

## Original research articles

## ResGraphNet: GraphSAGE with embedded residual module for prediction of global monthly mean temperature

Ziwei Chen<sup>a,b</sup>, Zhiguo Wang<sup>a,\*</sup>, Yang Yang<sup>b</sup>, Jinghui Gao<sup>b</sup><sup>a</sup> School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, 710049, Shaanxi, PR China<sup>b</sup> School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an, 710049, Shaanxi, PR China

## ARTICLE INFO

## Keywords:

Graph neural network

GraphSAGE

ResNet

Global temperature prediction

## ABSTRACT

Data-driven prediction of time series is significant in many scientific research fields such as global climate change and weather forecast. For global monthly mean temperature series, considering the strong potential of deep neural network for extracting data features, this paper proposes a data-driven model, ResGraphNet, which improves the prediction accuracy of time series by an embedded residual module in GraphSAGE layers. The experimental results of a global mean temperature dataset, HadCRUT5, show that compared with 11 traditional prediction technologies, the proposed ResGraphNet obtains the best accuracy. The error indicator predicted by the proposed ResGraphNet is smaller than that of the other 11 prediction models. Furthermore, the performance on seven temperature datasets shows the excellent generalization of the ResGraphNet. Finally, based on our proposed ResGraphNet, the predicted 2022 annual anomaly of global temperature is 0.74722 °C, which provides confidence for limiting warming to 1.5 °C above pre-industrial levels.

## 1. Introduction

The Global monthly mean temperature is a key climate indicator, which is usually expressed as a temperature “anomaly” that is just the difference from the average over a fixed period. Since 1850, as a typical time series, temperature series has been measured at weather stations, by ships and buoys and by satellites. The Paris Agreement is a legally binding international treaty on climate change, and its goal is to limit global warming to well around 2 °C, preferable to 1.5 °C, compared to pre-industrial levels. Thus, as a typical prediction problem of time series, it is significant to predict global mean temperature, which demonstrates the megatrend of global warming and climate changing. In the past few decades, prevalent climate models of the Earth have been proposed and worked well for weather and climate forecasting (Gordon et al., 2000; Stott and Kettleborough, 2002; Smith et al., 2007). However, not all significant physical and chemical processes can be explicitly resolved by these deterministic models (Bauer et al., 2015). Instead, in this study, the data-driven models (Pathak et al., 2018; Arcomano et al., 2020; Taylor et al., 2022) are focused on the near-term climate prediction (Kushnir et al., 2019) using time series of the global monthly mean temperature.

Nowadays, there are myriad data-driven methods to predict time series including global mean temperature series, such as the autoregressive integrated moving average (ARIMA) (Contreras et al., 2003), the seasonal autoregressive integrated moving average with exogenous

(SARIMAX) (Guin, 2006), the random forest (Breiman, 2001), the support vector regression (Smola and Schölkopf, 2004), the recurrent neural networks (RNN) (Hopfield, 1982), temporal convolutional network (Li et al., 2021), etc. In particular, as a deep learning model, the RNN and its variants, such as long-short term memory (LSTM) (Hochreiter and Schmidhuber, 1997) and gated recurrent unit (GRU) (Chung et al., 2014), have been widely applied in time series prediction in the Euclidean domain. However, different from many other time series, the global temperature generally shows a rising trend due to increasing accumulation of CO<sub>2</sub> in the atmosphere as a result of fossil fuel consumption, which leads to some bottlenecks in the performance of traditional data-driven methods. For example, because the global mean temperature series show rich relational structure in the long sequences, the inherent architecture of the RNN limits its representation ability with slow and complex training procedures. To improve the accuracy and efficiency of prediction, the graph neural network (GNN) (Scarselli et al., 2009) is an emerging approach to expand the representation of artificial neural network with the help of relevant concepts of graph theory in the non-Euclidean domain.

The current GNN models are mainly categorized as recurrent GNNs, convolutional GNNs, graph autoencoders, and spatial-temporal GNNs (Wu et al., 2021). As a pioneering work, Kipf and Welling proposed a simplified graph neural network model, graph convolution network

\* Corresponding author.

E-mail address: [emailwzg@mail.xjtu.edu.cn](mailto:emailwzg@mail.xjtu.edu.cn) (Z. Wang).<https://doi.org/10.1016/j.aiig.2022.11.001>

Received 28 October 2022; Received in revised form 26 November 2022; Accepted 26 November 2022

Available online 8 December 2022

2666-5441/© 2022 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

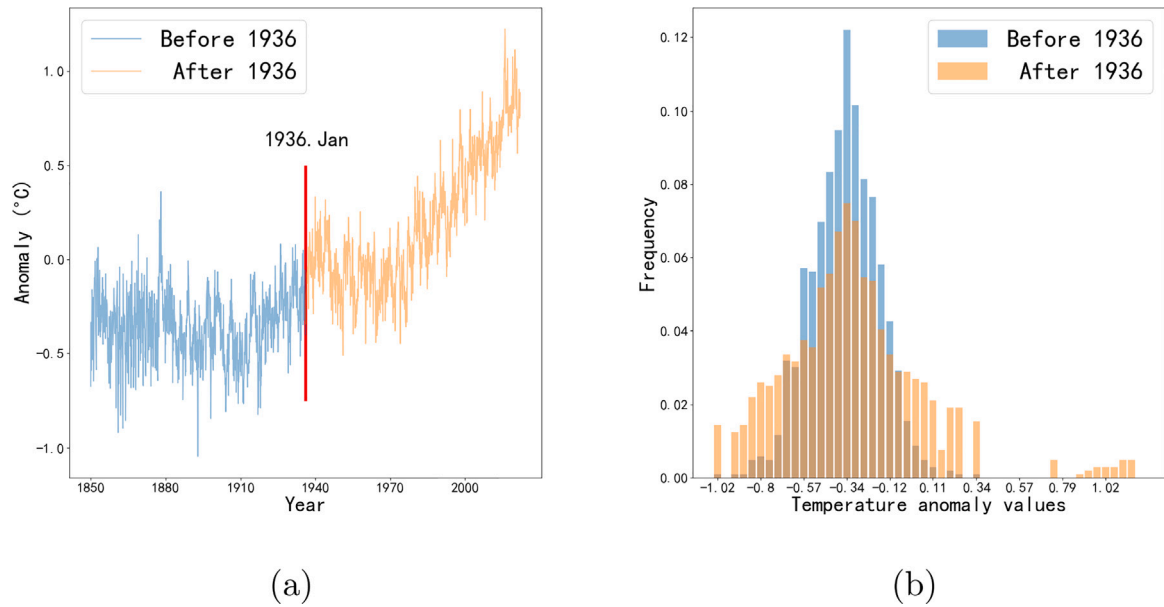


Fig. 1. (a) The global surface temperature dataset, HadCRUT5, includes a global mean temperature series from 1850 to 2022 in every month, and (b) its temperature anomalies distribution.

(GCN), that operates directly on a graph and induces embedding vectors of nodes based on the properties of their neighborhoods (Kipf and Welling, 2016). The major limitation of GCN is inherently transductive, GraphSAGE is an inductive framework that can quickly generate embeddings and predict for unseen information (Hamilton et al., 2017). Whereas, due to over-smoothing problem, it is difficult to improve the predictive performance with stacked up many GraphSAGE layers. Meanwhile, as a classical deep convolution neural network (CNN) with strong feature extraction, the residual neural network (ResNet) helps in tackling the vanishing gradient problem using identity mappings (He et al., 2016). Thus, the embedded ResNet module also can act as a buffer layer to reduce the over-smoothing problem of the GraphSAGE layers (Oono and Suzuki, 2020).

Therefore, we propose a specific deep neural network, called ResGraphNet, for the prediction of the global monthly mean temperature series. The ResGraphNet is an end-to-end architecture that consists of two GraphSAGE layers, an embedded ResNet layer and a full-connected (FC) layer. Meanwhile, the path graph is selected to represent the temperature time series, as the underlying topology of GraphSAGE. The subsequent experimental results show that compared with simply using the GraphSAGE or the ResNet as the network layers, our proposed ResGraphNet has a stronger accuracy and a generalization ability in predicting time series, especially monthly mean temperature series. Our source code is available at <https://doi.org/10.5281/zenodo.7213337>.

Compared with the previous models which has been commonly used in prediction of time series, our contributions of the proposed ResGraphNet mainly includes the following points:

1. To the best of our limited knowledge, this is the first study to model a temperature series as a path graph and learn the characteristics of the global temperature between the connection weights of each node in the GNN with embedded ResNet.
2. Results on several temperature time series datasets demonstrate that our method outperforms state-of-the-art temperature series methods, with possessing a shorter training time and faster convergence speed.
3. Our proposed ResGraphNet predicts that the 2022 annual anomaly of global temperature is 0.74722 °C, which provides a confidence for limiting warming to 1.5 °C above pre-industrial levels.

The rest of this paper is organized as follows. The Chapter II briefly introduces the basic concepts of the ResNet and the GraphSAGE, and then elaborates the architecture of the proposed ResGraphNet. The Chapter III shows experiments of datasets that reveal the performance of the proposed ResGraphNet. Finally, conclusions are drawn in the Chapter IV.

## 2. Data and method

### 2.1. Datasets of temperature time series

For understanding and predicting global surface temperature changes, a public time series *HadCRUT5* (Morice et al., 2021), the global surface temperature dataset from 1850.Jan to 2022.Mar in every month is analyzed mainly in this study. As shown in Fig. 1, since 1936, because of increasing accumulation of CO<sub>2</sub> in the atmosphere as a result of huge fossil energy consumption, the global temperature is increasing gradually (Arrow, 2007). This is a challenge to those methods that use past data to predict the future information (Dougherty et al., 2005; Hastings et al., 2014; Terando et al., 2014), because it is difficult to learn the drastic changes in temperature data in the last few years. Therefore, to tackle this problem, we build the novelty architecture of ResGraphNet.

In additional, ERSSTv4 (Smith et al., 2008), ERSSTv3b (Smith et al., 2008), Berkeley-Earth (Rohde et al., 2013), HadSST3 (Kennedy et al., 2011), and ERA5 (Hersbach et al., 2020) also are temperature data measured in different ways and in different areas. These datasets also are applied to verify the generalization ability of the proposed ResGraphNet.

### 2.2. Residual neural network

ResNet is a type of neural networks that applies identity mapping. It is known that simply stacking up CNN layers cannot indefinitely improve the performance of the CNN model, due to the vanishing gradient problem (Glorot and Bengio, 2010). To solve this problem effectively, the concept of a highway network is proposed as (Srivastava et al., 2015)

$$y = F(x, w_i) \cdot \tau(x, w_i) + x \cdot C(x, w_c) \quad (1)$$

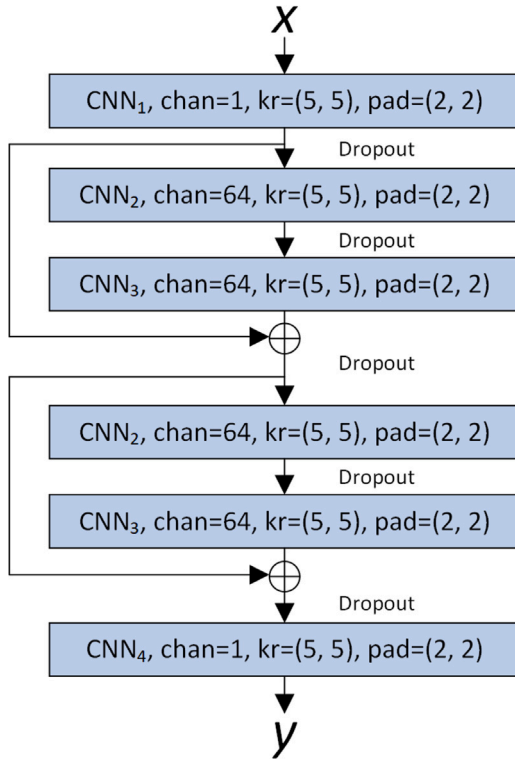


Fig. 2. The architecture of a ResNet module that is applied in this study.

where  $x$  and  $y$  are the input and output of the layers aimed, and the  $\tau(x, w_i)$  and  $C(x, w_c)$  are the transform gate and the carry gate, respectively. For simplifying parameters and reducing the risk of over fitting, the ResNet set these two gates as the identical mapping (i.e.,  $\tau(x, w_i) = 1$ ,  $C(x, w_c) = 1$ ) (He et al., 2016), then the forward propagation formula is

$$y = F(x, w_i) + x \quad (2)$$

where the function  $F(x, w_i)$  represents the residual mapping to be learned. As shown in Fig. 2, in this study, the ResNet module is mainly composed of 4 CNN layers, which can be expressed as

$$\begin{cases} h_1 = CNN_1(x) \\ h_2 = Dropout(h_1) \\ h_3 = CNN_2(h_2) \\ h_4 = Dropout(h_3) \\ h_5 = CNN_3(h_4) \\ h_6 = h_5 + h_1 \\ y = CNN_4(h_6) \end{cases} \quad (3)$$

where  $x$  and  $y$  is the input and output of our ResNet module (Fig. 2) and *Dropout* is the dropout layer. In the following content, the *ResNet* also denotes the forward propagation process in Eq. (3).

### 2.3. GraphSAGE neural network

In this study, a graph is defined as  $G = (V, E)$ , where  $V = \{v_0, v_1, \dots, v_{N-1}\}$  denotes the nodes and  $E$  denotes the edges.  $A \in \mathbb{R}^{N \times N}$  is the adjacency matrix. If there is an edge from  $v_i$  to  $v_j$ , that is  $\exists (v_i, v_j) \in E$ , then  $A_{ij} = 1$  otherwise  $A_{ij} = 0$ . And then the degree matrix  $D$  is defined as  $D_{ij} = \sum_{j=0}^{N-1} A_{ij}$ . Thus, the Laplacian matrix  $L$  is defined as  $L = D - A$ . In addition, the normalized form of Laplacian matrix  $\mathcal{L} = D^{-1/2} L D^{-1/2}$  is often used in practical applications. If data

information of node  $v_i$  is  $x_i$ , then GraphSAGE updates node information by sampling and aggregating adjacent points of each node  $v_i$ , that is:

$$x'_i = W_1 x_i + W_2 \cdot \text{mean}_{j \in N(j)} x_j \quad (4)$$

where  $N(j)$  is the set of adjacent points of a node  $v_i$ , which can be obtained by the adjacency matrix  $A$  of a graph  $G$ ;  $W_1$  and  $W_2$  are the learnable network weights; *mean* demonstrates that the aggregation method adopts the mean aggregation. Fig. 3 shows the running process of the GraphSAGE. For the random sampling process of nodes, the number of the first-order sampling nodes is 1 and the number of the second-order sampling nodes is 1. As shown in Fig. 3(b), the node aggregation takes the mean value after weighted summation of the information of adjacent points. Then, we update the information of the current central node, as shown in Fig. 3(c).

### 2.4. Proposed ResGraphNet

To pay attention to rich correlation in the temperature time dimension, we propose that the *ResNet* module is embedded in two GraphSAGE layers. Assuming that the input and the output of the network are  $x$  and  $y$ , its forward propagation can be expressed as:

$$\begin{cases} f_1 = \text{GraphSAGE}_1(x, A) \\ f_2 = \text{ResNet}(\text{Dropout}(f_1)) \\ \hat{y} = \text{GraphSAGE}_2(\text{Dropout}(f_2), A) \end{cases} \quad (5)$$

where  $\text{GraphSAGE}_i$  is the  $i$ th GraphSAGE layer,  $i = 1, 2$ .  $A$  is the adjacency matrix of the graph.

Here, notice that the GraphSAGE layers (including most other GNN layers) need a given topology structure of a graph in advance (i.e. the adjacency matrix  $A$ ). For a specific time series, it is an open topic to build an available graph (Lacasa et al., 2008, 2009; Wang et al., 2019; Chen et al., 2022). In this study, to preserve the inherent correlation and causality of time series, we consider adjacency matrix  $A_{\text{path}}$  of the path graph which satisfies

$$A_{\text{path}} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix} \quad (6)$$

For the path graph, each node  $v_i$  has an edge  $e_{i,i+1}$  pointing to the next node  $v_{i+1}$ , and the time series prediction task studied in this paper is to utilize the past data to predict future data. Based on this idea, a node  $v_t$  can be constructed for each time  $t$ , set  $x = [x_0, x_1, \dots, x_{L-1}]^T$ ,  $x_t \in \mathbb{R}^{M \times 1}$ ,  $t = 0, 1, \dots, L-1$ , then  $x_t$  is the value of node  $v_t$ ,  $A_{\text{path}} \in \mathbb{R}^{L \times L}$ . So the role of  $A_{\text{path}}$  is to make the value  $x_t$  of time series  $x$  at time  $t$  (the value of node  $v_t$ ) be related to the value  $x_{t+1}$  of next time  $t+1$  (the value of node  $v_{t+1}$ ).

Consequently, to further alleviate the gradient disappearance problem, a FC layer is applied to mitigate the instability of results due to the residual connection. From Fig. 4, the proposed ResGraphNet can be written as

$$\begin{cases} f_1 = \text{GraphSAGE}_1(x, A) \\ f_2 = \text{ResNet}(\text{Dropout}(f_1)) \\ f_3 = FC(f_1 + f_2) \\ \hat{y} = \text{GraphSAGE}_2(\text{Dropout}(f_3), A) \end{cases} \quad (7)$$

Ideally, for the proposed ResGraphNet, the main features of time series will be learned by the shallow layers (*GraphSAGE*), and more detailed features will be extracted by the deeper module (*ResNet*). To verify performance of the ResGraphNet, the experimental results are shown in Chapter 3.

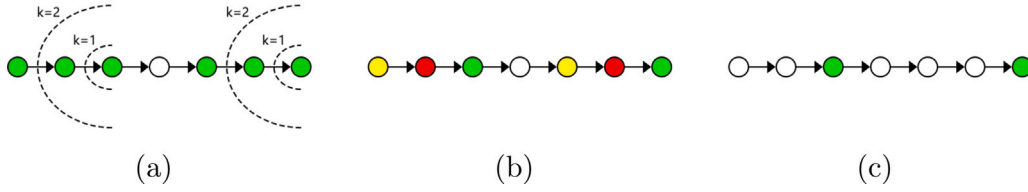


Fig. 3. The operation process of the GraphSAGE, involves (a) the sampling process, (b) the aggregation process and (c) the updating nodes.

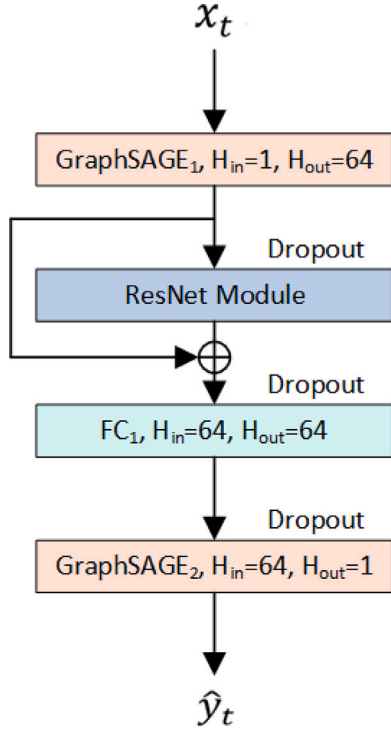


Fig. 4. Architecture of the proposed ResGraphNet is defined in Eq. (7), where the ResNet module is defined in Eq. (3).

### 3. Experiment and result

This chapter first shows the training dataset and the test dataset based on real time series data. Then, we build models based on different algorithms and set the parameters of these different models. Finally, the training dataset are applied to train each model, while the test dataset are used to evaluate the performance of models.

#### 3.1. Dataset samples

Now we build training datasets and test datasets based on *Had-CRUT5*. At the beginning, if we set the length of time series as  $l$ , then it can be denoted as  $D = [d_0, d_1, \dots, d_{l-1}]$ . The first 50% of the data are selected as the training dataset.  $D^{train}$  and the last 50% are selected as the test dataset  $D^{test}$ , whose lengths are  $l_{train}$  and  $l_{test}$  respectively. The target task of this paper is to predict time series. We hope to use the data before time  $t$  to predict the data at time  $t$  and the error between the prediction result and the real data will be as small as possible. To achieve this purpose, for each time  $t$ , we set the input data  $x_t = [d_{t-l_x}, d_{t-l_x+1}, \dots, d_{t-1}]$  and the input label  $y_t = [d_t]$ , and each pair of  $x_t$  and  $y_t$  constitute a sample  $s_t$ , then  $s_t = \{x_t, y_t\}$ , where  $l_x$  is the length of input data  $x$ . In other words, we use  $l_x$  data before time  $t$  to predict the data at time  $t$ . According to the above steps, the training dataset  $S^{train} = [s_0^{train}, s_1^{train}, \dots, s_{m^{train}-1}^{train}]$  can be constructed for  $D^{train}$  and the

test dataset  $S^{test} = [s_0^{test}, s_1^{test}, \dots, s_{m^{test}-1}^{test}]$  can be constructed for  $D^{test}$ , where  $m^{train}$  and  $m^{test}$  are the number of samples of  $D^{train}$  and  $D^{test}$ , respectively. Meanwhile,  $S^{train}$  and  $S^{test}$  can be divided into training data  $x^{train}$ , training labels  $y^{train}$ , test data  $x^{test}$  and test labels  $y^{test}$ . The construction process of samples is shown in Fig. 5.

#### 3.2. Parameters and indicator

For any sample  $s_t$ ,  $x_t$  is the input of the model and  $\hat{y}_t$  denotes the corresponding output. The loss function of the proposed model is the mean square error and the network weights are updated by the negative direction of the gradient decline of the loss function.

For regression problems, the Root Mean Square Error (RMSE) is generally selected as the performance evaluation indicator (Bellocchio et al., 2012). However, the RMSE is easily affected by the data itself. If there are great differences in the value distribution of different time series, the magnitude of their RMSE may be completely different, which makes it impossible to compare the testing results of these time series. Therefore, we further apply the determination coefficient (also named as  $R^2$  score) as our indicator. For any model, we denote the input data and predicted results as  $x^{train}$  and  $\hat{y}^{train}$  respectively, then we set  $x^{test}$  and  $\hat{y}^{test}$  for test set. Their  $R^2$  score is defined as  $r^{train}$ ,  $r^{test}$ :

$$\begin{cases} r^{train} = \frac{\sum_i (y_i^{train} - \bar{y}^{train})^2}{\sum_i (y_i^{train} - \bar{y}^{train})^2} \\ r^{test} = \frac{\sum_i (y_i^{test} - \bar{y}^{test})^2}{\sum_i (y_i^{test} - \bar{y}^{test})^2} \end{cases} \quad (8)$$

where  $\bar{y}^{train} = \frac{\sum y_i^{train}}{m^{train}}$ ,  $\bar{y}^{test} = \frac{\sum y_i^{test}}{m^{test}}$  are mean values of the  $y^{train}$  and the  $y^{test}$ , respectively. If  $r^{test}$  of one model is closer to 1, this model has a stronger time series prediction ability.

#### 3.3. Performances on different models

Now we train all the models based on  $S^{train}$ , and use  $S^{test}$  for performance testing where we set  $l_x = 60$  and  $l_y = 1$ . In Table 1, there are 12 models are applied, including the proposed ResGraphNet, the ResNet, the GraphSAGE, the GNN with Unified Message Passing Model (UniMP) (Shi et al., 2020), the GCN (Kipf and Welling, 2016), the Graph Isomorphic Network (GIN) (Morris et al., 2019), the RNN with LSTM, the RNN with GRU, the Random Forest Regression (Forest), the Linear Regression (Linear), the ARIMA and the SARIMAX. We calculate  $R^2$  score of each model, as shown in Table 1. Using climate spirals, we draw the true data and predicted results of the  $S^{train}$ , and the true data and predicted results of the  $S^{test}$  of 12 models, as shown in Fig. 6. From Table 1, the  $r^{test}$  of the proposed ResGraphNet is 0.9980, which is the best in all of 12 models. As shown in Fig. 6, the predicted climate spirals of the proposed ResGraphNet is closest to the true data, but the prediction of the Forest is the worst with the  $r^{test} = 0.6397$ .

Moreover, because the time series is too long, it is not intuitive to directly observe the curve between the prediction results and the true data. Hence, we define the error index  $e = [e_0, e_1, \dots, e_{m^{test}-1}]$  as

$$e = y^{test} - \hat{y}^{test} \quad (9)$$

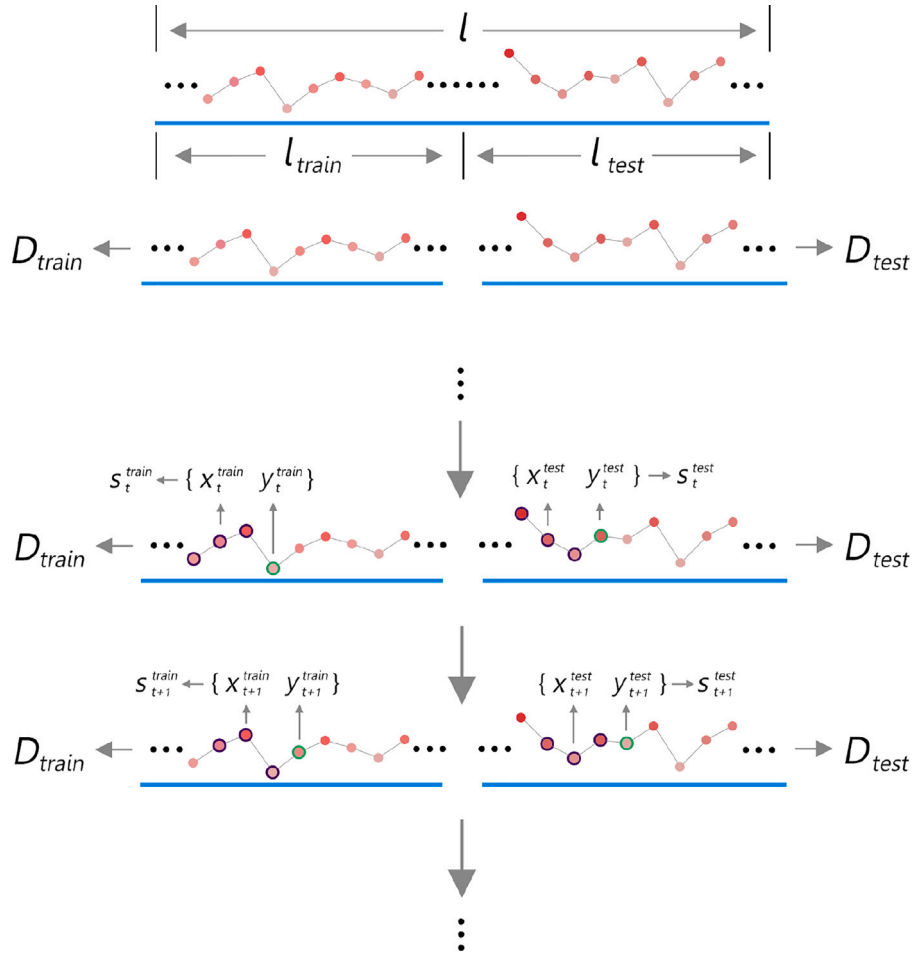


Fig. 5. Samples construction process of the time series (When  $l_x = 3$  and  $l_y = 1$ ).

Table 1

$R^2$  scores calculated by different models.

Model	ResGraphNet	ResNet	GraphSAGE	UniMP	GCN	GIN
$r_{train}^{train}$	0.9975	0.9138	0.8597	0.8430	0.8955	0.8621
$r_{test}^{train}$	0.9980	0.9039	0.8656	0.8440	0.8693	0.8664
Model	LSTM	GRU	Forest	Linear	ARIMA	SARIMAX
$r_{train}^{train}$	0.8871	0.8225	0.9464	0.9037	0.8782	0.8329
$r_{test}^{train}$	0.8030	0.8249	0.6397	0.9010	0.8214	0.8124

Table 2

Training time of different models.

Model	Time[s]
ResGraphNet	78.78
ResNet	123.57
GraphSAGE	11.87
LSTM	149.08
GRU	126.70
Forest	1.17
Linear	0.04

Then we can count and plot the frequency distribution of all elements  $e_i$  in the  $e$  as shown in Fig. 7. Most of the errors  $e_i$  computed by the ResGraphNet are distributed within the interval  $[-0.03, 0.06]$ , but errors  $e_i$  of other models are outside this interval. In other words, the error  $e_i$  computed by the ResGraphNet is smaller than that of other models.

Furthermore, to compare the computation time, we record the time spent on training different models, as shown in Table 2. From Table 2 it turns out that ResGraphNet can indeed converge faster than ResNet

Table 3

$r_{test}^{test}$  in different  $l_y$  and different models.

$l_y$	ResGraphNet	GraphSAGE	Linear	GRU
2	0.9966	0.8735	0.8513	0.8241
5	0.9891	0.8324	0.8273	0.7995
12	0.9858	0.7988	0.7768	0.7690

and RNN with LSTM and GRU. Due to the addition of *ResNet* module, the training time of ResGraphNet is longer than that of GraphSAGE. Because the traditional machine learning including Forest and Linear adopts a simpler calculation process than the Neural Network, the running time of them is also quite shorter. In this Scenario, the proposed ResGraphNet makes a certain and acceptable sacrifice for the computation time, in exchange for the highest prediction accuracy.

In the previous analysis, we set  $l_x = 60$  and  $l_y = 1$ . Actually, we hope that ResGraphNet can still maintain its good performance, when the length of predicted time series is becoming longer. The results  $r_{test}^{test}$  based on different  $l_y$  and different models are shown in Table 3.  $r_{test}^{test}$  decreases with the increase of  $l_y$ , but ResGraphNet outperforms other models. In other words, the length of time series, that can be predicted by the ResGraphNet, is limited in a way. When the temperature change of 12 months (i.e.  $l_y = 12$ ) in 2021 year is expected to be predicted, the indicator  $r_{test}^{test}$  between  $y^{test}$  and  $\hat{y}^{test}$  can remain above 0.98. Therefore, for the prediction of future temperature, the length of predicted time series  $l_y = 12$  is available.



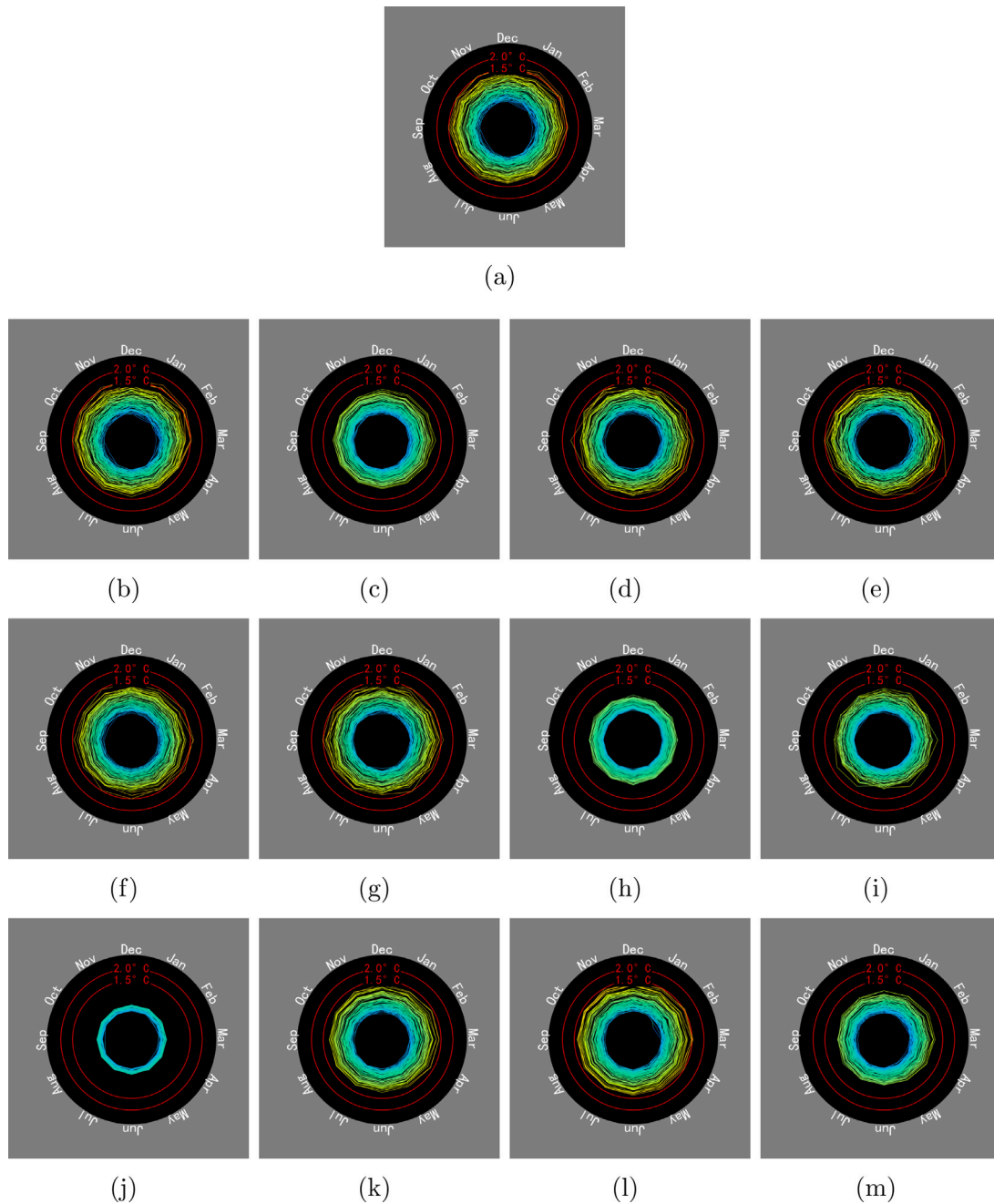


Fig. 6. Climate spirals of (a) true data and predicted results by (b) ResGraphNet, (c) ResNet, (d) GraphSAGE, (e) UniMP, (f) GCN (g) GIN, (h) LSTM, (i) GRU, (j) Forest, (k) Linear, (l) ARIMA, (m) SARIMAX models in  $S^{test}$ .

### 3.4. Prediction of future temperature

To show the prediction performance of the ResGraphNet for unknown future temperature, we firstly list the known true data and predicted results of 2021 in Table 4. As shown in Table 4, for HadCRUT5 dataset, the predicted result of annual mean temperature (0.76863 °C) is close to the measured value (0.76185 °C). Similarly, the temperatures in northern hemisphere HadCRUT5-N and southern hemisphere HadCRUT5-S are also tested.

Next, Table 5 and Fig. 8 show the results in 2022 year. The corresponding *mse* between the predicted results and measured data in 2022 January, February and March are 0.0015, 0.0109, 0.0048 respectively, as well as the high similarity of annual temperature. Thus, the results of prediction show the proposed ResGraphNet can accurately predict the unknown temperature data in the future to some extent. Note that,

based on our proposed model, the predicted 2022 annual anomaly of global temperature is 0.74722 °C, which is smaller than that of 2021. The predicted results of 2022 provide more confident to limit global warming to well below 2 °C, preferably to 1.5 °C above pre-industrial levels (red circle in Fig. 8).

### 3.5. Other datasets

Finally, to verify the generalization ability of the proposed ResGraphNet, we test the prediction performance of models based on other different public datasets, including HadCRUT5-N, HadCRUT5-S, ERSSTv4, ERSSTv3b, Berkeley-Earth, HadSST3 and ERA5. Meanwhile, we additionally consider some other time-series prediction models, i.e. the GraphSAGE, the Linear, the GRU, the ARIMA and the SARIMAX,

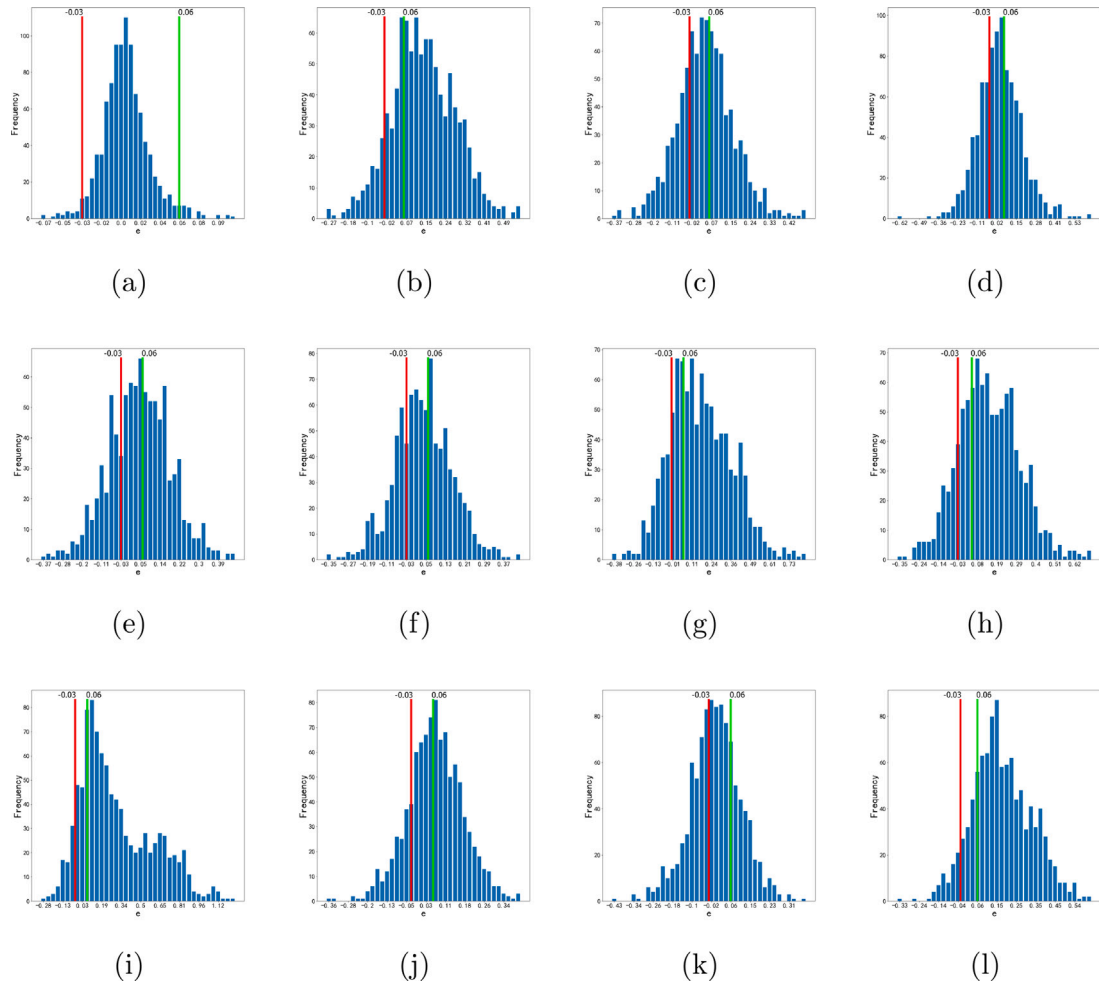


Fig. 7. The distributions of error  $e_i$  based on (a) ResGraphNet, (b) ResNet, (c) GraphSAGE, (d) UniMP, (e) GCN (f) GIN, (g) LSTM, (h) GRU, (i) Forest, (j) Linear, (k) ARIMA, (l) SARIMAX models, where marked red and green bars represent different  $e_i$  intervals.

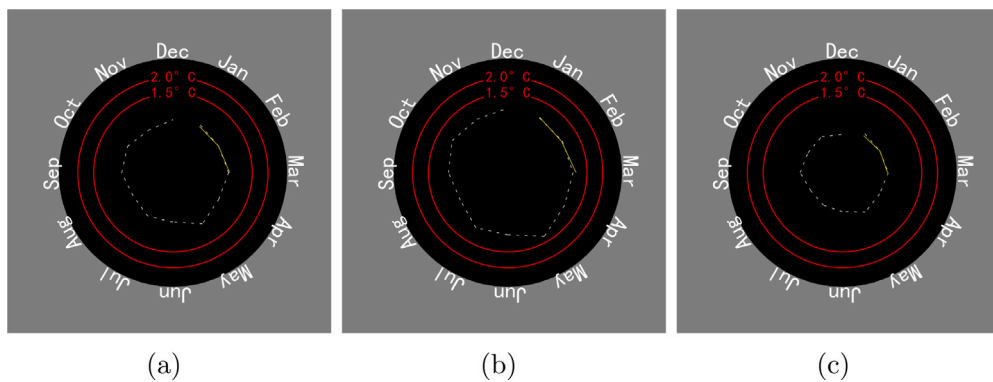


Fig. 8. The temperature of true data and predicted (a) HadCRUT5, (b) HadCRUT5-N, and (c) HadCRUT5-S results in 2022, where the white-dashed-lines are “Predict” and the yellow-solid-lines are “Known”. Annual predicted results provide more confident to limit warming to below 1.5 °C (red circle).

as the comparison methods. Table 6 reveals that the performance of the proposed ResGraphNet is the best among six prediction models.

When applied to seven different datasets, the  $r^{test}$  calculated by using the proposed ResGraphNet is still closer to 1 than that of other models. Therefore, the proposed ResGraphNet has a strong generalization ability in temperature prediction, and has potential to predict temperature data in many scenarios.

#### 4. Conclusion

This paper proposes a specific ResGraphNet that consists of two GraphSAGE layers, an embedded ResNet layer and a FC layer, which is an end-to-end architecture for time series prediction of the global monthly mean temperature. Based on some global mean temperature datasets, the performance of the proposed ResGraphNet is evaluated.

**Table 4**

The true and predicted temperature values in 2021.

Month	HadCRUT5		HadCRUT5-N		HadCRUT5-S	
	Predict	Known	Predict	Known	Predict	Known
January	0.6845	0.7008	1.0578	1.1269	0.3087	0.2747
February	0.6703	0.5645	1.2658	0.8643	0.2965	0.2647
March	0.8831	0.7262	1.2226	1.0232	0.4923	0.4292
April	0.7804	0.7601	1.1715	1.1503	0.4793	0.3698
May	0.7806	0.7064	1.1965	1.0403	0.4164	0.3724
June	0.7198	0.7132	1.1938	1.1113	0.3479	0.3151
July	0.7335	0.7916	1.0732	1.0753	0.5140	0.5079
August	0.7894	0.7994	1.0810	1.0170	0.5751	0.5818
September	0.7658	0.8674	1.0130	1.0672	0.6145	0.6676
October	0.9283	0.9073	0.9592	1.2139	0.5632	0.6008
November	0.7504	0.8536	1.1245	1.2119	0.4870	0.4952
December	0.7368	0.7513	1.0862	1.0331	0.4197	0.4694
Annual	0.7686	0.7618	1.1204	1.0779	0.4595	0.4457

**Table 5**

The true and predicted temperature values in 2022.

Month	HadCRUT5		HadCRUT5-N		HadCRUT5-S	
	Predict	Known	Predict	Known	Predict	Known
January	0.8300	0.7788	1.1400	1.1154	0.5140	0.4423
February	0.7628	0.7634	1.1290	1.0638	0.4636	0.4631
March	0.8569	0.8887	1.1028	1.2554	0.5425	0.5219
April	0.7702		1.1328		0.4518	
May	0.9628		1.5161		0.5386	
June	0.6580		1.2281		0.3322	
July	0.6843		1.1572		0.2894	
August	0.5555		0.9272		0.2483	
September	0.7179		1.0627		0.4310	
October	0.7553		1.1633		0.3561	
November	0.6440		1.0605		0.4123	
December	0.7684		1.0955		0.3096	
Annual	0.7472	0.8103	1.1429	1.1449	0.4236	0.4758

**Table 6**For seven datasets, the  $r^{test}$  scores are calculated by different models.

Dataset	ResGraphNet	GraphSAGE	Linear	GRU	ARIMA	SARIMAX
HadCRUT5-N	0.9589	0.8273	0.8591	0.8214	0.8533	0.8244
HadCRUT5-S	0.9569	0.8019	0.8245	0.8123	0.8391	0.8182
ERSSTv4	0.9492	0.8135	0.8221	0.8156	0.8380	0.8364
ERSSTv3b	0.9389	0.7883	0.8279	0.7915	0.8135	0.8124
Berkeley-Earth	0.9624	0.8609	0.8943	0.8538	0.8986	0.8809
HadSST3	0.9923	0.9658	0.9754	0.9454	0.9228	0.8681
ERA5	0.9162	0.8112	0.7927	0.8018	0.7756	0.7540

The results of the comparison methods show that the proposed Res-GraphNet has the best accuracy and generalization to predict monthly anomaly of temperature series. In addition, the predicted 2022 annual anomaly of global mean temperature is 0.74722 °C, which helps us understand the trend of global warming of 1.5 °C above pre-industrial levels.

However, there are still two problems that have not been well solved in this paper. First, there is still no better method to choose the best graph structure for specific data such as temperature time series. Second, this paper does not solve the problem of long-term prediction of temperature series. Therefore, We hope to further improve the performance of the current model.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Funding

This work was supported by the National Natural Science Foundation of China under Grant 41974137.

### References

- Arcomano, T., Szunyogh, I., Pathak, J., Wikner, A., Hunt, B.R., Ott, E., 2020. A machine learning-based global atmospheric forecast model. *Geophys. Res. Lett.* 47 (9), e2020GL087776.
- Arrow, K.J., 2007. Global climate change: A challenge to policy. *Econ. Voice* 4 (3), <http://dx.doi.org/10.2202/1553-3832.1270>.
- Bauer, P., Thorpe, A., Brunet, G., 2015. The quiet revolution of numerical weather prediction. *Nature* 525 (7567), 47–55.
- Bellocchio, F., Ferrari, S., Piuri, V., Borghese, N.A., 2012. Hierarchical approach for multiscale support vector regression. *IEEE Trans. Neural Netw. Learn. Syst.* 23 (9), 1448–1460.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32.
- Chen, Z., Wang, Z., Wu, S., Wang, Y., Gao, J., 2022. MagInfoNet: Magnitude estimation using seismic information augmentation and graph transformer. *Earth Space Sci.* 9 (12), <http://dx.doi.org/10.1029/2022EA002580>, e2022EA002580.
- Chung, J., Gulcehre, C., Cho, K., Bengio, Y., 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.
- Contreras, J., Espinola, R., Nogales, F., Conejo, A., 2003. ARIMA models to predict next-day electricity prices. *IEEE Trans. Power Syst.* 18 (3), 1014–1020. <http://dx.doi.org/10.1109/TPWRS.2002.804943>.
- Dougherty, M.R., Scheck, P., Nelson, T.O., Narens, L., 2005. Using the past to predict the future. *Memory Cogn.* 33 (6), 1096–1115.
- Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. JMLR Workshop and Conference Proceedings*, pp. 249–256.
- Gordon, C., Cooper, C., Senior, C.A., Banks, H., Gregory, J.M., Johns, T.C., Mitchell, J.F., Wood, R.A., 2000. The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley centre coupled model without flux adjustments. *Clim. Dynam.* 16 (2), 147–168.
- Guin, A., 2006. Travel time prediction using a seasonal autoregressive integrated moving average time series model. In: *2006 IEEE Intelligent Transportation Systems Conference*. pp. 493–498. <http://dx.doi.org/10.1109/ITSC.2006.1706789>.
- Hamilton, W., Ying, Z., Leskovec, J., 2017. Inductive representation learning on large graphs. *Adv. Neural Inf. Process. Syst.* 30.
- Hastings, S.N., Whitson, H.E., Sloane, R., Landerman, L.R., Horney, C., Johnson, K.S., 2014. Using the past to predict the future: Latent class analysis of patterns of health service use of older adults in the emergency department. *J. Am. Geriatr. Soc.* 62 (4), 711–715.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 770–778.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al., 2020. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* 146 (730), 1999–2049.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9 (8), 1735–1780. <http://dx.doi.org/10.1162/neco.1997.9.8.1735>.
- Hopfield, J.J., 1982. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci.* 79 (8), 2554–2558.
- Kennedy, J.J., Rayner, N.A., Smith, R.O., Parker, D.E., Saunby, M., 2011. Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties. *J. Geophys. Res.: Atmos.* 116 (D14), <http://dx.doi.org/10.1029/2010JD015218>.
- Kipf, T.N., Welling, M., 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.
- Kushnir, Y., Scaife, A.A., Arritt, R., Balsamo, G., Boer, G., Doblas-Reyes, F., Hawkins, E., Kimoto, M., Kolli, R.K., Kumar, A., et al., 2019. Towards operational predictions of the near-term climate. *Nature Clim. Change* 9 (2), 94–101.
- Lacasa, L., Luque, B., Ballesteros, F., Luque, J., Nuno, J.C., 2008. From time series to complex networks: The visibility graph. *Proc. Natl. Acad. Sci.* 105 (13), 4972–4975.
- Lacasa, L., Luque, B., Luque, J., Nuno, J.C., 2009. The visibility graph: A new method for estimating the Hurst exponent of fractional Brownian motion. *Europhys. Lett.* 86 (3), 30001.
- Li, J., Wu, Y., Li, Y., Xiang, J., Zheng, B., 2021. The temperature prediction of hydro-generating units based on temporal convolutional network and recurrent neural network. In: *2021 40th Chinese Control Conference. CCC*, pp. 8228–8233. <http://dx.doi.org/10.23919/CCC52363.2021.9549853>.
- Morice, C.P., Kennedy, J.J., Rayner, N.A., Winn, J., Hogan, E., Killick, R., Dunn, R., Osborn, T., Jones, P., Simpson, I., 2021. An updated assessment of near-surface temperature change from 1850: The HadCRUT5 data set. *J. Geophys. Res.: Atmos.* 126 (3), e2019JD032361.
- Morris, C., Ritzert, M., Fey, M., Hamilton, W.L., Lenssen, J.E., Rattan, G., Grohe, M., 2019. Weisfeiler and leman go neural: Higher-order graph neural networks. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, no. 01. pp. 4602–4609.
- Oono, K., Suzuki, T., 2020. Graph neural networks exponentially lose expressive power for node classification. In: *8th International Conference on Learning Representations. ICLR 2020*, Addis Ababa, Ethiopia, April 26–30, 2020.



- Pathak, J., Hunt, B., Girvan, M., Lu, Z., Ott, E., 2018. Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. *Phys. Rev. Lett.* 120 (2), 024102.
- Rohde, R., Muller, R., Jacobsen, R., Perlmutter, S., Rosenfeld, A., Wurtele, J., Curry, J., Wickham, C., Mosher, S., 2013. Berkeley earth temperature averaging process. *Geoinform. Geostat. An Overview* 1 (2), 1–13.
- Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M., Monfardini, G., 2009. The graph neural network model. *IEEE Trans. Neural Netw.* 20 (1), 61–80. <http://dx.doi.org/10.1109/TNN.2008.2005605>.
- Shi, Y., Huang, Z., Feng, S., Zhong, H., Wang, W., Sun, Y., 2020. Masked label prediction: Unified message passing model for semi-supervised classification. *arXiv preprint arXiv:2009.03509*.
- Smith, D.M., Cusack, S., Colman, A.W., Folland, C.K., Harris, G.R., Murphy, J.M., 2007. Improved surface temperature prediction for the coming decade from a global climate model. *Science* 317 (5839), 796–799.
- Smith, T.M., Reynolds, R.W., Peterson, T.C., Lawrimore, J., 2008. Improvements to NOAA's historical merged land–ocean surface temperature analysis (1880–2006). *J. Clim.* 2283–2296.
- Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. *Stat. Comput.* 14 (3), 199–222.
- Srivastava, R.K., Greff, K., Schmidhuber, J., 2015. Highway networks. *arXiv preprint arXiv:1505.00387*.
- Stott, P.A., Kettleborough, J.A., 2002. Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. *Nature* 416 (6882), 723–726.
- Taylor, J.A., Larraondo, P., de Supinski, B.R., 2022. Data-driven global weather predictions at high resolutions. *Int. J. High Perform. Comput. Appl.* 36 (2), 130–140. <http://dx.doi.org/10.1177/10943420211039818>.
- Terando, A.J., Costanza, J., Belyea, C., Dunn, R.R., McKerrow, A., Collazo, J.A., 2014. The southern megalopolis: using the past to predict the future of urban sprawl in the southeast US. *PLoS One* 9 (7), e102261.
- Wang, Z., Zhang, B., Gao, J., 2019. How to transform the seismic time series to a graph? In: *SEG Technical Program Expanded Abstracts 2019*. Society of Exploration Geophysicists, pp. 3459–3463.
- Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Yu, P.S., 2021. A comprehensive survey on graph neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* 32 (1), 4–24. <http://dx.doi.org/10.1109/TNNLS.2020.2978386>.