

基于 PCA-SVR 的冬小麦土壤水分预测^①

聂红梅^{1,2}, 杨联安^{1,2*}, 李新尧^{1,2}, 封涌涛³, 任丽^{1,2}, 张彬⁴

(1 西北大学陕西省地表系统与环境承载力重点实验室, 西安 710127; 2 西北大学城市与环境学院, 西安 710127; 3 宝鸡市农业技术推广服务中心, 陕西宝鸡 721001; 4 内江师范学院地理与资源科学学院, 四川内江 641000)

摘要: 土壤含水量状况是影响农作物生长的重要因素, 对农作物生长关键期土壤水分的精准预测是田间管理的重要内容。研究选取宝鸡市 2014 年至 2016 年冬小麦种植区 3—5 月的气象、地形和土壤属性 3 个方面共 15 个预测因子, 建立基于主成分分析(principal component analysis, PCA)的支持向量回归机(support vector regression, SVR)模型预测 0~20 cm 和 20~40 cm 土层的土壤水分, 并同时采用随机森林(random forest, RF)回归模型对同质数据进行预测分析, 以对比分析 PCA-SVR 模型的预测效果。结果表明: PCA-SVR 模型对宝鸡市冬小麦土壤水分的预测在 0~20 cm 和 20~40 cm 土层的平均预测精度分别为 92.899% 和 92.656%, RMSE 分别为 7.521 和 8.011; 随机森林回归预测模型在 0~20 cm 和 20~40 cm 土层的平均预测精度为 87.632% 和 87.842%, RMSE 分别为 10.759 和 11.042。因此, PCA-SVR 模型对宝鸡市冬小麦土壤水分具有更好的预测能力, 且模型在 0~20 cm 土层的预测效果略优于 20~40 cm 土层。

关键词: 支持向量回归机; 主成分分析; R 语言; 土壤水分; 冬小麦; 宝鸡市

中图分类号: S152.7; S572 **文献标识码:** A

土壤水分是农作物生长管理环节中指导灌溉的重要因素之一, 对于冬小麦生长的影响主要体现在土壤水分通过影响冬小麦根系对水分的吸收进而影响冬小麦的生长发育及产量^[1]。土壤水分及其变化是一个复杂的耦合系统, 不仅受土壤自身属性影响, 还受气候、地形、耕作方式等外界因素影响, 所以在不同区域因时因地而异地选取适宜的土壤水分监测指标, 构建科学合理的预测模型, 对实时准确获取冬小麦生长的水分状况、精确田间灌溉和冬小麦生长管理都具有重要意义^[2]。目前应用较多的土壤水分预测模型主要有基于实测数据的经验模型^[3]、遥感数据反演的多源数据模型^[4-6]、时间序列模型以及具有自学习能力的人工神经网络模型和支持向量机(support vector machine, SVM)模型^[7-9]。其中经验模型参数简单易得, 但时间和人力消耗大^[10]; 遥感等多源数据的使用加快了土壤水分大规模监测预测的发展, 但模型建立参数多样复杂且受土层深度和地表覆被等因素影响较大^[11-12]; 神经网络模型具有本质上的非凹性, 在

训练过程中易陷入局部最优^[13]; 而依据结构风险最小化原则提出的 SVM 模型, 其训练是一个凹二次规划问题, 有效保证了训练得到的解即为全局最优解^[14]。作为一种新型机器学习方法, SVM 结构简单, 泛化能力强, 能够很好地解决小样本、非线性、高维度的问题, 已经在国内外土壤水分预测研究中取得了良好应用, 且与其他预测模型相比表现出明显优势^[15-19]。

SVM 土壤水分预测模型建立的基础是合理选取输入变量, 即预测因子, 考虑到土壤水分影响因子过多易造成信息冗余, 导致模型训练时间过长, 本文采用主成分分析(principal component analysis, PCA)方法对 SVM 输入变量进行降维处理。结合 PCA 与支持向量回归机(support vector regression, SVR)建立土壤水分预测模型, 可以提高土壤水分预测精度, 分析气象、地形和土壤属性等要素对土壤水分的影响, 为土壤水分模拟预测提供科学依据。

基金项目: 国家重点研发计划课题(2016YFD0800307)、教育部人文社会科学研究规划项目(10YJA910010)、陕西省农业科技攻关项目(2011K02-11)、西安市科技计划农业技术研发项目(NC150201; NC1402)和西北大学研究生质量工程提升项目(YZZ17147; YZZ17151)资助。

* 通讯作者(yanglianana@163.com)

作者简介: 聂红梅(1993—), 女, 湖北武汉人, 硕士研究生, 主要从事 RS 与 GIS 在农田土壤方面的应用研究。E-mail: masterhongmei@126.com

1 研究区概况及数据来源

1.1 研究区概况

关中平原作为我国三大小麦主产区之一,小麦种植历史悠久。宝鸡市地处关中平原西部,属大陆性季风气候,南、北、西三面分别为秦岭、北山和关山山地,中部为渭河平原,冬、夏季风在春季的调整交替使得宝鸡地区 3—5 月气温上升迅速且降水少。宝鸡市主要属于秦岭以北的黄河水系,其中渭河流域面积占全市的 68.53%,全市水资源总量约为 48.8 亿 m³,其中地表水径流量 41.3 亿 m³,地下水可采量 7.5 亿 m³,总体上仍属缺水地区;在多种成土因素的综合作用下,形成以褐土、黑垆土和黄棕壤为主的土壤类型^[20]。宝鸡市春季农作物以冬小麦种植为主(图 1),3—5 月为冬小麦生长的拔节、抽穗和灌浆期,期间耗水量巨大,对土壤水分要求高。采用科学合理的模型对宝鸡市冬小麦种植区 3—5 月土壤水分的预测研究对本区冬小麦的科学种植管理和保质保产具有重要意义。

1.2 数据来源及预处理

因关中平原冬小麦关键生长期和需水期为 3—5 月,所用数据时间尺度为 2014—2016 年的 3—5 月。具体包括宝鸡市 2014 年至 2016 年春季采用土钻烘干法测量的土壤水分实测数据(相对含水量)、土壤属性统计数据以及相应时间的宝鸡市气象数据和数字高程模型(digital elevation model, DEM)数据。其中实测

土壤水分数据及各监测点土壤质地、潜水埋深、耕层厚度、土壤容重等数据由宝鸡市土肥站提供,土壤质地根据国际制土壤质地分级标准进行量化处理;所有气象数据均来源于国家气象科学数据共享服务平台的中国地面气候资料数据和产品,DEM 等数据来源于地理空间数据云平台的 GDEMDEM 30 m 分辨率数字高程数据,并根据 DEM 数据在 ArcGIS10.3 中提取坡度、坡向、地形湿度指数(TWI)等其他相关地形因子;对于少数几天缺失实测数据采取 k 邻近算法补全。对数据做基本预处理和归一化,以符合模型输入数据格式。

综合考虑宝鸡市复杂的地形、气候特征和土壤水分实测数据的可取性,本研究结合宝鸡市 2014 年至 2016 年的土壤水分实测数据、土壤属性统计数据、气象观测数据和 DEM 数据,选取海拔、坡度、坡向、TWI 等地形因子,水汽压、风速、日照时数、气压、气温、降水和相对湿度等气象因子以及土壤容重、土壤质地、潜水埋深和耕层厚度等土壤属性因子共同组成土壤水分支持向量回归因子集合。由于部分因子之间存在较强相关性,为避免数据冗余对预测因子集合进行 PCA 降维,选取累计贡献率大于 85% 的因子构成重要预测因子集合。然后建立土壤水分 SVR 预测模型,模型训练集和测试集按照 3 a 数据总数的 4 : 1 分配,按照训练集所得最优参数组合最优模型并用测试集进行精度验证。

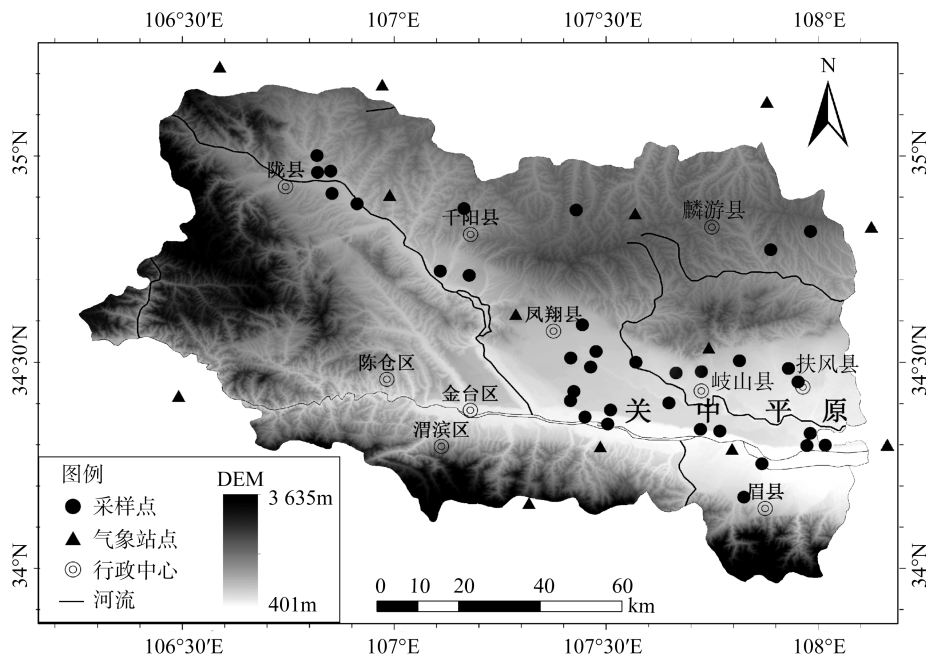


图 1 宝鸡市冬小麦种植区概况图
Fig.1 General situation of planting area of winter wheat in Baoji

2 SVR 原理和方法

2.1 SVR 原理

支持向量机(SVM)是由 Vapnik 等于 20 世纪 90 年代提出的一种数学推导严密、理论基础坚实的机器学习方法,由于其在计算机图像处理^[21-22]、非线性建模与预测^[23-25]、优化控制^[26-27]等领域中所表现出的特有优势而被学者们认为是继神经网络学习之后新的研究热点^[28]。SVR 是基于统计学习的 VC 维理论和结构风险最小化原则,在已有的样本信息中获取最佳学习模型,其核心思想是:通过对样本信息的学习获得因变量和自变量之间的复杂的非线性映射关系,即

基于 Mercer 核展开定理,通过非线性映射,把低维样本空间映射到一个高维乃至无穷维的特征空间,使其在特征空间中可以应用线性学习机的方法解决低维度样本空间中的高度回归等问题^[29-30]。

2.2 预测因子的选取和优化

在不考虑人为灌溉等因素的条件下,冬小麦种植区土壤水分主要受气象、地形和土壤自身属性的影响。本研究以宝鸡市 2014 年土壤水分设站监测以来共 3 a 每年 3—5 月共 306 组冬小麦生长期土壤水分实测数据的月均值为模型的目标值,以气象、地形和土壤属性等各因子的月均值为模型样本输入值。各预测因子详情如表 1 所示。

表 1 宝鸡市土壤墒情预测因子
Table 1 Prediction factors of soil moisture in Baoji

因子属性	预测因子	因子描述
气象因子	相对湿度(%)	平均相对湿度月均值
	降水量(mm)	平均降水量月均值
	气压(hPa)	平均气压月均值
	气温()	平均气温月均值
	风速(m/s)	平均风速均值
	水汽压(hPa)	平均水汽压月均值
	日照时数(h)	日照时数月均值
	地形因子	海拔(m)
	坡度(度)	采样点地形坡度
	坡向(度)	采样点坡面朝向
	TWI(%)	采样点地形湿度指数
土壤因子	质地(%)	采样点土壤砂粒(2~0.02 mm)占比
	容重(g/cm ³)	采样点土壤容重
	潜水埋深(m)	采样点潜水位埋藏深度
	耕层厚度(cm)	采样点土壤耕层厚度

表 2 预测因子主成分得分系数矩阵
Table 2 Matrix of principal component score coefficients of prediction factors

主成分	PC ₁	PC ₂	PC ₃	PC ₄	PC ₅	PC ₆	PC ₇	PC ₈
降水量	0.181	-0.144	0.299	0.098	-0.113	0.136	-0.088	0.091
相对湿度	0.171	-0.159	0.327	0.152	-0.076	0.002	-0.244	0.089
水汽压	0.331	-0.119	-0.040	0.005	0.014	0.013	0.100	-0.064
风速	-0.031	-0.083	-0.297	0.013	-0.250	0.400	-0.474	0.183
气温	0.0318	-0.072	-0.162	-0.051	0.045	0.007	0.194	-0.099
气压	-0.180	-0.211	0.031	0.120	-0.021	0.154	0.381	-0.281
日照时数	0.196	0.055	-0.355	-0.097	0.081	-0.048	0.104	-0.083
质地	0.041	0.145	0.153	-0.409	0.013	0.246	-0.206	-0.009
容重	0.056	0.212	0.165	-0.259	-0.076	0.256	-0.037	-0.401
潜水埋深	0.037	0.179	0.035	0.152	-0.253	0.286	0.635	0.434
耕层厚度	0.057	0.209	0.047	0.176	0.166	-0.480	-0.216	0.098
坡向	-0.008	-0.020	-0.083	0.286	0.453	0.409	-0.278	0.025
坡度	0.042	0.182	0.070	0.262	0.267	0.213	0.125	-0.551
DEM	0.071	0.276	-0.015	0.238	-0.164	0.083	-0.151	0.311
TWI	-0.009	-0.041	0.090	-0.229	0.557	0.111	0.191	0.573

表 3 预测因子主成分贡献率
Table 3 Contribution rates of principal components of prediction factors

主成份	特征值	贡献率(%)	累计贡献率(%)
PC ₁	2.767	18.444	18.444
PC ₂	2.681	17.871	36.315
PC ₃	1.949	12.994	49.309
PC ₄	1.645	10.970	60.279
PC ₅	1.249	8.328	68.607
PC ₆	1.144	7.627	76.234
PC ₇	0.876	5.840	82.073
PC ₈	0.808	5.386	87.459

由表 2、3 可知,PCA 提取出能够反映原始变量 87.459% 信息量的 8 个主成分,基本能够概括原有变量所反映的信息。故提取前 8 个主成分构建综合预测因子对冬小麦土壤水分进行建模预测。

3 PCA-SVR 模型的建立及评估

3.1 模型的建立和优化

PCA-SVR 模型的训练和测试是在 R 语言环境下加载 e1071 程序包完成的。模型建立的主要步骤为:

获取 0~20 cm 和 20~40 cm 土壤水分观测数据、气象要素观测数据,以月尺度统计宝鸡市 34 个土壤水分实测站点 2014—2016 年 3—5 月共 306 组数据,综合冬小麦种植区地形和土壤属性因子建立土壤水分预测因子集合;采用主成分分析法对预测因子集合进行降维优化,选出前 8 个主成分构建综合预测因子集合。将预处理后的数据集按照 4:1 的原则分为模型训练集和测试集进行交叉验证,以避免模型出现欠学习和过学习状态。选取的核函数为高斯径向基核函数,通过交叉验证发现核参数 $\gamma=100$,正则化参数 $\text{cost}=0.01$ 时预测误差最小且模型泛化能力最强。在 R 语言环境下进行模型训练所用主要函数如下:

训练函数: `model.svmtrain <- svm(), svmfit<-svm()`

参数寻优: `tuned<-tune.svm()`

预测函数: `preds <- predict()`

交叉验证: `cv.folds <- function()`

精度分析: `svm.score<-()`

3.2 预测结果及效果评估

本研究对宝鸡市 2014—2016 年共 3 a 的 3—5 月冬小麦生长需水关键期 0~20 cm 和 20~40 cm 土层的土壤水分进行预测,因随机森林(random forest,

RF)回归模型也可用于高维度数据集的处理应用且与支持向量机同属新型机器学习方法,为对比分析 PCA-SVR 模型的预测能力在 R 语言环境下加载 random Forest 程序包对同样的数据进行 RF 建模,将 3 a 共 306 组数据的实测值与模型预测值在 SPSS 21 中作散点图,并添加线性总计拟合线以直观对比拟合效果,如图 2 所示。从图 2 中可以发现,PCA-SVR 模型的拟合优度(R^2)无论是在 0~20 cm 土层还是 20~40 cm 土层都明显优于 RF 回归预测模型,表明 PCA-SVR 模型对于土壤水分具有更好的预测效果;且 PCA-SVR 模型在 0~20 cm 土层的拟合优度略高于 20~40 cm 土层,表明该 PCA-SVR 模型对宝鸡市冬小麦种植区的浅层土壤水分的拟合效果优于深层土壤,引起该现象的原因可能是本研究所搜集的土壤水分预测因子主要为影响表层土壤水分较明显的气象因子,而深层土壤对一般预测因子的变化响应具有明显的滞后性,且深层土壤还会受到地下水、壤中流和土壤生物活动等更多的不确定因素的影响。

对两种预测模型的模拟精度进行定量评价,采用决定系数(R^2)、均方根误差(RMSE)和预测精度 3 个评价指标,计算结果如表 4 所示,从表 4 中可以发现 RF 回归模型在 0~20 cm 土层土壤水分预测值与实测值的均方跟误差较 PCA-SVR 模型大;PCA-SVR 模型的预测精度在 0~20 cm 和 20~40 cm 土层均达到 90%以上,且都高于 RF 回归模型。由于预测因子的选取和深层土壤的不确定性等导致 0~20 cm 土层预测效果优于 20~40 cm 土层。

4 结论

1)PCA-SVR 模型在宝鸡市土壤水分预测中具有良好效果,可以运用到冬小麦乃至其他大田作物的生产管理中。应用两种预测模型研究,发现不同深度土层的预测精度有细微偏差,模型对浅层土壤的预测效果略优于深层土壤,且 PCA-SVR 模型在两个土层对宝鸡市土壤水分的预测效果优于 RF 模型。

2)采用在非线性、高维度研究方面具有优势的支持向量回归机模型来预测冬小麦土壤水分具有较高科学性;模型的时间尺度是冬小麦生长对水分要求最为严格的 3—5 月,且充分考虑到影响冬小麦土壤水分的各方面因素,较之以往单一依据气象要素的预测模型更为可靠。今后可加入生物和环境的交互影响等因素,并与遥感等卫星观测数据相结合,在提高模型预测精度的同时增强模型的应用范围。

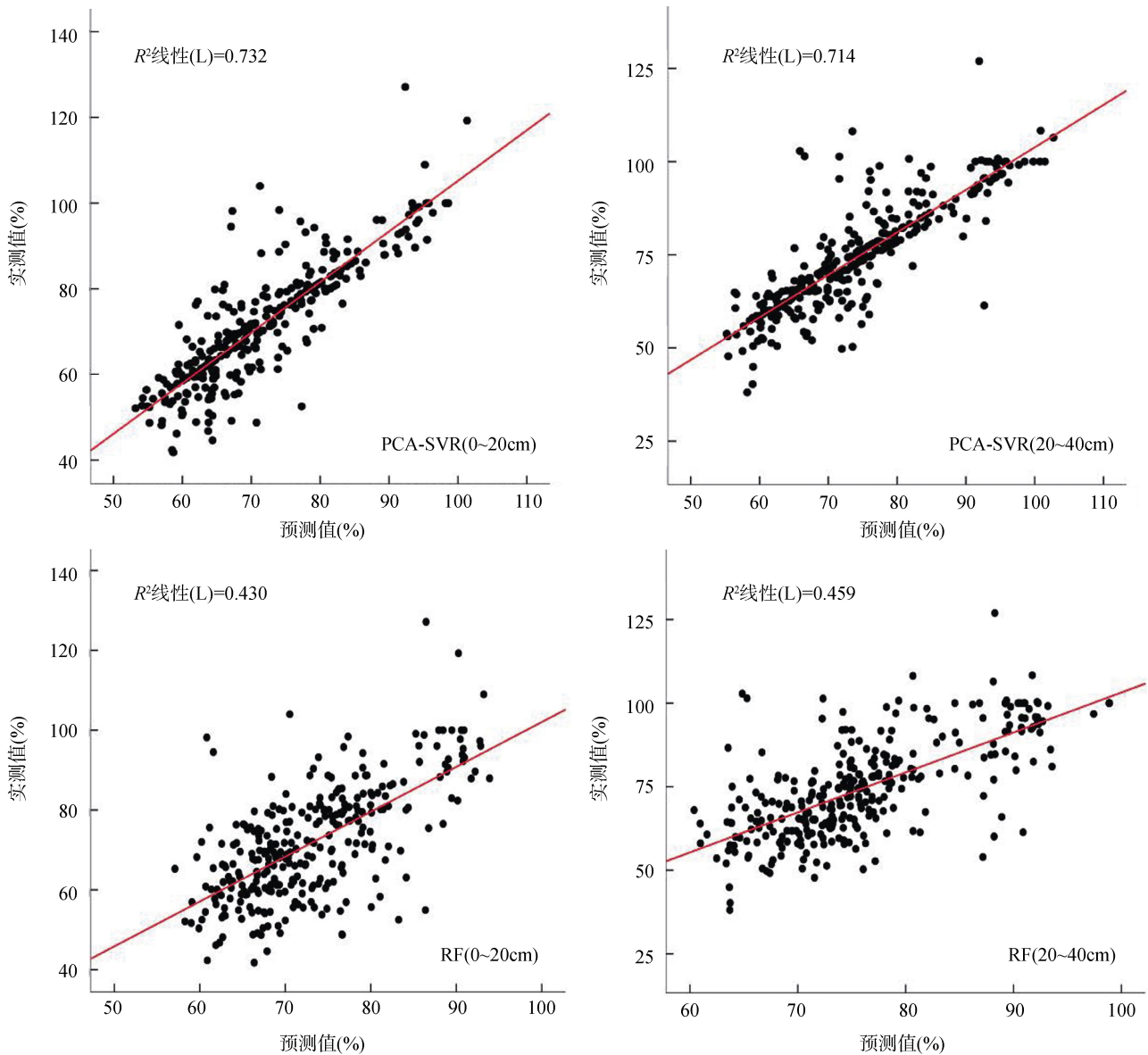


图 2 土壤水分实测值和预测值散点图

Fig. 2 Scatter plots of observed and forecasted soil moistures

表 4 模型预测精度评价

Table 4 Accuracies of predication models

预测模型	土层深度(cm)	R^2	RMSE	预测精度(%)
RF	0~20	0.430	10.759	87.632
	20~40	0.459	11.042	87.842
PCA-SVR	0~20	0.732	7.521	92.899
	20~40	0.714	8.011	92.656

参考文献:

- [1] 张明生, 王丰, 张国平. 中国农业用水存在的问题及节水对策[J]. 农业工程学报, 2005, 21(13): 1-6
- [2] 黄友昕, 刘修国, 沈永林, 等. 农业干旱遥感监测指标及其适应性评价方法研究进展[J]. 农业工程学报, 2015, 31(16): 186-195
- [3] 刘玉娇, 朱青, 吕立刚, 等. 基于时间稳定性和降维因子分析的土壤水分检测优化[J]. 土壤, 2016, 48(1): 186-192
- [4] Ahmad S, Kalra A, Stephen H. Estimating soil moisture using remote sensing data: A machine learning approach[J]. Advances in Water Resources, 2010, 33(1): 69-80
- [5] 蔡亮红, 丁建丽, 魏阳. 基于多源数据的土壤水分反演及空间分异格局研究[J]. 土壤学报, 2017, 54(5): 1057-1067

- [6] Tran A P, Vanclooster M, Zupanski M, et al. Joint estimation of soil moisture profile and hydraulic parameters by ground-penetrating radar data assimilation with maximum likelihood ensemble filter[J]. *Water Resources Research*, 2014, 50(4): 3131–3146
- [7] 张和喜, 杨静, 方小宇, 等. 时间序列分析在土壤墒情预测中的应用研究[J]. *水土保持研究*, 2008, 15(4): 82–84
- [8] 刘洪斌, 王伟, 魏朝富, 等. 土壤水分预测神经网络模型和时间序列模型比较研究[J]. *农业工程学报*, 2003, 19(4): 33–36
- [9] Liu Y, Mei L, Su K O. Prediction of soil moisture based on extreme learning machine for an apple orchard[C]// *IEEE, International Conference on Cloud Computing and Intelligence Systems*. IEEE, 2015: 400–404
- [10] 拉巴, 卓嘎, 陈涛. 藏北地区土壤水分遥感反演模型的研究[J]. *土壤*, 2017, 49(1): 171–176
- [11] 杨贵军, 岳继博, 李长春, 等. 基于改进水云模型和 Radarsat-2 数据的农田土壤含水量估算[J]. *农业工程学报*, 2016, 32(22): 146–153
- [12] 金慧凝, 张新乐, 刘焕军, 等. 基于光谱吸收特征的土壤含水量预测模型研究[J]. *土壤学报*, 2016, 53(3): 627–635
- [13] 陈昌华, 谭俊, 尹健康, 等. 基于 PCA-RBF 神经网络的烟田土壤水分预测[J]. *农业工程学报*, 2010, 26(8): 85–90
- [14] Vapnik V N. *The Nature of Statistical Learning Theory*[J]. *IEEE Transactions on Neural Networks*, 2002, 8(6): 1564–1564
- [15] 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
- [16] Gill M K, Asefa T, Kemblowski M W, et al. Soil moisture prediction using support vector machines[J]. *Journal of the American Water Resources Association*, 2006, 42(4): 1033–1046
- [17] 邓建强, 陈效民, 王伯仁, 等. 基于最小二乘向量机土壤水分动态模拟与分析[J]. *水土保持通报*, 2009(6): 119–122
- [18] 丁辉, 仲跃, 张俊, 等. 基于细菌觅食优化算法的支持向量机在土壤墒情预测中的应用[J]. *水土保持通报*, 2016, 36(6): 131–135
- [19] 王彤彤, 翟军海, 何欢, 等. BP 神经网络和 SVM 模型对施加生物炭土壤水分预测的适用性[J]. *水土保持研究*, 2017, 24(3): 86–91
- [20] 赵蓓. 生态示范区可持续发展评价研究[D]. 西安: 西北大学, 2009
- [21] 李春花, 凌贺飞, 卢正鼎. 基于支持向量机的自适应图像水印技术[J]. *计算机研究与发展*, 2007, 44(8): 1399–1405
- [22] 秦丰, 刘东霞, 孙炳达, 等. 基于深度学习和支持向量机的 4 种苜蓿叶部病害图像识别[J]. *中国农业大学学报*, 2017, 22(7): 123–133
- [23] 张豪, 罗亦泳, 张立亭. 基于遗传支持向量机的城市扩张非线性组合模型[J]. *地理学报*, 2010, 265(6): 656–664
- [24] 陈强, 蒋卫国, 陈曦, 等. 基于支持向量回归模型的水稻田甲烷排放通量预测研究[J]. *环境科学*, 2013, 34(8): 2975–2982
- [25] 张成成, 沈爱春, 张晓晴, 等. 应用支持向量机评价太湖富营养化状态[J]. *生态学报*, 2013, 33(23): 7563–7569
- [26] 叶林, 任成, 李智, 等. 风电场有功功率多目标分层递阶预测控制策略[J]. *中国电机工程学报*, 2016, 36(23): 6327–6336
- [27] 陈进东, 潘丰. 基于粒子群优化的在线支持向量回归预测控制方法[J]. *信息与控制*, 2013, 42(6): 723–728
- [28] 丁世飞, 齐丙娟, 谭红艳. 支持向量机理论与算法研究综述[J]. *电子科技大学学报*, 2011, 40(1): 2–10
- [29] 张波. 基于支持向量机的干旱预测研究[D]. 南京: 南京信息工程大学, 2012
- [30] 王定成. 支持向量机建模预测与控制[M]. 北京: 气象出版社, 2009

Prediction of Soil Moisture of Winter Wheat by PCA-SVR

NIE Hongmei^{1,2}, YANG Lian'an^{1,2*}, LI Xinyao^{1,2}, FENG Yongtao³, REN Li^{1,2}, ZHANG Bin⁴

(1 *Shaanxi Key Laboratory of Earth Surface System and Environmental Carrying Capacity, Northwest University, Xi'an 710127, China*; 2 *College of Urban and Environmental Sciences, Northwest University, Xi'an 710127, China*; 3 *Baoji Agricultural Technology Extension Service Centre, Baoji, Shaanxi 721001, China*; 4 *School of Geography and Resources Science, Neijiang Normal University, Neijiang, Sichuan 641000, China*)

Abstract: Soil moisture is one of the important factors affecting the growth of crops, accurate prediction of soil moisture in the critical period of crop growth is an important part of the field management. In this study 15 prediction factors were selected from meteorology, topography and soil properties from March to May in the winter wheat growing area from 2014 to 2016 in Baoji of Shaanxi Province, soil moistures in 0–20 cm and 20–40 cm soil layers were predicted and compared by using the established PCA-SVR (Principal Component Analysis-Support Vector Regression) model and Random Forest (RF) regression model. The results showed that prediction accuracies in 0–20cm and 20–40cm soil layers were 92.899% and 92.656% for PCA-SVR model, 87.632% and 87.842% for RF regression model, with the corresponding RMSEs of 7.521 and 8.011 for PCA-SVR model, 10.759 and 11.042 for RF regression model, respectively, indicating that PCA-SVR model had better predictive ability on soil moisture of winter wheat in Baoji, particularly for 0–20 cm soil layer.

Key words: Support vector regression; Principal component analysis; R language; Soil moisture; Winter wheat; Baoji City