Multiobjective Design of Water-Quality Monitoring Networks in River-Reservoir Systems

Mahyar Aboutalebi, S.M.ASCE[1]; Omid Bozorg-Haddad[2]; and Hugo A. Loáiciga, F.ASCE[3]

Abstract: This study develops and tests a method for multiobjective optimization of water-quality monitoring networks in river-reservoir systems. The optimization method identifies optimal sampling locations to detect the sudden release of contaminants to a reservoir, and meets two objectives: (1) minimizing the prediction error of methyl tert-butyl ether (MTBE) at the reservoir's outlet valve; and (2) minimizing the average time during which MTBE is detected at sampling locations. The optimization method considers 36 contaminant scenarios, corresponding to three volumes of contaminant release, three release locations, and four different seasonal release times. The MTBE pollutant chemograph was simulated at the outlet valve of the reservoir and at 16 possible sampling locations with the CE-QUAL-W2 model for each of the 36 scenarios of contaminant release. A support vector regression (SVR) tool is coupled to the nondominated sorting genetic algorithm II (NSGAII) to optimize the water-quality sampling locations. Implementation of the NSGAII-SVR method demonstrates its capacity to design water-quality monitoring networks that meet multiple objectives in a river-reservoir system. **DOI: 10.1061/(ASCE)EE.1943-7870.0001155.**

Author keywords: Monitoring network; CE-QUAL-W2; Support vector regression; NSGAII.

Introduction

Among various reservoir optimization schemes reported in the literature (Ahmadi et al. 2014; Bolouri-Yazdeli et al. 2014; Ashofteh et al. 2013a, 2015b), groundwater resources (Bozorg-Haddad et al. 2013b; Fallah-Mehdipour 2013a), conjunctive use operation (Fallah-Mehdipour 2013b), design-operation of pumped-storage and hydropower systems (Bozorg-Haddad et al. 2014a; Aboutalebi and Garousi-Nejad 2015), flood management (Bozorg-Haddad et al. 2015b), water project management (Orouji et al. 2014), hydrology (Ashofteh et al. 2013b), qualitative management of water resources systems, (Orouji et al. 2013; Bozorg-Haddad et al. 2015a; Shokri et al. 2014), water distribution systems (Seifollahi-Aghmiuni et al. 2013; Soltanjalili et al. 2013; Beygi et al. 2014), agricultural crops (Ashofteh et al. 2015c), sedimentation (Shokri et al. 2013), and algorithmic developments (Ashofteh et al. 2015a), a few dealt with the design of water-quality monitoring networks in river-reservoir systems considering multiple objectives.

Studies concerning the water-quality monitoring networks in river-reservoir system can be divided into three general categories. These categories are the performance assessment of water-quality monitoring networks, data mining in water resources, and water-quality monitoring optimization. Topics such as relocating sampling locations fall within the first category. The use of data mining tools and their performance in water resource problems belong in the second category. The development and optimization models and tools to optimize the design of monitoring network belong to the third category.

Concerning the first category, Ning and Chang (2001) studied the assessment of quality monitoring network in the Kao-Ping River located in southern Taiwan. Naddeo et al. (2007) analyzed the frequency of sampling of riverine water quality. Khalil and Ouarda (2009) examined statistical methods for assessing and redesigning existing surface water-quality monitoring networks. Afshar and Marino (2012) used the ant colony optimization algorithm (ACO) to assess monitoring stations in a water distribution network in Iran.

Data mining is a systematic search for new and valuable information in large databases. Data mining has two primary aims: classification and forecasting. The most important tools in data mining are artificial neural network (ANN), genetic programming (GP), and support vector machine (SVM). These three tools are widely used in classification and forecasting. A study involving these three tools in water resource systems is described in this paper. Behzad et al. (2009) evaluated the performance of ANN and support vector regression (SVR) in forecasting runoff of the Bakhtiyari River in Iran and showed that the SVR method exhibited better performance in runoff forecasting than ANN and hybrid ANN with genetic algorithm (ANN-GA). Yoon et al. (2011) applied ANN and SVR to predict the groundwater level in coastal aquifers located in Korea. Their results indicated the better performance of SVR than that of ANN. Singh et al. (2011) used SVR for water-quality management in the city of Lacknow, India. These authors used support vector classification (SVC) and SVR in the prediction of biochemical oxygen demand (BOD) and concluded that the SVR method had better performance than kernel versions of the discriminant analysis, kernel partial least squares, linear discriminant analysis, and partial least squares. Wei (2012) predicted water surface elevation of the Tanshui River in China by using a wavelet kernel function in SVR and reported that the predictive skill of SVR with wavelet kernel function was better than that of SVR with Gaussian kernel function.

[1]M.Sc. Graduate, Dept. of Irrigation and Reclamation Engineering, Faculty of Agricultural Engineering and Technology, College of Agriculture and Natural Resources, Univ. of Tehran, 31587 77871 Karaj, Iran. E-mail: Aboutalebi@ut.ac.ir

[2]Associate Professor, Dept. of Irrigation and Reclamation Engineering, Faculty of Agricultural Engineering and Technology, College of Agriculture and Natural Resources, Univ. of Tehran, 31587 77871 Karaj, Iran (corresponding author). E-mail: OBHaddad@ut.ac.ir

[3]Professor, Dept. of Geography, Univ. of California, Santa Barbara, CA 93106. E-mail: Hugo.Loaiciga@ucsb.edu

Note. This manuscript was submitted on December 7, 2015; approved on May 23, 2016; published online on July 18, 2016. Discussion period open until December 18, 2016; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Environmental Engineering***, © ASCE, ISSN 0733-9372.

Maity et al. (2013) evaluated the ability of SVR in predicting the monthly river flow discharge of the Mahanadi River in India and concluded that predictions with SVR were more accurate than those achieved with autoregressive moving average (ARMA). Aboutalebi et al. (2016) simulated the methyl tert-butyl ether (MTBE) pollutants chemograph in the Karaj river-reservoir located near the city of Karaj, Iran, by coupling the genetic algorithm (GA) with SVR (SVR-GA) to achieve optimal SVR parameters. Aboutalebi et al. (2016) compared the SVR-GA with the ANN and GP, and results indicated the superior performance of SVR-GA. More comprehensive discussions about applications of data mining tools in water resources management are found in Bozorg Haddad et al. (2013a, 2014b).

The design of optimal water-quality monitoring networks is essential to the management of surface and groundwater resources that supply water to municipalities, industry, and agriculture. Icaga (2005) applied the GA to optimize a water-quality monitoring network in the Gediz River, Turkey, and concluded the number of existing of monitoring stations could be reduced from 33 to 14 while maintaining the monitoring capacity of the network. Park et al. (2006) applied the GA and geographic information systems (GIS) to design water-quality monitoring networks in large rivers, showing that only 35 of 110 stations were properly located in the Nakdong River, South Korea. Kollat et al. (2008) presented a multiobjective evolutionary algorithm (MOEA) to design a groundwater monitoring network. The name of this algorithm was Epsilon-dominance hierarchical Bayesian optimization algorithm. Their results showed that the performance of the (ε-HBOA) was better than that of the Epsilon-dominance nondominated sorting genetic algorithm II (ε-NSGAII). Telci et al. (2009) proposed that the design of riverine water-quality monitoring network includes locating the sampling stations and determining the frequency of sampling and reported a new method to achieve optimal sampling locations based on minimizing the detection time of pollutants in the Altamaha river in Georgia. Yongqian et al. (2011) optimized water-quality monitoring with principal component analysis (PCA) and cluster analysis (CA) in the Yangtze river, China and concluded that only five stations are necessary to monitor several water-quality characteristics instead of the existing 16 stations. Mei et al. (2011) determined the optimal sampling locations in the Tangue River, China, by using continuous longitudinal monitoring data (CLMD) and cluster analysis (CA). The results indicated that the method is effective in determining the optimal locations for water-quality monitoring.

Khalil et al. (2011) divided the monitoring area for the Nile River's catchment area within Egypt into subcatchments and determined the optimal number of quality monitoring places in each subcatchment by using a stratified optimum sampling strategy, and results established that only 39 among 50 existing monitoring stations are necessary for river monitoring. Cetinkaya and Harmancioglu (2012) determined the optimal number of monitoring stations in the Gediz River, Turkey, by means of dynamic programming (DP), concluding that only 10 of 14 stations are properly located and that four monitoring stations could be removed from the network. Lee et al. (2014) reported an algorithm based on Shannon's entropy theory (Shannon 1948) and GA to optimize the sampling locations of water-quality monitoring in the Albert and Logan Rivers, Australia. The advantage of this algorithm relative to the GA and simulated annealing (SA) was that the algorithm was computationally faster and more accurate than the GA and SA.

A review of published works on water-quality monitoring network revealed that network optimization for the purpose of detecting the sudden release of pollutants has not been addressed. This study's primary goal was to locate the sampling stations of a water-quality monitoring network in a river-reservoir system to detect suddenly released pollutant (MTBE) into the system at various locations and times. The problem of designing the water-quality monitoring network has two objective functions: (1) minimizing the root-mean square error (RMSE) of the predictions of MTBE at the reservoir outlet; and (2) minimizing the average time during MTBE is detected at sampling locations. The first objective is justified because the most important chemograph of MTBE in the river-reservoir system is that at the outlet valve of the reservoir, which is used to devise pollutant-management strategies by operating the outlet valve to control water release. In this manner the chemographs monitored at sampling locations can be used to simulate the chemograph of the reservoir outlet with SVR. The second objective assures that the water-quality monitoring network can detect pollutants present in water. The SVR is applied to simulate the chemograph of MTBE at the reservoir outlet valve based on a database created from simulations of MTBE concentration in the river-reservoir system using the CE-QUAL-W2 model. NSGAII is applied to select the best water-quality sampling locations and to optimize the parameters of the SVR.

Methodology

This section presents a brief description of the CE-QUAL-W2 model, the theoretical foundations of SVR, and NSGAII. More detailed descriptions are available in Aboutalebi and Bozorg Haddad (2015) and Aboutalebi et al. (2015a, b).

CE-QUAL-W2

CE-QUAL-W2 is a two-dimensional water quality model that simulates water-quality characteristics in the longitudinal/vertical directions, and can predict water surface elevations, velocities, temperatures, and a number of water quality features. CE-QUAL-W2 routes water through cells in a grid network and solves multiple equations for each time step of the simulation. In this model, water bodies can be mapped through multiple branches and cells. Inflows to the water body and outflows from the water body can be point or nonpoint sources, branches, precipitation, and other variables. There is a convenient analysis tool in this model that provides options for calculating results with versatility.

Support Vector Machine

The theory of SVM was introduced by Vapnik et al. (1995) as a method for classification; and subsequently, Vapnik et al. (1998) developed SVM as a forecasting or predicting tool. The regression form of SVM (SVR) is described next.

Support Vector Regression

SVR was introduced by Vapnik et al. (1998) as a method for regression analysis. Vapnik et al. (1998) considered two functions for defining SVR. The first function is called the error function and calculates the SVR error between observed and predicted values of a variable of interest, in this case, MTBE concentrations in a reservoir. The second function predicts output values based on input data, weights, and bias. The epsilon insensitive (e-intensive) function or error function defined by Vapnik (1998) for SVR is as follows:

$$|y - f(\mathbf{x})| = \begin{cases} 0 & \text{if } |y - \mathrm{f}(\mathbf{x})| \le \kappa \\ |y - f(\mathbf{x})| - \kappa = \xi & \text{otherwise} \end{cases} \quad (1)$$

where $\mathbf{x}$ = vector of input variables; y = value of the observed output (in this study, this is the MTBE concentration simulated with the

water-quality model CE-QUAL-W2); $f(\mathbf{x})$ = predicted value of MTBE concentration calculated by SVR; κ = sensitivity of prediction error $|y - f(\mathbf{x})|$ (this is a SVR parameter); ξ = penalty for the values that are outside the range $(-\kappa, +\kappa)$; and $|\ldots|$ = absolute value operator. In the e-insensitive function, there is no error or penalty if y is located between $f(\mathbf{x}) + \kappa$ and $f(\mathbf{x}) - \kappa$. Otherwise, ξ is considered as a penalty.

The second (linear) function is the computational function of SVR that predicts MTBE concentrations to be compared with observed concentrations (simulated with CE-QUAL-W2). The second function is given by

$$f(\mathbf{x}) = \mathbf{w}^T \cdot \mathbf{x} + b \tag{2}$$

where the vector $\mathbf{w}$ = values of the weights of the vector of variables $\mathbf{x}$; b = value of the bias of the dot product $\mathbf{w}^T \cdot \mathbf{x}$; and T = transpose sign.

Calculation of $\mathbf{w}$ and b in SVR constitutes an optimization problem. The target of the optimization problem is minimizing the epsilon intensive function with respect to the weighting vector $\mathbf{w}$ and the bias b. Besides calculating these two objects ($\mathbf{w}$ and b), the calculated responses are located in the range $(-\kappa, +\kappa)$, which is defined as a constraint in the optimization model. There are two approaches to defining these constraints in SVR: the hard margin and the soft margin. This study relies on the soft margin approach of SVR, in which case the optimization model with constraints is defined as follows:

$$\text{Minimize} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{m}(\xi_i^- + \xi_i^+) \tag{3}$$

Subject to:

$$\begin{aligned} &(\mathbf{w}^T \cdot \mathbf{x} + b) - y_i < \kappa + \xi_i^+, \quad i = 1, 2, \ldots, m \\ &y_i - (\mathbf{w}^T \cdot \mathbf{x} + b) \le \kappa + \xi_i^-, \quad i = 1, 2, \ldots, m \qquad \xi_i^+, \quad \xi_i^- \ge 0 \end{aligned} \tag{4}$$

where C = coefficient of penalty (this is an SVR parameter); m = number of training data (75% of outlet chemographs in this study); ξ_i^-, ξ_i^+ = violation values that are located below and above the range of $(-\kappa, +\kappa)$, respectively; and $y_i = i$th observed (calculated with CE-QUAL-W2) value. To solve the optimization problem [Eqs. (3) and (4)], the Lagrange of objective function (L) is formed. Its partial derivatives with respect to variables b and $\mathbf{w}$ are formed and set equal to zero and solved for b and $\mathbf{w}$. The SVR-predicted MTBE value can then be obtained with Eq. (2) using the calculated values of b and $\mathbf{w}$. The values of C and κ are SVR parameters that are determined using the NSGAII according to methodology presented subsequently.

Nonlinear Support Vector Regression

The MTBE data can be transformed using transfer functions into a space in which linear functions can be fitted to the data whenever linear functions [Eq. (2)] do not fit the raw MTBE data. The transfer function is called the kernel function. Dibike et al. (2001) used different kernel functions to process rainfall-runoff with SVR and showed that the radial basis function (RBF) kernel had better performance than other functions. Han and Cluckie (2004) concluded that the RBF produced the best performance in a study of transformation effectiveness for regression analysis. The RBF equation is given by

$$K(x, x_i) = \exp\left(-\frac{|x - x_i|^2}{2\gamma^2}\right) \quad i = 1, 2, \ldots, m \tag{5}$$

where γ = RBF parameter, which is the parameter to handle nonlinear regression. The kernel functions convert the nonlinear behavior of the data to approximate linear behavior for fitting purposes using linear function of the type shown in Eq. (2).

Determination of the SVR Parameters

The SVR parameters are the SVR's C, κ, and the parameter γ of the RBF kernel function. The performance of SVR depends on these three parameters. Therefore, their optimal choice is essential. Several approaches, such as grid search algorithms and metaheuristic algorithm, are used to calculate the SVR parameters. The parameters are considered as decision variables, and the objective function is the maximization of the accuracy of the data mining tool in metaheuristic algorithms. The SVR parameters are determined by means of NSGAII in this study, following the work of Aboutalebi and Bozorg Haddad (2015).

Criteria for evaluating the results of SVR

Eq. (6) is used to measure the effectiveness of SVR

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(Hyd_{\text{obs}} - Hyd_{\text{sim}})^2}{n}} \tag{6}$$

where n = number of observed data in the MTBE chemograph at the outlet valve of the reservoir for testing date set; Hyd_{sim} = chemograph at the outlet valve of the reservoir predicted with SVR; and Hyd_{obs} = concentration of MTBE simulated with CE-QUAL-W2 model at the outlet valve of the reservoir.

Nondominated Sorting Genetic Algorithm II

Aboutalebi and Bozorg-Haddad (2015) linked the nondominated sorting genetic algorithm II (NSGAII) to SVR and proposed a new tool called SVR-NSGAII. This tool has the capability of calculating the optimized SVR parameters (C, κ, and γ) considering various objective functions. This tool is implemented in this study to optimize the three parameters of the SVR and to find the optimal locations of the sampling stations. The NSGAII introduced by Deb (2001) uses a concept called nondominated sorting. The NSGAII starts with the generation of a random parent population of solutions.

The NSGAII combines one generation of parent solutions and one of children solutions to produce a set of solutions R. Set R is determined based on crossover and mutation processes (Deb 2001). The best members of set R are chosen as the parents of the next generation, and the other members of the set R are removed from further consideration. The process of combining a generation of parent and children solutions to produce a new generation of parent solutions is called elitism in the NSGAII. Fig. 1 illustrates the process of elitism in the NSAGII.

After the new parent generation is produced in the NSGAII, the nominees for reproduction for the next generation are chosen by the use of a crowed tournament selection operator, and generation of children solutions is produced by using crossover and mutation operators.

A stopping criterion is needed to indicate the end of the search for a solution by the NSGAII. A commonly used stopping criterion is to set a maximum allowable number of iterations of the algorithm. Once the number of iterations of the algorithm reaches a

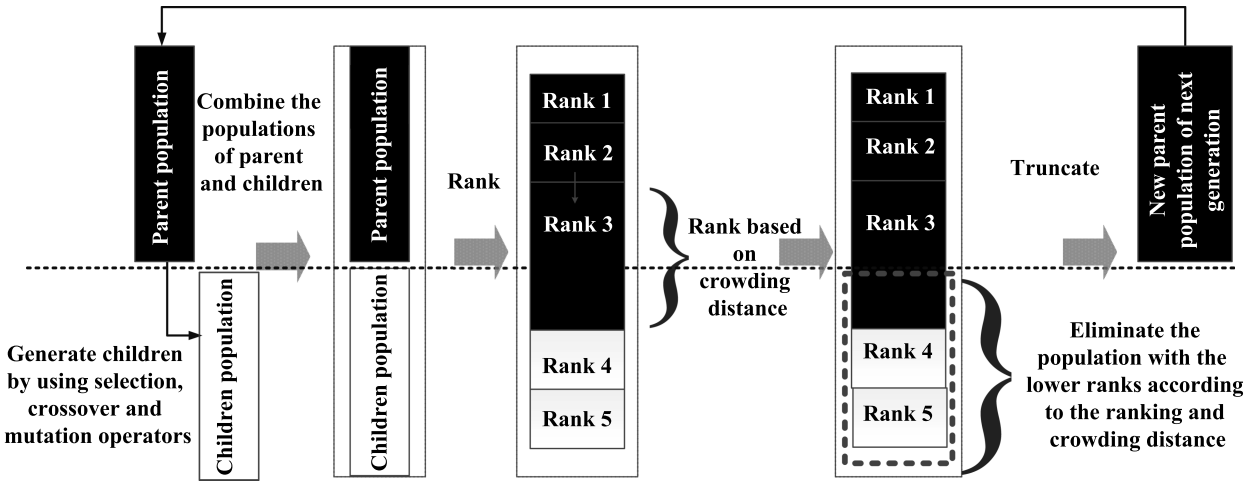

Fig. 1. Elitism process of the NSGAII

certain value, the algorithm is stopped, and the optimal Pareto frontier (the set of nondominated solutions) for the component objective functions of the overall objective function is calculated. Fig. 2 depicts the NSGAII steps.

According to Fig. 2, the initial population is produced in the NSGAII, and the objective function values are calculated and evaluated to produce a population. Next, the parent generations are modified, and crossover and mutation are applied to obtain the population of children solutions. At this stage, the objective function values are calculated and evaluated again, and the combination of parent and children generations are sorted. Then the crowding distance for all Pareto frontiers is calculated. Based on the criteria of crowding distance and ranking, the parent generation is selected. If the stopping condition is satisfied, the search algorithm stops; otherwise, the objective function value for the selected parent population is re-evaluated, and the search process for a solution continues.

Optimization Model

The optimization problem must be defined mathematically to integrate the SVR simulations with the NSGAII optimization. Eqs. (7)–(10) describe the optimization model

$$\min g_1 = \mathrm{RMSE}(Hyd_{\mathrm{sim.test}}, Hyd_{\mathrm{obs.test}}) \tag{7}$$

$$\min g_2 = \mathrm{Mean}(\boldsymbol{T}_{det}) \tag{8}$$

$$Hyd_{\mathrm{sim}} = \mathrm{Data\ mining}[Hyd(\boldsymbol{Samp}), Hyd_{\mathrm{obs}}, P_{DM}] \tag{9}$$

$$1 < \boldsymbol{Samp} \leq 16 \tag{10}$$

where g_1 = first objective function (accuracy of outlet valve chemograph simulation by SVR based on RMSE); $Hyd_{\mathrm{sim.test}}$ = chemograph

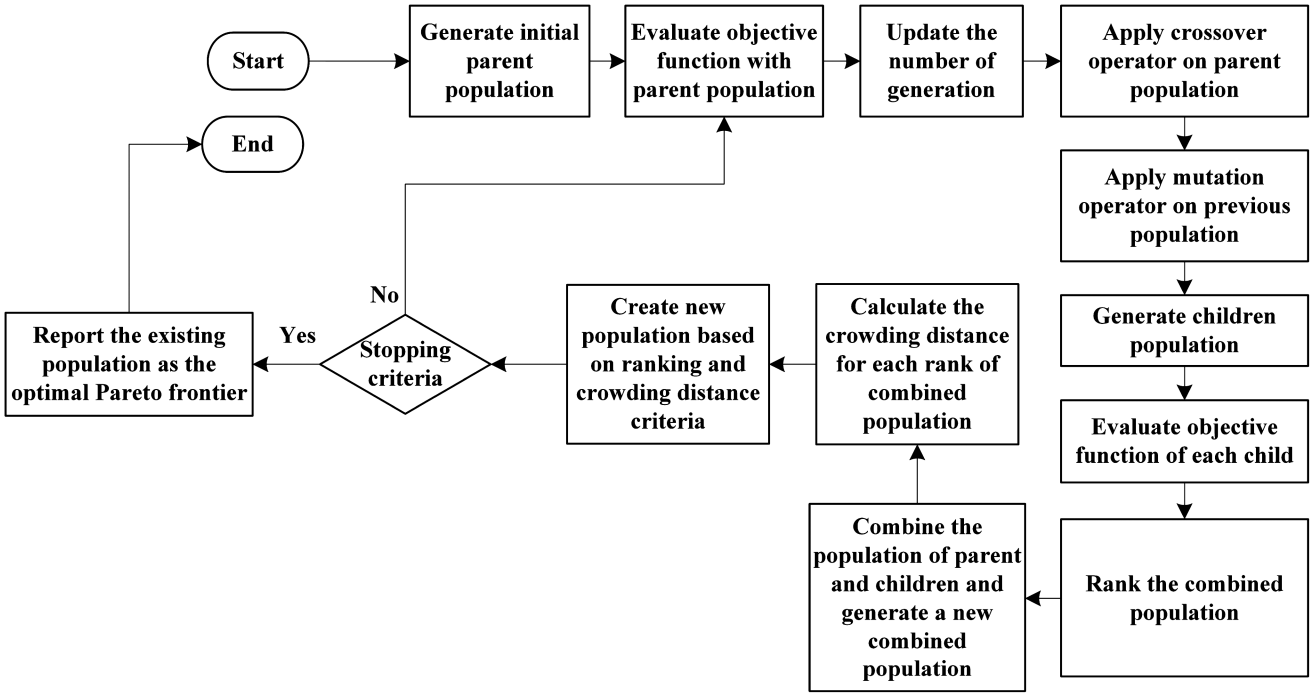

Fig. 2. Steps of the NSGAII

J. Environ. Eng., 2017, 143(1): 04016070

at reservoir outlet valve simulated by SVR for the testing data set; $Hyd_{obs.test}$ = chemograph at reservoir outlet valve simulated by CE-QUAL-W2 for the testing data set; g_2 = second objective function (mean detection time of MTBE); ***Samp*** = vector of sampling locations that are selected by NSGAII; $\boldsymbol{T_{det}}$ = vector of detection time of MTBE at sampling locations presented by NSGAII; Hyd_{sim} = chemograph of reservoir outlet valve that is simulated by SVR; P_{DM} = vector of SVR parameters; and Data mining denotes the data mining tool, which is SVR.

Case Study

The situation dealt with in this study is the accidental, sudden release of three possible volumes of MTBE equal to 15,000, 30,000, and 60,000 L at three possible places in the Karaj river-reservoir system (upstream location, middle of the reservoir, and downstream location), in four seasons (summer, autumn, winter, and spring). The combination of three volumes of release with three release locations and four release times creates $(3 \times 3 \times 4 =)36$ release scenarios in the Karaj river-reservoir system.

Simulation of MTBE Transport in the Karaj Reservoir

The model CE-QUAL-W2 was implemented to simulate MTBE transport in the Karaj river-reservoir (Shokri et al. 2014) for each of the 36 scenarios of MTBE release to the Karaj river-reservoir (Aboutalebi et al. 2016). Shokri et al. (2014) considered a grid composed of 21 segments and 46 layers for the Karaj river-reservoir system in the MTBE simulations with CE-QUAL-W2. Fig. 3 displays a cross-sectional view of the possible MTBE release locations and of the reservoir grid used in the simulations.

Cells numbered 1–16 in Fig. 3 are possible sampling locations, and Cell 17 is the location of the outlet valve of the reservoir. The simulation/optimization steps of the method applied in this study are depicted in Fig. 4.

First, the CE-QUAL-W2 model was run for the 36 release scenarios, and the results of the 36 simulations calculated in Cells 1–17 cells were stored in a database of MTBE concentrations. Cells 1–16 are possible sampling locations; Cell 17 is the location of the outlet valve of the reservoir. Then the data-mining SVR is connected to the optimization NSGAII. The evolutionary multiobjective optimization algorithm, NSGAII, randomly produces an initial population (a set of chromosomes). Each chromosome contains two parts: the first part contains the numbers of sampling cells (vary from 1 through 16, as shown in Fig. 3), and the second part contains the SVR parameters. Next, NSGAII extracts the MTBE chemograph and the detection time of MTBE from the database of simulated MTBE concentrations created with CE-QUAL-W2. Next, NSGAII uses SVR and the set of SVR parameters to predict the MTBE concentration chemograph [with Eq. (2)]. In this study, the NSGAII has as its first objective function (denoted by g_1) the minimization of the RMSE of predictions. Its second objective function (denoted by g_2) is the minimization of the average time during which MTBE is detected at sampling cells (average detection time). As the number of sampling location increases, the RMSE decreases, whereas the average detection time increases, thus creating a trade-off between the two objectives that is typical of Pareto solutions.

The optimization problem is solved for 16 possible sampling locations. In other words, it is first solved assuming that there is only one possible sampling location (whose cell number is between 1 and 16). Next, it is solved allowing at most two possible sampling locations, and so on, until solving the problem in which there are at most 16 possible sampling locations. The last case, when there are at most 16 sampling locations, has the largest feasible space of solutions. At the end of this sequential, incremental solution process, 16 Pareto optimal solutions are obtained. Each solution contains the numbers of the optimal sampling cells and values of the three SVR parameters used for MTBE prediction. These solutions are represented as a Pareto frontier.

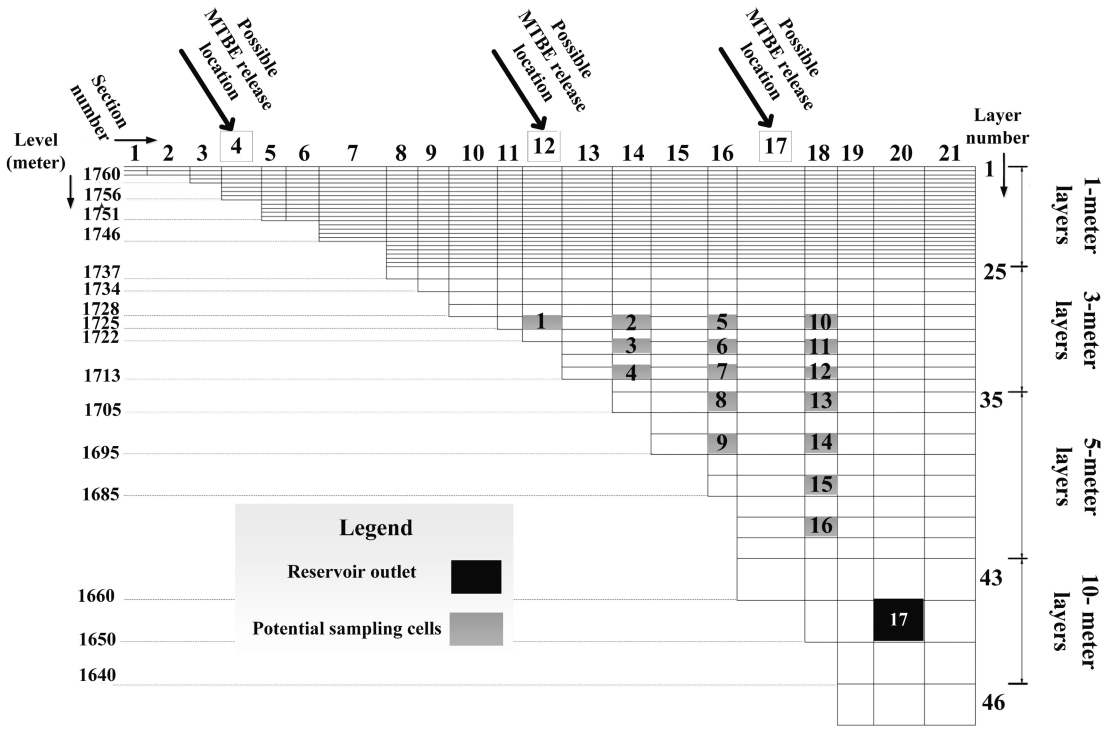

Fig. 3. Cross-sectional longitudinal view of Karaj reservoir discretization and possible release locations (Sections 4, 12, and 17)

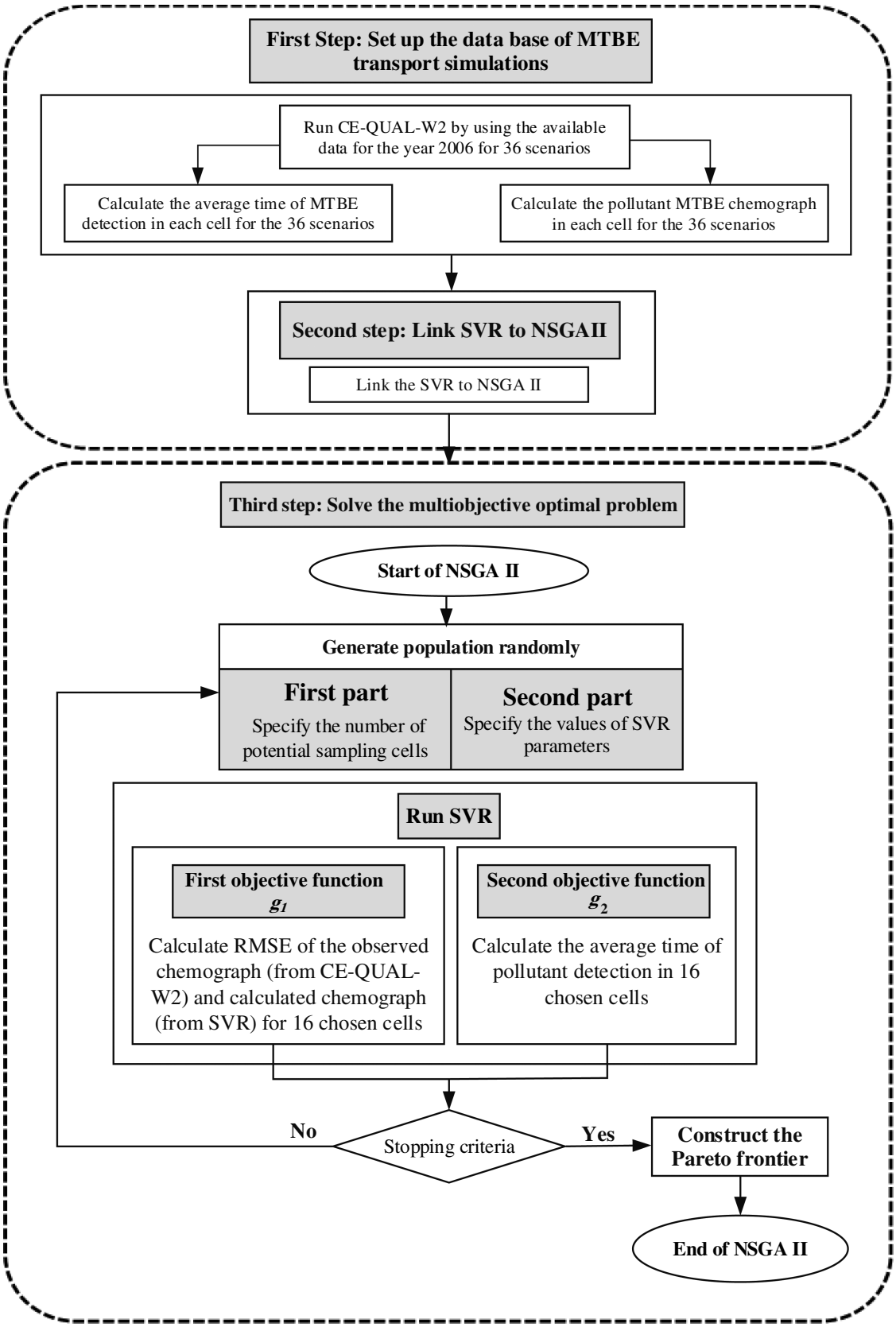

Fig. 4. Simulation/optimization steps of the study

The values of the NSGAII parameters (percentage of cross over, the percentage of mutations, number of iteration, and the number of initial population) were set equal to 70%, 15%, 1,000, and 50, respectively. The RBF function was used in SVR, and 75% of the data (chosen randomly) are considered as training data [m data values in Eq. (3)], and the remaining data are considered as the testing data (n data values). The *MATLAB* software was used for implementation of SVR.

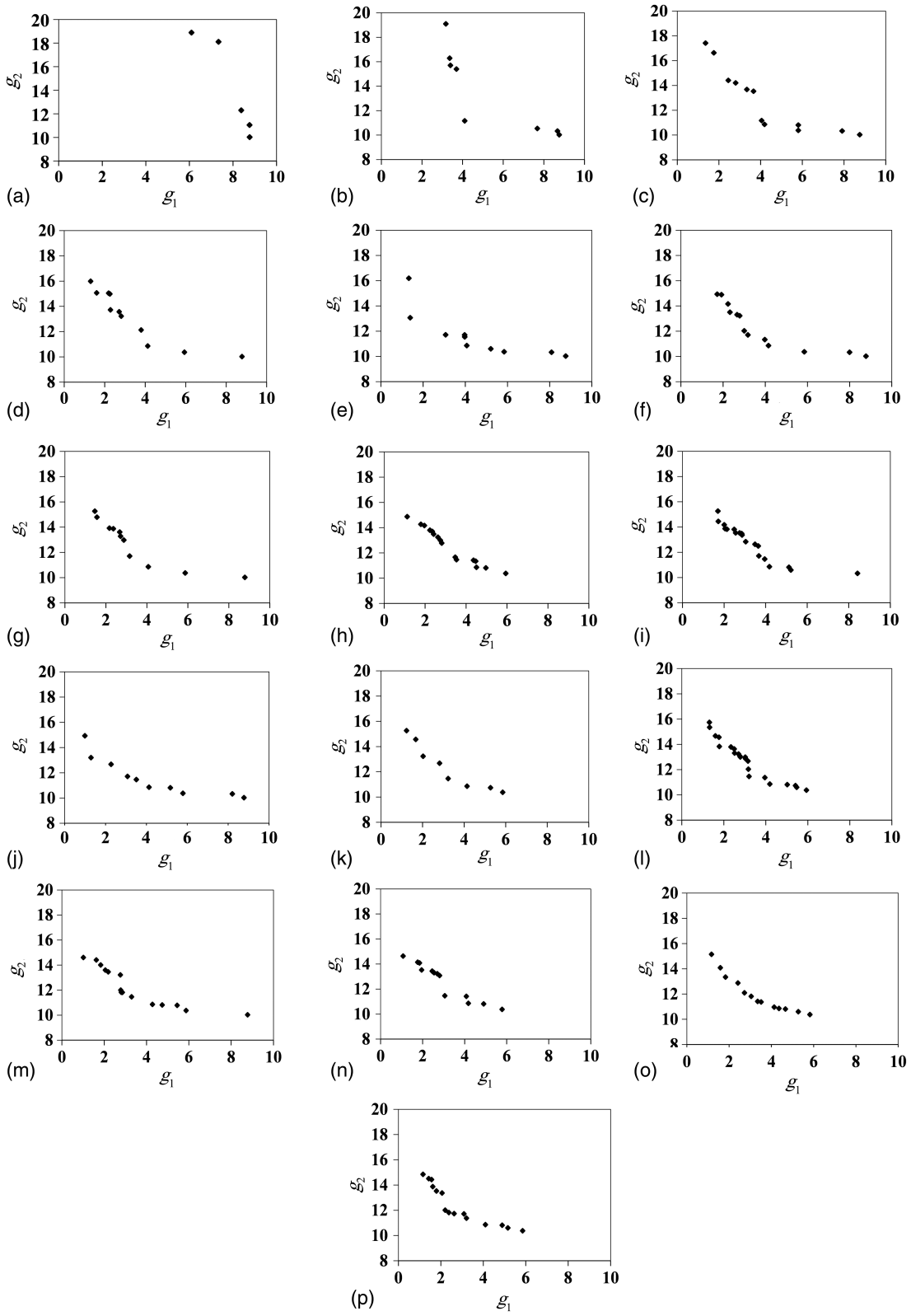

Fig. 5. Sixteen extracted Paretos for maximum allowed number of sampling locations (MANSL) 1–16; extracted Pareto for MANSL = (a) 1; (b) 2, . . . ; and (p) 16

Results and Discussion

The results of solving the 16 optimization problems described in the previous sections using NSGAII and the prediction tool SVR are shown in Fig. 5. There are 16 different Pareto frontiers in Fig. 5, each corresponding to the maximum number of sampling cells allowed in each problem solved. Each point on a Pareto front represents a solution to the corresponding problem. The value of the second objective function (minimize the detection time of the MTBE) varies between 8 and 20 h. The value of the first

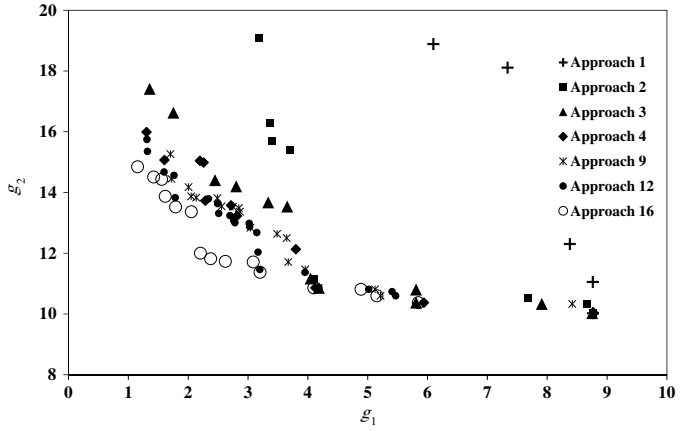

Fig. 6. Pareto solutions of Array 16 (at most 16 sampling cells) compared with the Pareto solutions of several other arrays

objective (minimize the RMSE of predictions) ranges between 1 and 10 units of RMSE calculated from the SVR for MTBE predictions and the CE-QUAL-W2 simulations. This range of accuracy indicates the high predictive skill of the SVR tool in simulating MTBE concentrations for the 36 considered scenarios. Fig. 6 portrays the Pareto fronts of the 16 different possible arrays of sampling sites, among which, the Pareto front of the 16th array is the most nondominated. This is so because this approach has the largest number of possible sampling locations, and therefore, has the largest feasible solution space.

Table 1 lists the values of the 15 Pareto solutions corresponding to Array 16. Table 1 shows that the most frequently and least frequently used sampling cells are Cells 12 and 2, respectively. The solutions in Table 1 are listed from top to bottom in order of increasing value of the first objective. Evidently, as the number of sampling locations decreases, the first objective function (minimization of prediction error, denoted by g_1) increases, whereas the

Table 1. 15 Pareto Solutions Calculated for Array 16 (at Most 16 Sampling Cells)

Array number	Sampling cell number																SVR parameters			Objective functions	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	κ	γ	C	g_1	g_2
1	1	1	1	1	1	1	1			—	1	1	—	1	1	1	0.005	5.95	106	1.15	14.84
2	—	1	—	1	1	—	—	—	1	—	1	—	1	—	—	—	0.017	12.66	28	1.42	14.50
3	—	1	—	1	1	—	—	—	1	—	—	1	1	—	—	—	0.018	16.06	33	1.55	14.43
4	—	—	—	1	1	1	—	—	—	—	—	—	1	—	—	—	0.028	40.05	69	1.61	13.87
5	—	1	—	—	1	—	1	1	1	—	1	—	1	—	—	—	0.001	29.29	39	1.79	13.52
6	—	1	—	—	—	1	1	1	1	—	1	—	1	—	—	—	0.026	29.17	48	2.05	13.36
7	1	1	—	1	1	—	—	1	1	—	—	—	—	—	—	—	0.026	39.35	65	2.20	12.00
8	1	1	—	1	—	1	—	1	1	—	—	—	—	—	—	—	0.021	39.23	43	2.37	11.81
9	—	1	1	1	1	—	—	1	—	—	—	—	—	—	—	—	0.020	27.31	83	2.62	11.73
10	—	—	—	1	—	1	—	1	—	—	—	—	—	—	—	—	0.016	40.07	51	3.08	11.71
11	1	1	—	—	—	1	—	1	1	—	—	—	—	—	—	—	0.024	29.36	55	3.20	11.37
12	—	—	1	—	—	1	—	—	—	—	—	—	—	—	—	—	0.004	54.67	73	4.09	10.86
13	—	1	1	—	—	1	—	—	—	—	—	—	—	—	—	—	0.052	56.88	52	4.88	10.81
14	—	1	—	—	—	1	—	1	—	—	—	—	—	—	—	—	0.020	15.89	23	5.15	10.60
15	—	1	—	—	—	1	—	—	—	—	—	—	—	—	—	—	0.004	46.45	48	5.84	10.37
16	—	1	—	—	—	1	—	—	—	—	—	—	—	—	—	—					

Note: g_1 = minimize prediction errors; g_2 = minimize the average detection time.

Table 2. Frequency of Selections of Sampling Cells with Arrays 1–16

Array number	Sampling cell number																Pareto points
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
1	0	0	0	0	0	1	0	1	1	0	0	1	0	1	0	0	5
2	0	0	0	0	0	5	2	1	2	1	0	0	1	1	0	2	8
3	0	2	4	1	0	8	3	1	3	3	0	0	2	3	0	0	12
4	1	1	6	2	2	6	2	1	4	1	2	0	3	3	0	1	11
5	2	4	2	2	1	7	2	5	2	1	0	0	0	0	0	3	10
6	1	1	5	4	3	10	3	2	7	0	2	0	3	2	1	0	13
7	0	2	2	6	3	10	0	4	3	2	2	0	4	0	0	1	11
8	2	12	9	4	9	6	3	5	7	0	1	1	2	1	3	4	16
9	4	2	4	1	3	3	3	2	4	0	2	1	4	0	0	0	19
10	0	5	3	2	4	7	2	3	3	0	1	0	1	0	0	2	10
11	0	4	4	2	6	4	1	2	5	0	2	0	4	0	0	0	8
12	5	20	12	2	10	18	10	15	12	3	8	7	2	9	1	6	22
13	4	10	9	3	8	7	6	7	8	2	2	3	2	2	1	0	16
14	4	11	6	2	7	10	5	7	8	1	6	2	7	1	1	1	14
15	6	10	4	5	4	9	4	5	8	1	1	0	2	0	0	3	13
16	4	12	4	8	7	10	3	8	7	0	4	2	5	1	1	1	15

Table 3. Relative Frequency of Selection of Sampling Cells for Arrays 1–16 (%)

Array number	Sampling location number																
	1	2	3	4	5	6	7	8	9	10	11	11	12	13	14	15	16
1	0	0	0	0	0	20	0	20	20	0	0	0	20	0	20	0	0
2	0	0	0	0	0	62	25	12	25	12	0	0	0	12	12	0	25
3	0	16	33	8	0	8	25	8	25	25	0	0	0	16	25	0	0
4	9	9	54	18	18	54	18	9	36	9	18	18	0	27	27	0	9
5	20	40	20	20	10	70	20	50	20	10	0	0	0	0	0	0	30
6	8	8	38	31	23	77	23	15	54	0	15	15	0	23	15	8	0
7	0	18	18	55	27	91	0	36	27	18	18	18	0	36	0	0	9
8	12	75	56	25	56	37	19	31	44	0	6	6	6	12	6	19	25
9	21	11	21	5	16	16	16	11	21	0	11	11	5	21	0	0	0
10	0	50	30	20	40	70	20	30	30	0	10	10	0	10	0	0	20
11	0	50	50	25	75	50	13	25	62	0	25	25	0	50	0	0	0
12	23	91	55	9	45	82	45	68	55	14	36	36	32	9	41	5	27
13	25	63	56	19	50	44	37	44	50	13	13	13	19	13	13	6	0
14	29	78	43	14	50	71	36	50	57	7	43	43	14	50	7	7	7
15	46	77	31	38	31	69	31	38	61	7	7	7	0	15	0	0	23
16	26	80	26	53	46	66	20	53	46	0	26	26	13	33	6	6	6

second objective function (maximization of the average detection time, denoted by g_2) decreases.

Table 2 lists the frequency of selection of each sampling cell corresponding to the 16 arrays. The last column of Table 2 lists the number of Pareto solutions of each array. Table 3 lists the frequency of selection of sampling cells for each array as a percentage. Table 3 shows that sampling Cells 2, 5, and 9 are frequent choices across all the 16 arrays. Fig. 7 shows the minimum, average, and maximum values of the SVR parameters κ, γ, and C associated with approaches 1–16. The optimal values of κ, γ, and C range from 0 to 0.06, from 2 to 70, and from 18 to 105, respectively. The average value of κ, γ, and C are 0.025, 35, and 60 nearly, respectively. Their maximum values are 12, 2, and 7, respectively.

Concluding Remarks

The aim of this study was to develop and test a multiobjective optimization method for a water-quality monitoring network in a river-reservoir system deployed to test sudden releases of MTBE or other pollutants. The method presented is novel and significant in several respects: (1) designing the monitoring network to detect the sudden release of pollutants by relying on two objective functions that maximize the accuracy of the simulated chemograph of pollutant at the release gate of a reservoir and minimize the average detection time of pollutant at sampling locations; (2) using SVR and improving the accuracy of pollutant predictions; and (3) linking the SVR tool with the NSGAII so that the prediction parameters

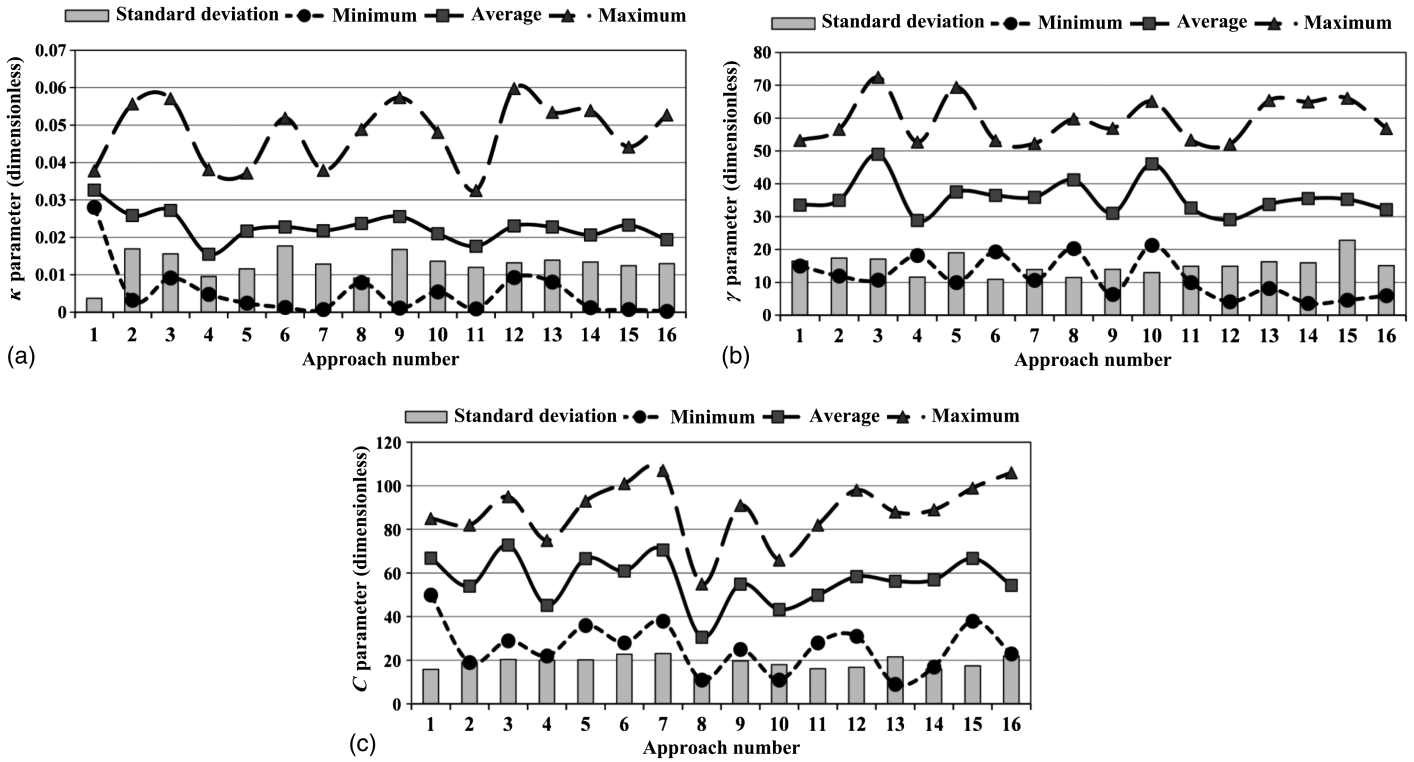

Fig. 7. Minimum, average, and maximum SVR parameters for Arrays 1–16: (a) κ parameter; (b) γ parameter; (c) C parameter

reach their optimal values during the extracting of optimal sampling locations. A database was first prepared based on the CE-QUAL-W2 model, and an SVR tool was used as a predictor to develop a rapid method for predicting pollutant concentrations in the Karaj river-reservoir system. The SVR and NSGAII were linked to find the best sampling locations for water-quality monitoring in the Karaj river-reservoir considering a various maximum number of allowed sampling locations.

References

Aboutalebi, M., and Bozorg Haddad, O. (2015). "Support vector machine with non-dominated sorting genetic algorithm for the monthly inflow prediction in hydropower reservoir." *Int. J. Civ. Struct. Eng.*, 2(1), 239–242.

Aboutalebi, M., Bozorg Haddad, O., and Loáiciga, H. A. (2015a). "Application of the SVR-NSGAII to hydrograph routing in open channels." *J. Irrig. Drain. Eng.*, 10.1061/(ASCE)IR.1943-4774.0000969, 04015061.

Aboutalebi, M., Bozorg Haddad, O., and Loáiciga, H. A. (2015b). "Optimal monthly reservoir operation rules for hydropower generation derived with SVR- NSGAII." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000553, 04015029.

Aboutalebi, M., Bozorg Haddad, O., and Loáiciga, H. A. (2016). "Simulation of methyl tertiary butyl ether concentrations in river-reservoir systems using support vector regression." *J. Irrig. Drain. Eng.*, 10.1061/(ASCE)IR.1943-4774.0001007, 04016015.

Aboutalebi, M., and Garousi-Nejad, I. (2015). "Discussion of 'Application of the water cycle algorithm to the optimal operation of reservoir systems' by Omid Bozorg Haddad, Mojtaba Moravej, and Hugo A. LoÆiciga." *J. Irrig. Drain. Eng.*, 10.1061/(ASCE)IR.1943-4774.0000947, 07015029.

Afshar, A., and Mariño, M. A. (2012). "Multi-objective coverage-based ACO model for quality monitoring in large water networks." *Water Resour. Manage.*, 26(8), 2159–2176.

Ahmadi, M., Bozorg-Haddad, O., and Mariño, M. A. (2014). "Extraction of flexible multi-objective real-time reservoir operation rules." *Water Resour. Manage.*, 28(1), 131–147.

Ashofteh, P., Bozorg-Haddad, O., Akbari-Alashti, H., and Mariño, M. (2015a). "Determination of irrigation allocation policy under climate change by genetic programming." *J. Irrig. Drain. Eng.*, 10.1061/(ASCE)IR.1943-4774.0000807, 04014059.

Ashofteh, P., Bozorg-Haddad, O., and Loáiciga, H. (2015b). "Evaluation of climatic-change impacts on multi-objective reservoir operation with multiobjective genetic programming." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000540, 04015030.

Ashofteh, P., Bozorg-Haddad, O., and Mariño, M. (2013a). "Climate change impact on reservoir performance indices in agricultural water supply." *J. Irrig. Drain. Eng.*, 10.1061/(ASCE)IR.1943-4774.0000496, 85–97.

Ashofteh, P., Bozorg-Haddad, O., and Mariño, M. (2015c). "Risk analysis of water demand for agricultural crops under climate change." *J. Hydrol. Eng.*, 10.1061/(ASCE)HE.1943-5584.0001053, 04014060.

Ashofteh, P.-S., Bozorg-Haddad, O., and Mariño, M. A. (2013b). "Scenario assessment of streamflow simulation and its transition probability in future periods under climate change." *Water Resour. Manage.*, 27(1), 255–274.

Behzad, M., Asghari, K., Eazi, M., and Palhang, M. (2009). "Generalization performance of support vector machines and neural networks in runoff modeling." *Exp. Syst. Appl.*, 36(4), 7624–7629.

Beygi, S., Bozorg-Haddad, O., Fallah-Mehdipour, E., and Mariño, M. A., (2014). "Bargaining models for optimal design of water distribution networks." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000324, 92–99.

Bolouri-Yazdeli, Y., Bozorg-Haddad, O., Fallah-Mehdipour, E., and Mariño, M. A. (2014). "Evaluation of real-time operation rules in reservoir systems operation." *Water Resour. Manage.*, 28(3), 715–729.

Bozorg Haddad, O., Aboutalebi, M., and Garousi-Nejad, I. (2014a). "Discussion of hydroclimatic stream flow prediction using least square-support vector regression." *ISH J. Hydraul. Eng.*, 20(3), 274–275.

Bozorg Haddad, O., Aboutalebi, M., and Marino, M. A. (2013a). "Discussion of prediction of missing rainfall data using conventional and artificial neural network techniques." *ISH J. Hydraul. Eng.*, 19(2), 76–77.

Bozorg-Haddad, O., Ashofteh, P.-S., Ali-Hamzeh, M., and Mariño, M. A. (2015a). "Investigation of reservoir qualitative behavior resulting from biological pollutant sudden entry." *J. Irrig. Drain. Eng.*, 10.1061/(ASCE)IR.1943-4774.0000865, 04015003.

Bozorg-Haddad, O., Ashofteh, P.-S., and Mariño, M. A. (2015b). "Levee's layout and design optimization in protection of flood areas." *J. Irrig. Drain. Eng.*, 10.1061/(ASCE)IR.1943-4774.0000864, 04015004.

Bozorg-Haddad, O., Ashofteh, P.-S., Rasoulzadeh-Gharibdousti, S., and Mariño, M. A. (2014b). "Optimization model for design-operation of pumped-storage and hydropower systems." *J. Energy Eng. (ASCE)*, 10.1061/(ASCE)EY.1943-7897.0000169, 04013016.

Bozorg-Haddad, O., Rezapour Tabari, M. M., Fallah-Mehdipour, E., and Mariño, M. A. (2013b). "Groundwater model calibration by meta-heuristic algorithms." *Water Resour. Manage.*, 27(7), 2515–2529.

Cetinkaya, C., and Harmancioglu, N. (2012). "Assessment of water quality sampling sites by a dynamic programming approach." *J. Hydrol. Eng.*, 10.1061/(ASCE)HE.1943-5584.0000420, 305–317.

Deb, K. (2001). "A fast and elitist multiobjective genetic algorithm: NSGAII." *IEEE Trans. Evol. Comput.*, 6(2), 182–197.

Dibike, Y. B., Velickov, S., Solomatine, D. P., and Abbott, M. B. (2001). "Model induction with support vector machines: Introduction and application." *J. Comput. Civ. Eng.*, 10.1061/(ASCE)0887-3801(2001)15:3(208), 208–216.

Fallah-Mehdipour, E., Bozorg-Haddad, O., and Mariño, M. A. (2013a). "Extraction of optimal operation rules in aquifer-dam system: A genetic programming approach." *J. Irrig. Drain. Eng.*, 10.1061/(ASCE)IR.1943-4774.0000628, 872–879.

Fallah-Mehdipour, E., Bozorg-Haddad, O., and Mariño, M. A. (2013b). "Prediction and simulation of monthly groundwater levels by genetic programming." *J. Hydro-Environ. Res.*, 7(4), 253–260.

Han, D., and Cluckie, I. (2004). "Support vector machines identification for runoff modeling." *Proc., 6th Int. Conf. on Hydroinformatics*, Singapore.

Icaga, Y. (2005). "Genetic algorithm usage in water quality monitoring networks optimization in Gediz (Turkey) river basin." *Environ. Monit. Assess.*, 108(1–3), 261–277.

Khalil, B., and Ouarda, T. B. (2009). "Statistical approaches used to assess and redesign surface water-quality-monitoring networks." *J. Environ. Monit.*, 11(11), 1915–1929.

Khalil, B., Ouarda, T. B., and St-Hilaire, A. (2011). "A statistical approach for the assessment and redesign of the Nile delta drainage system water-quality-monitoring locations." *J. Environ. Monit.*, 13(8), 2190–2205.

Kollat, J. B., Reed, P. M., and Kasprzyk, J. R. (2008). "A new epsilon-dominance hierarchical Bayesian optimization algorithm for large multiobjective monitoring network design problems." *Adv. Water Resour.*, 31(5), 828–845.

Lee, H., Paik, K., Yoo, D., and Kim, J. (2014). "Efficient method for optimal placing of water quality monitoring stations for an ungauged basin." *J. Environ. Manage.*, 132(1), 24–31.

Maity, R., Bhagwat, P., and Bhatnagar, A. (2013). "Potential of support vector regression for prediction of monthly streamflow using endogenous property." *Hydrol. Processes*, 24(7), 917–923.

MATLAB [Computer software]. MathWorks, Natick, MA.

Mei, K., Zhu, Y., Liao, L., Dahlgren, R., Shang, X., and Zhang, M. (2011). "Optimizing water quality monitoring networks using continuous longitudinal monitoring data: A case study of Wen-Rui Tang river, Wenzhou, China." *J. Environ. Monit.*, 13(10), 2755–2762.

Naddeo, V., Zarra, T., and Belgiorno, V. (2007). "Optimization of sampling frequency for river water quality assessment according to Italian implementation of the EU water framework directive." *Environ. Sci. Policy*, 10(3), 243–249.

Ning, S. K., and Chang, N. B. (2001). "Multi-objective, decision-based assessment of a water quality monitoring network in a river system." *J. Environ. Monit.*, 4(1), 121–126.

Orouji, H., Bozorg-Haddad, O., Fallah-Mehdipour, E., and Mariño, M. A. (2013). "Modeling of water quality parameters using data-driven models." *J. Environ. Eng.*, 10.1061/(ASCE)EE.1943-7870.0000706, 947–957.

Orouji, H., Bozorg-Haddad, O., Fallah-Mehdipour, E., and Mariño, M. A. (2014). "Extraction of decision alternatives in project management: Application of hybrid PSO-SFLA." *J. Manage. Eng.*, 10.1061/(ASCE)ME.1943-5479.0000186, 50–59.

Park, S. Y., Choi, J. H., Wang, S., and Park, S. S. (2006). "Design of a water quality monitoring network in a large river system using the genetic algorithm." *Ecol. Modell.*, 199(3), 289–297.

Seifollahi-Aghmiuni, S., Bozorg-Haddad, O., and Mariño, M. A. (2013). "Water distribution network risk analysis under simultaneous consumption and roughness uncertainties." *Water Resour. Manage.*, 27(7), 2595–2610.

Shannon, C. E. (1948). "A mathematical theory of communication." *Bell Syst. Tech. J.*, 27(4), 623–656.

Shokri, A., Bozorg-Haddad, O., and Marino, M. A. (2014). "Multi-objective quantity-quality reservoir operation in sudden pollution." *Water Resour. Manage.*, 28(2), 567–586.

Shokri, A., Bozorg-Haddad, O., and Mariño, M. A. (2013). "Reservoir operation for simultaneously meeting water demand and sediment flushing: Stochastic dynamic programming approach with two uncertainties." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000244, 277–289.

Singh, K. P., Basant, N., and Gupta, S. (2011). "Support vector machines in water quality management." *Anal. Chim. Acta*, 703(2), 152–162.

Soltanjalili, M., Bozorg-Haddad, O., and Mariño, M. A. (2013). "Operating water distribution networks during water shortage conditions using hedging and intermittent water supply concepts." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000315, 644–659.

Telci, I. T., Nam, K., Guan, J., and Aral, M. M. (2009). "Optimal water quality monitoring network design for river systems." *J. Environ. Manage.*, 90(10), 2987–2998.

Vapnik, V. (1995). *The nature of statistical learning theory*, Springer-Verlag, New York.

Vapnik, V. (1998). *Statistical learning theory*, Wiley, New York.

Wei, C. (2012). "Wavelet kernel support vector machines forecasting techniques: Case study on water-level predictions during typhoons." *Exp. Syst. Appl.*, 39(5), 5189–5199.

Yongqian, C., Qianwu, S., and Hongmei, M. (2011). "Research on optimization of water quality monitoring sites using principal component analysis and cluster analysis." *Int. Conf. on Computer Distributed Control and Intelligent Environmental Monitoring*, Changsha, China, 570–573.

Yoon, H., Jun, S., Hyun, Y., Bae, G., and Lee, K. (2011). "A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer." *J. Hydrol.*, 396(1–2), 128–138.