

# 基于最小二乘向量机土壤水分动态模拟与分析

邓建强<sup>1</sup>, 陈效民<sup>1</sup>, 王伯仁<sup>2</sup>, 黄晶<sup>2</sup>, 杜臻杰<sup>1</sup>, 张勇<sup>1</sup>

(1. 南京农业大学 资源与环境科学学院, 江苏 南京 210095;

2. 中国农业科学院 祁阳农田生态国家野外试验站, 湖南 祁阳 426182)

**摘要:** 土壤水分动态的模拟对水分循环与农业生产中水分的合理利用与管理具有重要的意义。应用最小二乘支持向量机对加入气象因子随机变量的红壤中土壤水分动态变化进行了训练、检验及模拟。结果表明, 最小二乘支持向量机相比与神经网络方法不论是模拟性能指标还是建模的数学意义都有更好的可靠性和优越性; 本研究应用最小二乘支持向量机对土壤水分动态日变化进行了模拟, 并采用 bior 3.3 小波函数 5 层分解提取日变化趋势图进而把该研究区土壤水分日变化划分为 4 个阶段, 其结果可为研究区水分合理利用和土壤墒情的预测预报提供科学依据。

**关键词:** 最小二乘向量机; 土壤水分动态模拟; 气象因子; 小波

文献标识码: A

文章编号: 1000-288X(2009)06-0119-04

中图分类号: S152.7, TV11, TP18

## Simulation and Analysis of Soil Water Dynamic Change Based on Least Square Support Vector Machine

DENG Jian-qiang<sup>1</sup>, CHEN Xiao-min<sup>1</sup>, WANG Bai-ren<sup>2</sup>, HUANG Jing<sup>2</sup>, DU Zhen-jie<sup>1</sup>, ZHANG Yong<sup>1</sup>

(1. College of Resources and Environmental Sciences, Nanjing Agricultural

University, Nanjing, Jiangsu 210095, China; 2. National Observation and Research Station of

Farmland Ecosystem of Qiyang, Chinese Academy of Agricultural Sciences, Qiyang, Hunan 426182, China)

**Abstract:** Soil water dynamic change is significant to water cycle research and agricultural production. The least square support vector machine and the meteorological factors were used to train, test, and simulate soil water dynamic change in red soil region. Results showed that the least square support vector machine had more reliabilities and advantages of simulation performance and mathematical meaning than the neural networks. Therefore, soil water dynamic change was simulated by the least square support vector machine and its trend was extracted by bior 3.3 with five layers of wavelet decomposition. The trend of soil water dynamic change can be divided into four stages which can provide a scientific basis for the water utilization and soil moisture prediction in the study region.

**Keywords:** least square support vector machine; soil water dynamic simulation; meteorological factor; wavelet

土壤水分是作物生长、植被恢复的关键性限制因子。如何准确、便捷获取土壤水分信息对农业生产、水分平衡及土壤-植物-大气水分循环的研究具有重要意义。土壤水分在地形、土壤、植被及前期气象等自然因素综合作用下, 不断地发生演变。对于特定区域而言, 除气象因素具有随机、混沌特性, 其它自然因素在一定时期是相对稳定的, 因而气象因子变化对土壤水分演变起决定性作用。目前, 土壤水分运动模

拟主要有两类, 即确定性模型及随机性模型<sup>[1-2]</sup>。国内外学者已对这两种模型进行了相关研究, 其中随机性模型主要集中在加入不同的随机变量应用线性拟合法或非线形模拟法对土壤水分动态变化进行模拟<sup>[3-8]</sup>。本研究在此基础上, 应用最小二乘支持向量机(LS-SVM)对加入气象资料的土壤水分进行动态模拟, 并对模拟结果进行了分析, 试图为研究区水分合理利用和土壤墒情的预测预报提供科学依据。

收稿日期: 2009-04-16

修回日期: 2009-05-21

资助项目: 国家重点基础(973 计划) 研究发展计划(2005CB121103); 中国科学院南京土壤研究所土壤与农业可持续发展国家重点实验室开放课题(0751010015)

作者简介: 邓建强(1984—), 男(汉族), 内蒙古自治区乌兰查布市察哈尔右翼后旗人, 硕士研究生, 主要从事土壤水分动态模拟。E-mail: djq1dq2@sina.com。

通信作者: 陈效民(1957—), 男(汉族), 江苏省张家港市人, 博士, 教授, 主要从事水土资源利用与管理及模型模拟方面的研究。E-mail: xmchen@njau.edu.cn。

# 1 原理

## 1.1 最小二乘支持向量机(LS-SVM)

支持向量机(SVM)是一种基于统计学习理论发展起来的机器学习方法<sup>[9]</sup>。该学习方法具有理论完备、训练时间短、泛化性能好,且对小样本强非线性映射能力强等优点,已在模式识别、函数估计等人工智能领域得到较好的应用。最小二乘支持向量机(LS-SVM)是 Suykens 于 1999 年提出的支持向量机器(SVM)改进算法<sup>[10]</sup>,是将标准支持向量机算法中不等式约束转化为等式约束而得到的,解决了样本数目过大时标准 SVM 算法求解相应的二次规划优化复杂的问题。其算法的数学描述如下:

设训练数据集 $\{x_i, y_i\}_{i=1}^N$ ,  $x_i \in R$  是第  $i$  个样本的输入模式;  $y_i \in R$  是对应于第  $i$  个样本的期望输出;  $N$  为训练样本数。LS-SVM 取如下形式:

$$y(x) = w^T \Phi(x) + b \quad (1)$$

式中非线性变换  $\Phi(x)$  将输入数据映射到高维特征空间。 $w$  的维数可以是无穷维。其目标函数描述为:

$$\min J(w, e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^N e_i^2 \quad (2)$$

高维空间优化中的内积运算采用满足 Mercer 条件的核函数  $k(x, x_i)$  代替,使得  $k(x, x_i) = [\Phi(x)^T, \Phi(x_i)]$ 。常见的满足 Mercer 条件的核函数有:(1) 多项式核函数。(2) 高斯径向基函数。(3) Sigmoid 函数。在本研究中选择高斯径向基函数(RBF):

$$k(x, x_i) = \exp\left[-\frac{(x-x_i)^2}{2\sigma^2}\right] \quad (3)$$

利用最小二乘支持向量机,求得非线性回归函数的表达式为:

$$f(x) = \sum_{i=1}^N a_i k(x, x_i) + b \quad (4)$$

式中:  $a$  ——拉格朗日乘子;  $b$  ——偏置值。

## 1.2 自适应神经模糊网络(ANFIS)

自适应神经模糊网络是在模糊逻辑理论及人工神经理论的基础上建立起来的推理系统。它是将模糊系统的模糊化、模糊推理和反模糊化 3 个基本过程用人工神经网络的结构表示出来,通过神经网络相应的学习算法实现模糊逻辑系统的推理功能,同时网络的权值也具有明确的模糊逻辑意义。其中, Jang 提出的自适应模糊神经网络(ANFIS)由于实现方便有效而得到广泛应用<sup>[11]</sup>。具体理论见文献[12]。

## 1.3 BP 神经网络(BP-ANN)

神经网络具有强大的非线性映射能力,目前比较成熟且应用广泛的网络为 BP(Back-Propagation)神经网络。该网络由输入层、隐含层及输出层组成。每层

都可以包含若干个神经元,同一层各神经元之间没有相互连接,相邻层的神经元之间通过权连接。BP 神经网络是一种误差反向传播的多层前馈式网络,即将网络输出误差反向传播对网络权重及阈值进行修正,从而实现网络的映射能力。具体算法见相关文献[13]。

## 1.4 模拟性能评价指标

为了比较模拟数据序列与观测数据的差异,选择以下几种误差评价函数从不同的侧面上反映模拟效果的好坏<sup>[14]</sup>。设样本数为  $n$ ,  $\hat{y}$  为模拟值,  $y(i)$  为实测值,  $\hat{y}_m$  为模拟平均值,  $y_m$  为实测平均值,  $\sigma_y$  为模拟值标准差,  $\sigma_y$  为实测值标准差,  $i$  为样本序号。

(1) 均方根误差(RMSE):反映了模拟值对实测值的平均偏离程度。取值大于或等于零,模拟无误差时等于零。

$$RMSE = \sqrt{\frac{\sum_{i=1}^n [\hat{y}(i) - y(i)]^2}{n-1}} \quad (5)$$

(2) 平均相对误差(MRERR):反映了模拟值对实测值的平均相对误差大小,模拟无误差时等于零。

$$MRERR = \frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}(i) - y(i)|}{y(i)} \quad (6)$$

(3) 模拟精度(PA):反映了模拟值在其均值附近的偏离与观测值在其均值附近的偏离之间的相关性。 $PA \in [-1, 1]$ ,当  $PA = 1$  表明模拟值与观测值分别对自身的均值和标准差归一化后是一样的,即模拟无误差。

$$PA = \frac{\sum_{i=1}^n \{[\hat{y}(i) - \hat{y}_m][y(i) - y_m]\}}{(n-1)\sigma_y\sigma_y} \quad (7)$$

# 2 建模方法

本研究采用 MATLAB 7.5 建立了随机性模型。建模过程包括:数据收集及整理、建模方法筛选和结果分析。

## 2.1 研究区概况及数据收集

研究区位于湖南省南部的中国农科院祁阳红壤实验站(东经  $111^\circ 53'$ , 北纬  $26^\circ 45'$ ),属于典型的红壤丘陵区。海拔高度为 100 ~ 300 m,年平均气温  $18^\circ\text{C}$ ,年降雨量 1 290 mm,无霜期 295 d,年日照约 1 613 h。研究区的面积为  $216 \text{ m}^2$ ,共设置了 12 个小区(旱田小区 4 个),每个小区为  $18 \text{ m}^2$  ( $3 \text{ m} \times 6 \text{ m}$ ,每个小区中安装两根中子管)。

试验期从 2007 年 3 月到 2008 年 7 月,每 7 d 利用中子仪测定一次 0—25 cm 土壤耕作层土壤含水量。由于耕作层土壤含水量变化受外界条件影响大,且该层次的土壤含水量对植被的生长有较强的影响,

所以对土壤耕作层含水量进行建模是必要的。

本研究应用聚类法提取出旱田小区 5 根典型中子仪样管测得数据为基础数据, 后对异常数据进行剔除, 取其平均值进行模拟。由于土壤水分变化对气象因子具有滞后现象, 本研究选取气象因子分别为: 1,

3, 7, 10 d 的累积降雨量(mm); 1, 3, 7, 10 d 的平均蒸发量(mm); 1, 3, 7, 10 d 平均温度(°C); 1, 3, 7, 10 d 平均最高温度(°C); 1, 3, 7, 10 d 最低温度(°C) 作为模型学习及检验的输入变量。表 1—2 分别为土壤耕作层的基本理化性质及原基础数据相对误差统计表。

表 1 供试土壤的基本性质

pH	有机质/	容重/	总孔隙	饱和容积持水量/	田间容积持水量/	萎蔫点容积含水量/	机械组成/(g·kg <sup>-1</sup> )		
	(g·kg <sup>-1</sup> )	(g·cm <sup>-3</sup> )	度/ %	(cm <sup>3</sup> ·cm <sup>-3</sup> )	(cm <sup>3</sup> ·cm <sup>-3</sup> )	(cm <sup>3</sup> ·cm <sup>-3</sup> )	黏粒	粉砂粒	砂粒
5.12	9.02	1.16	56.23	40.60	30.12	21.03	357.31	513.48	129.21

表 2 原基础数据相对误差统计

相对误差	< = 1%	< = 2%	< = 3%	< = 4%
累积贡献率	3%	25%	83%	100%

### 2.2 网络设计、训练及检验

本研究收集 2007 年 3 月到 2008 年 7 月的气象数据及土壤容积含水量数据共 55 组数据(随机抽取 12 组数据作为检验样本, 其余 43 组数据作为训练样本)应用 LS-SVM 对土壤水分进行非线性模拟, 以 ANFIS, BP-ANN 方法作为对比。输入变量为上述 20 个气象因子, 输出变量为相对应的土壤容积含水量。LS-SVM 模型的核函数选择高斯径向基函数(RBF), 参数优选采用交叉检验法(Cross-validation), 其优选结果  $r$  (正则化参数) = 41.677 3 及  $\sigma^2$  (径向基核参数) = 67.128 7。在 ANFIS 模型中, 由于输入变量较多所以采用减法聚类法对输入变量进行降维(聚类中心为 4), 学习法采用混合学习法, 最大迭代次数 40, 误差期望值 0.02。在 BP-ANN 模型中, 输入层及输出层传递函数分别采用双曲正切(TANSIG)传递函数及线性函数(PURELN), 训练函数采用归一化共轭梯度法算法(TRAINSCG), 自适应学习函数采用梯度下降学习算法(LEARNXDM), 最大迭代次数 100, 误差期望值 0.02, 对网络进行训练, 通过调整隐含层神经元个数来确定网络结构, 确定网络结构为 20: 5: 1。图 1 和表 3 分别为 3 种模型模拟的土壤水分模型检验图及检测样本模拟性能统计表。从图 1 可以看出 3 种模拟结果基本接近检验值。(1) 结合原基础数据相对误差统计表 2 可得出 3 种模型的平均相对误差(MRERR) 在允许误差范围之内; (2) LS-SVM 模型模拟精度(PA) 高于 BP-ANN 模型, 且显著高于 ANFIS 模型, 误差均低于其它两模型。在理论上 LS-SVM 的数学意义较为明确, 应用结构风险最小化原则克服了神经网络的维数灾难和过学习问题, 模拟结果比较稳定, 所以 LS-SVM 模型比其它两模型来说具有较强的优越性。

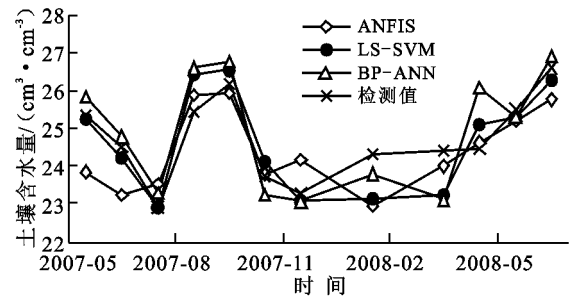


图 1 土壤水分模型检验

表 3 检验样本模拟性能统计

方法	RMSE	MRERR	PA
LS-SVM	0.657	0.019 8 ± 0.016 8	0.884
ANFIS	0.859	0.027 3 ± 0.020 1	0.733
BP-ANN	0.814	0.025 4 ± 0.019 7	0.867

### 2.3 模型模拟数据分析

根据 2.2 的分析可知应用 LS-SVM 方法对土壤水分模拟具有可靠性及可行性。应用此方法对 2007 年 3 月到 2008 年 6 月土壤水分日变化时间序列系统进行模拟, 模拟结果见图 2 和表 4。应用 bior 3.3 双正交小波对该系统进行 5 层分解以提取土壤水分日变化的趋势变化如图 3 所示, 与图 2 相比较可知, 峰与谷没有发生位移, 说明提取趋势方法是可行的<sup>[15]</sup>。通过图 3 可以把土壤水分动态变化划为 4 个阶段: (1) 土壤水分充盈期(3 月份下旬到 7 月上旬): 这个时期气温低, 蒸发量小, 且该时期降水频繁, 当土壤得到降水补给后, 土壤含水量逐渐增加并出现高峰。(2) 土壤水分亏缺期(7 月份中旬至 8 月份中旬): 这个时期降雨量少, 蒸发量高, 表层土壤迅速失水, 而下层土壤水分却未能迅速补充到上层, 所以土壤含水量出现了接近萎蔫点含水量的低谷。(3) 土壤水分补充恢复期(8 份中旬至 10 月中旬): 进入这个时期虽然温度高, 蒸发量较高, 但降雨量大, 土壤含水量逐渐回升, 同时出现含水量高峰。(4) 土壤水分干旱期

(10月中旬至3月份下旬): 进入本期后, 气温低, 降雨量少, 蒸发量低, 土壤含水量下降, 该一阶段土壤处于长时段的相对干旱期状态。

表 4 土壤水分模拟统计表

土壤含水量/ ( $\text{cm}^3 \cdot \text{cm}^{-3}$ )	$\leq 23$	$\leq 24$	$\leq 25$	$\leq 26$	$\leq 27$
累计贡献率	9.30%	34.2%	62.4%	87.1%	100%

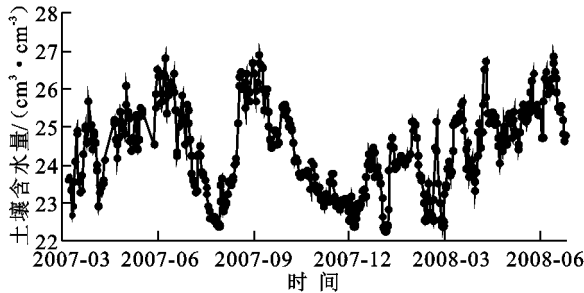


图 2 土壤水分 LS-SVM 模拟

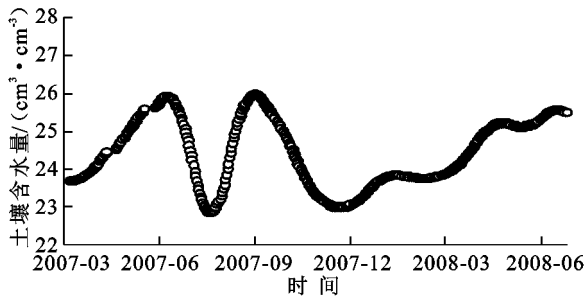


图 3 土壤水分动态变化趋势

### 3 结论

土壤水分动态变化过程具有复杂的混沌特性, 对其进行随机性模拟必须寻找各种非线性的模拟方法。本研究应用 LS-SVM 对土壤水分进行建模, 从模拟性能及建模数学意义来看, LS-SVM 建模法更具有较强可靠性。应用 LS-SVM 对土壤水分动态日变化进行模拟, 应用小波分析提取日变化趋势图的结果可以将土壤水分动态变化划分为 4 个阶段: 土壤水分充盈期、土壤水分亏缺期、土壤水分补充恢复期和土壤水分干旱期。土壤水分时期的合理划分可为研究区水分合理利用和土壤墒情预报提供科学依据。

### [ 参 考 文 献 ]

- [ 1 ] 申慧娟, 严昌荣, 戴亚平. 农田土壤水分预测模型的研究进展及应用[J]. 生态科学, 2003, 22 (4): 336-376.
- [ 2 ] 尚松浩. 土壤水分模拟与墒情预报模型研究进展[J]. 沈阳农业大学学报, 2004, 35(5/6): 455-458.
- [ 3 ] Amin Elshorbagy, Parasuraman K. On the relevance of using artificial neural networks for estimating soil moisture content[J]. Journal of Hydrology, 2008, 362(1/2): 1-18.
- [ 4 ] 薛晓萍, 王新, 张丽娟, 等. 基于支持向量机方法建立土壤湿度预测模型的探讨[J]. 土壤通报, 2007, 38(3): 427-433.
- [ 5 ] 周良臣. 土壤水分预测的 BP 神经网络方法及模型[J]. 西北农业学报, 2004, 13(4): 130-133.
- [ 6 ] 李兴旺, 冯宝平. 基于 BP 神经网络的土壤含水量预测[J]. 水土保持学报, 2002, 16(5): 117-119.
- [ 7 ] 刘洪斌, 武伟, 魏朝富. 基于神经网络的土壤水分预测建模研究[J]. 水土保持学报, 2003, 17(5): 59-62.
- [ 8 ] 邓建强, 陈效民, 方堃. 基于神经网络的混沌时间序列土壤墒情预测预报研究[J]. 水土保持通报, 2008, 28(6): 82-85.
- [ 9 ] Vapnik V, Golowich S, Smola A. Support vector method for function approximation, regression estimation and signal processing [C]//Advances in Neural Information Processing Systems 9. Cambridge: MIT Press, 1997: 281-287.
- [ 10 ] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers[J]. Neural Processing Letters, 1999, 9(3): 293-300.
- [ 11 ] Jang J S. ANFIS: Adaptive-network-based fuzzy inference systems[J]. IEEE Trans. on Sys., Man and Cybern., 1993, 23(3): 665-685.
- [ 12 ] 张国良, 曾静, 柯熙政, 等. 模糊控制及其 MATLAB 应用[M]. 西安: 西安交通大学出版, 2002: 6-41.
- [ 13 ] 张代远. 神经网络新理论与方法[M]. 北京: 清华大出版社, 2006: 22-26.
- [ 14 ] Chen J L, Islam S, Biswas P. Nonlinear dynamics of hourly ozone concentrations: nonparametric short term prediction [J]. Atmospheric Environment, 1998, 32(11): 1839-1848.
- [ 15 ] 郑治真. 小波变换及其 MATLAB 工具的应用[M]. 北京: 地震出版社, 2001: 93-109.