

第二讲 基本气候状态 及其诊断分析

黄嘉佑

(北京大学地球物理系, 100871)

数据的收集和整理是统计过程的第一步, 是作气候分析和预测的基础。本讲着重介绍如何作气象数据的基础处理。

1 如何表现气候状态

描述某地区的气候状态可以用定性资料或定量数据不同程度的表现来描述, 例如某地的县志等资料记载该地气象情况的描述资料, 对我国历史气温(降水量)的恢复可以根据记载情况把定性资料分为特冷(特旱)、冷(旱)、正常、暖(涝)和特暖(特涝)5个级别来描述。有时为描述简便, 把气候状态分为两级, 即冷和暖或旱和涝来描述一个时期中的气候态, 也可分3级, 即用冷、正常和暖来描述温度状态, 当然也可分7级或9级等。描述气候状态的时期依据研究问题而定, 即时间尺度可以是一季、一年、十年、百年或千年尺度, 短到以天为研究对象也可以, 如逐日的晴和雨状态等。为了研究方便, 定性资料常转化为定量资料, 其方法是根据定性的描述现象的严重程度用自然数表示其等级。对应5个气候级别用1、2、3、4和5转换为定量资料。为了便于数学的研究, 常把研究对象称为变量, 变量数据是离散的称为离散型变量。

对于由仪器观测得到的资料, 一般是由数值来描述的, 数值本身大小就是表现不同的气候状态。变量资料数据是用实数表示, 称连续型变量。为了研究方便, 常把连续型变量转换为离散型变量, 用某些数值为界限划分不同的级别即得。

气象数据的表示和进一步的分析不是我们研究的目的, 只是通过对它的研究分析来了解对象的全体。对象的全体也称为总体, 我们所取得的数据称为样本。气候状态的表现除了用变量样本数据的大小变化表示(常用列表形式给出)外, 还可以用变量值随时间变化曲线(用函数图像)表现。在数学上的表示常用变量 x 表示研究对象全体, 用

$$x_i (i = 1, 2, \dots, n) \quad (1)$$

表示样本, 其中 i 表示数据顺序, 对某个数据称为样品, n 表示数据的总个数, 也称为样本容量。如果数据是随时间变化的, 常称为时间序列, 把 i 改为 t 来表示不同时间所取得的样本。从几何学的角度可以把数据样本看成为 n 维空间的一个向量, n 个样品是它在 n 个轴上的投影分量。还可以用矩阵形式把样本数据表示出来, 当把 n 个数据排列成一行, 用 $X(1 \times n)$ 表示, 称为行向量。当然也可以把数据排成一列, 即用矩阵 X 的转置 X' 来表示, 称为列向量。可见对气候状态的表示有很多形式。

气候状态出现的机会是不同的, 某种状态出现的可能性大小用概率来度量。同样, 变量小于某一气候状态的事件出现的概率是不同的, 这种概率值随气候状态取值而变化, 称为分布函数。对连续型变量某事件出现的可能性可以通过对概率密度函数积分求得。不同气象要素有不同的分布。常见的分布是正态分布, 它的概率密度函数形状呈钟型。月平

均温度和年平均温度都可以看成正态分布，但是月降水量不一定是正态分布，特别是北方地区。此外气象要素中云量和风向风速等也不能看成是正态分布。变量的分布在统计方法使用中常常容易被人们忽视，因为统计方法理论基础的大部分是假定变量遵从正态分布，而气象灾害预测研究对象主要是旱涝，旱涝状态的分布是否符合这种假定是值得注意的。

上面所述的是单个气象要素的表现，很容易把它们推广到多个要素的情况。例如对单个测站的气温、降水量和气压的情况，或者是一个要素6、7和8月份的情况，或者一个地区几个测站同一要素的情况等。这时它们的数据常被表现成有行有列的表或图形，数学上则用矩阵表示。

2 怎样表示气候平均状态

我们所说的“气候”即为常见的气象要素的多年平均值。计算平均值的年代要足够长，一般来说用30年资料计算得的平均值作为气候要素值是合适的，也是世界气象组织所规定的统一要求，不同月份气象要素多年平均值也可用来表征气候要素的季节变化特征；广义的气候要素可包括具有能量意义的要素（如大气稳定性、大气透明度及太阳辐射强度等）的平均状态。当然，根据研究的时间尺度长短也可以用逐日、逐候的要素平均值来表现气候的平均状态，也即常见的气候态。把一地区不同台站同一要素的平均值点在一张图上形成平均场。

平均值是最常用的数据统计量，是用来表示常见气候状态的量。例如月平均温度可以表示一个地区气候冷暖程度的平均状态。但是，为什么不用月平均降水量来表示旱涝的平均状态呢？其原因主要是月气温可以近似遵从正态分布，而月降水量不遵从正态分布。如果变量遵从正态分布，其样本平均值是总体数学期望的最好估计值，在气候状态变化中，出现平均值附近的状态也是比较多

的，因而平均值能反映常见的气候状态，有较好的代表性。但是，若变量不是遵从正态分布，平均值的代表性就差，虽然月降水量与月平均温度同样是月气象要素测量值，但是其代表性是有很大差异的。试想，若30天中，有29天晴，有一天下了600mm雨量，其降水量平均值为20mm，它能反映平均降水状态么？因此，在月气象要素中没有平均降水量，而仅用月总降水量反映。

3 怎样表现气候变化

气候变化，泛指各种时间尺度的气候演变，变化的时间长度从最长的几十亿年至最短的年际变化，差别很大。据变化的时间尺度及性质可分为三类：①地质时期气候变化。其变化时间在万年以上，常称之为“气候变迁”；②历史时期气候变化。指冰后期即一万年以来，主要是近五千年的时段；③近代气候变化。指近百年或20世纪以来的时段。一些研究认为，其产生的原因除太阳活动和火山爆发外，还应考虑海气相互作用及人类活动的影响。研究尺度在数十年范围内，又称短期气候变化。

对连续数据气候态的异常程度可以用偏离平均状态的程度来反映。也即对原变量某些变换变量来度量。一般常用的变换变量有：

(1) 距平变量，即把原变量 x 减平均值 \bar{x} 后得到。即

$$x_d = x - \bar{x} \quad (2)$$

新变量的平均值为零。通常将它随时间变化的距平值绘成图，可以方便看见正负距平的变化情况，如果变量是降水量，即反映旱涝变化情况。对不同地点的距平变量序列绘在一张图上可以比较它们异常变化的相互关系。在气象中还常用时间-经度距平图，即把某一纬度要素距平值按时间和经度点在图上，从图上可分析正（负）距平区随时间向东或向西传播方向。还有距平波列图，在要素距平场的图上，正距平区反映正弦波的波峰，负距平区反映正弦波的波谷，正负距平区的中心连线

可以反映波列的走向。

将距平序列的各个数据值平方后再求平均值,所得的统计量称为方差。对方差再开方的量称标准差,也有称为均方差。表示为

$$s = \left[\frac{1}{n} \sum (x_i - \bar{x})^2 \right]^{1/2} \quad (3)$$

标准差是度量变量平均变化幅度,即平均异常程度,它与变量有相同的单位。将不同地点的降水量序列的标准差值绘在一张图上,可以看见哪个地区有最大的异常情况。标准差的代表性与平均值代表性一样,也与变量分布有很密切关系。当变量遵从正态分布时,标准差可以较好地反映其平均变化幅度。

在天气预报中,常用标准差来度量实况值与预报值在多次预报试验中的预报效果。通常把求出它们差值序列的标准差,称为均方误差,其值越小表明实况与预报值平均差异很小,有较好的预报效果。当然,作要素场预报时,可以把各测站的标准差值绘在一张图上,标准差小的地区表明有较好的预报效果。但是,对某气候要素的标准差场中,标准差大的地区对分析和预报上不一定是不好的地区。标准差大的地区表明它们受到某种外力作用有很强的变化反映,能包含外力因素影响的很多信息,用该地区的变量作预报量或预报因子,都能较好反映出外力的作用。

(2) 标准化变量,即把距平变量除以标准差后得到。

$$x_s = (x - \bar{x})/s \quad (4)$$

新变量的平均值为0,标准差为1。它是无单位的变量,可以将不同单位降水量和温度标准化变量序列绘在一张图上,比较它们异常变化的相互关系。对一年12个月的温度转换成标准化变量后,可以消除它们年变化的影响,也是一种滤波的作用。另外,对各月年序列标准化后可以将各月数值逐月连接起来,转换成月尺度的时间序列,起到扩大样本容量的效果。

(3) 距平百分率变量,即把距平变量除以平均值后得到。

$$x_b = (x - \bar{x})/\bar{x} \quad (5)$$

新变量的平均值为0,方差为原变量方差除

以平均值。它也是一个无单位的变量,可以消除不同地区降水量平均值的差异,在我国台站中常用。

(4) 0—1 变量,它定义为

$$x_0 = (x - x_m)/(x_m - x_e) \quad (6)$$

式中 x_m 和 x_e 分别为原变量序列中的最大值和最小值,它的作用是将原变量数据值的变化变换为0到1之间。它常被用于对变量作非线性变换前的数据处理,因为若数据中有负数时不能作变量的平方根变换。另外变换到小于1的数后,对于指数变换时可以避免数值太大使计算机计算溢出。

上述几个变量均是原变量的一种线性变换,用以改变原变量的平均值和变化幅度,达到相互比较的目的。

对于概括了解变量的变化情况,常用数据的统计量描述。除平均值和标准差外,还使用变异系数,它是把变量的标准差除以平均值得到。它是为了比较如气温和降水量不同变量或同一变量不同地区的变异程度所设计的。它的作用是在比较前先消除它们的单位,将变量的标准差除以平均值得到。即

$$v = s/\bar{x} \quad (7)$$

4 怎样诊断气候变化中的异常

在气候变化研究过程中,气候态出现异于常年状态时,称为气候异常。在气候级别划分中不是正常的级别均可以称为异常。由于气候是否异常与所考虑的时间尺度有关,在某一时段内的异常现象在更长的时段内也可能不算异常。因此,在报道异常气候时,常需指出是多少年一遇的异常。例如,世界气象组织曾规定30年一遇的天气气候为较短期的天气气候。就大范围气候来说,有极值的异常气候,有时空分布方面的异常气候。人们经常把大范围严寒,热浪,洪涝或干旱称为异常。例如,80年代初非洲撒哈拉地区的异常干旱,本世纪20年代开始的北半球地区气温异常增暖,1982—1983年由赤道东太平洋地区海温异常增高所引发的厄尼诺现象等。由于异常气候对人类活动和经济发展带来灾害性的影响,目前它是气象工作者十分关注的研究课题之一。

上面我们所描述的气候异常是指对正常状态的一种偏离程度，人们关注的异常是指气候的严重异常，通常指给人类带来严重灾害的气候异常态。也可以说是狭义的气候异常。

统计中诊断异常态的标准常用该状态出现的概率为标准。如果该状态出现的可能性很小，即所谓罕见，就可称为异常，在气象中通常用小于5%为标准。当然还可以用小于1%作为标准。应用气候状态这种判据的有：

(1) 资料数据的检错。对变量序列作标准化处理后，如发现其标准化值的绝对值大于3者，可以进而判断它是否有错。因为若变量遵从正态分布，变量的绝对值大于3的概率约为0.13%，一次观测能出现超过3的值，实属罕见，因而可以怀疑。当然也应注意有可能确是异常状态出现。

(2) 某一时期平均气候状态的异常。在气候变化分析中，我们常要了解在ENSO现象发生的年份，某一测站降水量是否有旱或涝的异常时，就可以求出ENSO年份的降水量平均值，但是这个平均气候状态是否可以算异常，还需作进一步的检验。检验的统计量常用t统计量

$$t = (x_a - x_m) / [s / (n - 1)^{1/2}] \quad (8)$$

式中， x_a 为 ENSO 年份降水量的平均值， x_m 为一般气候平均态，它可以用该地多年降水量平均值代替， s 为 ENSO 年份降水量的标准差， n 为 ENSO 年份数。 t 统计量遵从自由度为 $(n - 1)$ 的 t 分布。当统计的年份数在 30—100 时，统计量的值约大于 2 时，事件出现 4 的概率小于 5%，则可以判断为异常态。

例如，为了研究不同强度海温距平异常下我国降水的可预报性，使用类似统计量，其分子表示在海温异常（偏冷或偏暖，极冷或极暖）条件下，我国某测站降水量平均值（代表异常气候状态）与所有样品降水量平均值（代表正常气候状态）之差，分母为异常状态下的样本平均值的标准差。在变量遵从正态分布的前提下，当式(8)t 的绝对值大于 1.0 时认为有一定的可预报性。

如果把异常气候态和正常气候态认为是

不同的外力作用造成时，即把气候状态分为两类，异常态的样本归为一类，正常态的样本归为另一类，形成两组样本。异常态的判断归结为判断这两组样本平均值有无显著差异。这时可以用另一 t 统计量

$$t = (x_{m1} - x_{m2}) / [s_m / (1/n_1 + 1/n_2)^{1/2}] \quad (9)$$

式中， x_{m1} 为异常年份 n_1 气候态的平均值， x_{m2} 为正常态年份 n_2 的平均值， s_m 为两样本综合标准差，表示为

$$s_m = (n_1 s_1^2 + n_2 s_2^2) / (n_1 + n_2 - 2) \quad (10)$$

该 t 统计量遵从自由度为 $(n_1 + n_2 - 2)$ 的 t 分布。

把各站要素异常态的年份平均值绘在一张图上，称为合成图，还常用异常和正常态的差值绘在一张图上，在图上分析差值大的地区即为可能异常地区。但是，标准差大的地区，其值变化也大，即使不针对旱涝年份作差值图，差值的中心也仍可能出现在这些地区，因此很难认为这地区就是关键区。如果绘各站 t 统计量分布图则可避免这一缺点，而且还能从 t 值的统计检验中确定显著异常地区。

5 如何划分气候级别

正如上面指出，为了研究方便，常根据定性描述资料或定量数据转换成气候级别，以便表现不同的气候态。在对定量数据转换成自然数的级别过程中，不同程度的降低了数据变化的差异程度，能使级别变量趋于正态分布。在把降水量连续型变量转化为级别变量时，也称为旱涝指标。常见的几种旱涝指标有：

(1) 把降水量当成正态分布处理。用标准化降水量变量值来划分级别，把 <-2 、 $-2-1$ 、 $-1-1$ 、 $1-2$ 和 >2 的变量值称为特旱、旱、正常、涝和特涝的气候态。在这种划法中，特旱和特涝现象出现的概率约为 5%。也有用 5 个级别分界值 (-1.17 、 -0.33 、 0.33 、 1.17) 进行划分。

(2) 用降水量正负距平平均值划分。把距平取绝对值再平均并记为 d ，把 $<-2d$ 、 $-2d-1d$ 、 $-1d-1d$ 、 $1d-2d$ 和 $>2d$ 的

降水量称为特旱、旱、正常、涝和特涝。这种划法是把距平绝对值的平均值作为标准差的估计来进行划分。

(3)用降水量的百分比值划分,把降水量除以平均值得百分比值,以<40%、40%—80%、80%—120%、120%—160%和>160%来划分特旱、旱、正常、涝和特涝。这种划法是用距平百分率来进行划分,其界限值对应为<-60%、-60%—-20%、-20%—20%、20%—60%和>60%。

(4)用小于某一降水量值的概率进行划分,以<10%、10%—30%、30%—70%、70%—90%和>90%来划分特旱、旱、正常、涝和特涝。

由于降水量不遵从正态分布,以正态分布为基础的划分法会带来较大的误差。表1给出北京各季降水量概率分布的统计量估计值。

表1 北京降水量/mm 分布的偏度与峰度值
(1951—1980年间资料)

统计量	冬	春	夏	秋
平均值	12.8	61.1	482.7	87.6
标准差	8.51	39.48	208.82	53.78
偏度	0.76	1.02	1.24	0.46
峰度	0.36	0.25	2.09	-0.90

从表1可见,北京各季降水量均有异于0的偏度和峰度,实际统计检验表明它们均不遵从正态分布。因而若用常规划分法则可能会在识别旱涝状态上产生较大的误差。研究表明降水量能较好遵从Gamma分布。因此可以用Gamma分布计算出不同降水量对应的概率值、用概率值的划分法进行划分。

在气象要素序列演变中,不同状态的出现称为不同事件的发生。显然,在所研究的样本中,不同事件的发生频繁程度不同,可以用事件频率(用某事件发生的次数除以所有事件发生的总次数)度量它们。把不同事件发生的频率汇总成表,可以进一步分析不同事件发生的频率或频数分布状况。表2给出用上述几种划分级别方法的频数分布状况的比较(仍然用表1资料)。表中G表示用Gamma

分布概率值的划分法,N1表示用假定降水量为正态分布并把降水量转换为概率值的划分法,N2表示用标准化降水量值的划分法,5个级别分界值分别为-1.17、-0.33、0.33、1.17进行划分。从表中可见,用N1法划分时旱状态年份过多,而正常年份反而偏少。如果把4和5级合称为旱,把1和2级合称为涝,则各季旱均多于涝。这表明该划分法并不合乎常见的气候状态分布。用G和N2所划分的级别分布则较好地符合正态分布。但是检查各年这两种划法所得的级别差异发现,在N2法划分旱状态多于涝状态,且多集中在4级上面5级又过少。而正常气候状态下1与5级在30年中约出现3次。在N2划法中各季5级均少于3次,特别在春季竟无1次出现。例如1972年春季降水量为11.3mm,标准化值为-1.26,据N2法划为4级,但该年降水量是30年中最小的。该季30年中降水量次小值出现在1960年(19.1mm),N2法也把它划为4级。而使用G法把上述两年均划为5级。看来G法更合理些。又如夏季N2法把1972和1980年划为4级,降水量分别为227.2和242.3mm与特旱的年份1968年的203.7mm相差不大。在N2法中却划为4级。G法把它们划为5级似乎合理些。

表2 两种划分法的旱涝级别频数比较

级 别	1	2	3	4	5	涝	旱
冬	G	2	8	12	5	3	10
	N1	4	6	7	10	3	10
	N2	2	6	12	8	2	8
春	G	4	5	13	4	4	9
	N1	4	5	5	13	3	9
	N2	4	5	13	8	0	9
夏	G	3	4	16	2	5	7
	N1	3	5	9	10	3	8
	N2	3	3	17	5	2	6
秋	G	4	7	11	3	5	11
	N1	4	7	5	9	5	14
	N2	4	5	10	9	2	9
全年	G	13	24	52	14	17	37
	N1	15	23	26	42	14	38
	N2	13	19	52	30	6	32