

LPS Neural Operator (LPSNO): A Novel Deep Learning Framework to Predict the Indian Monsoon Low-Pressure Systems

K. S. S. Sai Srujan¹, Sukumaran Sandeep¹, and Hariprasad Kodamana¹

¹Indian Institute of Technology Delhi

April 16, 2024

Abstract

The synoptic scale variability of the Indian summer monsoon (ISM) is contributed by the weak cyclonic vortices known as low-pressure systems (LPSs). LPSs are the primary mechanism by which central Indian plains receive rainfall. Traditionally, synoptic variability is considered to have a low predictability. In the present study, we developed a framework, namely, LPS Neural Operator (LPSNO), using the neural operator-based deep learning to predict the spatial structure of daily mean sea level pressure anomalies over the Bay of Bengal at a resolution of $1^\circ \times 1^\circ$. The proposed neural operator extends the Fourier neural operator framework by employing convolutional LSTMs in the operator backbone. Further, the mean sea level pressure is reconstructed using the predicted anomaly and the climatology, which is then used to track the LPSs using a Lagrangian tracking algorithm. The median pattern correlation between the predicted and actual mean sea-level pressure anomalies over the BoB is about 88%, 60%, and 50% for 24, 48, and 72-hour forecasts, respectively. The proposed model improves the accuracy of predictions compared with the earlier ConvLSTM models. The pattern correlation between the observed and predicted synoptic activity index (SAI) is 0.94, 0.9, and 0.87 for 1, 2, and 3-day ahead predictions, respectively. A well-trained model of LPSNO takes only ~ 3.2 s to generate a one-day forecast on a single GPU node of Nvidia V100, which is computationally extremely cheap compared to the conventional numerical weather prediction models. The proposed LPSNO can advance operational weather forecasting substantially.

1 **LPS Neural Operator (LPSNO): A Novel Deep**
2 **Learning Framework to Predict the Indian Monsoon**
3 **Low-Pressure Systems**

4 **K. S. S. Sai Srujan¹, S. Sandeep^{1,2}, and Hariprasad Kodamana^{2,3}**

5 ¹Centre for Atmospheric Sciences, Indian Institute of Technology Delhi, New Delhi 110016, India

6 ²Yardi School of Artificial Intelligence, Indian Institute of Technology Delhi, New Delhi 110016, India

7 ³Department of Chemical Engineering, Indian Institute of Technology Delhi, New Delhi 110016, India

8 **Key Points:**

- 9 • A framework to predict the sea-level pressure anomaly over the Bay of Bengal is pro-
10 posed using the LPS Neural Operator for the first time
- 11 • The well-trained LPS Neural operator takes only a few seconds to generate a one-day
12 forecast over the Indian monsoon domain
- 13 • The pattern correlation between predicted and actual synoptic activity is 94%, 90%,
14 and 87% for 24, 48, and 72-hour forecasts
- 15
- 16
- 17

Corresponding author: S. Sandeep, sandeep.sukumaran@cas.iitd.ac.in

Corresponding author: Hariprasad Kodamana, kodamana@chemical.iitd.ac.in

Abstract

[The synoptic scale variability of the Indian summer monsoon (ISM) is contributed by the weak cyclonic vortices known as low-pressure systems (LPSs). LPSs are the primary mechanism by which central Indian plains receive rainfall. Traditionally, synoptic variability is considered to have a low predictability. In the present study, we developed a framework, namely, LPS Neural Operator (LPSNO), using the neural operator-based deep learning to predict the spatial structure of daily mean sea level pressure anomalies over the Bay of Bengal at a resolution of $1^\circ \times 1^\circ$. The proposed neural operator extends the Fourier neural operator framework by employing convolutional LSTMs in the operator backbone. Further, the mean sea level pressure is reconstructed using the predicted anomaly and the climatology, which is then used to track the LPSs using a Lagrangian tracking algorithm. The median pattern correlation between the predicted and actual mean sea-level pressure anomalies over the BoB is about 88 %, 60 %, and 50 % for 24, 48, and 72-hour forecasts, respectively. The proposed model improves the accuracy of predictions compared with the earlier ConvLSTM models. The pattern correlation between the observed and predicted synoptic activity index (SAI) is 0.94, 0.9, and 0.87 for 1, 2, and 3-day ahead predictions, respectively. A well-trained model of LPSNO takes only ~ 3.2 s to generate a one-day forecast on a single GPU node of Nvidia V100, which is computationally extremely cheap compared to the conventional numerical weather prediction models. The proposed LPSNO can advance operational weather forecasting substantially.]

Plain Language Summary

[The weak cyclonic vortices during the Indian Summer Monsoon (ISM) season, commonly known as Low-Pressure Systems (LPSs), are predominantly present over the head Bay of Bengal (BoB). More than half of ISM rainfall over the hugely populated Indo-Gangetic plains is contributed from LPSs, making them an important component of the hydrological cycle over South Asia with huge socio-economic impacts. Therefore, the prediction for LPS genesis will be helpful in better disaster preparedness and food security planning. A machine learning (ML) framework is developed initially to predict the spatial map of perturbations in mean sea level pressure (MSLP). Using the predicted perturbations and climatology, the full MSLP field is reconstructed. The LPSs are tracked from the MSLP field. The correlation of the spatial map of fluctuations in MSLP between actual and predicted is about 88%, 60%, and 50% at a lead time of 24, 48, and 72-hours. A well-trained ML model will be computationally efficient compared to traditional numerical weather prediction models.]

1 Introduction

The Indian summer monsoon (ISM) is an important component of the hydrological cycle of South Asia, which is essential to the water security of more than 1.5 billion inhabitants. Relatively weaker synoptic-scale cyclonic vortices embedded in the ISM circulation are known as the low-pressure systems (LPSs). Although the LPSs form in all monsoon regions around the world, they are most prominent in the ISM domain, with about 12 systems forming in each June - September period (Hurley & Boos, 2015). These are the main rain bearing systems, with a life span of 3–7 days and a diameter of 1000 – 2000 km (D. R. Sikka, 1977) contributing nearly half of the ISM rainfall over the Indo-Gangetic plains (Krishnamurthy & Ajayamohan, 2010; Praveen et al., 2015; Hunt & Fletcher, 2019; Sandeep et al., 2018; Thomas et al., 2021; Deoras et al., 2021). Conventionally, the propagation of LPSs have been identified manually using the surface pressure charts for which a long term archive exists (Mooley & Shukla, 1987; Sikka, 2006; Krishnamurthy & Ajayamohan, 2010). The dynamical and statistical models face difficulties in predicting the synoptic-scale rainfall (B. Wang et al., 2005, 2015; Saha et al., 2019). The chaotic nature of ISM makes the prediction of synoptic scale variability challenging (Goswami et al., 2006; Saha et al., 2019).

67 The conventional NWP models solve the prognostic partial differential equations using
 68 numerical methods, such as the finite difference method, which discretize time and space.
 69 Therefore, there is always a trade off between the grid resolution and the computational
 70 time. The computational stability of NWP models is measured by a condition called as CFL
 71 condition (Courant et al., 1967). However, data-driven deep learning models are not limited
 72 by these issues (Greenfeld et al., 2019; Kochkov et al., 2021; Li et al., 2020). For example,
 73 Convolutional Neural Networks (CNNs) in predicting the ENSO (Ham et al., 2019; Gupta
 74 et al., 2020) and estimating the intensity of the tropical cyclone (C. Wang et al., 2022;
 75 R. Zhang et al., 2020). The CNNs are known for the application of spatial data prediction.
 76 Also, the convolutions with the long short-term memory (ConvLSTM) show robust skills in
 77 nowcasting the precipitation (Shi et al., 2015) and predicting sea level pressure time series
 78 (Sinha et al., 2021).

79 The recent rapid advancement of data driven deep learning (DL) models suggest that
 80 they can be useful in the prediction of atmospheric and oceanic states (Q. Zhang et al.,
 81 2017; Ham et al., 2019; Gupta et al., 2020; Sinha et al., 2021; Andersson et al., 2021; Chen
 82 et al., 2022; Ling et al., 2022; Kurth et al., 2023). Recent reports suggest that the DL
 83 models are able to generate short and medium range weather forecasts globally with a skill
 84 that matches the best NWP models Bi et al. (2023); Lam et al. (2023); Y. Zhang et al.
 85 (2023). These developments suggest that the DL models can soon be cheaper alternatives
 86 to the computationally expensive NWP models. The DL models have also shown potential
 87 in sub-seasonal to seasonal scale forecasts as well (Weyn et al., 2021). They can also used
 88 in combination with the NWP models to improve the forecast skill (Rojas-Campos et al.,
 89 2023). Gupta et al. (2020) predicted the ENSO beyond the spring predictability barrier
 90 using the ConvLSTM, whereas the traditional models are unable to. Recently, a developed
 91 framework named Fourier Neural Operator (FNO) shows a robust skill in predicting partial
 92 differential equations (Li et al., 2020; Lu et al., 2021; Kossaifi et al., n.d.; Azzizadenesheli
 93 et al., 2023; Wen et al., 2023). However, the skill of DL models in forecasting the extreme
 94 weather events, such as tropical cyclones and monsoon LPSs, is yet to be proven.

95 Earlier studies predicted the MSLP time series, and used it as a proxy to predict the
 96 strength of active and break cycles of ISM (Sinha et al., 2021). However, the spatial map
 97 of MSLP anomalies are not yet predicted using a deep learning model. Here, we propose a
 98 framework to predict the genesis and track of LPSs by using a two step approach. Firstly, we
 99 predict the spatial pattern of daily MSLP anomalies using FNO. Secondly, we reconstructed
 100 the total MSLP field by adding the predicted anomalies to daily climatology. Then, the LPSs
 101 are tracked using an automated algorithm developed by Praveen et al. (2015), which mimics
 102 the manual tracking of trajectories of LPSs from the surface pressure charts. Though the
 103 model is used to predict the LPSs, it can also be potentially extended to predict tropical
 104 cyclones. The data and methodology are explained in section 2, the framework and the skill
 105 for predicting LPSs are shown in section 3, and the conclusions are presented in section 4.

106 2 Data and methods

107 Data

108 The daily mean sea-level pressure (MSLP) of the European Centre for Medium-Range
 109 Weather Forecasts fifth-generation (ERA5) reanalysis dataset (Hersbach et al., 2020) at a
 110 spatial resolution of $0.25^\circ \times 0.25^\circ$ from 1979–2018 is used in this study. For training and
 111 validation of the DL model 1979 – 2007 is used, and prediction is done for 2008 – 2018. The
 112 MSLP anomalies are computed by removing the long-term mean from 1979 – 2018. The
 113 region considered for the present study is $75^\circ\text{E} - 90^\circ$, $10^\circ\text{N} - 25^\circ\text{N}$. The LPSs are tracked
 114 using the algorithm developed by Praveen et al. (2015) from ERA5 using the daily MSLP.
 115 This LPS tracking algorithm identifies closed isobars at every one hPa interval from gridded
 116 MSLP data, and the storm’s center is identified as the centroid of the innermost isobar. The
 117 LPS centers identified from consecutive time intervals of gridded MSLP data are connected

118 to get the track. This algorithm mimics the manual tracking of LPSs from the pressure
 119 charts used by the India Meteorological Department and has been found to have a robust
 120 skill in tracking LPSs (Praveen et al., 2015).

121 LPS tracking

122 The LPS over the BoB (65° – 95°E and 0 – 23°N) and TC tracks over the WNP
 123 (110°E – 180°E and 0 – 30°N) from the model experiments are tracked using the algorithm
 124 developed by the Praveen et al. (2015), which mimics the conventional manual tracking
 125 algorithm based on sea level pressure closed isobars over the surface pressure charts. This
 126 algorithm searches for closed isobar at 1hPa interval at every time step around the grid
 127 of SLP minimum, and the storm center is taken as the centroid of the innermost closed
 128 isobar. The pressure depth (Δ SLP) is considered as the difference between the outermost
 129 and innermost closed isobar, and it signifies the intensity of the storm. Δ SLP \leq 2 hPa is
 130 called “low”, 2 hPa $<$ Δ SLP \leq 4 hPa is “depression”, 4 hPa $<$ Δ SLP \leq 10 hPa is “deep
 131 depression”, 10 hPa $<$ Δ SLP \leq 16 hPa is “cyclonic storm”, and Δ SLP $>$ 16 hPa is “severe
 132 cyclonic storm” (Mooley & Shukla, 1987; Sikka, 2006; Praveen et al., 2015).

133 Synoptic Activity Index

134 The genesis location, number of LPS days, and storm intensity of LPSs are together
 135 explained by defining an index named “Synoptic Activity Index” (SAI; (Ajayamohan et al.,
 136 2010)). The Δ SLP (pressure depth) measures the storm intensity. SAI is defined as the
 137 track density of LPS weighted by wind speed.

$$SAI = \sum_{n=0}^{n=l} \sum_{x-\Delta x}^{x+\Delta x} \sum_{y-\Delta y}^{y+\Delta y} U_{cat} \quad (1)$$

138 where l is the life span of an LPS in days, Δx and Δy are the grid spacing (1.5°) in
 139 X and Y directions, and x and y are the longitudinal and latitudinal positions of a storm
 140 center. The values of U_{cat} are 4.25, 11, 15, 20, 27.5 for the categories lows, depressions,
 141 deep depressions, cyclonic storms, and severe cyclonic storms, respectively (Ajayamohan et
 142 al., 2010; Sandeep et al., 2018).

143 ML layers

144 Convolutional layer

145 A convolutional layer is widely used in many computer vision algorithms, including
 146 CNNs that learn the spatial pattern robustly. A Convolution layer is also known as kernel
 147 convolution, where a kernel or filter (small matrix) is multiplied by an image or output from
 148 the previous layer. In a simple understanding, a convolutional layer acts like a spatial filter
 149 and extracts useful features from an image. The mathematical representation of an output
 150 from a convolutional layer is:

$$F[m, n] = X * h[m, n] \quad (2)$$

151

$$F[m, n] = \sum_j \sum_k h[j, k] X[m - j, n - k] \quad (3)$$

152 where $F[m, n]$ is the output feature matrix from a convolutional layer, $X[m, n]$ is the
 153 input image to a convolutional layer of width m and height n . The filter or kernel matrix is
 154 denoted by $h[m, n]$, which is multiplied by the input image.

155 The output from the convolutional layer ($F[m, n]$) is multiplied by a weights tensor
 156 (W), and bias (b) is added while training the model. Then, the output feature matrix

157 is passed to a nonlinear activation function. In each iteration, while optimizing the cost
 158 junction (J), the weights tensor (W) will be updated. The whole process in this layer is
 159 mathematically represented as:

$$Z^{[l]} = F^{[l]} * W^{T[l]} + b^{[l]} \quad (4)$$

$$Y^{[l]} = g(Z^{[l]}) \quad (5)$$

161 where $Y^{[l]}$ is the final output from a convolutional layer while training, $[l]$ denotes the
 162 iteration, and g is a nonlinear activation function.

163 ***Max-pooling layer***

164 A pooling layer aids in reducing the dimensions of a convolutional layer. In general, a
 165 pooling layer is placed just after a convolutional layer. The pooling layer involves sliding a
 166 filter along all channels in a feature matrix. In the case of max pooling, it picks a maximum
 167 value at a particular region of the sliding filter. The output dimensions from a pooling layer
 168 are:

$$(m - f + 1)/s * (n - f + 1)/s * c \quad (6)$$

169 where, m, n, c are the width, height, and number of channels of an image. s is the
 170 stride length and f is the size of the pooling filter. Pooling reduces the dimensions of the
 171 feature matrix from the convolutional layer therefore reducing the number of parameters
 172 to learn by the model and saving the amount of computational time. The pooling layer
 173 summarizes the features in a particular region instead of point-to-point or kernel-to-kernel
 174 mapping therefore helps the model learn robustly irrespective of the position and orientation
 175 of the features in an image.

176 ***Batch normalization layer***

177 A batch normalization layer reduces the covariance shift problem. It normalizes the
 178 intermediate output of each layer within the batch during the training of a model. This
 179 helps in stabilizing the optimization process and reduces the demand for dropout or other
 180 optimizations like the l2 norm. The mathematical representation of the batch normalization
 181 is:

$$X' = (x - M_b[x])/sqrt(var(x)) \quad (7)$$

182 where $M_b[x]$ is the mean of the mini-batch size and $var(x)$ is the variance of the mini-batch
 183 size and X' is the normalized component from the previous layer, and x is the output from
 184 the previous layer and input into the batch normalization layer.

185 It can be further developed as:

$$X'' = \gamma * X' + \beta \quad (8)$$

186 where X'' is the final output from the normalization layer. γ and β are the learnable
 187 parameters during the training of the model.

188 ***Dropout***

189 A dropout is a type of regularization that prevents the overfitting of the data during
 190 the training process. A dropout layer randomly makes the value of the nodes in a layer into
 191 zeros. The number of nodes to make zero depends on the input probability decided while
 192 tuning the hyperparameters.

193 **Flatten and dense layers**

194 As the name implies, a flatten layer flattens the multidimensional output from previous
195 layers to a simple two-dimensional matrix. The output dimensions from a flatten layer are
196 given below:

$$(M_b * m * n * f)_{output} = (M_b * (m * n * f))_{input} \quad (9)$$

197 Where M_b denotes the batch size and $m, n, \text{ and } f$ are image width, height, and kernel size
198 respectively. L.H.S. represents the output dimensions from a flatten layer and R.H.S. rep-
199 represents the input multidimensional matrix to a flatten layer. A dense layer is a regular fully
200 connected layer generally placed after a flatten layer. The operation done by a dense layer
201 is given below:

$$Y = g(\Sigma(X * W^T) + B) \quad (10)$$

202 Where Y is the output from a dense layer, g is the nonlinear activation function, X is the
203 input vector to a dense layer, W^T is a matrix of weights, and B is a bias vector.

204 **ConvLSTM and Fourier layer**

205 A ConvLSTM layer is a combination of a convolutional layer followed by a LSTM
206 layer. The LSTM layer is a type of recurrent neural network that learns the sequential
207 data, and the convolutional layer helps in understanding the pattern in the data. Therefore,
208 collectively, a ConvLSTM is useful in learning spatiotemporal data robustly (Gupta et al.,
209 2020; Sinha et al., 2021). The mathematical equations representing a ConvLSTM layer are
210 as follows (Shi et al., 2015):

$$i_t = g_1(w_{ix} * x_t + w_{ih} * h_{t-1} + w_{ic} \cdot c_{t-1} + b_i) \quad (11)$$

$$f_t = g_2(w_{fx} * x_t + w_{fh} * h_{t-1} + w_{fc} \cdot c_{t-1} + b_f) \quad (12)$$

$$o_t = g_3(w_{ox} * x_t + w_{oh} * h_{t-1} + w_{oc} \cdot c_t + b_o) \quad (13)$$

$$m_t = g_4(w_{mx} * x_t + w_{mh} * h_{t-1} + b_m) \quad (14)$$

$$c_t = f_t * c_{t-1} + i_t \cdot m_t \quad (15)$$

$$h_t = o_t \cdot \tanh(c_t) \quad (16)$$

216 Where t is the t^{th} step, g_i is the nonlinear activation function like sigmoid. $*$ indicates the
217 convolutional operation and \cdot denotes the element-wise multiplication. \tanh is an activation
218 function. $i_t, f_t, o_t,$ and m_t represent the input gate, forget gate, output gate, and modulation
219 gate. x_t is the input data to the ConvLSTM layer, and c_t and h_t are the cell and hidden
220 state, respectively.

221 The main principle of a Fourier layer is to decompose the signal of a time domain
222 into a frequency domain and to filter out the dominating frequency modes. The Fourier
223 decomposition involves representing the input signal into the sum of cosine and sine wave
224 components. The mathematical representation of a Fourier decomposition is given as follows:

$$f(x) = \sum_{i=1}^{\infty} 1/(\text{len}(x)/2)[a_i * \cos(i * 2\pi\omega x + \phi_i)] \quad (17)$$

225 The function $f(x)$ expresses the infinite linear combinations of sines and cosines of
226 different frequencies of input variable x , where a and ϕ determine the amplitude and phase
227 of the corresponding frequency (ω). The operation done in the Fourier layer is given below:

$$F' = FFT(X) * FFT(W) \quad (18)$$

228

$$F'' = IFFT(F') \quad (19)$$

229

230

231

232

233

234

235

236

where W and X are the randomly initialized weight matrix and the input X into the Fourier layer. FFT is the Fast-Fourier Transform and IFFT is the Inverse of FFT.

We employed a combination of a Fourier layer (F-layer), ConvLSTM layer, and convolutional layers. The Fourier layers convert the input into the frequency domain, and the weights are helpful in penalizing the dominant modes. Further, the ConvLSTM layers have a robust skill in predicting the sequential spatio-temporal data. We have also compared the skill of the LPS neural operator with the simple ConvLSTM without the Fourier layer.

237

3 Results and discussion

238

A Deep Learning Framework

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

The overall framework for predicting the MSLP anomaly is shown in Fig. 1. A sequential architecture uses both F-layer and ConvLSTM 2D layers as its first layer with five filters and ten filters, respectively, and Relu as an activation function in the ConvLSTM 2D layer. From recent studies, the ConvLSTM 2D is known for its efficiency in handling spatial-temporal data (Gupta et al., 2020). The input data into the model is a 5-dimensional tensor containing the length of the training data stack, input channels, latitude points, longitude points, and the stack of the input data for the past six days. The input data to the model is fed as stacked data, which means that the daily MSLP anomalies for the past six days are stacked and used to predict the next time step. The output from both the F-layer and the ConvLSTM 2D layer are concatenated and passed into two blocks of convolutional layer (Conv3D-1&2) having an activation of Relu, and five filters with a kernel size of 1×1 , and output is passed to a batch normalization layer (BatchNorm). Subsequently, the output from this step is passed to a Dropout (0.2), MaxPooling3D, Flattening layer, and two fully connected dense layers with 10 and 1 filters, respectively. Dropouts are added to the model wherever necessary during the parameter tuning to avoid overfitting the training data. The prominent features of the architecture are listed in Table 1. The model is optimized by tuning the hyper-parameters and the number of layers to obtain the best suitable combination of activation, number of filters, optimizers, dropouts, loss functions, etc. Satisfactory results were obtained with an epoch of 200 and a batch size of 160. Application of the MSE loss function yielded a model with good prediction capability. A total of 64 iterations are taken to learn the whole spatial map, satisfactorily. Therefore, in each iteration, the weights are not initialized randomly; rather, weights from the previous iteration were considered. The LPSNO is converged in the initial iteration of 200 epochs, as shown in Fig. 2. The shaded region shows the error bar (\pm standard deviation of loss function of all the 64 iterations from the initial iteration). The idea behind showing only the initial iteration of 200 epochs is the convergence of the model.

265

266

267

268

269

270

271

272

273

274

275

276

With the aid of the Fourier transforms, the MSLP anomaly is decomposed into a combination of sinusoidal waves, as shown in Fig. 3. The actual MSLP anomaly for the training period is shown in Fig. 3a, and the sinusoidal waves obtained from the Fourier decomposition are shown in Fig. 3b. Each sinusoidal wave has different amplitudes and phases; therefore, learning these high and low-frequency signals by an ML model helps in better prediction by considering the underlying weather modes. MSLP field can be reconstructed by combining the decomposed Fourier components (Fig. 3c). The power spectrum of daily MSLP anomaly for one JJAS season shows the maximum peak in the intraseasonal (30 - 60 day) period, and a secondary maximum in the synoptic and quasi-biweekly periods (Fig. 4). Therefore, the intraseasonal and synoptic scales are the two major components of the JJAS MSLP anomaly. The top nine sinusoidal components from Fig. 3b are shown in Fig. 5 for better visualizing the Fourier decomposition. The top

277 four panels in Fig. 5 show the Fourier components from intraseasonal oscillations, and the
 278 bottom four panels show the signals from the synoptic scale.

279 The Fourier layer introduced in the model architecture used in this study penalizes the
 280 important Fourier components by multiplying weights. Therefore, optimizing the Fourier
 281 weights helps the model learn the important weather modes. The Fourier layer primarily
 282 consists of three major layers: one is the Fourier transform of input time series, the second
 283 is the multiplication of weights to the Fourier transform, and the last one is the inverse
 284 transform into the time domain from the Fourier domain (Fig. 1b). The starting input to
 285 the model is a spatial map with latitude and longitude coordinates. However, the input
 286 map is fed into the model as an iterative 4×4 grid averaged time series out of a 64×64
 287 grid. The output from the model at each iteration of 200 epochs is compiled and depicted
 288 as a spatial structure again as a 16×16 grid size (1° resolution, Fig. 6).

289 The model predicts the daily MSLP anomalies at various lead times in a sequential
 290 fashion, i.e., the predicted one day lead is fed into the model to predict the day two, and so
 291 on. The LPSNO model was reasonably able to predict the MSLP anomaly spatial structure
 292 at a lead time of 3 days (Fig. 6). The right and left panels of Fig. 6 show the predictions of
 293 a low and high MSLP anomaly cases. The predictions are compared with the observations
 294 (Figs. 6a and e). The one day ahead prediction captured the spatial structure and magnitude
 295 of the MSLP anomalies reasonably well, for both negative and positive anomaly cases (Figs.
 296 6b and f). When the lead times are increased to two and three days, the quality of predictions
 297 weakened (Fig. 6c-h). Nevertheless, the overall structure of both the low and high pressure
 298 anomalies are predicted by the model at increased lead times.

299 The pattern correlation between the observed and the predicted spatial map of MSLP
 300 anomaly is shown in Fig. 7. The pattern correlation is defined as the Pearson product-
 301 moment coefficient of linear correlation between the two variables of the same dimensions.
 302 The prediction at the lead time of one day has a median pattern correlation of about 87%.
 303 Similarly, the median pattern correlation of lead two and lead three predictions is about
 304 $\sim 60\%$ and $\sim 50\%$, respectively. The correlation is weakening as time progresses; much lower
 305 values are observed in leads four and five (Fig. 7). An accurate prediction of the magnitudes
 306 and spatial pattern of MSLP anomalies is necessary to identify the intensity category and
 307 trajectory of LPSs, by reconstructing the full MSLP field using Eq. 6.16. The same strategy
 308 can be used to predict tropical cyclones as well. However, here, we focus only on LPSs.

$$MSLP = (MSLP)' + \overline{MSLP} \quad (20)$$

309 where $(MSLP)'$ is the anomaly of the MSLP and the \overline{MSLP} is the climatology (long-term
 310 mean).

311 Recent advancement of deep learning in this research area shows the ConvLSTM
 312 model's efficacy in handling spatial-temporal data (Gupta et al., 2020). Therefore, the
 313 current model of the Fourier layer variant is compared with the ConvLSTM model to see
 314 the prediction. Only the comparison of results from two iterations is shown in Table 2. The
 315 architecture of the LPS neural operator is the same as discussed above, with a Fourier layer
 316 concatenated with the ConvLSTM layer. Whereas in the case of the ConvLSTM model,
 317 there is no Fourier layer branch as shown in Fig. 1 and the rest of all architecture is the
 318 same; therefore, in ConvLSTM, the concatenation layer is also removed. The ConvLSTM
 319 model shows a correlation of around 0.77 in both the iterations between the observed and
 320 the predicted at lead 1. Whereas the LPSNO model shows a significant improvement in
 321 the prediction with a correlation of 0.84 in both iterations. The superiority of the proposed
 322 LPS neural operator over the simple ConvLSTM is seen.

323 The LPSs are tracked from the reconstructed daily MSLP field for 10 JJAS seasons.
 324 The lead 1 prediction captured 50 LPSs while 51 LPSs are observed during the same period.
 325 Although the lead time for the prediction is short, the accuracy of the model is remarkable.

326 The statistics for the different lead time predictions is shown in Table 3. The model’s
 327 ability to capture LPSs at higher leads diminishes gradually. One reason for this might be
 328 the deterioration of the skill in predicting the magnitude of the MSLP anomalies at greater
 329 lead times.

330 SAI is very useful for understanding the spatial distribution of LPS trajectories and
 331 their strength. The SAI for the observed period shows a maximum density over the head
 332 BoB, which is the core genesis region of the LPSs (Fig. 8a). The prediction at Lead 1 to
 333 5 days also shows the maxima over the head BoB, though with diminishing intensity with
 334 lead time of the prediction (Fig. 8b–f). At lead one, the model is also able to predict the
 335 propagation of LPSs in the northwest direction towards the continental India. The lead 5
 336 prediction captures the weakest synoptic activity, in line with the number of LPSs. The
 337 pattern of observed and predicted SAI is compared using the pattern correlation. The
 338 pattern correlation between the SAI of actual and the predictions at different lead times is
 339 also shown in Fig. 9. The pattern correlation between lead one predictions and observations
 340 shows the highest value of 0.94. Pattern correlations between observed and predicted SAI at
 341 lead 2 to 5 are 0.9, 0.87, 0.82, and 0.84, respectively (Fig. 9a). The pattern correlation alone
 342 is not a good measure of the skill of the model. The Root Mean Squared Error (RMSE)
 343 score between the predicted and observed SAI has a minimum value for lead time 1 and
 344 increases at subsequent lead times with the maximum RMSE score for lead 5 (Fig. 9b).
 345 When the pattern correlation and RMSE are taken together, the model skill in predicting
 346 the synoptic activity is quite low beyond a lead time of three days. The current LPS neural
 347 operator takes ~ 3.2 s to generate one day ahead prediction, which is significantly efficient in
 348 terms of computational resources required for a prediction using an NWP model. Further,
 349 the MSLP anomaly is predicted as a continuous variable in time, as discussed above, which
 350 makes it useful in operation weather forecast.

351 4 Conclusions

352 In this study, comprehensive deep learning framework to predict the spatial structure
 353 of the daily mean sea level pressure (MSLP) anomalies is proposed. Subsequently, synoptic-
 354 scale tropical storms known as “monsoon low pressure systems (LPS)” that contribute about
 355 60% of monsoon rainfall over the hugely populated Indo-Gangetic plains are tracked from
 356 the MSLP anomalies. To this extent, a start-of-the-art neural operator model comprised
 357 of a combination of Fourier and Convolutional Long Short Term Memory (LPS Neural
 358 Operator) is employed to predict the spatial MSLP anomaly map. A sequential prediction
 359 of MSLP anomalies is made using the prediction from the previous time step, similar to
 360 the conventional numerical weather prediction models. Median pattern correlations of 88%
 361 , 65% , and 50% , respectively, between the observed and predicted MSLP anomalies over
 362 the Bay of Bengal are obtained. Daily MSLP field is reconstructed by using the predicted
 363 anomalies and climatology. This MSLP field is used to track the LPSs over the BoB. The
 364 one day lead prediction captured almost the same number of LPSs as observed in a ten
 365 year period. At longer lead times, as expected, the model’s skill in capturing the LPSs
 366 diminished.

367 In the recent years, deep learning models are creating a revolution in the field of weather
 368 forecasting, with the models attaining the skill of the best operational numerical weather
 369 prediction models in the short and medium range forecasts Lam et al. (2023); Bi et al.
 370 (2023). However, the deep learning models are yet to prove their skill in capturing extreme
 371 weather phenomena such as tropical cyclones and LPSs. Here, we showed that a combination
 372 of Fourier and Convolutional Long Short Term Memory model is capable of accurately
 373 predicting the genesis of monsoon LPSs at one day lead time over a span of ten seasons.
 374 The predictions at lead times of up to three days are found to be reasonably well. Further
 375 improvements to this model will make it suitable for operational prediction of LPSs over
 376 the Indian region.

Acknowledgments

K. S. S. Sai Srujan is supported by the Prime Minister’s Research Fellowship by the Ministry of Education, Government of India. HK and SS acknowledge the funding support by the Ministry of Earth Sciences, Government of India through the project titled ”Prediction of Indian Summer Monsoon Active and Break Cycles using Deep Learning” (MoES/16/09/2022-RDESS).

Open Research

Data availability statement All data used in this study is freely available from public data repositories. The authors thank the developers of Matplotlib (Hunter, 2007) for making their code available on a free and open-source basis, which is used to generate all the figures. The codes and predictions of LPSNO (Srujan et al., 2024) can be accessed from <https://zenodo.org/doi/10.5281/zenodo.10499398>.

References

- Ajayamohan, R., Merryfield, W. J., & Kharin, V. V. (2010). Increasing trend of synoptic activity and its relationship with extreme rain events over central india. *Journal of Climate*, *23*(4), 1004–1013.
- Andersson, T. R., Hosking, J. S., Pérez-Ortiz, M., Paige, B., Elliott, A., Russell, C., ... others (2021). Seasonal arctic sea ice forecasting with probabilistic deep learning. *Nature Communications*, *12*(1), 5124.
- Azzizadenesheli, K., Kovachki, N., Li, Z., Liu-Schiaffini, M., Kossaiji, J., & Anandkumar, A. (2023). Neural operators for accelerating scientific simulations and design. *arXiv preprint arXiv:2309.15325*.
- Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023). Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, *619*(7970), 533–538.
- Chen, Y., Haywood, J., Wang, Y., Malavelle, F., Jordan, G., Partridge, D., ... others (2022). Machine learning reveals climate forcing from aerosols is dominated by increased cloud cover. *Nature Geoscience*, *15*(8), 609–614.
- Courant, R., Friedrichs, K., & Lewy, H. (1967). On the partial difference equations of mathematical physics. *IBM Journal of Research and Development*, *11*(2), 215–234. doi: 10.1147/rd.112.0215
- Deoras, A., Hunt, K., & Turner, A. (2021). The four varieties of south asian monsoon low-pressure systems and their modulation by tropical intraseasonal variability. *Weather*, *76*(6), 194–200.
- Goswami, B. N., Wu, G., & Yasunari, T. (2006). The annual cycle, intraseasonal oscillations, and roadblock to seasonal predictability of the asian summer monsoon. *Journal of Climate*, *19*(20), 5078 - 5099. doi: 10.1175/JCLI3901.1
- Greenfeld, D., Galun, M., Basri, R., Yavneh, I., & Kimmel, R. (2019). Learning to optimize multigrid pde solvers. In *International conference on machine learning* (pp. 2415–2423).
- Gupta, M., Kodamana, H., & Sandeep, S. (2020). Prediction of enso beyond spring predictability barrier using deep convolutional lstm networks. *IEEE Geoscience and Remote Sensing Letters*, 1–5. doi: 10.1109/LGRS.2020.3032353
- Ham, Y., Kim, J., & Luo, J. (2019). Deep learning for multi-year enso forecasts. *Nature*, *573*, 568–572. doi: 10.1038/s41586-019-1559-7
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... others (2020). The era5 global reanalysis. *Q J R Meteorol Soc*, *146*(730), 1999–2049.
- Hunt, K. M., & Fletcher, J. K. (2019). The relationship between indian monsoon rainfall and low-pressure systems. *Climate Dynamics*, *53*(3), 1859–1871.
- Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. doi: 10.1109/MCSE.2007.55

- 427 Hurley, J. V., & Boos, W. R. (2015). A global climatology of monsoon low-pressure systems.
428 *Q J R Meteorol Soc*, *141*(689), 1049-1064. doi: 10.1002/qj.2447
- 429 Kochkov, D., Smith, J. A., Alieva, A., Wang, Q., Brenner, M. P., & Hoyer, S. (2021). Ma-
430 chine learning–accelerated computational fluid dynamics. *Proceedings of the National*
431 *Academy of Sciences*, *118*(21), e2101784118.
- 432 Kossaifi, J., Kovachki, N. B., Azzadenesheli, K., & Anandkumar, A. (n.d.). Multi-
433 grid tensorized fourier neural operator for high resolution pdes, 2023. In *Url*
434 *https://openreview.net/forum*.
- 435 Krishnamurthy, V., & Ajayamohan, R. S. (2010). Composite structure of monsoon low
436 pressure systems and its relation to Indian rainfall. *Journal of Climate*, *23*(16), 4285-
437 4305. doi: 10.1175/2010JCLI2953.1
- 438 Kurth, T., Subramanian, S., Harrington, P., Pathak, J., Mardani, M., Hall, D., . . . Anand-
439 kumar, A. (2023). Fourcastnet: Accelerating global high-resolution weather fore-
440 casting using adaptive fourier neural operators. In *Proceedings of the Platform for*
441 *Advanced Scientific Computing Conference* (pp. 1–11).
- 442 Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., . . .
443 others (2023). Learning skillful medium-range global weather forecasting. *Science*,
444 eadi2336.
- 445 Li, Z., Kovachki, N., Azzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., & Anand-
446 kumar, A. (2020). Fourier neural operator for parametric partial differential equations.
447 *arXiv preprint arXiv:2010.08895*.
- 448 Ling, F., Luo, J.-J., Li, Y., Tang, T., Bai, L., Ouyang, W., & Yamagata, T. (2022). Multi-
449 task machine learning improves multi-seasonal prediction of the indian ocean dipole.
450 *Nature Communications*, *13*(1), 7681.
- 451 Lu, L., Jin, P., Pang, G., Zhang, Z., & Karniadakis, G. E. (2021). Learning nonlinear
452 operators via deepnet based on the universal approximation theorem of operators.
453 *Nature machine intelligence*, *3*(3), 218–229.
- 454 Mooley, D. A., & Shukla, J. (1987). *Characteristics of the westward-moving summer mon-*
455 *soon low pressure systems over the indian region and their relationship with the mon-*
456 *soon rainfall* (Tech. Rep.). Center for Ocean–Land–Atmosphere Interactions, 47 pp.
- 457 Praveen, V., Sandeep, S., & Ajayamohan, R. S. (2015). On the relationship between mean
458 monsoon precipitation and low pressure systems in climate model simulations. *Journal*
459 *of Climate*, *28*(13), 5305-5324. doi: 10.1175/JCLI-D-14-00415.1
- 460 Rojas-Campos, A., Langguth, M., Wittenbrink, M., & Pipa, G. (2023). Deep learning
461 models for generation of precipitation maps based on numerical weather prediction.
462 *Geoscientific Model Development*, *16*(5), 1467–1480.
- 463 Saha, S. K., Hazra, A., Pokhrel, S., Chaudhari, H. S., Sujith, K., Rai, A., . . . Goswami, B. N.
464 (2019). Unraveling the mystery of indian summer monsoon prediction: Improved es-
465 timate of predictability limit. *Journal of Geophysical Research: Atmospheres*, *124*(4),
466 1962-1974. doi: 10.1029/2018JD030082
- 467 Sandeep, S., Ajayamohan, R. S., Boos, W. R., Sabin, T. P., & Praveen, V. (2018). Decline
468 and poleward shift in indian summer monsoon synoptic activity in a warming climate.
469 *Proc. Natl. Acad. Sci. USA*, *115*(11), 2681–2686. doi: 10.1073/pnas.1709031115
- 470 Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-k., & WOO, W.-c. (2015).
471 Convolutional lstm network: A machine learning approach for precipitation now-
472 casting. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett (Eds.),
473 *Advances in neural information processing systems* (Vol. 28). Curran Associates,
474 Inc. Retrieved from [https://proceedings.neurips.cc/paper_files/paper/2015/
475 file/07563a3fe3bbe7e3ba84431ad9d055af-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2015/file/07563a3fe3bbe7e3ba84431ad9d055af-Paper.pdf)
- 476 Sikka. (2006). *A study on the monsoon low pressure systems over the indian region and their*
477 *relationship with drought and excess monsoon seasonal rainfall* (Tech. Rep.). Center
478 for Ocean–Land–Atmosphere Studies, Center for the Application of Research on the
479 Environment, 61 pp.
- 480 Sikka, D. R. (1977). Some aspects of the life history, structure and movement of monsoon
481 depressions. *PAGEOPH*, *115*, 1501–1529. doi: 10.1007/BF00874421

- 482 Sinha, A., Gupta, M., Srujan, K. S. S. S., Kodamana, H., & Sandeep, S. (2021). Prediction
 483 of synoptic-scale sea level pressure over the Indian monsoon region using deep learning.
 484 *IEEE Geoscience and Remote Sensing Letters*, 1-5. doi: 10.1109/LGRS.2021.3100899
- 485 Srujan, K. S. S. S., Sandeep, S., & Kodamana, H. (2024, January). *The Framework of*
 486 *Indian Summer monsoon Low-Pressure Systems Neural Operator (LPSNO)*. Zenodo.
 487 Retrieved from <https://doi.org/10.5281/zenodo.10499399> doi: 10.5281/zenodo
 488 .10499399
- 489 Thomas, T. M., Bala, G., & Srinivas, V. V. (2021). Characteristics of the monsoon low
 490 pressure systems in the indian subcontinent and the associated extreme precipitation
 491 events. *Climate Dynamics*, 56(5), 1859–1878.
- 492 Wang, B., Ding, Q., Fu, X., Kang, I.-S., Jin, K., Shukla, J., & Doblas-Reyes, F. (2005). Fun-
 493 damental challenge in simulation and prediction of summer monsoon rainfall. *Geophys-
 494 ical Research Letters*, 32(15), L15711. doi: <https://doi.org/10.1029/2005GL022734>
- 495 Wang, B., Xiang, B., Li, J., Webster, P. J., Rajeevan, M. N., Liu, J., & Ha, K. J. (2015).
 496 Rethinking indian monsoon rainfall prediction in the context of recent global warming.
 497 *Nature Communications*, 6, 7154. doi: 10.1038/ncomms8154
- 498 Wang, C., Zheng, G., Li, X., Xu, Q., Liu, B., & Zhang, J. (2022). Tropical cyclone inten-
 499 sity estimation from geostationary satellite imagery using deep convolutional neural
 500 networks. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1-16. doi:
 501 10.1109/TGRS.2021.3066299
- 502 Wen, G., Li, Z., Long, Q., Azzadenesheli, K., Anandkumar, A., & Benson, S. M. (2023).
 503 Real-time high-resolution co 2 geological storage prediction using nested fourier neural
 504 operators. *Energy & Environmental Science*, 16(4), 1732–1741.
- 505 Weyn, J. A., Durran, D. R., Caruana, R., & Cresswell-Clay, N. (2021). Sub-seasonal
 506 forecasting with a large ensemble of deep-learning weather prediction models. *Journal*
 507 *of Advances in Modeling Earth Systems*, 13(7), e2021MS002502.
- 508 Zhang, Q., Wang, H., Dong, J., Zhong, G., & Sun, X. (2017). Prediction of sea surface
 509 temperature using long short-term memory. *IEEE Geoscience and Remote Sensing*
 510 *Letters*, 14(10), 1745–1749.
- 511 Zhang, R., Liu, Q., & Hang, R. (2020). Tropical cyclone intensity estimation using two-
 512 branch convolutional neural network from infrared and water vapor images. *IEEE*
 513 *Transactions on Geoscience and Remote Sensing*, 58(1), 586-597. doi: 10.1109/TGRS
 514 .2019.2938204
- 515 Zhang, Y., Long, M., Chen, K., Xing, L., Jin, R., Jordan, M. I., & Wang, J. (2023). Skilful
 516 nowcasting of extreme precipitation with nowcastnet. *Nature*, 619(7970), 526–532.

Table 1. LPS NEURAL OPERATOR ARCHITECTURE DETAILS

<i>Layer (type)</i>	<i>Activation</i>	<i># Filters</i>	<i>Kernel size</i>	<i>Dropout</i>	<i>Bias</i>	<i>Pool size</i>
Fourier Layer	-	5	1x1	-	-	-
ConvLSTM2D	Relu	10	3x3	-	-	-
Conv3D-1	Relu	5	1x1	-	True	-
Conv3D-2	Relu	5	1x1	0.2	True	-
MaxPooling3D	-	-	-	0.2	-	1x1x1
Dense-1	Relu	10	-	-	True	-
Dense-2	Linear	1	-	-	True	-

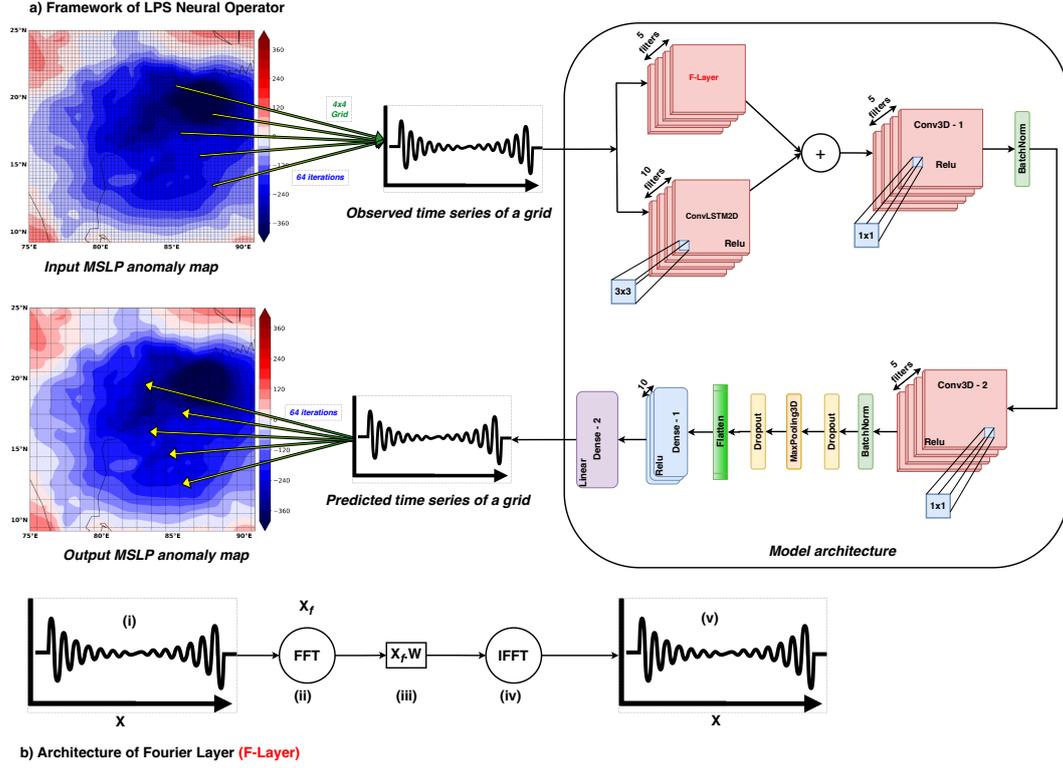


Figure 1. (a) Deep learning architecture used to predict the sea level pressure anomaly over the Bay of Bengal (details of the architecture are explained in Table 1). The plus symbol indicates the concatenation of two layers. The F-layer is the Fourier layer introduced in the deep learning model. The architecture of the Fourier Layer is shown in the bottom panel (b). The Roman numbers in (b) are explained as follows: (i) is the input time series, (ii) is the Fourier transform of (i, i.e., X_f), (iii) weight (W) multiplied to the X_f , (iv) is the inverse Fourier transform of the (iii), and (v) is the time series obtained from (iv).

Table 2. Comparison between ConvLSTM and F-layer ConvLSTM (LPSNO) at the lead time of 24 hrs

<i>Model</i>		<i>#Filters</i>	<i>Correlation</i>
LPSNO	iteration 1	5	0.83
	iteration 2	5	0.84
ConvLSTM	iteration 1	5	0.77
	iteration 2	5	0.76

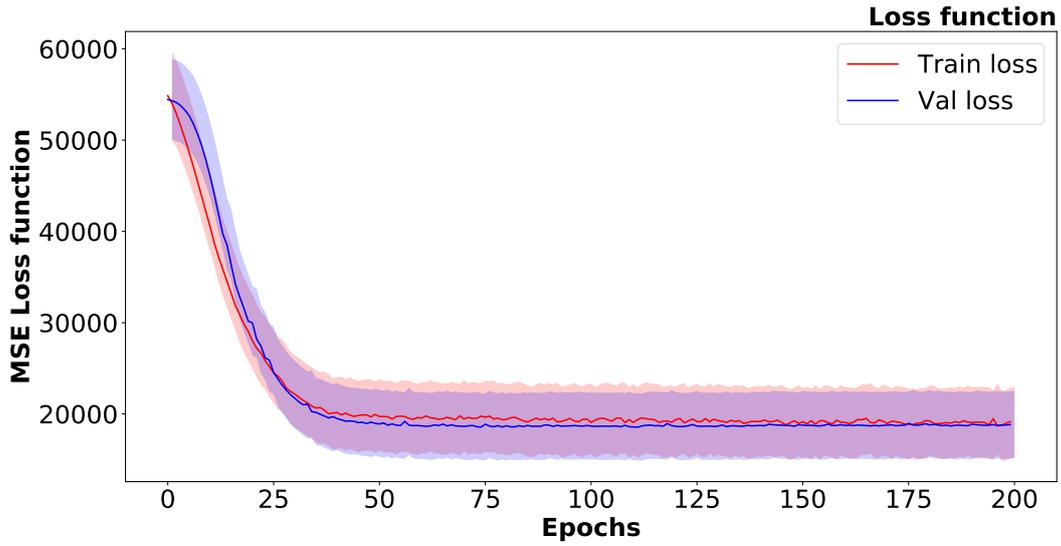


Figure 2. The loss function of the LPSNO model trained for the first step for 200 epochs is shown. Shading indicates the overall spread of the loss function computed as the total standard deviation in all steps. The red and blue indicate the training and validation curve, respectively.

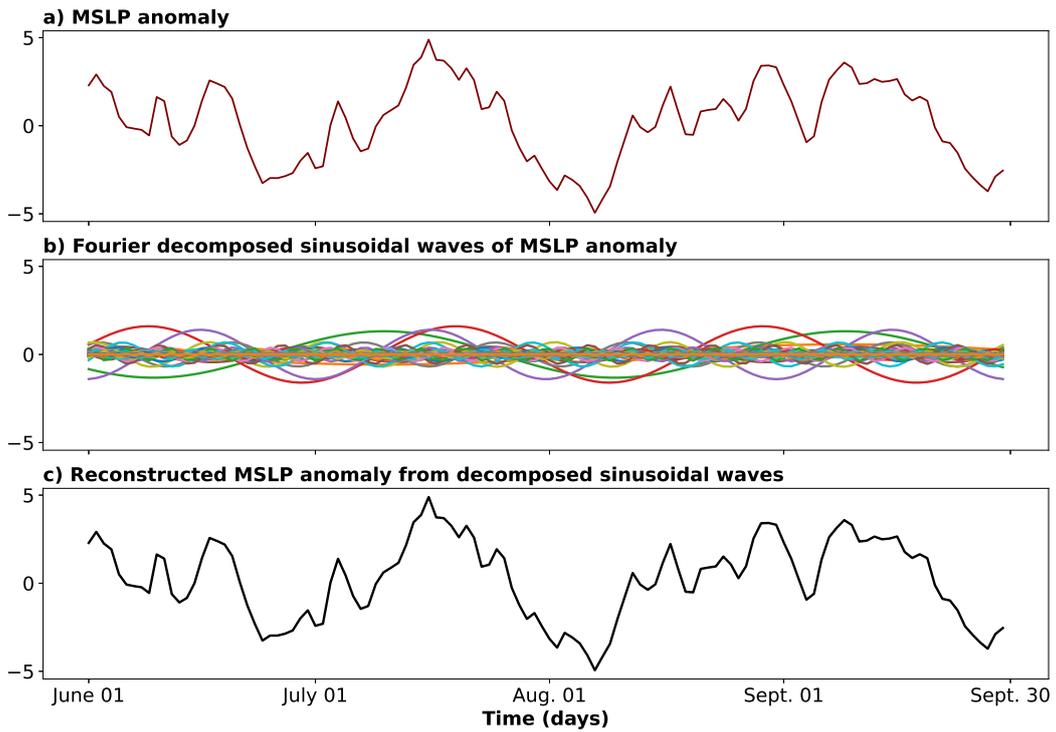


Figure 3. (a) Time series of MSLP anomaly area averaged over $75^{\circ}\text{E} - 90^{\circ}\text{E}$, $10^{\circ}\text{N} - 25^{\circ}\text{N}$ from 1979 – 2014 (period of training the model). (b) Fourier decomposed sinusoidal waves of MSLP anomaly from (a). (c) The reconstructed time series of MSLP anomaly using the decomposed Fourier components shown in (b)

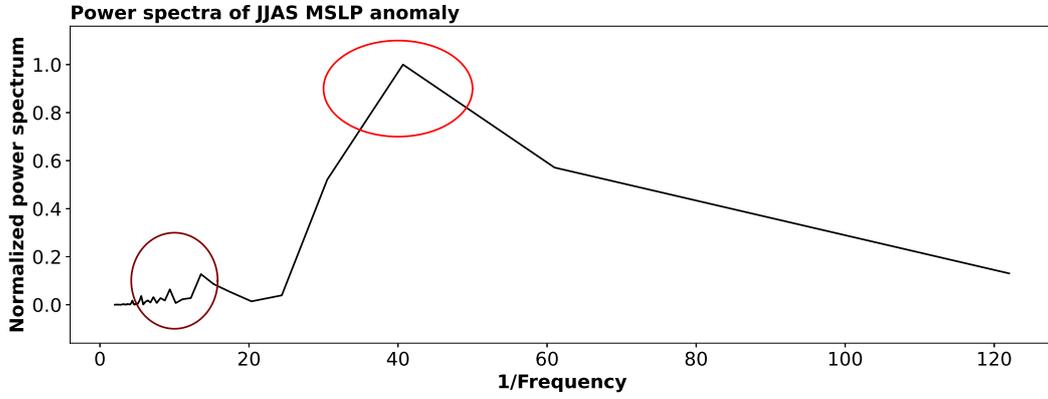


Figure 4. Normalized power spectrum MSLP anomaly for JJAS season of 1979. Red circle indicates the 1st maximum (represents the intraseasonal signal) and the maroon circle indicates the 2nd maxima (represents the synoptic signal) of normalized power spectrum

Table 3. Statistics of the actual and predicted MSLP anomaly at different lead times

<i>Type (actual/predicted)</i>	<i>Number of LPSs</i>
Actual	51
Lead 1	50
Lead 2	35
Lead 3	23
Lead 4	25
Lead 5	15

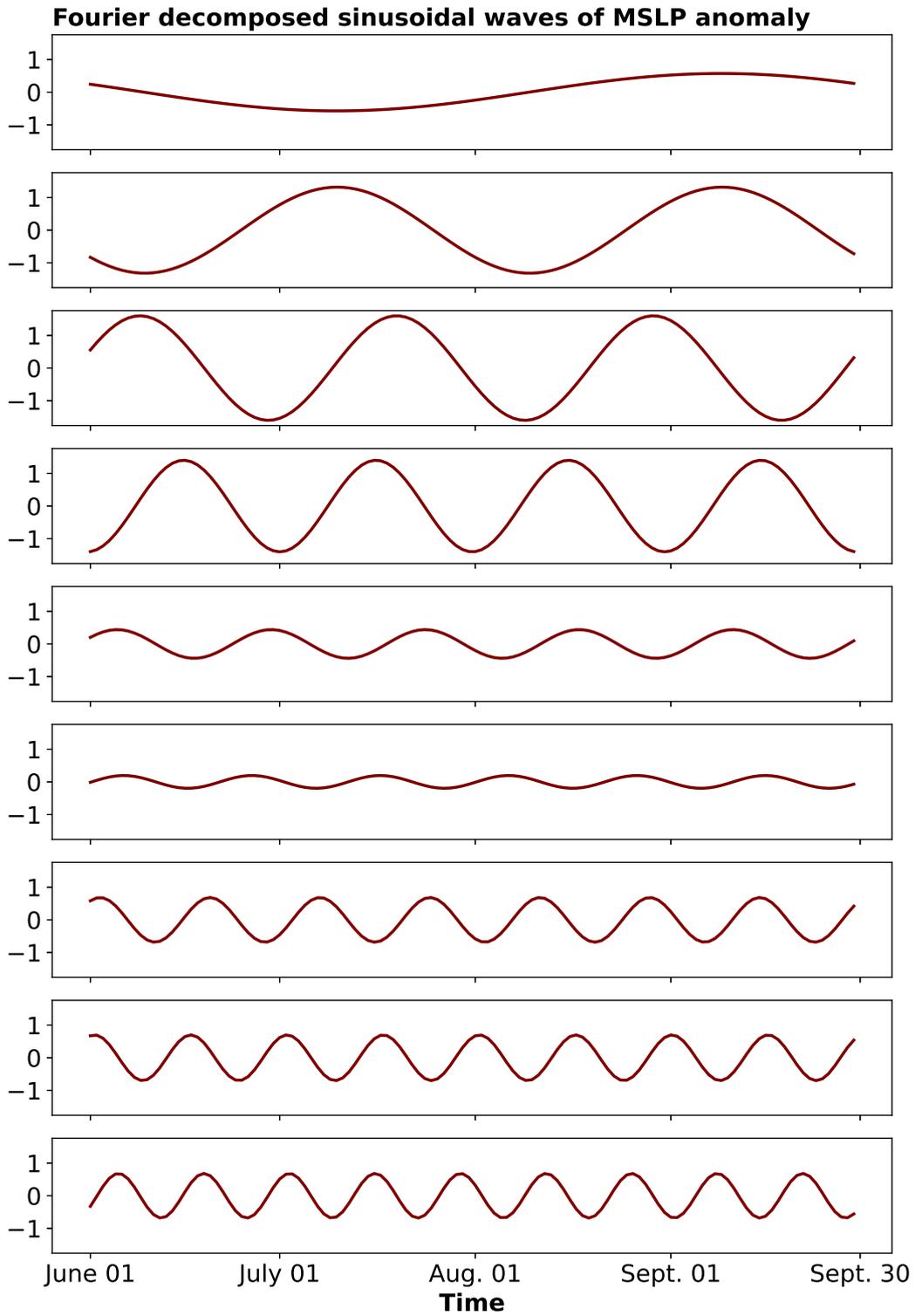


Figure 5. Top Nine sinusoidal wave components from the Fourier decomposition of MSLP anomaly as shown in Fig. 3. The y-axis shows the amplitude of the sinusoidal wave components, and the x-axis is the time (in days) of MSLP anomaly considered in this study, as mentioned in the data section.

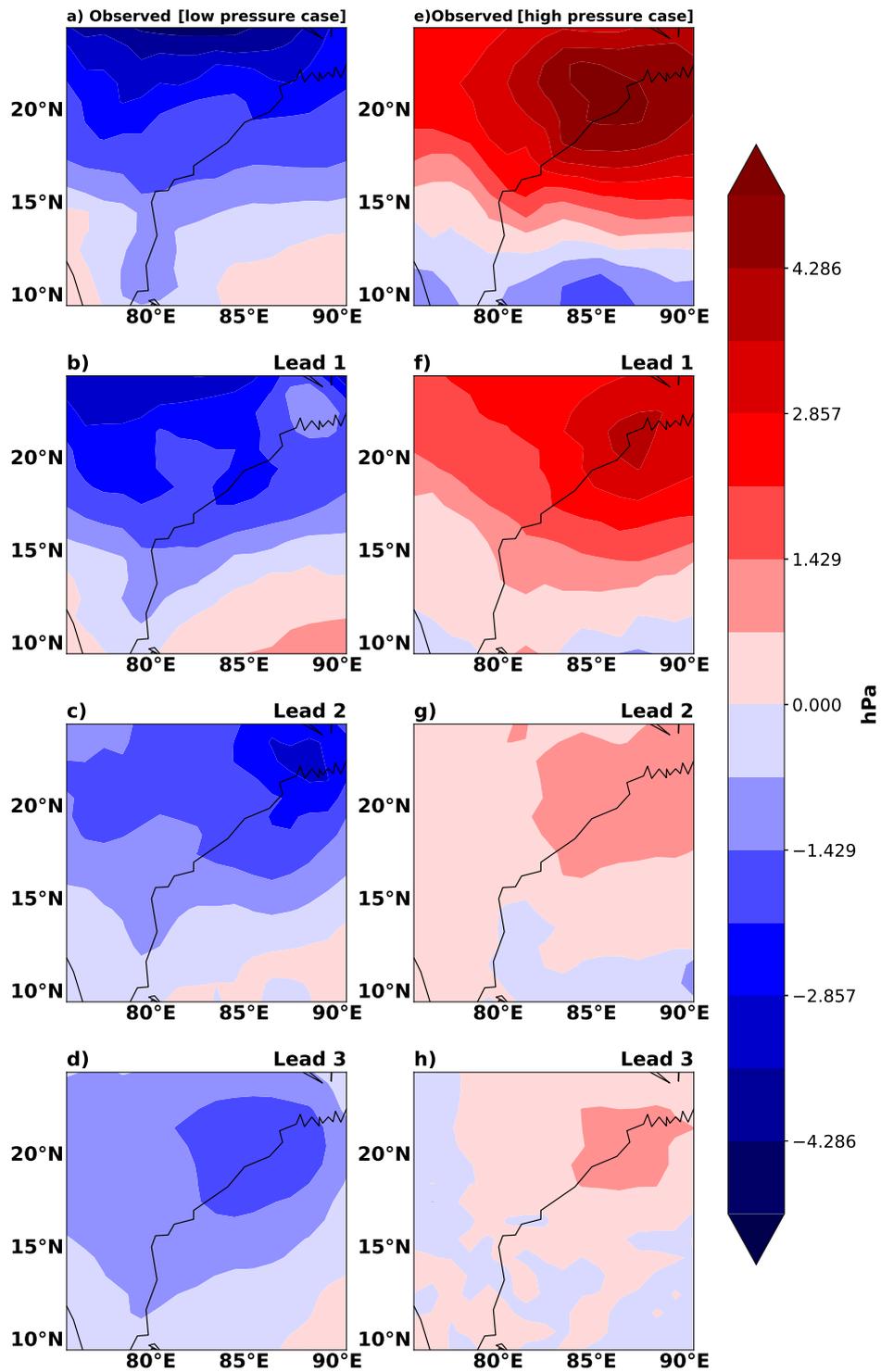


Figure 6. The actual (a,e) and predicted (b–d, f–h) spatial structure of sea-level pressure anomaly (units: hPa) over the Bay of Bengal at different lead times (24, 48, 72 hrs). The right and left panels are two different time steps.

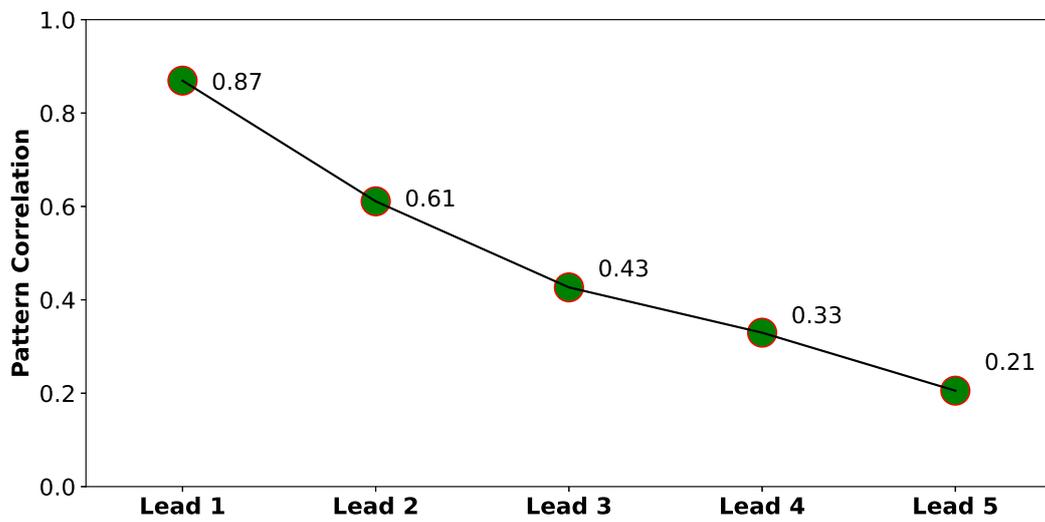


Figure 7. The median of pattern correlation between the observed and predicted sea level pressure anomaly at a lead time of 24, 48, 72, 96, and 120 hrs (Lead 1, Lead 2, Lead 3, Lead 4, and Lead 5 respectively) from 2008 – 2018.

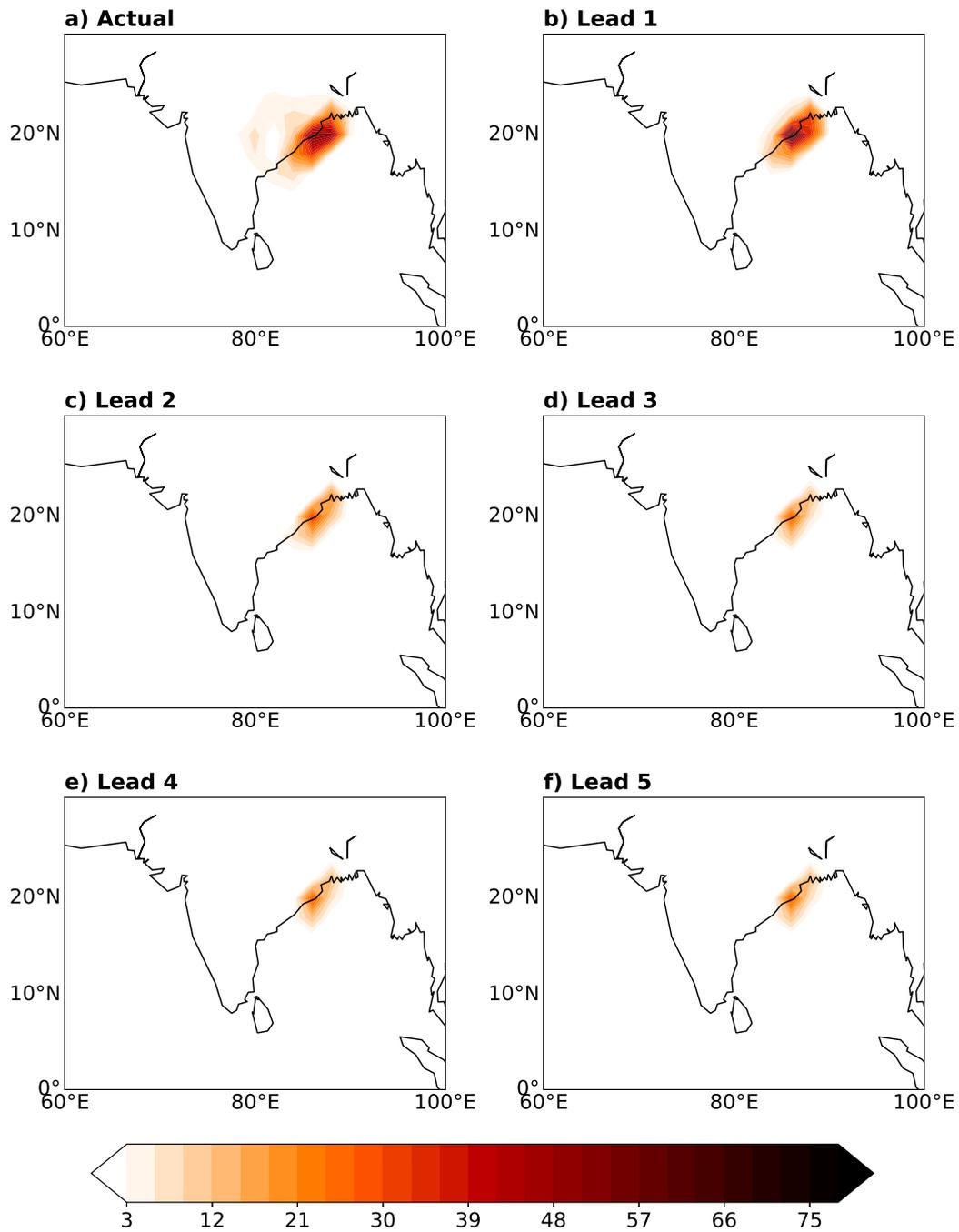


Figure 8. The synoptic activity index(SAI) computed for the life cycle of the LPSs for a) Actual and (b-f) predicted show the SAI at the lead times one to five days from 2008 – 2018.

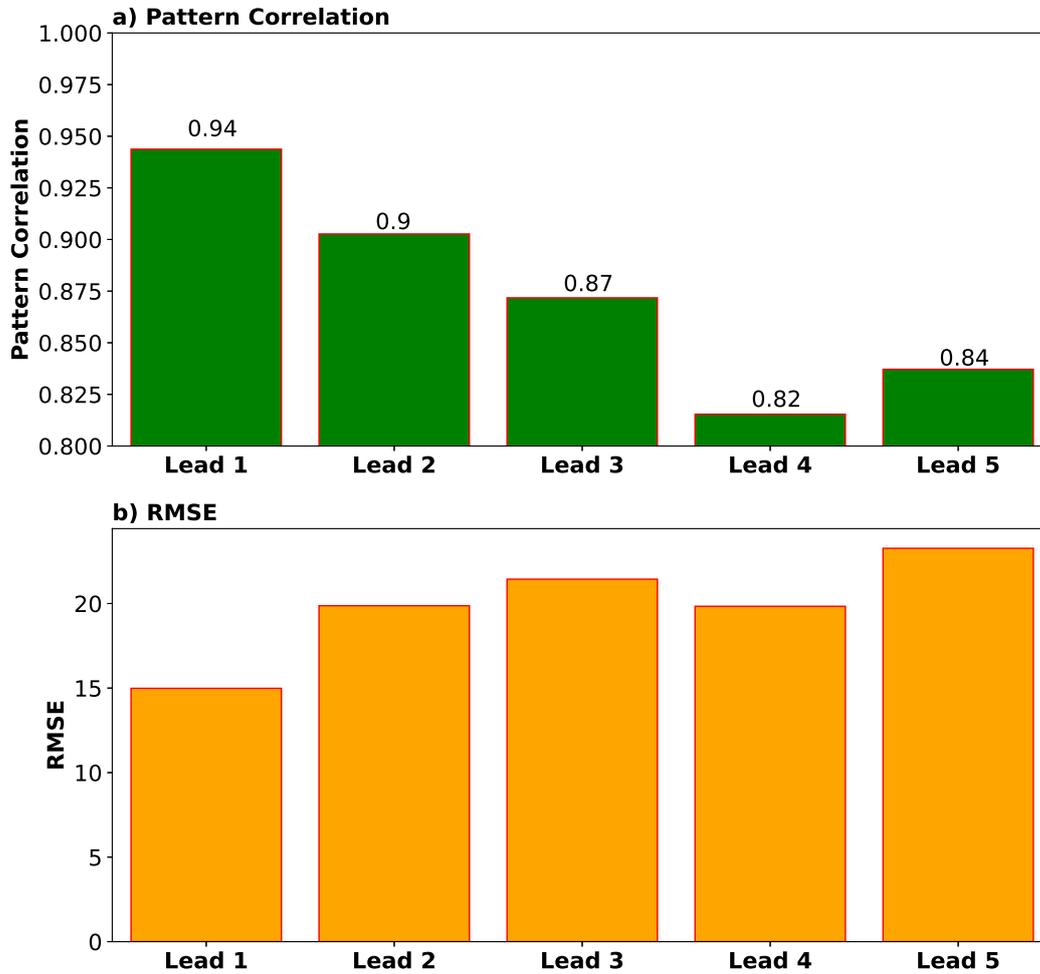


Figure 9. (a) The pattern correlation and (b) Root Mean Square Error (RMSE) score between the observed and predicted synoptic index computed using the life cycle of LPSs (from genesis to lysis) at a lead time of 24, 48, 72, 96, and 120 hrs (Lead 1, Lead 2, Lead 3, Lead 4, and Lead 5 respectively) from 2008 – 2018. (units of RMSE are the same as SAI)

1 **LPS Neural Operator (LPSNO): A Novel Deep**
2 **Learning Framework to Predict the Indian Monsoon**
3 **Low-Pressure Systems**

4 **K. S. S. Sai Srujan¹, S. Sandeep^{1,2}, and Hariprasad Kodamana^{2,3}**

5 ¹Centre for Atmospheric Sciences, Indian Institute of Technology Delhi, New Delhi 110016, India

6 ²Yardi School of Artificial Intelligence, Indian Institute of Technology Delhi, New Delhi 110016, India

7 ³Department of Chemical Engineering, Indian Institute of Technology Delhi, New Delhi 110016, India

8 **Key Points:**

- 9 • A framework to predict the sea-level pressure anomaly over the Bay of Bengal is pro-
10 posed using the LPS Neural Operator for the first time
- 11 • The well-trained LPS Neural operator takes only a few seconds to generate a one-day
12 forecast over the Indian monsoon domain
- 13 • The pattern correlation between predicted and actual synoptic activity is 94%, 90%,
14 and 87% for 24, 48, and 72-hour forecasts
- 15
- 16
- 17

Corresponding author: S. Sandeep, sandeep.sukumaran@cas.iitd.ac.in

Corresponding author: Hariprasad Kodamana, kodamana@chemical.iitd.ac.in

Abstract

[The synoptic scale variability of the Indian summer monsoon (ISM) is contributed by the weak cyclonic vortices known as low-pressure systems (LPSs). LPSs are the primary mechanism by which central Indian plains receive rainfall. Traditionally, synoptic variability is considered to have a low predictability. In the present study, we developed a framework, namely, LPS Neural Operator (LPSNO), using the neural operator-based deep learning to predict the spatial structure of daily mean sea level pressure anomalies over the Bay of Bengal at a resolution of $1^\circ \times 1^\circ$. The proposed neural operator extends the Fourier neural operator framework by employing convolutional LSTMs in the operator backbone. Further, the mean sea level pressure is reconstructed using the predicted anomaly and the climatology, which is then used to track the LPSs using a Lagrangian tracking algorithm. The median pattern correlation between the predicted and actual mean sea-level pressure anomalies over the BoB is about 88 %, 60 %, and 50 % for 24, 48, and 72-hour forecasts, respectively. The proposed model improves the accuracy of predictions compared with the earlier ConvLSTM models. The pattern correlation between the observed and predicted synoptic activity index (SAI) is 0.94, 0.9, and 0.87 for 1, 2, and 3-day ahead predictions, respectively. A well-trained model of LPSNO takes only ~ 3.2 s to generate a one-day forecast on a single GPU node of Nvidia V100, which is computationally extremely cheap compared to the conventional numerical weather prediction models. The proposed LPSNO can advance operational weather forecasting substantially.]

Plain Language Summary

[The weak cyclonic vortices during the Indian Summer Monsoon (ISM) season, commonly known as Low-Pressure Systems (LPSs), are predominantly present over the head Bay of Bengal (BoB). More than half of ISM rainfall over the hugely populated Indo-Gangetic plains is contributed from LPSs, making them an important component of the hydrological cycle over South Asia with huge socio-economic impacts. Therefore, the prediction for LPS genesis will be helpful in better disaster preparedness and food security planning. A machine learning (ML) framework is developed initially to predict the spatial map of perturbations in mean sea level pressure (MSLP). Using the predicted perturbations and climatology, the full MSLP field is reconstructed. The LPSs are tracked from the MSLP field. The correlation of the spatial map of fluctuations in MSLP between actual and predicted is about 88%, 60%, and 50% at a lead time of 24, 48, and 72-hours. A well-trained ML model will be computationally efficient compared to traditional numerical weather prediction models.]

1 Introduction

The Indian summer monsoon (ISM) is an important component of the hydrological cycle of South Asia, which is essential to the water security of more than 1.5 billion inhabitants. Relatively weaker synoptic-scale cyclonic vortices embedded in the ISM circulation are known as the low-pressure systems (LPSs). Although the LPSs form in all monsoon regions around the world, they are most prominent in the ISM domain, with about 12 systems forming in each June - September period (Hurley & Boos, 2015). These are the main rain bearing systems, with a life span of 3–7 days and a diameter of 1000 – 2000 km (D. R. Sikka, 1977) contributing nearly half of the ISM rainfall over the Indo-Gangetic plains (Krishnamurthy & Ajayamohan, 2010; Praveen et al., 2015; Hunt & Fletcher, 2019; Sandeep et al., 2018; Thomas et al., 2021; Deoras et al., 2021). Conventionally, the propagation of LPSs have been identified manually using the surface pressure charts for which a long term archive exists (Mooley & Shukla, 1987; Sikka, 2006; Krishnamurthy & Ajayamohan, 2010). The dynamical and statistical models face difficulties in predicting the synoptic-scale rainfall (B. Wang et al., 2005, 2015; Saha et al., 2019). The chaotic nature of ISM makes the prediction of synoptic scale variability challenging (Goswami et al., 2006; Saha et al., 2019).

67 The conventional NWP models solve the prognostic partial differential equations using
 68 numerical methods, such as the finite difference method, which discretize time and space.
 69 Therefore, there is always a trade off between the grid resolution and the computational
 70 time. The computational stability of NWP models is measured by a condition called as CFL
 71 condition (Courant et al., 1967). However, data-driven deep learning models are not limited
 72 by these issues (Greenfeld et al., 2019; Kochkov et al., 2021; Li et al., 2020). For example,
 73 Convolutional Neural Networks (CNNs) in predicting the ENSO (Ham et al., 2019; Gupta
 74 et al., 2020) and estimating the intensity of the tropical cyclone (C. Wang et al., 2022;
 75 R. Zhang et al., 2020). The CNNs are known for the application of spatial data prediction.
 76 Also, the convolutions with the long short-term memory (ConvLSTM) show robust skills in
 77 nowcasting the precipitation (Shi et al., 2015) and predicting sea level pressure time series
 78 (Sinha et al., 2021).

79 The recent rapid advancement of data driven deep learning (DL) models suggest that
 80 they can be useful in the prediction of atmospheric and oceanic states (Q. Zhang et al.,
 81 2017; Ham et al., 2019; Gupta et al., 2020; Sinha et al., 2021; Andersson et al., 2021; Chen
 82 et al., 2022; Ling et al., 2022; Kurth et al., 2023). Recent reports suggest that the DL
 83 models are able to generate short and medium range weather forecasts globally with a skill
 84 that matches the best NWP models Bi et al. (2023); Lam et al. (2023); Y. Zhang et al.
 85 (2023). These developments suggest that the DL models can soon be cheaper alternatives
 86 to the computationally expensive NWP models. The DL models have also shown potential
 87 in sub-seasonal to seasonal scale forecasts as well (Weyn et al., 2021). They can also used
 88 in combination with the NWP models to improve the forecast skill (Rojas-Campos et al.,
 89 2023). Gupta et al. (2020) predicted the ENSO beyond the spring predictability barrier
 90 using the ConvLSTM, whereas the traditional models are unable to. Recently, a developed
 91 framework named Fourier Neural Operator (FNO) shows a robust skill in predicting partial
 92 differential equations (Li et al., 2020; Lu et al., 2021; Kossaifi et al., n.d.; Azzizadenesheli
 93 et al., 2023; Wen et al., 2023). However, the skill of DL models in forecasting the extreme
 94 weather events, such as tropical cyclones and monsoon LPSs, is yet to be proven.

95 Earlier studies predicted the MSLP time series, and used it as a proxy to predict the
 96 strength of active and break cycles of ISM (Sinha et al., 2021). However, the spatial map
 97 of MSLP anomalies are not yet predicted using a deep learning model. Here, we propose a
 98 framework to predict the genesis and track of LPSs by using a two step approach. Firstly, we
 99 predict the spatial pattern of daily MSLP anomalies using FNO. Secondly, we reconstructed
 100 the total MSLP field by adding the predicted anomalies to daily climatology. Then, the LPSs
 101 are tracked using an automated algorithm developed by Praveen et al. (2015), which mimics
 102 the manual tracking of trajectories of LPSs from the surface pressure charts. Though the
 103 model is used to predict the LPSs, it can also be potentially extended to predict tropical
 104 cyclones. The data and methodology are explained in section 2, the framework and the skill
 105 for predicting LPSs are shown in section 3, and the conclusions are presented in section 4.

106 2 Data and methods

107 Data

108 The daily mean sea-level pressure (MSLP) of the European Centre for Medium-Range
 109 Weather Forecasts fifth-generation (ERA5) reanalysis dataset (Hersbach et al., 2020) at a
 110 spatial resolution of $0.25^\circ \times 0.25^\circ$ from 1979–2018 is used in this study. For training and
 111 validation of the DL model 1979 – 2007 is used, and prediction is done for 2008 – 2018. The
 112 MSLP anomalies are computed by removing the long-term mean from 1979 – 2018. The
 113 region considered for the present study is $75^\circ\text{E} - 90^\circ$, $10^\circ\text{N} - 25^\circ\text{N}$. The LPSs are tracked
 114 using the algorithm developed by Praveen et al. (2015) from ERA5 using the daily MSLP.
 115 This LPS tracking algorithm identifies closed isobars at every one hPa interval from gridded
 116 MSLP data, and the storm’s center is identified as the centroid of the innermost isobar. The
 117 LPS centers identified from consecutive time intervals of gridded MSLP data are connected

118 to get the track. This algorithm mimics the manual tracking of LPSs from the pressure
 119 charts used by the India Meteorological Department and has been found to have a robust
 120 skill in tracking LPSs (Praveen et al., 2015).

121 LPS tracking

122 The LPS over the BoB (65° – 95°E and 0 – 23°N) and TC tracks over the WNP
 123 (110°E – 180°E and 0 – 30°N) from the model experiments are tracked using the algorithm
 124 developed by the Praveen et al. (2015), which mimics the conventional manual tracking
 125 algorithm based on sea level pressure closed isobars over the surface pressure charts. This
 126 algorithm searches for closed isobar at 1hPa interval at every time step around the grid
 127 of SLP minimum, and the storm center is taken as the centroid of the innermost closed
 128 isobar. The pressure depth (Δ SLP) is considered as the difference between the outermost
 129 and innermost closed isobar, and it signifies the intensity of the storm. Δ SLP \leq 2 hPa is
 130 called “low”, 2 hPa $<$ Δ SLP \leq 4 hPa is “depression”, 4 hPa $<$ Δ SLP \leq 10 hPa is “deep
 131 depression”, 10 hPa $<$ Δ SLP \leq 16 hPa is “cyclonic storm”, and Δ SLP $>$ 16 hPa is “severe
 132 cyclonic storm” (Mooley & Shukla, 1987; Sikka, 2006; Praveen et al., 2015).

133 Synoptic Activity Index

134 The genesis location, number of LPS days, and storm intensity of LPSs are together
 135 explained by defining an index named “Synoptic Activity Index” (SAI; (Ajayamohan et al.,
 136 2010)). The Δ SLP (pressure depth) measures the storm intensity. SAI is defined as the
 137 track density of LPS weighted by wind speed.

$$SAI = \sum_{n=0}^{n=l} \sum_{x-\Delta x}^{x+\Delta x} \sum_{y-\Delta y}^{y+\Delta y} U_{cat} \quad (1)$$

138 where l is the life span of an LPS in days, Δx and Δy are the grid spacing (1.5°) in
 139 X and Y directions, and x and y are the longitudinal and latitudinal positions of a storm
 140 center. The values of U_{cat} are 4.25, 11, 15, 20, 27.5 for the categories lows, depressions,
 141 deep depressions, cyclonic storms, and severe cyclonic storms, respectively (Ajayamohan et
 142 al., 2010; Sandeep et al., 2018).

143 ML layers

144 Convolutional layer

145 A convolutional layer is widely used in many computer vision algorithms, including
 146 CNNs that learn the spatial pattern robustly. A Convolution layer is also known as kernel
 147 convolution, where a kernel or filter (small matrix) is multiplied by an image or output from
 148 the previous layer. In a simple understanding, a convolutional layer acts like a spatial filter
 149 and extracts useful features from an image. The mathematical representation of an output
 150 from a convolutional layer is:

$$F[m, n] = X * h[m, n] \quad (2)$$

151

$$F[m, n] = \sum_j \sum_k h[j, k] X[m - j, n - k] \quad (3)$$

152 where $F[m, n]$ is the output feature matrix from a convolutional layer, $X[m, n]$ is the
 153 input image to a convolutional layer of width m and height n . The filter or kernel matrix is
 154 denoted by $h[m, n]$, which is multiplied by the input image.

155 The output from the convolutional layer ($F[m, n]$) is multiplied by a weights tensor
 156 (W), and bias (b) is added while training the model. Then, the output feature matrix

157 is passed to a nonlinear activation function. In each iteration, while optimizing the cost
 158 junction (J), the weights tensor (W) will be updated. The whole process in this layer is
 159 mathematically represented as:

$$Z^{[l]} = F^{[l]} * W^{T[l]} + b^{[l]} \quad (4)$$

$$Y^{[l]} = g(Z^{[l]}) \quad (5)$$

161 where $Y^{[l]}$ is the final output from a convolutional layer while training, $[l]$ denotes the
 162 iteration, and g is a nonlinear activation function.

163 ***Max-pooling layer***

164 A pooling layer aids in reducing the dimensions of a convolutional layer. In general, a
 165 pooling layer is placed just after a convolutional layer. The pooling layer involves sliding a
 166 filter along all channels in a feature matrix. In the case of max pooling, it picks a maximum
 167 value at a particular region of the sliding filter. The output dimensions from a pooling layer
 168 are:

$$(m - f + 1)/s * (n - f + 1)/s * c \quad (6)$$

169 where, m, n, c are the width, height, and number of channels of an image. s is the
 170 stride length and f is the size of the pooling filter. Pooling reduces the dimensions of the
 171 feature matrix from the convolutional layer therefore reducing the number of parameters
 172 to learn by the model and saving the amount of computational time. The pooling layer
 173 summarizes the features in a particular region instead of point-to-point or kernel-to-kernel
 174 mapping therefore helps the model learn robustly irrespective of the position and orientation
 175 of the features in an image.

176 ***Batch normalization layer***

177 A batch normalization layer reduces the covariance shift problem. It normalizes the
 178 intermediate output of each layer within the batch during the training of a model. This
 179 helps in stabilizing the optimization process and reduces the demand for dropout or other
 180 optimizations like the l2 norm. The mathematical representation of the batch normalization
 181 is:

$$X' = (x - M_b[x])/sqrt(var(x)) \quad (7)$$

182 where $M_b[x]$ is the mean of the mini-batch size and $var(x)$ is the variance of the mini-batch
 183 size and X' is the normalized component from the previous layer, and x is the output from
 184 the previous layer and input into the batch normalization layer.

185 It can be further developed as:

$$X'' = \gamma * X' + \beta \quad (8)$$

186 where X'' is the final output from the normalization layer. γ and β are the learnable
 187 parameters during the training of the model.

188 ***Dropout***

189 A dropout is a type of regularization that prevents the overfitting of the data during
 190 the training process. A dropout layer randomly makes the value of the nodes in a layer into
 191 zeros. The number of nodes to make zero depends on the input probability decided while
 192 tuning the hyperparameters.

193 **Flatten and dense layers**

194 As the name implies, a flatten layer flattens the multidimensional output from previous
 195 layers to a simple two-dimensional matrix. The output dimensions from a flatten layer are
 196 given below:

$$(M_b * m * n * f)_{output} = (M_b * (m * n * f))_{input} \quad (9)$$

197 Where M_b denotes the batch size and $m, n, and f$ are image width, height, and kernel size
 198 respectively. L.H.S. represents the output dimensions from a flatten layer and R.H.S. rep-
 199 represents the input multidimensional matrix to a flatten layer. A dense layer is a regular fully
 200 connected layer generally placed after a flatten layer. The operation done by a dense layer
 201 is given below:

$$Y = g(\Sigma(X * W^T) + B) \quad (10)$$

202 Where Y is the output from a dense layer, g is the nonlinear activation function, X is the
 203 input vector to a dense layer, W^T is a matrix of weights, and B is a bias vector.

204 **ConvLSTM and Fourier layer**

205 A ConvLSTM layer is a combination of a convolutional layer followed by a LSTM
 206 layer. The LSTM layer is a type of recurrent neural network that learns the sequential
 207 data, and the convolutional layer helps in understanding the pattern in the data. Therefore,
 208 collectively, a ConvLSTM is useful in learning spatiotemporal data robustly (Gupta et al.,
 209 2020; Sinha et al., 2021). The mathematical equations representing a ConvLSTM layer are
 210 as follows (Shi et al., 2015):

$$i_t = g_1(w_{ix} * x_t + w_{ih} * h_{t-1} + w_{ic} \cdot c_{t-1} + b_i) \quad (11)$$

$$f_t = g_2(w_{fx} * x_t + w_{fh} * h_{t-1} + w_{fc} \cdot c_{t-1} + b_f) \quad (12)$$

$$o_t = g_3(w_{ox} * x_t + w_{oh} * h_{t-1} + w_{oc} \cdot c_t + b_o) \quad (13)$$

$$m_t = g_4(w_{mx} * x_t + w_{mh} * h_{t-1} + b_m) \quad (14)$$

$$c_t = f_t * c_{t-1} + i_t \cdot m_t \quad (15)$$

$$h_t = o_t \cdot \tanh(c_t) \quad (16)$$

216 Where t is the t^{th} step, g_i is the nonlinear activation function like sigmoid. $*$ indicates the
 217 convolutional operation and \cdot denotes the element-wise multiplication. \tanh is an activation
 218 function. $i_t, f_t, o_t,$ and m_t represent the input gate, forget gate, output gate, and modulation
 219 gate. x_t is the input data to the ConvLSTM layer, and c_t and h_t are the cell and hidden
 220 state, respectively.

221 The main principle of a Fourier layer is to decompose the signal of a time domain
 222 into a frequency domain and to filter out the dominating frequency modes. The Fourier
 223 decomposition involves representing the input signal into the sum of cosine and sine wave
 224 components. The mathematical representation of a Fourier decomposition is given as follows:

$$f(x) = \sum_{i=1}^{\infty} 1/(\text{len}(x)/2)[a_i * \cos(i * 2\pi\omega x + \phi_i)] \quad (17)$$

225 The function $f(x)$ expresses the infinite linear combinations of sines and cosines of
 226 different frequencies of input variable x , where a and ϕ determine the amplitude and phase
 227 of the corresponding frequency (ω). The operation done in the Fourier layer is given below:

$$F' = FFT(X) * FFT(W) \quad (18)$$

228

$$F'' = IFFT(F') \quad (19)$$

229

230

231

232

233

234

235

236

where W and X are the randomly initialized weight matrix and the input X into the Fourier layer. FFT is the Fast-Fourier Transform and IFFT is the Inverse of FFT.

We employed a combination of a Fourier layer (F-layer), ConvLSTM layer, and convolutional layers. The Fourier layers convert the input into the frequency domain, and the weights are helpful in penalizing the dominant modes. Further, the ConvLSTM layers have a robust skill in predicting the sequential spatio-temporal data. We have also compared the skill of the LPS neural operator with the simple ConvLSTM without the Fourier layer.

237

3 Results and discussion

238

A Deep Learning Framework

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

The overall framework for predicting the MSLP anomaly is shown in Fig. 1. A sequential architecture uses both F-layer and ConvLSTM 2D layers as its first layer with five filters and ten filters, respectively, and Relu as an activation function in the ConvLSTM 2D layer. From recent studies, the ConvLSTM 2D is known for its efficiency in handling spatial-temporal data (Gupta et al., 2020). The input data into the model is a 5-dimensional tensor containing the length of the training data stack, input channels, latitude points, longitude points, and the stack of the input data for the past six days. The input data to the model is fed as stacked data, which means that the daily MSLP anomalies for the past six days are stacked and used to predict the next time step. The output from both the F-layer and the ConvLSTM 2D layer are concatenated and passed into two blocks of convolutional layer (Conv3D-1&2) having an activation of Relu, and five filters with a kernel size of 1×1 , and output is passed to a batch normalization layer (BatchNorm). Subsequently, the output from this step is passed to a Dropout (0.2), MaxPooling3D, Flattening layer, and two fully connected dense layers with 10 and 1 filters, respectively. Dropouts are added to the model wherever necessary during the parameter tuning to avoid overfitting the training data. The prominent features of the architecture are listed in Table 1. The model is optimized by tuning the hyper-parameters and the number of layers to obtain the best suitable combination of activation, number of filters, optimizers, dropouts, loss functions, etc. Satisfactory results were obtained with an epoch of 200 and a batch size of 160. Application of the MSE loss function yielded a model with good prediction capability. A total of 64 iterations are taken to learn the whole spatial map, satisfactorily. Therefore, in each iteration, the weights are not initialized randomly; rather, weights from the previous iteration were considered. The LPSNO is converged in the initial iteration of 200 epochs, as shown in Fig. 2. The shaded region shows the error bar (\pm standard deviation of loss function of all the 64 iterations from the initial iteration). The idea behind showing only the initial iteration of 200 epochs is the convergence of the model.

265

266

267

268

269

270

271

272

273

274

275

276

With the aid of the Fourier transforms, the MSLP anomaly is decomposed into a combination of sinusoidal waves, as shown in Fig. 3. The actual MSLP anomaly for the training period is shown in Fig. 3a, and the sinusoidal waves obtained from the Fourier decomposition are shown in Fig. 3b. Each sinusoidal wave has different amplitudes and phases; therefore, learning these high and low-frequency signals by an ML model helps in better prediction by considering the underlying weather modes. MSLP field can be reconstructed by combining the decomposed Fourier components (Fig. 3c). The power spectrum of daily MSLP anomaly for one JJAS season shows the maximum peak in the intraseasonal (30 - 60 day) period, and a secondary maximum in the synoptic and quasi-biweekly periods (Fig. 4). Therefore, the intraseasonal and synoptic scales are the two major components of the JJAS MSLP anomaly. The top nine sinusoidal components from Fig. 3b are shown in Fig. 5 for better visualizing the Fourier decomposition. The top

277 four panels in Fig. 5 show the Fourier components from intraseasonal oscillations, and the
 278 bottom four panels show the signals from the synoptic scale.

279 The Fourier layer introduced in the model architecture used in this study penalizes the
 280 important Fourier components by multiplying weights. Therefore, optimizing the Fourier
 281 weights helps the model learn the important weather modes. The Fourier layer primarily
 282 consists of three major layers: one is the Fourier transform of input time series, the second
 283 is the multiplication of weights to the Fourier transform, and the last one is the inverse
 284 transform into the time domain from the Fourier domain (Fig. 1b). The starting input to
 285 the model is a spatial map with latitude and longitude coordinates. However, the input
 286 map is fed into the model as an iterative 4×4 grid averaged time series out of a 64×64
 287 grid. The output from the model at each iteration of 200 epochs is compiled and depicted
 288 as a spatial structure again as a 16×16 grid size (1° resolution, Fig. 6).

289 The model predicts the daily MSLP anomalies at various lead times in a sequential
 290 fashion, i.e., the predicted one day lead is fed into the model to predict the day two, and so
 291 on. The LPSNO model was reasonably able to predict the MSLP anomaly spatial structure
 292 at a lead time of 3 days (Fig. 6). The right and left panels of Fig. 6 show the predictions of
 293 a low and high MSLP anomaly cases. The predictions are compared with the observations
 294 (Figs. 6a and e). The one day ahead prediction captured the spatial structure and magnitude
 295 of the MSLP anomalies reasonably well, for both negative and positive anomaly cases (Figs.
 296 6b and f). When the lead times are increased to two and three days, the quality of predictions
 297 weakened (Fig. 6c-h). Nevertheless, the overall structure of both the low and high pressure
 298 anomalies are predicted by the model at increased lead times.

299 The pattern correlation between the observed and the predicted spatial map of MSLP
 300 anomaly is shown in Fig. 7. The pattern correlation is defined as the Pearson product-
 301 moment coefficient of linear correlation between the two variables of the same dimensions.
 302 The prediction at the lead time of one day has a median pattern correlation of about 87%.
 303 Similarly, the median pattern correlation of lead two and lead three predictions is about
 304 $\sim 60\%$ and $\sim 50\%$, respectively. The correlation is weakening as time progresses; much lower
 305 values are observed in leads four and five (Fig. 7). An accurate prediction of the magnitudes
 306 and spatial pattern of MSLP anomalies is necessary to identify the intensity category and
 307 trajectory of LPSs, by reconstructing the full MSLP field using Eq. 6.16. The same strategy
 308 can be used to predict tropical cyclones as well. However, here, we focus only on LPSs.

$$MSLP = (MSLP)' + \overline{MSLP} \quad (20)$$

309 where $(MSLP)'$ is the anomaly of the MSLP and the \overline{MSLP} is the climatology (long-term
 310 mean).

311 Recent advancement of deep learning in this research area shows the ConvLSTM
 312 model's efficacy in handling spatial-temporal data (Gupta et al., 2020). Therefore, the
 313 current model of the Fourier layer variant is compared with the ConvLSTM model to see
 314 the prediction. Only the comparison of results from two iterations is shown in Table 2. The
 315 architecture of the LPS neural operator is the same as discussed above, with a Fourier layer
 316 concatenated with the ConvLSTM layer. Whereas in the case of the ConvLSTM model,
 317 there is no Fourier layer branch as shown in Fig. 1 and the rest of all architecture is the
 318 same; therefore, in ConvLSTM, the concatenation layer is also removed. The ConvLSTM
 319 model shows a correlation of around 0.77 in both the iterations between the observed and
 320 the predicted at lead 1. Whereas the LPSNO model shows a significant improvement in
 321 the prediction with a correlation of 0.84 in both iterations. The superiority of the proposed
 322 LPS neural operator over the simple ConvLSTM is seen.

323 The LPSs are tracked from the reconstructed daily MSLP field for 10 JJAS seasons.
 324 The lead 1 prediction captured 50 LPSs while 51 LPSs are observed during the same period.
 325 Although the lead time for the prediction is short, the accuracy of the model is remarkable.

326 The statistics for the different lead time predictions is shown in Table 3. The model’s
 327 ability to capture LPSs at higher leads diminishes gradually. One reason for this might be
 328 the deterioration of the skill in predicting the magnitude of the MSLP anomalies at greater
 329 lead times.

330 SAI is very useful for understanding the spatial distribution of LPS trajectories and
 331 their strength. The SAI for the observed period shows a maximum density over the head
 332 BoB, which is the core genesis region of the LPSs (Fig. 8a). The prediction at Lead 1 to
 333 5 days also shows the maxima over the head BoB, though with diminishing intensity with
 334 lead time of the prediction (Fig. 8b–f). At lead one, the model is also able to predict the
 335 propagation of LPSs in the northwest direction towards the continental India. The lead 5
 336 prediction captures the weakest synoptic activity, in line with the number of LPSs. The
 337 pattern of observed and predicted SAI is compared using the pattern correlation. The
 338 pattern correlation between the SAI of actual and the predictions at different lead times is
 339 also shown in Fig. 9. The pattern correlation between lead one predictions and observations
 340 shows the highest value of 0.94. Pattern correlations between observed and predicted SAI at
 341 lead 2 to 5 are 0.9, 0.87, 0.82, and 0.84, respectively (Fig. 9a). The pattern correlation alone
 342 is not a good measure of the skill of the model. The Root Mean Squared Error (RMSE)
 343 score between the predicted and observed SAI has a minimum value for lead time 1 and
 344 increases at subsequent lead times with the maximum RMSE score for lead 5 (Fig. 9b).
 345 When the pattern correlation and RMSE are taken together, the model skill in predicting
 346 the synoptic activity is quite low beyond a lead time of three days. The current LPS neural
 347 operator takes ~ 3.2 s to generate one day ahead prediction, which is significantly efficient in
 348 terms of computational resources required for a prediction using an NWP model. Further,
 349 the MSLP anomaly is predicted as a continuous variable in time, as discussed above, which
 350 makes it useful in operation weather forecast.

351 4 Conclusions

352 In this study, comprehensive deep learning framework to predict the spatial structure
 353 of the daily mean sea level pressure (MSLP) anomalies is proposed. Subsequently, synoptic-
 354 scale tropical storms known as “monsoon low pressure systems (LPS)” that contribute about
 355 60% of monsoon rainfall over the hugely populated Indo-Gangetic plains are tracked from
 356 the MSLP anomalies. To this extent, a start-of-the-art neural operator model comprised
 357 of a combination of Fourier and Convolutional Long Short Term Memory (LPS Neural
 358 Operator) is employed to predict the spatial MSLP anomaly map. A sequential prediction
 359 of MSLP anomalies is made using the prediction from the previous time step, similar to
 360 the conventional numerical weather prediction models. Median pattern correlations of 88%
 361 , 65% , and 50% , respectively, between the observed and predicted MSLP anomalies over
 362 the Bay of Bengal are obtained. Daily MSLP field is reconstructed by using the predicted
 363 anomalies and climatology. This MSLP field is used to track the LPSs over the BoB. The
 364 one day lead prediction captured almost the same number of LPSs as observed in a ten
 365 year period. At longer lead times, as expected, the model’s skill in capturing the LPSs
 366 diminished.

367 In the recent years, deep learning models are creating a revolution in the field of weather
 368 forecasting, with the models attaining the skill of the best operational numerical weather
 369 prediction models in the short and medium range forecasts Lam et al. (2023); Bi et al.
 370 (2023). However, the deep learning models are yet to prove their skill in capturing extreme
 371 weather phenomena such as tropical cyclones and LPSs. Here, we showed that a combination
 372 of Fourier and Convolutional Long Short Term Memory model is capable of accurately
 373 predicting the genesis of monsoon LPSs at one day lead time over a span of ten seasons.
 374 The predictions at lead times of up to three days are found to be reasonably well. Further
 375 improvements to this model will make it suitable for operational prediction of LPSs over
 376 the Indian region.

Acknowledgments

K. S. S. Sai Srujan is supported by the Prime Minister’s Research Fellowship by the Ministry of Education, Government of India. HK and SS acknowledge the funding support by the Ministry of Earth Sciences, Government of India through the project titled ”Prediction of Indian Summer Monsoon Active and Break Cycles using Deep Learning” (MoES/16/09/2022-RDESS).

Open Research

Data availability statement All data used in this study is freely available from public data repositories. The authors thank the developers of Matplotlib (Hunter, 2007) for making their code available on a free and open-source basis, which is used to generate all the figures. The codes and predictions of LPSNO (Srujan et al., 2024) can be accessed from <https://zenodo.org/doi/10.5281/zenodo.10499398>.

References

- Ajayamohan, R., Merryfield, W. J., & Kharin, V. V. (2010). Increasing trend of synoptic activity and its relationship with extreme rain events over central india. *Journal of Climate*, *23*(4), 1004–1013.
- Andersson, T. R., Hosking, J. S., Pérez-Ortiz, M., Paige, B., Elliott, A., Russell, C., ... others (2021). Seasonal arctic sea ice forecasting with probabilistic deep learning. *Nature Communications*, *12*(1), 5124.
- Azzizadenesheli, K., Kovachki, N., Li, Z., Liu-Schiaffini, M., Kossaiji, J., & Anandkumar, A. (2023). Neural operators for accelerating scientific simulations and design. *arXiv preprint arXiv:2309.15325*.
- Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023). Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, *619*(7970), 533–538.
- Chen, Y., Haywood, J., Wang, Y., Malavelle, F., Jordan, G., Partridge, D., ... others (2022). Machine learning reveals climate forcing from aerosols is dominated by increased cloud cover. *Nature Geoscience*, *15*(8), 609–614.
- Courant, R., Friedrichs, K., & Lewy, H. (1967). On the partial difference equations of mathematical physics. *IBM Journal of Research and Development*, *11*(2), 215–234. doi: 10.1147/rd.112.0215
- Deoras, A., Hunt, K., & Turner, A. (2021). The four varieties of south asian monsoon low-pressure systems and their modulation by tropical intraseasonal variability. *Weather*, *76*(6), 194–200.
- Goswami, B. N., Wu, G., & Yasunari, T. (2006). The annual cycle, intraseasonal oscillations, and roadblock to seasonal predictability of the asian summer monsoon. *Journal of Climate*, *19*(20), 5078 - 5099. doi: 10.1175/JCLI3901.1
- Greenfeld, D., Galun, M., Basri, R., Yavneh, I., & Kimmel, R. (2019). Learning to optimize multigrid pde solvers. In *International conference on machine learning* (pp. 2415–2423).
- Gupta, M., Kodamana, H., & Sandeep, S. (2020). Prediction of enso beyond spring predictability barrier using deep convolutional lstm networks. *IEEE Geoscience and Remote Sensing Letters*, 1–5. doi: 10.1109/LGRS.2020.3032353
- Ham, Y., Kim, J., & Luo, J. (2019). Deep learning for multi-year enso forecasts. *Nature*, *573*, 568–572. doi: 10.1038/s41586-019-1559-7
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... others (2020). The era5 global reanalysis. *Q J R Meteorol Soc*, *146*(730), 1999–2049.
- Hunt, K. M., & Fletcher, J. K. (2019). The relationship between indian monsoon rainfall and low-pressure systems. *Climate Dynamics*, *53*(3), 1859–1871.
- Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. doi: 10.1109/MCSE.2007.55

- 427 Hurley, J. V., & Boos, W. R. (2015). A global climatology of monsoon low-pressure systems.
428 *Q J R Meteorol Soc*, *141*(689), 1049-1064. doi: 10.1002/qj.2447
- 429 Kochkov, D., Smith, J. A., Alieva, A., Wang, Q., Brenner, M. P., & Hoyer, S. (2021). Ma-
430 chine learning–accelerated computational fluid dynamics. *Proceedings of the National*
431 *Academy of Sciences*, *118*(21), e2101784118.
- 432 Kossaifi, J., Kovachki, N. B., Azzadenesheli, K., & Anandkumar, A. (n.d.). Multi-
433 grid tensorized fourier neural operator for high resolution pdes, 2023. In *Url*
434 *https://openreview.net/forum*.
- 435 Krishnamurthy, V., & Ajayamohan, R. S. (2010). Composite structure of monsoon low
436 pressure systems and its relation to Indian rainfall. *Journal of Climate*, *23*(16), 4285-
437 4305. doi: 10.1175/2010JCLI2953.1
- 438 Kurth, T., Subramanian, S., Harrington, P., Pathak, J., Mardani, M., Hall, D., . . . Anand-
439 kumar, A. (2023). Fourcastnet: Accelerating global high-resolution weather fore-
440 casting using adaptive fourier neural operators. In *Proceedings of the Platform for*
441 *Advanced Scientific Computing Conference* (pp. 1–11).
- 442 Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., . . .
443 others (2023). Learning skillful medium-range global weather forecasting. *Science*,
444 eadi2336.
- 445 Li, Z., Kovachki, N., Azzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., & Anand-
446 kumar, A. (2020). Fourier neural operator for parametric partial differential equations.
447 *arXiv preprint arXiv:2010.08895*.
- 448 Ling, F., Luo, J.-J., Li, Y., Tang, T., Bai, L., Ouyang, W., & Yamagata, T. (2022). Multi-
449 task machine learning improves multi-seasonal prediction of the indian ocean dipole.
450 *Nature Communications*, *13*(1), 7681.
- 451 Lu, L., Jin, P., Pang, G., Zhang, Z., & Karniadakis, G. E. (2021). Learning nonlinear
452 operators via deepnet based on the universal approximation theorem of operators.
453 *Nature machine intelligence*, *3*(3), 218–229.
- 454 Mooley, D. A., & Shukla, J. (1987). *Characteristics of the westward-moving summer mon-*
455 *soon low pressure systems over the indian region and their relationship with the mon-*
456 *soon rainfall* (Tech. Rep.). Center for Ocean–Land–Atmosphere Interactions, 47 pp.
- 457 Praveen, V., Sandeep, S., & Ajayamohan, R. S. (2015). On the relationship between mean
458 monsoon precipitation and low pressure systems in climate model simulations. *Journal*
459 *of Climate*, *28*(13), 5305-5324. doi: 10.1175/JCLI-D-14-00415.1
- 460 Rojas-Campos, A., Langguth, M., Wittenbrink, M., & Pipa, G. (2023). Deep learning
461 models for generation of precipitation maps based on numerical weather prediction.
462 *Geoscientific Model Development*, *16*(5), 1467–1480.
- 463 Saha, S. K., Hazra, A., Pokhrel, S., Chaudhari, H. S., Sujith, K., Rai, A., . . . Goswami, B. N.
464 (2019). Unraveling the mystery of indian summer monsoon prediction: Improved es-
465 timate of predictability limit. *Journal of Geophysical Research: Atmospheres*, *124*(4),
466 1962-1974. doi: 10.1029/2018JD030082
- 467 Sandeep, S., Ajayamohan, R. S., Boos, W. R., Sabin, T. P., & Praveen, V. (2018). Decline
468 and poleward shift in indian summer monsoon synoptic activity in a warming climate.
469 *Proc. Natl. Acad. Sci. USA*, *115*(11), 2681–2686. doi: 10.1073/pnas.1709031115
- 470 Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-k., & WOO, W.-c. (2015).
471 Convolutional lstm network: A machine learning approach for precipitation now-
472 casting. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett (Eds.),
473 *Advances in neural information processing systems* (Vol. 28). Curran Associates,
474 Inc. Retrieved from [https://proceedings.neurips.cc/paper_files/paper/2015/
475 file/07563a3fe3bbe7e3ba84431ad9d055af-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2015/file/07563a3fe3bbe7e3ba84431ad9d055af-Paper.pdf)
- 476 Sikka. (2006). *A study on the monsoon low pressure systems over the indian region and their*
477 *relationship with drought and excess monsoon seasonal rainfall* (Tech. Rep.). Center
478 for Ocean–Land–Atmosphere Studies, Center for the Application of Research on the
479 Environment, 61 pp.
- 480 Sikka, D. R. (1977). Some aspects of the life history, structure and movement of monsoon
481 depressions. *PAGEOPH*, *115*, 1501–1529. doi: 10.1007/BF00874421

- 482 Sinha, A., Gupta, M., Srujan, K. S. S. S., Kodamana, H., & Sandeep, S. (2021). Prediction
 483 of synoptic-scale sea level pressure over the Indian monsoon region using deep learning.
 484 *IEEE Geoscience and Remote Sensing Letters*, 1-5. doi: 10.1109/LGRS.2021.3100899
- 485 Srujan, K. S. S. S., Sandeep, S., & Kodamana, H. (2024, January). *The Framework of*
 486 *Indian Summer monsoon Low-Pressure Systems Neural Operator (LPSNO)*. Zenodo.
 487 Retrieved from <https://doi.org/10.5281/zenodo.10499399> doi: 10.5281/zenodo
 488 .10499399
- 489 Thomas, T. M., Bala, G., & Srinivas, V. V. (2021). Characteristics of the monsoon low
 490 pressure systems in the indian subcontinent and the associated extreme precipitation
 491 events. *Climate Dynamics*, 56(5), 1859–1878.
- 492 Wang, B., Ding, Q., Fu, X., Kang, I.-S., Jin, K., Shukla, J., & Doblas-Reyes, F. (2005). Fun-
 493 damental challenge in simulation and prediction of summer monsoon rainfall. *Geophys-*
 494 *ical Research Letters*, 32(15), L15711. doi: <https://doi.org/10.1029/2005GL022734>
- 495 Wang, B., Xiang, B., Li, J., Webster, P. J., Rajeevan, M. N., Liu, J., & Ha, K. J. (2015).
 496 Rethinking indian monsoon rainfall prediction in the context of recent global warming.
 497 *Nature Communications*, 6, 7154. doi: 10.1038/ncomms8154
- 498 Wang, C., Zheng, G., Li, X., Xu, Q., Liu, B., & Zhang, J. (2022). Tropical cyclone inten-
 499 sity estimation from geostationary satellite imagery using deep convolutional neural
 500 networks. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1-16. doi:
 501 10.1109/TGRS.2021.3066299
- 502 Wen, G., Li, Z., Long, Q., Azzadenesheli, K., Anandkumar, A., & Benson, S. M. (2023).
 503 Real-time high-resolution co 2 geological storage prediction using nested fourier neural
 504 operators. *Energy & Environmental Science*, 16(4), 1732–1741.
- 505 Weyn, J. A., Durran, D. R., Caruana, R., & Cresswell-Clay, N. (2021). Sub-seasonal
 506 forecasting with a large ensemble of deep-learning weather prediction models. *Journal*
 507 *of Advances in Modeling Earth Systems*, 13(7), e2021MS002502.
- 508 Zhang, Q., Wang, H., Dong, J., Zhong, G., & Sun, X. (2017). Prediction of sea surface
 509 temperature using long short-term memory. *IEEE Geoscience and Remote Sensing*
 510 *Letters*, 14(10), 1745–1749.
- 511 Zhang, R., Liu, Q., & Hang, R. (2020). Tropical cyclone intensity estimation using two-
 512 branch convolutional neural network from infrared and water vapor images. *IEEE*
 513 *Transactions on Geoscience and Remote Sensing*, 58(1), 586-597. doi: 10.1109/TGRS
 514 .2019.2938204
- 515 Zhang, Y., Long, M., Chen, K., Xing, L., Jin, R., Jordan, M. I., & Wang, J. (2023). Skilful
 516 nowcasting of extreme precipitation with nowcastnet. *Nature*, 619(7970), 526–532.

Table 1. LPS NEURAL OPERATOR ARCHITECTURE DETAILS

<i>Layer (type)</i>	<i>Activation</i>	<i># Filters</i>	<i>Kernel size</i>	<i>Dropout</i>	<i>Bias</i>	<i>Pool size</i>
Fourier Layer	-	5	1x1	-	-	-
ConvLSTM2D	Relu	10	3x3	-	-	-
Conv3D-1	Relu	5	1x1	-	True	-
Conv3D-2	Relu	5	1x1	0.2	True	-
MaxPooling3D	-	-	-	0.2	-	1x1x1
Dense-1	Relu	10	-	-	True	-
Dense-2	Linear	1	-	-	True	-

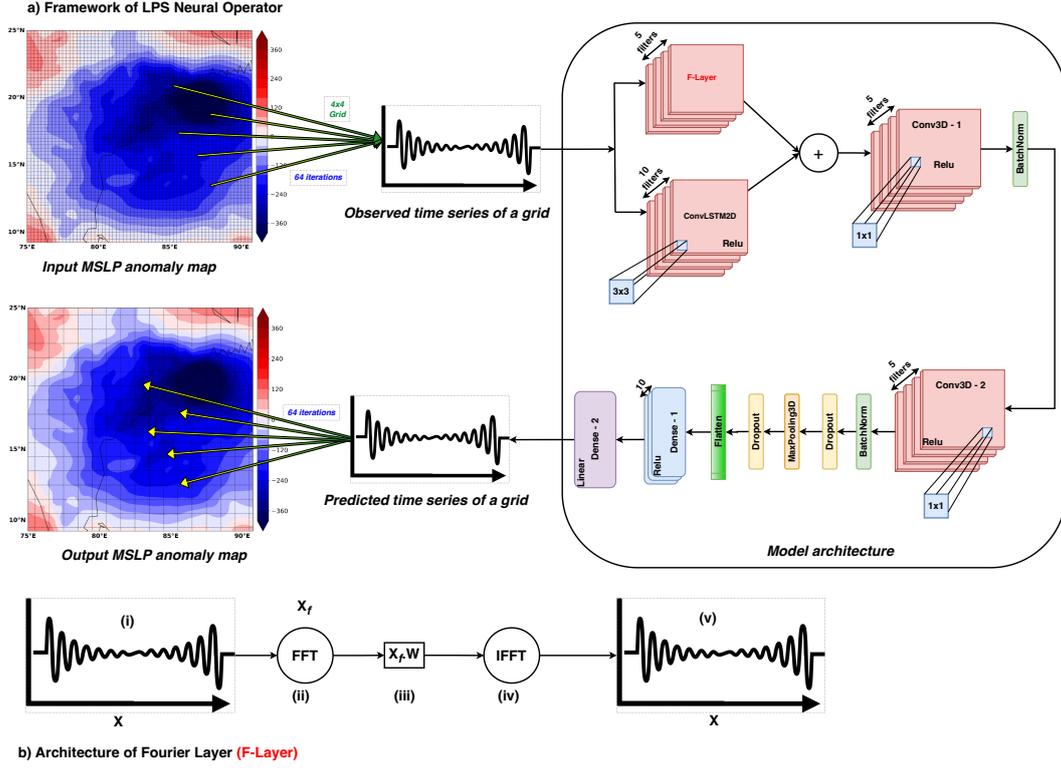


Figure 1. (a) Deep learning architecture used to predict the sea level pressure anomaly over the Bay of Bengal (details of the architecture are explained in Table 1). The plus symbol indicates the concatenation of two layers. The F-layer is the Fourier layer introduced in the deep learning model. The architecture of the Fourier Layer is shown in the bottom panel (b). The Roman numbers in (b) are explained as follows: (i) is the input time series, (ii) is the Fourier transform of (i, i.e., X_f), (iii) weight (W) multiplied to the X_f , (iv) is the inverse Fourier transform of the (iii), and (v) is the time series obtained from (iv).

Table 2. Comparison between ConvLSTM and F-layer ConvLSTM (LPSNO) at the lead time of 24 hrs

<i>Model</i>		<i>#Filters</i>	<i>Correlation</i>
LPSNO	iteration 1	5	0.83
	iteration 2	5	0.84
ConvLSTM	iteration 1	5	0.77
	iteration 2	5	0.76

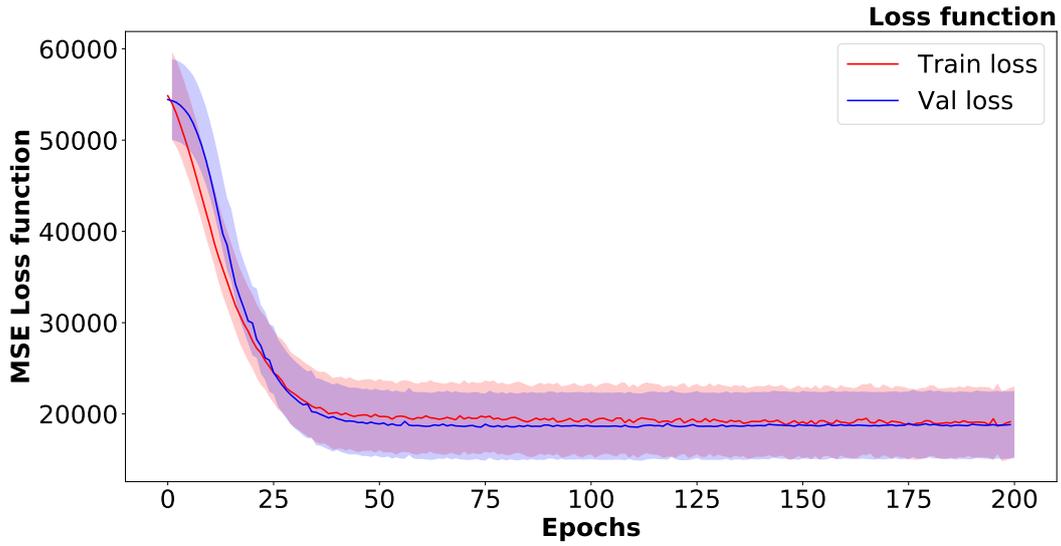


Figure 2. The loss function of the LPSNO model trained for the first step for 200 epochs is shown. Shading indicates the overall spread of the loss function computed as the total standard deviation in all steps. The red and blue indicate the training and validation curve, respectively.

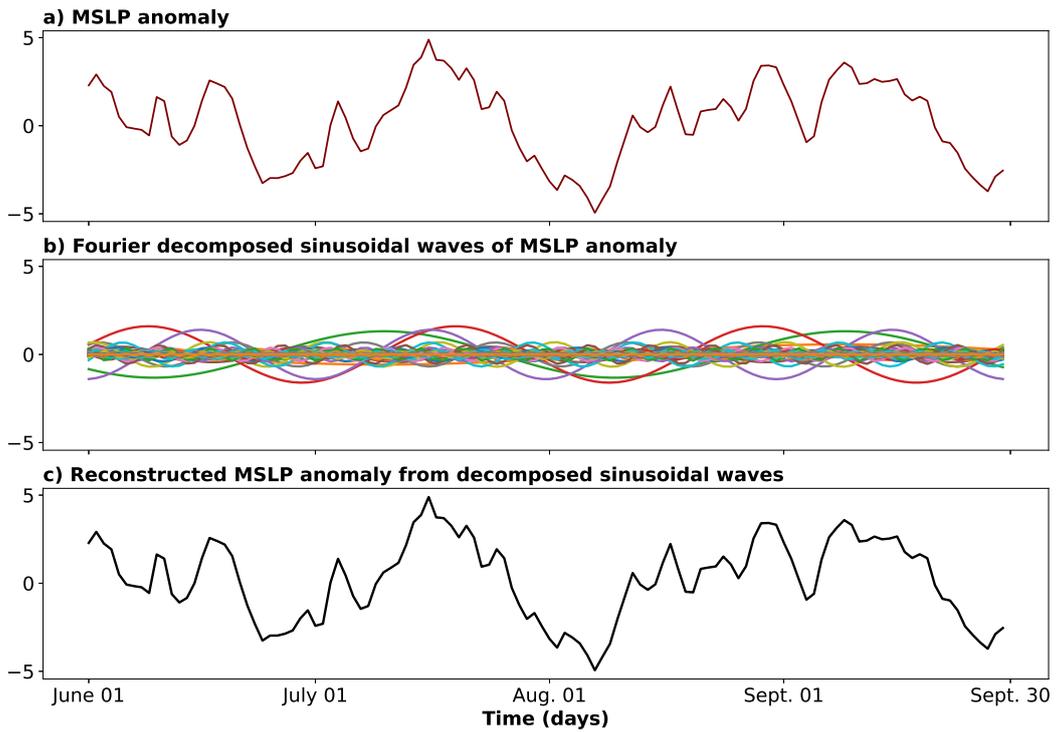


Figure 3. (a) Time series of MSLP anomaly area averaged over $75^{\circ}\text{E} - 90^{\circ}\text{E}$, $10^{\circ}\text{N} - 25^{\circ}\text{N}$ from 1979 – 2014 (period of training the model). (b) Fourier decomposed sinusoidal waves of MSLP anomaly from (a). (c) The reconstructed time series of MSLP anomaly using the decomposed Fourier components shown in (b)

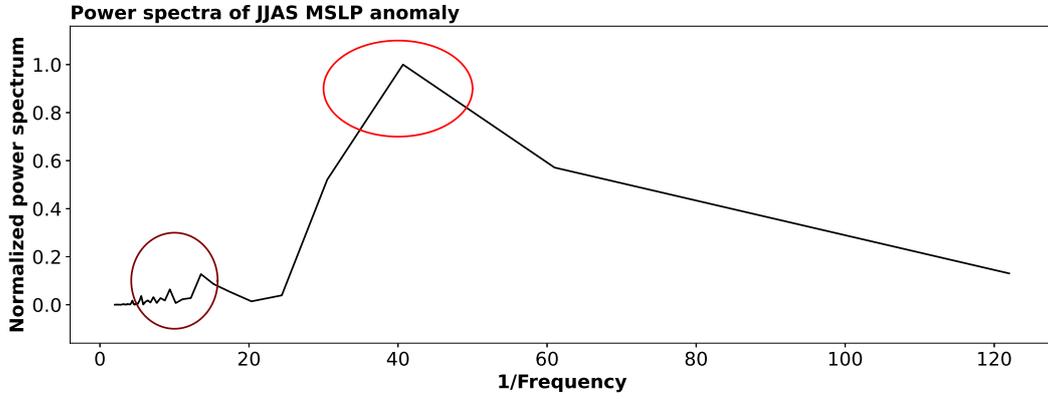


Figure 4. Normalized power spectrum MSLP anomaly for JJAS season of 1979. Red circle indicates the 1st maximum (represents the intraseasonal signal) and the maroon circle indicates the 2nd maxima (represents the synoptic signal) of normalized power spectrum

Table 3. Statistics of the actual and predicted MSLP anomaly at different lead times

<i>Type (actual/predicted)</i>	<i>Number of LPSs</i>
Actual	51
Lead 1	50
Lead 2	35
Lead 3	23
Lead 4	25
Lead 5	15

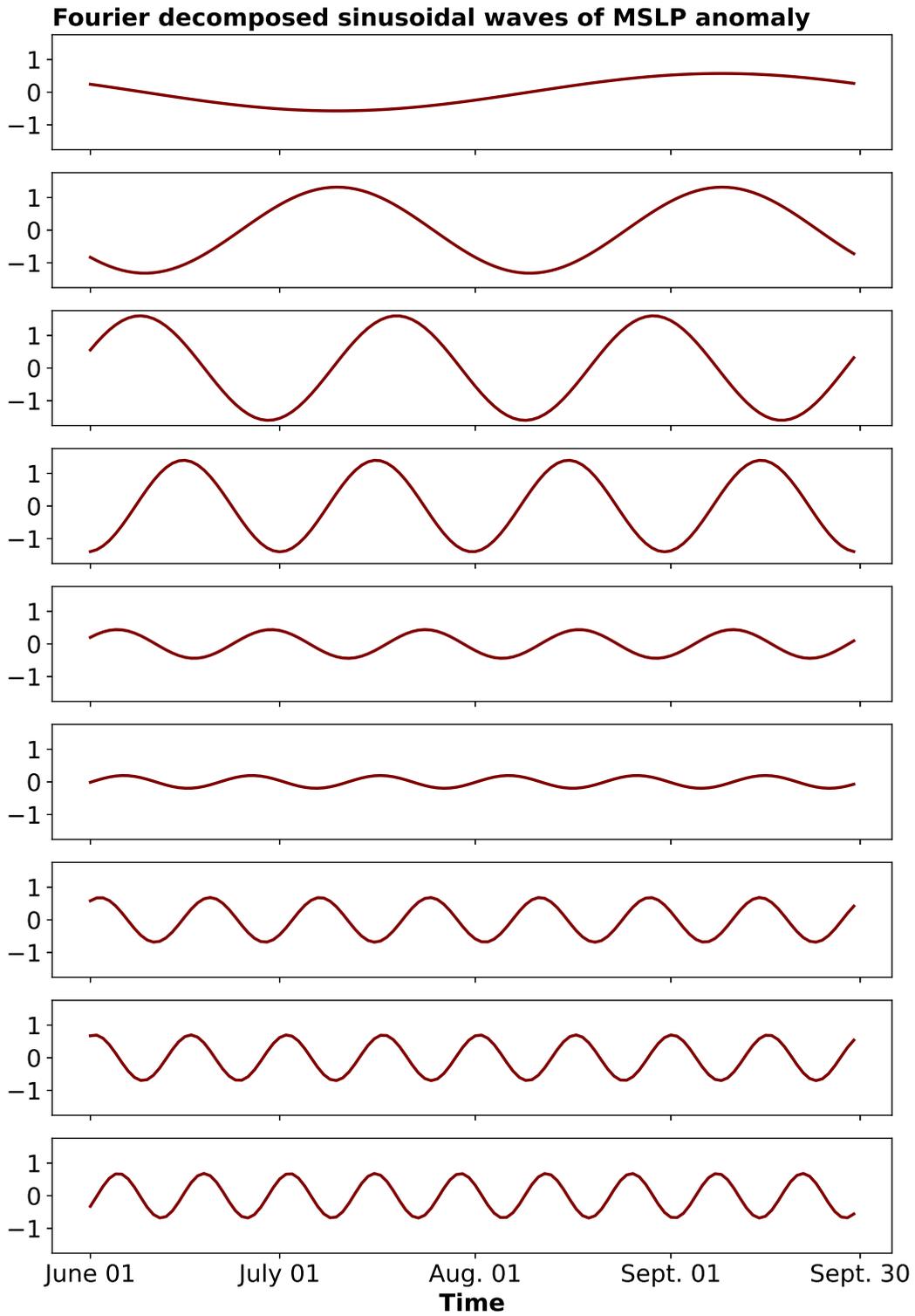


Figure 5. Top Nine sinusoidal wave components from the Fourier decomposition of MSLP anomaly as shown in Fig. 3. The y-axis shows the amplitude of the sinusoidal wave components, and the x-axis is the time (in days) of MSLP anomaly considered in this study, as mentioned in the data section.

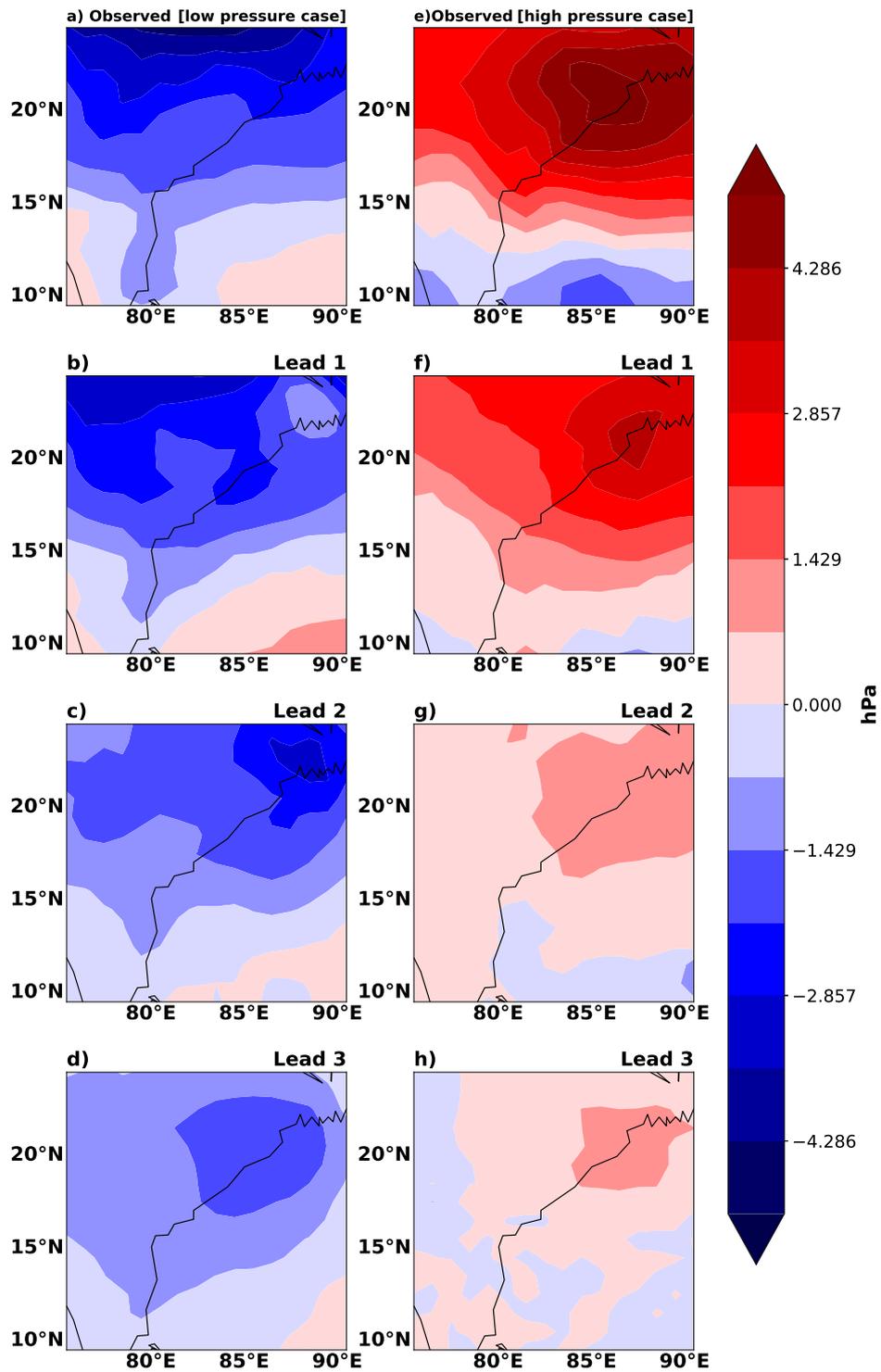


Figure 6. The actual (a,e) and predicted (b–d, f–h) spatial structure of sea-level pressure anomaly (units: hPa) over the Bay of Bengal at different lead times (24, 48, 72 hrs). The right and left panels are two different time steps.

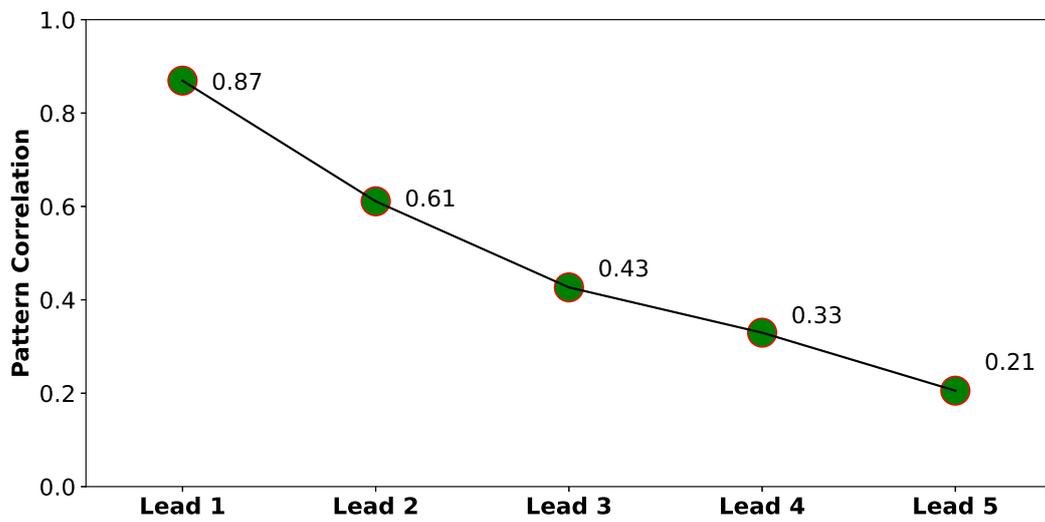


Figure 7. The median of pattern correlation between the observed and predicted sea level pressure anomaly at a lead time of 24, 48, 72, 96, and 120 hrs (Lead 1, Lead 2, Lead 3, Lead 4, and Lead 5 respectively) from 2008 – 2018.

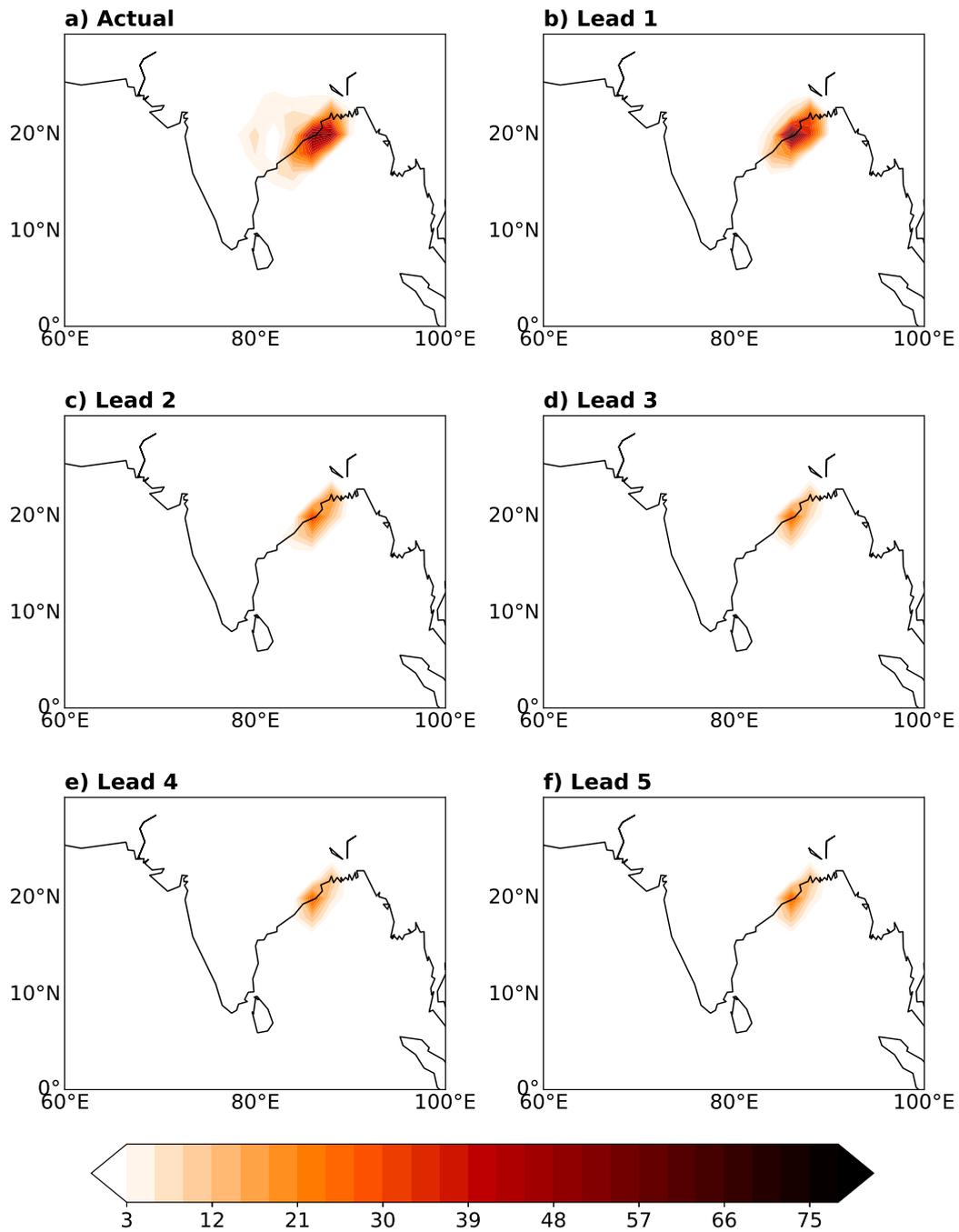


Figure 8. The synoptic activity index(SAI) computed for the life cycle of the LPSs for a) Actual and (b-f) predicted show the SAI at the lead times one to five days from 2008 – 2018.

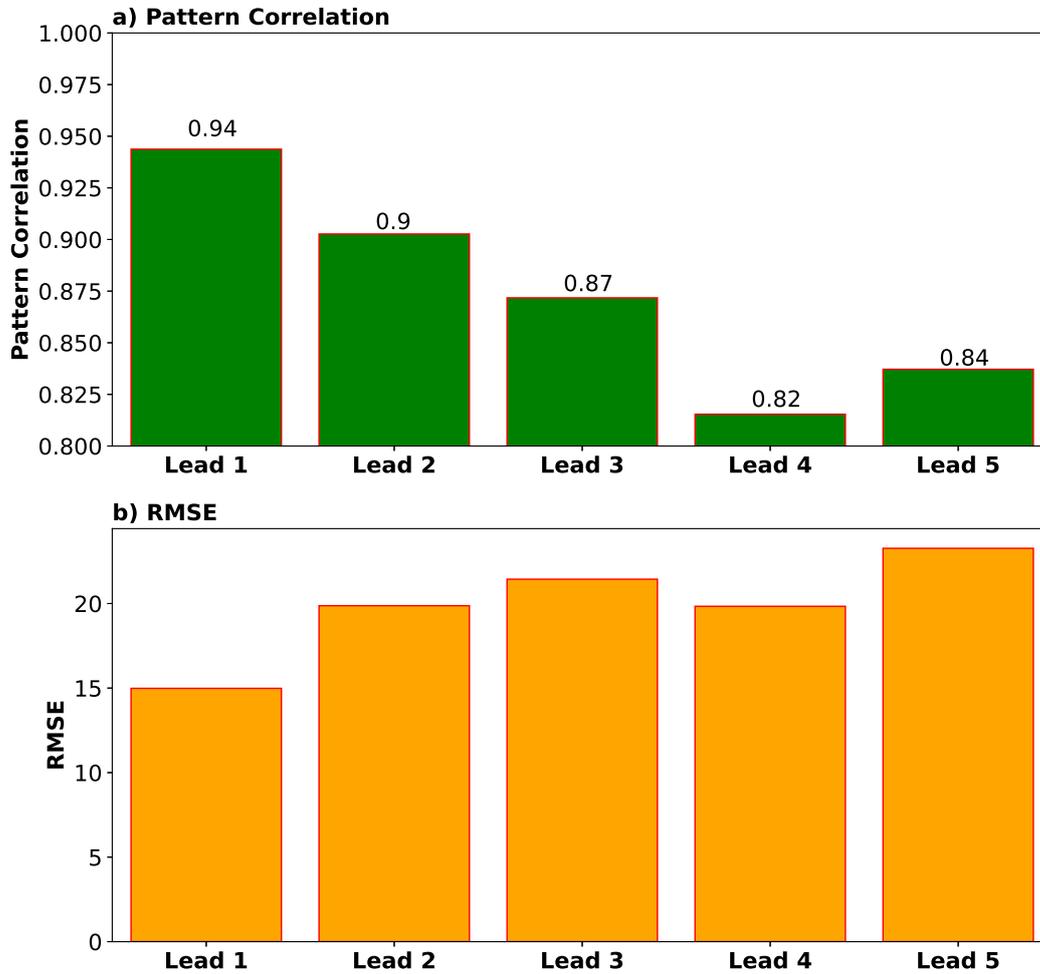


Figure 9. (a) The pattern correlation and (b) Root Mean Square Error (RMSE) score between the observed and predicted synoptic index computed using the life cycle of LPSs (from genesis to lysis) at a lead time of 24, 48, 72, 96, and 120 hrs (Lead 1, Lead 2, Lead 3, Lead 4, and Lead 5 respectively) from 2008 – 2018. (units of RMSE are the same as SAI)