

## 研究气温变化趋势的新方法——三组耐抗线

刘俊卿<sup>1,2</sup>, 杨军<sup>1</sup>, 强德厚<sup>2</sup>

(1. 南京信息工程大学应用气象学院, 江苏 南京 210044; 2. 日喀则地区气象局, 西藏 日喀则 857000)

**摘要:**针对气象资料难以满足高斯分布的情况,引进了具有耐抗性的三组耐抗线方法,分析了满足不同信度情况下最小二乘回归和三组耐抗线的差异,并用西藏近30 a气温资料进行了检验。结果表明三组耐抗线方法比最小二乘法研究气候变化趋势有明显优势;采用最小二乘法得到的西藏气候变暖估计偏高。

**关键词:**三组耐抗线;最小二乘法;气温变化

**中图分类号:**P468

**文献标识码:**A

## 引言

气候变化指气象要素在较长时间范围内的趋势变化。我国多数站点建站晚,气象资料行为缺乏大样本的简单性,并且难以满足经典统计最根本的高斯分布假设。虽然传统最小二乘回归方法代数运算简单,但是其推导完全建立在高斯分布基础之上,不提供耐抗性。要减小异常值对结果的破坏性影响,需要采用具有耐抗性的分析方法。而探索性数据分析方法<sup>[1-2]</sup>的目的在于提供耐抗性,即很重视数据的主体部分,几乎不重视离群值,数据的局部不良行为对拟合线的影响较小。即使少数点偏离基本模式很大,这些异常数据点一般只对拟合线有小的效应。西藏多数站点建站晚,难以满足经典统计最根本的高斯分布假设,故比较适合做最小二乘回归方法和三组耐抗线的比较研究。本文采用西藏1971~2000年平均气温和极端温度资料,针对不同信度检验情况对以上2种方法作对比分析,说明2种方法在气候趋势分析中的差异。

## 1 三组耐抗线(TGRL)算法介绍

对于数据 $(x_i, y_i), i=1, \dots, n$ ,拟合一条直线: $y = a + bx$ ,三组耐抗线基本思路:首先按 $x_i$ 的升序对数据集排列,并将其分为3组,其配置如表1所

示,使每组数据的样本个数尽可能相等。独立求取各组内 $x$ 和 $y$ 的中位数,得到3个总括点: $L(x_L, y_L), M(x_M, y_M), R(x_R, y_R)$ ,中位数提供对于 $x, y$ 的离群值的耐抗性。利用3个总括点得到斜率和截距分别为:

$$b_0 = \frac{y_R - y_L}{x_R - x_L} \quad (1)$$

$$a_0 = \frac{1}{3} [(y_L - b_0 x_L) + (y_M - b_0 x_M) + (y_R - b_0 x_R)] \quad (2)$$

$a_0$ 和 $b_0$ 都基于总括点的,是耐抗的。用斜率和中心值来表示拟合,在 $x = x_M$ 处初始直线为:

$$\hat{y} = a_0^* + b_0(x - x_M) \quad (3)$$

其中中心值( $a_0^*$ )是:

$$a_0^* = \frac{1}{3} \{ [y_L - b_0(x_L - x_M)] + y_M + [y_R - b_0(x_R - x_M)] \} \quad (4)$$

得到拟合直线的斜率的水平,下一步计算每个数据点的初始残差:

$$r_i^{(0)} = y_i - [a_0^* + b_0(x_i - x_M)] \quad (5)$$

残差本身能够揭示各种特性和模式,这里只利用它的一个一般性质:用残差代替原来 $y$ 值,即用 $(x_i, r_i)$ 代替 $(x_i, y_i)$ ,然后重复拟合过程,得到斜率和水平的调整值 $\delta_1$ 和 $\gamma_1$ ,将其分别与初始斜率和水平累加,继续下去就得到零拟合,即得到直线的零斜率和

零水平。这一过程即为迭代,  $j$  为迭代步数。

$$r_i^{(j)} = r_i^{(j-1)} - [\gamma_j + \delta_j(x_i - x_M)],$$

$$(i = 1, \dots, n) \quad (6)$$

如果迭代  $k$  步后, 当斜率调整值  $\delta_k$  的绝对值与初始斜率  $b_0$  的比值小于某个阈值(0.01%)时, 则可停止迭代, 得到的最终斜率和水平为:

$$b = b_0 + \delta_1 + \dots + \delta_k \quad (7)$$

$$a = a_0 + \gamma_1 + \dots + \gamma_k \quad (8)$$

相应的拟合直线为:

$$\hat{y} = a + b(x - x_M) \quad (9)$$

表1 各组的数据配置情况表

Tab.1 Data distribution of every group

组	$n=3k$	$n=3k+1$	$n=2k+2$
左	$k$	$k$	$k+1$
中	$k$	$k+1$	$k$
右	$k$	$k$	$k+1$

## 2 三组耐抗线(TGRL)方法与最小二乘回归线(LSL)的比较

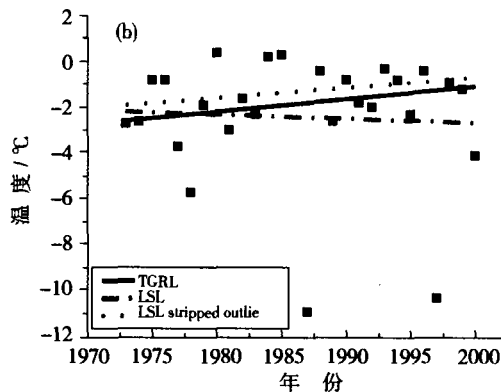
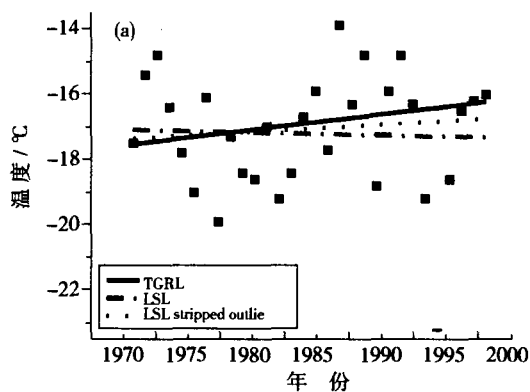


图1 帕里站1月月平均最低气温年变化(a), 改则站12月月平均最高气温年变化(b)

Fig.1 Yearly variations of  $T_{min}$  of Pali in January (a) and  $T_{max}$  of Gaize in December (b) during 1971 - 2000

表2 近30 a月平均极端温度  
线性拟合斜率(°C/100 a)比较

Tab.2 Comparison between variations from Three Group Resistance Line and those from Least Square Regression method in 30-year

站名	要素	月份	TGRL	舍点后 TGRL	LSL	舍点后 LSL
帕里	最低气温	1月	4.47	5.35	-0.78	2.1
改则	最高气温	12月	5.6	5.45	-1.77	4.51

### 2.1 TGRL方法与LSL方法斜率计算的比较

以1971~2000年帕里站4月  $T_{min}$  (月平均最低气温)和改则12月  $T_{max}$  (月平均最高气温)的线性趋势说明三组耐抗线与最小二乘回归线之间的差异。从图1可以看出, 用三组耐抗线法与最小二乘法分别得到了符号相反的变化趋势。为了说明2种方法的耐抗性, 舍去图1(a)的离群值(温度最小的1个极值点)和图1(b)中的离群值(温度最小的3个极值点), 因为这些值在次序统计量中明显偏离数据主体, 根据 Grubbs 检验法<sup>[3]</sup>判断此为离群值。重新计算斜率, 结果列于表2中。可以看到舍去离群值后, 最小二乘法拟合的斜率均向三组耐抗线的结果靠拢, 而且二者的斜率改变了符号; 并且三组耐抗线的斜率变化幅度明显小于最小二乘回归线。由此说明, 个别异常值对最小二乘回归线的斜率有很大影响, 甚至使斜率符号发生改变。而三组耐抗线方法对个别异常值不敏感, 能更好地显示数据的主要行为。因此用三组耐抗线方法分析西藏气候变化趋势具有更明显的优势。

### 2.2 不同信度情况下2种方法的比较

表3为  $\Delta T$ 、 $\Delta T_{min}$  满足99%、95%和90%信度的站点的最小二乘法和三组耐抗线的结果差值。由表3可知, 对于  $|\Delta T|$  随信度的分布情况, 在满足99%的信度情况下, 41.7%的  $|\Delta T| < 1 \text{ } ^\circ\text{C}/100 \text{ a}$ ,  $|\Delta T| > 1 \text{ } ^\circ\text{C}/100 \text{ a}$  的概率为58.3%。而信度为95%和90%时,  $|\Delta T|$  在  $< 1 \text{ } ^\circ\text{C}/100 \text{ a}$  的区间内出现的概率分别为25.5%、18.3%, 明显小于99%的

情况,而 $|\Delta T|$ 在 $>1\text{ }^{\circ}\text{C}/100\text{ a}$ 的区间内出现的概率分别为74.5%、81.7%,明显大于信度为99%的情况。由此可知,对于样本数据不能满足高斯分布时,信度越高, $|\Delta T|$ 的大值分布出现的概率越低,而信度较低时, $|\Delta T|$ 的大值分布出现的概率越小。对于 $|\Delta T|$ 随信度的分布情况与 $|\Delta T|$ 的分布情况类似,信度大小与 $|\Delta T|$ 的大值分布出现的概率呈相反关系。

由以上分析说明,最小二乘法和三组耐抗线的结果的分布有明显的不同;同时,信度越高,2种方法计算结果的差值越小,即最小二乘拟合线与三组耐抗线的斜率趋于一致;而信度越低时,最小二乘拟合线与三组耐抗线的斜率的差值越大。

表3 西藏30 a  $T$ 、 $T_{\min}$ 满足不同信度情况下最小二乘法和三组耐抗线的结果差值的分布(单位: $^{\circ}\text{C}/100\text{ a}$ )

Tab.3 Distribution of the differences between variations of  $T$  and  $T_{\min}$  for different degree of confidence from Three Group Resistance Line and those from Least Square Regression method in 30-year in Tibet

	99%	95%	90%
$ \Delta T  \leq 1$	41.7%	25.5%	18.3%
$1 <  \Delta T  \leq 2$	41.7%	52.9%	56.3%
$ \Delta T  > 2$	16.6%	21.6%	25.4%
$ \Delta T_{\min}  \leq 1$	35.9%	25.1%	24.5%
$1 <  \Delta T_{\min}  \leq 2$	50.0%	52.9%	55.2%
$ \Delta T_{\min}  > 2$	14.1%	22.0%	20.3%

### 3 西藏平均气温的变化趋势

西藏年平均气温(图2a)呈整体增温趋势,阿里、那曲东北部的增温较为明显,其中那曲\*( $0.48\text{ }^{\circ}\text{C}/10\text{ a}$ )增温幅度最大,达到0.05的信度水平(\*\*表示达到0.05的信度水平,\*表示达到0.10的信度水平,下同)其余在 $0.04\text{ }^{\circ}\text{C}/10\text{ a}$ 的范围内呈增温趋势;仅西藏东南部有小块降温。

平均气温变化趋势在各季不尽相同。春季呈增温趋势。阿里、那曲增温显著,其中那曲的班戈\*\*增温率最大为 $0.53\text{ }^{\circ}\text{C}/10\text{ a}$ ;西藏东南部的降温幅度为 $0.01\text{ }^{\circ}\text{C}/10\text{ a}$ 。夏季(图2b)是增温强度最大的季节。西藏的阿里、那曲增温显著,在班

戈、狮泉河附近各有一增温中心,其中增温率最大的为改则\*\*( $0.73\text{ }^{\circ}\text{C}/10\text{ a}$ ),其次是班戈\*\*; $0.68\text{ }^{\circ}\text{C}/10\text{ a}$ ;西藏东南部呈较弱的增温趋势。秋季呈增温趋势,强度比夏季略小。西藏的阿里、那曲及西藏东南部地区为增温显著区,以改则最显著( $0.56\text{ }^{\circ}\text{C}/10\text{ a}$ );降温区主要位于西藏的东北部及南部。冬季(图2c)是唯一出现较大面积降温区域的季节,以那曲东北部的索县降温趋势最大,降温率为 $0.45\text{ }^{\circ}\text{C}/10\text{ a}$ 。仅阿里、日喀则和林芝的部分地区为增温趋势,其中增温最显著的为拉萨\*( $0.3\text{ }^{\circ}\text{C}/10\text{ a}$ )。

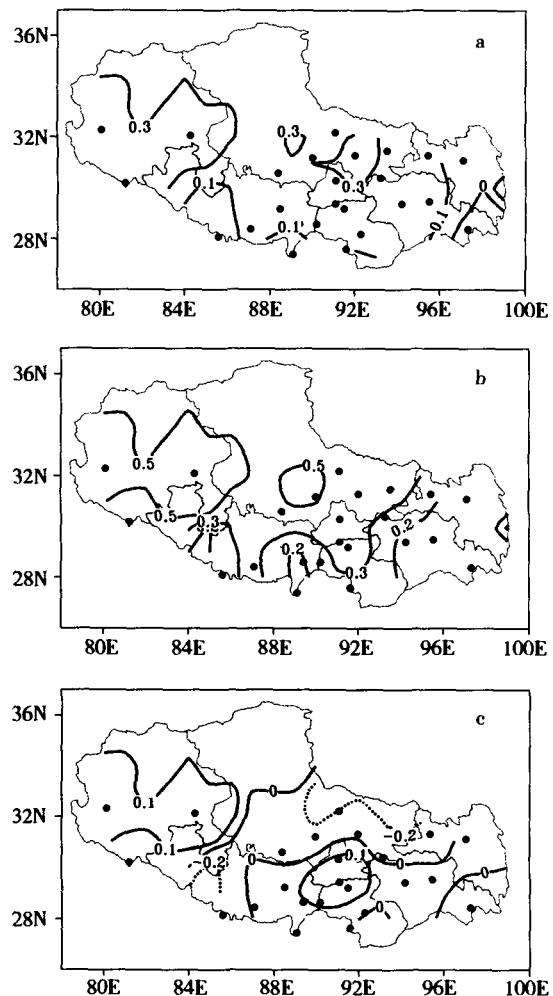


图2 西藏1971~2000年平均气温变化趋势分布 (a)全年;(b)夏季;(c)冬季;单位: $^{\circ}\text{C}/10\text{ a}$   
Fig.2 Variations of mean temperature in Tibet during 1971-2000, (a)Yearly; (b)Summer; (c)Winter (Unit:  $^{\circ}\text{C}/10\text{ a}$ )

上述结果与近年来他人研究西藏气温变化趋势比较,从年平均变化趋势看,本文与牛涛等<sup>[4]</sup>、杜军<sup>[5]</sup>的研究基本一致,但存在一些差异。牛涛指出藏北地区的增温率为 $0.4\text{ }^{\circ}\text{C}/10\text{ a}$ ,而本文的结果只有该地区的那曲站增温率为 $0.48\text{ }^{\circ}\text{C}/10\text{ a}$ ,其余站点的变温幅度在 $0.13\sim 0.24\text{ }^{\circ}\text{C}/10\text{ a}$ 之间;并且文献[4]中有阿里大部、那曲中西部呈 $0.46\sim 0.72\text{ }^{\circ}\text{C}/10\text{ a}$ 的速率增大。因此采用最小二乘法得到的对西藏气温增高的估计可能偏大。

## 4 结 论

(1)三组耐抗线具有较好的耐抗性,最小二乘法和三组耐抗线的结果的分布有明显的不同,信度越高, $\Delta T$ 较小的概率越高。因此,最小二乘拟合线与三组耐抗线的斜率趋于一致,当气象资料不能满足正态分布时,用三组耐抗线方法比最小二乘法具有更明显的优势。

(2)因为个别异常值对最小二乘回归线的斜率

有很大影响,甚至使斜率符号发生改变。而三组耐抗线方法对个别异常值不敏感,能更好地显示数据的主要行为。故采用最小二乘法计算西藏平均气温的变温强度结果偏大,从而导致对西藏气候变暖估计偏高。

### 参考文献:

- [1] Hoaglin D C, Mosteller F, Tukey J W. 探索性数据分析[M]. 陈忠琰,郭德媛译. 北京:中国统计出版社,1998. 135 - 146.
- [2] 杨军. 近50年中国区域地面温度趋势变化特征分析[A]. 中国地区大气气溶胶的基本特征及其气候效应[D]. 南京气象学院,2001.
- [3] 邱德仁. 工业分析化学[M]. 上海:复旦大学出版社,2003, 208 - 210.
- [4] Niu Tao, Chen Longxun, Zhou Zijiang. The characteristics of climate change over the Tibetan Plateau in the last 40 years and the detection of climate jumps[J]. Adv Atmos Sci, 2004, 21(2):193 - 203.
- [5] 杜军. 西藏高原近40年气温变化[J]. 地理学报,2001,56(6): 682 - 689.

## A New Method for Calculating Temperature Variation: Three Group Resistance Line

LIU Junqing<sup>1,2</sup>, YANG Jun<sup>1</sup>, QIANG Dehou<sup>2</sup>

- (1. Department of Applied Meteorology, Nanjing University of Information Science & Technology, Nanjing 210044, China;
2. Rikeze Meteorological Bureau, Rikeze 857000, Xizang, China)

**Abstract:** In the light of the question that meteorological surface data are unfit to Gaussian distribution, the variation tendencies computed with Tibet surface air temperature data by Three Group Resistance Line and Least Square Regression Method were compared in this paper. Results show that Three Group Resistance Line is not significantly influenced by outliers and is useful for tendency analysis of meteorological records sometimes without sufficient number of samples. The tendencies from Three Group Resistance Line were less than those from Least Square Regression method in Tibet during 1971 - 2000.

**Key words:** Three Group Resistance Line; Least Square Regression Method; temperature variation