

Comprehensive Model for Remediation of Sandy Soils Contaminated with Volatile Organic Compounds Using Thermal Enhancement of Soil Vapor Extraction Method

Mohammad Reza Sabour ·
Seyed Homam Seyed Jalali  · Ghorbanali Dezvareh

Received: 26 January 2017 / Accepted: 24 May 2017 / Published online: 8 June 2017
© Springer International Publishing Switzerland 2017

Abstract In today's world, remediation of the environmental pollutants including soil contaminations is among the main issues and concerns considered by environmental scientists. Vapor extraction method is an in situ method to clean up volatile and semi-volatile contaminants of soil especially in unsaturated areas. Thermal enhancement to extract vapors includes different technologies. Its purpose is to transfer heat to the subsurface of the soil to increase the vapor pressure of volatile organic compounds and, consequently, to increase the amount of extracted VOCs. In this study, modeling was done by using laboratory data after screening. Validation was also done with the help of an artificial neural network using the response surface methodology. After training and evaluating the model, it was found that this model determines the amount of contaminant removal rate according to available data and different temperatures by good measures. The correlation coefficient square was equal to 0.95 in the validation section by the neural network. This coefficient was equal to 0.99 in the original model. At the end, a contaminant removal formula for sandy soils has been presented. As a result, due to the proximity of the correlation coefficient to 1, this model can be used to

predict the removal rate of thermal enhancement in the relevant circumstances with a slight error.

Keywords Neural network modeling · Response surface methodology · Soil remediation · Soil vapor extraction · Thermal enhancement

1 Introduction

Nowadays, contamination of soil and groundwater becomes a serious problem due to the lack of control of pollutants from the source of production. Such sources include industrial and agricultural activities (EPA 1989). Every year, about 600 million tons of waste is produced by industrial activities, and this is growing at an annual rate of 10% (Ngo and Natowitz 2009). Contamination of soil and groundwater, at concentrations higher than the standard, will cause potential risk to health and the environment (Wuana and Okieimen 2011).

Soil remediation is done in different physical, chemical, and biological methods (Hamby 1996), and is divided into two major ex situ and in situ fields (Agarwal and Liu 2015). Soil vapor extraction is a physical in situ method for cleaning the volatile contaminants from the non-clay soils relying on their low evaporation temperature (EPA 2006). To improve the efficiency of the method, new initiatives have taken place and thermal enhancement is one of them (Air Force Center for Environmental Excellence 2001). Soil type and permeability, chemical

M. R. Sabour · S. H. Seyed Jalali (✉) · G. Dezvareh
Department of Civil Engineering, K.N. Toosi University of
Technology, No. 470, Mirdamad Blvd, Tehran 1996715433, Iran
e-mail: homamjalali@gmail.com

and physical properties of contaminants, environmental factors such as weather and hydrological conditions, pollutant volume, transfer ability, and toxicity are some factors affecting the level of soil contamination (Switzer and Kosson 2007; Truex et al. 2013).

In this study, the thermal enhancement method was used through injecting the water vapor and hot air. The reason for selecting these two ways is that these are the most common and widely used methods for thermal enhancement of vapor extraction in laboratory and industrial levels (Dablow et al. 2000; Di et al. 2002). Moreover, providing equipment and high temperatures is too expensive and with uncertainty (Beyke and Fleming 2005; Lowe et al. 2000).

Previously, some researches have been conducted on vapor extraction in laboratory scales and numerical forms. The laboratory researches include the efficiency of the SVE method in layered soils contaminated with TCE (Stinson 1989), the efficiency of this method in clay soils contaminated with thinner (Gibson et al. 1993), the influence of clay and temperature on remediation of sandy soil contaminated with diesel (Sabour et al. 2014), thermal remediation of light and heavy petroleum contaminations with electric heating (Beyke and Fleming 2005), and several other researches. Less attention has been paid to numerical study and modeling. Researches have been done to determine the mass transfer processes with the help of a three-dimensional model of vapor extraction (Nguyen et al. 2013). Vapor extraction data were also analyzed to determine the mass transfer limitations and predict mass flux from the source (Brusseau et al. 2010). In this study, a comprehensive model and formula has been presented for thermal enhancement removal rate in contaminated sandy soils.

2 Materials and Methods

Sandy soil was selected for modeling due to its having the highest cleaning efficiency and the lowest uncertainty (Albergaria et al. 2008; Peng et al. 2013). The contaminant used to obtain data is n-alkane (C_{13} – C_{19}) which is the most common range of volatile organic contaminant in petroleum products (Irwin et al. 2013; Li et al. 2016).

To obtain the data, the results of previous tests were used. These results are from the “Investigating the Influence of Clay & Heating in Remediation of Contaminated Soil from Gas-Oil by Soil Vapor Extraction” article (Sabour et al. 2014). For the validation part of

the modeling, data were extracted from the article “Predicting Hydrocarbon Removal from Thermally Enhanced Soil Vapor Extraction Systems” (Poppendieck et al. 1999). These data were extracted using chart analysis software 3D CAD CATIA, GRABIT toolbox in MATLAB, and GetData Graph Digitizer. A total amount of 67 packs of data were used for modeling. After screening and data processing, they were used for modeling. Screening was done using Plackett-Burman designs. This method helps to process data and minimize the errors. To apply this method, a Hadamard matrix in MATLAB was used.

After screening, the first model was created with MATLAB software with the help of a neural network for validation. Neural network modeling was chosen in view of the fact that since the relation between input data and objective function is not linear, a powerful and complex method is needed to understand this complicated relationship. Artificial neural network is a trainable method of modeling which consists of algorithms that can be used to perform nonlinear statistical modeling and has the ability to implicitly detect complex nonlinear relationships between dependent and independent variables, the ability to detect all possible interactions between predictor variables, and the availability of multiple training algorithms (Tu 1996).

Modeling was done using a feedforward backpropagation method and scaled conjugate gradient algorithm. This algorithm was used due to its having less dispersion and error compared to other algorithms. It is worth noting that the sigmoid function of the desired network was used due to the nature of the reviewed data.

After neural network modeling and data validation, the final model was analyzed and a removal formula was presented by using the response surface method (RSM) and modeling in historical data environment. According to Table 1 and regression coefficients obtained by data analysis, a second-order quadratic model was obtained for removal efficiency. In this equation, the response surface method considered the second-grade effects and interactions of tested variables in addition to linear effects of parameters.

3 Results and Discussion

In this study, by using a multilayer artificial neural network, the relation between time, temperature, and removal efficiency has been validated.

Table 1 The sum of the squares, df (degrees of freedom), mean square, *F* value, and *p* value for different modeling functions

Source	Sum of squares	df	Mean square	<i>F</i> value	<i>p</i> value	Probability > <i>F</i>
Mean vs total	142,138.6	1	142,138.6			
Linear vs mean	64,409.12	2	32,204.56	3533.469	<0.0001	
2FI vs linear	307.046	1	307.046	52.257	<0.0001	
Quadratic vs 2FI	12.410	2	6.205	1.057	0.351	
Cubic vs quadratic	267.636	4	66.909	22.087	<0.0001	Suggested
Residual	260.523	86	3.029342			
Total	207,395.3	96	2160.368			

As can be seen in Fig. 1, an artificial neural network with 30 hidden layers was used. In this network, two factors of time and temperature were used as input factors. Seventy percent of available data was used to train the network; 15% of data was used for validation and the rest for test.

According to the mean squared error curve (Fig. 2), the process of reducing the mean squared error (MSE) lost its downward trend after epoch 9 and then it increased. As a result, the process of network training for all three curves of training, validation, and test is in its best situation in stage 9.

The cease of the reduction error after stage 9 is due to the data screening before modeling. The data became uniform due to screening processes. The irrelevant data, which contains error, have been removed before entering the modeling process. Therefore, the model reached the best mode with the least errors after nine epochs.

In the neural network regression diagrams related to different aspects of training, validation, and test (Fig. 3), the accuracy of regression conducted in phases of training validation and test has been investigated. These correlation coefficients are equal to 0.96964, 0.98969, and 0.93453, respectively. This coefficient is equal to 0.96653 for the overall state of the model. By comparing the correlation coefficients, two points can be realized. First, the correlation coefficient at each stage is close to 1. Therefore, the model is accurately in compliance with reality. Second, the proximity of the coefficients in

various stages of modeling indicates that the model has followed a consistent trend at all stages of its implementation and no error has occurred. Therefore, the obtained values by the model are close to true values with good approximation.

The values of RMSE, MAE, and *R* indicate root mean square error, mean absolute error, and correlation coefficient, respectively. Their formulas are as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{1}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| = \frac{1}{n} \sum_{i=1}^n |e_i| \tag{2}$$

$$R = \sqrt{1 - \frac{\sum_{i=1}^n (y_i - f_i)^2}{\sum_{i=1}^n \left(y_i - \frac{1}{n} \sum_{i=1}^n y_i \right)^2}} \tag{3}$$

The errors related to the model are shown in Table 2. The lesser correlation coefficient in the test mode compared to the training and validation modes is due to the application of a variety of tests in this stage to check the quality of the model. The lesser correlation coefficient in

Fig. 1 The structure of the neural network for estimation

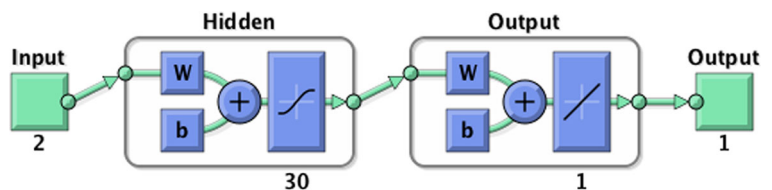
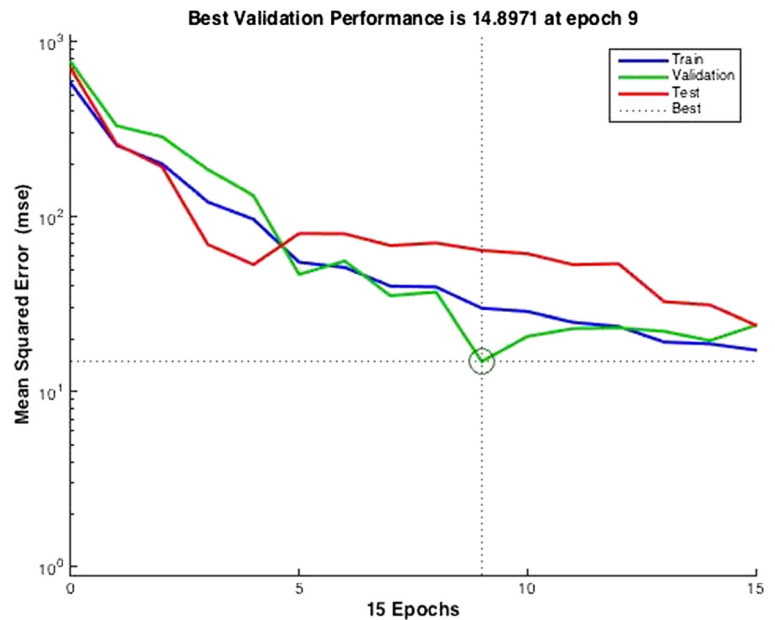


Fig. 2 Mean squared error curve



the training mode compared to the validation mode is due to large data volume and lack of network training. However, the proximity of the correlation coefficients and the square of the correlation coefficients to number 1 and each other shows the proper performance of the model.

As can be seen, the neural network model based on experimental data has the least error and acceptable correlation coefficient. The low error rate and the difference between predicted and actual data is due to the measurement errors in the laboratory, uncertainties related to contaminants, and the movement and accumulation of contaminants in the soil matrix.

After validation by the neural network model, another modeling was performed using response surface methodology. In this analysis, the obtained results of all the effects governing the issue including the first-grade effects, the second-grade effects, and interaction of two influencing factors were considered to reach an overall judgment about the significance of factors.

In Fig. 4, the process of efficiency changes with respect to time is shown. As can be seen, with the passage of time, the efficiency increased. The reason for this increase is that, with the passage of time, the contact area between the air and water vapor and contaminant increases. Consequently, the vapor pressure increases and the contaminant state changing from liquid to gas increases.

In Fig. 5, the process of efficiency changes with respect to temperature is shown. In this diagram, as expected, the removal rate increased with increasing of the temperature. Obviously, this increase is due to increase of vapor pressure and the proximity of the injected vapor and air temperature to the evaporation temperature of volatiles.

For better understanding, in Fig. 6, the actual values and the values provided by the model are shown in a diagram. In this diagram, the vertical axis shows the values predicted by the model and the horizontal axis indicates the values observed in the tests. Since the data concentration is on the fitted line, this means that values estimated by the model are close to laboratory values. The proximity of the obtained points to the axis $Y = X$ shows high adequacy of the model.

In order to check the validity of the model, the correlation coefficient (R^2), analysis of variance is investigated. In this section, the model competency is approved given the amount of p value < 0.05 . The possible value for each parameter can be seen. The results of the variance analysis in investigating the removal rate are shown in Table 3.

As can be seen, the possibility of a model regression equation is less than 0.0001 and smaller than 0.05 (considering the confidence level of 0.95). This shows that the second-order quadratic model complies with laboratory results. Regarding the regression equation, the low rate of p value ($\alpha \leq 0.5$) shows that the results

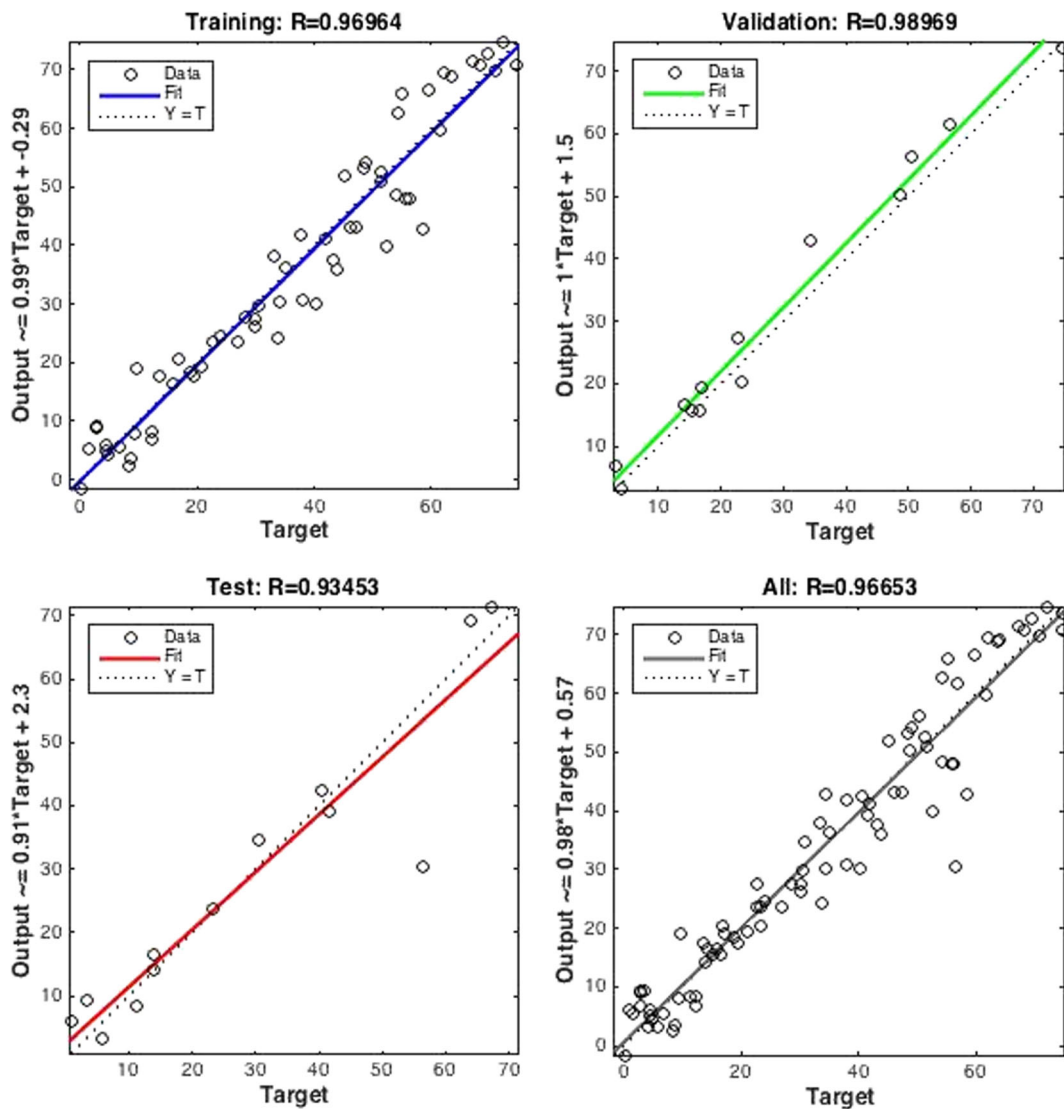


Fig. 3 Neural network regression diagram

are not random. The amount of p value (less than 0.5) indicates that the parameters are significant. According to the table above, the F value (test statistic) is equal to 2205.991. Therefore, it confirms the competency of the

Table 2 Calculated errors for neural network model

	MSE	RMSE	R	R^2	MAE
Training	29.973	5.475	0.970	0.940	–
Validation	14.897	3.859	0.990	0.979	–
Test	64.212	8.013	0.935	0.873	–
All	–	5.730	0.967	0.931	4.256

model. It also indicates that the model is significant and there is the possibility of a slight incompetence of the model. The F value for lack of fit is equal to 3.208121. This amount is appropriate since the p value is equal to 0.0953. This shows that the possibility of incompetence occurrence is 0.09%. According to the results of the data analysis in the variance analysis table, each of the factors and their interactions affect the test results.

The correlation coefficient R^2 is an indicator to express the precision of regression and relation between the experimental data and predicted responses. High level of correlation coefficient (close to 1) is desirable. High level of R^2 indicates that, in fitted models, the

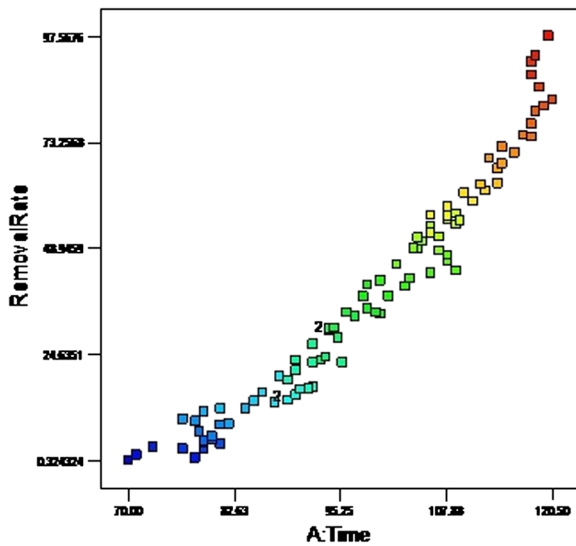


Fig. 4 Efficiency changes with respect to time

predicted and experimental responses have adapted well to each other. In results analysis, the correlation coefficients are as below (Table 4):

Figure 7 is related to two-dimensional contour, and Fig. 8 indicates the three-dimensional surfaces of the model. As shown, the pollutant removal efficiency increases as temperatures and time increase. According to the contour and surface diagram, it can be realized that the removal rate is more sensitive to temperature factor. Actually, temperature is more effective in the removal rate compared to time. Therefore, higher removal efficiency can be achieved by increasing the temperature

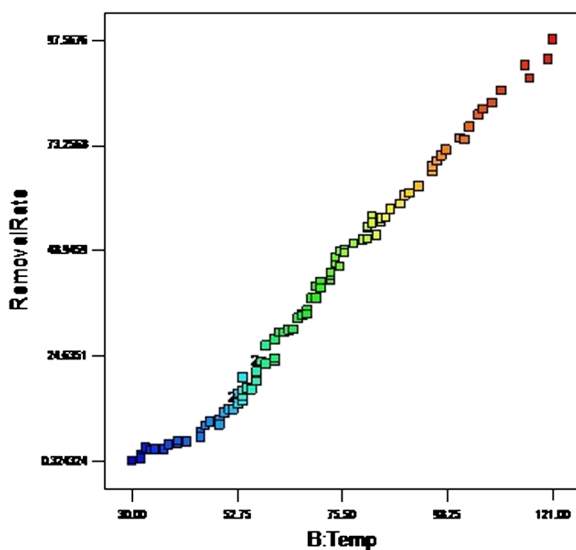


Fig. 5 Efficiency changes with respect to temperature

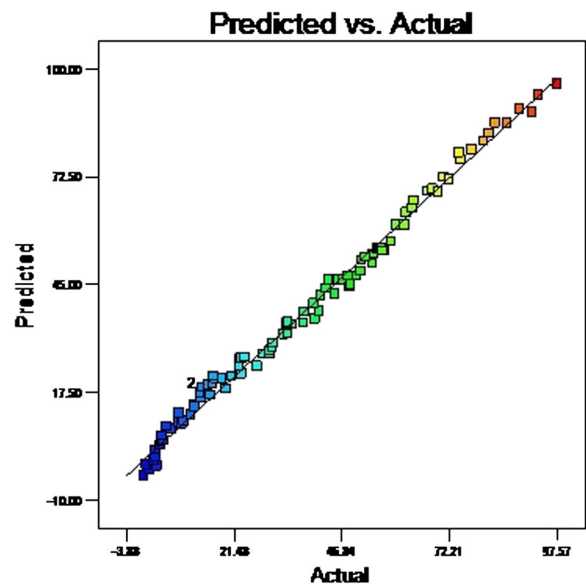


Fig. 6 Predicted and observed values in the model

(compared to time). In surfaces, it can be observed that the surface towards the axis of temperature is slightly curved. This reflects the greater impact of temperature on the removal efficiency.

Based on the mentioned points, the second-order polynomial overall model is as follows:

$$\begin{aligned} \text{Removal rate} = & -5.70387 - 0.4284 \times \text{Time} \\ & + 0.0099119 \times \text{Temp} + 0.009613 \\ & \times \text{Time} \times \text{Temp} + 0.002131 \\ & \times \text{Time}^2 - 0.00199 \times \text{Temp}^2 \quad (4) \end{aligned}$$

After simplifying and eliminating the parameters with lower effect coefficient, with the assumption that t represents the time measured by day passes from the start of the extraction, T represents the temperature measured by Celsius scale, and “R.R” represents the removal rate, the formula will be as follows:

$$\begin{aligned} \text{R.R} = & -5.703 - 0.428t + 0.01T(1 + t) \\ & + 0.002(t^2 - T^2) \quad t \geq 45, T \geq 65 \quad (5) \end{aligned}$$

4 Conclusion

In this study, the remediation process of the sandy soils contaminated with n-alkanes (C_{13} – C_{19}) has been

Table 3 Analysis of variance (ANOVA)

Source	Sum of squares	df	Mean square	F value	p value	Probability > F
Model	64,728.57	5	12,945.71	2205.991	<0.0001	Significant
A—time	3.937	1	3.937	0.671	0.415	
B—temp.	3.433	1	3.433	0.585	0.446	
AB	2.174	1	2.174	0.370	0.544	
A ²	0.149	1	0.149	0.025	0.874	
B ²	1.021	1	1.021	0.174	0.678	
Residual	528.159	90	5.868			
Lack of fit	518.649	85	6.102	3.208	0.095	Not significant
Pure error	9.510	5	1.902			
Cor total	65,256.73	95				

modeled based on the thermal optimization of vapor extraction through injecting the water vapor and pumping hot air. A neural network with 30 hidden layers and 15 epochs has been obtained to evaluate the data. This neural network provided a good and accurate prediction of time and removal efficiency.

The aim of the neural network was to validate data to prevent over-fitting. For this purpose, the existing data were divided into three categories: training, validation, and test. The mean error is equal to 4.2556, and the root mean square error is equal to 5.7303. It indicates the slight dispersion of modeling values around the actual values. The low rate of mean absolute error and root

mean square error indicates that the error rate is very low. It also shows that the observed values are very close to the predicted values. The high rate of the correlation coefficient and R^2 confirms the same issue. Moreover, the removal rate presented by the model is equal to laboratory and actual values with a slight error. The general correlation coefficient of the model is equal to 0.967, and the correlation coefficient square is approximately 0.931. This is very close to 1 and indicates the proximity of the actual and calculated values.

According to the contour and surface diagrams and response surface diagrams related to parameter interaction, it can be concluded that, in all interactions, the removal rate increases by increasing the temperature

Table 4 Error rates

Std. dev. ^a	2.422	R-squared	0.992
Mean	38.479	Adj. R-squared ^b	0.991
C.V. % ^c	6.296	Pred R-squared ^d	0.991
PRESS ^e	595.273	Adeq. precision ^f	164.917

^a Root mean square error

^b Adjusted R^2 (a type of R^2 including the number of factors of the model)

^c The coefficient of variation (rate of changes not explained or residual changes of data as a percentage of average response variable)

^d Predicted R -squared

^e The sum of squared prediction error (a measure for predicting the responses to a new test)

^f Adequacy of precision (derived by dividing the difference between the largest and smallest predicted responses by the average standard deviation of expected responses); the higher values of this statistic are more ideal, and values higher than 4 usually show a good performance of the model in the prediction phase. Number 164.917 indicates the proper design and lack of disturbance

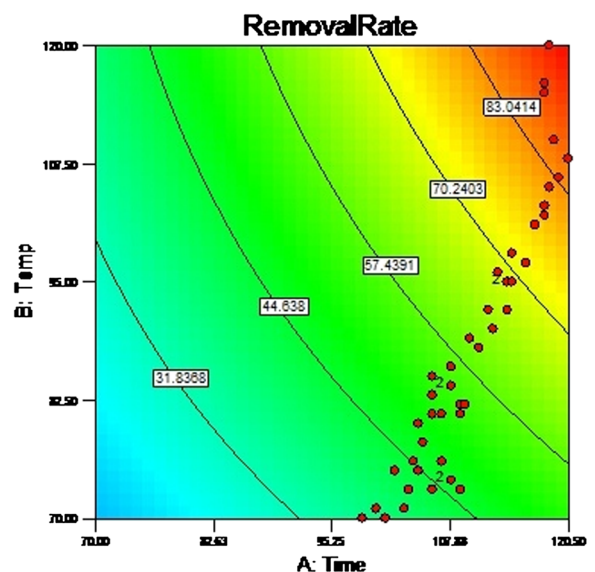


Fig. 7 Contour diagram of temperature and time reaction on the removal efficiency

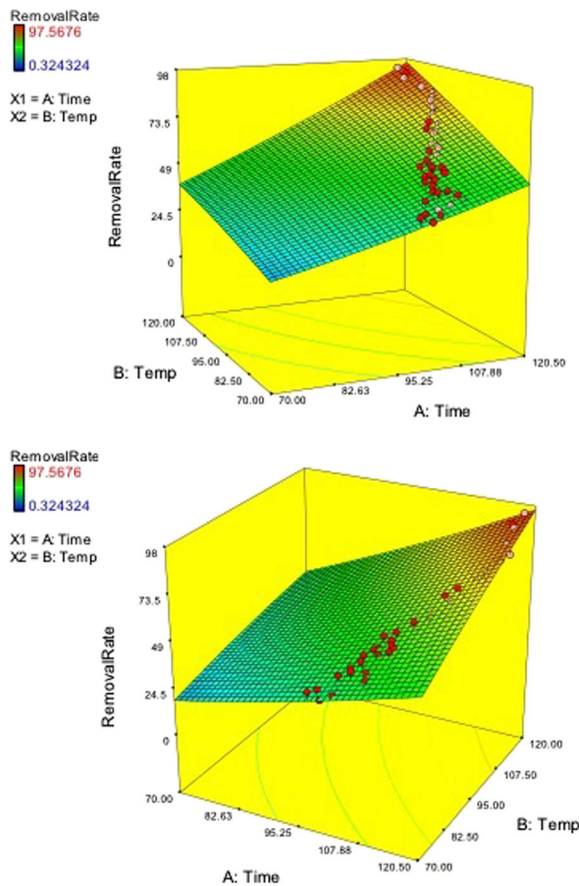


Fig. 8 The surface describing the temperature and time reaction on the removal efficiency

and time. By increasing the temperature, the vapor pressure of volatile organic compounds in the soil increases. Therefore, evaporation and state changing from liquid to gas increase in unit of time. Time increases the duration of contact between air and vapor pumped to the injection wells and contamination. Therefore, it increases the contaminant removal rate. Consequently, this model can be used to predict the optimization rate of cleaning sandy soils through the vapor extraction method.

The effect of time factor on removal efficiency rate will have a more severe slope after about 80 days; therefore, it will be more effective. The reason is that, after passing this period from the beginning of the removal, the vapor pressure of volatile substances will be so that the state change will occur and the volatile substances will be extracted. For that reason, after 80 days, the removal rate and efficiency will increase.

The slope of the effect of temperature on the removal of contaminant will decrease after almost 100 °C. The reason is that, as temperatures rise over 100 °C, the soil moisture evaporates and a significant amount of moisture in the soil is absorbed. Consequently, the contaminants that need to be solved in the water will not be absorbed and removed; hence, the removal efficiency will be reduced.

References

- Agarwal, A., & Liu, Y. (2015). Remediation technologies for oil-contaminated sediments. *Marine Pollution Bulletin*, 101(2), 483–490. doi:10.1016/j.marpolbul.2015.09.010.
- Air Force Center for Environmental Excellence. (2001). Environmental restoration program guidance on soil vapor extraction optimization.
- Albergaria, J. T., da Alvim-Ferraz, M. C. M., & Delerue-Matos, C. (2008). Soil vapor extraction in sandy soils: influence of airflow rate. *Chemosphere*, 73, 1557–1561. doi:10.1016/j.chemosphere.2008.07.080.
- Beyke, G., & Fleming, D. (2005). In situ thermal remediation of DNAPL and LNAPL using electrical resistance heating. *Remediation Journal*, 15(3), 5–22. doi:10.1002/rem.20047.
- Brusseau, M. L., Rohay, V., & Truex, M. J. (2010). Analysis of soil vapor extraction data to evaluate mass-transfer constraints and estimate source-zone mass flux. *Ground Water Monitoring and Remediation*, 30(3), 57–64. doi:10.1111/j.1745-6592.2010.01286.x.
- Dablow, J. F., Pearce, J. A., Johnson, P. C., Balshaw-Biddle, K., Oubre, C. L., & Ward, C. H. (2000). *Steam and electroheating remediation of tight soils*. Boca Raton: Lewis Publishers.
- Di, P., Chang, D. P. Y., & Dwyer, H. A. (2002). Modeling of polychlorinated biphenyl removal from contaminated soil using steam. *Environmental Science and Technology*, 36(8), 1845–1850. doi:10.1021/es010739o.
- EPA. (1989). *Risk assessment guidance for superfund* (Vol. I). Washington, DC.
- EPA. (2006). In situ treatment technologies for contaminated soil. *Epa*, pp. 15–20.
- Gibson, T. L., Abdul, A. S., Glasson, W. A., Ang, C. C., & Gatlin, D. W. (1993). Vapor extraction of volatile organic compounds from clay soil: a long-term field pilot study. *Ground Water*, 31(4), 616–626. doi:10.1111/j.1745-6584.1993.tb00595.x.
- Hamby, D. M. (1996). Site remediation techniques supporting environmental restoration activities—a review. *Science of the Total Environment*, 191(3), 203–224. doi:10.1016/S0048-9697(96)05264-3.
- Irwin, R. J., Mouwerik, M. V., Stevens, L., Seese, D. M., & Basham, W. (2013). Environmental contaminants encyclopedia. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699. doi:10.1017/CBO9781107415324.004.
- Li, Y., Zhang, R., Wang, T., Wang, Y., Xu, T., Li, L., et al. (2016). Determination of n-alkanes contamination in soil samples by

- micro gas chromatography functionalized by multi-walled carbon nanotubes. *Chemosphere*, 158, 154–162. doi:10.1016/j.chemosphere.2016.05.068.
- Lowe, D. F., Oubre, C. L., & Herb Ward, C. (2000). Soil vapor extraction using radio frequency heating: resource manual and technology demonstration. *Journal of Hazardous Materials*, 74(3), 218–219. doi:10.1016/S0304-3894(00)00170-9.
- Ngo, C., & Natowitz, J. (2009). *Our energy future: resources, alternatives and the environment*. Wiley survival guides in engineering and science (2nd Editio.). <http://www.dawsonera.com/depp/athens?url=http://www.dawsonera.com/depp/reader/protected/external/AbstractView/S9780470473788>. Accessed 5 Nov 2016.
- Nguyen, V. T., Zhao, L., & Zytner, R. G. (2013). Three-dimensional numerical model for soil vapor extraction. *Journal of Contaminant Hydrology*, 147, 82–95. doi:10.1016/j.jconhyd.2013.02.008.
- Peng, S., Wang, N., & Chen, J. (2013). Steam and air co-injection in removing residual TCE in unsaturated layered sandy porous media. *Journal of Contaminant Hydrology*, 153, 24–36. doi:10.1016/j.jconhyd.2013.07.002.
- Poppendieck, D. G., Loehr, R. C., & Webster, M. T. (1999). Predicting hydrocarbon removal from thermally enhanced soil vapor extraction systems: 2. Field study. *Journal of Hazardous Materials*, 69(1), 95–109. doi:10.1016/S0304-3894(99)00063-1.
- Sabour, M., Rezaei, B., & Jafargholi, A. (2014). Investigating the influence of clay and heating in remediation of contaminated soil from gas-oil by soil vapor extraction. *Journal of Civil Engineering (Journal of School of Engineering)*, 25(2), 53–65.
- Stinson, M. (1989). Epa site demonstration of the Terra Vac in situ vacuum extraction process in Groveland, Massachusetts.
- Switzer, C., & Kosson, D. S. (2007). Soil vapor extraction performance in layered vadose zone materials. *Vadose Zone Journal*, 6(2), 397. doi:10.2136/vzj2005.0131.
- Truex, M., Becker, D., Simon, M., Oostrom, M., Rice, A., & Johnson, C. (2013). Soil vapor extraction system optimization, transition, and closure guidance (PNNL-21843).
- Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology*, 49(11), 1225–1231. doi:10.1016/S0895-4356(96)00002-9.
- Wuana, R. A., & Okieimen, F. E. (2011). Heavy metals in contaminated soils: a review of sources, chemistry, risks and best available strategies for remediation. *ISRN Ecology*, 2011, 1–20. doi:10.5402/2011/402647.