Spatial-Temporal Forecast of the probability distribution of Oceanic Nino Index for various lead times

Jahnavi Jonnalagadda
Department of Information Science and Technology
George Mason University
Fairfax, Virginia, USA
jjonnala@gmu.edu

Mahdi Hashemi
Department of Information Science and Technology
George Mason University
Fairfax, Virginia, USA
mhashem2@gmu.edu, ORCID: 0000-0003-0212-0228

Abstract— **El Nino-Southern Oscillation (ENSO) is an irregular periodic oscillation in easterly winds and sea surface temperature (SST) over the tropical Pacific Ocean. El Nino and La Nina are warm and cold phases of ENSO. Oceanic Nino Index (ONI) determines ENSO events by calculating the three-month running mean of SST anomalies over the Nino 3.4 region (5°N-5°S and 120°W-170°W). El Nino refers to ONI greater than +0.5℃ and La Nina refers to ONI less than -0.5℃ for five consecutive months across the east-central equatorial Pacific. ENSO is one of the main drivers of Earth's inter-annual climate variability, which causes climate anomalies in the form of tropical cyclones, severe storms, heavy rainfall, and droughts. ENSO not only impacts global climate and oceanic conditions but also impacts food production, human health, and economy. Therefore, forecasting ENSO is of great importance. The main contribution of this study is proposing a convolutional long-short term memory that can capture spatial and temporal relationships between ENSO and environmental variables, such as SST, sea level pressure, meridional wind, and zonal wind. This study not only reports forecast accuracy but also quantifies the uncertainty associated with the forecast. Experimental results show that the proposed model improves the forecast accuracy by 14.8%, 10.4%, 11.8%, and 22.2% for lead times of 3, 6, 9, and 12 months, respectively.**

Keywords- Spatial-temporal forecast, Bayesian method, Probabilistic forecast, ENSO events, Climate anomalies, Variational inference

I. INTRODUCTION

El Nino-Southern Oscillation (ENSO) is characterized by irregular periodic variation in easterly winds and sea surface temperature over the Central and Eastern Pacific Ocean. It is one of the main drivers of Earth's interannual climate variability, which often causes a wide range of climate anomalies in the form of heavy rainfalls, severe droughts, heatwaves, unusual tropical storms, and other extreme weather conditions affecting both tropics and subtropics [1]. ENSO transitions between three phases: El Nino, neutral, and La Nina. Typical El Nino conditions in the east-central Pacific Ocean include a) above-average sea surface temperatures b) weak easterly winds, and c) deep oceanic thermocline than average. Typical La Nina conditions in the east-central Pacific Ocean include a) below-average sea surface temperature, b) strong easterly winds, and c)

shallow oceanic thermocline than average. Oceanic Nino Index determines ENSO events by calculating the three-month running mean of sea surface temperature anomalies over the Nino 3.4 region (5°N-5°S and 120°W-170°W). El Nino is observed in the Pacific Ocean if the ONI is greater than +0.5 ℃ for five consecutive months. La Nina is observed in the Pacific Ocean if the ONI is less than -0.5 ℃ for five consecutive months. Together, El Nino and La Nina can not only impact global weather, climate, and oceanic conditions but also food production, human health, economy, and water supply [1]. However, El Nino and La Nina differ from each other in terms of evolutionary patterns and impact on global climate. For instance, a prolonged El Nino or La Nina for more than two years has caused droughts in several regions of the United States, while the transition from El Nino to La Nina or La Nina to El Nino has caused flash floods in North-Eastern regions of Asia. Due to the impact of ENSO on the global climate, it is important to predict these events in advance.

Forecasting ENSO events in the literature can be classified into two types: a) dynamical and b) statistical models. Dynamical models use mathematical equations to describe physical laws governing interactions of atmosphere and ocean for forecasting ENSO. On other hand, statistical models learn patterns from historical data for forecasting ENSO. However, the latter models are often challenged by the complex and nonlinear nature of the ENSO. Machine learning models are statistical models that can extract salient features from high-dimensional data. Linear models such as autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) are too simple to capture the nonlinear and time-varying nature of the ENSO. To overcome the shortcomings of the linear models, artificial neural networks (ANN) were proposed in recent studies [2, 3]. Although ANNs can handle the nonlinearity in the data, they are not designed to explicitly handle the time-sequential dynamic interactions between variables. On the other hand, recurrent neural networks have a mechanism for capturing both nonlinear and time-varying dynamics of multivariate systems. However, RNNs suffer from vanishing gradient problems while backpropagating the error. A variant of RNN, known as long-short term memory (LSTM) was developed to overcome the vanishing gradient problem by

introducing several gates that would help the model decide what information to keep and what to forget.

Often, climate data exhibits both spatial and temporal autocorrelation, which means data from nearby locations are more similar compared to data from remote locations. Therefore, another challenge of statistical models is how to map the historical data in a meaningful way such that the model can learn spatial and temporal relationships between observed data and ENSO events. A fully connected LSTM can only handle long-term temporal dependencies between input variables, but they are not useful to capture spatial dependencies. Therefore, an extension of LSTM, which is convolutional long-short term memory (CLSTM) is introduced in [4] for precipitation forecasting. CLSTM replaces full connections in input and hidden layers of LSTM by convolution windows. In a recent study, ENSO is formulated as a spatial-temporal sequence forecasting problem, in which both input and output are sea surface temperature sequences [5]. However, their study only reported the forecast accuracy but did not quantify the uncertainty associated with the forecast. Estimating forecast uncertainty not only makes the forecasts reliable but also helps the decision-makers in the field to take appropriate actions.

In this study, spatial and temporal features of ENSO are derived by embedding ENSO predictors into a grid space, which fully expresses the spatial and temporal relationships with ENSO. Then these spatial-temporal features are fed to CLSTM to forecast ENSO events at three-month intervals up to one year. This study not only reports the forecast accuracy but also quantifies the uncertainty of the forecast.

II. LITERATURE REVIEW

Different machine learning and statistical analyses have been applied in literature for predicting environmental and urban phenomena [6, 7, 8, 9, 10, 11, 12, 13, 14, 15]. Forecasting ENSO in the literature is explored in this section under two categories: (a) deterministic methods and (b) probabilistic methods.

A. Deterministic methods

There have been many reviews done on ENSO predictability in recent years. Wu et al. [16] also used ANN and support vector regressor (SVR) to forecast sea surface temperature anomalies over the entire Tropical Pacific region with sea level pressure and sea surface temperature as predictors at a lead time 3-15 months. Ham et al. [17] used transfer learning to train a CNN first on historical data and subsequently on reanalysis data from 1871 to 1973. The forecasts were made for the years 1984 to 2017 and the correlation skill of CNN during these years is superior to that of state-of-art dynamical forecast systems. In a recent review, Dijkstra et al. [18] detailed the application of machine learning algorithms and their role in improving the prediction skill of ENSO.

B. Probabilistic methods

Recent studies use hybrid models by combining autoregressive integrated moving average and an ANN to predict the [19]. For a lead time up to six months, their model performed slightly higher than CSFv2 [20]. Whereas, for lead times beyond six months the prediction results are similar to that of shorter lead times. Another example of the hybrid model for forecasting ENSO events can be found in [21]. A suite of statistical and dynamical models [20, 22, 23, 24] are combined using the Bayesian model averaging. The weights for the models are derived using the expectation-maximization algorithm. The dynamical models, for example, CFSv2 [20], provides a single model ensemble forecast. It can give an estimation of the predictive uncertainty by making an ensemble of forecasts. Recent studies estimate the predictive uncertainty of forecasts, using an ensemble of standard neural networks (SNNs), called a deep ensemble (DE) [25], which can provide the mean and standard deviation of a Gaussian distribution instead of point estimates. Another probabilistic approach based on the Bayesian neural network is proposed in [26], which can give an estimate of the predictive uncertainty. However, the evaluation window for their model is very short (2015-2016). Peter et al. [27] experimented with the gaussian density neural network and quantile regression neural network to quantify the uncertainty of the forecasted ENSO. For the test period 1982-2001, the forecast accuracy is above 0.7 for 12 months lead time but drops to 0.4 for six months lead time for 2002-2017.

In this study, we propose CLSTM that can capture spatial and temporal relationships between ENSO and ENSO predictors that contributes in forecasting ENSO. The main difference between our work and previous works is that we not only report forecast accuracy but also quantifies uncertainty associated with the forecast.
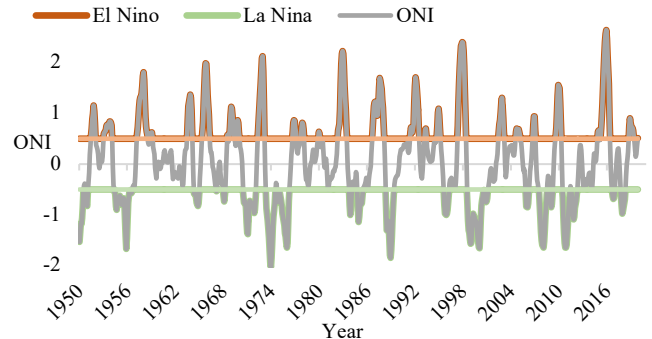

Figure 1. ONI variation from 1950 to 2019

III. DATA DESCRIPTION

ENSO events can be determined by forecasting ONI. The predictors used in this study are sea surface temperature (t: °C) sea level pressure (p: Nm^2), zonal wind speed (u: ms^{-1}), and meridional wind speed (v: ms^{-1}). The information about these variables is gathered from National Oceanic and Atmospheric Administration [28] and National Center for Environmental Prediction Reanalysis 1 [29]. The data are recorded at 500 different locations extending from (5°N-5°S and 120°W-170°W) from 1950-2019. The temporal resolution of the dataset is 1 month, and the spatial resolution of the dataset is 1° latitudinal and 1° longitudinal distance. Therefore, the dataset consists of a total of 420,000 records. Each record in the dataset represents the average of the three consecutive months. For example, the average sea surface temperature for three consecutive months (January, February, and March) is assigned

to January. All features (predictors) are normalized using standard scalar within the range of [0,1].

Figure 1 illustrates the variations of ONI from 1950-2019. The red and green lines are reference lines for El Nino and La Nina. Any value beyond the reference lines corresponds to El Nino or La Nina. From the figure, it is inferred that El Nino occurred more frequently than La Nina in the study period. The strongest El Nino recorded was in the year 2015, with a magnitude close to 3, and the strongest La Nina recorded was in the year 1973, with a magnitude close to -2. The data from 1950-2013 is used for training, while the data from 2013-2019 is used for testing.

IV. METHODOLOGY

The presence or absence of ENSO events spans across the Pacific region, but not at a single point in the Pacific region. Relating this with the dataset used in this study, all the data points belonging to the same time have the same output irrespective of their geographical location. This also means that the data points from the same time should be fed as a single input to the prediction model rather than feeding them as different inputs to the prediction model. In this way, the number of samples in the dataset is 840 (70 years multiplied by 12 months). We position ENSO predictors that belong to the same time on spatial grids, where each grid is for each predictor. These spatial grids are arranged next to each other to form a spatial feature map, and this spatial feature map is fed to the prediction model to determine ONI.

LSTMs are suitable for sequential prediction problems as they can remember the state of the previous time step with the help of memory gates. LSTM consists of three gates: input gate, forget gate, and output gate. With these gates, LSTM decides what information to store inside the memory unit and what information to throw away. More information about LSTM can be found in [30]. The flow of information through gates of LSTM is given by the following equations.

$$i_t = \sigma(W_{xi}X_t + W_{hi}H_{t-1} + W_{Ci}°C_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(W_{xf}X_t + W_{hf}H_{t-1} + W_{Cf} \circ C_{t-1} + b_f) \quad (2)$$

$$C_t = f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc}X_t + W_{hc}H_{t-1} + b_c) \quad (3)$$

$$o_t = \sigma(W_{x0}X_t + W_{ho}H_{t-1} + W_{Co} \circ +b_0) \quad (4)$$

$$H_t = o_t \circ \tanh(C_t) \quad (5)$$

Here i_t, f_t, o_t represents the input, forget, and output gates respectively, W with any suffix represents weight matrices, b with any suffix represents bias vectors and ∘ represents Hadamard product. Though LSTM has shown promising performance in sequence prediction problems, the full connections in input and hidden layers of LSTM cannot encode spatial information and therefore underperforms for spatial-temporal prediction problems. CNN has outperformed other neural network architectures in image processing and computer vision due to its ability to capture pixel dependencies in the image. Unlike LSTM, CNN does not have memory gates to remember the state of previous time steps and therefore underperforms for sequential data.

A. CLSTM

In a recent study, Shi et al. [4] introduced CLSTM for a spatial-temporal prediction problem. CLSTM combines the advantages of CNN and LSTM by replacing the full connections in the input and hidden layers of LSTM with convolutional windows. With this architecture, CLSTM can capture both spatial and temporal dependencies between predictors and the target variable. Therefore, we employ CLSTM in this study to forecast ONI. The flow of information through the gates of CLSTM is as follows.

$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{Ci}°C_{t-1} + b_i) \quad (6)$$

$$f_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{Cf}°C_{t-1} + b_f) \quad (7)$$

$$C_t = f_t° C_{t-1} + i_t° \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \quad (8)$$

$$o_t = \sigma(W_{x0} * X_t + W_{ho} * H_{t-1} + W_{Co}°C_{t-1} + b_0) \quad (9)$$

$$H_t = o_t ° \tanh(C_t) \quad (10)$$

Here * represents convolution operation and ∘ represents Hadamard product. The original CLSTM proposed in [4] follows encoding-forecasting structure for spatial-temporal sequence-to-sequence prediction. The dimensionality of the target is the same as the input. However, the encoding-forecasting structure cannot be implemented in this study due to different input and target dimensions. The architecture of the CLSTM consists of a 2D-CLSTM layer, a dropout layer, and three fully connected dense layers. The CLSTM layer consists of 32 filters, each filter with a size of 5×5. The size of the filter and the number of filters is determined using hyperparameter tuning. A dropout value of 0.5 means, only 50% of the output from the CLSTM layer is used as an input to successive layers of the network. Dropout avoids overfitting of training data. The first two layers of the fully connected layers consist of 128 and 64 neurons respectively, and the third layer (output layer) has as many neurons as the lead time. For example, the architecture for a lead time of 3 months has three neurons in its output layer.

B. Optimizing Model Parameters using Variational Inference

CLSTM provides the ENSO forecast when the model parameters ω are known. To quantify the uncertainty associated with the forecasts, we need to obtain the probability density function of ONI. This can be achieved when the true posterior distribution of model parameters given the observed data is known. Thus, the posterior distribution of model parameters according to the Bayesian rule is given as follows.

$$P(\omega|X,Y) = \frac{P(X,Y|\omega)P(\omega)}{\int P(X,Y|\omega)P(\omega)\,d\omega} \quad (11)$$

Here $P(\omega)$ is the prior of model parameters, $P(\omega|X,Y)$ is the true posterior distribution of model parameters given observed data, $P(X,Y|\omega)$ is the likelihood of observed data. However true posterior is generally intractable due to the multidimensional integrals in the denominator of (11). Variational inference (VI) is a technique used to approximate the posterior of model parameters $P(\omega|X,Y)$ by minimizing the distance between variational distribution $q_\theta(\omega)$ and true posterior. This is achieved by minimizing the Kullback-Leibler divergence (KL-divergence) between the variational distribution and the true posterior. Minimizing KL-divergence is the same as

maximizing the evidence lower bound (ELBO) concerning the variational parameters. Therefore, the final objective is to obtain the optimal distribution of model parameters by maximizing ELBO. The same can be formulated as follows.

$$ELBO = L(q) - KL\big(q_\theta(\omega)||P(\omega)\big) \qquad (12)$$

Here the first term L(q) is called expected log-likelihood and the second term is the KL distance between a variational distribution and true distribution of model parameters. Maximizing the first term would result in variational distribution $q_\theta(\omega)$ that explains the distribution of data well. Maximizing the second term would help the model from overfitting. The variational distribution $q_\theta(\omega)$ is assumed to be Gaussian with a predefined probability ρ and the standard deviation σ. The first term in (12) can be approximated by a Monte Carlo estimate using the mini-batches chosen randomly from the full dataset [X, Y] as follows.

$$ELBO \sim -\frac{N}{M}\sum_{m=1}^{M} ||Y^m - f_Y^{\hat{\omega}}(X^m)||^2 \qquad (13)$$

$$-KL(q_\theta(\omega)||P(\omega))$$

It is also assumed that prior follows a normal distribution with zero mean and unit variance $N(0, I)$. Therefore, the KL-divergence can be approximated as L2 regularization over the variational parameters θ. Once the optimal value of the variational distribution is obtained, the same can be used to approximate the true posterior. Now, given a test sample x^*, the probability of forecasting its output y^* is given by (14).

$$P(y^*|x^*, X, Y) = \int P(y^*|f_{y^*}^{\hat{\omega}}(x^*))\, q_\theta(\omega)\, d\omega \qquad (14)$$

To verify the performance of CLSTM, machine learning models, such as SNN, LSTM, CNN, and gaussian process regressor (GPR) are employed as baselines. The network settings used for baselines are as follows. The structure of SNN consists of three fully connected layers. The first two layers of SNN have 100 nodes each, followed by a third layer with as many neurons as the lead time. Similar to SNN, the structure of LSTM consists of three fully connected layers. The first two fully connected layers consist of 128 and 64 neurons respectively. The third fully connected layer has as many nodes as the lead time. Unlike CLSTM, where a sample is spatial maps of ENSO predictors of the same time, a sample in SNN and LSTM is ENSO predictors of a single geographical coordinate. Therefore, we need to train multiple LSTMs for data points recorded at the same time. The structure of CNN consists of two convolutional layers followed by a max-pooling layer after each convolutional layer and three fully connected dense layers. The first convolutional layer has 64 filters, each filter of size 5×5. The second convolutional layer has 32 filters, each filter of size 3×3. The first two fully connected layers have 100 nodes each, followed by a third layer with as many neurons as the lead time. For all neural network-based baselines, Relu is the activation function used in the hidden layers. GPR is a probabilistic method based on Bayesian theory. Multiple GPR models are trained for data points recorded at the same time. Each model corresponds to each geographical coordinate. In all baselines, mean squared error (MSE) is the loss function employed for error backpropagation.

V. RESULTS AND DISCUSSIONS

The results of hyperparameters are discussed as follows. The size of the filter and the number of filters determine the spatial information captured by the model.
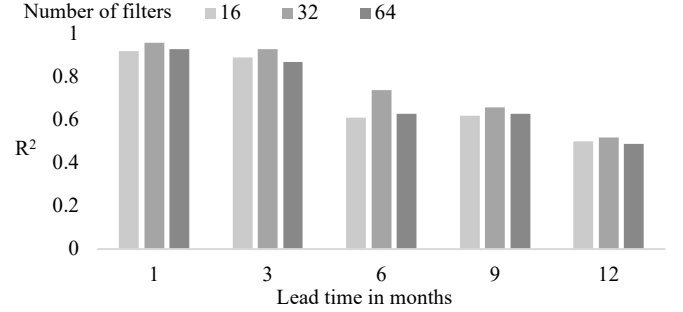

Figure 2. Coefficient of determination (R^2) in forecasting ONI by CLSTM for different number of filters

An increase in filter size increases the model's ability to capture multiple representations of the spatial dependencies. However, with more filters, the complexity of the model increases and they overfit training data quickly. Figure 2 illustrates the performance of CLSTM for various filter sizes and lead times. From the figure, it is evident that the R^2 value increases when the number of filters is changed from 16 to 32 and decreases when the number of filters is changed from 32 to 64. Since the performance of CLSTM with 32 filters is greater, the optimal choice for the number of filters is 32.
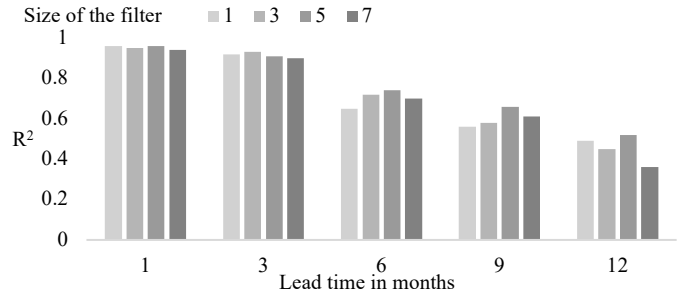

Figure 3. Coefficient of determination (R^2) in forecasting ONI by CLSTM for different sizes of the filter

As the filter size increases, CLSTM can capture a wide range of spatial dependencies between variables. Generally, odd-sized filters are used for modeling. The performance of CLSTM for various sizes of the filter is illustrated using Figure 3. A filter of size 1×1 means no spatial information is encoded for forecasting. For shorter lead times, filter size doesn't seem to have much effect on the model performance. For lead time beyond 6-months, the filter size of 5×5 performs best. From the results, we can conclude that for longer lead times, the model performs best when provided with larger amounts of spatial information.

The forecasting skill of all the experiments conducted in this study is verified using three metrics: root mean squared error (RMSE), mean absolute error (MAE), and coefficient of

determination (R^2). The metrics are defined by the following equations.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \qquad (15)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i) \qquad (16)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y_i})^2} \qquad (17)$$

Here, y_i and $\hat{y}_i$ denotes the actual and predicted values of the ith sample in the testing set, $\overline{y}$ and N denotes the mean and total number of samples in the testing set. The smaller is the values of RMSE and MAE, the better is the model performance. R^2 reaches its best at one and worst at zero.

TABLE I.   ACCURACY OF REGRESSION MODELS IN FORECASTING ONI

Model	Metric	Lead time in months				
		1	3	6	9	12
SNN	RMSE	0.77	0.82	0.87	0.95	1.15
	MAE	0.83	0.87	1.17	1.23	1.45
	R^2	0.66	0.58	0.45	0.32	0.28
LSTM	RMSE	0.45	0.51	0.53	0.76	0.97
	MAE	0.79	0.76	0.84	0.80	0.99
	R^2	0.89	0.81	0.67	0.59	0.42
CNN	RMSE	0.52	0.58	0.73	0.67	0.95
	MAE	0.71	0.78	0.72	0.88	1.23
	R^2	0.88	0.73	0.65	0.54	0.45
GPR	RMSE	2.38	3.21	4.78	5.02	7.87
	MAE	1.99	2.96	3.92	4.85	6.92
	R^2	0.22	0.18	0.13	0.11	0.10
CLSTM	RMSE	0.23	0.62	0.70	0.69	0.82
	MAE	0.67	0.72	0.70	0.75	0.85
	R^2	0.96	0.93	0.74	0.66	0.52

Continuous rank probability score (CRPS) is the metric used to assess the performance of probabilistic forecasts. It is defined by the following equation.

$$CRPS = \int_{-\infty}^{+\infty}[P(x) - H(x - y^{obs})]^2 \qquad (18)$$

Here $P(x)$ is the cumulative distribution function (CDF) of the forecast, y^{obs} is the observed value, and H is the Heaviside function, whose value is zero if $x - y^{obs} < 0$ and one if $x - y^{obs} > 0$. The CRPS achieves a perfect score (zero) if the observed and forecast values are the same.

Table 1 shows the comparison of different models for various lead times. The following observations are drawn from the results displayed in the table. The proposed model, CLSTM, outperforms other models in terms of R^2, RMSE, and MAE for various lead times. Particularly, R^2 values are close to one for lead times up to three months. Although the value of R^2 gradually drops for a lead time beyond six months, the proposed model is still superior when compared with others. CNN and LSTM almost performed equally in terms of R^2 for lead time up to six months. The better performance of CNN is due to the ability to capture spatial relationships between ENSO predictors and ENSO. The better performance of LSTM is due to the ability to capture temporal relationships between ENSO predictors and ENSO. The performance of CNN and LSTM drops beyond six months. SNN performed moderately for a lead time of three

months. Beyond three months, the performance of SNN drops due to the lack of proper handling of spatial-temporal data. GPR achieved lower R^2 and greater RMSE, and MAE scores. Experimental results demonstrate the ability of CLSTM in capturing both spatial and temporal relationships between ENSO predictors and ENSO. The results indicate the forecast accuracy is improved by 14.8%, 10.4%, 11.8%, and 22.2% for lead times of 3, 6, 9, and 12 months respectively.

TABLE II.   ACCURACY OF MODELS IN FORECASTING PROBABILISTIC DENSITY OF ONI

Model	Lead time in months				
	1	3	6	9	12
GPR	0.739	0.858	1.29	1.87	2.25
CLSTM	0.726	0.844	1.08	1.66	1.98

The CRPS scores for GPR and CLSTM are reported in Table 2. From the table, it is clear that CLSTM has the lowest CRPS score value for all lead times. The results conclude that the proposed model, CLSTM-BVI not only provides an accurate forecast but also a reliable probability density of the ONI.
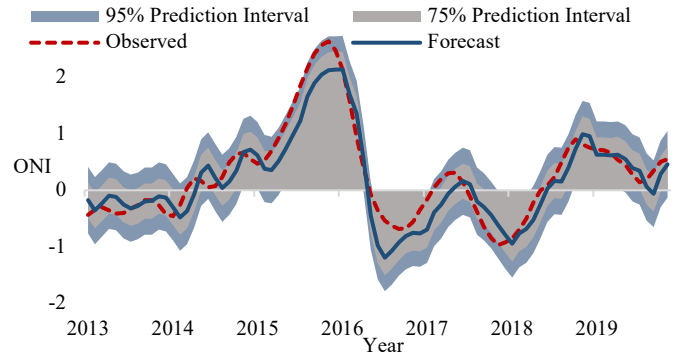

Figure 4.   Forecasted versus observed ONI

To estimate the forecast uncertainty obtained by CLSTM, Figure 4 is plotted for forecasted ONI and observed ONI. The light blue shade represents the 95% prediction interval, the light gray shade represents the 70% prediction interval. The blue solid line represents the forecasted ONI, and the dashed red line represents the observed ONI. As shown in the figure, the forecasted ONI follows the observed ONI, demonstrating the accuracy of CLSTM. Also, the observed ONI falls within the 95% prediction interval which means the proposed model is 95% confident that future values of ONI are contained within the 95% prediction interval.

VI.   CONCLUSIONS AND FUTURE DIRECTIONS

Experimental results demonstrate that CLSTM outperforms other models in terms of R^2, RMSE, and MAE. The R^2 value of CLSTM is close to one for shorter lead times (< 6 months), indicating the superiority of CLSTM in capturing spatial and temporal relationships between ENSO predictors and ENSO. However, the R^2 value gradually drops to 0.52 for longer lead times (> 6 months). An overall improvement in the forecast accuracy of CLSTM is observed. Experimental results indicate that the forecast accuracy of the proposed model improves by 14.8%, 10.4%, 11.8%, and 22.2% for lead times of 3, 6, 9, and

12 months respectively. The low CRPS value proves the superiority of the proposed model when compared with GPR. In addition to the forecast, the uncertainty associated with the forecast is also shown. The experimental results demonstrate that the CLSTM is 95% confident that the future values of ONI are contained within the 95% prediction interval. Therefore, the proposed model not only improves the forecast accuracy but also quantifies uncertainty associated with the forecast.

In future work, other environmental variables, such as warm water volume, westerly wind bursts, and upper ocean heat content can be added as predictors for forecasting ENSO. A robust framework can be built by adding more layers to the CLSTM network. Advanced methods can be proposed by integrating the existing methods with numerical weather prediction models.

REFERENCES

[1] S. Philander, "El Nino, La Nina, and the Southern Oscillation," *International Geophysics Series,* vol. 46, 1990.

[2] M. S. Baawain, M. H. Nour, A. G. El-Din and G. E.-D. Mohamed, "El Niño southern-oscillation prediction using southern oscillation index and Niño3 as onset indicators: Application of artificial neural networks," *Journal of Environmental Engineering and Science,* vol. 4, no. 2, pp. 113-121, 2005.

[3] F. T. Tangang, W. W. Hsieh and B. Tang, "Forecasting regional sea surface temperatures in the tropical Pacific by neural network models, with wind stress and sea level pressure as predictors," *Journal of Geophysical Research: Oceans,* vol. 103, no. C4, pp. 7511-7522, 1998.

[4] X. Shi, . Z. Chen, H. Wang, . D.-Y. Yeung, W.-k. Wong and W.-c. Woo, "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting," *Advancement in Neural Information Processing and Systems,* pp. 802-810, 2015.

[5] B. Mu, C. Peng, S. Yuan and L. Chen, "ENSO Forecasting over Multiple Time Horizons Using ConvLSTM Network and Rolling Mechanism," in *2019 International Joint Conference on Neural Networks (IJCNN)* , Budapest, Hungary, 2019.

[6] M. Hashemi, "A testbed for evaluating network construction algorithms from GPS traces," *Computers, Environment, and Urban Systems,* pp. pp. 96-109, 2017.

[7] M. Hashemi, "Automatic Inference of Road and Pedestrian Networks From Spatial-Temporal Trajectories," *IEEE Transactions on Intelligent Transportation Systems,* vol. 20, no. 12, pp. 4604-4620, 2019.

[8] M. Hashemi, "Intelligent GPS trace management for human mobility pattern detection," *Cogent Engineering,* vol. 4, no. 1, p. 1390813, 2017.

[9] M. Hashemi and A. A. Alesheikh, "A GIS-based earthquake damage assessment and settlement methodology," *Soil Dynamics and Earthquake Engineering,* vol. 31, no. 11, pp. 1607-1617, 2011.

[10] M. Hashemi and A. A. Alesheikh, "Development and implementation of a GIS-based tool for spatial modeling of seismic vulnerability of Tehran," *Natural Hazards and Earth System Sciences,* vol. 12, pp. 3659-3670, 2012.

[11] M. Hashemi and A. A. Alesheikh, "Spatio-Temporal Analysis of Tehran's Historical Earthquakes Trends," *Advancing Geoinformation Science for a Changing World, Springer,* pp. 3-20, 2011.

[12] M. Hashemi and H. A. Karimi, "Seismic Source Modeling by Clustering Earthquakes and Predicting Earthquake Magnitudes," *Smart City 360°,Springer,* pp. 468-478, 2016.

[13] M. Hashemi and H. A. Karimi, "Weighted machine learning for spatial-temporal data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 13, pp. 3066-3082, 2020.

[14] M. Hashemi, A. A. Alesheikh and M. R. Zolfaghari, "A spatio-temporal model for probabilistic seismic hazard zonation of Tehran," *Computers & Geosciences,* vol. 58, pp. 8-18, 2013.

[15] J. Jonnalagadda and M. Hashemi, "Forecasting Atmospheric Visibility Using Auto Regressive Recurrent Neural Network," in *2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI)*, Las Vegas, NV, USA, 2020.

[16] S. Aguilar-Martinez, W. W. Hsieh and L. Kantha, "Forecasts of Tropical Pacific Sea Surface Temperatures by Neural Networks and Support Vector Regression," *International Journal of Oceanography,* vol. 2009, pp. 1-13, 2009.

[17] Y.-G. Ham and J.-H. Kim, "Deep learning for multi-year ENSO forecasts," *Nature,* vol. 573, pp. 568-572, 2019.

[18] H. A. Dijkstra, P. Petersik, E. Hernández-García and C. López, "The application of machine learning techniques to improve El Niño prediction skill," *Frontiers of Physics,* vol. 7, pp. 1-13, 2019.

[19] P. D. Nooteboom, Q. Y. Feng, C. Lopez, E. Hernandez- Garcia and H. A. Dijkstra, "Using Network Theory and Machine Learning to predict El Nino," *Physics - Atmospheric and Oceanic Physics,* 2018.

[20] S. Saha, S. Moorthi, X. Wu, J. Wang, S. Nadiga, P. Tripp, D. Behringer, Y.-T. Hou, H.-Y. Chuang, M. Iredell, J. Meng, R. Yang, M. Mendez, H. Van Den Dool, Q. Zhang, W. Wang , M. Chen and E. Becker, "The NCEP Climate Forecast System version 2," *Journal of Climate,* vol. 27, no. 6, pp. 2185-2208, 2014.

[21] H. Zhang, C. Pao-Shin, L. He and D. Unger, "Improving the CPC's ENSO Forecasts using Bayesian model averaging," *Climate Dynamics,* vol. 53, no. 5-6, pp. 3373-3385, 2019.

[22] H. M. Van Den Dool, "Searching for analogues, how long must we wait?," *Tellus,* vol. 46, no. 3, pp. 314-324, 1994.

[23] Y. He and A. G. Barnston, "Long-lead forecasts of seasonal precipitation in the tropical Pacific Islands Using CCA," *Journal of Climate,* vol. 9, pp. 2020-2035, 1996.

[24] Y. Xue and A. Leetma, "Forecasts of tropical Pacific SST and sea level using a Markov model," *Geophysical Research Letters,* vol. 27, pp. 2701-2704, 2000.

[25] B. Lakshminarayanan, A. Pritzel and C. Blundell, "Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles," *Advances in neural information processing systems,* vol. 30, pp. 6402-6413, 2017.

[26] P. L. McDermott and C. K. Wikle, "Bayesian Recurrent Neural Network Models for Forecasting and Quantifying Uncertainty in Spatial-Temporal Data," *Entropy,* vol. 21, no. 2, p. 184, 2019.

[27] P. J. Petersik and H. A. Dijkstra, "Probabilistic Forecasting of El Niño Using Neural Network Models," *Geophysical Research Letters ,* vol. 47, no. 6, 2020.

[28] "National Oceanic and Atmospheric Administration," [Online]. Available: https://www.ncdc.noaa.gov/teleconnections/enso/indicators. [Accessed 10 January 2021].

[29] "NCEP/NCAR Reanalysis 1: Summary," [Online]. Available: https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html. [Accessed 10 January 2021].

[30] J. S. S. Hochreiter, "Long short-term memory," *Neural computation,* vol. 9, no. 8, pp. 1735-1780, 1997.