Environ. Sci. Technol. **2008,** *42,* 9237–9243

Health-Risk-Based Groundwater Remediation System Optimization through Clusterwise Linear Regression

L. HE,[†,‡] G. H. HUANG,[*,†,‡] AND H. W. LU[†,§]

Environmental Systems Engineering Program, Faculty of Engineering, University of Regina, Regina, Saskatchewan, Canada S4S 0A2

Received March 26, 2008. Revised manuscript received August 19, 2008. Accepted August 26, 2008.

This study develops a health-risk-based groundwater management (HRGM) model. The model incorporates the considerations of environmental quality and human health risks into a general framework. To solve the model, a proxy-based optimization approach is proposed, where a semiparametric statistical method (i.e., clusterwise linear regression) is used to create a set of rapid-response and easy-to-use proxy modules for capturing the relations between remediation policies and the resulting human health risks. Through replacing the simulation and health risk assessment modules with the proxy ones, many orders of magnitude of computational cost can be saved. The model solutions reveal that (i) a long remediation period corresponds to a low total pumping rate, (ii) a stringent risk standard implies a high total pumping rate, and (iii) the human health risk associated with benzene would be significantly reduced if it is regarded as constraints of the model. These implications would assist decision makers in understanding the effects of remediation duration and human-health risk level on optimal remediation policies and in designing a robust groundwater remediation system. Results from postoptimization simulation show that the carcinogenic risk would decrease to satisfy the regulated risk standard under the given remediation policies.

Introduction

Groundwater can be contaminated by petroleum hydrocarbons discharged into the subsurface due to leakage of underground storage tanks and pipelines. The contamination may pose a significant threat to human and ecological health since the petroleum hydrocarbons migrate through the subsurface environment as nonaqueous phase liquids (NAPLs). A large number of remediation techniques have been developed to clean up contaminated sites. One problem associated with these remediation actions is the deficiency in understanding processes controlling the fate of contaminants, probably leading to a large inflation of expenses (*1*).

Various simulation-based groundwater management models were therefore developed in order to improve remediation efficiency (*2*−*22*). Recently, risk-based decision analysis has been increasingly introduced into the models to ensure that the risks to human health and the environment associated with a contaminated site can be reduced to acceptable levels (*23*−*27*). For instance, Smalley (*23*) used a noise genetic algorithm to incorporate parameter uncertainty and variability within a risk-based framework for in situ bioremediation design. Wong and Yeh (*25*) presented a systematic approach for the management of groundwater supply systems based on stochastic health risk assessment. Yan and Minsker (*27*) proposed an adaptive neural network genetic algorithm to support optimal groundwater remediation design considering the effect of carcinogenic health risk for the removal of RDX (hexahydro-1,3,5-trinitro-1,3,5-triazine) and TNT (2,4,6-trinitrotoluene) in groundwater.

Many variables in groundwater management systems can be either continuous or discrete, and relations among them can be either linear or nonlinear (*28, 29*). Nevertheless, conventional methods such as regression and artificial neural networks can hardly reflect such complicated characteristics and relationships efficiently (*29*). Thus, Huang proposed a stepwise cluster analysis (SCA) method for dealing with discrete and nonlinear relationships in an air quality forecasting system (*28*). Huang et al. presented an integrated simulation, statistical, and nonlinear optimization system for supporting real-time dynamic modeling and process control of bioremediation systems at petroleum-contaminated sites (*29*); in the system, SCA was introduced to generate a number of approximated statistical models instead of the simulation models. He et al. also advanced a nonlinear stochastic programming model for optimizing surfactant-enhanced aquifer remediation processes under parameter uncertainty (*30*); the model was solved based on the proxy modules created by SCA.

However, SCA can only account for the differences between rather than within clusters, potentially leading to increased prediction errors. It is thus desirable to improve the method to mitigate the effects of prediction errors on decision results. This study aims to present a new semiparametric statistical method (clusterwise linear regression, CLR), through which a set of proxy modules can be created to capture the relations between remediation policies (e.g., pumping rate) and remediation performance (e.g., human-health risk level). To accelerate the optimization process, the modules are incorporated into a health-risk-based groundwater management (HRGM) model. A petroleum-contaminated site located in Western Canada is subsequently applied to illustrate the model's performance in identifying optimal groundwater remediation policies.

Materials and Methods

Study Site. The study site is located in west central Saskatchewan, north of Kindersley, approximately 350 km northwest of Regina, Canada (Figure 1). In the past, the site has acted as a gathering system, treatment center, and compressor station for natural gas. Due to leakage, drainage, and blow-down fluids from the underground facilities, significant hydrocarbon impacts have been presented; some of them have adversely affected the environmental quality of the soil and groundwater. The past monitoring program has suggested three principal contaminant sources (Figure S1 of the Supporting Information), including losses from the operation of the gas plant on the west side of the site (S1), a disposal pit formerly located in the northeast quadrant of

* Corresponding author phone: +1-306-585-4095; fax: +1-306-585-4855; e-mail: huang@iseis.org.

† Current address: Faculty of Engineering, University of Regina, Regina, SK, Canada S4S 0A2.

‡ Current address: Center for Studies in Energy and Environment, University of Regina, Regina, SK, S4S 0A2, Canada.

§ Current address: Chinese Research Academy of Environmental Science, North China Electric Power University, Beijing, 100012-102206, China.

10.1021/es800834x CCC: $40.75
Published on Web 11/14/2008

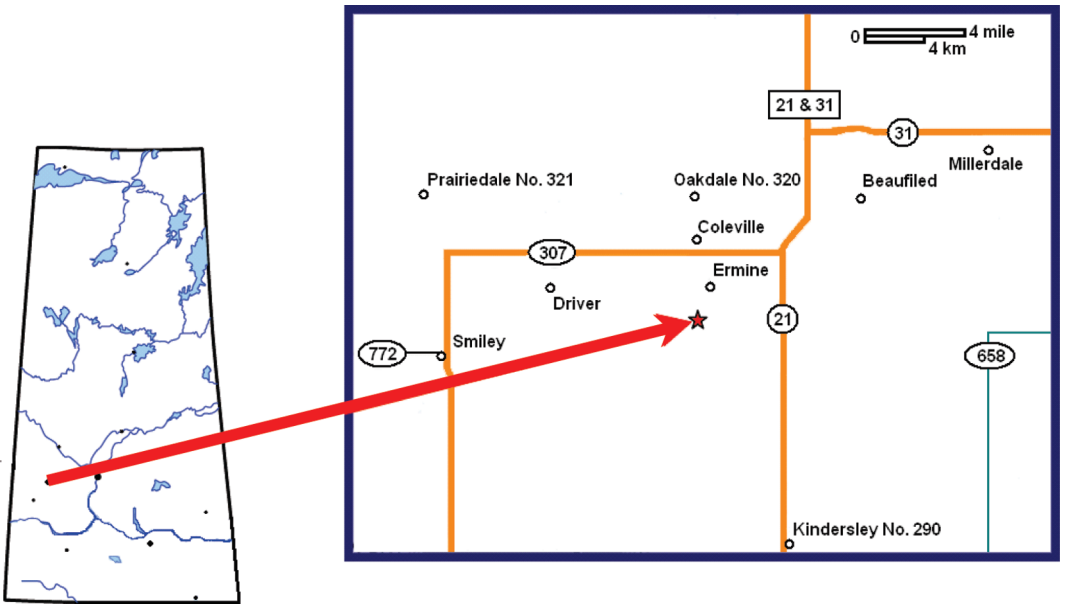

FIGURE 1. Site location map.

the site (S2), and losses from recirculating pump failures which is close to S1 (S3).

To remediate the site, a couple of remediation actions have been initiated since 2005. The remediation process was divided into two stages. In the first stage, dual phase vacuum extraction (DPVE) was commissioned on May 5, 2005 and operated until the pump seized on October 5, 2005. The DPVE system operated a total of approximately 1514 of a possible 3667 h over this duration. In this stage, approximately 3400 L of petroleum hydrocarbons (PHC) were recovered from the subsurface. The total PHC recovery volume was partitioned into the four phases as follows: 99% in the vapor phase, 1% in the biodegradation phase, less than 1% in the dissolved phase, and less than 1% in the liquid phase.

While the majority of PHC was typically removed, water samples from the site indicated that the DPVE system could only reduce total petroleum hydrocarbon concentrations to over 1400 μg/L. This level was much higher than the guidelines issued by the CCME (*32*) and SERM (*33*). Residual phase (trapped in the pore spaces) existing as LNAPL (light nonaqueous phase liquid) and BTEX (benzene, toluene, ethylbenzene, and xylenes) dissolved in groundwater were regarded as long-term and stable contamination sources. Therefore, a pump-and-treat (PT) system was recommended in the second stage for further removal of BTEX dissolved in the groundwater. This study focused on the optimal design of the proposed PT system based on the site characteristics after implementing a period of DPVE practice. Factors determining the efficiency of a PT system mainly can involve concentrations of the injected and extracted contaminants as well as the pumping rates at the remediation wells. Nonetheless, the remediation policy to be optimized only targeted the pumping rates of remediation wells. As shown in Figure S1 of the Supporting Information, 2 injection and 4 extraction wells were installed in or around the contaminant plume, while 8 wells were used to monitor contaminant concentrations.

The site investigation results revealed that nonflow boundary conditions were assigned at the top and bottom of the simulation domain, forming a steady groundwater from northeast to southwest. The hydraulic gradient was estimated to be 0.003. The site was considered as a three-dimensional heterogeneous domain with an area of 270 $\times$ 225 m^2 and a depth of 10 m. Horizontally, each layer was discretized into 54 $\times$ 45 grid blocks, with each one havign dimensions of 5 and 5 m in x and y directions, respectively.

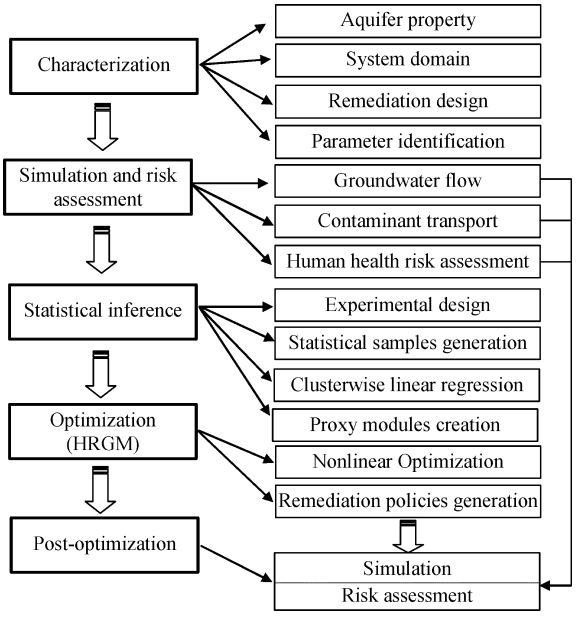

FIGURE 2. Framework of the study method.

Vertically, the simulation domain was discretized into four layers, with each one being 1, 2, 3, and 4 m in z direction, respectively. The total number of grids in the domain is 9720 (54 $\times$ 45 $\times$ 4). According to the previous investigation results, the site possesses complex soil types including clay till, silty clay, and sandy soil (*34*). Distribution of soil types is presented in Figure S2 of the Supporting Information. It was known that layers 2, 3, and 4 are located in the saturated zone, and layer 1 is situated in the unsaturated zone. Since the most serious plume in groundwater was mainly detected in the layer 2, the design effort was merely based on the simulation results for this layer.

Framework of the Study Method. Figure 2 shows the framework of the study method. The first step is site characterization, which attempts to identify aquifer property, system domain, remediation scenarios, and modeling parameters. Simulation (Section S1 of the Supporting Information) and health risk assessment (Section S2 of the Supporting Information) are performed to predict contaminant concentrations and the resulting carcinogenic and noncarci-

nogenic risk levels, respectively. Subsequently, numerical experiments are conducted to select and identify statistical samples comprising explanatory and response variables representing remediation policies and remediation performance, respectively. The statistical samples are obtained through computer-assisted randomly sampling within the ranges of pumping rates. Then, a set of proxy modules is created by CLR to capture the relations between remediation policies and remediation performance. With the obtained modules, health risk levels are estimated in terms of the outputs of the proxy modules, and then the HRGM model is solved through a simulated annealing algorithm (*31, 35*). Finally, postoptimization simulation is also employed to predict contaminant concentrations and evaluate the associated health risk under the identified optimal remediation policies.

Health-Risk-Based Groundwater Management Model. The model is formulated as follows:

Minimize

$$TR = \sum_{i=1}^{I} Q_{i,t}^{In} + \sum_{j=1}^{J} Q_{j,t}^{Ex} \tag{1a}$$

subject to

$$c_{k,t}(Q_{i,t}^{In}, Q_{j,t}^{Ex}) \le c_{\max} \text{ for all } k = 1, 2, ..., K \tag{1b}$$

$$ELCR_{k,t}(Q_{i,t}^{In}, Q_{j,t}^{Ex}) \le ELCR_{\max} \tag{1c}$$

$$HQ_{k,t}(Q_{i,t}^{In}, Q_{j,t}^{Ex}) \le HQ_{\max} \tag{1d}$$

$$0 \le Q_{i,t}^{In} \le Q_{i,\max}^{Ex} \text{ for all } i = 1, 2, ..., I \tag{1e}$$

$$0 \le Q_{j,t}^{Ex} \le Q_{j,\max}^{Ex} \text{ for all } j = 1, 2, ..., J \tag{1f}$$

$$\sum_{i=1}^{I} Q_{i,t}^{In} = \sum_{j=1}^{J} Q_{j,t}^{Ex} \tag{1g}$$

where TR is total pumping rate for all injection/extraction wells; $Q_{i,t}^{In}$ and $Q_{j,t}^{Ex}$ are pumping rates for the ith injection well and the jth extraction well for the t-year pumping period, respectively; $Q_{i,\max}^{In}$ and $Q_{j,\max}^{Ex}$ are maximum pumping rates for the ith injection well and the jth extraction well; $c_{k,t}$ is contaminant concentration of the kth monitoring well after t years of pumping, which is computed through the simulation module and determined by decision variables $Q_{i,t}^{In}$ and $Q_{j,t}^{Ex}$; $c_{\max}$ is the environmental standard denoted as maximum acceptable contaminant concentration; $ELCR_{k,t}$ and $HQ_{k,t}$ are excess lifetime cancer risk and hazard quotient at exposure location k and time t, respectively, and are estimated in terms of the health risk assessment modules which will be presented later; $ELCR_{\max}$ and $HQ_{\max}$ are risk standards denoted as maximum tolerable excess lifetime cancer risk and hazard quotient, respectively; i (= 1, 2, ... I), j (= 1, 2, ... J), k (= 1, 2, ... K) are indexes representing injection well, extraction well, and monitoring well, respectively.

Objective (1a) is represented as the total pumping rate of all remediation wells. Environmental constraint (1b) requires that the contaminant concentrations at the monitoring wells should be less than the environmental standard. Constraints (1c) and (1d) are incorporated to reduce the human-health risk to an allowable level (i.e., risk standard). Technical constraints (1e) and (1f) are provided to limit the injection and extraction pumping rates within a practical operating interval between a lower and an upper bound. The lower bound is set to zero in this study, while the upper bound is determined according to technical alternatives and site characteristics. Additionally, the sum of pumping rates at all injection wells should equal that at all extraction wells (i.e., constraint (1g)), such that the groundwater can flow directly toward the plume interior under a stable hydraulic gradient.

Clusterwise Linear Regression. The essence of this approach is to divide the fitting samples into a set of clusters that have significant differences, based on a given statistical criterion; each cluster is then assigned to a polynomial regression equation (this study used linear regression) representing a type of underlying relationship between variables. CLR is divided into five steps: criterion establishment for clusters splitting and mergence, cluster splitting, cluster mergence, regression analysis, and example forecasting. The first step is to provide a criterion for determining whether the samples can be divided into two clusters and whether the two clusters can be merged into one. The second step is to split the clusters in terms of the identified optimal cutting points. When all clusters are split, clusters mergence is then conducted in the third step, with the purpose of checking whether or not any of the two subclusters can be classified into one under a given F-test criterion. The fourth step is to use regression analysis to capture the relationships for each of the generated tip clusters. The last step is to forecast contaminant concentrations and carcinogenic risk under the given remediation policies. Tip cluster search and response level estimation should be conducted in this step. Section S3 of the Supporting Information details the procedures of the CLR method.

CLR is different from parametric approaches such as polynomial regression analysis which merely tackles isotropic functions (*36*) (i.e., the relation between remediation policies and the resulting remediation performance is fixed). CLR can output a set of anisotropic functions, which means that the functions may vary with remediation policies. It has advantages in (i) the capability to deal with continuous and discrete variables, as well as nonlinear relationships among the variables (*29, 30*); and (ii) the enhanced suitability in being applied to such cases where no knowledge about the specific relationships between explanatory and response variables is available.

CLR is also different from nonparametric approaches such as SCA and decision trees (DT) methods (*37*). By simply using the average response level of all samples classified into the same cluster as the forecasting value, SCA and DT methods cannot reveal implicated relationships within clusters. In contrast, CLR assigns a different regressor (linear or nonlinear) to each of the tip clusters, by which the differences not only between but also within the tip clusters can be simultaneously reflected. In general, CLR has the advantages in (i) avoiding the piecewise nature of the cluster trees (*37*); (ii) providing finer analysis for the difference not only between clusters but also within the clusters; and (iii) having a reasonable result interpretation since any variation of the explanatory variables will cause a change in the response level.

CLR also differs from the DT method in cluster splitting. On the one hand, cluster mergence is not considered by DT, which may generate unnecessary tip clusters that have no significant differences. On the other hand, criteria determining whether the cluster should be cut only rely on nonlinear optimization algorithm, probably lowering the convergence speed or causing solutions to be trapped to local minima. Conversely, CLR uses the exact F-test as the criterion, ensuring that the probability of each classification error is lower than an acceptable level (5% in this study). Artificial neural network (ANN) may also be used for creating proxy modules; however, they are very suitable for the cases with abundant fitting samples. When insufficient data are used, overfitting may occur, i.e., very small errors in the fitting process might result in large errors in the forecasting process (*38*).

Results and Discussion

Modeling calibration and verification, as important steps of this modeling study, were undertaken based on monitoring

data from 2001 to 2003. The site information and contamination conditions in 2001 and 2002 were used for estimating modeling parameters, while those in 2002 and 2003 were used to verify the simulation module. Table S1 of the Supporting Information shows part of the input parameters determined through such modeling calibration and verification. Errors of the simulated benzene concentrations range from −13.8 to 248.9 μg/L, with a mean absolute error of 72.78 μg/L, and a mean relative error of 30.02%. This revealed that the error levels are generally acceptable for the simulation of benzene.

Before producing proxy modules through CLR, the pumping rates (explanatory variables) were normalized between 0 and 1 by dividing their values by 100 to reduce computational errors. Also, the natural logarithms of contaminant concentrations were considered as the outputs of proxy modules to avoid scaling problems. Each proxy module merely outputs the concentrations at one monitoring well after a period of pumping. In this study, a total of 18 proxy modules were developed for the three pumping periods (i.e., 5, 10, and 15 years). Remediation schedules with longer or shorter periods were not considered because shorter periods could hardly guarantee the environmental and risk standards to be satisfied and longer ones would be less efficient.

In the optimization, the environmental standard (for benzene) was determined to be 500 μg/L (for agricultural areas) in terms of the guideline issued by SERM (*33*). To examine the effects of risk levels on the optimized pumping strategies, the carcinogenic risk standard was assumed to be varied from 3×10^{-5} to 1×10^{-4}. Note that the noncarcinogenic risk was not included in the optimization process, as it was found that the HQ level would be less than 1 once the ELCR level was lower than the standard.

Figure S4 of the Supporting Information shows the predicted benzene concentration distributions after 5, 10, and 15 years of natural attenuation. It is indicated that the predicted peak concentration would be around well BH118, with the respective concentrations being 2.79, 2.40, and 2.19 times the environmental standard. Figure S5 of the Supporting Information gives the predicted benzene carcinogenic risk distributions over the domain. The predicted peak ELCR levels would be, respectively, 145, 124, and 105 times a risk standard of 1.0×10^{-5}. Apparently, the times of peak ELCR levels exceeding risk standard are much greater than those of benzene concentrations exceeding environmental standard. Therefore, the human-health risk should be paid more attention than environmental quality when determining optimal remediation policies for the site.

Figure 3 presents the optimal remediation policies according to the three remediation periods. It is found that the six wells would play different roles in remediating the groundwater. For the 5-year remediation policy, well-pair M3 and M5 would be the most important contributors (Figure 3a). Approximately 30% of the total pumping rate would be from well M5, while well M3 would have to be fully operated (i.e., its pumping rate need to be equal to or near 100 m^3/d when the risk standard ranges from 3×10^{-5} to 1×10^{-4}. Well-pair M1 and M2 would have medium contribution to the remediation through extracting (or injecting) parts of contaminated (or clean) water. In comparison, well-pair M4 and M6 would not be operated due to their negative hydrogeological conditions. If the remediation duration increases to 10 years, the specific pumping policy would be slightly different from that under the 5-year duration (Figure 3b). It is well-pair M1 and M3 that would significantly affect the remediation. When the risk standard is 3×10^{-5}, the optimal pumping rates at wells M1 and M3 would be 100 and 73.49 m^3/d, respectively; when it rises to 1×10^{-4}, however, their pumping rates would be reduced by 82.19% and 60.55%, respectively. The pumping rates of well-pair M4

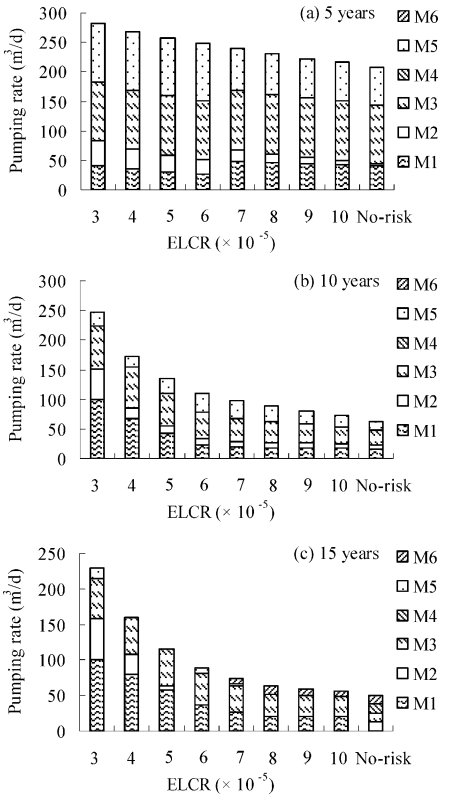

FIGURE 3. Optimal pumping of the wells for the 5-, 10-, and 15-year remediation policies, where no-risk means the policy obtained from the model without health-risk constraints.

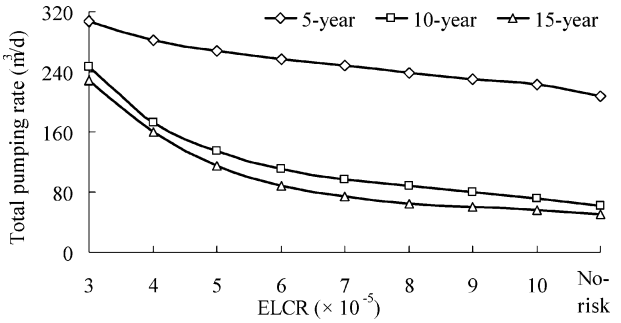

FIGURE 4. Optimal total pumping rate, where no-risk means the policy obtained from the model without health-risk constraints.

and M6 would still be zero under this policy. For the remediation duration of 15 years, the optimal remediation policy would be similar to that for the 10-year duration except that well M6 would be used when the risk standard is rather high (larger than 6×10^{-5}) (Figure 3c).

Figure 4 presents the optimal total pumping rate versus carcinogenic risk level for the three remediation periods. It is obvious that a long-term remediation period corresponds to a low total pumping rate. This could be interpreted by the following mechanism. When the pumping period is short, large amounts of contaminants need to be extracted to lower the benzene concentrations as the natural degradation of the aquifer cannot be fully capitalized. The growth of extraction rates would correspondingly enhance the injection rates in order to maintain a stable hydraulic gradient of the aquifer. The increases in both extraction and injection rates would thus cause the rise of total pumping rate. It is also observed that a stringent risk standard would lead to a high total pumping rate. For example, when the risk standard decreases from 1×10^{-4} to 3×10^{-5}, the total pumping rate would respectively increase by 31.02, 2.44, and 3.09% under

the 5, 10 and 15-year remediation policies. This is because a stringent risk standard requires an increased amount of contaminant to be removed, leading to the enhancement of the total pumping rate under a given remediation period.

Figure 4 compares the optimal remediation policies under health-risk consideration to those without such a concern. It is indicated that the total pumping rate would be intensively increased when health risk is considered. For instance, given a risk standard of 3×10^{-5}, the total pumping rate would be 307.15 m^3 per day under the 5-year remediation policy; this is 48.31% higher than that obtained from the model without risk-related constraints. Despite potential high remediation costs caused by the enhanced pumping rates, it could be much preferable to decision makers and local residents in terms of simultaneously improving environmental quality and reducing human-health risk.

To examine the performance of identified optimal remediation policies, postoptimization simulation was then conducted to evaluate the environmental quality (indicated by the predicted contaminant concentrations after the suggested remediation actions) and the resulting health risks. Figure S5 of the Supporting Information presents the predicted carcinogenic risks after a 5-year period of remediation. As shown in Figure S5b−d, the peak risk levels under the three risk standards would be less than 3×10^{-5}, 6×10^{-5}, and 1×10^{-4}, respectively. These are all lower than the peak risk level (i.e., 1.4×10^{-3}) obtained through the model without risk considerations (Figure S5a). This demonstrates that the health risk exposure to benzene would be significantly reduced through HRGH. Similar conclusions can also be summarized for the 10- and 15-year remediation policies (Figures S6 and S7). Although not shown in this study, the simulated benzene concentrations would meet the environmental standard of 500 μg/L, and the resulting noncarcinogenic risk (HQ) would be less than the HQ risk standard of 1.0.

The environmental standard should be determined in terms of governmental environmental guidelines, which vary for different countries, states, and provinces. For example, the guidelines for xylenes are 20, 40, 300, 400, 530, 1800, and 10,000 μg/L in Sweden, New Jersey, Canada, Japan, North Carolina, California, and Illinois, respectively (*39*). They are also dissimilar in different areas and functionalities. For instance, the guideline for benzene is 500 and 5000 μg/L, respectively, for agriculture and forest areas, as well as 300 and 5 μg/L for freshwater aquatic life and drinking water, respectively (*32*). Despite various alternatives, which one should be selected mainly relies on decision makers with regard to the goals and local nature. As the findings of this study were merely obtained based on the standard of 500 μg/L, cares should be taken when other environmental standards need to be satisfied.

According to World Health Organization (*40*), a value of 1×10^{-5} for ELCR is suggested as a tolerable carcinogenic risk level. This level implies that the probability of the impact on human health (i.e., developing cancer) would be less than 1 out of 100,000 people who use the onsite groundwater as the drinking water source. However, this regulation is not always exclusive, as a risk level between 1×10^{-6} and 1×10^{-4} or higher does not absolutely correspond to an observed adverse health effect (*41*). This study investigated the effects of risk standards from 3×10^{-5} to 1×10^{-4} on optimal remediation policies. If more stringent standards need to be satisfied, more applicable actions would be suggested, such as increasing remediation duration and the number of wells, implementing enhanced remediation techniques (e.g., air sparging, surfactant-enhanced remediation aquifer, chemical oxidation, etc.). Future works might thus be concentrated on examining the feasibility of applying these techniques, evaluating the associated operation and maintenance costs, and identifying the optimal remediation policies.

Heterogeneity of the contaminated aquifer was caused by spatially varied soil types; however, the aquifer properties were assumed to be homogeneous within an identical type of soil. Future studies would thus be undertaken by incorporating additional geophysical data into parameter estimation for better site characterization (*42*). Moreover, the statistical samples were obtained through the Monte Carlo (MC) sampling technique, so the optimization results could be affected by the quality of MC samples. Latin hypercube sampling (LHS), as another sampling technique, can also be used for dealing with random simulation problems and saving computational efforts. Examining whether LHS has higher computational efficiency than MC in handling random variables also deserves study.

Identification of proxy modules capable of evaluating the health risks associated with other contaminants existing as DNAPLs (e.g., TCE and PCE) also deserves attention in future studies. For PT systems, the remediation would last for a long time as the contaminants continuously dissolve in the groundwater until the residual oil phase is finally exhausted. Therefore, enhanced remediation techniques such as bioremediation would be useful in speeding up the remediation process. The results obtained from this study may provide implications in identifying optimal remediation policies for these techniques.

Prediction errors of proxy modules may exist, probably due to the limitation of the CLR method. To mitigate the effects of prediction errors on optimization results, integrated statistical techniques such as multiple regression and artificial neural network could be used to create proxy models. Through fusing the advantages of these techniques, the effect of estimation errors on decision making can be significantly reduced. Actually, any statistical method that performs exceptionally satisfactory in one situation may not be effective in others (*43*). This implies that any proxy module may hardly be of universal superiority under all situations. Therefore, when extending this effort to other sites, the model's training and predicting accuracy need to be retested.

Implications

Based on the simulation and risk assessment results, a health-risk-based groundwater management model was developed. It incorporated the considerations of environmental quality and human-health risks into the same framework. While this effort was specifically undertaken for optimizing petroleum-contaminated groundwater remediation, it could offer some useful implications for other systems with different remediation measures as well as environmental and health concerns. Many statistical methods such as parametric regression (*3, 44*) and nonparametric regression (*8, 9, 27, 45*) can also be used to create proxy modules; however, the difficulty in simultaneously dealing with nonlinear and discrete relations between remediation policies and remediation performance stimulated the development and application of the CLR method.

The proxy modules created by CLR provide a bridge directly linking the health risk assessment to the optimization process. This bridge is effective in mitigating the computational efforts in optimization processes. Through replacing the simulation and health risk assessment modules with the proxy ones, many orders of magnitude of computational cost was saved. Take this case for an example. One simulation requires an averare of 5 min CPU time to evaluate the risk levels under one remediation policy. Assuming that the optimization process needs 1500 simulation calls to find the optimal solution, a general optimization method not using proxy modules then requires about 125 h of CPU time. In

comparison, approximately 3000 simulations can be accomplished per second through the developed proxy modules; thus the optimal solution could be obtained within one second.

The optimization results revealed that (i) a long remediation period corresponds to a low total pumping rate, (ii) a stringent risk standard implies a high total pumping rate, and (iii) the human-health risk associated with benzene contamination would be significantly reduced if HRGM is used. Furthermore, although the remediation cost would rise due to the increased total pumping rate, it could be preferable to decision makers and local residents due to the model simultaneously improving environmental quality and reducing human-health risk. These implications would assist decision makers in understanding the effects of remediation duration and human-health risk level on optimal remediation policies and in designing a robust groundwater remediation system.

Results from postoptimization simulation indicated that the noncarcinogenic risk (HQ) would have satisfied the risk standard of 1.0 once the carcinogenic risk (ELCR) met the regulated risk standard. Therefore, HQ was not included in the optimization process in this study. However, care should be taken when extending this approach to other practical sites as the associated parameters for evaluating carcinogenic and noncarcinogenic risks could vary by case and be subject to complex uncertainties.

Acknowledgments

We thank the associate editor and anonymous reviewers for their helpful comments and suggestions. This research was supported by the Major State Basic Research Development Program of MOST (2005CB724200 and 2006CB403307) and the Natural Science and Engineering Research Council of Canada.

Supporting Information Available

Detailed descriptions of the simulation module, health risk assessment modules, and clusterwise linear regression, as well as figures and tables mentioned within the text.This material is available free of charge via the Internet at http://pubs.acs.org.

Literature Cited

(1) Chen, Z.; Huang, G H.; Chakma, A. Numerical modeling soil and groundwater contamination - A study for petroleum-contaminated site;Technical report prepared for TransGas - SaskEnergy, Canada, 1998.

(2) Gorelick, S. M.; Evans, B. J.; Remson, I. Identifying sources of ground-water pollution: An optimization approach. *Water Resour. Res.* **1983**, *19* (3), 779–790.

(3) Gorelick, S. M.; Voss, C. I.; Gill, P. E.; Murray, W.; Saunders, M. A.; Wright, M. H. Aquifer reclamation design: the use of contaminant transport simulation combined with nonlinear programming. *Water Resour. Res.* **1984**, *20* (4), 415–427.

(4) Lefkoff, L. J.; Gorelick, S. M. Design and cost analysis of rapid aquifer restoration systems using flow simulation and quadratic programming. *Ground Water* **1986**, *24* (6), 777–790.

(5) Wagner, B. J.; Gorelick, S. M. Optimal groundwater quality management under parameter uncertainty. *Water Resour. Res.* **1987**, *23* (7), 1162–1174.

(6) Ahfeld, D. P.; Mulvey, J. M.; Pinder, G. F. Contaminated groundwater remediation design using simulation, optimization, and sensitivity theory- 1. Model development. *Water Resour. Res.* **1998**, *24* (3), 431–441.

(7) Ahfeld, D. P.; Mulvey, J. M.; Pinder, G. F.; Wood, E. W. Contaminated groundwater remediation design using simulation, optimization, and sensitivity theory- 2. Analysis of a field site. *Water Resour. Res.* **1988**, *24* (3), 443–452.

(8) Rogers, L. L.; Dowla, F. U. Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling. *Water Resour. Res.* **1994**, *30* (2), 457–481.

(9) Rogers, L. L.; Dowla, F. U.; Johnson, V. M. Optimal field-scale groundwater remediation using neural networks and the genetic algorithm. *Environ. Sci. Technol.* **1995**, *29* (5), 1145–1155.

(10) McKinney, D. C.; Lin, M. D. Approximate mixed-integer nonlinear programming methods for optimal aquifer remediation design. *Water Resour. Res.* **1995**, *31* (3), 731–740.

(11) McKinney, D. C.; Lin, M. D. Pump-and-treat groundwater remediation system optimization. *J. Water Res. Pl. - ASCE* **1996**, *122* (2), 128–136.

(12) Culver, T. B.; Shoemaker, C. A. Dynamic optimal groundwater reclamation with treatment capital costs. *J. Water Res. Pl. - ASCE* **1997**, *123* (1), 23–29.

(13) Huang, C. L.; Mayer, A. S. Pump-and-treat optimization using well locations and pumping rates as decision variables. *Water Resour. Res.* **1997**, *33* (5), 1001–1012.

(14) Minsker, B. S.; Shoemaker, C. A. Differentiating a finite element biodegradation simulation model for optimal control. *Water Resour. Res.* **1996**, *32* (1), 187–192.

(15) Minsker, B. S.; Shoemaker, C. A. Dynamic optimal control of in-situ bioremediation of ground water. *J. Water Res. Pl. - ASCE* **1998**, *124* (3), 149–161.

(16) Mayer, A. S.; Kelley, C. T.; Miller, C. T. Optimal design for problems involving flow and transport phenomena in subsurface systems. *Adv. Water Resour.* **2002**, *25* (8−12), 1233–1256.

(17) Baú, D. A.; Mayer, A. S. Stochastic management of pump-and-treat strategies using surrogate functions. *Adv. Water Resour.* **2006**, *29* (12), 1901–1917.

(18) Baú, D. A.; Mayer, A. S. Data-worth analysis for multiobjective optimal design of pump-and-treat remediation systems. *Adv. Water Resour.* **2007**, *30* (8), 1815–1830.

(19) Guan, J. B.; Aral, M. M. Optimal remediation with well locations and pumping rates selected as continuous decision variables. *J. Hydrol.* **1999**, *221* (1), 20–42.

(20) Mulligan, A. E.; Ahlfeld, D. P. A new interior-point boundary projection method for solving nonlinear groundwater pollution control problems. *Oper. Res.* **2002**, *50* (4), 636–644.

(21) Zheng, C.; Wang, P. P. A field demonstration of the simulation-optimization approach for remediation system design. *Ground Water* **2002**, *40* (3), 258–266.

(22) Qin, X. S.; Huang, G. H.; He, L. Simulation and optimization technologies for petroleum waste management and remediation process control. *J. Environ. Manage.* **2008**, doi: 10.1016/j.jenvman.2008.07.002.

(23) Smalley, J. B. Risk-based in situ bioremediation design. Thesis, University of Illinois at Urbana-Champaign, Urbana, Illinois, 1998.

(24) Smalley, J. B.; Minsker, B. S.; Goldberg, D. E. Risk-based in situ bioremediation design using genetic algorithm. *Water Resour. Res.* **2000**, *36* (10), 3043–3051.

(25) Wong, H. S.; Yeh, Y. W. W.-G. Uncertainty analysis in contaminated aquifer management. *J. Water Res. Pl. - ASCE* **2002**, *128* (1), 33–45.

(26) DeKay, M. L.; Small, M. J.; Fischbeck, R. S.; Farrow, R. S.; Cullen, A.; Kadane, J. B.; LaveL.; Morgan, M. G.; Takemura, K. Risk-based decision analysis in support of precautionary policies. *J. Risk Res.* **2002**, *5* (4), 391–417.

(27) Yan, S.; Minsker, B. S. Optimal groundwater remediation design using an adaptive neural network genetic algorithm. *Water Resour. Res.* **2006**, *42* (5), doi: 10.1029/2005WR004303.

(28) Huang, G. H. Stepwise cluster analysis method for predicting air quality in an urban environment. *Atmos. Environ.* **1992**, *26B* (3), 349–357.

(29) Huang, G. H.; Huang, Y. F.; Wang, G. Q.; Xiao, H. N. Development of a forecasting system for supporting remediation design and process control based on NAPL-biodegradation simulation and stepwise-cluster analysis. *Water Resour. Res.* **2006**, *42* (6), doi: 10.1029/2005WR004006.

(30) He, L.; Huang, G. H.; Lu, H. W.; Zeng, G. M. Optimization of Surfactant-Enhanced Aquifer Remediation for a Laboratory BTEX System under Parameter Uncertainty. *Environ. Sci. Technol.* **2008**, *42* (6), 2009–2014.

(31) He, L.; Huang, G. H.; Lu, H. W. A simulation-based fuzzy chance-constrained programming model for optimal groundwater remediation under uncertainty. *Adv. Water Resour.* **2008**, doi: 10.1016/j.advwatres.2008.07.009.

(32) CCME. *Subsurface Assessment Handbook for Contaminated Sites*; Canadian Council of Ministers of the Environment: Ottawa, ON, 1994.

(33) SERM. *Risk Based Corrective Actions for Petroleum Contaminated Sites, Province of Saskatchewan*; Saskatchewan Environment and Resource Management: Regina, Saskatchewan, Canada, 2002.

(34) EEP. *Numerical simulation for contaminant flow and transport in subsurface- A study of soil and groundwater contamination at the Coleville Site*; Process Report, Energy and Environment Program, University of Regina: Regina, Saskatchewan, Canada, 2005.
(35) Motott, L. S.; Bartelt-Hunt, S. L.; Rabideau, A. J.; Flower, K. R. Application of heuristic optimization techniques and algorithm tuning to multilayered sorptive barrier design. *Environ. Sci. Technol.* **2006**, *40* (20), 6354–6360.
(36) Guan, Y.; Sherman, M.; Calvin, J. A. A nonparametric test for spatial isotropy using subsampling. *J. Am. Stat. Assoc.* **2004**, *99* (467), 810–821.
(37) Iorgulescu, I.; Beven, K. J. Nonparametric direct mapping of rainfall-runoff relationships: an alternative approach to data analysis and modeling. *Water Resour. Res.* **2004**, *40* (8), doi: 10.1029/2004WR003094.
(38) Elgaali, E.; Garcia, L. A. Using neural networks to model the impacts of climate change on water supplies. *J. Water Res. Pl. - ASCE* **2007**, *133* (3), 230–243.
(39) Li, J. B. Development of an inexact environmental modeling system for the management of petroleum-contaminated sites. Dissertation, University of Regina, Regina, Saskatchewan, 2003.
(40) WHO. *Guidelines for Drinking-Water Quality*, 2nd ed.; Volume 2-Health Criteria and other supporting information; World Health Organization: Geneva, 1996.
(41) Geomatrix Consultants Inc. *Baseline human health risk assessment of on-site soil and groundwater*; Prepared for Sierra Pacific Industries, Project No. 9329, Task 13, 2003.
(42) Cassiani, G.; Medina, M. A., Jr. Incorporating auxiliary geophysical data into ground-water flow parameter estimation. *Ground Water* **1997**, *35* (1), 79–91.
(43) Magdon-Ismail, M. No free lunch for noise prediction. *Neural Comput.* **2000**, *12* (3), 547–564.
(44) Weber, C. L.; VanBriesen, J. M.; Small, M. J. A stochastic regression approach to analyzing thermodynamic uncertainty in chemical speciation models. *Environ. Sci. Technol.* **2006**, *40* (12), 3872–3878.
(45) Lin, Y. P.; Huang, G. H.; Lu, H. W.; He, L. A simulation-aided factorial analysis approach for characterizing interactive effects of system factors on composting processing. *Sci. Total. Environ.* **2008**, 402(2–3), 268–277.

ES800834X