

基于神经网络-奇异谱分析的 ENSO 指数预测

严 军 刘健文

空军装备研究院航空气象研究所, 北京 100085

摘 要 采用奇异谱分析方法研究了 ENSO 指数及相关序列, 结果表明奇异谱分析能很好的对原始序列进行信噪分离, 增大了 ENSO 指数的可预报性。在此基础上, 提出了人工神经网络和奇异谱分析相结合的 ENSO 指数预测方法, 进行了不同因子组合的预报试验, 预报效果明显优于持续性预报, 超前 4 季的 Niño 3 区、Niño 4 区预报相关系数仍高于 0.5。

关键词 ENSO 指数 神经网络 奇异谱分析 时间序列

文章编号 1006-9895(2005)04-0620-07 **中图分类号** P461 **文献标识码** A

A Study of ENSO Index Prediction Based on Neural Network – Singular Spectrum Analysis

YAN Jun and LIU Jian-Wen

Beijing Aviation Meteorological Institute, Beijing 100085

Abstract ENSO indices and the related time series are studied by singular spectrum analysis (SSA). The results show that SSA can detect the signal and noise from the original time series, enhance the predictability of ENSO indices. And by that, a model based on neural network and singular spectrum analysis (NN-SSA) is built to forecast ENSO indices. The model is applied with different combinations of predictors from the time series. It is shown that NN-SSA model performance is higher than persistent forecasts evidently, the best in the zonal wind index at 850 hPa and Niño region indices as inputs. The correlation coefficient on Niño 3 and Niño 4 SSTA forecast is still above 0.5 at lead time of 4 quarters. There is also seasonal dependence on SSTA forecast skills in the NN-SSA model experiments. Compared with other statistic models (LR, CCA and Persistence etc.), NN-SSA model shows comparable or predominant skills on Niño 3 SSTA forecast, and the correlation coefficient decreases very slowly as the lead time increases.

Key words ENSO index, neural network, singular spectrum analysis (SSA), time series

1 引言

ENSO 现象是迄今为止发现的全球气候系统中最强的年际变化信号, 对全球气候产生着重要影响, 也是我国短期气候预测的重要研究内容之一。自 20 世纪 80 年代中期以来, 国内外研制了众多的 ENSO 预测模式^[1, 2], 虽已成功地预报了一些 EN-

SO 事件, 但仍存在许多不足之处。例如, 预报效果的稳定性随季节漂移, 预报时效的延伸长度随不同模式而有明显差异等等^[3]。ENSO 的发生发展过程中包含着非线性的物理过程和相互作用, 而目前多数的基于动力或统计的 ENSO 预测模式都较少处理非线性的作用。

近年来, 在人工智能领域中的神经网络(neural

network, 简称 NN)技术已逐步应用到气象研究领域,取得了一些有价值的研究成果^[4~6]。神经网络是一种抽象的数学模型,其实质是一门非线性科学,它具有大规模并行处理、分布存储和容错性、自学习自组织和自适应性等特点,在自动化控制、模式识别、专家系统和预报预测等领域都有很好的应用^[7, 8]。人工神经网络独特的性能在处理非线性问题方面有广阔的应用前景,直接将其运用在气象、海洋领域研究中,由于资料以及网络自身的特点,还有一些困难。但通过和其他统计方法相结合,其预报效果有很大的改进^[9]。本文直接从 ENSO 观测资料入手,利用奇异谱分析技术(singular spectrum analysis, 简称 SSA),采用时滞方法,重新构造时空动力学系统,提取 ENSO 指数序列中的主要振荡信号,然后通过神经网络建立模型,以提高对 ENSO 事件的预测能力。

2 原理和方法

在人工神经网络结构中,基于误差反向传播训练算法(Back Propagation, 简称 BP 算法)的网络是目前应用最为广泛的一种。BP 算法是在有指导的情况下,适合于多层神经网络的一种学习算法,它是建立在梯度下降法的基础上的。

典型的 BP 网络是由输入层、隐含层和输出层组成,各层之间实行全连接,同一层之间不存在相互连接(如图 1)。每一连接弧连接着两个神经元节点,并附有一数值作为权值,权值的大小表示节点间连接的强度。输入层从外部接收信息并将此信息传入网络;隐含层接收输入层的信息,然后对所有的信息进行处理;输出层接收网络处理后的信息,并将最后结果输出。误差反传算法的主要思想是把

学习过程分为两个阶段:第一阶段(正向传播过程),输入信息从输入层经过隐含层逐层处理并计算出各单元的实际输出值;第二阶段(反向过程),若在输出层不能得到期望的输出,那么逐层计算实际输出与要求输出之差值(即误差),以便据此差值调整权值,从后向前修正各层次之间的联系权重,在不断的学习和修正过程中,以使网络的学习误差达到最小^[7]。

本文所用的网络结构(图 1 所示)的多输入单输出的结构,其中输入层和隐含层间的传递函数为 sigmoid 型,隐含层和输出层之间为线性函数连接。输入和输出层的神经元数目依赖于预报因子和预报量的数目。本文在每次建模时的预报量为一个,输出层设一个神经元。选择网络的输入,即预报因子。超前时间定义为所要预报季节初始点超前于起报点的时间量^[10, 11]。模型设计是逐季预报的,比如资料截至 3 月份,超前 0 季的预报,即指预报 4、5、6 月份的值。起报点之前的 4 个季的预报因子构成网络的输入,预报量为 Niño 区指数,即网络的期望输出。

原始预报因子序列在进入网络之前进行奇异谱分析提取主要振荡分量作为网络的输入。SSA 是从时间序列的动力重构出发,并与经验正交函数相联系的一种统计技术。SSA 的分析对象是一维时间序列。采用动力系统分析中的办法,把序列时迟排列建立相空间。它计算出的特征向量和主分量分别称为时间特征向量(T-EOF)和时间主分量(T-PC)。SSA 易于识别和分离时间序列中的振荡分量。序列中的某一个振荡与奇异谱分析中一对接近退化的特征值相联系。一般,当下述三个条件成立时,可以逐一确定序列中在窗口长度分析范围内的振荡^[12, 13]:

- (1) 按序排列的二个相邻特征值近似相等;
- (2) 对应的时间特征向量具有准周期性,且周期相等,位相正交;
- (3) 相应的时间主分量位相正交。

时间主分量 T-PC 是原始序列的过滤形式。要研究振荡子集或恢复序列的主要特征时,需要进行主分量重构。使用重构的主分量代替时间主分量的主要优点,使位相延迟得到了复原。

从预报量和预报因子数据序列的奇异谱分析开始,到最终结果的输出,可通过图 2 的流程图来说

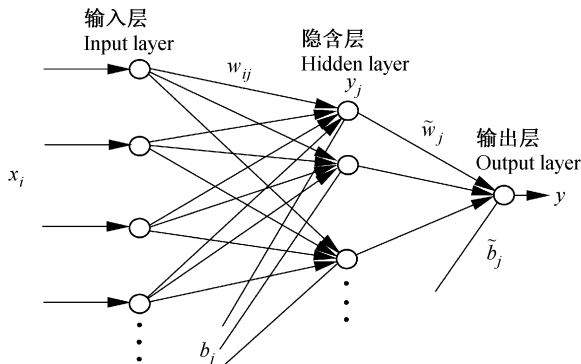


图 1 基于 BP 算法的神经网络结构示意图

Fig. 1 Neural network based on BP algorithm

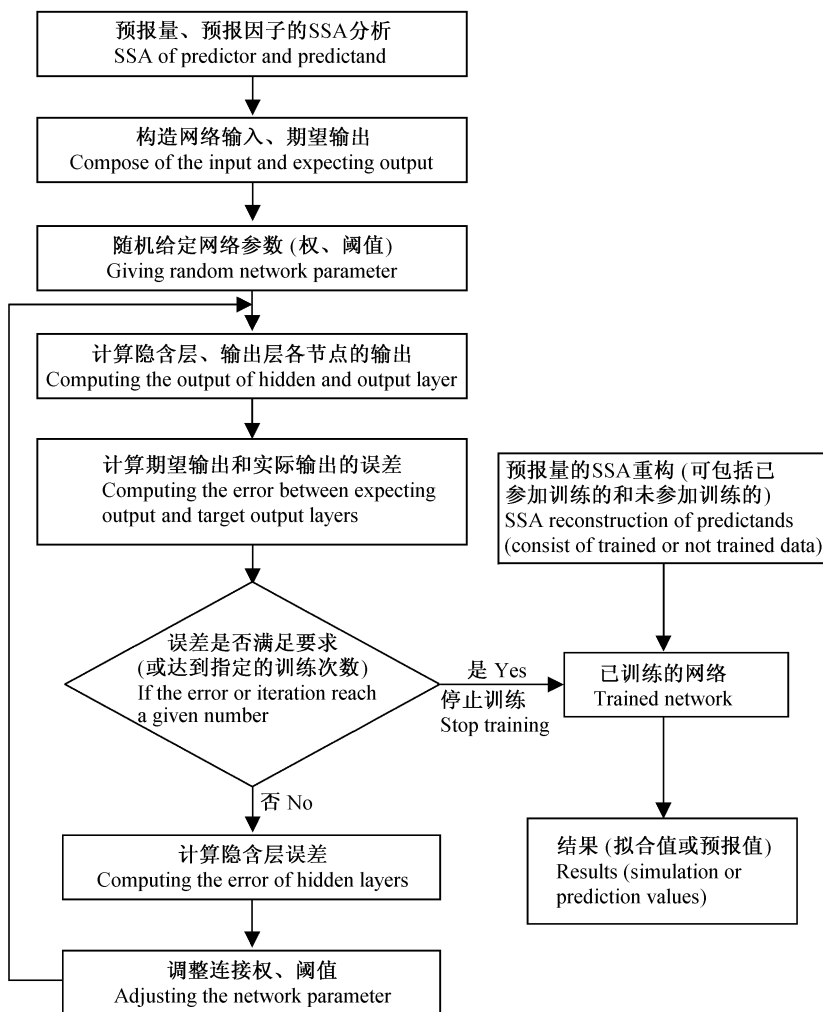


图 2 NN-SSA 实现流程图

Fig. 2 NN-SSA modeling flow chart

明。图中左边部分是网络训练过程,应用误差反传算法。循环次数依情况而定,若误差小于指定值或训练到一定的次数误差始终没有满足要求,则停止训练;图中右面部分是网络模拟,其中若输入的数据(预报因子)是已参加训练过的,则模拟的结果就是拟合值,若输入的是未参加训练的数据则得到的结果将是预报值。

通过对输入数据的奇异谱分析后,得到的重构数据完全拥有原始序列的有用信息,而去掉了“噪声”部分,就相当于一个数字滤波器,使预报建模效率提高,增大了 ENSO 指数的可预报性。

3 资料及其分析

本研究所用的资料均为月平均序列,由中国国

家气候中心气候诊断室整编提供。资料包括:时段为 1950 年 1 月~2002 年 2 月的 Niño 1+2 区、Niño 3 区、Niño 3.4 区, Niño 4 区的 SSTa, 南方涛动指数(SOI), 太阳黑子相对数(为后述方便简记为 SUNSPOT, 下同);时段为 1975 年 1 月~2002 年 2 月的 850 hPa 赤道西太平洋纬向风指数序列(简称 850WPZW), 范围 135°E~180°; 赤道中太平洋纬向风指数序列(简称 850MPZW), 范围 170°W~120°W; 200 hPa 赤道西太平洋纬向风指数序列(简称 200WPZW), 范围 135°E~180°; 赤道中太平洋纬向风指数序列(简称 200MPZW), 范围 175°W~120°W; 赤道区域(160°E~160°W)射出长波辐射时间序列(OLR)。

采用 SSA 方法,对上述数据序列进行逐一分

析, 结果见表 1。SSA 的一个关键问题是如何适当选取窗口长度 M 。这是一个经验变量, 取决于所讨论问题的时间尺度。试验表明, 当窗口长度为 M 时, 能较好地识别出周期为 $M/5 \sim M$ 的振荡。针对本文研究问题, 选用了 60 个月的窗口长度(太阳黑子相对数用了 150 个月)。

从表 1 中可以看出, Niño 1+2 指数, Niño 3 指数, Niño 3.4 指数和 SOI 均明显存在准 4 年和准 2 年周期的二个主要振荡; Niño 4 指数只存在准 4 年左右的一个主要振荡; 850MPZW 和 OLR 明显存在准 5 年和准 2 年的主要周期振荡; 850WPZW、200WPZW 和 200MPZW 只存在准两年的主要振荡, 而且 200WPZW 的准两年的特征值对子仅解释了总方差的 6%, 说明该 2 年振荡不是很突出; SUNSPOT 则明显存在准 11 年的周期, 其前 2 个特征值对子即解释了总方差的 81.5%。分析结果说明, 对 Niño 海区和 SOI 序列振荡分量的重构很好地反映了 ENSO 事件的时间和振幅特征。

根据以上分析, 最终确定预报量和预报因子的重构主分量的数目及其累计解释方差。其中重构序列包含上面各指数的主要振荡分量及部分显著主分量。参考 Tangang 等^[10, 11]对数据场作 EOF 分析和建模的方案, 本文为了使预报量序列更逼近原始序

列, 在预报量重构时选取了更多的主分量数目。SSA 重构的特征序列, 包含了原序列的主要周期振荡而滤去了噪音部分, 应具有更大的可预报性。

4 ENSO 指数预报试验

4.1 预报检验方法

为了客观检验已训练好的神经网络-奇异谱模型的泛化性能, 即对没有参加训练的数据的拟合能力, 采用交叉检验方法^[14]。交叉检验的做法是, 首先将输入数据集分成 k 个大小(近似)相等的子集, 在训练时, 每次从输入数据集中扣除这样一个子集, 用剩余的数据集输入网络进行训练。这样分别对剩余子集训练, 对扣除部分进行预报, 向前滑动构成整个预报, 最后对预报值和实况值作比较分析来说明模式的预报效果。本文所用建模数据集, 根据所选预报因子和预报量的不同, 有两个样本序列: 1950 年 1 月~1999 年 12 月, 样本容量为 600 个月; 1975 年 1 月~1999 年 12 月, 样本容量为 300 个月。在设计交叉检验时, 每次留出 5 年(共 60 个月)的连续样本, 用剩余的样本来建模, 然后用所建模型对留出的 5 年样本进行预报。这样向前滑动, 构成整个预报序列。

4.2 预报因子组合试验

神经网络模型中, 输入量(预报因子)的选择很重要, 直接影响着其预报推广能力。本文所用的因子包括反映海洋状况, 大气对流层低层、高层和大气对流状况的指数, 以及天文指数太阳黑子相对数。这些指数具有各自的物理意义, 不同的因子组合会得到不同的预报效果。通过组合试验可以确定最佳的预报因子集。

表 2 列出了分别引入南方涛动指数、纬向风指数、射出长波辐射、太阳黑子等因子和各海区海温指数组合在超前 0~4 季的预报效果。比较中采用相同的资料样本容量即 1975 年 1 月至 1999 年 12 月共 300 个月。

从各因子与海温指数组合的预报相关系数大小可以将这些因子在 0~4 季超前的预报能力分成 3 类。相关系数都在 0.6 以上的有 850MPZW、200MPZW、850MPZW+200MPZW、850MPZW+SOI、200MPZW+SOI、850MPZW+200MPZW+SOI, 其中各超前综合最好的是 850MPZW, 而 200MPZW+SOI、850WPZW 在 3、4 季长的超前

表 1 各指数的周期信号特征				
Table 1 Periodical characteristics of different indices				
指数 Index	振荡 Oscillation	特征值对 Eigenvalue pairs	解释方差 Explain variance / %	谱峰对应周期 (月)Period (months)
Niño 1+2	1	1, 2	41.4	50
	2	3, 4	22.5	29
Niño 3	1	1, 2	44.0	45
	2	3, 4	21.6	28
Niño 4	1	1, 2	43.5	55.6
Niño 3.4	1	1, 2	44.6	50
	2	3, 4	23.4	29
SOI	1	1, 2	26.0	50
	2	4, 5	19.5	29
850WPZW	1	4, 5	16.5	29
850MPZW	1	1, 2	33.6	62
	2	3, 4	21.3	30
200WPZW	1	5, 6	6.0	25
200MPZW	1	4, 5	17.0	29
SUNSPOT	1	1, 2	81.5	125
OLR	1	1, 2	36.7	62
	2	4, 5	18.9	29

表 2 不同因子与 Niño 区海温指数组合在 0~4 季超前预报的相关系数

Table 2 The forecasted correlation coefficients for different predictors combined with the sea surface temperature anomaly (SSTA) in the Niño region at lead time 0-4 quarters

预报因子	0 季	1 季	2 季	3 季	4 季
Predictor	quarter	quarter	quarter	quarter	quarter
SOI	0.84	0.72	0.62	0.57	0.55
850MPZW	0.84	0.73	0.64	0.62	0.63
850WPZW	0.73	0.68	0.59	0.52	0.53
200MPZW	0.80	0.67	0.61	0.63	0.65
200WPZW	0.82	0.67	0.53	0.46	0.43
850MPZW+200MPZW	0.82	0.70	0.61	0.61	0.63
850MPZW+850WPZW	0.84	0.72	0.59	0.48	0.41
200MPZW+200WPZW	0.78	0.64	0.56	0.57	0.58
850MPZW+SOI	0.83	0.71	0.62	0.61	0.63
200MPZW+SOI	0.79	0.68	0.62	0.63	0.64
850MPZW+200MPZW+SOI	0.81	0.71	0.63	0.62	0.64
OLR	0.83	0.69	0.57	0.52	0.51
OLR+SOI	0.83	0.71	0.60	0.53	0.51
SUNSPOT	0.84	0.68	0.54	0.48	0.45
SUNSPOT+SOI	0.82	0.66	0.51	0.45	0.45
SOI+OLR+850MPZW	0.82	0.71	0.60	0.55	0.56
SOI+OLR+SUNSPOT	0.82	0.66	0.52	0.46	0.45

中占优势。相关系数在 0.5 以上的有: 850WPZW、200WPZW、OLR、OLR + SOI、SOI + OLR + 850MPZW。剩余的相关系数也都在 0.4 以上, 最低的是太阳黑子 SUNSPOT 与海温的组合。结果说明, 具有不同物理意义的因子组合对预报效果的影响是有差异的, 并不是因子越多越好。其中对流层低层赤道中太平洋纬向风(850MPZW)的引入对 ENSO 的预报能力在各类因子组合中达到综合最高。对流层高层赤道中太平洋纬向风(200MPZW)与各因子组合的预报效果在长的超前中占优。在 ENSO 发展中, 对流层大气风场起着重要作用。在与太阳黑子相对数的因子组合的预报效果表现相对较低。射出长波辐射、南方涛动的引入的预报技巧在上述两者之间。

为了比较模式在不同海区、不同季节以及不同模式对比的预报效果, 在此预报试验的基础上, 下面的预报检验均选择 1950 年 1 月~1999 年 12 月时段的 SOI 指数和海温指数的组合, 对 Niño 区 SSTA 进行预报效果的对比分析。

4.3 预报效果的区域变化

表 3 是模式对各 Niño 区在 0~4 季超前预报的相关系数。从表3可见, 随着超前季节增长, 预报

表 3 各 Niño 区超前 0~4 季预报的相关系数
Table 3 The forecasted correlation coefficients for different Niño regions at lead time 0-4 quarters

超前时间(季)				
Lead time (quarters)	Niño 1+2	Niño 3	Niño 4	Niño 3.4
0	0.80	0.85	0.87	0.88
1	0.56	0.71	0.76	0.76
2	0.38	0.57	0.63	0.61
3	0.33	0.51	0.55	0.51
4	0.33	0.51	0.54	0.49

准确性逐渐降低, 不同海区的预报效果存在一定差异, 其中以 Niño4 区平均预报效果最好, 在超前 4 季的相关系数在 0.54, 其次是 Niño 3 区和 Niño 3.4 区, 预报效果最差的是 Niño 1+2 区, 在超前 4 季的预报相关系数为 0.33。

4.4 预报效果的季节变化

已有研究表明, ENSO 预测存在着季节依赖性问题(seasonal dependence)。图 3 中是各 Niño 区 SSTA 的预报相关系数随着超前时间和月份的变化。可以看出, 4 个 Niño 区的各个超前的相关变化都有一条“槽线”对应个超前的低值点, 其中 Niño 1+2 区的预报在春季预报效果最差, Niño 3 区的最低则在初夏季节, Niño 4 区和 Niño 3.4 区的最低值又推移至夏季到秋季。从上一节的各 Niño 区的总体预报效果分析, 最低值也存在着类似的变化趋势。

4.5 NN-SSA 与其他统计模式结果的比较

采用奇异谱分析和神经网络相结合, 就是运用奇异谱方法提取数据的主要特征, 提高模型对 EN-SO 事件的预报能力。本文比较了 NN-SSA 与单独采用神经网络(NN)、线性回归模式(LR)和典型相关分析(CCA)模式的预报技巧。LR、CCA 的结果引自文献[14], 其中采用的 2°×2°网格的 SLP (20°S~20°N)、SST(30°S~30°N)资料与本文不同。4 种方法的预报时段都是 1950 年 1 月~1997 年 12 月。预报技巧的评估采用交叉检验的相关系数。图 4 给出了 4 种方法及持续性预报在不同超前时段对 Niño 3 区 SSTA 的预报技巧。

从图 4 可看出数据经过 SSA 处理的 NN-SSA 模型, 在所有方法中是最优的。与直接用原始序列输入的网络预报效果(NN)相比, NN-SSA 随着预报时效的增长明显表现出优势。超前 1~3 季的 NN-SSA 预报相当或略优于 LR、CCA, 但超前 4

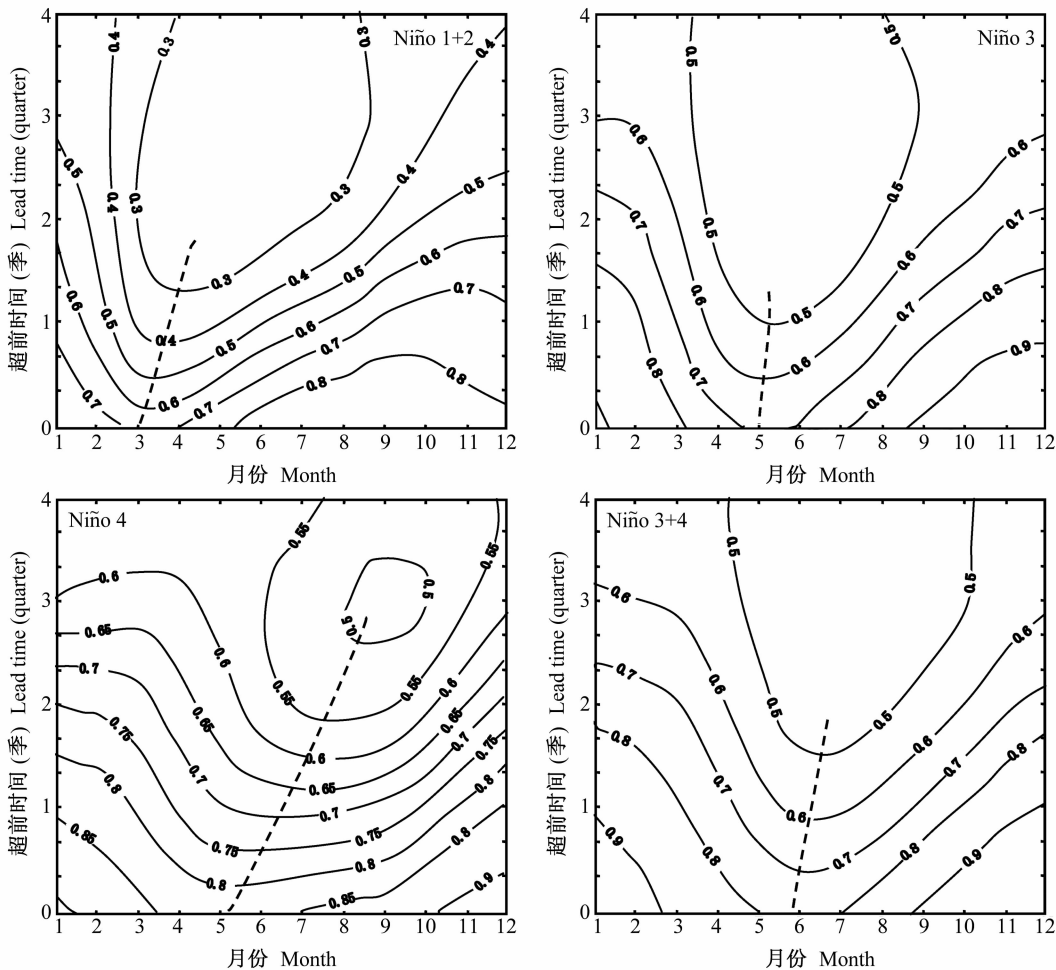


图 3 各 Niño 区 SSTA 超前 0~4 季预报相关系数的季节变化

Fig. 3 The seasonal variation of the correlation coefficients for SSTA forecast in different Niño regions at lead time 0-4 quarter

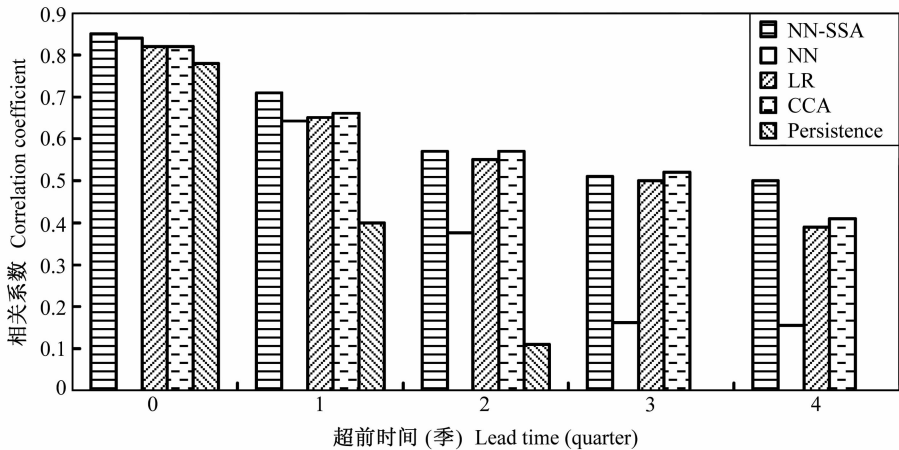


图 4 NN-SSA 与其他统计模式对 Niño 3 区 SSTA 的预报相关系数

Fig. 4 The forecast skill of SSTA in the Niño 3 region between the NN-SSA and other statistic models

季时明显优于 LR、CCA。持续性预报与 4 种方法比较在所有超前时段都是最低的，在超前 3、4 季

时甚至达到了负相关。
计算过程中还发现，在相同的计算条件下，不

经过任何预处理的网络训练时间要明显长于经过 SSA 处理的情况。前者一方面数据量很大, 另外其中还包含着各种噪音, 在计算中不仅费时, 而且占用大量的计算机内存。这也说明了 SSA 能很好的对原始序列进行信噪分离, 不仅增大了 ENSO 特征指数的可预报性, 而且使预报建模计算效率高, 效果稳定。

5 小结

(1) 在 ENSO 指数预测中, 基于神经网络奇异谱分析的预测模型是一种行之有效的统计方法。从建模理论到检验、比较, 都表明了该模型的可行性, 其预报效果稳定。

(2) 具有不同物理意义的因子组合对预报效果的影响是有差异的, 其中引入 850 hPa 赤道中太平洋纬向风指数对 ENSO 预报能力的提高在各类因子组合中达到综合最高。对流层高层赤道中太平洋纬向风指数与各因子组合的预报效果在长的超前中占优。

(3) 各 Niño 海区的预报效果存在一定差异, 其中以 Niño 4 区平均预报效果最好, 其次是 Niño 3 区和 Niño 3.4 区。预报效果还存在季节变化, 且随超前时效变化时预报技巧的变化呈现不同特点。从海温指数各预报时效平均情况来看, 其中 Niño 1+2 区的预报在春季预报效果最差, Niño 3 区的最低则在初夏季节, Niño 4 区和 Niño 3.4 区的最低值又推移至夏季到秋季。

(4) 与 NN 和其他统计预报的比较说明, NN-SSA 的预报占有优势, 尤其是在长的超前时段中。

人工神经网络技术在气象领域的应用还是初步的。在本文研究的基础上进一步研究网络揭示的非线性和 ENSO 机制的联系, 结合先进的统计技术, 优化网络结构等, 有望成为 ENSO 预测业务的一种新方案。

参考文献(References)

- [1] Barnston A G, Ropelewski C F. Prediction of ENSO episodes using canonical correlation analysis. *J. Climate*, 1992, **5** (11): 1316~1345
- [2] Barnston A G, Van den Dool H M, et al. Long-lead seasonal forecasts—Where do we stand? *Bull. Amer. Meteor. Soc.*, 1994, **75**: 2097~2114
- [3] 翟盘茂, 江吉喜, 张人禾. ENSO 监测和预测研究. 北京: 气象出版社, 2000. 108~114
- Zhai Panmao, Jiang Jixi, Zhang Renhe. *Studies on ENSO Monitoring and Prediction* (in Chinese). Beijing: China Meteorological Press, 2000. 108~114
- [4] 金龙, 秦伟良, 姚华栋. 多步预测的小波神经网络预报模型. 大气科学, 2000, **24**: 79~86
- Jing Long, Qin Weiliang, Yao Huadong. A multi-step prediction model of wavelet neural network. *Chinese Journal of Atmospheric Sciences* (in Chinese), 2000, **24**: 79~86
- [5] 张初, 蒋国荣. 赤道东太平洋海温及 El Niño/La Niña 的反演及预测. 解放军理工大学学报, 2000, **5**: 7~12
- Zhang Ren, Jiang Guorong. Retrieve and prediction of equatorial East-Pacific SST and El Niño/La Niña. *J. PLA Univ. Sci. Tech.* (in Chinese), 2000, **5**: 7~12
- [6] 张初. 基于前传式网络逼近的太平洋副热带高压活动的诊断预测. 大气科学, 2001, **25**(5): 650~660
- Zhang Ren. A diagnosis/prediction of the movement of the western pacific subtropical high based on the neural networks approximation with forth-propagation style. *Chinese Journal of Atmospheric Sciences* (in Chinese), 2001, **25**(5): 650~660
- [7] Hsieh W W, Tang Benyang. Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bull. Amer. Meteor. Soc.*, 1998, **79**: 1855~1870
- [8] 袁曾任. 人工神经元网络及其应用. 北京: 清华大学出版社, 1999. 1~27, 66~130
- Yuan Zengren. *Artificial Neural Network and Its Application* (in Chinese). Beijing: Tsinghua University Press, 1999. 1~27, 66~130
- [9] Tangang F T, Tang B, Monahan A H. Forecasting ENSO events: a neural network-extended EOF approach. *Journal of Climate*, 1998, **11**: 29~41
- [10] Tangang F T, Hsieh W W, Tang B. Forecasting the equatorial Pacific sea surface temperatures by neural network models. *Climate Dyn.*, 1997, **13**: 135~147
- [11] Tangang F T, Hsieh W W, Tang B. Forecasting regional sea surface temperatures in the tropical Pacific by neural network models, with wind stress and sea level pressure as predictors. *J. Geophys. Res.*, 1998, **103**(C4): 7511~7522
- [12] 丁裕国, 江志红. 气象数据时间序列信号处理, 北京: 气象出版社, 1998. 160~167
- Ding Yuguo, Jiang Zhihong. *Signal Processing of Time Series in Meteorology* (in Chinese). Beijing: China Meteorological Press, 1998. 160~167
- [13] Liu Jianwen, Dong Peiming. Short-range climate prediction experiment of the southern oscillation index based on the singular spectrum analysis. *Adv. Atmos. Sci.*, 2001, **18**: 873~881
- [14] Tang B Y, Hsieh W W, Monahan A H, et al. Skill comparisons between neural networks and canonical correlation analysis in predicting the equatorial Pacific sea surface temperatures. *J. Climate*, 2000, **13**: 287~293