Two-step approach based multi-objective groundwater remediation in highly heterogeneous media using enhanced random vector functional link integrated with evolutionary marine predator algorithm

Partha Majumder[1,2], Chunhui Lu[1,2*], T.I. Eldho[3]

[1]State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing, China

[2]Yangtze Institute for Conservation and Development, Hohai University, Nanjing, China

[3] Department of Civil Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India, E-mail: eldho@civil.iitb.ac.in

*Corresponding author: Chunhui Lu (clu@hhu.edu.cn)

Abstract

We herein propose two-step approach based simulation-optimization models for groundwater remediation using enhanced random vector functional link (ERVFL) and evolutionary marine predator algorithm (EMPA). The weighted least square method is used to improve the robustness of the ERVFL network, where weights are computed using the kernel density estimator (KDE). The EMPA is developed by modifying the marine predator algorithm (MPA) using elite opposition-based learning, biological evolution operators, and elimination mechanisms. In the multi-objective version of EMPA, the non-dominated solutions are stored in an external repository using an archive controller and adaptive grid mechanism to promote better convergence and diversity of the Pareto front. The performance evaluation of EMPA on several test functions suggests its superiority over other metaheuristics for both single-objective and multi-objective optimization. The ERVFL network is then used to approximate

the finite difference based groundwater flow and transport models to accelerate computational efficiency. The two-step approach based S-O models are then developed by integrating the simulation models directly or through the ERVFL network with the EMPA. The first step aims to find optimal pumping locations using EMPA with combinatorial optimization technique by minimizing the percentage of contaminant mass remained in the aquifer. In the second step, the ERVL based proxy simulator is coupled with EMPA and used for multi-objective optimization while explicitly using the pumping well locations as obtained in the first step. The multi-objective optimization generates a Pareto-optimal solution representing the relationship between the water extraction rates and the amount of contaminant mass in the aquifer. Further analyses suggest that the two-step approach shows a significant advantage over the traditional methods for multi-objective groundwater remediation.

Keywords: Groundwater remediation, Enhanced random vector functional link (ERVFL), Evolutionary marine predator algorithm (EMPA), Kernel density estimator (KDE)

51 **Notations**

C_R	Crossover operator
c_s^k	Dissolved contaminant concentration of species k (ML^{-3})
D	Search space dimension
$\mathbf{D}_h$	Coefficient of hydrodynamic dispersion (L^2T^{-1})
d	Hydraulic drawdown (L)
e_b	Bottom elevation (L)
e_t	Top elevation (L)
H_b	Aquifer thickness (L)
K	Kernel smoothing function
M	Contaminant mass (M)
m	Mutation scaling factor
N_d	Number of the input-output dataset
P	Penalty weight matrix
Q	Pumping rate (L^3/T)
q_s	Volumetric flux rate per unit volume representing sources or sinks of water (T^{-1})
R_L	Random number generated from Lévy distribution
R_B	Random number generated form normal distribution
R_d	Retardation coefficient
T	Number of stress period
t	Time (T)
$\mathbf{u}$	Seepage velocity (LT^{-1})

Δt	Time step (T)
$\sum r_n$	Chemical reaction term ($ML^{-3}T^{-1}$).
α_L	Longitudinal dispersivity (L)
α_T	Transverse dispersivity (L)
∇	Del operator
ϕ	Hydraulic head (L)
λ	Regularization factor
ε	Residuals
ε_f	Scaling factor

1. Introduction

The pump and treat (PAT) method is a physical process to remediate contaminated groundwater. Usually, simulation-optimization (S-O) models are used for designing PAT-based systems to remediate contaminated groundwater (McKinney and Lin 1994; Seyedpour 2019). In the S-O modeling, the simulation model is executed by the optimization model repeatedly to compute optimal values of the control parameters (i.e., pumping locations, extraction rates, injection rates) of the PAT system for achieving the remediation goals (McKinney and Lin 1994; Jiang and Na 2020; Majumder and Eldho 2020). Usually, surrogate/proxy simulators are used to approximate computationally expensive finite difference or finite element based simulation models to accelerate computational performance (Jiang and Na 2020). The most common surrogate simulators are the extreme learning machine (ELM), support vector machine (SVM), feed-forward neural network (FFNN), and random vector functional link (RVFL) (Pao et al. 1994; Kumar et al. 2013; Yadav et al. 2016; Majumder and Eldho 2020).

Among them, the RVFL network is quite popular due to its universal approximation ability of any continuous function with compact input-output datasets and high computational performance (Scardapane et al. 2015; Zhang and Suganthan 2016; Elaziz et al. 2020). The network structure of RVFL is similar to the extreme learning machine (ELM), and single-layer FFNN, except a direct link, exists between the input layer and the output layer (Pao et al. 1994). Such direct links between the input and output layers are standard practice in neural networks to avoid over-fitting issues during the training phase (Pao et al. 1994). In RVFL, the input weights and hidden layer thresholds are randomly sampled from a uniform distribution, and the output weights are computed by minimizing a loss function using Moore-Penrose pseudo-inverse or ridge regression (Zhang and Suganthan 2016). The RVFL network shows better generalization ability and superior computational efficiency than the SVM and single-

92 layer FFNN (Zhang and Suganthan 2016). Nevertheless, the training dataset with outliers
93 may reduce the generalization ability of the RVFL network. This issue can be resolved by
94 incorporating the weighted least square (WLS) approach in the RVFL network, where
95 weights are computed using the kernel density estimator (KDE) (Dai et al. 2015).

96 Some of the popular optimization methods for groundwater remediation and
97 management are cat swarm optimization (CSO), genetic algorithm (GA), particle swarm
98 optimization (PSO), grey wolf optimizer (GWO), differential evolution (DE), simulated
99 annealing (SA), harmony search (HS) and tabu search (TS) (McKinney and Lin 1994;
100 Sidiropoulos and Tolikas 2008; Tamer Ayvaz 2009; Kumar et al. 2013; Yang et al. 2013; Luo
101 et al. 2014; Majumder and Eldho 2016, 2020; Zhao et al. 2020). Recently, a metaheuristic,
102 namely the marine predator algorithm (MPA), has been proposed by emulating the
103 cooperative hunting strategy of the marine predators to capture prey (Faramarzi et al. 2020).
104 The performance of MPA is superior to various other metaheuristics (GA, DE, SA, and PSO)
105 in terms of exploration and exploitation ability (Faramarzi et al. 2020). Since its invention,
106 MPA has been getting lots of attention in various fields such as structural engineering,
107 electrical and power engineering, and energy (Elaziz et al. 2020; Ridha 2020; Sun et al.
108 2020). The performance of MPA can be further improved by various strategies such as:
109 incorporating elite opposition based learning, using biological evolution operators, and
110 incorporating elimination mechanisms (Wang and Li 2019; Dhargupta et al. 2020).

111 In groundwater remediation, decision-makers often encounter several conflicting
112 objectives, such as minimizing the cost of remediation, minimizing cleanup time, maximizing
113 reliability, and minimizing health risks (Erickson et al. 2002; Yang et al. 2013). Some of the
114 most common multi-objective optimization techniques are non-dominated sorting genetic
115 algorithm (NSGS-II) (Deb et al. 2002), multi-objective grey wolf optimizer (Mirjalili et al.
116 2016), multi-objective particle swarm optimization (MOPSO) (Coello et al. 2004), Pareto

117 archived evolution strategy (PAES) (Knowles and Corne 2000), and Niched Pareto tabu
118 search (NPTS) (Yang et al. 2013). A Pareto dominance scheme is used in all these algorithms
119 to obtain non-dominated/ Pareto-optimal solutions (trade-off among the conflicting
120 objectives) with good convergence and a selection criterion based on density to promote good
121 diversity among the non-dominated solutions. Now, the question arises, do we need another
122 new metaheuristic as there are so many already? The No-free launch (NFL) theory logically
123 proved that it is theoretically impossible to label a metaheuristic as the universal optimizer. A
124 metaheuristic showing superior performance for one class of problems may show quite
125 inferior performance to another (Wolpert and Macready 1997). This NFL theory thus
126 motivates researchers to develop new metaheuristics in a quest to find a better one.

127 Determining the optimal locations of pumping wells is crucial for designing an
128 effective PAT-based groundwater remediation system (Wang and Ahlfeld 1994; Seyedpour
129 2019). In most of the previous groundwater remediation studies on multi-objective
130 optimization, the pumping well locations are predefined/guessed along the centreline of the
131 plume for a homogeneous aquifer with a uniform and narrow plume (Erickson et al. 2002;
132 Yang et al. 2013, 2017; Luo et al. 2014; Jiang and Na 2020). However, determining optimal
133 pumping location is not straightforward for highly heterogeneous aquifers with wider
134 contaminant plume (Wang and Ahlfeld 1994; Huang and Mayer 1997; Guan and Aral 1999).
135 In a few previous studies, groundwater remediation systems were designed by incorporating
136 both the pumping rates and pumping well locations as continuous decision variables in the
137 management model (Wang and Ahlfeld 1994; Guan and Aral 1999). The Hermite
138 interpolation function is used to convert discrete pumping well locations of FDM into a
139 continuous function in space (Wang and Ahlfeld 1994; Guan and Aral 1999). However, the
140 interpolation step may lead to some amount inaccuracies in the management model.
141

Recently a novel two-step approach has been proposed for the remediation of contaminated groundwater that segregates a single objective optimization problem into two management models. The first management model uses the combinatorial optimization method to find optimal pumping locations (while keeping the extraction rate as a predefined constant value) by minimizing the amount of pollutant mass in the aquifer. The optimal locations of pumping wells obtained in the first step were used explicitly in the second step to achieve the remediation goal. The two-step approach can overcome the 'curse of dimensionality issue and is thus suitable for high-dimensional optimization problems (Mirjalili et al. 2014). The two-step approach was found to be computationally more efficient and accurate than the traditional approach. However, the efficacy of the two-step approach was not evaluated for multi-objective groundwater remediation. Partially motivated by this, the present study aims to check the efficacy of the two-step approach for multi-objective groundwater remediation in the highly heterogeneous aquifer. The present study also has many other novel features, as discussed in the next section.

The present study attempts to develop a two-step approach-based simulation-optimization model for multi-objective groundwater remediation using enhanced random vector functional link (ERVFL) and evolutionary marine predator algorithm (EMPA). The ERVFL network uses the weighted least squares (WLS) method to improve the robustness of the algorithm, where weights are computed using the kernel density estimator (KDE). The ERVFL network is used to approximate simulation models representing groundwater flow and contaminant processes to enhance computational performance. Further, EMPA is proposed by modifying the marine predator algorithm using elite opposition-based learning, biological evolution operator, and elimination mechanism. In the multi-objective version of EMPA, the archive controller and adaptive grid mechanism store the non-dominated

solutions in an external repository. Further, the single-objective and multi-objective versions of EMPA are used for groundwater remediation using the two-step approach.

2. Methodology

This study used the MODFLOW and MT3DMS codes for simulating groundwater flow and contaminant transport processes through the subsurface (Zheng and Wang 1999; Harbaugh, Arlen 2005). Further, a proxy model based on the ERVFL network is developed to approximate numerical flow and transport models. This study uses EMPA as an optimization model for the remediation of contaminated groundwater. The detailed methodologies discussing ERVFL and EMPA are discussed below.

2.1. Random vector functional link (RVFL)

The RVFL network consists of one hidden layer with additional direct links between input layers and hidden layers (Fig. 1). The direct links help avoid overfitting issues during the training phase (Pao et al. 1994; Vuković et al. 2018).

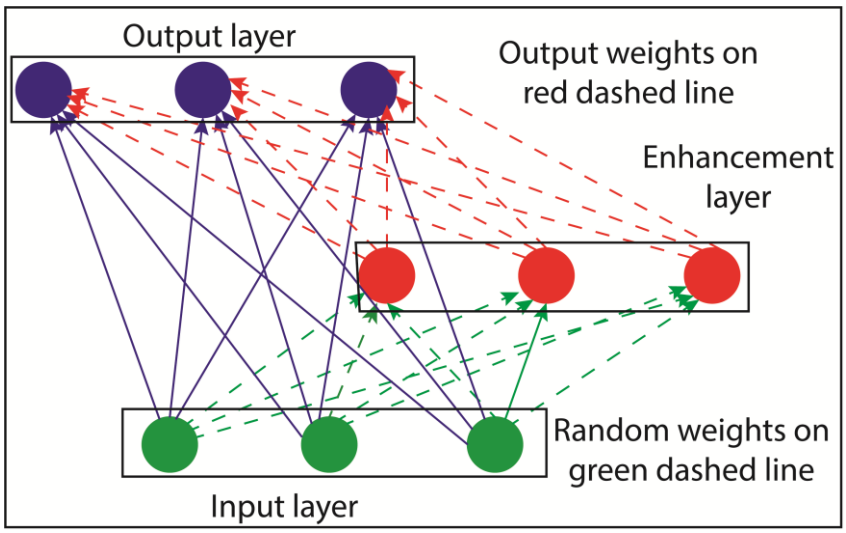

Fig.1. Pictorial representation of random vector functional link

Let us consider a distinct input-output dataset of dimensions n_d,

$$(x_j, y_j) \in \Re^n \times \Re^m; \quad j = 1, 2, ...n_d \tag{1}$$

Where x_j and y_j are the input and output datasets and can be represented as,

$$x_j = [x_{j1},\ x_{j2},\ x_{j3}, ...x_{jn}]_n^T \in \Re^n \tag{2a}$$

$$y_j = [y_{j1},\ y_{j2},\ y_{j3}, ...y_{jm}]_m^T \in \Re^m \tag{2b}$$

In RVFL, one fraction of the output is predicted by functional mapping of generic input x_j to a linear combination of a fixed number (L) of nonlinear transformations of the input itself. The other fraction of the output is predicted using the direct links between the input and the output layer (Vuković et al. 2018).

$$O_j = \sum_{i=1}^{L} \beta_i \times g(w_i \bullet x_j + b_i) + \sum_{k=1}^{n} \beta_k \times x_j \qquad j = 1, 2, ...N_d \tag{3}$$

Where L is the total hidden neurons; g is the activation function; $w_i = [w_{1i},\ w_{2i}, ...w_{ni}]_{n+L}^T$ is the input weight vector connecting hidden neuron i to the input neurons; b_i is the threshold of the hidden neuron i; $\beta_i = [\beta_{i1}, \beta_{i2}, ...\beta_{im}]_m^T$ is the output weight vector connecting hidden neuron i to output neurons; $\beta_k = [\beta_{k1}, \beta_{k2}, ...\beta_{km}]_m^T$ is the output weight vector connecting output neurons with input neuron k and $O_j \in R^m$ is the predicted output.

In matrix form, the Eq. (3) can be expressed as,

$$O = H\beta \tag{2a}$$

$$H = [H_1\ H_2]_{N \times (L+n)} \tag{2b}$$

$$\beta = [\beta_i\ \beta_k]_{(L+n) \times m} \tag{2c}$$

$$H_1 = \begin{bmatrix} g(w_1 X_1 + b_1) & \bullet & \bullet & g(w_L X_1 + b_L) \\ \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ g(w_1 X_N + b_1) & \bullet & \bullet & g(w_L X_N + b_L) \end{bmatrix}_{N \times L} \tag{2d}$$

$$H_2 = \begin{bmatrix} X_1 \\ X_2 \\ \bullet \\ X_N \end{bmatrix}_{N\times n} \tag{2e}$$

$$O = [O_1, O_2, O_3, ...O_N]^T_{N\times m} \tag{2f}$$

Where H_1 is the input matrix; H_2 is the hidden layer output matrix; H is a concatenated matrix of H_1 and H_2 ; O is the predicted output matrix.

The loss function is defined as the squared difference of the predicted output ($O = H\beta$) and actual output (Y) (Vuković et al. 2018).

$$Minimize \parallel H\beta - Y \parallel^2 \ \& \parallel \beta \parallel \tag{3}$$

The Eq.(3) can be formulated as a regularized least square (ridge regression) problem, which minimizes the loss function and norm of the output weights (Vuković et al. 2018).

$$\min_{\beta\in\Re^L} J(\beta) = \left\{ \frac{1}{2} \parallel H\beta - Y \parallel_2^2 + \frac{\lambda}{2} \parallel \beta \parallel_2^2 \right\} \tag{4}$$

Where, λ is known as the regularization factor, which help to minimize variance in the prediction of output weights (β)(Scardapane et al. 2015; Vuković et al. 2018).

The solution of Eq.(4) can be obtained by equating the gradient of $J(\beta)$ to zero (Vuković et al. 2018).

$$\frac{\partial J}{\partial \beta} = H^T H\beta - H^T Y + \lambda\beta = 0 \tag{5}$$

$$\hat{\beta} = \left(H^T H + \lambda I \right)^{-1} H^T Y \tag{6}$$

Where I is the identity matrix; H^T is the transpose matrix of H.

2.2. Enhanced random vector functional link (ERVFL)

The generalization ability of the RVFL can be further improved by incorporating the weighted least squares (WLS) approach into Eq.(4), where weights can be computed using kernel density estimator (KDE) (Dai et al. 2015).

$$\min_{\beta \in \Re^L} J_1(\beta) = P\left\{\frac{1}{2} \| H\beta - Y \|_2^2 + \frac{\lambda}{2} \| \beta \|_2^2\right\} \tag{7}$$

Where $P = diag(p_1, p_2, ..p_j ... p_N)$ is the penalty weight matrix representing the contribution of each sample to the loss function. Each element of the matrix p is computed according to the reliability of the sample. A suspected outlier will have very low reliability and thus a small p_j value. The assignment of a small p_j value to the outlier reduces its importance in the loss function (Eq.7). The residuals (ε) is computed using original RVFL as,

$$\varepsilon_j = \sum_{i=1}^{L} \beta_i \times g(w_i \bullet X_j + b_i) + \sum_{k=1}^{n} \beta_k \times X_j - Y_j \qquad j = 1, 2, ...N_d \tag{8}$$

The probability density function (PDF) of residuals can be computed using KDE as (Majumder and Eldho 2019)

$$f(x) = \frac{1}{N} \sum_{j=1}^{N} \frac{1}{h_b} \times K\left(\frac{x - \varepsilon_j}{h_b}\right) \tag{9}$$

Where $f(x)$ is the probability density function of the residuals; h_b is the bandwidth; N is the number of samples.

The kernel smoothing function (K) assuming it as Gaussian can be expressed as (Majumder and Eldho 2019),

$$K(x - \varepsilon_j, h_b) = \frac{1}{h_b \sqrt{2\pi}} e^{\left(-\frac{(x-\varepsilon_j)^2}{2h_b^2}\right)} \tag{10}$$

Further, a weight (p_j) is assigned to each residual (ε_j) with respect to the probability of residual, which is computed using KDE as,

$$p_j = f(\varepsilon_j) \tag{11}$$

After computing p_j, the output vector β is computed by minimizing the gradient of J_1 (Eq. 7) as,

$$\frac{\partial J_1}{\partial \beta} = PH^T H\beta - H^T PY + P\lambda\beta = 0 \tag{12}$$

$$\hat{\beta} = \left(H^T pH + \lambda pI\right)^{-1} H^T pY \tag{13}$$

2.3. Marine predator algorithm (MPA)

MPA imitates the cooperative foraging behavior of marine predators using the Brownian and Lévy movements (Faramarzi et al. 2020). The search processes of the MPA are divided into three phases with specified number of generations based on velocity ratio between predator and prey (Elaziz et al. 2020).

Exploration phase:

In the exploration phase, predators explore the search space using Brownian motion with high velocity in pursuit of finding suitable prey (Eq.14) (Faramarzi et al. 2020).

$$x_i^{t+1} = x_i^t + p_c \times r \times R_B(x_g^t - R_B \times x_i^t) \qquad 0 < t \le \frac{1}{3} \times t_{\max} \tag{14}$$

Where, $t_{\max}$ is the total number of iterations; x_i^t is the position of a candidate marine predator; x_i^{t+1} is new position of candidate marine predator; x_g^t is the position of best marine predator; p_c is a constant having value 0.5; $r \in U[0,1]$ and $R_B \in N[0,1]$.

Transition phase:

In this phase, the velocities of half of the marine predators are reduced using a parameter C_F to exploit the search space. The other half of the predators with higher velocities explore the

search space. A factor $p_f \in U[0,1]$ is used to distribute the population for exploration and exploitation. The position of a candidate predator is updated as (Faramarzi et al. 2020).

$$x_i^{t+1} = x_i^t + p_c \times r \times R_L (x_g^{t+1} - R_L \times x_i^t) \qquad if \;\; p_f \geq 0.5 \;\; \& \left(\frac{1}{3} \times t_{\max} < t \leq \frac{2}{3} \times t_{\max} \right) \qquad (15a)$$

$$x_i^{t+1} = x_g^t + p_c \times C_F \times R_B (R_B \times x_g^{t+1} - x_i^t) \qquad if \;\; p_f < 0.5 \;\; \& \left(\frac{1}{3} \times t_{\max} < t \leq \frac{2}{3} \times t_{\max} \right) \qquad (15b)$$

$$C_F = \left(1 - \frac{t}{t_{\max}} \right)^{\frac{2t}{t_{\max}}} \qquad (15c)$$

Where, R_L is a random number generated using Lévy distribution.

Exploitation

In this phase, the velocities of marine predators are low due to the use of the parameter CF. The position of a candidate marine predator is updated according to the equation below (Faramarzi et al. 2020).

$$x_i^{t+1} = x_g^t + p_c \times C_F \times R_L (x_g^{t+1} - R_L \times x_i^t) \qquad Where \qquad \left(\frac{2}{3} \times t_{\max} < t \leq t_{\max} \right) \qquad (16)$$

Eddy formation and fish aggregating devices' effect (F_{AD})

Environmental factors such as eddy formation and fish aggregating devices' (F_{AD}) may significantly affect the behavior of marine predators (Faramarzi et al. 2020). In a previous study, it is mentioned that 80% of the time, Shark spends in the vicinity of FADs, and the rest of the time, Shark makes many long jumps in various directions in a quest to search for prey. Such behavior of Shark can be incorporated in the MPA algorithm to avoid local minima entrapment. By considering eddy formation and F_{AD}, the positions of a candidate marine predator are updated as (Faramarzi et al. 2020)

$$x_i^{t+1} = x_i^t + C_F \{ x_{\min} + r \times (x_{\max} - x_{\min}) \} \times W \quad if \; r < F_{AD} \qquad (17a)$$

$$x_i^{t+1} = x_i^t + [F_{AD}(1-r) + r] \times (x_{r_1} - x_{r_2}) \qquad if \; r > F_{AD} \qquad (17b)$$

Where, $x_{\max}$ and $x_{\max}$ are vectors which contain lower and upper values for each dimension of candidate predator; $r \in U[0,1]$; FAD represent the probability of considering the effect of fish aggregating device with a value 0.2; W is a vector of binary variables 0 and 1. The elements of W is constructed by generating a random number from $r \in U[0,1]$ and setting the value of each element to zero if the element is smaller than FAD. Otherwise, set the value of the element to one if the element is greater than FAD (Faramarzi et al. 2020).

Marine Memory

During the search process, marine predators save the best position (x_g) obtained so far, and the corresponding fitness value ($f(x_g)$)(Faramarzi et al. 2020).

$$\begin{aligned} &if \quad f(x_g^{t+1}) < f(x_g) \quad then \\ &x_g = x_g^{t+1} \\ &f(x_g) = f(x_g^{t+1}) \end{aligned} \tag{18}$$

2.4. Elite opposition based learning (EOBL)

The EOBL is a standard procedure to improve the solutions in the field of artificial intelligence (Dhargupta et al. 2020). In EOBL, for every candidate solution, an opposite point (also known as the point of reflection) is generated using central symmetry (Soncco-Álvarez et al. 2019). Among them, the solution with better fitness value is selected for the next generation. Let us assume a candidate solution $x_i = [x_{i,1}, x_{i,2}, ...x_{i,D}]$ and the corresponding opposite point $x_e = [x_{e,1}, x_{e,2}, ...x_{e,D}]$. The following equation is used to compute the opposite point of the candidate solution x_i.

$$x_{i,j}^e = [lb_j + ub_j] - x_{i,j} \qquad j = 1, 2.....D \tag{19}$$

Where $x_{i,j} \in [lb_j, ub_j]$ is the j[th] element of the i[th] candidate; $x_{i,j}^e$ is the opposite point (solution); and ub_j and lb_j are the upper and lower bound of the j[th] element.

2.5. Biological evolution

Biological evolution is a natural process that makes living beings stronger/fitter over successive generations. The biological evolution process of marine predators can be modeled mathematically using differential evolution operators (mutation, crossover, selection) (Wang and Li 2019).

Mutation:

For each candidate solution (x_i^t), select three other candidates randomly from the population. Generate a muted solution (y_i^t) using the following equation (Storn and Price 1997).

$$y_i^t = x_{r1}^t + m \times (x_{r2}^t - x_{r3}^t) \qquad i = 1, 2, ...N \;\; \& \;\; t = 1, 2, ...T; \;\; x_{r1}^t \neq x_{r2}^t \neq x_{r3}^t \tag{20}$$

Where x_{r1}^t, x_{r2}^t, and x_{r3}^t are three distinct candidates selected randomly from the population; $m \in U[0.2, 0.8]$ is the mutation scaling factor; N is the size of the population; T is the total number of generation.

Crossover:

In this phase, generate an offspring solution (z_{ij}^t) by exchanging the elements of muted solution vector (y_{ij}^t) with the elements of candidate solution vector (x_{ij}^t) (Storn and Price 1997).

$$z_{ij}^t = \begin{cases} y_{ij}^t & C_R \geq r \;\; or \;\; j = d_r \\ x_{ij}^t & C_R < r \end{cases} \qquad i = 1, 2, ...N; \;\; j = 1, 2, ...D \tag{21}$$

Where $C_R = 0.2$ is the crossover operator; D is the dimension of the candidate/muted solution vector; $r \in U[0, 1]$; $d_r \in U[1, D]$.

Selection:

The candidate (x_i^t) and offspring solutions (z_i^t) are compared with respect to their respective fitness value, and the better one is selected for the next generation (Storn and Price 1997).

$$x_i^{t+1} = \begin{Bmatrix} z_i^t & f(z_i^t) < f(x_i^t) \\ x_i^t & f(z_i^t) \geq f(x_i^t) \end{Bmatrix} \quad i = 1, 2, ...N \tag{22}$$

Where $f(x_i^t)$ is the fitness function value computed at the point x_i^t .

2.6. Elimination mechanism

The survival of the fittest (SOF) principle states that only the fitter candidates survive and vulnerable candidates die in the successive generation due to various natural reasons (Wang and Li 2019). However, the new candidates will keep on joining the population due to natural birth or other reasons (Wang and Li 2019). The steps to implement the elimination mechanism in the MPA are stated below:

(a) Sort the candidates (population) in ascending order according to their respective fitness values.

(b) Eliminate some of the candidates with low fitness values. To facilitate this, generate a random integer (N_R) within the range $\left[\frac{N}{2\times\varepsilon_f}, \frac{N}{\varepsilon_f}\right]$.

Where, N is the size of the population (total number of candidates) and ε_f is the scaling factor. To preserve the fitter candidate for the successive generation, the value of ε should always be greater than 1.

(c) After removing N_R candidates, add the same number of new candidates with randomly assigned positions within the bounds of decision variables space.

2.7. Multi-objective Optimization (MO)

A multi-objective optimization (MO) problem, with more than one conflicting objectives problem can be expressed as (Deb 2012)

$$\left\{\begin{array}{lll} Minimize & F_m(x) & m=1,2,3,4,.....,M \\ Subject\,to & I_i(x)\geq 0 & i=1,2,3,4,.....,I \\ & E_k(x)=0 & k=1,2,3,4.....,K \\ & x_j^L \leq x_j \leq x_j^U & j=1,2,3,4,.....N \end{array}\right\} \quad (23)$$

Where, F_m is the m[th] objective function; $x=[x_1,x_2,x_3.....x_D]$ is a vector of decision variable of dimension D; N is the total number of decision variables; I_i are the j[th] linear/nonlinear inequality constraint; E_k are the k[th] linear/nonlinear equality constraint; x_j^L and x_j^U are the lower bound and upper bound of the decision variable x_i.

The MO finds the sets of solutions representing trade-offs between objectives known as Pareto optimal solution (Deb 2012).

Pareto dominance

Let us assume two solution vectors $x=[x_1,x_2,x_3.....x_D]$ and $y=[y_1,y_2,y_3.....y_D]$. The corresponding objective function vectors are $f(x)=[f_1^x, f_2^x ... f_m^x]$ and $f(y)=[f_1^y, f_2^y ... f_m^y]$. The solution vector x dominate y (denoted as $x \prec y$) if and only if (Mirjalili et al. 2016) :

$$\forall i \in \{1,2,.....D\}: f(x_i) \leq f(y_i) \wedge \exists i \in \{1,2,.....D\}: f(x_i) < f(y_i) \quad (24)$$

Pareto optimality

If a solution vector $x \in R^D$ is not dominated by any other solution in the feasible region of search space, then x is the Pareto optimal solution (Mirjalili et al. 2016).

$$\nexists y \in R^D \,|\, y \prec x \quad (25)$$

Pareto optimal set

The Pareto optimal set is the set of all Pareto optimal solutions (Mirjalili et al. 2016).

$$P_s = \{x \in R^D \mid \nexists\ y \in R^D, y \prec x\} \quad (26)$$

Pareto optimal front

The Pareto optimal front is the projections of the Pareto optimal set in the objective functions space (Mirjalili et al. 2016).

$$P_f = \{f(x) \mid x \in P_s\} \quad (27)$$

External repository

This study used an external repository (archive) to store non-dominated solutions using an archive controller and adaptive grid mechanism (Coello et al. 2004; Mirjalili et al. 2016).

Archive controller:

The archive controller controls the entry of non-dominated solutions into the repository during the course of iteration based on the following conditions (Mirjalili et al. 2016). Let us consider a non-dominated solution N_s willing to enter into the repository.

Case 1: Add N_s into the external repository if the archive is empty.

Case 2: If N_s is dominated by one or more member in the external repository, then discard N_s.

Case 3: If N_s is not dominated by any member in the external repository, then directly add it to the repository.

Case 4: If one or more members in the external repository are dominated by N_s then remove the dominated members and add N_s to the repository.

Adaptive grid mechanism

The purpose of the adaptive grid mechanism is to maintain diversity among non-dominated solutions (Mirjalili et al. 2016).

Case 5: If N_s fulfill the criterion (neither N_s nor any individual in the repository dominate each other) to enter into the external repository but the repository is full, then invoke the

adaptive grid mechanism. To make space for N_s, remove an individual from the most crowded segment of the repository using the roulette wheel selection method. If N_s lies outside the current bound of the grid, then recalculate the grid and relocate each individual of the grid (repository) to accommodate N_s. The removal of individuals from the most crowded space of the repository helps to improve the diversity of the Pareto optimal solutions. Various parameters are used to implement the adaptive grid mechanism such as grid inflation parameter ($\alpha = 0.1$), the number of grid per each dimension ($n_{grid} = 10$), best individual selection pressure parameter ($\beta = 4$), and repository individual selection parameter ($\gamma = 2$) (Liu et al. 2020).

All the possible cases discussed above are also be represented in pictorial form (Fig.2).

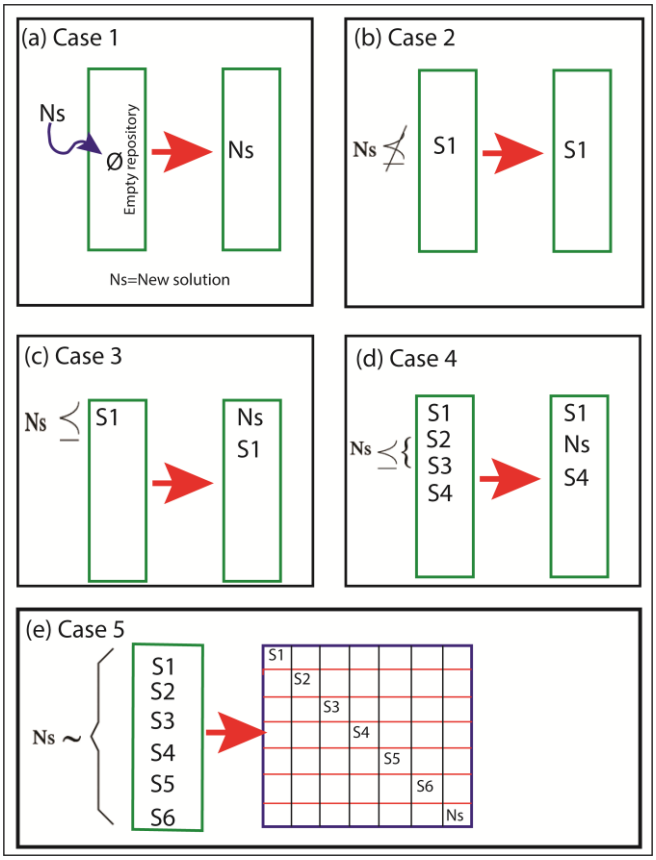

Fig. 2. Pictorial representation of all the possible cases of archive controller

Selection of best individual

The best individual selection mechanism chooses the least crowded segment of the search space using the roulette wheel selection method to emphasize better exploration (Mirjalili et al. 2016).

Performance metrics

In multi-objective optimization, performance matrices such as inverted generational distance (IGD), spacing (SP), and maximum spread (MS) are used to check the convergence and diversity of the non-dominated solutions with respect to the true Pareto-optimal solutions (Deb 2012).

3. Formulation of management models for groundwater remediation

The amount of contaminant mass extracted from the aquifer with a pumping strategy can be expressed as,

$$M_e = M_0 - M_r \tag{28}$$

Where M_0 is the total amount of contaminant mass in the aquifer before pumping, M_e is the amount the contaminant mass extracted from the aquifer with a pumping strategy, and M_r is the remaining amount of contaminant mass in the aquifer after pumping. Mathematically, M_0 and M_r can be expressed as,

$$M_0 = \sum_{i=1}^{N_g} \eta \times \Delta x_i \times \Delta y_i \times \phi_{i0} \times c_{i0} \tag{29a}$$

$$M_r = \sum_{i=1}^{N_g} \eta \times \Delta x_i \times \Delta y_i \times \phi_i \times c_i \tag{29b}$$

Where N_g is the number of spatial grids representing the aquifer domain; η is the porosity of the porous media; $\Delta x_i \times \Delta y_i$ is the area of the rectangular grid 'i' [L^2]; ϕ_{i0} is the hydraulic head value at grid 'i' prior to pumping [L]; c_{i0} is the contaminant concentration at grid 'i' prior to pumping [ML^{-3}]; ϕ_i is the hydraulic head value at grid 'i' after pumping [L]; c_i is the contaminant concentration value at grid 'i' after pumping.

The present study proposes two management models for the remediation of contaminated groundwater. In the first management model (Eq.30), the objective is to find optimal pumping locations with a constant pumping strategy by minimizing the percentage of remaining contaminant mass in the aquifer.

$$Min\ OF_0 = \frac{M_r}{M_0} \times 100 \tag{30a}$$

$$r_i \in [1, 2.....N_c] \tag{30b}$$

476 $$W_r = [r_1, r_2, ...r_i.....r_D] \quad (30c)$$

477 $$r_1 \neq r_2 \neq ..... \neq r_i ..... \neq r_D \quad (30d)$$

478 $$\phi_i = f(Q_{r_1}^1, Q_{r_2}^2, ...Q_{r_i}^i .....Q_{r_D}^D) \quad (31e)$$

479 $$c_i = f(Q_{r_1}^1, Q_{r_2}^2, ...Q_{r_i}^i .....Q_{r_D}^D) \quad (30f)$$

480 $$Q^1 = Q^2 = ... = Q^i ..... = Q^n = K_Q \quad (30g)$$

481 Where r_i is an integer generated randomly within the range $(1, N_c)$; N_c is the total number of
482 grids representing the locations of candidate pumping wells; W_r is a vector representing the
483 locations of pumping wells which are randomly generated; D is the total number of active
484 pumping wells; $Q_{r_i}^i$ denotes that in the r_i^{th} grid a pumping rate of value Q^i is assigned; and
485 K_Q is the constant pumping rate.

486 The second management model considers minimizing two conflicting objectives
487 simultaneously. The first objective minimizes water extraction rates, which can indirectly
488 minimize the pumping cost and treatment cost. The second objective is to minimize the
489 amount of contaminant mass in the aquifer.

490 $$Minimize\ OF_1 = \left(\sum_{t=1}^{T} \sum_{i=1}^{D_w} \left| Q_{i,t} \right| \right) \quad (31a)$$

491 $$Minimize\ OF_2 = \left(\frac{M_r}{M_0} \times 100 \right) \quad (31b)$$

492 $$Q_i^{\min} \leq Q_{i,t} \leq Q_i^{\max} \quad (31c)$$

493 $$d_i^T < d_{\max} \qquad i = 1, 2, 3, ...D \quad (31d)$$

494 $$c_i^T < c_{\max} \qquad i = 1, 2, 3.....D \quad (31e)$$

495 Where, T is the number of stress periods; D_w number of active pumping wells; $Q_{i,t}$
496 represent pumping rate from the i$^{\text{th}}$ extraction well in the t$^{\text{th}}$ stress period; $Q_i^{\min}$ and $Q_i^{\max}$ are

497 the upper and lower bounds of pumping rates for the ith pumping well; d_i^T is the hydraulic
498 drawdown in the location of ith extraction well; c_i^T is the contaminant concentration in the
499 location of ith extraction well; $d_{\max}$ is the permissible hydraulic drawdown; $c_{\max}$ is the
500 permissible contaminant concentration; and f_c is a multiplying factor which is used to give
501 the same weightage to both terms of the composite function.
502
503 **4. Model development**
504 In this study, finite difference method is used to simulate the subsurface flow and
505 solute transport processes (Zheng and Wang 1999; Harbaugh, Arlen 2005). The groundwater
506 remediation process is divided into two steps: (i) Determination of optimal pumping location
507 for groundwater remediation (ii) Multi-objective optimization to obtain Pareto optimal
508 solution. Both the steps are discussed below.
509
510 **4.1. Combinatorial optimization to determine the location of optimal pumping wells**
511 Here, the simulation model is directly integrated with EMPA to develop the S-O
512 model. The steps to find optimal pumping locations are:
513 i) Select a set of grids ($N_c = 231$) as the locations of candidate pumping wells. The
514 set of candidate pumping wells can be expressed as: $S = [1, 2, 3, ...N_c]$.
515 ii) Assign the pumping rate (Q_i) to zero to each pumping location. Mathematically,
516 set $Q_i = 0 \quad where, \quad i \in [1, N_c]$
517 iii) Generate a vector W_r of dimension '$D(=15)$' representing the set of active
518 pumping wells. Each element of W_r is an integer generated randomly within the
519 range [1, 231].

iv) Assign a constant pumping rate (K_Q) to each active pumping locations.

v) Using the Eq.(30), compute the fitness function value.

vi) Compute optimal fitness value and corresponding vector of active pumping wells (W_v) by iteratively repeating steps (ii-v) using EMPA.

4.2. Multi-objective optimization to find the Pareto optimal solution

The optimal pumping locations obtained in the first step are directly used for multi-objective groundwater remediation. Here, a proxy simulator is developed by approximating the simulation models using the ERVFL network. The proxy simulator is further coupled with the multi-objective version of EMPA to generate Pareto optimal solutions for groundwater remediation. The steps are enumerated below.

(i) In the simulation model, set the pumping well locations using the vector W_v as obtained in the 1st step.

(ii) Repetitively execute the simulation model to generate the input-output dataset to train EVRVL based proxy model. The input dataset are the extraction rates ($Q \in [LB, UB]$) within specified upper and lower bound, and the output dataset is the respective hydraulic drawdown (d), contaminant concentration (c), and the percentage of contaminant mass $\left(\frac{M_r}{M_0} \% \right)$ in the aquifer.

(iii) Train an ERVFL based proxy simulator using the input-output dataset. Check the accuracy of the proxy simulator.

(iv) Integrate the ERVFL based proxy simulator with EMPA to develop the S-O model.

(v) The management model for multi-objective optimization is the form of Eq. (31). Use the S-O models for obtaining Pareto optimal front representing the

relationship between optimal pumping rates and the percentage of remaining contaminant mass in the aquifer.

5. Numerical Experiments

In this section, several numerical experiments are carried out to check the performance of the evolutionary marine predator algorithm (EMPA) for both single-objective and multi-objective optimization.

5.1. Single objective optimization

We here consider six composite functions to check the performances of the single-objective version of the EMPA with respect to the other metaheuristics (Particle Swarm Optimization-PSO, Cat Swarm Optimization-CSO, Differential Evolution-DE, Grey Wolf Optimizer-GWO, Marine Predator Algorithm-MPA). The composite functions are constructed by shifting, hybridizing, and rotating primitive multimodal and unimodal benchmark functions (Liang et al. 2005). The surface plot of the composite functions resembles many real-world optimization problems (Fig.S2-Supplementary). Due to the presence of massive numbers of local minima, the composite functions are ideal for checking the capability of metaheuristics to escape local minima as well as exploitation and exploitation ability. Three statistical parameters, viz. best, mean, and standard deviation, are used to measure the metaheuristics performances. Dataset for statistical analysis is generated by executing each metaheuristic fifty times for each benchmark function. The results of composite test functions are shown in Table 1. The results show that in most cases, the performance of EMPA is better than other metaheuristics in obtaining the optimal solutions.

568 **Table 1.** Statistical comparisons of the optimal solutions obtained by EHSMS and other metaheuristics for composite benchmark functions

Function	Statistical Parameters	PSO	CSO	DE	GWO	MPA	EMPA
F10	Best	88.08	0.0321	97.84	0.0472	2.26	0.066
	SD	91.90	117.37	7.737	74.26	68.32	72.28
	Average	**193.12**	**140.02**	**110.96**	**86.73**	**93.99**	**53.95**
	Rank	**6**	**5**	**4**	**2**	**3**	**1**
F11	Best	88.05	0.0077	70.70	2.40	7.62	0.0422
	SD	67.80	87.55	18.94	71.77	50.88	41.66
	Average	**164.94**	**90.01**	**107.28**	**76.76**	**73.04**	**61.42**
	Rank	**6**	**4**	**5**	**3**	**2**	**1**
F12	Best	95.19	0.0047	102.24	0.25	0.602	0.01358
	SD	68.11	115.46	11.44	94.21	45.59	63.9718
	Average	**193.72**	**100.011**	**119.14**	**72.75**	**62.73**	**53.3714**
	Rank	**6**	**4**	**5**	**3**	**2**	**1**
F13	Best	88.05	0.0098	62.00	0.8258	2.52	0.0133
	SD	57.34	131.65	18.94	36.41	30.06	40.68
	Average	**163.23**	**120.01**	**105.00**	**63.36**	**55.17**	**60.03**
	Rank	**6**	**5**	**4**	**3**	**1**	**2**
F14	Best	0.072	0.0034	65.50	0.7961	2.430	0.0111
	SD	79.78	113.53	19.57	99.98	51.67	61.41
	Average	**164.30**	**120.012**	**105.49**	**46.02**	**55.02**	**45.69**
	Rank	**6**	**5**	**4**	**2**	**3**	**1**
F15	Best	21.50	82.61	89.27	3.620	2.281	0.0115
	SD	62.97	18.81	10.98	51.94	68.31	63.24
	Average	**156.19**	**108.65**	**103.37**	**54.71**	**60.58**	**56.03**
	Rank	**6**	**5**	**4**	**1**	**3**	**2**

569

5.2. Multi-objective optimization

The performance of MO-EMPA in terms of convergence and diversity of the non-dominated solutions is compared with the NSGA-II. The parameters of MO-EMPA and NSGA-II are given in Table S6 (Supplementary). Following Got et al.(2020), two performance metrics namely inverted generational distance (IGD) and spacing (SP), are considered here. The IGD quantitatively measures the convergence of the approximate Pareto optimal front with respect to the true Pareto optimal front. On the contrary, the spacings are used to quantitatively measure the diversity of the non-dominated solutions in the approximate Pareto-optimal front. We here consider nine test functions from the literature (Zhang et al. 2009; Deb 2012). The test functions are: ZDT1 (convex), ZDT2 (convex), ZDT3 (discontinuous), UF_1 (convex), UF_4 (non-convex), UF_5 (discontinuous), UF_6 (discontinuous), UF_9 (multi-modal) and UF_{10} (multi-modal). The ZDT test problems are relatively easy to solve. However, the UF test functions are the most challenging problems in the literature of multi-objective optimization. It is noteworthy to mention here that the smaller value of IGD and SP indicates the better convergence and better diversity of the approximate Pareto front. Fig.3 shows the non-dominated solutions obtained by MO-EMPA and NSGA-II.

The MO-EMPA and NSGA-II are evaluated twenty times for each test function to compare them statistically in terms of IGD and SP metrics. The statistical results in the form of box plots for both IGD and SP metrics are shown in Fig. 4 and Fig. 5, respectively. In both the figure, the box plots indicate superior performance (better convergence behavior and better diversity) of MO-EMPA over NSGA-II for most of the test functions.

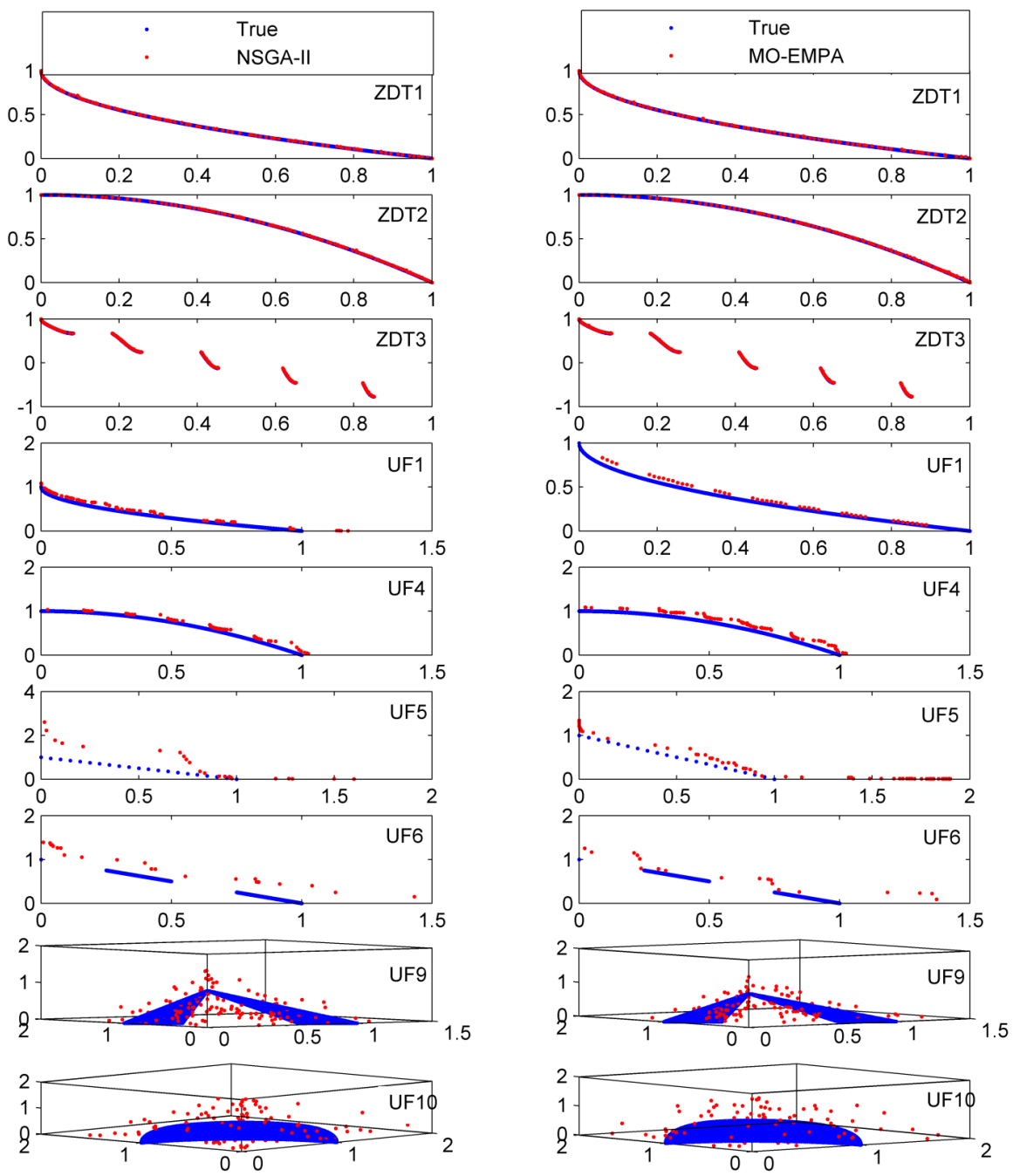

Fig.3. Approximate Pareto optimal fronts of the benchmark functions obtained by MO-EMPA and NSGA-II with respect to true Pareto optimal front

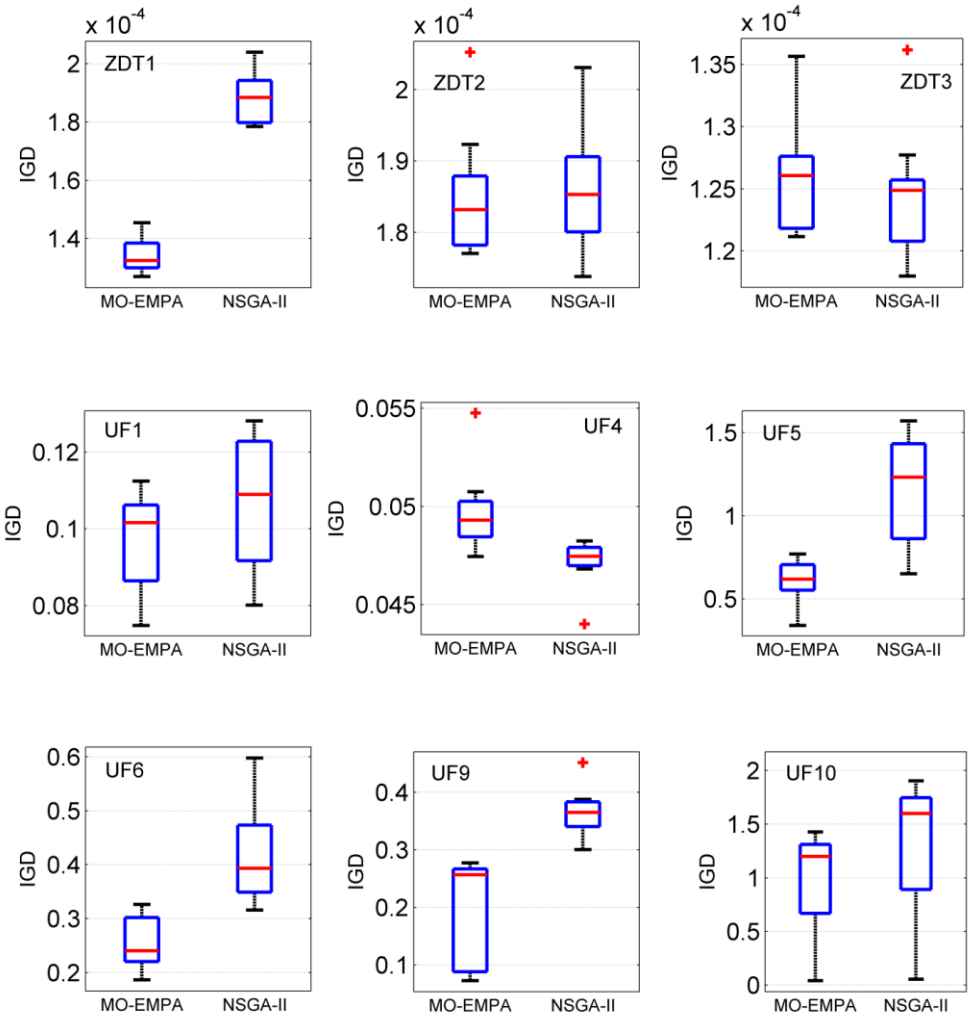

Fig.4. Box plot showing statistical results for IGD

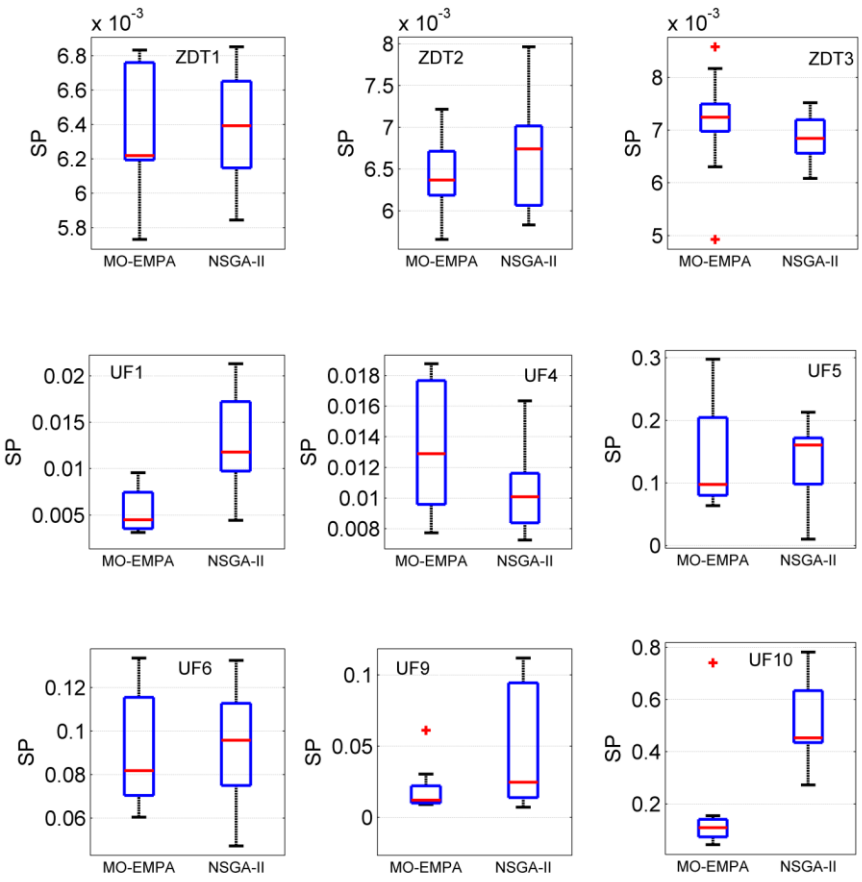

Fig.5. Box plot showing statistical results for SP

6. Case study

This study considers a highly heterogeneous unconfined aquifer for remediation of contaminated groundwater using the two-step approach. The domain of the aquifer and boundary conditions are almost identical to the East-Texas study area (Aquaveo 2018). However, many of the aquifer parameters are assumed and quite different than the East-Texas aquifer to make the aquifer more realistic and complex. The hydraulic conductivity field of the aquifer is generated using the truncated normal distribution shown in Fig.6. Fig.7a shows the aquifer domain with hydraulic conductivity fields and boundary conditions. Other aquifer parameters are: recharge (N_r) = 0.00006 m/d; bottom elevation (e_b) = 180 m; top elevation (e_t) = 230 m; thickness of the

aquifer (H_b) = 50 m; porosity (η) = 0.3; specific yield (S_y) = 0.2; transverse dispersivity (α_T) = 10 m; longitudinal dispersivity (α_L)= 50 m; total number of stress period (T) = 15; length of stress period (t) = 365 days; and time steps of each stress period (Δt) = 10 day. There are five injection wells in the aquifer that act as continuous contamination sources (Fig.7a). The injection wells are injecting contaminated water into the aquifer at a constant rate of 75 m^3/day for five years. Also, the concentration of contaminant in the injected water is 75000 µg /L. In this study, subsurface flow and transport of contaminants are simulated using MODFLOW and MT3DMS. The size of the spatial grid in the MODFLOW model is assumed to be 53.54 m×29.95 m. The contaminant contour after five year is shown in Fig.7b. Moreover, a zone for artificial recharge of area 720231.76 m^2 is considered to dispose of the extracted water during the remediation process (Fig.7a). The simulation models are further coupled with the metaheuristics based on the two-step approach to develop management models for remediating contaminated groundwater.

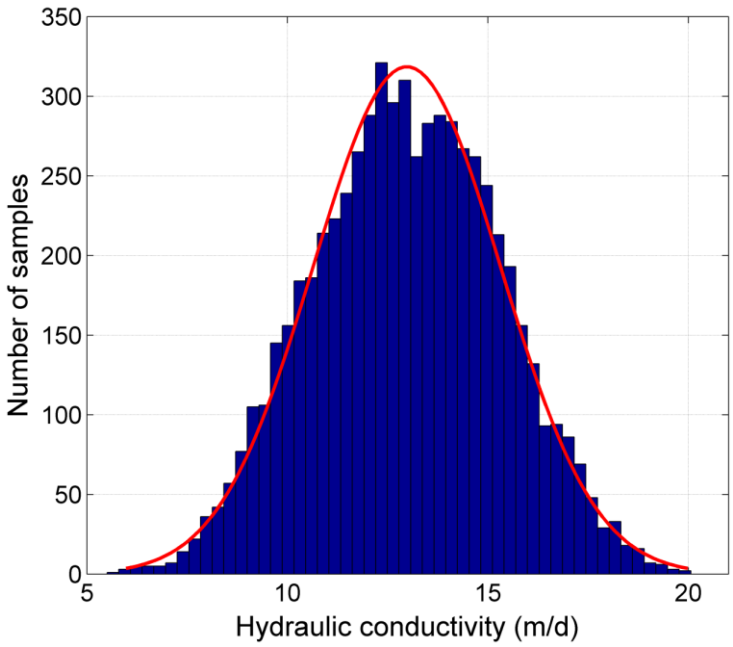

Fig.6. Histogram showing hydraulic conductivity distribution

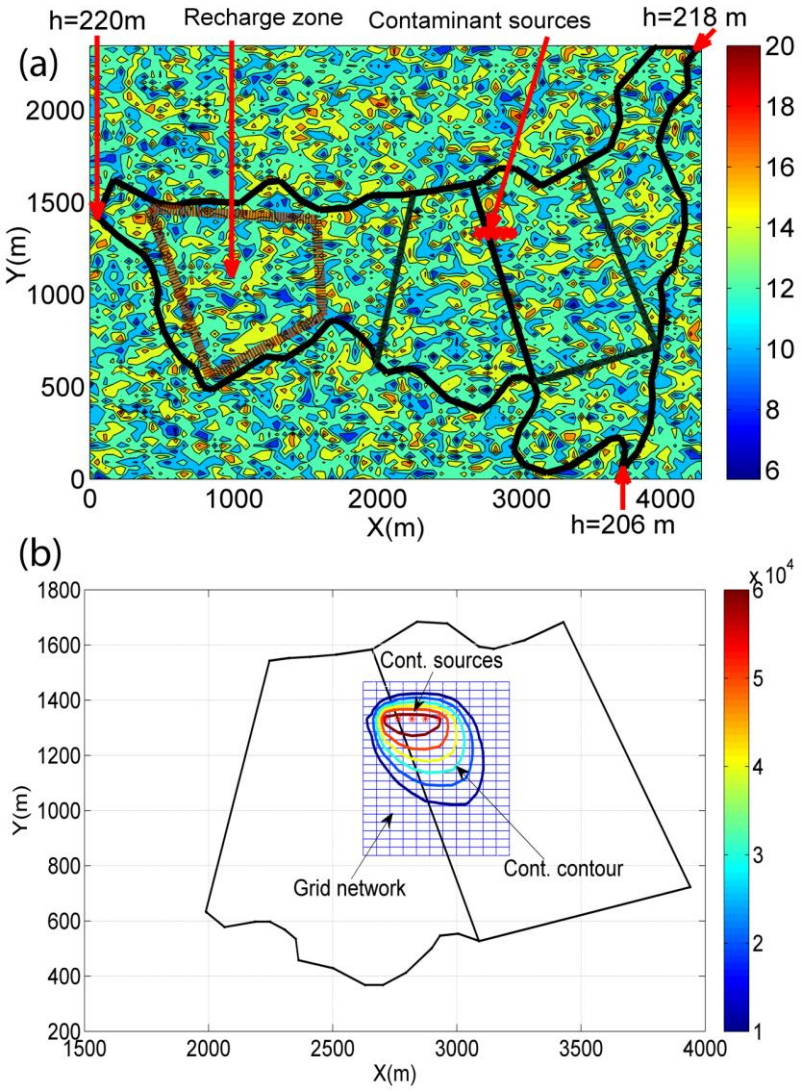

Fig.7. (a) An unconfined aquifer showing hydraulic conductivity field and boundary conditions **(b)** Contaminant contour due to continuous contamination for five years and rectangular nodes represnting probable pumping locations

6.1. Pumping well location optimization

The pumping well locations, which are the decision variables, are denoted by integers making it a combinatorial optimization problem. It should be noted that the pumping rates should be kept constant to find optimal locations for remediation. We here identified 231 finite-difference grids as the candidate pumping well locations [Fig.7b]. The aim here is to select 15

pumping wells out of 231 candidates by minimizing the Eq.(30). The total number of possible combinations is massive: ${}^{231}C_{15} = \frac{231!}{15!(231-15)!}$.

Due to such a massive number of combinations and highly nonlinear relationships between the decision variables (pumping well locations) and fitness values (amount of contaminant mass in the aquifer), the numerical simulation model cannot be approximated using a proxy simulator with a finite number of training data. Hence, the finite difference based simulation model is directly integrated with the metaheuristics to achieve the objective.

The pumping well locations obtained using EMPA and other metaheuristics are shown in Fig. 8. The fitness values corresponding to the optimal pumping locations are listed in Table 2. From the results, it can be deduced that the performance of EMPA is better than other metaheuristics in obtaining optimal pumping locations. Further, the performance of the metaheuristic are compared using the violin plots. A Violin plot is very similar to the box plot, except it also shows the probability distribution of numeric data using the kernel density estimator (KDE). In a Violin plot, the box and marker represent the interquartile range and median of the dataset, respectively. The whiskers represent the extreme values of the dataset, excluding outliers. The probability distribution plot also helps to assess whether the numeric dataset is sparse or multimodal. The dataset for violin plots is generated by executing each algorithm ten times to obtain the percentage of contaminant mass in the aquifer (Eq.30a). The violin plots obtained by EMPA and other metaheuristics are shown in Fig. 9. It is observed that the maiden value is least (minimum) for EMPA and maximum for PSO. The finding suggests the superior capability of EMPA in obtaining the median/average fitness value compared to other metaheuristics. The PSO is more prone to get entrapped in the local minima, which is the reason

for its poor performance. Further, we may get a fair idea of the data distribution by observing the interquartile range, whiskers, and probability density plots.

Table 2. Fitness values (OF_1) corresponding to optimal pumping locations obtained using EMPA and other metaheuristics **(Case study-2)**

	PSO	**CSO**	**DE**	**GWO**	**MPA**	**EMPA**
Contaminant mass remained- OF_0 **(%)**	15.010	14.948	14.634	14.2611	14.283	13.328
Rank	**6**	**5**	**4**	**2**	**3**	**1**

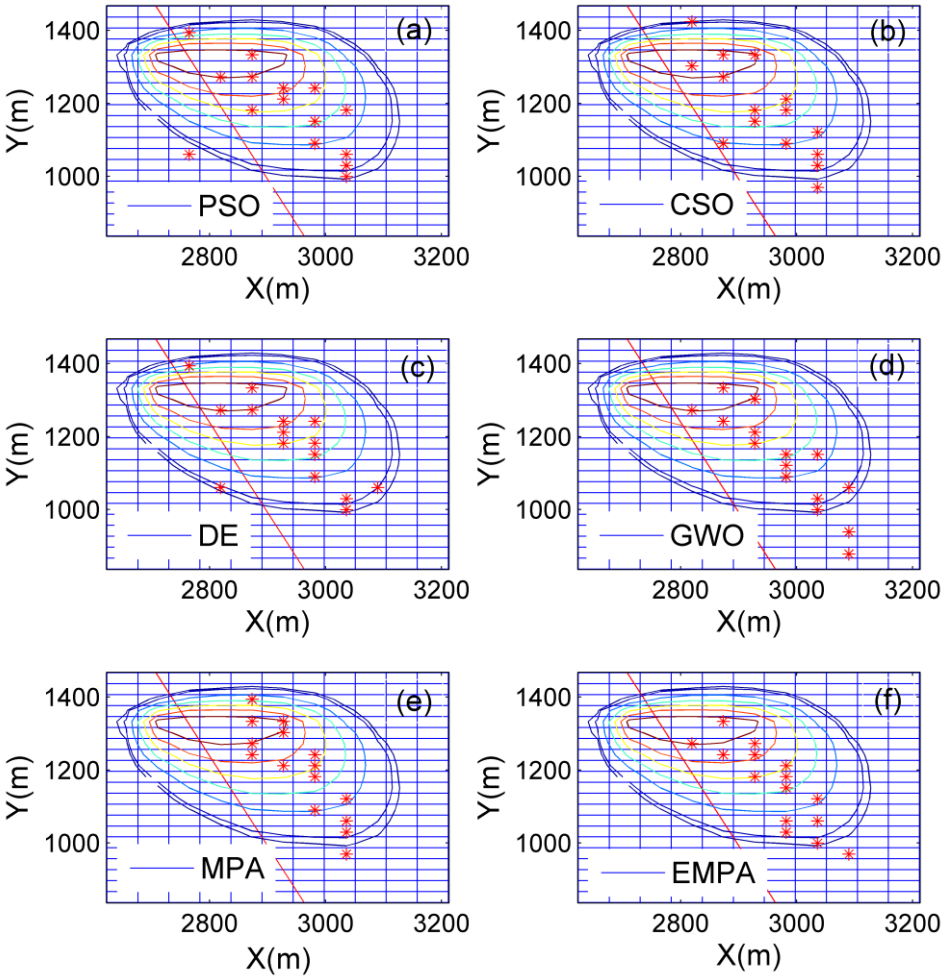

Fig.8. Optimal pumping locations (with contaminant contour prior to start pumping) obtained using EMPA and other metaheuristics

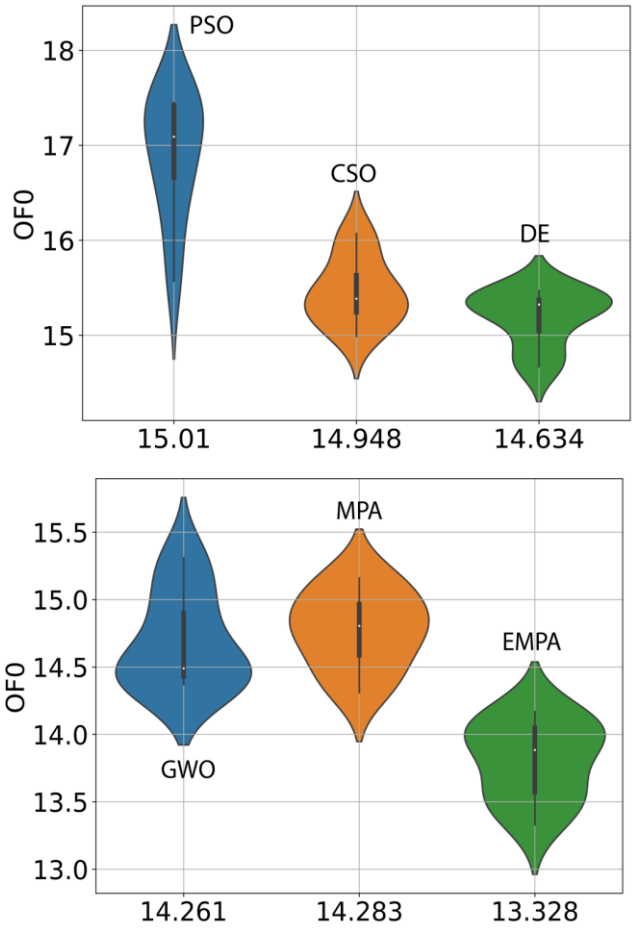

Fig. 9. Violin plot depicting the statistical results of the fitness value (OF_0) using EMPA and other metaheuristics

We further performed the Dunn and Sidák's test to compare the metaheuristics results statistically. The Dunn and Sidák's test results are interpreted in pictorial form (Fig.10a). In the figure, the lines represent the interval (range) of the data. A marker in the middle of the line shows the mean value of the dataset. Two different datasets are significantly different if their interval (line) is disjoint. The figure shows that the solution obtained by EMPA is significantly different and better in obtaining the optimal solutions than the other metaheuristics. Further, the convergence behavior of the three best algorithms (GWO, MPA, and EMPA) are shown in Fig. 10b

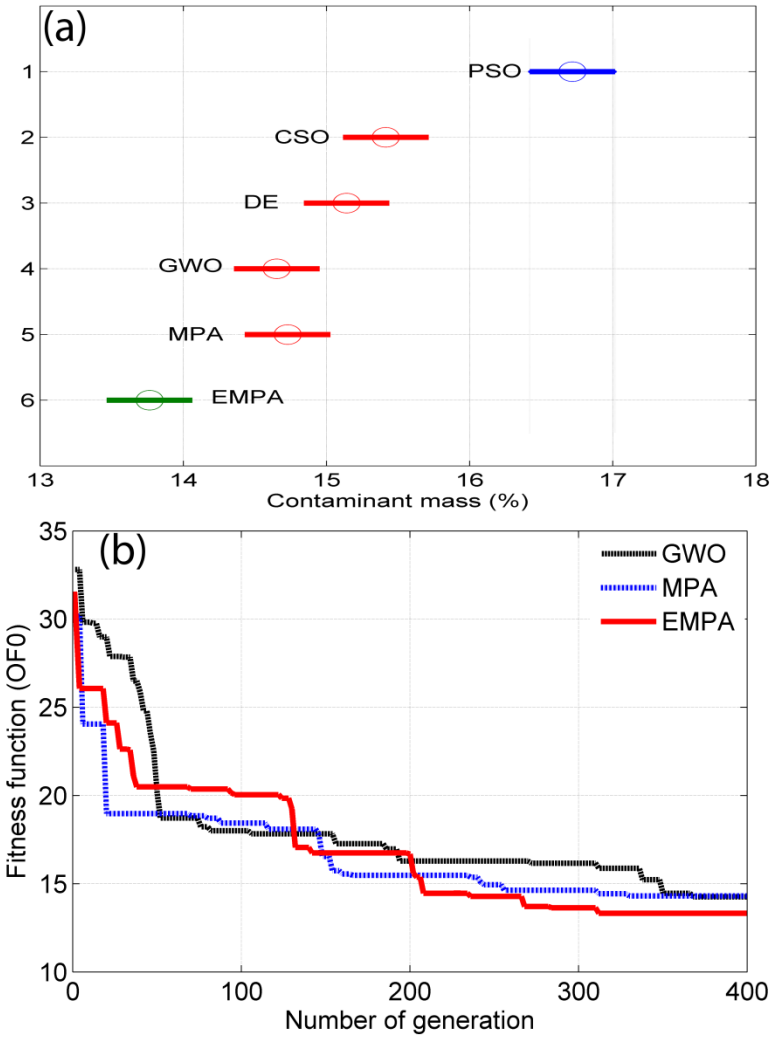

Fig.10.(a) Visual interpretation of Dunn and Sidák's test results **(b)** Variations of the fitness function (OF_0) with respect to iteration number

6.2. Multi-objective optimization for groundwater remediation

The optimal pumping locations obtained in the previous step are directly used here for multi-objective optimization. Here we first approximate the simulation models of groundwater flow and contaminant transport processes using the ERVFL network. The pumping rates (input dataset) are assumed to be in the range [-300, 0] m^3/day. The output datasets are the respective hydraulic drawdown, the contaminant concentration, and the percentage of extracted

contaminant mass from the aquifer. A total of 15000 input-output datasets are generated by repetitively executing the simulation models. The datasets are further subdivided into two parts to train and test the ERVFL network separately. The number of hidden neurons is assumed as 1000. Further, the cross-validation approach is used to compute the regularization parameter (C_r) (Scardapane et al. 2015; Vuković et al. 2018).

We search the regularization parameter (C_r) in the interval $2^j, j \in \{-15,-14,.....,15,14\}$ to obtain minimum RMSE of the testing dataset. Once we obtain the regularization parameter (C_r) in terms of 2^j, the value of C_r is further refined by searching around it. The cross-validation approach identified the values of the regularization parameter (C_r) as 0.76, 0.20, and 2.3 in approximating the hydraulic drawdown, contaminant concentration, and the total amount of extracted contaminant mass, respectively.

After training, the accuracy of the ERVFL network is further tested using the coefficient of correlation (R), and root mean square error ($RMSE$). In Fig. 11 (a, b, c), the coefficient of correlations in estimating hydraulic drawdown, contaminant concentration, and percentage of extracted contaminant mass using ERVFL are 0.992, 0.985, and 0.991, respectively. Also, in all the three cases, the RMSE values are small positive numbers [Fig. 11 (a, b, c)]. The results suggest the excellent generalization ability of the ERVFL model in approximating the numerical simulation model. The performance of ERVFL is further tested when the training dataset is corrupted due to outliers. We added one hundred outliers to corrupt the training dataset. Fig. 11 (d) shows very poor performance (R=0.793 and $RMSE$=758.40) of the original RVFL while computing the contaminant concentration. However, the performance of ERVFL is quite good when the training dataset is corrupted with outliers (R=0.962 and $RMSE$=53.40). The results

indicate the superior performance of ERVFL over RVFL in approximating datasets corrupted with outliers.

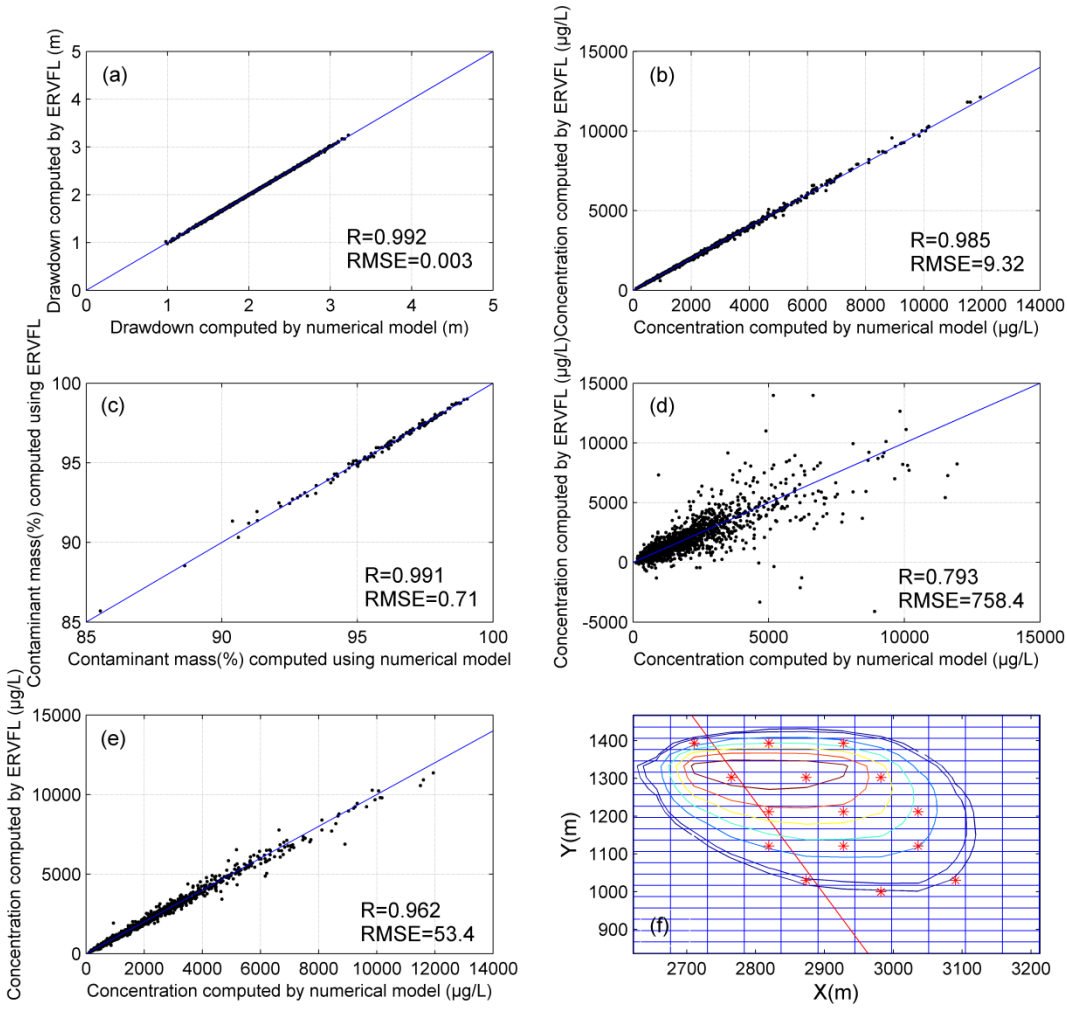

Fig.11. Scatter plot depicting the relationship between (a) Drawdown computed using numerical model vs ERVFL (b) Concentration computed using numerical model vs ERVFL (c) Percentage of extracted contaminant mass computed using numerical model vs ERVFL (d) Concentration computed using numerical model vs RVFL (Data corrupted with outliers) (e) Concentration computed using numerical model vs ERVFL (Data corrupted with outliers) (f) Predefined pumping well location for traditional approach

In the second step of the two-step approach, the aim is to obtain the Pareto-optimal solution of two conflicting objectives: minimization of total pumping rate and the minimization of the total amount of contaminant mass in the aquifer. The mathematical formulation of the management model is given in Eq. (31). Here we consider two metaheuristics: NSGA-II and MO-EMPA. The Pareto-optimal solutions obtained using both the metaheuristics are shown in Fig. 12. The Pareto fronts are generated for four different sets of generations: $T = 100$, $T = 250$, $T = 500$ and $T = 1000$. From the visual appearance of Pareto fronts, it is observed that the MO-EMPA is quicker (in terms of the number of generations) than NSGA-II to obtain the whole Pareto front. The Pareto fronts obtained using MO-EMPA after 100 and 250 iterations are spread over a broader range of the objective functions space than NSGA-II. However, the spread of Pareto optimal solutions after 1000 generations are similar for both the metaheuristics. It is noteworthy to mention here that we have relaxed the constraints mentioned in Eq. (31) to generate the whole Pareto front.

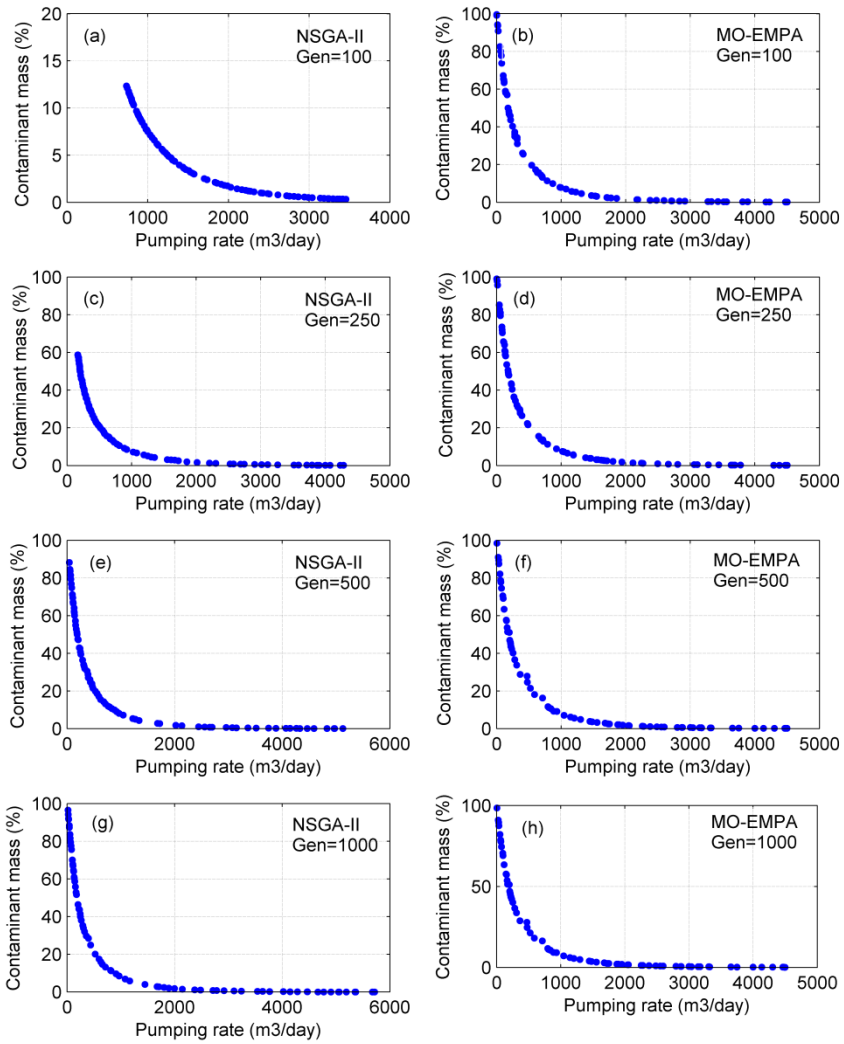

Fig.12. Pareto optimal solutions depicting the relationship between pumping rate and the percentage of contaminant mass in the aquifer using NSGA-II vs MO-EMPA with varying generation number

Here, we check the effectiveness of the two-step approach with respect to the traditional approach for multi-objective groundwater remediation. In the traditional approach, the researchers manually fix/guess the pumping well locations by studying the concentration plume (Erickson et al. 2002; Yang et al. 2013; Jiang and Na 2020). Fig.11d also shows the predefined pumping well location used in the traditional approach. The manually defined pumping well locations are never the best pumping locations for effective groundwater remediation. Fig.13 compares the Pareto-optimal solutions obtained using the two-step approach and the traditional

878 approach. Here also, we consider four scenarios by varying the range of decision variables
879 (pumping rates): (a) $-55 < Q < -45$ (b) $-60 < Q < -30$
880 (c) $-80 < Q < -20$ (d) $-300 < Q < 0$.
881 The figure shows that the percentage of contaminant mass in the aquifer using the two-step
882 approach is far less than the traditional approach for the same pumping rate. The difference in
883 the percentage of contaminant mass for the same pumping rates is almost 15% for cases (a) and
884 (b). Further, comparing the four scenarios, we also observed that the difference between the two
885 Pareto optimal fronts is less prominent when the ranges of pumping rates (decision variables) are
886 more. By increasing the range of pumping rates, we are giving more weightage to it in the
887 optimization model than the pumping locations. This is the reason for such behavior of the
888 Pareto front.
889
890
891
892
893
894
895
896
897
898
899
900

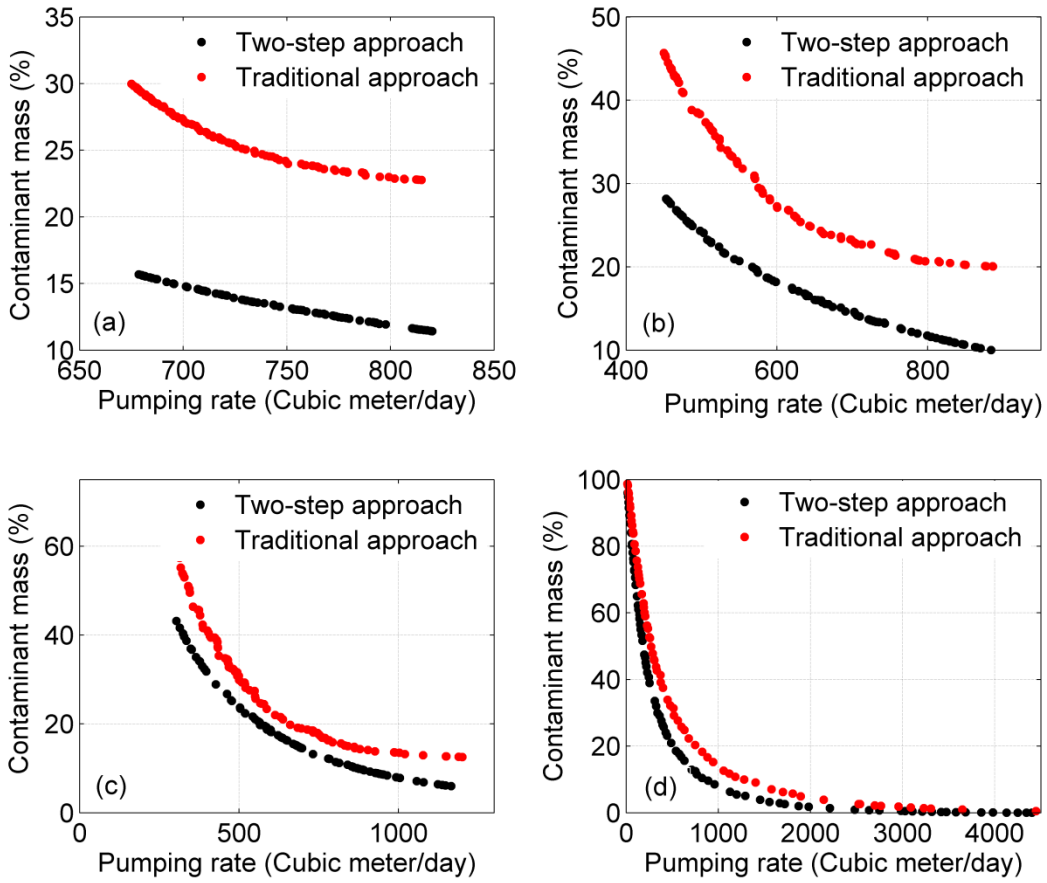

Fig.13. Pareto optimal fronts obtained using two-step approach vs traditional approach with varying decision variables range (a) $-55 < Q < -45$ (b) $-60 < Q < -30$ (c) $-80 < Q < -20$ (d) $-300 < Q < 0$

7. Conclusion

In this study, simulation-optimization (S-O) models are proposed for multi-objective groundwater remediation by integrating the evolutionary marine predator algorithm (EMPA) with enhanced random vector functional link (ERVFL). The performances of the single objective version of EMPA in obtaining the optimal values of composite test functions are found to be better than other metaheuristics. Further, the performance of the MO(multi-objective)-EMPA is tested on a series of benchmark functions, and the results were found to be relatively better than NSGA-II in terms of two performance metrics viz. spacing and inverted generational distance.

This study also proposes the ERVFL network as a proxy simulator to enhance the computational performance of simulation models. The ERVFL model showed excellent generalization ability in approximating the simulation model representing the groundwater flow and contaminant transport processes. Further, the ERVFL network performed significantly better than the original RVFL when the training dataset is corrupted with outliers.

The most novel contribution of the present study is the use of the two-step approach for groundwater remediation. In the first step, the optimal pumping locations are obtained for groundwater remediation using combinatorial optimization by minimizing the amount of contaminant mass in the aquifer while keeping the constant pumping rates. The optimal solutions obtained using EMPA are compared with other metaheuristics using violin plots and Dunn and Sidák's test. The results showed the superior performance of EMPA over other metaheuristics in obtaining optimal pumping locations.

The optimal pumping locations obtained in the first step are directly used in the second step to design multi-objective groundwater remediation strategies. The Pareto-optimal solutions (tradeof between optimal total pumping rates and percentage of contaminant mass in the aquifer) obtained using MO-EMPA is compared with NSGA-II. It is observed that the MO-EMPA generates the whole Pareto-optimal front with less number of generations compared to NSGA-II. The two-step approach is further compared with the traditional approach while varying the decision variables space. The comparison results show that the two-step approach is significantly better than the traditional approach for multi-objective groundwater remediation. The present study does not consider hydrological uncertainties for multi-objective groundwater remediation. In the future, hydrological uncertainties can be incorporated into the management model using chance constraints.

Acknowledgment

C. Lu acknowledges the financial support from the National Natural Science Foundation of China (51679067 and 51879088).

References

Aquaveo (2018) Groundwater Modeling System (version 10.4) Tutorial, MODFLOW - Conceptual Model Approach 1. http://gmstutorials-10.4.aquaveo.com/MODFLOW-ConceptualModelApproach1.pdf. 1–23

Batu V (2005). (2006) Applied flow and solute transport modeling in aquifers: fundamental principles and analytical and numerical methods. CRC Press, Boca Raton

Coello CAC, Pulido GT, Lechuga MS (2004) Handling multiple objectives with particle swarm optimization. IEEE Trans Evol Comput 8:256–279. https://doi.org/10.1109/TEVC.2004.826067

Dai W, Liu Q, Chai T (2015) Particle size estimate of grinding processes using random vector functional link networks with improved robustness. Neurocomputing 169:361–372. https://doi.org/10.1016/j.neucom.2014.08.098

Deb K (2012) Multi objective optimization using evolutionary algorithms. John Wiley & Sons, Ltd, New York

Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multi-objective genetic algorithm: NSGA-II. IEEE Trans Evol Comput 6:182–197. https://doi.org/10.1109/4235.996017

Dhargupta S, Ghosh M, Mirjalili S, Sarkar R (2020) Selective Opposition based Grey Wolf Optimization. Expert Syst Appl 151:113389. https://doi.org/10.1016/j.eswa.2020.113389

Elaziz MA, Shehabeldeen TA, Elsheikh AH, et al (2020) Utilization of random vector functional link integrated with marine predators algorithm for tensile behavior prediction of dissimilar friction stir welded aluminum alloy joints. J Mater Res Technol 9:11370–11381. https://doi.org/10.1016/j.jmrt.2020.08.022

Erickson M, Mayer A, Horn J (2002) Multi-objective optimal design of groundwater remediation systems: Application of the niched Pareto genetic algorithm (NPGA). Adv Water Resour

25:51–65. https://doi.org/10.1016/S0309-1708(01)00020-3

Faramarzi A, Heidarinejad M, Mirjalili S, Gandomi AH (2020) Marine Predators Algorithm: A nature-inspired metaheuristic. Expert Syst Appl 152:113377. https://doi.org/10.1016/j.eswa.2020.113377

Freezy RA, Cherry JA (1979) Groundwater. Prentice Hall, Englewood Cliffs, New Jersey

Got A, Moussaoui A, Zouache D (2020) A guided population archive whale optimization algorithm for solving multi-objective optimization problems. Expert Syst Appl 141:112972. https://doi.org/10.1016/j.eswa.2019.112972

Guan J, Aral MM (1999) Optimal remediation with well locations and pumping rates selected as continuous decision variables. J Hydrol 221:20–42. https://doi.org/10.1016/S0022-1694(99)00079-7

Harbaugh, Arlen W (2005) MODFLOW-2005 , The U . S . Geological Survey Modular Ground-Water Model — the Ground-Water Flow Process. Reston, Virginia

Huang C, Mayer AS (1997) Pump-and-treat optimization using well locations and pumping rates as decision variables. Water Resour Res 33:1001–1012. https://doi.org/10.1029/97WR00366

Jiang X, Na J (2020) Online surrogate multi-objective optimization algorithm for contaminated groundwater remediation designs. Appl Math Model 78:519–538. https://doi.org/10.1016/j.apm.2019.09.053

Knowles JD, Corne DW (2000) Approximating the nondominated front using the Pareto Archived Evolution Strategy. Evol Comput 8:149–172. https://doi.org/10.1162/106365600568167

Kumar D, Prasad RK, Mathur S (2013) Optimal design of an in-situ bioremediation system using support vector machine and particle swarm optimization. J Contam Hydrol 151:105–116. https://doi.org/10.1016/j.jconhyd.2013.05.003

Liang JJ, Suganthan PN, Deb K (2005) Novel composition test functions for numerical global optimization. Proc - 2005 IEEE Swarm Intell Symp SIS 71–78. https://doi.org/10.1109/SIS.2005.1501604

Liu J, Yang Z, Li D (2020) A multiple search strategies based grey wolf optimizer for solving multi-objective optimization problems. Expert Syst Appl 145:113134. https://doi.org/10.1016/j.eswa.2019.113134

1007 Luo Q, Wu J, Yang Y, et al (2014) Optimal design of groundwater remediation system using a
1008 probabilistic multi-objective fast harmony search algorithm under uncertainty. J Hydrol
1009 519:3305–3315. https://doi.org/10.1016/j.jhydrol.2014.10.023
1010 Majumder P, Eldho TI (2020) Artificial Neural Network and Grey Wolf Optimizer Based
1011 Surrogate Simulation-Optimization Model for Groundwater Remediation. Water Resour
1012 Manag 34:763–783. https://doi.org/10.1007/s11269-019-02472-9
1013 Majumder P, Eldho TI (2016) A New Groundwater Management Model by Coupling Analytic
1014 Element Method and Reverse Particle Tracking with Cat Swarm Optimization. Water
1015 Resour Manag 30:1953–1972. https://doi.org/10.1007/s11269-016-1262-5
1016 Majumder P, Eldho TI (2019) Reactive contaminant transport simulation using the analytic
1017 element method, random walk particle tracking and kernel density estimator. J Contam
1018 Hydrol 222:76–88. https://doi.org/10.1016/j.jconhyd.2019.01.006
1019 McKinney DC, Lin M -D (1994) Genetic algorithm solution of groundwater management
1020 models. Water Resour Res 30:1897–1906. https://doi.org/10.1029/94WR00554
1021 Mirjalili S, Mirjalili SM, Lewis A (2014) Grey Wolf Optimizer. Adv Eng Softw 69:46–61.
1022 https://doi.org/10.1016/j.advengsoft.2013.12.007
1023 Mirjalili S, Saremi S, Mirjalili SM, Coelho LDS (2016) Multi-objective grey wolf optimizer: A
1024 novel algorithm for multi-criterion optimization. Expert Syst Appl 47:106–119.
1025 https://doi.org/10.1016/j.eswa.2015.10.039
1026 Pao YH, Park GH, Sobajic DJ (1994) Learning and generalization characteristics of the random
1027 vector functional-link net. Neurocomputing 6:163–180. https://doi.org/10.1016/0925-
1028 2312(94)90053-1
1029 Ridha HM (2020) Parameters extraction of single and double diodes photovoltaic models using
1030 Marine Predators Algorithm and Lambert W function. Sol Energy 209:674–693.
1031 https://doi.org/10.1016/j.solener.2020.09.047
1032 Rizzo DM, Dougherty DE (1996) Design optimization for multiple management period
1033 groundwater remediation. Water Resour Res 32:2549–2561.
1034 https://doi.org/10.1029/96WR01334
1035 Scardapane S, Wang D, Panella M, Uncini A (2015) Distributed learning for Random Vector
1036 Functional-Link networks. Inf Sci (Ny) 301:271–284.
1037 https://doi.org/10.1016/j.ins.2015.01.007

1038 Seyedpour SM (2019) Optimal remediation design and simulation of coupled groundwater flow
1039 and contaminant transport using genetic algorithm and radial point collocation method (
1040 RPCM). Sci Total Environ 669:389–399.
1041 https://doi.org/10.1016/J.SCITOTENV.2019.01.409
1042 Sidiropoulos E, Tolikas P (2008) Genetic algorithms and cellular automata in aquifer
1043 management. Appl Math Model 32:617–640. https://doi.org/10.1016/j.apm.2007.01.005
1044 Soncco-Álvarez JL, Muñoz DM, Ayala-Rincón M (2019) Opposition-Based Memetic Algorithm
1045 and Hybrid Approach for Sorting Permutations by Reversals. Evol Comput 27:229–265.
1046 https://doi.org/10.1162/evco_a_00220
1047 Storn R, Price K (1997) Differential Evolution - A Simple and Efficient Heuristic for Global
1048 Optimization over Continuous Spaces. J Glob Optim 11:341–359.
1049 https://doi.org/10.1023/A:1008202821328
1050 Sun X, Wang G, Xu L, et al (2020) Optimal performance of a combined heat-power system with
1051 a proton exchange membrane fuel cell using a developed marine predators algorithm. J
1052 Clean Prod. https://doi.org/10.1016/j.jclepro.2020.124776
1053 Tamer Ayvaz M (2009) Application of Harmony Search algorithm to the solution of
1054 groundwater management models. Adv Water Resour 32:916–924.
1055 https://doi.org/10.1016/j.advwatres.2009.03.003
1056 Vuković N, Petrović M, Miljković Z (2018) A comprehensive experimental evaluation of
1057 orthogonal polynomial expanded random vector functional link neural networks for
1058 regression. Appl Soft Comput 70:1083–1096. https://doi.org/10.1016/j.asoc.2017.10.010
1059 Wang JS, Li SX (2019) An Improved Grey Wolf Optimizer Based on Differential Evolution and
1060 Elimination Mechanism. Sci Rep 9:1–21. https://doi.org/10.1038/s41598-019-43546-3
1061 Wang W, Ahlfeld DP (1994) Optimal groundwater remediation with well location as a decision
1062 variable: Model development. Water Resour Res 30:1605–1618.
1063 https://doi.org/10.1029/93WR03552
1064 Wolpert DH, Macready WG (1997) No free lunch theorems for optimization. IEEE Trans Evol
1065 Comput 1:67–82. https://doi.org/10.1109/4235.585893
1066 Yadav B, Ch S, Mathur S, Adamowski J (2016) Estimation of in-situ bioremediation system cost
1067 using a hybrid Extreme Learning Machine (ELM)-particle swarm optimization approach. J
1068 Hydrol 543:373–385. https://doi.org/10.1016/j.jhydrol.2016.10.013

Yang Y, Wu J, Sun X, et al (2013) A niched Pareto tabu search for multi-objective optimal design of groundwater remediation systems. J Hydrol 490:56–73. https://doi.org/10.1016/j.jhydrol.2013.03.022

Yang Y, Wu J, Wang J, Zhou Z (2017) An Elitist Multiobjective Tabu Search for Optimal Design of Groundwater Remediation Systems. Groundwater 55:811–826. https://doi.org/10.1111/gwat.12525

Zhang L, Suganthan PN (2016) A comprehensive evaluation of random vector functional link networks. Inf Sci (Ny) 367–368:1094–1105. https://doi.org/10.1016/j.ins.2015.09.025

Zhang Q, Zhou A, Zhao S, et al (2009) Multi-objective optimization test instances for the CEC 2009 special session and competition. 2009 IEEE Congr Evol Comput (CEC 2009) 1–30

Zhao Y, Qu R, Xing Z, Lu W (2020) Identifying groundwater contaminant sources based on a KELM surrogate model together with four heuristic optimization algorithms. Adv Water Resour 138:103540. https://doi.org/10.1016/j.advwatres.2020.103540

Zheng C, Wang PP (1999) MT3DMS: A Modular Three-Dimensional Multispecies Transport Model for simulation of advection, dispersion and chemical reactions of contaminants in groundwater systems. 1–239