Soft Periodic Convolutional Recurrent Network for Spatiotemporal Climate Forecast and Periodicity Analysis

Submitted to

Graduate School of Information Science and Technology

Osaka University

January 2022

Ekasit PHERMPHOONPHIPHAT

Abstract

Spatiotemporal prediction on climate data aims to predict future spatial data by learning from prior spatial sequence data. The study investigates the availability of the constructed architecture of convolutional layers and deconvolutional layers, and applies to convolutional long short-term memory (ConvLSTM). The ConvLSTM structure is utilized as a multivariate spatiotemporal climate prediction to predict the upper tropospheric circulations over the Northern Hemisphere, a geopotential height at 300 hPa (ZH300) variable. In addition, we compare the ConvLSTM with the baseline methods of convolutional neural network (CNN) and linear regression (LR). The results showed that the proposed model obtained a root mean square error (RMSE) of 77.36 meters (0.84% compared to the average ZH300 value) in short-term prediction. Meanwhile, CNN and LR models obtained RMSE of 109.35 (1.19%) and 153.61 (1.67%), respectively. The ConvLSTM maintains RMSE even in long-term prediction. Furthermore, the prediction features' investigation result showed that temperature at 300 hPa (T300) and prior ZH300 features are essential for ZH300 prediction. However, the ConvLSTM that considers only short-range sequential information may not be adequate to deal with periodic patterns such as seasonality. It is crucial for the climate prediction model to capture periodicity change.

This dissertation adopts a Periodic Convolutional Recurrent Network (Periodic-CRN) model to employ the periodicity component in the periodic representation dictionary (PRD). However, Periodic-CRN assumes stationary periodicity as PRD refers to exactly the same index of the previous cycle, which means "hard" periodicity. This paper focuses on "soft" yearly periodicity, including nearby months and a multiple-year cycle. It proposes a Soft Periodic-CRN (SP-CRN) with three proposals of utilizing periodicity components: nearby-time (PRD-1), periodic-depth (PRD-2), and periodic-depth differencing (PRD-3) representation to improve climate forecasting accuracy. The attention module in the SP-CRN can weigh the importance of periodic representations, which help capture the periodic pattern. In order to capture the spatial change over the periodicity phase, the dynamic spatial weights (DSW) on the attention module are proposed, which are multiple spatial weights for the attention module output, and switched according to the prediction month. This study experimented on geopotential height at 300 hPa (ZH300) and sea surface temperature (SST) datasets of ERA-Interim. The results showed the superiority of PRD-1 plus or minus one month of a prior cycle to capture the phase shift. In addition, PRD-3 considered that only the depth of one differencing periodic cycle (i.e., the previous year) could significantly improve the prediction accuracy of ZH300 and SST. The mixed method of PRD-1

and PRD-3 (SP-CRN-1+3) showed a competitive or slight improvement over their base models. We improve the prediction result drastically by adding the metadata component to indicate the month with one-hot encoding to SP-CRN-1+3. The results showed that the proposed method could learn four years of periodicity from the data, which may relate to the El Niño–Southern Oscillation (ENSO) cycle.

Acknowledgement

The dissertation would not have been possible without supports and helps of my professors, family, and friends.

I would like to express my very great appreciation to Professor Masayuki Numao for giving me the opportunities to study in Japan, for the continuous support of my study, and providing me with an excellent atmosphere for doing research.

I would also like to express my sincere gratitude to my advisor, Associate Professor Ken-ichi Fukui, for his patient guidance, motivation, enthusiastic encouragement, immense knowledge, insightful comments, corrected my writing, and supported me through this research work and my life.

My gratitude goes out to Professor Yasumasa Fujisaki and Professor Jun Tanida, the dissertation committee, for their great comments and contributions toward my doctoral degree.

I would also like to offer my special thanks to Associate Professor Tomohiko Tomita and Assistant Professor Takashi Morita for their valuable and constructive suggestions throughout this research work.

I would like to acknowledge the laboratory secretaries Ms. Megumi Tanabe, Ms. Mitsuyo Otsuka, Ms. Akiko Yamamoto, and Ms. Makiko Kuroki, Ms. Mamie Abe, Ms. Yumiko Morikawa, and Ms. Yohko Okada for their accommodating and assisting me with all of the university affairs.

I am grateful to Professor Boonserm Kijsirikul for an opportunity to study at Osaka University. Special gratitude also goes to Assistant Professor Sukree Sinthupinyo and Assistant Professor Denduang Pradubsuwun for their encouragement to pursue my doctorate study at Osaka University.

I would also like to thank The Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan for their continuous support in terms of the Japanese Government (MEXT) scholarship.

My sincere thanks also go to Dr. Nattapong Thammasan, Dr. Sirawit Sopchoke, Dr. Wasin Kalintha, Noppayut Sriwatanasakdi, Dr. Juan Lorenzo Hagad, Dr. Bassel Ali, Nattapat Boonprakong, Taweesak Emsawas, Nat Pavasant, Pongpisit Thanasutives, ex- and present students under Professor Numao's laboratory, for their encouragement, guidance, and insightful comments during the work.

I would also like to thank my friends and Thai students' association in japan under the royal

patronage (TSAJ) for their support and encouragement throughout my life in Japan.

Finally, I wish to thank Prasit Phermphoonphiphat, Arunee Phermphoonphiphat, Sureeporn Phermphoonphiphat, Sasithorn Phermphoonphiphat, Aekachai Phermphoonphiphat, Raksina Khanaroek, my family for great support, cheer up, and everything I need throughout my study and life in Japan.

Contents

List of Figures **iii**

List of Tables **v**

1 Introduction **1**
1.1 Research Motivation 1
1.2 Contributions 4
1.3 Organization of the Dissertation 6

2 Literature Review **7**
2.1 ML Climate Prediction Models 7
2.2 ML Climate Prediction Models with Periodicity Component 8
2.3 Base Methods 9
2.3.1 Convolutional LSTM 9
2.3.2 Convolutional GRU 10
2.3.3 Periodic Convolutional Recurrent Network 12
2.4 Summary 15

3 Climate Prediction using Multivariate Features by ConvLSTM **17**
3.1 Research Propose 17
3.2 Experiment Settings 18
3.2.1 Dataset 18
3.2.2 Comparison Methods 18
3.3 Results 19
3.3.1 Short-term prediction 19
3.3.2 Long-term prediction 24
3.3.3 Prediction features' investigation 24
3.4 Discussion 25

4 Soft Periodic Convolutional Recurrent Network **29**
4.1 Overview 29
4.2 Problem Statement 30

4.3 Soft Periodicity . . . 31
4.4 Soft Periodic Recurrent Neural Network . . . 33
4.4.1 CRN Component . . . 33
4.4.2 Periodic Representation . . . 33
4.4.3 Update Mechanism . . . 33
4.4.4 Load Mechanism . . . 34
4.4.5 Attention Module . . . 36
4.4.6 Fusion Module . . . 36
4.4.7 Metadata Component . . . 37
4.5 Experiment Settings . . . 37
4.5.1 ERA-Interim Dataset . . . 37
4.5.2 Evaluation Metric . . . 38
4.5.3 Comparison Methods . . . 39
4.5.4 Settings . . . 39
4.6 Results on ZH300 . . . 40
4.6.1 Input Sequence Length . . . 40
4.6.2 PRD Types . . . 41
4.6.3 Combination of PRD Types and Metadata . . . 42
4.6.4 Spatial Distribution Analysis . . . 43
4.6.5 Static Spatial Weights vs Dynamic Spatial Weights . . . 46
4.6.6 Periodicity Analysis . . . 46
4.7 Results on SST . . . 48
4.7.1 Input Sequence Length . . . 49
4.7.2 PRD Types . . . 50
4.7.3 Combination of PRD Types and Metadata . . . 51
4.7.4 Spatial Distribution Analysis . . . 52
4.7.5 Static Spatial Weights vs Dynamic Spatial Weights . . . 54
4.7.6 Periodicity Analysis . . . 55
4.8 Discussion and Limitation of the Proposed Method . . . 55

5 Conclusion and Future Work **59**
5.1 Summary . . . 59
5.2 Future Research Directions . . . 60

A List of Publication **69**
A.1 Journal . . . 69
A.2 International Conference/Symposium (with review) . . . 69
A.3 Domestic Workshop (without review) . . . 69

List of Figures

1.1 The conceptual diagram of the proposed PRD in case of monthly representation. Each cell indicates a saved representation for the future reference. Type PRD-1 refers adjacent months of the last year, and types PRD-2 and 3 refer the same month of prior consecutive years. 3

2.1 ConvLSTM structure on the experiment in Chapter 3 10
2.2 The Periodic-CRN architecture 13

3.1 The Northern Hemisphere ZH300 on prediction and error distribution on April, 2018 (the best month of ConvLSTM) 20
3.2 The Northern Hemisphere ZH300 on prediction and error distribution on February, 2018 (the worst month of ConvLSTM) 21
3.3 The predicted value minus ground truth in every month of 2018 (blue color is overestimated and red color is underestimated). In every month, left figure is CNN and right figure is ConvLSTM. 23
3.4 RMSE of long-term prediction 24
3.5 The comparison of ZH300 observed data in January from 2011 to 2018. 26
3.6 The comparison of ZH300 observed data in every month of 2018. 27

4.1 The SP-CRN (LSTM) architecture. (**a**) the CRN component, (**b**) the PRD, (**c**) update mechanism, (**d**) load mechanism, (**e**) the attention module, (**f**) fusion module, and (**g**) metadata component. 30
4.2 Soft periodicity 31
4.3 ACF plot of ZH300 32
4.4 ACF plot of SST 32
4.5 Examples of the three proposals for the load mechanism to predict the climate variable of July 1999 (index 1 starts from the top left of P). (**a**) Nearby-Time representation with the nearby representation of 2, (**b**) Periodic-Depth representation with a depth of 5, and (**c**) Periodic-Depth-Differencing representation with a depth of 5. 34

4.6 A sample of the monthly spatial plot of January 2018. (**a**) The Northern Hemisphere of ZH300 data, where deep blue represents the lower value of ZH300. (**b**) The entire globe in the spatial grid of SST data, where deep red represents high temperature, and the terrain area (white) is ignored. Note that these projections are just for visualization. 37
4.7 The average RMSE of the comparison methods in different length of sequential input for ZH300 prediction. The error bar indicates the SD of the RMSE. . . . 41
4.8 The prediction and error distribution of ZH300 on February 2017. 45
4.9 The ZH300 error distribution of SP-CRN-1 (1,1,0) SSW, SP-CRN-1 (1,1,0) DSW, SP-CRN-2 (0,4,0) SSW, and SP-CRN-2 (0,4,0) DSW in January 2018, with RMSE of 72.51 m, 63.16 m, 63.03 m, and 55.94, respectively. 47
4.10 The average attention weight of training data on SP-CRN-1 (1,1,0) for ZH300 prediction from the average of five runs. 48
4.11 The average attention weight of training data on SP-CRN-2 (0,4,0) for ZH300 prediction from the average of five runs. The attention weight of depth-1 to depth-4 was the value of the prediction month for one to four prior periodic cycles. 49
4.12 The average RMSE of comparison methods of different lengths of sequential input for SST prediction. The error bar indicates the SD of the SST. 50
4.13 The prediction and error distribution of SST on January 2018. 53
4.14 The SST error distribution of SP-CRN-1 (1,1,0) SSW, SP-CRN-1 (1,1,0) DSW, SP-CRN-2 (0,6,0) SSW, and SP-CRN-2 (0,6,0) DSW (RMSE of 0.60 K) in January 2018 with RMSE of 0.75 K, 0.60 K, 0.65 K, and 0.54 K, respectively. . 54
4.15 The average attention weight of training data on SP-CRN-1 (1,1,0) for SST prediction from the average of five runs. 55
4.16 The average attention weight of training data on SP-CRN-2 (0,6,0) for SST prediction from the average of five runs. 56

List of Tables

3.1 RMSE (meters) of 12 months in 2018. 22
3.2 Excluded features result of average RMSE of 12 months in 2018. 25

4.1 Evaluation of different settings on the ZH300 prediction for SP-CRN with three proposals. The settings ((x, y, z) are the numbers of the nearby-time of PRD-1, periodic-depth of PRD-2, and periodic-depth of PRD-3, respectively. The SP-CRN-1, SP-CRN-2, and SP-CRN-3 indicate each proposal in SP-CRN which was a nearby-time representation, periodic-depth representation, and periodic-depth differencing representation, respectively. The highlighted result in blue is the best prediction accuracy for each proposal, while the gray highlight in the SP-CRN baseline result. 42
4.2 Evaluation of different settings on the ZH300 prediction for SP-CRN and comparison methods. 43
4.3 Evaluation of different settings on SST predictions for SP-CRN with three proposals. The highlighted result in blue shows the best prediction accuracy for each proposal, while the highlight in gray is the SP-CRN baseline result. 51
4.4 Evaluation of different settings on the SST prediction for SP-CRN and comparison methods. 52

Chapter 1

Introduction

1.1 Research Motivation

Climate change is one of the most crucial problems on earth, which causes unusual climate events such as extreme typhoons and global temperatures to rise; therefore, an accurate climate forecasting model is important. The numerical weather prediction (NWP) model simulates the physical equations of the atmosphere to generate the forecasts. NWP considers the initial state of the atmosphere to produce the forecasting; however, all meteorological observations are imperfect, causing the errors to grow over time. NWP also needs data assimilation to update the model state by comparing the model forecast with newly received observation. Meanwhile, increasing climate remote sensing (e.g., weather satellite, AMeDAS), including re-analysis data that combine numerical simulations with observations, generates a massive amount of data. Furthermore, it consumes a lot of computational resources to make a forecast by using NWP models, and requires climatology knowledge to build a model, unlike machine learning (ML). Currently, many applications are using ML in the climate domain, such as tropical cyclone forecasting [34, 10], long-term rainfall prediction [13], and more [43, 42, 49].

Time-series forecasting models [28] have emerged to forecast climate variables by learning from the historical data. Long short-term memory (LSTM) [16] is a kind of recurrent neural network (RNN) capable of learning long-term dependencies [51, 50]. Recent works in ML [30, 29, 48] have proven that LSTM is beneficial for solving time-series climate forecasting problems.

The spatiotemporal forecasting problem in the climate domain given by the spatial and tem-

poral data is crucial for predicting the next time frame. The convolutional long short-term memory network (ConvLSTM) [36] is the first ML method to tackle spatiotemporal climate forecasting by extending a fully connected LSTM (FC-LSTM) with convolutional structures. ConvLSTM was tested on precipitation nowcasting (up to 6 hours) in Hongkong from radar echo dataset.

Meanwhile, the atmospheric variation in the upper troposphere is used for the weather prediction at the surface because the upper tropospheric circulations can be precursors of the weather prediction [40]. Moreover, the change of upper troposphere can cause the extreme winter cold spells [38] over northern Europe and the cold air outbreaks in the Northern Hemisphere [22].

Futhermore, periodic patterns, e.g., seasonality, exist in a global spatiotemporal climate pattern. For example, the positive and negative phases of the North Atlantic Oscillation (NAO) predominantly appear in boreal winter over the Northern Hemisphere [15]. In the study of Lanchang-Mekong [27], the sub-region periodicity pattern of temperature and precipitation significantly increased. Looking deeper into the periodic patterns, there are two periodic patterns that are considered in this research; (1) the phase shift, such as the late summer and early autumn [12], and (2) non-stationarity of periodicity [11, 18, 2]. Climate seasonal shifts in terms of spatial and intensity may not be the same every year [35]. The study of El Niño–Southern Oscillation (ENSO) [18] on South African rainfall showed that the periodic cycle may vary in many ranges of years. The study of upstream sub-basins of the Yangtze River [2] also shows statistically significant 2- to 4-year periodicities for mean areal precipitation. To tackle the problem of phase shift and non-stationarity of periodicity patterns, the model should consider the nearby time (months) and multiple cycles.

Although ConvLSTM and other novel spatiotemporal forecasting methods [44, 45, 7] are suitable for spatiotemporal forecasting, they do not consider periodicity, which is an essential component in climate forecasting. Periodic Convolutional Recurrent Network (Periodic-CRN) [54] has been proposed in a different domain (Taxi transportation) as a spatiotemporal forecasting model that makes use of the data's periodicity. The idea is to store the CRN component's output in the periodic representation dictionary (PRD) for re-use. Representations from the dic-

tionary were used through an attention mechanism [4] to improve the forecasting accuracy of the crowd density across two taxi datasets from Beijing and Singapore.

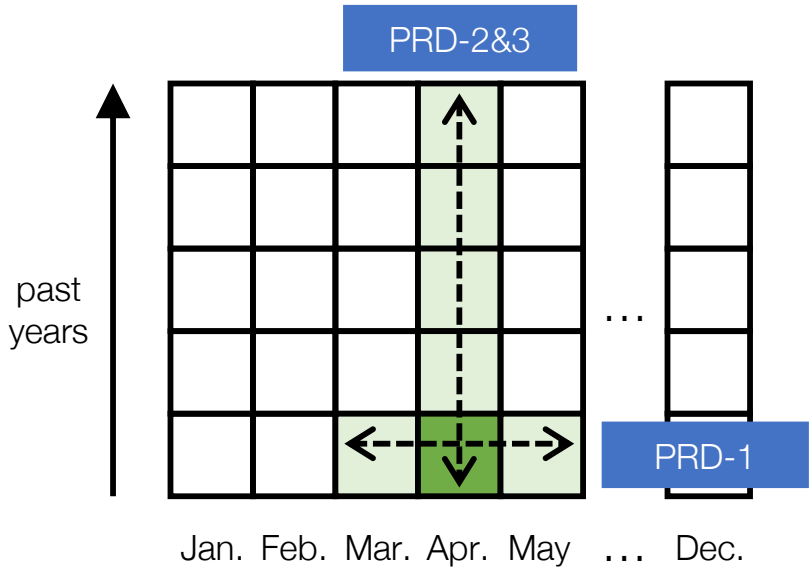


Figure 1.1: The conceptual diagram of the proposed PRD in case of monthly representation. Each cell indicates a saved representation for the future reference. Type PRD-1 refers adjacent months of the last year, and types PRD-2 and 3 refer the same month of prior consecutive years.

In this dissertation, the studies on ML climate predictions are divided into two aspects. The first is to investigate a multivariate spatiotemporal climate prediction for the upper tropospheric circulations over the Northern Hemisphere using ConvLSTM. However, the ConvLSTM that considers only short-range sequential information may not be adequate to deal with periodic patterns such as seasonality. Meanwhile, the Periodic-CRN refers to the same index of the previous cycle, which means "hard" periodicity. Therefore, this study focuses on "soft" yearly periodicity, including nearby time (month) and multiple-year cycles. That leads to the second aspect, Soft Periodic-CRN (SP-CRN) is proposed by adapting from Periodic-CRN with three types of PRD that improve the prediction accuracy by considering soft periodicity: Nearby-Time (PRD-1), Periodic-Depth (PRD-2), and Periodic-Depth-Differencing (PRD-3) representations, as shown in Fig. 1.1. In Periodic-CRN, the spatial weight on the attention module was the only single spatial weight, which is named the static spatial weights (SSW) on attention module in this study. In order to capture the spatial change over the periodicity phase, dynamic spa-

tial weights (DSW) on attention module are proposed. The experiments compared the model performances to predict one-month ahead for ZH300 and sea surface temperature (SST) in an ERA-Interim dataset.

1.2 Contributions

The main contributions of this thesis can be summarized as follows:

(i) **Multivariate spatiotemporal climate forecasting**

This dissertation proposed the utilization of ConvLSTM as a multivariate features climate prediction on predicting the upper tropospheric circulations over the Northern Hemisphere prediction. The ConvLSTM was constructed by an architecture of convolutional layers and deconvolutional layers to predict geopotential height at 300 hPa (ZH300). The experiments were compared with convolutional neural network (CNN) and linear regression (LR). The constructed predictive model ConvLSTM outperformed CNN and LR models in short- and long-term predictions. The result showed that ConvLSTM could maintain spatial correlation and remember important information through the network. The spatial error distribution of prediction output also showed that ConvLSTM has much smaller errors at the paths of subtropical and polar jets. It is conceivable that by considering the earlier six-months' datas the input of ConvLSTM and temporal variations in the ConvLSTM layer can archive the accuracy improvement compared to CNN and LR models.

(ii) **SP-CRN: the spatiotemporal prediction method by utilizing the periodicity component**

The novel method is proposed by extending Periodic-CRN to capture phase shift and non-stationarity of periodicity in climate data. The SP-CRN was proposed by introducing the soft periodicity component, which includes nearby months and a multiple-year cycle for consideration. The proposed SP-CRN has three types of PRD that improve the prediction accuracy by considering phase shift and non-stationarity of periodicity. The CRN component can be any spatiotemporal predictive model, which in this study, the ConvLSTM and

convolution gated recurrent unit (ConvGRU) [5] are utilized in the CRN component to be a comparison method. The experiments were conducted on ZH300 and SST by comparing with CNN, ConvLSTM, ConvGRU, SP-CRN with ConvLSTM and ConvGRU, and all combinations of PRD. The results showed that the prediction accuracy was drastically improved by including the periodicity component in their base model. Moreover, the PRD-3 with the considering change of (periodic) cycles performed the best compared to PRD-1 and PRD-2. Finally, including the metadata at the end of the model improved the prediction accuracy from its based model. The periodicity component proposed in the SP-CRN can better capture the periodicity from climate data and improve prediction accuracy.

(iii) **Dynamic spatial weights on attention module**

The SP-CRN was proposed with static spatial weights (SSW) on attention module, which weight the attention module's output regardless of the periodicity phase (month). In this study, dynamic spatial weights (DSW) on attention module was proposed to investigate and improve the prediction accuracy by assuming that the different spatial weights should be applied according to the periodicity phase. The experiments on ZH300 and SST compared all combinations of SP-CRN between SSW and DSW. The results showed that the DSW improved the prediction accuracy when without metadata.

(iv) **Periodicity analysis**

In order to understand the periodicity pattern for the SP-CRN, the analysis of attention weight is crucial. In this study, the periodicity pattern was analyzed using the attention weight that indicated the periodicity phase and the periodicity cycles. The analysis compared between SSW and DSW of the training dataset. The periodic pattern was similar for PRD-1 of SSW and DSW, and the one prior periodic cycle for one month ahead was more important than the other month. It may indicate that the phase shifts of periodicity have occurred in the range of one month. The PRD-2, which considers only the prior periodic cycle of ZH300 and SST, showed that the past four years' cycle of SSW was most significant, which possibly matches the time-scale variations of ENSO [39]. In contrast,

the DSW showed no significantly periodic pattern for PRD-2.

1.3 Organization of the Dissertation

In Chapter 2, the related literatures about ML prediction method, introduced ML to climate prediction which includes time series and spatiotemporal prediction. Moreover, the utilization of periodicity component in climate prediction model which improve the prediction accuracy.

Chapter 3, introduces the spatiotemporal temporal method called ConvLSTM. This chapter shows the study of upper tropospheric circulations over the Northern Hemisphere prediction using multivariate features by using the ConvLSTM.

Chapter 4, the SP-CRN method which includes the new ideas of periodicity and the extension of attention module. Then the analysis of periodicity component which is important to interpret the periodic pattern from the climate data.

Chapter 5, the last chapter of the thesis summarizes key findings, draws the conclusions along with interesting future research directions.

Chapter 2

Literature Review

2.1 ML Climate Prediction Models

The ML works on spatiotemporal climate forecasting are introduced in this section. Generally, a CNN is the method used to capture the spatial relationship of spatial input. However, it cannot by itself capture spatiotemporal relationships because there is no temporal memory for temporal changed patterns. A ConvLSTM [36] improves an FC-LSTM by capturing both spatial and temporal patterns, and was applied to precipitation nowcasting (short period, up to 6 hours). Since then, spatiotemporal climate forecasting has widely mentioned ConvLSTM as a baseline method [44, 45, 7] or adopted in other model architectures [20, 19, 21]. Predictive RNN (PredRNN) [44] adds extra connections between adjacent time steps in a core stack of spatiotemporal LSTM (ST-LSTM), which outperforms ConvLSTM in precipitation nowcasting. Memory-in-memory (MIM) networks [45] improve the learning of non-stationarity properties in spatiotemporal input, such as spatial correlations or temporal dependencies of local pixel values for low-level non-stationarity, and the accumulation, deformation or dissipation of radar echoes in precipitation forecasting for high-level non-stationarity. MIM captures those non-stationarity properties by adding stationary and non-stationary modules separately inside the ST-LSTM block [44]. Lastly, the most recent work on spatiotemporal climate forecasting is the spatiotemporal convolutional sequence-to-sequence network (STConvS2S) [7]. STConvS2S comprises three components, the sets of convolutional layers responsible for different tasks: the temporal block, which learns the temporal representation of sequential input by two approaches (temporal casual block [41] and temporal reversed block); spatial block, which extracts the

spatial features from the temporal block; and temporal generator block, designed to output a sequence of arbitrary length. The STConvS2S outperformed or was competitive with ConvLSTM, PredRNN, MIM models for temperature (CFSR[1]) and rainfall (CHIRPS[2]) prediction.

2.2 ML Climate Prediction Models with Periodicity Component

Secondly, the ML works on climate forecasting that utilize the periodicity component to improve prediction accuracy are introduced in this section. A seasonally integrated autoencoder (SSAE) [31] is proposed to improve short-term daily precipitation forecasting. SSAE architecture has two LSTM autoencoders: the short-term autoencoder, which learns the short-term sequential input (less than seven days), and the seasonal autoencoder which captures the seasonal change in a long-term sequential input long enough to cover the period of the wet or dry season (60–120 days). The other way is to introduce the metadata idea to the network. The prediction of a streamflow in a river [1] improves the prediction results by adding a digit for the month as one of inputs in the final position of the sequential input of LSTM, Extreme Learning Machines (ELM), and Random Forest (RF) algorithms. Committee ELM (Comm-ELM) [3] also forecasts the monthly standardized precipitation index (SPI) by adding a digit for the month to indicate yearly periodicity at the end of a sequential input to improve the prediction. However, all of these methods assume stationary periodicity, which means they can not capture the phase shift or non-stationarity of periodicity patterns.

Meanwhile, in time series analysis, the Autoregressive Integrated Moving Average (ARIMA), a statistical prediction model, is applied to a non-stationary time series to eliminate the non-stationary mean function (trend) corresponding to the "integrated" part of the model. ARIMA [25] is used in near-term regional temperature and precipitation forecasting. Seasonal ARIMA (SARIMA) [46] is an extension of ARIMA that supports a seasonal component for time series prediction. It adds difference operators to remove the seasonal component to transform the non-stationary time series into a stationary one. SARIMA analysis was found are acceptable for predicting the monthly mean temperature of Nanjing, China [9]. SARIMA was also used

[1] https://climatedataguide.ucar.edu/climate-data/climate-forecast-system-reanaly sis-cfsr
[2] https://chc.ucsb.edu/data/chirps

in the precipitation and temperature predictions at the Tehri and Uttarkashi stations in the Bhagirathi river basin in the state of Uttarakhand, India [14]. SARIMA was applied to a coastal Tuscany watershed to study hydrological cycle changes at the local scales (area) [53]. Furthermore, SARIMA was applied for the spatiotemporal analysis in the Highlands region, Algeria, to predict the local scale of precipitation [6]. However, SARIMA does not work well on multiple seasonalities [47], the daily and yearly patterns in the same time series of a climate variable (e.g., surface temperature). Empirical mode decomposition (EMD) [17] can also modify such a time series, from high to low frequency, into multi-resolution intrinsic mode functions (IMFs) while leaving the monotonous time series as the estimated trend time series. Ensemble EMD (EEMD) [23] is a variant of EMD, and is used to extract the seasonal component from a time series by selecting the IMFs of the decomposed original time series that corresponds to the seasonal averaged model.

2.3 Base Methods

2.3.1 Convolutional LSTM

Thirdly, this section introduces Convolutional LSTM (ConvLSTM) [36], one of the most popular and widely used spatiotemporal climate prediction methods. ConvLSTM has been developed and extended from fully connected LSTM (FC-LSTM) by making every gate in LSTM memory cell to be a 3D tensor that can maintain spatial correlation through the network. The advantage of LSTM structure inside ConvLSTM is an ability to remember important information while learning from the sequence data. Fig. 2.1 shows the architecture of the ConvLSTM that is used in Chapter 3.

Unlike recurrent neural network (RNN) that the error term in the backpropagation process of the long-term recurrent network is multiplied by less than one for many times along the recurrent network, causing long-term of the error term will not be back-propagated, this problem is called "vanishing gradient problem."

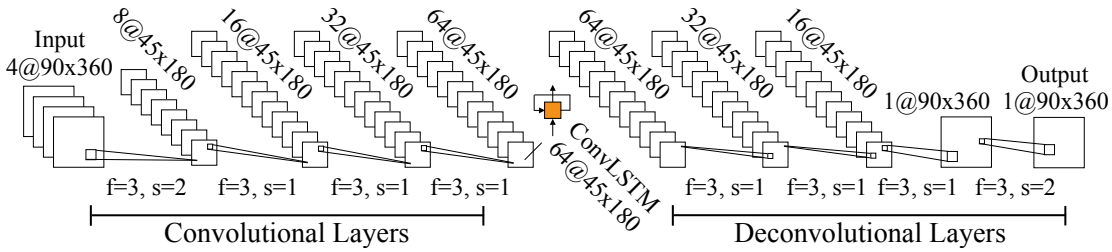


Figure 2.1: ConvLSTM structure on the experiment in Chapter 3

The spatial input data will be fed into ConvLSTM in each time step and can be seen as 3D tensor, where the first dimension is temporal information, and the last two dimensions are spatial information. Convolution operator has been used in this network to capture spatial correlation in every gate of the networks. Eqs. (2.3.1) are those of ConvLSTM, where W_x, W_h and W_c denote a set of weights (kernels) of current input, hidden state, and cell output, respectively. The 3D tensors $x^{(t)}, h^{(t)}, c^{(t)}$, and gates $i^{(t)}, f^{(t)}, o^{(t)}$ are input, hidden state, cell output, input gate, forget gate, and output gate, respectively. Moreover, b denotes bias, σ denotes an sigmoid function, $*$ denotes a convolution operator and $\circ$ denotes a Hadamard product.

$$
\begin{aligned}
i^{(t)} &= \sigma(W_{xi} * x^{(t)} + W_{hi} * h^{(t-1)} + W_{ci} \circ c^{(t-1)} + b_i) \\
f^{(t)} &= \sigma(W_{xf} * x^{(t)} + W_{hf} * h^{(t-1)} + W_{cf} \circ c^{(t-1)} + b_f) \\
c^{(t)} &= f^{(t)} \circ c^{(t-1)} + i^{(t)} \circ \tanh(W_{xc} * x^{(t)} + W_{hc} * h^{(t-1)} + b_c) \\
o^{(t)} &= \sigma(W_{xo} * x^{(t)} + W_{ho} * h^{(t-1)} + W_{co} \circ c^{(t)} + b_o) \\
h^{(t)} &= o^{(t)} \circ \tanh(c^{(t)})
\end{aligned}
\tag{2.3.1}
$$

The zero-padding was applied to remain an output to be the same size as the input after convolution operation inside the network. All hidden states were initialized as zero to ignore the state before training.

2.3.2 Convolutional GRU

Then, the Convolutional Gated Recurrent Unit (ConvGRU) [5] is considered as the spatiotemporal prediction model to use on climate data, which is introduced in this section. Gated-Recurrent-Unit (GRU) networks are a particular type of RNN to overcome the difficulty of RNNs to train

due to the exploding or vanishing gradient effect. The GRU and LSTM have empirically demonstrated their ability to model long-term temporal dependency in various tasks such as machine translation or image/video caption generation. GRU networks allow each recurrent unit to capture dependencies of different time scales adaptively. Eqs. 2.3.2 are those of GRU, where $\odot$ is an element-wise multiplication. z_t is an update gate that decides the degree to which the unit updates its activation, r_t is a reset gate, and σ is the sigmoid function. $\tilde{h}_t$ is a candidate activation that is computed similarly to that of the traditional recurrent unit in an RNN.

$$\begin{aligned} z_t &= \sigma(W_z x_t + U_z h_{t-1}) \\ r_t &= \sigma(W_r x_t + U_r h_{t-1}) \\ \tilde{h}_t &= \tanh(W x_t + U(r_t \odot h_{t-1})) \\ h_t &= (1 - z_t) h_{t-1} + z_t \tilde{h}_t \end{aligned} \tag{2.3.2}$$

However, fully-connected GRUs do not take advantage of the underlying structure of convolutional maps which are the spatiotemporal input. Convolution operations are added to the GRU to consider the spatiotemporal input. This structure named GRU-Recurrent Convolution Network (GRU-CRN) and named it as ConvGRU in this study (well-known name). Eqs. 2.3.3 are those of ConvGRU, where * denotes a convolution operation. Parameters W^l, W_z^l and W_r^l have a size of $k_1 \times k_2 \times O_x \times O_h$ where $k_1 \times k_2$ is the convolutioanl kernel spatial size. The candidate hidden representation $\tilde{h}_t(i,j)$, the activation gate $z_k(i,j)$ and the reset gate $r_k(i,j)$ are defined based on a local neighborhood of size $(k_1 \times k_2)$ at the location (i,j) in both the input data x_t and the previous hidden-state h_{t-1}.

$$\begin{aligned} z_t^l &= \sigma(W_z^l * x_t^l + U_z^l * h_{t-1}^l) \\ r_t^l &= \sigma(W_r^l * x_t^l + U_r^l * h_{t-1}^l) \\ \tilde{h}_t^l &= \tanh(W^l * x_t^l + U * (r_t^l \odot h_{t-1}^l)) \\ h_t^l &= (1 - z_t^l) h_t^l - 1 + z_t^l \tilde{h}_t^l \end{aligned} \tag{2.3.3}$$

The main difference between ConvLSTM and ConvGRU is that the ConvLSTM has three gates while the ConvGRU has two gates. The ConvGRU uses fewer training parameters and

therefore uses less memory and executes faster than ConvLSTM, whereas ConvLSTM is more accurate on a larger dataset. In this study, ConvGRU is used as the comparison method for spatiotemporal climate prediction model.

2.3.3 Periodic Convolutional Recurrent Network

Lastly, this study considered adopting spatiotemporal prediction model with periodicity component from other domain to perform on climate domain. This section introduces Periodic Convolutional Recurrent Network (Periodic-CRN) [54] which is the spatiotemporal prediction method for crowd density prediction in spatiotemporal domain. Specifically, the model predicts the next state of a geo-spatial space, given a sequence of previous states. The key idea is using periodic representations of previously observed recurring patterns with tensors, rather than using raw input data from those periods in the learning process. The periodic representations is used to indicate the periodicity of the data which can improve the prediction accuracy in taxi transportation dataset.

The Periodic-CRN consists of three components: (1) Convolutional recurrent network, (2) Periodic representation, (3) Fusion process. The entire architecture has shown in Fig. 2.2.

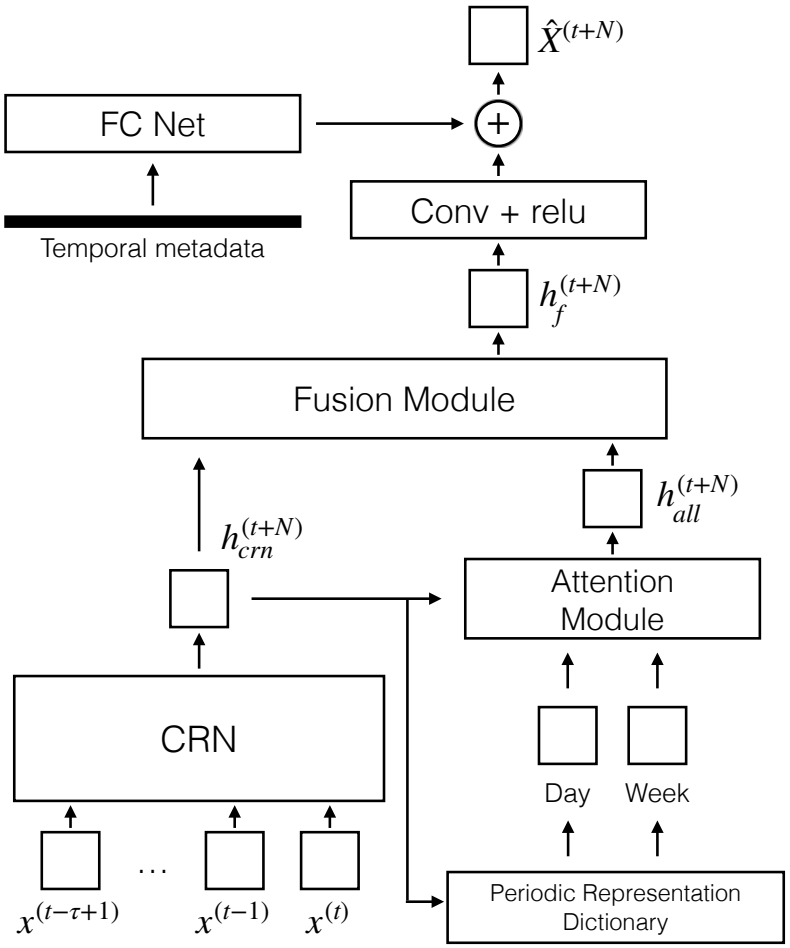


Figure 2.2: The Periodic-CRN architecture

Convolutional Recurrent Network

To capture spatial and temporal correlations is crucial for building accurate spatiotemporal prediction models. As in many geo-spatial problems such as crowd density or traffic congestion prediction, the state of traffic is one region in evitably influences nearby regions. The Periodic-CRN considered using Convolutional GRU [5], and adopt a pyramidal structure [8] for the multi-layer Convolutional GRUs. This modification reduces the sequence length of the upper layer by a factor of 2, which results in significant speedup without negatively affecting performance.

Periodic Representation

Periodic representation (P) is a hidden state from CRN component. To build a model that has an ability to consider a periodic pattern, Periodic-CRN stores the current periodic representation ($h_c^{(t+1)}$) to periodic representation dictionary (D). The length of dictionary (l_D) is a cycle of periodicity (e.g., a yearly periodic pattern has 12 bins of months). The periodic representation will be loaded and saved to D where the dictionary index is matched in Eq. (2.3.4). t denotes time index, N_P denotes index of periodicity cycle, and $h_P^{(t+1)}$ is a loaded periodic representation from D. $\Leftarrow$ and $\Rightarrow$ denote the direction of representation that load and save to D respectively.

$$\begin{aligned} h_P^{(t+1)} &\Leftarrow D[\mid t - (l_D * (N_P - 1)) \mid] \\ h_c^{(t+1)} &\Rightarrow D[\mid t - (l_D * (N_P - 1)) \mid] \end{aligned} \tag{2.3.4}$$

The nearby time index representation that loaded from D may improve accuracy result instead of using only a single periodic time index representation.

Fusion Process

This process is to merge periodic representation with the output from the CRN component, then merge with the temporal metadata.

Weighting Based Function: In this method, $h_c^{(t+1)}$ and $h_{P_j}^{(t+1)}$ are merged, where j denotes nearby time representation. The attention mechanism is used to weight multiple h_{P_j}. The scalar weight a_n is learned from the network indicating importance of each representation h_{P_j}, where $a_n \in [0,1]$ and $\sum_n a_n = 1$ according to the softmax operation. The $h_{all}^{(t+1)}$ denotes periodic representation output after attention mechanism process. Weight W_e is convolutional kernel size of $1\times1\times2$ that reduces the representation channels to one, and W_f size is equal to the size of flatten operation output.

Then periodic representation h_{all} and representation output $h_c^{(t+1)}$ are combined in Eq. (2.3.8).

$$e_j^{(t+1)} = W_f.flatten(W_e * [h_c^{(t+1)}, h_{P_j}^{(t+1)}]) + b_f \tag{2.3.5}$$

$$a_j^{(t+1)} = \frac{\exp(e_j^{(t+1)})}{\sum_{j'} \exp(e_j^{(t+1)})} \tag{2.3.6}$$

$$h_{all}^{(t+1)} = \sum_j a_j^{(t+1)} h_{P_j}^{(t+1)} \tag{2.3.7}$$

$$h_f^{(t+1)} = relu(W_c \circ h_c^{(t+1)} + W_P \circ h_{all}^{(t+1)}) \tag{2.3.8}$$

After the fusion process of attention mechanism, the convolution layer is used to transform the representation $h_f^{(t+1)}$ to be the same size as an input. In Eq. (2.3.9), the size of W_f and $h_f^{(t+1)}$ are the same size as the input.

$$h_f^{(t+1)} = relu(h_f^{(t+1)} * W_f) \tag{2.3.9}$$

Metadata Fusion: In this method, timestamp of records are used to indicate which periodic representation is currently used. One-hot encoded ($X_{meta}^{(t+1)}$) is defined as temporal metadata indicates month of the year, the size of $X_{meta}^{(t+1)}$ is 12×1 (for yearly periodic pattern). Then $X_{meta}^{(t+1)}$ is passed to two fully connected layers as shown in Eq. (2.3.10). After reshaping $h_{meta}^{(t+1)}$ from 1-D to 2-D to be the same size as the input, $h_{meta}^{(t+1)}$ with $h_f^{(t+1)}$ are merged to get the prediction output ($\hat{X}^{(t+1)}$).

$$h_{meta}^{(t+1)} = relu(relu(X_{meta}^{(t+1)}.W_{fc_1} + b_{fc_1}).W_{fc_2} + b_{fc_2}) \tag{2.3.10}$$

$$\hat{X}^{(t+1)} = relu(h_f^{(t+1)} + h_{meta}^{(t+1)}) \tag{2.3.11}$$

2.4 Summary

In summary, ConvLSTM and spatiotemporal climate forecasting models [44, 45, 7] consider only sequential input, which is the short sequence that does not cover the periodicity patterns. SSAE [31] focuses on the long sequence that covers the entire season, but the model does not regard changes in multiple periodic cycles. SARIMA [46], EMD [17], and EEMD [23] do not

address the spatial dependencies, and the decomposed time series by EMD and EEMD may not refer to the seasonality of the observations.

Meanwhile, Periodic-CRN [54] has been proposed as a spatiotemporal forecasting model in a taxi transportation domain, where it will use daily and weekly periodicity components along with the load and save mechanism in the PRD. Although Periodic-CRN improves prediction accuracy by using periodicity, it assumes stationary periodicity as PRD refers to exactly the same index of the previous cycle, which means "hard" periodicity. This dissertation proposed SP-CRN by adapting the PRD to be a decent mechanism in the non-stationary climate domain. The PRD approaches are novel ideas that consider phase shift and the non-stationarity of periodicity. This dissertation focuses on "soft" yearly periodicity, which includes nearby months and a multiple-year cycle, and use of three proposals from the PRD, satisfies climate data behavior.

Chapter 3

Climate Prediction using Multivariate Features by ConvLSTM

3.1 Research Propose

This chapter studies the prediction of upper tropospheric circulations over the Northern Hemisphere by predicting ZH300, which relates to the polar jet stream affecting the strength and the path of mid-latitude cyclones.

This chapter studies the availability of utilizing the deep learning methods for spatiotemporal climate prediction. The study includes the prediction of upper tropospheric circulations over the Northern Hemisphere by predicting ZH300, which relates to the polar jet stream affecting the strength and the path of mid-latitude cyclones. The study also investigates the availability of ConvLSTM in maintaining the spatiotemporal correlation through the network on short- and long-term prediction of ZH300. Furthermore, the sequential input on ConvLSTM may help the model capture the change of the season by investigating the projection of prediction output. Also, the advantage of sequential input on ConvLSTM in performing the long-term prediction over the model does not consider the sequential input. Moreover, the study also investigates the importance of prediction features to predict ZH300 using multivariate features for the ConvLSTM.

3.2 Experiment Settings

3.2.1 Dataset

This study uses ERA-Interim[1] dataset from the European Centre for Medium-Range Weather Forecasts (ECMWF) which is a global atmospheric reanalysis from 1979 as the main dataset. This dataset contains a monthly average of climate variables in spatial grid data. The study also use NINO 3 SST Index which is anomaly area-averaged sea surface temperature (SST) from 5°S-5°N and 150°-90°W, that indicates El Niño event. The four relevant features are selected for ZH300 prediction: temperature at 300 hPa (T300), T850, ZH300, and NINO 3 index. All features are extracted to 360×90 image resolution, and 480 months started from 1979 until 2018.

ZH300 is the geopotential height of pressure level 300 hPa, which is in different places have a different height of ZH300. ZH300 on the Northern Hemisphere can indicate the jet stream wind around the height of 9-12 km, which is essential to be an indicator of global seasonality and other events of climate such as cold air outbreaks in the Northern Hemisphere [22]. To make a prediction of ZH300, four relevant prediction features have been considered. T300 is the temperature at the same level of ZH300, the change of ZH300 is relative to the change of ZH300 because the jest stream wind is separate high (low temperature) and low (high temperature) pressure. T850 is the temperature at around the ground level, it also affect the change of jet stream wind. NINO 3 is the sea surface temperature index at the pacific ocean which is the most common used indices to define El Niño and La Niña. Lastly, the prior ZH300 that indicates the prediction variable itself (ZH300).

3.2.2 Comparison Methods

The experiments are conducted to study the prediction availabilities of ZH300 on the Northern Hemisphere, which has a spatial correlation for an entire area. It will be difficult for a model that is not able to capture spatial correlation to turn out the accurate pattern on prediction output. Three models were experimented as follows:

1. Linear regression model (LR): This model is the most straightforward linear prediction

[1] https://apps.ecmwf.int/datasets/data/interim-full-moda/levtype=pl/

algorithm that works well on linear relationship data. The model was considered as a baseline method for linear prediction algorithm. In this study, LR models were trained separated 360×90 so that the model could not capture the spatial correlation of the input data. The model was trained by learning from 300 months prior to predict the next month. In conclusion, this model used 4 features of 300 months, i.e., 1,200 prediction features.

2. Convolutional neural networks (CNN) [24]: This study constructed four convolutional layers to encode four channels of input data and decoded with four deconvolutional layers to obtain one output channel. 3×3 kernel size was applied to this model. The entire structure is shown in Fig. 2.1 but without ConvLSTM structure.

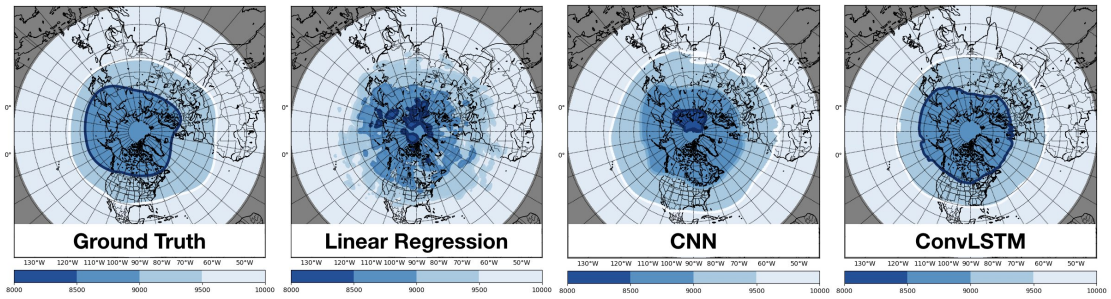
3. ConvLSTM: The same structure as CNN was used, but in the middle between convolutional and deconvolutional layers was ConvLSTM structure as shown in Fig. 2.1.

The input data was fed into both CNN and ConvLSTM were a spatial grid data. NINO 3 index was a scalar variable in each month, in order to apply the NINO 3 index to be the spatial grid data, the NINO 3 variable was applied as the same value for all grid points. In this experiment setting, CNN learned one-month data to predict next month, but ConvLSTM learned six-months to predict next month. The six-months input has been considered because it is a long period enough to the change of seasons. The models were trained with 462 months as a training set and 18 months as a test set aiming to predict 12 months in 2018. Then the models were trained until converged or reached the limit of 200,000 steps. Dropout was applied to the models to reduce overfitting and used adaptive moment estimation optimizer (Adam optimizer) as an optimizer.

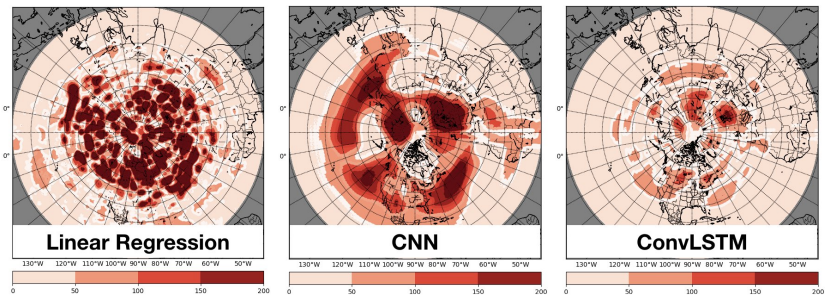
3.3 Results

3.3.1 Short-term prediction

Fig. 3.1a shows the output result of April 2018, which is the best month of ConvLSTM method. LR does not capture ZH300 pattern properly, while CNN and ConvLSTM perform better in capturing the spatial pattern.



(a) Ground truth and prediction result



(b) Error distribution

Figure 3.1: The Northern Hemisphere ZH300 on prediction and error distribution on April, 2018 (the best month of ConvLSTM)

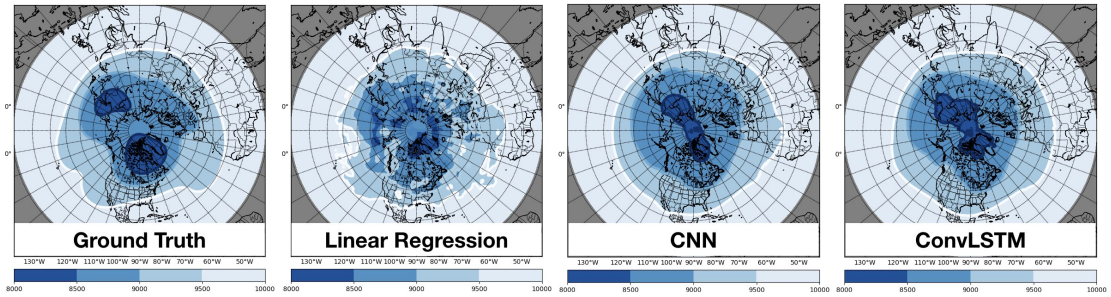
The spatial error distribution of CNN (Fig. 3.1b) clearly indicates that the error distribution of CNN is large at the paths of the subtropical and polar jets where the meridional ZH300 gradient is large, and the variability of ZH300 is also large. The large variability of ZH300 is coherent with the strong activity of extratropical cyclones there.

Thus, the CNN prediction with the one-month earlier data can not predict the variability of the subtropical and polar jet streams correctly, which is critical for climate prediction. The spatial error distribution of ConvLSTM, on the other hand, shows much smaller errors even on the paths of the subtropical and polar jets. It is conceivable that this improvement was achieved by considering the earlier six-months' data and these temporal variations in the ConvLSTM layer.

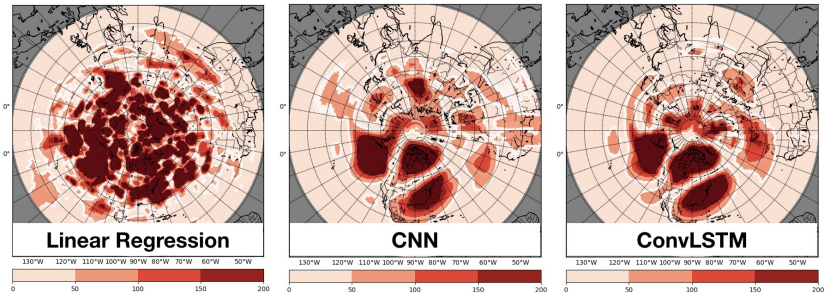
The experiments tested on six-months input to CNN, which was equivalent to ConvLSTM input. The four features of six-months spatial data are arranged as the sequential input of CNN. The result was worse than learning with one-month in short- and long-term prediction. This might because of the 24-channels of input data is too much for the network, and the network

can not capture a transition of the same feature in different month input.

The spatial error distribution of LR seems to be like the distribution of white noise in mid- and high-latitudes, and does not show a systematic distribution pattern. This distribution is probably because the LR prediction of this study did not include the spatial correlation.



(a) Ground truth and prediction result



(b) Error distribution

Figure 3.2: The Northern Hemisphere ZH300 on prediction and error distribution on February, 2018 (the worst month of ConvLSTM)

Table 3.1: RMSE (meters) of 12 months in 2018.

Month	LR	CNN	ConvLSTM
1	156.40	57.11	78.73
2	164.15	112.60	117.36
3	143.30	113.76	98.60
4	145.93	117.70	58.12
5	155.28	142.50	86.86
6	143.54	127.54	61.41
7	153.44	130.47	76.97
8	140.94	68.35	59.97
9	162.33	97.65	84.04
10	148.86	111.71	51.53
11	158.88	120.64	76.12
12	170.32	112.13	78.60
Average	153.61	109.35	77.36

Fig. 3.2 shows the result of February 2018, which is the worst month of ConvLSTM. The error distribution on Fig. 3.2b shows that high error of CNN and ConvLSTM are in the same area. It might because of a sudden stratospheric sarming that occurred on February, 2018 [32] that exceed the limit of CNN and ConvLSTM model to capture such kind of change.

Table 3.1 is RMSE of 12 months (compare to the ground truth) in 2018. The result shows that ConvLSTM outperforms with RMSE of 77.36 meters (0.84% comparing to average ZH300 value of 9,225.19 meters) over CNN and LR that have RMSE of 109.35 (1.19%) and 153.61 (1.67%) meters, respectively. LSTM network inside ConvLSTM helps the networks remember important information for a long period of time. However, the RMSE in each month of CNN and ConvLSTM do not indicate high accuracy on certain months.

The predicted value minus the ground truth for all months in 2018 were plotted to investigate over and underestimation of CNN and ConvLSTM models, as shown in Fig. 3.3. The overestimated months mostly appear in winter and underestimated months appear in summer for CNN. On the other hand, the ConvLSTM showed the over and underestimated at about the same amount in each month (the amount of red and blue area are similar for every month). Simple CNN would not even have an ability to predict the regular (or climatological) seasonal change in ZH300 which increases to summer and decreases to winter over the Northern Hemi-

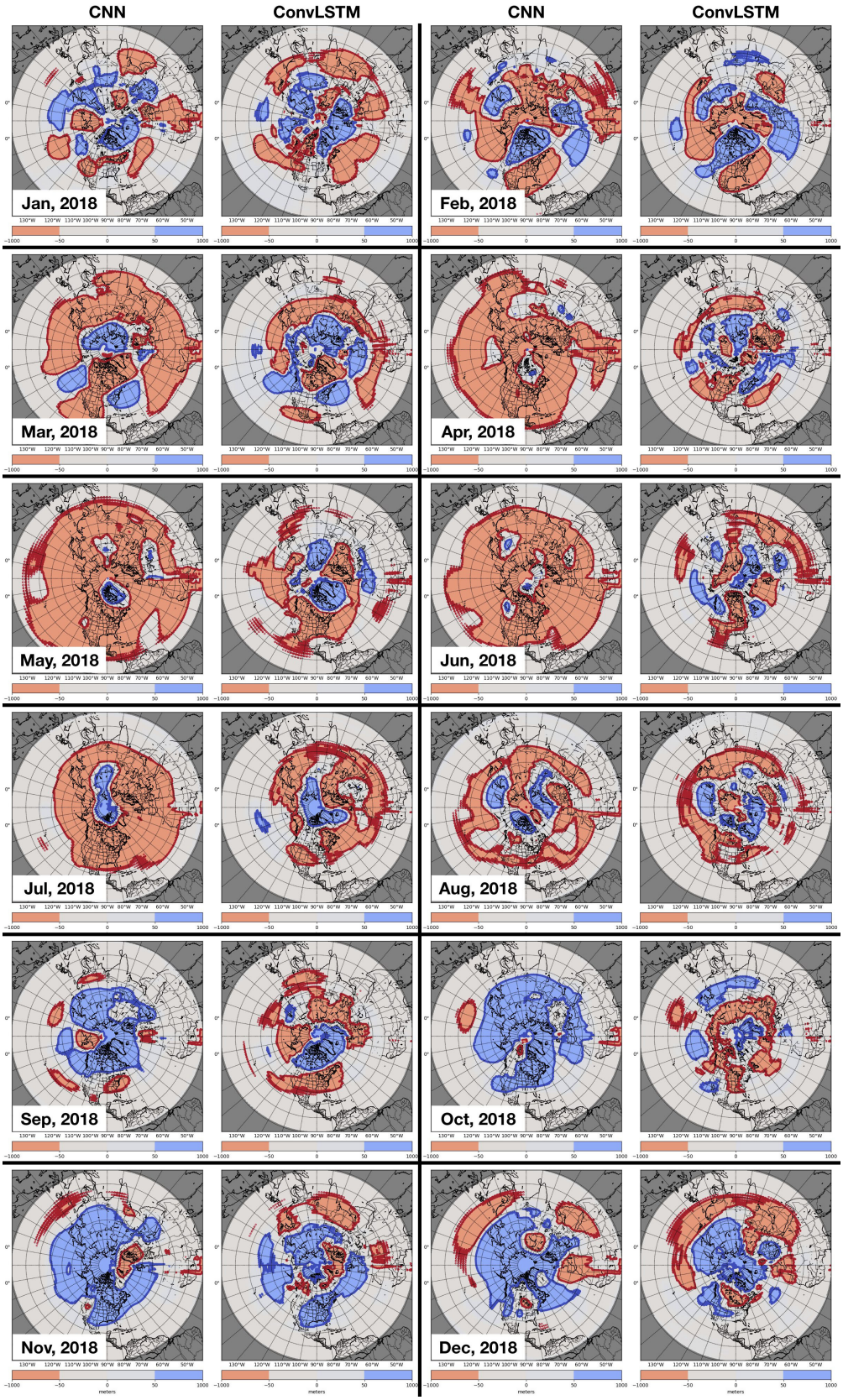


Figure 3.3: The predicted value minus ground truth in every month of 2018 (blue color is overestimated and red color is underestimated). In every month, left figure is CNN and right figure is ConvLSTM.

sphere polar region.

3.3.2 Long-term prediction

Next, models were applied to predict the next 1-5 months. The result is shown in Fig. 3.4, RMSE of ConvLSTM is only slightly increased, showing the robustness over the long-term prediction. The RMSE of ConvLSTM is less than 1% to average ZH300 value even for long-term prediction (the next 5 months prediction has RMSE of 1.02%, only). CNN has a better performance than LR in short-term prediction but getting worse in the long-term, this might because of the number of features for LR is 1,200 features (300 months prior $\times$ 4 features), while CNN uses only one month of 4 features. Although the LR maintains the RMSE well in long-term prediction, the output result shows poorly capturing ZH300 pattern. The experiment also tested on 6 months input to CNN, which was equivalent to ConvLSTM input, but the result was worse than learning with 1 month in short- and long-term prediction.

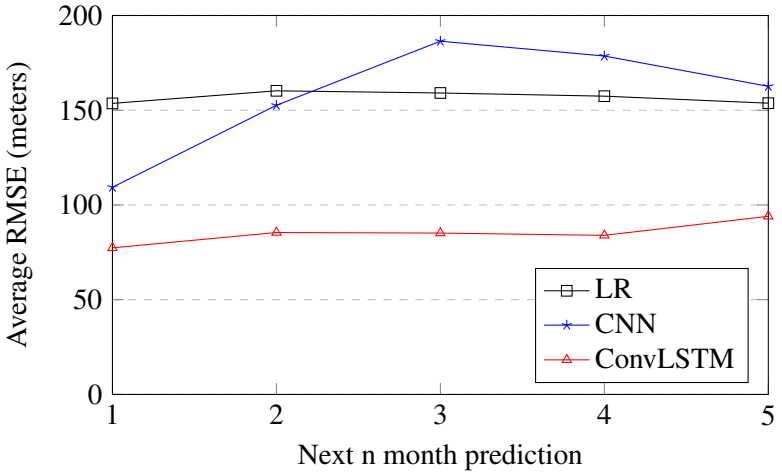


Figure 3.4: RMSE of long-term prediction

3.3.3 Prediction features' investigation

This study also investigated the importance of features to predict ZH300. The experiment tested with ConvLSTM on 4 datasets; each of them excluded 1 feature. If the excluded feature was important to ZH300 prediction, the RMSE result should be worse (RMSE higher than using 4 features).

Table 3.2 shows the result of excluding 1 feature out of 4 features. The standard deviation

(SD) shows the dispersion to the average RMSE value. By excluding T300 or ZH300, which are temperature at the same hight as ZH300 and prior ZH300, have higher RMSE of 79.42 and 81.30 meters than that of using all the 4 features, respectively. Therefore, these 2 features are relevant to ZH300 prediction. On the other hand, excluding T850 shows the lower RMSE of 74.59 meters, which means that T850 interferes with ZH300 prediction and should be excluded from the dataset.

Table 3.2: Excluded features result of average RMSE of 12 months in 2018.

Dataset (Exclude)	Avg. RMSE (m)	SD of RMSE (m)
ZH300	81.30	17.21
T850	74.59	15.11
T300	79.42	17.32
NINO 3	77.08	18.13
Includes all 4 features	77.36	18.55

Moreover, SD of excluding T850 is lowest that shows the error in every month is close to the average RMSE, which is good for prediction (low RMSE and low SD). The result of excluding NINO 3 has about the same RMSE of 77.08 comparing to included 4 features, which seems less relevance to ZH300 prediction. However, it might be the cause of how the scalar value of NINO 3 was implemented to be a spatial grid data that is the same value on every data point; different implementation of NINO 3 index may affect to the result.

Lastly, the experiment was conducted using ConvLSTM to predict ZH300 in short- and long-term prediction (next 1-5 months prediction) by using the most 2 relevant prediction features (T300 and ZH300). The result shows that using 2 relevant prediction features obtained the lower average RMSE of 82.92 (0.9%) comparing to average RMSE of 85.194 (0.92%) using 4 features.

3.4 Discussion

The results show that the deep learning methods were able to perform the spatiotemporal climate prediction. Especially, ConvLSTM obtained the best prediction accuracy compare to CNN

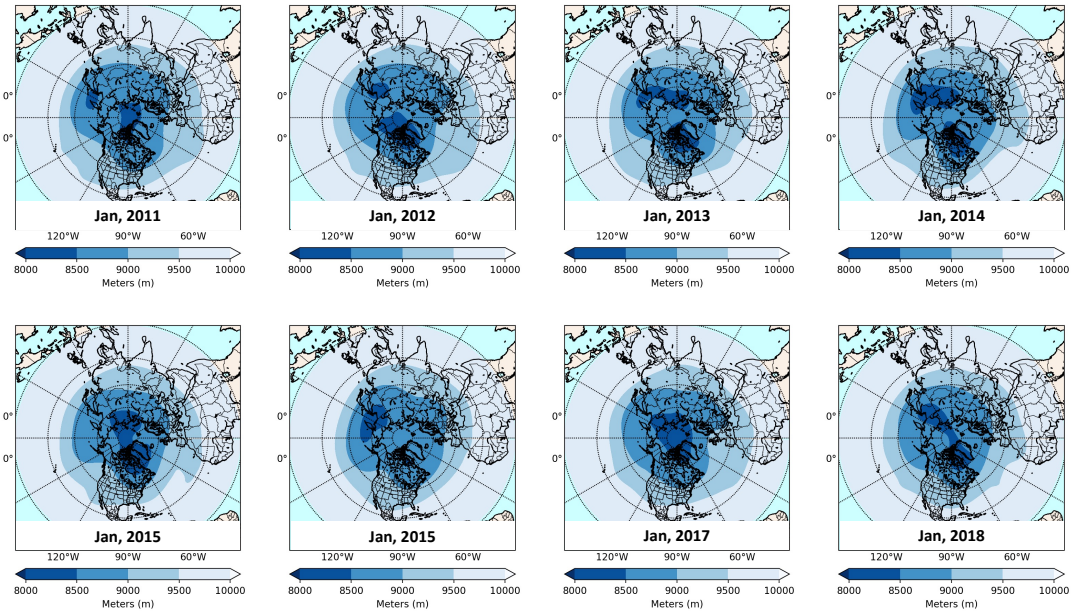


Figure 3.5: The comparison of ZH300 observed data in January from 2011 to 2018.

and LR. ConvLSTM also shows the performance of maintaining the spatiotemporal collation through the network on short- and long-term prediction. The results of investigation the importance of prediction features to predict ZH300 show that the prior ZH300 was the most important over T300, NINO 3, and T850, respectively.

Although ConvLSTM obtained a better prediction than CNN and LR, the model cannot capture the periodic patterns by learning from the sequential spatial input. Fig. 3.5 shows the comparison of ZH300 observed data in January from 2011 to 2018. The same month has similar spatial patterns, which indicate the periodic patterns across years. Also, Fig. 3.6 shows that the spatial patterns of every month in the same year are different. By learning the consecutive sequential input, the model can not learn the periodic patterns hidden in the sequential input. Despite extending the sequential input to cover the length of periodicity may not be helpful. To capture the periodicity patterns, the model needs some specific components to operate. Therefore, the following section introduces the spatiotemporal climate prediction model, which utilizes the periodicity component.

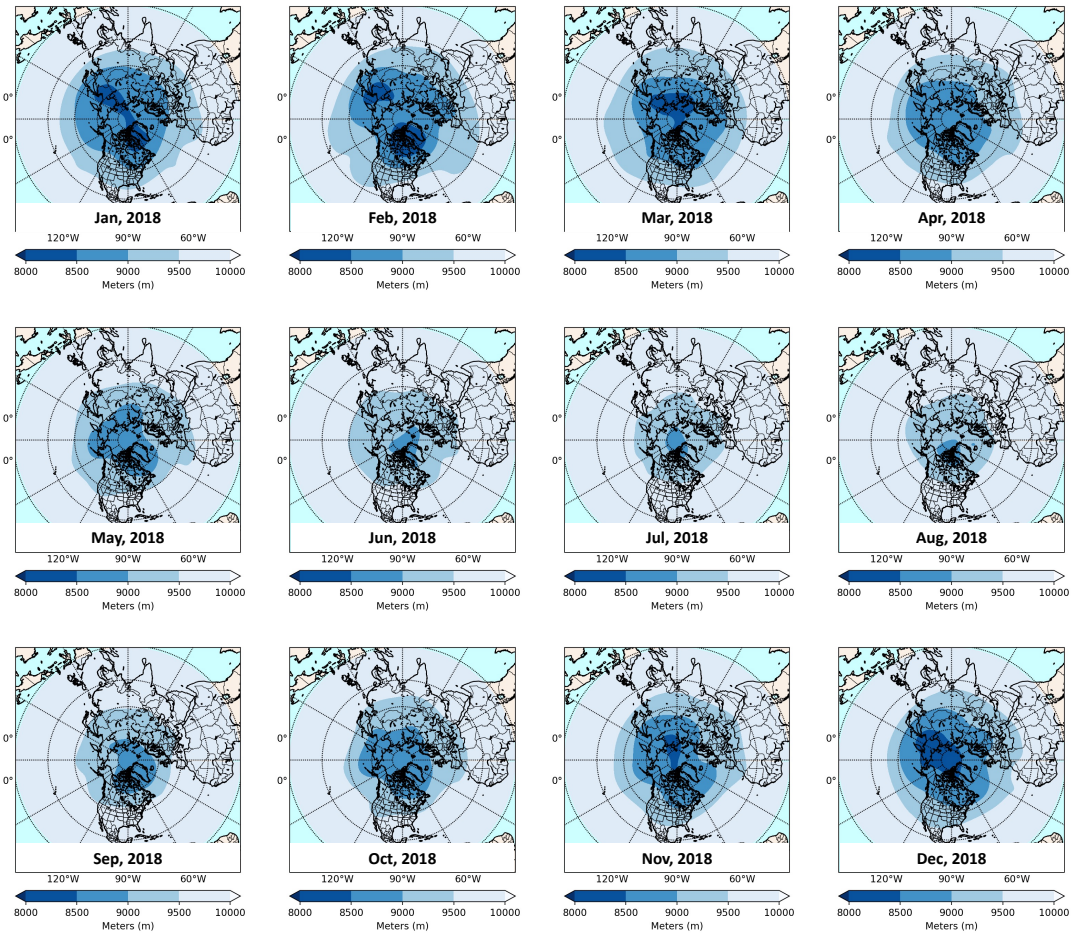


Figure 3.6: The comparison of ZH300 observed data in every month of 2018.

Chapter 4

Soft Periodic Convolutional Recurrent Network

4.1 Overview

In this research, the original Periodic-CRN was modified as two components: (1) CRN and (2) periodic representation. The entire architecture is shown in Fig. 4.1. The process of this method begins with learning the sequential input in the CRN component (Fig. 4.1(a)). Then, the hidden state output from the CRN component is saved as a periodic representation (Fig. 4.1(b)), and it is loaded according to the proposal described in Section 4.4.2. Next, to estimate the relevance of each periodic representation to the hidden state output, the representations are weighted by the attention module (section 4.4.5) as the representation output from the periodic representation component. Finally, the hidden state output of the two components is combined (fusion module in Fig. 4.1(f)), then reshaped with convolution operation as described in Section 4.4.6 to obtain the prediction.

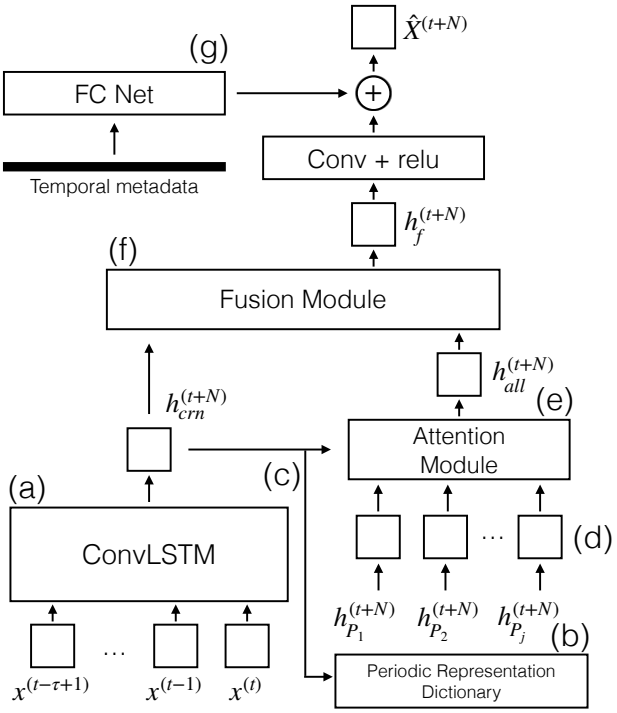


Figure 4.1: The SP-CRN (LSTM) architecture. (**a**) the CRN component, (**b**) the PRD, (**c**) update mechanism, (**d**) load mechanism, (**e**) the attention module, (**f**) fusion module, and (**g**) metadata component.

4.2 Problem Statement

Regarding climate prediction task as similar to a next video frame prediction [52], where input is a sequence of spatial 2-dimensional data, and the output is spatial data of the next timestep. Climate data for a geo-spatio-temporal domain is aggregated into multichannels of a 2-dimensional data $\hat{X}^{(t)} \in \mathbb{R}^{M*W*H}$, where M, W, H are the number of channels, width (longitude), and height (latitude), respectively, and t is a data time index.

Eq. (4.2.1) is a problem statement, where $\hat{X}^{(t+N)}$ is a spatial climate variable at time $t+N$, where N is a natural number. θ is model parameters of the function F; τ is a natural number of a pre-defined sequence length ($\tau = 1,2,3,...$), and $X^{(t)},...,X^{(t-\tau+1)}$ are spatial sequence input data.

$$\hat{X}^{(t+N)} = F(X^{(t)},...,X^{(t-\tau+1)};\theta) \tag{4.2.1}$$

4.3 Soft Periodicity

This section introduces soft periodicity, including the definition, example on a sine wave, and autocorrelation on real climate data.

Definition 1 (Soft Periodicity)**.** *A function f is said to be soft periodic if, for some nonzero constant period P and for some continuous function that changes according to k value $\Delta(k)$, it is the case that $f(k+P+\Delta(k)) = f(k)$ for all values of k in the domain. Here, $\Delta(k)$ is assumed to be much less than P ($\Delta(k) \ll P$). Also, the average of $\Delta(k)$ for all values of k is zero.*

This section demonstrates and describes the soft periodicity based on a sine wave. In Eqs. 4.3.1, $y(t)$ is a value at time t, where $\sin(t)$ denotes a sine function which represent a constant value of the waveform, $\Delta(t)$ denotes a continuous function that change according to time, and α is a constant value. The periodicity is hard if the $\Delta(t)$ is constant, while the periodicity is soft if the $\Delta(t)$ is not constant. Fig. 4.2 shows the soft periodicity of Eqs. 4.3.1 in which the phase of periodicity is shifted over time.

$$\begin{aligned} y(t) &= \sin(t+\Delta(t)) \\ \Delta(t) &= \alpha \sin(t) \end{aligned} \tag{4.3.1}$$

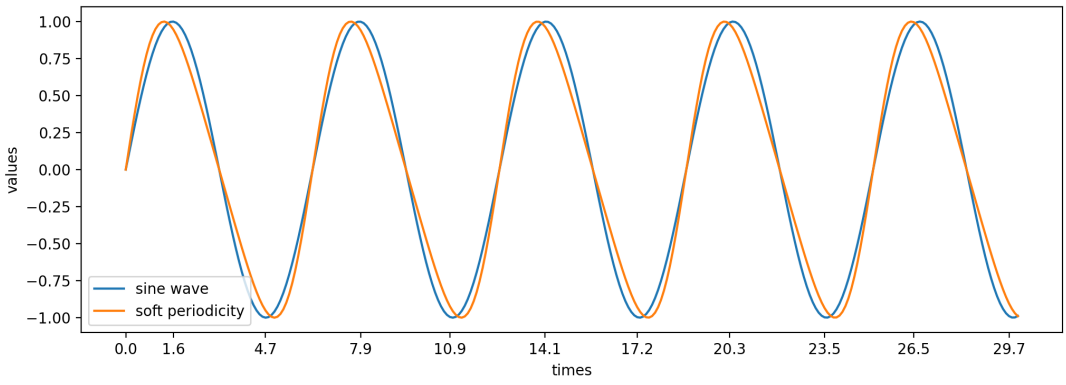


Figure 4.2: Soft periodicity

Furthermore, the autocorrelation function (ACF) is a visual way to show the serial correlation of the time series. ACF can be utilized to determine whether the climate data has a soft

periodicity or not. It can indicate the correlation over the seasonality lag by plotting the ACF plot. Figs. 4.3 and 4.4 show the ACF plot of geopotential height at 300 hPa (ZH300) and sea surface temperature (SST) which are the monthly data on 108 lags. The ZH300 and SST were calculated by averaging in spatial domain to transform the spatiotemporal data to time series data. The figures clearly show a high correlation of 12 lags (months) of yearly periodicity, which ZH300 shows stronger periodicity than SST. However, ZH300 peaks are not always every 12 lags; sometimes, the peak is at the 6^{th} lags of a cycle, which may indicate the soft periodicity.

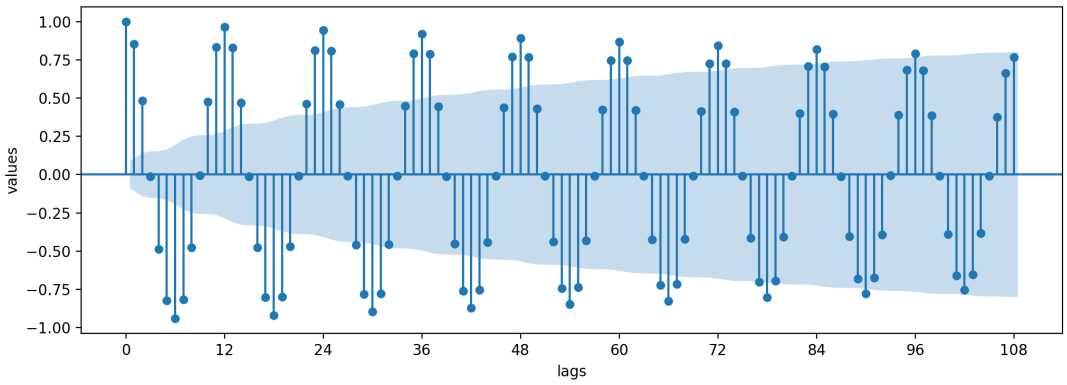


Figure 4.3: ACF plot of ZH300

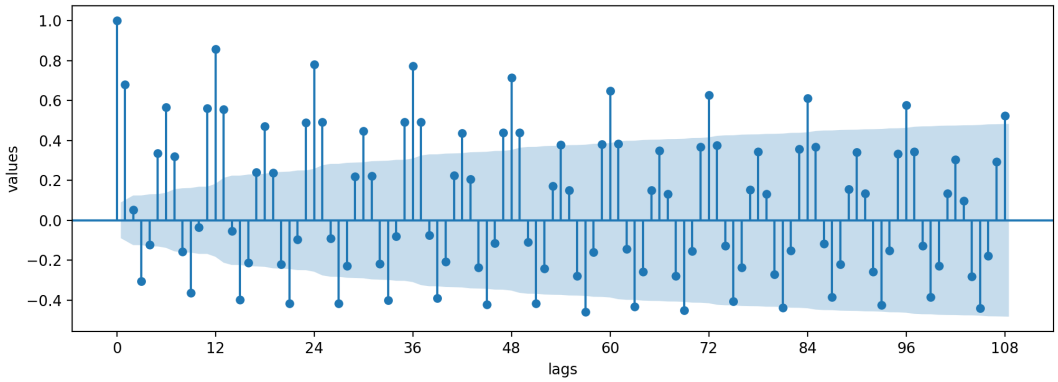


Figure 4.4: ACF plot of SST

4.4 Soft Periodic Recurrent Neural Network

4.4.1 CRN Component

To capture spatiotemporal patterns of an input sequence, spatial dependencies are needed to considered in the network. The convolution operator inside CRN is responsible for extracting spatial latent features from a low to high level. ConvLSTM is selected as a CRN component for this research, and named it SP-CRN (LSTM). The CRN component is shown in Fig. 4.1(a), and the detail of ConvLSTM is described in Section 2.3.1.

4.4.2 Periodic Representation

The purpose of this component is to consider a periodic pattern of the input sequence. A hidden state output from the CRN component ($h_{crn}^{(t+N)}$) is stored as a periodic representation (P), where *crn* refers to the CRN component. The periodic representation process of the load and update mechanisms is shown in Fig. 4.1(c) and (d).

P is an $m \times d$ matrix, where m is time index (month) of one periodic cycle; d is the depth of periodicity, and $p_{i,j}$ or $P[i][j]$ refers to a stored representation from the CRN component. The periodic representation is loaded and updated to P where the dictionary index (starting from 1) is matched with the indexing function, described in Sections 4.4.3 and 4.4.4. The expression $h_{P_j}^{(t+N)}$ represents loaded periodic representations from P, where j represents each of the loaded representation indices. I_P is a calculated time index of the cycle used in the load and update mechanism, where $I_P = t - (m \times \lfloor \frac{t}{m} \rfloor)$.

4.4.3 Update Mechanism

The output representation from the CRN component is stored as a periodic representation to the PRD and re-used in the next periodic cycle depending on the load mechanism explained in Section 4.4.4. In Periodic-CRN, the representation was saved to PRD as appending at the end of PRD. However, it is difficult for load mechanism of soft periodicity (Section 4.4.4) that is proposed in this study. Therefore, the new update mechanism is developed to update the representation to the PRD (2D matrix). The update mechanism is the depth-wise queue (First-In–First-Out), where it is repeated every time step (month by month). The process has two steps:

the first, shift representation, prepares the available space in P for the new representation, which is shifted from the bottom (the latest) to the top (the earliest) of P as shown in Eq. (4.4.1); the second, save new representation, is the new representation ($h_{crn}^{(t+N)}$) saved to the bottom of P as shown in Eq. (4.4.2); z refers to the depth of the matrix according to the indexing function.

$$P[I_P][z] \Rightarrow P[I_P][z-1] \quad (\text{for } z = 2, ..., d) \tag{4.4.1}$$

$$h_{crn}^{(t+N)} \Rightarrow P[I_P][d] \tag{4.4.2}$$

4.4.4 Load Mechanism

To use periodicity in the model, the representations from the PRD are loaded with three different proposals. Note that the load mechanism proceeds before the update mechanism. The representation at the bottom of P (where $z = d$) is the representation of the previous periodic cycle when the time index is I_P. Thus, $p_{i,z}$ (where $i < I_p$) are the representations of a current cycle. In contrast, $p_{i,z}$ (where $i \geq I_p$) are the representations of the previous cycle as shown in Fig. 4.5(a).

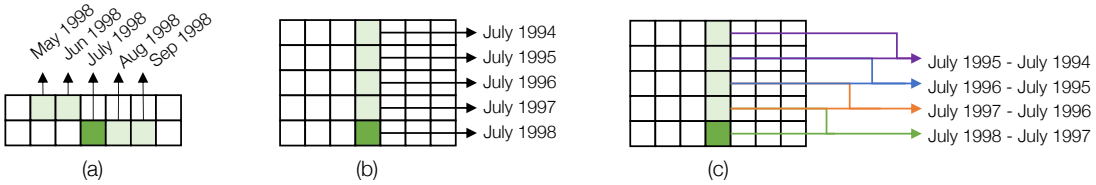


Figure 4.5: Examples of the three proposals for the load mechanism to predict the climate variable of July 1999 (index 1 starts from the top left of P). (**a**) Nearby-Time representation with the nearby representation of 2, (**b**) Periodic-Depth representation with a depth of 5, and (**c**) Periodic-Depth-Differencing representation with a depth of 5.

1. Nearby-Time Representation: In this work, to capture a phase shift that may change every year, the periodic attention mechanism component idea was extended from the original Periodic-CRN to include nearby (months) representations for the previous year as shown in Eq. (4.4.3).

In this proposal, d is 2, and $z \in \{1,2\}$ indicate which depth of P will be used for a load mechanism. Since the update mechanism is proceeded at every time step, representations in the same depth (z) may not be in the same periodic cycle. I_{Pn} is the nearby time index; n is

the incremental number indicating nearby time index representation; and n_{pre} is the pre-defined length. The example of the nearby-time representation proposal is shown in Fig. 4.5(a). To find a yearly periodic representation of July 1999, m was defined as 12 (yearly), t was 247 (when the time index started from January 1979); and $\lfloor \frac{t}{m} \rfloor$ was calculated as 20 (years); and I_P is 7 (July). If n_{pre} is defined as 2, then the representations of $p_{1,5}$, $p_{1,6}$, $p_{2,7}$, $p_{2,8}$, and $p_{2,9}$ which are May, June, July, August and September 1998 are loaded.

$$h_{P_n}^{(t+N)} \Leftarrow P[I_{Pn}][z] \quad (\text{for } n = -n_{pre}, ..., 0, ..., n_{pre}) \tag{4.4.3}$$

where

$$I_{Pn} = \begin{cases} (I_P + n) - m, & \text{if } I_P + n > m \\ m - (I_P + n), & \text{else if } I_P + n \leq 0 \\ I_P + n, & \text{otherwise} \end{cases}$$

$$z = \begin{cases} 2, & \text{if } n \geq 0 \\ 1, & \text{otherwise} \end{cases}$$

2. Periodic-Depth Representation: The periodic representation in the first proposal, the nearby time representation, considers only one year's prior representation. The representation was extended to be more than one period prior, and use only the same time index (phase) as shown in Eq. (4.4.4) where d is the depth of P representing the number of consecutive periodic cycles. The load mechanism loads all the representations that are in the same phase of the previous cycles.

An example of periodic-depth representation proposal is shown in Fig. 4.5(b). To find a periodic representation of July 1999, d is 5, t is 247, m is 12, $\lfloor \frac{t}{m} \rfloor$ is 20, then I_P is 7 (July). The representation of July 1998, 1997, 1996, 1995, and 1994 are loaded according to $d = 5$.

$$h_{P_z}^{(t+N)} \Leftarrow P[I_P][z] \quad (\text{for } z = 1, ..., d) \tag{4.4.4}$$

3. Periodic-Depth-Differencing Representation: Similar to the second proposal, the change over cycles were considered by taking the difference between the month of two adjacent years as shown in Eq. (4.4.5), in which a differencing of two years was used. The load mechanism

loaded the differencing of representation $p_{i,j}$.

The example of periodic-depth-differencing representation proposal is shown in Fig. 4.5(c). According to the parameters set in the second proposal, to find the periodic representation of July 1999, the representation of July 1998 minus 1997, 1997 minus 1996, 1996 minus 1995, and 1995 minus 1994 were loaded.

$$h_{P_z}^{(t+N)} \Leftarrow P[I_P][z] - P[I_P][z-1] \quad (\text{for } z = 2, ..., d) \tag{4.4.5}$$

4.4.5 Attention Module

In this method as shown in Fig. 4.1(e), the periodic representations from load mechanism (Section 4.4.4) are weighted with the attention mechanism as described in Eqs. (2.3.5–2.3.7) and Section 2.3.3. In this study, the different from Periodic-CRN is that the SP-CRN consider the single resolution of periodicity.

4.4.6 Fusion Module

In this method as shown in Fig. 4.1(f), the output from attention module and CRN component are combined as described in Eq. 2.3.8 and Section 2.3.3.

In Periodic-CRN, the spatial weight on the attention module was the only single (static) spatial weight (W_P in Eq. 2.3.8). In this study, we propose the dynamic spatial weight (DSW) by replacing SSW in Eq. 2.3.8. In Eq. 4.4.6, the DSW is the W_{P_k} with the size of dynamic weight is the number of time index in a periodicity cycle where k indicates the month of the year of the prediction.

$$h_f^{(t+1)} = relu(W_c \circ h_c^{(t+1)} + W_{P_k} \circ h_{all}^{(t+1)}) \tag{4.4.6}$$

After the fusion module, the convolution layer was used to transform the output $\hat{X}^{(t+N)}$ to the same size as an input. In Eq. (4.4.7), the size of W_f and $\hat{X}^{(t+N)}$ are the same size as the input, and the prediction of the model is $\hat{X}^{(t+N)}$. This study can obtain the prediction output without the metadata component (Section 4.4.7) to make a model focused on the periodicity component. Also, the ReLU function is unnecessary for this study because the prediction output can be

minus.

$$\hat{X}^{(t+N)} = h_f^{(t+N)} * W_f \tag{4.4.7}$$

4.4.7 Metadata Component

The idea of a metadata component is to indicate the month of the year (single resolution of periodicity) of the prediction output with one-hot encoding support. The metadata component was added at the end of the fusion module in Fig. 4.1(f), and Section 4.4.6. The metadata component is the same as Periodic-CRN, which is described in Eq. 2.3.11 by removing the ReLu function.

4.5 Experiment Settings

4.5.1 ERA-Interim Dataset

The description of ERA-Interim is described in Section 3.2.1. This study validates the proposed method performance by choosing two climate variables from ERA-Interim: geopotential at 300 hPa (Z300) and sea surface temperature (SST). In addition, the two datasets used the same periods of training, validation, and test set (1979–2015, 2016, and 2017–2018, respectively). Fig. 4.6 shows the sample of datasets.

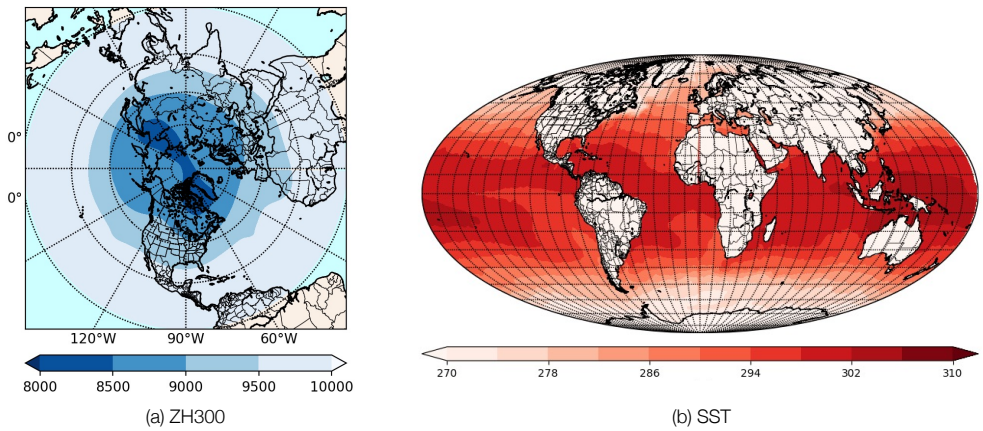


Figure 4.6: A sample of the monthly spatial plot of January 2018. (**a**) The Northern Hemisphere of ZH300 data, where deep blue represents the lower value of ZH300. (**b**) The entire globe in the spatial grid of SST data, where deep red represents high temperature, and the terrain area (white) is ignored. Note that these projections are just for visualization.

Z300 is at the same height as the jet stream, a strong upper tropospheric wind axis 9–12 km above sea level. The jet stream wind separates the high and low pressure systems representing hot and cold weather, respectively. Atmospheric variation in the upper troposphere is used for weather prediction at the surface because the air circulation can be a precursor to weather prediction [40]. Moreover, an upper troposphere change can cause extreme winter cold [38] over northern Europe and cold air outflows in the Northern Hemisphere [22]. In this dissertation, Z300 is converted to a geopotential height at pressure level 300 hPa (ZH300) divided by gravitational acceleration (9.8 m/s^2). The data was extracted to a 90 (latitude) $\times$ 360 (longitude) grid size that covered the Northern Hemisphere.

Sea surface temperature (SST) plays an important role in the interaction between the earth's surface and atmosphere. It is an important parameter in energy balancing at the earth's surface and is also a critical indicator of the heat of sea water. A global SST was chose with a 1 degree grid resolution, which is the spatial size of 180 (latitude) $\times$ 360 (longitude).

4.5.2 Evaluation Metric

As the area around the North and South Pole is less than the equator area, a latitude-weighted root mean square error (RMSE) was employed as an evaluation metric. A latitude-weighted RMSE was also the evaluation metric in WeatherBench [33], the first benchmark for data-driven, medium-range climate prediction.

$$RMSE = \frac{1}{N_{\text{forecasts}}} \sum_{i}^{N_{\text{forecasts}}} \sqrt{\frac{1}{N_{\text{lat}} N_{\text{lon}}} \sum_{j}^{N_{\text{lat}}} \sum_{k}^{N_{\text{lon}}} L(j)(f_{i,j,k} - t_{i,j,k})^2}, \tag{4.5.1}$$

where f is the model forecast, and t is the ground truth of the data. $N_{\text{forecasts}}$, N_{lat}, and N_{lon} are the numbers for forecasting, latitude, and longitude, respectively. $L(j)$ is the latitude weighting factor by the latitude (degree) at the j^{th} latitude index $\text{lat}(j)$:

$$L(j) = \frac{\cos(\text{lat}(j))}{\frac{1}{N_{\text{lat}}} \sum_{j}^{N_{\text{lat}}} \cos(\text{lat}(j))} \tag{4.5.2}$$

4.5.3 Comparison Methods

Convolutional neural networks (CNN) [24]: A CNN is a standard deep-learning method for next video frame prediction. It is also widely used and applied as a model for climate prediction. In this experiment used five convolutional layers, the last being the output layer with an output channel of one.

Convolution gated recurrent unit (ConvGRU) [5]: A ConvGRU was applied as a baseline method by replacing the first two layers with ConvGRU from the CNN baseline model; that is, two layers of ConvGRU and three layers of CNN. A prior consecutive spatial input was applied to the model.

ConvLSTM: To determine the benefit of using an LSTM structure that captures change over time, a ConvLSTM was adopted as a baseline method. The first two layers of the CNN were replaced by ConvLSTM layers, the same way as the ConvGRU comparison method.

SP-CRN (GRU) and SP-CRN (LSTM): As described in Chapter 4, the ConvGRU and the ConvLSTM were applied as the CRN component for the SP-CRN model. The SP-CRN (LSTM) will undergo further experimentation for the periodicity proposals .

4.5.4 Settings

The experiments were conducted to test the hyperparameters on CNN and ConvLSTM to find suitable ones for the experiments. The hyperparameters tested compared the kernel size of 3×3, 5×5, and 7×7, and the number of filters of 16, 32, and 64. Though the RMSEs on validation data were slightly different, the 3×3 kernel size and 16 filters were selected for the time of the computational resources. Then the CNN layer was replaced with a ConvLSTM layer. The same hyperparameters were used for all the methods because the baseline algorithms and the proposed method used the similar CNN base network. A dropout rate of 0.5, and ELU activation functions were applied to every layer except the output layer, which used the identity function.

The experiments performed all the models using an Adam optimizer with a learning rate of 10^{-4}, using mean square error (MSE) as a loss function. Shuffled training data was used to train CNN, ConvLSTM, and ConvGRU to be more generalized; on the other hand, SP-CRN did not use shuffled training data because the PRD component considers the order of training

data. The models adopted early stopping with 100 epoch patience on validation loss. The the models were trained and evaluated each model five times and computed the average RMSE from each model prediction result. The dataset was scaled using standardized (Z-score normalization) by calculating the mean and standard deviation (SD) from the whole of space and time in the training dataset. The Z-score normalization makes the learning a stable convergence and leads a better result.

All the experiments of SST were the same as for ZH300. The only difference was the size of the spatial data: 90×360 (ZH300) and 180×360 (SST). The spatial data of SST contained masking, which ignored all the terrain area (the white space in Fig. 4.6(b)). The terrain area was replaced with 0 and ignored it when using the standardize method in the training dataset, and MSE in the loss function. The same hyperparameters as ZH300 were used: a 3×3 kernal and 16 filters. The same set of comparison methods were used in SST experiments.

4.6 Results on ZH300

4.6.1 Input Sequence Length

The input sequence length affects the model's performance, and each climate variable may need a different length of the input. This experiment indicates the proper input sequence for each model to perform one step ahead prediction on ZH300. Fig. 4.7 shows the average RMSE prediction results comparing the models with the input sequence length of one to six months. The SP-CRN in this experiment is utilizing the PRD proposal of nearby-time representation, where $n_{pre} = 0$ (zero nearby-time representation). The results show that when increasing the length of sequential input, the RMSE of all models is significantly decreasing. Furthermore, the RMSE results tend to be converged after the sequence length of five or six months, indicating that the five to six months prior knowledge of ZH300 is essential for one step ahead of ZH300 prediction, which is consistent with the SST memory is about six months [26]. The SST is used to indicate ENSO, which impacts the prediction of weather and climate. The results also present superiority of the models consider periodicity component, which has the lower RMSE from their base model. The SP-CRN (GRU) and the SP-CRN (LSTM) have the competitive RMSE in every length of sequential input.

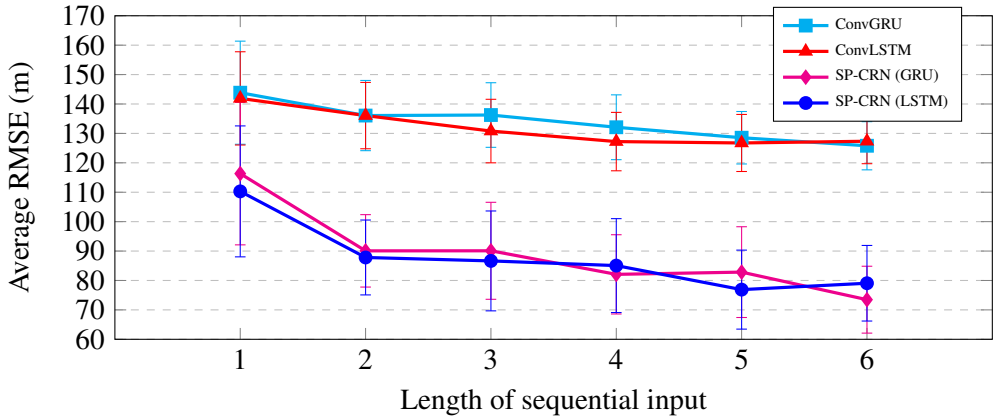


Figure 4.7: The average RMSE of the comparison methods in different length of sequential input for ZH300 prediction. The error bar indicates the SD of the RMSE.

4.6.2 PRD Types

The SP-CRN (LSTM) was used as the model with periodicity component to perform the PRD proposals. Note that SP-CRN refers to SP-CRN (LSTM) from now on. An input sequence length of five months was selected as the baseline setting because it was the best for the SP-CRN model in the previous experiment (Fig. 4.7). Table 4.1 shows the comparison of SP-CRN for the three proposals in different settings. The SD shows the dispersion to the average RMSE value. The SP-CRN-1 (1,1,0) with an RMSE of 74.22 m is slightly better than the SP-CRN-1 (0,1,0) baseline result (RMSE of 76.89 m). The result of SP-CRN-1 (1,1,0) indicated that only the nearby-time representation of plus and minus one month in one prior periodic cycle was more important than the more than two months ($n_{pre} > 1$) for ZH300 prediction. All settings for the SP-CRN-2 proposal results showed a worse prediction accuracy compared to the baseline except SP-CRN-2 (0,4,0) with an RMSE of 70.58 m, indicating the importance of the information of the previous four periodic cycles for ZH300 prediction and possibly for the four-year oscillation of ENSO [37]. All settings of SP-CRN-3 outperfored the baseline result, of which SP-CRN-3 (0,0,2) had the best with an RMSE of 67.97 m. The results clearly showed that the change over cycles significantly improved ZH300 prediction accuracy. In summary, using SP-CRN for periodicity in ZH300 prediction was most effective for periodic-depth differencing representation (PRD-3) followed by periodic-depth representation (PRD-2) and nearby-time representation (PRD-1).

Table 4.1: Evaluation of different settings on the ZH300 prediction for SP-CRN with three proposals. The settings ((x, y, z) are the numbers of the nearby-time of PRD-1, periodic-depth of PRD-2, and periodic-depth of PRD-3, respectively. The SP-CRN-1, SP-CRN-2, and SP-CRN-3 indicate each proposal in SP-CRN which was a nearby-time representation, periodic-depth representation, and periodic-depth differencing representation, respectively. The highlighted result in blue is the best prediction accuracy for each proposal, while the gray highlight in the SP-CRN baseline result.

Method	**Avg. RMSE (m)**	**SD of RMSE (m)**
SP-CRN-1 (0,1,0)	76.89	13.43
SP-CRN-1 (1,1,0)	74.22	15.44
SP-CRN-1 (2,1,0)	81.64	13.89
SP-CRN-1 (3,1,0)	84.42	17.67
SP-CRN-2 (0,2,0)	79.69	13.96
SP-CRN-2 (0,3,0)	94.85	14.05
SP-CRN-2 (0,4,0)	70.58	12.43
SP-CRN-2 (0,5,0)	103.93	12.62
SP-CRN-2 (0,6,0)	100.76	16.23
SP-CRN-3 (0,0,2)	67.97	11.95
SP-CRN-3 (0,0,3)	70.05	12.23
SP-CRN-3 (0,0,4)	69.53	12.16
SP-CRN-3 (0,0,5)	68.12	12.13
SP-CRN-3 (0,0,6)	68.26	12.47

4.6.3 Combination of PRD Types and Metadata

To indicate the performance of the periodicity component on both nearby-time and periodic-depth representation, the proposals of SP-CRN-1 (1,1,0), SP-CRN-2 (0,4,0), and SP-CRN-3 (0,0,2) because they performed the best in their settings from the previous experiment (Table 4.1). The mixed proposals were combined with all the combinations that were SP-CRN-1+2 (1,4,0), SP-CRN-1+3 (1,0,2), SP-CRN-2+3 (0,4,2), and SP-CRN-1+2+3 (1,4,2). The experiment also tested the idea of a metadata component by selecting SP-CRN-1+3 (1,0,2) as the mixed proposal to be added to the metadata component (SP-CRN-1+3-Meta (1,0,2)). Note that the results explained in this section were based on SSW, and did not include the metadata. The DSW explanation was explained in Section 4.6.5.

Table 4.2 shows the experimental results of each model setting that had the best prediction accuracy. All the proposals on SP-CRN had better prediction accuracy than SP-CRN (LSTM)

baseline, indicating that the periodicity utilization proposals helped improve prediction accuracy. The mixed proposal of SP-CRN-1+3 (1,0,2) was the best compared with the other mixed proposals, which improved SP-CRN-1 (1,1,0) and SP-CRN-3 (0,0,2) accuracy with an RMSE of 66.39 by considering the periodicity in both directions (nearby-time and periodic-depth). In contrast, the other mixed proposals with PRD-2 were worse than their baseline models. The metadata component in SP-CRN-1+3-Meta (1,0,2) was superior at indicating the phase of periodicity and had the lowest RMSE comparing to SP-CRN-1+3 (1,0,2) (base model).

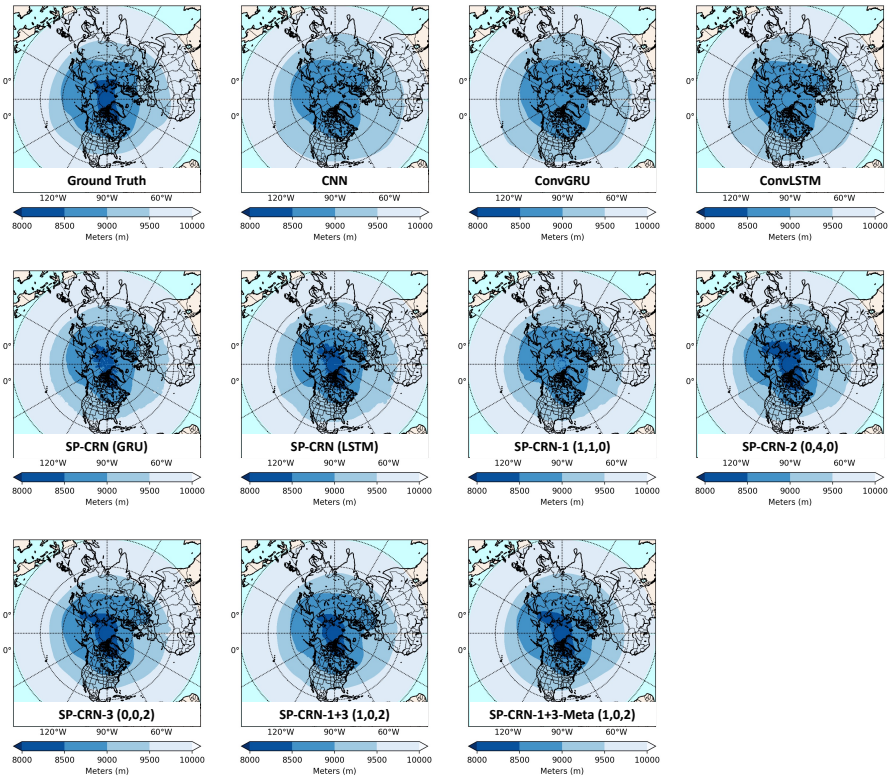
Table 4.2: Evaluation of different settings on the ZH300 prediction for SP-CRN and comparison methods.

Method	**RMSE**	**SD**
	SSW | DSW	**SSW | DSW**
CNN	141.79	16.54
ConvGRU	125.80	8.19
ConvLSTM	126.75	9.70
SP-CRN (GRU)	73.47	11.37
SP-CRN (LSTM)	76.89	13.43
SP-CRN-1 (1,1,0)	74.22 | 68.26	15.44 | 11.73
SP-CRN-2 (0,4,0)	70.58 | 59.91	12.43 | 10.82
SP-CRN-3 (0,0,2)	67.97 | 66.60	11.95 | 12.52
SP-CRN-1+2 (1,4,0)	82.89 | 56.91	13.85 | 11.25
SP-CRN-1+3 (1,0,2)	66.39 | 63.07	12.21 | 9.81
SP-CRN-2+3 (0,4,2)	88.44 | 57.47	10.56 | 11.13
SP-CRN-1+2+3 (1,4,2)	80.59 | 54.72	11.65 | 10.84
SP-CRN-1+3-Meta (1,0,2)	50.27 | 64.02	12.77 | 10.13

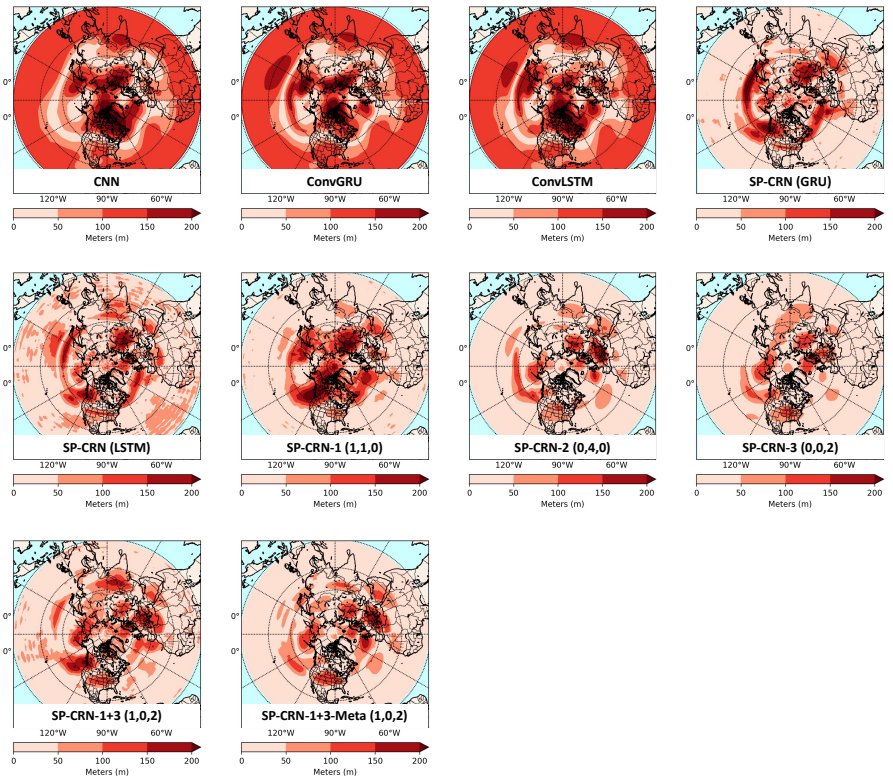
4.6.4 Spatial Distribution Analysis

Figure 4.8 shows the output result and error distribution of February 2017, in which the average RMSE of the month reflected the results of the PRD proposals in Table 4.2. Figure 4.8a shows the prediction output, in which all of the methods captured the global pattern compared to the ground truth. However, the low ZH300 value around the North Pole was only captured by the SP-CRN (LSTM) methods that considered the periodic-depth representation. The prediction results indicated that the periodicity component helped the models capture a wider range of ZH300 values. Moreover, the periodic-depth representation was more important than the

nearby-time representation for capturing the lowest peak value.



(a) Ground truth and prediction results.



(b) Error distribution (absolute error).

Figure 4.8: The prediction and error distribution of ZH300 on February 2017.

Figure 4.8b shows the spatial error distribution using the absolute error between ground truth and prediction results. The CNN, ConvGRU, and ConvLSTM results showed a large area of high error (100–150 m) around the tropical area. However, there was only a small error on the subtropical area of the CNN, ConvGRU, and ConvLSTM. Thus, the yearly change of this area was not much, which did not show a huge advantage for the model considering the periodicity components. The models that considered periodicity components, on the other hand, showed a much smaller error in the tropical area. The SP-CRN (GRU), SP-CRN (LSTM), and SP-CRN-1 could not capture the peak of low values in Figure 4.8a, in which they represented the high error distribution around the North Pole in Figure 4.8b. The SP-CRN-2 and SP-CRN-3, which considered only periodic-depth representation, showed some small area of high error distribution even when comparing SP-CRN-1+3 and SP-CRN-1+3-Meta with lower RMSE.

4.6.5 Static Spatial Weights vs Dynamic Spatial Weights

The experiments tested on all SP-CRN settings with DSW to check the availability of dynamic spatial weights for attention module output. The results in Table 4.2 clearly shows that DSW drastically improved for the SP-CRN results except for SP-CRN-1+3-M. SP-CRN-3 and SP-CRN-1+3 results of DSW were slightly better than the SSW results. The PRD-3 of DSW results may indicate that the change over cycle in each month was not much different, so the DSW did not affect the prediction accuracy. The DSW behavior was metadata-like which may confuse the network because the DSW of SP-CRN-1+3-M already included the metadata component.

The error distribution (Fig. 4.9) was plotted comparing the SSW and DSW for SP-CRN-1 and SP-CRN-2, and selected a month that the prediction accuracy was similar to the Table 4.2. The figures show that the high error areas of DSW were removed from the prediction results, especially the SP-CRN-2 that the lower errors were also removed in low latitude areas.

4.6.6 Periodicity Analysis

The weights in the attention module (Section 4.4.5) were analyzed to indicate the importance of the periodic representation collected in the PRD. The attention weights were from the trained model with 37 years (1979–2015) of the training set, then averaged each of the same months.

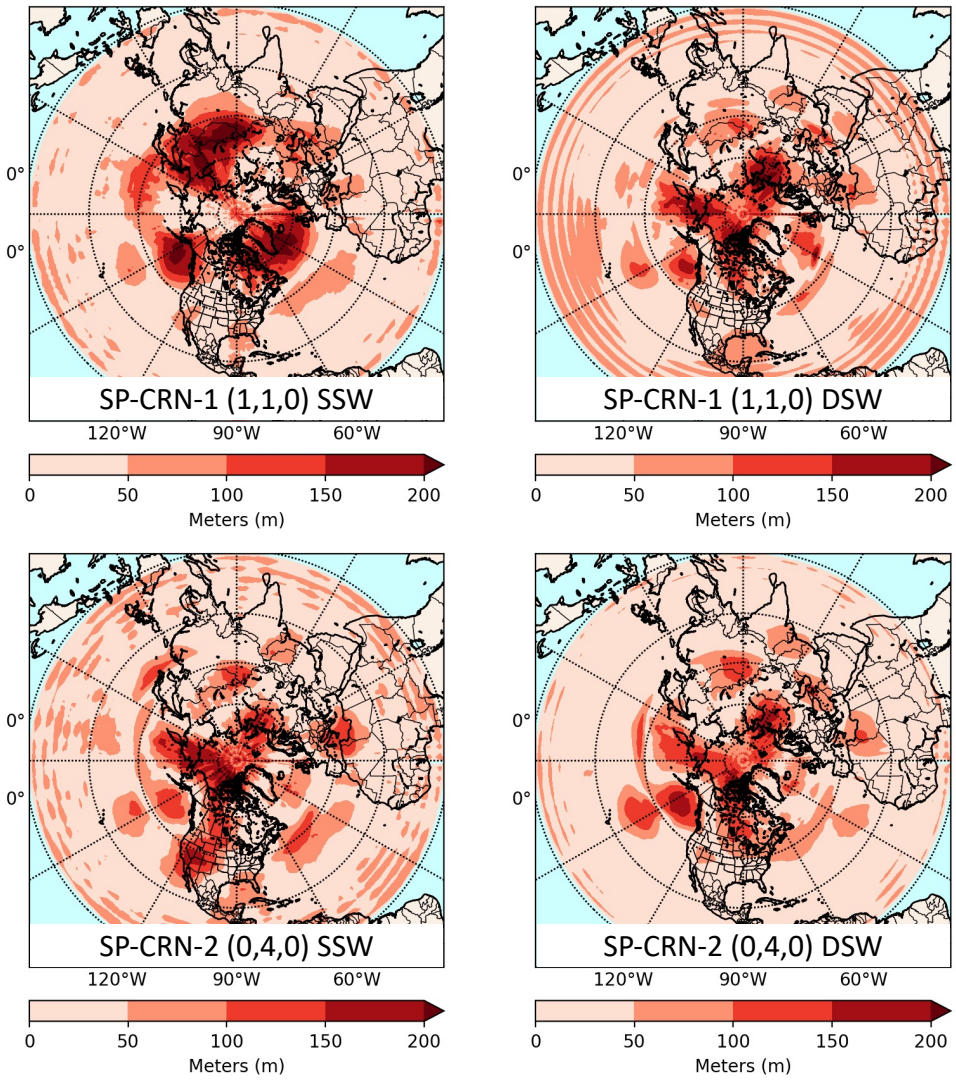


Figure 4.9: The ZH300 error distribution of SP-CRN-1 (1,1,0) SSW, SP-CRN-1 (1,1,0) DSW, SP-CRN-2 (0,4,0) SSW, and SP-CRN-2 (0,4,0) DSW in January 2018, with RMSE of 72.51 m, 63.16 m, 63.03 m, and 55.94, respectively.

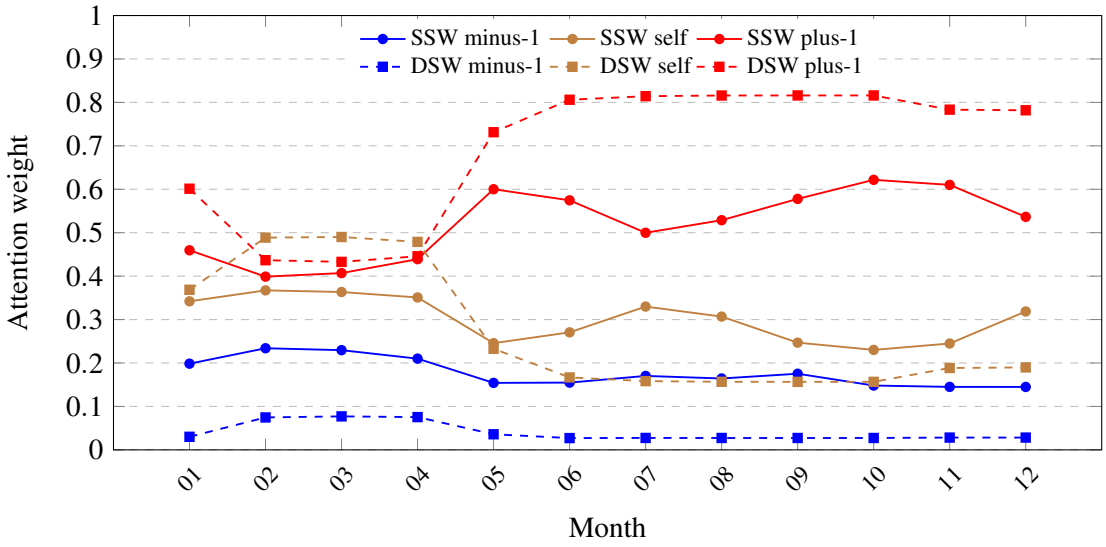


Figure 4.10: The average attention weight of training data on SP-CRN-1 (1,1,0) for ZH300 prediction from the average of five runs.

Fig. 4.10 shows the attention weight of the SP-CRN-1 (1,1,0) model, which was the best setting in SP-CRN-1 SSW of the ZH300 prediction. The phase of periodicity for one prior periodic cycle for one month ahead (plus-1) was more important than the same (self) or previous month (minus-1) of SSW and DSW, respectively. The attention weight of SSW plus-1 was high in May and October but low in February, while the attention weight for self was the opposite. Meanwhile, the DSW plus-1 was similar, but the height from May to December differed from SSW plus-1 from June to September. The minus-1 was smooth for all phases, indicating the same importance for every month of the ZH300 prediction.

Fig. 4.11 shows the attention weight for the ZH300 prediction with the SP-CRN-2 model. The periodic cycle of prior 1–4 years were considered together with the attention module. The overall weights were almost equal among the different depths between the weight of 0.1 to 0.4 excepts SSW depth-4, which indicated the importance of the fourth prior year over the other years in SSW. The single spatial weight of SSW, the depth-4 was significant. While the dynamic spatial weights of DSW, all the depths were not significantly different.

4.7 Results on SST

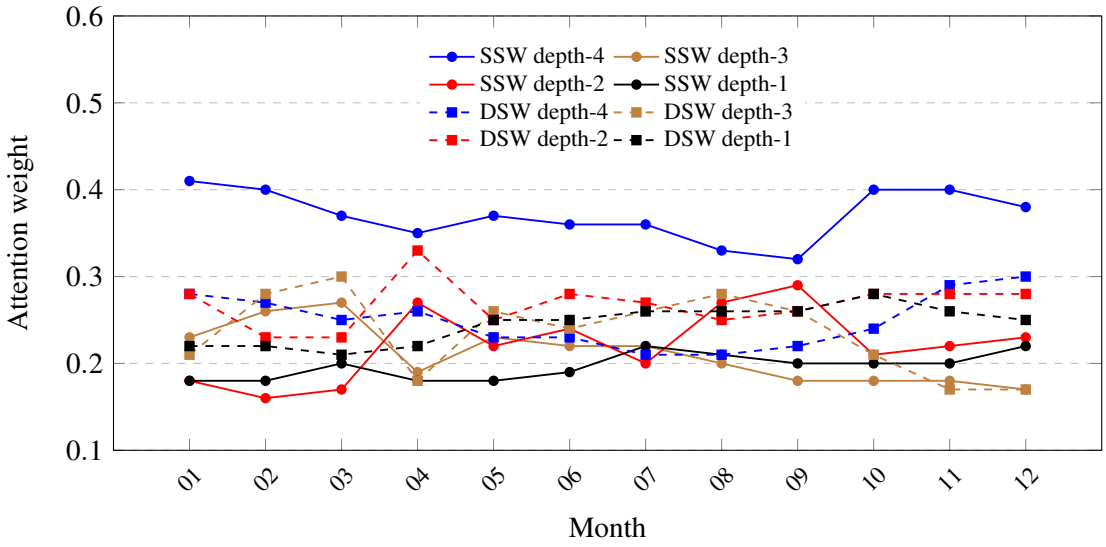


Figure 4.11: The average attention weight of training data on SP-CRN-2 (0,4,0) for ZH300 prediction from the average of five runs. The attention weight of depth-1 to depth-4 was the value of the prediction month for one to four prior periodic cycles.

4.7.1 Input Sequence Length

This experiment indicated the performance of comparison methods by considering the length of the sequential input. Fig. 4.12 shows the average RMSE from a sequence length of one to six. The ConvGRU and the ConvLSTM showed the competitive results as did the SP-CRN (GRU) and SP-CRN (LSTM), which used the zero nearby-time representation as the periodicity component, which can improve prediction accuracy of base models by the SST value of around two K. The increasing input sequence length slightly decreased the RMSE of the SP-CRN (GRU) and the SP-CRN (LSTM) for SST prediction. Except for the length of one compared to two, the RMSE dropped significantly. The results demonstrated the superiority of the periodicity component over the base models. However, the length of the sequential input improved the prediction accuracy slightly.

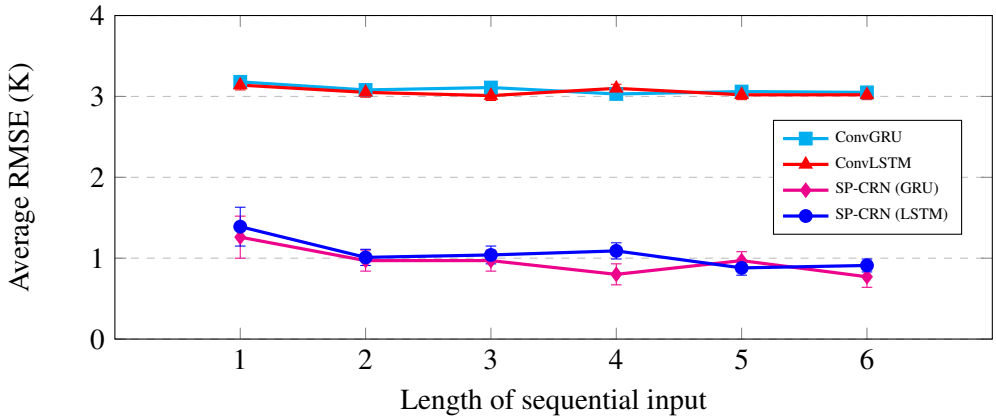


Figure 4.12: The average RMSE of comparison methods of different lengths of sequential input for SST prediction. The error bar indicates the SD of the SST.

4.7.2 PRD Types

The input sequence length of five was selected for this experiment as the SP-CRN had the lowest RMSE of 0.86 K in the previous experiment (Fig. 4.12). Table 4.3 shows the results of the three proposals compared with different settings for SP-CRN. The results for SP-CRN-1 were slightly better than for SP-CRN-1 (0,1,0) (baseline). The best result was for SP-CRN-1 (1,1,0) with an RMSE of 0.79 K. The SP-CRN-2 results indicated that the importance of a prior periodic cycle of more than four years was almost identical for an SST prediction. All the settings of SP-CRN-3 outperformed the baseline result, for which SP-CRN-3 (0,0,2) had the best RMSE of 0.55 K, and the change over cycles played an important role. The results showed that SP-CRN-2 with more than four years and SP-CRN-3 with the depth of two drastically improved prediction accuracy. In summary, the order of the effectiveness of SP-CRN in periodicity utilization for an SST prediction was PRD-3 over PRD-2 and PRD-1.

Table 4.3: Evaluation of different settings on SST predictions for SP-CRN with three proposals. The highlighted result in blue shows the best prediction accuracy for each proposal, while the highlight in gray is the SP-CRN baseline result.

Method	Avg. RMSE (K)	SD of RMSE (K)
SP-CRN-1 (0,1,0)	0.88	0.09
SP-CRN-1 (1,1,0)	0.79	0.10
SP-CRN-1 (2,1,0)	0.82	0.08
SP-CRN-1 (3,1,0)	0.84	0.12
SP-CRN-2 (0,2,0)	1.29	0.11
SP-CRN-2 (0,3,0)	0.96	0.05
SP-CRN-2 (0,4,0)	0.63	0.06
SP-CRN-2 (0,5,0)	0.63	0.05
SP-CRN-2 (0,6,0)	0.61	0.05
SP-CRN-3 (0,0,2)	0.55	0.04
SP-CRN-3 (0,0,3)	0.56	0.06
SP-CRN-3 (0,0,4)	0.57	0.05
SP-CRN-3 (0,0,5)	0.57	0.05
SP-CRN-3 (0,0,6)	0.56	0.05

4.7.3 Combination of PRD Types and Metadata

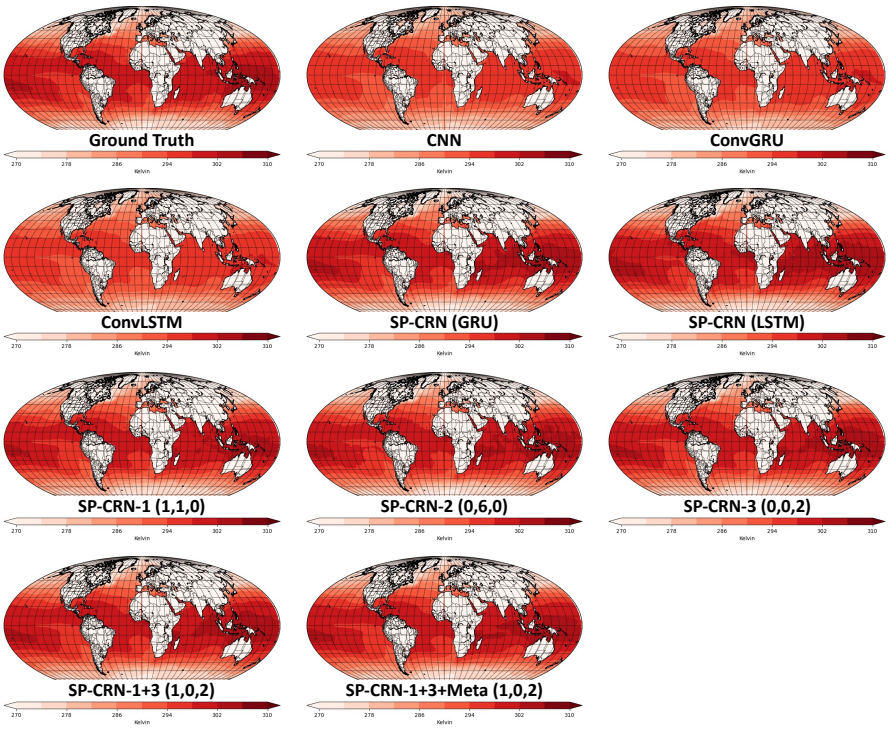
This section explained the results based on SSW. The DSW based results were explained in Section 4.7.5. The best setting of each proposal were mixed with a previous experiment (Table 4.3). These are SP-CRN-1 (1,1,0), SP-CRN-2 (0,6,0), and SP-CRN-3 (0,0,2). All the combinations of mixed proposal are SP-CRN-1+2 (1,6,0), SP-CRN-1+3 (1,0,2), SP-CRN-2+3 (0,6,2), and SP-CRN-1+2+3 (1,6,2). Table 4.4 shows the results of all the methods along with their best settings. SP-CRN-1+3 (1,0,2) was the best of all mixed proposals with an RMSE of 0.56 K, which was no statistical difference to SP-CRN-3 (0,0,2) (baseline). Mixing the proposals produced no accuracy improvement from baseline. Meanwhile, the metadata component was added to SP-CRN-1+3 (1,0,2) to indicate its performance for an SST prediction of the best mixed proposal. The metadata component of SP-CRN-1+3-Meta (1,0,2), on the other hand, demonstrated its superiority by improving prediction accuracy with an RMSE of 0.43 K and a small SD compared to SP-CRN-1+3 (1,0,2) (baseline).

Table 4.4: Evaluation of different settings on the SST prediction for SP-CRN and comparison methods.

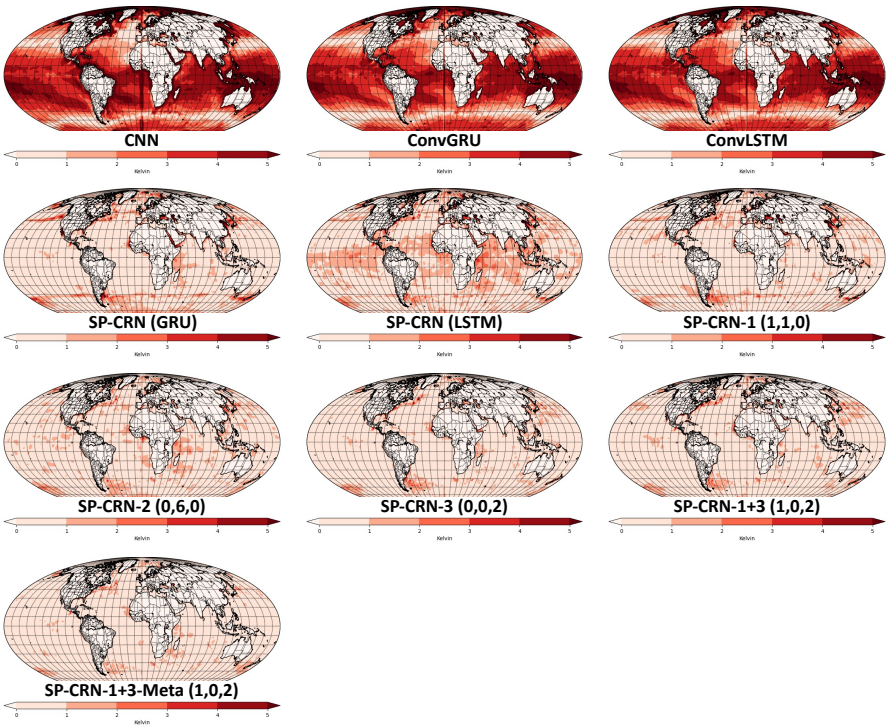
Method	**RMSE**	**SD**
	SSW | DSW	**SSW | DSW**
CNN	3.32	0.05
ConvGRU	3.03	0.06
ConvLSTM	3.02	0.06
SP-CRN (GRU)	0.77	0.13
SP-CRN (LSTM)	0.88	0.09
SP-CRN-1 (1,1,0)	0.79 | 0.56	0.10 | 0.04
SP-CRN-2 (0,6,0)	0.61 | 0.49	0.05 | 0.04
SP-CRN-3 (0,0,2)	0.55 | 0.56	0.04 | 0.05
SP-CRN-1+2 (1,6,0)	0.74 | 0.47	0.08 | 0.03
SP-CRN-1+3 (1,0,2)	0.56 | 0.59	0.05 | 0.04
SP-CRN-2+3 (0,6,2)	0.65 | 0.47	0.05 | 0.03
SP-CRN-1+2+3 (1,6,2)	0.61 | 0.48	0.05 | 0.04
SP-CRN-1+3-Meta (1,0,2)	0.43 | 0.46	0.03 | 0.03

4.7.4 Spatial Distribution Analysis

Figure 4.13 shows the output result and error distribution of January 2018 that reflects the order of the SP-CRN results in Table 4.4. Figure 4.13a shows the prediction output in which the methods that use periodicity components can capture changes in high and low temperature. Compared to the CNN, the ConvGRU, and ConvLSTM raely captured the Pacific Ocean's high temperature. Only the smooth change pattern was presented in the prediction output.



(a) Ground truth and prediction results.



(b) Error distribution (absolute error).

Figure 4.13: The prediction and error distribution of SST on January 2018.

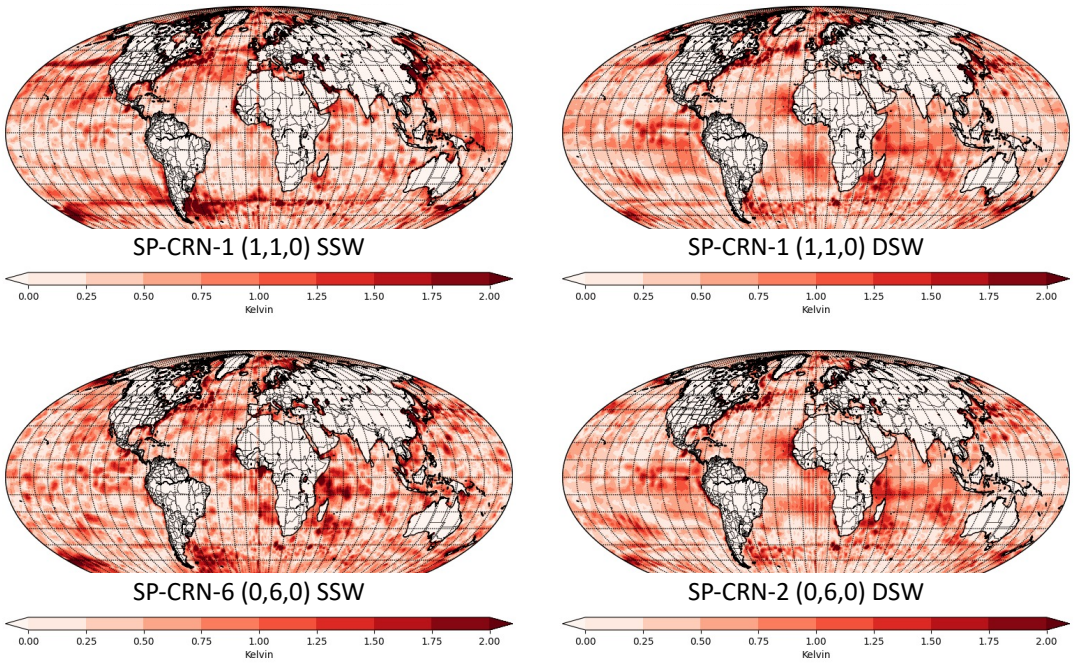


Figure 4.14: The SST error distribution of SP-CRN-1 (1,1,0) SSW, SP-CRN-1 (1,1,0) DSW, SP-CRN-2 (0,6,0) SSW, and SP-CRN-2 (0,6,0) DSW (RMSE of 0.60 K) in January 2018 with RMSE of 0.75 K, 0.60 K, 0.65 K, and 0.54 K, respectively.

Figure 4.13b shows the error distribution of comparison methods. The CNN, ConvGRU, and ConvLSTM showed that the extremely high temperature (Pacific Ocean) and low temperature (North and South Pole) have high error distribution, but for the model, considering the periodicity, it was not much different. SP-CRN (LSTM) showed a higher error distribution in more areas than the SP-CRN (GRU) and SP-CRN variant methods.

4.7.5 Static Spatial Weights vs Dynamic Spatial Weights

The results of DSW (Table 4.4) were similar to the ZH300 results which there was an accuracy improvement for almost the SP-CRN settings. Also, the PRD-3 of DSW in SST did not improve the accuracy in SP-CRN-3 and SP-CRN-1+3. However, including PRD-2 along with PRD-3 of DSW in SP-CRN-1+2+3 helped improve prediction accuracy. The DSW of SP-CRN-1+3-M showed worse prediction accuracy compared to SSW, which is the same trend as ZH300 result.

Fig. 4.14 show the error distribution of SP-CRN-1 and SP-CRN-2. The error distribution of DSW obviously showed lower error especially in the subtropical area.

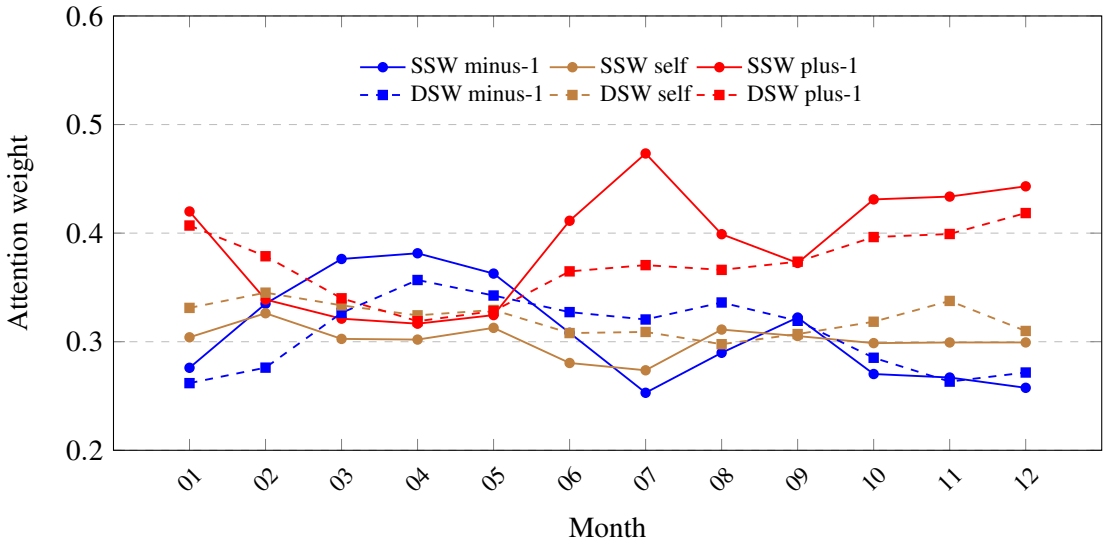


Figure 4.15: The average attention weight of training data on SP-CRN-1 (1,1,0) for SST prediction from the average of five runs.

4.7.6 Periodicity Analysis

Fig. 4.15 shows the attention weight of SP-CRN-1 (1,1,0) which was the best result in the SP-CRN-1 settings. The periodic pattern of SSW plus-1 was similar to DSW plus-1 (Fig. 4.10) in terms of one decrease in the early year and then two peaks. The plus-1 and minus-1 had occasional importance according to some months, while the attention weight of self was smooth every month. The difference in the attention weight pattern of plus-1 and drastically different RMSE between SSW and DSW, especially in June to August, indicated that each periodicity phase needs a specific periodicity pattern.

Fig. 4.16 shows the attention weight of SP-CRN-2 which showed the importance of periodic-depth from one to six years of a periodic cycle. The overall weights were about 0.1–0.2 for all months, which was about the average of six weights—0.167, except the late year of SSW depth-5 and DSW depth-1, which seems to have a periodic cycle.

4.8 Discussion and Limitation of the Proposed Method

First, a clear outcome of this study is that SP-CRN-1 (1,1,0) prediction results of ZH300 and SST are the best among the PRD-1. Considering only nearby one month of PRD-1 can indicate that the phase shifts of periodicity have occurred in the range of one month. During a season

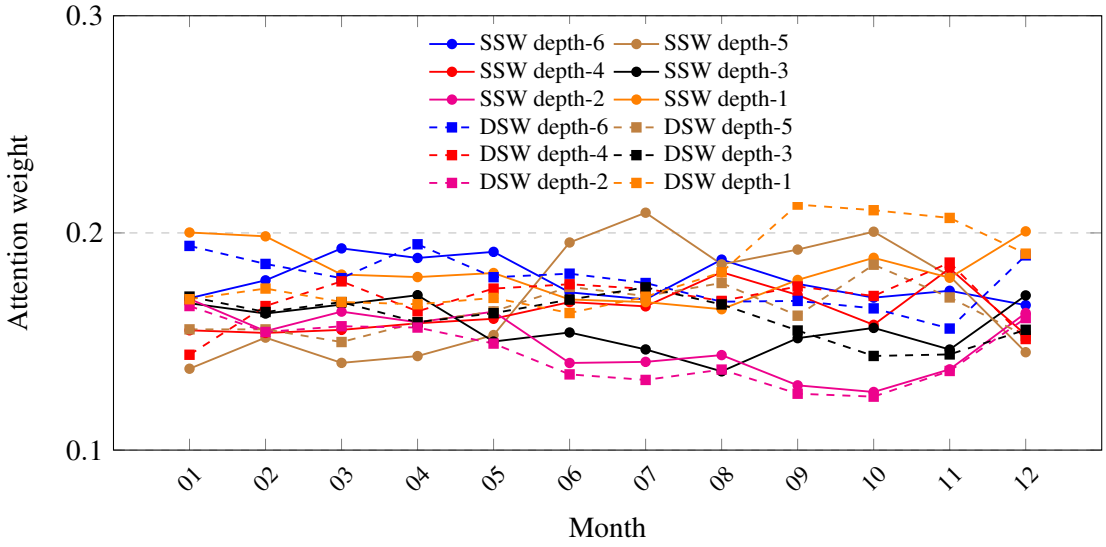


Figure 4.16: The average attention weight of training data on SP-CRN-2 (0,6,0) for SST prediction from the average of five runs.

(a term of three months), climate variables may have a similar changing tendency. In addition, for ZH300 (Fig. 4.10), plus-1 and self show obvious opposite relationship, while for SST (Fig. 4.15), plus-1 and minus-1 indicate opposite relationship, which may reflect the time-scale of persistence of each change. Second, the SP-CRN-2 (0,4,0), which considers the prior periodic cycle of four years, turns out the best prediction accuracy in PRD-2 of ZH300, and the drastic accuracy improvement when considering for more than four years of SST. Quasi-biennial and ENSO time-scale variations [39] are a possible cause of the four year periodic cycle that is occurred in SP-CRN-2 results.

Meanwhile, the proposed SP-CRN model has some limitations to consider. First, depending on the property of climate variable, the best combination of PRD type and its parameter is needed to be investigated. Second, this study considers the monthly data as the time resolution and the spatial input size of the Northern Hemisphere and the entire globe for the global scale forecasting. The higher time resolution causes the larger size in PRD (e.g., array size of 365 for daily data), and consumes the computational and memory resources. Third, extending the SP-CRN to multi-scale periodicity is possible as the original Periodic-CRN does (i.e., daily and weekly). However, multi-scale such as daily and yearly periodicity causes huge PRD as well. Therefore, some embedding on PRD to low dimensions is needed. Lastly, SP-CRN was

designed to work on the data with a periodic pattern. Therefore, other data with periodic patterns should be beneficial which need further investigation.

Furthermore, the proposed method of DSW can improve the prediction accuracy on ZH300 and SST when without metadata. The results of SSW with metadata were better than the DSW without metadata, which means the metadata with SSW is enough to obtain the best prediction result.

Chapter 5

Conclusion and Future Work

5.1 Summary

In this dissertation, the study of spatiotemporal climate prediction model, spatial and error distribution, and periodicity analysis have been made. The study experimented on ZH300 and SST, which are the important climate variables that can indicate global climate change.

Firstly, this study introduced the prediction of upper tropospheric circulations over the Northern Hemisphere by predicting ZH300 with multivariate features, which relates to the polar jet stream affecting the strength and the path of mid-latitude cyclones. The constructed predictive model ConvLSTM outperforms CNN and LR models in short- and long-term predictions. This result shows that ConvLSTM can maintain spatial correlation and remember important information through the network. Furthermore, the spatial error distribution of prediction output shows that ConvLSTM has much smaller errors at the paths of subtropical and polar jets. It is conceivable that considering the earlier six-months' data as the input of ConvLSTM and temporal variations in the ConvLSTM layer can improve accuracy compared to CNN and LR models. Then, the study investigated the importance of features to predict ZH300. The result shows that T300 and self prior of ZH300 are important features. Moreover, using only the important features can increase accuracy, which achieved less than 1% RMSE against ZH300 original value even for short- and long-term prediction.

Secondly, the Periodic-CRN model is adopted to use the periodicity component for climate forecasting, and proposed a Soft Periodic-CRN (SP-CRN) with three types of periodic representation dictionaries (PRD)—nearby-time (PRD-1), periodic-depth (PRD-2), and periodic-depth

differencing (PRD-3) representation—to overcome the phase shift and non-stationarity of periodic patterns. The experiments were conducted to perform periodicity on ZH300 and SST of the ERA-Interim dataset. The results showed the superiority of periodicity on SP-CRN (GRU) and SP-CRN (LSTM) over their base models (ConvGRU and ConvLSTM) in every length of sequential input. Furthermore, the PRD proposals showed that the effectiveness of SP-CRN in PRD-3 over PRD-2 and PRD-1, respectively, for ZH300 and SST predictions. Considering only the periodic-depth of two in PRD-3, drastically improved prediction accuracy. Moreover, the mixed proposal of SP-CRN-1+3 (1,0,2) slightly improved ZH300 prediction accuracy from their base models. The results of ZH300 and SST also showed that prediction accuracy improved when the PRD-2 considered the prior periodic cycle of four years. This may have indicated that the PRD-2 proposal could capture the four-year periodic cycle of ENSO. Lastly, in the attention weight analysis, the importance of one-month ahead of one prior periodic cycle had the most impact on the prediction.

5.2 Future Research Directions

The findings of this research may lead to future work in using periodicity in a spatiotemporal climate forecasting model. The study showed that including metadata on SSW in SP-CRN obtained the best prediction results, while the DSW drastically improved the accuracy compared to SSW when without metadata. However, the metadata together with DSW could not obtain the accuracy improvement. The metadata component with DSW (metadata-like) may conflict with each other in switching the periodicity phase. It would be interesting to conduct studies on combining DSW and metadata components to enhance the proposed method's performance. The idea would be to remove the metadata component and add convolutional layers after the DSW process before the fusion module would give more space to learn for metadata-like of DSW.

Another aspect is that the periodicity analysis on DSW. The DSW helped to capture the change of spatial weight, which is different from SSW that focuses on the attention weights inside the attention module. As the periodicity analysis on DSW in Sections 4.6.6 and 4.7.6, the rapid changed of attention weight in consecutive months were reduced. Meaning that there

would be drastic changes in spatial weights that needed further analysis.

Lastly, the proposed SP-CRN can be applied to multiple resolutions of periodicity as the original Periodic-CRN. It would lead to the utilization of the proposed method on climate data that has the multiple resolutions of periodicity. In general, climate data has multiple periodicity which is daily and yearly according to the earth rotation and revolution. However, the urban areas in which human activities mainly occurred on the ground level caused more periodicity resolutions (e.g., daily, weekly and monthly) for some climate data (e.g., the temperature at ground level). The SP-CRN with the multiple resolutions of periodicity component may help in capturing such kind of data.

Bibliography

[1] R. M. Adnan, M. Zounemat-Kermani, A. Kuriqi, and O. Kisi. *Machine Learning Method in Prediction Streamflow Considering Periodicity Component*, pages 383–403. Springer Singapore, 2021.

[2] N. Ahmed, G. Wang, M. J. Booij, A. Oluwafemi, M. Z.-u.-R. Hashmi, S. Ali, and S. Munir. Climatic variability and periodicity for upstream sub-basins of the yangtze river, China. *Water*, 12(3), 2020.

[3] M. Ali, R. C. Deo, N. J. Downs, and T. Maraseni. Multi-stage committee based extreme learning machine model incorporating the influence of climate parameters and seasonality on drought forecasting. *Computers and Electronics in Agriculture*, 152:149–165, 2018.

[4] D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate, *arXiv*, 2016.

[5] N. Ballas, L. Yao, C. Pal, and A. Courville. Delving deeper into convolutional networks for learning video representations, *arXiv*, 2016.

[6] I.-E. Bouznad, E. Guastaldi, A. Zirulia, M. Brancale, A. Barbagli, and D. Bengusmia. Trend analysis and spatiotemporal prediction of precipitation, temperature, and evapotranspiration values using the arima models: case of the algerian highlands. *Arabian Journal of Geosciences*, 13(24):1–17, 2020.

[7] R. Castro, Y. M. Souto, E. Ogasawara, F. Porto, and E. Bezerra. Stconvs2s: Spatiotemporal convolutional sequence to sequence network for weather forecasting. *Neurocomputing*, 426:285–298, 2021.

[8] W. Chan, N. Jaitly, Q. Le, and O. Vinyals. Listen, attend and spell: A neural network for large vocabulary conversational speech recognition. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4960–4964, 2016.

[9] P. Chen, A. Niu, D. Liu, W. Jiang, and B. Ma. Time series forecasting of temperatures using SARIMA: An example from nanjing. *IOP Conference Series: Materials Science and Engineering*, 394, 2018.

[10] R. Chen, W. Zhang, and X. Wang. Machine learning in tropical cyclone forecast modeling: A review. *Atmosphere*, 11(7), 2020.

[11] X. Chen and T. Zhou. Relative role of tropical sst forcing in the 1990s periodicity change of the pacific-japan pattern interannual variability. *Journal of Geophysical Research: Atmospheres*, 119(23):13,043–13,066, 2014.

[12] L. De Grandpré, J. C. Tardif, A. Hessl, N. Pederson, F. Conciatori, T. R. Green, B. Oyunsanaa, and N. Baatarbileg. Seasonal shift in the climate responses of pinus sibirica, pinus sylvestris, and larix sibirica trees from semi-arid, north-central mongolia. *Canadian Journal of Forest Research*, 41(6):1242–1255, 2011.

[13] J. Diez-Sierra and M. del Jesus. Long-term rainfall prediction using atmospheric synoptic patterns in semi-arid climates with statistical and machine learning methods. *Journal of Hydrology*, 586, 2020.

[14] T. Dimri, S. Ahmad, and M. Sharif. Time series analysis of climate variables using seasonal arima approach. *Journal of Earth System Science*, 129(1):1–16, 2020.

[15] S. B. Feldstein. The dynamics of nao teleconnection pattern growth and decay. *Quarterly Journal of the Royal Meteorological Society*, 129(589):901–924, 2003.

[16] S. Hochreiter and J. Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 1997.

[17] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear

and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971):903–995, 1998.

[18] R. Kane. Periodicities, ENSO effects and trends of some South African rainfall series: an update. *South African Journal of Science*, 105:199 – 207, 2009.

[19] S. Kim, S. Hong, M. Joh, and S.-k. Song. DeepRain: ConvLSTM network for precipitation prediction using multichannel radar data. *Proc. of International Workshop on Climate Informatics*, pages 89–92, 2017.

[20] S. Kim, J.-S. Kang, M. Lee, and S.-K. Song. Deeptc: Convlstm network for trajectory prediction of tropical cyclone using spatiotemporal atmospheric simulation data. *Neural Information Processing Systems Spatiotemporal Workshop*, 2018.

[21] S. Kim, J.-S. Kang, S.-k. Song, C.-G. Park, and B.-J. Kim. DeepRainK: ConvLSTM network for precipitation prediction using hybrid surface rainfall radar data. *Proc. of International Workshop on Climate Informatics*, pages 83–86, 2019.

[22] E. W. Kolstad, T. Breiteig, and A. A. Scaife. The association between stratospheric weak polar vortex events and cold air outbreaks in the northern hemisphere. *Quarterly Journal of the Royal Meteorological Society*, 136(649):886–893, 2010.

[23] Y.-l. Kong, Y. Meng, W. Li, A.-z. Yue, and Y. Yuan. Satellite image time series decomposition based on eemd. *Remote Sensing*, 7(11):15583–15604, 2015.

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.

[25] Y. Lai and D. A. Dzombak. Use of the autoregressive integrated moving average (arima) model to forecast near-term regional temperature and precipitation. *Weather and Forecasting*, 35(3):959 – 976, 2020.

[26] M. Latif and A. Grötzner. The equatorial atlantic oscillation and its response to enso. *Climate Dynamics*, 16(2):213–218, 2000.

[27] H. Li and W. Song. Characteristics of climate change in the lancang-mekong sub-region. *Climate*, 8(10), 2020.

[28] B. Lim and S. Zohren. Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2194), 2021.

[29] M. Liu, Y. Huang, Z. Li, B. Tong, Z. Liu, M. Sun, F. Jiang, and H. Zhang. The applicability of lstm-knn model for real-time flood forecasting in different climate zones in china. *Water*, 12(2), 2020.

[30] D. H. Nguyen, J.-B. Kim, and D.-H. Bae. Improving radar-based rainfall forecasts by long short-term memory network in urban basins. *Water*, 13(6), 2021.

[31] D. Ponnoprat. Short-term daily precipitation forecasting with seasonally-integrated autoencoder. *Applied Soft Computing*, 102, 2021.

[32] J. Rao, R. Ren, H. Chen, Y. Yu, and Y. Zhou. The stratospheric sudden warming event in february 2018 and its prediction by a climate system model. *Journal of Geophysical Research: Atmospheres*, pages 332–345, 2018.

[33] S. Rasp, P. D. Dueben, S. Scher, J. A. Weyn, S. Mouatadid, and N. Thuerey. Weatherbench: A benchmark data set for data-driven weather forecasting. *Journal of Advances in Modeling Earth Systems*, 12(11), 2020.

[34] X. Ren, X. Li, K. Ren, J. Song, Z. Xu, K. Deng, and X. Wang. Deep learning-based weather prediction: A survey. *Big Data Research*, 23, 2021.

[35] N. Savelieva, I. Semiletov, L. Vasilevskaya, and S. Pugach. A climate shift in seasonal values of meteorological and hydrological parameters for northeastern asia. *Progress in Oceanography*, 47(2):279–297, 2000.

[36] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. *In Proceedings of 28th Conference on Neural Information Processing Systems*, pages 802–810, 2015.

[37] M. Shiotani. Annual, quasi-biennial, and el niño-southern oscillation (enso)time-scale variations in equatorial total ozone. *Journal of Geophysical Research: Atmospheres*, 97(D7):7625–7633, 1992.

[38] L. Tomassini, E. P. Gerber, M. P. Baldwin, F. Bunzel, and M. Giorgetta. The role of stratosphere-troposphere coupling in the occurrence of extreme winter cold spells over northern europe. *Journal of Advances in Modeling Earth Systems*, 4(4), 2012.

[39] T. Tomita. The longitudinal structure of interannual variability observed in sea surface temperature of the equatorial oceans. *Journal of the Meteorological Society of Japan. Ser. II*, 78(4):499–507, 2000.

[40] T. Tomita and T. Yamaura. A precursor of the monthly-mean large-scale atmospheric circulation anomalies over the north pacific. *Scientific Online Letters on the Atmosphere*, 13:85–89, 2017.

[41] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu. Wavenet: A generative model for raw audio, arXiv, 2016.

[42] B. Wang, J. Lu, Z. Yan, H. Luo, T. Li, Y. Zheng, and G. Zhang. Deep uncertainty quantification: A machine learning approach for weather forecasting. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2087–2095. Association for Computing Machinery, 2019.

[43] C. Wang and Y. Hong. Application of Spatiotemporal Predictive Learning in Precipitation Nowcasting. In *AGU Fall Meeting Abstracts*, volume 2018, pages H31H–1988, 2018.

[44] Y. Wang, M. Long, J. Wang, Z. Gao, and P. S. Yu. Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 879–888, 2017.

[45] Y. Wang, J. Zhang, H. Zhu, M. Long, J. Wang, and P. S. Yu. Memory in memory: A predictive neural network for learning higher-order non-stationarity from spatiotemporal

dynamics. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9146–9154, 2019.

[46] B. M. Williams and L. A. Hoel. Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results. *Journal of Transportation Engineering*, 129(6):664–672, 2003.

[47] T. Xie and J. Ding. Forecasting with multiple seasonality. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 240–245, 2020.

[48] J. Yin, Z. Deng, A. V. Ines, J. Wu, and E. Rasu. Forecast of short-term daily reference evapotranspiration under limited meteorological variables using a hybrid bi-directional long short-term memory model (Bi-LSTM). *Agricultural Water Management*, 242, 2020.

[49] K. Yonekura, H. Hattori, and T. Suzuki. Short-term local weather forecast using dense weather station by deep neural network. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 1683–1690, 2018.

[50] R. Yu, S. Zheng, A. Anandkumar, and Y. Yue. Long-term forecasting using higher order tensor RNNs, *arXiv*, 2019.

[51] H. Zhou, E. Aizen, and V. Aizen. Constructing a long-term monthly climate data set in central asia. *International Journal of Climatology*, 38(3):1463–1475, 2018.

[52] Y. Zhou, H. Dong, and A. El Saddik. Deep learning in next-frame prediction: A benchmark review. *IEEE Access*, 8:69273–69283, 2020.

[53] A. Zirulia, M. Brancale, A. Barbagli, E. Guastaldi, and T. Colonna. Hydrological changes: are they present at local scales? *Rendiconti Lincei. Scienze Fisiche e Naturali*, 32(2):295–309, 2021.

[54] A. Zonoozi, J. jae Kim, X.-L. Li, and G. Cong. Periodic-crn: A convolutional recurrent model for crowd density prediction with recurring periodic patterns. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 3732–3738, 2018.

Appendix A

List of Publication

A.1 Journal

- Ekasit Phermphoonphiphat, Tomohiko Tomita, Takashi Morita, Masayuki Numao, and Ken-ichi Fukui. "Soft Periodic Convolutional Recurrent Network for Spatiotemporal Climate Forecast", Applied Sciences., 11 (20). 2021.

A.2 International Conference/Symposium (with review)

- Ekasit Phermphoonphiphat, Tomohiko Tomita, Masayuki Numao, and Ken-ichi Fukui. "A Study of Upper Tropospheric Circulations over the Northern Hemisphere Prediction Using Multivariate Features by ConvLSTM", Proc. the 23rd Asia-Pacific Symposium on Intelligent and Evolutionary Systems (IES-2019), pp. 130-141, Tottori, Japan, Dec. 2019.

A.3 Domestic Workshop (without review)

- Ekasit Phermphoonphiphat, Tomohiko Tomita, Masayuki Numao, and Ken-ichi Fukui. "Preliminary Study on Machine Learning Approach for Climate Change Prediction", 日本気象学会2017年度秋季大会, Oct. 2017.

- Ekasit Phermphoonphiphat, Tomohiko Tomita, Masayuki Numao, and Ken-ichi Fukui. "Climate Forecasting by ConvLSTM on Segmented Region", 電子情報通信学会人工知能と知識処理研究会 (AI), Aug. 2018.

- Ekasit Phermphoonphiphat, Tomohiko Tomita, Masayuki Numao, and Ken-ichi Fukui.

"Spatiotemporal Climate Forecasting with ConvLSTM", 日本地球惑星科学連合2019年大会, May 2019.

- Ekasit Phermphoonphiphat, Tomohiko Tomita, Masayuki Numao, and Ken-ichi Fukui. "A Prediction of Upper Tropospheric Circulations over the Northern Hemisphere Using ConvLSTM", 電子情報通信学会人工知能と知識処理研究会 (AI), Sep. 2019.

- Ekasit Phermphoonphiphat, Tomohiko Tomita, Masayuki Numao, and Ken-ichi Fukui. "Climate Prediction by Using Periodic Convolutional Recurrent Network", 日本地球惑星科学連合2020年大会, May 2020.

- Ekasit Phermphoonphiphat, Tomohiko Tomita, Masayuki Numao, and Ken-ichi Fukui. "Periodic Convolutional Recurrent Network for Climate Prediction", 人工知能学会全国大会2020年大会, Jun. 2020.

- Ekasit Phermphoonphiphat, Tomohiko Tomita, Takashi Morita, Masayuki Numao, and Ken-ichi Fukui. "Soft Periodic Convolutional Recurrent Network for Spatiotemporal Climate Forecast and Periodicity Analysis", 人工知能学会第20回データ指向構成マイニングとシミュレーション（SIG-DOCMAS）研究会, Nov. 2021.