Journal of Hydrology 669 (2026) 135002

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

# Robust discharge prediction of seasonal snow-influenced karst systems through hybridization of process-based and data-driven models

C. Sezen [a], N. Ravbar [b], A. Hartmann [c], Z. Chen [c,*]

[a] Faculty of Engineering, Ondokuz Mayıs University, 55139 Samsun, Turkey
[b] ZRC SAZU, Karst Research Institute, Titov trg 2, 6230 Postojna, Slovenia
[c] Institute of Groundwater Management, Technische Universität Dresden 01062 Dresden, Germany

ARTICLE INFO

This manuscript was handled by Renato Morbidelli, Editor-in-Chief, with the assistance of Junbing Pu, Associate Editor

ABSTRACT

Hydrological modeling of karst systems is difficult due to their unique recharge, drainage and discharge behavior, which is often highly dynamic and nonlinear. It becomes even more challenging for elevated karst catchments, where the recharge process is additionally influenced by snow accumulation and melting. In this study, an innovative modelling approach was developed that hybridizing a process-based model and a data-driven model for the karst systems influenced by seasonal snow cover and its application was tested to a large, complex karst system in the Unica River catchment in Slovenia. For this purpose, the process-based model Génie Rural à 6 paramètres Journalier, including the CemaNeige snow routine (CemaNeige GR6J), was hybridized with the Stacked Autoencoder Deep Neural Networks (SAE-DNN). A 60-year period of catchment discharge observations, from 1962 to 2021, was used for model development, testing and evaluation. The performance of the stand-alone models, CemaNeige GR6J and SAE-DNN, as well as the hybrid model CemaNeige GR6J-SAE-DNN, was systematically compared. The results show that the hybrid model clearly outperforms both stand-alone models, especially during the extreme flow conditions. Additionally, the hybrid model performs better for more recent modelling periods than for longer ones. This is due to changes in climate conditions in historical datasets, which the hybrid model is limited to capture. Overall, the proposed hybrid modeling approach offers an innovative way to robustly predict the daily discharge behavior of karst systems influenced by seasonal snow cover, especially during extreme flow conditions, and could be applied to other karst systems with similar complexity and characteristics to support robust decision making in karst water resource management.

## 1. Introduction

Karst regions cover approximately 15% of the global ice-free continental surface and a significant part of the world population benefits from the drinking water from karst aquifers (Stevanović, 2019; Goldscheider et al., 2020). Karst systems are primarily characterized by karstification processes of soluble rocks, such as carbonate and evaporite rock formations (Chen et al., 2017). They may differ based on climate characteristics, hydrological conditions, hydrogeological and hydrochemical settings (Chalikakis et al., 2011). It causes a high degree of heterogeneity and anisotropy in the hydraulic properties of karst systems, and thus affects specific recharge, storage and discharge processes within these systems (Ford and Williams, 2007). Depending on the availability of data and information, computer models are typically used to represent karst systems for effective management of karst water resources, with different modelling approaches ranging from black box (data-driven approach), grey box (process-based approach) to white box (physics-based approach). Often, due to the lack of data and information to adequately characterize karst systems and model them using the physics-based approach, the data-driven and process-based approaches are more commonly used in practice to simulate and predict karst system behavior to support decision making in karst water resource management, especially in the context of changing climate conditions leading to more extreme weather conditions and more frequent droughts and floods in the future (Hartmann et al., 2014).

Process-based models, also known as parametric or grey-box models, are based on the conceptualization of the relationship between variables using mathematical equations within the model structure (Devi et al., 2015). Process-based models have a relatively simple structure compared to the physics-based models, revealing more information

* Corresponding author.
*E-mail address:* zhao.chen@tu-dresden.de (Z. Chen).

https://doi.org/10.1016/j.jhydrol.2026.135002
Received 21 July 2025; Received in revised form 1 January 2026; Accepted 19 January 2026
Available online 22 January 2026

related to system processes than the data-driven models. However, a better understanding of the model and its simulation process requires a larger dataset. The calibration procedure must also be carried out carefully to ensure that the model parameters are determined correctly (Devi et al., 2015). Process-based models have been widely implemented to simulate karst catchment discharge behavior (e.g., Jukić and Denić-Jukić, 2009; Chang et al., 2017; Zhou et al., 2019; Sezen et al., 2019; Sezen and Šraj, 2024a). Jukić and Denić-Jukić (2009) proposed a process-based model comprising different components, including soil cover, epikarst zone, vadose and phreatic zones. They applied the model to the Jadro Spring in Croatia and stated that the computed groundwater balance indicating a significant storage capacity in the vadose and phreatic zones. Chang et al. (2017) employed a multi-model approach to identify the dominant hydrological processes and suitable process-based model to represent Yaji karst catchment in China. This work revealed that a simple process-based model structure with less complexity could be sufficient to simulate discharge behavior; however, data availability was an important factor in determining the appropriate process-based model structure. Zhou et al. (2019) coupled a two-reservoir-based karst model with the Xinanjiang (K-XAJ) and applied the coupled model approach for the Lijiang River Basin in China. They found that the proposed K-XAJ model simulated the rainfall-runoff process better than the traditional Xinanjiang (XAJ) model. Sezen et al. (2019) and Sezen and Šraj (2024a) implemented the Génie Rural à 4 paramètres Journalier (GR4J) (Perrin et al., 2003) and CemaNeige Génie Rural à 4 paramètres Horaires (CemaNeige GR4H) process-based models for daily and hourly rainfall-runoff modelling for the karst-dominated Ljubljanica River catchment in Slovenia, respectively. They revealed that process-based models produced promising simulation results when modelling the rainfall-runoff behavior of the Ljubljanica River karst catchment. However, the extreme discharge values could not be adequately simulated. Previous studies have demonstrated that process-based models can yield reliable simulation results, particularly with regard to revealing hydrological processes. However, challenges can arise when simulating extreme flows, and uncertainties can exist regarding input variables, model parameters and structure (Beven and Smith, 2015; Herrera et al., 2022; Sezen and Šraj, 2024b; Wei et al., 2024).

Data-driven models are based on the relationship between input and output data and usually do not consider any hydrological processes during modelling. Various data-driven modelling approaches have been extensively applied to hydrological or hydrogeological modelling, especially over the last two decades. These include artificial neural networks (ANN), support vector regression (SVR), adaptive neuro-fuzzy inference systems (ANFIS) and long short-term memory (LSTM) models (e.g., Kratzert et al., 2018; Gunathilake et al., 2021; Chiang et al., 2022; Wunsch et al., 2022; Rathnayake et al., 2023; Sabzipour et al., 2023). Data-driven models have also been applied to karst catchments (e.g., Song et al., 2022; Wunsch et al., 2022; Cinkus et al., 2023; Zhang et al., 2024). Song et al. (2022) employed the Shapley Additive exPlanation (SHAP) with the LSTM model to investigate the relationship between precipitation and karst spring discharge, and to estimate the spatial–temporal characteristics of karst spring in the Niangziguan Springs catchment in China. Wunsch et al. (2022) used convolutional neural networks (CNNs) with spatially distributed input data to model the discharge of several karst springs in the Alpine and Mediterranean regions of Europe. Their work clearly demonstrated the effectiveness of CNN models in modelling karst spring discharge. Cinkus et al. (2023) compared the performance of ANN and process-based models for simulating karst spring discharges in the Mediterranean and Alpine regions. They pointed out that ANN models offered flexibility in terms of the amount of input data and yielded robust simulation results for high-flow conditions. Zhang et al. (2024) used the LSTM model with different data decomposition techniques (ensemble empirical mode decomposition (EEMD) and variational mode decomposition (VMD)), optimization techniques (Bayesian optimization, particle swarm optimization, and genetic algorithms), input variables selection approaches (partial correlation, grey correlation analysis, and partial mutual information), and the autoregressive moving average (ARMA) for predicting discharge in three karst springs in China. They stated that using a combined approach in data-driven models could provide more robust and accurate discharge modelling results in karst springs. Previous studies demonstrate that data-driven models can help reveal nonlinear relationships between input and output variables. However, they also have limitations, such as the inability to disclose hydrological processes, which prevents a better understanding of the system at the modelling stage. Factors such as the adopted data-driven model type, the determination of the hyperparameters during the training stage, and the dataset used are crucial to modelling performance.

The downsides of stand-alone process-based and data-driven models as well as the idea of benefiting from the strengths of each model type, led to the development of the hybrid modelling approach (Humphrey et al., 2016; Mohammadi et al., 2022). The hybrid modelling approach is based on integrating at least two different modelling approaches to compensate for their disadvantages and enhance simulation ability by benefiting from their advantages. Distinctive hybrid modelling approaches have been developed to enhance the simulation performance compared to the stand-alone models (e.g., Okkan et al., 2021; Durgut and Ayvaz, 2023; Xu et al., 2024). Okkan et al. (2021) implemented several hybrid modelling approaches, namely nested and parallel coupling approaches, which integrate ANN, SVR data-driven models and a dynamic water balance process-based model to enhance monthly streamflow forecasting in Gediz River Basin, Türkiye. They revealed that the nested hybrid modelling approach, based on the simultaneous calibration of process-based and data-driven models, performed better than the parallel coupling approach, which involves the separate calibration procedures. They also stated that, although the nested hybrid modelling approach can have problems, such as over-parametrization, they can be a good alternative to the stand-alone and parallel coupling approaches, particularly for simulating high flows. Durgut and Ayvaz (2023) applied a hybrid modelling approach, which integrates the Technische Universität Wien (TUW) and SVR models to improve the daily rainfall-runoff modelling in the Gediz River basin, Türkiye. They used the integrated heuristic harmony search (HS) and Nelder-Mead Simplex (NMS) optimization approach to calibrate the model parameters. They found that the hybrid modelling approach enhanced the TUW model's simulation performance. Xu et al. (2024) combined the Xinanjiang (XAJ) conceptual model and random forests (RFs), SVR, extratrees (ET), and multilayer perceptron (MLP) models. They implemented a stacking-hybrid model to simulate hourly flood forecasting in the Huai River basin. They also investigated parameter uncertainty in the XAJ and data-driven models as well as structural uncertainty in the data-driven models. They pointed out that the stacking-hybrid model approach, which considers structure uncertainty and combines the predictions of data-driven models, could offer more reliable results and improve the peak flow estimation. The usefulness and efficiency of the hybrid modelling approach for discharge modelling in karst catchments have been demonstrated in recent years (e.g., Xu et al., 2022; Bouhafa et al., 2024). Xu et al. (2022) implemented a hybrid modelling approach that integrates the Utah Energy Balance (UEB) and convolutional long short-term memory (ConvLSTM), for discharge forecasting of a catchment in Utah, USA, with seasonal snowmelt and variably karstified carbonate bedrock. They revealed that the proposed hybrid modelling approach yielded more accurate discharge simulation results and accurately represented the spatial patterns of the karst catchment. Bouhafa et al. (2024) combined the Génie Rural à 5 paramètres Journalier (GR5J) and random forests models to simulate the discharge of a large karst aquifer in Bordeaux, France. They found that the GR5J-RF hybrid model remarkably improved the discharge simulation compared to the stand-alone GR5J and RF models.

This study presents a hybrid modelling approach that integrates the process-based CemaNeige GR6J model with the data-driven SAE-DNN model. This approach is implemented to simulate the highly variable

discharge of the Unica River catchment in Slovenia. The studied catchment is characterized by highly complex hydrological processes within a large karst system, additionally influenced by seasonal snow cover. The hybrid model's performance was comprehensively assessed against the stand-alone CemaNeige GR6J and SAE-DNN models. Differential Evolution (DE) was used for model calibration. To specifically evaluate the model performance during extreme flow conditions, various advanced metrics were employed, including recession analysis, flow duration curves, and long-term maximum values. Furthermore, SHAP analyses were conducted to improve our understanding about the CemaNeige GR6J model parameters and the input variables for the SAE-DNN model within the hybrid model structure. The additional novel aspect of this study is to use an extensive 60-year historical dataset with changing climate conditions. This will involve evaluating the model's performance across multiple time periods (10- to 60-year intervals) to better understand how robust the proposed hybrid modelling approach is against a dataset with changing input variable variability.

## 2. Data and methods

### 2.1. Study area and monitoring data

The Unica springs (450 m a.s.l.) are located on the southern edge of a karst polje in south-western Slovenia and form the outlets of a large and complex aquifer system. They are crucial for both biodiversity and water supply and represent a vital natural resource. The springs exhibit considerable hydrological variability, with typical flow rates ranging from 1 to 90 $m^3/s$ and an average flow rate of 20 $m^3/s$. They drain a total recharge area of approximately 820 $km^2$, fed by three distinct sub-catchments, each contributing unique hydrological and geological features to the system (Petric, 2010; Kogovšek et al., 2023). Due to flooding of the gauging station when water levels exceed the Unica riverbed, maximum discharge values should be interpreted with caution, as the rating curve above 90 $m^3/s$ is unreliable. In addition, the flooding makes it impossible to accurately monitor spring flow conditions.

The catchment is characterized by a moderate continental climate and predominantly forested land cover. The main recharge area is the heavily karstified Javorniki Plateau, which rises to 1800 m a.s.l. This plateau consists mainly of Cretaceous limestones with lesser occurances of dolomites and breccias, forming a highly permeable aquifer with considerable groundwater storage. The region's geological diversity is further increased by the presence of carbonate rocks from the Jurassic and Paleogene periods, which are present to a lesser extent. The recharge of this sub-catchment is primarily autogenic, with direct infiltration through the karstified surface contributing to spring discharge.

In the east, the hydrology is influenced by a strike-slip fault zone along which a chain of karst poljes has developed (between 500 and 700 m a.s.l.). These poljes are underlain by Upper Triassic dolomites, which merge into Jurassic limestones and dolomites in the southern and western parts of the area. The aquifers here exhibit fissure porosity with variable permeability ranging from very low to moderate. In some areas, this leads to the formation of a superficial river network (Fig. 1). The karst poljes are hydrologically connected in a descending series, forming a unified system characterized by transitions between surface and groundwater flows and frequent flooding (Mayaud et al., 2019). This connectivity has a significant influnce on the hydrological response of the Unica springs and contributes to their pronounced variability (Ćuk Đurović et al., 2022). In the western part of the catchment, the Pivka River Basin (between 500 and 700 m a.s.l.) contributes to the hydrological complexity. The northern part of this sub-catchment is dominated by poorly permeable Eocene flysch, which supports a surface river network. In contrast, the southern part consists of carbonate rock from the Cretaceous and Jurassic periods, which forms a shallow karst aquifer. At high water levels, the surface rivers come to the fore, supplemented by water from intermittent springs on the western foothills of the Javorniki Plateau. The sinking rivers primarily exhibit channel flow in the subsurface, which further illustrates the diverse flow mechanisms within the catchment (Gabrovšek et al., 2010; Ćuk Đurović et al., 2022).

The springs were selected for this study because they are capable of draining such a large and complex binary karst system. Their hydrology is well documented and supported by extensive hydro-meteorological data. Additionally, the system is significantly influenced by snow retention and melting, which are crucial for discharge dynamics,

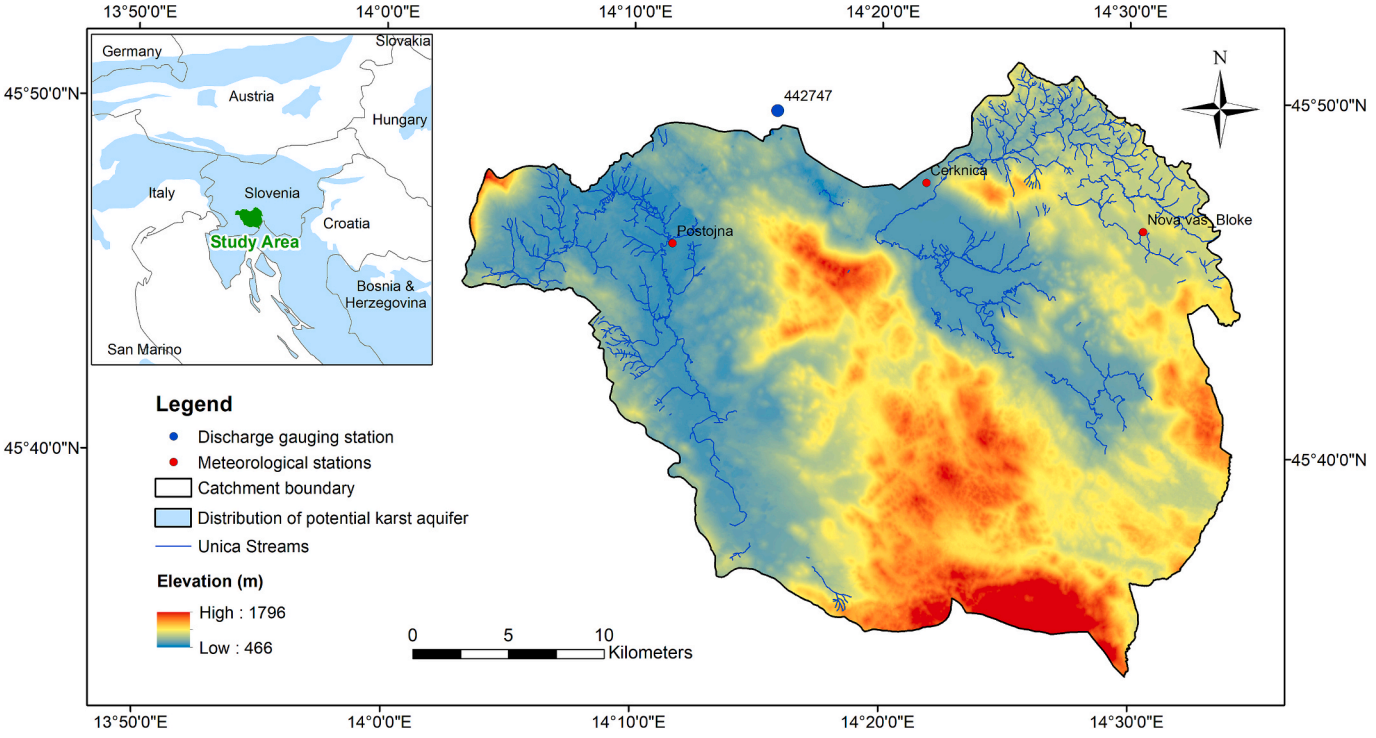


**Fig. 1.** The location of the Unica River catchment and monitoring stations selected for the current modelling work.

particularly in winter and early spring. The interconnectedness of the different sub-catchments, their varying lithology and hydrological behavior make the Unica springs an ideal natural laboratory for deepening the understanding of karst hydrology and improving hydrological modeling methods.

The daily precipitation, temperature, potential evapotranspiration, and discharge observation data were considered for the current modelling work and cover the period from 1 January 1962 to 31 December 2021, corresponding to 21,915 days. The period from 1 January 1962 to 31 December 1963 was used as a warm-up period in the process-based model. The period from 1 January 1964 to 7 August 2004 was used for the calibration/training process in the process-based, data-driven and hybrid models (i.e. 70% of the dataset). The period from 8 August 2004 to 31 December 2021 was applied for testing the process-based model (i.e. 30% of the dataset). The period from 8 August 2004 to 20 April 2013 was implemented for the validation process and the period from 21 April 2013 to 31 December 2021 was applied for the testing period in the data-driven and hybrid models (15% of the dataset for both the validation and testing stages). The sub-periods (i.e., 10-, 20-, 30-, 40-, and 50-years) were also used as modelling periods to analyze the variations in modelling performance based on the changes in climate and environmental conditions in the studied catchment. The areal precipitation of the catchment was calculated using the Thiessen polygon method based on precipitation data from the Postojna, Cerknica, and Nova vas Bloke meteorological stations. Potential evapotranspiration, snow accumulation and melt were calculated using temperature data from Postojna. The daily discharge data of the Unica Hasberg gauging station was used for catchment discharge simulation. The Oudin formula (Oudin et al., 2005), which considers the latitude of catchment and mean temperature data, was applied to calculate the daily potential evapotranspiration. The usefulness of the Oudin formula has been demonstrated in numerous rainfall-runoff modelling studies despite its simple structure (e.g., Flores et al., 2021; Bodian et al., 2024). Statistics related to precipitation, temperature, potential evapotranspiration, and discharge data, which contain minimum, mean, maximum, standard deviation, skewness, and kurtosis values, are given in Table 1. As shown in Table 1a, precipitation statistics showed significant variability compared to temperature, potential evapotranspiration, and discharge data. Furthermore, Pearson's correlation coefficients showing the relationship between the variables are given in Table 1b. The correlation between the precipitation and discharge data is approximately 0.25, reaching its maximum value for a time lag of t-1. This finding is consistent with those of Kovačič (2010) who conducted a time-series analysis of the study area. The correlations between precipitation and discharge decrease after a time lag of t-1, whereas the correlations between potential evapotranspiration and discharge, as well as temperature and discharge, are around −0.35 and −0.30, respectively, and less differentiated than the precipitation-discharge correlations.

## 2.2. Models

### 2.2.1. CemaNeige and GR6J

This research applied the GR6J hydrological model coupled with the CemaNeige snow module as the process-based model. The GR6J, which has six parameters in its model structure, was presented by Pushpalatha et al. (2011) to enhance the modelling performance of the GR4J and GR5J, particularly in terms of low flow catchment discharge simulation. The GR6J models' six parameters are related to production and routing stores, unit hydrographs, and intercatchment exchange. In the GR6J model, net precipitation and evapotranspiration are first computed, and then variables such as percolation and production store capacity are obtained. Finally, the outputs of the routing stores and direct flow are determined and the simulated catchment discharge is obtained by adding the flows from the routing stores and the direct flow together. The CemaNeige has two parameters, namely the weighting coefficient for the snowmelt factor and the cold content factor. Daily liquid equivalent total precipitation's water depth and temperature data are used as input data (Valéry et al., 2014). Thus, there are eight parameters to be calibrated in the CemaNeige GR6J model (Table 2). Detailed information related to the model structure and modelling procedure can be found in Perrin et al. (2003), Pushpalatha et al. (2011), Valéry et al. (2014) and Coron et al. (2017, 2022).

### 2.2.2. Stacked autoencoder deep neural networks

Deep learning is a powerful tool for revealing the relationship between input and output data. It implements successive layers of nonlinear transformations through the learning process, particularly in neural networks (Chollet, 2021; Tripathy and Mishra, 2024). There are several types of deep learning models, including LSTM, convolutional neural networks, generative adversarial neural networks (GANN), and Encoder-Decoder models, which are widely used in hydrological modelling (Liu et al., 2022; Le et al., 2023; Tripathy and Mishra, 2024). In this research, the stacked autoencoder deep neural network was applied. The encoder and decoder components include the transformations between the input and hidden layers as well as the transformations between the hidden and output layers, respectively. The encoder process can be defined as follows:

$$h_n = f(W_1 X_n + b_1) \tag{1}$$

where f stands for the encoding function, $X_n$ for the input vector, $h_n$ for the hidden encoder vector, $W_1$ for the weight matrix of the encoder, and $b_1$ for the bias vector. The decoder process can be expressed as follows:

$$\widehat{X}_n = g(W_2 h_n + b_2) \tag{2}$$

where $\widehat{X}_n$, g, $W_2$, and $b_2$ denote the vector of the output layer, the decoding function, the weight matrix of the decoder, and the bias vector, respectively. The autoencoder's parameters are optimized for minimizing the reconstruction error:

$$\varphi(\circledast) = \mathrm{argmin}_{\theta,\theta'} \frac{1}{n} \sum_{i=1}^{n} L\left(X^i, \widehat{X}^i\right) \tag{3}$$

where L stands for a loss function $L\left(X, \widehat{X}\right) = \|X - \widehat{X}\|^2$

The main work steps for the SAE-DNN model can be summarized as follows: a) Training of the first autoencoder from the input data and acquiring the first learned feature vector; b) Implementing the feature vector of previous layers as input for the following layers and repeating the same procedure until the completion of the training process; c) After completing the training process of all hidden layers, applying the backpropagation algorithm (BP) for obtaining ideal weights for the simulation (Liu et al., 2018). Several hyperparameters in the SAE-DNN

**Table 1a**
Statistics relating to the following data for the period 1962–2021: precipitation (P), temperature (T), potential evapotranspiration (PET) and discharge (Q).

| Variables | Statistics | | | | | |
|---|---|---|---|---|---|---|
| | Minimum | Mean | Standarddeviation | Maximum | Skewness | Kurtosis |
| P (mm/d) | 0 | 4.3 | 9.7 | 125.7 | 3.8 | 19.3 |
| T (°C) | −14.8 | 9.1 | 7.5 | 27.7 | −0.16 | −0.78 |
| PET (mm/d) | 0 | 1.8 | 1.4 | 5.4 | 0.43 | −1.1 |
| Q (mm/d) | 0.11 | 2.3 | 2.2 | 9.5 | 1.3 | 0.8 |

**Table 1b**
Cross correlations between precipitation, potential evapotranspiration, temperature and discharge data considering varied time lags (i.e., 0–10 days) for the period 1962–2021.

| | Cross-correlations | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| n | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| $P_{t\text{-}n}$-Q | 0.25 | 0.28 | 0.25 | 0.25 | 0.24 | 0.23 | 0.23 | 0.22 | 0.21 | 0.20 | 0.19 |
| $PET_{t\text{-}n}$-Q | −0.36 | −0.36 | −0.36 | −0.36 | −0.37 | −0.37 | −0.37 | −0.38 | −0.38 | −0.38 | −0.38 |
| $T_{t\text{-}n}$-Q | −0.33 | −0.32 | −0.31 | −0.31 | −0.31 | −0.31 | −0.31 | −0.31 | −0.32 | −0.32 | −0.32 |

**Table 2**
The definitions and ranges of the applied parameters (Flores et al., 2021; Mohammadi et al., 2025) in the CemaNeige GR6J model.

| Model name | Parameter name | Definition | Range |
|---|---|---|---|
| GR6J | $X_1$ | The production store's capacity [mm] | [0.5, 2000] |
| | $X_2$ | The coefficient of intercatchment exchange [mm/d] | [−5, 5] |
| | $X_3$ | The capacity of the routing store [mm] | [0, 1000] |
| | $X_4$ | Time constant of the unit hydrograph [d] | [0.5, 10] |
| | $X_5$ | Threshold of intercatchment exchange [−] | [−4, 4] |
| | $X_6$ | The coefficient of exponential store depletion [mm] | [0, 500] |
| CemaNeige | $X_7$ | The coefficient of weighting for snowpack thermal state [−] | [0, 1] |
| | $X_8$ | The coefficient of degree-day melt [mm/°C/d] | [0, 10] |

model need to be determined, including the learning rate and momentum. In the stand-alone SAE-DNN model and hybrid models, the related hyperparameters were determined using the differential evolution (DE) algorithm (see section 2.2.4 for more details). Table 3 presents an overview of the input variables, model configuration, activation function and hyperparameters for the stand-alone SAE-DNN model. The precipitation and potential evapotranspiration data with and without time lags (t, t-1, t-2 and t-3) were used as input variables for the SAE-DNN model. The hyperbolic tangent sigmoid function was implemented as the activation function and the number of neurons was determined using a trial-and-error approach.

### 2.2.3. Hybrid model

In this study, the process-based model (CemaNeige GR6J) and the data-driven model (SAE-DNN) were hybridized to simulate the daily discharge behavior of a large and hydrologically complex karst catchment in Slovenia. In the hybrid model structure, the SAE-DNN model utilizes hydrologically preprocessed input data by the CemaNeige GR6J model instead of precipitation and potential evapotranspiration. It means that the outputs of the CemaNeige GR6J model including actual evapotranspiration (AE), routing and exponential store components (QR and QRExp), and direct flow (QD), were applied as input variables for the SAE-DNN model. The DE algorithm was applied for the CemaNeige GR6J and SAE-DNN models to optimize the model parameters. The calibration of the CemaNeige GR6J model and training of the SAE-DNN model were conducted separately.

The proposed hybrid modelling approach uses a process-based model to preprocess the precipitation and temperature input data and transfers these signals to a data-driven model. The latter can then learn the nonlinear relationship between the hydrologically meaningful preprocessed variables and the observed catchment discharge. Fig. 2 illustrates the architecture of the proposed hybrid model. The following steps were adopted for the hybrid modelling procedure:

- Calibration of the CemaNeige GR6J model using the DE algorithm to obtain optimal model parameters;
- Obtaining the AE, QR, QRExp and QD variables from the CemaNeige GR6J model;
- Normalizing the above-mentioned variables between 0 and 1 before using them for the SAE-DNN model;
- Using different time lags (t-1, t-2 and t-3) for the AE, QR, QRExp and QD variables as input variables for the SAE-DNN model (see Table 3);
- Composing the SAE-DNN model structure and performing the training of the SAE-DNN model using 70% of the dataset and the DE algorithm and obtaining the optimal learning rate and momentum hyperparameters.

### 2.2.4. Differential evolution algorithm

Differential evolution (DE) is a well-established algorithm based on Darwin's evolution theory, and it is widely applied in optimization and engineering (e.g., Storn and Price, 1997; Ahmad et al., 2022). The DE algorithm has several advantages, including simplicity, and offers reliable optimization performance for challenging problems with nonlinearity, as well as large-scale and computationally burdened problems (Ahmad et al., 2022). The DE algorithm consists of five components: initialization, mutation, crossover, selection, and repetition (Storn and Price, 1997; Alizadeh and Yazdi, 2023). In the optimization process, the population is formed for each generation. The initial vector population is randomly selected. The initial population can be generated by adding normally distributed random deviations to the nominal solution. Then, new parameter sets are formed by the weighted difference between two population vectors to a third vector, a process known as the mutation process. The parameters of a population vector can be mixed with another predetermined vector and this parameter mixing process is called crossover. The ideal parameter set, which provides the minimum or maximum value of the cost function according to the chosen cost function, is selected for the optimization problem (Storn and Price, 1997). The procedure can be repeated until the optimum results are obtained. In this study, the root mean square (RMSE) was used as a cost function, and the parameter set that provides the minimum value of the

**Table 3**
The input variables and model configuration used for the standalone SAE-DNN model and the SAE-DNN part in the hybrid model.

| Model | Used input variables | SAE-DNNmodel structure | | | |
|---|---|---|---|---|---|
| | | Activation function | Number of hidden neurons | Hyperparameter used and their ranges | |
| | | | | Learning rate | Momentum |
| SAE-DNN | $(P, E)_{t\text{-}3,\ t\text{-}2,\ t\text{-}1,\ t}$ | Hyperbolic tangent sigmoid | (10,10) | [0.1, 0.5] | [0.1, 0.5] |
| CemaNeige GR6J-SAE-DNN | $(AE, QR, QRExp, QD)_{t\text{-}3,\ t\text{-}2,\ t\text{-}1,\ t}$ | | (20, 10) | [0.1, 0.5] | [0.1, 0.5] |

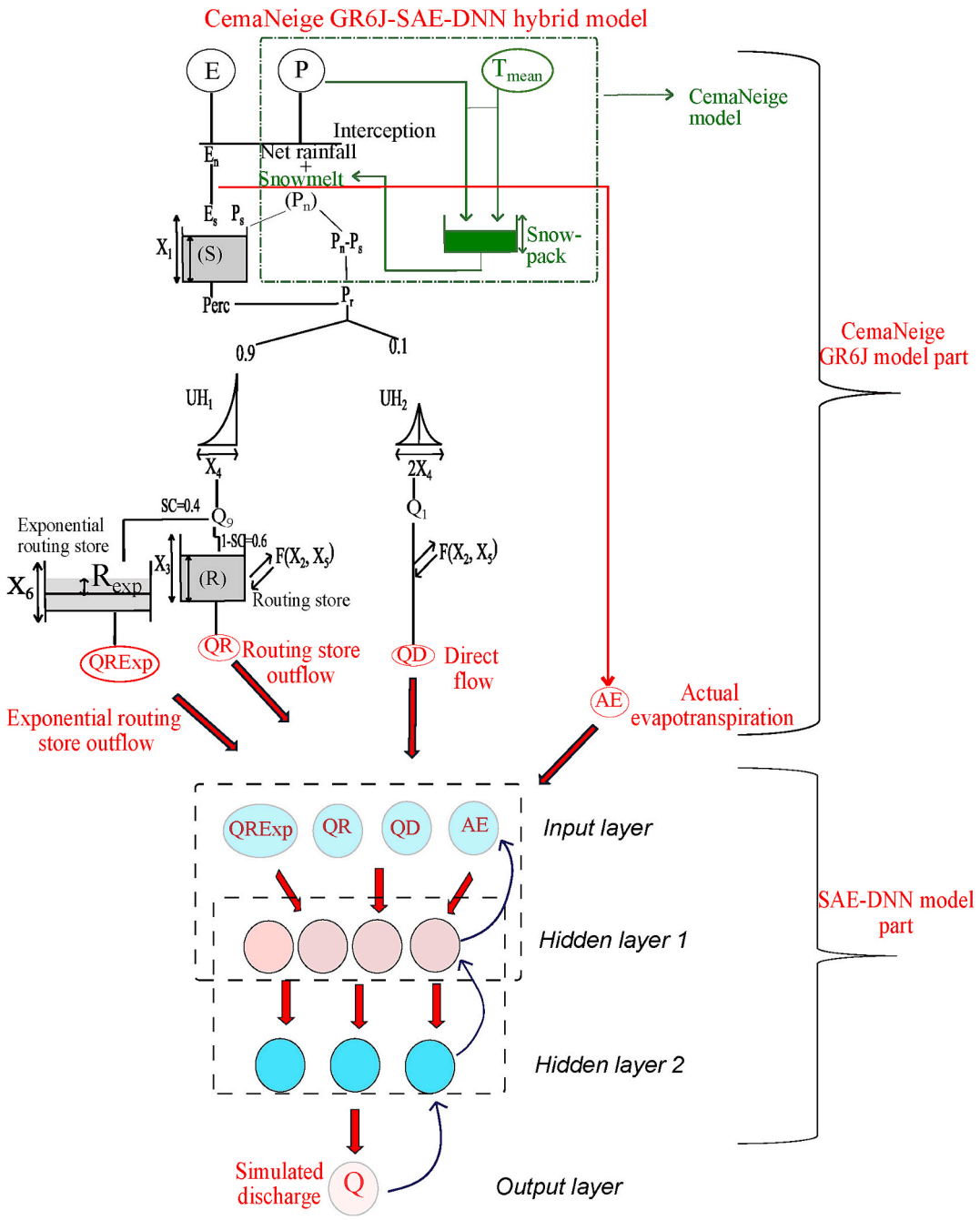


**Fig. 2.** The schematic illustration of the CemaNeige GR6J-SAE-DNN hybrid model structure (for the CemaNeige-GR6J model structure, partly adapted from Pushpalatha et al. (2011) and Sauquet et al. (2019), and for the SAE-DNN model structure, partly adapted from Liu et al. (2018)).

cost function was evaluated as the ideal parameter set for the solution. The learning rate and momentum parameters in the SAE-DNN model, the parameters of the CemaNeige GR6J model, and the same parameters in the hybrid model were determined using the DE algorithm.

### 2.3. Model evaluation

In this study, the performance of the models is evaluated by using multi-criteria, namely Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970), Kling-Gupta efficiency (KGE) (Gupta et al., 2009), index of agreement (d) (Willmott, 1981), and root mean square error (RMSE). The formula for the evaluation criteria, the value ranges and optimal values are shown in Table 4. The $Q_{obs}$, $Q_{sim}$, and $\overline{Q}_{obs}$ denote the observed discharge, simulated discharge, and mean of observed discharge, respectively, in the evaluation criteria's formula. In the KGE formula, r represents the correlation between the simulated and observed discharge values, β the ratio of the mean of the simulated discharge to the mean of the observed discharge, and α the ratio of the standard deviation of the simulated discharge to the standard deviation of the observed discharge. The optimal value is 1 for the NSE, KGE and d criteria, while it is 0 for RMSE.

The performance of the models was also assessed based on the simulation of extreme discharge values. In this regard, the recession and $Q_{95}$ percentile low discharge analysis were fulfilled to compare the low flow simulation performance of the models. The recession analysis is a helpful approach for investigating the low flow simulation performance (Gustard and Demuth, 2009). The recession constant was calculated for each model to analyze their low flow simulations. The following

**Table 4**
The statistical criteria used in the study to evaluate the performance of the models.

| Criterion | Formula | Range | Optimal value |
|---|---|---|---|
| Nash Sutcliffe efficiency (NSE) | $NSE = \frac{\sum_{i=1}^{N}(Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^{N}(Q_{obs,i} - \overline{Q}_{obs,i})^2}$ | [−∞, 1] | 1 |
| Kling Gupta efficiency (KGE) | $KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\alpha-1)^2}$ | [−∞, 1] | 1 |
| Index of agreement (d) | $d = 1 - \frac{\sum_{i=1}^{N}(Q_{sim,i} - Q_{obs,i})^2}{\sum_{i=1}^{N}(\lvert Q_{sim,i} - \overline{Q}_{obs}\rvert + \lvert Q_{obs,i} - \overline{Q}_{obs}\rvert)^2}$ | [0, 1] | 1 |
| Root mean square error (RMSE) | $RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Q_{sim,i} - Q_{obs,i})^2}$ | [0, ∞] | 0 |

equation was considered to calculate the recession constants (Maillet, 1905; Gustard and Demuth, 2009; Margreth et al., 2024):

$$k = Q_t/Q_{t-1} \quad (4)$$

$$Q_t = Q_{t-1}\, e^{-Ct} \quad (5)$$

$$C = -\frac{\ln(k)}{\Delta t} \quad (6)$$

where $Q_t$ stands for the discharge at time t, $Q_{t-1}$ for the discharge at the previous time interval, namely t-1, k for the slope of the curve fitting line based on the relationship between $Q_t$ and $Q_{t-1}$, and C for the recession constant. The $Q_{50}$ threshold level was implemented in the recession analysis. Moreover, the master recession curve (MRC) method was adopted for analysis. For further details regarding the recession analysis, refer to Gustard and Demuth (2009). Additionally, the $Q_{95}$ percentile low discharge values were obtained to evaluate the performance of low flow simulation. For high flow analysis, flow duration curves were constituted for each model, and 7-day rolling mean maximum daily discharge values were calculated. Furthermore, hydrographs showing the relationship between observed and simulated discharge values, scatter diagrams, and ridgeline plots were implemented to assess simulation performance.

The SHAP analysis was used to explore the importance of the parameters in the CemaNeige GR6J model and to assess the effect of hydrologically preprocessed input variables on the SAE-DNN model within the hybrid model structure. Proposed by Lundberg and Lee (2017), the SHAP method has several types, including KernelSHAP, LinearSHAP, DeepSHAP and SHAPtree (Makumbura et al., 2024). In this study, we decided to use the SHAPtree algorithm, which is based on tree-based approaches, such as random forests and offers several advantages, including the enhanced efficiency and the evaluation of interaction effects (Lundberg et al., 2019; Makumbura et al., 2024). In this study, extreme gradient boosting was implemented in the tree-based SHAP analysis. To better understand the importance of the CemaNeige GR6J model parameters, 2000 parameter sets (generated during the previous calibration process) and their RMSE values (calculated by comparing them to observed catchment discharge values) were used. To explore the importance of the input variables for the SAE-DNN model within the hybrid model structure, a further SHAP analysis was performed using the output obtained from the CemaNeige GR6J model with and without time lags (t, t-1, t-2 and t-3), as well as the final simulated catchment discharge values by the SAE-DNN model.

## 3. Results

### 3.1. Performance of the stand-alone and hybrid models for the period of 1962–2021

The model performance of CemaNeige GR6J, SAE-DNN and CemaNeige GR6J-SAE-DNN was initially investigated for the period from 1 January 1962 to 31 December 2021. A warm-up period from 1 January 1962 to 31 December 1963 was implemented for the process-based and hybrid models. Table 5 presents detailed information regarding the calibration/training, validation, and testing periods for each model. The eight parameters of the CemaNeige GR6J were calibrated via the DE algorithm, and the calibrated parameters were implemented in the testing period. The results related to the evaluation metrics are presented in Table 6a. Accordingly, the stand-alone CemaNeige GR6J model performs well, especially in the calibration period. The NSE and KGE values reach 0.73 and 0.77 in the calibration period, whereas they underperformed in the testing period (NSE and KGE values are 0.71 and 0.63). The hydrograph revealing the relationship between observed and simulated discharge showed that the CemaNeige GR6J model slightly overestimated low flow discharge values (Fig. 3). In addition, it overestimated and underestimated the high flow discharge values in different scenarios. Although the CemaNeige GR6J model yielded good performance according to the evaluation metrics, it did not simulate the discharge values during low and high flow conditions well enough in the studied catchment for an extended period.

The training process of the SAE-DNN model was conducted over the period from 1 January 1964 to 7 August 2004, utilizing the DE algorithm to optimize the learning rate and momentum hyperparameters. Then, the hyperparameters were obtained after using the validation dataset to avoid the overfitting problem. Precipitation (P), and potential evapotranspiration (PET) with different time lags (t, t-1, t-2, and t-3) were used as input data in the stand-alone SAE-DNN model. The SAE-DNN model performed poorly in the training, validation and testing periods, according to the evaluation metrics shown in Table 6b. It is also evident that the SAE-DNN model significantly overestimated the low-flow discharge values and underestimated the high-flow discharge values, as shown in Fig. 3. The precipitation and potential evapotranspiration data used as input variables could not help the SAE-DNN model to represent the discharge behavior of the Unica karst system. Further discussion regarding the performance of the SAE-DNN model was provided in section 4.1.

The calibration process of the CemaNeige GR6J model was initially completed, and then the model's outputs, namely AE, QR, QRExp, and QD, were acquired for the hybrid model. The related outputs were used as input variable for the SAE-DNN model in the hybrid model and then training and testing processes were performed, respectively. The simulation results indicated that the hybrid CemaNeige GR6J-SAE-DNN model remarkably improved the simulation performance compared to the stand-alone CemaNeige GR6J and SAE-DNN models. In this regard, the hybrid model enhanced the simulation results compared to the CemaNeige GR6J model by 16% and 21% (based on NSE) for the calibration and testing periods. In addition, the hybrid model improved performance compared to the SAE-DNN model by 174% and 136% (according to NSE) for the calibration/training and testing periods. The results demonstrate that the hybrid CemaNeige GR6J-SAE-DNN model significantly improved the discharge prediction during extreme flow conditions compared to the stand-alone models, as seen in Fig. 3. However, there are still certain events for which the model underestimates high flow discharge values.

### 3.2. Performance of the stand-alone and hybrid models for sub-periods

The performance of the stand-alone and hybrid models was also compared for different sub-periods (10-, 20-, 30-, 40-, and 50-year periods). Table 5 presents details regarding the warm-up, calibration/

**Table 5**
The selection of time periods for the calibration/training, validation and testing of the models.

| Modelling period | Modelling phases | Models | | |
|---|---|---|---|---|
| | | CemaNeige GR6J | SAE-DNN | CemaNeige GR6J-SAE-DNN |
| 1962–2021 | Warm-up | 01.01.1962–31.12.1963 | – | 01.01.1962–31.12.1963 |
| | Calibration / Training | 01.01.1964–07.08.2004 | 01.01.1964–07.08.2004 | 01.01.1964–07.08.2004 |
| | Validation | – | 08.08.2004–20.04.2013 | 08.08.2004–20.04.2013 |
| | Testing | 08.08.2004–31.12.2021 | 21.04.2013–31.12.2021 | 21.04.2013–31.12.2021 |
| 1971–2021 | Warm-up | 01.01.1971–31.12.1971 | – | 01.01.1971–31.12.1971 |
| | Calibration / Training | 01.01.1972–31.12.2006 | 01.01.1972–31.12.2006 | 01.01.1972–31.12.2006 |
| | Validation | – | 01.01.2007–02.07.2014 | 01.01.2007–02.07.2014 |
| | Testing | 01.01.2007–31.12.2021 | 03.07.2014–31.12.2021 | 03.07.2014–31.12.2021 |
| 1981–2021 | Warm-up | 01.01.1981–31.12.1981 | – | 01.01.1981–31.12.1981 |
| | Calibration / Training | 01.01.1982–31.12.2009 | 01.01.1982–31.12.2009 | 01.01.1982–31.12.2009 |
| | Validation | – | 01.01.2010–01.01.2016 | 01.01.2010–01.01.2016 |
| | Testing | 01.01.2010–31.12.2021 | 02.01.2016–31.12.2021 | 02.01.2016–31.12.2021 |
| 1991–2021 | Warm-up | 01.01.1991–31.12.1991 | – | 01.01.1991–31.12.1991 |
| | Calibration / Training | 01.01.1992–31.12.2012 | 01.01.1992–31.12.2012 | 01.01.1992–31.12.2012 |
| | Validation | – | 01.01.2013–02.07.2017 | 01.01.2013–02.07.2017 |
| | Testing | 01.01.2013–31.12.2021 | 03.07.2017–31.12.2021 | 03.07.2017–31.12.2021 |
| 2001–2021 | Warm-up | 01.01.2001–31.12.2001 | – | 01.01.2001–31.12.2001 |
| | Calibration / Training | 01.01.2002–31.12.2015 | 01.01.2002–31.12.2015 | 01.01.2002–31.12.2015 |
| | Validation | – | 01.01.2016–31.12.2018 | 01.01.2016–31.12.2018 |
| | Testing | 01.01.2016–31.12.2021 | 01.01.2019–31.12.2021 | 01.01.2019–31.12.2021 |
| 2011–2021 | Warm-up | 01.01.2011–31.12.2011 | – | 01.01.2011–31.12.2011 |
| | Calibration / Training | 01.01.2012–31.12.2018 | 01.01.2012–31.12.2018 | 01.01.2012–31.12.2018 |
| | Validation | – | 01.01.2019–01.07.2020 | 01.01.2019–01.07.2020 |
| | Testing | 01.01.2019–31.12.2021 | 02.07.2020–31.12.2021 | 02.07.2020–31.12.2021 |

**Table 6a**
Summary of model performance evaluation for different time periods for the process-based model.

| | | Calibration period | | | | Testing period | | | |
|---|---|---|---|---|---|---|---|---|---|
| Modelling period | Model | NSE | KGE | d | RMSE (mm/d) | NSE | KGE | d | RMSE (mm/d) |
| 1964–2021 | CemaNeige GR6J | 0.73 | 0.77 | 0.91 | 1.06 | 0.71 | 0.63 | 0.89 | 1.40 |
| 1972–2021 | | 0.74 | 0.79 | 0.92 | 1.0 | 0.70 | 0.62 | 0.89 | 1.48 |
| 1982–2021 | | 0.73 | 0.78 | 0.92 | 1.02 | 0.70 | 0.61 | 0.88 | 1.54 |
| 1992–2021 | | 0.74 | 0.77 | 0.92 | 1.09 | 0.69 | 0.61 | 0.88 | 1.57 |
| 2002–2021 | | 0.74 | 0.73 | 0.91 | 1.22 | 0.75 | 0.62 | 0.90 | 1.36 |
| 2012–2021 | | 0.74 | 0.70 | 0.91 | 1.47 | 0.72 | 0.57 | 0.88 | 1.34 |

training, validation and testing phases for each sub-period. It can be observed that the performance of the stand-alone CemaNeige GR6J model did not change considerably for the 50-, 40-, and 30-year periods (i.e., 1972–2021, 1982–2021, and 1992–2021), as shown in Table 6a. On the other hand, the performance of the CemaNeige GR6J model improved slightly particularly over the 20-years (i.e., 2002–2021). The performance of the stand-alone SAE-DNN model increased particularly for the 30-, 20- and 10-year periods (i.e., 1992–2021, 2002–2021, and 2012–2021) during the testing stages compared to other periods, as indicated by the KGE and d values (Table 6b). However, the SAE-DNN model performed remarkably poorly in all periods compared to the models CemaNeige GR6J and CemaNeige GR6J-SAE-DNN. The performance of the hybrid model CemaNeige GR6J-SAE-DNN improved slightly in the 50-, 40-, 30-, and 20-year periods compared to the period from 1964 to 2021 for the calibration phase (Table 6b). The performance of the hybrid model reached the peak for the 10-years encompassing the calibration, validation and testing phases, with NSE and KGE values exceeding 0.90.

### *3.3. Comprehensive analysis of stand-alone and hybrid models' performance during extreme flow conditions*

The performance of the stand-alone and hybrid models was comprehensively investigated. The SAE-DNN model yielded the worst performance in simulating low and high flow discharge values among used models (Fig. 4). Although the CemaNeige GR6J model outperformed the SAE-DNN model, it still did not simulate discharge values during extreme flow conditions very well. The performance of the SAE-DNN model improved the simulation of high flows, particularly in the periods 2002–2021 and 2012–2021, compared to other modelling periods. The hybrid CemaNeige GR6J-SAE-DNN model improved the performance of the stand-alone models during extreme flow conditions, as seen in Fig. 4. The enhancement in modelling performance is quite evident for the periods 2002–2021 and 2012–2021. The scatter plots indicating the relationship between the observed and simulated discharge values are presented for six different periods in Fig. 5. Accordingly, it was observed that the CemaNeige GR6J model overestimated low and high discharge values, while it underestimated mid-range discharge values. On the other hand, the performance of the

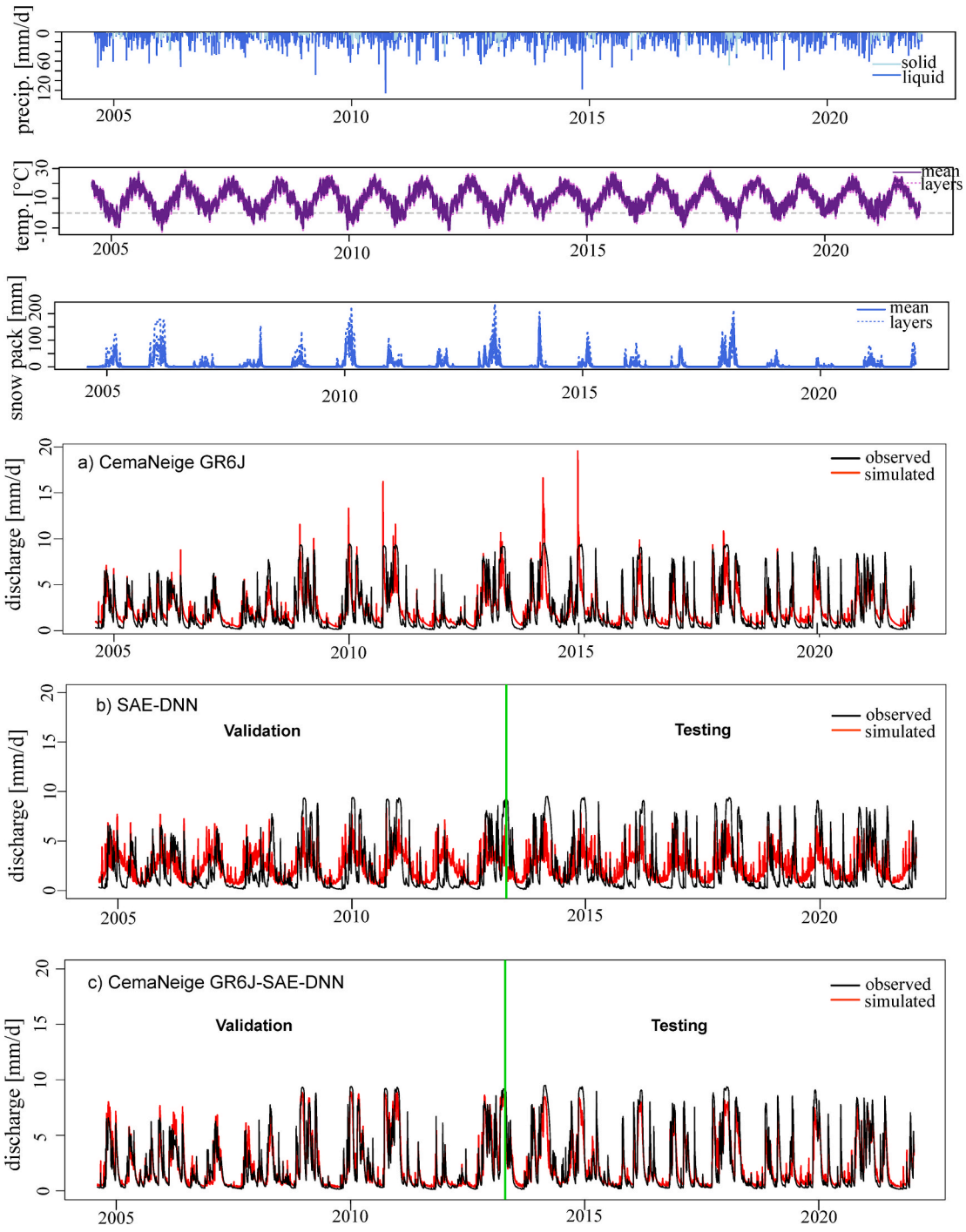


**Fig. 3.** Overview of selected model inputs and outputs compared with observed catchment discharge values from 08.08.2004 to 31.12.2021 during the modelling period (1962–2021) for the models: a) CemaNeige GR6J, b) SAE-DNN and c) CemaNeige GR6J-SAE-DNN.

CemaNeige GR6J model improved, especially for simulating peak discharge values in the 2002–2021 and 2012–2021 periods, as seen in Fig. 5. The SAE-DNN model underestimated high flow discharge values and overestimated low flow discharge values in all periods. The hybrid CemaNeige GR6J-SAE-DNN model performed better than the CemaNeige GR6J and SAE-DNN models during extreme flow conditions. This is quite evident, especially for the 2002–2021 and 2012–2021 periods, as shown in Fig. 5. The distribution of simulated and observed discharge was presented via ridgeline plots for different periods. As shown in Fig. 6, the simulated discharge for the hybrid model more closely aligns with the observed discharge than the stand-alone models. The tail and higher- density parts of the distribution plots are similar for the CemaNeige GR6J-SAE DNN hybrid model and observed discharge, especially for the 2002–2021 and 2012–2021 periods (Fig. 6e and f). The hybrid model simulated the low-flow discharge values more accurately than the process-based and data-driven models in all periods. However, the hybrid model could not sufficiently simulate extremely high discharge values, particularly in 1964–2021, 1972–2021, 1982–2021, and 1992–2021 (Fig. 6). The incompatibility of the distribution of the stand-alone SAE-DNN model's simulated discharge with the observed discharge is evident in all periods, according to Fig. 6. The CemaNeige GR6J model mostly has a better discharge distribution than the SAE-

**Table 6b**
Summary of model performance evaluation for different time periods for the deep neural network and hybrid models.

| Modelling period | Models | Calibration/Training period | | | | Validation period | | | | Testing period | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | NSE | KGE | d | RMSE (mm/d) | NSE | KGE | d | RMSE (mm/d) | NSE | KGE | d | RMSE (mm/d) |
| 1964–2021 | SAE-DNN | 0.31 | 0.38 | 0.68 | 1.70 | 0.30 | 0.40 | 0.70 | 2.05 | 0.36 | 0.37 | 0.69 | 2.2 |
| | CemaNeige GR6J- SAE-DNN | 0.85 | 0.89 | 0.96 | 0.81 | 0.89 | 0.91 | 0.97 | 0.80 | 0.85 | 0.74 | 0.95 | 1.07 |
| 1972–2021 | SAE-DNN | 0.34 | 0.42 | 0.71 | 1.61 | 0.33 | 0.38 | 0.69 | 2.23 | 0.37 | 0.36 | 0.69 | 2.14 |
| | CemaNeige GR6J- SAE-DNN | 0.86 | 0.90 | 0.96 | 0.74 | 0.89 | 0.90 | 0.97 | 0.91 | 0.86 | 0.76 | 0.95 | 1.0 |
| 1982–2021 | SAE-DNN | 0.35 | 0.42 | 0.71 | 1.61 | 0.32 | 0.38 | 0.69 | 2.37 | 0.40 | 0.38 | 0.70 | 2.12 |
| | CemaNeige GR6J-SAE-DNN | 0.86 | 0.91 | 0.96 | 0.73 | 0.89 | 0.91 | 0.97 | 0.95 | 0.86 | 0.75 | 0.95 | 1.03 |
| 1992–2021 | SAE-DNN | 0.36 | 0.49 | 0.75 | 1.72 | 0.34 | 0.38 | 0.69 | 2.33 | 0.41 | 0.46 | 0.75 | 2.1 |
| | CemaNeige GR6J-SAE-DNN | 0.87 | 0.91 | 0.96 | 0.77 | 0.88 | 0.90 | 0.97 | 1.0 | 0.88 | 0.80 | 0.96 | 0.97 |
| 2002–2021 | SAE-DNN | 0.30 | 0.39 | 0.69 | 2.02 | 0.45 | 0.59 | 0.81 | 2.14 | 0.37 | 0.42 | 0.72 | 2.0 |
| | CemaNeige GR6J-SAE-DNN | 0.87 | 0.89 | 0.96 | 0.88 | 0.91 | 0.93 | 0.97 | 0.88 | 0.90 | 0.87 | 0.97 | 0.81 |
| 2012–2021 | SAE-DNN | 0.38 | 0.43 | 0.73 | 2.26 | 0.32 | 0.42 | 0.72 | 2.07 | 0.36 | 0.52 | 0.76 | 2.01 |
| | CemaNeige GR6J-SAE-DNN | 0.90 | 0.92 | 0.97 | 0.92 | 0.94 | 0.91 | 0.98 | 0.62 | 0.94 | 0.96 | 0.98 | 0.64 |

DNN model, while its failure regarding capturing low and high flow discharge values is apparent compared to the hybrid model. The performance of the models in simulating extreme flow discharge values was specifically investigated through recession analysis and calculation of the $Q_{95}$ percentile low discharge values and 7-day maximum values. The recession constants were calculated for the observed and simulated discharge for the testing period of CemaNeige GR6J and validation and testing periods of the SAE-DNN and hybrid models, as shown in Table 7. The recession constant for 30% of the observed dataset in the 1964–2021 period is 13.1 days. On the other hand, the recession constants for the CemaNeige GR6J and SAE-DNN models are remarkably high (29.2 days) and low (3.9 days), respectively. The recession constant for the CemaNeige GR6J model decreases from the period 1964–2021 to 1992–2021, yielding better performance in closing the gap between the simulated discharge by the process-based model and observed discharge. Then, the recession constants for the CemaNeige GR6J model increase for the testing phase of the 2002–2021 and 2012–2021 periods. The recession constants for the simulated discharge by the SAE-DNN model are significantly lower for all periods compared to the observed discharge. The hybrid CemaNeige GR6J-SAE-DNN model yielded the best performance in the recession analysis (Table 7).

The average recession constant for the simulated discharge by the hybrid model is 10.5 days for all periods, and it is very close to the average recession constant of the observed discharge (i.e., 13.5 days) compared to the stand-alone process-based and data-driven models. Furthermore, the $Q_{95}$ percentile observed low discharge was compared with those of the stand-alone and hybrid models, as shown in Fig. 7. The SAE-DNN model yielded the worst performance by far in simulating the $Q_{95}$ percentile low discharge, although its performance improved in the 2002–2021 and 2012–2021 periods. The hybrid model obviously outperformed the stand-alone process-based and data-driven models regarding simulating the $Q_{95}$ percentile low flow discharge. It is also clear that the CemaNeige GR6J-SAE-DNN model has the most coherence with observed discharge for the 2012–2021 period compared to other periods, as seen in Fig. 7. The 7-day rolling mean maximum observed and simulated discharge plots were prepared for each month and long-term to compare the models' performance regarding simulating high flow discharge values, as seen in Fig. 8. The poor simulation performance of the SAE-DNN model is obvious. Accordingly, the SAE-DNN model underestimated maximum discharge values for each month and long-term. The CemaNeige GR6J model outperformed the SAE-DNN model, exhibiting a good performance, especially over long periods (i.e., 1964–2021, 1972–2021 and 1982–2021). However, the CemaNeige GR6J model underestimated maximum discharge values during the high-discharge season, especially in winter and autumn, whereas it overestimated maximum discharge values in summer for both the 2002–2021 and 2012–2021 periods. According to Fig. 8, the hybrid CemaNeige GR6J-SAE-DNN model outperformed the stand-alone models in simulating maximum discharge values, particularly during the 2002–2021 and 2012–2021 periods. On the other hand, it should be noted that the hybrid model was unable to sufficiently capture the maximum discharge values in the 1964–2021, 1972–2021, 1982–2021, and 1992–2021 periods.

### 3.4. SHAP analysis

The SHAP analysis was used to explore the importance of the input variables for the SAE-DNN model in the hybrid model structure. These variables were actual evapotranspiration (AE), the routing and exponential storage components (QR and QRExp) and the direct flow (QD), which were obtained from the CemaNeige GR6J with and without time lags (t, t-1, t-2 and t-3). Fig. 9 shows the mean SHAP values calculated for each of the considered input variables, along with their ranking. The results show that in general, the mean SHAP values of the input variables at time t are higher than those with time lags (t-1, t-2 and t-3). Furthermore, the mean SHAP values of the individual input variables varied across different modelling periods. For each period, the QR had a consistently greater impact than the other input variables. The QD, QRExp and AE appear in different orders to the QR on the ranking list and have varied mean SHAP values. Table S1 in the supplementary material provides a complete overview of the mean SHAP values and rankings of all the variables investigated. The SHAP analysis was also used to explore the importance of the CemaNeige GR6J model parameters, with the model run as a standalone. The results show that the model parameters $X_2$ and $X_5$, which are related to the intercatchment exchange, have the highest mean SHAP values for each modelling period (Fig. 10). In comparison, the mean SHAP values of other model parameters are much lower. The model parameters $X_2$ and $X_5$ are consistently ranked first and second. The remaining parameters are production store capacity ($X_1$), routing store capacity ($X_3$), snow module parameters

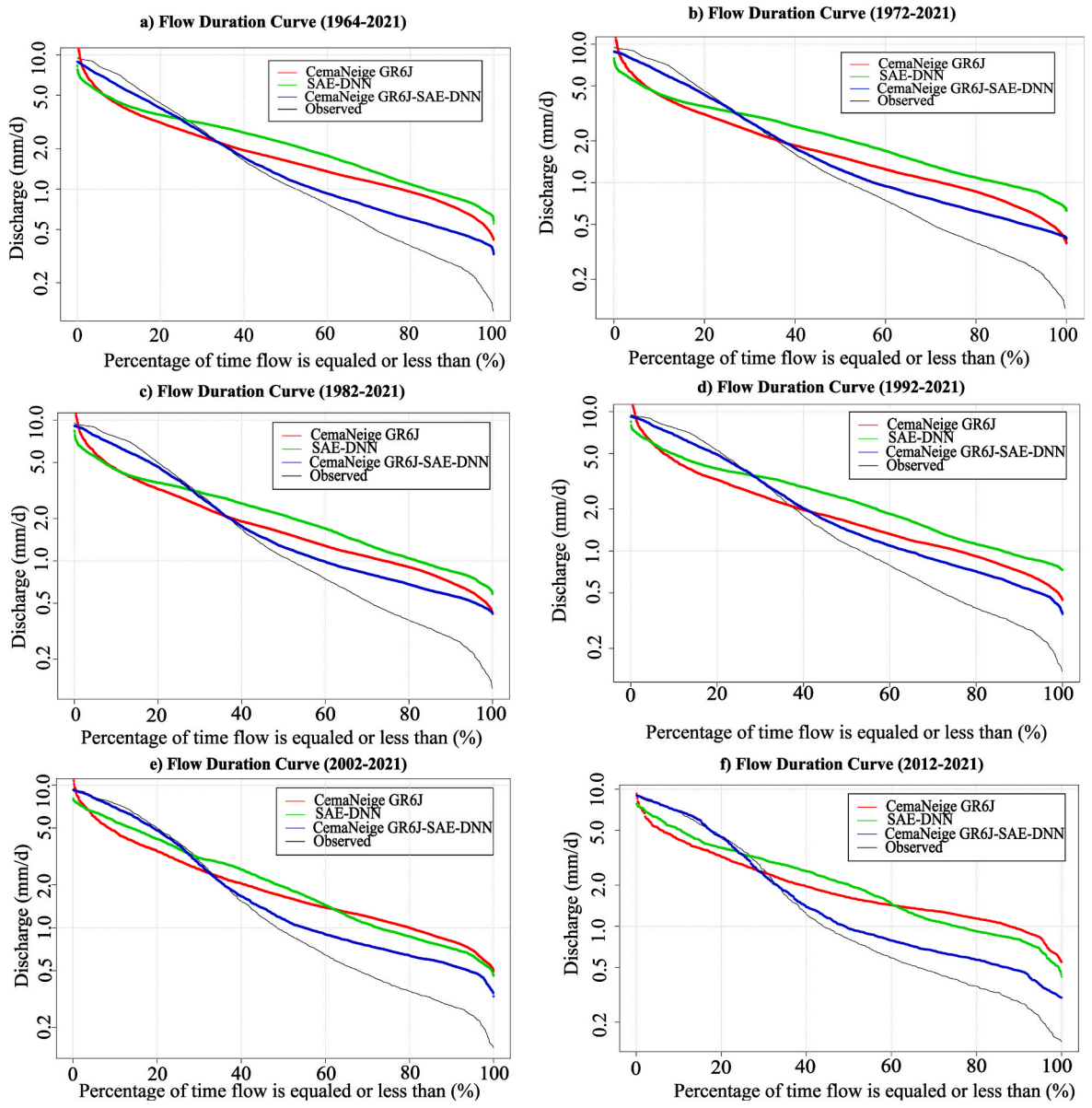


**Fig. 4.** The flow duration curves of the simulated and observed discharge values for the testing period of CemaNeige GR6J and validation and testing periods of the SAE-DNN and CemaNeige GR6J-SAE-DNN models of a) 1964–2021, b) 1972–2021, c) 1982–2021, d) 1992–2021, e) 2002–2021, and f) 2012–2021.

($X_8$ and $X_7$), exponential store depletion coefficient ($X_6$) and unit hydrograph time constant ($X_4$). This ranking changed slightly during the sub-periods.

### 3.5. *Impact of changing climate conditions on model performance*

To investigate model performance change across sub-periods, trends in hydrometeorological variables were analyzed using the Mann-Kendall and Pettitt tests. In this regard, the trends in annual precipitation totals, liquid and solid forms of precipitation obtained from the CemaNeige module, annual mean temperature, annual potential evapotranspiration totals, and annual total catchment discharge were investigated, as shown in Table S2. Accordingly, significant increasing trends were detected in temperature and evapotranspiration, while a significant decreasing trend was observed in the solid form of precipitation. Additionally, a significant change point was detected in the year 1996 for both temperature and potential evapotranspiration time series. Although trends in precipitation data are not statistically significant at $\alpha = 0.05$, a change point in precipitation is also detected in the year 1996. In this regard, significant trend changes in time series involved for the modelling, such as increases in temperature and potential evapotranspiration and decreases in solid precipitation, can be determined. These trends have a direct impact on the modelled snow cover behavior in the studied karst catchment. Increasing temperature will generally lead to less snow storage and more melting during snow season. Additionally, the change in temperature could also lead to spatial change of snow storage and melting pattern directly, and spatial–temporal change in storage and discharge of individual sub-catchments in the studied karst system. The calibrated parameters of the process-based model are presented in Table S3 to analyze whether there is a pattern change from the period 1964–2021 to 2012–2021. It can be observed that $X_2$ increased over the last 20 years, while $X_8$ decreased in the same period. It indicates that both model parameters controlling snow melting and surface–groundwater interaction had adapted the behavior change of snow cover and consequently change in catchment storage and discharge. This has a positive impact on the performance of catchment discharge simulations in the process-based and hybrid models during these periods, compared to previous ones.

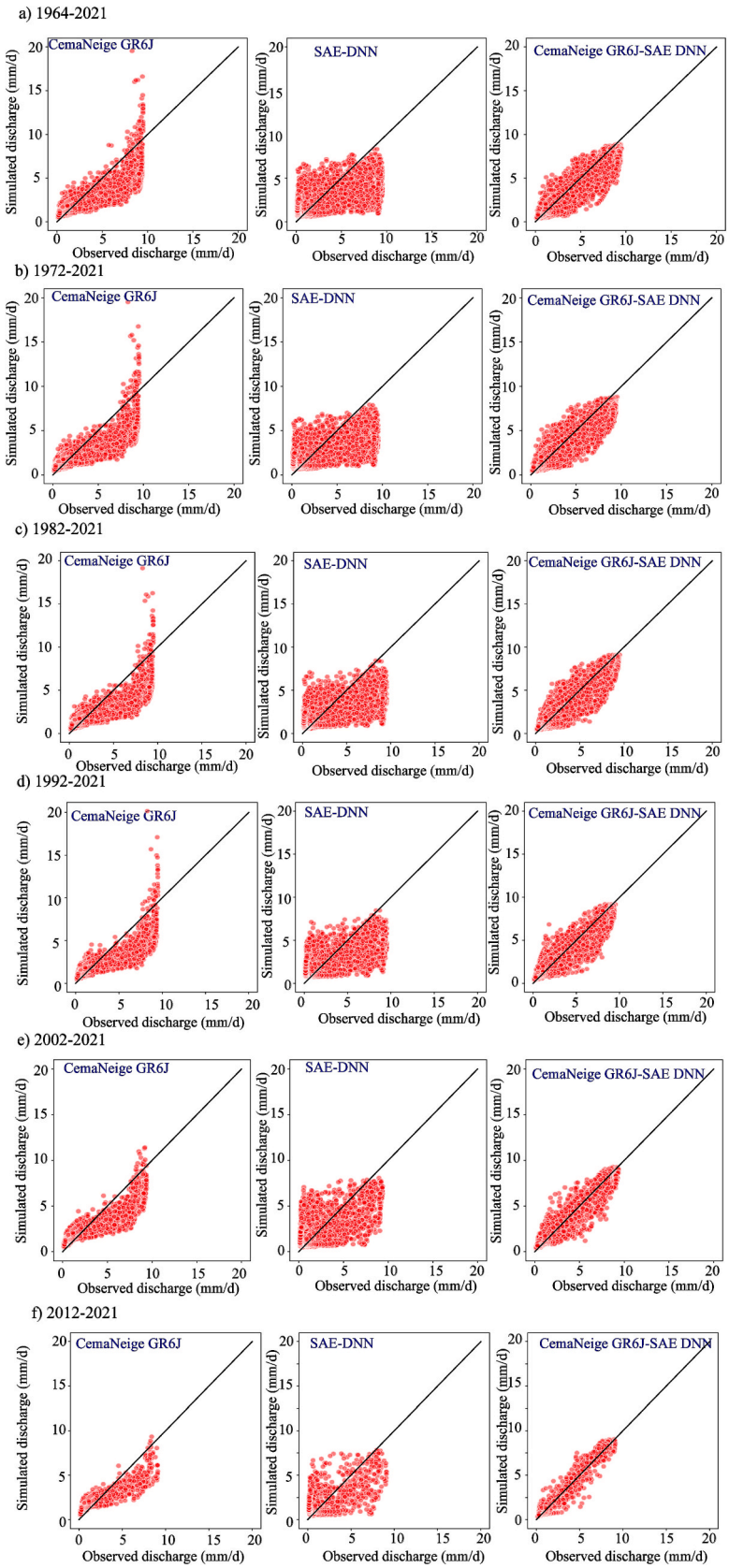


**Fig. 5.** The scatter plots of simulated and observed discharge values for the testing period of CemaNeige GR6J and validation and testing periods of the SAE-DNN and CemaNeige GR6J-SAE-DNN models of a) 1964–2021, b) 1972–2021, c) 1982–2021, d) 1992–2021, e) 2002–2021, and f) 2012–2021.

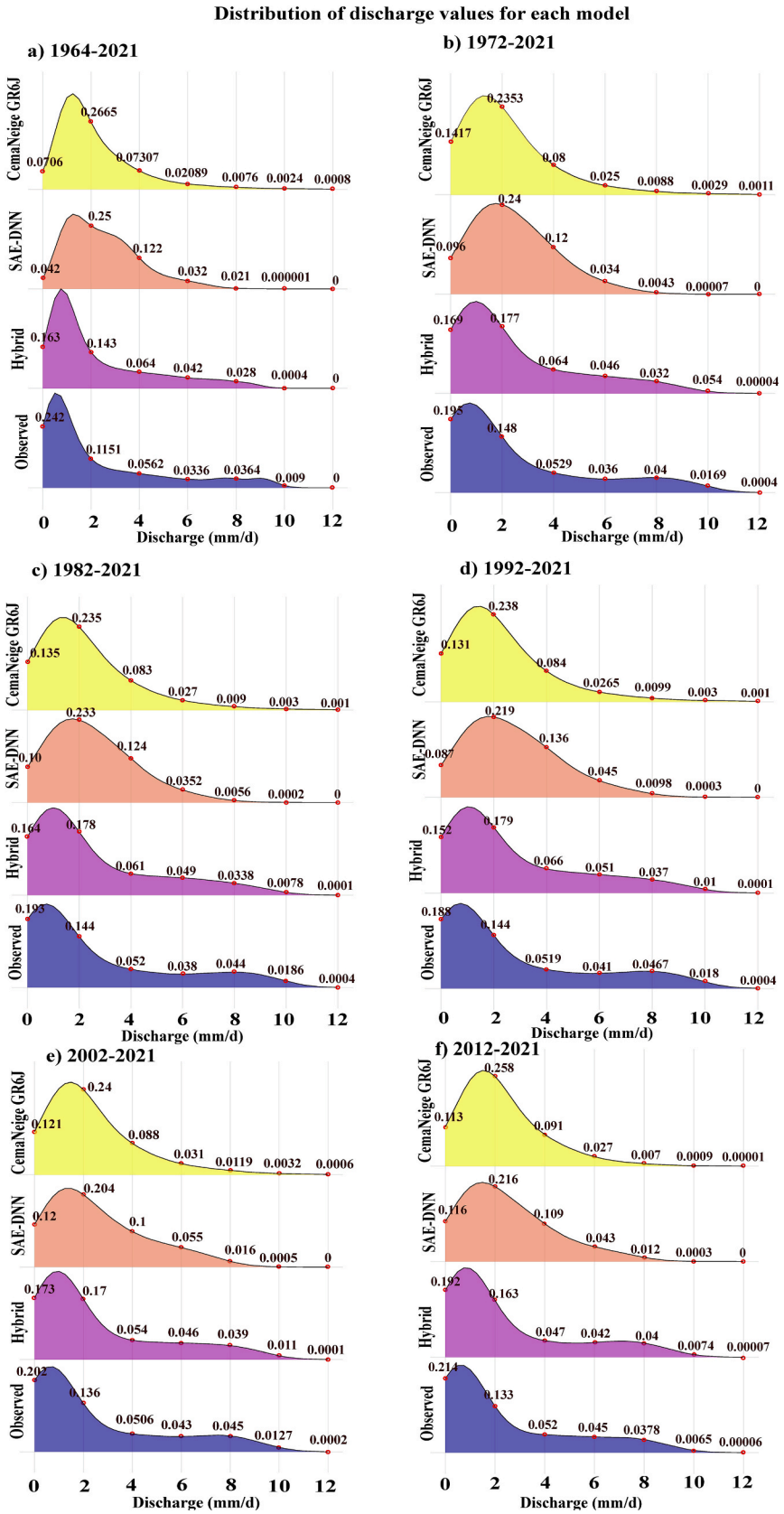


**Fig. 6.** Ridgeline plots illustrate the distribution of simulated and observed discharge values for the testing period of CemaNeige GR6J and validation and testing periods of the SAE-DNN and CemaNeige GR6J-SAE-DNN models of a) 1964–2021, b) 1972–2021, c) 1982–2021, d) 1992–2021, e) 2002–2021, and f) 2012–2021. The numbers around the red circles show the distribution values corresponding to the discharge values on the x-axis.

**Table 7**
Recession constants for the observed and simulated discharge for the testing period of CemaNeige GR6J and validation and testing periods of the SAE-DNN and hybrid models.

| | Recession constants, *C* (days) | | | | | |
|---|---|---|---|---|---|---|
| | 1964–2021 | 1972–2021 | 1982–2021 | 1992–2021 | 2002–2021 | 2012–2021 |
| CemaNeige GR6J | 29.2 | 27.2 | 25.7 | 19.3 | 35.5 | 39.9 |
| SAE-DNN | 3.9 | 3.7 | 3.5 | 4.3 | 3.2 | 4.1 |
| CemaNeige GR6J-SAE-DNN | 9.8 | 11.4 | 11.5 | 11.2 | 10.5 | 8.4 |
| Observed discharge | 13.1 | 13.3 | 13.6 | 13.6 | 13.5 | 13.6 |

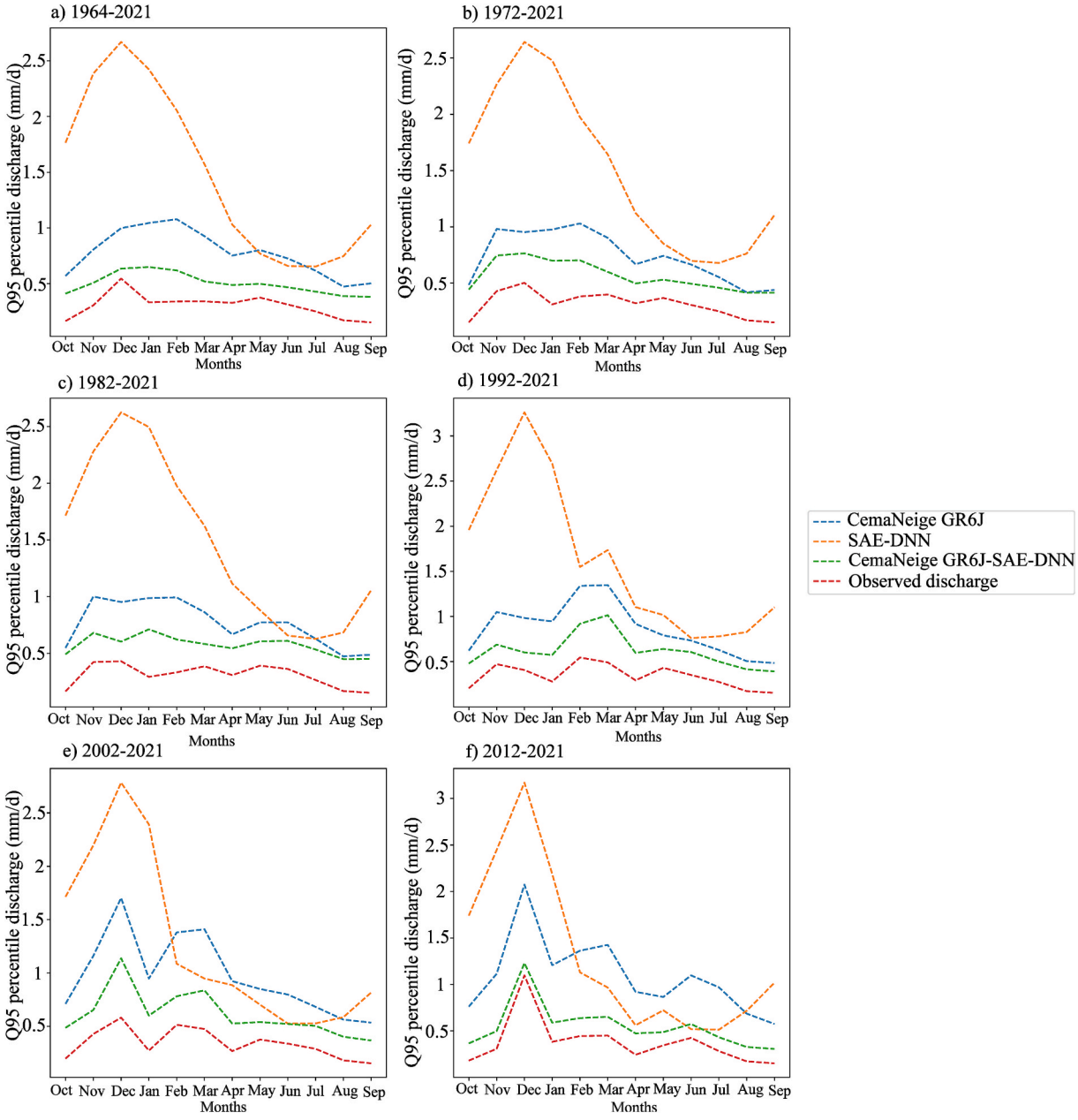


**Fig. 7.** The comparison of $Q_{95}$ percentile observed discharge and simulated discharge for the testing period of CemaNeige GR6J and validation and testing periods of the SAE-DNN and CemaNeige GR6J-SAE-DNN models of a) 1964–2021, b) 1972–2021, c) 1982–2021, d) 1992–2021, e) 2002–2021, and f) 2012–2021.

## 4. Discussion

### *4.1. Comparison between standalone and hybrid models*

One of the research focuses of this work was a comprehensive performance comparison between the CemaNeige GR6J, SAE-DNN, and CemaNeige GR6J-SAE-DNN models under both low and high flow conditions. According to the evaluation metrics and extreme flow discharge analyses, the hybrid CemaNeige GR6J-SAE-DNN model outperformed the stand-alone process-based and data-driven models. The hybrid

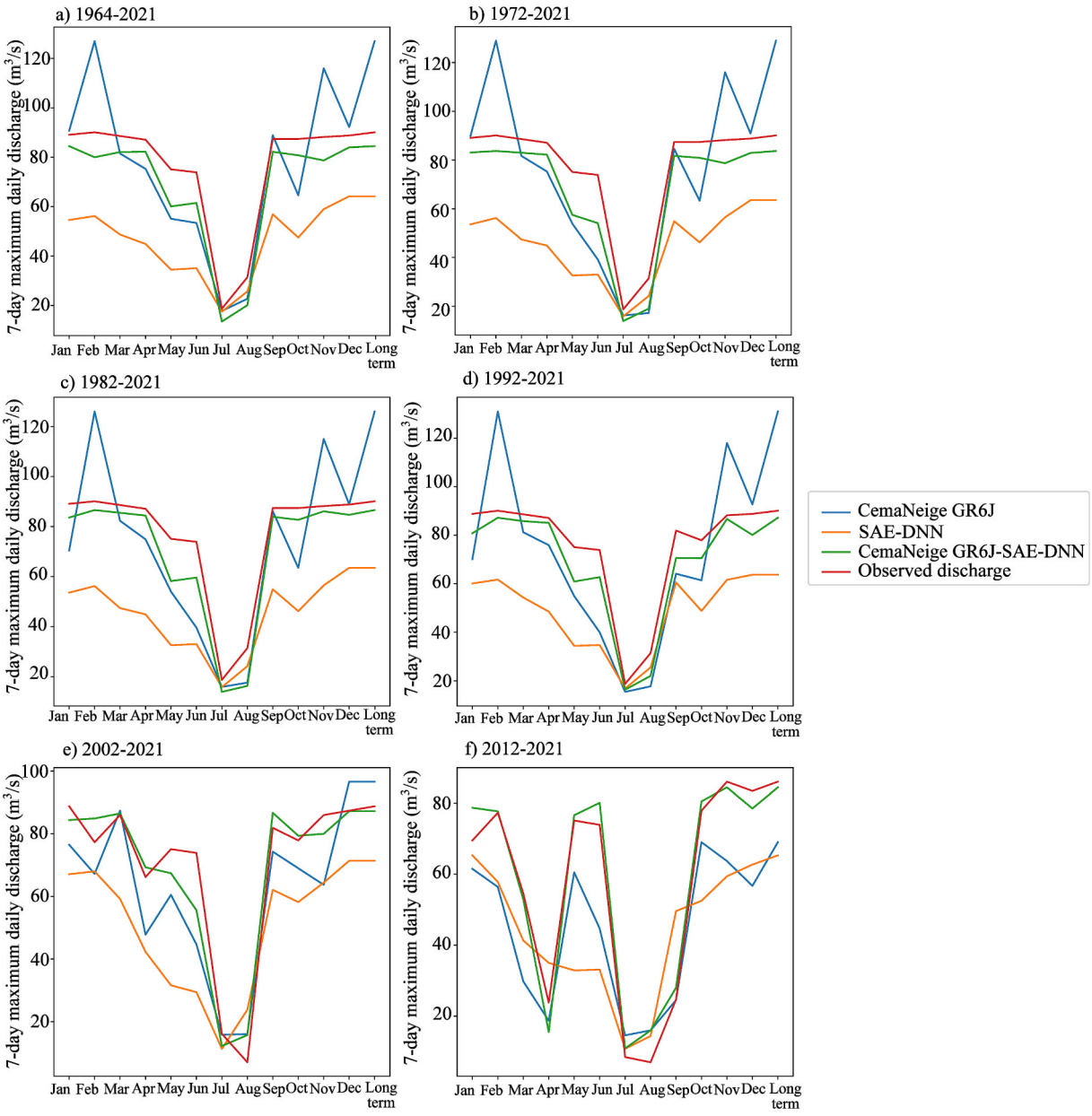


**Fig. 8.** The comparison of 7-day rolling mean maximum observed discharge and simulated discharge for the testing periods of CemaNeige GR6J and validation and testing periods of the SAE-DNN and CemaNeige GR6J-SAE-DNN models of a) 1964–2021, b) 1972–2021, c) 1982–2021, d) 1992–2021, e) 2002–2021, and f) 2012–2021.

model yielded very good performance, especially over the last 20 years (2002–2021) and the 10 years (2012–2021) periods. The low-flow analyses revealed that the stand-alone SAE-DNN model performed poorly, overestimating low flow discharge values. The CemaNeige GR6J model outperformed the SAE-DNN model; however, it also overestimated low flow discharge values. The calculated recession constants for the simulated discharge by SAE-DNN are considerably less than those for the observed discharge for each sub-period, and the recession constants for CemaNeige GR6J are remarkably higher than those of the observed discharge. Regarding the high flow analysis, the SAE-DNN model underestimated high-flow discharge values, whereas CemaNeige GR6J overestimated them, particularly peak discharge values.

The hybrid model significantly enhanced the performance of low and high-flow discharge simulations. Accordingly, the calculated recession constants of the CemaNeige GR6J-SAE-DNN model are more consistent with those of the observed discharge than those of other models. Flow duration curves, scatter diagrams, ridgeline plots, $Q_{95}$ percentile low discharge, and 7-day maximum discharge analysis also indicated greater coherence of the simulated discharge by the hybrid model with observed discharge than the stand-alone models in terms of either the distribution of the discharge values or high flow discharges.

In the hybrid model structure, the SAE-DNN model could fulfil the transformation process of actual evapotranspiration, routing and exponential store outflow and direct flow components obtained from the CemaNeige GR6J model into discharge output more reliably than the stand-alone model. After the CemaNeige GR6J model, the summation of the routing and exponential store outflow (i.e., QR and QRExp) and direct flow components yields the discharge output, whereas the SAE-DNN model can utilize these variables with different weights for discharge simulation and resulting in a more robust extreme flow discharge prediction. Thus, the effect of the variables obtained from the process-based model can be considered in light of their preponderance using the data-driven model within the hybrid model structure. However, the hybrid model significantly underestimated the extremely high

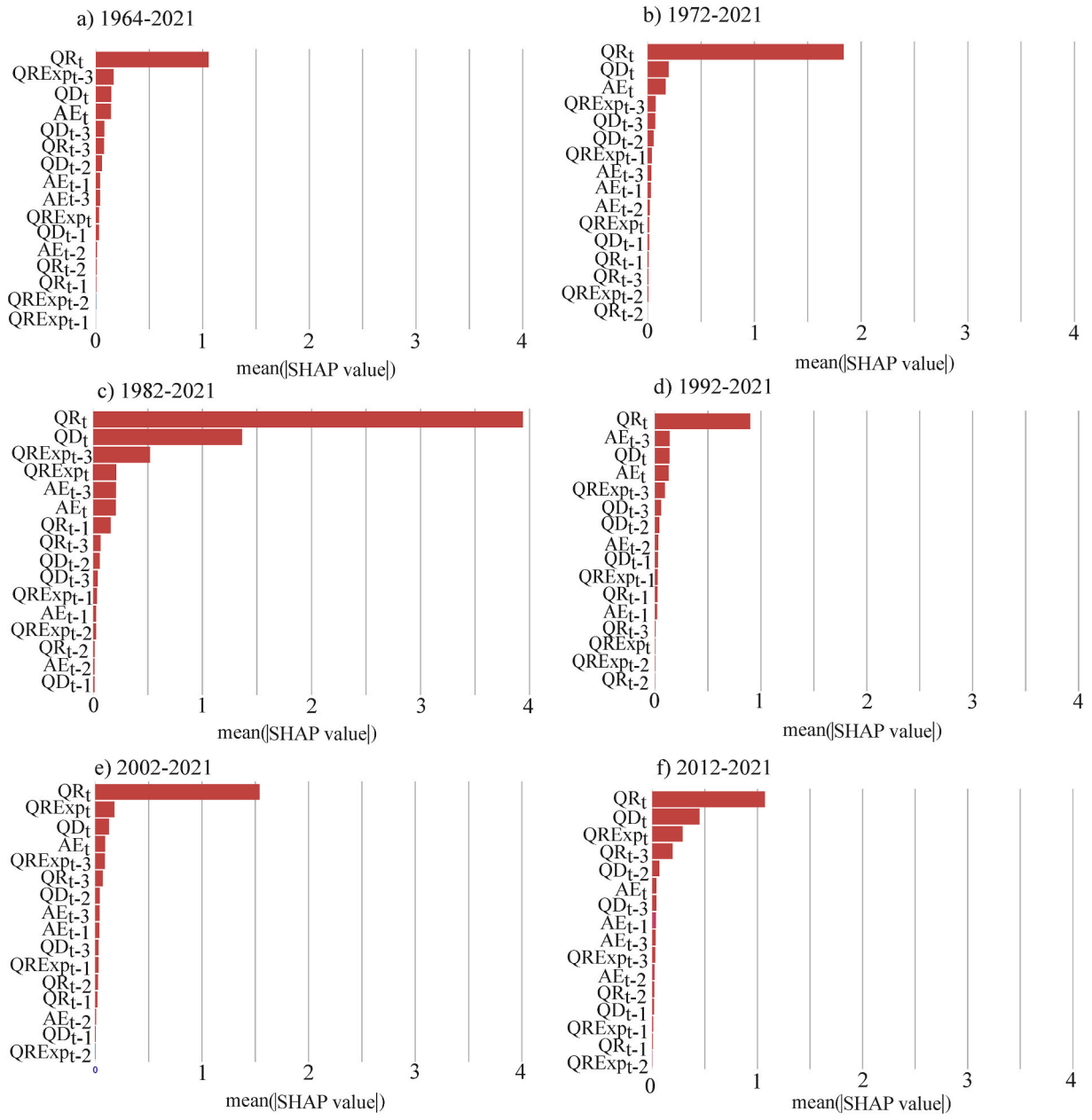


**Fig. 9.** Feature importance analysis of the SAE-DNN model input variables in the hybrid model structure for the training periods of a) 1964–2021, b) 1972–2021, c) 1982–2021, d) 1992–2021, e) 2002–2021, and f) 2012–2021. The subscripts in the input variables indicate the time lag (i.e., t, t-1, t-2, t-3) used in the input data for catchment discharge simulation.

discharge values, particularly during the periods 1964–2021 (58 years), 1972–2021 (50 years), 1982–2021 (40 years), and 1992–2021 (30 years). It can be considered that the performance of the hybrid model increased for the shorter periods of 2002–2021 (20 years) and 2012–2021 (10 years). Kovačič et al. (2020) reported an increase in precipitation and evapotranspiration, along with a decrease in snow cover, discharge regime and in significant vegetation cover change, in the Unica River catchment for the 5-year period 2014–2018, compared to the 1962–2013 period. It should be noted that the periodic and long-term variations in climatic variables and vegetation in the study area may affect the model's performance, as these changes were not considered during the calibration process. In particular, in the substantial change in vegetation cover from large-scale forest disturbance led to increased infiltration and groundwater recharge, as a result of reduced canopy interception and altered evapotranspiration (Vilhar et al. 2022). In addition, the decreasing snow cover and snowmelt effect in the Unica River karst system in recent years, noted also by Blatnik et al. (2024) could lead to a reduction in the highly nonlinear process. This process involves the transformation of snowmelt into discharge, which has enhanced the extreme flow discharge simulation performance of the hybrid model, particularly over the last 20 years.

The performance enhancement in extreme flow discharge simulation via using the hybrid modelling approach was also revealed in previous studies (e.g., Zhong et al., 2023; Nguyen et al., 2024; Wei et al., 2024). Zhong et al. (2023) integrated data-driven models (i.e., process-wrapped recurrent neural network and LSTM) and the EXP-HYDRO process-based model to simulate runoff responses to climate change in Alpine catchments. They revealed that the hybrid model approach outperformed the stand-alone models regarding simulating the $Q_{max}/Q_{min}$ ratio and could be useful in simulating the complex hydrological process under the effect of rapid climate change. Nguyen et al. (2024) utilized the hybridization of the Hydrologic Engineering Center-Hydrologic Modeling

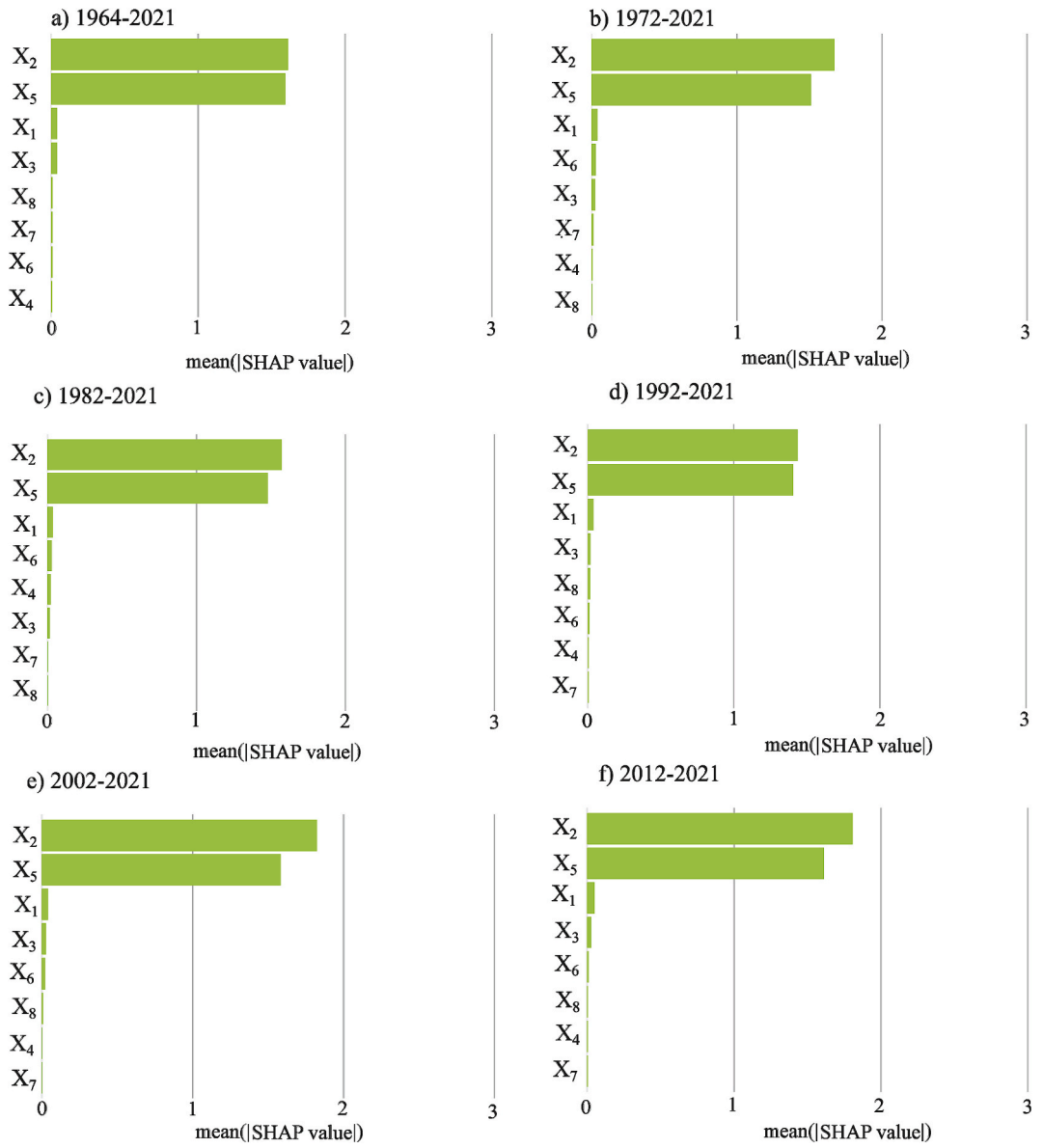


**Fig. 10.** Feature importance analysis of the CemaNeige GR6J model parameters for the calibration periods of a) 1964–2021, b) 1972–2021, c) 1982–2021, d) 1992–2021, e) 2002–2021, and f) 2012–2021.

System (HEC-HMS) hydrological model and an Encoder-Decoder-Long Short-Term Memory network for the flood forecasting in Vietnam. They found that the hybrid model enhanced the forecast ability to capture flood peaks and volume better than the HEC-HMS model. Wei et al. (2024) coupled the HBV and LSTM models and constituted three different hybrid model approaches. They revealed that the hybrid model approach provides more reliable simulation results for simulating extreme flows.

Although the stand-alone SAE-DNN model performed poorly in this study, this does not indicate a general failure of the deep learning model for discharge forecasting in karst systems. Wunsch et al. (2022) used station-based and gridded meteorological datasets as input data for modelling discharge in various karst systems, including the Unica karst system, using 1D and 2D convolutional neural networks (CNNs). They utilized precipitation, temperature, potential evapotranspiration, relative humidity, snow and new snow variables from the climate stations surrounding the Unica karst system as input variables in the 1D CNN model. They found that the 1D CNN model yielded satisfactory discharge simulation performance in the Unica karst system, although it was unable to capture the large peak events. What we learned from our own study is that a stand-alone deep learning model requires more than just precipitation and temperature data in order to reproduce correct discharge behavior of a hydrologically complex karst system, especially when seasonal snow cover is considered. Unlike previous studies, we used a process-based model to preprocess the precipitation and temperature variables in a hydrologically meaningful way and transfer these signals to a deep learning model. This work demonstrates that preprocessing the precipitation and temperature input data in a hydrological process-based manner significantly improves the performance of the applied deep learning model (SAE-DNN). This was achieved by coupling the deep learning model with the process-based model in sequence, without substantially changing the structure or parameterization of the deep learning model. It should also be noted that the type of deep learning model, its structure, and how its hyperparameters are determined are critical aspects of the performance of standalone deep learning models. This is an area that should receive more attention in future work.

### 4.2. Limitations of the proposed hybrid modelling approach

In previous works (Ćuk Đurović et al. 2022; Kogovšek et al., 2023), we know that the studied karst catchment is characterized by highly spatiotemporal variation in storage and discharge behavior of individual sub-catchments in eastern and southwestern part (dominated by strong interaction between groundwater and surface water due to karst poljes or shallow karst aquifer), in northwestern part (dominated by surface water), in central and southern part (dominated by groundwater in karst aquifer). The simulation results demonstrate that the proposed process-based model is able to reproduce the general catchment discharge behavior with accepted model performance for each modelling period. However, the proposed model structure is spatially lumped, and this limits the model to represent generally the individual sub-catchment specific storage and discharge processes and their interplay. Especially during high and low flow conditions, the process-based model provides the most deficit. With the extension in the hybrid model by coupling the SAE-DNN model in sequence, the deficit of the model performance can be improved consistently for each modelling period, i.e. the hybrid model could reveal the nonlinear transformation of preprocessed signals by the CemaNeige GR6J model into catchment discharge variability well despite the deficiency of the process-based model structure for the studied karst system. The current model structure does not require the detailed resolution of spatially varying hydrological processes, which would necessitate considerable effort in terms of site investigation and data/information availability. This is beneficial for practical work using the proposed hybrid model to reproduce highly variable catchment discharge behavior. However, the spatially lumped structure of the proposed hybrid model is limited in its ability to capture the effects of spatially varying weather conditions/parameters and their impact on catchment discharge behavior. In this study, we conducted additional time series analyses and identified changing climate conditions in the observation data sets. The hybrid model generally performs better during the periods with only “warm” condition (2002–2021 and 2012–2021) than the periods with mixed “cold” and “warm” conditions (1964–2021, 1972–2021, 1982–2021 and 1992–2021) due to the altered vegetation cover and changing snow cover behavior within the studied catchment area. To make the hybrid model structure more robust, we propose designing a parallel structure to involve multiple CemaNeige GR6J sub-models. This would enable the hybrid model to represent the storage and discharge behavior of each individual sub-catchment more specifically, while also capturing the spatially varying weather conditions across the studied catchment using a semi-distributed approach. We believe this would help to overcome the limitations of the current model structure in representing the interplay of spatially varying hydrological processes in individual sub-catchments.

Generally, the uncertainties in the input dataset, model parameters, and model structure should have a significant impact on the modelling performance of the process-based models (Beven and Smith, 2015; Teng et al., 2017; Herrera et al., 2022). There are various uncertainty resources related to the errors in observational data, parameter distribution, parameter estimation, model structure, and variability in the hydrometeorological variables for modelling by process-based models (Herrera et al., 2022). Similarly, uncertainties in determining the hyperparameters and uncertainties in input data in data-driven models can also lead to additional challenges in the hybrid modelling approach. In this regard, considering and analyzing the uncertainty resources in process-based and data-driven models is crucial for obtaining more reliable and robust hybrid modelling results. Testing different coupling strategies can also provide a more extensive outlook regarding how the simulation results can be enhanced and made more robust. Okkan et al. (2021) used the parallel and nested coupling strategies for monthly discharge forecasting. They revealed that the nested coupling approach, which involves the simultaneous calibration of all free parameters belonging to the process-based and data-driven models, could be a viable alternative to the parallel coupling approach, which considers the separate calibration of the parameters of the process-based and data-driven models. Even though the nested coupling approach can present challenges, such as overparameterization, it can be beneficial in offering diversified coupling approaches and enhancing the modelling results. Using different optimization algorithms for the calibration process of hybrid models can create a comprehensive perspective for parameter estimation, as shown in the study by Durgut and Ayvaz (2023). It would be beneficial for future work to include uncertainty and sensitivity analyses. Experimenting different optimization algorithms and coupling strategies in hybrid models would also improve our understanding of their limitations.

## 5. Conclusions

In this study, an innovative modelling approach has been developed that combines a process-based (CemaNeige GR6J) model and a data-driven (SAE-DNN) model to predict the discharge behavior of karst systems robustly. The modelling approach was tested on a large, hydrologically highly complex karst system in the Unica River catchment in Slovenia, where the catchment storage and discharge process is additionally influenced by seasonal snow cover, making catchment discharge prediction even more challenging. The performance of the stand-alone models CemaNeige GR6J, as well as SAE-DNN, and the hybrid model CemaNeige GR6J-SAE-DNN were systematically investigated and compared. The DE algorithm was used to calibrate and train the process-based, data-driven and hybrid models. A long-term historic dataset of precipitation, temperature and catchment discharge covering the period from 1962 to 2021 (almost 60 years) and its sub-periods (50 years, 40 years, 30 years, 20 years and 10 years) were used for current modelling work. This allowed the models to be tested and evaluated over a long period of time with significant variations in the inputs and outputs of the studied karst system. Model behavior during extreme flow conditions was systematically investigated using recession analysis, $Q_{95}$ percentile low discharge, 7-day maximum discharge analysis, flow duration curves, scatter plots and ridgeline plots. The results of this research can be summarized as follows:

- The hybrid model CemaNeige GR6J-SAE-DNN significantly enhanced the performance of the CemaNeige GR6J and SAE-DNN models’ performance across all modelling periods;
- Model performance can vary depending on the chosen sub-periods. The hybrid model's performance was found to be better in the last 20 years, which may be related to changing environmental conditions and, consequently, the changing storage and discharge behavior of the studied karst system;
- The hybrid model significantly enhanced catchment discharge simulation during both low and high flow conditions compared to the stand-alone models, providing the most robust discharge simulation among the models.

The current study demonstrates that the proposed hybrid modelling approach provides the most robust discharge prediction compared to process-based and data-driven models for the studied karst system. The influence of seasonal snow cover on catchment storage and discharge processes poses the greatest challenge to the performance of the data-driven model and is best predicted by the hybrid model. Although the hybrid model generally provides satisfactory performance during extreme low-flow conditions, its spatially lumped structure limits its ability to capture changing weather conditions within the studied catchment and to represent the varied hydrological processes of individuell sub-catchments, as well as their interplay. However, the proposed hybrid modelling approach offers significant benefits for practical work due to the low requirements for system understanding and observational data to set up the model. It can also be transferred to other karst systems with similar complexity and catchment characteristics, supporting robust decision-making in water resource management.

## CRediT authorship contribution statement

**C. Sezen:** Writing – original draft, Visualization, Validation, Software, Methodology, Funding acquisition, Formal analysis. **N. Ravbar:** Writing – review & editing, Writing – original draft, Validation, Data curation. **A. Hartmann:** Writing – review & editing, Validation. **Z. Chen:** Writing – original draft, Validation, Supervision, Methodology, Conceptualization.

## Funding

This research is part of the postdoctoral research of Dr. Cenk Sezen at the Institute of Groundwater Management at Technische Universität Dresden (TUD) in Germany, which was supported by the Scientific and Technological Research Council of Türkiye (TÜBİTAK) within the framework of the "TÜBİTAK 2219-International Postdoctoral Research Fellowship Program for Turkish Citizens". This research was also supported by the Slovenian Research and Innovation Agency (ARIS) through grants from the Slovenian National Research Programme Karst Research, No. P6-0119.

## 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.

## Acknowledgements

This research is part of the postdoctoral research of Dr. Cenk Sezen at the Institute of Groundwater Management at Technische Universität Dresden (TUD) in Germany, which was supported by the Scientific and Technological Research Council of Türkiye (TÜBİTAK) within the framework of the "TÜBİTAK 2219-International Postdoctoral Research Fellowship Program for Turkish Citizens". The authors would like to express their gratitude to TÜBİTAK for their support. They would also like to acknowledge the support of the Slovenian Research and Innovation Agency (ARIS) through grants from the Slovenian National Research Programme, Karst Research, No. P6-0119. The authors thank the four anonymous reviewers for improving the quality of the manuscript, as well as project DEAL for funding the open-access publication.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2026.135002.

## Data availability

The data are publicly available and can be obtained from the Slovenia Environment Agency (ARSO).

## References

Ahmad, M.F., Isa, N.A.M., Lim, W.H., Ang, K.M., 2022. Differential evolution: a recent review based on state-of-the-art works. Alex. Eng. J. 61, 3831–3872. https://doi.org/10.1016/j.aej.2021.09.013.

Alizadeh, Z., Yazdi, J., 2023. Calibration of hydrological models for ungauged catchments by automatic clustering using a differential evolution algorithm: the Gorganrood river basin case study. J. Hydroinf. 25, 645–662. https://doi.org/10.2166/hydro.2023.081.

Beven, K., Smith, P., 2015. Concepts of information content and likelihood in parameter calibration for hydrological simulation models. J. Hydrol. Eng. 20. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000991.

Blatnik, M., Gabrovšek, F., Ravbar, N., Frantar, P., Gill, L.W., 2024. Assessment of climatic and anthropogenic effects on flood dynamics in the Cerkniško Polje (SW Slovenia) based on a 70-year observation dataset. J. Hydrol.: Reg. Stud. 51, 101609. https://doi.org/10.1016/j.ejrh.2023.101609.

Bodian, A., Ndiaye, P.M., Diop, S.B., Diop, L., Dezetter, A., Ogilvie, A., Djaman, K., 2024. Evaluation and calibration of alternative methods for estimating reference evapotranspiration in the main hydrosystems of Senegal: Senegal, Gambia and Casamance River Basins. Proc. IAHS 385, 415–421. https://doi.org/10.5194/piahs-385-415-2024.

Bouhafa, N., Sakarovitch, C., Lalague, L., Goulard, F., Pryet, A., 2024. Hybrid modeling of karstic springs: error correction of conceptual reservoir models with machine learning. Water Supply 24, 1559–1573. https://doi.org/10.2166/ws.2024.092.

Chalikakis, K., Plagnes, V., Guerin, R., Valois, R., Bosch, F.P., 2011. Contribution of geophysical methods to karst-system exploration: an overview. Hydrgeol. J. 19, 1169–1180. https://doi.org/10.1007/s10040-011-0746-x.

Chang, Y., Wu, J., Jiang, G., Kang, Z., 2017. Identification of the dominant hydrological process and appropriate model structure of a karst catchment through stepwise simplification of a complex conceptual model. J. Hydrol. 548, 75–87. https://doi.org/10.1016/j.jhydrol.2017.02.050.

Chen, Z., Auler, A.S., Bakalowicz, M., Drew, D., Griger, F., Hartmann, J., Jiang, G., Moosdorf, N., Richts, A., Stevanovic, Z., Veni, G., Goldscheider, N., 2017. The World Karst Aquifer Mapping project: concept, mapping procedure and map of Europe. Hydrgeol. J. 25, 771–785. https://doi.org/10.1007/s10040-016-1519-3.

Chiang, S., Chang, C.-H., Chen, W.-B., 2022. Comparison of rainfall-runoff simulation between support vector regression and HEC-HMS for a rural watershed in Taiwan, 191–191 Water 14. https://doi.org/10.3390/w14020191.

Chollet, F., 2021. Deep Learning with Python, 2th ed. Simon and Schuster.

Cinkus, G., Wunsch, A., Mazzilli, N., Liesch, T., Chen, Z., Ravbar, N., Doummar, J., Fernández-Ortega, J., Barberá, J.A., Andreo, B., Goldscheider, N., Jourde, H., 2023. Comparison of artificial neural networks and reservoir models for simulating karst spring discharge on five test sites in the Alpine and Mediterranean regions. Hydrol. Earth Syst. Sci. 27, 1961–1985. https://doi.org/10.5194/hess-27-1961-2023.

Coron, L., Delaigue, O., Thirel, G., Dorchies, D., Perrin, C., Michel, C., 2022. airGR: Suite of GR Hydrological Models for Precipitation-Runoff Modelling.

Coron, L., Thirel, G., Delaigue, O., Perrin, C., Andréassian, V., 2017. The suite of lumped GR hydrological models in an R package. Environ. Model. Software 94, 166–171. https://doi.org/10.1016/J.ENVSOFT.2017.05.002.

Ćuk Đurović, M., Petrič, M., Jemcov, I., Mulec, J., Grudnik, Z.M., Mayaud, C., Blatnik, M., Kogovšek, B., Ravbar, N., 2022. Multivariate Statistical Analysis of Hydrochemical and Microbiological Natural Tracers as a Tool for Understanding Karst Hydrodynamics (the Unica Springs, SW Slovenia). Water Resour. Res. 58. https://doi.org/10.1029/2021WR031831.

Devi, G.K., Ganasri, B.P., Dwarakish, G.S., 2015. A review on hydrological models. Aquat. Procedia 4, 1001–1007. https://doi.org/10.1016/j.aqpro.2015.02.126.

Durgut, P.G., Ayvaz, M.T., 2023. A novel fully hybrid simulation-optimization approach for enhancing the calibration and verification performance of the TUW hydrological model, 128976–128976 J. Hydrol. 617. https://doi.org/10.1016/j.jhydrol.2022.128976.

Flores, N., Rodríguez, R., Yépez, S., Osores, V., Rau, P., Rivera, D., Balocchi, F., 2021. Comparison of three daily rainfall-runoff hydrological models using four evapotranspiration models in four small forested watersheds with different land cover in South-Central Chile, 3191–3191 Water 13. https://doi.org/10.3390/W13223191.

Ford, D., Williams, P., 2007. Karst Hydrogeology and Geomorphology. Wiley doi: 10.1002/9781118684986.

Gabrovšek, F., Kogovšek, J., Kovačič, G., Petrič, M., Ravbar, N., Turk, J., 2010. Recent results of tracer tests in the catchment of the Unica River (SW Slovenia). Acta Carsol. 39. https://doi.org/10.3986/ac.v39i1.110.

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. Hydrgeol. J. 28, 1661–1677. https://doi.org/10.1007/s10040-020-02139-5.

Gunathilake, M.B., Karunanayake, C., Gunathilake, A.S., Marasingha, N., Samarasinghe, J.T., Bandara, I.M., Rathnayake, U., 2021. Hydrological models and artificial neural networks (ANNs) to simulate streamflow in a tropical catchment of Sri Lanka. Appl. Comput. Intell. Soft Comput. 2021, 1–9. https://doi.org/10.1155/2021/6683389.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol. 377, 80–91. https://doi.org/10.1016/J.JHYDROL.2009.08.003.

Gustard, A., Demuth, S., 2009. Manual on Low-flow Estimation and Prediction. Koblenz.

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.

Herrera, P.A., Marazuela, M.A., Hofmann, T., 2022. Parameter estimation and uncertainty analysis in hydrological modeling. WIREs Water 9. https://doi.org/10.1002/wat2.1569.

Humphrey, G.B., Gibbs, M.S., Dandy, G.C., Maier, H.R., 2016. A hybrid approach to monthly streamflow forecasting: Integrating hydrological model outputs into a Bayesian artificial neural network. J. Hydrol. 540, 623–640. https://doi.org/10.1016/J.JHYDROL.2016.06.026.

Jukić, D., Denić-Jukić, V., 2009. Groundwater balance estimation in karst by using a conceptual rainfall–runoff model. J. Hydrol. 373, 302–315. https://doi.org/10.1016/j.jhydrol.2009.04.035.

Kogovšek, B., Jemcov, I., Petrič, M., 2023. Advanced application of time series analysis in complex karst aquifers: a case study of the Unica springs (SW Slovenia), 128976–128976 J. Hydrol. 626. https://doi.org/10.1016/j.jhydrol.2023.130147.

Kovačič, G., 2010. Hydrogeological study of the Malenščica karst spring (SW Slovenia) by means of a time series analysis. Acta Carsol. 39. https://doi.org/10.3986/ac.v39i2.93.

Kovačič, G., Petrič, M., Ravbar, N., 2020. Evaluation and quantification of the effects of climate and vegetation cover change on karst water sources: case studies of two springs in South-Western Slovenia, 3087–3087 Water 12. https://doi.org/10.3390/w12113087.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., Herrnegger, M., 2018. Rainfall–runoff modelling using long short-term memory (LSTM) networks. Hydrol. Earth Syst. Sci. 22, 6005–6022. https://doi.org/10.5194/hess-22-6005-2018.

Le, X.-H., Nguyen, D.H., Jung, S., Lee, G., 2023. Deep neural network-based discharge prediction for upstream hydrological stations: a comparative study. Earth Sci. Inf. 16, 3113–3124. https://doi.org/10.1007/s12145-023-01082-9.

Liu, G., Bao, H., Han, B., 2018. A stacked autoencoder-based deep neural network for achieving gearbox fault diagnosis. Math. Probl. Eng. 2018, 1–10. https://doi.org/10.1155/2018/5105709.

Liu, Y., Hou, G., Huang, F., Qin, H., Wang, B., Yi, L., 2022. Directed graph deep neural network for multi-step daily streamflow forecasting, 127515–127515 J. Hydrol. 607. https://doi.org/10.1016/j.jhydrol.2022.127515.

Lundberg, S.M., Erion, G.E., Lee, S.-I.L., 2019. Consistent individualized feature attribution for tree ensembles.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. In: Presented at the 31st Conference on Neural Information Processing Systems, pp. 1–10.

Maillet, E., 1905. Essais d'Hydraulique souterraine et fluviale. Nature 72, 25–26. https://doi.org/10.1038/072025a0.

Makumbura, R.K., Mampitiya, L., Rathnayake, N., Meddage, D.P.P., Henna, S., Dang, T. L., Hoshino, Y., Rathnayake, U., 2024. Advancing water quality assessment and prediction using machine learning models, coupled with explainable artificial intelligence (XAI) techniques like Shapley additive explanations (SHAP) for interpreting the black-box nature, 102831–102831 Results Eng. 23. https://doi.org/10.1016/j.rineng.2024.102831.

Margreth, M., Lustenberger, F., Hug Peter, D., Schlunegger, F., Zappa, M., 2024. Applying recession models for low-flow prediction: a comparison of regression and matching strip approaches. doi:10.5194/nhess-2024-78.

Mayaud, C., Gabrovšek, F., Blatnik, M., Kogovšek, B., Petrič, M., Ravbar, N., 2019. Understanding flooding in poljes: a modelling perspective. J. Hydrol. 575, 874–889. https://doi.org/10.1016/j.jhydrol.2019.04.092.

Mohammadi, B., Safari, M.J.S., Vazifehkhah, S., 2022. IHACRES, GR4J and MISD-based multi conceptual-machine learning approach for rainfall-runoff modeling, 12096–12096 Sci. Rep. 12. https://doi.org/10.1038/s41598-022-16215-1.

Mohammadi, B., Gao, H., Pilesjö, P., Tuo, Y., Guo, R., Duan, Z., 2025. Integrating machine learning with process-based glacio-hydrological model for improving the performance of runoff simulation in cold regions. J. Hydrol. 656, 132963. https://doi.org/10.1016/j.jhydrol.2025.132963.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — A discussion of principles. J. Hydrol. 10, 282–290. https://doi.org/10.1016/0022-1694(70)90255-6.

Nguyen, P.S., (Felix) Nguyen, T.H., Nguyen, T.H., 2024. A real-time flood forecasting hybrid machine learning hydrological model for Krong H'nang hydropower reservoir. River 3, 107–117. https://doi.org/10.1002/rvr2.72.

Okkan, U., Ersoy, Z.B., Ali Kumanlioglu, A., Fistikoglu, O., 2021. Embedding machine learning techniques into a conceptual model to improve monthly runoff simulation: a nested hybrid rainfall-runoff modeling, 126433–126433 J. Hydrol. 598. https://doi.org/10.1016/J.JHYDROL.2021.126433.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., Loumagne, C., 2005. Which potential evapotranspiration input for a lumped rainfall–runoff model? J. Hydrol. 303, 290–306. https://doi.org/10.1016/j.jhydrol.2004.08.026.

Perrin, C., Michel, C., Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. J. Hydrol. 279, 275–289. https://doi.org/10.1016/S0022-1694(03)00225-7.

Petric, M., 2010. Case study: characterization, exploitation, and protection of the Malenščica karst spring, Slovenia, in: Groundwater Hydrology of Springs. Elsevier, pp. 428–441. doi:10.1016/B978-1-85617-502-9.00021-9.

Pushpalatha, R., Perrin, C., Le Moine, N., Mathevet, T., Andréassian, V., 2011. A downward structural sensitivity analysis of hydrological models to improve low-flow simulation. J. Hydrol. 411, 66–76. https://doi.org/10.1016/J.JHYDROL.2011.09.034.

Rathnayake, N., Rathnayake, U., Chathuranika, I., Dang, T.L., Hoshino, Y., 2023. Cascaded-ANFIS to simulate nonlinear rainfall–runoff relationship, 110722–110722 Appl. Soft Comput. 147. https://doi.org/10.1016/j.asoc.2023.110722.

Sabzipour, B., Arsenault, R., Troin, M., Martel, J.-L., Brissette, F., Brunet, F., Mai, J., 2023. Comparing a long short-term memory (LSTM) neural network with a physically-based hydrological model for streamflow forecasting over a Canadian catchment, 130380–130380 J. Hydrol. 627. https://doi.org/10.1016/j.jhydrol.2023.130380.

Sauquet, E., Richard, B., Devers, A., Prudhomme, C., 2019. Water restrictions under climate change: a Rhône–Mediterranean perspective combining bottom-up and top-down approaches. Hydrol. Earth Syst. Sci. 23, 3683–3710. https://doi.org/10.5194/hess-23-3683-2019.

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

Sezen, C., Šraj, M., 2024a. Hourly rainfall-runoff modelling by combining the conceptual model with machine learning models in mostly Karst Ljubljanica River Catchment in Slovenia. Stoch. Environ. Res. Risk Assess. 38, 937–961. https://doi.org/10.1007/s00477-023-02607-w.

Sezen, C., Šraj, M., 2024b. Improving the simulations of the hydrological model in the karst catchment by integrating the conceptual model with machine learning models. Sci. Total Environ. 926, 171684. https://doi.org/10.1016/j.scitotenv.2024.171684.

Song, X., Hao, H., Liu, W., Wang, Q., An, L., Jim Yeh, T.-C., Hao, Y., 2022. Spatial-temporal behavior of precipitation driven karst spring discharge in a mountain terrain. J. Hydrol. 612, 128116. https://doi.org/10.1016/j.jhydrol.2022.128116.

Stevanović, Z., 2019. Karst waters in potable water supply: a global scale overview, 662–662 Environ. Earth Sci. 78. https://doi.org/10.1007/s12665-019-8670-9.

Storn, R., Price, K., 1997. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. J. Glob. Optim. 11, 341–359. https://doi.org/10.1023/A:1008202821328.

Teng, J., Jakeman, A.J., Vaze, J., Croke, B.F.W., Dutta, D., Kim, S., 2017. Flood inundation modelling: a review of methods, recent advances and uncertainty analysis. Environ. Model. Software 90, 201–216. https://doi.org/10.1016/j.envsoft.2017.01.006.

Tripathy, K.P., Mishra, A.K., 2024. Deep learning in hydrology and water resources disciplines: concepts, methods, applications, and research directions, 130458–130458 J. Hydrol. 628. https://doi.org/10.1016/j.jhydrol.2023.130458.

Valéry, A., Andréassian, V., Perrin, C., 2014. 'As simple as possible but not simpler': what is useful in a temperature-based snow-accounting routine? Part 2 – Sensitivity analysis of the Cemaneige snow accounting routine on 380 catchments. J. Hydrol. 517, 1176–1187. https://doi.org/10.1016/J.JHYDROL.2014.04.058.

Vilhar, U., Kermavnar, J., Kozamernik, E., Petrič, M., Ravbar, N., 2022. The effects of large-scale forest disturbances on hydrology – an overview with special emphasis on karst aquifer systems. Earth Sci. Rev. 235, 104243. https://doi.org/10.1016/j.earscirev.2022.104243.

Wei, Y., Wang, R., Feng, P., 2024. Improving hydrological modeling with hybrid models: a comparative study of different mechanisms for coupling deep learning models with process-based models. Water Resour. Manag. 38, 2471–2488. https://doi.org/10.1007/s11269-024-03780-5.

Willmott, C.J., 1981. On the validation of models. Phys. Geogr. 2, 184–194. https://doi.org/10.1080/02723646.1981.10642213.

Wunsch, A., Liesch, T., Cinkus, G., Ravbar, N., Chen, Z., Mazzilli, N., Jourde, H., Goldscheider, N., 2022. Karst spring discharge modeling based on deep learning using spatially distributed input data. Hydrol. Earth Syst. Sci. 26, 2405–2430. https://doi.org/10.5194/hess-26-2405-2022.

Xu, C., Zhong, P., Zhu, F., Xu, B., Wang, Y., Yang, L., Wang, S., Xu, S., 2024. A hybrid model coupling process-driven and data-driven models for improved real-time flood forecasting, 131494–131494 J. Hydrol. 638. https://doi.org/10.1016/j.jhydrol.2024.131494.

Xu, T., Longyang, Q., Tyson, C., Zeng, R., Neilson, B.T., 2022. Hybrid physically based and deep learning modeling of a snow dominated, mountainous, karst watershed. Water Resour. Res. 58, e2021WR030993. https://doi.org/10.1029/2021WR030993.

Zhang, W., Duan, L., Liu, T., Shi, Z., Shi, X., Chang, Y., Qu, S., Wang, G., 2024. A hybrid framework based on LSTM for predicting karst spring discharge using historical data. J. Hydrol. 633, 130946. https://doi.org/10.1016/j.jhydrol.2024.130946.

Zhong, M., Zhang, H., Jiang, T., Guo, J., Zhu, J., Wang, D., Chen, X., 2023. A Hybrid model combining the Cama-flood model and deep learning methods for streamflow prediction. Water Resour. Manag. 37, 4841–4859. https://doi.org/10.1007/s11269-023-03583-0.

Zhou, Q., Chen, L., Singh, V.P., Zhou, J., Chen, X., Xiong, L., 2019. Rainfall-runoff simulation in karst dominated areas based on a coupled conceptual hydrological model. J. Hydrol. 573, 524–533. https://doi.org/10.1016/j.jhydrol.2019.03.099.