基于支持向量回归（SVR）和多时相遥感数据的冬小麦估产

黎 锐[1,2]，李存军[1※]，徐新刚[1]，王纪华[1]，杨小冬[1]，黄文江[1]，潘瑜春[1]
（1．国家农业信息化工程技术研究中心，北京 100097； 2．首都师范大学信息工程学院，北京 100048）

摘 要：发展和构建高精度的作物遥感估产模型，对于国家制订粮食进出口政策和保障粮食安全具有重要意义。尝试利用支持向量回归方法（SVR）构建遥感估产模型，首先利用北京郊区 2004 年和 2007 年冬小麦主要生育期多时相 Landsat TM 影像生成的归一化植被指数，通过 SVR 构建遥感估产模型进行产量估算。然后针对模型的稳健型和预报能力进行交叉验证，并与常规的多元回归方法进行对比。结果表明，利用 SVR 方法构建的遥感估算模型有效地提高了估算精度，与多元回归方法相比，2004 年和 2007 年决定系数分别提高 0.2162、0.2158，均方根误差分别降低 0.1682、0.2912。因此基于 SVR 和多时相遥感数据构建估产模型用于冬小麦估产是可行、有效的，为应用多时相遥感数据进行冬小麦估产提供了一种方法。

关键词：SVR，多时相遥感，估产，NDVI
doi：10.3969/j.issn.1002-6819.2009.07.021
中图分类号：TP79：S127 **文献标识码**：A **文章编号**：1002-6819(2009)-7-0114-04

黎 锐，李存军，徐新刚，等. 基于支持向量回归（SVR）和多时相遥感数据的冬小麦估产[J]. 农业工程学报，2009，25(7)：114－117.
Li Rui, Li Cunjun, Xu Xingang, et al. Winter wheat yield estimation based on support vector machine regression and multi-temporal remote sensing data[J]. Transactions of the CSAE, 2009,25(7)：114－117.(in Chinese with English abstract)

0 引 言

客观、精确、实时地监测预报区域作物产量一直是国家粮食生产管理和粮食期货关心的问题。发展和构建高精度的大面积作物遥感估产模型，提高粮食估产精度，对于国家实现粮食宏观调控、制订粮食进出口政策、在国际农产品贸易中争取到主动权具有重要意义[1]。

目前，利用遥感技术进行作物估产的方法，依据所采用的模型特点可以大致分为 3 类：1）经验模型或统计模型方法，2）机理模型方法，3）半经验（半机理）模型方法。其中，机理模型方法充分考虑了作物的产量形成机理并与遥感结合，但需要较多的参数输入，在大面积区域应用时，因许多参数无法获取而受到限制。半经验模型方法对机理模型方法进行了适当简化，在区域应用中得到了较大发展，但往往需要高时间分辨率遥感影像如 AVHRR、MODIS 等，为获得较高的精度常常需要每日的遥感数据，导致数据处理量较大[1]。统计模型方法尽管对作物产量形成的机理解释性不强，但其操作实施简单、灵活，仍然是当前作物遥感估产的主要常规方法。

收稿日期：2009-04-09 修订日期：2009-05-30
基金项目：农业部公益性行业科研专项（200803037）；北京市自然科学基金（4092016）；北京市农林科学院青年基金（504-05-21）
作者简介：黎 锐（1983－），女，四川人，主要从事农业遥感和数据同化研究。北京 国家农业信息化工程技术研究中心，100097。
Email: peking973@163.com
※通信作者：李存军（1975－），男，湖北人，博士，副研究员，主要从事农业遥感应用研究。北京 国家农业信息化工程技术研究中心，100097。
Email: licj@nercita.org.cn

统计模型方法包括线性模型和非线性模型。线性模型简便，但作物产量形成通常具有非线性。所以基于遥感数据与产量统计关系的线性方法，通常存在经验特征强、精度不够高的缺点[2]。因此，非线性的遥感估产模型日益受重视，如神经网络法[3]。但神经网络是人脑神经元的近似模拟，其建模的精度往往受主观因素的制约，训练过程易陷入局部极值和过学习，影响预测的精度[4]。

支持向量机（support vector machine，SVM）是一种基于统计学习理论的机器学习法。支持向量机又分为两类，包括支持向量分类（SVC）和支持向量回归（SVR）[5]。最初，支持向量分类用于模式识别领域，农业上主要应用在病害诊断、品种识别、温室控制上面[6-8]。近年来，支持向量回归的应用也逐渐兴起，主要体现在非线性时间序列分析上，农业上已有这方面的报道[9-10]。然而，在农业遥感估产领域，比较常用的方法仍以偏最小二乘（PLS）、人工神经网络（ANN）为主，而利用支持向量回归还鲜有报道。

本文尝试通过支持向量回归（support vector mahine regression，SVR）方法，利用多时相 Landsat TM 影像获取的归一化植被指数 NDVI 构建遥感估产模型进行冬小麦估产，并对其精度进行了评价分析。

1 材料与方法

1.1 研究区域

研究分别于 2004 年和 2007 年两个生长季进行，对象是位于北京郊区昌平、顺义、通州等地的大田冬小麦，冬小麦于前一年约 9 月 28 日—10 月 8 日期间播种，第二

年 6 月 20 日左右收获。冬小麦样地在研究区域的分布如图 1 所示（图中黑点代表冬小麦样地）。

图 1 冬小麦样地分布图

Fig.1　Distribution of winter wheat sample plots

1.2　数据获取及处理

冬小麦产量数据是冬小麦收获前进行定点定位并实割实测获得。在试验地中选有代表性的 30 m×30 m 范围内收割约 5 m^2 冬小麦（中心和四角各取约 1 m^2），精确测量面积，脱粒晒干，称质量并根据实际取样面积计算单产。

遥感数据源为 Lanadsat TM 影像，利用历史几何校正好的 Lanadsat TM 影像为底图进行校正，误差小于 0.5 个像素。本研究中使用的 2004 年影像数据包括 4 月 1 日，4 月 17 日和 5 月 19 日；2007 年影像数据包括 4 月 10 日，4 月 26 日和 5 月 28 日，每年度 3 个时相。

1.3　模型构建方法

本文研究估产利用支持向量回归（SVR）方法进行，并与常规的多元线性回归（MLR）方法对比。

1.3.1　SVR 的基本原理

支持向量机起初是解决两类样本的分类问题，其核心思想是找到一个最优分类超平面 $w \cdot x+b=0$，使两类样本的分类间隔最大化。支持向量回归与支持向量分类相似，不同之处在于，回归所求超平面是使所有样本点到超平面的距离为最小。对于线性回归问题，实质上是寻求一个最优超平面，使得在给定精度 ε（$\varepsilon \geqslant 0$）条件下可以无误差的拟合 y，即所有样本点到最优超平面的距离都不大于 ε；考虑到允许误差的情况，可引入松弛变量 ξ 和 ξ^*（ξ、$\xi^* \geqslant 0$）以及惩罚参数 C（$C>0$），其寻优问题转化为相应的二次规划问题[11-12]。

对于非线性回归问题，可通过核函数变换将样本映射到一个高维特征空间中用线性回归来解决[13]。通常，特征空间维数很高甚至具有无穷维数，致使空间变换后计算量巨增而面临维数灾难等问题[14]。幸运的是支持向量机中待解的对偶问题只包含一个变换后特征空间的内积运算，而这种运算能在原空间中通过核函数来实现。根据 Mercer 定理可构造系列核函数，常见如线性核（t=0）、多项式核 1（t=1，d=2）、多项式核 2（t=1，d=3）、径向基核（t=2）和 sigmoid 核（t=3）等。

1.3.2　最优参数选取与模型确定

通常对样本先分割，再训练，之后搜索参数，最后确定相对最优值进行建模。本研究的实验数据为 2004 年与 2007 年各 3 个时相 Landsat TM 的 NDVI 数据和当年产量。其中 2004 年有 22 个样本、2007 年有 24 个样本，将样本分为 4 份，2004 年样本比例分别为 7∶5∶5∶5，2007 年样本比例分别为 6∶6∶6∶6，选取 3 份做训练集，留 1 份做测试集，交叉建模重复 4 次，预测 4 次。

采用 LIBSVM2.8，首先按上述的比例分割训练集和测试集，并归一化，再用 gridregression.py 自动搜索最优参数：最佳惩罚参数 c、属性数 g 及不敏感损失函数 p，它依据均方误差最小原则返回需要的参数。实际上，这 3 个参数和核函数类型 t，以及支持向量类型 s，决定着模型的优化性能。然而，gridregression.py 并不提供针对全部 N 个样本回代寻优，而是采取自动参数寻优，即 N 折交叉验证（实质为留一法），从而避免过拟合[15]。一方面，过高的回代拟合精度并无多大实际意义；另一方面，对估产预测模型，人们真正感兴趣的是实际预测能力而非回代结果。因此，本文以实际预测结果作为模型优劣的评价基准。之后，根据返回的参数，用 svm-train 建模和 svm-predict 预测。此时，搜索出的最优参数并非我们实际需要的参数，通常均方误差比较大。这就需要利用深度优先策略，适当调整 t 和 s，筛选出相对较好的参数。然后再在相对优异的参数 s、t 中，逐步改变 p、g，搜索最优组合，直到得到预测误差最小的模型，作为最优预测模型，建立精度相对较高的回归模型。

1.4　验证方法

本文采取统计学中常用的 2 种指标：均方根误差（*RMSE*）、决定系数（R^2）[16]评价冬小麦预测结果的优劣。

2　结果与分析

2.1　多时相遥感估产

分别通过多元线性回归（MLR）和支持向量回归（SVR）方法，得到了冬小麦产量预测值，并建立了实测值与预测值之间的关系。利用多元线性回归 ，2004 年 22 个冬小麦样地预测的均方根误差（*RMSE*）为 0.5975，2007 年 24 个样地预测的均方根误差（*RMSE*）为 0.8393。而通过支持向量回归，两年产量预测值的均方根误差均得到了有效的降低，2004 年均方根误差为 0.4293，2007 年均方根误差为 0.5481。此外，我们得到了两种方法预测值与实测值的对比，2004 年见图 2a、图 2b，2007 年见图 2c、图 2d。

对比图 2a、图 2b 可以看出利用 SVR 方法，2004 年数据得到的拟合值好于 MLR 方法，能更好的逼近实测值，决定系数较好达到了 0.6419，好于 MLR 方法得到的结果（决定系数仅为 0.4257）。

同样，对比图 2c、图 2d，利用 SVR 方法 2007 年数据得到的拟合值好于 MLR 方法，决定系数达到了 0.8483，好于 MLR 方法得到的结果（决定系数仅为 0.6325）。

上述结果表明，采用 SVR 估产，预测效果好于 MLR，当然，对于 2004、2007 年各自的冬小麦试验样本点中，

并不是每个样点采用 SVR 都能得到比 MLR 好的结果，但大样本的平均情况是符合统计学意义的。

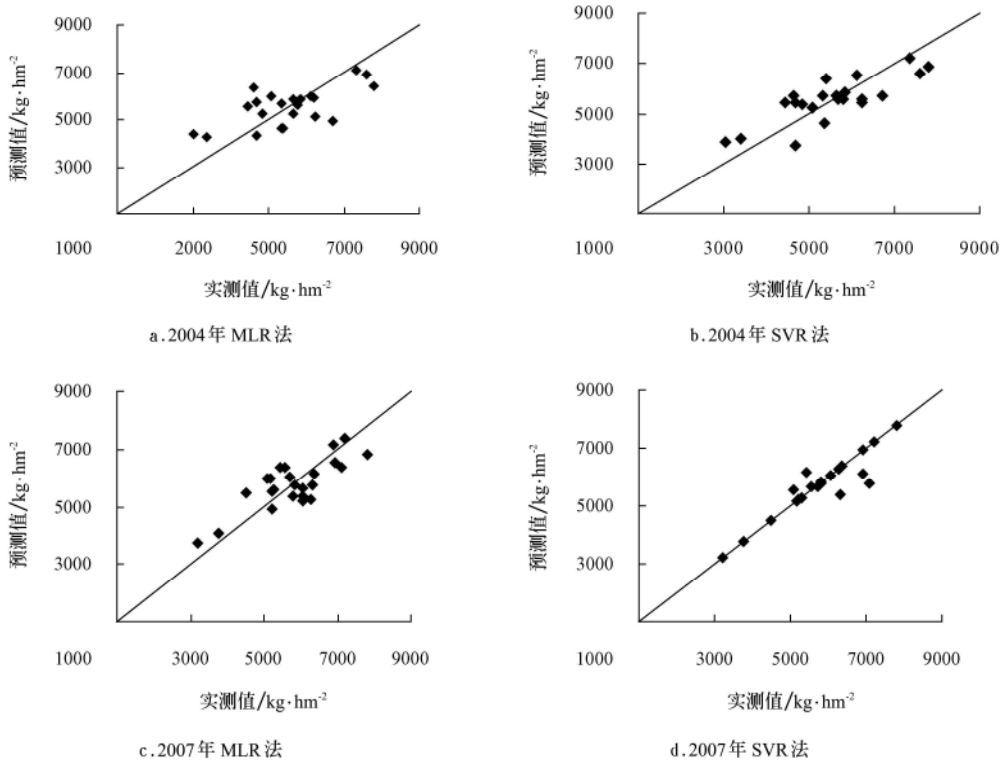

图 2 预测值与实测值散点图

Fig.2 Scatter of prediction values and the measured values

2.2 结果分析

从上述建模后的结果看出，两年的 SVR 方法均表现出较好的稳定性和鲁棒性。这是因为支持向量回归方法较好地解决了小样本、非线性、过拟合、维数灾和局部极小等问题，泛化推广能力优异[17]，这也是它比 MLR 能够较好的预测冬小麦产量的原因。并且 SVR 方法在参数设定后，能够有唯一解，优于人工神经网络（ANN）方法。

然而，SVR 在实际应用中尚存在一定的局限性，如其核函数的选取和核函数的一些参数确定带有经验性，在利用支持向量机方法进行农业估产上，支持向量回归（SVR）在其核函数、相应参数的选择上还有待进一步深入研究。由于支持向量机研究的兴起，多种与核函数有关的结合算法如核主成分分析（kernel PCA）[17]，核偏最小二乘分析（kernel PLS）[18]也相继提出，这些研究有望为核函数及其参数选取的客观性提供解决办法。

本研究中 2007 年遥感估产精度高于 2004 年，可能是因为 2007 年 3 个时期的遥感影像相对于 2004 年各晚了 9 天，2004 年的冬小麦分别处于返青期、拔节中期和灌浆初期，而 2007 年分别处于拔节期初期、旗叶期和灌浆中期，2007 年遥感影像的时相更有利于产量的估算。

3 结 论

本文研究了利用支持向量回归和多时相遥感数据进行冬小麦产量预测的可行性。结果表明，支持向量机回归方法具有较好的稳定性和鲁棒性，可有效提高作物估产的精度。本研究为应用多时相遥感数据进行冬小麦估产提供了一种新的方法尝试。

志谢：感谢宋晓宇和顾晓鹤两位博士在数据获取及预处理中给予的帮助！

［参 考 文 献］

[1] 徐新刚．农作物单产模型研究[D]．北京：中国科学院，2007．
Xu Xingang. Research of Crop Yield Models[D]. Beijing: Chinese Academy of Sciences，2007．(in Chinese with English abstract)

[2] Groten S M E. NDVI-crop monitoring and early yield assessment of Burkina Faso[J]. International Journal of Remote Sensing，1993，14：1495－1515．

[3] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报，2000，26(1)：32－42．
Zhang Xuegong. Introduction to statistical learning theory and support vector machine[J]. Acta Automatica Sinica，2000，26(1)：32－42．(in Chinese with English abstract)

[4] 边肇祺，张学工，等．模式识别[M]．北京：清华大学出版社，2002．

[5] 邓乃扬，田英杰．数据挖掘中的新方法－支持向量机[M]．北京：科学出版社，2004．

[6] 周立阳，费惠新，张孝羲．多维时间序列分析在稻纵卷叶螟长期预测预报上的试用[J]. 植物保护学报，1995，22(1)：1－6．
Zhou Liyang, Fei Huixin, Zhang Xiaoxi . The application of

multiple dimension time series analysis method in long-term forecasting of rice leaf roller[J]. Acta Phytophylacica Sinica，1995，22(1)：1－6．(in Chinese with English abstract)．

[7] 邹晓波，赵杰文．支持向量机在电子鼻区分不同品种苹果中的应用[J]．农业工程学报，2007，23(1)：146－149．
Zou Xiaobo, Zhao Jiewen. Distinguishing different cultivar apples by electronic nose based on support vector machine[J] Transactions of the CSAE，2007，23(1)：146－ 149. (in Chinese with English abstract)．

[8] 王定成．温室环境的支持向量机回归建模[J]．农业机械学报，2004，35 (5)：106－ 109．
Wang Dingcheng. Svm regression modeling for greenhouse environment[J]. Transactions of the Chinese Society for Agricultural Machinery，2004，35(5)：106－109．(in Chinese with English abstract)

[9] 袁哲明，张永生，熊洁仪．基于 SVR 的多维时间序列分析及其在农业科学中的应用[J]．中国农业科学，2008，41(8)：2485－2492．
Yuan Zheming, Zhang Yongsheng, Xiong Jieyi. Multidimensional time series analysis based on support vector machine regression and its application in agriculture[J]. Scientia Agricultura Sinica，2008，41(8)：2485－2492．(in Chinese with English abstract)

[10] 何丕廉，侯越先，常虹，等．基于神经网络的时间序列鲁棒预测[J]．控制与决策，2001，16(3)：333－336．
He Pilian, Hou Yuexian, Chang Hong, et al. Robust time series prediction of neural network[J]. Control and Decision，2001，16(3)：333－336．(in Chinese with English abstract)

[11] Burges C J C. A tutorial on support vector machines for pattern recognition[J]. Data Mining and Know ledge Discovery，1998，2 (2)：121－ 169．

[12] 梅虎，梁桂兆，周原．支持向量机用于定量构效关系建模的研究[J]．科学通报，2005，50(16)：1703－1708．
Mei Hu, Liang Guizhao, Zhou Yuan. Support vector machine applied in QSAR modeling[J]. Chinese Science Bulletin，2005，50(16)：1703－1708．(in Chinese with English abstract)

[13] Steve R. Gunn. Support vector machines for classification and regression[R]. Southampton: University of Southampton，1998：1－ 28．

[14] Smola A J, Scholkopf B. A tutorial on support vector regression[J]. Statistics and Computing，2004，14(3)：199－222．

[15] Sánchez A V D. Advanced support vector machines and kernel methods[J]. Neurocomputing，2003，55(1)：5－20．

[16] 唐启义，冯明光．实用统计分析及其 DPS 数据处理系统[M]．北京：科学出版社，2002．

[17] Tropsha A, Gramatica P, Gombar V. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models[J]. QSAR and Combinatorial Science，2003，22(1)：69－77．

[18] Tropsha A. Beware of q2![J] . Journal of Molecular Graphics and Modelling，2002，20(4)：269－276．

Winter wheat yield estimation based on support vector machine regression and multi-temporal remote sensing data

Li Rui[1,2], Li Cunjun[1※], Xu Xingang[1], Wang Jihua[1], Yang Xiaodong[1], Huang Wenjiang[1], Pan Yuchun[1]

(1. *National Engineering Research Center for Information Technology in Agriculture*, *Beijing* 100097, *China*;
2. *Information Engineering Institute*, *Capital Normal University*, *Beijing* 100048, *China*)

Abstract: Developing and establishing high accurate models for crop yield estimation using remote sensing is of great significance in decision making for national food import/export and food security. A machine learning methodology called support vector machine regression (SVR) was introduced to construct remote sensing estimation model. Firstly, NDVIs from multi-temporal Landsat TM for main growing stage of winter wheat in 2004 and 2007 in Beijing suburb were used to construct yield estimation model by remote sensing through SVR. Secondly, cross validation was made on the model's stability and forecasting ability, and then the performance of SVR methodology was compared with traditional multivariate linear regression (MLR) methodology. The results showed that yield estimation model by remote sensing based on SVR could increase the precision of yield prediction. The determination coefficients were increased by 0.2162 and 0.2158, respectively, while the root mean squared errors were decreased by 0.1682 and 0.2912 in 2004 and 2007 compared with the multivariate regression methodology. Therefore, it is feasible and effective to estimate winter wheat yield by constructing estimation model based on SVR and multi-temporal remote sensing data，which provides the method to estimate the winter wheat yield via multi-temporal remote sensing data.

Key words: support vector machine regression, multi-temporal remote sensing, yield estimation, NDVI