

单站统计天气预报方法的研究

中国科学院大气物理研究所集刊

第 3 号

单站统计天气预报方法的研究

科学出版社

内 容 简 介

本集系关于单站天气预报方法研究的工作总结,提供几种单站统计天气预报的方法。内容大致包括以下几个方面:(1)应用单站气象要素的分析,推断影响本站附近区域的天气系统,并在此基础上作出天气预报;(2)利用单站气象要素,根据当时的天气状况和各种要素在一段时期内的演变规律,寻找前期要素变化情况和后期天气的关系;(3)利用几个单站的气象资料作某一地区或一定范围的天气预报。同时,对于应用数理统计方法作天气预报的一些有关问题,如气象要素的处理、预报对象的分类、预报起始场的选取、预报因子的选择和综合方法等问题,本集所选的文章分别作了若干探讨。另外,还介绍了如何把不同预报结果进行综合的预报集成问题。

本集可供气象台站工作者、气象院校师生和气象研究工作者参考。

中国科学院大气物理研究所集刊

第 3 号

单站统计天气预报方法的研究

科学出版社出版

北京朝阳门内大街 137 号

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

1975 年 8 月第 一 版 开本: 787×1092 1/16

1975 年 8 月第一次印刷 印张: 5 3/4

印数: 0001—40,200 字数: 126,000

统一书号: 13031·286

本社书号: 448·13-15

定 价: 0.62 元

前 言

无产阶级文化大革命以来,我国的单站天气预报取得了新的更大的发展。近几年来,许多气象台站已经把统计预报的方法应用到日常的天气分析和预报中去,而且在预报因子的选择上,愈来愈多地考虑它们的天气意义,使之能够较好地反映天气过程发展的规律性,选取多个因子或一组特征过程。在预报因子预报结果的综合上,已经不是不分主次平等对待,而是考虑了各个预报因子和预报量之间的相关程度。在考虑预报因子和预报量之间关系时,不仅考虑各预报因子与预报量之间的直接相关,还考虑了各预报因子之间的相互关系及它们对预报量的关系,根据预报因子与预报量总的相关性大小进行综合,提出不少有关多因子综合的方法。统计预报方法比过去有很大的发展。为落实毛主席关于“备战、备荒、为人民”的伟大战略方针,适应形势发展的需要,使气象工作更好地为国民经济建设和国防建设服务,我们开展了单站统计天气预报方法的研究,初步摸索了几种单站统计天气预报的方法。本集就是这一工作的总结。

本集所选文章大致包括三个部分:第一部分是利用单站气象资料的分析来推断影响本站附近区域的天气系统,然后在此基础上作出天气预报;第二部分是直接利用单站气象资料,不经过推断影响本地附近的天气系统,而根据当时的天气状况和各种气象要素在一段时期内的演变规律,寻找前期要素变化情况和后期天气的关系作出预报。这两部分着重解决本站及本地区的天气预报问题。第三部分是在可以得到较多资料及要求作某一地区或一定范围的区域天气预报时,如何利用几个单站的气象资料,较全面地考虑有影响的天气过程的各个方面以及发展规律,从而找出它们的预报着眼点,应用统计预报的方法较精确地作出预报。本集所选文章还对应用数理统计方法作天气预报的若干问题,如资料处理、预报对象的分类、预报起始场的选择、预报因子的选取和综合方法、预报表格化等问题分别作了一些探讨和处理,使预报迅速、简便。同时,还就如何把不同的预报结果进行综合的预报集成问题作了介绍。

应该指出,全国各地气象台站的同志们创造了不少有关单站天气预报的好方法。我们开展单站天气预报的研究时间较短,水平低,认识肤浅,有不少问题没有考虑到,看法也不一定成熟,书中难免存在缺点和错误,恳请读者批评指正。

本集是我所部分同志与浙江省气象台暨地区台站、浙江农业大学农学系等单位同志共同工作的总结,由我所二室周家斌、陈烈庭、李吉顺、沈如金、孙淑清等同志执笔。

中国科学院大气物理研究所

目 录

要素序列图解综合预报法·····	(1)
用马尔可夫链试作逐日晴雨预报·····	(7)
浙江省汛期连雨过程中期预报方法·····	(24)
编码法·····	(38)
多因子交叉相关综合预报方法·····	(44)
多因子综合概率预报方法·····	(52)
关于预报集成问题·····	(62)
应用指标序列制作干旱季节开始日的长期预报·····	(72)
指标模式群综合中期预报方法·····	(78)
单站地面资料与汛期天气系统的关系·····	(81)

要素序列图解综合预报法

一、引言

天气是大气中发展起来的某些现象的综合结果,而这些现象在时间上和空间上是不断地发展演变着的,它具有一定的时空关系。对于单站统计预报来说,在空间上的联系方面,一般认为外在的影响多少已反映在本站的记录中,可以就分析本站要素序列的变化,找出其统计规律,做出预报。也就是说,把预报问题归结为某个变量 y 随时间的变化,这个随时间变化的函数 $y(t)$ 在任意有限时段内,它的值可以由相应的多元概率分布来确定。预报值 $y(t+\tau)$ 不是唯一的,而只是指明 $y(t+\tau)$ 的可能值的概率分布。这种考虑一方面是出于一般气象台站的具体情况,例如无法及时地获得外地的资料等,另方面实践证明它也是行之有效的。

时间上的联系问题,初始状态的考虑非常重要。一般预报的时效愈长,它所需要考虑的时段也应愈长。另外,对于初始状态,不仅要了解它的平均情况,更重要的是要了解它在一段时期的历史变化。在寻找前后期的各种联系时,应尽可能把这一段时期来的历史变化客观地、定量地表达出来。在这方面,山东省气象台提出了用“座标簇”来表达气象要素的历史变化,这种方法比较客观,而且简单易行。

本文所介绍的方法就是根据上述思路,采用气象要素的一级变量所组成的序列,用图解办法表示出其一段时期来的历史变化,及其与预报量的关系。然后,按L. 贝可夫斯基(Berkofsky)^[1]在作雾的预报时所采用的概率图确定其概率分布。此方法比较适于制作1—3天左右的逐日预报。

鉴于天气转折预报的重要性,下面我们特别选取晴雨转折预报作为研究对象。

二、资料处理和基本预报工具

(一) 根据不同预报对象,选取三至五个预报要素,如气压(P)、温度(T)、湿度(e)等,或几个气象要素按一定的物理考虑组成的新的物理量

本方法所用的基本资料,除预报量外,预报要素均采用气象要素的一级变量,并按时间顺序排成序列,预报量也同样按时间顺序排列。如表1所示。

(二) 基本预报工具的制作

对每一预报要素序列按图1所示,依次不断地取连续的四个一级变量,例如: ΔP_{t-3} , ΔP_{t-2} , ΔP_{t-1} , ΔP_t ($t=1, 2, 3, \dots, M-1$ *, M 是每年资料时段的总天数),加上制作预报当日 t 的预报量实况与预报量 $t+1$ 天的有无降水,做两组概率分布预报图。下面用

* 当作48小时的预报图时为 $M-2$,而当作72小时的预报图时为 $M-3$ 。

表 1

日 期 t		1	2	3	4	5	6	7		
预 报 要 素	ΔP_{24}									
	ΔT_{24}									
	Δe_{24}									
预 报 量 RR	08—08									
	20—20									

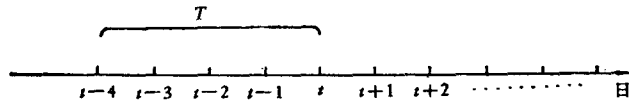


图 1

(图中 T 表示预报时所需之资料长度)

ΔP_{24} 序列预报有无降水为例具体说明如下:

1. 第一组概率分布预报图 (一张)

(1) 我们取 ΔP_{t-3} 为纵座标, ΔP_{t-2} 为横座标, 且 ΔP_{24} 的单位变化在纵横两座标上

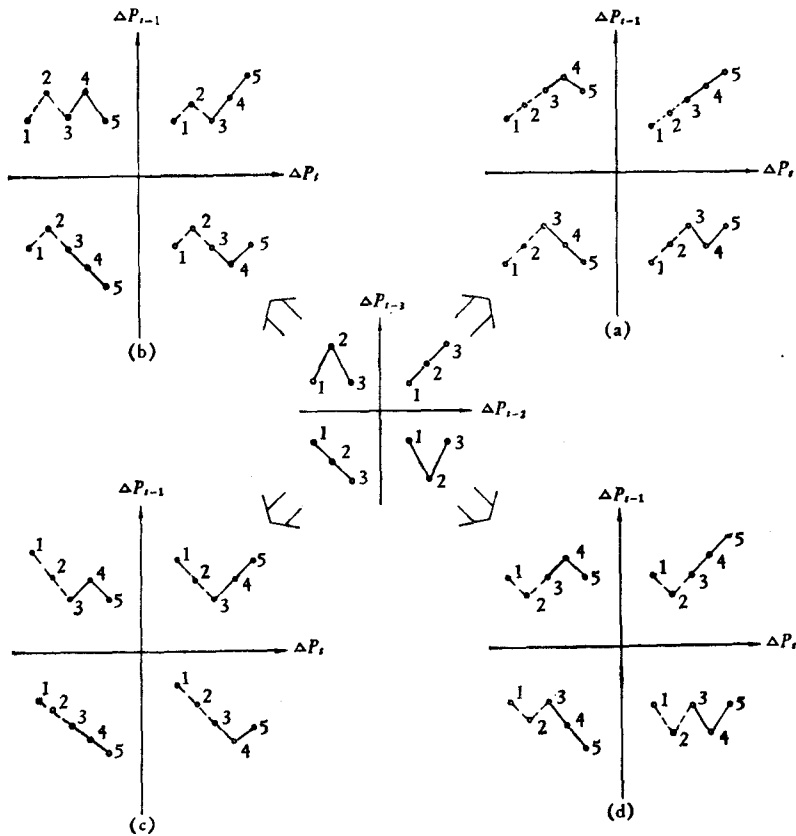


图 2

均取同样的座标间隔。然后根据 ΔP_{t-3} , ΔP_{t-2} 和第 t 天的降水实况做 $t+1$ 天有无降水的点聚图。当第 t 天是雨, $t+1$ 天还是雨; 或第 t 天是晴, $t+1$ 天还是晴时(即晴、雨维持), 在点聚图上用一种符号“。”表示。当第 t 天是雨, 而 $t+1$ 天是晴; 或第 t 天是晴, 而第 $t+1$ 天是雨时(即晴转雨、雨转晴), 在点聚图上用另一种符号“×”表示。

这样的点聚图, 由图 2 中间的 ΔP_{t-3} , ΔP_{t-2} 图可见, 由于纵座标和横座标是同一要素序列的连续二个时间的一级变量, 所以如果某一点落在第 I 象限, 则表示第 $t-3$, $t-2$ 天的气压是连升; 若某点落在第 III 象限, 则表示 $t-3$, $t-2$ 天的气压是连降; 第 II, IV 象限情况类推。

(2) 在点好上述点聚图的基础上进行概率的计算、平滑和等概率线的分析。

首先, 把点聚图的 ΔP_{t-3} 座标轴和 ΔP_{t-2} 座标轴按……—1.0—0.1 毫巴, 0.0—0.9 毫巴, 1.0—1.9 毫巴……分段, 把整个点聚图分成一个个的小方格。然后统计各个小方格中晴雨转折(即符号“×”)的个例数和总个例数之比, 并填在相应的小方格内。为了去掉某些偶然性影响, 以提高百分比数(近似概率值)的代表性, 把每四个小方格所组成的大方格逐个地进行滑动平均, 即统计该大方格中晴雨转折的个例数和总个例数之百分比(即晴雨转折概率), 把它填写在大方格中间。最后根据这些滑动的概率值分析等概率线。在分析等概率线时应注意: 既要考虑概率值的大小, 又要考虑不同符号的点子实际的分布情况。

2. 第二组概率分布预报图(共四张)

这一组图是按 ΔP_{t-3} , ΔP_{t-2} 图中四个象限的四种不同的变化情况分别制作的 ΔP_{t-1} , ΔP_t 与 $t+1$ 天的降水关系的四个概率分布预报图。即在 ΔP_{t-3} , ΔP_{t-2} 图中凡是落在第 I 象限的个例, 根据其后的 ΔP_{t-1} , ΔP_t 到图 2(a) 中做点聚, 点的符号(“。”或“×”)与 ΔP_{t-3} , ΔP_{t-2} 图中的相同。落在第 II 象限的个例, 则根据其后的 ΔP_{t-1} , ΔP_t 在图 2(b) 中做点聚。同样, 落在第 III、第 IV 象限的点子分别到图 2(c)、图 2(d) 上去做点聚。图 2(a)—(d) 的单位坐标间隔与图 2 中间的相同。

从图 2(a)—(d) 的各张图中可看出每个象限都反映出预报日前期五天来的不同的气压变化型式。但 $t-4$ 到 $t-2$ 天是考虑了趋势, 而 $t-2$ 到 t 天考虑了变化量。所做的点聚图则表示了前期 5 天来气压的各种变化型式和预报当天有雨或无雨的初始实况与第 $t+1$ 天降水的相关情况。

最后, 按前面所叙述的同样方法, 对这四张点聚图进行概率计算、平滑和等概率线的分析。

3. 进行预报

预报时利用前面作好的概率分布预报图, 先根据某一预报要素最近的四个一级变量, 将 $t-3$ 和 $t-2$ 天的一级变量点在该要素的第一组概率分布预报图上, 读取一个概率值。再根据 $t-3$, $t-2$ 天该要素所确定的象限, 把 $t-1$, t 天的两个一级变量点在第二组相应的概率分布预报图上, 又读取一个概率值。然后再取一个预报要素, 用上述同样方法在该要素的两组概率分布预报图上读取两个概率值。照此进行, 一直到读完所有选取的预报要素为止。若总共是三个预报要素, 则能读取六个概率值。把这六个概率值求平均, 并与天气转折气候概率值作比较。如果比转折气候概率值大, 又若作预报的当天是晴时, 则预报第二天要转雨; 当作预报的当天是雨时, 则预报第二天将转晴。而如果比转折

气候概率值小,且作预报的当天是晴时,则预报第二天还是晴天;当作预报的当天是雨时,则预报第二天还要下雨。

三、实 例

现将我们所试作的杭州5—6月份逐日晴雨预报的例子介绍如下:

所用的资料是杭州1956—1971年(共16年)从4月26日到6月30日的14时地面气压、温度、绝对湿度资料和降水(08—08时和20—20时的降水两种)资料。其中1957年、1964年和1971年三年资料作为试报用。

雨日定义:预报时效内(当天20时—第二天20时)的降水,微量不算雨日,对于 ≤ 0.5 毫米的降水,如果前一天20—20时是雨日,则认为该天为非雨日;而若前一天为非雨日,则认为该天为雨日。 >0.5 毫米的降水均算雨日。作为预报时初始状况的08—08时的降水,微量算非雨日, ≥ 0.1 毫米的降水均算雨日。

除三年资料作为试报用外,13年共有793个例子,其中晴转雨、雨转晴的个例有368

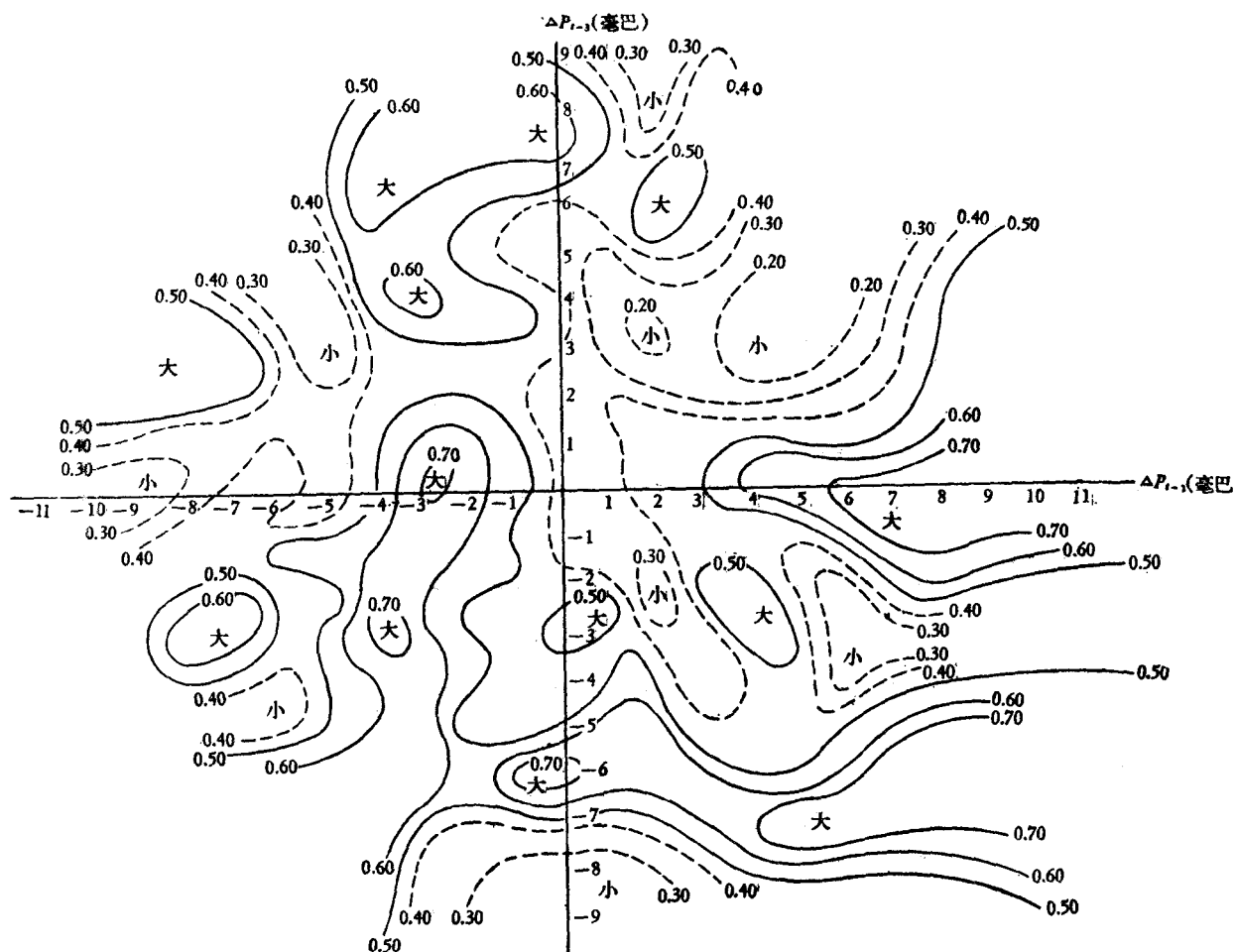


图 3

个, 因此转折的气候概率为 $\frac{368}{793} \approx 0.464$ 。

按前面所叙述的制作概率分布预报图的方法, 利用压、温、湿三个预报要素, 作出六组预报图(共 15 张)。由于篇幅关系, 我们这里只给出一张气压的概率分布预报图(图 3)。

试报结果: 1957, 1964, 1971 年 5—6 月, 分别报对 44, 45, 44 天, 三年共报对 133 天, 占总天数(183 天)的 73%。

四、讨 论

1. 这个方法, 可以理解为预报要素前期一段时期来的变化相似的情况下, 预报对象出现的条件概率。它比平常从单站曲线找相似的方法客观, 并能在图上反映出历史上各个个例的分布以及各点的条件概率在面上的分布。可以比较直观地看出图上前期的要素变化在什么样的范围内预报结果是相似的, 而超过了这个范围就不相似了。而条件概率则表示相似的程度。因此, 确定预报判据后, 能对历史资料进行概括, 根据概括率的大小来判定图的可用程度。而一般的找相似的办法是很难标出对历史资料的概括率的。

2. 概率分布预报图中点子很密集的区域, 算出的百分比可以近似地理解为预报对象出现的条件概率。在图的边缘, 点子较少。但是在边缘上的点子, 其数值是属于比较突出的异常特征。而抓住异常特征, 从历史资料中找相似来做预报, 是目前台站常用的一种预报思路, 虽然例子较少, 但常常能得到比较好的预报效果。对点子密集区和稀少区做这样两种理解, 我们认为图中的各个部位都是可用的。

3. 关于图的综合问题, 我们这里用的是将几张图的条件概率简单平均与气候概率相比较的方法。也可以根据历史资料概括率来确定判据, 使得根据这个判据概括率达到最高。还可以用别的方法对几张图进行综合: (1) 用图解法进行综合。以两张图的条件概率值(0—100%)做为两坐标, 根据个例在前两张图上的条件概率值, 将每个例子按天气转折或维持用不同符号点上去, 可以做出一张新的条件概率图, 以此图的条件概率和第三张图的条件概率值(0—100%)再做为两坐标, 同样作出一张条件概率图, 依此类推, 直到把要综合的几张图都利用了, 得到的条件概率, 即是这几张图的综合结果, 定出判据后即可做预报; (2) 经过一定的处理, 一般预报的综合办法(例如文献[2], [3]和本书第 101—121 页), 都可以对几张条件概率图进行综合, 得到综合的预报结果。

4. 要素序列图解法用两组图比较客观地、而且从量上考虑了预报要素最近 5 天的历史变化, 以此作为预报的初始状况。这对于制作比较长的(例如 3—5 天)预报是很重要的。但是这个方法无法表达一个时期内的各预报要素之间相互配置的情况, 也就是说, 不能把前期影响的天气系统全面地反映出来。因此, 我们考虑, 预报要素如果不用压、温、湿等气象要素本身, 而是运用它们按一定的物理考虑所组合成的一些物理量来作为预报因子, 预报效果可能会好一些。

5. 这个方法是建立在用点聚图找相似的基础上, 因此只要因子选得好, 长、中、短期预报都可应用, 但一定要要求个例比较多, 才能保证算出的条件概率有一定的代表性。

参 考 文 献

[1] L. Berkofsky, An objective determination of probability of fog formation, *Bull. Amer. Met. Soc.*,

Vol. 31, No. 5.1950.

[2] 王宗浩等：天气预报的概率统计方法，第四章，科学出版社，1974年。

[3] 中沢全一、能登正之：判別解析による降水形态の予測，研究时报，1971年，23，p. 79—86.

用马尔可夫链试作逐日晴雨预报

一、引言

二十多年来,马尔可夫链在气象学上的应用逐渐引起人们的注意。1947年,T. A. 萨里穆萨可夫(Сарымсаков)等^[1]用它研究了天气型的转换;后来K. A. 加布里埃尔(Gabriel)等^[2]和A. M. 费耶海姆(Feyerherm)等^[3,4]用之计算了干湿日及天气循环的概率;最近W. J. 蒙德尔(Maunder)等^[5]又用它研究了公路建设与气象条件的关系。然而,总的说来,这方面的应用大多数限于气候统计(本文不涉及马尔可夫链在大气物理方面的应用),只有个别文章讨论过天气预报问题(例如张家诚等^[6])。马尔可夫链之所以还未能广泛用于天气预报,看来是由于以下原因:简单链即一重链的无后效性太强,不合大气运动的实际,多重链的计算又十分复杂而困难(详论见后),这些都限制了它在天气预报上的应用。

因此,怎样才能不用一重链的假定而又避免直接引用多重链,就是一个值得深入研究的问题。么枕生^[7]的工作可以看作是这方面的一个尝试。本文以浙江北部(以下简称浙北)汛期(5—6月)逐日晴雨预报为例,进一步讨论这一问题。

另外,过去的工作都限于讨论一维情况。考虑到预报问题的需要,本文将对多维多重马尔可夫链用于天气预报的问题作一些讨论。

二、一维情况

我们首先讨论由浙北逐日晴雨组成的一维马尔可夫链。浙北逐日晴雨用浙北三站日降水量代表,凡三站日降水量均 ≥ 0.1 毫米,则该日定为雨日。其中两站日降水量 ≥ 0.1 毫米,且其和 ≥ 1.0 毫米,亦定为雨日。其他定为晴日(定义的若干细节从略)。由此,每日有晴雨两种可能状态,而浙北逐日晴雨组成状态数目有限时间疏散的马尔可夫链^[8]。

(一)状态定义

晴雨两种状态组成的简单链,因为只考虑前一天状态对后期的影响而不反映要素的历史演变,因而对预报不大适用。为改进计,参考么枕生^[7]的作法,对每日状态重新定义,使之反映历史演变,然后再用简单链处理。为了反映汛期中连晴、连雨过程的特点,我们将么枕生的状态定义加以改变,而采用如下定义:状态号码第一字为0表示晴,为1表示雨,号码第二字表示晴雨持续日数,即 RR_{0r} 表示连晴 r 日, RR_{1r} 表示连雨 r 日($r=1,2,3,\dots$)。

作为例子,表1第一列给出1972年5月26日至6月10日逐日晴雨定义(0表示晴,1表示雨)。第二列给出相应的状态定义。为使每种状态的频数不致过少,根据资料情况,规定连晴超过7日者归入 RR_{07} ,连雨超过6日者归入 RR_{16} ,故 RR_{07} 表示连晴7日以

上, RR_{16} 表示连雨 6 日以上。

表 1 1972 年 5 月 26 日至 6 月 10 日逐日晴雨、14 时要素及状态号码

项 目 期	晴 雨		气 压		温 度		湿 度		风 向	风速 (米/秒)	风 速 分 量			
	定义	状态 号码	P (毫巴)	状态 号码	T (°C)	状态 号码	e (毫巴)	状态 号码			u (米/秒)	状态 号码	v (米/秒)	状态 号码
5 月 26 日	0	01	05.1	13	24.3	02	15.6	13	ENE	2	1.8	01	0.8	11
27	0	02	05.7	01	24.5	03	17.2	01	NEE	4	3.7	02	1.5	01
28	0	03	04.6	11	27.1	04	23.4	02	SE	2	1.4	11	-1.4	-01
29	1	11	00.7	12	33.4	05	26.1	03	SW	6	-4.2	-01	-4.2	-02
30	1	12	13.1	01	17.6	11	13.9	11	SW	4	-2.8	-11	-2.8	-11
31	0	01	09.4	11	26.1	01	13.0	12	SSE	2	0.8	01	-1.8	-12
6 月 1 日	1	11	05.9	12	22.8	11	17.9	01	S	5	0	11	-5.0	-01
2	1	12	04.6	13	23.0	01	27.1	02	S	2	0	12	-2.0	-11
3	1	13	04.8	01	24.6	02	25.8	11	SSW	3	-1.1	-01	-2.8	-01
4	1	14	04.5	11	20.3	11	21.8	12	NNE	1	0	-11	0	-11
5	1	15	01.9	12	18.8	12	18.6	13	ESE	6	5.5	01	-2.3	-01
6	1	16	98.4	13	21.9	01	19.5	01	ENE	3	2.8	11	1.1	01
7	0	01	97.6	14	27.0	02	19.2	11	WNW	1	0	12	0	11
8	0	02	99.9	01	29.3	03	20.7	01	SSW	3	-1.1	-01	-2.8	-01
9	0	03	02.8	02	31.6	04	21.3	02	SE	2	1.4	01	-1.4	-11
10	0	04	00.0	11	32.8	05	20.8	11	SSE	6	2.3	02	-5.5	-01
5 月 25 日	1	13	07.3	12	20.7	01	16.8	12	N	4	0	13	4.0	02

注：末行列入 5 月 25 日的资料是为了确定 26 日的状态号码。风速 1 米/秒以静风计。连晴 7 日以上一律定为 RR_{07} ，连雨 6 日以上一律定为 RR_{16} ，对其他要素亦有类似规定。因此，从 RR_{07} 转向 RR_{07} ，从 RR_{16} 转向 RR_{16} 仍有一定的概率。

(二) 转移概率的计算

假定晴雨 RR 的马尔可夫链是简单而且均匀的，则一步转移概率的计算公式如下^[8]：

$$P_{\alpha, \beta} = \frac{\sum_{h=1}^N m_{\alpha, \beta}^{(h)}}{\sum_{h=1}^N m_{\alpha}^{(h)}} \quad (1)$$

其中 $P_{\alpha, \beta}$ 是状态 RR_{α} 转为状态 RR_{β} 的近似值。 $m_{\alpha, \beta}^{(h)}$ 是第 h 个观测列中状态 RR_{α} 之后实现状态 RR_{β} 的频数， $m_{\alpha}^{(h)}$ 为该观测列中状态 RR_{α} 的频数， N 为观测列总数。本文共用 14 年资料，故取 $N=14$ 。

由(1)式算得一步转移概率矩阵 $P^{(1)}$ 。为清楚起见，将其用表格形式列出(见表 2)。
 n 步转移概率矩阵 $P^{(n)}$ 可用矩阵乘法按下式求出^[8]：

$$P^{(n)} = P^{(1)n} \quad (2)$$

表 3 就是按(2)式求出的 2—5 步转移概率矩阵。

表 2 一步转移概率矩阵 $P^{(1)}$

$\beta \backslash \alpha$	01	02	03	04	05	06	07	11	12	13	14	15	16
01	0	0.702	0	0	0	0	0	0.298	0	0	0	0	0
02	0	0	0.650	0	0	0	0	0.350	0	0	0	0	0
03	0	0	0	0.710	0	0	0	0.290	0	0	0	0	0
04	0	0	0	0	0.796	0	0	0.204	0	0	0	0	0
05	0	0	0	0	0	0.550	0	0.450	0	0	0	0	0
06	0	0	0	0	0	0	0.650	0.350	0	0	0	0	0
07	0	0	0	0	0	0	0.606	0.394	0	0	0	0	0
11	0.318	0	0	0	0	0	0	0	0.682	0	0	0	0
12	0.456	0	0	0	0	0	0	0	0	0.544	0	0	0
13	0.315	0	0	0	0	0	0	0	0	0	0.685	0	0
14	0.405	0	0	0	0	0	0	0	0	0	0	0.595	0
15	0.500	0	0	0	0	0	0	0	0	0	0	0	0.500
16	0.455	0	0	0	0	0	0	0	0	0	0	0	0.545

表 3 二步至五步转移概率矩阵

$\beta \backslash \alpha$	01	02	03	04	05	06	07	11	12	13	14	15	16
$P^{(2)}$													
01	0.095	0	0.456	0	0	0	0	0.246	0.203	0	0	0	0
02	0.111	0	0	0.462	0	0	0	0.188	0.239	0	0	0	0
03	0.092	0	0	0	0.565	0	0	0.145	0.198	0	0	0	0
04	0.065	0	0	0	0	0.438	0	0.358	0.139	0	0	0	0
05	0.143	0	0	0	0	0	0.358	0.192	0.307	0	0	0	0
06	0.111	0	0	0	0	0	0.394	0.256	0.239	0	0	0	0
07	0.125	0	0	0	0	0	0.367	0.239	0.269	0	0	0	0
11	0.311	0.223	0	0	0	0	0	0.095	0	0.371	0	0	0
12	0.171	0.320	0	0	0	0	0	0.136	0	0	0.373	0	0
13	0.277	0.221	0	0	0	0	0	0.094	0	0	0	0.408	0
14	0.297	0.284	0	0	0	0	0	0.121	0	0	0	0	0.298
15	0.227	0.351	0	0	0	0	0	0.149	0	0	0	0	0.273
16	0.248	0.319	0	0	0	0	0	0.136	0	0	0	0	0.297

 $P^{(3)}$

01	0.171	0.066	0	0.324	0	0	0	0.160	0.168	0.111	0	0	0
02	0.169	0.078	0	0	0.367	0	0	0.127	0.129	0.130	0	0	0
03	0.136	0.065	0	0	0	0.311	0	0.282	0.099	0.107	0	0	0
04	0.177	0.046	0	0	0	0	0.285	0.172	0.244	0.076	0	0	0
05	0.201	0.100	0	0	0	0	0.217	0.184	0.131	0.167	0	0	0
06	0.190	0.078	0	0	0	0	0.238	0.189	0.175	0.130	0	0	0
07	0.198	0.088	0	0	0	0	0.223	0.182	0.163	0.146	0	0	0
11	0.147	0.218	0.145	0	0	0	0	0.171	0.065	0	0.254	0	0
12	0.194	0.120	0.208	0	0	0	0	0.163	0.093	0	0	0.222	0
13	0.233	0.195	0.144	0	0	0	0	0.160	0.064	0	0	0	0.204

续表 3

$\alpha \backslash \beta$	01	02	03	04	05	06	07	11	12	13	14	15	16
14	0.174	0.209	0.185	0	0	0	0	0.188	0.082	0	0	0	0.162
15	0.172	0.159	0.228	0	0	0	0	0.191	0.101	0	0	0	0.149
16	0.178	0.174	0.208	0	0	0	0	0.186	0.092	0	0	0	0.162

 $P^{(4)}$

01	0.162	0.120	0.043	0	0.258	0	0	0.140	0.110	0.091	0.076	0	0
02	0.140	0.119	0.051	0	0	0.202	0	0.243	0.087	0.069	0.089	0	0
03	0.168	0.096	0.042	0	0	0	0.202	0.172	0.192	0.054	0.074	0	0
04	0.190	0.124	0.030	0	0	0	0.173	0.181	0.117	0.133	0.052	0	0
05	0.171	0.141	0.065	0	0	0	0.131	0.181	0.126	0.071	0.114	0	0
06	0.180	0.133	0.051	0	0	0	0.145	0.178	0.129	0.095	0.089	0	0
07	0.178	0.139	0.057	0	0	0	0.135	0.178	0.125	0.088	0.100	0	0
11	0.187	0.103	0.142	0.103	0	0	0	0.162	0.117	0.035	0	0.151	0
12	0.205	0.136	0.078	0.148	0	0	0	0.160	0.111	0.051	0	0	0.111
13	0.173	0.164	0.127	0.102	0	0	0	0.179	0.109	0.035	0	0	0.111
14	0.172	0.121	0.136	0.131	0	0	0	0.178	0.128	0.045	0	0	0.089
15	0.174	0.120	0.104	0.162	0	0	0	0.173	0.130	0.056	0	0	0.081
16	0.175	0.125	0.113	0.147	0	0	0	0.174	0.127	0.050	0	0	0.089

 $P^{(5)}$

01	0.154	0.114	0.078	0.031	0	0.142	0	0.219	0.096	0.059	0.062	0.045	0
02	0.175	0.098	0.077	0.036	0	0	0.131	0.169	0.166	0.047	0.048	0.053	0
03	0.189	0.118	0.062	0.030	0	0	0.123	0.175	0.117	0.105	0.037	0.044	0
04	0.174	0.133	0.081	0.021	0	0	0.104	0.177	0.124	0.064	0.091	0.031	0
05	0.184	0.119	0.092	0.046	0	0	0.080	0.171	0.124	0.068	0.049	0.068	0
06	0.181	0.126	0.087	0.036	0	0	0.088	0.172	0.122	0.070	0.065	0.053	0
07	0.181	0.125	0.090	0.041	0	0	0.082	0.172	0.122	0.067	0.061	0.059	0
11	0.191	0.131	0.067	0.101	0.082	0	0	0.153	0.111	0.064	0.024	0	0.076
12	0.168	0.144	0.089	0.055	0.118	0	0	0.161	0.109	0.061	0.035	0	0.060
13	0.169	0.121	0.107	0.090	0.081	0	0	0.166	0.122	0.059	0.024	0	0.061
14	0.169	0.120	0.079	0.096	0.105	0	0	0.160	0.122	0.070	0.031	0	0.048
15	0.169	0.122	0.078	0.073	0.129	0	0	0.157	0.119	0.071	0.038	0	0.044
16	0.169	0.123	0.081	0.080	0.117	0	0	0.159	0.119	0.069	0.035	0	0.048

一般情况下,一步转移概率矩阵形式如下:

$$P^{(1)} = \begin{pmatrix} 0 & q_1 & 0 & \cdots & 0 & p_1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & q_2 & \cdots & 0 & p_2 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & q_{M-1} & p_{M-1} & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & q_M & p_M & 0 & 0 & \cdots & 0 \\ q_{M+1} & 0 & 0 & \cdots & 0 & 0 & p_{M+1} & 0 & \cdots & 0 \\ q_{M+2} & 0 & 0 & \cdots & 0 & 0 & 0 & p_{M+2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ q_{M+K-1} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & p_{M+K-1} \\ q_{M+K} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & p_{M+K} \end{pmatrix} \quad (3)$$

其中 M 为晴日状态的总数, K 为雨日状态的总数。连晴超过 M 天仍定为 R_{0M} , 连雨超过 K 天仍定为 R_{1K} 。若连晴日最多为 M 天, 连雨日最多为 K 天, 则 $P_{0M,0M}=q_M=0$, $P_{0M,11}=p_M=1$, $P_{1K,01}=q_{M+K}=1$, $P_{1K,1K}=p_{M+K}=0$ 。

我们应用数学归纳法于(2),(3)两式,求得矩阵 $P^{(n)}$ 的元素的递推公式。由于其表达式较长,列入附录中。

将表 2、表 3 所列各步转移概率加以比较,并分析附录中的普遍表达式,可以知道,随着矩阵阶数的增高,矩阵中非零元素的数目逐渐增多,当阶数 $n \geq \max(M+1, K+1)$ 时,矩阵 $P^{(n)}$ 就全部由非零元素组成了。因此,我们所研究的马尔可夫链是不可分解的非循环链^[8]。

(三) 理论值与实测值的比较

取第 t 日至第 $t+n$ 日体系的转移作为一次观测,则可按高阶转移概率的定义^[8]由历史资料用与(1)式类似的公式求得 n 步($n \geq 1$)转移概率^[6]。

由于我们的目的是预报晴雨,因此只要求出转晴转雨概率即可。为此,将转向 RR_{01} , $RR_{02}, \dots, RR_{07}$ 与 $R_{11}, R_{12}, \dots, R_{16}$ 的概率分别相加,并各自用 $P_{\alpha,0}^{(n)}$, $P_{\alpha,1}^{(n)}$ 记之。表 4 给出由表 2、表 3 算出的 1—5 步转雨概率的理论值及由实际资料算得的实测值。转晴概率 = 1—转雨概率,故略之。

表 4 $P_{\alpha,1}^{(n)}$ 的理论值与实测值

项 目 α	n	理 论 值					实 测 值			
		1	2	3	4	5	2	3	4	5
01		0.298	0.449	0.439	0.417	0.481	0.430	0.477	0.426	0.474
02		0.350	0.427	0.386	0.488	0.483	0.438	0.381	0.435	0.463
03		0.290	0.343	0.488	0.492	0.478	0.348	0.461	0.457	0.549
04		0.204	0.497	0.492	0.483	0.487	0.510	0.460	0.560	0.400
05		0.450	0.499	0.482	0.492	0.480	0.474	0.579	0.421	0.459
06		0.350	0.495	0.494	0.491	0.482	0.750	0.450	0.500	0.350
07		0.394	0.508	0.491	0.491	0.481	0.424	0.515	0.455	0.455
11		0.682	0.466	0.490	0.465	0.428	0.474	0.460	0.468	0.410
12		0.544	0.509	0.478	0.433	0.426	0.471	0.452	0.429	0.390
13		0.685	0.502	0.428	0.434	0.432	0.491	0.436	0.400	0.480
14		0.595	0.419	0.132	0.440	0.431	0.459	0.405	0.543	0.545
15		0.500	0.422	0.441	0.440	0.429	0.500	0.550	0.474	0.500
16		0.545	0.433	0.440	0.440	0.430	0.381	0.450	0.579	0.684

为了比较,我们用下式计算了理论值与实测值的离差:

$$\sigma^{(n)} = \sqrt{\frac{\sum_{\alpha} (P_{\alpha,1}^{(n)} - P_{\alpha,1}^{(n)'})^2}{M+K}} \quad (4)$$

式中 $\sigma^{(n)}$ 为 n 步转雨概率的离差, $\sum_{\alpha}$ 表示对所有状态号码求和。此处 $n=2,3,4,5$, $M=7, K=6$ 。由表 4 资料求得: $\sigma^{(2)}=0.081, \sigma^{(3)}=0.040, \sigma^{(4)}=0.063, \sigma^{(5)}=0.096$ 。

三、多 维 情 况

我们取浙北逐日晴雨 RR , 杭州逐日 14 时气压 p 、温度 T 、湿度 e 、风速分量 u, v 组成六维马尔可夫链, 用于预报晴雨 RR (即只报六维中的一维)。由于杭州位于浙北中心, 故以杭州要素变化代表浙北天气形势的演变 (关于地面要素与汛期天气系统的关系参看本书第 81—85 页)。

(一) 状 态 定 义

各要素的状态定义如下 (RR 的定义见前):

1. 气压 p 、温度 T 、湿度 e

先逐日确定 24 小时变量符号, 状态号码第一字为 0, 表示变量为正; 为 1, 表示变量为负。变量为 0 时, 以前一日符号计, 以示过程持续。状态号码第二字表示正负变量持续日数。例如, p_{0r} 表示气压连升 r 日, p_{1r} 表示气压连降 r 日 ($r=1, 2, 3, \dots$)。

2. 风速分量 u, v

将逐日 14 时风速向量分解成 u, v 两个分量。对 $u(v)$, 状态号码为正表示东 (北) 风, 为负表示西 (南) 风。分量为 0, 以前一日分量符号计。状态号码第一字为 0 表示该分量 24 小时变量为正, 为 1 表示变量为负。状态号码第二字表示正负变量持续日数。例如, $u_{0r}(u_{-0r})$ 表示东 (西) 风分量连增 r 日, $u_{1r}(u_{-1r})$ 表示连减 r 日。

作为例子, 表 1 给出 1972 年 5 月 26 日至 6 月 10 日逐日资料及相应的状态定义。图 1 给出相应的要素历史曲线。

(二) 转 移 概 率 计 算

多维情况下状态的转移, 一般是化成一维问题来作。例如, 对由 RR 与 p 组成的二维链, 若设 RR 与 p 各有两种状态: $RR_1, RR_2; p_1, p_2$, 则将 RR 与 p 的不同状态加以组合, 化成如下四种状态的一维链:

A_1 —— $RR_1 p_1$

A_2 —— $RR_1 p_2$

A_3 —— $RR_2 p_1$

A_4 —— $RR_2 p_2$

并求取如下转移矩阵:

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \\ P_{41} & P_{42} & P_{43} & P_{44} \end{bmatrix}$$

本文处理的是六维链, 每一维的状态数目很多, 化成一维问题处理不仅计算量大, 而且由于组合后状态数目极大, 使统计失去意义。因此, 我们采用简化方法, 计算不同要素的转晴、转雨概率。对上述简单例子, 可以分别计算 p_1, p_2 转向 RR_1, RR_2 的概率。即不再区分 $P_{11}, P_{12}, P_{31}, P_{32}$, 而代之以 p_1 转向 RR_1 的概率。同样, 以 p_1 转 RR_2 代替 P_{13} ,

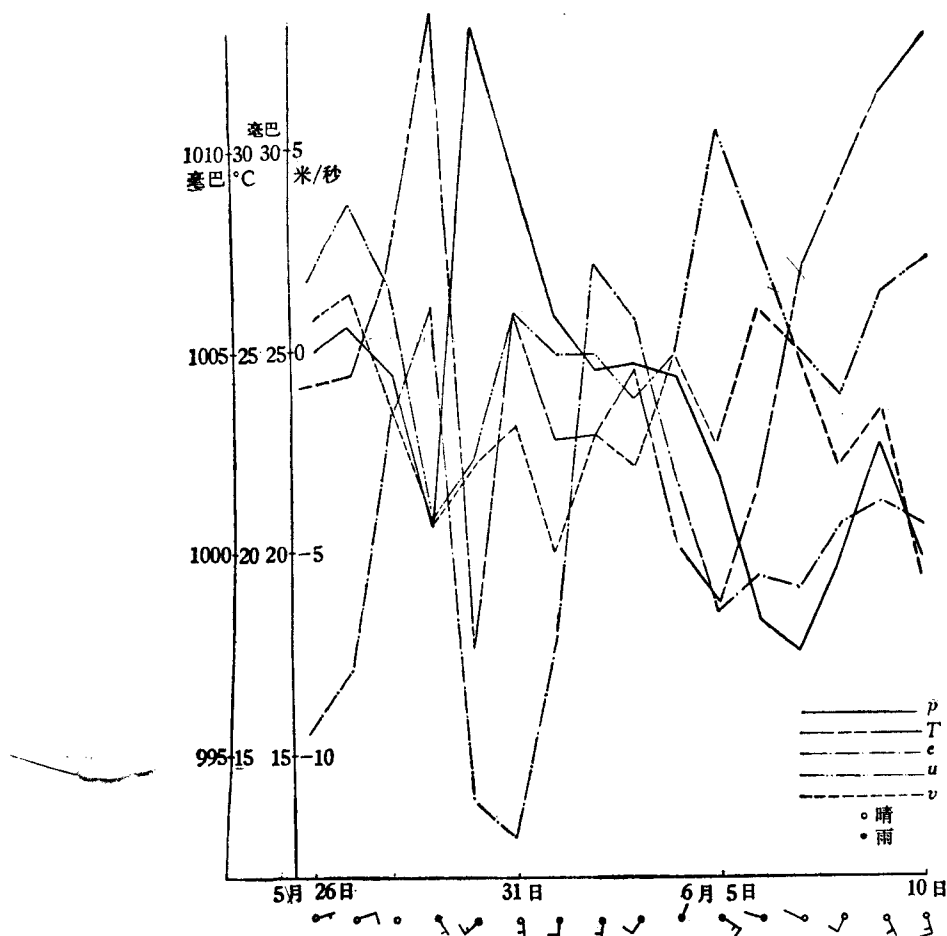


图 1 1972 年 5 月 26 日至 6 月 10 日浙北逐日晴雨及 14 时杭州要素的历史演变

P_{14}, P_{33}, P_{34} . 对 p_2 转 RR_1, p_2 转 RR_2 也是一样。

因此, 气压 p 一步转雨概率计算公式如下:

$$P_{\alpha,1} = \frac{\sum_{h=1}^N m_{\alpha,1}^{(h)}}{\sum_{h=1}^N m_{\alpha}^{(h)}} \quad (5)$$

式中 $P_{\alpha,1}$ 为某日状态为 p_{α} 时次日为雨日的转移概率, $m_{\alpha,1}^{(h)}$ 为第 h 个观测列中状态 p_{α} 转为雨日的频数, $m_{\alpha}^{(h)}$ 为该观测列中状态 p_{α} 的频数, N 为观测列总数。

气压一步转晴概率, 多步转晴、转雨概率, 其他要素多步转移概率, 其计算方法类之。

表 5 给出 p, e, u, v 的各步转雨概率。

四、预报公式及结果

(一) 24 小时 预 报

某要素 S , 预报某日为雨日的概率用下式算出: