中国农业大学学报 2012,17(4):148-153
Journal of China Agricultural University

基于 PCA-RBF 神经网络的森林碳储量遥感反演模型研究

张 超 彭道黎*
(北京林业大学 林学院,北京 100083)

摘 要 针对碳储量回归预测模型存在共线性和精度较低的问题,利用森林资源二类调查数据和 SPOT5 影像数据对北京市延庆县的杨树林进行碳储量反演研究。先对选取的 10 个指标进行主成分分析,在此基础上采用径向基函数(RBF)神经网络方法构建碳储量反演模型,用预留测试样本验证,并与实测值进行比较。研究结果表明:SPOT5 数据和二类数据可以很好地结合起来用于森林地上碳储量反演研究;PCA-RBF 神经网络森林碳储量遥感反演模型拟合精度为 99.90%,平均预测精度达到 96.71%,预估效果较理想;模型训练完成后,可以应用于延庆县森林地上碳储量反演。

关键词 森林碳储量; SPOT5; 主成分分析; 遥感反演; RBF 神经网络
中图分类号 S 771.8; S 779 **文章编号** 1007-4333(2012)04-0148-06 **文献标志码** A

Remote sensing retrieval model of forest carbon storage based on principal components analysis and radial basis function neural network

ZHANG Chao, PENG Dao-li*
(College of Forestry, Beijing Forestry University, Beijing 100083, China)

Abstract Aiming at the problem of multicollinearity and low precision predictions by the regression prediction model of carbon storage, this study used forest resource inventory data and SPOT5 image to retrieve the aboveground forest carbon storage of *Populus* forests in Yanqing County. Firstly, 10 factors were analyzed by principal components analysis. Then this paper introduced a method based on PCA and radial basis function (RBF) neural network for predicting forest carbon storage. The research results show that forest resource inventory data combined SPOT5 image is very useful for retrieving study of carbon storage of *Populus* forests; the fitting precision of the PCA-RBF neural network model was 99.90%, and the average prediction reached 96.71%. The model has a good retrieval accuracy, which can be well used for retrieval of regional aboveground forest carbon storage.

Key words forest carbon storage; SPOT5; principal component analysis; remote sensing retrieval; RBF neural network

21 世纪,人类社会面临着最严峻的气候变化挑战。自 1750 年以来,由于人类活动已引起全球大气中 CO_2 浓度明显增加,2005 年大气中 CO_2 和 CH_4 的浓度已远远超过了过去 65 万年的自然变化的范围[1]。因此,碳汇、碳循环、碳平衡等相关研究也成为全球气候变化研究的核心问题之一,开始受到世界各国的重视。作为陆地生态系统的主体,森林拥有最高的生物量,也是陆地生态系统的最大碳

收稿日期: 2012-03-07
基金项目: 国家"十一五"科技支撑计划(2006BAD23B05); 国家级林业推广项目(201145)
第一作者: 张超,硕士研究生,E-mail:izhangcici@qq.com
通讯作者: 彭道黎,教授,主要从事森林资源监测与评价研究,E-mail:dlpeng@bjfu.edu.cn

库，在调控全球碳循环和碳平衡中起着不可替代的作用。森林生物量是维持森林生态系统正常运行的能量和营养物质基础，是反映森林生态环境和评价森林生态系统生产力的重要指标，也是评估全球碳平衡的基础，及时、准确地获取森林生物量分布状况及发展趋势是深入了解生态系统变化规律的重要途径，准确估算森林生态系统的生物量对进一步研究陆地生态系统碳循环中的不确定性具有重要意义[2]。

传统的森林碳储量估算方法费时费力，仅适合小尺度的碳储量估测，结果不能大范围推广。近年来，遥感数据（包括多光谱数据、高光谱数据、微波遥感和雷达数据等）已经成为碳储量估算的主要数据来源[3-6]，其中，使用最广泛的为多光谱数据。综合应用 3S 技术和地面调查数据建立模型是定量评价森林碳储量的重要手段和趋势。现有的碳储量遥感模型常用建模方法可归纳为回归模型（线性回归、曲线估计）和神经网络模型两大类[7-10]，前者方法简单，但变量之间可能存在共线性问题，后者应用相对复杂，但适合于非线性模型构建，而 RBF 神经网络以其计算量小，学习速度快，不易陷入局部极小等诸多优点为建模提供了一种有效的手段。

本研究结合 SPOT5 数据和森林资源二类调查数据在主成分分析的基础上，利用 RBF 神经网络对北京延庆县杨树林地上部分碳储量进行定量反演研究，建立 PCA-RBF 碳储量遥感反演模型，并与实测值进行精度对比分析，验证 PCA-RBF 神经网络用于森林碳储量估测的可行性，以期为更准确快速获取区域森林地上碳储量信息提供新的思路。

1　研究区概况及数据获取

1.1　研究区概况

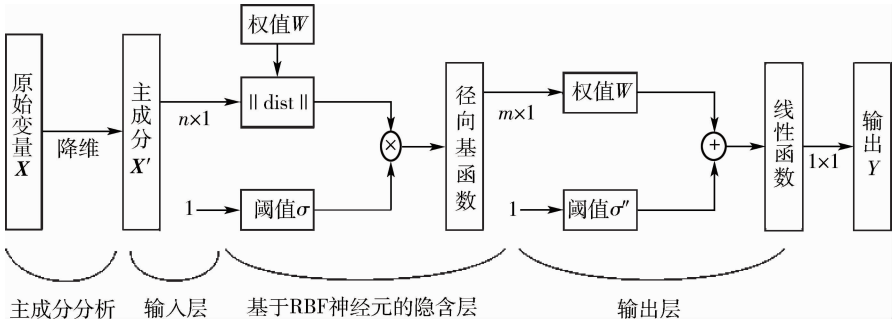
研究区域位于北京市西北部的延庆县，距北京市区 74 km，地理坐标为北纬 40°16′～40°47′，东经 115°44′～116°34′。该区域三面环山，一面临水，东邻怀柔，南连昌平，西面和北面与河北怀来县、赤城县接壤，西南是官厅水库。川区海拔 500～600 m，山区海拔 600～2 241 m，整个地势自东北向西南倾斜，北部群山因造山运动褶皱凸起，垂直起伏明显。山区由于近代侵蚀剧烈，形成沟壑纵横、滩涂交错的地貌特征。该县属于大陆性季风气候，是暖温带与中温带、半湿润与半干旱气候的过渡带，年平均气温 8.8 ℃，年平均降水量为 493 mm。原始植被类型为暖温带落叶阔叶林和温带针叶林，由于早期人为破坏，现已不多见。盆地、河川、沟谷部分台地植被为栽培植物，其余山区大部分为针叶林、阔叶林、针阔混交林、杂灌及草本群落。现今延庆地区主要分布的树种为侧柏（*Platycladus orientalis* （Linn.） *Franco*）、刺槐（*Robinia pseudoacacia* L.）、油松（*Pinus tabuliformis*）、柞树（*Xylosma racemosum*）、杨树（*Populus* spp.）等。

1.2　数据获取及处理

选取 SPOT5 影像的 4 个波段（波段 1，2，3，4）、差值植被指数（DVI）、比值植被指数（RVI）、归一化植被指数（NDVI）及坡度、坡向、海拔共 10 个指标作为碳储量模型构建的原始变量。获取覆盖延庆县的 2004-05-23 的 SPOT5 数据，图像清晰，成像质量较好，无阴云。太阳高度角为 64.4°，太阳方位角为 135.4°。采用 1∶5万地形图对影像进行几何校正，像元均方根误差为 0.5，满足精度要求。利用延庆县行政边界图对校正好的 SPOT5 影像进行掩膜处理得到研究区域影像。将 GPS 定位样点与 DEM、SPOT5 多光谱影像及三种植被指数影像进行叠加，提取各指标的样点值，其中地形因子包括坡度、坡向、海拔，光谱因子包括 DVI、RVI、NDVI 及 4 个原始波段值。收集延庆县 2005 年森林资源二类调查样地数据进行分析，随机选取 100 组数据作为样本数据，将前 80 组数据作为神经网络模型的训练样本，保留后 20 组数据作为测试样本，利用 Matlab 软件实现数据仿真。

2　PCA-RBF 神经网络模型原理

PCA-RBF 神经网络模型是主成分分析和径向基函数神经网络相结合的一种数据融合模型[11]（图 1），通过主成分分析对原始变量 $\boldsymbol{X}$ 进行预处理，从而得到主成分变量 $\boldsymbol{X}'$（作为 RBF 神经网络的输入层），然后利用 RBF 神经网络来对样本进行训练和仿真分析。



n 为输入矢量的维数；m 为隐含层神经元的个数；‖ dist ‖ 为求取输入矢量和权值的距离。

图 1 PCA-RBF 神经网络模型

Fig. 1 PCA-RBF neural network model

2.1 主成分分析

主成分分析也称主分量分析或主元分析[12]，是利用“降维”的思想，在损失很少信息的前提下把多个指标转化为几个综合指标的多元统计方法，通常把转化生成的综合指标称为主成分，各个主成分包含原始变量的主要信息，各主成分之间互不相关。

设有 n 个样本，每个样本有 m 个指标，得到原始样本矩阵 $\boldsymbol{X}=(x_{aj})_{n\times m}=(\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_j)$，其中 $\boldsymbol{X}_j=(x_{1j},x_{2j},\cdots,x_{nj})'$，$j=1,2,\cdots,m$。为避免原始变量量纲的影响，先对原始数据 $\boldsymbol{X}$ 进行标准化处理，即

$$\boldsymbol{X}_z=[\boldsymbol{X}-(1,1,\cdots,1)'\boldsymbol{M}]\cdot \operatorname{diag}(s_1^{-1},s_2^{-1},\cdots,s_m^{-1})$$

其中 $\boldsymbol{M}=(m_1,m_2,\cdots,m_m)$ 为变量 $\boldsymbol{X}$ 的均值，$\boldsymbol{S}=(s_1,s_2,\cdots,s_m)$ 为变量 $\boldsymbol{X}$ 的标准差。对 $\boldsymbol{X}_z$ 进行主成分分析，可以形成新的综合变量，用 $\boldsymbol{X}'$ 表示。计算矩阵 $\boldsymbol{X}_z$ 的相关系数矩阵 $\boldsymbol{R}$。

$$\boldsymbol{R}=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix}$$

求 $\boldsymbol{R}$ 的特征根 $\lambda_1\geqslant\lambda_p\geqslant 0$ 和对应的正则化特征向量 $\boldsymbol{a}_\beta(\beta=1,2,\cdots,p)$，得到主成分 $\boldsymbol{X}'_\beta=\boldsymbol{X}_z\boldsymbol{a}_\beta$。通常当前 k 个主成分累积方差贡献率达到 85% 以上时，即认为这 k 个主成分代表了原来 m 个指标的主要信息，主成分分析结束，将上述得到的 k 个主成分作为 RBF 神经网络的输入向量。

2.2 RBF 神经网络

径向基函数神经网络的基本构成包括输入层、隐含层和输出层，具有对非线性函数最佳逼近和全局最优的性能[12]。从输入层空间到隐含层空间的变化是非线性的，而从隐含层空间到输出层空间变换是线性的。RBF 网络不仅具有良好的泛化能力，而且计算量小，学习速度也比其他一般算法快，在一定程度上避免了 BP 神经网络中学习算法冗长的迭代计算过程和陷入局部极值的可能，在气象[13]、土壤[14]、植被[15]、工程控制[16]等众多领域得到广泛应用。

RBF 神经网络的输入层为主成分分析后得到的 k 维向量 $\boldsymbol{X}'=\{\boldsymbol{X}'_1,\boldsymbol{X}'_2,\cdots,\boldsymbol{X}'_k\}$。隐含层为 l 维向量 $\boldsymbol{D}=\{\boldsymbol{D}_1,\boldsymbol{D}_2,\cdots,\boldsymbol{D}_l\}$，隐含层节点个数一般通过不断试验来确定，直到达到误差满意为止。网络的输出为一维向量 $f(\boldsymbol{X}')$，其输出为

$$f(\boldsymbol{X}')=\sum_{i=1}^{l}W_{iu}D_i(\boldsymbol{X}')$$

式中：$D(\boldsymbol{X}')$ 为隐含层中的径向基函数；i 为隐含层的节点数目，$i=1,2,\cdots,l$；u 为输出神经元的个数，$u=1,2,\cdots,h$；W_{iu} 为第 i 个隐含层单元到输出单元的权值。隐含层作用函数采用径向基函数，一般选取高斯函数

$$D_i(\boldsymbol{X}')=\exp\left[-\frac{\|\boldsymbol{X}'-c_i\|^2}{2\sigma_i^2}\right]$$

式中：$\|\boldsymbol{X}'-c_i\|$ 为欧氏范数；c_i 为网络隐含层节点的中心；σ_i 为径向基函数的方差或宽度，用来调节网络的灵敏度。

RBF 神经网络学习算法需要求解的参数有 3 个[17]：基函数的中心 c_i、方差 σ_i、隐含层到输出层的权值 W_{iu}。

1)采用基于 K-均值聚类方法求取基函数的中心 c_i，随机选取 l 个训练样本作为聚类中心 c_i；将输入的训练样本集合按照最近邻规则分组：按照 $\boldsymbol{X}'$ 与中心为 c_i 之间的欧氏距离将 $\boldsymbol{X}'$ 分配到输入样本的各个聚类集合 $\boldsymbol{A}_b(b=1,2,\cdots,p)$ 中；重新调整聚类中心：计算各个聚类集合 $\boldsymbol{A}_b$ 中训练样本的平均值，即新的聚类中心 c_i，如果新的聚类中心不再发生变化，则所得到的 c_i 即为 RBF 神经网络最终的基函数中心。

2)该 RBF 神经网络的基函数为高斯函数，其方差 σ_i 求解公式为

$$\sigma_i = \frac{c_{\max}}{\sqrt{2l}}$$

3)利用最小二乘法计算得到隐含层到输出层的权值 W_{iu}

$$W_{iu} = \exp\left[\frac{l}{c_{\max}^2}\| \boldsymbol{X}' - c_i \|^2\right]$$

3　PCA-RBF 神经网络建模及结果分析

3.1　主成分分析建模结果

为避免原始指标量纲的影响，先将 80 组原始数据的 10 个指标标准化处理，再进行主成分分析，整个过程利用 SPSS 统计软件实现。从表 1 可以看出，对原始数据进行主成分分析后，前 4 个主成分的样本方差累积贡献率已高达 91.27%＞85.00%，几乎涵盖了原始变量的主要信息，符合主成分提取要求，所以选取前 4 个主成分作为森林碳储量建模的输入变量。

表 1　主成分特征值及其贡献率

Table 1　Eigenvalue of each principal component and its contribution rate

参 数	主成分 1	主成分 2	主成分 3	主成分 4
特征值	5.798	1.486	1.124	0.719
贡献率/%	57.98	14.86	11.24	7.19
累积贡献率/%	57.98	72.84	84.08	91.27

表 2 示出原始样本数据经过主成分分析后得到的主成分系数，据此可写出主成分分析得到的 4 个主成分的模型。

表 2　光谱及地形因子的主成分得分系数

Table 2　Principal component score coefficients of spectral and topographical factors

变 量	主成分 1	主成分 2	主成分 3	主成分 4
波段 1	0.063	0.469	0.242	−0.638
波段 2	−0.168	0.079	0.018	−0.060
波段 3	−0.161	0.054	0.070	−0.144
波段 4	−0.149	0.249	−0.066	−0.047
DVI	0.161	−0.124	0.101	−0.135
RVI	0.169	0.029	0.020	−0.135
NDVI	0.170	−0.018	0.030	−0.108
坡向	0.025	0.230	0.668	0.758
坡度	0.074	0.226	−0.558	0.579
海拔	0.050	0.511	−0.229	0.007

注：DVI 为差值植被指数，RVI 为比值植被指数，NDVI 为归一化植被指数。

$$\boldsymbol{X}_1' = 0.063x_1 - 0.168x_2 - 0.161x_3 - 0.149x_4 + 0.161x_5 + 0.169x_6 + 0.170x_7 + 0.025x_8 + 0.074x_9 + 0.050x_{10}$$

$$\boldsymbol{X}_2' = 0.469x_1 + 0.079x_2 + 0.054x_3 + 0.249x_4 - 0.124x_5 + 0.029x_6 - 0.018x_7 + 0.230x_8 + 0.226x_9 + 0.511x_{10}$$

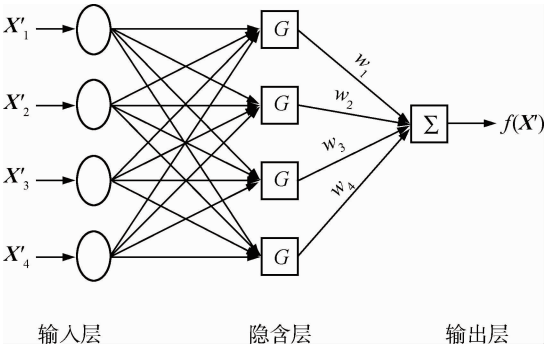
$$\boldsymbol{X}_3' = 0.242x_1 + 0.018x_2 + 0.070x_3 - 0.066x_4 + 0.101x_5 + 0.020x_6 + 0.030x_7 + 0.668x_8 - 0.558x_9 - 0.229x_{10}$$

$$\boldsymbol{X}_4' = -0.638x_1 - 0.060x_2 - 0.144x_3 - 0.047x_4 - 0.135x_5 - 0.135x_6 - 0.108x_7 + 0.758x_8 + 0.579x_9 + 0.007x_{10}$$

3.2 PCA-RBF 神经网络碳储量模型反演结果

3.2.1 网络设计

将主成分分析得到的4个新指标 X'_1、X'_2、X'_3 和 X'_4 作为网络输入样本，用矢量 $X'=\{X'_1, X'_2, X'_3, X'_4\}$ 表示。将森林碳储量实测值 $f(X')$ 作为网络输出值，可以建立输入层为4个神经元，输出层为1个神经元的RBF神经网络(图2)。



X'_1、X'_2、X'_3、X'_4 为主成分分析得到的4个新指标；G 为隐含层中的径向基函数；w_1、w_2、w_3、w_4 为第1、2、3、4个隐含层单元到输出层单元的权值；Σ 为输出层函数；$f(X')$ 为碳储量值。

图2 RBF神经网络设计结构

Fig. 2 Structure of RBF neural network

3.2.2 网络训练

将主成分分析后得到的80组主成分数据作为训练样本，先将样本数据进行归一化处理，在MATLAB的神经网络工具箱中用newrb函数创建这个径向基函数网络，newrb函数的调用方式为

net=newrb(P,T,GOAL,SPREAD,MN,DF)

其中：net为径向基函数网络对象；newrb为径向基函数；P 为网络输入样本向量矩阵；T 为输出目标向量矩阵；GOAL为网络均方误差目标值；SPREAD为径向基函数的分布系数；MN为神经元的最大个数；DF为2次显示之间所添加的神经元数目。

该函数设计的径向基函数网络net用作函数逼近，可以自动增加径向基RBF网络的隐含层神经元，直到均方误差在期望误差之下或者网络达到最大神经元数目为止。用newrb函数设计RBF神经网络是一个不断尝试的过程，在网络设计过程中，需要用不同的SPREAD值进行尝试，以确定一个最优值[12]。

网络训练时，设置GOAL为0.001，创建SPREAD为0.1、0.2、0.3、0.4、0.5的5个RBF神经网络，通过与真实值的误差分析对比选择1个最优值。经过试验，当SPREAD取0.5时，RBF网络的误差满足精度要求，逼近效果最好。

3.2.3 模型反演验证

为了验证建立的PCA-RBF神经网络模型是否具有实际意义，用预留的20组数据测试训练完成后的PCA-RBF神经网络模型。将20组测试样本原始变量主成分分析后的数据输入至模型中，利用建立好的神经网络模型进行预测，并与实测值进行比较分析，用以检验PCA-RBF神经网络碳储量预测的准确性和稳定性(图3)。

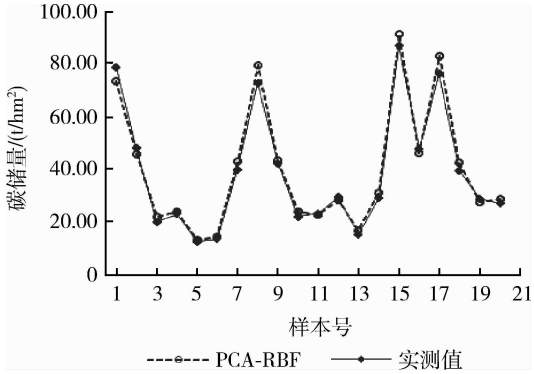


图3 PCA-RBF神经网络森林碳储量模型反演结果

Fig. 3 PCA-RBF neural network simulation results for forest carbon storage

4 结 论

本研究以北京市延庆县为例，通过融合主成分分析和最近邻聚类算法的RBF神经网络方法，构建了基于PCA-RBF神经网络的森林地上碳储量遥感估算模型，从而可以实现该区域的森林地上碳储量反演。主要结论如下：

1)PCA-RBF模型拟合精度为99.90%，用预留测试样本对模型进行反演验证，估测精度达到96.71%，预测效果较好。利用主成分分析方法，既可消除原始自变量的共线性，保留原始变量的主要信息，又能降低神经网络输入层的维数，缩小神经网络规模，简化模型，提高分析效率，是构建碳储量估测模型的一种有效方法。该PCA-RBF神经网络碳储量反演模型的建立为延庆县森林碳储量预测提供了重要的参考依据，对区域尺度的森林碳储量模型建立

有较好的借鉴作用。

2)将 SPOT5 影像数据与森林资源二类调查数据结合起来估算区域尺度的森林地上碳储量，获得了较好的预测结果，进一步验证了遥感数据结合地面数据估测森林碳储量是一种合理有效的方法。

参 考 文 献

[1] IPCC. 气候变化 2007：综合报告[R]. 日内瓦：政府间气候变化专门委员会，2008：1-104

[2] 李海奎，雷渊才. 中国森林植被生物量和碳储量评估[M]. 北京：中国林业出版社，2010

[3] Foody G M, Boyd D S, Cutler M E. Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions [J]. Remote Sensing of Environment, 2003, 85(4): 463-474

[4] Lim K, Treitz P, Baldwin K, et al. Lidar remote sensing of biophysical properties of tolerant northern hardwood forests [J]. Canadian Journal of Remote Sensing, 2003, 29(5): 658-678

[5] Smith B, Knorr W, Widlowski J L, et al. Combining remote sensing data with process modelling to monitor boreal conifer forest carbon balances[J]. Forest Ecology and Management, 2008, 255(12): 3985-3994

[6] 刘占宇，黄敬峰，吴新宏，等. 草地生物量的高光谱遥感估算模型[J]. 农业工程学报，2006，22(2)：111-115

[7] 徐天蜀，张王菲，岳彩荣. 基于 PCA 的森林生物量遥感信息模型研究[J]. 生态环境，2007，16(6)：1759-1762

[8] 汪少华，张茂震，赵平安，等. 基于 TM 影像、森林资源清查数据和人工神经网络的森林碳空间分布模拟[J]. 生态学报，2011，31(4)：998-1007

[9] 王淑君，管东生. 神经网络模型森林生物量遥感估测方法的研究[J]. 生态环境，2007，16(1)：108-111

[10] 王立海，刑雁秋. 基于人工神经网络的天然林生物量遥感估测[J]. 应用生态学报，2008，19(2)：61-266

[11] 关子明，常文兵. 基于 PCA-RBF 神经网络模型的航空备件预测方法[J]. 北京工商大学学报：自然科学版，2009，27(3)：60-64

[12] 何晓群. 多元统计分析[M]. 北京：中国人民大学出版社，2004

[13] 张德丰. MATLAB 神经网络应用设计[M]. 北京：机械工业出版社，2009

[14] 农吉夫，金龙. 基于 MATLAB 的主成分 RBF 神经网络降水预报模型[J]. 热带气象学报，2008，24(6)：713-717

[15] 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

[16] 李爽，张祖陆，周德民. 湿地植被地上生物量遥感估算模型研究：以洪河湿地自然保护区为例[J]. 地理研究，2011，30(2)：278-290

[17] 赵望达，徐志胜，吴敏. 基于主元分析和 RBF 神经网络的火灾模拟实验炉温软测量[J]. 中国工程科学，2007，9(1)：82-85

责任编辑：刘迎春