

基于 Γ 分布的气压序列非均一性检验方法初探^{*}

吴必文¹⁾ 温华洋¹⁾²⁾ 惠 军²⁾

¹⁾(安徽省气候中心,合肥 230031) ²⁾(合肥工业大学理学院,合肥 230009)

摘 要

通过对安徽省内 23 个代表台站 1961—2006 年平均气压序列均一性检验的研究,提出了一种基于 Γ 分布气压序列非均一性检验方法—— Γ 检验法。以桐城市数据为例,给出了检验和分析的具体过程,最后将该方法检验结果与常见的基于正态分布的 SNHT 法检验结果进行了对比分析,结果表明:该方法具有更广泛的使用范围和较高的准确性,可以检验大部分年平均气压的非均一性。检验结果表明:气压计高度的调整、台站迁址是安徽省内气压序列非均一性的主要原因。

关键词: Γ 检验法;非均一性检验;气压序列;变点

引 言

随着全球变暖的趋势越来越明显,气候变化已成为一个备受关注的问题。气候变化研究是建立在能够反映气候真实变化的均一性长序列数据基础上的。然而由于观测台站的迁移,周围环境的变化,观测方法和观测仪器变更等常会使序列数据出现非均一性^[1]。近年来国内外很多专家致力于气候资料的非均一性检验的研究^[2-8],Alexandresson^[2]发展了广泛应用的 SNHT(标准正态检验)法,Easterling 等^[3]发展了 E-P 技术,全球历史气候网采用该技术对全球气温和气压进行均一性检验,Potter^[4]用 Potter 法对 19 个台站 100 年降水序列进行均一性检验。国内气候资料的均一性研究起步较晚。刘小宁等^[5]、宋超辉等^[6]、鞠晓慧等^[7]分别针对降水、平均气温、太阳辐射的均一性检验做了总结,取得了一些有益的结论。他们应用的方法大多要求待检验的气候序列服从正态分布,如距平累加法^[6,8], t 检验法^[6,9],SNHT 法^[2,5],Potter 法^[4,7]等。但是在研究部分台站的气压资料发现,并非所有序列都是服从正态分布的,即无法用这些方法来检验,需要一种新的方法来处理此类数据。

在统计学上,所谓均一性序列,是指序列样本 $\{x_i, i=1, 2, \dots, n\}$ 取自同一总体因而具有相同的概

率分布函数以及相同的数字特征,如数学期望 u 。而非均一性序列 $\{x_i\}$,是指不同时间段的序列取自不同的样本总体,即存在变点,变点前后的序列有不同的数学期望 $u, u+\Delta$ 。 Γ 分布参数变点研究的难点在于它的数学期望和方差同时依赖两个参数 ν 和 λ ,关于它的变点研究问题已有一些结论,Hsu^[10]给出了 ν 已知时 λ 的检验程序;谭常春等^[11-12]利用滑窗思想研究了至多一个变点 Γ 分布的两个参数 ν 和 λ 同时存在变点的情况,给出了变点的位置的检验和区间估计以及相关的概率性质。本文利用文献^[13]中给出的 Γ 分布参数变点检验思想,提出了基于 Γ 分布的气压序列非均一性检验方法—— Γ 检验法,用该方法来检验安徽省部分台站 1961—2006 年年平均气压资料,依据相应台站的历史沿革对检验出来的变点产生的原因进行了分析,最后与国内普遍使用的 SNHT 法进行对比,并给出相应的结论。

文中,若 $x \sim \Gamma(x; \nu, \lambda)$,记 $u = E(x) = \nu/\lambda, \sigma^2 = V_{ar}(x) = \nu/\lambda^2$ 。

1 检验方法简介

定义 1 若随机变量 x 有概率密度函数^[14]

$$f(x) = \frac{\lambda^\nu}{\Gamma(\nu)} e^{-\lambda x} x^{\nu-1}, x > 0 \quad (1)$$

则称 x 服从 Γ 分布,记为 $x \sim \Gamma(x; \nu, \lambda)$, ν 称为形状

* 中国气象局气候变化专项(CCSF2007-20)资助。

2007-07-25 收到,2008-01-04 收到再改稿。

参数, λ 称为尺度参数, 且有 $\nu > 0, \lambda > 0$ 。其中 $\Gamma(\cdot)$ 是 Γ 函数 $\Gamma(\nu) = \int_0^{\infty} t^{\nu-1} e^{-t} dt$ 。

定义 2 设 x_1, x_2, \dots, x_n 为一列相互独立的随机变量序列, 若 x_1, x_2, \dots, x_k 相互独立同分布于 $\Gamma(x; \nu_1, \lambda_1), x_{k+1}, \dots, x_n$ 相互独立同分布于 $\Gamma(x; \nu_2, \lambda_2)$, 其中 k 未知, 称 k 为该序列的变点。

在定义 2 上考虑参数 ν 至多只有一个变点的检验, 假设检验 $H_0: \nu_1 = \nu_2$ 对立假设为 $H_1: \nu_1 \neq \nu_2$, 因为 $\frac{E(x)}{\sqrt{V_{ar}(x)}} = \sqrt{\nu}$ 与 λ 无关, 且至多只有一个变点, 无论 λ 是否发生变化, 下面定义的统计量 T_k 都可以作为 ν 是否有变点的统计量^[13]:

$$T_k = \left(\frac{k(n-k)}{n}\right)^{\frac{1}{2}} \left(\frac{1}{k} \sum_{i=1}^k \frac{x_i}{\sigma} - \frac{1}{n-k} \sum_{i=k+1}^n \frac{x_i}{\sigma}\right) \quad (2)$$

其中, 当 σ 未知, 用 σ 的估计 $\hat{\sigma}$ 来代替, $\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ 。

选取水平 α (一般取 $\alpha = 0.05$), 则当且仅当

$$\max_{1 \leq k \leq n} |T_k| > a(x(\alpha), \ln n) \quad (3)$$

时, 拒绝 H_0 , 即认为序列有变点, 且变点位置在 k 处。此时检验的渐进水平为 α 。其中

$$x(\alpha) = -\ln\left(-\frac{1}{2} \ln(1-\alpha)\right) \quad (4)$$

$$a(x(\alpha), \ln n) = (2 \ln(\ln n))^{-\frac{1}{2}} (x + 2 \ln(\ln n) + \frac{1}{2} \ln(\ln(\ln n)) - \frac{1}{2} \ln \pi) \quad (5)$$

变点位置 k 的相合性, 收敛速度等概率性质可以参见文献[13], 下面给出一个检验过程: 如序列长度 $n=46$ (本文所用资料序列长度), 取水平 $\alpha=0.05$, 代入式(4), 计算得 $x(\alpha)=x(0.05)=3.663$ 。代入式(5), 得 $a(x(\alpha), \ln n)=a(x(0.05), \ln 46)=3.615$ 。当检验统计量 T_k 中得最大值 $\max_{1 \leq k \leq n} |T_k| > 3.615$, 则认为序列存在变点, 且变点位置在 k 处。记 $T_{\max} = \max_{1 \leq k \leq n} |T_k|$, $\alpha=0.05$ 时, 记临界值 $T_{95} = a(x(0.05), \ln n)$ 。表 1 给出本文可能用到的不同 n 值时的 T_{95} 值。

表 1 不同 n 值时 T_{95} 的值

Table 1 The value of T_{95} in different n

n	11	15	20	22	25	30	35	38	43	46	50	70	100
T_{95}	3.612	3.601	3.599	3.6	3.602	3.605	3.608	3.61	3.613	3.615	3.617	3.627	3.637

多个变点情况只需检出第一个变点后, 变点前后的序列分别再检验即可, 不再赘述。

参数 λ 的变点检验和参数 ν 的变点检验类似, 而一般情况下, 在气象资料上若变点存在, $\nu_1 \neq \nu_2$ 且 $\lambda_1 \neq \lambda_2$ 此情况包含在对立假设 H_1 中, 则通过式(2)即可检验。

2 检验过程

2.1 参照台站的选取

从 1961 年以来气压资料较为完整的台站, 选择距离待测台站较近, 地理环境相似, 没有明显不均一点的台站计算它们之间相关系数。相关系数 ρ 通过显著性检验^[15] (本文取水平 $\alpha=0.05$), 本文选取 1961—2006 年的资料, 序列长度为 46, 即当 $\rho > 0.291$ 时可以通过检验, 然后将相关系数最大的 5 个台站作为参照台站。

2.2 序列资料的预处理

① 对于待测台站气压序列 x_i 作变换: $f(x_i) =$

x_i/x , 其中 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

② 参照台站气压序列函数 $g(y_i)$, 由选出来的 5 个参照台站加权平均获得: $g(y_i) = \sum_{j=1}^5 r_j \left(\frac{y_{ij}}{y_i}\right)$,

其中 $\bar{y}_j = \frac{1}{n} \sum_{i=1}^n y_{ij}$ 表示第 j 个参照台站月平均气压序列的均值, $r_j = \rho_j^2 / \sum_{i=1}^5 \rho_i^2$ 表示第 j 个参照台站加权因子, ρ_j 表示第 j 个参照台站与待测站相关系数。

③ 形成待测序列 z 。

$$z_i = f(x_i)/g(y_i) \quad (6)$$

2.3 Γ 分布检验

在使用 Γ 检验法进行非均一性检验时, 若序列存在变点则要求变点前后均服从 Γ 分布, 若不存在则序列本身服从 Γ 分布。本文直接利用 Matlab^[16] 提供的 Kolmogorov 拟合优度检验函数 (简记 K-S 法) kstest 检验序列是否服从 Γ 分布, 部分伪代码如下:

```
phat = gamfit(z, a);
```

```
%假定序列服从  $\Gamma$  分布, 获得参数  $\nu, \lambda$  的估计
```

```

p=gamcdf(z,phat(1),phat(2));
%利用参数  $\nu, \lambda$  估计,得到一个累积分布函数
[H,S]=kstest(z,[z,p],a);
%利用 K-S 法进行分布检验
if H==0
%H=0 表明数据服从  $\Gamma$  分布, $a$  为置信水平
disp('该序列服从  $\Gamma$  分布') %输出
else
disp('该序列不服从  $\Gamma$  分布')

```

end

3 实际应用

在江淮流域安徽省内 1961—2006 年年平均气压资料相对完整的 40 多个台站中,根据自然区域选取淮北、江淮、江南共计 23 个台站作均一性检验的代表性台站进行试验。同时用 SNHT 对这些台站进行非均一性检验,检验结果如表 2。

表 2 Γ 检验与 SNHT 检验结果及原因分析
Table 2 The results and causes of Γ -test and SNHT

台站号	台站名	Γ 检验	Γ_{\max}	SNHT	原因分析
58015	砀山		3.489	2000 年	原因不明
58102	亳州		2.367		
58118	蒙城		1.746	1964 年	原因不明
58122	宿县	1976 年	4.730	1976 年	1976 年调低 3.8 m
				2000 年	原因不明
58203	阜阳	1971 年	5.173	1971 年	1972 年调高 6.4 m
		1986 年	3.954	1986 年	1987 年迁址
58215	寿县	1988 年	5.507	1988 年	1988 年调低 3.7 m
58221	蚌埠	1975 年	4.565	1975 年	原因不明
58225	定远	1984 年	4.602	1984 年	1984 年仪器变更
				1963 年	原因不明
58236	滁州	1976 年	4.958	1976 年	1977 年调高 6.9 m
58240	天长	1984 年	5.446	1984 年	1984 年调高 7.7 m
		1998 年	3.651	1998 年	原因不明
58317	岳西	1972 年	5.203	1972 年	1973 年迁址,调高 71 m
58319	桐城	1981 年	6.573	无法检验	1982 年迁址,调高 35.7 m
		1975 年	3.844		原因不明
58321	合肥	2003 年	5.571	2003 年	2004 年迁址,调低 8.5 m
				1968 年	原因不明
58326	巢湖		2.854		
58329	无为	1980 年	3.827	1980 年	1981 年调高 3 m
58334	芜湖	1985 年	4.087	1985 年	1986 年调高 6.5 m
				1975 年	原因不明
58419	东至	1975 年	4.328	1975 年	原因不明
		2003 年	3.722	2003 年	2003 年调低 3.5 m
58424	安庆	1976 年	6.547	1976 年	1977 年调低 25.8 m
58433	宣城	1974 年	3.804	无法检验	原因不明
		1984 年	4.201		1985 年调低 3.3 m
58436	宁国	1999 年	4.989	1999 年	1999 年调高 3.2 m
				1994 年	原因不明
58437	黄山	1985 年	4.379	无法检验	1984 年仪器变更
58530	歙县	1971 年	5.917	无法检验	1972 年迁址,调高 36.3 m
58531	屯溪		3.407	1964,2003 年	原因不明

注:①“无法检验”表示该台站的气压序列不服从正态分布(平方根处理后也不服从),无法用 SNHT 法检验;②“调低”(“调高”)是指气压表高度做了向下(向上)调整;③ 存在多个变点的,按照检验结果前后次序给出;④ 分段检验若变点不存在 Γ_{\max} 没有给出。

4 检验分析

4.1 结果分析

对 23 个代表台站采用 Γ 检验法进行检验,检

验结果统计情况是:存在一个变点的台站有 13 个,存在两个变点的台站有 5 个,共有 23 个变点,涉及到 18 个台站,占总数的 78.3%;另外 5 个台站无变点(表 3)。

根据变点位置查阅相应台站的历史沿革,判断

产生变点的原因,其中探明原因个数 18 个,其中 16 个变点的原因是气压表高度的调整和站址的迁移,占变点总数的 69.5%,即气压表高度的调整和站址的迁移是造成序列不均一的主要原因,另外 5 个变点没有得到相应台站历史沿革的印证,原因不明,存在误判的可能性。

表 3 变点个数统计

Table 3 The number of change points

	Γ 检验	占总数 百分比/%	SNHT	占总数 百分比/%
变点总数	23		26	
涉及台站数	18	78.3	21	91.3
无法检验站数			4	17.4
20 世纪 70 年代变点数	10	43.5	8	30.8
20 世纪 80 年代变点数	7	30.5	6	23.1
探明原因个数	18	78.3	14	53.8

变点位置的年代际分布主要集中在 20 世纪 70,80 年代,其中 70 年代变点数为 10 个,占 43.5%,但是历史沿革上却无法印证部分变点的存在;其中 80 年代变点数为 7 个,占 30.5%,80 年代

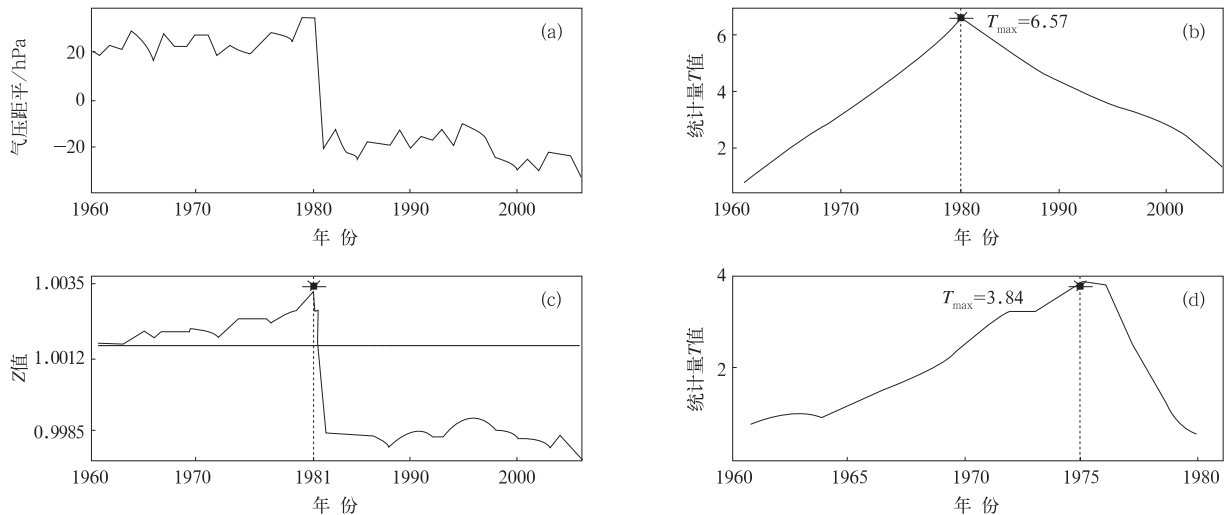
中期部分台站的水银气压表由定槽式变为动槽式也是这一阶段变点较多的原因之一。

4.2 个例分析

图 1a 给出了桐城站地面气压距平值,从图上可以看出在 1981 年出现了一个跳跃,为了检验是否存在变点,计算了检验统计量 T 值。由图 1b 中可以看出在 1981 年 $T_{\max}=6.573$ 远大于 $T_{95}=3.615$,图 1c 是由经过式(6)计算获得,从图中可以看出在 1981 年后 z 值全部在虚线以下,而 1981 年之前则大多数在虚线以上,佐证了该年存在变点的计算结果,查阅桐城站的历史沿革发现,1981 年 1 月该站从 $31^{\circ}03'N, 116^{\circ}57'E$ 迁到 $31^{\circ}04'N, 116^{\circ}57'E$,气压表高度由 46.1 m 调整为 81.8 m,升高了 35.7 m,气压表型号等没有发生变化,故认为台站的迁址以及气压表的高度调整是该变点产生的主要原因。

将序列分成前后两部分分别检验,发现 1975 年 $T_{\max}=3.844$ 大于 $T_{95}=3.604$,如图 1d,即在该年还存在一个变点,但没有探明其产生的具体原因。

在检验中发现桐城站年平均气压序列不服从正

图 1 桐城站地面气压距平序列(a), 1961—2006 年 T 值序列(b), Z 值序列(c), 1961—1981 年 T 值序列(d)Fig. 1 Pressure departure sequences (a), T sequences during 1961—2006 (b), Z sequences (c) and T sequences during 1961—1981 (d) of Tongcheng

态分布(用平方根法转换也不服从),SNHT 法无法检验出是否存在变点,而该序列经过式(6)预处理后形成的序列 z 的前后两段序列 1961—1981 年段,1982—2006 年段,均通过 Γ 分布检验,可以用 Γ 检验法来检验,显示了 Γ 检验使用范围的广泛性。

4.3 与 SNHT 法对比

本文将 Γ 检验法与 SNHT 法在进行气压序列

均一性检验的过程中作了较为全面的对比。

对原数据的要求,参考台站的选择等方面两者是相同的,而不同和优缺点主要体现在以下 3 个方面:

① 使用范围:SNHT 法要求原始序列服从正态分布,一旦原始序列不满足则无法使用,而 Γ 检验要求形成的检验序列服从 Γ 分布(若存在变点则要

求变点前后的两段序列服从 Γ 分布),而对原始序列没有要求。从表3中看到SNHT法有4个台站无法检验,而 Γ 检验则均可以。从本文的检验过程来看 Γ 检验适用范围更广。

② 检验结果:SNHT法比起 Γ 检验较为敏感,在有4个台站无法检验的情况下仍比 Γ 检验多检出3个变点,这样看来 Γ 检验存在着漏检的可能性,但是其探明原因的个数比 Γ 检验法少4个,即SNHT法存在大量无法得到台站历史沿革印证的变点,存在误判的可能性。

③ 计算复杂度:SNHT法在形成检验序列时要进行标准正态化, Γ 检验法没有要求,但 Γ 检验统计量计算的时间复杂度比SNHT法约高出两倍,在检验过程中程序的运行时间也证明了这一点。即 Γ 检验的计算复杂度偏高。

5 结论与存在的问题

1) 检验分析表明, Γ 检验法可以用于年平均气压序列的非均一性检验,相比SNHT法而言其使用范围更广,检验结果准确性较高,但计算复杂度较高,存在漏检的可能性。SNHT法高灵敏度的特点很好地补充了这个缺点,因此可以考虑将两者结合同时检验,相互补充,相互参照。

2) 在选择参考台站时发现,安徽省内台站气压的资料有较多的缺失,1961—2006年资料完全的台站约占总台站的一半。同时检验的结果表明,气压表高度的调整,以及台站迁址会引起气压序列不均一。由于安徽省多数台站气压表均有过调整或站址有过迁移,因此在使用气压资料时要慎重。

本文只考虑了 Γ 检验法在气温序列上的非均一性检验问题,能否将 Γ 检验在降水、气温、太阳辐射等气候资料的非均一性 Γ 检验上应用需要进一步探讨。

致 谢:中国科学技术大学谭常春博士,安徽省气候中心正

研级高工田红、徐光清为本文提供了技术上的支持,谨表谢忱。

参 考 文 献

- [1] 屠其璞,王俊德,丁裕国,等. 气象应用概率统计学. 北京:气象出版社,1984:488-489.
- [2] Alexandersson H. A homogeneity test applied to precipitation data. *International Journal of Climatology*, 1986,6: 661-675.
- [3] Easterling D R, Peterson T C. A new method for detecting and adjusting for undocumented in climatological time series. *International Journal of Climatology*, 1995,15: 369-377.
- [4] Potter K W. Illustration of a new test for detecting a shift in mean in precipitation series. *Mon Wea Rev*, 1981, 109:2040-2045.
- [5] 刘小宁,孙安健. 年降雨量序列非均一性检验方法探讨. *气象*,1995,21(8):3-6.
- [6] 宋超辉,刘小宁,李集明. 气温序列非一均性检验方法研究. *应用气象学报*,1995,6(3):289-296.
- [7] 鞠晓慧,屠其璞,李庆祥. 我国太阳总辐射月总量资料的均一性检验及订正. *南京气象学院学报*,2006,29(3):336-341.
- [8] Buishand T A. Some methods for testing the homogeneity of rainfall records. *Journal of Hydrology*, 1982, 58:11-27.
- [9] Worsley K J. On the likelihood ratio test for a shift in location of normal populations. *Journal of American Statistical Association*, 1979, 74:365-367.
- [10] Hsu D A. Detecting shifts of parameter in Gamma sequences with applications to stock price and air traffic flow analysis. *Journal of American Statistical Association*, 1979, 74: 31-40.
- [11] 谭常春,赵林城,缪柏其. 至多一个变点的 Γ 分布的统计推断及在金融中的应用. *系统科学与数学*,2007,27(1):2-10.
- [12] 谭常春,缪柏其. 至多一个变点 Γ 分布的统计推断. *中国科技大学学报*,2005,35(1):51-58.
- [13] 谭常春. 变点问题统计推断及其在金融中的应用. 合肥:中国科技大学,2007.
- [14] 孙俊锁. Gamma分布的特征函数及点估计. *鞍山钢铁学院院报*,2001,24(2):125-129.
- [15] 黄嘉佑. 气象统计分析与预报方法(第三版). 北京:气象出版社,2004:23-25.
- [16] 苏金明,阮沈用. Matlab 6.1实用指南(下册). 北京:电子工业出版社,2002:413-416.

The Inhomogeneity Test Method of Atmospheric Pressure Sequences Based on Γ -distribution

Wu Biwen¹⁾ Wen Huayang¹⁾²⁾ Hui Jun²⁾

¹⁾ (Anhui Provincial Climate Center, Hefei 230031)

²⁾ (School of Sciences, Hefei University of Technology, Hefei 230009)

Abstract

Climate change research is built on the basis of homogeneity sequence data which can reflect the real changes. Therefore the homogeneity and/or the inhomogeneity of the meteorological observation data should be tested before using the data for climate change research. Currently, the main methods of homogeneity test are based on the hypothesis that the data sequences obey the normal distribution.

However, in Anhui Province, it is found the average pressure sequence homogeneity test for 23 representative stations from 1961 to 2006 that some sequences are not subject to normal distribution, which can not be tested by the common method based on normal distribution. At the same time, it is found that some sequences obey Γ -distribution (pass the significant level of 0.05 Kolmogorov goodness-of-fit tests). A new homogeneity test method, Γ -test, for atmospheric pressure sequences is presented, in which the hypothesis should be met at first that the sequence is subject to Γ -distribution. Firstly, some reference stations are used to construct a reference sequence; then the ration sequences of stations which will be the test sequences to reference sequences are gotten; thirdly, the ration sequences are tested by Γ -test sequence homogeneity test; finally, the cause of the inhomogeneity sequences are analyzed by checking the history of the stations.

The method is briefly introduced, and the data of Tongcheng are taken as an example, the concrete process of Γ -testing and analysis are given. Furthermore, the result of Γ -test is compared with SNHT which is a widely use method. The results show that the 23 stations can be tested by Γ -test and the accuracy rate reaches 78.3%. In contrast, four stations can not be tested by SNHT and the accuracy only reaches 53.8%. In other words, Γ -test is superior to SNHT in scope and accuracy. Nevertheless, Γ -test also has some problems such as the high computational complexity and missing of some change points. The Homogeneity test results show that the change of barometer height and the station location is the main reason of the significant inhomogeneity of atmospheric pressure sequences in Anhui Province.

Key words: Γ -test method; atmospheric pressure series; inhomogeneity test; change points