Research Article

Simulation and Prediction of Ion Transport in the Reclamation of Sodic Soils with Gypsum Based on the Support Vector Machine

Jinman Wang,^{1,2} Zhongke Bai,^{1,2} and Peiling Yang³

¹ College of Land Science and Technology, China University of Geosciences, 29 Xueyuanlu, Handian District, Beijing 100083, China

² Key Laboratory of Land Consolidation and Rehabilitation, Ministry of Land and Resources, 37 Guanying Park West, Xicheng, Beijing 100035, China

³ College of Hydraulic and Civil Engineering, China Agricultural University, 17 Qinghua Donglu, Handian District, Beijing 100083, China

Correspondence should be addressed to Jinman Wang; wangjinman2002@163.com

Received 3 January 2014; Accepted 22 January 2014; Published 16 March 2014

Academic Editors: H. M. Conesa, P. Grenni, and B. R. Wilson

Copyright © 2014 Jinman Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The effect of gypsum on the physical and chemical characteristics of sodic soils is nonlinear and controlled by multiple factors. The support vector machine (SVM) is able to solve practical problems such as small samples, nonlinearity, high dimensions, and local minima points. This paper reports the use of the SVM regression method to predict changes in the chemical properties of sodic soils under different gypsum application rates in a soil column experiment and to evaluate the effect of gypsum reclamation on sodic soils. The research results show that (1) the SVM soil solute transport model using the Matlab toolbox represents the change in Ca^{2+} and Na^{+} in the soil solution and leachate well, with a high prediction accuracy. (2) Using the SVM model to predict the spatial and temporal variations in the soil solute content is feasible and does not require a specific mathematical model. The SVM model can take full advantage of the distribution characteristics of the training sample. (3) The workload of the soil solute transport prediction model based on the SVM is greatly reduced by not having to determine the hydrodynamic dispersion coefficient and retardation coefficient, and the model is thus highly practical.

1. Introduction

The support vector machine (SVM) method was developed based on the Vapnik statistical learning theory and in particular the statistical Vapnik-Chervonenkis Dimension theory and the Structural Risk Minimization Inductive Principle [1]. The basic principle of SVM is to solve the classification problem. Its guiding ideology is to map nonlinear separable sample data to a higher dimensional linear separable space through a kernel function and then to solve the partition hyperplane using an optimization method to determine the parameters of the decision-making function and minimize its structural risk [2]. Theory and practice have proved that the SVM can solve the practical problems of a small sample, nonlinearity, high dimensionality, and local minima points

and has good generalizability [3]. The SVM has become a hot topic in machine learning research and has been successfully applied in pattern recognition, data mining, regression, and approximation [4–6]. In the field of agriculture, the problem of reclaiming sodic soils has attracted much attention. Many studies have proved that the effect of the desulfurization byproducts used in the reclamation process on the physical and chemical properties of soil is nonlinear and multifactorial [6–11]. Šimůnek et al. constructed a multicomponent solute transport model of sodic soil reclamation and predicted the solute transport by using the UNSATCHEM and HYDRUS models [12–14]. However, the models use the major ion chemistry module, which have a relatively complex code and require a large number of input parameters that are not always readily available [15–17]. The SVM regression method

is applied here to overcome these issues and determine the variation in the chemical properties of sodic soils with reclaimed time and different application levels of a flue gas desulfurization byproduct (or gypsum) and to evaluate the effect on sodic soils of reclamation with gypsum.

2. Materials and Methods

2.1. Physical and Chemical Properties of Soils and CaSO_4 . The soil was sampled from the Changsheng Experimental Station of the Bayannur League Institute of Water Resources in the northwest of China. The soil had typical sodic soil characteristics, that is, a high pH and exchangeable sodium percentage and low hydraulic conductivity. The soil texture was clay and its physical and chemical properties are listed in Table 1. The soil was air-dried, crushed, and passed through a 2 mm sieve before the experiments. The pure CaSO_4 (molecular weight 136) was purchased from the Tianjin Wendaxigui Chemical Reagent Factory.

2.2. Ca^{2+} Penetration Experiment. The experimental device (Figure 1) was a 25-cm-high Plexiglass column with a 5 cm inside diameter (the soil height was 10 cm), and a Mariotte bottle was used to supply the water. The experiment consisted of four treatments (Table 2) with different CaSO_4 rates (0.5, 1, 1.5, and 2 $\text{g}\cdot\text{L}^{-1}$). The pure CaSO_4 was selected as the reagent and distilled water was used as the solvent to configure the solution.

The experimental procedure was as follows. The tested soil samples were poured homogeneously into the Plexiglass column at a dry soil bulk density of $1.45\text{ g}\cdot\text{cm}^{-3}$ to a depth of 10 cm. The soil columns were saturated with distilled water from the bottom up at a saturation rate of 2 cm per hour. Once the soils were saturated, the surface water was quickly drained with a vacuum pump and the Ca^{2+} solution that had been configured in the Mariotte bottle was immediately supplied. The Ca^{2+} and Na^+ concentrations of the leachate were measured once every 24 hours. The Ca^{2+} and Na^+ concentrations in the soil solution and soil colloid were measured after the penetration experiment.

2.3. Analytical Methods and Statistical Analyses. The soil samples were air-dried and passed through a 1 mm sieve. The EC, pH, soluble anions, and soluble cations were measured using 1:5 water extracts. The soluble cations were measured using an atomic absorption spectrophotometer (AAS-3620); soluble anions were determined by anion chromatography (ICS-900); soil pH was determined with the glass electrode method. The salt content was measured using a 1 cm conductivity cell, dip-type probe. Exchangeable cations were determined in a 1 M ammonium acetate (pH = 7) extract. Following this extraction and washing with 96% alcohol, the cation exchange capacity was determined by the removal of ammonium ions by distillation. Na^+ and K^+ were determined by flame emission spectroscopy (FP6400) in the extract, and Ca^{2+} and Mg^{2+} were determined by atomic absorption spectrophotometer (AAS-3620). Particle size distribution was determined with the hydrometer method. The concentrations

of Na^+ and Ca^{2+} in the leachate were measured using an atomic absorption spectrophotometer (AAS-3620).

2.4. SVM Theory. The SVM method was initially introduced to solve classification problems but can be extended to deal with regression problems [6, 11].

A linear and separable sample set $(x_i \in R^n, y_i \in \{1, -1\}, i = 1, \dots, l)$ is defined, and the hyperplane $(w^* \cdot x) + b = 0$ is constructed and divided. The two types of sample points must be correctly divided to achieve $y_i((w \cdot x_i) + b) \geq 1, i = 1, \dots, l$, but the largest spacing of classifications needs to be achieved, where $2/\|w\|$ is the maximum or $(1/2)\|w\|^2$ is minimum. w^* and b^* give the optimal solution to the following optimization problem:

$$\begin{aligned} \min_{w,b} \quad & \frac{1}{2}\|w\|^2, \\ \text{s.t.} \quad & y_i((w \cdot x_i) + b) \geq 1, \quad i = 1, \dots, l. \end{aligned} \quad (1)$$

This is the decision-making function $f(x) = \text{sgn}((w^* \cdot x) + b^*)$.

Equation (1) can be transformed into the dual problem based on duality theory [6]:

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j (x_i \cdot x_j) - \sum_{j=1}^l \alpha_j, \\ \text{s.t.} \quad & \sum_{i=1}^l y_i \alpha_i = 0, \quad \alpha_i \geq 0, \quad i = 1, \dots, l, \end{aligned} \quad (2)$$

where α_i is a Lagrange multiplier. The following relationship exists among the optimal solutions α^* in (2) and w^* and b^* in (1):

$$\begin{aligned} w^* &= \sum_{i=1}^l \alpha_i^* y_i x_i, \\ b^* &= y_j - \sum_{i=1}^l y_i \alpha_i^* (x_i \cdot x_j). \end{aligned} \quad (3)$$

Consider the following linear regression problem.

The regression hyperplane $y = (w \cdot x) + b$ is obtained for the sample set $T = \{(x_1, y_1), \dots, (x_l, y_l)\}$, $x_i \in R^n, y_i \in R, i = 1, \dots, l$. Intuitively, the difference between the hyperplane value at x_i and the actual input y_i value should be as small as possible. A parameter $\varepsilon > 0$ is introduced which requires

$$\begin{aligned} \min_{w,b,\varepsilon} \quad & \varepsilon \\ \text{s.t.} \quad & -\varepsilon \leq y_i - ((w \cdot x_i) + b) \leq \varepsilon, \quad i = 1, \dots, l. \end{aligned} \quad (4)$$

Equation (4) now needs to be converted into a classification problem. It can be envisaged that all of the sample points in the sample set T are, respectively, moved up and moved down ε along the y_i direction, so that the sample set $T^+ = \{(x_i, y_i + \varepsilon) \mid i = 1, \dots, l\}$ obtained after

FIGURE 1: Experimental device of Ca^{2+} penetration.

TABLE 1: Physical and chemical properties of the soil.

(a)							
Exchangeable cations/cmole _c ·kg ⁻¹					Soil bulk		
Na ⁺	K ⁺	Ca ²⁺	Mg ²⁺	CEC	Density/g·cm ⁻³	pH _{1:5}	
8.65	0.60	0.50	1.27	11.02	1.45	9.15	
(b)							
Soluble cations/mmol _c ·L ⁻¹					Soluble anions/mmol _c ·L ⁻¹		
Na ⁺	K ⁺	Ca ²⁺	Mg ²⁺	Cl ⁻	SO ₄ ²⁻	HCO ₃ ⁻	CO ₃ ²⁻
303.15	5.18	3.33	3.33	193.33	50.00	51.67	20.00
(c)							
Particle size distribution/%							EC _{1:5}
2.0–0.02		0.02–0.002		<0.002		dS·m ⁻¹	
23.2		0.95		42.1		0.95	

TABLE 2: Experimental treatments of Ca^{2+} penetration.

Experimental treatments	T1	T2	T3	T4
Concentration of $\text{CaSO}_4/\text{g} \cdot \text{L}^{-1}$	0.5	1.0	1.5	2.0
Concentration of $\text{Ca}^{2+}/\text{mmol}_c \cdot \text{L}^{-1}$	7.35	14.71	22.06	29.41

the upward movement is above the regression hyperplane (positive class) and the sample set $T^- = \{(x_i, y_i - \epsilon) \mid i = 1, \dots, l\}$ obtained after the downward movement is below the regression hyperplane (negative class point). The regression hyperplane is thus transformed into a partition hyperplane, and the regression problem is converted into a classification problem.

This problem can be solved by modifying (1):

$$\min_{w, \eta, b} \frac{1}{2} \|w\|^2 + \frac{1}{2} \eta^2, \quad \text{s.t.} \quad z_i ((w \cdot x_i) + \eta(y_i + z_i \epsilon) + b) \geq 1, \quad i = 1, \dots, 2l, \quad (5)$$

where η is the weight vector corresponding to the y component of the sample points:

$$z_i = \begin{cases} 1, & i = 1, \dots, l; \\ -1, & i = l+1, \dots, 2l. \end{cases} \quad (6)$$

The partition hyperplane $(w^* \cdot x) + \eta^* y + b^* = 0$ can be obtained according to the optimal solution (w^*, η^*, b^*) to (5), which corresponds to the regression function $y = (w \cdot x) + b$, where $w = -w^*/\eta^*$ and $b = -b^*/\eta^*$.

Letting $\epsilon^* = \epsilon - 1/\eta^*$, (5) is then equivalent to the following:

$$\min_{w, b} \frac{1}{2} \|w\|^2, \quad \text{s.t.} \quad \begin{aligned} (w \cdot x_i) + b - y_i &\leq \epsilon^*, & i = 1, \dots, l, \\ y_i - (w \cdot x_i) - b &\leq \epsilon^*, & i = l+1, \dots, 2l. \end{aligned} \quad (7)$$

The algorithm is further improved by allowing fitting error through the introduction of slack variables $\xi = \{\xi_i\}$ and a penalty parameter C , which modifies (5) as follows:

$$\begin{aligned} \min_{w, \eta, b, \xi} \quad & \frac{1}{2} \|w\|^2 + \frac{1}{2} \eta^2 + \frac{C}{l} \sum_{i=1}^{2l} \xi_i, \\ \text{s.t.} \quad & z_i ((w \cdot x_i) + \eta (y_i + z_i \varepsilon) + b) + \xi_i \geq 1, \\ & i = 1, \dots, 2l, \\ & \xi_i \geq 0, \quad i = 1, \dots, 2l. \end{aligned} \quad (8)$$

Provided that the optimal solution to (8) is $(w^*, \eta^*, b^*, \xi^*)$, then $\varepsilon = \varepsilon - 1/\eta^*$ and (8) is equivalent to

$$\begin{aligned} \min_{w, b} \quad & \frac{1}{2} \|w\|^2 + \frac{C}{l} \sum_{i=1}^l (\xi_i + \xi_i^*), \\ \text{s.t.} \quad & (w \cdot x_i) + b - y_i \leq \varepsilon^* + \xi_i, \quad i = 1, \dots, l, \\ & y_i - (w \cdot x_i) - b \leq \varepsilon^* + \xi_i^*, \quad i = 1, \dots, l, \\ & \xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, l. \end{aligned} \quad (9)$$

In the objective function of (9), minimizing $(1/2)\|w\|^2$ renders the regression function flat and increases the generalizability. $(C/l) \sum_{i=1}^l (\xi_i + \xi_i^*)$ is the total error of the regression function and should be minimized.

Setting the solution to (9) as $(\bar{w}, \bar{b})$, the regression function is then

$$y = (\bar{w} \cdot x) + \bar{b}, \quad \bar{w} = -\frac{w^*}{\eta^*}, \quad \bar{b} = -\frac{b^*}{\eta^*}. \quad (10)$$

The dual problem is then written as follows:

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_i) (\alpha_i^* - \alpha_i) (x_i \cdot x_j) \\ & + \varepsilon^* \sum_{i=1}^l (\alpha_i^* + \alpha_i) - \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i), \\ \text{s.t.} \quad & \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0, \\ & 0 \leq \alpha_i, \alpha_i^* \leq \frac{C}{l}, \quad i = 1, \dots, l. \end{aligned} \quad (11)$$

The solution to (11) is derived as $\bar{\alpha} = (\bar{\alpha}_1, \bar{\alpha}_1^*, \dots, \bar{\alpha}_l, \bar{\alpha}_l^*)^T$. In general, the solutions to (9) and (11) have the following relationship:

$$\begin{aligned} \bar{w} &= \sum_{i=1}^l (\bar{\alpha}_i^* - \bar{\alpha}_i) x_i, \\ \bar{b} &= y_j - \sum_{i=1}^l (\bar{\alpha}_i^* - \bar{\alpha}_i) (x_i \cdot x_j) \pm \varepsilon^*, \end{aligned} \quad (12)$$

where j is the sample index of the corresponding multiplier $\bar{\alpha} \neq 0$ (when it is positive) or the sample index of corresponding multiplier $\bar{\alpha}^* \neq 0$ (when it is negative).

As ε^* is a constant that is known in advance in the actual process, simplifying ε^* will not have any effect in (9), (11), and (12), and after this replacement the linear SVM machine of the insensitive loss function can be obtained.

For a wide range of nonlinear problems, the SVM maps samples to the high-dimensional Hilbert space by skillfully using nuclear transform methods and processing using linear methods. If the mapping is $\Phi : x \rightarrow \varphi(x)$, then it can be seen that (11) and (12) are only relevant to the inner product (dot product) of the sample points. Given this characteristic, the actual mapping function Φ need not be known: only the functional relationship of the sample points in the original space in the dot product of the high-dimensional space after mapping needs to be determined, which is represented by the fact that $K(x_i, x_j) = (\varphi(x_i) \cdot \varphi(x_j))$. $K(x_i, x_j)$ is a kernel function. The Gauss radial basis kernel, polynomial kernel, B-spline kernel, Fourier kernel, and Sigmoid kernel are commonly used kernel functions.

The complete algorithm of the SVM for dealing with nonlinear regression problems is as follows.

- (1) Select parameters ε and C and the appropriate kernel $K(x_i, x_j)$.
- (2) Construct the following optimization problem:

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_i) (\alpha_i^* - \alpha_j) K(x_i \cdot x_j) \\ & + \varepsilon \sum_{i=1}^l (\alpha_i^* + \alpha_i) - \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i), \end{aligned} \quad (13)$$

$$\begin{aligned} \text{s.t.} \quad & \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0, \\ & 0 \leq \alpha_i, \alpha_i^* \leq \frac{C}{l}, \quad i = 1, \dots, l. \end{aligned}$$

Find the optimal solution $\bar{\alpha} = (\bar{\alpha}_1, \bar{\alpha}_1^*, \dots, \bar{\alpha}_l, \bar{\alpha}_l^*)^T$.

- (3) Construct the decision function

$$f(x) = \sum_{i=1}^l (\bar{\alpha}_i^* - \bar{\alpha}_i) K(x_i \cdot x) + \bar{b}, \quad (14)$$

where $\bar{b} = y_i - \sum_{i=1}^l (\bar{\alpha}_i^* - \bar{\alpha}_i) (x_i \cdot x_j) \pm \varepsilon$; the positive sign is selected when $\bar{\alpha}_j \neq 0$, and the negative sign is selected when $\bar{\alpha}_i^* \neq 0$.

2.5. Construction of the SVM Regulation Model. The SVM method is applied to predict the reclamation effect of a desulfurization byproduct (gypsum) on sodic soils. The main factors influencing the reclamation process must first be determined, and then the sample datasets of observations are

FIGURE 2: Changes in the measured values of Ca^{2+} in the leachate with leaching times ((a) T1; (b) T2; (c) T3; (d) T4). The explanation of each condition (T1–T4) is reported in Table 2.

selected for training the SVM. Predictions can then be made according to the parameters obtained from the training.

Determination of the Factors of Influence. The reclamation of sodic soils with desulfurization byproducts is a multi-ion transportation and transformation process that is accompanied by ion exchange-adsorption and precipitation-dissolution reactions. Under one-dimensional vertical infiltration conditions, the factors that affect the soil solute transport include transport time, transport distance, soil texture, soil bulk density, soil moisture, ion concentration in the soil solution, ion concentration in the supply solution, and ion concentration in the solid exchange phase. When the ion concentration in the soil solution, ion concentration in the solid exchange phase, soil texture, soil bulk density, and soil texture are definite, the ion concentration in the supply solution, transport time, and transport distance can be used as the basis for predicting the soil solute transport. After determining the soil solution concentration in multiple profiles in a one-dimensional vertical transport experiment and training the model in terms of the transport time, transport distance, and corresponding soil solute content, the soil solute transport SVM control model can be established to simulate the spatial and temporal variations of the soil solute transport.

Construction of the Sample Dataset. To eliminate the effect of each factor caused by differences in dimension and units, the input and output parameters of the sample are normalized using the following equation:

$$y_i = \frac{2(z_i - z_{\min})}{z_{\max} - z_{\min}} - 1, \quad (15)$$

where z_i and y_i are the variables before and after normalization, respectively, and $z_{\min}$ and $z_{\max}$ are the maximum and minimum values, respectively, of the sample data z .

Learning and Training the SVM. The Matlab SVM toolbox is used to build and train the model. The main steps in this process include inputting the input and output parameter files, selecting the kernel function and kernel parameter, and data training and testing.

3. Results

3.1. Ca^{2+} and Na^+ Concentration in the Leachate. The measured values for the Ca^{2+} concentrations in the leachate are shown in Figure 2. When the Ca^{2+} concentrations in the supply water were 14.71, 22.06, and 29.41 mmol·L⁻¹, Ca^{2+} penetration was realized. When the Ca^{2+} concentration in

FIGURE 3: Change in the measured values of Na^+ in the leachate with leaching time ((a) T1; (b) T2; (c) T3; (d) T4). The explanation of each condition (T1-T4) is reported in Table 2.

the supply water was higher, Ca^{2+} penetration was faster. Ca^{2+} did not penetrate when the Ca^{2+} concentration was $7.35 \text{ mmol} \cdot \text{L}^{-1}$. In the initial stage, because the soil solution itself contained a certain amount of soluble Ca^{2+} , the Ca^{2+} in the supply water first reacted with the sodic salts and exchangeable Na^+ in the soil colloid in the top soil. Before the Ca^{2+} reached the bottom of the soil column, the Ca^{2+} in the soil solution was constantly leached from the soil column and the Ca^{2+} concentration in the leachate gradually decreased. However, as the Ca^{2+} gradually reached the bottom of the soil column, the Ca^{2+} concentration in the leachate gradually increased until it reached the Ca^{2+} concentration in the supply water, at which point it stabilized and realized penetration.

The measured values for the Na^+ concentrations in the leachate are shown in Figure 3. Because there was no Na^+ in the supply water, the Na^+ concentration in the leachate gradually decreased. Under higher leachate concentrations, the decrease in Na^+ was faster due to the increase in the soil's hydraulic conductivity capacity.

3.2. Prediction of the Ca^{2+} Concentration in a Penetration Experiment. An SVM model of the variations in Ca^{2+} concentration in sodic soils in a penetration test was constructed

FIGURE 4: Comparison of the measured values and predicted results for the Ca^{2+} concentration of the soil leachate in a penetration experiment.

and used for data prediction. The training process and analytical results are as follows.

The measured Ca^{2+} concentrations of the soil leachate from 0 h to 480 h in $0.5 \text{ g} \cdot \text{L}^{-1}$, $1.0 \text{ g} \cdot \text{L}^{-1}$, and $2.0 \text{ g} \cdot \text{L}^{-1}$ treatments of gypsum and from 240 h to 480 h in a $1.5 \text{ g} \cdot \text{L}^{-1}$ treatment of gypsum were used as the input data. The measured

FIGURE 5: Regression analysis of the measured values and predicted results for the Ca^{2+} concentration of the soil leachate in a penetration experiment.

FIGURE 7: Regression analysis of the measured values and predicted results for the Na^{+} concentration of the soil leachate in a penetration experiment.

FIGURE 6: Comparison of the measured values and predicted results for the Na^{+} concentration of the soil leachate in a penetration experiment.

Ca^{2+} concentrations of the leachate from 0 h to 240 h in a $1.5 \text{ g}\cdot\text{L}^{-1}$ gypsum treatment were used as the inspection data. The data were normalized, and the prediction model based on the SVM was constructed in Matlab with the following parameters: kernel function rbf (shaped as in (16)), $\delta = 0.8$, $C = 100$, $\varepsilon = 0.001$, and

$$\exp\left(\frac{-((x - x') * (x - x'))}{(2\delta^2)}\right). \quad (16)$$

The SVM model was applied to predict the data. The prediction results and measured data are shown in Figure 4. The results of a regression analysis of the measured data and prediction data are shown in Figure 5. The results show the measured data and prediction data to be strongly correlated. Clearly, the SVM model is able to correctly represent the changes in the Ca^{2+} concentration of the soil leachate in the reclamation process.

3.3. Prediction of the Na^{+} Concentration in a Penetration Experiment. The measured Na^{+} concentrations of the soil leachate from 0 h to 480 h in $0.5 \text{ g}\cdot\text{L}^{-1}$, $1.0 \text{ g}\cdot\text{L}^{-1}$, and $2.0 \text{ g}\cdot\text{L}^{-1}$ treatments of gypsum and from 240 h to 480 h in a $1.5 \text{ g}\cdot\text{L}^{-1}$ treatment of gypsum were used as the input data. The measured Na^{+} concentrations of the leachate from 0 h to 240 h in a $1.5 \text{ g}\cdot\text{L}^{-1}$ gypsum treatment were used as the inspection data. The data were normalized, and the prediction model based on the SVM was constructed in Matlab with the following parameters: kernel function rbf (shaped as in (16)), $\delta = 1.0$, $C = 100$, and $\varepsilon = 0.001$.

The SVM model was applied to predict the data. The prediction results and measured data are shown in Figure 6. The results of a regression analysis of the measured data and prediction data are shown in Figure 7. The results show the measured data and prediction data to be strongly correlated. Clearly, the SVM model is able to correctly represent the changes in the Na^{+} concentration of the soil leachate in the reclamation process.

4. Discussion

4.1. Validation of the SVM Model. The SVM method is a type of lumped parameter prediction method and has broad applicability in determining small changes in a parameter system [11, 18]. In situations where there is a lack of experimental data and system structure, such models are often able to make acceptable simulations and predictions of ion concentrations [6]. This study introduces the SVM to the complex ion change process that occurs during the application of desulfurization byproducts such as gypsum to improve sodic soils and achieves good prediction and simulation results. In addition to a visual check, measured values of Ca^{2+} and Na^{+} were compared with the results of SVM model prediction using the mean absolute error and the root mean square error. The mean absolute error (MAE) given by [19]

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |X_{mi} - X_{si}| \quad (17)$$

TABLE 3: Statistical parameters indicative of model performance.

Model	Ca ²⁺ concentration		Na ⁺ concentration	
	The root mean square error (RMSE)	The mean absolute error (MAE)	The root mean square error (RMSE)	The mean absolute error (MAE)
Support vector machine (SVM)	0.27	0.21	15.91	12.84
Artificial neural network (ANN)	0.52	0.39	26.15	18.81

describes the difference between the measured values (X_{mi}) and prediction results (X_{si}) in the units of a particular variable, with n being the number of measurement of Ca²⁺ and Na⁺. The root mean square error (RMSE) given by [19]

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (X_{mi} - X_{si})^2}{n-1}} \quad (18)$$

is the square root of the mean square error, also given in the units of a particular variable. In general, $\text{RMSE} \geq \text{MAE}$. The degree in which the RMSE value exceeds MAE is usually a good indicator of the presence and extent of outliers, or the variance of the differences between the predicted and observed values [19]. The RMSE between the measured values and prediction results of Ca²⁺ in leachate were 0.27 mmol·L⁻¹ and 15.91 mmol·L⁻¹ for Na⁺. All the RMSE were more than MAE in different degrees, but the degree in which the RMSE value exceeds MAE is low. So, the use of SVM model for prediction of ion transport in sodic soils reclaimed by CaSO₄ was reasonable.

4.2. Comparison with ANN Model. The application presented in the current paper was compared to a very well-known machine learning tool, the artificial neural network (ANN) model. The SVM can take full advantage of the distribution characteristics of the training sample, constructs a discriminant function according to the training sample, and does not need much prior information, which favors it over other nonlinear methods [18, 20]. The SVM finds the best compromise between model complexity and learning ability where sample information is limited and is highly generalizable. The ANN model uses a least squares loss function unlike SVMs which use ϵ -insensitive loss function as a fitness measure [6]. ANN method is more like an art and needs significant skill to obtain a relatively satisfactory result [21, 22]. ANN predictions are not stable and depend on the averages from various network initializations, which may yield a different result each time a model is trained. Converting the SVM algorithm into a quadratic optimization problem should in theory result in a globally optimal solution being obtained [18]. This effectively prevents the local minima problem that neural networks easily fall into, cleverly solves high-dimensional problems by nonlinear transforms and nuclear functions, makes the complexity of the algorithm independent of the sample dimension, and speeds up the training and learning. By contrast, the SVM results are stable and unique. Table 3 shows the results of the comparison between the SVM and ANN methods based on simulated

and predicted data. The RMSE using the SVM method was 0.27 mmol·L⁻¹ and 15.91 mmol·L⁻¹ for Ca²⁺ and Na⁺, respectively, but those of ANN were 0.52 mmol·L⁻¹ and 26.15 mmol·L⁻¹, indicating that SVM predicted better results than ANN. The maximum error, minimum error, and average error of the SVM results are smaller than those of the ANN [19]. Therefore, the SVM method has better capability to predict ion transport in sodic soils reclaimed by CaSO₄.

5. Conclusion

The simulation and prediction model of solute transport was constructed using Matlab toolbox based on nonlinear SVM theory, and the transport and transformation law of Ca²⁺ and Na⁺ in Ca²⁺ penetration process was carried out to simulate and predict. The following conclusions can be drawn from our findings.

- (1) The change of Ca²⁺ and Na⁺ in leachate can be reflected by SVM soil solute transport model by using Matlab toolbox, and the prediction accuracy is high.
- (2) Using SVM model to predict the spatial and temporal variations of soil solute content is feasible, and it does not require a specific mathematical model. SVM model can take full advantage of the distribution characteristics of training samples and does not require too much of a priori information and use skills.
- (3) The workload of soil solute transport prediction model based on SVM was greatly reduced without determination of the hydrodynamic dispersion coefficient and retardation coefficient, and the SVM model has a strong practicality.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (50749032), Beijing Higher Education Young Elite Teacher Project, and the Fundamental Research Funds for the Central Universities, China (2652012072).

References

- [1] S.-M. Cai, R.-Q. Zhang, L.-M. Liu, and D. Zhou, "A method of salt-affected soil information extraction based on a support vector machine with texture features," *Mathematical and Computer Modelling*, vol. 51, no. 11-12, pp. 1319–1325, 2010.
- [2] K. Lamorski, Y. Pachepsky, C. Sławiński, and R. T. Walczak, "Using support vector machines to develop pedotransfer functions for water retention of soils in Poland," *Soil Science Society of America Journal*, vol. 72, no. 5, pp. 1243–1247, 2008.
- [3] M. K. Gill, T. Asefa, M. W. Kemblowski, and M. McKee, "Soil moisture prediction using support vector machines," *Journal of the American Water Resources Association*, vol. 42, no. 4, pp. 1033–1046, 2006.
- [4] M. Khatibinia, M. J. Fadaee, J. Salajegheh, and E. Salajegheh, "Seismic reliability assessment of RC structures including soil-structure interaction using wavelet weighted least squares support vector machine," *Reliability Engineering & System Safety*, vol. 110, pp. 22–33, 2013.
- [5] N. K. C. Twarakavi, J. Šimůnek, and M. G. Schaap, "Development of pedotransfer functions for estimation of soil hydraulic parameters using support vector machines," *Soil Science Society of America Journal*, vol. 73, no. 5, pp. 1443–1452, 2009.
- [6] C. Ballabio, "Spatial prediction of soil properties in temperate mountain regions using support vector regression," *Geoderma*, vol. 151, no. 3-4, pp. 338–350, 2009.
- [7] J. M. Wang, Z. K. Bai, and P. L. Yang, "Sodic soil properties and sunflower growth as affected by byproducts of flue gas desulfurization," *PloS ONE*, vol. 7, Article ID e52437, 2012.
- [8] J. M. Wang, Z. K. Bai, and P. L. Yang, "Effect of byproducts of flue gas desulfurization on the soluble salts composition and chemical properties of sodic soils," *PloS ONE*, vol. 8, Article ID e71011, 2013.
- [9] S. Chun, M. Nishiyama, and S. Matsumoto, "Sodic soils reclaimed with by-product from flue gas desulfurization: corn production and soil quality," *Environmental Pollution*, vol. 114, no. 3, pp. 453–459, 2001.
- [10] U. Sahin, S. Eroğlu, and F. Sahin, "Microbial application with gypsum increases the saturated hydraulic conductivity of saline-sodic soils," *Applied Soil Ecology*, vol. 48, no. 2, pp. 247–250, 2011.
- [11] X.-Y. Guan, S.-L. Wang, Z.-Y. Gao, and Y. Lv, "Dynamic prediction of soil salinization in an irrigation district based on the support vector machine," *Mathematical and Computer Modelling*, vol. 58, no. 3-4, pp. 719–724, 2013.
- [12] M. C. Gonçalves, J. Šimůnek, T. B. Ramos, J. C. Martins, M. J. Neves, and F. P. Pires, "Multicomponent solute transport in soil lysimeters irrigated with waters of different quality," *Water Resources Research*, vol. 42, no. 8, 2006.
- [13] T. B. Ramos, J. Šimůnek, M. C. Gonçalves et al., "Field evaluation of a multicomponent solute transport model in soils irrigated with saline waters," *Journal of Hydrology*, vol. 407, no. 1-4, pp. 129–144, 2011.
- [14] J. Šimůnek and D. L. Suarez, "Sodic soil reclamation using multicomponent transport modeling," *Journal of Irrigation and Drainage Engineering*, vol. 123, no. 5, pp. 367–376, 1997.
- [15] L. P. Reading, T. Baumgartl, K. L. Bristow, and D. A. Lockington, "Applying HYDRUS to flow in a sodic clay soil with solution composition-dependent hydraulic conductivity," *Vadose Zone Journal*, vol. 11, no. 2, 2012.
- [16] J. C. Seaman, H. Chang, S. Goldberg, and J. Šimůnek, "Reactive transport modeling," *Vadose Zone Journal*, vol. 11, no. 2, 2012.
- [17] J. Šimůnek, M. T. van Genuchten, and M. Sejna, "HYDRUS: model use, calibration, and validation," *Transactions of the Asabe*, vol. 55, no. 4, pp. 1261–1274, 2012.
- [18] M. Kovačević, B. Bajat, and B. Gajić, "Soil type classification and estimation of soil properties using support vector machines," *Geoderma*, vol. 154, no. 3-4, pp. 340–347, 2010.
- [19] J.-D. Wang, S.-H. Gong, D. Xu, S. Juan, and J.-X. Mu, "Numerical simulations and validations of water flow and heat transport in a subsurface drip irrigation system using HYDRUS-2D," *Irrigation and Drainage*, vol. 62, no. 1, pp. 97–106, 2013.
- [20] R. M. Patel, S. O. Prasher, P. K. Goel, and R. Bassi, "Soil salinity prediction using artificial neural networks," *Journal of the American Water Resources Association*, vol. 38, no. 1, pp. 91–100, 2002.
- [21] A. Sarangi, M. Singh, A. K. Bhattacharya, and A. K. Singh, "Subsurface drainage performance study using SALTMOD and ANN models," *Agricultural Water Management*, vol. 84, no. 3, pp. 240–248, 2006.
- [22] S. Kanzari, M. Hachicha, R. Bouhlila, and J. Battle-Sales, "Characterization and modeling of water movement and salts transfer in a semi-arid region of Tunisia (Bou Hajla, Kairouan)—salinization risk of soils and aquifers," *Computers and Electronics in Agriculture*, vol. 86, pp. 34–42, 2012.