Study on the Influence of Attention Mechanism in Large-Scale Sea Surface Temperature Prediction Based on Temporal Convolutional Network

Yuan Feng, Tianying Sun(✉), and Chen Li

College of Information Science and Engineering, Ocean University of China, Qingdao 266100, China
suntianying@stu.ouc.edu.cn

Abstract. The short term and small-scale sea surface temperature prediction using deep learning has achieved good results. The long-term sea surface temperature prediction technology in large-scale sea area is limited by the large and complex data. So how to use deep learning to select more valuable data and realize high precision of sea surface temperature prediction is an important problem. In this paper, attention mechanism and Temporal convolutional network (TCN) are used to predict the Indian Ocean 40°E–110°E and –25°S–25°N from 2015 to 2018 with $1° \times 1°$ spatial resolution. The attention mechanism is used to distinguish the importance of the data, and the prediction models of full-feature (81 dimensions) and partial-feature (66 dimensions) are constructed. The experimental results show that the fitting degree of partial-feature models to sea surface temperature time series does not decrease significantly. The method proposed in this paper uses less data to ensure that the experimental accuracy does not decline significantly, and improves the long-term sea surface temperature prediction technology in large-scale sea area.

Keywords: Sea surface temperature · Attention mechanism · Temporal convolutional network

1 Introduction

Sea surface temperature is an important contributor to the health of regional marine ecosystems [1], and its changing trend may lead to the growth, reproduction and distribution of marine species. The rapid ocean warming trend will have a strong impact on marine fisheries [2]. The prediction of sea surface temperature has important guiding significance for large and medium-scale ocean physical phenomena. For example, the definition of the Indian Ocean Dipole (IOD) index is related to the abnormal changes of regional sea surface temperature. Large-scale annual sea surface temperature forecasts are helpful for climate monitoring, flood and drought risk warning and other aspects. The Indian Ocean will affect the surrounding areas, central South America, the southern

J. Xiong et al. (Eds.): MobiMedia 2021, LNICST 394, pp. 727–735, 2021.
https://doi.org/10.1007/978-3-030-89814-4_53

tip of Africa, southeastern Australia, Northeast Asia and other regions to have climate anomalies [3–6]. In this paper, the Indian Ocean is selected as the research object to make long-term forecast of sea surface temperature.

Dong et al. used the CFCC-LSTM neural network to predict the sea surface temperature 1 day, 7 days, and 30 days in advance on the Bohai Sea dataset (spatial resolution $0.05° \times 0.05°$). Experimental results show that the CFCC-LSTM model prediction error (RMSE) is 0.1466 °C, 0.2722 °C and 0.7260 °C [7]. Zhong et al. also conducted short-term predictions of sea surface temperature in a small area on the Bohai Sea dataset (spatial resolution 0.25° * 0.25°), and used the LSTM layer to model the time series data of sea surface temperature. The experimental results show that LSTM The model predicts RMSE for 1, 7, and 30 days to be 0.0767 °C, 0.3844 °C, and 0.3928 °C [8]. Yang et al. improved the fully connected LSTM model and made 1, 7 and 30-day lead time predictions for the Bohai Sea and the East China Sea. The prediction results show that the longer the prediction lead time, the larger the prediction area and the lower the prediction accuracy [9]. Guan et al. used the entire China Sea and its adjacent sea areas as the study area (spatial resolution $0.05° \times 0.05°$), classified the sea area into 130 small areas through the SOM algorithm, and established LSTM models to predict sea surface temperature. The RMSE is 0.5 °C (one month in advance), 0.66 °C (12 months in advance). Building and training a model for each small area requires a lot of calculations, especially in large research areas with high spatial resolution [10]. The current research selects daily data with small time granularity and high spatial resolution to predict short-term sea surface temperature in a small area of sea, and only uses sea surface temperature for prediction. There are relatively few studies on long-period sea temperature forecasting in the ocean.

The study of large and medium scale physical ocean phenomena needs to deal with a large number of complicated ocean data. This paper uses the monthly ocean-atmosphere data of the past ten years and five months to predict the sea surface temperature of large-scale sea areas, 7 months in advance. The attention mechanism selects special features and compares the effects of partial-feature models and full-feature models. Experiments show that the data set used has low spatial resolution and large time granularity. The attention mechanism reduces the data set but the experimental accuracy does not decrease. The method this paper proposed is more suitable for the study of physical ocean phenomena in the ocean.

2 Data and Method

2.1 A Subsection Sample

This paper faces the interior of the Indian Ocean, which is the third largest ocean in the world (30°E–135°E, 30°N–66.5°S), located between Asia, Oceania, Africa and Antarctica. This paper uses the reanalysis data set provided by the National Center for Meteorological and Environmental Prediction (NCEP) with a spatial resolution of $1° \times 1°$, which includes atmospheric temperature, geopotential height, vertical velocity, water vapor, east-west wind speed, north-south wind speed, and undersea data such as temperature, Ocean current velocity in east-west and north-south directions. These monthly ocean-atmosphere data are available at https://psl.noaa.gov/data/gridded/.

In this paper, we select monthly data with longitude (40°E–110°E) and latitude (−25°S–25°N) from 1980 to 2018, and organize two sets of data sets to model and forecast separately. This paper selects the data for ten consecutive years and January to May in the eleventh year to predict the sea surface temperature from June to December in the eleventh year. The surface of the ocean is dynamically affected by waves, wind shear, and heat exchange, and the mixing of thermal expansion, ocean circulation and turbulence from the interior of the ocean will also produce dynamic effects [11]. Therefore, each month includes atmospheric, sea surface, and subsea parameter factors (81 in total). This paper selects atmospheric temperature, geopotential height, vertical speed, water vapor, east-west wind speed, north-south wind speed, at different heights (1000 850 500 300 hPa), a total of 24 atmospheric factors. The sea surface parameters include the sea height of the center point (SSH), the sea surface temperature of the center point (SST) and the sea surface temperature of 15 points around (17 factors). The subsea parameters include temperature at different sea depths (5, 15, 25, 35, 45, 55, 65, 75, 85, 95 m), currents in the east-west direction (U), currents in the north-south direction (V), and salinity (SSS) at the center point (40 factors).The features are numbered in order, and the model constructed by the above method of selecting features becomes a full feature model. The total number of features is 10125.

This paper uses the attention mechanism and the 2015 full-feature model trained by TCN to get the feature importance ranking. The attention mechanism can obtain the proportion of the impact of the full feature on the predicted value. The sum of all feature influence ratios is 1. In this paper, through experiments, the least important 15-dimensional features are discarded from the monthly data, in order, the vertical height is 400 m, the 35, 65 m underwater temperature, the salinity at the depth of 25, 55, and 95 m, currents in the east-west direction (U) at the depth of 15 and 65 m, currents in the north-south direction (V) at the depth of 5, 15, 25, 95 m and the sea surface temperature of the three points farthest from the center point. The total number of features after selection is 8250.

The training set is divided into two parts: feature and label. The model learns the relationship between features and label. This paper organizes the data in a sliding window. The test set is used to test the effect of the model. The training set required for the prediction model in 2015 (a total of 25 * 2533 pieces of data, also known as the number of samples, m) is the 2533 effective data points of the Indian Ocean from 1980 to 1989 and from January to May in 1990 as features, Sea surface temperature from June to December in 1990 is used as a marker; data from 1981–1990 and January to May in 1991 is used as a feature, and sea surface temperature from June to December in 1991 is used as a marker. The training set finally takes the data from 2004–2013 and January-June in 2014 and the corresponding sea surface temperature from July to December in 2014. The test set is the data from 2005–2014 and from January to June in 2015 and the corresponding sea surface temperature from July to December in 2015, a total of 2533 data.

2.2 Method

This paper uses the method of combining TCN architecture proposed in [12] and attention mechanism. TCN applies a variety of ideas such as residual connection, one-dimensional convolution, dilated convolution and causal convolution, so that the TCN structure has more advantages when dealing with long time series problems. In this paper, the training set is processed into a three-dimensional matrix of (m, timesteps, feature_nums) (full feature model feature_nums = 81, partial feature model feature_nums = 66), m is the number of samples, timesteps = 125 represents how many months we will use, feature_nums is the number of one month of data.

This paper sets the hyper parameters in the TCN model. The size of the convolution kernel is 8 and the number of the convolution kernels is 24, dilations = [1, 2, 4, 8, 16, 32, 64, 128, 256]. The size of the convolution kernel determines the value of each cell in the feature map is related to which areas of the input. When convolving the input matrix, the convolution kernel extracts the information of 8 time steps of the input matrix every time it slides. After experiments, the initial size of the convolution kernel is set to 8. In each residual block, the convolution kernel becomes larger according to the expansion parameter list, and the experimental result is the best. Dong always uses 7-day, 20-day, and 50-day historical sea temperature data as features to predict the future sea surface temperature of 1 day, 7 days, and 30 days. This paper uses the sea temperature data of the past ten years to make annual forecasts, which is more applicable to the research requirements of large and medium-sized ocean physical phenomena. Using dilated convolution can consider longer historical information, and is more suitable for predicting internal changes in the ocean than ordinary convolution. The number of convolution kernels determines the number of feature maps generated by the convolutional layer, and each feature map contains different information. There are too many feature maps, and some accidental changes in the sea temperature data set will be learned by the model. There are too few feature maps, and it is difficult to learn the relationship between features and sea surface temperature. After experiments, the model has the best effect when the number of convolution kernels is set to 24.

The attention mechanism can make the neural network have the ability to focus on a subset of its input (or features). Since we have a large number of input features, in the process of neural network learning, the attention mechanism can learn the relationship between these features and the output by itself, and increase the weight of some features that contribute more to the output to better find the relationship between features and output. We add an attention mechanism before putting the data into the TCN. That is, add a DENSE layer, and use softmax to assign a value between 0–1 to each feature, and then multiply the obtained weight matrix with the original input to obtain new 81-dimensional feature data, and then put it into TCN for learning. In the training process, the neural network will automatically learn the weight matrix, amplify the important features, and reduce the weight of the unimportant features to improve the ability of the model. This paper discusses the importance of features, adds attention mechanism to all-feature 81-dimensional data, and generates feature influence ratio histogram.

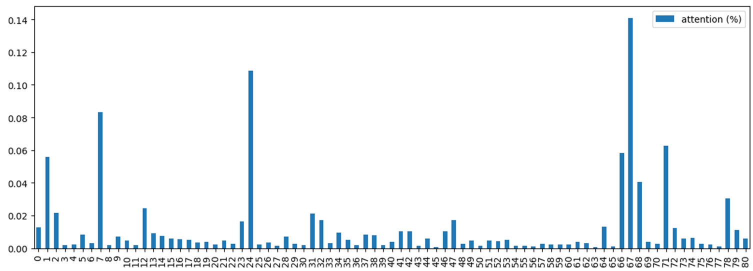

Fig. 1. The importance of whole features is marked by attention mechanism.

3 Result and Analysis

In this paper, a three-dimensional matrix (m, 125, feature_nums) is used to train the model, and 20% of the training set is randomly selected as the verification set, and the model parameters are adjusted to obtain the optimal model. This paper uses one evaluation mechanism-correlation, to compare the prediction accuracy of models trained with different features. The correlation degree represents the similarity between the predicted SSTs time series and the real SSTs time series. The higher the correlation degree, the higher the fit of the model to the true value and the higher the accuracy of the model. The long-term temperature changes in large-scale sea areas are affected by various factors, such as ocean currents, wind fields, and sea water velocity. The sea temperature changes are regular and periodic, and the ability of the model to learn the trend of sea temperature changes is measured by the similarity of the temperature time series.

Table 1 shows the average value of the correlations between all-feature and partial-feature models (8 in total) from 2015 to 2018. By studying the mean value of correlation in the sea area, the full-feature and part-feature models predict that the SSTs sequence fits the observed true value to a higher degree. It can be seen that the correlation degree of some feature models has decreased in 15 and 16 years, but the correlation degree has been improved compared with the full feature model in 17 and 18 years, which proves that the partial feature model proposed in this paper reduces the amount of data, the correlation degree does not decrease. In 2016, the reverse IOD phenomenon in the Indian Ocean showed that the average sea surface temperature in the west (50°E–70°E, 10°S–10°N) decreased, and the east (90°E–110°E, 10°S–0°) in the regional average sea surface temperature has risen. This abnormal sea surface temperature change is related to many factors such as ocean currents and wind fields. The wind field affects the seawater velocity, and the seawater flow affects the temperature field. The decline in correlation of some feature models in 2016 may be related to discarded features. The method of narrowing the data set through the attention mechanism is suitable for other normal years, but the fit for abnormally changed years is not enough.

Through the spatial distribution map of the correlation degree, the difference of the correlation degree in the area can be seen, so as to compare the effect of the full feature and the partial feature model. It can be seen from Fig. 1 that the correlation degree is higher overall from 2015 to 2016, and the correlation degree in 2017 and 2018 between

Table 1. Correlation comparison table of full feature and partial feature models.

Feature_nums	2015	2016	2017	2018
81dims	0.8806	0.9347	0.7717	0.8079
66dims	0.8660	0.8313	0.7856	0.8112

75°E–100°E, 7°S−−2°S, and there is a downward trend. From 2015 to 2016, the results of partial-feature models and full-feature models are close, and the error increase trend is not obvious. Partial- feature model performed better than the full-feature model in 2017, and the performance in 2018 was similar to the full-feature model. In general, the correlation between −10°S−−15°S and 87°E–100°E in the study area is lower than that of the surrounding sea area. This may be due to the complicated changes in sea temperature caused by the movement of regional ocean currents. It will make the model mistaken for noise, thus failing to learn the changes in sea temperature.

Figure 2 shows the comparison of the predicted SSTs time series and the real SSTs time series of two data points (−25°E, 63°S) and (−12°E, 73°S). The upper two figures are the prediction results of full-featured models, the two figures below are the prediction results of partial- feature model. The correlation degree is calculated from the 7-month real SSTs time series and the predicted SSTs time series, and the temperature fitting can be specifically observed through the curves of the real and predicted temperature values over the months. Observe the changes of the real temperature curve and the predicted curve at any data point, and you can measure the specific performance of the model on a single data point. It can be seen from the figure that the prediction results of partial-feature model and the full-feature model are close, and can roughly fit the trend of the true value of the data point. This paper uses the attention mechanism to select features. From the full-feature model to the partial feature, there is less data, but it does not affect the prediction result of the model. It can be seen from the figure that the error of the partial-feature model is basically close to the error of the full-feature model (Fig. 3).

Figure 2 Data points: July-December 2015 real sea surface temperature change curve and predicted sea surface temperature change curve

Through the comparison between the spatial distribution map of the correlation degree and the average correlation degree, the results show that this paper uses the attention mechanism to screen out more important features and reduce the total amount of data, but the correlation degree does not significantly decrease. Long-term SST information has far-reaching significance for some climate changes and the stability of ocean systems. The data required for this kind of research spans a long time and has complex features. Multi-source and multi-modal ocean-atmosphere data covers various factors that affect the temperature change of the ocean, but the data set is complex. It is important to distinguish the importance of features through attention mechanism. The method is less subjectively influenced by the researcher, and the prediction results rely on accurate calculation and objective analysis of the computer.

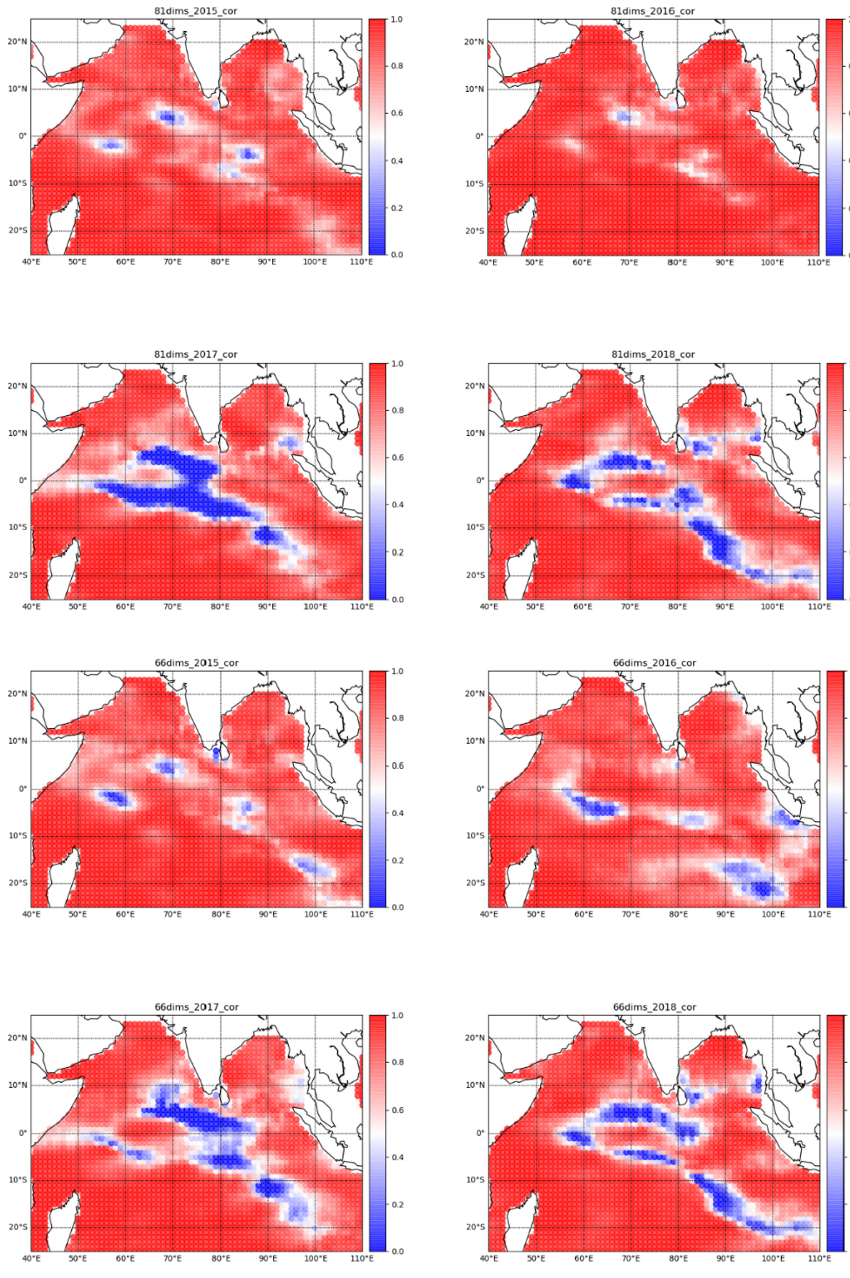

Fig. 2. Correlation distribution of full feature model and partial feature model from 2014 to 2018

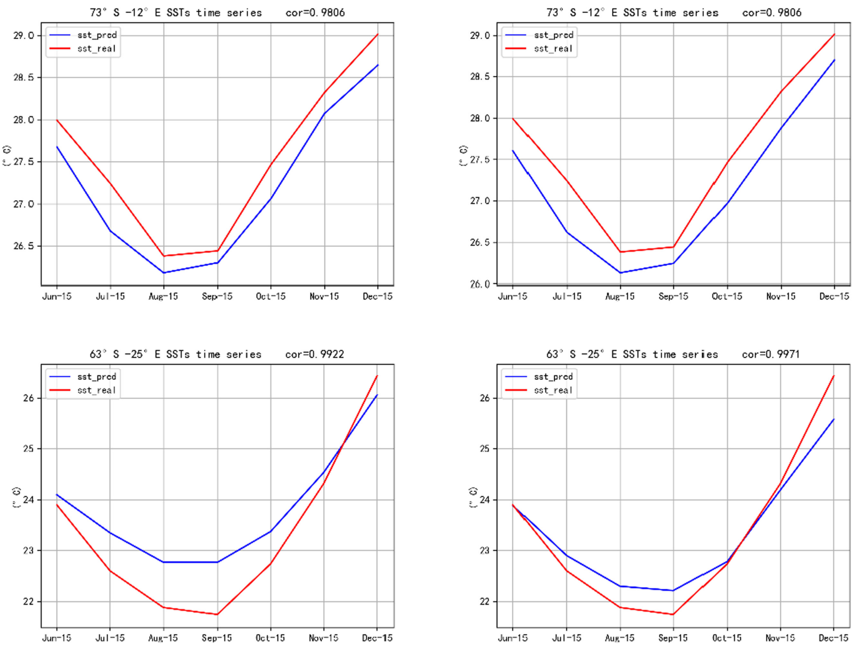

Fig. 3. Real sea surface temperature and forecast sea surface temperature

4 Summary

This thesis uses the attention mechanism to distinguish the importance of features, discards the least important 15 dimensions from all features, and compares the results of long-term sea surface temperature forecasting with all and some feature extraction schemes. The experimental results show that the TCN prediction model with some features maintains a stable high accuracy during the aging period. According to related research on sea surface temperature, the occurrence of large ocean climate phenomena is related to many factors. The full-feature model includes the regional influence of the wind field through the sea surface temperature of surrounding data points, and also considers the heat of ocean currents. The training data of full-feature model includes multi-factor and multi-level data, what need longer training time and higher hardware requirements. The result of sea surface temperature prediction used SSTs alone is not reliable enough, and the error in long-term sea surface temperature prediction is too large. Distinguishing the importance degree through the attention mechanism solves the problem of huge data set, and the accuracy is high. This paper uses the attention mechanism to complete the long-term sea surface temperature forecast of the large ocean basin structure in a partial feature manner. The model effect is stable and can better cope with the occurrence of abnormal sea temperature changes. It is useful for studying large-scale ocean physical phenomena.

This paper discusses the impact of reducing the size of the data set on the experimental results through TCN and attention mechanism. The results show that the data set is smaller and the accuracy is higher. Based on TCN system, historical ocean data were used to predict sea surface temperature 7 months in advance, and the attention mechanism was used to distinguish the importance of all features. Construct a new data set training model with partial features, and predict the sea surface temperature from July to December in 2015–2018. Through experiments, it is found that the experimental results of some feature models are better than the full feature model in 2016 and 2017. This method to reduce data is very effective in dealing with huge and detailed ocean data. This thesis uses attention machines to select features instead of relying on humans to distinguish. It has far-reaching significance for the combination of marine physics and deep learning.

References

1. Alexander, M.A., et al.: Projected sea surface temperatures over the 21st century: changes in the mean, variability and extremes for large marine ecosystem regions of Northern Oceans. Elem. Sci. Anthr. **6**, 9 (2018). https://doi.org/10.1525/elementa.191
2. Liang, C., Xian, W., Pauly, D.: Impacts of ocean warming on China's fisheries catches: an application of "mean temperature of the catch" concept. Front. Mar. Sci. **5**, 5 (2018). https://doi.org/10.3389/fmars.2018.00026
3. Ashok, K., Guan, Z., Saji, N.H., et al.: Individual and combined influences of ENSO and the Indian Ocean Dipole on the Indian summer monsoon. J. Clim. **17**(16), 3141–3155 (2004)
4. Ashok, K., Guan, Z., Yamagata, T.: Impact of the Indian Ocean Dipole on the relationship between the Indian monsoon rainfall and ENSO. Geophys. Res. Lett. **28**(23), 4499–4502 (2001)
5. Behera, S.K., Luo, J., Masson, S., et al.: Impact of the Indian Ocean Dipole on the East African short rains: a CGCM study. CLIVAR Exch. **27**, 43–45 (2003)
6. Chan, S.C., Behera, S.K., Yamagata, T.: Indian Ocean Dipole influence on South American rainfall. Geophys. Res. Lett. **35**(14), L14S12 (2008)
7. Yang, Y., Dong, J., Sun, X., Lima, E., Mu, Q., Wang, X.: A CFCC-LSTM model for sea surface temperature prediction. IEEE Geosci. Remote Sens. Lett. **15**(2), 207–211 (2018). https://doi.org/10.1109/LGRS.2017.2780843
8. Zhang, Q., Wang, H., Dong, J., Zhong, G., Sun, X.: Prediction of sea surface temperature using long short-term memory. IEEE Geosci. Remote Sens. Lett. **14**(10), 1745–1749 (2017). https://doi.org/10.1109/LGRS.2017.2733548
9. Yang, Y., Dong, J., Sun, X., Lima, E., Mu, Q., Wang, X.: A CFCC-LSTM model for sea surface temperature prediction. IEEE Geosci. Remote Sens. Lett. **15**, 207–211 (2018). https://doi.org/10.1109/lgrs.2017.2780843
10. Li, W., Guan, L., Qu, L., Guo, D.: Prediction of sea surface temperature in the china seas based on long short-term memory neural networks. Remote Sens. **12**(17), 2697 (2020). https://doi.org/10.3390/rs12172697
11. Wu, X., Yan, X.-H.: Estimation of subsurface temperature anomaly in the north Atlantic using a self-organizing map neural network. J. Atmos. Ocean. Technol. **29**(11), 1675–1688 (2012). https://doi.org/10.1175/jtech-d-12-00013.1
12. Bai, S., Kolter, J.Z., Koltun, V.: An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. ArXiv e-prints, April 2018