GCN-Att-GRU: Spatio-temporal Graph Convolution with Multi-Head Attention for Multi-step Urban Water Demand Forecasting

Muhua Hu
Mathematics and Digital Sciences Department of Chengyi College, Jimei University, XiaMen, FuJian,361021, China
E-mail: humuhuacy@163.com

Keywords: urban water demand prediction, china, temporal and spatial attributes, GCN-Att-GRU Model, multi-head attention model

Received: July 12, 2025

Urban water resource planning as well as the supply-demand balance are directly affected by immediate demand forecasts, making them vital for water resource management. The present approaches to water demand forecasting just consider temporal variables and do not take account of the possible impact of geographical characteristics on them. The reason for this is because short-term urban water demand forecasts are affected by several variables, many of which display complex nonlinear dynamic features. Predictions end up being inaccurate because of this. This research aims to address this problem by presenting a model that considers geographical and temporal characteristics in order to predict urban water consumption in the near term. The first step is to detect and fix anomalies using the Prophet model. In order to generate an adjacency matrix among variables, we use a maximum information coefficient, and to extract spatial attributes between variables, we use a graph convolutional neural network. Afterwards, a multi-head attention method is used to enhance crucial aspects of water consumption statistics while reducing the influence of unimportant components. The next phase involves projecting urban areas' short-term water demands using a Gated Recurrent Unit (GRUs). The suggested model achieves a Mean Absolute Percentage Error (MAPE) of 2.98% (±0.05) across a 7-day forecasting horizon, which is a significant 7.65% decrease in MAPE when compared to the optimized baseline. The mechanistic essentiality of the MIC-GCN and the MHA mechanism's integration for this enhanced performance has been confirmed by ablation experiments. Based on these findings, the GCN-Att-GRU may be confidently used to predict the operational water demand in cities.

Povzetek: Raziskava predstavi model, ki združi prostorske (grafovne) in časovne informacije za natančnejše kratkoročno napoved mestne porabe vode ter doseže bistveno nižjo napako (MAPE ~3%) kot obstoječe metode.

1 Introduction

Per capita freshwater availability is poor in China, even though the country has a lot of freshwater resources overall [1]. The rapid growth of China's urban areas has resulted in water shortages for the majority of cities. To address this issue, it is essential to carefully manage how urban residents utilize water in their homes [2]. The water consumption of urban residential areas is characterized by complicated nonlinear dynamics and is affected by a multitude of variables [3]. When it comes to complex water use trends, traditional approaches relying on human experience fall short [4]. However, urban water demand forecast may help uncover consumption patterns and guide efficient water resource allocation and use. There are three types of urban water demand predictions: short-term, medium-term, with long-term [5]. The immediacy and urgency of the short-term predictions make them especially important for water resource management [6]. Cities can increase the efficiency of their water delivery systems, save costs and energy usage, and plan for the future with more precision when they can accurately estimate water demand in the near term [7].

Several methods exist at the moment for forecasting future water demand. Grey models, SVMs, RFs, regression analysis designs, and traditional BPNNs are all part of this category [8]. These methods may be effective in certain contexts, but they lack the ability to generalize or make accurate predictions [9]. Predicting urban water demand has become more popular with the use of neural network along with deep learning techniques, thanks to recent advancements in this field. When it comes to predicting future water demands, the GRU model has performed quite well [10]. It excels at handling data with a high temporal resolution, as well as data with abrupt changes and uncertainty [11]. Additional details, such as the current day and national holidays, may also be included. However, there are still major obstacles to overcome, such as the effects of weather and information noise on the precision of predictions [12].

Current methods still mostly focus on extracting time series data's temporal properties, ignoring the impact of spatial variables on these temporal elements, despite substantial progress. Currently, multi-site analysis is the main emphasis of research on spatiotemporal characteristics [13].

Previous work has taken into account the impact of these aspects on water consumption data over time, but current efforts primarily aim at establishing correlations between spatial qualities at different observation locations rather than between the geographical features that comprise clear variables [14].

Given the above, GCN-Att-GRU, a model for attention processes in graph convolutional neural networks that include both spatial and temporal variables, is presented by the authors of this research [15]. Despite its complexity, this approach provides more accurate short-term predictions of urban water demand than earlier techniques [16]. Finding and fixing outliers is the first step in improving water data quality using the Prophet model [17]. Next, we use the greatest information coefficient to calculate the relationship between the weather and the water use statistics [18]. The next step is to use a multi-head attention mechanism to ensure that the graph convolutional network extracts an equal amount of geographic information on water usage and weather conditions [19]. Finally, the processed data is input into a three-layer GRU with residual connections added to improve the accuracy of the water demand estimate [20].

The following is the structure of the paper: The sources of the processed urban water data are detailed in Section 2. Section 3 explains the steps taken by the GCN-Att-GRU model. In Section 4, we examine the experimental data to verify the model's correctness. Section 5 wraps up the investigation by discussing the research outcomes.

2 Related work

Drawing on the work of researchers in [21], the authors of that paper develop a model that takes geographical and temporal factors into account in order to predict urban water demand for the near future. The first step is to detect and fix anomalies using the Prophet model. In order to generate an adjacency matrix among variables, we use a maximum information coefficient, and to extract spatial attributes between variables, we use a graph convolutional neural network. Afterwards, a multi-head attention method is used to enhance crucial aspects of water consumption statistics while reducing the influence of unimportant components. Predicting urban regions' short-term water consumption using a three-layer GRU structure is the next stage. By reducing the average absolute percentage error by 1.868-2.718% compared to existing prediction models, the study suggests a hybrid model that improves forecast accuracy and effectiveness. This one may aid municipalities in making better use of their water resources and will also pave the way for further research.

In a revolutionary move, the authors of [22] use data pre-processing and an Artificial Neural Network improved using the Backtracking Search Process to forecast monthly water consumption based on previous use. Information on monthly water use in Gauteng Province, South Africa, from 2007 to 2016 was utilized to design and evaluate the approach. Data pre-processing approaches were crucial for improving data quality before creating the prediction model. The best result was achieved by the BSA-ANN model, which had an efficiency coefficient of 0.979 and a root-mean-square error of 0.0099 mega liters. When compared to the Crow Search Algorithm, it was shown to be more efficient and dependable in terms of error scale. In sum, that article introduces a novel use case for the BSA-ANN hybrid model, which demonstrates how to accurately forecast water demand in a metropolis hit hard by both climate change with population boom.

In [23], the authors introduced a probabilistic approach to forecasting that accounts for systemic uncertainty to foretell future water use trends. That approach uses statistical methods to predict the distribution of possible future water demand situations. Better scheduling and handling of water resources is achievable because to the many options presented to decision-makers by that study's results. They developed a hybrid model for hourly water demand estimation using conformal predictions after comparing several machine learning approaches. That model integrates CNNs with bidirectional long short-term memories. Furthermore, they discuss important aspects of constructing a probabilistic forecasting framework, such as picking the right data and the right parameters for the model. For probabilistic water demand forecasting, the suggested model has been tested and shown to work in the actual world. Deterministic predictions improved by 10% and probabilistic forecasts by 26.7%, according to the results. Results show that method has promise for better resource management and decision-making.

Data preprocessing has the potential to fix the outlier issue, as stated by the authors of [24]. The results demonstrate that the proposed hybrid model outperforms similar machine learning models when it comes to prediction accuracy and resilience in the face of prediction uncertainty. Feature engineering improves the data input dimension before the LightGBM (Light Gradient Boosting Machine) method is used to predict future water demand. When it comes to missing value interpolation jobs, the findings show that cubic polynomial interpolation is better than both the Prophet model and the linear technique. The LightGBM model does a great job of predicting future patterns in water demand and has great forecasting performance overall. The test dataset has an assessment indication of 4.28% for MAPE (mean absolute percentage error) with an indicator of 0.94 for NSE (Nash-Sutcliffe efficiency coefficient). With these metrics as a foundation, water supply firms may make educated guesses about the near future.

The authors of [25] the classic techniques of water demand forecast are encountering difficulties, necessitating the development of new methods and instruments to enhance the precision and efficacy of their projections. From 1983 to 2024, the pertinent research on urban water demand projection was visually analyzed in the CNKI database with the Web of Science core literature database using CiteSpace software, bibliometrics, and visualization analysis methodologies. In order to pinpoint areas of concentrated research and developing patterns in urban water demand prediction, they will be using visual tools like keyword co-occurrence while clustering, keyword growth, etc. to compile their findings. That will serve as a roadmap for future work in that area.

In order to understand the regularity of urban water consumption and accomplish accurate water demand forecasting, an improved Legendre memory that features dual feature channels that are both linear and nonlinear was proposed by the experimenters in [26]. A precise mapping link between the target variable and exogenous variables is established in the information instantiation component with the introduction of an adaptive gating system that reallocates variable weights. Feature extraction becomes easier while high-frequency noise is suppressed because to the model's dual-channel topology, which simultaneously collects nonlinear and linear characteristics from historical data. That approach surpasses expectations, successfully predicting future water demand in a variety of contexts, according to real-world urban water demand statistics. The model routinely outperforms the competition across a variety of tasks in publicly available datasets, further demonstrating its high applicability.

The researchers of [27] provide a novel approach to forecasting urban water demand by integrating demographic variables such as population growth into a combination of deep learning as well as predictive modeling. weather variables like rainfall and temperature, and data on past water use. Since it can learn relationships between numerous observations in data, the GRU model is used for time-series prediction in that study. that model, in conjunction with the future water demand approach, anticipates shortages by comparing the expected demand with the actual quantity of water that is already available. Findings from that study stress the significance of demand forecast visualization for aiding resource deployment decisions and facilitating quick responses to water scarcity emergencies. The long-term sustainability and resilience of urban ecosystems may be enhanced by the application of predictive models in water management methods. More sustainable use of water resources, fewer waste, and better resource distribution are all possible outcomes.

In order to predict the monthly demand for water in cities, the authors of [28] propose looking at water use across many time periods. In addition to that ANFIS model, they were evaluated alongside hybrid crow search techniques with artificial neural networks because of recent accomplishments in applying these methods to other engineering optimization problems. Based on a variety of statistical metrics measurements, including the coefficient of efficacy (0.974, respectively), the research found that both ANFIS as well as CSA-ANN are statistically equal and can accurately anticipate the monthly demand for urban water. To get the most out of a denoising approach on raw time series data and to choose the right inputs for a model, data preparation is essential. In light of the increasing demand for urban water systems, policymakers may find this study helpful in making choices with reduced decision-related risk.

This study employed a hybrid model consisting of a convolutional neural network and a bidirectional long-term memory network, as suggested by the authors [29]. Special circumstances, such as weather-related calamities and vacations, are addressed with an appropriate remedial plan. The Bi-GRU method may be used to forecast urban water demand by entering the features retrieved from weather and water quantity data. To find out how weather and water use in cities are related, the researchers in that study looked at the correlation between the two. They used the water consumption data from the five days prior to and the daily maximum temperature as the basis for the vacation adjustment approach using the temperature correction model. The improved prediction results may be seen by comparing the models prior to and after the deviation was fixed. Based on comparisons with CNN, bidirectional long-term memory systems, sparse autoencoders, and CNN-GRU, the current study found that the CNN-Bi-GRU model decreased prediction error. Lastly, they compared the six models' training durations and convergence rates throughout the same training window. Although it takes more time than SAEs, CNN-Bi-GRU's training time is lower than that of Bi-GRU, CNN, GRU, and CNN-GRU.

The researchers set out to find out what elements in the weather affect water consumption (Ref. 30). It also sought to provide a decent and trustworthy method for predicting the demand for municipal water by combining the Gravitational Search Algorithms, the Backtracking Search Algorithm, and an Artificial Neural Network. In order to assess their effect on the water demand, eight meteorological variables were also included. The main takeaways from that study are that the annual and seasonal phases are best served by the hybrid GSA-ANN (Agent=40) model's fitness function, as measured by RMSE. Additional evidence from all years and seasons demonstrates that the GSA-ANN framework is capable of reliably predicting daily water usage patterns.

3 Proposed work

A. Problem definition

This research uses the graph to characterize the road network $G = (V, E, A)$, where V is a finite set signifying $|V| = N$ There are nodes in the urban network, and E is the collection of edges that link the nodes in graph G; $A \in \mathbb{R}^{N\times N}$ is the graph G's normalized adjacency matrix, which shows the distance and direction between nodes. The signal at time step t in graph G is $X_t = \{x_t^1, \dots, x_t^N\} \in \mathbb{R}^{N\times F}$, where $x_t^n (n \in \{1, \dots, N\})$ is the total number of structures seen at each node; are all features gathered through the n-th sensor at time step t. An additional connection is built into the deep network to mitigate the loss of spatio-temporal features. This connection may use 1×1 convolution for projecting the input towards the feature space of the result of the spatial feature extractor. The final product is the sum of the adaptive residual and the output $\widehat{H} = \{\widehat{h}_1, \dots, \widehat{h}_N\} \in \mathbb{R}^{N\times C\times T}$ the result that the ReLU function returns. The steps involved in the computation are as follows:

$$\widehat{H} = \mathrm{ReLU}\left(\widehat{H}_N + W_r \odot \Gamma_{\theta r}(x)\right) \quad (1)$$

where $\Gamma_{\theta_r}(\cdot)$ means that the 1×1 convolution operation is being performed using W_r as a learnable parameter and θr as the parameter.

At last, the output from the completely connected layer that corresponds to the projected target form is obtained after normalizing $\widehat{H}$.

B. Data sources

The study will take place in Beijing, a city in northern China situated on the Yellow River's eastern bank. In terms of geographic coordinates, Beijing may be found in 38°28′ N and 42°05′ N, as well as 1153°45′ E and 117°30′ E. Seasonal changes and low annual precipitation define its mild semi-humid continental climate. Economic, social, and environmental sustainability in metropolitan areas depend critically on water resource distribution in this climate.

The Beijing Water Authority's daily water use statistics from 1 September 2016 to 1 August 2023 are used as the data source for this research.

For training, we use 80% of the data, and for testing, we use 20%. We will split the training and evaluation sets on March 13, 2022. The National Meteorological Data Center's supplementary weather data for Beijing is shown in Table 1. The data includes fifteen different aspects of weather, such as average and sea level pressures, high and low temperatures, average and lowest dew points, average and highest rainfall, and maximum and minimum humidity, average and minimum humidity, average and sea-level visibility, average and meteorological station atmospheric pressures, average and minimum dew points, average and sea-level temperature, average and meteorological station atmospheric pressure, average and total precipitation, average and total precipitation, and average and total precipitation, which is given in fig 1, fig 2, and fig 3 respectively.

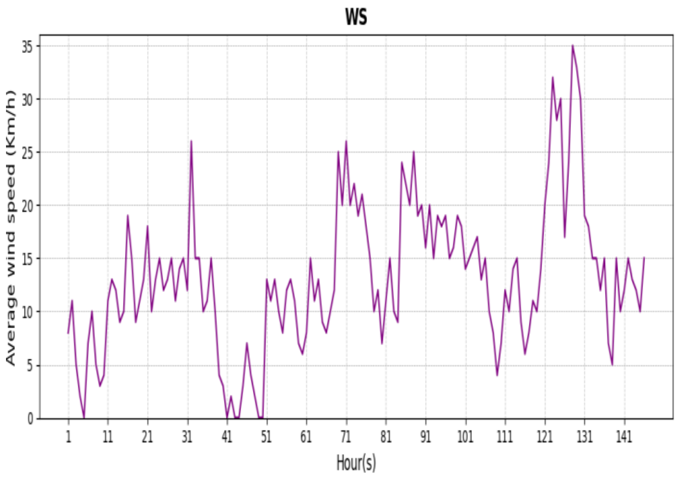

Figure 1: Average wind speed analysis with hours

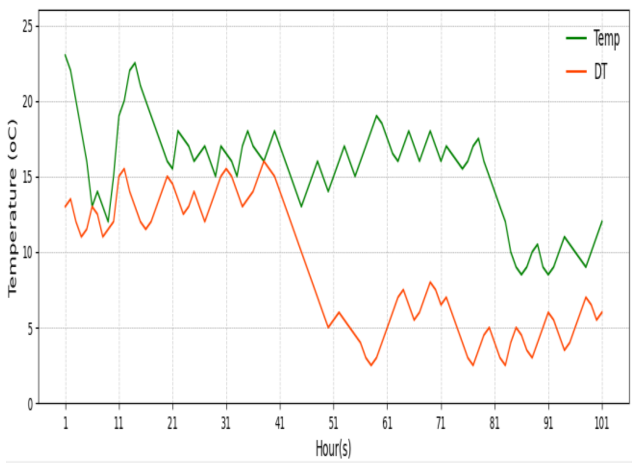

Figure 2: Temperature with hours

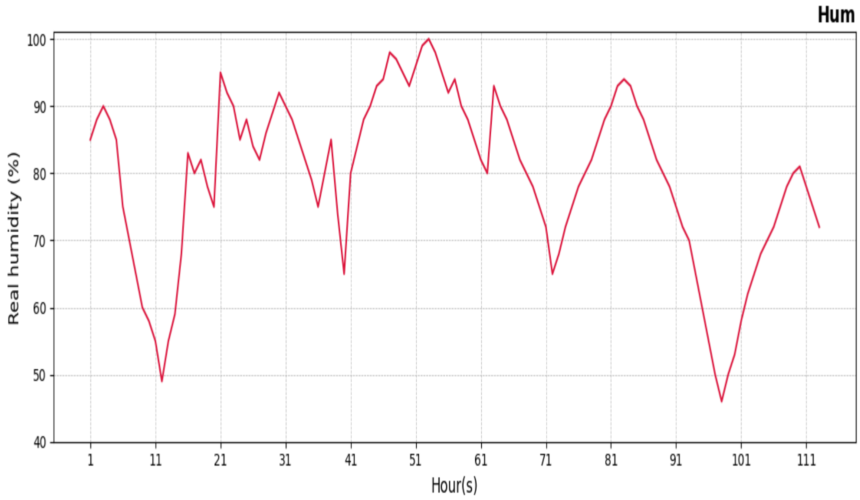

Figure 3: Real humidity with hours

Table 1: Urban water demand influencing variables

Feature	**Notation**	**Description**	**Unit**
Dew Point	f_3	Dew Point (max value)	(°C)
	f_4	Dew Point (min value)	(°C)
	f_5	Dew Point (mean)	(°C)
Humidity	f_7	Humidity (max value)	(%)
	f_8	Humidity (min value)	(%)
	f_9	Humidity (mean)	(%)
Rainfall	f_6	Rainfall (mean)	(mm)
	f_{14}	Rainfall (max value)	(mm)
Temperature	f_0	Temperature (max value)	(°C)
	f_1	Temperature (min value)	(°C)
	f_2	Temperature (mean)	(°C)
Visibility	f_{12}	Horizontal Visibility (mean)	(km)
Wind Speed	f_{13}	Wind Speed (mean)	$(m \cdot s^{-1})$
Atmospheric Pressure	f_{10}	Sea-Level Atmospheric Pressure (mean)	(mmHg)
	f_{11}	Meteorological Station Atmospheric Pressure (mean)	(mmHg)

Data analysis

This research painstakingly selected demand for water and related meteorological variables, such as f0-f5, f11, f13, and f14, in order to analyze the dynamic residential urban water demand. The following part lays out the criterion for choosing these variables.

As seen in Figure 4, visuals were generated for every element after the selection process. Water consumption may reach 3,758,000 m3/day at temperature as high as 41 °C. There has been a minimum per day water demand of 2.034 million cubic feet and a minimum recorded temperature of -19 °C. The average yearly rainfall is a pitiful 577 millimeters. The graph makes it quite evident that these variables are following predictable trends in their changes. We used the Mann-Kendall trend test to statistically evaluate the changes in urban water demand (y) as time passed in relation to several climatic parameters. If the Z-value is positive, then the trend in the time series data is increasing, and if it is unfavorable then the trend is decreasing. When p-values considered less than 0.05, trends were considered statistically significant at the 95% confidence level. Table 1 shows a compressed version of the data showing that f0, f2, f3, f4, f5, and f14 all became significantly different from one another, but y remained relatively constant. On the other hand, there was no statistically significant trend for variables f1, f11, and f13, which might be due to data issues, small sample size, or range limitations.

This study dug further into the yearly shifts in water resources by analyzing data segments that spanned the whole years of 2017–2022. Spring is March–May according to the solar calendar, summer is June–August, autumn is September–November, and winter is December–February. Figure 33 shows the yearly data on water consumption throughout that time period. It's easy to observe that water use drops in the winter and spring, but then spikes significantly in the summer. Using data collected from 2017 through 2022, we conducted a Mann-Kendall patterns test to get a better understanding of the yearly changes regarding water usage.

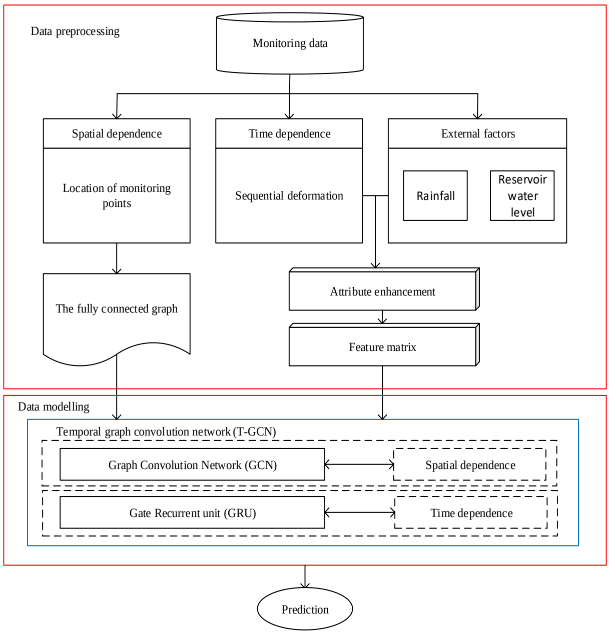

Figure 4: GCN with GRU model for urban water consumption prediction

Data processing

This research analyzes Beijing's water use over time and uses the Prophet model to spot and fix anomalies based on factors including seasonality, cyclicity, and vacations. The Prophet model is a potent tool for predicting time series data that is significantly affected by holidays and seasons. Seasonality, trend, holidays, and residual are the four main components of the time series that the model breaks down in order to examine and reproduce the data's inherent patterns. For time series decomposition specifically, this is the formula (given in Table 2):

$$y(t) = g(t) + s(t) + h(t) + \varepsilon(t) \quad (2)$$

The fact that g(t) represents the trend aspect and y(t) represents the chronology of daily urban water demand indicates that the chronology is non-cyclical. To show how the data has changed over time, it is often shown as a logistic or piecewise linear growth curve; the seasonal component, denoted as s(t), represents the effect with a constant cyclicity. By using Fourier series, Prophet is able to replicate the cyclical character of these patterns. The effect on the time series on specific days is shown by the holiday component, h(t). By entering a user-defined list of holidays, Prophet enables users to specifically include these occasions. The impact of these vacations on the time series information is then estimated by the model. The error term, denoted as ε(t), describes variations that were not anticipated by the model. Since the predicted normal distribution is not taken into consideration by the model's components (trend, seasonality, and holidays), it is used to capture the unpredictability in the time series data.

The Prophet model gets its prediction by first predicting each part of the time series independently, and then adding them all together. Creating confidence intervals for the predicting values is crucial for anomaly identification inside the Prophet model. An anomaly in a time series is a point that does not belong in the original data set and falls outside the confidence interval. Instead of interpolating missing data, the model fills in anomalies as well as missing values using forecasted findings.

Table 2: Model parameters for the prophet.

Parameter	Description	Parameter Value
Growth	A tendency that is linear	Linear
Seasonality_prior_scale	Seasonal component flexibility	10
Holidays_prior_scale	Adaptability of the vacation package	10
Changepoint_range	Plot illustrating possible shifts in the trend	0.8
Yearly_seasonality	Part of the year that is seasonal	True
Weekly_seasonality	Feature that changes with the seasons and runs weekly	True
Daily_seasonality	Seasonal component included daily	False
Changepoint_prior_scale	Manages the model's adaptability	0.09
Interval_width	Interval of confidence	0.99

The first step is to organize and classify all of the gathered data, which includes things like pressure, temperature, humidity, The force, direction, and velocity of the wind. Filling in any blanks in the dataset is the next stage. Excluding long-term missing information as well as linearly interpolating short-term and short-missing information allows us to be more exact. With the proportions a: b: c, the data is then divided into three collections: A for training, B for validation, while C for testing.

Finding the most influential input factors and the times in history when they had a substantial effect on the anticipated power is essential when working with several input variables. In order to determine the relationship between wind power and the values of each variable at various sampling periods, we use training set A to compute the Pearson correlation coefficient using the formula provided in (3).

$$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}} = \frac{E(X-E(X))E(Y-E(Y))}{\sqrt{D(X)}\sqrt{D(Y)}} \quad (3)$$

in where E stands for the mathematical expectations, D for the variance, and Cov(X, Y) for two separate bodies of evidence, X and Y, are referred to as the correlation of the sum of random variables used to quantify the overall inaccuracy. If the absolute value of |ρ_XY | is equal to or higher than 0.6, it suggests that variables X as well as Y are strongly correlated. Lastly, in order to produce training set A^', validation set B^', and evaluation set C^', the dataset is rebuilt and normalized using the Pearson correlation coefficients as a basis for variable and step selection. Equation (4) shows the normalizing procedure, which is Min-Max normalizing:

$$x_{std} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (4)$$

where x is the initial value, x_{std} is the value after normalization, besides x_{max} and x_{min} are the limits, correspondingly.

GCN-Att-GRU model construction

To accurately predict urban water demand in the near future, researchers proposed a GCN-Att-GRU model that integrates geographical and temporal variables.

The model started by using the Prophet framework to break down water usage data into its four component parts, and then it used forecasted values to find and fix anomalies. Data was also normalized using minmax to avoid overflow during calculations, which improved prediction efficiency.

The GCN and MIC analyses were fed 16 variables for spatial feature processing in this investigation. In the graph structure, parameter X comprises the variables f_0-f_14 that pertain to climate data and y, which pertains to water data. The next step was to use the MIC to find and show the indirect geographical correlations between water data and weather data. In addition, a 16H16 dimensional adjacency matrix A was built to depict the similarity between characteristics. In order to simplify things, a threshold τ was established to exclude similarity values below it, sparsifying the adjacency matrix. By combining data from nearby nodes, the GCN provides the water consumption node with geographically relevant features that are used to predict future water demand. The technique restricted the total amount of GCN layers to two due to the data's quantity, complexity, and the available computational resources. The multi-head attention method was able to dynamically calculate importance for each place and weighted aggregate the most relevant portions by applying it to spatial attributes.

A three-layer GRU was trained to anticipate future water demands using the retrieved spatial characteristics as input for temporal feature processing. To avoid overfitting and neural network deterioration, which may happen when using many GRU layers, we connected the initial and final GRU layers using a residual link to avoid the vanishing gradient issue. The complexity of the model and the likelihood of overfitting were both decreased by using dropout regularization after each GRU layer, which decreased neuron dependency. Y, the predicted water demand, was the end product. The whole model's process is shown in Figure 5.

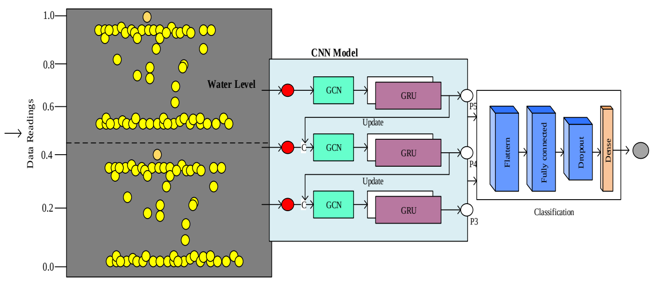

Figure 5: Urban water consumption prediction using GCN-ATT-GRU model

GCN Modeling

As a deep learning model specifically designed for graph data analysis, the graph convolutional network was introduced by Kipf and Welling in 2017. One major benefit of GCN is that it may adaptively change the way nodes are represented, making sure that every node takes nearby node information into account to an adequate degree. Its adaptability and effectiveness in handling graph data have led to its use in several fields, including as recommendation systems, natural language processing, and transportation networks with urban planning. A GCN architecture consists of input, hidden, as well as output layers. The characteristics of each node in this design change from X to Z, whilst the output is represented by Y. As you can see from Equations (5) and (6), the relevant calculation formulae are:

$$f(X_t^n, A) = Y, \quad (5)$$

$$A = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ a_{21} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{n-1n} \\ a_{n1} & \cdots & a_{nn-1} & 1 \end{bmatrix} \quad (6)$$

where X_t^n , where n is the total number of characteristics in the weather and water usage data, t is the period of the data, and Y is the result. Additionally, stands for the input time series. along with A symbol representing the matrix of adjacency that shows the relationships between nodes.

Multi-head attention

A deep learning-improved variant of the attention mechanism, the multi-head attention mechanism enhances the extraction of crucial information from sequential input. One common representation in traditional attention methods for the input sequence data is a single weighted context vector. A number of separate attention weights are used in the multi-head attention. With their own unique context vectors, these separate sets are capable of learning and capturing semantic information at varying depths. The result is obtained by joining these context vectors and then transforming them linearly. Q, Z, and X stand for the query, key, with value, respectively, which illustrates the architecture of the mechanism for multi-head attention.

This processing stage is shown in Formulas (6) and (7) for the multi-head attention mechanism:

$$\text{MultiHead}(Q, V, K) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W^H,$$
$$\text{head}_t = \text{Attention}\left(QW_t^Q, KW_t^K, VW_t^V\right), \quad (7)$$

where W_i^Q, W_i^K, W_i^V, and W^H symbolize the various head spaces, while head_i stands for the parameter matrix. After many heads have completed their computations, the results are combined using the Concat operation, which then undergoes a linear transformation.

Model parameters and settings

The learning rate, batch size, hidden unit count, and the amount of training iteration epochs are the four hyper-parameters of our suggested model. In the experiment, we used 32 batches and a learning rate of 0.001. It is recommended to conduct well-designed comparison experiments to identify the hidden units and the number of training iteration epochs, since these are critical factors that might impact the prediction accuracy. Each convolutional layer uses the ReLU as its activation, and the loss function is minimized using the Adam optimizer (Equation 8).

$$\text{loss} = \sum_{t=1}^{n} \left(Y_t - \hat{Y}_t\right)^2 / n \quad (8)$$

the predicted value, Y_t, and the actual measurement, $\hat{Y}_t$, are defined in relation to the length of the time patterns, n.

GCN-ATT-GRU Network

This part builds a GCN-ATT-GRU-based urban consumption interval forecasting framework to fully use the data's temporal and geographical properties. Firstly, GCN can extract geographically representative characteristics; secondly, GRU may process sequence information and successfully record the periodic variations in wind power output over time. Hence, a more thorough achievement of wind power interval forecasting may be accomplished by merging the built GCN-ATT-GRU network alongside the LUBE approach.

4 Results & discussion

The experimental tests were carried out using public datasets to test the method suggested in this research. The results of these tests were used to confirm the algorithm's correctness and efficacy.

4.1 Experiment preparation

Explanation of the experimental setting, data, and criteria for assessment is provided in this section. Figure 6 shows the GCN-ATT-GRU structure.

This section will perform a series of comparative tests comparing the suggested interval prediction technique of wind power with certain mainstream networks in order to validate its practicality and development. The open-source deep learning frameworks, Pytorch (version 1.7.1), Tensorflow (version 2.1.0), with Cuda (version 10.2) form the basis of all the experiments. In the experimental setting, we have an Intel Xeon I E5-2630 CPU, two Nvidia 1080Ti GPUs with 10 GB of RAM, and the Ubuntu 16.04.2 operating system.

In light of the aforementioned Pearson correlation finding, a GCN-ATT-GRU network-based interval prediction framework (see Table 3) for wind power is built using the suggested loss function Isoft_Loss. Their numbers include a Sigmoid function s softening factor of 80, a λ1 value of 0.0025, and a λ2 value that changes depending on the prediction step. More specifically, λ2 equals 0.1, 0.5, and 1 when the forecast time step is 1, 4, as well as 8, correspondingly. When training a model, the weights of the neural network are repeatedly updated using the Adam optimization process. The TPE algorithm optimizes a variety of GCN cells from 1 to 350. With a learning rate of 0.001, the iteration count is 60.

Table 3: Modifying the model's parameters.

Model	**Parameter**	**Range**	**Optimum values considered**
GCN-Att-GRU	Dropout	(0.2,0.3,0.4)	0.3
	GRU_hidden_size	(32,64,128)	128
	GCN_hidden_size	(132,64,128,256)	128
	GRU_num_layer	(2,3)	2
	Num_heads	(4,8,16)	8
	Lookback window	[7,10,14,20,30]	10
	GCN_out_size	(16,32,64)	16
	Learning rate	$[10^{-5},10^{-2}]$	10^{-3}
Att-GRU	Num_layers	(1,2,3)	1
	Hidden_size	(32,64,128)	64
	Lookback window	[7,10,14,20,30]	10
	Batch_size	(16,32,64)	16
	Learning rate	$[10^{-5},10^{-2}]$	10^{-3}
	Attention layers	(1-5)	5

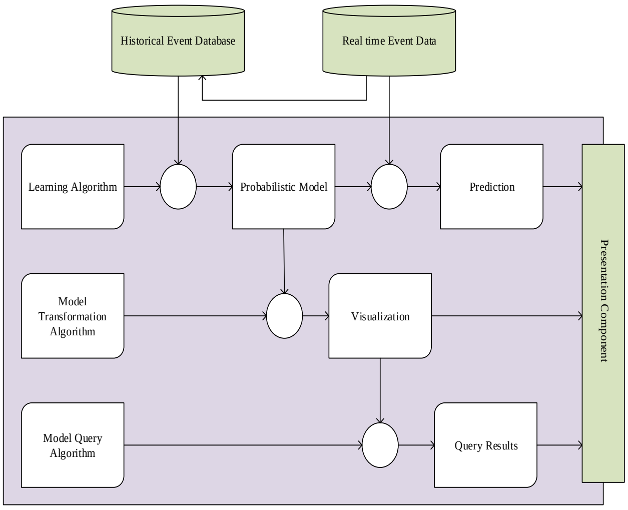

Figure 6: Modeling urban water consumption prediction

4.2 Evaluation index

In order to thoroughly assess the performance of the prediction models, this research included five important performance measures. Applications where substantial deviations are of concern are well-suited to the root-mean-square error (RMSE) because it emphasizes the significance of large mistakes. For cases when outliers don't make a difference, the mean absolute error (MAE) is useful. When accuracy in prediction is more important than absolute magnitude, the root-mean-square logarithmic error (RMSLE) is a good metric to utilize since it measures how well the ratio of expected and actual values matches up. When assessing relative mistakes, the mean absolute percentage error (MAPE) is a useful metric to use since it expresses prediction accuracy in percentage terms. As a measure of the model's explanatory capacity, the coefficient of determination (R^2) evaluates the predictability of the model in relation to the actual data. With each indicator illuminating a distinct facet of performance assessment, using them all together provides a holistic view of model performance. Here are the calculations for these indicators:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} (y_t - \hat{y}_t)^2}{n}},$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |\hat{y}_t - y_t|,$$

$$RMSLE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} \left(\ln(\hat{y}_t + 1) - \ln(y_t + 1)\right)^2},$$

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n} \left|\frac{y_t - \hat{y}_t}{y_t}\right|,$$

$$R^2 = 1 - \frac{\sum_t (\hat{y}_t - y_t)^2}{\sum_t (\bar{y}_t - y_t)^2},$$

where n depicts the total number of observations; $\hat{y}_i$ represents the expected value for the data point at index i; y_i agrees with the real worth; and $\bar{y}_l$ represents the average of all realistically measured values. Here is a table 4 where the **GCN-Att-GRU** model demonstrates a more significant performance advantage, and the newer GCN based models require more training time due to their complexity.

Table 4: Comparison of different prediction model results

Model	RMSE	MAE	RMSLE	MAPE (%)	R2	Epoch	Run Time for Each Epoch (s)

LSTM	0.05281	0.03966	0.03224	6.2832	0.85126	100	1.246
CNN-LSTM	0.04345	0.03304	0.02671	5.43400	0.87479	150	1.450
GCN-LSTM	0.04015	0.02800	0.02450	4.8000	0.88500	180	1.620
GCN-Att-LSTM	0.03600	0.02650	0.02100	4.1000	0.91000	190	1.750
GCN-Att-GRU	**0.02850**	**0.01950**	**0.01800**	**2.9500**	**0.94500**	200	1.880

Samples of the Predicted Data, Figure 7. Figure 8 shows the values of predictions. Model of the TSNE Visualization Embedding, seen in fig 9. Figure 10 shows the Accuracy in Relation to Epochs Fig 11 shows the Epochs vs Losses.

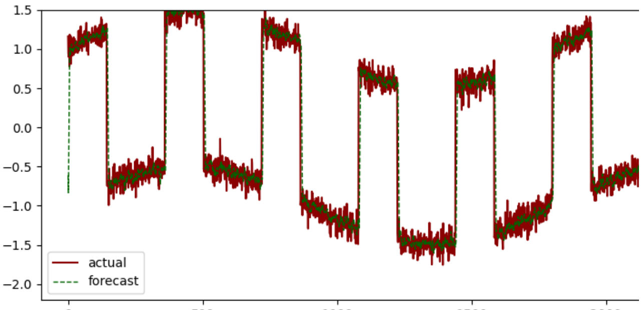

Figure 7: Predicted data samples (x axis – iterations, y axis – speed)

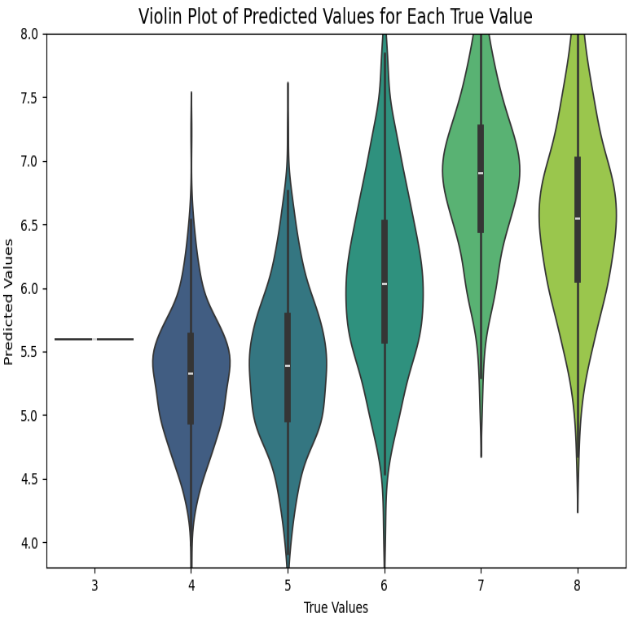

Figure 8: Predicted values

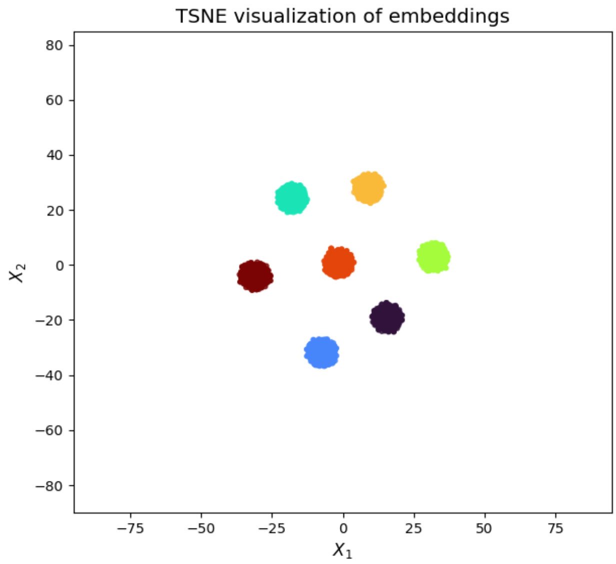

Figure 9: TSNE Visualization Embedding's Model

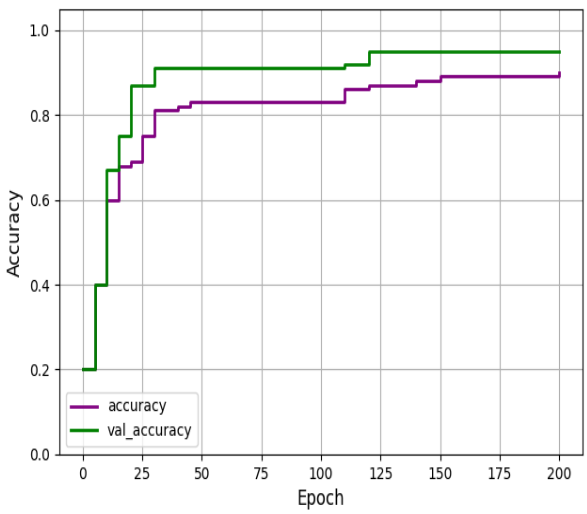

Figure 10. Accuracy vs. epochs

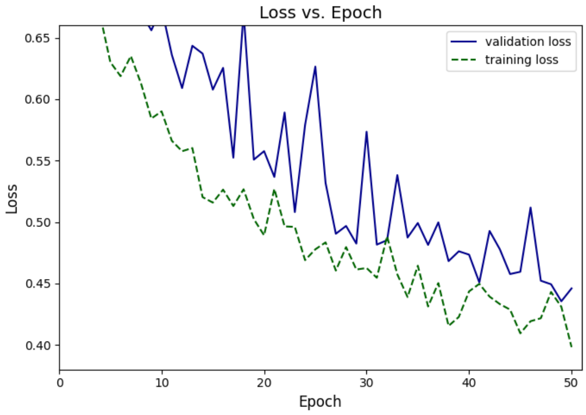

Figure 11: Loss vs. epochs

A wide range of prediction models, including simple linear models to intricate machine machine learning and deep learning techniques, are used in the process of estimating the amount of water that is consumed by metropolitan areas. For the purpose of estimating future water demand, these models make use of historical data on water usage, observations of weather patterns, information on population, and other pertinent aspects. The exact context, the availability of data, and the amount of information that is sought all play a role in determining which model to use and how accurate the forecasts turned out to be. Taking into consideration data preparation, the creation of features, and the measurement of prediction accuracy using proper metrics are all very significant aspects to take into consideration.

5 Conclusion

Planning for sustainable urban infrastructure and accurately predicting urban water demand are two of the most important aspects of environmental protection. Optimizing water distribution, improving resource efficiency, and reducing shortages all depend on accurate short-term projections. When trying to anticipate urban water demand, traditional models often fail because they focus only on the temporal aspects of the relationships between variables. In order to address the issue of short-term urban water demand forecasting, this research proposes the GCN-Att-GRU model. To provide better predictions of future urban water demands, this model takes into account both the geographical and temporal aspects of the relationships between variables.

The capacity of predictive models to perform and generalize is directly impacted by the quality of the data. The data quality of this research is improved since the Prophet model is used for anomaly detection as well as correction. The seasonal and cyclical character of water usage series is well captured by this model. Furthermore, the model leverages the graph convolutional network for obtaining spatial features, the multi-head attention system to amplify water-related features, the largest possible coefficient to look at the relationship among water consumption while weather data, and both temporal and spatial features. As a last step, we investigate the underlying characteristics of time series information with depth using an GRU model. Experiments showed that GCN-Att-GRU model reduced MAPE by 1.868-2.718%, making it more accurate than conventional GRU and CNN-GRU models. Research using ablation techniques confirmed the need of combining spatial and temporal characteristics, and it also provided greater clarity about the role of each model component.

This study presents the GCN-Att-GRU model, which not only shows how successful it is to combine many sophisticated methodologies, in addition to providing fresh perspectives and approaches to forecasting city water needs. Through the integration of spatial and temporal factors among variables, it enhances the precision of water demand forecasts, making it easier to understand complicated patterns in water usage data. Nevertheless, the GCN-Att-GRU is not easily interpretable due to its black-box aspect, which is common among sophisticated models.

Future research directions

If we want our model to be more accurate, resilient, and predictive, we need to do more study on it and incorporate more impacting elements.The GCN-Att-GRU model will be further developed and improved in three main areas in order to increase its operational resilience and capability in future research:

1. Improving Spatio-Temporal Feature Integration: Moving forward, our emphasis will be on incorporating exogenous aspects that are dynamic and predictive, rather than just adding more features. Among these changes is the substitution of short-term weather predictions for historical weather in the GCN's feature vector and the incorporation of real-time pressure sensor data as a dynamic feature impacting demand. This could improve the model's predictive power beyond its present 7.60% MAPE by allowing it to account for the causal impacts associated with network dynamics and future occurrences.

2. Despite the fact that the MIC-based adjacency (AMIC) has enhanced performance, the graph topology is still static after training in the current iteration of the adaptive adjacency matrix construction. Within the GCN layers, future studies will investigate dynamic graph learning methods, such as using a learnable metric function or a self-attention mechanism. To make the model more resilient to network anomalies, the spatial connections (A) might vary in real-time to account for things like developing consumption patterns or unexpected changes to the network architecture (such pipe bursts or maintenance).

3. Quantifying Uncertainty and Probabilistic Forecasting: Point forecasts, such as Mean Absolute Error or Root Mean Squared Error, are inadequate for critical infrastructure management. Using a Bayesian Neural Network or a Quantile Regression Loss, for example, future work will expand the regression output layer to a probabilistic framework. System operators will have a critical metric of predictive uncertainty for improved risk assessment and resource allocation thanks to the intervals the model can provide for predictions, such as 95% confidence limits.

References

[1] Mohamed Hussain, K., Sivakumaran, N., Radhakrishnan, T.K., Swaminathan, G., & Sankaranarayanan, S. (2024). Short-term water demand prediction using stacking regression–based machine learning. Water Practice & Technology. https://doi.org/10.2166/wpt.2024.292.

[2] Sahoo, B.B., Panigrahi, B., Nanda, T., Tiwari, M.K., & Sankalp, S. (2023). Multi-step Ahead Urban Water Demand Forecasting Using Deep Learning Models. SN Computer Science, 4. DOI:10.1007/s42979-023-02246-6

[3] Rezaali, M., Fouladi-Fard, R., & Karimi, A. (2023). Performance of TANN, NARX, and GMDHT Models for Urban Water Demand Forecasting: A Case Study in a Residential Complex in Qom, Iran. Avicenna Journal of Environmental Health Engineering. doi: 10.34172/ajehe.5283

[4] Jia, Z., Li, H., Yan, J., Sun, J., Han, C., & Qu, J. (2023). Dynamic Graph Convolution-Based Spatio-Temporal Feature Network for Urban Water Demand Forecasting. Applied Sciences. https://doi.org/10.1016/j.jpowsour.2025.236524

[5] Pu, Z., Yan, J., Chen, L., Li, Z., Tian, W., Tao, T., & Xin, K. (2022). A hybrid Wavelet-CNN-GRU deep learning model for short-term urban water demand forecasting. Frontiers of Environmental Science & Engineering, 17. DOI:10.1007/s11783-023-1622-3

[6] Guo, J., Sun, H., & Du, B. (2022). Multivariable Time Series Forecasting for Urban Water Demand Based on Temporal Convolutional Network Combining Random Forest Feature Selection and Discrete Wavelet Transform. Water Resources Management, 36, 3385-3400. DOI:10.1007/s11269-022-03207-z

[7] Patil, R., Alandikar, P., Chaudhari, V., Patil, P., & Deshpande, P.S. (2022). Water Demand Prediction Using Machine Learning. International Journal for Research in Applied Science and Engineering Technology. DOI:10.22214/ijraset.2022.47797

[8] Li, W., & Wang, X. (2023). Analysis and prediction of urban household water demand with uncertain time series. Soft Computing, 1-8. DOI:10.1007/s00500-023-09476-z

[9] Wu, S., Zhao, P., Bai, M., Wang, J., & Lan, Y. (2021). A Fine Grained Forecasting Method for Minute-Level Urban Water Demand. 2021 40th Chinese Control Conference (CCC), 8298-8303. DOI:10.23919/CCC52363.2021.9550182

[10] He, K., Liu, Y., Yuan, J., He, Z., Yin, Q., Xu, D., Zhao, X., Hu, M., & Lu, H. (2024). Water quality ensemble prediction model for the urban water reservoir based on the hybrid long short-term memory (GRU) network analysis. AQUA — Water Infrastructure, Ecosystems and Society. DOI:10.2166/aqua.2024.099

[11] Wang, D., Zhang, Y., & Yousefi, N. (2024). Urban Water-Energy consumption Prediction Influenced by Climate Change utilizing an innovative deep learning method. Scientific Reports, 14. DOI:10.1038/s41598-024-81836-7

[12] Li, J., & Song, S. (2023). Urban Water Consumption Prediction Based on CPMBNIP. Water Resources Management, 37, 5189-5213. DOI:10.1007/s11269-023-03601-1

[13] Wang, K., Ye, Z., Wang, Z., Liu, B., & Feng, T. (2023). MACLA-GRU: A Novel Approach for Forecasting Water Demand. Sustainability. https://doi.org/10.3390/su15043628

[14] Groppo, G.D., Costa, M.A., & Libânio, M. (2023). Predicting time series for water demand in the big data environment using statistical methods, machine learning and the novel analog methodology dynamic time scan forecasting. Water Supply. DOI:10.2166/ws.2023.008

[15] Liu, X., Zhang, Y., & Zhang, Q. (2022). Comparison of EEMD-ARIMA, EEMD-BP and EEMD-SVM algorithms for predicting the hourly urban water consumption. Journal of Hydroinformatics. DOI:10.2166/wst.2024.028

[16] Lučin, I., Lučin, B., Čarija, Z., & Sikirica, A. (2021). Data-Driven Leak Localization in Urban Water Distribution Networks Using Big Data for Random Forest Classifier. Mathematics. https://doi.org/10.3390/math9060672

[17] Sun, J., Guan, X., Sun, X., Cao, X., Tan, Y., & Liao, J. (2024). Water quality prediction and carbon reduction mechanisms in wastewater treatment in Northwest cities using Random Forest Regression model. Scientific Reports, 14. DOI:10.1038/s41598-024-83277-8

[18] Moon, J., Lee, J., Lee, S., & Yoon, H. (2022). Urban River Dissolved Oxygen Prediction Model Using Machine Learning. Water. https://doi.org/10.3390/w14121899

[19] Fu, L., Wang, J., Wang, S., Wang, H., & Hu, G. (2025). Analysis and prediction of water resources consumption: A case study in Yiwu city, China. Journal of Physics: Conference Series, 2958. DOI:10.1088/1742-6596/2958/1/012005

[20] Hou, Y., Zhang, A., Lv, R., Zhang, Y., Ma, J., & Li, T. (2023). Machine learning algorithm inversion experiment and pollution analysis of water quality parameters in urban small and medium-sized rivers based on UAV multispectral data. Environmental Science and Pollution Research, 30, 78913-78932. ; https://doi.org/10.3390/w16223163

[21] Liu, C., Liu, Z., Yuan, J., Wang, D., & Liu, X. (2024). Urban Water Demand Prediction Based on Attention Mechanism Graph Convolutional Network-Long Short-Term Memory. Water. DOI:10.3390/w16060831

[22] Zubaidi, S.L., Ortega-Martorell, S., Al-Bugharbee, H., Olier, I., Hashim, K.S., Gharghan, S.K., Kot, P., & Alkhaddar, R.M. (2020). Urban Water Demand Prediction for a City That Suffers from Climate Change and Population Growth: Gauteng Province Case Study. Water. https://doi.org/10.3390/w12071885

[23] Iwakin, O.M., & Moazeni, F. (2024). Improving urban water demand forecast using conformal prediction-based hybrid machine learning models. Journal of Water Process Engineering. DOI:10.1016/j.jwpe.2023.104721

[24] Yan, D., Tao, Y., Zhang, J., & Yang, H. (2024). Research on urban water demand prediction based on machine learning and feature engineering. Water Supply. https://doi.org/10.2166/ws.2024.157

[25] Sun, Z., Zhang, Y., & Wu, X. (2024). Visual analysis of urban water demand prediction research based on CiteSpace. Proceedings of the 2024 4th International Conference on Artificial Intelligence, Automation and High Performance Computing. https://doi.org/10.3390/land14112100

[26] Xia, S., Zhu, H., Zhang, N., Li, Y., & Zhou, C. (2024). A dual-feature channel deep network with adaptive variable weight reconstruction for urban water demand prediction. Sustainable Cities and Society. DOI:10.1016/j.jwpe.2023.104721

[27] Saradha, K., Sakthy S, S., A, A., & M, D.R. (2025). Urban Water Demand Prediction System Using GRU. 2025 International Conference on Computing and Communication Technologies (ICCCT), 1-6. DOI:10.1109/ICCCT63501.2025.11020461

[28] Zubaidi, S.L., Al-Bugharbee, H., Ortega-Martorell, S., Gharghan, S.K., Olier, I., Hashim, K.S., Al-Bdairi, N.S., & Kot, P. (2020). A Novel Methodology for Prediction Urban Water Demand by Wavelet Denoising and Adaptive Neuro-Fuzzy Inference System Approach. Water. ; https://doi.org/10.3390/w12061628

[29] Hu, P., Tong, J., Wang, J., Yang, Y., & Turci, L.D. (2019). A hybrid model based on CNN and Bi-GRU for urban water demand prediction. 2019 IEEE Congress on Evolutionary Computation (CEC), 1088-1094. DOI:10.1109/CEC.2019.8790060

[30] Zubaidi, S.L., Gharghan, S.K., Dooley, J., Alkhaddar, R.M., & Abdellatif, M. (2018). Short-Term Urban Water Demand Prediction Considering Weather Factors. Water Resources Management, 32, 4527-4542. DOI: 10.1007/s11269-018-2061-y