

内蒙古大学
硕士学位论文
数据挖掘技术在天气预报中的应用研究
姓名：李一平
申请学位级别：硕士
专业：计算机应用技术
指导教师：高光来
20030901

数据挖掘技术在天气预报中的应用研究

摘 要

本文是数据挖掘技术在天气预报中有针对性的应用。本文收集、整理并分析了大量降雹、沙尘以及降水天气资料。首先整理分析了近年从巴盟、包头及呼市各防雹办收集的降雹及雹情资料,且对相应的天气状况进行了初步分析;其次总结分析了全区 80 年代以来的沙尘暴资料,按能见度的程度及分布状况对沙尘暴个例进行了分类;另外初步分析了中西部地区各站点 1995 年后 4—8 月的日降水量,按不同量级和分布进行了统计,对其分门别类建立了样本库。在此基础上根据不同的数据挖掘技术的要求,对相应采用的数值预报产品进行客观处理。针对强对流天气(冰雹)将历史样本天气分为西北气流型、槽区型、西风气流型和西南气流型等 4 个类型,并建立其 HLAFS 资料 400hPa 四种特征场,按照模板匹配的原理,用相似分析方法在大、小两个关键区中对实时 HLAFS 预报产品进行计算、比较,再建立预报方程,得出结论;对沙尘暴天气将其按区域分为全区、西部、中部、中西部以及中东部 5 类,再按强弱分别分为强和一般 2 类,建立 ECMWF 的 3 个场(500hPa 高度场、850hPa 温度场和海平面气压场)的历史资料库,在用传真资料消空之后,用相似离度方法计算实时 ECMWF 资料,做出不同时次的预报;对降水天气按出现时间和区域分为适合和不适合飞机增雨作业 2 类,选择 T106 资料中恰当的气象物理量因子,用 BP 神经网络算法建立人工增雨降水预报模型,实际应用中将实时 T213 相关预报资料代入预报模型即可。这些工作表明,在完善气象信息数据的基础上,数据挖掘技术在气象预报,尤其在数值预报产品的释用中将有广阔的前景,在提高灾害性天气预报准确率的进程中,必将发挥愈益明显的作用。

关键词 数据挖掘, 模板匹配, 相似离度, 人工神经网络, 天气预报

APPLICATION STUDY OF DATA MINING ON WEATHER FORECAST

Abstract

In this paper , the data mining techniques have been applied to the weather forecast with a clear aim . The large quantity weather data about hail shooting , duststorm and precipitation have been collected , put in order and analyzed . First not only carefully analyzed the hail event data gathered from the Hail Suppress Offices of Bayan Naoer League , Baotou and Huhhot in recent years , but also preliminarily analyzed the relevant weather patterns . Second comprehensively analyzed the duststorm data since 1980's , classified them according to the visibility and distribution . Besides these , the precipitation data after 1995 within April and August in middle-west of Inner Mongolia were analyzed preliminarily and added up in accordance with different quantity degree and distribution . Futhermore set up the different sample databases . On this foundation , the numerical forecast products have been objective handled according to different require of every data mining technique . For strong convective weather (hail) , it classified all historical sample events into 4 weather patterns (like northwest , trough area , west wind and southwest current) , established 4 characteristic fields of 400hPa height of HLAFS , then according to the principle of the pattern match calculate and compare the real-time HLAFS forecast products

using the similar method inside the big and small key areas ,establish forecast equation , finally gain conclusion .To duststorm weather, the historical samples were sorted into duststorm and severe duststorm types in 5 regions (they are whole area, west, central, middle-west and middle-east of Inner Mongolia),moreover set up different sample databases about ECMWF fields (including 500hPa height, 850hPa temperature and sea-level pressure).In order to making duststorm forecast at different time level ,we first flitered the real-time data by FAX data ,and then used the method of similar range degree to compare the historical data to the actual data of ECMWF .To precipitation weather ,they were divided into 2 types that suit or unsuit airplane artificial precipitation stimulation in line with their emergence time and district .The appropriate weather physical factors come from T106 were chosen to establish the artificial precipitation stimulation prediction model .In the actual application ,we can get prediction result as long as use the real-time forecast data of T213 in the prediction model .These work indicate on the foundation of the data of perfect weather information ,be sure that the data mining techniques will have wide prospect in the weather forecast .By all means it will bring into play more function of obvious benefit in the progress of increasing the prediction accurate rate of disaster weather.

KEYWORDA data mining, pattern match, similarity degree, artificial neural networks, weather forecast

第一章 引言

在过去的 30 年中, 计算机硬件稳定的、令人吃惊的进步导致了功能强大的计算机、数据收集设备和存储介质的大量供应, 这些技术大大推动了数据库和信息产业的发展。然而随之而来的“数据丰富而知识贫乏”的问题却十分严重, 如何找到有关方法, 自动地分析数据、自动地对数据进行分类、自动地对数据汇总、自动地发现和描述数据中的趋势, 自动地标记异常成为必然。数据挖掘——从大量浩如烟海般的数据中用非平凡的方法发现有用的知识——就成了一种自然的需求。

近年来, 随着气象业务现代化建设的快速发展, 数值天气预报业务系统的不断升级, 现代通讯水平的不断提高和完善, 气象工作者可以通过 9210 工程网络获得越来越多的时间和空间分辨率不断升高的数值预报产品, 这些产品信息量十分庞大。相应地, 随着社会的发展, 人类对天气预报也提出了更高的要求。如经济建设中, 大型项目的长周期建设、船舶的洲际远航、商品的季节性调整、季节性疾病的预防、军事上的战略部署和战役准备等等, 需要的是准确的中长期天气预报; 而航空活动和火箭、卫星的发射等, 更需要的是包括云、风、能见度、气温和雷电等强对流天气的详细的中短期和短时预报。又如, 高速公路的发展, 使交通运输部门对大雾天气更为关注; 航天事业的发展, 要求气象工作者把天气预报的范围从地面延伸到中高层大气; 航海事业的发展 and 海底资源的开发, 则使海洋天气预报显得日趋重要。各种数值预报产品为我们提供了多层次、多要素、多时效、多种物理量的丰富资料, 为预报员开展各种释用工作奠定了良好的基础, 便于提高预报员的预报能力及准确率, 因而如何从中挖掘有用的信息就十分迫切。本项工作旨在将数据挖掘的一些分类和预测方法(诸如模板匹配、神经网络等)作为尝试应用于天气预报业务中, 以便挖掘大量气象信息(各种数值预报产品资料)中有用的知识, 进而提高天气预报的准确率, 满足各经济部门和人们生活及社会活动的各种需求, 减少自然灾害造成的损失, 保障社会和人民财产及生命安全。

第二章 数据挖掘概述

2.1 数据挖掘的重要性

需要是发明之母。

在过去的数十年中，我们产生和收集数据的能力已经迅速提高。起作用的因素包括条码在大部分商业产品中的广泛使用，许多商务、科学和行政事务的计算机化，以及由文本和图象扫描平台到卫星遥感系统的数据收集工具的进步。此外，作为全球信息系统的万维网的流行，已经将我们淹没在数据和信息的汪洋大海中。存储数据的爆炸性增长业已激起对新技术和自动工具的需求，以便帮助我们将海量数据转换成信息和知识。

近年来，数据挖掘引起了信息产业界的极大关注，其主要原因是存在大量数据，可以广泛使用，并且迫切需要将数据转换成有用的信息和知识。获取的知识可以广泛用于各种应用，包括商务管理、生产控制、市场分析、工程设计和科学探索等。

数据挖掘是信息技术自然演化的结果。现在，在一个大的机构（例如企业、医院、信用卡公司等）应用某些数据库技术时存在着一个趋势，它包含着“数据仓库”的概念，其中一个经典的数据搜索技术被称为“数据挖掘”。一个数据仓库就是一个包含着很多数据表的数据库系统，数据表中的内容会周期性的更新，它们包含着很多细节的历史信息，支持对于更高级数据的描述，还有很多的总结性工具，比如元数据等，也就是说，一些对于系统元素（例如名称、定义、结构等）的定位和描述的数据。

数据的丰富带来了对强有力的数据分析工具的需求，大量的数据被描述为“数据丰富，但信息贫乏”。快速增长的海量数据收集、存放在大型和大量数据库中，没有强有力的工具，理解它们已经远远超出了人的能力。结果，收集在大型数据库中的数据变成了“数据坟墓”——难得再访问的数据档案。数据挖掘工具进行数据分析，可以发现重要的数据模式，数据和信息之间的鸿沟要求系统地开发数据挖掘工具，将数据坟墓转换成知识“金块”。

2.2 什么是数据挖掘

简单地说，数据挖掘是从大量数据中提取或“挖掘”知识。数据挖掘是数据库系统和新的数据库应用的一个有希望的、欣欣向荣的学科前沿。数据挖掘通常又称数据库中知识发现（KDD），是自动的或方便的模式提取，这些模式代表隐藏在大型数据库、数据仓库或其他大量信息存储中的知识。而另一些人只是把数据挖掘视为数据库中的知识发现过程的一个基本步骤。我们同意这一看法，数据挖掘只是整个过程中的一步，尽管是最重要的一步，因为它发现隐藏的模式。知识发现过程由以下步骤组成：

数据清理（消除噪声或不一致数据）

数据集成（多种数据源可以组合在一起）

数据选择（从数据库中检索与分析任务相关的数据）

数据变换（数据变换或统一成适合挖掘的形式，如通过汇总或聚集操作）

数据挖掘（基本步骤，使用智能方法提取数据模式）

模式评估（根据某种兴趣度度量，识别表示知识的真正有趣的模式）

知识表示（使用可视化和知识表示技术，向用户提供挖掘的知识）

数据挖掘是从存放在数据库、数据仓库或其他信息库中的大量数据中挖掘有趣知识的过程。基于这种观点，典型的数据挖掘系统具有以下主要成分：

数据库、数据仓库或其他信息库：这是一个或一组数据库、数据仓库、电子表格或其他类型的信息库。可以在数据上进行数据清理和集成。

数据库或数据仓库服务器：根据用户的数据挖掘请求，数据库或数据仓库服务器负责提取相关数据。

知识库：这是领域知识，用于指导搜索，或评估结果模式的兴趣度。

数据挖掘引擎：这是数据挖掘系统基本的部分，由一组功能模块组成，用于特征化、关联、分类、聚类分析以及演变和偏差分析。

模式评估模块：通常，此成分使用兴趣度度量，并与数据挖掘模块交互，以便将搜索聚焦在有趣的模式上。

图形用户界面：模块在用户和数据挖掘过程系统之间通信，允许用户与系统交互，指定数据挖掘查询或任务，提供信息、帮助搜索聚焦，根据数据挖掘的中间结果进行探索式数据挖掘。

数据挖掘是一个交叉学科领域，它涉及多学科技术的集成，包括数据库技术、统计学、机器学习、高性能计算、模式识别、神经网络、数据可视化、信息检索、图象与信号处理和空间数据分析。

2.3 数据挖掘的数据

原则上讲，数据挖掘可以在任何类型的信息存储上进行。这包括关系数据库、数据仓库、事务数据库、高级数据库系统、展开文件和 WWW。高级数据库系统包括面向对象和对象——关系数据库；面向特殊应用的数据库，如空间数据库、时间序列数据库、文本数据库和多媒体数据库。挖掘的挑战和技术可能因存储系统而异。

2.4 数据挖掘的功能

数据挖掘功能用于指定数据挖掘任务中要找的模式类型。数据挖掘任务一般可以分两

类：描述和预测。描述性挖掘任务刻画数据库中数据的一般特性。预测性挖掘任务在当前数据上进行推断，以进行预测。数据挖掘系统要能够挖掘多种类型的模式，以适应不同的用户需求或不同的应用。数据挖掘功能以及它们可以发现的模式类型有：概念/类描述；特征化和区分；关联分析；分类和预测；聚类分析；孤立点分析；演变分析等。

数据挖掘技巧被用来从数据仓库中提取相关的信息，它被应用于各种不同的领域，例如：工程领域，例如设备故障预测，网络上的搜索引擎等；经济领域，例如对于某个投资的利润预测，检测一个消费线，评价一个贷款风险等；生物及医学领域，例如染色体中的蛋白质序列匹配，对于怀孕的几率的估计等。

这些技术都用到了模式识别中的一些方法，例如聚类分析、统计分析以及神经网络等。此外还应用了人工智能领域中的一些方法，例如知识表达、因果模型以及规则演绎等。

2.5 数据挖掘的发展趋势

鉴于数据、数据挖掘任务和数据挖掘方法的多样性，给数据挖掘提出了许多挑战性的课题。数据挖掘语言的设计，高效而有用的数据挖掘方法和系统的开发，交互和集成的数据挖掘环境的建立，以及应用数据挖掘技术解决大型应用问题，都是目前数据挖掘研究人员、系统和应用开发人员所面临的主要问题。其发展的一些趋势有：应用的探索；可伸缩的数据挖掘方法；数据挖掘与数据库系统、数据仓库系统和 Web 数据库系统的集成；数据挖掘语言的标准化；可视化数据挖掘；复杂数据类型挖掘的新方法；Web 挖掘；数据挖掘中的隐私保护与信息安全。

2.6 数据挖掘在气象应用中的必要性

随着气象信息数量和种类的迅速增长，如何存储、管理和检索利用信息就成为气象业务中面临的第一问题。如何从众多的信息中将有用的部分提取出来，并按照业务要求进行组织是需要解决的又一个问题。如何有效地综合运用包括数值预报、天气图、卫星、雷达等多方面的信息、知识，提高计算机对气象信息高度综合处理能力，进一步缩小计算机处理于专家知识二者之间的距离，拓宽计算机在气象中的应用领域，加速天气预报和气象业务现代化，是当前迫切需要解决的第三个问题。

从信息处理角度，人们更希望计算机帮助我们分析数据、理解数据，帮助我们基于丰富的数据做出决策，做人力所不能及的事情。数据挖掘技术就是利用计算机从大量数据中提取或“挖掘”此前并不知道的对有价值的知识。随着数值预报的发展，数值预报产品的应用已成为广大气象台站建立现代天气预报业务、提高天气预报精度的重要内容。如何采用各种释用

技术方法来充分挖掘数值预报产品中的有用信息,则是提高天气预报水平,尤其是灾害性、关键性天气预报准确率的重要途径。因而,尽快将数据挖掘技术应用于天气预报,以进一步提高预报准确率,满足社会各方面的需求是十分必要的。

第三章 模板匹配在冰雹预报中的应用

3.1 开展冰雹预报的背景及意义

冰雹是内蒙古中西部地区夏半年极具破坏性的天气之一,对农牧业、尤其对农业生产具有很大的危害,它是制约我区农牧业经济乃至国民经济的一个根本性因素。冰雹虽然受灾范围较小,但是突发性强且往往伴随强风的袭击,受灾的农作物轻者要减产、严重的会颗粒无收。据统计,我区平均年受雹灾农田面积约 130 万亩,最多可达 300 余万亩。内蒙古开展人工防雹始于七十年代初期,巴盟、乌盟农民在当地气象部门指导下首先搞起了土炮,此后包头、呼市、赤峰等地纷纷开展。80 年代初,甘肃省采用 37 高炮防雹,很快各省区也纷纷淘汰土炮,37 高炮迅速发展推广。90 年代初我区 37 高炮 150 门,发展至今已达 733 门。各盟市建立了指挥中心,九部雷达用于探测雹云指挥作业。90 年代我区受雹灾农田面积比防雹前大幅度减少,约 80 万亩。但目前由于冰雹预报尚未能在人工防雹作业指挥中发挥作用,所以实际防雹作业往往被动,时效难以保证,冰雹预报已是人工防雹工作中气象服务的一项迫切需要。

冰雹预报是一个难度很大的课题。多年来,各地许多科研人员对冰雹的形成、演变、降雹特点、天气形势、雷达冰雹云的特征等都作了大量不同程度的研究,而且还建立了一些模式。但因冰雹云发展和冰雹增长的因子非常复杂,又受地形和天气系统的制约,因此目前国内外还没有一种有效的预报方法能在短时间内预报出降雹区和强度。现在国外是用先进的探测通讯设备跟踪监视雹暴,做短时外推预报。国内的大量预报工作其着重点是利用常规天气图和单站探空资料及相关指标预报单站的冰雹情况。

如何采用各种释用技术方法来充分挖掘数值产品中的有用信息,将数值预报产品应用于冰雹预报中是各地势在必行的工作。我们的工作就是要建立一个以国家气象中心新的有限区域业务数值预报系统(HLAFS)的分析预报产品为基础,利用模板匹配的方法能在 Micaps 工作平台上完成资料自动采集、处理、分析、显示的客观化、量化的冰雹预报指导产品,形成一个具有实用价值的业务系统,增加人工防雹的科技含量。

3.2 使用资料和定义

3.2.1 使用资料

3.2.1.1 数值分析预报产品资料

国家气象中心新的有限区域业务数值预报系统(HLAFS)有以下特点:(1) 预报模式水平分辨率为 $1^{\circ} \times 1^{\circ}$ (目前已达到 $0.5^{\circ} \times 0.5^{\circ}$), 预报时次为每隔6小时一次。(2) 采用了四维同化技术, 使分析场的质量有了明显提高。(3) 初值化方案为非绝热非线性方案。(4) 模式的物理过程中引入了较为复杂的描写更为合理的参数化方案(如大尺度凝结、积云对流参数化、边界层过程、地面土壤过程、地面辐射能量收支等过程), 使得模式中对大气状况的描述更接近实际大气。(5) 分析的初始场及预报模式的侧边界统计均由T106中期预报模式提供。考虑到上述因素以及本课题的研究对象为冰雹天气过程, 故决定采用国家气象中心每日08时的HLAFS物理量分析场产品资料进行开发研究, 具体预报则要使用前一日20时HLAFS的12小时预报场。同时考虑到内蒙古中西部的地理位置及实际预报对物理量预报场的需要范围, 将使用资料的范围确定为($35^{\circ} \sim 55^{\circ} \text{ N}$, $95^{\circ} \sim 130^{\circ} \text{ E}$)。

3.2.1.2 降雹资料

使用从巴盟、包头及呼市各防雷办收集的降雹及雹情资料中获取的对应HLAFS资料时段的降雹资料。这些资料无论从数量上, 还是质量上均是最齐全的。

3.2.2 定义和分类

从资料齐备的角度出发, 1996年6月1日到8月15日及1998年6月1日到8月31日期间, 影响内蒙古中西部地区的降雹日共有67个。为了能有一定数量的样本, 定义降雹日为: 凡巴盟、包头及呼市防雷办所辖炮点有一处降雹即为一个降雹日。

由于冰雹预报的多样性和复杂性, 故对其发生的天气形势进行分类后再做预报研究是解决这一复杂问题的有效手段, 也是降低预报难度的主要途径。通过对67个雹日400hPa高度场进行全面仔细的分析, 将降雹天气分为西北气流型、槽区型、西风气流型和西南气流型等四个类型, 其中西北气流型有15个, 槽区型有22个, 西风气流型有16个, 西南气流有20个。

3.3 模板匹配简述

模式识别是一门研究对象描述和分类方法的学科。模板匹配实际上也属模式识别一类, 其目的是通过已建好的模板对事物进行分类。具体步骤先将某一类事物的历史样本进行分类, 建立各自的模板, 在此基础上, 对一个待分类的样本与各个模板进行比较, 求出最相

似的，就把当前的样本归为最相似的那一类。

3.3.1 模板建立及天气分型说明

对 67 个雹日中不同类型降雹天气的每日 08 时的 HLAFS 400 hPa 形势场进行合成平均，建立四种降雹天气类型的 400hPa 特征场。

西北气流型（见图 3.1）：其流型特点是主体槽已过河套地区，到达 120°E 左右，河套受槽后西北风气流控制，同时上游有弱冷空气沿槽后西北气流滑过河套地区，易产生冰雹天气。

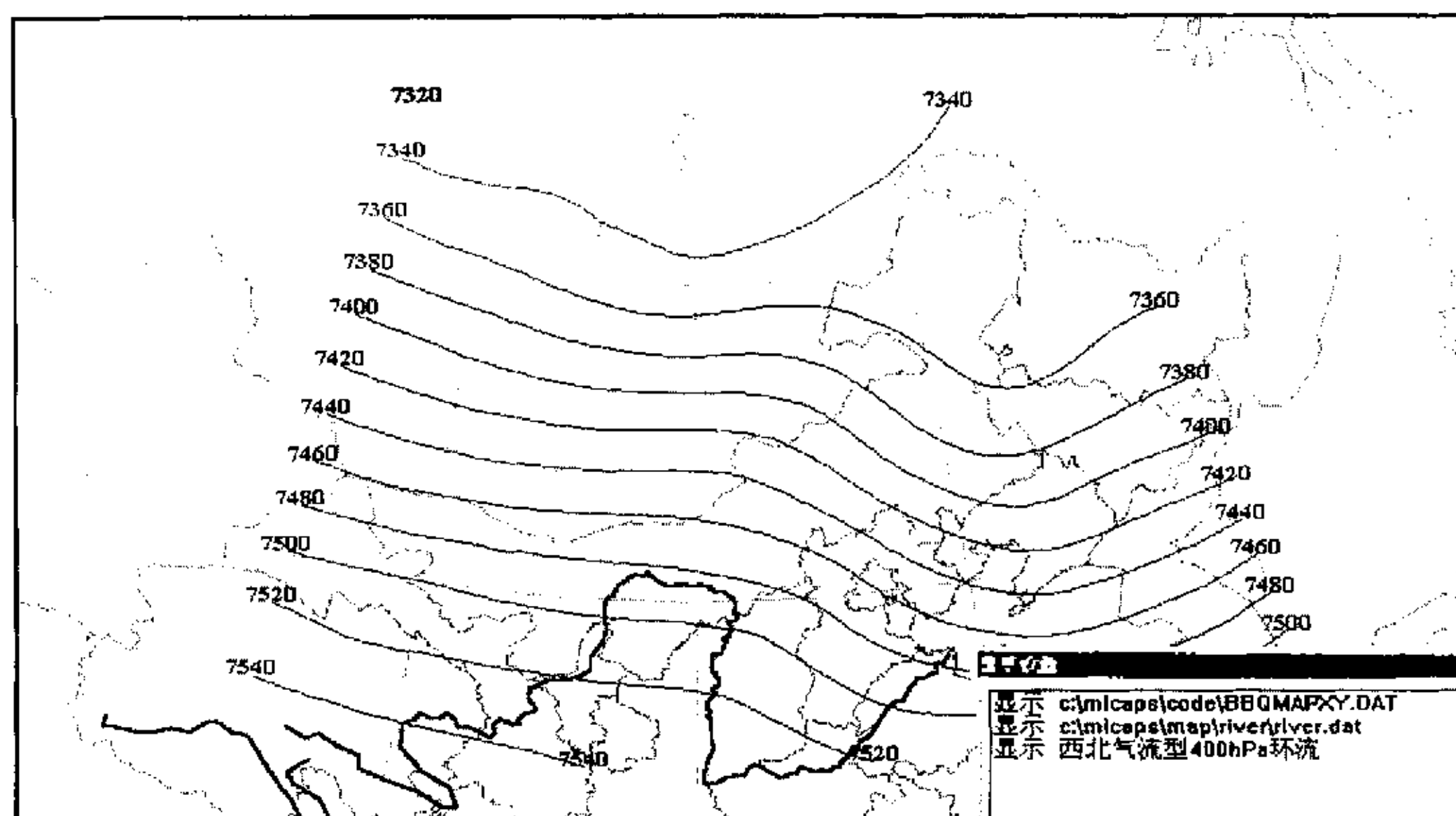


图 3.1 西北气流型 400hPa 特征场

槽区型（见图 3.2）：其流型特点是主体槽区或冷涡位于河套北部或略偏东，其南槽区正在河套上，有冷平流和正涡度平流。如河套地区有不稳定条件，则易产生冰雹天气。

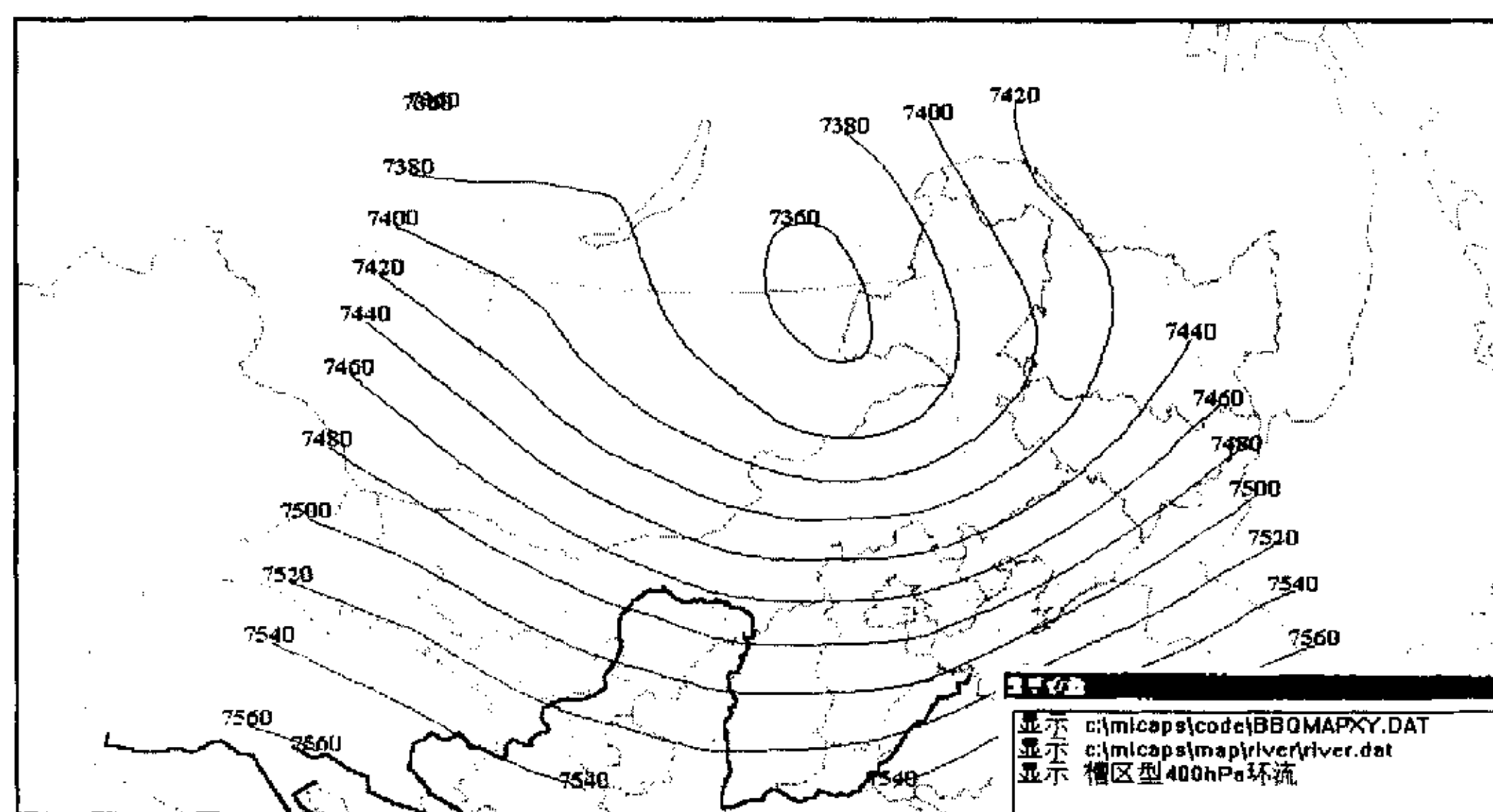


图 3.2 槽区型 400hPa 特征场

西风气流型（见图 3.3）：其流型特点是主体槽弱且偏西，400hPa 河套地区以西风气流为主，其上有小波动，另外河套地区气流有辐散特征，低层多为西南气流，可产生降雪条件。

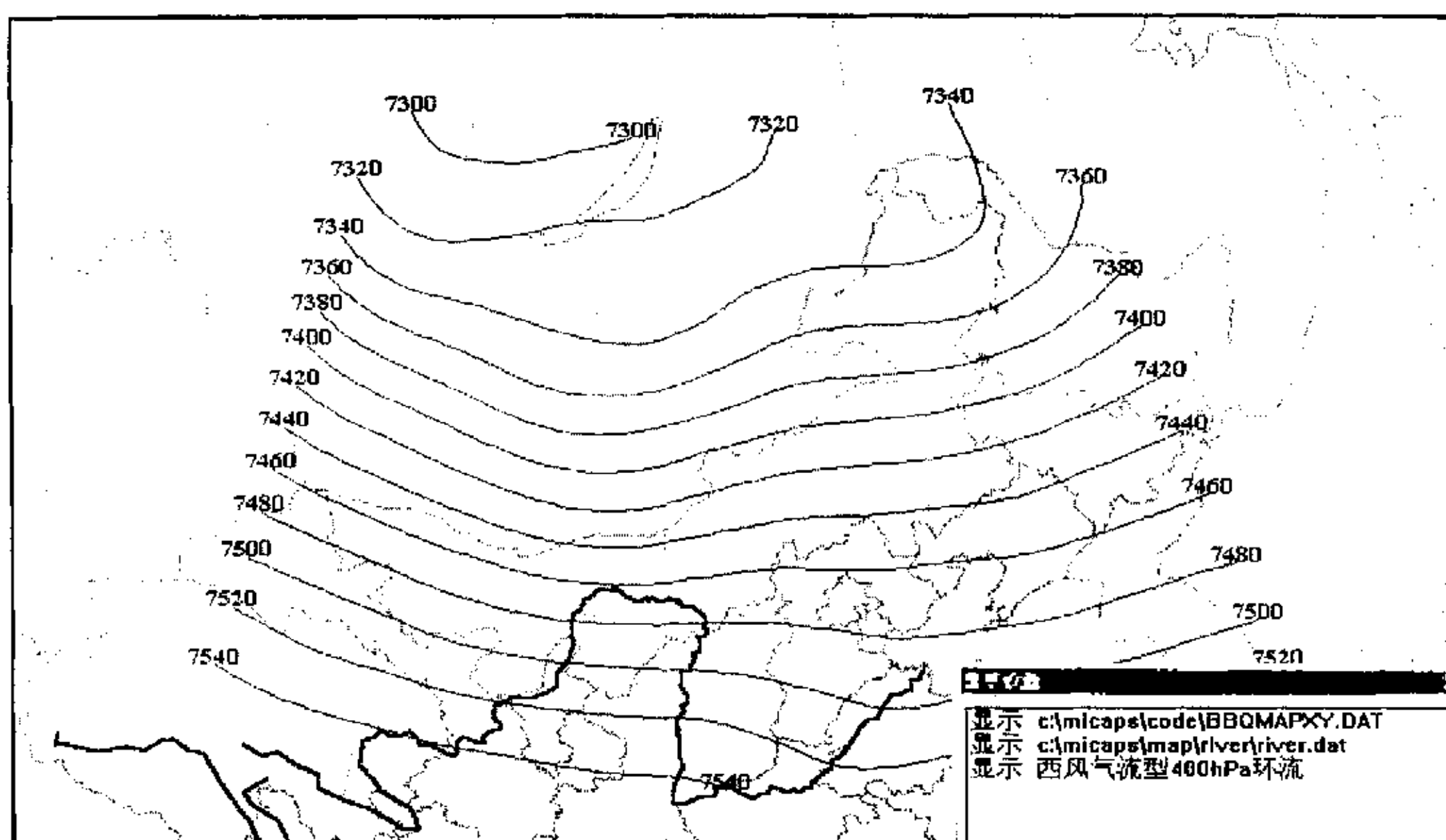


图 3.3 西风气流型 400hPa 特征场

西南气流型（见图 3.4）：其流型特点是主体槽强且位于河套西部 100° E 以西，河套地区受西南气流控制，低层多偏南和东南风。在有不稳定或低层湿度平流明显时，易产生冰雹天气，且雹灾较严重，多为暖切变影响。

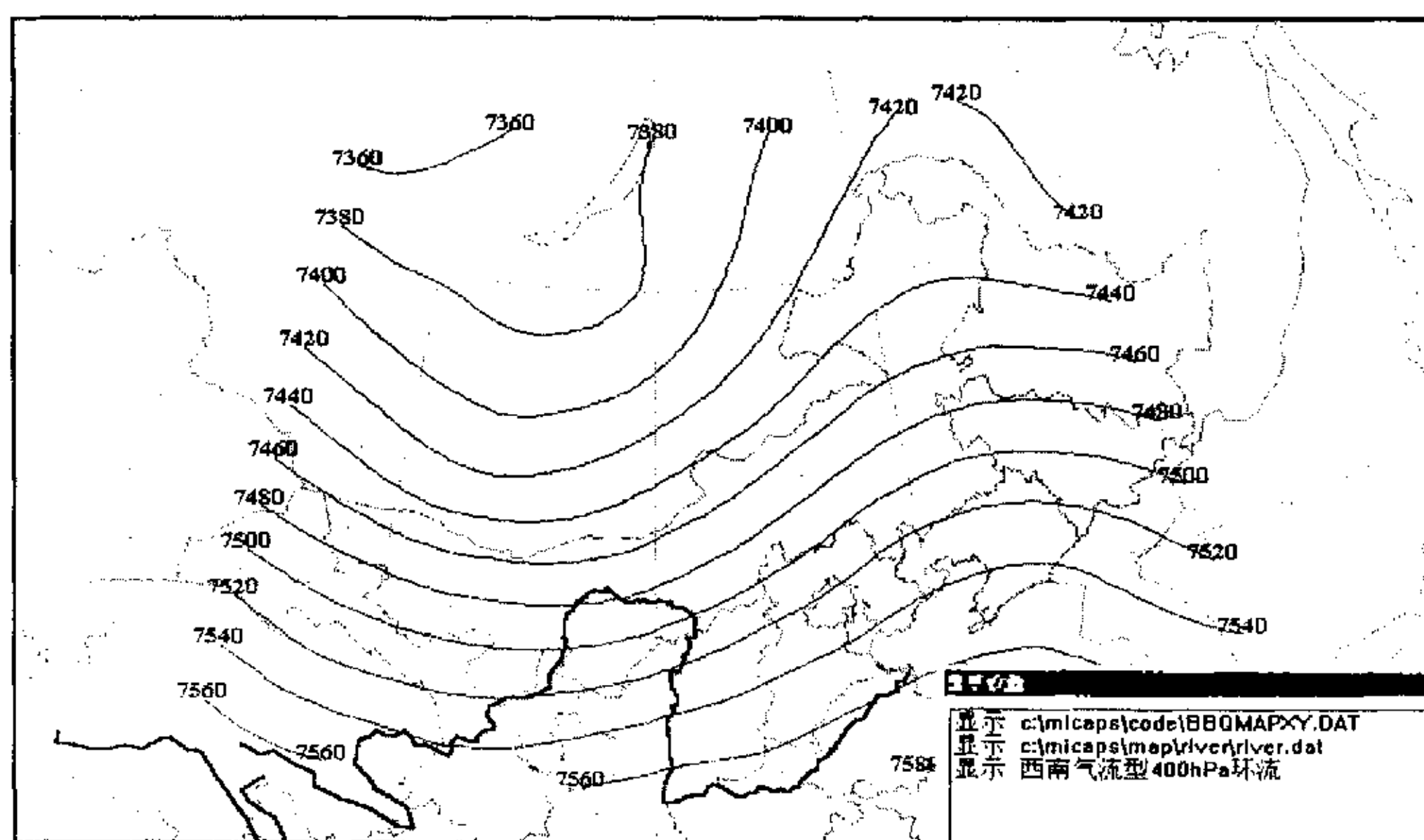


图 3.4 西南气流型 400hPa 特征场

对各类天气类型，500hPa、700hPa 和 850hPa 三层的特征也都很明显（图略）。

3.3.2 相似模型

对两个网格资料数据, n 为计算区域内数据的列数, m 为计算区域内数据的行数。某一型历史特征场矩阵形式为:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \quad (3.1)$$

实时资料矩阵形式为:

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix} \quad (3.2)$$

对两场求平均:

$$\overline{X} = \frac{1}{m \cdot n} \sum_{i=1}^n \sum_{j=1}^m x_{ij} \quad (3.3)$$

$$\overline{Y} = \frac{1}{m \cdot n} \sum_{i=1}^n \sum_{j=1}^m y_{ij} \quad (3.4)$$

求基差:

$$\Delta H = \overline{X} - \overline{Y} \quad (3.5)$$

特征场与其平均值的海明距离:

$$D_x = \sum_{i=1}^n \sum_{j=1}^m |x_{ij} - \overline{X}| \quad (3.6)$$

实时场与特征场的海明距离 (基差订正后)

$$D_y = \sum_{i=1}^n \sum_{j=1}^m |x_{ij} - y_{ij} - \Delta H| \quad (3.7)$$

定义场的相似度为:

$$R_x = \frac{D_x - D_y}{D_x} = 1 - \frac{D_y}{D_x} \quad (3.8)$$

当 $R_i > \delta$ 时, 为正相似;

当 $R_i < -\delta$ 时, 为负相似 (或反相似);

当 $-\delta \leq R_i \leq \delta$ 时, 为不相似。

其中 δ 为一适当的值, 如 0.1、0.05 等, 在本研究中取 0.1。对我们的工作只有正相似才有意义。

3.3.3 相似场的确定

对每日 20 时 HLAFS 资料 12 小时预报的或 08 时实况 400hPa 高度场选取大、小两个关键区与相应 400hPa 特征场 (模板) 进行相似判定。大区 ($50^\circ \sim 35^\circ \text{N}$, $100^\circ \sim 125^\circ \text{E}$), 相似判定大形势是否入选某种天气类型; 小区 ($44^\circ \sim 38^\circ \text{N}$, $105^\circ \sim 115^\circ \text{E}$), 相似判定河套地区上空的流型是否符合要求。分型如有差异, 则进行综合判定 (相似度作平均), 大者入选。

3.4 预报方程的建立

在了解本地区强对流发生、发展的大尺度特征的基础上, 结合强冰雹天气动力学特征, 通过大量的计算、普查、对照, 筛选出几个有代表性的物理量, 进行具体的统计和分类分析。在天气分型的基础上, 选择不同关键区对 HLAFS 资料自带的 15 个量和利用其资料新计算的 11 个量进行计算。通过分析发现, 各物理量在不同的天气类型下作用不同, 而且对有多层值的某些量其多层值之和变化明显, 故均取三层 (500 hPa、700 hPa、850 hPa) 之和 (因为日常工作中大多数量只能接收三层资料) 作为一个预报因子。通过对不同关键区各物理量取值的变化, 确定出主关键区 ($43^\circ \sim 39^\circ \text{N}$, $100^\circ \sim 115^\circ \text{E}$), 计算各物理量在不同天气类型下的合成平均, 选择代表性强、在不同关键区变化一致且与无雹天气差异大的物理量分别构建各种天气类型的预报方程。针对不同的天气类型通过方程做出有无降雹及预报因子符合条件多寡的落区预报。

3.4.1 预报因子说明

在研究过程中我们普查了 HLAFS 所提供的 15 个物理量 (包括: 高度、东西风、南北风、温度、相对湿度、涡度、散度、垂直速度、全风速、水汽通量散度、温度露点差、假相当位温、K 指数、海平面气压、相对辐散)。考虑到冰雹形成的条件和天气形势背景, 使用 HLAFS 资料又计算了对流有效位能 CAPE、温度平流、涡度平流、简化沙氏指数以及对流性稳定度 ($\theta_{se850} - \theta_{se500}$) 等, 说明一下由于资料本身的 K 指数分析结果与其意义相反 (可能有误), 又重新计算了 K 指数。我们选取了纬度不变 ($43^\circ \sim 39^\circ \text{N}$), 随经度变化的七个关键区。考虑到系统的移动, 物理量一般东西向差异大于南北向, 只选取了经度不变 ($105^\circ \sim 112^\circ \text{E}$),

随纬度变化的两个关键区。在关键区中对以上各物理量 500hPa、700hPa、850hPa 三层及三层之和在不同天气类型中进行平均,按纬度不变经度变化或经度不变纬度变化作图分析(图略)。分析表明,在纬度变化的情况下,绝大多数物理量的数值变化很小;而在经度变化的情况下,大部分物理量在不同天气类型下虽然值有差异,但变化趋势均一致。在以上分析的基础上,我们确定主关键区为 ($43^{\circ} \sim 39^{\circ} \text{ N}$, $105^{\circ} \sim 112^{\circ} \text{ E}$, 格点数为 5×8 , 共 40 个格点), 针对各物理量在不同的天气分型下作用的不同, 我们选择了代表性强、在不同关键区变化一致、与无雹天气差异大的物理量分别构建各种天气类型的预报方程。表 3.1 给出了所选物理量在不同天气类型下主关键区中的平均值。

表 3.1 关键区中预报因子在不同天气类型下的平均值

关键区为 ($43 \sim 39^{\circ} \text{ N}$, $105 \sim 112^{\circ} \text{ E}$)

天气类型 预报因子	西北气流 型	槽区型	西风气流 型	西南气流 型	无雹型
x_1 : du	98.9	139.9	195.9	154.3	131.5
x_2 : dv	44.6	-98.5	0.1	16.8	-73.1
x_3 : t0	203.1	144.5	183.4	184.8	162.9
x_4 : rh	1571.4	1730.4	1690.2	2029.7	1571.2
x_5 : di	-4.1	3.9	-3.6	-2.6	4.3
x_6 : vv	223.3	230.7	249	211.9	233.3
x_7 : ra	-30.1	38	-25.9	-18.1	40
x_8 : vo500	-19.9	6.1	-0.2	11.7	-6.5
x_9 : wp500	-38.9	45.9	-18.5	-26.6	44.4
x_{10} : th700	109.8	80.5	96.2	58.4	103.1
x_{11} : th850	98.1	72.2	75.4	44.4	83.3
x_{12} : tb58	2.8	6	5.6	6.7	2.9
x_{13} : cape	173.8	428.4	479.8	491.7	183.4
x_{14} : ssi	14.4	13.8	14.6	16.4	15.5
x_{15} : kk	26.1	30.4	28.7	33.3	25.4

表 3.1 中各预报因子分别代表: du、dv、to、rh、di、vv 和 ra 表示风场 U 分量、V 分量、温度、相对湿度、散度及水汽通量散度 500hPa 到 850hPa 三层之和, vo500 为 500hPa 涡度, wp500 为 500hPa 垂直速度, th700 和 th850 分别为 700hPa 和 850hPa 的温度露点差, tb58

为 $(\theta_{se850} - \theta_{se500})$, kk 为计算的 K 指数, $cape$ 为对流有效位能, ssi 为简化沙氏指数。

3.4.2 各类型预报方程的建立

3.4.2.1 四种类型天气各预报因子的阈值

分析表 3.1, 综合考虑降雹天气产生要求的水汽、不稳定条件及动力触发等因素, 以及各因子在不同天气类型中所起的作用的不同, 结合各因子在不同天气类型中阈值的差异, 对各种天气类型选取预报因子, 具体因子及相应阈值见表 3.2。

表 3.2 各预报因子在四种类型天气中的阈值

天气类型 预报因子	西北气流型	槽区型	西风气流型	西南气流型
x_1 : du	115.2		163.7	
x_2 : dv	-14.25	-85.8		-28.15
x_3 : $t0$	183	153.7		
x_4 : rh				1800.45
x_5 : di			0.35	
x_6 : vv			241.15	
x_7 : ra			-0.15	
x_8 : $vo500$		-0.2		2.6
x_9 : $wp500$	2.75			
x_{10} : $th700$		91.8		
x_{11} : $th850$	90.7			63.85
x_{12} : $tb58$			4.25	
x_{13} : $cape$		305.9	331.6	337.55
x_{14} : ssi	14.95	14.65		
x_{15} : kk				29.35

3.4.2.2 四种类型的预报方程

从表 3.2 中看出, 我们对四种天气类型分别选取了 6 个预报因子。具体对西北气流型选取的为 x_1 、 x_2 、 x_3 、 x_9 、 x_{11} 、 x_{14} , 对槽区型选取的为 x_2 、 x_3 、 x_8 、 x_{10} 、 x_{13} 、 x_{14} , 对西风气流型选取的为 x_1 、 x_5 、 x_6 、 x_7 、 x_{12} 、 x_{13} , 对西南气流型选取的为 x_2 、 x_4 、 x_6 、 x_{11} 、 x_{13} 、 x_{15} 。分别建立预报方程如下:

西北气流型: $Y_1 = x_1 + x_2 + x_3 + x_9 + x_{11} + x_{14}$

槽区型: $Y_2 = X_2 + X_3 + X_8 + X_{10} + X_{13} + X_{14}$

西风气流型: $Y_3 = X_1 + X_5 + X_6 + X_7 + X_{12} + X_{13}$

西南气流型: $Y_4 = X_2 + X_4 + X_8 + X_{11} + X_{13} + X_{15}$

再根据各预报因子相对无雹天气类型变化的趋势,我们给出各预报因子的 0、1 化规定(见表 3.3)。

表 3.3 各预报因子的 0、1 化规定

天气类型 预报因子	西北气流型 取值为 1	槽区型 取值为 1	西风气流型 取值为 1	西南气流型 取值为 1
x_1 : du	<115.2		>163.7	
x_2 : dv	>-14.25	<-85.8		>-28.15
x_3 : t0	>183	<153.7		
x_4 : rh				>1800.45
x_5 : di			<0.35	
x_6 : vv			>241.15	
x_7 : ra			<7.05	
x_8 : vo500		>-0.2		>2.6
x_9 : wp500	<2.75			
x_{10} : th700		<91.8		
x_{11} : th850	>90.7			<63.85
x_{12} : tb58			>4.25	
x_{13} : cape		>305.9	>331.6	>337.55
x_{14} : ssi	<14.95	<14.65		
x_{15} : kk				>29.35

当 $Y_1 \geq 3$ 时, 预报有降雹, 反之无降雹。

当 $Y_2 \geq 3$ 时, 预报有降雹, 反之无降雹。

当 $Y_3 \geq 3$ 时, 预报有降雹, 反之无降雹。

当 $Y_4 \geq 3$ 时, 预报有降雹, 反之无降雹。

在落区预报结果中通过等值线显示, 每满足一个因子加画一条线。

3.4.3 预报流程

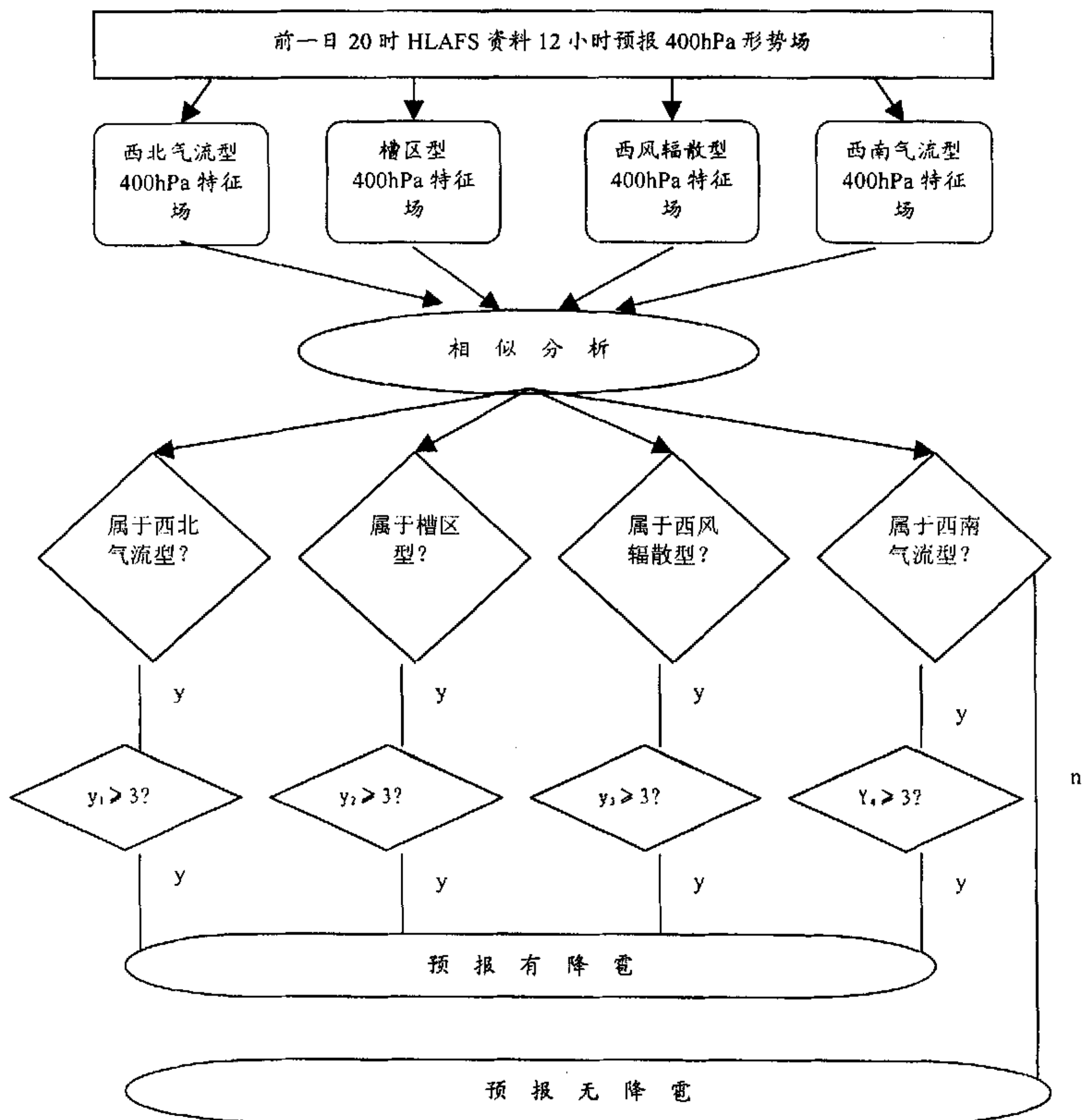


图 3.5 预报简易流程图

从图 3.5 可以了解其预报流程：先用前一日 20 时 HLAFS 资料 12 小时预报的 400 hPa 形势场与 400hPa 特征场做相似分析，判断天气类型，进行消空处理，再根据不同天气类型套用各自的预报方程，得出有无降雹的预报，具体落区通过等值线显示。

3.5 应用情况分析

作为冰雹预报预报工具之一，本方法在近 3 年的呼市地区防雹工作中起到预期的效果，基本可提前预报出对流天气的出现，为雷达指挥防雹提出了警示，使各部门人员都能及时作好准备，适时采取措施，将灾害损失降低到最低程度，收到了一定的经济和社会效益。当然使用数值预报产品或天气资料制作冰雹预报，只是人工防雹的第一步，具体准确的预报和指

挥防雹必须配合卫星云图、雷达、闪电定位仪等先进的探测手段方更准确、合理。

第四章 相似离度方法在沙尘暴预报中的应用

4.1 沙尘暴定义及其预报的必要性

从本质上来说,沙尘(暴)是发生在大陆干旱与半干旱地区的一种自然现象,古已有之,但近年来人类活动使这一现象更为加剧,并给人类生存环境带来巨大影响。沙尘暴发生需三个条件:地表丰富的松散干燥的沙尘、足够强劲持久的风力和强烈的大气垂直不稳定。根据气象观测规范定义,扬沙是指由于风力较大,将地面沙尘吹起,使空气相当混浊,水平能见度在 1—10 公里以内的天气现象;而沙尘暴是指强风将地面沙尘吹起,使空气特别混浊,水平能见度小于 1 公里的天气现象。

沙尘(暴)天气是我国北方(特别是西北地区)春季较为常见的一种灾害性天气现象,可对人类健康及社会经济造成一定的影响。关于扬沙和沙尘暴已有大量的观测研究,但大多数研究仅局限于沙尘暴成因和气候特征分析,比较分散、孤立,对沙尘暴的预报方法却很少,只限于遥感监测。目前亚洲沙尘暴已成为一种国际事件,引起广泛关注,而内蒙古中西部为我国沙尘暴两个多发高频中心之一,因而做好沙尘暴的预报是十分必要的。我们的工作就是利用相似离度方法建立我区沙尘暴预报业务系统,对我区沙尘暴天气预报进行一种新尝试。

4.2 预报方法—相似离度分析

在人工智能中,把应用两个或两类事物原理作预报的思维过程称之为类比推理。其基本思想是:当两个或两类事物在许多属性上都相同,可断定它们在其他一些属性上也相同。若用 A 和 B 分别代表两个或两类不同的事物,用 $a_1, a_2, \dots, a_n$, b 分别代表不同的属性,则类比推理可表示为:

A 与 B 有共同属性 $a_1, a_2, \dots, a_n$; A 有属性 b, 则 B 也有属性 b。

把类比推理的思想移植到计算机上,就可实现相似预报的人工智能化,在具体实现中相似准则的选择是其关键技术之一。

基于智能聚类的相似预报方法是表示物理量格点场之间的相似,一要考虑两个场之间数值的差异,另一方面要考虑格点场分析出的等值线形状之间的差异。相似预报方法是广大预报员熟悉的常用的预报方法之一。它所依据的相似原理认为:相似的天气形势反映了相似的物理过程,因而会出现相似的天气现象。预报员就是根据这一原理,把当前的天气形势与过

去曾出现过的历史个例进行比较, 从中找出最相似的某个个例, 然后把该个例相对应的天气情况作为当前预报的结论。

气象工作者把这种人工相似预报方法移植到计算机上来实现, 开发了不少有实用价值的系统, 在预报业务中发挥了重要作用, 但也存在某些不足之处: 有的描述相似的方法不够客观、全面。有了数值预报产品, 可以用预报的形势场到历史资料中找出相似个例或相似模型, 则该相似个例或相似模型对应出现的天气, 就是我们要预报的结论。可见, 应用了数值预报产品, 使我们把衡量天气形势和天气过程相似的标准, 从前期和当前推进到了未来, 无疑这对提高预报准确率是有利的。

衡量两个或两类事物属性的相似程度有许多方法, 相似系数法、各种距离法等都为广大气象工作者所熟悉, 它们在一定条件下适用, 都有局限性。如相似系数可反映环流型之间的位相差异 (即槽脊位置的不同), 而对同位相波动的强度 (振幅) 难以区分; 各种距离法可很好反映两样本间在数值上的差异程度 (值相似), 对描述系统的强度很有效, 但却不能很好地反映系统形态的相似与否。显然, 既可反映形态的相似也能描述数值的相似, 应是选择相似准则的基本要求。我们认为, 相似离度法就是符合这一基本要求的一种比较客观、合理的描述相似性方法。

对两样本 i 与 j , 相似离度的计算公式为:

$$C_{ij} = \frac{(\alpha \times R_{ij} + \beta \times D_{ij})}{\alpha + \beta} \quad (4.1)$$

式中:

$$R_{ij} = \frac{1}{m} \sum_{k=1}^m |H_{ij}(k) - E_{ij}| \quad (4.2)$$

$$D_{ij} = \frac{1}{m} \sum_{k=1}^m |H_{ij}(k)| \quad (4.3)$$

$$H_{ij}(k) = H_i(k) - H_j(k) \quad (4.4)$$

$$E_{ij} = \frac{1}{m} \sum_{k=1}^m H_{ij}(k) \quad (4.5)$$

其中: k 表示因子变量 ($k=1, 2, \dots, m$), 此为格点序号, m 为总因子数, 即所考察的

相似场的格点数, $H_i(k)$ 为第 i 个样本场的第 k 个因子, $H_j(k)$ 为第 j 个样本场的第 k 个因子, E_{ij} 表示 i 样本场对于 j 样本场所有因子之间的总平均差值, R_{ij} 反映两个样本中对它们每个因子场各自平均值离散程度的差异为形相似, D_{ij} 反映两个样本对应点之间的差异为值相似, α 、 β 分别为它们对总相似程度的贡献系数。一般来说, 不同的气象要素 (物理量) 场, 其分布形态和数值大小对产生的天气有不同的影响。如对高度场, 槽脊的地理位置 (形态) 一般比其强度 (数值) 有更明显的影响, 前者常决定某天气能否出现, 后者则主要影响天气的强度; 而对湿度场来说, 干湿的分布形态固然重要, 湿度大小具体数值的作用显然不次于前者。因此, 对不同的要素 (物理量) 场应取不同的 α 、 β 值。对高度场可取 $\alpha=2$ 、 $\beta=1$, 对湿度场可取 $\alpha=\beta$ 等。

4.3 使用资料

4.3.1 沙尘暴个例资料

作相似预报必须要有一批历史样本资料作为挑选相似个例的基础。故此我们选择了配合有 ECMWF 资料的 45 个沙尘暴个例, 对全区 106 个国家基准、基本站按照地域将其分为全区、西部、中部、中西部以及中东部 5 类, 再按强弱分别分为强和一般 2 类。具体区域定义、个例选择及强度划分标准如下:

沙尘暴区域划分标准:

西部: 指阿拉善盟、巴彦淖尔盟、鄂尔多斯市、包头市;

中部: 指包头市、呼和浩特市、乌兰察布盟、锡林郭勒盟;

中西部: 指阿拉善盟、巴彦淖尔盟、鄂尔多斯市、包头市、呼和浩特市、乌兰察布盟、锡林郭勒盟西部、南部;

中东部: 指包头市、呼和浩特市、乌兰察布盟、锡林郭勒盟、赤峰市、通辽市、呼伦贝尔市、兴安盟;

全区: 指全区大部。

个例选择标准 (满足如下条件为一个沙尘暴日):

西部: 24 小时内该区域至少有 5 个站出现沙尘暴;

中部: 24 小时内该区域至少有 5 个站出现沙尘暴;

中西部: 24 小时内该区域至少有 7 个站出现沙尘暴;

中东部: 24 小时内该区域至少有 7 个站出现沙尘暴;

全区: 24 小时内该区域至少有 12 个站出现沙尘暴。

沙尘暴强度划分标准:

所有出现沙尘暴的站有三分之一能见度大于 500 米小于等于 1000 米的为一般沙尘暴天气过程;

所有出现沙尘暴的站有三分之一能见度小于等于 500 米的为强沙尘暴天气过程。

4.3.2 历史样本资料选择

ECMWF 的形势预报是大家所公认的, 而且其格式也比较稳定, 同时它的三个因子场(地面气压、850hPa 温度和 500hPa 高度场)也能比较客观地反映沙尘暴的天气条件(风、温度等), 因而选择该资料作为历史资料场既客观又比较合理。考虑到从 88 年起才有比较完整的 ECMWF 资料, 故只选取 88 年以后的沙尘暴个例。历史场资料为 ECMWF 的 1988—2002 年 3—5 月对应沙尘暴个例的三个场的格点资料, 另外还有经过加工的可以在 Micaps 下显示的每个个例各站点日最低能见度分布情况。

4.3.3 实时资料

实时资料根据预报时段的要求分别选择从网上接收的 ECMWF 的地面气压场、850hPa 温度场和 500hPa 高度场资料, 时次为 48、72、96 三个时次。

4.4 预报思路

4.4.1 消空处理

根据多年的预报经验, 用相似离度分析方法制作沙尘暴预报之前, 先通过日本传真图进行消空, 选择满足什么类型, 如: 西部一般, 中部强等。具体做 24 小时预报要先看日本 24 小时和 36 小时 850hPa 风场预报(即传真图 JFXFE572 和 JFXFE573)进行消空, 做 48 小时预报要先看日本 24—48 小时地面气压场及降水预报(即传真图 JFSA04)进行消空, 做 48 小时预报要先看日本 48—72 小时地面气压场及降水预报(即传真图 JFSA07)进行消空。

4.4.1.1 日本传真图消空范围

西部: 100—110 E、38—47 N;

中部: 105—120 E、40—47 N;

中西部: 100—115 E、38—47 N;

中东部: 110—125 E、40—50 N;

全区: 100—125 E、38—50 N。

其中, E 表示东经多少度, N 表示北纬多少度。

4.4.1.2 日本传真图消空指标

24 小时：对应各区的消空范围中 572 和 573 图上 5 个纬距或经距内 850hPa 风场至少有两个大于等于 8m/s，则选择为一般沙尘暴；

24 小时：对应各区的消空范围中 572 和 573 图上 5 个纬距或经距内 850hPa 风场至少有两个大于等于 12m/s，则选择为强沙尘暴；

对于 48 小时和 72 小时预报只做一般沙尘暴的消空，用 S04 和 S07 图中地面气压梯度作判定。

48 小时：对应各区的消空范围 S04 图上在 5 个纬距或经距内出现大于等于 3 根等压线；

72 小时：对应各区的消空范围 S07 图上在 5 个纬距或经距内出现大于等于 3 根等压线。

4.4.2 相似类型选择

通过消空选择进行判定的相似类型，有：西部一般、西部强、中部一般、中部强、中西部一般、中西部强、中东部一般、中东部强、全区一般和全区强 10 类。

4.4.3 相似离度分析预报

4.4.3.1 关键区选择

按照不同类型区域做相似离度分析时应选择不同的关键区，对西部、中部及中西部选 80—120°E、30—60°N，共 9×7 个格点；对中东部选 100—135°E、35—65°N，共 8×7 个格点；对全区选 80—135°E、30—65°N，共 12×8 个格点。

4.4.3.2 预报制作

做 24 小时预报时，选择前一天欧洲 ECMWF 48 小时预报场的地面气压场、850hPa 温度场和 500hPa 高度场资料，用对应关键区的格点值与所选类型（西部一般、西部强、中部一般、中部强、中西部一般、中西部强、中东部一般、中东部强、全区一般和全区强等 10 类之一）的样本场关键区的格点值做相似离度分析，通过计算，取其相似离度值最小的为最佳相似个例。做 48 小时预报时，选择前一天欧洲 ECMWF 72 小时预报场的地面气压场、850hPa 温度场和 500hPa 高度场资料，用对应关键区的格点值与所选类型（西部一般、中部一般、中西部一般、中东部一般、全区一般等 5 类之一）的样本场关键区的格点值做相似离度分析，取其相似离度值最小的为最佳相似个例。做 72 小时预报时，选择前一天欧洲 ECMWF 96 小时预报场的地面气压场、850hPa 温度场和 500hPa 高度场资料，用对应关键区的格点值与所选类型（西部一般、中部一般、中西部一般、中东部一般、全区一般等 5 类之一）的样本场关键区的格点值做相似离度分析，取其相似离度值最小的为最佳相似个例。需说明一点历史资料的水平分辨率为 $1^{\circ} \times 1^{\circ}$ ，但实时资料则为 $0.5^{\circ} \times 0.5^{\circ}$ ，而且资料对应的范围也不同。

4.5 预报流程

通过上述分析、选择，建立沙尘暴预报业务系统粗框图（见图 4.1）

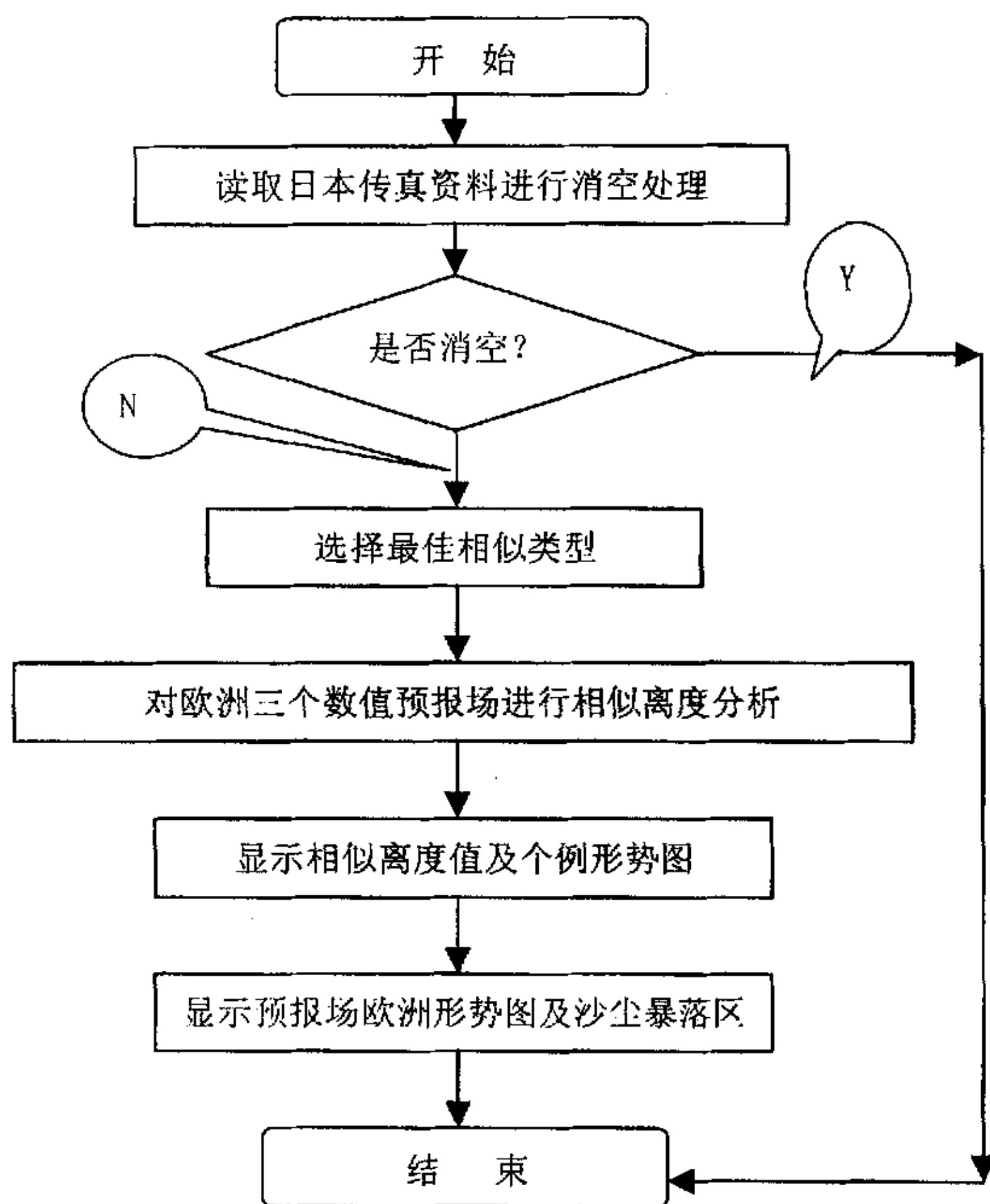


图 4.1 沙尘暴预报流程粗框图

4.6 试验结果讨论

沙尘暴天气目前已作为一种全球问题引起了世界各地（包括美国、日本、韩国等）对其的监测和研究，我国沙尘天气预警业务服务暂行规定也自 2002 年 3 月 20 日起开始执行，关于沙尘暴预报方法的研究随之变得十分迫切，但均处于摸索阶段。本方法的研制本着边应用边改进的原则，正在逐步完善。通过 2003 年 3—5 月的试用得到一些结论和启示。

2003 年春季沙尘天气过程相对很少，在我区满足标准的沙尘暴天气过程仅有 1 次（4 月 17 日，在观测的同一时次有 3 个站出现沙尘暴天气），即使同一天中有 3 个站出现沙尘暴天气现象的也只有 5 天（4 月 10 日、16 日、17 日、5 月 4 日、26 日）。在对 2003 年 3—5 月逐日的应用中，发现该方法对沙尘天气预报有一定的指导意义，对现有的样本个例，相似高度值小于 6 时，就可能局地出现扬沙或沙尘天气。分析实况，今春没有一个满足本文所列条件

的天气过程,这虽然不能对该方法进行很好的验证,但通过应用发现,不论是消空指标,还是样本个例,都应该进一步完善和充实,如相对沙尘天气出现的时段,对消空条件应做时段限制;在个例少的前提下,应减少分类,或者降低条件增加个例;另外可以考虑用数学方法进行分类,用欧洲3个形势场24小时的变化值进行相似离度分析,计算相似离度等等。相信随着今后应用中的不断改进和完善,再配合其他监测手段及预报方法,必定能更好地提高沙尘暴预报准确率。

第五章 神经网络在飞机人工增雨预报中的应用

5.1 飞机人工增雨目的及其预报的必要性

干旱是世界范围的严重问题。内蒙古自治区地域辽阔,干旱少雨,水资源严重匮乏,通过现有的人工影响天气科学合理的手段最大限度地开发空中水资源的潜力,对缓解水资源短缺和保护生态环境可以起到重要的辅助作用。飞机人工增雨是一项投资少、见效快、效益高、影响大的社会公益事业。我区飞机人工增雨从一九五八年已开始进行了四十余年。进入七十年代后,实行作业与科研相结合,使飞机人工增雨逐步走向业务化,在农牧业抗旱减灾工作中发挥了日益明显的作用,逐步为社会和政府肯定,成为了农牧业抗旱减灾的一项重要科技措施。近年来,由于西部大开发战略实施,积极开发利用空中水资源,我区飞机增雨的服务领域已经扩大到生态建设方面,飞机增雨基地由九十年代3处发展到6处。飞机人工增雨是以一定的天气条件为基础的,天气条件的好坏直接影响着人工增雨的效益。因此,提供适合人工增雨飞机作业的天气条件和有针对性的预报(如稳定性降水区域、增雨作业高效区预报等)对提高人工增雨作业效率是非常必要的。然而要想真正做到降水预报的客观定量,就目前的技术力量和探测手段还是有很大难度的。近年来,各种数值预报产品为我们提供了多层次、多要素、多时效、多种物理量的丰富资料,为预报员开展各种释用工作奠定了良好的基础,便于提高预报员的预报能力及准确率,加之有了Micaps这一强大的气象信息综合分析处理系统,可以把各种释用方法挂接在Micaps平台上,这也使得预报工作在人工影响天气工作中的作用更加突出出来。飞机人工增雨天气预报不仅关系着增雨作业时机的把握,而且对作业后的效果评估意义重大,准确的预报服务将明显提高人工增雨决策的科学性、先进性和客观化。

5.2 人工神经网络的基本工作原理

5.2.1 人工神经网络概述

人工神经网络是通过大量的但却是很简单的处理单元（神经元）广泛互连而成的复杂网络系统。它反映了人脑和神经系统功能的许多基本特性，但并非人脑和神经网络系统的真实写照，而只是用数学模型对其加工、存储和处理信息的机制的某种简化、抽象和模拟。人工神经网络的信息处理由神经元之间的相互作用来实现，并决定于各神经元连接权重的动态演化过程。

人工神经网络是高度复杂的非线性动力学系统。它除具有一般非线性系统的共性外，其显著优势在于：大规模并行处理；具有良好的容错性；自适应性，即具有很强的学习功能，网络中的大量参数均由学习（训练）获得，并能在使用中不断调整，从而使系统变得更“聪明”。

人工神经网络的数学模型很多，有代表性的有：感知器模型、反向传播模型（BP）、自适应共振理论模型、双向联想存储器、Kohonen网络、Hopfield网络等。其中目前研究最多，应用面最广的是BP模型，下面即以BP模型为例，介绍人工神经网络的基本工作原理。

5.2.2 BP网络的基本工作原理

BP模型由Rumelhart等人提出。其网络结构由输入层、隐层和输出层组成，工作过程中包括信息的正向传播和误差信号的反向传播，由于后者是区别于其他模型的主要特征，故相应的算法就称为反向传播学习算法（Back Propagation Learning Algorithm）。在正向传播过程中，输入信息从输入层经隐层逐层处理，并传向输出层，每一层神经元的状态只影响下一层神经元的状态。如果在输出层不能得到期望的输出（误差大于要求的精度），则将误差信号沿原来的连接通道作反向传播，通过修正各层神经元的权重，使误差减小。重复上述过程，直至误差最小或得到期望精度的输出。图5.1是一个BP网络示意图。

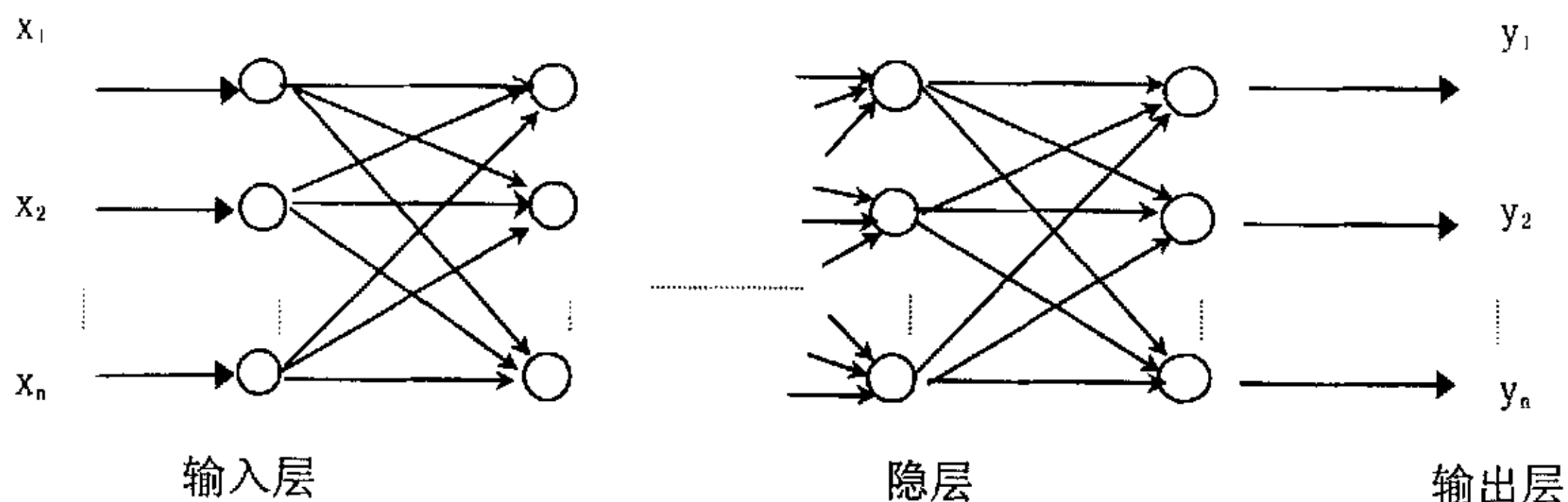


图 5.1 BP 网络示意图

BP 神经网络把一组样本的输入/输出问题作为非线性优化问题来处理，使用优化中最普通的梯度下降法，用迭代运算求解权重值，并通过引入隐节点的方法增加优化问题的可调参

数，从而使得得到更精确的解成为可能。

5.2.3 BP网络算法

5.2.3.1 神经网络的简单数学模型

图 5.2 是简单一个神经元的数学模型

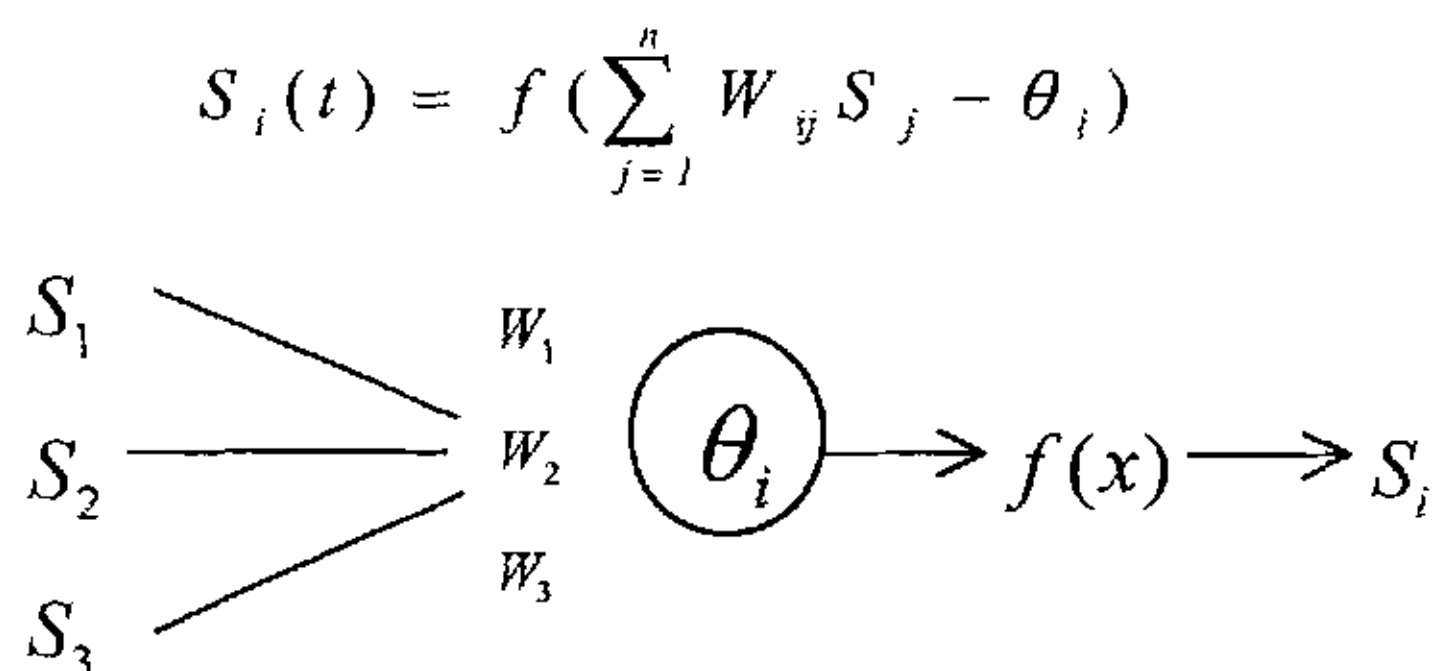


图 5.2 简单一个神经元的数学模型

θ_i 是阈值，大于则兴奋；小于则抑制， $f(x)$ 为激励函数， W_i 为权重。

其中，激励函数（也叫活化函数）可以是线性的，也可以是非线性的，在 BP 网络中一般选取的有三个，它们均是可导的：

线性激励函数：

$$y_k = w \cdot x_k \quad (\text{线性})$$

对数 S 形激励函数：

$$f(x) = \frac{1}{1 + e^{-x}} \quad (\text{log-sigmoid})$$

双曲正切 S 形激励函数：

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (\text{tan-sigmoid})$$

5.2.3.2 BP 网络算法

根据上述原理，BP 网络算法如下：

(1) 隐层节点输入

第 1 层节点输入：

$$net_j^l = \sum_{i=0}^{N-1} w_{ji}^l x_i - \theta_j^l = \sum_{i=1}^{N-1} w_{ji}^l x_i$$

第 L 层节点输入:

$$net_j^l = \sum_{i=0}^{N-1} w_{ji}^l x_i - \theta_j^l = \sum_{i=1}^{N-1} w_{ji}^l x_i$$

隐节点输出:

$$y_j^l = f(net_j^l) = f\left(\sum_{i=0}^{N_{l-1}-1} w_{ji}^l y_i^{l-1}\right)$$

(2) 输出层节点的输入:

$$net_j^0 = \sum_{i=0}^{N_{l-1}-1} w_{ji}^0 y_i^{l-1} - \theta_j^0 = \sum_{i=0}^{N_{l-1}-1} w_{ji}^0 y_i^{l-1}$$

输出层节点的输出:

$$y_k^0 = f(net_k^0)$$

(3) 输出层权值调整:

$$\Delta w_{kj}^0 = \eta y_j^{l-1} \delta_k^0$$

$$\delta_k^0 = (d_k - y_k) f'(net_k^0)$$

d_k 为当前样本期望输出的第 k 个分量;

(4) 隐层权值调整:

$$\Delta w_{kj}^l = \eta y_j^{l-1} \delta_k^l$$

$$\delta_k^l = (d_k - y_k) f'(net_k^l)$$

其中因子 η 为一个正常数, 它控制着沿梯度负方向移动的速度有多快, 称之为“学习速率”, 也叫学习因子。

5.2.4 BP 算法的改进

BP 算法的主要缺点为: 收敛速度慢; 局部极值; 难以确定隐层和隐结点个数。

BP 算法的改进主要有两种途径,一种是采用启发式学习方法,另一种则是采用更有效的优化算法。在 MATLAB 软件的神经网络工具箱中给出了 3 个函数用于对 BP 网络的训练。函数 `trainbpx()` 采用了动量法和学习率自适应调整的策略,从而提高了学习速度并增加了算法的可靠性。

动量法降低了网络对于误差曲面局部细节的敏感性,有效地抑制网络陷于局部极小。标准 BP 算法实质上是一种简单的最速下降静态寻优算法,在修正 $w(k)$ 时,只是按照 k 时刻的负梯度方向进行修正,而没有考虑到以前积累的经验,即以前时刻的梯度方向,从而常常使学习过程发生振荡,收敛缓慢。为此有人提出了改进算法:

$$w(k+1) = w(k) + \eta[(1-\alpha)D(k) + \alpha D(k-1)]$$

其中 $w(k)$ 既可表示单个权值,也可表示权值向量。 $D(k) = \frac{-\partial E}{\partial w(k)}$ 为 k 时刻的负梯度。 $D(k-1)$ 为时刻的 $(k-1)$ 负梯度。 η 为学习率, $\eta > 0$ 。 α 为动量因子, $0 \leq \alpha < 1$ 。这种方法所加入的动量项实质上相当于阻尼项,它减小了学习过程的振荡趋势,从而改善了收敛性。

自适应调整学习率有利于缩短学习时间。标准 BP 算法收敛速度慢的一个重要原因是学习率选择不当。学习率选得太小,收敛太慢;学习率选得太大,则有可能修正过头,导致振荡甚至发散。因此出现了自适应调整学习率的改进算法:

$$w(k+1) = w(k) + \eta(k)D(k)$$

$$\eta(k) = 2^\lambda \eta(k-1)$$

$$\lambda = \text{sign}[D(k)D(k-1)]$$

当连续两次迭代其梯度方向相同时,表明下降太慢,这时可使步长加倍;当连续两次迭代其梯度方向相反时,表明下降过头,这时可使步长减半。

函数 `trainlm()` 使用了 Levenberg-Marquardt 优化方法的权值调整率选为:

$$\Delta w = (J^T J + \mu I)^{-1} J^T E$$

其中 J 为误差对权值微分的 Jacobian 矩阵, E 为误差向量, μ 为一个标量,当 μ 很大时,上式就接近于梯度法;当 μ 很小时,上式就变成了 Gauss-Newton 法,在这种方法中, μ 也是自适应调整的。

对于较大的学习因子，我们将有可能得到差别很大的结果，或者在极小值点附近剧烈振荡，为了对这种振荡做一定程度的补救，通常我们在权重系数迭代公式中引入一个所谓动力项 α ，它取决于上一步迭代中权重系数的增量值，如下所示：

$$w_{ji}(k+1) = w_{ji}(k) + \alpha \Delta w_{ji}(k) + (1-\alpha) \Delta w_{ji}(k-1)$$

网络的目标函数：

$$E = (\delta_k^2) = \frac{1}{P} \sum_{k=1}^P (d_k - y_k)^2$$

5.3 BP 神经网络在飞机人工增雨预报中的应用

5.3.1 BP 神经网络在天气预报中的应用概述

由上述人工神经网络的基本工作原理可知，BP 神经网络的计算过程分为两个阶段：第一阶段为学习训练阶段，根据提供的样本资料，通过调整各层之间的权值。使之达到预定的拟合精度要求；第二阶段是应用阶段，通过输入层的输入，依据学习训练阶段得到的权值，给出的输出量即为预报值。可见，学习训练阶段是关键。其具体步骤及注意事项有：

5.3.1.1 资料的准备

包括选取预报对象、选取预报因子（系统的输入）及选取训练样本。

5.3.1.2 学习训练

（1）构造神经网络结构。用于天气预报神经网络多为 3 层，即除输入层和输出层外，常只取一个隐层。而隐层节点数 n （隐单元个数）的选取，一般可通过经验公式 $n = \log_2 M$ 来确定，其中 M 为输入节点数。

（2）设置网络的初始权值。一般可选某一随机小量为初始权值。

（3）确定学习精度，以决定训练学习的时机。

（4）逐个对输入的 N 个样本，计 L 个输出。

（5）从输入层到输入层的后一层（隐层），计算误差，反向传输修正权值。

（6）当输出与期望不一致时，重复（4）、（5）的学习过程，直至达到精度要求或产生振荡（不能收敛）。

学习结束后，权值保持不变。应用过程仅相当于上述学习过程的第（4）步，每输入一组“实时”资料，输出层的输出值即为预报值。

5.3.2 降水分布情况

针对人工增雨中部基地工作需求,详细统计了内蒙古中西部(包括从阿拉善盟到锡林郭勒盟西部的68个测站)各测站1995—2000年4—8月逐日降水情况,分析不同量级降水在年各月出现的情况,表5.1列出了1995—2000年4—8月各级雨量统计表。

表 5.1 历年中西部 68 个站 4—8 月各级雨量统计表

(1995-2000)(单位: mm)

雨量	<1.0	$1.0 \leq R < 10.0$	$10.0 \leq R < 25.0$	$25.0 \leq R < 50.0$	$50.0 \leq R < 100.0$	$R \geq 100.0$	合计
1995	759	1104	345	114	29	1	2352
1996	987	1353	355	114	14	1	2824
1997	735	1187	256	81	17	1	2277
1998	918	1560	388	125	37	2	3030
1999	772	1028	264	61	7		2132
2000	902	1214	267	59	3	1	2446
总计	5073	7446	1875	554	107	6	17056
年均	845.5	1241	312.5	92.3	17.8	1	2842.7

考虑飞机人工增雨的因素,从表 5.1 不难发现,我们将降水等级分为 6 级,若日降水量用 R 表示,规定 $R < 1.0$ 为 1 级, $1.0 \leq R < 10.0$ 为 2 级, $10.0 \leq R < 25.0$ 为 3 级, $25.0 \leq R < 50.0$ 为 4 级, $50.0 \leq R < 100.0$ 为 5 级, $R \geq 100.0$ 为 6 级。同时我们将各月各年各级雨量出现的情况做图如下:

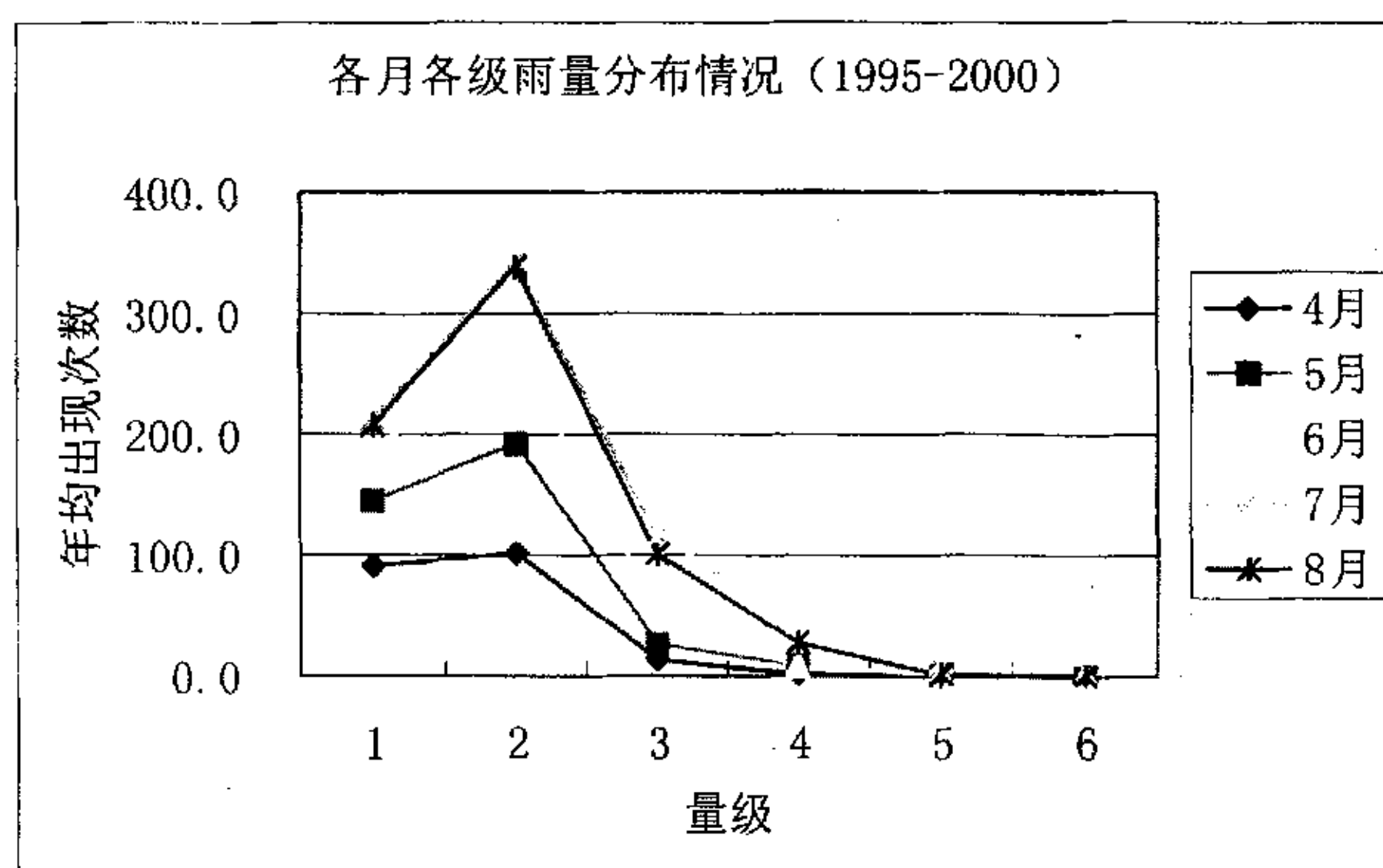


图5.1 各月各级雨量分布情况 (1995-2000年)

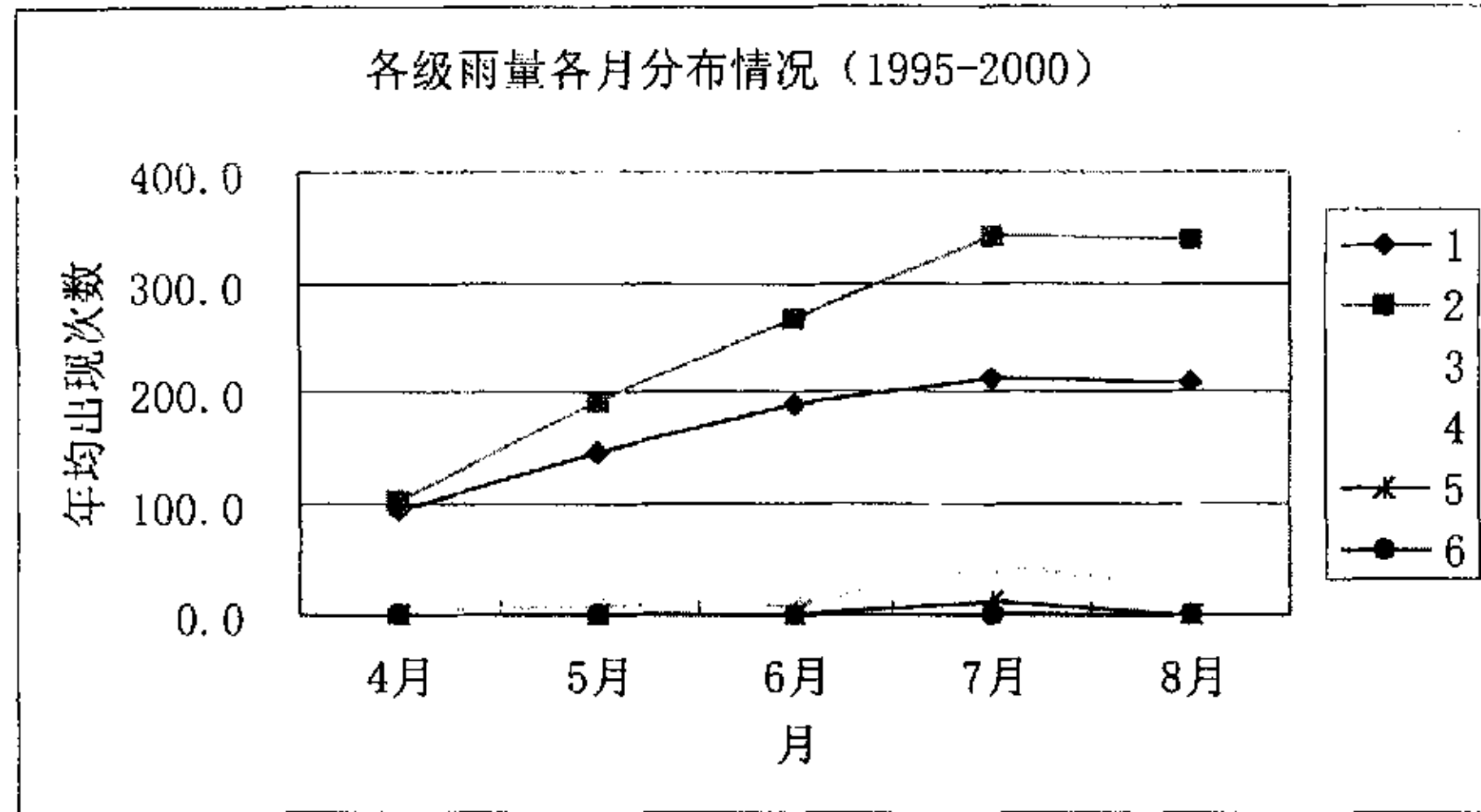


图5.2 各级雨量各月分布情况 (1995-2000年)

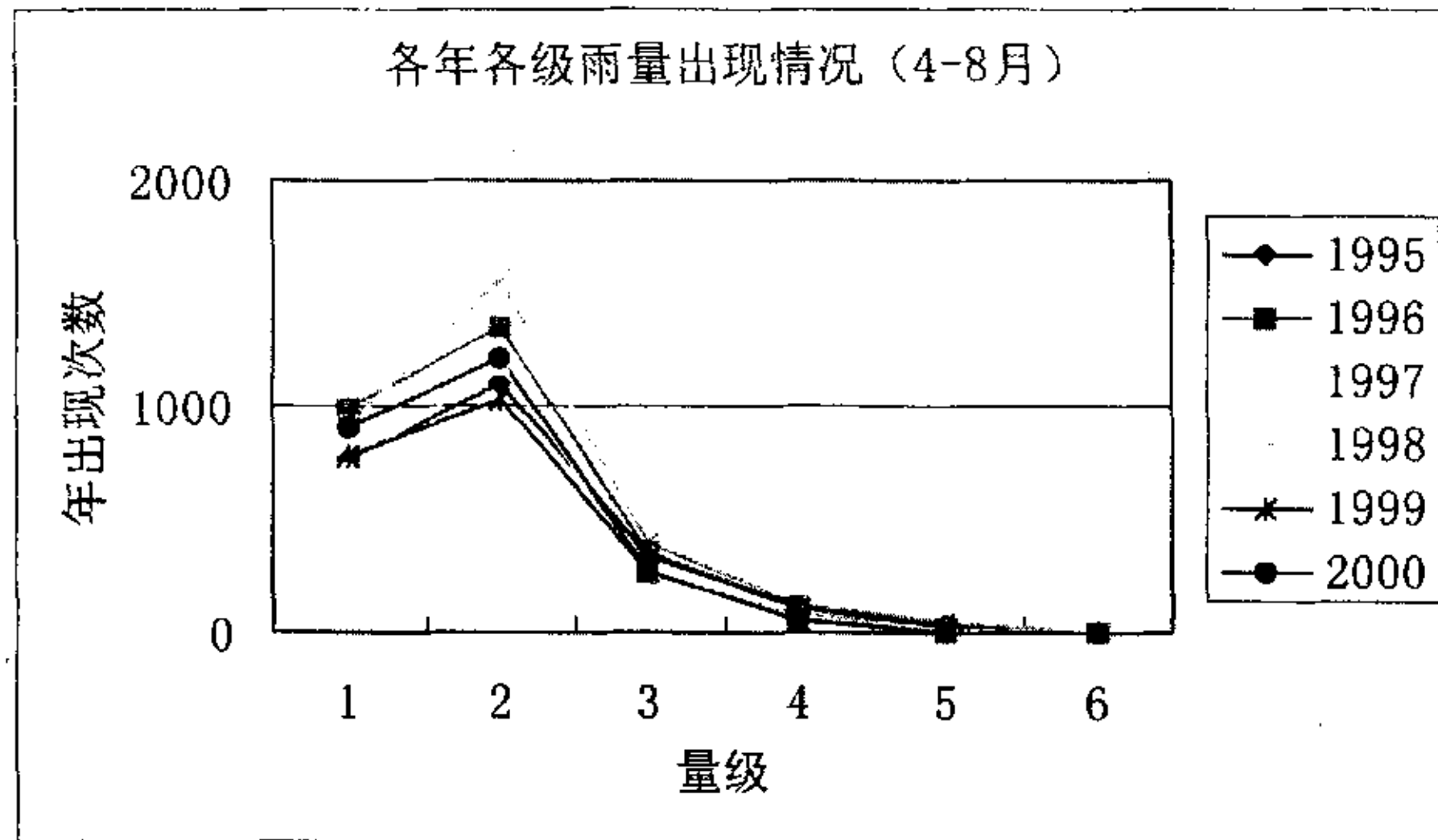


图5.3 各年各级雨量出现情况 (4-8月)

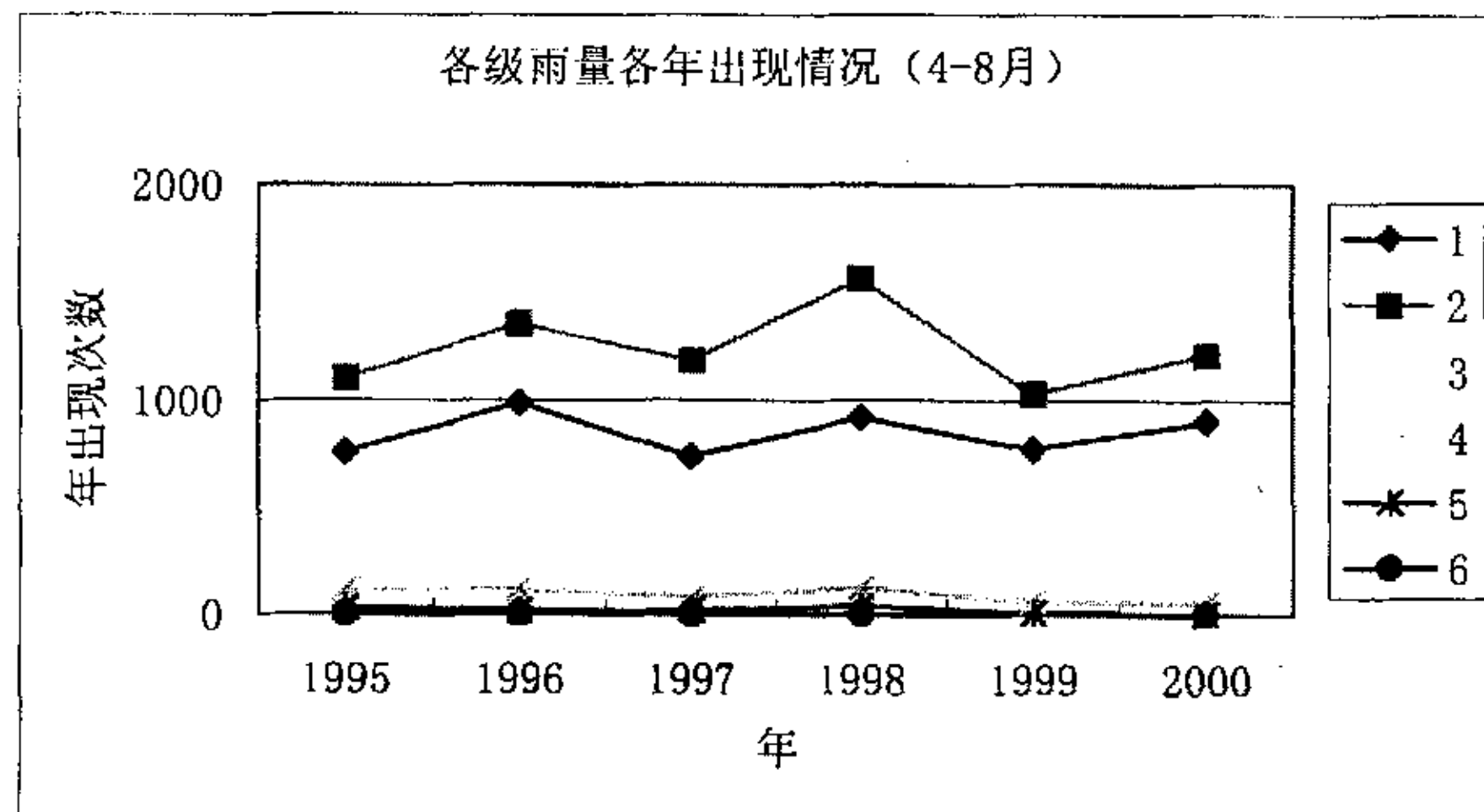


图5.4 各级雨量各年出现情况 (4-8月)

图5.1显示,降水主要集中在7、8月份,4、5、6月降水以1、2级为主。分析图5.2,2级降水次数最到级次之,4月分降水偏少,1、2级几乎各半,3级降水集中在7、8月,大约为2级降水的1/3,这在图5.3中也有表现。分析图5.4发现,1、2级降水的多少与当年降水多少成正比,且呈一年多一年少的特征。

5.3.3 天气过程确定

针对飞机增雨作业的实际要求,我们重点注意的是2、3级雨量出现的情况,而从上面的统计可发现,2、3级雨量占整个量级的一半以上。但因为飞机作业要有一定的范围,而且不是雷雨天气,并不是每一个2、3级雨量都适合飞机增雨,因而在此基础上,我们又详细分析了每一个2、3级雨量出现的区域及其时间,确定出适宜于人工增雨的天气过程(定义一天为一个过程)。

5.3.4 因子选取

T106数值预报产品包括22个要素,16层(有些要素不是每层都有),在大量的资料中我们选取了能表征降水天气的8个物理量共19个场,经过进一步加工分析确定700hPa水汽通量散度、700hPa相对湿度、850hPa温度露点差以及500hPa涡度作为预报因子。根据一般影响中西部降水天气系统的位置,我们选择预报关键区($36^{\circ} \sim 48^{\circ} \text{N}$, $100^{\circ} \sim 115^{\circ} \text{E}$)。T106数值预报产品提供的资料是 $1^{\circ} \times 1^{\circ}$,如果每一网格点上的资料都取的话,就有 $4 \times 13 \times 16$ 个预报因子,如此多的因子,对BP神经网络的学习不会有多大益处,相反却有收敛速度过慢,训练时间增加等弊端,若缩小所取范围,又会把影响降水的天气系统的特征场滤掉。鉴于以上原因,针对每一物理量场,我们每隔两个经度或纬度选一个点,这样对每个场取6个经度、5个纬度共 $6 \times 5 = 30$ 个网格点的资料,因此预报因子实际为120个。具体网格点选取见图5.5。

	100	103	106	109	112	115
	E	E	E	E	E	E
36 N						
39 N						
42 N						
45 N						
48 N						

图5.5 预报因子网格点选取(经纬度)

5.3.5 BP模型建立

5.3.5.1 资料预处理

由于一些最普通的激励函数,例如Sigmoid函数,它们会受到饱和现象的影响,应当将所有的输入特征通过比例变换转换到一个合适的范围内,否则具有较高值的输入将会驱动整个训练过程,从而掩盖了那些具有较低值的输入对于训练的贡献。通常来讲,预处理过程包含一系列的线性的特征比例变换,将所有的特征变换到 $[0, 1]$ 或者是 $[-1, 1]$ 区间内。在此我们采用下式对预报因子进行归一化处理,使每一个输入值包含在 $[0, 1]$ 之间。

$$x_{ij} = \frac{x_{ij} - x_{\min}}{x_{\max} - x_{\min}}$$

5.3.5.2 训练样本选取

根据上面天气过程的确定,选出无雨个例172个(标准:所选68站基本都无雨或最多有3个站雨量为1级或2个站雨量为2级),适合增雨个例160个(标准:所选68个站至少有20个站雨量为2级以上,且雨量均衡,至少有30站以上有雨,关键还不是雷雨天气)。对所选的样本我们进行4种方式的组合:第一种训练样本为后4年的(无雨113例、有雨99例)定义为A,第二种训练样本为前4年的(无雨106例、有雨107例)定义为B,第三种尽量按时间间隔选取无雨和小雨(大于3级的不足10个)(无雨65个,有雨51个)定义为WX,第四种在第三的基础上有雨个例再加上中雨日数(13个)定义为WXZ。

5.3.5.3 BP模型建立

对以上四种训练样本均采用一个隐层的两层BP网络,隐层和输出层函数都为logsig,对其分别采用3种训练函数(trainbp、trainbpx和trainlm),4种隐层结点数(4、6、8、10)共12种方式进行训练。3种训练函数性能对比结果见表5.2-5.4。

表5.2 函数trainbp训练结果

样本	A		B		WX		WXZ	
隐结 点数	训练 步数	误差	训练 步数	误差	训练 步数	误差	训练 步数	误差
4	30000	0.5128	30000	0.5071	30000	0.0116	30000	0.0103
6	30000	0.5091	30000	0.5088	30000	0.0100	30000	0.0117
8	30000	0.5081	30000	0.5069	30000	0.0010	30000	0.0010
10	30000	0.5087	30000	0.5081	30000	0.0010	30000	0.0010

表5.3 函数trainbpx训练结果

样本	A		B		WX		WXZ	
隐结 点数	训练 步数	误差	训练 步数	误差	训练 步数	误差	训练 步数	误差
4	10000	0.5009	10000	0.5171	269	0.0095	253	0.0099
6	10000	0.5008	10000	0.5016	258	0.0095	363	0.0098
8	10000	0.5014	10000	0.5000	255	0.0098	277	0.0098
10	10000	0.5006	10000	0.5006	258	0.0096	247	0.0098

表5.4 函数trainlm训练结果

样本	A		B		WX		WXZ	
隐结 点数	训练 步数	误差	训练 步数	误差	训练 步数	误差	训练 步数	误差
4	18	0.5	22	0.5	6	0.0089	9	0.0019
6	10	0.5	9	0.5	11	0.0032	12	0.0031
8	16	0.5	11	0.5	6	0.0067	8	0.0029
10	13	0.5027	11	0.5	8	0.0039	6	0.0035

由表5.2-5.4可见, 函数trainlm的训练速度明显比trainbp和trainbpx快, 而trainbpx也比trainbp快。另外训练速度与样本个数及质量有关, 从表中发现A、B比WX、WXZ的训练速度慢, 但A、B或WX、WXZ本身的差别不大。同时还表明对于同一输入层, 同一输出层, 不同隐层结点数, 训练样本的整体系统误差是不一样的。

为了检验样本的随机性导致的网络训练能力和预报能力的差别, 我们对上述几种训练结果都做了不同的检验, 对四种样本的12种情况都统计了它们对所有样本的拟合率, (注: 对拟合率选两种不同情况, 其中1指的是误差在0.01之内的检验结果的拟合率, 2指的是误差在0.1之内的检验结果的拟合率)。考虑到WX和WXZ两种训练样本的随机性及其对总体样本的拟合率本身比样本A和B低 (见表5.5), 对训练样本自身的拟合率、空报率、漏报率及TS评分都只对样本A和B进行了检验, 结果见表5.6和5.7。这四种检验都按误差分1、2两种情况 (类似对总体的拟合率), 对各种检验的具体计算方法如下:

总体拟合率=符合总体检验结果的样本数/总体样本数

自身拟合率=符合自身检验结果的样本数/自身样本数

空报率=空报次数/(空报次数+实际适合增雨次数)

漏报率=漏报次数/实际适合增雨次数

TS评分=预报正确的次数/(预报正确的次数+空报次数+漏报次数)

表5.5 四种训练样本对所有样本拟合率

训练模式	A		B		WX		WXZ	
	1	2	1	2	1	2	1	2
bp4	80.7%	94.0%	83.7%	94.3%	72.3%	88.0%	72.9%	86.7%
bpx4	90.7%	93.4%	93.1%	94.3%	74.1%	90.4%	74.4%	88.9%
lm4	92.5%	95.2%	93.7%	95.5%	71.4%	85.5%	85.2%	89.5%
bp6	80.4%	94.0%	79.2%	92.8%	68.4%	87.0%	69.6%	88.6%
bpx6	91.6%	94.3%	92.2%	94.6%	72.6%	87.3%	74.4%	87.3%
lm6	89.8%	94.0%	84.9%	92.2%	78.3%	87.0%	80.7%	85.5%
bp8	79.2%	94.0%	83.1%	94.3%	72.3%	86.4%	69.0%	85.5%
bpx8	90.7%	92.8%	93.4%	94.9%	72.0%	87.0%	74.7%	87.3%
lm8	93.4%	94.6%	93.1%	95.8%	65.1%	84.0%	78.6%	85.8%
bp10	80.1%	92.2%	81.9%	93.1%	72.3%	87.0%	72.9%	88.9%
bpx10	91.0%	93.4%	91.0%	93.7%	71.4%	88.6%	74.1%	88.6%
lm10	84.3%	90.1%	87.7%	93.4%	71.7%	85.5%	81.9%	91.3%

表5.6 A样本的自身拟合率、空报率、漏报率和TS评分

训练模式	自身拟合率		空报率		漏报率		TS评分	
	1	2	1	2	1	2	1	2
bp4	82.1%	99.1%	18.7%	14.1%	19.7%	13.1%	65.3%	74.6%
bpx4	98.6%	99.1%	18.7%	15.3%	23.0%	14.8%	62.7%	72.2%
lm4	99.1%	99.1%	15.3%	10.3%	19.7%	11.5%	68.1%	79.4%
bp6	85.4%	99.1%	22.8%	15.3%	26.2%	11.5%	57.0%	75.0%
bpx6	98.1%	99.1%	16.4%	14.1%	19.7%	11.5%	67.1%	76.1%
lm6	99.1%	99.1%	24.7%	14.1%	19.7%	13.1%	60.5%	74.6%
bp8	86.3%	99.1%	24.7%	10.3%	32.8%	18.0%	50.6%	73.5%
bpx8	98.6%	99.1%	17.6%	16.4%	24.6%	16.4%	62.2%	69.9%
lm8	99.5%	99.5%	7.6%	7.6%	26.2%	19.7%	68.2%	74.2%
bp10	84.0%	99.1%	21.8%	17.6%	24.6%	18.0%	59.0%	67.6%
bpx10	99.1%	99.1%	19.7%	14.1%	21.3%	16.4%	63.2%	71.8%
lm10	98.1%	99.1%	22.8%	15.3%	49.2%	32.8%	39.2%	56.9%

表5.7 B样本的自身拟合率、空报率、漏报率和TS评分

训练模式	自身拟合率		空报率		漏报率		TS评分	
	1	2	1	2	1	2	1	2
bp4	89.2%	99.5%	17.2%	5.4%	37.7%	28.3%	51.6%	67.9%
bpx4	99.1%	99.1%	8.6%	5.4%	30.2%	26.4%	63.8%	69.6%
lm4	100.0%	100.0%	10.2%	7.0%	28.3%	20.8%	64.4%	73.7%
bp6	86.9%	99.1%	29.3%	14.5%	35.8%	24.5%	45.3%	64.5%
bpx6	99.1%	99.1%	14.5%	8.6%	28.3%	20.8%	61.3%	72.4%
lm6	99.1%	99.1%	32.1%	15.9%	43.4%	26.4%	38.5%	61.9%
bp8	88.3%	99.1%	18.5%	5.4%	35.8%	26.4%	52.3%	69.6%
bpx8	99.1%	99.1%	11.7%	10.2%	24.5%	17.0%	66.7%	74.6%
lm8	99.1%	99.1%	7.0%	5.4%	32.1%	17.0%	63.2%	78.6%
bp10	86.9%	99.1%	20.9%	10.2%	34.0%	28.3%	52.2%	64.4%
bpx10	99.1%	99.1%	15.9%	10.2%	34.0%	24.5%	55.6%	67.8%
lm10	99.1%	99.1%	23.2%	11.7%	43.4%	24.5%	43.5%	66.7%

表5.5表明训练样本的多少直接影响拟合率的大小,一般样本越多,拟合率越高,从表中可见B样本的总体拟合率最高。再比较表5.6和5.7发现实际B样本的自身拟合率也比A样本高,但其TS评分整体却比A样本低,因为在实际预报中对(0、1)判定结果误差满足在0.1之内就可以,所以从表5.6和5.7的最后一列可以看到,BP模式的TS评分在70%左右,而且并不是隐层结点数越多越好,对A样本隐层数为4的trainlm函数TS评分结果最佳,是79.4%,对B样本隐层数为8的trainlm函数TS评分结果最佳,是78.6%,因此实际应用中我们可以选择这两个模式的训练结果。

5.4 讨论及试用

BP 人工神经网络方法对适合飞机人工增雨天气的识别,可以不分年、不分月,而且不须区分天气形势,这是因为它具有很强的自适应学习的能力,通过训练可以自动调整各层结点之间的连接权重和每个结点的阈值。

BP 算法由于具有非线性动力系统特征,能处理非线性问题,其拟合能力很强,试报能力也可取,因此可作为判别是否适合飞机人工增雨天气的一种预报工具。

该方法在今年6月后的人工增雨预报的指导工作中对中西部、尤其是中部增雨基地的工作起到了举足轻重的作用,中部基地的几次作业过程中都成功地预报了出来,使得增雨机组抓住有利时机,于6月22日等先后作业七架次。

第六章 各种系统在实际业务中的实现和应用

6.1 Micaps 简介

在讨论各种预报系统的业务应用之前,必须提到 Micaps,即气象信息综合分析处理系统, Micaps 以其友好的界面和集数据采集、转换、成图、分析、预报产品制作于一体的强大功能,深受广大预报人员的喜爱。它是由国家气象局组织开发的一个比较完善的、适于全国的预报工作平台。所以针对不同地区、不同目的,对其进行本地化处理是充分发挥该系统作用的一项重要工作。而各种业务用预报方法、工具在其上的挂接,也是本地化工作的一个方面。

6.2 强对流天气预报工作平台

为了便于操作,按照第 3 章的计算方法,我们将预报过程和结果显示采用 Fortran 和 VB 编程,先用前一日 20 时 HLAFS 资料 12 小时预报的 400 hPa 形势场与 400hPa 特征场做相似分析,判断天气类型,进行消空处理,再根据不同天气类型套用各自的预报方程,得出有无降雹的预报。每个处理程序通过 Micaps 定时器自动运行,计算结果自动生成,并且预报结果和相关物理量场均可通过 Micaps 用户界面直接调用、查看。图 6.1 即为强对流天气预报工作平台,按照图上提示进行操作就可完成强对流天气的预报。

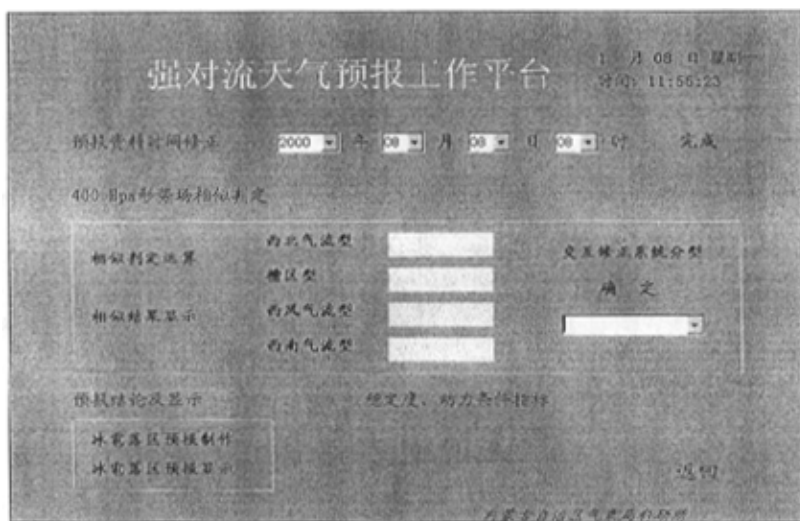


图 6.1 强对流天气预报工作平台

6.3 相似离度的沙尘暴预报业务系统

经过第 4 章的分析、选择,按照相似离度算法,利用 Fortran 编程计算相似离度,确定最佳相似个例,再用 VB 编制沙尘暴天气预报系统流程,建立沙尘暴预报业务系统(见图 6.2)。

12月19日 星期四
时间: 11:49:56

日本传真图消空范围 西部: 100-110° E, 38-47° N 中部: 105-120° E, 40-47° N 中西部: 100-115° E, 38-47° N 中东部: 110-125° E, 40-50° N 全区: 100-125° E, 38-50° N	查看日本传真图 24小时 572 573 48小时 S04 72小时 S07
--	--

西部一般

确定

消空指标 24h: 上述范围中572和573图上5个 纬距或经距内850hPa风场至少有两个 大于等于8m/s, 则为一般沙尘暴 24h: 上述范围中572和573图上5个 纬距或经距内850hPa风场至少有两个 大于等于12m/s, 则为强沙尘暴 48h: 上述范围S04在5个纬距或经 距大于等于3根等压线 72h: 上述范围S07在5个纬距或经 距大于等于3根等压线	欧洲形势预报相似高度分析 时间选择 2002 年 12 月 18 日 预报时段选择 24 相似高度预报 24小时 48和72
---	--

相似结果显示 定义相似日期及相似值 相似日期 相似高度值 910505 4.467	相似场形势及落区 欧洲形势场及预报落区 高空形势场及地面实况
--	---

内蒙古气象台

二〇〇二年

返回

图 6.2 沙尘暴天气预报系统流程

选择上章中最佳的网络结构（即A样本的隐层数为4的trainlm函数训练模式和B样本的隐层数为8的trainlm函数训练模式）以及由其训练出的各结点的权重和阈值，在实际应用中首先提取前一天20时T213（T106资料现已升级为T213）不同时次4个预报场（700hPa水汽通量散度、700hPa相对湿度、850hPa温度露点差以及500hPa涡度）对应关键区（ $36^{\circ} \sim 48^{\circ} \text{N}$ ， $100^{\circ} \sim 115^{\circ} \text{E}$ ）的对应格点资料，对其进行归一处理之后，代入BP神经网络输出公式，即可算

出对应预报时次24小时之内的预报。预报时次可以通过业务系统界面进行选择。

第七章 结 语

随着信息技术的飞速发展和人民生活水平的日益提高,天气预报引起了人们越来越多的关注,重要的关键的灾害性、转折性天气预报的准确与否对社会生产、经济及人们生活都有极大的影响,社会对气象服务的要求也越来越高,因此提高预报准确率和预报时效是预报工作迫切需要解决的问题。以上的工作只是将数据挖掘技术运用于天气专项预报中的初步尝试,还需进行大量的工作,如数据仓库的建立、规则选定等。

展望未来,随着气象知识的逐步积累,知识表示的日益加强,特别是计算机硬件环境的更趋完善,相信在进一步完善气象信息数据的基础上,数据挖掘技术在气象预报中将有广阔的前景,在加速气象业务现代化和提高灾害性天气预报准确率的进程中,必将发挥愈益明显的作用。

参考文献

- [1] Hanjiawei, Kamber, M. Data Mining : Concepts and Techniques (中译本书名: 数据挖掘概念与技术/译者: 范明等) 北京: 机械工业出版社, 2001
- [2] J.P.Marques de sa. Pattern Recognition : Concepts、Methods and Applications (中译本书名: 模式识别—原理、方法及应用/译者: 吴逸飞) 北京: 清华大学出版社, 2002
- [3] 孔玉寿 章东华. 现代天气预报技术. 北京: 气象出版社, 2002
- [4] 王耀生. 人工智能、模式识别在气象领域应用的现状与展望. 气象, 1994, 20 (6): 9-14.
- [5] 刘景涛, 罗孝逞. 内蒙古自治区天气预报手册(下册). 北京: 气象出版社, 1987: 122-125.
- [6] 吴鸿宾等. 内蒙古自治区主要气象灾害分析. . 北京: 气象出版社, 1990: 100-114.
- [7] 李一平, 薄玉华等. 内蒙古中西部地区降雹天气预报方法. 气象, 1999, 25 (7): 37-40.
- [8] 中央气象局. 地面气象观测规范. 北京: 气象出版社, 1979: 21-27.
- [9] 王式功. 中国北方地区沙尘暴变化趋势初探. 自然灾害学报, 1996, 5 (2): 86-94.
- [10] 徐启运, 胡敬松. 我国西北地区沙尘暴天气时空分布特征. 应用气象学报, 1996, 7 (4): 479-482.

- [11] 杨东贞, 房秀梅, 李兴生. 我国北方沙尘暴变化趋势的分析. 应用气象学报, 1998, 9 (3): 352-358.
- [12] 周自江. 近 45 年中国扬沙和沙尘暴天气. 第四纪研究, 2001, 21 (1): 9-17.
- [13] 叶笃正, 丑纪范, 刘纪远 等. 关于我国华北沙尘天气的成因于治理对策. 地理学报, 2000, 55 (5): 513-521.
- [14] 周自江, 王锡稳. 西北地区东部群发性强沙尘暴序列的建立与分析. 地理学报, 2002, 57 (4): 437-442.
- [15] 方宗义, 朱福康, 江吉喜 等. 中国沙尘暴研究. 北京: 气象出版社, 1997: 1-158.
- [16] 刘景涛, 杨耀芳, 李运锦 等. 中国西北地区 1993 年 5 月 5 日黑风暴的机理探讨. 应用气象学报, 1996, 7 (3): 371-376.
- [17] 周自江, 王锡稳, 牛若芸. 近 47 年中国沙尘暴气候特征研究. 应用气象学报, 2002, 13 (2): 193-200.
- [18] 方宗义, 张运刚, 郑新江 等. 用气象卫星遥感监测沙尘暴的方法和初步结果. 第四纪研究, 2001, 21 (1): 48-55.
- [19] 杨行峻, 郑君里. 人工神经网络. 北京: 高等教育出版社, 1992
- [20] Hagan, M. T. etc. Neural Network Design (中译本书名: 神经网络设计/译者: 戴葵等)/北京: 机械工业出版社, 2002
- [21] 张冬峰, 董存仙, 任璞等. 数值预报产品省级降水预报 B—P 神经元网络释用系统. 山西气象, 1999, 47 (2): 1-4.
- [22] 金龙, 陈宁, 林振山. 基于人工神经网络的集成预报方法研究和比较. 气象学报, 1999, 57 (2): 198-207.
- [23] 王繁强, 徐文金, 陈杰伦等. B-P 算法在青海降雨分区分级预报中的应用. 高原气象, 1997, 16 (1): 105-111.
- [24] 师春香, 吴蓉璋, 项续康. 多阈值和神经网络卫星云图云系自动分割试验. 应用气象学报, 2001, 12 (1): 70-78.
- [25] 郝家学, 张经珍, 孙秀忠等. 用神经网络法试报黄河三角洲汛期暴雨. 河南气象, 2000, (3): 19-20.
- [26] 李朝兴, 冶林茂, 乔春贵等. BP 神经元网络算法改进试验. 河南气象, 1999, (1): 41.
- [27] 楼顺天, 施阳. 基于 MATLAB 的系统分析与设计—神经网络. 西安: 西安电子科技大学出

版社, 2000.

[28] 罗建军, 杨琦. 精讲多练 MATLAB. 西安: 西安交通大学出版社, 2002.

[29] 王沫然. MATLAB 5. X 与科学计算. 北京: 清华大学出版社, 2000.

[30] 施阳, 李俊等. MATLAB 语言工具箱—TOOLBOX 实用指南. 西安: 西安西北工业大学出版社, 1998.

致 谢

感谢高光来教授悉心的指导。导师渊博的学识、严谨的治学态度、勤奋的敬业精神、对新思想新技术敏锐的洞察力, 给学生树立了治学和做人的典范。为了结合学生的预报工作实践和计算机应用专业, 从立题到设计导师都提出了不少合理的建议。

在完成该项工作的过程中, 我要深深地感谢我的家人, 他们给予了我极大的支持, 尤其是我的父母, 帮我承担了大部分家务, 使我能全身心地投入到论文的研究中。另外我还要感谢我的同事, 在第 3 章的工作中, 孙永刚副台长在算法和编程上给了很大的帮助, 在第 4 章的工作中, 康玲副台长提供了沙尘暴个例和消空指标, 在第 5 章的工作中, 薄玉华高级工程师帮助查阅了天气过程, 另外苏立娟工程师还帮我修改了英文摘要。同时还要感谢沈建国副局长, 他对我的工作表现了极大的热情, 为我提供了许多有关数据挖掘的材料。

攻读学位期间发表的学术论文目录

1. 一次有利于人工增雨作业的锋面降水天气过程分析、气象、26 卷 9 期、51-54、2000. 9、第三、二级核心刊物、中国气象科学研究院
2. 一次冰雹天气的综合分析、第十三次全国云、降水物理和人工影响天气科学讨论会文集、152-153、2000. 11、第一、一般刊物、内蒙古气象科学研究所
3. 一次人工增雨天气系统部位的选择、第十三次全国云、降水物理和人工影响天气科学讨论会文集、40-41、2000. 11、第二、一般刊物、内蒙古气象科学研究所
4. 一次飞机人工增雨宏观特征及效果分析、第十三次全国云、降水物理和人工影响天气科学讨论会文集、42、2000. 11、第二、一般刊物、内蒙古气象科学研究所
5. 人工防雹新技术与提高防雹总体效益的研究、第十三次全国云、降水物理和人工影响天气科学讨论会文集、140-141、2000. 11、第五、一般刊物、内蒙古气象局
6. 数值产品及卫星云图在降水预报中的综合应用、天气预报技术文集 (2000) (气象出版社)、

- 6-11、2001.4、第五、文集、内蒙古气象台
7. 1998年8月8~12日松嫩流域冷涡大暴雨个例研究、1998年长江嫩江流域特大暴雨的成因及预报应用研究（气象出版社）、447-455、2001.7、第五、文集、内蒙古气象台
8. 关于CIN的一次观测分析与数值模拟、国外强对流天气的应用研究（气象出版社）、116-122、2001.8、第一、译编专著、内蒙古气象科学研究所
9. 冷区高架强雷暴分析的初步结果、国外强对流天气的应用研究（气象出版社）、241-245、2001.8、第一、译编专著、内蒙古气象科学研究所
10. HLAFS数值预报产品在内蒙古中西部地区冰雹预报中的释用技术、北京区域气象技术开发资金论文和成果汇编（气象出版社）、122-131、2001.12、第一、成果汇编专著、内蒙古气象科学研究所
11. Micaps本地化开发与应用、内蒙古气象、148期、39-41、2002.4、第一、一般刊物、内蒙古气象台