






Graph-Based Deep Learning on the Swiss River Network

Benjamin Fankhauser¹, Vidushi Bigler², and Kaspar Riesen¹

¹ Institute of Computer Science, University of Bern, Bern, Switzerland
{benjamin.fankhauser,kaspar.riesen}@unibe.ch

² Institute for Optimisation and Data Analysis, Bern University of Applied Sciences,
Biel, Switzerland
vidushi.bigler@bfh.ch

Abstract. Major European rivers have their sources in the Swiss Alps. Data from these rivers and their tributaries have been collected for decades with consistent quality. We use GIS data to extract the structure of each river and link this structure to 81 river water stations (that measure both water temperature and discharge). Since the water temperature of a river is strongly dependent on the air temperature, we also include 44 weather stations (which measure, for instance, air or soil temperature). Based on this large data corpus, we present in this paper a novel graph representing the water network of Switzerland. Our goal is to accelerate the research of the complex relationships at the (Swiss) water bodies. In particular, we present different graph-based pattern recognition tasks that can be solved on the novel water body graph. In a first evaluation, we use graph-based methods to solve two of these tasks, outperforming current state-of-the-art systems by several percentage points.

Keywords: Water body graph · Water temperature · LSTM · Recurrent Neural Network · Graph data

1 Introduction

Water temperature – an important variable in our ecosystem – is mainly influenced by air temperature. That is, on the water surface a direct exchange with the surrounding air takes place. Thereby, solar radiation is either absorbed by particles in the water or the river bed, then transformed to heat and finally exchanged with the water. Other factors that influence the temperature of water bodies are snow melting, rain, ground water inflow, but also the rate of discharge. Last but not least, also human-made infrastructure plays a pivotal role. For instance, the climate regime shift (CRS) in the late 1980s, caused by anthropogenic and natural origin, led to a sudden increase in water temperature [1, 2].

Supported by Swiss National Science Foundation (SNSF) Grant Nr. PT00P2_206252. Data are kindly provided by the Federal Office for the Environment and MeteoSwiss.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
M. Vento et al. (Eds.): GbRPR 2023, LNCS 14121, pp. 172–181, 2023.
https://doi.org/10.1007/978-3-031-42795-4_16

Three major European rivers have their sources in Switzerland, namely Rhine, Rhône and Inn. In addition, the river Ticino rises in Switzerland, which contributes significantly to the river Po. Furthermore, a large part of the Alps, which separate the southern and northern climatic zones, lies in Switzerland. The Alps are in turn home to large glaciers and huge snow reservoirs, as well as human-made infrastructure such as power plants and dams. In addition, the topology of Switzerland also consists of hilly lowlands, where small rivers flow slowly and are influenced by both large cities and agriculture. Further towards the borders of Switzerland we have the big rivers which are less affected by small disturbances. There are also some medium sized lakes where the inflowing water stays for a long time and thus the outflowing water is only slightly influenced by the inflowing water (there the surface temperature is mainly influenced by the exchange in the atmosphere). Overall, we find that in Switzerland there is a fascinating network of water bodies that has a high complexity.

The present paper is concerned with water temperature predictions using air temperatures by means of graph-based pattern recognition and machine learning. Actually, concerning the climate crisis, rising water temperatures will have a big impact on the Swiss ecosystem. For instance, certain species of fish will not be able to reproduce anymore when the water temperature reaches a certain threshold [3]. Different climate models exist that project air temperature for the future in various versions [4]. Thus, our hypothesis is that it is rewarding to explore more accurate modelling of the air-water model, as this will also lead to better long-term projections of the water temperature.

The contribution of the present paper is threefold. First, based on data of the water bodies stemming from a Geographic Information System (GIS), as well as decades of measurements of dozens of stations, we create a novel and large-scale graph that aims to comprehensively capture and model the complexity of the Swiss water network. Hence, the basis of our research is similar in spirit to other important prediction tasks such as analyses of transportation networks [5], or predictions of loads on networks of power grids [6], to name just two examples. Second, the novel graph allows us to reconsider current approaches to predicting water temperature in rivers. We propose different tasks related to water temperature prediction that can potentially be solved with graph-based pattern recognition algorithms. Third, for two of these tasks, we propose a graph-based prediction system and show that this novel system significantly outperforms two current state-of-the-art methods.

The remainder of this paper is organized as follows. In Sect. 2, we describe two state-of-the-art methods that are currently used for water temperature prediction, viz. the Air2Stream method [7] as well the adaptation of LSTM neural networks [8]. In Sect. 3, we thoroughly describe the novel graph that models the Swiss water bodies and introduce the challenges in predicting water temperature. The novel method for water temperature prediction that employs a graph-based model is then presented and evaluated later in Sect. 4. Finally, in Sect. 5, we draw conclusions and propose possible rewarding avenues for future research activities.

2 Related Work

We are not the first to attempt to predict water temperatures based on air temperatures. In the following two subsections, we present two state-of-the-art models that are actually used as reference systems in our empirical evaluation.

2.1 Air2Stream

Air2Stream is a physically inspired model of the relationship between air and water temperature based on air temperature and discharge [7]. Basis of this model is a differential equation linearized using a Taylor series expansion. The resulting equation has eight tuneable parameters which are calibrated using training data. In particular, the original method employs a particle based optimisation scheme for training and is quite sensitive to the chosen hyper-parameter. In the present paper we use the predictions presented in [9].

2.2 LSTM on Water Data

Long short-term memory (LSTM) is a special type of a recurrent neural network (RNN) [10]. An RNN is a neural network that is applied to a time series on every time step. In addition, an LSTM keeps track of a hidden state and a memory state, two vectors which are inputs to the next time step and will be altered by the LSTM. Thus, the resulting backpropagation variant is called backpropagation through time [11]. To have a trade off between the time series and the update steps, one works with a certain time window, where at the end of every window the gradients are computed and an update step is made. LSTMs have been particularly designed to encounter the vanishing gradient problem. This problem occurs when the back propagation through time has to overcome a lot of time steps and repeated multiplications tend to unstable numeric conditions. In general LSTMs show state-of-the-art results on various time series data [12], and can be applied to the task of water temperature prediction [8, 13] or water level (discharge) prediction [14, 15].

3 The Swiss Water Body Graph

3.1 Construction of the Graph

One of the major contributions of the present paper is that we provide a novel graph based on the Swiss water body. We construct a knowledge graph containing information about the location of river beds (from a GIS), weather data of 44 weather stations (air temperature and more atmospheric measurements), and water data of 81 water stations (water temperature and discharge).

The knowledge Graph $G = (V, E, a_V, a_E)$ is a graph with nodes V and edges E . Each node $v \in V$ has an assigned type T_v . Currently we have three types of nodes i.e. $T_v \in \{\text{water station, universal river node, weather station}\}$. The universal river node is used to model the river itself with sources or river mouths.

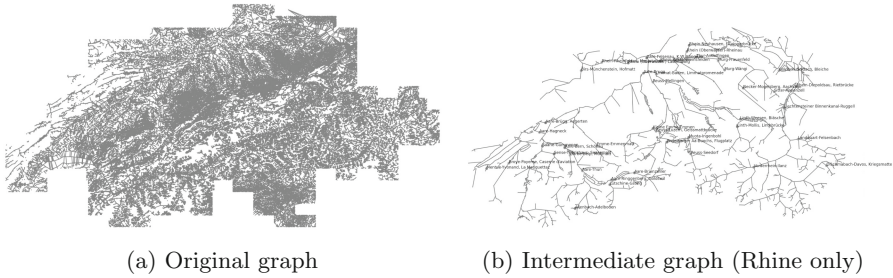


Fig. 1. (a) Original graph before pruning (all edges represent water). (b) Subgraph representing the river Rhine after pruning and added water stations. The illustrated graph is a tree where the water of a child station flows to its parent station.

There is currently only one edge type, which models the connectivity of the nodes. The functions $a_V : V \rightarrow \mathbb{R}^n$ and $a_E : E \rightarrow \mathbb{R}^m$ deliver additional attributes to nodes and edges. Function a_E assigns the edge length in meters to each edge. The nodes are attributed by the air temperature, or water temperature and discharge (depending on the actual type T_v of the node). The data basis for these attributes is thoroughly preprocessed. In particular the data is min-max normalised and outliers are removed by the Federal Office for the Environment (FOEN) as part of their quality control. Each weather station is manually connected by means of an edge to $n \geq 1$ water stations as proposed in [9].

At first glance, the considered data basis seems of natural origin, yet it is not. The current rivers are the product of decades of human intervention of stratification, city planning, power plants, and renaturation. Also the placement and running of the water and weather stations are obviously human based decisions and can change in future. Keeping a constant and high quality of measurements is challenging as it requires decades of stability in the corresponding country, which fortunately is the case in Switzerland.

The original GIS graph contains 258,103 edges and 258,191 nodes representing different types of river segments as well as lake contours. We apply the following preprocessing on this original graph. First, we prune the leaf nodes as not every side creek is important. Nodes that actually contain water stations are never purged. Then, we run a spanning tree algorithm in order to find the shortest paths of the water flow and remove any ambiguity in the graph (for example, when both sides of a lake are modelled as two edges in the graph). Finally, we collapse all edges such that only the connectivity between water stations is left in the form of a tree. This process allows us to compress edge information like river bed length as sum of all collapsed segments between two water stations.

This resulting graph consists of four trees (representing the rivers Rhine, Rhône, Inn, Ticino) with a total of 73 nodes and 69 edges. In Fig. 1(a) the original graph and in Fig. 1(b) the resulting graph is shown (representing the

river Rhine only). Note that in this illustration the edges are not yet collapsed to improve visualisability.

As we have man made changes over time on the underlying water body network, we can create snapshots of the graph at every useful point in time. Based on the visualisation of the available data from 1980 to 2021 (see Fig. 2), we see that many new stations were established after 2002. Hence, we propose two snapshots of the water body graph, viz. one graph that contains fewer nodes and measurements ranging from 1990 to 2021 to represent a long history of measurements and the other one from 2010 to 2021 where we include more stations but on a shorter period of time. In both snapshots we apply an approximate 80/20 training-test split at the end of 2014 and 2017, respectively. The two graphs are named G_{1990} and G_{2010} from now on. Both graphs will be made publicly available for research purpose on the Git Repository of our research group¹.

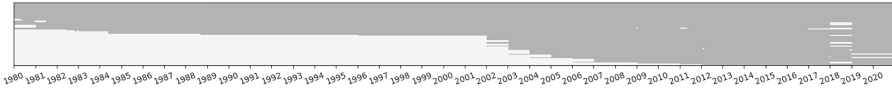


Fig. 2. Visualisation of the available data from 1980 to 2021. Each row represents one station. Dark grey pixels indicate that at a certain day the river water temperature, discharge and air temperature are available. Light grey pixels indicate that at least one of the three values is missing.

3.2 Proposed Water Challenges

The novel graph based representation defined above allows us to rethink current approaches for water temperature prediction. We propose five different benchmark tasks, that can potentially be solved on the basis of the novel graph.

Task 1 - Model Air Temperature Relationship: In this task the goal is to model the relationship between the air and water temperature. This challenge has already been extensively studied and it is what models like Air2Stream [7] or LSTMs [8] are aiming at. Formally, we have both air temperature a_0, \dots, a_t , the discharge q_0, \dots, q_t of $t + 1$ time steps and the goal is to find a model f that predicts the water temperature w_t at time t : $f(a_0, \dots, a_t, q_0, \dots, q_t) = w_t$.

Task 2 - k -Day Forecast: In this task, we do not have access to same day measurements anymore. Given the air temperature a_0, \dots, a_t and the discharge q_0, \dots, q_t of $t + 1$ time steps, the goal is to predict the water temperature w_{t+k} in k days (we define $k \in \{3, 7, 30\}$). Formally, we seek a model $f(a_0, \dots, a_t, q_0, \dots, q_t) = w_{t+k}$. Obviously, the larger k is chosen, the harder the problem (setting $k = 0$ results in Task 1).

Task 3 - Recover from Neighbours: Each water station is built at a certain construction time b_t . One problem of our graph is missing data for this station at times $t < b_t$. The goal of this task is to learn the data of a node for time points

¹ <https://github.com/Pattern-Recognition-Group-UniBe/swiss-river-network>.

$t < b_t$ based on the relationships with its neighbours. By filling in missing data, this procedure allows us to construct an estimated graph of water temperatures back to 1980 using all stations (although we cannot assess the quality of the estimates).

Task 4 - Work on Degenerated Data: A challenge for any sensing and recognition system are degenerated sensors. The fourth task is to detect and repair potentially corrupted data. Formally, we define a function drift $d(n) \in \mathbb{R}$, where n is the n -th day after construction and d is a function to model the amount of drift. The drift is then added to the water temperature during training: $w'_t = w_t + d(t - b_t)$, where w'_t is the degenerate training data and b_t is the construction time of the water station.

Task 5 - Few Shot Learning: The goal of this task is to minimise the effort required to collect water temperatures. Imagine a mobile sensor system that is moved from one place to another every month. When the mobile sensor system is on site, the data is available and can be used for training. The goal is to have as few of these mobile sensor systems in use as possible and still achieve a reasonable estimation of the water temperatures.

4 Proposed Method and Experimental Evaluation

4.1 Experimental Setup and Reference Models

In this paper, we use the snapshots of the graphs G_{1990} and G_{2010} as described in Sect. 3.1 and the corresponding training and test splits to solve Task 1 and Task 2 as defined in Sect. 3.2 (that is, predicting the water temperature in k days with $k \in \{0, 3, 7, 30\}$). To investigate the quality of the prediction, we measure and report widely used metrics, namely the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE) on the test set. In addition, we measure and report the Nash-Sutcliffe model Efficiency Coefficient (NSE), which is often used to assess the predictive skill of hydrological models. Formally, the three ratios are defined as follows

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

$$\text{NSE} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

where n describes the number of measurements, y_i the actual measured value, \hat{y}_i the value estimated by the model, and \bar{y} the mean of the actual measured values. For a perfect model with an estimation error variance equal to zero, the resulting NSE equals 1. That is, values of the NSE nearer to 1 suggest a model with more

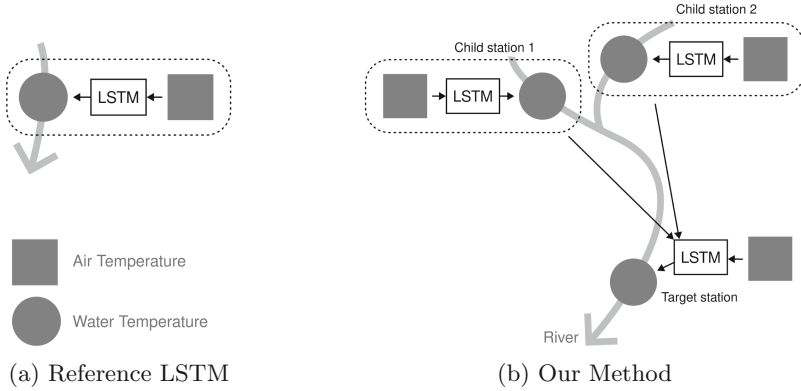


Fig. 3. (a) The reference method models the air to water relationship in a 1-to-1 manner [8]. (b) The proposed method makes use of the local neighbourhood on the novel water body graph. We adapt our LSTM architecture to the amount of child stations contributing to the target station and train one such LSTM per target station.

predictive skill. While for the errors, of course, values closer to 0 indicate good prediction quality.

For our evaluation, we use a total of three different reference models.

1. **Air2Stream:** The Air2Stream model as presented in Sect. 2.1. We only provide here the RMSE results from [9] (and we cannot use the other metrics for comparison).
2. **Baseline:** The baseline system refers to the unweighted average of the water temperature of the target station and the water temperatures of its child stations using the reference LSTMs (see below).
3. **LSTM:** For this reference system [8], we use LSTMs that take the air temperature as input (as described in Sect. 2.2 – see Fig. 3(a)). To find a suitable architecture, we perform a grid search on the width of the hidden layers, the depth of the LSTM, the learning rate, and the weight decay (we use the Adam optimiser). During validation we obtain the best results for 32 in width, 1 in depth, 0.01 for learning rate, and a weight decay of $1e-6$.

4.2 The Novel Graph-Based Model

For our new model, we use the four graphs described in detail in Sect. 3.1. The new model uses the locality of the graph structure to model the time series data and consists of two different nodes.

- Child station: Water station upstream to the target station
- Target station: Water station we want to predict.

We extract a subgraph for each target station with its c child stations. One such subgraph with one target station and $c = 2$ child stations is shown in Fig. 3(b).

In Task 1 and Task 2, we do not have access to measured water temperatures of the child stations as input, but we can estimate them using any air to water model. For each child station, we train a reference LSTM to obtain an estimate of the water temperature. Then we train an additional LSTM for the target station. This LSTM is given the estimated water temperatures of the child stations and the air temperature of the target station as input (see Fig. 3(b)).

More formally, the resulting recurrent neural network consists of an LSTM layer with the input size $c+1$. The LSTM uses a larger hidden space than its input size. The size of the hidden space is determined by a factor of the input size. After the LSTM layer, we project the hidden space to the desired output size using a linear layer. Our neural network models the function $f(\hat{w}_t^{(1)}, \dots, \hat{w}_t^{(c)}, a_t^{(ts)}) = \hat{w}_{t+k}^{(ts)}$ where $\hat{w}_t^{(x)}$ is the estimated water temperature at child station x and $a_t^{(ts)}$ is the air temperature at the target station ts at time t , and k depends on the current prediction task ($k \in \{0, 3, 7, 30\}$).

For the training of our model, we perform a grid search for both width and depth of the LSTM and use the Adam optimiser with a learning rate of 0.01 and a weight decay of $1e-6$.

Graph Neural Networks (GNNs) with message passing [16] are somehow related to the proposed method. Similar in spirit is, for instance, GraphSAGE [17]. However, while GraphSAGE uses an LSTM to handle a flexible amount of neighbours during the message aggregation phase, we have a fixed amount of child stations but a flexible amount of time steps to handle. Moreover, GNNs aim to process the graph as a whole input unit. In the proposed method we train a neural network individually per target station. This removes any inductive property as our trained networks do not generalise to other graphs.

4.3 Test Results

The results we obtain on both versions of the graph (i.e. G_{1990} and G_{2010}) are shown in Table 1. The metrics RMSE, MAE, and NSE are reported for the respective test years. In column $k = 0$, the results for estimations of the same day are shown (Task 1). In the columns $k = 3$, $k = 7$, and $k = 30$, we show the prediction results for 3, 7, and 30 days in the future, respectively (Task 2).

First, we observe that our new model performs best for both graphs and all four tasks (measured across all three evaluation metrics). On average, we outperform the state-of-the-art method Air2Stream by 23%. Moreover, on average, the novel location-based method outperforms the state-of-the-art LSTMs by about 5%, remarkably more at the most difficult task $k = 30$. The results of the baseline show that a simple average of locally connected water temperatures is a poor estimate.

Regarding the results, we conclude that the proposed method is a flexible extension to any system that models the relationship of air temperature and

Table 1. The results achieved on the test sets by our method and the reference systems on two versions of the graph (G_{1990} and G_{2010}). In the $k = 0$ column, we report the results for the same day relation (Task 1), and in the $k = 3$, $k = 7$, and $k = 30$ columns the forecasts for 3, 7, and 30 days in future, respectively (Task 2). The best result per metric, task and graph is shown in bold face. *The Air2Stream model uses similar years for training but a different set of test years.

| Metric | Method | Graph Version | | | | | | | |
|--------|-----------------------|---------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | | G_{1990} | | | | G_{2010} | | | |
| | | $k=0$ | $k=3$ | $k=7$ | $k=30$ | $k=0$ | $k=3$ | $k=7$ | $k=30$ |
| RMSE | Air2Stream* | 1.05 | - | - | - | 1.05 | - | - | - |
| | Baseline | 1.90 | 2.00 | 2.16 | 3.02 | 1.65 | 1.87 | 2.00 | 2.41 |
| | Reference LSTM | 0.80 | 1.10 | 1.37 | 2.29 | 0.91 | 1.24 | 1.43 | 2.24 |
| | Ours | 0.75 | 1.07 | 1.30 | 1.77 | 0.85 | 1.19 | 1.39 | 1.65 |
| MAE | Baseline | 1.58 | 1.65 | 1.79 | 2.38 | 1.36 | 1.53 | 1.62 | 1.95 |
| | Reference LSTM | 0.60 | 0.84 | 1.04 | 1.68 | 0.69 | 0.93 | 1.09 | 1.69 |
| | Ours | 0.56 | 0.81 | 0.99 | 1.36 | 0.63 | 0.89 | 1.05 | 1.26 |
| NSE | Baseline | 0.82 | 0.81 | 0.77 | 0.54 | 0.85 | 0.83 | 0.81 | 0.76 |
| | Reference LSTM | 0.97 | 0.94 | 0.90 | 0.74 | 0.96 | 0.93 | 0.91 | 0.77 |
| | Ours | 0.97 | 0.95 | 0.93 | 0.86 | 0.97 | 0.94 | 0.92 | 0.89 |

water temperature. We argue that our system is able to capture water temperature changes of upstream stations, which in general results in an improvement of the prediction accuracy. A more in depth analysis of the performance of individual stations, however, also reveals that there is no improvement for some individual stations. The reason for this observation is that some water stations have no dependence on their upstream water stations (e.g., when a lake lies between two stations).

5 Conclusion and Future Work

In this paper, we address the difficult task of analysing water networks in complex environments. This is indeed an important task, as the climate crisis is one of the greatest challenges facing humanity. We propose to model the complex water network of Switzerland using a graph. Based on this graph, we propose five different challenging tasks that can potentially be solved using graph-based pattern recognition or machine learning methods. Two of these five tasks are solved in this paper using a graph-based model built on LSTMs. In a large-scale experimental evaluation, we show that the proposed model can improve the widely used Air2Stream model by about 23% and an isolated (i.e., non-graph-based) LSTM by about 5%. We see many worthwhile future research activities. Currently we are working with the authorities to extend the graph with more water stations as well as other node types like cities, power plants and lakes. Moreover, we will tackle the remaining benchmark tasks and explore more possibilities of neural networks on our novel graph.

References

1. Reid, P.C., et al.: Global impacts of the 1980s regime shift. *Glob. Change Biol.* **22**(2), 682–703 (2016)
2. Woolway, R.I., Dokulil, M.T., Marszelewski, W., Schmid, M., Bouffard, D., Merchant, C.J.: Warming of central European lakes and their response to the 1980s climate regime shift. *Clim. Change* **142**, 505–520 (2017)
3. Dahlke, F.T., Wohlrab, S., Butzin, M., Pörtner, H.O.: Thermal bottlenecks in the life cycle define climate vulnerability of fish. *Science* **369**(6499), 65–70 (2020)
4. CH2018: Climate scenarios for Switzerland. Technical report, National Centre for Climate Services, Zurich (2018)
5. Yu, B., Yin, H., Zhu, Z.: Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. arXiv preprint [arXiv:1709.04875](https://arxiv.org/abs/1709.04875) (2017)
6. Khodayar, M., Liu, G., Wang, J., Khodayar, M.E.: Deep learning in power systems research: a review. *CSEE J. Power Energy Syst.* **7**(2), 209–220 (2020)
7. Toffolon, M., Piccolroaz, S.: A hybrid model for river water temperature as a function of air temperature and discharge. *Environ. Res. Lett.* **10**(11), 114011 (2015)
8. Qiu, R., Wang, Y., Rhoads, B., Wang, D., Qiu, W., Tao, Y., Wu, J.: River water temperature forecasting using a deep learning method. *J. Hydrol.* **595**, 126016 (2021)
9. Råman Vinnå, L., Zappa, M., Piccolroaz, S., Bigler, V.C., Epting, J.: Swiss-wide future river temperature under climate change, swissfurite (unpublished)
10. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997)
11. Werbos, P.J.: Backpropagation through time: what it does and how to do it. *Proc. IEEE* **78**(10), 1550–1560 (1990)
12. Siami-Namini, S., Tavakoli, N., Namin, A.S.: A comparison of ARIMA and LSTM in forecasting time series. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 1394–1401. IEEE (2018)
13. Jia, X., et al.: Physics-guided recurrent graph model for predicting flow and temperature in river networks. In: Proceedings of the 2021 SIAM International Conference on Data Mining (SDM), pp. 612–620. SIAM (2021)
14. Kim, D., Han, H., Wang, W., Kim, H.S.: Improvement of deep learning models for river water level prediction using complex network method. *Water* **14**(3), 466 (2022)
15. Zhao, Q., et al.: Joint spatial and temporal modeling for hydrological prediction. *IEEE Access* **8**, 78492–78503 (2020)
16. Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O., Dahl, G.E.: Neural message passing for quantum chemistry. In: International Conference on Machine Learning, pp. 1263–1272. PMLR (2017)
17. Hamilton, W., Ying, Z., Leskovec, J.: Inductive representation learning on large graphs. In: Advances in Neural Information Processing Systems, vol. 30 (2017)