Journal of
Hydroinformatics

Journal of Hydroinformatics Vol 26 No 7, 1648 doi: 10.2166/hydro.2024.010

Internal structure modification of a simple monthly water balance model via incorporation of a machine learning-based nonlinear routing

Umut Okkan[a,*], Zeynep Beril Ersoy[a,b] and Okan Fistikoglu[b]

[a] Department of Civil Engineering, Hydraulic Division, Balikesir University, Balikesir, Turkey
[b] Department of Civil Engineering, Hydraulic Division, Dokuz Eylul University, İzmir, Turkey
*Corresponding author. E-mail: umutokkan@balikesir.edu.tr

UO, 0000-0003-1284-3825; ZBE, 0000-0001-8362-5767; OF, 0000-0002-9483-1563

ABSTRACT

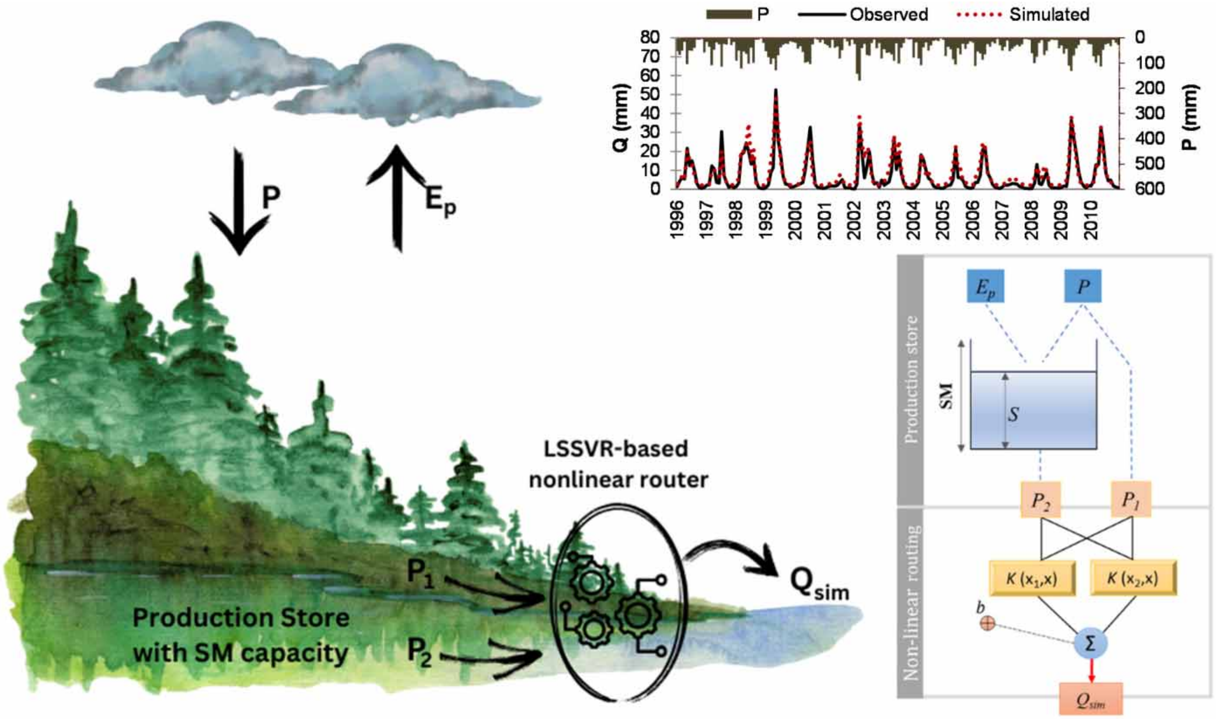
Among various monthly water balance models, one of the models that has the simplest structure and offers a well-behaved conceptual platform is the GR2M. Despite the widespread use of the model with two-free parameters, the fact that it tends to produce relatively large errors in peak flow months necessitates some modifications to the model. The reason for the mentioned simulation deficiencies could be that the relationship between the routing reservoir and the external environment of the basin is controlled by a single parameter, making the storage–discharge relationship linear. Therefore, in this study, least squares support vector regression, one of the nonlinear data-driven models, has replaced the routing part of the GR2M to enhance the monthly runoff simulation. The performance of the three-parameter hybrid model (GR3M), which was developed by considering the parameter parsimony point of view and including a machine learning (ML)-based nonlinear routing scheme, was examined in some locations in the Gediz River Basin in western Turkey. Statistical performance measures have shown that GR3M, which both leverages the capabilities of an ML model and blends conceptual outputs within a nested scheme, clearly outperforms the original GR2M. The proposed modification has brought significant improvements, especially to high-flow simulations.

Key words: data-driven models, Gediz River Basin, GR2M, hybrid water balance model, monthly water balance models, nonlinear routing

HIGHLIGHTS

- The routing component of the GR2M model has been replaced with least squares support vector regression (LSSVR).
- The model GR3M, which is entirely hybrid, has improved the monthly runoff simulation.
- Incorporating LSSVR into GR2M has made high-flow simulation more robust.
- GR3M also preserves the conceptual structure of GR2M with a nested hybridization.

GRAPHICAL ABSTRACT



1. INTRODUCTION

Monthly water balance models (MWBMs) are still of interest thanks to their limited demand for input data and the practical usability of their outputs in long-term hydrological drought assessments and reservoir operation studies. Despite some MWBMs capturing basin dynamics with more intensive parameter sets, most of them tend to oversimplify the hydrological process by conceptualizing watershed with several interconnected storages and generally have a number of parameters ranging from two to five (Bai *et al.* 2015). Notwithstanding a great deal of expertise pertaining to many models, there remains a need to identify and rectify their shortcomings (Xu & Singh 1998). For example, the majority of these conceptual-type models can suffer from linearity assumptions in generating runoff (Vidyarthi & Jain 2022). Moreover, the residuals of such models can be subject to seasonality and have limited capabilities, particularly in high-flow simulations (Fathi *et al.* 2023). Although there have been attempts to modify the evapotranspiration process to enable water balance models to produce more accurate runoff simulations (e.g., Bai *et al.* 2015), recent studies have shown that the main focus should be on nonlinearity in the mechanisms of the runoff components (Cheng *et al.* 2020; Fathi *et al.* 2023). However, there are a scant number of papers on the development of the runoff partitioning structure in MWBMs. For example, Cheng *et al.* (2020) claimed that the linear storage–discharge relationship on short timescales has a solid physical basis due to the negligibility of recharging from the current rainfall event. When analyzing MWBMs, however, recharge to groundwater storage from the current month's precipitation and therefore the baseflow generation process comes into consideration (Hrachowitz *et al.* 2014). To this end, Cheng *et al.* (2020) demonstrated that by arranging the baseflow components of several MWBMs in nonlinear exponential form, the storage–discharge dynamics could be better described, enhancing the monthly runoff simulations. In another study, Fathi *et al.* (2023) have integrated a seasonal component regulated by an exponent parameter into the simple GR2M model by allowing its routing coefficient to change from month to month as opposed to maintaining a constant value throughout the year. Their modified model has been verified to outperform the GR2M over 1,469 catchments in the USA.

Additionally, in recent years, studies on the integration of conceptual models with machine learning (ML) techniques have begun to demonstrate promise in improving simulation accuracy while preserving hydrological interpretability (Kapoor *et al.* 2023). Some of these models are of the coupled type and rely on employing multiple conceptual outputs derived from conceptual models as auxiliary predictors to strengthen ML-based training (e.g., Chen & Adams 2006; Humphrey *et al.* 2016;

Wang *et al.* 2021; Sezen & Partal 2022; Vidyarthi & Jain 2022). According to McIntyre (2013), performance enhancements are often the result of theoretical nonlinearity in the routing process, which merits greater consideration in conceptual modeling to contribute to process understanding. In the study presented by Vidyarthi & Jain (2022), Australian water balance model (AWBM) outputs were coupled with artificial neural networks (ANNs), intrinsically adopting the notion ascribed by McIntyre (2013). They employed AWBM to obtain runoff hydrographs at individual sub-catchment levels in upstream reaches of the Kentucky River Basin, USA, and subsequently trained ANNs in place of linear Muskingum to perform runoff routing at the basin outlet.

On the contrary, in a rather limited number of studies, the internal structures of conceptual models have been replaced with ML approaches (e.g., Kumanlioglu & Fistikoglu 2019; Okkan *et al.* 2021; Bhasme *et al.* 2022; Durgut & Ayvaz 2023; Kapoor *et al.* 2023). Kapoor *et al.* (2023), for instance, substituted modern deep learning models for the processes in the routing storage of the GR4J model. Given that the parameters of the considered part of the conceptual model and those of the ML model are calibrated jointly in this type of hybridization, accounting for potential parameter interactions results in a more robust simulation. Despite studies motivating the incorporation of conceptual and ML models into hybrid modeling frameworks in daily runoff simulation (e.g., Kumanlioglu & Fistikoglu 2019; Durgut & Ayvaz 2023), only a few studies have embedded ML tools into existing MWBMs to upgrade their runoff-generating mechanisms and enhance monthly simulations (Okkan *et al.* 2021; Bhasme *et al.* 2022). However, while adopting this synergistic manner, the parameter density of each model used might complicate the calibration process of MWBMs. From this standpoint, it would be more consistent not to deviate from the perspective of parameter parsimony and to search for hybridization facilities that can be easily adapted to operational studies. Therefore, our study's objectives are:

i. to replace the routing part of a simple MWBM, which considers the underlying physical process under the assumption of linearity, with a robust ML model;
ii. to propose a hybrid framework that preserves soil moisture storage within the physics-guided conceptual model structure and to question to what extent the complex relationships between intermediate outputs provided by the conceptual layer of the hybrid model and target runoff can be extracted with ML; and
iii. to evaluate an alternative parametric model that accounts for the nonlinear storage–discharge relationship against the efficacy of ML while managing the routing process in an MWBM, and to evaluate the performances of all models across various conditions.

For the objectives stated above, we examined potential modifications to the GR2M model, which is without a doubt one of the most parsimonious MWBMs. Due to its success in simulating monthly runoff and its simple structure, this two-parameter model is regarded as a benchmark for many basin-based studies conducted on different continents (Fathi *et al.* 2023). We replaced the linearly behaving routing storage of GR2M with least squares support vector regression (LSSVR), which can avoid several drawbacks of ANNs, and proposed a novel hybrid model with a total of three free parameters, maintaining the soil moisture (production) storage of the original model. The hybrid model (hereafter denoted as GR3M) is in line with the parsimonious structure of the modified GR2M suggested by Fathi *et al.* (2023) and deals with capturing the nonlinear rainfall–runoff relationship in a different way. The virtues of GR3M have been tested at the flow gauging stations across the Gediz River Basin (GRB), one of the most important agricultural basins in Turkey. To the best of our knowledge, there has been no prior investigation that has modified the internal structure of GR2M through employing any ML technique, with the exception of a scant number of papers (e.g., Okkan *et al.* 2021; Bhasme *et al.* 2022), which have integrated ML models into MWBMs with more intricate parameterization. Those studies have substituted a chosen ML model for malfunctioning or insufficiently functioning storages in the conceptual model, as implemented in this study. Nevertheless, the topologies within these models do not explicitly utilize nonlinear routing; rather, they treat much more conceptual outputs as refined inputs. Moreover, controlling the conceptual layer of the novel hybrid model with a singular parameter, while controlling the nonlinear routing layer utilizing LSSVR with two parameters could facilitate the calibration process for parsimonious model users. The novel aspects of the developed hybrid modeling framework are its ability to keep the original model's simplicity and successfully embed LSSVR into the conceptual structure as a nonlinear routing tool, thus having the potential to overcome the reported shortcomings of the GR2M model. This article is organized into five sections: the subsequent two sections detail the methodology adopted and the studied catchment and data, respectively. The results are then presented in Section 4, including discussions. Concluding remarks are given in the final section.

2. METHODOLOGY

2.1. Hybrid water balance model (GR3M) as an alternative to the GR2M

This section begins with an explanation of the two-parameter GR2M model and then describes how to configure the hybrid GR3M. The first reservoir used by both GR2M and GR3M contains a soil moisture reservoir with a maximum capacity of SM (the parameter in mm), while the GR2M has merely a routing reservoir with a capacity of 60 mm (see Figure 1). A nondimensional, positive parameter, RC, governs the connection between this routing store and the groundwater or external environment of the basin (Mouelhi *et al.* 2006; Fathi *et al.* 2023). First, a portion of the total precipitation (P) is entered into the production store. Thus, the initial soil moisture storage S rises to the S_1 level, and the remaining is considered excess precipitation (P_1). Then, the soil moisture storage decreases from S_1 to S_2 due to the impact of potential evapotranspiration (E_p). After the soil moisture storage is updated, a portion of water (P_2) is released from the production store (see w1 to w4 denoted in Figure 1). The operation described so far is common to both models. As for GR2M's next calculations, the level of the routing reservoir fed by the sum of P_1 and P_2 increases from R to R_1. Thereafter, R_1 is multiplied by RC to get the intermediate value of the routing store (R_2). Finally, the simulated runoff (Q_sim) is obtained as a function of R_2, and the updated value of R is reached for use in the next month (see w5 to w9 denoted in Figure 1).

The hybrid GR3M model is sequentially composed of a production store and an LSSVR-based nonlinear routing component, as depicted in Figure 1. While applying GR3M, the original production storage of the GR2M model, driven by P and E_p inputs and initial soil moisture storage, is preserved. Afterward, the excess precipitation and the released water from the production store are transmitted to the nonlinear routing part. Therein, the LSSVR, which has two parameters to be tuned, substitutes for the routing reservoir guided by the parameter RC and converts the production store outputs (P_1 and P_2) to the simulated monthly runoff at the catchment outlet. LSSVR has been favored among a pool of data-driven models when incorporating ML into an existing water balance model because its simplified algorithm requires less training

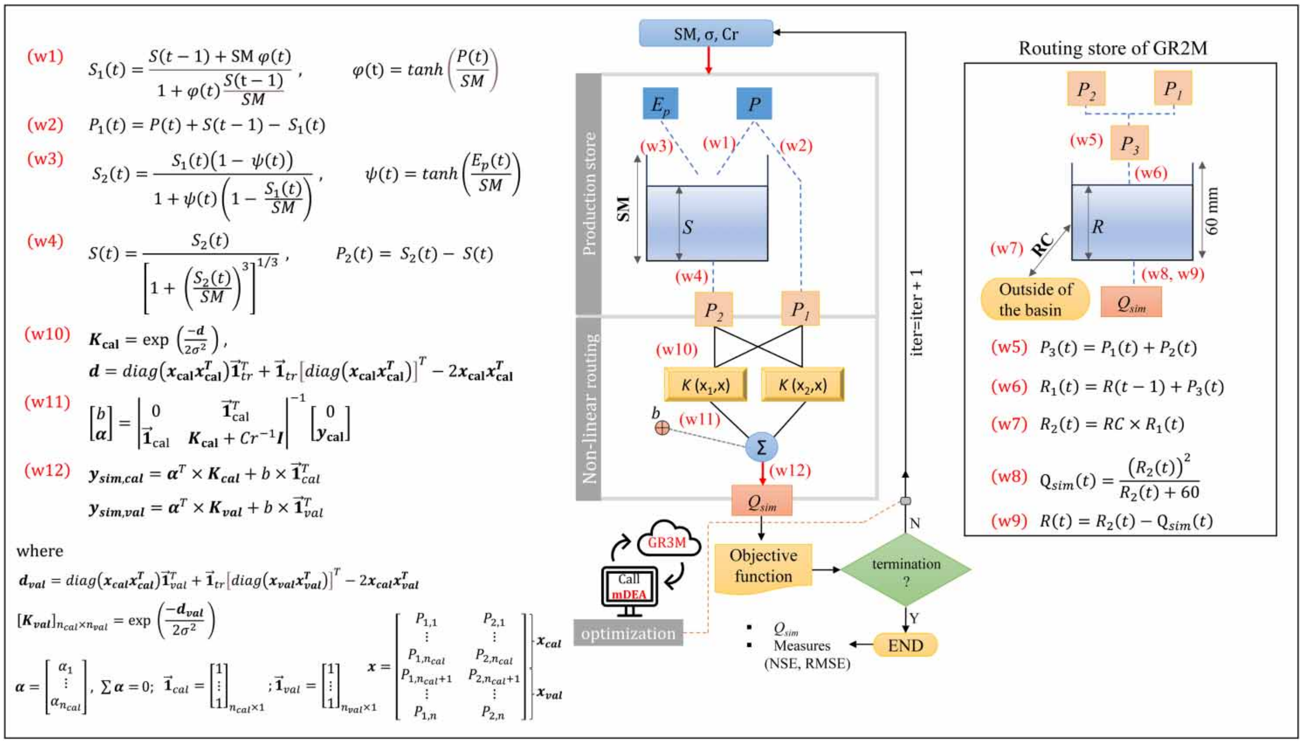

(w1) $S_1(t) = \frac{S(t-1) + \mathrm{SM}\,\varphi(t)}{1 + \varphi(t)\frac{S(t-1)}{SM}}$, $\varphi(t) = tanh\left(\frac{P(t)}{SM}\right)$
(w2) $P_1(t) = P(t) + S(t-1) - S_1(t)$
(w3) $S_2(t) = \frac{S_1(t)(1-\psi(t))}{1+\psi(t)\left(1-\frac{S_1(t)}{SM}\right)}$, $\psi(t) = tanh\left(\frac{E_p(t)}{SM}\right)$
(w4) $S(t) = \frac{S_2(t)}{\left[1+\left(\frac{S_2(t)}{SM}\right)^3\right]^{1/3}}$, $P_2(t) = S_2(t) - S(t)$
(w10) $\boldsymbol{K}_\mathbf{cal} = \exp\left(\frac{-\boldsymbol{d}}{2\sigma^2}\right)$,
$\boldsymbol{d} = diag(\boldsymbol{x}_\mathbf{cal}\boldsymbol{x}_\mathbf{cal}^{\boldsymbol{T}})\vec{\mathbf{1}}_{tr}^T + \vec{\mathbf{1}}_{tr}[diag(\boldsymbol{x}_\mathbf{cal}\boldsymbol{x}_\mathbf{cal}^{\boldsymbol{T}})]^T - 2\boldsymbol{x}_\mathbf{cal}\boldsymbol{x}_\mathbf{cal}^{\boldsymbol{T}}$
(w11) $\begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{vmatrix} 0 & \vec{\mathbf{1}}_\mathrm{cal}^T \\ \vec{\mathbf{1}}_\mathrm{cal} & \boldsymbol{K}_\mathbf{cal} + Cr^{-1}\boldsymbol{I} \end{vmatrix}^{-1} \begin{bmatrix} 0 \\ \boldsymbol{y}_\mathbf{cal} \end{bmatrix}$
(w12) $\boldsymbol{y}_{\boldsymbol{sim,cal}} = \boldsymbol{\alpha}^T \times \boldsymbol{K}_{\boldsymbol{cal}} + b \times \vec{\mathbf{1}}_{cal}^T$
$\boldsymbol{y}_{\boldsymbol{sim,val}} = \boldsymbol{\alpha}^T \times \boldsymbol{K}_{\boldsymbol{val}} + b \times \vec{\mathbf{1}}_{val}^T$
where
$\boldsymbol{d}_{\boldsymbol{val}} = diag(\boldsymbol{x}_{\boldsymbol{cal}}\boldsymbol{x}_{\boldsymbol{cal}}^{\boldsymbol{T}})\vec{\mathbf{1}}_{val}^T + \vec{\mathbf{1}}_{tr}[diag(\boldsymbol{x}_{\boldsymbol{val}}\boldsymbol{x}_{\boldsymbol{val}}^{\boldsymbol{T}})]^T - 2\boldsymbol{x}_{\boldsymbol{cal}}\boldsymbol{x}_{\boldsymbol{val}}^{\boldsymbol{T}}$
$[\boldsymbol{K}_{\boldsymbol{val}}]_{n_{cal}\times n_{val}} = \exp\left(\frac{-\boldsymbol{d}_{\boldsymbol{val}}}{2\sigma^2}\right)$
$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_{n_{cal}} \end{bmatrix}, \sum\boldsymbol{\alpha} = 0;\ \vec{\mathbf{1}}_{cal} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{n_{cal}\times 1};\vec{\mathbf{1}}_{val} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{n_{val}\times 1}$
$\boldsymbol{x} = \begin{bmatrix} P_{1,1} & P_{2,1} \\ \vdots & \vdots \\ P_{1,n_{cal}} & P_{2,n_{cal}} \\ P_{1,n_{cal}+1} & P_{2,n_{cal}+1} \\ \vdots & \vdots \\ P_{1,n} & P_{2,n} \end{bmatrix}$ $\boldsymbol{x}_{\boldsymbol{cal}}$ $\boldsymbol{x}_{\boldsymbol{val}}$
SM, σ, Cr
Production store
E_p
P
(w3)
(w1)
(w2)
SM
S
(w4)
P_2
P_1
Non-linear routing
(w10)
K (x₁,x)
K (x₂,x)
b
(w11)
Σ
(w12)
Q_{sim}
iter=iter + 1
N
Objective function
termination ?
Y
END
GR3M
Call mDEA
optimization
▪ Q_{sim}
▪ Measures (NSE, RMSE)
Routing store of GR2M
P_2
P_1
(w5)
P_3
(w6)
R
60 mm
(w7)
RC
Outside of the basin
(w8, w9)
Q_{sim}
(w5) $P_3(t) = P_1(t) + P_2(t)$
(w6) $R_1(t) = R(t-1) + P_3(t)$
(w7) $R_2(t) = RC \times R_1(t)$
(w8) $\mathrm{Q}_{sim}(t) = \frac{(R_2(t))^2}{R_2(t) + 60}$
(w9) $R(t) = R_2(t) - \mathrm{Q}_{sim}(t)$

Figure 1 | Structure of the hybrid GR3M in which the LSSVR substitutes for the routing reservoir. In the rightmost part of the figure, the routing storage of the original model and its governing equations are indicated. S denotes soil moisture equivalent storage, while R demonstrates routing storage. Bold variables in equations are those in the matrix form. x is the matrix containing all scaled values for P_1 and P_2, while parts of it allocated to calibration and validation are indicated as x_cal and x_val, respectively. n_cal and n_val are the data length used during the calibration and validation periods, respectively. I is the n_cal-by-n_cal identity matrix with ones on the diagonal and zeros elsewhere. The normalized runoff observations used in the calibration step are indicated as y_cal.

effort and offers a computational benefit over standard support vector regression. Moreover, it incorporates structural risk minimization as opposed to empirical risk minimization, which is utilized in many ML models such as ANNs, and exhibits strength in its efforts to enhance rainfall–runoff relationships (Katipoglu & Sarigol 2023; Xu *et al.* 2023).

In Figure 1, the equations for how to perform nonlinear routing with LSSVR are also specified. It should be noted that the production store outputs and observed runoff data are scaled to lie within the range [0, 1]. The radial basis kernel function (RBF) is employed to map nonlinearly the normalized outputs conveyed by the production reservoir into a higher-dimensional feature space (for the kernel matrix obtained for the calibration period, see w10 in Figure 1). It is anticipated that RBF can effectively address the nonlinear nature of the routing process, where optimal estimation of the kernel width parameter (σ) would play a crucial role. When the regularization parameter (C_r) is also taken into account, the Lagrangian multipliers (α) and the bias term (b) are estimated (see w11 in Figure 1). After the kernel matrix for the validation period (K_{val}), which may not be in square form, is constituted, the simulated outputs generated throughout the periods of calibration and validation are compiled (see w12 in Figure 1). Finally, all these values are converted back into runoff estimations that use the same unit as the original target.

2.2. Another nonlinear routing-themed model

Additionally, we sought to compare GR3M hybridized with ML against another nonlinear routing-themed model. In this concept, the two-storage GR2M was reconfigured as a single-storage by adopting the nonlinear storage–discharge relationship suggested by Peters & Aulenbach (2011) (hereafter referred to as PA). Cheng *et al.* (2020) also incorporated the PA into MWBMs with different structures. Their research indicates that MWBMs with nonlinear storage–discharge relationships are superior at capturing the dynamics of both the monthly baseflow component and the overall streamflow compared with typical MWBMs. In our study, the PA approach has been included into the GR2M by keeping the parametrization of production storage and the equations for estimating S, P_1, and P_2:

$$Q_{\text{sim}}(t) = P_1(t) + e^{(P_2(t)-b)/m} \tag{1}$$

In this alternative model, referred to as GR3M_PA, the constants m and b need to be calibrated together with SM so as to simulate total runoff. The nonlinear variability between water recharged from the production store and routed runoff is controlled by these new parameters.

2.3. Automatic calibration of the models

The hybridized model is linked to an optimization tool so that both a single conceptual parameter governing the production store (i.e., SM) and the hyperparameters (i.e., σ and C_r) governing the LSSVR-based nonlinear routing part are adjusted jointly. In recent years, metaheuristics have become increasingly prevalent in the automatic calibration of various types of hydrological models (Kumanlioglu & Fistikoglu 2019; Okkan & Kirdemir 2020; Durgut & Ayvaz 2023). The optimization algorithm employed is a differential evolution algorithm with a modified mutation scheme (hereinafter referred to as mDEA), which is a stochastic algorithm that maintains a population of np individuals. When calibrating models consisting of D numbers of parameters, mDEA, like conventional DEA proposed by Storn & Price (1997), utilizes three sequential stages: mutation, crossover, and selection. Before executing these operators, random parameter solutions for the given np are initially created:

$$par_{i,j} = \text{LB}_j + U(0,\ 1) \times (\text{UB}_j - \text{LB}_j), \qquad j = 1,\ 2,\ \ldots,\ D \tag{2}$$

where one individual in the population is represented by vector par_i ($i = 1,2,\ldots,np$); j is the index of any parameter; $U(0,1)$ is a randomly generated number from the uniform distribution with the range of [0, 1]; and LB_j and UB_j are the lower and upper limits of the jth parameter, respectively.

Unlike standard DEA, the mutation strategy followed when making random perturbations to parameter vectors is as follows (Gong & Cai 2013; Okkan *et al.* 2023):

$$Mpar_i(g) = par_i(g) + \text{SF}_1 \times (bestpar(g) - par_i(g)) + \text{SF}_2 \times (par_{r1}(g) - par_{r2}(g)) \tag{3}$$

where g is the generation index (i.e., iteration number); $Mpar$ is the mutant vector; and par_i and $bestpar$ denote the parameter solution of any individual and the global best solution in the population, respectively.

While the mutation process could gain stochastic search ability depending on two individuals randomly selected from the population ($r1, r2 \in \{1, 2,..., np\}$), the extent of mutation is regulated by the SF_1 and SF_2 scaling factors.

Following the mutation step, the crossover operator is utilized to produce candidate vectors as follows:

$$Cpar_{i,j}(g) = \begin{cases} Mpar_{i,j}(g), & \text{if } U(0,\ 1) < \text{CR} \text{ or } j = \text{random } (1, D) \\ par_{i,j}(g), & \text{otherwise} \end{cases} \quad (4)$$

where CR is the crossover constant that usually ranges from 0.5 to 1.0.

Finally, the selection operator determines the best-performing solution for the subsequent generation by comparing a trial vector with the parent solution in the current population:

$$par_i(g+1) = \begin{cases} Cpar_i(g), & \text{if the cost function improved} \\ par_i(g), & \text{otherwise} \end{cases} \quad (5)$$

It should also be noted that the steps of the mDEA, apart from the mutation operator, were executed as suggested by Storn & Price (1997).

2.4. Model evaluation criteria

In this study, we utilized the root mean squared error (RMSE) as the objective function during mDEA runs and evaluated the models based on the Nash–Sutcliffe efficiency (NSE) index. As noted by Moriasi *et al.* (2007) and Fathi *et al.* (2023), the model performances can be graded as very good ($0.75 \leq \text{NSE} < 1.00$), good ($0.65 \leq \text{NSE} < 0.75$), acceptable ($0.50 \leq \text{NSE} < 0.65$), and unsatisfactory ($\text{NSE} < 0.50$). Additionally, considering that models subjected to nonlinear routing modifications would be compared against the reference model GR2M, a percentage improvement in the overall performance of the model was revealed by PI index, following Senbeta *et al.* (1999). PI denotes the percentage of the initial variance that GR2M fails to account for; the modified model then attempts to account for this variance by incorporating nonlinear routing into the available model. The below equation formulates this criterion:

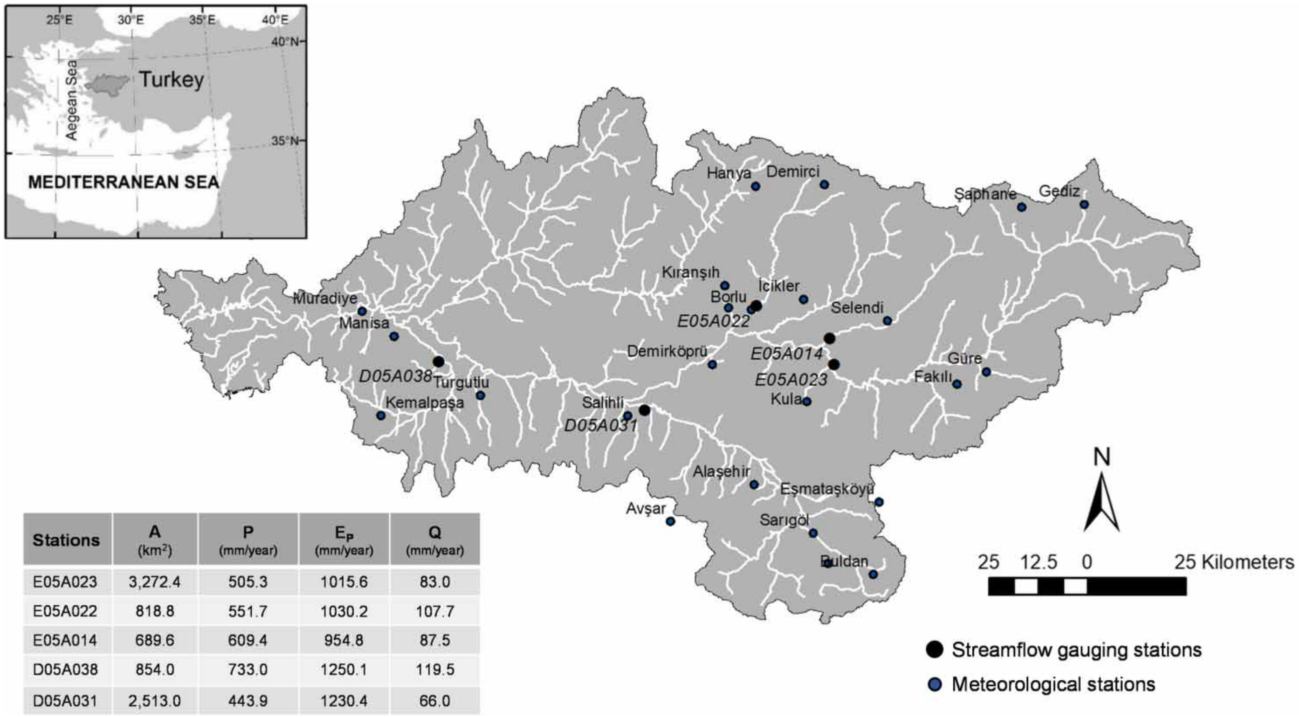
$$\text{PI} = \frac{\text{NSE}_\text{M} - \text{NSE}_\text{R}}{1 - \text{NSE}_\text{R}} \times 100 \quad (6)$$

where NSE_M and NSE_R are the NSE values obtained from the modified model and reference model GR2M, respectively.

Furthermore, FITEVAL software developed by Ritter & Muñoz-Carpena (2013) was implemented to evaluate the goodness-of-fit of the calibrated models since it can yield the NSE-based relative frequency distribution (RFD) produced by bootstrapping and the corresponding statistical significance. In order to facilitate comparisons and quantify discrepancies between the observed and simulated runoff, a Taylor diagram (Taylor 2001) is also utilized. This offers a brief statistical overview of the consistency between the simulated data and the observations, as measured by the root mean square difference (RMSD) and correlation coefficients. Section 4 elaborates on the application and discussion of the hybrid GR3M model.

3. STUDY AREA AND DATA

The GRB, which mostly satisfies the agricultural water demands of the western region of Turkey, was selected as the study region since robust water balance modeling is necessary for allocating water to the irrigation fields in the GRB (Okkan *et al.* 2021; Durgut & Ayvaz 2023). Due to the presence of several weirs and irrigated-operated dam reservoirs on the main river of the GRB, modeling investigations were conducted on a limited number of lateral tributaries where sufficient hydrometeorological data were available. The selected river tributaries are Murat, Demirci, Selendi, Alasehir, and Nif, on which flow gauging stations with the codes E05A23, E05A22, E05A14, D05A31, and D05A38 are operated, respectively (Figure 2). The natural streamflow data of these flow gauging stations covering a consecutive 31-year period from 1980 to 2010 have been compiled, and the related meteorological stations representing the catchments are given in Figure 2. While the average temperature is 17 °C over the Nif and Alasehir catchments, which are located at the near-coast and southern parts of the GRB, respectively, it is around 13 °C in the inner parts. The mean annual precipitation varies from



Stations	A (km²)	P (mm/year)	E_P (mm/year)	Q (mm/year)
E05A023	3,272.4	505.3	1015.6	83.0
E05A022	818.8	551.7	1030.2	107.7
E05A014	689.6	609.4	954.8	87.5
D05A038	854.0	733.0	1250.1	119.5
D05A031	2,513.0	443.9	1230.4	66.0

Figure 2 | Locations of the stations utilized and the corresponding hydrometeorological characteristics.

444 to 733 mm, and the runoff-to-precipitation ratio is between 0.14 and 0.20 (Figure 2). Considering E_p estimations derived from the Penman–Monteith equation and aridity indexes calculated by UNEP's formula (UNEP 1992), the climate regime over the catchments is dry sub-humid.

4. RESULTS AND DISCUSSION

In all models used, the production store was initially assumed to be at 10% capacity, while the initial value for the routing reservoir of the GR2M is fixed at 10 mm. To avoid any bias associated with the initial storage values, Fathi *et al.* (2023) have suggested implementing a 12-month warm-up period. The data set is therefore separated into three subsets: (i) the warm-up period from October 1979 to September 1980; (ii) the calibration period from October 1980 to September 1995; and (iii) the validation period from October 1995 to September 2010. As alluded to earlier, the control parameters of LSSVR, which supersedes the simple routing storage of GR2M, and the production store parameter SM were embedded into the GR3M scheme. Then, the metaheuristics optimizer, mDEA, assisted in the joint calibration of these parameters. Given the unknown physical limits of the LSSVR parameters responsible for controlling the nonlinear routing component of GR3M, it should be noted that the optimization process of these parameters has been initiated by generating random values within a specified range (i.e., [0.1, 100]). Subsequently, these values have not been constrained within a particular bound throughout the iterations. Additionally, we would like to highlight that recent research indicates that k-fold cross validation could potentially mitigate overfitting concerns and reduce uncertainties associated with input data set partitioning (Vu *et al.* 2022). The training set is subdivided into k-folds, and the validation procedure to be iterated k times, according to this concept. The mean value of fitness values throughout the k loops is the performance metric to be improved (Bazrkar & Chu 2022). In the study, we observed that by assigning the fold number k, population size of mDEA and maximum number of iterations to 5, 30, and 50, respectively, there could be sufficient convergence in the objective function without being affected by overfitting issue for GR3M. The other mDEA control settings were chosen in pursuit of several experiments. While both scaling factors were fixed at 0.5, CR was taken as 0.7. Under these settings, mDEA was executed 30 times during parameter optimization for each model. An iterative simulation–optimization procedure was implemented to calibrate randomly assigned parameter solutions in each independent run. Thereafter, the final parameter estimations were stored for any model, and the best solution from the solution pool could be extracted.

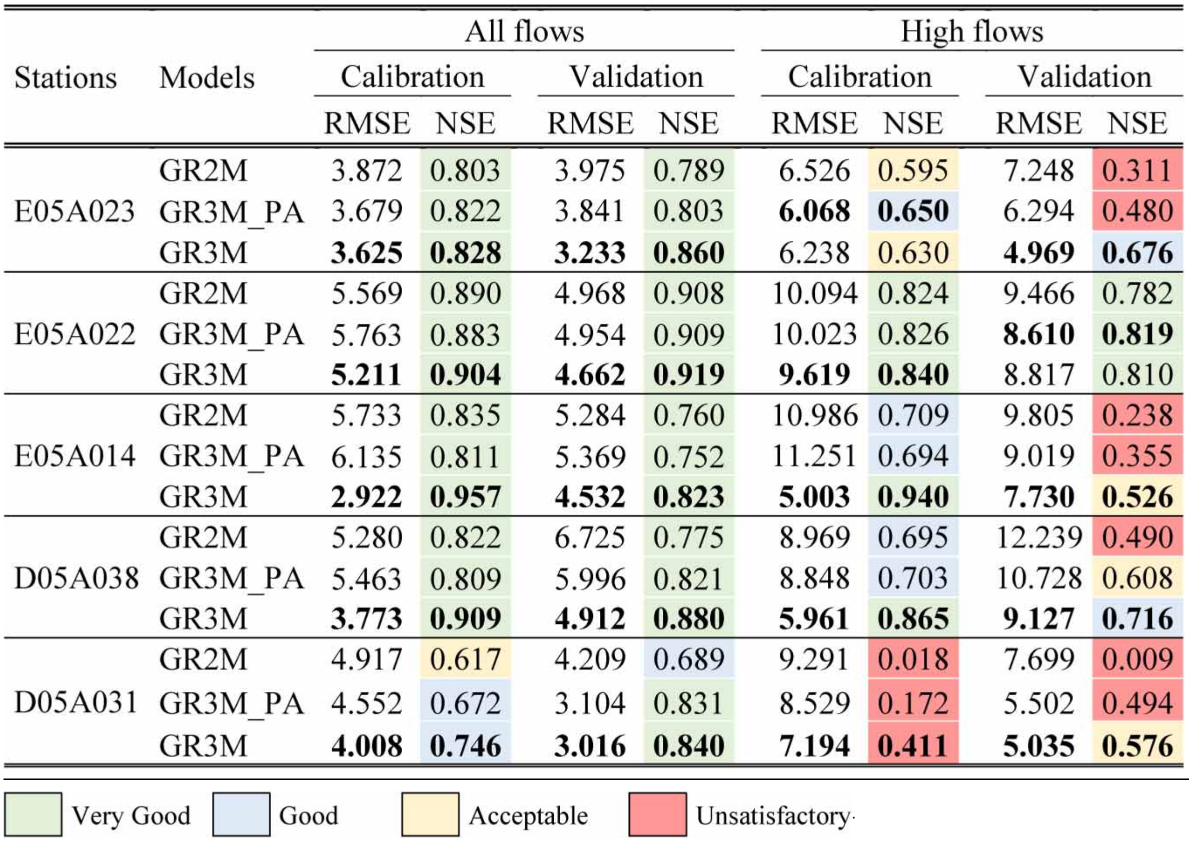
At the end of the multiple runs, the best parameter estimations are given in Table 1. The range in which the solution of the conceptual parameters is sought is stated as a footnote in Table 1. This table displays that there are also substantial discrepancies between the calibrated SM in the original model and those in GR3M. The fact that RC is a much more sensitive parameter in the GR2M may have resulted in underestimated SM parameters during the calibration phase. The performance evaluation of all models during calibration and validation is summarized in Table 2. Figure 3 displays the percentage improvement in the overall performance of the reference model using NSE values from this table. When an evaluation was made for all data points (all flows) allocated for calibration and validation, it was determined that GR2M performed very well at all stations, except for station D05A031. Nevertheless, it is obvious that GR3M enhances the calibration and validation

Table 1 | Calibrated parameters for GR2M, GR3M_PA, and hybrid GR3M

	GR2M		**GR3M_PA**			**GR3M**		
Stations	**SM**	**RC**	**SM**	***b***	***m***	**SM**	***σ***	**C_r**
E05A023	258.98	0.78	425.33	11.08	10.67	498.05	11.43	140.80
E05A022	216.08	0.73	406.62	14.52	5.97	409.29	1.14	148.55
E05A014	231.99	0.59	599.75	63.82	2.68	415.86	0.26	2,399.61
D05A038	645.61	0.66	1,140.15	21.69	16.15	861.42	0.50	1,694.24
D05A031	85.91	0.62	394.35	114.19	7.71	452.90	2.60	45.42

The SM parameter is in mm. The possible ranges of SM and RC were assigned as 10–1,000 mm and 0.01–1.0, respectively.

Table 2 | Performance assessment of models for the calibration and validation periods on each flow gauging station based on all flows and high flows



Stations	Models	All flows				High flows			
		Calibration		Validation		Calibration		Validation	
		RMSE	NSE	RMSE	NSE	RMSE	NSE	RMSE	NSE
E05A023	GR2M	3.872	🟩 0.803	3.975	🟩 0.789	6.526	🟨 0.595	7.248	🟥 0.311
	GR3M_PA	3.679	🟩 0.822	3.841	🟩 0.803	**6.068**	🟦 **0.650**	6.294	🟥 0.480
	GR3M	**3.625**	🟩 **0.828**	**3.233**	🟩 **0.860**	6.238	🟨 0.630	**4.969**	🟦 **0.676**
E05A022	GR2M	5.569	🟩 0.890	4.968	🟩 0.908	10.094	🟩 0.824	9.466	🟩 0.782
	GR3M_PA	5.763	🟩 0.883	4.954	🟩 0.909	10.023	🟩 0.826	**8.610**	🟩 **0.819**
	GR3M	**5.211**	🟩 **0.904**	**4.662**	🟩 **0.919**	**9.619**	🟩 **0.840**	8.817	🟩 0.810
E05A014	GR2M	5.733	🟩 0.835	5.284	🟩 0.760	10.986	🟦 0.709	9.805	🟥 0.238
	GR3M_PA	6.135	🟩 0.811	5.369	🟩 0.752	11.251	🟦 0.694	9.019	🟥 0.355
	GR3M	**2.922**	🟩 **0.957**	**4.532**	🟩 **0.823**	**5.003**	🟩 **0.940**	**7.730**	🟨 **0.526**
D05A038	GR2M	5.280	🟩 0.822	6.725	🟩 0.775	8.969	🟦 0.695	12.239	🟥 0.490
	GR3M_PA	5.463	🟩 0.809	5.996	🟩 0.821	8.848	🟦 0.703	10.728	🟨 0.608
	GR3M	**3.773**	🟩 **0.909**	**4.912**	🟩 **0.880**	**5.961**	🟩 **0.865**	**9.127**	🟦 **0.716**
D05A031	GR2M	4.917	🟨 0.617	4.209	🟦 0.689	9.291	🟥 0.018	7.699	🟥 0.009
	GR3M_PA	4.552	🟦 0.672	3.104	🟩 0.831	8.529	🟥 0.172	5.502	🟥 0.494
	GR3M	**4.008**	🟦 **0.746**	**3.016**	🟩 **0.840**	**7.194**	🟥 **0.411**	**5.035**	🟨 **0.576**

🟩 Very Good 🟦 Good 🟨 Acceptable 🟥 Unsatisfactory.

The RMSE values are in mm. The optimal results with regard to the performance indices assessed are marked in bold fonts.

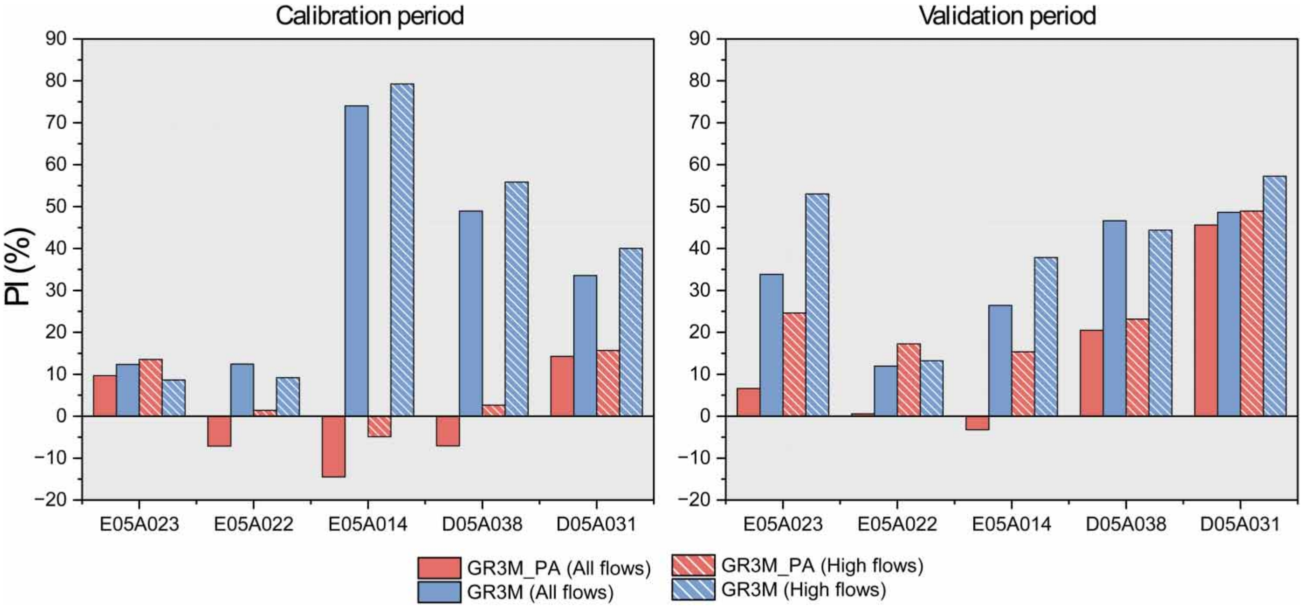


Figure 3 | Changes in the overall model performance following adjustments made to the reference model. The results extracted from calibration and validation data are shown in the left and right panels, respectively.

performances of all stations. As Senbeta *et al.* (1999) have argued that the modification to a model leads to substantial improvement when the PI exceeds 10%, Figure 3 confirms that the framework introduced by GR3M has a statistically significant influence on the monthly runoff simulation. The inclusion of free parameters (i.e., *b* and *m*), which were anticipated to introduce nonlinearity into the discharge–storage relationship, subsequent to the production store when composing GR3M_PA, had a negligible or adverse effect on the performance of all flows during the calibration period. While the performance of GR3M_PA was found to be significant for all flows during the validation period at stations D05A038 and D05A031, it is evident that it does not yield as consistent results across all stations and both periods as the hybrid GR3M.

Upon evaluating all data points, one could argue that the original GR2M is already sufficient and that the usage of GR3M is not essential, even if it improves overall performance. Nevertheless, it has been detected that GR2M's functionalities in simulating high flows are restricted, as Fathi *et al.* (2023) stressed. This determination was made based on the indices calculated between the points above the third quartile of the observed series and those in the corresponding simulations (Table 2). It is clear from Table 2 that the performance of the GR2M under high-flow conditions during the validation period was unsatisfactory at all stations, excluding E05A022. It is noteworthy that the GR3M tends to carry high-flow simulations into acceptable or good grades for the same stations. It is also intriguing, however, that GR3M_PA has gone beyond what it provided for all flows within the scope of high-flow simulations. While it ensures a PI greater than 10% at each station, except for E05A022 and D05A038, the high-flow simulations produced have remained unsatisfactory. In turn, it is worth mentioning that the PI values delivered by hybrid GR3M at stations E05A023, E05A014, and D05A038 surpass those of GR3M_PA by over 20% (Figure 3).

Moreover, FITEVAL software using MATLAB built-in functions was exploited to obtain the RFDs of the goodness-of-fit being within the considered performance grades. These values for the validation period were compiled for all stations, and the results were made interpretable through a boxplot graph across the GRB (Figure 4). As for the overall fit of the hydrographs, GR3M is 33% more likely to produce a very good simulation compared with that of GR2M. In addition, the relatively narrow interquartile range associated with GR3M reveals that the likelihood of obtaining a simulation in that category with this model does not vary considerably between stations (see leftmost panel of Figure 4). This indicator serves as evidence of the effectiveness of the hybrid model over the GRB. While the original model delivers simulations with a good grade for roughly a quarter of the current period, it is promising that almost all of this RFD shifts to the category of very good through the hybridized scheme. Interestingly, the RFDs regarding GR3M_PA appeared at an intermediate level between those of GR2M and GR3M. It is inferred from the same figure that 60% of high-flow months are under-represented with the original

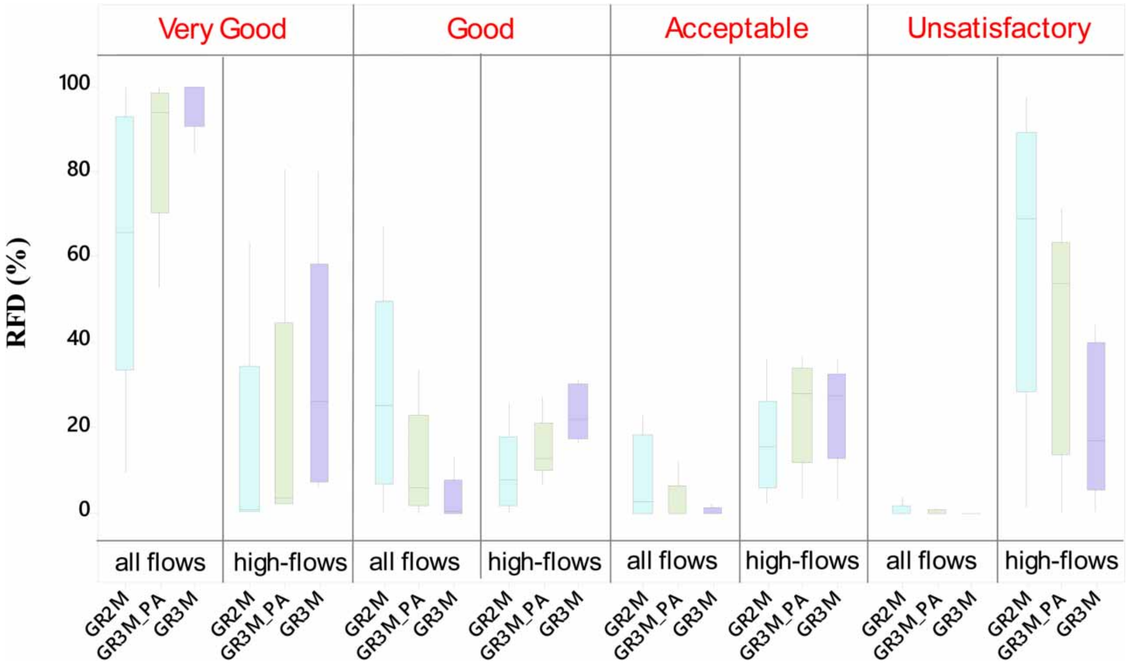


Figure 4 | Comparing the goodness-of-fit of the models according to the corresponding frequencies of being within the recommended performance ratings.

model. Conversely, as shown in the rightmost panel of Figure 4, hybrid GR3M dramatically decreased the relative frequency of these simulations that could be deemed unsatisfactory during peak flow conditions. During the validation phase at the E05A023 station, we also, as an extra illustration, focused on the runoff generation capabilities of the models in some wet years (Supplementary Figure S1). Evidently, the GR2M exhibited a tendency to overestimate the peak points in the water years 2002, 2009, and 2019 by 72, 50, and 32 Mm^3, respectively, despite successfully estimating the peak point in the 1999 water year. More effectively than GR3M_PA again, GR3M has mitigated these errors by around 80–90%.

Furthermore, Taylor diagrams were employed to assess the competence of runoff simulations at each station during the validation period, considering all flows (Supplementary Figure S2) and high flows (Supplementary Figure S3). It is discernible from these diagrams which MWBMs are in closest proximity to the observations. Accordingly, GR3M yielded RMSD values that were comparatively small, providing further evidence that the transition of the routing process to an ML-driven nonlinear form is consistent.

In addition to the advantages associated with hybrid models using ML, there are some obstacles that need to be addressed and research priorities. As simulations from these models are blended with ML approaches, they may inevitably inherit some of the limitations of these techniques (Slater *et al.* 2023). Frequently addressed limitations of ML models include the need for large data sets to train, sensitivity to collinearity, and overtraining (Ghaith *et al.* 2020). In our study, nearly 30 years of data were used, and the training and testing of GR3M seem not to be affected by the length of the data. Yet, it is advisable to validate the suggested model in other catchments employing data of varying durations. On the contrary, since the LSSVR embedded into the GR3M is routing with P_1 and P_2 from the production storage, the collinearity in this dimension is marginal. However, we recommend focusing on collinearity in cases where multiple variables computed from the conceptual layer of the hybrid model are serving as input for the data-driven phase (e.g., Kumanlioglu & Fistikoglu 2019).

We also believe that there are opportunities to increase the effectiveness of and disseminate nested hybrid models like GR3M. One avenue worth trying would be to incorporate physics-informed ML designs into the nested framework that tends to obey the law of conservation of mass. For instance, some recent articles focusing on physics-guided deep learning models (i.e., long short-term memory (LSTM) networks and convolutional networks) for runoff simulation attract attention (e.g., Xie *et al.* 2021; Bhasme *et al.* 2022; Frame *et al.* 2023; Zhong *et al.* 2023). Among these, Xie *et al.* (2021) revealed that the physics-guided LSTM network model outperformed the traditional LSTM in experiments conducted in more than 500 basins. Conversely, Frame *et al.* (2023) demonstrated that physics-guided deep learning models can exhibit weaker performance when confronted with biases detected in observations, as opposed to unconstrained deep learning models that account for residuals. Although mass conservation was taken as the basis for the production storage of GR3M proposed in this study, findings in the recent literature made us wonder how a nested model, which is hybridized LSTM with mass conservation constraints, could yield runoff simulations. We have the intention of conducting an investigation of this context in the future.

Furthermore, how to incorporate techniques for isolating noises in time series into the nested hybridization procedure will be the subject of another future study. For example, the wavelet transformation has been implemented in recent years to enhance the performance of runoff simulations by decomposing the periodic structure in conceptual model outputs (e.g., Wang *et al.* 2021; Sezen & Partal 2022) or hydroclimate records (e.g., Remesan *et al.* 2009; Bajirao *et al.* 2021). Thus, our further research might focus on evaluating the viability of this methodology as an intermediate preprocessing within a nested hybrid model.

In addition to all these, as reported by Slater *et al.* (2023), there may be concerns about how these hybrid models could complement or supersede traditional models in operational contexts. To persuade operational forecasters that hybrid models can bring value to the hydrology literature when compared with conventional ones, it is recommended to adopt the nested hybridization style as demonstrated in this paper. The success of blending the strengths of a conceptual model and an ML algorithm in one combinatorial scheme and ensuring that their parameters interact with one another has been revealed in a rare number of papers (e.g., Okkan *et al.* 2021). This study has shown that interfering with the linearly behaved components in even simpler MWBMs with the mentioned hybridization strategy can considerably enhance the monthly simulation accuracy and hopes to provide an insight for operational forecasters.

5. CONCLUSION

This study was conceived with the following two subsequent motivations in mind: (i) there are rather limited papers focusing on the modification of the runoff partitioning structure in MWBMs using ML techniques for better simulation performance. (ii) Given the significance of high-flow simulation, particularly in regions with intensive irrigation efforts (e.g., GRB), the fact that a commonly employed model (GR2M) generates biased estimates during periods of high precipitation could be a major worry. Despite a limited number of studies addressing these two concerns (e.g., Cheng *et al.* 2020; Fathi *et al.* 2023), ML techniques have not yet been widely utilized in the internal structural modification of MWBMs. Although it was previously discovered by Okkan *et al.* (2021) that there are ML-oriented improvements in more parameter-intensive MWBMs, this study reveals for the first time how the nested hybridization style yields outcomes in simpler structured and benchmark models such as GR2M.

In the study carried out considering all these motivations, a hybrid model termed GR3M was constructed by substituting only the routing store of GR2M with LSSVR, and it was studied to what extent this novel approach provided accurate monthly simulation. The performance measures suggest that the hybrid model manipulating nonlinear routing is more capable of capturing the rainfall–runoff relationship as opposed to the standard GR2M and GR3M_PA. The hybrid model, which is also not laborious to calibrate due to its simplicity, can be recognized as being rather robust in simulating both all flows and high flows. Nevertheless, diagnostics of the performance of the model over numerous catchments are undoubtedly essential for generalization. Moreover, the models in this hybridization style cannot claim to deal with basin dynamics in all their aspects. But the notion that ML techniques such as LSSVR, which contain few parameters to be tuned and have the flexibility to blend the conceptual outputs, can be utilized in MWBMs that require intervention in the runoff-generating mechanism has been criticized. The developed hybrid model has the potential to be adapted to operational studies as it preserves the conceptual structure of the original model with a nested hybridization.

AUTHORS' CONTRIBUTION

U.O. wrote the main manuscript text and provided the conceptualization of the study. U.O. and Z.B.E. prepared the MATLAB codes, all figures, and tables. U.O. and O.F. contributed to the development of the methodology proposed. U.O. and Z.B.E. revised the manuscript. All authors reviewed the manuscript.

FUNDING

No funding support was received for this study.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Bai, P., Liu, X., Liang, K. & Liu, C. 2015 Comparison of performance of twelve monthly water balance models in different climatic catchments of China. *Journal of Hydrology* **529**, 1030–1040. https://doi.org/10.1016/j.jhydrol.2015.09.015.

Bajirao, T. S., Kumar, P., Kumar, M., Elbeltagi, A. & Kuriqi, A. 2021 Potential of hybrid wavelet-coupled data-driven-based algorithms for daily runoff prediction in complex river basins. *Theoretical and Applied Climatology* **145** (3–4), 1207–1231. https://doi.org/10.1007/s00704-021-03681-2.

Bazrkar, M. H. & Chu, X. 2022 Development of category-based scoring support vector regression (CBS-SVR) for drought prediction. *Journal of Hydroinformatics* **24** (1), 202–222. https://doi.org/10.2166/hydro.2022.104.

Bhasme, P., Vagadiya, J. & Bhatia, U. 2022 Enhancing predictive skills in physically-consistent way: Physics informed machine learning for hydrological processes. *Journal of Hydrology* **615**, 128618. https://doi.org/10.1016/j.jhydrol.2022.128618.

Chen, J. & Adams, B. J. 2006 Semidistributed form of the tank model coupled with artificial neural networks. *Journal of Hydrologic Engineering* **11** (5), 408–417. https://doi.org/10.1061/(ASCE)1084-0699(2006)11:5(408).

Cheng, S., Cheng, L., Liu, P., Zhang, L., Xu, C., Xiong, L. & Xia, J. 2020 Evaluation of baseflow modelling structure in monthly water balance models using 443 Australian catchments. *Journal of Hydrology* **591**, 125572. https://doi.org/10.1016/j.jhydrol.2020.125572.

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

Fathi, M. M., Awadallah, A. G. & Aldahshoory, W. 2023 An improved monthly water balance GR2M model with a seasonally variable parameter. *Journal of Hydrology* **617**, 129127. https://doi.org/10.1016/j.jhydrol.2023.129127.

Frame, J. M., Kratzert, F., Gupta, H. V., Ullrich, P. & Nearing, G. S. 2023 On strictly enforced mass conservation constraints for modelling the rainfall-runoff process. *Hydrological Processes* **37** (3), e14847. https://doi.org/10.1002/hyp.14847.

Ghaith, M., Siam, A., Li, Z. & El-Dakhakhni, W. 2020 Hybrid hydrological data-driven approach for daily streamflow forecasting. *Journal of Hydrologic Engineering* **25** (2), 04019063. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001866.

Gong, W. & Cai, Z. 2013 Differential evolution with ranking-based mutation operators. *IEEE Transactions on Cybernetics* **43** (6), 2066–2081. https://doi.org/10.1109/TCYB.2013.2239988.

Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., Freer, J., Savenije, H. H. G. & Gascuel-Odoux, C. 2014 Process consistency in models: The importance of system signatures, expert knowledge, and process complexity. *Water Resources Research* **50** (9), 7445–7469. https://doi.org/10.1002/2014WR015484.

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

Kapoor, A., Pathiraja, S., Marshall, L. & Chandra, R. 2023 DeepGR4J: A deep learning hybridization approach for conceptual rainfall–runoff modelling. *Environmental Modelling & Software* **169**, 105831. https://doi.org/10.1016/j.envsoft.2023.105831.

Katipoglu, O. M. & Sarigol, M. 2023 Improving the accuracy of rainfall–runoff relationship estimation using signal processing techniques, bio-inspired swarm intelligence and artificial intelligence algorithms. *Earth Science Informatics* 1–17. https://doi.org/10.1007/s12145-023-01081-w.

Kumanlioglu, A. A. & Fistikoglu, O. 2019 Performance enhancement of a conceptual hydrological model by integrating artificial intelligence. *Journal of Hydrologic Engineering* **24** (11), 04019047. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001850.

McIntyre, N. 2013 Apportioning non-linearity in conceptual rainfall–runoff models: Examples from upland UK catchments. *Hydrology Research* **44** (6), 965–981. https://doi.org/10.2166/nh.2013.184.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D. & Veith, T. L. 2007 Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE* **50** (3), 885–900. https://doi.org/10.13031/2013.23153.

Mouelhi, S., Michel, C., Perrin, C. & Andréassian, V. 2006 Stepwise development of a two-parameter monthly water balance model. *Journal of Hydrology* **318** (1–4), 200–214. https://doi.org/10.1016/j.jhydrol.2005.06.014.

Okkan, U. & Kirdemir, U. 2020 Towards a hybrid algorithm for the robust calibration of rainfall–runoff models. *Journal of Hydroinformatics* **22** (4), 876–899. https://doi.org/10.2166/hydro.2020.016.

Okkan, U., Ersoy, Z. B., Kumanlioglu, A. A. & Fistikoglu, O. 2021 Embedding machine learning techniques into a conceptual model to improve monthly runoff simulation: A nested hybrid rainfall–runoff modeling. *Journal of Hydrology* **598**, 126433. https://doi.org/10.1016/j.jhydrol.2021.126433.

Okkan, U., Fistikoglu, O., Ersoy, Z. B. & Noori, A. T. 2023 Investigating adaptive hedging policies for reservoir operation under climate change impacts. *Journal of Hydrology* **619**, 129286. https://doi.org/10.1016/j.jhydrol.2023.129286.

Peters, N. E. & Aulenbach, B. T. 2011 Water storage at the Panola Mountain Research Watershed, Georgia, USA. *Hydrological Processes* **25** (25), 3878–3889. https://doi.org/10.1002/hyp.8334.

Remesan, R., Shamim, M. A., Han, D. & Mathew, J. 2009 Runoff prediction using an integrated hybrid modelling scheme. *Journal of Hydrology* **372** (1–4), 48–60. https://doi.org/10.1016/j.jhydrol.2009.03.034.

Ritter, A. & Muñoz-Carpena, R. 2013 Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments. *Journal of Hydrology* **480**, 33–45. https://doi.org/10.1016/j.jhydrol.2012.12.004.

Senbeta, D. A., Shamseldin, A. Y. & O'Connor, K. M. 1999 Modification of the probability-distributed interacting storage capacity model. *Journal of Hydrology* **224** (3–4), 149–168. https://doi.org/10.1016/S0022-1694(99)00127-4.

Sezen, C. & Partal, T. 2022 Two integrated conceptual–wavelet-based data-driven model approaches for daily rainfall–runoff modelling. *Journal of Hydroinformatics* **24** (5), 949–975. https://doi.org/10.2166/hydro.2022.171.

Slater, L. J., Arnal, L., Boucher, M. A., Chang, A. Y. Y., Moulds, S., Murphy, C., Nearing, G., Shalev, G., Shen, C., Speight, L. & Villarini, G. 2023 Hybrid forecasting: Blending climate predictions with AI models. *Hydrology and Earth System Sciences* **27** (9), 1865–1889. https://doi.org/10.5194/hess-27-1865-2023.

Storn, R. & Price, K. 1997 Differential evolution – A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization* **11**, 341–359. https://doi.org/10.1023/A:1008202821328.

Taylor, K. E. 2001 Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres* **106** (D7), 7183–7192. https://doi.org/10.1029/2000JD900719.

UNEP 1992 *World Atlas of Desertification*. Edward Arnold, London.

Vidyarthi, V. K. & Jain, A. 2022 Incorporating non-uniformity and non-linearity of hydrologic and catchment characteristics in rainfall–runoff modeling using conceptual, data-driven, and hybrid techniques. *Journal of Hydroinformatics* **24** (2), 350–366. https://doi.org/10.2166/hydro.2022.088.

Vu, H. L., Ng, K. T. W., Richter, A. & An, C. 2022 Analysis of input set characteristics and variances on *k*-fold cross validation for a recurrent neural network model on waste disposal rate estimation. *Journal of Environmental Management* **311**, 114869. https://doi.org/10.1016/j.jenvman.2022.114869.

Wang, J., Bao, W., Gao, Q., Si, W. & Sun, Y. 2021 Coupling the Xinanjiang model and wavelet-based random forests method for improved daily streamflow simulation. *Journal of Hydroinformatics* **23** (3), 589–604. https://doi.org/10.2166/hydro.2021.111.

Xie, K., Liu, P., Zhang, J., Han, D., Wang, G. & Shen, C. 2021 Physics-guided deep learning for rainfall–runoff modeling by considering extreme events and monotonic relationships. *Journal of Hydrology* **603**, 127043. https://doi.org/10.1016/j.jhydrol.2021.127043.

Xu, C. Y. & Singh, V. P. 1998 A review on monthly water balance models for water resources investigations. *Water Resources Management* **12**, 20–50. https://doi.org/10.1023/A:1007916816469.

Xu, D. M., Hu, X. X., Wang, W. C., Chau, K. W. & Zang, H. F. 2023 An enhanced monthly runoff forecasting using least squares support vector machine based on Harris hawks optimization and secondary decomposition. *Earth Science Informatics* 1–21. https://doi.org/10.1007/s12145-023-01018-3.

Zhong, L., Lei, H. & Gao, B. 2023 Developing a physics-informed deep learning model to simulate runoff response to climate change in Alpine catchments. *Water Resources Research* **59** (6), e2022WR034118. https://doi.org/10.1029/2022WR034118.

First received 2 January 2024; accepted in revised form 25 June 2024. Available online 8 July 2024