

Multivariate Regression based Forecast Model for Intraseasonal Oscillation

A Thesis

Submitted to

Indian Institute of Science Education and Research Pune
in partial fulfillment of the requirements for the
BS-MS Dual Degree Programme

by

Daman



Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

April, 2019

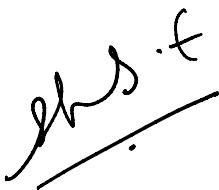
Supervisor: Dr. Suhas Ettammal
Department of Earth and Climate Science,
IISER Pune

© Daman. All rights reserved.

DEDICATED TO MY BELOVED PARENTS AND FRIENDS....

CERTIFICATE

This is to certify that this dissertation entitled "**Multivariate Regression based Forecast Model for Intraseasonal Oscillation**" towards the partial fulfilment of the BSMS dual degree programme at the Indian Institute of Science Education and Research, Pune, represents study/work carried out by Daman at Indian Institute of Science Education and Research, Pune, under the supervision of Dr. Suhas Ettammal, Assistant Professor, Earth and Climate Science, during the academic year 2018-2019.



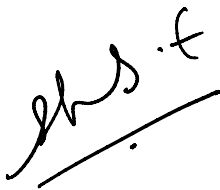
Dr. Suhas Ettammal
(Supervisor)



Daman
Reg. No. : 20141082

DECLARATION

I hereby declare that the research work presented in the report entitled ***“Multivariate Regression based Forecast Model for Intraseasonal Oscillation”*** have been carried out by me at the Earth and Climate Science, Indian Institute of Science Education and Research, Pune, under the supervision of Dr. Suhas Ettammal and the same has not been submitted elsewhere for any other degree.



Dr. Suhas Ettammal
(Supervisor)



Daman
Reg. No. : 20141082

ACKNOWLEDGEMENTS

I wish to thank and express my earnest gratitude to my guide Dr. Suhas Ettammal, Dept. of Earth and Climate Science, IISER Pune for his encouraging and valuable suggestion, discussion, never failing inspiring guidance throughout the year long project.

I would like to express my heartfelt thanks to Dr. Neena Joseph Mani, for providing me guidance throughout the project and giving me valuable suggestions and inputs.

I would like to thank my lab members: Daniel Simon, Salam Maheshwori Devi, Shreya Keshri, Dilip ,Snigdha Samantaray, Meera Mohan, Soumya Bohidar, Niyor Borah, Arijeet Dutta, Deepak Suryavanshi to help in conducting various experiments in the projects and giving valuable inputs.

I would like to thank IISER Pune and MHRD, govt. of India for providing me with computational facilities and financial assistance to carry out my computational experiments and analysis of the same.

I would like to thank my brothers Aman Bhonsle and Bhuneshwar Paswan and friends: Anurag Singh, Vaibhav Singh, Mangesh Daspute, Snigdha Samantaray, Prakash Kumar, Jainendra Singh, Akshat Yadav, Sagar Gupta and my family members for their guidance and support throughout the project.

CONTENTS

1. Abstract.....	1
2. Introduction.....	2
3. Data and Methods.....	6
3.1 Data.....	6
3.2 Methods.....	6
3.3 Model skill analysis metrics.....	9
4. Results and Discussion.....	11
4.1 Multivariate Regression model (MVR) forecast using RMM index.....	11
4.2 Skill analysis for RMM.....	24
4.3 MVR model using OMI index.....	27
4.3.1 Skill analysis for OMI Index.....	31
4.4 MVR model using MISO index.....	34
4.4.1 Skill analysis for MISO	38
5. Conclusion.....	41
6. References.....	42

List of Figures

Figure 1. Temporal correlation evolution within observed and t-model forecasted OLR anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal summer.

Figure 2. Temporal correlation evolution within observed and t-model forecasted OLR anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal winter.

Figure 3. Temporal correlation evolution within observed and j-model forecasted OLR anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal summer.

Figure 4. Temporal correlation evolution within observed and j-model forecasted OLR anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal winter.

Figure 5. Temporal correlation evolution within observed and t-model forecasted u850 anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal summer.

Figure 6. Temporal correlation evolution within observed and t-model forecasted u850 anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal winter.

Figure 7. Temporal correlation evolution within observed and j-model forecasted u850 anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal summer.

Figure 8. Temporal correlation evolution within observed and j-model forecasted u850 anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal winter.

Figure 9. Temporal correlation evolution within observed and t-model forecasted u200 anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal summer.

Figure 10. Temporal correlation evolution within observed and t-model forecasted u200 anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal winter.

Figure 11. Temporal correlation evolution within observed and j-model forecasted u200 anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal summer.

Figure 12. Temporal correlation evolution within observed and j-model forecasted u200 anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal winter.

Figure 13. Anomaly correlation Coefficient (ACC) is plotted against number of lead days for j-model (green color) and t-model (purple color) for Boreal winter (Left) and Boreal Summer (Right) calculated using RMM indices.

Figure 14. Root Mean Square Error (RMSE) is plotted against number of lead days for j-model (green color) and t-model (purple color) for Boreal summer (Left) and Boreal winter (Right) calculated using RMM indices.

Figure 15. Phase Error is plotted against number of lead days for j-model (green color) and t-model (purple color) for Boreal summer (Left) and Boreal winter (Right) calculated using RMM indices.

Figure 16. Temporal correlation evolution within observed and t-model forecasted OLR anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal summer.

Figure 17. Temporal correlation evolution within observed and t-model forecasted OLR anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal winter.

Figure 18. Temporal correlation evolution within observed and j-model forecasted OLR anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal summer.

Figure 19. Temporal correlation evolution within observed and j-model forecasted OLR anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal winter.

Figure 20. Anomaly correlation Coefficient (ACC) is plotted against number of lead days for j-model (green color) and t-model (purple color) for Boreal summer (Left) and Boreal winter (Right) calculated using OMI indices.

Figure 21. Root Mean Square Error (RMSE) is plotted against number of lead days for j-model (green color) and t-model (purple color) for Boreal summer (Left) and Boreal winter (Right) calculated using OMI indices.

Figure 22. Phase Error is plotted against number of lead days for j-model (green color) and t-model (purple color) for Boreal summer (Left) and Boreal winter (Right) calculated using OMI indices.

Figure 23. Temporal correlation evolution within observed and t-model forecasted precipitation anomalies are plotted for lead days from 3 to 15 with an increment of 3 days for boreal summer.

Figure 24. Temporal correlation evolution within observed and t-model forecasted precipitation anomalies are plotted for lead days from 3 to 15 with an increment of 3 days for boreal winter.

Figure 25. Temporal correlation evolution within observed and j-model forecasted precipitation anomalies are plotted for lead days from 3 to 15 with an increment of 3 days for boreal summer.

Figure 26. Temporal correlation evolution within observed and j-model forecasted precipitation anomalies are plotted for lead days from 3 to 15 with an increment of 3 days for boreal winter.

Figure 27. Anomaly Correlation Coefficient (ACC) is plotted against number of lead days for j-model (green color) and t-model (purple color) for Boreal summer (Left) and Boreal winter (Right) calculated using MISO indices.

Figure 28. RMSE is plotted against number of lead days for j-model (green color) and t-model (purple color) for Boreal summer (Left) and Boreal winter (Right) calculated using MISO indices.

Figure 29. Phase error is plotted against number of lead days for j-model (green color) and t-model (purple color) for Boreal summer (Left) and Boreal winter (Right) calculated using MISO indices.

1. ABSTRACT

The prediction of ISO using two multivariate linear regression (MVR) models were analyzed using various indices, which captures the ISO signal, as predictors and convective and circulations fields as predicted depending on the index under consideration. The MVR model using OLR MJO Index(OMI) is showing relatively high prediction skill during boreal winter as well as boreal summer compared to other considered models and could be preferred for operational prediction of ISO during both the summer and winter seasons. Comparison of prediction skill of ISO between j-model and t-model shows that t-model is showing better prediction skill as compared to j-model in most of the cases. Here, j-model is created based on the cross validation approach (Jiang et. al, 2008) which uses same dataset for training and validation whereas t-model keeps separate some data set to train the model and separate data set to validate the model. Prediction Skill of MVR model created using OMI and RMM index is relatively higher compared to that created using MISO index. The skill of ISO prediction remains sufficiently good for a lead time which is greater in boreal winter as compared to boreal summer in case of all season ISO indices (such as OMI and RMM) and this could be because of the multiple propagation directions in case of boreal summer as compared to its winter counterpart which complicates the prediction of the same. A time-related correlation between the predicted and observed u200 anomaly data is found to be relatively high and coherent as compared to OLR and u850 field data in case of MVR model created using RMM.

2. INTRODUCTION

Monsoon generally referred to the tropical and subtropical seasonal reversal of atmospheric circulation and the rainfall linked to it. Monsoonal rainfall undergoes an extended period of active (above seasonal mean rainfall) and break (below seasonal mean rainfall) conditions. Earlier studies (Dakshinamurthy and Keshavamurthy, 1976; Alexander et al., 1978) have associated it with the motion of Inter tropical convergence zone (ITCZ) over the core monsoon region (Krishnamurti et al 1976; Sikka and Gadgil 1980). The active and break episodes in monsoon is often considered as the manifestation of monsoon intraseasonal oscillations (MISOs). Two prominent modes of the MISOs are the westward and northward propagating modes with the periodicity of 10-20 and 30-60 days respectively (Goswami et al. 2018). The '30-60 days northward propagating mode' is prominent during northern hemispheric summer. The prediction of monsoon rainfall few weeks in advance is crucial for agricultural practice, economy, life and properties across the Indian subcontinent. The existence of large-scale organized, quasi-periodic propagating Intraseasonal modes provides an untapped source for extending the weather prediction beyond one week. Even though significant progress has been made in the simulation of MISOs, there are still significant biases in the extended range (10-30 days) prediction of rainfall using numerical models.

2.1 Depending on the time scale of occurrence of atmospheric processes, forecasting is divided into the following major categories:

2.1.1 Weather Forecasting: Weather forecasting is the branch of science which deals with the modern scientific technologies in predicting the future state of the weather. Following are the different types of weather forecasting: a. now casting (2-6 hours), b. Short- Range (2-3 days), c. Medium-Range (7-10 days) d. Extended-Range (10-30 days)

2.1.2 Long-Range Forecasting: Time range for this type of forecasting is from 30 days or 1 month to two years.

2.1.3 Climate Forecasting: Time range for this type of forecasting is beyond two years.

2.2 Different methods which are used for forecasting are listed below:

2.2.1 Statistical methods:

In these methods we attempt to find a relationship between predictand or dependent variable and predictors or independent variable using the available observed historical data with the help of statistical techniques. E.g. Methods based on regression, analog, Singular value decomposition (SVD) etc.

Regression is one of the fundamental methods of modeling and analyzing of data in which we try to minimize the error between the lines or the curve that we want to fit to the observed data values and the actual observations. Regression model, Logistic Regression, Polynomial Regression etc. are few examples of regression.

Analog techniques for prediction are based on the assumption that if state of two atmospheric systems has same initial condition, and then they are going to evolve in the similar fashion. This technique for prediction has been used extensively in different scales of forecasting. e.g. "General Circulation forecasting" (Gutzler and Shukla 1984; Van den Dool 1989)

2.2.2 Dynamical methods:

In dynamical methods, we use the solution of the mathematical equations which govern the Physics of Ocean and Atmosphere to get an estimate of the future state. Considering the highly non-linear behavior of the governing equations, we make certain assumptions and provide initial conditions in accordance with the system under study, e.g. CFSv2, ECMWF, NCEP's MRF model etc.

2.2.3 Hybrid methods:

These methods of forecasting use a combination of both the dynamic and statistical methods in the prediction of future state of the atmosphere.

One of the major reasons for errors in predicting the precipitation is attributed to the representation of clouds and cloud microphysics in the dynamical models. Also the dynamical models are unstable to perturbations and are sensitive to the initial conditions. These errors increase with time and give a poor and unreliable weather forecast beyond a period of 10 days. Hence we are interested in using the statistical method for prediction that is more skillful and compatible for extended range prediction.

The first statistical ISOs model was developed by Von Storch and Xu (1990) and it was based on the 'Principal Oscillation Patterns' (POPs) (Hasselmann, 1988) of equatorial 200 hPa velocity potential anomalies. They got a skillful forecast up to about 15 days. Later, a statistical model based on the 'singular value decomposition' (SVD) method was proposed (Waliser et al (1999)). By using past and present pentads (5-day average) of filtered outgoing Longwave Radiation (OLR), they were able to predict the future pentads of OLR. Because this model needs filtered data, we cannot use the same for real time application. Ever since, several efforts have been made to get a skillful forecast without using explicit time filtering. Lo and Hendon (2000), developed a lag regression model based on the observation that the MJO can be well represented by leading two EOFs (Empirical orthogonal functions) of OLR and leading three EOFs of stream function at 200mb. Separate methods were used to extract the signal of intraseasonal time scale by removing the annual cycle, interannual and high frequency component. They found a skillful forecast up to about 15 days after the application of model on the independent data and also better skill in the case of active MJO condition contrary to quiescent condition. Even after that numerous studies using statistical methods for the extended range prediction was also done (See 'Mo, 2001; Wheeler and Weichmann, 2001; Jones. et al, 2004; Maharaj and Wheeler, 2005; Jiang. et al, 2008.')

Recently, different approach was introduced to forecast the intraseasonal summer rainfall over China. The method involves the division of the entire domain into ten sub-regions using the rotated empirical orthogonal function analysis of intraseasonal (10-80 days) rainfall anomalies (Zhu and Li, 2016). The ISO signal extraction was done by "non-conventional filtering" following Hsu et al., 2015 methodology. The methodology allows the model to be used in real time. The forecasts were made using the spatial-temporal projection models (STPMs). Within

the group of ten predictors (OLR, zonal wind at 850 hPa (U850), zonal wind at 200 hPa (U200), geopotential height at 850 hPa (H850), geopotential height at 500 hPa (H500), geopotential at 200 hPa (H200), relative vorticity at 850 hPa (Curl850), relative humidity at 700 hPa (Rhum700), soil moisture (Soilm), surface precipitation (Precip), an optimal ensemble of the best five predictor was selected and then used for each region for the prediction. Results showed that the overall pattern of Intraseasonal rainfall over china can be reproduced with a good skill at 20 day lead using this model. Since the temporal resolution of this model (based on pentad data), is quite low, it will not be able to capture the Intraseasonal signal at 10-30 day period over the entire region. But this issue might be minimized by increasing the temporal resolution of the model. One more alternative is to conduct STPM particularly for 10-30 day mode. Similar study was done using STPM (*Zhu and Li, 2017*).

Since ISOs exhibit seasonality, the statistical models must capture the seasonality the ISOs. Roundy 2017 introduced an algorithm to access the regression slope coefficients which vary seasonally. For each day best fit regression slope coefficient value can be obtained by dividing the outcoming seasonal covariance with its variance. The proposed method was applied to estimate the seasonally varying relationship between the MJO and various circulation patterns.

Linear regression, a machine learning algorithm, does not need a complete understanding of physical processes of the atmospheric systems. We propose to develop a forecast model based on multivariate linear regression approach using various ISO indices. Predictors for the model will be the various indices representing different aspects of tropical Intraseasonal oscillations.

3. DATA AND METHODS

3.1 Data: Various observed as well as reanalysis products are used to develop a multivariate lag regression forecast model. The variables used are wind at 200 and 850 hPa pressure levels, precipitation and Outgoing Longwave Radiation (OLR).

3.1.1 OLR: The daily OLR data of 2.5°X 2.5° horizontal resolution for the time period of 1974-2015, are obtained from National Ocean and Atmosphere Administration (NOAA).

3.1.2 Wind data: We have used the four-times daily, 1.5°X 1.5° horizontal resolution; ERA-Interim reanalysis wind data at 850 and 200 hPa pressure levels for the period 1979-2015 (Dee DP, Uppala SM, Simmons AJ, et al. 2011).

3.1.3 GPCP Precipitation: The daily global precipitation climatology project (GPCP) of horizontal resolution 1°X 1° for the time period of 1997-2015 is obtained from the website: <https://precip.gsfc.nasa.gov/>

3.2 Methods:

A multivariate linear lag regression (MVR) forecast model is developed to predict various Intraseasonal dynamic and thermodynamic fields by following Jiang et. al, 2008. The general expression of the multivariate linear lag regression model is given below.

$$X(t_0 + \tau) = \beta_1 PC_1(t_0) + \beta_2 PC_2(t_0)$$

X is the predictand field (which is the field to be predicted), τ =Number of lead days, t_0 = time for which initial condition is given. PC_1 and PC_2 correspond to the bivariate indices which are used for creating the seasonally varying lag-regression coefficients β_1 and β_2 .

3.3 Procedure for creating Indices

Using the MVR model, we forecast the various modes of tropical intra-seasonal variability. The modes include the eastward and northward propagating intra-seasonal disturbances. The procedures for creating various indices that depict the circulation and convective features related to the ISO are (RMM, MISO, and OMI indices) are described below.

3.3.1 Real time Multivariate MJO (RMM) index:

RMM is a bivariate index, and it represents how the MJO is evolved (Wheeler, M.C. and H.H. Hendon, 2004). The procedure for creating the RMM index is briefly explained below.

The anomalies of OLR and zonal wind at 200 hPa and 850 hPa pressure levels from 15°S to 15°N latitudes and all longitudes for the period 1979 - 2001 (8401 days) are used for creating the RMM index. The anomalies are created by pulling out the climatological annual cycle (1st three harmonics of daily climatology) from the original data. The effects of major interannual variability are removed by subtracting the previous 120 day mean from the daily anomaly fields. Then, latitudinal average over 15°S – 15°N domain is performed for all the fields. All the fields are then normalized by their respective global variance.

Combined EOF (CEOF) analysis is done on the pre-processed zonal wind at pressure level of 200 hPa and 850 hPa (here after u200 and u850) and OLR anomalies. The spatial structure of the EOF's is a function of longitude alone. The first two modes of the CEOFs capture the planetary scale circulation and convective anomalies of the MJO which propagates eastward. One of the CEOF modes depicts the enhanced convection over the Maritime Continent, low-level westerly wind anomalies over the Indian Ocean and Maritime Continent region and baroclinic structure of wind anomalies over the Pacific region with easterlies at low level and westerlies at upper level.. The other mode captures the enhanced convection over the Pacific Ocean. The first two CEOF modes together explain 25% of the total variability. The projection of the OLR and zonal wind anomalies at 850 hPa and 200 hPa onto the two leading CEOF modes gives two time series (known as principle components (PCs)). These two PCs are named as RMM indices. These indices have been used for estimating the location and Amplitude of the MJO in real-time.

3.3.2 Monsoon Intraseasonal Oscillation (MISO) index:

Like the RMM index, the MISO index provides the phase and amplitude of the northward propagating Intraseasonal convective disturbances in real-time. A detailed description of the creation and use of the MISO index is given in Suhas et.al, 2013. The various steps involved in the calculation of MISO indices are briefly explained below:

First, daily precipitation anomalies are computed by withdrawing the climatological annual cycle from the original observed data. Since we want to extract the northward propagating Intraseasonal convective disturbances, the longitudinal average is done over the domain of (60.5°E - 95.5°E). Then the extended EOF analysis is carried out on the resultant data over the latitudinal domain 12.5°S-30.5°N, by keeping 15 days lagged copies of the data.

First two major modes of the EEOF analysis which explain the 23% of the total variability are chosen. The EEOFs prefer meridional wavenumber one structure which implies that opposite signal of convection will be present above Indian continental region and the Indian equatorial ocean region. The PC₁ and PC₂ which we get by projecting precipitation data to the two EEOF spatial structures chosen are named as MISO1 and MISO2 index. Since MISO1 and MISO2 index are related to the precipitation over the ISM domain we tried to find the spatial region over the ISM domain which is showing reasonable relation with MISO1 and MISO 2 evolution over time. The region chosen in this manner are,

BOX1. 15°N- 20°N and 60°E-100°E and **BOX2.** 0°N-14°N and 60°E -100°E

Where the area averaged the time series give reasonable relation with the MISO1 and MISO2. The correlation between MISO 1 and BOX2 is 0.48 and that between MISO2 and BOX1 is 0.54. For the rest of the analysis, instead of MISO1 and MISO2 index the BOX area averaged time series are used.

3.3.3 OLR Based MJO (OMI) Index

The OLR-based MJO index (OMI) describes the evolution of convective feature of the MJO. The MJO shows notable seasonality in their main properties. The eastward propagating feature of MJO is strong and frequent during boreal winter

as compared to that during boreal summer. The northward propagating convective signal over the ISM is another dominant feature during boreal summer. Hence, to incorporate seasonality the EOF analysis of 30-96 day filtered OLR with a time window of 120 days is calculated which gives the EOF as a function of day of the year as well as space. The EOF analysis of the mentioned field with time window less than 120 days was found to make the EOF structure degenerate while making it above 120 days was found to reduce the seasonality associated with MJO. So, an optimum time window as mentioned (120 days) was chosen while the EOF calculation was done. The first 2 dominant modes of EOF viz. EOF1 and EOF2 together represent the spatial structure of MJO. Then the OLR data for a particular calendar day was projected to the corresponding EOF1 and EOF2 spatial pattern to get the PC's which are called as OMI1 and OMI2 respectively.

3.4 Model skill analysis metrics

The performance of the MVR forecast model is quantified using the Root mean Square Error (**RMSE**), Bivariate Correlation (**COR**), and the Phase angle error (**ERR**) metrics. The definition and formula of these three metric are as follows:

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

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

$$ERR(\tau) = \frac{1}{N} \sum \tan^{-1} \left[\frac{a_1(t)b_2(t,\tau) - a_2(t)b_1(t,\tau)}{a_1(t)b_1(t,\tau) - a_2(t)b_2(t,\tau)} \right] \dots \dots \dots (3)$$

$a_1(t)$ and $a_2(t)$ are the observed RMM1 and RMM2 at time t and $b_1(t, \tau)$ and $b_2(t, \tau)$ are the respective forecasts for time t with a lead time of τ days or lag time of τ days. Also, N is the number of predictions.

3.4.1 Bivariate Correlation: Bivariate correlation (Eq.1) provides the information about the strength of the relationship between the two variables. Bivariate correlation is used for quantifying the strength of the relationship

between the predicted and observed indices. Higher the value of correlation is better will be the relationship and hence better will be the prediction.

3.4.2 e-folding time: The time (in days) corresponds to the correlation value of $\frac{1}{e}$ is called e-folding time which is taken as reference in all the ACC plots.

3.4.3 RMSE: Root mean Square Error represented by Eq. 2 calculates the error between the predicted value and observed value of the variable used in the study.

3.4.4 Phase angle Error: Phase angle Error is the measure of error associated with the phase angle between the predicted and observed value of the variables. Negative value of phase angle implies the slow propagation in prediction in comparison to the observed values and vice versa. It is represented by the formula given in Eq.3

4. RESULTS AND DISCUSSION

The MVR prediction model is expected to capture the basic spatial structure and northward and eastward propagating properties of the ISOs. The spatial pattern of the relation between predicted and observed fields (which show the convective and dynamic features associated with ISO) are analyzed. The model skills using various metrics are also analyzed.

4.1 Multivariate Regression model (MVR) forecast using RMM index

The spatial structure of temporal correlation within the model predicted and observed fields of OLR and Zonal Wind at pressure level of 850 hPa and 200 hPa for different time lags are shown in Fig1.

4.1.1 OLR.

Correlation using tmodel between predicted and observed jja olr

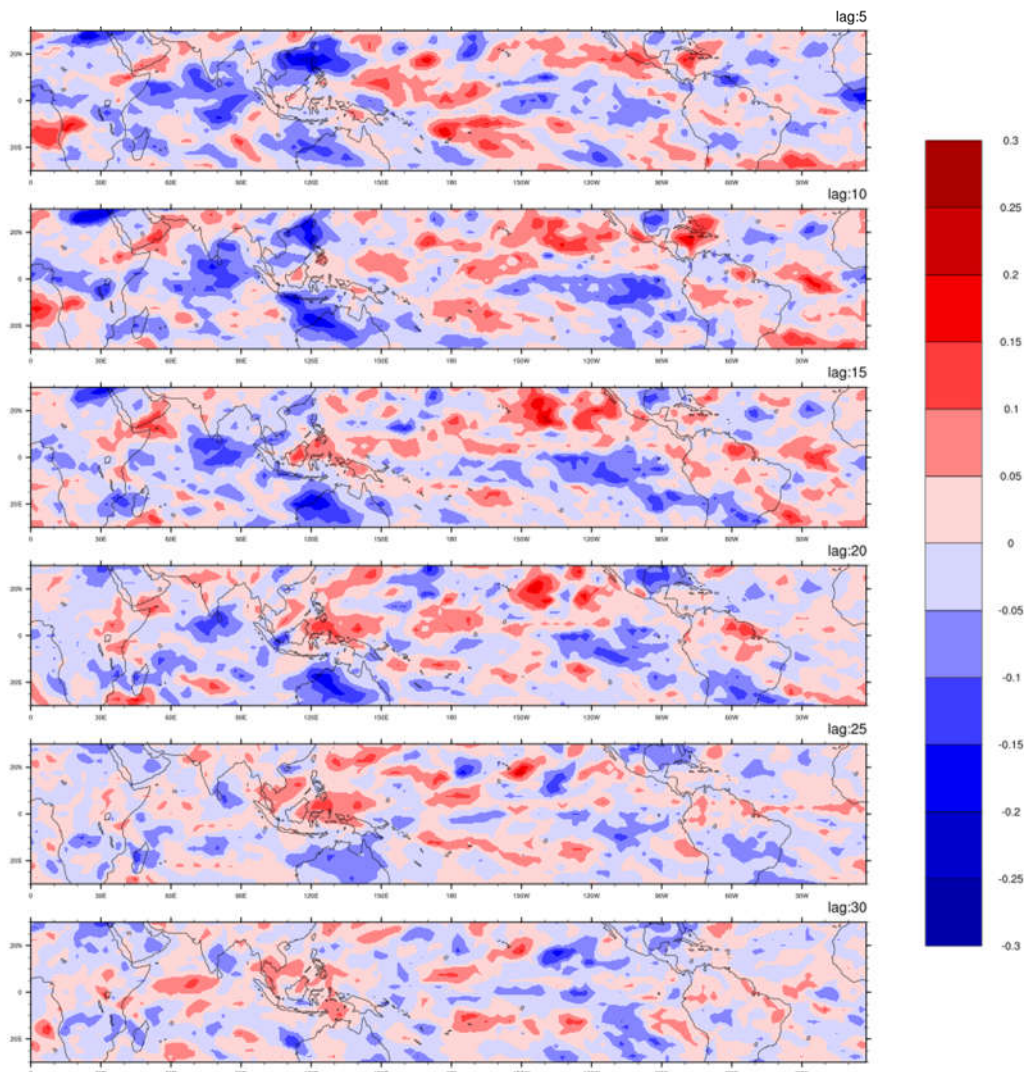


Figure 1. Temporal correlation evolution within observed and t-model forecasted OLR anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal summer.

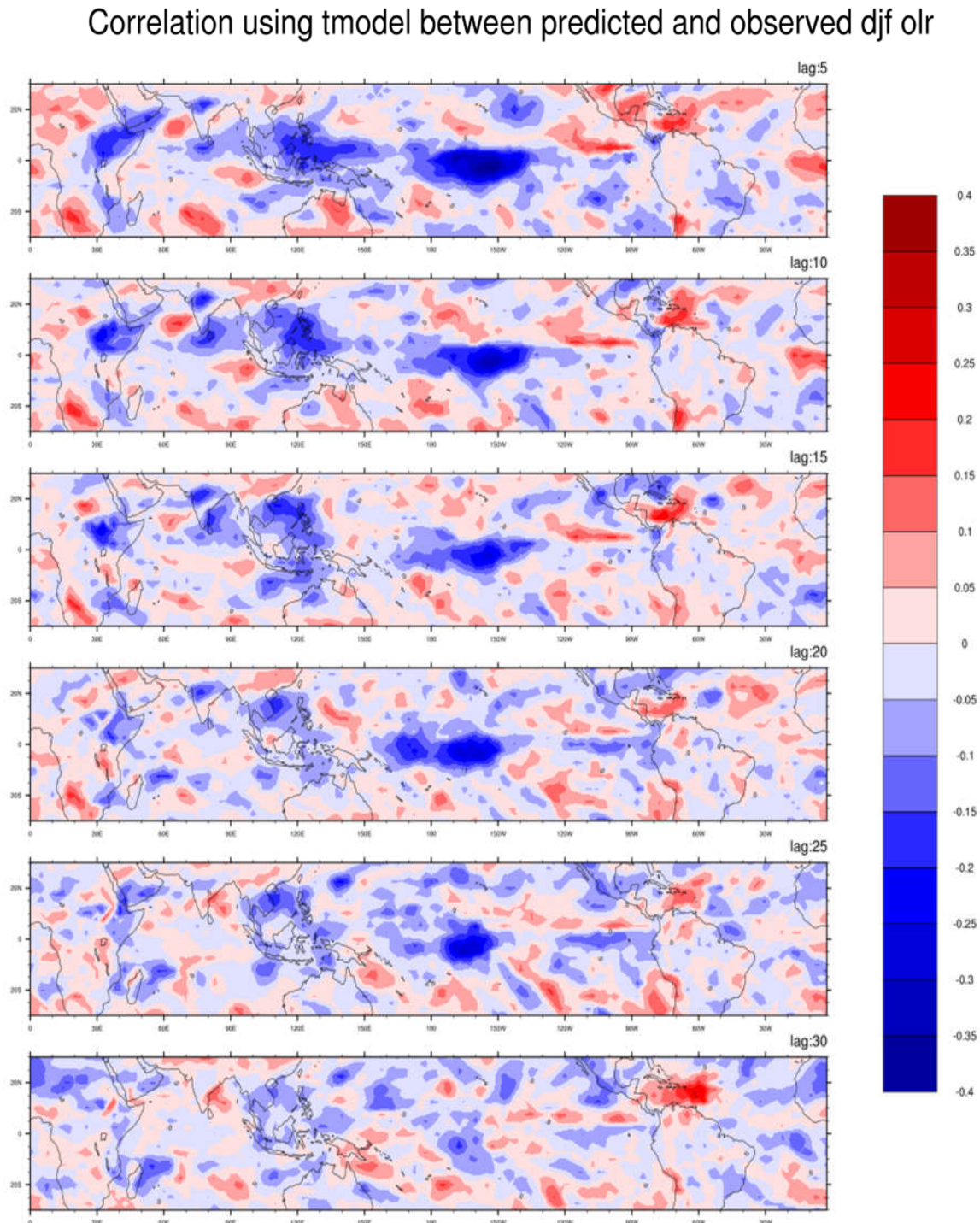


Figure 2. Temporal correlation evolution within observed and t-model forecasted OLR anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal winter.

Correlation using jmodel between predicted and observed jja olr

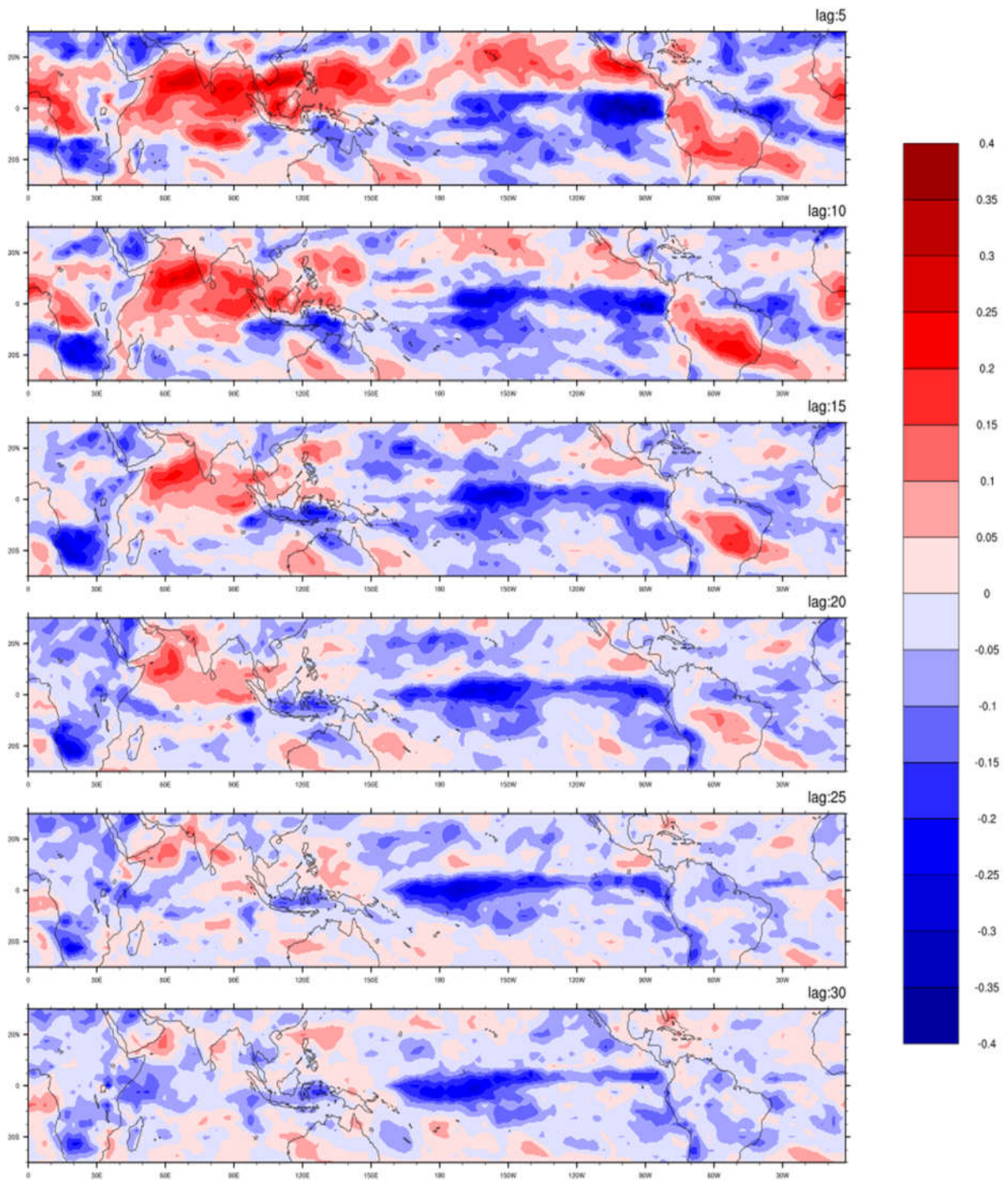


Figure 3. Temporal correlation evolution within observed and j-model forecasted OLR anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal summer.

Correlation using jmodel between predicted and observed djf olr

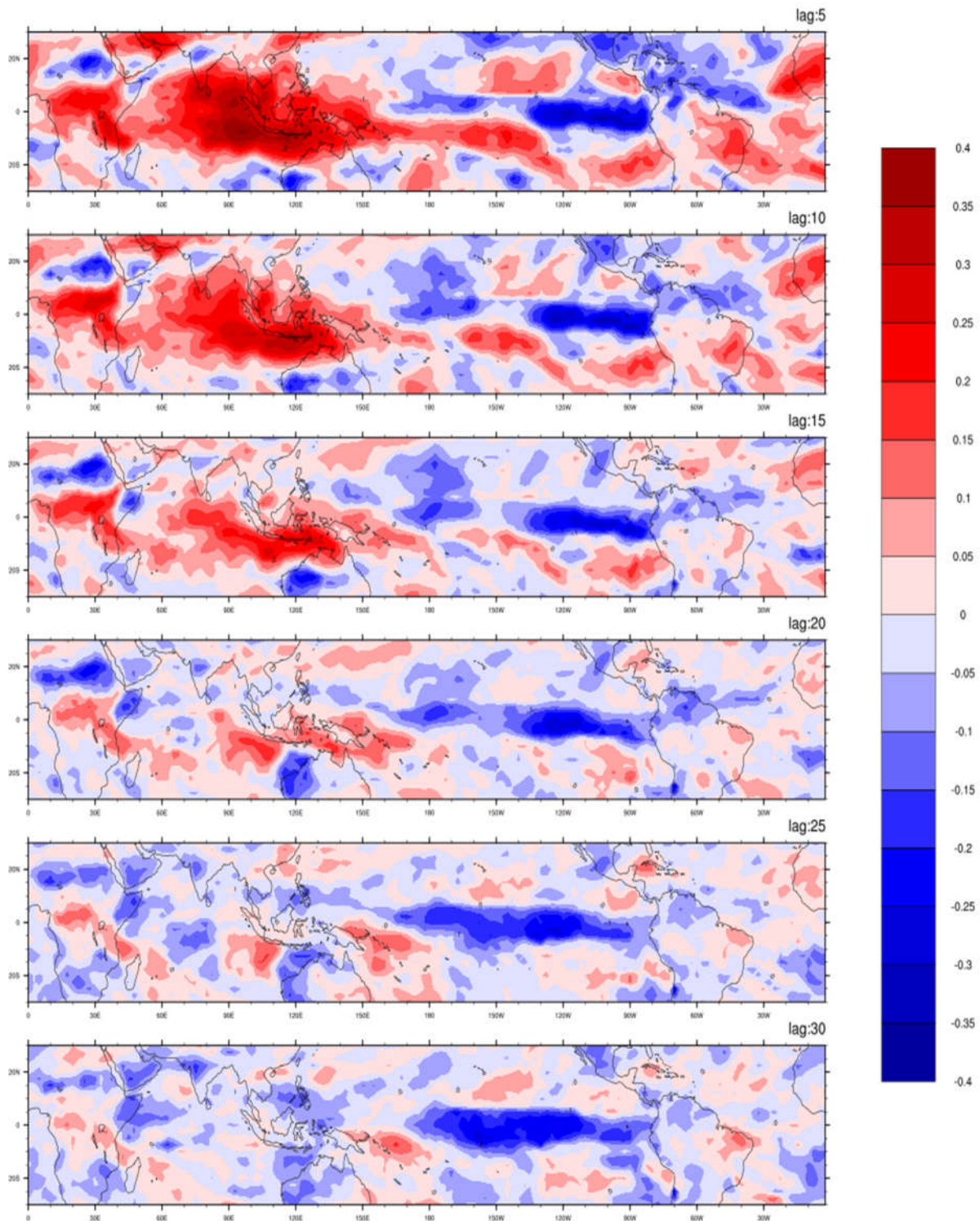


Figure 4. Temporal correlation evolution within observed and j-model forecasted OLR anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal winter.

In Fig.1, a strong negative correlation is found above central equatorial Indian Ocean and the northern Australia which is persistent even after 20 days of lag in both the cases. As the ISO signal during boreal summer has dominant northward propagating component and weak eastward propagating component we expect the spatial correlation to be reasonably strong above Indian summer monsoon region as well as over the equatorial Indian Ocean. Negative correlation is found over the Equatorial Indian Ocean and the strength reduces as the lead time increases. The RMM index captures the eastward propagating component better in comparison to the northward component of MJO. This may be a reason why the correlation plots are not showing significant correlation over the Indian Monsoon domain.

In Fig.2, moderately strong negative correlation is observed above east African region, western equatorial Pacific, central and eastern Equatorial Indian Ocean which are the regions where the convective anomalies of the MJO is clear during an MJO event. The correlation pattern is more coherent during the boreal winter season as compared to boreal summer season which can be seen clearly in the eastern equatorial Indian Ocean and central Pacific and the correlation value remains strong even after 25 days.

Fig.3 shows the spatial correlation between observed and Jiang et al 2008 model (here after j-model) predicted OLR. Strong positive correlation values are observed over the equatorial Indian Ocean, southern part of Indian monsoon region and some part of South America. Even for 15-20 day lag, the correlation remains moderately strong over the central and eastern equatorial Pacific region. It implies the potential for skillful prediction in this region.

Fig.4 shows the j-model predicted and observed OLR anomalies' spatial correlation during boreal winter season. The boreal winter ISO spatial pattern is best represented in this model in comparison to the t model. A strong positive correlation is observed from the equatorial Africa to central equatorial Pacific (where the convective signal in general dies off) which imply the potential of this model as an operational forecast model for predicting the boreal winter ISOs. Over the eastern equatorial and central pacific region, a strong negative correlation is observed which shows persistence for many lag days.

4.1.2 Since the model also captures the dynamic fields associated with the MJO, The spatial plots of temporal correlation for u850 and u200 for different time lags are shown in Fig. 5

4.1.2.1 u850

Correlation using tmodel between predicted and observed jja u850

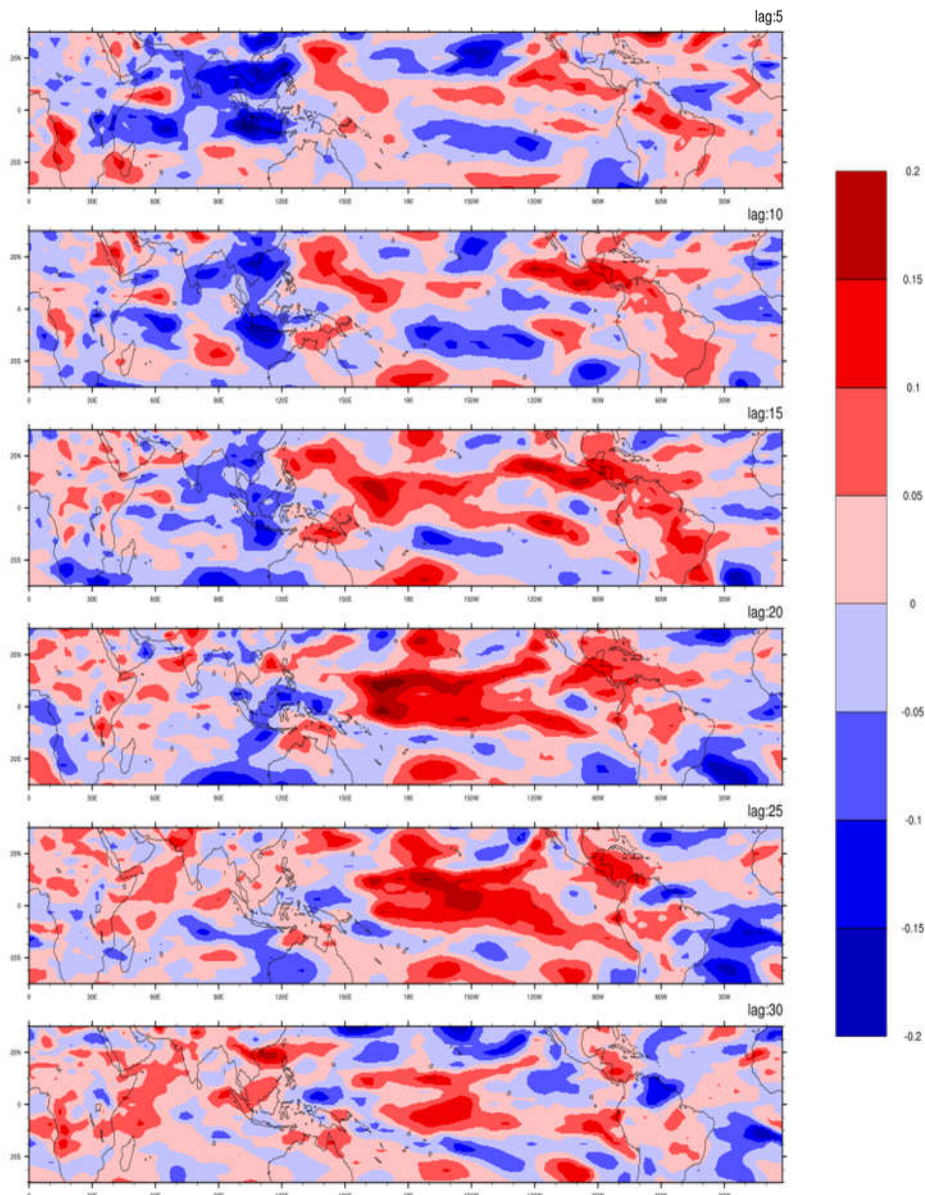


Figure 5. Temporal correlation evolution within observed and t-model forecasted u850 anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal summer.

Fig.5 shows the t-model predicted and observed u850 anomalies' spatial correlation during boreal summer. A very low correlation is observed over the tropics. The correlation persists over Indonesian region and it last for a lead of 20 days. An opposite behaviour is observed over the continental Indian region after 30 days lead.

Correlation using tmodel between predicted and observed djf u850

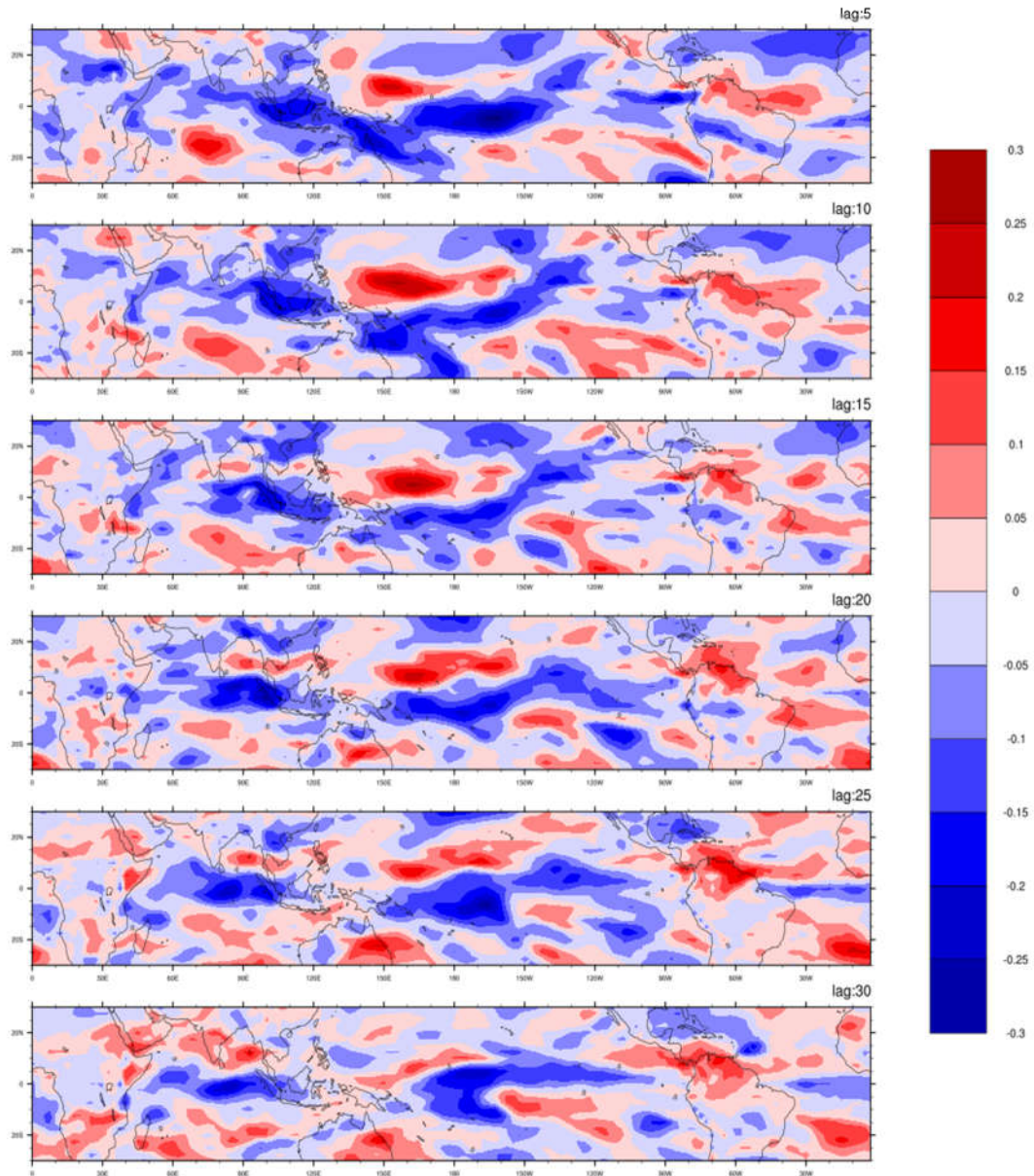


Figure 6. Temporal correlation evolution within observed and t-model forecasted u850 anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal winter.

Fig.6 shows the t-model predicted and observed u850 anomalies' spatial correlation during boreal winter. A moderately strong correlation coefficient is

observed over the eastern equatorial Indian Ocean and it persists for 30 days lead and the reversal of correlation sign is also observed. It implies the ability of the model to capture the propagating component of the MJO.

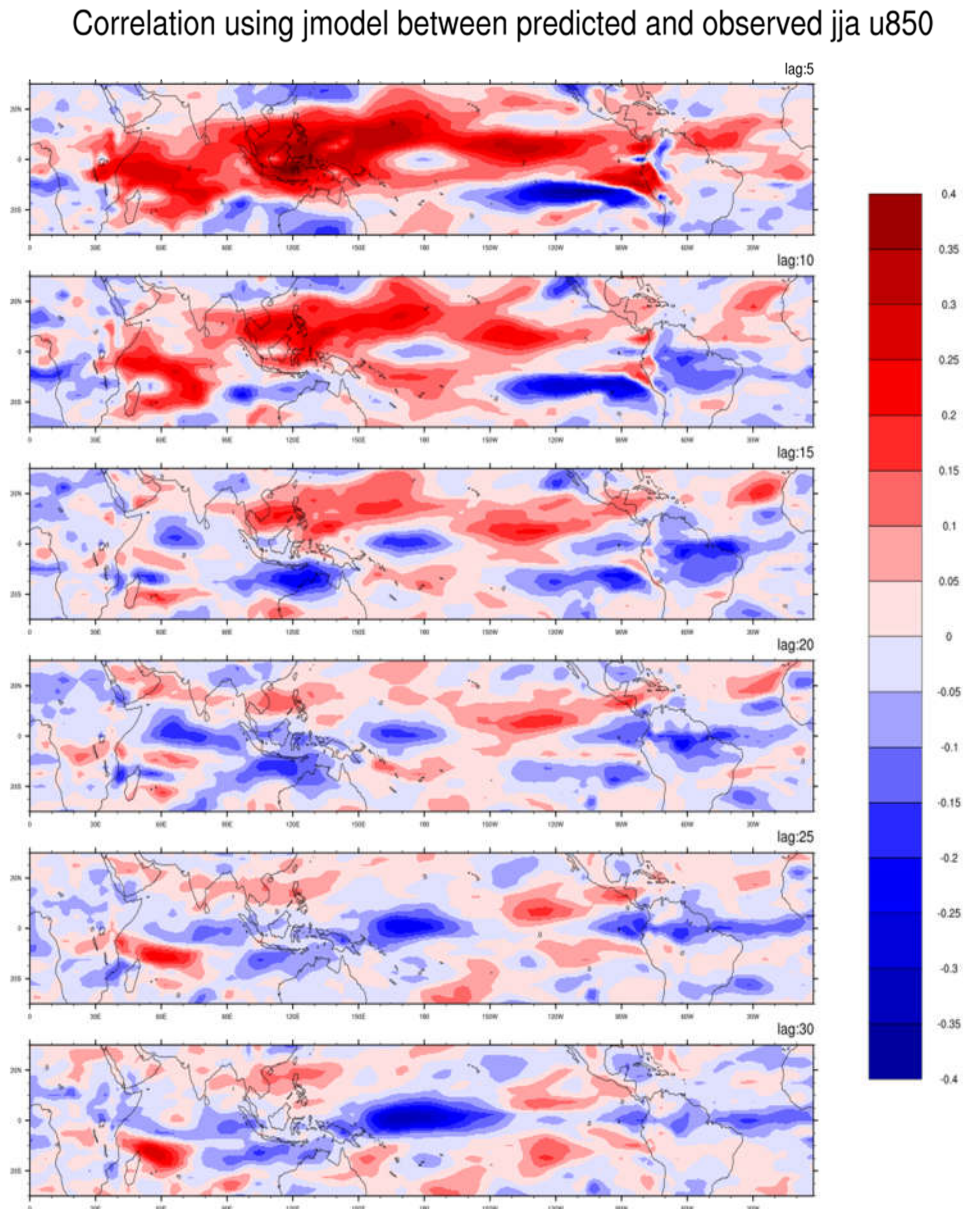


Figure 7. Temporal correlation evolution within observed and j-model forecasted u850 anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal summer.

Fig 7 shows the j-model predicted and observed u850 anomalies' spatial correlation during boreal summer. A strong positive correlation in the eastern and central Pacific region is observed and some patches of strong negative correlation is also observed in the western Pacific region.

Correlation using jmodel between predicted and observed djf u850

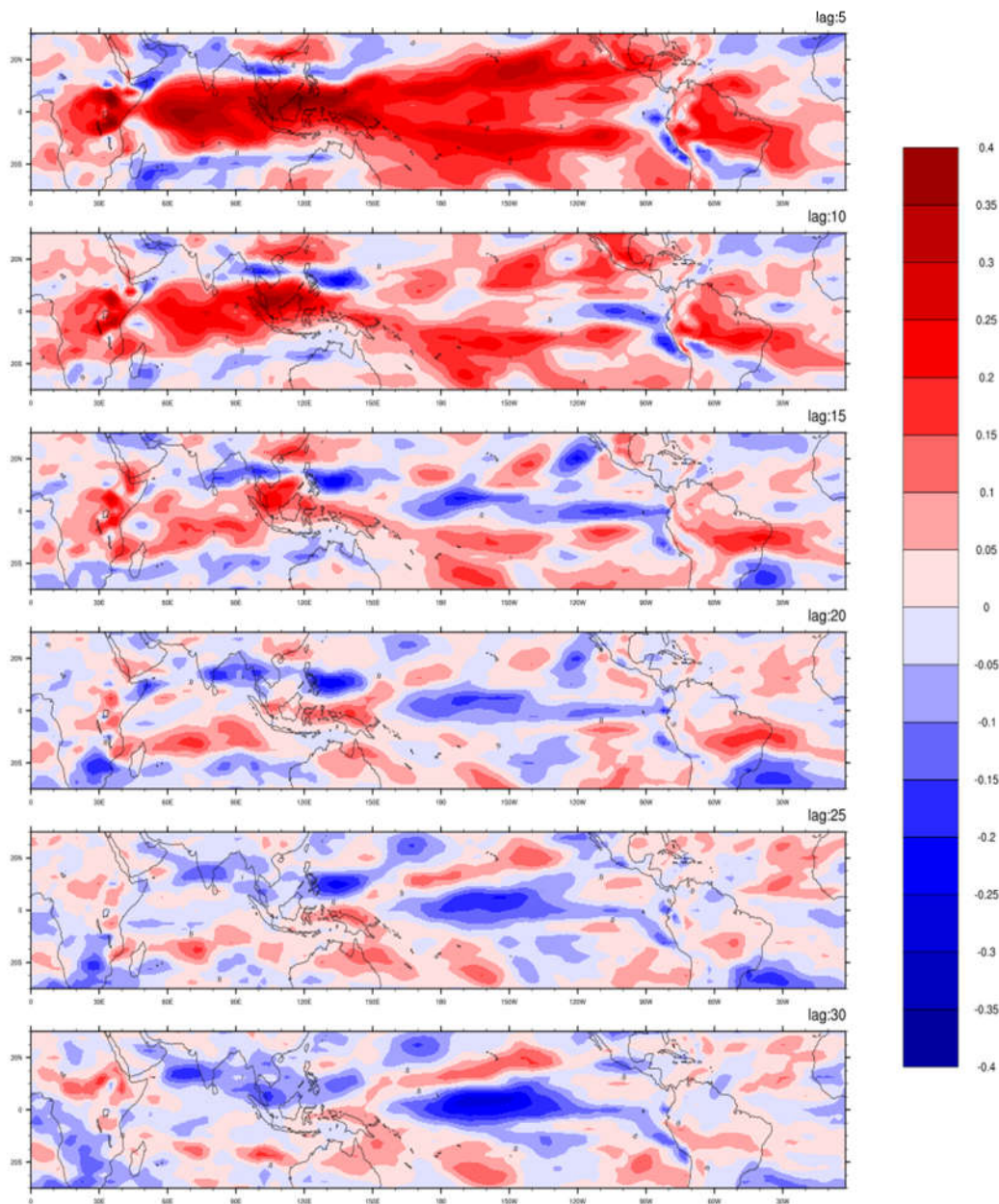


Figure 8. Temporal correlation evolution within observed and j-model forecasted u850 anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal winter.

Fig. 8 shows the j-model predicted and observed u850 anomalies' spatial correlation during boreal winter. Positive correlation values are more coherent as compared to the other cases of the t-model and boreal summer case of j-model. A strong positive correlation coefficient is observed over the South America and it persists for 20 day lead time.

4.1.2.2 u200

Fig. 9 shows the t-model predicted and observed u200 anomalies' spatial correlation during boreal summer. Very low values of correlation similar to that case of u850 for the same season can be observed. A negative correlation of 0.15-0.2 can be seen over the Indonesian region which persists for 15 days. As the correlation values are very less, the model might not give good prediction in that region.

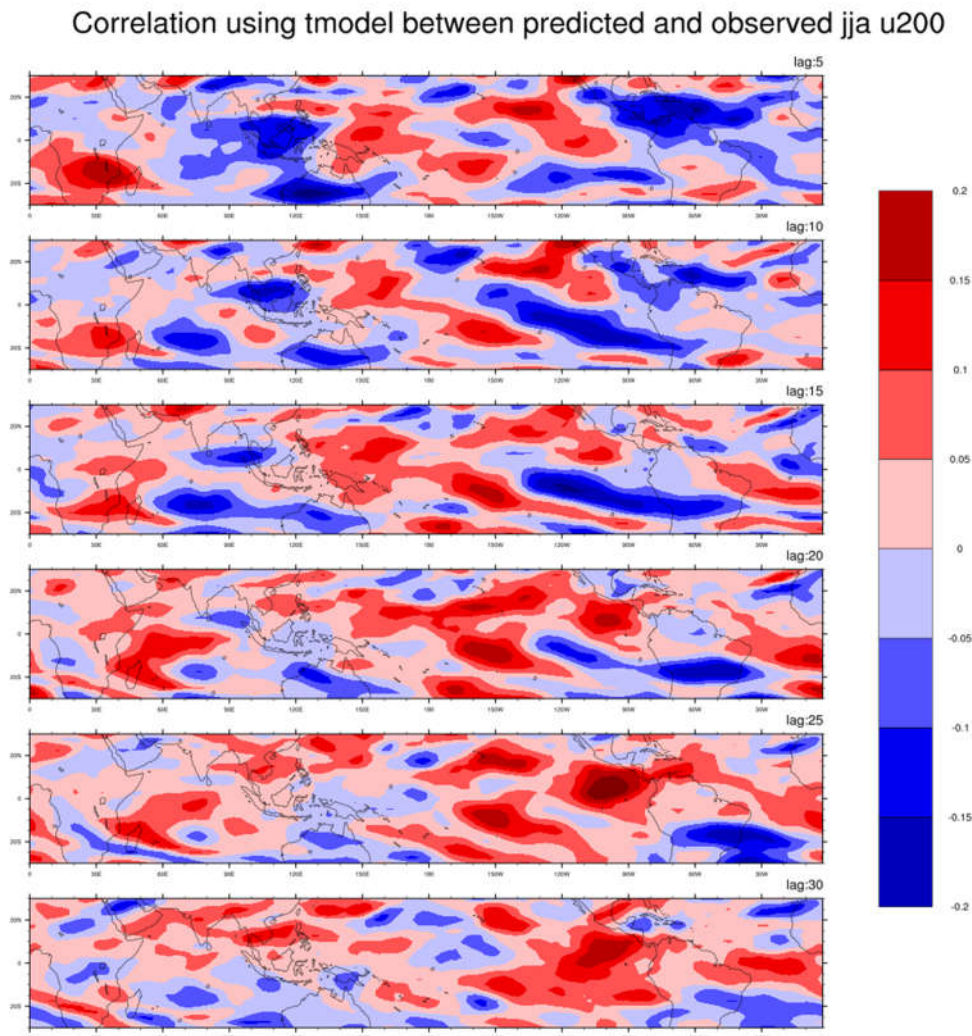


Figure 9. Temporal correlation evolution within observed and t-model forecasted u200 anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal summer.

Correlation using tmodel between predicted and observed djf u200

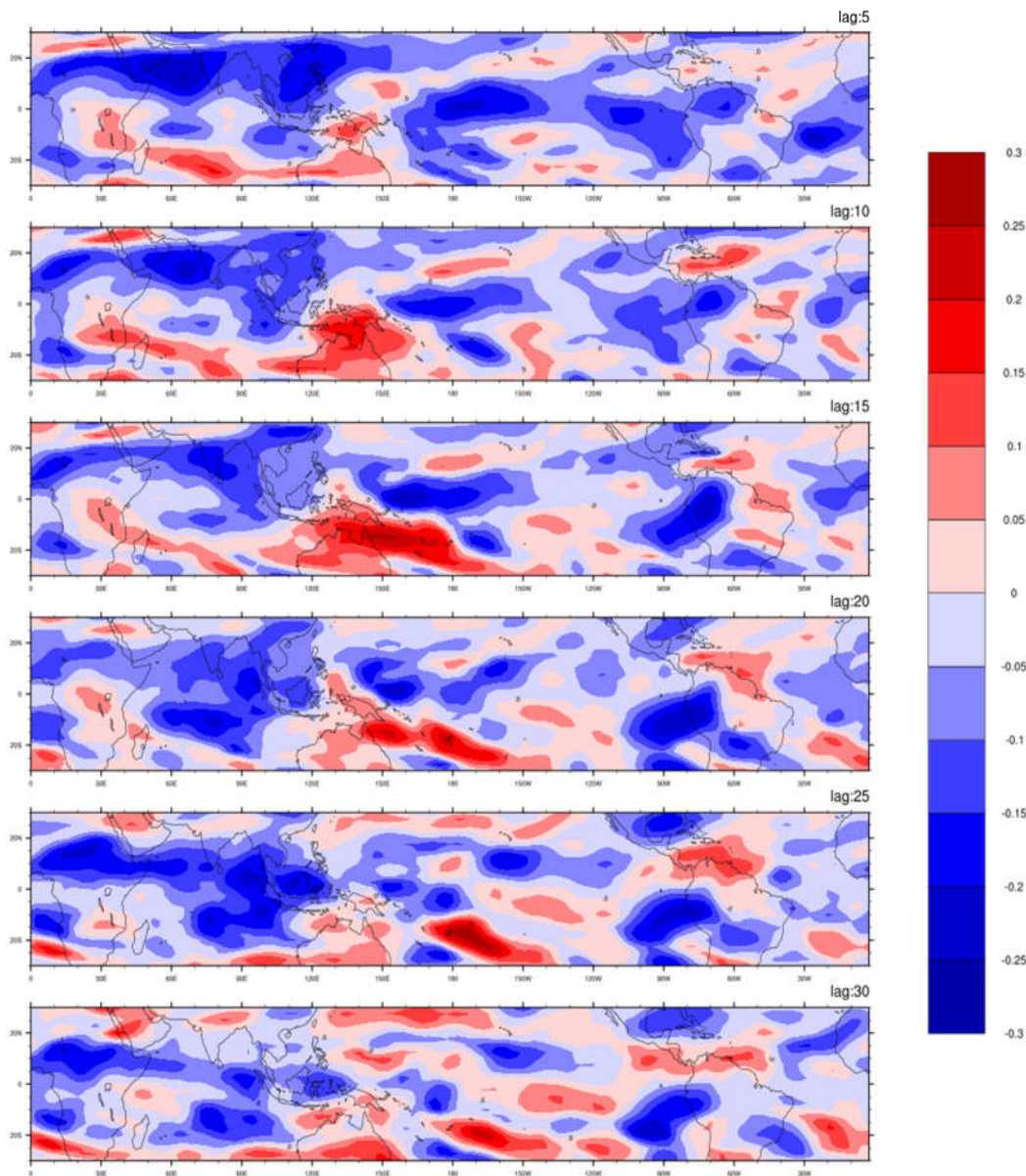


Figure 10. Temporal correlation evolution within observed and t-model forecasted u200 anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal winter.

Fig 10 shows the t-model predicted and observed u200 anomalies' spatial correlation during boreal winter. Negative correlation over the South Asian monsoon domain which is extending with increasing lead days over 20S-20N domain can be seen. Values are more during winter as compared to summer in case of t-model.

Correlation using jmodel between predicted and observed jja u200

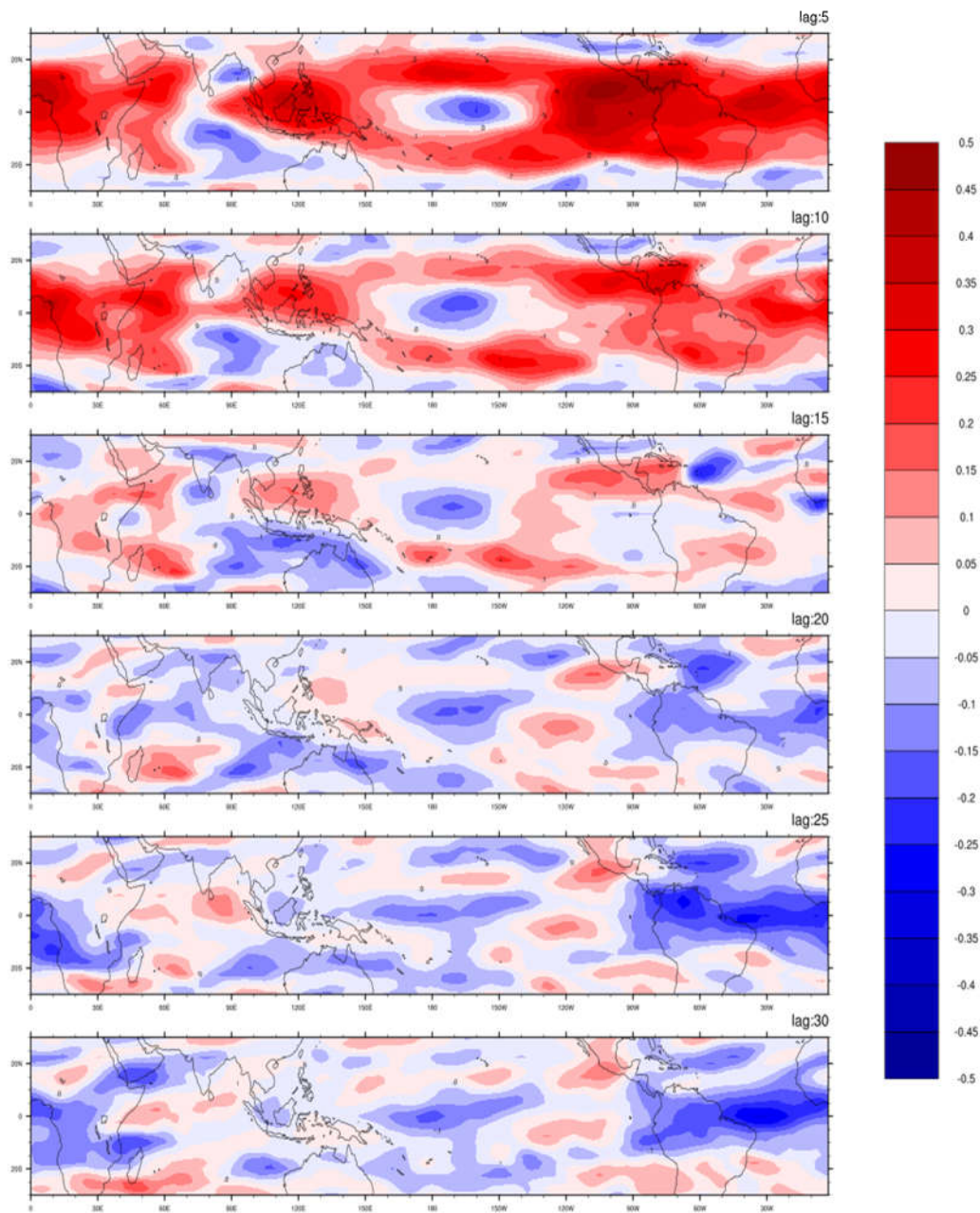


Figure 11. Temporal correlation evolution within observed and j-model forecasted u200 anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal summer.

Fig.11 shows the j-model predicted and observed u200 anomalies' spatial correlation during boreal summer. Correlation values for u200 are the highest compared to correlation in case of other fields mentioned above which is consistent with the past studies. Signal reversal over the north and South American region is much more clearly visible as compared to all other regions by 30 days lead time.

Positive correlation signal over East Africa and Indonesian domain can be seen but it stayed till 10 days lag.

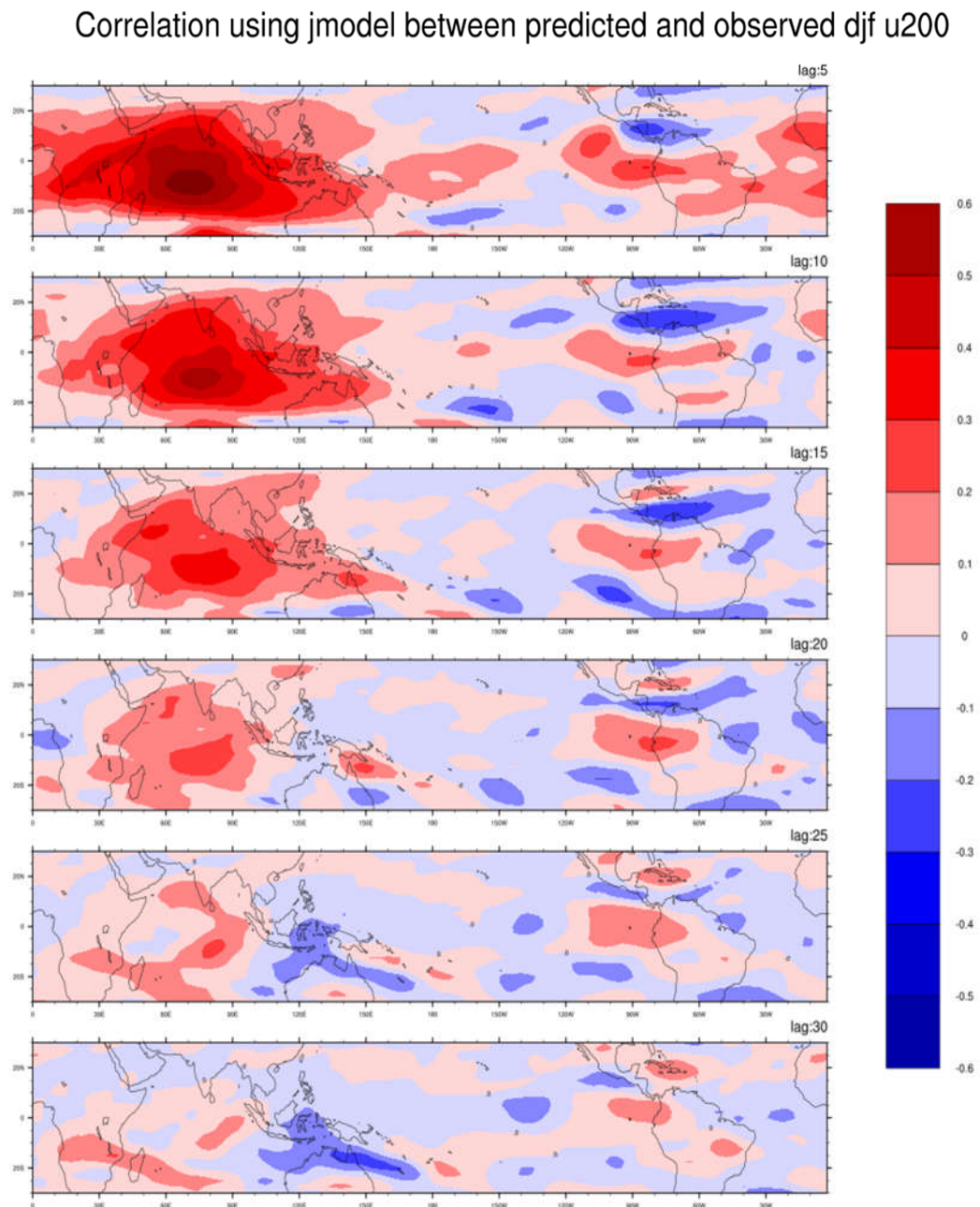


Figure 12. Temporal correlation evolution within observed and j-model forecasted u200 anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal winter.

Fig.12 shows the j-model predicted and observed u200 anomalies' spatial correlation during winter. u200 has the highest values of correlation within all the three used fields and temporal correlation is highest when we are using j-model for

winter u200. The highest positive correlation can be seen over the central equatorial Indian Ocean and it extends westwards till the equatorial African region and eastwards till the central equatorial pacific.

4.2 Skill analysis for RMM

4.2.1 Anomaly Correlation Coefficient

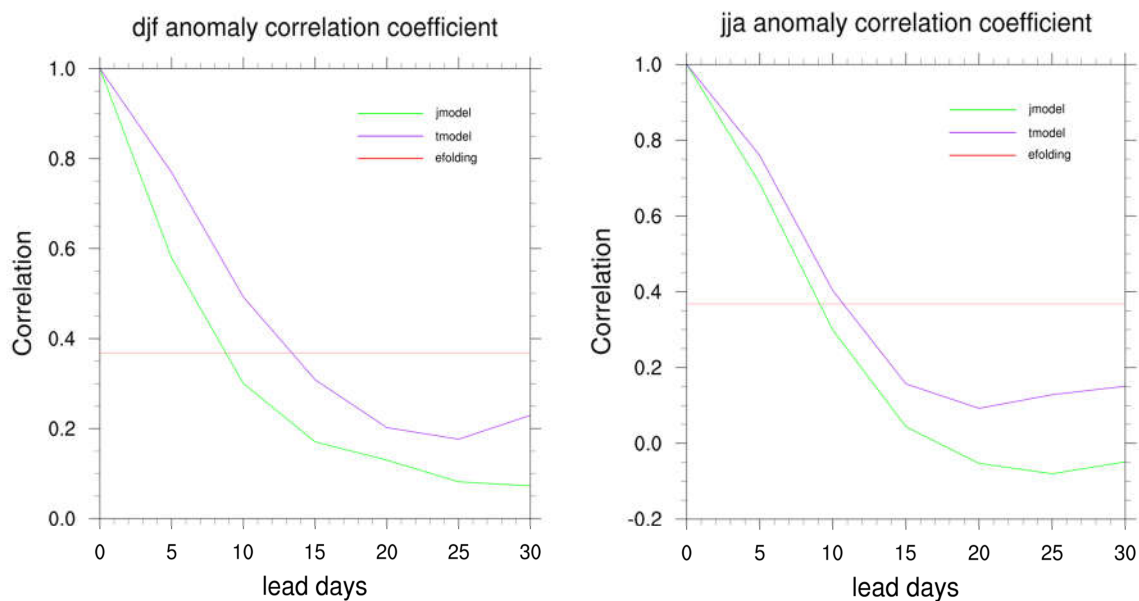


Figure 13. Anomaly correlation Coefficient (ACC) is plotted against number of lead days for j-model (green color) and t-model (purple color) for Boreal winter (Left) and Boreal Summer (Right) calculated using RMM indices.

Fig. 13 shows the lead day vs. correlation plots of seasonal Anomaly correlation coefficients for two models. During summer, as j-model skill (green) and e-folding time (red) are crossing each other at almost 8th day. So, prediction till 8th day can be done successfully with a good prediction skill. Similarly, 10 days is value for t-model which is showing greater prediction skill than j-model. During summer as well as winter the skill is same for j-model but for t-model it has been increased to 13 days (in case of winter).

4.2.2 Root Mean Square Error

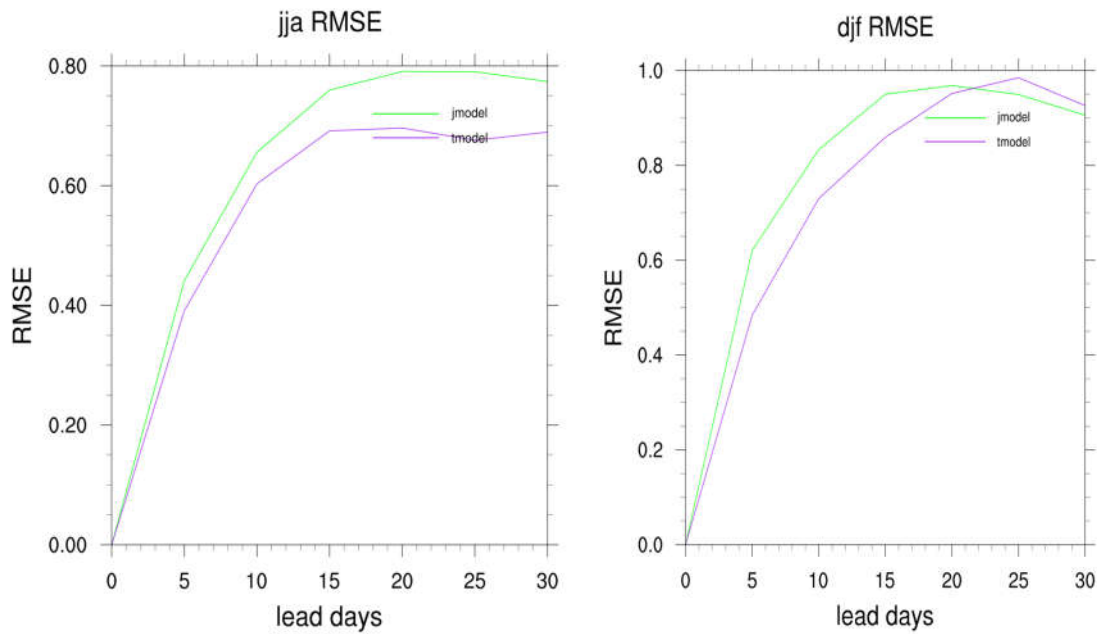


Figure 14. Root Mean Squared Error (RMSE) is plotted against number of lead days for j-model (green color) and t-model (purple color) for Boreal summer (Left) and Boreal winter (Right) calculated using RMM indices.

Fig. 14 shows the comparison of seasonal root mean square error for two models. Error increases with lead time which is we observe in general for most of the prediction systems. Same trend is being followed during both seasons. t-model is giving less value between the predicted and observed values than j-model with lead time for most of the time which is an implication of better prediction skill than t-model.

4.2.3 Phase Angle Error

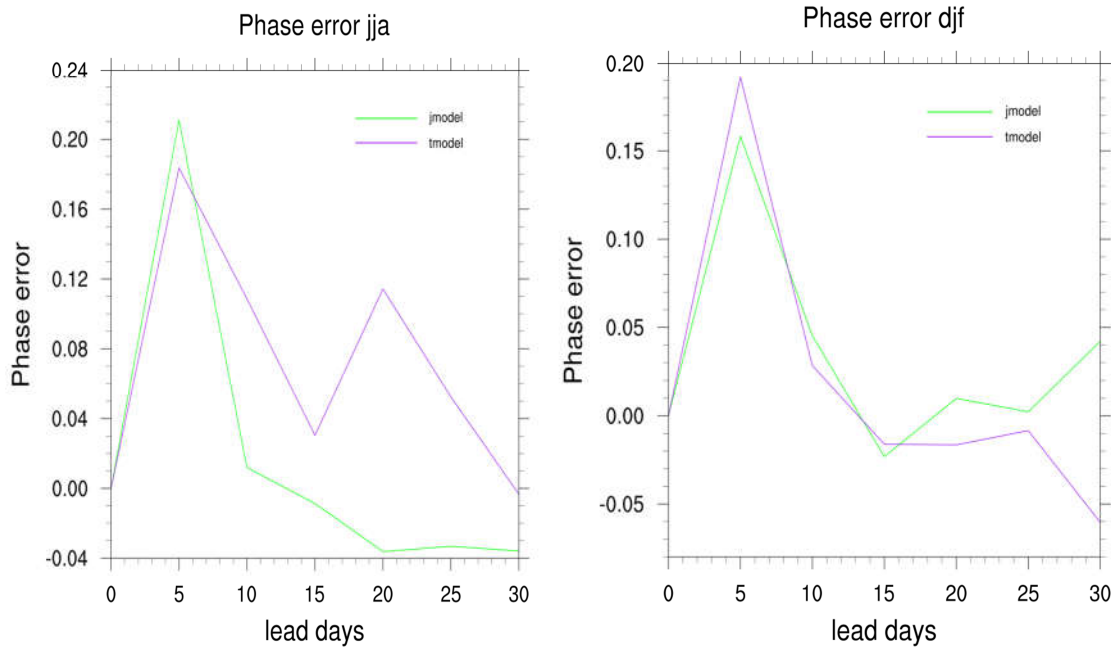


Figure 15. Phase Error is plotted against number of lead days for j-model (green color) and t-model (purple color) for Boreal summer (Left) and Boreal winter (Right) calculated using RMM indices.

Fig. 15 shows the comparison of seasonal phase angle error between two models. Values of phase error are fluctuating with the increasing number of lead days. But, following definition of phase error we can say that phase error in case of summer t-model values are positive which implies that predicted values are propagating faster than that of observed values.

4.3 MVR model using OMI indices

Correlation using omi-tmodel between predicted and observed jja olr

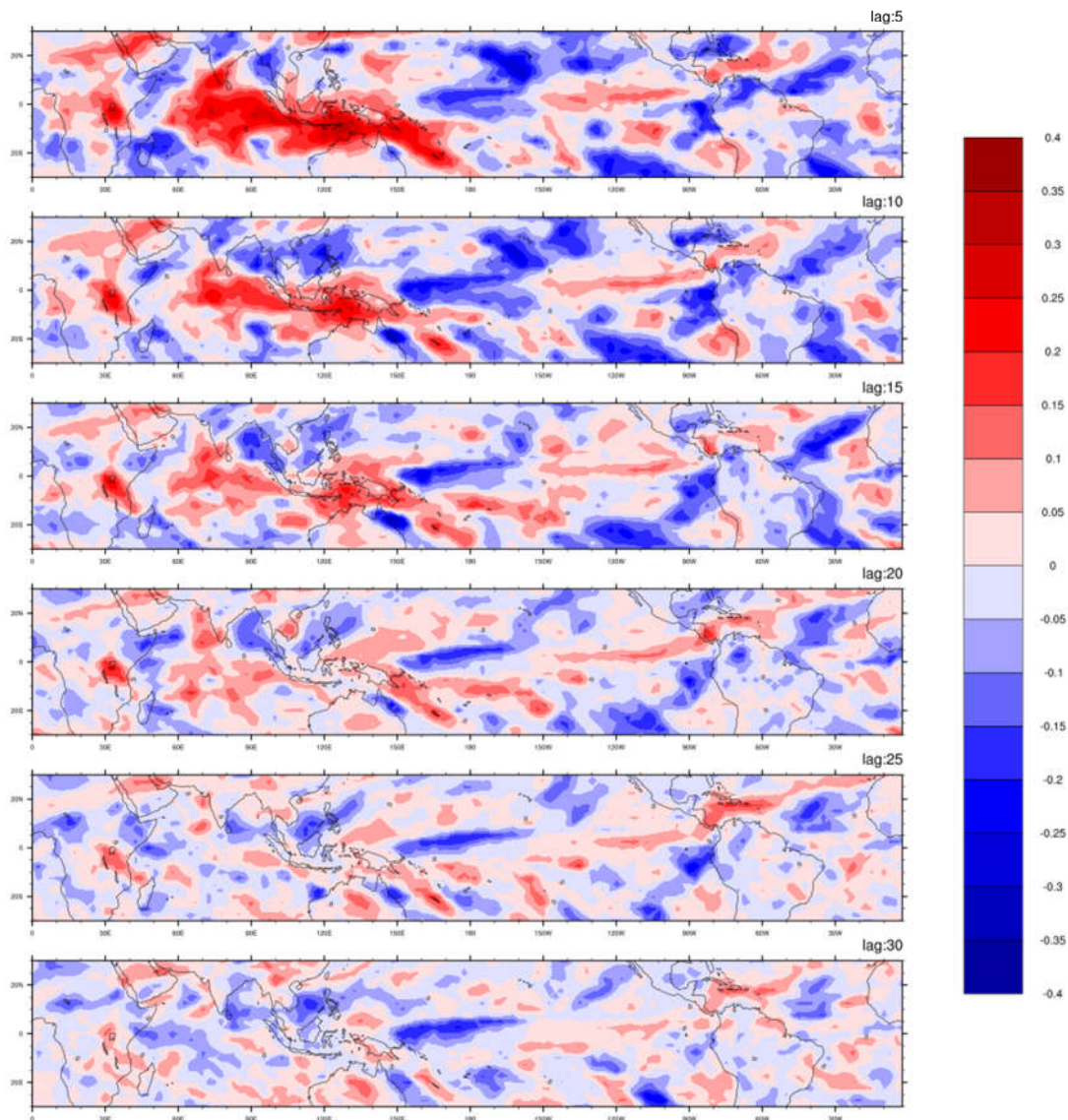


Figure 16. Temporal correlation evolution within observed and t-model forecasted OLR anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal summer.

Fig.16 shows the t-model predicted and observed OLR anomalies' spatial correlation during boreal summer. Total reversal of correlation signal over the continental Indian region and equatorial Indian Ocean after 20 days lag can be observed. Weakening of extent over the western Pacific region is also clear with increasing number of lead days.

Correlation using omi-tmodel between predicted and observed djf olr

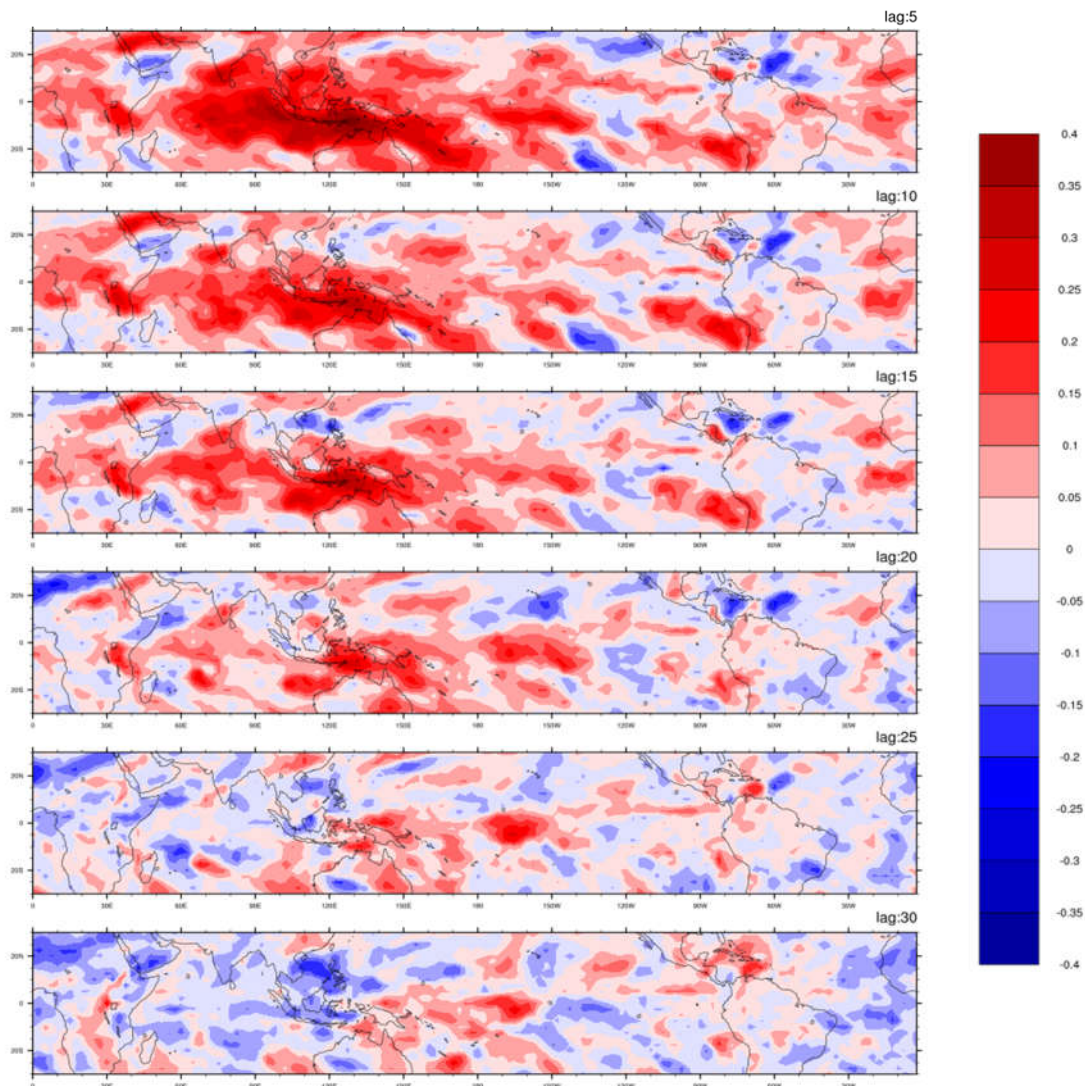


Figure 17. Temporal correlation evolution within observed and t-model forecasted OLR anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal winter.

Fig. 17 shows the t-model predicted and observed OLR anomalies' spatial correlation during winter. High positive correlation values covering almost whole of the equator at 5 day lead which further decreases with increasing number of lead days. Signal reversal is also clearly visible over the Indonesian region after 30 days lead.

Correlation using omi-jmodel between predicted and observed jja olr

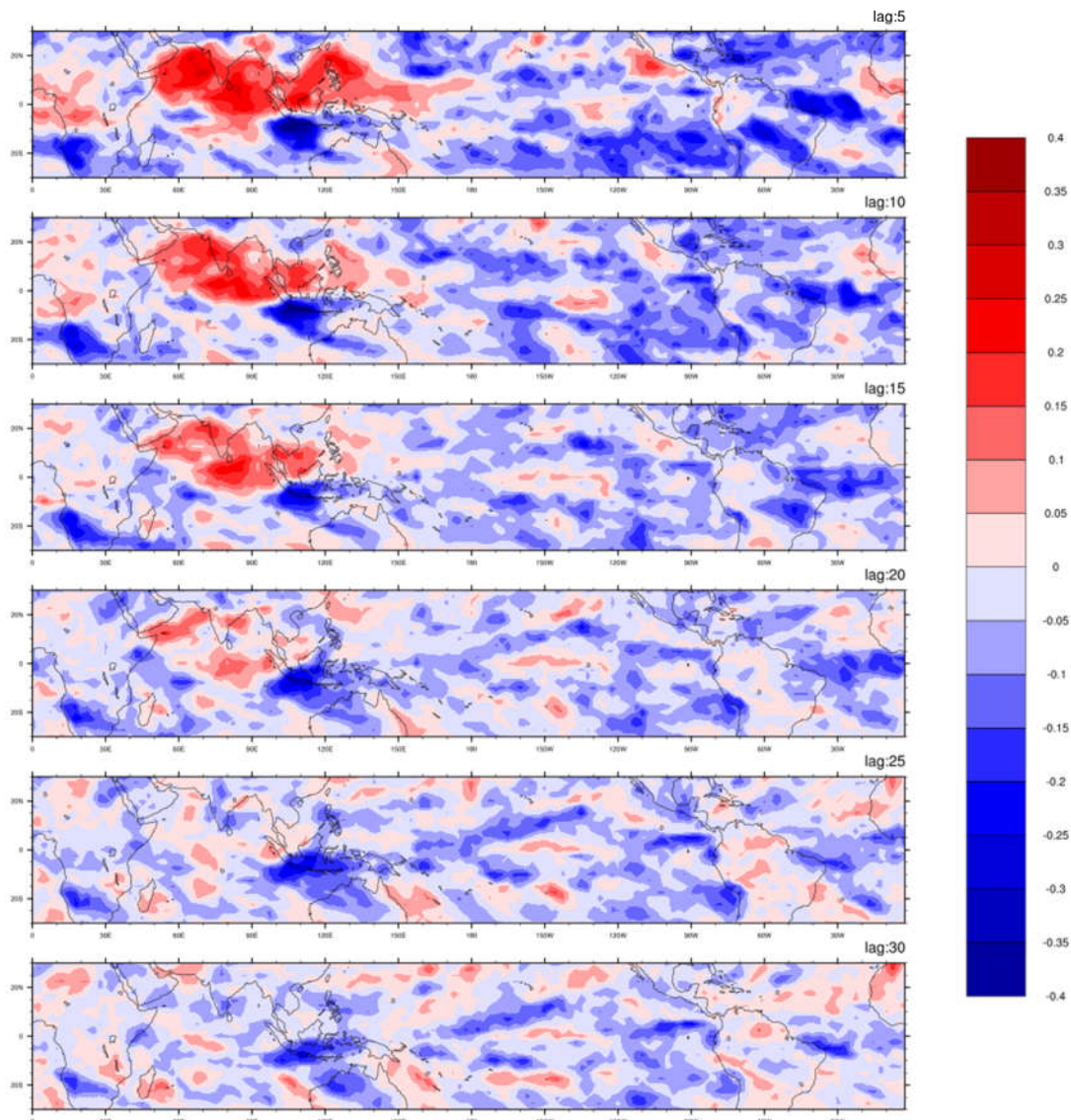


Figure 18. Temporal correlation evolution within observed and j-model forecasted OLR anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal summer.

Fig. 18 shows the j-model predicted and observed OLR anomalies' spatial correlation during summer. High positive correlation over the Indian summer monsoon domain which can be the implication of northward shift of MJO which is consistent with the past studies and persistent till 20 days lead. A consistent negative signal over the Indonesian region is also noticeable for 20 days lead.

Correlation using omi-jmodel between predicted and observed djf olr

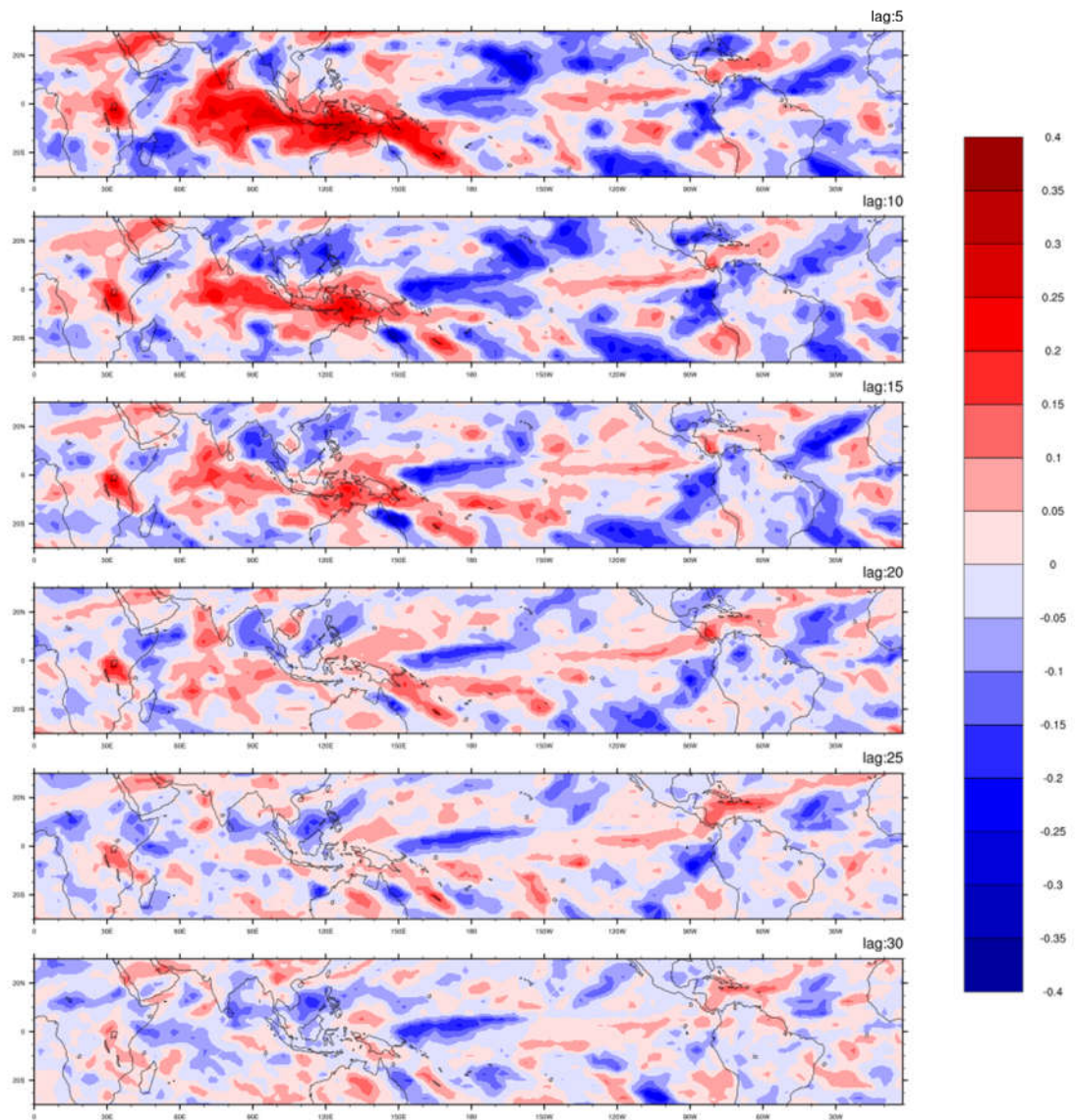


Figure 19. Temporal correlation evolution within observed and j-model forecasted OLR anomalies are plotted for lead days from 5 to 30 with an increment of 5 days for boreal winter.

Fig 19 shows the j-model predicted and observed OLR anomalies' spatial correlation during winter . High positive and coherent correlation signal can be seen which is mostly confined to the equator and is due to the reason that MJO is more confined to the equator during winter. A reasonably good positive correlation over east Africa is also visible which is decreasing with increasing number of lead days.

4.3.1 Skill Analysis of OMI Index

4.3.1.1 Anomaly correlation coefficients:

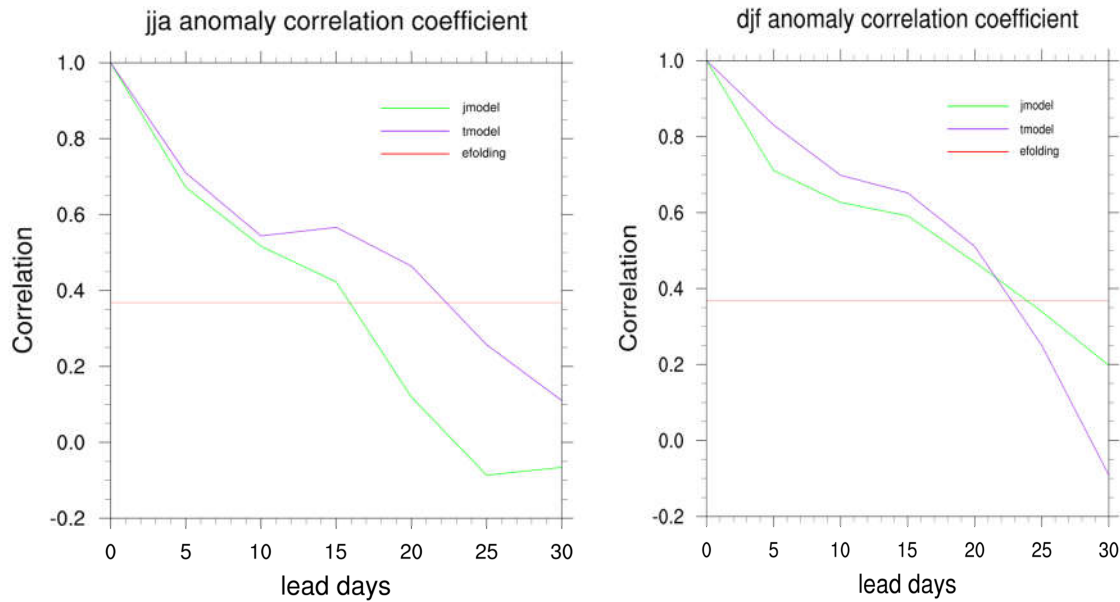


Figure 20. Anomaly correlation Coefficient (ACC) is plotted against number of lead days for j-model (green color) and t-model (purple color) for Boreal summer (Left) and Boreal winter (Right) calculated using OMI indices.

Fig. 20 shows the lead day vs. correlation plots of seasonal Anomaly correlation coefficients for two models. Both the models shows better prediction skill till 15-24 days as compared to the skill for RMM indices and here too t-model is comparable or better than the j-model in both the cases.

4.3.1.2 Root mean square error:

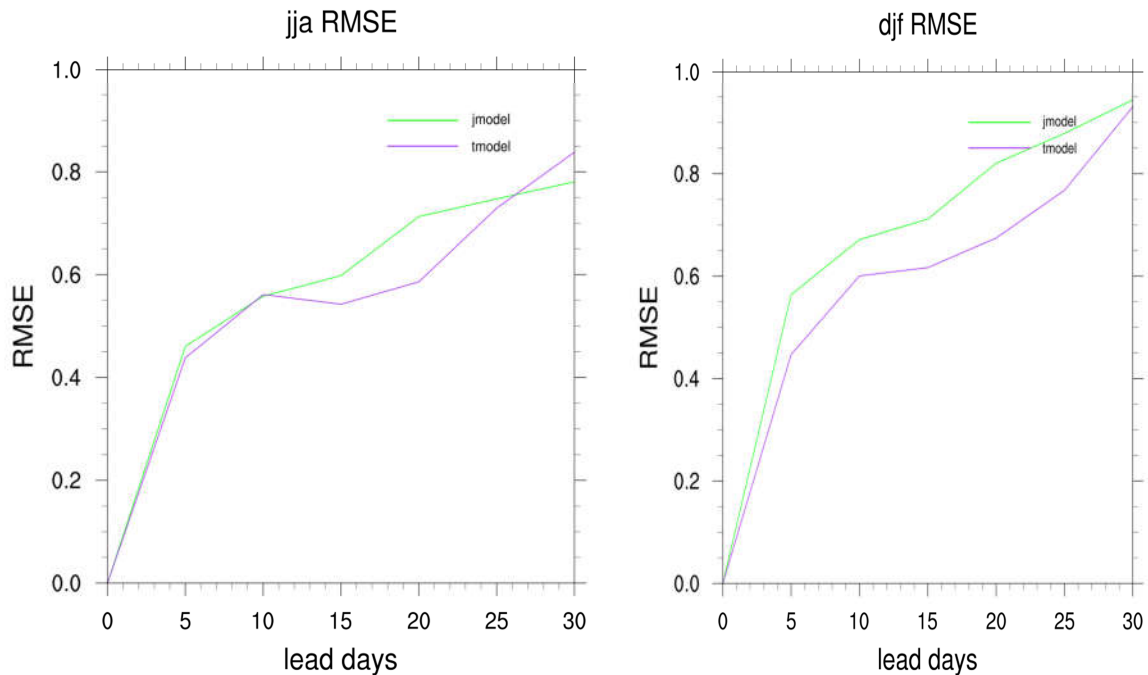


Figure 21. Root Mean Square Error (RMSE) is plotted against number of lead days for j-model (green color) and t-model (purple color) for Boreal summer (Left) and Boreal winter (Right) calculated using OMI indices.

Fig. 21 shows the comparison of seasonal root mean square error between two model for predicted and observed value of OMI indices. Increasing value of RMSE with increasing number of lead days can be observed which the trend we expect from any prediction model is. Values are high in case of winter season as compared to the summer season. Values observed for t-model are always comparable or less than j-model which conclude that t-model is better in prediction as compared to j-model for both the seasons.

4.3.1.3 Phase Angle Error:

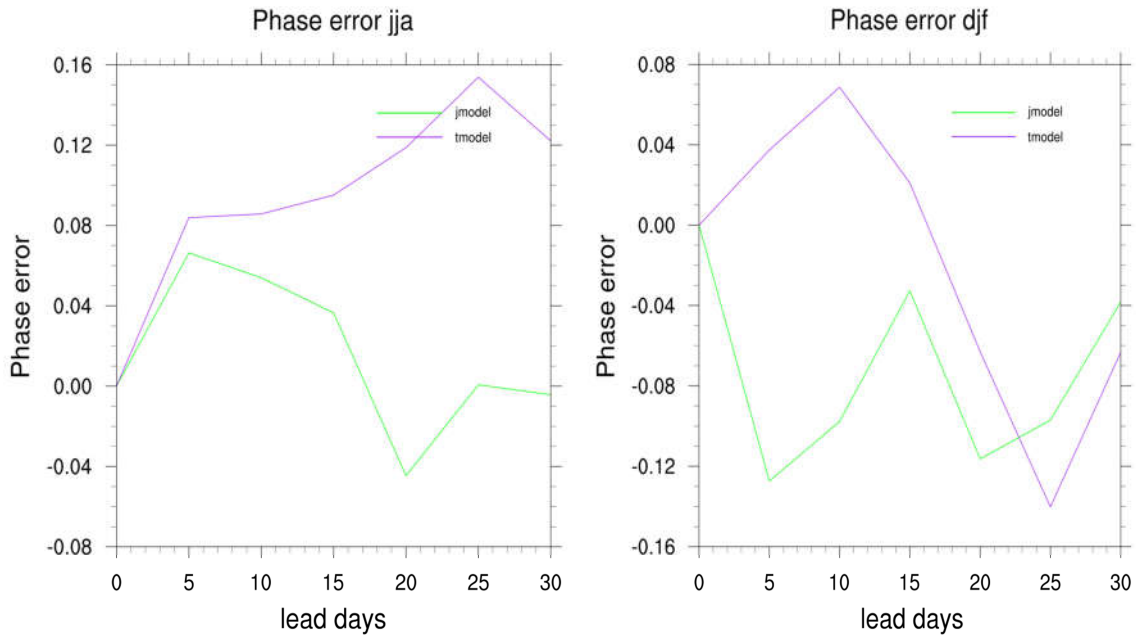


Figure 22. Phase Error is plotted against number of lead days for j-model (green color) and t-model (purple color) for Boreal summer (Left) and Boreal winter (Right) calculated using OMI indices.

Fig. 22 shows the comparison of seasonal phase angle error between two models for predicted and observed OMI indices. Values for t-model in case of summer season are always positive which implies faster propagation in prediction in comparison to the observed value. Similarly, other patterns can also be analyzed by using the definition of phase angle error.

4.4 MVR model using MISO index

The bivariate index called the MISO index represents position and strength of northward propagating component of ISO.

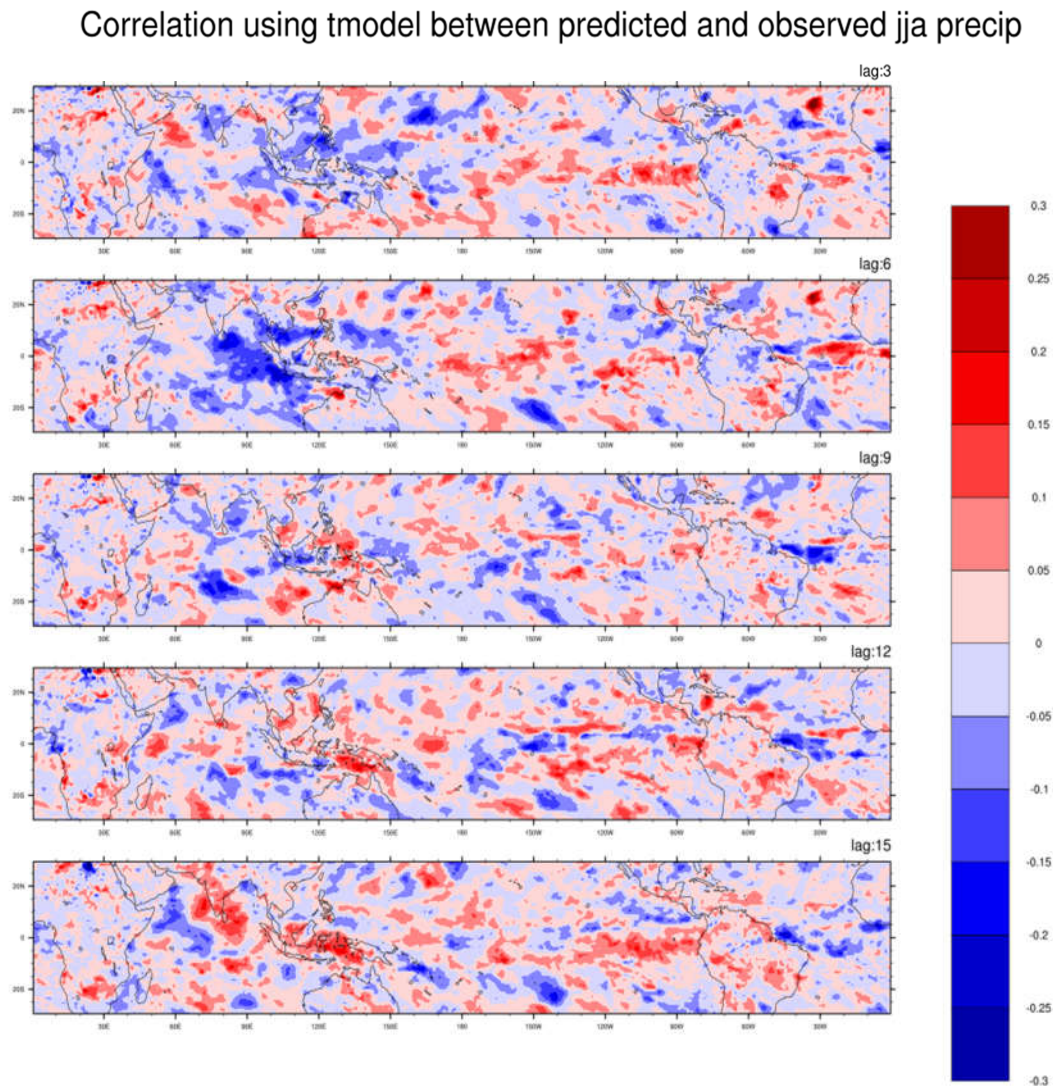


Figure 23. Temporal correlation evolution within observed and t-model forecasted precipitation anomalies are plotted for lead days from 3 to 15 with an increment of 3 days for boreal summer.

Fig. 23 shows the t-model predicted and observed Precipitation anomalies' spatial correlation during summer using t-model. We can observe the negative correlation signal over the western coast of India near Arabian Sea branch and positive correlation signal in Bay of Bengal branch at 3 day lag which got reversed after 15 days lag. This implies that reasonably good prediction of ISO signal over the

Indian summer monsoon domain can be done with lead of 15 days. We expect the MISO prediction skill during 15 days lead time to be reasonably good.

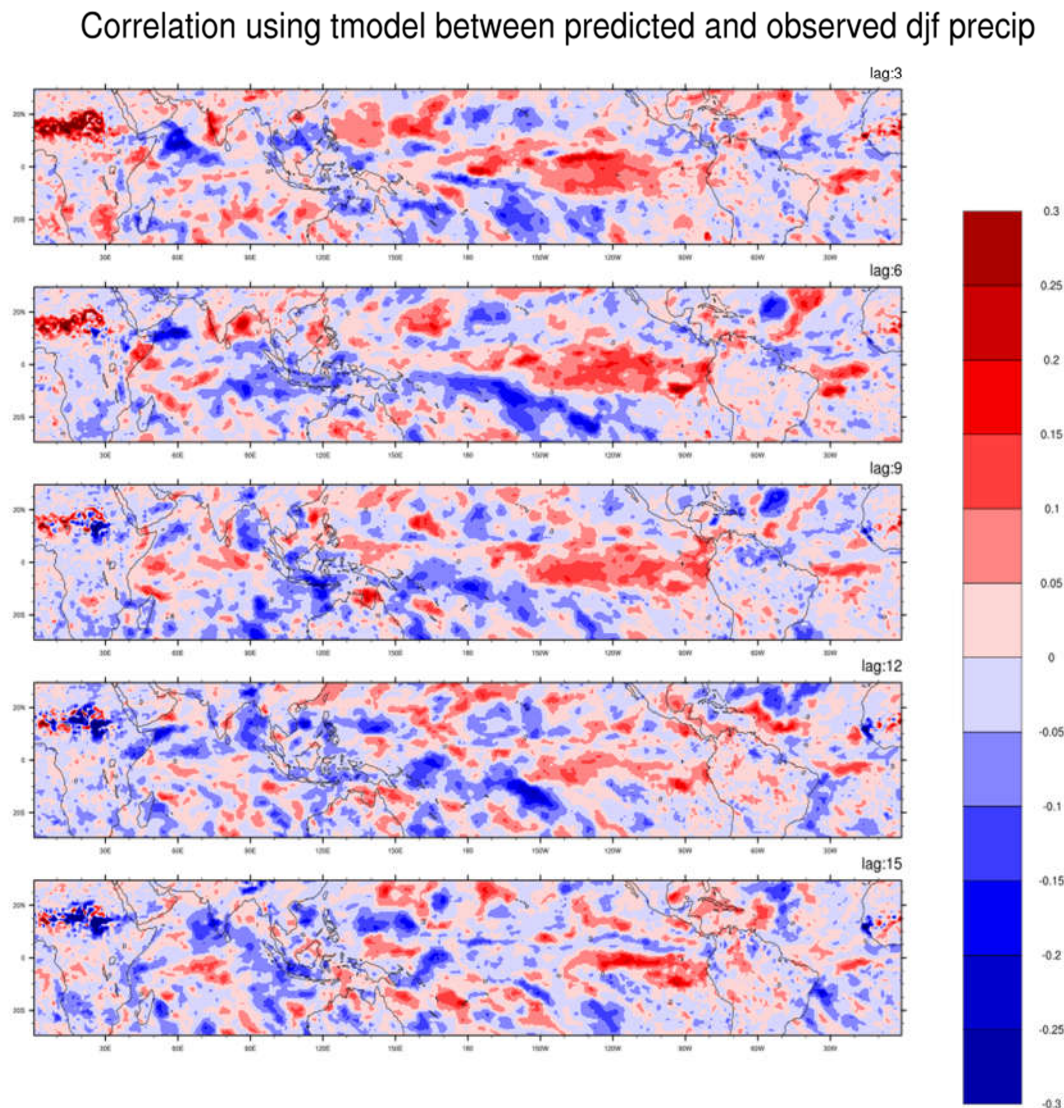


Figure 24. Temporal correlation evolution within observed and t-model forecasted precipitation anomalies are plotted for lead days from 3 to 15 with an increment of 3 days for boreal winter.

The spatial correlation of t-model predicted and observed precipitation (Fig.24) shows reasonably good positive correlation over the Southern Indian monsoon domain, north equatorial African, central and eastern equatorial Pacific regions with a lead of 3 days. A relatively strong negative correlation region can also be seen over the western Equatorial Indian Ocean. During boreal winter, as was observed in the case of boreal summer, the positive correlation over the south Indian

region flips its sign which imply that this relation is more or less independent of season.

Correlation using jmodel between predicted and observed jja precip

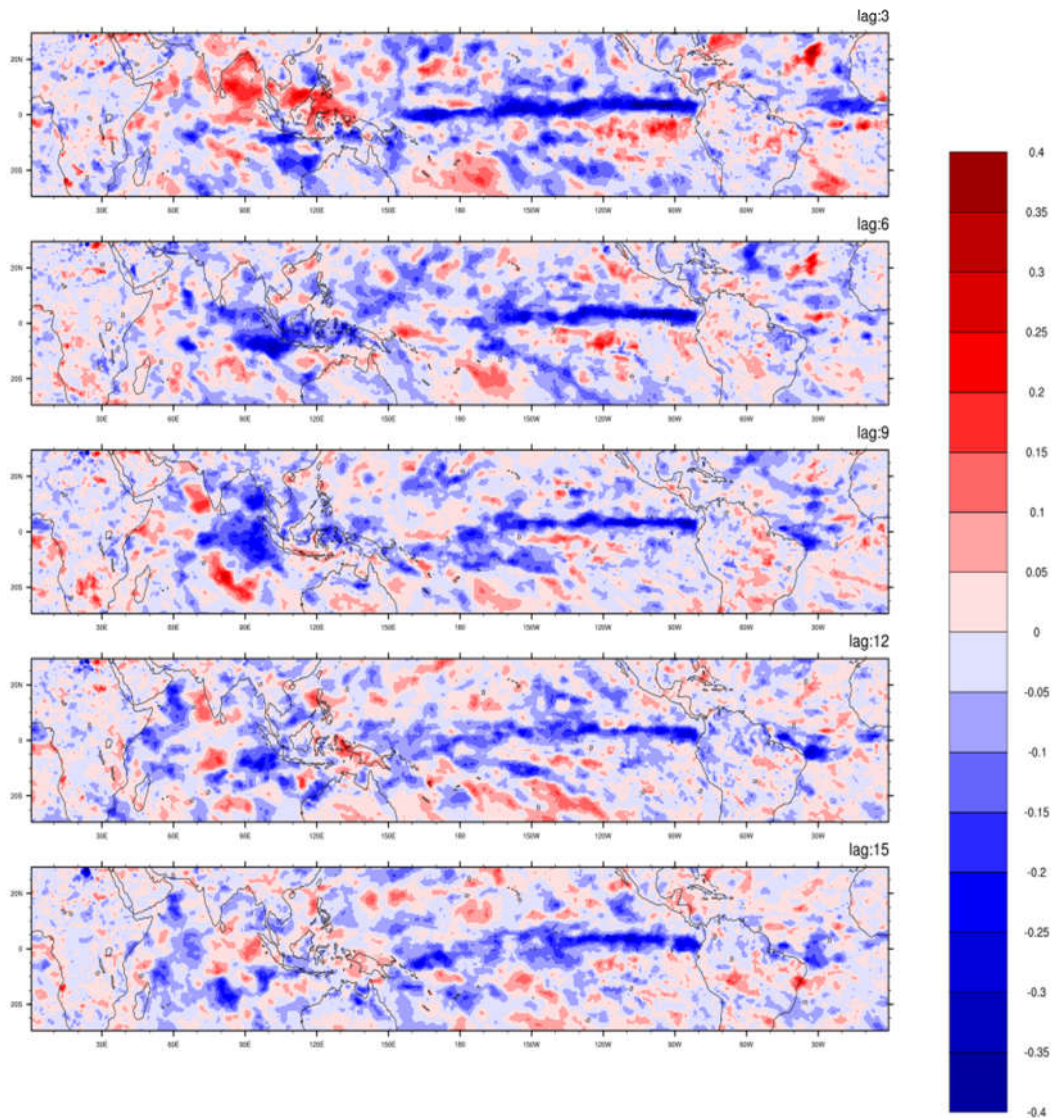


Figure 25. Temporal correlation evolution within observed and j-model forecasted precipitation anomalies are plotted for lead days from 3 to 15 with an increment of 3 days for boreal summer.

In case of spatial correlation between j-model predicted and observed precipitation during JJA (fig.25) we can see positive correlation over the Indian monsoon domain though it is not very strong for a lead of 3 days. The relatively strong positive correlation over the Bay of Bengal and eastern equatorial Indian Ocean is another striking feature which extends to western equatorial Pacific for 3

days lead. A relatively strong negative correlation over central and eastern equatorial Pacific can be seen which is persistent from 3 days to 15 days lead time. A reasonable negative correlation which spatially spreads from 6 days lead to 9 days lead can be observed over Bay of Bengal, maritime continent, and the eastern equatorial Indian Ocean regions.

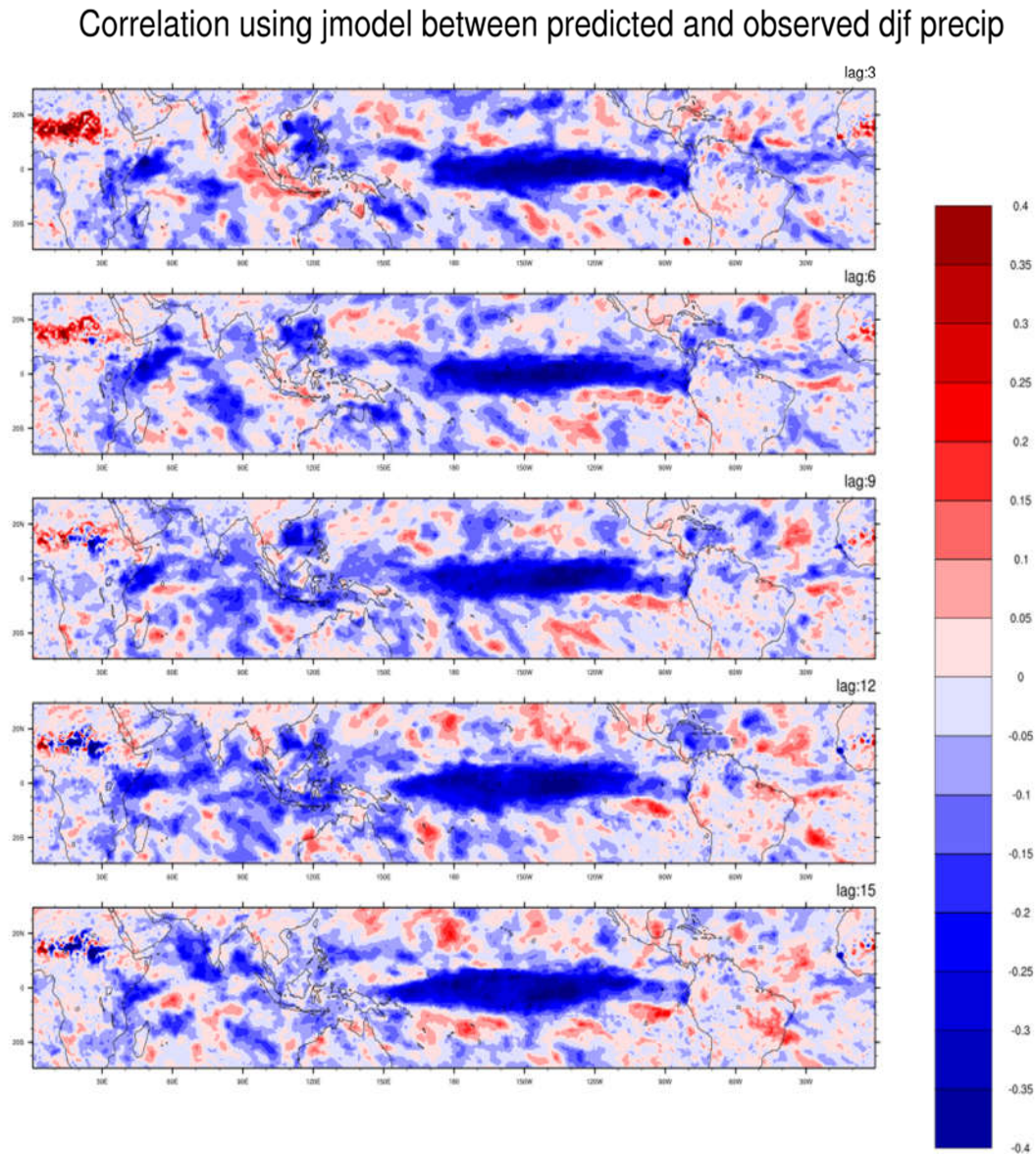


Figure 26. Temporal correlation evolution within observed and j-model forecasted precipitation anomalies are plotted for lead days from 3 to 15 with an increment of 3 days for boreal winter.

In case of spatial correlation between j-model predicted and observed precipitation during DJF (fig.26) relatively strong negative correlation over Equatorial

Pacific which extends from western equatorial Pacific till eastern equatorial Pacific can be seen for all leads beginning from 3 days to 15 days. The same feature can be seen over the eastern and western equatorial Indian Ocean as well. The correlation over the Indian monsoon domain is more or less negative for all the lead times from 3 to 15 days and it strengthens the relation as the lead time increases.

4.4.1 Skill Analysis for MISO

4.4.1.1 Anomaly Correlation Coefficients:

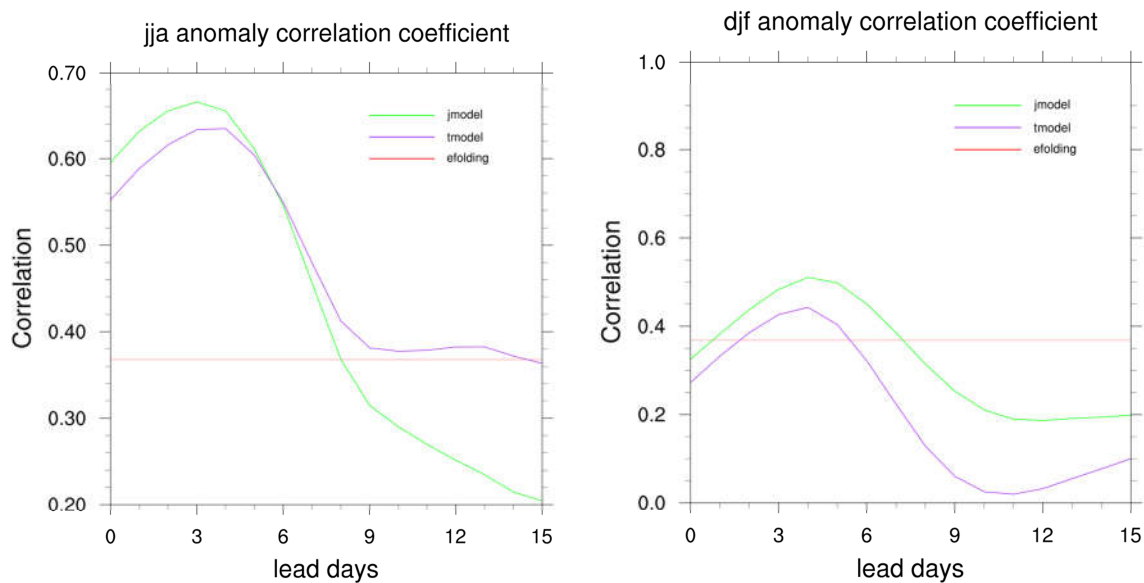


Figure 27. Anomaly Correlation Coefficient (ACC) is plotted against number of lead days for j-model (green color) and t-model (purple color) for Boreal summer (Left) and Boreal winter (Right) calculated using MISO indices.

The bivariate anomaly correlation coefficient in case of MISO index for the 2 models is compared and is plotted for boreal winter as well as summer. The Boreal summer MISO skill analysis shows that the j-model depicts relatively good skill till 6 days lead time and afterwards drops its skill rapidly with lead time and crosses the e-folding time at around 8 days lead time. t-model unlike j-model shows relatively less skill during the 1st few day of lead time and then reaches a steady state, which is above the e-folding time, after which the skill is more or less constant with further increase in lead time.

The MISO skill analysis for winter season shows that the skill is relatively less as compared to summer season as this index is created to capture the boreal summer ISO and the skill increases with lead time initially and then decreases which is a striking feature which require further investigation. The skill for both the models is maximum at a lead time of 4 days and then declines to lower values and crosses the e-folding time around 5 and 8 days respectively for t-model and t-model.

4.4.1.2 Root Mean Square Error:

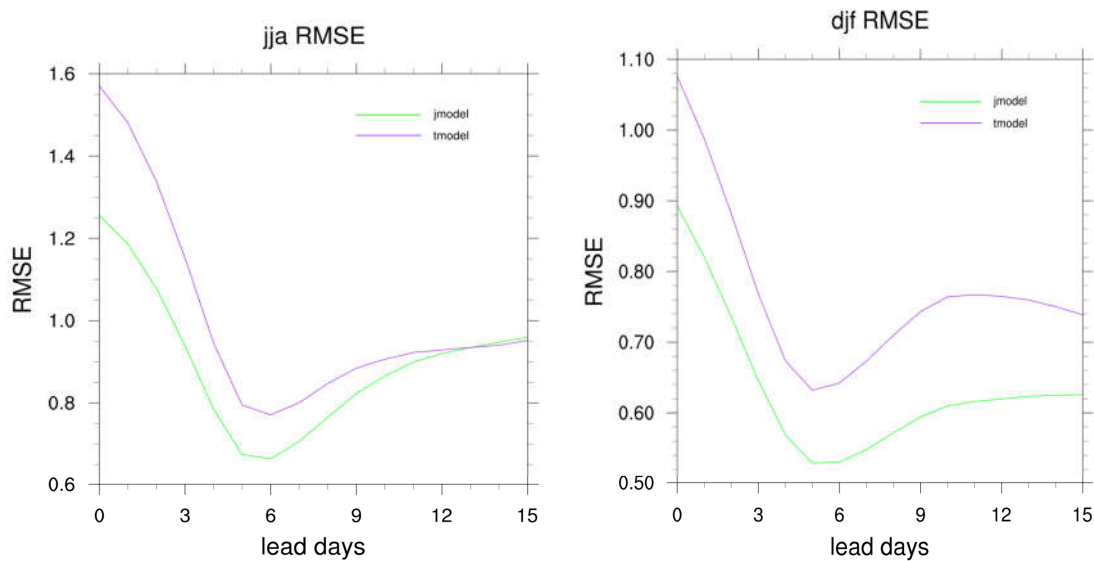


Figure 28. RMSE is plotted against number of lead days for j-model (green color) and t-model (purple color) for Boreal summer (Left) and Boreal winter (Right) calculated using MISO indices.

The RMSE first show a rapid decrease and then a gradual increase with lead time. The minimum RMSE corresponds to a lead time of 5 days lead time for both the considered MVR models.

4.4.1.3 Phase Angle Error:

The phase error analysis in case of t-model and j-model for summer season reveals that the MISO propagation is faster than observation in t-model throughout the prediction period. In case of j-model the MISO propagation is initially faster than observation and then it lags behind the observation and this trend is repeated as the lead time increases. During winter the MISO propagation in j-model and t-model are more or less slower than that of observation except for t-model around 3-7 days lead.

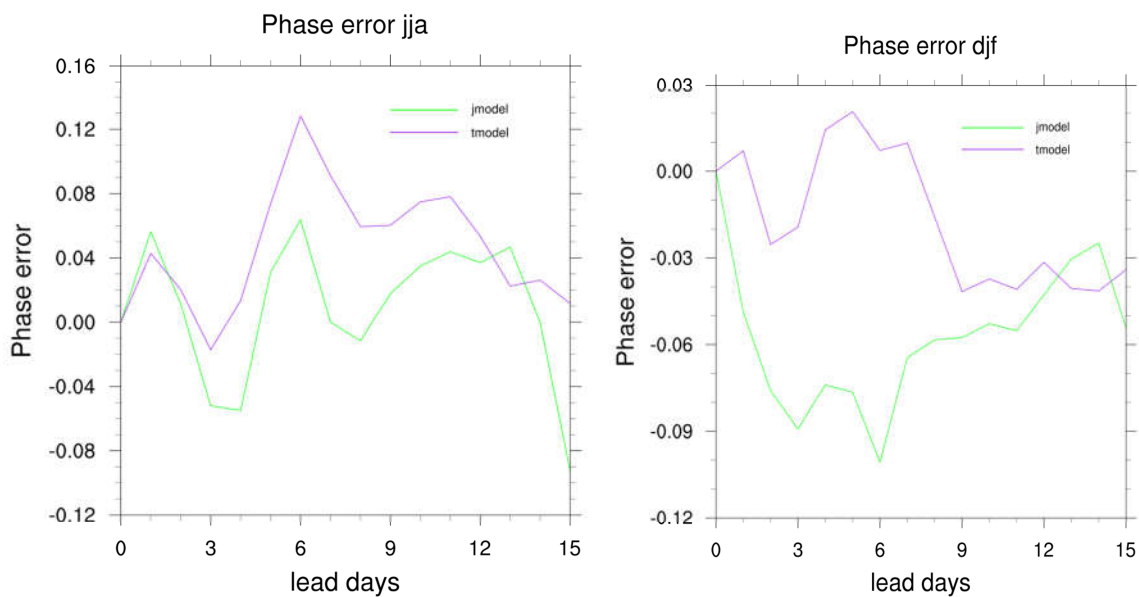


Figure 29. Phase error is plotted against number of lead days for j-model (green color) and t-model (purple color) for Boreal summer (Left) and Boreal winter (Right) calculated using MISO indices.

5. CONCLUSION

The following conclusions are made from the analysis carried out using the MVR forecast model.

1. The MVR forecast model using the OMI index shows higher prediction skill during boreal winter and boreal summer and could be preferred for operational prediction of the ISOs.
2. In general t-model is showing better prediction skill as compared to j-model.
3. The prediction skill of the OMI and RMM index based MVR model is higher than the MISO index based MVR model.
4. Temporal correlation for u200 is relatively high and coherent as compared to OLR and u850 fields of MJO in case of MVR model created using RMM.
5. The spatial correlation shows that the spatial structure of ISO during winter (summer) is captured by t (j)MVR model created using OMI.

MVR model had been studied extensively and had been employed to predict the ISOs. More new and efficient models will be explored and the existing ones will be refined to get better results and predications as well.

6. REFERENCES

- Alexander, G., R. Keshavamurty, U. De, R. Chellapa, S. Das, and P. Pillai (1978) Fluctuations of monsoon activity. *J. Meteorol. Hydrol. Geophys.*, **29**, 76–87.
- Dee DP, Uppala SM, Simmons AJ, et al (2011) The ERA-Interim reanalysis : configuration and performance of the data assimilation system. 553–597. doi: 10.1002/qj.828
- Dakshinamurthy, J. and R. N. Keshavamurthy (1976) On oscillations of period around one month in the Indian summer monsoon. *Ind. J. Meteorol. Geophys.*, **27**, 201–203.
- Goswami BN, Venugopal V, Chattopadhyay R (2018) South Asian Monsoon Extremes. *Trop Extrem* 15–49 doi:10.1016/b978-0-12-809248-4.00002-9
- Gutzler, D.S. and J. Shukla, 1984: Analogs in the Wintertime 500 mb Height Field. *J. Atmos. Sci.*, **41**, 177–189,
- Hsu P-C, Li T, You LJ, Gao JY, Ren HL (2015) A spatial-temporal projection method for 10–30-day forecast of heavy rainfall in Southern China. *Clim Dyn* 44:1227–1244
- Jones, C., D.E. Waliser, K.M. Lau, and W. Stern, 2004: The Madden–Julian Oscillation and Its Impact on Northern Hemisphere Weather Predictability. *Mon. Wea. Rev.*, **132**, 1462–1471,
- Jiang X, Wheeler MC, Jones C, et al (2008) Assessing the Skill of an All-Season Statistical Forecast Model for the Madden–Julian Oscillation. *Mon Weather Rev* 136:1940–1956. doi: 10.1175/2007mwr2305.1
- Kiladis, G.N., J. Dias, K.H. Straub, M.C. Wheeler, S.N. Tulich, K. Kikuchi, K.M. Weickmann, and M.J. Ventrice, 2014: A Comparison of OLR and Circulation-Based Indices for Tracking the MJO. *Mon. Wea. Rev.*, **142**, 1697–1715, <https://doi.org/10.1175/MWR-D-13-00301.1>
- Kim H, Vitart F, Waliser DE (2018) Prediction of the Madden-Julian oscillation:A review. *J Clim* 31:9425–9443. doi:10.1175/JCLI-D-18-0210.1

- Krishnamurti, T.N. and H.N. Bhalme, 1976: Oscillations of a Monsoon System. Part I. Observational Aspects. *J. Atmos. Sci.*, **33**, 1937–1954,
- K. Hasselmann Geophysical OF (1988) PIPs and POPs ' The Reduction of Complex Dynamical Systems Using Principal Interaction and Oscillation Patterns. **93**:
- Lo, F. and H.H. Hendon, 2000: Empirical Extended-Range Prediction of the Madden–Julian Oscillation. *Mon. Wea. Rev.*, **128**, 2528–2543,
- Mani NJ, Goswami BN (2014) An Indian monsoon intraseasonal oscillation (MISO) index for real time monitoring and forecast verification An Indian monsoon intraseasonal oscillations (MISO) index for real time monitoring and forecast verification. doi: 10.1007/s00382-012-1462-5
- Maharaj EA, Wheeler MC (2005) Forecasting an index of the madden-oscillation. *Int J Climatol* 25:1611–1618. doi: 10.1002/joc.1206
- Mo, K.C., 2001: Adaptive Filtering and Prediction of Intraseasonal Oscillations. *Mon. Wea. Rev.*, **129**, 802–817,
- North, G.R., T.L. Bell, R.F. Cahalan, and F.J. Moeng, 1982: Sampling Errors in the Estimation of Empirical Orthogonal Functions. *Mon. Wea. Rev.*, **110**, 699–706
- Roundy PE (2017) Diagnosis of seasonally varying regression slope coefficients and application to the MJO. 1946–1952. doi: 10.1002/qj.3054
- Sikka, D.R. and S. Gadgil, 1980: On the Maximum Cloud Zone and the ITCZ over Indian, Longitudes during the Southwest Monsoon. *Mon. Wea. Rev.*, **108**, 1840–1853,
- van den Dool, H.M., 1989: A New Look at Weather Forecasting through Analogues. *Mon. Wea. Rev.*, **117**, 2230–2247,
- von Storch, H. and J. Xu (1990) Principal oscillation pattern analysis of the 30- to 60-day oscillation in the tropical troposphere. *Climate Dynamics*, **4**, 175-190

- Waliser, D.E., C. Jones, J.E. Schemm, and N.E. Graham, 1999: A Statistical Extended-Range Tropical Forecast Model Based on the Slow Evolution of the Madden–Julian Oscillation. *J. Climate*, **12**, 1918–1939,
- Wheeler, M. and K.M. Weickmann, 2001: Real-Time Monitoring and Prediction of Modes of Coherent Synoptic to Intraseasonal Tropical Variability. *Mon. Wea. Rev.*, **129**, 2677–2694,
- Wheeler, M.C. and H.H. Hendon, 2004: An All-Season Real-Time Multivariate MJO Index: Development of an Index for Monitoring and Prediction. *Mon. Wea. Rev.*, **132**, 1917–1932,
- Zhu Z, Li T (2017) The statistical extended - range (10 – 30 - day) forecast of summer rainfall anomalies over the entire China. *Clim Dyn* **48**:209–224. doi: 10.1007/s00382-016-3070-2
- Zhu Z, Li T (2017) Statistical extended-range forecast of winter surface air temperature and extremely cold days over China. 1528–1538. doi: 10.1002/qj.3023