Groundwater Level Prediction with Deep Learning Methods

Hsin-Yu Chen 1, Zoran Vojinovic 2, Weicheng Lo 1,* and Jhe-Wei Lee 1

1 Department of Hydraulic and Ocean Engineering, National Cheng Kung University, No. 1 University Road, Tainan 701, Taiwan; n86094136@gs.ncku.edu.tw (H.-Y.C.); jheweilee@gmail.com (J.-W.L.)
2 UNESCO-IHE, Westvest 7, 2611 AX Delft, The Netherlands; z.vojinovic@un-ihe.org
* Correspondence: lowc@mail.ncku.edu.tw

Abstract: The development of civilization and the preservation of environmental ecosystems are strongly dependent on water resources. Typically, an insufficient supply of surface water resources for domestic, industrial, and agricultural needs is supplemented with groundwater resources. However, groundwater is a natural resource that must accumulate over many years and cannot be recovered after a short period of recharge. Therefore, the long-term management of groundwater resources is an important issue for sustainable development. The accurate prediction of groundwater levels is the first step in evaluating total water resources and their allocation. However, in the process of data collection, data may be lost due to various factors. Filling in missing data is a main problem that any research field must address. It is well known that to maintain data integrity, one effective approach is missing value imputation (MVI). In addition, it has been demonstrated that machine learning may be a better tool. Therefore, the main purpose of this study was to utilize a generative adversarial network (GAN) that consists of a generative model and a discriminative model for imputation. Although the GAN could not capture the groundwater level endpoints in every section, the overall simulation performance was still excellent to some extent. Our results show that the GAN can improve the accuracy of water resource evaluations. In the current study, two interdisciplinary deep learning methods, univariate and Seq2val (sequence-to-value), were used for groundwater level estimation. In addition to addressing the significance of the parameter conditions, the advantages and disadvantages of these two models in hydrological simulations were also discussed and compared. Regarding parameter selection, the simulation results for univariate analysis were better than those for Seq2val analysis. Finally, univariate was employed to examine the limits of the models in long-term water level simulations. Our results suggest that the accuracy of CNNs is better, while LSTM is better for the simulation of multistep prediction. Therefore, the interdisciplinary deep learning approach may be beneficial for providing a better evaluation of water resources.

Keywords: GAN; CNN; LSTM; imputation; groundwater prediction; alluvial fan of the Choushui River

1. Introduction

Due to its geographical and hydrological environment, Taiwan has suffered from insufficient water resources [1]. An insufficient supply of surface water is conventionally supported with groundwater. However, groundwater is a resource that accumulates over many years. Furthermore, excessive extraction of groundwater leads to land subsidence [2–8]. Hence, there remains a need to perform long-term water management. To be able to effectively allocate water resources, groundwater levels must be accurately predicted.

During the process of data collection, data may be missing due to various factors. Therefore, dealing with missing data is a challenge that needs to be overcome in any research field. To maintain data integrity, the missing value imputation (MVI) method is often chosen as a solution. Many missing value imputation methods lack validation. Thus, many studies have utilized machine learning techniques, simulating the learning process of neural networks, to impute missing values. Results have shown that model-based machine learning methods are more effective. Currently, artificial neural networks (ANNs) are
widely used in hydrological research as an important tool for groundwater management. However, in recent decades, the thrust of artificial intelligence has gradually changed from machine learning to deep learning [9]. Missing data is a problem that must be overcome in any research field. In recent years, research on using machine learning to supplement missing value imputation has gradually increased. In 2011, Buuren et al. [10] applied the MICE framework to a machine learning model and confirmed that the hybrid framework incorporated into the deep learning model can increase accuracy. Stekhoven et al. [11] established MissForest with a machine learning framework, labeled the data as a test set and training set, and achieved good simulation results.

Imputation models can be divided into discriminative models and generative models. Candes et al. [12] used matrix completion to restore a low-rank matrix identification model but could not handle sparse matrices with missing data. The denoising autoencoder (DAE) proposed by Vincent et al. [13] is a generative model. Although it has the ability to resist noise, it cannot handle data with a large amount of information. To solve the disadvantages of the two categories, in 2014, Goodfellow et al. [14] created the generative adversarial network (GAN), combining properties of the generative model and discriminative model, as shown in Figure 1. The two networks compete against each other to achieve Nash equilibrium, which greatly reduces the amount of data needed. It is the first network architecture capable of unsupervised deep learning.

Data missingness can be categorized into three types based on likelihood: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). In the case of groundwater level data, they fall under the MCAR category. This aligns with the focus of GAIN [15], which is specifically designed for this type of data. The study employs a framework with a priors (hints) mechanism to address the missing value situation in groundwater hydrology.

In the past, predecessors utilized widely applied physical models such as the Soil and Water Assessment Tool (SWAT) [17–19] and the modular three-dimensional groundwater flow model (MODFLOW) [20,21] to predict groundwater levels. In many cases, the lack of physical awareness of implicated processes might create problems in finding applicable models [22], and two additional barriers are the significant computational demands and the requirement for extensive hydrogeophysical data [23,24]. As a result, the modeling process remains fraught with uncertainty and challenges [25]. Moreover, under comparable simulation conditions, the modeling performance of ANNs surpasses that of MODFLOW [26]. Over the last decades, ANNs have sparked growing interest among researchers in various water resource issues and have obtained encouraging results [27–30], and numerous studies have employed various soft computing techniques such as a self-organizing map (SOM), an adaptive neuro-fuzzy inference system (ANFIS), GMDH, a supervised committee machine

**Figure 1.** GAN model (source: Pan et al. [16]).
with artificial intelligence (SCMAI), boosted regression trees (BRT), and Bayesian networks (BNs) to simulate time series. For instance, ANNs have been utilized for predicting groundwater level time series [31–35], estimating hydrogeological groundwater contamination, comprehending inter-annual dynamics of recharge [36–38], and capturing the influence of controlling factors on these processes [39,40].

Research on time series prediction began with regression equations, such as the autoregressive–moving average (ARMA) and autoregressive integrated moving average (ARIMA) models, which are the most basic modeling approaches [41,42]. However, achieving high-precision predictions for nonlinear data using complex models is challenging. Machine learning methods, on the other hand, can construct nonlinear models based on extensive historical data. Through iterative training and learning approximations, they can provide more accurate predictions compared with traditional statistical models. Typical methods include support vector machines (SVMs) [43], ANNs, and ensemble learning methods [44,45]. However, these aforementioned techniques often lack the effective handling of dependencies between variables, leading to limited predictive effectiveness [46].

Subsequently, LeCun [47] confirmed that deep learning can effectively mitigate the interference of outliers during the process of data feature extraction, thereby leading to a significant enhancement in technical capabilities across various domains [48,49]. As we know, in 1982, Hopfield [50] published a recurrent neural network (RNN) for sequence data, laying the foundation for deep learning models, and is often seen as the most efficient method of time series prediction. However, RNNs have vanishing gradients and exploding gradients in the process of backward transmission and cannot memorize long sequences. To improve the RNN problem, Hochreiter and Schmidhuber [51] developed long short-term memory (LSTM) to improve memory function, which is considered one of the most advanced methods for addressing time series prediction problems [52]. Hence, this study incorporated LSTM to simulate daily water levels.

In addition, LeCun et al. [53] combined the neocognitron to propose a convolutional neural network (CNN) and then applied the model to a one-dimensional CNN with Bengio et al. [54]. Learning features from sequence data gradually allows the network to achieve excellent performance in signal data and natural language. A CNN mainly extracts spatial information from image-like data, while LSTM was developed for sequence data. In this study, LSTM and a CNN were used to simulate water levels, and the differences between the two methods were compared. This study is expected to reduce the limitations of sequence data research methods and provide more possibilities for water resources research. Therefore, to address the problem of missing hydrological observation data, the purpose of the present study is to use a GAN for imputation. Complete hydrological data can improve the accuracy of groundwater level prediction. Accordingly, the performance of groundwater level prediction was compared using two interdisciplinary methods. One is the LSTM model specially developed for sequence data, while the second model is a CNN, which is good at processing image information. Figure 2 shows the flow of the entire research.
2. Materials and Methods

2.1. Case Study

The research area was in the Choushui River basin, the second-largest basin area in Taiwan. The slope increases from west to east, and the mainstream is 186.6 km long, making it the longest river in Taiwan. The elevation of the basin is approximately 0–100 m, the terrain is flat, and the groundwater subdivision contains rich water resources. The area is as large as 3156.9 square kilometers, which is the second-largest basin area in Taiwan. The mainstream comes from between the main peak and the east peak of Mount Hehuan, which is approximately 3200 m above sea level. Due to the large terrain changes, the information from the rainfall station of the Central Weather Bureau shows that the spatial distribution of annual rainfall decreases when moving from mountainous areas to the coast. The eastern side has steep mountains and abundant rainfall, while the western coast has less rainfall. Geographically south of the Tropic of Cancer, the Choushui River is affected by ocean currents and thus has a tropical marine climate, resulting in an extremely uneven spatial and temporal distribution of rainfall. Figure 3 shows the monthly average rainfall in the Choushui River. However, the temporal distribution of rainfall throughout the year is strongly heterogeneous. On average, 75% of total annual rainfall falls between May and September in the wet season (or the high-flow period), while October to April of the next year is the dry season (or the low-flow period), with only 25% of annual rainfall. Coupled with the impact of extreme climates in recent years, the difference between abundance and drought is particularly obvious.
The Central Geological Survey Report [55] noted that the strata in the Choushui River Basin are long and narrow from north to south and slightly westward in an arc-shaped distribution. According to soil properties and water content characteristics, it is divided into fan top, fan center, and fan tail [56]. The fan top is composed of unconsolidated sediments, with good strength and water permeability, and is easily recharged with surface water. It is an area rich in groundwater resources. The shallow layer in the central fan area is composed of coarse medium sand mixed with silt sand with well-developed pores and is the main aquifer. The deep layer is mostly a clay layer, and the thickness of the water-blocking layer increases gradually, and the denser layer leads to poor water permeability. From the center of the fan to the tail of the fan belongs to the middle and lower reaches of the mainstream, and the alluvial layer with a thickness of several hundred meters is formed by the alluvial flow of the river. It is mainly composed of clay and silt, and the aquifer is thinning. When one artificially overpumps the groundwater, the surface elevation is lowered, causing formation deformation and compression.

To explore the long-term groundwater variability, data collected from 2002 to 2021 by the Water Resources Agency (MOEA) were used. We knew the general situation and distribution of observation wells and detected missing 20-year long-term data. There were 63 stations with missing data [57,58]. The stations with too much missing data were eliminated, and the stations with 2–40 missing days of data were interpolated. Then, Thiessen’s polygon method was used to configure the rainfall data for observation wells, and the GAN was used to interpolate the missing values to obtain complete data.

Due to the high autocorrelation of the time series data, the univariate and Seq2val (sequence-to-value) frameworks were used to run the CNN and LSTM, respectively. We discussed the significance of the parameters and the advantages and disadvantages of the model and indicated that the two-cross-domain model can be applied to hydrology. Finally, the encoder–decoder framework proposed by Hamilton et al. [59] was incorporated to develop Seq2seq at the limit of long-term simulated water levels, as shown in Figure 4.
2.2. Methodology

To investigate the long-term groundwater variability, the measured groundwater level in the alluvial fan of the Choushui River from 2002 to 2021 was used as the examined data. First, the imputation of missing data was performed. For this purpose, five years of data were used to train the GAN. After our preliminary results were successfully verified, the missing part of the 20-year groundwater level data could be further imputed. In the current study, univariate and Seq2val were employed using the groundwater level and rainfall as hydrological parameters, which were inputted into the CNN and LSTM models, respectively. Then, the meaning of the parameters as well as the pros and cons of the models were discussed. Finally, the encoder–decoder was incorporated to explore the limit of the long-term simulated groundwater level in Seq2seq.

2.2.1. Water Level Prediction Data Preprocessing

To supervise the model to achieve a reasonable balance, the holdout method [60] was used to divide the data into three parts according to the following ratio: 70% for the training dataset (2002~2015), 10% for the validation dataset (2016~2017), and 20% for the test dataset (2018~2021), as shown in Table 1, Figures 5 and 6.

\[
X_{\text{nom}} = \frac{X - \mu}{X_{\text{max}} - X_{\text{min}}} \in [-1, 1] 
\]  

(1)

Table 1. Database.

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th>Validation</th>
<th>Test</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groundwater level (m)</td>
<td>5113</td>
<td>731</td>
<td>1461</td>
<td>7305</td>
</tr>
<tr>
<td>Precipitation (mm)</td>
<td>5113</td>
<td>731</td>
<td>1461</td>
<td>7305</td>
</tr>
</tbody>
</table>

Note: The unit is days.

The research parameters were the daily groundwater level and daily rainfall, and the representativeness and values of the two were completely different. Therefore, to ensure that the model training could be driven normally, the data had to be preprocessed via normalization, as shown in Figure 7.

2.2.2. Generative Adversarial Network

The generative adversarial network (GAN) was proposed by Goodfellow et al. [14] in 2014. It is the first deep learning application that enables unsupervised learning. Figure 8 shows that it learns the distribution of data in such a way that the discriminating network and generative network compete with each other without relying on any assumptions. A GAN generates samples using forward propagation to reduce computational costs and stops upon reaching Nash equilibrium. Research has incorporated cueing mechanisms to
reinforce the process [15]. We used the self-defined random variable $H|M = m$ to prompt the discriminating network and train the best neural network (NN).

$$L = E[\log P(F = \text{real} | X_{\text{real}})] + E[\log P(F = \text{fake} | X_{\text{fake}})]$$

(2)

Random noise ($z$) is sampled from the latent space following a Gaussian distribution. The generator $G$ produces $X_{\text{fake}}$, which approximates real values, while both $X_{\text{fake}}$ and $X_{\text{real}}$ are used as inputs to the discriminator network. The training objective function is a binary classification task aimed at distinguishing between $X_{\text{real}}$ and $X_{\text{fake}}$.

![Figure 5. Holdout method.](image)

<table>
<thead>
<tr>
<th>Training set</th>
<th>Validation set</th>
<th>Test set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training, tuning, and evaluation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Machine learning algorithm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictive Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final performance estimate</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 5.** Holdout method.

**Figure 6.** Datasets.
2.2.2. Generative Adversarial Network

The generative adversarial network (GAN) was proposed by Goodfellow et al. [14] in 2014. It is the first deep learning application that enables unsupervised learning. Figure 8 shows that it learns the distribution of data in such a way that the discriminating network and generative network compete with each other without relying on any assumptions. A GAN generates samples using forward propagation to reduce computational costs and stops upon reaching Nash equilibrium. Research has incorporated cueing mechanisms to reinforce the process [15]. We used the self-defined random variable $H|M=m$ to prompt the discriminating network and train the best neural network (NN).

$$L = \mathbb{E} \left[ \log P(F = \text{real} | X) \right] + \mathbb{E} \left[ \log P(F = \text{fake} | X) \right]$$ (2)

Random noise ($z$) is sampled from the latent space following a Gaussian distribution. The generator $G$ produces $X_{\text{fake}}$, which approximates real values, while both $X_{\text{fake}}$ and $X_{\text{real}}$ are used as inputs to the discriminator network. The training objective function is a binary classification task aimed at distinguishing between $X_{\text{real}}$ and $X_{\text{fake}}$.

Figure 7. Before (blue) and after (red) normalization of data.

Figure 8. GAN architecture.

2.2.3. Long Short-Term Memory

In 1997, Hochreiter and Schmidhuber [51] proposed LSTM based on the RNN and added cell states to improve the memory function, as shown in Figure 9. They increased the long-term dependence of the model and successfully solved the defects of RNN gradient vanishing and gradient explosion. LSTM is mainly composed of three stages: an input gate, an output gate, and a forget gate. Four kinds of information are obtained via splicing training with the current input and the previous transfer state. Among them, $f_t, i_t, o_t$ are normalized with sigmoid to $[0, 1]$, while $\tilde{C}_t$ is normalized with the Tanh activation function to $[-1, 1]$ to assist memory update and storage.
2.2.3. Long Short-Term Memory

In 1997, Hochreiter and Schmidhuber [51] proposed LSTM based on the RNN and added cell states to improve the memory function, as shown in Figure 9. They increased the long-term dependence of the model and successfully solved the defects of RNN gradient vanishing and gradient explosion. LSTM is mainly composed of three stages: an input gate, an output gate, and a forget gate. Four kinds of information are obtained via splicing training with the current input and the previous transfer state. Among them, $f_t$, $i_t$, $o_t$, $C_t$ are normalized with sigmoid to $[0, 1]$, while $C_{t-1}$ is normalized with the Tanh activation function to $[-1, 1]$ to assist memory update and storage.

2.2.4. Convolutional Neural Network

In 1989, LeCun et al. [53] proposed the CNN infrastructure LeNet-5, which mimics the neuron structure of the human visual cortex. It was based on the neocognitron proposed by Japanese scientist Fukushima et al. [61] in 1980, laying the foundation for CNNs. Through the continuous complexity of the model and the maturity of hardware technology, CNNs became the ImageNet champion in 2012. LeCun and Bengio [54] developed a one-dimensional (1D) CNN to learn features from segment sequence data, making the scope of the CNN more diverse. Whether a CNN is 1D, two-dimensional (2D), or three-dimensional (3D), it follows the same method and characteristics, and the only difference is in the method of extracting features, as shown in Figure 10. Figure 11 shows that a CNN mainly consists of convolutional layers, pooling layers, and fully connected layers. A CNN uses filters to extract information features and convolves data with sliding windows to reduce computational costs. The weights are computed using the chain rule and updated with backpropagation. This enables sequence data to preserve the location information of features and has the advantage of full-value sharing.

2.2.5. Objective Function

Written in Python version 3.9.5, packages such as NumPy, pandas, scikit-learn, and Matplotlib were used to perform basic operations. TensorFlow and Keras were also used for data imputation, training, and simulation. The most common mean absolute error (MAE), mean square error (MSE), and log-cosh metrics were selected for loss function screening, as shown in Figure 12. Finally, to detect whether the model exhibited underfitting or overfitting, we used MAE, RMSE, and R2 as evaluation indicators.

Figure 9. RNN structure (left) and LSTM structure (right).

Figure 10. Difference between 1D CNN and 2D CNN (source: Ackermann [62]).
2.2.5. Objective Function

Written in Python version 3.9.5, packages such as NumPy, pandas, scikit-learn, and Matplotlib were used to perform basic operations. TensorFlow and Keras were also used for data imputation, training, and simulation. The most common mean absolute error (MAE), mean square error (MSE), and log-cosh metrics were selected for loss function screening, as shown in Figure 12. Finally, to detect whether the model exhibited underfitting or overfitting, we used MAE, RMSE, and $R^2$ as evaluation indicators.

**MAE**

The mean absolute error (MAE) is a linear function that solely measures the average value of modeling errors. It is less susceptible to the influence of outlier data points, making it suitable for scenarios where anomalies are treated as damaged data. However, a
drawback is that its gradient remains constant, even when the loss value is small, and the
gradient remains large, which can hinder neural network learning and potentially lead to
missing the optimal point.

\[ \text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \]  

- MSE

Mean squared error (MSE) is a regression loss function, and the gradient changes
accordingly based on the increase or decrease in the loss. MSE is also sensitive to outliers,
which can lead to rapid amplification. It is suitable for datasets where outliers represent
significant anomalous situations.

\[ \text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]  

- Log-cosh loss

The advantages of Log-cosh, which combines the benefits of MSE and MAE, include its
tendency to converge near the optimal value as gradients progress and its robust handling
of outlier values. However, log-cosh may not be a perfect fit for this specific dataset due to
gradient issues arising from significant deviations in the modeled values.

3. Results
3.1. The Results of Imputation
3.1.1. Hyperparameter Settings of GAN

The hyperparameters of the GAN were the result of repeated adjustments. The model
only set the batch size and number of iterations. Adam in Table 2 was controlled by Alpha
and the Hint rate, and the nonlinear relationship was added using ReLU. The main objective
of this research’s target function was to minimize the loss function, where the model returns
the function difference to assist in gradually adjusting the parameters. A smaller function
value indicates that the simulated values are closer to the real values. Finally, the error was
returned by the MSE, and the weight parameters were gradually adjusted.

Table 2. Hyperparameter settings of GAN.

<table>
<thead>
<tr>
<th>Optimizer</th>
<th>Activation</th>
<th>Loss Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adam</td>
<td>ReLU</td>
<td>MSE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Iterations</th>
<th>Alpha</th>
<th>Hint Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>100,000</td>
<td>100</td>
<td>0.9</td>
</tr>
</tbody>
</table>

- Adam

Adam is one of the optimization methods used for model optimization and is an
extension of the gradient descent optimization algorithm. It was proposed by Kingma
and Ba in 2015 [63] and has been widely applied in various deep learning applications.
Compared with the traditional gradient descent method, which uses a fixed learning rate,
Adam can adjust the learning rate adaptively and remember the previous gradient states.
Adam performs well in both linear and non-linear functions and exhibits advantages over
other adaptive learning methods. It also requires less memory and is computationally
efficient, making it well suited for large datasets. Therefore, in this study, Adam was
selected as the optimization tool.

- ReLU

Rectified linear units (ReLU) is a piecewise linear function widely used in deep learn-
ing applications due to its ability to address the vanishing gradient problem. As indicated
in Equation (2), ReLU only activates input values that are greater than zero. This characteristic accelerates convergence and better aligns with the activation patterns observed in the human brain.

\[ f(x) = \max(0, x) \] (5)

3.1.2. Imputation Results

The research data were classified as “MCAR” (missing completely at random). In the literature on missing value imputation, only 10.8% of the articles imputed more than 50% of the missing values. To evaluate the stability and limits of the GAN, the data used a matrix of size [2192 × 46] composed of 46 stations with no missing records in 2015–2020. To imitate missing data scenarios, complete data can be randomly generated with missing rates of 10%, 20%, and up to 90%. Additionally, consecutive missing values for a duration of 2 to 30 days can be introduced, and then imputation techniques can be applied to fill in the missing values.

Figure 13 presents the imputation results using the GAN, evaluated based on the RMSE. The blue area indicates a good imputation performance, while the boundary between the white and red areas represents the limit of data imputation in this study. This result was used as the criterion for selecting imputed stations. Stations with missing data falling within the white or red areas were directly excluded from the imputation process and were not considered for imputing missing values.

![Figure 13. Imputation results of GAN (RMSE) (note: the unit is m) (The blue area indicates a good imputation performance, while the boundary between the white and red areas represents the limit of data imputation in this study).](image-url)
3.1.3. Practical Imputation Case Discussion

Water levels vary by geographic location. Figure 14 shows that the water level at the top of the fan (Stations 1–3) is high, the curve is smooth, and the simulation results are excellent. However, the water level significantly curves at the center (Stations 4–6) and tail (Stations 7–9) of the fan fluctuate, and the difference in local changes makes the interpolation error larger. In particular, Station 5 and Station 9 have the most drastic changes, and the peaks do not fit well, resulting in poor performance, as captured by the evaluation indicators. However, there is still an overall trend, so the feasibility of the model can be determined.

Figure 14. Supplementary results of fan top (Stations 1–3), fan central (Stations 4–6), and fan tail (Stations 7–9) water level stations.

Imputation was performed on the actual missing data. Figure 15 shows the serial trend can be captured in the smooth segment and that Station 2 is consistent with the normal regression segment. Each station was affected by different external factors, resulting in different shocks. However, since the model is trained with Gaussian random noise, it has a certain ability to resist noise. There are sudden rises and falls in the steep rising section of the daily water level. The water level of Station 3 rises gently and then is straight until the peak, which shows that the reaction of raw water is slightly slower. However, the series still maintains a stable change, which can be used to identify the overall trend.

3.2. Prediction Results

3.2.1. Hyperparameter Settings of CNN and LSTM

We referenced callbacks to monitor data changes and set EarlyStopping. When overfitting occurred or the model index did not improve and the convergence was too slow, the training was automatically terminated. When there was no abnormality in the training process, it automatically judged and stopped at the best epoch. Finally, ModelCheckpoint retained the best parameters to improve the model setting adjustments. Hyperparameter Settings of CNN and LSTM are shown in Tables 3–6.
Figure 14. Supplementary results of fan top (Stations 1–3), fan central (Stations 4–6), and fan tail (Stations 7–9) water level stations.

Imputation was performed on the actual missing data. Figure 15 shows the serial trend can be captured in the smooth segment and that Station 2 is consistent with the normal regression segment. Each station was affected by different external factors, resulting in different shocks. However, since the model is trained with Gaussian random noise, it has a certain ability to resist noise. There are sudden rises and falls in the steep rising section of the daily water level. The water level of Station 3 rises gently and then is straight until the peak, which shows that the reaction of raw water is slightly slower. However, the series still maintains a stable change, which can be used to identify the overall trend.

Figure 15. The actual missing data imputation results.

Table 3. Hyperparameter settings of LSTM.

<table>
<thead>
<tr>
<th>Optimizer</th>
<th>Activation</th>
<th>Loss Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adam</td>
<td>ReLU</td>
<td>MSE</td>
</tr>
</tbody>
</table>

Table 4. LSTM network architecture.

<table>
<thead>
<tr>
<th>Name</th>
<th>Dimension</th>
<th>Parameter Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input layer</td>
<td>(None, 3, 1)</td>
<td>0</td>
</tr>
<tr>
<td>LSTM</td>
<td>(None, 128)</td>
<td>66,560</td>
</tr>
<tr>
<td>Dense</td>
<td>(None, 100)</td>
<td>1290</td>
</tr>
<tr>
<td>Dense</td>
<td>(None, 5)</td>
<td>505</td>
</tr>
</tbody>
</table>

Total parameters: 79,965

Table 5. Hyperparameter settings of CNN.

<table>
<thead>
<tr>
<th>Optimizer</th>
<th>Activation</th>
<th>Loss Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adam</td>
<td>ReLU</td>
<td>MSE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Filters</th>
<th>Kernel Size</th>
<th>Stride</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

3.2.2. Results of Prediction

This section presents the results of the CNN and LSTM models. Root-mean-square error (RMSE), mean absolute error (MAE), correlation coefficient ($R^2$), scatter index (SI), and BIAS were used to assess the performance of the models. The formulas for these statistical parameters have been provided in the literature [64–66].
Table 6. CNN network architecture.

<table>
<thead>
<tr>
<th>Name</th>
<th>Dimension</th>
<th>Parameter Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input layer</td>
<td>(None, 3, 1)</td>
<td>0</td>
</tr>
<tr>
<td>Conv1D</td>
<td>(None, 3, 64)</td>
<td>192</td>
</tr>
<tr>
<td>Conv1D</td>
<td>(None, 3, 64)</td>
<td>8256</td>
</tr>
<tr>
<td>MaxPooling1D</td>
<td>(None, 2, 64)</td>
<td>0</td>
</tr>
<tr>
<td>Conv1D</td>
<td>(None, 2, 64)</td>
<td>8256</td>
</tr>
<tr>
<td>Conv1D</td>
<td>(None, 2, 64)</td>
<td>8256</td>
</tr>
<tr>
<td>MaxPooling1D</td>
<td>(None, 1, 64)</td>
<td>0</td>
</tr>
<tr>
<td>Flatten</td>
<td>(None, 64)</td>
<td>0</td>
</tr>
<tr>
<td>Dense</td>
<td>(None, 50)</td>
<td>3250</td>
</tr>
<tr>
<td>Dense</td>
<td>(None, 10)</td>
<td>510</td>
</tr>
</tbody>
</table>

Total parameters: 28,720

- Univariate and Seq2val

There was no significant difference between the simulated water levels of the CNN and LSTM, and the endpoint values were accurately simulated. The overall performance was excellent, and the $R^2$ values of the four models were all as high as 0.99 in Table 7. The accuracy of the CNN was not lower than that of LSTM, which means that the CNN does have strong potential for the prediction of one-dimensional data. Seq2val performed worse than univariate, indicating that the methods perform poorly on multivariate data. It is speculated that due to the large difference in parameter values, the weight of rainfall was too large, resulting in a slight shift in the RMSE.

Table 7. Evaluation indexes of CNN and LSTM.

<table>
<thead>
<tr>
<th></th>
<th>RMSE (m)</th>
<th>MAE (m)</th>
<th>$R^2$</th>
<th>SI</th>
<th>BIAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Univariate–CNN</td>
<td>0.007</td>
<td>0.005</td>
<td>0.998</td>
<td>0.0505</td>
<td>−0.0037</td>
</tr>
<tr>
<td>Univariate–LSTM</td>
<td>0.008</td>
<td>0.005</td>
<td>0.997</td>
<td>0.1265</td>
<td>−0.0086</td>
</tr>
<tr>
<td>Seq2val–CNN</td>
<td>0.0321</td>
<td>0.0194</td>
<td>0.9981</td>
<td>0.0506</td>
<td>0.0155</td>
</tr>
<tr>
<td>Seq2val–LSTM</td>
<td>0.0508</td>
<td>0.0342</td>
<td>0.9955</td>
<td>0.1856</td>
<td>0.0245</td>
</tr>
</tbody>
</table>

- Seq2seq

Since the above results indicate better performance using univariate analysis, these data were used to test the limits of the simulation. Multistep prediction of the CNN and LSTM with the Seq2seq framework was carried out. The peak value was slightly shifted within $T = 5$, but the $R^2$ was as high as 0.95 in Table 8, which indicated good performance. Although the $R^2$ of $T = 5–7$ was above 0.9, the RMSE was not excellent. It is speculated that where the curve changed sharply, the endpoint simulation was slightly extreme. After $T = 7$, the evaluation index gradually deteriorated, and the accuracy of the CNN decreased faster than that of LSTM, indicating that LSTM is still relatively good at remembering long-term information. Overall, the simulation accuracy of both models is good, and that of the CNN is slightly better. It is confirmed that the CNN can also be used as a tool for sequence data, and it has more potential in the field of hydrology.

Table 8. CNN and LSTM multistep prediction.

<table>
<thead>
<tr>
<th>Seq2seq–CNN</th>
<th>Seq2seq–LSTM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T = 1</td>
</tr>
<tr>
<td></td>
<td>RMSE (m)</td>
</tr>
<tr>
<td>Seq2seq–CNN</td>
<td>0.0304</td>
</tr>
<tr>
<td>Seq2seq–LSTM</td>
<td>0.0631</td>
</tr>
<tr>
<td>Seq2seq–CNN</td>
<td>0.0972</td>
</tr>
</tbody>
</table>
Table 8. Cont.

<table>
<thead>
<tr>
<th></th>
<th>Seq2seq–CNN</th>
<th></th>
<th></th>
<th>Seq2seq–LSTM</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE (m)</td>
<td>MAE (m)</td>
<td>$R^2$</td>
<td>RMSE (m)</td>
<td>MAE (m)</td>
<td>$R^2$</td>
</tr>
<tr>
<td>$T = 4$</td>
<td>0.1311</td>
<td>0.0879</td>
<td>0.9705</td>
<td>0.1338</td>
<td>0.0921</td>
<td>0.9690</td>
</tr>
<tr>
<td>$T = 5$</td>
<td>0.1634</td>
<td>0.1100</td>
<td>0.9529</td>
<td>0.1666</td>
<td>0.1151</td>
<td>0.9522</td>
</tr>
<tr>
<td>$T = 6$</td>
<td>0.1951</td>
<td>0.1331</td>
<td>0.9323</td>
<td>0.1956</td>
<td>0.1347</td>
<td>0.9320</td>
</tr>
<tr>
<td>$T = 7$</td>
<td>0.2252</td>
<td>0.1543</td>
<td>0.9087</td>
<td>0.2256</td>
<td>0.1559</td>
<td>0.9089</td>
</tr>
<tr>
<td>$T = 8$</td>
<td>0.2533</td>
<td>0.1737</td>
<td>0.8811</td>
<td>0.2538</td>
<td>0.1760</td>
<td>0.8810</td>
</tr>
<tr>
<td>$T = 9$</td>
<td>0.2803</td>
<td>0.1926</td>
<td>0.8482</td>
<td>0.2811</td>
<td>0.1951</td>
<td>0.8501</td>
</tr>
<tr>
<td>$T = 10$</td>
<td>0.3049</td>
<td>0.2097</td>
<td>0.8124</td>
<td>0.3058</td>
<td>0.2125</td>
<td>0.8160</td>
</tr>
</tbody>
</table>

Figures 16–18 show the simulation results of the three parts of the alluvial fan. $T = 1$ of Station 1 is in a mountainous area with a high daily water level and suffers less noise. However, Station 2, which is in the center of the fan, has more shocks, resulting in a slightly lower $R^2$. LSTM has high accuracy on the smooth segment but has a slightly extreme phenomenon on the simulated endpoint. The CNN evaluation indices are better. However, both CNN and LSTM are suitable tools for simulating groundwater, and both $R^2$ values are as high as 0.98 or more. $T = 5$ shows that the two models are worse at simulating extreme values, and there is a shift in the smoothing section, resulting in a rapid increase in the RMSE. This is especially the case for Station 2, which is more affected by noise: the simulation curve of LSTM fluctuates gradually, but the $R^2$ remains steady. The overall performance can capture trend changes, indicating that the model has a fairly good ability to grasp changes. The simulation at $T = 10$ deviates significantly. There is underestimation in the raw water section and overestimation in the receding water section. The simulation offset of LSTM at Station 2 is very severe, with an overestimation phenomenon as high as 0.69 m. Therefore, the model does not have the ability to make multistep forecasts 10 days in advance.

Figure 16. Station 1: top of alluvial fan.
Figure 16. Station 1: top of alluvial fan.

Figure 17. Station 2: center of alluvial fan.

Figure 18. Station 3: tail of alluvial fan.

4. Conclusions

The imputation method using a GAN is composed of a generative model and a discriminative model to fill in the missing data. Our results show that the trend of the sequence region could be reasonably simulated in the smooth section. Although some of the extreme values could not be captured in the undulating section of the groundwater level curve, there was still a certain trend so the feasibility of the model could be determined. Although the GAN could not capture the groundwater level endpoints in every section, the overall simulation performance was still excellent to some extent.

We note that if the characteristics of other models can be incorporated into the basis of the GAN architecture, the performance can be improved. Since the GAN was used in a section with drastic changes, it did not accurately capture the endpoint value. It is recommended to incorporate other mechanisms to strengthen the model to capture the change in the endpoint and improve the ability to respond faster in the sudden rise and fall section.

The CNN and LSTM were found to be excellent for hydrological estimation. The coefficient of determination for both was calculated to be approximately 0.99, as indicated in Table 7, which also lists the values of the RMSE and MAE. Regarding the parameter selection, the simulation results of the univariate analysis (RMSE = 0.007 and MAE = 0.005) were better than those of the Seq2val analysis (RMSE = 0.0321 and MAE = 0.0194). Upon closer inspection, compared with LSTM (RMSE = 0.008 and R = 0.997), the CNN was shown to be slightly better in the evaluation indexes (RMSE = 0.007 and R = 0.998). Both models underestimated the long-term trend. The CNN outperformed LSTM for a shorter stride, but for a longer stride, the drop in accuracy in the CNN was observed to be faster than that in LSTM.

In hydrology, the hysteresis of rainfall recharge has been the subject of many research discussions. From the perspective of the hydrological simulation, the CNN, which is good at processing two-dimensional image information, was shown to be better in accuracy and...
4. Conclusions

The imputation method using a GAN is composed of a generative model and a discriminative model to fill in the missing data. Our results show that the trend of the sequence region could be reasonably simulated in the smooth section. Although some of the extreme values could not be captured in the undulating section of the groundwater level curve, there was still a certain trend so the feasibility of the model could be determined. Although the GAN could not capture the groundwater level endpoints in every section, the overall simulation performance was still excellent to some extent.

We note that if the characteristics of other models can be incorporated into the basis of the GAN architecture, the performance can be improved. Since the GAN was used in a section with drastic changes, it did not accurately capture the endpoint value. It is recommended to incorporate other mechanisms to strengthen the model to capture the change in the endpoint and improve the ability to respond faster in the sudden rise and fall section.

The CNN and LSTM were found to be excellent for hydrological estimation. The coefficient of determination for both was calculated to be approximately 0.99, as indicated in Table 7, which also lists the values of the RMSE and MAE. Regarding the parameter selection, the simulation results of the univariate analysis (RMSE = 0.007 and MAE = 0.005) were better than those of the Seq2val analysis (RMSE = 0.0321 and MAE = 0.0194). Upon closer inspection, compared with LSTM (RMSE = 0.008 and R = 0.997), the CNN was shown to be slightly better in the evaluation indexes (RMSE = 0.007 and R = 0.998). Both models underestimated the long-term trend. The CNN outperformed LSTM for a shorter stride, but for a longer stride, the drop in accuracy in the CNN was observed to be faster than that in LSTM.

In hydrology, the hysteresis of rainfall recharge has been the subject of many research discussions. From the perspective of the hydrological simulation, the CNN, which is good at processing two-dimensional image information, was shown to be better in accuracy and speed. The model originally developed for 2D data still had good performance when applied to 1D data. Our results suggest that the interdisciplinary deep learning approach may be beneficial for providing a better evaluation of water resources. The accuracy of the CNN is better, while LSTM is better for the simulation of multistep prediction. It is recommended to combine multiple networks to improve their individual deficiencies. Trying to use other parameters to see if the accuracy can be improved is also one of the future research directions of hydrological information.

Author Contributions: Conceptualization, H.-Y.C. and J.-W.L.; methodology, H.-Y.C.; software, H.-Y.C.; validation, H.-Y.C.; formal analysis, H.-Y.C. and J.-W.L.; investigation, H.-Y.C.; resources, H.-Y.C.; data curation, H.-Y.C.; writing—original draft preparation, H.-Y.C.; writing—review and editing, H.-Y.C. and Z.V.; visualization, H.-Y.C.; supervision, W.L.; project administration, H.-Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: The first author is grateful for the support received from the Ministry of Science and Technology, Taiwan, under contract no. 108-2221-E-006-006-MY2.

Data Availability Statement: The data are available from the corresponding author.

Acknowledgments: The authors appreciate the Water Resources Office at the Water Resources Agency, Taiwan, and the Central Weather Bureau, Taiwan for providing the important information and data used in this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References


11. Stekhoven, D.J.; Bühlmann, P. MissForest—Non-parametric missing value imputation for mixed-type data. *Bioinformatics* 2012, 28, 112–118. [CrossRef]


22. Kirkner, J.W. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resour. Res.* 2006, 42. [CrossRef]


37. Fijani, E.; Nadiri, A.A.; Asghari Moghaddam, A.; Tsai, F.T.-C.; Dixon, B. Optimization of drastich method by supervised committee machine artificial intelligence to assess groundwater vulnerability for maragheh-bonab plain aquifer, Iran. *J. Hydrol.* 2013, 503, 89–100. [CrossRef]


**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.