

# APPLICATION OF SUPPORT VECTOR MACHINE BASED ON TIME SERIES FOR SOIL MOISTURE AND NITRATE NITROGEN CONTENT PREDICTION

Shaoe Yang<sup>1,2</sup>, Yuanfang Huang<sup>1,\*</sup>

<sup>1</sup> Key laboratory of plant-soil interactions, MOE; Key laboratory of soil and water, MOA; College of Resources and Environment, China Agriculture University, Beijing, P. R. China 100193

<sup>2</sup> Institute of remote sensing applications Chinese academy of sciences, Beijing, P. R. China 100101

\* Corresponding author, Address: College of Resources and Environment, China Agriculture University, Beijing, 100193, P. R. China, Tel: +86-010-62732963, E-mail: yfhuang@china.com

**Abstract:** Support Vector Machine based on Time Series (SVM-TS) was applied to predict soil moisture and nitrate nitrogen (NO<sub>3</sub><sup>-</sup>-N) content. For the prediction of soil moisture, the statistical result (t-test) indicate that there is no obvious difference between predicted and observed values in 0-20cm and 20-60cm soil layers, and that SVM-TS is capable for soil moisture prediction. For the prediction of NO<sub>3</sub><sup>-</sup>-N content, there is no obvious difference between predicted and observed values in 0-30cm soil layer, although the difference in 30-60cm soil layer is obviously, compared to the variability of observed value, a large predicted error is acceptable, SVM-TS is applicable for NO<sub>3</sub><sup>-</sup>-N content prediction.

**Keywords:** support vector machine, time series, soil moisture, nitrate nitrogen

## 1. INTRODUCTION

Prediction of soil moisture and nitrate nitrogen (NO<sub>3</sub><sup>-</sup>-N) content can supervise irrigation, fertilization and reduce pollution. However, alteration of

soil moisture and  $\text{NO}_3^-$ -N content is uncertain, nonlinear and complicated. Process-based predictive models rely on accurate, reliable physics theories, but their reliability can be diminished by the paucity of data, assumptions, and limited information (Zhang Shirong, 2002). To overcome these limitations, some approximation methods were utilized as surrogate for process-based models.

Many attempts have been made to simulate soil moisture in large scope (Peng Shizhang et al. 2004, Shen Huijuan et al. 2003). In specifically region, alterations of soil moisture and  $\text{NO}_3^-$ -N content close to time series principle cause by the well-regulated precipitation and traditional cultivate manners; it makes the simple time series predictive models practicable (Yang Shaohui et al. 2006, Liu Hongbin et al. 2004). Since fluctuant soil moisture and  $\text{NO}_3^-$ -N content affected by many environmental efforts, relatively factors should be considered in time series models for more accurate estimate. Artificial neural network (ANN) is a common tool for modeling complex "input-output" dependencies; it is another feasible method for soil moisture prediction (Zhou Liangchen et al. 2005, Liu Hongbin et al. 2003, Fan Aiwu et al. 2002, Jin Long et al.1998). The advantage of ANN is that it does not require knowledge of the mathematical form of the relationship between the inputs and corresponding outputs, but it is limited by the potential to fail to generalize well when trained with limited data (Khalil A. et al. 2005).

Support Vector Machine (SVM) was proposed by Vapnik which rely on the statistical learning theory (SLT) (Vapnik V. N.1995,1998a), it can effectively avoid overfitting and multidimensional problem. Good performances have been reported in many SVM applications (Khalil A. et al. 2005, Wang Jinglei et al. 2003, Li Qingguo et al. 2005, Shie-Yui liong et al. 2002).

The objective of this work is to combine the advantages of time series model and SVM algorithm, explore the applicability of Time Series Based Support Vector Machine (SVM-TS) model for prediction of soil moisture and  $\text{NO}_3^-$ -N content.

## 2. MATERIALS AND METHODS

### 2.1 Support vector machine

SVM's essential idea is the linearly separability, or hyperplane classifier. Assume a training set like  $(x_1, y_1)$ ,  $(x_2, y_2)$ , ...,  $(x_i, y_i)$  with the input  $x_i \in R^n$  and the output  $y_i \in R$ , the goal of linear SVM is to construct a function  $y = f(x) = \omega \cdot x + b$  With  $\omega, x \in R^n$ ,  $b \in R$ . According to

statistical learning theory and Lagrangian theory (Dongil Kim et al. 2008), linear SVM regression function can be deduced to formulate

as  $f(x) = \sum_{i=1}^l (a_i^* - a_i)(x_i \cdot x) + b$ . For nonlinear problems, SVM employ a

kernel function like  $K(x, y) = \varphi(x) \cdot \varphi(y)$  which follows Mercer's theory, it makes SVM regression function can be transformed

to  $f(x) = \sum_{i=1}^l (a_i^* - a_i)K(x_i, x) + b$ , each feature vector is projected into a

higher dimensional feature space; then SVM be able to utilized a hypothesis space of linear functions in a high dimensional feature space. The popular kernels contain linear kernel, polynomial kernel  $K(x, y) = [s(x \cdot y) + c]^d$

with  $d \in N$ , radial basis function (RBF)  $K(x, y) = \exp(-s\|x - y\|^2)$  and sigmoid kernel  $K(x, y) = \tanh[s(x \cdot y) + c]$ .

In this study, SVMdark which is based on SVMlight is chose as the SVM learning software (Alex Smola et al.). SVMdark provides four kernel functions (linear, polynomial, RBF, sigmoid) for both classification and regression. It can reduce the number of potential models by "zooming in" on those which perform well on the validation set according to the training mean squared error (MSE), and select the parameters that performed optimally on the validation set.

2.2 **Support vector machine based on time series**

A time series is a sequence of data consisting of continuous values changing with time. Any time series can to be split to trend component, periodical component and noisy component. Preparatory analysis showed that the fluctuation of soil moisture and  $\text{NO}_3^-$ -N content is a nonstationary time series process with the annual periodicity.

Assume  $x_i \in R^n$  are the variables for soil moisture and  $\text{NO}_3^-$ -N content prediction,  $y_i$  is the predictive value of soil moisture and  $\text{NO}_3^-$ -N content, the purpose of SVM is to establish the relationship between  $x_i$  and  $y_i$  as follow:

$$f: R^n \rightarrow R \tag{1}$$

$$y_i = f(x_i) (i = 1, \dots, k) \tag{2}$$

In this work,  $R^n$  is the trend component and periodical component of the time series in soil moisture and  $\text{NO}_3^-$ -N content fluctuation,  $R$  is the

predictive value of soil moisture and  $\text{NO}_3^-$ -N content. Trend component ( $T_{j,k}$ ) and periodical component ( $S_k$ ) was abstracted as follow:

$$T_{j,k} = \frac{1}{p} \sum_{k=1}^p x_{j,k} \quad (1 \leq j \leq q) \quad (3)$$

$$S_k = \frac{1}{q} \sum_{j=1}^q (x_{j,k} - T_{j,k}) \quad (1 \leq k \leq p) \quad (4)$$

Where  $j$  is the number of years,  $k$  is the number of observed values,  $x_{j,k}$  is the observed value,  $p$  is the number of observed value in one period,  $q$  is the number of periods.

### 2.3 Data measurement

Experiment was performed in Quzhou experimental station of China Agricultural University in Chinese Hebei province. Experimental area is one hectare and there are winter wheat-summer maize planting rotations. Three soil sections were set and sign as section-A, section-B and section-C. In these three sections, moisture was measured in 0-20cm and 20-60cm layers,  $\text{NO}_3^-$ -N content was measured in 0-30cm and 30-60cm layers. Soil moisture was measured by neutron probe and  $\text{NO}_3^-$ -N content was measured in Analytical Conssle-TRAACS2000 about once a week. There was one more measurement 24 hours later after fertilize, irrigate or rain. The experiment was last from 1999 to 2002 and March to October in each year.

For soil moisture study, 16 observed values were selected in each year in layer 0-20cm, 14 observed values were selected in each year in layer 20-60cm. For  $\text{NO}_3^-$ -N content study, 15 observed values were selected in each year in both layer 0-30cm and 30-60cm. Observed values extension of soil moisture and  $\text{NO}_3^-$ -N content is showed in table 1; and the difference between the three sections is estimated by relative error is showed in table 2. Relative error is calculated as follow:

$$\frac{\sum_{i=1}^n \left( \frac{|x_{p,i} - x_{q,i}|}{(x_{p,i} + x_{q,i})/2} \right)}{n} \quad (5)$$

Where  $p, q$  presents different section,  $i$  is the number of observed value,  $n$  is the total number of observed values.

Table 1. Observed value extension of soil moisture and NO<sub>3</sub><sup>-</sup>-N content

	Layer(cm)	Extension	Section-A	Section-B	Section-C
Moisture	0-20	Maximum(cm <sup>3</sup> /cm <sup>3</sup> )	0.419	0.443	0.487
		Minimum(cm <sup>3</sup> /cm <sup>3</sup> )	0.113	0.084	0.084
	20-60	Maximum(cm <sup>3</sup> /cm <sup>3</sup> )	0.424	0.402	0.394
		Minimum(cm <sup>3</sup> /cm <sup>3</sup> )	0.148	0.236	0.144
NO <sub>3</sub> <sup>-</sup> -N	0-30	Maximum(mg/l)	218.37	78.83	156.02
		Minimum(mg/l)	2.05	0.21	0.35
	30-60	Maximum(mg/l)	142.39	44.15	180.42
		Minimum(mg/l)	0.11	0.05	0.24

Table 2. Relative errors of soil moisture and NO<sub>3</sub><sup>-</sup>-N content observed values between soil sections

Layer(cm)	Section-A to section-B		Section-B to section-C		Section-C to section-A		
	Mean relative error(%)	Max relative error(%)	Mean relative error(%)	Max relative error(%)	Mean relative error(%)	Max relative error(%)	
Moisture	0-20	21.30	69.76	25.64	75.68	27.72	108.60
	20-60	10.39	69.58	22.86	52.37	18.53	58.60
NO <sub>3</sub> <sup>-</sup> -N	0-30	122.64	194.51	93.89	191.16	83.14	176.15
	30-60	131.74	204.45	105.09	206.66	91.22	197.47

## 2.4 Dataset setting

In this work, trend component and periodical component of soil moisture and NO<sub>3</sub><sup>-</sup>-N content were set as the inputs of SVM, soil moisture and NO<sub>3</sub><sup>-</sup>-N content are the outputs. For soil moisture study, in layer 0-20cm, 48 data groups between year 1999 and 2001 were setup as training set, 8 data groups between March and June in 2002 were setup as validation set, 8 data groups between July and October in 2002 were setup as test set; In layer 20-60cm, 42 data groups between year 1999 and 2001 were setup as training set, 7 data groups between March and June in 2002 were setup as validation set, 7 data groups between July and October in 2002 were setup as test set. For NO<sub>3</sub><sup>-</sup>-N content study, 45 data groups between year 1999 and 2001 were setup as training set, 8 data groups between March and June in 2002 were setup as validation set and other 7 data groups between July and October in 2002 were setup as test set, in both 0-30cm layer and 30-60cm layer.

## 2.5 WHNCS model

For the purpose of compared with physical model, here present the WHNCS model (Zhang Shirong. 2002). WHNCS is a component object model for the joint simulation of one dimensional soil water, heat and nitrogen behaviors, it is capable for soil moisture and NO<sub>3</sub><sup>-</sup>-N content

prediction in different soil layers. WHNCS is based on the physical models including crop evaporation model, soil water solution model, soil NO<sub>3</sub>--N transport model and soil temperature estimate model (Zhang Shirong. 2002).

### 3. RESULTS AND DISCUSSION

Trainings were based on the training sets and validation sets; and selected the kernels and parameters according to the training mean squared errors. Selected kernels and relevant parameters for soil moisture and NO<sub>3</sub>--N content prediction are showed in table 3.

Table 3. Selected kernels and parameters for soil moisture and NO<sub>3</sub><sup>-</sup>-N content prediction

	Section Layer(cm)	Kernel	<i>C</i>	$\epsilon$	<i>d</i>	<i>s</i>	<i>c</i>	
Moisture	A	0-20	polynomial	830256.1	0.105	4	2.130	3.813
		20-60	linear	195745.7	0.008	—	—	—
	B	0-20	linear	910702.8	0.016	—	—	—
		20-60	linear	540665.9	0.038	—	—	—
	C	0-20	polynomial	612353.9	0.060	3	4.215	3.486
		20-60	linear	497573.8	0.032	—	—	—
NO <sub>3</sub> <sup>-</sup> -N	A	0-30	polynomial	489455.9	0.184	1	0.823	0.866
		30-60	polynomial	620502.3	0.408	3	3.904	4.871
	B	0-30	polynomial	688345.0	0.411	1	0.651	2.784
		30-60	linear	424298.8	0.045	—	—	—
	C	0-30	linear	149357.6	0.057	—	—	—
		30-60	polynomial	305703.9	0.174	4	3.093	0.759

Apply the selected kernels and parameters to predict soil moisture and NO<sub>3</sub>--N content for the test sets. Predicted results are showed in table 4. For soil moisture prediction, all statistical results (t-test) present  $|t| < t_{0.05}$ , it indicates that there is no obvious difference between predicted values and observed values, SVM-TS is capable for soil moisture prediction in 0-20cm and 20-60cm layers. For NO<sub>3</sub>--N content prediction,  $|t| < t_{0.05}$  in 0-30cm layer, but  $|t| > t_{0.05}$  in 30-60cm layer in section-A and section-B, there is no obvious difference between predicted values and observed values in 0-30cm layer, but the difference is obvious in 30-60cm layer. It supposes that SVM-TS is capable for NO<sub>3</sub>--N content prediction in surface soil layer but not capable in deep layer. That may cause by the more complicated transformation in deep soil layer than in the surface soil.

Table 4. SVM-TS predicted results for soil moisture and NO<sub>3</sub><sup>-</sup>-N content

	Section	Layer (cm)	Mean relative error (%)	Max relative error (%)	t value
Moisture	A	0-20	18.41	26.26	-0.527
		20-60	26.59	77.82	0.056
	B	0-20	15.36	34.83	-0.318
		20-60	6.82	13.61	-1.551
	C	0-20	17.86	29.97	-0.491
		20-60	15.57	26.49	-0.679
NO <sub>3</sub> <sup>-</sup> -N	A	0-30	84.55	291.75	1.126
		30-60	31.38	44.07	-3.485**
	B	0-30	37.01	114.16	-0.363
		30-60	48.72	89.87	-3.117**
	C	0-30	50.80	75.15	-1.440
		30-60	69.14	92.01	-1.105

Note: in 0-20cm layer sample number is 8,  $t_{0.05}=2.145$ ,  $t_{0.01}=2.977$ ; else sample number is 7,  $t_{0.05}=2.179$ ,  $t_{0.01}=3.055$ .

Although most of the statistical results show that there is no obvious difference between predicted values and observed values, but the mean relative errors are quite large. Compared with the relative errors of observed values between different sections in table 2, the maximum of observed relative errors of soil moisture is 27.72%, and 131.74% of NO<sub>3</sub><sup>-</sup>-N content, predicted values' relative errors is smaller with the maximum of soil moisture is 26.59% and 84.55% of NO<sub>3</sub><sup>-</sup>-N content. It indicates that the large predicted mean relative errors are acceptable; SVM-TS can be another capable new method for soil moisture and NO<sub>3</sub><sup>-</sup>-N content prediction.

The data is also applied to WHNCS model to predict soil moisture and NO<sub>3</sub><sup>-</sup>-N content. Because of the absence of model parameters of soil water, NO<sub>3</sub><sup>-</sup>-N content and soil temperature, we only predict soil moisture and NO<sub>3</sub><sup>-</sup>-N content between June and December in 1999 in section-C; and compared the predicted mean relative error with SVM-TS in Fig.1 and Fig.2 They show that the SVM-TS predicted results are better than WHNCS model for both soil moisture and NO<sub>3</sub><sup>-</sup>-N content prediction.

WHNCS model is base on reliable physics theories, considering water, heat and solution conditions, but numbers of parameters are required and much strict limits for model parameters lead to the limit of accuracy. Reversely, SVM-TS model is simple but practical, superior for soil moisture and NO<sub>3</sub><sup>-</sup>-N content prediction.

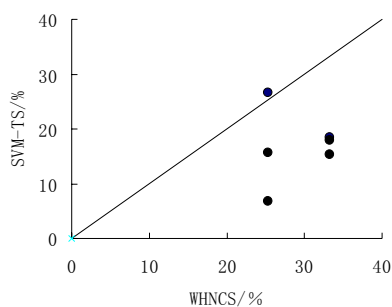


Fig. 1. Predicted mean relative error comparison between SVM-TS and WHNCS for soil moisture

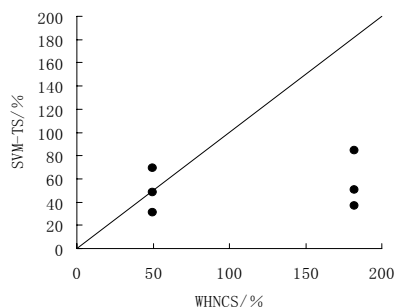


Fig. 2. Predicted mean relative error comparison between SVM-TS and WHNCS for NO<sub>3</sub><sup>-</sup>-N content

#### 4. SUMMARY AND CONCLUSIONS

Support vector machine base on time series (SVM-TS) was applied to predict soil moisture and NO<sub>3</sub><sup>-</sup>-N content in this work. Statistical results indicated that SVM-TS is capable for soil moisture prediction; and capable for NO<sub>3</sub><sup>-</sup>-N content prediction in surface layer but not capable in deep soil layer. Although the predicted mean relative errors are quite large, but it is acceptable compared with the relative errors of observed values between different sections. Meanwhile, predicted results indicate that the overall performance of SVM-TS was better than of the WHNCS physical model.

#### ACKNOWLEDGEMENTS

The authors acknowledge the financial support provided by the National Natural Science Foundation of China (No. 40671084), The 863 Program (2008AA10Z216, 2006AA10Z224), Beijing Natural Science Foundation (6072017), the National Key Technologies R&D Program (2006BDA10A01), and by the Program for New Century Excellent Talents in University (NCET-06-0107).

#### REFERENCES

- Alex Smola, Bernhard Schölkopf. <http://www.kernel-machines.org/index.html>  
 Dongil Kim, Hyoung-joo Lee, and Sungzoon Cho. Response Modeling with Support Vector Regression, *Expert Systems with Applications*, 2008 (34), 1102–1108



*Application of Support Vector Machine Based on Time Series for Soil Moisture and Nitrate Nitrogen Content Prediction* 2045

- Fan Aiwu, Liu Wei, Long Yan. Study on the Prediction of Soil Water Content Based on BP Network, *Journal of Huazhong University of Science and Technology (Natural Science Edition)*, 2002, 30(5):85-87
- Jin Long, Luo Ying, Miao Qilong, et al. Forecast Model of Farmland Soil Moisture by Artificial Neural Network, *Acta Pedologica Sinica*, 1998, 35(1):25-32
- Khalil A, Almasri M. N, McKee M, et al. Applicability of statistical learning algorithms in groundwater quality modeling. *Water Resources Research*, 2005, 41:5, W05010
- Li Qingguo, Chen Shouyu. A SVM regress forecasting method based on the fuzzy recognition theory, *Advances in Water Science*, 2005, 16(5):741-746
- Liu Hongbin, Wu Wei, Wei Chaofu, et al. Soil Water Dynamics Simulation by Autoregression Models, *Journal of Mountain Science*, 2004, 22(1):121-125
- Liu Hongbin, Wu Wei, Wei Chaofu. Study of Soil Water Forecast with Neural Network, *Journal of Soil and Water Conservation*, 2003, 17(5):59-62
- Peng Shizhang, Hu Ling, Zhang Lixin. Research Status and Dynamic Analysis of Irrigation Forecast in Domestic and Abroad, *Journal of Irrigation and Drainage*, 2004, 23(4):6-10
- Shen Huijuan, Yan Changrong, Dai Yaping. Progress and Application of Soil Moisture Monitoring and Forecasting Models, *Ecologic Science*, 2003, 22(4):366-370
- Shie-Yui liong, Chandrasekaran Sivapragasam. Flood stage forecasting with support vector machines, *Journal of the American Water Resources Association*, 2002, 38(1):173-186
- Vapnik V. N. *Statistical learning theory*. John Wiley and Sons Inc, New York, 1998a
- Vapnik V. N. *The nature of statistical learning theory*. Springer, New York, 1995
- Wang Jinglei, Wu Jingshe, Sun Jingsheng, et al. Application of Support Vector Machine Method in Prediction of Ground Water Level, *Journal of Hydraulic Engineering*, 2003, 5:122-128
- Yang Shaohui, Wang Yiming, Guo Zhengqin, et al. Research on Soil Moisture Forecast Based on ARIMA Model, *Agricultural Research in the Arid Areas*, 2006, 24(2):114-118
- Zhang Shirong. GIS-Based Simulation of Regional Soil Water and Nitrogen Behavior and Analysis of Agricultural Management, PhD candidate paper of China Agriculture University, Beijing, 2002
- Zhou Liangchen, Kang Shaozhong, Jia Yunmao. Application of BP Artificial Neural Network on Prediction of Soil Water Content, *Agricultural Research in the Arid Areas*, 2005, 23(5):98-102