



Hourly rainfall-runoff modelling by combining the conceptual model with machine learning models in mostly karst Ljubljana River catchment in Slovenia

Cenk Sezen¹ · Mojca Šraj²

Accepted: 28 October 2023 / Published online: 22 November 2023
 © The Author(s) 2023

Abstract

Hydrological modelling, essential for water resources management, can be very complex in karst catchments with different climatic and geologic characteristics. In this study, three combined conceptual models incorporating the snow module with machine learning models were used for hourly rainfall-runoff modelling in the mostly karst Ljubljana River catchment, Slovenia. Wavelet-based Extreme Learning Machine (WELM) and Wavelet-based Regression Tree (WRT) machine learning models were integrated into the conceptual CemaNeige Génie Rural à 4 paramètres Horaires (CemaNeige GR4H). In this regard, the performance of the hybrid models was compared with stand-alone conceptual and machine learning models. The stand-alone WELM and WRT models using only meteorological variables performed poorly for hourly runoff forecasting. The CemaNeige GR4H model as stand-alone model yielded good performance; however, it overestimated low flows. The hybrid CemaNeige GR4H-WELM and CemaNeige-WRT models provided better simulation results than the stand-alone models, especially regarding the extreme flows. The results of the study demonstrated that using different variables from the conceptual model, including the snow module, in the machine learning models as input data can significantly affect the performance of rainfall-runoff modelling. The hybrid modelling approach can potentially improve runoff simulation performance in karst catchments with diversified geological formations where the rainfall-runoff process is more complex.

Keywords Conceptual model with snow module · Hourly data · Hybrid modelling · Karst · Ljubljana River catchment · Machine learning

Abbreviations

ABC:	Artificial bee colony	d:	index of agreement
AE:	Actual evapotranspiration in CemaNeige GR4H model	db4:	Daubechies-4
AExch1	Actual exchange between catchments	DEM:	Digital elevation model
&AExch2:		DWT:	Discrete wavelet transformation
ANN:	Artificial neural network	E:	Evapotranspiration
CemaNeige	CemaNeige Génie Rural à 4 paramètres	GCM:	General circulation model
GR4H:	Horaires	G-ELM:	Geomorphology based integrated extreme learning machine
CWT:	Continuous wavelet transformation (CWT)	GMDH:	Group method of data
		GR:	Génie Rural
		GR4H:	Génie Rural à 4 paramètres Horaires
		GR4J:	Génie Rural à 4 paramètres Journalier
		GR6J:	Génie Rural à 6 paramètres Journalier
		GRNN:	Generalized regression neural network
		HBV:	Hydrologiska Byråns Vattenbalansavdelning
		I-ELM:	Integrated ELM

✉ Mojca Šraj
 mojca.sraj@fgg.uni-lj.si

¹ Faculty of Engineering, Ondokuz Mayıs University, 55139 Samsun, Turkey

² Faculty of Civil and Geodetic Engineering, University of Ljubljana, Jamova 2, Ljubljana, Slovenia

IHACRES:	Identification of unit hydrograph and component flows from rainfall, evapotranspiration, and streamflow	x3:	Routing store capacity in CemaNeige GR4H model
KGE:	Kling-Gupta efficiency	x4:	Unit hydrograph time constant
KNN:	K nearest neighbours regression	XGB:	Extreme gradient boosting
LR:	Linear regression		
LSTM:	Long short-term memory model (LSTM)		
MAE:	Mean absolute error		
Melt:	Average actual snow melt in CemaNeige GR4H model		
MLP:	Multilayer perceptron		
MLR:	Multiple linear regression		
MRA:	Multiresolution analysis		
MODWT:	Maximal over discrete wavelet transformation		
NSE:	Nash–Sutcliffe efficiency		
NWM:	National water model		
P:	Precipitation		
Perc:	Percolation		
Pliq:	Average liquid precipitation in CemaNeige GR4H model		
Pn-Ps:	The difference between net rainfall and the part of Pn filling the production store in CemaNeige GR4H model		
Psol:	Average solid precipitation in CemaNeige GR4H model		
Q:	Discharge		
QD:	Direct flow in CemaNeige GR4H model		
QR:	Routing store outflow in CemaNeige GR4H model		
RBFNN:	Radial basis function neural network		
RF:	Random forests		
SP:	Average snow pack (snow water equivalent) in CemaNeige GR4H model		
SMI:	Soil moisture index in CemaNeige GR4H model		
SOV:	Second-order Volterra series model		
SVM:	Support vector machine		
T:	Temperature		
WELM:	Wavelet-based extreme learning machine		
WGANN:	Wavelet-based genetic algorithm–artificial neural network		
W-M5 model:	Wavelet-based M5 model		
WRT:	Wavelet-based regression tree		
x1:	Production store capacity in CemaNeige GR4H model		
x2:	Groundwater exchange coefficient in CemaNeige GR4H model		

1 Introduction

The accurate forecasting of hydrological variables is crucial for water resources management and planning. Hydrological modelling can be critical not only to reveal the complex hydrological process but also taking measures against extreme hydrological events such as floods and droughts. Furthermore, hydrological modelling using high temporal resolution data e.g., on an hourly basis, has become increasingly important due to the sudden changes in hydrometeorological variables. In this regard, various hydrological modelling approaches such as conceptual, physically based, and data-driven models are widely implemented for rainfall-runoff modelling. The physically based models may require many hydrometeorological and catchment related variables, which represent the physical characteristics of the catchment (Abdulkareem et al. 2018). Conceptual models utilize simplified descriptions that incorporate the variables into hydrological processes. The data-driven models are based on the evaluation of the relationship between input and output data without considering the physical process in the catchment (Solomatine and Wagener 2011).

Conceptual models may have advantages for hydrological modelling because they require less data compared to the physically based models and considering the relationships between variables based on the hydrological process compared to the data-driven models. In this regard, conceptual models have been widely implemented for rainfall-runoff modelling in different catchments (Osuch et al. 2019; Al-Safi and Sarukkalige 2020; Mathevet et al. 2020; Lees et al. 2021). The Génie Rural (GR) lumped hydrological conceptual models, which have different versions for different time scales, such as Génie Rural à 4 paramètres Horaires (GR4H), and Génie Rural à 4 paramètres Journalier (GR4J), have become increasingly popular in the last two decades (Perrin et al. 2003; van Esse et al. 2013; Poncelet et al. 2017; Cantoni et al. 2022). The usefulness of the GR models has been demonstrated in various catchments (Lavtar et al. 2020; Darbandsari and Coulibaly 2020; Kumari et al. 2021). Lavtar et al. (2020) investigated the performance of the GR4J and Génie Rural à 6 paramètres Journalier (GR6J) in the sub-catchments of the Sava River

in Slovenia. They found that the GR6J performed better than GR4J, especially in larger sub-catchments. van Esse et al. (2013) implemented the SUPERFLEX and GR4H models for rainfall-runoff modelling in French catchments. They revealed that the performance of the conceptual models can be affected by catchment size, wetness, flashiness of flows and climatic characteristics. They stated that conceptual models performed better in larger and wetter catchments than in smaller and drier catchments. Valéry et al. (2014a, b) stated that using the snow accounting routine for snow-affected catchments could improve the modelling performance and they introduced the snow accounting routine named CemaNeige. Using the CemaNeige with GR models in the snow affected catchments can improve the performance of rainfall-runoff modelling as indicated by Sezen et al. (2018) and Hao et al. (2022). Accordingly, it can be pointed out that different factors such as catchment and climate characteristics as well as model structure can influence the modelling performance of the conceptual models.

Data-driven models are based on the relationship between the input and output variables without considering the physical process in the catchment. In this regard, data-driven models differ from conceptual and physically based models and are now widely used in hydrological modelling. The forecasting of hydrological variables such as precipitation, evapotranspiration, and runoff using data-driven models has increased in recent years (Unnikrishnan and Jothiprakash 2018; Hu et al. 2021; Sawaf et al. 2021; Wang et al. 2021; Chakraborty and Biswas 2023). These studies have demonstrated the reliability and applicability of the data-driven models for the estimation of the hydrological variables. To overcome the drawbacks of the data-driven models, such as the highly non-stationary data characteristics, techniques such as the wavelet transform have been used along with data-driven models for hydrological modelling (Nourani et al. 2014). Tayyab et al. (2019) analysed the performance of the integrated artificial neural network (ANN) with discrete wavelet transformation (DWT) for modelling rainfall-runoff in the Jinsha River catchment, China. They stated that DWT helped to improve the modelling performance and radial basis function neural network (RBFNN) with DWT yielded more satisfactory results. Roushangar et al. (2018) used the geomorphology based integrated extreme learning machine (G-ELM), an integrated ELM (I-ELM), and wavelet based extreme learning machine (WELM), and wavelet-based ANN (WANN) for modelling rainfall-runoff in the Ajichay catchment, Iran. They found that the WELM and I-ELM models performed better than the G-ELM model in predicting the peak values. They also pointed out that WELM gave more reliable results than WANN in dealing with the nonlinear characteristics of the rainfall-runoff process.

Nourani et al. (2019) compared the performance of the wavelet-based M5 model tree (W-M5 model) with WANN and the stand-alone M5 model tree for modelling rainfall-runoff in the Sardrud catchment in Iran. They found that the wavelet transform improved the modelling performance compared to the standalone M5 model tree and that the W-M5 model tree performed reliably for runoff simulation. It is clear that various data-driven modelling approaches continue to be used for rainfall-runoff modelling.

The hybrid modelling approaches, which integrate the conceptual and data-driven models have been used to overcome the drawbacks of both modelling approaches and improve the performance of rainfall-runoff modelling. The usefulness of the hybrid modelling approach for rainfall-runoff modelling has been shown in several studies (e.g., Sikorska-Senoner and Quilty 2021; Fattahi et al. 2022; Sezen and Partal 2022a). Sikorska-Senoner and Quilty (2021) implemented the Hydrologiska Byråns Vattenbalansavdelning (HBV) conceptual model, k nearest neighbours regression (KNN), multiple linear regression (MLR), second-order Volterra series model (SOV), ANN, random forests (RF), and extreme gradient boosting (XGB) data-driven models to analyse the performance of the hybrid model approach for modelling daily rainfall-runoff in the Dünneren, Kleine Emme, and Muota catchments in Switzerland. They indicated that the hybrid model approach yielded promising results in forecasting runoff and RF and XGB outperformed other data-driven models. Fattahi et al. (2022) integrated the identification of unit hydrograph and component flows from rainfall, evapotranspiration, and streamflow (IHACRES) model with group method of data (GMDH) for monthly runoff estimation in the Talesh-Anzali catchment, Iran. They pointed out that the integrated IHACRES-GMDH model approach improved the modelling performance compared to the IHACRES model. Sezen and Partal (2022b) combined the GR4J model with the wavelet-based genetic algorithm-artificial neural network (WGANN) for daily rainfall-runoff modelling in the Eastern Black Sea and Kızılırmak catchments in Turkey. They used different GR4J model outputs, such as soil moisture, routing store outflow, and direct runoff in the WGANN model as input data. They found that the hybrid modelling approaches performed better than the stand-alone GR4J model. They also stated that using the routing store outflow and direct runoff as input data provided more reliable results than using soil moisture as input data in the hybrid models. As can be seen, the hybrid model approach can not only have the advantage of using the conceptual model variables obtained by considering the hydrological process but also benefit from the data-driven models handling the complex relationship between input and output data.

The rainfall-runoff relationship and runoff could be more complex, particularly in karst catchments (Hartmann et al. 2014). In this regard, conceptual, physically based, and machine learning models have been implemented in karst catchments to improve rainfall-runoff modelling performance (Zhou et al. 2019; Sezen et al. 2019; Husic et al. 2022). Zhou et al. (2019) integrated the Xinanjiang conceptual model and a two reservoir-based karst model for predicting runoff in the karst and non-karst areas in the Lijiang River basin in China. They stated that Xinanjiang model integrated with karst model outperformed the conventional Xinanjiang model for simulating rainfall-runoff process. Sezen et al. (2019) compared the performance of the GR4J, ANN, deep neural network and regression tree models in the nonhomogeneous Ljubljana River catchment and its sub-catchments in Slovenia. They found that the ANN and deep neural network models yielded better performance than GR4J for modelling the hydrograph recession in the karst sub-catchments. Husic et al. (2022) compared the performance of the physically based Soil&Water Assessment tool, conceptual LUMP, and long short-term memory (LSTM) models in a fluvial stream and karst spring in Kentucky, USA. They stated that the LSTM model outperformed the physically based and conceptual models for the estimation of quick, intermediate, and slow flow contributions. However, they also pointed out that the process-based models yielded more exact time-fractal scaling of hydrological flow pathways than LSTM. Mayaud et al. (2023) investigated the water levels and flow pattern in Planinsko Polje (part of the Ljubljana River catchment) in Slovenia installing a monitoring system and implementing water balance analysis results in a conceptual model. They revealed that combining the water levels and flow measurements with a digital elevation model (DEM) can be useful for the evaluation of water balance and monitoring of floods in complex karst areas such as poljes. In most research, the conceptual, physically based and machine learning models were compared, or the enhanced versions of them were used to improve rainfall-runoff modelling in karst catchments. The hybrid modelling approach, including different model types, was rarely implemented, or not elaborately handled for the potential modelling improvement perspective in karst catchments. However, using the hybrid modelling approach can have the potential to overcome the drawbacks of the stand-alone conceptual, physically based and machine learning models in karst catchments where the rainfall-runoff process can be more complicated.

The main aim of this study is to integrate the GR4H model including the snow module (i.e., CemaNeige GR4H)

with the wavelet-based machine learning models, namely WELM and WRT for hourly rainfall-runoff modelling in a mostly karst catchment. In this respect, the objectives of the study can be stated as follows:

- Analysing the hybrid modelling performance for hourly rainfall-runoff modelling in a broad perspective via implementing distinctive data variables obtained from the conceptual model.
- Using the snow module variables of the conceptual model in a mostly karst and snow-affected catchment as input data in the machine learning models and observing the change in hybrid modelling performance.
- Analysing and comparing the performance of hybrid models and stand-alone conceptual and machine learning models in a mostly karst catchment.
- Determining the sensitivity of input variables in hybrid models.

In this regard, various outputs of the CemaNeige GR4H in the WELM and WRT models were used as input data. The performance of the hybrid models was compared with the stand-alone WELM, WRT, and CemaNeige GR4H models in the mostly karst Ljubljana River catchment in Slovenia. The artificial bee colony (ABC) optimization algorithm was used to determine parameters in the conceptual, data-driven, and hybrid models. Using different variables of the snow module, such as liquid and solid forms of precipitation, snowpack, and snowmelt, in addition to the outputs of the conceptual model, such as routing store outflow, direct runoff, actual evapotranspiration, and soil moisture in the machine learning models, represents the novelty of this study. In other words, the effects of different input variables, including snow-related data obtained from the conceptual model on hybrid modelling performance, were investigated comprehensively in a mostly karst and snow-affected catchment. Furthermore, implementing a hybrid modelling approach at an hourly scale and analysing the simulation performance of the conceptual, machine learning, and hybrid models at the extreme flows (i.e., low and high flows) also contributes to the novelty of the study. Testing the hybrid modelling performance using various wavelet-based data-driven models (i.e., WELM and WRT) is important to observe performance changes of the hybrid model. In addition, implementing the sensitivity analysis of the input variables in the hybrid models is one of the significant aspects of the study to investigate their effects on the rainfall-runoff modelling. In this way, the hybrid modelling approach in a mostly karst catchment for hourly rainfall-runoff modelling was analysed in a broad aspect.

2 Data and methods

In this section, firstly, the study area and used data characteristics were given in Sect. 2.1. The conceptual, stand-alone machine learning models and hybrid model structures were introduced in Sect. 2.2. Finally, the criteria for evaluating the models' performance were presented in Sect. 2.3.

2.1 Study area and data used

The hourly rainfall-runoff modelling was carried out in the Ljubljana River catchment in Slovenia, which has non-homogeneous characteristics. The elevation map and location of the study area were shown in Fig. 1. The Ljubljana River catchment has different features mostly based on the geological and karst area variety across the catchment (Sezen et al. 2019). The Ljubljana River catchment has approximately 1890 km² area and the altitude of the catchment ranges from 300 to 1800 m a.s.l. (Fig. 1) (Sapač et al. 2020). The Ljubljana River presents a significant area of the Sava River tributaries (Cotman et al. 2008). The river catchment mostly includes porous carbonate rocks and partly noncarbonate rocks, surface flows have mostly short duration, and rivers and streams submerge underground on several occasions across the flow path (Rusjan et al. 2019). The river catchment has transitional climate characteristics between sub-Mediterranean and continental climates with annual mean rainfall between 1400 and 2000 mm (Rusjan et al. 2019).

In this study, the period from 01 January 2016 00:00 to 31 December 2020 23:00 was used for hourly rainfall-runoff modelling. The first two months of the data period were used for warm-up process in the conceptual model. The rest of data was divided into two equal parts for the calibration (i.e., period between 01 March 2016 00:00 and 01 August 2018 11:00) and validation (i.e., period between 01 August 2018 12:00 and 31 December 2020 23:00) periods in conceptual, machine learning, and hybrid models. The data periods for the calibration and validation were determined according to the study carried out by Klemeš (1986). This research suggests the data partition as 50% for both calibration and validation periods if the data is long enough. The hourly precipitation (P) and temperature (T) data of the five meteorological stations, namely Topol pri Medvodah, Logatec, Postojna, Ljubljana, and Nova vas na Blokah were used in rainfall-runoff modelling (Fig. 1). The Thiessen polygons were applied to calculate the areal precipitation and temperature (Fig. 1). The hourly evapotranspiration (E) was calculated using the Oudin formula (Oudin et al. 2005). The usefulness of the Oudin formula for rainfall-runoff modelling was shown in the previous

studies (e.g., Kodja et al. 2020; Flores et al. 2021). The hourly discharge data (Q) of the gauging station Moste (Fig. 1) was used for modelling. The mean, standard deviation, minimum, maximum, skewness, and kurtosis statistics related to the hydrometeorological variables are given in Table 1. As can be seen in Table 1, average hourly precipitation, evapotranspiration, temperature, and runoff are about 0.2 mm/h, 0.1 mm/h, 10 °C, and 0.1 mm/h, respectively. In addition, the Pearson correlation coefficients between the precipitation, evapotranspiration, temperature, and runoff were given in Table 2. The Pearson correlation coefficients between the precipitation and runoff increase from 0.11 to 0.20 when the time lag increases from 0 to 10 h. In addition, the correlation coefficients between evapotranspiration and runoff are about -0.20 , and the correlation coefficients between temperature and runoff are about -0.33 as seen in Table 2.

2.2 Models

2.2.1 GR4H and CemaNeige GR4H conceptual models

The GR4H model is an hourly lumped conceptual model used for rainfall-runoff modelling (Mathevet 2005; Moine 2008). The GR4H model has four free parameters, namely, production store capacity (x_1 [mm]), groundwater exchange coefficient (x_2 [mm/h]), routing store capacity (x_3 [mm]), and unit hydrograph time constant (x_4 [h]). The precipitation and evapotranspiration are used as input data into the GR4H model. In this study, the CemaNeige snow module (Valéry et al. 2014a, b) was used together with the GR4H model for rainfall-runoff modelling. In the CemaNeige model, daily liquid equivalent water depth of total precipitations and daily air temperature are used as input data. The CemaNeige model has two free parameters, namely the cold content factor and the snowmelt factor (Valéry et al. 2014a). In this regard, the CemaNeige GR4H model has a total of six parameters that need to be calibrated. The ABC was used to calibrate the parameters of the CemaNeige GR4H model. The CemaNeige GR4H variables shown in Table 3, representing both GR4H and CemaNeige variables, were used as input data into the hybrid models. Further information on the implementation of the CemaNeige GR4H model can be found in Mathevet (2005), Moine (2008), Valéry et al. (2014a, b), Coron et al. (2017), and Coron et al. (2022).

2.2.2 Wavelet transformation

Wavelet transformation is a pre-processing technique that can help to reveal the characteristics of a time series, such as discontinuities and trends (Rezaie-Balf et al. 2021). Wavelet transformation can be useful to deal with the non-

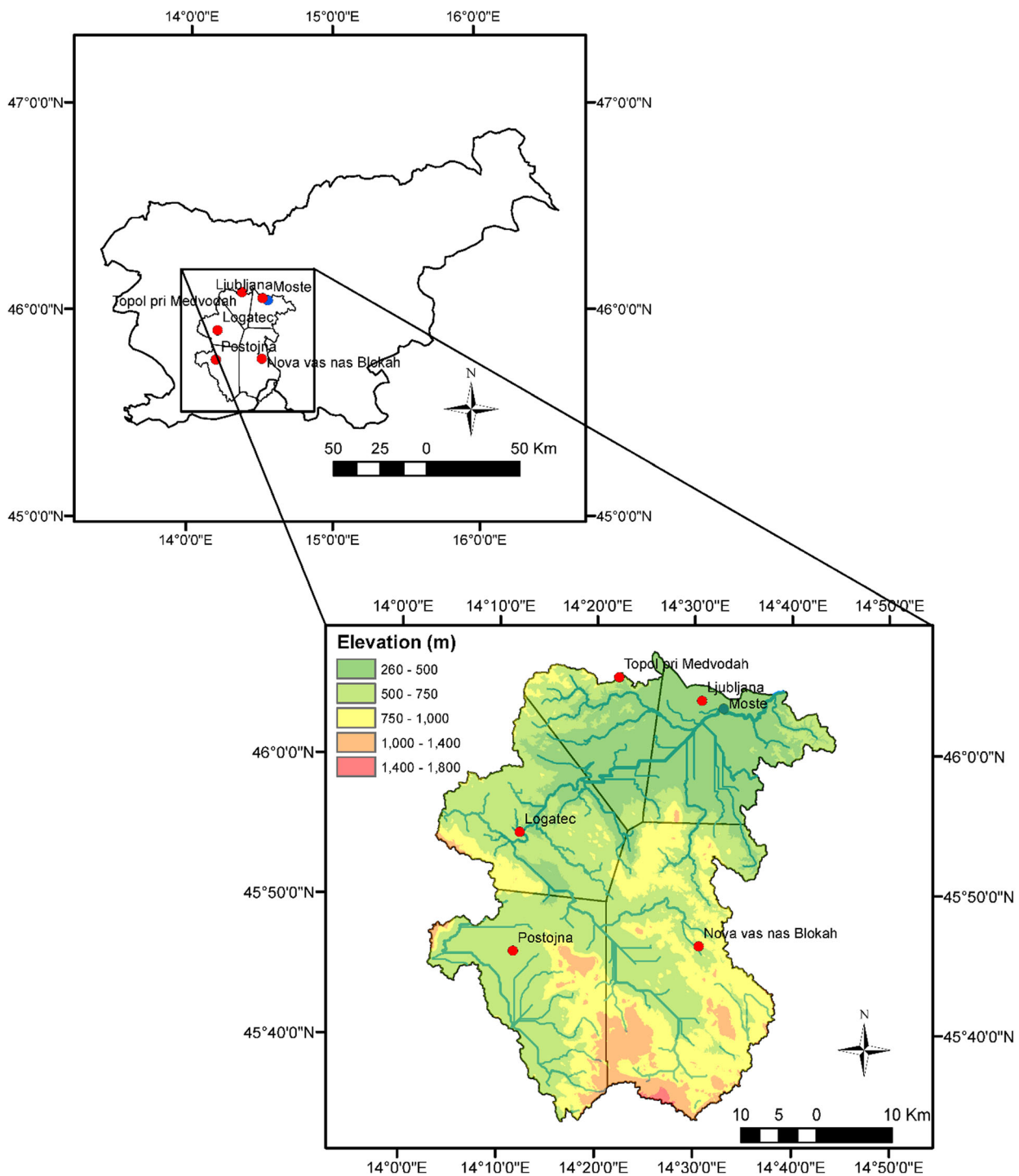


Fig. 1 The main characteristics and location of the study area. Red circles show considered meteorological stations and blue circle stands for the discharge gauging station

stationary features of the time series and show the temporal representation of the signal (Nourani et al. 2014). In this regard, the use of the wavelet transformation with the machine learning models has increased in hydrological modelling studies (Seo and Kim 2016; Sun et al. 2019; Graf et al. 2019; Bajirao et al. 2021). There are different

types of wavelet transformation, such as continuous wavelet transformation (CWT) and DWT. The DWT can provide sufficient information related to the time series and help to reduce the computational time compared to CWT (Wu and Kuo 2009). In this regard, the wavelet coefficients for specific scales are obtained in DWT. The DWT

Table 1 Basic statistics related to hourly precipitation, evapotranspiration, temperature, and discharge data in the selected period 2016–2020

Variables	Statistical data					
	Minimum	Mean	SD	Maximum	Skewness	Kurtosis
P (mm/h)	0	0.21	0.86	25.81	7.9	94.7
E (mm/h)	0	0.08	0.13	0.62	1.74	2.12
T (°C)	− 19.6	9.9	8.5	35.3	0.03	− 0.51
Q (mm/h)	0.01	0.11	0.11	0.55	1.54	1.57

Table 2 Pearson correlation coefficients of cross-correlations between hydrometeorological variables (precipitation (*P*), evapotranspiration (*E*), and temperature (*T*)) and discharge data (*Q*) for different time lags (*n* = 0–10 h)

n	Pearson correlation coefficients										
	0	1	2	3	4	5	6	7	8	9	10
$P_{t-n}-Q$	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20	0.20
$E_{t-n}-Q$	− 0.20	− 0.20	− 0.20	− 0.20	− 0.20	− 0.20	− 0.20	− 0.20	− 0.20	− 0.19	− 0.19
$T_{t-n}-Q$	− 0.34	− 0.34	− 0.34	− 0.33	− 0.33	− 0.33	− 0.33	− 0.33	− 0.33	− 0.33	− 0.32

Table 3 Variables obtained from the CemaNeige GR4H and used in the hybrid models as input data

Symbols	Variables
QR [mm/h]	Routing store outflow
QD [mm/h]	Direct flow
AE [mm/h]	Actual evapotranspiration
SMI	Soil moisture index
$Perc$ [mm/h]	Percolation
P_n-P_s [mm/h]	The difference between net rainfall and the part of P_n filling the production store
$AExch_1$ & $AExch_2$ [mm/h]	Actual exchange between catchments
$Pliq$ [mm/h]	Average liquid precipitation
$Psol$ [mm/h]	Average solid precipitation
SP [mm]	Average snow pack (snow water equivalent)
$Melt$ [mm/h]	Average actual snow melt

coefficients are obtained based on terms, such as decomposition level, time translation factor, fixed dilation step, and position parameter (Daubechies 1990). Thus, detail and approximation components of the time series are acquired using DWT. In other words, high- and low-frequency components of the time series are determined. In this study, maximal over discrete wavelet transformation (MODWT), which has the advantage of having no dyadic sample size limitation (Jiang et al. 2021), was used for the multiresolution analysis (MRA) and the mother wavelet of Daubechies-4 (db4) was implemented. The MRA analysis is based on the pyramid algorithm for time-series decomposition (Mallat 1989) and can help capturing the changes within different frequency intervals in the hydrological time series, such as streamflow data (Maslova et al. 2016).

The db4 was used as mother wavelet, which can be influential in generating time localization features for short memory time series (Nourani et al. 2014). In this study, the wavelet decomposition of level was selected as 14 according to the formula of $\text{int}(\log_2(N)-1)$ (Sang 2012), where *N* refers to the length of data.

The Boruta algorithm, which is a random forest-based feature selection technique, was implemented to determine significant wavelet components (Kursa et al. 2010; Kursa and Rudnicki 2010). The high performance of the Boruta algorithm for feature selection has been demonstrated in the previous studies (Prasad et al. 2019; Speiser et al. 2019). In this regard, important wavelet components of the input data detected by the Boruta algorithm were used in the stand-alone machine learning and hybrid models for runoff forecasting.

For this process, the maximum Z scores among shadow variables (MZSA) are determined and the attributes having lower and higher than MZSA were classified as unimportant and important, respectively (Kursa and Rudnicki 2010).

2.2.3 Artificial bee colony

The use of meta-heuristic algorithms such as swarm intelligence and evolutionary algorithms to solve optimization problems in water resources engineering has increased in the last decades (Reddy and Kumar 2020). Artificial bee colony, introduced by Karaboga (2005), is a swarm intelligence-based meta-heuristic algorithm that relies on the foraging behaviour of the honeybees. The forage selection, that triggers the collective intelligence of bee swarms, depends on three main components, namely food sources, employed foragers, and unemployed foragers. The exchange of information among bees is significant for the collective intelligence related to the quality of food resources, and the communication among bees occurs during the waggle dance in a dancing area (Karaboga 2005). In the ABC model, the artificial bee colony consists of three groups of bees, namely the employed bees, the onlookers, and the scouts. Accordingly, the ABC model consists of a process including determination of food sources and nectar amounts by employed bees, the preference and determination of food sources and nectar amounts by onlooker bees, the exploitation of the sources by the bees, the scouting of the search area and discovering new food sources and memorising the best food source found (Karaboga 2005). The ABC algorithm has been widely implemented for the calibration process in hydrological modelling studies (e.g., Huo et al. 2018; Farfán and Cea 2021). In this study, the ABC algorithm was used to calibrate the parameters in the conceptual, machine learning, and hybrid models. Detailed information about the ABC algorithm can be found in Karaboga (2005).

2.2.4 Extreme learning machine

Extreme learning machine can be considered as a training algorithm for the single hidden layer feedforward neural networks, which offers advantages such as fast convergence and satisfactory performance in classification, clustering, and regression problems (Wang et al. 2021). The training of ELM depends on random initialization and linear parameter solution. Initially, ELM uses random parameters for the weights (w), and the bias (b) in the hidden layer, which are frozen during the training stage. The input vector is mapped into a random feature space with random characteristics and nonlinear activation functions such as tangent sigmoid, sigmoid, and radial-basis functions. Second, the output weight (β) is

determined by the Moore–Penrose inverse as a linear problem $H\beta = T$, where H stands for the hidden layer output matrix and T for the training data target matrix (Huang et al. 2006; Wang et al. 2021). The usefulness of ELM, such as fast forecasting process and handling the complex problems for hydrological time series has been shown in many studies (e.g., Atiquzzaman and Kandasamy 2016; Zhu et al. 2019; Feng et al. 2022). In this study, the wavelet transformed input data were used in stand-alone ELM and hybrid models. In addition, the number of hidden neurons was determined using the ABC algorithm in stand-alone WELM and hybrid models including the WELM model. A detailed description and steps of the ELM model can be found in (Huang et al. 2006; Wang et al. 2021).

2.2.5 Decision trees

Decision trees can be used for classification or regression problems. Decision trees refer to the compact tree-like figurations of conditions, which determine when a decision should be implemented along with the action or decision (Dobra 2016). Decision trees include two basic node types, namely leaf node and non-leaf node (including root node). Each path from the root node to the leaf node can be regarded as a decision rule which includes a condition and a conclusion part (Berrar and Dubitzky 2013). Decision trees have two main processes, namely the growth and pruning phases. The growth phase refers to a recursive partitioning of the training dataset, where either each leaf node can be assigned to a class or a further portioning of the associated leaf node to its child nodes. The pruning phase targets the prevention of overfitting of the training dataset (Kotsiantis 2013). Decision trees have the advantage that the set of decision rules in the model structure is easy to understand, and the computational cost of the learning process is low to moderate (Berrar and Dubitzky 2013). In this regard, decision trees are widely used for the hydrological variables' estimation (e.g., Gharaei-Manesh et al. 2016; Nourani et al. 2019). In this study, regression trees were used to model hourly rainfall-runoff. Wavelet transformed input data were used in stand-alone RT and hybrid models including RT. Furthermore, the maximum depth of the tree was determined using the ABC algorithm in stand-alone WRT and hybrid models including WRT.

2.2.6 Hybrid models

Three different hybrid models were implemented for hourly rainfall-runoff modelling in this study. Accordingly, various outputs of the CemaNeige GR4H were used in three hybrid model structures. In this regard, the outputs of the conceptual model used in the hybrid model structures were presented in Table 4. The antecedent days of the

variables ($t-1$, $t-2$, $t-3$) were used as input data for rainfall-runoff modelling in stand-alone machine learning and hybrid models. Initially, the input variables were decomposed into components using the wavelet transform. Then, the Boruta algorithm was used to determine the important wavelet components, and these components were used in the ELM and RT models for runoff prediction. The hybrid model structures were shown in Fig. 2.

2.3 Model evaluation criteria

The performance of the conceptual, machine learning, and hybrid models was evaluated based on Nash–Sutcliffe efficiency (NSE) (Nash and Sutcliffe 1970), Kling–Gupta efficiency (KGE) (Gupta et al. 2009), index of agreement (d) (Willmott 1981), and mean absolute error (MAE) criteria. The formula, value ranges, and optimal values for each performance evaluation criterion are given in Table 5.

From Table 5, it can be seen that optimal performance occurs when the NSE, KGE, and d values approach 1. If the value of MAE value approaches 0, it means optimal performance of the models. In addition, when comparing the model performance, NSE values of 0.8 to 1, 0.7 to 0.8, 0.5 to 0.7, and less than 0.5 were considered as very good, good, satisfactory, and unsatisfactory performance, respectively, based on the study by Moriasi et al. (2015).

N stands for the data length, Q_{obs} for observed values, Q_{sim} for simulated values, \bar{Q}_{obs} for the mean of observed values in Table 5. Additionally, r stands for the correlation coefficients between the simulated and observed values, β for the ratio of the standard deviation of simulated values to the standard deviation of observed values, α for the ratio of mean of simulated values to the mean of observed values in the KGE formula.

Sensitivity analysis was performed to observe the effects of the input variables on rainfall-runoff modelling in the

Table 4 Used variables of the CemaNeige GR4H in hybrid models

Hybrid models	Used variables in hybrid models
CemaNeige GR4H-WELM ₁ , CemaNeige GR4H-WRT ₁	(AE, QR, QD) _{$t-1,t$} (AE, QR, QD) _{$t-2,t-1,t$} (AE, QR, QD) _{$t-3,t-2,t-1,t$}
CemaNeige GR4H-WELM ₂ , CemaNeige GR4H-WRT ₂	($SMI, Perc, P_n-P_s, AExch_1, AExch_2$) _{$t-1,t$} ($SMI, Perc, P_n-P_s, AExch_1, AExch_2$) _{$t-2,t-1,t$} ($SMI, Perc, P_n-P_s, AExch_1, AExch_2$) _{$t-3,t-2,t-1,t$}
CemaNeige GR4H-WELM ₃ , CemaNeige GR4H-WRT ₃	($Pliq, Psol, SP, Melt, AE, AExch_1, AExch_2$) _{$t-1,t$} ($Pliq, Psol, SP, Melt, AE, AExch_1, AExch_2$) _{$t-2,t-1,t$} ($Pliq, Psol, SP, Melt, AE, AExch_1, AExch_2$) _{$t-3,t-2,t-1,t$}

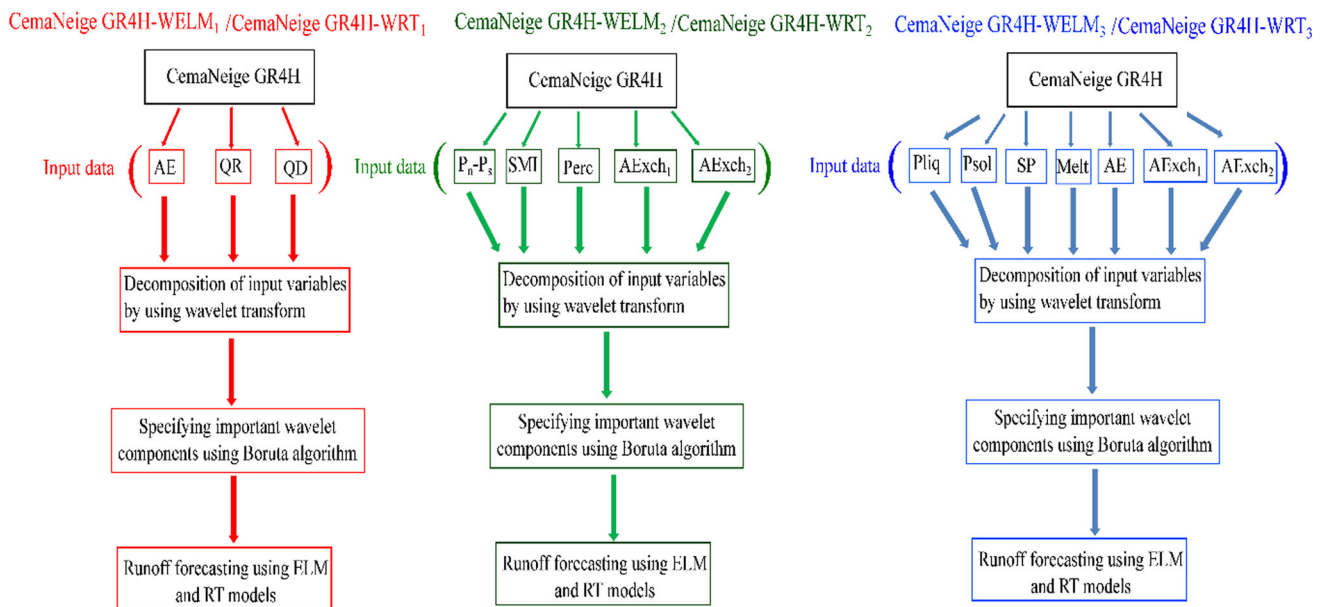


Fig. 2 The description of the hybrid models used in the study

Table 5 Model evaluation criteria used in the study

Evaluation criterion	Formula	Range	Optimal value
NSE	$NSE = \frac{\sum_{i=1}^N (Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^N (Q_{obs,i} - \bar{Q}_{obs,i})^2}$	$[-\infty, 1]$	1
KGE	$KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\alpha-1)^2}$	$[-\infty, 1]$	1
d	$d = 1 - \frac{\sum_{i=1}^N (Q_{sim,i} - Q_{obs,i})^2}{\sum_{i=1}^N (Q_{sim,i} - \bar{Q}_{obs} + Q_{obs,i} - \bar{Q}_{obs})^2}$	$[0, 1]$	1
MAE	$MAE = \frac{1}{N} \sum_{i=1}^N Q_{sim,i} - Q_{obs,i} $	$[0, \infty]$	0

hybrid models. In this regard, Bayesian Monte Carlo sensitivity analysis was used, based on a Latin hypercube sampling scheme for the prediction of the “Sobol” sensitivity indices described by Saltelli (2002). Accordingly, the calculation of the sensitivity indices is as follows:

First order for input i ,

$$S(i) = \text{Var}(E[f|x_i]) / \text{Var}(f) \quad (1)$$

where x_i stands for the i -th input and $S(i)$ for the first sensitivity indice. Furthermore, total effect for input i is computed as follows:

$$T(i) = \text{Var}(E[f|x_{-i}]) / \text{Var}(f) \quad (2)$$

where x_{-i} denotes all inputs except for the i -th input and $T(i)$ is the total effect sensitivity indice (Saltelli 2002; Gramacy 2007; Gramacy and Taddy 2010). The first-order indices quantify the portion of variability based on the variation in the main effects for each input variable, whereas total effect indices quantify the portion of variability based on the total variation in each input variable (Gramacy and Taddy 2022). For the implementation of Bayesian Monte Carlo sensitivity analysis, one can refer to Saltelli (2002) and Gramacy and Taddy (2022).

3 Results

3.1 Performance of the conceptual and data-driven models

Initially, hourly rainfall-runoff modelling was carried out using the GR4H and CemaNeige GR4H conceptual models. Simulation results for the calibration and validation periods were presented in Table 6. It was found that both the GR4H and CemaNeige GR4H models performed well in the Ljubljana River catchment. van Esse et al. (2013) investigated the performance of twelve different conceptual model structures from the SUPERFLEX framework and the GR4H model in 237 French catchments. They

revealed that conceptual hydrological models performed better in larger and wetter catchments than in smaller and drier catchments. They also stated that factors such as the flashiness of flows, and large variations in hydroclimatic conditions in the calibration and validation periods could negatively affect model performance. Newman et al. (2015) analysed the performance of the coupled Snow17 snow model and the Sacramento Soil Moisture Accounting Model in different sized catchments with various hydroclimatic characteristics across the contiguous United States. They found that model performance can vary by region and that factors such as aridity, precipitation intermittency, snowmelt, and runoff seasonality could affect the model performance. Accordingly, they pointed out that model performance could decrease in dry regions where snow is limited, while it could increase in catchments where snow water equivalent increases. The CemaNeige GR4H model yielded better simulation results than GR4H in both the calibration and validation periods, as shown in Table 6. The outperformance of the CemaNeige GR4H model compared to GR4H is more evident for the calibration period. This finding indicates that using the conceptual GR4H model with the CemaNeige snow module may be more reasonable than using a stand-alone GR4H model in the Ljubljana River catchment, which has highly elevated and snow-covered areas. The relationship between the simulated and observed streamflow for CemaNeige GR4H is shown in Fig. 3. Although the CemaNeige GR4H model performed well for rainfall-runoff modelling, it overestimated low flows and slightly underestimated high flows, as shown in Fig. 3. The scatter diagram, non-exceedance probability, and 30-days rolling mean plots clearly indicate overestimated low flows and underestimated high flows, as shown in Fig. 3. Low flows were overestimated, especially in the driest period of the year (i.e., June–September) in the Ljubljana River catchment. The nonhomogeneous characteristics of the catchment and discontinuities in surface flow across the flow path could hinder better forecasting performance of

Table 6 The performance of the conceptual, machine learning and hybrid models

Models	Input data	Model performance							
		Calibration				Validation			
		NSE	KGE	d	MAE (mm/h)	NSE	KGE	d	MAE (mm/h)
GR4H	$(P, E)_t$	0.80	0.84	0.94	0.035	0.84	0.79	0.95	0.03
CemaNeige GR4H	$(P, E)_t$	0.86	0.89	0.96	0.029	0.84	0.81	0.95	0.028
WELM	$(P, E)_{t-1,t}$	0.08	− 0.02	0.32	0.08	0	− 0.08	0.38	0.08
	$(P, E)_{t-2,t-1,t}$	0.10	0.04	0.40	0.08	0.01	− 0.01	0.42	0.08
	$(P, E)_{t-3,t-2,t-1,t}$	0.13	0.09	0.44	0.08	0.01	− 0.02	0.42	0.08
WRT	$(P, E)_{t-1,t}$	0.29	0.35	0.67	0.07	0.10	0.28	0.61	0.07
	$(P, E)_{t-2,t-1,t}$	0.31	0.38	0.69	0.07	0.12	0.31	0.63	0.07
	$(P, E)_{t-3,t-2,t-1,t}$	0.31	0.38	0.68	0.07	0.12	0.30	0.62	0.07
CemaNeige GR4H-WELM ₁	$(AE, QR, QD)_{t-1,t}$	0.92	0.94	0.98	0.022	0.89	0.91	0.97	0.022
	$(AE, QR, QD)_{t-2,t-1,t}$	0.92	0.94	0.98	0.022	0.89	0.91	0.97	0.022
	$(AE, QR, QD)_{t-3,t-2,t-1,t}$	0.92	0.94	0.98	0.022	0.89	0.91	0.97	0.022
CemaNeige GR4H-WRT ₁	$(AE, QR, QD)_{t-1,t}$	0.91	0.93	0.98	0.023	0.86	0.91	0.96	0.023
	$(AE, QR, QD)_{t-2,t-1,t}$	0.91	0.94	0.98	0.023	0.86	0.91	0.96	0.023
	$(AE, QR, QD)_{t-3,t-2,t-1,t}$	0.91	0.93	0.98	0.023	0.86	0.91	0.96	0.023
CemaNeige GR4H-WELM ₂	$(SMI, Perc, P_n-P_s, AExch_1, AExch_2)_{t-1,t}$	0.90	0.93	0.97	0.023	0.86	0.91	0.96	0.024
	$(SMI, Perc, P_n-P_s, AExch_1, AExch_2)_{t-2,t-1,t}$	0.91	0.93	0.97	0.021	0.87	0.91	0.96	0.022
	$(SMI, Perc, P_n-P_s, AExch_1, AExch_2)_{t-3,t-2,t-1,t}$	0.91	0.93	0.97	0.022	0.88	0.91	0.97	0.022
CemaNeige GR4H-WRT ₂	$(SMI, Perc, P_n-P_s, AExch_1, AExch_2)_{t-1,t}$	0.88	0.91	0.97	0.025	0.82	0.88	0.95	0.026
	$(SMI, Perc, P_n-P_s, AExch_1, AExch_2)_{t-2,t-1,t}$	0.88	0.91	0.97	0.025	0.82	0.88	0.95	0.026
	$(SMI, Perc, P_n-P_s, AExch_1, AExch_2)_{t-3,t-2,t-1,t}$	0.88	0.91	0.97	0.025	0.82	0.88	0.95	0.026
CemaNeige GR4H-WELM ₃	$(Pliq, Psol, SP, Melt, AE, AExch_1, AExch_2)_{t-1,t}$	0.89	0.92	0.97	0.024	0.84	0.86	0.95	0.025
	$(Pliq, Psol, SP, Melt, AE, AExch_1, AExch_2)_{t-2,t-1,t}$	0.89	0.92	0.97	0.024	0.85	0.85	0.96	0.024
	$(Pliq, Psol, SP, Melt, AE, AExch_1, AExch_2)_{t-3,t-2,t-1,t}$	0.89	0.92	0.97	0.024	0.85	0.85	0.96	0.024
CemaNeige GR4H-WRT ₃	$(Pliq, Psol, SP, Melt, AE, AExch_1, AExch_2)_{t-1,t}$	0.88	0.91	0.97	0.025	0.82	0.89	0.95	0.026
	$(Pliq, Psol, SP, Melt, AE, AExch_1, AExch_2)_{t-2,t-1,t}$	0.88	0.91	0.97	0.025	0.82	0.89	0.95	0.026
	$(Pliq, Psol, SP, Melt, AE, AExch_1, AExch_2)_{t-3,t-2,t-1,t}$	0.88	0.91	0.97	0.025	0.82	0.89	0.95	0.026

the conceptual model. Sapač et al. (2020) stated that variations in hydrogeological conditions in nonhomogeneous catchments such as the Ljubljana River catchment could significantly affect the rainfall-runoff process and surface runoff pattern.

The stand-alone WELM and WRT models performed poorly in hourly rainfall-runoff modelling (Table 6). The

performance of WELM is very poor in both the calibration and validation periods. Using different lags of the hydrometeorological variables did not noticeably improve the performance of the WELM model. Although the stand-alone WRT model performed better than the WELM model, it also yielded poor performance, especially in the validation period. Similar to the WELM, the

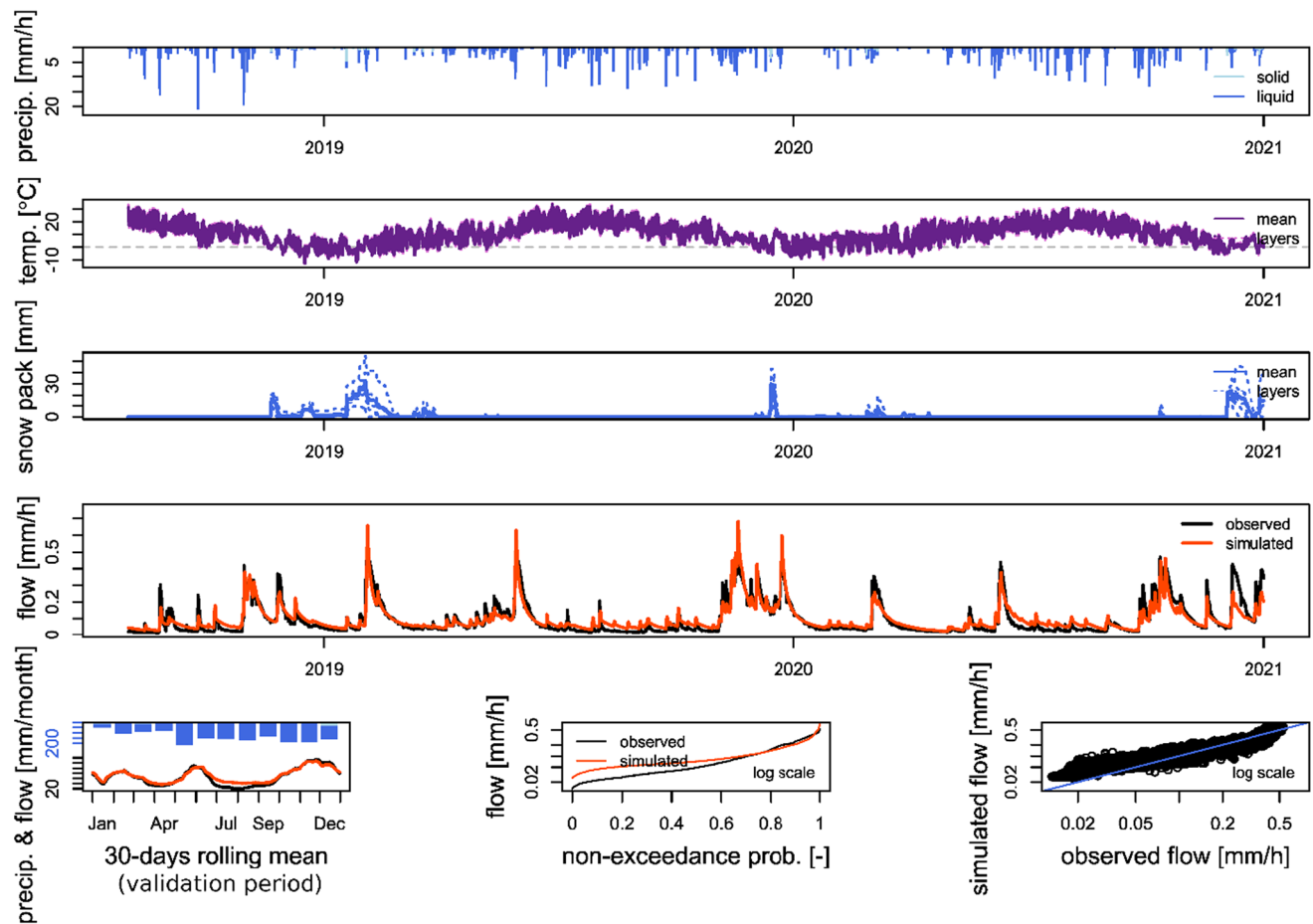


Fig. 3 Comparison between observed and simulated hourly runoff at the gauging station Moste on the Ljubljana River obtained as CemaNeige GR4H result

use of different lags in the hydrometeorological input data slightly improved the modelling performance. In other words, using only precipitation and evapotranspiration data in stand-alone machine learning models as input data did not yield satisfactory runoff prediction results. There are important factors, which can influence the machine learning models' performance, such as used input data, data length, and pre-processing techniques for runoff forecasting. Moosavi et al. (2022) investigated the role of input data, model type, data length, and pre-processing technique for runoff forecasting. They revealed that input variables and data length are more important than pre-processing technique and model type for runoff forecasting. In other words, if important explanatory input variables are not used in the machine learning models, the model type and pre-processing technique used could not provide satisfactory simulation results. In this regard, using only precipitation and evapotranspiration as input data without considering the hydrological process in stand-alone machine learning models did not provide promising simulation results.

3.2 Performance of the hybrid model approaches

Initially, hourly rainfall-runoff modelling was carried out using first the hybrid model approach, which includes *AE*, *QR*, and *QD* as input data. The wavelet components of *AE*, *QR*, and *QD* were obtained, and importance of each wavelet component was determined. The importance of the wavelet components is presented in Fig. 4. Accordingly, only the first components of *AE* were rejected, and it was found that the wavelet components of *AE* were more important than *QR* and *QD*. The CemaNeige GR4H-WELM₁ yielded very good performance based on the NSE criterion, reaching 0.92 in the calibration period and 0.89 in the validation period (Table 6). The other performance criteria (i.e., KGE, d, and MAE) also indicated the out-performance of the CemaNeige GR4H-WELM₁ compared to the stand-alone conceptual and data-driven models. The relationship between simulated and observed runoff, 30-days rolling mean, non-exceedance probability, and scatter diagrams for CemaNeige GR4H-WELM₁ are shown

Fig. 4 Determination of important wavelet components of actual evapotranspiration (AE), routing store flow (QR), and direct flow (QD) input variables via the Boruta algorithm for the CemaNeige GR4H-WELM₁ and CemaNeige GR4H-WRT₁ models. Blue, red, and green box plots with the circles having the same colour show important, unimportant, and shadow attributes, respectively

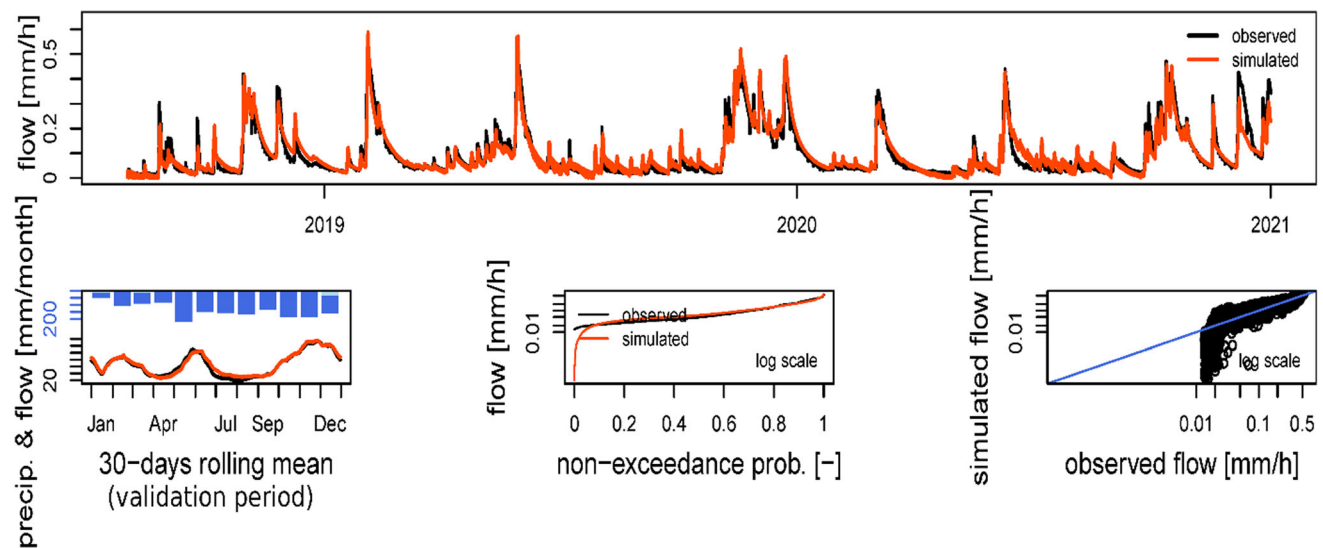
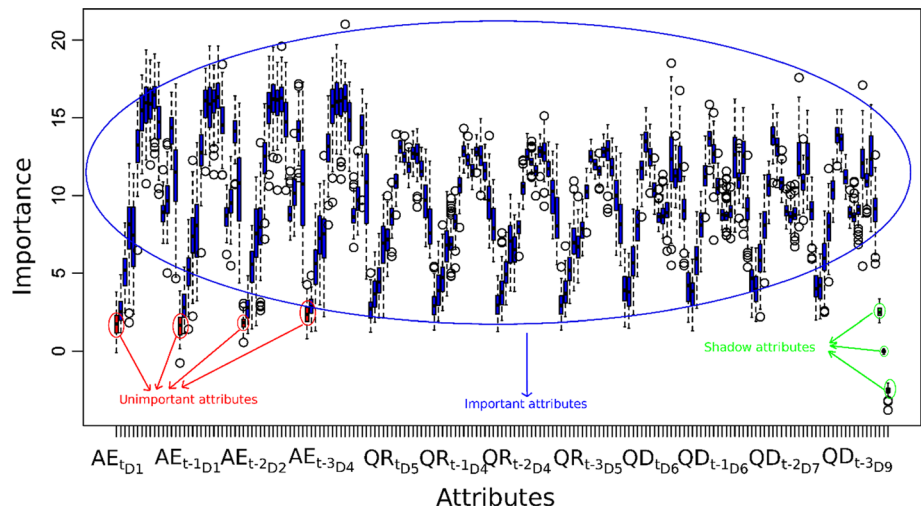


Fig. 5 Comparison between observed and simulated hourly runoff at the gauging station Moste on the Ljubljana River obtained as CemaNeige GR4H-WELM₁ result

in Fig. 5. Although low flows are underestimated, CemaNeige GR4H-WELM₁ performed better than the CemaNeige GR4H model in simulating low flows, especially in the dry period. Furthermore, the CemaNeige GR4H-WELM₁ model predicted high flows well, as can be seen in Fig. 5. The CemaNeige GR4H-WRT₁ model performed well in hourly runoff forecasting for the calibration and validation periods as shown in Table 6. Although CemaNeige GR4H-WRT₁ performed better than stand-alone conceptual and data-driven models, it was relatively weak compared to CemaNeige GR4H-WELM₁. The relationship between the simulated and observed streamflow is indicated for CemaNeige GR4H-WRT₁ in Fig. 6. Although data scatter is still present, the simulated and observed values in CemaNeige GR4H-WRT₁ are more seasonally

coherent than in CemaNeige GR4H according to the 30-days rolling mean plot (Fig. 6). In CemaNeige GR4H-WELM₁ and CemaNeige GR4H-WRT₁, using lags (i.e., implementing lags to $t-3$) of the hydrometeorological input variables did not produce more promising simulation results (Table 6). It can be concluded that using short-term lags such as 2 or 3 h for each input variable did not significantly affect the runoff forecasting performance.

Second, the performance of the hybrid models containing SMI , $Perc$, P_n-P_s , $AExch_1$, and $AExch_2$ as input data was investigated for hourly rainfall-runoff modelling. The importance of the wavelet components belonging to SMI , $Perc$, P_n-P_s , $AExch_1$, and $AExch_2$ is shown in Fig. 7. According to Fig. 7, it can be seen that the wavelet components of SMI , $Perc$, and $AExch_2$ have greater significance

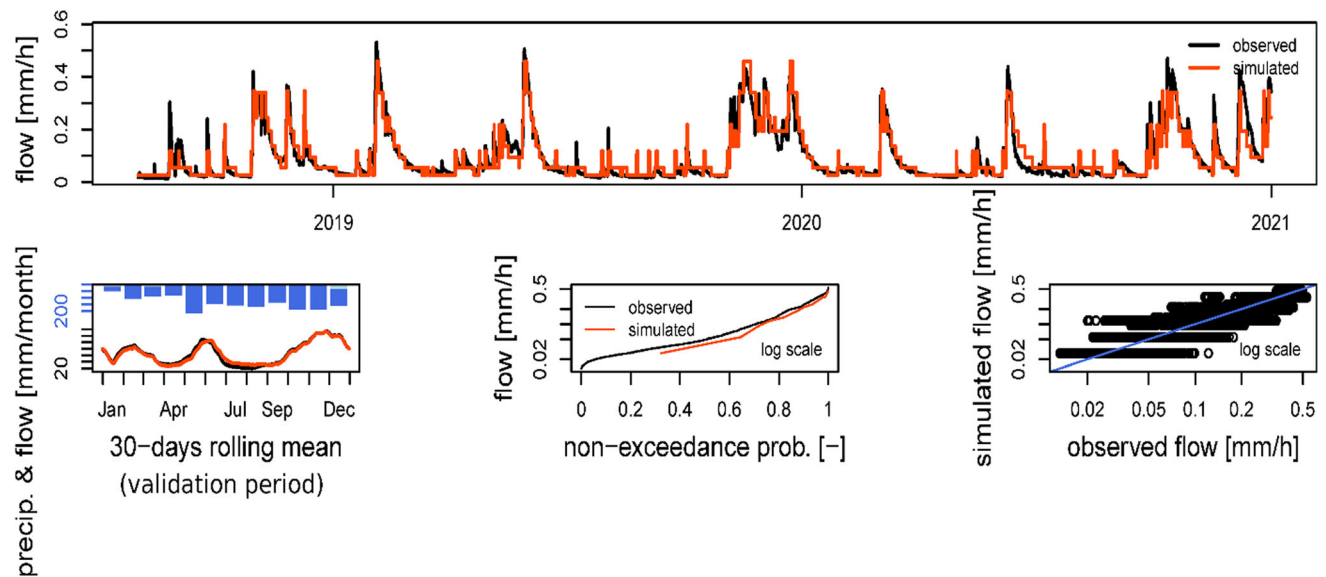
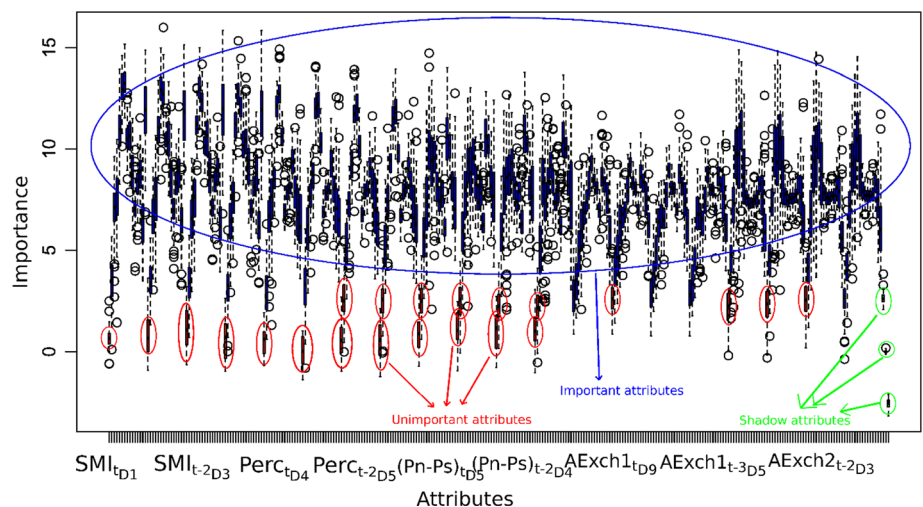


Fig. 6 Comparison between observed and simulated hourly runoff at the gauging station Moste on the Ljubljanica River obtained as CemaNeige GR4H-WRT₁ result

Fig. 7 Determination of important wavelet components of soil moisture index (*SMI*), percolation (*Perc*), and rainfall component (P_n-P_s), and actual exchange between catchments ($AExch_1$ and $AExch_2$) input variables via the Boruta algorithm for the CemaNeige GR4H-WELM₂ and CemaNeige GR4H-WRT₂ models. Blue, red, and green box plots with the circles having the same colour show important, unimportant, and shadow attributes, respectively



than other variables. In addition, most of the D_1 components of *SMI*, *Perc*, P_n-P_s , and $AExch_2$ were rejected, and these components were discarded from the input variables. The CemaNeige GR4H-WELM₂ yielded very good performance for the calibration and validation periods according to the NSE values (Table 6). Other performance criteria also show the satisfactory performance and out-performance of CemaNeige GR4H-WELM₂ against the stand-alone data-driven and conceptual models. However, the CemaNeige GR4H-WELM₂ model did not perform better than CemaNeige GR4H-WELM₁. The linkage between the simulated and observed streamflow for CemaNeige GR4H-WELM₂ is given in Fig. 8. The CemaNeige GR4H-WELM₂ model overestimated high flows, especially during the January–April, May–July, and October–December periods, as shown in Fig. 8. The low

flows in the July–September period, which is a critical period for the Ljubljanica River catchment, were well predicted by CemaNeige GR4H-WELM₂. The scatter and non-exceedance probability plots indicate the overestimation of high flows and underestimation of low flows by CemaNeige GR4H-WELM₂. The use of lags for each input variable slightly improved the runoff forecasting performance of CemaNeige GR4H-WELM₂ in the calibration and validation periods (Table 6). The CemaNeige GR4H-WRT₂ model performed very well for hourly runoff forecasting; however, it performed poorly compared to CemaNeige GR4H-WELM₂. In addition, Table 6 shows that the CemaNeige GR4H-WRT₂ model performed better than the stand-alone machine learning and conceptual models. In this regard, the superior performance of CemaNeige GR4H-WRT₂ over the stand-alone models was

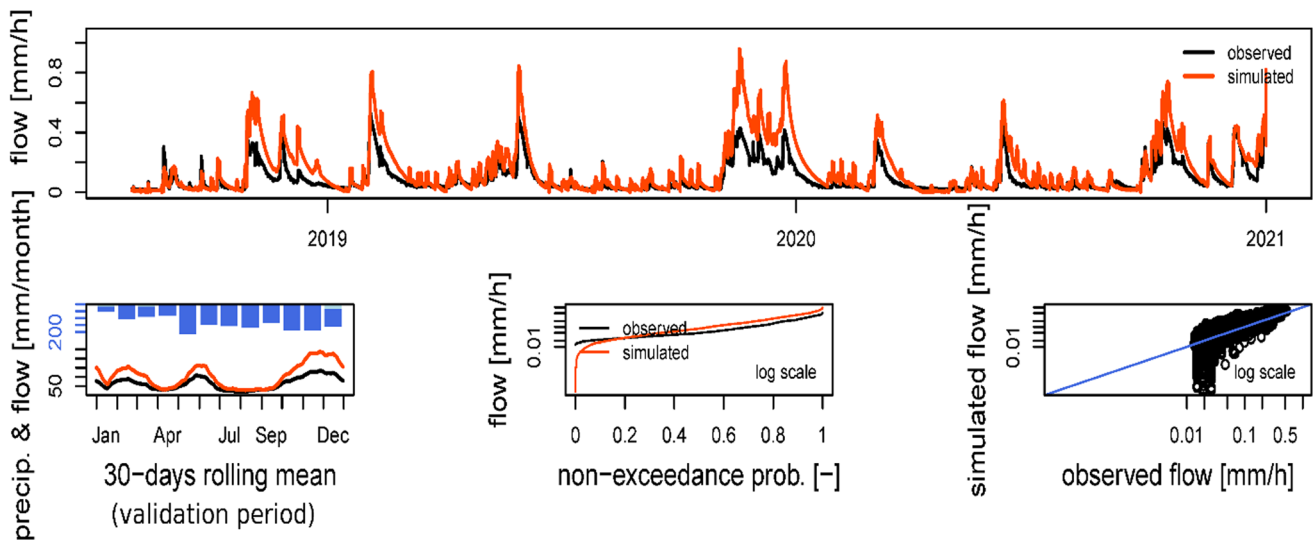


Fig. 8 Comparison between observed and simulated hourly runoff at the gauging station Moste on the Ljubljanica River obtained as CemaNeige GR4H-WELM₂ result

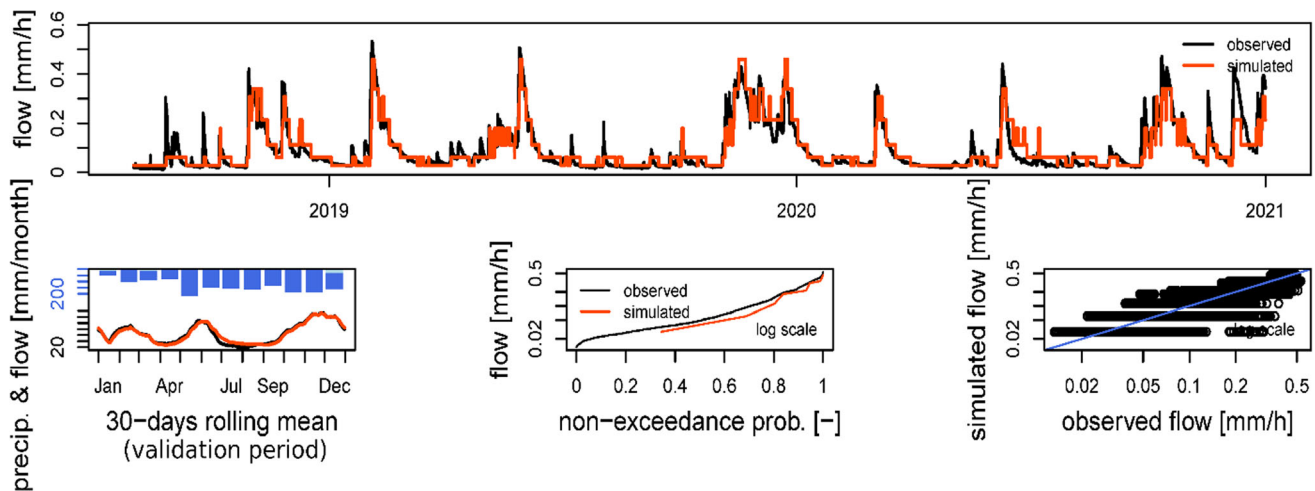


Fig. 9 Comparison between observed and simulated hourly runoff at the gauging station Moste on the Ljubljanica River obtained as CemaNeige GR4H-WRT₂ result

observed, especially in the calibration period according to NSE, KGE, d, and MAE. The relationship between the simulated and observed streamflow for CemaNeige GR4H-WRT₂ is presented in Fig. 9. Although the CemaNeige GR4H-WRT₂ model has scatter, it fits the observed flow pattern seasonally (Fig. 9). However, CemaNeige GR4H-WRT₂ slightly overestimated low flows, especially during the July–September period. Using various lags of each input variable did not improve the performance of the of CemaNeige GR4H-WRT₂ for hourly rainfall-runoff modelling (Table 6). Furthermore, the CemaNeige GR4H-WRT₂ model did not provide more reliable results than CemaNeige GR4H-WRT₁ according to the performance evaluation criteria.

In the third hybrid model approach, *Pliq*, *Psol*, *SP*, *Melt*, *AE*, *AExch₁*, and *AExch₂* were used as input data. The input variables were decomposed using the wavelet analysis and the importance of the wavelet components is shown in Fig. 10. According to Fig. 10, *Pliq*, *AE*, *AExch₁*, and *AExch₂* variables have more importance than other variables. Especially, the *D₁* and *D₂* components of the *Pliq*, *AE*, *AExch₁*, and *AExch₂* variables were found to be unimportant. Although CemaNeige GR4H-WELM₃ underestimated low flows according to the scatter plot, it yielded good runoff simulation results (Fig. 11). The performance evaluation criteria also indicated very good performance of CemaNeige GR4H-WELM₃ (Table 6). In this regard, the NSE values reached 0.89 and 0.85 in the calibration and validation periods, respectively. The

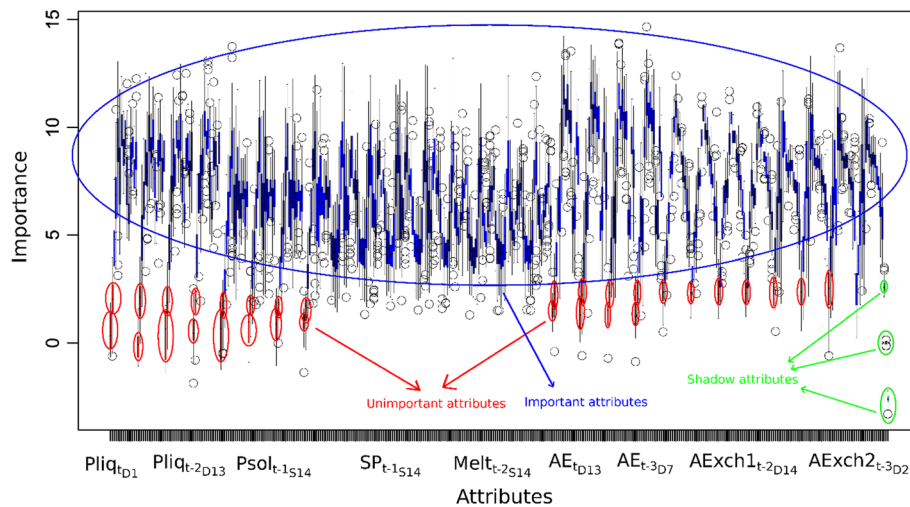


Fig. 10 Determination of important wavelet components of liquid precipitation (*Pliq*), solid precipitation (*Psol*), snowpack (*SP*), actual snow melt (*Melt*), actual evapotranspiration (*AE*), and actual exchange between catchments (*AExch₁* and *AExch₂*) input variables

via the Boruta algorithm for the CemaNeige GR4H-WELM₃ and CemaNeige GR4H-WRT₃ models. Blue, red, and green box plots with the circles having the same colour show important, unimportant, and shadow attributes, respectively

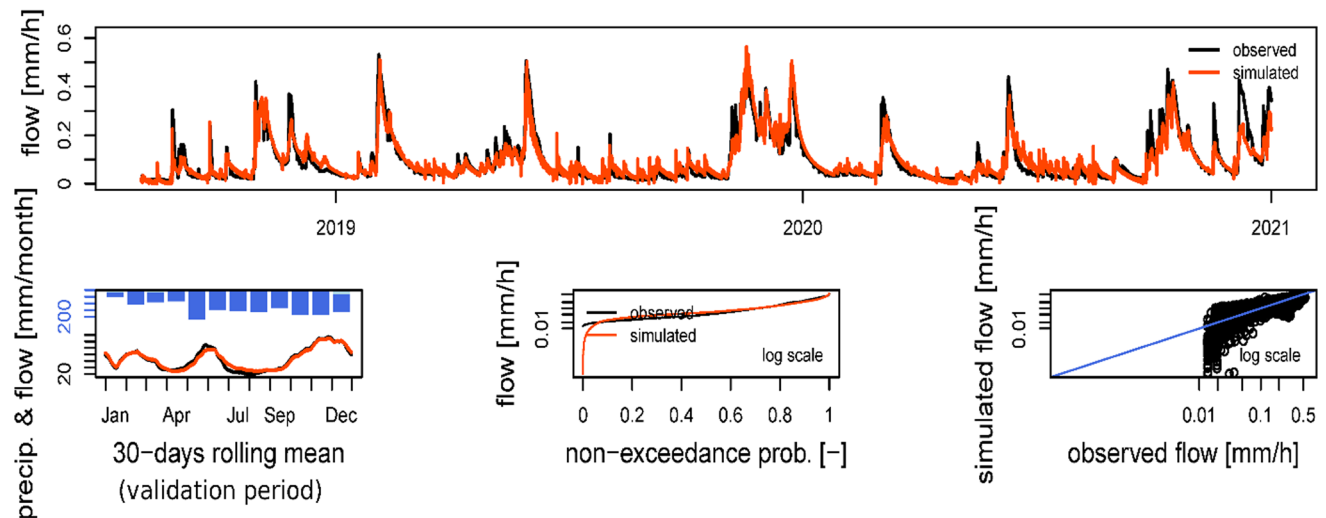


Fig. 11 Comparison between observed and simulated hourly runoff at the gauging station Moste on the Ljubljana River obtained as CemaNeige GR4H-WELM₃ result

relationship between the simulated and observed streamflow for CemaNeige GR4H-WELM₃ is shown in Fig. 11. The CemaNeige GR4H-WELM₃ model slightly overestimated flows during the July–September period; however, the simulated flows agreed with the observed flows on the monthly scale (Fig. 11). Besides, the CemaNeige GR4H-WELM₃ model performed poorly compared to CemaNeige GR4H-WELM₁ and CemaNeige GR4H-WELM₂ when the values of NSE, KGE, *d*, and MAE were considered. The CemaNeige GR4H-WRT₃ model outperformed the stand-alone conceptual and machine learning models (Table 6). The CemaNeige GR4H-WRT₃ model yielded poorer runoff simulation results than CemaNeige GR4H-WRT₁, whereas it performed similarly to CemaNeige GR4H-WRT₂. The

linkage between the simulated and observed streamflow for CemaNeige GR4H-WRT₃ is indicated in Fig. 12. According to the scatter plot, the CemaNeige GR4H-WRT₃ model underestimated low flows (Fig. 12). Moreover, the CemaNeige GR4H-WRT₃ simulated flows are compatible with observed flow on the monthly scale. Similar to CemaNeige GR4H-WELM₃, CemaNeige GR4H-WRT₃ slightly overestimated flows during the June–September period. Using several lags for each input variable slightly improved the modelling performance of the CemaNeige GR4H-WELM₃ model in the validation period, while it did not improve the performance of the CemaNeige GR4H-WRT₃ model.

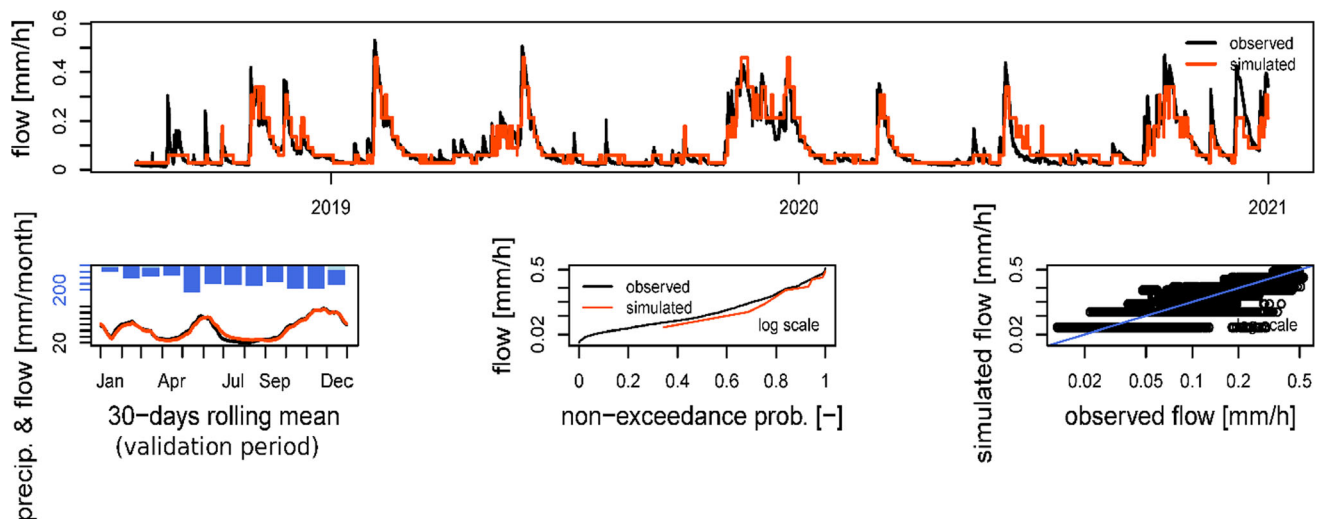


Fig. 12 Comparison between observed and simulated hourly runoff at the gauging station Moste on the Ljubljana River obtained as CemaNeige GR4H-WRT₃ result

Using several variables from CemaNeige GR4H in different hybrid models indicated that implementing different input variables could affect the performance of the hybrid model for hourly rainfall-runoff modelling. Accordingly, the most reliable performance in hourly runoff simulation was obtained when actual evapotranspiration, routing store outflow, and direct flow were used as input data in the WELM and WRT models in the Ljubljana River catchment. The similar findings were also obtained by Sezen and Partal (2022a, b). In this regard, Sezen and Partal (2022a) stated that using the outflows of the routing stores and direct flow obtained from the GR6J model performed better in the WGANN model than the use of the soil moisture index obtained via GR6J in the semi-arid Konya Closed basin, which has different climatic characteristics in Turkey. Similarly, using the routing store outflow and direct flow obtained from GR4J as input data to the WGANN model also provided better daily runoff simulation results than using the soil moisture index obtained from GR4J in the humid parts of the Kızılırmak and Eastern Black Sea basins in Turkey, as pointed out by Sezen and Partal (2022b). Implementing *SMI*, *Perc*, P_n - P_s , *AExch₁*, and *AExch₂* as input data in the second hybrid model approach improved the hourly runoff forecasting performance compared to the conceptual and machine learning models. Ancyil et al. (2004) implemented the soil moisture index obtained from GR4J as input data in the ANN model and found that using the soil moisture index as input data in the data-driven model enhanced the daily runoff forecasting performance. Tayfur et al. (2014) used soil moisture and rainfall as input data in a generalized regression neural network (GRNN) for hourly runoff forecasting in two small sub-catchments of the Tiber River catchment in Italy. They found that using soil moisture together with rainfall as

input data in GRNN remarkably improves the performance of hourly runoff forecasting performance. The performance improvement of the machine learning model when using soil moisture as input data has been found in similar studies (Casper et al. 2007; Ba et al. 2018). Using *Pliq*, *Psol*, *SP*, *Melt*, *AE*, *AExch₁*, and *AExch₂* as input data in the third hybrid model approach enhanced the runoff prediction performance compared to conceptual and machine learning models; however, the third hybrid model approach performed poorly compared to other hybrid model approaches. It could be stated that the implementation of input variables obtained considering the hydrological process stages in the conceptual model structure, such as routing store outflow, direct flow, and soil moisture, yielded better simulation results in the Ljubljana River catchment than the hybrid model approach using the liquid and solid form precipitation, snowpack, and melt as input data.

3.3 Comparing the performance of models and sensitivity analysis

To examine the performance of the models, initially, a raincloud plot was created for each model by converting the predicted hourly runoff to daily runoff and was presented in Fig. 13. The density and box plots for the simulated and observed flows showed that CemaNeige GR4H-WRT₁ and CemaNeige GR4H-WELM₁ were in better agreement with the observed values. The CemaNeige GR4H model has higher outliers than the observed flows, as shown in Fig. 13. The stand-alone WELM and WRT models performed poorly compared to the other models. Then, the level plots for daily runoff were prepared using the estimated hourly runoff to compare the minimum, maximum, and mean discharge values for each month and

Fig. 13 Raincloud plots for observed and simulated daily runoff by conceptual, machine learning, and hybrid models

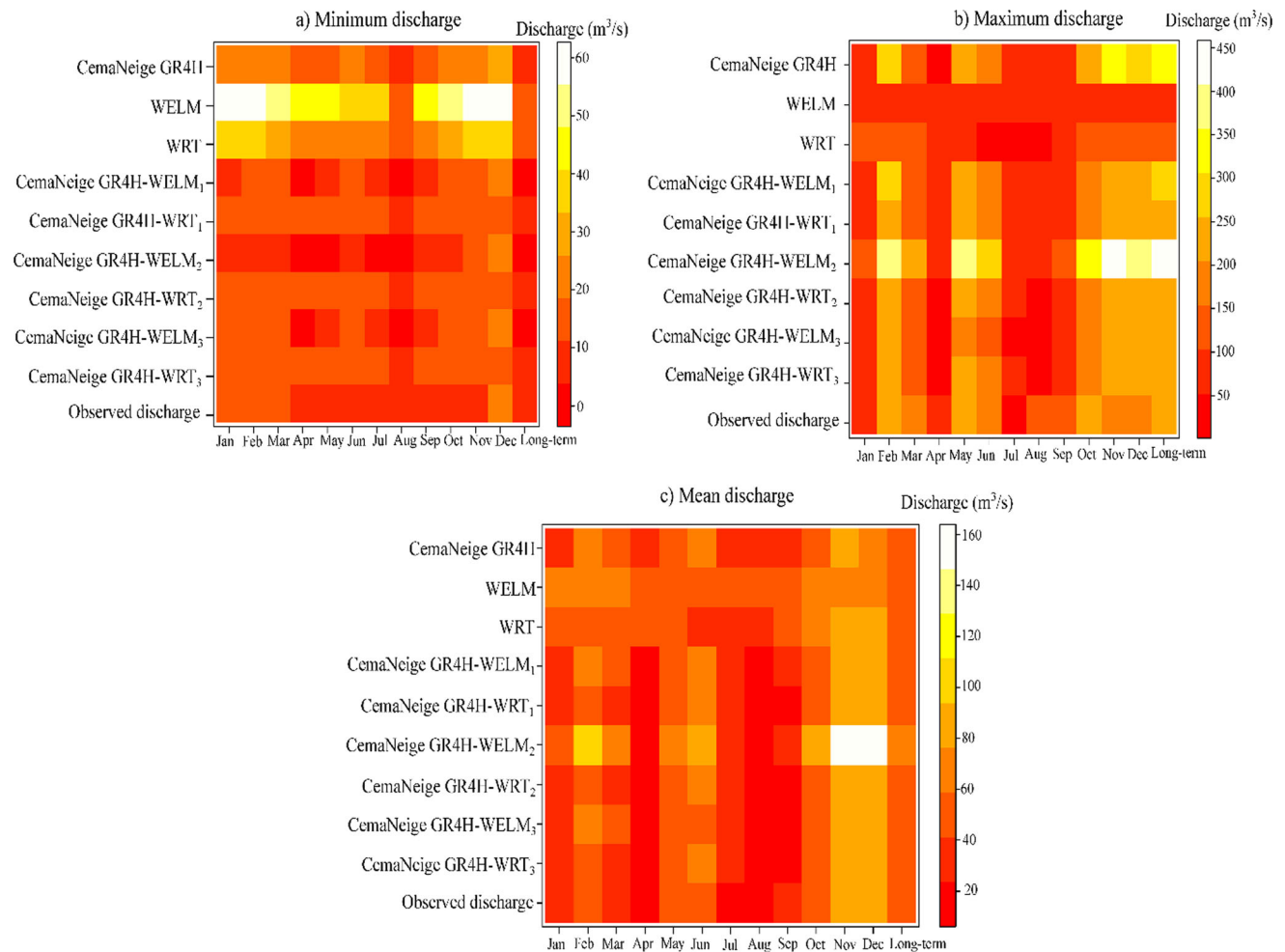
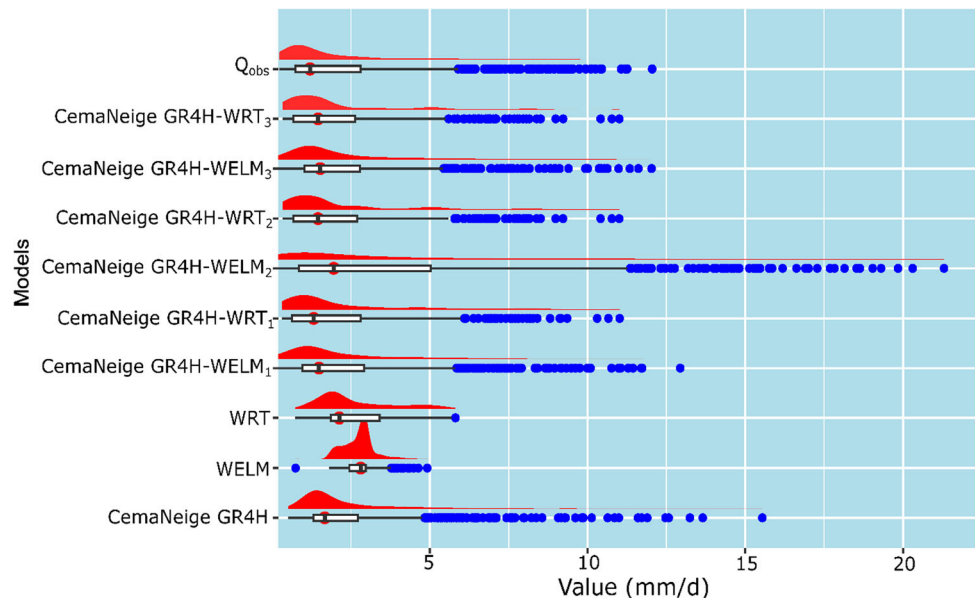


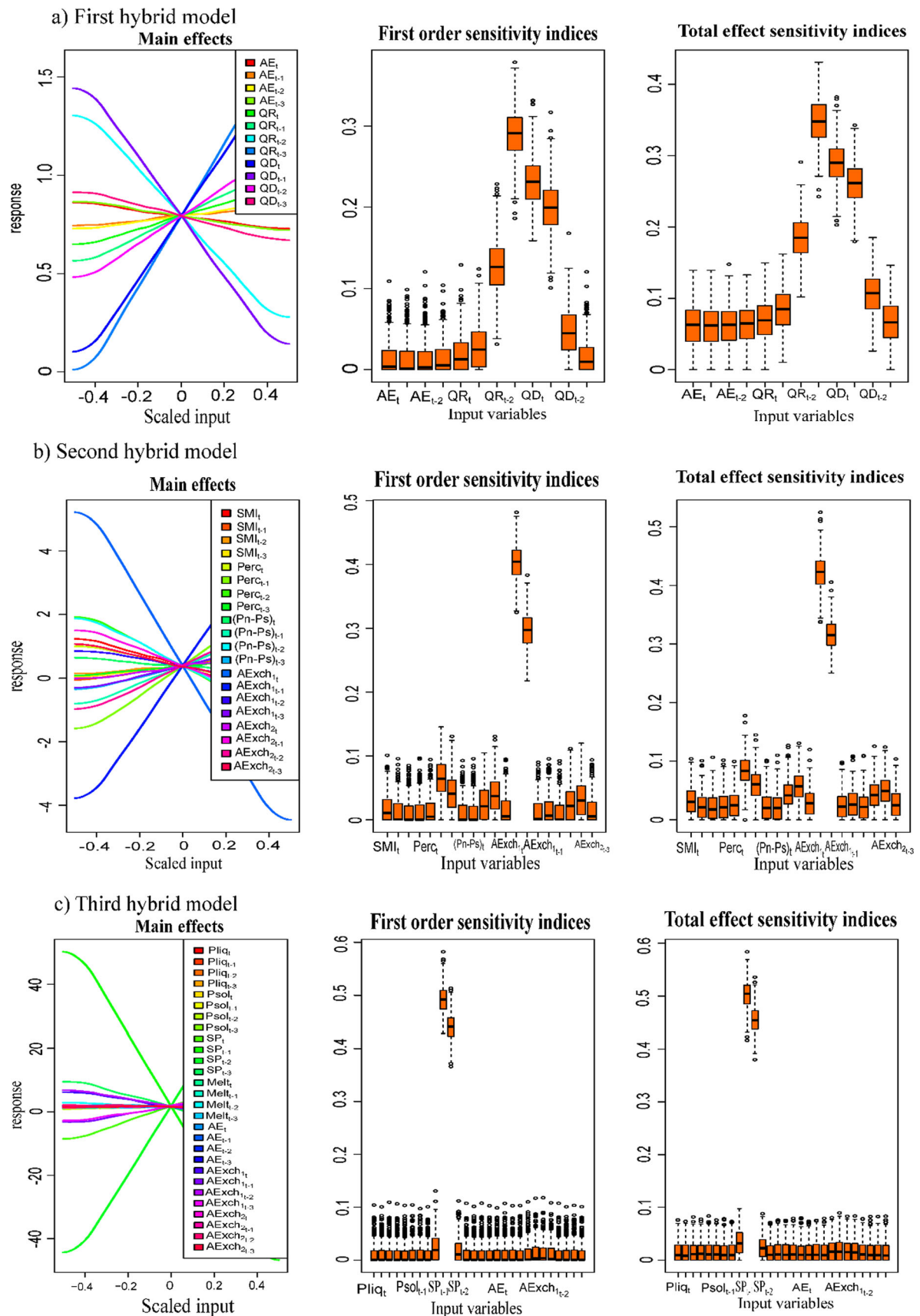
Fig. 14 Level plots of (a) minimum, (b) maximum, and (c) mean daily discharges for individual months and for long-term period (i.e., validation period)

were given in Fig. 14. The poor performance of the stand-alone machine learning models (i.e., WELM and WRT) is evident in predicting the minimum, mean, and maximum discharges as shown in Fig. 14. The conceptual model CemaNeige GR4H and hybrid models such as CemaNeige GR4H-WELM₁ and CemaNeige GR4H-WELM₃ performed well in predicting the minimum discharge. The CemaNeige GR4H-WRT₁ yielded better simulation results for maximum discharge than other models. In addition, the CemaNeige GR4H model performed poorly in predicting the maximum discharge compared to its performance in predicting the minimum discharge forecasting (Fig. 14a, b). In predicting mean discharge, the hybrid models performed well, with the exception of CemaNeige GR4H-WELM₂ (Fig. 14c). The CemaNeige GR4H model also performed well in predicting mean discharge, as shown in Fig. 14c. To observe the effects of input variables in the hybrid models for rainfall-runoff modelling, the Bayesian Monte Carlo sensitivity analysis was implemented and the findings were shown in Fig. 15. It was found that some input variables of the CemaNeige GR4H model had a greater impact on the rainfall-runoff modelling (Fig. 15). In this regard, it can be seen that QR_{t-2} , QR_{t-3} , QD_t , QD_{t-1} , and QD_{t-2} are the most efficient variables in the first hybrid model approach for rainfall-runoff modelling (Fig. 15a). In the second hybrid model approach, variables such as $Perc_{t-1}$, $Perc_{t-2}$, $(P_n - P_s)_{t-2}$, $(AExch_1)_t$, and $(AExch_1)_{t-1}$ influence the modelling of rainfall-runoff (Fig. 15b). Furthermore, SP_{t-1} and SP_{t-2} are the most influential variables for the runoff forecasting of the third hybrid model approach, as shown in Fig. 15c. Accordingly, it can be stated that the sensitivity of input variables changes for each hybrid model and affects the performance of the rainfall-runoff modelling.

4 Discussion

Implementing various variables related to the CemaNeige snow module and GR4H model structure as input data to the WELM and WRT machine learning models yielded more reliable results than the stand-alone conceptual and machine learning models for hourly rainfall-runoff modelling. Using actual evapotranspiration, routing store and direct flow components as input data in the first hybrid model approach yielded more accurate simulation results according to the evaluation criteria. The input variables such as soil moisture, percolation, actual exchange between catchments, and rainfall component in the second hybrid model approach also provided promising runoff forecasting results. Although the use of input variables such as liquid and solid precipitation, snowpack in the third hybrid model approach improved the performance of rainfall-runoff

modelling compared to the stand-alone models, it performed worse compared to the first and second hybrid model approaches based on the assessment criteria. These results indicated that implementing various outputs of CemaNeige GR4H as input data in the machine learning models helped improving the performance and provided different simulation results for each hybrid model approach. Similar conclusions regarding the implementation of the hybrid model approaches have been obtained in the previous studies. Accordingly, Humphrey et al. (2016) pointed out that using soil moisture as input data obtained from the GR4J in the ANN model enhanced the rainfall-runoff modelling performance. In addition, Kumanlioglu and Fistikoglu (2019) and Okkan et al. (2021) used different variables obtained from the conceptual models and revealed that the hybrid model approaches performed better than the stand-alone conceptual and machine learning models. In this study, the conceptual CemaNeige GR4H model yielded good performance, whereas the stand-alone WELM and WSVR models performed poorly in hourly runoff prediction. Using only the lags of precipitation and evapotranspiration as input data in the machine learning models did not yield reliable simulation results, especially for the WELM model. In addition, the WRT model outperformed the WELM model in hourly runoff prediction. Moreover, using WELM in the hybrid model resulted in better performance than the hybrid models containing the WRT model according to the evaluation criteria (Table 6). This indicates that used input variables have an important role on the performance of the machine learning models and implementing the outputs of the conceptual models as input data improved the prediction performance of the machine learning models. In this study, it was found that the implementation of WELM and WRT in the hybrid models could be useful to improve the performance of rainfall-runoff modelling, as also stated in previous studies (Yaseen et al. 2016; Nourani et al. 2019). As for extreme flows, it is obvious that hybrid models can be a good alternative for simulating minimum and maximum flows compared to stand-alone conceptual and machine learning models. The CemaNeige GR4H model performed better in simulating minimum flows than the maximum flows. The WELM and WRT models performed poorly in simulating extreme discharges compared to the other models. The sensitivity analysis showed that using different outputs of the conceptual model as input data in the machine learning models could significantly affect the runoff forecasting performance. In this regard, the effects and contributions of various variables obtained from the conceptual models on the rainfall-runoff modelling performance of different hybrid models will be further investigated in future studies. Furthermore, the implementation of different hybrid modelling approaches, such as nested hybrid models proposed



◀**Fig. 15** The sensitivity analysis of input variables for the (a) first hybrid model (i.e., CemaNeige GR4H-WELM₁ and CemaNeige GR4H-WRT₁), (b) second hybrid model (i.e., CemaNeige GR4H-WELM₂ and CemaNeige GR4H-WRT₂), (c) third hybrid model (i.e., CemaNeige GR4H-WELM₃ and CemaNeige GR4H-WRT₃)

by Okkan et al. (2021), will be investigated in further studies. As can be seen in this study, using different outputs of the conceptual models, applying various machine learning models, and different hybrid modelling approaches can be a guide to improve the performance of rainfall-runoff modelling.

Different factors, such as data availability, quality of dataset, uncertainty in the model structures, and used model types can affect the rainfall-runoff modelling performance remarkably. Westerberg and McMillan (2015) analysed the uncertainties in data and calculation methods on the hydrological signatures for rainfall-runoff process in the Brue catchment in UK and Mahurangi catchment in New Zealand. They revealed that the uncertainties could be large and highly variable between the hydrological signatures. Poulin et al. (2011) investigated the impacts of the model structure and parameter equifinality on the uncertainty for hydrological modelling by using the HYDROTEL physically based model and HSAMI conceptual model in Ceizur River basin, Canada. They pointed out that the uncertainty in hydrological model structure could be more important than parameter uncertainty. Her et al. (2019) examined the effects of the uncertainties related to the multiple general circulation models (GCMs) and multi-parameter ensembles on hydrological projections using the ABCD monthly water balance model in the Ohio River basin, USA. They detected that the impact of the multi-GCM ensemble uncertainty was prominent on rapid hydrological components such as direct runoff, whereas the uncertainty of model parameters was more efficient on slow components such as soil moisture and groundwater. Accordingly, it can be stated that the uncertainty of datasets, model structures, and parameters can significantly affect the rainfall-runoff modelling performance. In addition, using different machine learning model approaches could influence the modelling performance, remarkably. Han and Morrison (2022) implemented the LSTM deep learning model with National Water Model (NWM) in Russian River basin in California, USA. They revealed that the LSTM model approach contributed to the error correction of NWM model and improved the modelling performance of the NWM model. Rahimzad et al. (2021) compared the performance of linear regression (LR), multilayer perceptron (MLP), support vector machine

(SVM), and LSTM machine learning models for daily streamflow forecasting in the Kentucky River, USA. They stated the LSTM model yielded more reliable results than other conventional machine learning models. In this regard, using deep learning approaches can have a high potential to enhance rainfall-runoff modelling performance. As the limited part of this study, the uncertainty analysis, sensitivity analysis for the model outputs and unconventional machine learning models, such as deep learning approaches, are aimed to be implemented in further studies.

5 Conclusions

Rainfall-runoff modelling for a short-term period, e.g., at hourly and daily scales, is important for predicting extreme events. This task can be challenging, especially in the karst catchments, such as the mostly karst Ljubljana River catchment as well as changing climatic conditions. In this study, three different hybrid models based on the use of various outputs of the CemaNeige GR4H model as input data in WELM and WRT were implemented for hourly rainfall-runoff modelling in the Ljubljana River catchment. The performance of the hybrid models was compared with the stand-alone CemaNeige GR4H, WELM and WRT models. The findings of the present study can be summarised as follows:

- Using the CemaNeige snow module with GR4H improved the performance of the hourly rainfall-runoff modelling compared to the stand-alone GR4H model. However, the CemaNeige GR4H model overestimated low flows, especially during the dry period.
- The stand-alone WRT model outperformed the WELM model. However, both the WRT and WELM models performed poorly in runoff prediction. Implementing only meteorological variables as input data in the stand-alone machine learning models did not yield good performance in predicting extreme runoff.
- The first hybrid modelling approach (i.e., CemaNeige GR4H-WELM₁ and CemaNeige GR4H-WRT₁), which uses actual evapotranspiration, routing store outflow and direct flow obtained from the CemaNeige GR4H model as input data, performed the best among all hybrid models based on the evaluation criteria.
- The second hybrid modelling approach using the variables, such as soil moisture index and percolation as input data, improved the performance of the hourly rainfall-runoff modelling. Similarly, the third hybrid modelling approach using the variables, such as liquid precipitation, solid precipitation, and snowpack as input

data, also improved the modelling performance. However, the third modelling approach yielded the worst modelling performance among the hybrid models based on the evaluation criteria.

- The hybrid models generally performed well in predicting minimum and maximum flows for each month compared to stand-alone models. The CemaNeige GR4H model yielded better simulation results for minimum flows than for maximum flows, whereas the WELM and WRT models performed poorly in predicting minimum and maximum flows.
- Sensitivity analysis indicated that the *QR* and *QD* variables in the first hybrid modelling approach, *Perc*, *P_n-P_s* and *AExch_I* in the second hybrid modelling approach, and *SP* in the third hybrid modelling approach were influential variables for hourly runoff prediction.

As can be seen in this study, using different outputs obtained from the conceptual model as input data into the machine learning models can be useful to observe the improvement of modelling performance. Furthermore, the hybrid modelling approach may have the potential to improve the performance of rainfall-runoff modelling in karst catchments. Implementing a hybrid modelling approach can be useful for related stakeholders, researchers, and water management specialists regarding tackling the drawbacks of the stand-alone models and simulating the rainfall-runoff process in basins with nonhomogeneous characteristics and karst systems. In this regard, future studies will aim to use different conceptual and machine learning models in the hybrid modelling approach and implement the hybrid modelling approaches in the catchments with various climate and catchment characteristics to observe the transferability of related models. In addition, analysis of the uncertainties in input data, model structures and parameters and sensitivity analysis for the model outputs will be tackled in further studies.

Acknowledgements The authors would like to acknowledge the support of the Slovenian Research and Innovation Agency (ARIS) through grant P2-0180. The research carried out is also supported by the Slovenian national committee of the IHP UNESCO research programme (UNESCO IHP C3330-20-456010) and UNESCO Chair on Water-related Disaster Risk Reduction. The authors also wish to thank the Slovenian Environment Agency (ARSO) for making data publicly available.

Author contributions Conceptualization and data curation M.Š. and C.S.; Formal analysis C.S.; Funding acquisition and supervision M.Š.; Investigation, methodology and validation C.S. and M.Š.; Resources and Software C.S.; Validation C.S. and M.Š.; Visualization C.S.; Writing—original draft C.S.; Writing—review & editing M.Š.

Funding The authors would like to acknowledge the support of the Slovenian Research and Innovation Agency (ARIS) through grant P2-0180. The research carried out is also supported by the Slovenian

national committee of the IHP UNESCO research programme (UNESCO IHP C3330-20-456010) and UNESCO Chair on Water-related Disaster Risk Reduction.

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

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abdulkareem JH, Pradhan B, Sulaiman WNA, Jamil NR (2018) Review of studies on hydrological modelling in Malaysia. *Model Earth Syst Environ* 4:1577–1605. <https://doi.org/10.1007/S40808-018-0509-Y/TABLES/2>
- Al SMB, Kawanisi K, Jilati MN et al (2021) Extent of detection of hidden relationships among different hydrological variables during floods using data-driven models. *Environ Monit Assess* 193:1–14. <https://doi.org/10.1007/S10661-021-09499-9/FIGURES/5>
- Al-Safi HIJ, Sarukkalghe PR (2020) The application of conceptual modelling to assess the impacts of future climate change on the hydrological response of the Harvey River catchment. *J Hydro-Environ Res* 28:22–33. <https://doi.org/10.1016/J.JHER.2018.01.006>
- Ancitil F, Michel C, Perrin C, Andréassian V (2004) A soil moisture index as an auxiliary ANN input for stream flow forecasting. *J Hydrol (amst)* 286:155–167. <https://doi.org/10.1016/J.JHYDROL.2003.09.006>
- Atiquzzaman M, Kandasamy J (2016) Prediction of hydrological time-series using extreme learning machine. *J Hydroinf* 18:345–353. <https://doi.org/10.2166/HYDRO.2015.020>
- Ba H, Guo S, Wang Y et al (2018) Improving ANN model performance in runoff forecasting by adding soil moisture input and using data preprocessing techniques. *Hydrol Res* 49:744–760. <https://doi.org/10.2166/NH.2017.048>
- Bajirao TS, Kumar P, Kumar M et al (2021) Potential of hybrid wavelet-coupled data-driven-based algorithms for daily runoff prediction in complex river basins. *Theor Appl Climatol* 145:1207–1231. <https://doi.org/10.1007/S00704-021-03681-2/FIGURES/15>
- Berrar D, Dubitzky W (2013) Decision Tree. In: Dubitzky W, Wolkenhauer O, Cho KH, Yokota H (eds) *Encyclopedia of systems biology*. Springer, New York, pp 551–555
- Cantoni E, Trambly Y, Grimaldi S et al (2022) Hydrological performance of the ERA5 reanalysis for flood modeling in Tunisia with the LISFLOOD and GR4J models. *J Hydrol Reg Stud* 42:101169. <https://doi.org/10.1016/J.EJRH.2022.101169>

- Casper M, Gemmar P, Gronz O et al (2007) Fuzzy logic-based rainfall—runoff modelling using soil moisture measurements to represent system state. *Hydrol Sci J* 52:478–490. <https://doi.org/10.1623/HYSJ.52.3.478>
- Coron L, Thirel G, Delaigue O et al (2017) The suite of lumped GR hydrological models in an R package. *Environ Model Softw* 94:166–171. <https://doi.org/10.1016/J.ENVSOF.2017.05.002>
- Coron L, Delaigue O, Thirel G, et al (2022) airGR: Suite of GR hydrological models for precipitation-runoff modelling. <https://CRAN.R-project.org/package=airGR>. Accessed 20 Dec 2022
- Cotman M, Drolc A, Končan JZ (2008) Assessment of pollution loads from point and diffuse sources in Small River Basin: case study Ljubljana River. *Environ Forensics* 9:246–251. <https://doi.org/10.1080/15275920802122965>
- da Wu J, Kuo JM (2009) An automotive generator fault diagnosis system using discrete wavelet transform and artificial neural network. *Expert Syst Appl* 36:9776–9783. <https://doi.org/10.1016/J.ESWA.2009.02.027>
- Darbandsari P, Coulibaly P (2020) Inter-comparison of lumped hydrological models in data-scarce watersheds using different precipitation forcing data sets: case study of Northern Ontario. *Canada J Hydrol Reg Stud* 31:100730. <https://doi.org/10.1016/J.EJRH.2020.100730>
- Daubechies I (1990) The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans Inf Theory* 36:961–1005. <https://doi.org/10.1109/18.57199>
- Dobra A (2016) *Decision trees*. Springer, New York
- Farfán JF, Cea L (2021) Coupling artificial neural networks with the artificial bee colony algorithm for global calibration of hydrological models. *Neural Comput Appl* 33:8479–8494. <https://doi.org/10.1007/S00521-020-05601-3/TABLES/10>
- Fattahi P, Ashrafzadeh A, Pirmoradian N, Vazifedoust M (2022) Integrating IHACRES with a data-driven model to investigate the possibility of improving monthly flow estimates. *Water Supply* 22:360–371. <https://doi.org/10.2166/WS.2021.267>
- Feng BF, Xu YS, Zhang T, Zhang X (2022) Hydrological time series prediction by extreme learning machine and sparrow search algorithm. *Water Supply* 22:3143–3157. <https://doi.org/10.2166/WS.2021.419>
- Flores N, Rodríguez R, Yépez S et al (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. *Water (basel)* 13:3191. <https://doi.org/10.3390/W13223191>
- Gharraei-Manesh S, Fathzadeh A, Taghizadeh-Mehrjardi R (2016) Comparison of artificial neural network and decision tree models in estimating spatial distribution of snow depth in a semi-arid region of Iran. *Cold Reg Sci Technol* 122:26–35. <https://doi.org/10.1016/J.COLDREGIONS.2015.11.004>
- Graf R, Zhu S, Sivakumar B (2019) Forecasting river water temperature time series using a wavelet–neural network hybrid modelling approach. *J Hydrol (amst)* 578:124115. <https://doi.org/10.1016/j.jhydrol.2019.124115>
- Gramacy RB (2007) tgp: An R package for bayesian nonstationary, semiparametric nonlinear regression and design by treed gaussian process models. *J Stat Softw* 19:1–46
- Gramacy RB, Taddy M (2010) Categorical Inputs, Sensitivity Analysis, Optimization and Importance Tempering with tgp Version 2, an R Package for Treed Gaussian Process Models. *J Stat Softw* 33:1–48
- Gramacy RB, Taddy M (2022) Categorical inputs, sensitivity analysis, optimization and importance tempering with tgp version 2, an R package for treed Gaussian process models. 1–49
- Gupta HV, Kling H, Yilmaz KK, Martinez GF (2009) Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J Hydrol (amst)* 377:80–91. <https://doi.org/10.1016/J.JHYDROL.2009.08.003>
- Han H, Morrison RR (2022) Improved runoff forecasting performance through error predictions using a deep-learning approach. *J Hydrol (amst)* 608:127653. <https://doi.org/10.1016/j.jhydrol.2022.127653>
- Hao Y, Sun F, Wang H et al (2022) Understanding climate-induced changes of snow hydrological processes in the Kaidu River Basin through the CemaNeige-GR6J model. *Catena (amst)* 212:106082. <https://doi.org/10.1016/J.CATENA.2022.106082>
- Hartmann A, Goldscheider N, Wagener T et al (2014) Karst water resources in a changing world: review of hydrological modeling approaches. *Rev Geophys* 52:218–242. <https://doi.org/10.1002/2013RG000443>
- Her Y, Yoo S-H, Cho J et al (2019) Uncertainty in hydrological analysis of climate change: multi-parameter vs. multi-GCM ensemble predictions. *Sci Rep* 9:4974. <https://doi.org/10.1038/s41598-019-41334-7>
- Hu X, Shi L, Lin G, Lin L (2021) Comparison of physical-based, data-driven and hybrid modeling approaches for evapotranspiration estimation. *J Hydrol (amst)* 601:126592. <https://doi.org/10.1016/J.JHYDROL.2021.126592>
- Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. *Neurocomputing* 70:489–501. <https://doi.org/10.1016/J.NEUCOM.2005.12.126>
- Humphrey GB, Gibbs MS, Dandy GC, Maier HR (2016) A hybrid approach to monthly streamflow forecasting: Integrating hydrological model outputs into a Bayesian artificial neural network. *J Hydrol (amst)* 540:623–640. <https://doi.org/10.1016/J.JHYDROL.2016.06.026>
- Huo J, Liu L, Zhang Y (2018) An improved multi-cores parallel artificial Bee colony optimization algorithm for parameters calibration of hydrological model. *Futur Gener Comput Syst* 81:492–504. <https://doi.org/10.1016/J.FUTURE.2017.07.020>
- Husic A, Al-Aamery N, Fox JF (2022) Simulating hydrologic pathway contributions in fluvial and karst settings: an evaluation of conceptual, physically-based, and deep learning modeling approaches. *J Hydrol X* 17:100134. <https://doi.org/10.1016/j.hydroa.2022.100134>
- Jiang Z, Rashid MM, Johnson F, Sharma A (2021) A wavelet-based tool to modulate variance in predictors: an application to predicting drought anomalies. *Environ Model Softw* 135:104907. <https://doi.org/10.1016/J.ENVSOF.2020.104907>
- Karaboga D (2005) An idea based on honey bee swarm for numerical optimization. *Kayseri, Türkiye*
- Klemeš V (1986) Operational testing of hydrological simulation models. *Hydrol Sci J* 31:13–24. <https://doi.org/10.1080/02626668609491024>
- Kodja DJ, Akongnongbé AJS, Amoussou E et al (2020) Calibration of the hydrological model GR4J from potential evapotranspiration estimates by the Penman-Monteith and Oudin methods in the Ouémé watershed (West Africa). *Proc IAHS* 383:163–169. <https://doi.org/10.5194/PIAHS-383-163-2020>
- Kotsiantis SB (2013) Decision trees: a recent overview. *Artif Intell Rev* 39:261–283. <https://doi.org/10.1007/s10462-011-9272-4>
- Kourgialas NN, Dokou Z, Karatzas GP (2015) Statistical analysis and ANN modeling for predicting hydrological extremes under climate change scenarios: The example of a small Mediterranean agro-watershed. *J Environ Manage* 154:86–101. <https://doi.org/10.1016/J.JENVMAN.2015.02.034>
- Kumanlioglu AA, Fistikoglu O (2019) Performance enhancement of a conceptual hydrological model by integrating artificial intelligence. *J Hydrol Eng* 24:04019047. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001850](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001850)
- Kumari N, Srivastava A, Sahoo B et al (2021) Identification of suitable hydrological models for streamflow assessment in the

- Kangsabati River Basin, India, by using different model selection scores. *Nat Resour Res* 30:4187–4205. <https://doi.org/10.1007/S11053-021-09919-0/FIGURES/7>
- Kursa MB, Rudnicki WR (2010) Feature selection with the Boruta package. *J Stat Softw* 36:1–13
- Kursa MB, Jankowski A, Rudnicki WR (2010) Boruta – A system for feature selection. *Fundam Inform* 101:271–285. <https://doi.org/10.3233/FI-2010-288>
- Lavtar K, Bezak N, Šraj M (2020) Rainfall-runoff modeling of the nested non-homogeneous Sava river sub-catchments in Slovenia. *Water* 12:128. <https://doi.org/10.3390/W12010128>
- Le MN (2008) Le bassin versant de surface vu par le souterrain: une voie d'amélioration des performances et du réalisme des modèles pluie-débit?
- Lees T, Buechel M, Anderson B et al (2021) Benchmarking data-driven rainfall-runoff models in Great Britain: A comparison of long short-term memory (LSTM)-based models with four lumped conceptual models. *Hydrol Earth Syst Sci* 25:5517–5534. <https://doi.org/10.5194/HESS-25-5517-2021>
- Mallat SG (1989) A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans Pattern Anal Mach Intell* 11:674–693. <https://doi.org/10.1109/34.192463>
- Maslova I, Ticiavilca AM, McKee M (2016) Adjusting wavelet-based multiresolution analysis boundary conditions for long-term streamflow forecasting. *Hydrol Process* 30:57–74. <https://doi.org/10.1002/HYP.10564>
- Mathevet T, Gupta H, Perrin C et al (2020) Assessing the performance and robustness of two conceptual rainfall-runoff models on a worldwide sample of watersheds. *J Hydrol (amst)* 585:124698. <https://doi.org/10.1016/J.JHYDROL.2020.124698>
- Mathevet T (2005) Quels modèles pluie-débit globaux au pas de temps horaire? Développements empiriques et comparaison de modèles sur un large échantillon de bassins versants
- Mayaud C, Kogovšek B, Gabrovšek F et al (2023) Deciphering the water balance of poljes: example of Planinsko Polje (Slovenia). *Acta Carsologica*. <https://doi.org/10.3986/ac.v51i2.11029>
- Moosavi V, Gheisoori Fard Z, Vafakhah M (2022) Which one is more important in daily runoff forecasting using data driven models: Input data, model type, preprocessing or data length? *J Hydrol (amst)* 606:127429. <https://doi.org/10.1016/J.JHYDROL.2022.127429>
- Moriassi DN, Gitau MW, Pai N, Daggupati P (2015) Hydrologic and water quality models: performance measures and evaluation criteria. *Trans ASABE* 58:1763–1785
- Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I — a discussion of principles. *J Hydrol (amst)* 10:282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Newman AJ, Clark MP, Sampson K et al (2015) Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrol Earth Syst Sci* 19:209–223. <https://doi.org/10.5194/hess-19-209-2015>
- Nourani V, Hosseini Baghanam A, Adamowski J, Kisi O (2014) Applications of hybrid wavelet–Artificial Intelligence models in hydrology: a review. *J Hydrol (amst)* 514:358–377. <https://doi.org/10.1016/J.JHYDROL.2014.03.057>
- Nourani V, Tajbakhsh AD, Molajou A (2019) Data mining based on wavelet and decision tree for rainfall-runoff simulation. *Hydrol Res* 50:75–84. <https://doi.org/10.2166/NH.2018.049>
- Okkan U, Ersoy ZB, 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. *J Hydrol (amst)* 598:126433. <https://doi.org/10.1016/J.JHYDROL.2021.126433>
- Osuch M, Wawrzyniak T, Nawrot A (2019) Diagnosis of the hydrology of a small Arctic permafrost catchment using HBV conceptual rainfall-runoff model. *Hydrol Res* 50:459–478. <https://doi.org/10.2166/NH.2019.031>
- Oudin L, Hervieu F, Michel C et al (2005) Which potential evapotranspiration input for a lumped rainfall-runoff model?: Part 2—towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling. *J Hydrol (amst)* 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 (amst)* 279:275–289. [https://doi.org/10.1016/S0022-1694\(03\)00225-7](https://doi.org/10.1016/S0022-1694(03)00225-7)
- Poncelet C, Merz R, Merz B et al (2017) Process-based interpretation of conceptual hydrological model performance using a multinational catchment set. *Water Resour Res* 53:7247–7268. <https://doi.org/10.1002/2016WR019991>
- Poulin A, Brissette F, Leconte R et al (2011) Uncertainty of hydrological modelling in climate change impact studies in a Canadian, snow-dominated river basin. *J Hydrol (amst)* 409:626–636. <https://doi.org/10.1016/j.jhydrol.2011.08.057>
- Prasad R, Deo RC, Li Y, Maraseni T (2019) Weekly soil moisture forecasting with multivariate sequential, ensemble empirical mode decomposition and Boruta-random forest hybridizer algorithm approach. *Catena (amst)* 177:149–166. <https://doi.org/10.1016/J.CATENA.2019.02.012>
- Rahimzad M, Moghaddam Nia A, Zolfonoon H et al (2021) Performance comparison of an LSTM-based deep learning model versus conventional machine learning algorithms for streamflow forecasting. *Water Resour Manage* 35:4167–4187. <https://doi.org/10.1007/s11269-021-02937-w>
- Reddy MJ, Kumar DN (2020) Evolutionary algorithms, swarm intelligence methods, and their applications in water resources engineering: a state-of-the-art review. *H2Open J* 3:135–188
- Rezaie-Balf M, Kim S, Ghaemi A, Deo R (2021) Design and performance of two decomposition paradigms in forecasting daily solar radiation with evolutionary polynomial regression: wavelet transform versus ensemble empirical mode decomposition. *Predict Modell Energy Manag Power Syst Eng*. <https://doi.org/10.1016/B978-0-12-817772-3.00005-7>
- Roushangar K, Alizadeh F, Nourani V (2018) Improving capability of conceptual modeling of watershed rainfall-runoff using hybrid wavelet-extreme learning machine approach. *J Hydroinf* 20:69–87. <https://doi.org/10.2166/HYDRO.2017.011>
- Rusjan S, Sapač K, Petrič M et al (2019) Identifying the hydrological behavior of a complex karst system using stable isotopes. *J Hydrol (amst)* 577:123956. <https://doi.org/10.1016/J.JHYDROL.2019.123956>
- Saltelli A (2002) Making best use of model evaluations to compute sensitivity indices. *Comput Phys Commun* 145:280–297. [https://doi.org/10.1016/S0010-4655\(02\)00280-1](https://doi.org/10.1016/S0010-4655(02)00280-1)
- Sang YF (2012) A practical guide to discrete wavelet decomposition of hydrologic time series. *Water Resour Manage* 26:3345–3365. <https://doi.org/10.1007/S11269-012-0075-4/FIGURES/9>
- Sapač K, Rusjan S, Šraj M (2020) Assessment of consistency of low-flow indices of a hydrogeologically non-homogeneous catchment: a case study of the Ljubljanka river catchment. *Slovenia J Hydrol (amst)* 583:124621. <https://doi.org/10.1016/J.JHYDROL.2020.124621>
- Seo Y, Kim S (2016) River stage forecasting using wavelet packet decomposition and data-driven models. *Procedia Eng* 154:1225–1230. <https://doi.org/10.1016/J.PROENG.2016.07.439>
- Sezen C, Partal T (2022a) New hybrid GR6J-wavelet-based genetic algorithm-artificial neural network (GR6J-WGANN) conceptual-data-driven model approaches for daily rainfall-runoff

- modelling. *Neural Comput Appl* 34:17231–17255. <https://doi.org/10.1007/S00521-022-07372-5/TABLES/3>
- Sezen C, Partal T (2022b) Two integrated conceptual–wavelet-based data-driven model approaches for daily rainfall–runoff modelling. *J Hydroinf* 24:949–975. <https://doi.org/10.2166/HYDRO.2022.171>
- Sezen C, Bezak N, Šraj M (2018) Hydrological modelling of the karst Ljubljana river catchment using lumped conceptual model. *Acta Hydrotech* 31:87–100
- Sezen C, Bezak N, Bai Y, Šraj M (2019) Hydrological modelling of karst catchment using lumped conceptual and data mining models. *J Hydrol (amst)* 576:98–110. <https://doi.org/10.1016/J.JHYDROL.2019.06.036>
- Sikorska-Senoner AE, Quilty JM (2021) A novel ensemble-based conceptual-data-driven approach for improved streamflow simulations. *Environ Model Softw* 143:105094. <https://doi.org/10.1016/J.ENVSOF.2021.105094>
- Solomatine DP, Wagener T (2011) Hydrological modeling. treatise on water. Science 2:435–457. <https://doi.org/10.1016/B978-0-444-53199-5.00044-0>
- Speiser JL, Miller ME, Tooze J, Ip E (2019) A comparison of random forest variable selection methods for classification prediction modeling. *Expert Syst Appl* 134:93–101. <https://doi.org/10.1016/J.ESWA.2019.05.028>
- Sun Y, Niu J, Sivakumar B (2019) A comparative study of models for short-term streamflow forecasting with emphasis on wavelet-based approach. *Stoch Env Res Risk Assess* 33:1875–1891. <https://doi.org/10.1007/S00477-019-01734-7/TABLES/7>
- Tayfur G, Zucco G, Brocca L, Moramarco T (2014) Coupling soil moisture and precipitation observations for predicting hourly runoff at small catchment scale. *J Hydrol (amst)* 510:363–371. <https://doi.org/10.1016/J.JHYDROL.2013.12.045>
- Tayyab M, Zhou J, Dong X et al (2019) Rainfall-runoff modeling at Jinsha River basin by integrated neural network with discrete wavelet transform. *Meteorol Atmos Phys* 131:115–125. <https://doi.org/10.1007/S00703-017-0546-5/TABLES/6>
- Unnikrishnan P, Jothiprakash V (2018) Data-driven multi-time-step ahead daily rainfall forecasting using singular spectrum analysis-based data pre-processing. *J Hydroinf* 20:645–667. <https://doi.org/10.2166/HYDRO.2017.029>
- Valéry A, Andréassian V, Perrin C (2014a) ‘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 (amst)* 517:1176–1187. <https://doi.org/10.1016/J.JHYDROL.2014.04.058>
- Valéry A, Andréassian V, Perrin C (2014b) ‘As simple as possible but not simpler’: what is useful in a temperature-based snow-accounting routine? Part 1 – comparison of six snow accounting routines on 380 catchments. *J Hydrol (amst)* 517:1166–1175. <https://doi.org/10.1016/J.JHYDROL.2014.04.059>
- van Esse WR, Perrin C, Booij MJ et al (2013) The influence of conceptual model structure on model performance: a comparative study for 237 French catchments. *Hydrol Earth Syst Sci* 17:4227–4239. <https://doi.org/10.5194/hess-17-4227-2013>
- Wang J, Lu S, Wang SH, Zhang YD (2021) A review on extreme learning machine. *Multimed Tools Appl* 81:41611–41660. <https://doi.org/10.1007/S11042-021-11007-7/TABLES/8>
- Westerberg IK, McMillan HK (2015) Uncertainty in hydrological signatures. *Hydrol Earth Syst Sci* 19:3951–3968. <https://doi.org/10.5194/hess-19-3951-2015>
- Willmott CJ (1981) On the validation of models. *Phys Geogr* 2:184–194. <https://doi.org/10.1080/02723646.1981.10642213>
- Yaseen ZM, Jaafar O, Deo RC et al (2016) Stream-flow forecasting using extreme learning machines: a case study in a semi-arid region in Iraq. *J Hydrol (amst)* 542:603–614. <https://doi.org/10.1016/j.jhydrol.2016.09.035>
- Zeynoddin M, Bonakdari H, Azari A et al (2018) Novel hybrid linear stochastic with non-linear extreme learning machine methods for forecasting monthly rainfall a tropical climate. *J Environ Manage* 222:190–206. <https://doi.org/10.1016/J.JENVMAN.2018.05.072>
- Zhou Q, Chen L, Singh VP et al (2019) Rainfall-runoff simulation in karst dominated areas based on a coupled conceptual hydrological model. *J Hydrol (amst)* 573:524–533. <https://doi.org/10.1016/j.jhydrol.2019.03.099>
- Zhu S, Heddam S, Wu S et al (2019) Extreme learning machine-based prediction of daily water temperature for rivers. *Environ Earth Sci* 78:202. <https://doi.org/10.1007/s12665-019-8202-7>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.