

甘肃降水量正态性分析

陈学君¹, 胡文超², 杨瑜峰¹, 施涛¹, 唐国利³, 白冰¹

(1. 甘肃省气象信息与技术装备保障中心, 甘肃 兰州 730020; 2. 甘肃省气象服务中心, 甘肃 兰州 730020; 3. 中国气象局国家气象信息中心, 北京 100081)

摘要:利用1951~2010年甘肃省81个气象站月、年降水量资料, 采用Jarque-Bera检验与 χ^2 检验对原降水序列和经过Box-Cox变换后序列的正态性进行分析。结果表明: 甘肃省大多数气象站年降水量服从正态分布, 但一半以上台站的月降水不服从正态分布。其中12月、11月、1月的正态性最差, 而5月和9月最好; 从季节来看, 甘肃大多数气象站春、夏季正态性较好, 秋、冬季正态性较差; 从地域分布来看, 全省各地降水均存在非正态, 其中河西地区、陇中和陇东部分站点非正态现象尤其突出。经过数据变换后的序列, 站年正态性通过率从原来的81.5%~82.7%升高到97.5%~100%, 站月通过率从36.2%~68.5%升高到85.4%~94.0%, 大部分站点通过了正态性检验。

关键词:甘肃省; 降水量; Box-Cox变换; 正态性

文章编号:1006-7639(2013)-01-0107-06 doi:10.11755/j.issn.1006-7639(2013)-01-0107

中图分类号:P468.0+24

文献标识码:A

引言

甘肃省地处黄土高原、青藏高原和蒙古高原交汇地带, 地形复杂, 气候差异大。河西为干旱气候, 河东为季风气候, 甘南高原和祁连山为高寒气候。东南部温湿, 西北部冷干, 由东南到西北几乎包括了北亚热带到高原寒带的各种气候带和湿润、半湿润、半干旱、干旱、极干旱气候区。同时, 甘肃省气象灾害的种类繁多, 灾情也比较严重, 其中干旱是甘肃省最主要的气象灾害。因此降水研究对该区域农业生产和经济发展至关重要。作为气候研究的重要方面, 降水一直受到众多学者的关注, 成为气候研究的热点。在降水序列的非均一性检验^[1]、气候变化研究、统计分析和预测^[2]中, 常常要求降水量时间序列符合正态分布。然而事实上降水量一般并不能很好地服从正态分布, 因此对降水量进行正态性研究具有重要的理论价值和实际意义。

21世纪前, 降水量的正态性研究并未引起足够的重视, 只有少数的学者杨观竹^[3]、陈雅芬^[4]对陕西、江西部分站点的降水量正态分布进行了初步研

究。进入21世纪, 在全球气候变化的大背景下, 降水正态分布特征逐渐引起学者的重视。曹杰等^[5]对全国160个测站的月降水资料是否符合正态分布进行了检验分析, 指出中国夏季大多数测站降水量具有正态分布特征, 而冬季降水则多不符合正态分布; 黄河流域以南和江淮流域东北部分地区及云南部分地区基本符合正态分布。陶云^[6]、吴慧^[7]、胡文东^[8]、方建刚^[9]、王纪军^[10]、方媛^[11]、周雁翎^[12]、赵佩红^[13]分别对云南、海南、宁夏、陕西、河南、银川、粤西、新会等地就不同区域降水正态分布进行了相关研究, 陈学君^[14]以甘肃省部分站点为例对降水量数据的正态变换方法作了对比分析, 韩晖^[15]、张小明^[16]、刘卫民^[17]分别对甘肃省降水时空分布和气候特征进行了分析。

本文利用甘肃省81个气象站点从建站起至2010年的月、年降水观测资料, 对其降水的正态性进行检验与分析, 在此基础上使用Box-Cox数据变换并对其变换效果进行比较, 以期对降水统计分析预测、均一性研究及气候和气候变化研究提供基础。

收稿日期: 2012-11-06; 改回日期: 2012-12-26

基金项目: 国家公益性行业(气象)科研专项项目“中国近60年地面关键气候要素均一性检验与订正技术及站址变动影响研究”(GYHY201206013)资助

作者简介: 陈学君(1973-), 男, 甘肃天水人, 博士, 高级工程师, 主要从事数据分析与处理、气象数据质量控制工作。E-mail: xue-junchen1971@163.com

1 数据和方法

1.1 资料选取

利用甘肃省气象局提供经过质量控制(气候界限值检查、区域极值检查、时间一致性检查、空间一致性检查、内部一致性检查)的 1951~2010 年甘肃省 82 个气象站的月、年降水观测资料,降水量资料

的起始年份见表 1。这些台站空间分布相对均匀,包括部分国家基准、基本气象站,但多数为国家一般气象站。个别站点有缺测数据,但不影响数据分布的分析。由于兰山站从 2004 年开始观测,降水观测数据少,故而不在于统计计算中。实际统计站点数为 81 个站。

表 1 甘肃省各气象站年降水量资料的起始年份

Tab. 1 The start year for the annual precipitation data of each meteorological station in Gansu

站名	马宗山	敦煌	安西	玉门	鼎新	金塔	肃北	酒泉	高台	临泽	肃南	张掖	民乐	山丹
起始年份	1957	1951	1951	1952	1955	1957	1961	1951	1952	1957	1956	1951	1958	1952
站名	乌鞘岭	金昌	武威	民勤	古浪	永昌	景泰	皋兰	永登	兰州	兰山	靖远	白银	夏河
起始年份	1951	1991	1951	1953	1959	1958	1956	1957	1956	1951	2004	1951	1954	1952
站名	华家岭	东乡	广河	永靖	榆中	临夏	和政	临洮	康乐	会宁	安定	渭源	环县	庆城
起始年份	1951	1959	1967	1967	1951	1951	1959	1951	1967	1955	1957	1956	1957	1959
站名	静宁	通渭	崆峒	庄浪	西峰	灵台	镇原	泾川	华亭	崇信	华池	合水	正宁	宁县
起始年份	1957	1956	1951	1959	1951	1959	1956	1959	1956	1965	1965	1959	1959	1957
站名	张家川	碌曲	玛曲	甘南	临潭	卓尼	迭部	漳县	陇西	岷县	舟曲	宕昌	武都	文县
起始年份	1956	1973	1967	1957	1956	1976	1973	1959	1956	1951	1972	1956	1951	1959
站名	麦积区	甘谷	秦安	武山	天水	礼县	西和	清水	成县	康县	徽县	两当		
起始年份	1964	1955	1959	1959	1951	1956	1965	1959	1960	1958	1956	1970		

1.2 正态性检验

正态性检验方法的原假设为 H_0 : 数据服从正态分布; 相应的备择假设为 H_1 : 数据不服从正态分布, 在这种意义下, 这类检验有时也称非正态性检验 (non-normality test)。正态性检验方法有许多种^[18-19]。由于应用目的不同, 对正态性具体要求也不同, 这里主要采用气象上常用的基于偏度峰度联合检验的 Jarque - Bera 检验和 χ^2 检验 2 种方法对数据进行正态性检验并提供相应结论。在偏度峰度联合检验方法中, 为了消除样本容量引起的偏度和峰度偏差, 利用标准偏态系数和标准峰度系数来说明数据的正态性, 具体定义参见文献^[10]。2 种正态性检验方法设置信度均为 $\alpha = 0.05$, 若其 $P < 0.05$, 则否定原假设, 断定总体呈非正态分布。

1.3 Box - Cox 正态变换

Box - Cox 变换是 1964 年 Box 和 Cox 提出的在有效保留原有数据信息的基础上使数据服从正态分布(或者准正态化)的方法, 该方法变换简单且属于可使线性回归满足良好性质又不丢失信息的变换。其中包含对数变换($\lambda = 0$)、平方根变换($\lambda = 1/2$)和倒数变换($\lambda = -1$)等常用变换。针对具体数据

可选择最优的幂参数, 属于幂变换族^[20], 变换公式为:

$$y^{(\lambda)} = \begin{cases} \frac{y^{(\lambda)} - 1}{\lambda} & \lambda \neq 0 \\ \ln y & \lambda = 0 \end{cases} \quad \dots\dots\dots (1)$$

其中, y 为原始数据, $y^{(\lambda)}$ 为变换后数据, 变换要求 $y > 0$, 但当此条件不满足时, 只要作一个整体平移即可; 式中 λ 为待定参数, 可按极大似然估计得到。

1.4 数据处理过程

下面以华家岭站 1951~2010 年 2 月降水资料为例简要说明其处理过程, 详细说明参见文献[14], 其他资料的处理过程类似。首先对降水资料进行正态性检验, 由 Jarque - Bera 正态检验可知: 其检验 P 值 < 0.05 ($p = 0.0050$), 且偏度为 1.1571、峰度 4.2690; 而 χ^2 检验也可得到其检验 P 值 < 0.05 , 说明数据总体不符合正态分布, 表现出一定程度的正偏(图 1)。利用 Box - Cox 数据变换方法, 为寻找合适的参数 λ , 先给出一系列的 λ 值, 最优 λ 值可按极大似然估计得到。对于本例中当 λ 为 0.41 时, 其

标准差最小,故在此降水数据 Box - Cox 变换中,参数 λ 取值为 0.41。对数据变换后的序列经过 χ^2 检验与 Jarque - Bera 检验,其 P 值均 >0.05 ,偏度值与峰度值有明显降低,具体偏度为 -0.0440 、峰度为 2.8432,说明数据变换后的序列符合正态分布(图 1)。

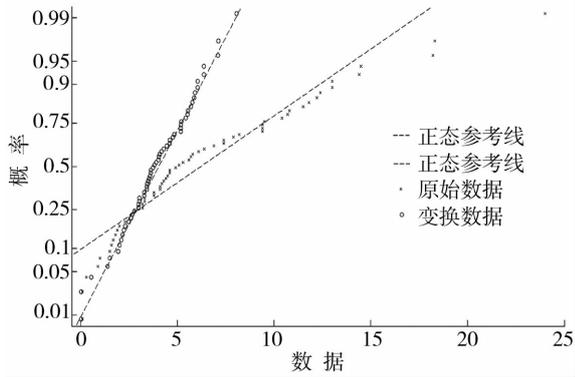


图 1 华家岭站 1951 ~ 2010 年 2 月降水资料正态概率图

Fig. 1 Normal probability figure of precipitation for Huajialing station in February during 1951 - 2010

2 结果分析

2.1 年降水量

2.1.1 正态性检验

甘肃省 81 个气象站点中原年降水序列有 15 个未通过 Jarque - Bera 检验(图 2),占总站点的 18.5%,有 14 个未通过 χ^2 检验,占总站点的 17.3%,在 JB 检验中分析表明,年降水数据以正偏为主,主要由于偏度不满足条件而未能通过正态性检验。

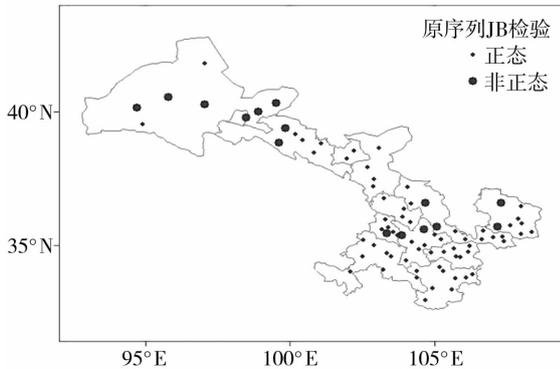


图 2 甘肃省 81 个台站 1951 ~ 2010 年降水量正态性检验分布图

Fig. 2 Distribution of normal distribution tests of the annual precipitation at 81 stations in Gansu Province during 1951 - 2010

同时,由图 1 可以看出,甘肃大多数气象站的年降水量服从正态分布。对于 JB 检验未通过站点的地域分析表明,河西西部有 8 站未能通过正态性检验(敦煌、安西、酒泉、鼎新、玉门、肃南、金塔、高台),陇中地区有 5 站(定西、会宁、和政、临洮、靖远),陇东地区有 2 站(环县、镇原)未通过检验,临夏仅有 1 站(和政)。对于 χ^2 检验未通过站点的地域分析结果与 JB 检验类似,只不过站点在上述地区的分布个别有差异。

经过数据变换后的序列,所有站点均通过 JB 检验,只有 2 站(永登、舟曲)未通过 χ^2 检验,站年正态性通过率从原来 81.5% ~ 82.7% 的升高到 97.5% ~ 100%,说明数据变换对于甘肃的年降水序列正态性改善是非常有效的。

2.1.2 偏度与峰度特征分析

甘肃省年降水量的偏态系数原始序列平均为 1.40,Box - Cox 数据变换后偏度系数明显减小为 0.05。站间标准差变化也比较明显,由 0.93 变为 0.06,偏度系数超过 2 的区域主要集中在酒泉市、陇中地区,陇东和张掖部分地区也有所表现。年降水量以正偏为主,中心强度超过 3 的站点有 6 个,分别为安西、敦煌、玉门、会宁、安定、高台,正偏中心为安西站,中心强度为 3.98。统计站点中仅有 6 个站为负偏,约占总数的 7.41%,且其绝对值均在 0.7 以下。甘肃省年降水量经过 Box - Cox 数据变换的偏度系数正偏与负偏相当,其中负偏站点有 39 个,81 个站点偏度系数的绝对值均在 0.4 以下。

甘肃省年降水量的原始序列的峰度系数平均值为 0.93,Box - Cox 数据变换后峰度系数减小为 0.83。站间标准差减小 1 倍左右,由 0.96 变为 0.52。甘肃省年降水量原始序列的峰度系数大多为负值,正值有 30 个站点,占总站点数的 37.0%。峰度系数绝对值超过 2 区域的与偏度系数超过 2 的区域类似,站点共有 7 个,约占总数的 8.64%,最大值为安西(5.19),其余 6 个站点为敦煌、玉门、会宁、永登、安定、文县。经 Box - Cox 数据变换后,峰度系数大多仍为负值,正值有 19 个站点,占 23.50%。峰度系数绝对值超过 2 的站点仅有 1 个,为永登(2.08),其余站点均在 2.0 以下。

由于甘肃省气候差异大,大部分处于半干旱及干旱气候区,导致年降水分布呈现非正态,但站年降水为年度级的累积,存在统计效应,相对消除了些站点降水分布的非正态现象,经过 Box - Cox 数据变换后则可基本达到正态。

2.2 月降水量

2.2.1 正态性检验

甘肃省各站逐月降水量通过正态性检验(Jarque - Bera 检验与 χ^2 检验)的站数随月份变化的关系如图3所示。

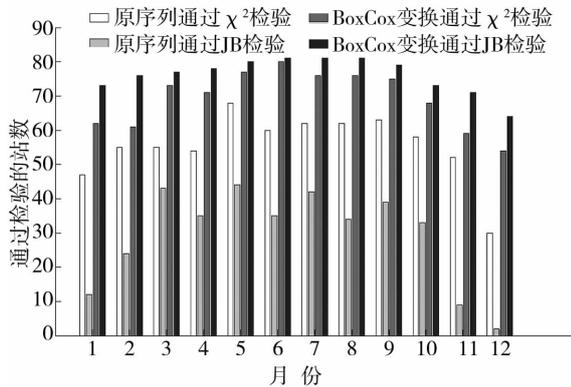


图3 甘肃省81个台站1951~2010年各月通过正态性检验的站数

Fig. 3 Numbers of weather stations passed the normal distribution test of monthly precipitation at 81 stations in Gansu Province during 1951 - 2010

由图3可以看出,对原降水序列 JB 检验表明,站月正态性通过率为36.2%,其中5月最好,有44个站通过检验,3月次之有43站通过,12月和11月最差,分别仅2站、9站通过。经过数据变换后的序列,同样用 JB 检验表明,站月正态性通过率为94.0%,其中6、7、8月81站全部通过检验,12月和11月最差,但也分别有64站、71站通过。对于原降水序列而言, χ^2 检验的结果表明,站月正态性通过率为68.5%,其中5月最好,有68个站通过检验,9月次之有63站通过,12月和1月最差,分别仅30站、47站通过。经过数据变换后的序列, χ^2 检验的结果为站月正态性通过率为85.6%,其中6月最好,有80个站通过检验,5月次之有77站通过,12月和11月最差,但也有54站、59站通过。

从地域的分析来看,全省各站点原降水序列均在不同月份存在非正态问题。对原降水序列 JB 检验表明超过6个月(含6个月)未能通过正态性检验的站点共有61个,占总站点的75.3%,在全省各地均有分布。经过数据变换后的序列超过6个月(含6个月)未能通过 JB 检验站点仅有3个站,占总站数的0.04%,分别为安西、敦煌、鼎新。 χ^2 检验的结果表明:超过6个月(含6个月)未能通过正态性检验的站点共有15个,占总站点的18.5%,主要分布在河西地区(特别是河西西部),在陇东和陇中

个别站点也超过6个月。经过数据变换后的序列超过6个月(含6个月)未能通过 χ^2 检验的站点数大大降低,只有3个,占总站数的0.04%,仍为安西、敦煌、鼎新。

分别按春、夏、秋、冬4季统计甘肃各站逐月降水正态性检验(Jarque - Bera 检验与 χ^2 检验),站数随季节变化的关系如图4所示。对原降水序列 JB 检验表明,春季、夏季正态性通过率较好,分别有122、111站通过检验,通过率分别达到50.2%、45.7%,冬季最差,仅有38站通过,通过率为15.6%。经过数据变换后的序列,同样用 JB 检验,春季、夏季正态性通过率提高到235、243站,通过率分别为96.7%和100%,冬季同样最差,但也有213站通过,通过率为87.7%。对于原降水序列而言, χ^2 检验的结果表明,夏季正态性通过率最好为75.7%,有184个站通过检验,春季次之,有177站通过,通过率为72.8%,冬季最差,有132站通过,通过率为54.3%。经过数据变换后的序列, χ^2 检验的结果夏季正态性通过率最好,为95.5%,有232个站通过检验,春季次之,有221站通过,通过率为90.9%,冬季最差,有177站通过,通过率为72.8%。

在四季中,从降水通过正态性检验的站数来看,原序列和变换后序列均表明,春、夏季正态性较好,秋、冬季正态性较差,变换后序列的正态性得到明显提高。

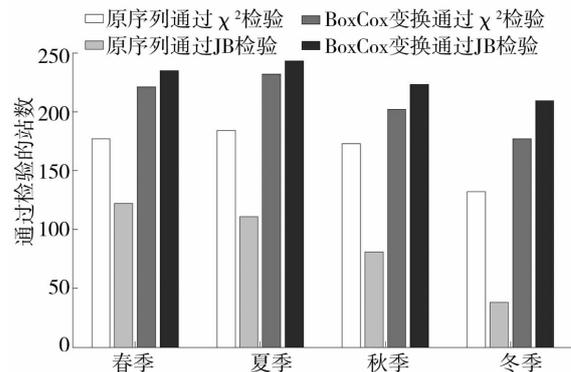


图4 甘肃省81个台站1951~2010年各季(逐月累积)通过正态性检验的站数

Fig. 4 Numbers of weather stations passed the normal distribution test of seasonal precipitation at 81 stations in Gansu Province during 1951 - 2010

总体而言,甘肃省全省气象站点月降水大部分不服从正态分布。从月份上讲,12月、11月、1月的站点正态性是在一年之中最差的,而5月和9月份是最好的;从季节看,春、夏季正态性较好,秋、冬季正态性较差;从地域分布来看,全省各地降水均存在

非正态,尤其是河西地区(特别是河西西部)、陇中和陇东部分站点存在严重的非正态现象,而经过Box-Cox数据变换后的序列,无论从时间还是从地域角度来看,站点总体正态性得到了很大提高,站月通过率从36.2%~68.5%升高到85.4%~94.0%,大部分站点通过了正态性检验。

2.2.2 偏度与峰度特征分析

甘肃省位于东经 $92^{\circ}13'$ ~ $108^{\circ}46'$ 和北纬 $32^{\circ}31'$ ~ $42^{\circ}57'$ 间,深居内陆,地形复杂,气候差异大,省域大部分处于我国半干旱及干旱气候区,全年降水少。由于站月降水分布差异较年降水更明显,故而导致站月降水呈现严重非正态现象。特别是甘肃河西地区降水量相对偏少,气候干旱,如敦煌站(见图5,敦煌站1951~2010年1~12月降水量箱线图)。以7月份为例,7月敦煌降雨量的中位数为6.85 mm,平均值为11.4 mm,但7月份降水量出现了3次较大值,分别为60.7 mm、45.4 mm和38.2 mm,这3次降水与中位数、平均值相比出现了很大的偏差,故而使其严重偏离正态。同样,对河西其他站点分析,情况类似,由量级分布情况分析,小量降水占大多数,但在个别情况下,某些时段内也会产生很大的降水。由于大降水出现的几率小,造成了小量级降水出现频率极高,中等量级降水频率较小,个别等级降水频率甚至为0,大量级降水频率极低的现象。在偏态系数上表现为向低值的强烈偏向情况,从而形成降水量的严重非正态分布。通过计算发现,未通过基于偏度与峰度的JB正态性检验者,主要由于两者系数均未达到阈值标准所致。而陇中和陇东个别地区由于偶尔会出现暴雨现象导致降水量偏大,降水分布不均匀从而产生非正态。

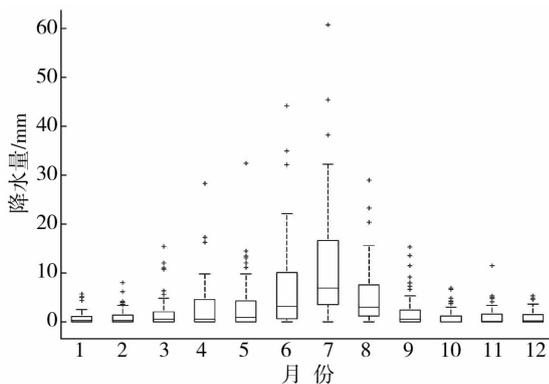


图5 敦煌站1951~2010年1~12月降水量箱线图
Fig. 5 Box plot of monthly precipitation at Donghuang station in Gansu Province during 1951-2010

3 结论

(1)甘肃大多数气象站的年降水量服从正态分布。少数不服从正态分布的站点主要位于河西西部、陇南、天水、陇中的个别地区。

(2)对于甘肃各站月降水序列而言,各站月降水一半多不服从正态分布。逐月检验表明,12月、11月、1月的站点正态性是在一年之中最差的,而5月和9月份是最好的。从地域分布来看,全省各地降水均存在非正态,尤其是河西地区(特别是河西西部)、陇中和陇东部分站点存在严重的非正态现象;从季节来看,甘肃大多数气象站春、夏季正态性较好,秋、冬季正态性较差。

(3)经过数据变换后的序列,站年正态性通过率从原来81.5%~82.7%的升高到97.5%~100%,站月通过率从36.2%~68.5%升高到85.4%~94.0%,大部分站点通过了正态性检验。

(4)未能通过正态性检验的降水样本,主要是由于降水量偏少(干旱原因),偏度系数存在强烈的向低值偏向的情况。经过数据变换后仍未能通过正态性检验,主要由于两者系数均未达到阈值标准所致。

参考文献:

- [1] Wang X L W Q H W. Penalized Maximal t Test for Detecting Undocumented Mean Change in Climate Data Series[J]. Journal of Applied Meteorology and Climatology, 2007,46(6):916-931.
- [2] 岳文泽,徐建华,徐丽华.基于地统计方法的气候要素空间插值研究[J].高原气象,2005(6):974-980.
- [3] 杨观竹.陕西省年、月降水量的理论频数分配[J].1983(2):36-41.
- [4] 陈雅芬,刘福基,张超美.江西省主要气象要素的正态分布检验[J].江西气象科技,1994(4):34-35.
- [5] 曹杰,陶云.中国的降水量符合正态分布吗?[J].自然灾害学报,2002(3):115-120.
- [6] 陶云,段旭.云南降水正态分布特征的初探[J].气象科学,2003(2):161-167.
- [7] 吴慧.海南省降水量的正态分布特征及正态化变换[J].广东气象,2005(2):12-13.
- [8] 胡文东,陈晓光,李艳春,等.宁夏月、季、年降水量正态性分析[J].中国沙漠,2006,26(6):963-968.
- [9] 方建刚,毛明策,程肖侠.陕西降水的正态分布特征分析[J].西北大学学报(自然科学版),2009(1):131-136.
- [10] 王纪军,任国玉,匡晓燕,等.河南省月和年降水量正态性分析[J].气候与环境研究,2010(4):522-528.
- [11] 方媛,李培月.银川市降水量正态分布特征[J].水利科技与经济,2010,16(8):873-875.
- [12] 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[A]:2007年中国气象学会年会论文集[C].广州,2007,204-208.

- [13] 赵佩红. 新会气候资料的正态分布检验及正态化处理[J]. 广东气象, 2007, 29(1): 27-28.
- [14] 陈学君, 苏仲岳, 李仲龙, 等. 年降水量数据的正态变换方法对比分析[J]. 干旱气象, 2012, 30(3): 459-464.
- [15] 韩晖, 孟丽霞. 甘肃省近 46a 年降水结构空间变化分析[J]. 干旱气象, 2010, 28(1): 30-34.
- [16] 张小明, 杨金虎, 高伟东, 等. 甘肃省近 50a 夏季极端强降水量的气候特征[J]. 干旱气象, 2008, 26(2): 48-52.
- [17] 刘卫民, 王陇, 贾海源. 甘肃省陇东南地区大到暴雨时空分布特征[J]. 干旱气象, 2006, 24(1): 44-47.
- [18] 梁小筠. 正态性检验[M]. 北京: 中国统计出版社, 1997.
- [19] 施能. 气象统计预报[M]. 北京: 气象出版社, 2009.
- [20] BOX G C D. An analysis of transformations[J]. Royal Statistical Society. Series B (Methodological), 1964, 26(2): 211-252.

Normality Analysis of Precipitation in Gansu Province

CHEN Xuejun¹, HU Wenchao², YANG Yufeng¹, SHI Tao¹, TANG Guoli³, BAI Bing¹

- (1. *Gansu Meteorological Information & Technique Support & Equipment Centre, Lanzhou 730020, China;*
 2. *Meteorological Service Centre of Gansu Province, Lanzhou 730020, China;*
 3. *National Meteorological Information Center, China Meteorological Administration, Beijing 100081, China*)

Abstract: Based on monthly and annual precipitation at 81 stations from 1951 to 2010 in Gansu Province, the normality of the observed precipitation and the precipitation series from Box – Cox transformation was analyzed by using Jarque – Bera test and the test of chi – square, respectively. The results show that annual precipitation followed the normal distribution at most of the stations but the monthly precipitation didn't obey that at half more stations, and in December, November, January the precipitation normality was worst and in May and September it was best. The normality was good in spring and summer but bad in fall and winter at most of stations. The precipitation exists non – normality in all Gansu Province, especially in the Hexi Corridor region, the center and east region of Gansu. Compared with the original series, the pass rate of normality about annual precipitation increased from 81.5% – 82.7% to 97.5% – 100% and monthly precipitation increased from 36.2% – 68.5% to 85.4% – 94.0% through the Box – Cox transformation, and most of stations followed the normality test.

Key words: Gansu; precipitation; Box – Cox transformation; normality analysis