DOI: 10.1515/jwld-2015-0022

JOURNAL OF WATER AND LAND DEVELOPMENT
2015, No. 27 (X–XII): 29–40
PL ISSN 1429–7426
Available (PDF): http://www.itp.edu.pl/wydawnictwo/journal; http://www.degruyter.com/view/j/jwld

Received 06.08.2015
Reviewed 10.09.2015
Accepted 13.10.2015

A – study design
B – data collection
C – statistical analysis
D – data interpretation
E – manuscript preparation
F – literature search

Multi-objective optimization of in-situ bioremediation of groundwater using a hybrid metaheuristic technique based on differential evolution, genetic algorithms and simulated annealing

Deepak KUMAR[1) ABCDEF], Sudheer CH[1) ABCDEF], Shashi MATHUR[1) ABCDEF], Jan ADAMOWSKI[2) ABCDEF]

1) Indian Institute of Technology in Delhi, India; e-mail: deepak.civil.iitdelhi@gmail.com; sudheer108@gmail.com; smathur@civil.iitd.ac.in
2) McGill University, Faculty of Agricultural and Environmental Sciences, Department of Bioresource Engineering, Quebec, Canada, H9X 3V9; e-mail: jan.adamowski@mcgill.ca

For citation: Kumar D., Ch S., Mathur S., Adamowski J. 2015. Multi-objective optimization of in-situ bioremediation of groundwater using a hybrid metaheuristic technique based on differential evolution, genetic algorithms and simulated annealing. Journal of Water and Land Development. No. 27 p. 29–40.

Abstract

Groundwater contamination due to leakage of gasoline is one of the several causes which affect the groundwater environment by polluting it. In the past few years, In-situ bioremediation has attracted researchers because of its ability to remediate the contaminant at its site with low cost of remediation. This paper proposed the use of a new hybrid algorithm to optimize a multi-objective function which includes the cost of remediation as the first objective and residual contaminant at the end of the remediation period as the second objective. The hybrid algorithm was formed by combining the methods of Differential Evolution, Genetic Algorithms and Simulated Annealing. Support Vector Machines (SVM) was used as a virtual simulator for biodegradation of contaminants in the groundwater flow. The results obtained from the hybrid algorithm were compared with Differential Evolution (DE), Non Dominated Sorting Genetic Algorithm (NSGA II) and Simulated Annealing (SA). It was found that the proposed hybrid algorithm was capable of providing the best solution. Fuzzy logic was used to find the best compromising solution and finally a pumping rate strategy for groundwater remediation was presented for the best compromising solution. The results show that the cost incurred for the best compromising solution is intermediate between the highest and lowest cost incurred for other non-dominated solutions.

Key words: *differential evolution, fuzzy logic, genetic algorithm, groundwater, hybrid algorithm, in-situ bioremediation, simulated annealing, support vector machine (SVM)*

INTRODUCTION

Generally one of the cleanest sources of water, groundwater can, however, see its quality compromised through leakage followed by leaching of organic and inorganic contaminants. These may enter the groundwater aquifer through a variety of sources, including toxic waste disposal sites, accidental chemical spills and improperly designed or maintained chemical transportation and storage facilities. Often released through the leakage of underground petroleum product storage tanks, the organic contaminant mix of benzene, toluene, ethyl benzene, and xylenes (BTEX) is of significant concern because of its harmful effects, even when present at very low concentrations. In the past, various techniques such as pump-and-treat, air sparging, and in situ bioremediation have been applied to groundwater remediation, but in recent times the latter technique has seen widespread use. It involves the use of injection and extraction wells installed at various locations on the site to check and remediate the concentration plume. Injection wells, located up gradient, are used to add oxygen and nutrients into the groundwater aquifer, whereas the extraction wells are located down gradient from the site. In a typical in situ bioremediation system, the location of injection and extraction wells, along with their pumping rates are decision variables determined after optimizing the system.

A review of the literature indicates that regression and metaheuristic based methods have served in assessing optimal groundwater remediation design under either single- and multi-objective optimization techniques. In an early approach to in situ bioremediation optimization for groundwater, MINSKER and SHOEMAKER [1998] applied the Successive Approximation Linear Quadratic Regulator (SALQR). Later, HU *et al.* [2004] developed a dynamic predictive control system for in situ bioremediation of groundwater, featuring the minimization of overall remediation cost as a single objective function. In their model, simulation and decision-making tools served as the two control systems involved in the optimization process. A genetic algorithm was used as a search method for optimization of the cost function as a single objective optimization. Similarly, SHIEH *et al.* [2005] studied the combined use of genetic algorithms and simulated annealing to search for optimal control of in situ bioremediation. In their model, BIOPLUME II was used for simulating aquifer hydraulics and the bioremediation process. Later, HU *et al.* [2007] developed a control system for in situ bioremediation of groundwater using a genetic algorithm for the multi-objective optimization method. Their objectives included minimizing the overall cost and maximizing system efficiency in the face of a bioremediation site's uncertain data. However, the present study proposes a novel procedure that employs a metaheuristic hybrid algorithm to solve a multi-objective optimization for in situ bioremediation of groundwater.

Metaheuristic algorithms have been extensively applied in optimization problems [MONTESINOS *et al.* 1999; PANIGRAHI *et al.* 2008; RAVIKUMAR PANDI, PANIGRAHI 2006]. Some of the metaheuristic algorithms most commonly employed in deriving a global optimal value in an optimization problem include: genetic algorithm (GA), simulated annealing (SA), differential evolution (DE), particle swarm optimization (PSO) and Tabu search (TS). However, metaheuristic optimizers can miss the global optimum, converging rather to a local, sub-optimal point. The occurrence of this serious problem [OLDENHUIS 2010] can be minimized through the use of artificial intelligence and operational research in conjunction with metaheuristic algorithms. The resultant algorithm is termed a hybrid algorithm. In such a hybrid algorithm, the weaknesses associated with each algorithm are negated by the strengths of others, while common strengths are cumulated. Hybrid algorithms can be developed by combining metaheuristic algorithms with one or more existing algorithms, such as dynamic programming [TSE *et al.* 2007], constraint programming [MAYER 2008; PRESTWICH 2002], tree search method [BLUM 2005; ROTHBERG 2007] or even with another metaheuristic algorithm [ELRAGAL *et al.* 2011; OLDENHUIS 2010; ZHANG *et al.* 2009].

Metaheuristic hybrid algorithms are thus hybrid algorithms with two or more coupled algorithms to increase the robustness of the algorithm. In such hybrid algorithms, a population-based algorithm tries to find a domain in which global optima may exist, followed by a local search, which can quickly identify the best solution. Both DE and PSO have been combined to yield hybrid algorithms [ALI *et al.* 2009; ELRAGAL *et al.* 2011; VINKO *et al.* 2007]. In the present study, GA, SA and DE were coupled to form a hybrid metaheuristic algorithm. This in turn was employed to solve an in situ bioremediation problem requiring multi-objective optimization. Our objective was to determine the best pareto on the basis of the algorithm's population size and evaluate different injection and extraction well-pumping patterns to determine the best compromise solution.

MATERIAL AND METHODS

DESCRIPTION OF STUDY AREA

A hypothetical contaminant site (Fig. 1) was considered in the present study. The aquifer was assumed to be homogeneous, with constant head boundaries on the west and the east sides of the domain having respective head values of 30.5 m and 27.7 m. The north and south side boundaries were assumed to be impervious, resulting in an initial west to east flow under an initial hydraulic gradient of 0.004. The groundwater

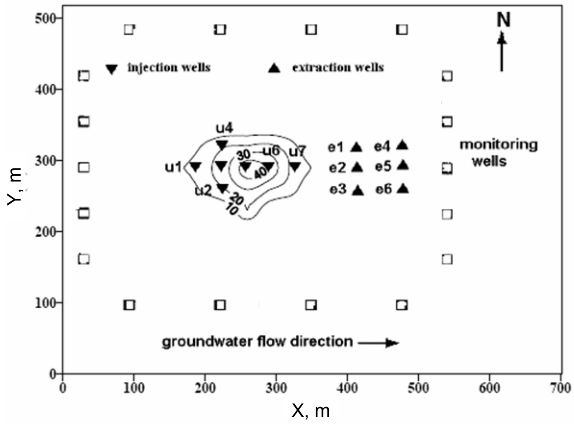

Fig. 1. Location of injection and extraction wells and the initial BTEX concentration (ppm) at the contaminated site; source: own elaboration

flow simulation was assumed to operate in a steady state and the representative organic pollutant was taken to be BTEX. Figure (1) depicts the initial contaminant concentration and a set of preselected injection and extraction well locations for an in situ bioremediation system.

The maximum concentration of organic pollutant was assumed to be 40 ppm and the site's hydraulic conductivity was taken to be $6{\cdot}10^{-6}$ m·s^{-1}. The transverse and longitudinal dispersivity of the 15 m thick aquifer was 2 m and 10 m, respectively. An effective porosity of 0.3 and a retardation and anisotropic factor of 1 were assumed. A three-year remediation period with an allowable maximum organic contaminant concentration of 5 ppm at its completion was selected for modeling.

SIMULATION OPTIMIZATION MODEL FOR IN-SITU BIOREMEDIATION

Groundwater flow and solute transport equations were simulated with BIOPLUME III, and its outputs for maximum and minimum hydraulic head and maximum contaminant concentration in the study area were calibrated and validated using Support Vector Regression (SVR). Finally, the SVR was coupled with an optimization algorithm to generate the paretos. After obtaining the paretos from the various algorithms (DE, DGS (Hybrid of DE, SA and GA), SA and GA), the performance of the algorithms was evaluated with respect to which eventually provided the best compromise solution.

BIOPLUME III

The BIOPLUME III program is a two-dimensional, finite difference model that simulates natural attenuation of organic contaminants in ground water due to advection, dispersion, sorption, and biodegradation. The model simulates the biodegradation of organic contaminants using a number of aerobic and anaerobic electron acceptors (oxygen, nitrate, iron (III), sulphate, and carbon dioxide). The transport equations for the contaminant and oxygen are expressed as

$$\frac{\delta(C.b)}{\delta t}=\frac{1}{R_c}\left(\frac{\delta}{\delta x_i}\left(b.D_{ij}\frac{\delta C}{\delta x_j}\right)-\frac{\delta}{\delta x_i}C.b.V_i\right)-\frac{C'.W}{\eta} \quad (1)$$

$$\frac{\delta(O.b)}{\delta t}=\frac{1}{R_c}\left(\frac{\delta}{\delta x_i}\left(b.D_{ij}\frac{\delta O}{\delta x_j}\right)-\frac{\delta}{\delta x_i}O.b.V_i\right)-\frac{O'.W}{\eta} \quad (2)$$

where:
- b = the saturated thickness, m;
- t = time, s;
- x_i, x_j = Cartesian coordinates, m;
- C = the concentration of hydrocarbons, mg·L^{-1};
- C' = the concentration of hydrocarbons in the source or sinks, mg·L^{-1};
- D_{ij} = the coefficient of hydrodynamic dispersion;
- O = the concentration of oxygen, mg·L^{-1};
- O' = the concentration of oxygen in the source or sink, mg·L^{-1};
- R_c = the retardation factor for hydrocarbons;
- V_i = the seepage velocity in the direction of x_i, m·s^{-1};
- W = the volume flux per unit area, L·s^{-1}·m^{-2};
- η = the effective porosity.

The biodegradation process is simulated by incorporating the instantaneous reaction model proposed by BORDEN and BEDIENT [1986]. It was assumed that the utilization of oxygen and contaminants by microorganisms in the subsurface zone could be simulated as an instantaneous reaction between the organic contaminant and the electron acceptors. The biodegradation of contaminants using aerobic electron acceptors was simulated using the principle of superposition [BORDEN, BEDIENT 1986], thus

$$\Delta C_{RC} = O/F\text{ ; } O = 0 \text{ where } C > O/F \quad (3)$$

$$\Delta C_{RO} = C.F\text{ ; } C = 0 \text{ where } O > C.F \quad (4)$$

where:
- ΔC_{RC} = calculated changes in concentrations of contaminant due to bioremediation, ppm;
- ΔC_{RO} = calculated changes in concentrations of oxygen due to bioremediation, ppm;
- F = the ratio of oxygen to hydrocarbon consumed.

Using BIOPLUME III, equations (1) to (4) were employed in simulating the contaminant concentration and maximum and minimum head at the end of each pumping period. A large number of pumping patterns was randomly generated and served as input to the BIOPLUME III model. The pumping patterns of the decision vector and the output of BIOPLUME III model were then respectively employed as input and output to calibrate and validate the Support Vector Machine.

SUPPORT VECTOR MACHINE (SVM) AS A VIRTUAL SIMULATOR

In a formal simulation-optimization strategy, an optimization algorithm is incorporated into the simulation model as a subroutine or a submodel [WANG, ZHENG 1997]. However, this process normally requires hundreds or even thousands of calls to the simulator, rendering the process extremely time-consuming and computationally intensive. To reduce the computational burden, a SVM was used in the present study as a virtual simulator. In machine learning, the SVM algorithm analyzes data and recognizes patterns and ultimately serves in classification and regression analysis. Numerous studies have shown the usefulness of the SVR method in hydrological forecasting (e.g. BELAYNEH *et al.* [2014]). In the present study, the SVM linked with the optimization algorithm was independent of the model runs.

POTENTIAL WELL LOCATIONS

For in situ remediation of the contaminant at the study area, a number of injection and extraction wells were installed. In the present study, the locations of the seven injection wells (u1, u2, u3, u4, u5, u6 and u7) and six extraction wells (e1, e2, e3, e4, e5 and e6) were preselected (Fig 1). Out of these seven injection wells and six extraction wells, three injection wells (u1, u2 and u4) and one extraction well (e2) were optimally selected for their potential use in developing the optimal design for in situ bioremediation, an important facet in solving groundwater remediation problems [GUAN, ARAL 1999; HUANG, MAYER 1997]. This optimization process was undertaken by using an ANN (Artificial Neural Network) embedded in a Monte Carlo approach [PRASAD, MATHUR 2008]. The same sets of well locations were then used in the present study to resort to further metaheuristic optimization.

MULTI-OBJECTIVE OPTIMIZATION

Since the cost of remediation depends on the allowable residual level of the contaminant, remediation costs can run very high if a low permissible contaminant level is adopted. Thus, the cost and contaminant concentration in the study area represent two conflicting objectives, which require a multi-objective optimization approach. The objective functions [SHIEH, PARELTA 2005] are:

Objective 1: Minimize cost, Z

$$Minimize\ Z = \sum_{k=0}^{K}\left(\frac{1}{1+i_r}\sum_{n=1}^{N} C^q(n).q(n,k)\right) + \\ + \sum_{n=1}^{N} C^{IP}(n).IP(n) + \\ + \max\left\{D\left(\sum_{n=1}^{N_i} q(n,k)\right)\right\}_{k=1}^{K} + \\ + \max\left\{E\left(\sum_{n=1}^{N_e} q(n,k)\right)\right\}_{k=1}^{K} \quad (5)$$

where:

i_r = the discount rate;
$q(n)$ = the injection/extraction rate of the n^{th} well, $L^3 \cdot T^{-1}$;
C = the concentration of contaminant remaining at the end of the 3-year management period, ppm;
$C^{IP}(n)$ = the installation cost of the well, \$ per well;
$C^q(n)$ = the cost parameter of injection/extraction, \$ per lps;
$D\left(\sum_{n=1}^{N_i} q(n,k)\right)$ = the oxygen and nutrient injection facility capital cost, \$;
$E\left(\sum_{n=1}^{N_e} q(n,k)\right)$ = the treatment facility capital cost, \$;
$IP(n)$ = the zero-one integer for well existence;
K = total duration of the remediation period, *i.e.*, 3 years, T;
N = the total number of injection and extraction wells;
N_e = the total number of extraction wells
N_i = the total number of injection well,
Z = the total cost of in situ bioremediation system, \$.

Objective2: Minimize concentration: C (6)

The cost parameter used in equation 5 is mentioned in Table 1.

Table 1. Values of cost parameters used for equation (5)

Coefficient	Value
i_r	0.05
C^{IP}	\$ 12,000 per well
C^q (for injection well)	\$ 4,755 (per Ls^{-1}-year)
C^q (for extraction well)	\$ 15,850 (per Ls^{-1}-year)
D (injection facility capital cost)	$D_{1.26Ls^{-1}}$ = \$ 20,000
E (treatment facility capital cost)	$E_{1.26Ls^{-1}}$ = \$ 30,000

Source: own study.

Equations (5) and (6) are subject to the constraints enumerated in Equations (7) through (12), and handled through a penalty function:

$$0 \le C_s^K(j_0) \le C_{\max}(j_0) \forall \quad j_0 \in \Phi \quad (7)$$

$$0 \le C_s^K(j_0) \le C_{\max}(j_0) \forall \quad j_0 \in \Psi \quad (8)$$

$$h^k(j_0) \le h^{\max}(j_o) \forall \quad j_0 \in \Phi \quad (9)$$

$$h^{\min}(j_0) \le h^k(j_o) \forall \quad j_0 \in \Phi \quad (10)$$

$$0 \le q_{ew}^k \le q_{ew}^k(\max) \quad (11)$$

$$0 \le q_{iw}^k \le q_{iw}^k(\max) \quad (12)$$

where:

$h^k(j_0)$ = the hydraulic head allowed at node j_0 in period k, m;
$h^{\max}(j_o)$ = the maximum hydraulic heads allowed at node j_0 in period k, m;
$h^{\min}(j_0)$ = the minimum hydraulic heads allowed at node j_0 in period k, m;
q_{ew}^k = the extraction pumping rates during period k, $L \cdot s^{-1}$;
q_{iw}^k = the injection pumping rates during period k, $L \cdot s^{-1}$;
$q_{ew}^k(\max)$ = the well capacity for extraction wells, $L \cdot s^{-1}$;
$q_{iw}^k(\max)$ = the well capacity for injection wells, $L \cdot s^{-1}$; $L^3 \cdot T^{-1}$;
$C_s^K(j_0)$ = the contaminant concentration/load at node j_0 at the end of the remediation period K, ppm;
$C_{\max}(j_0)$ = the water quality goal at location j_0, ppm;
Φ = represents a set of all study area nodes where water quality and hydraulic head compliance must be attained;

Ψ = a set of monitoring wells of the problem domain where water quality compliance endure over the management period.

DESCRIPTION OF HYBRID ALGORITHM

In the present study, a hybrid algorithm, which combines three metaheuristic algorithms (*i.e.*, DE, SA and NSGA II) implemented in MATLAB, was employed to solve the objective function (Eqs. 5, 6). The control parameters in DGS are listed in Table 2. In this hybrid algorithm, the full population was first split into N portions of random size and for each of these subpopulations a differential evolution, NSGA II and SA was performed randomly. The algorithm so developed is termed a DGS, as it is comprised of a hybrid algorithm formed by combining Differential evolution, Genetic algorithm and Simulated annealing approaches. The complete flowchart of the algorithm is depicted in Fig. 2.

Table 2. Control parameters for DGS algorithm

Algorithm	Parameter	Value
NSGA II	chromosome size	52
	crossover probability	0.5
	mutation probability	0.01
SA	cooling schedule constant	0.87
	reheating	5
DE	probability of crossover	0.95
	scaling factor	1

Source: own study.

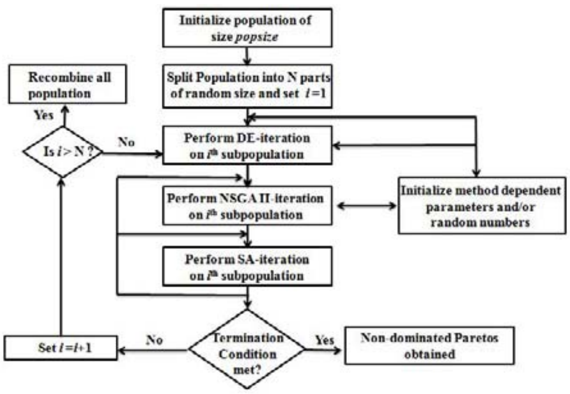

Fig. 2. Flowchart of the DGS algorithm; source: own elaboration

COUPLING SVM WITH OPTIMIZATION TOOL

During optimization of an in-situ bioremediation system, the pumping strategy is one of the most important decision variables on which the total cost of the remediation depends. The optimization algorithm searches the global optimal pumping strategy from a random set of pumping rates. For every random set of pumping strategy, the algorithm also checks if any constraint has been violated. Thus, for the specific problem considered in this study, one major constraint is that the maximum concentration in the aquifer should not be more than 5 ppm at the end of 3 years. Hence, to check this constraint, the simulation model needs to run for a particular pumping strategy so as to know whether there has been any violation of the constraint or not. The optimization algorithm would then call the simulation model for controlling the violation of constraint. Proxy simulators in place of real simulators can be used to reduce the computational time of the optimal design. In the present study, the trained SVM was coupled with the hybrid algorithm (Fig. 3). A coupling was required as the trained SVM was the only tool able to generate, during the optimization process, $h^{max}(j_o)$, $h^{min}(j_0)$ and maximum contaminant concentration at the contaminant site for randomly generated pumping rates. The output through the SVM was further employed to satisfy the constraints of Eqs. 7–12, by using a penalty function in the DGS algorithm. In the case where constraints were satisfied, the algorithm would find the paretos for the multi-objective function.

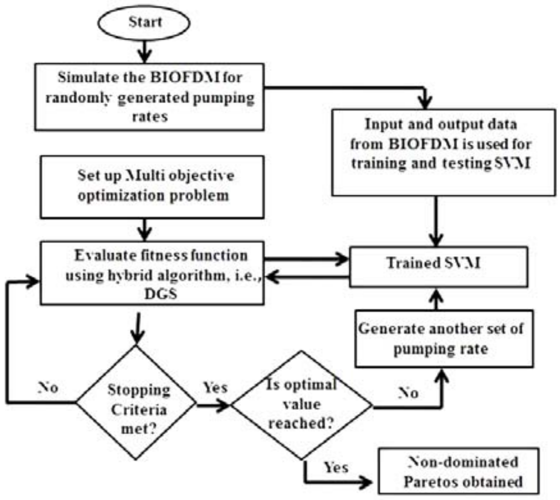

Fig. 3. Trained SVM coupled with the hybrid multi-objective algorithm source: own elaboration

DETERMINING THE BEST COMPROMISE SOLUTION

The Pareto front obtained from DGS was further employed in determining the best compromise solution. The mechanism, based on a fuzzy linear membership function, was applied in the present study to determine the best compromise solution. The membership value of each individual in the Pareto optimal set F_i was computed using the membership function [PANIGRAHI *et al.* 2010] defined as:

$$\mu_i \begin{cases} 1 & if\ F_i \le F_i^{\min} \\ \frac{F_i^{\max} - F_i}{F_i^{\max} - F_i^{\min}} & if\ F_i^{\min} < F_i < F_i^{\max} \\ 0 & if\ F_i \ge F_i^{\max} \end{cases} \tag{13}$$

where: μ_i = the membership value of the i^{th} function (F_i).

For each non-dominated solution k, the normalized membership value ($\mu[k]$) is defined as: [PANIGRAHI *et al*. 2010]

$$\mu[k] = \frac{\sum_{i=1}^{M} \mu_i[k]}{\sum_{j=1}^{N_{pareto}} \sum_{i=1}^{M} \mu_i[j]} \tag{14}$$

where:
M = the number of objectives,
N_{pareto} = the number of paretos in the optimal front.

The best compromise solution is one for which $\mu[k]$ is maximum.

RESULTS AND DISCUSSION

The study addressed five steps in the process of obtaining best solutions: (i) training a SVM for use as a virtual simulator in the optimization process, (ii) obtaining the best paretos on the basis of population size of the algorithm, (iii) comparing paretos obtained from various algorithms on the basis of IGD (Inverse Generational Distance), spacing and spread, (iv) deriving a best compromise solution using fuzzy logic and linear membership function, and (v) obtaining the best solution pumping rates.

TRAINING OF SVM AS A VIRTUAL SIMULATOR

A SVM served as a virtual simulator to BIOPLUME III. The input data for SVM were pumping discharge values selected using a random number generator, whereas its outputs were $h^{max}(j_o)$, $h^{min}(j_0)$ and maximum contaminant concentration. Pumping rates ranging between 0.45 and 1.26 $L \cdot s^{-1}$ were used as well discharge rates for potential injection and extraction wells, with the fate of the plume being simulated using BIOPLUME III. After each run, the $h^{max}(j_o)$ and $h^{min}(j_0)$ along with maximum contaminant concentration were obtained. The data set thus obtained was further used in calibrating and validating the SVM.

The number of data sets required for training and testing the SVM was determined using a trial and error method, such that 70% of the data was used in calibration, and the remainder for validation. The SVM was calibrated with 400 to 1200 data sets. Coefficient of determination (R^2) and *RMSE* (Root Mean Square Error) values were determined for each case. The R^2 varied between 0.85 to 0.98, and the RMSE between 0.56 and 0.15 as the number of data sets increased from 400 to 1200 (Fig. 4).

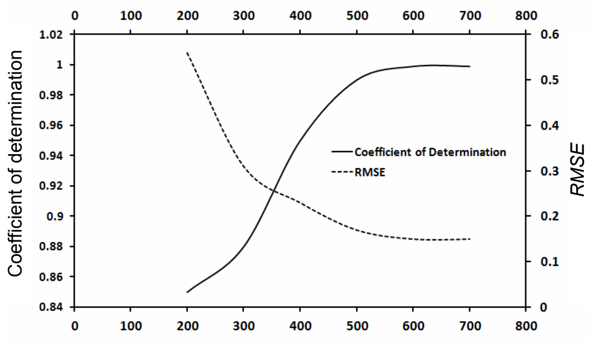

Fig. 4. Coefficient of determination and *RMSE* (Root Mean Square Error) for data; source: own study

As no change in the SVM performance occurred when the number of training data sets increased from 1000 to 1200, 1000 data sets was chosen as the optimum data set number for SVM model training. Of the 1000 data set points, 700 served in calibration and the remainder in validation of the SVM. The R^2 and *RMSE* values for actual output (from BIOPLUME III) and simulated output (from SVM) were 0.98 and 0.15, respectively, indicating that the SVM accurately predicted output values, and can thus be used effectively as a proxy simulator for BIOPLUME III.

SELECTION OF BEST PARETOS BASED ON POPULATION SIZE

The population size has always been a key parameter while determining the best Pareto solution. The proposed DGS algorithm was tested for population sizes ranging from 50 to 750. The DGS and other algorithms (DE, NSGA II and SA) were each run independently on roughly 40 occasions. The Paretos obtained for small population sizes of 50, 100 and 150 were not well-spaced (Fig. 5a–5c), while those obtained with a greater population size were more divergent (Fig. 5e, 5f).

Increasingly larger population sizes (500, 750 and more) increase the computational time and the complexity associated with the non-dominated sorting procedure for every generation. However, if the population size is too low (i.e. 50 or 100) the possibility of obtaining a local optima is much greater than that of obtaining a global optimal value. A population size of 250 yields well-spaced Paretos and results in a lower cost of remediation than population sizes of 50, 100 and 150. A similar response of Paretos was obtained for various population sizes while using the DE, NSGA II and SA algorithms. On this basis, a population size of 250 was adopted throughout the further steps of the compromise solution evaluation.

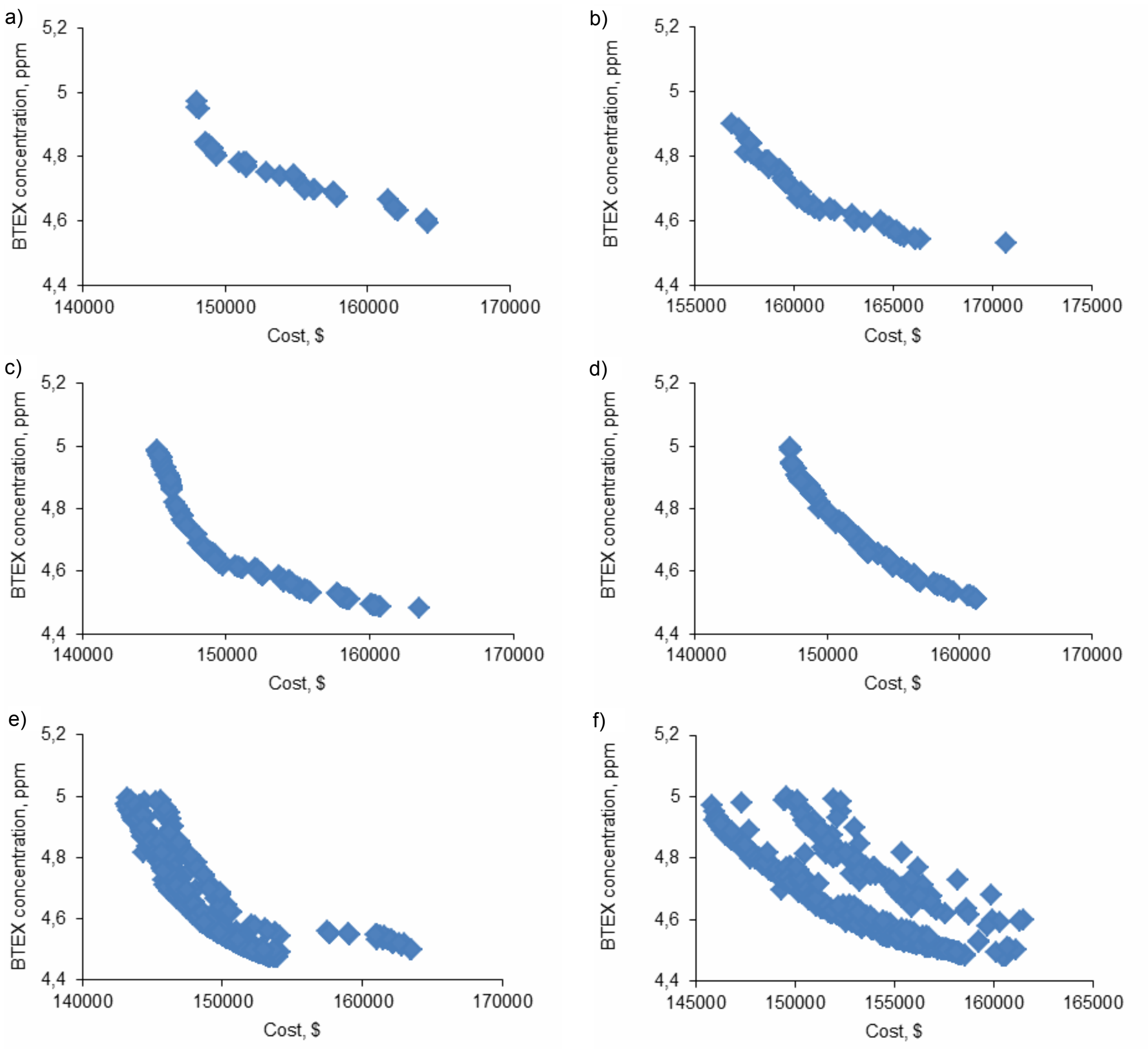

Fig. 5. Paretos obtained using DGS for population sizes of (a) 50, (b) 100, (c) 150, (d) 250, (e) 500 and (f) 750; source: own study

COMPARISON OF DGS WITH OTHER METAHEURISTIC ALGORITHMS

Low values of the Inverse Generational Distance (IGD), spacing and its spread (Tab. 2) were taken as representative of a better non-dominated solution when comparing the DE, NSGA and SA algorithms [DEB 2001]. The control parameter for the hybrid DGS algorithm (Tab. 2) and the control parameters used in NSGA, SA and DE algorithms were taken to be the same. The IGD matrix [KOTINIS 2010], Spacing matrix [SCHOOT 1995] and Spread matrix [DEB *et al.* 2000] are expressed as:

$$IGD = \frac{\sqrt{\sum_{k=1}^{P} d_k^2}}{|P|} \tag{15}$$

$$S = \frac{\sqrt{\sum_{i=1}^{|Q|} \left(d_i - \bar{d}\right)^2}}{|Q|} \tag{16}$$

$$D = \frac{\sum_{m=1}^{M} d_m^e + \sum_{i=1}^{|Q|} \left|d_i - \bar{d}\right|}{\sum_{m=1}^{M} d_m^e + |Q|\bar{d}} \tag{17}$$

where:

$\bar{d}$ = the mean value of all Euclidean distances;

d_m^e = the distance between extreme solutions of two different Paretos of the m^{th} objective function;

d_i = the Euclidean distance between the i^{th} solution and the nearest member of true pareto set,

d_k = the normalized distance between the true paretos and the computed paretos of the k^{th} objective function,

P = the number of objective functions,

Q = the set of objective functions.

For a population size of 250, the performance analysis of the various algorithms (DE, DGS, NSGA II and SA) was based on IGD, spacing and spread. In a metaheuristic algorithm, the Paretos could vary with each run; therefore, to avoid errors in the values associated with Paretos, around 40 independent runs were conducted for each algorithm. Mean and median values with box plots, the standard deviation, along with best and worst values (*i.e.*, respectively, smallest and largest IGD, spacing and spread values) for all the performance parameters were calculated after the 40 runs (Tab. 3). While IGD, spacing and spread values were similar across all algorithms, those of the DGS algorithm had the lowest values (Tab. 3). The box plot for the performance parameters (Fig. 6) represents the robust statistical parameters of median and quartile values. The lower, upper, median, 25% and 75% percentiles of the values of the performance matrix were determined. The median values of IGD for the DGS algorithm (Fig. 6a) were close to the minimum value obtained from 40 independent runs of the algorithm. For spacing and spread (Figs. 6b and 6c, the box plot for DGS is skewed upward, but its values remain less than those obtained using the other algorithms. Thus, the relatively low values of IGD, spacing and spread suggest that the DGS is more robust than the other algorithms used in the present study.

Table 3. Performance for DE, DGS, NSGA II and SA

Algorithm	Performance parameter	Worst	Best	Mean	Median	Standard deviation
DE	IGD	0.0289	0.0277	0.0285	0.0287	$0.6\cdot10^{-3}$
	Spacing	0.0210	0.0192	0.0200	0.0199	$0.6\cdot10^{-3}$
	Spread	0.0810	0.0751	0.0781	0.0779	$2.2\cdot10^{-3}$
DGS	IGD	0.0275	0.0262	0.0267	0.0269	$0.46\cdot10^{-3}$
	Spacing	0.0170	0.0140	0.0153	0.0151	$1.3\cdot10^{-3}$
	Spread	0.0579	0.0544	0.0560	0.0564	$1.39\cdot10^{-3}$
NSGA II	IGD	0.0321	0.0294	0.0303	0.0300	$1.05\cdot10^{-3}$
	Spacing	0.0201	0.0165	0.0180	0.0180	$1.40\cdot10^{-3}$
	Spread	0.0793	0.0758	0.0770	0.0766	$1.38\cdot10^{-3}$
SA	IGD	0.040	0.0350	0.0379	0.0382	$2.1\cdot10^{-3}$
	Spacing	0.0288	0.0247	0.0268	0.0268	$1.5\cdot10^{-3}$
	Spread	0.0760	0.069	0.0733	0.0740	$2.5\cdot10^{-3}$

Source: own study.

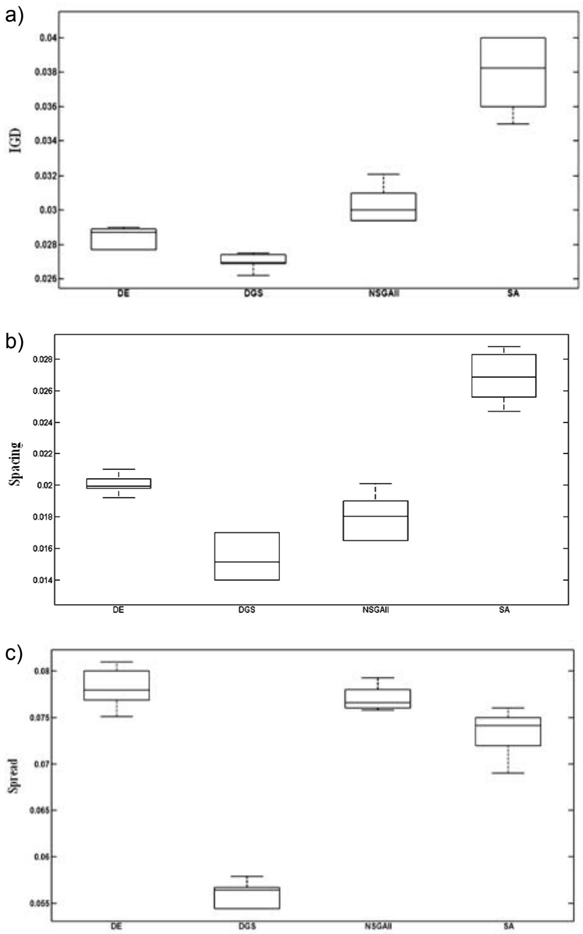

Fig. 6. Box plot comparison of (a) IGD, (b) spacing and (c) spread for various algorithms; source: own study

SELECTION OF THE BEST COMPROMISE SOLUTION

The main aim of this study was to develop a methodology to minimize the total cost of remediation as well as the residual contaminant concentration at the contaminant site. A Pareto front was obtained when the study's multi-objective optimization problem was solved through the use of the DGS algorithm. This Pareto front (Fig. 5) provides various solutions for the objective functions. All the solutions obtained are valid solutions and can thus be considered for implementation. However, out of all these possible solutions, a best compromise solution must be obtained. A fuzzy-based mechanism [ABIDO 2003; PANIGRAHI *et al.* 2010] was applied to obtain the best compromising solution from the Paretos obtained for a population size of 250 (Fig. 5d). To reduce the residual concentration to 5 ppm, an approximate cost of $147 151 is incurred, whereas, it requires $161 299 to reduce the residual concentration to 4.5 ppm (Fig. 7). The gradient of the obtained Pareto can be seen to increase substantially with an increase in cost although the reduction in residual concentration is relatively low (Fig. 7). For example, a reduction in the permissible limit of contaminant from 5 ppm to 4.6 ppm (0.4 ppm) results in an increase in cost of remediation of roughly $6000, whereas a reduction from 4.65 ppm to 4.5 ppm (0.15 ppm) results in a cost increase of $8197. It is therefore important to reach an optimal compromise solution so that the permissible concentration is reduced as much as possible without

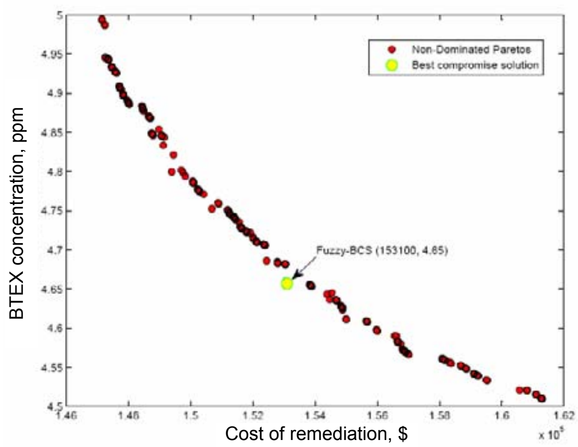

Fig. 7. Best compromising solution obtained using a DGS for population size 250; source: own study

incurring a substantial increase in remediation cost. The result obtained in terms of co-ordinate of the obtained Paretos suggests that the best compromising solution is obtained at (153100, 4.65). Thus the best strategy obtained in the present study is to remediate the contaminant site so that an initial concentration of 40 ppm will be reduced to a permissible level of 4.65 ppm after remediation, at a cost of $153 100.

PUMPING STRATEGY FOR THE BEST COMPROMISING SOLUTION

Management periods are groups of simulation time steps during which the pumping policy remains constant. However, over the course of the remediation period, the pumping rates are allowed to change. In this study, the pumping rates (negative for injection and positive for extraction) were allowed to change once in six months (Fig. 8), allowing the adjustment of pumping rates for injection and extraction wells so as to achieve the best compromise solution. The optimization algorithm set the pumping rate of the extraction wells to a near minimum value in the first three management periods. Injection pumping rates were seen to decrease in the latter half of the remediation period (Fig. 8). On the other hand, the extraction rate continuously increases from the first management period to the sixth management period (i.e., from 0.53 lps to 1.25 lps). The increased extraction rate checks the migration of plume towards monitoring wells.

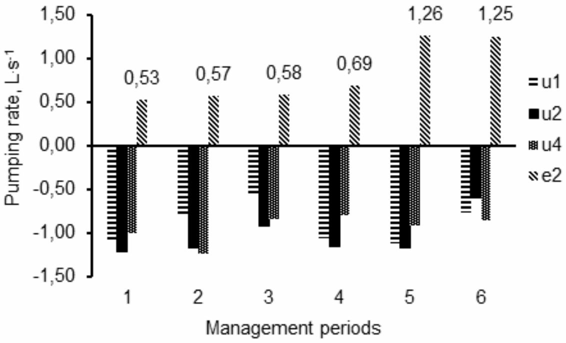

Fig. 8. Pumping rates for the best compromising solution; u = pumping well; e = extraction well; source: own study

The under engineered in-situ bioremediation condition following the pumping strategy for the best compromising solution BTEX concentration contour is shown in Fig. 8. From this figure, it can be seen that the BTEX concentration in the aquifer is 23 ppm, 12 ppm and 4.65 ppm at the end of the first, second and third year respectively (Fig. 9a, 9b, 9c).

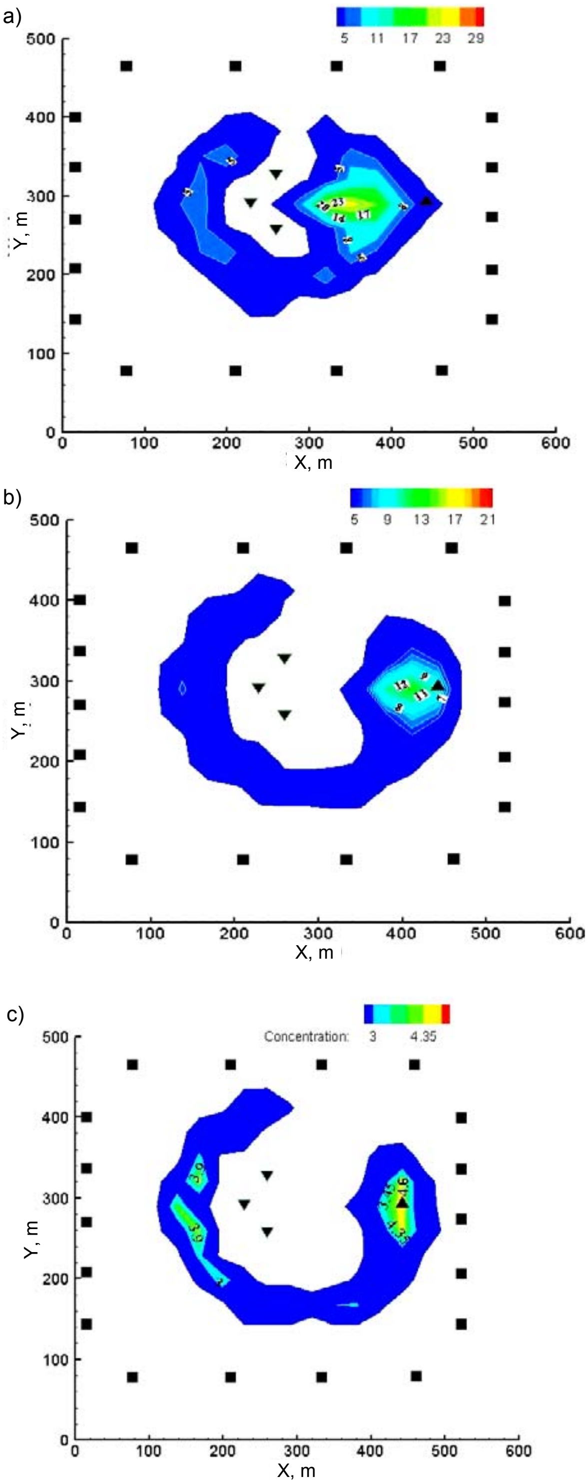

Fig. 9. BTEX contaminant contour (a) after 1 year (b) after 2 years (c) after 3 years; source: own study

CONCLUSIONS

Among the several remediation techniques available for remediation of groundwater, in-situ bioremediation is one of the most economic and eco-friendly technologies. Given the significant problems facing the water resources sector [ADAMOWSKI *et al*. 2009; 2010; 2012a; ADAMOWSKI, PROKOPH 2013; ARAGHI *et al*. 2015; BELAYNEH *et al*. 2014; CAMPISI *et al*. 2012; DANESHMAND *et al*. 2014; HAIDARY *et al*. 2013; NALLEY *et al*. 2012; 2013; NAMDAR *et al*. 2014; PINGALE *et al*. 2014; SAADAT *et al*. 2014; TIWARI, ADAMOWSKI 2014], new approaches such as the one proposed in this study are important to explore to facilitate the transition to more sustainable water resources management [BUTLER, ADAMOWSKI

2015; HALBE *et al*. 2013; INAM *et al*. 2015; KOLINJIVADI *et al*. 2014a, b; MAHMOOD *et al*. 2015; MEDEMA *et al*. 2014a, b; SAADAT *et al*. 2011; STRAITH *et al*. 2014]. In this study, to accelerate the biodegradation process, an engineered bioremediation process was studied using a set of injection and extraction wells. Simulation of BTEX contaminant and its biodegradation was studied using BIOPLUME III. Further, Support Vector Machine (SVM) was used as a proxy simulator to replace the actual simulation models so as to eventually enhance the computation efficiency of any algorithm. This study also presents an application of a hybrid algorithm which combines three metaheuristic algorithms (Differential evaluation, Genetic Algorithm, Simulated Annealing) in a single algorithm abbreviated as DGS. This hybrid algorithm served in multi-objective optimization of in situ bioremediation of groundwater, taking the cost of remediation and the permissible residual organic contaminant concentration as two separate objectives. Several sets of Paretos were then generated by varying the population size of the algorithm in this study. The results suggest that for a very small population size, the obtained Paretos are not uniformly spaced and diverged. A similar behaviour is also observed when very large population sizes of the algorithm are assumed. The Paretos obtained for a population size of 250 were found to be well spaced and well diverged for DGS. Further, the best compromise solution was obtained by adopting fuzzy logic and the associated pumping pattern was derived.

Acknowledgements

This research was partially funded by an NSERC Discovery and CFI grant held by Jan Adamowski.

REFERENCES

ABIDO M.A. 2003. Environmental/economic power dispatch using multi-objective evolutionary algorithms. IEEE Transactions on Power Systems. Vol. 18. Iss. 4 p. 1529–1537.

ADAMOWSKI J., ADAMOWSKI K., BOUGADIS J. 2010. Influence of trend on short duration design storms. Water Resources Management. Vol. 24 p. 401–413.

ADAMOWSKI J., CHAN H., PRASHER S., SHARDA V.N. 2012a. Comparison of multivariate adaptive regression splines with coupled wavelet transform artificial neural networks for runoff forecasting in Himalayan micro-watersheds with limited data. Journal of Hydroinformatics. Vol. 3 p. 731–744.

ADAMOWSKI J., PROKOPH A. 2013. Assessing the impacts of the urban heat island effect on streamflow patterns in Ottawa, Canada. Journal of Hydrology. Vol. 496 p. 225–237.

ADAMOWSKI K., PROKOPH A., ADAMOWSKI J. 2009. Development of a new method of wavelet aided trend detection and estimation. Hydrological Processes. Vol. 23 p. 2686–2696.

ADAMOWSKI K., PROKOPH A., ADAMOWSKI J. 2012b. Influence of the 11 year solar cycle on annual streamflow maxima in Canada. Journal of Hydrology. Vol. 442–443 p. 55–62.

ALI M., PANT M., ABRAHAM A. 2009. Inserting information sharing mechanism of PSO to improve the convergence of DE. World Congress on Nature and Biologically Inspired Computing. naBIC 2009. Coimbatore, India. 9–11 Dec. 2009. IEE p. 282–287.

ARAGHI A., ADAMOWSKI J., NALLEY D., MALARD J. 2015. Using wavelet transforms to estimate surface temperature trends and dominant periodicities in Iran based on gridded reanalysis data. Journal of Atmospheric Research. Vol. 11 p. 52–72.

BELAYNEH A., ADAMOWSKI J., KHALIL B., OZGA-ZIELINSKI B. 2014. Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet-support vector regression models. Journal of Hydrology. Vol. 508 p. 418–429.

BLUM C. 2005. Beam-ACO–hybridizing ant colony optimization with beam search: an application to open shop scheduling. Computers and Operations Research. Vol. 32. Iss. 6 p. 1565–1591.

BORDEN R.C., BEDIENT P.B. 1986. Transport of dissolved hydrocarbons influenced by oxygen-limited bioremediation. 1. Theoretical development. Water Resources Research. Vol. 22. Iss. 13 p. 1973–1990.

BUTLER C., ADAMOWSKI J. 2015. Empowering marginalized communities in water resources management: Addressing inequitable practices in Participatory Model Building. Journal of Environmental Management. Vol. 153 p. 153–162.

CAMPISI S., ADAMOWSKI J., ORON G. 2012. Forecasting urban water demand via wavelet- denoising and neural network models. Case study: City of Syracuse, Italy. Water Resources Management. Vol. 26 p. 3539–3558.

DANESHMAND F., KARIMI A., NIKOO M., BAZARGAN-LARI M., ADAMOWSKI J. 2014. Mitigating socio-economic--environmental impacts during drought periods by optimizing the conjunctive management of water resources. Water Resources Management. Vol. 28. Iss. 6 p. 1517–1529.

DEB K. 2001. Multi-objective optimization using evolutionary algorithms. John Wiley and Sons Ltd., England. ISBN 978-0-471-87339-6 pp. 515.

DEB K., PRATAP A, AGRAWAL S., MEYARIVAN T. 2002. A fast and elitist multi-objective genetic algorithm: NSGA-II. Technical Report, Indian Institute of Technology, Kanpur: Kanpur Genetic Algorithm Laboratory. Vol. 6. Iss. 2 p. 182–197.

ELRAGAL H.M., MANGOUD M.A., ALSHARAA M.T. 2011. Hybrid differential evolution and enhanced particle swarm optimisation technique for design of reconfigurable phased antenna arrays. IET Microwave Antennas Propagation. Vol. 5. Iss. 11 p. 1280–1287.

GUAN J., ARAL M.M. 1999. Optimal remediation with well locations and pumping rates selected as continuous decision variables. Journal of Hydrology. Vol. 221. Iss. 1–2 p. 20–42.

HAIDARY A., AMIRI B.J., ADAMOWSKI J., FOHRER N., NAKANE K. 2013. Assessing the impacts of four land use types on the water quality of wetlands in Japan. Water Resources Management. Vol. 27 p. 2217–2229.

HALBE J., PAHL-WOSTL C, SENDZIMIR J, ADAMOWSKI J. 2013. Towards adaptive and integrated management paradigms to meet the challenges of water governance. Water Science and Technology: Water Supply. Vol. 67 p. 2651–2660.

HU Z., CHAN C.W., HUANG G.H. 2004. Dynamic process control for in-situ bioremediation system. Electrical and Computer Engineering. Vol. 3 p. 1385–1388.

HU Z., CHAN C.W., HUANG G.H. 2007. Multi objective optimization for process control of the in-situ bioremediation system under uncertainty. Engineering Applications of Artificial Intelligence. Vol. 20. Iss. 2 p. 225–237.

HUANG C., MAYER A.S. 1997. Pump and treat optimization using well locations and pumping rate as decision variables. Water Resources Research. Vol. 33. Iss. 5 p. 1001–1012.

INAM A., ADAMOWSKI J., HALBE J., PRASHER S. 2015. Using causal loop diagrams for the initialization of stakeholder engagement in soil salinity management in agricultural watersheds in developing countries: A case study in the Rechna Doab watershed, Pakistan. Journal of Environmental Management. Vol. 152 p. 251–267.

KOLINJIVADI V., ADAMOWSKI J., KOSOY N. 2014a. Juggling multiple dimensions in a complex socioecosystem: The issue of targeting in payments for ecosystem services. GeoForum. Vol. 58 p. 1–13.

KOLINJIVADI V., ADAMOWSKI J., KOSOY N. 2014b. Recasting payments for ecosystem services (PES) in water resource management: A novel institutional approach. Ecosystem Services. Vol. 10 p. 144–154.

KOTINIS M. 2010. A particle swarm optimizer for constrained multi-objective engineering design problems. Engineering Optimization. Vol. 42. Iss. 10 p. 907–926.

MAHMOOD S., ADAMOWSKI J., ALAZBA A., EL-GINDY A. 2015. Rainwater harvesting for the management of agricultural droughts in arid and semi-arid regions. Paddy and Water Environment. In press.

MAYER B. 2008. Hybrids of constructive metaheuristic and constraint programming; A case study with ACO. In: Hybrid Metaheuristics: An Emerging Approach to Optimization. Berlin Heidelberg. Studies in Computational Intelligence. Vol. 114 p. 151–183.

MEDEMA W., LIGHT S., ADAMOWSKI J. 2014a. Integrating adaptive learning into water resources management. Journal of Environmental Engineering and Management. Vol. 13 p. 1801–1816.

MEDEMA W., WALS A., ADAMOWSKI J. 2014b. Virtual learning platforms for sustainable land and water governance: An innovative mechanism for facilitating multiloop social learning. Wageningen University Journal of Life Sciences. Vol. 69 p. 23-38.

MINSKER B.S., SHOEMAKER C.A. 1998. Dynamic optimal control of in-situ bioremediation of groundwater. Journal of Water Resources Planning and Management. Vol. 124. Iss. 3 p. 149–161.

MONTESINOS P., GARCIA-GUZMAN A., AYUSO J.L. 1999. Water distribution network optimization using modified genetic algorithm. Water Resources Research. Vol. 35. Iss. 11 p. 3467–3473.

NALLEY D., ADAMOWSKI J., KHALIL B. 2012. Using discrete wavelet transforms to analyze trends in streamflow and precipitation in Quebec and Ontario (1954–2008). Journal of Hydrology. Vol. 475 p. 204–228.

NALLEY D., ADAMOWSKI J., KHALIL B., OZGA-ZIELINSKI B. 2013. Trend detection in surface air temperature in Ontario and Quebec, Canada during 1967–2006 using the discrete wavelet transform. Journal of Atmospheric Research. Vol. 132/133 p. 375–398.

NAMDAR M., ADAMOWSKI J., SAADAT H., SHARIFI F., KHIRI A. 2014. A new approach to developing LULC maps for semi-arid regions using independent component analysis and expert classification. International Journal of Remote Sensing. Vol. 35 p. 8057–8073.

OLDENHUIS R.P.S. 2010. Trajectory optimization of a mission to the solar bow shock and minor planets [online]. M.Sc. thesis, Dept. of Astrodynamics and Space Missions, Delft University of Technology. [Access 06.08.2015]. Available at: http://repository.tudelft.nl/view/ir/uuid:6672aa53-6ecc-4571-85b2-2d4aafbba5bd/

PANIGRAHI B.K., RAVIKUMAR PANDI V., DAS S., DAS S. 2010. Multiobjective fuzzy dominance based bacterial foraging algorithm to solve economic emission dispatch problem. Energy. Vol. 35. Iss. 12 p. 4761–4770.

PINGALE S., KHARE D., JAT M., ADAMOWSKI J. 2014. Spatial and temporal trends of mean and extreme rainfall and temperature for the 33 urban centres of the arid and semi-arid state of Rajasthan, India. Journal of Atmospheric Research. Vol. 138 p. 73–90.

PRASAD R.K., MATHUR S. 2008. Potential well locations in in-situ bioremediation design using Neural Network Embedded Monte Carlo Approach. Practice Periodical of Hazardous, Toxic, and Radioactive Waste Management. Vol. 12. Iss. 4 p. 260–269.

PRESTWICH S. 2002. Combining the scalability of local search with the pruning techniques of systematic search. Annals of Operations Research. Vol. 115. Iss. 1–4 p. 51–72.

RAVIKUMAR PANDI V., PANIGRAHI B.K. 2006. An evolutionary algorithm approach to estimate the parameters of power quality signals. International Conference on Power Electronics, Drives and Energy Systems. New Dehli, India, 12–15 Dec. 2006. IEE p. 1–6.

ROTHBERG E. 2007. An evolutionary algorithm for polishing mixed integer programming solutions. INFORMS Journal on Computing. Vol. 19. Iss. 4 p. 534–541.

SAADAT H., ADAMOWSKI J., BONNELL R., SHARIFI F., NAMDAR M., ALE-EBRAHIM S. 2011. Land use and land cover classification over a large area in Iran based on single date analysis of satellite imagery. Journal of Photogrammetry and Remote Sensing. Vol. 66 p. 608–619.

SAADAT H., ADAMOWSKI J., TAYEFI V., NAMDAR M., SHARIFI F., ALE-EBRAHIM S. 2014. A new approach for regional scale interrill and rill erosion intensity mapping using brightness index assessments from medium resolution satellite images. Catena. Vol. 113 p. 306–313.

SCHOOT J.R. 1995. Fault tolerant design using single and multi-criteria genetic algorithm [online]. M.Sc Thesis, Dept. of Aeronautics and Astronautics, Massachusetts Institute of Technology. [Access 06.08.2015]. Available at: http://hdl.handle.net/1721.1/11582

SHIEH H.J., PERALTA R.C 2005. Optimal in-situ bioremediation design by hybrid genetic algorithm-simulated annealing. Journal of Water Resources Planning and Management. Vol. 131. Iss. 1 p. 67–78.

STRAITH D., ADAMOWSKI J., REILLY K. 2014. Exploring the attributes, strategies and contextual knowledge of champions of change in the Canadian water sector. Canadian Water Resources Journal. Vol. 39. Iss. 3 p. 255–269.

TIWARI M., ADAMOWSKI J. 2013. Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. Water Resources Research. Vol. 49. Iss. 10 p. 6486–6507.

TSE S.M., LIANG Y., LEUNG K.S., LEE K.-H., MOK T.S.K. 2007. A memetic algorithm for multiple-drug cancer chemotherapy schedule optimization. IEEE Transactions on Systems, Man, and Cybernetics. Part B. Vol. 37. Iss. 1 p. 84–91.

VINKO T., IZZO D., BOMBARDELLI C. 2007. Benchmarking different global optimisation techniques for preliminary space trajectory design. 58th Int. Astronautical Congress, Hyderabad. IAF p. 24–28.

WANG M., ZHENG C. 1997. Optimal remedial policy selection under general conditions. Ground Water. Vol. 35. Iss. 5 p. 757–764.

ZHANG S., GONG S.X., GUAN Y., ZHANG P.F., GONG Q. 2009. A novel IGA-EDSPSO hybrid algorithm for the synthesis of sparse arrays. Progress in Electromagnetics Research. Vol. 89 p. 121–134.

Deepak KUMAR, Sudheer CH, Shashi MATHUR, Jan ADAMOWSKI

Wielozadaniowa optymalizacja bioremediacji wód gruntowych in situ z zastosowaniem hybrydowej techniki metaheurystycznej opartej na zróżnicowanej ewolucji, algorytmach genetycznych i symulowanym wyżarzaniu

STRESZCZENIE

Słowa kluczowe: *bioremediacja in-situ, algorytm hybrydowy, algorytm genetyczny, logika rozmyta, maszyna wektorów nośnych (SVM), symulowane wyżarzanie, wody gruntowe, zróżnicowana ewolucja*

Zanieczyszczenie wód gruntowych wyciekami benzyny jest jedną z kilku przyczyn wpływających na środowisko wód podziemnych. W ostatnich latach bioremediacja *in situ* przyciągała uwagę badaczy z powodu jej zdolności do usuwania zanieczyszczeń w ich siedlisku i niskich kosztów procesu. Przedstawiona praca proponuje użycie nowego algorytmu hybrydowego do optymalizacji wielozadaniowej funkcji, która obejmuje koszty remediacji jako pierwsze zadanie i resztową zawartość zanieczyszczeń po zakończeniu procesu jako drugie z zadań. Algorytm hybrydowy powstał z połączenia metod różnicowej ewolucji, algorytmu genetycznego i symulowanego wyżarzania. Maszyna wektorów nośnych (SVM) została użyta jako wirtualny symulator biologicznej degradacji zanieczyszczeń w wodach gruntowych. Wyniki uzyskane z algorytmy hybrydowego porównano z wynikami zróżnicowanej ewolucji (DE), algorytmu genetycznego (NSGA II) i symulowanego wyżarzania (SA). Stwierdzono, że proponowany algorytm był w stanie zapewnić najlepsze rozwiązanie. Użyto metody z zakresu logiki rozmytej dla znalezienia najlepszego rozwiązania kompromisowego i na końcu przedstawiono dla tego rozwiązania strategię szybkości pompowania celem remediacji wód gruntowych. Wyniki pokazały, że koszty ponoszone na rozwiązanie kompromisowe są pośrednie między najwyższymi i najniższymi kosztami innych rozwiązań.