Environmental Modelling & Software 101 (2018) 183–193

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

A design of experiment aided stochastic parameterization method for modeling aquifer NAPL contamination

Zelin Li [a], Bing Chen [a, b, *], Hongjing Wu [a], Xudong Ye [a], Baiyu Zhang [a]

[a] The Northern Region Persistent Organic Pollution Control (NRPOP) Laboratory, Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, Newfoundland A1B 3X5, Canada
[b] College of Environmental Science and Engineering, Peking University, Beijing, China

ARTICLE INFO

Article history:
Received 4 December 2016
Received in revised form
19 December 2017
Accepted 20 December 2017
Available online 6 January 2018

Keywords:
Design of experiment
Stochastic
Parameterization
Subsurface NAPL contamination
Uncertainty

ABSTRACT

Numerical models have been widely applied in simulating subsurface Non-aqueous Phase Liquid (NAPL) contamination processes. In order to examine modeling uncertainties and improve simulation performance, a new hybrid stochastic - design of experiment (DOE) aided parameterization method was developed by using a coupled experimental and modeling approach. In a case study, an existing commercial groundwater modeling tool BioF&T 3D was applied to conduct numerical simulations of subsurface contamination processes based on flow cell experiments. Parameterization results indicated that porosity, distribution coefficient, and Henry's constant were the most significant parameters. The result also revealed their interactions. The DOE predicted responses were found reasonably close to the actual ones from the models' simulations. Monte Carlo simulation was applied to conduct uncertainty analysis within the narrowed parameters ranges, which were generated by centralizing the DOE optimized values, and the combinations of parameters were further updated when better responses were found. After parameterization, R^2 valued 0.80, 0.91, 0.89, and 0.90 for benzene, toluene, ethylbenzene, and xylene (BTEX), respectively. A good consistency ($R^2 = 0.76$ to 0.90 for BTEX) was also achieved during the model verification, which confirmed that after the parameterization processes, the simulation model can potentially be used for predictions under similar conditions.

1. Introduction

Subsurface Non-aqueous Phase Liquids (NAPLs) contamination from spill and leakage of petroleum products has become a major environmental concern. Due to the long persistence time and complicated treatment required, NAPLs are difficult to be removed from water and soil leading to a high potential risk to human health in a long-term timespan (Yang et al., 2012; Jácome and Van Geel, 2015). NAPLs leakage and contamination accidents are closely associated with human activities, and it can also raise the risk to cause severe environmental and health hazards (García-Junco et al., 2001). In order to achieve a better understanding of the fate of contaminants in aquifer, numerical simulation has been generally accepted as an effective tool and continuously studied (Kim and Corapcioglu, 2003; Yu et al., 2010). The knowledge can be further used to support decision making on monitoring and remediation practices. As a simplified representative, the numerical model is able to provide the simulated outputs for the modeled system. The simulated outputs based on various model settings can reflect different scenarios. Through numerical simulations, more resources, energy and time can be saved leading a more efficient decision making process. Ideally, numerical model should be able to reflect the realistic situations. However, there is often a lack of fit existing due to uncertainty in the modeling system (especially the imprecisely defined parameter uncertainties). Uncertainty is inevitable at the current stage due to limited knowledge and simplification of numerical models. In order to increase the reliability of the modeling results, it is essential to conduct uncertainty and sensitivity analysis, and properly calibrate the model, such that discrepancies between simulated and observed data can be minimized.

Parameter uncertainties have been extensively studied, particularly integrated with sensitivity analysis and model calibrations (Sin et al., 2011; He et al., 2012; Shen et al., 2012; Zhuo et al., 2013;

* Corresponding author. The Northern Region Persistent Organic Pollution Control (NRPOP) laboratory, Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, Newfoundland A1B 3X5, Canada.
E-mail address: bchen@mun.ca (B. Chen).

https://doi.org/10.1016/j.envsoft.2017.12.014
1364-8152/

Houska et al., 2014). To consider the parameter uncertainties, the optimal results from the different combinations of the parameter settings were analyzed. Monte Carlo simulation is one of the most common stochastic methods involving random sampling within certain types of distributions. Due to the ease of implementation and generalization, Monte Carlo methods have been widely applied to environmental systems by propagating the parameter uncertainties and evaluating their impacts on the model output (Huang and Loucks, 2000; Jing et al., 2013; Li et al., 2014).

One-factor-at-a-time (OFAT) is one of the traditional sensitivity analysis methods. This method simply adjusts one parameter at a time while keeping other parameters fixed. Its applications have been found in multiple studies conducted on various models (Holvoet et al., 2005; Jing and Chen, 2011). Garson (1991) introduced another method based on the concept of Artificial Neural Networks (ANN) by assigning the neural net weight matrix. This method has been further implemented in many other studies to find the relative importance of input variables in different processes modeling (Kermani and Ebadi, 2012; Nourani and Fard, 2012; Jing et al., 2014). However, both of these methods have their limitations. OFAT is incapable of revealing the interactions between parameters, which might lead to the ignorance of the potentially significant variables (Peeters et al., 2014). Garson's method is debatable due to the “black box” nature of ANN (Olden et al., 2004; Witek-Krowiak et al., 2014).

The traditional method of model calibration is essentially a trial and error process which uses iterations to adjust the relevant parameters until the simulated outputs are sufficiently close to the experimental data. This method is still popular and has been embedded in commercial modeling tools for automatic calibration (Sonnenborg et al., 2003; Mugunthan et al., 2005; Razavi and Tolson, 2012). Existing studies have focused on optimizing the mathematical algorithms to achieve a more efficient calibration process (Wu et al., 2012, 2014; Zhao et al., 2013). Calibration methods have also been improved by applying parameter estimation and global optimization within reasonable predefined intervals (Kang, 2014; Plasencia et al., 2014; Yen et al., 2014; Zhang et al., 2015). Despite that a good fit can be expected, it should not be ignored that some major limitations such as extensive computational requirements, low physical plausibility, and over-parameterization exist when traditional calibration methods are employed (Van Griensven et al., 2006; Whittaker et al., 2010; Okamoto and Akella, 2012).

To address this issue, design of experiment (DOE) provides a parameterization option. DOE is a well-known statistical methodology, which can unveil the interrelationships between parameters and the corresponding responses by conducting controlled experiments (Park, 2007; Strigul et al., 2009). By using DOE, it is possible to simultaneously study several parameters and their interactions (Veličković et al., 2013; Sarikaya and Güllü, 2015). DOE was originally developed to guide the planning and setup for physical experiments. However, considering that the complexity and cost can increase dramatically with the growing number of input variables, numerical simulation tools have been extensively involved. Especially in recent studies, Wu et al. (2012) used a DOE aided method to conduct sensitivity analysis and parameterization for a hydrological model SLURP and optimized the predicted regression equation, which has resulted in a greater goodness-of-fit value compared to the one achieved by the automatic calibration function within the model. Zahraee et al. (2013) introduced DOE in modeling a real-world construction process to achieve optimal resource levels and maximize the process productivity. In the study of Al-Shalabi et al. (2014), seven uncertain design parameters for a low salinity water injection process were screened by using DOE method, followed by the optimization of cumulative oil recovery using the Response Surface Methodology (RSM). Though DOE aided methods have proven advantages in conducting parameterization for numerical models, relationships between responses and stochastically distributed parameters are seldom integrated. Besides, it has rarely been used in groundwater and subsurface contamination models, in which uncertainties commonly exist and knowledge concerning complicated interactions between each parameter is far from adequate.

Targeting the subsurface hydrocarbon contamination, the objective of this study was to develop a new parameterization method to examine modeling uncertainties and improve simulation performance by using a coupled experimental and modeling approach. The research tasks entail: 1) to conduct a flow cell experiment to physically simulate subsurface hydrocarbon contamination and natural attenuation; 2) to employ the BioF&T 3D model to numerically simulate the contamination and natural attenuation processes; 3) to develop a new hybrid stochastic - DOE parameterization (HSDP) method for improving the modeling performance by quantifying the significance of modeling parameters and their interactions and evaluating the influence of uncertainties. Diesel fuel, a frequently used liquid fuel, has the advantage of easy access, strong volatility, short test period as well as time and cost saving. Thus, in this study, diesel fuel was selected as the appropriate NAPL contamination source with BTEX as the targeted compounds for lab analysis as well as numerical modeling.

2. Methodology

2.1. Flow cell experimental setups

In this study, the physical model was built based on a pre-manufactured flow cell, which was used to prescribe the simulation domain focused on the longitudinal and vertical directions, and provide data for parameterization and verification of the numerical model. As shown in Fig. 1, the flow cell was fabricated with organic glass and was installed on aluminum framed mobile base. 20 sampling ports with a space of 15 cm between each were placed on the front panel of the flow cell for aqueous sample collections. Dimensions of the flow cell are 82.5 cm × 55 cm × 4 cm. Water can be introduced from the inlet installed at the top left corner, and the effluent can be discharged through the outlet installed with a globe valve at the bottom right corner of the flow cell (Li et al., 2015; Li et al., 2017).

Prior to soil loading, water proofing of the flow cell was carefully checked for leakages. Commercial white play sand SHAW® was purchased and screened with a 2-mm mesh size sieve. The soil

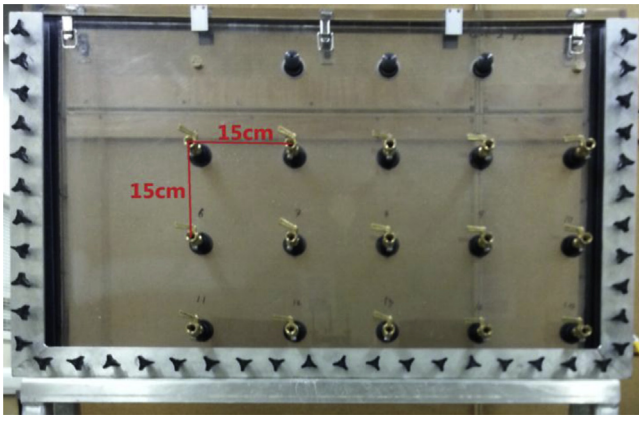

Fig. 1. Outlook of the flow cell unit.