

基于SARIMA-BP神经网络组合方法的MODIS叶面积指数时间序列建模
与预测

期刊:	光谱学与光谱分析
稿件 ID	151432
稿件类型:	论著
作者提交的日期:	2015-11-25
完整作者列表:	姜, 姜春雷 张, 张树清 张, 张策 李, 李华朋 丁, 丁小辉
关键词:	SARIMA, BP神经网络, LAI, SARIMA-BP神经网络组合方法, LAI时间序列建模与预测
研究方向:	遥感

SCHOLARONE™
Manuscripts

基于 SARIMA-BP 神经网络组合方法的 MODIS 叶面积指数时间序列建模与预测

姜春雷^{1, 2}, 张树清^{1*}, 张策³, 李华朋¹, 丁小辉^{1, 2}

1. 中国科学院东北地理与农业生态研究所, 吉林长春 130102;

2. 中国科学院大学, 北京 100049

3. 兰卡斯特大学环境中心, 兰卡斯特 4YQ, UK

摘要: 植被叶面积指数 (LAI) 时间序列的建模及预测是陆面过程模型和遥感数据同化方法的重要组成部分。MODIS LAI 时间序列包含线性部分和外在干扰产生的非线性部分, 单一的线性方法或非线性方法都不能对其精确建模和预测。本文采用季节自回归积分滑动平均 (SARIMA) 方法、BP 神经网络方法及二者的组合方法对 MODIS LAI 时间序列进行建模及预测。在 SARIMA-BP 神经网络组合方法中, SARIMA 方法用于建模及预测 LAI 时间序列中的线性部分, BP 神经网络方法用于对非线性残差部分进行建模及预测, 二者组合的结果可充分利用各自在线性与非线性建模的优势。实验结果表明: SARIMA-BP 神经网络组合方法对 MODIS LAI 时间序列具有更好的适应性, 其预测准确性高于 SARIMA 方法或 BP 神经网络方法。

关键词 SARIMA; BP 神经网络; LAI; SARIMA-BP 神经网络组合方法; LAI 时间序列建模与预测

中图分类号: 文献标识码: DOI:

引言

叶面积指数 (LAI) 指单位土地面积上植株叶片的单侧总面积^[1], 它是描述植被生长状态的重要指示因子, 是陆面过程模型及遥感数据同化的重要输入参数^[2-4]。上述应用中, LAI 历史时间序列的精确建模及准确预测将有利于提高相应计算的准确性^[5-8]。MODIS 数据产品 MOD15A2 提供了全球范围的历史 LAI 数据, 具有时间分辨率高、空间尺度大等独到优势, 是目前应用最为广泛的遥感数据源之一。然而 MODIS LAI 时间序列产品包含了一些低质量的数据。例如由于云层、气溶胶等的影响, 该产品在时间和空间上缺乏连续性^[9], 这些因素往往制约了传统方法的 MODIS LAI 建模精度和应用, 对其加以改进十分必要。

目前, LAI 时间序列建模主要分为两类: 一是基于物理模型, 如 Dickinson 等 (2008) 将温度和土壤湿度等做为驱动数据, 构建了基于物理模型的 DLM (dynamic leaf model) 方法用于 LAI 建模^[6]; 李喜佳等 (2013) 利用 DLM 方法描述 LAI 变化的动态过程^[4]。二是基于统计方法, 其中 SARIMA 是使用最为广泛的一种基于统计的时间序列建模方法, 如 Jiang 等 (2011) 分别采用动态谐波回归 (DHR) 方法、基于 loess 平滑的季节性长期趋势序列分解法 (STL) 和季节自回归积分滑动平均 (SARIMA) 三种方法对 MODIS LAI 时间序列数据

产品进行建模和预测^[7], 证明了 SARIMA 方法在建模与预测方面的优势; Xiao 等 (2011) 将 SARIMA 方法用于 LAI 建模并将建立的模型用于基于卡尔曼滤波的数据同化过程^[2]; 靳华安等 (2011) 利用 SARIMA 对 MODIS LAI 历史数据进行了建模并将其用于提高 LAI 反演精度^[3]。DLM 方法需要温度和土壤湿度等额外数据, 而这些数据在大尺度上往往难以获得, 另外由于所引入的温度和土壤湿度数据自身存在不确定性, 还可能带来额外误差。而 SARIMA 只对 LAI 历史数据进行拟合, 不需要额外的数据, 因此在实际应用中更具可行性。然而 MODIS LAI 时间序列包含了线性和非线性部分, SARIMA 作为一个线性方法^[10], 忽略了时间序列数据中非线性部分, 其预测结果与 MODIS LAI 时间序列产品之间必然存在一定误差, 针对 SARIMA 在 MODIS LAI 建模过程中未能拟合的非线性误差进行修正有望提高建模的精确性, 然而针对 LAI 建模文献中, 还未见非线性数据部分修正的报道。

BP 神经网络具有极强的学习能力和修正能力, 能够逼近任何非线性系统, 已经被成功地用于非线性系统的建模^[11-12]。本文首次将结合 SARIMA 和 BP 神经网络的方法应用到 LAI 时间序列建模, 以充分利用它们各自在线性拟合和非线性拟合中的优势, 并利用所建模型进行预测。

1 研究方法

收稿日期:

基金项目: 国家自然科学基金项目 (41271196), 中国科学院重点部署项目 (KZZD-EW-07-02) 资助

作者简介: 姜春雷, 1981 年生, 中国科学院东北地理与农业生态研究所博士研究生 e-mail: jiangchunlei@iga.ac.cn

*通讯联系人 e-mail: zhangshuqing@iga.ac.cn

由于云层和气溶胶等的干扰，本研究中使用的 MODIS LAI 数据产品中的时间序列产生了一些不平滑现象，也就是两个间隔时间点之间的 LAI 值向内凹陷或向外凸出，因而需要对原始数据进行平滑预处理：参照新华安等(2011)描述，首先利用 Savitzky-Golay (SG) 滤波平滑大部分受到干扰的 LAI 数据^[3]；再对小部分 SG 滤波未能平滑的部分，通过线性插值使其符合植物生长特性。平滑后的 MODIS LAI 时间序列数据仍包含线性部分和非线性干扰部分，且具有明显季节性特征。对于预处理后的时间序列数据，采用 SARIMA 方法进行建模，将 SARIMA 建模产生的残差序列用 BP 神经网络建模，建立的两个模型分别用于 MODIS LAI 时间序列中的线性部分和非线性部分预测，两者预测的结果之和被用做最终的结果。下面详细阐述所采用方法的原理：

1.1 季节自回归积分滑动平均 (SARIMA) 方法

ARIMA 方法假设一个变量未来的值同过去的若干个观测值和高斯噪声成线性关系，它最早由 Box 等公式化^[13]，记作 ARIMA (p, d, q)。带有季节性的 ARIMA 记作

SARIMA (p, d, q) (P, D, Q)^S, (P, D, Q)^S 是模型的季节性部分，它的公式如下^[13]：

$$\Phi(B^S)\phi(B)\nabla_S^D\nabla^d x(t) = \Theta(B^S)\theta(B)\varepsilon(t) \quad (1)$$

其中：

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p \quad (2)$$

$$\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q \quad (3)$$

$$\Phi(B^S) = 1 - \Phi_1 B^S - \dots - \Phi_p B^{pS} \quad (4)$$

$$\Theta(B^S) = 1 - \Theta_1 B^S - \dots - \Theta_q B^{qS} \quad (5)$$

$$\nabla_S^D\nabla^d x(t) = (1 - B^S)^D(1 - B)^d x(t) \quad (6)$$

$\Phi(B^S)\phi(B)$ 表示季节自回归模型，其中 $\phi(B)$ 表示 p 阶自回归多项式， $\phi_1, \phi_2, \dots, \phi_p$ 为非季节自回归参数。

$\Phi(B^S)$ 表示季节自回归多项式， $\Phi_1, \Phi_2, \dots, \Phi_p$ 为季节自回归参数。

$\Theta(B^S)\theta(B)$ 表示季节滑动平均模型，其中 $\theta(B)$ 表示 q 阶滑动平均多项式， $\theta_1, \theta_2, \dots, \theta_q$ 为非季节滑动平均参数。 $\Theta(B^S)$ 表示季节滑动平均多项式， $\Theta_1, \Theta_2, \dots, \Theta_p$ 为季节滑动平均参数。

$\nabla_S^D\nabla^d x(t)$ 为平稳时间序列。其中， B 表示滞后算子， ∇ 表示差分算子。

$\varepsilon(t)$ 表示服从独立正态分布 $N(0, \delta^2)$ 的高斯噪声。

EViews 7 软件被用于 MODIS LAI 时间序列的 SARIMA 建模。

1.2 BP 神经网络

BP 神经网络是一种按误差逆传播算法训练的多层前馈网络，它的主要特点是信号前向传输，误差被逆向传播。图

1 给出了三个输入节点的 BP 神经网络示意图，其中 x_1, x_2 和 x_3 为输入信息， y 为输出信息，在 BP 神经网络中，每层之间没有前馈循环，在同一层的每个节点之间也没有相互连接。输入的信息通过隐藏层后经输出层输出，如果输出结果同预期结果间的误差大于设定值，则根据误差调整传输路线上各节点的权值。神经网络相对其它非线性方法最重要的优势是神经网络能够全局近似，并且具有高度准确性。同时，神经网络的建模不需要事先假设，它很大程度上被数据的特

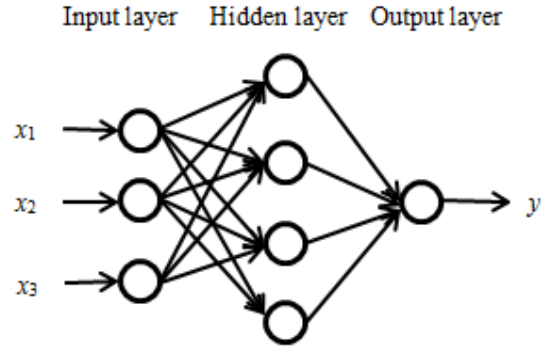


Fig.1 The structure of a three-layer back propagation neural network

性确定。BP 神经网络能描述复杂、非线性和不确定的系统^[14]。Hecht-Nielsen 证明了带有一个隐层的三层前馈神经网络能拟合任意多变量多项式函数^[15]。

三层前馈神经网络是应用最广泛的一种神经网络方法^[17]。它带有一个隐藏层、一个输出层和一个输入层。输入与输出间的关系可以通过下式表示：

$$x_t = w_0 + \sum_{j=1}^q w_j g(w_{0,j} + \sum_{i=1}^p w_{i,j} x_{t-i}) + e_t \quad (7)$$

$w_{i,j}$ ($i = 0,1,2, \dots, p, j = 1,2, \dots, q$) 和 w_j ($j = 0,1,2, \dots, q$) 表示连接权重；p 表示输入节点的个数，q 表示隐藏层节点的个数。

BP 神经网络被用于获取 MODIS LAI 时间序列中的非线性部分。为了用于比较，BP 神经网络也被单独作为一种方法，用于对 MODIS LAI 时间序列建模。BP 神经网络采用 MATLAB R2014b 软件实现。应用于本研究中的 BP 神经网络参数设置如表一：

Table.1 BP neural network parameter settings

参数	设置根据
隐藏层节点的数量	$l = \sqrt{m + n} + a, a \in [1,10]$ ^[17] l 是隐层节点的数量，m 是输入层节点的数量，n 是输出层节点的数量，a 为自适应变量。

归一化函数
$$u(z) = (h - g) \frac{z - X_{\min}}{X_{\max} - X_{\min}} + g$$

激活函数
$$f(\text{net}) = \frac{1}{1 + e^{-\text{net}}}$$
 [18]

学习速率 0.1 [19]

输入维度 3 [20]

1.3 组方法

神经网络既可以拟合线性部分，也可以拟合非线性部分，但对线性部分的拟合经常产生不稳定的结果 [21]。基于已有的相关研究 [10]，一个 MODIS LAI 时间序列可以看成由线性自相关结构和非线性部分组成。组方法可以被描述为：

$$y_t = f(L_t, N_t) \tag{8}$$

其中， L_t 表示 MODIS LAI 时间序列中的线性部分， N_t 表示 MODIS LAI 时间序列中的非线性部分。SARIMA 方法用于拟合线性部分 L_t ，BP 神经网络方法用于拟合非线性部分

N_t 。组方法分三步执行：

第一步：用 SARIMA 方法拟合时间序列并计算残差：

$$e_t = \hat{L}_t - y_t \tag{9}$$

\hat{L}_t 表示 SARIMA 方法的预测值。

第二步：使用 BP 神经网络方法对残差 e_t 时间序列建模，

然后将根据训练数据所建立模型用于对非线性残差部分进行预测。

第三步：将第一、二步结果之和作为最终预测结果。

1.4 预测质量验证

利用决定系数 R^2 、平均绝对误差 MAE (Mean Absolute Error) 和均方误差 (Mean Squared Error) 对方法的拟合及预测进行统计检验，并以此作为方法质量的评价指标。检验公式如下：

$$R^2 = 1 - \frac{\sum_{i=0}^n (y^i - \hat{y}^i)^2}{\sum_{i=0}^n (y^i - \bar{y})^2} \tag{10}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \tag{11}$$

$$MSE = \frac{1}{n} \sum_{i=1}^n w_i (y_i - \hat{y}_i)^2 \tag{12}$$

式中 y_i 表示 MODIS LAI 观测值， \hat{y}_i 表示 MODIS LAI 预测值， \bar{y} 表示 MODIS LAI 观测值的平均值， n 表示数据的数量。

2 实验结果与分析

2.1 实验数据

本研究数据采用了 8 天周期 1km 分辨率的 MODIS 数据产品 MOD15A2，时间范围为 2004-2013。研究点选在北纬 37°

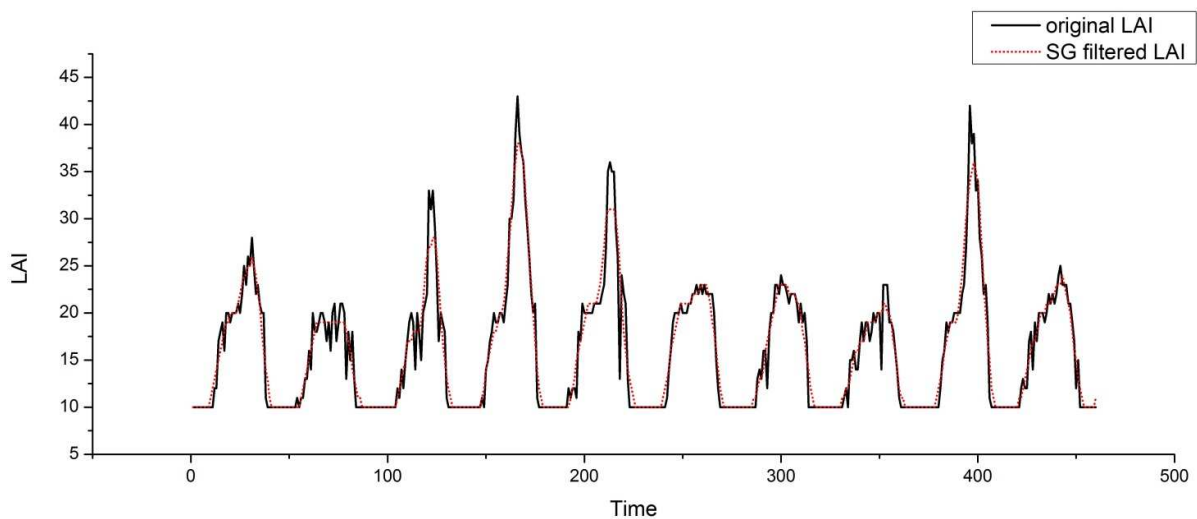


Fig.2 The comparison between SG filtered LAI and the original LAI

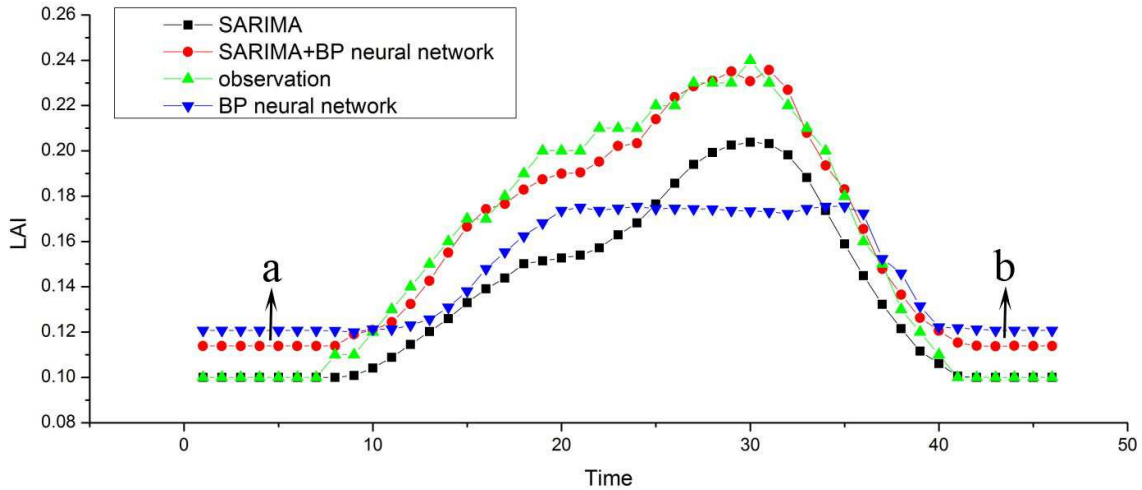


Fig.3 prediction curve of different method and observation

45' 东经 101° 22' 48"。依据新华安等^[3]及李喜佳等^[4]的描述，将研究点为中心的 10km×10km 范围内的 LAI 均值作为时间序列的值，在求取平均值时只对 MODIS LAI 数据中标记为高质量的值（即 MODLAND_QC=0）求取平均。

2.2 验证结果

图 2 对比了经过 SG 滤波和线性插值后的 MODIS LAI 时间序列和原始时间序列的差异。从图中可以看出，通过以上两步处理后的 LAI 相对原始 LAI 更为平滑，更符合植物生长的特征，这有利于建立更精确的模型和预测出更准确的结果。

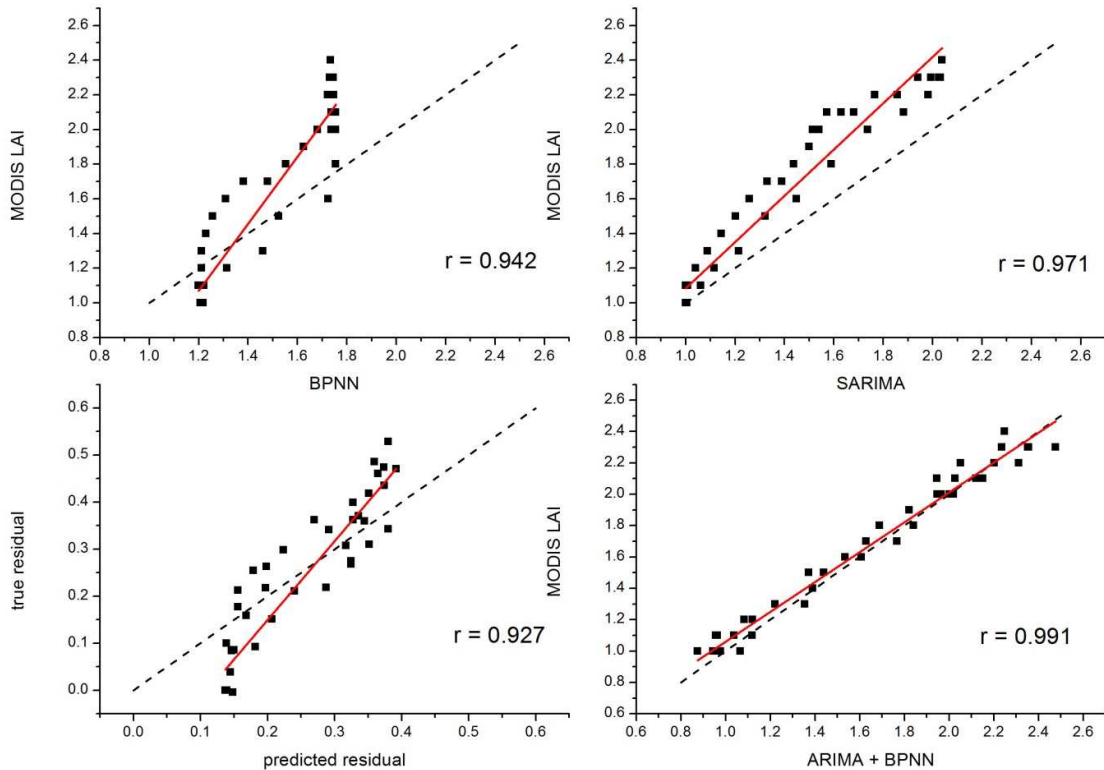


Fig.4 Relationship between the prediction values and the original values

SARIMA 方法采用 2004 年-2012 年 MODIS LAI 时间序列

建模，所建模型用于预测 2013 年 LAI。为了比较，BP 神经

网络也采用 2004 年-2012 年 MODIS LAI 时间序列建模, 并对 2013 年 LAI 进行预测。BP 神经网络也被用于对 SARIMA 模型残差序列建模, 并对 2013 年的 SARIMA 方法产生的残差进行预测。图 3 是 SARIMA、BP 神经网络和组合方法对 2013 年 LAI 的预测结果及 MODIS LAI 产品 2013 年的观测值。从图 3 可以看出, SARIMA 和 BP 神经网络的预测值总体上低于观测值, 其中 BP 神经网络在 LAI 值较大时的预测值明显低于观测值, 而在 LAI 观测值较小时 BP 神经网络的预测值相比 SARIMA 方法更接近观测值。以上结果说明构建的 BP 神经网络方法和 SARIMA 方法对 MODIS LAI 的预测具有不同的适应性, 即在 LAI 观测值较低时; 构建的 BP 神经网络模型预测准确度较高, 而在 LAI 观测值较高时, SARIMA 预测的结果更准确。这是由于 BP 神经网络在 MODIS LAI 值较大时, 不

能很好的拟合数据中的线性部分, 从而导致预测结果低于实际观测值; 而 SARIMA 方法由于包含了季节变化部分, 所以在整体趋势上拟合效果较好。SARIMA 同 BP 神经网络组合方法的预测准确率明显比单独的 SARIMA 方法和 BP 神经网络方法准确率高, 且结果更为光滑, 符合植物自然生长规律。从图 3 中可以看到另一个现象是组合方法在曲线起始和结束部分的预测值高于观测值(如图 3 中 a, b 所示), 造成这种现象的原因是由于 BP 神经网络基于全局近似, 而在图 3 中 a 段和 b 段部分的残差值比较小, 所以 BP 神经网络的预测值会偏高。但是, 在实际应用中, LAI 处于最低位较平坦时, 即位于 LAI 变化曲线起始或终端部分, 表征出了植物主要季相特征, 据此可方便地将高估部分识别出来而加以修正。

Table.2 Comparison of the performance of the proposed method with those of other method

MSE			MAE			R ²		
ARIMA	ANN	Hybrid	ARIMA	ANN	Hybrid	ARIMA	ANN	Hybrid
0.072	0.093	0.009	0.207	0.264	0.083	0.941	0.884	0.981

表二列出了用来衡量三种方法预测能力的 MSE、MAE 和 R²。组合方法的 MSE 和 MAE 值明显低于其它两种方法, 说明其预测误差较其它两种方法更小; 而组合方法的 R² 更接近于 1, 也说明该方法效果更好。组合方法的三种衡量指标均优于其它两种指标, 这是因为它充分利用了其它两种方法的优势, 即整合了它们各自对线性部分和非线性部分模式的识别能力。

对残差及预测值进行线性回归分析, 结果如图 4 所示, 相关系数为 0.927。对 SARIMA、BP 神经网络和组合方法的预测值同观测值进行线性回归分析, 结果如图 4 所示。SARIMA、BP 神经网络和组合方法的预测值同观测值之间的相关系数分别为 0.971、0.942 和 0.991, 进一步表明组合

方法的拟合度好于其它两种方法。

3 结论

SARIMA 和 BP 神经网络是广泛用于时间序列建模的两种方法。本文提出一种新的 MODIS LAI 建模方法: 分别通过使用 SARIMA 和 BP 神经网络方法来拟合 MODIS LAI 时间序列数据中的线性部分和非线性部分, 利用它们各自的优势对 MODIS LAI 历史时间序列数据进行建模及预测。试验结果表明, 结合 SARIMA 和 BP 神经网络所建立的模型预测准确率更高, 该方法的提出, 将为陆面过程模拟及遥感数据同化等应用提供更为可靠的遥感数据源。

References

1. Chen J M, Black T A. *Plant Cell and Environment*, 1992, 15(4): 421.
2. Xiao Z Q, Liang S L, Wang J D, et al. *Remote Sensing of Environment*, 2011, 115(1): 97.
3. Jin Hua-an, Wang Jin-di, Xiao Zhi-qiang, et al(靳华安, 王锦地, 肖志强, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2011, 31(9): 2485.
4. Li Xi-jia, Xiao Zhi-qiang, Wang Jin-di, et al(李喜佳, 肖志强, 王锦地, 等). *Journal of Remote Sensing(遥感学报)*, 2013, 18(1): 27.
5. Savoy P, Mackay D S. *Agricultural and Forest Meteorology*, 2015, 200: 46.
6. Dickinson R E, Tian Y H, Liu Q, et al. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 2008, 113(D16).
7. Jiang B, Liang S L, Wang J D, et al. *Remote Sensing of Environment*, 2010, 114(7): 1432.
8. Liu Q, Gu L H, Dickinson R E, et al., *Journal of Geophysical Research: Atmospheres (1984–2012)*, 2008, 113(D19).
9. Gao F, Morisette J T, Wolfe R E, et al. *Ieee Geoscience and Remote Sensing Letters*, 2008, 5(1): 60.
10. Zhang G P. *Neurocomputing*, 2003, 50: 159.

11. Cui Ri-xian, Liu Ya-dong, Fu Jin-dong, (崔日鲜, 刘亚东, 付金东). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2015. 9: 046.
12. Zhou Chun-yan, Hua Deng-xin, Le Jing, et al(周春艳, 华灯鑫, 乐静). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2015, 9: 053.
13. Box G E, Jenkins G M, Reinsel G C. Time series analysis: forecasting and control. John Wiley & Sons, 2011.
14. Cheng C, Cheng X S, Dai N, et al, Computers in biology and medicine, 2015, 66: 103.
15. Hecht-Nielsen R. Theory of the backpropagation neural network. In Neural Networks, 1989. IJCNN., International Joint Conference on. IEEE, 1989.
16. Haykin S, Multilayer perceptrons. Neural Networks: A Comprehensive Foundation, 1999. 2: 156.
17. Liu Y K, Xie F, Xie C L, et al. Annals of Nuclear Energy, 2015, 85: 566.
18. Chiang W K, Zhang D S, Zhou L N. Decision Support Systems, 2006, 41(2): 514.
19. Fu L M. Neural networks in computer intelligence. Tata: McGraw-Hill Education, 2003.
20. Khashei M, Bijari M. Applied Soft Computing, 2011, 11(2): 2664.
21. Faruk D Ö. Engineering Applications of Artificial Intelligence, 2010, 23(4): 586.

Modeling and predicting of MODIS LAI time series based on a hybrid SARIMA and BP neural network method

JIANG Chun-lei^{1, 2}, ZHANG Shu-qing^{1*}, ZHANG Ce³, LI Hua-peng¹, DING Xiao-hui^{1, 2}

1. Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130102, China

2. University of Chinese Academy of Sciences, Beijing 100049, China

3. Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK

Abstract: The modeling and predicting of vegetation Leaf area index (LAI) is an extremely important indication factor for growth status of vegetation. However, the time series of MODIS LAI include linear and nonlinear components, which cannot be accurately modeled and predicted by either linear method or nonlinear method. In this paper, SARIMA, BP neural network and a hybrid method of SARIMA-BP neural network were used for modeling and predicting MODIS LAI time series. The SARIMA-BP neural network combined both SARIMA and BP neural network, of which could model the linear component and the nonlinear component of MODIS LAI time series respectively. Thus the final result of SARIMA-BP neural network was the sum of results of the two methods. Experiments showed that the proposed SARIMA-BP neural network method performed the best in comparison with SARIMA and BP neural network, implying its well adaption to the LAI time series.

Keywords: SARIMA; BP neural network; MODIS LAI; SARIMA-BP neural network method; LAI time series modeling and predicting

*Corresponding author