

# 中国三峡地区汛期降水量的正态性研究\*

黄嘉佑 黄茂怡 张 印 朱 蕾

(北京大学物理学院大气科学系, 北京, 100871)

## 摘 要

为了分析三峡地区降水量序列的正态性和谐结构,对降水量的常见各种变换进行试验性研究。试验包括单站降水量、降水量的平方根、立方根和  $\Gamma$  分布概率值等非线性变换,以及级别变换。研究区域多站平均降水量、区域降水量的主分量、区域降水量的非线性变换后的主分量和区域降水量的级别变换后的主分量等序列。研究发现三峡地区单站降水量的各种变换不改变序列原始谱结构,仅影响概率分布的偏度和峰度,使其较好地遵从正态分布,其中以  $\Gamma$  分布的变换以及级别变换在分布的偏度上为最好。

区域降水量的各种变换的综合指数(区域平均值和主分量)正态性及谐结构分析表明,除区域平均值变换后不改变原单站序列的谱结构外,主分量的综合指数能改变原单站序列的谱结构,同时也影响概率分布的偏度和峰度,使其能较好地遵从正态分布。其中以降水量的立方根和  $\Gamma$  分布概率变换以及级别变换,在分布的偏度上有较好的效果。

关键词: 正态,非线性,主分量。

## 1 引 言

在统计处理气象资料中,很多统计分析方法使用的变量,常常假定遵从正态分布。在统计预测方法中预报量区间估计和显著性检验上均使用这一假定。从对气候数据的分析可以知道,气象要素所遵从的概率分布常见的有:二项分布、泊松分布、正态分布、对数正态分布、截尾正态分布、 $\Gamma$  分布、Weibull 分布、 $\beta$  分布等等。在对一个数据进行分析时,应该首先判断这个数据服从何种分布。一般,常用在气候诊断分析和预测中的月降水量并不很好地遵从正态分布<sup>[1]</sup>。使用它作为预报对象会在预测中带来一定的困难。其原因在使用统计预测模型中,例如回归方程预报中,预报值是因子出现条件下预报量的期望值。如果预报量为正态分布,期望值自然是出现概率较大的值。因为只有变量遵从正态分布,其数学期望和众数一致。即预报期望值恰好是最可能的出现值。如果预报量遵从的是偏态分布,预报

得到的期望值不能对应现象出现的是较大概率值。而且气候预测的因子大多是大气环流因子,例如高度、温度和气压等,它们一般均可看成为正态分布。因此为了与大多数的正态分布因子有较好的配合,把降水量变换为正态分布变量也是十分必要的。

另外,在汛期降水量的短期气候预测中,有时要使用几百年的资料数据,例如常用的中国 500 a 旱涝序列<sup>[2]</sup>。这一序列制作是为识别旱涝严重程度,从中国历史文献中由对旱涝现象的文字记载转换成数字,而近代的降水量观测值也转换为级别数字。划分时使级别遵从正态分布,一方面可以简化降水量的变化范围,另一方面也适应气候的旱涝现象的识别。传统上常把一定时期内(季或年)降水量的多年平均值作为划分旱涝标准分界线。把降水量划分为 5 个级别,即特旱、偏旱、正常、偏涝和特涝,并分别记为 5, 4, 3, 2 和 1。此外,有些计算相关时使用秩相关系数,其变量序列也作级别变换,只是把降水量变换的级别数与样本容量相等,此种序列称为秩

\* 初稿时间:2001 年 2 月 21 日;修改稿时间:2001 年 6 月 20 日。

资助课题:“国学重点基础研究发展规划项目”我国重大气候和天气灾害形成机理和预测理论的研究(G199804901)及 908 短期气候预测课题。

序列。秩序列变换则是按变量遵从均匀分布来划分级别的。这些级别变换后的序列正态性如何也是需要研究的。

如何把季节降水量转换为正态变量,目前比较简单的方法有两种。一种方法是把降水量划分不同的级别。把级别作为预报对象,它可以按人们意志任意改变序列数据的分布结构使其符合某种分布。划分时的原则就是使级别遵从正态分布<sup>[2]</sup>。另一种方法对降水量做非线性变换,例如平方根或立方根变换。现在世界上很多国家的气候公报中描述降水量分布使用的是 $\Gamma$ 分布,可以把降水量变换为 $\Gamma$ 的概率值。

三峡地区的建设是中国西部开发的重要组成部分,特别是三峡水库建设是中国国民经济建设的关键工程。因此,研究该地区的气候,特别是降水量的气候变化特点对该地区的短期气候预测无疑也有十分重要意义。本文将对中国三峡地区汛期降水量的正态性进行分析及对上述各种降水量正态性变换的优劣性等问题进行试验性研究,研究这些变换对降水量的正态性有什么影响,对降水量的原来序列结构有什么影响。

## 2 资料与方法

试验对象取长江上游三峡地区夏季汛期(6~8月)降水量作研究对象。选取该流域地区16个站夏季汛期(6~8月)降水量46 a(1952~1997年)资料组成预报对象场,对它们的正态性和序列结构进行研究,站点分别是宜昌、遵义、贵阳、毕节、兴仁、榕江、恩施、达县、酉阳、重庆、南充、内江、绵阳、成都、宜宾和雅安。

为了对16站汛期降水量作是否遵从正态分布的检验,根据参数检验要求<sup>[3]</sup>,在样本容量固定为46和显著水平5%下,只需比较各站降水量的偏度和峰度系数的绝对值是否分别大于0.664和0.178,若大于判别值,表示该站降水量不遵从正态分布,否则可以认为遵从正态分布。为了比较各站序列的结构,而序列的周期震动特征是描述序列变化的主要方面,因此试验还对序列做功率谱分析。序列的功率谱计算使用序列自相关计算的间接方法<sup>[3]</sup>,最大落后步长取15。降水量的变换尽量保留降水量的变化特征,把降水量划级别时,以多级别来反映降水量的变化性。因此把降水量按前述方法划分为5个级别,划分时不仅考虑降水量的平均值还考虑其

标准差;把降水量转化为标准化值后,5个级别划分的分界值分别为-1.17, -0.33, 0.33, 1.17<sup>[4]</sup>。在分级时有的也用均匀分布划级别,级别数最多与序列样本容量相同,称为秩序列。

由于较好的描述降水量的分布是 $\Gamma$ 分布,用降水量资料样本和最大或然法可得概率密度函数中的两参数的估计值。利用最大或然法通过实测降水量样本可以估计出这一概率密度函数的参数值,然后用数值积分方法计算事件‘降水量小于某一个降水量具体值’的概率估计值,从而完成降水量的 $\Gamma$ 分布概率变换<sup>[5]</sup>。

## 3 单站降水量序列正态性分析

首先对三峡地区各站降水量求偏度系数、峰度系数和主要周期,然后对各站降水量做距平百分率变换,再做试验。结果发现对各站降水量做距平百分率变换,其偏度系数和峰度系数的主要周期相同。说明距平百分率不改变降水量的原始数据结构。表1给出各站降水量距平百分率的偏度和峰度系数的值,以及各站主要周期。

表1 16站降水量的偏度和峰度系数的值和主要周期

站名	主要周期(a)	偏度系数	峰度系数
宜昌	30.0	0.14	-0.69
遵义	30.0	0.18	-0.17
贵阳	30.0	0.46	0.51
毕节	30.0	1.35	3.35
兴仁	100.0	0.54	0.22
榕江	30.0	0.89	0.51
恩施	30.0	0.43	-0.04
达县	30.0	1.22	1.45
酉阳	30.0	0.20	-0.34
重庆	30.0	1.06	1.66
南充	30.0	0.66	-0.35
内江	30.0	0.85	-0.05
绵阳	30.0	1.09	1.29
成都	30.0	1.04	1.15
宜宾	30.0	0.41	0.05
雅安	30.0	0.32	-0.39
平均		0.68	0.76

从表中可见,偏度系数检验中只有宜昌、遵义、贵阳、兴仁、恩施、酉阳、南充、宜宾和雅安等9站通过检验,但是,它们均为正偏度。峰度系数仅3站通过。平均来说,偏度和峰度系数均未能通过检验。谱分析发现绝大部分均表现为周期30 a的变化,其中有1站周期为无穷大(即为近似直线的很长的周

期变化趋势,表中用 100 a 表示)。

进一步计算各站降水量做平方根和立方根变换后的偏度和峰度系数及主要周期。序列变化主要周期仍为长周期,直线趋势的站数有所增加(平方根增为 3 站,立方根增为 4 站)。偏度和峰度系数比原降水量均减小(见表 2 和表 3)。

表 2 16 站降水量平方根变换的偏度和峰度系数的值及主要周期

站名	主要周期(a)	偏度系数	峰度系数
宜昌	30.0	-0.18	-0.58
遵义	30.0	-0.26	0.21
贵阳	30.0	-0.03	0.21
毕节	30.0	0.87	1.53
兴仁	100.0	0.17	0.03
榕江	30.0	0.45	0.10
恩施	100.0	0.00	-0.14
达县	30.0	0.78	0.47
酉阳	30.0	-0.13	-0.48
重庆	30.0	0.55	0.70
南充	30.0	0.36	-0.62
内江	100.0	0.58	-0.36
绵阳	30.0	0.61	0.19
成都	30.0	0.68	0.24
宜宾	30.0	0.06	-0.15
雅安	30.0	0.05	-0.57
平均		0.36	0.41

表 3 16 站降水量立方根变换的偏度和峰度系数的值及主要周期

站名	主要周期(a)	偏度系数	峰度系数
宜昌	30.0	-0.30	-0.49
遵义	100.0	-0.43	0.49
贵阳	30.0	-0.19	0.25
毕节	30.0	0.73	1.08
兴仁	100.0	0.04	0.06
榕江	30.0	0.29	0.07
恩施	100.0	-0.16	-0.05
达县	30.0	0.63	0.24
酉阳	30.0	-0.24	-0.47
重庆	30.0	0.39	0.49
南充	30.0	0.26	-0.68
内江	100.0	0.48	-0.43
绵阳	30.0	0.46	-0.04
成都	30.0	0.56	0.01
宜宾	30.0	-0.06	-0.15
雅安	30.0	-0.04	-0.59
平均		0.33	0.35

在分级时用正态分布划分级别,按常用 5 个级别,对 16 站的级别序列进行研究。发现各站均表现 30 a 以上的主要周期,其中直线趋势占 7 站,偏度系

数均通过检验,平均峰度系数也较降水量非线性变换要小,满足正态性的要求。

对三峡 16 站按均匀分布划分级别,即分别求各站降水量的秩序列。发现各站均表现 30 a 的主要周期,偏度系数均为 0,峰度系数均为 -1.2,全部满足正态性的要求。

对降水量做  $\Gamma$  分布转换成概率值,然后对各站概率值做分析。对 16 站的概率值序列进行研究(见表 4)。发现各站仍然表现为 30 a 的主要周期。偏度系数均通过检验,是几种变量正态性变换中平均值最小的。但是平均峰度系数却较大。

表 4 16 站降水量  $\Gamma$  分布概率值的偏度和峰度系数的值及主要周期

站名	主要周期(a)	偏度系数	峰度系数
宜昌	30.0	-0.12	-1.09
遵义	30.0	-0.06	-1.01
贵阳	30.0	-0.04	-1.11
毕节	30.0	0.07	-1.33
兴仁	30.0	0.10	-1.13
榕江	30.0	0.23	-0.96
恩施	30.0	-0.03	-1.04
达县	30.0	0.26	-0.95
酉阳	30.0	-0.20	-1.13
重庆	30.0	0.05	-0.94
南充	30.0	0.11	-1.12
内江	30.0	0.31	-1.04
绵阳	30.0	0.19	-1.08
成都	30.0	0.14	-1.21
宜宾	30.0	-0.02	-1.03
雅安	30.0	-0.07	-1.19
平均		0.12	-1.08

单站降水量的各种变换试验表明,变换不改变序列原始谱结构,仅影响概率分布的偏度和峰度,使其较好地遵从正态分布,其中以  $\Gamma$  分布的变换为最好。

#### 4 区域降水量的综合指标及其正态性分析

对于考察区域降水量综合指标的正态性,通常使用区域所有站降水量的平均序列和对区域所有站降水量做主分量分析。计算发现区域降水量平均序列的偏度系数为 -0.16,可以看成正态分布。但是峰度系数为 2.61,则偏离标准正态分布较多。功率谱分析发现,序列变化主要周期仍为 30 a 变化的长周期,说明平均的综合作用未能改变序列周期变化特征。

对三峡地区 16 站降水量作主分量分析,因为前

8 个主分量的累积解释方差已达 83%, 所以取前 8 个主分量进行分析。表 5 给出各主分量的累积解释方差, 偏度和峰度系数的值及主要周期。

表 5 区域降水量主分量的偏度和峰度系数的值及主要周期

主分量	累积方差	主要周期(a)	偏度系数	峰度系数
1	0.27	2.1	-0.16	-0.13
2	0.42	30.0	0.63	0.49
3	0.55	7.5	0.60	0.05
4	0.63	2.3	-0.23	0.16
5	0.69	2.1	-0.24	-0.27
6	0.74	3.8	0.19	-0.36
7	0.79	7.5	0.10	-0.28
8	0.83	7.5	0.37	-0.03
平均值			0.32	0.22

从表中可见, 所有主分量的偏度系数均通过检

验, 遵从正态分布。可见由该地区各站降水量的线性组合也是一种变换, 这种变换也能改善降水量的正态性。功率谱分析发现, 序列变化主要周期为准 2 a 的有 3 个主分量(占 38%), 10 a 以上周期的仅 1 个主分量(占 13%), 3~8 a 周期仍是大多数, 有 4 个(占 43%), 可见提取主分量 3~8 a 的类似 ENSO 周期表现也很明显。说明主分量的综合作用能改变序列谱结构特征。

进一步对三峡地区 16 站降水量做与上述类似的各种变换, 然后再做主分量分析, 仍对前 8 个主分量进行分析, 发现变换后主分量收敛性略有降低, 前 8 个主分量解释方差为 82% 左右。图 1 给出了各种变换后各主分量的主要周期比较。

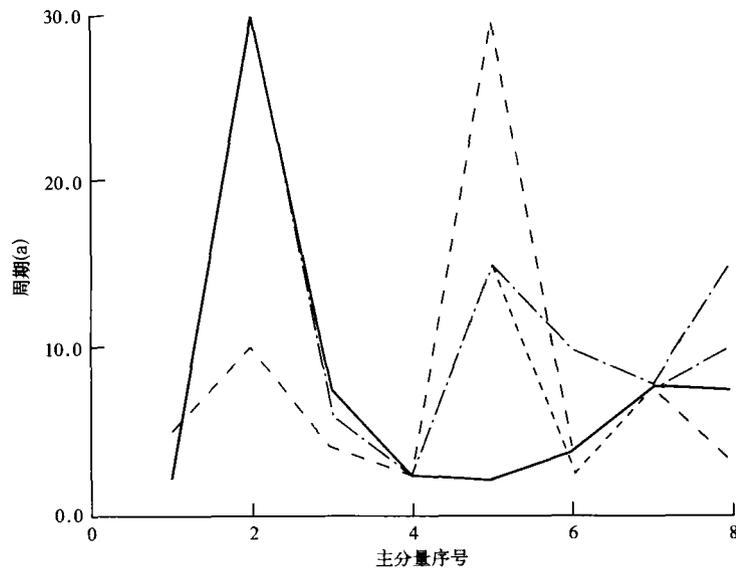


图 1 降水量不同变换各主分量的主要周期比较

(实线为原降水量, 长虚线为立方根变换,

短虚线为正态分级变换, 两点一划线表示秩变换)

从图可见, 主要周期表现在降水量的非线性变换的前 4 个主分量中变化不大, 只是在后 4 个主分量上有些变化。级别变换则有较大的变化, 表现在正态级别在第 2 主分量的主要周期由 30 a 减为 10 a, 而第 5 主分量的主要周期由短周期变成 30 a 的长周期。但是, 秩序列变换在前几个主分量上, 又与降水量及其非线性变换差不多(见表 6, 7 和 8)。

几种非线性变换的区域降水量主分量偏度和峰度系数的比较, 仅给出平方根变换、立方根和秩变换的情况(表 6, 7 和 8)。

表 6 区域降水量平方根变换后各主分量的偏度和峰度系数的值及主要周期

主分量	累积方差	主要周期(a)	偏度系数	峰度系数
1	0.27	2.1	-0.39	-0.17
2	0.42	30.0	0.30	-0.09
3	0.55	7.5	0.26	-0.65
4	0.63	2.3	-0.32	0.56
5	0.69	2.1	-0.32	-0.10
6	0.74	3.8	0.15	-0.31
7	0.78	7.5	-0.15	-0.59
8	0.82	10.0	-0.12	1.30
平均值			0.25	0.47

表7 区域降水量立方根变换后各主分量的偏度和峰度系数的值及主要周期

主分量	累积方差	主要周期(a)	偏度系数	峰度系数
1	0.28	2.1	-0.46	-0.13
2	0.42	30.0	0.21	-0.20
3	0.55	7.5	0.18	-0.78
4	0.62	2.3	-0.32	0.73
5	0.69	15.0	-0.27	-0.09
6	0.74	2.5	0.13	-0.24
7	0.78	7.5	-0.20	-0.55
8	0.82	10.0	-0.08	1.86
平均值			0.23	0.57

表8 区域降水量秩序列主分量的偏度和峰度系数的值及主要周期

主分量	累积方差	主要周期(a)	偏度系数	峰度系数
1	0.27	2.1	-0.33	-0.70
2	0.41	30.0	-0.02	-0.81
3	0.53	6.0	0.09	-0.99
4	0.60	2.3	0.09	0.77
5	0.67	15.0	-0.07	-0.23
6	0.72	10.0	0.34	-0.05
7	0.77	7.5	-0.39	0.42
8	0.82	15.0	-0.25	1.42
平均值			0.20	0.67

由表6,7和8发现降水量非线性变换的主分量偏度平均比降水量原值的主分量下降较多,但是各主分量偏度系数的变化十分相似。而秩序列变换在前几个主分量上,又与降水量及其非线性变换差不多。在峰度系数比较上,降水量非线性变换的主分量偏度平均比降水量原值的主分量下降较多,各种变换除与原降水量的前2个主分量有差别外,各主分量偏度系数的变化十分相似。降水量的非线性变换与原降水量相比,峰度系数的值则有些增加。秩序列变换的峰度增加最大。

由于降水量做平方根或立方根变换,并不改变数据在序列中的秩,没有必要对降水量做平方根或立方根变换,再做主分量分析,其主分量的类似分析结果是相同的。对降水量序列变换为 $\Gamma$ 分布概率值序列,再做主分量分析,取前8个主分量进行分析。表9给出各主分量的累积解释方差、偏度和峰度系数及主要周期。发现其收敛速度与降水量非线性变换差不多。主要周期表现也与降水量立方根变换类似,但是主要周期中2~3a的短期震动表现加强。偏度系数较降水量变换和秩序列变换还要小些,但峰度系数也比秩序列变换减少,完全能满足正态性的要求。

区域降水量的各种变换的综合指数正态性比较

表9 区域降水量 $\Gamma$ 分布概率值序列主分量的偏度和峰度系数的值及主要周期

主分量	累积方差	主要周期(a)	偏度系数	峰度系数
1	0.26	5.0	-0.35	-0.74
2	0.41	30.0	0.07	-0.81
3	0.53	7.5	0.02	-0.94
4	0.61	2.3	-0.03	0.78
5	0.67	2.1	-0.12	-0.06
6	0.73	30.0	0.30	0.26
7	0.77	7.5	-0.05	-0.01
8	0.81	2.7	-0.32	1.48
平均值			0.16	0.63

表明,变换后能改变原单站序列的谱结构,同时也影响概率分布的偏度和峰度,使其较好地遵从正态分布,降水量数值的立方根和 $\Gamma$ 分布概率值变换以及级别变换有较好的效果。

## 5 结论与讨论

为了分析区域降水量序列的正态性和谱结构,对降水量的常见各种变换进行试验性研究,试验包括单站降水量、区域多站平均降水量、区域降水量的主分量、区域降水量的非线性变换后的主分量、单站降水量的秩变换和区域降水量的秩变换后的主分量等序列正态性和谱结构研究。从上述试验研究中可以得到如下结论:

(1) 三峡地区16站降水量大部分不遵从正态分布,序列变化主要周期表现为30a以上的长周期;降水量的距平百分率变换未能改变序列的结构和正态性。

(2) 对降水量作平方根、立方根和 $\Gamma$ 分布概率值等非线性变换后,可以看成遵从正态分布。大多数序列变换并不改变原序列的谱结构。区域各站降水量的级别变换,变换中无论是正态化或均匀化的变换,都可以改善降水量的正态性,并能改变序列的谱结构。

(3) 区域所有站降水量的平均序列能改善降水量的正态性,但是序列谱结构不变。

(4) 区域降水量的主分量分析能改善降水量的正态性。主分量序列变化主要周期为准2a和3~8a周期。

(5) 区域各站降水量作平方根、立方根和 $\Gamma$ 分布概率值等非线性变换,然后再做主分量分析。这种变换也更加改善降水量的正态性。主分量序列变化主要周期与变换无太大关系,仍表现为准2a和3~8a周期。

(6) 区域各站降水量的级别变换后再做主分量分析,可以更加改善降水量的正态性。

(7) 在上述几种降水量的变换中,效果较好的

是降水量数值的立方根和  $\Gamma$  分布概率值变换以及级别变换。从操作简便来说,以立方根变换和级别变换为宜。

### 参考文献

- 1 徐尔源. 论年雨量之常态性. 气象学报, 1950, 21, 17~34
- 2 中央气象局等. 中国近五百年旱涝分布图集. 北京: 地图出版社, 1980. 332pp
- 3 黄嘉佑. 气象统计分析与预报方法. 北京: 气象出版社, 1990. 387pp
- 4 黄嘉佑, 李黄. 气象中的谱分析. 北京: 气象出版社, 1984. 318pp
- 5 黄嘉佑. 统计动力分析与预报. 北京: 气象出版社, 1993. 243pp

## THE STUDY OF NORMALITY ON SUMMER PRECIPITATION IN SANXIA AREA OF CHINA

Huang Jiayou Huang Maoyi Zhang Yin Zhu Rei

(Department of Atmospheric Sciences, the College of Physics, Peking University, Beijing 100871)

### Abstract

In short-range climatic prediction, the transformations for summer precipitation usually have been made in single station or in stations in some area in China. There are linear transformations, such as anomaly, anomaly percent, and normalized for single station. Otherwise, the average and principal component analysis based on stations in the area has been usually made. Sometime the nonlinear transformations are also used on the precipitation series for improving the prediction, for example, square root, cubice root, or class transformation and so on.

In order to investigate spectral structure and normality on summer precipitation in the original and transformed series, the data of summer (June to August) precipitation in Sanxia in China during the period 1952-1997 are selected in this paper. The transformations of precipitation are completed by nonlinear transformation. Those are square root, cubic root, probability (by Gamma distribution), class and rank transformation. Otherwise, the average, principal comonent of the variables in the 16 stations in the area is also analyzed. The method of power spectral analysis is used for the series. The parameters of skewness and kurtosis in distribution are calculated for the variables.

The results of the study show that: (1) The summer precipitation in most stations in the area is not following normal distribution. The main period longer than 30 years is represented in those series. The transformation of anomaly percent cannot change the spectral structure; (2) The main period in the spectral structure is not changed in most stations after the nonlinear transformation of precipitation. The skewness and the kurtosis in the transformation series can be changed. They are better than the primitive series in following normal distribution; (3) The transformation series of average series of precipitation in the stations is better than the single station in following normal distribution; (4) The principal components of precipitation in the stations in the area is different from the single station in the spectral structure. The main period is about 2-3 years in the first principal component. The skewness in the principal components is smaller than the single station; (5) The transformation of cubic root, probability of Gamma distribution, and rank for precipitation are better than others in those nonlinear transformations for normality, specially in skewness of distribution; (6) The cubice root and class transformation for precipitation are better in the operation.

**Key words:** Normality, Nonlinear transformation, Principal components of precipitation.