Soil Moisture Prediction of Caragana Shrub Based on Wavelet Analysis and NARX Neural Network

Bai Dong-Mei[1], Zhong-Sheng Guo[1,2*] and Man-Cai Guo[1]

[1]*Northwest A & F University, Yangling, China 712100, China.*

[2]*Institute of Soil and water conservation, Chinese Academy of Science and department of Water resoureces, Yangling 712100, China*

***Corresponding Author**
Zhong-Sheng Guo. Northwest A & F University, Yangling, China 712100, China.

Citation: Mei, B., D., Guo, Z., S., Guo, M., C. (2025). Soil Moisture Prediction of Caragana Shrub Based on Wavelet Analysis and NARX Neural Network. *J Water Research, 3*(3), 01-05.

Abstract
Soil moisture prediction is important for sustainable use of soil water resource and agriculture high-quality development in forestland of water-limited regions. There are some soil water models. However, there is not the best model to forecast the change of soil moisture with time in the caragana shrubland. In this research, the plant water relationship has been investigated at the same time for a long term in the caragana shrubland of semiarid region of the Loess Plateau of China. The data of soil moisture was divided and then NARX neural network was used to build model I and model II. For model I, low frequency component was the input variable, and for model II, low frequency component and high frequency component were predicted. The results showed the average relative error for model I is 3.5% and for model II is 0.3%. The average relative error of predicted soil moisture in100cm layer using model II is 0.8%, then soil water content in the 40 cm and 200 cm soil depth is selected and the forecast errors are 4.9 % and 0.4 %. The results showed that using model II to predict soil water is well． Predicting soil water using model II will be important for sustainable use of soil water resource and high-quality development.

Keywords: Semi-Arid Loess Hilly Region, Caragana Growth, Soil Water, Wavelet Analysis, NARX Recurrent Neural Network, Soil Moisture Forecast, Sustainable Use of Soil Water Resource, Agriculture High-Quality Development

1. Introduction
Soil water is one of the most basic elements of the soil fertility and the foundation for sustainable use of soil water resource and high-quality development [1-3]. The soil water resources are the soil water storage in a given soil depth, and soil water resources use limit by plant refers to the amount of water stored within the maximum infiltration depth, in which the soil moisture content in each layer is equal to the withering coefficient. When soil water resources in the maximum infiltration depth reduce to the soil water resources use limit by plant, soil drought is serious, which cause serious soil degradation, vegetation decline and crop failure because the limit of soil water carrying capacity for vegetation.

At this stage of plant growth, the plant water relationship has to be regulated and to get maximum yield and service and carry out high-quality development. So, soil moisture has been highly considered. Predicting the dynamics of soil moisture is the foundation for sustainable use of soil water resources and soil water management, drought assessment and yield estimation. There are different kinds of soil moisture models such as water balance model, the SPAC water transport model, the SPAC water heat transport model, the mathematical statistics model, the random water balance model and the stochastic soil water dynamics model and so on. Because meteorology, soil and crops have some random features in time and space, it is difficult to select the modelling parameters to predict [4].

Artificial neural network (ANN) is an analysis method dealing with global and nonlinear mapping relationship between inputs and outputs. Because of self-organizing, self-learning and

redundant fault-tolerant properties of the input data, ANN has a great advantage in the data fitting and function approximation and have widely been applied, such as the predication of soil moisture [4-7].

As a kind of LAN transformation of time and frequency, Wavelet analysis can extract the information effectively in the signal, and proceeds multi-scale refinement analysis of signal through scaling and translation to realize the high-resolution local positioning of the time domain and frequency domain. Combining with the artificial neural network and wavelet analysis method, wavelet neural network method has two sides of advantages and widely applied in the hydrological forecasting[7], water quality prediction[8]and so on. However, there are a few reports of the application of wavelet neural network method in soil moisture forecast, especially the predict accuracy of soil water content in forest land. In this paper, authors try to use wavelet and neural network model to forecast the soil moisture in caragana shrubland. The results will provide a powerful basis for estimating the soil water resources use limit by plant, soil water vegetation carrying capacity and the regulation of the relationship between plant growth and soil water and the vegetation restoration in loess hilly region [8,9].

2. Material and Methods

2.1. Site Description
This study was carried out at the Shanghuang Eco-experiment Station, located in the semiarid region of the Loess Plateau (35°59′-36°02′ N, 106°26′- 106°30′E) in Guyuan, Ningxia Hui Autonomous Region. The altitude ranges from 1,534 to 1,824 m above the sea level. Precipitation is scarce in the period from January to March, and the rainfall from June to September accounts for more than 70% of the annual precipitation. Mean rainfall measured between 1983 and 2001 was 415.6 mm with a maximum of 635 mm in 1984 and a minimum of 260 mm in 1991. The frost-free period is 152 days. The soil is mainly loamy loess, which is porous and widely distributed in the semiarid region of the Loess Plateau [6]. There is a little change of soil texture with the depth in the soil profile11. The experimental field was located in the Caragana bushland in the middle of Heici Mountain with a slope gradient of 3°, facing southeast at the station. The study object is 16-year-old Caragana (*Caragana korshinskii*) Plantation, under which the main plant species are *Stipa bungeana, Heteropappus altaicus, Artemisia giraldii and Thymus mongolicus.*

2.2. Measurement
The investigation of the existing Caragana shrubs was made at the study site in April of 2002. There are five 100m² (5×20m) standard runoff plots. The sowing amounts of Caragana were 2.0 kg, 1.5 kg, 1.0 kg, 0.5 kg and 0 kg/100m² (wasteland, control), respectively. A neutron probe, CNC503A (Beijing Nuclear Instrument Company, Beijing), was used to measure soil moisture in the soil layers from5 cm to 780 cm and measurement was made at every 20cm interval. The soil water content obtained for each measuring depth was taken to be representative for the soil layer that included the measuring point ±10 cm depth, apart from that for the 5 cm depth, which was taken to represent the upper 10 cm of soil. Neutron counts were

made for 16 seconds. The growing season of Caragana was in the period from April to October every year. Measurement was made at 15 days interval in growing season and one time per month in the period of dormancy. This paper selects the soil water contents measured in the experimental plot with the sowing amount of 1.5 kg /100 ㎡ as the research object to analysis and predict the dynamics of soil moisture in Caragana brushland.

2.3. Data Process
The MATLAB software was used to process the data. The data measured in this study basically reflect the soil moisture dynamic change trend, but it is short for neural network training and wavelet analysis, so cubic spline interpolation method was used to enlarge the measure frequency to one time a week and then got 113 sets of data.

2.3.1. Wavelet Analysis
Wavelet analysis, a kind of mathematical method based on Fourier analysis, can show the subtle structure and characteristics of measured data using wavelet transform through selecting proper scale and window functions. Wavelet transformation needs to choose suitable wavelet function such as Haar、Daubechies、Symlets、Coiflets and so on. The Daubechies wavelet function can be used to better analyze the problems of time series.

2.4. NARX Recursive Neural Network
Nonlinear Auto-regressive with Exogenous Inputs Model (NARX network) 1was a kind of recursive neural networks with delay unit and was referred to the nonlinear auto-regressive models with external input. The input layer of the network receives two kinds of signals. The inputs from the outside the network can be expressed as u(k), u(k-1), ..., u(k-p+1), and the outputs of feedback signals from the network can be expressed as y(k), y(k-1),..., y(k-p+1), and the nonlinear system of network dynamic can be expressed as y(k+1) = F(y(k), ..., y(k-p+1), u(k), ..., u(k-p+1)). The NARX network training algorithm adopted LM algorithm of BP network.

2.5. Model
Soil water data were divided and reconstruct based on the wavelet analysis method, and then authors got the general picture of the low-frequency components and high -frequency details. In model I, authors selected the low frequency component and the soil water content in a soil layer which is highly correlated with that in the predicted soil layers as input variables to build NARX recursive neural network prediction model I according to wavelet analysis. In Model II: NARX neural network was used to predict the general picture of low-frequency components and high frequency detail respectively, and then we have got the soil moisture forecast results by summing the network outputs.

3. Results and Analyses

3.1. Selection of Target Layers
The soil moisture contents in 5 to 300 cm soil layers were selected as the research target. The variation coefficient of soil moisture content in each soil layer were estimated. The variation coefficient of soil moisture content in 5 cm soil layer, CV5, is more than

30%, that is to say, CV5 > 30%, 20% < CV20-80 < 30%, 5% < CV100-180 < 20%, CV200-300 < 5%, respectively, so the 5 to 300 cm soil layers can be classified into four groups, 0 to 5 cm, 20 to 80 cm, 80 to 180 cm and 200 to 300 cm soil layer. Because soil moisture content in the 5 cm soil layer is surface soil and interfered by outside factors and the change trend of soil water in the soil layer was uncertainty in Caragana brushland, it was not be selected as the research target. Finally, the soil water contents in the rest three soil layers such as 40 cm, 100 cm and 200 cm soil layers were chosen respectively as the analysis object.

3.2. Wavelet Analysis of Soil Moisture
First, the soil water content in the 100 cm layer were selected as the research objects, and the interpolated 113 sets of data in 100 cm layer were selected as the original wavelet sequence s1, then db_2 was selected to do single scaling wavelet decomposition and reconstruction, as shown in figure 1.

In figure 1, s_1 is time series trend diagram of soil moisture data measured in 100 cm soil layer, a1 is the general picture part of low frequency and its variation trend is consistent with the original signal. The high frequency detail part is d1 after soil water data were decomposed using single scale wavelet, which reflects the change of the original signal frequency. s_1, $s_1 = a_1 + d_1$, is the synthesized reconstruction signal of decomposed by single scaling wavelet.

3.3. The Selection of Input Variable
This article selected the soil moisture content in the soil layers that highly correlated with s1 as the inputs, the correlation results at each soil layer are shown in Table1. Besides, a1 can be used as an input because it has the similar dynamic change trend with s1, so input variables of model I were as follows: 60cm, 80cm, 120cm, 140cm, a1 and output variables was s_1.

The selection of input variables in model II: (1) when predicting a_1, the soil water contents in 60cm, 80cm, 100cm, 120 cm and 140cm soil layers were inputs, a1 was outputs; (2) when the value of predicting d_1, s_1 is only related to s_1, d_1 was output variable.

3.4. Comparison Between Model I and Model II
The range of soil moisture content was volatile due to much affecting factors, so this article was only used for short-term prediction. The input variables of week N and output variables of week N + 1 form a sample set, and there was a total of 112 samples sets. Data sets was divided as follows: the data from 1 to 90 groups was taken as the training set, which used for fitting model; the data from 91 to 101 sets was taken as the validation set, which used for the prediction of error estimation in the selected model; the remaining data in II sets be used as a test set to ultimately evaluate the generalization error of selected model.

Figure 2 showed the simulation correlation analysis in training set s1 for model I and model II. of d_1 is only related to s_1, d_1 was output variable. It can be seen from Figure 2 that model II was better than model I in the overlap of the simulated values predicting d_1, s_1

was input variable for the change and the measured values and in the correlation analysis, showing that the NARX recursive neural network has better learning. ability than BP neural network. Table 2 is the validation set of forecast error analysis for the model I and model II. It can be seen from the Table 2 that average relative error of model I was 3.5% and that of model II was 0.3%, showing that the prediction accuracy using the model II was higher than that using model I.

3.5. The Applicant of Model II
To study the feasibility of actual operation of model II, in the paper soil water content from 103 to 113 weeks in the 100cm soil layer were forecasted, as shown in table 3.

Based on wavelet analysis, the model II was more suitable than model I for forecasting soil water content in the Caragana shrubland by comparing the results of correlation analysis of simulated and measured values and the prediction error analysis of validation set.

The average relative error of predicted soil water from 103 to 113 weeks is 0.8%, and absolute values of relative error of soil water in the 11 weeks were lower than 10%, showing that modelII is feasible in prediction of soil water content in the woodland.

The model after training (learning) right reflects the samples that did not occurred in training set. Learning is not simply to memorize the studied inputs, but to learn the inner regularity of environment itself embedded in the sample through the study of a finite number of being trained samples16. In order to further test the model generalization ability, moisture contents from 103 to 113 weeks in 40 cm and 200 cm soil layer were predicted. The correlation analysis of the predicted values and the measured values were shown in figure 3. The error analysis of prediction values was shown in table 4.

The table 4 showed that the curves of predicted values were consistent with the measured values in 40 cm and 200 cm soil layer. The correlation coefficient of the linear regression equation was 0.981, 0.984 respectively, and p< 0.05. The average relative errors were less than 5%, suggesting that the wavelet neural network model II has good generalization performance.

4. Discussion
After the original signal were divided into low-frequency components and high-frequency components, wavelet analysis was used to analyse the data, and fully highlighted the trend and frequency fluctuations of signals changing with time, and avoided bad learning problems of some high-frequency mutation data when directly using neural network to predict the original data.

The low-frequency component of the soil water content after wavelet analysis in the soil layers largely correlated with that were selected as the input variables in Model I to predict soil general picture of the low-frequency water content in target soil layer. Model II used NARX recursive neural network to predict the component and high frequency detail component, then combined

the outputs. Because of higher fitting precision and smaller error compared with the model I after training analysis and error analysis of verification set, model II can better predict soil moisture content in the Caragana brushland of loess hilly-gully region.

There were different change degrees of soil moisture in each soil layer. The changes of soil moisture content with soil depth gradually tend to be steady. The soil moisture in 100 cm soil layer was selected to forecast soil water content in other soil layers, and the soil moisture in 40 cm and 200 cm soil layer were chose to generalize the model. All of obtained results were ideal.

If we obtain the maximum infiltration depth and the change of withering coefficient with soil depth and soil water resources use limit by plant, the amount of water stored within the maximum infiltration depth, in which the soil moisture content in each layer is equal to the withering coefficient before forecasting soil water content in the the maximum infiltration depth in advance, we can estimate soil water resources within the maximum infiltration depth and then compare the soil water resources and soil water resources use limit by plant. If the soil water resources equal to soil water resources use limit by plant, soil drought is serious. At this stage of plant growth, the soil water seriously influences plant growth. The serious soil drought will cause soil degradation, plant death, crop failure, or waste land resource. In order to control the serious soil drought and obtain high production, we have to predict soil moisture content in the maximum infiltration depth and estimate soil water resources and soil water carrying capacity for vegetation, and then regulate the plant water relationship and improv the soil water condition by reducing plant density to increase the soil water supply from rainfall and reduce the soil water consumption if the soil water resources and soil water resources use limit by plant because soil water resources are renewable resources and only come from rainfall, there is not enough water resource to irrigation, and ground water is deep and cannot be used by plant. When regulate the relationship between soil water and plant growth, the amount of regulation equal the difference between present density minus soil water carrying capacity for vegetation to realize sustainable use of soil water resources [10-13].

5. Conclusion
According to the wavelet analysis, NARX recursive neural network model established in this paper has higher prediction accuracy of soil water content, and can well predict soil moisture content at a given soil layers with various fluctuating frequency, and has better generalization performance. This is a good model to predict soil moisture in the brushland. The results provide powerful basis for soil drying control, regulation of the relationship between soil moisture and plant growth and sustainable use of soil water in the water-limited regions.

If we estimate the maximum infiltration depth and the soil water content in the maximum infiltration depth and the change of wilting coefficient with soil depth, and estimated soil water resources, soil water resources use limit by plants in advance before predict soil

water at a given soil layers, such as maximum infiltration depth, then we can take effective measure to regulate plant water relation and control the soil and vegetation degradation by regulating the relationship between soil moisture and plant growth and realize the sustainable use of soil water and agriculture high quality development in the water-limited regions. This research needs to continue.

Authors' Contributions: Dong-Mei Bai and Zhong-Sheng Guo investigate the soil water data, Dong-Mei Bai analysed related date and wrote the paper. Zhong-Sheng Guo and edited and reviewed the paper.

Authors' Information: Bai Dong-Mei is the Lecturer, Zhong-Sheng Guo and Man-Cai Guo is the professor in the Northwest A & F University, China.

Reference
1. Guo Z.2022.Agriculture High-quality development. Encyclopedic forum in Chinese.2022,01,64-66.
2. Guo, Z. (2023). Forest restoration, resources sustainable use and high-quality sustainable management. *Glob J Ecol, 8*(1), 007-010.
3. Guo Z.S. 2025. "Innovation China" agricultural high-quality production industry service group, Chinese scientific and technological achievements.
4. Liu Hongbin, Wu Wei, & Wei Chaofu. (2003). Study of soil water forecast with neural network. *Journal of Soil and Water Conservation* , (5), 59-62.
5. Cai, L., Ma, S., Cai, H., Zhou, Y., & Liu, R. (2009). Prediction of SYM-H index by NARX neural network from IMF and solar wind data. *Science in China Series E: Technological Sciences, 52*(10), 2877-2885.
6. Liu, H., Xie, D., & Wu, W. (2008). Soil water content forecasting by ANN and SVM hybrid architecture. *Environmental monitoring and assessment, 143*(1), 187-193.
7. Guo, Q. C., & He, Z. F. (2012). Forecast model of soil water content based on artificial neural network. *J Shanxi Agric Sci, 40*, 892-895.
8. Guo, Z. S. (2021). Soil water carrying capacity for vegetation. *Land Degradation & Development, 32*(14), 3801-3811.

9. Guo, Z. (2021). Soil hydrology process and sustainable use of soil water resources in desert regions. *Water. 2021; 13* (17): 2377.
10. Guo, Z. (2020). Estimating method of maximum infiltration depth and soil water supply. *Scientific Reports, 10*(1), 9726.
11. FAO-UNESCO, I. S. R. T. C. (1988). Soil map of the world, revised legend. *World Soil Resources Report, 60*, 119.
12. Jiang, X., & Liu, H. (2004). Radial basis function networks based on wavelet analysis for the annual flow forecast. *J Appl Sci, 22,* 411-414.
13. Kseneman, M., Gleich, D., & Potočnik, B. (2012). Soil-moisture estimation from TerraSAR-X data using neural networks. *Machine Vision and Applications, 23*(5), 937-952.