Using the Hybrid Simulated Annealing-M5 Tree Algorithms to Extract the If-Then Operation Rules in a Single Reservoir

Nazak Rouzegari[1] · Yousef Hassanzadeh[2] · Mohammad Taghi Sattari[3,4]

Abstract

The environmental water demand of the Mahabad River in the Urmia Lake basin in Iran was first estimated, using the flow duration curve shifting method (FDC Shifting) in this study. Secondly, the optimal operating model of the reservoir was developed with the goals of decreasing the deficiencies and considering the downstream demands of the reservoir, especially the environmental water demands by employing the simulated annealing (SA) and non-linear programming (NLP) methods. The results of the SA algorithm were compared with those of the NLP model with the indices of reliability, resiliency velocity, vulnerability, and sustainability. Then, the optimum released water values in the current month, the optimum water storage values in the reservoir, reservoir inflows and monthly demands were considered as inputs of the M5 tree model, and the optimum values of released water in the next month were considered as outputs of the M5 model. Finally, the optimum operation rules from the reservoir were developed in the form of if-then linear rules for future uses. One of the main advantages of the M5 tree model is to present two operation rules as if-then rules for all the operating periods with relatively acceptable accuracy. The results showed that the SA-M5 tree model, as a method of data mining, can extract the operation rules with relatively high accuracy.

Highlights

- The environmental water demand is obtained by the FDC Shifting method.
- A metaheuristic algorithm (SA) and a classical method (NLP) are applied to optimize the reservoir operation.
- The SA algorithm has good performance in comparison with the NLP method.
- The combination of SA and M5 tree model is proposed.
- The results of SA-M5 have good performance in the optimizing reservoir operation.

✉ Yousef Hassanzadeh
yhassanzadeh@tabrizu.ac.ir

[1] Department of Water Engineering, Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran

[2] Department of Water Engineering, Center of Excellence in Hydroinformatics, Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran

[3] Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran

[4] Department of Farm Structures and Irrigation, Faculty of Agriculture, Ankara University, Ankara, Turkey

1 Introduction

Urmia Lake is one of the most important and valuable aquatic ecosystems in Iran. The decrease of rainfall, the increase of temperature and irregular construction of dams in this basin have caused the Lake to dry. If no serious action takes place, the gradual destruction of the lake is certain (Hassanzadeh et al. 2012). As a result, increasing the inflow to the lake is one of the solutions that are taken into consideration. One of the rivers flowing into the Urmia Lake is the Mahabad Chay river. Mahabad dam is constructed on this river in the Urmia Lake basin. It is necessary to assess the ecological flow (EF) of the river to prevent the negative long-term effects of the constructed dams on the ecosystem of the river and Urmia Lake (Azarnivand et al. 2015; Jia et al. 2019). It is possible to consider some constraints and minimum values for EF, which can be observed in some studies (Jager and Smith 2008), but in this research, it is necessary to provide a complete ecological demand to revive the Lake due to its critical condition (Azarnivand et al. 2015; Alagoz and Yasi 2018). One of the most important purposes in the reservoir management issues is to meet all the demands of the downstream. So, planning and management of the reservoir operation is an important problem to consider (Bozorg Haddad et al. 2018). There are different methods to optimize the reservoir operation, including the classical methods and metaheuristic algorithms. Classical methods can be used in specific situations (Bahrami et al. 2017). In the last decade, heuristic and metaheuristic methods were widely used as the optimization tools in various fields (Ehteram et al. 2017; Bahrami et al. 2017; Ma et al. 2018; Xu and Mei 2018; Asadieh and Afshar 2019; Wang et al. 2019). The SA algorithm proposed by Kirkpatrick et al. (1982) is an optimization technique that has been taken into consideration and reported as an effective model in reservoir operation problems (Li and Wei 2008; Khodabakhshi et al. 2009; Kangrang et al. 2012; Mehr et al. 2019). The performance of this algorithm depends on the cooling process and running time. This algorithm has been used considerably (Henderson et al. 2006). Data-driven models (DDMs) are other methods that have become common in recent years and have shown high performance in water resource issues (Solomatine and Ostfeld 2008). Applying the data mining techniques on the calculation of the optimal reservoir operation is a novel method in the field of water resource management (Bozorg Haddad et al. 2018). The decision tree algorithm, which is a branch of data mining, has been widely used in reservoir issues (Sudha et al. 2006; Jothiprakash and Kote 2010; Sattari et al. 2012; Kumar et al. 2013). The M5 model has been used in different issues: predicting the inflow and evaporation at the reservoir, flood control, predicting the runoff-river model, sedimentation model, climate changes impact models and water pollutant model (Allawi et al. 2018; Oyebode et al. 2014). Revision of the resources so far has shown that there have been no researches done regarding the use of the combination of a metaheuristic algorithm (SA) and a data-driven model (M5 tree model), which is the novelty of this study, to calculate the optimum operation rules in the reservoir. This paper is organized as follows. The first section provides the environmental water demand, which has not been considered in the Mahabad reservoir, using the FDC shifting method to prevent Urmia Lake from drying. The next section provides the operation rules for the reservoir by the SA and NLP methods and evaluates both methods in finding the optimum volume of agricultural water that should be released in each month for historical data (1992–2013). Finally, the last section presents the efficiency of employing the

outputs of the SA algorithm as the input of the M5 tree model to find a meaningful operation rule curve. The aim is to know how these policies are changed for future uses in the Mahabad reservoir in different conditions and seasons.

2 Material and Methods

2.1 Study Area

In this research, the data from 1992 to 2013 of the Mahabad reservoir (see Fig. 1) have been used. The reservoir volume at a normal level is 197.8 million cubic meters (MCM) and its active volume is 170.3 MCM. The objectives of the dam are to provide the agricultural, drinking and industrial water, and control the flood. The information about precipitation and evaporation, inflow, municipal and industrial, and agricultural water demands are presented in Figs. 2, 3, 4 and 5.

2.2 FDC Shifting Method

The FDC Shifting method was introduced by Smakhtin and Anputhas (2006) to assess the EF in a river system. This method aims to protect rivers in the desired ecological state. The environmental management classes (EMCs) are provided in Table 1 (Smakhtin and Anputhas 2006). The main steps of this method are (Karimi et al. 2012): Simulating the reference hydrological states, introducing the environmental management classes, creating the environmental flow duration
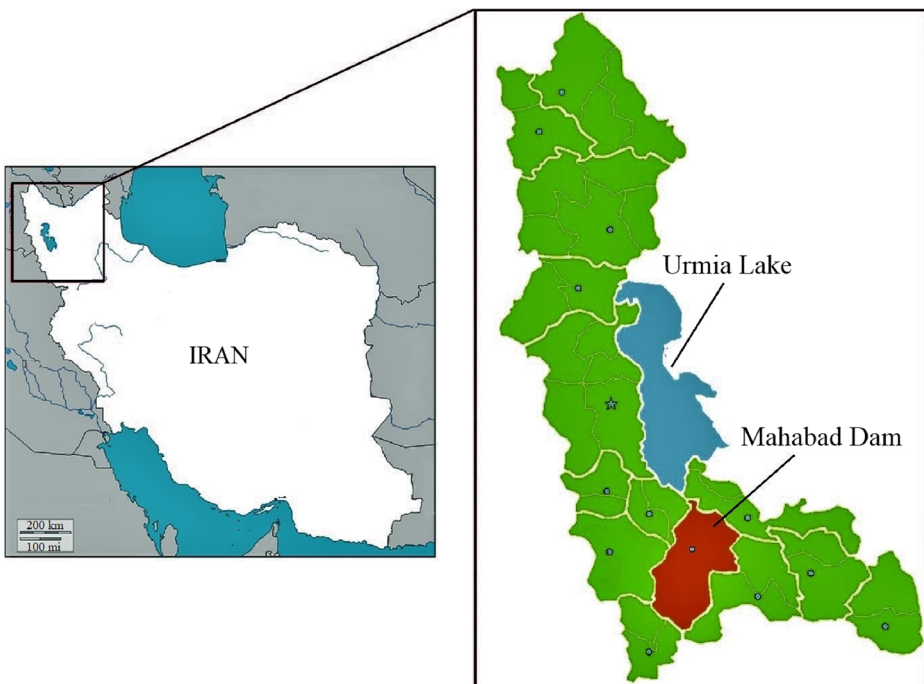

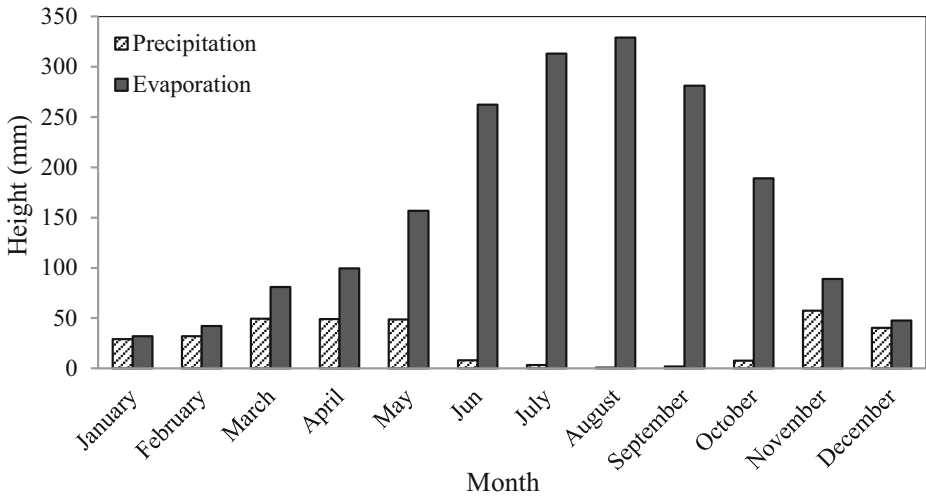**Fig. 1** The geographical position of Mahabad dam

Fig. 2 The average rate of precipitation and evaporation from Mahabad lake surface

curves and creating the continuous monthly EF time series. GEFC software is used in the flow duration curve method. It has been developed by the International Water Management Institute in 2007 and requires a long-time monthly flow data (at least 20 years).

2.3 SA

The SA algorithm is a known regional search method for optimization problems, having the feature of escaping local optima by choosing worse values of the objective function. The SA connects the dynamic behavior of solids with an optimization problem, in which the aim is to find the global minimum. The current and newly selected solutions and their objective function values are generated and compared at each iteration according to the dimensions of the problem in the space of possible solutions. Both improving and non-improving solutions are accepted based on some
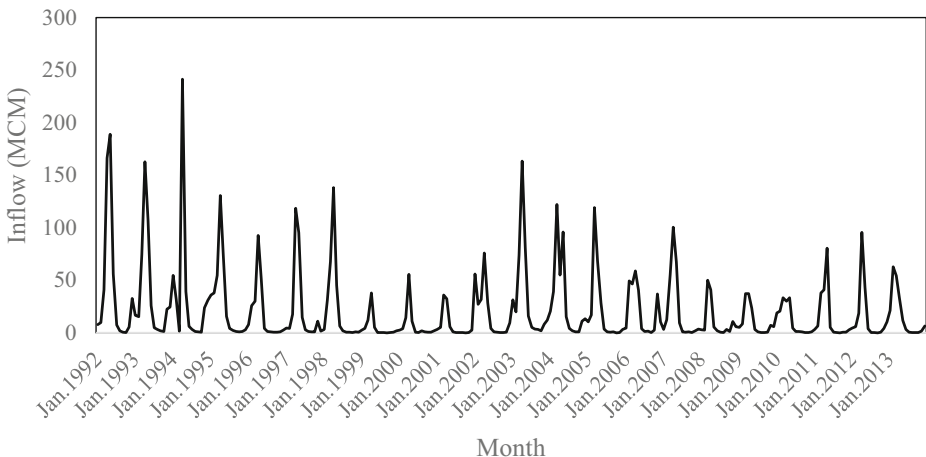

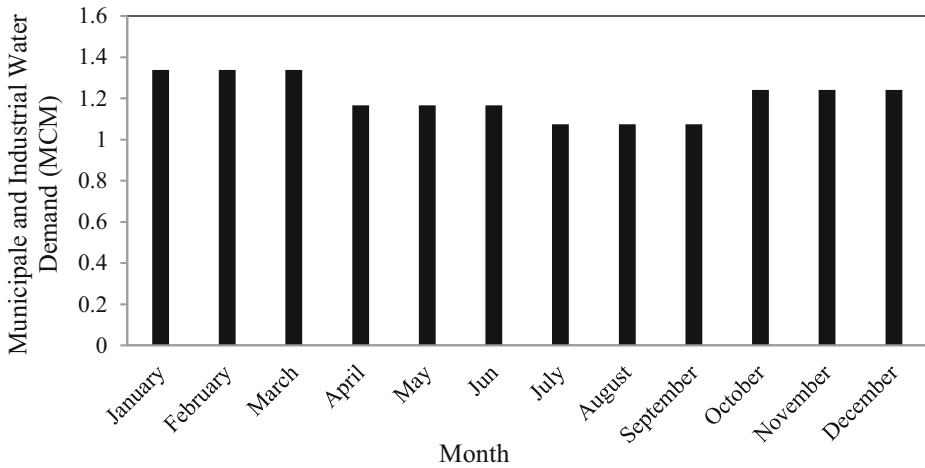**Fig. 3** Mahabad dam reservoir inflow during the statistical period

Fig. 4 Municipal and industrial water demand of Mahabad dam downstream

criterion, which is related to the temperature parameter that is fixed at each iteration (Henderson et al. 2006). The working process, the simulation model and the various SA parameters used in this research are presented in Fig. 6 and Table 2, respectively. In this table, T_0 is the initial temperature acting as the controlling parameter of the SA algorithm, and n is the repetition count or the Markov chain length.

2.4 NLP and GAMS

GAMS is an advanced program for solving complex linear and non-linear optimization problems with different solvers. In this program, solving methods are usually based on gradient. It is not a ready package, but it's a programming space with a specific coding language. To solve the NLP model with a classical method, a code is written and run in this program. Having an advanced programming language and being easily understood, GAMS
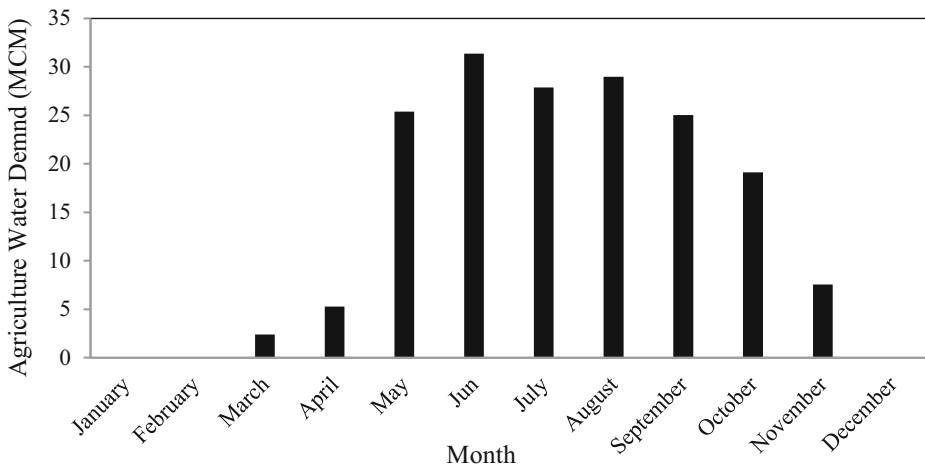

Fig. 5 Agricultural water demand of Mahabad dam downstream

Table 1 EMCs for FDC shifting (Smakhtin and Anputhas 2006)

EMC	Ecological description	Management perspective
A: Neutral	Pristine condition or minor modification of in-stream and riparian habitat	Protected rivers and basins. Reserves and national parks. No new water projects (dams, diversion, etc.) allowed
B: Slightly modified	Largely intact biodiversity and habitats despite water resources development and /or basin modifications	Water supply schemes or irrigation development present and/or allowed
C: Moder- ately modified	The habitats and dynamics of the biota have been disturbed, but basic ecosystem functions are still intact. Some sensitive species are lost and/ or reduced in extent. Alien species present	Multiple disturbances associated with the need for socio-economic development, e.g., dams, diversions, habitat modification, and reduced water quality
D: Largely modified	Large changes in natural habitat, biota, and basic ecosystem functions have occurred. A clearly lower than expected species richness. Much lowered presence of intolerant species. Alien species prevail	Significant and clearly visible disturbances associated with basin and water resources development, including dams, diversion, transfers, habitat modification, and water quality degradation
E: Seriously modified	Habitat diversity and availability have declined. A strikingly lower than expected species richness. Only tolerant species remain. Indigenous species can no longer breed. Alien species have invaded the ecosystem	High human population density and extensive water resources exploitation
F: Critically modified	Modifications have reached a critical level and ecosystem has been completely modified with the almost total loss of natural habitat and biota. In the worst case, the basic ecosystem functions have been destroyed and the changes are irreversible	This status is not acceptable from the management perspective. Management interventions are necessary to restore flow pattern, river habitats, etc. (if still possible/feasible) – to 'move' a river to a higher management category

can simply change the properties of the model and describe it. This software family consists of some other solvers including GAMS/MINOS, which is applicable both in linear and non-linear problems (Brooke et al. 1996).

2.5 Reservoir Operation Indices

Performance indices are used for comparing and evaluating the results of different methods. Reliability, resiliency, vulnerability, and sustainability indices are important indices in the water resource system as defined in Eqs. (1), (2), (3), and (4) (Hashimoto et al. 1982; Allawi et al. 2018).

$$\delta = 100 \left(1 - \frac{f}{F} \right) \tag{1}$$

$$\lambda = \left(\frac{f_s}{f} \right) \times 100 \tag{2}$$

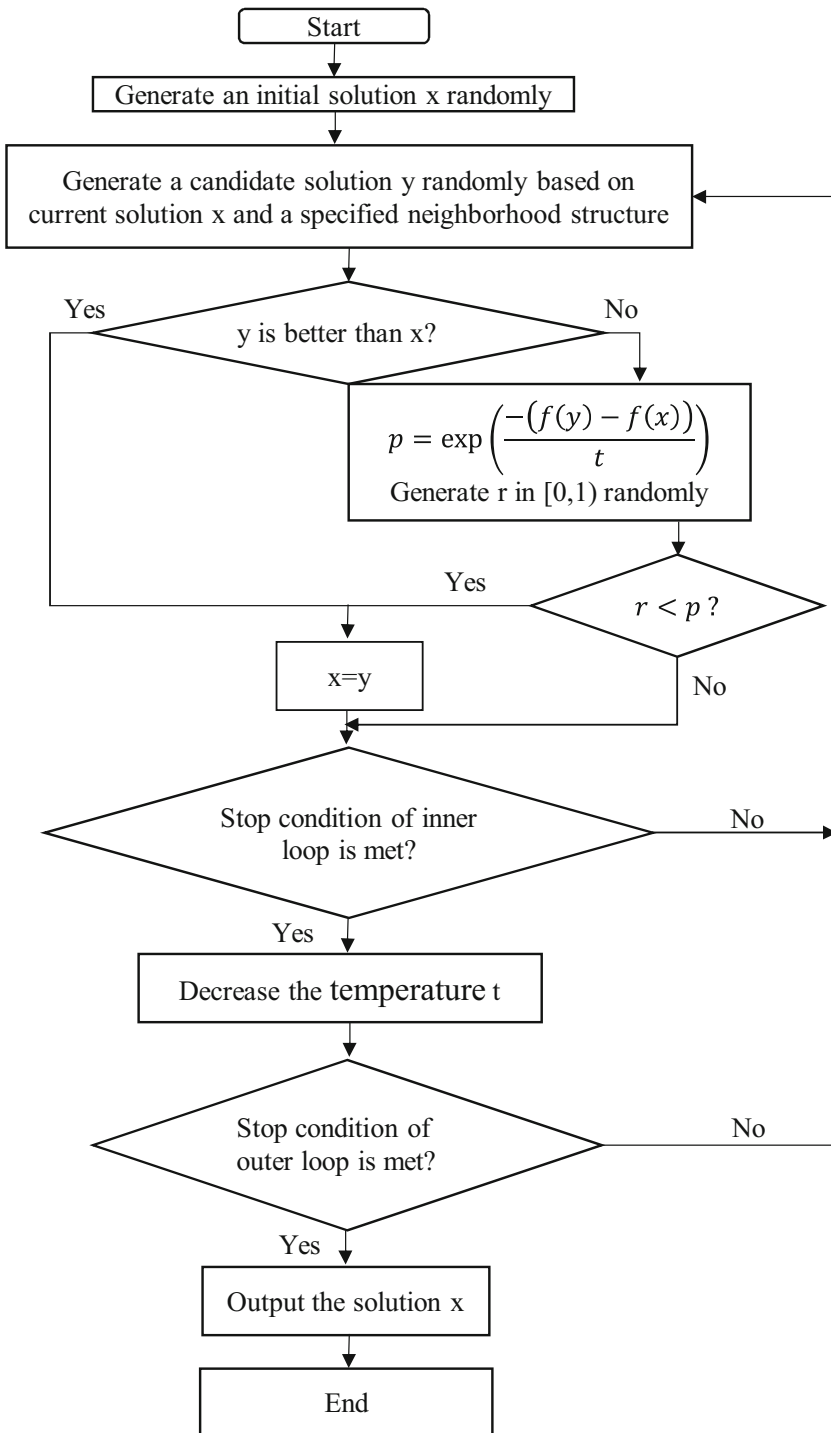$$\eta = \frac{\left(\sum_{t \in f} R_t^* - \sum_{t \in f} R_t \right)}{\sum R_t^*} \tag{3}$$

Fig. 6 SA algorithm flowchart (Zhan et al. 2016)

Table 2 SA algorithm parameters

Temperature function			Annealing function		T_0	n
Fast	Boltzmann	Exponential	Fast	Boltzmann	100	100

$$\varphi = \delta\gamma(1-\eta) \tag{4}$$

In these equations, δ is the reliability index ($0 \leq \delta \leq 1$), f is the total number of failure periods, F is the total number of utilization periods, λ is the resiliency velocity, f_s is the number of continuous failure period series, η is the vulnerability index, R_t^* is the intended demand in the tth failure period, R_t is the value of water release from the system during the tth failure period, and φ is the sustainability.

2.6 Optimization Model Development

In this research, the objective function includes the difference between the outflow rate and the downstream water demand according to the continuity equation. After setting this function, its minimum value calculates as Eq. (5) (Loucks et al. 1981).

$$Minimize \quad F = \sum_{i=1}^{N} \sum_{t=1}^{12} \left(\frac{DE_{(it)} - IR_{(it)}}{DE_{max}}\right)^2 \tag{5}$$

In this relationship, F is the value of the objective function; i is the year counter; N is the count of the statistical years; t is the month counter; IR_{it} is the volume of the withdrawn or agricultural water release from the dam reservoir in month t of the year N; DE_{it} is the agricultural water demand in month t of the year N, DE_{max} is the maximum monthly water demand during the operational period. The amount of water release from the dam reservoir in different months of the year is considered as the decision variable. The existing constraints for the dam reservoir's operational problem are defined as follows (Allawi et al. 2018):

$$ST_{t+1} = ST_t + QF_t + PP_t - EV_t - IR_t - DR_t - ECO_t - SPILL_t \tag{6}$$

$$A = A_0 + \alpha \times ST + \beta \times ST2 \tag{7}$$

$$ST_t \leq K_D + K_A \tag{8}$$

$$EV_t = e_t \times \left(A_0 + \alpha \times \left(\frac{ST_t + ST_{t+1}}{2}\right) + \beta \times \left(\frac{ST_t + ST_{t+1}}{2}\right)^2\right) \tag{9}$$

$$PP_t = P_t \times \left(A_0 + \alpha \times \left(\frac{ST_t + ST_{t+1}}{2}\right) + \beta \times \left(\frac{ST_t + ST_{t+1}}{2}\right)^2\right) \tag{10}$$

$$ST_t \geq K_D \tag{11}$$

$$IR_t \leq DE_t \tag{12}$$

$$SPILL_t = \left\{ \begin{array}{ll} ST_{t+1} - s_{max} & if \quad ST_{t+1} > s_{max} \\ 0 & if \quad ST_{t+1} < s_{max} \end{array} \right\} \tag{13}$$

$$penalty_t = \left\{ \begin{array}{ll} \left(\dfrac{ST_{t+1} - s_{min}}{s_{min}}\right)^2 & if \quad ST_{t+1} < s_{min} \\ 0 & if \quad ST_{t+1} > s_{min} \end{array} \right\} \tag{14}$$

In the above list, ST_{t+1} is the storage volume of the reservoir at the beginning of the $t+1^{th}$ period, QF_t is the volume of reservoir inflow in the tth period, PP_t is the precipitation volume on the lake surface in the tth period, EV_t is the evaporation volume from the lake surface in the tth period, IR_t is the withdrawn or agricultural water release volume in the tth period, DR_t is the required municipal water release volume, ECO_t is the required environmental water volume, $SPILL_t$ is the spillway volume in the tth period, A_0, α, and β are values obtained by drawing the surface-volume graph of the dam reservoir, and second degree curve fitting as Fig. 7 and they equal to 0.8878, 0.0808, and -0.0002, respectively. K_A and K_D are respectively the active and dead volumes of the dam, and e_t and p_t are the precipitation and evaporation heights during the tth period, respectively.

2.7 M5 Tree Model

The M5 tree model is proposed by Quinlan (1992). The output of these methods is a tree structure model that uses the input and output data and a linear relationship is presented for each class (Quinlan 1992). This model divides the domain of data into sub-areas (leaf) and a regression equation is given to each leaf (Pal et al. 2012) and it is based on two principles of the common decision tree and the linear regression functions. Two steps used in this model are

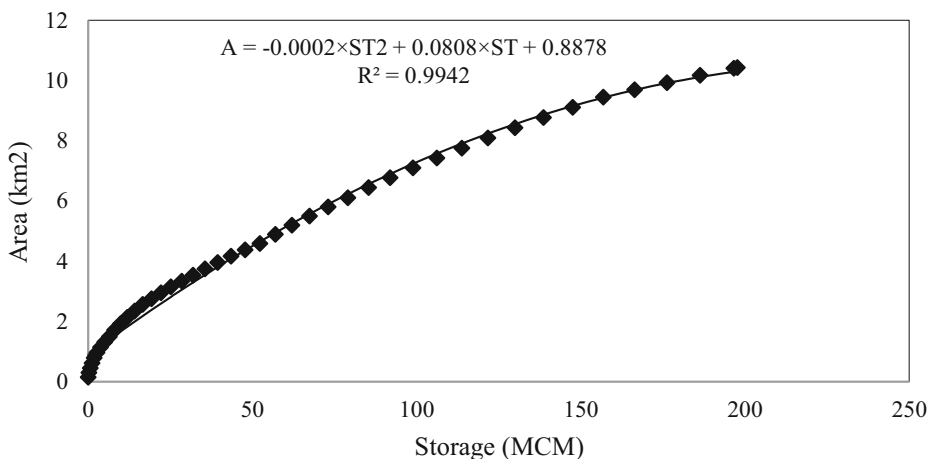$$A = -0.0002 \times ST2 + 0.0808 \times ST + 0.8878$$
$$R^2 = 0.9942$$

Fig. 7 Surface-volume diagram of the reservoir

splitting and pruning. The splitting criterion for the M5 algorithm is based on the standard deviation of the values of each class. The formula for measuring the standard deviation reduction (SDR) is as Eq. (15):

$$SDR = Sd(T) - \sum_{i=1}^{N} \frac{T_i}{T} SD(T_i) \qquad (15)$$

In this equation, T is a set of data entering each node; T_i indicates a subset of data that has the ith potential test result, N represents the number of the data and Sd is the standard deviation (Alberg et al. 2012). A bigger tree is going to be developed with this kind of splitting. Thus, the second step is pruning the tree model. The branches of the tree model are pruned and substituted with a regression function. The M5 model and Weka software (Witten and Frank 1999) developed at Waikato University in New Zealand, and are used in this research. The equation that is commonly used as reservoir operation rule curves can be in different forms (Bozorg Haddad et al. 2018). In this study, Eq. (16) has been used for the calculations in the derivation of dam reservoir operation rules.

$$IR_{t+1} = f(ST_t, QF_t, DE_t) \qquad (16)$$

In the above equation, f is a function sign of the decision variable, R_{t+1} is the amount of water release from the reservoir in the $t + 1^{th}$ period, QF_t is the amount of inflow to the reservoir in the tth period in MCM, ST_t is the storage volume of the reservoir at the beginning of the t period in MCM, and DE_t is the agriculture water demand at the begining of the t period in MCM.

3 Results and Discussion

In the process of the revival of Urmia Lake, allocating the EF of the Mahabad River, as one of the important catchments of the lake, is of high significance. Also, according to regional water reports, no water has been allocated to the environmental section from 1992 until 2013. Thus, the EF has been calculated for the first time and the optimum operating model has been
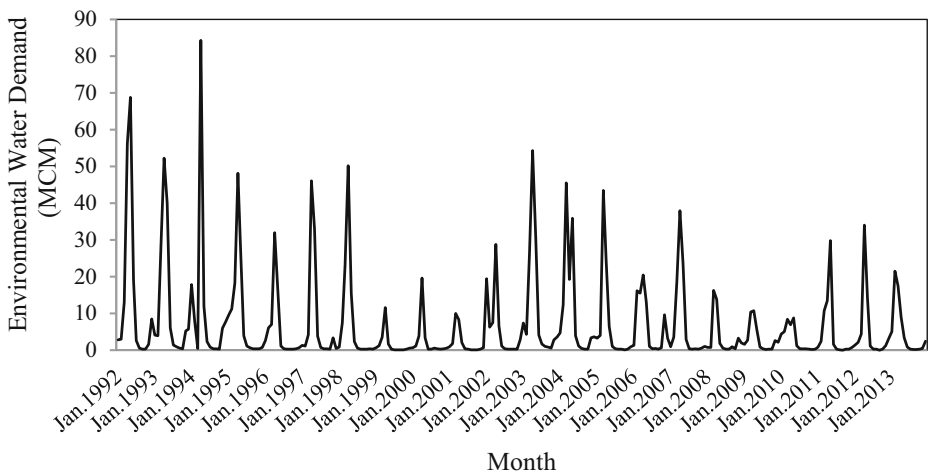

Fig. 8 Environmental water demand of Mahabad dam downstream

Table 3 Objective function values in 5 times of program running for SA algorithm

Temperature function	Annealing function	Iteration				
		1	2	3	4	5
Exponential	Fast	27.62	27.15	24.61	29.66	26.82
	Boltzmann	23.12	23.25	24.52	25.47	24.69
Boltzmann	Fast	26.25	25.26	27.8	27.04	29.79
	Boltzmann	21.36	21.57	20.42	**19.94**	20.56
Fast	Fast	23.06	23.94	22.97	21.27	24.28
	Boltzmann	24.78	24.3	26.66	23.97	23.16

developed in this research. The EF demand of the Mahabad River is estimated in this research, using the FDC shifting method. By using the FDC shifting method in the B management class (Fig. 8), a flow equal to 35% of the average annual flow, equal to 2.7 m^3/s, is needed for maintaining the Mahabad River in the least acceptable environmental condition. Using the SA optimization algorithm to optimize the Mahabad reservoir, a software code was first developed according to MATLAB programming. Values of the variables and the objective function for various SA algorithm parameters in 5 times of running the program and for 5000 iterations have been presented in Table 3. It is observed that the SA algorithm in the form of a Boltzmann temperature function, and the Boltzmann annealing function offer less value of the objective function which is equal to 19.94. Therefore, these parameters are chosen as the best options for the final run of the problem in the next step. According to Fig. 9, it is observed that the value of
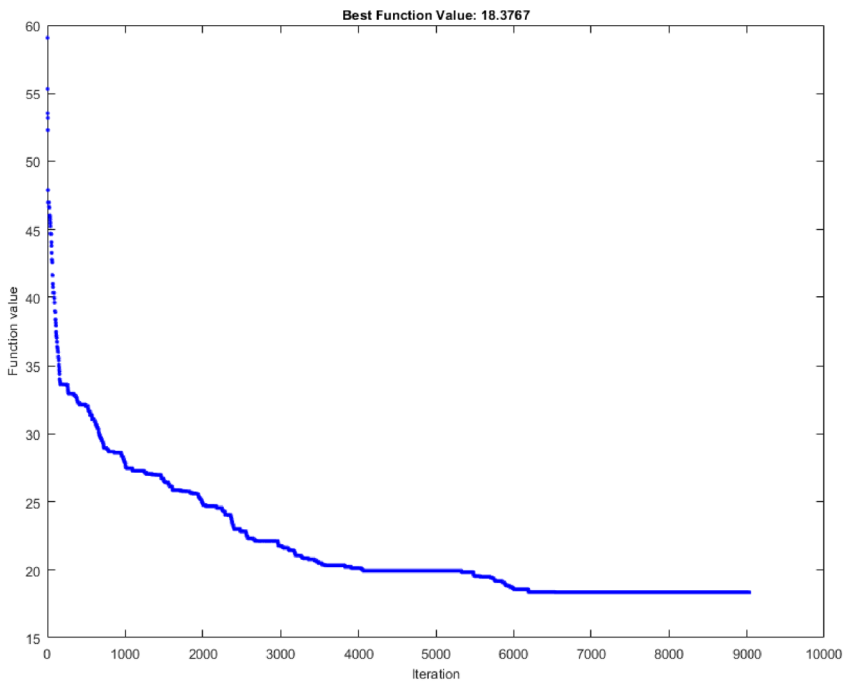

Fig. 9 The changes in the objective function value according to change in time in the SA algorithm

Table 4 Objective function value and the obtained performance indexes of both methods

Sustainability (%)	Vulnerability (%)	Resiliency Velocity (%)	Reliability (%)	Deficit (MCM)	Objective Function (MCM)	Method
0.002	0.99	39.82	59.05	2492.11	64.25	NLP
0.035	35.87	22.22	25	1426.49	18.37	SA

the objective function is obtained equal to 18.37 by using the SA. Table 4 shows the values of the objective function and the performance indices of both methods. It is inferred that the value of the objective function in the SA method is obtained less than the same value in the NLP. Moreover, according to Table 4, it is concluded that the SA method with the objective function value of 18.37 and the sustainability index of 0.035, has lesser objective function value, lower deficiency, and a higher stability index in comparison with the NLP method. According to Fig. 10, it is observed that the SA method has managed to reduce the deficiency rate up to an acceptable rate and has shown a proper performance. Comparing all the results, it is concluded that the SA algorithm has had a better performance in the optimum operation of the Mahabad dam reservoir and this method can be used to set optimum operational policies for the dam reservoir. The basic advantage of the SA algorithm is to use the worse results with a specific probability, and then moving towards the global optimum solution which is followed with escaping from the trap of the local optimum solution, although there is no warranty in finding a global optimum for metaheuristic algorithms, including SA algorithm. But the NLP solvers are trapped in the local optima in sophisticated or fairly complex functions due to the utilization of the classical methods. Most of the previous studies reported that the metaheuristic algorithms have much better performance compared to classical methods and have effectively met the demands (Allawi et al. 2018). There is the fact that metaheuristic algorithms don't always provide the global optimum solution by considering the results of the SA algorithm in this study, but at a reasonable computational time, it can provide acceptable results (Kumar and Reddy 2006). Figures 11 and 12 show the amounts of released and stored water volume from the Mahabad reservoir by using the methods of SA algorithm and NLP, respectively. According to Fig. 11, it is observed that the optimum amount of water released by the SA algorithm and the NLP method, in no periods go beyond the maximum water demand, which is approximately 31 MCM. However, the amount of water released by the operation of the
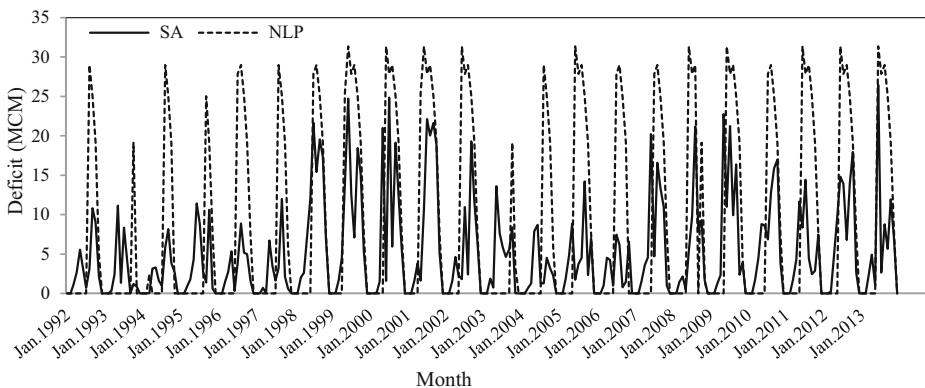
Fig. 10 Comparing the volume of the water deficit obtained from the SA and NLP methods
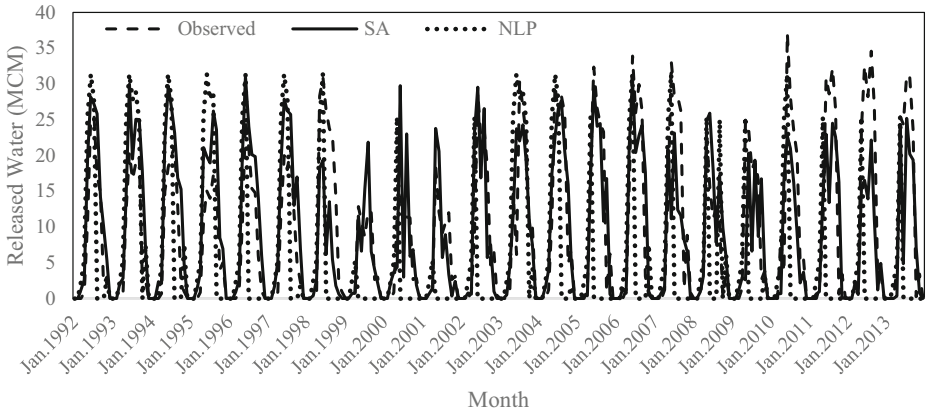
Fig. 11 Comparing the volume of the released water of reservoir obtained from NLP, SA and observational methods

dam has reached higher than this amount in some months, which is the evidence of misman-agement and lack of proper operation. It is obvious that these time series are alternative and show the seasonal changes. This indicates that the maximum amount of annual water release each year is related to the warm months of the year, and the minimum amount of water release corresponds to the colder months. According to Fig. 12, it is observed that the reservoir storage volume change in 22 years of the statistical period is between the maximum reservoir storage volume (197.8 MCM) and the minimum reservoir storage volume (25.26 MCM). According to Figs. 11 and 12, it seems that the operation of the dam has acted to store more water to avoid reservoir spill between the years 1992–1997 due to the existence of humid periods (Fig. 3). However, the agricultural water demand and the total water release in the downstream was low in these years due to not considering the environmental demand and lack of development in agricultural lands; consequently, less water was released from the reservoir to meet the needs of the downstream. Moreover, Fig. 11 indicates that after 2008, the operation of dam has been compelled to release more volume of water to meet the needs of the downstream due to the water shortage periods (Fig. 3) and increased cultivation area in recent years. This has caused the agricultural demand to be increased, and according to Fig. 12, it is observed that in these
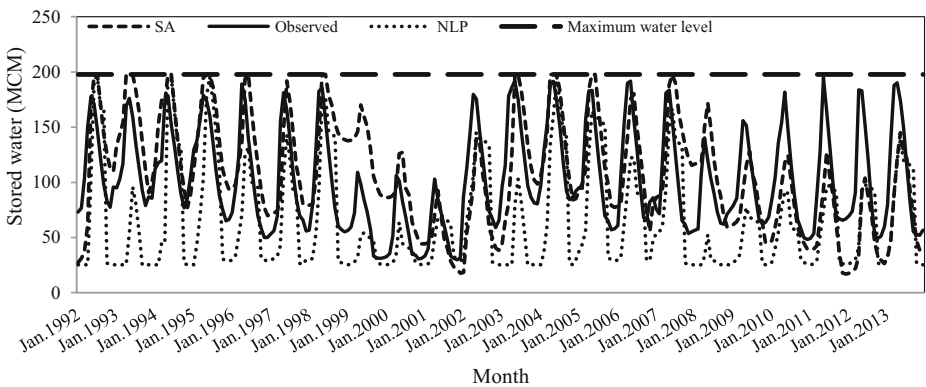

Fig. 12 Comparing the volume of the stored water of reservoir obtained from NLP, SA and observational methods
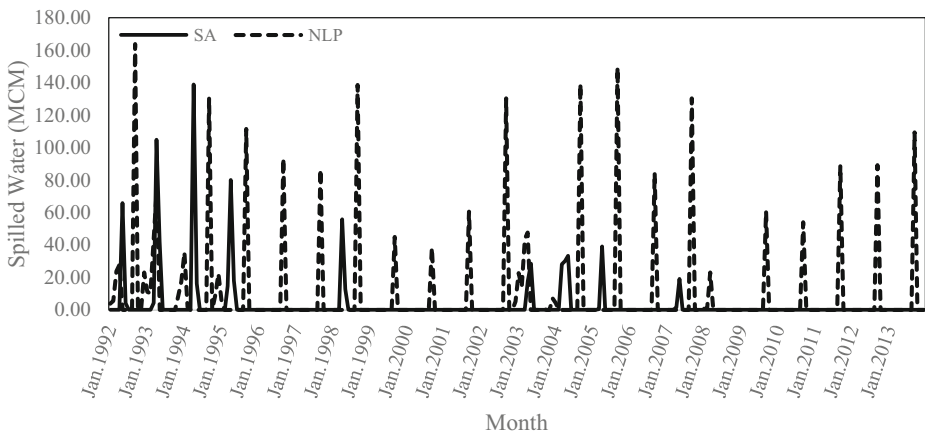
Fig. 13 The volume of spilled water obtained from NLP and SA methods M5 tree model

years, the amount of stored water behind the dam has increased just because the environmental demand of the downstream has not been faced and no water has not been released. Additionally, according to Fig. 12 and by referring to the reservoir inflow diagram (Fig. 3), it is inferred that the volume of the stored water in the reservoir reached its minimum amount only in those periods of low inflow and coincidence with water shortage periods. Considering that the main goal of the optimization model used in this research was to minimize the defects, so this optimization model has been utilized to lessen these deficiencies as far as possible by distributing them. As a result, a fundamental difference has occurred in the amount of the stored water and the released water between the optimizing and observational methods. Figure 13 presents the total volume of spilled water during a year from the Mahabad dam reservoir using the two SA and the NLP methods. It is observed that the amount of the spilled water obtained from the NLP method is more than the same in the SA method, which indicates that water waste has been prevented by using the SA method and optimum allocation of the water. Teegavarapu and Simonovic's (2002) research results, as the first ones who used the SA method to operate a four-reservoir system to provide agriculture and hydro-power demand show that SA, compared with NLP, has resulted in a more optimized objective function value than the NLP in dam reservoir operation. This is also compatible with the results of this research. Khodabakhshi et al. (2009) also showed that the SA algorithm can conclude better results than other metaheuristic algorithms in reservoir problems. Normally, the amount of the water release from the reservoir is considered as the decision variable in determining dam reservoir policies, which can be dependent on some parameters like the dam reservoir inflow, water demand in the downstream and the storage volume of the reservoir. The relationship

Table 5 The results of M5 tree model for four splitting data

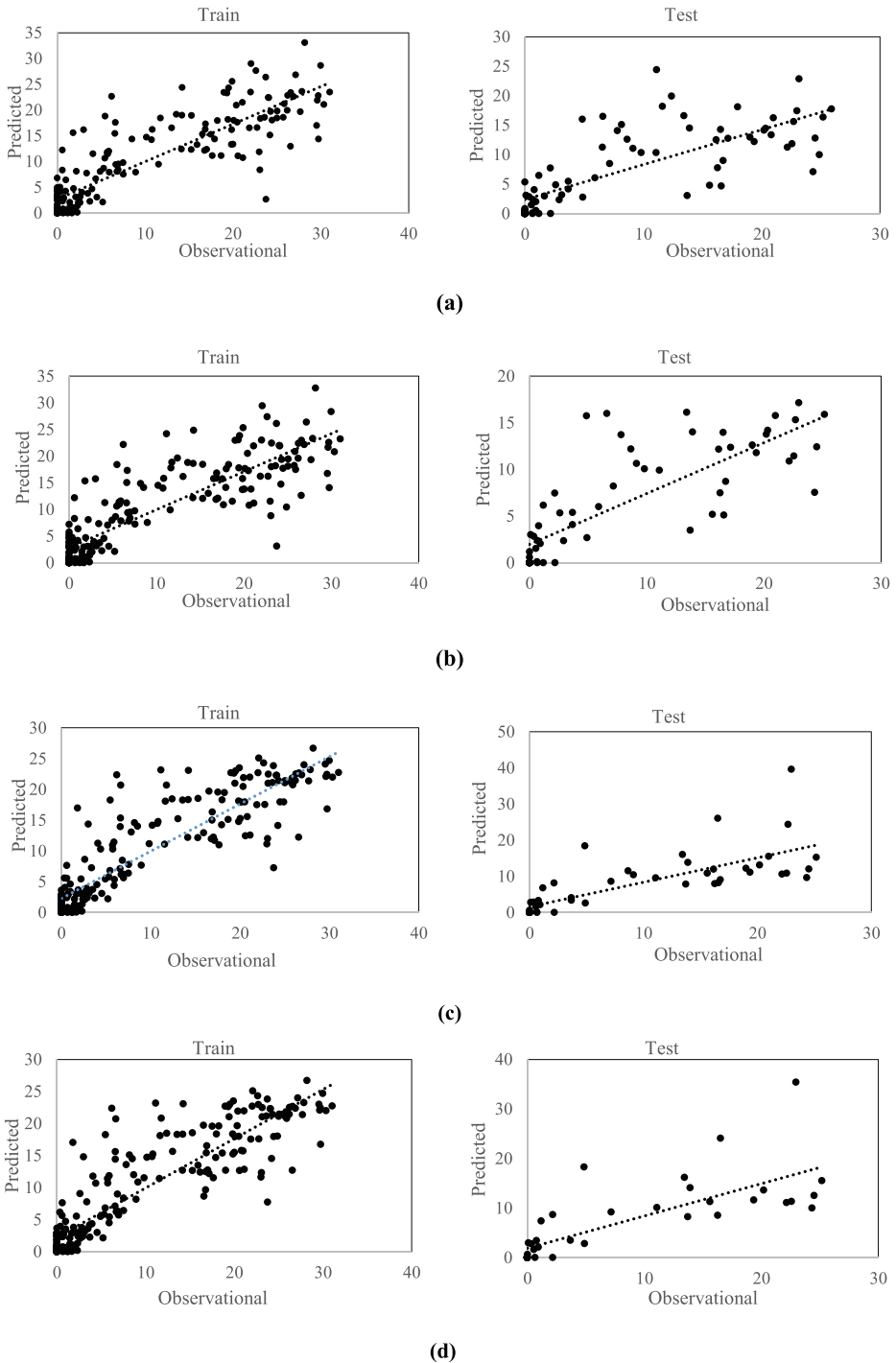| Number	Train(%)	Test(%)	For training			For testing		
			CC	RMSE	MAE	CC	RMSE	MAE
1	70	30	0.86	5.25	3.66	0.77	5.7	4.08
2	75	25	0.84	5.40	3.77	0.81	5.23	3.87
3	80	20	0.89	4.57	3.08	0.77	5.99	4.1
4	85	15	0.89	4.54	3.07	0.76	6.06	4.22

(a)

(b)

(c)

(d)

Fig. 14 Predicted released water volume distribution diagram of M5 vs. observational released water volume of SA for **a** 70% training, 30% testing, **b** 75% training, 25% testing, **c** 80% training, 20% testing, and **d** 85% training, 15% testing

Table 6 The results of operation rules obtained from M5Rules method (Training:75%, Testing: 25%)

If	Then
$DE_t \leq 3 \cdot 825$	$IR_{t+1} = 0.0118 \times ST_t + 0.0151 \times QF_t + 0.8012 \times DE_t - 0.9297$
$DE_t > 3 \cdot 825$	$IR_{t+1} = 0.0848 \times ST_t + 0.0852 \times QF_t + 0.4057 \times DE_t - 7.8463$

between the variables has been considered as Eq. (16) to derive the optimum operation rule curves in this study. The M5Rules option was used in the WEKA software to derive the if-then operation rules. The way it works is that, first, the optimum results obtained from the SA algorithm, including reservoir storage volume ST_t, downstream water demand DE_t, released water amount IR_{t+1} and also the volume of water inflow to the reservoir QF_t were input to the software as monthly data. Reservoir release as the objective variable was defined as a function dependent on the storage volume, downstream water demand and the volume of inflow water. Allawi et al. (2018) advised that the M5 tree model is used in four steps with different splitting data sets. In this research, 70, 75, 80 and 85% of total data were chosen as the training data and 30, 25, 20, 15% of total data were selected as the testing data, respectively. The results obtained from using the Weka software and the M5Rule option in predicting the amount of reservoir release shown in Table 5. Figure 14 indicates the scatter plot for the data of the predicted released water obtained from the M5 method compared with the data of observational release water obtained from the SA method. Different If-then rules are obtained for the four splitting data sets. By considering Fig.14 and Table 5, it can be concluded that in the case of 75% of data for training and 25% of data for testing, the CC value calculated for the testing data set has greater and the RMSE and MAE parameters have fewer values. These results indicate the high performance of the model in predicting the volume of released water from the reservoir in the future period, so the obtained results can be used as the optimum operation rules for the Mahabad reservoir. These rules are presented in Table 6. According to Table 6 and considering the linear relationships, it is concluded that the demand coefficient (DE_t) is more significant in comparison with other parameters. From this table, it can be seen that the M5 model is a piecewise linear model and can produce a simple, acceptable, transparent, provable and understandable model with a high convergence (Oyebode et al. 2014). Therefore, it is deduced that this parameter has more effect on water release and operation rules than any other parameters in the Mahabad dam.

4 Conclusion

Considering the drying of most lakes and ponds of Iran in recent years specifically, Lake Urmia, which its drying process has been intensified, studying how to allocate environmental flow right to the rivers flowing into the Lake Urmia with considering the hydrological, ecological, economic, social and agricultural conditions as well as allocating water to the agricultural section is one of the fundamental challenges of water management in Iran. In this research, the minimum EF of the Mahabad River with a goal of its full allotment from the Mahabad dam in Lake Urmia's revival has been calculated using the FDC shifting method. Based on field research, it seems that the current operating method of the Mahabad dam is dependent on the amount and condition of water flow in different months and seasons of the year and no particular rule curve based on optimization methods has been used in this dam. So, the NLP method with GAMS/MINOS solver and the metaheuristic SA algorithm is used for solving the optimum operational model of the single-reservoir system of

Mahabad dam in this research. The objective function was considered as minimization of differences between agricultural water demand and the volume of released water in case of providing complete environmental water demands (in providing the demand of Urmia Lake and a help to revive the lake), municipal and industrial water and some part of the agricultural section. According to the obtained results, it is observed that the NLP method failed to reach the global optima and it was trapped in the local optima and the SA metaheuristic method needs more time to find the optimum point compared with the NLP method. In this research, it was observed that although SA was able to present a more optimum solution than the NLP method, this gained point was not the global optima and it is essential that all the needed parameters in SA be chosen correctly in finding this point. This will cause a reduction in the program's running time. It cannot be claimed that the metaheuristic methods can find the global optimum and their convergence is proved if the number of their iterations tend to infinity. Finding the global optimum is not intended in many problems like optimization, the main goal is rather to find a satisfactory solution by spending a reasonable amount of time and calculations. Comparing all the results, it is possible to conclude that the SA algorithm had a better performance than NLP method in optimum operation of Mahabad dam reservoir and this method can be used in estimation of the optimum operation policies for the dam reservoir and the water needed for agricultural purposes in the downstream can be provided according to the results of this method. After proving the superiority of the SA method against MINOS solver, the outputs of the SA method have been used as inputs of the M5 tree model for the four different splitting data sets. It was observed that the M5 tree model presents the optimization operation rules of the reservoir in the coming period as linear relationships and it spends less time for calculation. It is suggested that the operation rules derived from this model can be used for future periods of the Mahabad reservoir according to the obtained results and due to the simplicity, practicality, and linearity of the M5 method. Presenting two operation rules as if-then rules for all operating periods with a partly acceptable precision is one of the main advantages of the M5 tree model, which eases the process of deciding the optimum allocation of water between different consumers.

References

Alagoz SH, Yasi M (2018) Urmia Lake environmental water right from Gadar River. Journal of Sigma 36(3): 861–870

Alberg D, Last M, Kandle A (2012) Knowledge discovery in data streams with regression tree methods. Journal of Wires Data Mining Knowledge Discovery 2(1):69–78

Allawi MF, Jaafar O, Hamzah FM, Abdullah SMS, El-shafie A (2018) Review on applications of artificial intelligence methods for dam and reservoir-hydro-environment models. Journal of Environmental Science and Pollution Research 25(14):13446–13469

Asadieh B, Afshar A (2019) Optimization of water-supply and hydropower reservoir operation using the charged system search algorithm. J Hydrol 6(1):5

Azarnivand A, Hashemi-Madani FS, Banihabib ME (2015) Extended fuzzy analytic hierarchy process approach in water and environmental management (case study: Lake Urmia Basin, Iran). Journal of Environmental earth sciences 73(1):13–26

Bahrami M, Bozorg Haddad O, Chu X (2017) Application of cat swarm optimization algorithm for optimal reservoir operation. J Irrig Drain Eng 144(1):04017057

Bozorg Haddad O, Aboutalebi M, Ashofteh PS, Loáiciga HA (2018) Real-time reservoir operation using data mining techniques. Journal of Environmental monitoring and assessment 190(10):594

Brooke A, Kendrick D, Mearaus A (1996) GAMS, a user guide. GAMS Development Corporation, Washington

Ehteram M, Karami H, Mousavi SF, El-Shafie A, Amini Z (2017) Optimizing dam and reservoirs operation based model utilizing shark algorithm approach. Journal of Knowledge-Based Systems 122:26–38

Hashimoto T, Stedinger JR, Loucks DP (1982) Reliability, resiliency, and vulnerability criteria for water resource system performance evaluation. J Water Resour Res 18(1):14–20

Hassanzadeh E, Zarghami M, Hassanzadeh Y (2012) Determining the main factors in declining the Urmia Lake level by using system dynamics modeling. J Water Resour Manag 26(1):129–145

Henderson D, Jacobson SH, Johnson AW (2006) The theory and practice of simulated annealing. In: Handbook of metaheuristics. Springer, Boston, MA, pp 287–319

Jager HI, Smith BT (2008) Sustainable reservoir operation: can we generate hydropower and preserve ecosystem values? Journal of River research and Applications 24(3):340–352

Jia W, Dong Z, Duan C, Ni X, Zhu Z (2019) Ecological reservoir operation based on DFM and improved PA-DDS algorithm: A case study in Jinsha river, China. Journal of Human and Ecological Risk Assessment: An International Journal 1–19

Jothiprakash V, Kote AS (2010) Effect of pruning and smoothing while using M5 model tree technique for reservoir inflow prediction. J Hydrol Eng 16(7):563–574

Kangrang A, Chaiyapoom W, Homwuttiwong S (2012) An application of simulated annealing algorithm with reservoir simulation model to improve reservoir rule curves. J Sci Technol MSU 31(5):578–584

Karimi SS, Yasi M, Eslamian S (2012) Use of hydrological methods for assessment of environmental flow in a river reach. Int J Environ Sci Technol 9(3):549–558

Khodabakhshi F, Ghirian AR, Khakzad N (2009) Applying simulated annealing for optimal operation of multi-reservoir systems. American Journal of Engineering and applied Sciences 2(1):80–87

Kirkpatrick S, Gelatt CD, Vecchi MP (1982) Optimization by simulated annealing. Journal of Science 220:671–680

Kumar DN, Reddy MJ (2006) Ant colony optimization for multi-purpose reservoir operation. J Water Resour Manag 20(6):879–898

Kumar SAR, Goyal MK, Ojha CSP, Singh RD, Swamee PK, Nema RK (2013) Application of ANN, fuzzy logic and decision tree algorithms for the development of reservoir operating rules. Journal of Water Resource Management 27:911–925

Li XG, Wei X (2008) An improved genetic algorithm-simulated annealing hybrid algorithm for the optimization of multiple reservoirs. Journal of Water Resource Management 22(8):1031–1049

Loucks DP, Stedinger JR, Haith DA (1981) Water resource systems planning and analysis, 1st edn. Prentice Hall, Englewood Cliffs, New York

Ma L, Wang H, Lu B, Qi C (2018) Application of strongly constrained space particle swarm optimization to optimal operation of a reservoir system. Journal of Sustainability 10(12):4445

Mehr AD, Jabarnejad M, Nourani V (2019) Pareto-optimal MPSA-MGGP: a new gene-annealing model for monthly rainfall forecasting. J Hydrol 571:406–415

Oyebode OK, Otieno FAO, Adeyemo J (2014) Review of three data-driven modelling techniques for hydrological modelling and forecasting. Journal of Fresenius Environmental Bulletin 23(7):1443–1454

Pal M, Singh NK, Tiwari NK (2012) M5 model tree for pier scour prediction using field dataset. KSCE J Civ Eng 16(6):1079–1084

Quinlan JR (1992) Learning with continuous classes. In 5th Australian Joint Conference on Artificial Intelligence 92: 343–348

Sattari MT, Apaydin H, Ozturk F, Baykal N (2012) Application of data mining approach to derive operation rules for the Eleviyan irrigation reservoir. Journal of Lake And Reservoir Management 28(2):142–152

Smakhtin VU, Anputhas M (2006) An assessment of environmental flow requirements of Indian river basins. IWMI Research Report 107, International Water Management Institute, Colombo, Sri Lanka

Solomatine DP, Ostfeld A (2008) Data-driven modelling: some past experiences and new approaches. J Hydroinf 10(1):3–22

Sudha V, Ambujam NK, Venugopal K (2006) A data mining approach for deriving irrigation reservoir operating rules. Conference on water observation and information system for decision support 23–26 may 2006. Orhid, Macedonia

Teegavarapu RSV, Simonovic SP (2002) Optimal operation of reservoir systems using simulated annealing. J Water Resour Manag 16(5):401–428

Wang J, Cheng Z, Ersoy OK, Zhang M, Sun K, Bi Y (2019) Improvement and application of chicken swarm optimization for constrained optimization. Journal of IEEE Access 7:58053–58072

Witten IH, Frank E (1999) Data mining: practical machine learning tools and techniques with Java implementations. Morgan Kaufmann, San Francisco

Xu Y, Mei Y (2018) A modified water cycle algorithm for long-term multi-reservoir optimization. Journal of Applied Soft Computing 71:317–332

Zhan SH, Lin J, Zhang ZJ, Zhong YW (2016) List-based simulated annealing algorithm for traveling salesman problem. Journal of Computational Intelligence and Neuroscience 5