

# 現代統計預報的進展

## 一、現代統計預報的發展

在氣象學中統計預報並不是一件新鮮東西。在20—40年代中曾經流行一時的瓦爾克、保爾、維澤等學派，都是採用統計方法作天氣預報的。現代統計預報在以下三方面又有了新的發展：

一、統計預報的基本原則是从大量歷史資料（子樣）中運用統計方法尋找統計規律和關係。現代統計預報採用了大型電子計算機，使它建立在客觀而穩定的統計關係上。

二、統計數學近20年來有很大發展。如在隨機函數理論中出現了概率理論的動力學部分，信息論，線性和動力規劃等。這給統計預報帶來了有力工具。

三、由於天氣預報本身發展，特別是動力氣象學、天氣學、數值預報等方面的发展，使人們對天氣演變的物理過程有了進一步的認識，這就為統計預報提供了物理基礎，使它更具有物理內容。

由於以上三方面原因，使統計預報在近20年來特別是最近10年來有了較大的發展。各國都很重視這方面工作，並取得了顯著成效。美國目前運用統計方法對降水，最高最低氣溫、海浪、海浪高度乃至能見度，雲高等方面作業預報已多年，尤其對降水和最高最低氣溫等預報效果較好。1970年美國提出計劃要在全國建立自動預報系統；美國氣象學會還把統計預報列為新學科。1968年到1972年四年中，美國召開過三次國際性統計預報會議，公開出版過統計預報專著。

蘇聯統計預報在理論上較有基礎，在這方面工作开展也較早。據了解，有三個全國性機構即：蘇聯水文氣象科學研究中心、地球物理觀象總台和蘇聯科學院大氣物理研究所設有統計預報研究的專門組織。1969年召開過全國性統計預報會議，出版了會議文集，水文氣象中心出版的期刊，主要發表統計預報研究成果。在業務預報方面，60年代初期開始用經驗影響函數作氣壓場預報，後來中央預報研究所開始作旬溫度距平的預報。60年代末在中央預報研究所、地球物理觀象總台，開始利用自然正交函數開展場作為預報因子，預報月降水距平，高空形勢場，最近又與數值預報結合作蘇聯歐洲地區降水短期報告。

在日本，統計預報開展較早，研究的領域也較廣，曾多次舉辦統計預報訓練班，對台風、雨、雪、能見度都作了統計預報試驗研究。

在60年代，我國就進行了一些統計預報研究工作。如中央氣象局和復旦大學關於我國長江流域汛期降水的时间序列研究；上海氣象台關於台風路徑預報；吉林省關於單站統計預報研究和簡化時間序列方法；北京大學、南京大學關於我國氣候特點，降水概率，正交分解和長期預報方面的研究等等。近三、四年來系統地有計劃地開展了統計預報的業務和研究工作。

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、武昌等气象台站的同志进行了统计预报的试验研究和服务工作。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。

## 二、统计预报的现状

在作统计预报时，首先遇到的问题是预报因子，也就是根据什么作预报的问题。在复杂的大气现象中，找出与预报对象有较好关系的物理因子，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客观的联系，即预报模型。它能否反映客观过程，预报是否有效，还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，或者出现别的问题，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，这就是进行统计预报的过程。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。以下简单介绍一下现代统计预报的现状。

### （一）因子选择

在统计预报中，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个重要的关键问题。现代统计预报采用大量资料，因子选择就成为重要的研究专题之一。

目前最常用的筛选方法是逐步回归方法，在各国业务预报中已广泛采用。如美国，目前降水、最高最低气温和其它一些天气预报（如云高、能见度等）都是用这种方法筛选因子。在我国内，如上海和浙江的大风、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。在发展过程中，技巧上是不断改进的。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；首先在所有成对因子中取出最好的一对，把其中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，每个因子组合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。这种方法是米勒（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变化；它比逐步回归简单，只需要计算一系列相关系数，每个因子可以重复选进，同时当完成因子选择时，预报方程也就推导出来了。当提供的可能筛选因子数量很大时（如作大面积预报时），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的预选，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。

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图形势场展开的切比雪夫多项式的主要项系数作为因子,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。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,苏联将数值预报所作等压面高度场、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量,效果不错。苏联还有人提出将三角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。大气物理所根据这个方法,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图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。

以上讲的是普遍挑选因子的方法。

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。我们在嘉定的试验中所用的挑选方法与此基本相同。

北欧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点作为预报因子,时间序列的平滑、过滤和其他处理方法,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。

## (二) 预报模型

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,归纳起来有以下几个方面:

1. 回归方程 是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预报都是用这种方法。我国上海、福建、浙江、湖北、安徽等省、市,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,效果比一般回归方程好些。

2. 事件概率方法 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。中国科学院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”原理预报天气的具体方案,有些地方还进行了试验,其中所谓子样“相似”概念的提出是一个重要结果。

3. 判别分析 是一个较老的方法,但是它与回归分析一样,也是一个基本的统计方法。

早期用这种方法的有美国的云向和单站降水预报，日本的台风、暴雨和其它天气预报；苏联用以分辨锋面存在和研究大尺度天气型等。近年来的工作有苏联欧洲部分 40 个站的月降水量预报，日本的降雪预报，根据雷达回波预报雷暴等。我们在上海嘉定，常熟，武昌等地降水预报试验后，江苏省和南京大学等预报冰雹也用了这种方法。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，并有所发展，苏联有人研究了非线性分辨。大气物理所在试验中根据天气预报特点在分辨基本原则进行了修正，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。

4. 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不同作法，20 年来苏联在这方面进行了许多工作，但真正在纯业务预报中运用的并不多。也有人根据动力方程提出相似判据，在业务预报中，英、美用来作长期预报。相似预报在业务预报中很有实用价值，在我国理论研究中几乎是个空白，这种状况应当改变。

5. 随机过程 它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也很普遍。60 年代中期在美国开始采用平稳时间序列预报气温，近来对降水和航空天气预报也逐步采用。我国中央气象局和复旦大学在 60 年代中期对上海某站 6、7 月降水和长江流域 5 个站汛期降水作过平稳时间序列预报。最近几年大气物理所也作了一系列这种预报。

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来预报天气，我们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一种尝试。

天气具有多年变化，利用前期时间序列作预报，时间较短，不能反应出这种年变化。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，我们用这种方法分析了上海春播期连阴雨波谱的特性，并对这段时间利用自然正交开展，进行了波谱分析。目前大气物理所和上海气象台正协作进行自然正交开展与多种滤波相结合的时间序列预报工作。

另一种随机过程是马尔可夫过程。在天气预报中较多的应用是马尔可夫过程即马尔可夫链。利用它计算晴、雨和其它天气变化概率已经有很久历史，苏联不少人利用马尔可夫链研究西风环流变化，还有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和稳定性。我国中央气象局也有人作了类似研究。大气物理所曾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。

### (三) 预报检验

在统计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，各有优缺点，其中百分比方法缺点较大，使预报人员不够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量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布赖尔评分方法（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。但现在常用的是奥布霍夫的准确性判据。

对检验统计方法，研究工作者提出了各种方案，目前认为最有成效的是信息论和熵的概念引入预报检验中来。

#### （四）趋势

当前统计预报在发展上有如下趋势：

##### 1. 统计预报的物理化

现代统计预报发展初期，有人认为只要有足够大的计算机，让机器从一切有关资料中寻找因子，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趋向由繁到简，预报因子一般都不多，方法多为回归和判别分析之类但着重在每个因子物理内容。这点可从美国最近产生的西部 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 4,500 个可能因子中先采用简化方法预选，每个因子选入前从天气形势、地理位置方面去考虑每个因子的物理意义，而不是所有相关系数高的都能选入。如选到第四个因子时 850 mb 在加利福尼亚东部 1,000 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，但考虑到涡度平流的物理意义，第四个因子采用了相关系数较小的那个点，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，然后再逐步回归选取 4 个作预报，每个因子均有明显的物理内容。所以在统计预报中力求方法简明，预报因子少而物理意义明确。

##### 2. 注意采用数值预报结果：

美国技术发展研究室提出三种统计预报方法：

i) 经典统计方法：用统计方法根据当天资料预报明天天气。

ii) 完全预报方法：根据历史资料将明天环流与明天天气构成统计预报关系，然后利用数值预报所预报的明天环流，作为信息代入方程预报次日天气。

iii) 不完全预报方法：与第二种方法不同，仅在于当天历史资料也是采用数值预报中积累的资料，而不是观测到的历史资料。

后二种方法都用到数值预报的结果，其基本思想是天气现象与环流的同时相关性远比时间先后的相关要好得多。后二种近来越来越多地采用，随着数值预报资料累积增加，这种应用将更为广泛。目前美国的降水，气温等业务预报，经多年试验都在不同程度上改用数值预报结果。气温和降水预报都采用六层模式的预报结果。

##### 3. 注意理论研究和采用近代统计数学新成果注意新概念新技术的引用。

### 三、几个基本理论问题

以下几个理论问题，在统计预报中普遍受到注意：

1. 大气可预报性：实质上是大气过程发展究竟在多大程度上具有统计规律性问题。在长时间内可以认为所有质点坐标和速度将遍历任何一组可能的值，于是初始条件的影响彼此抵消和互相抵消，所以在一段长时间间隔内来考察这类运动，则真正初始条件并不起任何作用，

因此说明这类运动具有统计规律，近年来由于统计预报和数值预报的发展，这个问题很受注意，但还未达到成熟阶段。

## 2. 大气大尺度运动的统计动力学：

这种观点认为既然数值预报和统计预报天气都有一定准确率，二者都能部分反映大气运动规律，因此二者结合，取长补短，形成统计动力学。

各种统计动力学模型已经提出，各种理论虽作了数值试验，但离应用还有一段距离，正在深入研究之中。

3. 运用学习机原理的概率逻辑理论和天气预报的自动控制的自学理论：这是多年来许多研究家共同感兴趣的问题。目前正处探索阶段，随着现代科学技术特别是逻辑数学，自动化理论和计算技术的发展迟早是会实现的，其结果是将天气预报完全建立在现代科学技术的基础上。

## 四、展 望

1. 首先在两三年内，美国业务预报的自动系统将完全建立，并在业务中广泛应用。除了目前降水、最高气温之外，其他天气如能见度，云高、风暴、乃至风浪、晴空、湍流都将广泛应用。

2. 自动控制系统的“自学”预报将逐渐得到发展。

3. 统计动力学的大气环流理论和天气预报方法将得到显著成果。

4. 统计预报业务将在世界各国普遍建立，我国的统计预报在业务上将成为一种重要方法。

(中国科学院大气所 李麦村供稿)