

长江流域月季降水量的概率分布特征分析

杜良敏¹ 刘绿柳² 高雅琦¹ 肖莺¹ 李杨³

(1 武汉区域气候中心, 武汉 430074; 2 国家气候中心, 北京 100081; 3 北京师范大学, 北京 100875)

摘要: 利用长江流域712站1951—2015年的月、季降水资料, 分析了站点降水在正态分布和 Γ 分布的空间服从情况。结果表明: 对于月降水, 大部分站点能服从 Γ 分布, 通过 Γ 检验的站点表现出时空差异, 两湖流域表现尤为突出, 两湖流域在1和7—12月 Γ 特性表现较好, 但在2—6月通过 Γ 检验的站点较少。季降水量服从正态分布的比例要高于 Γ 分布, 春、冬季降水的正态分布服从性要好于夏、秋季; 而对于 Γ 分布, 春、夏季的服从性要好于秋、冬季。月、季降水中都存在一定比例的站点同时服从两种分布或都不服从这两种分布。

关键词: 长江流域, 降水量, 正态分布, Γ 分布, 统计检验

DOI: 10.3969/j.issn.2095-1973.2018.04.010

Comparison and Analysis of Probability Distribution Characteristics of Monthly and Seasonal Precipitation in the Yangtze River Basin

Du Liangmin¹, Liu Lüliu², Gao Yaqi¹, Xiao Ying¹, Li Yang³

(1 Wuhan Regional Climate Centre, Wuhan 430074 2 National Climate Centre, Beijing 100081

3 Beijing Normal University, Beijing 100875)

Abstract: The normal distribution and Gamma distribution of precipitation in Yangtze River Basin were analyzed by using monthly and seasonal precipitation data at 712 stations from 1951 to 2015. The results show: most stations passed Gamma distribution test for monthly precipitation, and small part of stations for normal distribution test. The stations that passed the Gamma test showed spatio-temporal differences, such as those in Poyang Lake and Dongting Lake are good on Gamma distribution for January, July-December, and few sites that passed the Gamma test from February to June. For seasonal precipitation, the normal distribution is more popular than those in the Gamma distribution. The normality for spring and winter was better than that for summer and autumn. The Gamma distribution, in spring and summer are better than that in autumn and winter. a nother part of stations obeyed both distributions or disobeyed each of.

Keywords: Yangtze River Basin, precipitation, normal distribution, Gamma distribution, statistics test

0 引言

长江流域位于 $24^{\circ}27'—35^{\circ}54'N$, $90^{\circ}33'—122^{\circ}19'E$, 流域面积约180万 km^2 ^[1], 地理跨度大, 地势变化幅度大, 降水变率明显, 导致了长江流域的降水特征无论是在时间上或是空间上都存在较大差异, 给降水特征的研究带来诸多不便, 这对气候预测中统计方法的使用提出了更高要求。

在气候统计中, 如回归分析、判别分析等统计方法模型均要求预报对象服从正态分布^[2-3]。对于不符合正态分布的站点, 需要观察其频率直方图, 通过函数

拟合来判断其属于何种分布类型, 以便对数据进行正态化处理^[4-7]。通常关注的概率分布函数有正态分布和 Γ 分布。徐尔灏^[8]对我国月降水的理论分布进行研究检验, 发现我国各地年降水量基本上服从于正态分布。丁裕国^[9]从理论上证明了 Γ 概率分布模型在降水拟合上具有普适性, 对任意时段降水量概率分布的拟合效果都较好。梁莉等^[10]利用 Γ 分布拟合了淮河流域日降水量概率分布, 吴子君等^[6]的研究显示 Γ 分布函数在月、季降水中有较好的适用性。这些研究表明 Γ 分布对降水概率分布具有较好的描述能力。

目前针对分析长江流域在月和季节尺度的上的降水在正态分布和 Γ 分布的对比分析较少, 本文利用长江流域1951—2015年的月、季降水资料, 采用这两种分布, 分析了长江流域降水量分布特征, 对通过检验的站点开展了空间分析, 这一工作对长江流域降雨分

收稿日期: 2017年12月15日; 修回日期: 2018年7月4日

第一作者: 杜良敏(1976—), Email: duliangmin@qq.com

资助信息: 国家重点研发计划(2016YFE0102400, 2017YFC1502306); 中国长江电力股份有限公司项目资助

布特征分析和气候预测具有参考意义。

1 资料与检验方法

1.1 资料

本文使用国家气象信息中心提供的1951—2015年长江流域内经过地理信息筛选的712个气象站的逐日降水资料。因20世纪50年代前期观测站点较少，具体计算时以气象站建站时间为起点，剔除缺测降水数据，并将月降水数据处理成季降水数据，形成了月、季节尺度的降水历史数据序列。

1.2 降水概率分布检验方法

1.2.1 偏度系数、峰度系数检验法

目前常用的正态分布检验方法有 χ^2 拟合优度检验、偏度系数峰度系数检验法。 χ^2 检验法虽然是检验总体分布的较一般的方法，但是用它来检验总体的正态性时，没有用到正态的特性，第II类错误的概率往往较大^[1]，为此本文采用偏度系数、峰度系数检验法。

偏度 (Skewness) 和峰度 (Kurtosis) 是描述数据分布形态的统计量，偏度描述的是总体取值分布的对称性，峰度是描述总体取值分布的陡缓程度^[12]。

样本偏度定义为：

$$g_1 = \frac{m_3}{m_2^{3/2}} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^{3/2}}, \quad (1)$$

样本峰度定义为：

$$g_2 = \frac{m_4}{m_2^2} - 3 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^2} - 3, \quad (2)$$

式中， n 为样本长度， m_2 是二阶中心矩（即样本方差）， m_3 是三阶样本中心矩， m_4 是四阶样本中心矩， x_i 是第*i*个值， \bar{x} 是样本平均值。

若总体 X 为正态变量，则可证得当 n 充分大时，近似地有

$$g_1 \sim N\left(0, \frac{6(n-2)}{(n+1)(n+3)}\right), \quad (3)$$

$$g_2 \sim N\left(3 - \frac{6}{n+1}, \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}\right). \quad (4)$$

设 X_1, X_2, \dots, X_n 是来自总体 X 的样本，检验假设

$$H_0: X \text{ 为正态总体}. \quad (5)$$

记：

$$\sigma_1 = \sqrt{\frac{6(n-2)}{(n+1)(n+3)}}, \quad (6)$$

$$\sigma_2 = \sqrt{\frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}}. \quad (7)$$

当取显著性水平为 $\alpha=0.05$ 时， H_0 的拒绝域为

$$|g_1| > 1.96\sigma_1, \quad (8)$$

$$|g_2| > 1.96\sigma_2. \quad (9)$$

1.2.2 Kolmogorov–Smirnov 检验法

对于 Γ 分布通常使用Kolmogorov–Smirnov检验法，该方法是一种非参数统计分析的检验方法，简称为K-S检验法^[13-14]。K-S检验法既可以做单样本检验也可以做双样本检验。单样本检验是和特定的连续概率分布函数（正态分布或者其他分布函数）进行比较，而双样本K-S检验是比较两个样本是否来自同一个总体。双样本K-S检验是比较两个样本最有效的而且也是常用的非参数比较方法，其思想是，确定样本的相关经验概率分布方程和累积分布方程之间的距离，或者是确定来自两个样本的经验概率分布方程之间的距离，如果两者之间的距离很小，则可推断样本的概率分布和经验分布拟合得很好。具体步骤如下，假设 H_0 ：总体分布 X 来自理论分布。

K-S统计方程如下：

$$D_{n,n'} = \sup_x |F_{1,n}(x) - F_{2,n'}(x)|, \quad (10)$$

其中 $F_{1,n}$ 和 $F_{2,n'}$ 分别表示两个样本的经验概率分布方程， \sup 表示极大值函数（supremum function）。

原假设在显著性水平 α 拒绝域为：

$$D_{n,n'} > c(\alpha) \sqrt{\frac{n+n'}{nn'}}, \quad (11)$$

式中， n 和 n' 分别代表两个样本的长度，当显著性水平 α 取值为0.05时，通过查表得 $c(\alpha)=1.36$ 。当 $D_{n,n'}$ 值不在拒绝域时，接受假设 H_0 。

如果总体分布 X 的形式或者参数无法定义可以加入蒙特卡罗方法进行K-S检验。本文采用的方法是基于K-S检验的蒙特卡罗检验方法。

Γ 函数为：

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt, \quad (12)$$

Γ 概率密度方程如下：

$$f(t|A, C, D) = \frac{1}{C\Gamma(A)} \left(\frac{t-D}{C}\right)^{A-1} e^{-\frac{t-D}{C}}, \quad (13)$$

$$A > 0, C > 0, -\infty < D < +\infty, t > D$$

式中， A 、 C 和 D 分别代表形状参数（shape），尺度参数（scale）和阈值（threshold）三个参数。将观测的降水数据定义为 n ，使用最大似然法对 n 进行 Γ 参数拟合，得到三个参数 A 、 C 、 D 。将得到的3个参数代入式（13）生成一组大量符合 Γ 分布的伪随机数 n' ，将 n 、 n' 代入式（10）和式（11）计算出 $D_{n,n'}$ 和拒绝域。

1.3 降水量概率分布检验后处理

1.3.1 站点正态分布检验后处理

根据式(1)和式(2)可计算出长江流域712站的偏度和峰度值 g_1 、 g_2 ，取信度 $\alpha=0.05$ 的阈值，再根据式(8)和式(9)来确定该站点是否通过检验，通过检验的站点用“1”标示，未通过站点用“0”标示。

1.3.2 站点 Γ 分布检验后处理

根据生成的随机数 n' 和降水数据 n 做K-S双样本

检验，取信度 $\alpha=0.05$ 的阈值，根据蒙特卡罗的思想来确定站点是否通过检验，通过检验的站点用“1”标示，未通过站点用“0”标示。

1.3.3 空间插值

通过检验的站点赋值为“1”，没有通过检验的站点赋值为“0”。通过Kriging空间插值法得到检验结果的空间分布图，图1~图4中阴影代表通过统计检验的区域，空白是未通过统计检验的区域。

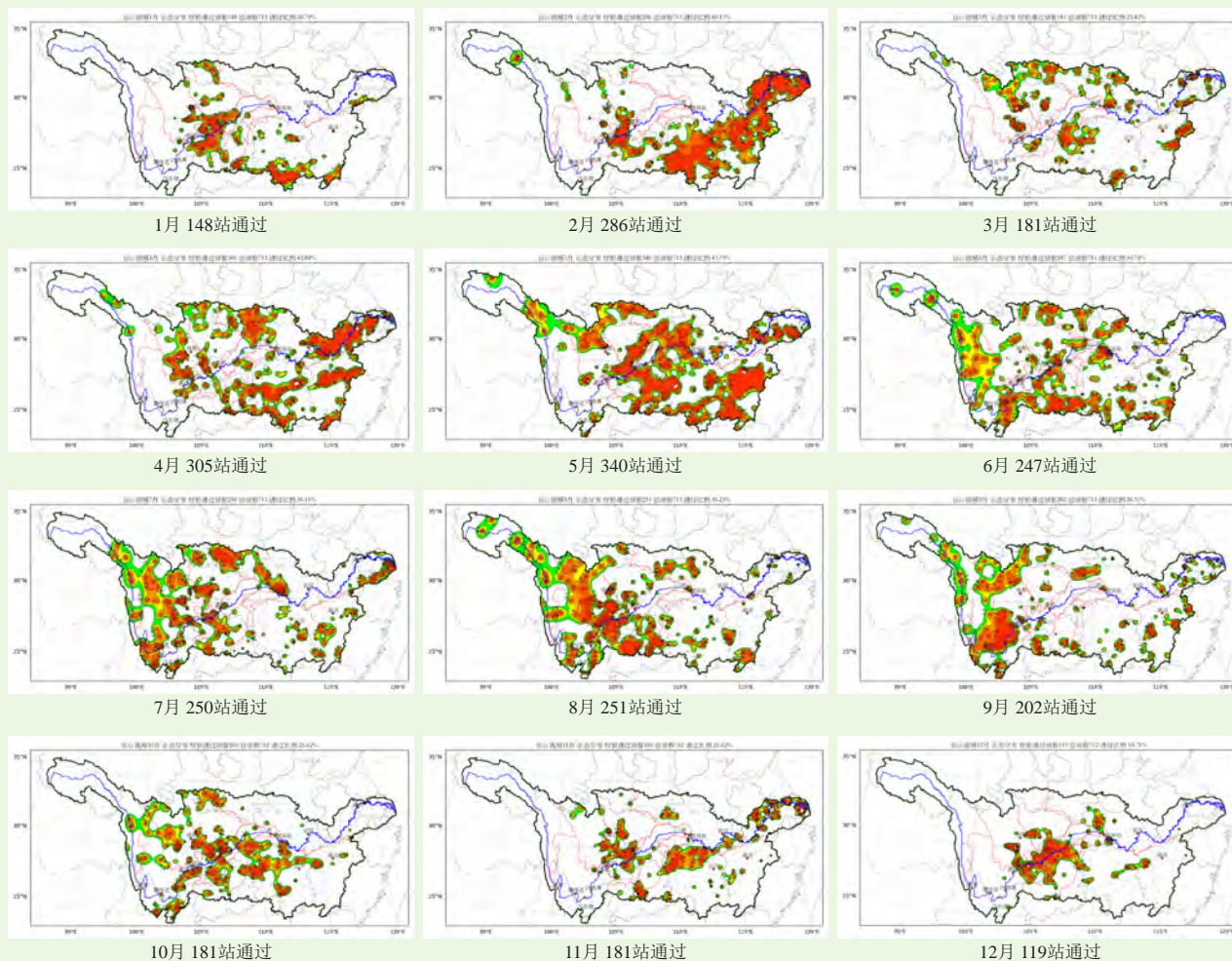


图1 正态分布偏度峰度检验空间分布

Fig. 1 Map of stations passed the normality test in monthly precipitation

2 结果分析

2.1 长江流域月降水量分布特征检验

长江流域气象站月降水量正态分布、 Γ 分布站点通过比例见表1，检验结果空间分布如图1和图2所示，通过检验的站点用颜色标注，没有通过检验的站点用白色标注。

由表1可见，长江流域站点的月降水量历年降水序列正态分布的假设检验通过率在16%~50%，均未

过半数，通过率最高的月份是5月，为47.8%；而长江流域通过 Γ 分布的假设检验通过率在50%以上的月份有11个，通过率最高的月份是11月，比率为75.8%；通过率最低的月份是6月，为47.5%， Γ 分布的假设检验通过率明显要好于正态分布的假设检验。

月降水通过正态检验的站点时空分布不均(图1)，长江中下干流区间仅在2、4和5月正态性表现较好，而金沙江流域在6—9月的正态性表现较好。同样，月

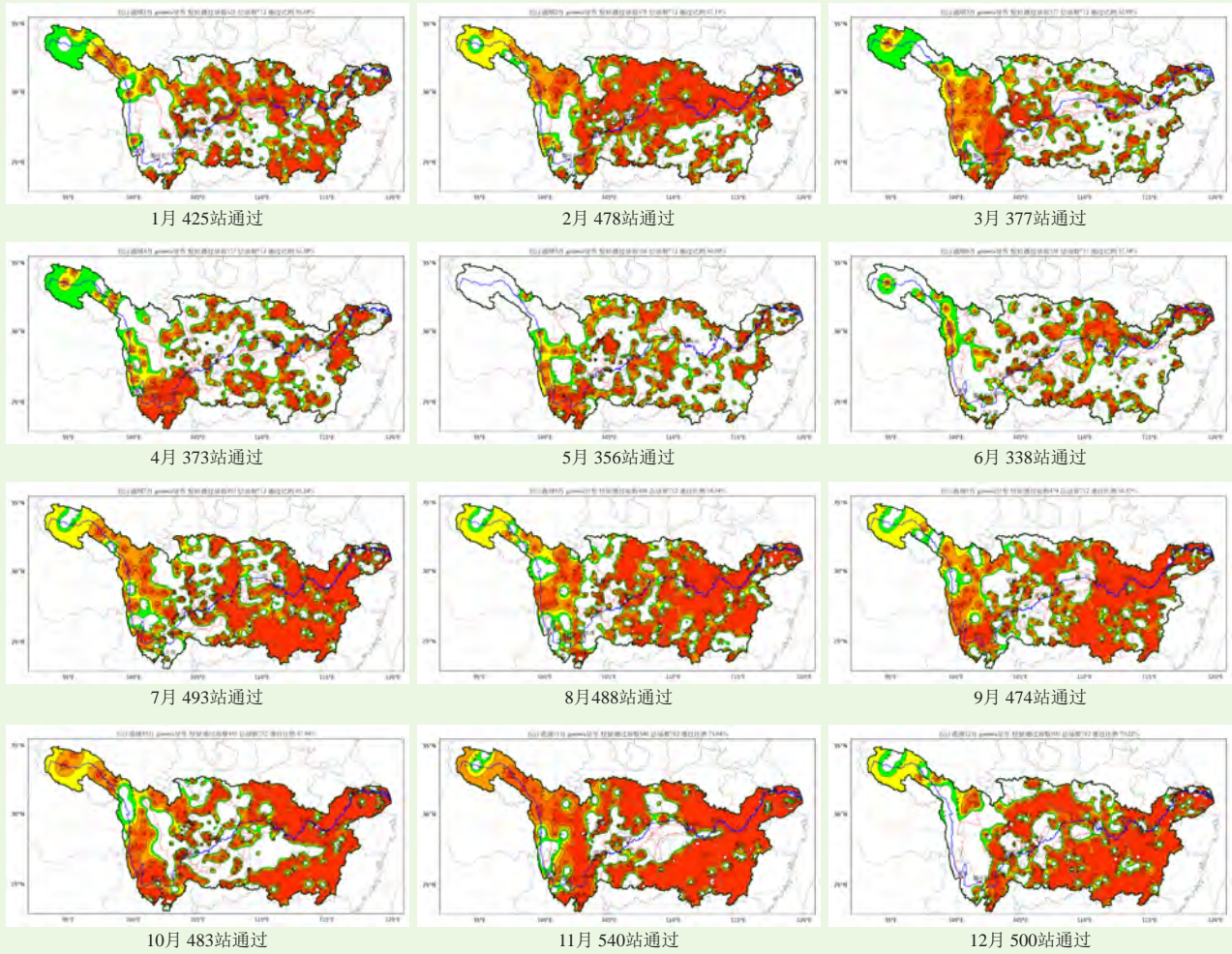


图2 Γ 分布K-S检验空间分布
Fig. 2 Map of stations passed Gamma test for monthly precipitation

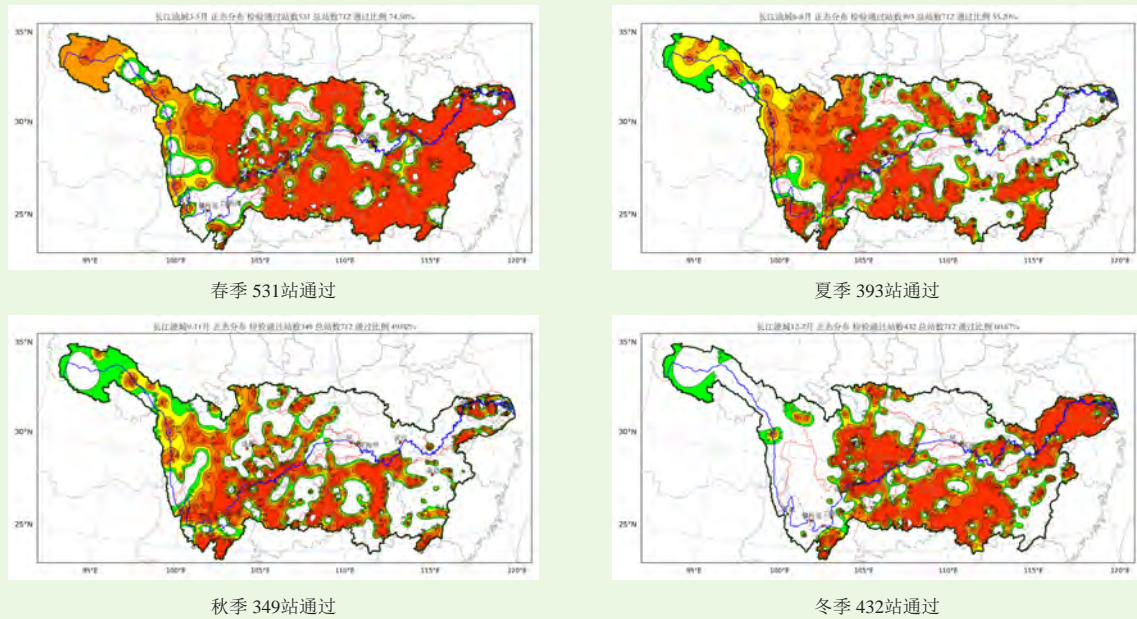


图3 正态分布偏度峰度检验空间分布
Fig. 3 Map of stations passed the normality test for seasonal precipitation

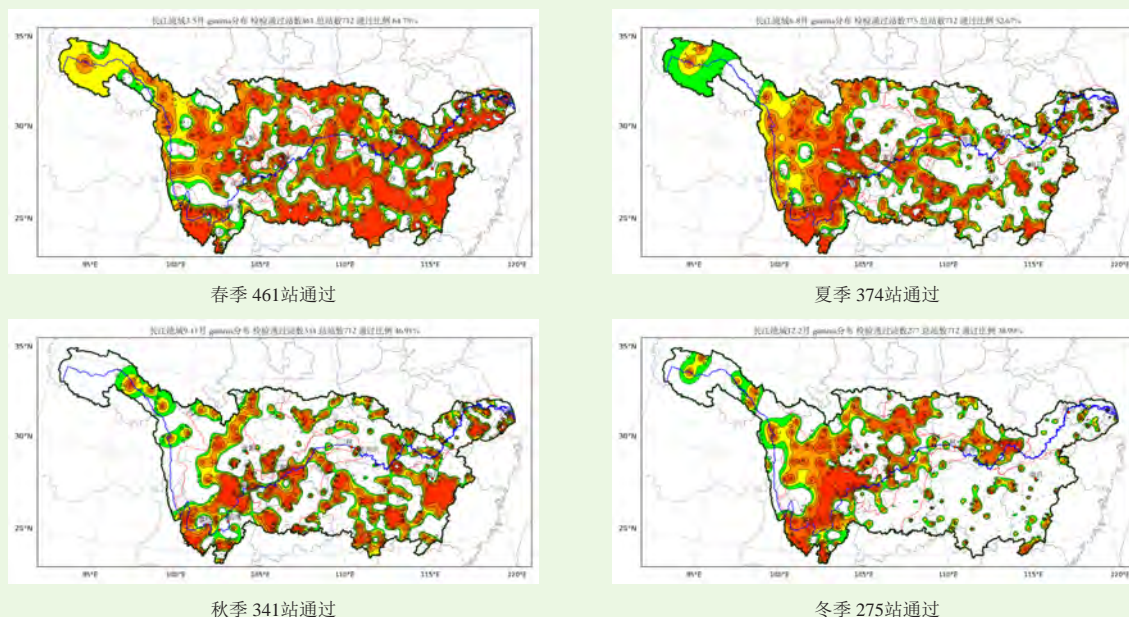


图4 Γ 分布K-S检验空间分布
Fig. 4 Map of stations passed the Gamma test for seasonal precipitation

表1 站点月降水量检验比例 (单位: %)

Table 1 The percentage of station number passed the distribution tests for monthly precipitation (unit: %)

月份	1	2	3	4	5	6	7	8	9	10	11	12
通过正态检验	20.8	40.2	25.4	42.8	47.8	34.7	35.1	35.3	28.4	25.4	25.4	16.7
通过 Γ 检验	59.7	67.1	53.0	52.9	50.0	47.5	69.2	68.5	66.5	67.8	75.8	70.2
通过正态检验但未通过 Γ 检验	9.1	18.6	11.9	19.3	22.9	16.7	14.2	15.9	12.8	13	10.8	9.1
未通过正态但通过 Γ 检验	48.0	45.5	39.5	28.9	25.1	29.5	48.3	49.2	51.0	55.5	61.5	62.6
两者同时通过比例	11.7	21.6	13.5	23.5	24.9	18.0	20.9	19.4	15.6	12.4	14.6	7.6
两者都不通过比例	31.2	14.3	35.1	28.2	27.1	35.7	16.6	15.6	20.6	19.1	13.3	20.6

降水通过 Γ 检验的站点也表现出时空差异(图2),其中两湖流域表现尤为突出,两湖流域在2—6月通过 Γ 检验的站点很少,但是在1和7—12月 Γ 特性表现较好。长江上游干流区间 Γ 特性一直表现较差。

图1、图2和表1所示,在2、8、10和11月通过正态检验站点与通过 Γ 分布检验站点具有一定的互补特征,但在1、3和6月站点既不符合正态分布也不符合 Γ 分布的比例超过30%,这些站点月降水量分布函数可能更适用于其他特征。同时也有一些站点既属于正态分布又属于 Γ 分布,这可能是该站点的月降水通过极大似然法估计出的 Γ 分布的形状系数较大,接近正态分布所致^[15]。

2.2 长江流域季降水量分布特征检验

取信度 $\alpha=0.05$ 的阈值,长江流域气象站季降水量正态分布、 Γ 分布站点通过比例见表2,检验结果空间分布如图3和图4所示。

结合表1和表2,长江流域季降水量正态性表现要

好于月降水的正态性,季降水量通过检验的站点比例平均值在59.8%,其中春季通过正态性检验的比例可以达到74.6%,秋季通过检验的比例最少为49.0%。与正态分布相比,通过 Γ 分布检验的比例略低,通过率最高季节是春季,春、夏季的表现好于秋、冬季。

对于月降水的分布检验来,大多服从 Γ 分布,但正态性的表现较差。而季降水的分布检验对比分析,正态表现与 Γ 特性差异较小,仅在冬季正态表现明显好于 Γ 特性。这可能是由于月降水年代际的变化大,

表2 站点季降水量检验比例 (单位: %)

Table 2 The percentage of station number passed the distribution tests for seasonal precipitation (unit: %)

季节	春季	夏季	秋季	冬季
通过正态检验	74.6	55.2	49.0	60.7
通过 Γ 检验	64.7	52.5	47.9	38.6
通过正态检验但未通过 Γ 检验	21.9	18.7	22.7	42.7
未通过正态但通过 Γ 检验	12.1	16.0	21.6	20.6
两者同时通过比例	52.7	36.5	26.3	18.0
两者都不通过比例	13.3	28.8	29.4	18.7

大量级降水和小量级降水均会出现,更容易为偏态分布,季节降水会平滑掉季节内的变化,年代际差异会较小,正态性和偏态性均有可能。

通过各个季节长江流域降水通过检验站点的空间分布图得出(图3和图4),对于正态分布检验,春季长江全流域大部通过检验;夏、秋两季,通过正态检验的站点主要位于长江中上游;而在冬季,通过正态检验的站点位于下游。对于 Γ 分布检验,春季长江流域大部通过检验,夏、冬两季在上游流域通过检验比例较高,秋季通过检验的站点分布较为分散。

与月降水类似,季节站点降水也有一些站点既不服从正态分布也不服从 Γ 分布,可能更适用于其他的分布函数。同时也有一些站点既服从正态分布又服从 Γ 分布,同时服从两种分布的比例要高于月降水(表1和表2)。

3 结论与讨论

了解长江流域月、季降水的概率特征,可为建立科学的降水预报模型提供可靠的理论依据。本文利用长江流域712个气象站的月、季降水资料,使用偏度、峰度和K-S检验法分别对其进行了正态分布检验与 Γ 分布检验,并对长江流域的降水量分布特征进行了空间分析。结果表明:

1) 长江流域站点月降水分布大都服从 Γ 分布,而服从正态分布站点的比例不超过50%。通过 Γ 检验的站点表现出时空差异,两湖流域表现尤为突出,1、7—12月 Γ 特性表现较好,但两湖流域在2—6月通过 Γ 检验的站点很少。

2) 长江流域季降水量分布特征是正态分布较好于 Γ 分布。春、冬季的正态性比夏、秋季表现要好;而对于 Γ 分布,春、夏季表现好于秋、冬季。

3) 从空间分布上分析,存在着同时服从和都不服从两种分布函数的情况,同时服从可能是通过降水分布估计出的 Γ 分布的形状系数较大,接近正态分布所致,同时不服从显示降水可能更适用于其他的分布函数。

本文对长江流域月、季降水进行了正态分布和 Γ 分布的服从性开展了探讨,但对不同区域内不同分布特征差异的成因分析,以及其他概率分布在长江流域的适用分析等还较为欠缺,这将在今后的工作中继续展开。

参考文献

- [1] 郝振纯,杨荣榕,陈新美,等. 1960—2011年长江流域潜在蒸发量的时空变化特征. 冰川冻土, 2013, 35(2): 408-419.
- [2] 方建刚,毛明策,程肖侠. 陕西降水的正态分布特征分析. 西北大学学报(自然科学版), 2009, 39(1): 131-136.
- [3] 赵佩红. 新会气候资料的正态分布检验及正态化处理. 广东气象, 2007, 29(1): 27-28.
- [4] 周雁翎,张琼雄,张练. 粤西降水量的分布特征及正态分布检验. 广东气象, 2007, 29(3): 32-34.
- [5] 马智博,朱建士,徐迺新. 有关正态分布的小样本可靠性评估. 核科学与工程, 2003, 23(4): 332-336.
- [6] 吴子君,张强,石彦军,等. 多种累积降水量分布函数在中国适用性的讨论. 高原气象, 2017, 36(5): 1221-1233.
- [7] 曹杰,陶云. 中国的降水量符合正态分布吗? 自然灾害学报, 2002, 11(3): 115-120.
- [8] 徐尔灏. 论年雨量之常态性. 气象学报, 1950, 21(1): 17-34.
- [9] 丁裕国. 降水量 Γ 分布模式的普适性研究. 大气科学, 1994, 18(5): 552-560.
- [10] 梁莉,赵琳娜,巩远发,等. 淮河流域汛期20 d内最大日降水量概率分布. 应用气象学报, 2011, 22(4): 421-428.
- [11] 周洪伟. 正态性检验的几种常用的方法. 南京晓庄学院学报, 2012, 28(3): 13-18.
- [12] 黄嘉佑. 气象统计分析与预报方法. 北京: 气象出版社, 2000: 19-27.
- [13] 南波. 基于K-S检验法的雪荷载统计分析. 兰州理工大学学报, 2012, 38(1): 115-119.
- [14] 石彦军. 新疆区域性极端干旱事件及其数值模拟和预测研究. 兰州: 兰州大学, 2014.
- [15] Ross S M. Introduction to Probability and Statistics for Engineers and Scientists. California: Elsevier Academic Press, 2004.