

基于支持向量机的土壤湿度模拟及预测研究*

张 强^{1,2} 黄生志^{1,2} 陈晓宏^{1,2}

(1 中山大学水资源与环境系, 广州 510275)

(2 中山大学 华南地区水循环与水安全广东省普通高校重点实验室, 广州 510275)

摘 要 基于中山大学珠海校区气象观测站日平均风速、日平均气温、日平均空气湿度、日平均水汽压、日平均总辐射量、日平均地表温度、日平均降雨量、日平均蒸发量以及日平均 10 cm、20 cm、30 cm 土壤的含水量, 利用支持向量机方法建立气象因子与土壤湿度统计关系, 并以此为基础建立土壤湿度模拟与预测模型。结果表明, 土壤湿度对气象因子有一定滞后相关性, 不同土层土壤湿度对气象因子的滞后相关性不同。研究发现考虑滞后相关性的预测模型在精度上要高于不考虑滞后相关性的预测模型。此外, 利用气象因子对地下 10 cm 的土壤湿度模拟与预测精度较高, 而对地下 20 cm、30 cm 的土壤湿度模拟精度较低。利用地下 10 cm 与 20 cm、20 cm 与 30 cm 的土壤湿度相关性大的特点, 可以考虑利用支持向量机方法以 10 cm 土壤湿度模拟与预测 20 cm 的土壤湿度, 以 20 cm 的土壤湿度模拟与预测 30 cm 的土壤湿度, 分析结果表明模拟精度较高。

关键词 支持向量机; 土壤湿度; 预测模型

中图分类号 S152.7 **文献标识码** A

土壤湿度在地表水文过程中扮演着重要角色, 它影响着地表与大气之间热量、水分、动量及化学物质的交换。土壤湿度与其他气候因子联系紧密, 可以影响土壤表面反射率及土壤热容量, 并可通过潜热与感热传导来影响短期气候变化^[1-6]。此外, 土壤湿度与农业密切相关, 土壤湿度过低, 易形成土壤干旱, 影响植物光合作用, 严重时还会导致作物凋萎甚至死亡; 土壤湿度过高会影响土壤通气性, 阻碍植物呼吸及生长等过程。因此, 对于土壤湿度的研究具有重要理论与现实意义。目前, 国内外学者对土壤湿度与蒸散发的关系及土壤湿度垂直演变特征研究较多^[1,7], 但利用气象因子对土壤湿度进行模拟与预测研究较少, 且过去的研究多利用多元线性回归方法对于土壤湿度进行模拟与预测, 陈志明^[8]利用影响草坪湿度的各因素对草坪湿度建立多元线性回归预测模型。但需要提及的是, 各气象因子间多为复杂非线性关系, 而非单一的线性关系。金龙等^[9-10]利用人工神经网络模型预报农田的土壤湿度, 虽然预报精度较高, 但是人工神经网络容易陷入维度灾, 取得局部最优值而非全局最优值。薛晓萍等^[11]利用适合解决非线性问题的支

持向量机方法建立预测精度较高的土壤湿度预测模型, 但却忽略了土壤湿度变化对各气象因子响应的滞后性, 且其预测模型参选取值是利用传统的试算方法, 运算效率低且难以找到参数的全局最优值。基于此, 此研究运用支持向量机方法对土壤湿度进行模拟与预测, 研究中考虑到土壤湿度变化对各气象因子的滞后相关性, 同时在具体建模过程中引入参数寻优算法, 达到提高运算效率、提高土壤湿度模拟与预测精度的目的。该研究可为土壤湿度模拟与分析提供思路, 相关成果可为利用气象因子模拟与预测土壤湿度提供解决途径, 在一定程度上解决水文模型中土壤湿度计算的理论难题。

1 数 据

研究数据取自中山大学珠海校区的 CR3000 气象观测站, 数据观测时间为 2011 年 3 月至 2011 年 9 月。所用数据为日平均风速、日平均气温、日平均空气湿度、日平均水汽压、日平均总辐射量、日平均地表温度、日平均降雨量、日平均蒸发量以及日平均 10 cm、20 cm、30 cm 土层土壤的含水量。排除异

* 国家自然科学基金项目(项目号:41071020;50839005)与新世纪优秀人才支持计划共同资助成果

作者简介:张 强(1974—),男,山东沂水人,博士,教授,博士生导师,主要从事流域气象水文学研究、旱涝灾害机理、流域地表水文过程及其对气候变化的响应机制与机理以及流域生态需水等领域的研究工作。E-mail:zhangq68@mail.sysu.edu.cn

收稿日期:2011-12-21;收到修改稿日期:2012-03-07

常数据,并考虑到各因子对土壤湿度的滞后影响,一共有 174 天的有效数据。用 111 天的数据进行模型构建与模拟,利用其余的 61 天的数据进行模拟与预测,检验模型模拟与预测精度。

2 研究方法

2.1 土壤湿度对气象因子变化的滞后相关性

取 4 月 4 日至 9 月 26 日的日平均风速、日平均气温、日平均空气湿度、日平均水汽压、日平均总辐射量、日平均地表温度、日平均降雨量、日平均蒸发量分别对日平均 10 cm、20 cm 以及 30 cm 土壤含水量求相关系数。先分别求出 4 月 4 日至 9 月 26 日的土壤湿度与各因子的相关系数,然后固定土壤湿度不变,其他因子分别逐个向 4 月 3 日、4 月 2 日直至 3 月 14 日依次前移,并分别求出其与土壤湿度的相关系数。然后根据各气象因子与土壤湿度的正负相关性,筛选出其相关系数绝对值最大点,该点对应的的时间即为土壤湿度对应该因子的滞后时间。

2.2 原始数据的标准化处理

在数值分析中,常用标准化方法主要有以下几种:

2.2.1 总和标准化 总和标准化由下式计算:

$$\hat{x} = \frac{x_{ij}}{\sum_{i=1}^m x_{ij}} \quad (i = 1, 2, \dots, m; j = 1, 2, \dots, n) \quad (1)$$

式中, x_{ij} 表示数据矩阵中第 i 行第 j 列的值,这种标准化方法所得到的新数据满足

$$\sum_{i=1}^m \hat{x}_{ij} = 1 \quad (j = 1, 2, \dots, n) \quad (2)$$

2.2.2 标准差标准化 其转换公式为:

$$\hat{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j} \quad (i = 1, 2, \dots, m; j = 1, 2, \dots, n) \quad (3)$$

式中, S_j 为第 j 列数据的标准差, \bar{x}_j 为第 j 列数据的平均值。由这种标准化方法所得到的新数据各要素的平均值为 0, 标准差为 1。

2.2.3 极大值标准化 其转换公式为:

$$\hat{x}_{ij} = \frac{x_{ij}}{\max\{x_{ij}\}} \quad (i = 1, 2, \dots, m; j = 1, 2, \dots, n) \quad (4)$$

经标准化所得的新数据极大值为 1。

2.2.4 极差标准化(正规化方法) 其转换公式为:

$$\hat{x}_{ij} = \frac{x_{ij} - \min_i\{x_{ij}\}}{\max_i\{x_{ij}\} - \min_i\{x_{ij}\}} \quad (i = 1, 2, \dots, m; j = 1, 2, \dots, n) \quad (5)$$

标准化后的数据极大值为 1, 极小值为 0。

以上四种数据标准化方法均可以消去参数量纲及数量级不同而产生的误差。本文所采用的标准化方法为极差标准化法。

2.3 数据的随机排列

本文所用时间序列长度为 2011 年 3 月至 2011 年 9 月, 序列长度较短, 本文对标准化后的数据做随机采样, 以增强数据代表性、一致性及可靠性。

2.4 基于支持向量机的土壤湿度模拟与预测方法

2.4.1 支持向量机方法^[11-12] 支持向量机(Support Vector Machines, 简称 SVM)是由 Vapnik 等在 20 世纪 90 年代得出并迅速发展起来的一种基于统计学习理论的机器学习算法。支持向量机以统计学习理论为理论体系, 通过寻求结构风险最小化来实现风险的最小化, 追求在有限信息的条件下得到最优结果^[12]。以往困扰机器学习方法的很多问题, 如模型的选择与学习问题、非线性和维数灾难问题、局部极小值问题等, 均可以通过支持向量机的方法得到一定程度的解决。近几年, 支持向量机理论的不断发展和完善, 加上神经网络等学习方法在理论上进展缓慢, 因此支持向量机越来越受到广泛的关注^[13]。

SVM 方法是基于 Mercer 核展开定理, 通过非线性映射 φ 将样本空间映射到一个高维的特征空间(Hilbert 空间), 从而使样本空间的非线性分类与回归等问题能在特征空间中通过应用线性学习机的方法解决^[11]。其中, 支持向量是 SVM 的核心问题。在回归问题上, 最终的目的是能够找到一个回归超平面。它是将 Vapnik 提出的 ε (不敏感损失函数)作为误差度量, 从而使得对训练样本的数据拟合达到最好的状态。关于非线性的回归问题, 就是运用非线性映射 φ 的方法, 将非线性的样本空间映射到一个高维的线性特征空间, 进而将线性最优化问题转变为一定约束条件下的最大化函数^[11]。

2.4.2 影响因子的选择与处理^[11] 土壤湿度除了与自身的土壤结构息息相关以外, 还与其背景气候因子联系密切。土壤湿度变化的不确定性主要源于大气因子的随机性。因此, 本文从中山大学珠海校区的气象观测站中抽取 3 月至 9 月的逐日日平均风速、日平均气温、日平均空气湿度、日平均水汽压、日平均总辐射量、日平均地表温度、日平均降雨

量以及日平均蒸发量 8 个要素,对地下 10 cm、20 cm 以及 30 cm 土壤的含水量进行模拟与预测。

2.4.3 模型的建立 支持向量机主要有以下两项关键算法:(1)基于 SRM 原则的具有最大间隔的最优分类面设计。线性最优分类面被设计在高维特征空间中,并且输入空间的非线性学习算法是通过核函数得到;(2)核函数方法。核函数法就是用非线性变换 φ 将 n 维向量空间中随机矢量 x 映射到高维特征空间,在高维特征空间中设计线性学习算法,若其中各坐标分量间的相互作用仅限于内积,则不需要知道非线性变换 φ 的具体形式,只要用满足 Mercer 条件的核函数替换线性算法中的内积,就能得到原输入空间中对应的非线性算法。常用的满足 Mercer 条件的核函数有多项式函数、径向基函数和 Sigmoid 函数等,选用不同的核函数可构造不同的支持向量机^[13]。本文选用径向基函数,其核函数为:

$$K(x, x_i) = \exp\left\{-\frac{|x - x_i|^2}{\sigma^2}\right\} \quad (6)$$

构造的支持向量机的判别函数为:

$$f(x) = \text{sgn}\left\{\sum_{i=1}^s \alpha_i \exp\left\{-\frac{|x - x_i|^2}{\sigma^2}\right\} - b\right\} \quad (7)$$

其中 s 个支持向量机 x_i 可确定径向基函数的中心位置, s 为中心数目的。

2.4.4 用非线性支持向量机进行回归预测^[13] 首先使用一非线性映射将数据映射到一个高维特征空间,再到高维特征空间进行线性回归。关键问题就是核函数 $K(x, y)$ 的使用。此处,优化问题

$$\omega = \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i \quad 0 \leq \alpha_i, \alpha_i^* \leq C \quad (8)$$

$i = 1, \dots, l$

成为在式(1)约束下最优化函数:

$$w(\alpha_i, \alpha_i^*) = -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*) K(x_i \cdot x_j) + \sum_{i=1}^l (\alpha_i - \alpha_i^*) y_i - \sum_{i=1}^l (\alpha_i + \alpha_i^*) \varepsilon \quad (9)$$

此时:

$$\omega = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \phi(x_i) \quad (10)$$

函数 $f(x)$ 可直接表示为

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x, x_i) + b \quad (11)$$

按照 Kuhn-Tucker 定理有:

$$\begin{aligned} \alpha_i [\varepsilon + \zeta_i - y_i + f(x_i)] &= 0, i = 1, \dots, l \\ \alpha_i [\varepsilon + \zeta_i^* - y_i + f(x_i)] &= 0, i = 1, \dots, l \end{aligned} \quad (12)$$

$$\begin{aligned} \zeta_i y_i &= 0, i = 1, \dots, l \\ \zeta_i^* y_i^* &= 0, i = 1, \dots, l \end{aligned} \quad (13)$$

由线性支持向量机可得:

$$\begin{aligned} C - \alpha_i - y_i &= 0, i = 1, \dots, l \\ C - \alpha_i^* - y_i^* &= 0, i = 1, \dots, l \end{aligned} \quad (14)$$

由式(12)可见, $\alpha_i \alpha_i^* = 0$, 即任何一组的 α_i 和 α_i^* 均不会同时为非 0。由式(13)和式(14)知:

$$\begin{aligned} (C - \alpha_i) \zeta_i &= 0, i = 1, \dots, l \\ (C - \alpha_i^*) \zeta_i^* &= 0, i = 1, \dots, l \end{aligned} \quad (15)$$

由此可见,对应于 $\alpha_i = C$ 或者 $\alpha_i^* = C$ 的 $f(x_i)$ 与 y_i 的误差可能大于 ε , 对应于 $\alpha_i \in (0, C)$ 或 $\alpha_i^* \in (0, C)$ 的 $f(x_i)$ 与 y_i 的误差必然等于 ε , 即 $\zeta_i = 0$ 或 $\zeta_i^* = 0$, 因此有:

$$\begin{aligned} \varepsilon - y_i + f(x_i) &= 0, \alpha_i \in (0, C) \\ \varepsilon + y_i + f(x_i) &= 0, \alpha_i^* \in (0, C) \end{aligned} \quad (16)$$

由式(16)可以求出 b 。

2.4.5 模型参数的释义 不敏感损失参数 ε 、惩罚因子 C 以及核函数的系数 g 等参数的选取将直接影响土壤湿度的预测精度,即参数的变化会对土壤湿度样本的模拟度产生影响。

不敏感损失函数参数 ε 是模拟误差度量,它控制着支持向量的个数, ε 越大,决定最优模型的支持向量越少,模型将变得简单但是误差却增大。相反,当决定最优模型的支持向量多,虽然拟合精度增大,但会增加求解时间,且会导致拟合过度的发生。 C 是正则化参数,它是用来调节模型复杂度和训练误差的。 C 是对不敏感区域外因子的惩罚力度,它受数据的随机噪音影响较大。 C 的取值小,模型对超出不敏感损失区域的样本的惩罚力度变小,导致训练误差变大,模型的预测精度不高; C 的取值大,模型泛化能力变差。 g 与高斯核的宽度参数有关,模型预报能力与其值选取有关。当 g 很小时,核宽度系数大,模型对样本识别能力不高,学习推广能力很弱;当 g 很大时,核宽度系数小,支持向量间的联系不够密切,学习机器也比较复杂,泛化推广能力不强。此外,参数迭代步长的选择对模型选优也有较大影响。

本文通过 libsvm-mat 的参数寻优函数对 C 和 g 进行最优值寻找,其中 C 、 g 值的搜索范围设置为 2^{-8} 至 2^8 , 迭代步长设为 0.5。

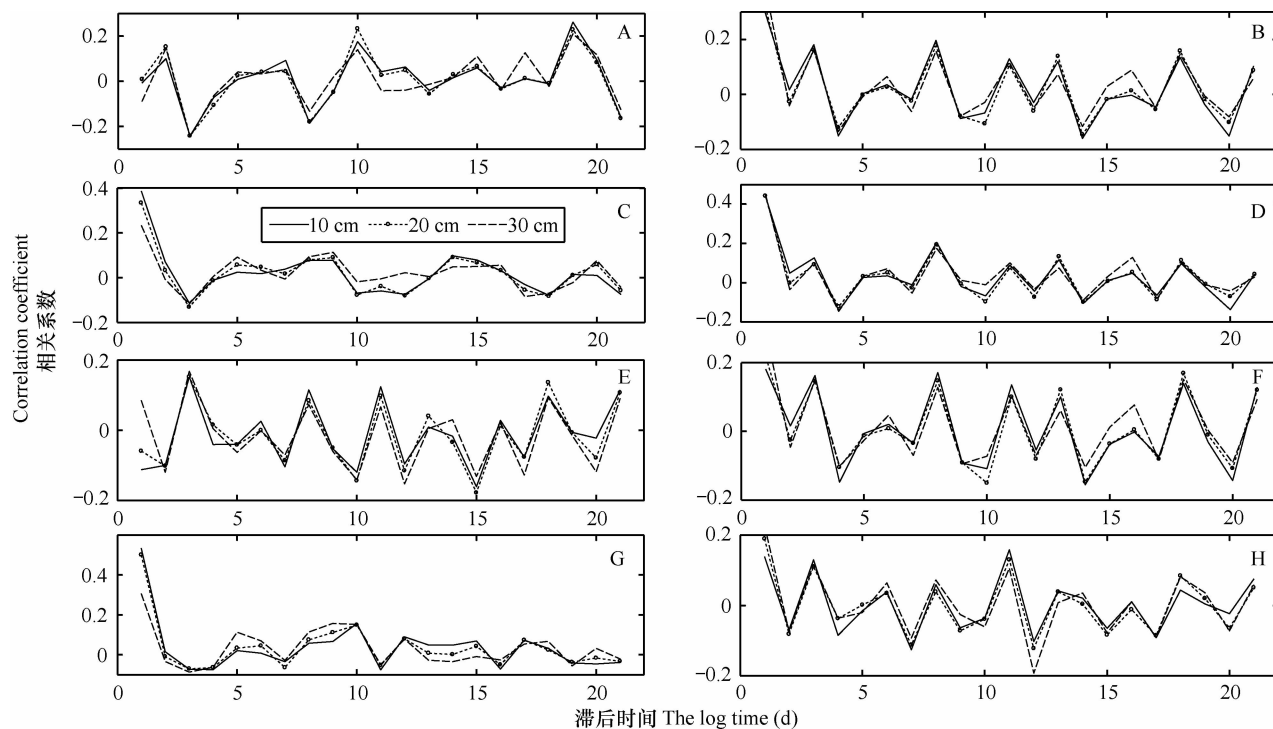
2.4.6 模型参数的确定 关于参数惩罚因子 C 以及核函数系数 g 的确定,传统的方法是通过不断地改变 C 和 g 值试算,然后取训练精度最高的点所对应的 C 和 g 值为最优。但是,传统方法运算烦

琐,运算量较大,而且未必能找到全局最优参数。基于此,本文通过参数寻优算法,结合 libsvm-mat,设置 C 和 g 值的搜索范围,并将各 C 和 g 值所对应的误差用三维图直观表达出来,可简单、准确且形象地找到最优 C 和 g 值。

3 结果

3.1 土壤湿度对各气象因子的滞后时间

分别将 4 月 4 日至 9 月 26 日的 10 cm、20 cm



注: A: 风速, B: 气温, C: 空气湿度, D: 水汽压, E: 总辐射量, F: 地表温度, G: 降雨量, H: 蒸发量 Note: A: Wind speed, B: Air temperature, C: Air humidity, D: Vapour pressure, E: Total solar radiation, F: Surface air temperature, G: Rainfall, H: Evaporation

图 1 气象因子与土壤湿度的相关系数

Fig. 1 Correlations between soil moisture and meteorological factors

由图 1 可知,不同气象因子与土壤湿度的相关系数不相同,其中大气温度、水汽压、空气湿度以及降雨量与土壤湿度的相关系数比较大,风速及总辐射量、地表温度以及蒸发量与土壤湿度的相关系数较小。

以及 30 cm 土壤湿度与日平均风速、日平均气温、日平均空气湿度、日平均水汽压、日平均总辐射量、日平均地表温度、日平均降雨量以及日平均蒸发量逐个向 4 月 3 日、4 月 2 日直至 3 月 14 日依次前移,并求其相关系数。关于土壤湿度对各因子的滞后时间的确定,先判断各因子与土壤湿度的相关性的正负。当其为正相关时,取相关系数为正的最大值的点对应的时间为滞后时间。相反,当其为负相关时,取相关系数为负的最小值的点对应的时间为滞后时间。结果如图(1)。

其中,随着土壤深度的不断增大,各因子与土壤湿度的在滞后时间点相关系数不断变小,说明气象因子对土壤湿度的影响随着深度的增大而减小。各土壤深度的土壤湿度对各气象因子的滞后时间如表 1。

表 1 各土壤深度的土壤湿度相对各气象因子的滞后时间

Table 1 Time of soil moisture at different soil depth lagging behind each meteorological factors

土壤深度 Soil depth	滞后时间 The log time (d)							
	风速 ^①	气温 ^②	空气湿度 ^③	水汽压 ^④	总辐射量 ^⑤	地表温度 ^⑥	降雨量 ^⑦	蒸发量 ^⑧
10 cm	2	3	0	0	14	3	0	6
20 cm	2	3	0	0	14	3	0	6
30 cm	2	3	0	0	14	3	0	11

①Wind speed, ②Air temperature, ③Air humidity, ④Vapour pressure, ⑤Total solar radiation, ⑥Surface air temperature, ⑦Rainfall, ⑧Evaporation

由表 1 可以发现同一因子随着土层深度的增加,其滞后时间基本不变,这有违常理。由于该实验采用的是日均值,以天(d)为单位,而 10 cm 与 20 cm、20 cm 与 30 cm 土层之间的各因子滞后时间差可能在一天以内,所以在大尺度的时间下,土壤湿度对应各气象因子的滞后性在深度上没有体现出来。

3.2 SVM 参数寻优的结果——C 与 g 的最优值

利用 libsvm-mat 中关于回归预测的参数寻优函数的调用,分别得到各土层土壤湿度在考虑滞后性与不考虑滞后性的情况下 C 与 g 参数的最优值,两参数寻优过程如图 2 所示,参数值如表 2 所示:

各模型的参数寻优结果如表 2 所示。

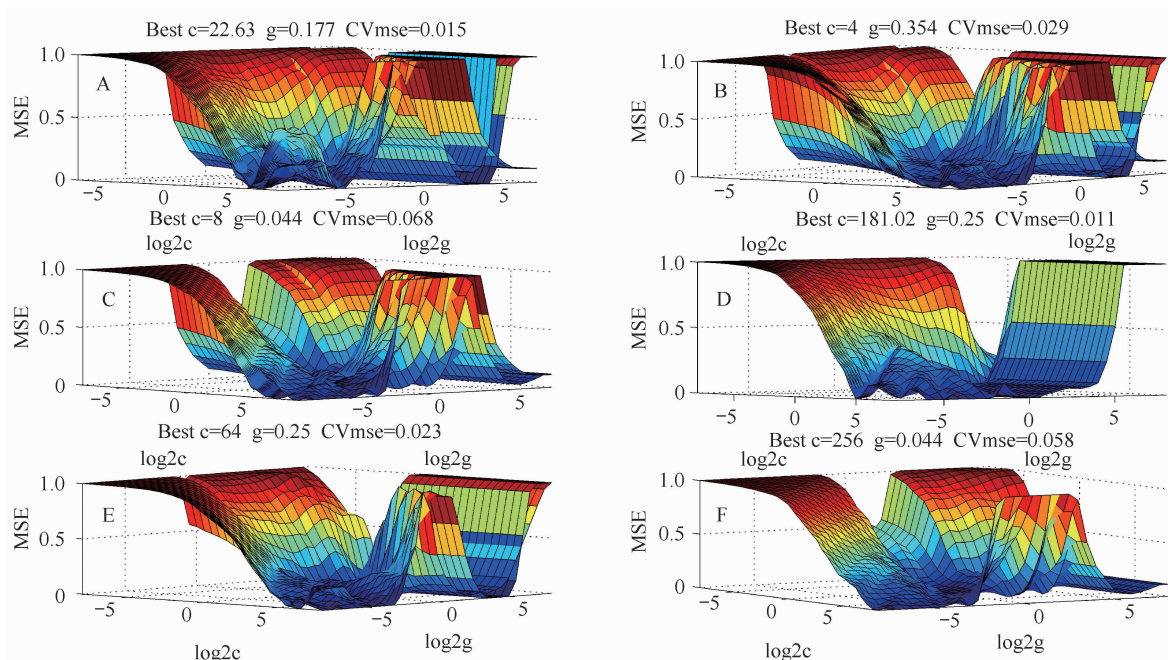


图 2 各土壤深度的土壤湿度预测模型参数寻优过程图(A, B, C:分别为不考虑滞后性 10 cm、20 cm、30 cm 的土壤深度; D, E, F:分别为考虑滞后性 10 cm、20 cm、30 cm 的土壤深度)

Fig. 2 Charts of parameters optimization of forecast of soil moisture at different soil depths (A, B and C represents the soil depth of 10 cm, 20 cm, and 30 cm, respectively, taking into account lag time respectively; D, E and F represents the soil depth of 10 cm, 20 cm, and 30 cm without taking into account lag time)

表 2 各模型的 C、g 最优值

Table 2 Optimized C and g values of different models

条件 Condition	土壤深度 Soil depth (cm)	最优 C 值 Optimal value of C	最优 g 值 Optimal value of g	平均均方差 MSE
不考虑滞后性 Regardless of lag time	10	22.63	0.177	0.015
	20	4.000	0.353	0.029
	30	8.000	0.044	0.068
考虑滞后性 Taking into account lag time	10	181.0	0.250	0.011
	20	64.00	0.250	0.023
	30	256.0	0.044	0.059

由表 2 可知考虑滞后性以后在参数寻优过程中的平均均方差较不考虑滞后性的平均均方差要小,

对数据的拟合程度要好。

3.3 用 SVM 方法对预测样本的预测结果

最优参数找到后,说明在该参数条件下对样本进行训练可以达到理想的精度。开始对土壤湿度数据的预测样本进行模拟与预测计算,并与实际观测值进行对比,分析模拟误差等,具体结果详见图 3、图 4。

由图 3、图 4 可以清楚地看到,考虑滞后性对土壤湿度的预报精度要较不考虑滞后性高很多。随着土壤深度的不断增大,其预测精度不断减小。主要原因是气象因子对土壤湿度的影响随着土壤深度的增大而减小。各模型的预测情况如表 3 所示。

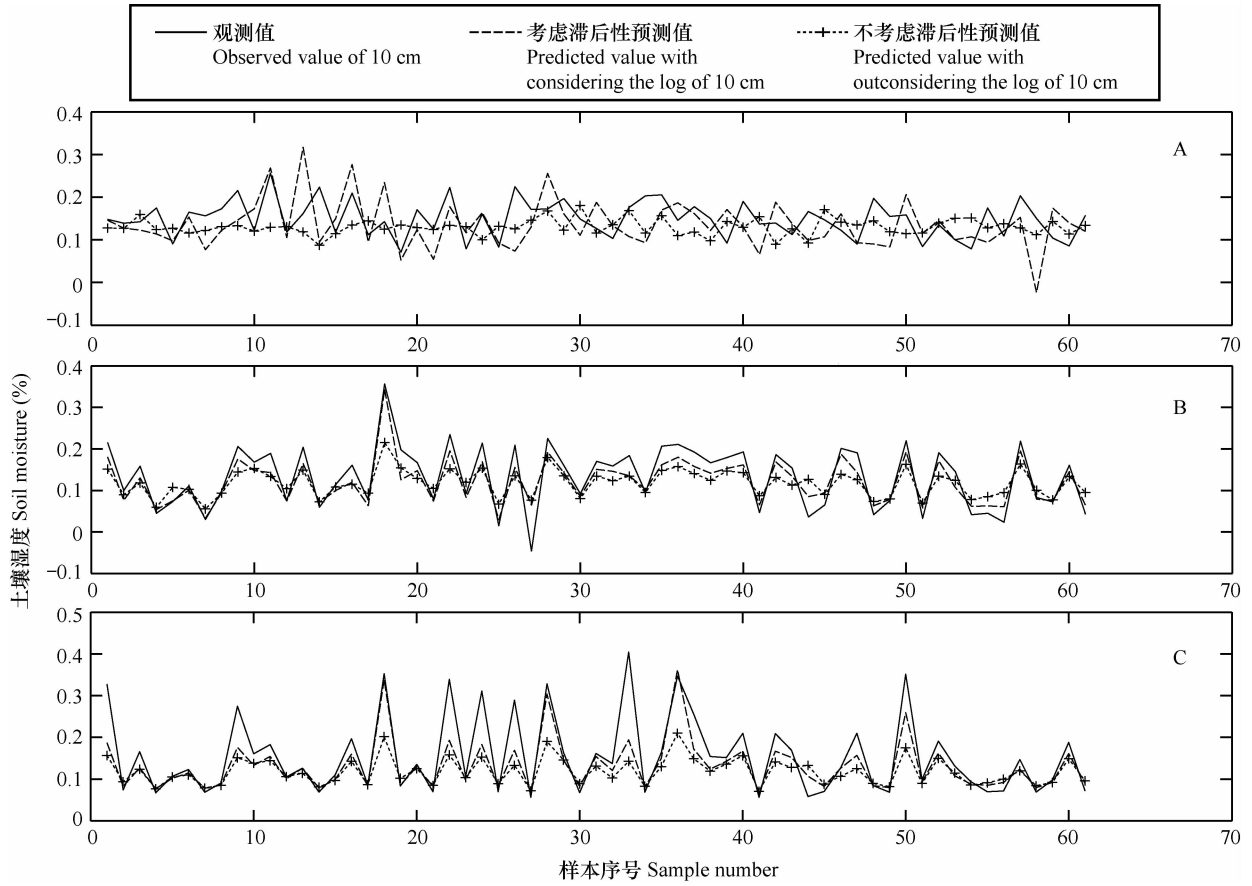


图 3 10 cm、20 cm、30 cm 土壤湿度预测值与实际值的对比图

Fig. 3 Simulated and observed soil moisture contents in soils at 10 cm, 20 cm and 30 cm in depth

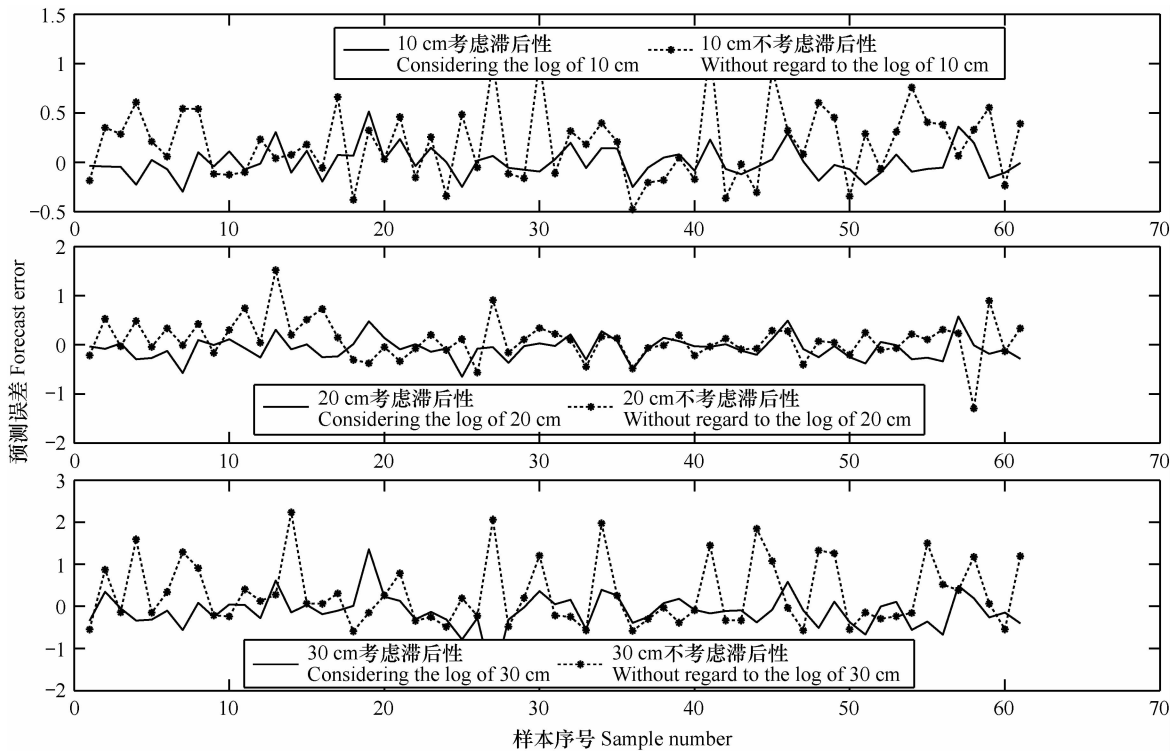


图 4 10 cm、20 cm、30 cm 土壤湿度预测误差图

Fig. 4 Errors in prediction of soil humidity at 10 cm, 20 cm and 30 cm in depth

表 3 各模型的预测情况

Table 3 Prediction variance of the models in simulation with/without taking into account lag time

条件 Condition	土壤深度 Soil depth (cm)	平均均方差 Mean squared error	平方相关系数 Squared correlation coefficient
不考虑滞后性 Regardless of lag time	10	0.012	0.359
	20	0.022	0.341
	30	0.045	0.331
考虑滞后性 Taking into account lag time	10	0.003	0.737
	20	0.011	0.668
	30	0.027	0.637

由表 3 可以直接看出考虑滞后性较不考虑滞后性在对各土层土壤湿度的预测的平均均方差要小,平均相关系数要大。表明,考虑滞后性其预测模型的预测精度将进一步提高。此外,随着土壤深度的增加,预测的均方差不断增大,平方相关系数不断减小。其中,10 cm 土壤湿度的预测均方差为 0.003,平方相关系数为 0.737,精度较高。但是,对 20 cm 以及 30 cm 土壤湿度预测则不是很理想。要提高对 20 cm 以及 30 cm 土壤湿度预测精度,可以尝试运用已经预测出来的 10 cm 土壤湿度来预测

20 cm 以及 30 cm 土壤湿度。其中,10 cm 与 20 cm 的相关系数为 0.953,10 cm 与 30 cm 土壤湿度的相关系数为 0.814,20 cm 与 30 cm 土壤湿度的相关系数为 0.878。因此,以预测出的 10 cm 土壤湿度为自变量,20 cm 土壤湿度为因变量,预测出的 20 cm 土壤湿度为自变量,30 cm 土壤湿度为因变量,运用支持向量机方法,按照前面的步骤,分别预测 20 cm、30 cm 土壤湿度。其预测及误差如图 5。

由图 5 可知,通过利用已经预测出来的 10 cm 土壤湿度进行预测 20 cm 土壤湿度,以及利用已经预测出来的 20 cm 土壤湿度进行预测 30 cm 土壤湿度,其预报效果比较理想。其中,用 10 cm 土壤湿度进行预测 20 cm 土壤湿度,其平均均方差为 0.003,平均相关系数为 0.929;用 20 cm 土壤湿度进行预测 30 cm 土壤湿度,其平均均方差为 0.003,平均相关系数为 0.956。由此可以看出,在得出 10 cm 土壤深度的土壤湿度的基础上,利用 10 cm、20 cm、30 cm 不同层次间土壤湿度相关关系,可以进一步由 10 cm 土壤深度的土壤湿度来模拟与分析 20 cm 与 30 cm 土壤深度处的土壤湿度值,这一思路是可行的,在模拟预测精度上也是可以接受的。

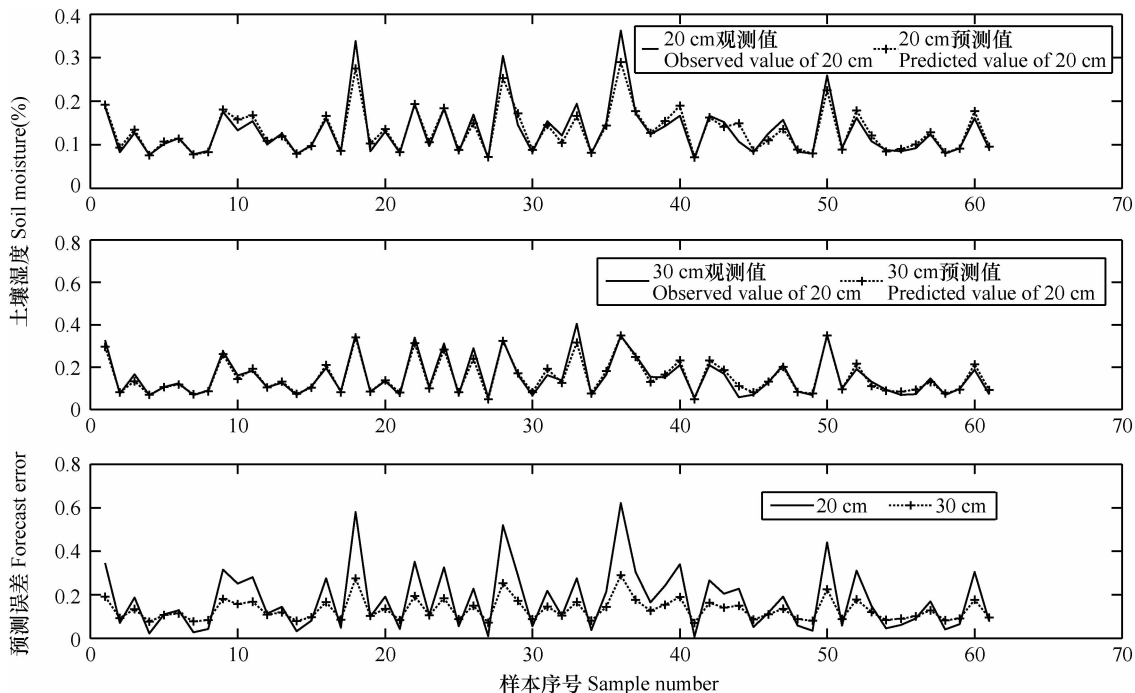


图 5 20 cm、30 cm 土壤湿度观测与预测对比图及其预测误差图

Fig. 5 Comparison of predicted values with observed values of the soil humidity at 20 cm and 30 cm in depth and errors in prediction

4 讨论与结论

本文利用 SVM 法,利用中山大学珠海校区气象观测站的逐日日平均风速、日平均气温、日平均空气湿度、日平均水汽压、日平均总辐射量、日平均地表温度、日平均降雨量以及日平均蒸发量作为自变量,模拟与预测日平均地下 10 cm、20 cm、30 cm 土壤湿度,得出以下结论:

1) 土壤湿度与背景气象因子联系密切,其变化的不稳定性主要源于气象因子的随机性、复杂性以及非线性。利用气象因子对土壤湿度进行预测,有较高的精确度,因此,通过建立统计关系与统计模型,利用气象因子分析某一土层深度的土壤湿度变化是可行的,这一结论可为水文模型中土壤湿度的计算提供理论依据。

2) 土壤湿度的变化对气象因子的变化有一定滞后性,而且滞后时间随土层深度的增加而增大,通过本文的模拟分析,在利用气象因子来模拟与预测土壤湿度时,这种滞后性是必须要考虑的,考虑滞后性较不考虑滞后性,利用气象因子模拟土壤湿度的精度会有较大幅度的提高。

3) SVM 的预测精度与其 C 、 g 值密切相关,传统试算法计算效率低,且往往很难找到局部最优值,利用 libsvm-mat 的参数寻优函数,能够提高运算效率,同时可以很方便地找到 C 与 g 值的全局最优值。

总之,SVM 方法的引入,将有利于样本的分析、因子的筛选、信息的压缩以及知识的挖潜。将 SVM 回归方法应用到土壤湿度预测当中,给土壤湿度的模拟与预测提供了新的思路与方法,也为水文模型中土壤湿度的计算提供了便利。

参考文献

- [1] 王磊,文军,韦志刚,等. 中国西北区西部土壤湿度及其气候响应. 高原气象, 2008, 27(6): 1 257—1 266. Wang L, Wen J, Wei Z G, et al. Soil moisture over the west of northwest China and its response to climate (In Chinese). Plateau Meteorology, 2008, 27(6): 1 257—1 266
- [2] Chahine T M. The hydrological cycle and its influence on climate. Nature, 1992, 359: 373—380
- [3] Walker J M, Rowntree P R. The effect of soil moisture on circulation and rainfall in a tropical model. Quart J Roy Meteor Soc, 1977, 103 (43S): 29—46
- [4] Shukla J, Miniz Y. Influence of landsurface evapotranspiration on the earth's climate. Science, 1982, 215: 1 498—1 501
- [5] Hebrard O M, Voltz Andrieux P. Spatio-temporal distribution of soil surface moisture in a heterogeneously farmed Mediterranean catchment. Hydrology, 2006, 329: 110—121
- [6] Wen J, Su Z B, Ma Y M. Determination of land surface temperature and soil moisture from tropical rainfall measuring mission/microwave imager remote sensing data. J Geophys Res, 2003, 108(D2): 4038, doi:10.1029/2002JD002176
- [7] US National Research Council. GOALS (Global Ocean-Atmosphere-Land System) for predicting seasonal-to-interannual climate. Washington D C: National Academy Press, 1994: 103—104
- [8] 陈志明. 基于多元回归分析的草坪土壤湿度模型研究. 江苏农业科学, 2010(6): 279—281. Chen Z M. Based on the multiple regression analysis model of soil moisture lawn (In Chinese). Jiangsu Agricultural Sciences, 2010(6): 279—281
- [9] 金龙,罗莹,缪启龙,等. 农田土壤湿度的人工神经网络预报模型研究. 土壤学报, 1998, 35(1): 25—32. Jin L, Luo Y, Miu Q L, et al. Forecast model of farmland soil moisture by artificial neural network (In Chinese). Acta Pedologica Sinica, 1998, 35(1): 25—32
- [10] 黄俊,吴普特,赵西宁. 基于神经网络及灰色关联分析的土壤入渗研究. 土壤学报, 2011, 48(6): 1 282—1 286. Huang J, Wu P T, Zhao X N. Research on soil infiltration based on BP network and gray correlation analysis (In Chinese). Acta Pedologica Sinica, 2011, 48(6): 1 282—1 286
- [11] 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. 土壤通报, 2007, 38(3): 427—433. Xue X P, Wang X, Zhang L J, et al. Prediction model of soil moisture based on support vector machines (In Chinese). Chinese Journal of Soil Science, 2007, 38(3): 427—433
- [12] 陈永义,俞小鼎,高学浩,等. 处理非线性分类和回归问题的一种新方法(I)——支持向量机方法简介. 应用气象学报, 2004, 15(3): 345—354. Chen Y Y, Yu X D, Gao X H, et al. A new method for non-linear classify and non-linear regression I: Introduction to Support Vector Machine (In Chinese). Journal of Applied Meteorological Science, 2004, 15(3): 345—354
- [13] 白鹏,张喜斌,张斌,等. 支持向量机理论及工程应用实例. 西安:西安电子科技大学出版社, 2008. Bai P, Zhang X B, Zhang B, et al. Support vector mechanism deals with examples of engineering application (In Chinese). Xi'an: Xi'an University of Electronic Science and Technology Press, 2008

SIMULATION AND PREDICTION OF SOIL MOISTURE BASED ON SUPPORT VECTOR MACHINE TECHNIQUE

Zhang Qiang^{1,2} Huang Shengzhi^{1,2} Chen Xiaohong^{1,2}

(1 Department of Water Resources and Environment, Sun Yat-sen University, Guangzhou 510275, China)

(2 Guangdong University Key Laboratory of Water Cycle and Security in South China, Sun Yat-sen University, Guangzhou 510275, China)

Abstract Based on observed meteorological data, such as daily mean wind speed, daily mean air temperature, daily mean air humidity, daily mean water vapor pressure, daily mean total radiation, daily mean land surface temperature, daily mean rainfall, and daily mean evaporation, and daily mean soil moisture at 10 cm, 20 cm and 30 cm in depth, statistical relationships were established between meteorological variables and soil moisture using the Support Vector Machine (SVM) technique, and on such a basis, models for simulation and prediction of soil moisture were built up. It was found that responses of soil moisture to meteorological variables somewhat lagged behind, and were affected by soil depth. The model for prediction of soil moisture taking into account the lag correlation was more accurate than the one that did not count the lag correlation. Besides, using the meteorological variables, the model was more accurate in simulating and predicting the soil moisture at 10 cm in depth than in doing the soil moisture at 20 cm or 30 cm in depth. By taking into account the close relationships between the soils at 10 cm and 20 cm and between the soils at 20 cm and 30 cm in soil moisture, it is advisable to use the support vector machine technique in simulating and predicting soil moisture at 20 cm or 30 cm on the basis of the soil moisture at 10 cm or 20 cm. The findings indicate that the model for simulation of soil moisture is very high in accuracy

Key words Support vector machine; Soil moisture; Prediction model

(责任编辑:陈德明)