

Article

Development on Surrogate Models for Predicting Plume Evolution Features of Groundwater Contamination with Natural Attenuation

Yajing Wang ¹, Mingyu Wang ^{1,*} and Runfeng Liu ²

¹ College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 101408, China; wangyajing18@mailsucas.ac.cn

² College of Mining, Liaoning Technical University, Fuxin 123000, China; llmkongjianguo@163.com

* Correspondence: mwang@ucas.ac.cn

Abstract: Predicting the key plume evolution features of groundwater contamination are crucial for assessing uncertainty in contamination control and remediation, while implementing detailed complex numerical models for a large number of scenario simulations is time-consuming and sometimes even impossible. This work develops surrogate models with an effective and practicable pathway for predicting the key plume evolution features, such as the distance of maximum plume spreading, of groundwater contamination with natural attenuation. The representative various scenarios of the input parameter combinations were effectively generated by the orthogonal experiment method and the corresponding numerical simulations were performed by the reliable Groundwater Modeling System. The PSO-SVM surrogate models were first developed, and the accuracy was gradually enhanced from 0.5 to 0.9 under a multi-objective condition by effectively increasing the sample data size from 54 sets to 78 sets and decreasing the input variables from 25 of all the considered parameters to a smaller number of the key controlling factors. The statistical surrogate models were also constructed with the fitting degree all above 0.85. The achieved findings provide effective generic surrogate models along with a scientific basis and investigation approach reference for the environmental risk management and remediation of groundwater contamination, particularly with limited data.

Keywords: groundwater contamination plume; natural attenuation; key plume features; surrogate models; GMS numerical simulations; prediction uncertainty; support vector machine; multiple regression



Citation: Wang, Y.; Wang, M.; Liu, R. Development on Surrogate Models for Predicting Plume Evolution Features of Groundwater Contamination with Natural Attenuation. *Water* **2024**, *16*, 2861. <https://doi.org/10.3390/w16192861>

Academic Editor: Jianjun Ni

Received: 12 March 2024

Revised: 18 April 2024

Accepted: 22 April 2024

Published: 9 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, the intensification of national industrial evolution and urban expansion has markedly exacerbated groundwater contamination, which may include various anthropogenic sources such as mining, smelting, and industrial activities [1,2]. Organic contaminants seep through various channels [3–5]. Particularly, some Southeast Asian countries are affected by the deterioration of their water environment [6,7]. Risk assessment studies has been widely applied to water environments [8]. However, the assessment of groundwater contamination risk necessitates a thorough evaluation of contributing factors to control the risk within tolerable limits, thereby facilitating the achievement of site decontamination objectives [9,10]. Natural attenuation plays a vital role in environmental restoration by leveraging processes like the convective dispersion of contaminants in groundwater and the adsorptive and degradative capacities of chemical and biological agents, thus diminishing contaminant levels [11]. Distinguished by its cost-effectiveness and minimal environmental impact, monitored natural attenuation (MNA) emerges as the foremost strategy in contemporary groundwater contamination risk management [12,13]. Accordingly, the research into predicting the plume evolution features of organic con-

taminants in groundwater is crucial for assessing contamination levels, refining control measures, and fostering the sustainable use of groundwater resources.

Early investigations into the natural attenuation of groundwater at contaminated sites concentrated on understanding the processes and mechanisms affecting individual contaminants. Recent studies, however, have shifted towards a quantitative analysis and predictive modeling of natural attenuation processes. Groundwater Modeling System (GMS), an advanced three-dimensional numerical modeling tool for groundwater analysis, has gained widespread and long-period adoption for assessments and predictive studies in groundwater research, providing correct simulation results with reasonable parameter settings [14,15]. Valivand and Katibrh employed GMS to develop a 3D predictive model for nitrate contamination in the groundwater of the Uruguay Plain, enhancing the precision of numerical groundwater simulations and offering a reference for further research on inorganic salt solutes in aquatic environments [16]. Similarly, Tong Xiaoxia et al. utilized GMS to identify key contaminants from a landfill site to evaluate their impact on the surrounding soil and groundwater. Their findings, derived from numerical simulations, offered crucial data and a scientific framework for mitigating water and soil pollution in landfill settings [17]. Gao Qifeng et al. applied GMS to create a model for water flow and solute transport within an industrial park, assessing the potential effects of contaminants from sewage treatment facilities on the groundwater under abnormal conditions [18]. Notably, current research often focuses on simulating the transport and fate of specific contaminants without assessing the comprehensive contamination scenarios and their future tendency. This study leverages GMS 10.4 version software simulations to statistically analyze the evolution of contamination plumes under natural attenuation conditions, and three types of data can be obtained from every simulation experiment, the distance of maximum plume spreading (DMPS), time to reach the DMPS (T-DMPS), and the mean concentration within the DMPS (MC-DMPS), which are selected to be the key features to represent every simulation scenario. These critical features are pivotal for understanding the natural attenuation process, offering vital theoretical insights and scientific contributions to the decision-making process in managing groundwater contamination risks at contaminated sites.

Surrogate models summarize uncertain or complex input–output relations through reasonable and accurate simple functions, which make an effort to reduce computational load for various large-scale complicated system research studies [19,20]. In the last few years, the continuous progression in computer technology and statistical methodologies has led to the widespread application of various statistical methods and intelligent algorithms, including multiple regression, support vector machines (SVMs), and particle swarm optimization (PSO), in environmental modeling [21–24]. For instance, Aradhi et al. incorporated a range of influential factors such as industrial activities, land use, and meteorological conditions, employing a multiple linear regression model to forecast pollution levels in the groundwater of an industrial area [25]. Muhammad et al. utilized several SVM methods for the classification and recycling application in automatic systems, elucidating the types of dry waste on solid waste and showcasing the proficiency of SVMs in complex classification challenges [26], while the optimization of kernel functions and their accompanying parameters for the SVM algorithm were not given in detail. Gai Rongli, Pang Xi, and Xu Genqi et al., by analyzing the characteristics of environmental sample data and the shortcomings of existing optimization methods, proposed the PSO-SVM algorithm to conduct research on water quality assessments by calculating the weight coefficient to determine the four main parameters with 100 sets of training data and 20 sets of testing data, air quality classification with 150 sets of training data and 215 sets of testing data after normalization processes under only one dimension, and debris flow disaster prediction with six primary inputs and 1000 sets of monitoring data. The reliability and effectiveness of the PSO-SVM model in environmental research studies with a certain number of sample data with multi-dimensional features were investigated [27–29]. However, there is still a lack of relevant further research on PSO-SVM modelling in groundwater environments.

Wang Bilian et al. applied GMS for simulating the transport and fate from a constant pollution source within an aquifer, subsequently establishing a quantitative relationship between the characteristic variables and the related controlling factors through multiple regression analysis [30]. Nevertheless, their research did not sufficiently account for the potential influence of contamination source degradability and the characteristics of the aquitard and the confined aquifers on the plume evolution of the contaminated aquifer. Consequently, this study therefore adopts the PSO-SVM approach, takes organic contaminants as its pollution sources, and selects a more comprehensive set of factors potentially influencing the natural attenuation of groundwater contamination as input variables to conduct a further investigation. The DMPS, T-DMPS, and MC-DMPS at this juncture are employed as output variables. This methodology endeavors to develop a surrogate model for predicting the key features of the organic contamination plume evolution in groundwater environments with natural attenuation, incorporating a wider array of factors to enhance the understanding of the contaminant transport and fate dynamics.

Since the surrogate models to be developed are aimed to apply to typical groundwater systems, this paper comprehensively considers a wide range of hydrogeological and geochemical parameters that may affect the natural attenuation of the contaminant plumes. In addition, the representative various scenarios of groundwater contamination plume evolution are effectively generated by the orthogonal experiment method, and the corresponding numerical simulations are performed by the reliable Groundwater Modeling System in order to achieve the reliable basic datasets for the establishment of the surrogate models. The PSO-SVM surrogate models are first developed, and the statistical surrogate models are also constructed by multiple regression based on the same dataset used for the PSO-SVM model. Furthermore, this research endeavors to investigate model precision by enlarging the sample size and changing the input factors. A comparative evaluation between PSO-SVM models and statistical regression models is performed to offer an observation on the difference between the two types of surrogate models as to prediction accuracy performance.

2. Materials and Methods

2.1. Conceptual and Baseline Models for Constructing Different Representative Simulation Scenarios

The conceptual model for constructing different simulation scenarios was established based on a typical contaminated area which is a chemical industrial cluster in the Shandong Yellow River alluvial plain. The average annual rainfall is approximately 700 mm, and the average annual evaporation is around 1650 mm. The groundwater flows from west to east, the overall topography is flat, and the annual groundwater level change does not exceed 2 m. The region has a wide distribution and diverse types of contaminants, the concentration of part of which exceeds the corresponding water quality standards, mainly dominated by organic contaminants such as benzene and toluene.

By taking the site conditions as references, the conceptual model for this investigation was established as showed in Figure 1. The groundwater system basically consisted of, from top to bottom, the phreatic aquifer, aquitard, confined aquifer, and impermeable bottom layer. The modeling area was 7000 m × 8000 m and had a thickness of 50 m, and the corresponding grading was taken with meeting sufficient numerical computation accuracy. The hydraulic property of the groundwater system was assumed to be homogeneous and isotropous for each layer, with the fixed head boundaries on east and west sides and the specific head differences between the boundaries. In addition, groundwater contamination plume developed with a degraded contamination source in the phreatic aquifer. Furthermore, the contaminant was an easily degraded and dissolved organic compound represented by the benzene series. The contaminant transport and fate processes included advection, hydrodynamic dispersion, adsorption, and degradation. The permeability coefficient and other parameters in the baseline model were set based on the actual groundwater system, parameters, and empirical values in the investigated area, as shown in Table 1.

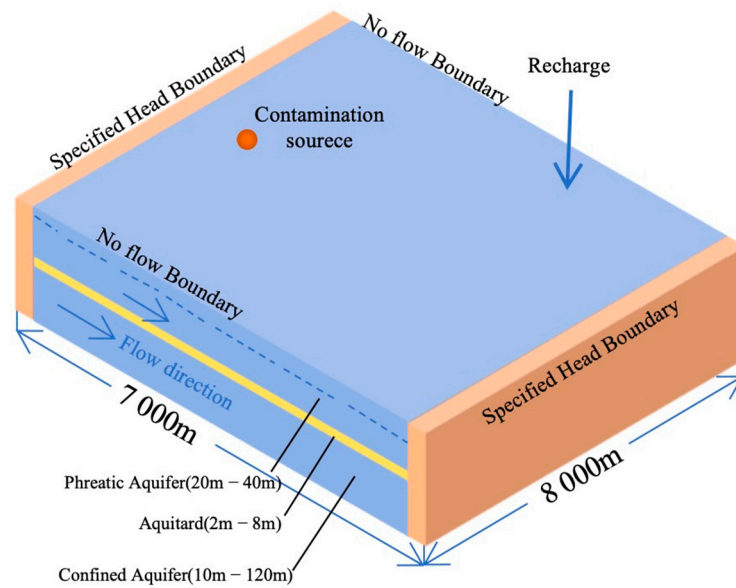


Figure 1. Conceptual model diagram.

A baseline model was first established based on the investigated area for constructing the different representative simulation scenarios used for the development of the surrogate models. For the preliminary examination, typical numerical simulation results were obtained from the baseline model, shown in Figure 2. The total simulation period was set to be 40 years with a set of outputs every 180 days. In Figure 2, taking benzene as the contaminant of interest, the concentration cut-off value was set for the plume frontier as 0.01 mg/L [31].

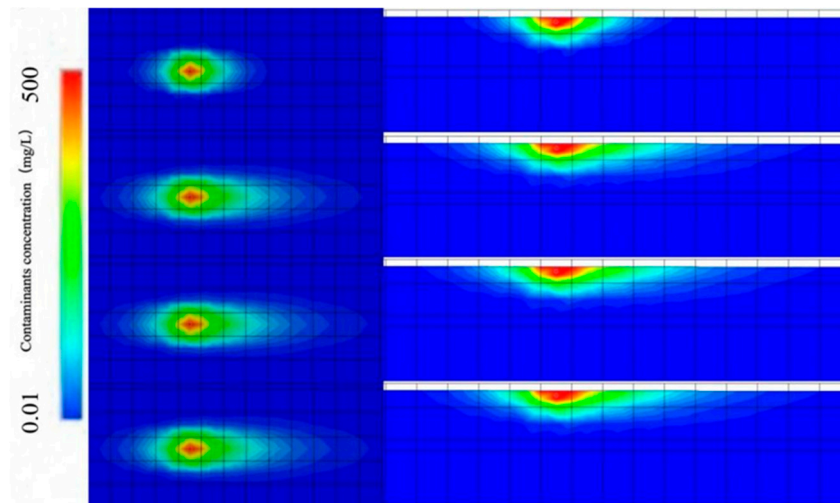


Figure 2. Natural attenuation of contamination plumes at 360 d, 3600 d, 7380 d, and 14,600 d of the investigated site.

Table 1. Baseline model parameters of GMS numerical simulations from the investigated site.

Parameters	Values	Parameters	Values
Heads between Upstream and Downstream Boundaries ΔH_h (m)	8	Effective Porosity P_1	0.25
Heads between Phreatic Aquifer and Confined Aquifer ΔH_v (m)	1	Average Thickness M_1 (m)	20
Recharge Rate R (m/d)	0.0002	Phreatic Aquifer Permeability Coefficient Kw_1 (m/d)	50
Concentration of Source Contamination C_0 (mg/L)	500	Adsorption Coefficient Ka_1 (m^3/kg)	0.0005
Area of Source Contamination A (m^2)	900	Degradation Coefficient Kd_1 (1/d)	0.003

Table 1. Cont.

Parameters		Values	Parameters		Values
Ratio of Source Contamination Thickness to Phreatic Aquifer		0.25	Dispersion D_1 (m)		60
Aquitard	Effective Porosity P_2	0.15	Confined Aquifer	Effective Porosity P_3	0.25
	Average Thickness M_2 (m)	5		Average Thickness M_3 (m)	25
	Permeability Coefficient Kw_2 (m/d)	0.008		Permeability Coefficient Kw_3 (m/d)	50
	Adsorption Coefficient Ka_2 (m^3/kg)	0.0005		Adsorption Coefficient Ka_3 (m^3/kg)	0.0005
	Degradation Coefficient Kd_2 (1/d)	0.003		Degradation Coefficient Kd_3 (1/d)	0.003
Dispersion D_2 (m)		0.4	Dispersion D_3 (m)		60

2.2. Orthogonal Experiment

The orthogonal experiment, as a statistical method, is frequently used to address experimental design issues with multiple factors and levels. Its main idea is to analyze a portion of representative experimental results in order to understand the overall experimental situations [32]. By utilizing a regular orthogonal array design, representative level combinations are selected for all factors and level combinations to feature high efficiency, high speed, and economy [33].

A total of 25 parameters that impact the natural attenuation of groundwater contamination plume were selected, including the potential degradation coefficient of the contaminant source itself. Due to a multitude of parameters, an L54 ($2^1, 3^{24}$) orthogonal array was adopted to design the experimental plans according to the design rules for orthogonal experiments, with each parameter set to three levels beside the porosity of aquitard with two levels. Values for each parameter were assigned sequentially from low to high based on their possible ranges. The range of the degradation and adsorption coefficients referred to the BIOSCREEN user manual for contamination plume simulation software [34], with the maximum source concentration set above the maximum solubility of benzene contaminants (Toluene 25 °C–542 mg/L). The DMPS was calculated by counting the number of grid points with concentrations more than 0.01 mg/L in the phreatic aquifer when the contamination plume reached the farthest distance. The T-DMPS was based on the time required for the DMPS. The MC-DMPS represented the contaminant quantity per unit volume of water within the contaminated range in the phreatic aquifer. The different value levels of the 25 parameters are presented in Table 2.

Table 2. Different value levels of each parameter set.

Level	ΔH_h (m)	ΔH_v (m)	R (m/d)	Cs (1/Y)	C_0 (mg/L)	A (900 m^2)	r
1	4	1	0.0001	0.1	100	1	0.1
2	8	3	0.0002	1	300	25	0.5
3	12	5	0.0004	10	800	100	1
Phreatic Aquifer							
Level	P_1	M_1 (m)	Kw_1 (m/d)	Ka_1 (m^3/kg)	Kd_1 (1/Y)	D_1	
1	0.15	20	10	0.0001	0.1	10	
2	0.25	30	50	0.0005	1	30	
3	0.35	40	100	0.001	10	60	
Aquitard							
Level	P_2	M_2 (m)	Kw_2 (m/d)	Ka_2 (m^3/kg)	Kd_2 (1/Y)	D_2	
1	0.1	2	0.005	0.0001	0.1	0.2	
2	0.15	5	0.01	0.0005	1	0.4	
3	0.25	8	0.03	0.001	10	0.6	

Table 2. Cont.

Level	P ₃	M ₃ (m)	Confined Aquifer			D ₃
			Kw ₃ (m/d)	Ka ₃ (m ³ /kg)	Kd ₃ (1/Y)	
1	0.15	10	10	0.0001	0.1	10
2	0.25	50	60	0.0005	1	30
3	0.35	120	120	0.001	10	60

Notes: A total of 25 parameters that may influence the numerical simulation experiment of groundwater contamination were selected, which include hydrogeological, hydrochemical and several specific parameters associated with the contamination source, where ΔH_h(m) is the heads between upstream and downstream boundaries; ΔH_v(m) is the heads between phreatic aquifer and confined aquifer; R (m/d) is the recharge rate; Cs (1/Y) is the degradation coefficient of source contamination; C₀ (mg/L) is the concentration of source contamination; A (m²) is the area of source contamination; r is the ratio of source contamination thickness to phreatic aquifer; P is the effective porosity; M (m) is the average thickness; Kw (m/d) is the permeability coefficient; Ka (m³/kg) is the adsorption coefficient; Kd (1/Y) is the degradation coefficient; and D (m) is the dispersion.

2.3. Support Vector Machine (SVM)

The SVM approach was established based on the VC (Vapnik Chervonenkis) dimension theory of statistical learning and the principle of minimizing structural risk. It seeks optimal results between the complexity of the model and its learning ability using limited sample information [35]. Initially introduced with the goal of finding the optimal hyperplane in the sample space, maximizing the distance between samples and the hyperplane to enhance model generalization, SVM has evolved into a prediction method employed in various classification models [36].

The SVM algorithm determines the hyperplane for classifying data points from different categories, and this hyperplane serves as a decision boundary. According to the binary classification mechanism of SVM, data points are divided into two classes defined as 0 and 1 [37]. Thus, when classifying new data points, their category can be determined based on their position relative to the hyperplane, achieving effective predictive outcomes.

For a training set, $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i)\} \in (X \times Y)^l$, where the sample size is l , and $x_i \in X \in \mathbf{R}^n, y_i \in Y \in \mathbf{R}, i = 1, 2, \dots, l$. Assuming $f(x) = \omega \cdot x + b$, there will be errors in the prediction results inevitably represented by ζ_i , and taking ε as the error insensitive function, then the SVM can be represented as follows:

$$\min_{\omega, b, \zeta^*} \tau(\omega, b, \zeta^*) = \frac{1}{2} \|\omega\|^2 + \frac{C}{l} \sum_{i=1}^l (\zeta_i + \zeta_i^*)$$

$$\text{s.t.} \begin{cases} y_i - (\omega \cdot x_i + b) \leq \varepsilon + \zeta_i \\ (\omega \cdot x_i + b) - y_i \leq \varepsilon + \zeta_i^* \\ \zeta_i^* \geq 0 \end{cases} \quad (1)$$

where C and ε are both positive real numbers, representing the penalty factors and fitting accuracy of the control parameters respectively.

When dealing with linearly inseparable data, it is common to use kernel functions to map the features of data samples into high-dimensional space. Choosing the correct and appropriate kernel function can avoid directly computing complex calculations in high-dimensional space, thereby reflecting the results of the classification computed in low-dimensional space onto high-dimensional space. The selection of kernel functions in SVM significantly affects the predictive outcomes. Currently, commonly used kernel functions include the Sigmoid kernel function, the radial basis function (RBF), the polynomial kernel function, and the linear kernel function. Research has shown that the RBF kernel function possesses strong learning capabilities and is widely applied. Therefore, this paper selects the following RBF kernel function:

$$k(x_1, x_2) = \exp(-g \|x_i - x_j\|^2) \quad (2)$$

where g is the kernel parameter.

By employing the Lagrangian method, the optimization problem is transformed into its dual form:

$$\begin{aligned} & \min \left[\frac{1}{2} \alpha^T Q \alpha - e^T \alpha \right] \\ & \text{s.t.} \begin{cases} y^T \alpha = 0 \\ 0 \leq \alpha_i \leq C, i = 1, \dots, n \end{cases} \end{aligned} \quad (3)$$

where α is the lagrange multiplier, representing the importance of each support vector.

Finally, the decision function is obtained [19].

$$\text{sgn} \left[\sum_{i=1}^l y_i \alpha_i k(x_i, x) + b^* \right] \quad (4)$$

And, the optimal solution is transformed by the following:

$$\sum_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \exp(-g |x_i - x_j|^2) \quad (5)$$

2.4. Optimization of SVM Parameters

SVM needs to set two parameters manually before running, penalty factor C and the kernel parameter gamma [28]. C represents the tolerance for errors, and a higher C indicates a smaller tolerance for errors, easily overfit, while a smaller C may lead to underfitting. On the other hand, the kernel parameter gamma determines the distribution of data mapped to a new feature space. The larger the gamma result, the fewer the support vectors. Therefore, the value of gamma influences the performance of training and prediction by affecting the number of support vectors.

Under the conditions of large-scale and multi-dimensional spatial data, to mitigate the impact of the penalty factor C and kernel parameter gamma on SVM, particle swarm optimization (PSO) is employed. Compared to other commonly used optimization algorithm, PSO has strong parallel processing capabilities in parameter optimization, a fast convergence speed, and global convergence advantages, and the best parameter combination of C and gamma can be searched globally [28].

The basic concept of PSO is to search for the optimal solution through cooperation and information sharing among individuals in a swarm. Each particle has only two properties, velocity and position, and each particle searches for the optimal solution in the search space individually, recording it as its personal best. By sharing individual extrema with other particles in the entire particle swarm, the optimal individual extremum is found as the current global optimal solution of the total particle swarm. All particles will change their motion speed and current position continuously based on their position relationship with the optimal particle [23]. The velocity update formula for each particle in the swarm is given by the following:

$$v_i(t+1) = \omega v_i(t) + c_1 \text{rand}_1(pbest_i(t) - x_i(t)) + c_2 \text{rand}_2(g(t) - x_i(t)) \quad (6)$$

where c_1 and c_2 are learning factors, ω is the inertia factor, $v_i(t)$ represents the velocity of the i -th particle, rand_1 and rand_2 are random numbers uniformly distributed in the interval $[0,1]$, $x_i(t)$ denotes the position of the i -th particle, and t indicates the iteration number.

The update position formula for each particle in the swarm is as follows:

$$x_i(t+1) = v_i(t+1) + x_i(t) \quad (7)$$

The main algorithmic flow of PSO-SVM is illustrated in Figure 3.

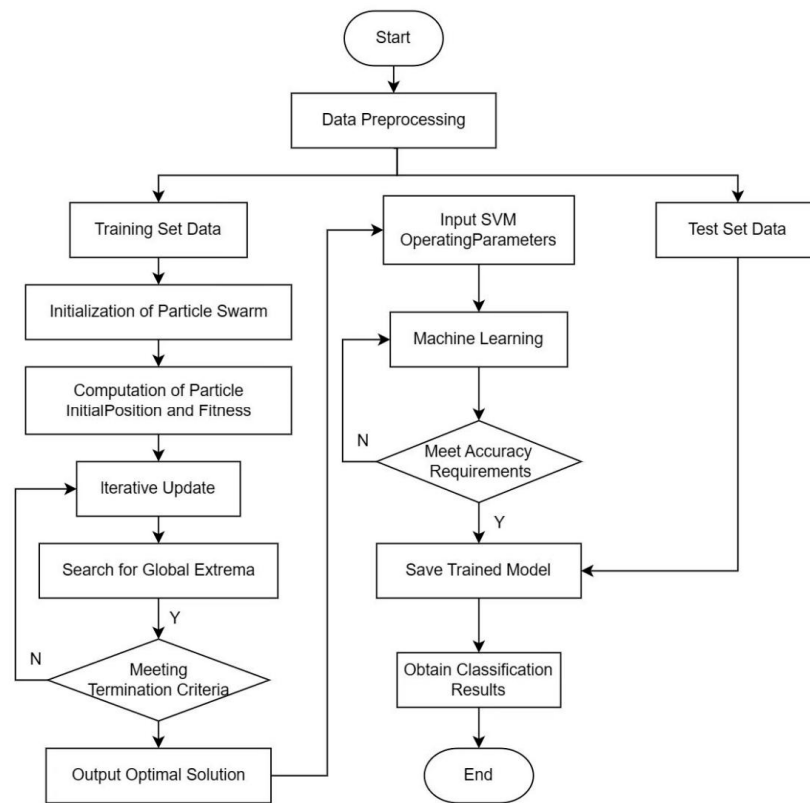
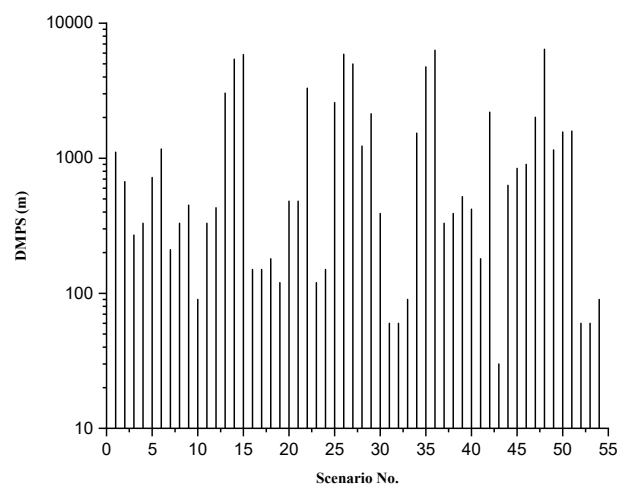


Figure 3. PSO-SVM algorithm flowchart.

3. Results and Discussion

3.1. GMS Numerical Simulation Results of Orthogonal Experiment Scenarios

The natural attenuation process of organic contaminants in groundwater was simulated with the investigation site, and numerous parameters were involved. Therefore, this paper employed a scenario analysis approach to explore these parameters, altering their assigned values with each set of values consistent with a specific scenario in order to generalize the prediction model of groundwater contamination plume evolution features. According to the orthogonal experiment, 54 scenarios with natural attenuation were simulated, with the simulation results represented by the key features of contaminant plume, which included the DMPS, T-DMPS, and MC-DMPS of each scenario. The experiment results of 54 sets of scenario simulation are shown in Figure 4.



(a)

Figure 4. Cont.

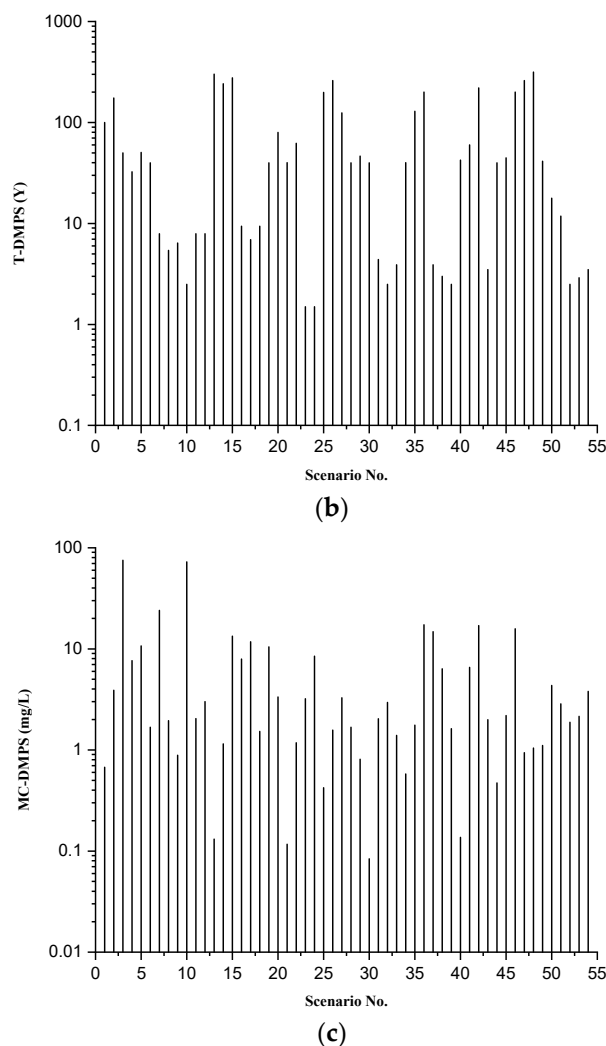


Figure 4. The numerical simulation results from 54 representative scenarios based on the orthogonal experiment design. (a) Simulated DMPS values for 54 representative scenarios; (b) simulated T-DMPS values for 54 representative scenarios; (c) simulated MC-DMPS values for 54 representative scenarios.

3.2. Identification of Key Controlling Factors

Analysis of variance (ANOVA), a widely utilized statistical technique, is often applied to assess the differences in mean values across two or more groups, assuming the homogeneity of variance. This method divides the total variance observed within an experiment into contributions from factor effects and experimental errors, thereby offering a quantitative evaluation of the relative significance of each factor variability on the overall variation. Through F-tests conducted within the SPSS, it is possible to examine the aggregate variance attributable to the factors versus random errors, determining the factor significance on the outcomes via p -values [38]. The results show that the data meet the Bonferroni standard deviation confidence interval of 0.95 and the homogeneity of variance test, proving that all data were normally independent and with the same variance. The ANOVA findings related to the factors affecting the DMPS, T-DMPS, and MC-DMPS are presented in Table 3. Additionally, the data were performed by logarithmic transformation.

The ANOVA results revealed that, under the criterion of $p < 0.01$, there are six factors significantly associated with the T-DMPS according to the simulation outcomes: the phreatic aquifer's effective porosity, its permeability coefficient, its adsorption coefficient, and its degradation coefficient, and the effective porosity and the adsorption coefficient of the aquitard. Regarding the DMPS, eight factors exhibit significant correlations: the head difference between the upstream and downstream boundaries, the phreatic aquifer's

permeability coefficient, its degradation coefficient, the adsorption coefficient and degradation coefficient of the aquitard, the thickness and permeability coefficient of the confined aquifer, and the concentration of the contamination source. Moreover, nine factors significantly influence the MC-DMPS: the head difference between the upstream and downstream boundaries, the phreatic aquifer’s effective porosity, its degradation coefficient, the aquitard thickness, the confined aquifer’s effective porosity, its dispersivity, the contamination source concentration, its area, and the proportion of the contamination source thickness relative to the phreatic aquifer’s thickness. With a *p*-value threshold set below 0.05, the count of factors significantly associated with the DMPS, T-DMPS, and MC-DMPS of the contamination plume increases to ten, nine, and fifteen, respectively.

The analysis of variance of all factors indicates that in the dynamics of groundwater contaminant natural attenuation, which encompasses the transportation, dispersion, and natural degradation of contaminants within the phreatic aquifer, there is a significant interrelation with the hydrogeological attributes of both the aquitard and the confined aquifer. This approach addresses and ameliorates the limitations observed in both domestic and international groundwater contamination studies, which often focus solely on the hydrogeological or hydrochemical parameters of phreatic aquifers.

Table 3. The total parameter analysis of variance for three different plume key features based on their individual numerical simulation results.

Parameters	ln(T-DMPS)			ln(DMPS)			ln(MC-DMPS)		
	Mean Square Error	F Value	<i>p</i> Value	Mean Square Error	F Value	<i>p</i> Value	Mean Square Error	F Value	<i>p</i> Value
<i>lnΔH_h</i>	0.199	1.02	0.316	4.589	29.69	0.000 **	3.474	53.35	0.000 **
<i>lnΔH_v</i>	0.087	0.45	0.507	4.750	2.84	0.099	0.388	5.96	0.018 *
<i>lnP₁</i>	1.483	7.63	0.008 **	0.189	1.19	0.282	2.638	40.50	0.000 **
<i>lnM₁</i>	0.484	2.49	0.120	0.110	0.68	0.412	0.341	5.24	0.026 *
<i>lnK_{w1}</i>	1.507	7.75	0.007 **	27.717	173.24	0.000 **	1.621	24.89	0.000 **
<i>lnK_{a1}</i>	2.963	15.25	0.000 **	0.024	0.15	0.699	0.404	6.20	0.016 *
<i>lnK_{d1}</i>	73.055	376.04	0.000 **	41.968	262.31	0.000 **	7.856	120.62	0.000 **
<i>lnD₁</i>	0.002	0.01	0.917	0.007	0.05	0.833	0.014	0.21	0.648
<i>lnP₂</i>	1.431	7.37	0.009 **	0.858	5.36	0.025 *	0.396	6.08	0.017 *
<i>lnM₂</i>	0.000	0.00	0.982	0.012	0.07	0.790	0.690	10.60	0.002 **
<i>lnK_{w2}</i>	0.482	2.48	0.121	0.001	0.00	0.945	0.013	0.20	0.658
<i>lnK_{a2}</i>	2.437	12.55	0.001 **	1.770	11.07	0.002 **	0.026	0.40	0.530
<i>lnK_{d2}</i>	0.511	2.63	0.111	1.408	8.80	0.005 **	0.013	0.20	0.659
<i>lnD₂</i>	0.324	1.67	0.202	0.122	0.76	0.388	0.021	0.33	0.569
<i>lnP₃</i>	0.003	0.02	0.897	0.037	0.23	0.634	0.493	7.57	0.008 **
<i>lnM₃</i>	0.143	0.74	0.394	2.072	12.95	0.001 **	0.245	3.77	0.058
<i>lnK_{w3}</i>	0.437	2.25	0.140	1.772	11.07	0.002 **	0.051	0.79	0.378
<i>lnK_{a3}</i>	0.007	0.03	0.854	0.130	0.81	0.372	0.002	0.02	0.878
<i>lnK_{d3}</i>	0.085	0.44	0.512	0.360	2.25	0.140	0.435	6.68	0.013 *
<i>lnD₃</i>	0.691	3.56	0.065	0.737	4.61	0.509	0.758	11.64	0.001 **
<i>lnR</i>	0.521	2.68	0.108	0.597	3.73	0.044 *	0.001	0.01	0.922
<i>lnC₀</i>	0.225	1.16	0.287	0.011	0.07	0.000 **	61.765	948.33	0.000 **
<i>lnA</i>	0.842	4.33	0.042 *	3.429	21.43	0.329	13.293	204.10	0.000 **
<i>lnC_s</i>	0.973	5.01	0.030 *	0.156	0.97	0.792	0.121	1.86	0.178
<i>lnr</i>	1.213	6.24	0.016 *	0.195	1.22	0.276	16.048	264.40	0.000 **

Note: ** indicates that the *p*-value is less than 0.01, indicating that parameter is a very significant parameter in this model. * indicates that the *p*-value is less than 0.05 but greater than 0.01, indicating that parameter is a significant parameter in this model.

3.3. PSO-SVM Surrogate Model

3.3.1. Model Development Based on the Dataset from Orthogonal Experiment Scenarios

In the study on groundwater contamination plume evolution, with the natural attenuation capabilities of organic contaminants, the surrogate prediction models were developed

by employing the PSO-SVM approach. Those models incorporated twenty-five variables influencing natural attenuation as input parameters, with the DMPS, T-DMPS, and MC-DMPS as their output metrics. By utilizing the classification potential of SVM, the model facilitates both single- and multi-objective predicting processes. For scenarios with a single output metric, the classification of the output data was conducted, with values above a predefined threshold marked as 0 and those below as 1, creating three distinct classification models. In addressing challenges with two output metrics, the data were paired and classified, generating three predictive models for the feature variables including the following: the DMPS, T-DMPS, and MC-DMPS. For dual-output scenarios, a two-dimensional coordinate system was employed, exemplified by the DMPS and MC-DMPS model, where the X-axis represented the MC-DMPS data and the Y-axis represented the DMPS data. This approach generated a two-dimensional dataset, marking a point as 1 when both concentration and area criteria are met and 0 otherwise, effectively converting two disparate output metrics into a binary classification challenge suitable for SVM. In cases involving three outputs for multi-objective decision making, a trio of dependent variables from simulated scenarios was plotted in a three-dimensional space, creating a dataset where the points meeting all three criteria are labeled as 1 and all others as 0. The mapping of the GMS simulation outcomes is depicted in Figure 5.

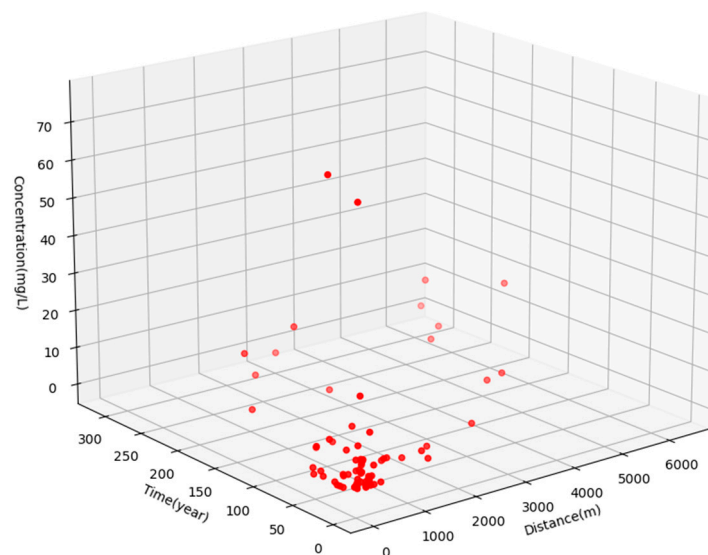


Figure 5. The three-dimensional spatial distribution of simulation data. Each point represents a piece of data. X-axis represents T-DMPS, Y-axis represents DMPS, and Z-axis represents MS-DMPS, which means every point is a three-dimension datum under SVM calculation mechanism.

In terms of machine learning, subsequent to the data processing steps described, along with utilizing the classification mechanism inherent in support vector machines, it becomes feasible to segregate data points within two-dimensional or three-dimensional spaces by distinctly defining classification thresholds for each output parameter. For example, by establishing criteria such that the time must be less than 20 years, the distance less than 20,000 m, and the concentration less than 5 mg/L for a data point to be concurrently satisfied, such points were categorized as '1', which meant class '1', indicating that the plume feature performance met the specific criteria. The other plume feature performances not meeting these criteria were assigned as class '0'. The classification could have been changed by actual remediation targets. However, at that stage, the accuracy of the surrogate model hovered around 0.5. Consequently, efforts to improve the accuracy of the PSO-SVM model were directed towards augmenting the volume of sample data and minimizing the count of input variables.

3.3.2. Model Enhancement Based on the Additional Sample Data

The further investigation incorporated more sample data, emphasizing the scenarios primarily accounting for key controlling factors with more importance. As a result, several specific parameters associated with the source of contamination, the degradation rate C_s , initial concentration C_0 , and source area A were assigned two levels each. By employing a complete permutation and combination approach, eight distinct combinations were derived. These combinations, when integrated with the other parameters set at three levels within the simulation framework for the natural attenuation of groundwater contamination, culminated in a total of 24 unique scenarios. The training results are shown in Figure 6.

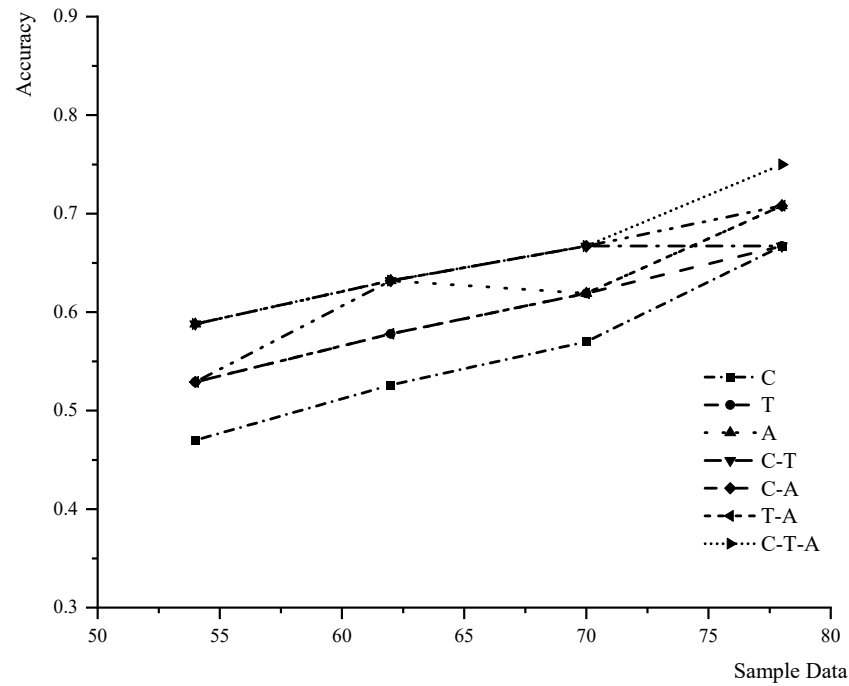


Figure 6. The model training with sample data added.

3.3.3. Model Reliability Increasing by Dimension Deduction

By leveraging the ANOVA results to streamline the input factors of the PSO-SVM models, the accuracy of various models was enhanced from approximately 0.5 to over 0.7. Utilizing the ANOVA results, concurrently increasing the sample size and decreasing the number of input factors, the prediction outcomes of each PSO-SVM model are detailed in the subsequent Table 4.

Table 4. The comparison of training results from each PSO-SVM model with different outputs.

$p < 0.01$							
Outputs	C (mg/L)	T (Year)	D (m)	T-C (Year-mg/L)	T-D (Year-m)	D-C (m-mg/L)	T-D-C (Year-m-mg/L)
Classification	$C < 5$	$T < 20$	$D < 2000$	$C < 5, T < 20$	$T < 20, D < 2000$	$D < 2000, C < 5$	$T < 20, D < 2000, C < 5$
Accuracy	0.833	0.833	0.875	0.917	0.833	0.875	0.965
$p < 0.05$							
Outputs	C (mg/L)	T (Year)	D (m ²)	T-C (Year-mg/L)	T-D (Year-m)	D-C (m-mg/L)	T-D-C (Year-m-mg/L)
Classification	$C < 5$	$T < 20$	$S < 2000$	$C < 5, T < 20$	$T < 20, S < 2000$	$S < 2000, C < 5$	$T < 20, S < 2000, C < 5$
Accuracy	0.833	0.833	0.875	0.917	0.875	0.833	0.965

From the standpoint of model training outcomes, the implementation of critical factor selection through the application of significance values derived from the variance analysis markedly improves model accuracy. Furthermore, in the prediction model of contamination plume evolution features with multiple target outputs—namely, the DMPS, T-DMPS, and MC-DMPS—the model accuracy surpasses 0.9. Comparing with other SVM prediction or classification models, the accuracy of this surrogate model can be improved to 0.9 only with 54 sets of training data, while similar levels of accuracy were achieved by generally applying more than 100 training datasets in other reported SVM models [27–29].

Taking the T-D-C multi-objective PSO-SVM model as an example, when analyzing the 24 test scenarios from 78 simulation scenarios, the maximum iteration step was 100. The input parameters of the three-objective model were combined with the key factors of each signal feature, so there were 20 input parameters totally. After a number of iteration steps, the particle swarm found the global optimal fitness value, indicating that the particle swarms were moving towards the optimal position after every iteration step. The importance of the input factors are presented in Figure 7. The result showed that class ‘1’ scenarios appeared 6 times, and class ‘0’ scenarios appeared 18 times. There were 22 samples of correct decisions and 2 samples of incorrect decisions, and the model accuracy was 0.917. This demonstrates the high feasibility of employing the PSO-SVM model in the research of groundwater contamination plume evolution feature prediction.

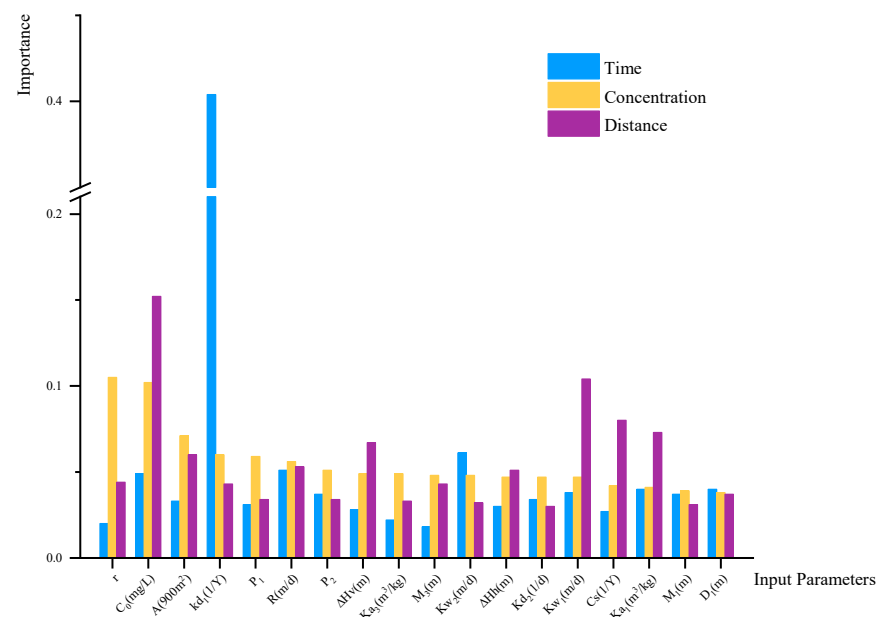


Figure 7. The importance of input parameters.

3.4. Multiple Regression Statistical Surrogate Model

Multivariate statistical regression models are pivotal in statistical analysis, facilitating not just the exploration of the effects of several independent variables on dependent variables but also the discernment of data patterns through the examination of positive and negative correlations and the relative importance of regression coefficients among variables [39]. Additionally, these models enable the estimation and prediction of future values for dependent variables using data from existing independent variables. In this research, employing 78 simulated experimental datasets, three evolution features were considered as dependent variables (Y). A multivariate regression analysis was executed using SPSS to develop predictive statistical surrogate models for the DMPS, T-DMPS, and MC-DMPS. These models were utilized to forecast the natural attenuation of groundwater organic contamination, with the importance and significance of parameters confirmed through standard regression coefficients. The developed multivariate regression predictive model for the natural attenuation of the contaminant plume demonstrates significant

predictive capability [25]. The general form of the multiple regression prediction models for the natural attenuation of contamination plumes presents as follows:

$$\ln Y = \ln \rho + \alpha_1 \ln \Delta H_h + \alpha_2 \ln \Delta H_v + \alpha_3 \ln P_1 + \alpha_4 \ln \Delta M_1 + \dots + \alpha_{25} \ln r \tag{8}$$

Then, the multiple regression prediction models can be transformed as follow:

$$Y = \rho \times \Delta H_h^{\alpha_1} \times \Delta H_v^{\alpha_2} \times P_1^{\alpha_3} \times M_1^{\alpha_4} \times \dots \times r^{\alpha_{25}} \tag{9}$$

where Y stands for the plume evolution features of groundwater contamination, including the DMPS, T-DMPS, and MC-DMPS.

The regression formulas for the DMPS are transformed and obtained as follows:

$$\begin{aligned} \text{DMPS} = & -2.39 \times \Delta H_h^{0.725} \times \Delta H_v^{0.1568} \times P_1^{-0.183} \times M_1^{-0.188} \times Kw_1^{0.7961} \times Ka_1^{0.0253} \\ & \times kd_1^{-0.5124} \times D_1^{0.0176} \times P_2^{-0.628} \times M_2^{0.0264} \times Kw_2^{-0.006} \times Ka_2^{0.204} \\ & \times kd_2^{0.094} \times D_2^{-0.112} \times P_3^{0.078} \times M_3^{0.206} \times Kw_3^{0.202} \times Ka_3^{0.055} \times kd_3^{0.046} \\ & \times D_3^{0.169} \times R^{-0.221} \times C_0^{0.203} \times A^{0.030} \times C_S^{0.001} \times r^{-0.066} \end{aligned} \tag{10}$$

$$R^2 = 0.939$$

$$\begin{aligned} \text{DMPS} = & 0.634 \times \Delta H_h^{0.741} \times Kw_1^{0.792} \times kd_1^{-0.521} \times P_2^{-0.403} \times Ka_2^{0.209} \times kd_2^{0.079} \times M_3^{0.182} \\ & \times kw_3^{0.154} \times R^{-0.168} \times C_0^{0.251} \end{aligned} \tag{11}$$

$$R^2 = 0.917$$

$$\begin{aligned} \text{DMPS} = & 0.634 \times \Delta H_h^{0.741} \times Kw_1^{0.792} \times kd_1^{-0.521} \times Ka_2^{0.209} \times kd_2^{0.079} \times M_3^{0.182} \times kw_3^{0.154} \\ & \times C_0^{0.251} \end{aligned} \tag{12}$$

$$R^2 = 0.909$$

The regression formulas for the T-DMPS are achieved as follows:

$$\begin{aligned} \text{T-DMPS} = & 7.1 \times \Delta H_h^{-0.131} \times \Delta H_v^{-0.059} \times P_1^{-0.466} \times M_1^{-0.328} \times Kw_1^{0.169} \times Ka_1^{0.238} \\ & \times kd_1^{-0.605} \times D_1^{-0.008} \times P_2^{0.705} \times M_2^{0.002} \times Kw_2^{-0.124} \times Ka_2^{0.216} \\ & \times kd_2^{-0.052} \times D_2^{0.167} \times P_3^{-0.022} \times M_3^{-0.049} \times Kw_3^{0.084} \times Ka_3^{0.011} \\ & \times kd_3^{0.021} \times D_3^{-0.151} \times R^{-0.17} \times C_0^{-0.047} \times A^{-0.062} \times C_S^{-0.105} \times r^{0.152} \end{aligned} \tag{13}$$

$$R^2 = 0.909$$

$$\begin{aligned} \text{T-DMPS} = & 5.509 \times P_1^{-0.570} \times Kw_1^{0.127} \times Ka_1^{0.192} \times kd_1^{-0.626} \times P_2^{0.442} \times Ka_2^{0.170} \\ & \times A^{-0.056} \times C_S^{-0.015} \times r^{0.109} \end{aligned} \tag{14}$$

$$R^2 = 0.864$$

$$\begin{aligned} \text{T-DMPS} = & 5.569 \times P_1^{-0.537} \times Kw_1^{0.139} \times Ka_1^{0.205} \times kd_1^{-0.620} \times P_2^{0.500} \times Ka_2^{0.183} \\ & \times C_S^{-0.014} \end{aligned} \tag{15}$$

$$R^2 = 0.851$$

The regression formulas for the MC-DMPS are established as follows:

$$\begin{aligned} \text{MC-DMPS} = & 8.043 \times \Delta H_h^{0.548} \times \Delta H_v^{0.124} \times P_1^{-0.622} \times M_1^{0.275} \times Kw_1^{0.176} \times Ka_1^{-0.088} \\ & \times kd_1^{-0.198} \times D_1^{-0.021} \times P_2^{-0.371} \times M_2^{-0.192} \times Kw_2^{0.020} \times Ka_2^{0.022} \\ & \times kd_2^{0.008} \times D_2^{0.043} \times P_3^{-0.269} \times M_3^{-0.064} \times Kw_3^{0.025} \times Ka_3^{0.006} \times kd_3^{0.047} \\ & \times D_3^{0.158} \times R^{-0.006} \times C_0^{0.782} \times A^{0.247} \times C_S^{-0.004} \times r^{0.553} \end{aligned} \tag{16}$$

$$R^2 = 0.977$$

$$\begin{aligned} \text{MC-DMPS} = & 0.123 \times \Delta H_h^{0.555} \times \Delta H_v^{0.131} \times P_1^{-0.613} \times M_1^{0.292} \times Kw_1^{0.180} \times Ka_1^{-0.083} \\ & \times kd_1^{-0.197} \times P_2^{-0.302} \times M_2^{-0.184} \times P_3^{-0.260} \times M_3^{-0.061} \times kd_3^{0.049} \\ & \times D_3^{0.165} \times C_0^{0.767} \times A^{0.247} \times r^{0.558} \end{aligned} \tag{17}$$

$$R^2 = 0.975$$

$$MC - DMPS = 0.164 \times \Delta H_h^{0.557} \times P_1^{-0.611} \times Kw_1^{0.181} \times kd_1^{-0.197} \times M_2^{-0.183} \times P_3^{-0.258} \times D_3^{0.166} \times C_0^{0.786} \times A^{0.246} \times t^{0.559} \quad (18)$$

$$R^2 = 0.962$$

Incorporating the outcomes from variance analysis, the comprehensive full-factor regression model, and the multivariate regression models with p -values above 0.05 and p -values above 0.01, the R^2 values reach more than 0.85. This indicates a strong correlation between the plume evolution feature variables under natural attenuation conditions and various impact factors. The Durbin–Watson test values for the regression models are all below 2, signifying the absence of significant autocorrelation within the data and affirming the statistical significance of the models.

3.5. Discussion

3.5.1. Comparisons between the Two Types of the Developed Surrogate Models

By using an area–time–concentration multi-objective output model as an example case, selecting the same test dataset, and taking the same classification criteria as the PSO-SVM surrogate model, a comparison was performed for the model prediction accuracy between the PSO-SVM and statistical regression models. The comparison results are shown in Table 5.

Table 5. Comparison of prediction accuracy between two types of models.

Outputs	C (C < 5 mg/L)	T (T < 20 Y)	D (D < 2000 m)	C-T-D (C < 5 mg/L T < 20 Y D < 2000 m)
Accuracy of total factors	0.625	0.333	0.917	0.75
Accuracy when $p < 0.05$	0.583	0.333	0.833	0.75
Accuracy when $p < 0.01$	0.417	0.333	0.734	0.672
Accuracy of PSO-SVM	0.833	0.833	0.875	0.965

Specifically, as to the plume evolution features, the DMPS, T-DMPS, and MC-DMPS, a study case was established using the same test data and classification criteria as those applied in both PSO-SVM and statistical regression models, followed by an accuracy comparison. The comparison findings indicated that with contamination plume thresholds defined as a DMPS under 2000 m, T-DMPS below 20 years, and MC-DMPS less than 5 mg/L, the accuracy of the regression model for multiple outputs stood at 0.75. Nonetheless, the accuracy of the regression surrogate models experienced a notable decline with the identified key controlling factors, rendering the performance generally inferior to or at best comparable to that of the PSO-SVM surrogate model. This stressed the superior performance of the PSO-SVM model when operating with a limited number of key controlling factors. By taking the data volume into consideration, the PSO-SVM surrogate model also offered a good performance based on a limited dataset, which meant this surrogate mode took a wider application in situations where the data were laborious to obtain. In addition, it was indicated that the PSO-SVM model could achieve more accurate results under conditions of limited parameter availability compared to the conventional statistical approaches, while other published research studies were less focused on the comparison between different types of prediction models based on the same test data [17–19]. Overall, the PSO-SVM surrogate model demonstrated greater reliability of application for predicting plume evolution features concerning groundwater contamination, particularly regarding the natural attenuation of organic contaminants with a limited data size, than the traditional regression models.

3.5.2. Further Elucidation about the Validation on the Built Surrogate Models

The statistical surrogate models and PSO-SVM surrogate models were both built up based on the datasets from the numerical simulation results, obtained through the reliable Groundwater Modeling System, of many representative modeling scenarios. This means that the numerical simulation data used in this investigation can be treated as the “real accurate data” for establishing those surrogate models. As a result, we should firstly validate our surrogate models using additional numerical simulation data from the representative modeling scenarios which were not included and utilized for building the surrogate models. In fact, in this investigation, we conducted validation tests, by using an additional 24 datasets, on both the statistical surrogate models and PSO-SVM surrogate models. In this sense, we would say that our surrogate models have been tested in respect to the real prediction of experimental results. Certainly, it is reasonable and valuable to perform further examinations and confirmations on those established surrogate models using the specific site field data in future practical applications.

3.5.3. Practicability and Advantage of Prediction Uncertainty Assessment

Upon considering the uncertainty in obtaining the factor parameter values, there are different degrees of uncertainty regarding the predicted results of plume evolution feature variables. The uncertainty analyses by Monte Carlo simulations can evaluate the possible range of variation for the prediction results obtained from the established surrogate models. Therefore, giving the parameter error range of related impact factor variables and their distribution types, a large number of possible different parameter combination scenarios (such as 500 scenarios) can be generated. By using the surrogate models, the prediction results of corresponding plume feature variables for all the scenarios can be efficiently calculated, and the relative frequency distribution histogram of the predicted results can be rapidly obtained. Based on the relative frequency distribution histogram, under a given confidence probability, the possible variation range of the plume feature variables can be offered quickly, providing a quantitative basis for the upper and lower boundaries of plume feature variables in contamination control and risk management decision making. Comparably, it is time-consuming, even impossible, to implement detailed complex numerical models since a large number of simulations of scenarios are required.

4. Conclusions

This work developed surrogate models with an effective and practicable pathway for predicting the key plume features, including the DMPS, T-DMPS, and MC-DMPS, of groundwater contamination with natural attenuation. The developed models were aimed to apply to typical groundwater systems, and a wide range of the hydrogeological and geochemical parameters were comprehensively considered that may affect the natural attenuation of the contaminant plumes. The main conclusions drawn from this investigation are given as follows:

- (1) According to the numerical simulations and variance analysis, the key controlling factors affecting the DMPS, T-DMPS and MC-DMPS of contaminant plumes are different. It is indicated that the transport and fate of contaminants in the aquifer are significantly correlated not only with hydrological parameters but also with certain parameters of the aquitard and confined aquifer. The degradation coefficient of the phreatic aquifer is a crucial factor determining the natural attenuation of the contaminants.
- (2) The PSO-SVM model prediction accuracy can be gradually enhanced by implementing the measures of effectively increasing the sample data sizes and replacing all of considered input variables with the identified key controlling factors. It is interesting to note that the final developed PSO-SVM models still can present good reliability with the utilization of the limited sample data.
- (3) The statistical surrogate models are also constructed by multiple regression based on the same dataset used for the PSO-SVM model. The statistical regression surrogate

models also exhibit pretty good fitting accuracy, while in comparison, the PSO-SVM models offer generally higher prediction accuracy than the statistical regression models, particularly by taking the key controlling factors as input variables.

- (4) The findings of this study offer effective generic surrogate models along with a scientific basis and investigation approach reference for environmental risk management and remediation pertaining to the commonly existing groundwater contamination.

Author Contributions: Conceptualization, Y.W. and M.W.; methodology Y.W. and M.W.; software, Y.W. and R.L.; formal analysis, Y.W. and M.W.; writing—original draft preparation, Y.W.; writing—review and editing, M.W.; supervision, M.W.; project administration, M.W.; funding acquisition, M.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Key Research and Development Program of China (2020YFC1807102, 2020YFC1807100).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Demirak, A.; Yilmaz, F.; Tuna, A.L.; Ozdemir, N. Heavy metals in water, sediment and tissues of *Leuciscus cephalus* from a stream in southwestern Turkey. *Chemosphere* **2006**, *63*, 1451–1458. [[CrossRef](#)] [[PubMed](#)]
- Muhammad, S.; Shah, M.T.; Khan, S. Arsenic health risk assessment in drinking water and source apportionment using multivariate statistical techniques in Kohistan region, northern Pakistan. *Food Chem. Toxicol.* **2010**, *48*, 2855–2864. [[CrossRef](#)]
- Wang, W.; Jia, J.; Zhang, B.; Xiao, B.; Yang, H.; Zhang, S.; Gao, X.; Han, Y.; Zhang, S.; Liu, Z.; et al. A review of Sustained release materials for remediation of organically contaminated groundwater: Material preparation, applications and prospects for practical application. *J. Hazard. Mater. Adv.* **2024**, *13*, 100393. [[CrossRef](#)]
- Cundy, A.; Bardos, R.; Church, A.; Puschenreiter, M.; Friesl-Hanl, W.; Müller, I.; Neu, S.; Mench, M.; Witters, N.; Vangronsveld, J. Developing principles of sustainability and stakeholder engagement for “gentle” remediation approaches: The European context. *J. Environ. Manag.* **2013**, *129*, 283–291. [[CrossRef](#)] [[PubMed](#)]
- Ravindiran, G.; Rajamanickam, S.; Sivarethinamohan, S.; Sathaiyah, B.K.; Ravindran, G.; Muniasamy, S.K.; Hayder, G. A Review of the Status, Effects, Prevention, and Remediation of Groundwater Contamination for Sustainable Environment. *Water* **2023**, *15*, 3662. [[CrossRef](#)]
- Hossain, M.; Bhattacharya, P.; Jacks, G.; von Brömssen, M.; Ahmed, K.M.; Hasan, M.A.; Frape, S.K. *Sustainable Arsenic Mitigation—from Field Trials to Implementation for Control of Arsenic in Drinking Water Supplies in Bangladesh. Best Practice Guide on the Control of Arsenic in Drinking Water; Metals and Related Substances in Drinking Water Series; IWA Publishing: London, UK, 2017; pp. 99–116.*
- Lone, S.A.; Jeelani, G.; Mukherjee, A.; Coomar, P. Arsenic fate in upper Indus river basin (UIRB) aquifers: Controls of hydrochemical processes, provenances and water-aquifer matrix interaction. *Sci. Total Environ.* **2021**, *795*, 148734. [[CrossRef](#)]
- Qaiser, F.U.; Zhang, F.; Pant, R.R.; Zeng, C.; Khan, N.G.; Wang, G. Characterization and health risk assessment of arsenic in natural waters of the Indus River Basin, Pakistan. *Sci. Total Environ.* **2023**, *857*, 159408. [[CrossRef](#)]
- Liu, G.; Niu, J.; Zhang, C.; Guo, G. Characterization and assessment of contaminated soil and groundwater at an organic chemical plant site in Chongqing, Southwest China. *Environ. Geochem. Health* **2016**, *38*, 607–618. [[CrossRef](#)]
- Bardos, R.P.; Bone, B.D.; Boyle, R.; Evans, F.; Harries, N.D.; Howard, T.; Smith, J.W. The rationale for simple approaches for sustainability assessment and management in contaminated land practice-ScienceDirect. *Sci. Total Environ.* **2016**, *563*, 755–768. [[CrossRef](#)]
- Li, Y.; Wang, S.; Zhang, M.; He, Z.; Zhang, W. Research progress of monitored natural attenuation remediation technology for soil and groundwater pollution. *Zhongguo Huanjing Kexue/China Environ. Sci.* **2018**, *38*, 1185–1193.
- U.S. Environmental Protection Agency. *Use of Monitored Natural Attenuation at Superfund, RCRA, Corrective Action and UST Sites EPA's Office of Solid Waste and Emergency Response (OSWER) Directive 9200.4-17*; U.S. Environmental Protection Agency: Washington, DC, USA, 1999.
- Rügner, H.; Finkel, M.; Kaschl, A.; Bittens, M. Application of monitored natural attenuation in contaminated land management—A review and recommended approach for Europe. *Environ. Sci. Policy* **2006**, *9*, 568–576. [[CrossRef](#)]
- Amiri, S.; Rajabi, A.; Shabanlou, S.; Yosefvand, F.; Izadbakhsh, M.A. Prediction of groundwater level variations using deep learning methods and GMS numerical model. *Earth Sci. Inform.* **2023**, *16*, 3227–3241. [[CrossRef](#)]
- Valivand, F.; Katibeh, H. Simulation and Prediction of Groundwater Pollution Based on GMS: A Case Study in Beijing, China. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *826*, 012014.
- Valivand, F.; Katibeh, H. Prediction of nitrate distribution process in the groundwater via 3D modeling. *Environ. Model. Assess.* **2020**, *25*, 187–201. [[CrossRef](#)]

17. Tong, X.X.; Ning, L.B.; Dong, S.G. GMS model for assessment and prediction of groundwater pollution of a garbage dumping site in Luoyang. *Miner. Explor.* **2012**, *35*, 197–201.
18. Gao, Q.F.; Jiao, J.N.; Zhang, Y.Z.; Du, J.X. Application of numerical simulation in groundwater environmental impact assessment based on GMS: A case study of an industrial park. *Miner. Explor.* **2023**, *14*, 1236–1243.
19. Zou, Y.; Yousaf, M.S.; Yang, F.; Deng, H.; He, Y. Surrogate-Based Uncertainty Analysis for Groundwater Contaminant Transport in a Chromium Residue Site Located in Southern China. *Water* **2024**, *16*, 638. [[CrossRef](#)]
20. Davis, S.E. Efficient Surrogate Model Development: Impact of Sample Size and Underlying Model Dimensions. *Comput. Aided Chem. Eng.* **2018**, *44*, 979–984.
21. Kanu, O.P.; Ugwoha, E.; Udeh, N.U.; Amah, V. Modelling Groundwater Quality of Aba in Abia State Using Principal Component Analysis and Multiple Linear Regression. *J. Eng. Res. Rep.* **2023**, *25*, 39–54. [[CrossRef](#)]
22. Alshahri, A.H.; Elbisy, M.S. Assessment of Using Artificial Neural Network and Support Vector Machine Techniques for Predicting Wave-Overtopping Discharges at Coastal Structures. *J. Mar. Sci. Eng.* **2023**, *11*, 539. [[CrossRef](#)]
23. Guneshwor, L.; Eldho, T.I.; Kumar, A.V. Identification of Groundwater Contamination Sources Using Meshfree RPCM Simulation and Particle Swarm Optimization. *Water Resour. Manag.* **2018**, *32*, 1517–1538. [[CrossRef](#)]
24. Gad, M.; Gaagai, A.; Eid, M.H.; Szűcs, P.; Hussein, H.; Elsherbiny, O.; Elsayed, S.; Khalifa, M.M.; Moghanm, F.S.; Moustapha, M.E.; et al. Groundwater Quality and Health Risk Assessment Using Indexing Approaches, Multivariate Statistical Analysis, Artificial Neural Networks, and GIS Techniques in El Kharga Oasis, Egypt. *Water* **2023**, *15*, 1216. [[CrossRef](#)]
25. Krishna, A.K.; Satyanarayanan, M.; Govil, P.K. Assessment of heavy metal pollution in water using multivariate statistical techniques in an industrial area: A case study from Patancheru, Medak District, Andhra Pradesh, India. *J. Hazard. Mater.* **2009**, *167*, 366–373. [[CrossRef](#)] [[PubMed](#)]
26. Nuzul, N.B.M.; Hassan, M.K.; Norrima, M.; Amirul, W.M.W.; Heshalini, R.; Tarmizi, A.; Jafferi, J. Automatic dry waste classification for recycling purpose. *Int. Conf. Artif. Life Robot.* **2022**, *3*, 1003–1010.
27. Gai, R.; Guo, Z. A water quality assessment method based on an improved grey relational analysis and particle swarm optimization multi-classification support vector machine. *Front. Plant Sci.* **2023**, *14*, 1099668. [[CrossRef](#)]
28. Pang, X.; Hu, H.R.; Zhao, C.; Bai, N.B. Research on air quality classification based on PSO-SVM algorithm. *Environ. Sci. Surv.* **2023**, *42*, 63–67.
29. Xu, G.; Yan, X.-E.; Cao, N.; Ma, J.; Xie, G.; Li, L. Debris Flow Prediction Based on the Fast Multiple Principal Component Extraction and Optimized Broad Learning. *Water* **2022**, *14*, 3374. [[CrossRef](#)]
30. Wang, B.L.; Wang, M.Y.; Pang, Y.T. Primary controlling Factors and statistical modeling of plum stability for BTEX in Typical phreatic aquifers. *Earth Sci.* **2023**, *48*, 3454–3465.
31. *GBT 14848-2017; Quality Standard for Groundwater*. Standardization Administration of the People’s Republic of China: Beijing, China, 2017.
32. Vapnik, V. *The Natural of Statistical Learning Theory*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1995.
33. Samandi, V.; Mukhopadhyay, D. Workflow scheduling in cloud computing environment with classification ordinal optimisation using SVM. *Int. J. Comput. Sci. Eng.* **2021**, *24*, 563–571. [[CrossRef](#)]
34. Fan, Z.; Chen, W.G.; Zou, J.X.; Yang, D.K.; Shi, H.Y. Quantitative Analysis of Furfural Dissolved in Transformer Oil Based on Raman Spectroscopy. *J. Light Scatt.* **2018**, *30*, 46–50.
35. Li, M.Z.; Zhai, Y.Z.; Zuo, R. Sensitivity Analysis of Parameters in Numerical Simulation of Solute Transport in Groundwater. *S.-N. Water Transf. Water Sci. Technol.* **2014**, *3*, 133–137.
36. Zhang, S.; Liu, Z.; Sun, R.; Liu, W.; Chen, Y. Orthogonal Experimental Study on Remediation of Ethylbenzene Contaminated Soil by SVE. *Sustainability* **2023**, *15*, 1168. [[CrossRef](#)]
37. Newell, C.J.; Mcleod, R.K.; Gonzales, J.R. *BIOSCREEN: Natural Attenuation Decision Support System. User’s Manual Version 1.3*; U.S. Environmental Protection Agency, Office of Research and Development: Washington DC, USA, 1996.
38. Ni, H.; Liu, Y.R.; Long, Z.G. Application of orthogonal design to sensitivity analysis of landslide. *Chin. J. Rock Mech. Eng.* **2002**, *21*, 989–992.
39. Li, L. Quantifying TiO₂ Abundance of Lunar Soils: Partial Least Squares and Stepwise Multiple Regression Analysis for Determining Causal Effect. *J. Earth Sci.* **2011**, *22*, 549–565. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.