

郭品文,徐同,居丽丽.近 53 a 中国气温异常分布的非线性特征[J].南京气象学院学报,2009,32(1):11-16.

近 53 a 中国气温异常分布的非线性特征

郭品文¹,徐同²,居丽丽³

(1.南京信息工程大学气象灾害省部共建教育部重点实验室,江苏南京 210044;
2.中国气象局上海台风研究所,上海 200030; 3.南汇区气象局,上海 201300)

摘要:运用一种基于神经网络的非线性主成分分析法(nonlinear principal component analysis,NLPCA)对中国 1951—2003 年 53 a 四季气温距平场(surface air temperature anomaly,SATA)进行分析,NLPCA 第一模态结果显示中国四季气温异常具有一定的非线性特征,并且具有显著的季节性差异,即春、夏两季的非线性较强,秋、冬两季较弱。一维 NLPCA 对原始气温距平场的近似比一维 PCA(principal component analysis)更好地反映了气温场的实际分布情况。

关键词:气温;非线性;主成分分析;非线性主成分分析

中图分类号:P732 **文献标识码:**A **文章编号:**1000-2022(2009)01-0011-06

Nonlinear Characteristics of Surface Air Temperature Anomalies over China in Recent 53 Years

GUO Pin-wen¹, XU Tong², JU Li-li³

(1. Key Laboratory of Meteorological Disaster of Ministry of Education, NUIST, Nanjing 210044, China;
2. Shanghai Typhoon Institute, China Meteorological Administration, Shanghai 200030, China;
3. Nanhui Meteorological Bureau, Shanghai 201300, China)

Abstract: Seasonal surface air temperature(SAT) anomalies over China from 1951 to 2003 is investigated by applying a neural-network-based nonlinear principal component analysis(NLPCA) method. The results of the first NLPCA mode show that the seasonal SAT anomalies have some nonlinear characters; its nonlinearity is stronger in spring(MAM) and summer(JJA), and weaker in autumn(SON) and winter(DJF). The SAT approximation by 1-D NLPCA is closer to the observations than that by 1-D PCA.

Key words: temperature; nonlinearity; principal component analysis; NLPCA

0 引言

自 20 世纪 80 年代后期以来,人工神经网络的理论和应用研究都有了较大的发展,目前在国内外大气科学领域中有关神经网络方法的应用研究内容涵盖了气象学科的很多方面,其中主要包括中短期天气预报、短期气候预测、气候模式、温度时间序列预报等。

近年来,国外的气候学者将神经网络应用到非线性主成分分析(nonlinear principal component analysis,NLPCA)中。Kramer^[1]指出,主成分分析

(principal component analysis,PCA)仅能识别变量间的线性特征,而 NLPCA 则能同时反映变量间的线性和非线性相关,而不仅限于识别数据集中的非线性特征。Monahan^[2]将 NLPCA 应用到 Lorenz 数据集中,结果表明 NLPCA 能得到比 PCA 更好的一维和二维近似。NLPCA 一维和二维近似分别解释了 Lorenz 数据系统数值积分数据集总方差的 76% 和 99.5%,而 PCA 仅分别解释 60% 和 95%。Monahan^[3]运用 NLPCA 分析了热带太平洋海表温度和海平面气压,结果表明热带太平洋海表温度的 NLPCA 一维近似要优于 PCA 一维近似。海表温度

收稿日期:2007-04-13;改回日期:2007-07-09

基金项目:江苏省“六大人才高峰”项目

作者简介:郭品文(1962—),男,江苏苏州人,博士,教授,博士生导师,研究方向为短期气候预测,guo@nuist.edu.cn.

场存在着较显著的非线性,而海平面气压场则不明显。

国内方面,金龙等^[4-7]、林建玲等^[8]、李永华等^[9]对神经网络方法在大气科学领域中的应用做了总结,并将神经网络模型应用到数值预报和短期气候预测中。同时,神经网络方法也被应用到降水、温度的气象要素预报中^[10-12]。韦莹莹等^[13]在国内最早采用 NLPCA 分析了北半球冬季大气环流遥相关型的非线性特征,发现北半球冬季大气环流的遥相关型有明显的非线性结构特征,特别是 EU 型和 PNA 型,当非线性主成分取正负极值时,EU 和 PNA 型并不是完全的呈现反位相结构。本文采用 NLPCA 非线性统计模式对中国四季气温场异常变率进行研究,探讨了中国四季气温变化的非线性特征。

1 资料和方法

1.1 资料来源及预处理

本文所用的全国 160 站逐月气温资料来自国家气候中心资料集,资料长度为 1951 年 1 月—2003 年 12 月共 53 a。资料的处理方法为按春(3—5 月)、夏(6—8 月)、秋(9—11 月)、冬(12 月—次年 1 月)四季,求出平均气温场。

1.2 非线性主成分分析法(NLPCA)

非线性主成分分析法(NLPCA)是 Monahan^[2]于 2000 年首先提出的,这是一种用人工神经网络提取多变量资料集的主要特征的方法,该方法由 5 层前向型 BP 神经网络^[14]实现(图 1),网络的非线性最优化由 MATLAB 自带函数“fminu”完成。图 1 是由 5 个相互平行的层组成,依次为输入层(包含 1 个神经元)、编码层(包含 m 个神经元)、瓶颈层(包含 1 个神经元)、译码层和输出层,数据压缩由瓶颈层实现,用 U 表示瓶颈层神经元即非线性主元。输

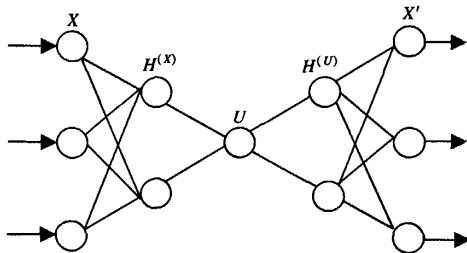


图 1 实现非线性主成分分析的神经网络示意

Fig. 1 A schematic diagram of the neural network model for calculating NLPCA

入资料集 $X(t_n)$ 的近似 $X'(t_n) = f\{S_f[X(t_n)]\}$, 其中,压缩变换 S_f 和展开变换 f 是非线性的双曲正切函数,即允许原始数据到特征空间的映射是非线性的。文中网络迭代次数为 50,随机选取权重和偏移系数。同时随机选取 20% 的数据用于网络测试,余下的用于网络训练。当测试数据的均方误差比训练数据大时,网络停止运行以避免得到过适应的解。从 0~0.18 每隔 0.006 取一惩罚系数值,隐层神经元重复取 2,3 进行运行,选择均方误差(即整个资料集上未被解释的方差百分比)最小的解为所用的解。

2 EOF 分析

在 NLPCA 分析之前,先对四季的温度距平场进行 EOF 分析(即 PCA 分析),这样做利用了 EOF 在数据压缩降维方面的优势,EOF 将气象要素场分解为特征向量和时间系数两部分。特征向量反映场的空间分布特点,它不随时间变化,时间系数描述场的时间变化规律,EOF 可将要素场的主要特征集中在前几个特征向量和时间系数中。表 1 显示了前 4 个主模态的方差贡献及累积方差贡献,各季的 EOF 第一模态均有较大的解释方差,从 29%~59%,冬季最大,夏季最小,前 4 个特征向量累积方差贡献最小也达到了 62.8%,最大达到了 80.3%,这反映出我国冬季气温的影响因子单一且作用强烈,夏季则受多种因素的影响。

表 1 1951 年 1 月—2003 年 12 月四季气温距平 EOF 前 4 个模态的方差贡献

Table 1 Variance contributions of the first four EOF modes of seasonal SAT anomalies

from January 1951 to December 2003					%
季节	模态 1	模态 2	模态 3	模态 4	合计
春季	42.7	16.9	8.3	5.8	73.7
夏季	28.8	16.2	12.6	5.2	62.8
秋季	47.9	9.7	6.8	4.8	69.2
冬季	58.5	12.2	6.0	3.6	80.3

3 中国四季气温距平的 NLPCA 分析

3.1 四季 NLPCA 三维空间图及非线性强弱特征

EOF 分析通常用于提取资料中的最优线性特征,如果资料的主要结构原本就是非线性的,则 EOF 就只能获得次优的低维特征,神经网络的非线性映射方法可以较好地解决这个问题。利用图 1

所示的神经网络能实现该方法,文中所用的 NLPCA 有 4 列输入和输出, m 取 2 和 3,为了防止网络过适应,在目标函数中加入权重系数 P 。

将前 4 个主成分 (principal component, PC) 作为非线性主成分分析的对象,得到四季气温距平 (surface air temperature anomaly, SATA) 的 NLPCA 第一模态,为了更直观地了解和分析非线性特征,采用三维空间图 (图 2)。图 2 中,不仅能看到在 PC1-PC2-PC3 三维空间中的 NLPCA 模态近似曲线,该曲线由一系列重叠的空心方框构成 (PCA 近似模态表现为一直线),同时能看到其在 PC1-PC2、PC1-PC3 和 PC2-PC3 平面上的投影。

图 2a 显示的春季的情况,可以在 PC1-PC2 平

面上看到一条明显的曲线 (PCA 表现为直线),表现出较强的非线性。夏季 (图 2b) 与春季较为相像,秋、冬两季则不明显。衡量非线性强弱的重要指标均方误差 (mean square error, MSE) 的比率 (MSE 是指一数据点与它的 PCA 近似模态投影或 NLPCA 近似模态投影之间距离的均方差,它反映了近似的误差程度),也理解为 NLPCA 第 1 模态未解释的方差百分比与 PCA 第 1 模态未解释的方差百分比的比值,该比值越小,说明非线性越强,反之若越接近 1,则非线性越弱。春、夏、秋、冬四季的均方误差比率分别为 0.738、0.798、0.969、0.934。由此可见,春、夏两季的非线性较强,而秋、冬两季较弱。与 PCA 第一模态提取的特征相比, NLPCA 第一模态

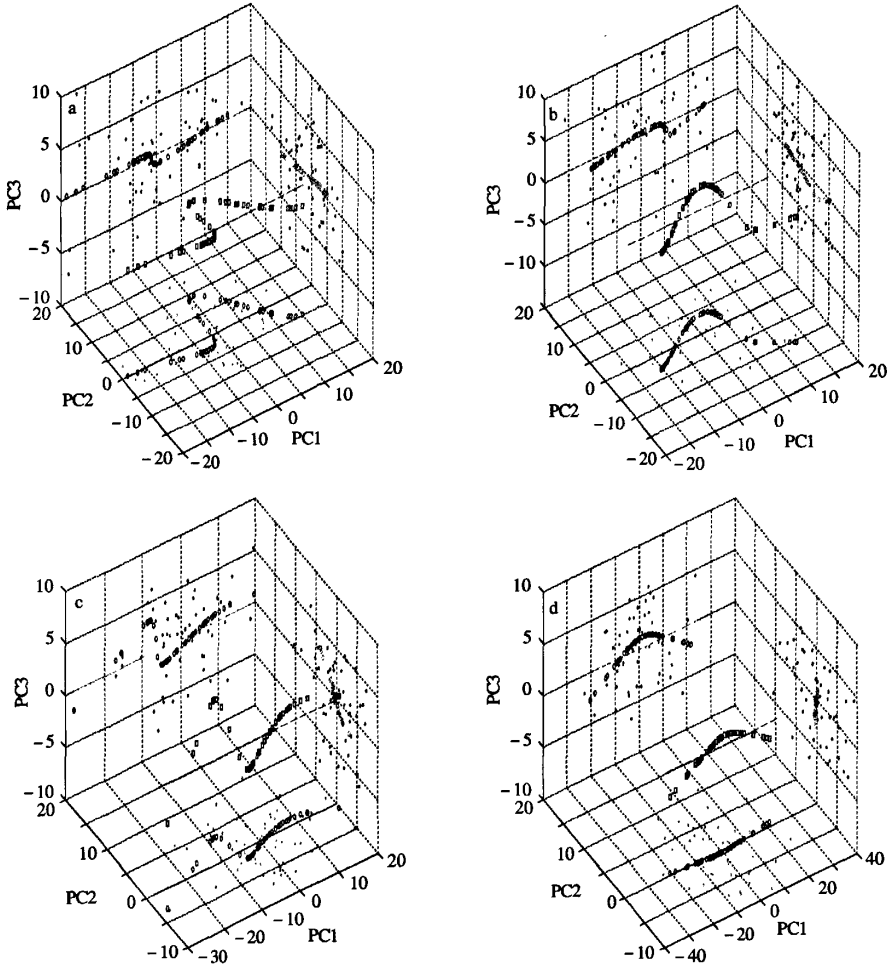


图 2 春季(a)、夏季(b)、秋季(c)和冬季(d)气温距平的 NLPCA 第一模态在 PC1-PC2-PC3 三维空间中的投影

Fig. 2 Projections of the first NLPCA modes of (a) spring, (b) summer, (c) autumn and (d) winter SAT anomaly in the PC1-PC2-PC3 3-D space

分别解释了原数据集方差的 55.20%、39.63%、49.25%、61.89%，均高于 PCA 第一模态解释的方差，其中春季和夏季表现的最为明显。

依旧取前 4 个时间系数序列作为分析的对象，同时将瓶颈层的神经元个数取为 2，可以提取中国四季气温距平的第 2 个模态(图略)。与 NLPCA 第 1 模态相比，四季 MSE 比率分别为春季 0.699、夏季 0.878、秋季 0.857、冬季 0.940，春、夏两季依旧表现出很强的非线性特征，秋季的非线性特征较第 1 模态有所增强。而冬季依旧表现为较强的线性特征。由此可见，冬季的非线性是最弱的。四季的 NLPCA 第二模态分别解释了原数据集总方差的 27.9%、31.4%、24.1% 和 22.9%，也均高于 PCA 第二模态解释的方差，表明 NLPCA 比 PCA 更好地提取了原资料集中的主要特征。

从主成分分析原理知道，PC1 和 PC2 是线性不相关的，但这里的不相关是指没有线性关系，它们之间还可能非线性关系，那么这种非线性关系是如何表现出来的呢？

以夏季为例，如图 3 所示，PC1、PC2 的非线性主成分近似 $X'_1(t_n)$ 、 $X'_2(t_n)$ 的样本点在图 3 上投影聚集为一条曲线，当 $X'_1(t_n)$ 取较大的正值和较大的负值时， $X'_2(t_n)$ 都为负值。PC1 和 PC2 之间的这种非线性关系结合 EOF1、EOF2 空间型就能反应气温距平场时空变率的不对称性。

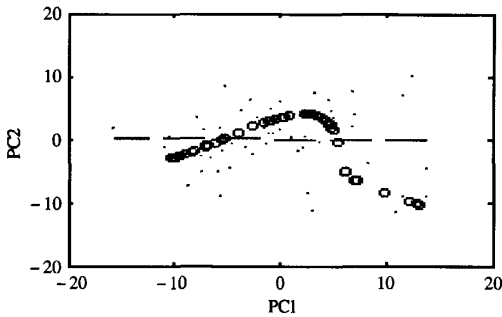


图 3 夏季气温距平场 NLPCA 第一模态近似在平面 PC1-PC2 上的投影(SATA 投影以实心点表示,NLPCA 第一模态近似以空心圆表示)
Fig. 3 The projection of the first NLPCA mode approximation of the summer SAT anomaly on the PC1-PC2 plane, where solid dots denote the projections of surface air temperature anomalies and open circles denote those of the first NLPCA mode approximation

从夏季 EOF1(图 4a)可以看出，华北和东北地

区为负距平区，因此 PC1 取强的负值和正值，分别表示气温正异常和负异常。从 EOF2(图 4b)可以看出，华北和东北地区也为负距平区，根据 PCA 原理，由于前两个模态解释的方差贡献较大，因此可用 PC1 乘 EOF1 加 PC2 乘 EOF2 近似地表示原始 SA-TA 场。由于 PC1 和 PC2 之间的非线性关系主要表现为当 PC1 取强的正负值时，PC2 都倾向于取负值，于是这两个成分的合成表示出的华北和东北地区的 SATA 特征为，气温偏高时的正异常值大于气温偏低时的负异常值，春季的情况与夏季较为相像，振荡在强度上的这种不对称性正是通过 PC1 和 PC2 之间的这种非线性关系反映出来的。

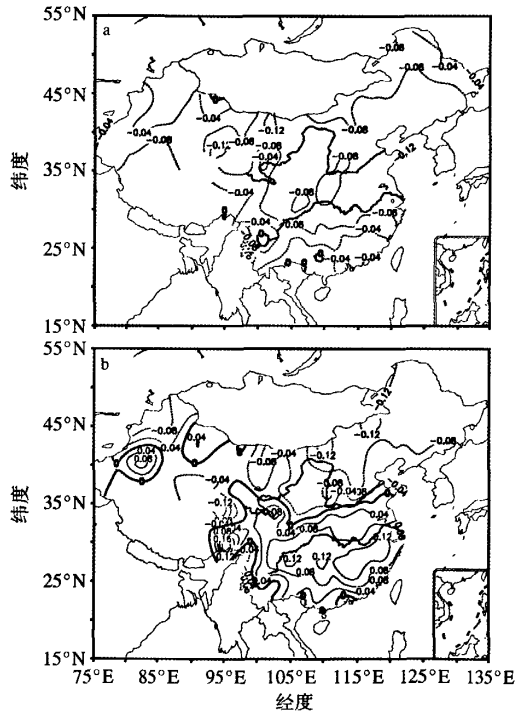


图 4 夏季气温距平场 EOF1(a)和 EOF2(b)的空间分布型

Fig. 4 Spatial patterns of (a) the EOF1 and (b) EOF2 of summer SAT anomalies

3.2 空间型的对称性和非对称性

NLPCA 的另一个优势在于，它能够表现出不同空间振荡型的结构。将神经网络瓶颈层神经元输出的时间序列称为非线性主成分时间序列 U ，对于一个特定的非线性主元 U ，利用神经网络将其从瓶颈层投影到输出层的 X' 再结合原来的前 4 个 EOF，可以返回生成多个不同位相下气温距平场的空间分布。当 U 取最大值和最小值时所对应的空间型是

非对称的,不再是镜像对应关系。随着非线性程度的加强,空间非对称性也有所增强。空间对称性的程度也是非线性强弱的另外一种表现形式。

图 5 显示了夏季和冬季 NLPCA 第 1 模态的空间型情况。在夏季,当 U 取最小值时(图 5b),全国气温异常几乎一致为负,且与夏季的 EOF1 分布型类似,而当 U 取最大值时(图 5a),正异常区主要位于我国长江以北的区域,长江以南的大部分地区为负异常区,气温的分布型为南负北正型,与 EOF2 的空间型分布类似,非对称性明显。

冬季当 U 取最小值时(图 5d),全国气温几乎一致为负,相反,当 U 取最大值时(图 5c),全国均表现为正异常,异常的空间分布型表现出很好的镜像关系。这也从另一个角度证明了冬季气温异常的非线性特征很弱。

为了对比一维 NLPCA 和一维 PCA 近似的结果,可用两近似场与原始距平场的相关系数地理分布图进行比较。以夏季为例,图 6a 是观测的 SATA

与一维 PCA 近似返回到地理空间所得的对 SATA 的近似之间的相关系数分布,图 6b 是观测的 SATA 与一维 NLPCA 近似之间的相关系数地理分布。如图 6 所示,除在西南很小的区域外,全国大部分地区 NLPCA 近似与 SATA 的相关比 PCA 近似与 SATA 的相关高。可见,一维 NLPCA 对原始 SATA 的近似比一维 PCA 近似更加逼真。

4 结论

(1) 四季气温 NLPCA 前两个模态解释的方差百分比均大于传统 PCA 解释的方差百分比,表明 NLPCA 能更好地提取原始资料中的主要特征。同时将夏季 PCA 和 NLPCA 两近似场与原始距平场的相关系数地理分布进行比较,发现除在西南很小的区域外,全国大部分地区 NLPCA 近似与 SATA 的相关比 PCA 近似与 SATA 的相关高。可见,一维 NLPCA 对原始 SATA 的近似比一维 PCA 近似更为逼真。

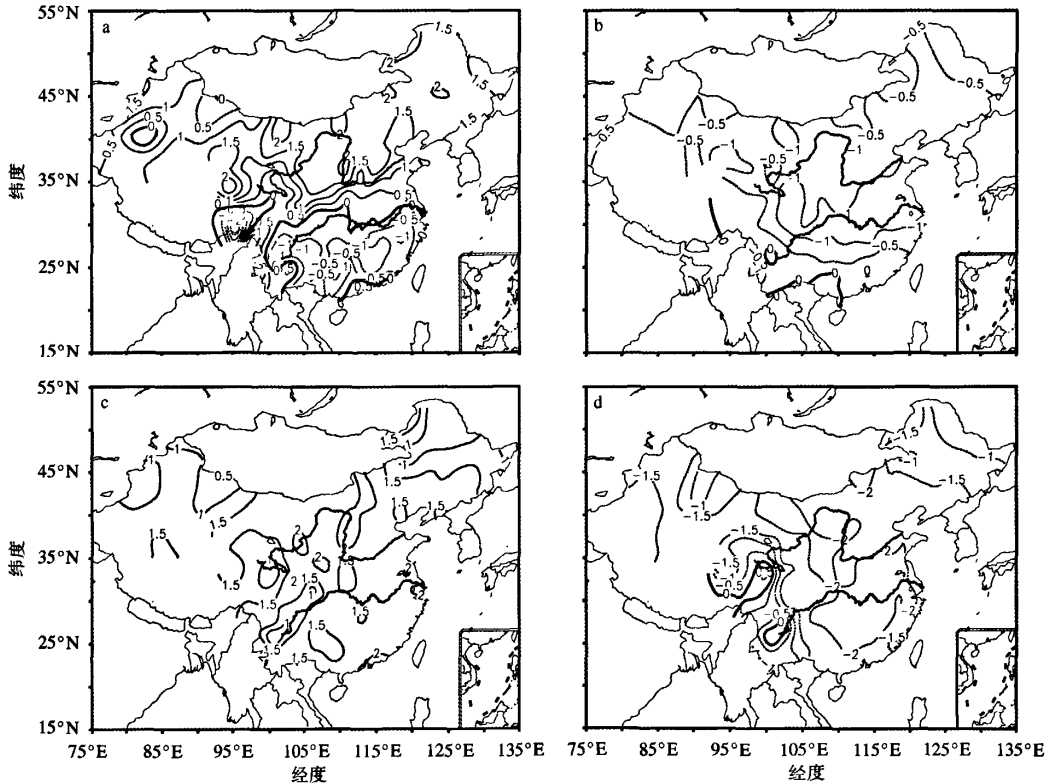


图 5 NLPCA 第一模态的非线性主元 U 取最大值(a,c)和最小值(b,d)时气温异常的空间分布型
a,b.夏季;c,d.冬季

Fig. 5 SAT anomaly patterns when the U of the first NLPCA mode is (a,c) maximum in (a,b) summer or (b,d) minimum in (c,d) winter

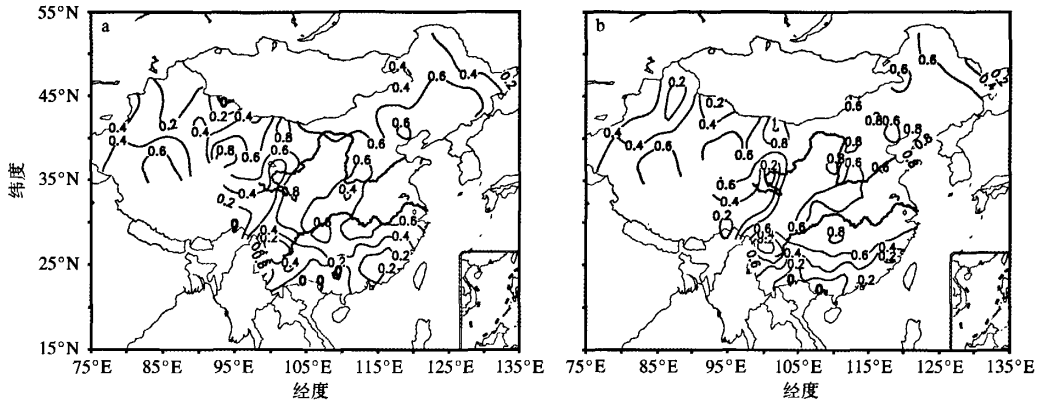


图 6 夏季气温距平场与一维 PCA 近似(a)及一维 NLPCA 近似(b)之间的相关系数分布

Fig. 6 Correlation coefficients between the summer SAT anomaly and (a) the 1-D PCA approximation, and (b) the 1-D NLPCA approximation, respectively

(2) 从 NLPCA 第一模态的三维空间投影图、均方误差 (MSE) 比率以及非线性主元 U 极值对应的空间型都可以看出四季气温异常变化的非线性具有季节性差异, 即春、夏两季较强, 秋、冬两季较弱。第二模态的 MSE 比率反映出春、夏两季非线性依然很强, 秋季的非线性较第一模态有所增强, 冬季仍表现出很强的线性特征。

参考文献:

- [1] Kramer M A. Nonlinear principal component analysis using autoassociative neural networks [J]. *AIChE Journal*, 1991, 37(2): 233-243.
- [2] Monahan A H. Nonlinear principal component analysis by neural networks: Theory and application to the Lorenz system [J]. *J Climate*, 2000, 15(2): 821-835.
- [3] Monahan A H. Nonlinear principal component analysis: Tropical Indo-Pacific sea surface temperature and sea level pressure [J]. *J Climate*, 2001, 14(1): 219-233.
- [4] 金龙. 人工神经网络技术发展及在大气科学领域中的应用 [J]. *气象科技*, 2004, 32(6): 385-392.
- [5] 金龙, 吴建生, 林开平, 等. 基于遗传算法的神经网络短期气候预测模型 [J]. *高原气象*, 2005, 24(6): 981-987.
- [6] 金龙, 林熙, 金健, 等. 模块化模糊神经网络的数值预报产品释用预报研究 [J]. *气象学报*, 2003, 61(1): 78-84.
- [7] 金龙, 罗莹, 王业宏, 等. 月降水量的神经网络混合预报模型研究 [J]. *高原气象*, 2003, 22(6): 619-623.
- [8] 林健玲, 金龙, 林开平. 神经网络方法在广西日降水预报中的应用 [J]. *南京气象学院学报*, 2006, 29(2): 215-219.
- [9] 李永华, 金龙, 缪启龙, 等. 基于 SSA-MGF 的 BP 神经网络多部预测模型 [J]. *南京气象学院学报*, 2005, 28(4): 550-555.
- [10] 姚志刚, 陈洪滨. 利用神经网络从 118.75GHz 附近通道亮温反演大气温度 [J]. *气象科学*, 2006, 26(3): 252-259.
- [11] 陈仁芳, 刘静. 人工神经网络方法预报长江上游流域雨量的探讨 [J]. *气象科学*, 2004, 24(4): 483-487.
- [12] 郑凤琴, 孙崇志, 林金渝. 神经网络在水位预报中的研究 [J]. *气象科学*, 2006, 26(1): 53-57.
- [13] 韦莹莹, 吴洪宝, 沈桂凤, 等. 北半球冬季大气环流遥相关型的非线性特征 [J]. *南京气象学院学报*, 2007, 30(1): 17-25.
- [14] 吴洪宝, 吴蕾. 气候变率诊断和预测方法 [M]. 北京: 气象出版社, 2005.

(责任编辑: 张福颖)