International Journal of GEOMATE, May., 2021, Vol.20, Issue 81, pp.79-85
ISSN: 2186-2982 (P), 2186-2990 (O), Japan, DOI: https://doi.org/10.21660/2021.81.6209
Geotechnique, Construction Materials and Environment

CHARACTERIZATION OF GROUNDWATER POLLUTION SOURCES BY KRIGING BASED LINKED SIMULATION OPTIMIZATION

*Anirban Chakraborty[1] and Om Prakash[1]

[1]Department of Civil and Environmental Engineering, Indian Institute of Technology Patna

* Corresponding Author, Received: 08 Apr. 2020, Revised: 12 Jan. 2021, Accepted: 13 Feb. 2021

ABSTRACT: Often untreated wastes are buried underground as convenient method of disposal. These clandestine disposal methods are adopted to offset the cost incurred in proper waste management. Though there might be short term economic gain to the disposer but the long-term impact of such disposal manifests in form of groundwater pollution, which if left unchecked would potentially pollute the entire aquifer. Any contaminated aquifer treatment technique requires prior information about the pollutant sources characteristics. Thus, characterization of unknown groundwater pollution sources becomes imperative in remediation process. The first step involves the identification of number of sources and their locations. In a clandestine scenario the entire study area is treated as a potential source and the search for the actual source can be extensively rigorous in space and time. To overcome this challenge a kriging based zoning is adopted to narrow down the search space for the actual sources in a limited data availability scenario. A comparative analysis of two methods is presented; first method where only (Simulated Annealing based Linked Simulation Optimization) SALSO is used, and second method where kriging based zoning plus SALSO is adopted for pollution source characterization. It is found that kriging method clubbed with SALSO gives significantly better results compared to initial results. A boundary of Indian Institute of Technology Patna with hypothetical hydro-geological parameters is used to test the developed method. The location of the pollutant sources is easily identified when a kriging based Simulated Annealing Linked Simulation Optimization (KSALSO) method is adopted.

Keywords: Linked simulation optimization; *Groundwater pollution source identification*; *Simulated annealing*; *Kriging*

INTRODUCTION

In several parts of the world, ground water is one of the most important drinking sources. Pollution of these groundwater aquifers is the biggest threat to sustainability of environmental aspects and even more difficult to detect. Clandestine dumping of toxic wastes underground to offset the cost incurred in proper waste management has further complicated the scenario. Finding such sources of pollution is highly challenging due to their untracebility. Presences of such pollutant sources are unknown until pollution is first detected in some sparsely located wells. In designing an effective remediation strategy for these polluted aquifers, accurate knowledge of pollution source characteristics in terms of their numbers, their locations with release flux histories are most vital. It may be possible to decipher the number of sources from the pattern of land use, but the greatest challenge is to estimate the history of release flux and location of these pollution sources.

Nonlinear regression model was applied to identify the release flux histories of groundwater pollution sources [1]. Linked Simulation Optimization (LSO) models are very popular in identifying the release histories of pollution sources [2-5]. Surrogate models are efficiently used to recover the release histories of groundwater pollution sources when locations of sources are well known [6-7]. These methods are valid for only some simple case studies, which cannot be used for complex scenarios of source identification problems involving clandestine sources.

There are very few methods which simultaneously identify the locations and release flux histories of groundwater pollution sources [8-9]. One of the biggest limitations of these methods are that these methods assume the location of the sources to be anywhere in the study area, which makes these methods computationally expensive. However, these developed methods fail to yield any fruitful result when the study area is large and there is no initial guess about the number of sources and their location. In such scenarios the number of possibilities to be evaluated in LSO is very large and may not converge to an optimal solution. [10] used a geo-statistical kriging method to identify the number of sources and their locations using pollutant concentration measurement from a sequentially designed monitoring network.

In order to reduce the computational burden and accurately identify groundwater pollution sources, a KSALSO methodology is proposed. This interpolation technique termed as kriging method is applied to determine the pollutant concentration profile over the study area. This concentration profile gives an overview of location of pollutant sources which reduces the area of search. Thereafter, LSO method is used after kriging interpolation for source identification. Here, Simulated Annealing (SA) is used as optimization algorithm. A comparative study is also presented between the developed KSALSO model and conventional SA based SALSO technique in identifying unknown source characteristics.

METHODOLOGY

In order to estimate Location and Release Flux histories (LRF) of these pollutant sources, two different pathways are adopted. In the first case conventional SALSO is used. Conventional SALSO was mostly used previously to identify the release flux histories of pollutant sources. In the current study, SALSO methods are used to identify both location and release flux histories. Here, a newly developed KSALSO method is implemented for simultaneous identification of LRF of pollutant sources. The search space for the sources is narrowed down by kriging based pre-processing to reduce the computational budget.

Linked Simulation-Optimization

Linked Simulation Optimization (LSO) is widely used in groundwater pollution source identification problems [11-12]. A linked simulation-optimization is initialized with suitable optimization control parameters. Each iteration of the optimization model decreases the residual error by suitably adjusting the values of the explicit decision variables based on the algorithm control parameters i.e. initial temperature, temperature reduction factor, etc. in case of SA. Decision variables constitute of the unknown source characteristics i.e. the grid locations of the sources, and flux release history such that the optimal values of these decision variables gives the optimal estimate of the unknown source characteristics. The initial values of pollutant flux of each source are taken randomly within the prescribed bounds depending on the problem scenario. The iterative process is repeated until the solution converges to an optima or reaches the stopping criteria as specified by the user. The proposed methodology uses meta-heuristic optimization algorithm clubbed with a simulation model (Equation 1 & Equation 2) to reconstruct LRF of pollutant sources.

$$k_i \nabla^2 h + R_h = S_s \frac{\partial h}{\partial t} \tag{1}$$

Equation 1 represents groundwater flow equation and is based on Darcy's law. The hydraulic head is denoted by h; k_i is the hydraulic conductivity in *i* th direction R_h is the storitivity of the aquifer; *t* is the elapsed time.

$$\frac{\partial C}{\partial t} = D_{i,j} \nabla^2 C - \frac{\partial h}{\partial t} \nabla c + C_s q_s \tag{2}$$

Equation 2 simulates the transport of pollutant in groundwater system. *C* represents the concentration of pollutant sources; $D_{i,j}$ is the disversivity coefficient at *i*, *j*; the term $C_s q_s$ represents the source sink mixing in groundwater system;

In LSO approach, the residual error between simulated and observed pollutant concentration values are minimized in Equation 3. Simulation model replicates flow and pollutant transport in groundwater system [13]. The unknown LRF are considered as explicit unknown decision variables. For different combinations of source numbers, locations and fluxes, simulation model is solved iteratively, such that the optimal solution represents the best estimate of the unknown decision variables. SA is used to solve the optimization problem. In LSO method, location and release flux histories (LRF) of pollutant sources are estimated by formulating optimization problem.

$$F = Min \sum_1^k \sum_{iob}^{mob} (Cobs_{iob}^k - Cest_{iob}^k) \tag{3}$$

Subjected to,

$$L_{min} \le L \le L_{max}; \tag{4}$$

In Equation 3, objective function is minimized at any arbitrary location *iob* and time *k; mob* is the maximum number of observation well $Cobs_{iob}^k$ is the observed concentration and $Cest_{iob}^k$ is the simulated concentration at measurement time step *k* at location *iob;* $Cobs_{iob}^k$ is obtained from sampling an observation well; and *;* $Cest_{iob}^k$ is obtained from Equation (5);

$$Cest_{iob}^k = f(L, F, H) \tag{5}$$

L is the location (grid) of pollutant sources and it depends on the study area and constrained by Equation 4, *F* is the release concentration of the sources and *H* represents the hydrogeological characteristics of the polluted aquifer study area. In the optimization formulation *L* and *F* are the decision variables. The SALSO method links the simulation model to the SA optimization algorithm. SALSO approach incorporates the groundwater flow and pollutant transport simulation models as set of binding constraints in the optimization formulation which is solved iteratively.

In this optimization problem, maximum number of observation *mob* is considered as 25. Total number of observation time steps *k*, is considered as 4.

Description of Kriging

Kriging is popular interpolation technique, which is applied in the developed methodology to identify the groundwater pollution source locations to reduce the computational burden in identifying the LRF. Previously self-organizing map related surrogate models were used to identify the unknown characteristics of pollutant source characteristics [14]. In geo-statistics, kriging or Gaussian regression is originally an interpolation procedure for which the interpolated values are determined by a Gaussian mechanism controlled by prior covariance [15]. Application of Geo-statistics is very common in the field of pollution [16-17]. Kriging as a tool is often used to determine the quality of groundwater with the help of intensity profile. The approach is commonly used in the spatial analysis context and in machine experiments.

Generally, L values of the pollutant source depend on the size of the study area chosen. If the study area is large, the variable L increases the computational burden as more number of source combination scenarios need to be evaluated to find the optimal result. In order to avoid the excessive load in computing, Ordinary Kriging (OK) is incorporated as a preprocessor. This preprocessor tool is used to reduce the load on optimization. The OK equation is stated in Equation 6.

$$G(L_0) = F(L_0) + \mu \qquad (6)$$

$G(L_0)$ is the concentration obtained by OK; $F(L_0)$ is the unknown concentration value interpreted by combined neighborhood function at the location L_0; and the μ is the standard normalized error occurred at same location Equation 6.

Kriging Based Zoning of KSALSO Method

The proposed methodology (KSALSO) uses two step processes. In the 1st step, pollutant concentration data from monitoring wells are collected. From these concentration data, a kriging interpolation is carried out to generate concentration profile over the entire study area, which is further used as input for the source identification model. In this 2nd step, higher concentration region is designated as potential source location zones (Equation 7) for LSO model. If,

$$G^{max}(L_{max}) \gg G^{min}(L_0); \qquad (7)$$

$G^{max}(L_{max})$ is the maximum concentration obtained after creating the variogram at location L_{max}, L_0 is reduced potential grid after Kriging based zoning; L_0 is not considered as a potential source location in LSO. After OK the L^T_{max} variable is deduced. L^T_{max} is maximum length of study area of KSALSO method;

$$L^T_{max} \le L_{max}; \qquad (8)$$

Equation 4 is replaced by Equation 8 to forms Equation 9.

$$L_{min} \le L \le L^T_{max}; \qquad (9)$$

These variable L^T_{max} become the final constraint of the optimization. Thus, the search space is reduced from entire study area to smaller region for actual source locations in LSO. The schematic chart representing the flow of developed KSALSO is given in Fig. 1.

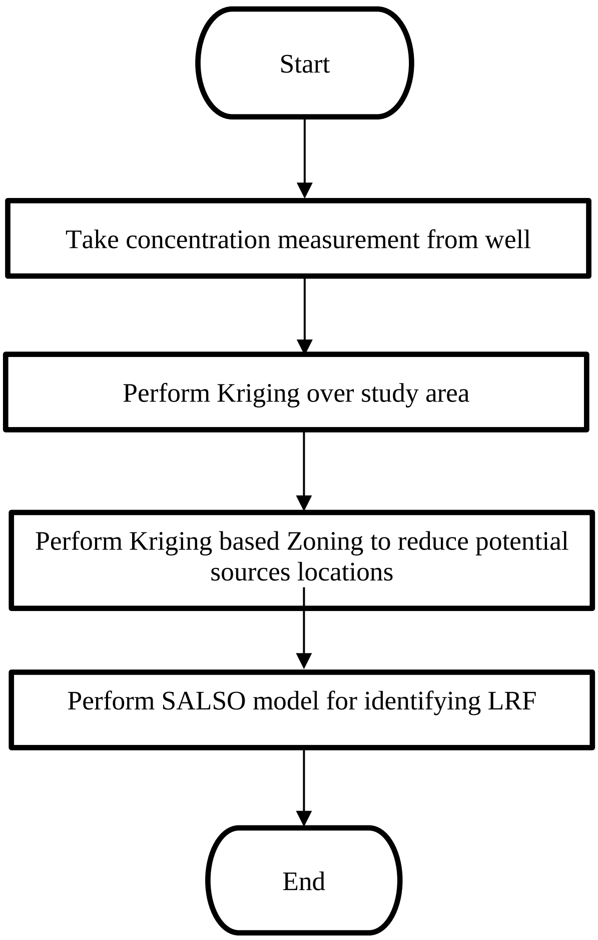

Fig 1: Flowchart of KSALSO

PERFORMANCE EVALUATION

Table 1: Hydro-geological parameters of study area

Name of the parameter	Unit	Value
Length of the area	m	5410
Breadth of the area	m	3628
Thickness of the area	m	30.6
Length of the grid	m	50
Breadth of the grid	m	50
Thickness of the grid	m	30.5
Hydraulic conductivity	m/s	21.34
Horizontal anisotropy	Unit less	0.8
Size of stress period	Days	365
Porosity	Unit less	0.3
Source grid	-	(16,7), (20,6)
Well Number	-	16

In order to test the performance of the developed methodology, site of Indian Institute of Technology Patna is chosen. The location of the site is 25.5357° N, 84.8512° E (Fig. 2). It is assumed that there are two pollutant sources and there are sixteen points where the concentration data is recorded. The hydro-geological parameters of the study area are given in table 1. It is assumed that all these sixteen observation locations are impacted by pollution coming from two point sources. It is assumed that pollutant is conservative, and the sources are active for seven stress periods. Each stress period is of one year. Source fluxes remain constant throughout each stress periods.

A three-dimensional unsteady transient pollutant transport model was developed to estimate the fate and transport behavior of polluted aquifer. It was assumed that the initial concentration is zero. The simulation is performed with the implicit assumption that the starting time of these sources are known. In the LSO model SA is used as optimization algorithm to estimate the source flux release history and source locations using pollutant concentration measurements from few monitoring locations. The fundamental idea of SA comes from thermodynamics. Groundwater pollution source identification problems are categorized as multivariate optimization problem. The non-linearity phenomenon makes the solution more complex. The formulation of such complex problem has multiple local optima. SA works efficiently in getting out from this trap of local optima. Using hill-climbing moves to avoid local optima makes SA is effective in solving problems of non-convex optimization. The simplicity of application to the complex problem and convergence to an optimum global solution extends the competency to solve the ill-posed inverse problems, as in the case with the problem.

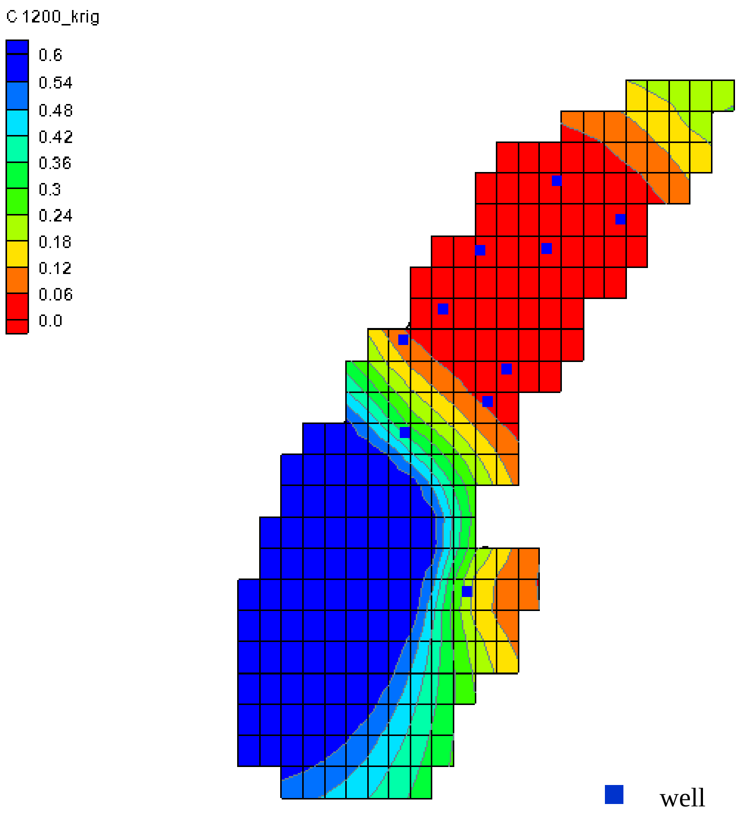

Fig 2: Plan view and interpolated concentration in the study area

From Figure. 2, the likelihood of sources to be present are in the higher concentration zone which is denoted by deep blue color. By kriging, the potential source locations are reduced from 108 active grids to 36 active grids (refer: table 2). Thereafter, SALSO model is applied to identify the sources characteristics in the limited search area. Basically, Kriging based zoning method helps SALSO method to perform in reduced search area for location of pollutant sources, which invariably decreases the plausible number of source characters' combinations to be evaluated in the SA optimization algorithm in search for the optimal LRF.

Table 2**:** Comparison of optimization parameters between SALSO and KSALSO

Method	Number of decision variables	*L*	Source location combinations for search
SALSO	18	108	5778
KSALSO	18	36	630

In this study area, the maximum number of sources are two. Total number of stress period for each sources are seven. So, the number of variables to be estimated is eighteen (2 × 2 + 7 × 2). The description of and *L* variable is given in table 2.

RESULTS AND DISCUSSIONS

The identified source characteristics result is given in table 3 and Fig. 3. Basically, the location and release flux histories are identified with the KSALSO method within minimal computational time. The result of conventional SALSO method [4-5], which was used by several researchers is also presented in identifying location and release flux histories. In case of KSALSO, out of two sources, location of one source matches exactly with the actual location of the source, and the location of other sources is slightly shifted (one grid shift) from the actual source location of the second source. The location is expressed in terms of grids.

The release flux histories are also presented in the Fig. 3, where X dimension represents the stress period and Y dimension represent the release history of each sources. In the X dimension, there are seven stress period plotted for each sources. Blue bar chart represents the actual release histories of each sources and red and green bar chart represent the identified release flux by KSALSO and SALSO methods respectively. It is observed from Fig. 3 that KSALSO method give better recovery of release flux of the identified sources compared to SALSO.

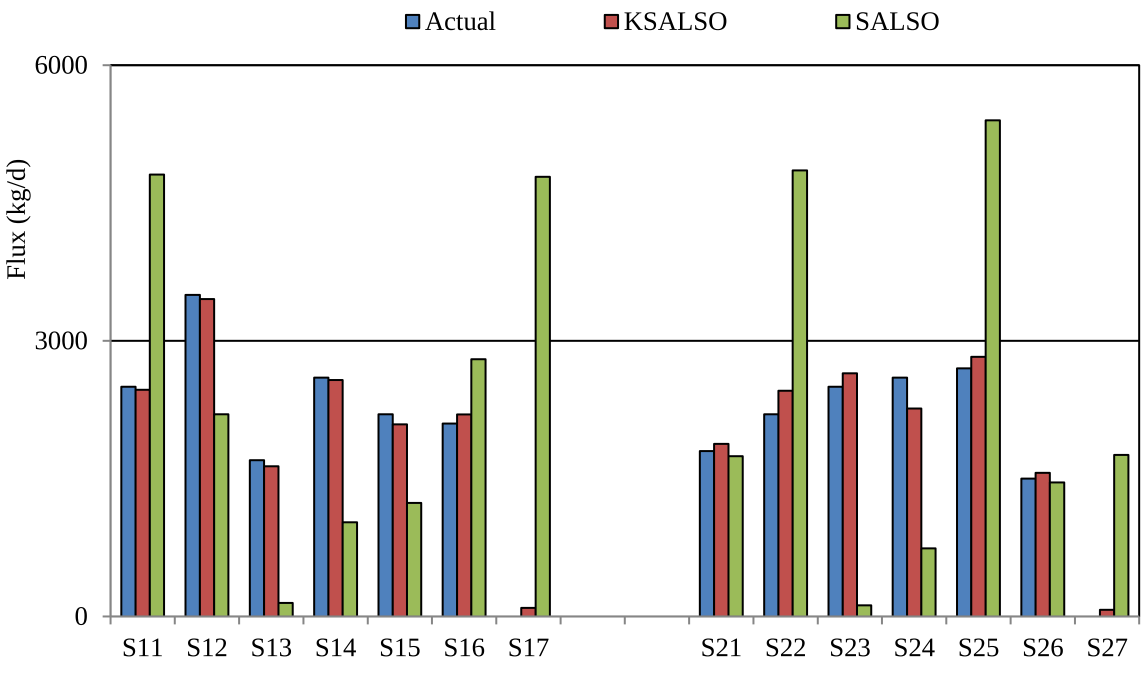

Fig 3: Release flux histories of identified sources

A comparative study is also presented between developed KSALSO method and conventional SALSO methods. From the table 3, it is obvious that KSALSO give significantly superior result than that of SALSO in terms of identifying location of the pollutant sources. KSALSO method converged with less number of iteration compared to SALSO method. That is why elapsed time for completing KSALSO method is also lower than that of SALSO method.

Table 3: Characteristics of identified sources

Characteristics	Actual	KSALSO	SALSO
Co-ordinates (S1)	(16,7)	(16,7)	(12,8)
Co-ordinates **(S2)**	(20,6)	**(**21,6)	(18,8)
Time taken (sec)	-	31294	591432

Computational time taken by the developed method KSALSO, and the conventional method SALSO are analyzed to estimate the efficiency in using kriging for potential source location zoning. From the estimates it is seen that performance of KSALSO model is superior in efficiency and reduces the computational burden by 97.40%.

By comparing the breakthrough curves, the accuracy of the estimated source fluxes using both KSALSO and SALSO methodology is examined in Figure 4. Estimated source flux values for producing the breakthrough curves are used in a forward simulation for both KSALSO and SALSO method.

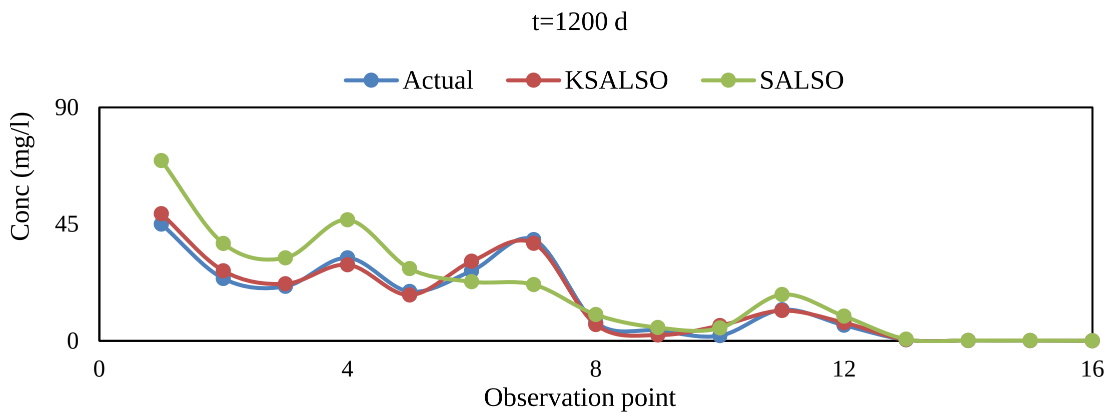

Fig 4: Breakthrough curve comparison

The concentration versus observation location is generated at 1200 days after pollution started. In Fig 4, the blue line represents the actual concentration at different monitoring point at 1200 days of pollution being started. Similarly, the red line and green line demonstrate the estimated concentration at same monitoring location and same time using KSALSO and SALSO method respectively. All estimated concentration values for both KSALSO and SALSO methods are compared with the actual concentration at same time and it is found that KSALSO gives greater accuracy in recovery of breakthrough curve than that of SALSO method.

It is very evident that the developed KSALSO model outperforms the conventional SALSO technique for identifying the location and release flux histories. Moreover, groundwater pollution source characterization is very time taken procedure, when SALSO techniques are employed to solve the problem. This newly developed KSALSO technique is very efficient to reduce the computational burden and more accurately identify the pollutant source characteristics than the previous methods.

CONCLUSIONS

KSALSO is significantly more efficient than SALSO in identifying the sources characteristics. The location of each sources are identified with 90% precision. The recovery of flux histories are also satisfactory (87% precision). KSALSO method works 97% faster than conventional SALSO method, which is a major advantage of developed KSALSO method. One of the major challenges in groundwater pollution is to identify the locations of pollution sources. Most of the previous literature based on the implicit assumption that the location of groundwater pollution sources is well known or the small potential number of sources location are known. The newly developed KSALSO technique is able to relax the existing gap in the field of groundwater pollution source characterization by taking entire study area as potential source location. The number of sources is not taken as unknown explicit variable in this developed methodology

REFERENCES

[1] Gorelick, S. M., Evans, B., &Remson, I. (1983). Identifying sources of groundwater pollution: An optimization approach. Water Resources Research, 19(3), 779-790.

[2] Mahar, P. S., & Datta, B. (2000). Identification of pollution sources in transient groundwater systems. Water Resources Management, 14(3), 209-227.

[3] Chakraborty, A., & Prakash, O. (2020). Identification of clandestine groundwater pollution sources using heuristics optimization algorithms: a comparison between simulated annealing and particle swarm optimization. Environmental Monitoring and Assessment, 192(12), 1-19.

[4] Prakash, O., & Datta, B. (2014). Multiobjective monitoring network design for efficient identification of unknown groundwater pollution sources incorporating genetic programming–based monitoring. Journal of Hydrologic Engineering, 19(11), 04014025.

[5] Prakash, O., & Datta, B. (2015). Optimal characterization of pollutant sources in contaminated aquifers by integrating sequential-monitoring-network design and source identification: methodology and an application in Australia. Hydrogeology Journal, 23(6), 1089-1107.

[6] Xing, Z.X., Qu, R.Z., Zhao, Y., Fu, Q. , Ji, Y. , Lu, W.X , 2019. Identifying the release history of a groundwater pollution source based on an ensemble surrogate model. J. Hydrol. 572, 501–576.

[7] Zhao, Ying, et al. "Identifying Groundwater Contaminant Sources Based on a KELM Surrogate Model Together with Four Heuristic Optimization Algorithms." Advances in Water Resources (2020): 103540.

[8] Sophia, L., & Bhattacharjya, R. K. (2020). A GA Based Iterative Model for Identification of Unknown Groundwater Pollution Sources Considering Noisy Data. In Nature-Inspired Methods for Metaheuristics Optimization (pp. 303-321). Springer, Cham.

[9] Ayvaz, M. T. (2010). A linked simulation–optimization model for solving the unknown groundwater pollution source identification problems. Journal of Contaminant Hydrology, 117(1-4), 46-59.

[10] Prakash, O., & Datta, B. (2013). Sequential optimal monitoring network design and iterative spatial estimation of pollutant concentration for identification of unknown groundwater pollution source locations. Environmental monitoring and assessment, 185(7), 5611-5626.

[11] Datta, B., Petit, C., Palliser, M., K Esfahani, H., & Prakash, O. (2017). Linking a simulated annealing based optimization model with PHT3D simulation model for chemically reactuve transport processes to optimally characterize unknown contaminant sources in a former mine site in Australia. Journal of Water Resource and Protection, 9, 432-454.

[12] Chakraborty, A., & Prakash, O. Identification of clandestine groundwater pollution source locations and their release flux history. In IOP Conference Series: Earth and Environmental Science (Vol. 626, No. 1, p. 012003). IOP Publishing.

[13] Zheng, C. (1992). MT3D: A modular three-dimensional transport model for simulation of advection, dispersion and chemical reactions of contaminants in groundwater systems. SS Papadopoulos & Associates.

[14] Hazrati Y, S., & Datta, B. (2017). Self-organizing map based surrogate models for contaminant source identification under parameter uncertainty. International Journal of GEOMATE, 13, 11-18.

[15] Hazrati Yadkoori, S., & Datta, B. (2017). Evaluation of unknown groundwater contaminant sources characterization efficiency under hydrogeologic uncertainty in an experimental aquifer site by utilizing surrogate models. Journal of Water Resource and Protection, 9, 1612-1633.

[16] Prakash, O., & Datta, B. (2014). Optimal monitoring network design for efficient identification of unknown groundwater pollution sources. International Journal of GEOMATE, 6(1 SERL 11), 785-791.

[17] Cressie, N. (1990). The origins of kriging. Mathematical geology, 22(3), 239-252.