Conditional Operation Rules For Optimal Conjunctive Use of Surface And Groundwater

Mina Khosravi
Iran University of Science and Technology
Abbas Afshar
Iran University of Science and Technology
Amir Molajou (amolajou@yahoo.com)
Iran University of Science and Technology

Research Article

Keywords: Conjunctive Use, Decision Tree Algorithm, Conditional Operation Rules, Uncertainty

Posted Date: December 8th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-994212/v1

Conditional Operation Rules for Optimal Conjunctive Use of Surface and Groundwater

Mina Khosravi[1], Abbas Afshar[2,3], and Amir Molajou[4*]

[1]Ph.D. Candidate, Department of Civil Engineering, Iran University of Science and Technology,

Email: mina.khosravi100@gmail.com

[2]Professor, Department of Civil Engineering, Iran University of Science and Technology,

[3]Department of LAWR, University of California, Davis, US

Email: a_afshar@iust.ac.ir

[4]Ph.D., Department of Civil Engineering, Iran University of Science and Technology,

*Corresponding author: amolajou@yahoo.com; Tel: +98 9141163014

Abstract

The current study presents an efficient method for deriving precise operation rules from all subsystems of a distributed conjunctive use system (CUS), including aquifer, river, and reservoir. Distributed aquifer simulation has been performed using the URM method. Given that the historical flow time series can only represent one of the possible situations in the future and its use to determine the performance of the CUS is certainly not very reliable, in this study, river flow uncertainties are implicitly considered. To develop the operation rules, the time series of river flow were generated using autoregressive model. Then, the operation optimization model of the system was implemented with the objective function of minimizing water shortage for different river flow time series. 70% of the data was used for model training and 30% for model validation. Finally, using the decision tree algorithm (M5Rules), the conditional operation rules were extracted and compared with the single linear regression operation rules. Using five efficiency criteria CC, MAE, RMSE, RAE, and RRSE, the comparison of conditional and single linear regression operating rules has been done. The results showed that the the conditional operation rules reduces relative absolute error by a minimum of 39% and a maximum of 71%. If the system is operated according to the conditional rules, in the worst case, the amount of water shortage imposed will be 16.61 MCM over ten years.

Keywords: Conjunctive Use. Decision Tree Algorithm. Conditional Operation Rules. Uncertainty

1. Introduction

Given that the optimal use of water resources systems requires operation rules, it is necessary to provide valid methods for their extraction (Nayak et al., 2018; Nourani et al., 2020). Operation

rules are one of the appropriate tools in water resources planning that estimate the time and amount of optimal harvesting from different water sources (Fallah-Mehdipour et al., 2015).

Uncertainty analysis is inevitable to extract operation rules because the extracted operation rules must be reliable for future uncertain conditions. In other words, operating rules that are presented only for a historical series of the river flow (as a representative of the state of the river flow in the future) do not provide a realistic assessment of the system (Schoups et al., 2006; Yang et al., 2017; Zhao et al., 2014).

So far, relatively few studies have been conducted to provide rules for the optimal operation of conjunctive use system (CUS) of surface water (SW) and groundwater (GW). A group of these studies have not considered the uncertainty of SW resources and have derived the operation rules only for a historical series of flow, as a representative of the future flow (Fallah-Mehdipour et al., 2013, 2015; Pan et al., 2016; Afshar et al., 2010). The second group of these studies has mainly adopted two approaches to extract operation rules that are accurate and optimal for possible runoff time series in the future (Schoups et al., 2006): (i) extract the operation rules based on explicit stochastic optimization (ESO), and (ii) extract the operation rules based on implicit stochastic optimization (ISO). In the first approach, uncertainties in the series of inflows to the reservoir are explicitly included in the objective function (Buras, 1963; Philbrick and Kitanidis, 1998; Onta et al., 1991). One of the most popular ESO methods is stochastic dynamic programming (SDP) which suffers from a curse of dimensionality as the optimization problem grows (Loucks et al., 1981). For this reason, ISO models are important in the analysis of water resources systems (Draper, 2001). In the ISO approach, two methods are mainly used. In the first method, the operation rules formulated with unknown coefficients are explicitly and directly embedded in the optimization model. Then, by implementing the optimization model for a historical flow series, the decision

variables of the optimization problem, which are the unknown coefficients of the operating rules, are obtained. Finally, the validity of these rules is measured by comparing them with the outputs of the optimization model for other probable river flow time series (Schoups et al., 2006; Afshar et al., 2008). In the second method, the deterministic optimization model is implemented for several short-term artificial river flow time series or one long-term historical time series. Then the outputs obtained from the solution of the model can be presented as operation rules using different methods such as Fuzzy Inference System (Milan et al., 2018; Safavi et al., 2013; Chang et al., 2013), Adaptive Neuro-based Fuzzy Inference System (Safavi et al., 2013), and Bayesian Networks (Rafipour-Langeroudi et al., 2014). Finally, the validity of these rules is confirmed by validation data.

Each of these studies, which presents the operation rules of a CUS, has shortcomings and needs to be examined in more depth. Some of these studies did not address uncertainties in the development of operating rules (Fallah-Mehdipour et al., 2013, 2015; Pan et al., 2016; Afshar et al., 2010). Some of them have provided the operation rules for lumped CUSs (Buras, 1963; Philbrick and Kitanidis, 1998; Onta et al., 1991). Another set of studies has developed operation rules for only one element of the CUS (Schoups et al., 2006; Safavi et al., 2013; Milan et al., 2018). However, water resources management with the view of integrated management of SW and GW resources needs to provide valid rules for all subsystems in a CUS.

Our reviews show that the presentation of operating rules for all subsystems of a distributed river-reservoir-aquifer system with regard to hydrological uncertainties remains intact. Therefore, in this study, the development of a simulation-optimization framework for extracting the operation rules of a complete and distributed CUS has been considered. In this study, we have used decision tree algorithm to derive the conditional operation rules (multi-linear regression operation rules or

89 M5Rules). Using the concepts of ISO, the operation rules of different parts of the CUS are
90 presented as a function of available SW and GW. In this regard, first 50 river flow time series were
91 produced using autoregressive model. Then, by implementing the optimization model with the
92 objective function of minimizing water shortage in the agricultural sector, the decision variables
93 of the optimization problem that indicate the optimal water transmission between various parts of
94 the system were extracted. 70% of the output of the optimization model was used as training data
95 and 30% as validation data. To extract the operation rules, in addition to fitting the multi-regression
96 model, a single-regression model was used to confirm the efficiency of the decision tree algorithm
97 in this field. It should be noted that in this study, distributed aquifer simulation has been performed
98 using the unit response matrix (URM) method.

99

100 ## 2. Methodology

101 ### 2.1. Case Study

102 The study region includes the Abhar River basin and Kinevers Dam (SW reservoir), which is
103 located in Iran. The studied system consists of a local aquifer with an area of 80 km^2. Table 1
104 shows the historical series of the Abhar River in 40 seasonal time steps from 2008 to 2018. SW is
105 conjunctively used with GW for urban water supply and agricultural purposes in the region.
106 Seasonal environmental, urban, and agricultural water demands are shown in Table 2.

107 **Table 1.** Seasonal inflow to the reservoir (MCM)

Year	1	2	3	4	5	6	7	8	9	10
Fall	2.34	2.85	3.59	5.15	15.84	4.96	4.95	3.96	2.77	1.84
Winter	6.54	5.96	8.63	15.70	13.60	11.75	7.89	10.41	5.51	5.17
Spring	7.37	33.59	9.48	15.32	19.58	49.20	6.17	17.85	2.57	5.80

Summer	0.58	1.16	0.87	0.94	1.76	1.45	0.91	0.93	0.52	0.24

Table 2. Agricultural, urban, and environmental water demands (MCM)

Season	Agriculture	Urban	Environmental
Fall	0.834	2.056	0.262
Winter	0.000	1.464	0.262
Spring	8.169	2.850	0.542
Summer	6.597	4.030	0.542
Annual	15.600	10.400	1.608

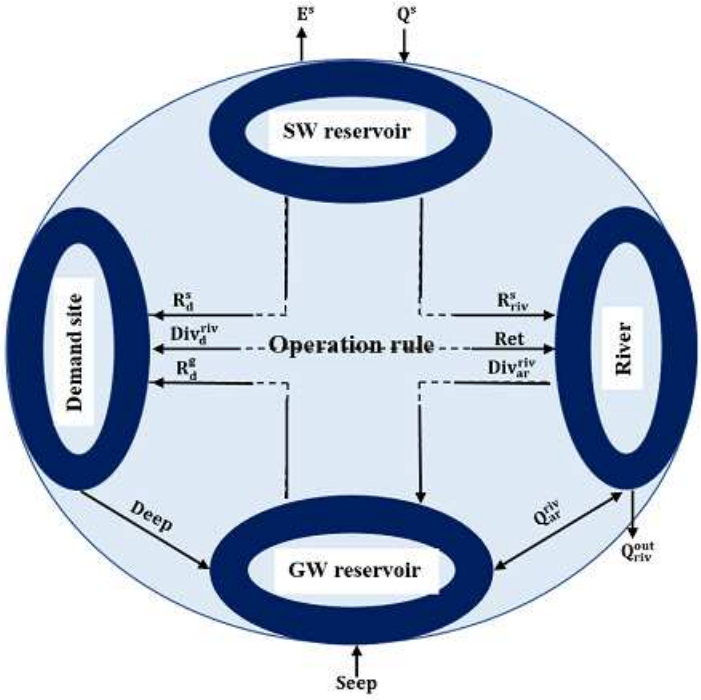
According to studies, the best area for artificial recharge is the pumping wells. The maximum drawdown in wells is 10 meters. The minimum and maximum recharging and pumping are equal to 3 MCM/season. According to field data, 10% of precipitation and 10% of the supplied water recharge the aquifer. The river is considered as a rectangular section with a width of 20 meters and a length of 20 kilometers. Manning coefficient and river slope are equal to 0.02 and 0.0001, respectively. 10% of the supplied water returns to the river. More information is available in the study of Afshar et al., 2021.

2.2. Conjunctive Use System (CUS)

Figure 1 shows a simplified model of the CUS under study. A description of all the components is shown in Table 3. According to Figure 1, the CUS is a complete system including SW reservoir, river, and GW reservoir that meet the region's water demands. The amount of transmission between different parts of the CUS is directly or indirectly determined by the operation rules. For example, the amounts of $R_d^s(t)$, $R_d^g(t)$, and $Div_d^{riv}(t)$ are determined directly, and the amount of deep percolation into the aquifer is indirectly determined by the operation rules. In fact, the deep

126 percolation into the aquifer is a function of total values $R_d^s(t)$, $R_d^g(t)$, and $Div_d^{riv}(t)$ and the

127 amounts of $R_d^s(t)$, $R_d^g(t)$, and $Div_d^{riv}(t)$ are determined by the operation rules.



128

129 **Fig. 1.** CUS elements and their interactions

130

131 **Table 3.** Description of CUS components

Component	Description
$\mathbf{R_{riv}^s}$	Water release from the dam to the river
$\mathbf{Q^s}$	Inflow to the dam
$\mathbf{E^s}$	Evaporation from the dam
$\mathbf{R_d^s}$	Water transmission from the dam to area of demand
$\mathbf{R_d^g}$	GW pumping
$\mathbf{Div_d^{riv}}$	Water diversion from the river to area of demand
$\mathbf{Q_{ar}^{riv}}$	Seepage from the aquifer to the river, or vice versa
$\mathbf{Div_{ar}^{riv}}$	Water diversion from the river to the area of artificial recharge
Ret	Return water flow from demand sites to the river reaches
Seep	Natural recharge from infiltrated precipitation

Deep	Deep percolation of irrigation water
$\mathbf{Q_{riv}^{out}}$	SW leaving the system's boundary

2.3. Simulation-Optimization Model of CUS

In previous studies, modeling of CUSs has been mainly simulation-optimization (Song et al., 2020; Kerebih and Keshari, 2021; Heydari et al., 2016). In simulation-optimization approach, using various optimization techniques, the best management plans are extracted from the CUS simulation model. Given that the present study focuses on a complete CUS, the four subsystems of reservoir, river, aquifer, and demand area and the interactions of these subsystems should be simulated. The following presents a simulation-optimization model of the CUS. The decision variables of the optimization problem are $R_d^s(t)$, $Div_d^{riv}(t)$, $Div_{ar}^{riv}(t)$, $R_d^g(t)$, and $R_{riv}^s(t)$. The problem of nonlinear non-convex optimization is solved with Lingo software.

Objective Function:

$$\text{Minimize} \sum_{n=1}^{NT} WS(t) \tag{1}$$

The objective function of model is to minimize the water shortage (WS) in NT time steps.

Reservoir Water Mass Balance:

$$S(t+1) = S(t) + \Delta S(t) \tag{2}$$

$$\Delta S(t) = Q^s(t) - E^s(t) - R_d^s(t) - R_{riv}^s(t); \quad \forall t \tag{3}$$

In which $\Delta S\,(t)$ = dam storage volume changes in time step t; $S\,(t)$ = dam storage volume at the beginning of the time step t.

Aquifer Water Mass Balance:

$$S^{g}\,(t+1) = S^{g}\,(t) + \Delta S^{g}\,(t);\ \ \forall t \tag{4}$$

$$\Delta S^{g}\,(t) = \sum_{n=1}^{NT} Div_{ar}^{riv}(t) - \sum_{n=1}^{NT} R_{d}^{g}\,(t) + \sum_{r=1}^{NR} kqv(t).\,Q_{ar}^{riv}(r,t) + Prc(t).\,Seep.\,AQA + Deep.\,Supply(t);\ \forall t \tag{5}$$

$$\sum_{w=1}^{NW} q_{p}(w,t) = R_{d}^{g}\,(t);\ \ \forall t \tag{6}$$

$$\sum_{w=1}^{NW} q_{ar}(w,t) = Div_{ar}^{riv}(t);\ \ \forall t \tag{7}$$

In which $\Delta S^{g}\,(t)$ = GW volume changes in time step t; $S^{g}\,(t)$ = GW volume at the beginning of the time step t; $q_{ar}(w,t)$ = volume of recharge to the well w; $q_{p}(w,t)$ = volume of pumping from well w; $kqv(t)$ = conversion factor (discharge to volume) in time step t; AQA = the aquifer surface area; $NW,\ NR$ = number of total wells and river reaches, respectively; Prc = precipitation depth.

River Water Mass Balance:

$$\Delta S_{riv}\,(r,t) = \left(Q_{riv}^{in}(r,t) + ql_{riv}(r,t) - Q_{riv}^{out}(r,t)\right).kqv(t);\ \ \forall r,t \tag{8}$$

$$ql_{riv}(r,t) = \frac{Area(r).Prc(t) - Div_{d}^{riv}(t) - Div_{ar}^{riv}(t) + Retr(r).Sup(t)}{kqv(t)} + Q_{ar}^{riv}(t);\ \ \forall r,t \quad (9)$$

$$\Delta S_{riv}(r,t) = Area(r).dh_{riv}(r,t);\ \ \forall r,t \quad (10)$$

$$Q_{riv}^{in}(1,t) = R_{riv}^{s}(t)/kqv(t);\ \ \forall t \quad (11)$$

$$Q_{riv}^{in}(r+1,t) = Q_{riv}^{out}(r,t);\ \ \forall r,t \quad (12)$$

154

155 In which $\Delta S_{riv}(r,t)$ = river storage volume changes; $Q_{riv}^{in}(r,t)$ = river inflow to the river
156 reach r; $ql_{riv}(r,t)$ = summation of lateral inflows or outflows along the river reach r; $dh_{riv}(r,t)$ =
157 river depth change in reach r in time step t; $Area(r)$ = river surface area. $Q_{riv}^{in}(r,t)$ and $Q_{riv}^{out}(r,t)$
158 are a function of river inflow and outflow depths, respectively.

Demand Site Water Mass Balance:

$$Supply(t) = R_{d}^{g}(t) + R_{d}^{s}(t) + Div_{d}^{riv}(t);\ \ \forall t \quad (13)$$

$$Demand(t) = WS(t) + Supply(t);\ \ \forall t \quad (14)$$

159

160 In which $WS(t)$ = water shortage; $Demand(t)$ and $Supply(t)$ are refer to water demand
161 and supply in time step t, respectively.

162 **Water Table Fluctuations in Aquifer:**

163 The amount of fall/rise in wells and river reaches can be calculated using the URM method
164 (Afshar et al., 2021). Details of this method are available in the studies of Afshar et al., 2020 and
165 Khosravi et al., 2020. Relation 14 states the basic equation of the URM method.

$$D(x,n) = \sum_{t=1}^{n} \sum_{j=1}^{NS} \beta_x(x,j,n-t+1).P(j,t) \quad (15)$$

In which, $D(x,n)$ is the drawdown at node x at the end $n^{th}$of the time period, $\beta_x(x,j,n-t+1)$ or unit response coefficient is the change of water table in node x at the end of n^{th} the time period with unit stimuli at the node j at the end of t^{th} time period, $P(j,t)$ is the amount of stimuli at node j, and time period t and NS is the total number of stimuli nodes.

River–Aquifer Interactions:

$$Q_{ar}^{riv}(r,t) = Criv(r).\left(h_{riv}^{s}(r,t) - h_{riv}^{g}(r,t)\right) \quad \text{if } h_{riv}^{g}(r,t) > h_{riv}^{bot}(r);\ \forall r,t \quad (16)$$

$$Q_{ar}^{riv}(r,t) = Criv(r).\left(h_{riv}^{s}(r,t) - h_{riv}^{bot}(r)\right) \quad \text{if } h_{riv}^{g}(r,t) > h_{riv}^{bot}(r);\ \forall r,t \quad (17)$$

In which $Criv(r) =$ the hydraulic conductance of the stream-aquifer interconnection which is a function of semi pervious streambed hydraulic conductivity, length, width and thickness of river reach r; $h_{riv}^{s}(r,t) =$ hydraulic head in the river reach r; $h_{riv}^{bot}(r) =$ elevation of semi pervious streambed bottom; $h_{riv}^{g}(r,t) =$ elevation of the aquifer water table below river reach r.

Sustainable Abstraction:

$$S\,(NT+1) \geq S\,(1) \quad (18)$$

$$S^{g}\,(NT+1) \geq S^{g}\,(1) \quad (19)$$

2.4. M5 Model Tree

Quinlan first proposed the M5 tree model in 1992 based on the tree classification method to establish the relationship between independent and dependent variables. Unlike many decision tree model algorithms, which are used for qualitative data, this model can be used for both qualitative and quantitative data.

The M5 model is similar to isolated linear functions, which is a combination of linear regression and tree regression models and is widely used in various sciences. The regression model provides a regression equation for the entire data space, but in the tree regression model, the data range is divided into sub-areas called leaves, and each leaf is given a numerical label. Replacing the linear regression equation with a label in nodes is a method used in the M5 model that can predict or estimate continuous numerical variables (Nourani et al., 2019a, 2019b).

In the M5 model, two branches branch from each parent node. The construction of the decision tree model is done in two steps. In the first stage, the decision tree is formed by branching the data. The branching criterion in the M5 model is to maximize the reduction of the standard deviation of the data in the child node. When it is not possible to reduce the standard deviation of the child node data, the parent node does not branch and does not reach the final node or leaf. This division often produces a large tree-like structure that may cause overfitting (Nourani & Molajou 2017). Consequently, the tree must be pruned back. So, the second stage would involve pruning the overgrown tree and replacing the subtrees with linear regression functions. This technique of generating the model tree splits the parameter space into subspaces and builds a linear regression model in each of them (Nourani et al., 2019b).

3. Results and Discussion

3.1. Stochastic Streamflow Generation

Classic black box models such as Autoregressive Integrated Moving Average (ARIMA) are widely used to predict hydrological time series. The essence of this model is linear and is based on the assumption that the data is static. This model has the ability to identify complex patterns in the data and provides the possibility of predicting the future in accordance with the input data of the past (Valipour, 2015; Molajou et al., 2021). ARIMA (p,d,q) has three trend elements: i) p: trend autoregression order ii) d: trend difference order iii) q: trend moving average order.

The Seasonal-ARIMA (SARIMA) model is also one of the classic black box models and is the general version of the ARIMA model. The SARIMA model, unlike the ARIMA model, also involves the seasonality of the data in modeling (Molajou et al., 2021). SARIMA (p,d,q)(P,D,Q) has three seasonal elements that are not part of ARIMA (p,d,q): i) P: Seasonal autoregressive order ii) D: Seasonal difference order iii) Q: Seasonal moving average order. In the current study, the autocorrelation function (ACF) plot is used to identify the time series data structure (see Figure 2).

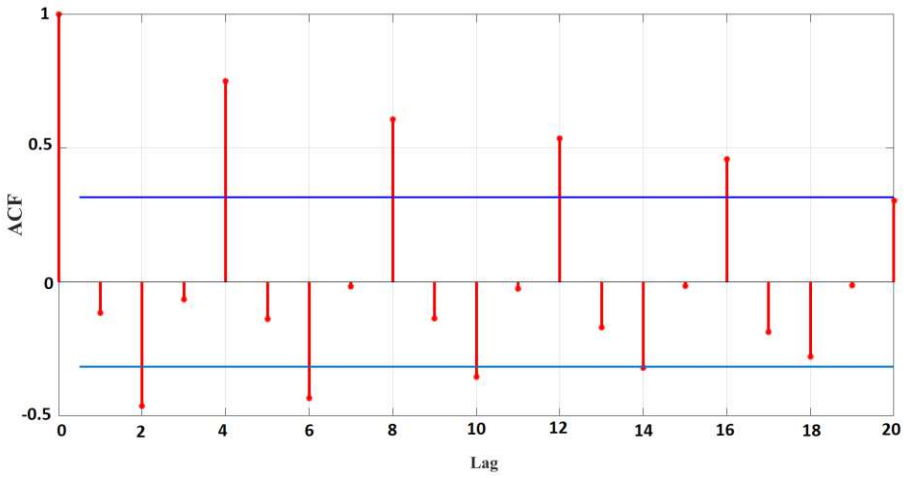


Fig. 2. The autocorrelation function (ACF) plot for historical time series of river flow

3.2. Operation Rules

This section presents the operation rules for the different subsystems of the CUS. In order to apply uncertainty, the concepts of the ISO method are used. The steps for generating operation rules of the CUS are as follows:

(i) Generation of 50 time series of river flow using an autoregressive model. Based on the data, the best autoregressive model, SARIMA, was selected.

(ii) Implementation of the operation optimization model for the time series produced in the first step considering capacities for various subsystems of the CUS (see Table 4).

Table 4. Known capacities of subsystems

Subsystem	Capacity
Dam	10.5 MCM
Water transfer from the dam to demand area	2.9 MCM/season
Water diversion from the river to demand area	1.8 MCM/season
Water diversion from the river to the artificial recharge area	1.2 MCM/season

(iii) Fitting the multi-regression and single regression models to 70% of optimization model outputs obtained from the second step.

It should be noted that 30% of the data is used to verify operation rules. Available SW ($S(t) + Q^s(t)$) and available GW ($S^g(t)$) are considered as independent variables in single-regression and multi-regression models. Five efficiency criteria are used to evaluate the performance of the operating rules, which are: Correlation coefficient (CC), Mean absolute error (MAE), Root mean squared error (RMSE), Relative absolute error (RAE), Root relative squared error (RRSE). The multi-regression model is fitted to the output data using the M5 algorithm of the decision tree. Relation 20 shows the fundamental relationship between dependent variables

238 ($R_b^a(t)$) and independent variables in single and multiple regression models. In which $R_b^a(t) =$

239 water transmission from element a to element b in time step t.

$$R_b^a(t) = a * (S(t) + Q^s(t)) + b * S^g(t) + C \quad (20)$$

240

241 By implementing the three steps mentioned above, the results for the single regression

242 model are presented in Tables 5 and 6. Table 5 shows that $Div_d^{riv}(t)$ has no dependence on the

243 volume of water stored in the aquifer. Table 6 shows that the fit of the single linear regression

244 model is not very accurate except for the variable $R_{riv}^s(t)$. In addition, the predicting error of

245 variable $Div_{ar}^{riv}(t)$ in the single-regression method is significant. Given that 50-time series, both

246 low water and high water, have been generated to apply uncertainty in the future, it seems that the

247 high-efficiency criterion in the multi-regression method is due to the classification of independent

248 and dependent data.

249 **Table 5.** Coefficients of operation rules in the single regression model

Decision variable	a	b	c
$\mathbf{Div_{ar}^{riv}}$	0.0121	-0.0136	4.0949
$\mathbf{Div_{d}^{riv}}$	0.0325	0.0000	0.8686
$\mathbf{R_d^g}$	0.0955	0.1830	-40.4194
$\mathbf{R_d^s}$	0.0541	0.0246	-4.7780
$\mathbf{R_{riv}^s}$	0.7030	-0.0737	14.1188

250

251 **Table 6.** Prediction errors of single regression model for training and verification data

Decision variable		CC	MAE	RMSE	RAE(%)	RRSE(%)
$\mathbf{Div_{ar}^{riv}}$	**train**	0.3092	0.2179	0.3339	99.46	95.10
	verify	0.2774	0.2161	0.3322	100.11	96.11
	train	0.4162	0.4955	0.5825	91.30	90.93

$\mathbf{Div_d^{riv}}$	**verify**	0.4134	0.508	0.5921	91.76	91.10
$\mathbf{R_d^g}$	**train**	0.3912	2.6975	2.9768	86.21	92.03
	verify	0.3994	2.6946	2.9757	85.85	91.69
$\mathbf{R_d^s}$	**train**	0.408	0.9559	1.0693	87.46	91.30
	verify	0.4199	0.9437	1.0602	86.73	90.81
$\mathbf{R_{riv}^s}$	**train**	0.8979	2.2303	2.7902	57.59	44.02
	verify	0.8795	2.2473	2.7893	60.56	47.71

By fitting the multi regression model to the variables $\mathrm{Div}_{ar}^{riv}(t)$, $\mathrm{Div}_{d}^{riv}(t)$, $\mathrm{R}_{d}^{g}(t)$, $\mathrm{R}_{d}^{s}(t)$, and $\mathrm{R}_{riv}^{s}(t)$ the 14, 19, 31, 22, and 28 conditional operation rules (M5Rules) were obtained, respectively. For example, the first two operating rules for the variable $\mathrm{Div}_{ar}^{riv}(t)$ were obtained as follows:

$$\text{Rule1: If } S(t) + Q^{s}(t) > 14.631$$
$$\text{Then } \mathrm{Div}_{ar}^{riv}(t) = 0.0005 \times \left(S(t) + Q^{s}(t)\right) - 0.0005 \times S^{g}(t) + 1.3258 \quad (21)$$

$$\text{Rule2: If } S^{g}(t) \leq 228.978$$
$$\text{Then } \mathrm{Div}_{ar}^{riv}(t) = 0.0027 \times \left(S(t) + Q^{s}(t)\right) - 0.0008 \times S^{g}(t) + 1.3747 \quad (22)$$

Given that the M5 decision tree algorithm classifies the environment of decision variables and presents a regression model in each class, it is obvious that the rules obtained from this method will be better than the single-regression method. The results of the decision tree method for decision variables are showed in Table 7. As can be seen in Table 7, the errors of the multi-regression method are significantly reduced compared to the single-regression method. For example, in the multi-regression method, the value of the efficiency criteria of CC (for training data) for the variable $\mathrm{Div}_{ar}^{riv}(t)$ is improved by about 60%. By comparing Tables 6 and 7, it can be

seen that the least improvement is for variable $R^{s}_{riv}(t)$. A comparison of the actual and predicted values of both models for this variable is shown in Figure 3.

Table 7. Prediction errors of multi regression model for training and verification data

Decision variable		CC	MAE	RMSE	RAE(%)	RRSE(%)
Div^{riv}_{ar}	**train**	0.9035	0.0614	0.1518	28.01	43.22
	verify	0.7859	0.0801	0.2153	37.08	62.29
Div^{riv}_{d}	**train**	0.7841	0.2606	0.399	48.03	62.29
	verify	0.7003	0.3096	0.4642	55.94	71.42
R^{g}_{d}	**train**	0.9317	0.7838	1.2359	25.05	38.21
	verify	0.7862	1.1747	2.0067	37.43	61.83
R^{s}_{d}	**train**	0.8908	0.3931	0.5446	35.96	46.50
	verify	0.7672	0.5066	0.7515	46.56	64.37
R^{s}_{riv}	**train**	0.9854	0.7145	1.0916	18.45	17.22
	verify	0.935	1.121	2.0797	30.21	35.57

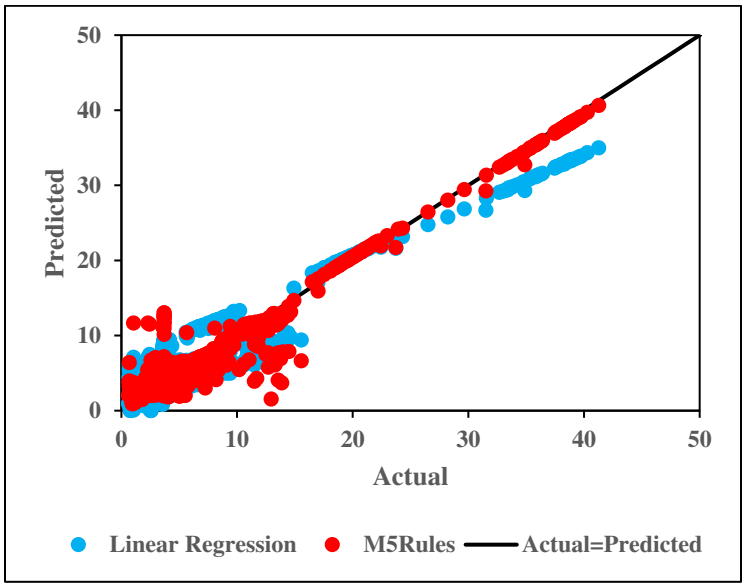


Fig.3. Actual and predicted values for decision variable R^{s}_{riv}

3.3. Water Supply to the Demand Area

The purpose of this section is to show how reliable the multi-regression operation rules for allocating water to the demand area. Figure 4 shows actual and predicted water shortage values for 50 river flow series. The actual values are obtained from the optimization model output for each inflow to the SW reservoir. In fact, by implementing the optimization model for each of the 50 river flow series, the objective function is shown in Figure 4. The predicted values have been obtained from the output of M5rules for each river flow. As can be seen in Figure 4, the minimum and maximum water shortage for actual data are 0 and 11.725 MCM, respectively, and the minimum and maximum water shortage for predicted data are 0 and 25.908, respectively.

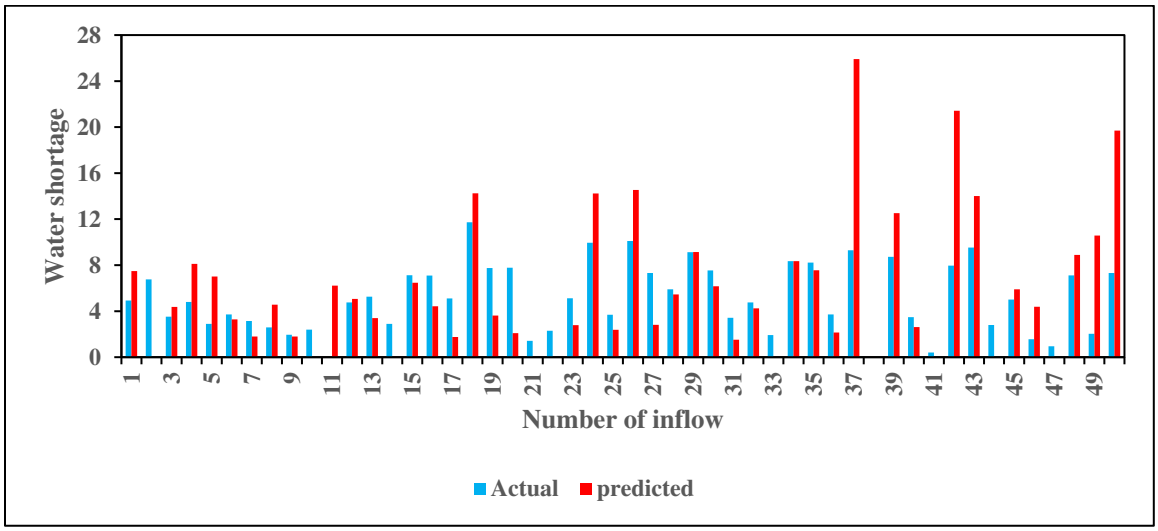


Fig. 4. Actual and predicted values of water shortage at the end of 10-years planning horizon for 50 river flow series

The actual and predicted values of water shortage for river flow No. 37 are 9.299 and 25.908 MCM, respectively. In other words, if we assume that flow No. 37 occurs in reality, with the allocation of water resources using the M5rules, the amount of water shortage will be 25.908 MCM during the 10-year planning period. This is while it can be done in such a way that water shortage is reduced by 9.299 MCM. This means that the M5rules impose about 16.609 MCM of

water shortage more than the optimal management plan. Given that the maximum difference between actual and predicted data has occurred for flow No. 37, it can be concluded that the maximum cost of M5rules will be the same (i.e., 16.609 MCM).

For river flow No. 2, the conditions are quite the opposite of river flow No. 37. In river flow No. 2, unlike river flow No. 37, the actual water shortage is greater than the predicted water shortage. The actual and predicted values of water shortage for river flow No. 2 are 6.763 and 0 MCM, respectively. It is logical that there will never be operation rules that satisfy all the constraints of the optimization model and perform better than the optimal management plan (outputs from the optimization model). We attribute it to the model error. If you pay attention to constraints 18 and 19, you will notice that the operation of SW and GW resources is done in such a way that the amount of water stored in SW and GW reservoirs at the end of the simulation period is equal to their initial. In fact, these two reservoirs will be operated sustainably. The fact that acting according to the M5rules can reduce the amount of water shortage by 6.763 MCM indicates that at the end of the10-year period, the total water stored in SW and GW reservoirs will be 6.763 million MCM less than their initial amount. It is important to know that among the 50 river flow series, the maximum error of the multi-regression model is for river flow No. 2. One of the notable points of Figure 4 is that in 28 cases, the water allocation conditions are the same as river flow No. 2 and the average model errors for all 28 flows are 2.06 MCM. Therefore, it can be said with reasonable accuracy that the proposed operating rules, in addition to guaranteeing the water allocation sustainability, also guarantee the SW and GW resources sustainability. In addition, in 22 cases, the water allocation conditions are the same as river flow No. 37, and the average cost of M5rules for all 22 flows are 4.33 MCM. Therefore, it seems that, in general, the proposed multi-

312 regression operation rules have a high performance, and the water allocation to the demand area
313 based on these rules can be suggested.
314
315 **4. Conclusion**
316 In this article, a simulation-optimization model is developed to extract the operating rules from a
317 distributed CUS. The operation optimization model of CUS tries to minimize water shortage while
318 satisfying physical, environmental, and social constraints. The single and multi-linear regression
319 operation rules of the SW reservoir, pumping from the GW reservoir, aquifer artificial recharge,
320 and water diversion from the river to the demand area have been extracted as a function of the
321 available SW and GW. Based on surveys, the SARIMA statistical model is the best autoregressive
322 model for river flow prediction. In order to take into account SW flow uncertainties in the
323 development of rules, 50 time series of artificial river flow were first generated using the SARIMA
324 model. Then, the operating rules were extracted from the fit of single and multi-linear regression
325 models to the outputs of the optimization model. The M5 tree algorithm is used to derive the
326 conditional operation rules. The results show the remarkable efficiency of the decision tree
327 algorithm in extracting operating rules. For example, the CCs (for training data) for the aquifer
328 artificial recharge, water diversion from the river to demand area, GW pumping, water transfer
329 from the reservoir to demand area, and release from the SW reservoir to the river improved by
330 59%, 37%, 54%, 48%, and 9%, respectively. In the single regression method, the CC was obtained
331 desirable only for release from the dam, and for other variables, this coefficient was less than 0.5.
332 Conditional operation rules ensure the sustainability of SW and GW water resources and
333 the sustainability of water allocation to the demand site. The maximum cost of conditional
334 operation rules was obtained 16.61 MCM of water shortage. Considering that the average cost of

335 M5rules was estimated to be about 4 MCM during the ten years of the planning horizon, the use
336 of conditional operation rules is highly recommended.

337 The present study can be improved by considering other existing uncertainties such as
338 rainfall, aquifer hydraulic characteristics, climate change and water demands.

339

340 **Declarations**

341 **Funding**

342 This work was supported by Iran National Science Foundation (INSF), grant number :96010175.

343 **Conflicts of interest/Competing interests**

344 There is no conflict of interest.

345 **Ethics approval**

346 Not applicable

347 **Consent to participate**

348 Not applicable

349 **Consent for publication**

350 The authors give their full consent for the publication of this manuscript.

351 **Availability of data and material**

352 Not applicable

353 **Code availability**

Not applicable

Authors' contributions

Mina Khosravi: Conceptualization, Methodology, Writing - original draft, Software.

Abbas Afshar: Conceptualization, Methodology, Supervision, Funding acquisition.

Amir Molajou: Conceptualization, Methodology, Software, Writing - original draf.

Acknowledgment

Authors appreciate Iran National Science Foundation (INSF) for their partial support through grant number 96010175.

Reference

Afshar A, Khosravi M, Molajou A (2021) Assessing adaptability of cyclic and non-cyclic approach to conjunctive use of groundwater and surface water for sustainable management plans under climate change. Water Resour Manag 1-17. https://doi.org/10.1007/s11269-021-02887-3

Afshar A, Khosravi M, Ostadrahimi L, Afshar A (2020) Reliability-based multi-objective optimum design of nonlinear conjunctive use problem; cyclic storage system approach. J Hydrol 588: 125109. https://doi.org/10.1016/j.jhydrol.2020.125109

Afshar A, Ostadrahimi L, Ardeshir A, Alimohammadi S (2008) Lumped approach to a multi-period–multi-reservoir cyclic storage system optimization. Water Resour Manag 22(12):1741-1760. https://doi.org/10.1007/s11269-008-9251-y

373 Afshar A, Zahraei A, Mariño MA (2010) Large-scale nonlinear conjunctive use optimization
374 problem: decomposition algorithm. J Water Resour Plan Manag 136(1):59-71.
375 https://doi.org/10.1061/(ASCE)0733-9496(2010)136:1(59)

376 Buras N (1963) Conjunctive operation of dams and aquifers. J Hydraul Div 89(6):111-131.
377 https://doi.org/10.1061/JYCEAJ.0000949

378 Chang LC, Chu HJ, Chen YW (2013) A fuzzy inference system for the conjunctive use of surface
379 and subsurface water. Adv Fuzzy Syst. https://doi.org/10.1155/2013/128393

380 Draper AJ (2001) Implicit stochastic optimization with limited foresight for reservoir systems.
381 University of California, Davis PhD dissertation

382 Fallah-Mehdipour E, Bozorg Haddad O, Mariño MA (2013) Extraction of optimal operation rules
383 in an aquifer-dam system: genetic programming approach. J Irrig Drain Eng 139(10):872-879.
384 https://doi.org/10.1061/(ASCE)IR.1943-4774.0000628

385 Fallah-Mehdipour E, Haddad OB, Alimohammadi S, Loáiciga HA (2015) Development of real-
386 time conjunctive use operation RULES for aquifer-reservoir systems. Water Resour Manag
387 29(6):1887–1906. https://doi.org/10.1007/s11269-015-0917-y

388 Heydari F, Saghafian B, Delavar M (2016) Coupled quantity-quality simulation-optimization
389 model for conjunctive surface-groundwater use. Water Resour Manag 30(12):4381-4397.
390 https://doi.org/10.1007/s11269-016-1426-3

391 Kerebih MS, Keshari AK (2021) Distributed simulation-optimization model for conjunctive use
392 of groundwater and surface water under environmental and sustainability restrictions. Water
393 Resour Manag 1-19. https://doi.org/10.1007/s11269-021-02788-5

394 Khosravi M, Afshar A, Molajou A (2021) Reliability-based design of conjunctive use water
395 resources systems: comparison of cyclic and non-cyclic approaches. Journal of Water Wastewater
396 31(7):90-101. https://dx.doi.org/10.22093/wwj.2020.201234.2924 (in Farsi).

397 Loucks DP, Stedinger JR, Haith DA (1981) Water resource systems planning and analysis.
398 Prentice-Hall

399 Milan SG, Roozbahani A, Banihabib ME (2018) Fuzzy optimization model and fuzzy inference
400 system for conjunctive use of surface and groundwater resources. J Hydrol 566:421-434.
401 https://doi.org/10.1016/j.jhydrol.2018.08.078

402 Nayak MA, Herman JD, Steinschneider S (2018) Balancing flood risk and water supply in
403 California: policy search integrating short-term forecast ensembles with conjunctive use. Water
404 Resour Res 54(10):7557-7576. https://doi.org/10.1029/2018WR023177

405 Nourani V, Molajou A (2017) Application of a hybrid association rules/decision tree model for
406 drought monitoring. Glob Planet Change 159:37-45.
407 https://doi.org/10.1016/j.gloplacha.2017.10.008

408 Nourani V, Razzaghzadeh Z, Baghanam AH, Molajou A (2019a) ANN-based statistical
409 downscaling of climatic parameters using decision tree predictor screening
410 method. Theor Appl Climatol 137(3):1729-1746. https://doi.org/10.1007/s00704-018-2686-z

411 Nourani V, Tajbakhsh AD, Molajou A (2019b) Data mining based on wavelet and decision tree
412 for rainfall-runoff simulation. Hydrol Res 50(1):75-84. https://doi.org/10.2166/nh.2018.049

413 Nourani V, Rouzegari N, Molajou A, Baghanam AH (2020) An integrated simulation-optimization
414 framework to optimize the reservoir operation adapted to climate change scenarios. J Hydrol

587:125018. https://doi.org/10.1016/j.jhydrol.2020.125018

Molajou, A., Nourani, V., Afshar, A., Khosravi, M., & Brysiewicz, A. (2021). Optimal Design and Feature Selection by Genetic Algorithm for Emotional Artificial Neural Network (EANN) in Rainfall-Runoff Modeling. Water Resources Management, 35: 2369–2384. https://doi.org/10.1007/s11269-021-02818-2

Onta PR, Gupta AD, Harboe R (1991) Multistep planning model for conjunctive use of surface- and ground-water resources. J Water Resour Plan Manag 117(6):662-678. https://doi.org/10.1061/(ASCE)0733-9496(1991)117:6(662)

Pan CC, Chen YW, Chang LC, Huang CW (2016) Developing a conjunctive use optimization model for allocating surface and subsurface water in an off-stream artificial lake system. Water 8(8):315. https://doi.org/10.3390/w8080315

Philbrick CR, Kitanidis PK (1998) Optimal conjunctive-use operations and plans. Water Resour Res 34(5):1307-1316. https://doi.org/10.1029/98WR00258

Quinlan JR (1992, November) Learning with continuous classes. In 5th Australian joint conference on artificial intelligence (Vol. 92, pp. 343-348)

Rafipour-Langeroudi M, Kerachian R, Bazargan-Lari MR (2014) Developing operating rules for conjunctive use of surface and groundwater considering the water quality issues. KSCE J Civ Eng 18(2):454-461. https://doi.org/10.1007/s12205-014-1193-8

Safavi HR, Chakraei I, Kabiri-Samani A, Golmohammadi MH (2013) Optimal reservoir operation based on conjunctive use of surface water and groundwater using neuro-fuzzy systems. Water Resour Manag 27(12):4259-4275. https://doi.org/10.1007/s11269-013-0405-1

436 Schoups G, Addams CL, Minjares JL, Gorelick SM (2006) Reliable conjunctive use rules for
437 sustainable irrigated agriculture and reservoir spill control. Water Resour Res 42(12).
438 https://doi.org/10.1029/2006WR005007

439 Song J, Yang Y, Sun X, Lin J, Wu M, Wu J, Wu J (2020) Basin-scale multi-objective simulation-
440 optimization modeling for conjunctive use of surface water and groundwater in northwest
441 China. Hydrol Earth Syst Sci 24(5):2323-2341. https://doi.org/10.5194/hess-24-2323-2020

442 Valipour M (2015) Long-term runoff study using SARIMA and ARIMA models in the United
443 States. Meteorol Appl 22(3):592-598. https://doi.org/10.1002/met.1491

444 Yang G, Guo S, Liu P, Li L, Liu Z (2017) Multiobjective cascade reservoir operation rules and
445 uncertainty analysis based on PA-DDS algorithm. J Water Resour Plan Manag 143(7):4017025.
446 https://doi.org/10.1061/(ASCE)WR.1943-5452.0000773

447 Zhao T, Zhao J, Lund JR, Yang D (2014) Optimal hedging rules for reservoir flood operation from
448 forecast uncertainties. J Water Resour Plan Manag 140(12):4014041.
449 https://doi.org/10.1061/(ASCE)WR.1943-5452.0000432

450