



Prediction of sea surface temperatures using deep learning neural networks

Partha Pratim Sarkar¹  · Prashanth Janardhan¹ · Parthajit Roy¹Received: 1 April 2020 / Accepted: 21 July 2020 / Published online: 30 July 2020
© Springer Nature Switzerland AG 2020

Abstract

Sea surface temperature (SST) prediction has widespread applications in the field of marine ecology, fisheries, sports and climate change studies. At present, the real-time SST forecasts are made by numerical models which are categorically based on physics-based assumptions subjected to boundary and initial conditions. They are more suited to a large spatial region than in a specific location. In this study, location-specific SST forecasts were made by combining deep learning neural networks with numerical estimators at five different locations around India for three different time horizons (daily, weekly and monthly). Firstly, forecasts were made with traditional neural networks (NNs) and then through deep learning neural networks. The NNs significantly improved on the results achieved by numerical forecasts which were further enhanced by the deep learning long short-term memory (LSTM) neural network over all timescales and at all the selected sites. The model was performed successfully in terms of various statistical parameters with correlation values nearing 1.0 while minimizing the errors. Additionally, a comparative study with a linear system, the autoregressive integrated moving average with exogenous input was made. The predictive skills of deep learning LSTMs are found to be more attractive than the other existing techniques (linear or other NNs) due to their ability of learning long time dependencies and extracting features from a sample space.

Keywords Deep learning · Long short-term memory network · Prediction · Neural network · Sea surface temperature

1 Introduction

Sea surface temperature (SST) observations are very important to understand the interaction of Ocean with Earth's atmosphere. SST provides principal data on the global atmospheric framework. It is a very important parameter in weather prediction and for the study of marine ecosystems. SSTs are especially important to determine the onset of El Nino events. They stand out as an important factor which helps in the identification of ENSO events as the temperature fluctuations averaged over a period of time in the Pacific region are more or less responsible for these events. These events have impacts on the climate all over the world. There are also other operational

benefits of SST measurements such as validation of atmospheric models, tracking of marine animals, evaluation of coral bleaching, tourism, fishery industries, militia and defense studies and sports. However, the precise prediction of SST is a tough task due to aberrations in the heat radiation and flux [12, 22]. The wind patterns over the sea surface also have uncertain nature as elucidated by [2]. The prediction of SSTs can be done using numerical based or data-driven methods. The numerical approach is better suited for predictions over a wider area, while the data-driven techniques are more applicable for location-specific studies. There are many such data-driven approaches ranging from familiar statistical methods to the newest machine learning and artificial intelligence mechanisms. In

✉ Partha Pratim Sarkar, Parthasarkar.nits@gmail.com; Prashanth Janardhan, prashnitk.hsn@gmail.com; Parthajit Roy, parthajit_roy@yahoo.co.in | ¹Department of Civil Engineering, National Institute of Technology, Silchar, Assam 788010, India.



such statistical approaches, the Markov model [25], regression models [9] and pattern inquisitive models [1] were employed. Artificial neural networks (ANNs) have become a very popular alternative among current data-driven methods. Artificial neural networks (ANNs), which derive their functionality from the human brain, have evolved as a very popular and capable method in modeling non-linearity and fitting random data. However, Martinez and Hsieh [11] and Wu et al. [23] also attempted to predict SSTs using genetic algorithms and support vector machines.

Many researchers in the past have employed ANNs to predict SST. The work of Tangang et al. [18] to predict sea surface temperatures over a specific location in the Pacific Ocean using neural networks (NN) can be considered pivotal in this field, as it was the first study of such kind. They made seasonal SST predictions by using empirical orthogonal functions of wind stress and SST anomalies as inputs fed into an ANN. Tang et al. [17] did a skill comparison study between linear regression (LR), canonical correlational analysis (CCA) [8], a more sophisticated version of LR and neural networks (NN). A multi-layer perceptron neural network model was developed by Wu et al. [23] developed a multi-linear feed-forward NN to forecast SST abnormalities across the tropical pacific. Tripathi et al. [19] used ANNs over a small area of Indian Ocean (27° to 35° S and 96° to 104° E) to predict sea surface temperature anomalies (SSTA). In this study, 12 networks were developed for each month of a year and the training of the NN was done on the area average values. The model was a feed-forward NN with logistic activation used for the neurons in the hidden layer to impart nonlinearity to the model. Gupta and Malmgren [5] did a comparison study on the forecasting abilities of various approaches which were dependent on certain training algorithms, regression and ANN. It was observed that NN performed better than other methods. Patil et al. [14] exhibited the usefulness of nonlinear autoregressive (NAR) type of NN models to predict monthly SST anomalies at six different locations for time lead of 1–12 months. In a similar work, Mohongo and Deo [10] predicted SST anomalies (monthly and seasonal) in the western Indian Ocean using nonlinear autoregressive with exogenous input (NARX) NN. Patil and Deo [13] exhibited the usefulness of a special kind of NN known as wavelet neural network by making SST predictions at six different locations around India. Xiao et al. [24] showed the usefulness of LSTM network in SST measurements using satellite data.

The review of the above-mentioned publications clearly shows the advantage of NNs in predicting SSTs. However, all the above-mentioned studies are site specific, either SST prediction is restricted to a few locations or the SST values were averaged over a region. It can be said that ANNs cannot consider spatial and temporal variability of

SST at once for a selected region. Nonetheless, for applications such as fishing, sporting events, coral bleaching, satellite measurement calibrations and tracking of marine animals, site-specific information is important. This is where the use of ANNs can come in handy although it requires a great deal of experimentation to reach its full capability. Most agencies across the globe utilize numerical models to provide real-time forecasts of SST. These numerical models employ the process of heat exchange across the oceans and the atmosphere for modeling SSTs. In this study, we investigated the prediction capability of such existing numerical models to forecast site-specific SSTs, and if not, whether this can be effectively achieved by artificial neural networks and deep learning NNs. A further comparative analysis was done with a linear system, namely the autoregressive integrated moving average with exogenous input (ARIMAX) model.

2 Study location and SST data

Normally for such SST prediction studies, the required data can be derived from various numerical model products of the ocean and atmosphere that are detailed over geographical grids of certain dimension. The data acquired as such are normally again polished by incorporating in situ based or satellite-based observations. Thermometers and thermistors, connected to static or moving buoys, record the onsite data. Satellites measurements and observations offer better understanding of its spatial and temporal variability. Both these methods of observation have reasonable agreement with each other.

The pre-processing of data was done by converting the unprocessed monthly data into aberrations by subtracting the absolute or raw values from their subsequent mean values calculated over the whole sample. These aberrations were then fed into the model as inputs. This helped in removing the small variations around the average as compared to the fluctuations in absolute values. The average for daily, monthly and weekly data was different. For example, say, January, the mean SST value of all the January months was calculated for the entire dataset. The anomaly was obtained by subtracting this mean value from the unprocessed January value and so on. This same process was repeated to obtain daily, weekly and monthly SST anomaly error series. Figure 1 shows the location of the five sites used in this study around India and within the Indian Ocean. The SST values were extracted for each of these sites and analyzed. The satellite data obtained from the different sources were then interpolated to a similar resolution of $1^\circ \times 1^\circ$ for each of the five different locations to produce meaningful results. Figure 1 also shows the

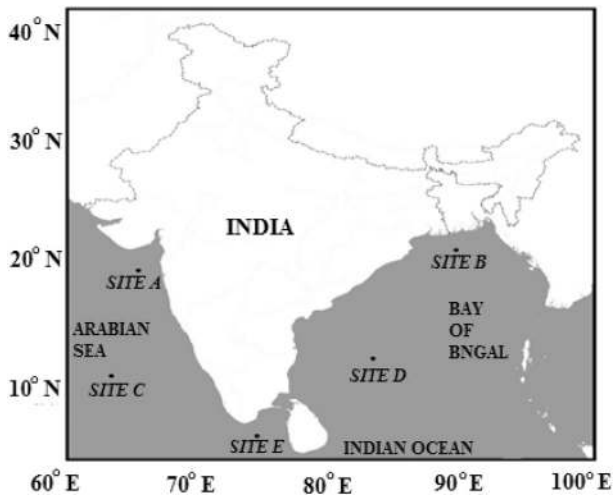


Fig. 1 The site locations (site A: 19° N–68° E, site B: 20° N–90° E, site C: 12° N–64° E, site D: 13° N–84° E and site E: 8° N–75° E)

coordinates of these specific sites and are named as sites A, B, C, D and E, respectively.

For the daily prediction, data sample size was 36 months (2015–18), for weekly prediction, data sample size was 38 years (1981–2018), and for monthly analysis, sample size was 149 years (1870–2018). The detailed features of these data are described in the next section. Table 1 shows the different data sources employed as per their availability.

2.1 Numerical dataset

The Canadian Centre for Climate Modeling and Analysis (CCCma) generated this numerical dataset from the

Table 1 Description of the datasets used in the study

Forecast horizon	Nature of data	Source of data	Time period
Daily dataset	Numerical	CCCma	36 months (2015–18)
	Observed	NOAA	36 months (2015–18)
Weekly dataset	Numerical	CCCma	38 years (1981–2018)
	Observed	HadISST	38 years (1981–2018)
Monthly dataset	Numerical	CCCma	149 years (1870–2018)
	Observed	NOAA	149 years (1870–2018)

second generation Earth System model with a spatial resolution of 0.93° N × 1.40° E. These data catered for the daily, weekly and monthly studies (<https://climate-modeling.canada.ca/climatemodeldata/cgcm4/cgcm4.shtml>).

2.2 Observational dataset

The National Oceanic and Atmospheric Administration (NOAA), USA, provides these observational SST datasets. The data benefit from the usage of Advanced Very High Resolution Radiometer (AVHRR) infrared satellite SST. Among various such NOAA products, we have used OISST version 2 dataset because of their massive sample size and close to surface measurements. These also take in account diurnal characteristics of the SST and are more suited for modeling shorter-time-ahead forecasts. They have grid resolution of 0.25° × 0.25°. These data cater for daily and monthly analysis (<https://www.esrl.noaa.gov/psd/data/gridded/data.noaa.oisst.v2.html>).

The Hadley SST data from the U.K. Hadley Centre's sea ice and sea surface temperature (HadISST) are used as the dataset for monthly analysis. The spatial resolution of these data is 1° × 1°. They assimilate observations acquired from the Global Telecommunications System (GTS) (<https://www.metoffice.gov.uk/hadobs/hadisst/>).

3 Methodology

The principle of time series forecasting is used in this study for the predicting future SST values. ANNs in such kind of prediction consider previous values, analyze the unknown patterns present, adjust itself according to the patterns in the dataset, and mimic and make future prediction one-step ahead. The obtained time-ahead values can be then validated using different statistical evaluation parameters to analyze the accuracy of prediction. In other words, if such a model can predict the preceding observations, we can say that it can be used for multiple-time-ahead forecasts. The generalization and adaptive capability of ANNs come in very handy.

3.1 Artificial neural network

A neural network (NN) is comprised of a large number of fundamental units known as neurons, which operate parallel to each other, organized in layers. The initial layer gathers the raw input data, drawing similarity to ocular nerves present in human beings, which help in visual processing. Thus, each layer in NN receives the output from the preceding layer. The last layer gives the output of the system as shown in Fig. 2 below. These layers are highly

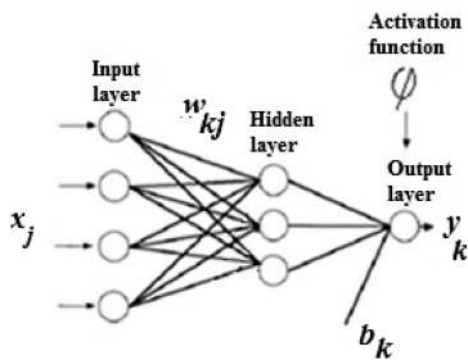


Fig. 2 Functioning of an ANN

interconnected, and every neuron can store information. One of the important properties of NN is adaptivity, which means that they can modify themselves according to the changes in their surroundings. This learning is achieved through the initial training and with every training they gain more knowledge of the system. Neurons contributing to better answers are assigned higher weights.

We may describe a neuron K in mathematical terms with the help of the following equations:

$$u_k = \sum_{j=1}^m w_{kj}x_j \tag{1}$$

$$y_k = \varphi(u_k + b_k) \tag{2}$$

where $x_1, x_2, x_3, \dots, x_m$ are the initial input signals; $w_{k1}, w_{k2}, w_{k3}, \dots, w_{km}$ are the connection strengths of neuron k , also known as the synaptic weights or weights; u_k is the linear combiner of the input signals to their corresponding weights; b_k is known as the bias; $\varphi(\cdot)$ is called the activation function; and y_k is the neuron's final output. Tan hyperbolic (\tanh), sigmoid, linear function, etc., are few commonly used activation functions. The function of the bias (a constant) is to adjust the output u_k obtained so that the overall model can fit best for the given data. The activation function analyses, whether a neuron should be initialized or not by evaluating the synaptic weights associated with the neuron and adding bias with it. Activation function adds nonlinearity into the output of the neuron.

NNs can be generally categorized into three categories, namely single-layer feed-forward neural networks, multiple-layer feed-forward neural networks (FFNN) and recurrent neural networks (RNN). Many NN applications in Oceanic studies show significant use of FFNNs, observed from the review of literature of similar works. A FFNN involves an input layer, one or several hidden layer and an output layer. As discussed above, the training or activation function and no. of hidden layers solely depends on the

objective. The network has to be trained with sufficient input–output pairs to produce good results, i.e., achieving the goal of error minimization. A large amount of experimentation with numbers of neurons in the hidden layer, different types of transfer and training algorithms, learning rate, momentum rate, etc., needs to be done to produce accurate results. In this study, a traditional feed-forward back-propagation (FFBP) NN along with nonlinear autoregressive exogenous (NARX) NN with two different training algorithms, namely Levenberg Marquardt and Bayesian algorithms, was used. FFBP is commonly applied for regression analysis due to its simple architecture and error minimization capabilities. However, to further improve the accuracy of the forecasts, another network known as NARX was employed. NARX has feedback connections enclosing several layers of the network. As part of the standard NARX architecture [16] the output is fed back to the input of the feed-forward neural network which results in accurate input for the next time step. One another advantage is that the resulting network has a distinct feed-forward architecture, and static back-propagation can be used for training. Levenberg Marquardt algorithm converges faster on its own and handles models with multiple free parameters efficiently. For large datasets, the training can be a bit slower, but the algorithm is capable enough to find an optimal solution. Similarly, the Bayesian algorithm is easier to implement and trains faster. It requires comparatively lesser data and is highly scalable and can make probabilistic predictions as well. The readers are referred to the textbooks of Haykin [6], Wasserman [20] and Wu [21] for detailed information on neural networks and their training algorithms.

3.1.1 Long short-term memory neural networks

Long short-term memory NNs commonly known as LSTMs are a special kind of RNNs. It has the ability to learn long time dependencies. The variation of LSTM from RNN is due to the changes in their repeating module [4]. In a standard RNN, the repeating module may consist of a very simple structure, e.g., tan hyperbolic layer, but LSTM has certain gates, namely forget gate, input gate and output gate, to control the cell state. These three gates define a single time step for LSTM. Forget gates determine whether past information has to stored or deleted. Input gates prioritize parts of current input vector. The output gate filters and determines what data should be the output to the hidden versus stored state. In other words, these gates control the entry of information, when to be used as an output or when to be removed from the cell memory [15] as shown in Fig. 3 below. This feature of LSTM is unique and can help store information in memory for longer time durations. LSTMs give us more controlling ability and better results

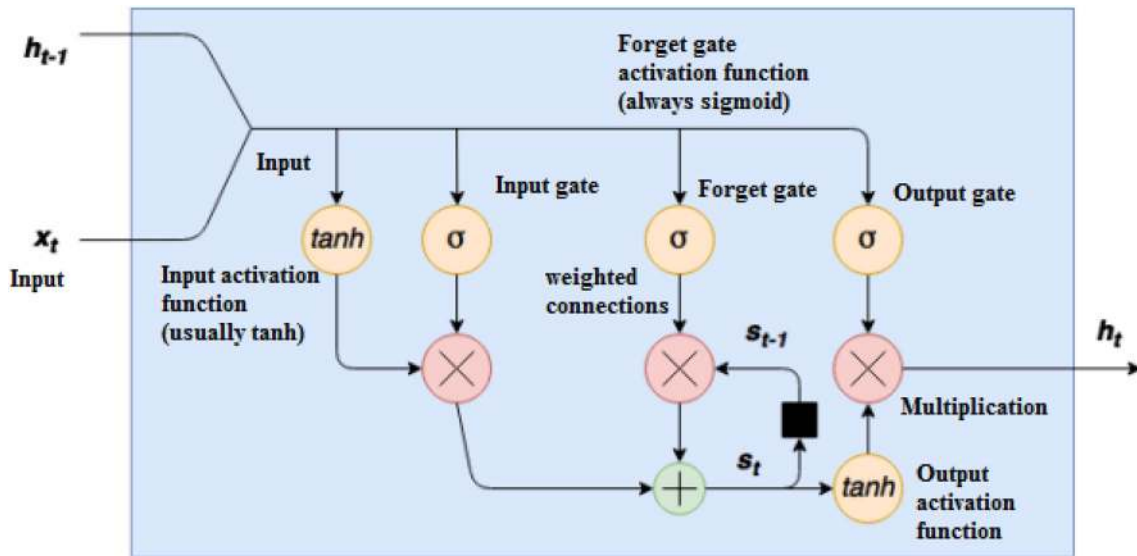


Fig. 3 Typical LSTM architecture

but certainly increase the complexity. Hence, to achieve high accuracy a lot of experimentation within the LSTM network architectures has to be done. For detailed information on the working of LSTMs, readers are referred to Hochreiter and Schmidhuber [7].

3.2 Autoregressive integrated moving average with exogenous input (ARIMAX) model

ARIMA stands for autoregressive integrated moving average model, a stochastic modeling approach that can be used to measure the probability of a future value lying between two certain limits. Box and Jenkins [3] developed the ARIMA model, whereas in the autoregression (AR) process the series current values are dependent on its previous values, the integration (I) involves subtracting an observation from an observation from a previous time step and the moving average (MA) is dependent on the observation and the obtained residual error. It can be expressed mathematically as follows:

$$Y^* = \theta_1 Y_{t-1}^* + \dots + \theta_p Y_{t-p}^* + \epsilon_t + \phi_1 \epsilon_{t-1} + \dots + \phi_q \epsilon_{t-q} \quad (3)$$

where ϕ and θ are unknown parameters and ϵ terms are autonomous similarly shared error terms with zero average. Here, Y^* is only expressed in terms of its past values and the current and past values of error terms. This is also known as ARIMA (p, d, q) model of Y . The AR behavior of the model is denoted by p , which is the number of lagged values of Y^* . The MA nature of model is represented by q , which is the number of lagged values of the error term and d is the number of times Y has to be differenced to produce the stationary Y^* . The term integrated implies that

in order to obtain a forecast of Y , we have to combine the values of Y^* because Y^* are the differenced values of the original series Y .

The transfer function of ARIMAX used in this study bears close resemblance to ARIMA. It improves the competence of ARIMA due to the introduction of transfer functions, innovations and additional explanatory variables shown in Eq. (4), where β can be a new transfer function and X can be any variable.

$$Y^* = \beta \cdot X + \theta_1 Y_{t-1}^* + \dots + \theta_p Y_{t-p}^* + \epsilon_t + \phi_1 \epsilon_{t-1} + \dots + \phi_q \epsilon_{t-q} \quad (4)$$

4 Model results

4.1 Numerical predictions

The statistics of observed SST for the five sites (site A to site B), minimum, maximum, average, standard deviation and coefficient of variation values in $^{\circ}\text{C}$, respectively, are shown in Table 2. From the table, we can observe the variations across the different locations.

The variations are more at sites C and E as compared to the other locations. These variations are due to changes in their regional environmental characteristics. The predictive skill of numerical SST forecasts versus observed SST was investigated for the various timescales as elucidated in the previous study area and data section. This analysis was done using various statistical parameters such as coefficient of correlation r , mean absolute error (MAE) and root mean square error (RMSE). Different statistical parameters provide a better perspective of the variations between the

Table 2 The basic statistics of SST datasets across the five locations over the three different timescales

Location		Minimum (°C)	Maximum (°C)	Average (°C)	Standard deviation (°C)	Co-variance (%)
Site A	Daily	25.760	31.100	27.789	1.036	3.138
	Weekly	26.742	32.115	27.802	1.038	3.141
	Monthly	27.780	32.324	27.811	1.039	3.143
Site B	Daily	26.480	31.000	28.181	0.843	2.990
	Weekly	27.443	31.015	28.195	0.845	2.991
	Monthly	28.501	31.224	29.203	0.848	2.994
Site C	Daily	25.270	30.820	28.162	1.283	3.845
	Weekly	26.252	30.835	28.176	1.284	3.849
	Monthly	27.290	30.944	28.184	1.287	3.850
Site D	Daily	26.020	30.800	28.360	0.987	3.485
	Weekly	27.013	30.815	28.374	0.989	3.743
	Monthly	27.040	30.924	28.382	0.992	3.778
Site E	Daily	23.530	30.800	27.829	1.604	5.782
	Weekly	25.340	31.915	28.842	1.609	5.792
	Monthly	27.548	32.824	29.851	1.612	5.798

Table 3 Single-step-ahead numerical versus observed forecast of SST anomalies in terms of *r*, RMSE and MAE (values in °C)

Site location		<i>r</i>	RMSE	MAE
Site A	Daily	0.33	0.69	0.56
	Weekly	0.12	0.75	0.54
	Monthly	0.11	0.56	0.66
Site B	Daily	0.21	0.72	0.45
	Weekly	0.31	0.81	0.38
	Monthly	0.15	0.67	0.59
Site C	Daily	0.23	0.55	0.66
	Weekly	0.18	0.78	0.34
	Monthly	0.09	0.43	0.52
Site D	Daily	0.38	0.78	0.61
	Weekly	0.11	0.83	0.42
	Monthly	0.12	0.59	0.49
Site E	Daily	0.29	0.43	0.39
	Weekly	0.20	0.72	0.52
	Monthly	0.07	0.34	0.48

observed and predicted values. *r* shows the strength of the relationship between the 2 datasets. RMSE and MAE show the deviations between the observed and predicted values. RMSE will always be greater or equal to the MAE as the errors are squared before they are averaged. RMSE gives more significance to large errors which is not the case for MAE. As compared to RMSE, MAE is a more natural measure of average error. Therefore, RMSE and MAE together would provide a better perspective to the analysis in terms of predictive accuracy. Table 3 shows a comparison of the anomalies (SST) between the unprocessed numerical

forecasts and their subsequent observations in terms of the parameters mentioned above.

The results (SSTA) obtained through numerical predictions clearly highlight the inaccuracy of the forecasts, which is evident from the table. The *r* values are very low than the much expected value of 1.0. Similarly, RMSE and MAE are also much higher than expected. Hence, the physics-based numerical predictions are not much suited for site-specific predictions. This limitation can be sorted out with data-driven approaches, also highlighting the fact that by adopting a hybrid model. The model would employ both numerical and data-driven technique. The following segment exhibits the development of a similar hybrid model, which combines neural networks with numerical predictions to produce enticing results for different lead times in the future.

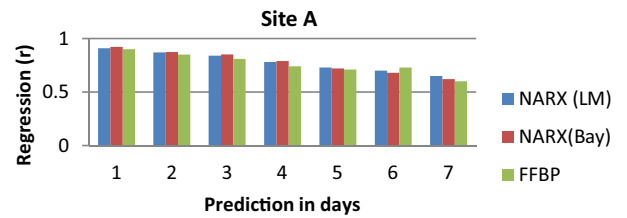
4.2 Numerical and neural network (data-driven) hybrid forecasts

In this method, we obtain the errors, which are nothing but the differences between the numerical forecasts and their corresponding observations. Simultaneously, we create a time series and this time series serves as an input for the neural networks that helps in forecasting anomalies at future lead times. The benefit of this method is that it incorporates the advantages of both the approaches.

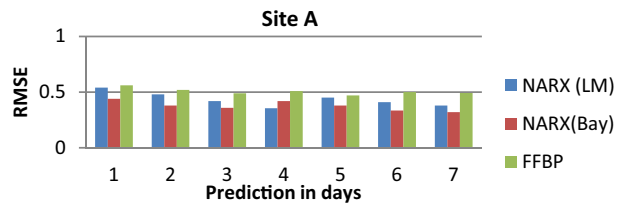
The selection of suitable neural network architecture depends on a lot of experimentation, done in this case also. Finally, a traditional feed-forward back-propagation network (FFBP) along with a non-autoregressive neural network with exogenous output (NARX) was adopted

for the modeling purpose. The NARX model was further tested with two different training algorithms, namely Bayesian algorithm and Levenberg–Marquardt algorithm. Table 4 shows the optimal architecture of the 2 NNs which resulted in the best forecasts for the three timescales. Validation failures help to prevent the network from performing poorly during validation and testing for unseen data. The training would cease to an end if the validation performance degrades for the max. limit set during the iterations. The parameters, namely performance goal and performance gradient, help in optimizing the network to achieve better accuracy. They are dependent on the type of dataset to be used.

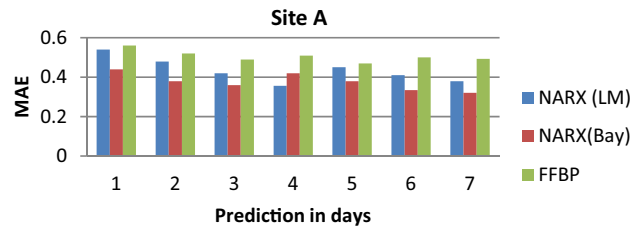
A lot of emphasis was given to avoid over fitting of the data, which can lead to loss of generalization capability of the network. After successful training, the model was tested for unknown input output parameters. The dataset was split into two parts; the latest 20% for testing and remaining 80% was divided into training and validation. The datasets pertaining to the three timescales (daily, weekly and monthly) were fed as inputs into the model. The sample size was 1080 (30 × 36), 1824 (4 × 12 × 38) and 1788 (12 × 149) for daily, weekly and monthly analysis, respectively. The results obtained improved on the previous numerical predictions. Figure 4a–c shows the results for the different neural networks mentioned above in terms of *r*, RMSE and MAE for site A as an example. The other four sites showed similar improved results. The *r* values considerably increased nearing 1.0, but RMSE and MAE were still much higher than expected. For the weekly and monthly timescales, the results showed a declining trend highlighting the loss of predictive capability with an increase in time duration.



(a) NN model performance for the daily predictions in terms of *r*



(b) NN model performance for the daily predictions in terms of RMSE



(c) NN model performance for the daily predictions in terms of MAE

Fig. 4 **a** NN model performance for the daily predictions in terms of *r*. **b** NN model performance for the daily predictions in terms of RMSE. **c** NN model performance for the daily predictions in terms of MAE

4.3 Predictions using LSTM

The forecasts of SSTA obtained thus far still leave scope for improvement. The numerical predictions came short

Table 4 Optimal architecture of the NNs used

Type of ANN	Training algorithm	Transfer function	No. of input, hidden and output layers	No. of iterations	Performance goal (during training)	Min. performance gradient (during training)	Max. validation failures (during training)
FFBP	Levenberg Marquardt	Sigmoid	Input-1, hidden-1 (10 neurons), output-1	500	0	1.00E−06	10
	Bayesian	Sigmoid	Input-1, hidden-1 (12 neurons), output-1	500	0	1.00E−06	10
NARX	Levenberg Marquardt	Sigmoid at hidden and linear activation function at the output	Input-1, hidden-2 (8, 10 neurons), output-1	500	0	1.00E−05	12

in terms of the statistical evaluation parameters, while the other two popular NNs improved the forecasting capabilities to an extent but still not quite as expected. This is further investigated with the application of LSTM model the advantages of which are described in the previous section. The same input used in the previous section was introduced into the LSTM network. The datasets were also split in the same manner as previously described. The determination of hyperparameters of the model requires adequate experimentation. Every LSTM layer was accompanied by a dropout layer. This layer helped to prevent over fitting by ignoring randomly selected neurons during training and hence reduced the sensitivity to the specific weights of individual neurons. A 20% testing set was used as a good compromise between retaining model accuracy and preventing over fitting. The activation function is introduced in the final layer. Therefore, after a lot of trials the hidden neurons were set at 200 for the optimum performance of the LSTM network and number of iterations were kept at 250. The initial learning rate was set at 0.005, and it was dropped after half of the iterations by

a factor of 0.2. After successful training, the network was reset which prevented past predictions from affecting the new data predictions. The results are evaluated in terms of r , RMSE and MAE. The forecasts are made on daily, weekly and monthly basis.

4.3.1 Daily predictions

The performance of LSTM networks during the testing phase of daily predictions is shown in Fig. 5 for the five different sites. The figure shows a better fit and correlation during the testing phase, while the spread is even for the sites A, C and E. LSTM improved on the previous forecasts obtained using NNs.

The network forecasts for the three time horizons in terms of r , RMSE and MAE are shown in Tables 5, 6 and 7, respectively. For the daily forecasts, the r values are very close to the value of 1.0 and MAE and RMSE have reduced significantly. For the 7 days forecast in the future, all the values of MAE and RMSE were found to be less than

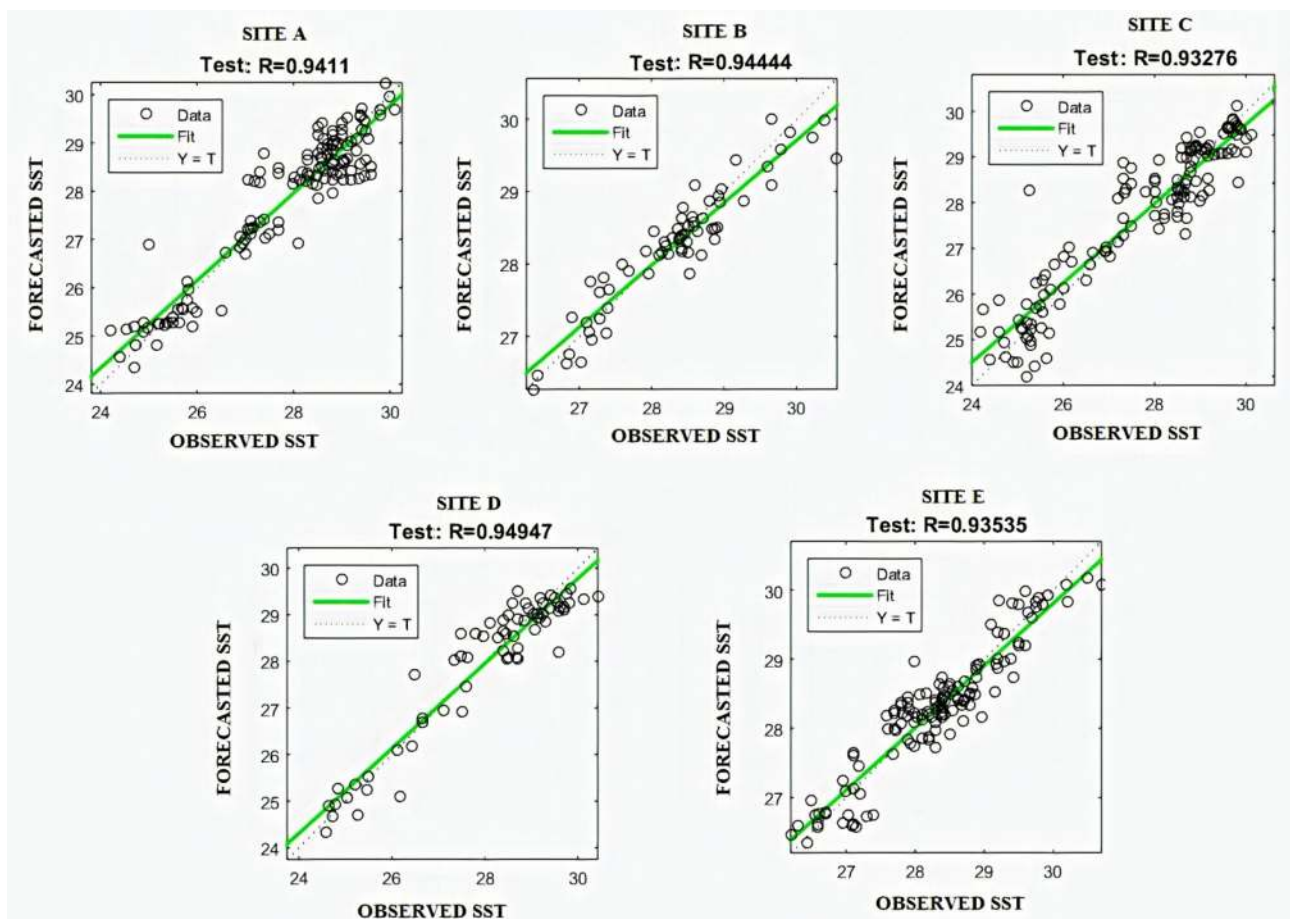


Fig. 5 LSTM network performance during testing phase (5 time steps ahead) at the five different sites

Table 5 LSTM network performance for the daily predictions in terms of r , RMSE and MAE

Forecast in days		1	2	3	4	5	6	7
Site A	r	0.99	0.97	0.96	0.95	0.94	0.92	0.91
	RMSE	0.06	0.08	0.12	0.15	0.17	0.19	0.21
	MAE	0.04	0.05	0.07	0.11	0.15	0.18	0.20
Site B	r	0.99	0.97	0.96	0.94	0.92	0.91	0.90
	RMSE	0.08	0.11	0.14	0.18	0.21	0.23	0.25
	MAE	0.05	0.07	0.10	0.14	0.17	0.20	0.23
Site C	r	0.99	0.98	0.95	0.93	0.91	0.89	0.86
	RMSE	0.06	0.10	0.11	0.15	0.19	0.22	0.24
	MAE	0.05	0.08	0.09	0.13	0.15	0.17	0.21
Site D	r	0.98	0.95	0.93	0.92	0.91	0.89	0.89
	RMSE	0.04	0.08	0.11	0.16	0.19	0.23	0.25
	MAE	0.05	0.07	0.10	0.14	0.15	0.18	0.22
Site E	r	0.98	0.98	0.97	0.97	0.96	0.94	0.92
	RMSE	0.04	0.05	0.08	0.11	0.14	0.18	0.20
	MAE	0.03	0.04	0.06	0.09	0.11	0.15	0.18

Table 6 LSTM network performance for the weekly predictions in terms of r , RMSE and MAE

Forecast in weeks		1	2	3	4	5	6	7
Site A	r	0.98	0.97	0.95	0.92	0.90	0.88	0.84
	RMSE	0.06	0.11	0.15	0.19	0.23	0.25	0.29
	MAE	0.05	0.08	0.12	0.15	0.18	0.22	0.26
Site B	r	0.99	0.97	0.96	0.92	0.89	0.86	0.83
	RMSE	0.05	0.10	0.14	0.18	0.23	0.26	0.29
	MAE	0.03	0.09	0.11	0.16	0.21	0.23	0.28
Site C	r	0.98	0.96	0.94	0.89	0.86	0.83	0.81
	RMSE	0.06	0.10	0.12	0.16	0.20	0.24	0.29
	MAE	0.04	0.08	0.11	0.15	0.18	0.23	0.27
Site D	r	0.97	0.96	0.94	0.89	0.86	0.84	0.80
	RMSE	0.05	0.11	0.15	0.17	0.21	0.24	0.28
	MAE	0.04	0.10	0.12	0.15	0.20	0.23	0.26
Site E	r	0.98	0.96	0.94	0.92	0.89	0.86	0.84
	RMSE	0.05	0.09	0.14	0.18	0.22	0.25	0.28
	MAE	0.04	0.10	0.12	0.16	0.20	0.23	0.27

0.26 °C. When the results within the same period were compared with the previous numerical SST and NN prediction, there was a considerable increase in prediction capability. The numerical SST forecasts produced $r=0.34$, $RMSE=0.51$ °C and $MAE=0.41$ °C, while the NN forecasts produced $r=0.9$, $RMSE=0.45$ °C and $MAE=0.32$ °C for the same testing period. The drop in accuracy of the numerical forecasts is due to the fact that they assume many physics based assumptions and are subjected to boundary and initial conditions. Therefore, numerical models are better suited for large spatial regions than for site-specific locations. Table 4 shows the network performance of LSTM over the future time steps. The error statistics improved

significantly with only a few anomalies that occurred with increase in future time leads.

4.3.2 Weekly predictions

The weekly predictions also far better in terms of the error statistics in comparison with the previous methods employed. The r values were higher and close to the desired value of 1, which shows that the forecasted values highly correlated with the observed values. The exceptions were at longer time leads. The five sites produced more or less similar kind of results. The RMSE and MAE values were less than 0.30 °C for all the sites across different time

Table 7 LSTM network performance for the monthly predictions in terms of *r*, RMSE and MAE

Forecast in months		1	2	3	4	5	6	7
Site A	<i>r</i>	0.96	0.87	0.84	0.79	0.71	0.66	0.61
	RMSE	0.09	0.14	0.19	0.23	0.27	0.31	0.36
	MAE	0.07	0.11	0.16	0.19	0.23	0.28	0.32
Site B	<i>r</i>	0.97	0.89	0.85	0.81	0.75	0.71	0.68
	RMSE	0.09	0.13	0.16	0.20	0.25	0.29	0.32
	MAE	0.06	0.10	0.13	0.17	0.21	0.24	0.28
Site C	<i>r</i>	0.97	0.87	0.80	0.74	0.69	0.66	0.62
	RMSE	0.06	0.18	0.21	0.28	0.33	0.37	0.40
	MAE	0.05	0.14	0.17	0.22	0.26	0.31	0.36
Site D	<i>r</i>	0.98	0.87	0.82	0.77	0.70	0.65	0.61
	RMSE	0.06	0.14	0.16	0.20	0.23	0.31	0.36
	MAE	0.05	0.11	0.13	0.16	0.18	0.23	0.31
Site E	<i>r</i>	0.97	0.91	0.87	0.81	0.76	0.70	0.65
	RMSE	0.07	0.15	0.17	0.24	0.29	0.32	0.36
	MAE	0.05	0.12	0.13	0.18	0.21	0.27	0.33

intervals. For the same test period, the numerical SST forecasts produced $r=0.29$, $RMSE=0.61\text{ }^{\circ}\text{C}$ and $MAE=0.58\text{ }^{\circ}\text{C}$, while the NN forecasts produced $r=0.78$, $RMSE=0.35\text{ }^{\circ}\text{C}$ and $MAE=0.31\text{ }^{\circ}\text{C}$. Table 5 shows the test performance of the model at all the sites and over different time steps ahead in the future.

4.3.3 Monthly predictions

The monthly predictions at longer lead times showed comparatively declining results but were still much better than numerical and the NN predictions for the same period. The values of *r* were found to be greater than 0.6 across all the time steps. The RMSE and MAE were less than 0.37 $^{\circ}\text{C}$ across all the intervals. The numerical predictions produced $r=0.12$, $RMSE=0.91\text{ }^{\circ}\text{C}$ and $MAE=0.88\text{ }^{\circ}\text{C}$, and the NN forecasts produced $r=0.48$, $RMSE=0.69\text{ }^{\circ}\text{C}$ and $MAE=0.61\text{ }^{\circ}\text{C}$ for the same test period. The monthly performance is shown in Table 6. After 4 time steps, there is a significant decrease in forecasts. As the time span

increases, the network loses the associative memory. This can be viewed as a drawback for such networks where it loses forecasting ability at longer time leads. This drawback is due to the fact that monthly data are smooth data as compared to daily or weekly data. The data smoothness has an indirect relationship with the performance of the network (LSTM or ANN).

A further investigation was done to compare the predictive skills of the LSTMs with a linear model (ARIMAX). ARIMAX performed better than numerical forecasts, but LSTM model clearly outperforms it. Figures 6, 7 and 8 show the performance of ARIMAX and LSTM models in terms of the above-mentioned statistical parameters for the various locations with varying timescale.

The other two site locations produced similar results. The ARIMAX model is not able to capture the long-term dependency and hence loses prediction capabilities at higher time leads. The much lesser values of *r* for all the timescales show the lack of correlation between the forecasted and observed SST.

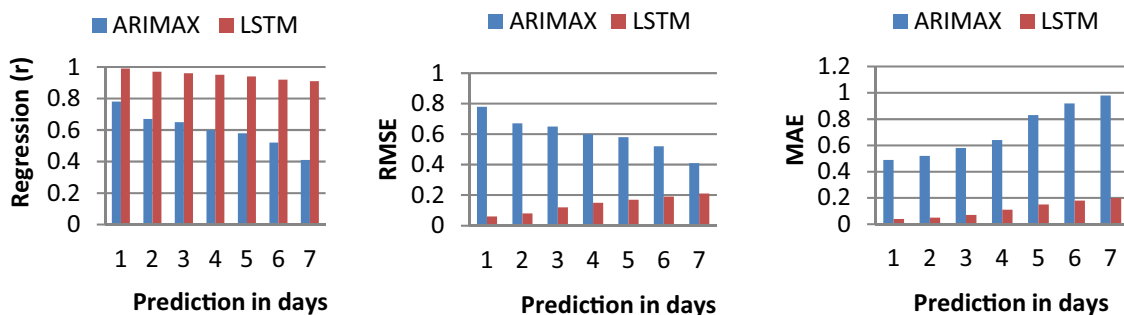


Fig. 6 Performance of ARIMAX and LSTM in terms of *r*, RMSE and MAE for the daily forecasts at site A

4.4 Real-time forecasting using LSTM network

The real-time predictive ability of LSTM network is then analyzed in terms of observed SSTs. The LSTM model is used to predict future SSTs for a period of 170 days starting from the January 1, 2019, to May 19, 2019. The observed versus forecasted SSTs for the different site locations are shown in Figs. 9, 10, 11, 12, 13. The network shows an adequate fit during the testing phase which is reflected by the closeness of the observed and predicted values.

5 Discussions

In most of the past studies, researchers have mainly focused on seasonal predictions but in this work daily, weekly and monthly prediction horizons have been explored. The suitability of nonlinear methods in forecasting site-specific SSTA can be clearly seen from the results obtained. The work done by Tangang et al. [18] was pivotal in this domain as it was one of the first studies where ANN was introduced as an alternative for such climate change predictions. The predictions were done

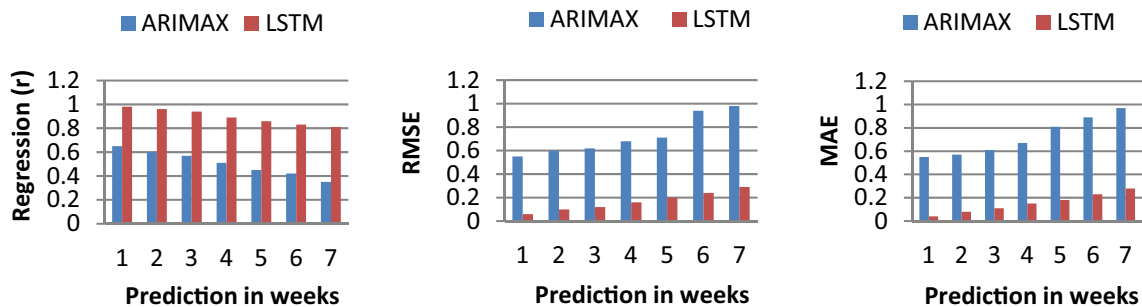


Fig. 7 Performance of ARIMAX and LSTM in terms of *r*, RMSE and MAE for the weekly forecasts at site C

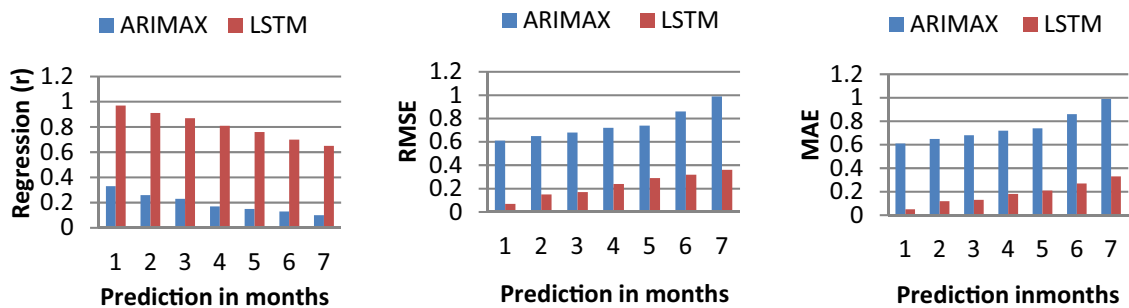


Fig. 8 Performance of ARIMAX and LSTM in terms of *r*, RMSE and MAE for the monthly forecasts at E

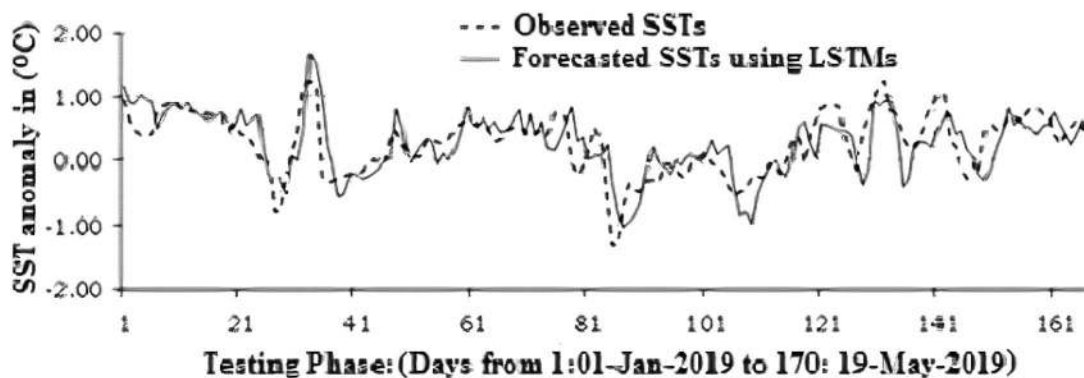


Fig. 9 Observed versus forecasted SSTs at site location A

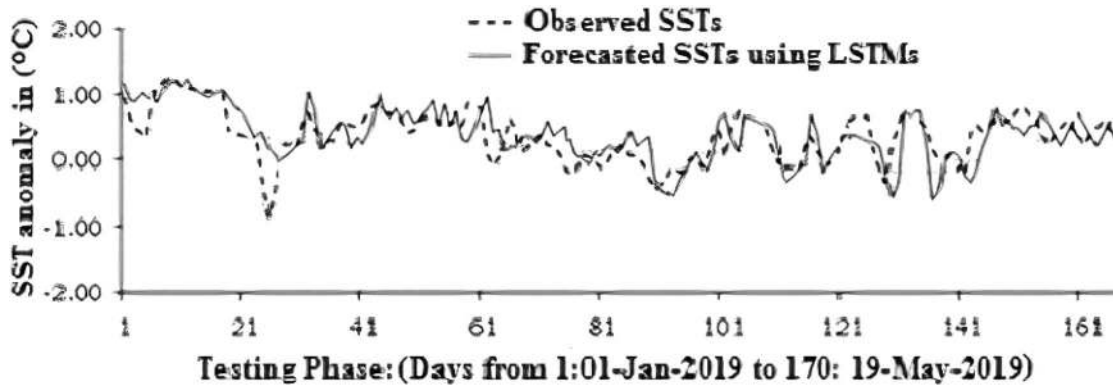


Fig. 10 Observed versus forecasted SSTs at site location B

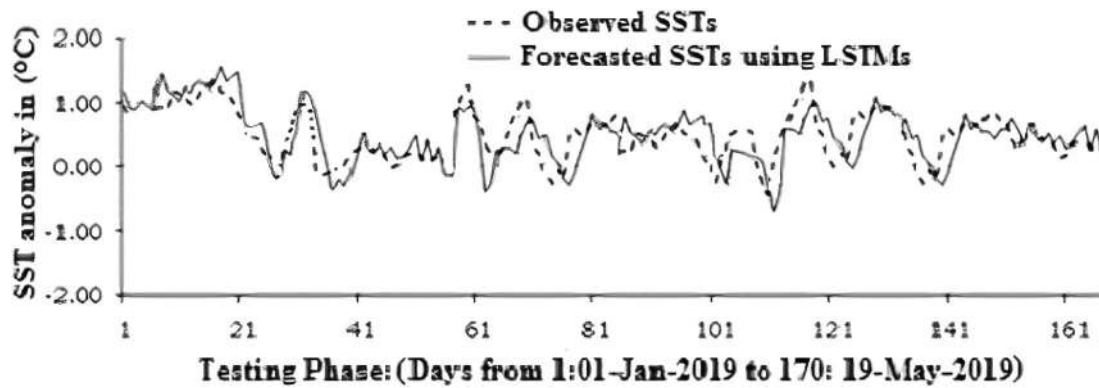


Fig. 11 Observed versus forecasted SSTs at site location C

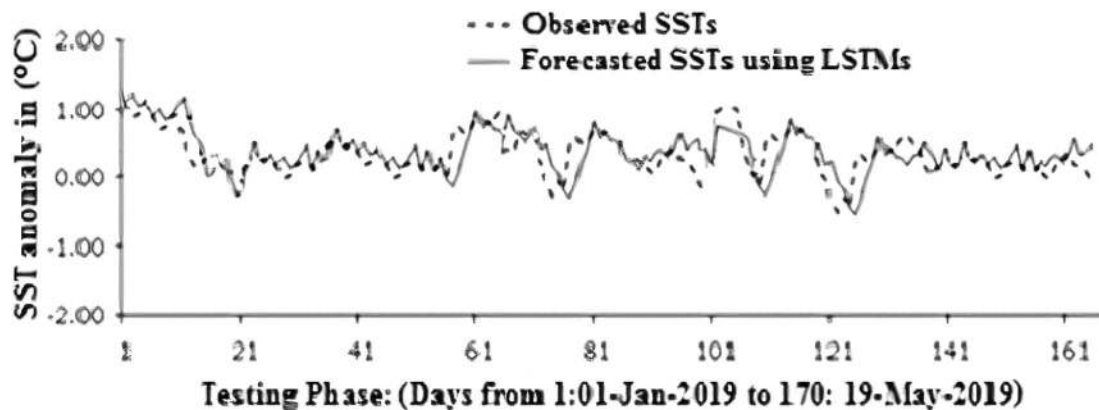


Fig. 12 Observed versus forecasted SSTs at site location D

for 3 months ahead in the future with coefficient of correlation (0.83) values well short of 1.0. Tripathi et al. [19] improved the error statistics but involved traditional NNs for the study. The networks were also much suited for seasonal forecasts. They were not able to forecast for smaller time intervals with greater accuracy. The similar trend was observed in the study of Martinez and Heish

[11] when they forecasted tropical Pacific SSTs. They employed two nonlinear approaches, namely support vector machines (SVMs) and Bayesian NN, though the performance of the models reduced drastically at higher time leads. Mahongo and Deo [10] used a NAR NN for the two locations in the Indian Ocean and achieved correlation values up to 0.85. All the studies mentioned above

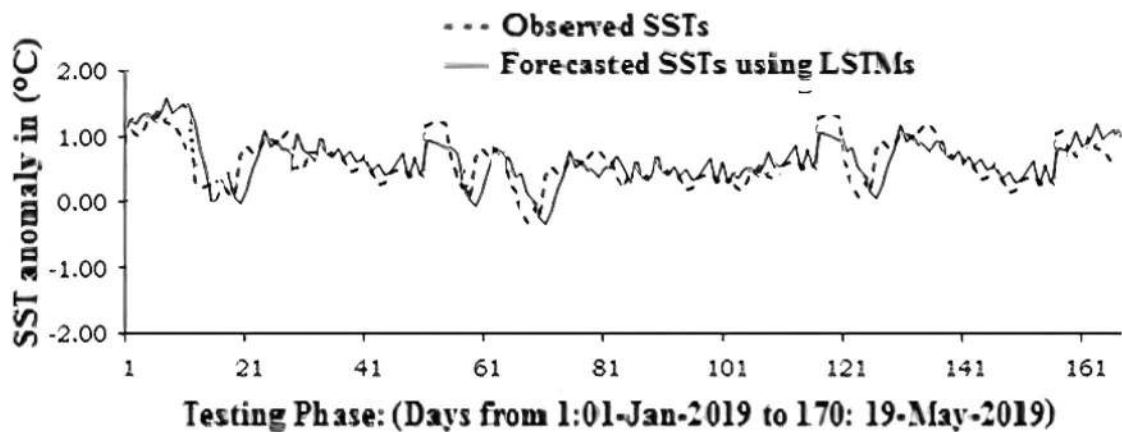


Fig. 13 Observed versus forecasted SSTs at site location E

showed the similar trend of forecast results at higher time intervals. Deep learning networks such as LSTMs can improve on the forecast performances and address the problems faced in the above-mentioned studies. LSTM is insensitive to gap lengths, which makes it capable to forecast at smaller time horizons and at higher time intervals. They are very capable of extracting patterns from the input space where the input data span over long durations. For such networks, sometimes the factor of overtraining becomes very decisive. Over-fitting or over-training leads to loss in generalization capability and can lead to inaccurate forecasts. In the present study, the error statistics show that the model did not encounter the difficulty of over training. The correlational values for all timescales were much higher even at longer time durations. However, for successful training of such networks, a large amount of data are required. Only sufficient data can lead to successful training which in turn can produce accurate test results. In case of lesser data during time series analysis, more importance has to be given to network calibration and small variations in the datasets to achieve meaningful forecasts. This aspect was very similar to our work in the present study. In spite of this limitation, the results we obtained have stood firm when evaluated by different statistical metrics. We can derive that the results obtained depend on the datasets which are polished and incorporated in situ or through satellite-based observations can directly or indirectly affect the final outcome, which can be seen as a drawback in such site-specific studies. Satellite products are usually a combination of different nighttime and daytime satellite observations. Data from buoy, mooring and commercial ships are also added to them. After arranging all data in temporal similarity format, different agencies across the world use optimal interpolation to

come up with suitable gridded data. Parameters like the wind speed and sea level pressure which cause uncertainties in SST measurements are also reflected in the SST daily (diurnal) variations. Therefore, the results could not be superimposed to other homogenous regions. The above-mentioned features will be incorporated during the modeling as well and measuring these SST diurnal variations accurately can only reduce the effect of such uncertainties.

6 Conclusions

Site-specific forecasts of SSTs have widespread applications in different domains, and the commonly applied method of numerical forecast of SSTs shows large deviations when applied for site-specific studies. Moreover, at longer time leads, the accuracy of the numerical estimation drops considerably. A hybrid approach combining both numerical and data-driven methods (NNs) tends to address this drawback and produces improved results. The above-mentioned approach employing the standard and traditional NNs significantly improves the results but still leaves scope for further improvements. The use of deep learning LSTM NN addresses this problem efficiently for all the three timescales. The network produced meaningful real-time forecasts during testing as well. A comparison with a linear system (ARIMAX) also showed that linear models are not much suited for studies for different time horizons and longer duration due to their lack of generalization capability in modeling nonlinear behavior which is a staple of time series prediction. The LSTM NN which is insensitive to gap lengths and has higher data extraction capability from a given input space makes it an attractive

alternative in time series prediction studies. The capability of suggested approach was also assessed to produce real-time SST forecasts in the future. The suggested method requires large amount of experimentation with the architecture and learning parameters to yield successful results.

Acknowledgements This study utilizes datasets from the National Oceanic and Atmospheric Administration (NOAA), U.K. Hadley Centre's sea ice and sea surface temperature and the Canadian Centre for Climate Modeling and Analysis (CCCma).

Code availability The modeling and analysis of the above-mentioned approaches are performed in MATLAB R2018b, specifically the Neural Network toolbox and the Deep Learning toolbox.

Compliance with ethical standards

Conflict of interest The authors hereby declare no conflict of interest in this study.

References

- Agarwal N, Kishtawal CM, Pal PK (2001) An analogue prediction method for global sea. *Curr Sci* 80(1):49–55
- An SI, Jin FF (2001) Collective role of thermocline and zonal advective feedbacks in the ENSO mode. *J Clim* 14:3421–3432
- Box GEP, Jenkins GM (1976) *Time series analysis: forecasting and control*, 2nd edn. Holden-Day, San Francisco, CA, p 553
- Gers FA, Schmidhuber E (2001) LSTM recurrent networks learn simple context-free and context-sensitive languages. *IEEE Trans Neural Netw* 12(6):1333–1340. <https://doi.org/10.1109/72.963769>
- Gupta SM, Malmgren BA (2009) Comparison of the accuracy of SST estimates by artificial neural networks (ANN) and other quantitative methods using radiolarian data from the Antarctic and Pacific Oceans. *J Earth Sci India* 2:52–75
- Haykin S (1998) *Neural networks: a comprehensive foundation*. Prentice Hall PTR, New Jersey, United States, p 842
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hsieh WW (2000) Nonlinear canonical correlation analysis by neural networks. *Neural Netw* 13:1095–1105
- Laepple T, Jewson S (2007) Five-year ahead prediction of sea surface temperature in the tropical Atlantic: a comparison between IPCC climate models and simple statistical methods. *arXiv preprint physics/0701165*
- Mahongo SB, Deo MC (2013) Using artificial neural networks to forecast monthly and seasonal sea surface temperature anomalies in the western Indian Ocean. *Int J Ocean Climate Syst* 4:133–150. <https://doi.org/10.1260/1759-3131.4.2.133>
- Martinez SA, Hsieh WW (2009) Forecasts of tropical Pacific sea surface temperatures by neural networks and support vector regression. *Int J Oceanogr* 2009:1–13. <https://doi.org/10.1155/2009/167239>
- Monahan AH (2001) Nonlinear principal component analysis: Tropical Indo-Pacific Sea surface temperature and sea level pressure. *J Clim* 14:219–233
- Patil K, Deo MC (2017) Prediction of daily sea surface temperature using efficient neural networks. *Ocean Dyn* 67:357–368. <https://doi.org/10.1007/s10236-017-1032-9>
- Patil K, Deo MC, Ghosh S, Ravichandran M (2013) Predicting sea surface temperatures in the North Indian ocean with nonlinear autoregressive neural networks. *Int J Oceanogr* 2013:1–11. <https://doi.org/10.1155/2013/302479>
- Schmidhuber J (2015) Deep learning in neural networks: an overview. *Neural Netw* 61(January):85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>
- Siegelmann HT (1997) Computational capabilities of recurrent NARX neural networks, *IEEE transactions on systems, man, and cybernetics—part b: cybernetics*, vol 27(II)
- Tang B, Hsieh WW, Monahan AH, Tangang FT (2000) Skill comparisons between neural networks and canonical correlation analysis in predicting the equatorial Pacific sea surface temperatures. *J Clim* 13:287–293
- Tangang FT, Hsieh WW, Tang B (1997) Forecasting the equatorial Pacific sea surface temperatures by neural network models. *Clim Dyn* 13:135–147. <https://doi.org/10.1007/s003820050156>
- Tripathi KC, Das ML, Sahai AK (2006) Predictability of sea surface temperature anomalies in the Indian Ocean using artificial neural networks. *Indian J Marine Sci* 35(3):210–220
- Wasserman PD (1993) *Advanced methods in neural computing*. Van Nostrand Reinhold, New York, p 255
- Wu KK (1994) *Neural Networks and Simulation Methods*. Marcel Dekker, pp.456.
- Wu A, Hsieh WW (2002) Nonlinear canonical correlation analysis of the tropical Pacific wind stress and sea surface temperature. *Clim Dyn* 19:713–722
- Wu A, Hsieh WW, Tang B (2006) Neural network forecasts of the tropical Pacific sea surface temperatures. *Neural Netw* 19:145–154. <https://doi.org/10.1016/j.neunet.2006.01.004>
- Xiao C, Chen N, Hu C, Wang K, Gong J, Chen Z (2019) Short and mid-term sea surface temperature prediction using time-series satellite data and LSTM-AdaBoost combination approach. *Remote Sens Environ* 233:111358
- Xue Y, Leetmaa A (2000) Forecasts of tropical Pacific SST and sea level using a Markov model. *Geophys Res Lett* 27(17):2701–2704

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.