

曹丽娟,鞠晓慧,刘小宁. PMFT 方法对我国年平均风速的均一性检验[J]. 气象,2010,36(10):52-56.

# PMFT 方法对我国年平均风速的均一性检验<sup>\*</sup>

曹丽娟 鞠晓慧 刘小宁

国家气象信息中心,北京 100081

**提 要:** 利用惩罚最大  $F$  检验(PMFT)方法并结合详尽的台站历史沿革信息对我国 701 个气象观测台站年平均风速资料进行了均一性检验,结果表明该检验方法能够应用于对年平均风速资料进行均一性检验,我国年平均风速资料的不连续点主要集中在 1~2 个,已检验的 701 个台站中有 61.3% 的台站年平均风速序列是均一的,表明我国大部分台站风速资料均一性情况良好。仪器变化和台站迁移是引起年平均风速序列非均一性的主要原因,而仪器变化中的仪器换型是造成年平均风速序列非均一性的最重要原因。

**关键词:** 年平均风速,均一性,PMFT

## Penalized Maximal $F$ Test for the Homogeneity Study of the Annual Mean Wind Speed over China

CAO Lijuan JU Xiaohui LIU Xiaoning

National Meteorological Information Centre, Beijing 100081

**Abstract:** The annual mean wind speed data of 701 meteorological observation stations have been detected by using Penalized Maximal  $F$  Test (PMFT) and the detailed metadata archive. The results show that the homogeneity detection method works well for the annual mean wind speed data over China. The number of change points of the annual mean wind speed may be one or two. The annual mean wind speed series of 61.3 percent are homogeneous among the detected 701 stations, showing that the homogeneity of the annual mean wind speed is good in most areas of China. The changes of instrument and location are the two main causes for the non-homogeneity, and the changes in the type of the observation instrument are the most important causes for the non-homogeneity of the annual mean wind speed over China.

**Key words:** annual mean wind speed, homogeneity, PMFT (penalized maximal  $F$  test)

## 引 言

在气候变化研究中,均一性的长序列资料是进行研究的基础,均一性的气候数据有益于真实可靠的评估历史气候趋势和变率,尤其是对于气候态和极端事件的研究非常重要,然而长序列的气候数据记录不可避免的存在由于观测仪器改变、观测方式改变、台站迁移等非气候因素造成的不连续点,制约

了资料同化系统的发展和气候变化模式预报、预测和预估的准确性。国内外许多专家在气候资料序列均一性检验和订正方面做了许多研究工作,并已经取得了重大进展,许多方法应用于各国气候资料业务、科研部门的研究,极大地推进了气候研究的发展<sup>[1-5]</sup>,值得我们借鉴参考。近年来我国不少研究人员也开始注重这方面的研究<sup>[6-10]</sup>,并且分别对我国的探空资料、气温、降水资料以及风速资料等进行了均一性检验与订正研究。在我国,对于气温数据的

\* 中国气象局气象新技术推广项目(CMATG2006Z03,CMATG2008Z12),科技部气象科学数据共享中心 01 课题,公益性行业(气象)科研专项项目(GYHY(QX)2007-6-5)和国家自然科学基金项目(No:40805060)共同资助

2009 年 3 月 1 日收稿; 2010 年 6 月 3 日收修定稿

第一作者:曹丽娟,从事气候资料分析及气候变化研究. Email:caolj@cma.gov.cn

均一化研究已经取得了一些突破性进展,建立了均一化的数据集,而对于降水和风速序列的均一性研究仍然停留于一些尝试性的水平,没有形成经过订正的均一化的数据集,许多相关领域的研究仍然基于原始的没有经过均一化处理的观测数据<sup>[11-13]</sup>,导致研究结论中存在着相当多的不确定性。因此,深入研究气候资料不同要素的均一性检验与订正方法,努力探索气候资料均一化业务实施手段,仍是当前气候学和气候变化研究面临的难点之一。

国际上针对不同的气候要素或时间尺度,发展了众多均一性研究方法,WMO 气候项目出版的指导手册至少列举了 14 种均一化研究方法,各种方法特点各异,或是引用的气候学基础不同,或是探查的时间尺度不同,应用于同一序列甚至会得到互相矛盾的结论。因此,通过试验对不同要素选取不同的均一性检验与订正方法,并根据其特点实施订正,不失为一个好方法。目前,均一化研究越来越关注对“无记录”间断点的检验和订正<sup>[14-15]</sup>，“无记录”间断点指的是没有元数据支持的间断点,而原有一些均一性检验方法对这些间断点的检验能力有限。

随着我国风电产业的发展,气候变暖背景下风能资源的可持续利用问题倍受关注,然而,对于风速资料的均一性研究还没有较为成熟的方法,对我国近 50 年风速变化特征的分析仍然基于原始的没有经过均一化的气候观测序列,导致研究结论中存在着相当多的不确定性。已有研究表明,风速的概率分布并非完全是标准正态分布,而是正偏态分布或韦布尔分布,如采用 SNHT(标准正态均一检验)等方法进行均一性检验则需要首先进行序列的正态转换<sup>[7]</sup>。我国历史上风速仪器存在多次同期批量换型的情况,建站初期气象台站使用的风向风速观测仪器大多是维尔德风压板,其构造简单、灵敏性差,所测资料缺乏准确性,部分台站配备了达因式风向风速计等其他仪器。1966 年开始,国产 EL 型电接风向风速计陆续装备全国各气象台站,有效提高了风速观测的精度,20 世纪 90 年代初期,部分台站使用的 EL 型电接风向风速计又被 EN 型系列测风数据处理仪所替换,该仪器具有测量准确可靠,使用安装简便等优势。仪器的同期换型增加了参考序列的建立难度,此外,即使参考序列的选取对序列的均一性检验有很大帮助,然而如果选取的参考序列本身就存在非均一性问题,则会大大影响均一性检验的效果。王小兰等<sup>[16]</sup>在对二相回归统计检验方法进行

了系列改进之后,提出了惩罚最大  $F$  检验方法(PMFT),该方法能够有效避免非均一的参考序列带来的检验误差,便于对风速资料等受局地环境影响较大的要素进行检验。本研究选用该均一化检验方法开展我国共 701 个基本、基准气象观测台站年平均风速数据的均一性研究,通过均一性检验,研究并分析了引起我国气象台站年平均风速数据非均一性的原因。

## 1 资料与均一化检验方法

本研究利用由加拿大环境部气候研究中心建立的均一化检验方法 PMFT 结合历史台站沿革信息开展我国年平均风速资料的均一性检验,资料来源于国家气象信息中心气象资料室建立的中国地面气候资料年值数据集,已经过质量控制,时间范围为 1951—2006 年,共包括 753 个基本、基准站的资料。

PMFT 方法是基于惩罚最大  $F$  检验<sup>[16]</sup>,经验性的考虑了时间序列的一阶滞后自相关,并嵌入回归检验算法,能够用于检验、订正包含一阶自回归误差的数据序列的多个间断点(平均突变)。通过应用经验的惩罚函数,使得误报警率和检验能力的非均匀分布问题也大大减少,利用该方法进行间断点检验,可以不使用参考序列。

检验过程如下:

假设  $\epsilon_t$  代表变量,该变量均值为 0,方差为  $\sigma^2$ ,对于存在线性趋势  $\beta$  的时间序列  $\{X_t\}$ ,要检验在  $t = k$  时刻是否存在一个平均突变,

原假设:

$$H_0: X_t = \mu + \beta t + \epsilon_t, 1, 2, \dots, N \quad (1)$$

备择假设:

$$H_a: \begin{cases} X_t = \mu_1 + \beta t + \epsilon_t, & t \leq k \\ X_t = \mu_2 + \beta t + \epsilon_t, & k - 1 \leq t \leq N \end{cases} \quad (2)$$

$\mu_1 \neq \mu_2$ ,当  $H_a$  为真时, $t = k$  时的点被称作间断点。 $\Delta = |\mu_1 - \mu_2|$  被称作平均突变的大小,最可能的间断点服从以下分布:

$$PF_{\max} = \max_{1 \leq k \leq N-1} [P(k)F_c(k)] \quad (3)$$

其中: $P(k)$  是建立的经验性的惩罚因子,其建立方法可参照文献<sup>[16]</sup>。

$$F_c(k) = \frac{(SSE_0 - SSE_A)}{SSE_A / (N - 3)}$$

$$SSE_A = \sum_{t=1}^k (X_t - \hat{\mu}_1 - \hat{\beta}t)^2 + \sum_{t=k+1}^N (X_t - \hat{\mu}_2 - \hat{\beta}t)^2$$

$$SSE_0 = \sum_{t=1}^N (X_t - \hat{\mu}_0 - \hat{\beta}_0 t)^2$$

其中  $\hat{\mu}_0$  和  $\hat{\beta}_0$  是在  $\mu_1 = \mu_2 = \mu$  时的估计值。

通过回归检验算法来检验出多个间断点,首先在  $t = \{N_{\min}, N_{\min+1}, \dots, (N - N_{\min})\}$  时段找出最可能的间断点  $C_0$ ,类似的,分别在时段  $t = \{N_{\min}, \dots, (C_0 - N_{\min})\}$  和  $t = \{(C_0 + 1 + N_{\min}), \dots, (N - N_{\min})\}$  找出次可能的间断点  $C_{01}$  和  $C_{03}$ ,接下来在  $C_{01}$  和  $C_{03}$  之间的时段找出最可能的间断点  $C_{02}$ ,然后在  $C_{01}, C_{02}$  和  $C_{03}$  之间寻找最可能的间断点,即依次分段找出序列中各段最可能的间断点,计算所有间断点的统计量  $PF_{\max}$ ,估计自相关和  $P$  值的大小,找出最大的  $PF_{\max}$  值对应的间断点,如果是显著的,该突变就被认为是找到的第一个间断点,该间断点被记为  $N_c = 1$ 。寻找该间断点位置之后每段最可能的间断点,估计其显著性,找出下一个可能的间断点,重复该过程,将间断点按照显著性由大到小排列,形成间断点列表,判断最小的间断点是否显著,当不显著时剔除该间断点,需要再次评估剩余间断点的显著性,最终保留都统计显著的间断点即为序列检验得到的间断点,受分段检验的影响,该方法能够检验的最短序列长度是 20,详细过程参见文献[17]。

## 2 检验结果分析

年平均风速是逐年振动很小的气象要素,检验其逐年平均记录,更容易检查出由于台站迁移、仪器换型、周围环境变化等引起的非均一。本研究针对全国 753 个基本、基准气象观测站年平均风速资料进行了均一性检验,检验结果如下:753 个台站中,因资料长度较短(不足 20 年)而未能进行检验的台站有 52 个,占台站总数的 6.9%。在已检验的 701 个台站中,共检验出 271 个站有间断点,占被检验台站的 38.7%,间断点总个数为 356 个,其空间分布如图 1,由图中可以看出,全国各地年平均风速资料都存在非均一性的问题,非均一性台站在全国范围有均匀分布的特点,表明我国风速资料在空间上普遍存在不同程度的非均一性问题,因此有必要进行风速资料的均一性检验与订正。已检验的 701 个台站中有 231 个站年平均风速资料质量较好,序列是均一的,另有 199 个台站虽然都检验出了间断点,但是由于检验出的间断点并不显著,因此,认为该部分台站资料在给定的置信度水平下不存在显著性间断

点,这些台站资料可以近似认为是均一的。资料序列均一的台站和检验出间断点不显著台站共 430 个,占已检验台站的 61.3%,由此可见,虽然风速数据在全国各地普遍存在非均一的现象,但是就台站数量而言,仍然有超过 60% 的台站年平均风速序列是均一的,表明我国大部分台站风速资料均一性情况良好。

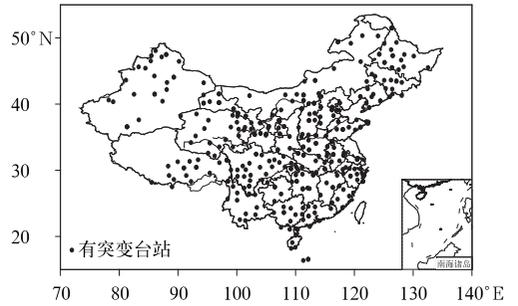


图 1 存在间断点台站分布

Fig. 1 Spatial distribution of observation stations with the discontinuities

对于经过检验的 701 个基本、基准站,共 271 个站有间断点,其中,存在 1 个间断点的台站有 195 个,占 27.8%,存在 2 个间断点的台站有 69 个,占 9.8%,存在 3 个间断点的台站有 7 个,占 1.0%,这 7 个台站年平均风速资料均一性较差,7 个台站共检验出 21 个间断点,其中有 15 个是由于仪器换型、仪器更换或仪器高度变化引起,只有 2 个间断点是由于台站迁移引起,而其余 4 个间断点找不到元数据的支持。由以上分析可以看出,我国基本、基准站年平均风速序列的间断点主要集中在 1~2 个,仪器变化是引起个别台站间断点较多(达到 3 个)的重要原因。

综合分析所有检测出的 356 个间断点,其中有 68 个间断点是由于台站迁移引起,占总间断点的 19.1%;因风速仪器换型引起的间断点有 153 个,占总间断点的 43%;因风速观测仪器高度变化引起的间断点有 55 个,占 15.4%;由于仪器故障等原因造成仪器更换引起的间断点有 37 个,占 10.4%;原因不明的间断点有 41 个,占 11.5%;由于台站环境改变引起的间断点仅有两个。由此可见,仪器换型与台站迁移是引起年平均风速资料序列非均一的主要原因(图 2),此外,因仪器高度变化、同类型仪器更换、观测场环境变化等也造成部分台站风速资料序列的不连续。

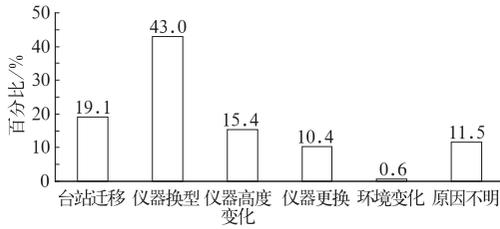


图 2 不同类型间断点百分比分布

Fig. 2 Percentage distribution of different types of the discontinuities

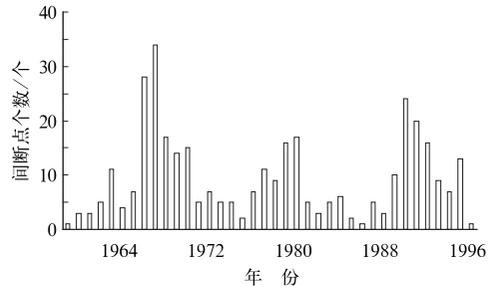


图 3 各年间断点分布

Fig. 3 Inter-annual distribution of the discontinuities

图 3 为我国基本、基准站各年间断点个数的统计分布,可以看到出现间断点最多的年份是 1968 年,其次是 1967 年,另一个间断点出现的高峰时段是 1991 年和 1992 年。根据历史台站沿革信息,可以看到我国在 1967—1970 年间风速观测仪器普遍进行了换型,由风压板换型为 EL 型电接风向风速计,而在 1991 年和 1992 年,存在由 EL 型电接风向风速计向 EN 型电接风向风速计的换型。许多研究结果都表明,风速资料的均一性同观测仪器的变化尤为密切,而仪器换型是仪器变化最主要的因素,本研究结果也证明我国多年平均风速资料的非均一性主要是由于风速观测仪器变化引起。

图 4 给出山东惠民站和四川稻城站订正前后年平均风速序列,山东惠民站检验出的两个间断点分别在 1967 年和 1991 年,检查元数据信息发现,该站在我国历史上两次风速仪器批量换型时期的 1968 和 1992 年都进行了仪器的换型,1968 年风速观测仪器由维尔德测风器换型为 EL 型电接风向风速计,1992 年风速观测仪器由 EL 型电接风向风速计换型为 EN 型测风数据处理仪。四川稻城站检验出一个间断点位于 1971 年,而该站在 1972 年存在由维尔德测风器向 EL 型电接风向风速计的换型。可以看出,采用该方法能够检验出由于仪器换型引起的显著突变。

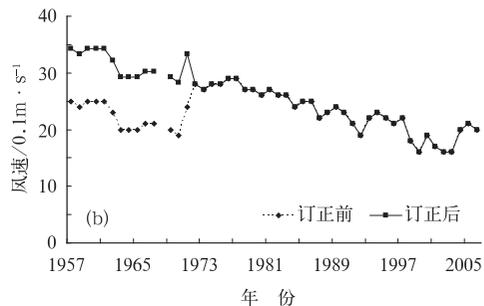
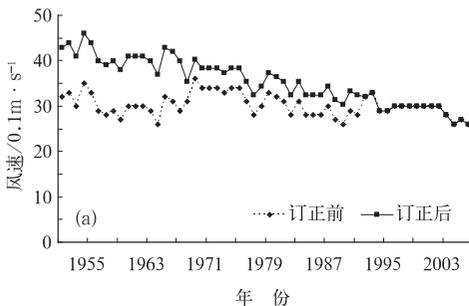


图 4 惠民站(a)和稻城站(b)订正前后年平均风速序列

Fig. 4 The time series of annual mean wind speed at Stations Huimin (a) and Daocheng (b) before (dashed curve) and after (solid curve) adjustment

### 3 结论与讨论

本研究应用 PMFT 检验方法对我国 753 个基本、基准气象观测台站年平均风速数据进行了均一性检验,该方法提供了一种不依赖于参考序列的检验方法,检验结果表明,本方法能够用于对我国年平均风速资料进行均一性检验。

我国各地区年平均风速资料都存在不同程度的

非均一现象,但是就台站数量而言,仍然有超过 60% 的台站年平均风速序列是均一的,表明我国大部分台站风速资料均一性情况良好。造成年平均风速资料非均一性的主要原因是风速观测仪器的变化,包括仪器换型、同类仪器高度的变化以及仪器故障之后的更换等,其中仪器换型是引起风速数据非均一性的主要原因,其次是台站迁移的影响较大。环境因素对年平均风速数据的非均一性影响似乎并不显著,这种环境因素的改变可能同时被台站的迁

移或者仪器的变化所掩盖。总之,在使用年平均风速数据进行历史趋势和变率的分析与评估时,必须充分考虑风速的非均一性问题。

本研究是应用 PMFT 检验方法对风速资料进行均一性检验的初步试验,虽然取得了较为满意的结果,但是由于风速资料本身的特殊性,其概率分布的非正态特性并不能完全符合均一性检验的假设,此外风速资料受地形、地貌及障碍物影响等的局地性特点,未来需要对该检验方法进行多个实验的测试和改进,同时需要利用更加详尽的元数据信息对订正结果进行分析,在此基础上开展深入的均一性研究。对于因序列较短未能进行检验的台站,将选取新的均一性检验与订正方法并利用月平均或日平均风速资料进行检验,以期得到完整的均一化检验结果。

## 参考文献

- [1] Alexandersson H. A homogeneity test applied to precipitation data[J]. *International Journal of Climatology*, 1986, 6: 661-675.
- [2] Jones P D, Bradley R S, Diaz H F et al. Northern Hemisphere surface air temperature variations: 1851 - 1984 [J]. *Journal of Climate and Applied Meteorology*, 1986, 25: 161-179.
- [3] Easterling D R, Peterson T C. A new method for detecting and adjusting for undocumented discontinuities in climatological time series [J]. *International Journal of Climatology*, 1995, 15: 369-377.
- [4] Herzog J and Mülle W G. Homogenization of various climatological parameters in the German weather service [R]. In: *Proceedings of the first seminar for homogenization of surface climatological data*, 1996: 101-111.
- [5] Peterson T C, Easterling D R, Karl T R, et al. Homogeneity adjustments of in situ atmospheric climate data: A review [J]. *International Journal of Climatology*, 1998, 18: 1493-1517.
- [6] 翟盘茂. 中国历史探空资料中的一些过失误差及偏差问题 [J]. *气象学报*, 1997, 55(5): 563-572.
- [7] 刘小宁. 我国 40 年年平均风速的均一性检验 [J]. *应用气象学报*, 2000, 11(2): 27-34.
- [8] Li Q X, Liu X N, Zhang H Z et al. Detecting and adjusting on temporal inhomogeneity in Chinese mean surface air temperature dataset [J]. *Adv Atmos Sci*, 2004, 21(2): 260-268.
- [9] 李庆祥, Menne M J, Williams C N, Jr, 等. 利用多模式对中国气温序列中不连续点的检测 [J]. *气候与环境研究*, 2005, 10(4): 736-742.
- [10] 张智, 林莉, 梁培, 等. 宁夏年气温资料的均一性检验研究 [J]. *气象*, 2009, 35(10): 79-83.
- [11] 毛飞, 卢志光, 郑凌云, 等. 近 40 年那曲地区日照时数和风速变化特征 [J]. *气象*, 2006, 32(9): 77-83.
- [12] 刘学锋, 任国玉, 梁秀慧, 等. 河北地区边界层内不同高度风速变化特征 [J]. *气象*, 2009, 35(7): 46-53.
- [13] 傅玮东, 姚艳丽, 李迎春. 新疆吉木萨尔县 45 年气候变化特征分析 [J]. *气象*, 2007, 33(6): 96-101.
- [14] Lund R, and Reeves J. Detection of undocumented change-points: A revision of the two-phase regression model [J]. *Journal of Climate*, 2002, 15: 2547-2554.
- [15] Wang X L. Comments on 'Detection of Undocumented Change points: A revision of the two-phase regression model' [J]. *Journal of Climate*, 2003, 16: 3383-3385.
- [16] Wang X L. Penalized maximal  $F$ -test for detecting undocumented mean-shifts without trend-change [J]. *J Atmos Oceanic Tech*, 2008, 25: 368-384.
- [17] Wang X L. Accounting for autocorrelation in detecting mean shifts in climate data series using the Penalized Maximal  $t$  or  $F$  Test [J]. *Journal of Applied Meteorology and Climatology*, 2008, 47(9): 2423-2444.