Title

Development of the FA-KNN hybrid algorithm and its application to reservoir operation

Journal

Theoretical and Applied Climatology, 155(2)

ISSN

0177-798X

Authors

Azadi, Firoozeh

Ashofteh, Parisa-Sadat

Shokri, Ashkan

et al.

Publication Date

2024-02-01

DOI

10.1007/s00704-023-04688-7

Peer reviewed

Theoretical and Applied Climatology (2024) 155:1261–1280
https://doi.org/10.1007/s00704-023-04688-7

RESEARCH

Development of the FA-KNN hybrid algorithm and its application to reservoir operation

Firoozeh Azadi[1] · Parisa-Sadat Ashofteh[1] · Ashkan Shokri[1,2] · Hugo A. Loáiciga[3]

Received: 10 July 2023 / Accepted: 24 September 2023 / Published online: 24 October 2023

Abstract

This study presents a method to address the issue of burdensome computations in water resources optimization based on a hybrid algorithm derived from the firefly algorithm (FA) and the K-nearest neighbor (KNN) algorithm, herein named the FA-KNN algorithm. The FA-KNN algorithm introduced in this work is tested with three standard test problems (the Ackley, Rosenbrock, and Sphere problems), and with a reservoir operation problem that minimizes the relative agricultural water-supply deficit under baseline and climate-change conditions. The efficiency indexes of the reservoir system are calculated to evaluate the performance of the FA-KNN algorithm and its accuracy. The results demonstrate the operational policy obtained with the FA-KNN algorithm has better performance in terms of computational burden than the FA's. This work's findings establish the FA-KNN hybrid algorithm reduces the computational time by 60% with acceptable accuracy compared with the FA algorithm. The findings indicate a reduction in run-time of 99.5, 94, and 92% for solving the Ackley, Rosenbrock, and Sphere test problems achieved with the FA-KNN algorithm while maintaining a high level of accuracy when contrasted with solutions derived from both deterministic methodologies and the FA approach. The volumetric reliability and flexibility in the reservoir problem calculated under the baseline conditions outperformed those obtained with the climate change conditions by 10 and 3.5%, respectively. Moreover, a notable discrepancy emerged in terms of the main simulator's invocation frequency between the FA-KNN and FA methods (the former exhibited a mere 0.3 ratio compared to the latter). The application of the FA-KNN approach yielded a reduction exceeding 60% in run-time for the reservoir problem.

1 Introduction

Johari et al. (2013) evaluated applications of the FA in optimization problems. The results of the evaluation show that the FA algorithm performs better than other meta-heuristic algorithms. Chou and Ngo (2017) proposed a modified firefly algorithm (MFA) algorithm to optimize the design of multidimensional structures and improve optimization capabilities by improving the search process and setting the attractiveness parameter. The results showed the superiority of MFA algorithm compared to FA and its higher convergence speed. Garousi-Nejad et al. (2016b) applied the firefly algorithm (FA) to optimal operation of reservoirs with the purpose of irrigation water supply and hydropower production. Their results demonstrated the superior performance of the FA compared to the genetic algorithm (GA) in terms of the convergence rate to global optima. Azizipour et al. (2016) applied the invasive weed optimization (IWO) algorithm to the optimal operation of hydropower reservoir systems. The results were compared with the results obtained with particle swarm optimization (PSO) and the GA. Their results established the IWO was more efficient and effective than PSO and the GA. SaberChenari et al. (2016) adopted the PSO for solving the operation of the multipurpose Mahabad reservoir in northwestern Iran. Hossain et al. (2018) applied and compared the artificial bee colony (ABC) algorithm, PSO, the GA, and neural

✉ Parisa-Sadat Ashofteh
PS.Ashofteh@qom.ac.ir

Firoozeh Azadi
f.azadi@stu.qom.ac.ir

Ashkan Shokri
Ashkan.Shokri@monash.edu

Hugo A. Loáiciga
hloaiciga@ucsb.edu

[1] Department of Civil Engineering, University of Qom, Qom, Iran

[2] Australian Bureau of Meteorology, Melbourne, Australia

[3] Department of Geography, University of California, Santa Barbara, CA 93016-4060, USA

network-based stochastic dynamic programming (SDP) in solving for optimal operation of the Aswan reservoir in Egypt. Their results indicated the water release policy calculated with the ABC algorithm outperformed the comparison algorithms. Garousi-Nejad et al. (2016a) reported a modified firefly algorithm (MFA) and applied it to solve reservoir operation problems. Three well-known benchmark multireservoir operation problems were optimized for energy production. The results of the MFA were compared with LP, differential dynamic programming (DDP), and discrete DDP (DDDP), the GA, the multi-colony ant algorithm (MCAA), the honey-bee mating optimization (HBMO) algorithm, the water cycle algorithm (WCA), the bat algorithm (BA), and the biogeography-based optimization (BBO) algorithm. The MFA was found to be more effective than alternative optimization methods. Ahmadianfar et al. (2017) extracted the optimal policies of hydropower multi-reservoir systems employing the enhanced differential evolution (EDE) algorithm. Azizipour and Afshar (2018) presented a novel hybrid GA and cellular automata (CA) method for solving reliability-based reservoir operation problems. The proposed method was implemented to calculate monthly water supply and hydropower generation at the Dez reservoir in Iran. Their results showed the proposed method was more accurate and efficient than the GA. Ahani et al. (2018) applied four common data-driven modeling techniques including multiple linear regression, KNN, artificial neural networks (ANN), and adaptive neuro-fuzzy inference systems to runoff forecasting. Their results indicated the selected KNN model yielded the best performance. Yaseen et al. (2018) introduced and applied the hybrid approach artificial fish swarm algorithm (AFSA) and applied the particle swarm optimization algorithm (PSOA) to optimize the Karun-4 reservoir operation for energy generation and minimizing downstream water shortages. Their results indicated the hybrid algorithm (HA) performed with higher reliability, lower vulnerability, and higher resiliency compared with the AFSA and the PSOA. Also, the HA was top ranked according to multi criteria decision making. Karami et al. (2018) introduced an improved version of the krill algorithm (KA) for reservoir operation. The KA converged faster to the near-optimal solution than PSO and the GA. The improved KA could meet 97% of irrigation demands and produced the lowest value of the vulnerability index among GA, PSO, and the simple KA. The average solution of the improved KA was closer to the global solution than those produced by the alternative optimization methods. Samadi-koucheksaraee et al. (2019) introduced the gradient evolution (GE) algorithm as an efficient solution for water resource management. Fang et al. (2021) proposed new optimization algorithms based on the accelerated gradient-based (AGBO) for multi-reservoir system. Ashofteh et al. (2021) developed and applied a bi-objective genetic programming (BO-GP) algorithm to optimize the operating rules of the Aidoghmoush reservoir (in Iran). Their results indicated a successful performance of the BO-GP algorithm in minimizing vulnerability and maximizing reliability of water supply. Hu et al. (2019) proposed an improved cloud adaptive quantum-inspired binary social spider optimization algorithm for short-term hydropower generation scheduling of three Gorges hydropower station. The results demonstrated the effectiveness of the proposed algorithm. Ahmadianfar et al. (2022) extracted the nonlinear operating rules of multi-reservoir systems using the self-adaptive teaching learning-based algorithm with differential evolution (SATLDE). Samadi-koucheksaraee et al. (2022) emphasized the significance of meta-heuristic and evolutionary algorithms in various engineering domains. Shirvani-Hosseini et al. (2022) applied data mining methods to modeling in water science. Ahmadianfar et al. (2023) calculated optimal operating rule curves for hydropower multi-reservoir systems by influential flower pollination algorithm (IFPA).

Complex optimization problems are commonly not solvable with classic methods, such as linear programming (LP), and are beset by burdensome computations, which in many instances fail to converge to optimal solutions. Meta-heuristic and evolutionary algorithms (MHEAs) were introduced to tackle these otherwise intractable optimization problems (see, e.g., Bozorg-Haddad et al. 2017). Experience has shown MHEAs have performed well in solving complex water resources problems.

Many hydrologic problems are beset by computational burden and dimensionality. This means the complexity of the problem is compounded by the computational burden and the amount of storage space needed to solve it. Several researchers have evaluated methods to overcome the complexities of hydrological models. For example, Kou et al. (2008) examined a new hybrid three-dimensional computational solution for groundwater solute transport in a layered aquifer. The three-dimensional solute transport was reduced to a hybrid three-dimensional model by combining the two-dimensional computations and the one-dimensional analytical solutions. The resultant hybrid three-dimensional method was capable of large computational time steps coarse spatial grids and was thus suitable for simulating solute transport in large-scale sites of layered aquifers. This method was verified by analytical solutions as well as by the numerical model MT3D. Schoups et al. (2008) compared three model complexity control methods, and their results indicated that simulation of water flow using non-physically-based models achieved better calibration. Jato-Espino et al. (2017) presented a simulation–optimization method to model urban catchment located in Espoo

(southern Finland) under non-stationary extreme rainfall events under climate change conditions. The results showed their approach achieved suitable calibration and prediction of extreme precipitation. Shokri et al. (2018) applied the patient rule induction method (PRIM) to detect hydrologic model behavioral parameters and quantify uncertainty, and reduced the size of the decision space applying PRIM.

The criteria for choosing water resources models have relied on evaluating their technical traits, data requirements, and system complexity (Arkesteijn and Pande 2013). Höge et al. (2018) examined the role of model complexity in choosing a model among several ones that are often used in water resources, and reported a classification procedure for the correct interpretation of complexity and model choice. Li and Yue (2018) proposed an inverse method to calibrate the spatially and temporally varying Root Density Distribution Function (RDDF) in unsaturated water flow modeling. Model calibration was formulated as an optimization problem in the framework of the Tikhonov regularization theory to overcome complexity, adding constraints to the objective function. Arkesteijn and Pande (2013) examined the complexity of hydrologic models and the effect of uncertainty on predictions. The latter authors presented an algorithm for assessing the complexity of hydrological models, and demonstrated that hydrological model complexity has a geometric interpretation in terms of the model's output space. Hong et al. (2003) accelerated the convergence of evolutionary algorithms (EAs) by means of multi-layer artificial neural networks (ANNs). The proposed method was verified by a numerical example. Shokri et al. (2013) proposed a method based on ANNs to reduce the number of simulations required by EAs. The performance of the proposed method was examined by integrating it with the non-dominated sorting genetic algorithm (NSGAII) in multi-objective problems. Their results demonstrated that use of the NSGAII-ANN hybrid algorithm reduced the required time for optimization up to 50 times compared with the standard NSGAII. Pseudo models are employed in the statistical approximation and uncertainty analyses of time-consuming and complex models. The pseudo models apply emulators that yield solutions close to those obtained with the implementation of the main model simulators. Moreno-Rodenas et al. (2018) implemented an emulator as a substitute for the main model for hydro-dynamic simulations under the influence of precipitation and parametric scenarios. Chong et al. (2021) reported a thorough review of optimization application problems (OAPs) in engineering and scientific fields, such as economic dispatch, structural design, and water resources. Lai et al. (2022a) provided a comprehensive review of reservoir operation optimization, highlighting its benefits in terms of efficient energy production, flood prevention, cost reduction, and addressing water scarcity. Almubaidin et al. (2022) reviewed reservoir operation optimization focusing on the intricacies posed by non-linearities, complex constraints, and multiple variables. They examined various metaheuristic algorithms (MHAs) for minimizing water deficits and aiding decision-making in reservoir operation. Lai et al. (2022b) reported a study focused on the KLang Gate Dam (KGD) and its operation under changing climate scenarios using sim-heuristic techniques. The latter authors explored the impact of climate change on reservoir water resources by considering an ensemble of general circulation models (GCMs) from the CMIP5 dataset.

The FA was introduced by Yang (2008). Many studies have proved the high efficiency of the FA algorithm and improved the algorithm, as demonstrated by Yang (2008, 2010, 2011), Hassanzadeh et al. (2011), Yan et al. (2012), Nandy et al. (2012), Afnizanfaizal et al. (2012), Silva et al. (2013), Garousi-Nejad et al. (2016a, b), among others. Many simulation–optimization processes used to solve complex problems involve a heavy computational burden. Orabona et al. (2010) introduced the online independent support vector machine (SVM). The SVM method has been widely applied in various fields because of its high computational efficiency. However, SVM cannot solve online classification problems effectively due to the involved computational burden of this process (Wang et al 2013). The standard SVM and most of the modified SVMs are in essence batch learning. Such SVMs cannot process large-scale data effectively because they are costly in terms of memory and computing effort. Under some conditions support vectors (SVs) are generated, which generally means a long testing time. The training of traditional SVM is time and storage-space consuming because the SVM training usually is posed as a quadratic programming problem (Zheng et al 2013).

Some water resource management problems involving slow simulator models and long-term studies such as climate change studies are beset by long run time in the optimization process. This work overcomes the time complexity of models and improves the run-time in a dynamic and online simulation–optimization process by developing the hybrid firefly K-nearest neighbor algorithm (FA-KNN).

The FA-KNN algorithm is herein introduced to accelerate the solution of complex problems. The FA-KNN algorithm features two operators known as "oblivion" and "sampling" that render it comparatively efficient in solving complex problems. The FA-KNN is tested, and its performance evaluated with the Ackley, Rosenbrock, Sphere standard test problems and with a water resources problem that minimizes the relative deficit in supplying water demand under baseline and climate-change conditions by the Gharanghu reservoir in East Azerbaijan, Iran. The FA and KNN algorithms are

described next, followed by a description of the methodology herein employed leading to the hybrid FA-KNN algorithm and a discussion of its properties. In addition, this work evaluates the application of the FA-KNN algorithm with a water resource management problem.

2 Methodology

The firefly algorithm is inspired by the behavior of fireflies (Yang 2008). The FA rests on three assumptions: (1) the fireflies are mono-sexual and sexuality does not affect inter-firefly attraction; (2) the degree of attractiveness between any two fireflies is governed by their brightness and is inversely related to their separation distance; and (3) the brightness of a firefly is determined by its objective function. The parameters used in the algorithm, including the light absorption coefficient (γ), the initial value of the absorption coefficient (β_0), and the random parameter (α) are specified according to the range given by Yang (2008).

An initial population of fireflies is created in the FA. Each firefly moves towards a more attractive firefly. Light intensity (representing the value of the objective function) is proportional to the inverse of the square of the distance from fireflies:

$$I = \frac{I_0}{r^2} \tag{1}$$

in which, I_0 = light intensity at the source and I = light intensity at a distance r from the light source.

Equation (1) has a singularity at $r = 0$. To this singularity another equation has been defined for light intensity:

$$I = I_0 e^{-\gamma \cdot r^2} \tag{2}$$

in which, γ = light absorption coefficient.

Equation (3) is obtained for light intensity using a truncated Taylor series in Eq. (2):

$$I \approx \frac{I_0}{1 + \gamma \cdot r^2} \tag{3}$$

Similarly, the equation of firefly attractiveness is defined as follows:

$$\beta = \beta_0 e^{-\gamma \cdot r^m} \approx \frac{\beta_0}{1 + \gamma \cdot r^m} \tag{4}$$

in which, β_0 = the initial magnitude of attractiveness, β = attractiveness at a distance r; m is a coefficient whose value ranges between zero and two.

The attractiveness approaches a constant value as the m coefficient tends to zero, in which case the light source acts as a surface light source (with infinite surface area). If m equals one, the light source acts as a linear light source that is proportional to the inverse of the distance. If m equals two, it acts as a point light source that corresponds to the inverse of the square of the distance.

When firefly i moves towards firefly j (provided $I_i < I_j$), the new position of firefly i ($newX_i$) is calculated with the following equation:

$$newX_i = X_i + \beta(X_j - X_i) \tag{5}$$

in which X_i and X_j denote the current positions of i-th and j-th fireflies, respectively. These positions are 3-dimensional vectors.

Fireflies randomly move towards more attractive fireflies, and their path is not always straight; therefore, there is randomness in the way fireflies move towards the more attractive firefly. This is captured with Eq. (6):

$$\text{new}X_i = X_i + \beta(X_j - X_i) + \alpha\varepsilon_i \tag{6}$$

in which $\alpha\varepsilon_i$ = a random vector. During the optimization search $\alpha\varepsilon_i$ $\alpha\varepsilon_i$ is adjusted to assure convergence to a solution. The process of optimizing, correcting the position of fireflies, and moving towards a more attractive firefly continues until the stopping criterion is met.

Figure 1 shows the flowchart of the optimization process for the firefly algorithm. The convergence speed of the FA is increased in this work by hybridizing it with the KNN algorithm. The KNN is a relatively simple, yet, very effective tool in data mining and machine learning. In fact, the KNN algorithm does not learn; rather, it remembers data that are used for classification and regression problems. Figure 2 provides a generic graphic comparison of the FA-KNN algorithm (Fig. 2(a)) and of the FA (Fig. 2(b)). It is seen in Fig. 2 that when using the FA algorithm alone (Fig. 2(b)), all simulations are performed by the main simulator. On the other hand, the FA-KNN algorithm (Fig. 2(a)) adds the "decision" section to the algorithm that chooses between the main or hybrid simulators based on predetermined conditions that increase the efficiency and accuracy of the algorithm, whose run time is decreased by reducing the number of main simulator calls.

The "main simulator" is the part of the model that is directly involved in calculating the value of the objective function. In other words, the parameters in the objective function formula are calculated by the main simulator. This simulator is a set of equations and codes or hydrological model simulation software that are interfaced with MATLAB to participate in the optimization process.

The execution time in the simulation process is optimized by the main simulator. This study applies a hybrid simulator to reduce the model's execution time. The "hybrid simulator" has features that are described in detail in the rest of the article (including the use of KNN, the forgetting index,

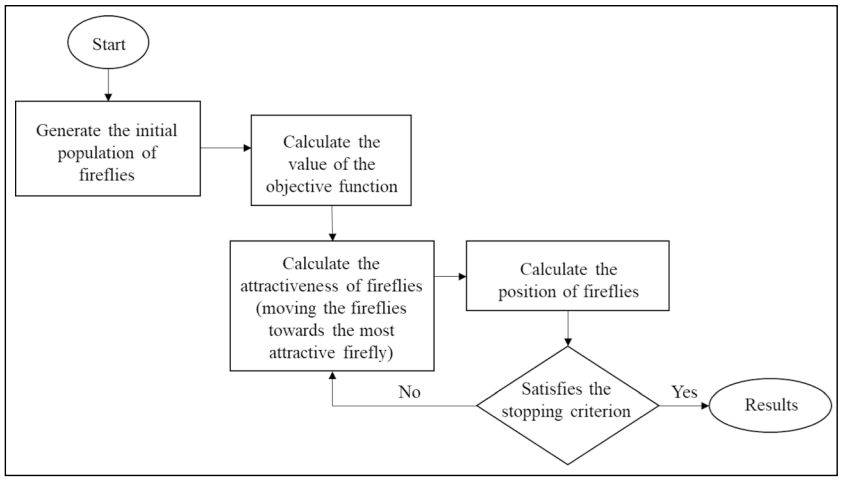


Fig. 1 Flowchart of the firefly algorithm (FA)

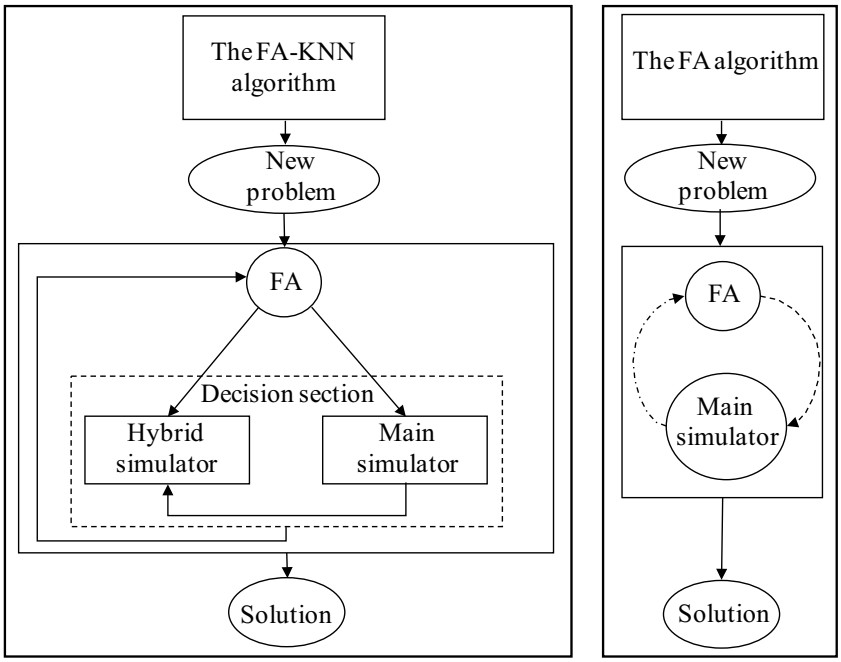


Fig. 2 Simplified flowcharts of **a** the proposed method and **b** the classic method

and the sampling index), which reduces the number of calls to the main simulator and the computational burden in the optimization.

2.1 The FA-KNN hybrid algorithm

A set of solutions simulated with the main FA is stored in a database called the information archive. The KNN measures proximity using standard Euclidean distances between input data (randomly generated possible solutions) and the data in the information archive. FA-KNN solutions are calculated with the weighted averaging (distance-weighted KNN) method. Figure 3 illustrates the flowchart of the FA-KNN algorithm proposed in this work. The initial population of fireflies (possible solutions) is randomly generated within the range of the decision variables. This constitutes the input data. A database named the information archive stores the values of the decision variables and the values of the corresponding objective function. The first step of the FA-KNN calls the FA (original), and its results are saved in the information archive. The next step modifies the initial population and creates a new population of fireflies. This step calculates the Euclidean distances between the data in the archive and the FA-KNN solutions in the input data. The ID of the nearest neighbors is calculated by the knnsearch function defined in MATLAB. An Euclidean distance equal to zero means the solution is already in the input data. Therefore, the value of the objective function [F(x)] corresponding to the decision

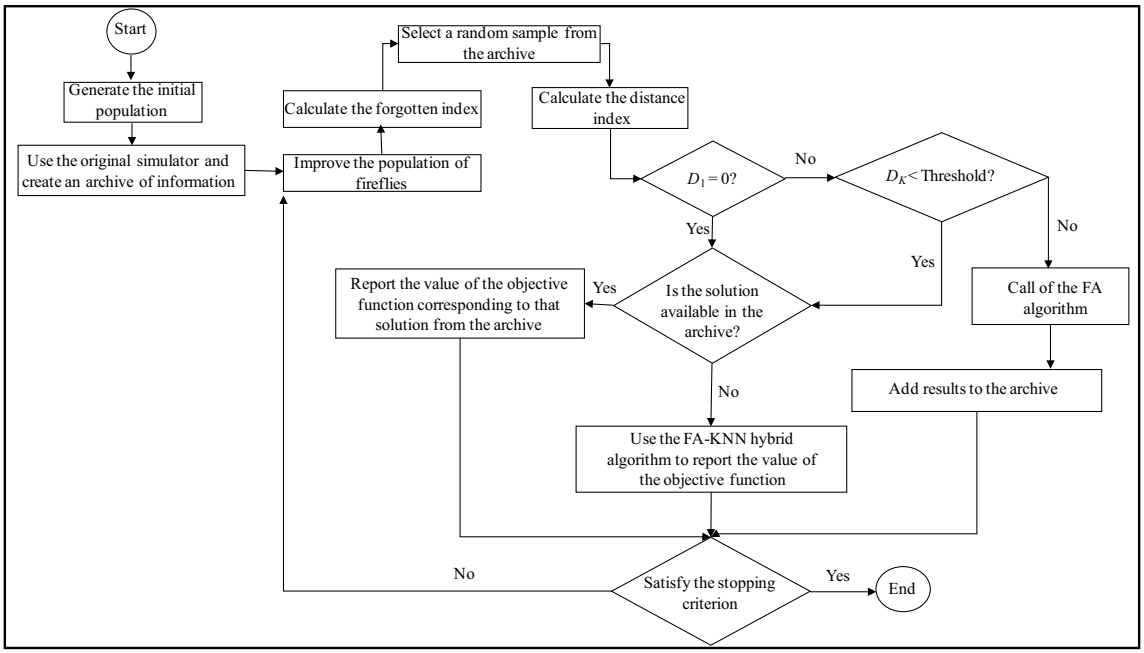


Fig. 3 Flowchart of the FA-KNN algorithm

variable is already stored in the archive. An Euclidean distance equal to less than a threshold value (that is sufficiently small relative to the decision variables) means the value of the objective function related to this decision variable is reported using the K-nearest solutions to the present solution. The FA is called, and its solution is added to the information archive whenever the Euclidean distance exceeds the threshold. All the data in the information archive are obtained with the FA. Data obtained with the FA-KNN algorithm are not entered in the archive. The threshold value is small enough to yield accurate results; yet, it must not be too small to avoid excessive computational time by the FA-KNN.

An "oblivion" index is added to increase the speed of the FA-KNN algorithm whereby the algorithm "forgets" (or ignores) some of the main data in the archive, and these data are deleted from the archive. The oblivion index removes the initial values from the archive to reduce the number of algorithmic comparisons thus increasing the convergence speed of the algorithm. In addition, when a database size reaches a specified size, a random sample with smaller size is selected from the archive. This process is named "sampling" in this paper. The sampling is performed using the "randperm" command. This reduces the number of comparisons carried out by the KNN without introducing virtually any error in the algorithmic results. The newly added "oblivio" operator causes the algorithm to forget some of the original data in the archive and removes these data from the archive. As the model convergences through consecutive iterations, it is no longer necessary to compare the input data with the original data in the archive because they are probably far from optimal. The oblivion index causes the model to remove the initial values from the archive when it is converging, thus reducing the number of comparisons and increasing the speed of the algorithm.

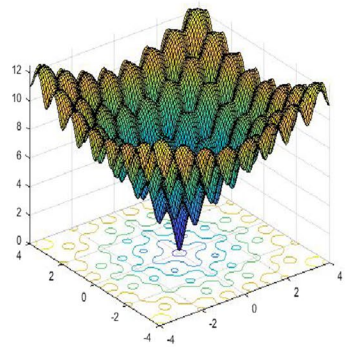
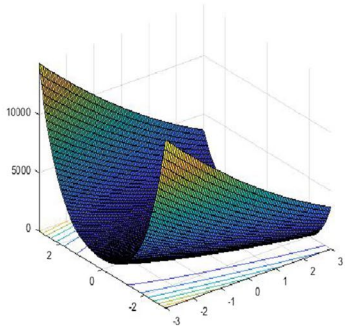
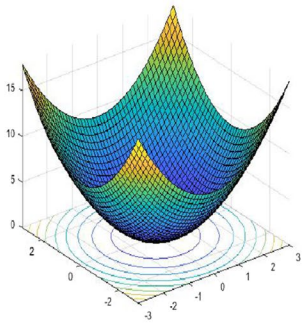
2.2 Evaluation of the FA-KNN algorithm

The FA-KNN algorithm was herein tested with three standard problems: the Ackley, Rosenbrock, and Sphere problems, and the results were compared with the FA's. Table 1 lists the characteristics of the test problems. The global optimal values of the test problems' objective functions are equal to zero, which are compared with those of the FA and the FA-KNN algorithm. The optimal value of a real-world optimization problem was obtained with the LINGO 12 software.

The initial solutions (initial populations of fireflies) are randomly generated. Therefore, each run of the algorithm produces a set of optimized decision variables and corresponding objective function value. The FA-KNN algorithm was run several times, and the optimal solutions were expressed as an average of the solutions obtained in the runs. This work implemented five independent runs. Test problems were solved in sets of runs listed in Table 2, in which the test problems were classified into groups and sets according to the solver chosen (the FA or the FA-KNN) and the algorithmic parameters implemented in each case.

The test problems were classified into groups I and II. The thresholds (i.e., stopping criteria) of groups I and II are respectively equal to 0.1 and 0.2. Each group has features

Table 1 Characteristics of the test functions

Ackley	Rosenbrock	Sphere
$Minimize\ f(x_1,\cdots,x_n)$ $= -20\exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2}\right)$ $-\exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i)\right)+20+e$ $-4 < x_i < 4 \quad i=1,2,\cdots,n$ 	$Minimize\ f(x_1,\cdots,x_n)$ $=\sum_{i=1}^{n-1}\left[100\left(x_{i+1}-x_i^2\right)^2+(x_i-1)^2\right]$ $-3 < x_i < 3 \quad i=1,2,\cdots,n$ 	$Minimize\ f(x_1,\cdots,x_n)$ $=\sum_{i=1}^{n}x_i^2$ $-3 < x_i < 3 \quad i=1,2,\cdots,n$ 
The Ackley function is one of the most widely used mathematical functions to evaluate the performance of algorithms that are applicable to multiple dimensions (several variables). The Ackley function consists of linear, exponential, and cosine functions, and the objective is to minimize it; its minimum is equal to f=0 at point (0, 0, 0, …0). The Ackley function has abundant local optimal, and its global minimum is found in the center of the curve.	The Rosenbrock function, also known as the Valley Function, and the Banana Function, is a single-peak function of the minimization. The absolute optimum of the Rosenbrock function is shaped inside a long valley, and its value is equal to f = 0 at point (1, 1, 1, …, 1). This test function is useful for assessing the performance of the optimization algorithms.	The Sphere function is a continuous, strongly convex, and single-peak (Unimodal), and multivariate function (i.e., has multiple decision variables) and therefore, almost all optimization algorithms find an absolute optimal solution for it. The Sphere function is a quadratic equation.

problems classified as A, B, C, D. Each problem in set A is solved by the FA algorithm. The same problem in set B is solved using the FA-KNN algorithm and with the stopping criterion equal to that of set A. In fact, all the parameters, including the number of iterations, the population of the fireflies, and other algorithm parameters, are the same for the sets A and B. The only difference is that set A uses the main simulator and set B uses the hybrid simulator. The sets C and D are exactly the same as set A and are solved with the FA, except that in set C the number of main simulator calls is equal to the average of the number of main simulator calls in five iterations of set B, and in set D, the number of main simulator calls is equal to the maximum number of main simulator calls in five iterations of set B. This is done by adjusting the population of the fireflies and the maximum number of iterations. The new data are compared to the data in the data bank, in which case the smaller the threshold, the greater the density around the new data; in this instance, the hybrid simulator is run by the algorithm. Otherwise, the main simulator runs. Therefore, the smaller the threshold, the higher the accuracy of the model and the longer the run time compared with that arising when a larger threshold is applied.

2.3 The reservoir operation problem

Gharanghu Dam is located 26 km southwest of the city of Hashtrood in East Azerbaijan Province, Iran (Fig. 4). It provides agricultural water and drinking water. The length of the Gharanghu River is 120 km and its average annual discharge is 149×10^6 m^3, and the area of the river basin equals 3590 km^2.The dam is built with compacted earth

Table 2 Characteristics of the test functions' optimization sets

Problem	Groups and sets		Algorithm	Threshold value	Stopping criterion (number of iterations)	Description
Ackley	I	A	FA	-	200	Populations = 15 Number of decision variables = 2 $\beta_0 = 1$ $\alpha = 0.1$ $\gamma = 0.01$
		B	FA-KNN	0.1	200	
		C	FA	-	-	
		D	FA	-	-	
	II	B	FA-KNN	0.2	200	
		C	FA	-	-	
		D	FA	-	-	
Rosenbrock	I	A	FA	-	200	Populations = 10 Number of decision variables = 2 $\beta_0 = 1$ $\alpha = 0.01$ $\gamma = 0.01$
		B	FA-KNN	0.1	200	
		C	FA	-	-	
		D	FA	-	-	
	II	B	FA-KNN	0.2	200	
		C	FA	-	-	
		D	FA	-	-	
Sphere	I	A	FA	-	200	Populations = 10 Number of decision variables = 2 $\beta_0 = 2$ $\alpha = 0.2$ $\gamma = 1$
		B	FA-KNN	0.1	200	
		C	FA	-	-	
		D	FA	-	-	
	II	B	FA-KNN	0.2	200	
		C	FA	-	-	
		D	FA	-	-	

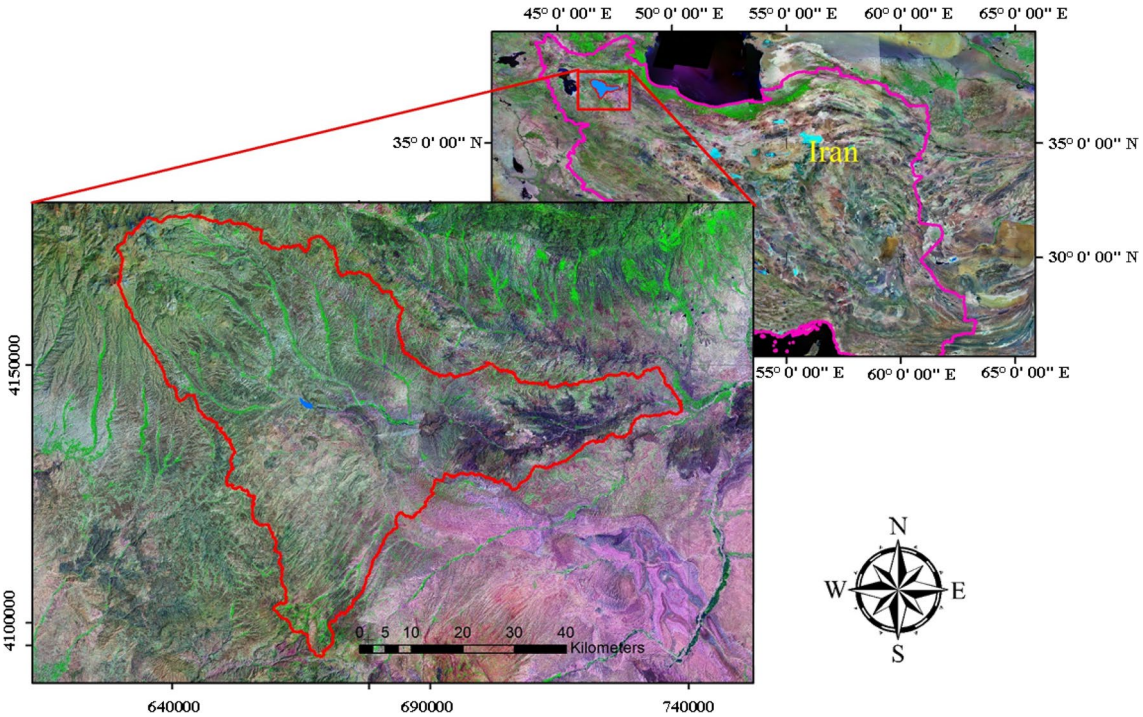


Fig. 4 Location of the study area

featuring a clay core with a crown length of 450 m and a height of 47 m. The normal volume of the reservoir equals 165×10^6 m^3, and its active volume equal to 135×10^6 m^3 with a dead volume equal to 17×10^6 m^3 (Golfam et al. 2019). The reservoir serves 14,500 ha of cultivated land. The objective function of the reservoir operation problem under the baseline and climate-change conditions minimizes the sum of squares of the relative water-supply deficiency during 30 years of operation. Existing meteorological stations have data for 1971–2000. This was the baseline period. The period 2040–2069 constitutes the future period under climate-change conditions. Climate projections (rainfall and surface air temperature) were performed with the HadCM3 (Gordon et al. 2000) model under the A2 emission scenario (envisions the most critical greenhouse gases emission scenario). Projected surface temperature and rainfall under climate change scenarios are presented in Fig. 5(a). Climate projections were input to the streamflow simulation model [IHACRES (Identification of unit hydrographs and component flows from rainfall, evapotranspiration and streamflow)] (Littlewood et al. 1997) and the crop-water demand model [Cropwat (crop water requirement)] (Clarke et al. 2000) to create future hydrologic and agricultural water use projections in the study area. Figure 5(b) shows the streamflow and water demand for the baseline and climate-change periods. The results indicate a 25% reduction in reservoir inflow and a 20% increase in water demand over the future 30 years relative to the baseline. The climate projections were made by Ashofteh et al. (2017) in association with a multi-objective reservoir operation study that assessed system performance under climate-change conditions.

The objective function of reservoir operation and the constraints of the problem minimize the sum of the squared normalized water supply deficits:

$$\text{Minimize} f = \sum_{t=1}^{T} \left(\frac{D_e(t) - R_e(t)}{D_{e\max}}\right)^2 \tag{7}$$

The reservoir continuity equation:

$$S(t+1) = S(t) + Q(t) - R_e(t) - \text{Loss}(t) - Sp((t) \tag{8}$$

Evaporative loss:

$$\text{Loss}(t) = A(t) \cdot Ev(t) \tag{9}$$

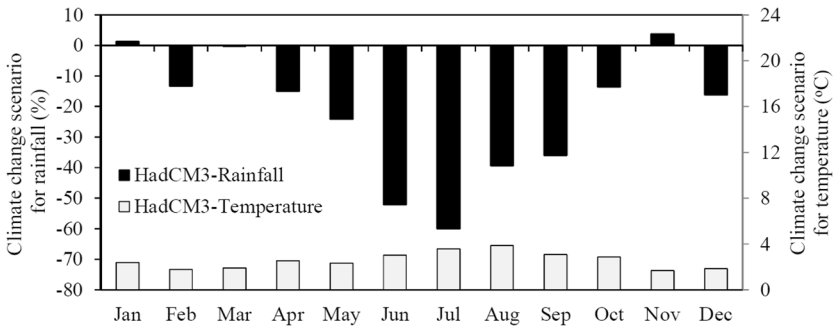
Reservoir lake surface area as function of reservoir storage:

$$A(t) = g[S(t)] \tag{10}$$

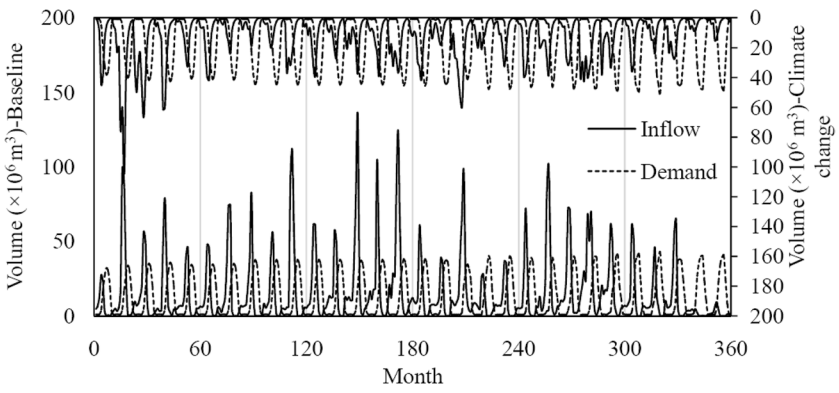
Constraints on reservoir storage:

$$S_{\min} \le S(t) \le S_{\max} \tag{11}$$

Constraints on release:



(a)



(b)

Fig. 5 **a** Projected temperature and rainfall under climate change scenarios. **b** Reservoir inflow and water demand under baseline and climate-change conditions

$$0 \le R_e(t) \le D_e(t) \tag{12}$$

in which the time index $t = 1, 2, \ldots, T$; $D_e(t)$ = downstream water demand during operating period t;$R_e(t)$ $R_e(t)$ = water release from the reservoir during operating period t;D_{emax} D_{emax} = maximum water demand during operating period; $S(t)$ = reservoir storage at the beginning of period $t+1$; S(t) S(t) = reservoir storage at the beginning of period t,Q(t) Q(t) = river inflow to the reservoir during period t;Loss(t) Loss(t) = water loss from the reservoir during period t;Sp(t) Sp(t) = reservoir spill during period t;A(t) A(t) = lake surface area of reservoir at the start of period t, which is a function of storage S(t) S(t); Ev(t) Ev(t) = evaporation depth during period t; T = total number of operation periods; and t = period counter.

The performance or efficiency indexes of reservoir operation are the reliability, vulnerability, and resiliency. Reliability is defined as the probability that failure does not occur during the operating period (Hashimoto et al. 1982). The reliability is investigated in two ways, one for the volume of water supply (volumetric reliability, *VR*) and one for the number of successful supply periods (*NR,* numeric reliability ranges between 0 and 1).

$$VR = \frac{\sum_{t=1}^{T} R_e(t)}{\sum_{t=1}^{T} D_e(t)} \times 100 \tag{13}$$

$$NR = \frac{\sum_{t=1}^{T} \begin{cases} R_e(t) \ge D_e(t) \\ \text{otherwise} = 0 \end{cases}}{T} \times 100 \tag{14}$$

The vulnerability (V ranges between 0 and 1) measures the magnitude of system water-supply deficit (Hashimoto et al. 1982), and is defined as follows:

$$V = \frac{1}{T} \sum_{t=1}^{T} \frac{\text{Deficit}(t)}{D_e(t)} \times 100 \tag{15}$$

Deficit (t) = water deficit in period t is equal to the water demand in period t minus the reservoir release in period t or equals zero if the water demand is less than the reservoir release in period t. The resiliency equals the probability that the system returns to the desired state after a failure (Hashimoto et al. 1982). Herein, the resiliency (Res ranges between 0 and 1) denotes the number of times the system has changed from a failure state to a satisfactory state over the total number of times the system fails written as follows (Ashofteh et al. 2017):

$$\text{Res} = \frac{\sum_{t=1}^{T} (D_e(t+1) = 0 \big| D_e(t) > 0)}{\sum_{t=1}^{T} N(D_e(t) > 0)} \times 100 \tag{16}$$

The flexibility index (FL) is a general criterion for expressing system conditions combining the three preceding indexes (Loucks 1997). The flexibility index is an effective benchmark for determining the overall performance of a water resources system:

$$FL = (1 - V) \times \text{Res} \times (VRorNR) \times 100 FL = (1 - V) \times \text{Res} \times (VRorNR) \times 100 \tag{17}$$

3 Results and discussion

Table 3 lists the run times (T1) required to obtain the optimization results for the test functions: Ackley, Rosenbrock, and Sphere. The optimizations were carried out in a high-performance computer (HPC) with an Intel (R) Xeon (R) CPU E5-2860 v4 @ 2.00 GHz 2.00 GHz (2processors) and 16 GB RAM. It is seen in Table 3 the optimization time for the Ackley, Rosenbrock, and Sphere test problems is shorter by implementing the FA-KNN compared with the FA solutions. T1 in Table 3 denotes the time required to simulate the main simulator once. The simulation times invested to solve the test problems were negligible. In the three problems, the run times seen in the first row for T1 = 0 represent the real times required to optimize the test functions. These sets optimized by the FA (sets A, C, and D) are often close to zero. On the other hand, for sets B-I and B-II, the run times are longer. The run times are required by KNN in the FA-KNN algorithm. For example, it took 9 min and 46 s for the Ackley function in the B-I, and 9 min and 1 s for the B-II. It is seen in Table 3 that the time required for one simulation by the main simulator was increased artificially by 0.5, 1, 3, 12, 60, and 300 s. The results show that with an increase of 0.5 s of simulation time in the main simulator, the FA-KNN performed much better than the FA. In addition, concerning sets BI and B-II, the optimization of the set B-II takes less time than the set BI because the threshold value of the former is smaller than that of the latter, and stricter conditions prevail when executing the model. Table 3 lists the optimization times for the Ackley, Rosenbrock, and Sphere test problems implementing the FA-KNN, which decreased by 95.5, 94, and 92%, respectively, compared to the non-hybrid optimization. The time reduction values for the Ackley, Rosenbrock, and Sphere problems were calculated by averaging the reduction time achieved with the B-I and B-II models (which had the best accuracy among the group-sets studied with FA-KNN) compared to the A model. The execution time for each of the models B-I and B-II were subtracted from the execution time of model A, and the result was divided by the execution time of model A. This ratio was expressed as

Table 3 Run time required for optimizations. See Table 2 for the designation of runs

Ackley							
T1 (secs)	A-I (hrs:mins:secs)	B-I (hrs:mins:secs)	C-I (hrs:mins:secs)	D-I (hrs:mins:secs)	B-II (hrs:mins:secs)	C-II (hrs:mins:secs)	D-II (hrs:mins:secs)
0	00:00:02	00:09:46	00:00:00	00:00:00	00:09:01	00:00:00	00:00:00
0.5	05:26:04	00:24:59	00:15:13	00:15:15	00:24:13	00:15:12	00:15:15
1	10:52:05	00:40:12	00:30:26	00:30:30	00:30:24	00:30:24	00:30:29
3	32:36:09	01:41:04	01:39:18	01:31:30	01:40:13	01:31:12	01:31:27
12	130:24:36	06:14:58	06:05:12	06:06:00	06:13:49	06:04:48	06:05:48
60	652:03:00	30:35:46	30:26:00	30:30:00	30:33:01	30:24:00	30:29:00
300	3260:15:00	153:40:46	152:10:00	152:30:00	152:09:01	152:00:00	152:25:00
Rosenbrock							
T1 (secs)	A-I (hrs:mins:secs)	B-I (hrs:mins:secs)	C-I (hrs:mins:secs)	D-I (hrs:mins:secs)	B-II (hrs:mins:secs)	C-II (hrs:mins:secs)	D-II (hrs:mins:secs)
0	00:00:03	00:03:35	00:00:00	00:00:00	00:02:29	00:00:00	00:00:00
0.5	01:43:50	00:09:42	00:06:07	00:06:08	00:08:35	00:06:06	00:06:07
1	03:27:33	00:15:49	00:18:21	00:18:24	00:14:41	00:18:18	00:18:22
3	10:22:48	00:40:17	00:36:42	00:36:48	00:39:05	00:36:36	00:36:45
12	41:31:03	02:30:23	02:26:48	02:27:12	02:28:53	02:26:24	02:27:00
60	207:35:00	12:17:00	12:14:00	12:16:00	12:14:29	12:12:00	12:15:00
300	1037:55:03	61:13:35	61:10:00	61:20:00	61:02:29	61:00:00	61:15:00
Sphere							
T1 (secs)	A-I (hrs:mins:secs)	B-I (hrs:mins:secs)	C-I (hrs:mins:secs)	D-I (hrs:mins:secs)	B-II (hrs:mins:secs)	C-II (hrs:mins:secs)	D-II (hrs:mins:secs)
0	00:00:03	00:10:31	00:00:00	00:00:00	00:08:56	00:00:00	00:00:00
0.5	02:26:34	00:21:11	00:10:40	00:10:41	00:19:34	00:10:38	00:10:40
1	04:53:04	00:31:51	00:21:20	00:21:22	00:30:13	00:21:17	00:21:20
3	14:39:06	01:14:31	01:04:00	01:04:06	01:12:47	01:03:51	01:04:00
12	58:36:15	04:26:31	04:16:00	04:16:24	04:24:20	04:15:24	04:16:00
60	293:01:03	21:30:31	21:20:00	21:22:00	21:25:56	21:17:00	21:20:00
300	1465:05:03	106:50:31	106:40:00	106:50:00	106:33:56	106:25:00	106:40:00

a percentage. This was done for all the B-I and B-II results, which were then averaged. The averages are listed in Table 3.

These three test problems are known to have global optima equal to zero. Table 4 lists the difference between the calculated optimal objective function and the global optimum for each problem. Table 4 indicates the final values of the objective function are closer to the global optima by implementing the FA-KNN compared with the FA.

Figure 6 displays the maximum, average, and minimum number of the FA iterations in five runs for several sets of the three test problems. The number of the FA iterations of set A-I was much larger than the other sets. Results indicate that the number of iterations decreases when the FA-KNN algorithm is implemented.

Figure 7 shows the run time and the difference between the objective function and global optima for all problem sets. The results in Fig. 7 establish shorter run times and better accuracy when implementing the FA-KNN. It is seen in Fig. 7 that the set A (solved by FA) exhibits the longest time, and the sets C and D have the shortest times among the sets. The sets A, B-I, and B-II achieved the highest precision and the smallest difference with the absolute optima, and these three sets' results are very close to each other in terms of computational time and accuracy. In solving the Ackley test problem. the accuracy of the set B is higher than set A's.

Table 4 The difference between calculated objective functions and global optima for test problems. See Table 2 for the designation of runs

Problem	A-I	B-I	C-I	D-I	B-II	C-II	D-II
Ackley	0.0000265	0.00000578	0.013	0.000735	0.0000258	0.063	0.000884
Rosenbrock	0.039	0.0566	0.065	0.057	0.0567	0.058	0.061
Sphere	6.62E-10	7.14E-10	1.04E-07	5.87E-08	7.52E-10	1.75E-07	9.76E-08

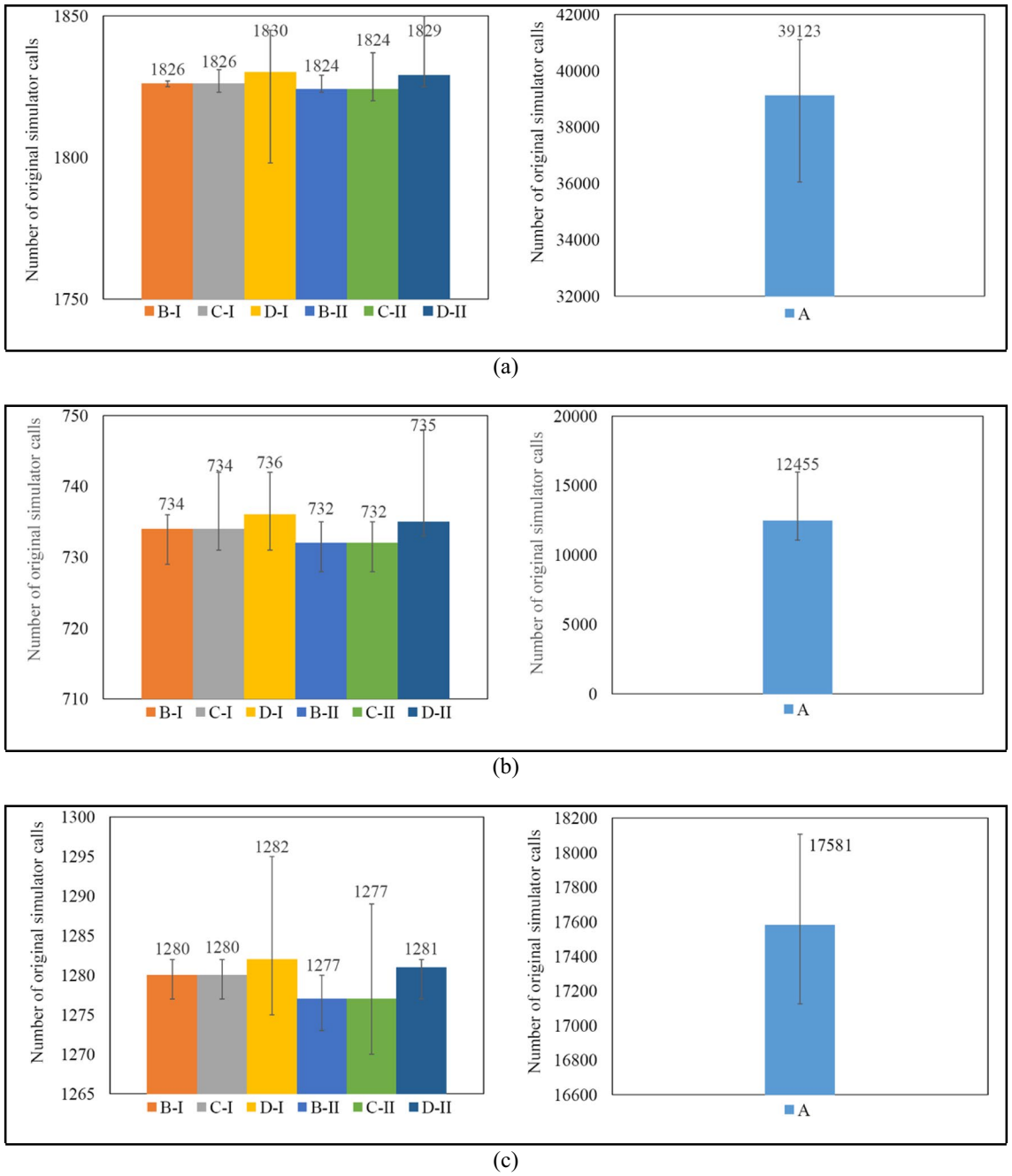


Fig. 6 The maximum, average, and minimum number of the number of calls of the main simulator corresponding to sets A-I, BI, B-II, CI, C-II, DI, and D-II for the **a** Ackley, **b** Rosenbrock, and **c** Sphere functions

Generally, the set C has the lowest accuracy among the sets under consideration. Therefore, by considering the factors of time and accuracy, the results show better performance of the two sets B-I and B-II.

Figure 8 depicts the convergence history of the sets B-I and B-II in comparison with the set A-I for the test functions. Generally, the FA-KNN converges faster to the solution than the FA. Figure 8 shows the convergence curve of the two sets B-I and B-II in comparison with the set A for all three test functions. Figure 8 shows the convergence curve of the two sets B-I and B-II are approximately similar to set A. In Fig. 8(a), the convergence number for the two sets B is less than that of the set A, and in the other two figures is larger. In all the figures, it is seen that the degree of overlap at the beginning of the diagram increases, and it shows that the algorithm is searching the entire solution space. The

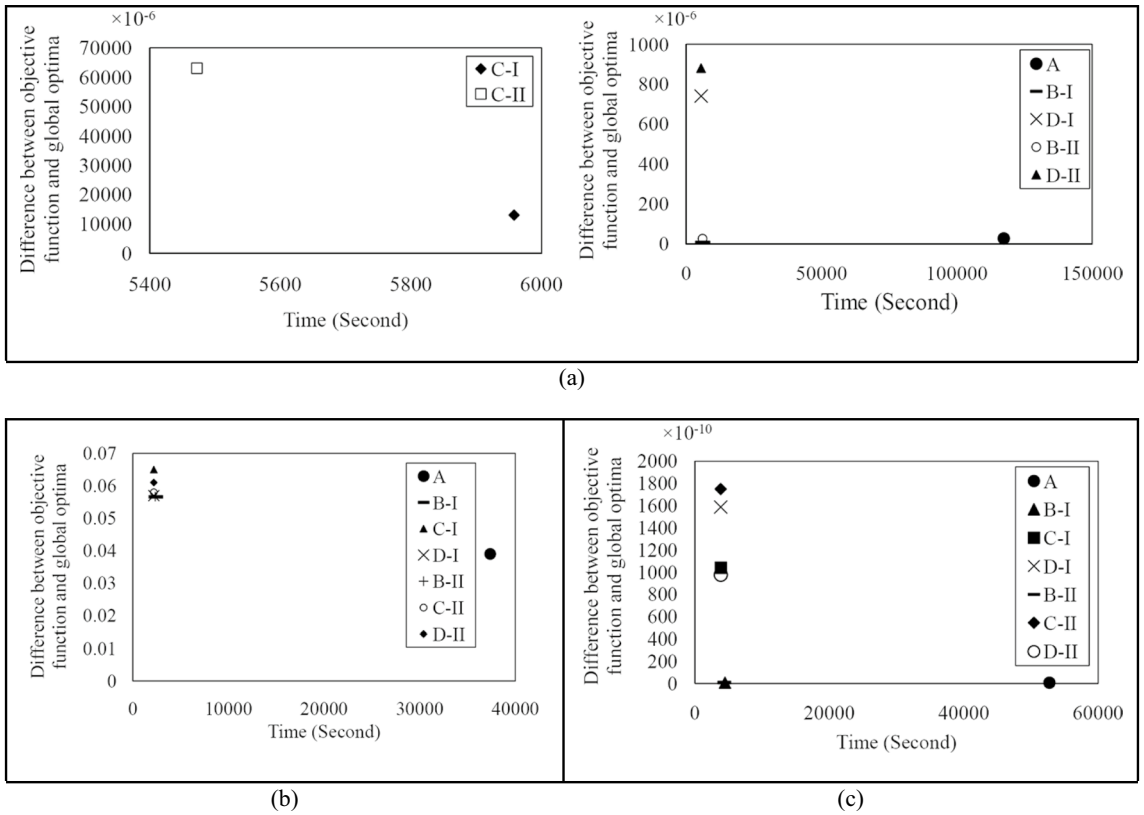


Fig. 7 The difference between the objective function and the global optima versus time for the **a** Ackley, **b** Rosenbrock, and **c** Sphere functions

set B-I approaches the optimal solution more rapidly in the initial iterations relative to the other two sets. Also, the end range of the graph is limited to small numbers, and the range of variation is very small, showing that the firefly population is well modified and approaches the optimal solution. In addition, the set B-I converges more quickly and performs better than the set B-II.

The reservoir problem was solved under baseline and climate-change conditions with five problem sets. Table 5 lists the problem sets evaluated and the characteristics of each problem set.

The FA-KNN accuracy was evaluated with the root mean square error (RMSE). Tables 6 and 7 list the performance of the system under baseline and climate-change conditions for all five sets examined. In general, the volumetric reliability under the baseline conditions was about 10% higher than under climate-change conditions, and the vulnerability was about 10% less under baseline conditions than under climate-change condition. The flexibility of the system under the baseline conditions is about 3.5% higher than under climate-change conditions. In addition, the number of calls to the main simulator in the FA-KNN algorithm was much lower than with the FA (about 0.3 times that of the FA). The number of iterations is equal for of the FA and FA-KNN. Among the sets under baseline conditions, the set CBS-I has a superior performance with higher flexibility, while under climate-change condition, the sets Sc-II and CBSs-II have better volumetric and numeric flexibly than other sets, respectively.

Figure 9 depicts the run time and error in the objective function for the problem sets under baseline and climate-change conditions. The shortest optimization time belongs to the sets solved by FA-KNN. The problem sets solved by FA-KNN have a greater error than the solutions set obtained with the FA due to the use of the hybrid simulator and the interpolation between the solutions in the database. Figure 10(a) and (b) depict the maximum, average, and minimum convergence histories of the five independent runes for the CsA set (case study A) under baseline and climate-change conditions, respectively. Figure 10(c) and (d) display the convergence history for the mean of five runs under climate-change and baseline conditions, respectively. These graphs establish favorable convergence speed of the FA-KNN compared with the FA.

Figure 11 presents a comparison between the obtained releases from the five problem sets examined and the water demand under baseline and climate-change conditions. The water deficiency under climate-change exceeds that under

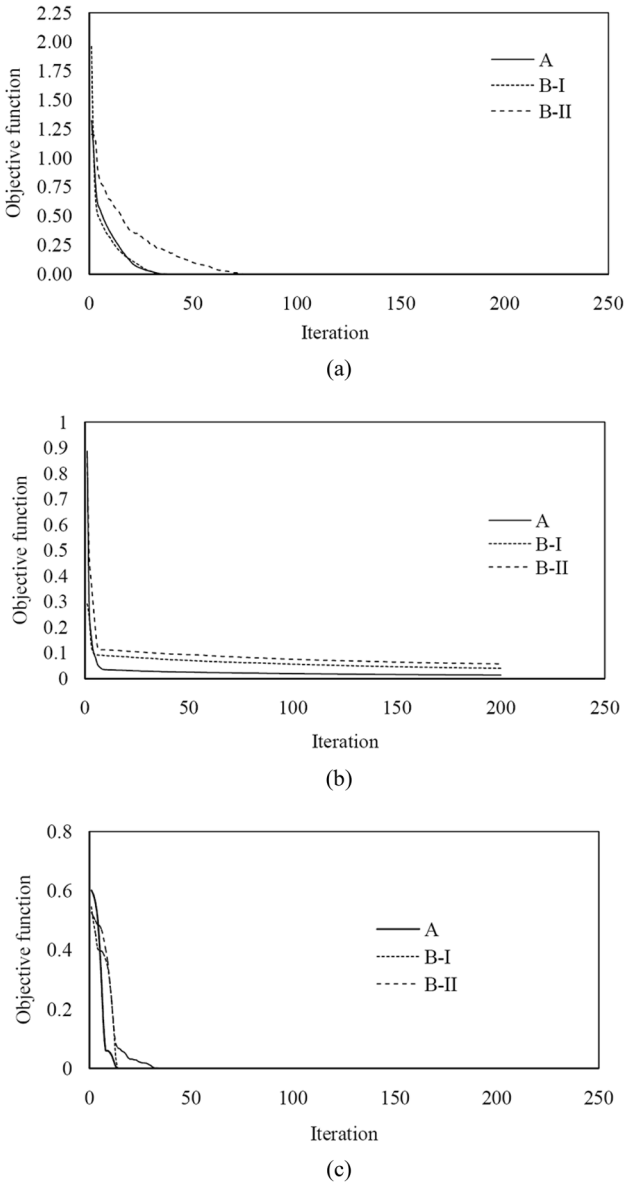


Fig. 8 The convergence curve for sets A-I, B-I, and B-II for **a** Ackley,(**b** Rosenbrock, and **c** Sphere functions

baseline conditions, rendering the system more vulnerable in the future period.

The results for the problem set CsB-I displayed in Figs. 12 and 13 compare the reservoir capacity, volume of spill, and volume of water deficit with the maximum and minimum reservoir capacity, reservoir inflow, water release, and water demand volume under baseline and climate-change conditions, respectively. Figure 12 establishes the deficit of water supply increases under climate-change relative to the baseline conditions, and the deficit grows over time. Figure 13 indicates reservoir storage under climate-change decreases relative to baseline condition, the storage peaks are close to the maximum volume of the reservoir, and the volume and number reservoir spills under in climate-change conditions are smaller than under baseline conditions.

4 Discussion

The FA-KNN algorithm is simpler than support vector machine (SVM) and does not require training as the SVM does. It also bypasses the need for learning, which is advantageous. The FA-KNN algorithm updates its database in real time and dynamically, which increases the accuracy of the algorithm, while SVM must cope with online learning. The SVM is not appropriate as a classification technique for online learning because of storage space and computational requirements.

The accuracy of the solutions and the speed of the algorithm presented in this paper can be adjusted by the user, and this is done by trial and error. The authors attempted calculating highly accurate solutions by considering thresholds that are very small relative to the magnitudes of the decision variables. The accuracy of the results and the execution time are reduced with larger thresholds or lower k. Thus, although the accuracy of the results obtained with the FA-KNN algorithm is lower than the accuracy of the results obtained with the FA (as shown in Fig. 9), the execution time is also

Table 5 Characteristics of the reservoir problem's optimization sets

Problem	Set	Algorithm	K	Threshold	Stopping criterion (number of iterations)	Description
(Case study A) CsA	-	FA	-	-	2000	Population = 10 Number of decision variables = 360 $\beta_0 = 2$ $\alpha = 10$ $\gamma = 0.01$
(Case study B) CsB	I II	FA-KNN	5	0.8 0.4	2000	
(Case study C) CsC	I II	FA-KNN	3	0.8 0.4	2000	

Table 6 Performance indexes, run time, and number of FA-KNN iterations associated with problem sets corresponding to baseline conditions

Problem	Set	Volume reliability (%)	Number reliability (%)	Vulnerability (%)	Resiliency (%)	Flexibility (%)		Run time (seconds)	The number of calls to main simulator
						Volume	Number		
(Case study A) CsA	-	88.46	26.94	12	15.20	11.9	3.62	57,814.002	115,446
(Case study B) CsB	I	86.88	27.78	12	15.77	12	3.8	20,904.702	38,679
	II	87.31	26.94	12.11	14.83	11.37	3.51	23,281.002	43,405
(Case study C) CsC	I	87.02	26.67	12.17	15.15	11.58	3.54	20,487.498	38,108
	II	87.40	27.50	12	15.32	11.75	3.69	22,823.202	42,811

Table 7 Performance indexes, run time, and number of FA-KNN calls associated with problem sets corresponding to climate-change conditions

Problem	Set	Volume reliability (%)	Number reliability (%)	Vulnerability (%)	Resiliency (%)	Flexibility (%)		Run time (seconds)	The number of calls to the main simulator
						Volume	Number		
(Case study A) CsA	-	70.04	26.94	21.29	14.83	8.17	3.14	59,604.4	119,031
(Case study B) CsB	I	69.62	26.94	21.10	15.21	8.35	3.233	22,271.16	42,067
	II	69.68	27.50	21.163	14.94	8.20	3.239	25,072.86	47,663
(Case study C) CsC	I	69.68	27.22	21.161	15.27	8.38	3.27	22,267.38	41,819
	II	70.06	26.67	20.92	15.15	8.39	3.19	24,407.58	46,127

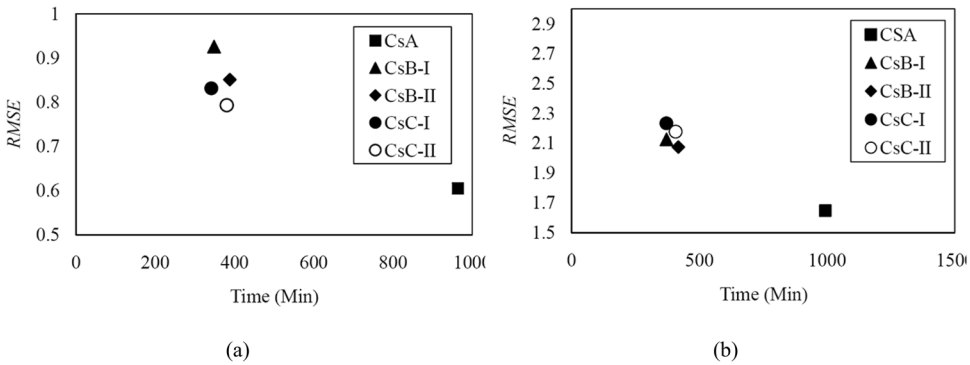


Fig. 9 Comparison of the run time and error associated with sets CsA, CsB-I, CsB-II, CsC-I, and CsC-II corresponding to (**a**) baseline conditions and (**b**) climate-change conditions

significantly reduced. The performance of the FA-KNN algorithm must consider run time and accuracy criteria so that the user can select the appropriate threshold value based on specific preferences using trial and error. This paper's objective is to provide an algorithm that overcomes the computational burden of time-consuming algorithms, thus its focus on execution time and accuracy. Climate change studies commonly involve long simulation periods, which excise costly computational burden.

It is seen in Fig. 11 that the baseline and climate change periods feature deficits in water supply, which may be due to reduced inflow to the reservoir in the final years of the baseline and climate change periods. Water demand under climate change conditions would be higher than under baseline conditions, and reservoir inflow would be reduced under climate change conditions compared to baseline conditions. This means that full supply of the water demand would not occur in most years of the climate change period.

It is seen in Fig. 12(a) the reservoir is in a better condition under the baseline conditions than under climate change conditions, and the deficit in downstream water demand from month 330th onwards (from year 27 onwards) exceeds than 10×10^6 m^3, compared to prior supply shortage between 0 and5×10^6 m^3. This may be due to the fact that as one approaches the final years of the baseline period (especially the last 3 years), the reservoir inflow decreases. The maximum inflow under baseline in the 30-year period is 135.37×10^6 m^3. Figure 12(b) shows that under climate change, the inflow to the reservoir is reduced, and the downstream water demand increases compared to the baseline conditions. The amount of inflow in the 30-year climate change period under study varies from 0.03 to 104.6×10^6

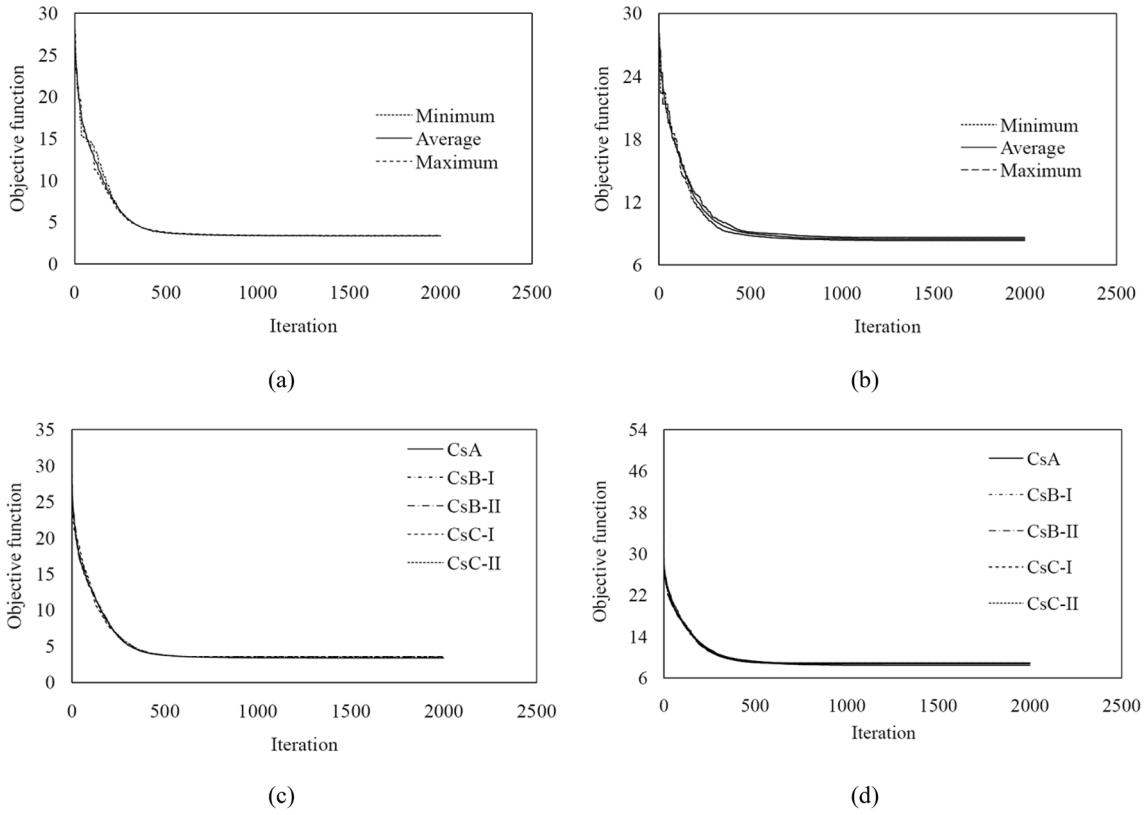


Fig. 10 The convergence history for **a** and **c** baseline condition, **b** and **d** climate-change conditions. (**a** and **b** represent the CsA, **c** and **d** represent all sets)

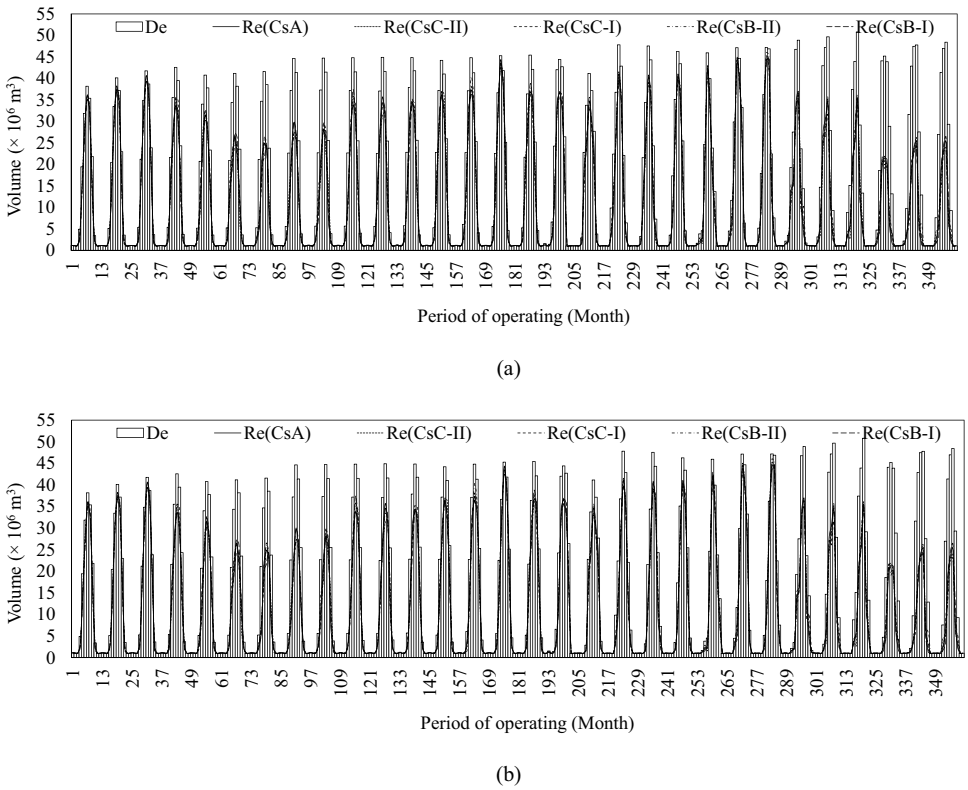


Fig. 11 Comparison between the water releases and the water demand corresponding to **a** the baseline and **b** the climate-change conditions for all sets

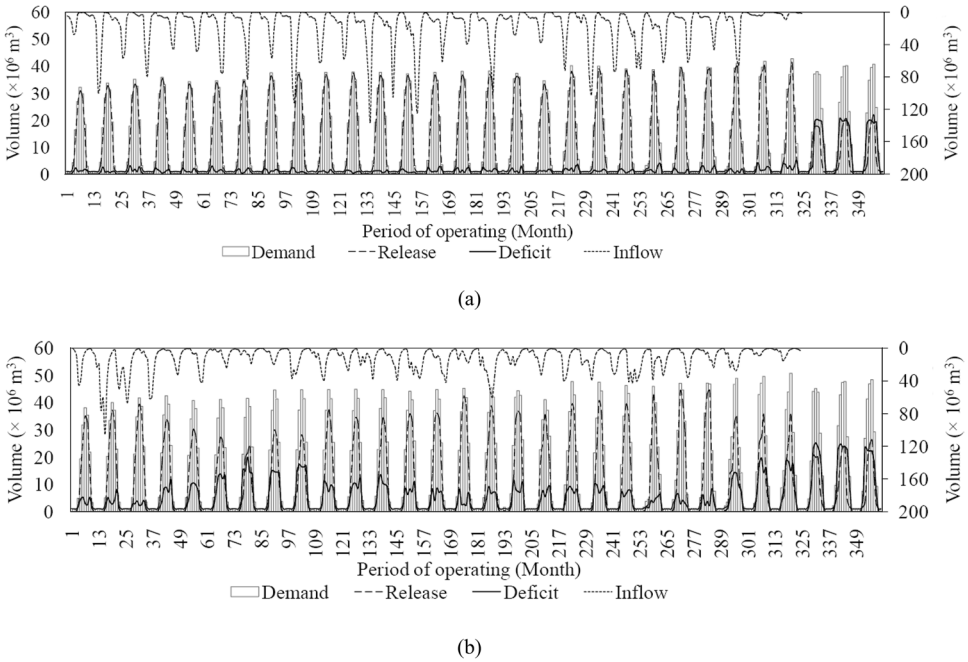


Fig. 12 Comparison of the water release and (water demand, water deficit, reservoir inflow) for the CsB-I set under **a** baseline and **b** climate-change conditions

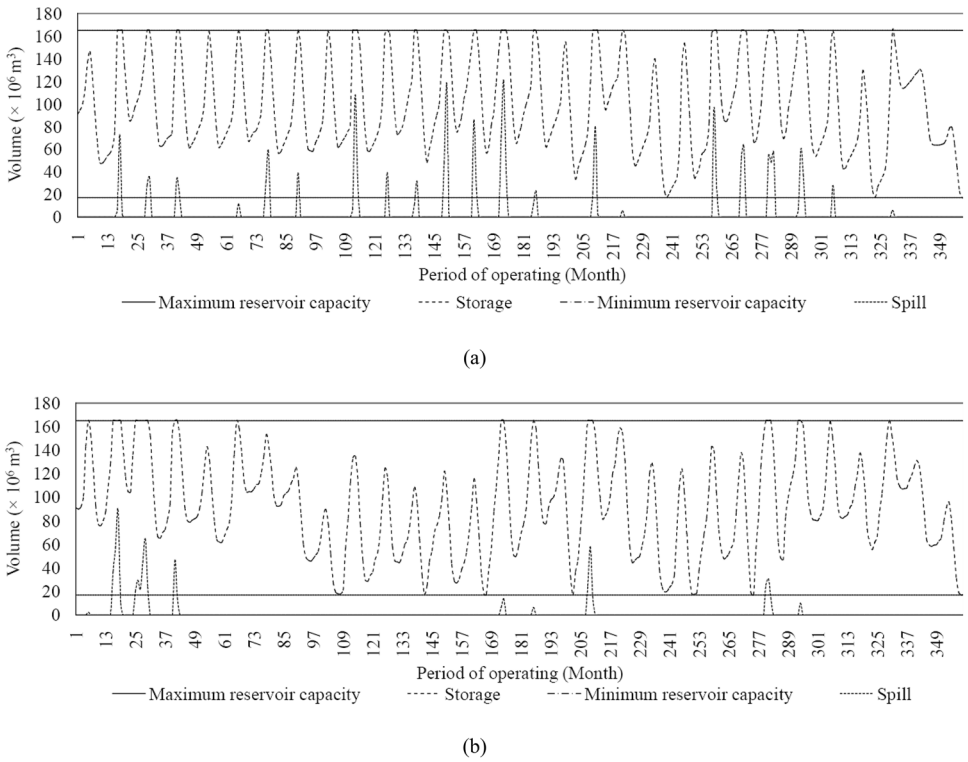


Fig. 13 Comparison of the storage and reservoir spill obtained for the CsB-I set showing the maximum and minimum capacity of the reservoir under **a** the baseline and **b** climate-change conditions

m^3). The deficit in water supply increases under climate change compared to the baseline conditions, and the amount of the deficit increases with the passage of time.

Figure 13 shows the reservoir volume fluctuates between its maximum (165×10^6 m^3) and the minimum (17×10^6 m^3) volumes. It is seen in Fig. 13 that 30 peaks (corresponding to 30 years) are observed, most of which (23 years) are close to the maximum reservoir capacity under baseline conditions, and the reservoir storage volume is generally larger than during the climate change period. The storage volume in the reservoir decreases during the climate change period, and there are only 13 years when the peaks approach the maximum reservoir

volume. The amount of spill and the number of overflows from the reservoir under climate change conditions are less than under the baseline condition. In the last years of the 30-year climate change period, due to the decrease in reservoir inflow, the storage volume of the reservoir decreases. The overflow volume for the baseline period ranged from zero to 120.71×10^6 m^3) and for the climate change period, it ranged from zero to 89.36×10^6 m^3).

5 Concluding remarks

This research proposed an approach to reduce run-time in simulation–optimization methods. The presented approach is not limited to the use of a specific optimization algorithm. The literature review showed the FA had a good performance compared to other meta-heuristic algorithms. For this reason, this paper's linked the developed operators of the optimization algorithms with FA. This paper approach can be linked with all other optimization algorithms to increase their speed of convergence to the solution in problems that feature time and computational complexity. The approach presented in this paper is not limited to use with the FA. Any meta-heuristic algorithm can be improved with this approach to solve optimization problems beset by computational complexity. Possibly, areas of application for this paper's method are linking water quality simulation with optimization, or solving optimization problems using Big Data under climate change conditions.

This study evaluated the convergence speed, accuracy, and applicability of the proposed FA-KNN algorithm with three standard problems (Ackley, Rosenbrock, and Sphere) and a reservoir operation problem. Results from the test problems established very high accuracy, and reductions equal to 95.5, 94, and 92% of the FN-KNN's run-time compared to the FA's for the Ackley, Rosenbrock, and Sphere problems, respectively.

The run time of the FA-KNN method in solving the reservoir operation problem was reduced by more than 60% relative to the FA's. In general, the system's volumetric reliability is higher and the vulnerability lower under baseline conditions than under climate-change conditions, respectively. The flexibility of the system is about 3.5% higher under baseline conditions than under climate-change conditions. Calculation of the efficiency indexes of the model proved that the release policy obtained with the proposed FA-KNN had better performance than with the FA method. The reservoir problem under baseline conditions featured RMSE of the FA algorithm equal to 0.6×10^6 m^3, and the largest error among the sets solved with FA-KNN is equal to 0.92×10^6 m^3. Under climate change conditions, the error of the FA equals 1.64×10^6 m^3; the largest error of the FA-KNN is equal to 2.23×10^6 m^3. This work's results show the FA-KNN hybrid algorithm can solve complex optimization problems beset by a large computational burden. Algorithmic speed, the bypassing of model training, faster convergence, high accuracy, and reduced optimization time are the strengths of the proposed algorithm. It is suggested that future research consider the uncertainty in projections of climate change and its effects on reservoir operation. Limitations of presented methodology are as follows:

(1) Problem complexity: the FA-KNN algorithm demonstrated improved execution speed and acceptable accuracy for the problems we evaluated. Nevertheless, it is possible that the algorithm's performance could differ when applied to more complex problems, or to problems with other characteristics;
(2) Data availability: the assumptions, model simplifications, and data input used in the reservoir problem may not be applicable to other problems;
(3) Climate change scenarios: potential limitations related to climate change scenarios and the role of uncertainty related to climate change predictions could affect the robustness of the results obtained for future periods;
(4) Future research: the FA-KNN algorithm performs better than the FA algorithm in terms of speed and accuracy, although comparisons with other algorithms and other optimizations problems would be worthwhile in future investigations.

Author contribution Firoozeh Azadi developed the theory and performed the computations. Parisa-Sadat Ashofteh verified the analytical methods and encouraged Firoozeh Azadi to investigate a specific aspect. Parisa-Sadat Ashofteh supervised the findings of this work, and Loáiciga helped supervise the project and reviewed the paper contents. All authors discussed the results and contributed to the final manuscript. Ashkan Shokri helped the algorithmic part of the research. Firoozeh Azadi wrote the manuscript with support from Parisa-Sadat Ashofteh, and especially Loáiciga. Parisa-Sadat Ashofteh conceived the original idea.

Data availability My manuscript has no associated data.

Declarations

Ethical approval The paper is not currently being considered for publication elsewhere. The paper reflects the authors' own research and analysis in a truthful and complete manner.

Informed consent Informed consent is not applicable.

Competing interests The authors declare no competing interests.

References

Afnizanfaizal A, Safaai D, MohdSaberi M, SitiZaiton MH (2012) A new hybrid firefly algorithm for complex and nonlinear problem. Adv Intell Soft Comput 151:637–680

Ahani A, Shourian M, Rahimi Rad P (2018) Performance assessment of the linear, nonlinear and nonparametric data driven models in river flow forecasting. Water Resour Manage 32(2):383–399. https://doi.org/10.1007/s11269-017-1792-5

Ahmadianfar I, Samadi-koucheksaraee A, Bozorg-Haddad O (2017) Extracting optimal policies of hydropower multi-reservoir systems utilizing enhanced differential evolution algorithm. Water Resour Manage 31:4375–4397. https://doi.org/10.1007/s11269-017-1753-z

Ahmadianfar I, Samadi-koucheksaraee A, Asadzadeh M (2022) Extract nonlinear operating rules of multi-reservoir systems using an efficient optimization method. Sci Rep 12:18880. https://doi.org/10.1038/s41598-022-21635-0

Ahmadianfar I, Samadi-Koucheksaraee A, Razavi S (2023) Design of optimal operating rule curves for hydropower multi-reservoir systems by an influential optimization method. Renew Energy 211:508–521. https://doi.org/10.1016/j.renene.2023.04.113

Almubaidin MA, Najah AM, Sidek LM, Elshafie A (2022) Using metaheuristics algorithms (MHAs) to optimize water supply operation in reservoirs: a review. Arch Comput Methods Eng 29(4). https://doi.org/10.1007/s11831-022-09716-9

Arkesteijn L, Pande S (2013) On hydrological model complexity, its geometrical interpretations and prediction uncertainty. Water Resour Res 49(10):7048–7063. https://doi.org/10.1002/wrcr.20529

Ashofteh P-S, Rajaee T, Golfam P (2017) Assessment of water resources development projects under conditions of climate change using efficiency indexes (EIs). Water Resour Manag 31(12). https://doi.org/10.1007/s11269-017-1701-y

Ashofteh P-S, Bozorg-Haddad O, Loáiciga HA (2021) Application of bi-objective genetic programming (BO-GP) for optimizing irrigation rules using two reservoir performance criteria. J River Basin Manag 19(1). https://doi.org/10.1080/15715124.2019.1613415

Azizipour M, Afshar MH (2018) Reliability-based operation of reservoirs: a hybrid genetic algorithm and cellular automata method. Soft Comput 22(19):6461–6471. https://doi.org/10.1007/s00500-017-2698-0

Azizipour M, Ghalenoei V, Afshar MH, Solis SS (2016) Optimal operation of hydropower reservoir systems using weed optimization algorithm. Water Resour Manage 30(11):3995–4009

Bozorg-Haddad O, Solgi M, Loáiciga HA (2017) Meta-heuristic and evolutionary algorithms in engineering optimization. John Wiley & Sons, New York

Chong KL, Lai SH, Najah AM, El-Shafie A (2021) Review on dam and reservoir optimal operation for irrigation and hydropower energy generation utilizing meta-heuristic algorithms. IEEE Access 9:19488–19505. https://doi.org/10.1109/ACCESS.2021.3054424

Chou JS, Ngo NT (2017) Modified firefly algorithm for multidimensional optimization in structural design problems. Struct Multidisc Optim 55:2013–2028. https://doi.org/10.1007/s00158-016-1624-x

Clarke D, Smith M, El-Askari Kh (2000) CropWat for Windows: User guide, 1–43.

Fang Y, Ahmadianfar I, Samadi-koucheksaraee A, Azarsa R, Scholz M, Yaseen ZM (2021) An accelerated gradient-based optimization development for multi-reservoir hydropower systems optimization. Energy Rep 7:7854–7877. https://doi.org/10.1016/j.egyr.2021.11.010

Garousi-Nejad I, Bozorg-Haddad O, Loáiciga HA (2016a) Modified firefly algorithm for solving multireservoir operation in continuous and discrete domains. J Water Resour Plan Manag 142(9). https://doi.org/10.1061/(ASCE)WR.1943-5452.0000644

Garousi-Nejad I, Bozorg-Haddad O, Loáiciga HA, Mariño MA (2016b) Application of the firefly algorithm to optimal operation of reservoirs with the purpose of irrigation supply and hydropower production. J Irrig Drain Eng 142(10). https://doi.org/10.1061/(ASCE)IR.1943-4774.0001064

Golfam P, Ashofteh P-S, Rajaee T, Chu X (2019) Prioritization of water allocation for adaptation to climate change using multi-criteria decision making (MCDM). Water Resour Manage 33(10):3401–3416. https://doi.org/10.1007/s11269-019-02307-7

Gordon C, Cooper C, Senior CA, Banks HT, Gregory JM, Johns TC, Mitchell JFB, Wood RA (2000) The simulation of SST, sea ice extents and ocean transport in a version of the Hadley Centre coupled model without flux adjustments. Clim Dyn 16:147–168

Hashimoto T, Stedinger JR, Loucks DP (1982) Reliability, resiliency and vulnerability criteria for water resources system performance evaluation. Water Resour Res 18(1):14–20. https://doi.org/10.1029/WR018i001p00014

Hassanzadeh T, Meybodi MR, Mahmoudi F (2011) "An improved firefly algorithm for optimization in static environment", The 5th Iran Data Mining Conference. Amirkabir University of Technology, Tehran, Iran, Dec, pp 13–14

Höge M, Wöhling Th, Nowak W (2018) A primer for model selection: the decisive role of model complexity. Water Resour Res 54(3):1688–1715. https://doi.org/10.1002/2017wr021902

Hong Y-S, Lee H, Tahk M-J (2003) Acceleration of the convergence speed of evolutionary algorithm using multi-layer neural networks. Eng Optim 35(1):91–102. https://doi.org/10.1080/03052150310000069672

Hossain MS, El-Shafie A, Mahzabin MS, Zawawi MH (2018) System performances analysis of reservoir optimization–simulation model in application of artificial bee colony algorithm. Neural Comput Appl 30(7):2101–2112. https://doi.org/10.1007/s00521-016-2798-2

Hu H, Yang K, Liu L, Su L, Yang Zh (2019) Short-term hydropower generation scheduling using an improved cloud adaptive quantum-inspired binary social spider optimization algorithm. Water Resour Manage. https://doi.org/10.1007/s11269-018-2138-7

Jato-Espino D, Sillanpää N, Charlesworth SM, Rodriguez-Hernandez J (2017) A simulation-optimization methodology to model urban catchments under non-stationary extreme rainfall events. Environ Model Softw. https://doi.org/10.1016/j.envsoft.2017.05.008

Johari NF, Zain AM, Mustaffa NH, Udin A (2013) Firefly Algorithm for Optimization Problem. Appl Mech Mater 421:512–517

Karami H, Mousavi SF, Farzin S, Ehteram M, Singh VP, Kisi O (2018) Improved krill algorithm for reservoir operation. Water Resour Manage 32(10):3353–3372. https://doi.org/10.1007/s11269-018-1995-4

Kou Y-Ch, Huang LH, Tsai T-L (2008) A hybrid three-dimensional computational model of groundwater solute transport in heterogeneous media. Water Resour Res 44(3). https://doi.org/10.1029/2007WR006084

Lai V, Huang YF, Koo CH, Ahmed AN, El-Shafie A (2022) Conceptual sim-heuristic optimization algorithm to evaluate the climate impact on reservoir operations. J Hydrol 614:128530. https://doi.org/10.1016/j.jhydrol.2022.128530

Lai V, Huang YF, Koo CH, Ahmed AN, El-Shafie A (2022a) A review of reservoir operation optimisations: from traditional models to metaheuristic algorithms. Arch Comput Methods Eng 29(5). https://doi.org/10.1007/s11831-021-09701-8

Li N, Yue XY (2018) Calibrating the spatiotemporal root density distribution for macroscopic water uptake models using Tikhonov regularization. Water Resour Res 54(3):1781–1795. https://doi.org/10.1002/2017WR020452

Littlewood IG, Down K, Parker JR, Post DA (1997) IHACRES: catchment-scale rainfall streamflow modelling (PC version) Version 1.0—April 1997. Institute of Hydrology, Centre for Ecology and Hydrology, Wallingford, Oxon. http://www.nwl.ac.uk/ih/www/products/mswihacres.html.

Loucks DP (1997) Quantifying trends in system sustainability. Hydrol Sci J 42(4):513–530

Meshram SG et al (2019) New approach for sediment yield forecasting with a two-phase feedforward neuron network-particle swarm optimization model integrated with the gravitational search algorithm. Water Resour Manage 33:2335–2356

Moreno-Rodenas AM, Bellos V, Langeveld JG, Clemens FHLR (2018) A dynamic emulator for physically based flow simulators under varying rainfall and parametric condition. Water Res 142:512–527. https://doi.org/10.1016/j.watres.2018.06.011

Nandy S, Sarkar PP, Das A (2012) Analysis of nature inspired firefly algorithm based back- propagation neural network training. Int J Comput Appl 43(22):8–16

Orabona F, Castellini C, Caputo B, Jie L, Sandini C (2010) On-line Independent Support Vector Machines. Pattern Recog 43(4). https://doi.org/10.1016/j.patcog.2009.09.021

SaberChenari K, Abghari H, Tabari H (2016) Application of PSO algorithm in short-term optimization of reservoir operation. Environ Monit Assess. https://doi.org/10.1007/s10661-016-5689-1

Samadi-koucheksaraee A, Ahmadianfar I, Bozorg-Haddad O, Asghari-Pari SA (2019) Gradient evolution optimization algorithm to optimize reservoir operation systems. Water Resour Manage 33:603–625. https://doi.org/10.1007/s11269-018-2122-2

Samadi-Koucheksaraee A, Shirvani-Hosseini S, Ahmadianfar I, Gharabaghi B (2022) Optimization algorithms surpassing metaphor", In: Bozorg-Haddad, O., Zolghadr-Asli, B. (eds) Computational intelligence for water and environmental sciences. Studies in computational intelligence, Vol 1043. Springer, Singapore. https://doi.org/10.1007/978-981-19-2519-1_1.

Schoups G, van de Gisen NC, Savenije HJG (2008) Model complexity control for hydrologic prediction. Water Resour Res 44:W00B03. https://doi.org/10.1029/2008WR006836

Shirvani-Hosseini S, Samadi-Koucheksaraee A, Ahmadianfar I, Gharabaghi B (2022) "Data mining methods for modeling in water science" In: Bozorg-Haddad, O., Zolghadr-Asli, B. (eds) Computational intelligence for water and environmental sciences. Studies in computational intelligence, Vol 1043. Springer, Singapore. https://doi.org/10.1007/978-981-19-2519-1_8.

Shokri A, Bozorg-Haddad O, Mariño MA (2013) Algorithm for increasing the speed of evolutionary optimization and its accuracy in multi-objective problems. Water Resour Manage 27(7):2231–2249. https://doi.org/10.1007/s11269-013-0285-4

Shokri A, Walker JP, Dijk AIJM, Wright AJ, Pauwels VRN (2018) Application of the patient rule induction method to detect hydrologic model behavioural parameters and quantify uncertainty. Hydrol Process 32(8):1005–1025

Silva DO, Vieira LGM, Lobato FS, Barrozo MAS (2013) Optimization of hydro cyclone performance using multi-objective firefly colony algorithm. Sep Sci Technol 48(12):1891–1899

Wang D, Qiao H, Zhang B, Wang M (2013) Online support vector machine based on convex hull vertices selection. IEEE Trans Neural Netw Learning Syst 24(4). https://doi.org/10.1109/TNNLS.2013.2238556

Yan X, Zhu Y, Wu J, Chen H (2012) An improved firefly algorithm with adaptive strategies. Adv Sci Lett 16(1):249–254

Yang XS (2008) Firefly algorithm, nature-inspired meta-heuristic algorithms. Wiley Online Libr 20:79–90

Yang XS (2011) Chaos-enhanced firefly algorithm with automatic parameter tunning. J Swarm Intell Res 2(4):1–11

Yang XS (2010) "Firefly algorithm, Lévy flights and global optimization", Research and Development in Intelligent Systems XXVI, 209–218.

Yaseen ZM, Karami H, Ehteram M, Mohd NS, Mousavi SF, Hin LS, Kisi O, Farzin S, Kim S, El-Shafie A (2018) Optimization of reservoir operation using new hybrid algorithm. KSCE J Civ Eng 22(11):4668–4680. https://doi.org/10.1007/s12205-018-2095-y

Zheng J, Shen F, Fan H, Zhao J (2013) An online incremental learning support vector machine for large-scale data. Neural Comput Appl 22(5). https://doi.org/10.1007/s00521-011-0793-1