Neural Network Modeling of Hydrocarbon Recovery at Petroleum Contaminated Sites

J. B. Li[1], G. H. Huang[1], Y. F. Huang[1], A.Chakma[1], and G. M. Zeng[2]
[1] Faculty of Engineering, University of Regina, Regina, Sask S4S 0A2, Canada
Tel: 1-306-5854095, Fax: 1-306-5854855, E-mail: huangg@uregina.ca
[2] Department of Environmental Engineering, Hunan University, Changsha, Hunan 410082, China
Tel: 86-8823701, Fax: 86-8823413, E-mail: zgming@mail.hunu.edu.cn

Abstract: A recurrent artificial neural network (ANN) model is developed to simulate hydrocarbon recovery process at petroleum-contaminated site. The groundwater extraction rate, vacuum pressure, and saturation hydraulic conductivity are selected as the input variables, while the cumulative hydrocarbon recovery volume is considered as the output variable. The experimental data for establishing the ANN model are from implementation of a multiphase flow model for dual phase remediation process under different input variable conditions. The complex nonlinear and dynamic relationship between input and output data sets are then identified through the developed ANN model. Reasonable agreements between modeling results and experimental data are observed, which reveals high effectiveness and efficiency of the neural network approach in modeling complex hydrocarbon recovery behavior.

1. Introduction

Subsurface contamination by leakage and spill of low-density nonaqueous phase liquids (LNAPLs) from petroleum industry has been of much concern in recent years [1]. Among various remediation measures for cleaning up such contaminations, dual-phase vacuum extraction (DPVE) is a popular cost-effective emerging technology to enhance remediation efficiency by recovering petroleum hydrocarbon from the subsurface [2]. This technology applies a high vacuum system to remove various combinations of contaminated groundwater, free product, and hydrocarbon vapor from the subsurface, and the extracted liquids and vapor are collected and then treated above ground. However, appropriate system design should be desired to improve system performance, and it is recognized that the effectiveness and reliability of the remediation system design should depend on suitable multiphase flow and transport models [3]. Many researchers have devoted enormous efforts to propose many mathematical models based on various assumptions to predict the system behavior under different remediation activities, but few models have been developed and applied to simulate the DPVE process during the past years [4].

Generally speaking, the reliable multiphase flow and transport model development and application depend on the availability of a set of important site-specific data, and the model should be able to describe the nonlinear and dynamic characteristics of the DPVE remediation process. Unfortunately, it is very difficult to characterize the complete set of input data through field investigations due to excessive costs, and usually the site investigations are only limited to a selected number of monitoring wells and field tests [5]. In addition, it is also difficult to reflect the complex nonlinear and dynamic features of the DPVE process, and the numerical model is always complicated for the non-professional users, while the model implementation is usually time consuming.

While the multiphase flow and transport models are of importance in the understanding of the remediation processes, there are many practical situations where the main concern is to make reasonable predictions of the system performances such as hydrocarbon recovery forecasting under various system configurations. In such situations, the engineers or users may prefer not to spend excessive time and efforts to develop and implement a complicated multiphase flow numerical model and instead implement a simpler remediation system model that can also reflect the nonlinear and dynamic features of the DPVE process. Recently, significant progress in the fields of nonlinear and dynamic pattern recognition and system control theory have been made through advances of artificial neural network modeling (ANN). ANN is a nonlinear and dynamic mathematical structure with the ability to represent the complex nonlinear and dynamic processes that relate the inputs and outputs of any system [6]. The successful application of neural network modeling in many sicence and engineering fields may suggest that it is an alternative to multiphase flow and transport model to offer an effective means for remediation system design [7, 8]. However, few attempts have been made to apply this technique to remediation simulation work, especially to the DPVE process modeling [9].

In this paper, a recurrent artificial neural network model is developed to simulate hydrocarbon recovery process for a petroleum-contaminated site that is undergoing the DPVE remediation, and the complex nonlinear and dynamic relationship between input and output data sets are identified. The networks are trained from a great number of experimental data obtained from careful implementation of a well-developed multiphase flow model associating recovery system configuration variations (i.e., extraction rate, vacuum pressure, soil type) to selected outcome (i.e., total hydrocarbon recovery). The developed neural network model can then be applied for hydrocarbon recovery simulation with only little conceptual understanding of the remediation process complexities.

2. Methodology

ANNs are computing tools composed of many simple interconnected elements called neurons with a unique ability of recognizing underlying relationships between input and output events. A typical neuron is shown in Figure 1. A neuron has two components [10]: (1) a weighted sum which performs a weighted summation of the inputs with components (X_1, X_2, X_3, ..., X_n), i.e., $s = \sum w_i X_i + b$, where b is the bias of the networks, and (2) a linear, nonlinear or logic function which gives an output corresponding to s. Here, many kinds of functions can be used, including threshold (logic), sigmoid, hyperbolic tangent and Gaussian functions. In this paper, each of them is examined at each neuron during the training process in order to get desired ANNs.

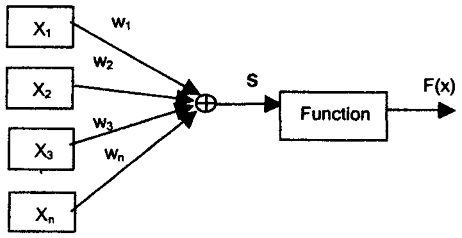

Figure 1 Basic components of a neuron

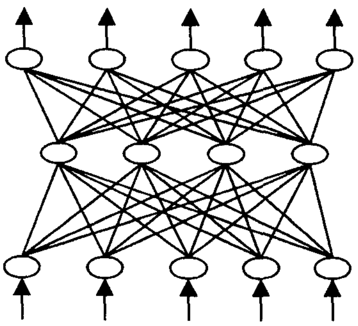

Figure 2 A fully interconnected three-layered back-propagation network

In a typical ANN, there are three types of neurons [6]: (a) input neurons which may receive external data, (b) output neurons which send data out of the ANN, and (c) hidden neurons whose signals remain within the ANN. There are three types of layers corresponding to the types of neurons. The hidden neurons may form one or more hidden layers. The neurons in each layer are usually fully interconnected with neurons from neighboring layers. The importance of each inter-neuron connection is determined by its numerical value. A three-layered back-propagation network structure is depicted in Figure 2 [6, 10]. The ANN shown in Figure 2 has an input layer, an output layer, and one hidden layer. The input layer contains an array of variables into which the input data of the system are read from an external source. Similarly, the predicted data or results, which can be multiple vectors, are written in the output layer. Initially, the input layer receives the input and passes it to the hidden layer. If more hidden layers exist, the processed information from the first hidden layer is then passed the next hidden layer for processing. Finally, the output layer receives information from the last hidden layers. In this paper, the number of hidden layer is not fixed. The training tool will automatically select suitable number of hidden layer to get the desired ANN model.

When an ANN is constructed, small numbers (weights) are assigned randomly to the connections between neurons. In general, the output from neuron j in layer k can be calculated by the following equation:

$$u_{jk} = F_k\left(\sum_{i=1}^{N_{k-1}} w_{ijk} u_{i(k-1)} + b_{jk}\right) \tag{1}$$

where coefficients w_{ijk} and b_{jk} are connection weight and bias of the network, respectively; they are fitting parameters of the model. The purpose is to obtain a mapping from an input vector to an output one. It is desired that the difference between the predicted and the observed (actual) values in the output vector be as small as possible. The fitting parameters are modified until an error criterion between the input and the output is satisfied based on the topology of the ANN and the learning technique. The adjustment of the weights is defined as the learning process. The ANN is tested with input/output values used in training. After training and testing, the network is ready to perform tasks such as pattern recognition, classification, or function approximation.

There are mainly two types of networks, feed-forward networks and recurrent networks. In this paper, the back-propagation technique with momentum is used. The fitting procedure from which weights w_{ijk} are determined is performed using a least-squares minimization routine. In this routine, the sum of root-squared relative errors between the calculated and the experimental data is to be minimized.

In general, the back-propagation method uses the following steps [10] : (a) read a specific input and calculated its corresponding output; (b) if the error between the produced output and the desired output is acceptable, then stop; (c) if the error is unacceptable in step (b), then the weights are adjusted for all of the interconnections that go into the output layer. An error value is calculated for all of the units in the hidden layer that is just below the output layer, and the weights are then adjusted for all interconnections that go into the hidden layer. The process is continued until the last layer of weights has been adjusted. Typically an application of back-propagation requires both a training set and a test set. Both the two sets contain input/output pattern pairs. While the training set is used to train the network, the test set is used to assess the performance of the network after the training is complete. To provide the best test of network performance, the test set should be different from the training set. The most successful ANN architecture is the one which has the smallest prediction error on a data set for which it was not trained.

The system performance of dual phase remediation process is affected by a number of factors such as groundwater extraction rate at the recovery well, the vacuum pressure, the soil type at the site, initial free product thickness, groundwater gradient, etc.. In order to develop a neural network model for dual phase hydrocarbon recovery process, the groundwater extraction rate, vacuum pressure, and saturated hydraulic conductivity are selected as input variables, while the cumulative hydrocarbon recovery volume is selected as the output variable. The data used for developing the ANN model are from careful implementation of a multiphase flow model for DPVE process [11].

3. Case Study

A petroleum-contaminated site that is undergoing dual phase hydrocarbon recovery program is illustrated in Figure 3 [4]. The site domain is 400 m × 400 m. The initial free product thickness is shown in Figure 4. In order to recover hydrocarbon from the subsurface, three recovery wells are installed at (160, 200), (180, 160), and (180, 220), respectively. The multiphase flow model [4, 11] is then implemented under various groundwater extraction rate, vacuum pressure, and hydraulic conductivity conditions, to generate a number of pairs of input and output values (cumulative hydrocarbon recovery volume at 300 days later after system initiation) for establishing the ANN model.

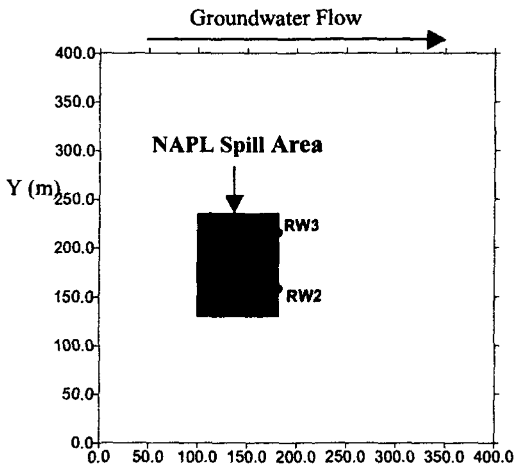

Figure 3 Site domain for dual phase hydrocarbon recovery simulation

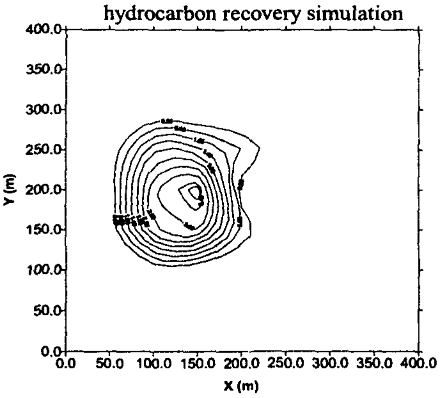

Figure 4 Initial free product thickness (m) for the study site

4. Results Analysis

4.1 Trainning Results

The scatter plots in Figures 5 provide comparisons of the measured cumulative oil volume at t = 300 days with the ANN-derived ones. Figures 5 presents the error analysis results for ANNs.

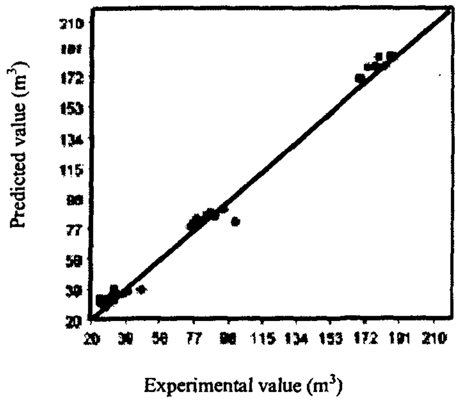

Figure 5 The experimental versus predicted cumulative oil volume (m3) at t=300 days

Table 1 shows the outputs of statistical analyses for calibration results from the ANN. It is indicated that the ANNs models have been well developed with low calibration errors. In detail, for the ANN model, the calibrated relative errors are 5.80%, the minimum relative error is 0.3%, the maximum relative error is 24.2%, the standard deviation is 9.87 (m^3), and the correlation coefficient is 0.996. Detailed trainning results are listed in Table 3 in appendix.

Table 1 Statistical analysis for calibration results from the ANN

Method	Average relative error (%)	Minimum relative error (%)	Maximum relative error (%)	Standard deviation (m^3)	Correlation coefficient
ANN	5.80	0.3	24.2	9.87	0.996

4.2 Application to cumulative oil volume forecasting

After the ANNs model was established, they could then be used for cumulative oil volume forecasting under a variety of operation conditions. With measured data sets that were not used for training, the modeling outputs could then be compared with measured values to verify the model's accuracy.

The predicted results of cumulative oil volume is showed in Figures 6. Totally, 14 combination of recovery operation condition are used as the input of the developed ANNs model. As shown, high prediction accuracies was obtained by the ANNs models.

Table 2 shows the results of error analyses for prediction outputs from the developed ANNs. Compared with the calibration results as shown in Table 1, the relative errors and standard deviations become higher, while the correlation coefficients are lower. In detail, the average relative errors are 1.76%, the minimum relative error is 0.24%, the maximum relative error is 3.59, the standard deviation is 21.4 (m^3), and correlation coefficients are 0.999. Detailed predicted results are listed in Table 4 in appendix. It indicates that the developed ANNs model can be efficiently used to replace of the complexty subsurface model for simulation, further uncertainty analysis and other decision analysis even with lower accuracy than the trainning results.

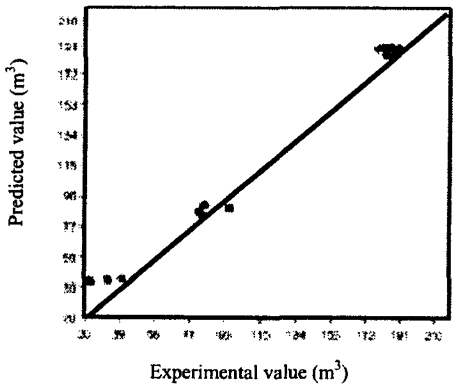

Figure 6 The experimental versus predicted cumulative oil volume (m3) at t=300 days

Table 2 Error analysis for prediction outputs

Method	Average relative error (%)	Minimum relative error (%)	Maximum relative error (%)	Standard deviation (m^3)	Correlation coefficient
ANN	17.96	0.10	89.56	9.93	0.999

5. Conclusions

A methodology based on artificial neural network was proposed in this paper to examine the hydrocarbon recovery system performance at the petroleum-contaminated site under various system configurations for the dual phase remediation process. The groundwater extraction rate, vacuum pressure, and saturation hydraulic conductivity are considered as the input variables, and the cumulative hydrocarbon recovery volume is selected as the output variable. The multiphase flow model developed for the dual phase remediation process is implemented to generate a great number of experimental data that contain a lot of pairs of input and output values. The experimental data are then used for network training and test, and the back-propagation technique is used for establishing the ANN model while the sum of root-squared relative errors between the calculated and the experimental data is minimized. The impacts of system parameters on the remediation process performance can be examined through the developed model, and the potential of ANN model for simulating the hydrocarbon recovery process behavior is then illustrated. Application of the proposed model can provide guidance on DPVE remediation system design and operation to determine optimum operation conditions, evaluate system performances, so as to prevent system failures and provide major cost savings. The proposed method will act as an effective tool for for site remediation design and management practices.

References

[1] Chen, Z., Huang, G. H., and Chakma, A., "Integrated environmental risk assessment for petroleum-contaminated sites: a North American case study;" *Water Sci. Tech.*, Vol.38, no.4/5, pp. 131-138, 1998.

[2] US Army Corps of Engineers, *Engineering and Design - Multi-Phase Extraction, EM 1110-1-4010, Washington, DC, 1999.*

[3] Charbeneau, R. J., Johns, R. T., Lake, L.W., and McAdams, M. J., "Free-product recovery of petroleum hydrocarbon liquids." *Ground Water Monitoring and Remediation*, Vol. 20, no.3, pp. 147-168, 2000.

[4] Draper Aden Environmental Modeling Inc., *MOVER Multiphase Organic Vacuum Enhanced Recovery Simulator: Technical Documentation & User Guide*, Blacksburg, VA, 1997.

[5] United States Environmental Protection Agency, *Abstracts of Remediation Case Studies: Volume 5*, EPA 542-R-01-008, Washington, DC, 2001.

[6] Hsu, K.-L., Gupta, H. V., and Sorooshian, S., "Artificial neural network modeling of the rainfall-runoff process." *Water Resour. Res.*, Vol. 31, no. 10, pp. 2517-2530, 1995.

[7] Baxter, C. W., Zhang, Q., Stanley, S. J., Shariff, R., Tupas, R.-R. T., and Stark, H. L., "Drinking water quality and treatment: the use of artificial neural networks." *Can. J. Civ. Eng.*, Vol. 28, no. 1, pp. 26-35, 2001.

[8] Perez, P., and Reyes, J., "Prediction of particulate air pollution using neural techniques." *Neural Comput. Applic.*, no. 10, pp. 165-171, 2001.

[9] Morshed, J., and Kaluarachchi, J. J., "Parameter estimation using artificial neural network and genetic algorithm for free-product migration and recovery." *Water Resour. Res.*, Vol. 34, no. 5, pp. 1101-1113, 1998.

[10] Dayhoff J.E., *Neural Network Architectures*, Van Nostrand Reinhold, New York, 1990.

[11] Li, J. B., Huang, G. H., Chakma, A., and Zeng, G. M., "Numerical simulation of dual phase vacuum extraction for the removal of nonaqueous phase liquids in subsurface: a Canadian case study." To be appeared in *Proceedings of Canadian International Petroleum Conference 2002*, June 11-13, 2002, Calgary, Canada, 2002.