EVALUATION OF CLIMATE PREDICTABILITY FOR MULTIPLE CLIMAE MODELS AT VARIOUS TIME SCALES

by

Waqar Younas

M.Phil. (Physics), Quaid-e-Azam University, Pakistan, 2007
M.Sc. (Physics), Quaid-e-Azam University, Pakistan, 2004
B.Sc., Punjab University, Pakistan, 2002

DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN
NATURAL RESOURCES AND ENVIRONMENTAL STUDIES

UNIVERSITY OF NORTHERN BRITISH COLUMBIA

December 2014

© Waqar Younas, 2014
Abstract

The predictability of the Pacific North American (PNA) pattern is evaluated on time scales from days to months using state-of-the-art dynamical multiple model ensembles including the Canadian Historical Forecast Project (HFP2) ensemble, the Development of a European Multimodel Ensemble System for Seasonal-to-Interannual prediction (DEMETER) ensemble, and the Ensemble Based Predictions of Climate Changes and their Impacts (ENSEMBLES). Some interesting findings in this study include (i) Multiple-model ensemble (MME) skill was better than skill from most of the individual models; (ii) both actual prediction skill and potential predictability increased as the averaging time scale increased from days to months; (iii) There is no significant difference in actual skill between coupled and uncoupled models, in contrast with the potential predictability where coupled models performed better than uncoupled models; (iv) relative entropy ($R_E$) is an effective measure in characterizing the potential predictability of individual predictions, whereas the mutual information (MI) is a reliable indicator of overall prediction skill; (v) Compared with conventional potential predictability measures of the signal-to-noise ratio, the MI-based measures characterized more potential predictability when the ensemble spread varied over initial conditions. It is also confirmed that from monthly to seasonal time scales, the potential predictability of PNA is teleconnected with ENSO.

The predictive skill on intra-seasonal time scales in the tropics is linked to Madden-Julian Oscillations (MJO). Using recently developed framework of potential predictability, information-based and ensemble based predictability measures were explored on multiple time scales for MJO predictability. Results show that there is no significant difference in the
simulation of MJO in coupled (CanCM3) and uncoupled (GCM3) models. Both models simulated the tropical low frequency variability reasonably well compared with observations with some positive bias in CanCM3 in simulating the precipitation, whereas GCM3 could not capture the upper zonal wind variability on eastern Pacific. The MJO prediction skill is significantly better in CanCM3 than in GCM3 in terms of correlation and Root Mean Square Error (RMSE). In terms of potential predictability of MJO, coupled models forecast skill again dominated uncoupled models prediction skill. It was found that $A_{C,M}$ estimate more potential predictability than $A_{C,P}$. MI is seen to be reliable predictor of model overall skill. MJO index is found to be strong component of internal variability in terms of DC. There is no significant difference found in MJO prediction skill with time average (both for actual and potential skill), where time average skill is calculated using 3-days, 5-days and 7-days mean.
# TABLE OF CONTENTS

Abstract .............................................................................................................................................. ii

Table of Contents ............................................................................................................................. iv

List of Tables ..................................................................................................................................... vi

List of Figures ................................................................................................................................... vii

Glossary .............................................................................................................................................. xii

Acknowledgements .......................................................................................................................... xiii

## CHAPTER 1: Introduction

1.1 Predictability ........................................................................................................................................ 1

1.2 Pacific North American (PNA) Index ................................................................................................ 2

1.3 Madden-Julian Oscillation (MJO) .................................................................................................... 4

1.4 Objectives and Outlines ......................................................................................................................... 9

## CHAPTER 2: Models data and Experimental Framework

2.1 Ensemble Prediction Products and Observed Data for PNA study ................................................ 12

2.2 Ensemble Prediction Products and observed data for PNA study ................................................. 13

2.3 Measures of Skill for PNA Predictability ......................................................................................... 14

2.3.1 Actual Measures of Predictability ............................................................................................ 14

2.3.2 Potential Predictability Measures ............................................................................................. 15

2.3.3 Predictable Component Analysis (PrCA) .................................................................................... 19

2.4 Measures of Predictability for MJO ................................................................................................. 20

## CHAPTER 3: PNA Predictability at Various Time Scales

3.1 Properties of Ensemble Systems ...................................................................................................... 22

3.2 PNA Actual Skill at different time scales ......................................................................................... 23

3.3 PNA Potential Predictability ............................................................................................................. 25

3.3.1 MI based Predictability ............................................................................................................. 29

3.3.2 Comparison between MI and SNR ............................................................................................ 30
3.4 Predictability between Coupled and Uncoupled Model 32
3.5 Temporal Variation of PNA Predictability and its Source 34
3.6 Predictable Component Analysis (PrCA) 36
3.7 Discussion and Conclusion 37

CHAPTER 4: Prediction Skill of Madden-Julian Oscillation in AGCM and CGCM Ensemble Forecasts 59
4.1 MJO Index for Prediction Skill 59
4.2 Sub-seasonal Low Frequency Variability 62
4.3 Actual Prediction Skill of MJO bi-variate Index 64
4.4 Potential Prediction Skill of MJO bi-variate Index 67
   4.4.1 Relative Entropy, Predictive Information and Predictive Power 67
   4.4.2 Comparison of Actual and Potential Prediction Skill 71
4.5 Potential Predictability Source for MJO Index 72
4.6 Summary and Conclusion 72

CHAPTER 5: Prediction Skill Comparison between PNA and MJO 90
5.1 Coupling Role in the Prediction of PNA and MJO indices 90
5.2 Coupling Role in the Potential Predictability of PNA and MJO 91
5.3 Relative Role of Signal and Noise in Potential Predictability at Various Time Scales 92
5.4 The Impact of Time Average on Prediction Utility at Various Time Scales 94
5.5 Capability of Potential Predictability Measures in Quantifying the Prediction Skill at Various Time Scales 95

References 110
Appendix 119
List of Tables

Table 3.1: Actual and potential prediction skill ($AC_M$) over different time scales using MME_H, MME_E (Correlation value of 0.5 is chosen as threshold for 41
List of Figures

Figure 3.1 NAO and PNA patterns obtained using rotated EOF analysis of 500mb monthly NCEP geopotential height data. Contour interval is 10m with negative contours dashed. Dark (light) shading indicates values <10m (>10m) 42

Fig. 3.2 RMSE and Ensemble spread as a function of lead time over daily time scale, for different models and their MME’s. Here the ENSEMBLES models are shown with “_E” and DEMETER models are shown with “_D”. CERFACS and LODYC are DEMTER models 43

Fig. 3.3 Correlation and RMSE as function of lead time for different time scales, for different models. Persistent skill and climatological spread is also shown for each time scale. Here different colors are indicating different models 44

Fig. 3.4 Same as Fig. 2 but using only MME_H, MME_D and MME_E. Here vertical error bars are the sample standard deviation calculated using bootstrap experiment at each lead time 45

Fig. 3.5 Correlation coefficients of different MME over different time scales. Here the daily correlation values are averaged over one week (weekly_avg), 2 weeks (bi_weekly_avg) and month (monthly_avg) to compare with corresponding weekly, bi-weekly and monthly time scale correlation 46

Fig. 3.6 a) Anomaly correlation for MME_H over different time scales as function of lead time, here the daily correlation is shown for first 20 days only. (b) Climatological variance of averaging time by observation (stars) and MME_H (open diamonds), (c) Persistence
Correlation at different time scales. The legends for (a) and (c) are same and shown in the bottom.

Fig. 3.7 Comparison of $AC_M$ with $AC_P$ over different time scales for MME_H and MME_E

Fig. 3.8 Actual and potential skill for MME_H and MME_E. Here the vertical bars are the sample standard deviation calculated using bootstrap experiment (see context).

Fig. 3.9 The PDF of prediction for different randomly chosen lead times for weekly time scale for ECMWF models and HFP2 models. Observed PDF for weekly time scale is also shown as reference.

Fig. 3.10 $AC_M$ as function of lead time over different time scales for MME_E and MME_H is presented. Here vertical error bars are the sample standard deviation calculated using bootstrap experiment (see text).

Fig. 3.11 Monthly correlation skill of SST using persistence (left panel) and MME_E (right panel) at different lead months.

Fig. 3.12 Averaged RE as function of initial time is shown at different time scales as function of initial time for MME_E.

Fig. 3.13 Variations in averaged SC and DC as function of initial time for MME_E over different time scales.

Fig. 3.14 Spatial pattern Correlation between predicted ENSEMBLES SST and PNA of MME_E (left panel) for different time lags (based on selected initial conditions in which
signal component was high). Here SST is leading and PNA is lagging behind. Right: Same as left but using observed SST and PNA of MME_H

Fig. 3.15 The leading PrCA modes for different lead months for MME_E which capture the PNA related variabilities. In the title of each pattern, the variance explained based on total predictability by that pattern is also shown

Fig. 3.16 Same as Fig. 14 except using MME_H

Fig. 3.17 The projection of PrCA patterns of Fig. 14 and Fig. 15, on observed SST

Fig. 4.1 Longitudinal distribution of EOF1 and EOF2 from the combined analysis of prlr, u850 and u200 of NCEP-NCAR reanalysis

Fig. 4.2 Variability patterns of low-frequency (20-100 days) of u850 for the NCEP-NCAR reanalysis, GCM3, and CanCM3. Areas with variances greater than 6m²/s² are shaded

Fig. 4.3 Variability patterns of low-frequency (20-100 days) of u200 for the NCEP-NCAR reanalysis, GCM3, and CanCM3. Areas with variances greater than 24m²/s² are shaded

Fig. 4.4 Variability patterns of low-frequency (20-100 days) of prlr for the GPCP, GCM3, and CanCM3. Areas with variances greater than 24m²/s² are shaded

Fig. 4.5 Correlation and RMSE of MJO bi-variate Index of GCM3 and CanCM3 from 1979-2001

Fig. 4.6 Seasonal dependence of MJO correlation skill in coupled and uncoupled model ensemble forecast from 1979-2001
Fig. 4.7 Comparison of $A_{CM}$ and $A_{CP}$ between GCM3 and CanCM3

Fig. 4.8 Potential predictability $A_{CM}$ and $A_{CP}$ comparison between CanCM3 and GCM3 for MJO bi-variate index for the period 1979-2001

Fig. 4.9 Variation of average $RE_a$, PI and PP as function of initial time for GCM3 (left panel) and CanCM3 (right panel) for the period 1979-2001. Here average is done over first 20 lead days only

Fig. 4.10 Variation of average $SC$ (left panel) and average $DC$ (right panel) as function of lead time and initial time both for GCM3 and CanCM3. Here average is done over first 20 lead days only

Fig. 4.11 Correlation of $RE$ with $SC$ and $DC$, for GCM3 (left panel) and for CanCM3 (right panel)

Fig. 4.12 Comparison of Actual and potential skill between GCM3 and CanCM3. The potential skill is estimated using MI

Fig. 4.13 Daily SST anomaly correlation for Persistence (left paned) and predicted (right panel) for different lead days

Fig. 5.1 Variations in PI for PNA index with time average for MME_H (left panel) and for MME_E (right panel)

Fig. 5.2 Variations in PP for PNA index with time average for MME_H (left panel) and for MME_E (right panel)

Fig. 5.3 Correlation of SC and DC with RE at various time scales for PNA index using MME
Fig. 5.4 Same as Fig. 5.2 but for MME of HFP2 (uncoupled) models

Fig. 5.5 Time average prediction skill (correlation) of MJO index using 3-days, 5-days and 7-days mean over daily time scale

Fig. 5.6 Time average potential prediction skill of MJO index using 3-days, 5-days and 7-days mean over daily time scale. $AC_{MI}$ is used as potential prediction skill

Fig. 5.7 Scatterplots of Mutual information $MI$ vs $r$ (left) at different time scales and $MI$ vs RMSE (right) at different time scales, for MME_H

Fig. 5.8 Same as Fig. 5.7 but for MME_E

Fig. 5.9 Scatterplots of mutual information $MI$ vs estimated $MI$ from $r$ for MME_H (left) and MME_E (right) at different time scales

Fig. 5.10 Scatterplots of Mutual information $MI$ vs $r$ and RMSE for GCM3 (upper panel) and CanCM3 (lower panel), of MJO index at daily time scale

Fig. 5.11 Scatterplots of mutual information $MI$ vs estimated $MI$ from $r$ for MJO daily index using GCM3 (left) and CanCM3 (right)
Glossary

PNA: Pacific North American

MJO: Madden-Julian Oscillation

HFP2: Historical Forecast Project 2

CHFP2: Coupled Historical Forecast Project 2

AGCM: Atmospheric General Circulation Model

CGCM: Coupled General Circulation Model

RE: Relative Entropy

PI: Predictive Information

PP: Predictive Power

MI: Mutual Information

AC$_MI$: Anomaly Correlation based on Mutual Information

AC$_P$: Anomaly correlation based on perfect model assumption

MME$_E$: Multi Model Ensemble Mean of ENSEMBLES

MME$_H$: Multi Model Ensemble Mean of HFP2

RMSE: Root Mean Square Error

SC: Signal Component

DC: Dispersion Component

NAO: North Atlantic Oscillation

REOF: Rotated Empirical Orthogonal Functions
Acknowledgements

First and foremost, I would like to show my deepest gratitude to my supervisor Dr. Youmin Tang and co-supervisor Dr. Peter Jackson for their guidance, mentoring, and constant encouragement during my study period in UNBC. Their keen and vigorous academic observation enlightens me not only in this thesis but also in my future study. I would like to express my gratitude to the other committee members, Dr. Bin Yu, Dr Pranesh Kumar and Dr. Jueyi Sui, for their suggestions in my thesis proposal and during thesis research period. Thanks to the anonymous reviewers for constructive comments on my paper.

This thesis greatly benefits from Youmin’s courses, from which I learnt many helpful statistical methods and tools. Thanks Dr. Ken Otter, Dr. Laurie Chan and Dr. Phil Burton for their valuable teachings with interesting scientific topics of environmental sciences during my course study.

I am grateful for the numerical computation supports of Dr. Jean Wang at the HPC Lab of UNBC. I owe my sincere gratitude to other group members Zhiyu Wang, Xiaoqin Yan, Yanjie Cheng, Ziwang Deng, Dejian Yang, Siraj Ul Islam, and Manoj Nambiar, for their friendships and discussions on research and careers. I also thank to all my friends in Canada.

My deepest gratitude to my parents, without their constant help and encouragement, I would not be able to achieve this goal. They faced hardships in their life so that their children can attain what they want. Especial thanks to my mother (Zahida Sultana), who prayed and practically spent her life in nourishing her children dreams. Without constant support of my family in Pakistan, I would not be able to finish my thesis.
This work is supported by NSERC Discovery Grant and UNBC Graduate Entrance Scholarship.
Chapter 1: Introduction

1.1. Predictability

Predictability is the study of the extent to which events can be predicted (e.g. DelSole 2004). The basic idea is that the initial state of a system is not known exactly and therefore is represented by a probability distribution function (PDF) (Lorenz, 1965, 1969; Epstein 1969). This distribution evolves with time under the principle that total amount of probability will always equals unity. The distribution before observations become available is called “prior” distribution while the distribution after the observations become available is called “posterior” distribution (DelSole 2004). The system is predictable if the posterior distribution differs in any way from the prior distribution. The above definition provides a single, consistent framework for quantifying different kinds of predictability (DelSole 2004).

Lorenz (1975) classified two kinds of climate predictability: the first kind refers to the initial value problem of predicting the evolution of the climate system given some estimate of its current state while the second kind refers to the boundary value problem of assessing a change in climate due to some external forcing (Collins 2002). The focus of this study will be on evaluating first kind of predictability only ranging from weather prediction to seasonal climate prediction. An example of second kind of predictability is global warming as a result of changing the concentration of CO2 in the atmosphere.

The current predictability limit on daily time scale is of the order of two weeks, initially estimated by Lorenz and others (Lorenz 1965; Lorenz 1969; Leith 1971; Buizza 1997; Kalnay 2003; Van den Dool et al. 2003; Tribbia and Baumhefner 2004). These studies
demonstrated that for realistic cases (i.e. available atmospheric observations and dynamical models), the limit for making skillful forecasts of mid-latitude weather systems is estimated at about two weeks. The motivation to have the atmospheric predictability beyond the deterministic predictability limit (e.g. 2 weeks) is primarily based upon the fact that a major part of the atmospheric variability is found at low frequencies. The separation between the low-frequency and high-frequency transient phenomena is natural since they result from different physical processes. The former has large-scale, barotropic structure whereas the latter are produced by synoptic baroclinic processes. It is precisely the large scales planetary waves that one is tempted to forecast in the long run. For atmospheric predictability on intraseasonal time scales and longer, the initial conditions contain information with much longer time scales than the dominant atmospheric instabilities. For example, the initial conditions contain information beyond the atmospheric predictability limit, and include details on the states of the ocean and land surface.

1.2. Pacific North American (PNA) Index

The Pacific North American (PNA) pattern is one of the most prominent modes of low frequency climate variability in the Northern Hemisphere during winter (e.g., Wallace and Gutzler 1981; Woodhouse 1997). It significantly affects the weather and climate anomalies over North America. For example, during the PNA positive phase, above-normal temperatures can be witnessed in the western United States. In its negative phase, dry and warm conditions are detected in the eastern United States while relatively dry and cold conditions may be experienced in the west (Yarnal and Diaz 1986; Leathers et al. 1991). It has been argued that, in the midlatitude region, the seasonal climate prediction skill is mainly determined by the prediction skill of the PNA and the North Atlantic Oscillation (NAO),
another important mode in mid-to-high latitudes (e.g., Hurrell 1995; Hurrell et al. 2003; Doblas-Reyes et al. 2003; Vitart 2004).

Past studies on PNA predictability covered several different aspects using models and observations, including model skill evaluation and analysis of the sources of predictability. While it has been well reported that PNA predictability, on monthly to seasonal time scales, is mainly from slowly varying external forcing, (e.g., the El Niño Southern Oscillation (ENSO) forcing) (e.g., Horrel and Wallace 1981; Hoskins and Karoly 1981; Simmons 1982; Sardeshmukh and Hoskins 1988; Kumar et al. 1996; Straus and Shukla 1997; Shukla et al. 2000), some other work also suggested that a considerable part of PNA predictability comes from the internal dynamics of midlatitude atmosphere inherent to atmospheric instabilities (baroclinic in nature in extratropics) and nonlinear interactions between large scale and synoptic scale atmospheric process (e.g., Lau 1981; Wallace and Blackmon 1983; Chervin 1986; Palmer 1993; Straus and Shukla 2002).

In the field of statistical predictability of seasonal climate, it has been challenging to separate predictability from internal variability which often is treated as stochastic forcing compared with large-scale slowly varying signals. Recently, information-based potential predictability measures were applied to study statistical predictability, providing a convenient way to explore the relative role of internal and external forcing on predictability. For example, Abramov et al. (2005) found, using a highly simplified model, that the ensemble spread associated with the internal dynamics is the main contributor to PNA potential predictability. On the other hand, Kleeman (2008) demonstrated in another simplified model, that the variation in midlatitude atmospheric predictability with respect to initial conditions is mainly
determined by ensemble signal related to external forcing. Thus, the source of PNA predictability is still an open question. A reasonable answer may be expected by exploring the relative contributions of external forcing and internal dynamics to the PNA potential predictability at different time scales using realistic atmospheric or climate models, which seems absent in the literature.

Another issue in the past studies of PNA predictability is the lack of a comprehensive evaluation of actual and potential predictability with time averaging. Usually the PNA predictability was evaluated using either actual skill (Renwick and Wallace 1995; Lin and Derome 1996; Nakaegawa and Kanamitsu 2006) or potential predictability (e.g., Barnett et al. 1997; Phelps et al. 2004; Abramov et al. 2005) separately, mainly at monthly to seasonal scales. Very few studies have focused on the evaluation of the PNA predictability at various time scales from days to months, in particular for potential predictability. One exception is the work of Johansson (2007) where he investigated the actual correlation skill of PNA and NAO from daily to seasonal time scales using five years of winter predictions by NCEP CFS from 2001 to 2005. The PNA actual and potential predictability study at an integral framework at various time scales is absent in the literature. These gaps will be main focus of this study.

1.3. Madden Julian Oscillation (MJO)

MJO is considered as a result of internal atmospheric dynamics involving the interaction

---

1 For a clarification, we define the potential predictability used in this study as the prediction skill of a 'perfect' forecast system (model), which does not make use of observations to define the actual state whereas the actual prediction skill is the prediction skill evaluated using observations. The predictability, used as a general expression, means both the potential predictability and actual prediction skill in this study.
between tropical convection and large scale circulation. It propagates eastward at a period of 30-90 days (Reichler and Roads, 2005). There is clear evidence that MJO influences not only the tropics but also plays an important role in the extra-tropics especially in the Pacific North American and the Atlantic sectors. For example, the tropical cyclone activity in the western and eastern north Pacific, in the Gulf of Mexico, and in southern Indian Ocean and Australia, is all related to the MJO active phase (Vitart 2009). In the extra-tropics, Lin and Derome (2009) found a lagged response of the MJO to the NAO. For seasonal and longer term climate prediction, the MJO represents an important component of stochastic forcing, for example westerly wind bursts and the development of El-Nino. MJO also serves as a key component of the atmospheric “noise” that can limit the skill associated with the long term forecasts (e.g. McPhaden, 1999; Moore and Kleeman 1998; Kessler and Kleeman, 2000).

Current global climate models suffer severe deficiencies in representing the variabilities related to MJO, especially in the form of convection (Waliser et al. 2008; Kim et al. 2009; Hung et al. 2013). Several studies on MJO simulation and variability using Atmospheric General Circulation Models (AGCMs) documented that the capability of AGCMs to simulate MJO is limited (Slingo and coauthors 1996; Kang and coauthors 2002; Wu et al. 2002; Waliser et al. 2003). In the presence of these difficulties in the representation of MJO, the prediction skill of the MJO is generally reported in the range of 15-20 days (cf. review by Waliser 2005; Jones et al. 2000; Lin et al. 2006; Ling et al. 2014).

Although there have been improvements in the GCMs to correctly simulate and forecast MJO, these improvements are largely model dependent and limited (Seo et al. 2009; Lin et al. 2006; Fu et al. 2007; Kim et al. 2009). The apparent deficiencies in MJO simulation and forecasts can be attributed to many factors which are not limited to air-sea coupling (Seo and
W. Younas: Evaluation of Climate Predictability

Wang 2010). For example, the deep convection scheme applied in the model is another most sensitive factor required to correctly simulate the MJO. A few other studies have shown that the addition of moisture triggers to cumulus convection schemes significantly improves MJO simulation (Wang and Schlesinger 1999). But this is beyond the scope of this study, so the role of coupling in MJO prediction skill, will be main emphasis in this study. Thus this work using different models may assist to evaluate the role of coupling in the prediction skill and predictability of MJO.

The role of air-sea coupling in maintaining and propagating MJO has been the topic of many studies (Waliser et al. 1999; Woolnough et al. 2000; Webster et al. 2002; Fu et al. 2003; Fu and Wang 2004; Rajendran and Arakawa 2004; Zheng et al. 2004; Rajendran and Kitoh 2006; Fu et al. 2007; Kim et al. 2010). These studies demonstrate that interactive air-sea coupling needs to be included in numerical models to obtain a reasonable representation of the MJO. However, some other studies suggested that inclusion of air-sea coupling does not lead to significant improvements in intraseasonal simulation (e.g. Hendon 2000; Inness and Slingo 2003; Bellon et al. 2008; Newman et al. 2009). So it is still an open question whether the coupling plays an important role in the prediction of MJO. Seo et al. 2009 assessed the effect of interactive air sea coupling on MJO forecasts by comparing forecasts from NCEP's fully coupled operational forecast model (CFS) and its atmospheric component Global Forecast system (GFS) model. They found that the coupled model marginally but consistently outperforms the uncoupled GFS skill. Another study by Pegion and Kirtman 2008 investigated the impact of air-sea interactions on the predictability of tropical intraseasonal oscillations by using the NCEP operational model in coupled, uncoupled and perfect model experiments. They found that the uncoupled model had similar skill to that of
the coupled models. Woolnough et al. (2007) also focused on the role of coupling in the prediction skill of MJO using the operational monthly forecast model of the European Centre for Medium Range Weather Forecasts (ECMWF). They concluded that the dynamical coupled model has improved skill compared with the uncoupled model with a persistent SST forecast, indicating that the role of coupling is important in the prediction skill of MJO. A more recent study by Fu et al. (2013) explored the differences and similarities in forecasting MJO using NCEP CFS and GFS forecast models. They found that air sea coupling extended MJO skill by about one week.

In terms of potential predictability of MJO, Waliser et al. 2003 estimated the potential predictability of the MJO at around 25-30 days using NASA Goddard Laboratory for the Atmosphere (GLA) AGCM. Rashid et al. (2011) using the POAMA model reported the potential predictability of MJO to be about 40 days. This upper bound of MJO skill has not yet been confirmed in other studies as the dynamical predictions of the MJO remain limited by model deficiencies and imperfect initial conditions. Using observational data, Ding et al. 2010 estimated the potential predictability of MJO at around 35 days. Moreover, the above mentioned studies used a perfect model framework in order to estimate the potential predictability of MJO. The perfect model potential prediction skill is often termed a “signal to noise ratio” based potential prediction skill. It is argued in Yang et al. (2012) that these measures underestimate the potential predictability because they measure a linear relationship between prediction (ensemble mean) and ‘observation’ (an ensemble member), which underestimates the true potential predictability that is statistically defined as the coherence between prediction and initial (boundary) conditions. On the other hand, information theory based Mutual Information (MI) potential predictability measures both the
linear and non-linear statistical dependence between prediction and observation (Yang et al. 2012; Tang et al. 2013). Thus a re-examination of potential predictability of MJO using new measures such as information theory is needed.

In this study, we examine the difference and similarities of MJO variability and predictability between coupled and atmospheric-only models; however our work is significantly different from previous studies. One distinctive difference is to study the MJO potential prediction skill in coupled and uncoupled models using newly developed advanced statistical techniques (such as information theory), which to our knowledge is absent in literature. Information theory based potential predictability measures were therefore applied to study statistical predictability. These measures provide a convenient way to explore the relative role of external and internal forcing on predictability. Previously, these measures were extensively used in other studies related to El-Nino Southern Oscillation (Tang et al. 2008) and for midlatitude predictability studies (Kleeman 2008) but have not yet been applied to study the potential predictability at intraseasonal time scales.

The second way in which our work is distinctly different is related to the lack of time averaged MJO prediction skill analysis in other studies. It is well established that instantaneous states of weather can be predicted beyond two weeks (due to slowly moving boundary forcing) but this predictability signal is difficult to detect due weather noise. Time averaging the instantaneous states reduces the amplitude of weather noise, without appreciably reducing the predictable part, thus increasing the signal-to-noise ratio. We will incorporate the time averaged predictability of MJO using 3-day, 5-day and 7-day means and expect the predictability to improve.
To achieve our goals, we will use simulation results from state of the art coupled and uncoupled global climate models from CCCma (the Canadian Centre for Climate Modeling and Analysis). Specifically simulations from the CHFP2 coupled climate model (CanCM3) and its atmospheric component used in HFP2, called a third generation general circulation model (GCM3) will be used. This will allow us to compare the ensemble products of coupled and uncoupled models, with same atmospheric component, to rigorously explore the impact of coupling on MJO predictability.

1.4. Objectives and Outline

The study is carried out using global climate model data sets produced at different climate centers in the world, namely: the Canadian Centre for Climate Modeling and Analysis (CCCma) coupled (CHFP2) and uncoupled models (HFP2) ensemble forecasts, Development of a European Multi-model Ensemble for seasonal to Inter-annual Prediction (DEMETER) five coupled ensemble forecasts, and ECMWF ENSEMBLES three coupled ensemble forecasts. The main emphasis of the thesis to explore the prediction uncertainty at time scales from that of weather to seasons, and to measure the useful information provided by the predictions. A newly developed set of theoretical tools will be used to explore some essential issues related to PNA and MJO predictability at various time scales, including dominant precursors of forecast skill and degree of confidence that can be placed in an individual prediction.

The objective of this research is to explore the actual and potential predictability of PNA and MJO at various time scales, using newly developed advanced statistical techniques. In addition, the source of PNA and MJO will be identified at monthly (MJO&PNA) and
seasonal time scales (PNA only). The specific objectives of this research are:

i) The actual and potential predictability of PNA and MJO will be evaluated, through the use of retrospective ensemble forecasts of state of the art global coupled and uncoupled models. This will provide an opportunity to compare the prediction skill of coupled and uncoupled models.

ii) Multimodel Ensemble Mean (MME) skill will be assessed and an appraisal of the prediction skill between individual models and their MME will be carried out. It is often argued the MME offset the individual model errors and enhances the predictability limit.

iii) The advantage of time averaged prediction skill will be assessed over daily unfiltered time scales both for PNA and MJO. For PNA, the time average of daily prediction skill will be studied for weekly, bi-weekly and monthly time scales and for MJO, the daily data will be studied for 3-day, 5-day and weekly time scales. As already mentioned, time averaging reduces the weather noise so that a predictable signal beyond two weeks is more visible.

iv) Potential predictability at various time scales will be studied mainly using information theory to derive robust measures to estimate the uncertainly in PNA and MJO predictions and to identify the source of predictability at various time scales.

v) In addition, some good and reliable information-based measures will be identified at various time scales, both for PNA and MJO and their relationships with actual prediction skill measures will be explored.

vi) Predictable Component Analysis (PrCA) will be used to decompose the predictability into various patterns that explain different contributions to the total predictability.
PrCA is especially useful if predictability is dominated by few patterns, which allows us to focus on few predictable structures instead of various structures that are not predictable.

The thesis comprises four main chapters (chapter 2 to chapter 5) to achieve the aforementioned goals. In chapter 2, detailed information about the ensemble forecasts from different centers will be given. Moreover, different techniques and measures to estimate the actual and potential predictability will be presented. Chapter 3 will be devoted to the prediction skill of PNA at various time scales using actual and potential prediction skill measures and the source of PNA potential predictability at monthly to seasonal time scale. In chapter 4, MJO prediction skill will be evaluated and emphasis will be placed on the difference and similarities of MJO prediction skill between coupled and uncoupled models. The focus of chapter 5 will be to draw the connection between PNA and MJO prediction skill in terms of time average prediction skill, to get an insight of which potential predictability measures is reliable (or not consistent) to estimate the predictability both at subseasonal to seasonal time scale, a quite interesting research question to explore for the climate prediction community.
Chapter 2: Models, Data, and Experimental Framework

2.1. Ensemble prediction products and observed Data for PNA study

Three ensemble prediction datasets were used to study PNA predictability including HFP2, DEMETER and ENSEMBLES stream 2. HFP2 is a collaborative project among some Canadian universities and government laboratories, whose objective is to test the extent to which the potential predictability of mean seasonal conditions could be achieved (Kharin et al. 2009; Derome and Coauthors 2001). This product includes ensembles of four global atmospheric models: the second generation atmospheric general circulation (AGCM2; McFarlane et al. 1992) and third generation general circulation model (AGCM3; Scinocca et al. 2008), a reduced resolution version of medium-range weather forecast global spectral model (SEF, Ritchie 1991), and the Global Environmental Multiscale (GEM) model (Lin et al. 2008). Each model produces an ensemble of 10 parallel integrations of four-month duration for the period of 1969-2003 from the beginning of each month. The integrations are initialized from the NCEP/NCAR reanalysis (Kalnay et al. 1996) lagged at 12-hour intervals prior to the initial time of prediction (ITF). That is the first member is initialized 12 hours before the ITF, the second member is initialized 24 hours prior to the ITF etc., and the 10th member is initialized 5 days prior to the ITF. The oceanic forcing used in the atmospheric prediction is from the persistent prediction of the global sea surface temperature anomaly (SSTA), namely that, the SSTA of the initial month of the prediction persisted through the entire prediction period, and this was superimposed onto the climatological SST of the target month of prediction.

DEMETER stands for Development of a European Multi-model Ensemble for seasonal to
Inter-annual Prediction. It includes seven state-of-the-art global coupled ocean-atmosphere models which produce 6 month forecasts starting from February 1st, May 1st, August 1st and November 1st of each year, over a common period 1980-2001 (Palmer et al. 2004). Four SST perturbations are added to and subtracted from initial conditions of the hindcast to represent the uncertainty in the SST. These SST perturbations are based on two quasi-independent SST analyses. Each hindcast has been integrated for 6 months and comprises an ensemble of 9 members. Two models, INVG and MPI, which have very poor skill over the PNA region, were excluded in this study. Thus, the multi-model of DEMETER is formed by merging 5 models having 45 ensemble members.

The ENSEMBLES seasonal forecasts used in this study are from its stream 2 experiment. In this experiment, five globally coupled general circulation models were run, including European Centre for Medium Range Weather Forecasts (ECMWF), the Leibniz Institute of Marine Sciences at Kiel University (IFM-GEOMAR), Meteo France (MF), the UK Met Office (UKMO) and the Euro-Mediterranean Centre for Climate Change in Bologna (CMCC-INGC). Each ensemble comprises 9 runs initialized with different sets of ocean reanalysis generated from wind stress and SST perturbations. For each year, 7-month long seasonal forecasts starting on 1st of February, May, August and November have been initialized which covers the 46-years 1960-2005 (Weisheimer et al. 2009). The November initial condition is extended to 14-month duration of forecast except for CMCC-INGV. Among five models, IFM-GEOMAR and CMCC-INGV were not used due to their relatively poor performance than other models.

2.2. Ensemble prediction products and observed Data for MJO study
To study the MJO forecast skill, the ensemble forecasts of one atmosphere only model and one coupled model will be used. The atmospheric model is the third generation of the general circulation model (GCM3) which was used in seasonal forecast experiment conducted for HFP2. As GCM3 belongs to the HFP2 product, the details of which were provided above, only coupled model forecasts will be briefly mentioned. The coupled model CanCM3 belongs to Coupled Historical Forecast Project (CHFP2). The atmospheric component of the CanCM3 is GCM3 whereas CanOM4 served as the ocean model in CanCM4 (Merryfield et al. 2013). As the atmospheric component of CanCM3 is same i.e. GCM3 which has been described above (it should be emphasized here that GCM3 used in CanCM3 has different initialization of atmosphere than GCM3 used in HFP2), we will briefly mention the ocean component of CanCM3 here. The vertical resolution of the CanOM4 is 40 levels ranging from 10m near ocean surface to >300m at abyssal depths. The coupling between atmospheric and oceanic components of CanCM3 is done on daily basis, where ocean component received daily mean surface heat, freshwater and momentum fluxes computed by the atmospheric component, and after one day, the daily mean Sea Surface Temperature (SST) are updated are substituted back to atmosphere. There are precisely six ocean grid cells under each atmospheric grid cell whereas beneath each atmospheric cell, the surface is entirely ocean or land. This configuration simplifies the model physics and interpolation of coupling fields but it enforces relatively low atmospheric model resolution. For more details refer to Merryfield et al. (2013).

2.3. Measures of Skill for PNA Predictability

2.3.1. Actual Measures of Predictability

In general, there are two groups of predictability measures: actual measures that make use of
observations, and potential predictability measures that do not make use of observations. The actual skill of ensemble mean prediction over 22 years is measured by anomaly correlation (r) and root mean square error (RMSE), defined below

\[
 r(t) = \frac{\sum_{i=1}^{N} (T_i^p(t) - \mu^p)(T_i^o(t) - \mu^o)}{\sqrt{\sum_{i=1}^{N} (T_i^p(t) - \mu^p)^2 \sum_{i=1}^{N} (T_i^o(t) - \mu^o)^2}}
 \] (2.1)

\[
 RMSE(t) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} [T_i^p(t) - T_i^o(t)]^2}
 \] (2.2)

where \( T \) is the index of interest, \( t \) is the prediction lead time (varies with time scale), superscript \( p \) is the predicted index, \( o \) is the observed counterpart. The \( \mu^p \) is the mean of the forecasts, \( \mu^o \) is the mean of observations and \( N \) is the number of initial conditions.

RMSE indicates the mean ‘distance’ between forecasts and observations over the verification time period. It usually increases with lead time and asymptotically approaches a “saturation” value. The saturation value is equivalent to the mean difference between two randomly chosen fields from the system (e.g., DeSole 2004).

2.3.2. Potential Predictability Measures

The measures of potential predictability include traditional metrics such as signal-to-noise ratio (SNR), ensemble mean, ensemble spread and information-based metrics. As it is
straightforward to calculate ensemble mean and ensemble spread from ensemble predictions so here only information-based potential predictability measures will be introduced. The idea behind the information-based measures is to use the difference between two entropies, climatological entropy and predictive entropy, to quantitatively measure extra information that the prediction brings, as expressed by Relative Entropy (RE). The average of RE over all initial conditions is called mutual information (MI), quantifying the overall predictability of dynamic systems.

a) RELATIVE ENTROPY

Relative Entropy or Kullback-Leibler divergence is a measure used to calculate the difference between two distributions, such as the difference between climatological and predictive distribution, as used in this study. If $q(v)$ denotes the climatological distribution for random variable $v$ and $p(v|\Theta)$ denotes for forecast given the initial or boundary condition of $\Theta$, then for a continuous set of states, relative entropy ($RE_A$) is defined as

$$RE_A = \int p(v|\Theta) \ln \left( \frac{p(v|\Theta)}{q(v)} \right) dv$$

(2.3)

$q(v)$ is also interpreted as prior distribution (climatological distribution) and $p(v|\Theta)$ is described as posterior distribution (forecast distribution). For a Gaussian PDF approximation, $RE_A$ can be calculated exactly in terms of predictive and climatological variances and a difference between their means. In that case, the expression for relative entropy is given by (Kleeman 2002) as

$$RE_A = \frac{1}{2} \left[ \ln \left( \frac{\text{det}(\sigma_p^2)}{\text{det}(\sigma_q^2)} \right) + \text{tr} \left( \left( \sigma_p^2 \right)^{-2} \right) - n + (\mu_p - \mu_q)^T (\sigma_q^{-2})^{-1} (\mu_p - \mu_q) \right]$$

(2.4)
Here, $\sigma_q$ and $\sigma_p$ are the climatological and ensemble variances, respectively, while $\mu^e$ and $\mu^q$ are the ensemble and climatological mean of the system, and $n$ is the number of degrees of freedom. As shown in Eq. (2.4), RE$_A$ is composed of two components: (i) a reduction in climatological uncertainty by the prediction (first three terms of Eq. 2.4, also called Dispersion component) and (ii) a difference in the predictive and climatological means (last term of Eq. 2.4 which is also called Signal component). These components are interpreted as components of utility of prediction (Kleeman 2002). A large value of RE$_A$ indicates that more information that is different from the climatological distribution is being provided by the prediction, which can be interpreted as a more reliable prediction.

PI is defined as the entropy difference between posterior (e.g. prediction) and prior (e.g. climatology) distributions. If we denote posterior distribution as $p$ and prior distribution as $q$ then PI can be written as

$$PI = -\int q(x) \ln(q(x)) \, dx - \int -p(x) \ln(p(x)) \, dx$$  \hspace{1cm} (2.5)$$

The first term of Eq. 2.5 denotes the absolute entropy of the posterior or predicted distribution measuring the uncertainty of the observations and associated prediction, whereas the second term measures the entropy of the prior or climatological distribution, estimating the uncertainty of a prior time when no extra information is available. Thus large PI indicates the decrease in uncertainty in a predicted distribution because useful information is being provided by a prediction, which means the prediction is likely to be reliable.

PP is defined in the same way PI i.e. large PP corresponds to small uncertainty and vice versa. For a univariate case with climatological mean of zero, the RE, PI and PP can be written mathematically in terms of prediction and climatological covariance matrices as
W. Younas: Evaluation of Climate Predictability

(DelSole 2004):

$$RE_A = \frac{1}{2} \left[ \ln \left( \frac{\sigma_q^2}{\sigma_p^2} \right) + \frac{\sigma_p^2}{\sigma_q^2} - 1 + \frac{\mu_p^2}{\sigma_q^2} \right]$$ (2.6)

The mathematical representation of Signal Component (SC) and Dispersion Component (DC) can be written as

$$DC = \frac{1}{2} \left[ \ln \left( \frac{\sigma_q^2}{\sigma_p^2} \right) + \frac{\sigma_p^2}{\sigma_q^2} - 1 \right]$$ (2.7)

$$SC = \frac{1}{2} \left( \frac{\mu_p^2}{\sigma_q^2} \right)$$ (2.8)

$$PL = \frac{1}{2} \ln \left( \frac{\sigma_q^2}{\sigma_p^2} \right)$$ (2.9)

$$PP = 1 - \left( \frac{\sigma_q^2}{\sigma_p^2} \right)$$ (2.10)

Here, $\sigma_q$ and $\sigma_p$ are the climatological and predictive covariance matrices, respectively, while $\mu_q$ and $\mu_p$ are the climatological and predictive mean state vectors.

Another measure is Mutual Information (MI), which is defined as the average of $RE_A$, over all initial conditions (e.g., DelSole 2004; Yang et al. 2012). Interestingly, there is a theoretical relationship between MI based anomaly correlation skill ($AC_{MI}$ hereafter) and the conventional STR (signal to total variance ratio) based counterpart ($AC_p$ hereafter), if the prediction and climatological PDFs are Gaussian, as below (Yang et al. 2012; Cheng et al. 2011) (also see the APPENDIX).

$$AC_p \leq AC_{MI}$$ (2.11)
AC_p can also be estimated using a "perfect model" scenario: taking one model realization (an ensemble member) as a true observation and measuring the forecast skill using the mean of the remaining ensemble members. The equality holds in Eq. 2.11 when ensemble variance is not a function of initial conditions; otherwise, the AC_m_l measures more predictability than AC_p. This is because the AC_p measures a linear relationship between prediction (ensemble mean) and ‘observation’ (an ensemble member), which underestimates the true potential predictability that is statistically defined as the coherence between prediction and initial (boundary) conditions. On the other hand, AC_m_l measures the statistical dependence, both linear and nonlinear, between prediction and observation (Yang et al. 2012; Tang et al. 2013).

For Gaussian distribution the Mutual Information (MI) is also related to a theoretical correlation coefficient r as (DelSole 2004):

\[
MI = -\frac{1}{2} \ln(1 - r^2)
\]

(2.12)

2.3.3. Predictable Component Analysis (PrCA)

PrCA is analogous to the traditional principle component analysis (PCA), which decomposes the total variance into different structures (eigenvectors), whereas PrCA decomposes the total predictability into various patterns that explain different contributions to the total predictability. PrCA is especially useful if predictability is dominated by a few patterns, which allows us to focus on a few predictable structures instead of various structures that are not predictable.

In practice, the PrCA analysis is performed in the state space of truncated PCA modes, so that the covariance matrix that is used to solve the eigenvalue equation is of full rank. This
problem arises due to the fact that practically, the number of grid points should be much larger than the number of total samples in climate studies. In this study, the first 30 PCA modes for truncation were used. A detailed discussion of the PrCA algorithm can be found in Delsole and Tippett (2007) and Tang et al. (2013).

2.4. Measures of Skill for MJO Predictability

The potential predictability measures used to explore the MJO prediction skill and predictability will be same as used for PNA predictability. However, the actual skill measures e.g. correlation and RMSE are modified due to the fact that the MJO index is described using first two PCs instead of one. For MJO, these PCs are often termed as Real Time Multivariate MJO series (RMM1 & RMM2). The purpose of using RMM1 and RMM2 is that MJO is mainly described by the first two leading EOF modes, to represent the MJO eastward movement. These RMM indices can be used to define the MJO bivariate index both for observation and forecast. Following Rashid et al. 2011 and Lin et al. 2008, we can mathematically formulate the actual skill measures in terms of correlation (Corr) and Root Mean Square Error (RMSE) to calculate MJO bivariate index skill as:

\[
Corr(t) = \frac{\sum_{t=1}^{N}[a_1(t)b_1(t, \tau) + a_2(t)b_2(t, \tau)]}{\sqrt{\sum_{t=1}^{N}[a_1^2(t)]} \sqrt{\sum_{t=1}^{N}[b_1^2(t, \tau) + b_2^2(t, \tau)]}}
\]

\[
RMSE(t) = \sqrt{\frac{1}{N} \sum_{t=1}^{N}[a_1(t) - b_1(t, \tau)]^2 + [a_2(t) - b_2(t, \tau)]^2}
\]

Where \(a_1(t)\) and \(a_2(t)\) are the verification (observation) RMM1 and RMM2 at time \(t\), and
$b_1(t, \tau)$ and $b_2(t, \tau)$ are the respective forecasts a time $t$ for a lead time of $\tau$ days. $N$ is the total number of forecasts from 1979-2001.
Chapter 3: PNA predictability at various time scales

In this study, the PNA predictability will be systematically explored over time periods ranging from days to months using actual and potential predictability measures. The purpose of this study is to comprehensively analyze the PNA predictability and its variation at different time scales, using long-term ensemble predictions from multiple global coupled and uncoupled climate models, where the coupled models mean coupled atmospheric and oceanic general circulation models (also called one-tier forecasts in seasonal prediction community), whereas the uncoupled models mean the atmospheric circulation models driven by persistent boundary (ocean) forcing (two-tier forecasts). This allows us to derive statistically robust, generalizable and realistic conclusions. To achieve this, the analysis is done using ensemble predictions of 500 mb geopotential height from multiple model ensembles including CCCma (the Canadian Centre for Climate Modeling and Analysis) HFP2 product of four different global models, the ECMWF DEMETER product from five global coupled models, and the ECMWF ENSEMBLES product from three global coupled models. The multiple-model ensemble (MME), or super-ensemble, of each product was used in this study. It has been argued that MME is usually better than a single model ensemble (SME) since the uncertainties associated with the different model frameworks can offset each other and can be diminished by a large number of ensemble members (e.g., Krishnamurti et al., 1999, 2000; Palmer et al., 2004; Yan and Tang 2012).

PNA predictability study focuses on the 500 mb geopotential height ensemble prediction for the common 21-year period of 1980-2001. The hindcasts are initialized, respectively, on February 1, May 1, August 1, and November 1, and last 4 months. For validation purposes,
the NCEP-National Centre for Atmospheric Research (NCAR) reanalysis dataset (Kalnay et al. 1996), at horizontal resolution of 2.5×2.5, was used.

Originally the PNA was defined by the four-point teleconnection patterns (Wallace and Gutzler, 1981). However, the classic definition seems insufficient in describing the large scale spatial-temporal structure of PNA pattern, thus many researchers prefer to use the leading modes of PCA (principle component analysis) of the northern hemisphere 500mb geopotential height anomalies to define PNA such as Thompson and Wallace (1998, 2000). In this study we follow this strategy -- , applying the rotated PCA onto the observed (NCEP) monthly mean 500mb geopotential height anomalies over the domain between the equator to 87.5°N from 1950-2000, to define the PNA. The definition is the same as one that has been used operationally at the NOAA Climate Prediction Center (Chen and Van Den Dool 2003).

Shown in Fig. 3.1 are the first two modes of the PCA, clearly indicating that they well characterize the NAO (North Atlantic Oscillation) and PNA structure, respectively. The PNA pattern derived from monthly data (mode 2) is very similar to those from the data of other time scales such as weekly or daily data (Johansson 2007), thus for simplicity, we use the monthly PNA pattern for all time scales in this study. The predicted PNA index is obtained by projecting ensemble prediction of 500mb geopotential height anomalies on the PNA pattern for all time scales from days to month.

3.1. Properties of ensemble systems

As a starting point, the ensemble spread and the root mean square error of the ensemble mean prediction (Ensemble RMSE hereafter) are examined for all the individual models and the MME. Ensemble spread and the Ensemble RMSE are considered to be the most basic
quantities used to evaluate an ensemble prediction system. The former, to some extent,
diagnoses the sensitivity of model error growth to initial uncertainty, whereas the latter
directly measures the accuracy of the ensemble mean prediction against observations. The
forecast error growth from initial uncertainty would saturate after some time due to the
nonlinear nature of the atmospheric system, which, on average, is about two weeks on daily
time scales (Lorenz 1969). One of the motivations behind ensemble forecasting is to estimate
the forecast uncertainty using ensemble-derived variables such as ensemble spread, a
quantity often used in weather forecasts. Stensrud et al. (1999) argued that forecasts with
larger ensemble spread are probably less certain than forecasts with smaller ensemble spread.
This is based on the perfect model assumption which is usually made for potential
predictability studies. Under this assumption, one arbitrary realization of the forecast
distribution is used as the hypothetical observation instead of real observation. In this case
the actual observation is statistically indistinguishable from members of the forecast
ensemble, and the ensemble spread is equivalent to the RMSE (Begnsston et al. 2008; Cheng
et al. 2011). Thus, for daily time scales, the upper limit of weather predictability defined by
Lorenz (1969) depends on the saturation of RMSE which in turn is determined by the mean
climatological variance (Shukla and Kinter 2006).

The variations of the ensemble spread and Ensemble RMSE for all individual models and the
MME of each ensemble product, as a function of lead time, for daily time scale are shown in
Fig. 3.2. It should be noted here that RMSE values were normalized. The climatological
variance obtained using observations during the period 1980-2001 is also presented as a
reference. Among the four uncoupled models of HFP2, GEM, GCM3 and GCM2 have
sufficient spread at initial lead times as their ensemble spread and Ensemble RMSE lie close
to each other for most of the lead times. For SEF, the ensemble spread is far from the Ensemble RMSE, suggesting large biases in SEF as found in Kharin et al. (2009). Regarding the coupled models, only ECMWF from ENSEMBLES has an ensemble spread close to its Ensemble RMSE. In all the other coupled models, either from ENSEMBLES or DEMETER, the ensemble spread is considerably smaller than the Ensemble RMSE at shorter lead times. This is probably due to insufficient initial perturbation (Bengtsson et al. 2008). A typical behavior in many current ensemble forecast systems is the underestimation of the ensemble spread, since many possible sources of model-related uncertainty, such as parameter uncertainty, are often not well considered (Jolliffe and Stephenson 2003).

The MME of HFP2 (MME_H hereafter), DEMETER (MME_D hereafter) and ENSEMBLES (MME_E hereafter) apparently has a better relationship between Ensemble RMSE and ensemble spread than individual models. The central argument for the MME superiority over the single model ensemble has two points: 1) the MME takes a holistic consideration of uncertainties from both the initial conditions and the model uncertainties (e.g., Palmer and Shukla, 2000; Palmer et al. 2004); and 2) a lack of understanding of atmospheric behavior could possibly be offset by different models in the MME. Thus, most of the analyses in this study focus on MME, except that individual models are required for the purpose of skill validation.

3.2. PNA actual skill at different time scales

In this section, the actual prediction skill of individual models at different time scales will be evaluated. The bootstrap method is used to perform the statistical significant test of correlation skill, instead of the student's t test that requires the effective number of degrees of
freedom, a difficult quantity to estimate for a sample with a temporal average. The bootstrap experiment was designed as follows: 1) given the prediction lead time, the observation and prediction are paired based on the same prediction target time, i.e., constructing the sample of observation-prediction pair for the entire period from 1980-2001; 2) randomly choosing 95% of the samples of the prediction-observation pair, and calculating their correlation coefficient; 3) repeating 2) 1000 times to obtain 1000 correlation coefficients, whose standard deviation is used as a threshold value at a given confidence level (i.e., the error bar in figures).

The skill in predicting the PNA index as a function of lead time is shown in Fig. 3.3 for different models at different time scales. The persistence skill is also provided as a reference. As can be seen in this figure, most of the models performed better than persistence, with modest improvement at the weekly time scale (i.e. correlation coefficient at lead 1 week (Fig. 3.3c) were compared to mean of daily correlation coefficients (Fig. 3.3a) and so on). The increase in skill is significant at the bi-weekly time scale (Fig. 3.3e), compared to mean of daily values (Fig. 3.3a). (e.g. first lead time at the bi-weekly time scale is compared with mean of daily correlation coefficients from 1-14 days, and so on). For monthly time scale, the correlation skill at the lead time of 1 (month) ranges from 0.46 to 0.65 among all models, which is meaningfully higher than the mean correlation skill of daily prediction over the first 30 days (Fig. 3.3g). Some monthly predictions have significant correlation beyond one season. These results are consistent with a widely recognized concept that predictability can be enhanced by taking the spatial or temporal average (e.g., Lorenz 1969; Van den Dool and Saha 1990). This is because large-scale variability does not change rapidly and the growth of initial errors is relatively slow for low-frequency components.

Fig. 3.3 shows that most of the models have RMSE smaller than persistence at all lead times,
over multiple time scales. The model prediction skill is considered to be useful only if it is better than climatology; thus, the bottom limit of predictability occurs when the RMSE approaches the climatological spread (Kimoto et al. 1991). As can be seen in Fig. 3.3b, most of the models at daily time scales reached this limit around day 12. The bootstrap experiment found that correlation values less than 0.2 were not significant. Therefore, correlation coefficient at 0.2 can define the limit of predictability. It is notable that the predictability limit defined in this way is longer compared with the predictability limit defined in terms of RMSE. This difference might be attributed to the significance level used for the correlation skill, i.e. 95%. By changing the level to 99% for correlation, the RMSE and correlation have the almost equivalent predictability limit. For weekly time scales, the predictability limit remained comparable to daily time scale limit (Fig. 3.3d) whereas the improvement was substantial at the bi-weekly time scale, where the predictability limit increased to 4-6 weeks (Fig. 3.3f). The impact of slowly varying boundary forcing was prominent at monthly time scales, as the skill in terms of RMSE lingered until two months (Fig. 3.3h).

The prediction skill for MME of the three ensemble products: HFP2, DEMETER and Ensembles, is shown in Fig. 3.4, respectively. The overall features of MME skill can be stated as: all the MMEs beat the persistence skill at almost all lead times over all time scales, demonstrating the advantage of multiple models since some individual models performed worse than persistence at longer lead times as shown in Fig. 3.3. Using bootstrap experiments, the prediction skill of MME was found to be significant for the first 20 days at a daily scale, 3-4 weeks for the weekly time scale, 1-2 months for bi-weekly and 3 months for monthly time scales, indicating that the forecast skill increased with the increase of averaging time.
The RMSE skill of MME is presented in Fig. 3.4b, d, f, and h over different time scales. Compared with individual models, the RMSE of MME reached error saturation (i.e., RMSE = climatological spread) at longer lead times. Similar to the correlation skill, the RMSE skill reflects the increase in predictability with the time average, as indicated by the lead time at which error growth saturation is reached. For example, the time limit of predictability, in terms of the error growth saturation, increased from 14 days for daily scale to 2 months for monthly scale.

The increase in predictability with averaging time scales can be further demonstrated by comparing correlation coefficients of two different scales at an equivalent lead time. For example, in Fig 3.5, the average correlation coefficients of daily scale over day 1 to day 7 (or day 8 – day 14) are compared with the correlation skill at the 1-lead time (or 2-lead times) of weekly scale. It can be observed that the weekly skill is always larger than, or at least equivalent to, the value of the weekly average (the average of daily skill over 7 days) weekly scale. For bi-weekly scale, the enhancement in prediction skill is substantial even for 1-lead time (e.g., the correlation coefficient was over 0.73-0.8 for all three MMEs) as compared to the mean of corresponding daily values of the first 14 days (the correlation coefficient was 0.73 for MME_H, 0.8 for MME_D and 0.8 for MME_E). A similar approach is applied to compare monthly time scale correlation at different lead times, with mean of daily correlation coefficients. The enhancement in correlation skill continued at monthly time scale where the correlation was significant for lead times up to three months.

Thus, the actual prediction skill from the three ensemble products show that the skill increases with time averaging. Tribbia and Baumhefner (1988) argued that time averaging impacts predictability in two ways. First, it reduces the phase decorrelation rates and
alleviates high frequency noise (noise variance), which improves predictability; and second, time averaging reduces the climatological variance (e.g., total variance) which opposes the greater predictability trend. Thus, a net effect of time averaging is the competition between both contributions to predictability. Following Tribbia and Baumhefner (1988), the decrease in climatological variance and the variation of decorrelation with time averaging are estimated and compared with the correlation skill at different time scales (Fig. 3.6). Here, only MME_H is randomly selected for comparison due to the similarity of these models’ prediction skill. The impact of time averaging is clearly observable in Fig. 3.6. For example, weekly, bi-weekly and monthly time scale forecasts are more predictable than daily forecast skill in Fig. 3.6a, whereas the decrease in climatological variance and autocorrelation (persistence) increase with time average as shown in Fig. 3.6b and 6c respectively. A comparison among Fig. 3.6a, b, and c, reveals that the increase in the length of decorrelation (persistence skill) overweights the decrease in climatological variance so that the former dominates the predictability. Thus, the increase in predictability with averaging time scale, even in the presence of a decrease in climatology variance, is due to two reasons: i) the averaging lessens the noise, thereby increasing the signal; ii) the role of slowly varying boundary forcing becomes more and more influential with time average.

3.3. PNA potential predictability

3.3.1. MI based predictability

Fig. 3.7 shows AC_MI and AC_p for ENSEMBLES and HFP2 for different time scales. Two obvious features can be observed in this figure. First, the coupled model ENSEMBLES has significantly better skill than the uncoupled model HFP2 at multiple time scales, in contrast
with the comparison of actual skills (Fig. 3.4). Second, the $AC_{MI}$ is always larger than $AC_{P}$, as indicated by Eq. 2.11, suggesting that the conventional predictability measure of signal to noise ratio (SNR) underestimates the potential predictability. These two features will be discussed further after we examine the variation in potential predictability with the averaging time scale, where ENSEMBLES and $AC_{MI}$ as the target of analysis are used due to their better representation than HFP2 and $AC_{P}$.

Similar to actual skill, the potential predictability of PNA also increased with the time scale, as shown in Fig. 3.7. The improvement in potential predictability over weekly time scale, (Fig. 3.7c,d) is significant compared to daily time scale (Fig. 3.7a,b). The enhancement in predictability persisted at the bi-weekly time scale (Fig. 3.7e,f) and monthly time scale (Fig. 3.7g,h), where the potential predictability is meaningful for all four months. Table 3.1 shows the maximum lead time that remains at a correlation skill of 0.5\(^2\) for all time scales. Table 3.1 show that $AC_{MI}$ remained significant until 40 days for MME_E and 28 days for MME_H at daily time scale. In contrast with actual skill, the improvement in skill is also quite substantial even at weekly time scales (16 weeks for MME_E and 5 weeks for MME_H).

### 3.3.2. Comparison between MI and SNR

We also explored predictability using the SNR, where signal and noise are estimated using Rowell’s scheme (Rowell 1998). The results show that the predictability by SNR has features similar to those by MI; namely, that the SNR increased with time averaging.

Fig. 3.7 shows the $AC_{MI}$ and $AC_{P}$ as a function of lead time over different time scales. The

\(^2\) The correlation value of 0.5 is partially arbitrarily set, but is often used as a threshold of useful prediction skill in the seasonal prediction community.
most prominent features here are the skill decreases with lead time, and the $A_{CM}$ is higher than $A_{CP}$. The former is in reasonable agreement with the general conclusion that predictability declines with lead time in chaotic or stochastic dynamical systems whereas the latter is consistent with the theoretical formula (Eq. 2.11). As shown in Fig. 3.7, at the daily time scale, both $A_{CM}$ and $A_{CP}$ were close to each other for the first few lead times (Fig. 3.7a and b), suggesting that the ensemble variance does not differ much from one forecast to another for these lead times. For subsequent lead times, $A_{CM}$ was significantly larger than $A_{CP}$, simply because the ensemble variance was no longer approximately constant after sufficient development with lead times. The same phenomena can be observed at weekly and bi-weekly time scale. Over monthly time scales, even at a 1-month lead time, $A_{CM}$ was higher than $A_{CP}$ (Fig. 3.7g, and h).

The comparison of actual and potential skill using MME_E and MME_H at different time scales is shown in Fig. 3.8. Here potential skill is measured using $A_{CM}$. It can be observed that the potential prediction skill is higher than the actual skill. Table 3.1 also shows the skill limit for both actual and potential predictability. It is clear that the potential predictability limit is quite large compared to actual skill at multiple time scales for both MME_H and MME_E. It is not surprising since the potential predictability represents an upper limit of prediction skill that a perfect model can achieve. The fact that the potential predictability is higher than actual skill suggests that there may be a lot of room to improve PNA prediction skill.

It should be emphasized here that the potential predictability may be smaller than the actual skill if the model imperfectly estimates the noise variance. For example, Batté and Deque (2011) found that if the ensemble is over-dispersive and ensemble variance is overestimated,
the potential predictability could be underestimated compared with the actual skill. On the other hand, if model error only exists in the ensemble mean (signal) as random error, the potential predictability is greater than the actual skill, as proven by Sardeshmukh et al. (2000). Over all, there are three kinds of errors impacting the atmospheric predictability: nonlinear chaos, model random error and model systematic error. The difference between actual and potential predictability is due to inclusion of these three factors in estimating the actual skill.

Another possible reason for the difference between actual and potential skill shown in Fig. 3.8 is due to the error of Gaussian assumption used in calculating potential predictability. To examine this, we estimated the probability distribution function (PDF) for all lead times and all time scales for HFP2 and ECMWF, respectively. The results show that the Gaussian assumption is always approximately held for any case. The One-sample Kolmogorov-Smirnov test further validated the normality of each distribution at each time scale for different lead times (not shown). Shown in Fig. 3.9 are the PDFs of different randomly chosen lead times for the weekly time scale. Fig. 3.9 also shows that there is no essential difference of Gaussian approximation between coupled and uncoupled models, indicating that the difference of potential skill between coupled and uncoupled models is not due to their disparity in assuming Gaussian approximation. This issue will be discussed in detail in next subsection.

3.4. Predictability of coupled and uncoupled models

It is intriguing to explore the difference in predictability between coupled and uncoupled models. First, the difference of their actual prediction skill is examined. Using bootstrap test,
it is found that their difference in actual skill, shown in Fig. 3.4, is not statistically significant as indicated by the sampling error bars exceeding these differences. This is consistent with the findings of Johansson (2007), where he mentioned that coupling is not relevant in enhancing the prediction skill of PNA.

Fig. 3.10 compares the difference of the potential predictability $AC_{M1}$ between coupled model (MME_E) and uncoupled model (MME_H) for different averaging times. Again, the bootstrap method is used to estimate the extent of uncertainty due to sampling error, as shown by the bars in this figure. A significant difference beyond the sampling uncertainty is witnessed in potential skill between MME_E and MME_H for all time scales as shown in this figure, in contrast with actual skill results. Thus, one compelling question is why coupling does not lead to a better PNA prediction skill although it has higher potential predictability? To shed lights on this issue, the prediction skill of SSTA (70S-70N) is calculated from MME_H and MME_E at monthly time scale. The SSTA prediction is provided by the persistent scheme for uncoupled models (MME_H) and the oceanic component of the coupled models (MME_E). Fig. 3.11 shows that the MME_E SST prediction skill is better than MME_H (persistence) in many areas especially in the tropics and over North America (PNA index). With the increase in lead time, MME_E has higher skill for SSTA prediction than MME_H. Thus, the reason that coupled models do not lead to better PNA prediction than uncoupled models is probably due to the bias of atmospheric models, which can cause the coupling inconsistent with the observation, resulting in the skill in coupled model is not necessarily better than that of uncoupled model. One may understand the prediction of the MME_E and MME_H like two atmospheric model runs: one forced with persistent SST, the other forced with coupled model SST. The coupled model SST is
better than the persistent SST (uncoupled model). If the atmospheric model was perfect, the atmosphere model would have more realistic forcing (prediction) when it is forced by the coupled model SST. However if the atmospheric model has bias, this conclusion may not held. It should be noted that the fact that the coupled model SST is better than the persistent SST prediction of the uncoupled model does not conflict with the above argument since the SST prediction forced by the biased atmosphere may still be better than the persistent SST prediction. It should be emphasized here that we are not considering any error in the coupling schemes and assume that the error source is just due to the atmosphere.

3.5. Temporal variation in PNA predictability and its source

In the last section, the overall potential predictability is discussed using mutual information. In this section, the emphasis will be put on the temporal (e.g., interannual or decadal) variation in potential predictability and its source. Shown in Fig. 3.12 is the averaged $RE_A$ over the effective lead times for all individual predictions by ENSEMBLES. Here, the effective lead time is equivalent to the maximum lead time beyond which the prediction skill is not significant and is determined by bootstrap experiment. A striking feature in Fig. 3.12 is that there are temporal variations in potential predictability of PNA in terms of $RE_A$, where large $RE_A$'s are mainly located in fewer predictions, such as 1982-83 and 1997-98 (strong ENSO events). For most of other predictions, $RE_A$ is small or exhibiting small variations with initial time. Next, we will examine what determined these variations in $RE_A$, at different time scales.

As described in chapter 2, $RE_A$ is the sum of DC and SC (Eq. 2.6). The variations in both signal and dispersion components for MME_E over different time scales are shown in Fig.
3.13, for the period 1980-2001. As can be seen in Fig. 3.13, the signal component is large in the predictions from 1980-1985 and 1995-2001. In between, the signal is low (1985-1995). The variation feature might be related to variation in the tropical Pacific SST forcing. For example, the two periods with large signal components have the strongest El Nino events (1982/83 and 1997/98). To explore the relationship between SST forcing and the potential predictability, lead-lag temporal correlation between PNA index and SSTA is calculated for each grid at monthly time scale over the region 30N-30S and 150E-90W, using the periods of high signal component, i.e., 1980-1985 and 1996-2001, as shown in Fig. 3.14.

Fig. 3.14 (a, c, e and g) shows the correlation pattern between the PNA index and predicted SST (ensemble mean prediction of SST) for ENSEMBLES. It can be observed that the significant correlation mainly appears in the tropical central eastern Pacific (ENSO region). The correlation patterns resemble the typical ENSO pattern, implying that the ENSO is a major source of PNA potential predictability. This result also favors the notion that the role of tropical SST forcing is to amplify the atmospheric variability such as PNA, by impacting its signal (ensemble mean) (e.g. when forcing is high, the PNA signal is strong, and the prediction skill is high, and vice versa). In particular, the maximum correlation is found when SST was leading by two and three months (Fig. 3.14e and g) (e.g., 0.6-0.8 in ENSO region, which is significant at 95% confidence level). A high correlation is also found when the PNA observed index is lagged-correlated with observed SSTA by 2-3 months (not shown). The lagged correlation of 2-3 month of ENSO and PNA has also been reported by Munoz et al. (2010). A similar correlation pattern is also obtained for MME_H where the persistent prediction of SSTA was correlated with PNA index as shown in Fig.3.14b. d. f and h.
3.6. Predictable Component Analysis (PrCA)

In the preceding sections, we discussed the PNA predictability at various time scales and its possible sources related to ENSO forcing. In these analyses, the PNA index defined by the time series of the leading PCA mode has been used. This simplifies the analysis and easily captures the main feature of PNA predictability. However such a one dimensional index may not well characterize some features of multiple space of atmospheric circulation, especially the atmosphere response to SST forcing. Thus, in this section, we will explore the predictability of the atmospheric variability over the northern hemisphere, especially the position and role of the PNA predictability in the total predictability of various atmospheric variabilities.

To achieve the above goal, we performed the PrCA analysis using 500mb geopotential height monthly anomaly data of MME_E and MME_H over the northern hemisphere for each lead time of prediction. As discussed in section 2.3.3, we obtained multiple PrCA modes for each prediction lead time like multiple modes in PCA. In this study, we only consider the modes that can characterize the PNA. Shown in Fig 3.15 and Fig. 3.16 are the PNA-like modes of PrCA using MME_E and MME_H at different lead months, chosen from the first four modes, as higher modes have very small and negligible contribution towards total predictability. As can be seen in these figures, the PNA-like pattern usually appears as the second mode in the PrCA except as the first mode in the prediction of lead time of 1 month for MME_E. This indicates that the PNA is one of the most predictable structures among low-frequency atmospheric variability. It should be noted that the first pattern in PrCA,

3 A further examination found that most of first modes of the PrCA characterize a NAO-like pattern, which we will not discuss here.
unlike PCA analysis, does not necessarily explain the most variability, as shown at the parenthesis of each panel of Fig. 3.15 and Fig. 3.16. This is due to the fact that PrCA seeks the optimal modes based on predictability rather than on explained variance (variability).

To get further insight into the source of PNA predictability, we projected the time series of PrCA patterns of Fig. 3.15 and Fig. 3.16 onto the observed SSTA of the tropical Pacific region. Shown in Fig. 3.17 are the projected patterns, which resemble the warm phase of ENSO. Fig. 3.17 builds a direct bridge between the predictable components of PNA with the ENSO, namely that, the most predictable component of PNA is mainly due to the tropical SST forcing.

3.7. Discussion and Conclusion

In this study, the PNA predictability has been explored using actual and potential predictability measures for individual and MME of three ensemble prediction datasets, HFP2, DEMETER and ENSEMBLES stream 2. The first one is the ensemble of uncoupled models whereas the latter two are the products of fully coupled models. The primary purpose of this study was to examine the efficacy of time averaging with regard to extracting a predictable signal from day-to-day weather fluctuations. The emphasis was put on actual skill and potential predictability over different time scales from days to seasons, and the intercomparison of different scales and different models. The actual skill measures include correlation and RMSE, and the potential skill measures include signal to noise ratio and information-based metrics.

The comparison of MME skill with individual models in terms of actual skill is first evaluated. It was found that the MME prediction skill was better than most of the individual
models. As a result, the MME is used to study PNA predictability. A comprehensive analysis reveals that, on daily time scale, the PNA predictability limit is around 20 days. A modest improvement can be observed at a weekly time scale. The effect of time averaging was more pronounced at the bi-weekly time scale, and persisted in monthly time scale to 3 months. Overall, the predictability is seen to be improved with time averaging over longer periods. The phase de-correlation rate decreased with time averaging, which resulted in an increase in predictability, but a reduction in climatological variability with the time average, on the other hand, restricted further predictability improvements of time averaging. Thus, the predictability of the filtered fields is a trade-off between these two factors.

For the first time, the PNA potential predictability is evaluated at multiple time scales. The \( R_{E \alpha} \) and \( AC_{MI} \) measured potential skill at multiple time scales using MME\_H and MME\_E. These measures were found to be consistent with the notion that model prediction skill and predictability generally decreased with lead time (monotonicity). A practical comparison between information-based \( AC_{MI} \) with SNR-based \( AC_{P} \) with averaging time affirm their theoretical relationship [Eq. 2.11] that \( AC_{MI} \) can measure more potential predictability than \( AC_{P} \). One advantage of using \( AC_{MI} \) is that it can measure the statistical dependence, linear or nonlinear, between the ensemble mean and an ensemble member (hypothetical observation) whereas \( AC_{P} \) measures only linear dependence. When the prediction and climatological distributions are Gaussian and ensemble spread is invariant with forecast, the two measures are equivalent.

The \( R_{E \alpha} \) is an effective measure in distinguishing the individual forecasts from each other where some forecasts are more predictable than others (It is not explicitly confirmed in this study but using \( R_{E \alpha} \) as measure potential predictability the large \( R_{E \alpha} \) corresponds to more
prediction skill than low RE_A). It has been observed that RE_A can detect the signals related to SST forcing in terms of its inherent association with the signal component, whereas other information-based metrics are lacking this property. This property of RE_A makes it preferable compared to other potential predictability measures. In this study, this association has been reconfirmed for the PNA region, irrespective of time scale of interest, by observing temporal variabilities in RE_A and signal component. The high amplitude of the signal component in some particular initial conditions, especially at monthly to seasonal time scale, was related to SST anomalies in the tropical Pacific (ENSO) region.

Comparison of predictability between coupled and uncoupled models reveals that the potential predictability measured using coupled models is larger than uncoupled models for the PNA index, at multiple scales, in contrast with the actual skills which show no significant difference between coupled and un-coupled models. To shed light on the possible reason for this, the SSTA prediction skill is evaluated for coupled and uncoupled models on monthly time scale at different lead times. Persistent SSTA is used for uncoupled models whereas predicted SSTA from coupled models is analyzed. It was found that the SST prediction skill in coupled models is much better even at lead time of four months, especially in the tropical and PNA regions. Thus, the atmospheric bias is the most probable reason responsible for weakening or offsetting the advantage of coupled model in predicting PNA.

The source of PNA potential predictability from monthly to seasonal time scale has also been investigated in this study. It is well known that the variations in the North Pacific region are teleconnected with variations in the tropical SST forcing (Horrel and Wallace 1981; Hoskins and Karoly 1981; Simmons 1982; Sardeshmukh and Hoskins 1988; Straus and Shukla 1997; 2002). Thus, emphasis was placed on the role of ENSO in PNA predictability, which was
explored using lead-lag correlation between PNA index and the SST signal. It was found that the tropical SST forcing impacts the PNA potential predictability mainly by changing the amplitude of the ensemble mean. The SSTA-PNA correlation patterns in the models resemble the typical ENSO pattern, suggesting that the ENSO is the main source of PNA seasonal predictability. Their maximum correlation occurred at a lag of two to three months, which is consistent with some past studies.

PrCA analysis explores the PNA predictability and its source in the data space with the temporal and spatial variation. It was found that the PNA is one of two most predictable patterns among low frequent atmospheric variability. Its main source of predictability is due to the tropical SSTA forcing, i.e., ENSO.
Table 3.1: Actual and potential prediction skill ($\text{AC}_{\text{M}}$) over different time scales using MME_H and MME_E (Correlation value of 0.5 is chosen as threshold for significance).

<table>
<thead>
<tr>
<th></th>
<th>$\text{AC}_{\text{M}}$ MME H</th>
<th>$\text{AC}_{\text{M}}$ MME E</th>
<th>Actual skill MME H</th>
<th>Actual skill MME E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily</td>
<td>28</td>
<td>40</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>Weekly</td>
<td>5</td>
<td>16</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Bi_weekly</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Monthly</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Fig. 3.1 NAO and PNA patterns obtained using rotated EOF analysis of 500mb monthly NCEP geopotential height data. Contour interval is 10m with negative contours dashed. Dark (light) shading indicates values <10m (>10m).
Fig. 3.2 RMSE and Ensemble spread as a function of lead time over daily time scale, for different models and their MME's. Here the ENSEMBLES models are shown with "_E" and DEMETER models are shown with "_D". CERFACS and LODYC are DEMTER models.
Fig. 3.3 Correlation and RMSE as function of lead time for different time average, for different models. Persistent skill and climatological spread is also shown for each time scale. Here different colors are indicating different models.
Fig. 3.4 Same as Fig. 3.3 but using only MME_H, MME_D and MME_E. Here vertical error bars are the sample standard deviation calculated using bootstrap experiment at each lead time.
Fig. 3.5 Correlation coefficients of different MME over different time scales. Here the daily correlation values are averaged over one week (weekly_avg), 2 weeks (bi_weekly_avg) and month (monthly_avg) to compare with corresponding weekly, bi-weekly and monthly time scale correlation.
Fig. 3.6 a) Anomaly correlation for MME_H over different time scales as function of lead time, here the daily correlation is shown for first 20 days only. (b) Climatological variance of averaging time by observation (stars) and MME_H (open diamonds). (c) Persistence Correlation at different time scales. The legends for (a) and (c) are same and shown in the bottom.
Fig. 3.7 Comparison of $AC_M$ with $AC_P$ over different time scales for MME_H and MME_E.
Fig. 3.8 Actual and potential skill for MME_H and MME_E. Here the vertical bars are the sample standard deviation calculated using bootstrap experiment (see context).
Fig. 3.9 The PDF of prediction for different randomly chosen lead times for weekly time scale for ECMWF models and HFP2 models. Observed PDF for weekly time scale is also shown as reference.
Fig. 3.10 $AC_{MI}$ as function of lead time over different time scales for MME_E and MME_H is presented. Here vertical error bars are the sample standard deviation calculated using bootstrap experiment (see text).
Fig. 3.11 Monthly correlation skill of SST using persistence (left panel) and MME_E (right panel) at different lead months.
Fig. 3.12 Averaged RE as function of initial time is shown at different time scales as function of initial time for MME_E.
Fig. 3.13 Variations in averaged SC and DC as function of initial time for MME_E over different time scales.
Fig. 3.14 Spatial pattern Correlation between predicted ENSEMBLES SST and PNA of MME_E (left panel) for different time lags (based on selected initial conditions in which signal component was high). Here SST is leading and PNA is lagging behind. Right: Same as left but using observed SST and PNA of MME_H.
Fig. 3.15 The leading PrCA modes for different lead months for MME_E which capture the PNA related variabilities. In the title of each pattern, the variance explained based on total predictability by that pattern is also shown.
Fig. 3.16 Same as Fig. 3.15 except using MME_H.
Fig. 3.17 The projection of PrCA patterns of Fig. 3.15 and Fig. 3.16, on observed SST.
Chapter 4: MJO prediction skill and predictability in AGCM and CGCM ensemble forecasts

After presenting PNA predictability at various time scales, this chapter will focus on the actual and potential predictability of Madden-Julian Oscillation (MJO) using one coupled and one uncoupled model ensemble forecasts. The emphasis will be placed on the potential predictability using the information theory based measures introduced in chapter 2. The literature review on MJO predictability has been presented in chapter 1 and measures of actual and potential predictability are introduced in chapter 2. This chapter will emphasize on MJO variability and predictability analysis.

4.1. MJO index for Prediction skill

The MJO index is calculated using combined Empirical Orthogonal Functions (EOFs) based on the technique of Wheeler and Hendson 2004. The ensemble forecast data is taken from one Atmospheric General Circulation model (AGCM) and one Coupled General Circulation model (CGCM). The detailed description about AGCM (GCM3 from HFP2) and CGCM (CanCM3 from CHFP2) has already been given in chapter 2. To represent tropical convection, we used Precipitation (PR) instead of Outgoing Long Wave Radiation (OLR) because OLR from GCM3 was not available in the Historical Forecast Project 2 (HFP2) archive. To compute combined EOFs of lower and upper zonal winds (u850&u200 respectively) and PR, NCEP-NCAR reanalysis data has been utilized from 1979-2002. At first, the temporal mean (1979-2002) and first three harmonics of the daily climatology are removed at each grid point. Then the previous 120 days mean is subtracted from each input field. Following Lin et al. 2008 and Rashid et al. 2010, we neglected the complicated step of
removing the component that is linearly associated with ENSO, before removing previous 120 days mean. As will be discussed later, this step is insensitive to the definition of ENSO index. Next, a meridional band average (15°S-15°N) is performed for each of u850, u200 and PR anomaly data, while retaining the longitudinal variation. The individual latitudinally-averaged fields are then divided by the normalizations computed from each variable’s own zonal average of temporal standard deviation and then three fields are combined. This last step is necessary to ensure that each field contributes equally to the total variance of the combined field.

The longitudinal distributions of EOF1 and EOF2 of u850, u200 and PR are shown in Fig. 4.1. EOF1 explains 12.8% and EOF2 explained 11.2% of total combined variance respectively and they are well distinguished from the remaining EOFs which each explain less than 5% of the variance. The upper level zonal winds are out of phase with the low-level winds, indicating the baroclinic structure of wind circulation. The zonal fields change their sign at 150°E and the Greenwich meridian for EOF1. At near 150°E, low level convergence (westerlies from the west and easterlies from the east), and upper level divergence (easterly wind anomalies are to the west over Indian Ocean and westerly wind anomalies are to the east across the Pacific) occur, which favors enhanced convection and precipitation at this longitude (the EOF sign is arbitrary, depending on the PC). At the Greenwich meridian, the opposite situation occurs: due to upper zonal convergence and low level divergence, the air flows downward which favors low precipitation. In EOF2, the regions where both zonal winds change their signs move eastward, indicative of eastward movement of MJO oscillations. In terms of precipitation, enhanced precipitation is observed near the Maritime Continent in EOF1, consistent with low level convergence and upper level divergence near
150°E. An opposite pattern – weakened precipitation – can be seen in EOF2 in the eastern Indian Ocean due to the low level divergence mentioned above. This detail of EOF1 and EOF2 has been included following Lin et al. 2008.

The EOF1 and EOF2 of u200 and u850 show most of the variability found by Wheeler and Hendon (2004, WH04 hereafter), even though they removed the variability part linearly associated with ENSO before EOF analysis, while the present study did not. The temporal correlation between the daily principal component time series of the first EOF (PC1) with that of WH04 is 0.96 and it is 0.97 for PC2 (This confirms that extra step of removing the component that is linearly associated with ENSO is not needed). Since no temporal filtering is used here in obtaining the combined EOFs, these PC time series are suitable for real-time forecast. To be consistent with WH04, our pair of PC time series are also named as the real-time multivariate MJO (RMM) series 1 (RMM1) and 2 (RMM2).

The RMM indices for model hindcasts are obtained by projecting daily forecast anomaly data onto EOF1 and EOF2. Forecast anomalies are produced by the removing the daily forecast climatology for the period of 1979-2001. Next, the interannual variability is removed by subtracting the mean from the previous 120 days, but now this mean is created for a forecast at lead time n days as the mean of the previous n days of the forecast plus the previous 120-n days of observation (NCEP) up to the start of the forecast (Rashid et al. 2010; Lin et al. 2008). Prior to projecting the model anomalies onto observed EOFs, the individual latitudinally-averaged model anomaly fields are divided by observed normalizations, following WH04. After the projection, the resulting RMM values are divided by the square root of the respective observed eigenvalues. The whole procedure described above is consistent with the approach that is being recommended by the United States Climate
4.2. Sub-seasonal low frequency variability

In this section, the low-frequency tropical variabilities from 30S-30N simulated by GCM3 and CanCM3 are compared with the observations. It is important to first analyze the MJO simulation in the models, as most of the current climate models experience difficulties in simulating the tropical convection related to MJO. It should be emphasized here that following Lin et al. 2008, the forecast data for first month is discarded from each forecast and next three months of forecast data is used to represent the long-term climatological behavior in the models. Particularly in this section, the analyses are divided into different seasons, winter (November-April) and summer (May-October). The procedure of presenting low frequency intraseasonal variability has also been influenced by Lin et al. 2008.

Tropical low frequency variability is detected using a Lanczos filter that passes the variability with the period of 20-100 days, for both observations and model outputs. The variances for the filtered u850 from the NCEP-NCAR reanalysis data and models in winter and summer are represented in Fig. 4.2 (a, b). In winter, the greater part of low-frequency variability in low level zonal wind is found in the tropical Indian Ocean, the Maritime Continent, and the western Pacific region. In summer, the activity of maximum variance is shifted to 10° N, representing variability related to Indian Monsoon circulation. This could be related to the fact that the intra-seasonal variability related to MJO is strongly connected to variation between active and quiescent monsoon periods. Pai et al. 2009 evaluated the impact of MJO on the intraseasonal variation of summer monsoon in India. They found that there is a strong connection observed between MJO and the onset of break and active monsoon events over
India. It was shown that during an active monsoon period, the MJO activity is rather weak (in terms of both amplitude and phase), which could shift the maximum variability to 10° N in summer. Thus, both GCM3 and CanCM3 characterize the low-level zonal wind low-frequency variability reasonably well in terms of longitudinal variation only. It seems no significant difference between CGCM and AGCM models in terms of low level zonal wind.

The low-frequency variability of upper zonal wind in the form of u200 is presented in Fig. 4.3. The observed u200 (Fig. 4.3a, b) shows that the upper level zonal wind variability is small in the tropics and large in mid-latitudes. The minimum variance was found over the Indian Ocean and the western Pacific. During winter season, maximum variance can be observed over the equatorial eastern Pacific, which may be linked to an extratropical influence and variability in the zonal outflow of MJO-related convection in the Maritime Continent and western Pacific regions (Lin et al. 2008). Apparently, during winter season, CanCM3 simulated the upper zonal wind variability better than GCM3 especially in the eastern Pacific region and some parts in western Pacific (120W-90W and 0-20N). In summer, both models (Fig. 4.3d, f) could not simulate the upper level variability well as compared with observations (Fig. 4.3b).

Fig. 4.4 presents the variances of filtered precipitation rates in the two model simulations (Fig. 4.4c-f) and in observation (Fig. 4.4a-b). It should be noted here that to represent precipitation in observation, Global Precipitation Climatology Project (GPCP) data is used, which is available from 1997-2009. The low-frequency variability in PR is consistent with that of the low-level zonal wind (Fig. 4.2), in general. Large precipitation activity can be seen in the Indian Ocean, Maritime Continent and western Pacific regions. CanCM3 seems to be more wet (positive bias) than GCM3 in comparison with observations. The probable reason
may be due to the fact that analysis period for precipitation was short (1997-2001).

In summary, both CanCM3 and GCM3 simulated the tropical low-frequency variability comparable to observation. CanCM3 showed a positive bias in simulating PR whereas GCM3 could not capture the pattern of upper zonal wind variability in the eastern Pacific. Overall, both models apparently have close resemblance with observation in simulating intraseasonal variability (ISV) in Indian Ocean, Maritime Continent and western Pacific in zonal winds and precipitation rates.

4.3. Actual Prediction skill of MJO bi-variate index

In this section, the actual prediction skill of coupled and uncoupled models will be evaluated. The bootstrap experiment is used to perform the statistical significant test of correlation skill. The bootstrap experiment was designed as follows: 1) the observation and prediction are paired to construct a sample of observation-prediction pairs for the entire period from 1979-2001, at a given prediction lead time; 2) 95% of the samples are of the prediction-observation pairs are randomly chosen and used to calculate their correlation coefficient; 3) repeating 2) 1000 times to obtain 1000 correlation coefficients, whose standard deviation is used as a threshold value at a given lead day (i.e., the error bar in figures).

Fig. 4.5 shows the correlation coefficient and root mean square error (RMSE) of the MJO bi-variate index from dynamical predictions using the ensemble mean (cf. chapter 2) of CanCM3 and GCM3 for the period 1979-2001. The skill scores presented are calculated for annual data (12 months). The persistence skill is better for a one day lead time for CanCM3 and a 2 day lead time for GCM3. Beyond that, the model skill is either comparable with or better than the persistence skill. Prediction skill in the form of correlation drops to 0.5 at
about 16 days for CanCM3 and at 12 days for GCM3 (Fig. 4.5a). CanCM3 managed to have a correlation very close to 0.5 until 21 days. Using bootstrap results (error bars in Fig. 4.5a), the coupled model (CanCM3) correlation skill is significantly higher than that of the uncoupled model (GCM3), consistent with the previous findings of Seo et al. (2009).

In terms of RMSE (Fig. 4.5b), the persistence skill is better (smaller) than model's skill for one to two days, but thereafter the ensemble mean forecast beats persistence. The model prediction skill is considered to be useful only if it is better than a prediction based on climatology; thus the bottom limit of predictability in terms of RMSE occurs when RMSE approaches the climatological spread (Kimoto et al. 1991). It should be emphasized here that the RMSE values were normalized for Fig. 4.5b. As can be seen from Fig. 4.5b, CanCM3 and GCM3 attained this limit at around 10 days and 7 days respectively.

In the context of MJO prediction skill in CGCM and AGCM models, CanCM3 outperformed GCM3 significantly at all lead times. In Fig. 4.5, coupled and uncoupled model prediction skill were close to each other only for the one day lead time forecast, beyond that the coupled model prediction skill significantly surpasses the uncoupled model skill (Fig. 4.5a). For short lead time, the difference in GCM3 and CanCM3 prediction skill is probably attributed to different initialization processes of these models. For example, in CanCM3 the atmosphere is initialized by the data assimilation process (Merryfields et al. 2013) whereas AGCM atmosphere was initialized by lagged forecast (Kharin et al. 2009). After 2 weeks, the difference in the prediction skill is more probably from the impact of coupling, due to the initial information lost. Fu et al. 2007 directly compared intraseasonal predictability in a fully coupled model with its atmospheric component and found that interactive atmosphere-ocean coupling extends intraseasonal predictability by at least a week longer than the
atmosphere only model. This is also evident from Fig. 4.5a: if we choose 0.5 as an arbitrary threshold correlation value below which prediction skill is not significant, then CanCM3 prediction skill at around 16 days is better than GCM3 prediction skill which is around 12 days.

Fig. 4.6 shows the correlation skill of CanCM3 (Fig. 4.6a) and GCM3 (Fig. 4.6b) for winter and summer, respectively. It can be seen in Fig. 4.6a that the coupled model has similar prediction skill for both seasons. In the uncoupled model (Fig. 4.6b), both winter and summer MJO skill remained close to each other for few lead days, and after 14 days, winter MJO skill is much better than summer skill. The uncoupled model has poor prediction skill at long leads in summer is probably because it cannot detect the weak signal and resolve the impact of monsoon on MJO well. During the summer season, the MJO signals are weak and partially connected with the intraseasonal variation in Asian summer monsoon (Hendon and Liebmann 1990; Lawrence and Webster 2002).

As we introduced in chapter 1, it is still an unsettled issue regarding whether atmosphere ocean coupling is important for MJO prediction skill and predictability. We may argue from the discussion above, that air-sea interaction may be important for MJO prediction skill, at least in the context of actual skill, since the coupled model's skill is either comparable to or better than the uncoupled model's skill. The difference in prediction skill is found to be significant between CanCM3 and GCM3. Specifically, the CanCM3 MJO prediction skill had a correlation value above 0.5 for 16 days and even beyond. This result is noteworthy because according to Merryfield et al. 2014, MJO is not well represented in CanCM3, mainly due to small variability in the western Indian Ocean and western Pacific relative to variability in the central Indian Ocean and Maritime Continent. This implies that if these biases and
errors are reduced in CanCM3 simulation, its prediction skill can be further enhanced.

4.4. Potential prediction skill of MJO bi-variate index

4.4.1. Relative Entropy, Predictive information, and Predictive power

The actual prediction skill of the MJO index is evaluated in previous sections, and it is found that the coupled model prediction skill is significantly better than that of the uncoupled model. In this section, our emphasis will be on evaluating the MJO potential predictability between coupled and uncoupled models; mainly using information theory based potential predictability measures. These measures include relative entropy ($R_{EA}$), predictive information (PI), predictive power (PP), mutual information (MI) etc. As mentioned in chapter 2 of the thesis, $R_{EA}$ is composed of two components, Signal Component (SC) and Dispersion Components (DC). The mathematical representations of these components are also presented in chapter 2. Also, the relationship between the model prediction skill and potential predictability will be identified and to seek a predictor of forecast skill by which we can assess the degree of confidence that can be placed in an individual prediction. It should be noted here that these information theory measures have not previously been applied to MJO prediction skill (at a sub-seasonal scale) and their previous application was limited to seasonal climate prediction (ENSO) or weather predictability. We expect that information based potential predictability measures at sub-seasonal time scales will provide important precursors of predictability. We will start by finding the potential predictability measure that is best suited to assess MJO index predictability.

Fig. 4.7 shows the $AC_{MI}$ and $AC_{P}$ (see chapter 2 for more details) predictability measures of the MJO bi-variate index for GCM3 and CanCM3 as a function of lead time. Here $AC_{P}$ is
estimated using a "perfect model" scenario: taking one model realization (an ensemble member) as a true observation and measuring the forecast skill using the mean of the remaining ensemble members. The perfect model correlation skill is equivalent to conventional signal to noise ratio (Tippett et al. 2007). Figure 4.7 can be discussed in two different ways. First the coupled model potential prediction skill is significantly better than that of the uncoupled model, consistent with actual prediction skill results. We will be back to this issue later in this section. Second, $AC_{MI}$ is larger than $AC_P$ both for GCM3 (Fig. 4.7a) and CanCM3 (Fig. 4.7b), suggesting that the conventional predictability measure of signal to noise ratio (SNR) underestimates the potential predictability. Again if we take 0.5 as an arbitrary value for a significant potential prediction skill, the coupled and uncoupled model’s skill remained above 0.5 for all lead times both for $AC_{MI}$ and $AC_P$.

Figure 4.8 compares the difference of the potential predictability $AC_{MI}$ between coupled model (CanCM3) and uncoupled model (GCM3). Again, the bootstrap method is used to estimate the extent of uncertainty due to sampling error, as shown by the bars in Fig. 4.8. A significant difference beyond the sampling uncertainty is seen in potential skill between CanCM3 and GCM3 for all lead days, which is consistent with the results using actual skill measures. For comparison, the potential predictability in terms of $AC_P$ is also shown in Fig. 4.8b (Here error bars are not drawn as Fig. 4.8b is shown just for comparison). Again, the coupled model potential prediction skill surpassed the uncoupled model skill. As Mutual Information (MI) is measure of model overall potential predictability and $AC_{MI}$ is overall model prediction skill so next, we will focus on $RE_A$, PI, PP (cf. chapter 2) as potential predictability measures of individual prediction.

Figure 4.9 shows the variation of the average $RE_A$, PI and PP over the effective lead times for
GCM3 and CanCM3. Here effective lead time is equivalent to the maximum lead time beyond which the prediction skill is not significant and is determined by bootstrap experiment. Some apparent features can be identified as follows: For most of the predictions, $RE_A$ has relatively similar variation from one prediction to another, both for GCM3 (Fig. 4.9a) and CanCM3 (Fig. 4.9b), where CanCM3 has larger potential predictability in terms of $RE_A$. Some of the predictions which have large $RE_A$ might be related to the onset of an ENSO event (1997/98, 2000/2001), which will be interesting to explore. Similar patterns of potential prediction skill in terms of $PI$ (Fig. 4.9c&d) and $PP$ (4.9e&f) can be seen. The relatively similar variation of $RE_A$ with different initial conditions (i.e. through time) is quite interesting result because previously $RE_A$ was observed to be large in predictions related to ENSO and smaller for non-ENSO predictions (Tang et al. 2008). This was attributed to the presence of the ensemble mean in the definition of $RE_A$ (cf. chapter 2), which can quantify the impact of external forcing on the predictability. On the other hand, this contribution (ensemble mean) is absent both in $PI$ and $PP$ and they only depend on ensemble and climatological spread (cf. chapter 2). Next we will examine the variations of SC and DC with initial time of prediction.

The variations both in SC and DC for CanCM3 and GCM3 are shown in Fig. 4.10 as a function of initial time. Both SC and DC are averaged over the effective lead days (first 20 days). As can be seen in Fig. 4.10 the signal component of both GCM3 (Fig. 4.10a) and CanCM3 (Fig. 4.10c) is relatively large in some predictions whereas for most predictions, SC is small. If we consider the amplitude of SC >1 as anomalous then SC was large in early 1997 and late 1999, both for GCM3 (Fig. 4.10a) and CanCM3 (Fig. 4.10c). Interestingly, these events occurred one year before the start of strong ENSO events (1997/98 and
W. Younas: Evaluation of Climate Predictability

2000/2001). But on the other hand, no anomaly is seen in SC before another strong ENSO event (1982/83). This suggests that strong MJO activity may lead to a strong ENSO event, but the result cannot be generalized to all ENSO events. Also, this study consists of two general circulation models only, so inclusion of more models from different centres will be helpful in generalizing the importance of SC of MJO index as precursor of onset of ENSO.

The DC component of GCM3 (Fig. 4.10b) and CanCM3 (Fig. 4.10d) has relatively smooth variations from forecast to forecast but the amplitude is comparable to the SC amplitude. As DC is estimated using ensemble and climatological spreads (cf. chapter 2), which in turn are linked with internal variability, we may argue that MJO is a strong internal variability phenomena. Moreover, the DC of CanCM3 (Fig. 4.10b) is larger than the DC of GCM3 (Fig. 4.10d). Thus on one hand, MJO index may be used to predict the onset of ENSO (which needs to be confirmed with more studies). On the other hand, the predictability of MJO may be dependent on internal dynamics. To further test these hypotheses, we will next present the correlation of SC and DC with REA, both for coupled and uncoupled models, to see which of SC or DC contributes more in estimating REA.

Fig. 4.11 shows the correlation of REA with SC and DC both for GCM3 (left panel) and CanCM3 (right panel). It should be emphasized here that SC is associated with the ensemble mean and with differences between the ensemble mean and the climatological mean. On the other hand, DC is associated with the ensemble spread and understood as a prediction utility associated with reduction in uncertainty (Tang et al. 2005; Kleeman 2008). It is clear from Fig. 4.11, that correlations of SC and DC with RE are significantly high. Although, SC has strong association with REA, the contribution of SC to MJO potential predictability is not clear. But we may argue that in terms of DC (which is mainly determined by the ensemble
variance, which in turn is a representation of internal variability) MJO has a strong component of internal variability. In Numerical Weather Prediction (NWP), the variation in spread is related to initial perturbation of ensemble members and the growth of initial perturbations with time with both of these factors playing an important role. Whereas in Ensemble Climate Prediction (ECP), initial conditions have a relatively small contribution to the growth of initial error and growth is mainly controlled by model behavior leading to less variation of prediction ensemble spread. In present study, i.e. at sub-seasonal time scales, which is a bridge between NWP and ECP, we found that both ensemble spread (DC) and ensemble mean (SC) may be important precursor of MJO potential predictability (Although the contribution in terms of SC is not vibrant and requires more attention).

4.4.2. Comparison of Actual and Potential Prediction skill

The comparison of actual and potential predictability using CanCM3 and GCM3 is shown in Fig. 4.12. Here potential predictability is measured using only $AC_{\text{mi}}$. The difference between actual and potential predictability can be viewed as the extent by which the predictive skill can be expected to increase through improved initial conditions and reduced model error. It can be seen that potential predictability is always significantly higher than actual prediction skill. For example, the potential prediction skill remains above the arbitrary correlation value of 0.5 even beyond 40 days (not shown) whereas actual skill drops to a correlation value of 0.5 at 12 days for GCM3 and 16-21 days for CanCM3. This is not surprising since potential predictability represents an upper limit of prediction skill that a perfect model can achieve (Rashid et al. 2010; Waliser et al. 2003).
4.5. Potential Predictability Source for MJO index

In this section, we will examine why coupled model prediction skill was found to be significantly better than uncoupled model skill. In this context, comparing SST forecast skill using persistence with a coupled model SST forecast can shed light on the possible reasons.

To investigate this, the prediction skill of daily Sea Surface Temperature Anomaly (SSTA) is calculated from 30S-30N for the time period 1982-2001. The SSTA prediction is provided by the persistence scheme for uncoupled model (GCM3) and by the oceanic component of the coupled model (CanCM3). The NOAA OI SST V2 high resolution dataset is used for observed data. Starting from unfiltered daily observed SST data from 1982-2002, the seasonal cycle of the daily climatology is removed at each spatial grid point. Next, spatial correlation is calculated at each grid point for different lead days.

Fig. 4.13 shows the prediction skill of daily SSTA using persistence and coupled model forecast (CanCM3), for different lead days. As expected, the persistence skill is initially better than the coupled model prediction skill, but beyond ten days, the CanCM3 SST prediction skill was meaningfully better than persistence (Fig. 4.13). The procedure is also repeated for intraseasonal SST skill (by removing 120 days mean) but similar conclusion is obtained (figure not shown). Thus, a probable reason why the coupled model leads to better MJO prediction than the uncoupled model is the better SST prediction skill of the coupled model.

4.6. Summary and Conclusion

In this study, the MJO prediction skill is evaluated from the AGCM and CGCM models in the Canadian seasonal hindcast (HFP2) and Canadian coupled seasonal hindcast (CHFP2)
experiments. MJO prediction skill has been studied using actual and potential predictability skill measures. The central focus of this study was to assess the difference and similarities between AGCM and CGCM ensemble forecasts. For comparison of actual skill, correlation and Root Mean Square Error (RMSE) measures were used, whereas information theory based potential predictability measures were used to assess the potential prediction skill. In terms of potential prediction skill, a comparison is also drawn between the usual signal to noise ratio based measure and MI based potential prediction skill.

A significant difference in both actual and potential prediction skill is found between coupled and uncoupled model ensemble mean prediction skill. CanCM3 predicted the MJO index successfully in terms of correlation for about 16 days and beyond (the threshold limit of prediction skill was taken as a correlation value of 0.5) whereas GCM3 predicted MJO index successfully for 12 days. As argued above, the difference in prediction skill between GCM3 and CanCM3 is probably attributed to coupling impact and different initialization processes involved in both models, depending on the lead time. CanCM3 clearly has better prediction skill than GCM3 at all lead days until the predictability limit is achieved. It can be argued that air-sea interaction is necessary to better predict MJO as indicated by many other studies (Waliser et al. 1999; Woolnough et al. 2004; Webster et al. 2002; Fu et al. 2003; Fu and Wang 2004; Rajendran et al. 2004; Fu et al. 2007; Kim et al. 2010).

In terms of potential predictability of the MJO index, first, the Mutual Information is used as an overall predictor of prediction skill. The theoretical relationship between MI based potential predictability skill $\text{AC}_\text{MI}$ and the usual signal to noise ratio based predictability skill ($\text{AC}_\text{P}$) is examined for predictions of the daily MJO index. These measures were found to be consistent with the notion that model prediction skill and predictability generally decreased
monotonically with lead time (not shown). A practical comparison between information-based $AC_{MI}$ with SNR-based $AC_p$ affirm their theoretical relationship that the former can measure more potential predictability than the latter. One advantage of using $AC_{MI}$ is that it can measure the statistical dependence, linear or nonlinear, between ensemble mean and an ensemble member (hypothetical observation) whereas $AC_p$ measures only linear dependence. Again, CanCM3 prediction skill surpassed the GCM3 potential prediction skill significantly based on the bootstrap test.

The other measures of Information theory based potential predictability, like average $RE_A$, PI and PP were used to estimate the MJO potential predictability and a comparison was drawn between coupled and uncoupled model prediction skill. It is found that all of these predictability measures have relatively similar variations from forecast to forecast. Some of the MJO forecasts have large $RE$, one year before strong ENSO events (1997/98 and 2000/2001). As $RE_A$ can be decomposed into SC and DC components, the variation of SC and DC has also been presented as function of initial time. For most of the predictions, DC showed relatively similar variations with initial time, whereas SC had anomalously large amplitude in some predictions one year before strong ENSO events but SC did not show any anomaly in some other ENSO events.

To further dig down the association of SC and DC with $RE_A$, SC and DC are correlated with $RE_A$ independently for uncoupled and coupled models. It was shown that the correlation was high both for SC and DC. It should be emphasized that both SC and DC have different interpretations in terms of precursor of forecast utility. SC is associated with ensemble mean and is coupled with shifts in the means and DC is connected with ensemble spread and understood as prediction utility associated with reduction in uncertainty. The conclusion in
terms of SC (i.e. using SC as precursor of onset of ENSO) is not clear and need more consideration in terms of incorporating more models. On the other hand, we may argue that the predictability of MJO index is mainly influenced by internal dynamics (in terms of DC). This is quite interesting result in terms of potential predictability of MJO, which was not explored previously using information theory based measures. Previously, information theory measures were applied either to seasonal prediction (Tang et al. 2008) or weather predictability (Kleeman 2008); and it was shown that either ensemble mean (SC) is more important precursor for the predictability on seasonal time scale or ensemble spread is more reliable precursor for weather predictability. This is the first time that information theory measures were applied to sub-seasonal scale predictability and it was shown that DC is reliable predictor for predictability at sub-seasonal time scale. This conclusion is more pronounced in coupled model prediction skill (e.g. correlation between DC with RE_A in CanCM3 is 0.60 whereas same correlation in GCM3 is 0.54).

As coupling involves the interaction between atmosphere and ocean, the next logical step was to determine the prediction skill of tropical daily SST anomaly data and compare it with persistence SST skill. It was shown that persistence skill was better than predicted skill for few lead days and after that predicted SST performed better than persistence. The conclusion remained same when we repeated the same result using intraseasonal and monthly SST data. Thus a probable reason why the coupled model prediction skill of MJO index is better than uncoupled model, may be due to the skill difference between predicted SST skill and persistence skill.

In summary, there is significant difference found in the prediction skill of AGCM and CGCM ensemble forecast skill of MJO index. It may be argued that the air-sea interaction is quite
important in the predictability of MJO. Also, the large difference between actual and potential predictability reveals that further improvements in general circulation models are necessary in order to bridge the gap between actual and potential prediction skill of MJO index.
Fig. 4.1 Longitudinal distribution of EOF1 and EOF2 from the combined analysis of prlr, u850 and u200 of NCEP-NCAR reanalysis.
Fig. 4.2 Variability patterns of low-frequency (20-100 days) of u850 for the NCEP-NCAR reanalysis, GCM3, and CanCM3. Areas with variances greater than $6\text{m}^2/\text{s}^2$ are shaded.
Fig. 4.3 Variability patterns of low-frequency (20-100 days) of u200 for the NCEP-NCAR reanalysis, GCM3, and CanCM3. Areas with variances greater than 24m²/s² are shaded.
Fig. 4.4 Variability patterns of low-frequency (20-100 days) of prlr for the GPCP, GCM3, and CanCM3. Areas with variances greater than $24m^2/s^2$ are shaded.
Fig. 4.5 Correlation and RMSE of MJO bi-variate Index of GCM3 and CanCM3 from 1979-2001.
Fig. 4.6 Seasonal dependence of MJO correlation skill in coupled and uncoupled model ensemble forecast from 1979-2001.
Fig. 4.7 Comparison of $A_{CM}$ and $A_{CP}$ between GCM3 and CanCM3.
Fig. 4.8 Potential predictability $A_{CM}$ and $A_{CP}$ comparison between CanCM3 and GCM3 for MJO bi-variate index for the period 1979-2001.
Fig. 4.9 Variation of average RE, PI and PP as function of intial time for GCM3 (left panel) and CanCM3 (right panel) for the period 1979-2001. Here average is done over first 20 lead days only. Here the RE, PI and PP values are averaged over the lead days 1-20.
Fig. 4.10 Variation of average SC (left panel) and average DC (right panel) as function of lead time and initial time both for GCM3 and CanCM3. Here average is done over first 20 lead days only.
Fig. 4.11 Correlation of RE with SC and DC, for GCM3 (left panel) and for CanCM3 (right panel).
Fig. 4.12 Comparison of Actual and potential skill between GCM3 and CanCM3. The potential skill is estimated using MI.
Fig. 4.13 Daily SST anomaly correlation for Persistence (left panel) and predicted (right panel) for different lead days.
Chapter 5: Time Average Prediction skill Comparison between PNA and MJO

In previous chapters, we have evaluated the prediction skill and potential predictability for two prominent modes of low frequency atmospheric variability, PNA and MJO. In this chapter, we will further discuss them. Emphasis is placed on summarizing how several important factors of predictability; using the PNA and MJO as examples, impact predictability of two different time scales, including the coupling role of air-sea interaction, the measures of predictability and the time average. This chapter will also serve as summary chapter for the whole thesis with some additional evaluation of PNA and MJO potential predictability with time average. First two sections of this chapter will be devoted to summarize the work presented in the previous chapters (chapter 3&4).

5.1. Coupling role in the Prediction of PNA and MJO

In chapter 3, PNA actual skill at various time scales was presented. To explore the role of coupling in PNA predictability, coupled and uncoupled models from different research centres were used. Correlation and RMSE skill measures were used to estimate the actual skill. The bootstrap test was used to assess differences between coupled and uncoupled models' prediction skill. It was found that there is no significant difference of PNA prediction skill among coupled and uncoupled models (Figs. 3.3&3.4). Even MME prediction skill, which is often considered to offset the individual model errors, did not change this conclusion. This finding is attributed to the atmospheric model bias which may serve as the most probable reason responsible for weakening or counterweighing the advantage of a coupled model in predicting PNA.
In addition, MJO prediction skill, measured by correlation and RMSE, is estimated using one coupled and one uncoupled model from CCCma: HFP2 and CHFP2. The bootstrap test, mentioned above, is applied to distinguish the role of coupling in the predictability of MJO. The correlation value of 0.5 is considered as a threshold below which the forecast is not reliable. It was found that the coupled model predicted the MJO index for 16 days and longer whereas the uncoupled model predicted it for 12 days (Fig. 4.5). Bootstrap test results reveal that coupled model prediction skill is significantly better than uncoupled model prediction skill. The importance of air-sea interaction in predicting the MJO index has been addressed in many studies (e.g. Waliser et al. 1999; Woolnough et al. 2004; Webster et al. 2002; Fu et al. 2003; Fu and Wang 2004; Rajendran et al. 2004; Fu et al. 2007; Kim et al. 2010). The results of the present study may also serve to state the importance of the role of coupling in the predictability of MJO.

5.2. The Coupling role in the Potential Predictability of PNA and MJO

The potential predictability study in chapters 3&4 showed some consistent properties for both PNA and MJO modes although some of their features, in particular these related to precursors of predictability, were different between PNA and MJO. The consistent properties will be briefly described first.

PNA potential predictability is found to be significantly higher in coupled models than in uncoupled models (Fig. 3.10). This is in contrast with PNA actual prediction result (Figs. 3.3&3.4), which showed no significant difference between coupled and un-coupled models. On the other hand, significant difference in potential prediction skill has also been found between coupled and uncoupled models for MJO potential predictability (Fig. 4.8) which is
consistent with its actual prediction result (Fig. 4.5).

To examine why the coupling role is not clear in PNA prediction skill (i.e. why there are contrasting results for actual and potential prediction skill) and why it has significant impact on MJO prediction skill, we scrutinized the SST prediction skill in coupled models and compared it with persistence skill. It should be emphasized here that the models used to predict PNA and MJO were different. It was found, that SST prediction skill in coupled models is much better than persistence in all models used for both indices (Fig. 3.11&4.13). Thus for PNA, we may conclude that model biases weaken or diminish the role of coupling in its prediction skill, whereas for MJO, we may argue that the coupling role is important and it is confirmed by SST prediction skill results.

5.3. Relative Role of Signal and Noise in Potential Predictability at Various Time Scales

In this section, the relative role of the signal and noise in estimating potential predictability at various time scales will be summarized using results of chapter 3&4. The $R_{EA}$ and $AC_{MI}$ served the purpose of measuring potential skill at multiple time scales. A practical comparison between information-based $AC_{MI}$ and SNR-based $AC_P$ with time average asserted their theoretical relationship that the former can measure more potential predictability than the latter. One advantage of using $AC_{MI}$ is that it can measure the statistical dependence, both linear and nonlinear, between ensemble mean and an ensemble member; whereas $AC_P$ can only estimate linear dependence. These two measures are equivalent in two conditions 1) if distributions are Gaussian 2) if ensemble spread is invariant with forecast initial time.

In terms of averaged $R_{EA}$, PI and PP for PNA index, it is found that variations in $R_{EA}$ with
time average (Fig. 3.12), are opposite to variations in PI with time average (Fig. 5.1). It should be noted here that $\text{REA \ and PI \ are \ related \ to \ each \ other \ as \ REA \ also \ contains}$ information provided by PI (Tang et al. 2008). This is also visible from Eq. 2.6 in which the first term of $\text{REA}$ is basically PI (Eq. 2.9). Comparing Fig. 3.12 with Fig. 5.1 it can be seen that when $\text{REA}$ is large, PI is small and vice versa. Also, $\text{REA}$ is large in some forecasts which correspond to strong ENSO events (1982/82 and 1997/98). A similar pattern is visible when comparing variations in $\text{REA}$ (Fig. 3.12) with variations in PP (Fig. 5.2). $\text{REA}$ is composed of two components, SC and DC (Eq. 2.6). These in turn are estimated by ensemble mean and ensemble spread, whereas PI and PP by definition, are only dependent on ensemble spread (Eq. 2.9&2.10) and they lack the contribution from ensemble mean. So it would be interesting to plot the variations of SC and DC with initial time to see the impact of external forcing (in terms of ensemble mean) and the role of internal dynamics (in terms of ensemble spread) on PNA predictability.

The variations in SC and DC with time average revealed that SC is large in those predictions when $\text{REA}$ (1982/82 and 1997/98) is large and vice versa, whereas DC is large when $\text{REA}$ is small and vice versa (Fig. 3.13). The variation feature in SC was related to SST forcing in the tropics and it was shown that the correlation between PNA index and SSTA was high mainly in tropical central Pacific region (Fig. 3.14). On the other hand, the correlation of SC and DC with $\text{REA}$ (Fig. 5.3&5.4) showed that $\text{REA}$ is mainly determined by SC, and DC has negligible contribution in estimating $\text{REA}$. Thus for the PNA index, SC can be used to assess the impact of external forcing on PNA potential predictability. Next, the variations in averaged $\text{REA}$, PI and PP will be examined for the MJO index.

In terms of the MJO index, $\text{REA}$ showed relatively similar variations with initial time for
most of the predictions (Fig. 4.9a,b). Some of the predictions with large $R_{E_A}$ were found to be related to the onset of strong ENSO events (1997/98 and 2000/2001). On the other hand, PI and PP also showed relatively similar variations with initial time (Fig. 4.9b-f). To further elaborate these results, the variations in SC and DC as a function of lead time were also presented (Fig. 4.10). It is found that SC was large in a few predictions whereas for most of other predictions, SC was small. Interestingly, it was noted that the SC was anomalously large one year before the occurrence of strong ENSO events (1997/98 and 2000/2001). Thus apparently, it may be argued that strong MJO activity can lead to strong ENSO events and SC of MJO index can be used as precursor to forecast the onset of a strong ENSO. But this result cannot be generalized for two reasons: 1) SC of MJO index was not large for another strong ENSO event (1982/83) 2) only one coupled and one uncoupled model were used in this study. On the other hand, DC showed relatively smooth variations with initial time but interestingly the amplitude of DC was comparable to SC. Next, the contribution of SC and DC toward $R_{E_A}$ will be examined for MJO index.

The correlation of $R_{E_A}$ with SC and DC for MJO index (Fig. 4.11) revealed that both measures have high correlation with $R_{E_A}$ (especially SC). The contribution of SC in terms of MJO potential predictability is not clear but it can be argued in terms of DC that MJO is strong component of internal variability.

5.4. The Impact of Time Average on Prediction utility for various time scales

In this section, the predictability results with time averaging will be presented both for PNA and MJO indices. For PNA, the daily data is averaged over weekly, bi-weekly and monthly time scales. Actual prediction skill results show that PNA predictability is seen to benefit
from time averaging (Figs. 3.3&3.4). In terms of correlation, a modest improvement in forecast skill is found on weekly time scales, which was more pronounced at the bi-weekly time scale and persisted for 3 months over monthly time scale. The decrease in phase decorrelation rate was also compensated by reduction in climatological variability which restricted the predictability to improve further. The potential predictability of PNA has also been evaluated over multiple time scales, using $A_{CMI}$. The potential predictability is also seen to improve with time averaging, consistent with actual predictability results. We summed up time average results of actual and potential predictability for PNA index in a table 3.1 which showed the maximum lead time that remains at a correlation skill of 0.5 for all time scales. According to the table, $A_{CMI}$ remained significant until 40 days for MME$_E$ and 28 days for MME$_H$ at daily time scale. In contrast with actual skill, the improvement in skill is also quite substantial even at weekly time scales (16 weeks for MME$_E$ and 5 weeks for MME$_H$).

To study the impact of time average on MJO predictability, the MJO daily index is averaged over 3-days, 5-days and 7-days. It is found that the time averaging MJO index (Fig. 5.5&5.6), did not show any improvement over the daily time scale. A probable reason could be that the MJO index has already filtered out high frequency weather noise and low frequency variabilities related to ENSO and only retains variabilities at intra-seasonal time scale. The time average of intra-seasonal time scales may not improve the skill.

5.5. Capability of potential predictability measures in quantifying prediction skill at various time scales

In this section, different measures of potential predictability for the PNA and MJO indices
will be studied in the context of quantifying precursors of potential predictability. The actual and potential predictability measures defined in chapter 2 will be used in this section. The key point is to investigate which potential measures have skill that is closer to actual skill both for PNA and MJO with time average. As there is no improvement found in the prediction skill of MJO with time average both in terms of actual and potential skill, the MJO results will be evaluated only at the daily time scale.

PNA potential predictability results with time average will be discussed first. The scatter plots of MI versus actual prediction skill $r$ and RMSE, evaluated by the Multimodel Ensemble Mean (MME), are shown in Fig. 5.7, in order to examine the relationship of potential predictability and actual skill. Here MME_H, which refers to MME of 4 HFP2 uncoupled models, is used and results are shown with different time averages. Fig. 5.7 indicates that MI is a good predictor of model skill for MME_H and model skill is well correlated with MI with time average. When MI is large, the skill in terms of $r$ and RMSE is good (high correlation and low RMSE), whereas when MI is small, the skill is usually low.

Fig. 5.8 shows the same conclusion but for MME of ENSEMBLES data, which used 3 coupled models to study the PNA index. Comparing Fig. 5.7 with Fig. 5.8, we can observe that model prediction skill drops slowly in Fig. 5.8 as compared Fig. 5.7 which is a confirmation of the results of chapter 3 i.e. the coupled models prediction skill is better than uncoupled models.

Fig. 5.9 compares the MI from (A1) and MI estimated from (2.12) for the PNA index with different time averages. MME_H results are shown in the left column (Fig. 5.9a,c,e,g) and MME_E results are shown in the right column (Fig. 5.9b,d,f,h). As can be seen, the two estimates of MI are in good agreement with each other both for coupled and uncoupled
models. Their correlations are around 0.98 from daily to monthly time scale. Thus MI is good indicator of overall skill. Next, the relationship between information-based measures and model predictability for the daily MJO index will be examined.

Fig. 5.10 shows the scatterplots of MI versus model skill measured in terms of $r$ and RMSE both for GCM3 and CanCM3 using the daily MJO index. Again, it can be perceived that for the MJO index too, MI is a good predictor of model skill, especially for the coupled model (CanCM3), where model skill drops slowly as compared to GCM3 skill. When MI is large, the correlation is high and RMSE is low (high skill) and when MI is small, the model skill is poor. Next, MI is estimated from (2.12) for the MJO index and is compared with MI estimated from (A1). Fig. 5.11 shows the scatterplots of both estimates for uncoupled (Fig. 5.11a) and coupled model (Fig. 5.11b). The high correlation between two estimates for both coupled and uncoupled models shows that MI is a good indicator of measuring the overall skill of the MJO index.

In conclusion, in this chapter, a thesis summary is presented by presenting the overall comparison between PNA and MJO prediction skill. As per the objectives of the thesis presented in chapter 1, the emphasis is placed on comparison of actual and potential predictability between coupled and uncoupled models, to find a precursor of forecast skill, potential predictability source and impact of time average on prediction skill of both indices. Information based potential predictability measures were used to estimate the potential predictability whereas correlation and RMSE measures were utilized to measure the actual skill. A comparison is also made between MI based correlation skill with usual signal to noise ratio based potential predictability.
In terms of actual skill, there is no significant difference between coupled and uncoupled models for the PNA index, whereas coupled model performed better than uncoupled model for the MJO index. On the other hand, coupled models potential predictability was significantly better than uncoupled models, both for PNA and MJO indices. The comparison between $AC_{M1}$ and $AC_P$ revealed that the usual signal to noise ratio base measures underestimate the potential predictability for both indices. SC is found to be an important precursor for PNA potential predictability whereas DC is found to be important precursor for MJO potential predictability. It is argued that PNA is teleconnected with ENSO whereas MJO is an internal dynamic mode. The predictability is seen to improve for PNA index with time average whereas the MJO predictability did not show any significant difference with time average.
Fig. 5.1 Variations in PI for PNA index with time average for MME_H (left panel) and for MME_E (right panel)
Fig. 5.2 Variations in PP for PNA index with time average for MME_H (left panel) and for MME_E (right panel)
Fig. 5.3 Correlation of SC and DC with RE at various time scales for PNA index using MME of ENSEMBLES (coupled) models.
Fig. 5.4 Same as Fig. 5.3 but for MME of HFP2 (uncoupled) models.
Fig. 5.5 Time average prediction skill (correlation) of MJO index using 3-days, 5-days and 7-days mean over daily time scale.
Fig. 5.6 Time average potential prediction skill of MJO index using 3-days, 5-days and 7-days mean over daily time scale. AC_MI is used as potential prediction skill.
Fig. 5.7 Scatterplots of Mutual information MI vs $r$ (left) at different time scales and MI vs RMSE (right) at different time scales, using MME_H for PNA index.
Fig. 5.8 Same as Fig. 5.7 but for MME_E.
Fig. 5.9 Scatterplots of mutual information MI vs estimated MI from r for MME_H (left) and MME_E (right) at different time scales for PNA index.
Fig. 5.10 Scatterplots of Mutual information MI vs r and RMSE for GCM3 (upper panel) and CanCM3 (lower panel), of MJO index at daily time scale.
Fig. 5.11 Scatterplots of mutual information MI vs estimated MI from \( r \) for MJO daily index using GCM3 (left) and CanCM3 (right).
References


W. Younas: Evaluation of Climate Predictability


NOAA High Resolution SST data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.esrl.noaa.gov/psd/.


114
Ph.D. Dissertation: University of Northern British Columbia


W. Younas: Evaluation of Climate Predictability


Vitart, F., 2009: Impact of the MJO on tropical storms in ECMWF EPS.


Yang, D., Y. Tang, Y. Zhang, and X. Yang, 2012: Information-based potential predictability of


Appendix

If the prediction and climatological PDFs are Gaussian, the MI can be written as below (e.g., Yang et al. 2012, Tang et al. 2013)

\[
M_I = \frac{1}{2} \left( \ln \sigma_q^2 - \ln \sigma_p^2 \right) \geq \frac{1}{2} \left( \ln \sigma_q^2 - \ln(\sigma_p^2) \right) = -\frac{1}{2} \ln \left( \frac{\sigma_q^2}{\sigma_p^2} \right) = -\frac{1}{2} \ln(1 - \text{STR}) \tag{A1}
\]

\[
\text{STR} = \frac{\sigma_q^2}{\sigma_p^2} \tag{A2}
\]

Where \( \sigma_q^2, \sigma_p^2 \) and \( \sigma_s^2 \) are climatological, ensemble and signal variances respectively. The inequality in (A1) is due to the fact that arithmetic mean is larger than or equal to the geometric mean. A detailed derivation can be found in Yang et al. (2012). The STR can be interpreted as the perfect correlation skill \( AC_p \) that is defined by the correlation between ensemble mean and a random ensemble member (e.g., Tang et al. 2013), as shown below.

\[
AC_p = \sqrt{\text{STR}} \tag{A3}
\]

Using (A3), (A1) can be re-written as

\[
M_I \geq -\frac{1}{2} \ln(1 - AC_p^2) \tag{A4}
\]

Namely,

\[
AC_p \leq \sqrt{1 - \exp(-2 \cdot M_I)} \tag{A5}
\]

The MI-based correlation \( AC_{MI} \) is defined by (e.g., Delsole 2004; Yang et al. 2012)

\[
AC_{MI} = \sqrt{1 - \exp(-2M_I)} \tag{A6}
\]

Thus, we can have the relationship between \( AC_{MI} \) and \( AC_p \) as expressed by (2.11).