Convolutional GRU Network for Seasonal Prediction of the El Niño-Southern Oscillation

Lingda Wang[1,*]**, Savana Ammons**[2,*]**, Vera Mikyoung Hur**[2,*]**, Ryan L. Sriver**[3,*]**, and Zhizhen Zhao**[1,*]

[1]Department of Electrical and Computer Engineering, University of Illinois Urbana-Champaign, Urbana, IL, 61820, USA
[2]Department of Mathematics, University of Illinois Urbana-Champaign, Urbana, IL, 61801, USA
[3]Department of Atmospheric Sciences, University of Illinois Urbana-Champaign, Urbana, IL, 61820, USA
[*]{lingdaw2, savanaa2, verahur, rsriver, zhizhenz}@illinois.edu

ABSTRACT

Predicting sea surface temperature (SST) within the El Niño-Southern Oscillation (ENSO) region has been extensively studied due to its significant influence on global temperature and precipitation patterns. Statistical models such as linear inverse model (LIM), analog forecasting (AF), and recurrent neural network (RNN) have been widely used for ENSO prediction, offering flexibility and relatively low computational expense compared to large dynamic models. However, these models have limitations in capturing spatial patterns in SST variability or relying on linear dynamics. Here we present a modified Convolutional Gated Recurrent Unit (ConvGRU) network for the ENSO region spatio-temporal sequence prediction problem, along with the Niño 3.4 index prediction as a down stream task. The proposed ConvGRU network, with an encoder-decoder sequence-to-sequence structure, takes historical SST maps of the Pacific region as input and generates future SST maps for subsequent months within the ENSO region. To evaluate the performance of the ConvGRU network, we trained and tested it using data from multiple large climate models, including a pre-industrial simulation spanning approximately 1300 years from the Community Climate System Model version 4 (CCSM4) and a 30-member historical hindcast ensemble during the years 1921-2100 using the NOAA Seamless system for Prediction and EArth System Research (SPEAR) model. We also compare and contrast the prediction skill of the ConvGRU network against existing models. The results demonstrate that the ConvGRU network significantly improves the predictability of the Niño 3.4 index compared to LIM, AF, and RNN. This improvement is evidenced by extended useful prediction range, higher Pearson correlation, and lower root-mean-square error. The proposed model holds promise for improving our understanding and predicting capabilities of the ENSO phenomenon and can be broadly applicable to other weather and climate prediction scenarios with spatial patterns and teleconnections.

Introduction

The El Niño-Southern Oscillation (ENSO) phenomenon over the tropical Pacific region is the most energetic driver of climate variability on the seasonal to interannual timescales[1]. It plays a crucial role on global oceanic and atmospheric dynamics, particularly during its irregular warming (El Niño) and cooling (La Niña) phases. The impacts of ENSO are widespread, leading to anomalous temperature and precipitation patterns on a global scale[2–4], as well as changing extreme and hazardous weather conditions on a regional scale, such as winter to early spring tornado outbreaks in the United States[5], tropical cyclone intensity changes in northwestern Pacific[6], and unusual fire weather in Australia[7] and the United States[8]. Consequently, accurate prediction of sea surface temperature (SST) maps within the ENSO region and its associated Niño indices—for instance, Niño 1+2, 3, 3.4, and 4[9–11]—has become a critical area of research. Reliable ENSO prediction can provide valuable insights for decision-making processes in various sectors, including government agencies, food and insurance industries, and transportation, enabling them to prepare for the associated impacts[12,13].

Prediction models for SST maps within the ENSO region and the associated Niño indices can be broadly classified into two types: dynamical models and statistical models. Dynamical models, such as the North American Multi-Model Ensemble (NMME)[14], are commonly used for seasonal ENSO prediction. However, these model ensembles are computationally intensive, sensitive to initialization conditions, and expensive to run. In contrast, this study focuses on statistical models due to their simplicity and comparable prediction skill to dynamical models. One widely used statistical ENSO prediction model is the linear inverse model (LIM)[15,16], which employs principal components analysis and Markov prediction to approximate trends and predict future states based on empirical orthogonal functions, similar to linear regression (LR)[17]. However, LIM fails to capture nonlinear ENSO dynamics—for instance, surface-subsurface interactions and surface winds[18]—and can lead to underestimation of ENSO variability. Another type of statistical prediction model is based on Lorenz's analog forecasting

(AF)[19]. Initially, AF based models use observed or free-running model data as libraries of states. Predictions are then generated by matching states in the library that are very similar to observed data at prediction initialization, and following the evolution on these so-called analogs. Advantages of AF based models include avoiding expensive and unstable initialization systems and reducing structural model error. The kernel analog forecasting (KAF) model[17,20,21], as a generalization of conventional AF based models, utilizes nonlinear kernels to better capture nonlinearity in ENSO dynamics. Recently, with the development of deep learning techniques, the convolutional neural network (CNN)[22] and long short-term memory (LSTM)[23,24] network have been used for predicting Niño indices, but their prediction skills have not yet been extended to capturing spatial patterns in SST variability within the ENSO region.

In this study, we propose the use of a Convolutional Gated Recurrent Unit (ConvGRU) network, inspired by and modified from the original developments[25–27], to predict SST maps within the ENSO region, along with the Niño 3.4 index as a downstream task. The ConvGRU network has an encoder-decoder sequence-to-sequence (Seq2Seq) structure, with both the encoder and the decoder consisting of multi-layer ConvGRU cells. The encoder compresses the input SST maps of the Pacific region into hidden states across all layers, and the decoder unfolds the hidden states from the encoder to generate predictions within the ENSO region. The ConvGRU cell, a key component of both the encoder and the decoder, incorporates several 2-D convolutional layers. This architecture enables the ConvGRU network to take historical SST maps of the Pacific region as inputs and generate future SST maps of the ENSO region for subsequent months, taking into consideration of spatio-temporal correlation of the SST maps. Moreover, this architecture significantly reduces the number of network parameters while accelerating the training process.

To evaluate the performance of the ConvGRU network, we conduct numerical experiments and compare it against existing models, such as KAF, LIM, Seq2Seq with GRU, and LR, using global climate ensembles and atmospheric reanalysis datasets. These datasets include two SST datasets and one surface air temperature dataset. The comparison results demonstrate that the ConvGRU network achieves significant improvements over the other models in terms of useful prediction range, Pearson correlation (PC), and root-mean-square error (RMSE).

By developing an improved prediction model that accurately captures the complex dynamics and spatial patterns of SST within the ENSO region, this study aims to contribute to better understanding and prediction of ENSO-related climate phenomena. Further research can explore further enhancements to the network architecture and investigate its applicability to other climate-related features and prediction tasks.

The rest of this paper is organized as follows: In preliminaries, we provide an overview of the ENSO region prediction problem and discuss existing models for spatio-temporal sequence prediction. In methodology, we present the ConvGRU network, including the ConvGRU cells and the encoder-decoder Seq2Seq structure, and describe the training process. In results and discussion, we discuss the performance of the ConvGRU network on the ENSO region spatio-temporal sequence prediction task as well as its comparison with existing models.

Preliminaries

ENSO region prediction problem

We address the ENSO region prediction problems, which involves predicting future SST map sequences within the ENSO region of the Pacific, given previously observed gridded SST maps of the Pacific region. Suppose that SST maps of the Pacific region are sampled and averaged monthly on a grid of size $M \times N$, representing an SST map of the Pacific region as a matrix in $\mathbb{R}^{M \times N}$ for a specific month. As monthly records of SST maps of the Pacific region are accumulated, a sequence of such matrices is obtained, $\tilde{\boldsymbol{X}}_1, \ldots, \tilde{\boldsymbol{X}}_t, \ldots (\in \mathbb{R}^{M \times N})$, where t denotes a specific month. Given the previous J-month (referred to as the condition range) observed SST maps of the Pacific region, including the current one, represented as $\tilde{\boldsymbol{X}}_{t-J+1:t} \in \mathbb{R}^{J \times M \times N}$, the ENSO region spatio-temporal sequence prediction problem at month t aims to predict the most likely K-month (referred to as the prediction range) future SST maps within the ENSO region. These predicted maps are denoted as $\hat{\boldsymbol{Y}}_{t+1}, \ldots, \hat{\boldsymbol{Y}}_{t+K} (\in \mathbb{R}^{M \times N})$, abbreviated as $\hat{\boldsymbol{Y}}_{t+1:t+K} \in \mathbb{R}^{K \times M \times N}$. Formally, the problem can be stated as follows:

$$\hat{\boldsymbol{Y}}_{t+1:t+K} = \underset{\boldsymbol{Y}_{t+1:t+K}}{\operatorname{argmax}} \mathbb{P}\left(\boldsymbol{Y}_{t+1:t+K} \mid \tilde{\boldsymbol{X}}_{t-J+1:t}\right). \tag{1}$$

The extent of the ENSO region for the predicted maps can cover the entire Pacific region or any other region within the Pacific, depending on the downstream task, for example, the south Pacific decadal oscillation[28]. In real-world applications, SST maps of the Pacific region are typically sampled and averaged monthly on a latitude-longitude grid of a specific resolution, such as $1° \times 1°$ per latitude-longitude grid cell, and the prediction range spans 12 and 24 months.

The ENSO region prediction problem encompasses a series of downstream tasks, such as predicting Niño indices. It holds potential applications in other climate-related features such as fire weather and drought indices[7,29].

Models for spatio-temporal sequence prediction

There exists a range of approaches for spatio-temporal sequence prediction, including machine learning and traditional statistical models. We categorize these models into autoregressive models and statistical models, highlighting their applicability to weather and climate related tasks.

Autoregressive models

Autoregressive models have found widespread usage in time series prediction problems, including recurrent neural networks (RNNs) such as vanilla RNN[30], LSTM[23], and GRU[31]. One multivariate variant of general-purpose RNN, known as fully-connected RNN (FC-RNN)[25, 32], was among the earliest models employed for spatio-temporal sequence prediction. For instance, FC-LSTM takes vectorized inputs (spatio maps) and utilizes LSTM cells. The main equations for FC-LSTM can be summarized as follows:

$$i_t = \sigma_i(W_{ix}x_t + W_{ih}h_{t-1} + b_i), \qquad \bullet \text{ Input gate}$$
$$f_t = \sigma_f(W_{fx}x_t + W_{fh}h_{t-1} + b_f), \qquad \bullet \text{ Forget gate}$$
$$o_t = \sigma_o(W_{ox}x_t + W_{oh}h_{t-1} + b_o), \qquad \bullet \text{ Output gate}$$
$$\tilde{c}_t = \sigma_{\tilde{c}}(W_{\tilde{c}x}x_t + W_{\tilde{c}h}h_{t-1} + b_{\tilde{c}}), \qquad \bullet \text{ New memory cell}$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, \qquad \bullet \text{ Final memory cell}$$
$$h_t = o_t \odot \sigma_h(c_t), \qquad \bullet \text{ Hidden state}$$

where $\odot$ represents the element-wise product (Hardmard product), and σ is either the sigmoid or tanh function.

However, FC-LSTM has limitations in efficiently capturing spatial correlations. To overcome this drawback, ConvLSTM[25] was introduced, which incorporates 2-D convolutional layers within an LSTM cell. ConvLSTM has been further enhanced with various variants and successors such as TrajGRU[26], CDNA[33], PredRNN[34]. Another approach to autoregressive models for spatio-temporal sequence prediction is based on Transformers[35, 36]. Additionally, autoregressive models have been combined with other techniques, such as graph neural networks[37, 38] and generative models[39], leading to significant achievements in short to medium-range weather prediction and other spatio-temporal sequence prediction applications.

Statistical climate models

Statistical climate models employ statistical techniques tailored for climate-related data analysis and prediction. Examples include LIM[15, 16] and KAF[17, 20, 21].

LIM assumes that the dynamics of a system can be described by a linear stochastic differential equation of the form:

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{B}\boldsymbol{x} + \xi.$$

Here $\boldsymbol{x}(t)$ represents the state of the system at time t, $\boldsymbol{B}$ is a time-independent operator, and ξ is stationary white noise. For stationary statistics, $\boldsymbol{B}$ must be dissipative, meaning that its eigenvalues have negative real parts, and

$$\boldsymbol{C}(\tau) = \boldsymbol{G}(\tau)\boldsymbol{C}(0) \quad \text{and} \quad \boldsymbol{G}(\tau) = \exp(\boldsymbol{B}\tau),$$

where $\boldsymbol{C}(0)$ and $\boldsymbol{C}(\tau)$ are covariances of $\boldsymbol{x}$ at lags 0 and τ, respectively. In prediction problems, $\boldsymbol{G}(\tau)\boldsymbol{x}(t)$ represents the best linear prediction of the state at time $t + \tau$, given the state at time t. The matrices $\boldsymbol{B}$ and $\boldsymbol{G}$ can then be determined as $\boldsymbol{B} = \tau^{-1}\ln(\boldsymbol{C}(\tau)C(0)^{-1})$.

KAF is a generalization of AF[19, 40], incorporating both nonlinear kernel methods and operator-theoretic ergodic theory[41]. By establishing a rigorous connection with Koopman operator theory[42] for dynamical systems, KAF can generate statistically optimal predictions as conditional expectations. KAF is particularly useful when dealing with noisy and partially observed data during prediction initialization.

These models form the foundation for spatio-temporal sequence prediction and have been applied to various weather and climate problems. However, we should notice that i) LIM is not capable of capturing nonlinear dynamics with the ENSO region; ii) KAF is usually applied to Niño index prediction task, and is not capable of predicting the spatial pattern within the ENSO region. Those limitations of statistical models motivate the proposal of the ConvGRU network in this study.

Methodology

We now propose our ConvGRU network, which is inspired by and modified from the ConvGRU model[26, 27], for the ENSO region spatio-temporal sequence prediction. The ConvGRU network incorporates 2-D convolutional layers in both input-to-(hidden) state and (hidden) state-to-(hidden) state transitions within a ConvGRU cell. This modification offers several advantages over
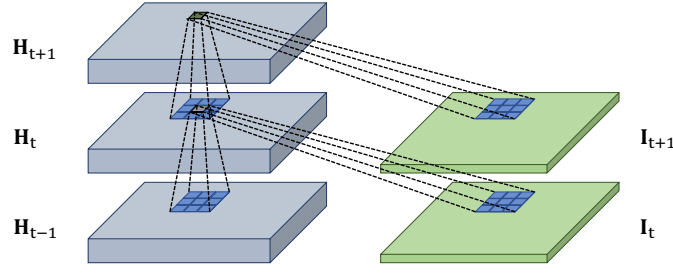
Figure 1. Illustration of 2-D convolutional layers within a ConvGRU cell. Convolutional layers are applied to update gate, reset gate, and new memory cell (see Eq. (2)).

FC-GRU cells, efficiently capturing spatial correlations of SST maps and reducing the number of network parameters. The ConvGRU network is composed of multiple stacked ConvGRU cells and follows an encoder-decoder Seq2Seq structure. During the training process, samples are generated from fixed-length windows with different starting points.

Convolutional GRU cell

If we were to tackle the ENSO region spatio-temporal sequence prediction problem in Eq. (1) using a network built with FC-GRU cells, we would need to vectorize inputs and hidden states before performing matrix multiplication. These steps are essentially equivalent to fully connected layers in a neural network. However, the vectorization and matrix multiplication steps are not required when using ConvGRU cells. Instead, a ConvGRU cell employs 2-D convolutional layers, which offers several advantages, including extracting meaningful spatial correlation features, reducing the number of network parameters, and speeding up the training process.

Following the structure of FC-GRU cells, the equations for the ConvGRU cell can be expressed as follows:

$$
\begin{aligned}
z_t &= \sigma_z(\boldsymbol{W}_{zx} * \boldsymbol{I}_t + \boldsymbol{W}_{zh} * \boldsymbol{H}_{t-1} + b_z), && \bullet \text{ Update gate} \\
r_t &= \sigma_r(\boldsymbol{W}_{rx} * \boldsymbol{I}_t + \boldsymbol{W}_{rh} * \boldsymbol{H}_{t-1} + b_r), && \bullet \text{ Reset gate} \\
\tilde{\boldsymbol{H}}_t &= \sigma_{\tilde{h}}(\boldsymbol{W}_{\tilde{h}x} * \boldsymbol{I}_t + \boldsymbol{W}_{\tilde{h}rh} * (r_t \odot \boldsymbol{H}_{t-1}) + b_{\tilde{h}}), && \bullet \text{ New memory cell} \\
\boldsymbol{H}_t &= (1 - z_t) \odot \tilde{\boldsymbol{H}}_t + z_t \odot \boldsymbol{H}_{t-1}, && \bullet \text{ Hidden state}
\end{aligned}
\tag{2}
$$

where $*$ represents the convolution operator. Here, input $\boldsymbol{I}_t$, hidden state $\boldsymbol{H}_t$, update gate z_t, reset gate r_t, and new memory cell $\tilde{\boldsymbol{H}}_t$ are all 3D tensors, with the last two dimensions representing the spatial dimensions (rows and columns). Figure 1 illustrates the application of 2-D convolutional layers within a ConvGRU cell for both the input-to-(hidden) state and (hidden) state-to-(hidden) state transitions. This allows the future hidden state in a specific grid cell to extract relevant information locally from its neighboring inputs and past hidden states. The size of the neighbors considered by a grid cell is determined by the size of the convolutional kernel. A large kernel is recommended for fast-evolving spatio-temporal sequences, while a small kernel is more suitable for slow-varying sequences.

Encoder-Decoder Seq2Seq structure

We utilize the ConvGRU cells in Eq. (2) as a key component to construct our ConvGRU network for ENSO region spatio-temporal sequence prediction. We recognize this as a Seq2Seq learning problem (see Eq. (1)) that can be effectively addressed using the encoder-decoder Seq2Seq structure[31, 43].

Figure 2 illustrates the architecture of the ConvGRU network for a 3-layer example, although the number of layers can be adjusted based on performance considerations, such as RMSE. The ConvGRU network consists of two main parts: multi-layer encoder and multi-layer decoder.

The encoder, depicted in Figure 2a, utilizes ConvGRU cells and 2-D convolutional layers. Each layer's hidden states are initialized as an all 0-tensor and updated using inputs and the previous hidden states, following Eq. (2). The first layer takes SST maps of the Pacific region as inputs, while the subsequent layers receive the output hidden states from the previous layer. Convolutional layers are applied before each ConvGRU cell, to adjust the number of the input channels and the size of the input spatial dimensions, enhancing feature extraction.

The decoder, depicted in Figure 2b, consists of ConvGRU cells and 2-D deconvolutional layers. A crucial step that connects the encoder and decoder is that the hidden states of each layer in the decoder is copied from the last output hidden states of the corresponding layer in the encoder. The decoder architecture is similar to the encoder, but the flow direction of hidden states among layers is reversed. This enables the adoption of 2-D deconvolutional layers, which are the reverse operation of 2-D
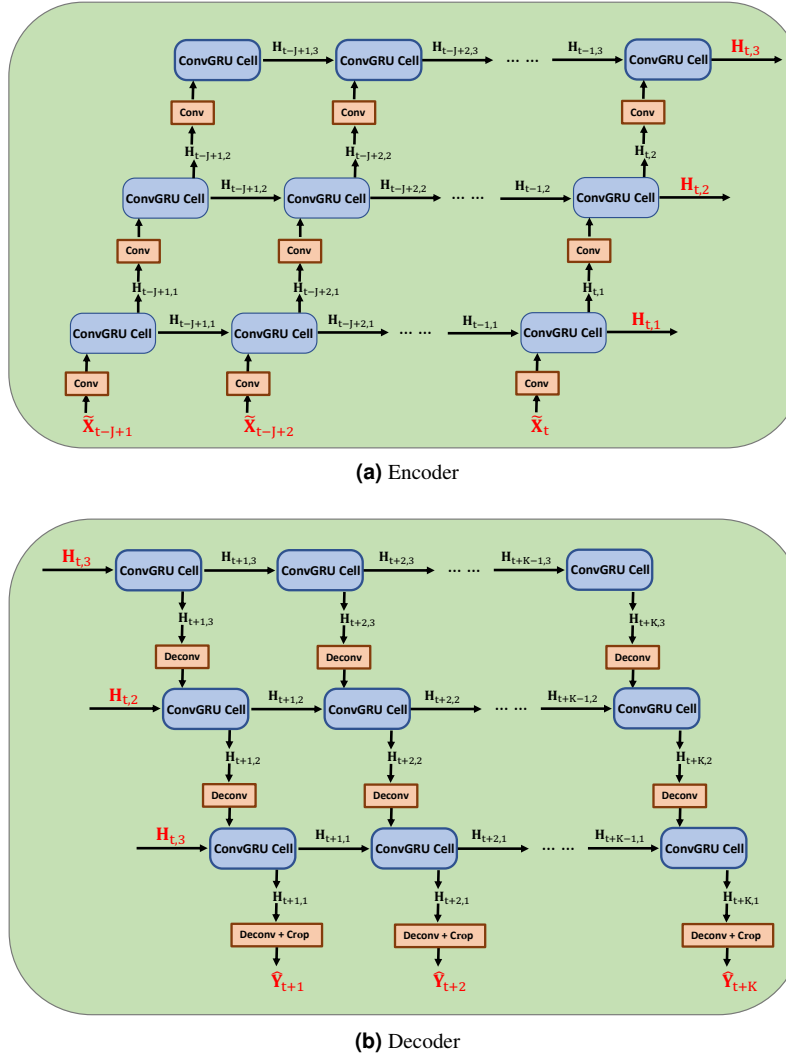
**(a)** Encoder

(b) Decoder

Figure 2. Three-layer ConvGRU network, where the initial hidden states of the decoder are copied from the last hidden states of the encoder. (a) Encoder architecture utilizing ConvGRU cells and 2-D convolutional layers. (b) Decoder architecture constructed with ConvGRU cells and 2-D deconvolutional layer.

convolutional layer. They ensure that the inputs and network parameters in each decoder layer's ConvGRU cells are consistent with those in the encoder, so that the last output hidden states from the encoder can be utilized by the decoder. For the outputs of the first layer of the decoder, the grid is cropped to the ENSO region.

The encoder-decoder Seq2Seq structure of the ConvGRU network can be interpreted as follows: the encoder compresses the input SST maps of the Pacific region into hidden states across all layers, while the decoder unfolds the hidden states from the encoder to generate predictions for the ENSO region. Consequently, the ConvGRU network approximates the problem stated in Eq. (1):

$$
\begin{aligned}
\hat{\boldsymbol{Y}}_{t+1:t+K} &= \underset{\boldsymbol{Y}_{t+1:t+K}}{\arg\max} \; \mathbb{P}\left(\boldsymbol{Y}_{t+1:t+K} \mid \tilde{\boldsymbol{X}}_{t-J+1:t}\right) \\
&\approx \underset{\boldsymbol{Y}_{t+1:t+K}}{\arg\max} \; \mathbb{P}\left(\boldsymbol{Y}_{t+1:t+K} \mid f_{\mathrm{ENC}}\left(\tilde{\boldsymbol{X}}_{t-J+1:t} \mid \boldsymbol{W}_{\mathrm{ENC}}\right)\right) \\
&\approx g_{\mathrm{DEC}}\left(f_{\mathrm{ENC}}\left(\tilde{\boldsymbol{X}}_{t-J+1:t} \mid \boldsymbol{W}_{\mathrm{ENC}}\right) \mid \boldsymbol{W}_{\mathrm{DEC}}\right),
\end{aligned}
\tag{3}
$$

where $f_{\mathrm{ENC}}(\cdot|\boldsymbol{W}_{\mathrm{ENC}})$ and $g_{\mathrm{DEC}}(\cdot|\boldsymbol{W}_{\mathrm{DEC}})$ represent the encoder and decoder, respectively, with network parameters $\boldsymbol{W}_{\mathrm{ENC}}$ and $\boldsymbol{W}_{\mathrm{DEC}}$.

(a) Training   **(b)** Testing
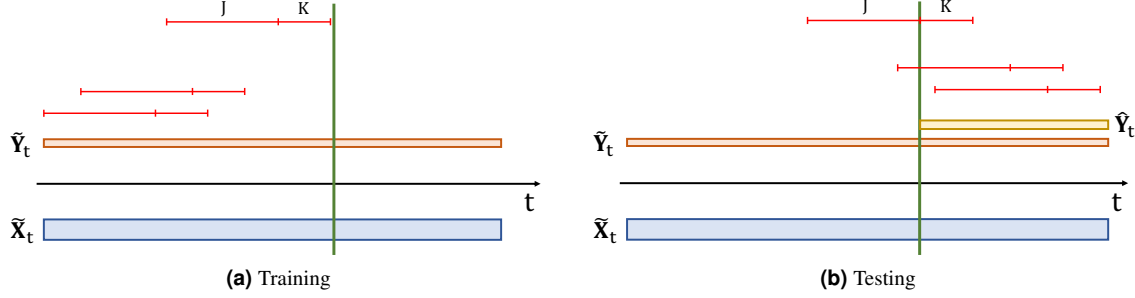
Figure 3. Data setup for training and testing. The green vertical lines divide the entire dataset into the training data and the testing data. The performance metric is evaluated to the right of the green line, and no data in this part is used in training. (a) Data setup during the training process. The red lines depict the training windows of $\{(\tilde{\boldsymbol{X}}_t, \tilde{\boldsymbol{Y}}_t)\}_{t=1}^{T}$, where the left part represents the condition range ($t-J+1$ to t for some t), and the right part represents the prediction range ($t+1$ to $t+K$). Note that all training windows are to the left of the green line. (b) During the testing process, the prediction range is strictly to the right of the green line.

Training process

Given a training dataset consisting of SST maps, denoted as $\{(\tilde{\boldsymbol{X}}_t, \tilde{\boldsymbol{Y}}_t)\}_{t=1}^{T}$, of the Pacific and ENSO regions, respectively, the network parameters $\boldsymbol{W}^*_{\text{ENC}}$ and $\boldsymbol{W}^*_{\text{DEC}}$ of the encoder and decoder can be learned by minimizing the difference between the predicted sequence $\hat{\boldsymbol{Y}}_{t+1:t+K}$ and the ground truth sequence $\tilde{\boldsymbol{Y}}_{t+1:t+K}$. The optimization process can be described as follows:

$$
\begin{aligned}
\boldsymbol{W}^*_{\text{ENC}}, \boldsymbol{W}^*_{\text{DEC}} &= \operatorname*{argmin}_{\boldsymbol{W}_{\text{ENC}}, \boldsymbol{W}_{\text{DEC}}} \sum_{t=J}^{T-K} \mathscr{L}\left(\tilde{\boldsymbol{Y}}_{t+1:t+K}, g_{\text{DEC}}\left(f_{\text{ENC}}\left(\tilde{\boldsymbol{X}}_{t-J+1:t} \mid \boldsymbol{W}_{\text{ENC}}\right) \mid \boldsymbol{W}_{\text{DEC}}\right)\right) \\
&= \operatorname*{argmin}_{\boldsymbol{W}_{\text{ENC}}, \boldsymbol{W}_{\text{DEC}}} \sum_{t=J}^{T-K} \mathscr{L}\left(\tilde{\boldsymbol{Y}}_{t+1:t+K}, \hat{\boldsymbol{Y}}_{t+1:t+K}\right),
\end{aligned}
\tag{4}
$$

where the loss function $\mathscr{L}$ can be chosen, for instance, as the mean squared loss. The optimization problem in Eq. (4) can be solved using stochastic gradient descent algorithms, such as Adam[44] and Adagrad[45].

During the training process, multiple training windows (instances) of length $J+K$ are generated from the training dataset, each with different start points. The condition (J) and prediction (K) ranges remain fixed for all training windows. For instance, if the training dataset spans from month 1 to month 10000, training windows can be created with t in Eq. (4) ranging from J to $10000-K$. Figure 3a illustrates the generation of training windows. Once the ConvGRU network is trained, it can be evaluated on a testing dataset $\{(\tilde{\boldsymbol{X}}_t, \tilde{\boldsymbol{Y}}_t)\}_{t=T+1}^{T+F}$, as depicted in Figure 3b.

# Results and discussions

Experimental setup

Experiment results and discussions of the ConvGRU network on various global climate and atmospheric reanalysis datasets are presented. The reanalysis datasets used consist of two SST datasets and one air temperature dataset. The performance of the ConvGRU network is evaluated on the ENSO region spatio-temporal sequence prediction task for the SST datasets. Additionally, the performance is compared with several existing models on a downstream task of predicting the Niño 3.4 index, which is calculated based on the aforementioned ENSO region spatio-temporal sequence prediction task. The ConvGRU network is also evaluated on the spatio-temporal sequence prediction task for the air temperature dataset, which covers almost 2/3 of the global surface.

For the numerical experiments, the ConvGRU network is implemented using PyTorch[46]. The experiments are conducted on a Linux server equipped with a single GPU, either NVIDIA GeForce GTX 1080Ti or NVIDIA RTX A6000. The implementation codes of the ConvGRU network are publicly available, and more detailed information can be found in the data availability section.

CCSM4 dataset

The CCSM4 dataset is a modeled SST dataset derived from a 1300-year, pre-industrial control integration of the Community Climate System Model version 4 (CCSM4)[47]. The dataset is sampled and averaged monthly on the model's native ocean grid with a normal resolution of approximately $1° \times 0.5°$ (longitude-latitude). The SST maps of the Pacific and ENSO regions are

(a) Ground truth

(b) Prediction
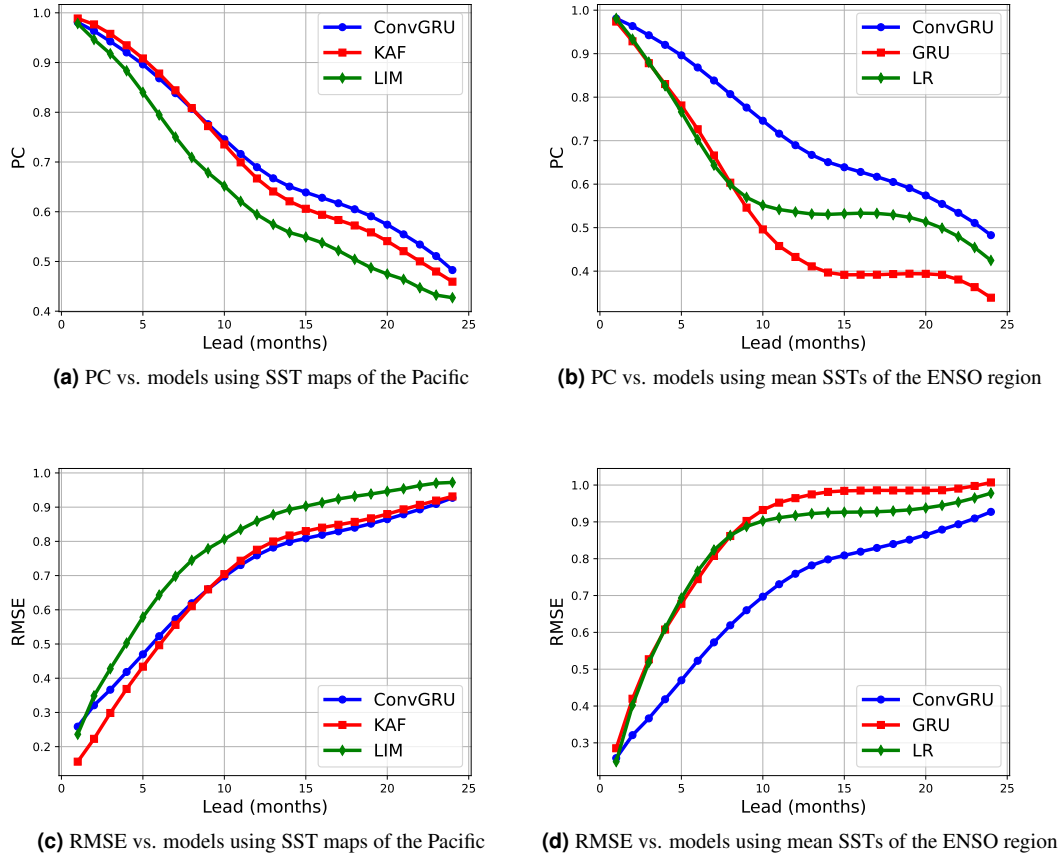

(c) Diff.

(d) RMSE and PC
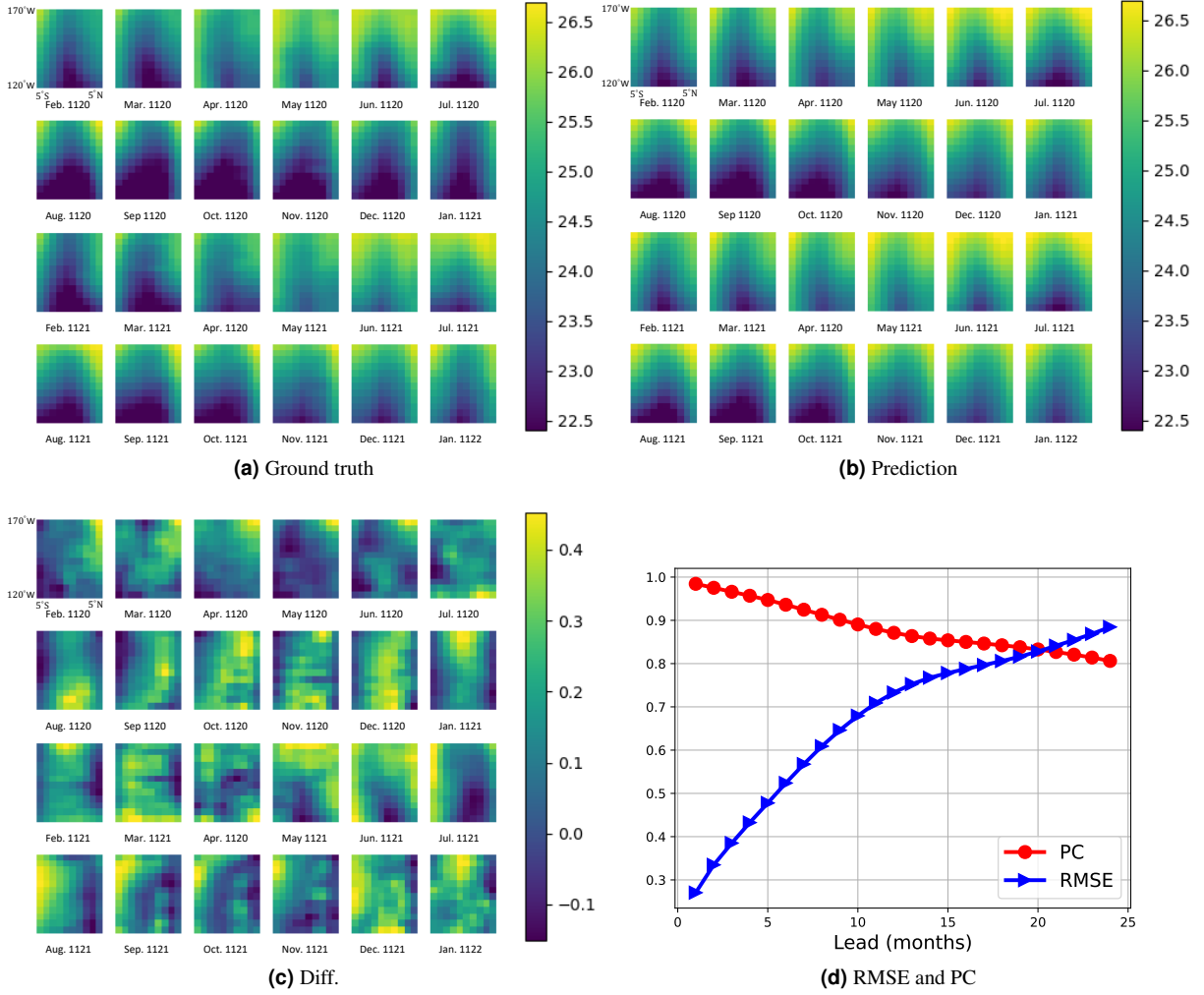
Figure 4. Performance on the ENSO region spatio-temporal sequence prediction task. (a): Sample ground truth of the ENSO region starting from February 1120. (b): Sample prediction of the ENSO region starting from February 1120. (c): Sample difference between the ground truth and prediction of the ENSO region start from February 1120. (d) RMSE per grid cell and PC as a function of lead time computed in the testing period.

extracted from specific longitude-latitude boxes. The Pacific region covers 16°E-56°W and 69°S-32°N (256 × 256 grid), while the ENSO region covers 170°-120°W and 5°S-5°N (45 × 38 grid). To reduce computational complexity and GPU memory usage, the SST maps of the Pacific and ENSO regions are down-sampled to 64 × 64 and 12 × 10 grids, respectively. The CCSM4 dataset is split into disjoint training and testing data periods, with years 1-1099 allocated for training and years 1100-1300 for testing.

For experiments on the CCSM4 dataset, a 3-layer ConvGRU network is implemented with parameters provided in the `NetParams.py` in the `ConvGRU_CCSM4` folder. The condition range (J) and the prediction range (K) are set to 48 and 24 months, respectively.

Figure 4 illustrates the performance of the ConvGRU network on the ENSO region spatio-temporal sequence prediction task. Figures 4a, 4b and 4c include a sample comparison, starting from February 1120, between the ground truth and the network's prediction for the ENSO region. The patterns in both the ground truth and the prediction exhibit high similarity. Figure 4d presents the prediction skill assessed using RMSE and PC metrics over the entire testing period as a function of lead time. The RMSE values are averaged over all possible start points in the testing data split and grid cells of the ENSO region. For PC, the ground truths and predictions are vectorized and concatenated over all possible start points in the testing data split. The RMSE and PC results demonstrate the high correlation and low error characteristics of predictions generated by the ConvGRU network.

(a) PC vs. models using SST maps of the Pacific

(b) PC vs. models using mean SSTs of the ENSO region

(c) RMSE vs. models using SST maps of the Pacific

(d) RMSE vs. models using mean SSTs of the ENSO region

Figure 5. Performance of the ConvGRU network against other models on predicting the Niño 3.4 index in the CCSM4 dataset during 1100-1300. (a) and (c) PC and RMSE, respectively, as a function of lead time, compared to KAF and LIM. (b) and (d) PC and RMSE, respectively, as a function of lead time, compared to Seq2Seq with GRU and LR.
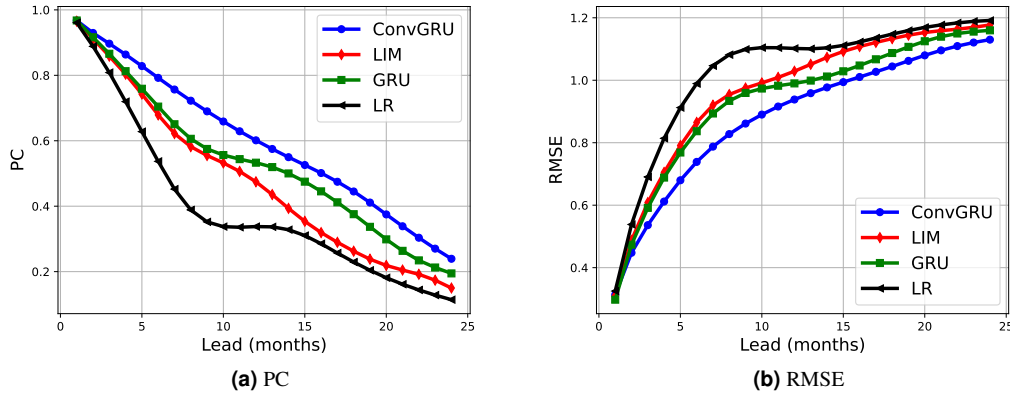
Next, the performance of the ConvGRU network is compared with existing models for predicting the Niño 3.4 index in the CCSM4 dataset. Here Niño 3.4 indices mean SST anomalies relative to monthly climatology (average SST) of the ENSO region. The models selected for comparison include KAF[17, 20, 21], LIM[15–17], Seq2Seq with GRU[31, 43], and LR. KAF and LIM utilize SST maps of the Pacific region as input (predictor) variables, while Seq2Seq with GRU and LR use mean SSTs of the ENSO region. Seq2Seq with GRU is implemented using the DeepAR model[48] from the GluonTS package[49], with a 1-layer GRU network with a 20-dimensional hidden state, and the condition and prediction ranges (*J* and *K*) the same as the ConvGRU network. For LR, *K* = 24 separate models are trained for lead months 1-24, using *J* = 48 months lagged mean SSTs of the ENSO region (including the current month) as input features.

Figure 5 illustrates the performance of the ConvGRU network compared against other models in predicting the Niño 3.4 index in the CCSM4 dataset over the testing period of 1100-1300, with the training period from 1 to 1099. A threshold of PC = 0.6 is commonly used to differentiate useful from non-useful predictions[17]. The comparison demonstrates that although the performance of the ConvGRU network deteriorates with longer lead times, it consistently outperforms the competing models in terms of both PC and RMSE, particularly in the long-term prediction range. When considering the useful prediction range using the PC threshold of 0.6, the ConvGRU network achieves the longest useful range of 18-19 months, surpassing KAF, LIM, Seq2Seq with GRU, and LR by 3-4, 6-7, 10-11, and 10-11 months, respectively.

NOAA-GDFL-SPEAR dataset
The NOAA-GDFL-SPEAR dataset used in the numerical experiment is a simulated monthly averaged SST dataset with a nominal resolution of $1° \times 1°$ (longitude-latitude) from the GFDL SPEAR large ensembles. This includes 30-member ensembles of climate change simulations covering the period 1921-2100 using the SPEAR-MED climate model[50]. The simulations are forced with historical radiative forcings from 1921 to 2014 and SSP5-8.5 projected radiative forcings[51, 52] from 2015 to

Figure 6. Performance averaged over 30 ensembles of the ConvGRU network against other models on predicting the Niño 3.4 index in the NOAA-GDFL-SPEAR dataset during 2051-2100. (a) and (b) PC and RMSE, respectively, as a function of lead time, compared to LIM, Seq2Seq with GRU, and LR.

2100. The SST maps of the Pacific and ENSO region are extracted from the longitude-latitude boxes 150°E-82°W, 69°S-59°N (128 × 128 grid) and 170°-120°W, 5°S-5°N (50 × 10 grid), respectively. Similar to the previous comparison, the SST maps in both regions are down-sampled to 64 × 64 and 25 × 5 grids, reducing computational complexity and GPU memory usage.

The NOAA-GDFL-SPEAR dataset in each ensemble is divided into disjoint training and testing data splits, with the training period covering years 1921-2050 and the testing period covering years 2051-2100. For experiments on this dataset, a 3-layer ConvGRU network is implemented using the parameters specified in the `NetParams.py` file in the `ConvGRU_SPEAR` folder. The ConvGRU network is trained using data from all 30 ensembles that end on or before the year 2050, and then tested on data from all 30 ensembles starting from the year 2051. The condition range (J) and the prediction range (K) are 48 and 24 months, respectively.

Similar to the previous comparison, the performance of the ConvGRU network is compared to several existing models for predicting the Niño 3.4 index in the NOAA-GDFL-SPEAR dataset. The selected models for comparison include LIM, Seq2Seq with GRU, and LR. For LIM and LR, separate models are trained for each ensemble in the dataset, since the NOAA-GDFL-SPEAR dataset contains data from 30 ensembles, and the metrics computed in the testing period are averaged over all ensembles. Seq2Seq with GRU utilizes the DeepAR model to handle multiple time series with a single model, and the detailed and fine-tuned settings remain the same in the CCSM4 dataset.

Figure 6 presents the performance averaged over 30 ensembles of the ConvGRU network compared against other models in predicting the Niño 3.4 index in the NOAA-GDFL-SPEAR dataset over the testing period of 2051-2100, using the training period of 1921-2051. The results of the experiments demonstrate that the ConvGRU network significantly outperforms the competing models in terms of both PC and RMSE, with the longest useful range of 12 months, surpassing LIM, Seq2Seq with GRU, and LR by 4-5, 4, and 7 months, respectively.

NOAA-CIRES air temperature dataset

The NOAA-CIRES air temperature dataset used in this experiment is a simulated monthly ensemble mean air temperature dataset at the 2m level with a nominal resolution of approximately 2° × 2° (longitude-latitude). It is from the NOAA-CIRES 20th-Century Reanalysis Version 2c[53,54], which provides comprehensive global atmospheric circulation dataset spanning the years 1850-2014. For the NOAA-CIRES air temperature dataset, we aim to demonstrate that the ConvGRU network is capable of accurately predicting other climate and atmospheric spatio-temporal sequence beyond the ENSO region. For this experiment, the target region is the longitude-latitude box 120°E − 1°W, 60°N − 60°S (128 × 64 grid), covering almost two-thirds of the total global surface.

The NOAA-CIRES air temperature dataset is divided into disjoint training and testing periods, with the training period covering the years 1851-1980 and testing period covering 1981-2014. In this experiment, a 3-layer ConvGRU network is implemented using the parameters specified in the `NetParams.py` file in the `AirTmp_M` folder. The condition range (J) and the prediction range (K) are set to 24 and 12 months, respectively.

Figure 7 illustrates the performance of the ConvGRU network on the air temperature spatio-temporal sequence prediction task. Figures 7a, 7b, and 7c presents a sample comparison starting from January 1987 between the ground truth and the prediction, revealing a very similar pattern in both the ground truth and the prediction. Figures 7d and 7e present the prediction

(a) Ground truth

(b) Prediction

(c) Diff.

(d) PC

(e) RMSE
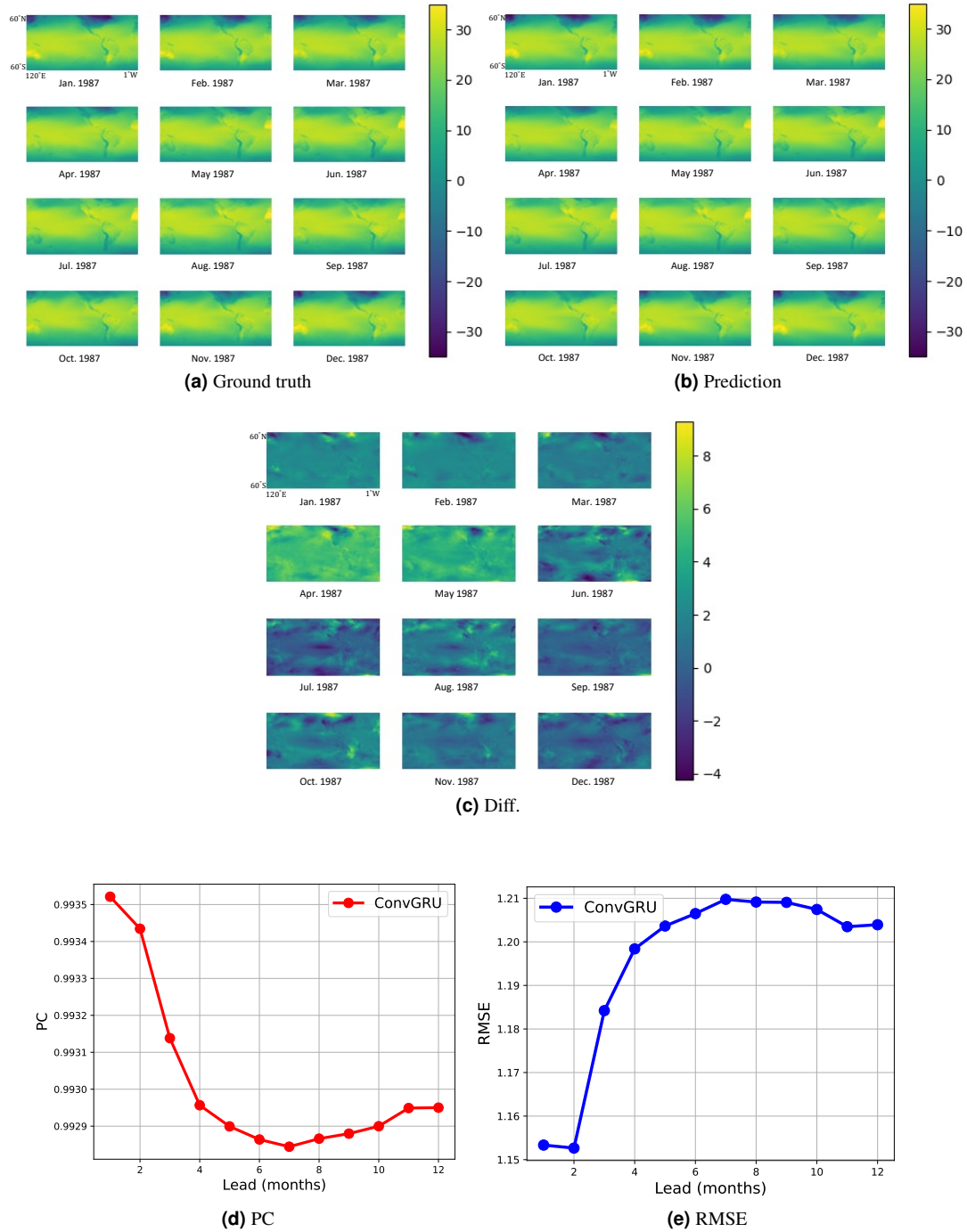
Figure 7. Performance on the air temperature spatio-temporal sequence prediction task. (a) Sample ground truth starting from January 1987. (b) Sample prediction starting from January 1987. (c) Sample difference between ground truth and prediction starting from January 1987. (d) PC as a function of lead time computed in the testing period. (e) RMSE per grid cell as a function of lead time computed in the testing period.

skill assessed using RMSE and PC, similar to Fig. 4, over the entire testing data split and as a function of lead time. The PC result demonstrates a significantly high correlation of over 99% between the ground truth and the prediction within a prediction range of 12 months. The RMSE ranges from approximately 1.1 °C to 1.2 °C.

Conclusion

We addressed the ENSO region spatio-temporal sequence prediction problem by proposing a modified ConvGRU network, as well as its downstream task of predicting the Niño 3.4 index. The ConvGRU network incorporated 2-D convolutional layers within a ConvGRU cell and employed an encoder-decoder Seq2Seq structure, offering advantages over existing models such as LR, LIMs, KAF, and Seq2Seq with GRU. These advantages include the ability to output future SST maps of the ENSO region, rather than ENSO indices, and modelling approximate nonlinear dynamics.

Through experiments on various climate and atmospheric reanalysis datasets, we demonstrated the effectiveness of the ConvGRU network in predicting future SST maps in the ENSO region. The ConvGRU network outperformed existing models in various scenarios, including the Niño 3.4 index prediction, showcasing its capabilities in downstream applications. We also evaluated the performance of the network in predicting other climate-related tasks, such as predicting monthly air temperature over a large portion of the global surface, which further demonstrate its potential for accurate spatio-temporal sequence predictions.

Overall, the proposed ConvGRU network offers a promising approach for ENSO region spatio-temporal sequence prediction and hold potential for advancing climate and atmospheric prediction and related domains. There are opportunities for further exploration, such as variants or successors of the ConvGRU network and leveraging Transformer-based models to enhance prediction skills. Additionally, this study lays the foundation for constructing real-time forecast systems with monthly updates based on the availability of observational data.

Data availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request. Additionally, the codes and detailed information about the datasets can be found at the following public GitHub repository: https://github.com/LingdaWang/ConvGRU_ENSO_Forecast.

References

1. McPhaden, M. J., Zebiak, S. E. & Glantz, M. H. ENSO as an integrating concept in earth science. *Science* **314**, 1740–1745 (2006).

2. Fraedrich, K. & Müller, K. Climate anomalies in Europe associated with ENSO extremes. *Int. J. Climatol.* **12**, 25–31 (1992).

3. Kumar, K. K., Rajagopalan, B. & Cane, M. A. On the weakening relationship between the Indian monsoon and ENSO. *Science* **284**, 2156–2159 (1999).

4. Glantz, M. H. *et al. Currents of change: Impacts of El Niño and La Niña on climate and society* (Cambridge University Press, 2001).

5. Cook, A. R., Leslie, L. M., Parsons, D. B. & Schaefer, J. T. The impact of El Niño–Southern Oscillation (ENSO) on winter and early spring US tornado outbreaks. *J. Appl. Meteorol. Climatol.* **56**, 2455–2478 (2017).

6. Camargo, S. J. & Sobel, A. H. Western North Pacific tropical cyclone intensity and ENSO. *J. Clim.* **18**, 2996–3006 (2005).

7. Squire, D. T. *et al.* Likelihood of unprecedented drought and fire weather during Australia's 2019 megafires. *Clim. Atmospheric Sci.* **4**, 64 (2021).

8. Trouet, V., Taylor, A. H., Carleton, A. M. & Skinner, C. N. Interannual variations in fire weather, fire extent, and synoptic-scale circulation patterns in northern California and Oregon. *Theor. Appl. Climatol.* **95**, 349–360 (2009).

9. Rasmusson, E. M. & Carpenter, T. H. Variations in tropical sea surface temperature and surface wind fields associated with the Southern Oscillation/El Niño. *Mon. Weather. Rev.* **110**, 354–384 (1982).

10. Trenberth, K. E. The definition of El Niño. *Bull. Am. Meteorol. Soc.* **78**, 2771–2778 (1997).

11. Trenberth, K. E. & Stepaniak, D. P. Indices of El Niño evolution. *J. Clim.* **14**, 1697–1701 (2001).

12. Hansen, J. W., Challinor, A., Ines, A., Wheeler, T. & Moron, V. Translating climate forecasts into agricultural terms: Advances and challenges. *Clim. Res.* **33**, 27–41 (2006).

13. Hammer, G. *et al.* Advances in application of climate prediction in agriculture. *Agric. Syst.* **70**, 515–553 (2001).

14. Barnston, A. G., Tippett, M. K., Ranganathan, M. & L'Heureux, M. L. Deterministic skill of ENSO predictions from the North American Multimodel Ensemble. *Clim. Dyn.* **53**, 7215–7234 (2019).

15. Penland, C. & Magorian, T. Prediction of Niño 3 sea surface temperatures using linear inverse modeling. *J. Clim.* **6**, 1067–1076 (1993).

16. Penland, C. & Sardeshmukh, P. D. The optimal growth of tropical sea surface temperature anomalies. *J. Clim.* **8**, 1999–2024 (1995).

17. Wang, X., Slawinska, J. & Giannakis, D. Extended-range statistical ENSO prediction through operator-theoretic techniques for nonlinear dynamics. *Sci. Reports* **10**, 1–15 (2020).

18. Chapman, D., Cane, M. A., Henderson, N., Lee, D. E. & Chen, C. A vector autoregressive ENSO prediction model. *J. Clim.* **28**, 8511–8520 (2015).

19. Lorenz, E. N. Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmospheric Sci.* **26**, 636–646 (1969).

20. Zhao, Z. & Giannakis, D. Analog forecasting with dynamics-adapted kernels. *Nonlinearity* **29**, 2888 (2016).

21. Burov, D., Giannakis, D., Manohar, K. & Stuart, A. Kernel analog forecasting: Multiscale test problems. *Multiscale Model. & Simul.* **19**, 1011–1040 (2021).

22. Ham, Y.-G., Kim, J.-H. & Luo, J.-J. Deep learning for multi-year ENSO forecasts. *Nature* **573**, 568–572 (2019).

23. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).

24. Huang, A., Vega-Westhoff, B. & Sriver, R. L. Analyzing El Niño–Southern Oscillation predictability using long-short-term-memory models. *Earth Space Sci.* **6**, 212–221 (2019).

25. Shi, X. *et al.* Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Adv. Neural Inf. Process. Syst.* **28** (2015).

26. Shi, X. *et al.* Deep learning for precipitation nowcasting: A benchmark and a new model. *Adv. Neural Inf. Process. Syst.* **30** (2017).

27. Ballas, N., Yao, L., Pal, C. & Courville, A. Delving deeper into convolutional networks for learning video representations. *arXiv preprint arXiv:1511.06432* (2015).

28. Zhang, Y., Wallace, J. M. & Battisti, D. S. ENSO-like interdecadal variability: 1900–93. *J. Clim.* **10**, 1004–1020 (1997).

29. Hess, J. C., Scott, C. A., Hufford, G. L. & Fleming, M. D. El Niño and its impact on fire weather conditions in Alaska. *Int. J. Wildland Fire* **10**, 1–13 (2001).

30. Medsker, L. R. & Jain, L. Recurrent neural networks. *Des. Appl.* **5**, 64–67 (2001).

31. Cho, K., van Merriënboer, B., Bahdanau, D. & Bengio, Y. On the properties of neural machine translation: Encoder–decoder approaches. In *Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-8)*, 103–111 (2014).

32. Graves, A. Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850* (2013).

33. Finn, C., Goodfellow, I. & Levine, S. Unsupervised learning for physical interaction through video prediction. *Adv. Neural Inf. Process. Syst.* **29** (2016).

34. Wang, Y. *et al.* Predrnn: A recurrent neural network for spatiotemporal predictive learning. *IEEE Transactions on Pattern Analysis Mach. Intell.* **45**, 2208–2225 (2022).

35. Fonseca, A. H. d. O., Zappala, E., Caro, J. O. & van Dijk, D. Continuous spatiotemporal transformers. *arXiv preprint arXiv:2301.13338* (2023).

36. Yang, Z., Yang, X. & Lin, Q. TCTN: A 3D-temporal convolutional transformer network for spatiotemporal predictive learning. *arXiv preprint arXiv:2112.01085* (2021).

37. Keisler, R. Forecasting global weather with graph neural networks. *arXiv preprint arXiv:2202.07575* (2022).

38. Lam, R. *et al.* Graphcast: Learning skillful medium-range global weather forecasting. *arXiv preprint arXiv:2212.12794* (2022).

39. Ravuri, S. *et al.* Skilful precipitation nowcasting using deep generative models of radar. *Nature* **597**, 672–677 (2021).

40. Ding, H., Newman, M., Alexander, M. A. & Wittenberg, A. T. Skillful climate forecasts of the tropical Indo-Pacific ocean using model-analogs. *J. Clim.* **31**, 5437–5459 (2018).

41. Eisner, T., Farkas, B., Haase, M. & Nagel, R. *Operator Theoretic Aspects of Ergodic Theory*, vol. 272 (Springer, 2015).

42. Giannakis, D. Data-driven spectral decomposition and forecasting of ergodic dynamical systems. *Appl. Comput. Harmon. Analysis* **47**, 338–396 (2019).

43. Chung, J., Gulcehre, C., Cho, K. & Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. In *NIPS 2014 Workshop on Deep Learning, December 2014* (2014).

44. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *Int. Conf. on Learn. Represent.* (2015).

45. Duchi, J., Hazan, E. & Singer, Y. Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.* **12** (2011).

46. Paszke, A. *et al.* PyTorch: An imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* **32** (2019).

47. Gent, P. R. *et al.* The community climate system model version 4. *J. Clim.* **24**, 4973–4991 (2011).

48. Salinas, D., Flunkert, V., Gasthaus, J. & Januschowski, T. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *Int. J. Forecast.* **36**, 1181–1191 (2020).

49. Alexandrov, A. *et al.* GluonTS: Probabilistic and neural time series modeling in Pyhon. *The J. Mach. Learn. Res.* **21**, 4629–4634 (2020).

50. Delworth, T. L. *et al.* SPEAR: The next generation GFDL modeling system for seasonal to multidecadal prediction and projection. *J. Adv. Model. Earth Syst.* **12** (2020).

51. Van Vuuren, D. P. *et al.* A new scenario framework for climate change research: Scenario matrix architecture. *Clim. C* **122**, 373–386 (2014).

52. O'Neill, B. C. *et al.* The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geosci. Model. Dev.* **9**, 3461–3482 (2016).

53. Compo, G. P. *et al.* The twentieth century reanalysis project. *Q. J. Royal Meteorol. Soc.* **137**, 1–28 (2011).

54. Krueger, O., Schenk, F., Feser, F. & Weisse, R. Inconsistencies between long-term trends in storminess derived from the 20CR reanalysis and observations. *J. Clim.* **26**, 868–874 (2013).

Acknowledgements

Author contributions

Z.Z., R.S., and V.H. designed the research. L.W., Z.Z., R.S., and V.H. proposed the methodology, L.W. and S.A. performed numerical experiments. L.W. wrote the main manuscript. All authors reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to L.W., R.S., or Z.Z.