IJCNN 2019. International Joint Conference on Neural Networks. Budapest, Hungary. 14-19 July 2019

ENSO Forecasting over Multiple Time Horizons Using ConvLSTM Network and Rolling Mechanism

Bin Mu
School of Software Engineering
Tongji University
Shanghai, China
binmu@tongji.edu.cn

Cheng Peng
School of Software Engineering
Tongji University
Shanghai, China
pengcheng@tongji.edu.cn

Shijin Yuan*
School of Software Engineering
Tongji University
Shanghai, China
yuanshijin2003@163.com

Lei Chen
Shanghai Central Meteorological Observatory
Shanghai, China
qqydss@163.com

***Abstract*—El Niño Southern Oscillation (ENSO) event is characterized by sea surface temperature (SST) anomalies in the tropical Pacific and mainly identified with Oceanic Niño Index (ONI). ENSO forecasting is very challenging owing to the existence of predictability barrier and chaos of climate variability. Recently, machine learning approaches have received considerable attention besides conventional numerical models for this task. However, these existing works mostly focus on investigating single ONI data, neglecting the spatial and temporal dependencies of SST data, and the skill of predictions reduce significantly beyond a lag time of 6 months. With the goal of capturing the spatial and temporal dependencies of SST simultaneously and improving the skill of prediction over longer time horizon, we propose ConvLSTM-RM model, which is a hybrid of convolutional LSTM and rolling mechanism, and use it to build an end-to-end trainable model for ENSO forecasting problem in this paper. Specifically, ENSO forecasting is formulated as a spatiotemporal sequence forecasting problem in which both the input and the output are SST sequences, and ONI can be acquired based on the output. Experiments on historical SST dataset demonstrate that ConvLSTM-RM outperforms seven well-known methods over multiple time horizons (6-, 9- and 12-month). We also apply this model to predict the latest ENSO event during 2015/2016. The results show that our model gets a relatively reliable prediction both from SST grid maps and ONI.**

***Keywords*—Convolutional LSTM, Rolling Mechanism, Spatiotemporal Sequence Forecasting, ENSO**

I. Introduction

Approximately every 4 years, the sea surface temperature (SST) is higher than average in the tropical Pacific. This phenomenon is called El Niño [1]. Operationally[1], an El Niño event is said to occur when the 3-month running mean of SST anomalies over the Niño3.4 region (5°N-5°S, 170°W-120°W, Fig. 1), known as Ocean Niño Index (ONI), is above the threshold of +0.5°C for at least 5 consecutive months. El Niño is the ocean phenomenon of the large-scale ocean-atmosphere interaction between the equatorial Pacific and the global atmosphere [2], referred to as the El Niño Southern Oscillation (ENSO). ENSO forecasting from ocean is to forecast El Niño event. As the most important driver of climate variability, ENSO can trigger disasters in various parts of the globe [3]. Since the tremendous influences and the inner oscillation mechanism of ENSO is still hard to well described, the ENSO forecasting problem is quite challenging and has emerged as a hot research topic in the meteorology community [4].

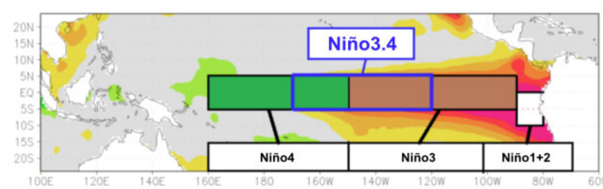

Fig. 1. Niño3.4 region (5°N-5°S, 170°W-120°W), other regions over tropical Pacific (Niño1+2, Niño3 and Niño4) are interconnected with this region.

Existing methods for ENSO forecasting can roughly be categorized into two classes, namely, numerical weather prediction (NWP) based methods and statistical methods.

For the former, NWP models use partial differential equations which are based on the laws of physics to predict ENSO [5]. However, improving skill of numerical prediction is affected by many complicated factors, such as sub-grid parameterization, and this kind of work is very time and effort consuming. Meanwhile, ENSO is not predicted well enough up to 6 months ahead due to the existence of the so-called predictability barrier [6, 7].

For the latter, conventional statistical methods analyses ENSO related data from statistical perspective [8], but these approaches lack strong nonlinear mapping ability to extract inherent characteristics from large amounts of data. Machine learning provides a prominent way to properly handle these issues. In essence, ENSO forecasting is a spatiotemporal sequence forecasting problem with the past ENSO related data (e.g., SST grid maps) as input to predict the future evolution.

[1] http://www.cpc.ncep.noaa.gov/

978-1-7281-2009-6/$31.00 ©2019 IEEE paper N-19743.pdf

However, such a learning problem is nontrivial in the first place due to the high dimensionality of spatiotemporal sequence, unless the spatiotemporal information of ENSO related data is well described by prediction model. Moreover, building an effective prediction model is even more challenging due to the chaotic of climate system.

Recent advances in deep learning, especially deep learning approaches for spatiotemporal sequence forecasting [9], provide some useful insights on how to analysis this problem. Sufficient data and suitable model are considered as two essential points of applying machine learning methods effectively. Specifically, The ENSO forecasting problem satisfies the data requirement as it is easy to collect a huge amount of historic data. What is needed is an appropriate model for end-to-end learning. The key of this part is how to capture the spatial and temporal information simultaneously in one learning model. This task is very challenging as it is affected by the following two complex factors (we take SST data as example for illustration):

(1) Spatial dependencies. The SST of one region (one grid point of Fig. 1) is affected by nearby region as well as distant region. Likewise, the change of this region would affect other region due to the complicated interactions of global climate system.

(2) Temporal dependencies. The SST in a region is affected by recent time intervals, both near and far. More specifically, the evolution of SST is effected by the near month, and is also modulated by the interannual climate variations. Furthermore, for ENSO forecasting problem, there is a strong demand to offer skillful predictions at different time scales (e.g., 6~12 month ahead), while multi-step forecasting is still an open challenge in time series forecasting tasks [10].

To tackle these challenges, we propose ConvLSTM-RM model, which integrates convolutional long short-term memory (ConvLSTM) and rolling mechanism (RM), to predict ENSO over different time horizons (6-, 9- and 12-month). Our contributions are three-fold:

- We formulize the ENSO forecasting problem as a spatiotemporal sequence forecasting problem. In particular, we view the SST grid data as 2-D grid maps based on longitude and latitude, and a grid point donates a region of the Niño3.4 region. In this way, the raw data is converted to a series of image-like matrices which contain rich physical information before being fed into the learning algorithm.
- We propose ConvLSTM-RM, a hybrid of convolutional LSTM and rolling mechanism. ConvLSTM-RM can capture the spatial and temporal information of SST data effectively and provide a better forecasting result over longer time horizon. ConvLSTM-RM is not limited to ENSO forecasting problem and it is applicable to other spatiotemporal sequence forecasting tasks.
- We evaluate our model in real SST dataset with 6-, 9- and 12-month ahead respectively. The results demonstrate that our model outperforms all other neural network models and conventional statistical models. We also apply the model to predict the latest ENSO event, which occurred over 2015/2016, from the generated SST grid maps and ONI. The results show that our model provide a more reliable forecasting compared with NWP models.

The remainder of the paper is structured as follows. Section 2 briefly introduces the related work. Section 3 formulizes the ENSO forecasting problem and describes ConvLSTM-RM model. Experiments and result analysis are shown in Section 4. Finally, Section 5 concludes the paper and describes the future work.

II. Related Work

A. *Machine Learning for ENSO Forecasting*

ENSO is related to dynamics and thermodynamics of air-sea coupled system, which contains diverse physical data. More and more researchers have tried to apply machine learning methods for ENSO forecasting problem. Those studies can be broadly categorized into two types: (1) Conventional machine learning methods and related variants. Silvestre et al. [11] used neural network and support vector regression to forecast SST from 3 to 15 months. Feng et al. [12] applied the neural network and climate network theory to predict ONI. Nooteboom et al. [13] predicted the ONI with 6-, 9- and 12-month ahead by integrating neural network and auto-regressive integrated moving average (ARIMA) model. (2) Deep learning methods, especially recurrent neural networks (RNN) have been applied for this problem. McDermott et al. [14, 15] used two improved RNNs, ensemble quadratic echo state network (ensemble QESN) and Bayesian spatiotemporal recurrent neural network (BAST-RNN), to predict ONI 6 months ahead. Zhang et al. [16] applied LSTM to forecast SST over coastal seas of China. All those papers show that machine learning is an effective tool to handle temporal relations among ENSO related data.

However, most of the researches above-mentioned focused on investigating single ONI data and only modeled the temporal process. As a spatiotemporal process, ENSO also contains much spatial information, which can be used to study the SST characteristics further. However, not much research has taken spatial correlation into consideration to the best of our knowledge.

B. *Spatiotemporal Modeling with Deep Learning*

From a higher-level perspective, ENSO forecasting is intrinsically a spatiotemporal sequence forecasting problem in which both the input and output are spatiotemporal sequences. Many deep learning based models have been proposed to tackle this topic, and those models have achieved good performance in many real problems like video prediction [17, 18, 19], precipitation nowcasting [20, 21], traffic speed forecasting [22, 23, 24] and crowd flows prediction [25, 26]. More details can be found in a recent survey paper [9] and the references therein.

In general, those deep learning models based on two major challenges: (1) how to learn a model for multi-step forecasting; (2) how to adequately model the spatial and temporal structures. For the former, there exist two basic learning strategies: Direct Multi-step (DMS) and Iterative Multi-step (IMS) [27]. As an effective practice of IMS, rolling mechanism is preferred for long-term forecasting, which meets the forecasting requirement of ENSO. For the latter, experiments show that ConvLSTM has strong representational power to make predictions in complex dynamical systems [20], which indicates its great potential to handle spatiotemporal dependencies of ENSO.

III. Methodology

A. Formulation of ENSO Forecasting Problem

The goal of ENSO forecasting is to use the previously observed SST grid maps to forecast a variable length of the future grid maps of Niño3.4 region, the output length here depends on actual forecast demands (e.g., 6-, 9- or 12-month ahead). In real application, the grid maps are usually recorded from satellites and reconstructed with analysis schemes [28]. In this study, we partition the whole Niño3.4 region into an $I * J$ grid map based on the longitude and latitude where a grid denotes a region, as shown in Fig. 2.

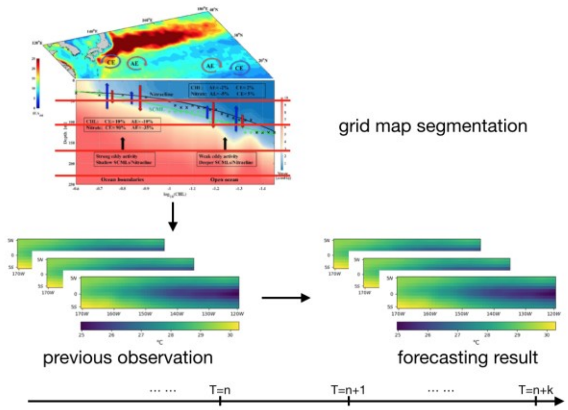

Fig. 2. Formulation of ENSO forecasting problem.

If we record the observations periodically, we will get a sequence of tensors $\hat{X}_1, \hat{X}_2, \dots, \hat{X}_t$. The spatiotemporal sequence forecasting problem is to predict the most likely length-K sequence in the future given the previous J observations which include the current one:

$$\hat{X}_{t+1}, \dots, \hat{X}_{t+K} = \underset{\hat{X}_{t+1}, \dots, \hat{X}_{t+K}}{\arg\max}\, p\left(\hat{X}_{t+1}, \dots, \hat{X}_{t+K} \mid \hat{X}_{t-j+1}, \dots, \hat{X}_t\right) \quad (1)$$

It is worth noting that comparing with conventional video prediction tasks, the data of a point here contains particular physical significance rather than a single grayscale value, and further researches can make much of the output (e.g., calculate the ONI).

B. ConvLSTM-RM Model

We now present our ConvLSTM-RM model. The key of this part is how to capture the spatiotemporal dependencies simultaneously into one forecasting model and improve the prediction skills over longer time horizon. For temporal modeling, fully connected LSTM (FC-LSTM) structure [29] shows great potential in capturing long-term dependencies over different sequence data, and it is built by LSTM cells (Fig. 3). The main equations of LSTM are shown in (2), where '∘' denotes the Hadamard product. We can take LSTM cell as the SST temporal accumulator in our model.

$$\begin{aligned}
i_t &= \sigma(W_{xi}X_t + W_{hi}H_{t-1} + W_{ci} \circ C_{t-1} + b_i) \\
f_t &= \sigma(W_{xf}X_t + W_{hf}H_{t-1} + W_{cf} \circ C_{t-1} + b_f) \\
C_t &= f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc}X_t + W_{hc}H_{t-1} + b_c) \\
o_t &= \sigma(W_{xo}X_t + W_{ho}H_{t-1} + W_{co} \circ C_t + b_o) \\
H_t &= o_t \circ \tanh(C_t)
\end{aligned} \quad (2)$$

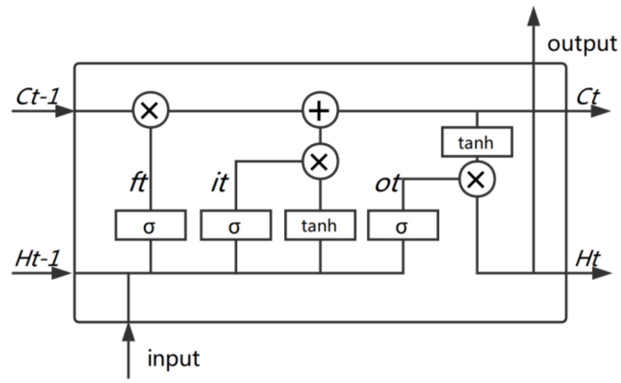

Fig. 3. Structure of LSTM cell (*ft*, *it* and *ot* represent forget gate, input gate and output gate respectively).

SST data is spatial-correlated. Intuitively, the SST of nearby region may affect each other, which can be effectively handled by the convolutional neural network (CNN), which has shown its powerful ability to hierarchically capture the spatial structural information [30]. The usage of full connection in FC-LSTM, which does not encode spatial information, can be replaced by convolution operation, then form the ConvLSTM structure. The key equations of ConvLSTM are shown in (3), where '*' denotes the convolution operation:

$$\begin{aligned}
i_t &= \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i) \\
f_t &= \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + b_f) \\
C_t &= f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \\
o_t &= \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \circ C_t + b_o) \\
H_t &= o_t \circ \tanh(C_t)
\end{aligned} \quad (3)$$

With the convolution operation, ConvLSTM can determine the future of a certain SST grid by the inputs and the past state of its local neighbors (Fig. 4), so the spatial information between different regions is encoded into our model.

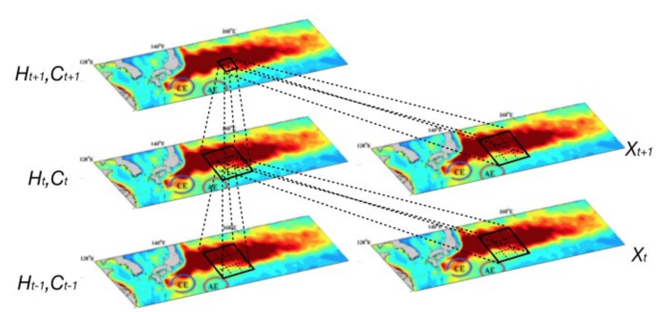

Fig. 4. ConvLSTM structure, modeling the spatial and temporal information of SST collectively.

Like many other network structures (e.g., FC-LSTM), ConvLSTM can be adopted as a basic building block for deeper structure, and it is essential to build a network with many layers in our ENSO forecasting problem for the following two reasons: (1) Owing to interactions of global climate system, we should not only model the neighbor spatial relation, but also capture the spatial dependency of any region. However, one convolution only accounts for spatial near dependencies, limited by the size of their kernels; (2) ENSO is a periodicity phenomenon, so both the short time-span and long time-span temporal information need to be considered into the whole learning model. To handle these two problems, we use the structure shown in Fig. 5 which consists of two networks, an encoding network and a forecasting network. Like in [20], the encoding network compresses the whole input sequence into a hidden state tensors while the forecasting network unfolds this hidden state to provide the final prediction. Both networks are formed by stacking several ConvLSTM layers. Since the network has multiple stacked ConvLSTM layers, it has strong nonlinear mapping ability which makes it suitable to provide reliable prediction in complex ENSO forecasting problem.

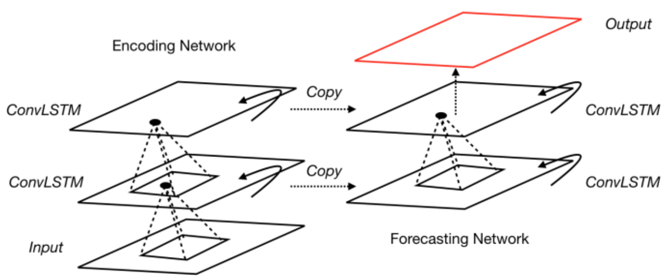

Fig. 5. Encoding-Forecasting model.

Unlike many single-step forecasting problems, ENSO forecasting only makes sense over relatively long time horizon. As direct multi-step forecasting is often unstable and more difficult to train [9, 27], we need a more suitable learning strategy to improve prediction skills of longer time horizon. Previous researches show that RM is preferable to handle chaotic data and it is an efficient techniques to increase forecasting accuracy [31, 32]. ENSO forecasting problem meets the data requirement and the application scenario.

The main idea of RM process is utilize the most recent data for future prediction. We can group the SST data as the set of new inputs via:

Where n = 1, 2,…, $Q_{(t-1)}$, …, $Q_{(t-n)}$ and $Q_{(t)}$ are the input and output data respectively. The symbol n denotes the lagged values of SST time series. Fig. 6 shows the difference between RM forecasting and direct forecasting.

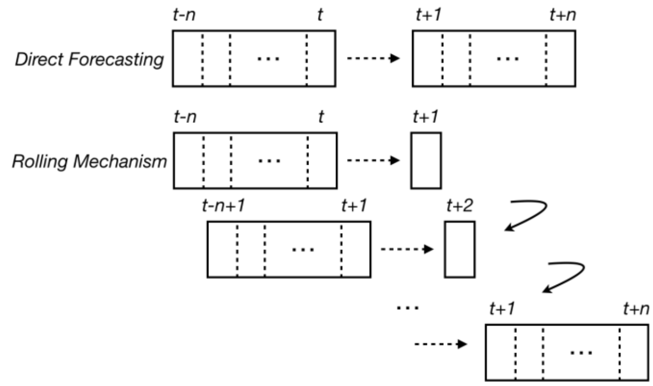

Fig. 6. Rolling mechanism and direct forecasting.

$$\{Q_{(t-1)}, Q_{(t-2)}, \dots, Q_{(t-n)}\}\ ConvLSTM\ \rightarrow Q_{(t)} \tag{4}$$

Fig. 7 illustrates the schematic structure of whole ConvLSTM-RM model. It consists of 4 parts, data input, data preprocess, model and data output. The historical SST data comprise of measurements taken at multiple time horizons (6-, 9-, 12-month). The data preprocess is then performed in 2 parts, including the initial normalization and RM. As a normal process in data-driven modeling, the scaling of the input variables is applied prior to the forecasting in order to avoid the attributes denoted with larger numeric ranges dominating the smaller numeric range [33, 34]. Therefore, the input SST data x^* normalized in the range [0, 1] as follows:

$$x^* = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{5}$$

Then we put our data into neural network model. As mentioned above, to extract much spatiotemporal information as possible, we apply multiple stacked ConvLSTM layers, and use ReLU as the activation function between each layer. To retain more raw information, we do not use sub-sampling [35]. A Conv3D layer is applied to adjust the size of regression output. We get a grid map that indicates a one-step prediction each time, then apply RM to get the next grid map iteratively. Finally, we concatenate those generated grid maps to form a sequence as the data output.

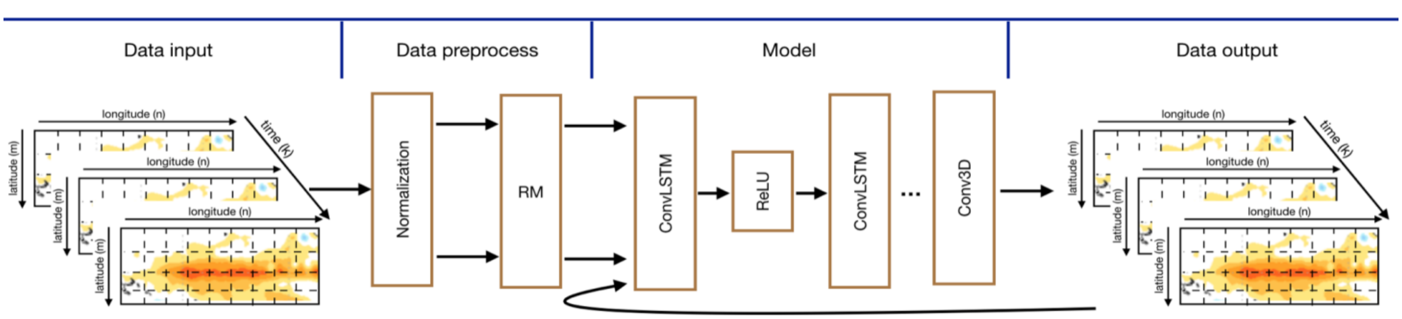

Fig. 7. Architecture of ConvLSTM-RM model.

TABLE I. PERFORMANCE COMPARISON OF DIFFERENT MODELS

T	Metric	HA	ARIMA	SVR	FNN	CNN	FC-LSTM	ConvLSTM	ConvLSTM-RM
6-month	RMSE	1.555	1.300	2.056	1.261	1.048	1.341	0.947	**0.729**
	MAE	1.271	1.053	1.767	0.860	0.777	1.004	0.749	**0.555**
	MAPE	4.78%	3.95%	6.44%	3.15%	2.96%	3.85%	2.72%	**1.45%**
9-month	RMSE	1.506	1.314	2.056	1.248	1.147	1.313	0.976	**0.807**
	MAE	1.224	1.051	1.791	0.997	0.920	0.981	0.769	**0.605**
	MAPE	4.59%	3.95%	6.54%	3.73%	3.38%	3.75%	2.86%	**2.27%**
12-month	RMSE	1.251	1.158	2.119	1.295	1.039	1.079	1.033	**0.789**
	MAE	0.969	0.905	1.882	1.034	0.801	0.814	0.805	**0.607**
	MAPE	3.64%	3.39%	6.86%	3.86%	3.00%	3.06%	2.98%	**2.25%**

IV. EXPERIMENTS

A. Experiment Settings

We conduct experiments on the real-world monthly (1850.01-2015.12) SST grid dataset[2], which covers the whole Niño 3.4 region with 1° latitude multiplied by 1° longitude. To evaluate the prediction performance over multiple time horizons, we construct different sliding windows with length of 6-, 9- and 12-month as the input sequences respectively. 80% of data is used for training, 10% is used for testing while the remaining 10% for validation. To evaluate the performance of our model, we adopt 3 commonly used metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). The metrics are defined as follow, and all of them are the smaller the better.

$$\mathrm{RMSE}(\boldsymbol{x}, \widehat{\boldsymbol{x}}) = \sqrt{\frac{1}{|\boldsymbol{\Omega}|}\sum_{i\in\boldsymbol{\Omega}}(x_i - \widehat{x_i})^2} \tag{6}$$

$$\mathrm{MAE}(\boldsymbol{x}, \widehat{\boldsymbol{x}}) = \frac{1}{|\boldsymbol{\Omega}|}\sum_{i\in\boldsymbol{\Omega}}|x_i - \widehat{x_i}| \tag{7}$$

$$\mathrm{MAPE}(\boldsymbol{x}, \widehat{\boldsymbol{x}}) = \frac{1}{|\boldsymbol{\Omega}|}\sum_{i\in\boldsymbol{\Omega}}\left|\frac{x_i - \widehat{x_i}}{x_i}\right| \tag{8}$$

In the following experiments, we first compare different forecasting models over different prediction horizons (6-, 9- and 12-month), then discuss the effects of different hyperparameter settings of ConvLSTM-RM model to the final result. Finally, we take the ENSO event during 2015/2016 as the case, investigating the effectiveness of model both from the generated SST grid maps and ONI. All neural network based approaches are implemented using Keras, and the source code is available at GitHub[3]. We run all the experiments on a computer with two NVIDIA 1080Ti GPUs.

B. Performance Comparison of Different Models

We compare ConvLSTM-RM network with other widely used time series regression models, including (1) HA: Historical Average, which models the development of SST as a periodic variation, and uses the historical observation as the prediction; (2) ARIMA: Auto-Regression Integrated Moving Average, which is a well-known model for understanding and forecasting future; (3) SVR: Support Vector Regression, which applies linear support vector machine for the regression task. The following deep neural network based models are also included: (4) Feed forward Neural Network (FNN); (5) Convolutional Neural Network (CNN); (6) Recurrent Neural Network with fully connected LSTM hidden units (FC-LSTM); (7) ConvLSTM network without RM (ConvLSTM). Those deep neural networks all with roughly same amount of trainable parameters, and they are fully trained with the same number of epochs (e.g., 10, 000 epochs) for fair comparison. Adam optimizer [36] is applied during training process.

Table 1 shows the comparison of different models for 6-, 9- and 12-month ahead forecasting. We observe the following phenomena on this process: (1) ConvLSTM-RM outperforms all other baselines regarding all the metrics for all forecasting horizon, which suggests its effectiveness of handling spatiotemporal dependencies over long time horizon. (2) Deep neural network based methods, especially CNN, ConvLSTM and ConvLSTM-RM, tend to have a better performance than other baselines. One intuitive reason is that the development of SST is irregular and highly spatial-correlated, so it is hard for a model to give accurate predictions on test set without learning the inner dynamics development of the climate system. (3) The performances of different models show no obvious consistency with the increase of forecasting horizon. The reason may be that the most uncertainties of ENSO prediction are caused by the period of March-April-May, which is included in all experiments.

C. Comparison of Different ConvLSTM-RM Structures

To further investigate the impact of different network structures to the final result and to figure out the best network setting, we compare ConvLSTM-RM from the following two aspects: (1) Number of layers, which is a fundamental setting for deep neural network structure. (2) Number of kernels and kernel size, which can represent attributes of spatial information collector.

[2] https://www.esrl.noaa.gov/psd/data/gridded/data.cobe2.html
[3] https://github.com/KrisCheng/Deep4Cli

Fig. 8 shows the comparison between different model layers. We find that: (1) deeper models can produce better results with fewer parameters; (2) More layers does not always reach better performance, the RMSE of results first decrease gradually and then increase over all different time horizons. The MAE and MAPE also show the same tendency. Actually, experiments in previous studies show that the layer is not the more the better, and the training data is not sufficient enough for so many parameters. 4 layers can be considered an appropriate choice in this experiment.

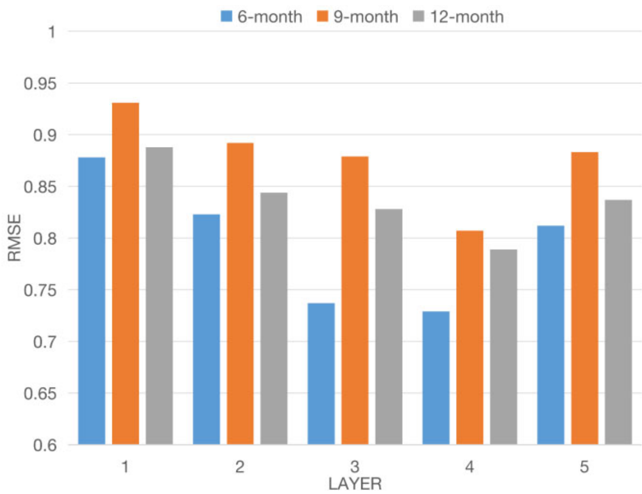

Fig. 8. RMSE of different layers with 6-, 9- and 12-month ahead.

Next, we explore the choice of different number of kernels and kernel size. Fig. 9 shows the performance of different kernel number with a 4 layers network. In general, more kernels enable the model to capture multiple representations of spatial dependencies, and each kernel can be considered as one representation of the entire spatial information. However, with more kernels, the cost of learning complexity increasing rapidly, and over-fitting problem occurs. We find that with the increase of kernel number, the error on the testing dataset first quickly decrease, and then slightly increase. 8 kernels of each layer perform best with 6- and 12-month ahead while 16 kernels of each layer perform best with 9-month ahead.

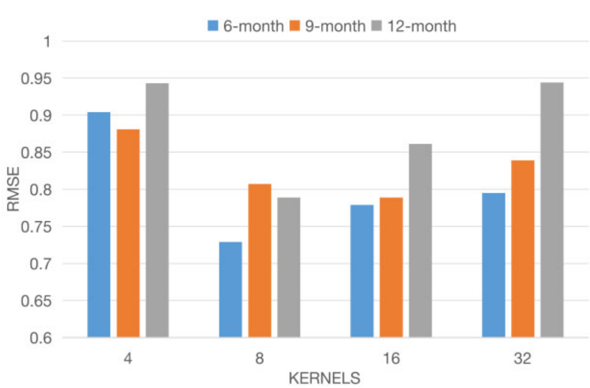

Fig. 9. RMSE of different kernels with 6-, 9- and 12-month ahead.

With the fixed layer and kernel number, we go deep to study the influence of different kernel size, which can be viewed as the volume of spatial correlation collector. Larger kernel size enables the model to capture broader spatial dependency over the Niño3.4 region. Nevertheless, the experiment result (Fig. 10) shows that kernel size is also not the bigger the better. The kernel size of 3*3 performs best with 12-month lead, while the kernel size of 4*4 works best with 6- and 9-month lead. When kernel size is 1*1, no spatial information is encoded into the model. We have also tried other methods to reduce the loss on validation set (e.g., dropout, add batch normalization layers [37]), but the performance of final result has not improved significantly.

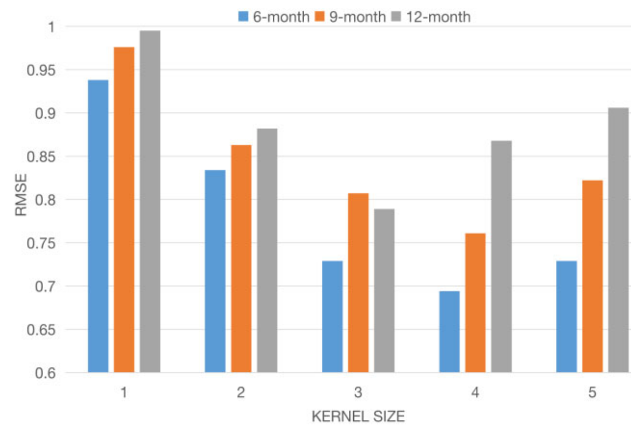

Fig. 10. RMSE of different kernel size with 6-, 9- and 12-month ahead.

D. *Effectiveness of Model on ENSO Event Over 2015/2016*

In order to evaluate the performance of our model on real ENSO event and better understand the model behavior, we take the recent ENSO event occurred during 2015/2016 as the case, which is considered as one of the most extreme ENSO since records began [4]. The fluctuation of SST during this period is a suitable candidate to test the effectiveness and robustness of our model. This period is not included in the training process.

Fig. 11 shows the SST grid maps generated by our model along with the ground truth from 2015.01 to 2015.12. We have the following observations: (1) From the perspective of SST development, the warmer region (yellow part of Fig. 11) was first generated near 160°W, then reduced in the following months. (2) The cooler region (blue part of Fig. 11) generated near 140°W in May, then diffused westward gradually, but the kernel (deep blue part) still stayed in the east region. Our prediction patterns show this same tendency in general.

To quantify the prediction result further. We calculate the ONI with a regressive operation during this period, and compare it with prediction of numerical model from IRI CPC[4]. Fig. 12 shows the development of ONI from MAM (March-April-May) to NDJ (November-December-January). It is worth mentioning that IRI CPC only provides a 9-month ahead forecasting, so we extract a 9-month prediction during the peak value of ENSO 2015/2016 for fair comparison. We can see that even though our model indicates a higher deviation at the beginning, but quickly captures the increasing trend of ONI and

[4] https://iri.columbia.edu/our-expertise/climate/enso/

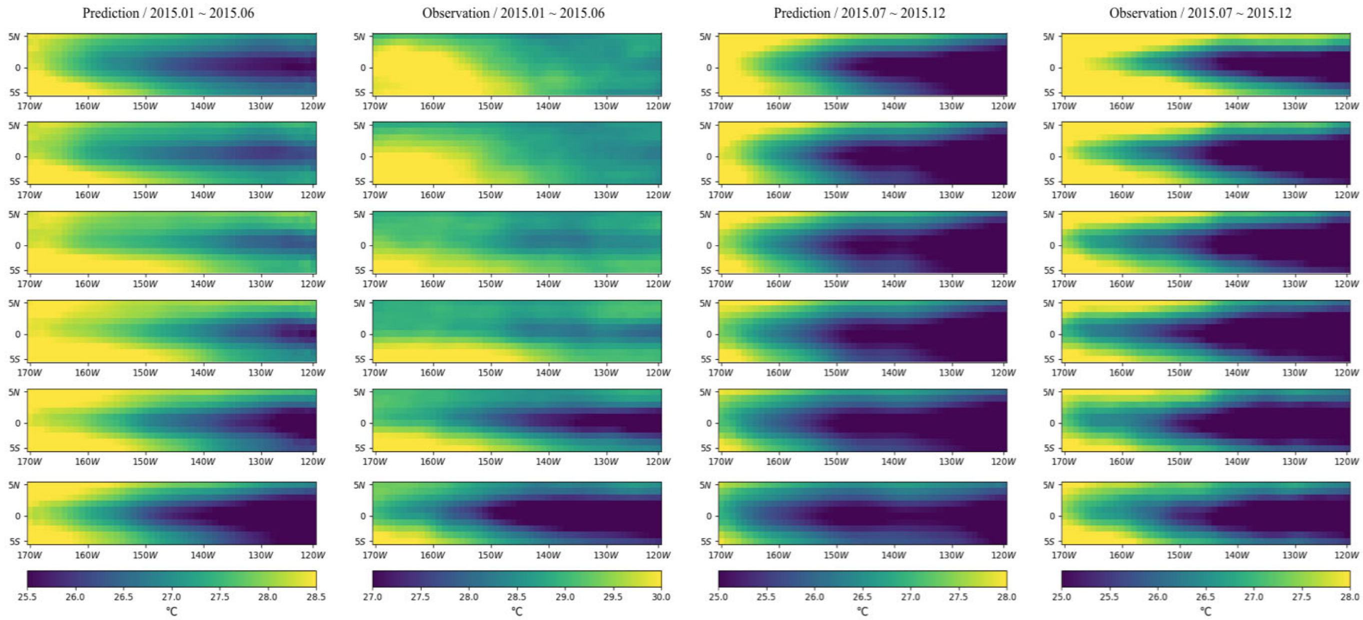

Fig. 11. Visualization of SST pattern over Niño3.4 region, one map stands for a monthly average SST distribution. Four columns left to right are respectively the prediction from 2015.01 to 2015.06, the observation from 2015.01 to 2015.06, the prediction from 2015.07 to 2015.12 and the observation from 2015.07 to 2015.12.

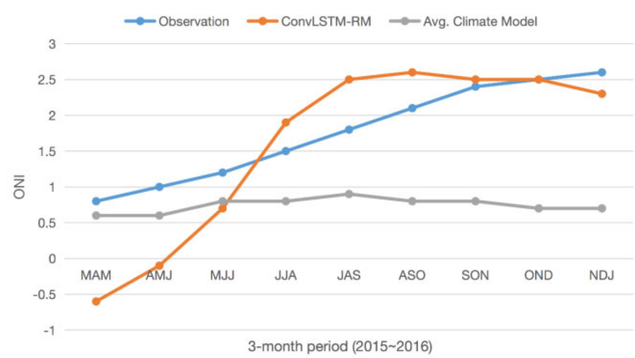

Fig. 12. ONI from MAA to NDJ of 2015, the horizontal ordinate represents the 3-month running mean.

forecasts the peak value of ENSO with a relatively high accuracy compared with climate models. As the ENSO over 2015/16 is an abnormal climatic event over the past three decades, it is reasonable that almost all the climate models underestimate the peak value. However, our model performs fairly reasonable in this case, both from the generated SST grid maps and the ONI, and can be considered as a reliable prediction of ENSO in this period.

V. Conclusion and Future Work

In this paper, we explore a new data-intensive direction for ENSO forecasting problem, which involving spatial and temporal dependencies collectively. Specifically, we formulate the ENSO forecasting as a spatiotemporal sequence forecasting problem. By combining convolutional LSTM and rolling mechanism, we build the ConvLSTM-RM model to capture the spatiotemporal dependencies of ENSO over multiple time horizons. When evaluated on real-world SST dataset, our approach obtained significantly better prediction than baselines.

For future work, we will investigate the following two aspects: (1) constructing a multichannel dataset, which involves more related measurements (e.g., sea level pressure, westerly wind bursts), to model the spatiotemporal process of ENSO more comprehensive; (2) integrating the ConvLSTM-RM approach with different numerical models (e.g., ZC, GFDL CM) for better prediction result.

VI. Acknowledgment

The research is supported by the Fundamental Research Funds for the Central Universities of China in 2017.

References

[1] Wunsch C, "El Nino, La Nina, and the Southern Oscillation," *Science*, 248(4957):904-906, 1990.

[2] Bjerknes, Jakob, "Atmospheric teleconnections from the equatorial Pacific," *Monthly weather review,* 97(3), 163-172, 1969.

[3] J. Ludescher et al, "Very early warning of next El Niño," *Proceedings of the National Academy of Sciences*, vol. 111, (6), pp. 2064-2066, 2014.

[4] A. Santoso, M. Mcphaden and W. Cai, "The Defining Characteristics of ENSO Extremes and the Strong 2015/2016 El Nino," *Reviews of Geophysics*, vol. 55, (4), pp. 1079-1129, 2017.

[5] A. Barnston et al, " Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing?" *Bulletin of the American Meteorological Society*, vol. 93, (5), pp. 631-651, 2012.

[6] W. Duan and C. Wei, "The 'spring predictability barrier' for ENSO predictions and its possible mechanism: results from a fully coupled model," *International Journal of Climatology*, vol. 33, (5), pp. 1280-1292, 2013.

[7] L. Goddard et al, "Current approaches to seasonal to interannual climate predictions," *International Journal of Climatology*, vol. 21, (9), pp. 1111-1152, 2001.

[8] J. Kug et al, "A statistical approach to Indian Ocean sea surface temperature prediction using a dynamical ENSO prediction," *Geophysical Research Letters*, vol. 31, (9), 2004.

[9] Shi X, Yeung D Y. "Machine Learning for Spatiotemporal Sequence Forecasting: A Survey," *arXiv preprint arXiv:1808.06865*, 2018.

[10] S. Ben Taieb et al, "A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition," *Expert Systems with Applications*, vol. 39, (8), pp. 7067-7083, 2012.

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

[12] Feng Q Y, Vasile R, Segond M, et al. "ClimateLearn: A machine-learning approach for climate prediction using network measures," *Geoscientific Model Development*, 2016.

[13] P. D. Nooteboom et al, "Using network theory and machine learning to predict El Niño," *Earth System Dynamics*, vol. 9, (3), pp. 969-983, 2018.

[14] McDermott P L, Wikle C K. "Bayesian Recurrent Neural Network Models for Forecasting and Quantifying Uncertainty in Spatial-Temporal Data, " *arXiv preprint arXiv:1711.00636*, 2017.

[15] P. McDermott and C. Wikle, "An ensemble quadratic echo state network for non-linear spatio-temporal forecasting," *Stat*, vol. 6, (1), pp. 315-330, 2017.

[16] Q. Zhang et al, "Prediction of Sea Surface Temperature Using Long Short-Term Memory," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, (10), pp. 1745-1749, 2017.

[17] Finn, Chelsea, Ian Goodfellow, and Sergey Levine. "Unsupervised learning for physical interaction through video prediction," in *Advances in neural information processing systems*. 2016.

[18] Jia, Xu, et al. "Dynamic filter networks," in *Advances in neural information processing system*. 2016.

[19] Wang Y, Gao Z, Long M, et al. "PredRNN++: Towards A Resolution of the Deep-in-Time Dilemma in Spatiotemporal Predictive Learning," *arXiv preprint arXiv:1804.06300*, 2018.

[20] Xingjian, S. H. I., et al. "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Advances in neural information processing systems*. 2015.

[21] Shi, Xingjian, et al. "Deep learning for precipitation nowcasting: A benchmark and a new model," in *Advances in neural information processing systems*. 2017.

[22] Yu, Rose, et al. "Deep learning: A generic approach for extreme condition traffic forecasting," in *Proceedings of the 2017 SIAM International Conference on Data Mining*. Society for Industrial and Applied Mathematics, 2017.

[23] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," in *International Conference on Learning Representations*, 2018.

[24] Zhang J, Shi X, Xie J, et al. "GaAN: Gated Attention Networks for Learning on Large and Spatiotemporal Graphs," *arXiv preprint arXiv:1803.07294*, 2018.

[25] Zhang, Junbo, et al. "DNN-based prediction model for spatio-temporal data," in *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2016.

[26] Zhang, Junbo, Yu Zheng, and Dekang Qi. "Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction," in *Association for the Advance of Artificial Intelligence*. 2017.

[27] G. Chevillon, "Direct multi-step estimation and forecasting," *Journal of Economic Surveys*, vol. 21, (4), pp. 746-785, 2007.

[28] S. Hirahara, M. Ishii and Y. Fukuda, "Centennial-Scale Sea Surface Temperature Analysis and Its Uncertainty," *Journal of Climate*, vol. 27, (1), pp. 57-75, 2014.

[29] Graves A. "Generating sequences with recurrent neural networks," *arXiv preprint arXiv:1308.0850*, 2013.

[30] Y. Lecun et al, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, (11), pp. 2278-2324, 1998.

[31] Z. Zhao et al, "Using a Grey model optimized by Differential Evolution algorithm to forecast the per capita annual net income of rural households in China," *Omega*, vol. 40, (5), pp. 525-532, 2012.

[32] D. Akay and M. Atak, "Grey prediction with rolling mechanism for electricity demand forecasting of Turkey," *Energy*, vol. 32, (9), pp. 1670-1675, 2007.

[33] K. Hsu et al, "Self-organizing linear output map (SOLO): An artificial neural network suitable for hydrologic modeling and analysis," *Water Resources Research*, vol. 38, (12), 2002.

[34] M. Dehghani et al, "Monthly stream flow forecasting via dynamic spatio-temporal models," *Stochastic Environmental Research and Risk Assessment*, vol. 29, (3), pp. 861-874, 2015.

[35] Jain, Viren, et al. "Supervised learning of image restoration with convolutional networks," in *International Conference on Computer Vision*, 2007.

[36] Kingma D P, Ba J, "Adam: A method for stochastic optimization, " *arXiv preprint arXiv:1412.6980*, 2014.

[37] Ioffe S, Szegedy C, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015