Simulation of Spring Discharge Using Deep Learning, Considering the Spatiotemporal Variability of Precipitation

Chunmei Ma[1], Haoyu Jiao[1], Yonghong Hao[2], Tian-Chyi Jim Yeh[2,3], Junfeng Zhu[4], Huiqing Hao[2], Jiahui Lu[1], Jiankang Dong[2]

[1]School of Computer and Information Engineering, Tianjin Normal University, Tianjin, China.

[2]Tianjin Key Laboratory of Water Resources and Environment, Tianjin Normal University, Tianjin, China.

[3]Department of Hydrology and Atmospheric Sciences, The University of Arizona, Tucson, AZ 8 5721-0011, USA.

[4]Kentucky Geological Survey, University of Kentucky, Lexington, KY, USA

Correspondence to: Yonghong Hao, haoyh@sxu.edu.cn

Key Points:

- A generative variational autoencoder is applied to augment precipitation data to improve a LSTM network for spring discharge prediction.
- Augmenting precipitation data improves learning generalization and predictive capability of various deep learning models.
- The generative variation autoencoder offers a novel solution to address data scarcity issue across diverse research domains.

Abstract

Precipitation data collected from sparse monitoring stations in numerous karst catchment pose a challenge for hydrologic models to accurately capture spatial and temporal correlation between precipitation and karst spring discharge, hindering the development of robust simulation models. To address this data scarcity issue, this study employs a coupled deep learning model that integrates a variation autoencoder (VAE) for augmenting precipitation data and a long short-term memory (LSTM) network for karst spring discharge prediction. The VAE contributes by generating synthetic precipitation data through an encoding-decoding process. This process generalizes the observed precipitation data by deriving joint latent distributions with improved preservation of temporal and spatial correlations in the data. The combined VAE-generated precipitation and observation data are used to train and test the LSTM for predicting the spring discharge. Applied to Niangziguan spring catchment in northern China, our coupled VAE/LSTM model demonstrated significantly higher predictive accuracy compared to a LSTM model using only field observations. We also explored temporal and spatial correlations in the observed data and the impact of different ratios of VAE-generated precipitation data to actual data on model performances. Additionally, our study evaluated the effectiveness of VAE-augmented data on various deep learning models and compared VAE with other data augmentation techniques. Our study demonstrates that the VAE offers a novel approach to address data scarcity and uncertainty, improving learning generalization and predictive capability of various hydrological models.

Plain Language Summary

Millions of people around the world use spring as their water sources. To protect these precious springs, water resources managers need to have a good understanding on how spring discharges change in the future under the stress of climate change and human activities. A common tool to help improve this understanding is a computer model. A trustworthy computer model requires plenty of quality data, which are unfortunately not available for many springs. To address this data scarcity issue, we applied a computer-based learning technique, called variation autoencoder (VAE), that learned the patterns of real-world data and generated data that complied with the learned pattern. We then combined the generated data and real-world data to train a computer-based learning model, called long short-term memory (LSTM) network, that is excellent in simulating spring discharges. We tested our method using Niangziguan spring in the northern China, demonstrating that adding VAE-generated data significantly improved the LSTM model. In addition, we investigated the effectiveness of the VAE in improving other common models. The study shows that our model is accurate in predicting spring discharges and VAE is a very helpful tool in improving our model.

1 Introduction

Continuous and discontinuous carbonate rocks cover 15.2% of the global land surfaces (Goldscheider et al.,2020). The karst aquifers formed by carbonate rock formations provide fresh water for approximately 678 million or 9.2% of the world's population (Stevanović,2019). Carbonate rocks cover 3.44 million km^2 of China territory (He et al.,2019). Many large karst groundwater catchments feed big karst springs in northern China (Han et al., 2006). Over the past decades, human activities and environmental problems have led to declining groundwater levels or karst spring dry-ups in many regional karst groundwater catchments (Hao et al., 2009). Therefore, accurate simulation and prediction of karst spring discharge is essential for the sustainable management of water resources in the region.

To gain comprehension and accurate predictive capacity regarding these intricate hydrological processes, Labat et al. (2000) developed a rainfall-runoff model that integrates linear and steady-state rainfall-runoff models to identify and simulate the processes. Similarly, Juki et al. (2009) introduced a conceptual rainfall-runoff model to estimate the components of groundwater balance, including the dynamic catchment boundaries and the subsurface flow influences within the catchment area. Notwithstanding the significant advancements achieved by conventional hydrological models, limitations persist when confronted with spatiotemporal nonlinearity and nonstationarity (Wunsch et al., 2022; Çallı et al., 2022).

The recent development of deep learning (DL) methods has made significant strides in modeling the spatiotemporal behavior of rainfall-discharge processes. For example, Artificial Neural Networks (ANNs) have been employed in investigating karst hydrological processes (Yaseen et al., 2015) and have emerged as a prominent tool in hydrology. Wunsch et al. (2022) utilized Convolutional Neural Networks (CNNs) to simulate the flow within three karst spring

catchment areas in the Alps and the Mediterranean region. Nevertheless, exploring time series data, a vital component in hydrological research, has remained somewhat limited (Yin et al., 2022). To address this gap, Song et al. (2022), Yin et al. (2022), and Zhou et al. (2022) have implemented Long Short-Term Memory (LSTM) networks to simulate karst spring flow and effectively manage time series data. Additionally, Cheng et al. (2021) conducted a comparative evaluation of three machine learning methods (Multi-Layer Perceptron (MLP), LSTM, and Support Vector Machine (SVM) to enhance our understanding of the mechanisms behind the fluctuations in spring discharge in the Longzici spring's karst area and its relationship with precipitation. Their research underscores that artificial neural networks are the preferred approach for simulating and predicting karst spring discharge (Zhou et al., 2022; Gai et al., 2023).

However, deep learning models often demand substantial training data for robust performance (Shorten et al., 2021; Tang et al., 2022). In the context of hydrological process studies, the high costs associated with data collection and the inherent spatial and temporal randomness or stochasticity of the natural phenomena present formidable challenges (Dugdale et al., 2022). In remote areas, data collection tasks can become even more arduous, thus constraining a comprehensive grasp of hydrological processes (Mengistu et al., 2022). Nowhere is this challenge more pronounced than in karst regions, where highly spatially variable groundwater flow and the intricate dynamics of groundwater flow and spring discharge complicate sampling and monitoring efforts (Hartmann et al., 2014). Consequently, the sparsely sampled data brings significant challenges to apply deep learning in spring discharge simulation.

Some traditional methods have already been employed to deal with data sparsity in hydrology. For example, Yeh et al. (2015) and Yeh et al. (2023) have summarized many

advances over the past decades in applying the Bayesian statistic concept to characterize aquifer heterogeneity and to predict groundwater flow and solute transport processes in spatially variable geologic media with spare data. In particular, they promoted collecting more spatially non-redundant data cost-effectively using hydraulic tomography to reduce uncertainty in predictions (Yeh and Liu, 2000**;** Zhu and Yeh, 2005). Varouchakis et al. (2013) and Smith et al. (2021) proposed interpolation techniques to enhance model performance with sparse data. Bruckmann et al. (2020) employed sequential Gaussian simulation for the statistical modeling of groundwater flow, enabling the quantification of overall uncertainty in sparsely sampled hard rock aquifers. Sun et al. (2020) also applied it to meteorological observations or regions with limited or no data on the Qinghai-Tibet Plateau by inversely evaluating the reconstructed precipitation with a glacier-hydrology model. This approach contributes to catchment hydrological modeling and forecasting research. Similarly, Grundmann et al. (2019) introduced an inverse surface hydrologic modeling approach to reconstruct spatial-temporal rainfall patterns stochastically and applied it to areas with data scarcity and poor catchment measurements. These methods, however, typically rely on complex mathematical models governed by partial differential equations and algorithms. Moreover, their limited adaptability to diverse hydrological data domains and types hampers their cross-domain applicability and restricts flexibility and scalability, posing challenges in addressing evolving hydrological research needs.

While the methods above have improved prediction accuracy in hydrology under sparse sampling conditions, they often struggle to handle the large-scale datasets required by deep learning models (Ghorbanidehno et al., 2020; Addor et al., 2020). Among the diverse branches of deep learning, generative models such as the Variational Autoencoder (VAE), presented in this study, can be precious when confronted with challenges like data scarcity, inadequate

sampling (Lin et al., 2023), and the intricacies of hydrological process modeling (Chen et al., 2022). Specifically, this study addresses predicting spring discharge time series from spatially limited observed precipitation time series over a large-scale karst catchment, where unknown spatial and temporal variability of precipitations and hydrologic processes exist. In particular, a deep-learning model that integrates a VAE and a Long short-term memory (LSTM) network developed based on the spatiotemporal statistics of precipitation and discharge to predict the most likely spring discharges and address their uncertainty in a karst catchment. This paper is organized as follows.

(1) It first introduces the problems of predicting spring discharge at the Niangziguan Springs karst catchment in China with sparse precipitation observation stations. (2) It then investigates the spatial and temporal statistics of seven spatially sparse precipitation time series to demonstrate that Bayesian statistics can generate many possible precipitation time series in the latent space. These generated time series would represent those not observed at other parts of the catchment and have the same spatiotemporal statistics as those observed. Besides, the study examines the statistical relationship of the observed precipitation time series with the spring discharge time series. Afterward, it uses their spatiotemporal relationships to analyze the most likely discharge time series and address the uncertainty as the stochastic predictive and inverse models in subsurface hydrology discussed in Yeh et al. (2015) and Yeh et al. (2023). (3) This study proposes a deep-learning network (VAE/LSTM) model to avoid complex classical governing surface and subsurface flow partial differential equations and characterization of the hydraulic properties of the karst aquifer. (4) This study presents the results of the model's application to the Niangziguan Karst catchment. It demonstrates the model's validity: argumentation of more synthetic precipitation time series, capturing the precipitation's spatial

variability, improves the spring discharge prediction. (5) At last, this study discusses its scientific insights.

2 Statement of Problems

The Niangziguan Springs composite, one of the largest karst springs in northern China, is located in east Shanxi Province, China, with an annual average spring discharge of 9.35 m^3/s from 1959 to 2019. The Karst aquifer is an Ordovician carbonate aquifer, sandwiched by Quaternary loess deposits, Permian shale, Carboniferous argillaceous limestone with coal seams on top, and Cambrian dolomite on the bottom (Gai et al., 2023). The karst groundwater flows eastward, and when groundwater meets with the low-permeable dolomite strata of the Cambrian at the Mian River valley, it perches on the surface, and Niangziguan Springs occurs (Figure 1).

The landforms of the Niangziguan Springs catchment are rough hilly terrains and gentle sloping river terraces, where the elevation ranges from 2149m to 362m above mean sea level (MSL). Precipitation is the primary aquifer recharge source, with an annual average precipitation of 534.6 mm from 1959 to 2019 (Gai et al., 2023).

Before 1971, the Niangziguan Springs catchment was a remote rural mountainous area with a well-developed river water system. Residents mainly used river and spring water for their water supply. Karst groundwater was not developed and remained in natural condition during that time. Although the residents utilized water from the rivers and streams connected to the subsurface groundwater, widespread pumping development of the karst aquifer did not occur. That is, the impacts of human activities on groundwater were negligible (Hao et al., 2016; Song et al., 2022).

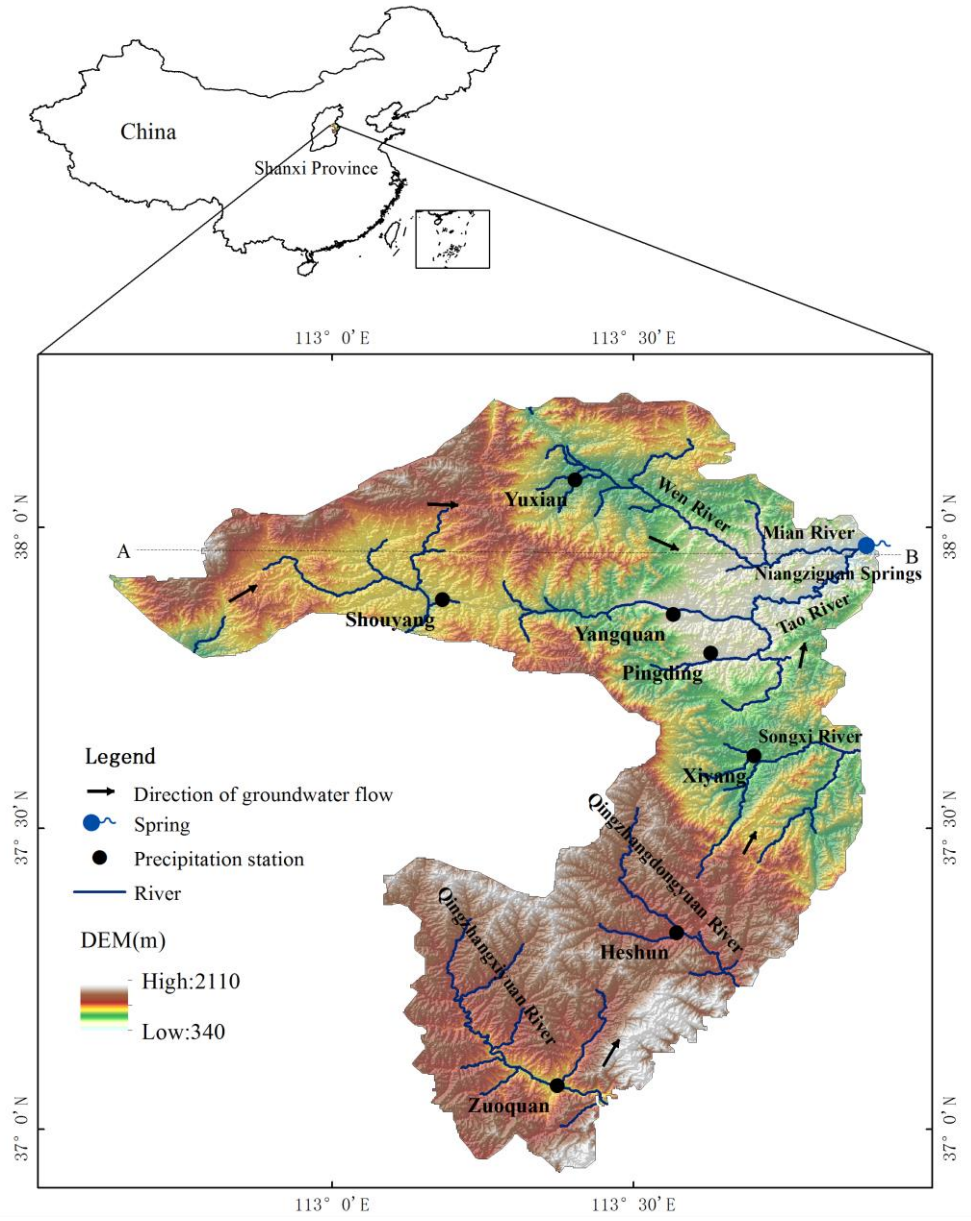

Figure 1. Niangziguan Springs catchment, China.

With China's reforming and opening up from the beginning of the 1970s, regional economic development and population growth increased. The Niangziguan Springs catchment has become one of the national heavy industrial zones for coal mining, power generation, and metallurgy. Specifically, after the early 1970s, groundwater of the catchment began to be developed for industrial, municipal, and irrigation uses. The regional economic and social

development largely influenced karst groundwater in the Niangziguan Springs catchment and the Niangziguan Spring discharge accordingly.

Figure 2 displays the monthly spring discharges measured at the Niangziguan Springs gauging station in Mian River (point B in Figure 1) from 1959 to 2019. The discharge records exhibit a two-scale variability, a large-scale declining trends from 1959 to 2019, depicted by the polynomial and linear regression lines, and local-scale variability at the monthly level. As illustrated in the figure, the points of intersection between the linear regression line and the polynomial regression line occur at two distinct time points, namely 1971.10 and 2006.12. Based on the two intersecting points, we divide the discharge records into the nature period (1959.1-1970.12), the overexploitation period (1971.1-2006.12), and the recovery period (2007.1-2019.12). This segmentation method comprehensively analyzes and compares spring discharge characteristics in these distinct phases while exploring potential patterns and trends in karst spring spatiotemporal variation.

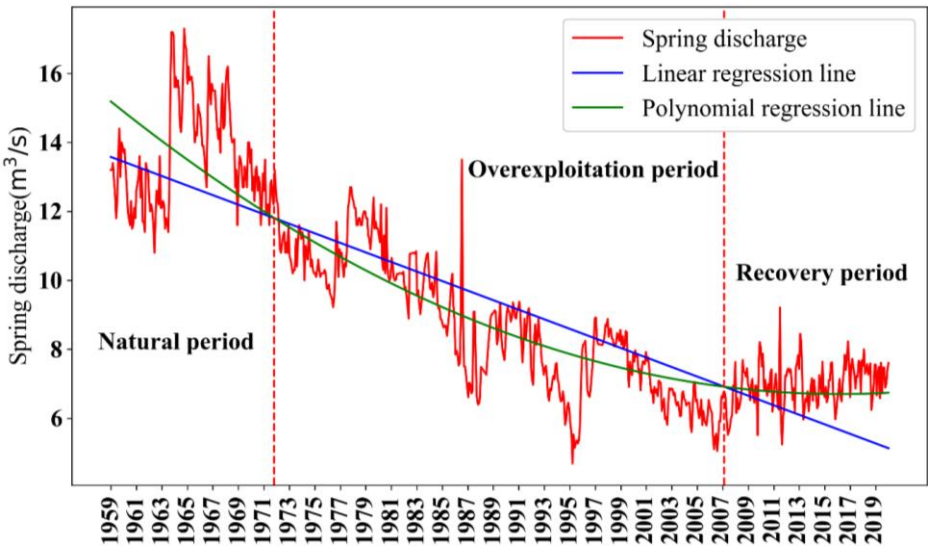

Figure 2. Spring discharge data partition of Niangziguan spring from 1959 to 2019.

From 1971 through 2006 (the overexploitation period), the region underwent rapid development due to China's economic reform and open-door policy. Due to a substantial demand for groundwater, spring discharge exhibited a significant declining trend (An et al., 2020; He et al., 2019).

Post-2006, influenced by China's sustainable development strategy, local authorities actively promoted industrial transformation (Liu et al., 2020). As a result, natural resource consumption and groundwater exploitation decreased, leading to a gradual recovery in spring discharge, and we call this period the recovery period.

Precipitation over the catchment is the primary source of the spring discharge. Figure 3 illustrates the monthly precipitations as a function of time (from 1959 to 2019) recorded at seven meteorological stations located at Yuxian, Shouyang, Pingding, Xiyang, Heshun, and Zuoquan counties, and Yangquan city (i.e., the black dots in Figure 1) (Gai et al., 2023). These figures exhibit the variability of the monthly precipitation values, which do not exhibit the trend in the discharge in Figure 2. Absence of the trend confirms our interpretation of groundwater usage's influence on the spring discharge due to the catchment development.

The mean values and standard deviations of the precipitation corresponding to Figure 3 are also listed in Table 1.

Table 1. The means and standard deviations of the monthly precipitation time series (1959-2019) at the seven stations.

	Yangquan	Pingding	Yuxian	Shouyang	Xiyang	Heshun	Zuoquan
Mean(mm)	45.2	44.8	46.0	41.0	45.5	45.6	43.7
Variance	3493.9	3426.7	3388.1	2521.2	3626.4	3273.3	2888.4

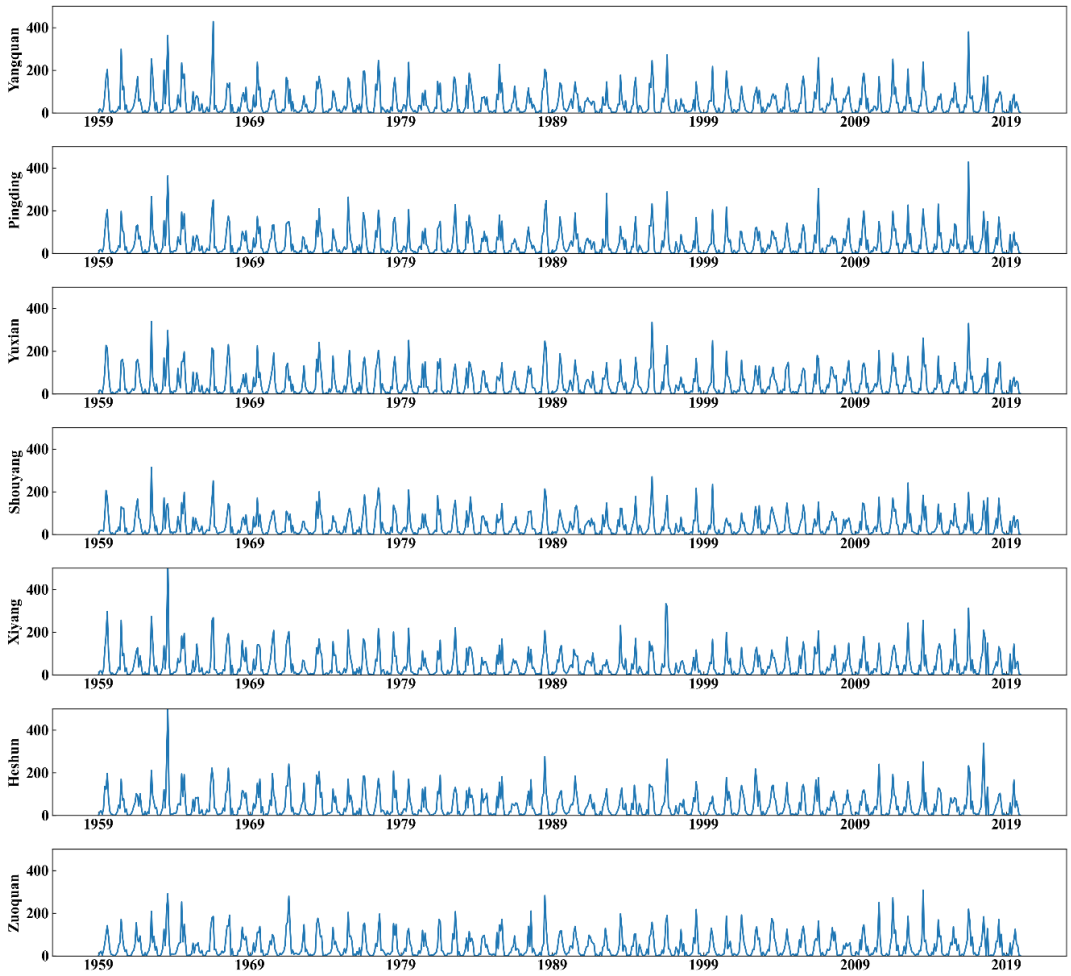

Figure 3. The monthly precipitations as a function of time (from 1959 to 2019) recorded at seven meteorological stations.

As indicated in Table 1, the mean values of the monthly precipitations (1959-2019) at the seven stations are almost the same. Their standard deviations are close, indicating a slight spatial variation in the monthly precipitation. However, the precipitation records manifest a periodic behavior: heavy rainfalls in the middle of every year (the temporal variability).

Figure 4a illustrates the temporal autocorrelation of the recorded precipitation time series (from 1959 to 2019) at Yangquan station. The maximum or minimum autocorrelation values decrease as the lag (separation) time increases. The autocorrelation behaviors of the precipitation time series at the other six stations have similar patterns, showing the same 12-month periodicity

(Figure 4b). This periodicity implies that if this month's precipitation is higher than the mean, the precipitation 12 months later will likely be higher than the mean or vice versa. The negative covariances indicate that if a given month's precipitation value is higher than the mean, the precipitation separated by a time lag between 3 to 9 months will be below the mean value of the entire record (Table 1). Such a periodic autocorrelation suggests that the precipitations over the catchment lasted less than a month. As a result, the monthly accumulated precipitations at different locations over the catchment do not exhibit significant deviations.

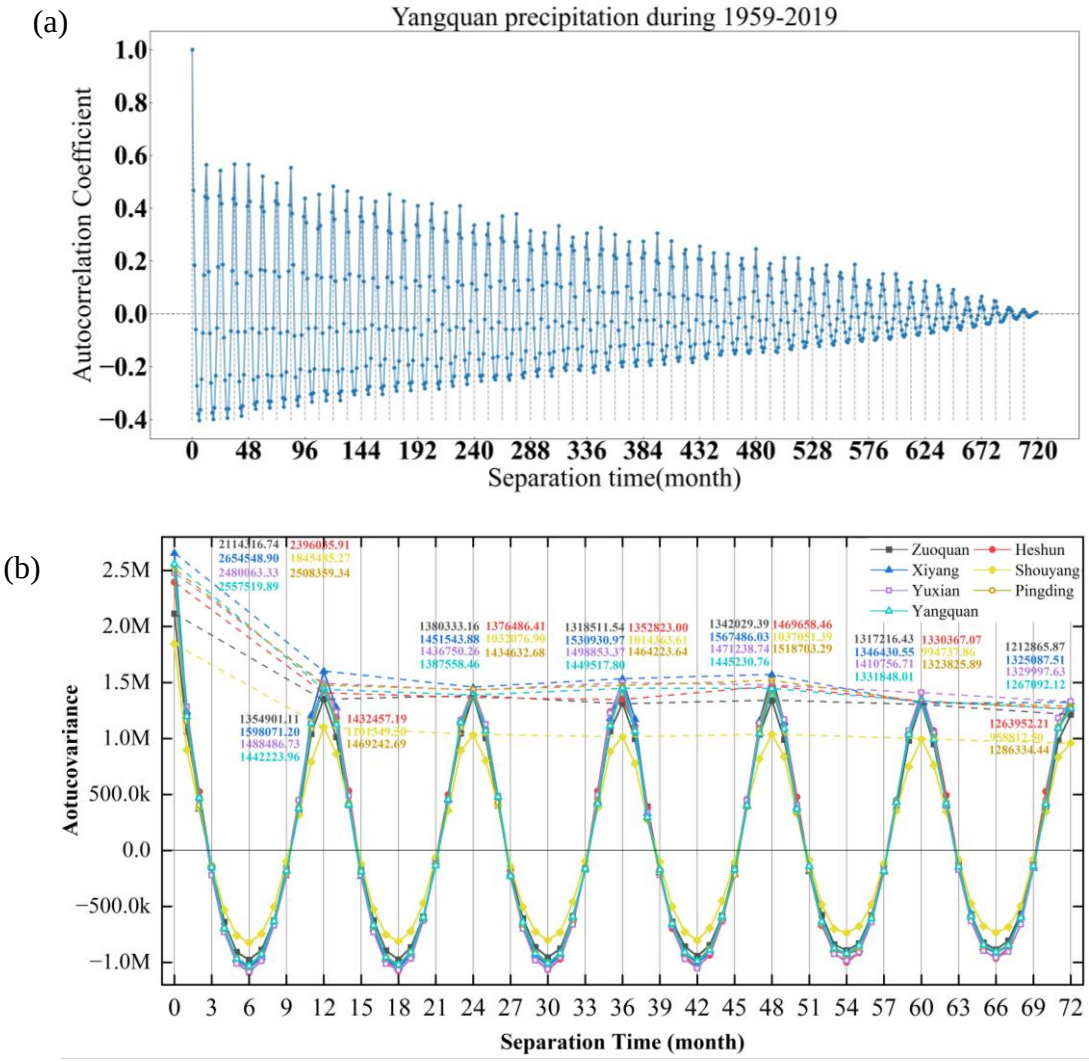

Figure 4. (a) illustrates the autocorrelation of the precipitation from 1959 to 2019 at Yangquan station. (b) shows the autocovariance of the monthly precipitation perturbations as a function of time recorded at seven meteorological stations (1959 to 1964).

Figure. 4b shows that the time series at the seven stations behave similarly. The similarity is expected since the time series are monthly precipitations, which integrate the spatial variation of the precipitation over the time intervals of less than a month. However, they are slightly different, unveiling spatial variations in the precipitation over the entire catchment. This spatial variation is also confirmed by Figure 5, which plots the correlation values of the recorded precipitation among the seven stations, revealing that the precipitations are correlated well between adjacent stations and less with far away stations.

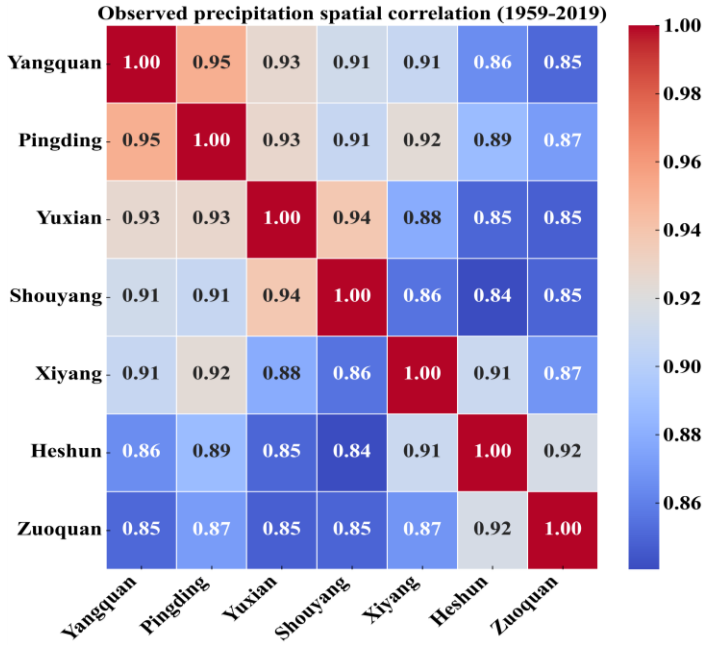

Figure 5. The correlations between the precipitation data recorded at the seven stations illustrate their spatial similarity and variation.

We also carried out the autocorrelation analysis for the spring discharge time series in Figure 2. The autocorrelation of the spring discharge without detrending is shown in Figure 6a. and that of the discharge after removing the polynomial trend (Figure 2) is illustrated in Figure 6b, which exhibits some periodicities at large and small scale. The small-scale variations have one year's periodicity, corresponding to the above mentioned precipitations. It suggests that an appropriate cross-correlation analysis of the precipitation and spring discharge data must use the

detrended spring discharge data. Table 2 lists the means and standard deviations of the spring discharge time series (1959-2019) before and after detrending. Figures 6a and 6b and Table 2 indicate that human activity heavily influences spring discharge time series during the over-exploration period.

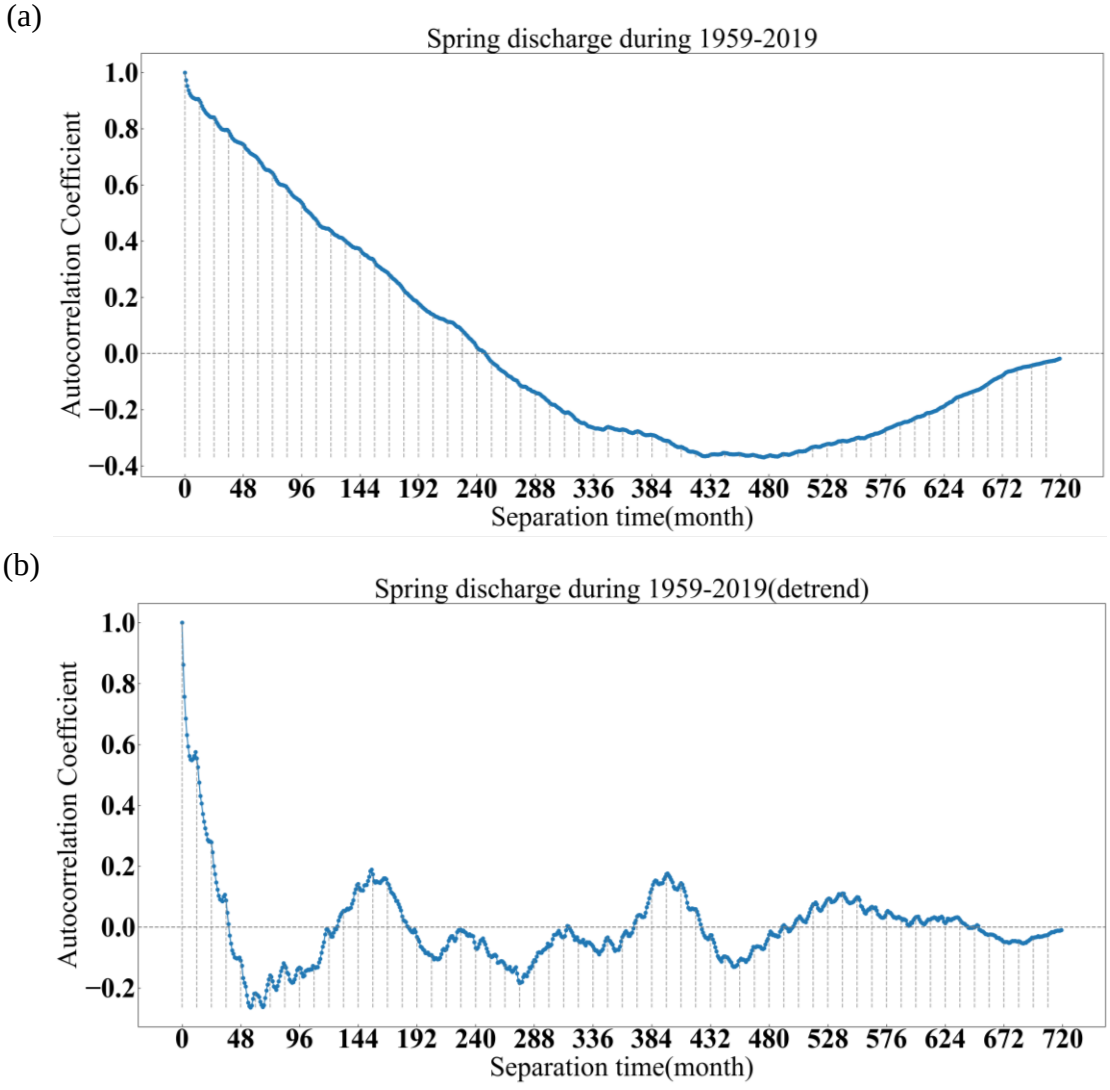

Figure 6. Auto-correlation of spring discharge. (a) non-detrended. (b) detrended data.

Table 2. Means and variances of non-detrened and detrended spring discharge data.

Spring discharge	Non-detrened	Detrended
Mean (mm)	9.35	0
Variance (mm^2)	7.99	1.49

The analyses in Figures 3, 4, 5, and 6 suggest that the precipitation over the catchment and the spring discharge time series can be conceptualized as spatiotemporal stochastic processes (Preciley, 1979) or latent processes (AI jargon): they are random but correlated in space and time.

Since the precipitation is the primary source of the spring discharge, relationships between the precipitation and discharge data must exist. As such, we instigated the cross-covariance between the discharge and the precipitation at each of the seven stations as a function of separation time in Figure 7. The discharge time series (1959-2019) were detrended, and the residuals were then analyzed.

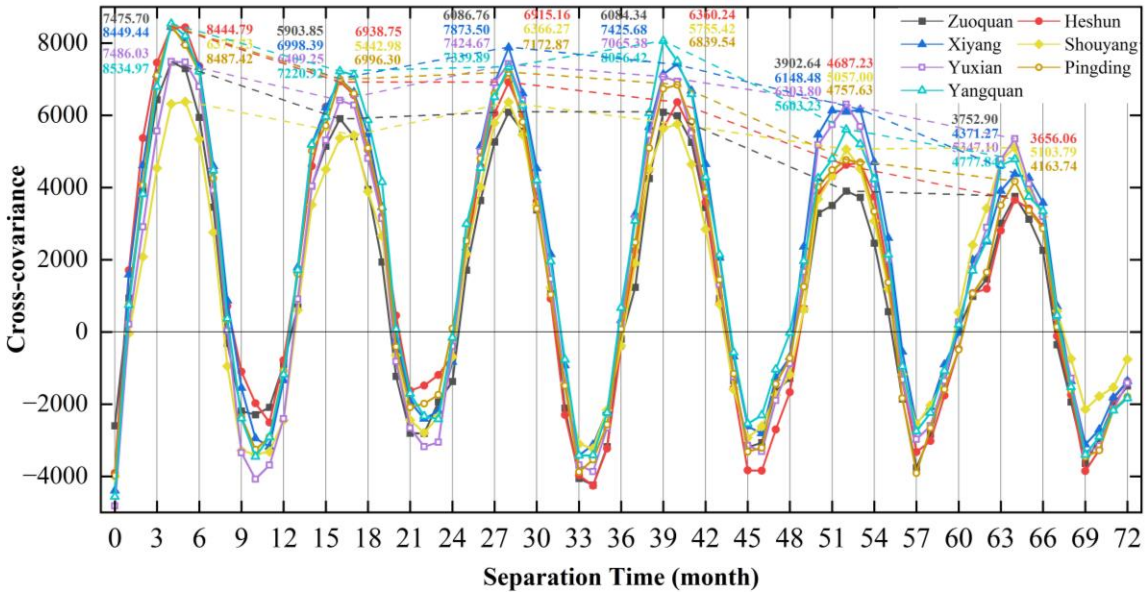

Figure 7. The cross-correlation between the discharge (detrended) and the precipitation at each of the seven stations as a function of separation time (month).

The cross-covariance in Figure 7 shows a periodicity of almost 12 months as the autocorrelation of the precipitation time series (after human impacts were removed), suggesting that the precipitations closely control the spring discharge. Notice that the maximum cross-correlation occurs four or five months after the precipitation perturbation greater than the mean

value. These four or five months may present the average travel time for the precipitation to reach the spring discharge location.

The above discussion and Figure 1 demonstrate that the catchment covers an area with significant topographic variations, implying precipitations likely vary in time and space, not captured by the monthly cumulative precipitation data from the seven stations. The problems of sparse measurement and spatiotemporal variability processes are not new. Many scientists have employed stochastic approaches to deal with this issue in hydrology. For instance, Yeh et al. (2015) and Yeh et al. (2023) introduced the approach that conceptualizes the spatial variation of hydraulic properties in geologic media as a spatial stochastic process. Given some observed values, the process is exemplified by a joint posterior distribution with the mean, variance, and autocorrelation function as a function of correlation scales.

Within this probability distribution, many possible heterogeneous aquifer parameters exist and can be used to simulate the statistically most likely flow and transport processes and to derive the associated uncertainty due to the effects of unknown heterogeneity missed by sampling. This approach is called Monte Carlo (MC) simulation in hydrology. This approach, however, requires solving 3-D partial differential equations governing the surface and subsurface flow with many parameters, boundaries, and initial conditions. Such an approach is a complex numerical simulation task.

Of course, this study could have followed the stochastic MC simulation approach, but it faces many difficulties. An alternative is to use a deep-learning machine approach, skipping solving 3-D partial differential equations. However, it needs to deal with sparsely distributed temporal varying precipitation information.

For this reason, this study proposes a new deep-learning network model combining VAE and LSTM to predict precipitation-driven spring discharge. Figure 8 displays the architecture of the proposed network. VAE is a data generation model that can generate new data of similar spatiotemporal statistics (mean, covariance, and probability distribution) from the input data via the encoding and decoding mechanism in VAE. On the other hand, LSTM is a model that exploits spatiotemporal cross-correlation statistics between precipitation and spring discharge data to predict spring discharge at different times.

3 Variational Autoencoder (VAE) and Long Short-Term Memory (LSTM) network: VAE / LSTM model

3.1 Variational AutoEncoders (VAE)

The VAE is a data-generative model that consists of two parts: an encoder and a decoder. The encoder derives latent posterior distributions of input data by neural networks and samples from the distributions to obtain latent samples. The decoder converts the latent samples back to the data that have similar statistical features with input data by neural networks. Different from MC simulation, this process is controlled by the neural network parameters trained by the input data. Details are available in Supporting Information S1 (Figure S1).

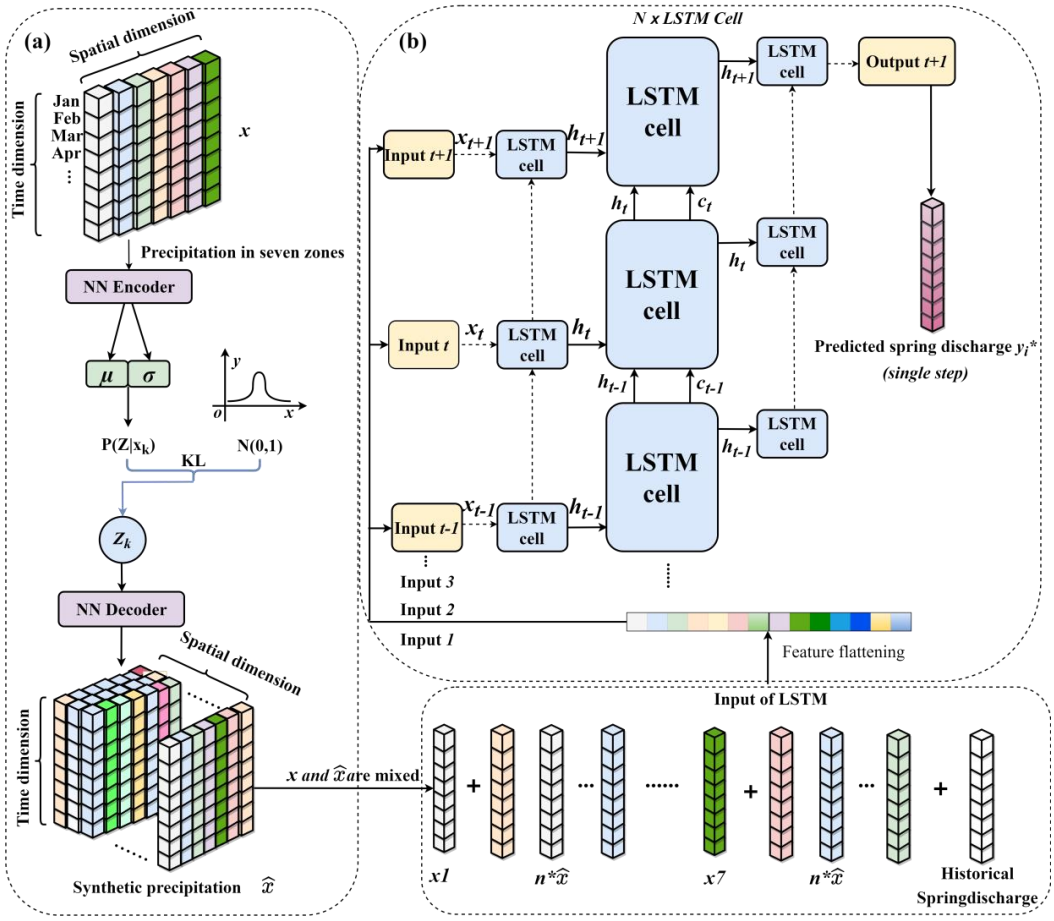

Figure 8. The architecture of the VAE/LSTM for precipitation-driven spring discharge prediction. (a) is the module of the VAE for new precipitation data generation that has similar spatiotemporal statistics with the observed precipitation. (b) is the module of LSTM for spring discharge prediction using the observed precipitation, synthetic precipitation and historical spring discharges.

The VAE encoder is conditioned on the observed precipitation from the seven stations to derive seven latent posterior probability distributions with their means and covariances. Specifically, let X_{ik} be the precipitation data set of all the i-th month at the k observation station, where i=1,2,...,12, k =1,2,...,7 (representing 12 months of the observed precipitation data in seven zones of Niangziguan Springs catchment). The encoder maps X_{ik} to a latent state distribution $P_{\theta_E}(Z_{ik} \mid X_{ik})$ with the mean u_{ik} and the variance σ_{ik}^2, which is the posterior probability distribution

of $P(X_{ik})$. θ_E is the parameter matrix of the encoder neural network associated with all zone precipitation. Thus, the latent distribution is a joint posterior probability distribution of the observed precipitation data, which can characterize other unknown locations by mean, variance, and covariance (i.e., spatiotemporal relationship between the precipitations at different locations in the entire catchment). Let z_{ik} be the sample of $P_{\theta_E}(Z_{ik} \mid X_{ik})$. Then the VAE decoder converts z_{ik} to a new precipitation data, $\hat{x}_{ik}$ as $P_{\theta_D}(\hat{x}_{ik} \mid z_{ik})$, where θ_D is the parameter matrix of the decoder neural network.

The VAE, in essence, conceptualizes the multi-scale spatiotemporal variable precipitation as stochastic processes, similar to the stochastic conceptualization of multi-scale heterogeneous geologic media (Chapter 4 in Yeh et al., 2015). Based on this process, the VAE generates many possible synthetic data, representing those spatiotemporal variating precipitations not observed at the seven stations (similar to the above MC simulation). This approach is the so-called "data augmentation" in the machine learning field.

3.2 Long Short-Term Memory (LSTM) Network

As sufficient spatially varying precipitation data over the catchment becomes available, next, the deep learning network of LSTM uses 12 consecutive months of the observed precipitation data in seven zones $[\boldsymbol{X}]_{12\times7}$, and their synthetic precipitation $\hat{X}_{12\times7\times M}$ (M is the number of synthetic data of each zone) as well as the observed spring discharge data Y_{12} from the past 12 months to exploit their spatiotemporal cross-correlations features to predict the spring discharge in the following months. Again, this is analogous to the conditional stochastic approaches of cokriging or SLE (Yeh et al., 2005).

LSTM has been widely employed for modeling time series data and proved well-suited for hydrological data, especially in Karst regions, for predictive tasks (Song et al., 2022). The core innovation of LSTM incorporates memory cells and gate mechanisms. Memory cells enable the network to store and retrieve long-term memory information, while gate mechanisms oversee and control the data read and write operations of these memory cells. These mechanisms are similar to stochastic hydrology's spatial and temporal correlation concepts (i.e., data are correlated over short or long temporal or spatial distances, depicted by the auto or cross-correlation, Yeh et al., 2015; Yeh et al., 2023). The LSTM process, controlled by its network parameters θ_{LSTM} of all the gates, acquires high-dimensional positively or negatively correlated spatiotemporal features of spring discharges for all data as $F_{\theta_{LSTM}}$. Finally, a nonlinear calculation of $F_{\theta_{LSTM}}$ by a fully connected neural network outputs the spring discharge for the following month or months denoted as *Y**, which can be given as

$$Y^* = C_{\theta_{FC}}(F_{\theta_{LSTM}}), \quad (1)$$

where θ_{FC} is the parameter matrix of the fully connected neural network. Details are available in Supporting Information S2 (Figure S2).

In our model, the VAE and LSTM are coupled. The output of VAE, the observed precipitation, and spring discharges are fed to the LSTM to predict the spring discharge. The initial parameters of θ_E, θ_D, $F_{\theta_{LSTM}}$ and θ_{FC} are zero and are iteratively updated by the coupled model training until some criteria about the agreement between the observed and synthetic precipitation and the observed and simulation spring discharge are met. Afterward, the algorithm and the optimal parameters were verified through the testing (verification) phase, where the observed discharge time series were not used in the training.

Unlike traditional statistical methods, our approach exploits the results of synthetic precipitation and spring discharge prediction work together to update the model's parameters. This update enables LSTM to explore spatiotemporal variation cross-correlation features between precipitation and spring discharge while enabling VAE to generate more realistic spatial variation precipitation. That is, it can leverage the cross-correlation between spring discharge and precipitation to generate more reasonable data instead of relying solely on the statistical process of the precipitation itself. The detailed algorithm of VAE and LSTM can be seen in the Supporting Information S1 and S2 (Figures S1 and S2).

3.3 Evaluation Metrics

To assess the predictive accuracy of the model, we employed the following evaluation metrics: root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and Nash-Sutcliffe efficiency coefficient (NSE). They are defined below:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - y_i^*)^2} \tag{2}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}| y_i - y_i^* | \tag{3}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}| (y_i - y_i^*) / y_i | \times 100\% \tag{4}$$

$$NSE = 1 - \frac{\sum_{i=1}^{N}(y_i - y_i^*)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{5}$$

The y_i in the above equations represents the observed spring discharge value, $\bar{y}$ is the mean of the observed discharges, the y_i* is the predicted, and N is the number of values used.

4 Results

4.1 Model Implementation

This study employs the proposed coupled model of VAE/LSTM to predict spring discharge. Since the parameters of the neural network model control its performance, parts of the observed data are used for the model training to obtain the optimal parameters, and the remaining is used for model testing (verification).

In this paper, the ratio of the training data to the testing data is 2:1 for each divided period (i.e., nature, overexploitation and recovery periods in Figure 2) as shown in Figure 2. Specifically, during the nature period, the training phase used the time series from 1959.1 to 1966.12, and the testing phase employed data from 1967.1 to 1970.12. During the overexploitation period, 1971.1- 1994.12 is the training phase, 1995.1-2006.12 is the testing phase. For the recovery period, 2007.1-2014.12 is the training phase, 2015.1-2019.12 is the testing phase.

In addition, the LSTM time step used in this paper is 12 months (Song et al., 2022), which is consistent with 12 month periodicity in the autocorrelation of precipitation data (Figure 4) and the cross-correlation between the precipitation and spring discharge (Figure 7). In other words, the model continuously employs cross-covariance of the past 12 months of precipitation and spring flow to predict the next month's spring discharge.

More synthetic precipitation data reduces the uncertainty of spatial precipitation distribution, which can improve the ability of the LSTM to capture the cross-correlation features between precipitation and spring discharge. However, this will lead to higher data dimension calculations and reduce the generalization ability of the LSTM to diverse data during the testing phase. Therefore, it systematically increases the proportion of synthetic data during the training

process to determine the optimal numbers of the synthetic data in each period. This incremental approach enables the model to grasp precipitations' spatiotemporal characteristics over the watershed progressively.

4.2 Effectiveness of VAE/LSTM model for Single-Step Spring Discharge Prediction

Figures 9a, b, and c show the evaluation performance of single-step spring discharge prediction with various proportions of the synthetic of the three periods in the testing phase. These figures underscore the influence of additional spatiotemporal precipitation data on the model's predictive capabilities. As shown in the figure, VAE/LSTM model yields better-predicted spring discharges as the proportion of synthetic precipitations gradually increases, which reflects that the model is exposed to an expanding pool of spatiotemporal varying precipitation data, enabling it to acquire broader spatiotemporal precipitation attributes. Furthermore, the predicted performance becomes stable when the proportion increases to a certain extent. Although there are differences in the cross-covariance between precipitation and spring discharge in different zones (as shown in Figure 7), when a certain amount of spatial precipitation data is employed, LSTM can capture their relevant features well, implying that the ergodic condition is reached.

Figure 9 displays the single-step spring discharge prediction. It shows that the evaluation of NSEs of our model in the three periods can reach 0.92, 0.94 and 0.82, respectively, indicative of the excellent performance of our model. It is worth noting that the ratios of synthetic data that make the prediction optimal are different for different periods: 7:28 in the natural period, 7:21 in the overexploitation period, and 7:42 in the recovery period. Many reasons could contribute to this variation of the three periods. One likely reason is that human intervention has disrupted the natural correlation between precipitation and spring discharge.

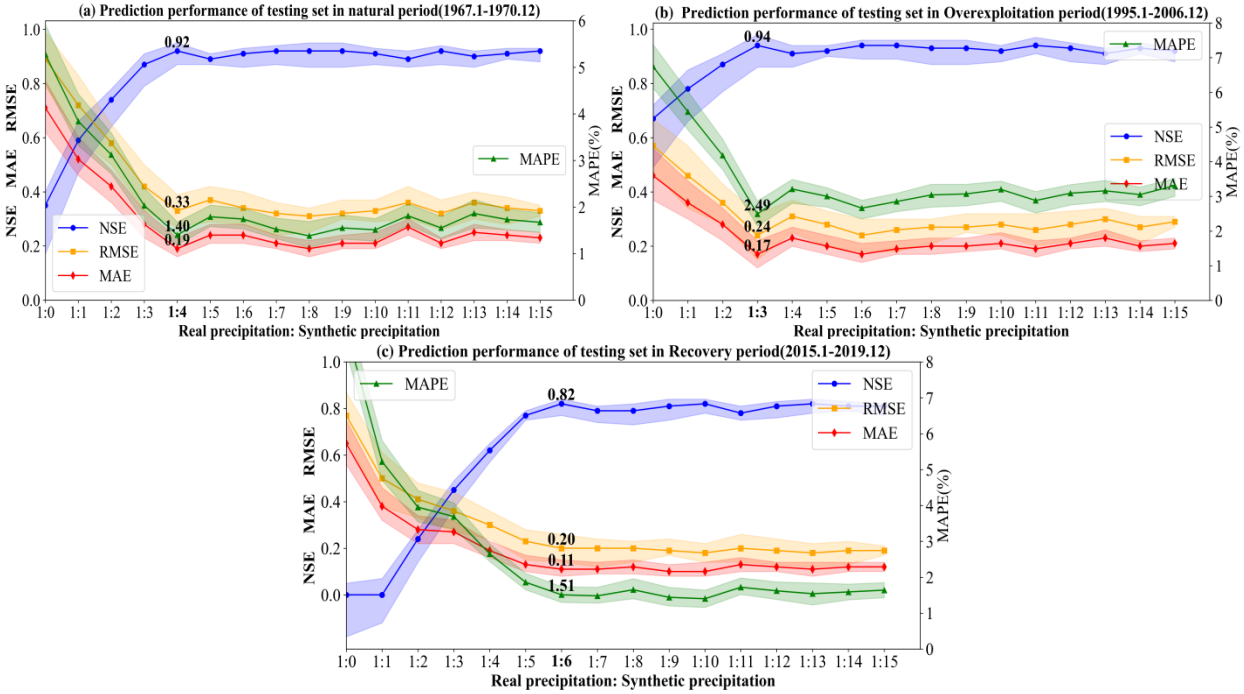

Figure 9. Performance evaluation of LSTM Model with VAE data augmentation at different data ratios. The ratio means the amount of data by which one precipitation data is enhanced.

Figure 9 also illustrates that model performance slightly deteriorates when the proportion of synthetic precipitation becomes high. The slight deterioration may be due to insufficient experiments. This result may imply that LSTM becomes overly sensitive to specific details and noise. Theoretically, more synthetic precipitations with corrected spatiotemporal statistics should yield stable predictions.

Figure 10 visually compares the observed and the single-step predicted spring discharge in the training and testing phases for the three periods. Figure 10a-c are the results of the natural period (1959.1-1970.12). Figure 10d-f are the results of the overexploitation period (1971.1-2006.12). Figure 10g-i are the results of the recovery period (2007.1-2019.12). In these figures, the red curve is the observed spring discharge, the blue curve denotes the predicted spring discharge during training, and the green curve represents the predicted spring discharge in the testing phase. The scatter plots compare the observed and predicted discharges in the training

and testing data set. The shaded area (in Figures 10a, 10d, and 10g) and the area between the red lines in the scatter plots are the predictions where their NSEs are with a 95% confidence interval. That means that when evaluating the model, we can have a 95% confidence level to believe that the real NSE value falls within the calculated confidence interval. This range provides a statistical measure of the model's performance and helps to determine the model's credibility under different conditions.

These figures indicate that the NSEs of training for the three periods are all higher than 0.9. Further, for the testing phase, the NSEs of the three periods reaches 0.92, 0.94, and 0.82, respectively. The R^2 of the scatter plots comparing predicted and observed spring discharge of the three periods for the training phase are 0.97, 0.96, and 0.88. They are 0.92, 0.95, and 0.84 for the testing phase. These results demonstrate outstanding predictive performance during the testing periods. For the different periods, these results show some differences. Such differences arise because these results depend on LSTM's training, which is influenced by the amount of observed data in each period and the differences in data distribution between the testing and training phases. As shown in Figure 2, the amount of observed data during each period and the distribution also varies.

Table 3 presents NSE, RMSE, MAE, and MAPE values for the VAE/LSTM model during the training and testing phases of spring discharge for each of the three periods, using optimal synthetic precipitation ratios. These results underscore the robustness of the proposed model, which is enhanced by precipitation data augmentation (considering spatiotemporal variation) through VAE.

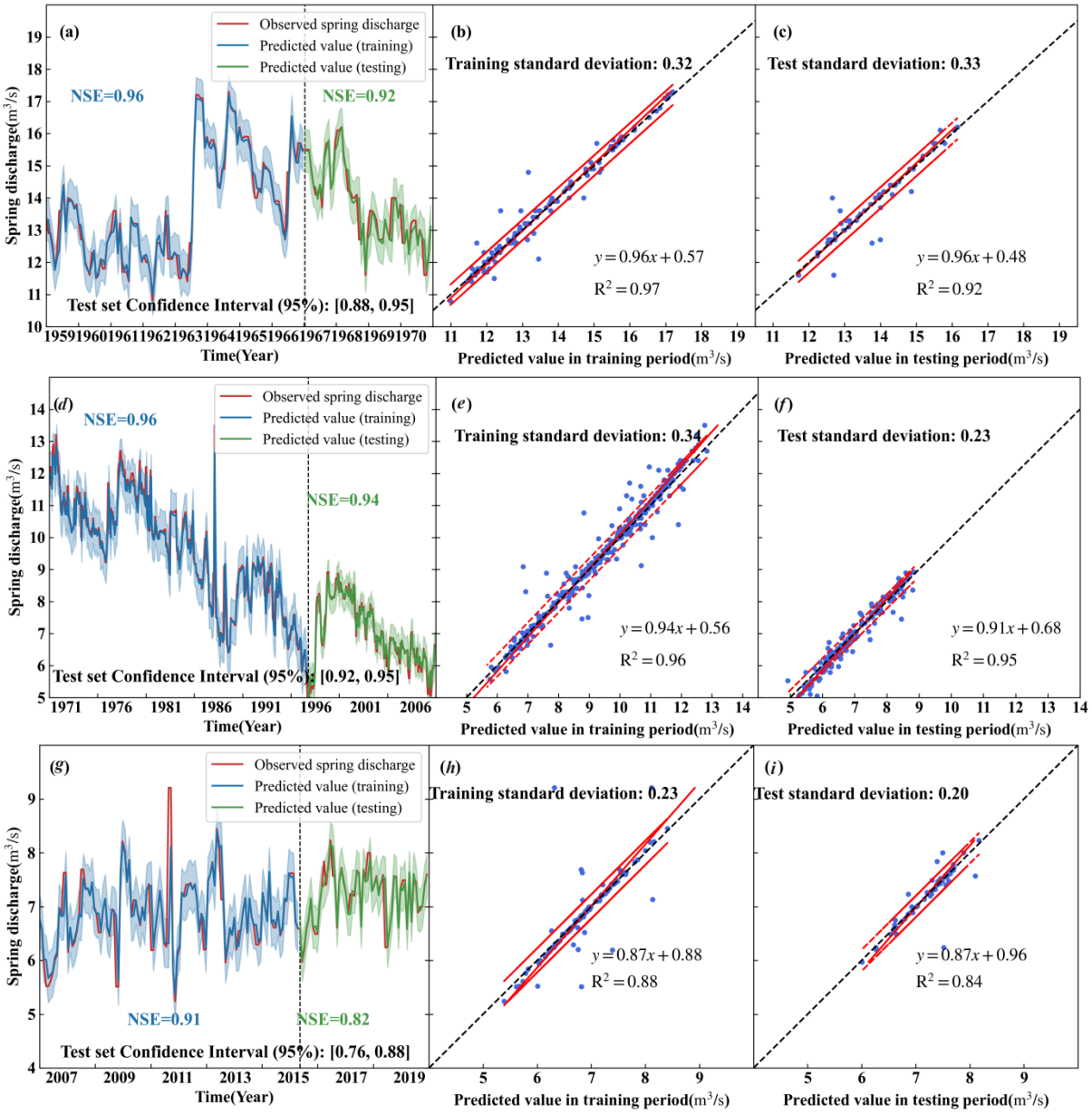

Figure 10. Comparison between observed and predicted spring discharge in training and testing for the three periods.

485 **Table 3.** Performance of VAE/LSTM for single-step spring discharge prediction during training
486 and testing phases of the three periods.

Period	Data Ratio	Phase	NSE	RMSE	MAE	MAPE(%)
Natural	1:4	Training	0.96	0.32	0.18	1.31
1959-1970		Testing	0.92	0.33	0.19	1.40
Overexploitation	1:3	Training	0.96	0.37	0.19	1.97
1971-2006		Testing	0.94	0.24	0.17	2.49
Recovery	1:6	Training	0.91	0.20	0.09	1.31
2007-2019		Testing	0.82	0.20	0.11	1.51

487 4.3 Effectiveness of VAE/LSTM Model for Multi-Step Spring Discharge Prediction

488 Next, this paper expands its predictive horizon from forecasting one month of spring
489 discharge to predicting spring discharge for the next six months. The results for the test data set
490 of the three periods depicted in Figure 11 show a gradual decrease in the NSE values as the
491 prediction step extends. When comparing the prediction outcomes across the three periods, it is
492 evident that the NSE values for one-month and six-month predictions are as follows: 0.92 and
493 0.28 (Natural period), 0.94 and 0.45 (overexploitation period) and 0.82 and 0.07 (Recovery
494 period), respectively, which have a reduction of 69.57%, 52.13%, and 91.46%, correspondingly.
495 This pattern might be linked to the climatic conditions of the Niangziguan spring. Niangziguan
496 Spring is in northern China and is influenced by the northern monsoon and continental climate
497 (He et al., 2019). During abundant rainfall, increased surface runoff can diminish the proportion
498 of spring recharge, thereby impacting spring discharge (Fiorillo et al., 2010). Furthermore,
499 predicting the next six months encompasses the summer high-temperature season, when rising
500 temperatures can result in the evaporation of certain springs. Evaporative losses curtail the

effective recharge of spring (Mustafa et al., 2015), consequently diminishing the accuracy of forecasting future spring discharge.

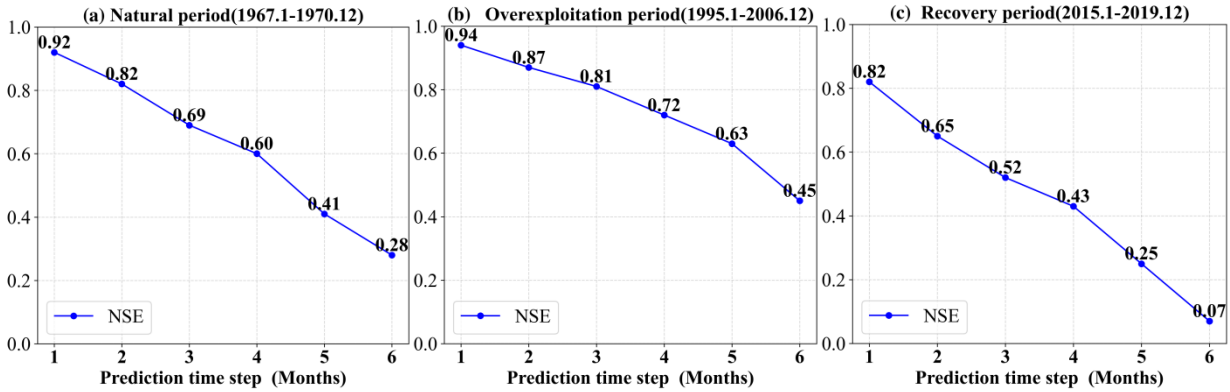

Figure 11. NSE performance for different forecast horizons in multi-step prediction of spring discharge for the three periods.

In contrast to single-step prediction, multi-step prediction provides longer-term projections, aiding in exploring the evolving patterns of karst spring discharge. However, considering long-term climate trends and seasonal variations at Niangziguan Spring, expanding the prediction horizon may include a greater range of seasons and climate variations. Different seasons may exhibit significant differences in precipitation patterns, increasing the complexity of predictions. Hence, choosing an excessively long prediction horizon may result in less accurate long-term LSTM predictions than shorter ones. Opting for smaller prediction horizons, such as two and three months, yields better predictive accuracy (Table 4). Smaller prediction timeframes encompass similar seasons and climatic conditions, aiding in capturing the trends in spring discharge more effectively. The previous cross-covariance analysis (Figure 7) should also explain these results.

518 **Table 4.** Performance of VAE/LSTM for 2 months and 3 months spring discharge predictions on
519 the test data set for the three periods using the optimal parameters

Period	Prediction Steps	NSE	RMSE	MAE	MAPE(%)
Natural	2 Months	0.82	0.53	0.43	3.14
1967.1-1970.12	3 Months	0.69	0.64	0.53	3.92
Overexploitation	2 Months	0.87	0.36	0.28	4.01
1995.1-2006.12	3 Months	0.81	0.44	0.34	5.30
Recovery	2 Months	0.65	0.26	0.18	2.46
2015.1-2019.12	3 Months	0.52	0.32	0.21	2.88

520 **5 Discussions**

521 5.1 Comparing VAE-Augmented Deep Learning Models in Spring Discharge Prediction

522 This study also evaluates various deep learning models (RNN, ANN, GCN, and
523 Transformer) to validate the universality of the data augmentation method proposed and its
524 potential value in other hydrological applications. The introduction of these approaches are
525 available in supporting information S3 and Figure S3. The experimental results in Table 5
526 demonstrate that various deep learning models, following precipitation enhancement, achieve
527 favorable single-step predictive outcomes. It is crucial to emphasize that the LSTM model yields
528 the best predictive results (Table 3). This is attributed to its outstanding capability for time series
529 modeling, its internal state to retain past information, and its proficiency in extracting temporal
530 features. Therefore, it remains the preferred model for handling hydrological data, especially in
531 karst terrains.

532

533 **Table 5.** Performance comparison of various models with VAE based data augmentation

Models	Periods	NSE	RMSE	MAE	MAPE(%)
RNN	Original(No Periods)	0.33	0.54	0.38	5.74
	Augmented(natural)	0.92	0.33	0.18	1.34
	Augmented(Overexploitation)	0.95	0.23	0.14	2.06
	Augmented(Recovery)	0.82	0.21	0.11	1.56
	Augmented(Mean)	**0.89**	**0.26**	**0.14**	**1.65**
ANN	Original(No Periods)	-0.42	0.79	0.64	9.84
	Augmented(natural)	0.77	0.55	0.41	3.02
	Augmented(Overexploitation)	0.81	0.44	0.32	4.90
	Augmented(Recovery)	0.77	0.23	0.13	1.83
	Augmented(Mean)	**0.78**	**0.41**	**0.29**	**3.25**
GCN	Original(No Periods)	0.05	0.64	0.43	6.41
	Augmented(natural)	0.73	0.59	0.44	3.13
	Augmented(Overexploitation)	0.82	0.43	0.31	4.82
	Augmented(Recovery)	0.60	0.31	0.15	2.11
	Augmented(Mean)	**0.71**	**0.44**	**0.30**	**3.35**
Transformer	Original(No Periods)	-0.15	0.71	0.57	8.23
	Augmented(natural)	0.90	0.36	0.25	1.89
	Augmented(Overexploitation)	0.87	0.36	0.26	3.97
	Augmented(Recovery)	0.78	0.23	0.15	2.14
	Augmented(Mean)	**0.85**	**0.32**	**0.22**	**2.67**

Specifically, RNN and LSTM excel in handling time series data because they can capture temporal dependencies within hydrological data through internal states. Conversely, ANN is typically designed as a feedforward neural network and lacks internal states for handling temporal dependencies, potentially resulting in inferior performance when dealing with time series data.

Furthermore, GCN can be employed in hydrological prediction for spatiotemporal graph data, where nodes represent distinct geographical locations, and edges represent connections between these locations. GCN's strength lies in its capacity to consider the relationships between geographical locations (Gai et al., 2023). However, in this study, the data primarily focuses on time series information, leading to slightly lower predictive performance for GCN than LSTM.

Simultaneously, the predictive performance of the Transformer model is not as satisfactory as that of LSTM. This discrepancy may arise because Transformers tend to perform better under conditions of large-scale datasets (Vaswani et al., 2017), whereas the dataset in this research comprises only thousands of records even after augmentation. Therefore, Transformer models may struggle to leverage their advantages when dealing with smaller datasets. In summary, in the context of sparse and uncertain hydrological data, the approach presented in this study significantly improves the performance of various deep-learning models in hydrological prediction.

5.2 Performance Comparison of Various Data Augmentation Approaches

This study compares the performance of the LSTM model using various data augmentation approaches. The results of performance metrics are shown in Table 6. We initiated experiments for spring discharge prediction using linear regression and ARIMA (AutoRegressive

Integrated Moving Average) models. These traditional prediction models yielded unsatisfactory results (as indicated in the first two rows of the table). This outcome is attributed to the fact that, in hydrological forecasting, the relationship between spring discharge and various factors, including meteorological and hydrological variables, is often nonlinear and nonstationary.

Furthermore, this involves statistical methods, like augmenting precipitation data to conform to a uniform distribution of its extremes (Fourth row of Table 6). It also includes precipitation augmentation using the tophat kernel function and density estimation (fifth row of Table 6). Linear interpolation is applied by estimating values between known precipitation data points, resulting in new augmented values (sixth and seventh row of Table 6). The methods for data augmentation using deep learning include the LSTM model with direct data augmentation without period division (eighth row of the Table 6) and the LSTM model with a two-period division before and after human activities in 1971 for data augmentation (ninth row of the Table 6). The experimental results highlight that the LSTM model enhanced by VAE exhibits lower errors in spring discharge prediction than the baseline LSTM model. Moreover, it demonstrates superior predictive performance compared to traditional data augmentation methods.

Traditional statistical methods often rely on a single distribution estimate, which proves challenging for capturing the intricate spatiotemporal characteristics of precipitation due to its inherent complexity and uncertainty. In contrast, the VAE can derive joint latent distributions to learn the nuanced representation of data by the encoding-decoding process, resulting in improved preservation of temporal and spatial correlations in precipitation. Unlike precipitation interpolation methods, which are limited by local information and may lead to information loss in global spatiotemporal correlations, VAE excels at comprehensively preserving the multimodal nature and global correlations of precipitation data, thereby providing enhanced information for

579 the LSTM. In augmenting precipitation with VAE across various periods, the enhanced model
580 improves the NSE to 0.55 (NSE=0.38 vs. NSE=0.93) and MAE by 0.25 (MAE=0.40 vs.
581 MAE=0.15). Moreover, the decrease in RMSE (RMSE=0.54 vs. RMSE=0.15) further
582 substantiates the superior fit between the enhanced model and actual observations. Consequently,
583 the three-period LSTM model with data augmentation yields enhanced performance in spring
584 discharge prediction. Additionally, period-wise predictions based on regression analysis
585 significantly augment the overall predictive capabilities of the model.

586 **Table 6.** LSTM model validation with different data augmentation strategies on validation set.

Data Augmentation Approach	NSE	RMSE	MAE	MAPE(%)
LinearRegression	0.04	1.53	1.26	9.04
ARIMA	-2.84	1.15	1.01	13.93
Only LSTM (Baseline)	0.38	0.54	0.40	6.03
LSTM with precipitation Uniform Distribution	-0.51	4.13	3.36	32.42
LSTM with Kernel Density Estimation (tophat)	0.67	0.65	0.49	3.63
LSTM with Linear Interpolation (No Periods)	0.49	0.87	0.69	5.04
LSTM with Linear Interpolation (Three-Period)	0.80	0.50	0.41	3.02
LSTM with VAE Data Augmentation (No Periods)	0.57	0.39	0.22	3.10
LSTM with VAE Two-Period Data Augmentation	0.72	0.31	0.15	2.11
LSTM with VAE Three-Period Data Augmentation	**0.93**	**0.26**	**0.15**	**1.80**

5.3 Visualization and Quantitative Results of the VAE

Next, we demonstrate that the VAE encoder captures the posterior probability distribution, mean, and autocovariance functions of observed precipitation data, and then the decoder generates the synthetic data with the same distribution, mean, and autocovariance functions. Figure 12a is the scatter plots showing that the synthetic series are unbiased compared with the observed data for the seven stations from 1959-2019. Figure 12b demonstrates that the probability distribution of the synthetic time series is consistent with that of the observed.

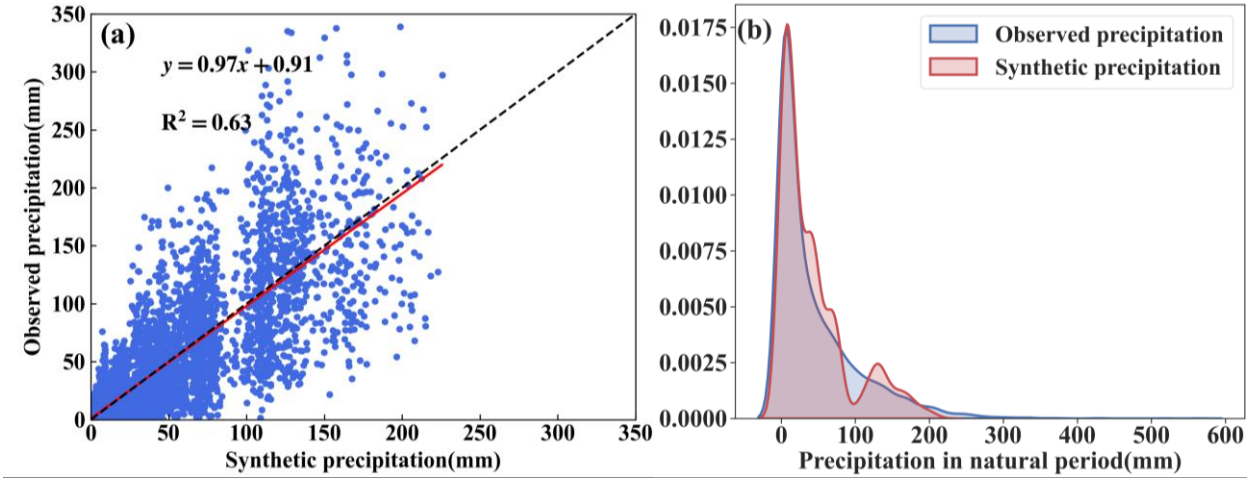

Figure 12. The observed and synthetic precipitation distribution for the seven stations from 1959-2019.

Figure 13 displays the correlations between the augmentation data at seven precipitation observation stations. Compared with Figure 5, the synthetic data reduce the gap of spatial precipitation and thus enhance the correlation between the precipitation observation points. Notice that intuitively, hourly or daily precipitations may vary throughout the region. The cumulative precipitations over a month likely filter out the spatial variability in precipitation. As such, the correlation between monthly precipitation data at different locations in the catchment is similar, indicating that the monthly precipitation spatial pattern over the entire catchment is

almost uniform (small spatial variability). This result may suggest that the temporal variability of precipitation may play a more critical role in predicting the spring discharge than its spatial variability. Figure 14 shows the temporal auto-correlation of the seven stations' observed and synthetic precipitation time series (from 1959 to 2019). They illustrate that the VAE also enhances the temporal correlation of the precipitation.

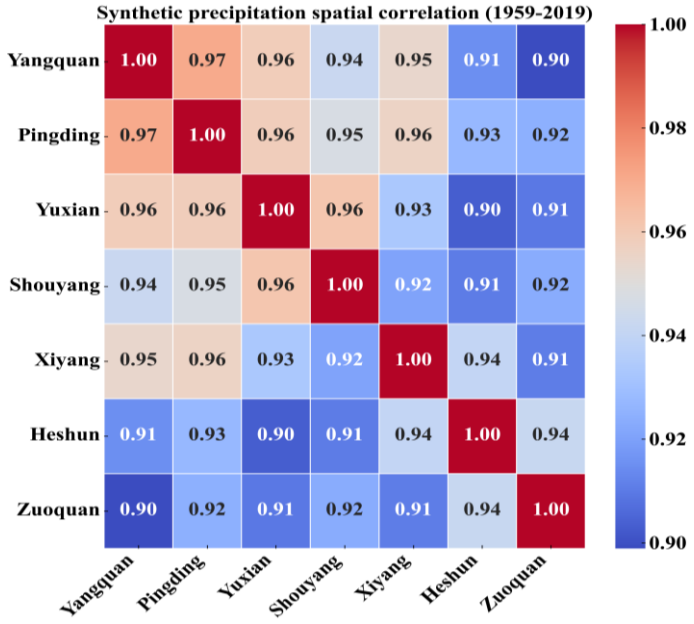

Figure 13. The correlations between the augmentation data at 7 precipitation observation stations

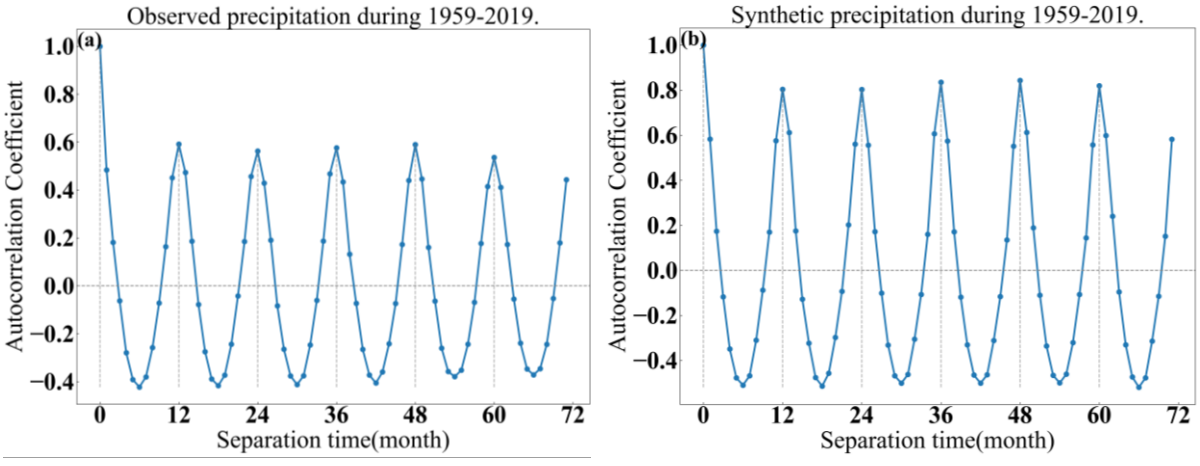

Figure 14. The temporal auto-correlation of the observed and synthetic precipitation (from 1959 to 2019). (a) is for the observed precipitation. (b) is for the synthetic precipitation.

6 Conclusions

This study utilized a VAE model to generate synthetic precipitation data and applied it to the simulation of spring discharge in data-sparse karst regions. Augmenting the precipitation data improved the learning capabilities and predictive performance of various deep learning models (e.g., LSTM, RNN, ANN, GCN, Transformer, see supplementary section) for spring discharge prediction.

We applied the VAE-augmented precipitation to the LSTM model for three periods of Niangziguan Spring in China: a natural period, a subsequent overexploitation period, and finally, a recovery period. The results of the augmented model demonstrated a significant advantage in single-step spring discharge prediction. When validated against actual observations, the augmented model exhibited notably higher predictive accuracy than the baseline LSTM model.

In the multi-step prediction of LSTM after data augmentation, which considered spring discharge across three distinct periods, this study had revealed that opting for smaller prediction horizons (e.g., two and three months) leaded to enhanced predictive accuracy. We concluded that these narrower prediction timeframes encompassed more closely aligned seasons and climatic conditions, thus facilitating the model in capturing the patterns in spring discharge more effectively.

Our precipitation data augmentation strategy was further validated through multi-model generalization experiments, demonstrating the versatility and effectiveness of this approach in addressing challenges related to data scarcity in hydrology, particularly in regions with limited data availability. This strategy is not restricted to specific LSTM models but can be extended to a broader range of models depending on various hydrological contexts and research questions.

While we applied this strategy to a karst spring, its generality allows it to meet the demands of diverse research inquiries across different domains. It offers researchers a novel solution to address issues related to data scarcity and sampling difficulties, enabling them to select appropriate models for prediction and analysis tailored to their specific requirements. We present a fresh perspective on handling the challenges of data scarcity and uncertainty in hydrological data.

Acknowledgments

The work of Yonghong Hao is partially supported by the National Natural Science Foundation of China 42072277, 41272245, 40972165, 42307088 and 40572150. Chunmei Ma is partially supported by the Scientific Research Project of Tianjin Education Commission under Grant (No.2021KJ186). Yeh is partially supported by US NSF grant 000316729.

Open research

Data Availability Statement

The test data and test codes of our hybrid model are available at: https://github.com/csmcm/spring-discharge-prediction.

References

Addor, N., Do, H. X., Alvarez-Garreton, C., Coxon, G., Fowler, K., & Mendoza, P. A. (2020). Large-sample hydrology: recent progress, guidelines for new datasets and grand challenges. *Hydrological Sciences Journal*, 65(5), 712-725.

An, L., et al. 2020. Simulation of karst spring discharge using a combination of time–frequency analysis methods and long short-term memory neural networks. *Journal of hydrology*, 589, 125320.

Bruckmann, J., & Clauser, C. (2020). Ensemble-based stochastic permeability and flow simulation of a sparsely sampled hard-rock aquifer supported by high performance computing. *Hydrogeology Journal*, 28(5), 1853-1869.

Çallı, S. S., Çallı, K. Ö., Yılmaz, M. T., & Çelik, M. (2022). Contribution of the satellite-data driven snow routine to a karst hydrological model. *Journal of Hydrology*, 607, 127511.

Chen, Q., Cui, Z., Liu, G., Yang, Z., & Ma, X. (2022). Deep convolutional generative adversarial networks for modeling complex hydrological structures in Monte-Carlo simulation. *Journal of Hydrology*, 610, 127970.

Cheng, S., Qiao, X., Shi, Y., & Wang, D. (2021). Machine learning for predicting discharge fluctuation of a karst spring in North China. *Acta Geophysica*, 69, 257-270.

Dugdale, S. J., Klaus, J., & Hannah, D. M. (2022). Looking to the skies: realising the combined potential of drones and thermal infrared imagery to advance hydrological process understanding in headwaters. *Water Resources Research*, 58(2), e2021WR031168.

Fiorillo, F., & Guadagno, F. M. (2010). Karst spring discharges analysis in relation to drought periods, using the SPI. *Water resources management*, 24, 1867-1884.

Gai, Y., Wang, M., Wu, Y., Wang, E., Deng, X., Liu, Y.,Tian-Chyi Jim Yeh, & Hao, Y. (2023). Simulation of spring discharge using graph neural networks at Niangziguan Springs, China. *Journal of Hydrology*, 130079.

Ghorbanidehno, H., Kokkinaki, A., Lee, J., & Darve, E. (2020). Recent developments in fast and scalable inverse modeling and data assimilation methods in hydrology. *Journal of Hydrology*, 591, 125266.

683 Goldscheider, N., Chen, Z., Auler, A. S., Bakalowicz, M., Broda, S., Drew, D., & Veni, G.
684 (2020). Global distribution of carbonate rocks and karst water resources. *Hydrogeology Journal*,
685 28, 1661-1677.

686 Grundmann, J., Hörning, S., & Bárdossy, A. (2019). Stochastic reconstruction of spatio-
687 temporal rainfall patterns by inverse hydrologic modelling. *Hydrology and Earth System*
688 *Sciences*, 23(1), 225-237.

689 Han D, Xu H, Liang X. 2006. GIS-based Regionalization of a karst water system in Xishan
690 Mountain area of Taiyuan Basin, North China. *Journal of Hydrology*, 331(3–4): 459–470.

691 Hao, H., Hao, Y., Liu, Y., Yeh, T. C. J., Zhang, M., Wang, Q., & Fan, Y. (2023). Anomaly
692 of glacier mass balance in different vertical zones and responses to climate modes: Urumqi
693 Glacier No. 1, China. *Climate Dynamics*, 60(1-2), 493-509.

694 Hao, Y., Huo, X., Duan, Q., Liu, Y., Fan, Y., Liu, Y., & Yeh, T. C. J. (2015). A Bayesian
695 analysis of nonstationary generalized extreme value distribution of annual spring discharge
696 minima. *Environmental Earth Sciences*, 73, 2031-2045.

697 Hao, Y., Zhang, J., Wang J., Li, R., Hao, P., & Zhan, H. (2016). How does the
698 anthropogenic activity affect the spring discharge? *Journal of hydrology*, 540, 1053-1065.

699 Hao Y, Zhu Y, Zhao Y, Wang W, Du X, Yeh TJ. 2009. The role of climate and human
700 influences in the dry-up of the Jinci Springs, China. *Journal of the American Water Resources*
701 *Association*, 45(5): 1228–1237.

702 Hartmann, A., Goldscheider, N., Wagener, T., Lange, J., & Weiler, M. (2014). Karst water
703 resources in a changing world: Review of hydrological modeling approaches. *Reviews of*
704 *Geophysics*, 52(3), 218-242.

705 He, X., Wu, J., & Guo, W. (2019). Karst spring protection for the sustainable and healthy
706 living: the examples of Niangziguan spring and Shuishentang spring in Shanxi, China. *Exposure*
707 *Health*, 11, 153-165.

708 Juki, D. , & Deni-Juki, V. (2009). Groundwater balance estimation in karst by using a
709 conceptual rainfall-runoff model. *Journal of Hydrology*, 373(3-4), 302-315.

710 Labat, D., Ababou, R., Mangin, A., 2000. Rainfall-runoff relations for karstic springs. Part I:
711 convolution and spectral analyses. *Journal of Hydrology*, 238, 123–148.

712 Lin, M., Teng, S., Chen, G., & Bassir, D. (2023). Transfer Learning with Attributes for
713 Improving the Landslide Spatial Prediction Performance in Sample-Scarce Area Based on
714 Variational Autoencoder Generative Adversarial Network. *Land*, 12(3), 525.

715 Liu, B., et al. 2020. Measurement of sustainable transformation capability of resource-based
716 cities based on fuzzy membership function: A case study of Shanxi Province, China. *Resources*
717 *Policy*, 68, 101739.

718 Mengistu, A. G., Woldesenbet, T. A., & Dile, Y. T. (2022). Evaluation of observed and
719 satellite-based climate products for hydrological simulation in data-scarce Baro-Akob River
720 Basin, Ethiopia. *Ecohydrology & Hydrobiol*ogy, 22(2), 234-245.

721 Mustafa, O., Merkel, B., & Weise, S. M. (2015). Assessment of hydrogeochemistry and
722 environmental isotopes in karst springs of Makook Anticline, Kurdistan Region, Iraq. *Hydrology*,
723 2(2), 48-68.

724 Shorten, C., Khoshgoftaar, T. M., & Furht, B. (2021). Text data augmentation for deep
725 learning. *Journal of big Data*, 8, 1-34.

726 Smith, R., & Li, J. (2021). Modeling elastic and inelastic pumping-induced deformation
727 with incomplete water level records in Parowan Valley, Utah. *Journal of Hydrology*, 601,
728 126654.

729 Song, X., Hao, H., Liu, W., Wang, Q., An, L., Yeh, T. C. J., & Hao, Y. (2022). Spatial-
730 temporal behavior of precipitation driven karst spring discharge in a mountain terrain. *Journal of*
731 *Hydrology*, 612, 128116.

732 Stevanović, Z. (2019). Karst waters in potable water supply: a global scale overview.
733 *Environmental Earth Sciences*, 78(23), 662.

734 Sun, H., & Su, F. (2020). Precipitation correction and reconstruction for streamflow
735 simulation based on 262 rain gauges in the upper Brahmaputra of southern Tibetan Plateau.
736 *Journal of Hydrology*, 590, 125484.

737 Tang, T., Jiao, D., Chen, T., & Gui, G. (2022). Medium-and long-term precipitation
738 forecasting method based on data augmentation and machine learning algorithms. *IEEE Journal*
739 *of Selected Topics in Applied Earth Observations and Remote Sensin*g, 15, 1000-1011.

Varouchakis, E. A., & Hristopulos, D. T. (2013). Comparison of stochastic and deterministic methods for mapping groundwater level spatial variability in sparsely monitored basins. *Environmental Monitoring&Assessment* 185: 1-9.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing system*s, 30.

Wunsch, A., Liesch, T., Cinkus, G., Ravbar, N., Chen, Z., Mazzilli, N., & Goldscheider, N. (2022). Karst spring discharge modeling based on deep learning using spatially distributed input data. *Hydrology and Earth System Sciences*, 26(9), 2405-2430.

Yaseen, Z. M. , El-Shafie, A. , Jaafar, O. , Afan, H. A. , & Sayl, K. N. . (2015). Artificial intelligence based models for stream-flow forecasting: 2000-2015. *Journal of Hydrology*, 530, 829-844.

Yin, H., Wang, F., Zhang, X., Zhang, Y., Chen, J., Xia, R., & Jin, J. (2022). Rainfall-runoff modeling using long short-term memory based step-sequence framework. *Journal of Hydrology*, 610, 127901.

Zhou, R., & Zhang, Y. (2022). On the role of the architecture for spring discharge prediction with deep learning approaches. *Hydrological Processes*, 36(10), e14737.

Zhu, J., & Yeh, T. C. J. (2005). Characterization of aquifer heterogeneity using transient hydraulic tomography. *Water Resources Research*, 41(7): W07028.