

局地天气预报的 数据分析方法

曹鸿兴 编著

气象出版社

局地天气预报的 数据分析方法

曹鸿兴 编著

气象出版社

内 容 简 介

本书论述运用数据分析作局地天气预报的原理和方法。内容包括适合于手算的简易预报方法,逻辑代数、数字滤波和学习机原理等在预报中的应用,以及某些七十年代以来在气象中使用的概率统计方法,如聚类分析、数量化方法、对应分析、非线性映象和最大熵谱分析等。书中还综述了在我国发展的一些预报原理和方法,如预报指标的逻辑结构、模糊数学的应用等。大部分方法都给出了详列计算过程的实例。

全书共分八章。第一章阐述数据分析方法的基本特征,第二章介绍图和表的统计分析,第三章论述时间序列分析的实用方法,第四章至第七章针对不同的数据类型(0-1变量、连续变量和分类变量)讨论各自的分析方法。最后一章论述模糊数学在天气预报中的应用。

本书读者对象为气象工作者和水文、海洋、地震、地质、农林和生物等领域中的统计工作者。也可供有关专业的高等院校师生参考。

局地天气预报的数据分析方法

曹鸿兴 编著

*

气象出版社出版

(北京西郊白石桥路46号)

北京印刷一厂印装 新华书店北京发行所发行

*

开本: 787×1092 1/32 印张: 10.5625 字数: 240 千字

1983年3月第1版 1983年3月第1次印刷

印数: 1—4,000 统一书号: 13194·0090

定价: 1.10元

序

六十年代以来在许多学科中发展了数据分析方法。这是由于实验、观测手段的现代化,人们能取得的数据越来越多,而电子计算机的广泛应用,又使得能通过对大量数据的分析得到手工分析无法获得的有价值的结果。数据分析的内容按学科本身的特点而千差万别,如空间科学中的滤波技术、地球物理中的谱分析、图象识别中的特征抽取原理等,但或多或少都和概率论和数理统计相联系。在气象学中,由于有着大量的观测资料,因此数据分析一直是人们研究气象现象的基本手段。广义地说,天气图就是一种有技巧性的数据分析的产品。天气图分析是一个滤波过程,输入的是原始资料,输出的是包含各种图形的天气图。或者说输入的是信号加噪声(干扰),输出的是信号的估计。如果没有天气图分析,即使在仅一个时次的全球观测的千万个数据中也难理出个头绪来。

运用本地前期征兆来作天气预报已有悠久的历史,从天气谚语到单站预报,都是利用对局地气象的或非气象的数据进行分析来预报未来天气的。在气候变迁的研究中,因受到资料的限制,很多结果是利用单点记录得到的。我国气象台站运用前期征兆来预报局地天气的方法和经验尤为丰富和精细,值得深入总结和提高。作者在总结自己工作成果的基础上,搜集了国内外的有关文献资料,就借助数据分析制作局地天气预报这一专题编著成本书。

为了叙述的方便,我们把气象要素,如气压、温度、湿度;非气象要素,如水温、地电;大气物理量,如涡度、散度等都统称为

变量。在涉及预报问题时，则称用作预报依据的变量为预报因子，要预报的变量称为预报量或预报对象。把一个气象台或气象站能收集到的所有目测和器测结果，如天象、物象以及本站和邻近站的记录乃至传真接收的数值预报制品和卫星资料，均称为数据。数据分析的含意当然是极为广泛的，但我们只论及预报局地天气的这类方法，所以显然与运用大范围空间和多个时刻的资料的天气图方法不同，亦与只运用一个初始场的求解动力-热力方程的数值预报方法相异，与从数学角度来区分的概率统计方法也不尽相同。就数学原理而言，除了象回归、判别分析和时间序列分析等统计方法外，还叙述了模糊数学、逻辑代数、人工智能中的学习机原理等，我们把这些方法统称为数据分析方法，同时把论题限定在预报局地天气上。

从编著本书的上述宗旨出发，必然要用到一些必要的数学和天气学知识。假定读者已有这类知识或自行参考有关的书籍，仅需掌握操作使用的读者，可以跳过理论叙述，直接阅读例子。因为编写时注意到了使实际例子相对于数学原理有一定的独立性，有些方法则直接用实例来阐述，以便于从事实际工作的同志参考。

在编写过程中得到我所张家诚、陈国范等同志的指导和帮助，王淑凤、张树荣同志为本书绘制了插图，作者深表谢意。本书引用了不少自印材料，不再一一列出，在此一併致谢。本书从初稿写成至今，几经易稿，疏忽和错误在所难免，望读者批评指正。

编著者

1981年8月于北京

目 录

第一章 绪论	1
§ 1 问题的提法和一般思路	1
§ 2 单站数据的某些基本特性	3
§ 3 数据的类别及其统计处理	9
§ 4 信息意义下的天气预报	11
§ 5 统计预报过程	15
§ 6 可预报性与可达水平	19
第二章 图表分析	25
§ 1 表格的分析与综合	26
§ 2 图形的数学表示	32
§ 3 多点聚图准确性的概率估计	34
§ 4 多变量作图分析	41
§ 5 非线性映象	44
第三章 时间序列分析	51
§ 1 周期与韵律	51
§ 2 转移概率的应用	55
§ 3 数字滤波	65
§ 4 平稳时间序列和最大熵谱分析	73
§ 5 多维时间序列	92
第四章 0-1 数据分析	103
§ 1 用 0-1 数据作因子粗选	103
§ 2 逻辑组合因子的运用	109
§ 3 预报指标的逻辑结构	116
§ 4 0-1 型综合	125
第五章 连续变量的多因子预报	133

§ 1	多元回归的基本原理	133
§ 2	逐步回归的原理和实例	139
§ 3	一个节省计算量的正交筛选方案	149
§ 4	数量化方法	156
§ 5	用方差分析估计预报量	163
§ 6	序数相关和回归	170
§ 7	正交函数分解	175
§ 8	对应分析	191
§ 9	数值预报结果在局地天气预报中的使用	200
第六章	分类预报	211
§ 1	划类和判类的基本原理	211
§ 2	计数法	215
§ 3	线性学习机原理在预报中的应用	221
§ 4	不同准则的判别分析	232
§ 5	逐步判别	244
第七章	客观分类	252
§ 1	天气预报中的客观分类问题	252
§ 2	聚类方法及其应用	259
§ 3	有序样本的聚类分析	266
§ 4	相似分析	276
第八章	模糊数学在预报中的应用	282
§ 1	模糊集的基本原理	283
§ 2	运用隶属原则和择近原则识别天气类型	293
§ 3	多变量的模糊综合	300
§ 4	模糊聚类分析	310
附录		326
表 1 至表 6		326

第一章 绪 论

为了使运用数据分析作局地天气预报的各种各样的方法有一定的联系,不但看到一个一个的具体方法,而且能看到其总的图象,在本章我们将论述这些方法所具有的共同的基本特征。

§ 1 问题的提法和一般思路

我们的论题是运用本站所能获取的数据来制作局地天气预报。因此,它既不象数值预报那样,把预报问题提为一个空间场的初值问题;也不象天气图方法那样,从气象要素的空间分布上追踪系统随时间的移动和演变,以推断未来天气。自然我们会从分析数据随时间的变化上着眼,即把预报问题提为一个数据序列问题或局地要素的演变问题。

设用作预报因子的有:如本站气压 P 、温度 T ,某动物的反应 r ;某指标站的500毫巴高度 H ;数值预报的涡度 ξ 等。要预报的对象如降水 R 、风力大小 W 等。按时间顺序将它们排列为:

$$\begin{pmatrix} P_0 & P_1 & \dots & P_t & P_{t+l} \\ T_0 & T_1 & \dots & T_t & T_{t+l} \\ r_0 & r_1 & \dots & r_t & r_{t+l} \\ H_0 & H_1 & \dots & H_t & H_{t+l} \\ \xi_0 & \xi_1 & \dots & \xi_t & \xi_{t+l} \\ \dots & \dots & \dots & \dots & \dots \\ R_0 & R_1 & \dots & R_t & R_{t+l} \\ W_0 & W_1 & \dots & W_t & W_{t+l} \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

把预报因子记为 $X=(x_0, x_1, \dots, x_l, x_{l+1})$, 预报对象记为 $Y=(y_0, y_1, \dots, y_l, y_{l+1})$, ($l=1, 2, \dots, L$, 预报时效)。现在的问题是, 根据 t 时刻以前的数据来推断 $t+l$ 时刻的 y 值, 显然问题归结为对向量 x 和 y 随时间的变化进行分析, 寻找函数关系。

$$y_{l+1}=f(x, y) \quad (l=1, 2, \dots, L)$$

为了获得与实际情况相一致的函数关系, 需要对数据序列进行物理分析或统计分析。物理分析能使我们了解对影响本地的天气过程有一了解, 而统计分析则能得到一定的数量关系。

数据序列分析可以比喻地用“温故而知新”这一成语来表达, 其形象的直观图象如图 1.1。立足于现在, 一只眼睛瞧着过去, 一只眼睛盯着将来, 通过对过去和现在数据的分析来预测未来。

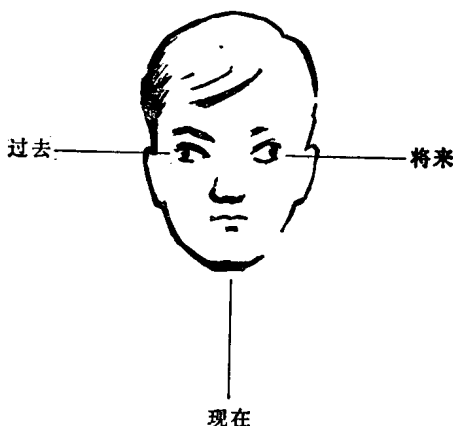


图 1.1 数据序列的形象表示

运用过去和现在数据作未来天气预报的方法, 依其历史发展阶段和固有特征分为三类:

第一类是经验性推断, 即对前期征兆通过分析获得一些经

验性规则,实际使用时主观地运用这些规则进行判断,例如看天象、物象用天气谚语作预报就属于这一类。

第二类是资料图表法,以点聚图、要素曲线图、综合时间剖面图等图表为主要工具,寻找要素变化的特征,如峰、谷、交叉点、变率周期性等,由此建立针对不同天气的预报模式和指标,所以也称为模式指标法。

第三类是定量计算法,凭藉不同的数学模型,对大量资料进行数字处理,通过计算建立适合于不同预报项目的公式和判据,以此来预报未来天气。

这三类方法表示了由经验、定性到定量的发展过程,但其共同特点都是对表征前期征兆的历史数据进行统计分析以获得定性或定量的规则(或公式),由此来推断天气的未来。

本书将着重论述第三类方法,即从不同的数学原理出发,建立适合于不同性质的数据的模型,叙述建立这类不同于动力模式和天气学模式的预报模型的原理以及凭借这些模型来制作局地天气预报的方法。这些模型可以分成自因分析和他因分析。自因分析是对变量自身演变规律的分析,在概率统计中属于时间序列分析的范围。他因分析着眼于不同变量之间的关系。就预报而言,要分析预报因子与预报量之间的关系,多元统计分析、布尔代数和模糊集原理等都是研究这类问题的。

§ 2 单站数据的某些基本特性

单站数据是指一个测站对气象要素或与气象有关的量所作观测结果的集合,它可以是人工观测的原始记录或加工数字,可以是自记仪器上的自动记录,也可以是拍摄的照片或录音,乃至由卫星、雷达观测尔后传递来的关于本站气象实况的情报。单站天气数据是不同天气系统影响本站的各物理量的测量结果。

粗看起来,它仅仅是一串数字、一条曲线或几张图片;但从辩证唯物论的观点对单站数据作一番科学分析,就会发现它包含着反映本地天气和气候特征的许多信息,它反映了大气运动的规律。这些信息和规律正是我们作预报时所需要的。下面分几个方面来说明单站数据的基本特性。

1. 决定性*和随机性

如果把几年的逐日 14 点气温 T_{14} 点成一根曲线,就会看到由冬到夏 T_{14} 逐渐升高,由夏到冬 T_{14} 逐渐降低,有一个显著的年变程。这是由于地球绕太阳公转而造成四季变化,这种变化为万有引力定律所决定,对单站来说可以认为是一种确定型变化。在这根曲线上叠加有许多扰动,譬如说 4 月份的气温是逐渐上升的,但由于一次强冷空气侵袭,可能在 4 月中旬某日气温达到该月的最低点。这一变化相对于气温的年变程来说是一种随机型变化,因为产生这一最低点可能在上旬也可能在下旬,只是发生在上旬的机率最大罢了。

决定性和随机性在理论上虽有明确的含意,但就具体的一串观测数据来说,区分决定性和随机性是相对的。例如上述造成气温最低点的冷空气是前期大气运动的结果,伴随冷空气的天气系统可能是一个范围千公里的冷高压,在高空环流操纵下自西向东移动时路经本站,温度达最低是冷高压影响本站的必然结果。因此就短期天气过程而言,这一温度最低点又是决定型的。

其实,决定性和随机性的概念在我们日常预报中是经常使用的,例如为了突出天气系统的影响常把日变程、年变程去掉。

* 或称确定型、决定论的,这是一种由拉普拉斯等创立的认识自然的观点。它认为所有事物的现在状态是先前状态和制约这状态的原因的发展的必然结果。如果我们获得了给定时刻的全部知识,就有可能准确无误地知道事物的过去和将来。

为了使一个数字序列接近平稳随机序列，采用一级变量、二级变量的预处理。滑动平均则是把要素的随机性排除，突出决定型变化。

2. 信息与干扰

这里所谓的信息与干扰是在预报意义下使用这一词的。把天气数据中包含的对未来天气有指示意义的那部分称为信息，对预报无用的部分称为干扰。从这一定义就可想到信息与干扰的划分是相对的，某一要素的某种变化对长期预报来说是一种干扰，对中期预报恰是有用的信息。因此就这种意义来说，建立模式、找预报指标都是为着把天气数据划分为信息和干扰，也就是说，天气分析或统计分析的目的就是要求准确地排除干扰，提取对预报有意义的信息。

如果我们用 10 个因子点了五张点聚图，对图上点子分布作了一番分析，选择了其中分布有序的两张。其余三张点子分布零乱，如用它们作预报只会起到干扰作用，这是我们凭经验划分了信息和干扰。

在用本站要素演变图建立某种天气预报模式的时候，我们通常把图形相似、指标相同的属同一类天气的个例放在一起，画出一张典型图。显然这时综合了这些个例的共同特征，忽略了各自的细节，这张典型图就是预告某类天气出现的信息，而那些细节则是干扰。

在寻找预报指标时，日变程的破坏是对未来天气变化的良好预示，因此对预报来说它就是有用的信息。但对日变程规律来说，这种破坏是一种干扰。所以，提取信息和排除干扰是相反相成的，在排除干扰时要注意抓住对预报有用的信息，在提取信息过程中要善于识别干扰并排除它。

信息和干扰虽是一个抽象的概念，但掌握了它对具体工作

是有益的，它使我们开拓思路。充分利用信息，要求我们获取除本地资料以外的其他数据，如邻近站的、天气图的乃至数值模式的计算结果，促使我们掌握更多的提取信息的有效方法。

3. 综合性与可分性

单站数据是各种天气的和非天气的因素的综合作用结果，如太阳辐射、流经本站的空气属性、本地下垫面的性质等因素。单站数据是全球大气运动在单点的观测值，是整体和局部的矛盾统一的结果，既有天气变化的普遍性，又有本地天气的特殊性。在时间的长河中，它是整个长趋势变化中的一个环节，又是特定时段中的瞬时值。在空间的广袤中，它是大尺度大气运动在一个点的反映，又迭加有本地中小尺度运动的结果。

这种既综合又分解的特性在自然界中是广泛存在的。如太阳光看起来是白色的，经过分光镜后，可以分成红、橙、黄、绿、青、兰、紫等七种颜色，它是多种波长光的混合。同样，观测数据亦可视为多种因素综合作用的结果，是各种时-空尺度运动在单站的综合影响。自然，我们可将它分解，分解结果可用以作相应时-空尺度的天气预报。

单站数据的综合性和可分性，是能用单站数据作不同时效预报的一个理论支柱。这就是说，若滤除数据中的高频部分，就能用它来作中长期预报；反之突出了高频部分，就能用它来预报诸如冰雹、飚线等局地性小尺度天气现象。

4. 空间分布与时间序列

制约天气变化的方程(如动量守恒方程、质量守恒方程、热流量方程)均可写为

$$\frac{\partial x(t_i)}{\partial t} = F(x, \alpha, t_i) \quad (1)$$

式中， x 为某物理量， α 为物理参数，如柯氏参数，粘性系数，静力

稳定度等。当观测为离散时 $t_i = t_0, t_1, \dots$ 。 $F(x, \alpha, t_i)$ 表示 t_i 时刻的空间项, 其中包括如平流项、梯度力项、柯氏力项等。

显然, 单站预报只利用了(1)式左端项即局地物理量在时间上的变化信息, 也就是包含于 $x(t_0), x(t_1), \dots$ 这一串数据中的信息。是从同一地点不同时间的物理量的变化出发作预报的。而数值预报运用特定时刻 t_i 时的(1)式, 即该时刻时间项与空间项的转换关系, 只利用某一时刻包含于空间项中的场信息。是从同一时间、同空间的物理量的相互影响出发作天气预报的。就两个时刻 t_i, t_{i+1} 对(1)分别积分得

$$\begin{aligned} x(t_1) - x(t_0) &= \int_{t_0}^{t_1} F(x, \alpha, t) dt \\ x(t_2) - x(t_1) &= \int_{t_1}^{t_2} F(x, \alpha, t) dt \end{aligned} \quad (2)$$

.....

$$x(t_{n+1}) - x(t_n) = \int_{t_n}^{t_{n+1}} F(x, \alpha, t) dt$$

从(2)式看出, 对单站来说, 分析 $x(t_0), x(t_1), \dots, x(t_{n+1})$ 与对空间项积分的分析是等价的。因此可以说单站要素是全空间大气运动在局地的指示, 它含于空间之中。反之, 气象场是由一个个单点组成的不可分割的整体, 空间场的变化总要反映到局地中来, 亦即空间场耦于局地之中。要素的空间分布与单站要素的时间序列是分别对空间坐标和时间坐标展开的。但由于大气运动的统一性, 其间存在着一定的对应关系和转换关系。一方面, 对空间场变化的连续分析将为单站预报提供依据; 另一方面, 通过对单站时间序列的分析, 将在一定程度上推知空间场的变化。

大气无论在空间上或时间上都具有内在统一性。大气的局地变化和全球大气运动息息相关。一个地区的天气演变势必影

响其他地区大气的运动。这种局地变化的传播是与天气系统的尺度密切有关的。一个小系统如飑线等,只能影响邻近地区,范围在几十公里到几百公里。一个气旋尺度的系统则要在洲际间影响,范围在几百公里到几千公里。大气运动在时间上的变化是连续的,这一时刻的大气运动状态是由上一时刻大气运动演变来的。这表现在不同形势场的前后的承续性,但这种演变的速率是很不相同的。就单站来说,锋面过境时要素的变化与高压控制时要素的变化差异是很大的。气象量在空间与时间上的对应、长时间与短时间的关联、空间与局部的相关,是我们研究局地预报的基本论题之一^[1]。

可以把大气运动尺度粗分为四大类:时间尺度在一小时以下的湍流尺度,一小时至数周左右的天气尺度,数周至自然季节的大天气尺度以及年以上的气候尺度。在天气尺度以下,大气过程的时间尺度与空间尺度近似地保持线性对应关系。也就是说,一个大系统它的生命史也长,一个小系统只有短暂的生命。就空间分布与时间序列而言,一个空间分布广的天气系统在单站记录上经历的时间也长,占有空间范围小的系统在单站记录上只经历短暂的时间。强天气在单站记录上往往表现出大的振幅,弱天气系统则只有小的振幅。这说明,在天气尺度范围内,空间波动性和时间的周期性在一定程度上是等价的。

在大天气尺度和气候尺度,空间坐标为主的天气分析让位于时间坐标为主的气候分析。这时,各种时间长度的平均值将成为基本的分析量。平均量的空间分布是单站均值在空间上的有序排列,是单点值的地理分布。显然,平均量空间分布与单点时间序列间的关系与瞬时情况是不同的,既有物理联系也有统计关系,平均时间越长,统计关系越明显。

关于单站数据基本特性的知识,无疑是有助于作好天气预

报的。

§ 3 数据的类别及其统计处理

气象数据可以粗分为连续变量和非连续变量,前者也称为量的数据,如气压、温度、等压面高度等;后者也叫质的数据,如霜冻的有无、风力等级等。因数据性质不同,对它的统计处理方法必然亦相异,因此研究数据的类别是很重要的。为此进一步把用作天气预报的数据细分为如下四类:

1. 连续变量

气象要素的大部分是连续变量,这可以从要素自记曲线上看出,因为通常大气可视为一个连续介质。连续变量中有的有绝对零点,如温度采用绝对温标,这一类变量间的关系称为间隔尺度;另一类如长度、重量等,需以某一标准来对比以确定其数值,称为比例尺度。

连续变量无论用动力学方法或统计学方法处理都比较容易,因为许多数学、力学研究是针对连续变量的。

2. 名义变量

或称二值变量,0-1 变量,它只取两个值,无大小、次序的概念,如“下雨”和“不下雨”、正距平和负距平,龙卷风的“有”“无”。

3. 有序变量

这一类变量可以用自然数和它相对应,在次序上可排列出,如台风等级,小雨、中雨、大雨、暴雨、大暴雨等的降水量等级。又如我们把某一要素按其历史值的大小顺次排列起来给以值序号,称为序数,也就成为一个有序变量。

4. 分类变量

如大型环流型 W、C、E 型,天气过程分为冷锋型、切变型、气旋波型;锋分为冷锋、暖锋、静止锋、锢囚锋等。这一类变量无

次序而言,亦无大小可比,它与 0-1 变量的区别是取两个以上的值。分类变量是不可排的,而有序变量是可排的,这是两者的区别。

从数学的观点来看,天气预报问题是根据一组已观测到的变量 $(x_1, x_2, \dots, x_m)$ 来预报另一组变量 $(y_1, y_2, \dots, y_n)$ 。前者视为天气变化的原因或预报的凭据,称为说明变量或预报因子;后者视为天气变化的结果,称为基准变量或预报对象、预报量。预报因子和预报对象均可取上述四种变量之一,因此这就构成了不同变量之间的转换和沟通。

四种不同变量间的转换在天气预报中是经常进行的。出于服务的需要,我们常把连续的风速转换为蒲福风级(有序变量)。为了计算简便,把预报因子如某月气温按一界限划分为 0-1。出于分析的需要,计算气压的正、负距平。反之,也可将非连续变量转换为连续变量,如数量化方法、预报因子取 0-1 值时的判别分析、事件概率回归等。

从气象数据随时间的变化来分类,则可分为周期性和非周期性数据。年变化和日变化是大气受地球公转和自转决定的周期现象。此外,人们还找到了不少准周期现象,如低纬高度场的准三年周期,环流指数循环的数周的周期。在单站资料中,人们从不同角度对不同季节找到了不同长度的准周期变化。这种周期或准周期变化对预报是十分有用的。例如,由于知道天气存在年变化,就不会把夏季的温度预报成比冬季还低。一种最简单的周期数据是正弦波(或余弦波),其函数关系为:

$$x(t) = A \sin(2\pi ft + \theta)$$

式中, A 为振幅; f 为频率,即单位时间内的循环数(振动次数); θ 为相角,相对于时间原点的初始相角。正弦数据完成一个循环所需要的时间,称为周期 T 。频率与周期的关系是