

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

Phermphoonphiphat Ekasit¹, Tomohiko Tomita², Takashi Morita³,
Masayuki Numao³, and Ken-ichi Fukui³

¹ Graduate School of Information Science and Technology, Osaka University

² Faculty of Advanced Science and Technology, Kumamoto University

³ SANKEN (The Institute of Scientific and Industrial Research), Osaka University

Abstract: Many machine-learning applications and methods are emerging to solve problems associated with spatiotemporal climate forecasting; however, a prediction algorithm that considers only short-range sequential information may not be adequate to deal with periodic patterns such as seasonality. Phase shifts and non-stationarity of periodicity are the key components in the model to support. In this paper, we propose 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 could weight the importance of periodic representations, which helped in capturing the periodic pattern. In order to capture the spatial change over the periodicity phase, we also propose dynamic spatial weights (DSW) on attention module, which are multiple spatial weights for the attention module output, and switched according to the prediction month. We experimented on geopotential height at 300 hPa (ZH300) and sea surface temperature (SST) datasets of ERA-Interim. The results showed that the proposed method improved the prediction accuracy and could learn the periodicity from the data, which capture the phase shift and non-stationarity of periodicity in ZH300 and SST.

1 Introduction

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 numerical weather prediction (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 and long-term rainfall prediction [1].

The spatiotemporal forecasting problem in the climate domain given by the spatial and temporal data is crucial for predicting the next time frame. The convolutional long short-term memory network (ConvLSTM) [2] 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, 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 [3]. Looking deeper into the periodic patterns, there are

two periodic patterns that we consider in this research; (1) the phase shift, such as the late summer and early autumn, and (2) non-stationarity of periodicity. Climate seasonal shifts in terms of spatial and intensity may not be the same every year. 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 are suitable for spatiotemporal forecasting, they do not consider periodicity. The Committee ELM (Comm-ELM) [4] added a digit for the month as metadata as one of inputs to indicate yearly periodicity. Meanwhile, Periodic Convolutional Recurrent Network (Periodic-CRN) [5] 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 dictionary were used through an attention mechanism [6] to improve the forecasting accuracy of the crowd density across two taxi datasets from Beijing and Singapore. However, all of these methods assume stationary periodicity, which means they can not capture the phase shift or non-stationarity of periodicity patterns.

Based on the idea of Periodic-CRN, we propose Soft Periodic-CRN (SP-CRN) [7] with three types of

PRD that improve the prediction accuracy by considering phase shift and non-stationarity of periodicity: Nearby-Time, Periodic-Depth, and Periodic-Depth-Differencing representations. In the work [7], the spatial weight on attention module was only single spatial weight which we named as the static spatial weights (SSW) on attention module in this paper. In order to capture the spatial change over the periodicity phase, we propose the dynamic spatial weights (DSW) on attention module. The experiments compared the model performances to predict one-month ahead for a geopotential height at pressure 300 hPa (ZH300) and sea surface temperature (SST). Then we analyzed the periodicity of attention weight to indicate the importance of periodic representation for the prediction. The accurate prediction and the understanding of periodic patterns may help to indicate the threats of nature.

2 Materials and Methods

2.1 Overview

The entire architecture is shown in Fig. 1. The process of this method begins with learning the sequential input in the CRN component (Fig. 1(a)). Then, the hidden state output from the CRN component is saved as a periodic representation (Fig. 1(b)), and it is loaded according to the proposal described in Section 2.4. Next, to estimate the relevance of each periodic representation to the hidden state output, the representations are weighted by the attention module (section 2.5) as the representation output from the periodic representation component. The hidden state output of the two components is combined (fusion module in Fig. 1(f)), also the extended idea of DSW is explained in Section 2.6. Finally, the metadata component (Fig. 1(g)) is added as the original Periodic-CRN has, then combined to the output from fusion module as described in Section 2.7 to obtain the prediction.

2.2 Problem Statement

We regard our climate prediction task as similar to a next video frame prediction, 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. (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 ($\tau = 1, 2, 3, \dots$), and $X^{(t)}, \dots, X^{(t-\tau+1)}$ are spatial sequence input data.

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

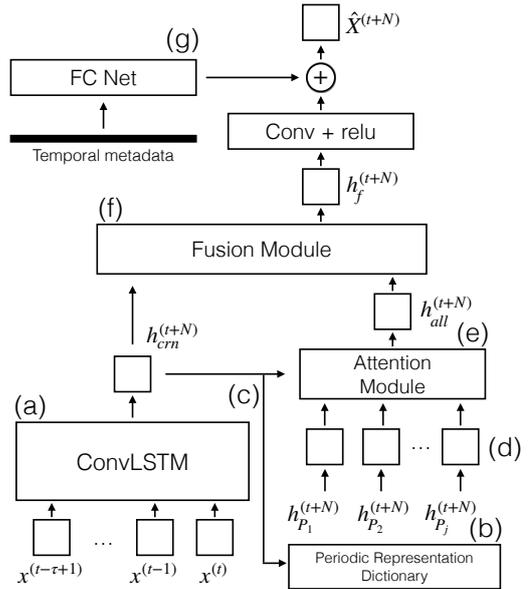


Figure 1: The SP-CRN (LSTM) architecture. (a) the CRN component, (b) the PRD, (c) update mechanism, (d) load mechanism, (e) the attention module, (f) fusion module, and (g) metadata component.

2.3 CRN Component

To capture spatiotemporal patterns of an input sequence, we need a network that considers spatial dependencies. The convolution operator inside CRN component (Fig. 1(a)) is responsible for extracting spatial latent features from a low to high level. We selected ConvLSTM as a CRN component for this research, and named it SP-CRN (LSTM).

2.4 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. 1(c),(d).

2.4.1 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 2.4.2. 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 ;

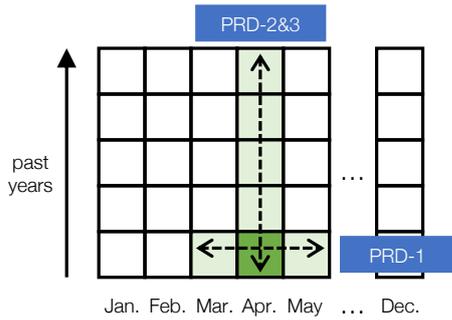


Figure 2: The conceptual diagram of the proposed PRD in case of monthly representation. Each cell indicates a saved representation for the future reference. Type PRD-1 refers adjacent months of the last year, and types PRD-2 and 3 refer the same month of prior consecutive years.

the second, save new representation, the new representation is saved to the bottom of P .

2.4.2 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 is the representation of the previous periodic cycle. Fig. 2 shows the conceptual diagram of the proposed PRD.

Nearby-Time Representation (PRD-1): In this work, to capture a phase shift that may change every year, we extended the periodic attention mechanism idea from the original Periodic-CRN to include nearby (months) representations for the previous year. For example, to find a yearly periodic representation of July 1999 where the nearby month is 2, the representations of May, June, July, August, and September of 1998 are loaded.

Periodic-Depth Representation (PRD-2): The periodic representation in the first proposal, the nearby time representation, considers only one year’s prior representation. We extended the representation to be more than one period prior, and use only the same time index (phase). The load mechanism loads all the representations that are in the same phase of the previous cycles. For example, to find a periodic representation of July 1999 where the depth is 5, the representation of July 1998, 1997, 1996, 1995, and 1994 are loaded.

Periodic-Depth-Differencing Representation (PRD-3): Similar to the second proposal, we considered the change over cycles by taking the difference between the month of two adjacent years. In this proposal, we used a differencing of two years. The load mechanism loaded the differencing of representation. For example, to find a periodic representation of July 1999 where the depth is 5, the

representation of July 1998 minus 1997, 1997 minus 1996, 1996 minus 1995, and 1995 minus 1994 are loaded.

2.5 Attention Module

In this method, as shown in Fig. 1(e), we merged the output from CRN component and the loaded representations from the PRD. The attention mechanism [6] was used to weight the multiple representations, which indicated the importance of each representation according to the softmax operation.

2.6 Fusion Module

The periodic representation from attention module and the representation output from CRN component are now combined in this module as shown in Fig. 1(f). Then transform the output shape with convolution operation to obtain the prediction output.

The SP-CRN [7] uses static spatial weights (SSW) on the output from the attention module before input to the fusion module. However, the SSW is a spatial weight regardless of the prediction month and may not denote the change over the periodicity phase. This paper proposes dynamic spatial weights (DSW) on attention module, a set of spatial weights with the size of periodicity phase (the size of 1 year periodicity is 12), which will be switched according to the prediction month during both training and inference.

2.7 Metadata component

The metadata component idea is added to the model as it is in the original Periodic-CRN [5] to indicate the month of the year with one-hot encoding. The metadata component was added at the end of the fusion module in Fig. 1(f), and Section 2.6. The metadata component may help the model indicating the phase of periodicity better to increase prediction accuracy.

3 Experiment Settings

3.1 ERA-Interim Dataset

ERA-Interim¹ is a dataset from the European Centre for Medium-Range Weather Forecasts (ECMWF), an objective re-analysis of a global atmospheric variable from 1 January 1979 to 31 August 2019. The dataset includes a 4-dimensional variational analysis (4D-Var): latitude, longitude, air pressure (height), and time. The monthly means of daily means was used as a time resolution. To validate the proposed method performance, we chose 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. 3 shows the sample of datasets.

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

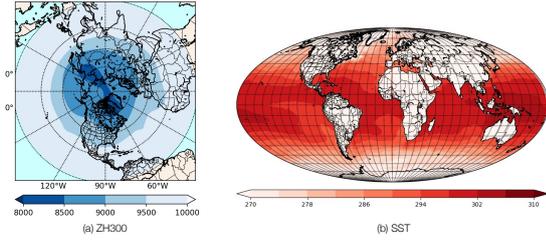


Figure 3: A sample of the monthly spatial plot of January 2018. (a) The Northern Hemisphere of ZH300 data, where deep blue represents the lower value of ZH300. (b) The entire globe in the spatial grid of SST data, where deep red represents high temperature, and the terrain area (white) is ignored. Note 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 [8]. In this paper, Z300 is converted to a geopotential height at pressure level 300 hPa (ZH300) divided by gravitational acceleration (9.8 m/s^2). Then we extracted the data 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. We chose a global SST with a 1 degree grid resolution, which is the spatial size of 180 (latitude) \times 360 (longitude).

3.2 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 [9], the first benchmark for data-driven, medium-range climate prediction.

3.3 Comparison Methods

Convolutional neural networks (CNN): 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, we used five convolutional layers, the last being the output layer with an output channel of one.

ConvLSTM: To determine the benefit of using an LSTM structure that captures change over time, we adopted a ConvLSTM as a baseline method. We replaced the first two layers with ConvLSTM from the

CNN baseline model; that is, two layers of ConvLSTM and three layers of CNN. Then we applied consecutive a prior spatial input to the model.

SP-CRN (LSTM): As described in Section 2, we applied the ConvLSTM as the CRN component for the SP-CRN model. The SP-CRN (LSTM) will undergo further experimentation for our periodicity proposals.

3.4 Settings

We conducted the experiments to test the hyperparameters on CNN and ConvLSTM to find suitable ones for our 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, we selected the 3×3 kernel and 16 filters for the time of our computational resources. Then we replaced the CNN layer 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.

We 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 and ConvLSTM 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. We adopted early stopping with 100 epoch patience on validation loss. We 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. 3b). We replaced the terrain area with 0 and ignored it when using the standardized method in the training dataset and MSE in the loss function. The same hyperparameters as ZH300 were used: a 3×3 kernel and 16 filters. We also used the same set of comparison methods.

4 Results

4.1 Results on ZH300

4.1.1 PRD Types

We used the SP-CRN (LSTM) as the model with periodicity component to perform the PRD proposals. Note that SP-CRN refers to SP-CRN (LSTM) from

Table 1: Evaluation (average RMSE and SD of average RMSE in meter) of different settings on the ZH300 prediction for SP-CRN and comparison methods. 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.

Method	RMSE		SD	
	SSW	DSW	SSW	DSW
CNN	141.79		16.54	
ConvLSTM	126.75		9.70	
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-M (1,0,2)	50.27	64.02	12.77	10.13

now on. We have tested with varying input sequence lengths and confirmed that five months is enough; therefore, an input sequence length of five months was selected as the baseline setting. We experimented on the SP-CRN for the three proposals in different settings. We have tested various settings [7] and showed only the best setting in each PRD type in Table 1. Note that the results explained in this section were based on SSW, and did not include the metadata.

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 result of SP-CRN-1 indicated that the nearby-time representation of plus and minus one month in one prior periodic cycle improved the prediction accuracy. SP-CRN-2 proposal result indicated that the importance of the information of the previous four periodic cycles for ZH300 prediction and possibly for the four-year oscillation of ENSO [10]. SP-CRN-3 outperformed the baseline result, of which SP-CRN-3 had an RMSE of 67.97 m. The result 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).

4.1.2 Combination of PRD Types and Metadata

We mixed the proposals of SP-CRN-1, SP-CRN-2, and SP-CRN-3, and combined with all the combinations. Then, we selected SP-CRN-1+3 as the mixed proposal to be added to the metadata component (SP-CRN-1+3-M).

The mixed proposal of SP-CRN-1+3 was the best compared with the other mixed proposals, which improved SP-CRN-1 and SP-CRN-3 accuracy with an

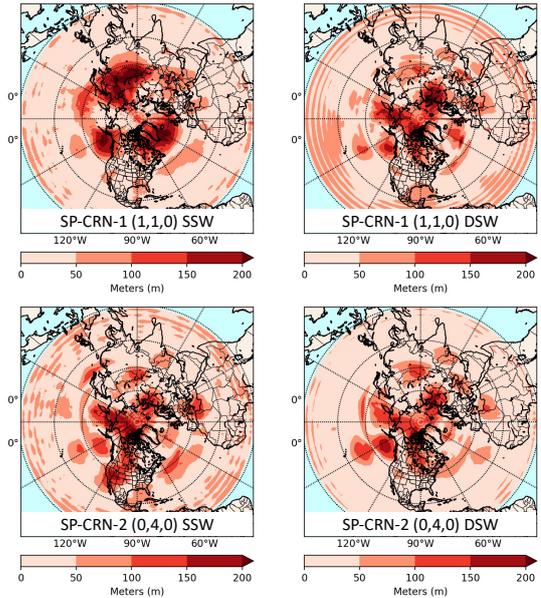


Figure 4: The ZH300 error distribution of SP-CRN-1 (1,1,0) SSW, SP-CRN-1 (1,1,0) DSW, SP-CRN-2 (0,4,0) SSW, and SP-CRN-2 (0,4,0) DSW in January 2018, with RMSE of 72.51 m, 63.16 m, 63.03 m, and 55.94, respectively.

RMSE of 66.39 m 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-M was superior at indicating the phase of periodicity and had the lowest RMSE comparing to SP-CRN-1+3 (base model).

4.1.3 Static Spatial Weights vs Dynamic Spatial Weights

We experimented on all SP-CRN settings with DSW to check the availability of dynamic spatial weight for attention module output. The results in Table 1 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.

We plotted the error distribution (Fig. 4) 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 1. 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.

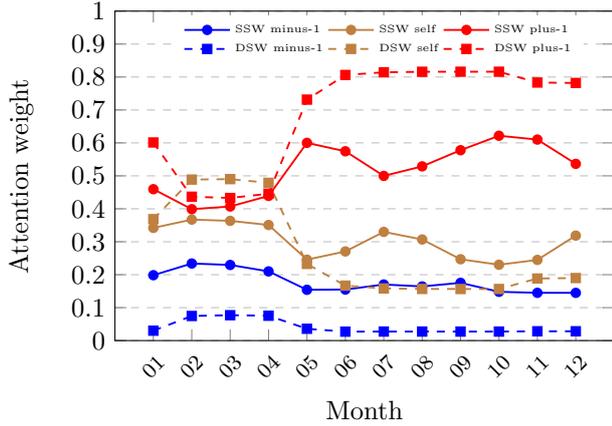


Figure 5: The average attention weight of training data on SP-CRN-1 (1,1,0) for ZH300 prediction from the average of five runs.

4.1.4 Periodicity Analysis

We analyzed the weight in the attention module (Section 2.5), which indicated the importance of the periodic representation collected in the PRD. Fig. 5 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. 6 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 weight of DSW, all the depths were not significantly different.

4.2 Results on SST

4.2.1 PRD Types

We have also tested with varying input sequence lengths, and confirmed that the five months is enough, which was the same input length as ZH300 experiments. We experimented for the three proposals compared with various settings for SP-CRN, and selected the best setting result for each proposal. Table 2 shows the best result of each PRD type and

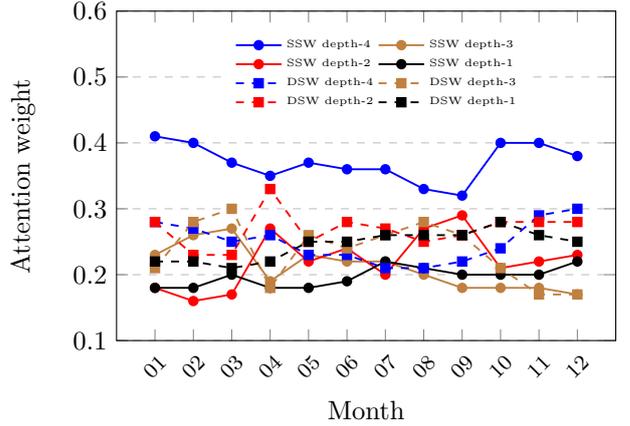


Figure 6: The average attention weight of training data on SP-CRN-2 (0,4,0) for ZH300 prediction from the average of five runs. The attention weight of depth-1 to depth-4 was the value of the prediction month for one to four prior periodic cycles.

Table 2: Evaluation (average RMSE and SD of average RMSE in Kelvin) 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	
ConvLSTM	3.02		0.06	
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-M (1,0,2)	0.43	0.46	0.03	0.03

combinations of SP-CRN. The results for SP-CRN-1 were slightly better than for SP-CRN (baseline). SP-CRN-3 outperformed the baseline result, for which SP-CRN-3 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.

4.2.2 Combination of PRD Types and Metadata

Table 2 shows the results of all mixing proposals. SP-CRN-1+3 was the best of all mixed proposals with an RMSE of 0.56 K, which was no statistical difference to SP-CRN-3 (baseline). Mixing the proposals produced no accuracy improvement from baseline. Meanwhile, the metadata component was added to SP-CRN-1+3

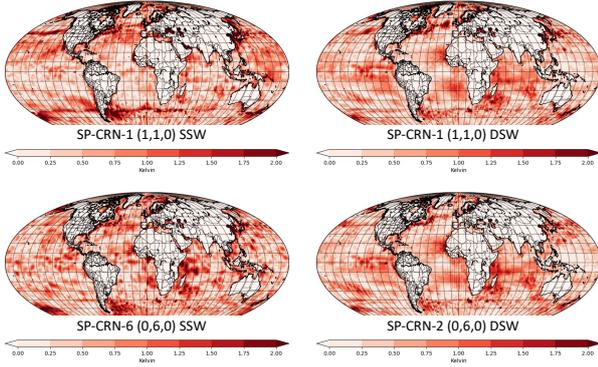


Figure 7: The SST error distribution of SP-CRN-1 (1,1,0) SSW, SP-CRN-1 (1,1,0) DSW, SP-CRN-2 (0,6,0) SSW, and SP-CRN-2 (0,6,0) DSW (RMSE of 0.60 K) in January 2018 with RMSE of 0.75 K, 0.60 K, 0.65 K, and 0.54 K, respectively.

to indicate its performance for an SST prediction of the best mixed proposal. The metadata component of SP-CRN-1+3-M, 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 (baseline).

4.2.3 Static Spatial Weights vs Dynamic Spatial Weights

The results of DSW (Table 2) 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. 7 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.

4.2.4 Periodicity Analysis

Fig. 8 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. 5) 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. 9 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

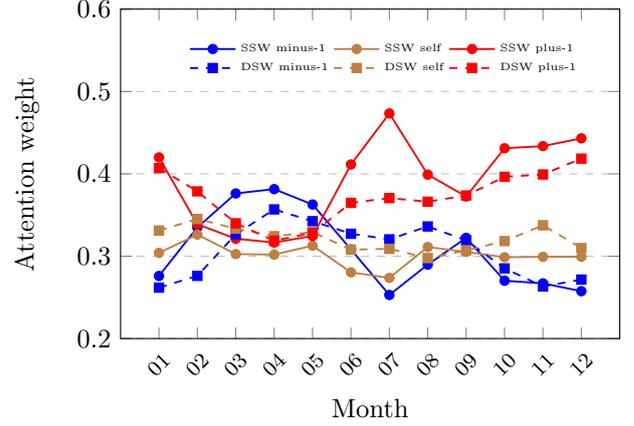


Figure 8: The average attention weight of training data on SP-CRN-1 (1,1,0) for SST prediction from the average of five runs.

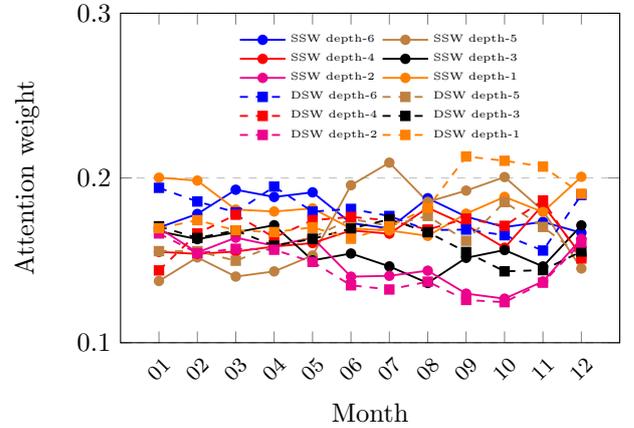


Figure 9: The average attention weight of training data on SP-CRN-2 (0,6,0) for SST prediction from the average of five runs.

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.

5 Discussion

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 (a term of three months), climate variables may have a similar changing tendency. In addition, for ZH300 (Fig. 5), plus-1 and self show obvious opposite relationship, while for SST (Fig. 8), plus-1 and minus-1 indicate opposite relationship, which may reflect the time-scale of persistence of each change.

Second, the SP-CRN-2, which considers the prior periodic cycle of four years, turns out the best predic-

tion 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 [11] are a possible cause of the four year periodic cycle that is occurred in SP-CRN-2 results.

Lastly, 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. The metadata component with DSW (metadata-like) may conflict with each other in switching the periodicity phase. On the other hand, regarding periodicity analysis, we need further analysis of how spatial weights affect attention weights on the DSW.

6 Conclusions

In this paper, we adopted the Periodic-CRN model 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 (LSTM) over their base model (ConvLSTM). Moreover, our PRD proposals showed that the effectiveness of SP-CRN in PRD-3 over PRD-2 and PRD-1, respectively, for ZH300 and SST predictions. Furthermore, the dynamic spatial weights (DSW) on attention module were applied to the model, and improved prediction results over the static spatial weights except for the SP-CRN with metadata component. Lastly, in the attention weight analysis, the importance of one-month ahead of one prior periodic cycle had the most impact on the prediction.

Acknowledgment

This work was supported in part by the Network Joint Research Center for Materials and Devices and by JSPS KAKENHI Grant Number 19K22876.

References

[1] Diez-Sierra J., del Jesus M.: Long-term rainfall prediction using atmospheric synoptic patterns in semi-arid climates with statistical and machine learning methods, *J. Hydrol.*, Vol. 568, (2020)

[2] Shi X., Chen Z., Wang H., Yeung D.Y., Wong W.K., Woo W.C.: Convolutional LSTM network: A machine learning approach for precipi-

tation nowcasting, In *Proceedings of 28th Conference on Neural Information Processing Systems*, pp. 802–810, (2015)

[3] Feldstein S.B.: The dynamics of NAO teleconnection pattern growth and decay, *Q. J. R. Meteorol. Soc.*, Vol. 129, No. 589, pp. 901-924, (2003)

[4] Ali M., Deo R.C., Downs N.J., Maraseni T.: Multi-stage committee based extreme learning machine model incorporating the influence of climate parameters and seasonality on drought forecasting., *Comput. Electron. Agric.*, Vol. 152, pp. 149-165, (2018)

[5] Zonoozi A., jae Kim J., Li X.L., Cong G.: 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*, pp. 3732-3738, (2018)

[6] Bahdanau D., Cho K., Bengio Y.: Neural Machine Translation by Jointly Learning to Align and Translate, *arXiv*, (2016)

[7] Phermphoonphiphat E., Tomita T, Morita T., Numao M., Fukui K.: Soft Periodic Convolutional Recurrent Network for Spatiotemporal Climate Forecast, *Applied Sciences*, Vol. 11, No. 20, (2021)

[8] Tomita T. and Yamaura T.: A Precursor of the Monthly-Mean Large-Scale Atmospheric Circulation Anomalies over the North Pacific., *Sci. Online Lett. Atmos.*, Vol. 13, pp. 85-89, (2017)

[9] Rasp S., Dueben P.D., Scher S., Weyn J.A., Mouatadid S., Thuerey N.: WeatherBench: A Benchmark Data Set for Data-Driven Weather Forecasting, *J. Adv. Model. Earth Syst.*, Vol. 12, (2020)

[10] Shiotani M.: Annual, quasi-biennial, and El Niño-Southern Oscillation (ENSO) time-scale variations in equatorial total ozone., *J. Geophys. Res. Atmos.*, Vol. 97, No. D7, pp. 7625-7633, (1992)

[11] Tomita T.: The Longitudinal Structure of Interannual Variability Observed in Sea Surface Temperature of the Equatorial Oceans., *Journal of the Meteorological Society of Japan. Ser. II*, Vol. 78, No. 4, pp. 499-507, (2000)