中国农业科学 2008,41(8):2485-2492
Scientia Agricultura Sinica doi: 10.3864/j.issn.0578-1752.2008.08.039

基于 SVR 的多维时间序列分析及其在农业科学中的应用

袁哲明，张永生，熊洁仪

（湖南农业大学生物安全科学技术学院，长沙 410128）

摘要：【目的】建立一种基于结构风险最小、既反映样本集动态特征又体现环境因子影响的高精度非线性多维时间序列预测方法。【方法】耦合支持向量机回归（SVR）和带受控项的自回归模型（CAR），以留一法基于 MSE 最小原则实施模型定阶和变量筛选，以一步预测法检验新模型 SVR-CAR 的有效性，并通过强制汰选给出各保留变量对预测的相对重要性次序。【结果】3 个农业科学实例验证表明，SVR-CAR 在 7 种参比模型中预测精度最高，且可更精细地反映样本集的非线性动态特征，依各保留变量对预测的相对重要性次序及其动态变化可赋予保留变量部分解释能力。【结论】SVR-CAR 是一种基于 SVR 并融合时间序列分析和回归分析的非线性多维时间序列分析方法，具结构风险最小、非线性、适于小样本，能有效克服过拟合、维数灾和局极小，非线性定阶和非线性筛选变量，自动选择核函数及其相应参数，泛化推广能力优异、预测精度高等诸多优点，在农业科学、生态学、经济学等领域有广泛应用前景。

关键词：多维时间序列；支持向量机回归；预测；均方误差

Multidimensional Time Series Analysis Based on Support Vector Machine Regression and Its Application in Agriculture

YUAN Zhe-ming, ZHANG Yong-sheng, XIONG Jie-yi

(*Bio-safety Science and Technology College, Hunan Agricultural University, Changsha 410128*)

Abstract: 【Objective】To construct a novel nonlinear multidimensional time series approach based on structural risk minimization, which shows the dynamic characteristics of sample set as well as the effect of environmental factors.【Method】Integrated controlled autoregressive (CAR) into support vector machine regression (SVR), a novel nonlinear multidimensional time series model named as SVR-CAR was proposed. After estimated the order and screened the variable by leave-one-out method based on the minimum mean square error, the reliability of SVR-CAR was validated by one-step prediction method.【Result】 The prediction results of the agricultural sample set showed that SVR-CAR had the highest prediction precision in all reference models, characterized the nonlinear dynamic of the sample sets subtly, and explained the effect of variable to prediction partly according to the order of the restrained variable screened compulsorily. 【Conclusion】As a novel nonlinear multidimensional time series analysis approach integrated CAR into SVR, SVR-CAR had the advantages of structural risk minimization, non-linear characteristics, avoiding the over-fit, strong generalization ability and high prediction precision, *etc*. SVR-CAR, can be widely used in the prediction area of agriculture, ecology and economics.

Key words: Multidimensional time series; Support vector machine regression; Forecast; Mean square error

0 引言

【研究意义】多维时间序列分析是在考虑因变量时序变动的基础上融入多个自变量控制作用的一种建模方法[1]。自然和社会经济现象中存在大量非线性、高维特征的复杂时间序列，如农业科学领域涉及的粮食产量、病虫害发生量、农业气象、农业生产指数、动植物生长等[1~3]。预测是认识和决策的依据，多维时

收稿日期：2006-12-14；**接受日期：**2007-01-19
基金项目：国家自然科学基金（30570351）和教育部新世纪优秀人才支持计划（NCET-06-0710）
作者简介：袁哲明（1971-），男，湖南岳阳人，教授，博士，研究方向为昆虫生态及预测预报。Tel：0731-4618163；E-mail：zhmyuan@sina.com

间序列预测的准确性目前仍是一个巨大挑战，发展高精度的多维时间序列预测分析方法意义重大[4]。【前人研究进展】Box 等[5]、Hannan[6]最早提出经典的线性多维时间序列分析模型－带控制项的自回归滑动平均模型（controlled autoregressive integrating moving average，CARMA）；Boker 等[7]给出了其定阶的 F 检验判别法；邓自立等[8]进一步建立了阶、子阶和时滞的 F 检验判决器，形成了对 CARMA 模型结构的完整辨识，并给出了其简化形式－带受控项的自回归模型（controlled autoregressive，CAR）。人工神经网络（artificial neural networks，ANN）具有很好的非线性逼近能力，基于 ANN 的非线性时间序列分析或非线性回归分析已有大量报道[9,10]。Vapnik 基于统计学习理论提出的支持向量机（support vector machine，SVM）是目前发展最快的机器学习方法[11,12]，它最初用于模式识别（SVC），随 ε 不敏感损失函数的引入，现已扩展到用于非线性时间序列分析或非线性回归分析（SVR）[13~19]；SVR 基于结构风险最小，较好地解决了小样本、非线性、过拟合、维数灾和局极小等问题，泛化推广能力优异[11,12]。【本研究切入点】多维时间序列数据往往既隐含大量的动态特征，又受环境因子的影响，同时具有高度的非线性性，因此宜融合时间序列分析和回归分析进行非线性建模。CARMA 和 CAR 虽然都融合了时间序列分析和回归分析，但其基于线性进行模型定阶和逐步回归变量筛选（stepwise linear regression，SLR）获得的模型阶数和保留变量往往并非最优，实际应用中预测能力较弱，因而建立一种非线性的模型定阶和变量筛选方法是必要的[9]。ANN 是非线性的，但融合时间序列分析和回归分析的 ANN 模型尚不多见；即或有，ANN 也存在基于经验最小化、模型结构难以确定、易于出现过度训练和训练不足、陷入局部最小、对连接权初值敏感、过度依赖设计技巧等诸多缺陷。SVR 是非线性且基于结构风险最小的，但基于 SVR 并融合时间序列分析和回归分析的非线性多维时间序列分析方法国内外未见报道；同时，SVR 在实际应用中尚存在一定的局限性：一是其核函数的选取是经验性的[17,20]；二是有些 SVR 程序其核函数相应参数的选取也是经验性的[20]；三是与 ANN 一样，SVR 模型的可解释性较差[20]。【拟解决的关键问题】本研究基于 SVR 和 CAR，拟依训练集留一法交叉测试均方误差（mean squared error，MSE）最小原则实施模型定阶，依 MSE 最小原则以多轮末尾淘汰法实施变量筛选，从而建立非线性的模型定阶和变量筛选方法。同时，拟依 MSE 最小原则自动、动态选用核函数，采用 LIBSVM2.8（http://www.csie.ntu.edu.tw/ ～ cjlin/libsvm）结合 gridregression.py 自动搜索并非经验性地确定最佳惩罚参数、灵敏度及径向宽度等核函数参数，通过变量强制筛选并依淘汰顺序给出变量的相对重要性次序以赋予 SVR 部分的可解释性，从而克服或部分克服 SVR 的局限性。在此基础上，为粮食产量、病虫害发生量等预测领域建立一种基于 SVR、融合时间序列分析和回归分析、预测精度高、适用范围广的非线性多维时间序列分析方法（SVR-CAR），在计算机上程序化实现并以 3 个农业科学实例验证 SVR-CAR 的有效性。

1 研究方法

1.1 SVR-CAR 建模过程

1.1.1 SVR 基本原理 SVM 起初是用于解决线性可分情况下两类样本的分类问题（SVC），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。SVR 与 SVC 相似，但 SVR 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，给定样本集（x_i，y_i），其中 $i=1,\ldots,n$；$x\in R^d$；$y\in R$，问题变为寻求一个最优超平面，使得在给定精度 ε（ε≥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 $\xi^*\geqslant 0$ 以及惩罚参数 C＞0，其寻优问题转化相应的二次规划问题为：

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi i+\xi i^*)$$

$$\text{约束条件}\begin{cases} y_i - w\cdot x - b \leqslant \varepsilon + \xi_i \\ w\cdot x + b - y_i \leqslant \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geqslant 0 \end{cases}$$

将该优化问题转化成对偶问题后可解得最优回归函数为：

$f(x) = w\cdot x + b = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)(x\cdot x_i) + b$，其中 $0\leqslant \alpha_i$，$\alpha_i^* \leqslant C$。

仅有少部分样本的 Lagrange 乘子 α 不为零，因此决定该超平面的样本只能是这些支持向量。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。通常，特征空间具很高甚至无穷维数，致使空间变换后计算量巨增而面临维数灾难等问题。幸运的是 SVM 中待解对偶问题

只包含一个变换后特征空间内积运算，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（t=1，d=2）、多项式核（t=1，d=3）、径向基核（t=2）和 sigmoid 核（t=3）等。有关 SVM、SVR 的详细内容参见文献[11,12,20,21]。

1.1.2　模型定阶与核函数选取　假定一多输入单输出回归模型有 N 个样本、一个因变量、m-1 个自变量，由低阶到高阶递增地以 SVR 进行留一法测试（原始变量 svmscale 规格化到[-1，+1]），并依 MSE 最小标准决定拓展阶数与否。对待比较的相邻两模型 SVR（n）和 SVR（n+1），记 $MSE_{SVR(n)}$为 SVR（n）的均方误差，$MSE_{SVR(n+1)}$为 SVR（n+1）的均方误差。若 $MSE_{SVR(n)}$＞$MSE_{SVR(n+1)}$，继续拓阶；若 $MSE_{SVR(n)}$≤$MSE_{SVR(n+1)}$，拓阶终止，取 SVR（n）为定阶后模型。

对给定的 5 种常用核函数，依次依训练集留一法交叉测试 MSE 最小标准进行模型定阶，并以 MSE 最小标准确定最优核函数及相应模型阶数。

1.1.3　变量筛选及其强制汰选　假定多输入单输出回归模型最优核函数和最高阶 n 确定后有 N'个样本、p 个输入变量，现以多轮末尾淘汰法从包含全部输入变量的 SVR 模型中以留一法（原始变量 svmscale 规格化）依 MSE 最小标准逐次剔除对提高预测精度有不利影响的变量。

对第一轮筛选，记 $MSE_{(x1,x2,\ldots,xi,\ldots,xp)}$为 p 个输入变量的均方误差，$MSE_{(x1,x2,\ldots,xi\text{-}1,xi+1,\ldots,xp)}$为剔除第 i 个输入变量后的均方误差。如 $\min[MSE_{(x1,x2,\ldots,xi\text{-}1,xi+1,\ldots,xp)}]$＞$MSE_{(x1,x2,\ldots,xi,\ldots,xp)}$，表明没有变量可剔除，汰选结束；反之，剔除第 i 个变量后进入下一轮筛选（注意此时 p 变为 p－1），直至没有变量可剔除为止[22]。汰选结束后的保留变量用于后续建模预测。

为基于 SVR-CAR 给出保留变量的相对重要性次序，可进一步采用多轮末尾淘汰法对保留变量进行强制汰选，每轮淘汰一个$[MSE_{(x1,x2,\ldots,xi\text{-}1,xi+1,\ldots,xp)}]$最小的保留变量，直到只剩下一个保留变量为止。

1.1.4　预测评价指标　对全部 N 个样本，遍历核函数参数 c、g、p 组合（c∈[－1，6]，g∈[－8，0]，p∈[－8，－1]，步长均为 1），经 svmtrain 建模并以 svmpredict 回代，能搜索到一组最优核函数参数组合使 SVR 如 ANN 一样对样本集的拟合达到极高精度，但使用该参数组合建模往往实际预测效果很差。事实上，在 SVR 中 gridregression.py 并不提供针对全部 N 个样本回代的自动参数寻优，而以 n-fold 交叉验证（其极限是留一法）避免过拟合。一方面，至少对 ANN 和 SVR 而言过高的回代拟合精度并无多大实际意义；另一方面，对预测模型特别是多维时间序列模型人们真正感兴趣的是实际预测能力而非回代拟合结果。因此，本文以实际预测结果作为模型优劣的评价基准。

为避免单个样本预测的偶然性，视时间序列长短，规定至少连续选取时间序列最后 5 个以上样本作为预测样本。在预测第 i 个样本时，其自身及后续未来样本不得参与建模训练；在预测第 i+1 个样本时，第 i 个样本加入训练样本（一步预测法）。预测结果优劣采用 MSE、平均绝对误差百分率（mean absolute percentage error，MAPE）和 Q^2_{ext}作为评价指标：

$$\mathrm{MSE}=\frac{\sum(y_i-\hat{y}_i)^2}{n}$$

$$\mathrm{MAPE}\%=\frac{\sum|y_i-\hat{y}_i|/y_i}{n}\times 100$$

$$Q^2_{ext}=1-\frac{\sum(y_i-\hat{y}_i)^2}{\sum(y_i-\overline{y}_{(i,\mathrm{train})})^2}$$

式中，y_i为真值，$\hat{y}_i$为预测值，n 为预测样本数，$\overline{y}_{(i,\mathrm{train})}$为第 i 个待测样本的训练样本（即前 i－1 个样本）真值的平均[23]。MSE 仅适用于同一数据集不同模型间的比较，MAPE%可用于不同数据集间的比较。但对同一数据集，如 A 模型与 B 模型相比虽 MAPE%较大而 MSE 较小，则 A 模型预测更为稳健；因此，MSE 为主要评价指标。Q^2_{ext}可对单个模型的外部预测能力是否达到统计所需精度给出度量，一般认为 Q^2_{ext}＞0.6 模型才有实际预测价值[23]。

SVR-CAR 以自编 C++程序通过调用 LIBSVM2.8 实现并经验证通过。

1.2　参比模型

所有参比模型同样采用一步预测法。多元线性回归（multiple linear regression，MLR）和逐步线性回归（stepwise linear regression，SLR；Backward 法，剔除变量 α=0.2）由 SPSS13.0 给出。CAR 由 DPS6.55 给出（默认参数设置）[22]。时间序列趋势分析包括一次滑动平均、一次指数平滑、线性回归、二次滑动平均、一次平滑、二次指数平滑、三次指数平滑等 7 个亚模型，在预测第 i 样本时，取前 i－1 个样本 y 值分别以 DPS6.55 拟合建模（默认参数设置）[22]，以拟合 MSE 最小亚模型预测值作为该样本时间序列趋势分析最终预测结果。SVR 由 LIBSVM2.8 给出，自变量既不拓阶也不筛选，原始变量 svmscale 规格化后以留一法

gridregression.py 寻优，以 5 个常用核函数中 MSE 最小者为最优核函数建模预测。基于 SVR 的逐步非线性回归（stepwise nonlinear regression，SNR）以自编 C++程序通过调用 LIBSVM2.8 实现并经验证通过，除不拓展阶数外与 SVR-CAR 相同。显然，除时间序列趋势分析外的 6 种模型明显可分为 3 对：MLR－SVR、SLR－SNR、CAR－SVR－CAR，每对模型中前者为线性模型，后者为基于 SVR 的对应非线性模型。

2 实例分析

2.1 中国农业总产值指数预测

表 1 数据引自文献[22]，后 10 年农业总产值指数一步预测结果如表 2。（1）回归模型和时间序列趋势分析模型 Q^2_{ext} 均大于 0.6，这表明中国农业总产值指数既受农业劳动力、粮食产量和农业税诸因子的影响，又隐含动态特征，宜融合回归分析和时间序列分析统一建模。（2）从 MSE、MAPE%和 Q^2_{ext} 看，在不拓阶（即只考虑回归而忽略样本集动态特征）时，MLR 优于 SVR，SLR 优于 SNR，这表明当年农业劳动力、粮食产量和农业税诸因子对当年农业总产值指数的影响更多地呈现为线性关系。（3）拓阶即统一考虑回归分析和时间序列分析的 CAR 反劣于 MLR 和 SLR，这

表 1 中国 1952～1980 年农业总产值指数与诸因子数据

Table 1 Index of agricultural total output value and its correlative factors from 1952 to 1980 in China

年份 Year	农业总产值指数 (y) Index of agricultural total output value (y)	农业劳动力 (x_1,万人) Agricultural labor force (x_1, ten thousand persons)	粮食产量 (x_2, 万吨) Crop yield (x_2, ten thousand tons)	农业税 (x_3, 亿元) Agricultural taxes (x_3, 100 million *yuan* RMB)
1952	100	17317	16392	27
1953	103.1	17748	16683	27.1
1954	106.6	18152	16952	32.8
1955	114.7	18593	18394	30.5
1956	120.5	18545	19275	29.7
1957	124.8	19310	19505	29.7
1958	127.8	15492	20000	32.6
1959	110.4	16273	17000	33
1960	96.4	17019	14350	28
1961	94.1	19749	14750	21.7
1962	99.9	21278	16000	22.8
1963	111.6	21968	17000	24
1964	126.7	22803	18750	25.9
1965	137.1	23398	19453	25.8
1966	149	24299	21400	29.6
1967	151.2	25167	21782	29
1968	147.5	26065	20906	30
1969	149.2	27119	21097	29.6
1970	166.3	27814	23996	32
1971	171.4	28400	25014	30.9
1972	171.1	28286	24048	28.4
1973	185.5	28861	26494	30.5
1974	193.2	29222	27527	30.1
1975	202.1	29460	28452	29.5
1976	207.1	29448	28631	29.1
1977	210.6	29345	28273	29.3
1978	229.6	29426	30477	28.4
1979	249.4	29425	33212	29.5
1980	259.1	30211	32056	27.7

可能与 CAR 至少必然拓展一阶有关，如果由 SVR（0）拓展到 SVR（1）时 MSE 增大，则这种必然拓展一阶的做法显然是不合适的。拓阶后保留变量间（如当年粮食产量和上年粮食产量）可能存在的共线性同时使得 CAR 模型的可解释性变差。（4）经非线性定阶和非线性变量筛选后的 SVR-CAR 在表 2 所列模型中预测最优，这表明非线性地统一考虑回归分析和时间序列分析是必要的。与 CAR 不同，SVR-CAR 并不必然拓展一阶而要求一个给定的最低概率保证，并确保在 α=1 时也不会出现由 SVR（0）拓展到 SVR（1）时 MSE 增大反而拓阶的情形。拓展阶数、筛选变量后系统非线性化程度增强，SVR-CAR 精细地刻画了样本集的这种非线性变化，因而预测精度最优。

系统考察 1952～1970、1952～1971、……、1952～1980 年期的变量筛选与保留变量强制汰选过程及其动态变化（表 3 仅给出 1952～1980 年期细节），结果表明：当年粮食产量始终稳定的对农业总产值指数的预测影响最大。农业劳动力至 1977 年止对农业总产值指数的预测都有较大影响，但随后逐渐下降，至 1980 年已基本无影响变为非保留变量（表 3）；这可能反映 20 世纪 70 年代末期，随“联产承包责任制”的实施，农业劳动力已显露过剩苗头。1973 年以后，农业税特别是上年农业税对农业总产值指数的预测有一定影响。上年农业总产值指数对当年农业总产值指数预测的影响明显阶梯式地增大，至 1977 年其影响已仅次于当年粮食产量；这可能与研究期间中国农业总产值

表 2　农业总产值指数预测

Table 2　Predication of the index of agricultural total output value

年份 Year	真值 True value	MLR	SVR	SLR	SNR	CAR	SVR-CAR	时间序列预测 Time series analysis
1971	171.4	176.1	171.3	176.1	162.6	189.1	176.1	194.8
1972	171.1	167.3	163.9	167.3	166.9	169.9	166.9	171.3
1973	185.5	185.0	173.6	186.3	170.8	186.1	180.9	171.1
1974	193.2	191.8	179.3	191.8	178.9	167.8	193.2	185.2
1975	202.1	198.0	198.4	198.0	170.9	201.0	199.9	193.0
1976	207.1	200.3	196.2	201.6	197.4	206.2	204.5	201.9
1977	210.6	199.7	197.3	200.1	197.5	207.8	207.1	207.0
1978	229.6	217.5	205.0	217.6	199.9	222.7	219.3	213.3
1979	249.4	242.2	234.1	240.8	245.4	249.3	240.4	229.2
1980	259.1	236.9	226.8	235.0	236.8	248.5	248.9	273.5
	MSE	91.2	257.6	99.3	315.2	112.7	37.4	182
	MAPE%	3.31	6.05	3.37	7.24	3.39	2.40	5.51
	Q^2_{ext}	0.9842	0.9553	0.9828	0.9454	0.9804	0.9935	0.9684

表 3　全部样本拓展一阶后变量筛选与保留变量强制汰选过程及其 MSE 值

Table 3　Screening variables and their MSE values in different steps according to table 1

阶段 Stage	轮次 Step	汰选前 Before screen	x_1	x_2	x_3	上年 y y of last year	上年 x_1 x_1 of last year	上年 x_2 x_2 of last year	上年 x_3 x_3 of last year	淘汰变量 Variables screened
变量筛选 Screening variables	1	21.6	21.2	81.4	22.0	52.2	16.2	21.9	31.7	上年 x_1 x_1 of last year
	2	16.2	12.7	70.5	17.3	47.3	-	16.1	26.2	x_1
保留变量强制汰选 Screening retained variables compulsorily	3	12.7	-	58.2	12.8	30.0	-	13.1	17.1	x_3
	4	12.8	-	54.9	-	42.9	-	13.5	16.3	上年 x_2 x_2 of last year
	5	13.5	-	55.0	-	62.6	-	-	20.2	上年 x_3 x_3 of last year
	6	20.2	-	68.2	-	50.2	-	-	-	上年 y y of last year

指数的时间序列特征愈趋明显有关。

2.2 某地春粮产量预测

18 年春粮产量（y，5×10^5 kg）数据引自文献[24]，4 个初始自变量为春粮播种面积（x_1，万亩）、化肥施用量（x_2，5×10^5 kg）、饲养肥猪头数（x_3，万头）和水稻扬花期降水（x_4，10 mm）。后 5 年春粮产量一步预测结果如表 4。（1）MLR 模型和时间序列趋势分析模型 Q^2_{ext} 大于 0.6，表明预测当地春粮产量同样宜融合回归分析和时间序列分析统一建模。从 MSE、MAPE%和 Q^2_{ext} 看，（2）在不拓阶时，MLR 优于 SVR，SLR 优于 SNR，说明当年春粮播种面积、化肥施用量、饲养肥猪头数和水稻扬花期降水诸因子对当年春粮产量的影响更多地呈现为线性关系。（3）至少必然拓展一阶的 CAR 模型明显劣于 MLR 和 SLR；且 Q^2_{ext}=0.3168＜0.6，模型预测失真。表明拓阶后样本集已不能由线性模型预测。（4）SVR-CAR 在表 4 所列模型中预测最优，表明非线性地统一考虑回归分析和时间序列分析同样是必要的。拓展阶数、筛选变量后系统非线性化程度明显增强，SVR-CAR 准确地捕捉到了样本集的这种非线性变化，因而预测精度最优。

表 4 春粮产量预测

Table 4 Predication of crop yields in spring

年序 No. of year	真值 True value	MLR	SVR	SLR	SNR	CAR	SVR-CAR	时间序列预测 Time series analysis
14	618	624.5	816.8	618	616.9	659	595.9	629.5
15	742	656.7	524.0	630.4	524.0	801.3	787.0	642.6
16	805	903.6	511.7	856	974.9	833.7	880.5	721.7
17	859	848.2	845.3	852.7	846.1	415.2	846.4	773.6
18	855	861.9	759.8	892.6	843.8	941.9	843.8	904.4
	MSE	3441	36467	3302	15341	42106	1698	5337
	MAPE%	5.4	22.1	5.3	10.7	16.0	4.4	8.3
	Q^2_{ext}	0.9442	0.4083	0.9464	0.7511	0.3168	0.9724	0.9134

因本例预测年份较少，仅考察了年序 1～18 年期的变量筛选与保留变量强制汰选过程（限于篇幅，过程未列出），结果表明：很自然地，上年春粮播种面积和上年水稻扬花期降水对春粮产量预测基本无影响（非保留变量）。保留变量中，对春粮产量预测影响的重要性从大到小排序如下：当年饲养肥猪头数＞当年化肥施用量＞当年春粮播种面积＞上年化肥施用量＞上年春粮产量＞上年饲养肥猪头数＞当年水稻扬花期降水。这表明，肥料是影响当地春粮产量预测准确性的关键因素，且肥料的影响存在至少一年的后效；由于当地春粮播种面积年际间较为稳定，其对春粮产量预测重要性比想象的要小；上年春粮产量对当年春粮产量预测有一定影响，当地春粮产量存在一定的时间序列特征；当年水稻扬花期降水对春粮产量预测虽有影响但较弱，这从第 4、8、15 年扬花期降水量较大而春粮产量未受明显影响可得到进一步佐证。

2.3 二代玉米螟危害程度预测

21 年二代玉米螟危害程度（y，虫株率%）数据引自文献[25]，2 个初始自变量为一代残虫基数（x_1，头/百株）和 7 月上旬的温雨系数（x_2，mm/℃）。后 5 年二代玉米螟危害程度一步预测结果如表 5。同样地，从 MSE、MAPE%和 Q^2_{ext} 看，SVR-CAR 均明显

表 5 二代玉米螟危害程度预测

Table 5 Forecasting damage degree of the 2nd generation corn bore

年序列 No. of year	真值 True value	MLR	SVR	CAR	SVR-CAR	时间序列预测 Time series analysis
17	10	15.8	30.7	12.1	16.4	25.1
18	38	32.5	33.6	26.6	34.0	23.1
19	27	42.9	39.5	41.8	31.0	27.4
20	12	18.0	22.9	13.2	8.1	25.2
21	31	36.8	31.1	34.5	32.0	25.7
	MSE	77.3	144.6	73.4	17.8	130.5
	MAPE%	40.01	71.21	25.42	25.01	63.76
	Q^2_{ext}	0.6464	0.3384	0.6640	0.9184	0.4028

优于参比模型。由于初始自变量较少，本例未进行 SLR 和 SNR 预测；限于篇幅，SVR-CAR 变量筛选与保留变量强制汰选过程从略。

3　讨论

与 ANN 一样，由于不存在一个解析的表达式，SVR、SVR-CAR 对因子欠缺解释能力[20]。对 MLR 模型，回归方程的系数反映了每个自变量对因变量值边际贡献率大小；在不存在共线性时，系数的正负指明了自变量对因变量增强或减弱的方向[26]。ANN 模型只建立了抽象的函数，不能用与回归系数同样的方式来解释神经网络的权重；若要用 ANN 分析输入对于输出的影响，则必须以灵敏度分析的形式，通过改变输入水平来观察对应的输出水平[26]。对于 SVR-CAR 模型，本文提出的保留变量强制汰选方法可给出各保留变量对预测的相对重要性次序，并在样本充分大时可观察到各保留变量排序次序变化的动态过程，为尝试解释因子提供了一种新的途径，但其合理性有待进一步研究。必需指出，对 SVR 和 SVR-CAR 建立的非线性模型，单个自变量与因变量的单调递增或递减关系往往仅在自变量的某个取值范围内成立，并受其它自变量的影响；各保留变量对预测的相对重要性次序也不能生硬地对应为 MLR 模型中各自变量回归方程系数绝对值的排序，而仅反映留一法建模时对预测精度的影响大小。现阶段，SVR-CAR 对于预测可能比对决策分析更为有用，对因子的进一步解释可考虑结合保留变量强制汰选与灵敏度分析进行。

4　结论

提出了一种新的基于 SVR 的非线性多维时间序列分析方法（SVR-CAR）并在计算机上程序化实现。SVR-CAR 融合了时间序列分析和回归分析，具有基于结构风险最小，非线性，适于小样本，避免过拟合（核函数参数较少且以留一法建模）、维数灾和局极小，泛化推广能力优异，非线性定阶、非线性筛选变量，基于 MSE 最小自动选择核函数、基于 gridregression.py 自动搜索确定最优核函数参数，操作较 ANN 相对简便等许多优点；通过保留变量强制汰选，SVR-CAR 可给出保留变量对预测的相对重要性次序从而赋予因子部分解释能力；基于严格的一步预测法，实例验证表明 SVR-CAR 预测精度高，在农业科学、生态学、经济学等领域有广泛应用前景。

References

[1] 吴承祯, 洪　伟. 林木生长的多维时间序列分析. 应用生态学报, 1999, 10(4): 395-398.
Wu C Z, Hong W. Multidimensional time series analysis on tree growth. *Chinese Journal of Applied Ecology*, 1999, 10(4): 395-398. (in Chinese)

[2] 吴承祯, 洪　伟. 长苞铁杉种群个体年龄与胸径的多维时间序列模型研究. 植物生态学报, 2002, 26(4): 403-407.
Wu C Z, Hong W. A proposed multidimensional time series model of individual age and diameter in tsuga longibracateata. *Acta Phytoecologica Sinica*, 2002, 26(4): 403-407. (in Chinese)

[3] 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. 植物保护学报, 1995, 22(1): 1-6.
Zhou L Y, Fei H X, Zhang X X. The application of multiple dimension time series analysis method in long-term forecasting of rice leaf roller. *Acta Phytophylacica Sinica*, 1995, 22(1): 1-6. (in Chinese)

[4] 向书坚. 20 世纪 90 年代时间序列预测领域主要研究动态. 中南财经大学学报, 2001, (2): 31-36.
Xiang S J. Main research dynamics in time series forecasting in 1990s. *Journal of Zhongnan University of Finance and Economics*, 2001, (2): 31-36. (in Chinese)

[5] Box Q E P, Jenkins G M. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-day Press, 1970.

[6] Hannan E J. The estimation of the order of an ARMA process. *Annals of Statistics*, 1980, 8(5): 1071-1081.

[7] Boker J, Keviczky L. Structural properties and structure estimation of vector difference equations. *International Journal of Control*, 1982, 36(3): 461-476.

[8] 邓自立, 郭一新. 动态系统分析及其应用. 沈阳: 辽宁科学技术出版社, 1985: 31-130.
Deng Z L, Guo Y X. *Dynamic System Analysis and Its Application*. Shenyang: Liaoning Science & Technology Press, 1985: 31-130. (in Chinese)

[9] 何丕廉, 侯越先, 常　虹, 孙学军. 基于神经网络的时间序列鲁棒预测. 控制与决策, 2001, 16(3): 333-336.
He P L, Hou Y X, Chang H, Sun X J. Robust time series prediction of neural network. *Control and Decision*, 2001, 16(3): 333-336. (in Chinese)

[10] Chakraborty K, Mehrotra K, Kmohan C, Ranka S. Forecasting the behavior of multivariate time series using neural networks. *Neural Networks*, 1992, (5): 961-970.

[11] Vapnik V N. *The Nature of Statistical Learning Theory*. New York: Springer Verlag Press, 1995.

[12] 邓乃扬, 田英杰. 数据挖掘中的新方法－支持向量机. 北京: 科学出版社, 2004: 77-162, 224-272.
Deng N Y, Tian Y J. *Support Vector Machine－A New Method in Data Mining*. Beijing: Science Press, 2004: 77-162, 224-272. (in Chinese)

[13] 马晓光, 胡 非. 利用支撑向量机预报大气污染物浓度. 自然科学进展, 2004, 14(3): 349-353.
Ma X G, Hu F. Forecasting the concentration of air pollutant using support vector machine. *Progress in Natural Science*, 2004, 14(3): 349-353. (in Chinese)

[14] Pai P F, Hong W C. Support vector machines with simulated annealing algorithms in electricity load forecasting. *Energy Conversion and Management*, 2005, 46: 2669-2688.

[15] Thissena U, van Brakela R, de Weijerb A P, Melssena W J, Buydens L M C. Using support vector machines for time series prediction. *Chemometrics and Intelligent Laboratory Systems*, 2003, 69: 35- 49.

[16] Pai P F, Lin C S. A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega*, 2005, 33(6): 497-505.

[17] Liang Y C, Sun Y F. An improved method of support vector machine and its applications to financial time series forecasting. *Progress in Natural Science*, 2003, 13(9): 696-700.

[18] Chang B R. Applying nonlinear generalized autoregressive conditional heteroscedasticity to compensate ANFIS outputs tuned by adaptive support vector regression. *Fuzzy Sets and Systems*, 2006, (157): 1832-1850.

[19] Rojo-Álvarez J L, Camps-Valls G, Martínez-Ramón M, Soria-Olivas E, Navia-Vázquez A, Figueiras-Vidal A R. Support vector machines framework for linear signal processing. *Signal Processing*, 2005, (85): 2316-2326.

[20] 梅 虎, 梁桂兆, 周 原, 李志良. 支持向量机用于定量构效关系建模的研究. 科学通报, 2005, 50(16): 1703-1708.
Mei H, Liang G Z, Zhou Y, Li Z L. Support vector machine applied in QSAR modeling. *Chinese Science Bulletin*, 2005, 50(16): 1703-1708. (in Chinese)

[21] Cristianini N, Shawe-Taylor J. 李国正, 王 猛, 曾华军, 译. 支持向量机导论. 北京: 电子工业出版社, 2004: 82-139.
Cristianini N, Shawe-Taylor J. Translated by Li G Z, Wang M, Zeng H J. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Beijing: Electronics Industry Press, 2004: 82-139. (in Chinese)

[22] 唐启义, 冯明光. 实用统计分析及其 DPS 数据处理系统. 北京:科学出版社, 2002: 525-585.
Tang Q Y, Feng M G. *DPS Data Processing System for Practical Statistics*. Beijing: Science Press, 2002: 525-585. (in Chinese)

[23] 禹新良, 王学业, 高进伟, 李小兵, 王寒露. 用量子化学参数研究烯烃聚合物定量构效关系. 化学学报, 2006, 64(7): 629-636.
Yu X L, Wang X Y, Gao J W, Li X B, Wang H L. QSPR studies on vinyl polymers based on quantum chemical descriptors. *Acta Chimica Sinica*, 2006, 64(7): 629-636. (in Chinese)

[24] 裴鑫德. 多元统计分析及其应用. 北京: 北京农业大学出版社, 1991: 479.
Pei X D. *Multivariate Statistic Analysis and Its Application*. Beijing: Beijing Agricultural University Press, 1991: 479. (in Chinese)

[25] 周潘金, 甄玉起, 刘广德. 利用列联表分析预报二代玉米螟危害程度. 昆虫知识, 1986, 23(5): 203-205.
Zhou P J, Zhen Y Q, Liu G D. Forecast the damage degree of the second generation corn bore with contingency table. *Entomological Knowledge*, 1986, 23(5): 203-205. (in Chinese)

[26] 潘大丰, 李 群. ANN 预测方法应用研究. 情报学报, 1999, 18(2): 105-112.
Pan D F, Li Q. Study on application of ANN forecasting method. *Journal of the China Society for Scientific and Technical Information*, 1999, 18(2): 105-112. (in Chinese)

（责任编辑 毕京翠）