<table>
<thead>
<tr>
<th>項目</th>
<th>内容</th>
</tr>
</thead>
<tbody>
<tr>
<td>タイトル</td>
<td>予測技術と予測性の評価方法 - マッデン・ジュリアン振動のパーセプションスベース</td>
</tr>
<tr>
<td>著者</td>
<td>市川悠衣子</td>
</tr>
<tr>
<td>発行日</td>
<td>2018年03月22日</td>
</tr>
<tr>
<td>DOI</td>
<td>10.14943/doctoral.k13135</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/2115/69416">http://hdl.handle.net/2115/69416</a></td>
</tr>
<tr>
<td>ファイル情報</td>
<td>Yuiko_Ichikawa.pdf</td>
</tr>
<tr>
<td>学術機関</td>
<td>北海道大学学術情報学会 : HUSCAP</td>
</tr>
</tbody>
</table>
Doctoral Dissertation

Evaluation methods of prediction skill and predictability in
the Madden-Julian oscillation phase space

Yuiko Ichikawa
Division of Earth and Planetary Dynamics, Department of Natural
History Sciences, Graduate School of Science, Hokkaido University,
Sapporo, Japan

March, 2018
Abstract

The Madden-Julian oscillation (MJO) is the dominant component of planetary-scale intraseasonal variability in the tropics and consists of a convective system coupled with atmospheric dynamics. Through its local influence in the tropics and the teleconnections to higher latitudes, the MJO is thought to be an important agent for subseasonal forecasts, because of its typically regular behavior of moving eastward with its speed of about 5 m/s across the Indian Ocean and the Pacific. Weather forecast is usually evaluated in light of prediction skill, which is a measure of how much the forecast is matched with the truth in a particular lead time. Another important measure is predictability, which represents the forecast lead time when the initial-value error expands enough to mask out the prediction signal in the perfect model. This dissertation examines prediction skill and predictability on the MJO phase space, which is two-dimensional space customarily utilized to represent the phase and amplitude of the MJO.

Prediction skill in the MJO phase space has been evaluated by the combination of the bivariate root-mean-square error (RMSE) and the bivariate anomaly correlation coefficient (BCC). However, the BCC is not suitable for models in which the MJO amplitude systematically damps in some specific phases, because the BCC strongly depends on the MJO amplitude. To overcome this drawback, the mean-error vector is introduced in this dissertation to associate a model’s erroneous mean tendency with the RMSE. For example, the Japan Meteorological Agency (JMA) forecast has a leftward mean error vector field uniformly distributed over the MJO phase space with its amplitude related to the RMSE. Mean error vector explains a fraction of RMSE as model bias which can be removed by simple bias correction of subtracting mean error.
vector from forecast states. The RMSE should then be used with the mean error vector for evaluating the MJO prediction skill.

Predictability has been conventionally estimated with spread of an ensemble experiment, the members of which start their forecasts from the initial conditions with small, optimally-growing perturbations. This approach assumes that the forecast model is perfect. However, this perfect model assumption is not always reasonable for the MJO prediction due to the model bias. This dissertation relieves the problem by introducing an alternative method to estimate the potential predictability. Based on a theoretical consideration, a maximum possible value of the initial-value error is related to the co-variance between analysis and bias-corrected ensemble-mean forecast. This approach does not assume the perfect model and has a merit that it enables us to estimate potential predictability from the ensemble-mean forecast data, not from their spread. The proposed method is hence readily applicable to the multi-model average of ensemble-mean forecasts. Applied to the operational models from European Centre for Medium Range Weather Forecasting (ECMWF), the JMA, and National Centers for Environmental Prediction (NCEP), the new method estimates that the predictability is higher when MJO amplitude exceeds unity, consistent with conventional method in which the error is evaluated as the ensemble-forecast spread.

The dissertation originally developed a new evaluating method of prediction skill (Chapter 2) and that of predictability (Chapter 3), motivated by phenomena, like the MJO, that the forecast model has a large bias, and examined the validity in terms of multi-model forecast data projected onto two-dimensional MJO phase space. We recommended the use of mean error vector that directly delineated the systematic model bias, instead of the BCC that is highly sensitive to the MJO amplitude. Even a simple
bias correction by removing the mean error vector from forecast state vector could provide us a plausible behavior on the MJO phase space over the forecast lead time. The consideration of external agents that possibly affect the MJO forecast, such as the Rossby wave penetration into the MJO genesis region, may enhance efficiency of bias correction; however, it remains unaddressed. On the other hand, the new method on potential predictability realized the first publication on the multi-model evaluation of the MJO predictability in the season-to-subseasonal intercomparison project that was launched in 2015 and is now ongoing. The results suggested that the “maritime continent barrier” is not related to the intrinsic prediction limit but to the insufficient representation of the MJO in numerical models. This implied that a further model development could resolve this barrier. Though the above results are robust, there is possibility that the method is often sensitive to the choice of criterion of expanding initial-vale error; it should be cautioned when one applied this method to other phenomena.
日本語要旨
マッデン・ジュリアン振動(MJO)は熱帯の主要な季節間内変動であり、惑星規模の循環場が対流系と結合した現象である。MJOはテレコネクションを通じて世界各地の気象に影響を及ぼし、また多くの場合規則的に5メートル毎秒で東進する性質から、季節内予報においてMJOは重要であると考えられている。気象予測は一般的に、ある予報時間において予測が真値とどの程度一致するかを評価する予測精度と、初期値誤差の拡大によって予測対象のシグナルが不分明になるまでの期間を評価する予測可能性の観点から評価されてきた。本博士論文は慣例的にMJO相空間と呼ばれるMJOの状態を表現する2次元相空間上において予測精度と予測可能性の一般的な評価手法について調査した。

予測精度の評価には、二変数二乗平均平方誤差(bivariate root-mean-square error, 以下RMSE)と二変数アノマリ相関係数(bivariate anomaly correlation coefficient, 以下BCC)の組み合わせが慣例的に使われてきたが、これらの評価方法はモデルバイアスを評価できない。それだけでなく、BCCはMJOの振幅に強く依存するので、ある相でMJOのシグナルが減衰する傾向にあるモデルにおいてBCCを使うことは適切ではない。そこで平均誤差ベクトルによってモデルの平均移動速度誤差とRMSEを結びつけることでこの問題を解決する。たとえば気象庁の予報モデルは、MJO相空間上で一様な左向きベクトルであらわされる平均誤差を持ち、その振幅はRMSEと関係づけられる。この場合、MJOの予報精度の評価においては、RMSEと平均誤差の組み合わせを用いるべきである。
予測可能性は慣例的に完全モデルを仮定し、初期値に小さな、成長しやすいよう最適化した擾乱を与えたアンサンブル予報のスプレッドを用いて推定されてきた。しかしながら、完全モデルの仮定はMJO相空間においてバイアスをもつモデルに対しては必ずしも成立しない。また完全モデルの仮定のために、マルチモデルアンサンブル予報を用いて予測可能性を推定するとは困難だった。ここでは完全モデルの仮定を用いず、初期値依存性誤差を取り除く最大値に結びつける。この手法は完全モデルを仮定しないので、スプレッドを用いず、マルチモデル平均予報から予測可能性を推定することが可能である。それに基づく予測可能性の推定値は、ヨーロッパ中期予報センター(ECMWF), 日本気象庁(JMA), 米国国立環境予測センター(NCEP)いずれの現業モデルにおいても相空間上での振幅が1以上の時に長くなった。これはスプレッドを用いた従来の手法による推定値と大まかに一致する結果である。この手法はマルチモデル平均予報に対しても適用可能であるので、マルチモデル解析も行われた。初期値依存性のMJO相空間上の相依存性も議論された。

この博士論文はMJOのような数値予報においてバイアスの大きい現象に適した予測精度(Chapter 2)と予測可能性(Chapter 3)の新しい評価手法を独自に提案し、それらを2次元のMJO相空間上に射影したマルチモデル予報データに適用することで検証した。MJOの振幅に対して不安定なBCCに代わって、モデルの系統的な誤差の指標となる平均誤差ベクトルを用いることが推奨された。平均誤差ベクトルを予報の状態ベクトルから除去するという簡易なバ
イアス補正であってもある程度の予報の改善が見込めることが分かった。例えば中高緯度からのロスビー波の侵入がMJOの生成に与える影響のような外的要因を考慮することによって、さらにバイアス補正の効率を高めることも可能と考えられるが、この研究では行わなかった。その一方、予測可能性推定の新手法は2015年に立ち上がった季節内から季節予報プロジェクトにおいて、マルチモデル解析を行った初の試みである。その結果からは、“maritime continent barrier”と知られる問題が潜在的な予報限界によるものではなく、数値モデルにおける不十分なMJOの再現を原因とするものであることが示唆された。これは同時に、モデルの改善によってこの問題が解消されうることを意味する。これらの結果は十分に確からしいが、初期値依存性誤差の時間成長の目安の与え方に依存しているため、ほかの現象にこの手法を適用する際には注意が必要である。
Contents

Abstract ........................................................................................................................ iii
日本語要旨 ................................................................................................................... vi
Chapter 1. General introduction ................................................................................. 1
Chapter 2. Methods to evaluate prediction skill in the MJO phase space ............. 4
  2.1. Introduction ...................................................................................................... 4
  2.2. Data .................................................................................................................. 6
  2.3. MJO phase space and mean tendency vector ................................................... 7
  2.4. Evaluation methods ........................................................................................ 13
    i. BCC and bivariate RMSE ..................................................................................... 13
    ii. Mean error ............................................................................................................. 14
  2.5. Results ............................................................................................................ 14
    i. Comparison of BCC and RMSE ........................................................................... 14
    ii. Mean tendency in reanalysis and forecast ............................................................ 21
  2.6. Summary and discussion ................................................................................ 24
Chapter 3. An Alternative Estimate of Potential Predictability on the MJO Phase
  Space Using S2S Models ............................................................................................. 27
  3.1. Introduction .................................................................................................... 27
  3.2. Data ................................................................................................................ 29
    i. Verification and phase space ................................................................................ 29
    ii. Forecast Data ........................................................................................................ 30
    iii. Alternative method to estimate potential predictability .................................... 35
  3.3. Results ............................................................................................................ 41
    i. The prediction skill and bias correction ................................................................. 41
    ii. Prediction limit and its comparison with a conventional method ......................... 47
    iii. The multi-model ensemble analysis .................................................................... 53
  3.4. Concluding remarks ....................................................................................... 57
Chapter 4. General summary .................................................................................... 61
Acknowledgments .......................................................................................................... 63
References .................................................................................................................... 64
Chapter 1. General introduction

The Madden-Julian oscillation (MJO) is the dominant component of planetary-scale intraseasonal variability in the tropics and consists of a convective system coupled with atmospheric dynamics. The MJO propagates eastward at an average speed of $5 \text{ m s}^{-1}$ across the equatorial Indian and Pacific Oceans with a timescale of 30–100 days and a zonal extent of 12,000 to 20,000 km (Zhang 2005). The MJO modulates tropical depressions in western hemisphere (Mo 2000; Maloney and Hartmann 2000) and Asian-Australian monsoons (Wang and Rui 1990; Hendon and Liebmann 1990; Lau and Waliser 2011), and excites teleconnection patterns such as North Atlantic Oscillation (Cassou 2008; Frederiksen and Lin 2013; Lin et al. 2015) and Pacific North American pattern (Liebmann and Hartmann 1984; Higgins and Mo 1997; Mori and Watanabe 2008). After discovered by Madden and Julian (1971), the numerical simulation of the MJO has outpaced advances in conceptual understanding (Emanuel 2007). Some of the recent numerical models realize an MJO forecast with the lead time of nearly a month (Miyakawa et al. 2014; Vitart 2014). Moreover, an extended-range forecast even in the mid-latitudes could be achieved through an accurate prediction of the MJO in the tropics (Jones et al. 2004; Cassou 2008). Stimulated by this expectation, prediction projects were organized with their focus on the representation of the MJO (Matsueda and Endo 2011; Brunet et al. 2010; Zhang 2013; Vitart et al. 2017).

The MJO prediction has been evaluated as two aspects; prediction skill and predictability. The prediction skill refers to a model’s ability to accurately simulate the real atmosphere, namely how well the model describes essential processes through its dynamics and physical parametrization schemes. On the other hand, predictability does
not depend on model settings, but limited by intrinsic uncertainty due to chaotic nature of the atmosphere. This kind of uncertainty arises from negligibly small error in the initial value of forecasts because of the nonlinearity in governing dynamics, and therefore it is referred to as initial-value error (Lorenz 1969; Yoden 2007). Several metrics have been proposed to evaluate prediction skill and predictability, and some of them are accepted as standard. However, typical metrics for prediction skill do not indicate how the model is biased and have singularity around the origin; similarly, predictability has been estimated with an assumption which may not be reasonable for the MJO prediction as will be described in the main text.

The purpose of this dissertation is to examine prediction skill and predictability of the MJO, and to propose new methods to compensate drawbacks in the conventional ones. We hereafter discuss prediction skill and predictability with analysis and forecast data projected onto the MJO phase space, summarized the MJO’s large scale circulation field, defined as two leading empirical orthogonal function (EOF) modes for intraseasonal variability in equatorial zonal wind (Wheeler and Hendon 2004). For example, an eastward propagation of the MJO is represented as counter-clockwise movement and strength of convective activity as distance from the origin.

The rest of this dissertation is organized as follows. Chapter 2 describes the methods to evaluate prediction skill in the MJO phase space based on Ichikawa and Inatsu (2016), which is subdivided into the specific introduction on the evaluation of prediction skill in the MJO phase space in section 2.1, a description of the Japan Meteorological Agency’s one-month forecast and verification in section 2.2, the definition of the MJO phase space and mean tendency vector in section 2.3, definitions and characteristics of evaluation methods in section 2.4, results in section 2.5, and
discussion in section 2.6. Chapter 3 proposes an novel method to estimate potential predictability using numerical forecasts provided by European Centre for Medium Range Weather Forecasting (ECMWF), the Japan Meteorological Agency (JMA), and National Centers for Environmental Prediction (NCEP) based on Ichikawa and Inatsu (2017). To these numerical models the evaluation methods described in Section 2 was applied as well. Section 3.1 gives a brief introduction to the estimation of predictability in the MJO phase space. Section 3.2 describes the forecast models whose output was used in the analysis. Section 3.3 gives the estimate of potential predictability of the MJO using the new method and compares it with that from the conventional method. Section 3.4 concludes Chapter 3. The general summary is given in Chapter 4.
Chapter 2. Methods to evaluate prediction skill in the MJO phase space

2.1. Introduction

The prediction skill, meaning model accuracy against the insufficient representation of the MJO, was typically evaluated by bivariate anomaly correlation coefficient (BCC) between the verification and ensemble mean forecast that measures the similarity of the spatial pattern, and the root-mean-square error (RMSE) that measures the magnitude of difference between the verification and forecast. For instance, Vitart et al. (2007) used a linear correlation coefficient for each component of a real-time multivariate MJO index (RMM; Wheeler and Hendon, 2004). Lin et al. (2008) proposed the use of BCC and bivariate RMSE based on RMM1 and RMM2 to investigate the amplitude and phase dependency on the MJO phase space. Most of the recent publications have followed Lin et al. (2008), and the BCC and RMSE are standard metrics for evaluating the prediction skill for the MJO (Gottschalck et al. 2010; Wang et al. 2013; Kim et al. 2014; Vitart 2014). The combination of BCC and bivariate RMSE provides information of prediction skill about the amplitude of the MJO signal and the moving speed toward the east. Additionally, the prediction limit of numerical models can be defined by setting a criterion. Usually the prediction limit is the day when BCC reaches 0.5 (Hollingsworth et al. 1980) or RMSE reaches $\sqrt{2}$ (Rashid et al. 2011).

The BCC around the origin of the MJO phase space should be used with caution because the BCC has a strong singularity at the origin. We discuss two simple examples with a set of single verifications and their corresponding forecasts. If the verification is
(1, 0) and the forecast is (0.01, 0), then BCC is unity. In contrast, if the verification is 
(1, 0) and the forecast is (−0.01, 0), then BCC is −1. Because RMSE in the former case 
is 0.99 and in the latter case it is 1.01, the prediction skill is similar, being consistent with the similarity of the states. If the verification is interchangeable with the forecast in these examples, the results do not change. Thus, it is obvious that BCC is not a suitable index for evaluating the prediction skill. Due to this fact, the hypothesis that the prediction skill is low during inactive MJO periods cannot be examined by BCC (Lin et al. 2008; Xiang et al. 2015). It should be examined by other indices (Jones et al. 2000). Kumar and Hoerling (2000) showed that the expected value of BCC decreases with amplitude in a context different from the MJO. The traditional anomaly correlation coefficient between verification and forecast patterns of some climatic variables removes the singularity because the method measures the similarity of the non-MJO anomaly around the origin of the phase space.

Moreover, the combination of the BCC and RMSE does not provide the difference between verification and forecast. For example, if the verification is (1,0) and the forecast is ($\sqrt{3}/2, 1/2$), then BCC is $\sqrt{3}/2 = \cos 30^\circ$, which indicates a difference of 30° in the MJO phase, and RMSE in this case is $(\sqrt{6} - \sqrt{2})/2$. If the verification is interchanged with the forecast, BCC and RMSE do not change. This means that the two indices do not show whether the forecast MJO is faster or slower than the verification MJO. The lack of this information is usually compensated for by the mean phase error, although this also has a singularity at the origin (Rashid et al. 2011). Hence, another metric is required to overcome these drawbacks of these two indices. It is also desirable that the new metric can be used for any MJO phase irrespective of whether the MJO is active or inactive.
Recently, studies have compared the prediction skill of models (Neena et al. 2014; Kim et al. 2014; Matsueda and Endo 2011). The combination of BCC and RMSE was frequently used in these studies because of its clarity and ease of calculation. It may be especially useful for multi model analysis of prediction skill, if there is a metric that is as easily calculated as BCC and RMSE, yet still holds the information of model bias, to be used when BCC cannot be applied due to the singularity problem. The mean error vector, defined as the vector difference between model and analysis, compensates for the lack of information on the MJO’s moving speed and dumping.

This chapter introduces a new suite of metrics to evaluate prediction skill that can be used regardless of MJO status. Because the BCC is not always suitable around the origin of the MJO phase space, another index is necessary. We examine the JMA model, and use the one-month ensemble forecast to consider our metric. Moreover, the relation between new and old metrics will be clarified. Because we intend to evaluate the prediction skill methods, but not the potential predictability, we do not discuss initial disturbance and forecast spread in this chapter.

### 2.2. Data

This study used the one-month ensemble forecast provided weekly by the JMA (The Japan Meteorological Agency 2013). The forecast model has a horizontal resolution of TL159 (comparable to an approximately 110 km mesh) with 60 vertical levels up to the top of 0.1 hPa. An initial perturbation was not imposed in the tropics before February 2007 in the JMA forecast system (Kubota and Mukougawa 2005). However, the version after March 2007 has several members with an initial perturbation made by the breeding of growing-mode method with a norm of an equatorial 200 hPa velocity potential
(Chikamoto et al. 2007). Therefore, we used the data from January 2008 to December 2012. Although 25 runs start from Wednesday and the other 25 runs start from Thursday in the same week, we regard 49 out of those 50 runs as a single forecast. We used 24 perturbed runs started from Thursday and discarded a control run for Thursday because it gave a similar result to the control run of Wednesday. Hereafter, the ensemble average of the 49 runs is regarded as a forecast result. We also used the JRA25/JCDAS JMA reanalysis dataset from January 1990 to December 2012 (Onogi et al. 2007) for verification.

2.3. MJO phase space and mean tendency vector

Following the method in Wheeler and Hendon (2004) the MJO phase space was spanned by the two gravest combined-EOF modes in the tropical intraseasonal variations of zonal wind at 200 and 850 hPa levels of the reanalysis data, averaged between 15°S and 15°N after removing seasonal and interannual variations (Fig. 2.1). We did not use outgoing longwave radiation data in the combined EOF analysis (Straub 2013), and we did not use special treatments to remove El Niño/Southern Oscillation signals (Lin et al. 2008; Rashid et al. 2010). Conventionally the MJO forecast evaluation is performed for each MJO phase. The MJO phase $k$ ($k = 1, 2, \ldots, 8$) is defined as the state where $r \geq 1$ and $\frac{\pi}{4} k - \frac{5\pi}{4} \leq \theta < \frac{\pi}{4} k - \pi$, where $(r, \theta)$ is the polar coordinate representation in the MJO phase space. Provided that $\phi(\tau)$ is an evaluation quantity for a forecast with lead time $\tau$, $\phi(\tau)$ is calculated separately for the forecasts started from phase $k$. We call this representation the phase separation method, and $\phi(\tau)$ for phase $k$ is referred to as $\phi_k(\tau)$. Phase separation is widely used to summarize the
temporal variation of model performance (Neena et al. 2014; Kim et al. 2014; Matsueda and Endo 2011). The MJO phase space is also subdivided into small square blocks. The block \((i, j)\) is defined as the area \((i − 1)Δ ≤ \text{RMM1} < iΔ\) and \((j − 1)Δ ≤ \text{RMM2} < jΔ\), where \(Δ\) is the length of one side of a small square block and \((i, j)\) are integers. \(\phi(\tau)\) is also calculated separately for the forecast in which the MJO state is initially located in block \((i, j)\). We call this representation the block separation method, and \(\phi(\tau)\) for block \((i, j)\) is written as \(\phi_{ij}(\tau)\). In this paper, block separation is applied with \(Δ = 0.25\), balancing the resolution with the limitations of the amount of data. Figure 2.2 compares the phase separation and block separation methods. The block separation method is superior for displaying the two-dimensional map because of its resolution; for example, there are 61 blocks inside the unit circle (Fig. 2.2b). However, it is difficult to draw time-series graphs with this method because it has more segments than the phase separation method (Fig. 2.2a). Although the averaging for each block is weighted with the probability density function (PDF) estimated from data by using the method in Kimoto and Ghil (1993) after Section 2.5, simple averaging is used in the following explanation for simplicity.
Figure 2.1: (a) First and (b) second modes of combined empirical orthogonal function calculated for the ±15° latitudinally averaged anomalies of the zonal wind at (dotted line) 850 hPa and (solid line) at 200 hPa. Each mode explains 16.5% or 15.4% of the variance in the zonal wind anomaly.
Figure 2.2: Schematic illustration of (a) phase and (b) block separation methods demonstrated using dummy data. Color contours denote quantity according to the scale at the bottom of the figure.
We determine the averaged motion of the MJO state by drawing the mean tendency vector in the MJO phase space with the block separation method. Figure 2.3a shows the mean tendency vector of verification with a 5-day lead time, which is expressed as

\[
V^0_{ij}(\tau) = \frac{1}{N_{ij}} \sum_{l=1}^{N_{ij}} x^0(t^l_{ij} + \tau) - x^0(t^l_{ij})/	au,
\]

(2.1)

Here, \(x^0(t)\) is the verification state vector, \(t^l_{ij}\) is the \(l\)-th verification that starts from block \((i, j)\), and \(N_{ij}\) is the total number of samples that starts from the block \((i, j)\). The verification mean tendency vector is almost axisymmetric around the origin, corresponding to the regular eastward movement of the MJO signal while maintaining its amplitude. The calculation of the time evolution of the MJO state based on this mean tendency vector is a kind of statistical MJO prediction (Maharaj and Wheeler 2005). In contrast, the mean tendency vector in the 5-day lead time forecast by the JMA model is written as

\[
V^f_{ij}(\tau) = \frac{1}{N_{ij}} \sum_{l=1}^{N_{ij}} x^f(t^l_{ij}, \tau) - x^0(t^l_{ij})/	au,
\]

(2.2)

and is shown in Fig. 2.3b, where \(x^f(t, \tau)\) is the forecast-ensemble mean RMM vector with lead time \(\tau \) (\(\tau = 5\) days, as in Fig. 2.3a). The forecast mean tendency vector is also axisymmetric although its center is near \((-0.5, 0.5)\). A slight inward component is also found in the vector field. The difference in the mean tendency vector between the verification and forecast suggests that there is a systematic bias in the model and that it cannot maintain the MJO amplitude for the 5-day lead time. This fact motivates us to investigate bias in detail and remove it efficiently, to that purpose a new metric is needed.
Figure 2.3: (a) Mean tendency vector based on JRA25/JCDAS reanalysis data. (b) Mean tendency vector for the forecast ensemble mean at a 5-day lead time. The reference vector of 0.3 day$^{-1}$ is shown in the bottom right.
2.4. Evaluation methods

i. BCC and bivariate RMSE

With the same notation as in Section 2.3, BCC with lead time $\tau$ of block $(i, j)$ is defined as

$$C_{ij}(\tau) = \frac{\sum_{l=1}^{N_{ij}} X_f(t_{ij}^l, \tau) \cdot X^0(t_{ij}^l + \tau)}{\sqrt{\sum_{l=1}^{N_{ij}} \|X_f(t_{ij}^l, \tau)\|^2} \sqrt{\sum_{l=1}^{N_{ij}} \|X^0(t_{ij}^l + \tau)\|^2}}$$

(2.3)

and bivariate RMSE with lead time $\tau$ in block $(i, j)$ is defined as

$$E_{ij}(\tau) = \frac{1}{N_{ij}} \sum_{l=1}^{N_{ij}} \|X_f(t_{ij}^l, \tau) - X^0(t_{ij}^l + \tau)\|^2.$$

(2.4)

Note that block $(i, j)$ can be replaced with phase $k$. After some algebraic manipulations, the relation between the BCC and RMSE can be written as

$$E_{ij}^2(\tau) = \sigma_{ij}^2[X_f](\tau) + \sigma_{ij}^2[X^0](\tau) - 2\sigma_{ij} [X_f](\tau)\sigma_{ij} [X^0](\tau) C_{ij}(\tau),$$

(2.5)

which is the well-known cosine formula (Taylor 2001). The variance of the verification and of the forecast are written as

$$\sigma_{ij}^2[X_f](\tau) = \frac{1}{N_{ij}} \sum_{l=1}^{N_{ij}} \|X_f(t_{ij}^l, \tau)\|^2,$$

(2.6)

and

$$\sigma_{ij}^2[X^0](\tau) = \frac{1}{N_{ij}} \sum_{l=1}^{N_{ij}} \|X^0(t_{ij}^l + \tau)\|^2,$$

(2.7)

respectively. This demonstrates that BCC becomes smaller for the inactive cases of the MJO if the RMSE is constant.
ii. \textit{Mean error}

We introduce the mean error vector as a new metric to evaluate the MJO prediction skill. The mean error in block $(i, j)$ for lead time $\tau$ is defined as

$$B_{ij}(\tau) = \frac{1}{N_{ij}} \sum_{i=1}^{N_{ij}} X^f(t_{ij}^1, \tau) - X^0(t_{ij}^1 + \tau).$$

(2.8)

From Eqs. (2.1) and (2.2), we see that the mean error vector is proportional to the mean tendency vector difference between the verification and forecast:

$$B_{ij}(\tau) = \tau[V_{ij}^f(\tau) - V_{ij}^0(\tau)].$$

(2.9)

The mean error represents the damping or amplifying bias of the MJO signal in the model as an inward or outward tendency in the radial direction, respectively. The mean error also shows the slower or faster speed of the MJO signal in the MJO as a counterclockwise or clockwise tendency in the rotational direction, respectively.

Because the mean squared error, $E_{ij}^2$, is decomposed into squared mean error norm, $\|B_{ij}\|^2$, covariance of forecast and verification, variance of forecast, and variance of verification (Murphy 1988), the RMSE should have a pattern similar to the norm of the mean error vector if the mean error norm is the dominant term. The mean error vector is generally equal to the average amplitude and phase error introduced by Rashid et al. (2010).

2.5. \textbf{Results}

i. \textit{Comparison of BCC and RMSE}

Figure 2.4 shows the BCC over the MJO phase space. The BCC for a 5-day lead time exceeds 0.6 for any block in the phase space and has a minimum around (0.5, 0.3)
(Fig. 2.4a). The BCC indicates that the prediction started from Phases 8-1 is better than the prediction started from Phases 4-5. The BCC is lower near the origin of the phase space, where the MJO amplitude is smaller. The JMA forecast is the worst for a 10-day lead time, when the forecast initiated at Phases 3-4 (Fig. 2.4b). In the other phases, the BCC is larger than 0.5. The amplitude dependency of BCC becomes unclear because the difference of amplitude becomes large among forecasts from same block during the 10-day lead time.

The bivariate RMSE provides different results for the same JMA forecast (Fig. 2.5). For a 5-day lead time, RMSE ranges from 0.4 to 1.1, largely depending on the MJO phases. The RMSE exceeds 0.9 for Phases 4-5, whereas it is nearly 0.5 in Phases 8-1. For the 10-day lead time, the RMSE asymptotically approaches the climatological variance of $\sqrt{2}$, particularly in Phases 3-4. However, a global character over the phase space is maintained as in the 5-day forecast RMSE.
Figure 2.4: Bivariate anomaly correlation coefficient (BCC) for (a) 5-day lead time and (b) 10-day lead time with contour interval of 0.05. The value is plotted at the initial position of the forecast. The area where the probability density function (PDF) is less than 0.02 is masked out.
Figure 2.5: Root-mean square error (RMSE) for (a) 5-day and (b) 10-day lead time, plotted at each position for the initial forecast time. Contour interval is 0.05.
By comparing the BCC with bivariate RMSE in the JMA forecast (Figs. 2.4, 2.5), it is found that the evaluation of the prediction skill seems inconsistent in some places in the phase space. For example, in Phases 8-1, the BCC indicates good prediction skill for the 10-day prediction, whereas bivariate RMSE is as large as the climatological variance. There is a small difference in mean tendency between the reanalysis and forecast in those Phases (Fig. 2.3). In contrast, BCC and RMSE provide consistent evaluations of prediction skill in other places. For example, both indices evaluate a high skill in Phases 4-5 for a 5-day lead time. In those phases, there is a significant inward bias in mean tendency (Fig. 2.3). These results suggest that whether the BCC and RMSE provide consistent prediction skills depends on the mean tendency error between verification and forecast, when the MJO amplitude is sufficiently large.

The BCC, RMSE, and amplitudes are shown as a function of forecast lead time using the phase separation method in Fig. 2.6. For forecasts initialized in Phases 2-3, the amplitude rapidly decreases with the lead time (Fig. 2.6a). This is consistent with the inward mean tendency vector of the forecast (Fig. 2.3b). The BCC subsequently decreases, especially after a 5-day lead time, and the bivariate RMSE increases almost linearly during first 10 days of forecast. The BCC and RMSE attain the saturation levels of 0.5 and $\sqrt{2}$, respectively, at the end. For forecasts initialized in Phases 4-5, the BCC decrease follows the amplitude decrease over 5 days (Fig. 2.6b). This may be linked to the result for Phases 2-3, because the mean tendency vector indicates a possible translation of forecast MJO from Phases 2-3 to Phase 4-5 in a week. In contrast with Phases 2-5, the forecast initialized in Phases 6-7 and Phases 8-1 shows that the amplitude remains around 1.5 for a couple of weeks, and the BCC exceeds 0.8 in this period. The feature is also consistent with a small difference between verification and
forecast in two-dimensional maps (Figs. 2.4 and 2.5). This result confirms that the BCC depends strongly on the MJO amplitude.
Figure 2.6: BCC (dotted line), bivariate RMSE (dash-dotted line), and root mean squared amplitude of forecast (black solid line) and verification (gray line) as a function of forecast lead time for the forecasts initialized in (a) Phases 2-3, (b) Phases 4-5, (c) Phases 6-7, and (d) Phases 8-1.
Mean tendency in reanalysis and forecast

From the results in the previous subsection, it is considered that the error in the mean tendency, namely, the mean error, is a key to exploring the consistency between the BCC and RMSE. Figure 2.7 shows the mean error of the JMA forecast. The mean error for the 5-day and 10-day lead time has a leftward tendency distributed almost uniformly over the MJO phase space. This tendency indicates that the forecast MJO signal tends to move slower in the Indian Ocean, damps around the Maritime Continent, and moves faster compared with verification (see also Fig. 2.3). The norm of the mean error vector is largest in Phases 4-5 and smallest in Phases 8-1. This is a similar spatial pattern to the RMSE (Fig. 2.5), but with slightly smaller magnitudes of 0.2–1.0. This indicates that a large part of the RMSE is attributed to the mean error norm. Based on the relationship between the BCC and RMSE, the mean error vector provides simple classifications. When the mean error vector is directed inward to the origin of the MJO phase space, as in Phases 4-5, the BCC tends to be low due to its amplitude dependency. When the mean error vector is small, as for Phases 8-1, the BCC may remain high because the amplitude of the MJO tends to remain large. Additionally, when the MJO status is near the origin, the BCC tends to be low regardless of the mean error. Hence, the BCC can evaluate prediction skill only if the model MJO maintains the amplitude for the lead time; otherwise, it overemphasizes the damping effect. Furthermore, the BCC does not provide information about the systematic error in the model; therefore, we recommend the use of the mean error instead of the BCC. Therefore, it is concluded that the use of the mean error vector is more suitable for the evaluation of the MJO forecast than the use of the BCC.
Figure 2.7: Mean error vector field for (a) 5-day and (b) 10-day lead time. The reference vector of 1.0 day\(^{-1}\) is showed in the bottom right. Contour denotes the norm of the mean error vector and its interval is 0.1.
Figure 2.8: The PDF of the state for the ensemble forecast for (black) 1-day, (blue) 5-day, and (red) 10-day lead time. The contour interval is 0.05.
The mean error vector provides information about the temporal evolution of the PDF as a function of forecast lead time, although the PDF should be stationary in a perfect model. Reflecting the leftward movement of the mean error, the forecast PDF shifts leftward with the lead time (Fig. 8). The PDF is concentrated around its center reflecting the horizontal convergence of the mean error vector slightly east of the origin. This PDF shift also indicates that the JMA model error can be related to a leftward movement of the MJO status that is independent of phase. Because the mean error norm explains about 70% of the mean square error, the systematic bias correction is needed to enhance the usefulness of prediction skill evaluated with the bivariate RMSE.

2.6. Summary and discussion

We have assessed the mean error vector, and BCC and RMSE as indices to evaluate the MJO prediction skill. By projecting reanalysis and JMA forecast data onto the MJO phase space (Wheeler and Hendon 2004), we demonstrated that the BCC, widely used in previous studies, is not suitable when MJO signal is damped in some phases. This is because the BCC strongly depends on the MJO amplitude. This feature causes a discrepancy in prediction skill evaluation between BCC and RMSE. In the JMA forecast model, we found that the mean error vector is almost uniformly directed toward the negative phase of RMM1, independent of the MJO phase. This means that the JMA’s forecast MJO then tends to be damped around the Maritime Continent, such that the prediction skill of BCC and RMSE is poor. In contrast, when the initial location of the MJO is located in Phases 8-1, the mean error is smaller, and the prediction skill of the BCC indicates a nearly perfect prediction, even though the RMSE is larger. To
avoid this strong dependency of the BCC on the MJO amplitude, we recommend the use of mean error vector instead of BCC.

The mean error vector helps us understand how the forecast model is biased over the phase space. The leftward tendency of mean error vector for the JMA forecast is associated with the error of the circulation and convection in the tropics. The error composite of the 200 hPa zonal wind for Phases 4-5 (Fig. 9a) shows an erroneous westerly in the Indian Ocean and an erroneous easterly in the central Pacific for the 5-day forecast. This results in an erroneous downdraft in the position where convection should occur. However, we could not find an obvious error for the 850 hPa zonal wind (Fig. 9b). This implies that the error may stem from the upper troposphere. However, we need more evidence to support this idea for reducing the model bias comprehensively.

We suggest that using the mean error vector can facilitate understanding of the model bias in the multi-model ensemble. Recently, a project comparing subseasonal-to-seasonal predictions was launched to improve monsoon prediction (Brunet et al. 2010), and the MJO prediction is an important analysis target. Because some models can forecast the MJO well and maintain the amplitude, whereas others cannot, the BCC cannot explore how much worse a model prediction is, except through verification. Bias correction by using mean error may help identify initial-value predictability. Moreover, although the method is not mathematically trivial (Yoden 2007), a multi-model ensemble mean without an erroneous mean tendency has better skill in seasonal forecasting than a single-model ensemble (Krishnamurti et al. 1999; Palmer et al. 2000). It is valuable to attempt the bias correction with mean tendency error in a multi-model ensemble; this will be investigated in chapter 3.
Figure 2.9: Error composites of (a) 200 and (b) 850 hPa zonal winds for a 5-day lead time for forecasts initialized in Phases 4-5. Both are normalized by their standard deviation of (a) 5.43 and (b) 1.19 m s\(^{-1}\). The color scale is given at the bottom of the figure.
Chapter 3. An Alternative Estimate of Potential Predictability on the MJO Phase Space Using S2S Models

3.1. Introduction

The potential predictability, representing the intrinsic uncertainty due to chaotic nature of the atmosphere, was usually estimated from single-model ensemble forecast by calculating spread among members initialized with small different perturbations. However, because most of current GCMs could not sustain the convective activity of the MJO, the ensemble members tended to damp into the origin of MJO phase space. Waliser et al. (2003) focused on this damping problem and estimated the intrinsic limit of predictability of the MJO as the forecast lead time at which spread times $\sqrt{Z}$ attains the MJO amplitude. Their estimate of the predictability was about 25-30 days for 200-hPa velocity potential. Neena et al. (2014), applying the same method to each of eight GCMs, estimated that the potential predictability of the MJO is 20-30 days in the MJO phase space. They also found that the predictability of the MJO also depends on its initial phases. Although a similar metric was used for an ensemble forecast analysis, Kim et al. (2014) suggested that the potential predictability was insensitive to the initial MJO phase. The reason for these contradictory results has not been clarified yet.

Here, we want to emphasize that all of the above studies implicitly assumed that a forecast model perfectly represented the MJO time-series in their estimates. However, this perfect model assumption is not always reasonable for the MJO prediction, because
most GCMs still underestimate the MJO amplitude and then distort the climatological PDF in the MJO phase space like Fig. 2.8 (Zhang et al. 2006; Subramanian and Zhang 2014; Wang et al. 2013; Vitart 2014). If we want to develop an alternative approach to evaluate the potential predictability without the perfect-model assumption, we should extricate ourselves from model dependence in the estimation. The theoretical consideration by Kumar and Hoerling (2000) may help us relate the magnitude of the initial-value error to the statistics of analysis and imperfect-model forecast data. If one removed the model bias averaged over the forecasts (Ichikawa and Inatsu 2016), the potential predictability can be related to the manner how the forecasts are co-variated with the analyses, as will be discussed later (Kumar et al. 2014; Scaife et al. 2014; Eade et al. 2014). Recently, the sub-seasonal to seasonal (S2S) prediction project has been launched and posed the MJO forecast as one of the most important targets (Brunet et al. 2010; Vitart et al. 2012, 2017). We can access a set of hindcasts of operational GCMs, each of which has different resolution and different physical parameterizations. Multi-model estimate of potential predictability has not been established yet, because the perfect-model assumption obstructs a reasonable compilation of results. Hence a new approach without the perfect-model assumption would be applicable to multi-model ensemble analysis.

This chapter introduces an alternative method to estimate the potential predictability without assuming the perfect model. The new method is completely different from a conventional method that evaluates the spread of initial perturbations in an ensemble forecast. The aim of this chapter is to develop a strategy to cope with averaged time-series of the ensemble members, posing the analysis time-series as its counterpart. A set of analyses and forecasts projected onto the MJO phase space starting
from many various dates are then prepared. In order to achieve this purpose, we will first formulate the estimation method of potential predictability based on a set of analyses and forecasts. The method newly developed in this study is applicable to the multi-model ensemble analysis.

### 3.2. Data

#### i. Verification and phase space

The verification in this chapter is an average of ERA-Interim of ECMWF (Dee et al. 2011), JRA55 of the JMA (Kobayashi et al. 2015), and the Climate Forecast System (CFS) reanalysis version 2 of NCEP (Saha et al. 2010). Presumably because of sparseness of surface observation and deficit of model parameterizations, a discrepancy among reanalyses is not negligible in the tropics (Trenberth et al. 2011; Kim and Alexander 2013). While most of previous works evaluated GCMs’ performance with their own reanalysis as the verification, averaging reanalyses enables us to fairly compare the prediction among models (Park et al. 2008), and is expected to be a more optimal estimate than a single analysis (Wei et al. 2010).

The MJO phase space is spanned by two gravest combined-EOF modes of zonal wind at 200 hPa and at 850 hPa averaged between 15°S and 15°N based on NCEP’s CFS reanalysis version 2 from 1979 to 2015 (Wheeler and Hendon 2004). If the data are replaced with the reanalysis average, the MJO phase space does not change utterly. We excluded the outgoing longwave radiation from the combined EOF analysis for simplicity, since the two gravest EOF modes with and without outgoing longwave radiation are not different largely (Matsueda and Endo 2011; Straub 2013). Climatology
and 120-day average were removed from zonal wind prior to the EOF analysis to extract the intraseasonal variations, and the signal of the El Niño/Southern Oscillation was automatically removed in this processing (Lin et al. 2008). We classified the MJO phase as convention and the non-MJO phase is defined as the state where the MJO amplitude is less than unity.

ii. **Forecast Data**

We used reforecast data of three GCMs joining the S2S project ([http://s2sprediction.net/](http://s2sprediction.net/)): Operational models of ECMWF (Vitart 2014), GSM1403C of the JMA (Japan Meteorological Agency 2013) and CFS version 2 of NCEP (Saha et al. 2014). The details are provided in Table 3.1. Hereafter each model is referred to by organization name (e.g. the JMA model for GSM1403). The period of data is from Jan 1999 to Dec 2009. Model forecast data are also projected on the phase space, after a removal of the seasonality and 120-day running average using the reanalysis provided with the corresponding model. In S2S reforecast data, an initial perturbation was created using singular vector and ensemble data assimilation perturbation by ECMWF, using bred vector and lagged averaging (LAF) method by the JMA, and using LAF method for NCEP.
Table 3.1: Subseasonal-to-seasonal (S2S) forecast models selected in this model.

<table>
<thead>
<tr>
<th>Organization</th>
<th>ECMWF</th>
<th>JMA</th>
<th>NCEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecast period</td>
<td>32 days</td>
<td>33 days</td>
<td>44 days</td>
</tr>
<tr>
<td>Initial perturbation</td>
<td>Singular vector and ensemble data assimilation perturbation</td>
<td>Bred vector and lagged averaging method</td>
<td>Lagged averaging method</td>
</tr>
<tr>
<td>Ensemble size</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Initial time</td>
<td>00 UTC</td>
<td>12 UTC</td>
<td>00 UTC</td>
</tr>
<tr>
<td>Frequency</td>
<td>Twice a week</td>
<td>3 times a month</td>
<td>Everyday</td>
</tr>
<tr>
<td>Total number of forecasts</td>
<td>1,397</td>
<td>396</td>
<td>3,972</td>
</tr>
<tr>
<td>Horizontal resolution</td>
<td>TL639 (TL319 after day 10)</td>
<td>TL319</td>
<td>T126</td>
</tr>
<tr>
<td>Vertical layers</td>
<td>91</td>
<td>60</td>
<td>64</td>
</tr>
</tbody>
</table>
A stumbling block when using the reforecast data joining the S2S project is a different interval of initial date and time, which makes it difficult to find a set of models initialized on the same date. The ECMWF model has 1,397 forecasts that were initialized at 00 UTC of the calendar date corresponding Thursdays between Jan 2015 and May 2016 or Mondays from 14 May 2015 to 30 May 2016. The JMA model has 396 forecasts that ran from 12 UTC of 10th, 20th and the last day of each month. The NCEP model has 3,972 forecasts started from 00 UTC of every date during the period. For the multi-model ensemble forecast analysis, reforecasts of ECMWF, JMA, and NCAR are selected only if the difference of initial times is less than 12 hours; for example JMA model’s reforecast initialized on 20th March at 12 UTC is grouped into the multi-model analysis with ECMWF and NCEP models’ reforecasts initialized on 21th March at 00 UTC. This selection is effective to match the initial date among the forecasts, and we realized multi-model ensemble analysis with 250 independent forecasts. The list of the calendar dates is provided in Table 3.2, and the number of single model and multi-model forecasts initialized in each MJO phase are shown in Table 3.3. It is noted that the forecasts initialized on 1 Jan 1999, 1 Feb 2003, and 11 Feb 2003 are excluded for a technical reason.
Table 3.2: A list of months and days on which the multi-model ensemble analysis is performed in 1999-2009.

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan</td>
<td>1</td>
<td>11</td>
<td>21</td>
</tr>
<tr>
<td>Feb</td>
<td>1</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Mar</td>
<td>10</td>
<td>21</td>
<td>31</td>
</tr>
<tr>
<td>Apr</td>
<td>11</td>
<td>21</td>
<td>30</td>
</tr>
<tr>
<td>May</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jun</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jul</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aug</td>
<td>10</td>
<td>20</td>
<td>31</td>
</tr>
<tr>
<td>Sep</td>
<td>10</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Oct</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nov</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dec</td>
<td>10</td>
<td>21</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.3: The number of individual forecasts, each with its own forecast period starting from each piecewise area in the Madden-Julian Oscillation (MJO) phase space.

<table>
<thead>
<tr>
<th>Phase Category</th>
<th>ECMWF</th>
<th>JMA</th>
<th>NCEP</th>
<th>Multi-model ensemble</th>
</tr>
</thead>
<tbody>
<tr>
<td>All MJO phase</td>
<td>827</td>
<td>236</td>
<td>2,311</td>
<td>143</td>
</tr>
<tr>
<td>Phases 2-3</td>
<td>182</td>
<td>51</td>
<td>511</td>
<td>41</td>
</tr>
<tr>
<td>Phases 4-5</td>
<td>218</td>
<td>64</td>
<td>638</td>
<td>43</td>
</tr>
<tr>
<td>Phases 6-7</td>
<td>196</td>
<td>58</td>
<td>529</td>
<td>31</td>
</tr>
<tr>
<td>Phases 8-1</td>
<td>231</td>
<td>63</td>
<td>633</td>
<td>28</td>
</tr>
<tr>
<td>Non-MJO phase</td>
<td>570</td>
<td>160</td>
<td>1,661</td>
<td>107</td>
</tr>
</tbody>
</table>
iii. An alternative method to estimate potential predictability

Consider the time-series of the true value, \( X^o \), and the perfect model hypothetically (Fig. 3.1). We first clip a part of the time-series \( X^o \) in the range from \( t_0 \) to \( t_0 + T \). This time segment is renamed as \( X_1^o(\tau) \) for \( 0 \leq \tau \leq T \). We numbered the time segment (see below) in the subscript. In addition, the superscript \( o, f, \) and \( e \) mean the truth, the ensemble mean, and the initial-value error, respectively, almost following the notation of the data assimilation literature (Ide et al. 1997)]. The perfect model, as a thought experiment, is repeatedly integrated from \( X_1^o(0) \) added with initial error, the magnitude of which is infinitesimally small and randomly given. If an average of infinite ensemble members of the above experiment is taken, we obtain the time segments of ensemble mean perfect-model forecast written as \( X_1^f \) here. We now emphasize that the nonlinearity intrinsic to the system makes the difference between \( X_1^o \) and \( X_1^f \) for \( \tau > 0 \). This difference, expanding in time, is regarded as initial-value error \( X_1^e(\tau) = X_1^o(\tau) - X_1^f(\tau) \). As \( \tau \) goes to infinity, the ensemble members of the perfect model scatter over the phase space, following the climatological probability density function (PDF) of the true value. The ensemble-mean \( X_1^f(\tau) \) is asymptotic to the origin of the phase space, while the true value \( X_1^o(\tau) \) does not converge to a certain value as \( \tau \to \infty \).
Figure 3.1: A schematic of time evolution of a single time segment of the true value and perfect model’s ensemble forecast. (Solid line and closed circles) True value and (open circles) the ensemble-mean perfect-model forecast from the initial to the infinite lead time $\tau$. The forecast expansion of the ensemble members is represented by the thick lines enclosed by their mean.
Next, we labeled a set of time segments of the true value as
\[ \{X_1^o(\tau), X_2^o(\tau), \ldots, X_N^o(\tau)\}, \quad (3.1) \]
and that of corresponding ensemble-mean forecasts as
\[ \{X_1^f(\tau), X_2^f(\tau), \ldots, X_N^f(\tau)\}, \quad (3.2) \]
where \( N \) is the total number of selected time segments. As \( X_j^f(\tau) \) is uncorrelated with the initial-value error \( X_j^e(\tau) \) for all time segments \( 1 \leq j \leq N \), the variance of the true values among a set of time segments [Eq. (3.1)], \( \sigma^2[X^o(\tau)] \), can be represented by the variance of the ensemble-mean forecasts among a set of time segments [Eq. (3.2)], \( \sigma^2[X^f(\tau)] \), added by the initial-value error spread (DelSole and Feng 2013; Williams et al. 2014; Kumar et al. 2014):
\[ \sigma^2[X^o(\tau)] = \sigma^2[X^f(\tau)] + \sigma^2[X^e(\tau)]. \quad (3.3) \]
At \( \tau = 0 \), \( \sigma^2(X^o) \) is equal to \( \sigma^2(X^f) \) as the initial-value error should be zero. As \( \tau \) approaches to infinity, \( \sigma^2(X^o) \) becomes the variance of climatological PDF. For this limit, since \( X_j^f \to 0 \) for all time segments \( j \), the variance is exactly zero. Now we introduce the anomaly correlation coefficient (ACC) between true value and ensemble-mean forecast as
\[ r = \frac{1}{N - 1} \sum_{j=1}^{N} \frac{(X_j^o - \langle X^o \rangle) \cdot (X_j^f - \langle X^f \rangle)}{\sigma(X^o) \sigma(X^f)}, \quad (3.4) \]
where the angle bracket denotes the average over the time segments. The ACC is high if the true value is well co-variated with the forecast (Fig. 3.2a); otherwise, it is low (Fig. 3.2b). By using the relation \( X_j^o(\tau) = X_j^f(\tau) + X_j^e(\tau) \) and the independence between the sets of forecasts and initial-value error, the ACC is reduced to the ratio of the standard deviation of forecast to that of the true value:
Figure 3.2: A schematic describing forecast groups with (a) high and (b) low anomaly correlation coefficient skill. The open circles are the forecasts and the filled circles are the verifications. The number on them denotes the time segment label.
Finally we relieve the assumption that $X_f$ would be created with the perfect model. If $X_f$ is redefined as the ensemble-mean of an imperfect model forecast, we add one extra term of covariance between forecast and error, $C(X_f, X^e)$, as

$$r = \frac{\sigma(X_f)}{\sigma(X^o)} + \frac{C(X_f, X^e)}{\sigma(X^o)\sigma(X_f)}. \tag{3.6}$$

Because the initial-value error deteriorates the forecast, the extra cross-correlation term is in general negative. Therefore the equality in Eq. (3.6) is replaced with the inequality as below (Eade et al. 2014):

$$r \leq \frac{\sigma(X_f)}{\sigma(X^o)}. \tag{3.7}$$

The ACC is unity at the initial time and almost monotonically decreases with the lead time. Within the range of a positive ACC, using Eq. (3.3), we can derive the inequality that determines a possible maximum value of initial-value error based on a set of time segments of the true value and the ensemble-mean forecasts as

$$\sigma(X^e) \leq \sigma(X^o)\sqrt{1 - r^2}. \tag{3.8}$$

Thus the prediction limit can be evaluated as the time when $\sigma(X^o)\sqrt{1 - r^2}$ reaches a threshold value. On the MJO phase space, we use the threshold value of unity, that is about 70 percent of the climatological variance. It is worthwhile remarking that the new method only uses the ensemble-mean forecast data unlike the conventional method that uses each ensemble member. Expanding Eq. (3.8) to multi-model analysis with $M$ models, the inequality should be replaced with $\sigma(X^e) \leq \sigma(X^o)\sqrt{1 - R^2}$. Here $R$ is the maximum ACC among the models and their mean, because our evaluation only provides the lower bound of the predictability.
For a practical use of the above theoretical consideration in this paper, we replace the true value with verification defined in Section 3.2, say the reanalysis average. In this case, the discrepancy in the initial state between verification and ensemble-mean forecast was not negligible. We hence constantly offset the time segment of forecast $X_j^f(\tau)$, so as to match its initial state $X_j^f(0)$ with the verification $X^o_j(0)$ for each time segment:

$$X_j^f(\tau) = X_j^f(\tau) - (X^f(0) - X^o(0)).$$

(3.9)

Furthermore, a bias correction is made, the details of which will be explained in next section. Since we focus on the MJO behavior, $X^o$ and $X^f$ are hereafter two-dimensional vector on the MJO phase space.

A caution is made that the ACC introduced by (3.4) is different from the BCC (2.3), in chapter 2, that was often used for the evaluation of prediction skill on the MJO phase space (Ichikawa and Inatsu 2016). The BCC is

$$\rho = \frac{1}{N - 1} \sum_{j=1}^{N} \frac{X^o_j \cdot X^f_j}{s(X^o) s(X^f)}$$

(3.10)

where $s(X)$ is mean amplitude $\sqrt{1/N \sum_{j=1}^{N} |X_j|^2}$. The BCC is formerly related to the ACC as

$$\rho = r + \frac{1}{N - 1} \sum_{j=1}^{N} \frac{\langle X^o \rangle \cdot \langle X^f \rangle}{s(X^o) s(X^f)}$$

(3.11)

only when the lead time is sufficiently large so that $\sigma(X) = s(X)$. However, the BCC is defined in the zero-mean RMM state as the background. Since we will evaluate the predictability for forecasts initialized on each blocks of the MJO phase space, the mean
of state is generally non-zero. Hence the BCC cannot simply be related to calculate the initial-value error in the way described in our formulation.

3.3. Results

i. The prediction skill and bias correction

This section evaluates the predictability estimated using the new method developed in the above. Information of a maximum possible value of initial-value error is provided based on a pair of analysis and ensemble-mean forecast datasets. To effectively test the new method, the model bias is removed from the forecast data in advance. Hence we first discuss the prediction skill for each model and then make the bias correction by removing the bias vector averaged over each MJO phase. The prediction skill is now evaluated as error on the MJO phase space. Each separated block in figures of this paper corresponds to the MJO phase: the convective center is located over the Indian Ocean in Phases 2-3, the Maritime Continent in Phases 4-5, and the western Pacific in Phases 6-7, and Phases 8-1 correspond to suppressed or formation stages. Figure 3.3 displays the prediction skill at 10-day lead time calculated for each forecast as the dot plot at the initial-state position on the MJO phase space, with the root-mean-square error (RMSE) as the statistics among forecasts. The best model was ECMWF’s, with the RMSE less than unity at this lead time. The RMSE is 1.07 for NCEP model and 1.15 for JMA’s. These two models (Figure 3.3b, c) especially provided erroneous forecasts with the RMSE exceeding 1.4 when they are initialized in Phases 4-5, when convection is active over the Maritime Continent. In contrast, ECMWF model predicted the MJO state more
accurately when initialized there, while it provided erroneous forecasts initialized in Phases 8-2, when the signal resides between Africa and central Indian Ocean.
Figure 3.3: The prediction skill measured by two-dimensional error vector on the MJO phase space for each forecast at 10-day lead time plotted at the initial-state position for (a) European Centre for Medium range Weather Forecasting (ECMWF), (b) the Japan Meteorological Agency (JMA), and (c) National Centers for Environmental Prediction (NCEP) models. The NCEP model plot is every 10 days for clarity. The magnitude of error is represented with gray scale as per the reference in the right. The MJO phases are labeled and partitioned by with dashed lines in the plot. The number of forecasts and the root-mean-square error among all forecasts are provided in the top of the panels.
The model bias as a function of lead time is defined for each initial phase as the bias vector averaged over the forecasts starting from each phase (Fig. 3.4). The bias vector at 10-day lead time of the JMA model shows a clockwise tendency with a leftward translation for all phases (Fig. 3.4b), moderately consistent with Ichikawa and Inatsu (2016). The bias vector of the NCEP model tends to direct toward the positive direction of the second principal component with a weak clockwise motion (Fig. 3.4c). The bias vector of ECMWF is characterized by clockwise rotation uniformly over the space, corresponding to a bias of slower eastward propagation of the MJO (Fig. 3.4a). The bias correction was made for each model forecast, by subtracting the bias vector shown in Fig. 3.4 from the original forecast state vector. Figure 3.5 shows the error plot after the bias correction. By the bias correction the error becomes phase-independent commonly in three models. The RMSE decreases by 0.1 for ECMWF and NCEP models (Fig. 3.5a, c). In the JMA model, the RMSE dropped from 1.15 to 0.97 as a result of the bias correction (Fig 3.5b).
Figure 3.4: The error vector at 10-day lead time averaged over forecasts initially located in each MJO phase for (a) ECMWF, (b) JMA, and (c) NCEP models. The magnitude of error vector is as measured by the scale (bottom right).
Figure 3.5: Same as Fig. 3.3, but the error based on the bias-corrected data.
ii. Prediction limit and its comparison with a conventional method
Predictability is evaluated as a prediction limit in this paper. The proposed method provides a maximum possible value of initial-value error variance as a function of the lead time. Therefore, it provides a minimum possible value of the prediction limit defined as the lead day just before the initial-value error variance attains unity, half of climatological variance of the data on the 2-dimensional phase space. As the time goes to infinity, the variance of analysis time segments is asymptotic to 2. The new method in essence evaluates the prediction limit when a possible maximum of initial-value error times $\sqrt{2}$ reaches this limitation. This is analogous to the conventional method in which the prediction limit is defined as the time when the ensemble spread times $\sqrt{2}$ reaches the MJO amplitude. It is worthwhile noting that $\sigma(X^0) \rightarrow \sqrt{2}$ can be used in the new method but the spread goes to the MJO amplitude in the limit due to model’s imperfectness (Neena et al. 2014). In order to be fairly compared with the method proposed in this paper, the prediction limit is also evaluated following Neena et al., except that we did not take a 51-day running average for amplitude due to the limited length of prediction. As the new method gives the minimum possible value of prediction limit, it can be said that the conventional method provides a fake short limit when it is shorter than the new method estimation.

We first test the new method in light of the classification of forecasts on the initial MJO amplitude. Averaging over forecasts of which the initial MJO amplitude exceeds unity, a minimum possible value of prediction limit evaluated by the new method is 14 days based on ECMWF model, and 10 days based on JMA and NCEP models (Figs. 3.6a-c). In contrast, averaging over forecasts starting from non-MJO state, a minimum-possible prediction limit is 12 days based on ECMWF model, 9 days based
on JMA’s, and 10 days based on NCEP’s (Figs. 3.6d-f), slightly shorter than forecasts from MJO states. The result suggests that the forecasts starting from MJO state is more predictable in average than those starting from non-MJO state. This is consistent with the consequence from the conventional method. The prediction limit in ECMWF model is estimated as 23 days for forecasts from MJO state, while it is 18 days for forecasts from non-MJO state (Fig. 3.6a, d). Similarly, the prediction limit for the NCEP model is 23 days for initial MJO state and 21 days for initial non-MJO state (Figs 3.6c, e). Although the conventional method did not provide the prediction limit for the JMA model due to no crossing of amplitude and spread lines, the prediction limit seems shorter in forecasts with initial non-MJO state (Fig 3.6d). This result was not substantially changed even if we classified forecasts by the average MJO amplitude through forecast period, and we again found a longer prediction limit for the forecasts for which average amplitude exceeding unity (not shown).
Figure 3.6: (a-c) (Solid line) MJO amplitude, (dashed line) ensemble spread times $\sqrt{2}$ on the MJO phase space, and (dotted line) a maximum possible value of initial-value error evaluated with the method proposed in this paper, averaged over forecasts of which the MJO amplitude initially exceeds unity for (a) ECMWF, (b) JMA, and (c) NCEP models. The prediction limit (day) is denoted as the vertical line for the new and conventional methods if any. (d-f) Same as (a-c), but for forecasts from the initial state with the MJO amplitude less than unity.
Our method can be utilized to estimate the dependence of MJO predictability on the initial phase. Here the results are shown in Figure 3.7 for NCEP model and Table 3.4 for ECMWF model. A minimum-possible prediction limit based on NCEP model is longest for forecasts from Phases 6-7 and shortest for forecasts from Phases 2-3. This phase dependence is coherent with the prediction limit estimated with ensemble spread and the low prediction limit for forecasts initialized in Phases 2-3 may be due to an irregular amplitude change over the Maritime Continent. The ECMWF model provides a minimum-possible prediction limit at 27 days for forecasts from Phases 4-5, much longer than that for forecasts from other phases. This is likely because there is little possibility to initiate another MJO event under the dry environment that occurs after the MJO passage over the Indian Ocean and the Maritime Continent. This is also in line with the prediction limit estimated with the conventional method, though the difference among phases is not so prominent.
Figure 3.7: (Solid line) MJO amplitude, (dashed line) the ensemble spread times $\sqrt{2}$, and (dotted line) a maximum possible value of the initial-value error for NCEP model, averaged over forecasts initially starting from (a) Phases 2-3, (b) Phases 4-5, (c) Phases 6-7, and (d) Phases 8-1.
Table 3.4: The list of prediction limit for ECMWF model (day) for forecasts from each MJO phase with the proposed and conventional methods.

<table>
<thead>
<tr>
<th>Initial state</th>
<th>Proposed method</th>
<th>Conventional method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phases 2-3</td>
<td>15</td>
<td>17</td>
</tr>
<tr>
<td>Phases 4-5</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>Phases 6-7</td>
<td>16</td>
<td>25</td>
</tr>
<tr>
<td>Phases 8-1</td>
<td>14</td>
<td>17</td>
</tr>
</tbody>
</table>
iii. The multi-model ensemble analysis

The multi-model analysis combines the estimates of prediction limit from each of single models and multi-model average. Multi-model average is expected, in general, to provide a better forecast than the single model. Figure 3.8 shows the prediction skill of forecasts averaged among ECMWF, JMA, and NCEP models after a bias correction. Although the sample number is only 250 as pointed out in section 3.2, the RMSE is less than any other models’ (Fig. 3.5), and the multi-model analysis gives a longer estimate of potential predictability than the single model analysis. Considering this point, we now regard the multi-model average as one of the model results, and a minimum-possible prediction limit is defined as the lead day just before the smallest estimate of initial-value error among the models reaches unity, in order to get as a long minimum value as possible. The new method can extend to multi-model analysis as this manner, while the conventional method cannot.
Figure 3.8: Scatter plot of the prediction skill at the 10-day lead time for each forecast based on the multi-model average.
A minimum-possible prediction limit in the multi-model analysis is shown in Figure 3.9. For forecasts of which the initial amplitude exceeds unity, the smallest estimate of initial-value error reaches unity at 18 days (Fig. 3.9a). For forecasts starting from non-MJO state (Fig. 3.9b), in contrast, a minimum-possible prediction limit is 12 days. The multi-model analysis also supports the notion that the potential predictability for forecasts from MJO state is longer than that from non-MJO state. Classified the initial state like the previous section, we investigate the dependence of the potential predictability on MJO phase in the multi-model analysis. A minimum-possible prediction limit for forecasts initialized in Phases 4-5 is 28 days, while that for forecasts initialized in other phases is 15-18 days. This is because the variance $\sigma^2(X_0)$ is so small that the initial-value error can keep less than unity even with the relatively low ACC.
Figure 3.9: (a, b) Initial-value error for (mark E) ECMWF, (mark J) JMA, and (mark N) NCEP models and (mark M) multi-model average, averaged over forecasts of which the initial MJO amplitude is (a) more than and (b) less than unity. (c-f) Initial-value error averaged over forecasts starting from (c) Phases 2-3, (d) Phases 4-5, (e) Phases 6-7 and (f) Phases 8-1.
It is cautioned that there is an estimation error in the evaluation above caused by the small sample size in multi-model analysis. For example, there are only 31 and 28 forecasts initialized in Phases 6-7 and Phases 8-1, respectively (Table 3.3). In these cases, the estimation error of the initial-value error could be as large as ~0.05-0.08 around the lead time when it reaches unity, according to the boot-strap method (not shown).

3.4. Concluding remarks

This study has purposed an alternative method to estimate potential predictability without making perfect model assumption. The formulation started from the definition of the initial-value error as the difference between analysis and ensemble-mean forecast. Their covariance among a bundle of time segments is a key to develop the new method. Applying it to the two-dimensional state on the MJO phase space, the potential predictability for three operational models participating the S2S project was evaluated. It was found that forecasts from MJO state had a longer prediction limit in average than forecasts from non-MJO state. The new method also detected a longer prediction limit for forecasts initialized in Phases 4-5. We obtained consistent results from a conventional method in which the initial-value error was estimated with the spread of ensemble members normalized by the amplitude of MJO signal. Though we obtained a similar result to the conventional method, our new method is superior to the conventional method in three points: it can be used for phenomena that are not well forecasted by models; it can be used for multi-model ensemble analysis; and it can identify a fake short prediction limit obtained using the conventional method. Another different point from the conventional method is the use of the ensemble spread. The
ensemble members or their spread might have more information than their mean. Though they might give some statistical stability, this point is actually beyond the scope of this paper and will be reported elsewhere.

We have already compared the new method with the conventional method, though the former only gives a minimum possible value of prediction limit. As reviewed in Introduction, several publications were devoted to exploring potential predictability related to the MJO with the conventional method. Even with different metrics or different models, Waliser et al. (2003) and Neena et al. (2014) found that the prediction limit for forecasts from MJO state is a few days longer than that for forecasts from non-MJO state, quite consistent with our result. In terms of the phase-dependency, however, our result that an initial state in Phases 4-5 is preferable to predictability is different from previous studies: Neena et al. (2014) found slightly higher potential predictability for forecasts initialized in Phases 2-3, and Phases 6-7; Kim et al. (2014) indicated little dependence of potential predictability on the initial phase. We consider that more forecast samples are necessary to get stable statistics about the phase dependency. Moreover, if one used the precipitation pattern as the metric to estimate predictability, the prediction limit would be much shorter (Waliser et al. 2003). Similarly, the prediction limit is only 4 days for unfiltered zonal wind that contains small scale variability (Jones and Dudhia 2017). It should be kept in mind that our estimate focused only on the planetary-scale aspect of MJO that was well represented by the projection onto the two-dimensional phase space.

Although the phase dependency on potential predictability is still unsolved, the comparison with prediction skill is interesting because it indicates the possibility of improving numerical models. At least in our result based on ECMWF model and multi-
model analysis, forecasts from Phases 4-5 provide a longer prediction limit. In contrast, they provide a low prediction skill, which is probably related to the so-called "maritime continent barrier", which prevents the model MJO from propagating eastwards beyond the Maritime Continent (Vitart 2014; Matsueda and Endo 2011; Wang et al. 2013). This difference of phase dependency between prediction skill and prediction limit in Phases 4-5 suggests that “maritime continent barrier” is not related to the intrinsic prediction limit but to the insufficient representation of the MJO in numerical models. This is in line with the suggestion by Neena et al. (2014), Kim et al. (2014) and Seo et al. (2009) with different analysis from ours. We finally discuss a possible effect of the initial variance of time segments to prediction limit estimation in this study. Since the initial standard deviation of the state, \( \sigma(X^o) \), is initially 1.6 in the MJO state and 0.7 in the non-MJO state (Fig. 3.10), the latter case allows a smaller ACC if \( \sigma(X^o) \) slowly grows in the lead time. However, \( \sigma(X^o) \) for the forecasts initialized with the MJO and non-MJO states rapidly converge to the same value around 14 days, which is related to the system memory over the phase space. The prediction limit is solely decided by ACC after that. Because the ACC was higher for forecasts from MJO state (not shown), its prediction limit turns out to be slightly longer than that from non-MJO state, in spite of a larger \( \sigma(X^o) \) in the initial time. Hence, the difference in initial amplitude has little effect to prediction limit estimation.
Figure 3.10. The standard deviation of analysis time segments as a function of lead time (black) for forecasts initialized at the MJO state and (gray) for forecasts initialized at the non-MJO state.
Chapter 4. General summary

This dissertation investigated evaluation methods of prediction skill and predictability in the MJO phase space. In Chapter 2, typical evaluation methods of prediction skill in the MJO phase space was investigated, and to compensate their drawbacks, the use of mean error vector instead of BCC was recommended. The prediction skill has been commonly evaluated with the combination of RMSE and BCC, though neither of them indicates how the model is biased, namely whether forecasted MJO tends to be too strong or too weak, moving too fast or too slow. Moreover, BCC has singularity at the origin. Because BCC evaluates the phase error, it becomes low when the forecasted signal resides around the origin, where small fluctuation of location could result in large phase difference. Drawbacks from these characteristics were demonstrated by evaluating the JMA forecast with JRA25/JCDAS reanalysis. The mean error vector that we proposed was found to be useful to indicate how the model is biased and it related RMSE to bias. The JMA model had leftward error vector everywhere in the MJO phase space, with its amplitude explaining 70 % of RMSE. The leftward tendency in mean error for the JMA forecast was associated with the error of the circulation and convection in the tropics.

In Chapter 3, we proposed an alternative method to estimate potential predictability without making perfect model assumption. Based on a theoretical consideration, the initial-value error was related to ACC between analysis and ensemble-mean forecast. It was found that the estimated prediction limit was generally similar between proposed and conventional method. The new method indicated that the forecasts initialized in Phases 4-5 are the most predictable, suggesting the difficulty in predicting the passage of the MJO over the Maritime Continent is not related with
initial-value error. It is also noted that recent operational prediction models do not show apparent damping of the MJO signal and that multi-model ensemble of ECMWF, JMA and NCEP models performed better than single models.
Acknowledgments.

My deepest gratitude goes to Prof. Masaru Inatsu; without his help, this dissertation could not be completed. I would like to thank deeply to Prof. Shoshiro Minobe, Dr. Yoshinori Sasaki and Dr. Hanna Na, who has encouraged me throughout my Ph. D. program. I am greatly indebted to Prof. Hitoshi Mukougawa for his advises and help. I also thank him for providing model output from JMA’s one-month forecast. Advises from Prof. Yukari Takayabu, Prof. Masahide Kimoto, Prof. Masaki Sato, Prof. H. H. Hendon, Dr. T. Nakazawa, Mr. S. Maeda, Dr. S.-W. Son and Dr. C. Jones was invaluable in completing the research works on which this dissertation is based. Discussion with Dr. H. Miura, Prof. K. Yasunaga, Dr. A. Hamada., Dr. S. Yokoi, Dr. T. Miyakawa, Ms. T. Suematsu, Mr. D. Takasuka and Mr. S. Matsugishi has greatly improved our early work. Mr. A. Shimpo and Mr. A. Takaya introduced the S2S project and the reference of the JMA operational model. I am most grateful for support from Mr. Jinji Koike and colleagues at Tokyo Aviation Weather Service Center while I was completing my Ph.D. program. Finally, I would like to thank to the members of Laboratory of Meteorology, Hokkaido University for their encouragements.
References


