

中国东部月降水量分布的统计特征

汤燕冰

(杭州大学地理系)

提 要

本文利用中国东部(105°E以东)110个站的30年(1953—1982)月降水量资料,分析了月降水量分布的统计特征,经计算,月降水量的变差系数大多在0.5—1.0之间,大部分站点月降水量的频数分布呈较明显的正偏形式,但偏斜程度因时间、地点的不同而异。假设检验的结果表明:在统计意义下($\alpha=0.05$),各地都有一些站点的月降水记录遵从正态分布。这些站点的时空分布反映了我国东部干湿季明显的季风气候特点。各地遵从正态分布站点数较高的月份大体上与该地变差系数较小的月份相对应,而这些月份均位于各地的多雨时段。对于其它一些无法用正态分布拟合的月降水的频数分布,绝大部分可用三种偏斜分布模式较好地拟合。

一、前 言

降水是最重要的气候要素之一。在国民经济各部门,尤其在农业生产和水利工程方面,都需要各种不同时段降水统计分析资料。在各种不同时段降水量中,月降水占有较重要的地位。这是因为对一地的降水特点来说,月降水量的代表性优于年或日降水量。它不但可以较好地反映不同季节的降水特点,而且可以反映降水的季节变化情况。已有人^[1-3]从不同的角度对月降水量的分布及其特点进行了研究,但与对年降水量的研究工作相比,还差得很远。本文的目的在于通过较全面地分析月降水量的统计特征,为月降水资料的分析和应用提供一些依据。

二、资料处理

本文选用中国东半部(105°E以东)110个站点的30年月降水量资料。资料取自《中国降水量变率资料》及《中国地面气象记录月报》。为保证各站记录年代一致以便于比较,故所选的30年时段为1953—1982。考虑到南北气候上的差异,为了讨论问题的方便,根据不同的地理位置,自北向南将我国东部分为五个区,这五个区依次为:东北、华北、长江中下游、长江上游和华南(见图1)。

由于各区所处的地理位置不同,因此降水的年变程有明显差异^[4]。I、II两区的最多雨月为7、8月,干季为10—4月。III区的最多雨月为4—6月,干季为10—2月,7和8月为伏旱季节。IV区的最多雨月为6—9月,干季为11—3月。V区的最多雨月为5—6月(其中雷州半岛以南地区为8—10月),干季为10—3月。

三、分析结果

我们依次对各站点的30年逐月降水总量计算其基本统计量. 这些统计量主要为:

$$\text{均值: } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\text{方差: } s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\text{偏态系数: } r_1 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 / s^3$$

$$\text{峰态系数: } r_2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4 / s^4 - 3$$

$$\text{变差系数: } c_v = s / \bar{x}$$

同时还计算了各月降水量的经验分布函数, 由此来讨论月降水量分布的统计特征.

1. 频数分布形式

从经验分布函数的计算结果来看, 除个别站的个别月份外, 绝大部分站点的月降水总量的频数分布呈较明显正偏形式, 但对于不同的时间及地点, 偏斜程度各异. 表1给出了各区逐月偏态系数 r_1 的平均值.

由表1可见, 各区雨季的偏度基本上比干季小, 也就是说雨季月降水的频数分布更

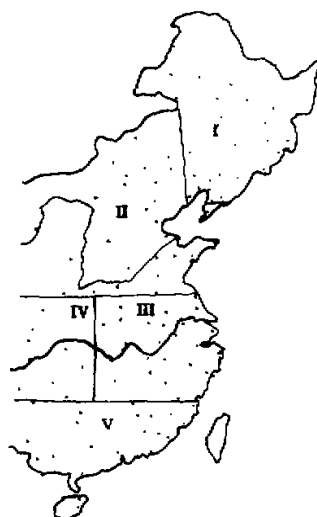


图1 分区及站点分布示意图

分区后, 110个站点的分布情况为: I区29个; II区28个; III区22个; IV区9个; V区22个.

表1 各区逐月偏态系数 r_1 的平均值

	1	2	3	4	5	6	7	8	9	10	11	12
I	1.3748	1.5090	1.2335	0.7162	0.7884	0.6252	0.5386	0.7529	0.5774	0.9304	0.9612	1.7351
II	1.5023	1.3495	1.1677	1.2270	1.2105	1.0569	0.9395	0.9029	0.8448	1.1060	1.3085	1.5922
III	0.9088	0.7110	0.5115	0.7818	0.7197	0.9620	1.1204	1.0025	0.9417	1.0337	0.9778	1.1791
IV	0.4760	0.8026	0.8178	0.8229	0.6656	1.0886	0.8554	1.0526	0.7013	0.6401	0.7994	0.6920
V	1.3708	1.3146	0.5006	0.9837	0.4849	0.7346	0.9184	0.6376	0.8637	0.9904	1.3786	1.4387

接近于正态分布(正态分布的 r_1 为零). 我们知道, 若一随机变量是许多微小独立随机因素的总后果, 而在正常情况下, 每种因素都不能起压倒一切的主导作用, 则该变量遵从正态分布. 而一地雨季的形成, 正是受到许多因素的共同影响, 固然在每一次降水过程中, 总有某一个或几个因素起主导作用, 但月降水量往往包括了多次降水过程的降水量, 因此雨季月降水量较易接近正态分布. 与其相反, 一地干季的形成常常有某个因素起主要作用, 如冬季各区普遍为干季, 则是因为冬季风盛行, 气流干燥. III区(长江中下游地区)7, 8月的伏旱则是由于副热带高压控制该区的结果. 因此, 干季月降水量的偏态系数 r_1 相对于雨季来讲就较大. 春秋过渡季节为冬夏大气活动中心调整时期, 移动性气压系统活跃, 气旋和反气旋活动频繁, 在这种环流背景下, 就易使得 r_1

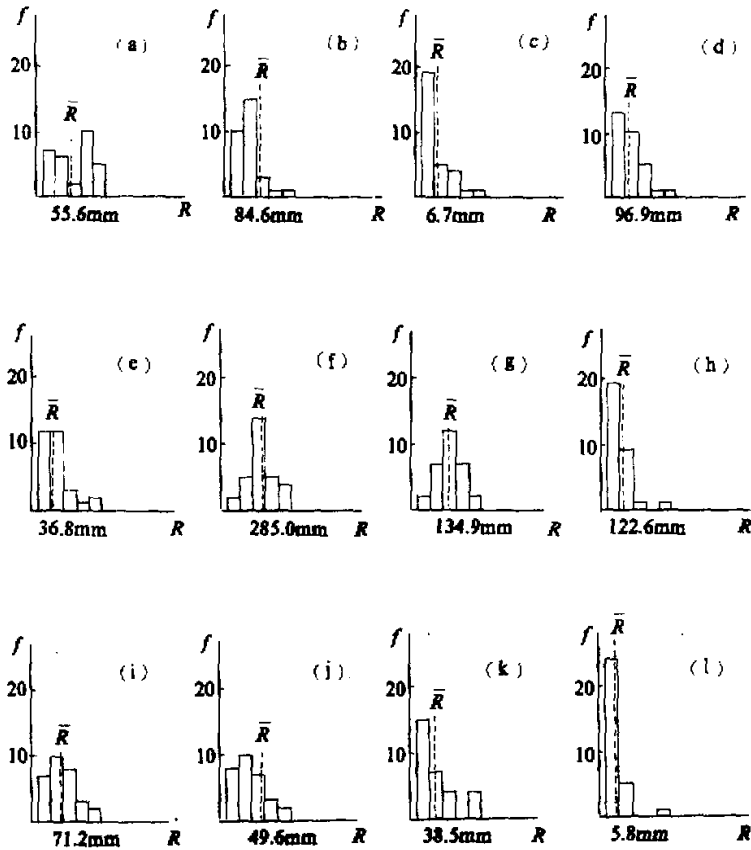


图2 12站月降水量频数分布图

- a: 一月, 吉安, 27.07°N , 114.58°E , $r_1 = -0.1443$, $r_2 = -1.4283$;
 b: 二月, 杭州, 30.14°N , 120.50°E , $r_1 = 1.4384$, $r_2 = 2.6553$;
 c: 三月, 天津, 39.06°N , 117.10°E , $r_1 = 1.7114$, $r_2 = 2.9629$;
 d: 四月, 南宁, 22.49°N , 108.21°E , $r_1 = 1.4091$, $r_2 = 2.4641$;
 e: 五月, 济南, 36.41°N , 116.59°E , $r_1 = 1.4061$, $r_2 = 1.4841$;
 f: 六月, 永安, 25.58°N , 117.21°E , $r_1 = 0.6647$, $r_2 = 0.1422$;
 g: 七月, 前郭尔罗斯, 45.07°N , 124.50°E , $r_1 = 0.4766$, $r_2 = -0.0472$;
 h: 八月, 芷江, 27.27°N , 109.41°E , $r_1 = 2.9779$, $r_2 = 10.9903$;
 i: 九月, 郑州, 34.43°N , 113.39°E , $r_1 = 0.9942$, $r_2 = 0.6533$;
 j: 十月, 上海, 31.10°N , 121.26°E , $r_1 = 1.0653$, $r_2 = 0.4461$;
 k: 十一月, 广州, 23.08°N , 113.19°E , $r_1 = 0.9092$, $r_2 = -0.4272$;
 l: 十二月, 哈尔滨, 45.41°N , 126.37°E , $r_1 = 3.7356$, $r_2 = 15.4148$.

较小, III, IV 区的最小值出现在3月, I, II 区出现在9月就证实了这一点。

由于频数分布的形式为正偏, 因此月降水量出现负距平的次数高于正距平, 这样正距平值往往就较大, 而较大的距平值对于分布的峰态影响很大。经统计分析, 某站某

月只要出现过大于三倍标准差的正距平值的年份，则该月的频数分布为高峰态(通常所讲的高峰态是指峰态系数 r_2 大于零的情况，而这里所讲的高峰态其 r_2 均超过1)，概括率为85%。由此可见，由于某些极端异常值的影响，可以使总体分布大大偏离正态分布(正态分布的 r_2 亦为零)。图2给出了各种典型的频数分布形式。

2. 分布模式拟合

利用月降水量的偏、峰态系数对月降水量的总体分布形式做统计假设检验。结果表明，在统计的意义下($\alpha=0.05$)，各个不同地区各月都有一些站点的月降水量记录遵从正态分布(表2)。

表2 各区逐月服从正态分布的站点数及变差系数 <0.5 的站点数统计表

项目	月份 站点数												
		一	二	三	四	五	六	七	八	九	十	十一	十二
I 分布		4	1	8	17	14	18	23	16	22	6	14	7
区													
(29) 变差系数		1			2	9	13	20(19)	6	3	1	1	2
II 分布		3	6	8	3	7	5	15	14	14	9	7	1
区													
(28) 变差系数					1	2	2	12(8)	3	1			
III 分布		12	14	19	12	14	9	8	8	9	7	9	4
区													
(22) 变差系数			1	15(14)	16	12	8						
IV 分布		5	4	5	4	6	5	4	5	6	7	4	6
区													
(9) 变差系数		2	1	5	5	7(6)	5	5		2	4(3)	3	1
V 分布		5	6	17	9	18	13	9	16	13	8	4	2
区													
(22) 变差系数			1	8	9	15(14)	13	3	10	1			

注：表中各系数栏括号中的数为所统计的二项皆满足的站点数。

从表2可以看出，各区普遍在多雨月份里服从正态分布的站点数较高，而干季则较低。这与前面所提到的雨季月降水分布的偏斜程度较干季小的结论是一致的。由此说明降水量的频数分布形式在很大程度上受到了各影响系统的共同作用，以及它们之间的相互关系的影响。

至于相当一部分不符合正态分布的月降水记录，可以用各种偏斜分布模式去拟合。我们所采用的模式主要为皮尔逊第III型频数模式(Γ 模式也属于此类^[5])，对数正态分布模式和格兰姆-沙里叶级数模式。么枕生曾对这几种模式的拟合计算作过详细的介绍^[6,7]。

在实例拟合计算过程中，我们发现这三类模式虽然同为正偏模式，都可以用来拟合通不过正态性检验的月降水记录，但在实际应用中还是有所区别的，也就是说要根据频数分布的不同偏、峰态来选择合适的模式。

皮尔逊III型频数模式要求： $2\beta_2 - 3\beta_1 = 6$ (式中的 $\beta_1 = r_1^2$ ， $\beta_2 = r_2 + 3$)。上述条件成立时，则频数分布可以用皮尔逊III型频数模式较好地拟合。由于我们在计算中

是用样本的偏、峰态系数来代替总体的偏、峰态系数,故上述差值只要达到5左右,

表3 皮尔逊Ⅲ型曲线拟合计算结果

杭 州 (2月) $2\beta_2 - 3\beta_1 = 5.10$		南 宁 (4月) $2\beta_2 - 3\beta_1 = 4.97$	
理论频数	实测频数	理论频数	实测频数
3.06	5	2.96	3
14.50	16	14.52	15
7.16	5	7.29	9
2.74	3	2.78	2
0.94	1	0.95	1

即可收到较好的拟合效果.表3中所给出的两个拟合实例的差值分别为5.10(杭州,2月)及4.97(南宁,4月).若该值在4.5以下,则皮尔逊Ⅲ型模式就难以较好地拟合该月降水记录.

对数正态分布对于正偏高峰态($r_1 > 0, r_2 > 0$)的频数分布有较好的拟合效果.若 $r_2 < 0$,则对数正态分布模式就不适用.这种情况可以采用格兰姆-沙里叶级数模式去拟合.表4和表5分别为上述二种模式拟合的实例.

表4 对数正态分布模式拟合计算结果

天 津 (3月) $r_1 = 1.7114 \quad r_2 = 2.9629$		芷 江 (8月) $r_1 = 2.9779 \quad r_2 = 10.9903$		郑 州 (9月) $r_1 = 0.9942 \quad r_2 = 0.6533$	
理论频数	实测频数	理论频数	实测频数	理论频数	实测频数
10.06	12	18.31	19	5.97	7
11.67	10	8.47	9	10.79	10
5.12	5	2.15	1	7.57	8
1.91	2	0.66	0	3.43	3
1.10	1	0.42	1	1.85	2

表5 格兰姆-沙里叶级数模式拟合计算结果

济 南 (5月) $r_1 = 1.4061 \quad r_2 = 1.4841$		上 海 (10月) $r_1 = 1.0653 \quad r_2 = 0.4461$		广 州 (11月) $r_1 = 0.9092 \quad r_2 = -0.4272$	
理论频数	实测频数	理论频数	实测频数	理论频数	实测频数
12.51	12	7.56	8	13.47	15
10.90	12	10.07	10	7.16	7
3.54	3	6.88	7	4.52	4
1.36	1	2.87	3	2.68	0
1.70	2	2.62	2	2.18	4

当然对于 $r_2 > 0$ 的情况,格兰姆-沙里叶级数模式也有较好的效果.一般来说,当 $r_1 > 1, r_2 > 1$ 时,采用对数正态模式较为合适,反之则以格兰姆-沙里叶级数模式为佳.

为了进一步说明上述三种偏斜分布模式对月降水记录的拟合能力,我们同时还对记录年代较长的上海站的记录进行拟合分析,以进行比较.表6为三种模式对上海100年月降水记录拟合适度的 χ^2 检验计算结果.从表中可见,对于1月,皮尔逊Ⅲ型模式的拟合效果最好,8月为格兰姆-沙里叶级数模式,10月则首推对数正态分布模式.由此表明我们对30年记录分析所得的结论是有代表性的.

表 6 三种模式对上海 (1873—1972) 月降水记录拟合适度的 χ^2 检验计算结果

χ^2 模 式	月 份	一 月	八 月	十 月
		$r_1 = 1.3689$ $r_2 = 2.6808$ $2\beta_2 - 3\beta_1 = 5.74$	$r_1 = 0.6911$ $r_2 = -0.0666$ $2\beta_2 - 3\beta_1 = 4.43$	$r_1 = 1.7$ $r_2 = 3.2433$ $2\beta_2 - 3\beta_1 = 3.82$
	皮尔逊 III 型频数模式	7.1972	7.4474	11.5235
	对数正态分布模式	10.6764	5.2195	7.2826
	格兰姆—沙里叶级数模式	28.9629*	3.6816	53.8039*

注：表中带“*”号者表示通不过显著性检验 ($\alpha = 0.05$)。

然而，仅正态分布及上述三种偏斜分布模式还不足以全面描述所选各点的各月降水频数分布形式。如图 2 中所示的吉安 1 月份降水记录的频数分布形式为较明显的负偏分布，这就需要用其它偏斜分布模式来拟合。又如哈尔滨 12 月份的频数分布形式也是很有代表性的。这种由于一个异常大值(距平为标准差的 4.8 倍)的出现，使频数分布呈正偏且异常尖峰态的情况，用以上几种偏斜分布模式是无法拟合的。曾有人推测这主要是由于样本均值对于异常值是非稳健的^[8]，因此采用中位值代替均值进行模式拟合计算，但没有得到令人满意的结果。若去掉这一异常值，则其频数分布遵从正态分布。我们在分析上海资料时也遇到类似的问题。利用上海 30 年(1953—1982)逐月降水记录，对总体分布形式进行统计假设检验时，有 8 个月(2, 3, 4, 5, 6, 7, 9, 11 月)可通过正态分布的显著性检验。若利用 100 年(1873—1972)资料，则只有 2 个月(5 和 7 月)可通过正态分布检验。检查 2, 3, 4, 6, 9, 11 月记录，发现各月均有异常大值(距平为标准差的 3 倍以上)出现过。去掉这一异常大值，则 2, 3, 4, 6 月记录均可通过正态分布的显著性检验，9, 11 月接近显著水平。可见异常值对于总体分布确有很大的影响。对于这一类受异常值影响的频数分布的模式拟合问题，还要进一步探索。关键问题在于：弄清是否由于这些异常值的出现而歪曲了总体的分布形式，或是必须用一些更复杂的偏斜分布模式(必须是具有长尾的)才能较好地拟合这一类的频数分布形式。

3. 变差系数

利用各站逐月降水的变差系数(相对标准差)来比较不同地区月降水的年际变化情况。经计算，所选各站点的逐月变差系数大多在 0.5—1.0 之间，最大的为 2.39 (承德, 1 月)，最小的为 0.3088 (长沙, 4 月)。表 7 列出了各区逐月平均变差系数。表 7 表明：各区变差系数的最小值都出现在最多雨月，最大值都出现在干季。

表 7 各区逐月平均变差系数

	1	2	3	4	5	6	7	8	9	10	11	12
I	0.9294	1.0349	0.8503	0.6628	0.5781	0.5104	0.4580	0.5557	0.5902	0.7235	0.7537	1.0579
II	1.2327	1.0708	0.8709	0.8393	0.7627	0.6732	0.5364	0.6077	0.7314	0.8137	0.9723	1.2393
III	0.6713	0.5970	0.4765	0.4626	0.4920	0.5880	0.7399	0.7189	0.7353	0.8214	0.8040	0.8655
IV	0.5394	0.6094	0.5247	0.4722	0.4396	0.5393	0.5529	0.6357	0.5895	0.4930	0.6078	0.6452
V	0.9243	0.7519	0.6030	0.6078	0.4773	0.4910	0.6130	0.5110	0.6665	0.8962	1.0541	1.0786

必须指出的是：在月降水量能用正态分布拟合的站点数较高的月份里，各站点相应时段的变差系数往往普遍较小，这在表 2 中体现得很清楚。以 I 区为例，7 月

份全区有 23 个站的降水记录可用正态分布拟合, 而变差系数小于 0.5 的有 20 个站, 其中 19 个站是二者兼备的. 另外几个站(服从正态分布, 但变差系数大于 0.5)的情况较特殊. 如克山(48.03°N , 125.53°E) 7 月份的降水服从正态分布, 而变差系数为 0.5378, 略大于 0.5, 但这已是该站的最小值. 这说明月降水的分布形式在某种程度上受到了降水量年际变化的影响.

表 8 上海 4 个月月降水量正态化处理的计算结果

		未 处 理	平方根法处理	立方根法处理
1 月	r_1	0.8296 *	-0.2634	-1.1976
	r_2	0.2609	0.3111	2.9576 *
8 月	r_1	1.3879 *	0.0863	-0.7631
	r_2	2.3601	0.1970	1.6581 *
10 月	r_1	1.0653 *	0.1584	-0.1432
	r_2	0.4461	-0.8936	-0.9295
12 月	r_1	0.9492 *	0.1425	-0.2303
	r_2	0.1224	-0.7453	-0.4868

注: 表中带“*”号处表示该月通不过正态分布显著性检验.

4. 正态化处理

由于许多常用的统计函数是建立在正态分布的假设之上, 因此在对月降水量进行分析和研究时, 就必须考虑到正态化处理的问题. 降水量的正态化处理方法较多, 如平方根法、立方根法、函数变换法等. 我们以上海通不过正态性检验的 4 个月(1, 8, 10, 12)的月降水量记录为例, 分别用平方根法和立方根法进行处理, 结果列在表 8 中. 经比较, 我们发现: 平方根法优于立方根法. 因为月降水量的偏斜度介于日降水量与年降水量之间, 立方根法对于偏斜不是很大的情况, 往往容易处理过头, 使其由正偏形式变为负偏形式, 如上海的 1 月与 8 月, 因而达不到处理的目的.

四、结 语

由于月降水量的时空变化很大, 因此月降水的分布是很复杂的. 本文仅仅作了一些初步的探索, 许多问题还有待于今后进一步深入研究.

致谢: 本文曾得到王德翰教授、姚樟荣副教授的指导和审阅. 谨此致谢.

参 考 文 献

- [1] 么枕生, 1959, 中国东部境内各月降水量的保证机率, 南京大学学报(自然科学版), No. 6, 49—60.
- [2] 朱炳海, 1957, 中国的降水变率和旱涝, 南京大学学报(自然科学版), No. 4, 23—44.
- [3] 孙安健, 1986, 我国平均降水相对变率和降水量统计关系的研究, 气象学报, 44, 102—107.
- [4] 张家诚等, 1985, 中国气候, 上海科学技术出版社, 176—186.
- [5] Essenwanger, O., 1976, *Applied Statistics in Atmospheric Science*, Elsevier, Scientific Publishing Company, 57—62.
- [6] 么枕生, 1962, 气候统计, 科学出版社, 160—162.
- [7] 么枕生, 1984, 气候统计学基础, 科学出版社, 177—190.
- [8] 复旦大学, 1979, 概率论, 第二册第二分册, 人民教育出版社, 291—295.