hPa, but the mid-level relative humidity was higher over a larger area for the DEV cases than both the RND and BND composites. Having sufficient moisture is important in the development of TCs as it keeps the mid-troposphere buoyant thus reducing the occurrence of downdrafts that can weaken a nascent vortex located under it (Dunion and Velden 2004; Wu et al. 2006). Therefore, in addition to the lower-tropospheric features mentioned above, a relatively moist mid-troposphere as well as strong upward vertical motion and enhanced convection are also included in the conceptual model.

3) THE UPPER-TROPOSPHERE

The largest differences between the composites at upper levels are the upper-level divergence and associated northerly winds to the south of the vortex. The divergence at upper-levels is greater for the developing cases, specifically with regards to the northerly winds to the south of the vortex. Further investigation of the upper-levels showed that upper-tropospheric environment of the RND and BND composites was characterized by larger areas of stronger absolute vorticity to the south (and north in the case of the BND composite) of the low-level vortex position. Since this type of environment tends to inhibit the distance that the air can travel to the south, the DEV composite was characteristic of faster southerly winds (and faster winds in general) at this level. Evacuating mass quickly at upper-levels increases the convergence at low-levels thus deepening the storm (Riehl 1948); therefore, the upper-

level divergence is included in the conceptual model as is the tendency of having low absolute vorticity equatorward of the low-level vortex's location.

4.6 Time progression of dynamical features

In addition to examining the differences between nascent vortices that develop and those that do not, it is useful to investigate whether the strongest vortices 48 hours prior to genesis remain the strongest at their time of declaration and study how the physical features surrounding those vortices that become the strongest tropical storms progress through time. To that end, a separate data is compiled. Recall, this new composite consists of the strongest vortices at the time of their declaration as a tropical storm provided they existed at least 48 hours prior to that time. Composites of these 25 vortices are made at 24 and 48 hours prior to genesis. Before examining the physical progression of the genesis events, it is important to note the differences between the 48-hour time period of the new data set and the one used previously in this the study. Slightly less than half (12) of the DEV cases that are among the strongest tropical storms overlap with the strongest vortices 48 hours prior to declaration, leaving the majority of the cases in the data sets to be different. This suggests that the magnitude of the vorticity itself alone is not a good indicator of the future vortex strength; rather the environment (and how it changes) surrounding the vortex plays a key role in vortex development.

Many of the lower-tropospheric environmental features surrounding the two vortex composites are similar (Figure 4.16). Both composites show a westerly jet to their south extending to the east, (though the jet in the new data set is slightly weaker to the east and stronger to the west) and both composite vortices exhibit a slight tilt along the shear line of the jet. The lower-tropospheric relative humidity is higher for the stronger vortex composite,

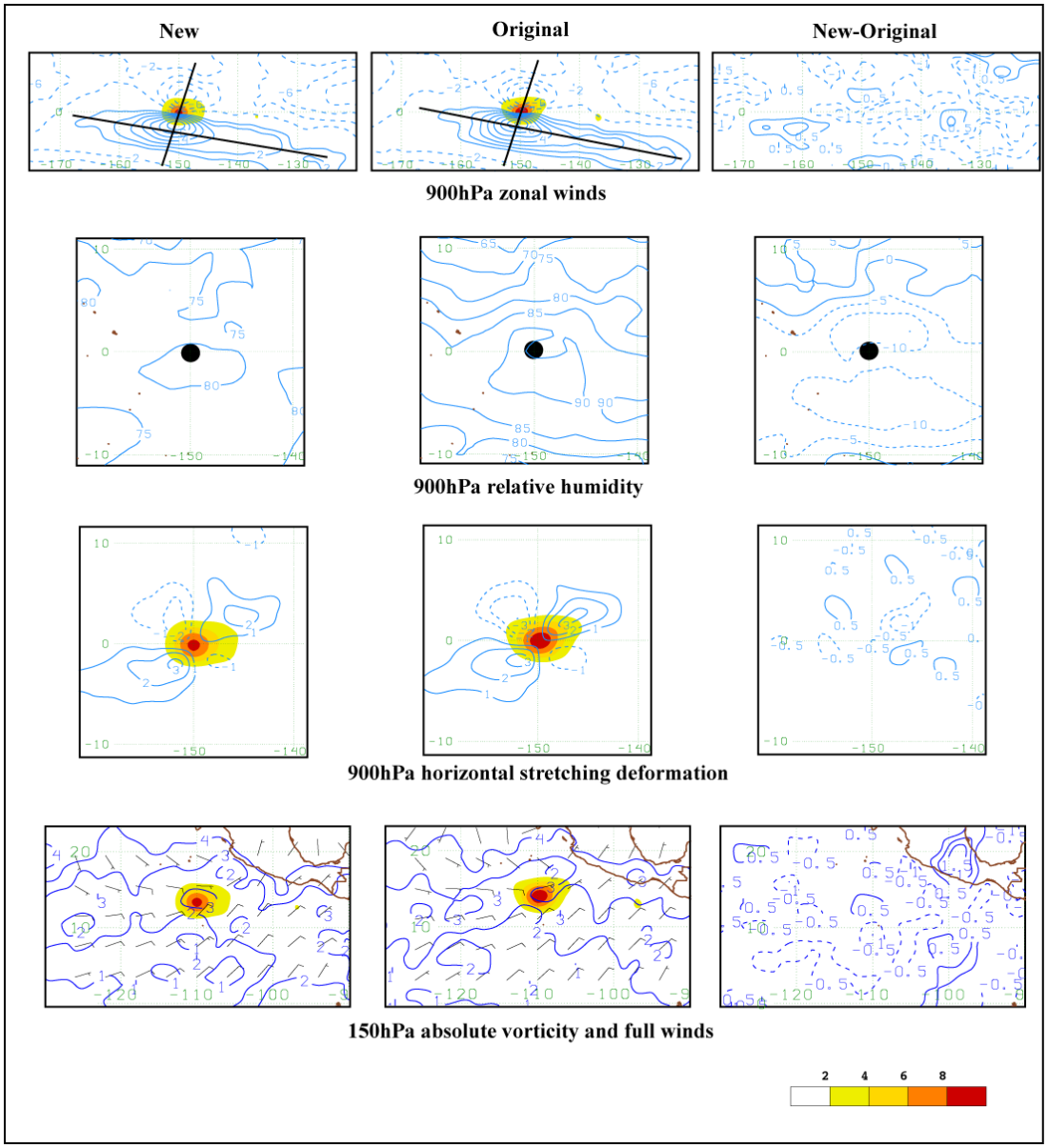

Figure 4.16 The 900hPa zonal winds (top row, contoured in ms^{-1}), relative humidity (second row, contoured every 5%), horizontal stretching deformation (third row, contoured in $10^{-5}s^{-1}$), and 150hPa winds (barbs) and absolute vorticity (bottom row, contoured in $10^{-5}s^{-1}$) for the new DEV composite (first column), original DEV composite (second column), and their difference (third column). Composites are for 48 hours prior to tropical storm declaration.

but slightly lower relative humidity exists in the upper-troposphere (upper-levels not shown). The 900hPa horizontal stretching deformation is also similar for both composites with a slightly stronger field surrounding the original vortex composite, which is likely contributed by the stronger vorticity of the vortex. At upper-levels (150hPa) the winds in this new composite are stronger to the south of the vortex than those in the original DEV composite, which is likely due to the lower absolute vorticity south of the vortex in the new composite.

An interesting difference that exists between the two composite sets at 48 hours prior to declaration is the upper-tropospheric divergence, which is slightly larger for the strong tropical storm composite, even though the relative vorticity is weaker in the new data set at 48 hours prior to declaration than that in the original developing composite (Figure 4.17). In addition the relative humidity at upper levels is above 70% for a much larger expanse. The difference between both the relative humidity and divergence in the upper-troposphere suggests that the upper-level environment may play a crucial role in the intensification of a low-level vortex.

The progression of several features associated with the 25 case composite are shown in Figure 4.18. At low levels the convergence remains strong for the duration of the vortex up to its time of declaration. The relative humidity ranges from 75% to 80% over the vortex center 48 hours prior to declaration and increases in the subsequent 24 hours to above 85% before decreasing slightly at the time of declaration. The stretching deformation keeps its quadrupole shape throughout the 48 hour time period as well. The magnitude of both the negative and positive deformation sectors increase in a similar fashion counteracting one another. In addition, the vorticity of the vortex is stronger than its surrounding deformation field as indicated by the Okubo-Weiss parameter. Upper-tropospheric divergence remains

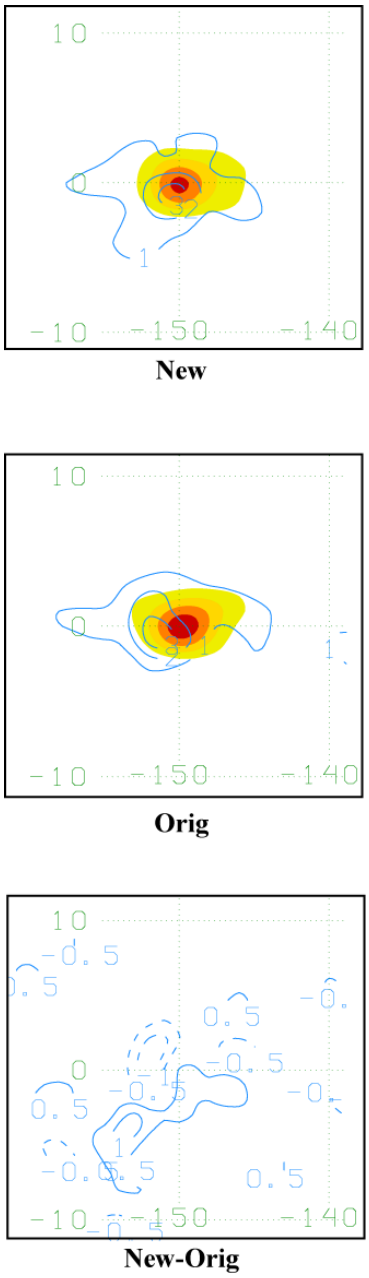

Figure 4.17 The 900hPa relative vorticity (shaded) and 150hPa divergence field for the strongest 25 tropical storms (that also exist 48 hours prior to declaration) 48 hours prior to declaration (top panel), the strongest 25 vortices at 48 hours prior to declaration (middle panel), and their difference (bottom panel).

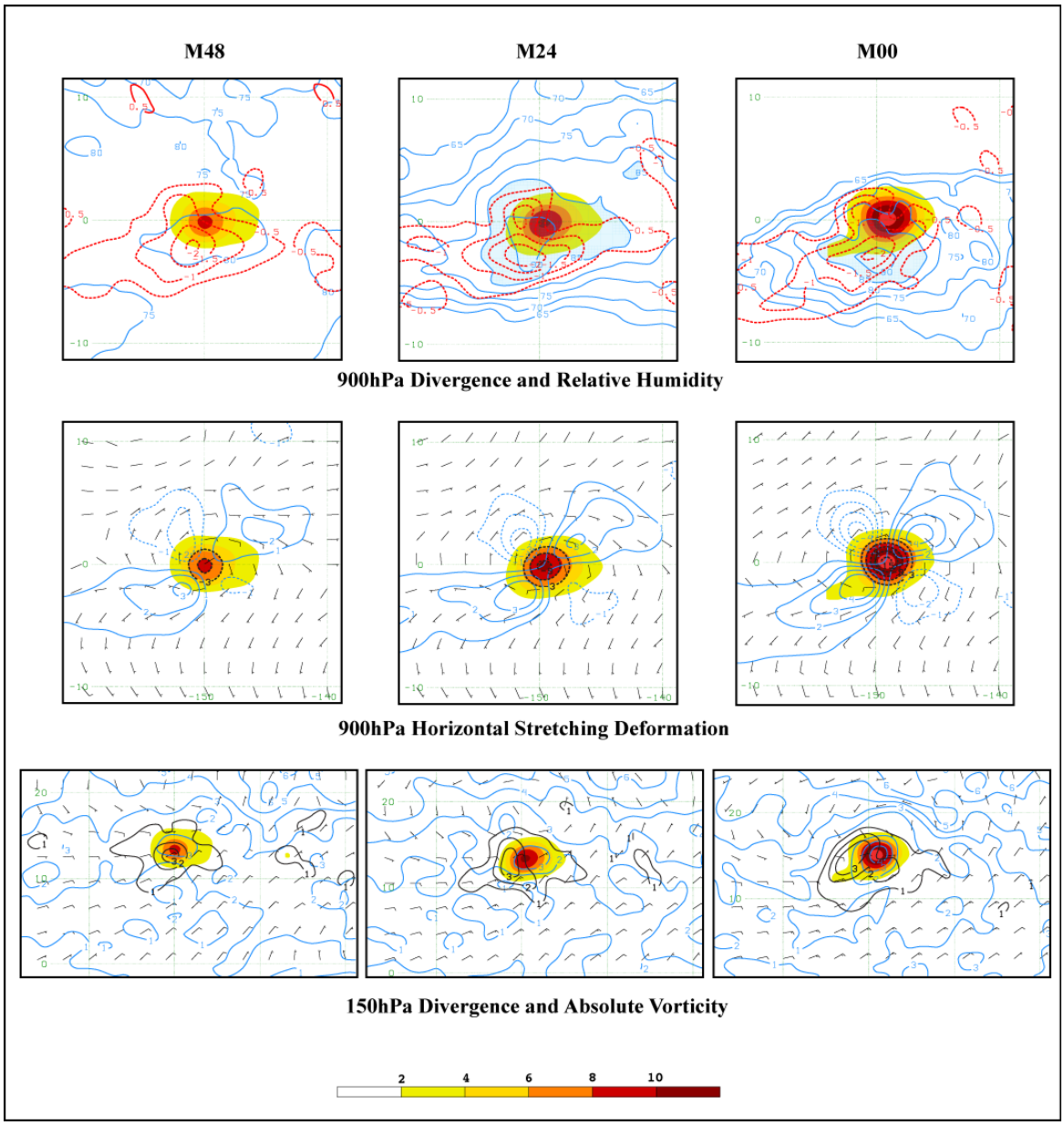

Figure 4.18 Time progression composites of the new composite set for 900hPa divergence and relative humidity (top row), 900hPa horizontal stretching deformation (middle row), and 150hPa divergence and absolute vorticity at 48 hours prior to declaration (left column), 24 hours prior to declaration (middle column), and the time of declaration (right column).

strong through the 48 hours leading up to declaration, while the absolute vorticity to the south of the vortex remains low allowing for strong northerly winds to the south of the vortex.

Another view of this process can be shown by a north-south cross section through the center of the vortex (Figure 4.19). Westerly winds extend well to the south (approximately 16˚-24˚) of the vortex and result from the vortex. The relative humidity remains above 75% over the vortex core for the whole 48 hours, though it decreases (especially to the north) through time. The upward vertical motion increases from 48 to 24 hours prior to genesis.

In the upper-troposphere, an anti-cyclonic rotation surrounding the vortex (indicated by the purple contours) increases in strength leading up to genesis as indicated by the winds normal to the cross-section, which increase (especially to the north) as the composite vortex reaches its declaration time. In addition, anomalously high divergence aloft at 48 hours prior to declaration gradually weakens over time to resemble the original composite. This could suggest the existence of a transient outflow feature or preconditioning due to earlier convection. However, more investigation is required to determine what mechanism is responsible.

Another compelling aspect of this cross-section is the upper-tropospheric divergence. The difference between the original DEV composite and current composite divergence seen at 150hPa is more pronounced in the cross-section. Both the magnitude and vertical depth of the divergence are greater for the strongest 25 tropical storms composite at 48 hours prior to genesis than that found in the original DEV composite. In addition, the relative humidity surrounding the vortex in (Figure 4.19) is higher over a larger area than that in Figure 4.13. As previously stated, these differences in the upper-troposphere suggest that upper-level features (rather than those in the lower-troposphere) may have a substantial impact on the

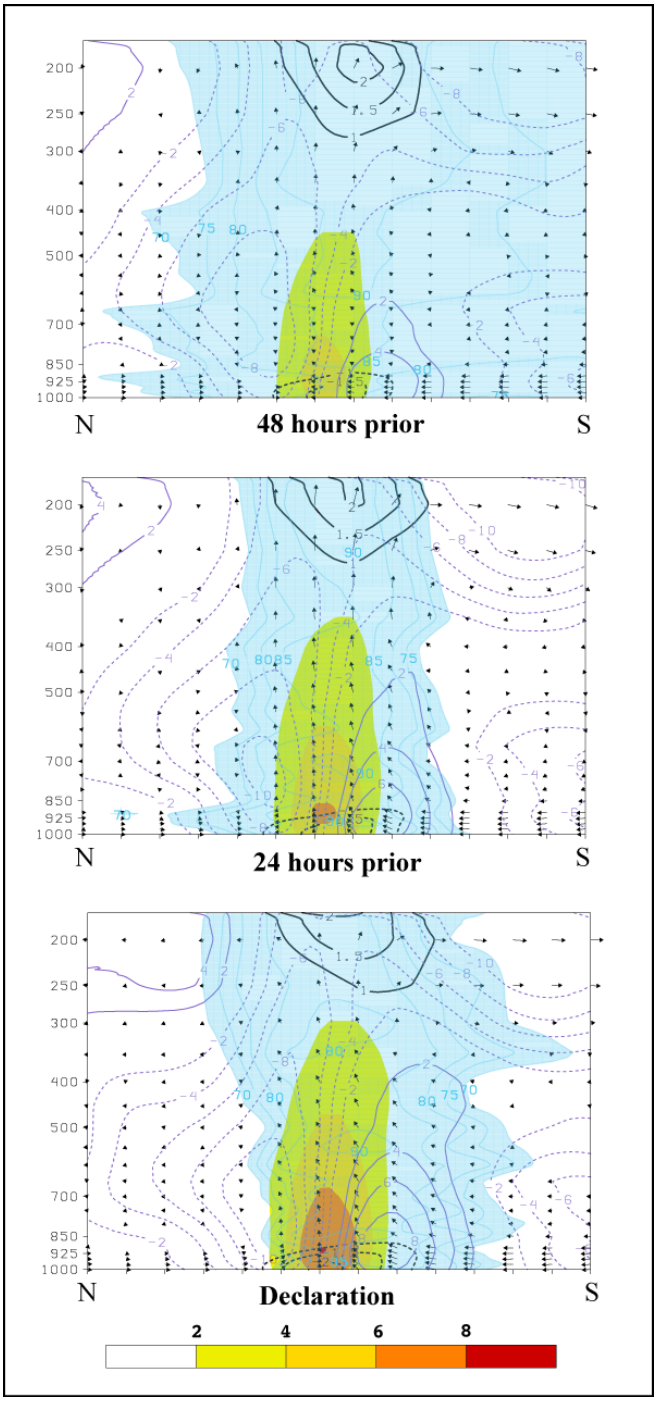

Figure 4.19 North-south cross-section of the vortex vorticity (shaded in warm colors every $2x10^{-5}s^{-1}$ starting with $2x10^{-5}s^{-1}$), relative humidity (contoured in blue every 5% with values higher than 70% shaded), divergence (black contours), and winds normal to the cross section (contoured in purple every $2ms^{-1}$). Winds in the plane of the cross section are represented by the black arrows.

development of vortices days before declaration that may supersede the strength of an extant vortex.

4.7 Conclusions and future work

Previous studies have shown that specific climatological features influence the development of TCs. These features include relative humidity, vertical wind shear, cyclonic shear, and distance from the equator. In this study, a composite analysis comparing the environmental differences between DEV and ND nascent vortices demonstrated that the former group of vortices were associated with higher relative humidity, stronger vortex vorticity, weaker vertical shear of the zonal wind, and locations more poleward than the non-developing group. These findings are consistent with observations cited by Gray (1968, 1979) and Palmén (1968).

In addition to the above climatological features, a number of kinematic variables also seem to influence the development of TC in the east Pacific. For example, Maloney and Hartmann (2000) found that more TCs occur during periods in which the low-level westerly winds persist across the basin. Findings by Davis et al. (2008) found similar results during the 2005 and 2006 seasons stating that vortices tend to form in the presence of a zonally confined low-level westerly jet. The composite results presented in this study show that a climatological low-level jet persists from June to October with its strength peaking in September. Most of the DEV cases tend on the cyclonic shear side of the jet. While the cyclonic shear provided by the jet provides the vorticity necessary for genesis, it may also aid in the development of a vortex through barotropic growth.

The horizontal stretching deformation field at 900hPa showed notable differences between the DEV and ND composite with stronger vortices exhibiting a quadrupole

deformation and weaker vortices having a more linear configuration of deformation. In addition, the magnitude of the vortex vorticity was stronger than the magnitude of the surrounding deformation field for the DEV composite, while the magnitude of the BND composite vortex vorticity was roughly the same as its deformation field. This indicates that the strength of the vortex vorticity compared to its surrounding deformation field is influential to vortex development. However, as mentioned in Chapters 2 and 3, TS Enrique was able to develop despite the strong deformation field surrounding it, which could mean that another "atypical" mode for tropical cyclogenesis could exist for the east Pacific.

Another feature that varies between the DEV, RND, and BND composites is the upper-level absolute vorticity, specifically to the south of the vortex center. The DEV composite showed weaker absolute vorticity to the south of the vortex, which may account for the higher northerly winds to the south of the vortex at upper-levels. The faster wind speed could increase the divergence aloft and lead to a stronger vertical circulation through the vortex core.

Looking at a time progression of some of the strongest tropical storms that occurred from 2001-2009, it is evident that these favorable features persist for a sufficient period of time. Other than a slight decrease in relative humidity, many of the features mentioned above and shown in Figure 4.15 remained consistent or became more favorable over the 48 hours leading up to vortex declaration. Perhaps an important aspect of developing strong tropical storms is the effect of the upper-tropospheric relative humidity and divergence days before their declaration; vortices that grew to become the strongest tropical storms tended to have enhanced upper-level divergence and relative humidity compared to vortices that were already strong 48 hours prior to declaration.

There are still many questions left unanswered from this study that leave interesting avenues for future work. For example, there are storms such as TS Enrique (2009) that develop despite being embedded in a hostile environment characterized by strong deformation. It would be interesting to identify a group of storms like Enrique to better understand the physical underpinnings that allow a vortex to overcome unfavorable conditions and develop into a tropical storm, perhaps identifying a mode of tropical storm genesis that is poorly captured by the conceptual model proposed in this study (i.e. secondary mode of genesis taking advantage of a barotropic growth mechanism along the low-level westerly jet). It would also be useful to consider another set of ND cases to compare with the DEV vortices in the basin, specifically with regards to the low-level jet. Since, the vortices used in this composite were located throughout the basin, it may be useful to limit the ND vortices to include only those that develop in the MDR.

Another approach that evolves naturally from this study as well as the work described in Chapters 2 and 3 considers those vortices that were incorrectly forecast by the statistical models. Comparing composites of the false alarms and missed developing cases could shed further light on the physical features that are responsible for the missed forecasts.

Chapter 5: Conclusion

5.1 Statistical Models

a. Linear Discriminant Analysis algorithm (LDA)

A number of statistical techniques have been used to attempt to predict TC formation in various tropical basins. One of the more common techniques is LDA. This technique has been used to identify developing tropical cloud clusters in the West Pacific (Perrone and Lowe 1986), predicting tropical depression formation in the Atlantic (Hennon and Hobgood 2003), and identifying developing from non-developing disturbances in the east Pacific basin (Kerns and Zipser 2009).

In Chapter 2, LDA was used to predict rather than identify (as done in the Kerns and Zipser 2009) eastern Pacific tropical cyclogenesis using large-scale dynamical, thermodynamical, and kinematic fields as predictors. Some of the fields (relative humidity, relative vorticity, and the Coriolis parameter) were chosen because they demonstrated skill with this technique in the earlier studies listed above (and subsequently re-confirmed in this present study), while other features (e.g. MJO indices and the Okubo-Weiss parameter) were additions to the previous studies. Using vortices that developed during the peak seasons from 2001 to 2009, various sets of predictors were tested with the algorithm. The most skillful set included the latitude of the vortex center, averages of relative vorticity, horizontal stretching and total deformation, zonal wind; various vertical shears averaged about the center of the vortex (900hPa-500hPa zonal and 900hPa-200hPa equatorward meridional), and the Okubo-Weiss parameter.

Interestingly, a meridional barotropic conversion term, considered as a parameter in the algorithm, did not improve on the predictive skill of the model and was therefore excluded

as a parameter. However, TS Enrique may have been classified correctly if such a parameter was included in the algorithm, since it seems to have developed from mechanisms attributable to barotropic growth. One potential avenue for future research could be to develop a "secondary" LDA algorithm that would be calibrated to forecast TC genesis from a barotropic growth mechanism. The purpose of this additional algorithm would be to consider those vortices that are classified as non-developing from the initial algorithm and re-classifying them according to features associated with barotropic growth (e.g. barotropic conversion terms like those discussed in Maloney and Hartmann 2001). Perhaps such a secondary algorithm could decrease the number of misclassified cases produced by the original algorithm.

b. Bayes probabilistic model

As with LDA, the use of a *naïve* Bayes probabilistic model has been successfully employed in the atmospheric sciences. As with LDA, this model is used to classify two groups of events, but it does so by finding a *conditional probability* that an event belongs to one group rather than another based on a combination of parameters. While this model has not been as widely used as LDA in the atmospheric sciences, it has been found to improve seasonal forecast skill of the El Nino-3.4 index when compared to empirical persistence forecasts and bias-corrected coupled model forecasts (Coelho et al. 2003). In relation to tropical meteorology, it was shown to skillfully predict secondary eye-wall formation in eastern Pacific hurricanes (Kossin et al. 2009). However, this model has not been used to predict tropical cyclogenesis in the east Pacific basin. As a consequence, in addition to the deterministic forecasts provided by the LDA algorithm, a Bayesian probabilistic model was employed in Chapter 3.

Like the LDA algorithm, this Bayesian probabilistic model showed significant forecast skill for predicting tropical cyclogenesis in the east Pacific basin for forecast lead times as long as 48 hours. However, the probabilistic nature of this model provides more forecast information than does a binary model like the LDA algorithm. Rather than producing a strict "yes" or "no" forecast, this model is able to generate the likelihood that a given vortex may develop. In addition, the verification tools, such as the attributes diagram, used to assess the model provide more detailed information regarding the calibration of the model.

Since the Bayes model does not rely on the difference between the parameter means of the two groups to make a classification (as does the LDA algorithm), some additional features may increase the skill of the model. One example resulting from the study in Chapter 3 is the longitude of the vortex center. Since the means between the DEV and ND groups are similar, the addition of this parameter in the LDA algorithm did not add forecast skill. However, since the distribution of the longitude was appreciably different between the DEV and ND groups, Bayes model was able to gain forecast skill from this parameter. One line of future research could explore the use of other features that have similar means, but different distributions to see if they could add forecast skill to this model.

5.2 Summary and discussion of results

Tropical cyclones (TCs) are complex weather phenomena at every stage of their lifecycle, and their genesis stage is no exception. This body of work has attempted to identify – using a variety of statistical methods – those features that best distinguish between eastern Pacific vortices that develop into tropical storms and those that do not. Building from the findings of the statistical models presented in Chapters 2 and 3, an investigation of the underlying kinematic dynamics of developing and non-developing vortices was performed.

Many of the findings of this study are consistent with previous work, while some new aspects regarding genesis in the east Pacific have been revealed. This chapter will briefly review the findings of this work in context with previous studies and outline potential avenues for future research.

a. Climatological features

A number of features were found to add forecast skill to both the LDA algorithm (Chapter 2) and Bayes probabilistic model (Chapter 3). These features included the 900hPa relative vorticity of the vortex, 900hPa-500hPa relative humidity surrounding the vortex, and the latitude of the vortex center. All of these features have been found to impact tropical cyclogenesis across the tropics with strong vortex vorticity at latitudes poleward of 5˚ and high values of relative humidity favoring genesis (Gray 1968, 1979; Davis et al. 2008).

The findings of this study are consistent with the observations made by Gray. All vortices, whether developing or non-developing, formed well north of 5˚N and developing vortices were typically located further to the north than the non-developing cases. Developing cases were also associated with larger values of relative humidity and relative vorticity than their non-developing counterparts.

b. Low-level kinematic interactions

Some kinematic features that have been found to distinguish between developing and non-developing vortices in the east Pacific basin include low-level (900hPa) horizontal stretching deformation associated with a lower-tropospheric westerly wind maximum. For example, more vortices (both developing and overall) tend to form when the east Pacific basin is in a westerly wind regime (Maloney and Hartmann 2001; Davis et al. 2008). The findings in the current body of work also show that the developing vortices are typically embedded in

either westerly or weak easterly winds, while the non-developing cases are largely surrounded by easterly winds. Further inspection of these westerly winds revealed that they are part of a low-level jet, typically centered near 10˚N.

The results from the statistical models also showed that the magnitude of the 900hPa-200hPa equatorward vertical shear of the meridional wind was stronger for the developing vortices confirming the findings of Davis et al. (2008). The vertical shear of the meridional wind to the south of the vortex center can be thought of as a surrogate for convection over the vortex. These results indicate that developing vortices are associated with stronger convection than the non-developing vortices.

The final kinematic feature that Davis et al. (2008) found to differ significantly between the two groups of vortices is the 900hPa horizontal stretching deformation. Just as found in the Davis study, on average, the developing cases were associated with weaker deformation fields. However, there is evidence (as in the case of TS Enrique 2009 – STRDPROG Chapter 2) that vortices can still develop despite strong stretching deformation over the vortex. Therefore, it appears that it may be more important that the vortex vorticity be stronger than the deformation over the vortex as demonstrated by the Okubo-Weiss parameter (Figure 2.9).

c. Madden-Julian Oscillation

The MJO has been shown to modulate the environment of the east Pacific such that active phases create favorable conditions for genesis while suppressed phases tend to inhibit development (Maloney and Hartmann 2000). Specifically, during active phases, the east Pacific basin is characterized by westerly winds at 850hPa and enhanced convection, both of which are associated with increased TC activity over the basin (Maloney and Hartmann 2000;

Hartmann and Maloney 2001; Davis et al. 2008). The suppressed phases of the MJO leave easterly winds at low levels and decreased convection over the basin and are associated with less TC activity. For these reasons the MJO indices, as defined by Wheeler and Hendon (2004), were considered as parameters in both the LDA algorithm and the Bayes probabilistic model with the hypothesis that they would add forecast skill to the models. However, neither of the models resulted in increased skill with the addition of the MJO parameters.

It is thought that the cyclical nature of the MJO may be the reason behind the lack of skill in the LDA algorithm. Of the eight MJO phases, three (phases 1, 2, and 8) are favorable in the east Pacific basin while the remaining phases are unfavorable for TC development. Therefore, the group means of the MJO phase were very similar; in fact, the difference between their group means did not pass a two sample *t*-test at the 95% confidence level at all forecast times. Since the LDA algorithm uses the difference between the means of the two groups to classify unknown cases, parameters whose means are similar between the two groups could potentially decrease classification skill of the algorithm. Therefore, the similarity between the two group means for the MJO phase may explain why this parameter did not add skill to the algorithm. However, this reasoning does not explain the lack of skill provided by the MJO phase to the Bayes probabilistic model.

The Bayes model does not rely as heavily on the parameter means between the two groups. Therefore, a similarity between the two group means does not automatically result in no skill being added to the model predictions provided the distribution of the parameter values is different for each group. Indeed, the majority of the DEV cases occurred during phases 1, 2, or 8, while the ND cases occurred during all phases (Figure 3.9). However, even with the differences in the phase distributions, the parameter did not add skill to Bayes model. This

lack of additional skill may have resulted from the fact that the zonal wind and OLR are the two features used to define the phase of the MJO. Since the zonal wind is already used as a parameter in the model, the addition of the MJO phase as a parameter is redundant and does not add skill.

5.3 Discussion and future work

The present study has used a number of statistical techniques to better predict and understand eastern Pacific tropical cyclogenesis. Both the LDA and Bayes probabilistic models demonstrated significant skill in predicting genesis out to 48 hours. A composite analysis of the physical features that influence genesis revealed a number of differences between developing and non-developing nascent vortices. In addition, an initial analysis of TS Enrique suggested that an "atypical" mode of genesis might exist in the basin. Future research could include performing an extensive case study of TS Enrique and similar storms to determine how this type of formation occurs. Another line of future work would focus on the origin of the developing and non-developing vortices in the basin since the east Pacific has a myriad of internal (ITCZ breakdown, MCSs) and external (EWs, KWs, and the MJO) features that could generate or intensify nascent vortices (Shreck et al. 2012). Below, the results of the dissertation are briefly placed in the context of existing tropical cyclogenesis theories and possible future work is discussed.

a. Lower-tropospheric vortex initiation and tropical cyclogenesis

One aspect of genesis that was not focused on in the current body of work is the origin of the lower-tropospheric vortices in the basin. While the body of work described in the previous chapters defines the tropical cyclogenesis as the transition of a low-level vortex into a tropical storm, other studies (e.g., Montgomery et al. 2006) define the process to occur at an

earlier stage, specifically when a mid-level vortex transforms into a low-level vortex. Indeed, a branch of research has been devoted to identifying those physical mechanisms that lead to the formation of a lower-tropospheric vortex. The most prominent theory concerns the development of a surface vortex that forms as a result from the intensification and penetration of a vortex that forms in the mid-troposphere (Davidson et al. 1990, Bister and Emanuel 1997; Raymond et al. 1998; Davis and Bosart 2001, 2002; Nolan, 2007). The downward development of the midlevel vortex may result from two distinct processes.

One process is based on the concept of vortical hot towers (VHT), which are individual strong updrafts that form in a cumulus field that create intense positive and negative [potential] vorticity anomalies (Hendricks et al. 2004). The merger of a number of these VHTs into one organized more intense cyclone occurs from the tilting of the ambient vorticity field (Montgomery et al. 2006). The resulting stronger vortex can develop a greater penetration depth manifesting itself as a lower-troposphere vortex that could become a TC (Hendricks et al. 2006).

Using the thermal wind balance perspective, Bister and Emanuel (1997) suggest that a mid-level vortex can reach the lower-troposphere via precipitation and evaporative cooling. Using a case study of Hurricane Guillermo, which formed during the Tropical Experiment in Mexico (TEXMEX)[7] field campaign, a conceptual model for the transformation of a mid-tropospheric vortex to one near the surface was outlined as follows: diabatic heating in the anvil of a mesoscale convective system (MCS) and cooling due to evaporation below the anvil produce a mid-tropospheric vortex. From the thermal wind balance, this vortex is associated with a cold (warm) core in the lower (upper) troposphere. The cold core under the

[7] A brief description of the TEXMEX campaign can be found online at www.aoml.noaa.gov/hrd/hfp96/cyclo/.

vortex becomes increasingly cool and moist over time due to the evaporation of precipitation, while the air in the lowest portion of the troposphere becomes relatively warm and dry resulting in the formation of downdrafts that extend the cold-core vortex downward. If the vortex core penetrates the boundary layer, it can decrease the lower tropospheric stability by decreasing the equivalent potential temperature needed to trigger convection in the lower-troposphere. Simultaneously, the winds associated with the vortex act to increase the surface heat flux at the ocean-atmosphere interface and thereby increase the near surface equivalent potential temperature. These two processes decrease the vertical gradient of equivalent potential temperature and create an environment for the re-development of convection (and concomitant latent heat release) in the lower-troposphere, which would act to increase the surface vorticity through diabatic redistribution of potential vorticity.

Consistent with either of the competing (and in some ways complimentary) views of tropical cyclogenesis, the composite environment within which the surface vorticity intensifies is characterized by a relatively moist middle and lower troposphere, cyclonic (shear) vorticity, and evidence of recent convection. That the developing vortices had to have a minimum threshold for development (or a favorable Okubo-Weiss parameter value) is consistent with the results of Rotunno and Emanuel (1987) who indicate the finite-amplitude nature of tropical cyclogenesis.

The existence of the lower-tropospheric jet appears to be an important aspect for intensification of *existing* vortices in the east Pacific Basin. The jet provides an environment of vorticity spin-up not only through the background cyclonic shear vorticity north of the jet axis, but also indirectly from that shear vorticity as it creates an environment favorable for Ekman boundary layer pumping that would help moisten the lower-troposphere. Furthermore,

as discussed below, the very existence of shear provides a potential source for barotropic growth of favorably tilted vortices north and south of the jet.

Whether through VHT mergers or the downward extension of a vortex from downdrafts, recent studies have indicated that the formation of a surface vortex stems from an initial vortex in the mid-troposphere. Therefore, a natural extension of the work presented in the previous chapters would be to investigate the origins or initiation of the surface vortices tracked in the prior chapters. There are a number of potential source regions and/or initiation mechanisms present in the basin that could spawn vortices in the middle troposphere that have the potential to develop in a favorable environment. Vortices can spawn from breakdowns in the ITCZ (Lipps 1970; Ferreira and Schubert 1997) or form from an MCS over the basin (Bister and Emanuel 1997). Still others may propagate into the basin from easterly waves (Hartmann and Maloney 2001; Fu et al. 2007, Schreck et al. 2012). In fact, a number of vortices may result from or be influenced by easterly waves (EW), Kelvin waves (KW), or active phases of the MJO (Figures 5.1-5.3[8]).

For example, during the 2004-2006 peak seasons, 31 tropical storms were declared, and the majority of these storms were likely influenced by one or more of these three features. A subjective analysis of the Hovmöller diagrams for the three seasons can be used to determine whether a vortex is associated with one or more of these features. For example, if a vortex is located in a region where an enhanced phase of the MJO (red dashed lines) or an EW (black solid line) exists, it may have originated from the EW and intensified from the

[8] Hovmöller diagrams are provided by Kyle Griffin (University of Wisconsin – Madison) The waves represented in the Hovmöller diagrams are OLR filtered. Kelvin waves are determined using a 2.5-20 day window and wave numbers of 1 to 14 (according to Wheeler and Kiladis (1999)). The MJO contours are constructed using a 30-96 day filter for wave numbers 1-9 (according to Kiladas et al. (2005)). Finally, the easterly waves are filtered over 2-10 days using wave numbers -20 to -6 (Roundy and Frank (2004)).

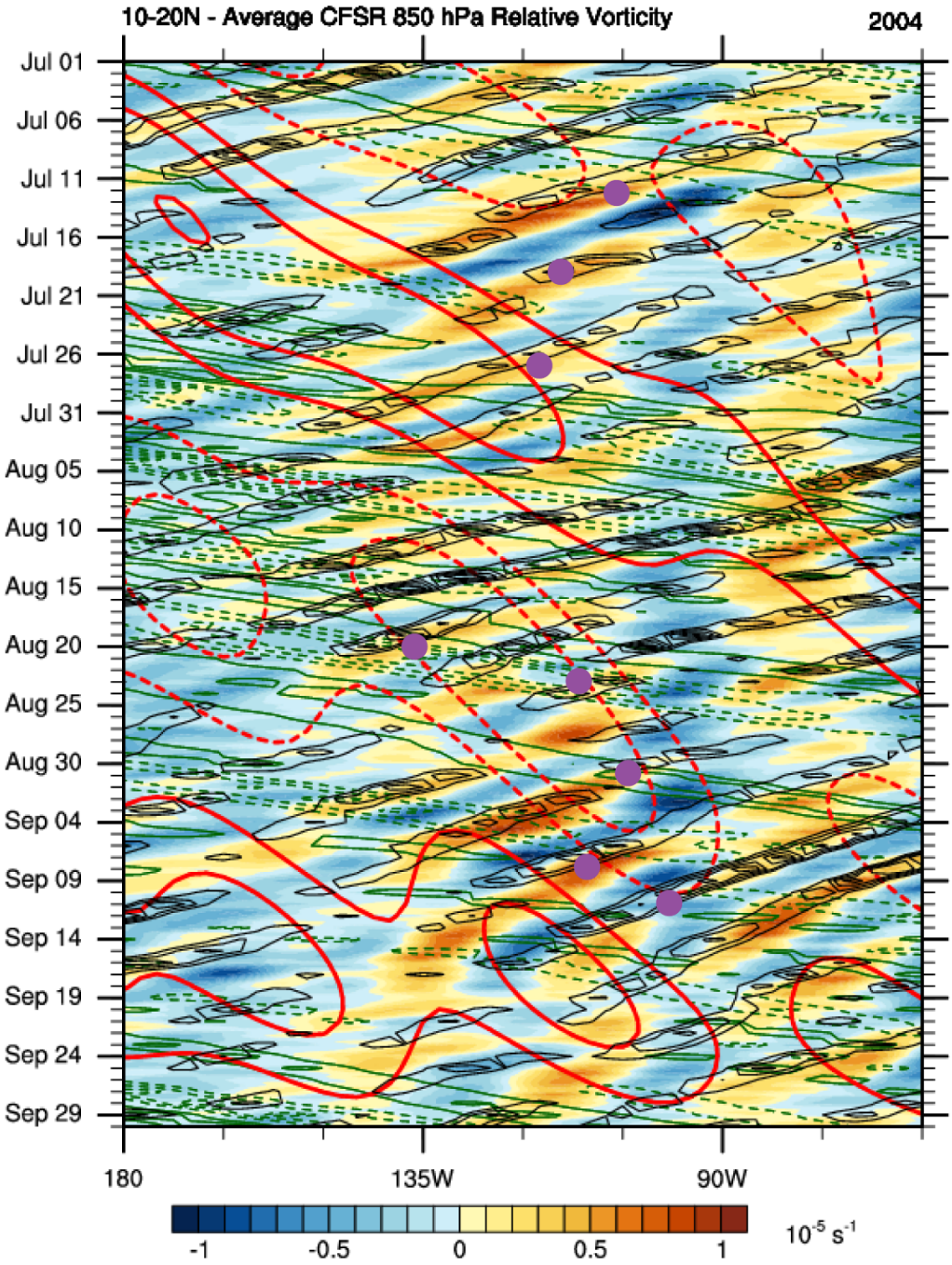

Figure 5.1 Hovmöller diagram (for the 2004 peak hurricane season) of the 850hPa relative vorticity (fill pattern; warm (cool) colors indicate cyclonic (anti-cyclonic) vorticity averaged over 10˚N-20˚N. Contoured are the suppressed (solid contours) and enhanced (dashed contours) MJO phases (red contours) and Kelvin waves (green contours). Also plotted are easterly waves (represented by black solid contours) and the latitudinal locations of vortices at their time of declaration as tropical storms (purple circles).

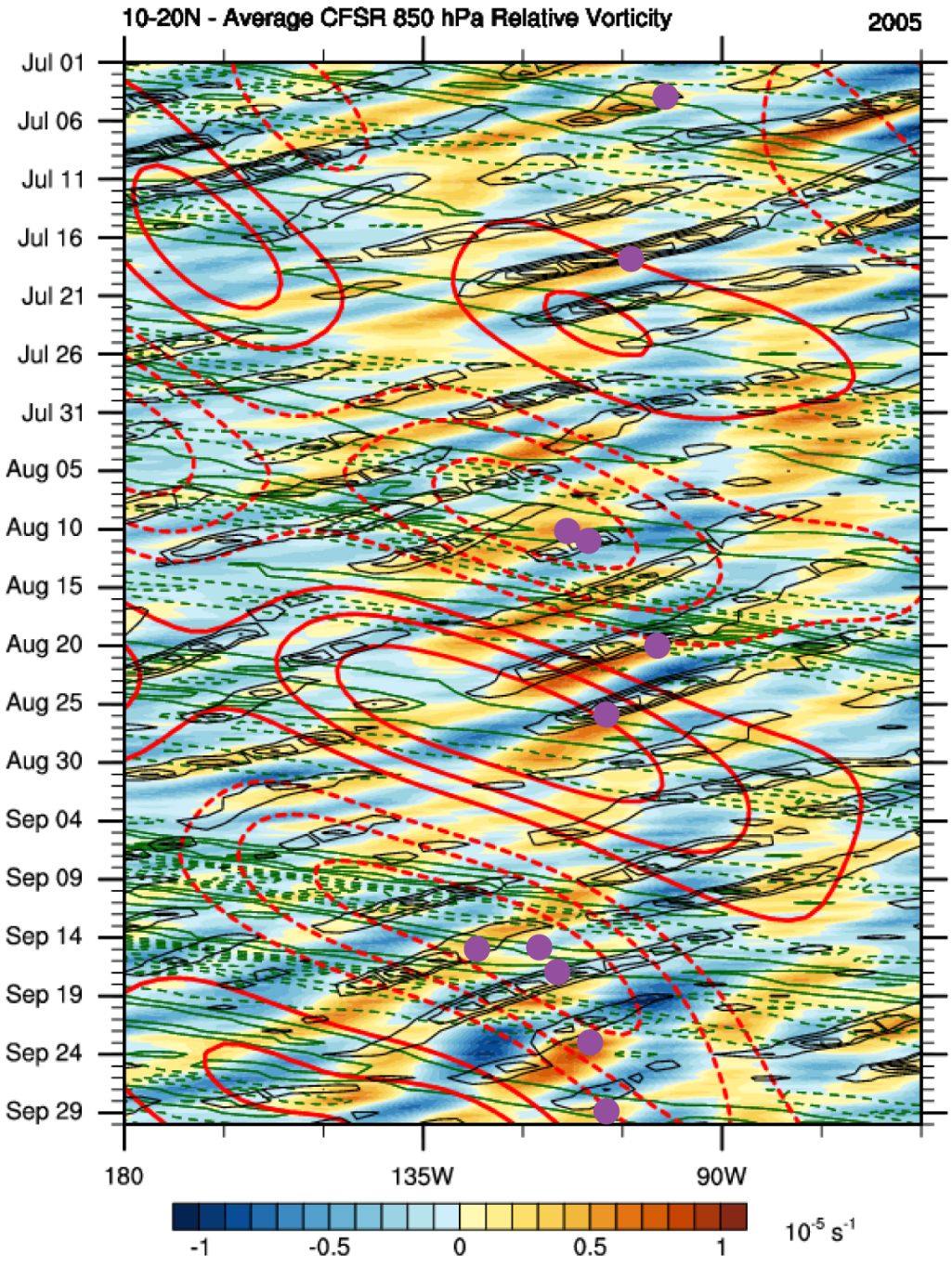

Figure 5.2 Same as a Figure 5.1, but for the 2005 peak hurricane season.

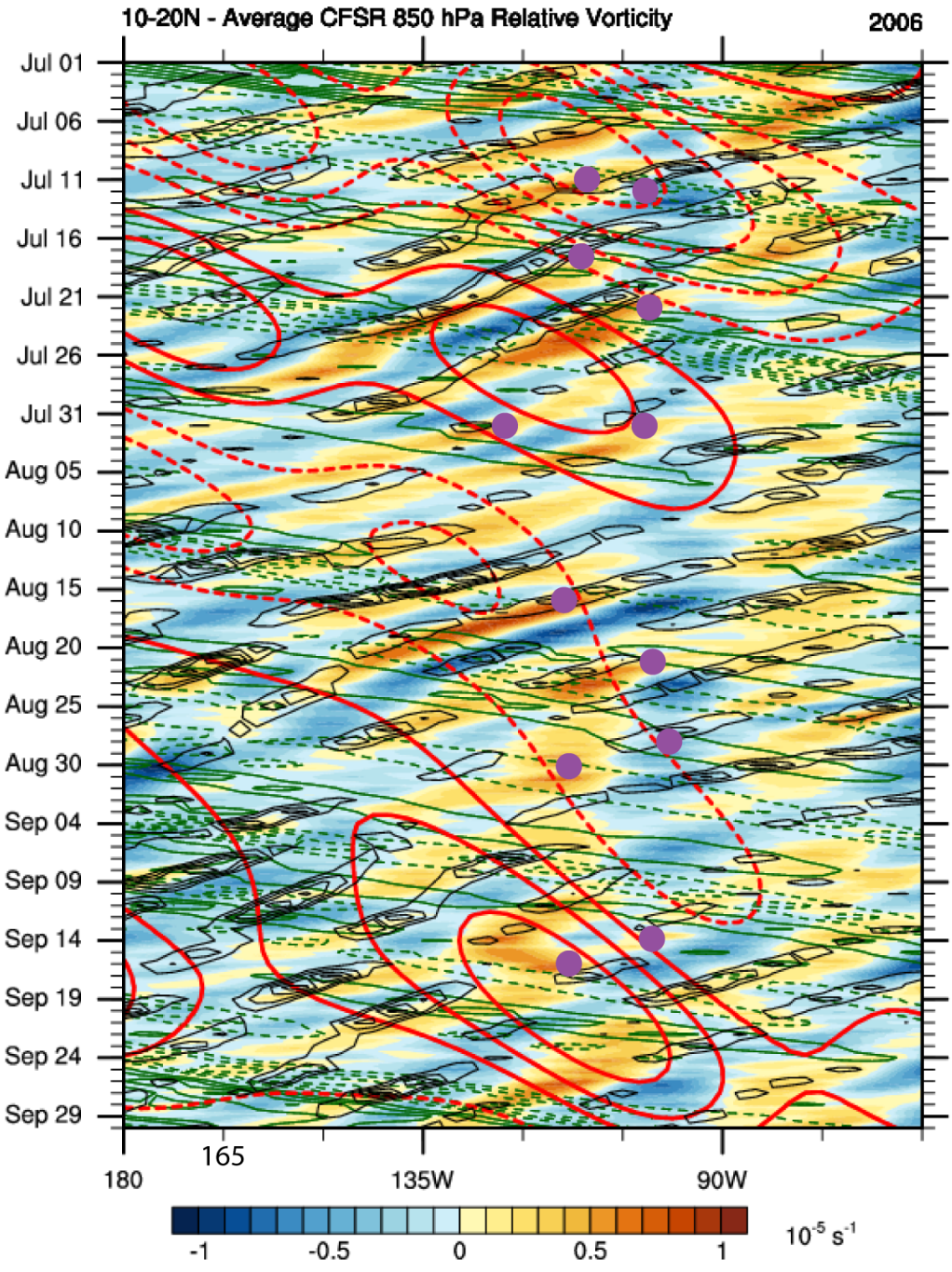

Figure 5.3 Same as a Figure 5.1, but for the 2006 peak hurricane season.

environment modulated by the MJO. Such an analysis of the 2004-2006 Hovmöller diagrams shows that favorable MJO phases, Kelvin waves, and easterly waves likely influenced 21, 11, and 10 of the storms, respectively (Table 5.1). The reason that the total number of storms influenced by these phenomena is larger than the total number of storms present during the three seasons is due to the fact that many of the storms are influenced by a combination of the features. While this analysis is not rigorous, it does suggest that further investigation of these features would be useful for future research.

b. "Atypical" formation

A comparison between TS Gil and TS Enrique revealed that a potential subset of TC formation, which results from a vortex developing in a strong shear flow just north of the jet, may lead to poor forecasts by the LDA algorithm. In the case of TS Enrique, the vortex is misclassified by the strong horizontal shearing deformation surrounding the vortex. However, there is evidence the vortex was able to grow at the expense of the jet through barotropic growth.

The fact that both statistical models incorrectly predicted Enrique as a non-developing case is interesting. The parameters selected for use in the models are selected to maximize the number of hits, yet the predictor values for Enrique mostly seemed to indicate that the vortex would not develop. While the vortex associated with pre-Enrique was in a strongly sheared environment, it did intensify into a tropical storm. One feature that seemed important to TS Enrique's development was the barotropic energy conversion term, $-\frac{\partial \overline{U}}{\partial y}\overline{u'v'}$, which (as stated earlier) was not found to add skill to either model and was subsequently discarded. This could indicate that a subset of TCs in the east Pacific basin may intensify barotropically

	MJO	EW	KW	other	total
2004	7	5	4	0	8
2005	7	3	3	0	11
2006	7	3	3	5	12
total	21	11	10	5	

Table 5.1 The number of east Pacific developing vortices that appear to be influence by the Madden-Julian Oscillation (MJO), easterly waves (EW), or Kelvin waves (KW) for the 2004, 2005, and 2006 peak seasons. Those vortices, which do not appear to be associated with any of these phenomena, are placed in the "other" category.

in an environment resembling that shown in Figure 4.15 by extracting energy from the mean shear. TS Enrique may prove to be an interesting case study for this type of growth mechanism. An in-depth case study of TS Enrique's development could guide future work for this "atypical" development mechanism in the east Pacific basin.

c. Additional predictors for statistical models

A final avenue of future research would stem from the statistical models used in Chapters 2 and 3, specifically the inclusion of additional parameters that may increase the skill of the model forecasts. For example, since the SST was not used as a predictor in either statistical model, it may be useful to investigate if the inclusion of the SSTs surrounding the low-level vortex would increase the model forecast skill. Another interesting predictor to consider, from an energetics perspective, would be Emanuel's maximum potential intensity (MPI) (Emanuel 1991), which has been shown to be skillful for the northwest Pacific and Atlantic basins (Tonkin et al. 2000). It would be interesting to adapt Emanuel's MPI as a predictor in the statistical models to see if it could add skill to the forecasts produced by the models.

The findings of the statistical models in Chapters 2 and 3 of this dissertation produced a solid list of predictors that demonstrate skill in forecasting tropical storm formation in the east Pacific basin. However, this list of predictors is just the beginning. Using additional predictors (such as the two examples listed above) in the models could lead to improve forecast skill and/or identify other features that are influential to TC formation.

Bibliography

Agee, E. M., 1972: Note on ITCZ wave disturbances and formation of Tropical Storm Anna. *Mon. Wea. Rev.,* **100,** 733–737.

Aiyyer, A. and J. Molinari, 2008: MJO and tropical cyclogenesis in the Gulf of Mexico and Eastern Pacific: case study and idealized numerical modeling. *J. of Atmos. Sci.*, **65**, 2691-2704.

Barrett, Bradford S., Lance M. Leslie, 2009: Links between Tropical Cyclone Activity and Madden–Julian Oscillation Phase in the North Atlantic and Northeast Pacific Basins. *Mon. Wea. Rev.*, **137**, 727–744.

Bister, M., and K.A. Emanuel, 1997: The genesis of Hurricane Guillermo: TEXMEX analyses and a modeling study. *Mon. Wea. Rev.*, **125**, 2662–2682.

Bermowitz, R.J. and E.A. Zurndorfer, 1979: Automated guidance for predicting quantitative precipitation. *Mon. Wea. Rev.*, **107**, 122-128.

Beven, J.L., 2011: Tropical Cyclone report: Tropical Storm Agatha. National Hurricane Center.

Bochkova, T.M., 1978: Use of parametric linear discriminant analysis to forecast major air temperature anomalies. *Soviet Meteor. Hydrol.*, 84-86.

Brier, G.W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1–3.

Briegel L.M. and W.M. Frank, 1997: Large-scale influences on tropical cyclogenesis in the Western North Pacific. *Mon. Wea. Rev.*, **125**, 1397-1413.

Briggs, W., and D. Ruppert, 2005: Assessing the skill of yes/no predictions. Biometrics, 61, 799–807.

Brown, D.P., 2008: Tropical Cyclone report: Tropical Storm Alma. National Hurricane center.

Charney, J.G., and A. Eliassen, 1964: On the growth of the hurricane depression. *J. Atmos. Sci.,* **21**, 68-75.

Coelho, C.A. et al., 2003: Skill of coupled model seasonal forecasts: a Bayesian assessment of ECMWF ENSO forecasts. *Technical Memorandum*, **462**, 16pp.

Colon, J.A., and W.R. Nightingale, 1963: Development of tropical cyclones in relation to circulation patterns at the 200 mb level. *Mon. Wea. Rev.,* **91**, 329-336.

Camargo, S.J. et al., 2007: Use of a genesis potential index to diagnose ENSO effects on tropical cyclone genesis. *J. of Clim.*, **20**, 4819-4834.

Camargo, S.J. et al., 2008: Clustering of eastern North Pacific tropical cyclone track: ENSO and MJO effects. *Geochemistry Geophysics Geosystems*., **9**, 23pp.

Corbosiero, K.L., M.J. Dickinson, and L.F. Bosart, 2009: The contribution of Eastern North Pacific tropical cyclones to the rainfall climatology of the southwest United States. *Mon. Wea. Rev.*, **137**, 2415-2435.

Davidson, N.E., H.H. Hendon, 1989: Downstream Development in the Southern Hemisphere Monsoon during FGGE/WMONEX. *Mon. Wea. Rev.*, **117**, 1458–1470.

Davis, C.A., and L.F. Bosart, 2001: Numerical simulations of the genesis of Hurricane Diana (1984). Part I: Control Simulation. *Mon. Wea. Rev.*, **129**, 1859-1881.

Davis. C.A., and L.F. Bosart, 2002: Numerical Simulations of the genesis of Hurricane Diana (1984). Part II: Sensitivity of track and intensity prediction. *Mon. Wea. Rev.*, **130**, 1100-1124.

Davis, C.A. et al., 2008: A vortex-based perspective of eastern Pacific tropical cyclone formation. *Mon. Wea. Rev.*, **136**, 2461-2477.

Dewey, K.F., 1975: The prediction of Lake Huron lake-effect snowfall systems. *J. of Applied Meteorology*, **14**, 3-7.

Dunkerton, T. J., M. T. Montgomery, and Z. Wang (2009), Tropical cyclogenesis in a tropical wave critical layer: Easterly waves. *Atmos. Chem. Phys., 9,* 5587-5646.

Dunion, J.P. and C.S. Velden, 2004: The impact of the Saharan air layer on Atlantic tropical cyclone activity. *BAMS*, **85**, 353-365.

Elsberry R.L. and P.A. Harr, 2008: Tropical cyclone structure (TCS08) field experiment science basis, observational platforms and strategy. *Asia-Pacific J. Atmos. Sci.*, **44**, 23pp.

Elsner, J.B., 2003: Tracking Hurricanes. *BAMS*, **84**, 353 -356.

Emanuel, K.A., 1989: The finite-amplitude nature of tropical cyclogenesis. *J. Atmos. Sci.*, **46**, 3431-3456.

Emanuel, K.A., 1991: The theory of hurricanes. Annu. Rev. Fluid. Mech., 23, 179-196.

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

Frank, N.L., 1970: Atlantic tropical systems of 1969. *Mon. Wea. Rev.*, **98**, 307-314.

Frank, N.L., 1976: Atlantic tropical systems of 1976. *Mon. Wea. Rev.,* **104**, 466-474.

Frank, W.M., and P.E. Roundy, 2006: The role of tropical waves in tropical cyclogenesis. *Mon. Wea. Rev.*, **134**, 2397–2417.

Fu, B. et al. 2007: Analysis of Tropical Cyclogenesis in the Western North Pacific for 2000 and 2001. *Wea. and Forecasting*, **22**, 763-780.

Godfrey, R.A., 1982: An application of model output statistics to the development of a local wind regime forecast procedure. *J. of Applied Meteorology*, **21**, 1786 -1791.

George, J.E. and W.M. Gray, 1977: Tropical cyclone recurvature and nonrecurvature as related to surrounding wind-height fields. *J. of Applied Meteorology*, **16**, 34 -42.

Gerrish, H. P., and M. Mayfield, 1989: Eastern North Pacific tropical cyclones of 1988. *Mon. Wea. Rev.,* **117,** 2266–2277.

Gray, W.M., 1968: Global view of the origin of tropical disturbances and storms. *Mon. Wea. Rev.*, **96**, 669-700.

Gray, W.M., 1984: Atlantic seasonal hurricane frequency. Part I: El Nino and 30 mb quasi-biennial oscillation influences. *Mon. Wea. Rev.,* **112**, 1649-1688.

Gunther E.B. and R.L. Cross, 1985: Eastern North Pacific tropical cyclones of 1984. *Mon. Wea. Rev.,* **113**, 1393-1410.

Gunther E.B. and R.L. Cross, 1986: Eastern North Pacific tropical cyclones of 1985. *Mon. Wea. Rev.,* **114**, 1931-1949.

Gunther E.B. and R.L. Cross, 1987: Eastern North Pacific tropical cyclones of 1986. *Mon. Wea. Rev.,* **115**, 2507-2523.

Hand, D.J., and K. Yu, 2001: Idiot's Bayes: Not so stupid after all? *Int. Stat. Rev.*, **69**, 385-398.

Hartmann, D. and E.D. Maloney, 2001: The Madden-Julian Oscillation, Barotropic Dynamics, and North Pacific tropical cyclone formation. Part I: observations. *J. Atmos. Sci*., **58**, 2545-2558.

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

Hoover, B.T., 2010: Identifying barotropic growth structures in east Pacific tropical cyclogenesis with adjoint-derived sensitivities. *Dynamical sensitivity analysis of tropical cyclone steering and genesis using an adjoint model*, 85-121.

Hux, J.D., P.C. Knappenberger, P.J. Michaels, P.J. Stenger, H.D. Cobb, and M.P. Rusnak, 2001: Development of a discriminant analysis mixed precipitation (DAMP) forecast model for mid-Atlantic winter storms. *Wea. Forecasting*, **16**, 248–259.

Janowiak, J.E., et al., 1995: An examination of the east Pacific ITCZ rainfall distribution. *J. of Climate*, **8**, 2810-2823.

Jennrich, R.I., and P.F. Sampson, 1977: BMDP7M, stepwise discriminant analysis. *BMDP Biomedical Computer Programs,* W.J. Dixon, Ed., Los Angeles, University of California Press, 711-733 and 833-839.

Kerns, B., and E. Zipser, 2009: Four years of tropical ERA-40 vorticity maxima tracks. Part II: Differences between developing and nondeveloping disturbances. *Mon. Wea. Rev.*, **137**, 2576-2591.

Kieu, C.Q., and D.L. Zhang, 2008: Genesis of Tropical Storm Eugene (2005) from merging vortices associated with ITCZ breakdowns. Part I: observational and modeling analyses. *J. Atmos. Sci.*, **65**, 3419–3439.

Kiladis, G.N. et al. 2005: Zonal and vertical structure of the Madden-Julian Oscillation. *J. Atmos. Sci.*, **62**, 2790-2809.

Kumar, V. and Krishnan, R.R. et al., 2005: On the association between the Indian summer monsoon and the tropical cyclone activity over northwest Pacific. *Current Science*, **88**, 602 -612.

Kossin, J.P., and M. Sitkowski, 2009: An objective model for identifying secondary eyewall formation in hurricanes. *Mon. Wea. Rev.*, **137**, 876-892.

Kuleshov, Y. et al. 2009: Tropical cyclone genesis in the Southern Hemisphere and its relationship with the ENSO. *Ann. Geophys.*, **27**, 2523-2538.

Landsea, C.W., 1995: "Climate Variability table — Tropical Cyclones". Atlantic Oceanographic and Meteorological Laboratory, NOAA. Retrieved 2012-08-01.

Landsea, C.W., et al. 1998: The extremely active 1995 Atlantic hurricane season: environmental conditions and verification of season forecasts. *Mon. Wea. Rev.*, **126**, 1174-1193.

Leroy, Anne, Matthew C. Wheeler, 2008: Statistical Prediction of Weekly Tropical Cyclone Activity in the Southern Hemisphere. *Mon. Wea. Rev.*, **136**, 3637–3654.

Lipps, F.B., 1970: Barotropic stability and tropical disturbances. *Mon. Wea. Rev.*, **98**, 122–131.

Madden R.A. and P.R. Julian, 1971: Detection of a 40-50 day oscillation in the Zonal wind in the Tropical Pacific. *J. Atmos. Sci.*, **28**, 702 -708.

Madden R.A. and P.R. Julian, 1994: Observations of the 40-50-day oscillation – a review. *Mon. Wea. Rev.*, **122**, 814-837.

Maloney, E.D. and D. Hartmann, 2001: The Madden-Julian Oscillation, Barotropic Dynamics, and North Pacific tropical cyclone formation. Part I: observations. *J. Atmos. Sci*., **58**, 2545-2558.

Mason, S.J., and G.M. Mimmack, 2002: Comparison of some statistical methods of probabilistic forecasting of ENSO. *J. Climate*, **15**, 8–29.

Miller, R.G., 1962: Statistical prediction by discriminant analysis. *Meteor. Monogr.*, **4**, No. 25, 1-54.

Molinari, J. et al., 2000: Origins and mechanisms of eastern Pacific tropical cyclogenesis: a case study. *Mon. Wea. Rev.*, **128**, 125-139.

Montgomery, M. T. et al. 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*, 9879-9900.

Montgomery, M.T. et al., 2006: A vortical hot tower route to tropical cyclogenesis. *J. of Atmos. Sci.*, **63**, 355-386.

Moser J.B. and J.A. Zehnder, 1994: Cluster analysis of eastern North Pacific tropical cyclogenesis precursors. *JGR*, **99**, 8085-8093.

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

Palmén, E., 1948: On the formation and structure of tropical Hurricanes.

Pasch, R.J. et al., 2006: An evaluation and comparison of predictions of tropical cyclogenesis by three global forecast models. *27th Conference on Hurricanes and Tropical Meteorology*

Pasch, R.J. et al., 2008: Performance of the GFS in predicting tropical cyclone genesis during 2007. *28th Conference on Hurricanes and Tropical Meteorology*, Orlando, FL, AMS. [Available online at *http://ams.confex.com/ams/28Hurricanes/techprogram/paper_138218.htm*.]

Pasch, R.J., E.S. Blake, L.A. Avila, J.L. Beven, D.P. Brown, J.L. Franklin, R.D. Knabb, M.M. Mainelli, J.R. Rhome, and S.R. Stewart, 2009: Eastern North Pacific hurricane season of 2006. *Mon. Wea. Rev.*, **137**, 3–20.

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

Philippon N., P. Camberlin, and N. Fauchereau, 2002: Empirical predictability study of October-December East African rainfall. *Q.J.R. Meteorol. Soc.*, **128**, 2239 -2256.

Pratt, A.S. and J.L. Evans, 2004: An evaluation of the Global Forecasting System (GFS) and Navy Operational Global Atmospheric Prediction System (NOGAPS) forecasting skill of tropical cyclogenesis. *26th Conference on Hurricanes and Tropical Meteorology*, 354 -355.

Ramage, C.S., 1959: Hurricane development. *J. of Meteor.*, **16**, 227-237.

Randerson, D., 1977: Determining the relative frequency of occurrence of local cumulonimbus activity through discriminant analysis. *Mon. Wea. Rev.*, **105**, 709-712.

Raymond, D.J. and C. Lopez-Carrillo, 1998: Case-studies of developing east Pacific easterly waves. *Q.J.R. Meteorol. Soc.*, **124**, 2005.

Raymond, D. J., S. L. Sessions, and C. Lopez Carrillo, 2011: Thermodynamics of tropical cyclogenesis in the northwest Pacific, J. Geophys. Res., 116, D18101, doi:10.1029/2011JD015624

Reed, R.J. and E.E. Recker, 1971: Structures and Properties of Synoptic-scale wave disturbances in the equatorial western Pacific. *J. Atmos. Sci*., **28**, 1117-1133.

Reihl, H., 1950: A Model of Hurricane Formation. *J. of Applied Physics*, **21**, 917-925.

Ritchie, Elizabeth A., Greg J. Holland, 1999: Large-Scale Patterns Associated with Tropical Cyclogenesis in the Western Pacific. *Mon. Wea. Rev.*, **127**, 2027–2043.

Ropelewski, C.F. and M.S. Halpert, 1987: Global and Regional scale precipitation patterns associated with the El Niño/Southern Oscillation. *Mon. Wea. Rev*., **115***,* 1606-1626.

Rotunno, R. and Emanuel, K.A. 1987. An air-sea interaction theory for tropical cyclones. Part II: Evolutionary study using a nonhydrostatic axisymmetric numerical model. *J. Atmos. Sci.,* **44**, 542- 61.

Roundy, P.E, and W.M. Frank, 2004: Effects of low-frequency wave interaction on intraseasonal oscillations. *J. Atmos. Sci.*, **61**, 3025-3040.

Sadler, J.C., 1964: Tropical Cyclones of the eastern North Pacific as revealed by TRIOS observations. *J. of Applied Meteorol.*, **3**, 347-366.

Schreck III, C.J. et al. 2012: A global view of equatorial waves and tropical cyclogenesis. *Mon. Wea. Rev.*, **140**, 774-788.

Shapiro, L.J., 1977: Tropical storm formation from easterly waves: a criterion for development. *J. Atmos. Sci.*, **34**, 1007-1021.

Thorncroft, C.D. and B.J. Hoskins, 1994: An idealized study of African easterly waves. I: A linear view. *Q.J.R. Meteorol. Soc.*, **120**, 953-982.

Thompson, O. E., and J. Miller, 1976: Hurricane Carmen: August - September 1974: Development of a wave in the ITCZ. *Mon. Wea. Rev.,* **104,** 1194–1199.

Tippet, M.K. et al. 2011: A Poisson regression index for tropical cyclone genesis and the role of large-scale vorticity in genesis. *J. of Clim.*, **24**, 2335-2357.

Tonkin, H. et al. 2000: An evaluation of thermodynamic estimates of climatological maximum potential tropical cyclone intensity. *Mon. Wea. Rev.*, **128**, 746-762.

Velasco, I. and M.J. Fritsch, 1987: Mesoscale convective complexes in the Americas. *JGR*, **29**, 9591-9613**.**

Vimont, D.J. and J.P. Kossin, 2007: The Atlantic meridional mode and hurricane activity. *Geo. Res. Let.*, **34**, L07709.

Walker, G. T., 1923: correlations in seasonal variations of weather, VIII. A preliminary study of world weather I, *Mem. India. Meteorol. Dept.,* **23,** 75-131.

Watterson I.G., J.L. Evans, and B.F. Ryan, 1995: Seasonal and interannual variability of tropical cyclogenesis: Diagnostics from large-scale fields. *J. Climate*, **8**, 3052-3066.

Wheeler, M.C. and G.N. Kiladis, 1999: Convectively coupled equatorial waves: analysis of clouds and temperature in the wavenumber-frequency domain. *J. Atmos. Sci.*, **56**, 374-399.

Wheeler, M.C., and H.H. Hendon, 2004: An all-season real-time multivariate MJO index: development of an index for monitoring and prediction. *Mon. Wea. Rev.*, **132**, 1917–1932.

Whitney, L.D., and J.S. Hobgood, 1997: The relationship between sea surface temperatures and maximum intensities of tropical cyclones in the eastern north pacific ocean. *J. of Clim.*, **10**, 2921-2930.

Wilks, D.S., 2006: *Statistical methods in the atmospheric sciences*. 2nd ed. International Geophysics Series, Vol. 91, Academic Press, 627 pp.

Wu, L., et al. (2006), Simulating the formation of Hurricane Isabel (2003) with AIRS data, Geophys. Res. Lett., 33, L04804, doi:10.1029/2005GL024665.

Zehnder J.A. et al., 1999: The interaction of easterly waves, orography, and the intertropical convergence zone in the genesis of eastern Pacific tropical cyclones. *Mon. Wea. Rev.*, **127**, 1566-1585.

Zehnder J.A. and R.L. Gall, 1991: Observational studies of tropical cyclone initiation in the eastern North Pacific, *Atmosfera*, **4**, 37-51.

Zhang, H., 2006: On the optimality of naïve Bayes with dependent binary features. *Pattern Recog. Lett., **27***, 830-837.