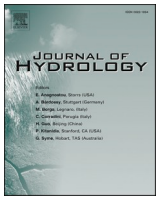
Journal of Hydrology 645 (2024) 132235

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol



Research papers

Interpretable multi-step hybrid deep learning model for karst spring discharge prediction: Integrating temporal fusion transformers with ensemble empirical mode decomposition

Renjie Zhou [a,*], Quanrong Wang [b], Aohan Jin [b], Wenguang Shi [b], Shiqi Liu [c]

[a] Department of Environmental and Geosciences, Sam Houston State University, Huntsville, TX 77340, USA
[b] School of Environmental Studies, China University of Geosciences, Wuhan, Hubei 430074, PR China
[c] Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, PR China

ARTICLE INFO

This manuscript was handled by Yuefei Huang, Editor-in-Chief

Keywords:
Temporal Fusion Transformers (TFT)
Transformers
Ensemble Empirical Mode Decomposition (EEMD)
Deep learning
Rainfall-runoff relationship
Karst Hydrology

ABSTRACT

Karst groundwater is a critical freshwater resource for numerous regions worldwide. Monitoring and predicting karst spring discharge is essential for effective groundwater management and the preservation of karst ecosystems. However, the high heterogeneity and karstification pose significant challenges to physics-based models in providing robust predictions of karst spring discharge. In this study, an interpretable multi-step hybrid deep learning model called selective EEMD-TFT is proposed, which adaptively integrates temporal fusion transformers (TFT) with ensemble empirical mode decomposition (EEMD) for predicting karst spring discharge. The selective EEMD-TFT hybrid model leverages the strengths of both EEMD and TFT techniques to learn inherent patterns and temporal dynamics from nonlinear and nonstationary signals, eliminate redundant components, and emphasize useful characteristics of input variables, leading to the improvement of prediction performance and efficiency. It consists of two stages: in the first stage, the daily precipitation data is decomposed into multiple intrinsic mode functions using EEMD to extract valuable information from nonlinear and nonstationary signals. All decomposed components, temperature and categorical date features are then fed into the TFT model, which is an attention-based deep learning model that combines high-performance multi-horizon prediction and interpretable insights into temporal dynamics. The importance of input variables will be quantified and ranked. In the second stage, the decomposed precipitation components with high importance are selected to serve as the TFT model's input features along with temperature and categorical date variables for the final prediction. Results indicate that the selective EEMD-TFT model outperforms other sequence-to-sequence deep learning models, such as LSTM and single TFT models, delivering reliable and robust prediction performance. Notably, it maintains more consistent prediction performance at longer forecast horizons compared to other sequence-to-sequence models, highlighting its capacity to learn complex patterns from the input data and efficiently extract valuable information for karst spring prediction. An interpretable analysis of the selective EEMD-TFT model is conducted to gain insights into relationships among various hydrological processes and analyze temporal patterns.

1. Introduction

Karst aquifers play a crucial role in serving freshwater to numerous regions worldwide, contributing approximately 25% of global groundwater resource and nearly 40% of the U.S. freshwater resource (Goldscheider et al., 2020). In comparison to other aquifer types, karst aquifers are known for their highly heterogeneous hydraulic properties and complex hydrological behaviors caused by the dissolution of soluble bedrocks and distinctive subsurface features including caves, sinkholes, and conduits (Hartmann et al., 2014). The interconnections of various porous spaces including micropores, fissures, fractures and conduits create complex networks of preferential flow pathways that are difficult to characterize, further increasing the heterogeneity and complexity of the karst system. Consequently, understanding and managing karst aquifers present distinct challenges (Bakalowicz, 2005).

The karst spring discharge data provides valuable information for

* Corresponding author.
E-mail address: renjie.zhou@shsu.edu (R. Zhou).

https://doi.org/10.1016/j.jhydrol.2024.132235
Received 2 May 2024; Received in revised form 10 September 2024; Accepted 8 October 2024
Available online 21 October 2024

understanding hydrodynamics of karst aquifers and managing the freshwater resources in karst environments (Duran et al., 2020). Karst spring discharge can be affected by various factors, including climate variability, anthropogenic activities, the inherent heterogeneity of the karst system, and hierarchical permeability structures (Ghasemizadeh et al., 2012; Zhou and Zhang, 2023a). As a result, predicting karst spring discharge patterns requires specialized approaches and methodologies capable of addressing these complexities. A variety of physics-based hydrological models are developed to simulate the discharge values of karst springs (de Rooij et al., 2013; Duran et al., 2020; Fleury et al., 2007). For example, de Rooij et al. (2013) provided a physics-based distributed model with a flexible spatial discretization to incorporate the complicated conduit networks and simulate flow in the karst system. Such a discrete-continuum model coupled flow between various domains including rock matrix, overland, conduit and channels for simulating flow dynamics within karst environments. Birk et al. (2006) investigated how spring responses reflected geometric properties of the conduit system and proposed a process-based model coupling a pipe-network model with MODFLOW to simulate flow and transport in the karst systems. The physics-based models can estimate the physical dynamics and offer valuable insights into hydrological processes within karst environments. However, these models often requires extensive input data, sophisticated mathematical tools and a detailed characterization of internal structures, such as the geometry of the conduit system and the interaction between conduits and the fractured rock matrix, which are challenging to obtain and characterize for the karst system due to karstification and its tremendous heterogeneity (Sezen et al., 2019). Thus, while physics-based models can potentially provide valuable insights, their practical use in karst environments can be limited by these constraints. In addition to physics-based models, empirical models, such as traditional regression models and machine learning models, are widely used in hydrological studies as they exhibit strong data adaptability and establish direct mapping relationships between input features and the output as a "black box" without a detailed characterization of internal structures and physical processes (Jin et al., 2024a). These qualities make data-driven models an alternative method to tackle complex hydrological behaviors in environments where comprehensive data and internal structures are difficult to obtain (Gao et al., 2020). Karst systems present unique challenges for spring discharge prediction. While spring discharge is influenced by rainfall, the complex hydrological systems within karst landscapes, characterized by complex behaviors, multiple discharge sources, and intricate responses, often make traditional time series regression models insufficient (Goldscheider et al., 2014; Hartmann et al., 2014; Zhou and Zhang, 2023a). For example, spring discharge in the karst system is often nonlinearly related to precipitation due to complex flow paths, variable storage capacities, and threshold effects (Labat et al., 2000). A small increase in rainfall might lead to a disproportionately large increase in spring discharge. Rainfall and spring discharge in the karst system often exhibit nonstationarity due to seasonal variations, long-term climate trends, and changes in land use or karst system properties, leading to changes of their statistic properties over time.

In recent decades, machine learning and deep learning models have been extensively developed and employed for analyzing and predicting runoff in various hydrological systems, such as support vector machine (SVM) (Bray and Han, 2004; Feng et al., 2020) and artificial neural network (ANN) (An et al., 2020; Cui et al., 2022; Gao et al., 2020; Zhou et al., 2024; Zhou and Zhang, 2023b). In particular, various ANN architectures, including recurrent neural network (RNN) and its variants, were specifically designed for predicting time sequential data with high computational efficiency and accuracy, and had achieved good performance for hydrological models. For example, Zhang et al. (2021) employed deep RNN models to predict daily runoff at Muskegon River and Pearl River with meteorological data and its principal component analysis (PCA) components as inputs. They concluded that deep RNN models with multiple meteorological input data obtained higher prediction accuracy compared to single meteorological data input. Zhou and Zhang (2022a) proposed hybrid deep learning models that combines a long short-term memory model (LSTM), a gated recurrent unit model (GRU) and a simple RNN model with an encoder-decoder architecture for predicting runoff. The findings suggested that the deep learning models with an encoder-decoder architecture obtained superior prediction performance compared to those lacking such an architecture. In addition to RNN models, Transformer models proposed by Vaswani et al. (2017) have become widely adopted for predicting and analyzing sequential data because of its excellent ability to capture complex dependencies and patterns within data. Yin et al. (2022) proposed a runoff-rainfall model based on Transformer and the attention mechanism called RR-Former. It created direct connections between two arbitrary positions in time series using the attention mechanism and provided more flexibility than RNN models. Lim et al. (2021) introduced the Transformer-based Temporal Fusion Transformer model (TFT), which utilized the self-attention mechanism to capture the complex temporal dynamics of multi-horizon time sequences. This innovative TFT model demonstrated substantial performance improvement over other sequence-to-sequence models in established benchmarks. Compared to Local Interpretable Model-agnostic Explanations (LIME) and other Explainable AI (XAI) methods which do not account for the types and the order of input features, the TFT model can tackle multiple different input types in multi-horizon forecasting and offer interpretable insights into temporal dynamics by assessing the significance of each input variable in prediction outcomes. The TFT model has been successfully applied to multiple disciplines as a valuable tool for predicting and understanding the underlying patterns (Wu et al., 2022; Zhang et al., 2022). For example: Wu et al. (2022) employed the TFT model for wind speed forecasting and obtained satisfying performance. The model ranked the importance of multiple distinct meteorological data, while also offering an attention-based analysis of various forecast horizons.

Although deep learning models have shown impressive results in handling various sequence modeling tasks, the inherent strong nonlinearity and nonstationarity in complex systems can significantly impact the accuracy and robustness of predictions (Liu et al., 2019). Multiple time–frequency analysis methods, such as empirical model decomposition (EMD), ensemble empirical model decomposition (EEMD), and singular spectrum analysis (SSA) were employed to preprocess input data, reduce noises, facilitate the analysis of nonstationary signals, and enhance the simulation performance (Jin et al., 2024b; Nourani et al., 2009; Wang et al., 2012). Wang et al. (2012) analyzed the applications of EMD and EEMD on time–frequency analyzing behaviors of nonlinear and nonstationary seismic signal. The comparison revealed that EEMD could better decompose the signal into multiple components without the mode mixing phenomenon occurred in EMD. Apaydin et al. (2021) integrated SSA and ANN models for predicting streamflow and concluded that the SSA-ANN hybrid model improved streamflow prediction accuracy by 24.11%. Zhou and Zhang (2022b) decomposed the precipitation data into multiple components using EEMD, which later served as the input of LSTM, convolutional neural network (CNN) and CNN-LSTM models to reconstruct the missing discharge data.

In the present study, we proposed an interpretable two-stage selective EEMD-TFT hybrid model to predict multi-step ahead karst spring discharge. In the first stage, the nonlinear and nonstationary precipitation data is decomposed into multiple intrinsic model functions (IMFs) and a residual function using EEMD. The decomposed components, air temperature, and categorical variables including year, month, and day are served as the input of the TFT model, which examines and ranks the importance of input variables on karst spring discharge. In the second stage, the IMFs and residual function with high importance are selected to serve as the TFT model's input features along with temperature and categorical variables to eliminate the redundant information, highlight the useful characteristics, improve the prediction performance and efficiency of the model. The selective EEMD-TFT hybrid model presented in this study offers several unique contributions to the field of

hydrological modeling, particularly for karst systems. Firstly, it synergistically combines the strengths of EEMD in decomposing nonlinear and nonstationary signals with TFT's capability to capture complex temporal dependencies for multi-step ahead karst spring forecasting and interpretability. This combination is particularly suited to the challenges posed by karst systems, where traditional methods often struggle with the inherent complexities and heterogeneities. Secondly, our two-stage approach incorporates an interpretable analysis to adaptively select the most relevant decomposed components, enhancing simulation efficiency and performance. This feature is valuable in karst landforms, where identifying the most relevant timescales and inputs can provide insights into the underlying hydrological processes. Furthermore, the newly proposed model demonstrates superior performance in multi-step ahead predictions for long forecast horizons, addressing a key challenge in hydrological forecasting. Lastly, the interpretability offered by this approach provides valuable insights into the relative importance of different timescales and input variables in karst spring discharge prediction, contributing to a better understanding of these complex systems.

2. Study area

To evaluate the prediction performance, the newly proposed model is applied for predicting multi-step ahead discharge values of the Barton Springs (USGS 08155500) in Austin, Central Texas (Fig. 1). The Barton Springs is a natural discharge point of the Edwards aquifer. The Barton Springs segment of the Edwards aquifer consists of highly karstified limestone and dolomite formations of Early Cretaceous age with extensively fractures, caves, and sinkholes. It supplies freshwater for municipal, industrial, commercial, recreation, agricultural and domestic uses in local communities, and supports the development of the economy and population in surrounding communities (Scanlon et al., 2003). The Edwards aquifer comprises both confined and unconfined zones: the confined zone is overlain by the clay rich Del Rio formation, while the extensive unconfined zone directly connects to the surface and functions as the recharge zone (Mahler and Bourgeais, 2013). The Barton Springs monitoring station, operated by the U.S. Geological Survey (USGS), collects daily data on karst spring discharge and water quality. Situated within the unconfined area of the Edwards aquifer, it resides in a subtropical, humid climate zone with major rainstorms occurring in the spring and fall seasons. The local karst spring receives recharge through fractures and fissures in the streambed when surface water flows across the unconfined aquifer, storm runoff and infiltration from the surround upland areas (Hauwert, 2016). The spring discharge values may fluctuate quickly in response to rainfall events. Meteorological data is monitored by the Camp Mabry weather station (GHCND: USW00013958) managed by National Oceanic and Atmospheric Administration (NOAA). Precipitation is directly related to spring discharge and considered as an essential input variable. Temperature also plays an important role in the regional hydrological cycle and is closely linked to various hydrogeological processes. For example, temperature influences evapotranspiration and can be related to rainfall events, thereby affecting water infiltration, runoff patterns and spring discharge (Trenberth et al., 2015). Thus, both temperature and precipitation are used as input variables, allowing our model to capture the complex hydrological interactions. Given its proximity to Barton Springs and its data coverage, the meteorological information collected at the Camp Mabry weather station is adopted to represent the local weather conditions of the studied area.

In this study, daily spring discharge, air temperature, and precipitation data from 11/27/2002 to 03/25/2023 with 7,424 data points are adopted as the training and testing datasets of the newly proposed model (see Fig. 2). The date is decomposed into three separate categorical features: "year", "month", and "day". To associate categorical date features with chronological time, a relative time index is created to represent the chronological ordering of the data points and help the model understand the chronological progression of time, while the categorical date features can provide periodic information about seasons and yearly patterns.

3. Methods

3.1. Workflow of the selective EEMD-TFT model

In this study, an interpretable multi-step hybrid deep learning model that leverages the advantages of EEMD and TFT algorithms is developed for addressing the challenges of the karst system and forecasting spring discharge. Karst aquifers are characterized by highly heterogeneous flow paths, ranging from slow matrix flow to rapid conduit flow. The EEMD model decomposes nonstationary precipitation signals into

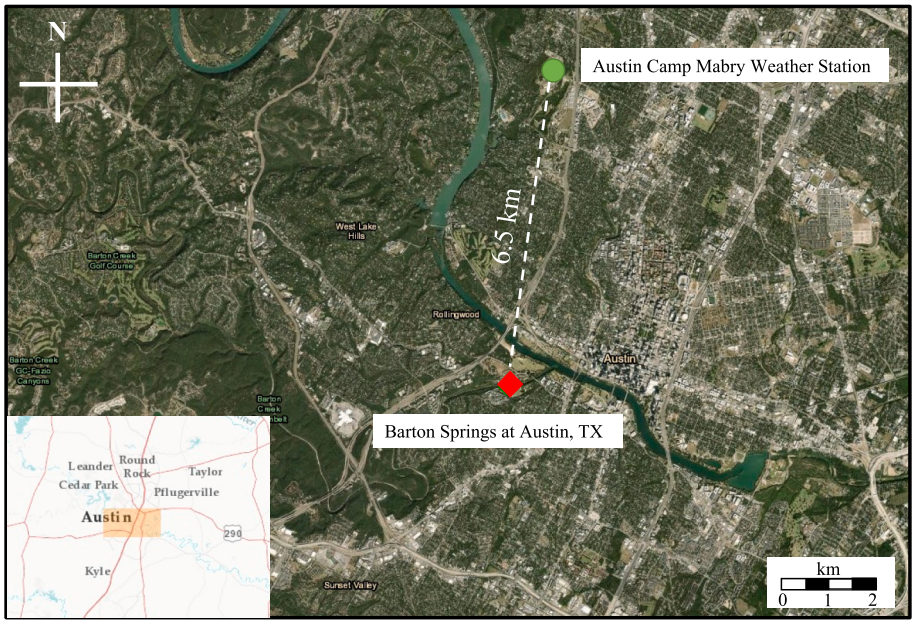


Fig. 1. The locations of the Barton Springs monitoring station (USGS 08155500) and the Camp Mabry weather station (GHCND: USW00013958) at Austin, Texas.

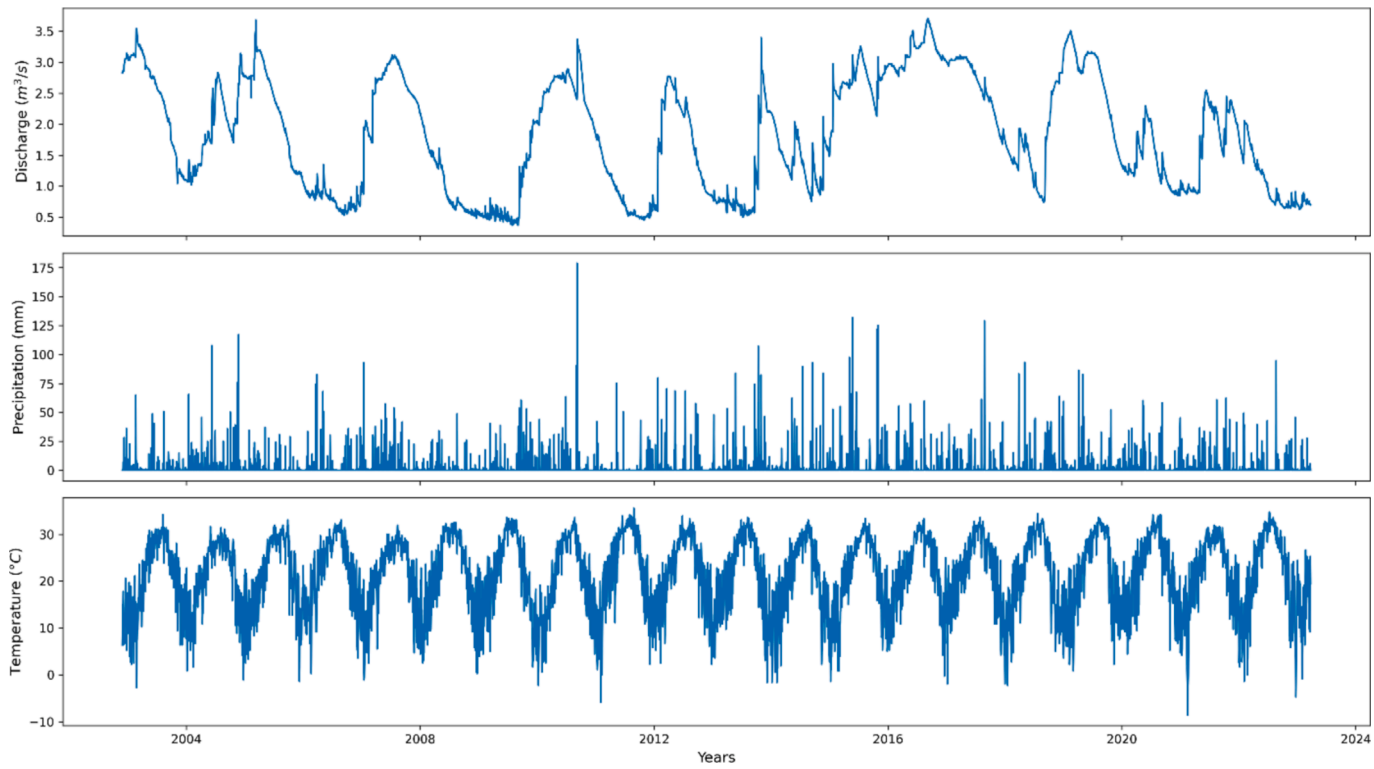


Fig. 2. Daily karst spring discharge, precipitation, and temperature from 11/27/2002 to 03/25/2023 measured at Barton Springs, Texas.

multiple intrinsic mode functions and effectively captures various temporal scales of rainfall patterns that may influence discharge differently. It allows the model to distinguish between rapid responses (potentially related to conduit flow) and slower, more gradual changes (possibly associated with matrix flow). Karst systems often exhibit nonlinear responses and threshold behaviors due to the complex interplay of conduits, fractures, and matrix porosity. The TFT model with its attention mechanism and variable selection networks can map complex and nonlinear relationships between various input variables and spring discharge across different time scales. The hybrid approach also leverages the interpretable mechanism of the TFT model to quantify the importance of different IMFs and select the most important IMFs. This selective and adaptive process can improve prediction accuracy by focusing on the most relevant components and provide valuable insights into the karst system's behavior. The derived attention weight patterns can provide interpretable insights into the temporal dynamics of karst spring discharge, reveal the relative importance of different historical time steps for prediction, and help identify key time steps and temporal patterns. By integrating these methods and incorporating a selective process, the proposed hybrid model overcomes the limitations of individual methods and results in more robust predictions and deeper understanding tailored to the complexities of the karst system.

As illustrated in Fig. 3, the selective EEMD-TFT hybrid deep learning model consists of several steps: data collection and preparation, decomposition of daily precipitation using EEMD, training and tuning of TFT, forecasting with TFT, interpretable analysis, selection of important decomposed components, and re-training and re-tuning of TFT. The detail processes of the selective EEMD-TFT hybrid model are described as follows. Step 1: The daily karst spring discharge, precipitation, and temperature data are collected from USGS and NOAA. Step 2: as the daily precipitation data exhibits strong nonstationary and nonlinear characteristics, the daily precipitation data is decomposed into several IMFs and a residual function using EEMD in order to better extract the inherent patterns and dynamics. Note that the daily temperature data is not decomposed because of its apparent seasonal patterns and periodic

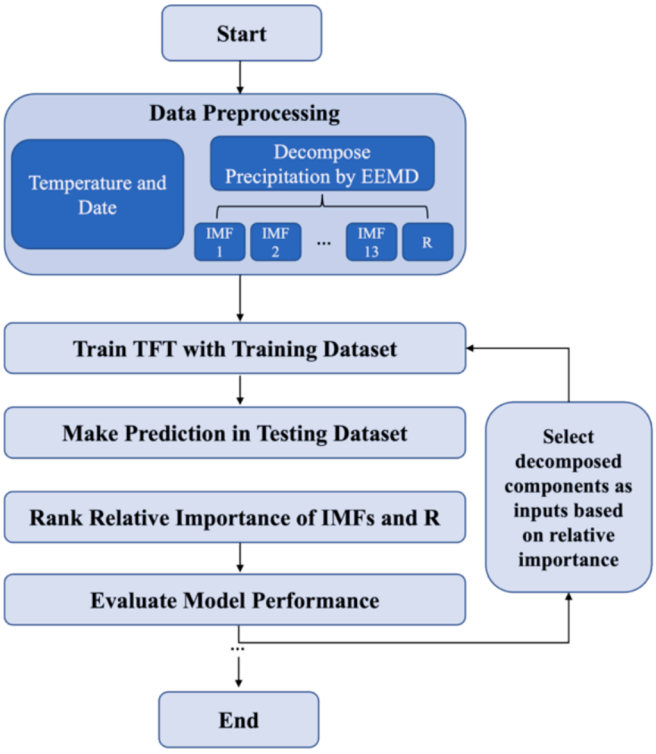


Fig. 3. Flowchart of the selective EEMD-TFT model.

characteristics as shown in Fig. 2. Step 3: Temperature and all decomposed components, including IMFs and R, are fed to TFT as known variables, while year, month and day are inputted to the TFT model as known categorical variables. Step 4: the TFT parameters, such as batch size, learning rate, number of hidden layers, number of hidden layer

neurons and attention of head size, are tuned to obtain the optimal performance. Step 5: The model is then deployed to the testing samples and assessed by evaluation metrics. Step 6: An interpretable analysis of IMFs and R is conducted to quantify and rank the relative importance of individual input variables. The decomposed components are added as input variables for the subsequential step based on their importance, from highest to lowest. The selected high-importance decomposed components and temperature serve as input. Steps 4 & 5 are repeated until the optimal performance is achieved, determining the final combination of decomposed components. The model hyperparameters are tuned again to optimize the model based on the selected input variables for the final prediction. It is worthwhile noting that the number of selected decomposed components is determined through trial and error: multiple models with various decomposed components are simulated and compared, and the model with the best performance will be adopted to decide the number of decomposed components that should be included in the final model.

3.2. Ensemble empirical mode decomposition (EEMD)

Empirical model decomposition (EMD), introduced by Huang et al. (1998), is a time–frequency analysis method to decompose time series $x(t)$ into a finite number of oscillatory modes from high to low frequencies through a sifting process. These oscillatory modes are referred to as intrinsic mode functions $C_i(t)$ (IMFs) and a residual trend component R:

$$x(t) = \sum_{i=1}^{n} C_i(t) + R \tag{1}$$

where n is the total number of IMFs after decomposition. Every IMF must meet two conditions: 1) IMF has the same number of extrema and zero crossings or differs at most by one; 2) IMF is symmetric with respect to the local zero mean, which means that the averages of the maxima and minima must be zero at any data point. After removing the effects of the noise and oscillatory behaviors, the residual component exhibits a monotonic trend or a curve with single extremum. Compared to wavelet transform methods that require a proper wavelet basis function, the EMD method is self-adaptive and automatically adapts to the data, making it particularly suitable for handling nonlinear and nonstationary time series and capturing the inherent patterns and dynamics (Huang et al., 1998; Wang et al., 2015). However, EMD may misrepresent characteristics of the original data caused by a potential mode-mixing problem, which can manifest in two ways: 1) a single IMF may comprise multiple components with widely disparate scales; 2) a signal of a specific scale may reside in various IMFs (Wang et al., 2012).

To improve the decomposition reliability, address the mode-mixing issue, and provide a more accurate representation of the original data's characteristics, Wu and Huang (2009) introduced a white noise-assisted data-analysis technique known as the Ensemble Empirical Mode Decomposition (EEMD). By incorporating Gaussian white noises, EEMD creates a consistent reference background and eliminates the potential mode-mixing problem in EMD. It projects the original data with varying scales onto proper scales for decomposition and smooths out the exceptional events like pulse interferences (Wang et al., 2012).

The extraction of IMFs and R from the original data using EEMD involves the following steps: 1) the amplitude of the Gaussian white noise and the number of realizations are set; 2) a Gaussian white noise $w_1(t)$ is added to the investigated data $x(t)$, creating noise-embedded data: $x_1(t) = x(t) + w_1(t)$, where $x_1(t)$ is the white noise-added data; 3) the local minima and maxima that define the lower envelope $e_l(t)$ and the upper envelope $e_u(t)$ are labeled and connected by a cubic spline interpolation to create the envelope; 4) the average of the local minima and maxima is calculated: $m_1(t) = [e_l(t) + e_u(t)]/2$, if the average $m_1(t)$ is close enough to zero, one can subtract it from $x_1(t)$: $h_1(t) = x_1(t) - m_1(t)$. If $h_1(t)$ satisfies the two conditions of IMFs previously mentioned, $h_1(t)$ is considered an IMF $C_1(t)$. If not, $x_1(t)$ is replaced with $h_1(t)$ as a new time series data and the process is repeated until the stopping criteria are met; 5) after extracting the first IMF component $C_1(t)$, it is subtracted from $x_1(t)$ to compute the residual $R_1(t)$: $R_1(t) = x_1(t) - C_1(t)$, which is then treated as a new time series for further decomposition into additional IMFs. This iterative decomposition is repeated until the residue $R_n(t)$ only contains a monotonic function or a minimal number of local extremes, from which no further IMF can be derived; 6) The entire procedure is then repeated for N iterations, each with different Gaussian white noises added to the original data. The final decomposed components are the ensemble means of the corresponding IMFs and the residual components. The original time series $x(t)$ is decomposed into m IMFs and a residue: $x(t) = \sum_{i=1}^{m} C_i(t) + R$. Each instinct component represents a distinct oscillatory mode or frequency component in the original data series, facilitating the analysis of underlying patterns and characteristics. In this study, the daily precipitation data from 11/27/2002 to 03/25/2023 is decomposed into 13 IMFs and 1 residue function.

3.3. Temporal fusion transformer (TFT)

The TFT model is an attention-based deep learning model designed for tackling multi-horizon time series forecasting while offering interpretable insights into temporal dynamics (Lim et al., 2021). As shown in Fig. 4, it utilizes self-attention layers to capture long-term dependencies and incorporates recurrent layers to train and predict local patterns and temporal dependencies, which allows the model to focus on more relevant parts of the input sequences for the prediction and handle complex and diverse temporal structures. In addition, TFT employs the gating mechanism to regulate the information flow through the network and prioritize relevant features for improving time series forecasting performance. It contains five significant components including gating mechanisms, variable selection networks, static covariate encoders, temporal processing, and prediction intervals.

3.3.1. Gating mechanisms

Gated Residual Network (GRN) is adopted to provide adaptive depth and network complexity, facilitating nonlinear processing between input variables and the target under a wide range of scenarios. It offers efficient information flow with skip connections and gating layers. GRN blocks contain a primary input a and an optional context vector c, described as follows:

$$GRN_{\omega}(a, c) = LayerNorm(a + GLU_{\omega}(\eta_1)) \tag{2}$$

$$\eta_1 = W_{1,\omega}\eta_2 + b_{1,\omega} \tag{3}$$

$$\eta_2 = ELU(W_{2,\omega}a + W_{3,\omega}c + b_{2,\omega}) \tag{4}$$

where *LayerNorm* denotes the standard layer normalization; ω is an index of weight sharing; GLU is gated linear units, which is also the key element for the controlling function and network flexibility. With an input γ, $\mathrm{GLU}_{\omega}(\gamma)$ can be calculated with the following equation: $GLU_{\omega}(\gamma) = \sigma(W_{4,\omega}\gamma + b_{4,\omega}) \odot (W_{5,\omega}\gamma + b_{5,\omega})$, where γ refers to the input, $\odot$ denotes the element-wise matrix multiplication product; ELU refers to the activation function of Exponential Linear Unit; η_1, η_2 refer to the intermediate layers: $\eta_1, \eta_2 \in R^{d_{model}}$; d_{model} is the hidden state size; $b_{(.,.)}$ and $W_{(.,.)}$ are the biases and weights; σ is the sigmoid activation function.

3.3.2. Variable selection networks

Since the relationship between input variables and the target is typically unknown in advance, variable selection networks are designed to determine which input variables are more relevant and should be focused on at each time step. Entity embedding is adopted to represent categorical variables. Linear transformation is used to transform each input continuous variable into a vector with dimensions in subsequent layers for skip connections. It is worthwhile to note that the static, past,

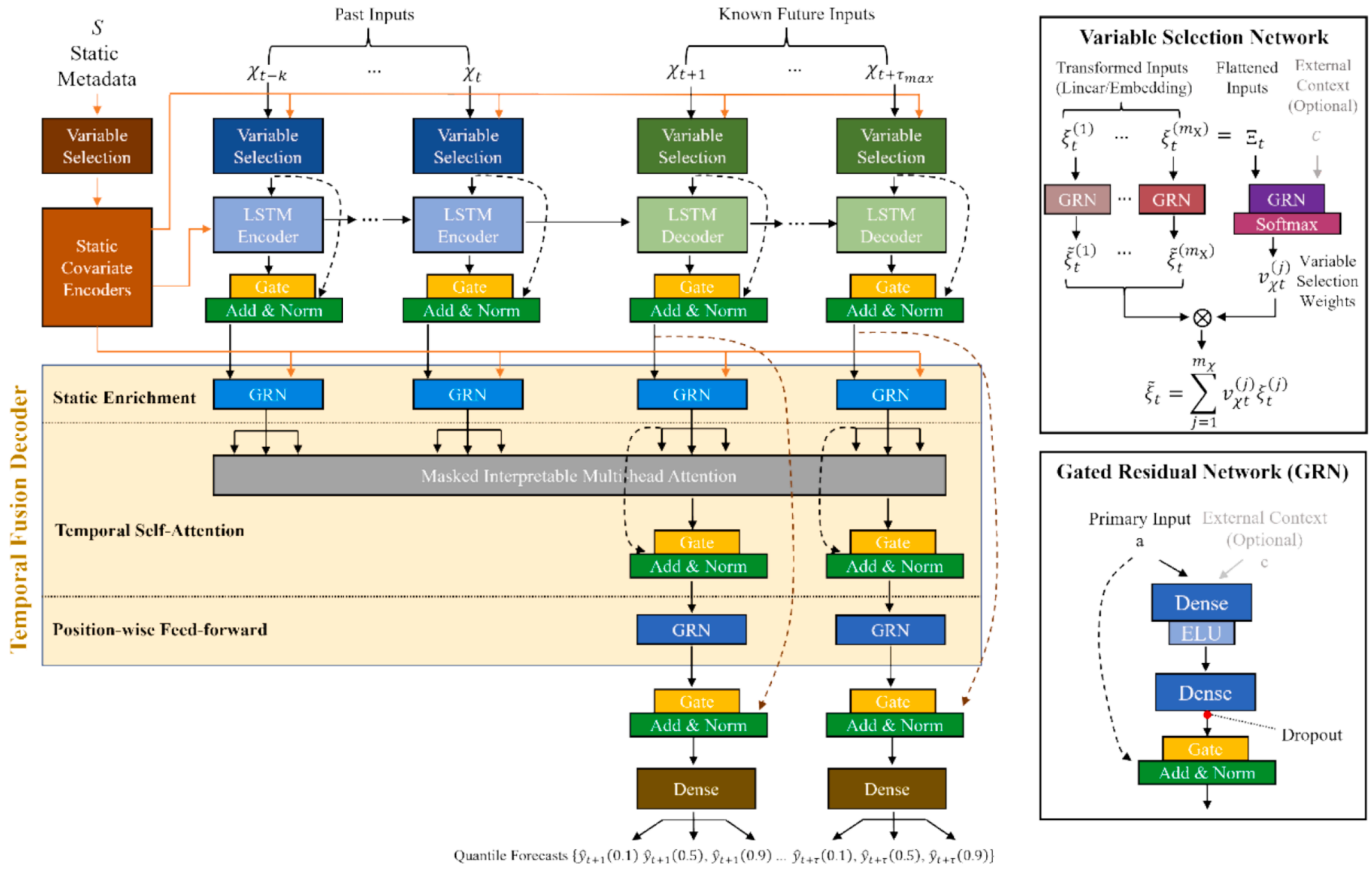


Fig. 4. The architecture of the TFT model. Static metadata, time-varying past inputs, and time-vary known future inputs serve as inputs of the TFT model. Variable selection network can select what features are more salient at each time step. Gated residual network (GRN) has the grating mechanism with gating layers and skip connections that can increase the efficiency of information flow and skip unnecessary parts of the networks if needed. It provides adaptive depth and network complexity for a wide range of scenarios.

and future inputs will have separate variable selections with distinct weights, marked by different colors in Fig. 4. The flattened vector of all past inputs at time step t can be calculate as:

$$\Xi_t = \left[\xi_t^{(1)^T}, \ldots, \xi_t^{(m_X)^T}\right]^T \tag{5}$$

where $\xi_t^{(j)}$ is the transformed input of j-th variable at time step t. The flattened vector Ξ_t and an external context vector c_s serve as inputs to a Softmax layer containing a GRN, which is used to determine variable selection weights in Eq. (6). Each $\xi_t^{(j)}$ is individually processed through a nonlinear layer by its own GRN at every time step, as shown in Eq. (7). Subsequently, the processed features $\tilde{\xi}_t$ are weighted according to their variable selection weights as outlined in Eq. (8).

$$v_{Xt} = softmax(GRU_{v_X}(\Xi_t, c_s)) \tag{6}$$

$$\tilde{\xi}_t^{(j)} = GRN_{\tilde{\xi}^{(j)}}\left(\xi_t^{(j)}\right) \tag{7}$$

$$\tilde{\xi}_t = \sum_{j=1}^{m_\chi} v_{\chi t}^{(j)} \tilde{\xi}_t^{(j)} \tag{8}$$

where v_{Xt} is a vector of variable selection weights; $v_{\chi t}^{(j)}$ is the j-th element of vector v_{Xt}; $\tilde{\xi}_t^{(j)}$ is the processed feature vector for variable j.

In this study, the static metadata consists of the categorical location identifier, labeled "Barton". The past and future inputs include the same variables: categorical date features, relative time index, temperature, and decomposed precipitation components.

3.3.3. Static covariate encoders

In contrast to RNN models for time series forecasting, the TFT model can also extract and learn useful information from static input variables. To integrate the stable variables into the networks, it produces four different context vectors through different GRN encoders. The contexts vectors include temporal variable selection (c_s), temporal features enrichment (c_e), and temporal features local processing (c_c, c_h). This encoding process involves passing the static features through GRN encoders, which transform the static information into a format that can be effectively used throughout the network. For instance, if ξ is considered as the output from the static variable selection network, the context for temporal features enrichment is defined using $c_e = GRN_{c_e}(\xi)$. As depicted in Fig. 4, the context vectors are then integrated to various layers and locations in the model where static variables may play a significant role.

3.3.4. Interpretable multi-head attention

The TFT model learns short- and long-term temporal characteristics and patterns from observations and known time-varying inputs. The local processes are handled with a sequence-to-sequence layer. The long-term dependencies across different time steps are trained and learned through an interpretable transformer-based multi-head attention architecture with the self-attention mechanism. The self-attention mechanism scales values V based on relationships between queries Q and keys K, and can be calculated in Eq. (9):

$$Attention(Q, K, V) = A(Q, K)V \tag{9}$$

where $A()$ is a normalization function. $A(Q, K)$ is usually estimated by

the scaled dot-product attention: $A(Q,K) = softmax\left(QK^T/\sqrt{d_{attn}}\right)$, where d_{attn} denotes the dimension of the attention layer. The initial values for the Q, K, and V matrices are derived from the output of the LSTM layers and the static enrichment process. These are learned parameters that are initialized randomly and then optimized during the training process. Multi-head attention is a technique adopted here to enhance the learning capacity of the attention mechanism by utilizing multiple heads, each focusing on different representation of subspace within the input data:

$$MultiHead(Q,K,V) = [H_1,\cdots,H_{m_H}]W_H \tag{10}$$

$$H_h = Attention\left(QW_Q^{(h)}, KW_K^{(h)}, VW_V^{(h)}\right)$$

where W_H linearly combines outputs concatenated from all heads H_h; $W_Q^{(h)}$, $W_K^{(h)}$, $W_V^{(h)}$ are the head-specific weights for keys, queries, and values. Due to the distinct values in each head, attention weights alone may not be indicative of a particular feature's importance. To address this issue, the multi-head attention is modified to share values across all heads and employ the additive aggregation of all heads.

$$IMA(Q,K,V) = \widetilde{H}W_H \tag{11}$$

$$\widetilde{H} = \widetilde{A}(Q,K)VW_V \tag{12}$$

$$= \left\{\frac{1}{m_H}\sum_{h=1}^{m_H} A\left(QW_Q^{(h)}, KW_k^{(h)}\right)\right\}VW_V \tag{13}$$

$$= \frac{1}{m_H}\sum_{h=1}^{m_H} Attention\left(QW_Q^{(h)}, KW_k^{(h)}, VW_V\right) \tag{14}$$

Where IMA refers to interpretable multi-head attention; $\widetilde{H}$ and $\widetilde{A}$ refer to the modified heads and normalization function, W_V is the shared value weights across all heads; W_H is employed for the final linear mapping; m_H is the number of heads. Consequently, $\widetilde{A}(Q,K)$ efficiently enhances the representation capacity and enables interpretability studies through the analysis of attention weights.

3.3.5. Temporal fusion decoder

In the temporal fusion decoder, multiple layers are designed to learn temporal relevance in the dataset (see Fig. 4), including: 1) a sequence-to-sequence layer that handles for enhancing locality and extracting local patterns. By assigning $\widetilde{\xi}_{t-k:t}$ to the encoder and $\widetilde{\xi}_{t+1:t+\tau_{\max}}$ into the decoder, a set of uniform temporal features is generated, serving as the input of the temporal fusion decoder itself; 2) a static enrichment layer that facilitates temporal features with static variables, which may have a notable influence in time series forecasting; 3) a temporal self-attention layer is employed after the static enrichment layer to maintain causal information flow through masking and learn long dependencies within the dataset; 4) a position-wise feed-forward layer using GRNs is implemented to nonlinearly process the outputs from the previous self-attention layer. A gated residual connection is also offered, which allows for skipping over the entire transformer block and providing a direct path to the sequence-to-sequence layer. This approach yields a simpler model when additional complexity is not necessary and provides an efficient and adaptable process of the input data.

In summary, the TFT model consists of several interconnected layers that process the input data sequentially: 1) Variable Selection Network uses GRNs to determine the importance of each input variable; 2) Static Covariate Encoders generate context vectors using GRNs, which are then used to condition the processing of temporal data; 3) Temporal Processing Layers use LSTMs to process past and future inputs separately and incorporate information with static enrichment layers to enhance the temporal features; 4) Temporal Fusion Decoder includes a temporal self-attention mechanism to capture long-term dependencies in the data and additional GRN layers to further process and refine the attention output. The output of each layer serves as input to subsequent layers, creating a flow of information through the network. Both GRN and the multi-head attention mechanism are important components of the TFT architecture and work in conjunction. The GRN is used in various parts of the network, such as the variable selection network and the static enrichment. Its output serves as input to the multi-head attention layers in the temporal fusion decoder. Specifically, the GRN processes input features and produces transformed representations, which are then used as the input for the multi-head attention mechanism. The multi-head attention operates on these GRN-processed inputs to capture complex temporal dependencies.

3.4. Model calibration and performance evaluation metrics

Root mean squared error (RMSE) and mean absolute error (MAE) are adopted to assess the prediction performance in the testing samples as evaluation metrics. They can be calculated as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Q_i - \widehat{Q}_i)} \tag{15}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|Q_i - \widehat{Q}_i| \tag{16}$$

where N refers to the size of output samples; Q_i and $\widehat{Q}_i$ are the observed karst spring discharge and the predicted discharge values, respectively. RMSE quantifies how spread-out the prediction errors are. It emphasizes larger errors, as they contribute more significantly to the final RMSE value. On the other hand, MAE measures the mean of the absolute errors. This metric evaluates how closely the predictions align with the ground truth and treats all errors equally. Smaller RMSE and MAE values indicate better performance by the model, with predictions that are closer to the ground truth.

4. Results

4.1. The selection of decomposed components

The multi-step hybrid deep learning model in this analysis uses a maximum encoder length of 365 days for prediction and a forecast horizon of 30 days. This means that for each prediction, the model considers up to 365 days of historical input variables to forecast Barton spring discharge for the next 30 days. The training dataset comprises daily spring discharge, temperature, and precipitation data collected from November 27, 2002, to February 23, 2023, which is used to train and fine-tune the model. The daily precipitation data at Barton Springs is decomposed into 13 IMFs from high to low frequencies and a residual function using EEMD (Fig. 5A). At first, all decomposed components, temperature and date serve as the input of the TFT model. An interpretable analysis off each input variable is then conducted using TFT to quantify and rank the importance of all IMFs and a residual function on the predicted karst spring discharge, as shown in Fig. 5B. The components with high importance values are selected to feed the TFT model again, while the redundant information and noisy components are abandoned, enabling the TFT model to focus more on the valuable information and learn the data's intrinsic characteristics. From the most to the least important components, we run the TFT model multiple times and compare the results of evaluation performance. The performance of variable models with different numbers of decomposed components is presented in Fig. 6: from the smallest to the greatest number on the x-axis, 1 means that only the most important decomposed component is selected while 14 means that all the decomposed components are selected. The smallest RMSE and MAE values are observed when the

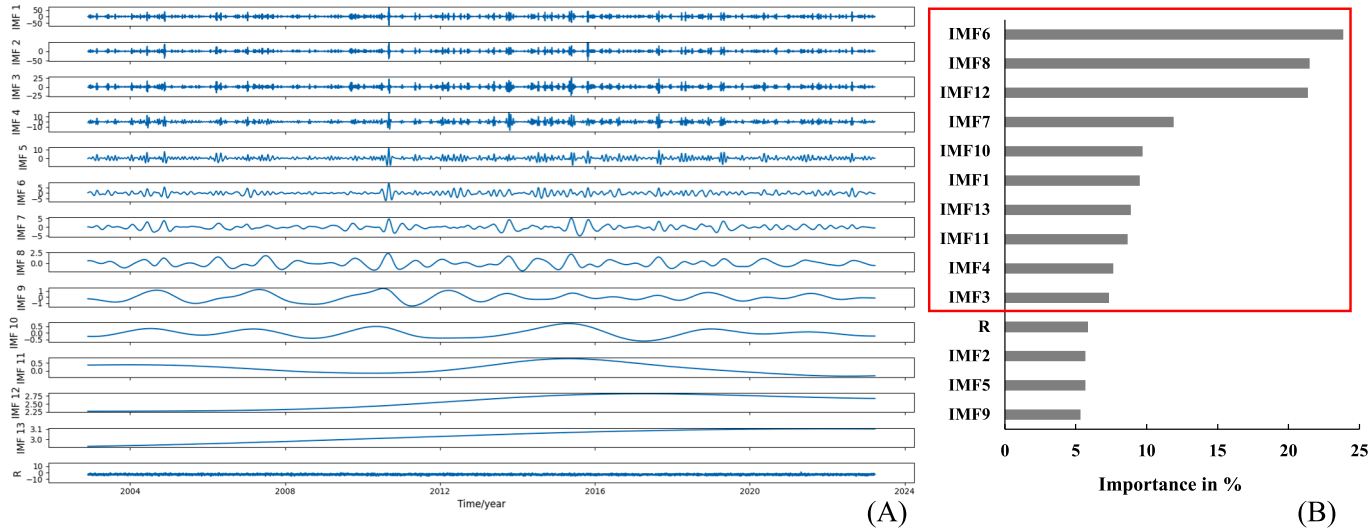


Fig. 5. Precipitation IMFs and residue decomposed by EEMD (A) and their importance on karst spring discharge according to the TFT model (B).

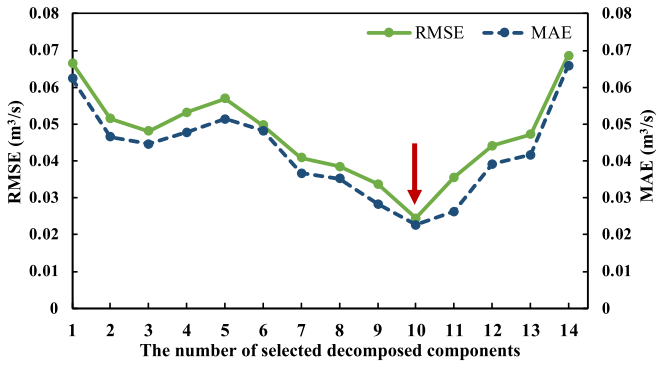


Fig. 6. The prediction performance of the selective EEMD-TFT hybrid model with various numbers of selected decomposed components as input. On the *x*-axis, the numbers range from smallest to largest, where 1 indicates that only the most important decomposed component is selected, and 14 signifies that all IMFs and a residue are included.

most significant ten decomposed components are selected, including IMF6, IMF8, IMF12, IMF7, IMF10, IMF1, IMF13, IMF11, IMF4, and IMF3. When only a few decomposed components are selected, the TFT model has less satisfying performance due to the lack of necessary information. As additional decomposed components are included, the prediction performance improves because of the incorporation of valuable information, leading to reduced RMSE and MAE values. The optimal prediction performance is achieved when the ten most important decomposed components are selected with RMSE and MAE values of 0.0255 m^3/s and 0.0224 m^3/s. However, the prediction performance tends to decline as more decomposed components are added, likely caused by the introduction of redundant information and unnecessary noises. In the following sections, the selective EEMD-TFT model will adopt the top ten decomposed components for the optimal prediction performance.

4.2. *The influence of the forecast horizon*

To comprehensively assess the model performance, we extensively compare the prediction performance between the selective EEMD-TFT hybrid model and three sequence-to-sequence deep learning models for multi-step ahead predictions in karst spring discharge. The models compared are: 1) a multi-step ahead LSTM model with an encoder-decoder architecture (Zhou and Zhang, 2022a); 2) a standard TFT model that uses raw precipitation and temperature as input variables without EEMD preprocessing; 3) an EEMD-TFT model that uses all decomposed components of daily precipitation data without the selection process. All models are trained and fine-tuned on the same datasets with a same input step of 365 days.

As illustrated in Fig. 7, the prediction performance is evaluated with RMSE and MAE across various forecast horizons. The forecast horizon refers to the length of time into the future for which the models generate predictions. For all four models, an increase of the forecast horizon corresponds to a decline of prediction performance. As the forecast horizon extends, the computational complexity and prediction difficulty significantly increase. The sequence-to-sequence models need to learn and predict more complex patterns and characteristics in the dataset and are more susceptible to exposure bias, leading to reduced performance.

Among the four models, the LSTM model with the encoder-decoder architecture has the least satisfactory prediction performance, as indicated by the highest RMSE and MAE values. Its performance is highly sensitive to the forecast horizon: its RMSE and MAE values increase from 0.0324 m^3/s and 0.0248 m^3/s, respectively, when the forecast horizon is 5 days, to 0.0938 m^3/s and 0.0849 m^3/s at a 30-day forecast horizon. The standard TFT model has better performance than the LSTM model with lower RMSE and MAE values. With the help of its attention mechanism and variable selection networks, the TFT model can capture complex and nonlinear relationships that might be challenging for traditional deep learning approaches and dynamically select the most relevant features for each time step that allows the model to focus on the most important information. It is less sensitive to the values of the forecast horizon with its RMSE and MAE ranging from 0.0159 m^3/s and 0.0101 m^3/s to 0.0690 m^3/s and 0.0657 m^3/s, respectively. The EEMD-TFT model and the selective EEMD-TFT models obtain more robust results and are less affected by the forecast horizon. It demonstrates that that the EEMD component can help mitigate random noises and allow the subsequent TFT models to better capture intrinsic patterns at multiple temporal scales that might be overlooked by conventional deep learning models, which substantially improve prediction performance. Compared to the EEMD-TFT model, the selective EEMD-TFT model incorporates a feature selection step and gains more accurate and robust predictions on karst spring discharge. It has the lowest RMSE and MAE values, ranging from 0.0144 m^3/s and 0.0131 m^3/s to 0.0307 m^3/s and 0.0250 m^3/s over various forecast horizons, respectively. By making use of the interpretability of the TFT model and focusing on the most relevant components, the selective EEMD-TFT model has following advantages over other deep learning approaches: 1) it reduces noise and redundant information, enhancing the signal-to-noise ratio in the input

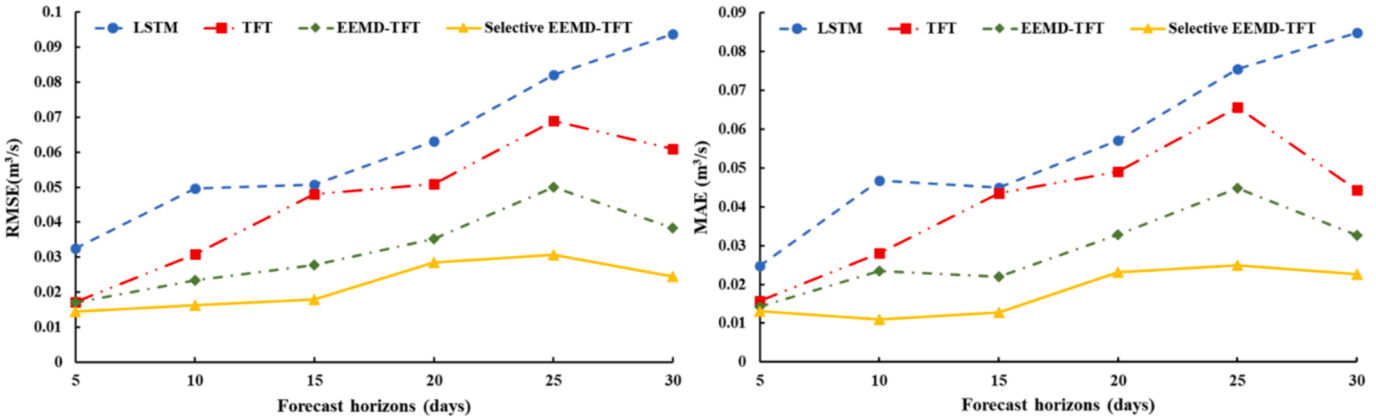


Fig. 7. The comparison between LSTM, TFT, EEMD-TFT and selective EEMD-TFT at various forecast horizons. Two evaluation matrices are adopted: RMSE and MAE.

data and allowing the model to capture more robust and meaningful relationships between spring discharge and input variables; 2) it avoids potential overfitting by eliminating less relevant inputs, and thus improves generalization. As different IMFs may correspond to various hydrological processes operating at different time scales, such as rapid flow through conduits and slow flow through the matrix, this selective approach is particularly beneficial for karst spring discharge prediction, given the intricate nature of the karst landform. By identifying and focusing on the most important IMFs, the model can better capture the dominant component influencing spring discharge in karst systems.

Overall, the selective EEMD-TFT model outperforms other three sequence-to-sequence models and has obtained more accurate and robust prediction performance. It is less sensitive to the forecast horizon compared to other sequence-to-sequence models and demonstrates a stronger ability to learn intricate patterns from the training dataset and efficiently extract useful information for prediction.

4.3. Interpretable analysis

An interpretable analysis of the selective EEMD-TFT model is conducted to assess variable importance, provide insights into the relationships among various hydrological processes, and analyze temporal patterns. To reveals which temporal scales of rainfall patterns and other input variables are most influential in predicting discharge, the model quantifies the importance of different input variables, including decomposed precipitation components and other variables. As shown in Fig. 8A, IMF3 with high frequency and IMF13 with low frequency are the most important decomposed components and contribute greatly to the karst spring prediction. This suggests that the model extracts more information from the decomposed components with certain frequencies. The short-term fluctuations represented by IMF3 and long-term trends represented by IMF13 play crucial roles in the karst system's behavior. Moreover, "Year" and "Month" play important roles in the prediction of karst spring discharge, while the contribution of "Day" is relatively insignificant. The general trend of attention weight patterns learned from raw training data is depicted in Fig. 8B. This approach sheds light on the length of time steps necessary for an intervention to have an impact and identifies the key historical time steps that the model depends on for its prediction. In this case, the greatest attention weights were observed about 330–360 steps prior to prediction steps. The observation is consistent with the previous analysis regarding the importance of "Year", indicating that the model places more emphasis on annual patterns when making predictions about karst spring discharge.

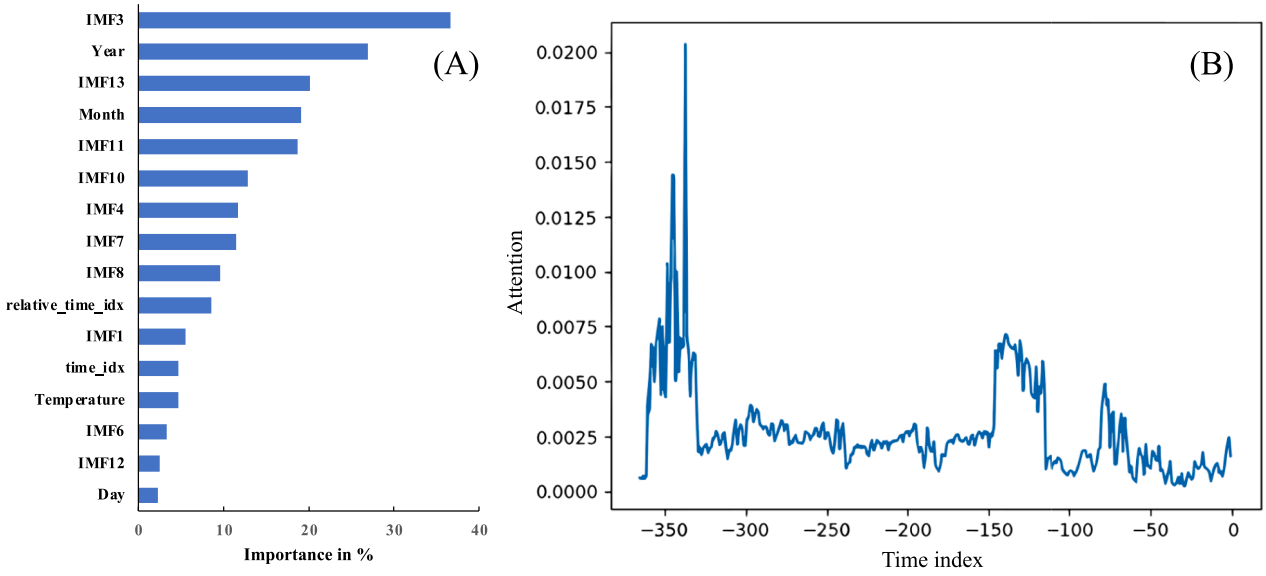


Fig. 8. Interpretable analysis of the selective EEMD-TFT model: (A) the overall importance of input variables consisting of the encoder and decoder variables' importance; (B) the attention weight patterns.

5. Conclusions

In this study, an interpretable two-stage selective EEMD-TFT hybrid deep learning model is proposed for predicting multi-step spring discharge. The novel approach incorporates the advantages of EEMD and TFT techniques and provides an improved multi-step ahead prediction of karst spring discharge. The selective EEMD-TFT model mainly includes two stages: in the first stage, the daily precipitation data is decomposed into multiple intricate components using EEMD, which extracts the patterns and characteristics from nonlinear and nonstationary precipitation data. All decomposed components, daily temperature and date will be fed to TFT as input features for predicting karst spring discharge. The model then quantifies and ranks the importance of each input variable on kart spring discharge. In the second stage, the most important ten decomposed components are selected from all IMFs and R based on their importance. These components, along with temperature and date, are inputted to the TFT model again for predicting karst spring discharge. The selection process emphasizes key features, eliminates redundant signals, and thus improves prediction performance and efficiency. While the selective EEMD-TFT model is developed and tested using data from the Barton Springs karst system, its underlying principles and structure suggest potential for broader applications in hydrology and beyond. For example, the proposed model could be applicable to other karst systems which share common characteristics of high heterogeneity, nonlinear responses, and complex flow paths. Beyond karst systems, the proposed hybrid model can be applied in other hydrological systems characterized by complex dynamics and multiple influencing factors. For instance, glacial watersheds, which also exhibit nonlinear responses and are influenced by multiple variables such as temperature, precipitation, solar radiation, could potentially benefit from the proposed approach. The primary conclusions are summarized as follows:

1. A comparative analysis is conducted on the prediction performance of various models with various numbers decomposed components as input (see Fig. 6). It is observed that the model demonstrated less satisfying performance with a limited number of decomposed components because of the lack of necessary information. As more decomposed components are incorporated, the prediction performance improves, evidenced by decreased RMSE and MAE values. The optimal performance is obtained when the ten most important decomposed components are selected, resulting in RMSE and MAE values of 0.0255 m^3/s and 0.0224 m^3/s respectively. However, the further addition of decomposed components causes a decline in the model's prediction performance, likely due to the redundant data and unnecessary noises.
2. The selective EEMD-TFT deep learning model is compared and benchmarked against other multivariate muti-step deep learning models, including a LSTM model with an encoder-decoder architecture, a TFT model, and an EEMD-TFT model without the selection process. For all four sequence-to-sequence models, an increase in forecast horizons is related to a reduction in prediction performance. This is due to an increase in computational complexity and learning difficulty as the forecast horizon increases. These models need to learn and predict more intricate patterns in the dataset and are more susceptible to exposure bias, thereby affecting performance negatively.
3. Compared to the other sequence-to-sequence models, the selective EEMD-TFT model demonstrates more accurate and robust prediction performance. It is less sensitive to the forecast horizon than other models because of its ability to effectively learn intrinsic patterns and extract valuable information from the training data for prediction.

CRediT authorship contribution statement

Renjie Zhou: Writing – original draft, Visualization, Methodology, Investigation, Conceptualization, Funding acquisition. **Quanrong Wang:** Writing – review & editing, Resources, Conceptualization. **Aohan Jin:** Writing – review & editing, Visualization, Validation. **Wenguang Shi:** Writing – review & editing, Resources. **Shiqi Liu:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was supported by the U.S. National Science Foundation (Award# 2407963) and the internal grants received from the Sam Houston State University Office of Research and Sponsored Programs.

Data availability

Data will be made available on request.

References

An, L., Hao, Y., Yeh, T.-C.-J., Liu, Y., Liu, W., Zhang, B., 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. https://doi.org/10.1016/j.jhydrol.2020.125320.

Apaydin, H., Taghi Sattari, M., Falsafian, K., Prasad, R., 2021. Artificial intelligence modelling integrated with Singular Spectral analysis and Seasonal-Trend decomposition using Loess approaches for streamflow predictions. Journal of Hydrology 600, 126506. https://doi.org/10.1016/j.jhydrol.2021.126506.

Bakalowicz, M., 2005. Karst groundwater: a challenge for new resources. Hydrogeol J 13, 148–160. https://doi.org/10.1007/s10040-004-0402-9.

Birk, S., Liedl, R., Sauter, M., 2006. Karst Spring Responses Examined by Process-Based Modeling. Ground Water 44, 832–836. https://doi.org/10.1111/j.1745-6584.2006.00175.x.

Bray, M., Han, D., 2004. Identification of support vector machines for runoff modelling. Journal of Hydroinformatics 6, 265–280. https://doi.org/10.2166/hydro.2004.0020.

Cui, Z., Zhou, Y., Guo, S., Wang, J., Xu, C.-Y., 2022. Effective improvement of multi-step-ahead flood forecasting accuracy through encoder-decoder with an exogenous input structure. Journal of Hydrology 609, 127764. https://doi.org/10.1016/j.jhydrol.2022.127764.

de Rooij, R., Perrochet, P., Graham, W., 2013. From rainfall to spring discharge: Coupling conduit flow, subsurface matrix flow and surface flow in karst systems using a discrete–continuum model. Advances in Water Resources 61, 29–41. https://doi.org/10.1016/j.advwatres.2013.08.009.

Duran, L., Massei, N., Lecoq, N., Fournier, M., Labat, D., 2020. Analyzing multi-scale hydrodynamic processes in karst with a coupled conceptual modeling and signal decomposition approach. Journal of Hydrology 583, 124625. https://doi.org/10.1016/j.jhydrol.2020.124625.

Feng, Z., Niu, W., Tang, Z., Jiang, Z., Xu, Y., Liu, Y., Zhang, H., 2020. Monthly runoff time series prediction by variational mode decomposition and support vector machine based on quantum-behaved particle swarm optimization. Journal of Hydrology 583, 124627. https://doi.org/10.1016/j.jhydrol.2020.124627.

Fleury, P., Plagnes, V., Bakalowicz, M., 2007. Modelling of the functioning of karst aquifers with a reservoir model: Application to Fontaine de Vaucluse (South of France). Journal of Hydrology 345, 38–49. https://doi.org/10.1016/j.jhydrol.2007.07.014.

Gao, S., Huang, Y., Zhang, S., Han, J., Wang, G., Zhang, M., Lin, Q., 2020. Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. Journal of Hydrology 589, 125188. https://doi.org/10.1016/j.jhydrol.2020.125188.

Ghasemizadeh, R., Hellweger, F., Butscher, C., Padilla, I., Vesper, D., Field, M., Alshawabkeh, A., 2012. Review: Groundwater flow and transport modeling of karst aquifers, with particular reference to the North Coast Limestone aquifer system of Puerto Rico. Hydrogeol J 20, 1441–1461. https://doi.org/10.1007/s10040-012-0897-4.

Goldscheider, N., Chen, Z., Auler, A.S., Bakalowicz, M., Broda, S., Drew, D., Hartmann, J., Jiang, G., Moosdorf, N., Stevanovic, Z., Veni, G., 2020. Global distribution of carbonate rocks and karst water resources. Hydrogeol J 28, 1661–1677. https://doi.org/10.1007/s10040-020-02139-5.

Goldscheider, N., Drew, D. (Eds.), 2014. Methods in Karst Hydrogeology, 0 ed. CRC Press. https://doi.org/10.1201/9781482266023.

Hartmann, A., Goldscheider, N., Wagener, T., Lange, J., Weiler, M., 2014. Karst water resources in a changing world: Review of hydrological modeling approaches. Rev. Geophys. 52, 218–242. https://doi.org/10.1002/2013RG000443.

Hauwert, N.M., 2016. Stream Recharge Water Balance for the Barton Springs Segment of the Edwards Aquifer. Journal of Contemporary Water Research & Education 159, 24–49. https://doi.org/10.1111/j.1936-704X.2016.03228.x.

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C., Liu, H.H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. r. Soc. Lond. A 454, 903–995. https://doi.org/10.1098/rspa.1998.0193.

Jin, A., Wang, Q., Zhan, H., Zhou, R., 2024a. Comparative Performance Assessment of Physical-Based and Data-Driven Machine-Learning Models for Simulating Streamflow: A Case Study in Three Catchments across the US. J. Hydrol. Eng. 29, 05024004. https://doi.org/10.1061/JHYEFF.HEENG-6118.

Jin, A., Wang, Q., Zhou, R., Shi, W., Qiao, X., 2024b. Hybrid Multivariate Machine Learning Models for Streamflow Forecasting: A Two-Stage Decomposition-Reconstruction Framework. J. Hydrol. Eng. 29, 04024026. https://doi.org/10.1061/JHYEFF.HEENG-6254.

Labat, D., Ababou, R., Mangin, A., 2000. Rainfall–runoff relations for karstic springs. Part II: continuous wavelet and discrete orthogonal multiresolution analyses. Journal of Hydrology 238, 149–178. https://doi.org/10.1016/S0022-1694(00)00322-X.

Lim, B., Arık, S.Ö., Loeff, N., Pfister, T., 2021. Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting 37, 1748–1764. https://doi.org/10.1016/j.ijforecast.2021.03.012.

Liu, F., Cai, M., Wang, L., Lu, Y., 2019. An Ensemble Model Based on Adaptive Noise Reducer and Over-Fitting Prevention LSTM for Multivariate Time Series Forecasting. IEEE Access 7, 26102–26115. https://doi.org/10.1109/ACCESS.2019.2900371.

Mahler, B.J., Bourgeais, R., 2013. Dissolved oxygen fluctuations in karst spring flow and implications for endemic species: Barton Springs, Edwards aquifer, Texas, USA. Journal of Hydrology 505, 291–298. https://doi.org/10.1016/j.jhydrol.2013.10.004.

Nourani, V., Komasi, M., Mano, A., 2009. A Multivariate ANN-Wavelet Approach for Rainfall-Runoff Modeling. Water Resour Manage 23, 2877–2894. https://doi.org/10.1007/s11269-009-9414-5.

Scanlon, B.R., Mace, R.E., Barrett, M.E., Smith, B., 2003. Can we simulate regional groundwater flow in a karst system using equivalent porous media models? Case study, Barton Springs Edwards aquifer, USA. Journal of Hydrology 276, 137–158. https://doi.org/10.1016/S0022-1694(03)00064-7.

Sezen, C., Bezak, N., Bai, Y., Šraj, M., 2019. Hydrological modelling of karst catchment using lumped conceptual and data mining models. Journal of Hydrology 576, 98–110. https://doi.org/10.1016/j.jhydrol.2019.06.036.

Trenberth, K.E., Fasullo, J.T., Shepherd, T.G., 2015. Attribution of climate extreme events. Nature Clim Change 5, 725–730. https://doi.org/10.1038/nclimate2657.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is All you Need. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems. Curran Associates Inc.

Wang, W., Chau, K., Qiu, L., Chen, Y., 2015. Improving forecasting accuracy of medium and long-term runoff using artificial neural network based on EEMD decomposition. Environmental Research 139, 46–54. https://doi.org/10.1016/j.envres.2015.02.002.

Wang, T., Zhang, M., Yu, Q., Zhang, H., 2012. Comparing the applications of EMD and EEMD on time–frequency analysis of seismic signal. Journal of Applied Geophysics 83, 29–34. https://doi.org/10.1016/j.jappgeo.2012.05.002.

Wu, Z., Huang, N.E., 2009. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Adv. Adapt. Data Anal. 01, 1–41. https://doi.org/10.1142/S1793536909000047.

Wu, B., Wang, L., Zeng, Y.-R., 2022. Interpretable wind speed prediction with multivariate time series and temporal fusion transformers. Energy 252, 123990. https://doi.org/10.1016/j.energy.2022.123990.

Yin, H., Guo, Z., Zhang, X., Chen, J., Zhang, Y., 2022. RR-Former: Rainfall-runoff modeling based on Transformer. Journal of Hydrology 609, 127781. https://doi.org/10.1016/j.jhydrol.2022.127781.

Zhang, J., Chen, X., Khan, A., Zhang, Y., Kuang, X., Liang, X., Taccari, M.L., Nuttall, J., 2021. Daily runoff forecasting by deep recursive neural network. Journal of Hydrology 596, 126067. https://doi.org/10.1016/j.jhydrol.2021.126067.

Zhang, H., Zou, Y., Yang, X., Yang, H., 2022. A temporal fusion transformer for short-term freeway traffic speed multistep prediction. Neurocomputing 500, 329–340. https://doi.org/10.1016/j.neucom.2022.05.083.

Zhou, R., Zhang, Y., 2022a. On the role of the architecture for spring discharge prediction with deep learning approaches. Hydrol. Process. 36. https://doi.org/10.1002/hyp.14737.

Zhou, R., Zhang, Y., 2022b. Reconstruction of missing spring discharge by using deep learning models with ensemble empirical mode decomposition of precipitation. Environ. Sci. Pollut. Res. https://doi.org/10.1007/s11356-022-21597-w.

Zhou, R., Zhang, Y., 2023a. Predicting and explaining karst spring dissolved oxygen using interpretable deep learning approach. Hydrol. Process. 37, e14948. https://doi.org/10.1002/hyp.14948.

Zhou, R., Zhang, Y., 2023b. Linear and nonlinear ensemble deep learning models for karst spring discharge forecasting. J. Hydrol. 627, 130394. https://doi.org/10.1016/j.jhydrol.2023.130394.

Zhou, R., Zhang, Y., Wang, Q., Jin, A., Shi, W., 2024. A hybrid self-adaptive DWT-WaveNet-LSTM deep learning architecture for karst spring forecasting. J. Hydrol. 634, 131128. https://doi.org/10.1016/j.jhydrol.2024.131128.