

天气预报的概率统计基础

(上册)

河北省气象局統計预报训练班

1974. 1.

目 录

一、现代统计预报的兴起原因和发展

二、现状

1. 因子选择

2. 预报模型

3. 预报检验

4. 趋势

三、几个基本理论问题

1. 可预报性

2. 统计动力学

3. 概率误差

四、前景的展望

现代统计预报的进展

一、现代统计预报的兴起

在气象学中统计预报并不是一件新鲜东西，在20-40年代中曾经流行一时的瓦尔德（Walker）保尔（Baur），维洛（Bjelle）等学派，都是采用统计方法作天气预报的。但是，现代统计预报对于上述古典统计预报在下一方面有了主要发展。

第一，统计预报的基本原理是从大量历史资料（样品）中应用统计方法寻找统计规律和关系。现代统计预报采用了大数定律和统计假设这一原理得以实现，并使它建立在一些客观可靠和稳定的统计假设上。

第二，统计预报在二十年来有很大的发展。不仅是经典统计，这个部分，而且新的分支，如多元统计。例如，在随机函数理论中又引进了多元统计的初步概念，如多元，线性和多元规划等，这些都给统计工作带来了有力工具。

第三，由于天气学本身的发展，特别是动力气象，天气学，数值预报等的发展，使人们对天气更复杂的物理过程有了进一步的认识，这就使统计预报更具有物理内容，为摆脱形式主义提供以物理基础。

由于以上三方面原因，使统计预报在最近二十年来特别是

最近十年来有了较大的发展。各国都很重视这方面工作，因而取得了显著成效。在业务预报中已经是可与数值预报相比拟的一种客观预报方法。特别是预报天气现象方面效果更为突出。美国目前运用统计方法对降水、最高最低气温、海浪、海浪高度乃至能见度，云高等方面业务预报，已逾数年。尤其是降水和最高最低气温等预报方案，通过实践，几经修改，从最近收到的报告来看，预报效果较好，并计划大力发展。这个国家除了许多高等学校研究统计预报外，还有两个机构大量进行统计预报的研究与服务。一个是气象局内的技术发展研究室，另一个是“旅行者研究中心”的研究室。二者结合数值预报和气象学的现代成就，制作统计预报。1970年美国提出计划，要在全国建立自动预报系统，预报结果随观测电报发到全国各地气象台，并建立“天气计算机”自动回答人们利用电话询问的各类天气预报。实践的需要推动了学术研究，1970年美国气象学会把统计预报列为新学科，提请会员注意发展近代统计预报，引用新概念新工具，推动统计预报向前发展。1968年到1971年三年中，美国召开过两次国际性统计预报会议，公开出版过的统计预报专著至少有三本。

如果说美国统计预报技术较为发达，那么苏联在理论上则较有基础，他们许多统计数学家同时就是气象学家，有一批理

论家从事这方面研究，他们这方面工作开展也较早，据目前了解，有三个全国性机构设有统计预报研究的专门组织或个人。它们是苏联水文气象科学研究中心，地球物理现象总台和苏联科学院大气物理研究所。此外在乌克兰，中亚细亚等各加盟共和国科学院分院也有专门研究人员。除了在数值预报大气环流等会议上讨论统计预报外，1969年还专门召开过全国性统计预报会议，出版了会议文集，苏联从六十年代末期在水文气象中心开始出版一种期刊，《ТРУДЫ ИНСТИТУТА - ГЕОФИЗИКИ СССР》去要发表统计预报研究成果。此外，别的气象刊物也常发表专门研究结果。从1967年以来至少出版了五本统计预报方面的书。

在业务预报方面，在六十年代初期，开始用经验影响函数作气压形势场预报，后来中央预报研究所开始作温度有距平的预报，至今仍在应用和改进中。六十年代末期在中央预报研究所，地球物理现象总台，开始利用自然正交函数开展涛涛利预报因子，预报月降水距平，高空形势场，最近又与数值预报站合作，苏联欧洲地区降水短期预报。

在日本统计预报开展也较早，统计预报研究的领域也较广，他们比较注意培养干部，曾多次举办统计预报训练班，对台风，降水，雪，能见度都作了统计预报试验，对判别分析有广泛研

究。

此外，在联合国气象组织和瑞士，英国等国家都广泛开展了统计预报。

在六十年代我国就有一些统计预报研究工作出现。例如中央气象局和复旦大学关于我国长江流域汛期降水的时间序列研究；上海气象台关于台风路径预报；吉林省关于单站统计预报研究和简化时间序列方法；北京大学、南京大学关于我国气候特征、降水概率、正交分解和长期预报方面的研究著作。但是，系统地有计划地开展统计预报的业务和研究工作，还是最近三、四年的事。从1969年起，大气物理所在过去单站预报的基础上，分别与上海嘉定、常熟、杭州、平阳、武昌等气象台站同志进行了统计预报的试验研究和业务工作；目前，上海、江苏、浙江、安徽、湖北、云南、贵州、四川、湖南、福建等省市普遍开展了这项工作，对降水、暴雨、连阴雨、大风、台风、冰雹等天气都进行了预报。一个有全国各地气象台站的实际工作者参加的统计预报训练班将在今年内举行。为了满足我国统计预报工作蓬勃发展的迫切需要，南京气象学院翻译了统计理论方面的专著，大气物理所以也编著了一本统计预报的专门书籍，将在今年出版。

二、统计预报的现状

在作统计预报时，首先遇到的问题是什么因子，也就是根据什么作预报的问题。在一大堆复杂的大气现象中，找出与预报对象有较好关系的有物理内容的那些现象，作为预报的根据，然后采用一定的统计数学方法，如回归判别等工具，将所选择的因子与预报量之间建立起量观的联系，即预报模型。但是这个预报模型能否反映客观过程，预报是否有效，还有待于实践去检验。这就要看它与气象历史资料相符合的程度以及在预报过程中所表现的正确程度，等等。如果两者不相符合，或出现问题，就要重新选择因子，或修改预报模型。如此循环多次，一次比一次完善，这就是作一个统计预报的过程。从这个过程，我们可以看到一个完整的统计预报系统是由以下三个部分组成的，即：（1）因子选择，资料处理；（2）统计数学模型；（3）预报效果检验。以下我们将按照上述三个方面简单介绍一下现代统计预报的现状。

I、因子选择：

在系统预报中，数学模型是一种信息的加工厂。所以因子选择对于统计预报的成败是一个很重要的关键问题。同时，由于现代统计预报采用大量资料，如何利用机器从浩繁的资料从

中选取主要的因子就是十分必要的了。因此 在现代统计预报中，因子选择就成为重要的、继而列为研究的专题。

目前最常用筛选方法是逐步回归方法，在各国土气预报中已广泛采用。如美国 目前降水、最高最低气温和月总一送大气预报（如云量、能见度等），都是用这种方法筛选因子。在我国国内，如上海和新江的大风、台风预报，福建和安徽等省的降水预报以及大气物理所在武昌所作的降水预报等，也都是采用这种方法。当然，在其发展过程中，在技巧上是在不断的有所改进的。例如：美国在作 108 个城市的降水预告中是采用成对因子筛选方法；首先在原有成对因子中取出最好的一对，把成对因子中的最好的一个因子保留在回归方程中，然后将剩下的所有成对因子与第一个已选因子联合进行选择，再将最好的一对因子中较好的一个因子保留在回归方程中。如此反复进行选择，直到满意为止。在大气物理所所作的试验中，则采用几个因子组合的办法进行选择。这个方法的特点是：每个组合因子全部作为一个因子全部进入回归方程，而不是部分地进入回归方程。

另一种筛选因子方法是正交筛选。这种方法是 Miller 首先引入统计预报中的。它要求因子之间互相正交。从预报信息论角度来看，似乎较好，但在最近大气物理所的试验中，已经

证明，正交筛选与逐步回归筛选基本上是一致的。

不久前，在统计预报中又有人提出了一种简化的选择因子方法。这种方法不需要计算相关矩阵和正交化，它比逐步回归简单。只需要计算一系列相关系数，每个因子可以重复筛选，同时当完成因子选择时，预报方程也就推导出来了。最近，这个方法的使用者之一，在美国西部10个站降水预报中，采用这个方法与回归相结合从4500个因子中筛选出5个因子作为预报降水量的因子，而得到好效果。当提供可能筛选因子数很大时（例如作面上预报时就是这种情况），用逐步回归来筛选往往需要相当长的时间，而先采用这种简化方法进行筛选，然后再做回归筛选就可以避免这个问题，用这种简化方法进行因子筛选，叫做阶段回归。两者结合使用的方法叫做阶段回归与逐步回归联合筛选。

十多年以前就有人从信息论角度来改进正交筛选和逐步回归。但在实际上没有得到广泛应用。

综上所述，无论正交筛选或简化方法，基本上都是逐步回归方法的延伸和发展，都是属于回归这一类概念的。但是，最近从正交函数来选择的方法得到了迅速发展。正交函数中以切比雪夫多项式和自然正交函数最为流行。我国中央气象局和苏联都有人利用天气各形场展开的切比雪夫多项式的主要项係

数作为因子，用马尔可夫链预报天气形势。然而，近年来用得较多的是自然正交函数组。苏联有人曾证明在分组选择因子中，自然正交组是最佳组合。美国在十多年前曾提出自然正交函数开展进行统计动力预报，但未取得可行的结果。目前，苏联广泛采用这种开展作为因子，用回归方法预报西伯利亚、苏联欧洲部分等地区的短、中期天气形势。最近两年，苏联将数值预报的作等压面高度场温度场与垂直速度场进行自然正交开展，用主要项作为因子预报苏联欧洲部分的降水，效果不错。苏联还有人提出将三阶级数与自然正交函数相结合进行场分解的工作。场的单要素与单站多元时间序列在自然正交开展上具有相同性质。大气物理所根据这样想法于1971年在上海嘉定对多元时间序列进行自然正交分解来作春播期和梅雨期的预报，其后在常熟也开展了类似的工作。最近美国和苏联在河流流量和热带波动过程方面都开展了类似的工作。

我国广大气象台站常用的点聚明和单相关系数挑选预报因子方法具有很大的实用价值，在日本这种方法较为流行。

以上讲的是着题挑选因子的方法。

最近，苏联乌克兰有人提出分辨法（判别分析）专有的挑选因子方法。日本有人在降雪预报中采用分辨法原则来挑选单因子。大气物理所在嘉定的试验中所有的挑选方法基本相同。

此外许多人利用时间序列的相关函数分析选取相关性最大的几个因子作为预报因子。时间序列的平滑、过滤和其它处理方法，在近年来得到了很好的发展。几种过滤气象噪声的方法已经提出，其中卡门滤波在气象中的应用是一个值得注意的方法。

II. 预报模型

所谓预报模型就是综合预报因子作出预报判断的方法，即一般所谓统计预报方法，归纳起来有以下几方面：

1. 回归方程

2. 事件概率方法

3. 判别分析

4. 相似演算

5. 随机过程 马尔可夫链随机过程和类似马尔可夫过程

现就以上五个方面分别论述之

回归方程是种统计预报中应用最广泛的一种方法，目前美国、日本、苏联在业务预报中大多数都采用回归方程作预报。如美国108个城市降水预报，西部10个站的总降水量预报，最高最低气温预报，苏联欧洲部分降水预报以及形势场预报都是用这种方法。我国上海、福建、浙江、湖北、安庆等省在预报

大风、台风、降水等也用这种方法。最近有人提出所谓事件概率回归方法。预报方程仍是回归方程，但预报结果是事件发生的概率，而不是预报对象本身数量，这一方法被认为是回归方法的一个发展，经过对降水、云高等试验，效果不错，大气物理所在上海、武汉、成都等地作了试验，结果比一般回归方程好些，1971年美国对108个城市降水预报也做了这样试验。

引用概率逻辑方法预报天气最近有了发展，首先利用马尔可夫链计算事件的转移概率，然后利用统计方法估计天气事件出现的概率，近年来在英美和瑞士进行了许多事件条件概率计算，用贝叶斯定理和其它统计方法预报的工作愈来愈多。大气物理所提出的多因子综合相关法也是概率预报的一种变形。利用概率逻辑运算便于在电子计算机上进行天气预报，目前在苏联和美国已经开始研究，提出了几种根据自动控制系统的“自学”原理 (самобучающаяся, адаптивная система) 预报天气的具体方案，有些地方还进行了试验，其中得到子样“相似”概念的提示是一个重要结果。

统计决策函数理论在天气预报中受到了重视，这种方法能估计所作天气预报有多大把握，目前在试行天气预报开始应用，对雷暴预报已有了一定的结果。

一个天气预报的统计“集成”简单方案由美国人提出，大

物理过程进行了深入研究。

判别分析是一个较老方法，但是直到四十年代才广泛作为一个基本的统计方法。早期用这种方法例有美国的气象和海洋学天气预报、日本台风、暴雨和其它天气预报，苏联则用来分辨锋面存在和识别大尺度天气型等，近年来例有苏联欧洲部分各个站的气象预报、日本的降雪预报，根据雷达回波预报暴雨等。大气物理所在上海嘉定、常熟、武昌等地降水预报试验以后，江苏省和南京大学等预报水害也用了这种方法。苏联列宁格勒观测台近年来正在从多方面进行这项工作。判别分析在理论研究上也很活跃，美国和日本有人研究了这种方法与回归分析关系，并发展了这一方法。苏联有人研究了非线性分辨。大气物理所在嘉定试验中根据天气预报性质，在分辨基本原则上进行了修正，提出了两种改进方案，在预报实践中收到了一定效果。最近苏联又有人提出利用概率逻辑与判别分析结合进行天气预报的自动化和天气气候区域的自动划分。

相似方法在预报中应用较广历史较长，但在选择什么相似指标和相似判据问题上历来就有很不同的作法，近年来苏联在这方面进行了许多工作，但是真正在业务预报中能接受的虽不多。至今苏联气象学界这方面仍是一个很活跃领域，有人根据动力方程，提出相似判据，可能是一个重要进展。在业务预报

中美英用来做长期预报，很受重视，由于这方面不是本文范围，长期预报，另有专文讨论。这就不多说了，只有一点应当提出，相似预报在业务预告中很有实用价值，在我国理论研究中几乎是空白，这种状况应当改变。

随机过程对天气预报具有重要意义。时间序列在气象学中有很多分析，平稳时间序列作天气预报亦很普遍。六十年代中期在美国开始采用平稳时间序列预报气温，近来降水和航空气象预报也逐步采用。我国中央气象局和复旦大学六十年代中期有人对上海一站六、七月降水和长江流域五站汛期降水作过平稳时间序列预报，最近几年大气物理仍在上海、武昌、常熟作了一系列这种预报。苏联对高空环流指数作了谱分析，并进行预报试验。

由于天气有时具有突变的特征，所以非平稳过程预报对天气预报具有一定意义，苏联采用函数逼近谱分析的结果，得出气象要素时间变化的函数，以预报天气是这种方向的一个努力，大气物理而采用自然正交函数分解时间序列并用回归方法进行预报，和采用时间序列马尔可夫过程预报，都是企图克服这种困难的一个尝试。

天气具有多年变化，利用前期时间序列作预报时间较短，不能反映出这种年变化。北欧的几个国家有人研究了时间序列

前期与多年变化相结合的时间序列预报方法。

用周期图，频谱分析作预报在各国都流行，最近美国有人用功率谱分析热带波动，大气物理所用这种方法分析了上海春播期连阴雨波浪特性。并对这段时期利用自然正交开展，进行了波浪分析，最近看到美国和苏联亦发展了类似工作。一次自然正交开展与多种谐波相结合的时间序列预报工作已正在由大气物理所和上海气象台协作开展。

另一种随机过程是马尔可夫过程。在天气中较多的应用还只限于最简单马尔可夫过程即马尔可夫链。利用马尔可夫链计算干日与湿日（即晴、雨）和其它天气变化概率已经有很久历史，至今仍在进行。苏联不少人利用马尔可夫链研究了西风环流变化，又有人根据等压面形势场的切比雪夫系数研究了形势场的转变，根据马尔可夫遍历性理论研究了天气形势的稳定性。我国中央气象局也有人作了类似研究。此外，在么槐生关于马尔可夫链初始状态定义的启发下，大气物理所有人进行了时间序列马尔可夫链转移概率的计算，对浙江地区晴雨作了预报试验。

III 预报检验

在系统预报中对于预报检验是十分重视的。在实际工作中常用的有：

成败百分比 绝对误差, 相对误差, 均方差, 标准误差, 相关系数。这些方法都很简单, 各有其缺点, 其中百分比方法缺点较大。正以许多研究后指示, 这种方法使得预报人员不重视稀有天气的预报。

在业务预报中美国常用复相关系数的平方, 而苏联直接应用相关系数, 或经验回归系数。但是各种特殊天气预报常用特殊方法, 如预报温度和气压则用绝对误差, 预报风场则采用均方向角误差与平均风速之比。如果把风速换成低压中心移动速度, 则上述公式同样可作为台风和低压中心预报检验的方法。在概率预报中常采用的是 Brier, Score 方法 (概率预报中概率的标准误差的平方)。正反预报检验方法是目前检验最成熟的方法, 两类表很通用。但现在常用的是 $O\delta y \times O\delta$ 的准确性判据。

各种检验的统计方法已被揭示, 研究工作着设计了各种方案, 目前认为最有成效的是使熵论和熵的概念引入预报检验中来, 这种思想在我国很早就有人研究过。

各种思想不断深入, 其中在预报检验中不仅考虑预报本身, 而且还要估价预报使用经济效果的思想被公认为是必要的。

四、趋势

在分析了各国统计预报现状后, 我们得到如下印象, 当前统计预报在发展趋势上有如下趋势:

(1) 统计预报的物理化

现代统计预报兴起的头几十年, 曾经有过这样的乐观情绪, 认为只要有了足够大的计算机, 让机器从一切有关资料中寻找因

天气预报可迎刃而解, 经过几年试验, 目前各国业务预报

一致性的由浅到深，预报因子一般都不多，方法多为回归和判别分析之类，似着重在每个因子物理内容，所以因子选择上十分讲究，这更可以从美国发展起来的西部10个地方降水预报方法为代表。在回归方程中只用了5个因子。在4500个可能因子中先采用简化方法进行筛选，每个因子选入以前从天气形势、地理位置方面去考虑每个因子物理意义，并不是所有相关系数高的都能选入，还要看是否有物理意义，例如当选到第四个因子时，850 mb 在 California 东部1000英里地方相关系数为0.192，比这个地方西部300英里的地方相关系数0.159要高，但考虑到温度平流的物理意义，则第四个因子采用相关系数较小的上游那个点的温度作为第四个因子，又如在同一报告中对另外一种降水预报公式用上述办法仔细挑选了9个，然后在逐步回归时选取4个作预报，每个因子均有很明显物理内容。所以当前在统计预报中力求方法简明，预报因子少而物理意义明确。所以在统计预报中广泛进行天气动力学分析，力求预报具有明显物理内容，避免繁琐数学是十分重要的。

(2) 注意采用数值预报结果

美国技术发展研究室提示三种统计预报方法：

- 1) 经典统计方法：用统计方法根据当天资料预报明天天气。