



Conservative strategy-based ensemble surrogate model for optimal groundwater remediation design at DNAPLs-contaminated sites



Qi Ouyang^{a,b}, Wenxi Lu^{a,b,*}, Jin Lin^c, Wenbing Deng^d, Weiguo Cheng^e

^a Key Laboratory of Groundwater Resources and Environment, Ministry of Education, Jilin University, Changchun 130021, PR China

^b College of Environment and Resources, Jilin University, Changchun 130021, PR China

^c State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing 210029, PR China

^d Cores and Samples Center of Land and Resources, China Geological Survey, Yanjiao 065201, PR China

^e Shenyang Academy of Environmental Sciences, Shenyang 110000, PR China

ARTICLE INFO

Keywords:

Conservative strategy
Groundwater remediation
Optimization
Surrogate
Uncertainty

ABSTRACT

The surrogate-based simulation-optimization techniques are frequently used for optimal groundwater remediation design. When this technique is used, surrogate errors caused by surrogate-modeling uncertainty may lead to generation of infeasible designs. In this paper, a conservative strategy that pushes the optimal design into the feasible region was used to address surrogate-modeling uncertainty. In addition, chance-constrained programming (CCP) was adopted to compare with the conservative strategy in addressing this uncertainty. Three methods, multi-gene genetic programming (MGGP), Kriging (KRG) and support vector regression (SVR), were used to construct surrogate models for a time-consuming multi-phase flow model. To improve the performance of the surrogate model, ensemble surrogates were constructed based on combinations of different stand-alone surrogate models. The results show that: (1) the surrogate-modeling uncertainty was successfully addressed by the conservative strategy, which means that this method is promising for addressing surrogate-modeling uncertainty. (2) The ensemble surrogate model that combines MGGP with KRG showed the most favorable performance, which indicates that this ensemble surrogate can utilize both stand-alone surrogate models to improve the performance of the surrogate model.

1. Introduction

Dense nonaqueous-phase liquids (DNAPLs) are now frequently detected in groundwater throughout the world, because of the widespread use, improper disposal, accidental spills and leaks of petrochemical products (Kueper and Mcworter, 1991). Because of their low solubility, low mobility and high density in water, DNAPLs may remain in aquifers for long periods of time, and thus ultimately become long-term continuous sources of groundwater contamination (Qin et al., 2007). Surfactant-enhanced aquifer remediation (SEAR), an enhancement to the conventional pump-and-treat technique, is a promising way to remove DNAPLs from aquifers (Schaerlaekens et al., 2005). By adding surfactants to the water, the solubility and mobility of DNAPLs in an aquifer can be increased (Delshad et al., 1996), which makes SEAR more efficient than the conventional pump-and-treat technique. Because the cost of the SEAR process is high, optimizing design for cost-effectiveness is of great value.

Simulation-optimization (S/O) techniques have been used extensively to solve such problems (Jiang et al., 2015; Luo et al., 2013).

When such techniques are employed, the numerical simulation model would be called thousands of times before the optimal design is obtained, which is computationally expensive, and may be prohibitive (Hou et al., 2015). Using surrogates (also known as meta-models or proxy models) to replace the computationally expensive simulation models has become commonplace.

However, no matter how well the surrogate model approximates the simulation model, errors caused by surrogate-modeling uncertainty exist. In constrained optimization (constraints being surrogate models), the obtained solution may be infeasible because of surrogate errors (Viana et al., 2010). He et al. (2010) regarded the error between simulation model and surrogate model as a stochastic variable and adopted the chance-constrained programming (CCP) method to incorporate it into the optimization model for groundwater remediation design. However, before this method is adopted, the hypotheses of normality and zero-means for the errors generated by surrogate models should be tested, which cannot always be achieved.

Recently, researchers have focused on conservative strategy-based surrogate models (also called conservative surrogates), which push the

* Corresponding author at: Key Laboratory of Groundwater Resources and Environment, Ministry of Education, Jilin University, Changchun 130021, PR China.
E-mail address: luwenxi@jlu.edu.cn (W. Lu).

optimal solution into the feasible region (Pan et al., 2012). In many engineering problems, there is an incentive to obtain approximations that are as close as possible but on the safer side in terms of the actual response (Picheny et al., 2008). In groundwater remediation design optimization, such a response may represent the minimum allowable value of the contaminant removal rate that, to avoid failure, must not be overestimated. In this paper, we call a surrogate model conservative when the estimations are lower than the true responses. That is to say, conservative surrogates tend to underestimate target values. In contrast, general surrogate models are unbiased; that is, the estimations are equally likely to be lower and higher than the actual value (Pan et al., 2012). To date, conservative surrogate models have been used in structural analysis of vehicle engineering (Zhu et al., 2013) and aircraft engineering (Acar et al., 2007), but have not previously been applied in optimization of groundwater remediation design.

Many techniques for surrogate modeling have been proposed, such as artificial neural networks (ANNs) (Luo et al., 2013), Kriging (KRG) (Zhao et al., 2016), support vector regression (SVR) (Ouyang et al., 2017), and extreme learning machines (ELMs) (Jiang et al., 2015). More recently, multi-gene genetic programming (MGGP) (Hinchliffe et al., 1996; Searson et al., 2007) has been designed to develop the input–output relationship of a system, and has attracted the attention of many researchers across a broad range of fields (Pan et al., 2013; Pandey et al., 2015; Mohammadzadeh et al., 2016). The main advantage of MGGP is its ability to develop a compact and explicit prediction equation in terms of different model variables without assuming a prior form of the existing relationships (Muduli and Das, 2015). A previous study (Ouyang et al., 2017) demonstrated the superiority of MGGP over KRG and SVR. Recently, researchers have tended to combine multiple surrogate models in ensembles instead of selecting only the best model and discarding the rest (Acar and Rais-Rohani, 2008; Goel et al., 2006; Viana et al., 2009). However, the combination of MGGP surrogate modeling with other techniques has not previously been evaluated.

The aim of the present study is to determine an optimal groundwater remediation design for DNAPL-contaminated sites with minimum costs under certain constraints. To address the abovementioned concerns, this study 1) combines an MGGP surrogate model with other surrogate models to form ensembles and make comparisons between them, and 2) adopts a conservative strategy to address surrogate-modeling uncertainty in case of failure. The CCP method is used for comparison.

2. Methodology

2.1. MGGP

MGGP is a robust variant of genetic programming (GP) and is designed to generate empirical mathematical models of the input–output relationship from the datasets. GP is based on the evaluation of a single gene, whereas MGGP is constructed from a number of genes (Gandomi and Alavi, 2012; Searson et al., 2007). Each gene evolved by MGGP is a structured tree composed of functions and terminals (Searson et al., 2007), as can be seen in Fig. 1. The function set can include elements such as arithmetic operators (i.e., plus, minus, divide, multiply), non-linear functions (e.g., sin, cos, tan, exp., tanh, and log) or Boolean operators (Garg et al., 2014). The terminal set consists of constants and variables of the model (Garg et al., 2014).

2.2. Ensemble of surrogates

Based on previous studies by Bishop (1995) and Acar and Rais-Rohani (2008), Viana et al. (2009) proposed a new method to construct an ensemble of surrogates (also called a weighted average surrogate, WAS), named OWS_{diag} (optimal weighted surrogate), and demonstrated that this method is able to stabilize the addition of poorer surrogates up

to the point of using all created surrogates. Therefore, this method is adopted in the present study to construct an ensemble of surrogates for simulation of the SEAR process. This method is based on minimizing the mean square error (MSE):

$$\text{MSE}_{\text{WAS}} = \frac{1}{V} \int_V e_{\text{WAS}}^2(\mathbf{x}) d\mathbf{x} = \mathbf{w}^T \mathbf{C} \mathbf{w} \quad (1)$$

where \mathbf{w} is the weight vector of the weighted surrogate, $e_{\text{WAS}}(\mathbf{x}) = y(\mathbf{x}) - \hat{y}_{\text{WAS}}(\mathbf{x})$ is the error associated with the prediction of the WAS model, and \mathbf{C} is the covariance matrix, the elements of which can be calculated as

$$c_{ij} \cong \frac{1}{p} \tilde{\mathbf{e}}_i^T \tilde{\mathbf{e}}_j \quad (2)$$

where p is the number of training data points, $\tilde{\mathbf{e}}$ is the vectors of leave-one-out cross-validation errors, and i and j indicate different surrogate models.

Given the \mathbf{C} matrix, the optimal weighted surrogate (OWS) is obtained from minimization of the MSE as:

$$\min \text{MSE}_{\text{WAS}} = \mathbf{w}^T \mathbf{C} \mathbf{w} \quad (3a)$$

subject to

$$\mathbf{1}^T \mathbf{w} = 1 \quad (3b)$$

The solution is obtained using Lagrange multipliers:

$$\mathbf{w} = \frac{\mathbf{C}^{-1} \mathbf{1}}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}} \quad (4)$$

In this equation, only the diagonal elements of matrix \mathbf{C} are used to avoid negative weights, or weights larger than one. Therefore, this method is called OWS_{diag}.

2.3. Chance-constrained programming (CCP)

CCP was proposed by Charnes and Cooper (1959) as a means of handling uncertainty by specifying a confidence level at which the stochastic constraints are desired to be held. A typical CCP model can be written as follows:

$$\min f(\mathbf{x}) \quad (5a)$$

subject to

$$\Pr[g(\mathbf{x}, \delta) \geq b] \geq \alpha \quad (5b)$$

$$\mathbf{x} \in \mathbb{R}^n \quad (5c)$$

where \mathbf{x} is the decision variable vector; $f(\mathbf{x})$ is the objective function; $g(\mathbf{x}, \delta)$ is the constrained function, which contains the stochastic variable δ ; $\Pr[g(\mathbf{x}, \delta) \geq b]$ indicates the probability that $g(\mathbf{x}, \delta)$ is no less than the predefined value b ; and $\alpha \in [0, 1]$ is the confidence level, which means that the constraint is satisfied with a probability of at least α . If the probability density function (pdf) of δ is known, then Eq. (5b) can be substituted with its deterministic equivalent (He et al., 2010; Tiedeman and Gorelick, 1993):

$$E[g(\mathbf{x}, \delta)] - \Phi^{-1}(\alpha) \cdot \sigma \geq b \quad (6)$$

where $E[\cdot]$ is expected value; $\Phi^{-1}(\alpha)$ is value of the standard normal cumulative distribution corresponding to the confidence level of α ; and σ is the standard deviation of δ .

2.4. Conservative strategy-based surrogate

For a given set of sample points, a general unbiased surrogate, $\hat{g}(\mathbf{x})$, can be constructed to approximate the actual value using surrogate modeling techniques. A conservative surrogate can be obtained by multiplying or adding a constant to the unbiased surrogate:

$$\hat{g}_c(\mathbf{x}) = \hat{g}(\mathbf{x}) + s_m \quad (7)$$

Download English Version:

<https://daneshyari.com/en/article/5765850>

Download Persian Version:

<https://daneshyari.com/article/5765850>

[Daneshyari.com](https://daneshyari.com)