Journal of Hydrology 534 (2016) 352–363

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Multi-objective optimization of long-term groundwater monitoring network design using a probabilistic Pareto genetic algorithm under uncertainty

Qiankun Luo [a], Jianfeng Wu [b,*], Yun Yang [b,c], Jiazhong Qian [a], Jichun Wu [b]

[a] School of Resources and Environmental Engineering, Hefei University of Technology, Hefei 230009, China
[b] Key Laboratory of Surficial Geochemistry, Ministry of Education, Department of Hydrosciences, School of Earth Sciences and Engineering, Nanjing University, Nanjing 210023, China
[c] Huai River Water Resources Commission, Bengbu 233001, China

ARTICLE INFO

Article history:
Received 26 October 2015
Received in revised form 3 January 2016
Accepted 6 January 2016
Available online 12 January 2016
This manuscript was handled by Geoff Syme, Editor-in-Chief

Keywords:
Groundwater management
Multi-objective optimization
Monitoring network design
Probabilistic Pareto genetic algorithm
Monte Carlo analysis

SUMMARY

Optimal design of long term groundwater monitoring (LTGM) network often involves conflicting objectives and substantial uncertainty arising from insufficient hydraulic conductivity (K) data. This study develops a new multi-objective simulation–optimization model involving four objectives: minimizations of (i) the total sampling costs for monitoring contaminant plume, (ii) mass estimation error, (iii) the first moment estimation error, and (iv) the second moment estimation error of the contaminant plume, for LTGM network design problems. Then a new probabilistic Pareto genetic algorithm (PPGA) coupled with the commonly used flow and transport codes, MODFLOW and MT3DMS, is developed to search for the Pareto-optimal solutions to the multi-objective LTGM problems under uncertainty of the K-fields. The PPGA integrates the niched Pareto genetic algorithm with probabilistic Pareto sorting scheme to deal with the uncertainty of objectives caused by the uncertain K-field. Also, the elitist selection strategy, the operation library and the Pareto solution set filter are conducted to improve the diversity and reliability of Pareto-optimal solutions by the PPGA. Furthermore, the sampling strategy of noisy genetic algorithm is adopted to cope with the uncertainty of the K-fields and improve the computational efficiency of the PPGA. In particular, Monte Carlo (MC) analysis is employed to evaluate the effectiveness of the proposed methodology in finding Pareto-optimal sampling network designs of LTGM systems through a two-dimensional hypothetical example and a three-dimensional field application in Indiana (USA). Comprehensive analysis demonstrates that the proposed PPGA can find Pareto optimal solutions with low variability and high reliability and is a promising tool for optimizing multi-objective LTGM network designs under uncertainty.

1. Introduction

The most important objective for long-term groundwater monitoring (LTGM) applications is to ensure that the multiple management/remediation objectives are being achieved while minimizing the monitoring cost and the risks to human health and environment (Chien et al., 2004). However, the LTGM requires substantial capital expenditure. The report of the Air Force Center for Engineering and Environment (AFCEE) highlighted that the cost for long term monitoring was about 35% of the capital cost for contaminant remediation system (Phil Hunter, 2011). Over-sampling is a common problem encountered in performing LTGM that will add unnecessary expenditure (Chien et al., 2004; Wu et al., 2005; Reed et al., 2013). Therefore, the most urgent task for the LTGM is to provide the best possible map of contaminant concentrations at a particular instant while minimizing the monitoring cost under the framework of simulation–optimization models. On the other hand, for a real aquifer simulation model, uncertainty is always inevitable due to the limited knowledge of the hydrogeological setting and parameter distributions, in which hydraulic conductivity (K) is commonly regarded as one of the most important contributors to uncertainty in model outputs (Wu et al., 2006). To a great extent, the spatial variation of K dominates the transport and fate of contaminants in groundwater systems (Wu et al., 2006; Singh and Minsker, 2008). Thus, addressing uncertainty to an aquifer simulation model must be considered in implementing the optimal design of LTGM problems.

Optimization of the LTGM network design has attracted increasing attention to achieve the cost-effective LTGM and

* Corresponding author.
E-mail addresses: jfwu@nju.edu.cn, jfwu.nju@gmail.com (J. Wu).

http://dx.doi.org/10.1016/j.jhydrol.2016.01.009
0022-1694/

overcome the complexity of characterizing groundwater contaminant plume over long time periods. For example, Reed et al. (2000) presented a simulation–optimization methodology for cost-effective monitoring network design which is used to examine the feasibility of reducing long-term monitoring costs. On the basis of the work of Reed et al. (2000), Wu et al. (2005) developed a methodology for cost-effective sampling network design associated with LTGM problems at large-scale field sites. The difference between Reed et al. (2000) and Wu et al. (2005) is that the first and second moments of a three-dimensional contaminant plume were introduced by the latter as new constraints in the optimization formulation. With the new moment constraints, the plume interpolated from the sampled data can sufficiently characterize the spatial features of the plume output by the flow and transport model. Furthermore, Wu et al. (2006) developed a noisy genetic algorithm for cost-effective sampling network design in the presence of uncertainty in the *K*-field. Compared with Monte Carlo simple genetic algorithm, the noisy genetic algorithm can maintain acceptable accuracy in global mass and higher moment estimations while achieving a notable cost savings. More recently, Dhar and Datta (2010) developed a logic-based mixed-integer linear optimization model for redundancy reduction in groundwater monitoring network design. Chadalavada et al. (2011) proposed an uncertainty-based optimal monitoring network design for a chlorinated hydrocarbon contaminated site in which the uncertainty in the study area was quantified by using the concentration estimation variance at all of the potential monitoring locations. Further, Khader and Mckee (2014) presented a relevance vector machine for groundwater monitoring network design under uncertainties of aquifer properties, population transport processes, and climate.

However, most of the studies above dealt with the groundwater monitoring design as a single objective optimization problem. In reality, decision makers often need to simultaneously consider several competing objectives such as minimizing monitoring cost, maximizing monitoring accuracy, and minimizing health risks. These multiple competing objectives will lead to a series of compromised solutions, known as non-dominated solutions (Pareto optimal solutions), *i.e.*, solutions that one objective cannot be improved without degrading the performance in another objective (Deb, 2001; Deb et al., 2002). Thus, the LTGM network design problem should be formulated as a multi-objective optimization model that leads to a set of Pareto optimal solutions.

In recent years many multi-objective evolutionary algorithms have been successfully employed to solve groundwater monitoring design problems to avoid running multiple times with different weighted or constrained levels to identify the entire Pareto optimal solutions (trade-off curve) for the multi-objective optimization problem by single optimization method. For instance, Reed and Minsker (2004) combined quantile kriging and the non-dominated sorted genetic algorithm-II (NSGAII) to solve LTGM problem with four objectives. The optimization result indicated that high-order Pareto optimization has a significant potential as a tool that can be used in the LTGM problems. Reed et al. (2007) added ε-dominance archiving and automatic parameterization techniques to the NSGA-II (ε-NSGAII) to significantly reduce computational demands for a four-objective groundwater monitoring design problem. Kollat et al. (2008) presented a new multi-objective evolutionary algorithm, the epsilon-dominance hierarchical Bayesian optimization algorithm (ε-hBOA), which represents a new class of probabilistic model building evolutionary algorithms to solve LTGM problems. More recently, Gopalakrishnan et al. (2011) used the primary data source (concentrations in soil samples) and secondary data source (concentrations in plant branch tissue) to design monitoring network for Phytoremediation system at Argonne National Laboratory. They showed that use of the composite of both the primary and secondary data can result in effective monitoring designs, even at sites where the data with significant error can be used for the transformation model. Reed and Kollat (2013) developed a framework for simultaneously assessing solution quality and scalability for massively parallel multi-objective evolutionary algorithm-based search, in which visual analytics were used to evaluate how changes in search metric performance relate to actual decision relevant changes in the Pareto approximate set.

However, most of the previous works in the field of multi-objective optimization of LTGM problems were taken under deterministic hydrogeological conditions. In this study, a new multi-objective simulation–optimization model is developed using flow and transport codes, MODFLOW (Harbaugh and McDonald, 1996) and MT3DMS (Zheng and Wang, 1999), for optimal design of LTGM under considering uncertainty of both the simulation model and the optimization model. Furthermore, a new probabilistic Pareto genetic algorithm (PPGA) is developed to search for Pareto optimal solutions with low variability and high reliability to the multi-objective simulation–optimization models. We also aim to demonstrate the applicability and efficiency of PPGA in the handling of the *K*-field uncertainty through a two-dimensional hypothetical problem and a three-dimensional field application at Granger in Indiana (USA). It is noteworthy that the methodology developed in this study is specifically designed for sites with over-sampling locations in the pre-existing LTGM networks during the aquifer remediation. The goal of this study is to reduce over-sampling costs by eliminating data redundancy at sites with numerous monitoring/sampling locations. Thus, it is not intended for contaminant plume detection and is assumed that the contaminant plume at initial time can be geostatistically interpolated using historical data and/or samples taken from all available monitoring well locations being considered (Reed et al., 2000; Wu et al., 2005, 2006).

2. Methodology

In this study, the new evolutionary multi-objective simulation–optimization model for LTGM network design includes three parts: a flow and transport simulation model, a geostatistics-based estimator for global contaminant plume estimation (global mass estimation and spatial moment estimation) under different potential monitoring strategies, and a new multi-objective evolutionary algorithm, the PPGA based optimization model for multi-objective optimal design of LTGM network under uncertainty of *K*-field distribution.

2.1. Flow and transport simulation model

In this study, the three-dimensional finite-difference groundwater flow code, MODFLOW (Harbaugh and McDonald, 1996), and its solute transport companion, MT3DMS (Zheng and Wang, 1999), are used as the flow and transport simulation model to characterize the future contaminant plume to be monitored. For a particular optimal design of LTGM application, the main program of the classical version of MODFLOW and MT3DMS were modified into modular subroutines so that they can be called by the optimization program repeatedly. Note that the contaminant plume simulated by flow and transport models under the true *K*-field is assumed to represent the true distribution of contaminant to be monitored.

2.2. Global contaminant plume estimation

The global contaminant plume estimation includes three spatial moments of the plume: the zeroth moment, the first moment and