Forecasting Water Demand With the Long Short-Term Memory Deep Learning Mode

Junhua Xu, Zhejiang Tongji Vocational College of Science and Technology, China*

ABSTRACT

Traditional methods often fall short in modeling the nonlinear, seasonally variable nature of urban water demand. This proposed solution is an integrated ARIMA-LSTM deep learning model, combining ARIMA's proficiency in linear trend and seasonal modeling with LSTM's strength in capturing nonlinear time dependencies. In these experiments, the authors trained and evaluated using daily water demand data from 2015 to 2020, with its performance validated for the year 2021. The ARIMA-LSTM model demonstrates promising results, outperforming individual models in terms of accuracy. In validation, it achieves a high coefficient of determination (R^2) of 0.98 and a significantly low root mean square error (RMSE) of 2.94. These metrics indicate an excellent fit to the data and a high level of precision in its predictions. The significance of this research lies in its potential to advance the field of urban water demand forecasting, ultimately contributing to better water resource management and sustainability in urban areas.

KEYWORDS

ARIMA, Deep Learning, LSTM, Urban Water Demand, Water Resource Management

INTRODUCTION

The rapid expansion of urban areas underscores the increasing necessity for intelligent urban water supply systems, which are crucial for the development of modern smart cities (Wu et al., 2018). A key component of such systems is the prediction of urban water demand, a task that is both significant and challenging. Accurate predictions of water demand are essential for optimal operation of valves and pumps in waterworks and for identifying potential leakages in the supply network. Effective water demand forecasting can not only lower water supply costs but also ensure that urban water needs are met. However, few models are practically applicable to urban water demand prediction. This task is complex due to the multifaceted influencing factors and inherent unpredictability of urban water demand (Du et al., 2021).

The accurate forecasting of urban water demand is a critical aspect of sustainable urban development and efficient resource management. As urban populations grow and climate patterns

DOI: 10.4018/IJITSA.338910
*Corresponding Author

change, the need for robust forecasting models becomes even more pronounced. In recent years, the integration of machine learning and deep learning techniques in this field has shown promising results, leading to more accurate, adaptive, and efficient forecasting models (Khozani et al., 2022; Xu et al., 2022). Urban water demand forecasting is not just a technical challenge but also a key component in addressing broader issues such as water conservation, infrastructure development, and environmental sustainability (Butler & Memon, 2005; Donkor et al., 2014). Accurate forecasting helps in reducing wastage, optimizing distribution, and planning for future needs, particularly in the context of climate change and urban expansion (Zubaidi et al., 2020). The deep learning models commonly used in the field are as follows (Adamowski & Karapataki, 2010): Linear Regression models are the most basic form of predictive analysis and are used for their simplicity and interpretability (Pulido-Calvo et al., 2007; Sebri, 2016). However, they often fail to capture complex nonlinear relationships inherent in water demand data. ARIMA (auto-regressive integrated moving average) is a popular traditional statistical model for time series forecasting, which is effective for data with a clear trend or seasonal pattern but struggles with non-linear data (Pulido-Calvo et al., 2007). Support vector machines (SVMs) are effective for classification and regression challenges. SVMs work well with high-dimensional data but can be computationally intensive and less effective with very large datasets (Peña-Guzmán et al., 2016). The random forest is an ensemble learning method that is robust to overfitting and can handle both linear and nonlinear data (Chen et al., 2017). However, it can be less interpretable and may require a significant amount of data for optimal performance. The LSTM (long short-term memory network) is a type of recurrent neural network (RNN) particularly well-suited for time series data due to its ability to capture long-term dependencies. However, LSTMs can be complex to tune and train and may require substantial computational resources (Kühnert et al., 2021).

Banjac et al. (2015) developed a water demand prediction system utilizing ANNs, which was a significant step towards understanding the complex dynamics of water demand series. Further, Yin et al. (2018) demonstrated the superiority of ANNs over traditional statistical models in water-energy demand forecasting, highlighting the enhanced predictive accuracy and capability of ANNs in handling complex datasets. Additionally, other machine learning methods have also been explored and applied in the field of water demand forecasting, signifying a shift towards more sophisticated, data-driven approaches in this area. However, these methods, particularly those with shallow architectural designs commonly found in classic intelligent models, face limitations in efficiently handling large-scale data. This restriction hinders their ability to effectively mine and interpret complex data features, presenting a significant challenge in the field of predictive analytics, where understanding and leveraging intricate data patterns is crucial (Du et al., 2021).

Given the limitations of traditional models in capturing the complexities of urban water demand, there is a clear need for models that can assimilate both linear and nonlinear patterns efficiently. This led to the proposal of combining ARIMA and LSTM into a cohesive model, aiming to utilize the strengths of both methods: ARIMA's efficiency in modeling linear aspects and seasonality and LSTM's prowess in capturing complex, long-term nonlinear dependencies. The ARIMA-LSTM model is designed as a hybrid approach, where initially, ARIMA is employed to model and remove the linear trends and seasonality from the time series data. This processed data is then fed into an LSTM network to model the residual (non-linear) components. The ARIMA component handles the linear and seasonal aspects of the time series data, making the residuals more stationary and easier for the LSTM to model. The LSTM component focuses on the non-linear patterns in the residuals, leveraging its recurrent nature and long-term memory capabilities. By combining these two approaches, the ARIMA-LSTM model aims to provide a more comprehensive and accurate forecast of urban water demand, addressing both short-term fluctuations and long-term trends. This hybrid approach is expected to outperform traditional models, especially in scenarios with complex demand patterns influenced by multiple factors such as weather, urban activities, and policy changes.

The main contribution of this article lies in introducing an integrated approach that combines the traditional ARIMA time series model with the deep learning LSTM model. This integration harnesses the strengths of ARIMA in capturing seasonality and cyclical patterns in time series data and combines them with LSTM's ability to model long-term dependencies. As a result, the proposed model achieves improved prediction accuracy for urban water demand forecasting. Additionally, the article provides insights into data preprocessing and optimization techniques used to enhance the model's performance. These techniques include appropriate data differencing and standardization to mitigate seasonality and trends, ultimately improving the model's predictive capabilities. The empirical study conducted in the article showcases the practical effectiveness of the integrated ARIMA-LSTM model by applying it to real-world urban water demand data. This empirical evidence underscores the value of the proposed approach for urban planners and water resource managers in addressing the complexities and uncertainties of water resource demands in urban settings. Furthermore, the versatility of the integrated ARIMA-LSTM method is emphasized, as it can be applied to various time series forecasting tasks beyond urban water demand. This includes domains such as energy demand prediction and stock market price forecasting, expanding the applicability and relevance of the approach.

RELATED WORK

Autoregressive Integrated Moving Average

The ARIMA model plays a significant role in the field of forecasting urban water demand due to its ability to capture and model time series data effectively. In this context, ARIMA is applied as follows. ARIMA captures the temporal patterns and trends in historical urban water consumption data. It is particularly effective in identifying seasonality and cyclic behavior, which are essential components of water demand patterns (Kofinas et al., 2014; Oliveira et al., 2017; Pandey et al., 2021). For example, Liao et al. used an ARIMA model to predict water supply and demand in Shandong Province over the next 15 years. Oliveira et al. (2017) used a bi-quadratic ARIMA model to predict a single day's water demand in a district metering area (DMA).

By differencing the data (i.e., subtracting the previous value from the current value) to achieve stationarity, ARIMA can then model these temporal dependencies. One of the notable advantages of ARIMA in this domain is its interpretability. It provides insights into the underlying factors contributing to water demand fluctuations, such as daily, weekly, or monthly patterns. This interpretability is valuable for urban planners and policymakers as it allows them to understand the drivers of water demand changes. However, ARIMA also has its limitations. It may struggle to capture long-term dependencies and non-linear relationships in the data, which are characteristics often present in complex urban water demand scenarios. Additionally, ARIMA models can be sensitive to outliers and require careful preprocessing of the data to ensure accurate results. In conclusion, ARIMA models are valuable tools for forecasting urban water demand, particularly in capturing temporal patterns and seasonality. Their interpretability makes them useful for gaining insights into water consumption behavior. However, their limitations include difficulty in modeling long-term dependencies and susceptibility to outliers, which may necessitate integrating other models like LSTM for more accurate and comprehensive predictions in this field. It primarily relies on historical data and normal distribution assumptions to predict future trends. Despite improvements, such as integrating seasonal factors in seasonal ARIMA models, these methods still face limitations in handling complex and nonlinear time series. Particularly in predicting water demand series with stochastic characteristics, their accuracy may be insufficient. Therefore, while ARIMA models have been foundational and instructive in the field of water demand forecasting, their effectiveness in complex real-world scenarios remains limited.

Long Short-Term Memory Networks

In the field of forecasting urban water demand, the LSTM deep learning model has been applied in a detailed manner, offering both advantages and limitations. LSTM has found extensive utility in this domain due to its remarkable ability to capture temporal dependencies and patterns within time series data (Kühnert et al., 2021; Nasser et al., 2020). It excels in modeling sequences of data, making it well-suited for forecasting tasks where historical water consumption patterns play a crucial role. By analyzing past consumption data and considering factors such as seasonality, trends, and cyclic patterns, LSTM can provide accurate and context-aware predictions of future urban water demand. One notable advantage of LSTM is its ability to handle long-term dependencies, which are common in water demand data influenced by factors like climate changes and population growth. Unlike traditional statistical models, LSTM can learn and adapt to intricate relationships between historical data points, enabling it to make more precise forecasts in dynamic urban environments.

However, LSTM also comes with some limitations. It can be computationally intensive and require substantial amounts of data for training to prevent overfitting (Lee et al., 2019). Additionally, LSTM models may struggle when faced with noisy or incomplete data, which can be a challenge in real-world urban water demand datasets. Furthermore, interpreting the inner workings of LSTM models can be complex, making it challenging to provide transparent explanations for forecasting results, which may be a concern in decision-making processes. LSTM's detailed application in forecasting urban water demand capitalizes on its strengths in capturing temporal dependencies and modeling intricate relationships within time series data. Its ability to handle long-term dependencies is a valuable asset in this field. However, it is essential to consider the computational demands and challenges associated with noisy data when deploying LSTM models for water demand forecasting.

The ARIMA-LSTM Hybrid Model

The integrated ARIMA-LSTM model plays a pivotal role in the field of urban water demand forecasting, offering a comprehensive and effective approach. This hybrid model combines the strengths of two distinct methodologies, ARIMA and LSTM, to address the complexities of urban water demand prediction. One of the primary advantages of integrating ARIMA and LSTM is the improved prediction accuracy it brings to the domain of urban water demand forecasting. ARIMA excels at capturing seasonality, cyclical patterns, and short-term dependencies in time series data. In contrast, LSTM is proficient in modeling long-term dependencies and capturing complex temporal relationships. By combining these capabilities, the integrated model can provide highly accurate predictions that consider both short-term fluctuations and long-term trends in water demand. Another notable benefit of the ARIMA-LSTM model is its adaptability to diverse data patterns commonly encountered in urban water demand forecasting. It can effectively handle data with multiple variables, including historical water consumption, meteorological factors, and demographic information. This versatility allows the model to capture the multifaceted factors influencing water demand in urban areas.

MATERIALS AND METHODS

Overview

In pursuing accurate urban water demand forecasting, our research introduces a cutting-edge approach: an integrated ARIMA-LSTM deep learning model. This model showcases distinctive characteristics, leverages core technologies, and revolutionizes the landscape of water demand prediction. Our model stands out with its ability to combine the best of both models: the traditional strength of ARIMA and the cutting-edge capabilities of LSTM neural networks. It marries the temporal modeling expertise of ARIMA with the sequential data prowess of LSTM, resulting in precise and adaptable forecasting. In addition to utilizing the ARIMA-LSTM model for predictions, we conducted separate forecasts using both the ARIMA and LSTM models and compared the results through evaluation.

At the heart of our model lies the synergy between ARIMA and LSTM. ARIMA excels at capturing temporal patterns and seasonality, while LSTM specializes in modeling intricate dependencies in data. ARIMA's output can be effectively integrated into LSTM, combining the strengths of both models for enhanced time series forecasting. Forecasting water demand using an integrated ARIMA-LSTM deep learning model is a complex process that involves several stages:

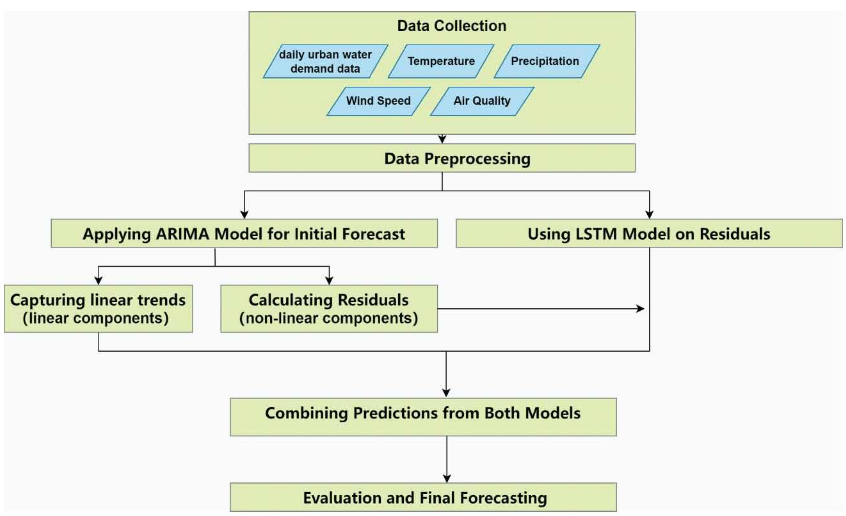
1. **Data collection:** daily meteorological data (e.g., temperature, precipitation, wind speed), air quality index, and water demand data (2015–2021).
2. **Data processing:** cleaning the data (handling missing values, removing outliers), and possibly transforming it (normalization, scaling) to make it suitable for use in the models.
3. **Model training and testing:** using three models—ARIMA, LSTM, and an integrated ARIMA-LSTM. Using data from 2015 to 2020 to train the models and a loss function that measures the accuracy of your forecasts. Common choices include mean squared error (MSE) or mean absolute error (MAE). Use optimization algorithms (Adam) to update the parameters of your models, especially for the LSTM and integrated model. The ARIMA model's parameters are usually determined during the model fitting process. The models are then tested using data from the year 2021 to evaluate their forecasting accuracy.
4. **Model evaluation and comparison:** Comparing the performance of ARIMA, LSTM, and ARIMA-LSTM models using metrics like RMSE, MAE, or others. Evaluate based on how well they capture trends and seasonality and react to external variables like weather data (see Figure 1).

ARIMA

The ARIMA model is a fundamental time series forecasting method employed in predicting urban water demand. It plays a pivotal role in capturing temporal patterns and seasonality, making it an integral component of the water demand forecasting methodology. At its core, the ARIMA model operates on the principle of differencing the time series data to make it stationary. The model comprises three primary components: the autoregressive, integrated, and moving average components.

- **Autoregressive (AR) component:** models the relationship between the current observation and past observations at lag intervals (Shumway et al., 2017). It quantifies how the current water demand depends on its own previous values.

Figure 1. Overview of the framework



- **Integrated (I) component:** involves differencing the time series data to make it stationary. This step eliminates trends and seasonality, ensuring that the model captures the stationary component of the data.
- **Moving average (MA) component:** models the relationship between the current observation and a lagged moving average of past observations. It accounts for the influence of past white noise (error terms) on the current value (Liang et al., 2022).

where represents the observed water demand at time *t*; represents a constant term; are autoregressive coefficients; are lagged values of the water demand time series; represents the differenced time series after applying the integrated component; are moving average coefficients; and represents the error term at time *t*.

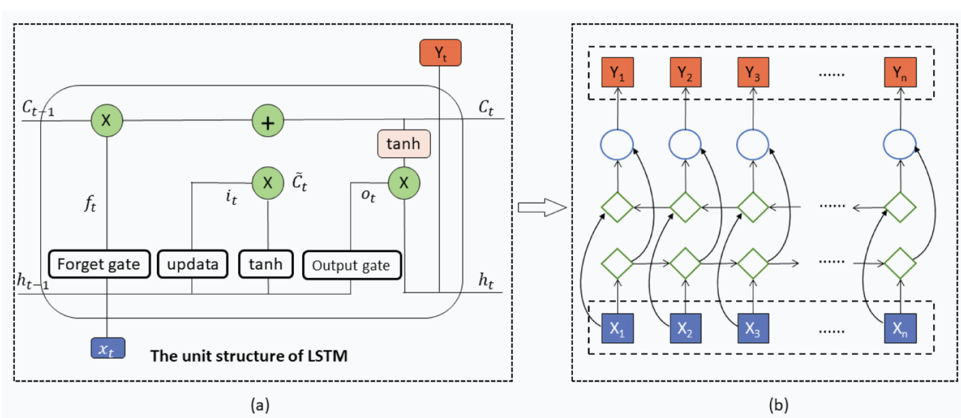
The ARIMA model plays a critical role in capturing temporal patterns and seasonality in urban water demand data by employing autoregressive, integrated, and moving average components. It is a foundational element of the water demand forecasting methodology, facilitating accurate predictions.

LSTM

LSTM networks (a subset of RNNs) are particularly adept at handling time-series data (Hochreiter & Schmidhuber, 1997). RNNs are known for their directed cycle structure, which allows output transfer from one layer to the same layer in the next cycle. This structure facilitates the identification of time-series features. However, RNNs often face the vanishing gradient problem, leading to reduced accuracy. LSTM networks address this issue with a unique architecture comprising gating mechanisms and cell states. These mechanisms include three different logic gates—the forget gate, input gate, and output gate—alongside cell states connected to each element (see Figure 2). A series of mathematical equations govern the operation of these gates and the updating of cell states (Olah, 2015). The LSTM network effectively captures temporal dynamics, making it ideal for predicting variables that change over time, such as water levels and quality. The bias and weight matrices within the LSTM network are crucial components, enabling the network to learn and adapt based on the input data. The flexibility and robustness of LSTM in handling time-related data make it a powerful tool in environmental modeling and prediction tasks (Yamak et al., 2019).

The LSTM model comprises several key components, each with its set of equations and variables, as follows.

Figure 2. LSTM structure



Input Gate

The input gate *i* determines which information from the current input should be stored in the cell state.

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \tag{1}$$

where i_t is the output of the input gate at time step *t;* $\tilde{C}_t$ is the candidate value for the cell state; W_i and W_C are the weights of the input gate and candidate value, respectively; and b_i and b_C are the biases of the input gate and candidate value, respectively.

Forget Gate

The forget gate *f* decides what information from the cell state should be discarded or kept.

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \tag{2}$$

$$\tilde{C}_t = \tanh\left(W_C \cdot \left[h_{t-1}, x_t\right] + b_C\right) \tag{3}$$

where f_t is the output of the forget gate at time step *t*; σ represents the sigmoid function. are the weights of the forget gate; W_f is the output from the previous time step; h_{t-1} is the input at the current time step; x_t is the bias of the forget gate; and b_f is the bias of the forget gate.

Cell State Update

Cell state *c* stores information over time where C_t is the cell state at time step *t* and $*$ denotes element-wise multiplication.

Output Gate

The output gate *o* determines the next hidden state and the output prediction:

$$\begin{gathered} o_t = \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right) \\ h_t = o_t * tanh\left(C_t\right) \end{gathered} \tag{4}$$

where o_t is the output of the output gate at time step *t*; h_t is the final output of the LSTM unit at time step *t*; W_o is the weight of the output gate; b_o is the bias of the output gate; σ denotes the sigmoid function; and $\tanh$ is the hyperbolic tangent function. These equations collectively describe how information flows through an LSTM unit. These components and equations collectively enable the LSTM model to capture complex temporal dependencies and evolving trends within the urban water demand time series data. The variables and activations within the LSTM cells are adjusted during training to optimize the model's predictive capabilities (Zheng et al., 2021).

ARIMA-LSTM

In time series forecasting, the distinction between linear and nonlinear models is crucial. Linear models are adept at identifying and working with linear patterns within time series data, but they fall short when it comes to recognizing and mining nonlinear relationships. Nonlinear models, on the other hand, excel in this area. Research and practical applications. Using a single model, whether linear or nonlinear, can be effective when dealing with time series data that predominantly consist of either linear or nonlinear components (Xu et al., 2022). However, these models encounter limitations when faced with more complex problems where the time series data contain a mix of both linear and nonlinear components. In such cases, relying solely on a single type of model, be it linear or nonlinear, is insufficient for accurate and effective forecasting.

In the integrated ARIMA-LSTM model for forecasting urban water demand, the process begins with the ARIMA model. The ARIMA model is first employed to forecast the time series trends, capturing and modeling linear trends and seasonal patterns inherent in the data (Dave et al., 2021; Xu et al., 2022). This step hinges on the ARIMA model's capability to analyze historical data and project it into the future based on identified patterns. Once the ARIMA model has made its predictions, the next step involves calculating the residuals, or errors, which are the differences between the ARIMA model's predictions and the actual data. These residuals are crucial because they represent the aspects of the data that the ARIMA model, typically adept at handling linear relationships, has failed to capture or explain. This is where the LSTM model comes into play. The LSTM, a type of recurrent neural network, is particularly adept at capturing non-linear patterns and dependencies in data, especially those that unfold over time. By applying the LSTM model to the residuals generated by the ARIMA model, it effectively addresses and models the non-linear patterns that were not accounted for by the ARIMA model. The final step in the forecasting process is the combination of predictions, and formulas are as follows:

$$Y_t = L_t + A_t \tag{5}$$

where Y_t is a combination of linear L_t and nonlinear A_t parts.

The prediction from the ARIMA model, representing the linear trends and seasonality, is added to the prediction of the residuals from the LSTM model. This integration results in the final forecast, which encompasses both the linear and non-linear aspects of the historical water usage data, thereby providing a more comprehensive and accurate prediction of future urban water demand. This integrated ARIMA-LSTM approach effectively harnesses the strengths of both models: ARIMA's proficiency in linear analysis and LSTM's capability in capturing complex, non-linear relationships, resulting in a robust forecasting tool (Fan et al., 2021; Pierre et al., 2023).

EXPERIMENT

Experimental Design

In this study, three experiments were designed to forecast urban water demand: the ARIMA model, the LSTM model, and an integrated ARIMA-LSTM model. All experiments utilized daily water usage data and corresponding meteorological elements (i.e., temperature, precipitation, wind speed, and air quality index) from 2015 to 2022 as inputs. The dataset from 2015 to 2020 was used for training the models, while the data for 2021 was employed for validation. The ARIMA-LSTM model initially employed the ARIMA model to predict time series trends, capturing and modeling linear trends and seasonal patterns. Subsequently, the residuals, which are the differences between the ARIMA model's predictions and the actual data, were calculated. The LSTM model was then applied to these residuals to predict non-linear patterns that the ARIMA model failed to explain. Finally, the predictions from the ARIMA model were combined with the residuals predicted by the LSTM model to obtain the final forecast. This integrated approach aimed to leverage the strengths of both ARIMA and LSTM models for a more accurate prediction of urban water demand.

In this study, the experimental environment was configured to support efficient data processing and model training. A high-performance computer from the AMD Ryzen series was used, equipped with NVIDIA Tesla series GPUs to accelerate deep learning tasks. The software environment included the Python programming language, combined with Pandas and NumPy libraries for efficient data processing. For machine learning, Statsmodels was used to implement the ARIMA model, and TensorFlow for building and training the LSTM model. Matplotlib was utilized for data analysis and result presentation, and integrated development environments such as PyCharm and Jupyter Notebook were used for code writing and debugging.

In this study, specific parameter settings were established for the three models. For the ARIMA model, we chose a differencing order of $d = 1$ to ensure time series stationarity, an autoregressive term of $p = 2$ to capture long-term dependencies in the data, and a moving average term of $q = 2$ to smooth the forecast errors. For the LSTM model, a single hidden layer with 50 neurons was set. The learning rate was fixed at 0.001 to optimize the training process, with a batch size of 32 and an iteration count (Epochs) of 100. As for the integrated ARIMA-LSTM model, we applied the same ARIMA parameter settings and replicated the LSTM configuration. Additionally, data input to the LSTM model was normalized to ensure effective training and forecasting. With these carefully chosen parameter settings, our aim was to achieve accurate predictions of urban water demand.

In evaluating the test results, six performance metrics were used: mean absolute percentage error (MAPE), root mean square error (RMSE), mean absolute error (MAE), Mean squared error (MSE), root mean squared logarithmic error (RMSLE), and the coefficient of determination (R^2). The performance metrics for the forecast results were calculated using the following formulas, where n is the number of data points, Y_i is the actual value, $\hat{Y}_i$ is the predicted value, and $\overline{Y}$ is the mean of the actual observations:

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \widehat{Y}_i}{Y_i}\right| \times 100\% \tag{6}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2} \tag{7}$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2 \tag{8}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \widehat{Y}_i\right| \tag{9}$$

$$\mathrm{RMSLE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\log\left(Y_i + 1\right) - \log\left(\hat{Y}_i + 1\right)\right)^2} \tag{10}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2} \tag{11}$$

These metrics provide a comprehensive assessment of the model's forecasting accuracy and error magnitude, enabling a thorough evaluation of the model's performance in predicting urban water demand.

Dataset

The focus of this study is a reservoir located in Yu County, Shanxi Province, China. It is located in the southern part of the mountainous basin in the northwest of Hebei Province, belongs to the Yongding River basin, and is endowed with abundant water resources. The total water resources of Yu County amount to 198 million cubic meters. The surface water resource volume is 147 million cubic meters, with a utilizable amount of 110 million cubic meters. The groundwater resource volume is 108 million cubic meters, with an extractable amount of 82 million cubic meters. The annual average industrial and agricultural water usage in the county is about 100 million cubic meters, leaving nearly 100 million cubic meters of water resources available for utilization. The Huli River Reservoir, located in the western part of the county, is a medium-sized reservoir with a capacity of 87 million cubic meters and controls a catchment area of 1,717 square kilometers. Yu County has a temperate continental monsoon climate. The average annual temperature is 6.5°C, which is about 6 C lower than in Beijing and roughly 2 C lower than in the Zhangjiakou city area. The frost-free period lasts for 152 days. The county receives an annual precipitation of 400 millimeters. The efficient and effective utilization of water resources plays a vital role in the socioeconomic development and urban growth of societies.

We have selected daily original water usage data from the Yu County Water Bureau, spanning from January 1, 2015, to December 31, 2021. The water data, recorded in 24-hour intervals, provides a comprehensive overview of the water consumption patterns in the region. Additionally, meteorological data for Yu County, sourced from the Yu County Meteorological Bureau, includes four key meteorological elements: temperature, wind speed, precipitation, and air quality.

The impact of temperature, wind speed, precipitation, and air quality on water demand can be summarized as follows. Temperature has a direct influence on water demand. High temperatures often lead to increased water consumption for cooling and hydration, while low temperatures can result in reduced water usage. Wind speed, along with wind direction, affects local meteorological conditions, which in turn impact water demand. Changes in wind patterns can influence temperature and humidity, influencing water consumption patterns; Precipitation levels play a vital role in water demand. Sufficient rainfall can reduce the need for irrigation and outdoor water use, potentially lowering water demand. Conversely, drought conditions can increase water demand for irrigation. Air quality indirectly affects water demand through its impact on public health. Poor air quality can lead to health issues, increasing the need for water for drinking and hygiene purposes. Cleaner air quality may have a lesser direct impact on water demand (Bougadis et al., 2005; Gato et al., 2007). Partial data is visible in Table 1.

These factors are interconnected and can vary by region and season, making them important considerations for effective water resource management and urban planning. This dataset not only offers valuable insights into the factors affecting water usage but also aids in developing more accurate and efficient models for water demand forecasting, considering the interplay between water consumption and varying weather conditions. This integrated approach is essential for developing sustainable water management strategies that support the balance between resource availability and urban development needs.

Table 1. Input data for different variables

Time	Temperature (°C)	Precipitation (mm)	Wind speed (m/s)	Air quality (AQI)	Water demand (10^3 m^3d^{-1})
2023-06-01	28.0	2.5	3.0	45	132.34
2023-06-02	29.5	1.0	2.8	50	135.59
2023-06-03	30.2	0.5	2.7	48	142.76
2023-06-04	31.5	0.0	2.5	42	134.45
2023-06-05	32.8	1.2	3.2	55	142.45
2023-06-06	30.7	3.5	3.5	60	132.41
2023-06-07	29.8	2.0	2.8	58	132.89

Comparison Study Results and Analysis

Comparing the daily predicted water usage with the actual values for the year 2021, The provided model performance data indicates that the ARIMA-LSTM integrated model outperforms the standalone ARIMA and LSTM models across all evaluation metrics. Specifically, the ARIMA-LSTM model exhibits the highest accuracy with a mean absolute percentage error (MAPE) of just 1.3%, significantly lower root mean square error (RMSE) at 2.94 (in units of 10^3 m^3d^{-1}), and the lowest mean absolute error (MAE) and mean squared error (MSE), indicating minimal average prediction errors. Its root mean squared logarithmic error (RMSLE) of 0.039 is the lowest among the three models, reflecting superior performance in predicting extreme values. Moreover, with a coefficient of determination (R^2) of 0.98, the ARIMA-LSTM model nearly perfectly fits the data, demonstrating its robustness and reliability (see Figure 3). In contrast, while the standalone ARIMA and LSTM models show decent predictive capabilities, the ARIMA model has a MAPE of 2.6%, RMSE of 9.95 (in 10^3 m^3d^{-1}), MAE of 4.52 (in 10^3 m^3d^{-1}), MSE of 27.80, RMSLE of 0.055, and an R^2 of 0.75. The LSTM model improves on these figures with a MAPE of 1.6%, RMSE of 8.02, MAE of 3.78, MSE of 22.21, RMSLE of 0.043, and an R^2 of 0.84. They are significantly outclassed by the integrated approach, highlighting the effectiveness of combining ARIMA's linear analysis strengths with LSTM's proficiency in capturing complex, non-linear relationships (see Table 2).

In the context of forecasting urban water demand using data from 2015 to 2020 and predicting for the year 2021, a comparative analysis of the ARIMA, LSTM, and ARIMA-LSTM models revealed intriguing trends in their respective performances. In the comparison of the ARIMA, LSTM, and ARIMA-LSTM deep learning models for forecasting water demand, significant differences were observed in their performances. The line graph comparisons show that while the ARIMA model captures the overall trend of water demand to some extent, it falls short of capturing finer fluctuations. The LSTM model, in contrast, shows improved prediction accuracy, especially in capturing non-linear patterns in the time series. Most notably, the ARIMA-LSTM model closely follows the actual water demand data, demonstrating superior predictive capabilities. The relative error graphs further

Table 2. Comparison of results from different prediction models

Model	MAPE/%	RMSE/ (10^3 m^3d^{-1})	MAE/ (10^3 m^3d^{-1})	MSE	RMSLE	R^2
ARIMA	2.6	9.95	4.52	27.80	0.055	0.75
LSTM	1.6	8.02	3.78	22.21	0.043	0.84
ARIMA-LSTM	0.9	2.94	3.31	18.68	0.039	0.98

Figure 3. Comparison of the predictions for urban water demand made by different models (ARIMA, LSTM, and ARIMA-LSTM)

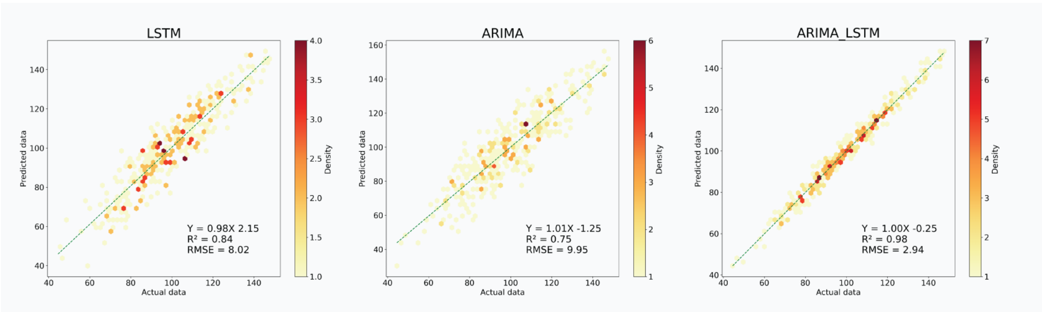


Figure 4. ARIMA-predicted monthly average water demand (2015–2021)

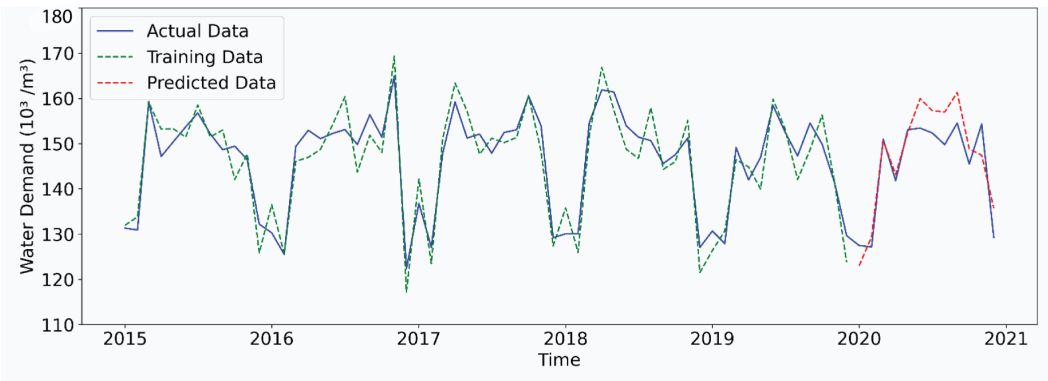
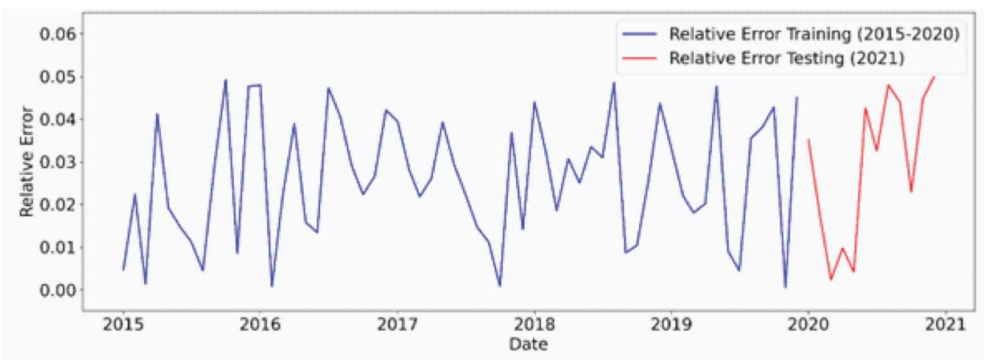


Figure 5. Relative error of ARIMA-predicted water demand



corroborate this finding, with the ARIMA-LSTM model exhibiting the lowest prediction error, indicating its significant advantage in accurately forecasting water demand.

These results suggest that while traditional ARIMA models have their strengths in handling linear time series, their performance is limited when dealing with more complex, dynamically changing data. LSTM, as a more advanced deep learning approach, shows higher adaptability in such scenarios,

Figure 6. LSTM-predicted monthly average water demand (2015–2021)

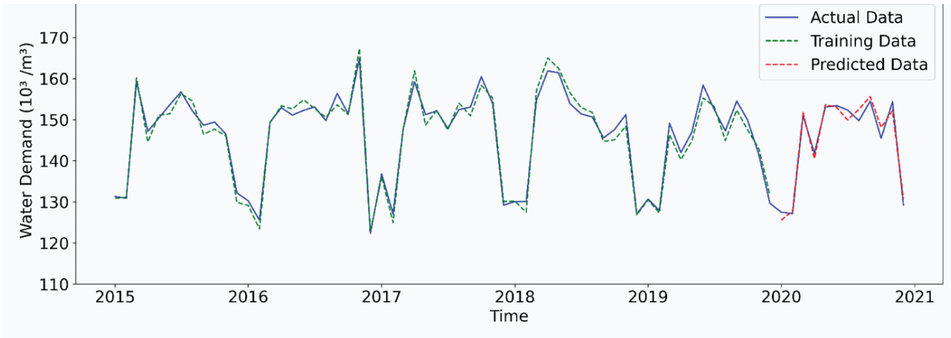


Figure 7. Relative error of LSTM-predicted water demand

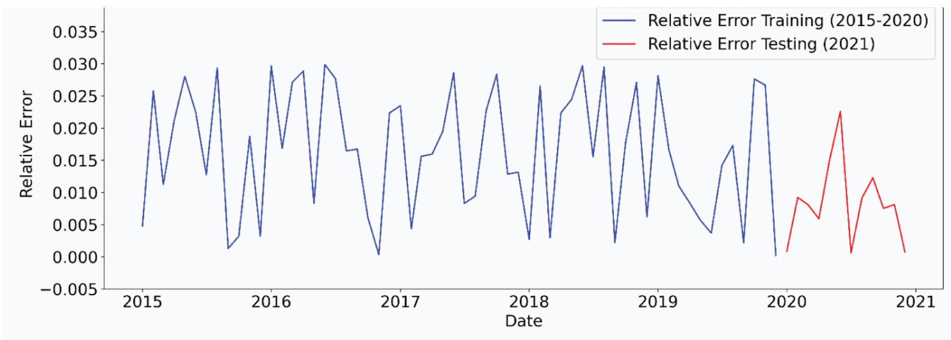
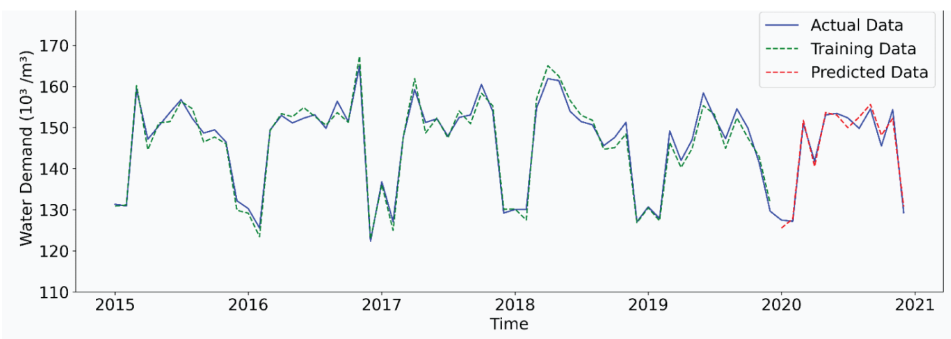


Figure 8. ARIMA-LSTM-predicted monthly average water demand (2015–2021)



particularly in understanding and predicting long-term dependencies. The integrated ARIMA-LSTM model successfully combines the strengths of both approaches, capable of capturing both linear trends and understanding complex non-linear patterns. The advantage of this integrated approach lies in its flexibility and adaptability, allowing for more accurate predictions of dynamic changes in water demand, which is crucial for water resource management and planning.

Furthermore, we conducted an analysis of the predictive variables, including an assessment of the importance of various indicators. Figure 10 illustrates the varying importance of different

Figure 9. Relative error of ARIMA-LSTM water demand

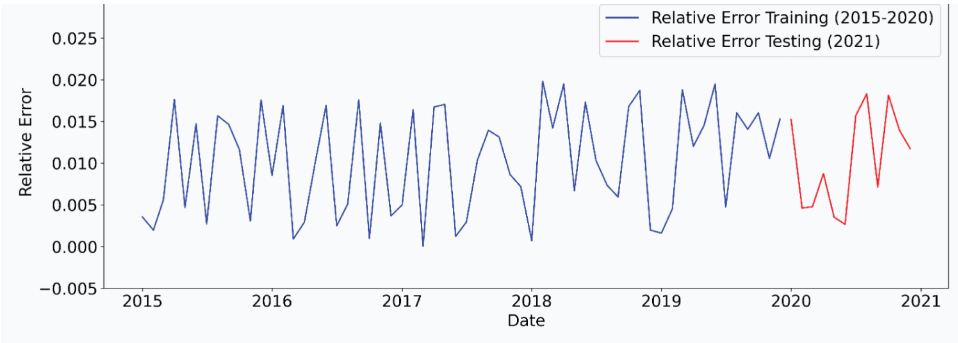
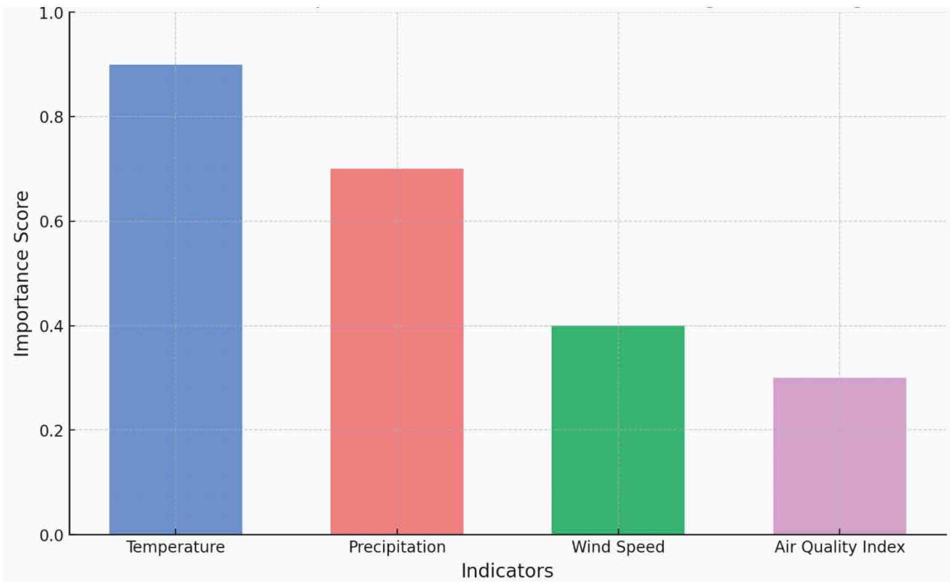


Figure 10. Indicator importance of the ARIMA-LSTM model for urban water demand forecasting



indicators in water usage forecasting. Temperature is the most significant factor, indicating its strong influence on water demand patterns. Precipitation follows as the second most important indicator, highlighting its impact on water usage. Wind speed and air quality index, though included, have a lesser impact compared to temperature and precipitation, suggesting their roles are more marginal in predicting water demand.

CONCLUSION

The primary objective of this study was to address the challenge of accurately forecasting water demand, a critical task for efficient water resource management. Accurate predictions are essential for ensuring sustainable water supply, especially in regions facing irregular rainfall patterns and growing urbanization. To tackle this problem, we proposed an integrated approach combining auto-regressive integrated moving averages (ARIMA) with long short-term memory (LSTM) networks. This hybrid model leverages the strengths of both ARIMA and LSTM: ARIMA's proficiency in analyzing linear time series data and LSTM's ability to capture long-term dependencies and non-linear patterns in complex datasets. The experiment involved collecting historical water usage data, including factor weather conditions (temperature, precipitation, wind speed, air quality index). This data was then processed and fed into three different models: a standalone ARIMA model, a standalone LSTM model, and the proposed ARIMA-LSTM integrated model. Each model was trained and tested on the dataset, with the performance evaluated based on metrics like mean absolute error (MAE), root mean squared error (RMSE), and others relevant to time series forecasting. The proposed methods involved leveraging the strengths of each model to capture linear and non-linear patterns in the historical data. The experiments were conducted by training these models on daily water demand data from 2015 to 2020 and then evaluating their predictive performance for the year 2021. The results demonstrated that while the standalone ARIMA model was adept at capturing linear trends, it struggled with complex, non-linear patterns. The LSTM model significantly improved the handling of these complexities, indicating its strength in modeling time-dependent sequences. However, the most notable performance was observed in the ARIMA-LSTM integrated model. This hybrid model not only captured both linear and non-linear patterns effectively but also outperformed the standalone models in almost all performance metrics. The accuracy of water demand forecasting was significantly higher with the ARIMA-LSTM model, indicating its potential as a robust tool for predicting water usage in various scenarios.

In this study, there are two potential shortcomings to consider. Firstly, the dependence and limitations of the data used in the research are critical. The effectiveness of the study is highly contingent on the quality and scope of the dataset. If the dataset lacks sufficient historical information or misses key variables such as extreme weather events or population growth rates, the model may fail to capture all the crucial dynamics affecting water demand. Additionally, the complexity of deep learning models, particularly LSTM networks, raises the risk of overfitting to historical data, suggesting that while the model may perform well on the training set, its performance could diminish when applied to new or unseen data. Secondly, there are limitations to the model's generalizability. The model might be overly tailored to the dataset of a specific region, limiting its applicability in different geographical areas or water demand scenarios. For instance, a model trained on urban data may not accurately predict water demands in rural areas. Moreover, water demand is influenced by a plethora of factors, including socioeconomic conditions, policy changes, and climate change. The model might not adequately adapt to these dynamically changing factors, especially in long-term forecasts. These drawbacks highlight that while the integrated ARIMA-LSTM model shows potential in predicting water demand, caution must be exercised in applying this model for decision-making support, and it should be complemented with other methods and expert knowledge for a more comprehensive and reliable prediction.

In the foreseeable future, the field of forecasting water demand using integrated ARIMA-LSTM models holds significant potential for advancements. These advancements could materialize by incorporating a wider array of data sources, including real-time consumption data, meteorological forecasts, demographic trends, and socio-economic indicators, all contributing to heightened predictive accuracy. Refining the model architecture and optimizing its parameters, coupled with exploring alternative deep learning configurations, may further enhance its performance. Moreover, adaptability to climate change and the ability to predict the impact of extreme weather events on water demand represent crucial areas of development. Integrating climate models and creating scenarios based on predicted climate changes could facilitate a better understanding of their influence on future water demand patterns. Real-time prediction and adaptability are also paramount, ensuring that the model can continuously update forecasts based on the latest data, rendering predictions more relevant for immediate decision-making in water resource management. Efforts should be made to generalize the model for different geographic locations and be scalable for both urban and rural communities. Rigorous testing and validation in diverse environments, along with necessary modifications to accommodate local conditions and data availability, are essential steps. Ultimately, seamless integration into decision support systems used by urban planners, water resource managers, and policymakers is the ultimate goal, enabling more informed and proactive planning and management of water resources. These collective advancements hold the promise of providing more accurate, adaptable, and actionable insights to support sustainable water resource management practices.

COMPETING INTERESTS

The authors of this publication declare there are no competing interests.

FUNDING

This research is supported by the Scientific Research Plan project of the Zhejiang Provincial Department of Education, "Research on Resource Management Collaboration Mechanism of Water Resources Internet of Things Based on Mobile Edge Computing" (Y202045220).

REFERENCES

Adamowski, J., & Karapataki, C. (2010). Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: Evaluation of different ANN learning algorithms. *Journal of Hydrologic Engineering*, *15*(10), 729–743. doi:10.1061/(ASCE)HE.1943-5584.0000245

Banjac, G., Vašak, M., & Baotić, M. (2015). Adaptable urban water demand prediction system. *Water Science and Technology: Water Supply*, *15*(5), 958–964. doi:10.2166/ws.2015.048

Bougadis, J., Adamowski, K., & Diduch, R. (2005). Short-term municipal water demand forecasting. *Hydrological Processes*, *19*(1), 137–148. doi:10.1002/hyp.5763

Butler, D., & Memon, F. A. (2005). *Water demand management*. Iwa Publishing.

Chen, G., Long, T., Xiong, J., & Bai, Y. (2017). Multiple random forests modelling for urban water consumption forecasting. *Water Resources Management*, *31*(15), 4715–4729. doi:10.1007/s11269-017-1774-7

Dave, E., Leonardo, A., Jeanice, M., & Hanafiah, N. (2021). Forecasting Indonesia exports using a hybrid model ARIMA-LSTM. *Procedia Computer Science*, *179*, 480–487. doi:10.1016/j.procs.2021.01.031

Donkor, E. A., Mazzuchi, T. A., Soyer, R., & Alan Roberson, J. (2014). Urban water demand forecasting: Review of methods and models. *Journal of Water Resources Planning and Management*, *140*(2), 146–159. doi:10.1061/(ASCE)WR.1943-5452.0000314

Du, B., Zhou, Q., Guo, J., Guo, S., & Wang, L. (2021). Deep learning with long short-term memory neural networks combining wavelet transform and principal component analysis for daily urban water demand forecasting. *Expert Systems with Applications*, *171*, 114571. doi:10.1016/j.eswa.2021.114571

Fan, D., Sun, H., Yao, J., Zhang, K., Yan, X., & Sun, Z. (2021). Well production forecasting based on ARIMA-LSTM model considering manual operations. *Energy*, *220*, 119708. doi:10.1016/j.energy.2020.119708

Gato, S., Jayasuriya, N., & Roberts, P. (2007). Temperature and rainfall thresholds for base use urban water demand modelling. *Journal of Hydrology (Amsterdam)*, *337*(3–4), 364–376. doi:10.1016/j.jhydrol.2007.02.014

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. doi:10.1162/neco.1997.9.8.1735 PMID:9377276

Khozani, Z. S., Banadkooki, F. B., Ehteram, M., Ahmed, A. N., & El-Shafie, A. (2022). Combining autoregressive integrated moving average with Long Short-Term Memory neural network and optimisation algorithms for predicting ground water level. *Journal of Cleaner Production*, *348*, 131224. doi:10.1016/j.jclepro.2022.131224

Kofinas, D., Mellios, N., Papageorgiou, E., & Laspidou, C. (2014). Urban water demand forecasting for the island of Skiathos. *Procedia Engineering*, *89*, 1023–1030. doi:10.1016/j.proeng.2014.11.220

Kühnert, C., Gonuguntla, N. M., Krieg, H., Nowak, D., & Thomas, J. A. (2021). Application of LSTM networks for water demand prediction in optimal pump control. *Water (Basel)*, *13*(5), 644. doi:10.3390/w13050644

Lee, G. H., Jung, S. H., & Lee, D. E. (2019). Applicability and limitation of a deep-learning algorithm, LSTM for hydrologic time-series analysis. *Proceedings of the Korea Water Resources Association Conference*. Research Gate.

Liang, F., Zhang, H., & Fang, Y. (2022). The analysis of global RMB exchange rate forecasting and risk early warning using ARIMA and CNN model. *Journal of Organizational and End User Computing*, *34*(8), 1–25. doi:10.4018/JOEUC.300762

Nasser, A. A., Rashad, M. Z., & Hussein, S. E. (2020). A two-layer water demand prediction system in urban areas based on micro-services and LSTM neural networks. *IEEE Access : Practical Innovations, Open Solutions*, *8*, 147647–147661. doi:10.1109/ACCESS.2020.3015655

Oliveira, P. J., Steffen, J. L., & Cheung, P. (2017). Parameter estimation of seasonal ARIMA models for water demand forecasting using the Harmony search algorithm. *Procedia Engineering*, *186*, 177–185. doi:10.1016/j.proeng.2017.03.225

Pandey, P., Bokde, N. D., Dongre, S., & Gupta, R. (2021). Hybrid models for water demand forecasting. *Journal of Water Resources Planning and Management*, *147*(2), 04020106. doi:10.1061/(ASCE)WR.1943-5452.0001331

Peña-Guzmán, C., Melgarejo, J., & Prats, D. (2016). Forecasting water demand in residential, commercial, and industrial zones in Bogotá, Colombia, using least-squares support vector machines. *Mathematical Problems in Engineering*, 2016.

Pierre, A. A., Akim, S. A., Semenyo, A. K., & Babiga, B. (2023). Peak electrical energy consumption prediction by ARIMA, LSTM, GRU, ARIMA-LSTM and ARIMA-GRU approaches. *Energies*, *16*(12), 4739. doi:10.3390/en16124739

Pulido-Calvo, I., Montesinos, P., Roldán, J., & Ruiz-Navarro, F. (2007). Linear regressions and neural approaches to water demand forecasting in irrigation districts with telemetry systems. *Biosystems Engineering*, *97*(2), 283–293. doi:10.1016/j.biosystemseng.2007.03.003

Sebri, M. (2016). Forecasting urban water demand: A meta-regression analysis. *Journal of Environmental Management*, *183*, 777–785. doi:10.1016/j.jenvman.2016.09.032 PMID:27652580

Shumway, R. H., Stoffer, D. S., Shumway, R. H., & Stoffer, D. S. (2017). ARIMA models. Time series analysis and its applications: With R examples (pp. 75–163).

Wu, Y., Zhang, W., Shen, J., Mo, Z., & Peng, Y. (2018). Smart city with Chinese characteristics against the background of big data: Idea, action and risk. *Journal of Cleaner Production*, *173*, 60–66. doi:10.1016/j.jclepro.2017.01.047

Xu, D., Zhang, Q., Ding, Y., & Zhang, D. (2022). Application of a hybrid ARIMA-LSTM model based on the SPEI for drought forecasting. *Environmental Science and Pollution Research International*, *29*(3), 4128–4144. doi:10.1007/s11356-021-15325-z PMID:34403057

Yamak, P. T., Yujian, L., & Gadosey, P. K. (2019). A comparison between ARIMA, LSTM, and GRU for time series forecasting. *Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence*. ACM. doi:10.1145/3377713.3377722

Yin, Z., Jia, B., Wu, S., Dai, J., & Tang, D. (2018). Comprehensive forecast of urban water-energy demand based on a neural network model. *Water (Basel)*, *10*(4), 385. doi:10.3390/w10040385

Zheng, Y., Zuo, X., Zuo, W., Liang, S., & Wang, Y. (2021). Bi-LSTM+GCN causality extraction based on time relationship. *Journal of Jilin University Science Edition*, *59*(3), 643–648.

Zubaidi, S. L., Ortega-Martorell, S., Al-Bugharbee, H., Olier, I., Hashim, K. S., Gharghan, S. K., Kot, P., & Al-Khaddar, R. (2020). Urban water demand prediction for a city that suffers from climate change and population growth: Gauteng province case study. *Water (Basel)*, *12*(7), 1885. doi:10.3390/w12071885