

区域降水数值预报产品人工神经网络释用预报研究

林健玲¹ 金龙² 彭海燕³

(1 南京信息工程大学, 南京 210044; 2 广西壮族自治区气象局气象减灾研究所, 南宁 530022; 3 江苏省气象局, 南京 530022)

摘要 利用 T213、日本细网格降水预报等数值预报产品,采用神经网络方法进行预报释用。通过聚类分析方法对广西自治区测站进行分类,简化预报对象,对数量众多的 T213 数值预报产品采用自然正交分解(EOF)方法,浓缩大量因子的有效信息,并结合日本降水预报因子建立广西 5~6 月区域降水量级的逐日人工神经网络预报模型。运用与实际业务预报相同的方法进行逐日预报试验。结果表明,用这种数值预报产品释用方法建立广西 3 个预报区域的 BP 人工神经网络预报模型对中雨以上降水量级预报的 TS 评分分别为 0.55、0.5 和 0.26,比目前业务预报中参考使用的 T213 和日本数值预报产品降水预报具有更好的预报效果。

关键词 神经网络 EOF 数值预报产品释用 降水量级预报

引言

人工神经网络方法是一种人工智能的方法,它是以模拟人脑神经网络的结构和接受、处理信息为基础的智能算法。人工神经网络最明显的优点是具有较好的自适应学习能力和非线性映射能力,比较适合处理那些物理机制复杂、因果关系和推理规则难以确定的非线性问题,它的这些特点为其运用到天气预报中提供了坚实的基础。该方法自 20 世纪 80 年代中期开始被应用于气象领域以来,得到了积极的开发和利用,已广泛应用于如资料同化、资料解释、预报制作、预报质量保证、数据库开发、决策支持、混合数据处理以及影响评价等各个方面,在大气科学领域应用方面有广泛的应用前景。

广西地处亚热带季风区,5~6 月降水不仅受中高纬西风环流影响,而且还受西太平洋副高、台风、东亚季风和印度西南季风环流的影响。由于这些因子之间的非线性作用十分复杂,降水预报难度较大,至今尚未研制出一种预报稳定且水平较高的业务预报方法。本文以数值预报产品资料为基础尝试用人工神经网络模型对广西逐日降水量级进行预报,以

寻求区域降水量级预报的新途径。

1 人工神经网络基本原理

本文采用 BP 神经网络模型^[1],各神经元是分层排列的,每个神经元只与前一层的神经元相连接,网络中除输入层单元和输出层单元外,还有不与外层直接联系的中间层单元,称为隐层单元,一般隐层是由多个节点构成的。网络第一层单元的状态只影响下一层单元的状态,也就是每一层单元只接收来自前一层单元的输入信号,并对后一层单元输入,形成前馈式特点(如图 1 所示)。

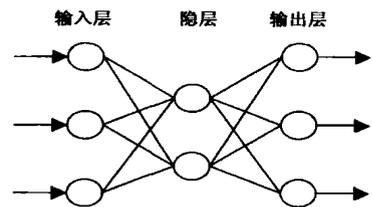


图 1 人工神经网络结构示意图

前馈式网络模型物理概念清晰,通用性好,具有高度灵活可变的拓扑结构,具有很强的非线性映射能力。前馈式网络最常使用的算法是误差反向传播算

广西自然科学基金项目(0339025) 国家自然科学基金项目(40075021) 共同资助

作者简介:林健玲,女,1976 年生,硕士生,主要从事短期天气预报研究,Email: linjl@tom.com

收稿日期:2004 年 11 月 15 日;定稿日期:2005 年 1 月 25 日

法(back propagation algorithm),这时的前馈式网络就称为误差反传神经网络,简称BP神经网络。在理论上已经证明,一个3层的前馈神经网络能够实现任意精度的连续函数映射^[2]。它的基本思想是:如果利用已有连接权和阈值正向传播得不到期望的输出,则反向传播学习,反复修改各节点的连接权和阈值,逐步减小代价函数,直到代价函数不再减小或达到预先设定的误差标准,此时确立输入与输出映射关系。

2 预报建模

2.1 预报对象的处理

聚类分析的主要依据是把相似的样本归为一类,而把差异大的样本区分开来。采用聚类分析能够在一定程度上把复杂的广西降水分成具有共同特点的类,从而降低预报难度。在聚类过程中,我们选用欧氏距离来度量类与类之间的相似程度,聚类方法采用类平均法。另外,在聚类过程中会出现部分站点独自成类的现象,我们再根据该站点所在位置把它人为地归入距其最近的区。

根据1951~2000年50年5、6月广西89个测站逐日降水资料,经过聚类分析,把广西分成3个区(如图2所示)。分别计算3个区域内所有站的平均降水量,并进行分级,1级为无雨,2级为小雨(0.1~9.9 mm),3级为中雨(10.0~24.9 mm),4级为大雨(25.0~49.9 mm),5级为暴雨以上降水(≥ 50.0 mm)。经过聚类分析和分级处理后,预报对象转化为3个分区的降水量级(1~5)的预报。神经网络输出的是各分区的降水分级预报。

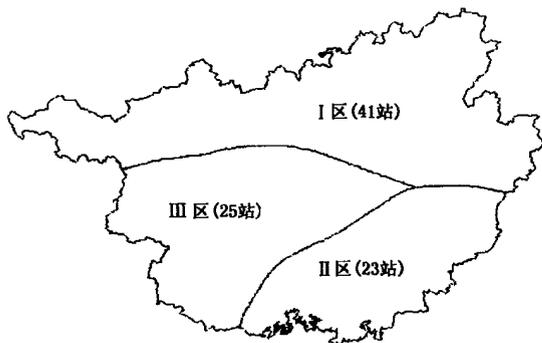


图2 广西降水聚类分析分区图

2.2 预报因子

2.2.1 相关普查

本文所选用的数值预报产品均为48 h预报场,

包括:①T213各标准层17个常规气象要素及物理量要素场($100^{\circ} \sim 120^{\circ} \text{E}, 15^{\circ} \sim 30^{\circ} \text{N}, 1^{\circ} \times 1^{\circ}$,共336个格点);②日本细网格降水预报场($100^{\circ} \sim 120^{\circ} \text{E}, 15^{\circ} \sim 30^{\circ} \text{N}, 1.25^{\circ} \times 1.25^{\circ}$,共221个格点)。

对2002年、2003年5~6月数值预报产品场与预报对象进行场相关普查,以成片的高于0.05显著性水平的格点区为基础,在高相关区内选2个相邻格点的平均值作为该相关区的代表值并成为待选因子。另外,在选因子时,对其中部分因子还进行了组合,即对相关而符号相反的两个相邻或相近区域,将这两个区域的代表格点值相减,获得组合预报因子。预选因子时,以达到或超过0.01置信度水平为入选标准,最终得到I区的预报因子46个(T213因子45个,日本降水预报格点因子1个),II区的预报因子42个(T213因子41个,日本降水预报格点因子1个),III区的预报因子42个(T213因子41个,日本降水预报格点因子1个)。其中,数量最多的T213因子包括海平面气压和各标准层物理量,如散度、涡度、比湿、相对湿度、水汽通量散度、水汽通量、假相当位温、涡度平流、垂直速度等。各预报区最终入选的因子,既含有与降水有关的大尺度形势场及能反应降水特征的物理量,也包含有实践中预报结果较好的模式对降水的预报。

2.2.2 EOF方法

神经网络预报模型的建立过程,是通过预报因子输入和期望输出之间对应关系的反复学习和训练,不断调整网络输入层与隐层、隐层与输出层之间的连接权值和阈值,找出网络模型的最佳权值矩阵和阈值向量,从而建立起两者之间映射关系预报模型。理论上,取较多的预报因子及较长的训练数据序列,对模型预报效果会有所提高。但网络结构增大,容易产生过拟合现象,同时也影响网络模型的泛化性能和模型的误差函数收敛速度,导致模型的预报能力下降^[3]。而由于影响天气未来状况的因素较多,必须考虑多种预报因子对预报对象的影响。同时,从统计预报的角度考虑,众多预报因子之间并不是独立的,往往存在复共线性关系,因子各自所带的噪声信息叠加,会造成信息重复、噪声增加,从而影响到预报模型的预报能力。因此,需要构造数量少但能包含多种影响未来天气状况变化的因素,即不仅要求神经网络学习矩阵的输入节点数尽可能少,又要要求输入矩阵含有更多的预报信息。基于

以上分析,本文对预报因子采用 EOF 方法进行降维去噪处理,减少预报因子的数量并尽可能保留其所带的有用信息,以建立网络结构较小的预报模型。自然正交展开法(EOF)是一种有效的压缩空间场资料成为少数几个主要模态的方法,在大气科学中有广泛的应用^[4-6]。EOF 方法的基本思想是将气象要素场序列分解成正交的时间函数与正交的空间函数乘积之和。具体的计算过程见参考文献^[7]。

EOF 方法有以下优点:①EOF 分析后得到的是相互正交的矢量;②各个矢量对原有场的方差贡献可定量地由其特征值表示;③分析后可用数目较少的几个主分量表示出原有场的主要特征,且因为这些量是正交的,不会有多余的重复信息。一方面,神经网络的结构不会因为输入节点的增加而加大,同时,神经网络又不会因为减少输入节点而舍弃大量有用的信息。

表 1 3 个区的神经网络模型预报因子及其相关系数、方差贡献

| 预报对象 | 所选因子 | 相关系数 | 方差贡献率/% |
|-------|----------|---------|---------|
| I 区 | 第 1 主分量 | -0.6864 | 45.89 |
| | 第 2 主分量 | 0.1659 | 8.46 |
| | 第 3 主分量 | 0.2083 | 4.60 |
| | 日本格点降水预报 | 0.6743 | |
| II 区 | 第 1 主分量 | -0.6896 | 43.45 |
| | 第 2 主分量 | 0.3194 | 10.53 |
| | 第 3 主分量 | -0.1586 | 6.47 |
| | 日本格点降水预报 | 0.6338 | |
| III 区 | 第 1 主分量 | -0.6097 | 42.89 |
| | 第 2 主分量 | -0.1449 | 9.90 |
| | 日本格点降水预报 | 0.6464 | |

由上一节广西 3 个区预报量和预报因子相关普查计算结果可以看到,每个预报量与 T213 数值预报产品的相关达到 0.01 置信度以上的预报因子均超过 40 个。对各预报区的 T213 预报因子进行 EOF 分析,并计算每个主成分与预报对象的相关系数。I 区预报对象有 45 个主分量,其中第 1、2、3 主分量与预报对象的相关系数分别为 -0.6864、0.1659、0.2083,这前 3 个主分量共解释了 59.95% 的方差。加上 I 区的日本格点预报因子,最终有 4 个预报因子作为 I 区的建模输入因子。这样, I 区的神经网络模型的输入节点只有 4 个,比原来的 46 个因子少了一个数量级,但仍包含了原有 T213 的

45 个因子和日本格点降水预报因子共 46 个因子的主要信息,既兼顾了神经网络结构不能过大的要求,也考虑了输入信息全面性的需要。与 I 区相似, II、III 区也对各自的 T213 因子作 EOF 分析,计算各主成分与预报量的相关系数,并综合考虑所占方差贡献及相关系数选取因子。各区最终以表 1 所列因子作为神经网络输入因子建立模型。

3 模型试验及结果分析

根据上述计算分析确定的 3 个预报区的神经网络预报模型输入,我们以 2002 年、2003 年 5~6 月的资料作为神经网络模型的训练样本,3 个预报区的训练样本长度分别为 104、106 和 104 天,对 2004 年 5~6 月进行逐日的预报试验,预报天数分别为 55、55 和 56 天,其中未作预报的日期主要是缺数值预报产品资料。同时,为了便于比较,神经网络模型在作每天的预报时,模型参数——学习因子和动量因子统一取 0.9 和 0.7 不变,训练次数也均为 3000 次。在剔除数值预报产品资料不全的日期后,我们作了 2004 年 5~6 月逐日中雨以上降水预报试验,中雨以上降水过程是指预报区域平均 24 h 降水量达到 10 mm 以上,即降水量级大于等于 3 以上。神经网络预报模型的预报输出值大于等于 2.5 为预报一次中雨以上降水。数值预报产品的降水量级预报值则根据以下方法确定:按聚类分析时所确定的预报区域,分别求区域内所有格点的 T213、日本预报的平均降水量,并按相同的降水量分级规则分级,得到 T213、日本对区域降水量级的预报值。对预报产品进行统一的处理后,就能够把神经网络预报与数值预报产品 T213、日本预报按照相同的评分规则进行评定。表 2 为 I 区神经网络模型 2004 年 5 月、6 月逐日中雨以上降水预报及其与 T213、日本预报的比较。

当区域内平均降水量达到 10 mm 以上时(I 区为 41 个站平均, II 区为 23 个站平均, III 区为 25 个站平均),一般为所在区域受到较强天气系统的影响,出现了较明显的降水过程,准确预报区域内中雨以上降水过程对防灾减灾有重要作用。根据表 2,统计 2004 年 5~6 月 I 区逐日预报结果(见表 3)。按 TS(对的次数/(对+空+漏)总次数),计算预报评分, I 区神经网络模型的中雨以上降水预报 TS

表 2 2004 年 5 月、6 月 I 区神经网络模型逐日预报及其中雨以上降水预报与 T213、日本预报比较

| 5 月 | 实况 | 模型 | | T213 | | 日本 | | 6 月 | 实况 | 模型 | | T213 | | 日本 | |
|------|----|------|----|------|----|----|----|------|----|------|----|------|----|----|----|
| | | 预报 | 评定 | 预报 | 评定 | 预报 | 评定 | | | 预报 | 评定 | 预报 | 评定 | 预报 | 评定 |
| 1 日 | 3 | 3.72 | ✓ | 3 | ✓ | 3 | ✓ | 1 日 | 2 | 1.83 | | 2 | | 2 | |
| 2 日 | 2 | 1.58 | | 2 | | 1 | | 2 日 | 2 | 2.08 | | 2 | | 3 | △ |
| 3 日 | 2 | 1.40 | | 2 | | 1 | | 4 日 | 2 | 2.03 | | 4 | △ | 2 | |
| 4 日 | 2 | 1.61 | | 2 | | 2 | | 5 日 | 1 | 1.73 | | 3 | △ | 2 | |
| 5 日 | 3 | 3.64 | ✓ | 3 | ✓ | 3 | ✓ | 6 日 | 1 | 1.62 | | 2 | | 2 | |
| 6 日 | 2 | 1.65 | | 2 | | 2 | | 7 日 | 1 | 1.45 | | 2 | | 1 | |
| 7 日 | 1 | 1.48 | | 1 | | 2 | | 8 日 | 1 | 1.47 | | 2 | | 2 | |
| 8 日 | 1 | 1.61 | | 2 | | 2 | | 9 日 | 2 | 1.48 | | 2 | | 1 | |
| 9 日 | 2 | 2.17 | | 3 | △ | 3 | △ | 10 日 | 2 | 1.51 | | 2 | | 2 | |
| 12 日 | 2 | 2.71 | △ | 3 | △ | 2 | | 11 日 | 2 | 1.58 | | 2 | | 2 | |
| 13 日 | 4 | 2.51 | ✓ | 3 | ✓ | 3 | ✓ | 12 日 | 2 | 1.90 | | 2 | | 2 | |
| 14 日 | 4 | 3.41 | ✓ | 5 | ✓ | 3 | ✓ | 14 日 | 4 | 3.01 | ✓ | 4 | ✓ | 4 | ✓ |
| 15 日 | 2 | 2.47 | | 3 | △ | 2 | | 15 日 | 3 | 2.44 | × | 3 | ✓ | 2 | × |
| 16 日 | 3 | 2.61 | ✓ | 5 | ✓ | 2 | × | 16 日 | 2 | 2.41 | | 3 | △ | 2 | |
| 17 日 | 2 | 2.92 | △ | 4 | △ | 2 | | 17 日 | 3 | 2.77 | ✓ | 3 | ✓ | 3 | ✓ |
| 18 日 | 2 | 1.96 | | 2 | | 2 | | 18 日 | 2 | 3.66 | △ | 4 | △ | 3 | △ |
| 19 日 | 2 | 1.92 | | 1 | | 1 | | 19 日 | 2 | 2.56 | △ | 3 | △ | 3 | △ |
| 20 日 | 1 | 1.44 | | 2 | | 2 | | 20 日 | 2 | 2.43 | | 2 | | 3 | △ |
| 21 日 | 1 | 1.49 | | 2 | | 2 | | 21 日 | 3 | 2.13 | × | 2 | × | 2 | × |
| 22 日 | 1 | 1.54 | | 2 | | 2 | | 22 日 | 3 | 2.40 | × | 2 | × | 3 | ✓ |
| 23 日 | 2 | 1.67 | | 2 | | 2 | | 23 日 | 2 | 2.61 | △ | 3 | △ | 3 | × |
| 25 日 | 2 | 2.07 | | 2 | | 2 | | 24 日 | 2 | 2.29 | | 3 | △ | 2 | |
| 26 日 | 3 | 2.55 | ✓ | 2 | × | 2 | × | 25 日 | 3 | 1.85 | × | 2 | × | 2 | × |
| 27 日 | 3 | 2.66 | ✓ | 3 | ✓ | 2 | × | 26 日 | 2 | 1.58 | | 2 | | 2 | |
| 28 日 | 3 | 2.87 | ✓ | 3 | ✓ | 3 | ✓ | 27 日 | 1 | 1.70 | | 2 | | 2 | |
| 29 日 | 3 | 3.66 | ✓ | 5 | ✓ | 3 | ✓ | 28 日 | 2 | 2.12 | | 2 | | 2 | |
| 30 日 | 2 | 1.86 | | 2 | | 1 | | 29 日 | 2 | 1.64 | | 2 | | 2 | |
| | | | | | | | | 30 日 | 2 | 1.82 | | 2 | | 2 | |

注:表中“✓”表示报对一次,“△”表示空报一次,“×”表示漏报一次。

达到 0.55,比 T213 的 0.44 和日本的 0.45 均高出不少。说明对于 I 区的中雨以上降水,神经网络模型的预报有较好的参考价值。II、III区按与 I 区相同的方法对预报进行评定,结果见表 3。从表 3 的

统计结果中可以看到,II区预报模型的中雨以上降水 TS 也达到 0.5,好于 T213、日本的预报。III区模型的中雨以上降水 TS 则仅为 0.26,但仍稍好于 T213、日本的预报。

表 3 2004 年 5~6 月 3 种方法的中雨以上降水预报统计结果和 TS 评分

| | 神经网络方法 | | | | T213 预报 | | | | 日本降水模式预报 | | | |
|-------|--------|----|---|------|---------|----|---|------|----------|---|---|------|
| | 对 | 空 | 漏 | TS | 对 | 空 | 漏 | TS | 对 | 空 | 漏 | TS |
| I 区 | 11 | 5 | 4 | 0.55 | 11 | 11 | 3 | 0.44 | 9 | 5 | 6 | 0.45 |
| II 区 | 9 | 7 | 2 | 0.50 | 8 | 8 | 3 | 0.42 | 7 | 5 | 4 | 0.43 |
| III 区 | 7 | 18 | 2 | 0.26 | 5 | 11 | 6 | 0.23 | 4 | 8 | 6 | 0.22 |

进一步分析 3 个预报模型的预报效果还发现,III区神经网络模型预报与 T213、日本相比,只是略有提高。同时,由表 3 可以看出,对于 III区的预报,T213、日本和本文提出的预报模型 TS 评分均远不及 I 区、II区的预报效果。本文在做逐日预报试验

时,采用的是动态模型,即做后一次预报时,把前一次的资料加入作为模型的训练样本,因此,每次预报时模型的训练样本数均比前一次预报增加一个训练样本。我们根据 3 个区逐日的拟合平均相对误差值,分析模型的预报性能,见图 3。计算 I、II、III区

模型(I、II区为55次,III区为56次)逐次平均拟合相对误差的平均值分别为:22.12%、19.43%、24.78%,并计算了各模型逐次预报的平均拟合相对误差的方差:0.3569、0.7591、1.4597。由计算结果不难看出,III区模型的平均拟合相对误差的平均值和方差都是3个模型中最大的。从预报效果看,方差最小的I区模型的预报效果最好,方差和平均拟合相对误差最大的III区模型的预报效果最差。并且从图中看到,I、II区模型的平均拟合相对误差曲线并未随预报次数的增加而增大,且波动也没有加大的趋势,相比I、II区模型而言,III区模型的逐日误差曲线随着预报次数增加而有上升的趋势,且曲线的波动也有逐渐增大的趋势,说明III区预报模型随着时间增加预报模型表现不够稳定,因此在实际预报应用时可能要考虑随着时间增长,误差变大,模型有必要重建。另外,III区预报模型的平均拟合相对误差随着预报次数增加而增大,方差也较大的原因,可能与建立模型时,III区是由相似性不太好的邻近站点组成,致使预报区本身一致性较差的“先天缺陷”有关,最终导致日本、T213及本文提出的模式预报效果均不够理想。

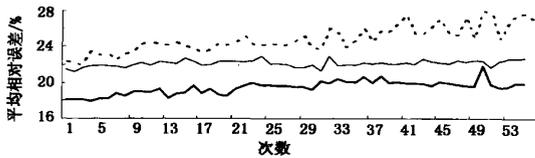


图3 3个神经网络模型的逐日平均拟合相对误差
(细实线为I区,粗实线为II区,虚线为III区)

4 结论

(1)为建立广西逐日降水量级神经网络预报模型,根据各站降水气候特征的相似程度运用聚类分析对广西89个站点分成3个预报区(I区41站,II区23站,III区25站),区域内所有站点的24h平均降水量可以反映较强天气系统的影响,有利于提高区域降水预报准确率。

(2)在构造各预报区的神经网络预报模型学习

矩阵输入时,对用相关普查得到的初选因子中大量的数值预报产品T213因子采用EOF方法进行处理,建立“混合因子场”,提取主分量作为预报因子,浓缩了大量预报因子的有效信息,起到了显著的降维作用,并减少了节点之间重复信息和噪声输入,使神经网络模型输入节点大大减少,有效减少神经网络模型中存在的较多自由参数,使模型的非线性不稳定得到抑制。采用EOF方法对数值预报产品处理并建立神经网络模型,对今后业务中大量的数值预报产品的释用有较好的参考意义。

(3)运用与实际业务预报相同的方法进行逐日预报试验,预报结果表明,所建立的3个神经网络预报模型中的I、II区模型对中雨以上降水过程的预报效果明显好于目前业务中使用的T213、日本模式预报效果,有较好的参考价值,在业务运用中有较好的应用前景。

(4)分析III区神经网络模型预报中雨以上降水过程效果只是略好于T213、日本预报的原因,可能主要与所建模型的平均拟合相对误差随着时间增加有增长的趋势,使后期预报出现较大的波动振荡,以及在建立模型时,预报区内各站点间的降水相似性不是很好有关。如何提高这种区域内站点降水相似性差的预报结果问题,有待于进一步研究。

参考文献

- [1] 孙照渤,谭桂容,赵振国.神经网络方法在夏季降水预报中的应用[J].南京气象学院学报,1998,21(1):47-51.
- [2] Jin Long, Luo Ying, Guo Guang, et al. Study on mixed model of neural network for farmland flood/drought prediction[J]. Acta Meteorologica Sinica, 1997,11(3):364-373.
- [3] 金龙,况雪源,黄海洪,等.神经网络预报模型的过拟合研究[J].气象学报,2004,62(1):62-70.
- [4] 伍红雨,王谦谦.近49年贵州降水异常的气候特征分析[J].高原气象,2003,22(1):65-70.
- [5] 王业宏,金龙.基于自然正交展开的神经网络长期预报模型[J].自然灾害学报,2003,12(2):127-132.
- [6] 李跃清.相空间EOF方法及其在气候诊断中的应用[J].高原气象,2001,20(1):88-93.
- [7] 施能.气象科研与预报中的多元分析方法[M].北京:气象出版社,1995.205-207.

Application of Numerical Forecast Products to Regional Rainfall Forecasting by Artificial Neural Network

Lin Jianling¹ Jin Long² Peng Haiyan³

(1 Nanjing University of Information Science & Technology, Nanjing 210044; 2 Guangxi Research Institute of Meteorological Disasters Mitigation, Nanning 530022; 3 Jiangsu Provincial Meteorological Bureau, Nanjing 210008)

Abstract: The numerical forecast products (T213, Japanese fine grid rainfall forecast, etc.) are used in forecasting by means of the artificial neural network technology. Classification is conducted on all stations over Guangxi by use of Cluster Analysis. A large amount of valuable information about numerous factors is concentrated in terms of the Empirical Orthogonal Function (EOF). Three artificial neural network day-by-day regional rainfall forecast models for May and June in Guangxi Province are established in combination with the rainfall forecast of Japan. A day-by-day forecast test is carried out using the operational method. The TS evaluation of the three models is 0.55, 0.5, 0.26, respectively, for moderate rainfall prediction. These results suggest that the three artificial neural network regional rainfall forecast models perform better than the operationally used numerical forecast products — T213 and Japanese model.

Key words: Artificial Neural Network, EOF, Numerical forecast product, application, graded rainfall forecast