Seasonal Forecasting of Tropical Cyclone Formation in the Australian Region

Angelika Werner
Dipl. Met.

A thesis submitted in fulfilment
of the requirements for the degree of
Doctor of Philosophy (Ph.D)

MACQUARIE
UNIVERSITY

FACULTY OF SCIENCE

Department of Environment and Geography
Faculty of Science
Macquarie University
Sydney, Australia

May 2011

“I've lived in good climate,
and it bores the hell out of me.
I like weather rather than climate.”

John Steinbeck, American Writer (1902-1968)

CONTENTS

ABSTRACT

The hazard of tropical cyclones (TCs) is a very relevant topic to the Australian economy and to the welfare of its northern population. Australia's climate and the interannual variability of Australian region TC formation (genesis; TCG) is strongly dominated by the ocean-atmosphere interannual climate pattern El Niño-Southern Oscillation (ENSO) and to a lesser extent by other climate modes of variability. This thesis investigates new ways of seasonal forecasting Australian region TCG counts and distribution by identifying potentially skilful climatological predictors and applying more advanced statistical modelling methods than previous models for the region.

ENSO is known to be the most important predictor of seasonal variations in TCG for the Australian region. To investigate the ENSO-independent effects of the Indian Ocean Dipole (IOD) on Australian TCG, a simple, but effective method has been developed to separate the IOD from ENSO. Results demonstrate, that there is reasonable individual forecast skill afforded by the influence of the isolated IOD. In combination with common ENSO metrics, however, the IOD does not significantly improve seasonal forecasting of seasonal TCG counts in the Australian region or subregions.

A Poisson regression model using Bayesian inference and the Markov chain Monte Carlo (MCMC) method was developed to forecast seasonal TCG counts in the Australian region. The final three-predictor model based on derived indices of subtropical Central Pacific June-July-August average convective available potential energy (CAPE), the tropical northeast Pacific May-June-July average meridional winds at 850 hPa (v_{850}) and subtropical central South Pacific June-July-August geopotential height at 500 hPa performs best with the corresponding correlation coefficient between observed annual TCG totals and cross-validated model hindcasts of r = 0.73 over the 40-year record between 1968/89-2007/08. The model is adaptable for hindcasting seasonal TCG totals in Australia's Eastern (Coral Sea) TC subregion, while it lacks skill in the Western (eastern Indian Ocean) TC subregion (r = 0.79 and r = 0.38 respectively). To improve forecasts of annual TCG counts in the Western region (90°-135°E), a separate model was developed with correlations between cross-validated hindcasts and observed annual TCG count of r = 0.67 using the June-July-August tropical Central Pacific sea level pressure (SLP) and the above used index of v_{850} as predictors.

A logistic regression approach applied in the Bayesian seasonal forecast model was found to be successful in forecasting spatial probabilities of Australian region TCG on a 2.5° x 2.5° grid for the upcoming season. The most skilful model is based on the SLP, NINO4 and v_{850} indices, combined with spatial information from CAPE and shows an average improvement over the climatological average of 25%. The average distribution of TCG probabilities over the study period, as well as the hindcasted strong variations of probabilities and distribution of TCG during ENSO events match remarkably well against observations over most of the study domain.

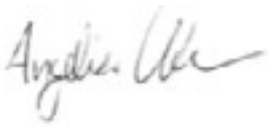
Results demonstrate that the combination of dynamic with synoptic and/or thermodynamic features is most useful to identify climatic influences on the seasonal frequency and spatial distribution of TC development in the Australian region. Independent forecasts using the three introduced models and comparisons with current operational models demonstrate the relatively high skill of the models presented in this thesis.

DECLARATION

I certify that the work in this thesis entitled "Seasonal Forecasting of Tropical Cyclone Formation in the Australian Region" has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree to any other university or institution other than Macquarie University.

I also certify that the thesis is an original piece of research and it has been written by me. Any help and assistance that I have received in my research work or the preparation of the thesis itself have been appropriately acknowledged.

This thesis differs from a traditional thesis by including, in their traditional form, a series of authored and co-authored journal articles in II Papers.



Angelika Werner
Student Number 41251393

ACKNOWLEDGEMENTS

This Ph.D was a fantastic adventure and in hindsight I am very glad I got this opportunity, even though it wasn't that easy at times. Studying so far from home proved to be quite a challenge and it would not have been possible without the help and support of supervisors, colleagues and friends.

First and foremost, I would like to thank my supervisors, Assoc Prof Neil Holbrook and Dr Angela Maharaj for their constant support, guidance, patience and encouragement and for always making themselves available for discussions throughout my Ph.D candidature at Macquarie University. I would also like to thank Neil for offering me the opportunity to undertake my Ph.D at Macquarie University under the funding of the MQRES scholarship and let me follow my interests regarding the selection of the topic. Even though he left to the University of Tasmania at the start of my candidature, he kept strongly involved and interested in my work and more than once his advice with thesis directions was incredibly helpful and important. Many thanks also to my co-supervisor, Dr. Kevin Cheung, for his cooperation and help throughout my studies and in particular when he took over the official supervision after Neil and Angela both have left Macquarie University.

I also want to thank the Higher Degree Research Council at Macquarie University for not only providing my scholarship, but also for the various sources of funding during my candidature. It enabled me to attend many national and international conferences and workshops, which elevated my work to a higher level and helped me to maintain a large international network, and also gain the skill of communicating my results with increasing confidence. When I finally moved to Hobart over the last year of my candidature, UTAS generously provided office space and access to computational resources.

Many thanks also to Steve Wright for his help in explaining the fine miracles of Bayesian modelling and Michael Baxter and Phil Ingram for helping to solve the many UNIX and other computational problems. Thanks also my fellow students and friends who have come and gone during my time at Macquarie and UTAS (Nik, Erla, Christina, Anna, André, Richard, Supriya, Suresh, Kerstin, Pipp, Eric, Shirley, Mauro and many more) many of whom turned into very good friends.

I want to thank my parents, brother and grand parents for backing me when I first decided to move to Australia for my Ph.D studies. I know they had quite a hard time when I made this decision and throughout those years, but they nevertheless supported me all the way through. Also million thanks for the backups of my many friends from home (Moni, Moni, Chris, Flo, Karin, Conny, Stefan, Richi, Hansi, Ana, HP, Antonia, Lynn) who always had an open ear when I was suffering of homesickness. Finally I would like to thank my dear friend Nicola Smith and her fantastic family. Nik and her family gave me a warm welcome to a second home and it would have been infinitely harder to get through those years without knowing her.

Thank you all for keeping a smile on my face.

I INTRODUCTION AND METHODOLOGY

1 INTRODUCTION

The hazard of tropical cyclones (TCs) affecting the Australian region is a topic of considerable importance to the Australian economy and to the welfare of its northern populations. TCs are one of the most devastating natural disasters affecting Australia every year. With winds of up to 200-300 kmh^{-1}, accompanying storm surges and severe rainfall, severe damage and loss of life are expected. In the most recent 2010/11 tropical cyclone season, the Queensland coast was crossed by a number of TCs, with two storms causing severe economic loss. In December 2010, TC Tasha and its remnants caused widespread flooding, which marked the beginning of the now well-known Queensland floods of 2011. On February 2nd 2011, TC Yasi made landfall just south of Innisfail on Australia's northeast coast as a Category 5 tropical cyclone, producing wind gusts as high as 300 kmh^{-1} and an accompanying storm surge of 3 m and more. The fact that there was no loss of life was due to a remarkably impressive model performance of the track and intensity forecasts, and a good preparedness and early action by authorities regarding the hazard. However, despite the model performance of this single event being particularly impressive, seasonal forecasts of tropical cyclone numbers for the Australian region in the upcoming season have considerable room for improvement. For the most recent Australian TC season, 2010/11, the Australian Bureau of Meteorology (BoM) forecast 20-22 TCs for the Australian TC region (0°-30°S and 90°-160°E) due to the very strong La Niña conditions (http://www.bom.gov.au/climate/ahead/tc.shtml) evident and developing at the time. The actual observed number of TCs originating in, or moving into the Australian region, turned out to be 11 which fell well short of the BoM forecast total, and was only slightly above the climatological average.

Australia's climate is strongly affected by the ocean-atmosphere interannual climate pattern El Niño-Southern Oscillation (ENSO). ENSO modulates the climate in the tropical and subtropical Pacific Ocean regions, but also shows strong climatic connections with the Indian Ocean region. Most previous research on Australian

region TCs have concentrated on tropical cyclone relationships with classical ENSO indices and therefore seasonal TC forecast models in the Australian region are mostly limited to the potential forecast skill of common ENSO-measures. Large-scale ENSO-related changes to the climatic environment have been mostly neglected regarding the possible effects on Australian region TCs. Further, the possible Indian Ocean effects on Australian region TCs has not previously been investigated in any comprehensive way.

The key aims of this thesis are:

1. to improve our understanding of Australian region TC formation (genesis: TCG) and the relationship with large-scale climate variables;
2. to investigate the importance of ENSO-independent variables on TCG, specifically the role of the ENSO-independent Indian Ocean Dipole; and
3. to build statistical forecasting models that improve seasonal forecast skill of TCG for the Australian region, both in terms of annual counts and spatial probabilistic estimates.

These aims set the scene for the project to build a set of statistical seasonal forecast models for the number of TCs expected to form in the Australian region and subregions in the upcoming season, but also the spatial distribution where TCs are likely to form. The project aims to build tools that substantially improve upon previous and existing operational forecast models, so that vulnerable populations might better prepare for tropical cyclone risks in the upcoming season(s).

Chapter 1 is structured into four sections. Section 1 addresses the leading atmospheric and oceanic modes in the Indo-Pacific region and their large-scale effects on the climatology of the Australian region. Section 2 describes the physical mechanisms of a tropical cyclone and the necessary conditions for development. Section 2 also provides a statistical overview of TCs in the Australian region. Section 3 gives a history of statistical seasonal forecasting and Section 4 presents the objectives and structure of the thesis.

1.1 Interannual Ocean-Atmosphere Modes in the Tropical Indo-Pacific Region

Australia's tropical climate is dominated by two large-scale interannual climate modes: ENSO in the Pacific basin and Indian Ocean Dipole (IOD) in the Indian Ocean (e.g., Philander 1990; Saji et al. 1999; Webster et al. 1999). The dominant mode of monthly SSTA (1979-2004) in the tropical Pacific (30°N-30°S, 110°E-70°W) explains 45% of the tropical Pacific SSTA variance, and identifies the classic canonical (cold tongue) El Niño pattern (Ashok et al. 2007). In the Indian Ocean, 30% of the total SST variability can be attributed to ENSO, while the IOD explains around 12 %. (Saji et al. 1999) These modes and their effects on the Australian climate have been the subject of numerous studies (e.g., Cai et al. 2001; Ashok et al. 2003; Meyers et al. 2007; Ramsay et al. 2008; Hendon et al. 2009; Ummenhofer et al. 2009). Importantly, rainfall variability in Australia is shown to be associated with the IOD almost as significantly as with ENSO (Saji and Yamagata 2003). Nevertheless, there remains much ongoing debate over the extent to which the IOD is indeed a unique 'mode' of the climate system and hence whether it is therefore really independent of ENSO (e.g., Saji et al. 1999; Allan et al. 2001; Meyers et al. 2007; Risbey et al. 2009). In the following both of these large-scale ocean-atmosphere signals, ENSO and IOD, and their interannual modulation of the large-scale tropical and subtropical atmosphere and the effects on the Australian climate are discussed. As this study focuses on seasonal variations, intra-seasonal signals, such as the Madden-Julian-Oscillation are not further investigated.

1.1.1 The El Niño-Southern Oscillation

ENSO is the most important ocean-atmosphere phenomenon in the Pacific and affects the large-scale climate all over the globe (e.g., Philander 1985, 1990, 1991; Lau 1985; Cane et al. 1986; Ropelewski and Halpert 1986; Allan 1988). Its effects, particularly on precipitation distribution, and also on tropical cyclones, are felt in many regions, particularly in Central and South America, South Asia and Australia.

The normal conditions in the tropical Pacific can be described by an extended warm pool area with a deep thermocline (around 200m) in the western equatorial Pacific and a cold dipole extending zonally from the tropical western into the central Pacific. The reason for this (around 6°C colder) 'cold tongue' is a shallow thermocline in the western tropical Pacific allowing upwelling of cold deep water. Due to this ocean temperature dipole, a low pressure system and convection occurs in the West Pacific warm pool region and a high pressure system and subsidence at the eastern rim of the tropical Pacific. These pressure systems form a closed circulation, called the Walker circulation – or the Southern Oscillation. The induced surface-near trade winds further enhance the SST dipole by piling up water in the tropical West Pacific and further pushing down the thermocline in that region (Philander 1990).

The mechanism of ENSO can be traced back to a positive feedback on an initial perturbation at the equatorial thermocline (Bjerknes 1969). The equatorial SST dipole in the Pacific initiates zonal easterly trade winds towards the tropical Western Pacific warm pool. This enables a further enhancement of the SST dipole, as the slope in the tropical thermocline increases, allowing more upwelling of cold water in the East Pacific and establishes at the same time the warm pool in the Western Pacific. How strong this Bjerknes feedback between ocean and atmosphere in the normal climatological conditions or in the ENSO phases is therefore strongly dependent on the structure of the tropical thermocline. Applying the Bjerknes feedback theory to explain an El Niño event, consider a warm perturbation in the equatorial East Pacific. The resulting reduction in SST gradient between the Pacific SST dipoles leads to a weakening of the pressure gradient induced by the SSTs and therefore a weakening of

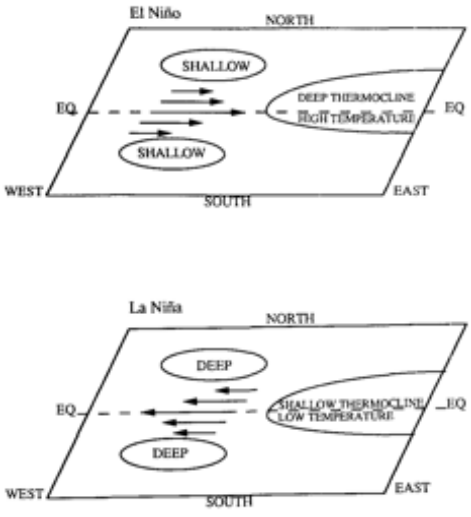


Figure 1.1 A schematic diagram showing spatial distributions of surface wind stress, sea surface temperature, and heat storage at peak phases of El Niño and La Niña. The arrows indicate wind anomalies during El Niño. They depress the thermocline in the east where sea surface temperature anomalies are large. In the west they elevate the thermocline primarily off the equator, in the areas indicated by the contours, because of their curl (Philander 1991).

the easterly trade winds and a flattening of the thermocline slope. This mechanism eventually ends with a complete breakdown of the Walker circulation and an interruption of the deepwater upwelling in the Eastern Pacific leading to El Niño conditions. Accordingly a La Niña event is induced with a cold source perturbation in the equatorial East Pacific. A resulting enhancement of the SSTA gradient and subsequently a stronger pressure gradient enforces the easterly trade winds. Figure 1.1 shows an idealized schematic of the interaction between the thermocline and the induced wind stress leading to El Niño and La Niña conditions (Philander 1991).

Various studies have shown the interannual nature of the ENSO signal, in the range of 2-8 years. The duration of a single ENSO event is 1.5-2 years (Rasmusson 1984; Wright et al. 1985) and tends to oscillate with the seasonal cycle (Philander, 1985). An event can first be recognized during late austral spring as SST anomaly in the tropical Southwest Pacific regions (Nicholls 1984a,b, 1985b). Figure 1.2 shows the two leading empirical orthogonal functions (EOFs) of monthly SSTs over 1979-2004. While the first EOF clearly reveals the classical ENSO pattern with the equatorial cold-tongue and out-of–phase off-equatorial boomerang (Rasmusson and Carpenter 1982) explaining 45% of the total monthly SST anomalies, the second EOF shows the El Niño Modoki pattern (Ashok et al. 2007) which accounts for 12% of the total SST variations. This pattern is associated with an anomalous SST in the tropical central Pacific with a horse-shoe-shaped extension into higher latitudes, while the eastern and western equatorial Pacific have the same sign, but out-of-phase with the central Pacific anomaly. In this study, we want to focus on the classical ENSO pattern, due to its known strong impact on the Australian climate, which will be reviewed below.

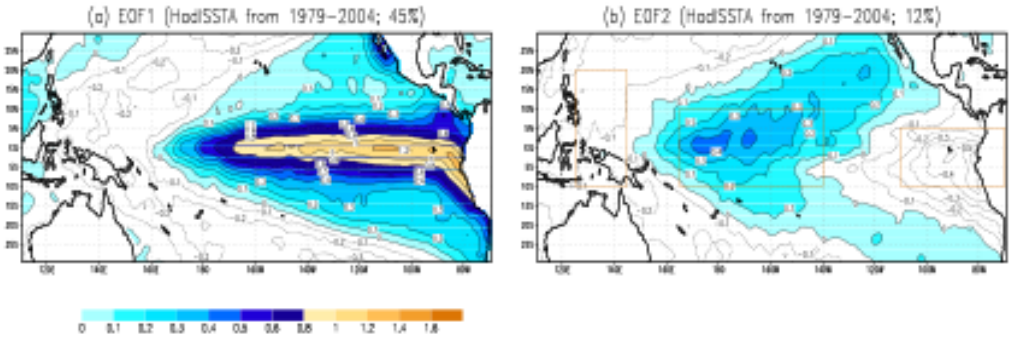


Figure 1.2 Top two EOF modes of tropical Pacific SSTA (1979-2004) multiplied by respective standard deviations of the principal components; units in °C (Ashok et al. 2007).

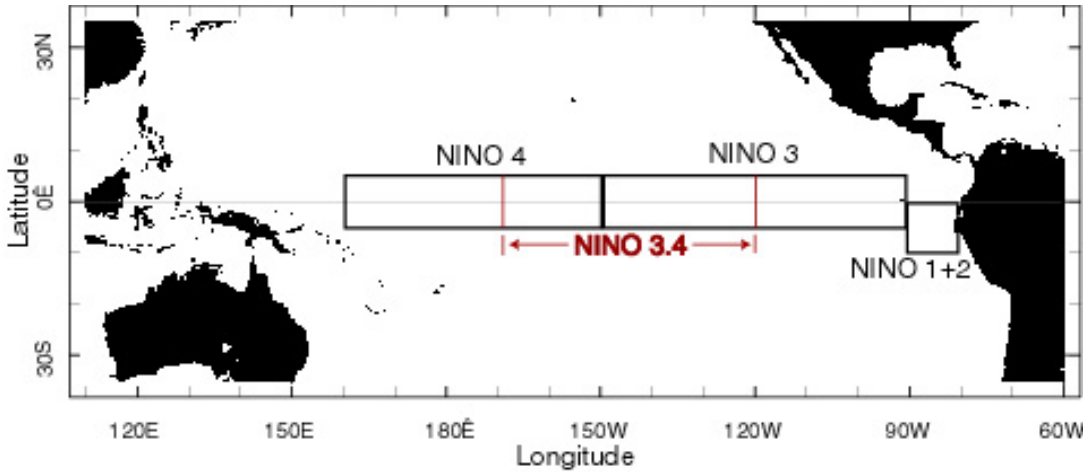


Figure 1.3 Location of prominent SSTA ENSO indices (http://iri.columbia.edu/climate/ENSO/background/monitoring.html).

Many classical ENSO indices describe the signal based on the sea level pressure (SLP) or SST anomalies (SSTA) responsible for the large-scale changes in the climatology due to ENSO. The classical index used in many older studies is the Southern Oscillation Index (SOI), which is defined as the pressure gradient between Darwin in Australia and Tahiti and was already recognized by Sir Gilbert Walker in the early 20th century. Recent studies prefer to use SST indices, as the ocean is considered the physical driver of ENSO. Figure 1.3 shows the locations of the SSTA indices NINO 1+2 (0°-10°S, 80°-90°W), as the region where it is first recognized if the upwelling process is interrupted, NINO 3 (5°S-5°N; 150°-90°W) as the usual location of the East pacific cold tongue and the more central to central East Pacific indices NINO 3.4 (5°S-5°N; 170°-120°W) and NINO4 (5°S-5°N; 160°E-150°W). Another SSTA measure is the larger cold tongue index defined from 6°N-6°S; 180°-90°W. There are also various newer ENSO indices based on different climate variables such as zonal wind components (trade wind index) or outgoing longwave radiation (ORL).

There are various studies looking into the large-scale modulation of the atmosphere as a result of ENSO on different time-scales (e.g., van Loon and Shea 1985, 1987; Karoly 1989; Drosdowsky and Williams 1991; Zhang et al. 1997; Garreaud and Battisti 1999; Seager et al. 2003; Lu et al. 2008). As this thesis focuses on seasonal forecasting, we focus henceforth on the inter-annual ENSO variations. Seager et al. (2003) showed an El Niño-accompanied consistent warming of the tropical atmosphere as a result of the anomalous ocean-atmosphere heat flux from

increased SSTs. Poleward from the tropics, between 25° and 40° in the Northern Hemisphere (NH) and 35° and 50° in the Southern Hemisphere (SH) an eddy-driven upward motion cools the troposphere in a boomerang-shape around the central Pacific (Lu et al. 2008). Also during El Niño conditions the Hadley circulation is more compressed and intensified, resulting in an equatorward shift of the strengthened Subtropical Jets. The change in the quasi-barotropic zonal winds poleward of 20°N and 30°S can be traced back to a feedback mechanism (Robinson 2002; Seager et al. 2003). Gaurreaud and Battisti (1999) investigated the interannual effects of ENSO on climate variables. A 6-year high-pass filter was applied to the cold tongue index to filter out all possible interdecadal variations (CT*). Spatial fields of climate variables were then regressed on the CT* time series (Figure 1.4). The regressed SST field (Figure 1.4a) reveals the above shown (see Figure 1.2) classical ENSO SSTA-tongue pattern with the boomerang shaped out-of-phase SSTAs in the central Pacific centred around 30 degrees. In the Indian Ocean an extended anomaly of the same sign as the cold-tongue is associated with the ENSO. The regressed surface winds (Figure 1.4b) show the change in trade winds along the equator, but also the increased meridional inflow in the central and East Pacific from the Northern Hemisphere during warm events. Also strong anomalies can be seen along the East Australian Current with an increased surface inflow in the central and west Pacific from the Southern Hemisphere. In the midlatitudes, the strongest anomalies can be seen in both hemispheres around the subtropical highs. ENSO related surface wind anomalies in the other ocean basins appear marginal. The SLP pattern shows the strong dipole between the tropical and subtropical East Pacific with the other pole extending from the tropical and subtropical West Pacific across the Australian continent into the Western Indian Ocean. In the mid-latitudes we find the dominant patterns centred around 60°S with a barotropic character. During El Niño, this pressure anomaly blocks the westerly flow in conjunction with the weakening of the Subtropical High (Figure 1.4c; Garreaud and Battisti 1999). Another salient characteristic during an El Niño can be observed in the 500 hPa geopotential height anomalies. A midlatitude trough with an accompanying poleward high pattern in the geopotential height field (Figure 1.5) has been intensively discussed in the literature (e.g., van Loon and Shea, 1985; Garreaud and Battisti 1999; Lu et al. 2008). In the Southern Hemisphere, this trough is located east of New Zealand. The pattern is strongest in austral winter and

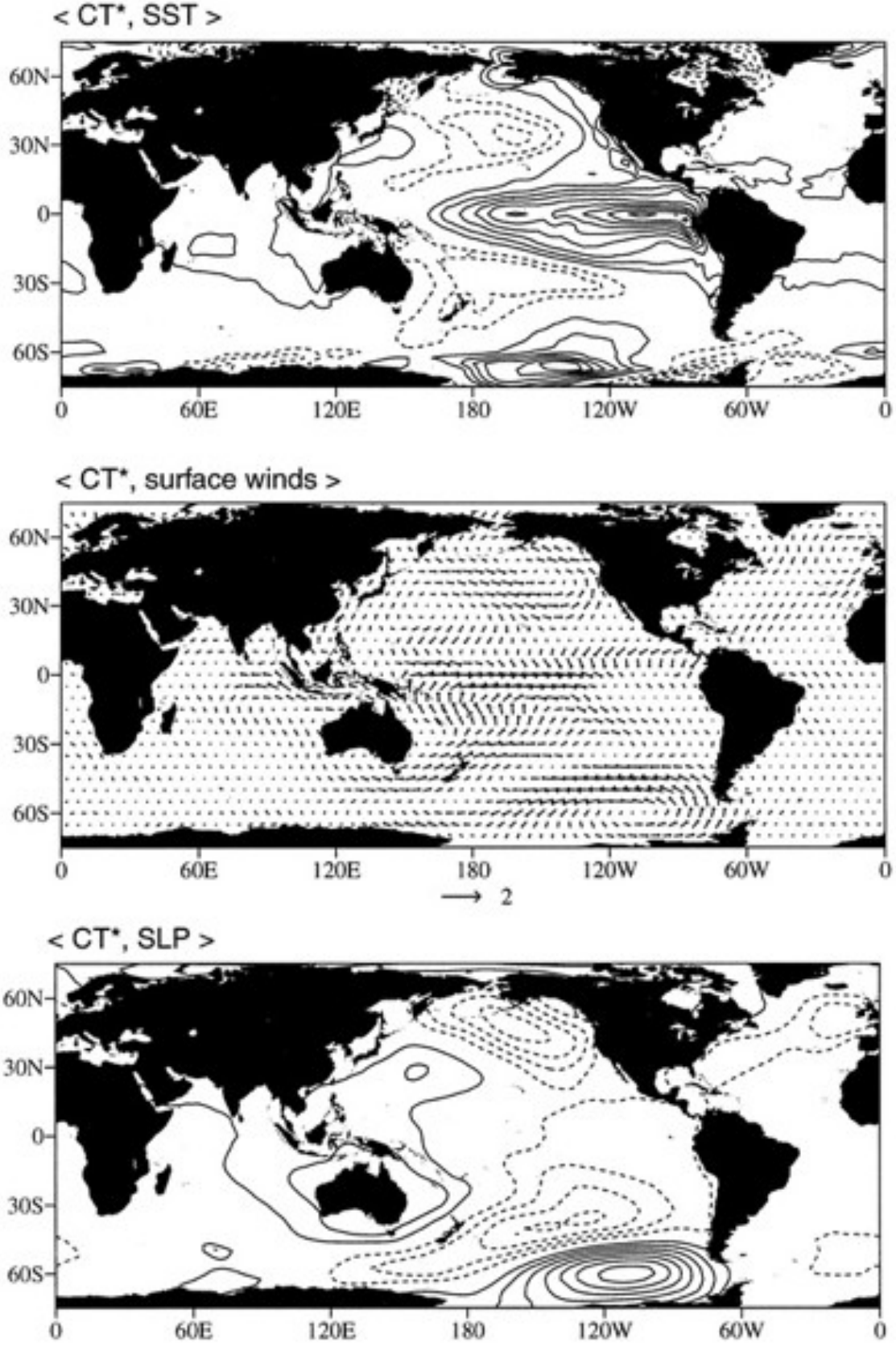

< CT*, SST >
60N
30N
0
30S
60S
0
60E
120E
180
120W
60W
0
< CT*, surface winds >
60N
30N
0
30S
60S
0
60E
120E
180
120W
60W
0
⟶ 2
< CT*, SLP >
60N
30N
0
30S
60S
0
60E
120E
180
120W
60W
0

Figure 1.4 Global reanalyzed fields regressed upon CT*. (top) Sea surface temperature (SST). Contour interval is 0.1 K per standard deviation of CT* (std dev)$^{-1}$. The zero contour is omitted and negative contours are dashed. (middle) Surface winds. The reference vector is 1.5 m s^{-1} (std dev)-1. (bottom) Sea level pressure (SLP). Contour interval is 0.25 hPa (std dev)$^{-1}$. The zero contour is omitted and negative contours are dashed. Regressions are based on all calendar months, from Jan 1958 to Dec 1993 (Garreaud and Battisti 1999).

autumn and fades towards summer. It was found that during La Niña-like conditions, these midlatitude anomalies are positive with a lesser-pronounced negative summer-anomaly during an El Niño-like state (Garreaud and Battisti 1999).

ENSO affects synoptic conditions in the Australian region in numerous ways. Evans and Allan (1992) found that during El Niño the Australian monsoon trough displaced equatorward with a convection maximum northwest of Australia. During La Niña, the monsoon trough is stronger and more coherent while monsoonal activity over northern Australia is enhanced. Simultaneously, an equatorial convergence zone is active over the Southeast Asian region, which during El Niño, is located over the West Pacific. The impact of ENSO on Australian precipitation has long been known (e.g., Nicholls 1979a, 1984a,b; Streten 1981, 1983; Wright 1984; Meehl 1987; Hackert and Hastenrath 1987). Pittock (1975) was the first to find strong correlations

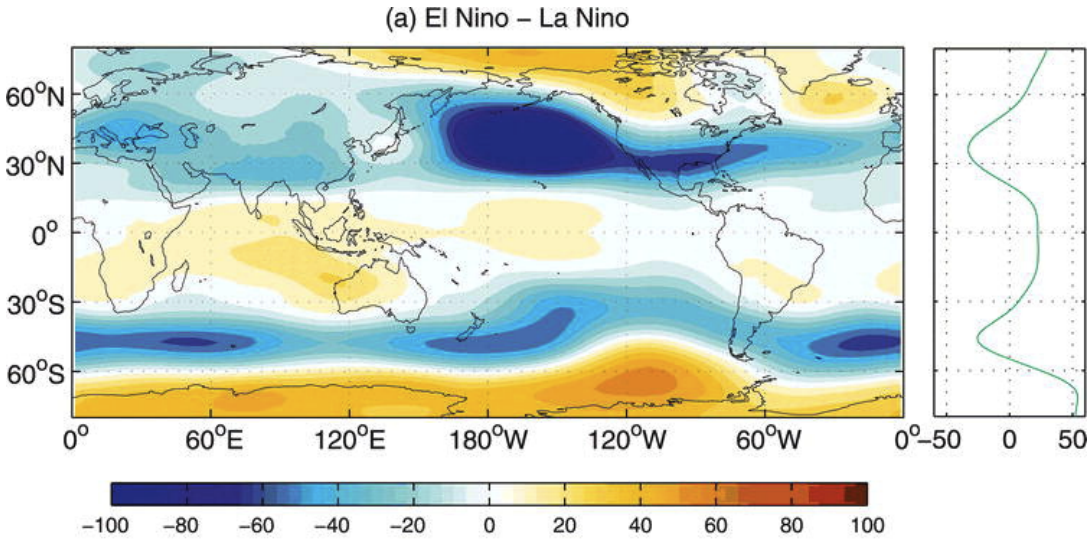


Figure 1.5 DJF 500 hPa geopotential height field (m) for El Niño–La Niña composite. The right side panel is the zonal mean of the corresponding field (Lu et al. 2008).

between SOI and precipitation in North and East Australia. Later, McBride and Nicholls (1983) investigated lagged correlations between Australian rainfall and SOI and demonstrated that the most persistent and strong correlations were to be found during winter and spring. Nicholls (1984c) stipulated that SST anomalies north of Australia were highly correlated with both ENSO and East Australian rainfall. From very early on, the importance of ENSO for Australian precipitation rate and distribution was understood to be related to the seasonal changes to the tropical Pacific Warm Pool and associated changes in ocean-atmosphere coupling over Australasia (e.g., Nicholls 1984a,b; Hackert and Hastenrath 1986). Additionally, the temperature variations over Australia were shown to be strongly related to ENSO (e.g., Coughlan 1979; Kiladis and Diaz 1989; Jones 1991; Halpert and Ropelewski 1992; Lough 1995; Jones 1999) with El Niño conditions leading to higher than normal temperatures and La Niña conditions to negative anomalies. However, there is a well-known covariability of rainfall and surface temperature with a negative relationship between rainfall and maximum temperatures (Deacon 1953; Jones 1991; Lough 1995; Nicholls et al. 1997; Power et al. 1998; Jones 1999). Model results link this relationship to local changes in the latent heat flux and changes in the radiation budget (Simmonds and Lynch 1992; Watterson 1997; Power et al. 1998).

1.1.2 The Indian Ocean Dipole

In the Indian Ocean a basin-scale internal variability exists with anomalously low sea surface temperatures off Sumatra and high sea surface temperatures in the western Indian Ocean, known as the Indian Ocean Dipole IOD (Saji et al. 1999, Webster et al. 1999). An IOD event is accompanied by large-scale wind, moisture transport and precipitation anomalies (Saji et al. 1999, Ummenhofer et al. 2009). The SST mode and its evolution from early austral winter to early austral summer is shown in Figure 1.6. The dipole intensifies rapidly over the austral summer and usually peaks around September/October with a short phase of rapid decaying thereafter (Yamagata et al. 2004). The dipole behaviour is so robust that an SST gradient index can be built between a western pole (50°-70°E, 10°N-10°S) and an eastern pole (90°E-110°E, 0-10°S). This index is highly correlated (>0.7) with the 2nd EOF in the Indian Ocean, which accounts for 12% of the SST variations (Saji et al.

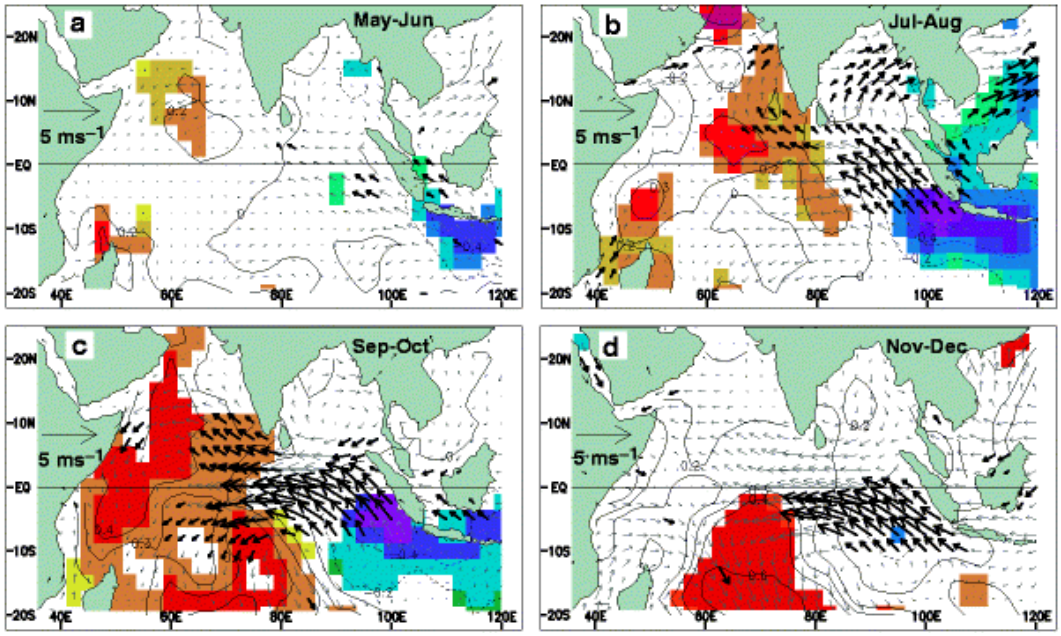


Figure 1.6 A composite dipole mode event. a-d) Evolution of composite SST and surface wind anomalies from May-June a) to Nov-Dec. The statistical significance of the analysed anomalies were estimated by the two tailed t-test. Anomalies of SSTs and winds exceeding 90% significance are indicated by shading and bold arrows, respectively (Saji et al. 1999).

1999). A positive dipole event is when anomalously cool SSTs occur in the eastern pole, while the western pole appears anomalous warm. The zonal equatorial winds in the eastern and central Indian Ocean are highly correlated with the SSTA pattern and correspond to easterlies during the positive IOD event, blowing from cooler to warmer waters. There is some co-occurrence with ENSO events observed with positive IOD events tending to occur during the same year as El Niño and negative IOD events during La Niña years (Yamagata et al. 2004).

Normally, in the Indian Ocean basin, southeast trade winds converge into the South equatorial trough and cause high precipitation in the Southeast Asian region. In Figure 1.7 composites of 6 positive events are presented: 1961, 1967, 1972, 1982, 1994 and 1997. In the early austral summer, the first cold SST anomalies in the tropical East Pacific can be observed with the convection in that area weakening. This leads to a higher surface pressure and a subsequent extension and shift of the southeast trade winds. During the austral winter the cold anomalies intensify and extend towards the equator and further westward so that more moisture is transported to the downstream end of the trade winds. At the same time the western Indian Ocean develops anomalous warm SSTs around the equator and zonal easterly winds pick up.

Rapid decay of the signal begins after the peak of the mode around September/October, (Saji et al. 1999). It has been shown that upwelling in the Java–Sumatra region is an essential controlling factor in the growth of IOD anomalies (Meyers et al. 2007). Prior to this, the thermocline experiences large vertical displacements beneath both poles of the dipole, and the displacements are correlated to the local SST anomalies (Rao et al. 2002; Xie et al. 2002; Feng and Meyers 2003).

There is much debate over the extent to which the Indian Ocean Dipole mode, is independent of ENSO (e.g., Saji et al. 1999; Allan et al. 2001; Ashok et al. 2003; Meyers et al. 2007; Risbey et al. 2009). Webster et al. (1999) initially showed that both signals appear to be independent despite the strong positive IOD event in 1997/98 co-occurring with a strong El Niño in that year. To investigate the dependency of IOD on ENSO, various studies have explored methods to separate the pure IOD mode from ENSO, or otherwise investigated the effect of the IOD on key climate variables over Australia such as precipitation and temperature (e.g., Ashok et al. 2003; Saji et al. 2005; Meyers et al. 2007; Risbey et al. 2009; Ummenhofer et al. 2009). For example, Meyers et al. (2007) applied a lagged empirical orthogonal function (EOF) analysis to remove the direct or lagged effects of ENSO from the IOD. This approach has since been applied in other studies (Risbey et al. 2009; Ummenhofer et al. 2009), although it only removes ENSO at a single lag from the IOD. Table 1.1 shows recent IOD and ENSO events as classified by Meyers et al. (2007). An ENSO year is defined as exceeding one standard deviation for two or more consecutive months between and including June and February of the following year from the derived lagged ENSO EOF time series. A year is classified as an IOD year if the time series derived by Meyers et al. (2007) exceeds one standard deviation for two or more consecutive months between and including June and December. The IOD time series are corrected between 1954-1960 due to a climate anomaly and both time series are shifted to compensate for ocean warming effects. The detailed description of the IOD time series and the corrections can be found in Meyers et al. (2007). Other studies have used partial correlations to remove the direct ENSO and IOD effects from Australian temperature and rainfall data (Ashok et al. 2003; Saji et al. 2005; Risbey et al. 2009). In short, these studies have been able to demonstrate some degree of independence of the IOD from ENSO regarding the timing of Australian region precipitation rate variability, despite the difficulties in removing the

more complete lag/lead effects of ENSO.

The effect of the IOD on Australian climate has been the subject of a series of studies (e.g., Saji et al. 1999; Saji and Yamagata 2003; Ashok et al. 2003, Meyers et al. 2007; Ummenhofer et al. 2009, Risbey et al. 2007). They all found a particularly strong effect on the South and Southeast Australian winter rainfall with drought conditions during positive IOD events. The large-scale effects can be associated with a baroclinic response to the low-level anticyclonic circulation over the eastern and central Indian Ocean. This results in anomalous subsidence and therefore reduced precipitation over the southern parts of the Australian continent (Ashok et al. 2003). Meyers et al. (2007) related the low rainfall rate during positive IOD events to changes in the Northwest Cloud Band over Australia (e.g., Tapp and Berell 1984). Recent studies concentrated on IOD induced changes in the large-scale variability of

TABLE 1.1 Classification of years when El Niño or La Niña and/or positive or negative Indian Ocean dipole occurred. Boldface (lightface) indicates a higher (lower) level of certainty in the classification as explained in the text. The classification is given lower certainty if either the ENSO phenomenon or the IOD phenomenon is not clear (Meyers et al 2007).

	Negative IOD	**No event**	**Positive IOD**
El Niño	**1930**	**1877, 1888, 1899, 1911, 1914**, 1918, 1925, **1940,** 1941, **1965, 1986, 1987**	**1896, 1902, 1905, 1923,** 1957, **1963, 1972, 1982, 1991, 1997**
No event	**1880, 1958, 1968, 1974, 1980, 1985, 1989, 1992**	1881, **1882, 1883, 1884, 1895, 1898, 1901, 1904, 1907, 1908, 1912,** 1915, 1920, **1921, 1927, 1929,** 1931, **1932, 1934, 1936, 1937, 1939, 1943, 1947, 1948,** 1951, 1952, **1953,** 1956, 1959, 1960, **1962, 1966, 1969, 1976, 1979, 1990,** 1993, **1995**	**1885, 1887, 1891, 1894,** 1900, 1913, **1919, 1926, 1935, 1944, 1945, 1946, 1961, 1967,** 1977, **1983,** 1994
La Niña	**1906, 1909, 1910,** 1916**, 1917, 1928, 1933, 1942, 1950, 1975,** 1981	**1878,** 1879, 1886, 1889, **1890, 1892, 1893, 1897, 1903,** 1922, 1924, **1938, 1949,** 1954, 1955, **1964,** 1970, 1971, **1973,** 1978, 1984, **1988,** 1996, 1998	

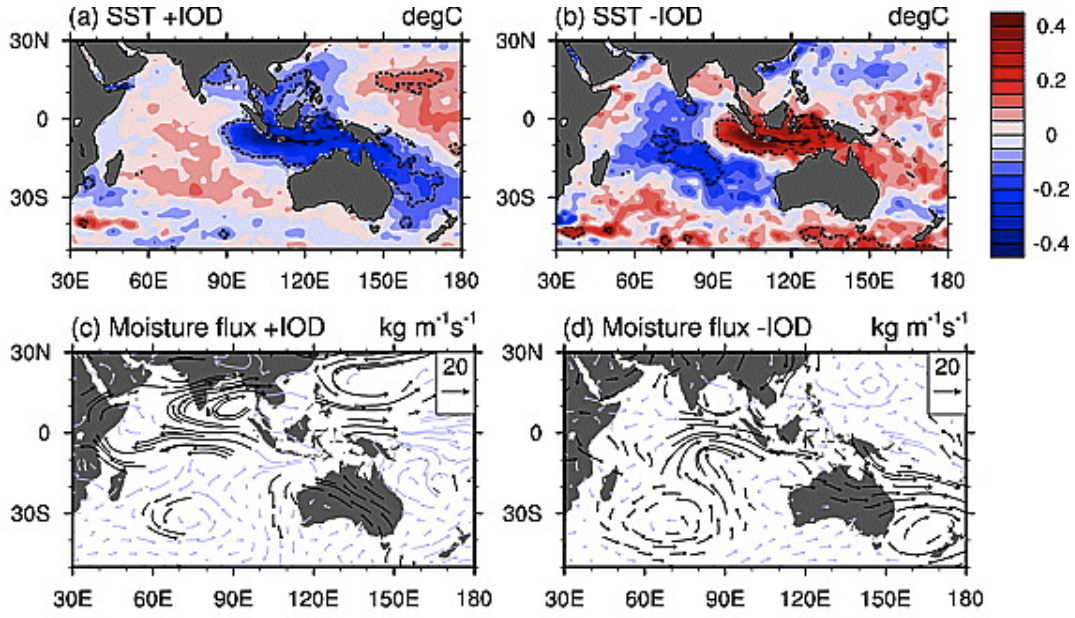


Figure 1.7 Characteristic climate conditions during pure IOD years. Composite of anomalies in (a, b) SST (°C) and (c, d) moisture flux (kg m−1s−1) June–October months during pure positive and negative IOD years (Ummenhofer et al. 2008).

the atmospheric thickness between 1000 and 500 hPa as response to anomalous SSTs (Risbey et al. 2009) and moisture flow integrated to 500 hPa (Ummenhofer et al. 2009). Both studies confirmed the strong effects of the IOD and Southeast Australian precipitation rate and related drought periods in this area to positive IOD events. Figure 1.7 shows the composites of SSTs and the corresponding moisture flux and precipitation rate over the Australian continent during positive and negative IOD events. All these studies confirmed a strong sensitivity of the atmosphere over Australia due to large-scale SST changes in both, the South Pacific and the Indian Oceans.

1.2 Tropical Cyclone Development

1.2.1 Tropical Cyclone Structure

In the Southern Hemisphere, tropical cyclones (TCs) are clockwise rotating tropical storm systems exceeding a predefined minimum central pressure and

Table 1.2 Tropical cyclone categories and corresponding values of approximate average maximum wind speeds and central pressures, modified for the Australian region from the Saffir-Simpson scale (Simpson 1974; Dare and Davidson 2004).

Category	Avg max wind (km h^{-1})	Central pressure (hPa)
1	63-90	>985
2	90-125	985-970
3	125-165	970-955
4	165-225	955-930
5	>225	<930

maximum wind speeds. Since 1973, the intensity of these storms is traditionally classified in 5 categories distinguished by the intensities of their sustained winds by the Saffir-Simpson-Scale (Simpson 1974). The Australian region utilises a more contemporary intensity scale for all TCs of gale intensity and higher, with 10-minutes average mean surface winds (MSW) exceeding 34 kts (63 km h^{-1}). Severe TCs are then defined with MSWs exceeding 65 kts (125 km h^{-1}; Table 1.2).

The system is divided into three parts, the eye, the eyewall and the widespread rainbands. The eye can be identified by a dry adiabatic downstream flow, in which convection is suppressed. The strongest winds can be found in the eyewall near the surface, where the highest pressure gradients towards the storm centre can be found. The radius of maximum winds (RMW) varies strongly from system to system and between the different TC basins from 10-100 km with an increase of the radius with height. Inside the RMW an approximate solid-body rotation can be observed, while outside the RMW the winds successively weaken with increasing radius following the r^x-law. The cloud and rainbands around the eyewall are a common structure, organised as spiral bands with approximately 10 km width and a height between 3 and 15 km (Emanuel 1991). The energy source of a mature TC is the thermodynamic disequilibrium between the lower atmosphere and the oceans (Kleinschmidt 1951). This disequilibrium is not reflected in an actual temperature difference between the lower troposphere and the ocean, but in the undersaturation of surface-near air. Emanuel (1986) described an idealized TC as a Carnot machine, in which heat gets converted into mechanical energy. To achieve a more thermodynamic equilibrium of the ocean and lower troposphere, surface air flows cyclonically (Southern Hemisphere (SH)) towards the boundary layer of the storm center. There it rises close to the moist adiabatic state into the mid- and high troposphere (Emanuel 1991). Therefore

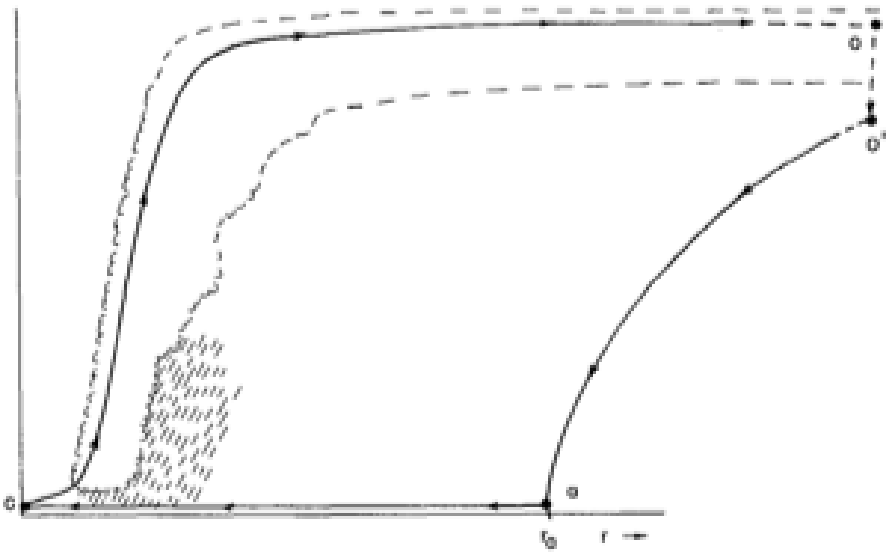


Figure 1.8 The hurricane Carnot cycle. Air begins spiraling in toward the storm center at point *a*, acquiring entropy from the ocean surface at fixed temperature T_S. It then ascends adiabatically from point *c*, flowing out near the storm top to some large radius, denoted symbolically by point *o*. The excess entropy is lost by export or by electromagnetic radiation to space between *o* and *o'* at a much lower temperature T_0. The cycle is closed by integrating along an absolute vortex line between *o'* and *a*. The curves *c*-*o* and *o'*-*a* also represent surfaces of constant absolute angular momentum about the storm's axis (Emanuel 1991).

increased surface winds lead to increased heat transfer from the ocean and to a subsequent intensification and so on. The idealised process is shown in Figure 1.8.

1.2.2 Tropical Cyclogenesis

Tropical cyclone (TC) formation only occurs when certain atmospheric and oceanic conditions are present. Gray (1968) defined basic favourable environmental conditions for tropical cyclone formation including sea surface temperature (SST), conditional instability in the lower to mid-troposphere, absolute vorticity in the lower troposphere, mid-troposphere relative humidity, divergence in the upper troposphere, and vertical shear of the horizontal winds between the lower and upper troposphere. These parameters were further enhanced in several studies (e.g., Gray 1975, 1992, 1993; McBride and Zehr 1981), but the general dynamic and thermal requirements on the state of the atmosphere and ocean remains the same.

Dynamic parameters include (i) environmental vertical wind shear (EVWS), (ii) low-level relative vorticity and (iii) planetary vorticity and the thermal parameters, consist of (iv) mid-tropospheric relative humidity, (v) ocean thermal energy and (vi)

equivalent potential temperature gradient. All six conditions need to be met to allow the development of a TC. (i) It is imperative that the EVWS between the lower and upper atmosphere is weak, otherwise the convective system of the eyewall cannot develop or persist. The critical value for the magnitude of EVWS is 10 ms^{-1} (Landsea 2000). (ii) A pre-existing negative (SH) low-level relative vorticity anomaly, like a small anticyclonic atmospheric disturbance, a tropical wave or a monsoonal trough with convergence provides the initial conditions to develop a TC. The attendant upper-troposphere divergence supports the development of a deep convection, which in turn intensifies the disturbance into a low-pressure system (Cheung 2004). (iii) The Coriolisparameter f is the measure for the planetary vorticity and is strongly dependent on the latitude. Up to a distance of about 500 km from the equator (< 5°S), the Coriolis force is not strong enough to keep a system rotating. The Coriolis effect causes rotating systems to turn polewards in the absence of strong steering currents. The westerly winds on the equatorward portion of the cyclone pull slightly to lower latitudes, but as the Coriolis effect weakens towards the equator, the net drag on the cyclone is poleward. Thus, TCs in the SH usually turn south, before being blown east, when no other effects counteract the Coriolis effect. (iv) Relatively moist layers in the lower- and mid-troposphere (relative humidity at 500 hPa) are essential, as dry mid levels suppress the continuing development of widespread deep convection. Further, a dry lower troposphere cools down due to evaporation, which leads to a slight descent (Cheung 2004). (v) The Ocean Heat Content (OHC) given by the area of sea surface temperature (SST) warmer than 26°C and the depth of the 26°C isotherm (Shay et al. 2006) determines the amount of warm water supporting tropical evaporation. A minimum of 26.5°C throughout a sufficient depth of at least 50m (Gray 1968) guarantees the required heat-flux into the troposphere. Additionally the evaporation contributes to intensify the low-level disturbance. (vi) The gradient of equivalent potential temperature between 1000 hPa and 500 hPa describes the enthalpy between these two layers and hence the likelihood to form cloud clusters. It is necessary that the atmosphere cools fast enough with height such that it is potentially unstable to moist convection (Gray 1968). However, understanding the physical processes of TC formation during these favourable conditions is still a subject of ongoing research, particularly with respect to the various climate signals, which also influence the conditions.

1.2.3 Tropical Cyclones in the Australian Region

The Australian TC region is usually defined from 90°-160°E and 0-30°S (e.g., Kuleshov and de Hoedt 2003; Ramsay et al. 2008) with greater variability of TC frequency, speed and trajectory than in other TC basins (Bessafi et al. 2002). This is a result of some unique climatological features including the major landmass in the region (Holland 1984), the existence of the Australian monsoon trough (McBride and Keenan 1982, Evans and Allan 1992) and the close approach of the midlatitude westerlies to low latitudes (Dare and Davidson 2004). Dare and Davidson (2004) performed an extensive study on the TC climatology in the Australian region, as well as on its three subregions West (90°-125°E), North (125°-142.3°E) and East (142.3°-160°E) by comparing the characteristics of 500 TCs in the time period from 1963/64-2002/03. The seasons are defined as from June to July in the following year with the highest TC activity observed from December to April with an average of 12.5 storms per season but with significant individual seasonal standard deviations. In general Australian TCs have an average lifetime of 7.5 days, reach its maximum intensity after 3.5 days and originate between 5° and 15°S. The mean lifetime is longer in the Pacific regions, as storms tend to stay over the ocean, while in the Indian Ocean the storms recurve towards the Australian landmass. However, the most intense storms were found over the Indian Ocean, while the storms over the Pacific have faster translation speeds, further steering up after intensification.

Most TCs form in a worldwide band of thunderstorm activity, for example, within the ITCZ or a monsoon trough. The Australian region TC behaviour is characterised by its strong dependence on the monsoon trough (McBride and Keenan 1982; Evans and Allan 1992; Dare and Davidson 2004; Ramsay et al. 2008). The monsoon trough is an area of strong negative low-level vorticity. The upper-level ridge is located around 12°S, accompanied by weak environmental vertical wind shear leading to a zone extremely favourable for TC development and intensification (Dare and Davidson 2004). Additionally Australian TC development is affected by continental effects. If warm, dry continental air in Western Australia extends offshore, it implies extensive dry ambient conditions for Australian west coast TCs. On the other hand, when the monthly average of relative humidity is concentrated in the west

coast region, extremely intense TCs can be expected (Tonkin et al. 1997). During the Southern Hemisphere TC season, Camargo et al. (2007) showed that mid-tropospheric relative humidity and vertical wind shear are important contributors to variations in TCG in the Southern Hemisphere. While investigating ENSO effects on TCG in Australia, Kuleshov et al. (2009) found relative humidity at 500hPa and low-level relative vorticity to be most important for TCG.

The deviation of mean TCs in the Australian region depends strongly on atmospheric and oceanic patterns, primarily ENSO. There have been numerous studies on the influence of ENSO on tropical cyclone formation in the Australian region (e.g. Nicholls 1984d, 1992; Basher and Zheng 1995; Kuleshov and de Hoedt 2003; Kuleshov et al. 2009)**.** Ramsay et al. (2008) presented a comprehensive study on the statistical effects of ENSO on seasonal TCG frequency in the Australian region. There it was shown, that out of all ENSO indices, August-September-October (ASO) averaged NINO4 captured Australian region TCG variability best with a correlation coefficent of -0.73 between 1970-2005. The link between ENSO dynamics

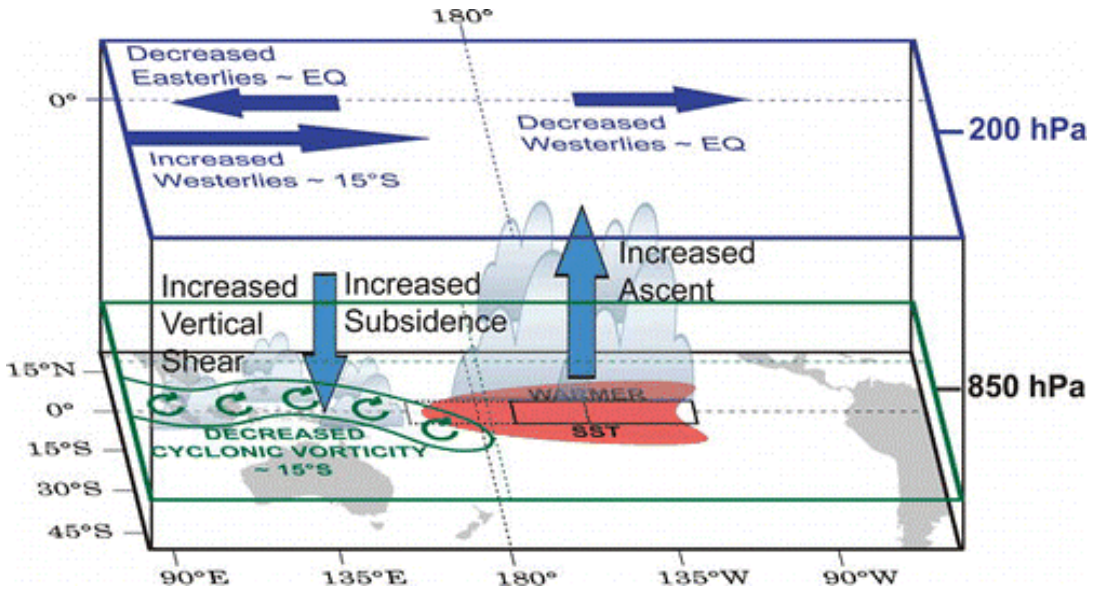


Figure 1.9 Schematic showing the connection between anomalously warm SST in the Niño 3.4 and Niño4 regions (pink shading) associated with El Niño events and the corresponding atmospheric response: increased 200 hPa zonal westerly winds (blue horizontal arrows) around 15°S resulting in increased vertical shear of the zonal wind, co-located with decreased 850 hPa relative cyclonic vorticity (green) associated a weakened monsoon trough over the Australian region. Light blue vertical arrows indicate anomalous subsidence over the northern Australian region and anomalous ascent over the central equatorial Pacific (Ramsay et al. 2008).

and TCG occurrence is the strong relationship between the Australian monsoon trough and ENSO through atmospheric bridge processes (Figure 1.9; Evans and Allan 1992; Ramsay et al. 2008). Nicholls (1992) also found correlations of r = 0.72 between September-October-November (SON) values of the Southern Oscillation Index and the 'first differences' of consecutive seasons instead of the total number of TC counts between 1959/60 and 1990/91. Liu and Chan (2010) identified ENSO-related strong correlations between Australian region TCG count and the July-August-September (JAS) averaged trade wind index (5°N-5°S, 135°-180°E) of r = 0.68 and the ASO averaged equatorial longwave outgoing radiation between 160°E-160°W with r = 0.60. During la Niña events an increase of TC numbers east of 70°E has been observed (Kuleshov and de Hoedt 2003) with the maximum TC frequency in the Southern Hemisphere occurring at the end of January during La Niña years and about one month later during El Niño years. Also the onset of the South Indian Ocean TC season occurs one month earlier during La Niña years. The distribution of TCG shifts eastward during warm ENSO phases (Evans and Allan 1992; Basher and Zeng 1995; Sinclair et al. 1997; Kuleshov and de Hoedt 2003; Camargo et al. 2007). This leads to increased TC formation in the central Pacific during El Niño years. In contrast, during La Niña events TCG shifts westward resulting in a higher likelihood of TCs making landfall along the Queensland coastline (Evans and Allan 1992). Also increased activity can then be observed over the Southeast Indian Ocean (Sinclair et al. 1997; Kuleshov et al. 2008). Figure 1.9 highlights the interaction between central tropical Pacific SST anomalies (here Niño 3.4 region) and the atmospheric response and modulation of TC development conditions in the Australian TC region. Next to the strong correlations of direct ENSO indices, an early study by Nicholls (1984) showed correlations of r = 0.78 between SON SST in a region north of Australia (5°–15°S, 120°–160°E) and Australian region TC counts from 1964-1982. However this relationship has been shown to degrade over time and is not robust for the more recent years (Ramsay et al. 2008). Liu and Chan (2010) also showed a strong correlation between Australian region TCG and the September-October averaged DMI of 0.60.

The relationship of western Australian TCs with ENSO is weaker than the one for the whole Australian region (Broadbridge and Hanstrum 1998; Goebbert and Leslie 2010). However, Liu and Chan (2010) also found significant relationships between western Australian (90°-135°E) TCG and NINO4, the trade wind index, the

OLR index and DMI. In contrast, due to the weak ENSO correlations Goebbert and Leslie (2010) suggest indices based on persistent and significant relationships with spatial atmospheric variables to better explain TCG variations in the Southeast Indian Ocean basin. They found the strongest correlations between variations in northwestern Australian TCG (105°-135°E) and April-May-June averaged 700 hPa geopotential height over North America and JJA 925 hPa geopotential height over the South Atlantic Ocean basin.

Finally, the tracks of TCs in the Australian region are also affected by ENSO. During El Niño, the westerly zonal component is dominant, which leads to more landfalling storms in northwestern and northern Australia, while the landfall rate on the Queensland coast is reduced. Westerly flow is weaker and less coherent during la Niña and the westerly zonal motion is not observed. Then the dominant path is along the east Australian coast and TCs persist further south in both, the Indian Ocean and the Pacific (Figure 1.5; Evans and Allan 1992) basins.

1.3 Statistical Forecast Modelling of Tropical Cyclones

1.3.1 Statistical Seasonal Forecasting

Statistical modelling of variations of TCG counts is a common approach of developing seasonal forecast models. As we have seen in the previous section, the development of TCs is strongly dependent on the climatic environment and this in turn is modulated by intra-seasonal to inter-decadal climate modes and signals, but also by long-term changes of the climate. The advantage of using climate variations prior to the onset of a TC season is at the same time its limitation. A detailed knowledge of the climate signals itself, as well as how they modulate the atmospheric and oceanic conditions for TC development is essential to make the most use of the statistical approach. Another limitation is the fact that TC development is a rather random event. Even though the conditions may be favourable on a large-scale, the actual TC development is dependent on more than just purely seasonal climatic variations, but also on local conditions.

Since the early 1980s a number of statistical seasonal forecast schemes have been developed and improved to predict TC activity in various basins and sub-basins (Klotzbach et al. 2010). Seasonal forecast modelling of TC activity was first undertaken by Nicholls (1979a) for the Australian region and Gray (1984) for the North Atlantic. In later studies by Gray et al. (1992, 1994) climatic relationships with hurricane activity in the North Atlantic are based on metrics, such as the Quasi-Biennial Oscillation and African rainfall. A relationship between intense hurricanes and the Sahel monsoon rainfall was also established (Landsea and Gray 1992). The skill of Gray's operational Atlantic seasonal TC forecasts for the analyzed period from 1984–2001 relative to climatology and persistence was confirmed and improved (e.g., Owens and Landsea 2003; Saunders and Lea 2005; Klotzbach 2007). Other relevant North Atlantic statistical forecasts include model predictions of hurricane counts using Poisson regression models (e.g., Elsner and Schmertmann 1993; Lehmiller et al. 1997). The Poisson method was later extended using a Bayesian approach to investigate seasonal TC counts and landfall over the USA (e.g., Elsner and Jagger 2004, 2006). This approach was also used most recently to improve multi-season forecasting of Atlantic hurricane activity (Elsner et al. 2008), and seasonal forecasting of TCs affecting the Fiji, Samoa and Tonga regions (Chand et al. 2010) and the central North Pacific (Chu and Zhao 2007). In the northwest Pacific and Australian region, projection pursuit regression (PPR) has been used to forecast seasonal TC totals and associated TC predictands (e.g., Chan et al. 1998; Chan and Shi 1999; Chan et al. 2001; Liu and Chan 2010). Also, a statistical scheme based on ENSO related indices has been developed for predicting the annual number of TCs making landfall along the south China coast (Liu and Chan 2003). Most recently, modes from an empirical orthogonal analysis of climate factors have been used as predictors of TC behavior in a statistical model also for the South China region (Goh and Chan 2010).

One major difficulty with developing seasonal tropical cyclone prediction equations for other basins lies in the lack of good data. Except for the North Atlantic and the western North Pacific, direct measurements of tropical cyclone intensity (e.g., by aircraft reconnaissance) are not available and the resulting data base has to be considered unreliable (Holland, 1981). The use of satellite estimates since the 1970s has improved the quality enormously, but questions remain on the reliability of

analyses of intense tropical cyclones.

1.3.2 Statistical Seasonal Forecasting in the Australian Region

For the Australian region, Nicholls (1979b) showed that the austral winter to spring anomalies of sea level pressure at Darwin are highly correlated with early season Australian region tropical cyclone activity, and to a lesser extent with total seasonal TC activity. Subsequent research and operational testing, confirmed the strong link with the ENSO-related Southern Oscillation index (SOI; Nicholls 1984d, 1985a, 1992; Drosdowsky and Woodcock 1991; Ready and Woodcock, 1992). Solow and Nicholls (1990) presented the first Poisson regression based statistical forecast model for the Australian region. They used the SOI as a predictor of Australian region total TC counts. More recently, a Poisson regression model using the September lead saturated equivalent potential temperature gradient between 1000hPa and 500hPa and SOI was developed to forecast upcoming season TCG totals across the Australian region (McDonnell and Holbrook 2004a,b). This model has also been applied to forecast subregional TCG totals in the eastern Indian Ocean, Northern Australia and southwest Pacific regions (e.g., McDonnell et al. 2006). A more recent study of Liu and Chan (2010) presented a seasonal forecast model for the Australian region based on a PPR using generally known predictors such as the NINO4, trade wind index, DMI and the outgoing long-wave radiation index from the U.S. Climate Prediction Centre. For the validation of the model results, the jackknife method is used. The model skill is then measured as the improvement of the RMSE over the RMSE calculated for the climatology. This skill score was 51% for Australian region (90°-160°E), 39% for Western region (90°-135°E) and 37% for Eastern region (135°-160°E) TCG counts. For seasonal TC variations in the northwestern Australian region (105°-135°E), a preliminary linear multiple regression model result was presented by Goebbert and Leslie (2010). They pointed out the different climatology for this particular region only showing weak correlations with commonly known climate indices. As predictors they suggest derived indices of geopotential height at 1000hPa and 850hPa, as well as Southern Hemisphere meridional winds at 100hPa. Using a skill score based on the improvement of MSE over MSE of climatology they achieved a score of 64%. On intra-seasonal time scales, Leroy and Wheeler (2008) developed a

logistic regression model for TC development in the Australian region. As predictors, they used the two dominant varimax rotated SSTA modes for the Indo-Pacific region, as well as an index describing variations in the Madden-Julian Oscillation.

Despite all these model studies, only three models have been used, at least in an experimental capacity, for operational forecasts of Australian region TCG counts and its distribution (McDonnell and Holbrook 2004a,b; Liu and Chan 2010). A separate model, based on the ENSO indices Southern Oscillation index (SOI) and NINO3.4, is in use at the Australian Bureau of Meteorology (BoM 2011: http://www.bom.gov.au/climate/ahead/tc.shtml). Although the McDonnell and Holbrook (2004a,b) model is the only one that has made serious attempts to capture the spatial distribution of TCGs, a spatial bias has restricted its effective use for operational spatial probabilistic predictions. In recent years, this model has not been used further in experimental forecasts, albeit that its skill in forecasting annual aggregated counts is reasonably good. The Liu and Chan (2010) and BoM models failed in forecasting correctly the only available reference forecast of the previous Australian TC season of 2010/11.

1.4 Thesis Objectives and Structure

As seen in the review above, to investigate the dependency of the IOD on ENSO, various studies have explored methods to separate the pure IOD mode from ENSO, or otherwise investigated the effect of the IOD on key climate variables over Australia such as precipitation and temperature (e.g., Ashok et al. 2003; Saji et al. 2005; Meyers et al. 2007; Risbey et al. 2009; Ummenhofer et al. 2009). Even though these studies have been able to demonstrate some degree of independence of the IOD from ENSO regarding the timing of Australian region precipitation rate variability, they had difficulties in removing the more complete lag/lead effects from ENSO. Also, the more advanced methods (Meyers et al. 2007; Ummenhofer et al. 2009) are quite laborious to repeat. A more recent study (Liu and Chan 2010) combined various ENSO indices with the IOD index as predictors in a seasonal forecast model. However, the quantification if, or of how strong the DMI improves an ENSO-only forecast, had not been investigated.

In the past three decades, seasonal TC forecasting was often limited to MLR or Poisson regression models using ENSO indices only (e.g., Nicholls 1979a, 1984d, 1985a, 1992; Solow and Nicholls 1990; Drosdowsky and Woodcock 1991; Ready and Woodcock, 1992; McDonnell and Holbrook 2004a,b; Goebbert and Leslie 2010). Liu and Chan (2010) recently applied a project-pursuit regression (PPR) technique, including the DMI index as predictor. Bayesian model approaches, as used in various other TC basins (e.g., Elsner and Jagger 2004, 2006; Elsner et al. 2008; Chu and Zhao 2007; Chand et al. 2010), have not been applied to the Australian region yet. Also, forecasting the spatial distribution of TCs in the Australian region is a very sparse investigated field in the literature. Only McDonnell and Holbrook (2004a,b) applied a Poisson regression model approach on a 5°longitude x 2°latitude grid, but had limited success regarding grid-scale forecasting.

The main aim of this thesis is to improve our understanding of how climatic signals and variables affect TC occurrences observed in the Australian region. Of special interest is how to quantify the role of interannual large-scale climate signals such as ENSO and IOD in variations of TCG counts. Also, our belief is that the use of indices based purely on ENSO or IOD dynamics is limited. A particular goal was then to improve existing seasonal statistical forecast models to evaluate the TC hazard for the Australian region for the upcoming season. On the basis of the analysed climatic signals and variables, a statistical seasonal TC forecast model for TC counts, and also for the spatial probability of TCs forming was developed. The main aims are as follows:

1. Develop a simple, but effective method to separate the IOD from ENSO to estimate the statistical independency of IOD from ENSO.

2. Quantify the individual IOD and with ENSO measures combined forecast skill of the original IOD signal and the ENSO-independent IOD predictor of seasonal variations in TCG for the Australian region and subregions.

3. Identify important prediction schemes of climate variables for seasonal variations of TCG counts in the Australian region and subregions.

4. Develop a skilful seasonal forecast model for TCG counts in the Australian region and subregions with a Bayesian Poisson regression model.

5. Develop a skilful seasonal forecast model for spatial probabilities of TCG in the Australian region with a Bayesian logistic regression model.

The rest of the thesis is organised as follows. Chapter 2 gives a short review of the statistical methods and tools used in this study. Chapter 3 is presented in the form of a paper currently under review at the Journal *Climate Dynamics (under minor revisions)* that addresses whether the Indian Ocean Dipole plays a role as a potential predictor for seasonal TCG counts in the Australian region. For this reason the ENSO-independent DMI component was extracted with a lagged regression method and compared to the contributions of the original time series. Finally a multiple linear regression model was applied to investigate the roles of ENSO and IOD alone, but also in combination with each other as potential seasonal forecast predictors. Chapter 4 comprises an accepted paper from the *Journal of Climate* that exploits a more sophisticated, stochastically advanced, Bayesian seasonal forecast model for TCG counts in the Australian region. The model uses a Poisson regression and incorporates predictor indices extracted from analysed climate variables. The data is validated on a 40-year observational data set and presents significant improvements over existing statistical seasonal forecast models for that region. Also presented is a paper being prepared for submission to the *Geophysical Research Letters* introducing a separate model for seasonal forecasting of Southeast Indian Ocean TCG counts. Chapter 5 investigates the potential to forecast seasonal variations via spatial probability of TCG occurring on the basis of a Bayesian logistic regression model. This chapter is a manuscript in preparation for the Journal *Climate Dynamics*. Chapter 6 discusses the results with respect to the literature and applies the developed forecast models operationally to the last three seasons, which are all outside of the training data set. Finally some of preliminary results and future research is outline before we summarize the key achievements of this study.

2 STATISTICAL TOOLS

Understanding the climate system, its changes and effects on different time- and space scales is a complex research field. Nevertheless, statistical techniques provide very useful tools for investigating climate system relationships and possible interactions between variables, such as temperature, precipitation and storm frequency. In the present thesis, interannual variations in different large-scale climate parameters across the Indo-Pacific are used to build a forecast system for Australian region tropical cyclone genesis (TCG). As this thesis has a strong statistical focus, some of the standard expressions and regression methods are explained here. A basic introduction to machine learning is also presented.

2.1 Basic Statistical Expressions

2.1.1 Time Series Analysis

First, various standard techniques are used in this thesis to estimate the scale of a data set and its relative variation from the mean. The standard deviation has the same physical dimensions as the underlying data and is defined as

$$\sigma = \left[\frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \bar{x} \right)^2 \right]^{\frac{1}{2}} , \qquad (2.1)$$

where n is the sample size, x_i is the ith data point, and $\bar{x}$ is the mean value of the n data points. The squaring process in equation (2.1) results in a special weight on data points far away from the data mean as the difference. The standard deviation is an integrated part in a set of statistical expressions. The square of the data standard deviation, σ^2, is the sample variance of a time series or data set (Hsieh 2009). The variance can also be used to quantify an extracted or decomposed fraction of a time series against its background signal.

When measuring the variance of two time series x and y together, the covariance is calculated, which is defined as

$$Cov(x,y) = \frac{1}{n-1}\sum_{i=1}^{n}\left[(x_i - \bar{x})(y_i - \bar{y})\right] \quad . \tag{2.2}$$

The sample variance is the special case of the covariance, when $x = y$. The correlation coefficient calculates the similarities of the variations within the two time series x and y. In the present study, the Pearson product-moment coefficient of linear correlation is applied. One way – and the way it used in this thesis - is to view the Pearson correlation as the ratio of the sample covariance of the two variables to the product of the two standard deviations

$$r_{xy} = \frac{Cov(x,y)}{\sigma_x \sigma_y} \quad . \tag{2.3}$$

The Pearson product-moment correlation coefficient takes on values between -1 and 1. It is important to note, that nonlinear relationships between the two time series may not be recognised. Also, as the Pearson correlation coefficient is essentially a non-dimensionalised covariance, it can be extremely sensitive to one or a few outlying point pairs (Wilks 1995). To evaluate if a correlation is significant, e.g., different from 0, a null hypothesis test can be performed. A common approach is the Student's t-test, in which t is defined as

$$t = r\sqrt{\frac{n-2}{1-r^2}} \quad , \tag{2.4}$$

with $f = n-2$ identifies the number of degrees of freedom, assuming independent sampling across all data values (Student 1907). On basis of t and f, the p-value can be calculated as calculated as

$$p^2 = \frac{t^2}{f + t^2} \quad . \tag{2.5}$$

The confidence level is defined as 1-p. The smaller p is, the higher the significance of the correlation - taking account of the number of degrees of freedom. In geophysical data, such as climate data (e.g., temperature, humidity, ocean salinity, etc), serial correlation in time (time dependency) can be expected. For that purpose, often a serial correlation analysis is applied to address and remove the temporal autocorrelations of the variable with itself to obtain a better estimate of the effective number of degrees of freedom (e.g., Hsieh 2009). The value of n used in the significance tests will then have to be adjusted to represent the effective sample size. To compute the serial correlation, lagged data pairs n-k are integrated into the formula for the Pearson correlation (Formula 2.3) so that

$$r_k = \frac{\sum_{i=1}^{n-k}\left[\left(x_i - \bar{x}_-\right)\left(x_{i+k} - \bar{x}_+\right)\right]}{\left[\sum_{i=1}^{n-k}(x_i - \bar{x}_-)^2 \sum_{i=k+1}^{n}(x_i - \bar{x}_+)^2\right]^{1/2}} \quad , \tag{2.6}$$

where k is the lag autocorrelation and the sample mean of the first $n-k$ values are denoted with the subscript "−", while that of the last $n-k$ values are denoted with the subscript "+" (Wilks 1995).

2.1.2 Validation Techniques

To validate the quality of model hindcasts against the observed data, various validation methods are applied in this thesis. A measure, which is based on the standard deviation between two time series, is the standard error. The standard error *se* is written as

$$se = \frac{\sigma}{\sqrt{n}} \tag{2.7}$$

(Everitt 2003). The standard error does not take into account how good variations within the time series are captured (as the correlation coefficient does), but is a measure of how strong the two time series differ depending on the length of the time series. Similar measures are the mean-squared error (MSE) defined as

$$MSE = \frac{1}{n}\sum_{i-1}^{n}\left(h_i - o_i\right)^2 \quad , \tag{2.8}$$

and the root-MSE (RMSE), which is simply the square-root of the MSE, with h_i being the hindcast for the data point i and o_i the corresponding observation (Hsieh 2009). The RMSE is a commonly used metric for the potential utility of a predictor or predictor combination in a model, with small values indicating good performance. If hindcasts are obtained as a hindcast probability distribution, and not as fixed numbers (i.e., the probability for a number of events occurring), the accuracy of the hindcasts can be estimated by taking the probability distribution into account. In this case, the MSE is defined as

$$MSE = \frac{1}{n}\sum_{i-1}^{n}\sum_{k=0}^{\infty} p_i(k)\left(k - o_i\right)^2 \quad , \tag{2.9}$$

with p_i being the predicted probability that k events occur at the data point i. For comparison purposes, the skill-score (SS) provides an excellent measure. There, the

improvement of the hindcast over the climatology is calculated using the MSE_{Clim} for the climatology average field and MSE for a chosen predictor combination both calculated from equation (2.8). The skill score is then

$$SS = 1 - \frac{\sum_{i=1}^{n} MSE(i)}{\sum_{i-1}^{n} MSE_{C\lim}(i)} \quad . \tag{2.10}$$

2.2 Regression Techniques

Regression is used to estimate the relationship between a dependent variable, y, and one or more independent variables, **x** (Wilks 1995). Moreover, regression is a prediction tool, where the given predictor(s) is/are **x** and y is the response variable, or predictand. There are two essentially different regression methods, linear and nonlinear regression. The linear regression group only contains the single predictor linear regression and the multiple linear regression approach, with two or more predictors – both typically using a least squares fit approach. Nonlinear regression approaches are applied when the predictand is defined in a limited space and /or the residuals are not distributed in a Gaussian way. The fitting, in this case, uses a maximum likelihood estimate approach. The most commonly used nonlinear regressions are the logistic and Poisson regression.

2.2.1 Multiple Linear Regression

The multiple linear regression (MLR) is one of the simplest regression methods, and is particularly useful for quick estimates of predictor skill. The response variable, y, is then regressed on the predictors, **x**, in the form

$$y_i = \beta_0 + \sum_{j=1}^{k} x_{ij}\beta_j + e_i \quad , \tag{2.11}$$

where β_o and β_j are the corresponding regression coefficients, k is the number of predictors, and e_i is the error or residual unaccounted for by the regression. Graphically (or geometrically), β_o is also called the intercept, while β_j are the slopes.

To estimate the optimal coefficients, the least squares method is applied to minimise the error e_i. For that purpose, the sum of the squared errors (SSE) is calculated as

$$SSE = \sum_{i=1}^{n} e_i^2 \quad . \tag{2.12}$$

Additionally, the sum of the errors has to be zero, so that the residual distributions are centred on the predictands (Wilks 1995). A limitation of MLR is the possibility of 'overfitting' the data. That means, even though a large number of possible predictors may be available, not all of them would have a significant effect. This can result in fitting 'noise' into the data and the model performance degrades (Hsieh 2009).

2.2.2 Logistic Regression

The logistic regression model is an appropriate method that can be used to model the probabilities of an event, i.e., when the outcome variable is binary. The difficulty when events are given as integer numbers, rather than probabilities, can be solved with Regression Estimation of Event Probabilities (REEP; Glahn 1985). In REEP, an assumption based on the underlying data set or experience is made, so the predictand takes on values between 0 and 1. One issue associated with this approach is that dependent on the assumption, the resulting forecasts can lie outside of the interval [0 1]. This is the case if the limit of the assumption is close around the range of the data. However, this is usually a marginal problem as the final forecast will not differ strongly from the probability interval and can be approximated towards the limits (Wilks 1995). With REEP, a Binomial distribution can be applied to the observational data as

$$P(Y_i = y) = \left(\frac{n!}{k!(n-k)!} \right) p_i^{\,k} (1 - p_i)^{n-k} \quad , \tag{2.13}$$

where n is the REEP assumption for the total number defined as $p = 1$ and k is the number of observed events. P in 2.13 is obtained from the logistic regression

$$P(Y_i = y \mid x_i, \beta) = \frac{\exp(\mu_i)}{1 + \exp(\mu_i)} \quad , \tag{2.14}$$

where μ_i provides the multiple linear regression estimates on the predictors. The logistic regression is defined for all positive and negative discrete numbers. When the model coefficients are obtained with the approach described above, forecasts can be obtained by using the Bernoulli distribution on the logistic regression. The Bernoulli

distribution is a special case of the Binomial function, in which $n = 1$, and allows us to obtain a result for the probability of a 0 and 1.

2.2.3 Poisson Regression

The Poisson distribution is often used to model the occurrence of rare, discrete events, such as tornado counts and the occurrences of droughts or cold spells (e.g., Wilks 1995). The Poisson distribution also restricts the possible outcomes to non-negative integers, making it ideal for modelling tropical cyclone occurrences (Elsner and Schmertmann 1993). The Poisson distribution is defined as

$$P(Y_i = y) = \frac{\mu_i^{\,y} \exp(-\mu_i)}{y!} \quad , \mathrm{y} = 0,1,2,\ldots,\infty \quad , \tag{2.15}$$

where $y = 0, 1, \ldots, \infty$ are the number of events, and

$$\mu_i = \exp(\beta_o + \sum_j (\beta_j x_{ij})) \quad . \tag{2.16}$$

Here, if Y has a Poisson distribution, the logarithm of the expected number of TCG occurrences, μ, can be modelled as a linear combination of the predictors, x_{ij}, with j being the specified predictor during the season i. β_j is the corresponding Poisson regression coefficient, β_0 is the intercept, and y is the observed TCG count. In a Poisson model, the variance is equal to the mean (μ), and the standard deviation is the square root of μ.

2.3 Machine Learning

2.3.1 Bayesian Inference

Bayes' theorem represents a quantification of uncertainty provided by probabilities. By comparison, in the frequentist approach probabilities are seen in terms of frequencies of random repeatable events (Wilks 1995). In this study, the Bayesian approach is used to predict the seasonal number or spatial probabilies of TCG occurrences. The Bayesian theorem is applied to find the best possible model coefficient representation, and this information is used to predict the seasonal TCG totals. The observed predictor set is denoted as $x_{1:T} = \{x_1,\ldots,x_T\}$ and the corresponding

seasonal number of TCG occurrences, or TCG probability, as $y_{1:T} = \{y_1,\ldots,y_T\}$ during the observations 1:T. We are interested to find the conditional probability of y_{T+1} given x_{T+1} and the model coefficients β, $p(y|x_{T+1},\beta)$. The model coefficients, β, are estimated using the posterior distribution $p(\beta|x_{1:T},y_{1:T})$.

The assumptions about the prior knowledge of the model coefficients are stated before observing the data $x_{1:T}$ and $y_{1:T}$ in the form of a prior probability $p(\beta)$. As we have no, or only little, prior information on the climatic effects of our chosen predictors on TCG occurrences, we chose the conservative way of defining the prior probability of the model coefficients as almost flat priors. The priors are defined as a Gaussian distribution

$$p\left(\beta_j\right) = N(\beta_j \mid \mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(\beta_j - \mu\right)^2\right), \qquad (2.17)$$

with the mean selected as $\mu = 0$ and the standard deviation $\sigma = 100$, with β_j representing the j model coefficients. For the TCG occurrences model, we consider the Poisson distribution

$$p(y_i \mid x_i,\beta) = Poiss(y_i \mid x_i,\beta) \quad . \qquad (2.18)$$

Following Bayes' rule, we get the posterior distribution

$$p(\beta \mid x_{1:T}, y_{1:T}) \propto p(\beta)\prod_{i=1}^{T} p(y_i \mid x_i,\beta) \quad . \qquad (2.19)$$

The posterior probability (Equation 2.19) allows us to take uncertainties into account and predictions are then obtained from

$$p(y_{T+1} \mid x_{T+1},\beta) = \frac{1}{N}\sum_{i=1}^{N} p(y_{T+1} \mid x_{T+1},\beta) \quad , \qquad (2.20)$$

with N being the total number of obtained samples from the posterior distribution. When applying a logistic regression, the likelihood function is replaced as

$$p(y \mid x,\beta) = \left[\frac{\exp(\mu)}{1+\exp(\mu)}\right]^{y}\left[1-\frac{\exp(\mu)}{1+\exp(\mu)}\right]^{1-y} \quad , \qquad (2.21)$$

where the assumption of REEP is included, so the Bernoulli distribution can be applied in the sample estimate of the posterior distribution. Besides these adjustments, the model set-up stays the same.

2.3.2 Markov Chain Monte Carlo

To usefully apply the Bayesian approach, and obtain the appropriate values for the model coefficients, the use of a sampling method like the Markov Chain Monte Carlo (MCMC; Hastings 1970) method is indispensable (Larget and Simon 1999). The MCMC method simulates direct draws from a probability distribution. There, the previous sample values generate randomly the next sample value, which means generating a Markov chain. Unlike previous studies (e.g., Elsner and Jagger 2004, 2006; Chu and Zhao 2007; Chand et al. 2010), which applied the Gibbs sampler via the open source software WinBUGS, we instead used the multivariate slice sampler (Neal 2003), which is a form of auxiliary variable technique (Roberts and Rosenthal 1999). The slice sampler (Neal 2003) avoids specifying the proposal densities as in Metropolis-Hastings algorithms (e.g., Hastings 1970; Gelman 1992). In that way, after finding the appropriate augmentation scheme, the method can be applied to all data sets without any computational difficulties (Ntzoufras 2009). The slice sampler is defined by

$$\hat{p}(\beta,u) = \begin{cases} \dfrac{1}{Z_p} & \text{if } 0 \le u \le \tilde{p}(\beta) \\ 0 & \text{otherwise} \end{cases}, \tag{2.22}$$

where $Z_p = \int \tilde{p}(\beta)d\beta$. The marginal distribution over β is given by

$$\int \hat{p}(\beta,u)du = \int_0^{\tilde{p}(\beta)} \frac{1}{Z_p} du = \frac{\tilde{p}(\beta)}{Z_p} = p(\beta) \ , \tag{2.23}$$

so we can sample from $p(\beta)$ by sampling from $\hat{p}(\beta,u)$ and then ignoring the u values. Given the value of β we evaluate $\tilde{p}(\beta)$ and then sample u uniformly in the range $0 \le u \le \tilde{p}(\beta)$. Afterward, u is uniformly fixed from the 'slice' through the distribution defined by $\{\beta: \tilde{p}(\beta) > u\}$. Slice sampling is applied to multivariate distributions by repeatedly sampling each of the n variables in turn, in the manner of the Gibbs sampling (Bishop 2006), where one needs n iterations to get from $\beta_j^{(i)}$ to $\beta_j^{(i+1)}$

$$\left.\begin{aligned}
&1:\beta_1^{(i+1)} \sim p(\beta_1 \mid \beta_2^{(i)},\beta_3^{(i)},...,\beta_n^{(i)}) \\
&2:\beta_2^{(i+1)} \sim p(\beta_2 \mid \beta_1^{(i+1)},\beta_3^{(i)},...,\beta_n^{(i)}) \\
&\quad\vdots \\
&j:\beta_j^{(i+1)} \sim p(\beta_j \mid \beta_1^{(i+1)},\beta_2^{(i+1)},...,\beta_{j-1}^{(i+1)},\beta_{j+1}^{(i)},...,\beta_n^{(i)}) \\
&\quad\vdots \\
&n:\beta_n^{(i+1)} \sim p(\beta_n \mid \beta_1^{(i+1)},\beta_2^{(i+1)},...,\beta_{n-1}^{(i+1)})
\end{aligned}\right\}. \qquad (2.24)$$

The model standard deviation was then calculated from the expected number of events $E(y_{T+1} \mid x_{T+1},\beta)$ as estimated using the average of the mean μ obtained from the MCMC samples. The model standard deviation σ is then defined as

$$\begin{aligned}\sigma(y_{T+1} \mid y_{1:T}) = E[\sigma(y_{T+1} \mid x_{T+1},\beta_i) \mid y_{1:T}] \\ + \sigma[E(y_{T+1} \mid x_{T+1},\beta) \mid y_{T:1}]\end{aligned}, \qquad (2.25)$$

where the first term is the average process standard deviation and the second term is the coefficient uncertainty (Peters et al. 2008). The hindcasted TCG totals are taken as the number with the maximum probability from the hindcast distribution.

II PAPERS

3 THE SEPARATED EFFECTS OF THE EL NIÑO-SOUTHERN OSCILLATION AND INDIAN OCEAN DIPOLE ON AUSTRALIAN REGION TROPICAL CYCLONE COUNTS

3.1 Chapter Overview

This chapter addresses whether the Indian Ocean Dipole (IOD) is a useful predictor of seasonal variations in Australian region annual tropical cyclone (TC) counts. The Australian region TCG counts were taken from a 40-year record (1968/69-2007/08) in the region 0-30°S, 90°-170°E. To be able to investigate the IOD effects on TCs independently from ENSO, a simple but effective measure was introduced to separate the IOD from ENSO dependence. The then ENSO-independent IOD index, DMI_{NOENSO}, was utilised in a simple multiple linear regression model to analyse the hindcast skill of the ENSO-independent IOD, as well as with ENSO.

The main text of this chapter is a paper under minor revisions for the Journal *Climate Dynamics* (Werner et al. 2011: A new method for extracting the ENSO-independent Indian Ocean Dipole: application to Australian region tropical cyclone counts, *Climate Dynamics, under minor revisions*).

Candidate's contribution to this paper

The experiment design and analysis methods were the candidate's ideas and then jointly discussed between Dr Holbrook, Dr Maharaj and myself. I also performed all of the data analysis, however input from Dr Holbrook and Dr Maharaj was attained regularly throughout the process. All sections of the coauthored *Climate Dynamics* paper and the ongoing review process were led by myself under the guidance of both coauthors.

A new method for extracting the ENSO-independent Indian Ocean Dipole: application to Australian region tropical cyclone counts

Angelika Werner
Department of Environment and Geography, Macquarie University, NSW 2109 Australia

Angela M. Maharaj

Department of Environment and Geography, Macquarie University, NSW 2109 Australia

Neil J. Holbrook
School of Geography and Environmental Studies, University of Tasmania, TAS 7001 Australia

Under Minor Revisions for *Climate Dynamics*
(24 April 2011)

Corresponding author address: Assoc Prof Neil J Holbrook,
School of Geography and Environmental Studies, University of Tasmania, TAS 7001 Australia
Phone: +61 3 6226 2027 | Fax: +61 3 6226 7628 | E-mail: **Neil.Holbrook@utas.edu.au**

Keywords Climate variability, IOD, ENSO, Tropical Cyclones, Australia

Abstract

We introduce a simple but effective means of removing ENSO-related variations from the Indian Ocean Dipole (IOD) in order to better evaluate the ENSO-independent IOD contribution to Australian climate – specifically here interannual variations in Australian region tropical cyclogensis (TCG) counts. The ENSO time contribution is removed from the Indian Ocean Dipole Mode index (DMI) by first calculating the lagged regression of the DMI on the sea surface temperature anomaly (SSTA) index NINO3.4 to maximum leads and lags of 8 months, and then removing this ENSO portion. The new ENSO-independent time series, DMI_{NOENSO}, correlates strongly with the original DMI at r = 0.87 (significant at >99% level). Despite the strength of the correlation between these series, the IOD events classified based on DMI_{NOENSO} provide important differences from previously identified IOD events, which are more closely aligned with ENSO phases. IOD event composite maps of SSTAs regressed on DMI_{NOENSO} reveal a much greater ENSO-independence than the original DMI-related SSTA pattern. This approach is used to explore relationships between Australian region TCG and IOD from 1968-2007. While we show that both the DMI and DMI_{NOENSO} have significant hindcast skill (on the 95% level) when used as predictors in a multiple linear regression model for Australian region annual TCG counts, the IOD does not add any significant hindcast skill over an ENSO-only predictor model, based on NINO4. Correlations between the time series of annual TCG count observations and ENSO+IOD model cross-validated hindcasts achieve r = 0.68 (significant at the 99% level).

1 Introduction

Australia's tropical climate is dominated by two large-scale interannual climate modes: El Niño-Southern Oscillation (ENSO) in the Pacific basin and Indian Ocean Dipole (IOD) in the Indian Ocean (e.g., Philander 1990; Saji et al. 1999; Webster et al. 1999). These modes and their effects on Australian's climate have been the subject of numerous studies (e.g., Cai et al. 2001; Ashok et al. 2003; Meyers et al. 2007; Ramsay et al. 2008; Hendon et al. 2009; Ummenhofer et al. 2009). Importantly, rainfall variability in Australia is shown to be associated with the IOD almost as significantly as with ENSO (Saji and Yamagata 2003). Nevertheless, there remains much ongoing debate over the extent to which the IOD is indeed a unique 'mode' of the climate system and hence whether it is therefore really independent of ENSO (e.g., Saji et al. 1999; Allan et al. 2001; Meyers et al. 2007; Risbey et al. 2009).

Various studies have explored methods to separate the pure IOD mode from ENSO, or otherwise investigated the effect of the IOD on key climate variables over Australia such as precipitation and temperature (e.g., Ashok et al. 2003; Saji et al. 2005; Meyers et al. 2007; Risbey et al. 2009; Ummenhofer et al. 2009). For example, Meyers et al. (2007) applied a lagged empirical orthogonal function (EOF) analysis to remove the direct or lagged effects of ENSO from the IOD. This approach has since been applied in other studies (Risbey et al. 2009; Ummenhofer et al. 2009), although it only removes ENSO at a single lag from the IOD. Other studies have used partial correlations to remove the direct ENSO and IOD effects from Australian temperature and rainfall data (Ashok et al. 2003; Saji et al. 2005; Risbey et al. 2009). In short, these studies have been able to demonstrate some degree of independence of the IOD

from ENSO regarding the timing of Australian region precipitation rate variability, despite the difficulties in removing the more complete lag/lead effects of ENSO.

There have also been numerous studies of the influence of ENSO on tropical cyclone formation in the Australian region (e.g., Nicholls 1984; Solow and Nicholls 1990; Basher and Zheng 1995; Kuleshov and de Hoedt 2003; Ramsay et al. 2008; Kuleshov et al. 2009). We are aware of only one previous study that has investigated TC activity in the Australian region considering the IOD as well as ENSO (Liu and Chan 2010). Their study inferred that there may be an IOD contribution to Australian region TC activity, degeneracy in the empirical orthogonal functions presented together with ENSO/IOD linkages in the composite analyses. However, Liu and Chan (2010) had not been separated the signals into true ENSO and IOD modes, moreover they have not quantified the contribution of IOD to ENSO in terms of predictability of annual variations of TC measures.

In this paper, we introduce a simple method to remove the direct and lagged ENSO signal from the IOD - a method that can be usefully applied more generally to the separation of climate mode effects from the time series of different parameter sets. Our approach uses lagged regression and allows us to quantify the contributions from a relatively 'pure' Indian Ocean Dipole on annual TCG counts in the Australian region, i.e., independently from ENSO. While we show that the timing of the ENSO-independent IOD index, DMI_{NOENSO}, is significantly correlated (at the 99% confidence level) with variations in annual TCG counts in the Australian TC region, application of a multiple linear regression model demonstrates that the IOD alone has significant hindcast skill (on the 95% confidence level), but does not add any significant skill for hindcasting Australian region TCG counts over an ENSO-only predictor model. Correlation coefficients between the cross-validated hindcasts and

the of the combined ENSO+IOD model and TCG counts are as high as r = 0.68, which translates to approximately 46% of the explained variance in Australian region annual TCG counts.

The paper is structured as follows. Section 2 describes the data and variables used. Our climate signal separation method and its advantages over previous techniques will be explained in section 3. In section 4, we investigate the effect of the ENSO-independent IOD on variations in Australian region TCG counts. The approach uses multiple linear regression to build a statistical forecast model that investigates the importance of the IOD on Australian region annual TCG counts using leave-one out cross-validation with the ENSO-independent IOD (DMI_{NOENSO}) as a predictor. Finally, section 5 discusses and summarises the quality and improvements of the method and provides some conclusions regarding the importance of the large-scale IOD as a climate mode forcing for Australian region tropical cyclone formation.

2 Data and variables

2.1 Ocean data and indices

The statistical analysis is based on standard sea surface temperature anomaly (SSTA) indices. These were calculated from SST data provided in the Hadley Centre Global Sea Ice and Sea Surface Temperature (HadISST1) dataset (Rayner et al. 2003), compiled by the UK Met Office Hadley Centre. HadISST1 is a combination of global monthly SST fields and sea ice concentrations, and on a 1° × 1° grid from 1870 to the present. To quantify ENSO timing from an oceanic perspective, we used region-averaged NINO3.4 (5°N-5°S, 120°W-170°W) and NINO4 (5°N-5°S, 160°E-150°W) monthly SSTAs calculated from HadISST1. ENSO events are classified according to

the definition used by the U.S. National Weather Service (<http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml>). El Niño (La Niña) events are defined by the NINO3.4 SSTA exceeding thresholds of ±0.5^{o}C for a minimum of five consecutive three-month average overlapping periods. The NINO3.4 index enables us to capture both the classical "cold-tongue" ENSO and the more ENSO Modoki-like variations. The Dipole Mode index (DMI) is used to quantify the timing of IOD variations and represents the SSTA gradient between the Western (50°-70°E, 10°N-10°S) and Eastern Indian Ocean (90°-110°E, 0°-10°S) (Saji and Yamagata 2003).

All monthly SSTAs are calculated against the climatological monthly means derived from the 30-year base period 1970-1999. Correlation analyses take into account serial correlations and the effective number of degrees of freedom in the time series according to Davis (1976), and significance is tested at the 95% confidence level.

2.2 Tropical cyclogenesis observations

This study takes advantage of the global tropical cyclone (TC) best track data set IBTrACS (Knapp et al. 2010) provided by the U.S. National Oceanic and Atmospheric Administration (NOAA). TCG is defined as the spatial location where a tropical storm system with winds exceeding 34 knots (17.5 ms^{-1}) was first recorded.

For the purpose of this study, the Australian tropical cyclone region is defined as spanning between 0°-30°S and 90°-170°E. TCG occurrences identified over land have been removed in the quality control process. Also, only TCG events during the Australian TC season from November to April are taken into account. Overall, a total of 570 TCs during the 40-year period from 1968/69 to 2007/08 are analysed following

the quality control. Fig. 1 shows the spatial distribution of all TCG occurrences included in this study together with the time series of Australian region annual TCG counts from 1968/1969 to 2007/2008.

3 A method for removing ENSO

3.1 Lagged regression

In an early study of the independence of the IOD, Allan et al. (2001) argued that the IOD is highly correlated with ENSO. Fig. 2 shows the absolute value of cross-correlations between the three-monthly averaged NINO3.4 SSTA and DMI at monthly leads and lags up to a maximum of 10-months over the 40-year period 1968-2007. Statistically significant correlations (at the $p<0.05$ level) can be seen between the one-month lead to five-months lagged NINO3.4 to the DMI. We find that by regressing the DMI on the monthly NINO3.4 time series at leads and lags up to eight months, and subtracting the sum of these contributions from the IOD time series, the method is most effective and efficient in removing the ENSO signal from the Indian Ocean Dipole SSTA index. This technique takes account of the significant lead and lag effects of the ENSO signal that we wish to remove from the complementary co-existing (non-ENSO) variable, here the DMI. Hence, our ENSO-independent monthly SSTA Dipole mode index is calculated as:

$$\mathrm{DMI}_{\mathrm{NOENSO}}(t) = \mathrm{DMI}(t) - \sum_k b_k \,\mathrm{NINO3.4}(t+\mathrm{m}(k)) \ , \qquad (1)$$

where t is the month in the full time series, k = 1, 2, …, 17 is the regression number, m = -8, …, 8 is the corresponding regression monthly lead or lag, and b_k is the regression coefficient on the k^{th} NINO3.4 predictor. The original DMI time series and

the newly derived time series (Fig. 3a) are significantly correlated at the 99% confidence level, with a correlation coefficient of r = 0.87. Cross-correlating the three-month averaged DMI_{NOENSO} to other ENSO metrics, including the SSTA indices NINO3, NINO4 and the sea level pressure gradient index the Southern Oscillation index, we found no remaining significant correlation within lead and lags of up to 12 months. Conversely, the residuals between DMI and DMI_{NOENSO} were highly and significant correlated with NINO3.4 at r = 0.62 at the 99% confidence level (Fig. 3b). The individual variance of the DMI_{NOENSO} time series is reduced to 76% of the original DMI time series. From this point on, DMI_{NOENSO} refers to our newly formulated ENSO-independent normalised Indian Ocean Dipole SSTA index monthly time series.

3.2 Identification of ENSO-independent Indian Ocean Dipole events

The normalised DMI_{NOENSO} SSTA time series (Fig. 3a) is used here to identify the IOD event years. Following Meyers et al. (2007), an IOD event is considered to have occurred when two consecutive three-month average overlapping periods in any eight-month period between June and the following January exceed values higher than one standard deviation. In total, we identified 10 positive and 14 negative IOD events during the study period (Table 1). Differences in IOD event classification compared with Ummenhofer et al. (2009) are identified. Our ENSO-independent IOD time series identifies IOD event years quite different to previous studies (Saji et al. 2005; Meyers et al. 2007; Ummenhofer et al. 2009). While positive (negative) IOD events in these previous studies rarely coincide with La Niña (El Niño), we identify both positive and negative IOD events occurring throughout all ENSO-phases, which may be described as a greater ‘randomness’ in the co-variances between IOD and

ENSO events (Table 1). In our study period from 1968-2007, we only match 50% of both the positive negative events as identified in Ummenhofer et al. (2009).

Fig. 4a,b shows the SSTA composite maps of the tropical and subtropical Indo-Pacific Ocean for the peak IOD period of September-October-November (SON; Saji and Yamagata 2003). The SSTA maps corresponding to IOD event year composites are characteristic of the corresponding EOF patterns described previously (e.g. Saji et al. 1999). However, we believe these composite maps do not characterise the IOD pattern in isolation from other climate signals – in particular, ENSO. Rather, these IOD event year composites contain substantial ENSO dependence, which dominates the Indo-Pacific region, as well as SSTA contributions from the subtropical Indian Ocean. In the following section, we provide and trial a method that isolates the Indian Ocean Dipole signal from ENSO in the spatial domain.

3.3 Decomposing climate data

For a better understanding of the net influence of the primary Indian Ocean climate signal on the larger Indian to Southwest Pacific Ocean region, we regressed monthly SSTAs across all spatial grid cells (2.5˚x2.5˚) across the region, on DMI and DMI_{NOENSO}. This approach allows us to examine the ENSO-independent IOD event year contributions as a composite of SSTAs across the Indo-Pacific region. The on DMI_{NOENSO} regressed SSTA fields (Fig. 4e,f), corresponding to the peak timing (SON) of the characteristic IOD pattern, reveal that most of the spatial pattern contributions from the Pacific that were evident in the original (cf. Fig. 4a,b) and on the DMI regressed composite SSTA fields (cf. Fig. 4c,d) for the IOD event years have been filtered out. The extension of the dipole pattern towards the subtropical regions of the Indian Ocean that existed previously within the IOD composite maps also

disappears, suggesting that this extended feature is an ENSO-linked response. This method has great utility as it can be applied usefully to any climate variable in order to extract the ENSO-independent climatic field contributions in space and time. Further the method is generic as it can be applied to separate out any climate signals, specified by the framing and context of the task.

4 IOD effect on Australian region TCG counts

4.1 Effects of ENSO and IOD on Australian region TCG counts

The effect of ENSO on TC occurrences in the Australian region has been analysed extensively (e.g., Nicholls 1984; Evans and Allan 1992; Basher and Zheng 1995; Kuleshov et al. 2008; Ramsay et al. 2008; Kuleshov et al. 2009). Ramsay et al. (2008) demonstrated that the NINO4 index provides a useful simple metric of ENSO that relates most strongly with TCG in the Australian region. Taking account of this finding, we also used NINO4 in the present study as the regional-specific ENSO index most appropriate to forecast Australian region TCG.

Fig. 5 provides an inventory of correlations between three-monthly means of the selected climate indices and Australian region annual TCG occurrences. In Table 2 the strongest correlations between annual TCG counts and the three-month averaged climate indices prior to the Australian TC season onset (November) are provided. It is not surprising that NINO4 provides the strongest correlations with annual TCG counts in the Australian region with a correlation coefficient of r = -0.69 during August-September-October (ASO). These results corroborate findings from previous studies (e.g., Nicholls 1984; Ramsay et al. 2008).

Notwithstanding the dominant pre-seasonal ENSO contribution to Australian region annual TCG counts provided by NINO4, which explains 48% of the variance in annual TCG frequency, the DMI_{NOENSO} correlates most strongly with annual TCG counts around its peak also during ASO ahead of the upcoming season (r = -0.45; significant at the 99% confidence level), with the ENSO-independent IOD explaining 20% of the variance in annual TCG counts. The original (standard) DMI time series is mere strongly correlated with annual TCG counts during JAS/ASO with correlation coefficients as high as -0.6. While the sign (-) and timing is consistent with our DMI_{NOENSO} from early austral winter to early summer, the far stronger correlations between the standard DMI and TCG over DMI_{NOENSO} is due to the implicit contribution from ENSO in the original DMI time series that represents the stronger predictor of Australian region TCG occurrences.

4.2 Predictive model of TCG counts

Our results suggest that a combination of complementary Pacific and Indian Ocean centred indices may have the potential to forecast annual TCG counts in the Australian region better than an ENSO-only based model. To investigate this further, we perform a multiple linear regression analysis of annual TCG counts on the ENSO and non-ENSO predictor indices and their combinations. We chose the TC pre-seasonal three-monthly indices averaged over ASO, as then all indices correlated most strongly with the Australian region TCG time series (Table 2; Fig. 5). These predictor indices were used in the multiple linear regression model $TCG_{totals}(t) = b_o + \sum b_n * predictor_n(t)$, where $predictor_n$ is the nth SSTA predictor time series for the upcoming TC season t, and b_o and b_n are the corresponding regression coefficients. To quality assure the TCG forecast (hindcast) estimates and to identify the best model, a

leave-one-out cross-validation technique was applied. This method trains the model on n-1 years to 'hindcast' the left-out year (e.g., Stone 1974) and quantifies the potential predictive skill of the selected climate predictor indices by calculating the root-mean-squared error (RMSE; e.g., Elsner and Jagger 2006) and the standard error (se). Correlation coefficients between the observations and the cross-validated hindcasts are indicative of the quality and potential forecast utility of the selected climate variables for TCG occurrence (Table 2). Despite the stronger individual hindcast skill of DMI over DMI_{NOENSO}, we find the NINO4+DMI_{NOENSO} model performs just as good as NINO4+DMI, with both not adding significant skill to the NINO4-only model. Fig. 6a presents the observed annual TCG counts and leave-one out cross-validated model hindcasts of annual TCG counts for the NINO4+DMI_{NOENSO} model. Fig. 6b illustrates the relatively small differences between the combined and the NINO4-only models via the residuals between the TCG observations and model hindcasts. In 10 of the 40 investigated seasons (25%), the DMI_{NOENSO} improves the hindcast, hile in 10 seasons (25%) the NINO4-only model is more skilful. Hence, overall the ENSO-independent IOD provides only marginal additive skill that is insignificant above the ENSO variance. Correlations between the observed Australian region TCG counts and the cross-validated model hindcasts improved only slightly from r = 0.67 (NINO4-only model) to r = 0.68 (NINO4+DMI_{NOENSO} model), with corresponding small reductions in the standard error and RMSE (Table 2).

5 Summary and Discussion

The dominant large-scale modes of interannual SST variablility in the tropical Pacific and Indian Oceans during austral spring and summer are El Niño-Southern

Oscillation (ENSO) and the Indian Ocean Dipole (IOD). Both climate modes have been shown to affect the climate and precipitation of Australia (Meyers et al. 2007). In the present study, we isolated and combined the contributions from the IOD and ENSO to understand their respective roles in influencing tropical cyclone formation (genesis; TCG) annual counts in the Australian region.

Previous research suggests that El Niño usually occurs in combination with positive IOD events, and conversely La Niña with negative IOD events, albeit with exceptions (Meyers et al. 2007). Various studies have attempted to untangle the relationship between ENSO and the IOD, and have endeavoured to evaluate the relative independence of these climate signals as modes of variability (e.g., Saji et al. 2005; Meyers et al. 2007; Ummenhofer et al. 2009). For example, previous studies have shown that using a partial correlation approach applied to rainfall or temperature variations (Ashok et al. 2003; Saji et al. 2005; Risbey et al. 2009) removes only the direct (zero-lag) effects of ENSO (in terms of the widely used sea level pressure gradient, the Southern Oscillation index) or IOD (through the Dipole Mode index). Hence, lag and lead effects were ignored. Meyers et al. (2007) instead used a lagged EOF approach to remove lagged correlations, by firstly shifting the time series describing ENSO and IOD toward the highest correlations with the NINO3 SSTA time series, and then decomposing the data using an EOF analysis on the entire time series. The first EOF (the ENSO mode) was then removed. While this method demonstrates one way of addressing the problem of lagged interactions between the IOD and ENSO, it only removes one fixed lag or lead-time shift despite its complexity. It is also well understood that ENSO and IOD events do not evolve at exactly the same ‘rate’, i.e., ENSO and IOD events peak at different times (Meyers et al. 2007). Consequently, we cannot assume that lagged interactions between events,

taken individually or collectively, will always occur on the same timescale. Hence, this method is unable to capture the substantial variances in physical interactions between ENSO and the IOD spanning over multiple phases. We further note that using the EOF method to decompose these time series means that interannual and inter-decadal variations in the Pacific will be expected to separate into two different modes. Even after removing the leading EOF, decadal variations of ENSO are still present in the IOD time series (Meyers et al. 2007).

All these problems are avoided in the approach introduced in the present study, which applies a lagged regression of NINO3.4 on the IOD index, the DMI, and removes the ENSO-dependent lagged components to produce an ENSO-independent IOD index, which we call DMI_{NOENSO}. Lead and lag relationships between ENSO and the IOD are taken into account up to eight months prior and post, and can be readily extended to shorter or longer lead/lags where deemed to be important. Further, the method can be potentially applied usefully to any climate variable in order to extract the ENSO-independent IOD climatic field contributions in both space and time. Another advantage of this method is its utility to extract any climate signal in order to isolate selected climate effects. In our application of the method, we found significant cross-correlations between NINO3.4 and the DMI of from one-month lead to five-months lag. We found that by regressing the DMI on the NINO3.4 time series at leads and lags up to eight months, and subtracting the sum of these contributions from the DMI time series, the method is effective and efficient in removing the ENSO signal from the Indian Ocean Dipole SSTA index. In comparison to the previous study by Ummenhofer et al. (2009), we found important differences in the classification of IOD events. Notwithstanding, the September-October-November averaged SSTA composite maps of IOD-events presented here (Fig. 4) clearly show the isolated IOD

SSTA variations as being concentrated in the tropical Indian Ocean. This method is successful in better separating the influences of ENSO and IOD in order to investigate the importance of the Indian Ocean Dipole on Australian region TCG.

Ramsay et al. (2008) recently showed that out of all ENSO indices readily available, NINO4 is the best climate index predictor of Australian region annual TCG counts with a correlation coefficient between August-September-October (ASO) averaged NINO4 region SSTA and annual TCG counts of r = -0.73 for the 36-year period 1970-2005. For our longer 40-year period of 1968-2007, there remains a strong, albeit weaker, relationship between ASO averaged NINO4 and the Australian region annual TCG counts with r = -0.69 (significant at the 99% level). We attribute a proportion of the difference between these correlations to the longer time period of this study, and our use of the more recent released and quality-improved IBTrACS TC data set. We also found significant correlations (at the 99% level) between the pre-seasonal ASO averaged DMI and DMI_{NOENSO} with Australian region TCG, with r = -0.60 and r = -0.45 respectively.

To evaluate the contribution of various climate drivers on Australian region TCG, we developed a multiple linear regression model using a leave-one-out cross-validation approach. We found that IOD has significant hindcast skill (on the 95% confidence level) of Australian annual TCG counts, with the original DMI performing better than DMI_{NOENSO} (Table 2). An ENSO+IOD multiple linear regression model provides a slight improvement of potential forecast skill over an ENSO-only model (Fig. 6). Correlations between model cross-validated hindcasts and observations of TCG counts are r = 0.68 for the NINO4+DMI_{NOENSO} model compared to r = 0.67 for the NINO4-only model. Liu and Chan (2010) included both the original DMI and NINO4 as predictors in their seasonal forecast model for Australian region TCG

counts, however they remained inconclusive about the role of the IOD on TC activity and did not investigate the additional contribution of DMI over NINO4. Also they indicate that while the IOD may influence TC activity in the Australian region, the mechanisms are unclear and that further work was required to understand any potential IOD mechanisms. Our study provides extra work showing that a truly ENSO-independent IOD is relatively unimportant to Australian region TCG compared with ENSO, but nevertheless improves the model slightly, albeit insignificantly. More importantly, the ASO averaged original DMI time series alone shows good hindcast skill of observed Australian TCG counts, but did not add more value to the ENSO-models as did the ENSO-independent DMI_{NOENSO}. This highlights, that the better TCG hindcast skill of DMI over DMI_{NOENSO} is solely due to the original DMI's inter-dependency to ENSO. We note that, this analysis is simply, nevertheless one application of the approach introduced in this paper.

In summary, we have presented a simple and effective method for removing ENSO-related variations from the IOD signal using lagged regression. A new ENSO-independent Indian Ocean Dipole SSTA index (DMI_{NOENSO}) has been generated that has important differences to the original Dipole Mode index, providing an IOD event classification that differs from previous definitions. Regression of DMI_{NOENSO} on SST across space and time isolates a much purer Indian Ocean pattern contribution than has been shown previously. We believe there is great utility in this approach presented, with the new time series being of value for other public research. Finally, our study shows that Australian region tropical cyclone formation appears to be influenced by our purer Indian Ocean 'mode' (significant at the 99% confidence level). Further it has significant hindcast skill (on the 95% confidence level) of Australian TCG counts and adds value to an ENSO-only TCG forecast model.

However, ENSO remains the dominant climate player in Australian region TCG, with the IOD only adds marginal hindcast skill that is not significant above the background ENSO variance.

Acknowledgements A. Werner was supported by MQRES and PGRF Scholarships from Macquarie University, Australia. AW also thanks the School of Geography and Environmental Studies at the University of Tasmania for providing computing resources and office space during her visits to the University of Tasmania. Finally, we wish to thank Dr. Peter C. McIntosh from the Centre for Australian Weather and Climate Research for his constructive comments on this work.

References

Allan RJ, Chambers D, Dosdrowsky W, Hendon HH, Latif M, Nicholls N, Smith I, Stone RC, Tourre Y (2001) Is there an Indian Ocean dipole, and is it independent of the El Niño-Southern Oscillation? CLIVAR Exchanges 6: 18-22

Ashok K, Guan Z, Yamagata T (2003) Influence of the Indian Ocean Dipole on the Australian winter rainfall. Geophys Res Lett 30: 1821, doi:10.1029/2003GL017926

Basher RE, Zheng X (1995) Tropical cyclones in the southwest Pacific: Spatial patterns and relationships to Southern Oscillation and sea surface temperature. J Climate 8: 1249-1260

Cai W, Whetton PH, Pittock AB (2001) Fluctuations of the relationship between ENSO and northeast Australian rainfall. Clim Dyn 17: 421–432

Davis RE (1976) Predictability of sea surface temperatures and sea level pressure anomalies over the North Pacific Ocean. J Phys Oceanogr 6: 249-266

Elsner JB, TH Jagger (2006) Prediction models for annual U.S hurricane counts. J. Climate 19: 2935–2952

Evans JL, Allan RL (1992) El Nino/Southern Oscillation modification to the structure of the monsoon and tropical cyclone activity in the Australasian region. Int J Climatol 12: 611-623

Hendon HH, Hudson D, Wang G, Alves O (2009) Dynamical forecasts of inter-El Nino variations of tropical SST and Australian springtime rainfall. Mon Wea Rev: 137: 3796-3810

Knapp KR, Kruk MC, Levinson DH, Diamond HJ, Neumann CJ (2010) The international best track archive for climate stewardship (IBTrACS). Bull Amer Meteor Soc 91: 363-376

Kuleshov Y, Chane Ming F, Qi L, Chouaibou I, Hoareaux C, Roux F (2009) Tropical cyclone genesis in the southern hemisphere and its relationship with the ENSO. Ann Geophys 27: 2523–2538

Kuleshov Y, de Hoedt G (2003) Tropical cyclone activity in the Southern Hemisphere. Bull Austral Meteor Oceanogr Soc, 16, 135-137

Kuleshov Y, Qi L, Fawcett R, Jones D (2008) On tropical cyclone activity in the Southern Hemisphere: Trends and the ENSO connection. Geophys Res Lett 35: L14S08, doi:10.1029/ 2007GL032983

Meyers G, McIntosh PC, Pigot L, Pook M (2007) The years of El Niño, La Niña, and interactions with the tropical Indian Ocean. J Climate 20: 2872-2880

Nicholls N (1984) The southern oscillation, sea-surface temperature, and interannual fluctuations in Australian tropical cyclone activity. J Climatol. 4: 661-1149

Philander SGH (1990) El Niño, La Niña and the Southern Oscillation. Academic Press, San Diego, CA: 289 pp

Ramsay HA, Leslie LM, Lamb PJ, Richman MB, Leplastrier M (2008) Interannual variability of tropical cyclones in the Australian region: Role of large-scale environment. J Climate 21: 1083-1103

Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. J Geophys Res 108: (D14)

Risbey JS, Pook MJ, McIntosh PC, Wheeler MC, Hendon HH (2009) On the remote drivers of rainfall variations in Australia. Mon Wea Rev 137: 3233-53

Saji NH, Ambrizzi T, Ferraz SET (2005) Indian Ocean Dipole Mode events and austral surface air temperature anomalies. Dym Atmos Oceans 39: 87-101

Saji NH, Goswami BN, Vinayachandran PN, Yamagata T (1999) A dipole mode in the tropical Indian Ocean. Nature 401: 360-363

Saji, NH, Yamagata T (2003) Structure of SST and surface wind variability during Indian Ocean dipole mode years: COADS observations. J Climate 16: 2735–2751

Solow A, Nicholls N (1990) The relationship between Southern Oscillation and tropical cyclone frequency in the Australian region. J Climate 3: 1097-1101

Stone M (1974) Cross-validatory choice and assessment of statistical predictions. J Royal Stat Soc 36(2): 111–147

Ummenhofer CC, England MH, McIntosh PC, Meyers GA, Pook MJ, Risbey JS, Sen Gupta A, Taschetto AS (2009) What causes Southeast Australia's worst droughts? Geophys Res Lett 36: L04706, doi:10.1029/2008GL036801

Webster PJ, Moor AM, Loschnigg JP, Leben RR (1999) Coupled ocean–atmosphere dynamics in the Indian Ocean during 1997–98. Nature 401: 356-360

Table 1 Positive and negative IOD events and their coincidence with El Niño or La Niña years between 1968-2007. The year identified as El Niño/La Niña corresponds to its year of onset. Also indicated are the IOD event years as classified by Ummenhofer et al. (2009), with matching event years indicated in bold. The +/- sign in brackets indicate the positive/negative IOD event as classified in that previous study

	El Niño	**neutral**	**La Niña**
IOD+	1972, **1976**, **1991**, **1994**, **1997,** 2006		**1983**, 1999, 2000**,** 2007
neutral	**1977**, 1982(+), **1987**	1978(+),**1979**,**1993, 2001**, **2003**, **2005**	**1970**, **1971**, **1973**, 1975(-), 1988(-), **1995**, **1998**
IOD-	**1968**, 1969, 1986, 2002, **2004**	**1980**, 1981, 1989, 1990, **1992**, 1996	**1974**, **1984, 1985**

Table 2 Strongest correlation coefficients r(max) between three-month SSTA index means (prior to the Australian TC season onset; here ASO) and Australian region annual TCG counts (all significant at the 99% confidence level). Also shown are correlation coefficients r(CV) between the model cross-validated hindcasts of annual TCG counts (using the leave-one out method) and the observed annual TCG counts over the 40-year record 1968/69-2007/08 (all significant at the 95% confidence level). The standard error (se) and root-mean-squared error (RMSE) for each of the models are also provided. The strongest models are indicated in bold. Statistical significance takes account of serial correlation according to Davis (1976)

	r(max)	r(CV)	se	RMSE
NINO3.4	-0.58	0.50	0.68	4.23
NINO4	-0.69	0.67	0.59	4.65
DMI	-0.60	0.54	0.66	4.13
DMI_{NOENSO}	-0.45	0.39	0.72	4.50
NINO3.4+DMI		0.58	0.64	4.02
NINO3.4+DMI_{NOENSO}		0.58	0.64	4.02
NINO4+DMI		**0.68**	**0.58**	**3.60**
NINO4+DMI_{NOENSO}		**0.68**	**0.58**	**3.60**

FIG. CAPTIONS

Fig.1 (a) Spatial distribution of the first recorded location of tropical storm systems with winds exceeding 34 knots (17.5 ms^{-1}) in the Australian region from 1968/69-2007/08. (b) The corresponding time series of annual Australian region TCG counts

Fig.2 Absolute values of cross-correlation coefficients between the three-monthly averaged NINO3.4 SSTA and DMI from 1968-2007, at monthly leads and lags up to a maximum of 10 months. Dashed lines show the p-values of the correlations, while the solid horizontal line indicates the p = 0.05 level. Positive (negative) lags indicate the DMI leading (lagging) NINO3.4. Note that the p-value axis is flipped for visualisation purposes

Fig.3 Normalised three-month running mean SSTA time series for the 40-year period 1968-2007. Dashed lines indicate one standard deviation in the time series, to be used as thresholds for the identification of ENSO-independent IOD events. (a) DMI and DMI_{NOENSO} time series are shown (r = 0.87, significant at 99% level). (b) The NINO3.4 and 'residual' time series, $DMI_{RESIDUAL}$ (removed following lagged regression of DMI on NINO3.4) are shown (r = 0.62, significant at the 99% level)

Fig.4 Composite maps of SSTA (°C) during the peak IOD (SON) period for (a) positive IOD events, and (b) negative IOD events. Events are classified as in Table 1. (c),(d) as in (a),(b) but with SSTA (°C) regressed on the original DMI and (e),(f) SSTA (˚C) regressed on DMI_{NOENSO}

Fig.5 Correlation coefficient between Australian region annual TCG counts and selected SSTA index time series from January-February-March (JFM) prior to the

onset of the TCG season through to the post-TCG period of April-May-June (AMJ). The vertical dashed lines indicate the start and end of the IOD season

Fig.6 (a) Time series of observed annual TCG counts (solid line) and hold-one out cross-validated hindcasts of Australian region annual TCG counts (dashed line) for the NINO4+DMI_{NOENSO} model. Shading indicates the variance in the model results. (b) Residuals between the annual TCG observations and model hindcasts of the NINO4-only and the NINO4+DMI_{NOENSO} models

FIGURES

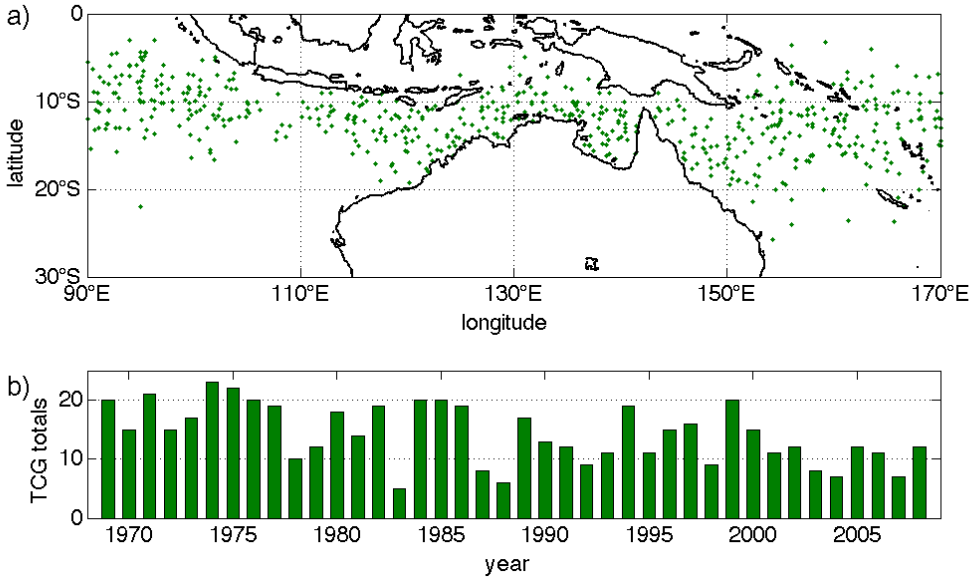


Fig.1 (a) Spatial distribution of the first recorded location of tropical storm systems with winds exceeding 34 knots (17.5 ms^{-1}) in the Australian region from 1968/69-2007/08. (b) The corresponding time series of annual Australian region TCG counts

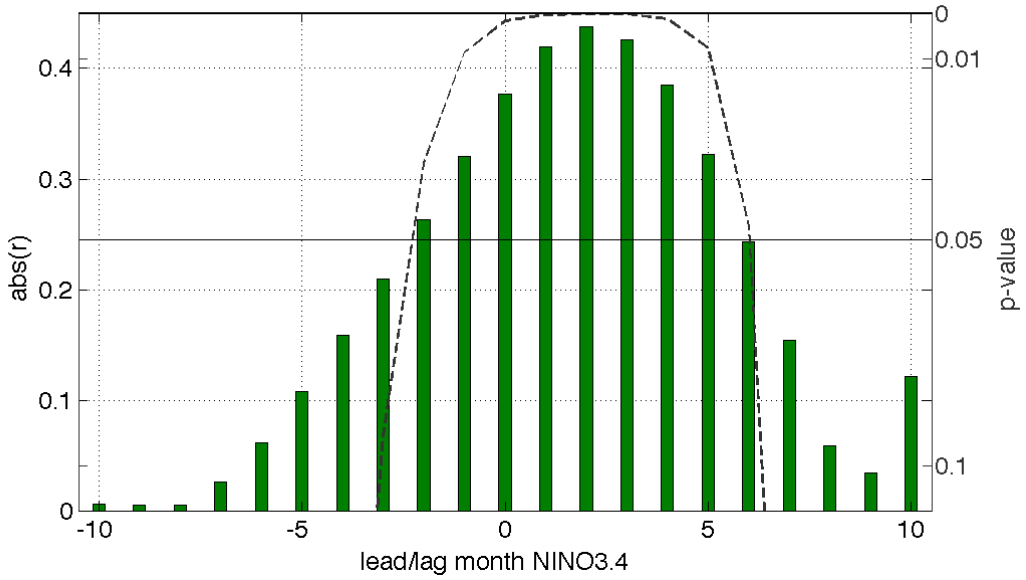


Fig.2 Absolute values of cross-correlation coefficients between the three-monthly averaged NINO3.4 SSTA and DMI from 1968-2007, at monthly leads and lags up to a maximum of 10 months. Dashed lines show the p-values of the correlations, while the solid horizontal line indicates the p = 0.05 level. Positive (negative) lags indicate the DMI leading (lagging) NINO3.4. Note that the p-value axis is flipped for visualisation purposes

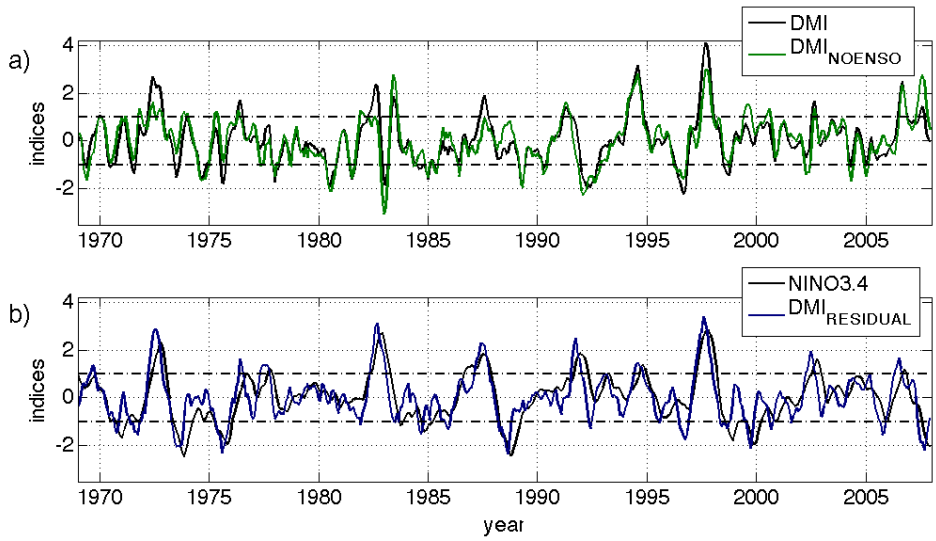


Fig.3 Normalised three-month running mean SSTA time series for the 40-year period 1968-2007. Dashed lines indicate one standard deviation in the time series, to be used as thresholds for the identification of ENSO-independent IOD events. (a) DMI and DMI_{NOENSO} time series are shown (r = 0.87, significant at 99% level). (b) The NINO3.4 and 'residual' time series, $DMI_{RESIDUAL}$ (removed following lagged regression of DMI on NINO3.4) are shown (r = 0.62, significant at the 99% level)

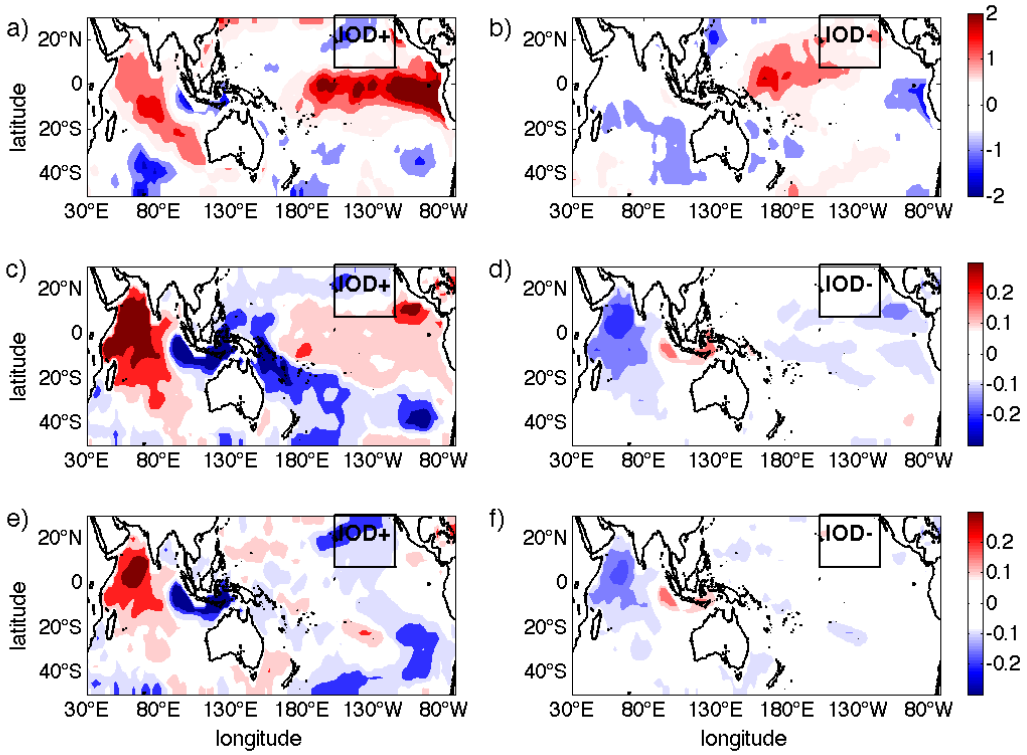


Fig.4 Composite maps of SSTA (°C) during the peak IOD (SON) period for (a) positive IOD events, and (b) negative IOD events. Events are classified as in Table 1. (c),(d) as in (a),(b) but with SSTA (°C) regressed on the original DMI and (e),(f) SSTA (˚C) regressed on DMI_{NOENSO}

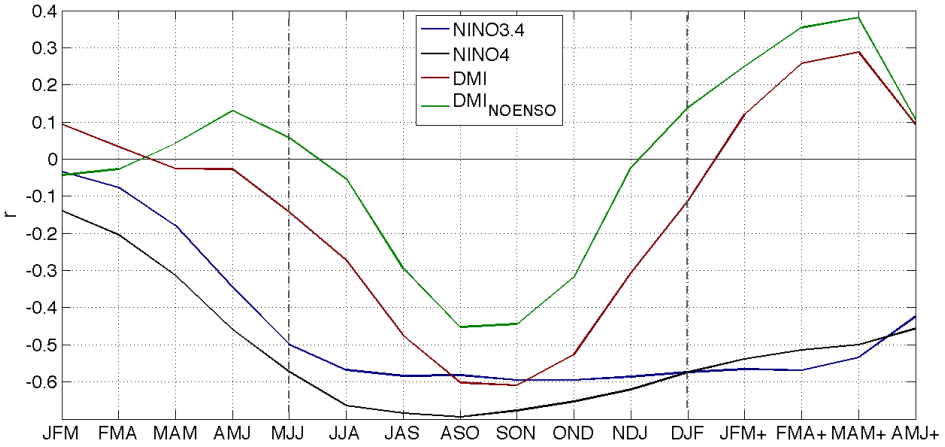


Fig.5 Correlation coefficient between Australian region annual TCG counts and selected SSTA index time series from January-February-March (JFM) prior to the onset of the TCG season through to the post-TCG period of April-May-June (AMJ). The vertical dashed lines indicate the start and end of the IOD season

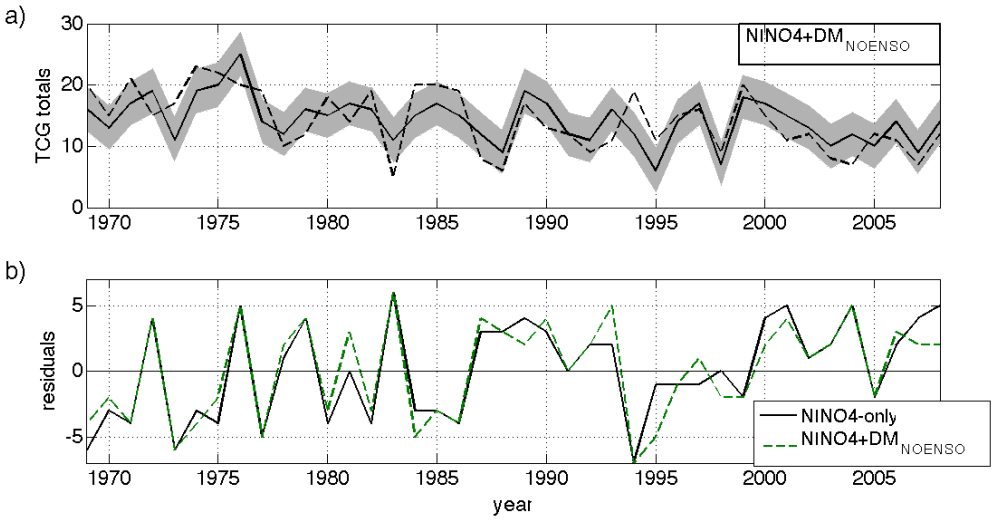


Fig.6 (a) Time series of observed annual TCG counts (solid line) and hold-one out cross-validated hindcasts of Australian region annual TCG counts (dashed line) for the NINO4+DMI$_{NOENSO}$ model. Shading indicates the variance in the model results. (b) Residuals between the annual TCG observations and model hindcasts of the NINO4-only and the NINO4+DMI$_{NOENSO}$ models

3.3 Further Discussion

The IOD is a large-scale ocean-atmosphere phenomenon in the tropical Indian Ocean, with impacts that extend over to the tropical and subtropical Pacific regions. Therefore it is of interest if the IOD is a useful predictor, not only for the all-Australian TC region but also for seasonal variations of TCG counts in its subregions. For this reason we present here the method introduced in the previous section to test the seasonal forecast skill of the ENSO-independent IOD for seasonal variations of TCG count in the Western and Eastern Australian subregions. Following Kuleshov et al. (2010), the two subregions are divided at 135°E where the least number of TC tracks were crossed in the historical records.

Fig. 3.1 provides an inventory of correlations between three-monthly means of the selected climate indices and annual TCG occurrences in the Australian subregions. In Table 3.1 the strongest correlations between annual TCG counts and the three-month averaged climate indices prior to the Australian TC season onset (November) are provided. While correlation coefficients between climate indices and annual TCG counts in the Western region show similar correlation coefficients with both NINO3.4 and NINO4 during August-September-October (ASO), NINO4 correlates strongly

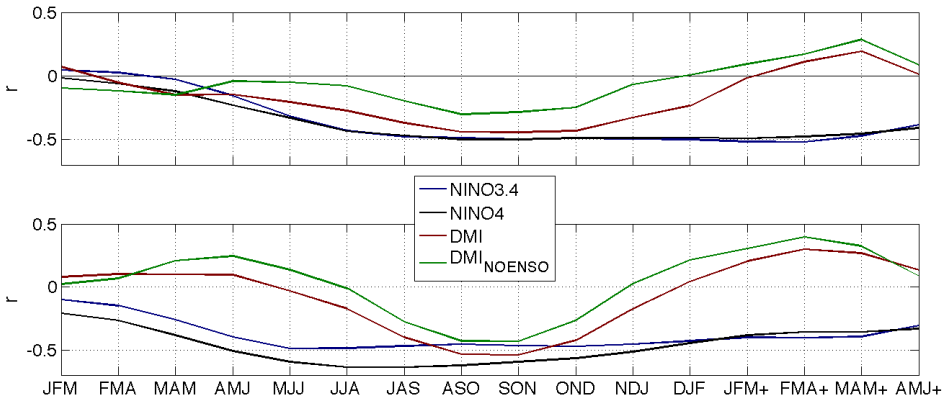


Figure 3.1 Correlation coefficient between Australian region annual TCG counts and selected SSTA index time series from January-February-March (JFM) prior to the onset of the TCG season through to the post-TCG period of April-May-June (AMJ). The vertical dashed lines indicate the start and end of the IOD season (JFM) prior to the onset of the TCG season through to the post-TCG period of AMJ.

with Eastern region TCG during May-June-July (MJJ) with r = -0.64 (significant at the 95% confidence level) explaining 41% of the total variance of seasonal TCG counts. As for the seasonal variation of TCG counts in the Australian region, the original (standard) DMI time series is more strongly correlated with annual TCG counts during ASO/September-October-November (SON) with correlation coefficients as high as r = -0.54 for the Eastern region and r = -0.44 for the Western region, explaining 29% and 19% of the total seasonal TCG count variance respectively. As for correlations of Australian region TCG, the sign (-) and timing is consistent with our DMI_{NOENSO} from early austral winter to early summer, the far stronger correlations between the standard DMI and TCG over DMI_{NOENSO} is due to the implicit contribution from ENSO in the original DMI time series that represents the stronger predictor of Australian region TCG occurrences (see Figure 3.1).

Table 3.1 Strongest correlation coefficients r(max) between three-month SSTA index means (prior to the Australian TC season onset; here ASO) and Western and Eastern Australian region annual TCG counts (all significant at the 95% confidence level are indicated bold). Also shown are the correlations, r(CV), between the model cross-validated hindcasts of annual TCG counts (using the leave-one out method) and the observed annual TCG counts over the 40-year record 1968/69-2007/08 (all significant at the 95% confidence level). The standard error (se) and root-mean-squared error (RMSE) for each of the models are also provided. The strongest models are indicated in bold. Statistical significance takes account of serial correlation according to Davis (1976).

	Western region				**Eastern region**			
	R(max)	**R(CV)**	**se**	**RMSE**	**r(max)**	**r(CV)**	**se**	**RMSE**
NINO3.4	-0.49 (ASO)	0.42	0.43	2.69	-0.49 (MJJ)	0.40	0.46	2.85
NINO4	-0.50 (ASO)	0.38	0.44	2.77	-0.64 (JJA)	0.56	0.41	2.55
DMI	-0.44 (ASO)	0.39	0.44	2.73	-0.54 (ASO)	0.48	0.43	2.72
DMI_{NOENSO}	-0.30 (ASO)	0.16	0.47	2.96	-0.43 (ASO)	0.39	0.45	2.84
NINO3.4+DMI		0.38	0.44	2.76		0.47	0.44	2.77
NINO3.4+DMI_{NOENSO}		0.37	0.44	2.77		0.48	0.44	2.74
NINO4+ DMI		0.35	0.45	2.83		0.56	0.41	2.58
NINO4+DMI_{NOENSO}		0.39	0.44	2.77		0.61	0.39	2.44

A multiple linear regression was applied and the cross-validated (leave-one out method) forecast skill of the predictor indices are shown in Table 3.1. In the Western Australian (eastern Indian Ocean) region, the NINO3.4 model provides the strongest correlations between observed annual TCG counts and cross-validated hindcasts with r = 0.42, a skill score of SS = 18.3% and the smallest RMSE = 2.69. Despite the stronger individual hindcast skill of DMI over DMI_{NOENSO}, we find the NINO4+DMI_{NOENSO} model performs best for the Eastern Australian (southwest Pacific) subregion with the correlation between annual TCG count observations and cross-validated hindcast of r = 0.61 and a RMSE = 2.44 (Table 3.1). A total skill score of SS = 36.9% was achieved with the NINO4+DMI_{NOENSO} model in the eastern Australian TC region over SS = 31.4% with the NINO4-only model.

In summary, the IOD – while showing some individual hindcast skill - does not have any additional hindcast skill over ENSO for annual TCG counts in the Western Australian region. In the Eastern region, the IOD adds skill, with the ENSO-independent IOD being even more skilful than the original ENSO-dependent DMI time-series. There the strongest predictor combination was shown to be NINO4+DMI_{NOENSO}.

3.4 Chapter Summary

This chapter addressed the thesis aims of the development of a simple, but effective method to separate the IOD from ENSO to estimate the statistical independency of IOD from ENSO. Also the individual and with ENSO measures combined forecast skill of the original IOD signal and the ENSO-independent IOD predictor of seasonal variations in TCG for the Australian region and subregions were quantified.

In detail we have presented a new method for extracting ENSO from the IOD and have shown that the IOD has a statistical ENSO-independent component important for Australian region TCG count variations. The IOD has relevant pre-seasonal cross-validated hindcast skill of annual TCG counts in the Australian region, but also in its subregions. When combining the skill with ENSO indices in a multiple linear regression model, additional hindcast skill of the IOD is marginal over the standard ENSO predictors for the Australian region and Western Australian subregion. In the Eastern Australian subregion, DMI_{NOENSO} adds substantial skill and hence improves the cross-validated TCG count hindcasts.

4. DEVELOPMENT OF STATISTICAL SEASONAL FORECAST MODELS FOR TROPICAL CYCLONE OCCURRENCES IN THE AUSTRALIAN REGION AND SUBREGIONS

4.1 Chapter Overview

This chapter introduces a Bayesian forecast model of Australian region annual TCG counts based on observational data from 1968/69-2007/08. To include climatological information in the most meaningful way possible, spatial correlations (prior to the TC season) between climate variables and seasonal TCG counts were analysed on the basis of persistent pattern indices derived and tested as predictors. A step-by-step predictor selection based on the probabilistic root-mean squared error (RMSE) ensured the most skilful model was taken into account. The final model, based on indices of convective available potential energy, meridional winds at 850hPa and geopotential height at 500hPa, shows considerable skill in hindcasting annual TCG counts for both the Australian region and Eastern Australian subregions. A separate model for the Western Australian subregion, based on indices of sea level pressure and meridional winds at 850hPa, substantially improves the hindcast skill in the eastern Indian Ocean (Western Australian) region.

The main text of this section is a paper accepted by the *Journal of Climate* (Werner and Holbrook 2011a: A Bayesian forecast model of Australian region tropical cyclone formation - accepted by the *Journal of Climate*). The second part of this chapter discusses the differences between the frequentist and Bayesian approaches, and the model skill for Australian subregions when divided at 135°E. The final part of this chapter is a paper being prepared for submission to the *Geophysical Research Letters* (Werner and Holbrook 2011b: How to improve seasonal forecast modelling of tropical cyclone formation in the southeast Indian Ocean, *in preparation for Geophysical Research Letters*) with a subsequent further discussion.

Candidate's contribution to the papers

The experiment design and analysis methods were the candidate's ideas and then jointly discussed between Dr Holbrook and myself. I also performed all of the data analysis. All sections of the coauthored *Journal of Climate* and *Geophysical Research Letters* papers and the ongoing review process were led by myself under the guidance of Dr Holbrook.

A Bayesian forecast model of Australian region tropical cyclone formation

Angelika Werner[1]
Department of Environment and Geography, Macquarie University, Sydney NSW, Australia

Neil J. Holbrook
School of Geography and Environmental Studies, University of Tasmania, Hobart TAS, Australia

Submitted to the Journal of Climate 23 November 2010;
Revised 6 May 2011

Accepted for publication 13 May 2011

---

[1] *Corresponding author address*: Angelika Werner, Department of Environment and Geography, Macquarie University, Sydney, New South Wales, 2109, Australia.

E-mail: **angelika.werner.la@gmail.com**

Abstract

A new and potentially skilful seasonal forecast model of tropical cyclone formation (genesis, TCG) is developed for the Australian region. The model is based on Poisson regression using the Bayesian approach. Predictor combinations are chosen using a step-by-step predictor selection. The three-predictor model based on derived indices of June-July-August average convective available potential energy, May-June-July average meridional winds at 850 hPa (v_{850}) and July-August-September geopotential height at 500 hPa produces the smallest standard error (se = 0.36) and root-mean-squared error (RMSE = 5.20) for the leave-one-out cross-validated TCG hindcasts over the 40-year record between 1968/89-2007/08. The corresponding correlation coefficient between observed annual TCG totals and cross-validated model hindcasts is r = 0.73. Using *four*-fold cross-validation, model hindcast skill is robust with 85% of the observed seasonal TCG totals hindcast within the model standard deviations. Seasonal TCG totals during ENSO events are typically well captured with RMSE = 5.14 during El Niño and RMSE = 6.04 during La Niña years. The model is shown to be valuable in hindcasting seasonal TCG totals in the Eastern Australian subregion (r = 0.73) and also provides some skill for the Western Australian region (r = 42), while it not useful for the Northern region. In summary, we find that the three-predictor Bayesian model provides substantial improvement over existing statistical TCG forecast models, with remarkably skilful hindcasts (forecasts) of Australian region and subregional seasonal TCG totals provided one month ahead of the TC season.

1. Introduction

Australia's tropical climate is dominated by the El Niño-Southern Oscillation (ENSO), that is driven largely from the Pacific basin (e.g., Allan et al. 1996). The relationship between ENSO and Australian region tropical cyclone formation (genesis, TCG) has been reported extensively (e.g., Nicholls 1984; Basher and Zheng 1995; Kuleshov and de Hoedt 2003; Ramsay et al. 2008; Kuleshov et al. 2009). Ramsay et al. (2008) argue that, next to sea surface temperature (SST), vertical zonal wind shear from 850 hPa to 200 hPa and low-level relative vorticity are the main ENSO-related factors affecting Australian region TCG. Kuleshov et al. (2009) confirmed these results, but added relative humidity in the mid-troposphere as a major contributor. Seasonally, the Australian monsoon trough (Intertropical Convergence Zone) plays an important role in TCG in this region (McBride and Keenan 1982). It has also been shown that the effect of ENSO-linked dynamics on TCG occurs through a strong relationship between the monsoon trough and ENSO via atmospheric bridge processes (Evans and Allan 1992).

Since the late 1970s/early 1980s, a number of statistical seasonal forecast schemes have been developed and improved to predict TC activity in various basins and sub-basins (Klotzbach et al. 2010). In particular, seasonal forecast modeling of TC activity was first undertaken by Nicholls (1979) for the Australian region and Gray (1984) for the North Atlantic. In later studies by Gray et al. (1992, 1994), climatic relationships with hurricane activity in the North Atlantic are based on metrics, such as the Quasi-Biennial Oscillation and African rainfall. A link between intense hurricanes and the Sahel monsoon rainfall was also established (Landsea and Gray 1992). The skill of Gray's operational Atlantic seasonal TC forecasts for the

analyzed period from 1984–2001 relative to climatology and persistence was confirmed and improved (e.g., Owens and Landsea 2003; Saunders and Lea 2005; Klotzbach 2007). Other relevant North Atlantic statistical forecasts include model predictions of hurricane counts using Poisson regression models (e.g., Elsner and Schmertmann 1993; Lehmiller et al. 1997). The Poisson method was later extended using a Bayesian approach to investigate seasonal TC counts and landfall over the USA (e.g., Elsner and Jagger 2004, 2006). This approach has also been used most recently to improve multi-season forecasting of Atlantic hurricane activity (Elsner et al. 2008) and seasonal forecasting of TCs affecting the Fiji, Samoa and Tonga regions (Chand et al. 2010) and the central North Pacific (Chu and Zhao 2007). In the northwest Pacific, projection pursuit regression has been used to forecast seasonal TC totals and associated TC predictands (e.g., Chan et al. 1998; Chan and Shi 1999; Chan et al. 2001). A statistical scheme based on ENSO related indices was later developed for predicting the annual number of TCs making landfall along the south China coast (Liu and Chan 2003). Most recently, modes from an empirical orthogonal analysis of climate factors have been used as predictors of TC behavior in a statistical model also for the South China region (Goh and Chan 2010).

For the Australian region, Nicholls (1979) showed that the austral winter to spring anomalies of sea level pressure at Darwin are highly correlated with early season Australian region tropical cyclone activity, and to a lesser extent with total seasonal TC activity. Subsequent research, and operational testing, confirmed the strong link with the ENSO metric, the Southern Oscillation index (SOI; Nicholls 1984, 1985, 1992; Drosdowsky and Woodcock 1991; Ready and Woodcock, 1992). Solow and Nicholls (1990) presented the first Poisson regression based statistical forecast model for the Australian region. They used the SOI as the predictor of

Australian region total TC counts. More recently, a Poisson regression model using SOI and the September lead saturated equivalent potential temperature gradient between 1000 hPa and 500 hPa was developed to forecast upcoming season TCG totals across the Australian region (McDonnell and Holbrook 2004a,b). This model has also been applied to forecast subregional TCG totals in the eastern Indian Ocean, Northern Australia and southwest Pacific regions (e.g., McDonnell et al. 2006). On intra-seasonal time scales, Leroy and Wheeler (2008) developed a logistic regression model for TC development in the Australian region. As predictors, they used the two dominant varimax rotated modes of SST anomalies for the Indo-Pacific region, as well as an index describing variations in the Madden-Julian Oscillation.

This paper presents a new Australian region statistical seasonal TCG forecasting scheme that shows considerable promise based on a comprehensive assessment of its cross-validated hindcast skill, with high correlations identified between hindcast and observed seasonal TCG counts (r = 0.73) and a low standard error (se = 0.36). This is a substantial improvement in cross-validated hindcast skill over previous studies with correlations between cross-validated hindcasts and observations of TC counts ranging from r = 0.44 to r = 0.60 (e.g., Solow and Nicholls 1990; Nicholls 1992; McDonnell and Holbrook 2004b). Following previous successful studies across different basins (e.g., Elsner and Jagger 2006; Chu and Zhao 2007; Chand et al. 2010), the model developed here is based on the Poisson regression using a Bayesian approach. The Bayesian inference enables us to characterize the uncertainties of the model parameters by a posterior distribution after taking observed data into account. Predictors for the model are carefully selected indices of atmospheric parameters known to affect TC formation. A step-by-step predictor selection based on the RMSE calculated from the cross-validated hindcasts ensures

the best combination of predictors. We show that this model makes significant advances on previous statistical schemes used in the Australian region.

The paper is structured as follows. Section 2 describes the data and variables used. Section 3 introduces the prediction schemes applied as the basis for predictor selection. The model set-up and the techniques used are outlined in section 4. Section 5 presents the model results, and finally section 6 discusses and summarizes the quality and improvements of the models presented over existing models.

2. Data

a. Tropical cyclone observations

This study takes advantage of the global TC best track data set IBTrACS.v02 (Knapp et al. 2010) provided by the U.S. National Oceanic and Atmospheric Administration. TCG is defined to occur when and where a tropical storm system with winds exceeding 34 knots (17.5 ms^{-1}) is first recorded.

The Australian (tropical cyclone) region is defined here as spanning between 0°-30°S and 90°-170°E. Following Dare and Davidson (2004), the Australian region is also divided into three subregions: a Western region from 90°-125°E, a Northern region from 125°-142.5°E, and an Eastern region from 142.5°-170°E (Fig. 1). TCG occurrences identified poleward of 30°S, or over land, have been removed in the quality assessment process. Only storms during the Australian TC season from November to April are taken into account. Overall, a total of 570 TCs during the 40-year period from 1968/69-2007/08 are analyzed following the quality control. Figure 1 shows the spatial distribution of quality assured TCG points and corresponding time series of seasonal TCG totals in the 40-year record.

b. Oceanic and atmospheric data

In this study, SST data were taken from uniformly gridded temperature observations provided in the Hadley Centre Global Sea Ice and Sea Surface Temperature (HadISST1) dataset (Rayner et al. 2003), compiled by the UK Met Office Hadley Centre. HadISST1 is a combination of global monthly SST fields and sea ice concentrations on a 1° × 1° grid from 1870 to the present.

The atmospheric data analyzed in this study are from the NCEP/NCAR monthly mean upper-air reanalyses, with 2.5° horizontal resolution on 17 pressure levels (Kalnay et al. 1996). In total, eight variables were analyzed as potential TCG predictors describing the thermodynamic and dynamic condition of the ocean and atmosphere across a large portion of the Indo-Pacific region from 30°N-50°S, 30°E-70°W. Monthly anomalies of all variables were determined against a 30-year base period of 1970-1999. Statistical significances of the correlation coefficients are based on the reduced effective number of degrees of freedom method outlined by Davis (1976).

c. Thermodynamic and dynamic parameters

1) THERMODYNAMIC PARAMETERS

Four thermodynamic parameters were selected. These are SST, geopotential height at 500 hPa (GPH), convective available potential energy calculated between 850 hPa and 300 hPa (CAPE), and the (non-saturated) equivalent potential temperature gradient between 1000 hPa and 500 hPa (EPT). The ocean temperature, and therefore the potentially available moist convection due to evaporation, is described by the SST (°C or K). To identify the low to mid-troposphere temperature, we examined the GPH (m). The stability of the troposphere, which characterizes the likelihood of deep

convection, is measured by CAPE (m^2 s^{-2}) defined as

$$CAPE = g \int_{z(850hPa)}^{z(300hPa)} \frac{T_m - T}{T} dz \quad , \tag{1}$$

where *z(850hPa)* is assumed the approximate level of free convection and *z(300hPa)* the level of neutral buoyancy. T is a function of pressure level height (z) and is defined as the environmental temperature, T_m is the temperature of an idealized rising air parcel which is assumed to be saturated at the 850 hPa level, and g = 9.81 m s^{-2} is the standard gravity constant.

EPT (K) describes the enthalpy between two layers and, in the present context, the likelihood to form cloud clusters. It is defined as

$$EPT = \nabla \left[\Theta \exp\left(\frac{L_v q_v}{c_p T} \right) \right]_{z(1000hPa)}^{z(500hPa)} \quad , \tag{2}$$

with the vapor pressure

$$q_v = \frac{RHe_{str} \exp\left(\frac{L_v}{R_v} \left(\frac{1}{T_{tr} - T} \right) \right)}{100} \quad , \tag{3}$$

where T (K) is the temperature and RH (%) the relative humidity – with both variables dependent on the pressure level. The constants are: the latent heat of vaporization L_v = 2.5 x 10^6 J kg^{-1}, the specific heat at constant pressure c_p = 1004 J kg^{-1} K^{-1}, the saturation vapor pressure at triple point e_{str} = 6.11 hPa, the gas constant for vapor R_v = 461 J kg^{-1} K^{-1}, and the triple point temperature T_{tr} = 273.16 K. Relatively moist layers in the lower- and mid-troposphere (RH between 700 hPa and 500 hPa) are essential, as dry mid levels suppress the continuing development of widespread deep convection – i.e., EPT is highly dependent on RH.

2) DYNAMIC PARAMETERS

Four dynamic parameters were also analyzed. These are the zonal and meridional winds at 850 hPa (u_{850}, v_{850}), the environmental vertical wind shear between 850 hPa and 200 hPa (EVWS), and the relative vorticity at 850 hPa (RV). The 850 hPa pressure level was chosen as the lower dynamic level to avoid effects of the boundary layer and thus focus on atmospheric interior geostrophic flows. It is imperative that EVWS (m s^{-1}), defined as

$$EVWS = \sqrt{\left(u_{200} - u_{850}\right)^2 + \left(v_{200} - v_{850}\right)^2} \quad , \tag{4}$$

is weak, otherwise convection within the TC eyewall cannot develop or persist. Also an existing negative (Southern Hemisphere) RV - such as a small cyclonic atmospheric disturbance, a tropical wave, or a monsoonal trough with convergence - is a necessary initial factor to develop a TC. RV (s^{-1}) is defined as:

$$RV = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \quad , \tag{5}$$

where u,v (m s^{-1}) are the east and north components of velocity, and x,y (m) are the east and north Cartesian displacement directions respectively. Attendant strong divergence supports the development of deep convection, which in turn intensifies the disturbance into a low-pressure system.

d. ENSO definitions and effects on TCG

Ramsay et al. (2008) found the NINO4 SST anomaly (SSTA) index to be the strongest ENSO predictor of interannual TC frequency in the Australian region and is therefore also included in the present study. NINO4 is defined as the SSTA time series averaged spatially between 5°S-5°N and 160°E-150°W.

ENSO events are classified according to the definition used by the U.S. National Weather Service

(<http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml>) using the three-month running mean in the NINO3.4 region. El Niño (La Niña) events are defined by the NINO3.4 SSTA exceeding thresholds of ±0.5°C for a minimum of five consecutive three-month average overlapping periods (see APPENDIX 1).

3. Prediction schemes

Figure 2 shows spatial correlation maps between Australian region annual TCG totals and the individual predictor variables shown for the June-July-August (JJA) period prior to the upcoming TC season. The correlation patterns with the thermal variables EPT, CAPE and SST (see corresponding panels in Fig. 2) are characterized by the developing ENSO (e.g., Drosdowsky and Chambers 2001) 'boomerang' pattern across the tropical/subtropical Pacific. This pattern defines regions of enhanced convection, contrasted by suppressed convection over the Indian Ocean. GPH shows a strong negative correlation pattern throughout the tropics that is maximized in the central Indian Ocean. The pattern is consistent with variations in the Intertropical Convergence Zone (ITCZ) whereby colder lower- and mid-troposphere air masses support convection and the development of TCs. The positive correlations in the central subtropical South Pacific are located at the southern tip of the South Pacific Convergence Zone and the warm advection region of the semi-permanent South Pacific subtropical High. However, interannual variations of the air column in that region are mainly driven by changes of the trough in the mid-latitude westerlies between 15°S and 45°S in the central South Pacific (Van Loon and Shea 1985). The dynamic variables tend to show similar patterns as CAPE, EPT and SST, with a change of sign at the eastern boundary of the Australian TC region (bottom four

panels of Fig. 2). These patterns describe changes in the Walker and Hadley circulations due to changes in ENSO phases. The correlation pattern in the tropical Pacific and Indian Ocean between annual TCG totals and u_{850} imply enhanced TCG in the Australian region with a strengthening of the Walker circulation and trade winds. The v_{850} correlation patterns North of the equator in the far eastern Pacific and western Indian Ocean suggest enhanced TCG with increased meridional surface inflow into the equatorial regions of the eastern and central Pacific, but also into the tropical western Indian Ocean. In contrast, the correlation patterns of v_{850} also describe enhanced TCG with weakening of the Hadley circulation in the Pacific warmpool and West Pacific regions. The correlation pattern of EVWS and RV with Australian region TCG totals mostly reflects the correlation pattern of u_{850} over the tropical central Pacific. Our results imply a stronger (weaker) Walker circulation leading to increased (decreased) convection in the Western Pacific Warm Pool area, which is more (less) favorable to Australian region TCG.

To achieve climatologically relevant predictors, we derived indices for each climate predictor variable on the basis of persistence and strength of pre-seasonal spatial correlation coefficients. Spatial correlation maps for October-November-December (OND; Fig. 3) help to identify patterns, which continue to be significant after the austral winter till the start of the Australian TC season. The most persistent correlations observed for CAPE, EPT and GPH are all located in the subtropical central South Pacific, while the used SSTA index (NINO4), u_{850} and EVWS describe variations in the equatorial central Pacific region. The v_{850} and RV indices correspond to the northern inflow region of the Walker cell in the tropical East Pacific.

Figure 4 presents the correlation coefficients from all three-month overlapping periods from January-February-March (JFM), ahead of the following TC season,

through to OND at the start of the TC season. We found that all thermal predictors generate a strong increase of correlations from March-April-May (MAM) to JJA and stay highly correlated with around r = +/-0.5 to the start of the TC season. Correlation coefficients between annual TCG totals and CAPE, EPT and NINO4 are all statistically significant (at the 95% confidence level) from early spring on, while correlation coefficients with GPH only reach 95% significance by May-June-July (MJJ). The two wind component indices generate increased correlations with annual TCG totals from April-May-June (AMJ; v_{850}) and MJJ (u_{850}), and peak during MJJ and June-August-September (JAS), respectively. Correlations between TCG totals and u_{850} are statistically significant from AMJ while the ones with v_{850} only reach correlation coefficients significant on 90% confidence level from AMJ on. EVWS shows a strong increase in the correlation coefficients from AMJ to JJA (significant at the 95% confidence level from MJJ on), and peaks during JAS, slowly decreasing thereafter. RV reaches its maximum correlations during JAS (significant at the 95% confidence level from AMJ on) and reduces subsequently. The definition of each predictor index, as well as the three-month mean of each index best relating to variations in TCG totals during the upcoming season, are provided in Table 1. In total, eight predictor indices were further investigated. Possible combinations of predictors should be complementary and contribute information that adds value. For that reason, predictor combinations with correlations of more than ± 0.8 with each other were not included in any further analysis due to their collinearity. Here, collinearity was observed between RV and v_{850}, as well as between CAPE and EPT.

4. Bayesian Regression Model

a. Poisson Regression

The Poisson distribution is often used to model the occurrence of rare, discrete events, such as tornado counts and the occurrences of droughts or cold spells (e.g., Wilks 1995). The Poisson distribution also restricts the possible outcomes to non-negative integers, making it ideal for modeling tropical cyclone occurrences (Elsner and Schmertmann 1993).

Following previous studies (e.g., Solow and Nicholls 1990; Elsner and Schmertmann 1993; McDonnell and Holbrook 2004a,b; Elsner and Jagger 2006; Chu and Zhao 2007; Chand et al. 2010), we applied a Poisson regression approach to model TCG totals satisfying

$$P(Y_i = y) = \frac{\mu_i^{\,y} \exp(-\mu_i)}{y!} \quad , \mathrm{y} = 0,1,2,\ldots,\infty \, , \tag{6}$$

where

$$\mu_i = \exp(\beta_o + \sum_j (\beta_j x_{ij})) \quad . \tag{7}$$

Here, if Y has a Poisson distribution, the logarithm of the expected number of TCG occurrences, μ, can be modeled as a linear combination of the predictors x_{ij}, with j being the specified predictor during the season i. β_j is the corresponding Poisson regression coefficient, β_0 the intercept and y the observed TCG count. In a Poisson model, the variance is equal to the mean (μ), with the standard deviation being the square root of μ.

b. Bayesian analysis

Bayes' theorem represents a quantification of uncertainty provided by probabilities. By comparison, in the frequentist approach, probabilities are seen in terms of frequencies of random repeatable events (Wilks 1995). In this study, the Bayesian approach is used to predict the seasonal number of TCG occurrences. The Bayes'

theorem is applied to find the best possible model coefficient representation and this information is then used to predict the seasonal TCG totals. The observed predictor set is denoted as $x_{1:T} = \{x_1,\ldots,x_T\}$ and the corresponding seasonal number of TCG occurrences, or TCG probability, as $y_{1:T} = \{y_1,\ldots,y_T\}$ during the observations 1:T. We are interested to find the conditional probability of y_{T+1} given x_{T+1} and the model coefficients β, $p(y|x_{T+1}, \beta)$. The model coefficients, β, will be estimated using the posterior distribution $p(\beta\,|x_{1:T}, y_{1:T})$.

The assumptions about the prior knowledge of the model coefficients are stated before observing the data x_i and y_i in the form of a prior probability $p(\beta)$. As we have no, or only little, prior information on the climatic effects of our chosen predictors on TCG occurrences, we chose the conservative way of defining the prior probability of the model coefficients as almost flat priors. The priors are defined as a Gaussian distribution

$$p(\beta_j) = N(\beta_j \mid \mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(\beta_j - \mu)^2\right), \tag{8}$$

with the mean selected as μ = 0 and the standard deviation σ =100, with β_j representing the j model coefficients. For the TCG occurrences model, we consider the Poisson distribution

$$p(y_i \mid x_i, \beta) = Poiss(y_i \mid x_i, \beta). \tag{9}$$

Following Bayes' rule we get the posterior distribution

$$p(\beta \mid x_{1:T}, y_{1:T}) \propto p(\beta)\prod_{i=1}^{T} p(y_i \mid x_i, \beta). \tag{10}$$

The posterior probability (Equation 10) allows us to take uncertainties into account and predictions are then obtained from

$$p(y_{T+1} \mid x_{T+1}, \beta) = \frac{1}{N}\sum_{i=1}^{N} p(y_{T+1} \mid x_{T+1}, \beta), \tag{11}$$

with N being the total number of obtained samples from the posterior distribution.

To usefully apply the Bayesian approach, and obtain the appropriate values for the model coefficients, the use of a sampling method like the Markov Chain Monte Carlo (MCMC; Hastings 1970) method is indispensable (Larget and Simon 1999). The MCMC simulates direct draws from a probability distribution. There, the previous sample values generate randomly the next sample value, which means generating a Markov chain. Unlike previous studies (e.g., Elsner and Jagger 2004, 2006; Chu and Zhao 2007; Chand et al. 2010), which applied the Gibbs sampler via the open source software WinBUGS, we instead used the multivariate slice sampler (Neal 2003), which is a form of auxiliary variable technique (Roberts and Rosenthal 1999; see APPENDIX 2). The model standard deviation was then calculated from the expected number of events $E(y_{T+1} \mid x_{T+1}, \beta)$ as estimated using the average of the mean μ obtained from the MCMC samples. The model standard deviation σ is then defined as

$$\begin{aligned} \sigma(y_{T+1} \mid y_{1:T}) = & E[\sigma(y_{T+1} \mid x_{T+1}, \beta) \mid y_{1:T}] \\ & + \sigma[E(y_{T+1} \mid x_{T+1}, \beta) \mid y_{1:T}] \end{aligned}, \tag{12}$$

where the first term is the average process standard deviation and the second term is the coefficient uncertainty (Peters et al. 2009). The hindcasted TCG totals are taken as the number with the maximum probability from the hindcast distribution.

To discard the effects of the chosen initial conditions we applied a model burn-in of 500 iterations. This relatively short iterative burn-in achieves a quick convergence (see Fig. 5). Also, to avoid high autocorrelations and gain statistically independent samples out of the iteration process, the samples were thinned so that only every fifth sample was taken into account. Finally 5,000 samples were used to estimate the model coefficients and obtain the predictions.

c. Model skill

To compare the skill of the possible model predictors, a leave-one-out cross-

validation (e.g., Stone 1974; Elsner and Schmertmann 1993) was performed. In this method, the model gets trained using n-1 seasons to hindcast the number of storms expected for the one season that has been withheld from the training data set. The train-and-test approach is successively repeated to hindcast every season across the 40-year data set. This enables us to perform an independent hindcast of every season. For a better understanding of the robustness of the model results over time, we also performed a *k*-fold cross-validation technique with $k = 4$. The method splits the data into *k* equal-sized subsets (e.g., Stone 1974; Efron and Gong 1983). For the *k*th subset, the model is developed using the other *k*-1 data subsets, and then the fitted model is used to predict the *k*th data subset. To validate the skill of the models, the root-mean-square error (RMSE) of each model hindcast was calculated (e.g., Elsner and Jagger 2006; Chand et al. 2010). The MSE is defined as

$$MSE = \frac{1}{n}\sum_{i}\sum_{k} p_i(k)\left(k - o_i\right)^2 \quad , \tag{13}$$

where n is the total number of seasons and i is the particular season being hindcast and used for cross validation, $i = 1,\ldots,n$. p_i is the predicted probability that k TCs develop in the hindcast season i, and o_i is the number of TCs that were actually observed to form in that season. The RMSE is a commonly used metric for the potential utility of a predictor or predictor combination in a probabilistic model, with small values indicating a good model. It is calculated using the probabilities of the independent hindcasts in the leave-one-out cross-validation method.

The standard error (se) and the cross-correlation (r) between the predicted and observed number of TCs formed in each region provide a measure of the overall hindcast skill at the first-order level of the seasonal time series. The final model coefficients were estimated based on data from the training period between 1968/69-

2007/08.

5. Results

a. Model predictor selection

The kernel distributions of the estimated posterior densities of the model coefficients aid verification of the likely merit of the predictors. They give an indication of the quality of each of the tested predictors. In the ideal case, all sampled coefficients lie on either side of the zero line. This shows that the chosen predictor is playing a significant role in predicting the events given the sign of the chosen predictor.

The posterior densities of the climatology and the tested single-predictors over the 40-year record are shown in Figure 6. We find that all indices in the present study are suitable as TCG predictors. Using the leave-one-out cross-validation technique, we calculated the RMSE to help evaluate the model skill. The climatology-only model was determined as the uncertainties of the intercept and hindcasts of TCG numbers are close to the observational long-term average, with a high RMSE = 6.20 and se = 0.38. The climatology-only model does not contain any predictive power beyond the background state. The best single-predictor is CAPE (RMSE = 5.39, se = 0.33), while the predictor with the least utility as a single predictor was found to be GPH (RMSE = 5.70, se = 0.41; Table 2a).

In an attempt to further improve the model, we investigated predictor combination models using a step-by-step predictor selection based on the calculated RMSEs. CAPE was used as the key single-predictor based on its lowest RMSE and strongest correlation with TCG. With CAPE as the base, we found that the two-predictor combinations, CAPE+v_{850}, CAPE+EVWS and CAPE+RV provided further reductions to the RMSE to 5.21, 5.33 and 5.21 respectively. The three-predictor

model CAPE+v_{850}+GPH provided the lowest errors, with RMSE = 5.20 and se = 0.36 (Table 2b/3a). Figure 7 shows the 40-year leave-one-out cross-validated (hereafter CV40) hindcasts for that model plotted against the observed total number of TCs formed in each season. The hindcasted annual TCG count is the number of TCG occurrences with the maximum probability in the hindcast distribution as obtained by Equation (11). We find that the CAPE+v_{850}+GPH model captures the variability in number of cyclones formed within its boundaries of standard deviation, with 80% success rate of the CV40 hindcasts and very favourable performance against the observed total of TCG occurrences, with the correlation coefficient being r = 0.73 (Table 2b). This is a substantial improvement in cross-validated hindcast skill of at least 21.5% over previous models with correlations between cross-validated hindcasts and observations of TC counts ranging from r = 0.44 to r = 0.60 (e.g. Solow and Nicholls 1990; Nicholls 1992; McDonnell and Holbrook 2004b).

b. Four-fold cross-validation

For a better evaluation of the robustness of the predictor-set, we applied a *four*-fold cross-validation technique. In this method, the data are split into *four* consecutive 10-year subsets. The model is then trained on three of the four subsets to hindcast the left-out 10 years. This procedure was used to hindcast the 10-year periods 1968/69-1977/78, 1978/79-1987/88, 1988/89-1997/98 and 1998/99-2007/08. The RMSEs for each independent 10-year hindcast were calculated for the CAPE+v_{850}+GPH model and compared to the average RMSE of the CV40 hindcast results (Figure 8a). We found that the RMSEs calculated on the hindcasts did not deviate strongly from the CV40 hindcasts. We also found that the lowest RMSE corresponded to the 1988/89-1997/98 10-year hindcast. The highest RMSE was for the 1968/69-1977/78 decade.

Figure 8b shows the annual RMSEs for the CAPE+v_{850}+GPH model as obtained from the *four*-fold cross validated hindcast distributions. We note that there is a slight trend towards reduced errors of the model hindcasts over time. The model captures ENSO seasons quite well, except for the 1985/86 La Niña event year (RMSE = 7.67), when the TCG totals of 19 storms were substantially underestimated and the 1987/88 El Niño event year (RMSE = 7.44) in which the low TCG occurrences of 6 was not reproduced. The highest RMSEs can be observed during the La Niña of 1971/72 (RMSE = 9.47), in which a rather low number of events (15) was observed and 1974/75 (RMSE = 8.26), for which the RMSE is particularly high due to the high number of TCG and therefore broad hindcast probability. The other periods showing very high RMSEs (> 7) are the underestimated ENSO-neutral years of 1979/80 and 1993/94 and the 1973/74 La Niña year which had a high number of TCG occurrences and broad hindcast probability.

Figures 9 and 10 present the hindcast distributions obtained from the *four*-fold cross-validation. We find that the hindcast distributions of Australian region TC seasons with an expected high number of TCG events are broader than the ones for seasons with a lower number of occurrences hindcasted. This is particuarly the case in the first hindcast decade and explains the high RMSE for the 1968/69-1977/78 decade, in which six out of 10 seasons had 19 TCG occurrences or more. Also during this decade, we observe large values of standard deviation, especially in the process uncertainty term (not shown), indicating lower skill of the model hindcasts. Figure 11 shows the four independent *four*-fold 10-year hindcasts, as well as the fitted 30-year hindcasts on which the model was trained. The results are compelling with the model displaying considerable skill. The independent hindcasts of the left-out 10 years are mostly very accurate, with only six out of 40 (15%) of the seasonal hindcasts outside

of the model standard deviation. The model lacks skill when an unusually high number of TCGs occurred during ENSO-neutral or El Niño conditions (1979/80, 1993/94; see Figs. 9 and 10). Another issue is correctly capturing the very low number of cyclones during the identified unusual weak seasons (1982/83, 1987/88, 2003/04) - a recognized issue in trying to capture extremes using regression modelling. Only the 1985/86 La Niña event year were not well presented by the model.

In the following section, we demonstrate that despite ENSO years being by their nature somewhat 'extreme', the model performs very well in capturing the TCG occurrences across most of these events.

c. ENSO years

To assess the skill of hindcasting TCG totals during strong El Niño and La Niña years between 1968/69-2007/08, we evaluate the model performance based on the CV40 hindcasts (refer Fig. 7) for the CAPE+v_{850}+GPH model. We find the model performs very well during ENSO event years (Figure 8a), with RMSEs only little higher than the 40-year average. The model skilfully hindcasts the low TCG numbers during El Niño years (RMSE = 5.14), as well as the high TCG numbers during La Niña years (RMSE = 6.04; Fig. 8). Figure 12 shows the hindcast distribution for TCG totals using the CAPE+v_{850}+GPH model for all El Niño and La Niña event years. The model accurately hindcasts the number of TCG occurrences during El Niño years (79% success) and is particularly skilful in hindcasting TCG totals during La Niña event years, with 84% of the La Niña year TCG hindcasts falling within hindcast standard deviation statistics. We note that the higher RMSE for La Niña compared to El Niño

years results from a broader hindcast distribution of the annual hindcasts due to the higher numbers of TCG occurrences, rather than a poor performance of the model.

d. Subregional hindcasts

We also examined the skill of the CAPE+v_{850}+GPH model for the three Australian subregions: in the West, North and East. To achieve that aim, the model was trained on the subregional seasonal TCG totals with the model skills verified using the CV40 approach (Fig. 13). In Table 3 the subregional model coefficients, as well as their standard deviations, are listed. We find that the model performs well at the sub-regional scale for the Western and Eastern Australian region with 87.5% of the observed seasonal TCG counts being hindcast within the model standard deviation. For the Northern region, TCG hindcast skill was found to be little better than climatology-only. Due to the different averages of annual TCG counts in each region, the RMSEs and standard errors cannot be used as a comparative measure, as both are directly linked to the average or the standard deviation of the TCG totals and hence have different amplitudes for each subregion. Correlating the CV40 hindcasts with the observed seasonal TCG totals for each subregion however provides a measure of the model's skill in capturing the correct phase of the hindcast variability. Correlation coefficients in the Eastern region are as high as r = 0.73 (se = 0.22). For the Western region, the model is also strong with correlations between the hindcasts and observations of r = 0.42 (se = 0.27). The Northern region TCG totals, however, are hindcast poorly with model outputs varying marginal from the average TCG occurences (Fig. 13b).

6. Summary and Discussion

This study represents a substantial improvement in the potential for more accurate statistical seasonal forecasting of tropical cyclone formation (genesis, TCG) for the Australian region. It is well understood that TCs in the Australian region are strongly affected by the phase of El Niño-Southern Oscillation (ENSO). Nevertheless, models based purely on ENSO metric predictors typically fail to forecast seasonal TCG totals in both the eastern Indian and southwest Pacific Ocean regions.

Our approach comprises of an extended analysis using well-known climate indices together with the identification of new and potentially skilful indices that represent metrics important for predicting TC formation. These include: three subtropical central South Pacific indices of the convective available potential energy between 850 hPa and 300 hPa (CAPE), the (un-saturated) equivalent potential temperature gradient between 1000 hPa and 500 hPa (EPT) and geopotential height at 500 hPa (GPH); two tropical central Pacific indices of the zonal winds at 850 hPa (u_{850}) and environmental vertical wind shear between 850 hPa and 200 hPa (EVWS); and, two tropical northeast Pacific indices including the meridional wind at 850 hPa (v_{850}) and low-level relative vorticity at 850 hPa (RV; see Fig. 3). Additionally, the ENSO sea surface temperature anomaly (SSTA) index NINO4 was used. As a result of correlating pre-seasonal three-month climate index means with TCG totals in the upcoming season, we found that three-month austral winter mean indices showed the greatest overall potential for forecasting Australian region TCG totals in the upcoming TC season. The best eight predictor variables (Table 1) were incorporated into a Poisson regression model developed on TCG totals. Following recent studies (e.g., Elsner and Jagger 2004, 2006; Chu and Zhao 2007; Elsner et al. 2008; Chand et al. 2010), we applied a Bayesian approach using the Markov Chain Monte Carlo method. The Bayesian approach is beneficial as it allows for incorporation of prior beliefs, and

is convenient to account for the uncertainties in model parameters. The final model runs generated a total of 5,000 independent samples using a slice sampler with a burn-in of 500 and a thinning of five.

Using a three-predictor Bayesian model on key indices of CAPE, v_{850} and GPH, we have been able to create a substantial improvement in cross-validated hindcast skill over previous studies. Previous published studies produced correlations between cross-validated hindcasts and observations of TCG counts ranging from r = 0.44 to r = 0.60 (e.g. Solow and Nicholls 1990; Nicholls 1992; McDonnell and Holbrook 2004b). Nicholls (1992) also found correlations of r = 0.72 between September-October-November (SON) values of the Southern Oscillation Index and the 'first differences' of consecutive seasons instead of the total number of TC counts between 1959/60 and 1990/91. An earlier study by Nicholls (1984) showed correlations of r = 0.78 between SON SST in a region North of Australia (5°–15°S, 120°–160°E) and Australian region TC counts from 1964-1982. However this relationship has been shown to degrade over time and is not robust for the more recent years (Ramsay et al. 2008). With the leave-one-out cross-validated hindcast approach we achieve correlations of r = 0.73 and a standard error of se = 0.36 using a three-predictor model of CAPE+v_{850}+GPH for the period 1968/69-2007/08. In comparison to the skill of the climatology-only model, the CAPE+v_{850}+GPH model shows a 19% improvement in RMSE. Using a *four*-fold cross-validation approach, we achieved highly skilled hindcasts for entire decades that were left out, highlighting the robustness and potential skill of the model.

The CAPE, v_{850} and GPH indices derived in this study are ENSO-linked. Combining convective available potential energy and geopotential height from the subtropical central South Pacific with the lower-troposphere meridional tropical

inflow from the Northern Hemisphere in the eastern Pacific, was found to produce a valuable and complementary predictor set. Figure 14 shows the anomalies of all model-relevant predictor variables (CAPE, wind flow at 850 hPa, GPH) between active TC seasons (TCGs ≥ 17) and rather inactive TC seasons (TCGs ≤ 11) and indicates the index locations used in the three-predictor model. Active and inactive seasons were here chosen to represent around 25% of the investigated seasons, respectively. The CAPE index is located in the subtropical central South Pacific and embedded in the South Pacific Convergence Zone of ENSO-related convection. CAPE, as a measure of the instability of the atmosphere, provides a likelihood metric of deep convection within the zone of TC formation. The relationship between TCG occurrence and the ITCZ-linked North Pacific index, v_{850}, is negative, which means a strengthening of the low-level equatorward trade wind inflow (Walker circulation) leads to enhanced TCG. Finally, the positive sign of the GPH model coefficient, as an indication for heating and atmospheric layer expansion, relates warmer airmasses in the lower and mid-troposphere of the subtropical central Pacific during austral winter to enhanced TCG in the following season. The GPH index is located in the warm advection zone of the South Pacific subtropical High. Van Loon and Shea (1985) identified this region to be linked to the Southern Oscillation. In autumn (fall) and winter prior to a warm event, the trough in the westerlies, located west of New Zealand, fails to develop to its usual amplitude, allowing colder air from the mid-latitudes intrude the Australian Pacific region. These conditions are suppressed the year after a warm event took place.

Based on the assessment of various quality controls in the leave-one-out and *four*-fold cross-validated hindcasts, we found that the three-predictor CAPE+v_{850}+GPH Poisson model is the most skilful in our model suite. The lowest

RMSE corresponded to the 1988/89-1997/98 decade (RMSE = 4.41), while the largest corresponded to the 1968/69-77/78 decade (RMSE = 6.34). During ENSO event years (El Niño or La Niña), the three-predictor model demonstrated high skill, performing generally very well during El Niño, with the RMSE slightly below the average (RMSE(CV40) = 5.20; RMSE(El Niño) = 5.14). The higher than average RMSE during La Niña (RMSE(La Niña) = 6.04) is due to a broader hindcast distribution of high numbers of TCG totals rather than a deficiency in the model. In fact, all except two of the hindcasts during La Niña event years captured the observed seasonal TCG totals within its standard deviation. Further, the independent hindcasts of the four separate decades using *four*-fold cross-validation is impressive, with only six seasons (15%) out of a total of 40 being outside of the model standard deviation. We believe these results are quite compelling.

Finally, we tested the three-predictor model on three smaller subregions, West, North and East. We found that the model is adaptable to forecast seasonal TCG totals in the Eastern region (r = 0.73), while there is also some skill for TCG in the Western region (r = 0.42). For the Northern region, however, the hindcasts are not very different from the average number of seasonal TCG occurrences adding up to correlations between hindcasts and observed TCG totals of r = -0.23. The regional scale complexities associated with this shallow sea region are problematic for our large-scale model. It may also be, that a component of the poor performance in this Northern region is due to northern Australian TCG being inherently unpredictable for statistical schemes.

In summary, we have developed a new and potentially skilful seasonal forecast model of tropical cyclogenesis for the Australian region. We find that a three-predictor CAPE+v_{850}+GPH Poisson model produces remarkably skilful hindcasts of

Australian region seasonal TCG totals by September of each year, one month prior to the onset of the Australian region TC season (November-April). The predictor variables identified in this study are physically meaningful and appropriate to condition the model forecasts of TCG. By combining information from useful dynamic and thermal variables as predictors in a Bayesian approach Poisson model system, we are able to demonstrate skilful cross-validated hindcasts of Australian region seasonal TCG totals with high correlation (r = 0.73) and low RMSE (5.20) against a 40-year record of observations.

Acknowledgements: A. Werner was supported by a MQRES Scholarship from Macquarie University, Australia. AW also thanks the School of Geography and Environmental Studies at the University of Tasmania for providing computing resources and office space during her visits to UTAS. Finally, we wish to thank Dr. John McBride and Dr. Pavel Shevchenko for their constructive comments and helpful discussions on this work.

APPENDIX 1

ENSO event years

This study defined ENSO events as defined by the Climate Prediction Center of the National Weather Service from three-month SSTA means in the NINO3.4 region.

El Niño: {1968/69, 1969/70, 1972/73, 1976/77, 1977/78, 1982/83, 1986/87, 1987/88, 1991/92, 1994/95, 1997/98, 2002/03, 2004/05, 2006/07}

La Niña: {1970/71, 1971/72, 1973/74, 1974/75, 1975/76, 1983/84, 1984/85, 1985/86, 1988/89, 1995/96, 1998/99, 1999/2000, 2000/01, 2007/08}

APPENDIX 2

Slice sampling

The slice sampler (Neal 2003) avoids specifying the proposal densities as in Metropolis-Hastings algorithms (e.g., Hastings 1970; Gelman 1992). In that way, after finding the appropriate augmentation scheme, the method can be applied to all data sets without any computational difficulties (Ntzoufras 2009). The slice sampler is defined by

$$\hat{p}(\beta,u) = \begin{cases} \dfrac{1}{Z_p} & \text{if } 0 \le u \le \tilde{p}(\beta) \\ 0 & \text{otherwise} \end{cases}, \tag{A1}$$

where $Z_p = \int \tilde{p}(\beta)d\beta$. The marginal distribution over β is given by

$$\int \hat{p}(\beta,u)du = \int_0^{\tilde{p}(\beta)} \frac{1}{Z_p} du = \frac{\tilde{p}(\beta)}{Z_p} = p(\beta) \tag{A2}$$

so we can sample from $p(\beta)$ by sampling from $\hat{p}(\beta,u)$ and then ignoring the u values. Given the value of β we evaluate $\tilde{p}(\beta)$ and then sample u uniformly in the range $0 \le u \le \tilde{p}(\beta)$. After u is uniformly fixed from the 'slice' through the distribution defined by $\{\beta: \tilde{p}(\beta) > u\}$. Slice sampling is applied to multivariate distributions by repeatedly sampling each of the n variables in turn, in the manner of the Gibbs sampling (Bishop 2006), where one needs n iterations to get from $\beta_j^{(i)}$ to $\beta_j^{(i+1)}$

$$\left.\begin{array}{l} 1: \beta_1^{(i+1)} \sim p(\beta_1 \mid \beta_2^{(i)}, \beta_3^{(i)}, ..., \beta_n^{(i)}) \\ 2: \beta_2^{(i+1)} \sim p(\beta_2 \mid \beta_1^{(i+1)}, \beta_3^{(i)}, ..., \beta_n^{(i)}) \\ \quad \vdots \\ j: \beta_j^{(i+1)} \sim p(\beta_j \mid \beta_1^{(i+1)}, \beta_2^{(i+1)}, ..., \beta_{j-1}^{(i+1)}, \beta_{j+1}^{(i)}, ..., \beta_n^{(i)}) \\ \quad \vdots \\ n: \beta_n^{(i+1)} \sim p(\beta_n \mid \beta_1^{(i+1)}, \beta_2^{(i+1)}, ..., \beta_{n-1}^{(i+1)}) \end{array}\right\}. \tag{A3}$$

REFERENCES

Allan R., J. Lindesay, and D. Parker, 1996: *El Niño: Southern Oscillation and Climatic Variability*. CSIRO Publishing, 416pp.

Basher, R. E., and X. Zheng, 1995: Tropical cyclones in the southwest Pacific: Spatial patterns and relationships to Southern Oscillation and sea surface temperature. *J. Climate,* **8,** 1249-1260.

Bishop, C. M., 2006: *Pattern Recognition and Machine Learning*. Springer, 738pp.

Broadbridge, L. W., and B. N. Hanstrum, 1998: The relationship between tropical cyclones near Western Australia and the Southern Oscillation Index. *Austr. Meteorol. Mag.*, **47**(3), 183-89.

Chan, J. C. L., J. Shi, and C. M. Lam, 1998: Seasonal forecasting of tropical cyclone activity over the western North Pacific and the South China Sea. *Wea. Forecasting*, ***13*****,** 997–1004.

Chan, J. C. L., and J. E, Shi, 1999: Prediction of summer monsoon rainfall over South China. *Int. J. Climatol.,* **19,** 1255-1265.

Chan, J. C. L., J. E. Shi, and C. M. Liu, 2001: Improvements in the seasonal forecasting of tropical cyclone activity over the western North Pacific. *Wea. Forecasting,* **16,** 491–498.

Chand, S., K. J. E. Walsh, J. C. L. Chan, 2010: A Bayesian Regression Approach to

Seasonal Prediction of Tropical Cyclones Affecting the Fiji Region. *J. Climate,* **23,** 3425-3445.

Chu, P.-S., and X. Zhao, 2007: A Bayesian regression approach for predicting tropical cyclone activity over the central North Pacific. *J. Climate,* **20,** 4002–4013.

Dare, R. A., and N. E. Davidson, 2004: Characteristics of Tropical Cyclones in the Australian Region. *Mon. Wea. Rev.*, **132,** 3049-3065.

Davis, R. E., 1976: Predictability of sea surface temperatures and sea level pressure anomalies over the North Pacific Ocean. *J. Phys. Oceanogr.,* **6,** 249-266.

Drosdowsky, W., and L. E. Chambers, 2001: Near global scale sea surface temperature anomalies as predictors of Australian seasonal rainfall. *J. Climate,* **14,** 1677-1687.

Drosdowsky, W. and F. Woodcock, 1991: The South Pacific and southeast Indian Ocean tropical cyclone season 1988-1989. *Aust. Met. Mag.,* **39,** 113-129.

Efron, B. and G. Gong, 1983: A leisurely look at the bootstrap, the jackknife, and cross- validation. *The American Statistician*, **37,** 36-48.

Elsner, J. B., and C. P. Schmertmann, 1993: Improving extended-range seasonal predictions of intense Atlantic hurricane activity. *Wea. Forecasting*, **8,** 345–351.

Elsner, J. B., and T. H. Jagger, 2004: A hierarchical Bayesian approach to seasonal hurricane modeling. *J. Climate,* **17,** 2813–2827.

Elsner, J. B., and T. H. Jagger, 2006: Prediction models for annual U.S hurricane counts. *J. Climate,* **19,** 2935–2952.

Elsner, J. B., T. H. Jagger, M. Dickinson, and D. Rowe, 2008: Improving Multiseasonal Forecast of North Atlantic Hurricane Activity. *J. Climate,* **21,** 1209-1219.

Evans, J. L., and R. L. Allan, 1992: El Nino/Southern Oscillation modification to the structure of the monsoon and tropical cyclone activity in the Australasian region. Int. *J. Climatol.,* **12,** 611-623.

Gelman, A., 1992: Iterative and Non-Iterative Simulation Algorithms. *Computing Science and Statistics (Interface Proceedings)*, **24,** 433-438.

Goh, A. Z. C., and J. C. L. Chan, 2010: An improved statistical scheme for the prediction of tropical cyclones making landfall in South China. *Wea. Forecasting,* **25,** 587-593.

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

Gray, W. M., 1975: Tropical cyclone genesis. *Dept. of Atm. Sci. Paper No. 234, Colorado State University, Fort Collins, CO,* 121 pp [Available from Department of Atmospheric Sciences, Colorado State University, Ft. Collins, CO 80523].

Gray, W. M., 1984: Atlantic seasonal hurricane frequency. Part II: Forecasting its variability. *Mon. Wea. Rev.,* **112,** 1669–1683.

Gray, W. M., C. W. Landsea, P. W. Mielke Jr., and K. J. Berry, 1992: Predicting Atlantic seasonal hurricane activity 6–11 months in advance. *Wea. Forecasting,* **7,** 440–455.

Gray, W. M., C. W. Landsea, P. W. Mielke Jr., and K. J. Berry, 1994: Predicting Atlantic basin seasonal tropical cyclone activity by 1 June. *Wea. Forecasting,* **9,** 103-115.

Hastings, W. K., 1970: Monte Carlo Sampling Methods Using Markov Chain and Their Applications. *Biometrica,* **57,** 97-109.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437-471.

Klotzbach, P. J., 2007: Revised prediction of seasonal Atlantic basin tropical cyclone activity from 1 August. *Wea. Forecasting,* **22,** 937-949.

Klotzbach P. J., A. Barnston, G. Bell, S.J. Camargo, J.C.L. Chan, A. Lea, M. Saunders, and F. Vitart, 2010: Seasonal Forecasting of Tropical Cyclones. *Global Guide to Tropical Cyclone Forecasting, 2nd edition*, C. Guard, editor, WMO, *in press*.

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

Kuleshov, Y., and G. de Hoedt, 2003: Tropical cyclone activity in the Southern

Hemisphere. *Bull. Austr. Met. Oceanogr. Soc.,* **16,** 135-137.

Kuleshov, Y., F. Chane Ming, L. Qi, I. Chouaibou, C. Hoareaux, and F. Roux, 2009: Tropical cyclone genesis in the southern hemisphere and its relationship with the ENSO. *Ann. Geophys.,* **27,** 2523–2538.

Landsea, C. W. and W. M. Gray, 1992: The strong association between western Sahel monsoon rainfall and intense Atlantic hurricanes. *J. Climate*, **5**, 435-453.

Larget, B. and D. Simon, 1999: Markov chain Monte Carlo algorithms for the Bayesian analysis phylogenetic trees. *Mol. Biol. Evol.,* **16,** 750-759.

Lehmiller, G. S., T. B. Kimberlain, and J. B. Elsner, 1997: Seasonal prediction models for North Atlantic basin hurricane location. *Mon. Wea. Rev.,* **125,** 1780–1791.

Leroy, A., and M. C. Wheeler, 2008: Statistical Prediction of Weekly Tropical Cyclone Activity in the Southern Hemisphere. *Mon. Wea. Rev.,* **136,** 3637-3654.

Liu, K. S., and J. C.-L. Chan, 2003: Climatological characteristics and seasonal forecasting of tropical cyclones making landfall along the South China Coast. *Mon. Wea. Rev,* **131,** 1650–1662.

McBride, J. L., and T. D. Keenan, 1982: Climatology of tropical cyclone genesis in the Australian region. *Int. J. Climatol.,* **1,** 13-33.

McDonnell, K. A., and N. J. Holbrook, 2004a: A Poisson regression model of tropical cyclogenesis for the Australian-southwest Pacific Ocean region. *Wea. Forecasting,*

19, 440-455.

McDonnell, K. A., and N. J. Holbrook, 2004b: A Poisson regression model approach to predicting tropical cyclogenesis in the Australian/southwest Pacific Ocean region using the SOI and saturated equivalent potential temperature gradient as predictors. *Geophys. Res. Lett.,* **31,** L20110, doi:10.1029/2004GL020843.

McDonnell, K. A., N. J. Holbrook and H. Shaik, 2006: Seasonal forecasts of tropical cyclone numbers in the Australian/southwest Pacific Ocean region using a new Poisson regression model: verification of 2005/06 season forecast and forecast for 2006/07 season. *Bull. Austr. Met. Oceanogr. Soc.,* **19(6),** 126-128.

Neal, R. M., 2003: Slice Sampling. *Ann. Stat.,* **31,** 705-767.

Nicholls, N., 1979: A possible method for predicting seasonal tropical cyclone activity in the Australian region. *Mon. Wea. Rev*., **107,** 1221–1224.

Nicholls, N., 1984: The southern oscillation, sea-surface temperature, and interannual fluctuations in Australian tropical cyclone activity. *Int. J. Climatol.,* **4,** 661-1149.

Nicholls, N., 1985: Predictability of interannual variations of Australian seasonal tropical cyclone activity. *Mon. Wea. Rev.,* **113,** 1144–1149.

Nicholls, N., 1992: Recent performance of a method for forecasting Australian seasonal tropical cyclone activity, *Aust. Meteor. Mag*., **40,** 105–110.

Ntzoufras, I., 2009: *Bayesian modeling using WinBUGS.* Wiley series in

Computational Statistics, 698, John Wiley and Sons, 520pp.

Owens B. F., and C. W. Landsea, 2003: Assessing the Skill of Operational Atlantic Seasonal Tropical cyclone Forecasts. *Wea. Forecasting,* **8,** 45-54.

Peters G. W., P. V. Shevchenko, and M. V. Wüthrich, 2009: Model uncertainty in claims reserving within Tweedie's compound Poisson models. *ASTIN Bull.,* **39,** 1-33.

Ramsay, H. A., L. M. Leslie, P. J. Lamb, M. B. Richman, and M. Leplastrier, 2008: Interannual Variability of Tropical Cyclones in the Australian Region: Role of Large-Scale Environment. *J. Climate,* **21,** 1083-1103.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.,* **108,** (D14).

Ready, S., and F. Woodcock, 1992: The South Pacific and southeast Indian Ocean tropical cyclone season 1989-1990. *Aust. Met. Mag.,* **40,** 111-121.

Roberts, G. O., and J. S. Rosenthal, 1999: Convergence of slice sampler Markov chains. *J. R. Statist. Soc.,* **61,** 643-660.

Saunders, M. A., and A. S. Lea, 2005: Seasonal prediction of hurricane activity reaching the coast of the United States. *Nature,* **434,** 1005-1008.

Solow, A., and N. Nicholls, 1990: The relationship between the Southern Oscillation and tropical cyclone frequency in the Australian region. *J. Climate*, **3,** 1097–1101.

Stone M., 1974: Cross-validatory choice and assessment of statistical predictions. *J. Royal Stat. Soc.,* **36(2),** 111–147.

Van Loon, H., and D. J. Shea, 1985: The Southern Oscillation. Part IV: The Precursors of 15°S to the Extremes of the Oscillation. *Mon. Wea. Rev.,* **113,** 2063-2074.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 467 pp.

TABLE CAPTIONS

TABLE 1. Regional average defined by all predictor indices, as well as the three-month average index period with the strongest correlation (r) with annual TCG totals in the upcoming season. All correlations presented here are significant at the 95% confidence level.

..... Page 42

TABLE 2. a) Comparison of the model skill calculated from the leave-one out cross-validated hindcasts of the Climatology and the single-predictors. Root-mean-square errors (RMSEs) are used to evaluate the quality of the predictor. Next to the RMSEs, standard errors (se) and significant correlation coefficients (r) between hindcasts and observed number of TCs help to further validate the skill of the model. Bold indicates the model with the lowest RMSE. b) as a) but for two- and three-predictor combinations based on step-by-step model selection. For the multi-predictor models, bold indicates improvement over the best single predictor CAPE model and the best three-predictor model.

..... Page 43

TABLE 3. Posterior means of model coefficients and standard deviations for selected predictor combinations, as well as the mean Bayesian model coefficients and standard deviations for the subregional three-predictor CAPE+v_{850}+GPH TCG model.

..... Page 44

FIGURE CAPTIONS

TABLES

TABLE 1. Regional average defined by all predictor indices, as well as the three-month average index period with the strongest correlation (r) with annual TCG totals in the upcoming season. All correlations presented here are significant at the 95% confidence level.

Index	**Latitude**	**Longitude**	**3-month mean**	**r**
CAPE	20°-30°S	160°E-160°W	JJA	0.73
EPT	15°-25°S	160°E-160°W	JAS	0.68
GPH	25°-45°S	170°-130°W	JAS	0.57
NINO4	5°N-5°S	160°E-150°W	JAS	-0.68
u_{850}	5°N-5°S	170°-130°W	JAS	-0.65
v_{850}	15°-5°N	120°-80°W	MJJ	-0.65
EVWS	0°-10°S	170°-130°W	JAS	0.67
RV	20°-10°N	125°-85°W	JAS	-0.65

TABLE 2. a) Comparison of the model skill calculated from the leave-one out cross-validated hindcasts of the Climatology and the single-predictors. Root-mean-square errors (RMSEs) are used to evaluate the quality of the predictor. Next to the RMSEs, standard errors (se) and significant correlation coefficients (r) between hindcasts and observed number of TCs help to further validate the skill of the model. Bold indicates the model with the lowest RMSE. b) as a) but for two- and three-predictor combinations based on step-by-step model selection. For the multi-predictor models, bold indicates improvement over the best single predictor CAPE model and the best three-predictor model.

a) ***predictor***	**RMSE**	**se**	**r**
Climatology	6.20	0.38	0.24
CAPE	**5.39**	**0.33**	**0.65**
EPT	5.53	0.37	0.61
GPH	5.70	0.41	0.53
NINO4	5.59	**0.33**	0.58
u_{850}	5.54	0.37	0.61
v_{850}	5.61	0.36	0.56
EVWS	5.50	0.38	0.60
RV	5.55	0.40	0.59

b) ***predictor***	**RMSE**	**se**	**r**
CAPE+GPH	5.42	0.36	**0.66**
CAPE+NINO4	5.48	0.34	**0.64**
CAPE+ u_{850}	5.38	0.37	**0.67**
CAPE+ v_{850}	**5.21**	0.35	**0.72**
CAPE+EVWS	**5.33**	0.35	**0.69**
CAPE+RV	**5.21**	0.36	**0.72**
CAPE+ v_{850} + GPH	**5.20**	**0.36**	**0.73**
CAPE+ v_{850} +EVWS	5.26	0.38	0.71
CAPE+ RV+ GPH	5.33	0.37	0.69
CAPE+ RV+ EVWS	5.24	0.36	0.72

TABLE 3. Posterior means of model coefficients and standard deviations for selected predictor combinations, as well as the mean Bayesian model coefficients and standard deviations for the subregional three-predictor CAPE+v_{850}+GPH TCG model.

predictor	$\overline{\beta}_0$	**SD(β_0)**	$\overline{\beta}_1$	**SD(β_1)**	$\overline{\beta}_2$	**SD(β_2)**	$\overline{\beta}_3$	**SD(β_3)**
CAPE	2.650	0.042	0.018	0.003				
CAPE + v_{850}	2.676	0.043	0.014	0.003	-0.222	0.082		
CAPE + EVWS	2.649	0.042	0.012	0.004	-0.015	0.007		
CAPE + RV	2.644	0.042	0.013	0.004	-0.011	0.006		
CAPE + v_{850} + GPH	**2.666**	**0.044**	**0.010**	**0.004**	**-0.244**	**0.084**	**0.004**	**0.002**
CAPE + RV + EVWS	2.643	0.042	0.011	0.004	-0.014	0.007	0.005	0.009
CAPE + v_{850} + GPH (West)	1.777	0.069	0.001	0.007	-0.197	0.127	0.008	0.003
CAPE + v_{850} + GPH (North)	1.063	0.094	0.006	0.010	-0.116	0.183	-0.002	0.005
CAPE + v_{850} + GPH (East)	1.667	0.074	0.020	0.007	-0.385	0.139	0.002	0.004

FIGURES

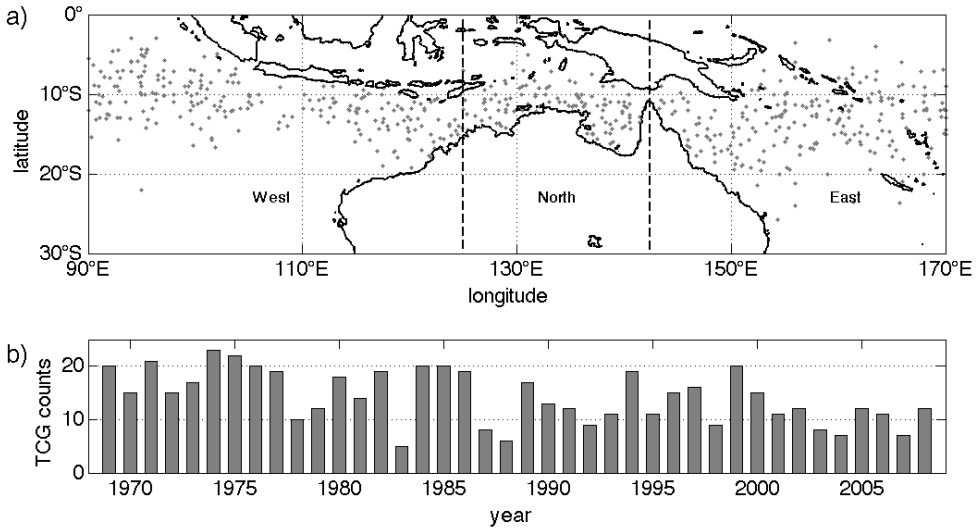


FIG 1. a) Spatial distribution of the first recorded location of tropical storm systems that later develop into tropical cyclones in the Australian region from 1968/69-2007/08 with subregions indicated (separated by the dashed lines), and b) corresponding time series of annual TCG totals across the region.

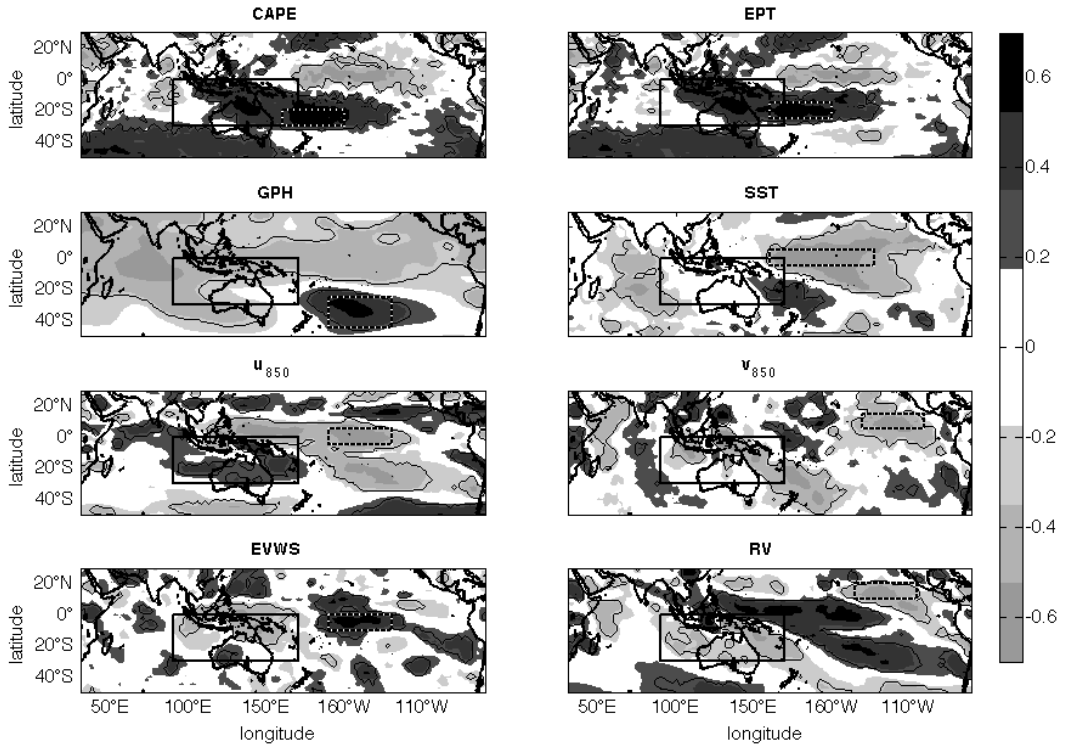


FIG 2. Spatial map of correlations between annual TCG totals and dynamic and thermal predictor variables for June-July-August (JJA). The bold box indicates the Australian TC region and the dashed box identifies the predictor region with maximum relationship with Australian region seasonal TCG. The thin lines outline spatial pattern correlations significant at the 95% level.

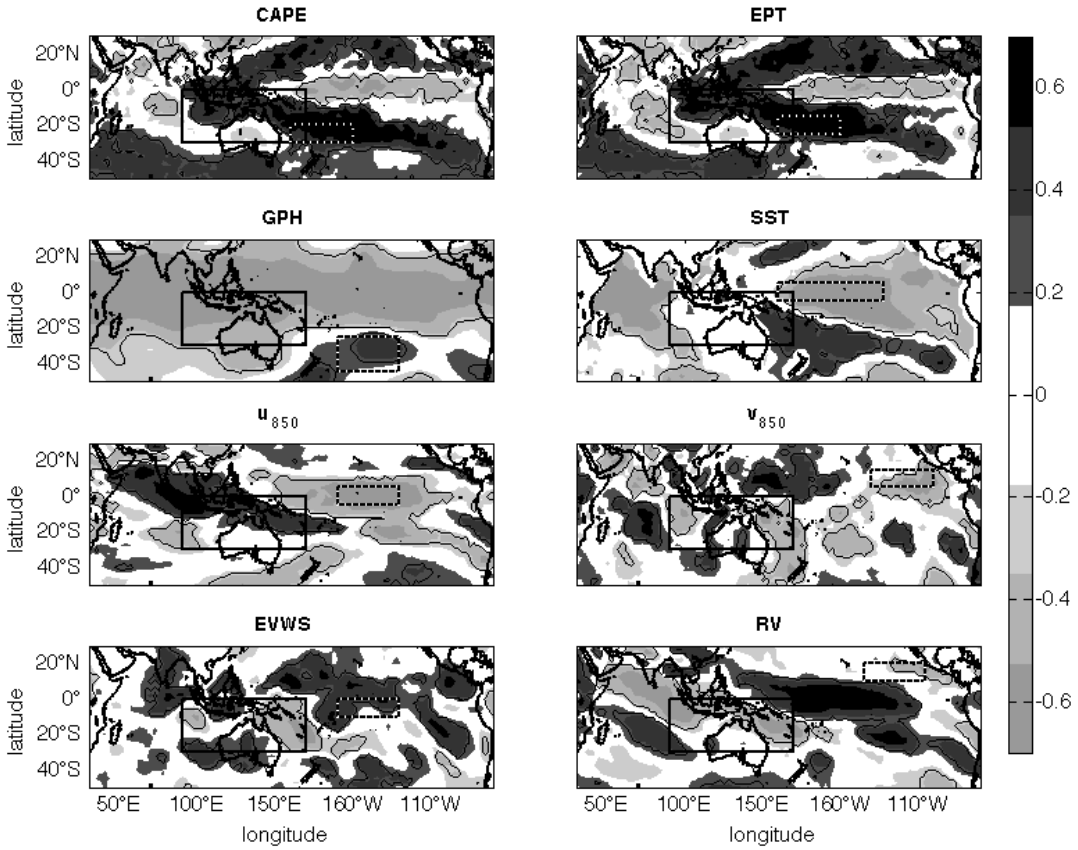


FIG 3. As in Fig. 2, but for October-November-December (OND).

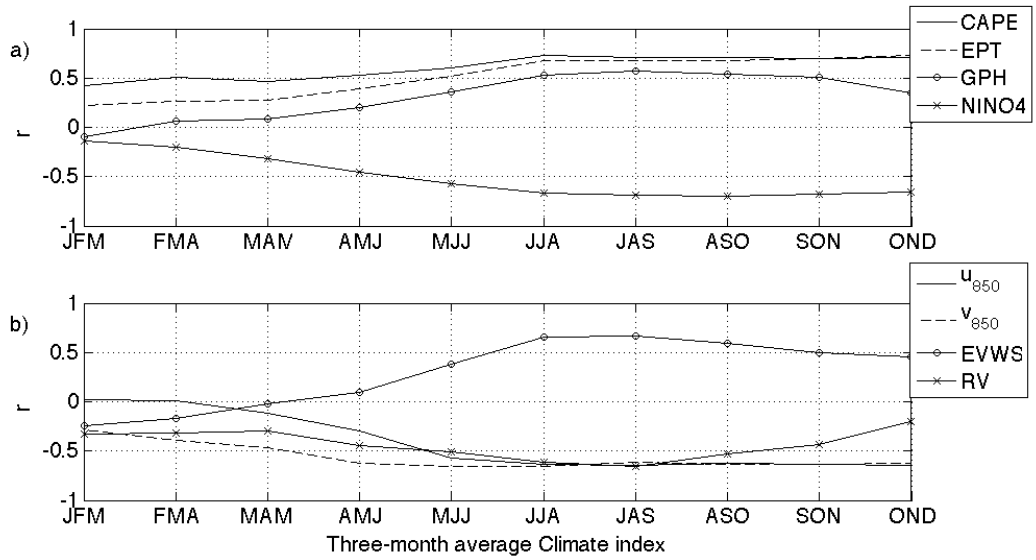


FIG 4. Correlations between annual TCG totals and the a) thermal predictor indices and b) dynamic predictor indices – where the climate indices are three-monthly averages leading the TC season.

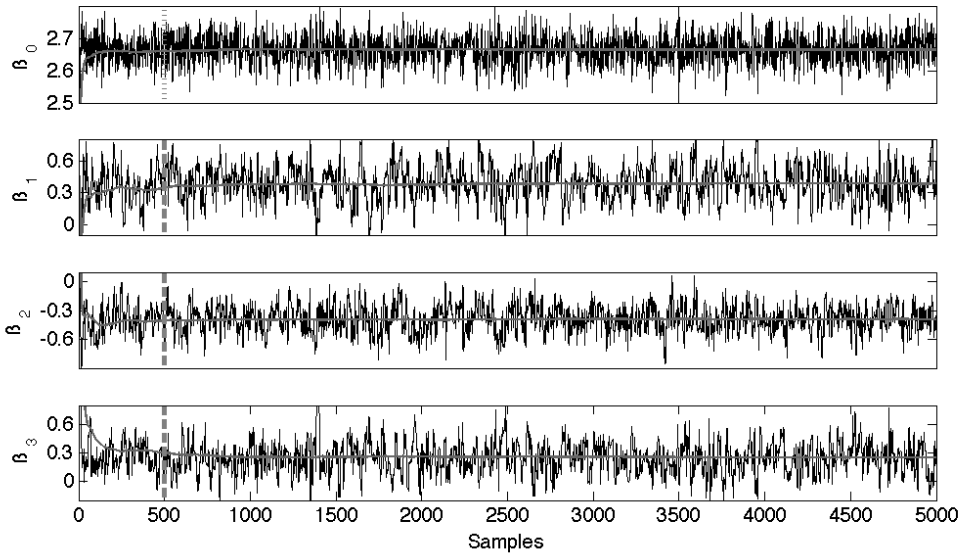


FIG 5. Example of first 5,000 iterations of the CAPE+V_{850}+GPH model. Thin black lines show the sampled model coefficients, with β_0 the model intercept and β_1, β_2 and β_3 being the corresponding model coefficients for the predictors CAPE, v_{850} and GPH. The grey line gives the cumulative means of the coefficients. Also indicated is the chosen burn-in at 500 iterations.

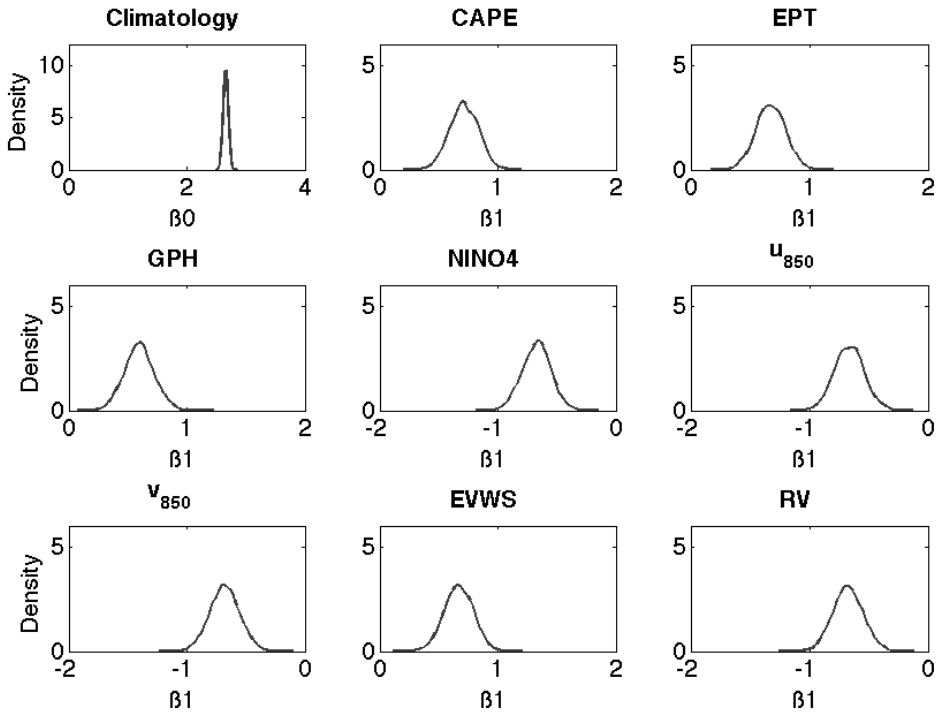


FIG 6. Corresponding Kernel densities of posterior distributions for Climatology and single-predictor variables for the MCMC samples.

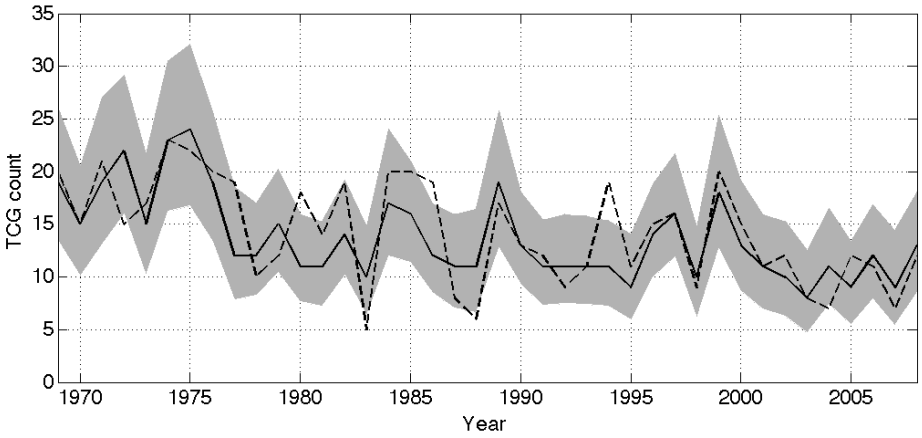


FIG 7. Variations in observed annual TCG totals (dashed line) and leave-one out cross-validation hindcasts (solid line) using the CAPE+v_{850}+GPH predictor model over the 40-year record from 1968/69-2007/08. Model standard deviations are indicated by the shading.

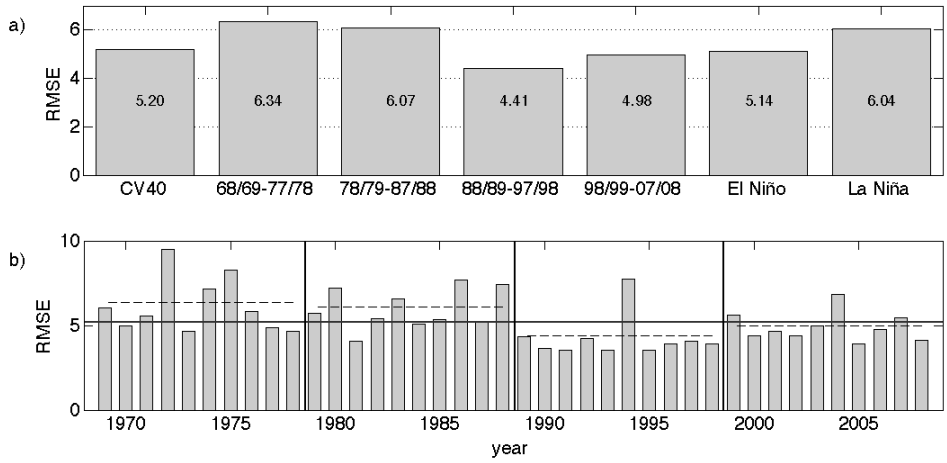


FIG 8. a) Comparison of RMSEs between the leave-one-out (CV40) and *four*-fold cross-validated hindcasts, as well as for ENSO events based on the leave-one-out cross-validation of the CAPE+v_{850}+GPH predictor model. b) Annual RMSEs of the four *four*-fold cross-validated hindcasts. The bold solid horizontal line represents the averaged RMSE from the leave-one-out cross-validated data. The thinner lines represent the averaged RMSEs of each of the four 10-year *four*-fold cross-validation hindcasts.

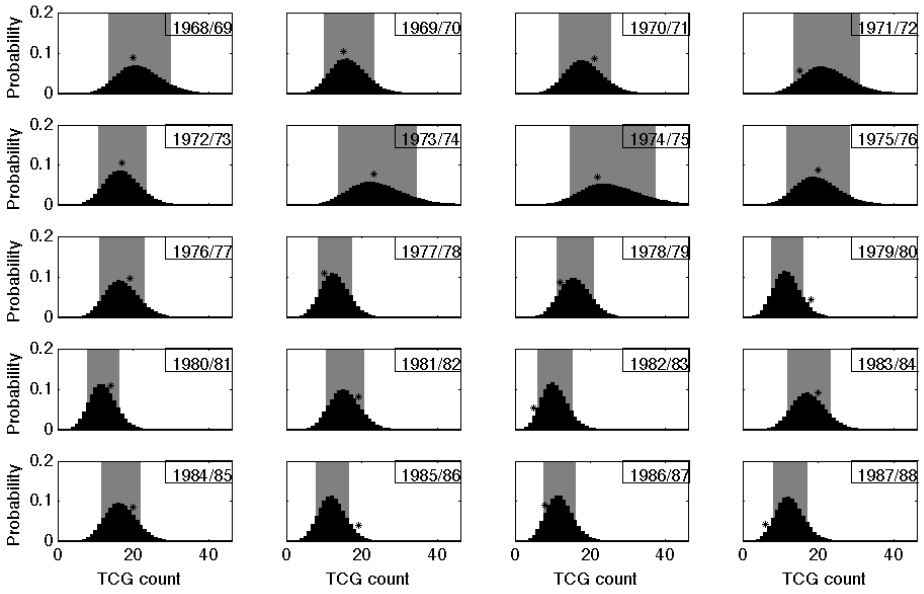


FIG 9. Probability distributions of the annual total number of TCG occurrences independently hindcast (*four*-folded cross-validation) for individual TC seasons between 1968/69-1987/88 occurrences using the three-predictor CAPE+V_{850}+GPH model. Asterisks indicate the observed number of TCGs in that season, while the model standard deviations are indicated by the shading.

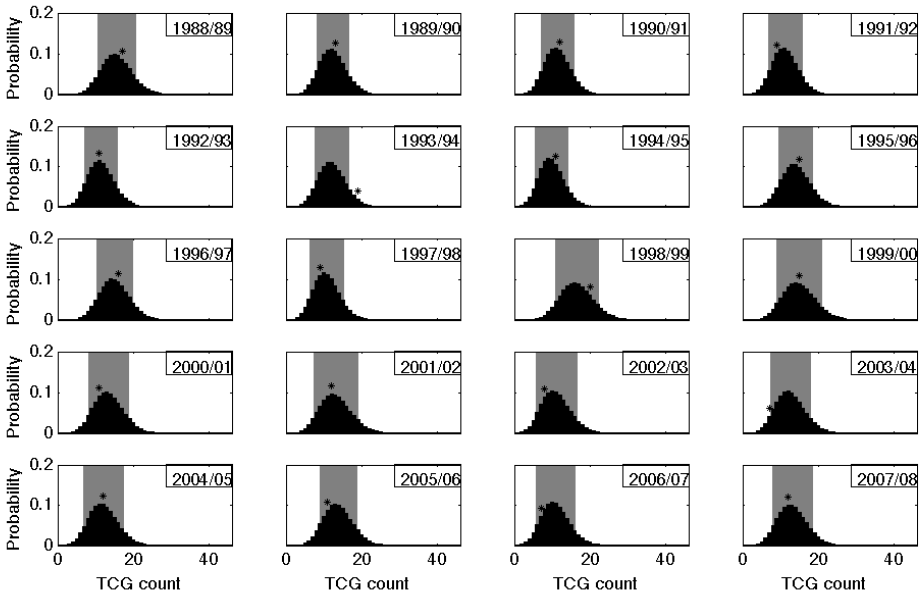


FIG 10. As in Fig. 9, but for individual TC seasons between 1988/89-2007/08.

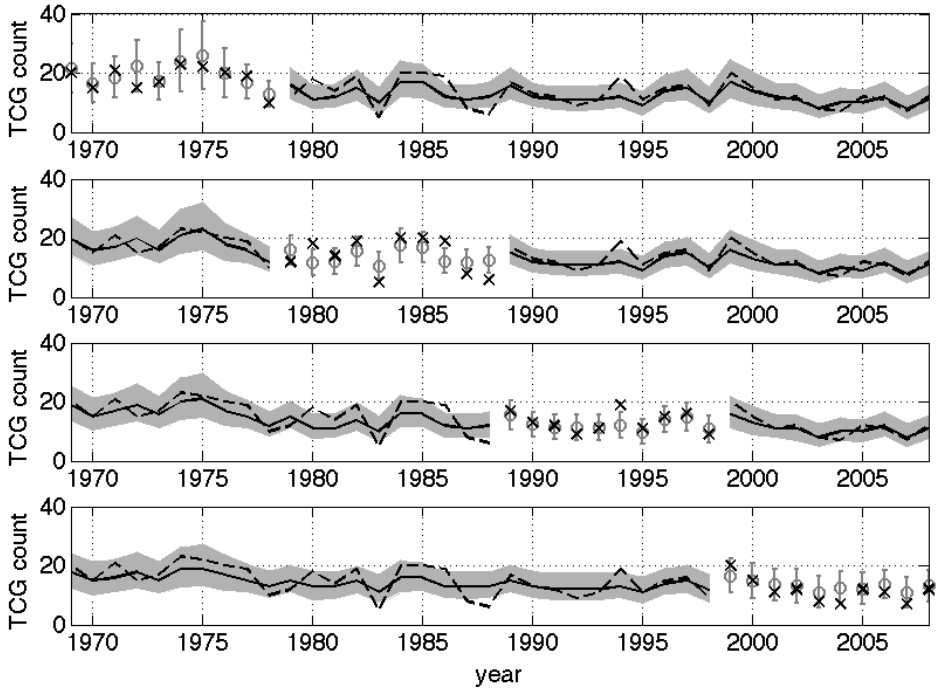


FIG 11. Variations in observed seasonal TCG totals (dashed line) and regressed hindcasts (solid line) using the CAPE+v_{850}+GPH predictor model trained on 30 years and independently hindcasting the 10 left-out years (circle; *four*-fold cross-validation), with error bars indicating the standard deviation. Crosses indicate the observed TCG totals during the independent hindcast period. Model standard deviations of the 30-year training period are indicated by the shading.

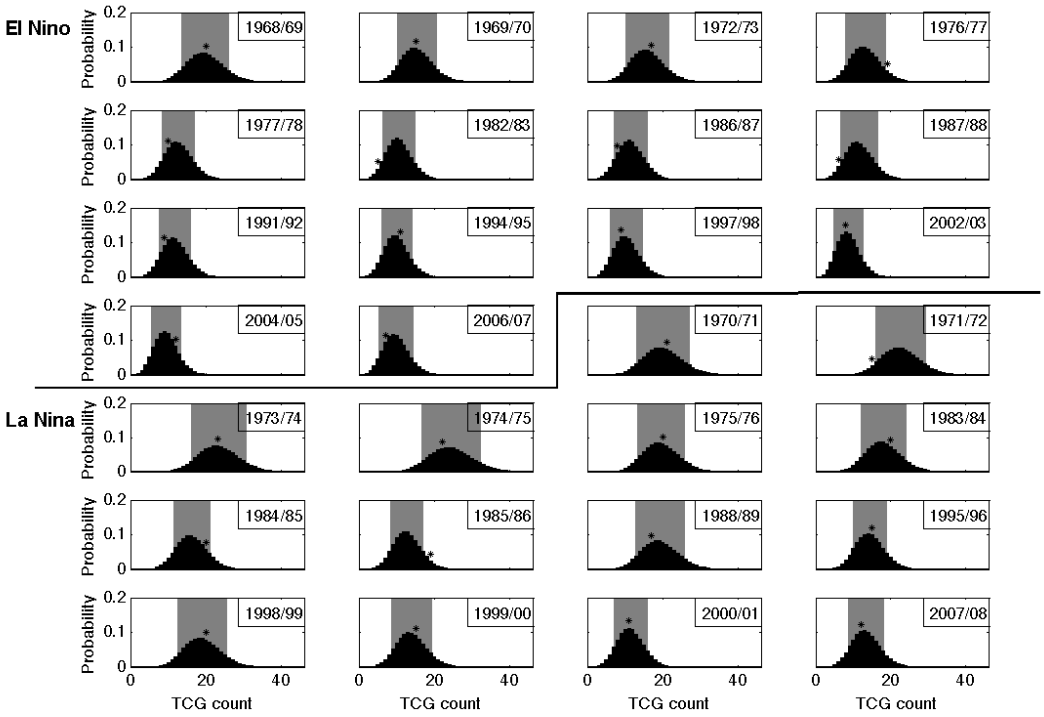


FIG 12. As in Fig. 9, but for individual TC seasons during El Niño and La Niña event years.

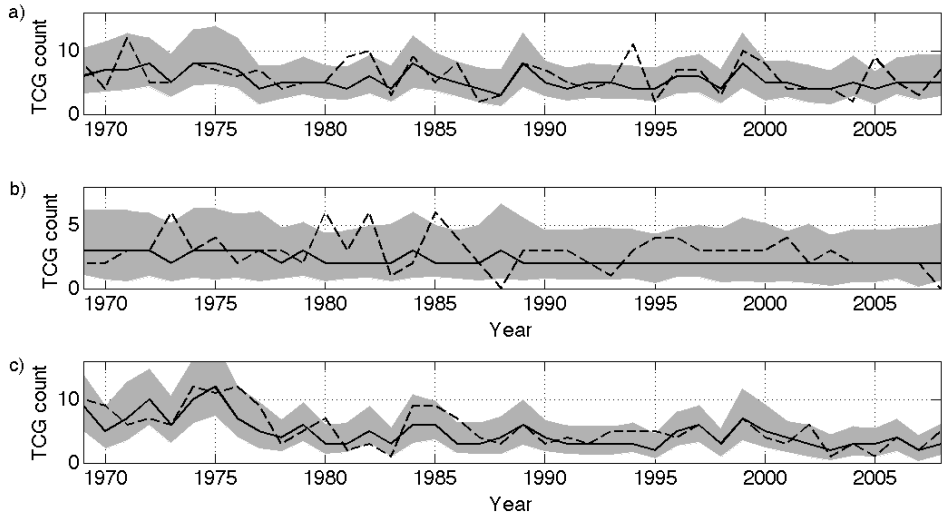


FIG 13. As in Fig. 6, but for the CAPE+v_{850}+GPH model applied to Australia's subregions a) West, b) North and c) East.

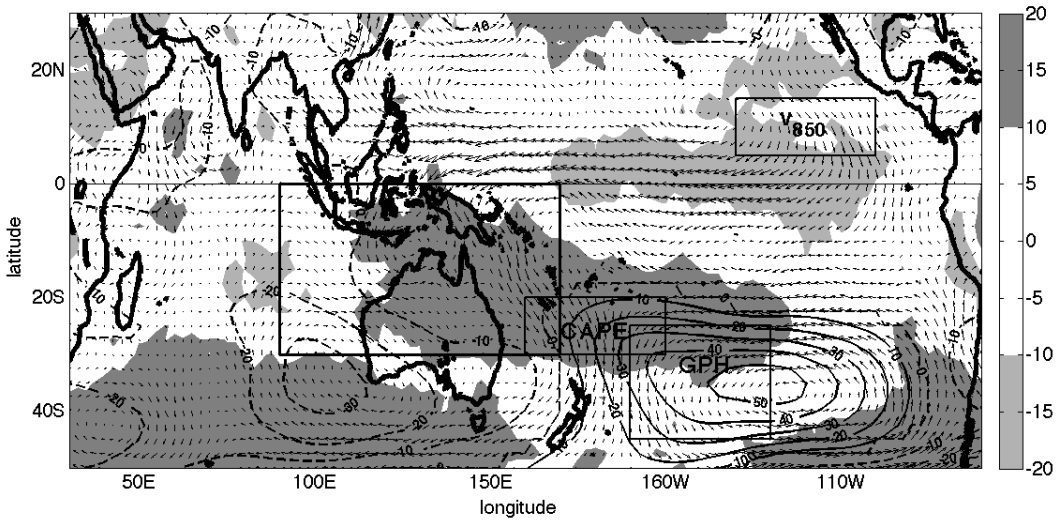


FIG 14. Map of composite anomalies for active Australian TC seasons minus composite anomalies for inactive Australian TC seasons during June-July-August. Wind vectors describe the wind flow anomalies at 850 hPa. Grey shaded areas represent changes of CAPE (m^2 s^{-2}), while full contour lines show positive anomalies and dashed lines negative anomalies of geopotential height anomalies at 500 hPa (m). Also indicated are the locations of the predictor indices CAPE, v_{850} and GPH and the Australian TC region.

4.3 Further Discussion

The advantage of the Bayesian approach over the frequentist is that it takes into account the uncertainties of the model coefficients. This enables us to predict TCG counts also outside the boundaries of our observed data set and improves the consideration of uncertainties of the model hindcasts. Additionally, the Markov Chain Monte Carlo (MCMC) method helps to successively train the model to get a better estimate of the model posterior distributions. Figure 4.1 shows the prior distribution, likelihood probability based on the observed data, and the posterior distribution of the four model coefficients of the Bayesian model, presented in the previous section. We find a more Gamma distribution-like posterior probability of the model coefficients subsequent to the MCMC method. In the following section, the Bayesian model presented in the *Journal of Climate* paper is compared to a Poisson regression frequentist model using the same predictor combination. Also analysed is the subregional hindcast skill of the Bayesian model for the Western and Eastern Australian subregions divided at 135°E.

The frequentist Poisson model, using the selected predictor combination CAPE+v_{850}+GPH, is applied to the available observed TCG data and the leave-one

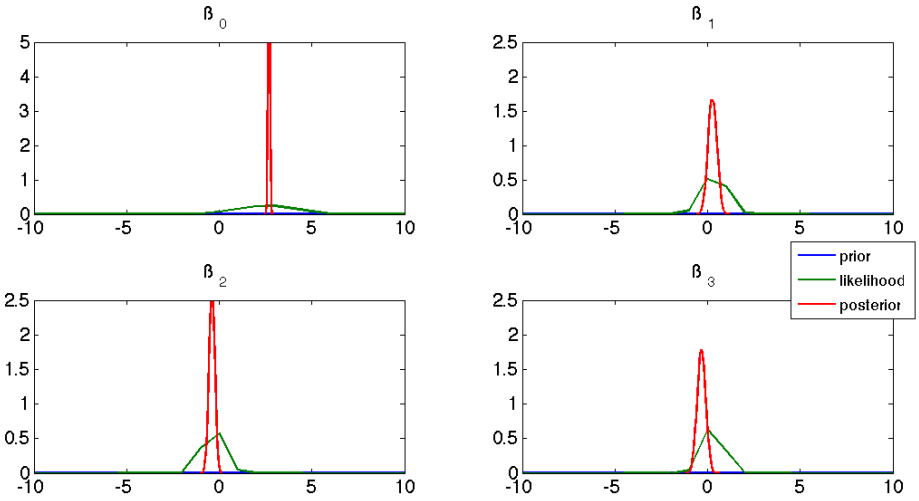


Figure 4.1 Prior probabilities, likelihood estimate from the observed data, and the posterior distribution for the model coefficients of the Bayesian model.

out cross-validated model hindcasts are compared to those generated by the previously introduced Bayesian model. The frequentist approach results in model coefficient estimates on the basis of least squares regression. As we obtain the Poisson hindcast distribution by one single coefficient combination, the hindcast distribution is narrower than the one obtained from the complete sample set of uncertainties from the posterior distribution in the Bayesian model. This means RMSE is not an adequate measure to compare the two statistical modelling methods. While both approaches result in the same correlation coefficient, the skill score between the model hindcasts and the annual TCG observations of the Bayesian model improves to SS_{Bay} = 51.5% from SS_{Fre} = 42.8% with the frequentist model. Figure 4.2 shows the residuals between the annual TCG counts and hindcasts for both the frequentist and Bayesian models. It can be seen that the Bayesian model hindcasts tend to produce a lower number of TCG occurrences than the frequentist model hindcasts and shows reduced residuals between hindcasts and TCG count observations. The reduced residuals in the cross-validated hindcasts imply a possible improvement in forecasting potential using the Bayesian model over the frequentist method.

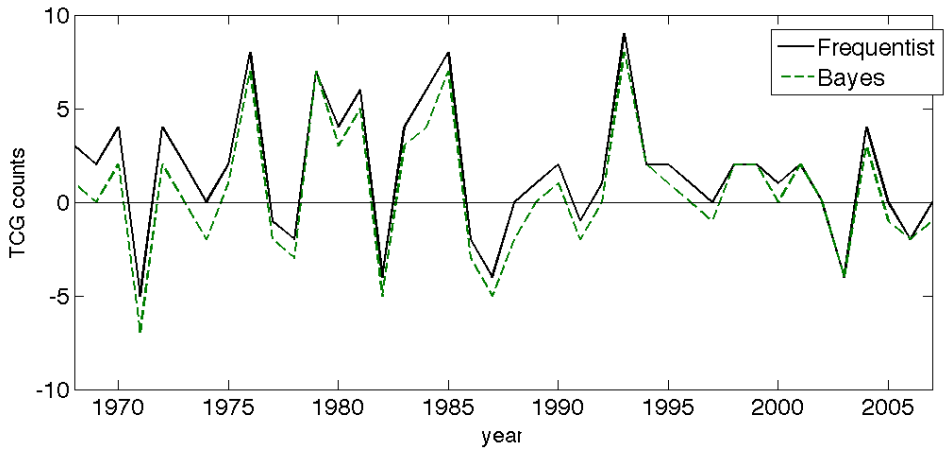


Figure 4.2 Residuals of the annual TCG count observations and the leave-one out cross-validated hindcasts from the Frequentist and Bayesian models from 1968/69-2007/08.

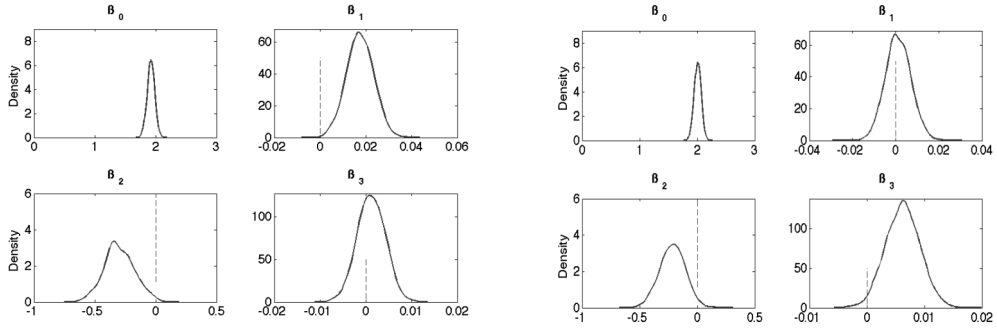


Figure 4.3 Posterior distributions of model coefficients for a) the Western Australian, and b) Eastern Australian TC regions. The dashed line indicates the zero-line for the model coefficients.

The subregional skill was further shown for three subregions, West, North and East separated at 125°E and 142.5°E. Here, TCG variations in the Northern region appear to be particularly difficult to hindcast. This lack of skill led to the choice of the two-subregion approach presented previously, with the Pacifc Ocean and Indian Ocean TC regions separated at 135°E.

Figure 4.3 shows the posterior distributions of the model coefficients when the CAPE+v_{850}+GPH predictor combination model is applied to hindcast annual TCG counts in the Western and Eastern Australian subregions. For TCG variations in the eastern Indian Ocean (Western) region, we find that the CAPE predictor has the least skill, with the sampled model coefficients changing signs in almost 50% of the sampled cases. Conversely, the v_{850} and GPH predictors appear to have good skill. For the Eastern Australian region, CAPE and v_{850} appear to be very useful predictors, while the contribution from GPH is lower. In Figure 4.4 hindcast results from the leave-one out cross-validated hindcasts are shown. We find the presented model is adaptable for forecasting southwest Pacific Ocean (Eastern) region TCG counts with a correlation coefficient between cross-validated hindcats and annual TCG count observations of r = 0.79, with 87.5% of the observed TCG counts being within the model's standard deviation and a skill score of SS = 57.8%. However, the skill of hindcasting eastern Indian Ocean TCG counts appears to be limited with correlations of only r = 0.38 and SS = 9.0%. This suggests that it is beneficial to build a separate seasonal forecast model for that region on the basis of different predictor indices.

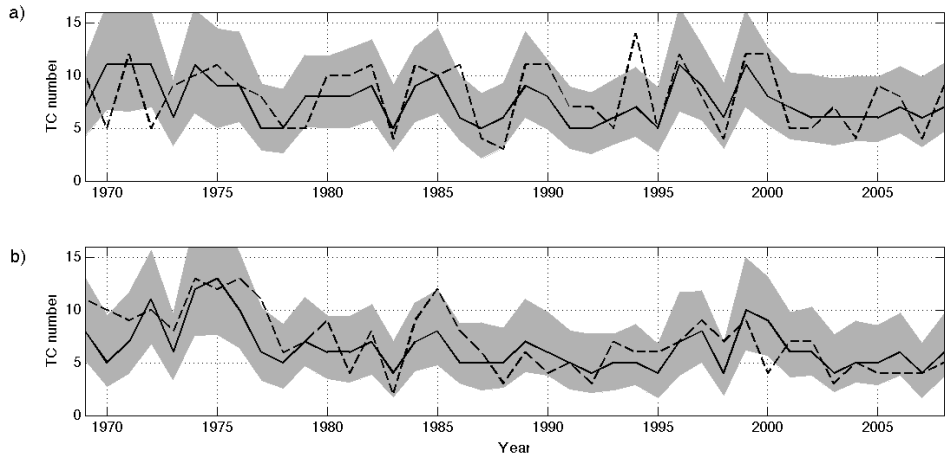


Figure 4.4 Variations in observed annual TCG totals (dashed line) and leave-one out cross-validation hindcasts (solid line) using the CAPE+v_{850}+GPH predictor model over the 40-year record from 1968/69-2007/08 applied to Australia's subregions a) West and b) East. Model standard deviations are indicated by the shading.

How to improve seasonal forecast modelling of tropical cyclone formation in the southeast Indian Ocean

Angelika Werner
Department of Environment and Geography, Macquarie University, Sydney NSW, Australia

Neil J. Holbrook
School of Geography and Environmental Studies, University of Tasmania, Hobart TAS, Australia

For submission to *Geophysical Research Letters*

Corresponding author address:
Assoc Prof Neil J Holbrook
School of Geography and Environmental Studies,
University of Tasmania,
TAS 7001 Australia.
E-mail: Neil.Holbrook@utas.mq.edu.au
Phone: +61-(0)3-6226-2027
Fax: +61-(0)3-6226-7628

Abstract

A new and potentially skilful seasonal forecast model of tropical cyclone formation (genesis, TCG) is developed for the Southeast Indian Ocean region. The model is based on Poisson regression using the Bayesian approach. Predictor combinations are chosen using a step-by-step predictor selection. The two-predictor model based on derived indices of June-July-August sea level pressure and May-June-July meridional winds at 850hPa in the tropical East Pacific Ocean produces the smallest RMSE = 3.74 for the leave-one-out cross-validated TCG hindcasts over the 40-year record between 1968/89-2007/08. The corresponding correlation coefficient between observed annual TCG totals and cross-validated model hindcasts is r = 0.57 with 82.5% of the observed seasonal TCG totals hindcast within the model standard deviations. In summary, we find that the two-predictor Bayesian model provides substantially improved skill over the climatology, with remarkably skilful hindcasts (forecasts) of Australian region and subregional seasonal TCG totals provided two months ahead of the TC season.

1. Introduction

The southeast Indian Ocean is the most active Australian tropical cyclone (TC) subbasin, with the most intense TCs occurring and a higher chance of the storms to make landfall [*Goebbert and Leslie*, 2010; *Dare and Davidson*, 2004]. Also the western Australian region is of high economic importance due to extensive reservoirs of oil and other natural resources. Nevertheless there have been very few studies on TC activity in the southeast Indian Ocean or northwest Australian region [e.g., *Broadbridge and Hanstrum*, 1998; *Goebbert and Leslie*, 2010]. They found that the relationship of western Australian TCs with ENSO is weaker than the one for the whole Australian region. That complicates accurate seasonal forecasting of tropical cyclones in the southeast Indian Ocean, as commonly used ENSO metrics are shown to have no significant correlation with TC activity in that region [e.g., *McDonnell and Holbrook*, 2004b; *Goebbert and Leslie*, 2010; *Werner and Holbrook*, 2011].

Here, we develop a very recent statistical seasonal forecast model of Australian region tropical cyclone formation [genesis; TCG; *Werner and Holbrook,* 2011] for the smaller southeast Indian Ocean subregion (0-30°S, 90°-135°E). While the larger-scale Australian region TC statistical forecast model performed well in Australia's East TC subregion (i.e., the southwest Pacific Ocean), the model cross-validated skill of seasonal (annual) TCG counts in Australia's West TC subregion (i.e., the southeast Indian Ocean) was poor – consistent with a previous statistical forecast study of TC activity in this subregion [*McDonnell and Holbrook*, 2004b]. Previous studies demonstrate that seasonal TCG count variations off northwest Australia (southeast Indian Ocean) are less related to El Niño-Southern Oscillation (ENSO) than for the entire Australian TC region, where it otherwise dominates [*Broadbridge and Hanstrum*, 1998; *Goebbert and Leslie*, 2010]. Most recently it has

been shown that other climate indices, describing large-scale climate patterns in both the Southern and Northern Hemisphere, are not significantly correlated with northwest Australian seasonal TCG counts, suggesting that it will be important to look for new measures in spatial atmospheric data [*Goebbert and Leslie*, 2010].

Since the early 1980s, a number of statistical seasonal forecast schemes have been developed and improved to predict TC activity in various basins and sub-basins [*Klotzbach et al.*, 2011]. Statistical seasonal forecast modelling of TC activity was first undertaken by *Nicholls* [1979] for the Australian region and *Gray* [1984] for the North Atlantic – both using linear regression methods. *Solow and Nicholls* [1990] presented the first nonlinear statistical forecast model of TC activity based on the Poisson regression for the Australian region. There, they used the large-scale atmospheric sea level pressure gradient time series between Tahiti and Darwin – the Southern Oscillation index (SOI) - as a predictor of Australian region annual TC counts. More recently, a Poisson regression model using spatial grid-point estimates (on a 5^{o} x 2.5^{o} longitude-latitude grid) of the September lead saturated equivalent potential temperature gradient between 1000hPa and 500hPa and the SOI was developed to forecast upcoming season TCG totals across the Australian region [*McDonnell and Holbrook*, 2004a,b]. Their study was informed by *Gray*'s [1968] seasonal genesis parameters. To extend this work, *Werner and Holbrook* [2011] also developed a Poisson regression model for forecasting seasonal TCG events, but instead used a Bayesian approach and explored different predictor variables scoped out through a large-scale correlation analysis on Australian region TCG. The final predictor variables selected and derived by virtue of a systematic step-by-step process include indices (time series) of the subtropical central South Pacific convective available potential energy, meridional winds at 850 hPa in the tropical East Pacific

(the Walker circulation inflow region), and the central South Pacific geopotential height at 500 hPa. Both the *McDonnell and Holbrook* (2004b) and *Werner and Holbrook* [2011] models have been applied to forecast TCG totals in Australian region subsets (subregions). However, while both models showed considerable improvement in annual TCG count forecasts over climatology of 25% [*McDonnell and Holbrook*, 2004b] and 58% [*Werner*, 2011] in the southwest Pacific Ocean (Australian East subregion), only the *Werner and Holbrook* [2011] model some skill in forecasting annual TCG count forecasts in the southeast Indian Ocean region (Australian West subregion). *Liu and Chan* [2010] applied a project-pursuit regression model to forecast seasonal variations in TCG counts for the Australian region and its two important subregions using well-known climate indices – the western equatorial Pacific sea surface temperature anomaly (SSTA) index - NINO4, the Indian Ocean Dipole Mode index (DMI), the trade wind index, and outgoing long-wave radiation (OLR) index. Using the jackknife method, they achieved a 51% improvement in the root-mean-square error (RMSE) over climatology for the Australian region and a 39% improvement for the northwest Australian region (southeast Indian Ocean). For the southeast Indian Ocean-only region, *Goebbert and Leslie* [2010] also presented some very preliminary results from a multiple linear regression model approach investigating TC frequency and TC days in the northwest Australian region (105°-135°E, 0-35°S). As results using the more standard climate indices were found not to be significant, the model instead took advantage of derived indices and achieved considerable skill in doing so. However, it is important to note that since the results were only preliminary, the predictors were not comprehensively investigated due their climatological relevance or persistence of the correlation patterns. So, we are cautious here about the robustness of the apparent model skill presented in their results, which may be artificial.

This paper presents an application of the *Werner and Holbrook* [2011] Australian-region statistical TCG count forecast model developed further for the southeast Indian Ocean region (90°-135°E, 0-30°S). The forecasting scheme presented here shows considerable promise based on a comprehensive assessment of its cross-validated hindcast skill, with relatively high correlations identified between observed and hindcast seasonal TCG counts (r = 0.57) and a low standard error (se = 0.26) – an improvement in mean-squared error (MSE) of 30% over the climatological average. We demonstrate that this model makes significant advances on previous statistical schemes used to forecast TC activity in the southeast Indian Ocean (Australian West subregion).

2. Data

a. Tropical cyclone observations

This study takes advantage of the global TC best track data set IBTrACS.v02 [*Knapp et al.*, 2010] provided by the U.S. National Oceanic and Atmospheric Administration. TCG is defined as the spatial location where a tropical storm system with winds exceeding 34 knots (17.5 ms^{-1}) is first recorded.

The southeast Indian Ocean (tropical cyclone) region is defined here as spanning between 0°-30°S and 90°-135°E. The 135°E eastern border was chosen following *Kuleshov et al.* [2010], as it represents the longitude with the fewest TC tracks crossing it. TCG occurrences identified poleward of 30°S have been removed in the quality assessment process. Only storms during the Australian TC season from November to April are taken into account. Overall, a total of 322 TCs during the 40-year period from 1968/69-2007/08 are analyzed following the quality control. Figure 1 shows the spatial distribution of the first noted locations of tropical storm systems

that later developed into TCs and the corresponding time series of seasonal TCG totals in the 40-year record.

b. Oceanic and atmospheric data

In this study, sea surface temperature (SST) data were taken from the uniformly gridded observations provided in the Hadley Centre Global Sea Ice and Sea Surface Temperature (HadISST1) dataset [*Rayner et al.*, 2003], compiled by the UK Met Office Hadley Centre. HadISST1 is a combination of global monthly SST fields and sea ice concentrations on a 1° × 1° grid from 1870 to the present.

The atmospheric data analyzed in this study are the NCEP/NCAR monthly mean upper-air reanalysis, with 2.5° horizontal resolution on 17 pressure levels [*Kalnay et al.*, 1996]. In total, eight variables were analyzed as potential TCG predictors, describing the thermodynamic and dynamic condition of the ocean and atmosphere across a large portion of the Indo-Pacific region identified as 30°N-50°S, 30°E-70°W. The four thermodynamic parameters considered are SST, sea level pressure (SLP), geopotential height at 500 hPa (GPH), and the (non-saturated) equivalent potential temperature gradient between 1000hPa and 500hPa (EPT). The four dynamic parameters investigated are the meridional winds at 850hPa (v_{850}), zonal winds at 200hPa (u_{200}), the environmental vertical wind shear between 850hPa and 200hPa (EVWS), and the relative vorticity at 850hPa (RV). Monthly anomalies of all variables were determined against a 30-year base period of 1970-1999. Statistical significance of the correlation coefficients takes account of serial correlation in the time series and is based on the reduced effective number of degrees of freedom method outlined by *Davis* [1976].

c. Statistical model

Following previous studies [e.g., *Elsner and Jagger*, 2006; *Chu and Zhao*, 2007; *Chand et al.*, 2010; *Werner and Holbrook*, 2011], we applied a Bayesian Poisson regression approach to predict the seasonal number of TCG occurrences. Bayes' theorem allows us to find the best possible model coefficient representation – specifically, this information is used here to develop model forecast estimates of southeast Indian Ocean region seasonal TCG totals. Following *Werner and Holbrook* [2011], we used diffusive prior information, selecting a Gaussian distribution with zero mean and standard deviation of 100 for all model coefficients. A Markov Chain Monte Carlo (MCMC) method was performed using slice sampling. The posterior probability quantifies the uncertainty in the model coefficients provided by the probabilities. To discard the effects of the chosen initial conditions, we applied a model burn-in of 500 iterations. To avoid high autocorrelations, aiming to achieve effective statistical independence of the samples in the iteration process, the samples were thinned so that only every n+2 sample was taken into account, with n being the number of predictors in the model. In summary, 5,000 samples were used to estimate the model coefficients and obtain the model forecasts. For detailed information about the modelling approach, see *Werner and Holbrook* [2011].

To assess potential model skill using the selection of predictors, a leave-one-out cross-validation was performed [e.g., *Stone*, 1974; *Werner and Holbrook*, 2011]. Here, the model is trained using n-1 tropical cyclone seasons to hindcast the number of TCG events expected for the single tropical cyclone season that has been withheld from the training data set. This train-and-test approach is successively repeated to hindcast each TC season across the 40-year data set. This enables us to perform an independent hindcast of every TC season. To evaluate model skill, the probabilistic RMSE of each model hindcast was calculated [e.g., *Elsner and Jagger*, 2004, 2006;

Chu and Zhao, 2007; *Chand et al*., 2010; *Werner and Holbrook*, 2011]. The RMSE is a commonly used metric for the potential utility of a predictor (or predictor combination) in a probabilistic model, where a smaller RMSE typically means a better model. The RMSE is calculated using the probabilities of the independent hindcasts in the leave-one-out cross-validation method. Another measure is the skill-score (SS), calculated from the common MSE improvement of the cross-validated model hindcasts over the climatological average MSE. The climatology-only model (as used in table 2a) was determined as the uncertainties of the intercept and with hindcasts of TCG numbers close to the observational long-term average.

The standard error (se) and cross-correlation (r) between predicted and observed number of TCs formed in the region provide measures of the first-order hindcast skill. The final model coefficients were estimated based on data from the training period 1968/69-2007/08.

3. Prediction schemes

Figure 2 shows spatial correlation maps between southeast Indian Ocean region annual TCG totals and the individual predictor variables shown for the June-July-August (JJA) period prior to the upcoming TC season. The correlation patterns with the thermal variables EPT and SST (see corresponding panels in Figure 2) are characterized by the developing ENSO [e.g., *Gaurreaud and Battisti*, 1999; *Drosdowsky and Chambers*, 2001] 'boomerang' pattern across the tropical/subtropical Pacific. This pattern defines regions of enhanced convection in the Pacific, contrasted by suppressed convection over the Indian Ocean. SLP shows a strong positive correlation pattern throughout the Pacific that is strongest in the tropical and subtropical central Pacific, while negative correlations exist over much of the

Australian continent, south of Australia, and in the southeast Indian Ocean associated with large-scale ENSO. The positive correlations of GPH with southeast Indian Ocean TCG counts in the central subtropical South Pacific are located at the southern tip of the South Pacific Convergence Zone and the warm advection region of the semi-permanent South Pacific subtropical High. However, interannual variations of the air column in that region are driven mainly by changes of the trough in the mid-latitude westerlies between 15°S and 45°S in the central South Pacific [*Van Loon and Shea*, 1985].

The dynamic variables u_{200}, EVWS and RV, tend to show similar patterns to GPH, EPT and SST, with a change of sign in the southeast Indian Ocean (bottom four panels of Figure 2). These patterns describe changes in the Walker and Hadley circulations due to ENSO phase changes. The correlation pattern in the tropical Pacific and Indian Ocean between annual southeast Indian Ocean TCG totals and u_{200} imply enhanced TCG with a strengthening of the Walker circulation and trade winds. The correlation patterns of v_{850}, EVWS and RV with southeast Indian Ocean region TCG totals imply a stronger (weaker) Walker circulation leading to increased (decreased) convection in the Warm Pool, which is more (less) favorable to TCG in the southeast Indian Ocean.

To achieve climatologically relevant predictors, we derived single indices for each climate predictor based on the persistence and magnitude of the spatial correlation coefficients. Spatial correlation maps for October-November-December (OND; not shown) also help to identify patterns, which remain significant after the austral winter till the start of the Australian region TC season. The most persistent correlation patterns observed for GPH and EPT with southeast Indian Ocean TCG are both located in the subtropical central South Pacific, while the NINO4 SSTA index,

SLP, u_{200} and EVWS describe variations in the equatorial central-western Pacific region. The v_{850} and RV indices correspond to the northern inflow region of the Walker cell in the tropical East Pacific.

We found that all predictors generate persistent strong correlations with upcoming seasonal TCG by JJA at the latest, with correlations $r > |\pm 0.4|$. The strongest correlations between the climate index three-month means and annual TCG totals are provided in Table 1. All correlations are significant at the 95% confidence level. Predictor combinations should be complementary and contribute information that adds value. For that reason, predictor combinations where $r > |\pm 0.8|$ with each other were not included in any further analysis. In total, eight predictor indices were further investigated.

4. Results

Using the leave-one-out cross-validation technique, we calculated the RMSE as a metric of model skill. The best single-predictor is SLP (RMSE = 3.79, se = 0.26; Table 2a). In an attempt to further improve the model, we investigated predictor combination models using a step-by-step predictor selection based on the calculated RMSEs. SLP was used as the key single-predictor based on its lowest RMSE and strongest correlation with annual TCG totals. The two-predictor combination of SLP+v_{850} provided a further 1% reduction in RMSE to 3.74 and se = 0.26 (Table 2b). No further reductions were gained from extra predictors. Figure 2 shows the 40-year leave-one-out cross-validated hindcasts for the two-predictor SLP+v_{850} model plotted against the observed total number of TCs formed in each season. The hindcasted annual TCG count identifies the total number of TCG occurrences with maximum probability in the hindcast distribution. We find that this two-predictor model captures

the variability in number of cyclones formed within the boundaries of standard deviation, with 82.5% success rate of the cross-validated hindcasts (Figure 3). The model performed well overall, with favourable cross-validated hindcast totals (using the leave-one out method) against the observed annual TCG totals, with r = 0.57 (Table 2b) and a skill score of 30%. While the model is skilful for this otherwise challenging region of low predictability, there are nevertheless problems capturing unusually high or low TCG counts (1971/72, 1985/86, 1986/87, 1993/94, 2003/04) or unusually high counts during certain El Niño years (1972/73, 2004/05). We find, however, that the phase of ENSO does not appear to be as important for the number of observed southeast Indian Ocean TCG occurrences as it is for the southwest Pacific (not shown), which is consistent with previous analyses [*Broadbridge and Hanstrum*, 1998; *Goebbert and Leslie*, 2010].

5. Discussion and Summary

This paper presents a new Bayesian seasonal forecasting model of tropical cyclone formation (genesis, TCG) for the southeast Indian Ocean region that provides substantial improvement (SS = 30%) over climatology. It is well understood that TC activity in the Australian region is strongly affected by the phase of ENSO. While there is clear evidence for this relationship across the broader Australian tropical cyclone region, recent work indicates that there are no significant correlations between standard ENSO indices and seasonal TCG variations in the northwest Australian region [*Goebbert and Leslie*, 2010].

In the present study, we extended the statistical analysis to include a suite of other predictor variables. These included: two subtropical central South Pacific indices - the geopotential height at 500 hPa (GPH), and (un-saturated) equivalent

potential temperature gradient between 1000 hPa and 500 hPa (EPT); three central tropical Pacific indices - the sea level pressure (SLP), zonal winds at 200 hPa (u_{200}), and environmental vertical wind shear between 850 hPa and 200 hPa (EVWS); and two northeast tropical Pacific indices - the meridional wind speed at 850 hPa (v_{850}), and low-level relative vorticity at 850 hPa (RV; see Figure 2). Additionally, the ENSO SSTA index NINO4 was used. As a result of correlating pre-TC season three-month climate index means with southeast Indian Ocean TCG totals in the upcoming season, we found that the three-month austral winter mean indices showed the greatest overall TCG forecast potential. The best eight predictor variables (Table 1) were incorporated into a Poisson regression model developed on TCG totals. Following recent studies [e.g., *Elsner and Jagger*, 2004, 2006; *Chu and Zhao*, 2007; *Chand et al.,* 2010; *Werner and Holbrook*, 2011], we applied a Bayesian approach using the Markov Chain Monte Carlo method. This approach is beneficial as it permits the incorporation of prior beliefs, and is convenient to account for the uncertainties in model parameters.

Using a two-predictor Bayes' model on key indices of SLP and v_{850}, we provide substantial improvement in cross-validated hindcast skill over climatology. Recently published seasonal forecast modelling studies of southeast Indian Ocean and northwest Australian region TC activity showed improvements of 39% in RMSE over climatology [90°-135°E; *Liu and Chan*, 2010] using a project-pursuit regression model and the jackknife validation method and preliminary results by *Goebbert and Leslie* [2010] using a multiple linear regression model approach suggest 64% in MSE over climatology in the northwest Australian region defined as 105°-135°E. While the second result is apparently impressive, the authors clearly note the preliminary nature of their model approach, in which predictors are derived on basis of correlation

patterns between climate variables and annual TCG variablility in the northwest Australian region. However, the predictors are located in large distance to the investigated region and are not analysed due their persistence of the correlation patterns and their climatological relevance to TCs in the northwest Australian region. Here, using the leave-one-out cross-validated hindcast approach, we achieve correlations of r = 0.57 with observed TCG totals using the two-predictor model of SLP+v_{850} for the period 1968/69-2007/08. In comparison to the skill of the climatology-only model, the SLP+v_{850} model shows a 30% improvement in MSE.

It is interesting to finally note that while ENSO is arguably less important to southeast Indian Ocean region TC activity [*Broadbridge and Hanstrum*, 1998; *Goebbert and Leslie*, 2010], at least the SLP index derived in this study is ENSO-linked, while the v_{850} index appears to vary due to a different mechanism. Combining SLP with the lower-troposphere meridional inflow from the northeast tropical Pacific produced a valuable and complementary predictor set. The East Pacific tropical SLP index is embedded in a large-scale pressure pattern associated with ENSO. The significant negative relationship between southeast Indian Ocean TCG annual totals and the ITCZ-linked North Pacific v_{850} index means a strengthening of the low-level equatorward trade wind inflow (Walker circulation) leads to enhanced TCG.

In summary, we have showed improved seasonal forecasting of TCG counts in the southeast Indian Ocean region by using meaningful indices of climate variables in a Bayesian Poisson model. We find that a two-predictor SLP+v_{850} Poisson model produces relatively skilful cross-validated hindcasts of southeast Indian Ocean region annual TCG totals, with the 'forecasts' available at least two months prior to the onset of the Australian (including southeast Indian Ocean) region TC season from November-April. The predictor variables identified in this study are physically

meaningful and appropriate to condition the model TCG forecasts. With the Poisson model system using a Bayesian approach, we generate cross-validated hindcasts of southeast Indian Ocean region seasonal TCG totals (r = 0.57) with low RMSE (3.74) against a 40-year record of observations – that compare favourably against previous model attempts in the literature for this region.

References

Broadbridge, L. W., and B. N. Hanstrum (1998) The relationship between tropical cyclones near Western Australia and the Southern Oscillation Index. *Austr. Meteorol. Mag.*, *47*(3), 183-89.

Chand, S., K. J. E. Walsh, J. C. L. Chan (2010) A Bayesian Regression Approach to Seasonal Prediction of Tropical Cyclones Affecting the Fiji Region. *J. Climate, 23,* 3425-3445.

Chu, P.-S., and X. Zhao (2007) A Bayesian regression approach for predicting tropical cyclone activity over the central North Pacific. *J. Climate, 20,* 4002–4013.

Davis, R. E. (1976) Predictability of sea surface temperatures and sea level pressure anomalies over the North Pacific Ocean. *J. Phys. Oceanogr., 6,* 249-266.

Drosdowsky, W., and L. E. Chambers (2001) Near global scale sea surface temperature anomalies as predictors of Australian seasonal rainfall. *J. Climate, 14,* 1677-1687.

Elsner, J. B., and T. H. Jagger (2004) A hierarchical Bayesian approach to seasonal hurricane modeling. *J. Climate, 17,* 2813–2827.

Elsner, J. B., and T. H. Jagger (2006) Prediction models for annual U.S hurricane counts. *J. Climate, 19,* 2935–2952.

Garreaud, R. D., and D. S. Battisti (1999) Interannual (ENSO) and interdecadal

(ENSO-like) variability in the Southern Hemisphere tropospheric circulation. *J. Climate, 12,* 2113–2123.

Goebbert, K. H., and L. M. Leslie (2010) Interannual Variability of Northwest Australian Tropical Cyclones. *J. Climate, 23,* 4538-4555.

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 II: Forecasting its variability. *Mon. Wea. Rev., 112,* 1669–1683.

Kalnay, E., and Coauthors (1996) The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc., 77,* 437-471.

Klotzbach, P. J., A. Barnston, G. Bell, S.J. Camargo, J.C.L. Chan, A. Lea, M. Saunders, and F. Vitart (2011) Seasonal Forecasting of Tropical Cyclones. *Global Guide to Tropical Cyclone Forecasting, 2nd edition*, C. Guard, editor, WMO, *in press*.

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

Kuleshov, Y., R. Fawcett, L. Qi, B. Trewin, D. Jones, J. McBride, and H. Ramsay (2010) Trends in tropical cyclones in the South Indian Ocean and the South Pacific Ocean. *J. Geophys. Res., 115,* D01101, doi:10.1029/2009JD012372.

Liu, K. S., and J. C.-L. Chan (2010) Interannual variation of Southern Hemisphere tropical cyclone activity and seasonal forecast of tropical cyclone number in the Australian region. *Int. J. Climatol. (2010)*, DOI: 10.1002/joc.2259.

McDonnell, K. A., and N. J. Holbrook (2004a) A Poisson regression model of tropical cyclogenesis for the Australian-southwest Pacific Ocean region. *Wea. Forecasting, 19,* 440-455.

McDonnell, K. A., and N. J. Holbrook (2004b) A Poisson regression model approach to predicting tropical cyclogenesis in the Australian/southwest Pacific Ocean region using the SOI and saturated equivalent potential temperature gradient as predictors. *Geophys. Res. Lett., 31,* L20110, doi:10.1029/2004GL020843.

Nicholls, N. (1979) A possible method for predicting seasonal tropical cyclone activity in the Australian region. *Mon. Wea. Rev*., *107,* 1221–1224.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res., 108,* (D14).

Solow, A., and N. Nicholls (1990) The relationship between the Southern Oscillation and tropical cyclone frequency in the Australian region. *J. Climate*, *3,* 1097–1101.

Stone, M. (1974) Cross-validatory choice and assessment of statistical predictions. *J. Royal Stat. Soc., 36(2),* 111–147.

Van Loon, H., and D. J. Shea (1985) The Southern Oscillation. Part IV: The Precursors of 15°S to the Extremes of the Oscillation. *Mon. Wea. Rev., 113,* 2063-2074.

Werner, A. (2011) Seasonal forecasting of Australian region tropical cyclone formation. Ph.D thesis, Macquarie University. 240pp (submitted, 19 May 2011).

Werner, A., and N. J. Holbrook (2011) A Bayesian forecast model of Australian region tropical cyclone formation. *J. Climate* (accepted, 13 May 2011).

Figures and Tables

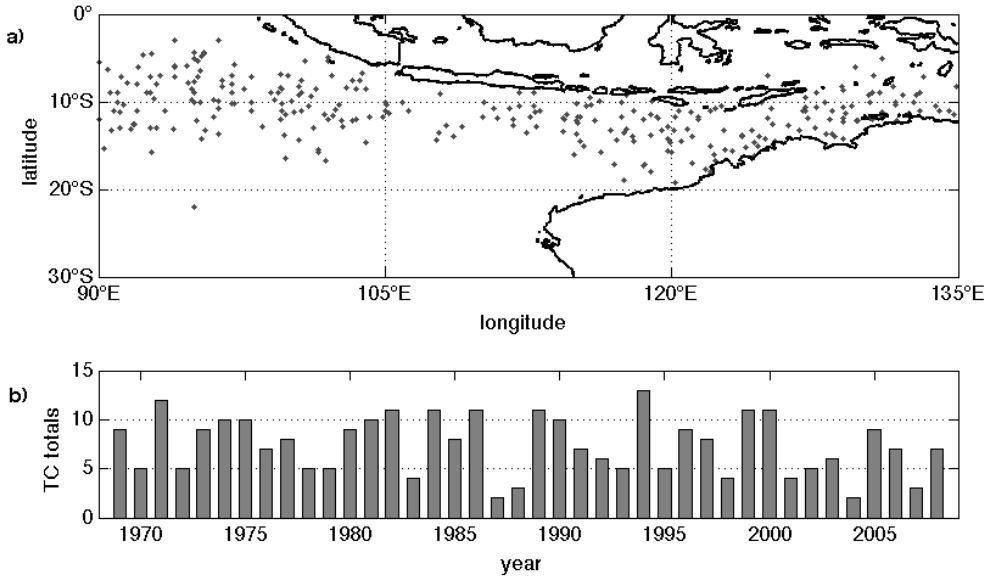


Figure 1: a) Spatial distribution of the first recorded location of tropical storm systems that later develop into tropical cyclones in the southeast Indian Ocean region from 1968/69- 2007/08, and b) corresponding time series of annual TCG totals across the region.

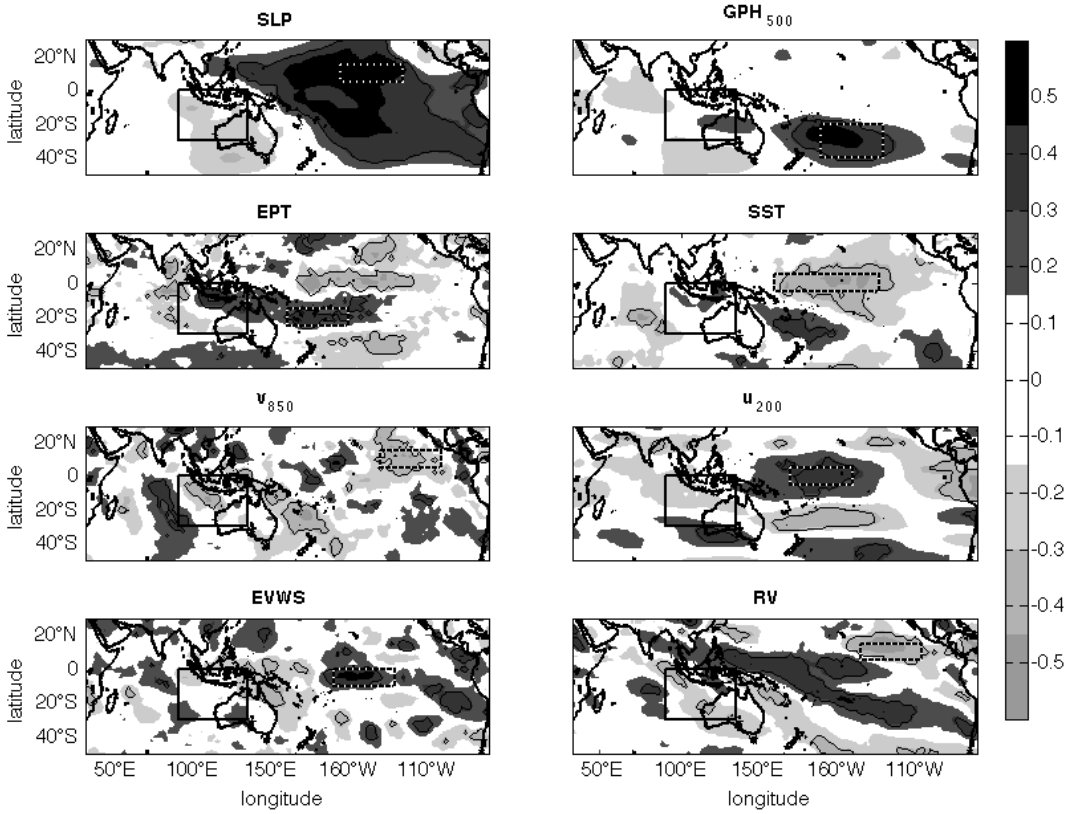


Figure 2: Spatial map of correlations between annual TCG totals and dynamic and thermal predictor variables for June-July-August (JJA). The bold box indicates the southeast Indian Ocean TC region and the dashed box identifies the predictor region with maximum relationship with southeast Indian Ocean TCG. The thin lines outline spatial pattern correlations significant at the 95% level.

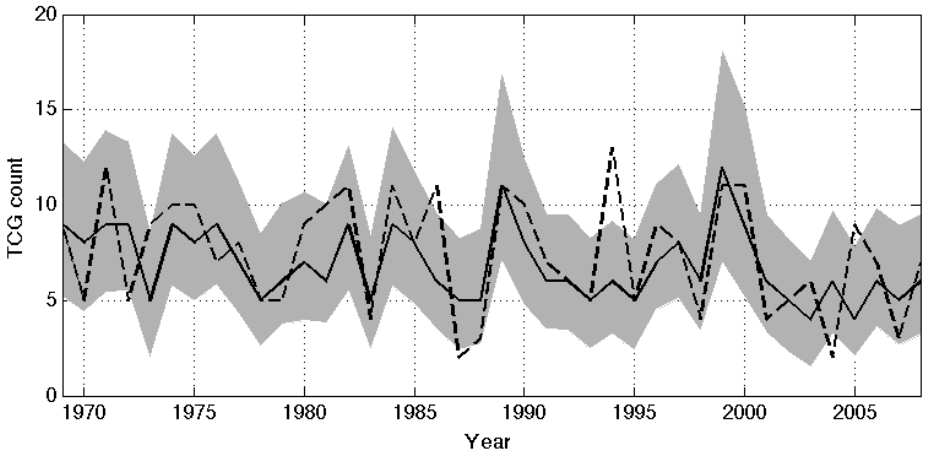


Figure 3: Variations in observed seasonal TCG totals (dashed line) and leave-one out cross-validation hindcasts (solid line) using the SLP+v_{850} predictor model over the 40-year record from 1968/69-2007/08. Model standard deviations are indicated by the shading.

Table 1: Regional average defined by all predictor indices, as well as the three-month average index period with the strongest correlation (r) with annual TCG totals in the upcoming season. All correlations presented here are significant at the 95% confidence level.

Index	**Latitude**	**Longitude**	**3-month mean**	**r**
SLP	5°-15°N	165°-125°W	JJA	0.57
GPH	25°-45°S	170°-130°W	JAS	0.47
EPT	20°-30°S	160°E-160°W	JAS	0.38
NINO4	5°N-5°S	160°E-150°W	JAS	-0.47
v_{850}	15-5°N	120°-80°W	MJJ	0.37
u_{200}	5°N-5°S	170°E-150°W	JJA	0.38
EVWS	0°-10°S	170°-130°W	JAS	0.50
RV	15°-5°N	145°-105°W	JJA	-0.49

Table 2: a) Model cross-validated hindcast skill (based on the leave-one out method and three skill metrics) for different predictors and combinations. RMSEs are used to evaluate the quality of the predictor. Next to the RMSEs, standard errors (se) and significant correlation coefficients (r) between hindcasts and observed number of TCs help to further validate the skill of the model. Bold indicates the model with the lowest RMSE. b) as a) but for two- and three-predictor combinations based on step-by-step model selection. For the multi-predictor models, bold indicates improvement over the best single predictor SLP model and the best two-predictor model.

a) ***predictor***	**RMSE**	**se**	**r**
Climatology	4.04	0.24	NaN
SLP	**3.79**	**0.26**	**0.53**
GPH	3.92	0.25	0.37
EPT	4.06	0.27	0.29
NINO4	4.04	0.28	0.31
v_{850}	4.03	0.26	0.17
u_{200}	4.03	0.28	0.24
EVWS	3.96	0.27	0.47
RV	3.87	0.27	0.39

b) ***predictor***	**RMSE**	**se**	**r**
SLP+GPH	3.86	0.26	0.50
SLP+EPT	3.88	0.28	0.47
SLP+NINO4	3.92	0.28	0.49
SLP+ v_{850}	**3.74**	0.26	**0.57**
SLP+ u_{200}	3.87	0.27	0.49
SLP+EVWS	3.79	0.26	0.55
SLP+RV	3.83	0.27	0.50

4.5 Further Discussion

For a better evaluation of the robustness of the predictor-set for annual TCG counts in the eastern Indian Ocean region, we applied a *four*-fold cross-validation technique. In this method, the data are split into *four* consecutive 10-year subsets. The model is then trained on three of the four subsets to hindcast the left-out 10 years. This procedure was used to hindcast the 10-year periods 1968/69-1977/78, 1978/79-1987/88, 1988/89-1997/98 and 1998/99-2007/08. Figure 4.5 shows the four independent *four*-fold 10-year hindcasts, as well as the fitted 30-year hindcasts on which the model was trained. The results are compelling with the model displaying considerable skill. The independent hindcasts of the left-out 10 years are mostly very accurate, with only five out of 40 (12.5%) of the annual (seasonal) hindcasts outside of the model standard deviation. As for the leave-one out cross-validated data, the model lacks skill when an unusually high or low number of TCGs occurred (1985/86, 1986/87, 1993/94, 2003/04). For the year 2004/05, the unusually high number of 9 TCs during an El Niño year was not well represented by the model. The fewer seasons in which the observed number of TCG counts is within the model standard deviation are due to larger model coefficient uncertainties.

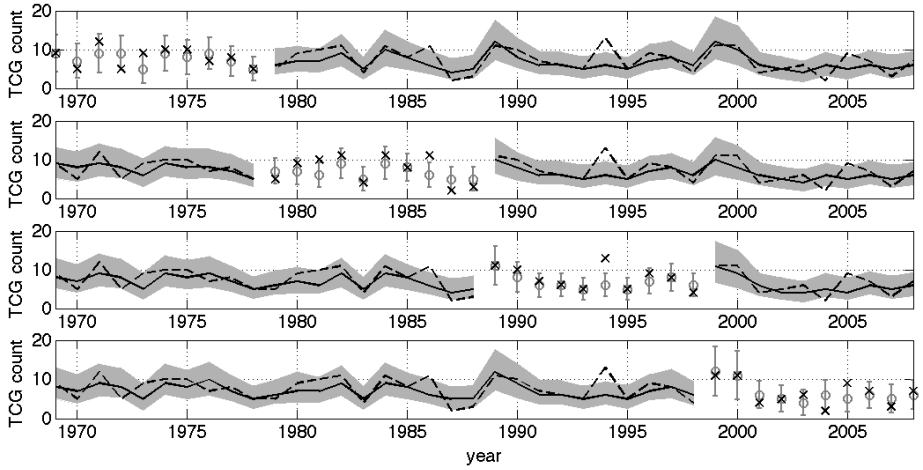


Figure 4.5 Variations in observed annual TCG totals (dashed line) and regressed hindcasts (solid line) using the SLP+v_{850} predictor model trained on 30 years and independently hindcasting the 10 left-out years (circle; *four*-fold cross-validation), with error bars indicating the standard deviation. Crosses indicate the observed TCG totals during the independent hindcast period. Model standard deviations of the 30-year training period are indicated by the shading.

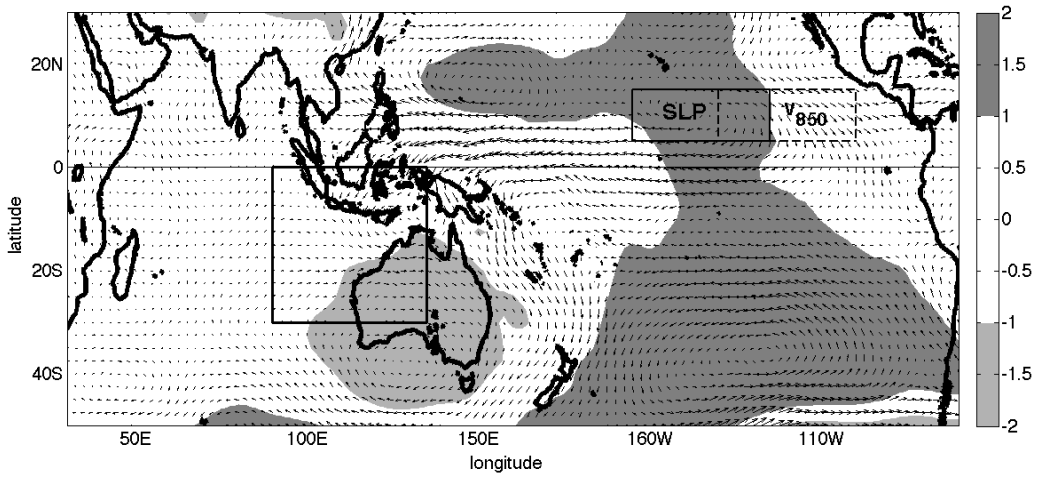


Figure 4.6 Map of composite anomalies for active Australian TC seasons minus composite anomalies for inactive eastern Indian Ocean TC seasons during June-July-August. Wind vectors describe the wind flow anomalies at 850 hPa. Grey shaded areas represent changes of SLP (hPa). Also indicated are the locations of the predictor indices SLP and v_{850} and the eastern Indian Ocean TC region.

The SLP and v_{850} indices derived in this study are ENSO-linked. Combining central tropical North Pacific sea level pressure with the lower-troposphere meridional tropical inflow from the eastern North Pacific, was found to produce a valuable and complementary predictor set. Figure 4.6 shows the anomalies of the model-relevant predictor variables (SLP, wind flow at 850 hPa) between hindcasted active TC seasons (TCGs $\geq$ 9) and relatively inactive TC seasons (TCGs $\leq$ 4) and indicates the index locations used in the two-predictor model. Based on this definition, active and inactive seasons represent around 25% of the investigated seasons, respectively. The SLP index location is embedded in an ENSO-related large-scale SLP anomaly. The relationship between TCG occurrence and the ITCZ-linked North Pacific v_{850} index is negative, which means a strengthening of the low-level equatorward inflow (Hadley circulation) leads to enhanced TCG.

4.6 Chapter Summary

This chapter has addressed the thesis aims identifying important prediction schemes of climate variables for seasonal variations of TCG counts in the Australian region and subregions and on its basis, seasonal forecast models for TCG counts using a Bayesian Poisson model was built.

We presented a new Bayesian seasonal forecast model of Australian region TCG counts. We showed substantial improvement over previous seasonal forecast models of TCG counts in the Australian region, as well as the advantage of the Bayesian method over the frequentist approach. The model has been shown to be adaptable for forecasting of southwest Pacific (Eastern Australian) region TCG counts. Applying the model to hindcast seasonal TCG counts in the eastern Indian Ocean (Western) region lacks sufficient skill and suggests that it is beneficial to build a separate seasonal forecast model for that region on the basis of different predictor indices.

5 DEVELOPMENT OF A SEASONAL FORECAST MODEL FOR THE SPATIAL PROBABILITY OF TROPICAL CYCLONE OCCURRENCES IN THE AUSTRALIAN REGION

5.1 Chapter Overview

This chapter applies the Bayesian forecast model introduced in Chapter 4 to build a forecast model for the spatial probability of TCs forming in the Australian region. As we are now interested in the probability of an event occurring, instead of the expected number of events, the Poisson likelihood function is replaced by a logistic regression applied with a Bernoulli distribution. An accurate representation of the seasonal variations of the probability of TCs forming is achieved, as well as the spatial variations. The model is based on the previously introduced indices of SLP and v_{850}, as well as the ENSO SSTA index NINO4 and spatial information CAPE.

The main text of this chapter is a paper being prepared for submission to the Journal *Climate Dynamics* (Werner and Holbrook 2011c: A spatial statistical forecast model of seasonal probabilities for Australian region tropical cyclone formation - in preparation for submission to *Climate Dynamics*).

Candidate's contribution to the papers

The experiment design and analysis methods were the candidate's idea and then jointly discussed between Dr Holbrook and myself. I also performed all of the data analysis. All sections of the coauthored paper in preparation were led by myself under the guidance of Dr Holbrook.

A spatial statistical forecast model of seasonal probabilities for Australian region tropical cyclone formation

Angelika Werner
Department of Environment and Geography, Macquarie University, NSW 2109
Australia

Neil J. Holbrook
School of Geography and Environmental Studies, University of Tasmania, TAS 7001
Australia

For submission to *Climate Dynamics*

Corresponding author address: Assoc Prof Neil J Holbrook,
School of Geography and Environmental Studies, University of Tasmania, TAS 7001
Australia
Phone: +61 3 6226 2027 | Fax: +61 3 6226 7628 | E-mail: **Neil.Holbrook@utas.edu.au**

Keywords Tropical Cyclone Formation, Seasonal Forecasting, Spatial, Probabilities

Abstract

A new and potentially skilful seasonal forecast model for the spatial distribution of tropical cyclone development (genesis; TCG) in the Australian region on a 2.5° x 2.5° grid is developed for the Australian region. The model is based on logistic regression using the Bayesian approach. Predictor combinations are chosen using a step-by-step predictor selection. The four-predictor model based on indices of June-July-August averaged tropical central Pacific sea level pressure, July-August-September averaged NINO4, May-June-July averaged tropical North Pacific meridional winds at 850 hPa and the spatially-varying (grid-point estimates) of April-May-June averaged convective available potential energy shows an average improvement over the climatological average of 25.2% and a spatially averaged root mean-squared error of 0.41. The average distribution of TCG probabilities over the study period, as well as the hindcasted strong variations of probabilities and distribution of TCG during El Niño-Southern Oscillation events match remarkably well against observations over most of the study domain, with a high skill for the whole range of probabilities.

1. Introduction

Australia's tropical climate is dominated by El Niño-Southern Oscillation (ENSO), driven largely from the Pacific basin (e.g., Allan et al. 1996). The number of observed Australian region tropical cyclones (TCs) in any particular year depends strongly on ENSO. Consequently, numerous studies have been undertaken previously regarding the influence of ENSO on tropical cyclone formation in this region (e.g., Nicholls 1984, 1992; Basher and Zheng 1995; Kuleshov and de Hoedt 2003; Ramsay et al. 2008; Kuleshov et al. 2009)**.** The Australian TC region has greater variability of TC frequency, speeds and trajectories compared with other TC basins (Bessafi et al. 2002). This is the result of unique climatological and other features, including the Australian landmass in the region (Holland 1984), the existence of the Australian monsoon trough (McBride and Keenan 1982; Evans and Allan 1992), and the relatively close approach of the mid-latitude westerlies into the tropics (Dare and Davidson 2004). While the monsoon trough is an area of strong negative low-level relative vorticity, the upper-level ridge located around 12°S is accompanied by weak environmental vertical wind shear in a zone that is extremely favourable for TC formation and intensification (Dare and Davidson 2004). Additionally, Australian region TC formation (genesis; TCG) is affected by continental effects. When warm, dry continental air from Western Australia extends offshore, extensive dry ambient conditions are less conducive to the formation of Australian's west coast TCs. On the other hand, when the monthly-averaged relative humidity is increased and concentrated in the west coast region, intense TCs can be expected (Tonkin et al. 1997). During the tropical cyclone season, mid-tropospheric relative humidity and

vertical wind shear are important contributors to variations in TCG in the Southern Hemisphere (Camargo et al. 2007). Furthermore, relative humidity at 500hPa and low-level relative vorticity appear to be the most important variables for Australian region TCG on interannual time scales (Kuleshov et al. 2009).

Since the late 1970s/early 1980s, a number of statistical seasonal forecast schemes have been developed and improved to predict TC activity in various basins and sub-basins (Klotzbach et al. 2010). Seasonal forecast modelling of TC activity was first undertaken by Nicholls (1979) for the Australian region and Gray (1984) for the North Atlantic. The first nonlinear statistical forecast model of Australian region tropical cyclone frequency was developed 20 years ago, based on the Poisson regression (Solow and Nicholls 1990). In that study, the large-scale atmospheric sea level pressure gradient between Darwin and Tahiti – the Southern Oscillation index (SOI) – was used as the predictor of Australian region TC counts. More recently, a Poisson regression model using spatial grid-point estimates of the September lead saturated equivalent potential temperature gradient between 1000hPa and 500hPa, as well as the SOI, was developed to forecast upcoming season TCG totals across the Australian tropical cyclone region (McDonnell and Holbrook 2004a,b). Werner and Holbrook (2011) built on that study with the development of a Poisson regression model using a Bayesian approach, with predictor variables that include the subtropical central South Pacific convective available potential energy, meridional winds at 850 hPa in the tropical East Pacific (the Walker circulation inflow region), and the central South Pacific geopotential height at 500 hPa. Both the McDonnell and Holbrook (2004b) and Werner and Holbrook (2011) models have also been applied to forecast TCG totals within three Australian subregions. In each case, the TCG model forecasts showed considerable improvement over climatology of 25% (McDonnell and

Holbrook 2004b) and 58% (Werner 2011) for the southwest Pacific, while neither achieves significant skill in the eastern Indian Ocean subregion. Liu and Chan (2010) applied a project-pursuit regression model to forecast seasonal variations in TCG counts for the Australian region and two subregions using well-known climate indices - NINO4, the Indian Ocean Dipole Mode index (DMI), the trade wind index, and outgoing long-wave radiation (OLR) index. Using the jackknife method, they achieved 51% improvement in the root-mean-square error (RMSE) over climatology for the Australian region and 39% improvement for the Western Australian (eastern Indian Ocean) region. McDonnell and Holbrook (2004a,b) took account of (grid-scale) spatial variations in saturated equivalent potential temperature gradient in their Poisson regression model on a 5°longitude x 2°latitude grid. While they achieved a 22% improvement of RMSE over climatology in expected TCG totals at the basin-wide scale, they were unable to accurately capture the spatial distribution of seasonal TCG probabilities at the grid-point scale. Most recently, a statistical model developed for the Australian subregions by Leroy and Wheeler (2008), based on logistic regression, has had some success in forecasting TCG probabilities of occurrence at the intra-seasonal time-scale.

The present paper introduces a new and novel spatial statistical forecast model of Australian region TCG seasonal probabilities on a 2.5° x 2.5° grid. Our logistic regression model is developed here using a Bayesian Markov Chain Monte Carlo (MCMC) approach, and takes account of recently derived climate indices shown to be useful for predicting Australian region TCG (Werner and Holbrook 2011). As an average, the model provides a 25% improvement over climatology. The model utility is in the fact that we are able to now provide forecasted TCG event probabilities at the grid-point scale across the entire Australian tropical cyclone region. Model skill is

also assessed through hindcasts of TCG probabilities during ENSO events (both El Niño and La Niña year), where we find that TCG hindcast probabilities match remarkably well against observations over most of the study domain.

The paper is structured as follows. Section 2 describes the data and model predictors used. The model details are provided in Section 3. Section 4 presents the model results, and finally Section 5 discusses and summarizes the quality and improvements of the models presented over existing models.

2. Data

a. Tropical cyclone observations

This study takes advantage of the global TC best track data set IBTrACS.v02 (Knapp et al. 2010) provided by the U.S. National Oceanic and Atmospheric Administration. TCG is here defined as the location where a tropical storm system with winds exceeding 34 knots (17.5 ms^{-1}) was originating.

The Australian tropical cyclone region is defined here as 0°-30°S and 90°-170°E. TCG occurrences identified poleward of 30°S, or over land, have been removed in the quality assessment process. Only storms during the Australian TC season from November to April are taken into account. Overall, a total of 570 TCs during the 40-year period from 1968/69-2007/08 are analyzed following the quality control. Figure 1 shows the spatial distribution of the first noted locations of tropical storm systems that later developed into TCs and the corresponding time series of seasonal TCG totals in the 40-year record. TCG locations were subsequently binned into their corresponding 2.5° x 2.5° grid cells. This results in a 33 x 13 grid cells within the Australian TC region. With an average of 14 TCGs occurring per season, this leads to very few data points with TCG ≥ 1 in the time-space grid. To increase the

spatial information, but also to smooth the randomness of the location of TCG due to favourable local conditions, we applied a spatial running grid of three zonal and meridional grid boxes. This results in spatial running grid consisting of a total of nine grid cells, for which the TCG occurrences are added up to better estimate the likelihood for TCG in the centred box.

b. Oceanic and atmospheric data

Sea surface temperature (SST) data were taken from uniformly gridded temperature observations provided in the Hadley Centre Global Sea Ice and Sea Surface Temperature (HadISST1) dataset (Rayner et al. 2003), compiled by the UK Met Office Hadley Centre. HadISST1 is a combination of global monthly SST fields and sea ice concentrations on a 1° × 1° grid from 1870 to the present.

The atmospheric data analyzed in this study are from the NCEP/NCAR monthly mean upper-air reanalyses, with 2.5° horizontal resolution on 17 pressure levels (Kalnay et al. 1996). Monthly anomalies of all variables and indices were determined against a 30-year base period of 1970-1999. Statistical significances of the correlation coefficients are based on the reduced effective number of degrees of freedom method outlined by Davis (1976).

c. ENSO definitions and effects on TCG

Ramsay et al. (2008) found the NINO4 SST anomaly (SSTA) index to be the strongest ENSO predictor of interannual TC frequency in the Australian region and is therefore also included in the present study. NINO4 is defined as the SSTA time series averaged spatially between 5°S-5°N and 160°E-150°W.

ENSO events are classified according to the definition used by the U.S. National

Weather Service (<http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml>) using the three-month running mean in the NINO3.4 region. El Niño (La Niña) events are defined by the NINO3.4 SSTA exceeding thresholds of ±0.5°C for a minimum of five consecutive three-month average overlapping periods.

d. Model predictors

Next to the ENSO metric NINO4, two additional indices are used, derived on the basis of significant and persistent correlation patterns between sea level pressure (SLP) and meridional winds at 850 hPa (v_{850}) with seasonal variations of Australian region TCG counts. Figure 2 shows spatial correlation maps between Australian region annual TCG totals and June-July-August (JJA) averaged SST, SLP and v_{850}, prior to the upcoming TC season. SLP shows a strong positive correlation pattern throughout the Pacific that is strongest in the tropical and subtropical central Pacific, while negative correlations exist over much of the Australian continent, south of Australia, and in the southeast Indian Ocean associated with large-scale ENSO (Garreaud and Battisti 1999). The SST correlation pattern is characterized by the developing ENSO (e.g., Drosdowsky and Chambers 2001) with a weak boomerang-shaped negative SSTA pattern featured in the central Pacific. An extended anomaly, of the same negative sign as in the central Pacific, is also observed in the Indian Ocean. The correlation patterns with v_{850} also describe enhanced Australian region TCG being related to a weakening of the Hadley circulation in the Western Pacific Warm Pool (WPWP) region, and a strengthening of the Hadley circulation in the northeast Pacific.

To use climatologically relevant predictors, we applied indices derived in

previous studies on the basis of the persistence and strength of the pre-seasonal spatial correlation coefficients between climate variables and Australian and southeast Indian Ocean region TCG variability (Werner and Holbrook 2011a,b). The three-month averaged predictors were chosen based on the strongest correlations between pre-TC season indices and upcoming season Australian region TCG. We found that the strongest predictor indices based on our assessment are the central tropical Pacific SLP (SLP; 5°-15°N, 165°-125°W), the western Pacific ENSO SSTA index - NINO4 (5°N-5°S, 160°E-150°W), and the meridional winds at 850 hPa in the northeast tropical Pacific (v_{850}; 15-5°N, 120°-80°W). Also used are the 2.5° x 2.5° grid-scale spatial estimates of April-May-June (AMJ)-averaged convective available potential energy calculated between 850 hPa and 300 hPa (CAPE). CAPE ($m^2\ s^{-2}$) describes the stability of the troposphere, and is defined as

$$CAPE = g \int_{z(850hPa)}^{z(300hPa)} \frac{T_m - T}{T} dz \quad , \qquad (1)$$

where z(850hPa) is assumed to approximate the level of free convection, and z(300hPa) the level of neutral buoyancy. T (°C) is a function of the pressure level height (z) – and defined as the environmental temperature, T_m (°C) is the temperature of an idealized rising air parcel which is assumed to be saturated at the 850 hPa level, and g = 9.81 m s^{-2} is the standard gravity constant.

Figure 3 shows correlation maps between the four selected predictor variables and the spatial running means of upcoming season TCG totals across the Australian region. While SLP and NINO4 correlate significantly with TCG in the Pacific and parts of the Indian Ocean, v_{850} and CAPE only show broad significant relationships in the southwest Pacific sector. There are two regions where none of the indices or CAPE appears to have any skill on the basis of the spatial correlation coefficients.

These regions are located in the Indian Ocean between 105° and 115°E and over the northern Australia between 125° and 145°E, and therefore in proximity or around local minima of TC activity (Goebbert and Leslie 2010).

3. Bayesian Regression Model

a. Logistic Regression

The logistic regression model is an appropriate method that can be used to model the probabilities of an event, i.e., the outcome variable is binary. Following previous studies (e.g., Leroy and Wheeler 2008), we applied the logistic regression approach to model the probability of TCG events. The apparent difficulty that seasonal TCG events are integer numbers, rather than probabilities, can be solved with Regression Estimation of Event Probabilities (REEP; Glahn 1985). In REEP, an assumption based on the underlying data set or experience is made, so the predictand takes on values between 0 and 1. This enables us to treat the predictand as the probability of an event occurring, and therefore the application of a Bernoulli distribution (Wilks 1995). The relevant assumption made in our case study is that, for the nine grid boxes used in the spatial running grid, the TCG event probabilities within each running spatial grid-space are divided by nine. If the total number of observed TCG occurrences then exceeds 9, the probability of occurrence would be theoretically higher than one, but is assumed as 1. With REEP, a Binomial distribution can be applied to the observational data as

$$P(Y_i = y) = \left(\frac{n!}{k!(n-k)!}\right) p_i^{\,k} (1 - p_i)^{n-k} \quad , \qquad (2)$$

where n is the REEP assumption for the total number defined as p=1, k identifies the number of observed events and p_i is the logistic regression as in

$$P(Y_i = y \mid x_i, \beta) = \frac{\exp(\mu_i)}{1+\exp(\mu_i)} \quad , \tag{3}$$

where μ_i provides the multiple linear regression estimates on the predictors. The logistic regression is defined for all positive and negative discrete numbers. With the here described approach, we can consider the response variable Y_i as a set of i independent Bernoulli trials with different success probabilities of an event occurring.

b. Bayesian analysis

Bayes' theorem allows us to find the best possible model coefficient representation – specifically, this information is used here to develop model forecast estimates of Australian region TCG annual totals. Following Werner and Holbrook (2011), we used diffusive prior information, selecting a Gaussian distribution with zero mean and standard deviation of 100 for all model coefficients. For the likelihood function, we consider the Bernoulli model resulting from Equation (2) and (3) in

$$p(y_i \mid x_i, \beta) = \left[\frac{\exp(\mu_i)}{1+\exp(\mu_i)}\right]^{y} \left[1 - \frac{\exp(\mu_i)}{1+\exp(\mu_i)}\right]^{1-y} . \tag{4}$$

A Markov Chain Monte Carlo (MCMC) method was then applied using slice sampling. The posterior probability quantifies the uncertainty in the model coefficients provided by the probabilities. To discard the effects of the chosen initial conditions, we applied a model burn-in of 500 iterations. To avoid high autocorrelations, and hence aiming to achieve effective statistical independence of the samples in the iteration process, the samples were thinned so that only every sixth sample was taken into account. In summary, 5,000 samples were used to estimate the model coefficients and obtain the model forecasts. For detailed information about the modelling approach, see Werner and Holbrook (2011).

d. Model skill

The final model coefficients were estimated based on data from the 40-year training period of 1968/69-2007/08. To evaluate the model skill, a skill-score (SS) is calculated from the mean-squared error (MSE) improvement of the model hindcasts over the climatological-average MSE. The root mean-squared error (RMSE) and correlation coefficient (r) between the predicted and observed number of TCs formed in the region provides an additional measure of the first-order hindcast skill.

4. Results

Based on a systematic step-by-step assessment of the utility of the model predictors, and through a series of cross-validated hindcasts, we found a four-predictor SLP+NINO4+v_{850}+CAPE logistic regression model using Bayesian probabilities displays the greatest potential for skilful spatial forecasts of TCG seasonal probabilities. Figure 4 shows the 40-year average observed and hindcasted probabilities for Australian region TCG. Clearly, the average distribution is accurately captured in the model hindcasts. The model shows improvements (30-50%) over climatology in both the Pacific and western border region of the eastern Indian Ocean, with a mean average skill score of SS=25.2% (Figure 5). The model provides modest improvements in the Australian North subregion and eastern Indian Ocean between 105° and 115°E. The RMSEs (Fig. 5, lower panel) highlight first-order hindcast skill and the relative TC activity spatially, with the area-averaged RMSE = 0.41. The highest RMSEs, and hence the strongest deviation of model hindcasts and observations, are in the far western region around 90°E. High RMSEs are also found in the Australian North subregion, where skill levels are typically poor – not only here, but also in all previous literature we have been able to scope. This is further

quantified by correlations between the observations and hindcasts of $r < 0.3$ in that subregion. Nevertheless, across most of the southwest Pacific and eastern Indian Ocean regions, the very good correlations affirm that the model performs well overall as a spatial forecasting tool of seasonal TCG probabilities at scales of only a few degrees of latitude and longitude. Furthermore, we find the model is potentially skilful across a range of probabilities (Figure 6), albeit that there appears to be a tendency to underestimate higher probability TCG events. However, despite this apparent bias, the overall hindcast performance is remarkably good.

Due to the strong influence of ENSO climate variations on TCG counts and their distribution across the tropical Indo-Pacific Ocean regions, we also assess the spatially-varying hindcast skill of seasonal TCG probabilities during El Niño and La Niña years from 1968/69-2007/08. To evaluate model performance, we compare El Nino and La Nina event composites of observed and hindcast TCG event probabilities using the four-predictor SLP+NINO4+v_{850}+CAPE logistic model (Figure 7). We find that the spatial distributions of TCG-event probabilities in both phases of ENSO are remarkably well captured overall. Despite a slight underestimation of the probabilities just off the North Queensland coast, and in the Australian North subregion for both phases of ENSO, the model performs very well in capturing the changing TCG probabilities and distributions according to the large climatic swings associated with ENSO event phase-changes.

5. Summary and Discussion

This paper provides a new and novel Bayesian seasonal forecasting model, based on the logistic regression, giving spatially-varying grid-scale forecasted estimates of seasonal tropical cyclone formation (genesis; TCG) probabilities across

the Australian tropical cyclone region. We have demonstrated that this model provides substantial improvement in potential forecast skill over climatology (SS = 25.2%, RMSE = 0.41).

We are aware of only one previous body of work that has attempted to develop a spatial forecast model of TCG for the Australian region (McDonnell and Holbrook 2004a,b). In that study, they applied a Poisson regression model on a 5°longitude x 2°latitude grid based on the SOI and grid-scale estimates of the saturated equivalent potential temperature as predictors. Despite providing a 22% improvement of RMSE over climatology in expected TCG totals at the basin-wide scale, they were unable to accurately capture the spatial distribution of seasonal TCG probabilities at the grid-point scale.

To avoid the problem of data sparsity and randomness, a spatial running-mean grid of three horizontal and vertical grid boxes was applied. This enables us to regress on probabilistic predictands by dividing the total number of seasonal TCG events in each of the running grids by the number of subgrids, where we assume the probability represented in the centred box of the spatial running grid is representative of the aggregate of boxes. Following recent studies (e.g., Elsner and Jagger 2006; Chu and Zhao 2007; Chand et al. 2010; Werner and Holbrook 2011), we applied Bayes' theorem using the Markov Chain Monte Carlo method to estimate seasonal TCG event totals. The Bayesian approach is beneficial as it permits the incorporation of prior beliefs, and is convenient to account for the uncertainties in model parameters. To respond to the binary outcome variable, a logistic regression model is then applied. We found that the most skilful predictor combination of both standard and derived indices (i.e., area-averaged time-series that summarize large-scale climate changes) are the June-July-August (JJA)-averaged central tropical Pacific SLP, July-August-

September (JAS)-averaged NINO4, May-June-July (MJJ)-averaged tropical North Pacific v_{850}, as well as the spatially-varying (grid-point estimates) April-May-June (AMJ)-averaged CAPE.

The SLP and NINO4 indices used in the model are ENSO-linked, while the v_{850} index appears to vary due to a different mechanism. Combining central tropical Pacific SLP with central tropical Pacific SST anomalies and the lower-troposphere meridional tropical inflow from the Northern Hemisphere into the eastern tropical Pacific, produced a valuable and complementary predictor set. The East Pacific tropical SLP index is embedded in a large-scale pressure pattern associated with ENSO. The NINO4 index provides a time history of SSTA changes in the western tropical Pacific (at the western end of the equatorial cold tongue connected to interannual ENSO variations) that has been previously shown to be important for capturing Australian region TCG variability (e.g., Ramsay et al. 2008; Liu and Chan 2010). The significant negative relationship between eastern Indian Ocean TCG annual totals and the ITCZ-linked North Pacific index, v_{850}, means that a strengthening of the low-level equatorward inflow from the Northern Hemisphere leads to enhanced TCG. CAPE, as a measure of tropospheric instability, is used to provide grid-scale (spatially-varying) estimates of the tendency for deep convection.

The model developed here shows improvements (30-50%) over climatology in both the Pacific and western border region of the eastern Indian Ocean, with a mean average skill score of SS=25.2%. This is a substantial improvement over the frequentist Poisson regression model results provided by McDonnell and Holbrook (2004b), who showed an RMSE improvement over the climatology of 22%. Further, the almost grid-scale seasonal TCG probabilistic estimates accurately provided here, including ENSO TCG event probabilities that are well-captured as composite El Nino

and La Nina phases, should represent a substantial step forward in forecast potential for Australian region TCG. We note, however, important forecast difficulties evident in the eastern Indian Ocean region between 105° and 115°E and in the Northern Australian region. This historically low forecast skill eastern Indian Ocean region is also notably an area of local minimum TC activity around 105°E (Goebbert and Leslie 2010). In our attribution diagram (Fig. 6), we show that overall our model provides relatively high skills for all selected probabilities, albeit that there is a trend towards slightly underestimating the higher probability events that are observed.

In summary, we have developed a new and potentially skilful seasonal forecast model for the spatial probability of tropical cyclogenesis for the Australian region on a 2.5° x 2.5° resolution. We find that a four-predictor SLP+NINO4+v_{850}+CAPE logistic regression model produces remarkably skilful hindcasts of Australian region seasonal spatial TCG probability by September of each year, one month prior to the onset of the Australian region TC season (November-April). The predictor variables identified in this study are physically meaningful and appropriate to condition the model forecasts of TCG. By combining information from useful dynamic and thermal and synoptic variables as predictors in a Bayesian Bernoulli approach, we are able to demonstrate skilful cross-validated hindcasts of Australian region seasonal spatial TCG probability with strong average improvement over the climatological average of 25% against a 40-year record of observations.

References

Allan R, Lindesay J, Parker D (1996) El Niño: Southern Oscillation and Climatic Variability. CSIRO Publishing, 416pp

Basher RE, Zheng X (1995) Tropical cyclones in the southwest Pacific: Spatial patterns and relationships to Southern Oscillation and sea surface temperature. J Climate 8: 1249-1260

Bessafi M, Lasserre-Bigorry A, Neumann CJ, Pignolet-Tardan F, Payet D, Lee-Ching-Ken M (2002) Statistical prediction of tropical cyclone motion: An analog-CLIPER approach. Wea Forecasting, 17, 821–831

Camargo SJ, Emanuel KA, Sobel AH (2007) Use of a genesis potential index to diagnose ENSO effects of tropical cyclone genesis. J Climate 20: 4819–4834, doi:10.1175/JCLI4282.1

Chand S, Walsh KJE, Chan JCL (2010) A Bayesian Regression Approach to Seasonal Prediction of Tropical Cyclones Affecting the Fiji Region. J Climate, 23: 3425-3445

Chu P-S, Zhao X (2007) A Bayesian regression approach for predicting tropical cyclone activity over the central North Pacific. J Climate, 20, 4002–4013

Dare RA, Davidson NE (2004) Characteristics of Tropical Cyclones in the Australian Region. Mon Wea Rev, 132, 3049-3065

Davis RE (1976) Predictability of sea surface temperatures and sea level pressure anomalies over the North Pacific Ocean. J Phys Oceanogr 6: 249-266

Drosdowsky W, Chambers LE (2001) Near global scale sea surface temperature anomalies as predictors of Australian seasonal rainfall. J Climate, 14, 1677-1687

Elsner JB, TH Jagger (2006) Prediction models for annual U.S hurricane counts. J. Climate 19: 2935–2952

Evans JL, Allan RL (1992) El Nino/Southern Oscillation modification to the structure of the monsoon and tropical cyclone activity in the Australasian region. Int J Climatol 12: 611-623

Garreaud RD, Battisti DS (1999) Interannual (ENSO) and interdecadal (ENSO-like) variability in the Southern Hemisphere tropospheric circulation. J Climate, 12, 2113–2123

Glahn HR (1985) Statistical weather forecasting. In: A.H. Murphy and R.W. Katz, eds., Probability, Statistics, and Decision Making in the Atmospheric Sciences. Boulder, Westview, 289–335

Goebbert KH, Leslie LM (2010) Interannual Variability of Northwest Australian Tropical Cyclones. J Climate, 23, 4538-4555

Gray WM (1984) Atlantic seasonal hurricane frequency. Part II: Forecasting its variability. Mon Wea Rev, 112, 1669–1683

Holland GJ (1984) On the climatology and structure of tropical cyclones in the Australian/southwest Pacific region: I. Data and tropical storms. Aust Meteor Mag, 32, 1–15

Kalnay E et al (1996) The NCEP/NCAR 40-Year Reanalysis Project. Bull. Amer. Meteor. Soc. 77: 437-471

Klotzbach, PJ, Barnston A, Bell G, Camargo SJ, Chan JCL, Lea A, Saunders M, Vitart F (2010) Seasonal Forecasting of Tropical Cyclones. Global Guide to Tropical Cyclone Forecasting, 2nd edition, C. Guard, editor, WMO, in press

Knapp KR, Kruk MC, Levinson DH, Diamond HJ, Neumann CJ (2010) The international best track archive for climate stewardship (IBTrACS). Bull Amer

Meteor Soc 91: 363-376

Kuleshov Y, de Hoedt G (2003) Tropical cyclone activity in the Southern Hemisphere. Bull Austral Meteor Oceanogr Soc, 16, 135-137

Kuleshov Y, Chane Ming F, Qi L, Chouaibou I, Hoareaux C, Roux F (2009) Tropical cyclone genesis in the southern hemisphere and its relationship with the ENSO. Ann Geophys 27: 2523–2538

Leroy A, Wheeler MC (2008) Statistical Prediction of Weekly Tropical Cyclone Activity in the Southern Hemisphere. Mon Wea Rev, 136, 3637-3654

Liu KS, Chan JCL (2010) Interannual variation of Southern Hemisphere tropical cyclone activity and seasonal forecast of tropical cyclone number in the Australian region. Int J Climatol (2010), DOI: 10.1002/joc.2259

McBride JL, Keenan TD (1982) Climatology of tropical cyclone genesis in the Australian region. Int J Climatol, 1, 13-33

McDonnell KA, Holbrook NJ (2004a) A Poisson regression model of tropical cyclogenesis for the Australian-southwest Pacific Ocean region. We. Forecasting, 19, 440-455

McDonnell KA, Holbrook NJ (2004b) A Poisson regression model approach to predicting tropical cyclogenesis in the Australian/southwest Pacific Ocean region using the SOI and saturated equivalent potential temperature gradient as predictors. Geophys Res Lett, 31, L20110, doi:10.1029/2004GL020843

Nicholls N (1979) A possible method for predicting seasonal tropical cyclone activity in the Australian region. Mon Wea Rev, **107,** 1221–1224

Nicholls N (1984) The Southern Oscillation, sea-surface temperature, and interannual fluctuations in Australian tropical cyclone activity. J Climatol. 4: 661-1149

Nicholls N (1992) Recent performance of a method for forecasting Australian seasonal tropical cyclone activity, Aust Meteor Mag, 40, 105–110

Ramsay HA, Leslie LM, Lamb PJ, Richman MB, Leplastrier M (2008) Interannual variability of tropical cyclones in the Australian region: Role of large-scale environment. J Climate 21: 1083-1103

Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. J Geophys Res 108: (D14)

Solow A, Nicholls N (1990) The relationship between the Southern Oscillation and tropical cyclone frequency in the Australian region. J Climate 3: 1097-1101

Stone M (1974) Cross-validatory choice and assessment of statistical predictions. J Royal Stat Soc 36(2): 111–147

Tonkin H, Landsea C, Holland GJ, Li S (1997) Tropical Cyclones and Climate Change: A preliminary Assessment. Assessing Climate Change: Results from the Model Evaluation Consortium for Climate Assessment (Ed. W. Howe and A. Henderson-Sellers). Gordon and Breach, Australia, ISBN: 90- 5699-067-5, pp 323-354

Werner A (2011) Seasonal Forecasting of Australian region tropical cyclone formation. Ph.D thesis, Macquarie University. 240pp (submitted, 23 May 2011)

Werner A, Holbrook NJ (2011a) A Bayesian forecast model of Australian region tropical cyclone formation. J Climate (accepted, 13 May 2011)

Werner A, Holbrook NJ (2011b) How to improve seasonal forecast modeling of tropical cyclone formation in the southeast Indian Ocean. in preparation

Wilks DS (1995) Statistical Methods in the Atmospheric Sciences. Academic Press, 467 pp

FIG. CAPTIONS

Fig.1 Spatial distribution of the origins of tropical storm systems with winds exceeding 34 knots (17.5 ms^{-1}) in the Australian region from 1968/69-2007/08

Fig. 2 Spatial maps of correlations between three different June-July-August (JJA)-averaged large-scale climate variables with upcoming Australian region TCG seasonal totals. The bold box indicates the Australian TC region and the dashed box identifies the predictor region with maximum relationship with Australian region seasonal TCG. The thin lines outline the spatial pattern correlations significant at the 95% confidence level, taking account of serial correlation

Fig. 3 Spatial maps of correlations between the spatial running grid (binned in nine 2.5° x 2.5° grid-cells) of observed TCG counts across the Australian tropical cyclone region and four separate climate predictor variables. The thin lines outline spatial pattern correlations significant at the 95% confidence level, taking account of serial correlation

Fig. 4 Spatial distribution of 40-year record-average TCG: (a) observed probabilities; and (b) hindcasted probabilities; across the Australian region from 1968/69-2007/08

Fig. 5 (upper panel) Spatial distribution of skill scores based on model hindcasts of seasonal TCG event probabilities using the four-predictor SLP+NINO4+v_{850}+CAPE model. (lower panel) Spatial distribution of corresponding RMSEs shaded in red - the overlying contour lines outline the correlation coefficient distribution between the observed and hindcasted TCG event probabilities for the period 1968/69-2007/08

Fig. 6 Attributes diagram of model hindcasts. In the subpanel, the bar diagram summarizes the total distribution of hindcast probabilities, while the large outer diagram compares the observed probabilities with the hindcasted probabilities, within the probability range

Fig. 7 El Nino and La Nina event composite maps of observed and hindcasted TCG probabilities across the Australian region

FIGURES

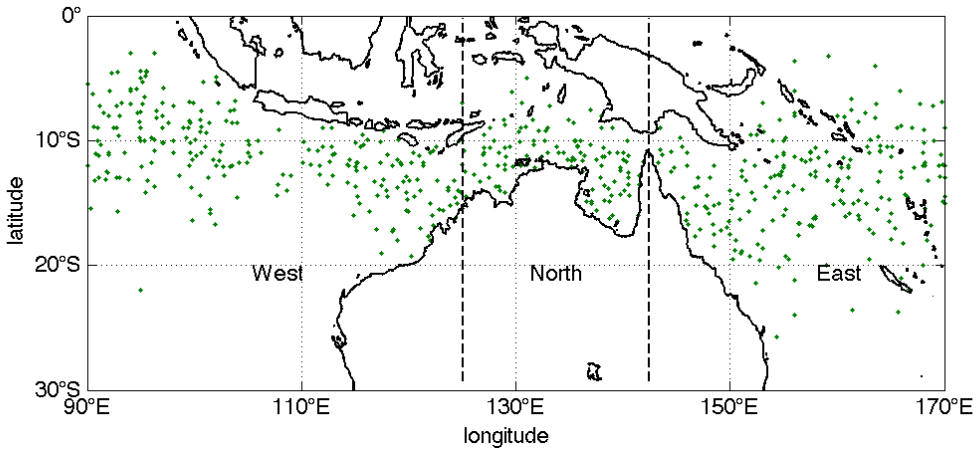


Fig.1 Spatial distribution of the origins of tropical storm systems with winds exceeding 34 knots (17.5 ms^{-1}) in the Australian region from 1968/69-2007/08

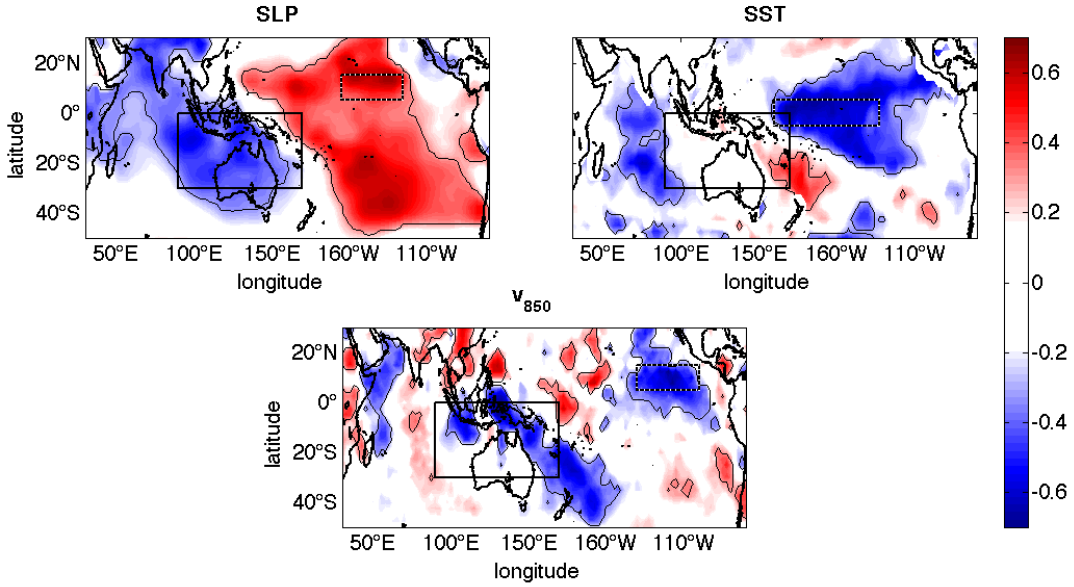


Fig. 2 Spatial maps of correlations between three different June-July-August (JJA)-averaged large-scale climate variables with upcoming Australian region TCG seasonal totals. The bold box indicates the Australian TC region and the dashed box identifies the predictor region with maximum relationship with Australian region seasonal TCG. The thin lines outline the spatial pattern correlations significant at the 95% confidence level, taking account of serial correlation

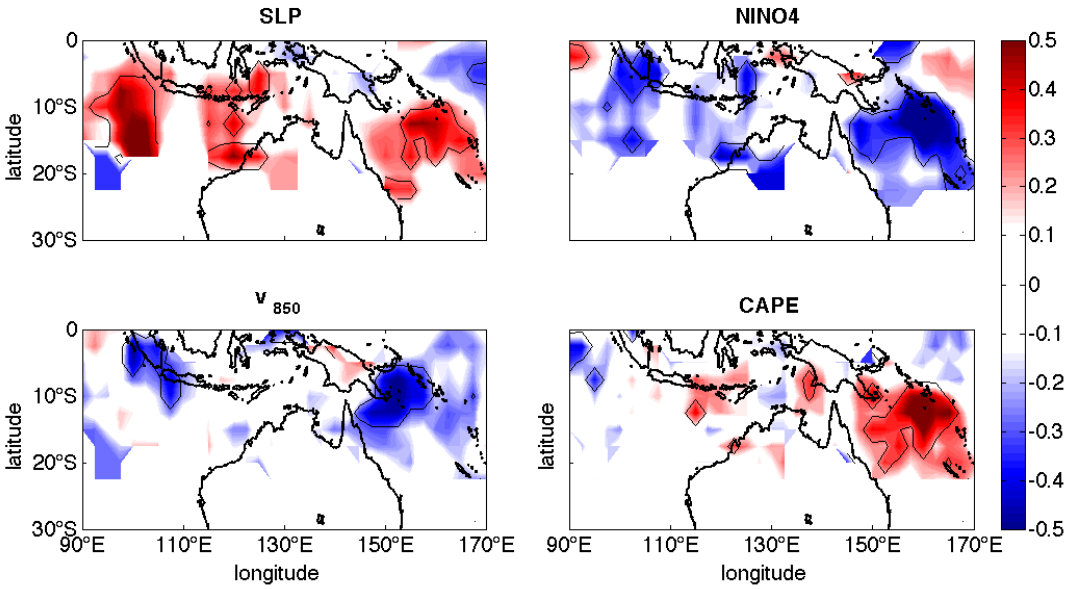


Fig. 3 Spatial maps of correlations between the spatial running grid (binned in nine 2.5° x 2.5° grid-cells) of observed TCG counts across the Australian tropical cyclone region and four separate climate predictor variables. The thin lines outline spatial pattern correlations significant at the 95% confidence level, taking account of serial correlation

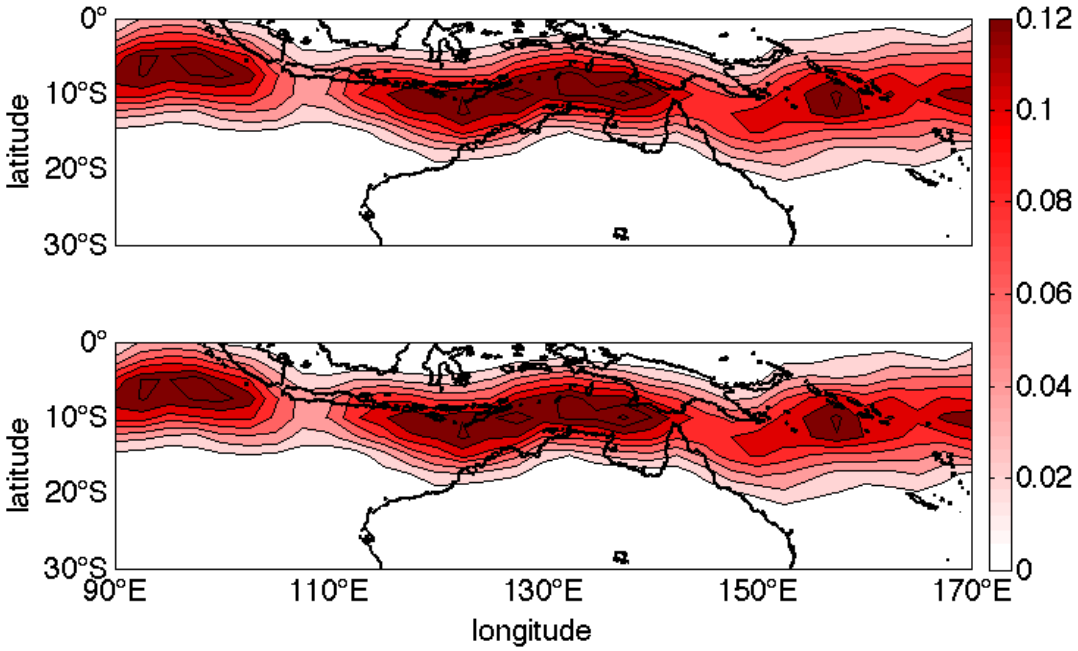


Fig. 4 Spatial distribution of 40-year record-average TCG: (a) observed probabilities; and (b) hindcasted probabilities; across the Australian region from 1968/69-2007/08

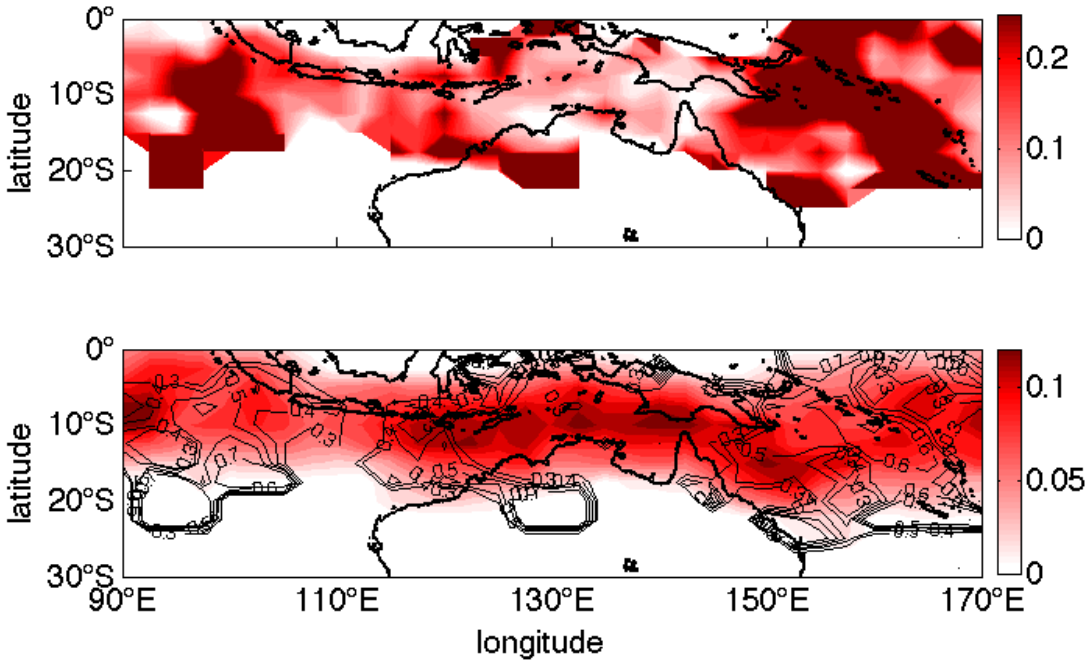


Fig. 5 (upper panel) Spatial distribution of skill scores based on model hindcasts of seasonal TCG event probabilities using the four-predictor SLP+NINO4+v_{850}+CAPE model. (lower panel) Spatial distribution of corresponding RMSEs shaded in red - the overlying contour lines outline the correlation coefficient distribution between the observed and hindcasted TCG event probabilities for the period 1968/69-2007/08

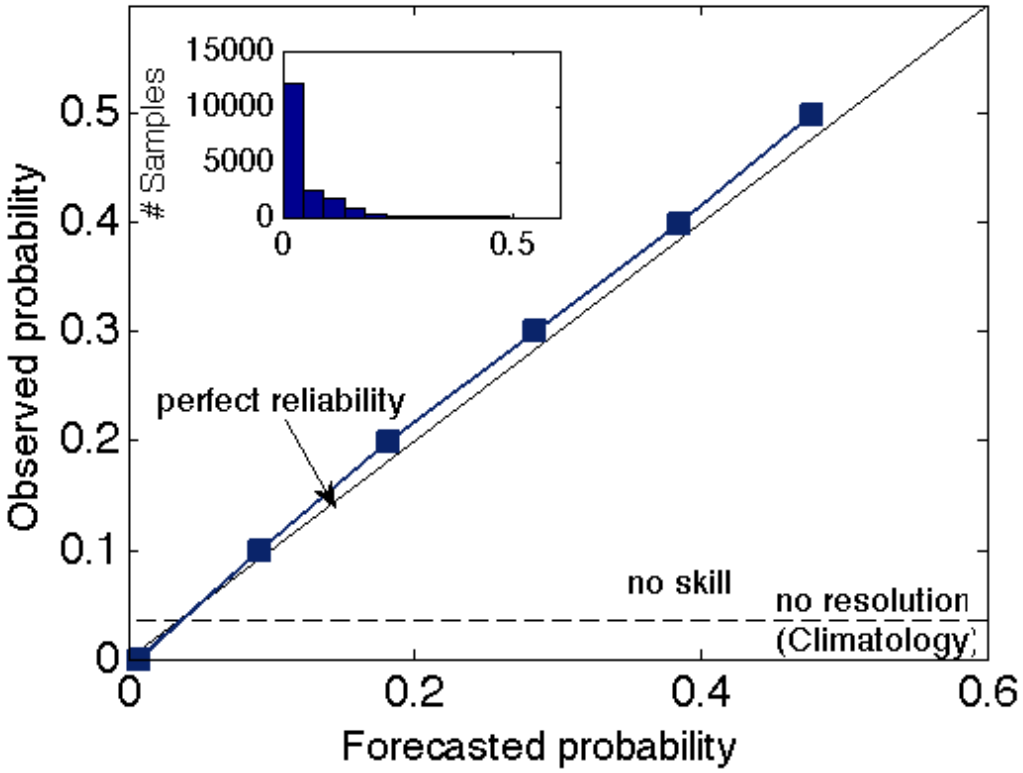


Fig. 6 Attributes diagram of model hindcasts. In the subpanel, the bar diagram summarizes the total distribution of hindcast probabilities, while the large outer diagram compares the observed probabilities with the hindcasted probabilities, within the probability range

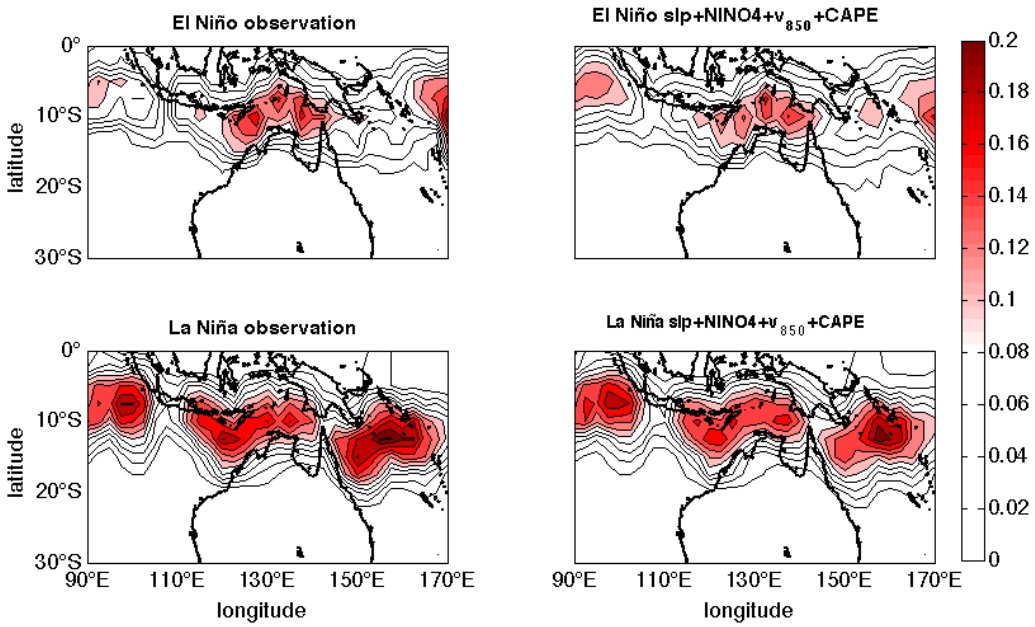


Fig. 7 El Nino and La Nina event composite maps of observed and hindcasted TCG probabilities across the Australian region

5.3 Chapter Summary

This chapter has addressed the thesis aim to develop a seasonal forecast model for the spatial probability of TCG.

The model applies the Bayesian approach on a logistic regression model with a Bernoulli distribution. The model shows substantial cross-validated skill in hindcasting the average spatial distribution and magnitude of the seasonal variations. In particular the strong probability variations due to ENSO events are accurately captured.

III DISCUSSION AND SUMMARY

6 DISCUSSION OF RESULTS

6.1 Objectives

This thesis comprises studies that have attempted to improve seasonal forecasting of Australian region tropical cyclone formation (genesis: TCG). First the contribution from, and forecast skill of, classical El Niño-Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) indices on Australian region annual TCG counts were quantified. Prediction schemes were also identified to build skilful seasonal forecast models of TCG for the Australian region, its subregions (West and East), and their probabilistic distribution. For this purpose, it was important to validate the importance of large-scale climate signals for Australian region TCG, but also to improve the understanding of seasonal variations of the Australian region climatology and how it is affecting the formation of TCs. Skilful seasonal tropical cyclone forecasts are essential for populations in vulnerable regions to provide advanced information for emergency services and insurance industries to be better prepared for the risks associated with the upcoming TC season.

The Australian climate is affected by two tropical large-scale ocean-atmosphere modes operating in the Indo-Pacific region, ENSO and the IOD. A strong link between ENSO and TCG counts in the Australian region is well-established (e.g., Nicholls 1984d, 1992; Basher and Zheng 1995; Sinclair et al. 1997; Kuleshov and de Hoedt 2003; Ramsay et al. 2008; Kuleshov et al. 2009) and therefore ENSO metrics are the basis for all seasonal TC model forecasts for the Australian region. The earlier seasonal forecast models for the Australian region all used SLP-based ENSO indices such as the SOI (e.g., Nicholls 1979b, 1984d, 1985a, 1992; Solow and Nicholls 1990; Drosdowsky and Woodcock 1991; Ready and Woodcock, 1992; McDonnell and Holbrook 2004a,b) with McDonnell and Holbrook (2004a,b) additionally

incorporating spatial information of saturated equivalent potential temperature (EPT) to forecast TCG annual totals. A more recent study (Liu and Chan 2010) included various ENSO indices such as NINO4, the trade wind and outgoing longwave radiation (OLR) index, as well as the IOD index (the Dipole Mode index, DMI) as predictors in a seasonal forecast model. However, the quantification of, or of how strong the DMI improves an ENSO-only forecast, had not been investigated. Goebbert and Leslie (2010), however, have recently discussed the possible use of alternative indices on the basis of correlations between TCG variations in the Northwestern Australian region and climate variables, in a preliminary model approach. The seasonal TC forecast models techniques used for the Australian region differ between multiple linear regression (MLR) models (e.g., Nicholls 1979a, 1984d, 1985a, 1992; Drosdowsky and Woodcock 1991; Ready and Woodcock, 1992; Goebbert and Leslie 2010) and Poisson regression models (Solow and Nicholls 1990; McDonnell and Holbrook 2004a,b), all using the frequentist approach. Liu and Chan (2010) recently applied a project-pursuit regression (PPR) technique.

Despite all these model studies, only three models have been used, at least in an experimental capacity, for operational forecasts of Australian region TCG counts and its distribution (McDonnell and Holbrook 2004a,b; Liu and Chan 2010). A separate model, based on the ENSO indices Southern Oscillation Index (SOI) and NINO3.4, has been used recently by the Australian Bureau of Meteorology (BoM 2011: http://www.bom.gov.au/climate/ahead/tc.shtml). Although the McDonnell and Holbrook (2004a,b) model is the only one that has made serious attempts to capture the spatial distribution of TCGs, a spatial bias has restricted its effective use for operational spatial probabilistic predictions. In recent years, this model has not been used further in experimental forecasts, albeit that its skill in forecasting annual aggregated counts is reasonably good. The Liu and Chan (2010) and BoM models failed in correctly forecasting the only available reference forecast of the previous Australian TC season 2010/11. In summary it appears that all these models have some reasonable skill over their training period. However, the inaccuracy of operational forecasts of TCG for the most recent TC season (2010/11), highlight the need for an improved model.

The goal of this thesis has been to investigate whether there is an important contribution from the IOD on seasonal variations in Australian region or subregional

TCG counts, and whether this adds skill to that provided by classical ENSO metrics. Moreover, the primary objective of this thesis was to develop an advanced statistical seasonal forecast system for TCG counts in the Australian region and subregions, as well as to develop a seasonal forecast model of the spatial distribution of TCG probabilities. For this reason, we have focused on important prediction schemes identified in this thesis and incorporated these in a Bayesian model approach to enable skilful forecasts for the Australian TC season regarding the number and distribution of TCG events.

6.2 Key Findings

To address the key objective of this thesis, the influence of the two tropical large-scale ocean-atmosphere modes, ENSO and IOD, operating in the Indo-Pacific region on Australian region TCG were investigated. Previous attempts to separate out the two signals either concentrated purely on the climatological effects (e.g. Saji et al. 2003, Ashok et al. 2004; Saji et al. 2005; Risbey et al. 2009) or are rather laborious to repeat (Meyers et al. 2007; Ummenhofer et al. 2009). Alternative prediction schemes for seasonal forecasting of Australian region TCG on the basis of various synoptic, dynamic and thermal climate variables, were not investigated before. Instead seasonal forecasting was based on traditional ENSO- or IOD metrics (e.g., Nicholls 1979b 1984d, 1985a, 1992; Solow and Nicholls 1990; Drosdowsky and Woodcock 1991; Ready and Woodcock, 1992; McDonnell and Holbrook 2004a,b; Liu and Chan 2010). Forecast models of seasonal TCG occurrences in the Australian region were thus far based on frequentist statistics applying MLR or Poisson regression schemes, with only Liu and Chan (2010) having most recently used the Monte Carlo based jackknife method for validation of their PPR model. In previous years a couple of seasonal TCG Poisson forecast models have been introduced using Bayesian statistics in various TC basins (e.g., Elsner and Jagger 2004, 2006; Elsner et al. 2008; Chu and Zhao 2007; Chand et al. 2010), however, this approach had not yet been applied to the Australian region. In the past, there have been only a few statistical seasonal forecast models of

the spatial distribution of TCG developed. For the Australian region, McDonnell and Holbrook (2004a,b) applied a Poisson regression model approach on a 5°longitude x 2°latitude grid, but had limited success regarding grid-scale forecasting. More recently, Leroy and Wheeler (2008) developed an intra-seasonal TCG forecast model for the Australian subregions. To correspond to the low numbers of TCG events expected at the shorter time scales, the model is based on the logistic regression forecasting the probability of occurrence.

In Chapter 3 a simple but effective statistical mean to remove ENSO from the IOD was presented, which extracts the ENSO-independent component of the IOD and its effects on seasonal variations of Australian region TCG. There is much debate over the extent to which the Indian Ocean Dipole mode is independent of ENSO (e.g., Saji et al. 1999; Allan et al. 2001; Ashok et al. 2003; Meyers et al. 2007; Risbey et al. 2009). To investigate the dependency of the IOD on ENSO, various studies have explored methods to separate the pure IOD mode from ENSO, or otherwise investigated the effect of the IOD on key climate variables over Australia such as precipitation and temperature (e.g., Ashok et al. 2003; Saji et al. 2005; Meyers et al. 2007; Risbey et al. 2009; Ummenhofer et al. 2009). Even though these studies have been able to demonstrate some degree of independence of the IOD from ENSO regarding the timing of Australian region precipitation rate variability, they had difficulties in removing the more complete lag/lead effects from ENSO. Also, the more advanced methods (Meyers et al. 2007; Ummenhofer et al. 2009) are quite laborious to repeat. We applied a lagged regression of NINO3.4 on DMI and removed the NINO3.4 contributions within leads and lags of up to 8 months from the 40-year (1969-2008) record. From the original DMI time series, 24% of the variations were removed in that process. We regressed the resulting DMI_{NOENSO} time series on a spatial SST field and were able to confirm the successful removal of ENSO from the spatial SST pattern, when isolating the IOD-typical SST anomalies during the peak-season in austral spring (Saji et al. 1999). Subsequently, we analysed the forecast potential from NINO3.4 and NINO4, as well as from the original DMI and the ENSO-independent DMI_{NOENSO} for seasonal variations of Australian region and subregional TCG. No additional skill was found from the DMI or DMI_{NOENSO} when applying a MLR model to forecast seasonal variations of Australian region TCG.

When applying a leave-one out cross-validation, the skill score (SS) over climatology only improves from 44% for the NINO4-only model to 45.5% for the NINO4+DMI_{NOENSO} model. For the Eastern Australian subregion, some additional skill was gained from the NINO4+DMI_{NOENSO} model (SS = 36.9%) over the NINO4-only model (SS = 31.4%) when forecasting seasonal TCG counts, while TCG variations in the western Australian region were best forecasted by the NINO3.4-only predictor model.

In Chapter 4, a Bayesian seasonal TCG forecast model for the Australian region was developed. Previous models were all based on well-known ENSO indices derived from anomalies of seas surface temperature (SST), sea level pressure (SLP), SLP gradients, the trade winds or OLR (e.g., Nicholls 1984d; Solow and Nicholls 1990; McDonnell and Holbrook 2004a,b; Liu and Chan 2010; BoM 2011: http://www.bom.gov.au/climate/ahead/tc.shtml). Our belief is that the use of indices based purely on ENSO dynamics is limited. To study the effects of ENSO, or the large-scale climatology in general, on Australia's climate more effectively, it is beneficial to look into possible relationships between various patterns of climate variables and how they affect Australia's climate, e.g., TCs, rainfall patterns. Here various prediction schemes were identified based on persistent spatial correlation patterns between Australian region annual TCG totals and climate variables of convective available potential energy (CAPE), EPT, geopotential height at 500 hPa (GPH), zonal and meridional winds at 850 hPa (u_{850}, v_{850}), environmental vertical wind shear between 850 hPa and 200 hPa, and low-level relative vorticity (RV). The derived indices were then tested in a Bayesian MCMC model applied with a Poisson regression, to evaluate potential forecast accuracy. A step-by-step predictor selection ensured that the most skilful model was chosen. The final three-predictor model is based on June-July-August (JJA) averaged central subtropical South Pacific CAPE, May-June-July (MJJ) averaged tropical Northeast Pacific meridional winds at 850 hPa (v_{850}), and July-August-September (JAS) averaged central subtropical South Pacific GPH. To evaluate our model results, a leave-one out cross-validation (LOOCV) method was applied. For the LOOCV TCG hindcasts over the 40-year record from 1968/89-2007/08, the corresponding probabilistic RMSE = 5.20 and the correlation coefficient between observed annual TCG totals and LOOCV model

hindcasts is r = 0.73. This is a substantial improvement in cross-validated hindcast skill of at least 21.5% over previous models, with correlations between cross-validated hindcasts and observations of TC counts ranging from r = 0.44 to r = 0.60 (e.g., Solow and Nicholls 1990; Nicholls 1992; McDonnell and Holbrook 2004b). Using *four*-fold cross-validation, model hindcast skill is robust with 85% of the observed seasonal (annual) TCG totals hindcast within the model standard deviations. Also, the seasonal TCG totals during ENSO events are typically well captured. Comparing the LOOCV results of the Bayesian model for final predictor selection with the LOOCV results of a Poisson regression frequentist model, the SS shows an improvement over the climatological average - ranging from SS = 42.8% for the frequentist approach to 51.5% for the Bayesian model. These results are similar to improvements shown by Liu and Chan (2010). However, their 51% improvement over climatology is based on a slightly smaller region (90°-160°E), and the PPR technique and the jackknife validation method, making it difficult to directly compare the results. The model presented here is also shown to be valuable in hindcasting seasonal TCG totals in the Eastern Australian subregion (r = 0.73, SS = 49.3%) and provides some skill for the Western Australian region (r = 42, SS = 10.7%), while it not useful for the Northern region. Applying the model to only two subregions, subdivided at 135°E, the model shows remarkable skill in the southwest Pacific (r = 0.79, SS = 57.8%), while forecasts show only little improvement over climatology in the eastern Indian Ocean region (r = 0.38, SS = 9.0%). For this reason, we looked for different predictors for the Bayesian model used in the eastern Indian Ocean (Australian Western) subregion (90°-135°E). We found that a two-predictor model based on JJA-averaged tropical central Pacific SLP and the above-mentioned MJJ-averaged v_{850} index was the most useful predictor combination. The corresponding correlation coefficient between observed annual TCG totals and LOOCV model hindcasts is r = 0.57, with SS = 29.6%. The subregional results presented in Chapter 4 differ quite strongly from the ones presented by Liu and Chan (2010). They found improvements of 37% and 39% for the Western and Eastern Australian subregions, respectively, when using the PPR technique with the jackknife validation method.

In Chapter 5, a model for the spatial distribution of TCG probabilities in the Australian region on a 2.5°x2.5° grid is presented. The challenge in forecasting the

spatial distribution of TCG probabilities, is to find skilful predictors for all grid cells and to solve the issue of sparse data points. There has been only one previous study that has made serious attempts to spatially forecast TCG occurrences for the Australian region - that is, by McDonnell and Holbrook 2004(a,b). They applied a Poisson regression model on a 5°longitude x 2°latitude grid based on the SOI and spatially-dependent grid-point EPT values as predictors. However, despite achieving a 22% improvement of RMSE over climatology at the subregional scales, they had difficulties in providing unbiased forecast estimates of the grid-scale *spatial distribution* of TCs formed. The model presented in this thesis has little in common with the McDonnell and Holbrook (2004a,b) formulation. First, to better take account of data sparsity, a spatial running mean of three horizontal and vertical grid boxes was applied. This enables us to regress probabilistically on the predictands, which are estimated at the centred box of the spatial mean. To model the probabilities, logistic regression was applied using a Bayesian MCMC approach to each of the spatial means. The most skilful model predictor combination uses the JJA-averaged tropical central Pacific SLP, JAS-averaged NINO4, MJJ-averaged v_{850} and the spatial information of AMJ-averaged CAPE. The model shows an average skill-score improvement over climatology of SS = 25.2%. The average distribution, as well as the changing probabilities for occurrences during ENSO events, is well-captured while there are forecast difficulties for the eastern Indian Ocean region between 105° and 120°E. An attribution diagram (Chapter 5.2; Fig. 6) identifies relatively high skill for all probabilities. However, there is a trend towards slight underestimation for the higher probability occurrences that are observed.

6.3 Operational Model Forecasts

The robust outcome of this thesis is not only a relatively skilful and improved TCG statistical forecast model system for the annual number of TCG events in the Australian region and its subregions, but also a new probabilistic model of the likely location of origin of tropical cyclones in the upcoming season. In this section, these forecast models are tested in an experimental ‘operational’ forecast mode for the

Australian region tropical cyclone seasons 2008/09, 2009/10 and 2010/11 - seasons that are outside of our model training data set. Additionally, results from the most recent Australian TC season of 2010/11 are discussed in comparison to the two available operational seasonal forecast models used respectively by the Australian BoM (2011: http://www.bom.gov.au/climate/ahead/tc.shtml) and the City University of Hong Kong (Liu and Chan 2011: http://weather.cityu.edu.hk/tc_forecast/2010_forecast_NOV.htm) - forecasts that were made and released prior to the onset of the 2010/11 Australian tropical cyclone season. While the BoM model uses the SOI and NINO3.4 as predictors, the latter model is based on the Liu and Chan (2010) formulation. Observations are taken from the UNISYS (http://weather.unisys.com/hurricane/) database, which include TC Best track data provided by the Joint Typhoon Warning Center. Note that the observed TCG locations here are locations where the system first reached TC strength (maximum sustained winds ≥ 34 kts), while the model was trained on storm origins. This may result in a slight shift of the TCG numbers, but certainly results in a southward shift of the observed TCG locations against the model forecasts.

6.3.1 TCG Forecasting of the 2008/09 Season for the Australian Region

a. Climatic conditions

The climatic conditions during austral winter (JJA) 2008, and therefore prior to the onset of the Australian TC season 2008/09, are classified as an ENSO neutral year based on classifications from the National Oceanic and Atmospheric Administration (NOAA) using NINO3.4 anomalies (http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml). Figure 6.1 shows the climatologies of all climate variables used to derive our model predictors. Interestingly, the SST pattern shown in the lower panel reveals almost negative El Niño Modoki conditions in the Pacific, with weak positive SLP anomalies across most of the tropical and subtropical Pacific. There are strong SLP anomalies in the Southern mid-latitudes which are also evident in the geopotential height field at 500 hPa (upper panel), indicating stronger pressure systems as a result of increased blocking of the mid-latitude air masses towards lower latitudes. The central Indian Ocean was warmer than normal with weak negative SLP anomalies in

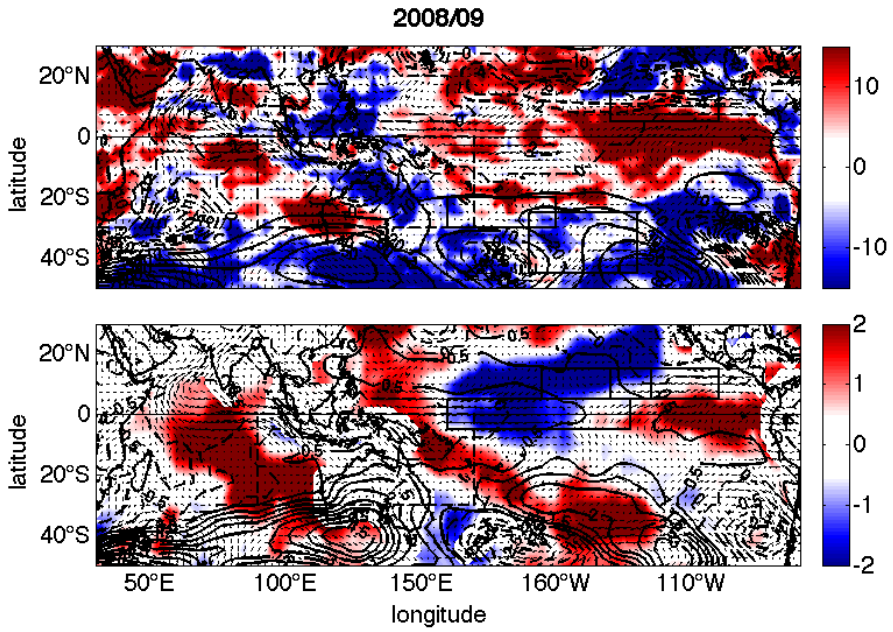


Figure 6.1 upper panel: Map of anomalies during June-July-August 2008. Wind vectors describe the wind flow anomalies at 850 hPa. Shaded areas represent changes of CAPE (m^2 s^{-2}), contour lines show positive anomalies and dashed contour lines negative anomalies of geopotential height anomalies at 500 hPa (m). Also indicated are the locations of the predictor indices CAPE, v_{850} and GPH and the Australian TC region. lower panel: As above, but shaded areas represent changes of SST (°C), contour lines show positive anomalies and dashed contour lines negative anomalies of SLP (hPa). Also indicated are the locations of the predictor indices NINO4, v_{850} and SLP and the Australian TC region.

the Arabian Sea and in the subtropical western Indian Ocean. There are clock-wise surface wind anomalies in the Arabian Sea around the negative SLP anomalies. Also, a reduced trade wind component can be seen in the tropical east Pacific corresponding to the warm SST anomalies and the reduced SLP gradient between East and West Pacific. The CAPE anomaly pattern (upper panel) shows enhanced convection in the tropical central to East Pacific and in the tropical central and subtropical eastern Indian Ocean. The mid-latitudes show negative anomalies throughout - a result of the lower intensity of the atmospheric Rossby wave train.

For the predictors used in our models, the CAPE index is almost neutral, v_{850} is slightly positive indicating a reduced inflow into the East Pacific tropics, and GPH is positive resulting in the increased blocking of mid-latitudinal air-masses. The SLP index is neutral and NINO4 is weak negative. Spatially, in the TC formation regions, there are very neutral conditions of CAPE, with a negative anomaly over North

Australia and weak positive anomalies at the outer East and West boundaries of the Australian region. Generally, all predictors are close to neutral conditions with only GPH suggesting more than average activity. In this case, therefore, the 2008/09 forecasts are expected to be close to the climatological average.

b. Model Forecasts

Figure 6.2 shows the model forecasts for the Australian TC season of 2008/09 as obtained by the models presented in Chapters 4.2, 4.4 and 5.2. The forecasted 2008/09 TCG total is 13 for the Australian region, with 7 TCG events forecast for the Western region and 5 for the Eastern region. The spatial pattern shows TCG probability maxima in the eastern Indian Ocean around 10°-15°S, 95°E and 5°-15°S, 115°-135°E. Table 6.1 lists all observed TCG events and locations for the Australian TC season 2008/09. Comparing the model forecast results with observations (Fig. 6.2, Table 6.1), we note that 12 TCG events were observed in the Australian region with 5 and 7 in the Western and Eastern regions, respectively, are all are within the model

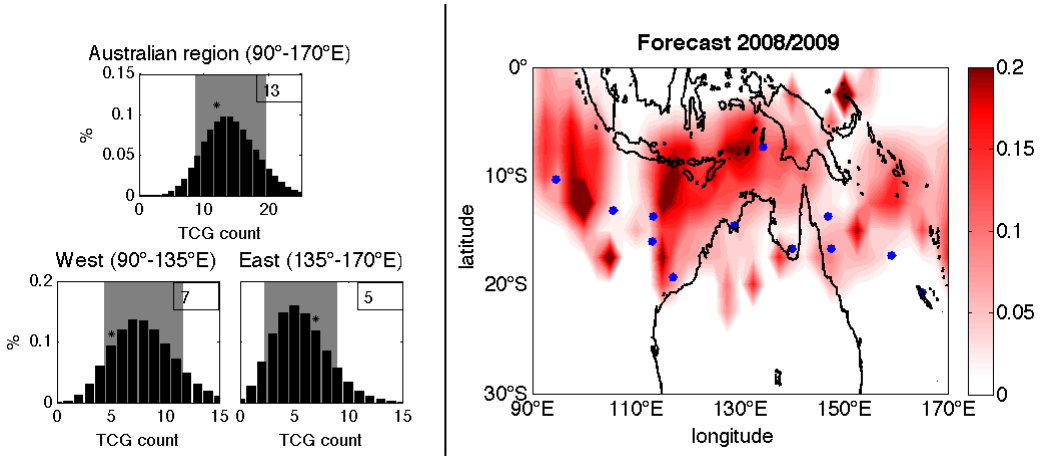


Figure 6.2 left side: Probability distributions of the annual total number of TCG occurrences forecasted for the season 2008/09 for the Australian TC region and its subregions West and East. Asterisks indicate the observed number of TCGs in that region, while the model standard deviations are indicated by the shading. Right side: Spatial (2.5°x2.5° resolution) probability of TCG forecasted for the season 2008/09 for the Australian TC region. Blue dots indicate observed location of TCs first noted 34 kts.

Table 6.1 Tropical Cyclogenesis points (taken from UNISYS 2011) observed in the Australian region (0-30°S, 90°-170°E) during the Australian TC season 2008/2009 (November to April). Here noted location is, where the storm system first reached 34 kts. Maximum sustained winds (MSW) and the Category (Cat) describe the maximum observed Intensity of the TCs.

2008/09											
South Indian Ocean						**South Pacific**					
Name	**Date**	**lat**	**long**	**MSW**	**Cat**	**Name**	**Date**	**lat**	**long**	**MSW**	**Cat**
Anika	18/11/08	-10.3	94.5	50	2	Charlotte	11/1/09	-16.7	140.0	35	1
Billy	18/12/08	-14.5	128.9	105	4	Ellie	31/1/09	-16.7	147.5	40	1
Dominic	25/1/09	-19.3	117	50	2	Innis	17/2/09	-20.7	165.1	35	1
Freddy	6/2/09	-16.0	113.1	55	2	Hamish	5/3/09	-13.7	146.9	130	5
Gabrielle	2/3/09	-13.2	105.6	35	1	Jasper	24/3/09	-17.3	159.2	45	1
Ilsa	18/3/09	-13.7	113.3	100	4						
Kirrily	26/4/09	-7.3	133.0	40	1						

standard deviations. The genesis locations for the observed events have the expected southward shift, but the general distribution is well captured.

We conclude that the TCG total for the 2008/09 season was successfully forecasted for the Australian region and its two subregions. The spatial distribution of TCG probabilities is, besides the southward shift of observed TCG locations due to the different definition, also remarkably well captured.

6.3.2 TCG Forecasting of the 2009/10 Season for the Australian Region

a. Climatic conditions

The climatic conditions during the astral winter (JJA) of 2009, prior to the onset of the Australian TC season of 2009/10, were evolving towards an El Niño year based on classifications from NOAA using NINO3.4 anomalies. Figure 6.3 shows the climatologies of all climate variables used to derive our model predictors. In the SST pattern (lower panel), the El Niño conditions are evident with warming throughout the Indo-Pacific tropics. However, tropical SLP anomalies are weak, with strong negative anomalies of SLP and geopotential height at 500 hPa in the southern mid-latitudes, and decreased blocking of the mid-latitudinal air. In the Indian Ocean, a weak SLP dipole is evident with positive anomalies in the subtropical eastern Indian Ocean and negative anomalies in the tropical western Indian Ocean. Surface winds appear as anomalies from the positive SLP pole to the negative one. Also a weakening of the

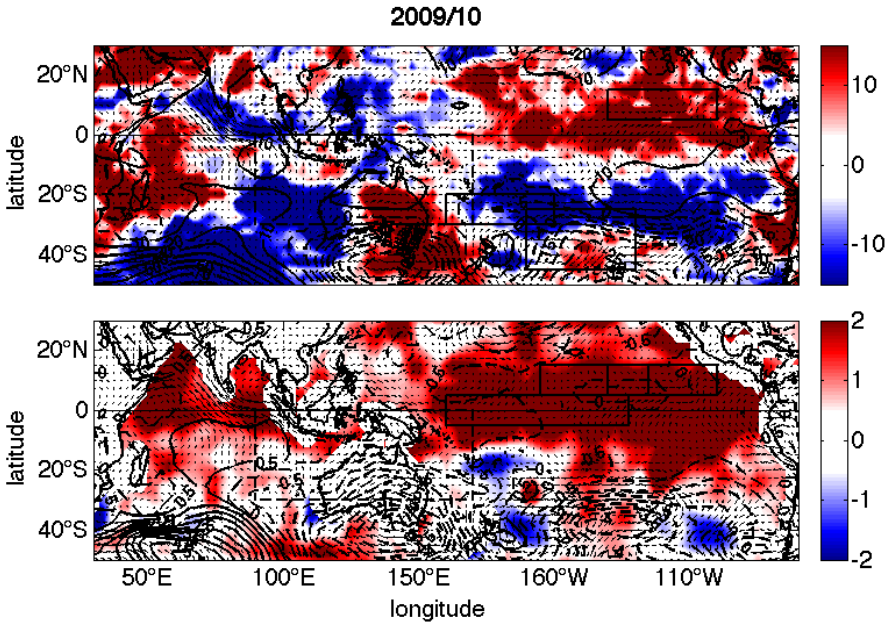


Figure 6.3 As Figure 6.1, but for 2009.

trade winds is observed in both the tropical Pacific and the Indian Ocean. The CAPE anomaly pattern (upper panel) shows enhanced convection in the tropical central to East Pacific and a mirrored negative anomaly in the subtropical South Pacific. There is a CAPE dipole over the subtropical to mid-latitude Australian continent with positive anomalies in the east and negative in the Western Australian regions extending into the Southern Ocean. Positive anomalies of CAPE can also be observed in the western Indian Ocean.

For the predictors used in our models, the CAPE index is weak negative, v_{850} shows positive anomalies indicating a reduced inflow into the East Pacific tropics, and GPH is negative resulting in the decreased blocking of mid-latitudinal air-masses. The SLP index is neutral, while NINO4 is strong positive. Spatially, in the TC formation regions, there are very neutral conditions of CAPE, with the above-mentioned dipole in the subtropics.

b. Model Forecasts

Figure 6.4 shows the model forecasts for the Australian TC season of 2009/10 as obtained by the models presented in Chapters 4.2, 4.4 and 5.2. The TCG count forecasts are 11 for the entire Australian region, with 6 TCG events forecast for the Western region and 4 in the Eastern region. The spatial pattern shows probability

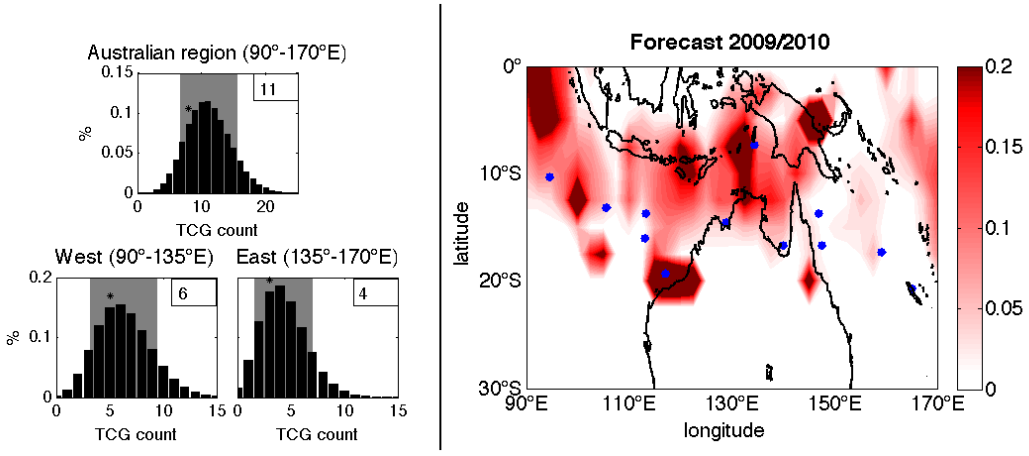


Figure 6.4 As Figure 6.2, but for season 2009/10.

maxima for TCG mostly in the eastern Indian Ocean at the northwestern corner of the Australian region, around 10°-15°S, 95°E, at the western Australian coastline around 20°S, 115°S, and in the northern Australian regions. Table 6.2 lists all observed TCG events and locations for the Australian TC season of 2009/10. Comparing the model forecast results with observations (Fig. 6.4, Table 6.2), we note that there were 8 TCG events observed in the Australian region with 5 and 3 in the Western and Eastern regions, respectively, all being within the model standard deviations. The locations of the observed TCG events have the expected southward shift. However, the two Pacific TCs formed in regions with very low TCG probability forecasted. In contrast the Indian Ocean TCG events appear to have their origins in the proximity of probability maxima.

TABLE 6.2 As Table 6.1, but for season 2009/10. Storm in brackets developed outside of the Australian region, but in its close proximity (<1° longitude).

2009/10											
South Indian Ocean						**South Pacific**					
Name	**Date**	**lat**	**long**	**MSW**	**Cat**	**Name**	**Date**	**lat**	**long**	**MSW**	**Cat**
Laurence	13/12/09	-12.7	128.0	115	4	Olga	23/1/10	-16.6	150.2	50	2
Magda	20/1/10	-13.1	122.5	60	2	Ului	11/3/10	-14.7	166.4	140	5
(Imani	22/3/10	-11.5	89.2	70	3)	Paul	27/3/10	-12.9	136.6	60	2
Robyn	2/4/10	-11.3	92.5	60	2						
Sean	22/4/10	-11.8	115.7	55	2						

We conclude that the TCG totals for the TC season of 2009/10 were successfully forecasted for the Australian region and its two subregions. The spatial distribution of TCs formed is, besides the southward shift of the observed TCG location due to the different definitions, well captured in the eastern Indian Ocean, but had difficulties in the Pacific regions.

6.3.3 TCG Forecasting of the 2010/11 Season for the Australian Region

a. Climatic conditions

The climatic conditions during the austral winter (JJA) of 2010, prior to the onset of the Australian TC season of 2010/11, were evolving towards a moderate to strong La Niña year based on classifications from NOAA using NINO3.4 anomalies (http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml). Figure 6.5 shows the climatologies of all climate variables used to derive our model predictors. In the SST pattern (lower panel), the strong La Niña conditions are evident with warming in the subtropical Pacific, the West Pacific, and all of the Indian Ocean. However, tropical SLP anomalies are weak with negative anomalies over central and northern Australia and slight positive anomalies in the eastern

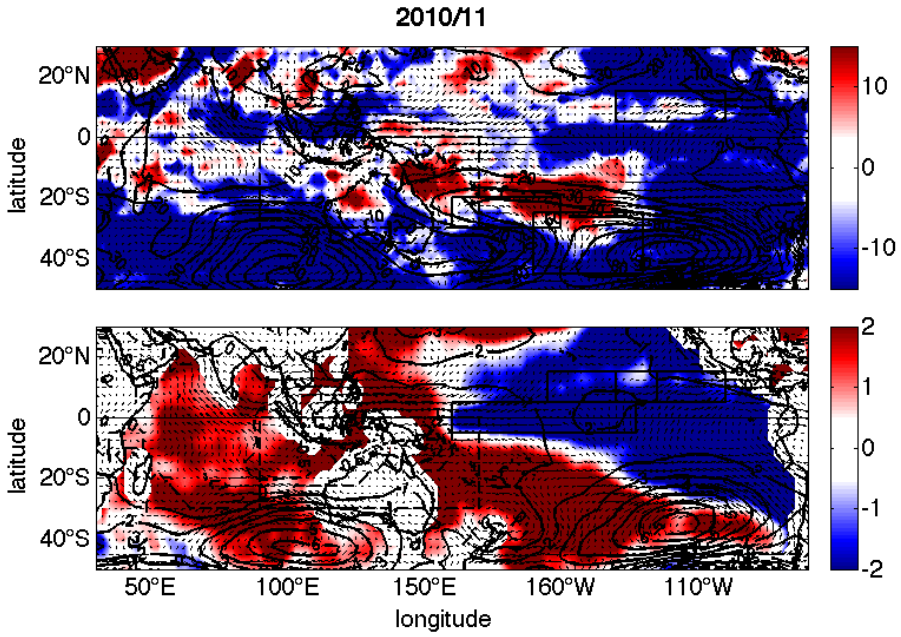


Figure 6.5 As Figure 6.1, but for 2010.

Pacific. Strong positive anomalies of SLP and geopotential height at 500 hPa are observed in the southern mid-latitudes, with reduced blocking of the mid-latitude air. Surface winds indicate a weakening of the trade winds in the eastern Pacific and Indian Ocean and a strengthening in the western Pacific. There are strong wind anomalies around the stronger than normal semi-stationary high-pressure cells in the South Indian Ocean and Pacific increasing blocking of the mid-latitude air masses. The CAPE anomaly pattern (upper panel) is strongly negative over the Southern Ocean and East Pacific. There is only a small line of enhanced instability from the subtropical southwest Pacific into the central Pacific. Anomalies in the Indian Ocean are weak with a tendency towards a more stable mid-upper troposphere.

For the predictors used in our models, the CAPE index is weak positive, v_{850} shows positive anomalies indicating a reduced inflow into the East Pacific tropics, and GPH is strongly positive resulting in the enhanced blocking of mid-latitude air-masses. The SLP index is slightly positive, while NINO4 is strong negative. Spatially, in the TC formation regions, there are neutral conditions of CAPE, with a positive anomaly east of northern Queensland.

b. Model Forecasts

Figure 6.6 shows the model forecasts for the Australian TC season of 2010/11 as obtained by the models presented in Chapters 4.2, 4.4 and 5.2. The TCG count

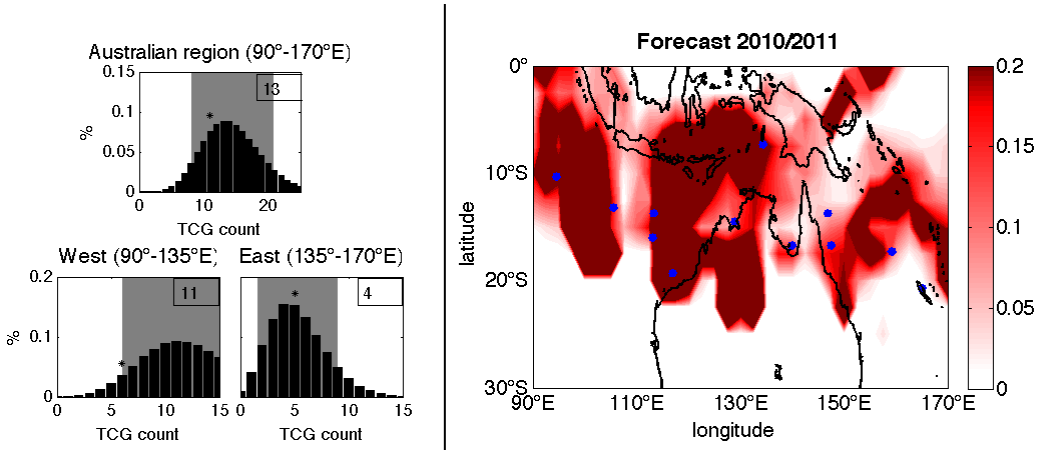


Figure 6.6 As Figure 6.2, but for 2010/11.

forecasts are 13 for the entire Australian region, with 11 TCG events forecast for the Western region and 4 for the Eastern region. The spatial forecast pattern shows high TCG probabilities east and west of 110°E in the Indian Ocean. In the Pacific, there is a prominent probability maximum between 10° and 15°S slightly off the coast in North Queensland. For the most recent season of 2010/11, two model forecasts were provided for each of the expected number of TCG counts in the Australian region and its subregions. Both forecasts predicted above normal TCG counts due to the strong La Niña conditions expected. The BoM released a forecast of 20-22 TCG counts for the Australian region (90°-160°E) in the 2010/11 season, with 11-12 for the Western region (90°-125°E), 5 for the Northern region (125°-142.5°E), and 6-7 for the Eastern region (142.5°-160°E). Liu and Chan (2011) forecasted 19 TCG events for the Australian region (90°-160°E), with 14 in the Western region (90°-135°E) and 7 in the Eastern region (135°-160°E).

Table 6.3 lists all observed TCG events and their locations for the Australian TC season of 2010/11. Comparing the model forecast results with observations (Fig. 6.6, Table 6.3) we note 11 TCG events were observed in the Australian region, with 6 and 5 in the Western and Eastern regions, respectively. While the TCG count observations in the Australian and Eastern regions are within the model standard deviations, the 6 TCG events observed in the Western region are just outside the model uncertainty. Nevertheless, the seasonal forecasts provided here are substantially more accurate than those released by the BoM and using the Liu and Chan (2011) schemes, which both strongly overestimated the activity of the Australian TC season for 2010/11. The observed TCG event total for the Australian

TABLE 6.3 As Table 6.1, but for season 2010/11. Storm in brackets developed outside of the Australian region, but in its close proximity (<1° longitude).

2010/11

South Indian Ocean						**South Pacific**					
Name	**Date**	**lat**	**long**	**MSW**	**Cat**	**Name**	**Date**	**lat**	**long**	**MSW**	**Cat**
Vince	12/1/11	-15.1	108.3	40	1	Tasha	24/12/10	-17.1	146.3	35	1
Bianca	25/1/11	-17.4	121.8	115	4	Vania	11/1/11	-19.0	168.9	35	1
Carlos	15/2/11	-12.3	130.7	65	2	Zelia	14/1/11	-12.2	150.9	90	4
Dianne	16/2/11	-19.1	111.2	85	3	Anthony	23/1/11	-15.5	148.7	45	1
Twenty	2/4/11	-13.5	127.6	35	1	(Yasi	30/1/11	-13.2	170.5	135	5)
Errol	15/4/11	-13.1	126.3	55	2	Atu	18/2/11	-15.4	169.4	115	4

TABLE 6.4 Seasonal forecasts of tropical cyclone counts for the Australian region and individual subregions from the three models from the BoM (2011), Liu and Chan (2011) and as obtained from the models presented in this thesis. In brackets and italic are the corresponding observed tropical cyclone counts for each region (taken from UNISYS 2011).

TC season 2010/11

	BoM 2011 Forecast *(Obs.)*	**Liu & Chan 2011 Forecast *(Obs.)***	**Werner & Holbrook 2011 Forecast *(Obs.)***
Australia region	20-22 *(9)*	19 *(9)*	13 *(11)*
Western region	11-12 *(3)*	14 *(6)*	11 *(6)*
Northern region	5 *(3)*	xxx	xxx
Eastern region	6-7 *(3)*	7 *(3)*	4 *(5)*

region (90°-160°E) was 9, with 3 in each of the BoM subregions, and 6 and 3 in the Liu and Chan (2011) subregions, respectively. The locations of observed TCG events are remarkably well captured using our model. It is also interesting to note that even though TC Yasi developed outside of the Australian region (13.2°S, 170.5°E), its TCG location was observed around the northeastern side of the Pacific probability maximum for TCG. In summary, both operational models, BoM (2011) and Liu and Chan (2011), forecasted a much more active TC season by comparison with the models presented here in this thesis – the observed and forecasted numbers of each of the models is also shown in Table 6.4.

We conclude that the TCG totals in the 2010/11 season was successfully forecasted for the Australian region and the Eastern subregion. Even though the TCG count forecasted for the Western Australian subregion was overestimated, it was still far closer to the observed number of TCs in comparison to other operational seasonal TC forecast models. The two discussed operational models also failed to forecast TCG counts in the Australian region and all other subregions. However, the 2010/11 Australian TC season was particularly difficult to forecast. Despite the presence of extremely strong La Niña conditions, suggesting very high TC activity, only average TC activity, in terms of total counts (not intensity!), was observed. The spatial distribution of cyclone formation was also well forecasted.

6.4 Future Work

An important outcome of this thesis is the identification of skilful prediction schemes and the development of the Bayesian forecast model. The results of Chapter 5 show that the Bayesian Poisson model introduced in Chapter 4 can be adapted to statistical forecasting using logistic regression. Compared to previously introduced prediction schemes the spatial probability of seasonal TCG is successfully forecasted.

Future efforts are under way to adapt the presented model approach to forecast the seasonal TC hazard to different categories for the Australian region. To build the model on the most useful predictors, the best predictor selection for the spatial TC hazard independently from the intensity was identified using the Bayesian Poisson regression model. Then, to forecast the probability of the spatial hazard of different categories the selected predictor combination is applied with a logistic regression to each of the intensity subcategories. Preliminary results suggest the best predictor selection is a three-predictor model using previously identified indices of low-level relative vorticity, environmental wind shear and SLP (see in Chapter 4.2, 4.4) reaching a skill score of SS = 58% for the total spatial TC hazard in Australia for the regressed data between 1968/69-2007/08.

Another interesting future potential is to further investigate the prediction schemes identified here. It is of interest to understand how the areas of the indices are modulated by ENSO, but also if and how they interact with each other. Possible future efforts could then regard the predictors not purely independently as in this study, but use the information of the dependencies between the model predictors in the model approach.

6.5 Summary

The aim of this thesis was to improve understanding of how climatic signals and variables affect TCG occurrences observed in the Australian region and to develop, on this basis, skilful statistical seasonal forecast models of TCG counts and probability distribution. Of special interest was to quantify the role of interannual large-scale climate signals such as ENSO and IOD in informing TCG count forecasts. The thesis questions and analyses are addressed using a variety of regression methods, climate analysis tools and Bayesian statistics for the seasonal forecast models. The main findings are as follows:

1. A simple, but effective method has been developed to separate the IOD from ENSO. This thesis demonstrates that it is possible to isolate a statistically significant ENSO-independent IOD that is potentially useful for future climate studies, and that has been used here to investigate the role of IOD on TCG in the Australian region.

2. ENSO is known to be the most important predictor of seasonal variations in TCG for the Australian region. This thesis demonstrates, nevertheless, that there is also reasonable individual forecast skill afforded by the influence of the IOD. In combination with common ENSO metrics, however, the IOD does not add significant forecast skill of seasonal TCG counts for either the Australian region or Western subregion. Marginal improvements were found for hindcasted TCG counts in the Eastern subregion.

3. TCG counts in the Australian region and subregions can be successfully forecasted with a Bayesian Poisson regression model using a Markov chain Monte Carlo method applied with a multivariate slice sampler.

4. Next to NINO4, four climate indices resulting from persistent prediction schemes of CAPE, v_{850}, GPH and SLP were found to represent skilful predictors of TCG for the Australian region and its subregions.

5. A logistic regression approach applied in a Bayesian seasonal forecast model was found to be successful in generating spatial probabilities of TCG for the upcoming season. The most skilful model is based on SLP, NINO4 and v_{850} indices, combined with spatial information from CAPE.

6. Independent forecasts using the three introduced models and comparisons with current operational models demonstrate the relatively high skill of the models presented in this thesis.

These results demonstrate that ENSO-related derived indices, on the basis of persistent correlation patterns between climate variables and seasonal variations of TCG, add important skill to the more-standard ENSO metrics as predictors of Australian region seasonal TCG count and spatial probability. As such, we believe that a more comprehensive understanding of the effects of ENSO on climate variables can help to improve predictions of seasonal TCG variability in the Australian region. Further to this, we believe that with some refinement, the proposed Bayesian framework for seasonal TCG forecasting could be used to forecast the spatial seasonal TC hazard in the Australian region.

REFERENCES

Allan R., 1988: El Niño Southern Oscillation influences in the Australasian region. *Progress in Phys. Geogr.,* **12**, 313 – 348.

Allan R. J., D. Chambers, W. Dosdrowsky, H. H. Hendon, M. Latif, N. Nicholls, I. Smith, R. C. Stone, and Y. Tourre, 2001: Is there an Indian Ocean dipole, and is it independent of the El Niño-Southern Oscillation? *CLIVAR Exchanges,* **6,** 18-22.

Ashok K., Z. Guan, and T. Yamagata, 2003: Influence of the Indian Ocean Dipole on the Australian winter rainfall. *Geophys. Res. Lett.,* **30,** 1821, doi:10.1029/2003GL017926.

Basher R. E., and X. Zheng, 1995: Tropical cyclones in the southwest Pacific: Spatial patterns and relationships to Southern Oscillation and sea surface temperature. *J. Climate,* **8,** 1249-1260.

Bessafi M., A. Lasserre-Bigorry, C. J. Neumann, F. Pignolet-Tardan, D. Payet, and M. Lee-Ching-Ken, 2002: Statistical prediction of tropical cyclone motion: An analog-CLIPER approach. *Wea. Forecasting,* **17,** 821–831.

Bishop C. M., 2006: *Pattern Recognition and Machine Learning*. Springer, 738pp.

BoM, cited 2011: 2010/11 Australian Tropical Cyclone season outlook. [Available online at http://www.bom.gov.au/climate/ahead/tc.shtml.]

Broadbridge L. W., and B. N. Hanstrum, 1998: The relationship between tropical cyclones near Western Australia and the Southern Oscillation Index. *Austr. Meteorol. Mag.*, **47**(3), 183-89.

Bjerknes J., 1969: Atmospheric teleconnections from the equatorial Pacific. *Mon. Wea. Rev.,* **97,** 163-172.

Camargo S. J., A. G. Barnston, P. J. Klotzbach, and C. W. Landsea, 2007: Seasonal tropical cyclone forecasts. *WMO Bull.,* **56,** 297–309.

Cane M. A., Zebiak, *S.* E., and Dolan, S.C., 1986: Experimental forecasts of El Niño. *Nature,* **321,** 827-832.

Chan J. C. L., J. Shi, and C. M. Lam, 1998: Seasonal forecasting of tropical cyclone activity over the western North Pacific and the South China Sea. *Wea. Forecasting*, ***13*****,** 997–1004.

Chan J. C. L., and J. E, Shi, 1999: Prediction of summer monsoon rainfall over South China. *Int. J. Climatol.,* **19,** 1255-1265.

Chan J. C. L., J. E. Shi, and C. M. Liu, 2001: Improvements in the seasonal forecasting of tropical cyclone activity over the western North Pacific. *Wea. Forecasting,* **16,** 491–498.

Chand S., K. J. E. Walsh, and J. C. L. Chan, 2010: A Bayesian Regression Approach to Seasonal Prediction of Tropical Cyclones Affecting the Fiji Region. *J. Climate,* **23,** 3425-3445.

Cheung K. K. W., 2004: Large-Scale Environmental Parameters Associated with Tropical Cyclone Formations in the Western North Pacific. *Journal of Climate*, **17,** 466-484.

Chu P.-S., and X. Zhao, 2007: A Bayesian regression approach for predicting tropical cyclone activity over the central North Pacific. *J. Climate,* **20,** 4002–4013.

Coughlan M. J., 1979: Recent variations in annual-mean maximum temperature over Australia. *Quart. J. Roy. Meteor. Soc.,* **105,** 707-719.

Dare R. A., and N. E. Davidson, 2004: Characteristics of Tropical Cyclones in the Australian Region. *Mon. Wea. Rev.*, **132,** 3049-3065.

Deacon E. L., 1953: Climatic change in Australia since 1880. *Aust. J. Physics,* **6,** 209-218.

Drosdowsky W., and M. Williams, 1991: The Southern Oscillation in the Australian Region. Part I: Anomalies at the Extremes of Oscillation. *J. Climate,* **4,** 619-638.

Drosdowsky W., and F. Woodcock, 1991: The South Pacific and southeast Indian Ocean tropical cyclone season 1988-1989. *Aust. Met. Mag.,* **39,** 113-129.

Elsner J. B., and C. P. Schmertmann, 1993: Improving extended-range seasonal predictions of intense Atlantic hurricane activity. *Wea. Forecasting*, **8,** 345–351.

Elsner J. B., and T. H. Jagger, 2004: A hierarchical Bayesian approach to seasonal hurricane modeling. *J. Climate,* **17,** 2813–2827.

Elsner J. B., and T. H. Jagger, 2006: Prediction models for annual U.S hurricane counts. *J. Climate,* **19,** 2935–2952.

Elsner J. B., T. H. Jagger, M. Dickinson, and D. Rowe, 2008: Improving Multiseasonal Forecast of North Atlantic Hurricane Activity. *J. Climate,* **21,** 1209-1219.

Emanuel K. A., 1986: An air-sea interaction theory for tropical cyclones. Part I. J. *Atmos. Sci.,* **43,** 585-604.

Emanuel K. A., 1991: The Theory of Hurricanes. *Annu. Rev. Fluid Mech., **23,*** 179-96.

Evans J. L., and R. L. Allan, 1992: El Nino/Southern Oscillation modification to the structure of the monsoon and tropical cyclone activity in the Australasian region. Int. *J. Climatol.,* **12,** 611-623.

Everitt B.S. 2003: *The Cambridge Dictionary of Statistics*, CUP.

Feng M., and G. Meyers, 2003: Interannual variability in the tropical Indian Ocean—A two-year time scale in the Indian Ocean dipole. *Deep-Sea Res. II,* **50,** 2263–2284.

Garreaud R. D., and D. S. Battisti, 1999: Interannual (ENSO) and interdecadal (ENSO-like) variability in the Southern Hemisphere tropospheric circulation. *J. Climate,* **12,** 2113–2123.

Gelman A., 1992: Iterative and Non-Iterative Simulation Algorithms. *Computing Science and Statistics (Interface Proceedings)*, **24,** 433-438.

Glahn H. R., 1985: *Statistical weather forecasting*. In: A.H. Murphy and R.W. Katz, eds., Probability, Statistics, and Decision Making in the Atmospheric Sciences. Boulder, Westview, 289–335.

Goebbert K. H., and L. M. Leslie, 2010: Interannual Variability of Northwest Australian Tropical Cyclones. *J. Climate,* **23,** 4538-4555.

Goh A. Z. C., and J. C. L. Chan, 2010: An improved statistical scheme for the

prediction of tropical cyclones making landfall in South China. *Wea. Forecasting,* **25,** 587-593.

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

Gray W. M., 1975: Tropical cyclone genesis. *Dept. of Atm. Sci. Paper No. 234, Colorado State University, Fort Collins, CO,* 121 pp [Available from Department of Atmospheric Sciences, Colorado State University, Ft. Collins, CO 80523].

Gray W. M., 1984: Atlantic seasonal hurricane frequency. Part II: Forecasting its variability. *Mon. Wea. Rev.,* **112,** 1669–1683.

Gray W. M., C. W. Landsea, P. W. Mielke Jr., and K. J. Berry, 1992: Predicting Atlantic seasonal hurricane activity 6–11 months in advance. *Wea. Forecasting,* **7,** 440–455.

Gray W. M., C. W. Landsea, P. W. Mielke Jr., and K. J. Berry, 1993: Predicting Atlantic basin seasonal tropical cyclone activity by 1 August. *Wea. Forecasting,* **8,** 73–86.

Gray W. M., C. W. Landsea, P. W. Mielke Jr., and K. J. Berry, 1994: Predicting Atlantic basin seasonal tropical cyclone activity by 1 June. *Wea. Forecasting,* **9,** 103-115.

Hackert E. C., and Hastenrath, S., 1986: Mechanisms of Java rainfall anomalies. *Mon. Wea. Rev.,* **114,** 745-757.

Halpert M. S., and C. F. Ropelewski, 1992: Surface temperature patterns associated with the Southern Oscillation. *J. Clim.,* **5,** 577-593.

Hastings W. K., 1970: Monte Carlo Sampling Methods Using Markov Chain and Their Applications. *Biometrica,* **57,** 97-109.

Holland G. J., 1981: On the quality of the Australian tropical cyclone data base. *Aust. Meteor. Mag.,* **29,** 169–181.

Holland G. J., 1984: On the climatology and structure of tropical cyclones in the Australian/southwest Pacific region: I. Data and tropical storms. *Aust. Meteor. Mag.,*

32, 1–15.

Hsieh W. W., 2009: *Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels.* Cambridge University Press, 364pp.

Jones P. D., 1991: Historical records of cloud cover and climate for Australia. *Aust. Meteor. Mag.,* **39,** 181-189.

Jones P. D., 1999: Characteristics of Australian land surface temperature variability. *Theor. Appl. Climatol.,* **63,** 11-31.

Karoly D. J., 1989: Southern Hemisphere circulation features associated with El Niño–Southern Oscillation. *J. Climate,* **2,** 1239– 1252.

Kiladis G. N., and H. F. Diaz, 1989: Global climatic anomalies associated with extremes in the Southern Oscillation. *J. Clim.,* **2,** 1069-1090.

Kleinschmidt E. Jr., 1951: Principles of the theory of tropical cyclones. *Arch. Meteor. Geophys., Bioklimatol.,* **4**, 53-72.

Klotzbach P. J., 2007: Revised prediction of seasonal Atlantic basin tropical cyclone activity from 1 August. *Wea. Forecasting,* **22,** 937-949.

Klotzbach P. J., A. Barnston, G. Bell, S.J. Camargo, J.C.L. Chan, A. Lea, M. Saunders, and F. Vitart, 2010: Seasonal Forecasting of Tropical Cyclones. *Global Guide to Tropical Cyclone Forecasting, 2nd edition*, C. Guard, editor, WMO, *in press*.

Kuleshov Y., and G. de Hoedt, 2003: Tropical cyclone activity in the Southern Hemisphere. *Bull. Austr. Met. Oceanogr. Soc.,* **16,** 135-137.

Kuleshov Y., L. Qi, R. Fawcett, and D. Jones, 2008: On tropical cyclone activity in the Southern Hemisphere: Trends and the ENSO connection. *Geophys. Res. Lett.,* **35,** L14S08, doi:10.1029/ 2007GL032983.

Kuleshov Y., F. Chane Ming, L. Qi, I. Chouaibou, C. Hoareaux, and F. Roux, 2009: Tropical cyclone genesis in the southern hemisphere and its relationship with the ENSO. *Ann. Geophys.,* **27,** 2523–2538.

Kuleshov Y., R. Fawcett, L. Qi, B. Trewin, D. Jones, J. McBride, and H. Ramsay, 2010: Trends in tropical cyclones in the South Indian Ocean and the South Pacific Ocean. *J. Geophys. Res.,* **115,** D01101, doi:10.1029/2009JD012372.

Landsea C. W., and W. M. Gray, 1992: The strong association between western Sahel monsoon rainfall and intense Atlantic hurricanes. *J. Climate*, **5**, 435-453.

Larget B. and D. Simon, 1999: Markov chain Monte Carlo algorithms for the Bayesian analysis phylogenetic trees. *Mol. Biol. Evol.,* **16,** 750-759.

Lau K-M., 1985: Elements of a stochastic-dynamical theory of the long-term variability of the El Niño/Southern Oscillation. *J.* Atmos. *Sci.,* **42,** 1552-1558.

Lehmiller G. S., T. B. Kimberlain, and J. B. Elsner, 1997: Seasonal prediction models for North Atlantic basin hurricane location. *Mon. Wea. Rev.,* **125,** 1780–1791.

Leroy A., and M. C. Wheeler, 2008: Statistical Prediction of Weekly Tropical Cyclone Activity in the Southern Hemisphere. *Mon. Wea. Rev.,* **136,** 3637-3654.

Liu K. S., and J. C.-L. Chan, 2003: Climatological characteristics and seasonal forecasting of tropical cyclones making landfall along the South China Coast. *Mon. Wea. Rev,* **131,** 1650–1662.

Liu K. S., and J. C.-L. Chan, 2010: Interannual variation of Southern Hemisphere tropical cyclone activity and seasonal forecast of tropical cyclone number in the Australian region. *Int. J. Climatol. (2010)***,** DOI: 10.1002/joc.2259.

Liu K. S., and J. C.-L. Chan, cited 2011: 2010/11 Predictions of Seasonal Tropical Cyclone Activity in the Australian Region. [Available online at http://weather.cityu.edu.hk/tc_forecast/2010_forecast_NOV.htm.]

Lough J. M., 1995: Temperature variations in a tropical- subtropical environment: Queensland, Australia 1910-1987. *Int. J. Climatol.,* **15,** 77-95.

Lu J., G. Chen, and D. M. W. Frierson, 2008: Response of the Zonal Mean Atmospheric Circulation to El Niño versus Global Warming. *J. Climate,* **21,** 5835-5851. doi: 10.1175/2008JCLI2200.1

Meehl G.A., 1987: The annual cycle of interannual variability in the tropical Pacific

and Indian Ocean regions. *Mon. Wea. Rev.,* **115,** 27-50.

McBride J. L., and T. D. Keenan, 1982: Climatology of tropical cyclone genesis in the Australian region. *Int. J. Climatol.,* **1,** 13-33.

McBride J. L., and N. Nicholls, 1983: Seasonal relationships between Australian rainfall and the Southern Oscillation. *Mon. Wea. Rev.,* **111,** 1998-2004.

McBride J. L. and Zehr R., 1981: Observational analysis of tropical cyclone formation. Part II: Comparison of non-developing versus developing systems *J. Atmos. Sci*., **38,** 1132-51.

McDonnell K. A., and N. J. Holbrook, 2004a: A Poisson regression model of tropical cyclogenesis for the Australian-southwest Pacific Ocean region. *Wea. Forecasting,* **19,** 440-455.

McDonnell K. A., and N. J. Holbrook, 2004b: A Poisson regression model approach to predicting tropical cyclogenesis in the Australian/southwest Pacific Ocean region using the SOI and saturated equivalent potential temperature gradient as predictors. *Geophys. Res. Lett.,* **31,** L20110, doi:10.1029/2004GL020843.

McDonnell K. A., N. J. Holbrook, and H. Shaik, 2006: Seasonal forecasts of tropical cyclone numbers in the Australian/southwest Pacific Ocean region using a new Poisson regression model: verification of 2005/06 season forecast and forecast for 2006/07 season. *Bull. Austr. Met. Oceanogr. Soc.,* **19(6),** 126-128.

Meyers G., P. C. McIntosh, L. Pigot, and M. J. Pook, 2007: The years of El Niño, La Niña, and interactions with the tropical Indian Ocean. *J. Climate,* **20,** 2872-2880.

Neal R. M., 2003: Slice Sampling. *Ann. Stat.,* **31,** 705-767.

Nicholls N., 1979a: A simple air-sea interaction model. *Q. J. R. Meteorol.* Soc., **105,** 93-105.

Nicholls N., 1979b: A possible method for predicting seasonal tropical cyclone activity in the Australian region. *Mon. Wea. Rev*., **107,** 1221–1224.

Nicholls N., 1984a: El Niño Southern Oscillation and North Australian sea surface temperature. *Trop. Ocean-Atmos. Newsl.,* **24,** 11-12.

Nicholls N., 1984b: The Southern Oscillation and Indonesian sea surface temperature. *Mon. Wea. Rev.,* **112,** 424-432.

Nicholls N., 1984c: A system for predicting the onset of the north Australian wet-season. *J. Climatol.,* **4,** 425-435.

Nicholls N., 1984d: The southern oscillation, sea-surface temperature, and interannual fluctuations in Australian tropical cyclone activity. *Int. J. Climatol.,* **4,** 661-1149.

Nicholls N., 1985a: Predictability of interannual variations of Australian seasonal tropical cyclone activity. *Mon. Wea. Rev.,* **113,** 1144–1149.

Nicholls N., 1985b: Towards the prediction of major Australian droughts. *Austr. Meteor. Mag.,* **33,** 161-166.

Nicholls N., 1992: Recent performance of a method for forecasting Australian seasonal tropical cyclone activity, *Aust. Meteor. Mag*., **40,** 105–110.

Nicholls N., Drosdowsky W., and B. Lavery, 1997: Australian rainfall variability and change. *Weather,* **52**, 66–71.

NOAA, cited 2011: Cold & Warm Episodes by Season. [Available online at http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml.]

Ntzoufras I., 2009: *Bayesian modeling using WinBUGS.* Wiley series in Computational Statistics, 698, John Wiley and Sons, 520pp.

Owens B. F., and C. W. Landsea, 2003: Assessing the Skill of Operational Atlantic Seasonal Tropical cyclone Forecasts. *Wea. Forecasting,* **8,** 45-54.

Peters G. W., P. V. Shevchenko, and M. V. Wüthrich, 2009: Model uncertainty in claims reserving within Tweedie's compound Poisson models. *ASTIN Bull.,* **39,** 1-33.

Philander S. G. H., 1985: El Niño and La Niña. *J. Atmos. Sci.,* **42**, 2652–2662.

Philander S. G. H., 1990: *El Niño, La Niña and the Southern Oscillation,* Academic Press, San Diego, Calif., 293 pp.

Philander S. G. H., 1985: Ocean-Atmosphere interactions in the tropics: A review of

recent theories and models. *J. Applied Meteor.,* **31,** 938-945.

Pittock A.B., 1975: Climatic change and the patterns of variation in Australian rainfall. *Search,* **6,** 498-504.

Power S., Tseitkin, F., Torok, S. J., Lavery, B., Dahni, R., and B. McAvaney, 1998: Australian temperature, Australian rainfall and the Southern Oscillation, 1910±1992: Coherent variability and recent changes. *Aust. Meteor. Mag.,* **47,** 85-101.

Ramsay H. A., L. M. Leslie, P. J. Lamb, M. B. Richman, and M. Leplastrier, 2008: Interannual Variability of Tropical Cyclones in the Australian Region: Role of Large-Scale Environment. *J. Climate,* **21,** 1083-1103.

Rao S. A., S. K. Behera, Y. Masumoto, and T. Yamagata, 2002: Interannual variability in the subsurface tropical Indian Ocean. *Deep-Sea Res. II,* **49,** 1549–1572.

Rasmusson E. M., 1984. 'El Niño: the ocean/atmosphere connection', *Oceanus,* **27,** 36-39.

Rasmusson E.M., and T. H. Carpenter, 1982: Variations in tropical sea surface wind fields associated with the Southern Oscillation/El Niño. *Mon. Wea. Rev.,* **110,** 354-384.

Ready S., and F. Woodcock, 1992: The South Pacific and southeast Indian Ocean tropical cyclone season 1989-1990. *Aust. Met. Mag.,* **40,** 111-121.

Risbey J. S., M. J. Pook, P. C. McIntosh, M. C. Wheeler, and H. H. Hendon, 2009: On the remote drivers of rainfall variations in Australia. *Mon. Wea. Rev.,* **137,** 3233-3253.

Roberts G. O., and J. S. Rosenthal, 1999: Convergence of slice sampler Markov chains. *J. R. Statist. Soc.,* **61,** 643-660.

Robinson W. A., 2002: On the midlatitude thermal response to tropical warming. *Geophys. Res. Lett.,* **29,** 1190, doi:10.1029/ 2001GL014158.

Ropelewski C. F., and M. S. Halpert, 1996: Quantifying Southern Oscillation–Precipitation relationships. *J. Climate,* **9,** 1043–1059.

Saji N. H., T. Ambrizzi, and S. E. T. Ferraz, 2005: Indian Ocean Dipole Mode events and austral surface air temperature anomalies. *Dym. Atmos. Oceans.,* **39,** 87-101.

Saji N. H., B. N. Goswami, P. N. Vinayachandran, and T. Yamagata, 1999: A dipole mode in the tropical Indian Ocean. *Nature,* **401,** 360-363.

Saji N. H., and T. Yamagata, 2003: Structure of SST and surface wind variability during Indian Ocean dipole mode years: COADS observations. *J. Climate,* **16,** 2735–2751.

Saunders M. A., and A. S. Lea, 2005: Seasonal prediction of hurricane activity reaching the coast of the United States. *Nature,* **434,** 1005-1008.

Simmonds I., and M. H. Lynch, 1992: The influence of pre-existing soil moisture content on Australian winter climate. *Int. J. Climatol.,* **12,** 33-54.

Simpson R. H., 1974: The hurricane disaster potential scale. *Weatherwise,* **27,** 169-86.

Sinclair M. R., J. A. Renwick, and J. W. Kidson, 1997: Low-frequency variability of Southern Hemisphere sea level pressure and weather system activity. *Mon. Wea. Rev.,* **127,** 2531–2542.

Solow A., and N. Nicholls, 1990: The relationship between the Southern Oscillation and tropical cyclone frequency in the Australian region. *J. Climate*, 3, 1097–1101.

Streten N. A. 1981: Southern Hemisphere sea surface temperature variability and apparent associations with Australian rainfall. *J. Geophys. Res.,* **86,** 485-497.

Streten N. A. 1983: Extreme distributions of Australian annual rainfall in relation to sea surface temperature. *J. Climatol.,* **3,** 143-153.

Student 1907: On the error of counting with a haemacytometer. *Biometrika,* **5,** 351–360.

Tapp R. G., and S. L. Barrel, 1984: The north-west Australian cloud band: Climatology, characteristics and factors associated with development. *Int. J. Climatol.*, ***4*,** 411–424.

Tonkin H., C. Landsea, G.J. Holland, and S. Li, 1997: Tropical Cyclones and Climate

Change: A preliminary Assessment. *Assessing Climate Change*: *Results from the Model Evaluation Consortium for Climate Assessment* (Ed. W. Howe and A. Henderson-Sellers). Gordon and Breach, Australia, ISBN: 90- 5699-067-5, pp 323-354.

Ummenhofer C. C., M. H. England, P. C. McIntosh, G. A. Meyers, M. J. Pook, J. S. Risbey, A. Sen Gupta, and A. S. Taschetto, 2009: What causes Southeast Australia's worst droughts? *Geophys. Res. Lett.,* **36,** L04706, doi:10.1029/2008GL036801.

UNISYS, cited 2011: Hurricane/Tropical Data. [Available online at http://weather.unisys.com/hurricane/.]

Van Loon, H., and D. J. Shea, 1985: The Southern Oscillation. Part IV: The Precursors of 15°S to the Extremes of the Oscillation. *Mon. Wea. Rev.,* **113,** 2063-2074.

Van Loon H., and D. J. Shea, 1987: The Southern Oscillation. Part VI: Anomalies of sea level pressure on the Southern Hemisphere and of Pacific sea surface temperature during the development of a warm event. *Mon. Wea. Rev.,* **115,** 370–379.

Watterson I. G., 1997: The diurnal cycle of surface air temperature in simulated present and doubled CO2 climates. *Clim. Dynam.,* **13,** 533-545.

Webster P. J., A. M. Moor, J. P. Loschnigg, R. R. Leben, 1999: Coupled ocean–atmosphere dynamics in the Indian Ocean during 1997–98. *Nature,* **401,** 356-360.

Wilks D. S., 1995: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 467 pp.

Wright P. B., 1984: Relationship of SST near northern Australia to the Southern Oscillation. *Trop. Ocean-Atmos. Newsl.,* **27,** 17-18.

Wright P. B., Mitchell, T. P., and J. M. Wallace, 1985: *Relationships between Surface Observations over the Global Oceans and the Southern Oscillation,* NOAA, Seattle, WA, Data Report ERL PMEL-I2, 61 pp.

Xie, S. P., H. Annamalai, F. A. Schott, and J. P. McCreary, 2002: Structure and mechanisms of south Indian Ocean climate variability. *J. Climate,* **15,** 867–878.

Yamagata T., S. Behera, J.-J. Luo, S. Masson, M. R. Jury, and S. A. Rao, 2004: Coupled ocean-atmosphere variability in the tropical Indian Ocean. *Earth's Climate: The Ocean-Atmosphere Interaction*, *Geophys. Monogr.,* **147,** edited by C. Wang, S.-P. Xie, and J. Carton, pp. 189–211, AGU, Washington, D. C.

Zhang Y., J. M. Wallace, and D. S. Battisti, 1997: ENSO-like interdecadal variability: 1900–93. *J. Climate,* **10,** 1004–1020.