

利用观测资料求区域平均值过程中 应注意的一个问题*

牛 涛 刘洪利 陈隆勋

(中国气象科学研究院,北京 100081)

摘 要

为了得到描述一个区域整体特征的时间序列,客观分析和区域化求平均是在气象、海洋等领域普遍应用的资料处理方法。作者分析了应用 Barnes 方法和先区域化再求平均值方法得到的青藏高原地温、气温、气压等资料的近 40 年变化曲线,可以看出在地形较复杂地区,如果观测资料连续性不好,这两种方法在实际应用中都必须消除地区差异的影响,才能得到较为可信的结论,否则就有可能得到与事实相背的结果。这一问题应该引起足够的重视。

关键词: 客观分析 地区差异 资料不连续

引 言

资料预处理是否合理会直接影响分析结果的可信度,所以对观测台站资料的处理,一直受到各界关注。由于台站分布不均,建站时间不一致等许多因素造成资料处理存在一定困难,但台站资料又是许多科学研究中必不可少的,所以正确处理和合理应用台站资料显得尤为重要。自从 Mitchell^[1](1961 年)做了开创性的工作以来,许多科研工作者根据不同资料用不同方法建立了半球和全球平均地面温度序列。Hansen^[2](1987 年)把地球表面分为面积相等的 80 个区,每个区又分为 100 个小区,先建立小区序列,再合为大区序列。建立小区序列时,先用最长的序列,再把短的序列逐步加进来,这样做是为了充分利用所有观测资料。Jones^[3](1988 年)用插值法计算每 5°(纬度)×10°(经度)上的气温值,再求全球及半球平均。Vinnikov^[4](1990 年)采用客观分析方法,计算格点值,再求平均得到了北半球最早的序列。IPCC 半球及全球平均地面温度序列是在 Jones^[3]等序列的基础上加以改进的,主要是使用 COADS 资料(综合海洋大气数据集)对海表温度做了订正,而 COADS 资料也是 2°×2°的区域化资料。由此可见对台站观测资料做区域化处理或客观分析是非常重要的,在气象、海洋等领域广泛应用。没有区域化处理或客观分析就得不到描述一个区域(小至一个地区,大至全球)整体特征的时间序列。中国许多气候研究工作中也都采用了这些方法^[5-7]。Barnes^[8]方法是目前最普遍使用的客观分析方法,本文在做青藏高原气候变化研究中采用此方法将高原台站资料计算到格点上,再求高原平均,得到了高原平均气温、地温、气压等序列,但作者却发现了一个 Barnes 方法和区域化方法

* 本文受 973 项目(GI998040907)和 863 项目(2002 AAI 35360)的联合资助。
2003-01-02 收到,2003-06-17 收到修改稿。

在实际应用中出现的问题。

1 计算方法及资料

为了得到描述一个区域整体特征的时间序列,最常用的有“Barnes 方法”和“区域化再求平均值”方法。

方法 1:“Barnes 方法”的核心是将分布不规则的测站资料计算到规则格点上。该方法在计算格点值时采用“与格点距离(平方)成反比做为权重,求扫描半径范围内测站值的加权平均,做为该格点值。得到了各个格点值,也就得到了一个规则分布的要素场,有了规则分布的要素场,再求区域平均等就很容易。

方法 2:“区域化再求平均”方法的核心是将大区域划分成规则的一个个小区域,用小区域内站点的平均值做为该小区域的值,将小区域的值再做面积加权合为大区域的平均值。

本文取青藏高原地区(包括青海、西藏以及云南、贵州、四川、甘肃、新疆临近高原的一些地区)120 个观测台站 1961~1998 年单站月平均气候资料,这些台站的位置分布见图 1。月平均气候资料

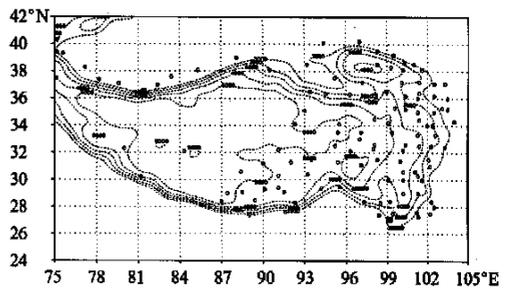


图 1 本文所用青藏高原 120 个观测台站位置分布
(圆圈为测站位置,虚线为海拔高度等值线)

包括温度、最高温度、最低温度、降水量、积雪日数、气压、云量、地温等共 14 个气象要素,对这些要素分别求出单站年平均值,再用 Barnes 方法将其插值到格点上,再求高原全区平均值,这样一来可以得到 14 个表征高原总体年季变化特征的时间序列。用“区域化再求平均”方法也可以得到表征高原总体年季变化特征的时间序列。

2 问题的提出

2.1 问题的发现

在这 14 个时间序列中,气温(图 2a)、最高气温(图略)、最低气温(图略)序列均有明显的增温趋势,而 0 cm 地温序列却表现为下降趋势(见图 2b)。0 cm 地温和气温真有这么大差别吗? 0 cm 地温序列的这种下降趋势是否可信呢? 如果可信,怎样解释地-气之间的湍流交换呢? 为了回答这些问题,本文对各个台站的地温资料和气温资料的连续性情况进行了分析,发现气温资料的连续性很好,而地温资料的连续性很差,图 2c 给出了地温的有效起始年分布(本文定义第一次出现连续 3 年有观测值的年为有效起始年),可见高原中东部大部分台站开始地温观测的时间都较晚,青南高原最明显。这些台站海拔高度相对较高,大多在 4000 m 以上,比那些较早开始地温观测的台站海拔高度明显偏高。高海拔测站的 0 cm 地温会明显低于低海拔测站的 0 cm 地温,由此可以看出在用 Barnes 方法做客观分析时,随着短序列不断引入(而这些短序列中海拔高度较高的台站占了较大

比例),也就是说相对较低的地温被不断引入客观分析中,所以致使每一个客观分析场都有不同的母场,利用这些母场将台站资料插值到格点上,求出逐年高原平均值,再将这些平均值看做一个时间序列时,必然导致序列出现下降趋势。而气温序列由于绝大部分台站1961年开始都有观测,其连续性较好,所以不存在上述问题,因而认为气温序列的增温趋势是可信的,地温序列的下降趋势完全是由于资料长短不一,台站分布地区差异较大造成的。

图2中的虚线是应用“区域化方法”得到的高原平均气温和地温序列,可以看出与应用“Baners方法”得到的实线非常近似,所以下文仅给出Baners方法得到的曲线。

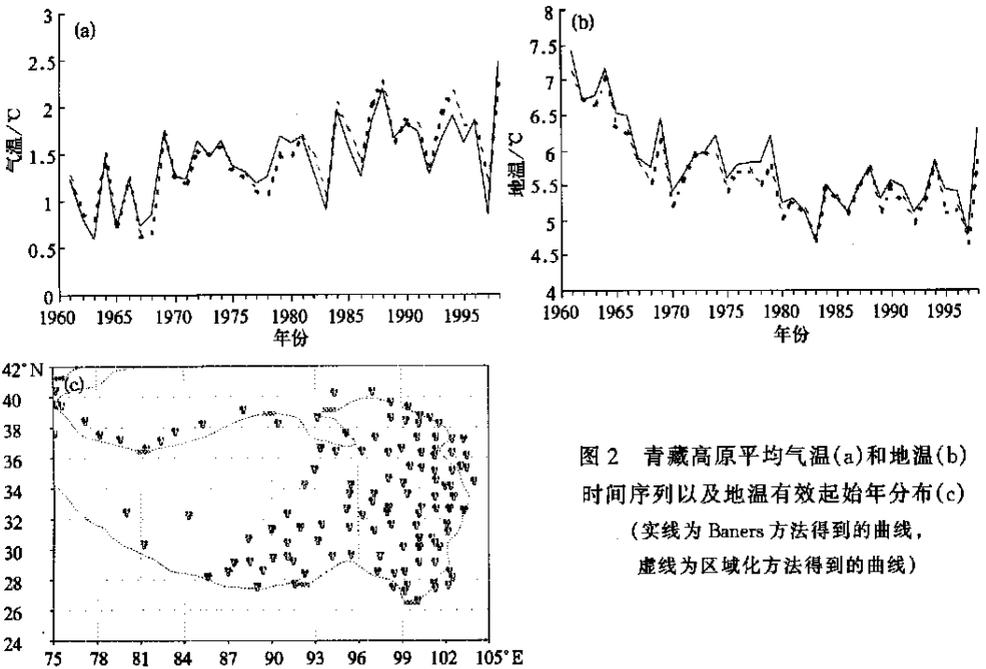


图2 青藏高原平均气温(a)和地温(b)时间序列以及地温有效起始年分布(c) (实线为Baners方法得到的曲线,虚线为区域化方法得到的曲线)

2.2 问题的深入研究

对于2.1节作出的推论,需要做进一步的验证。为了做到这一点,本文将气温观测值和地温观测值进行“逐站对比”,对于同一台站若在某一年地温缺测,则将气温也设为缺测,否则气温取原气温观测值,这样可以得到各个台站缺测情况与地温缺测一致的气温序列(本文称这一过程为缺测化处理),利用缺测化处理后的各站气温序列做客观分析,得到格点值,再做高原平均,得到一个新的时间序列(见图3),可以看出新的气温时间序列呈下降趋势,而从图2a可知气温时间序列原本呈上升趋势(图2a是未做缺测化处理的原气温序列,由于其资

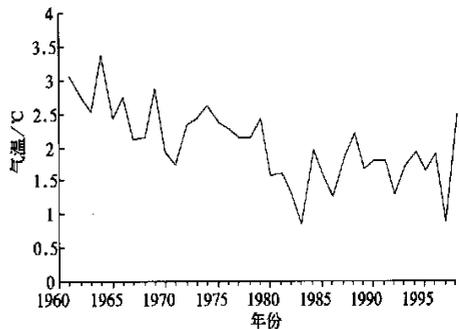


图3 缺测处理后应用Baners方法再求平均得到的气温序列

料长短较一致,连续性好,所以认为是可信的)。由此可见长短不一,缺测较多的气温序列(经缺测化处理),用 Barnes 方法做客观分析再求平均,得到了与原事实并不相符的结论,也就是说,不是什么资料拿来都可以做客观分析再求平均的。因此在做客观分析时,应对台站的分布情况、资料的连续情况首先有个了解,对于那些长短较一致,连续性好的资料,其客观分析再求平均的结论是可信的;对于那些长短不一致,连续性差的资料,其客观分析再求平均的结论不一定可信,尤其是那些台站分布地区差异较大的资料更应注意。

将气压等连续性较好的资料也与地温进行逐站对比,做缺测处理,得到各个台站缺测情况与地温缺测一致的气压序列。利用缺测化处理后的各站气压序列做客观分析,得到格点值,再做高原平均,同样可得到一个与原气压时间序列完全不一致的新序列(图略)。这又一次验证了 2.1 节得出的推论。

同样,用“区域化”方法重复做以上步骤,也可得出与“Barnes”方法一样的结论(图略)。可见,台站资料的结构会直接影响客观分析求平均和区域化求平均的结果。

3 问题的初步解决

由上文讨论可知,对于那些长短不一或缺测较多的资料,在求较大区域平均值时存在一定的困难。这些资料如果不能充分利用,无疑是一种遗憾和浪费。怎样合理利用这些资料呢?由前面的讨论可以知道,这些问题主要是由于地温观测的起始年不一致,资料连续性差,台站分布地区差异较大(如海拔高度差异较大)造成的。“起始年不一致,连续性差”已是事实,无法改变,如果消除“地区差异”即对观测资料做标准化处理,结果会怎样呢?

以资料长短较一致,连续性好的“气温”资料为例,首先对各站气温观测资料做标准化处理,再用 Barnes 方法将其插值到格点上,再求高原全区平均值,得到图 4a 中的实线,可以看出这条实线与图 2a 中的实线几乎完全一致。再将缺测处理后的各站气温序列(同 2.2 节)做标准化处理,得到各个台站缺测情况与地温缺测一致的标准化气温序列。利用这些序列做客观分析,得到格点值,再做高原平均,得到图 4a 中的虚线,它与图 4a 中的实线很一致。而虚线是在缺测较严重情况下得出的,可见“标准化”基本克服了观测的起始

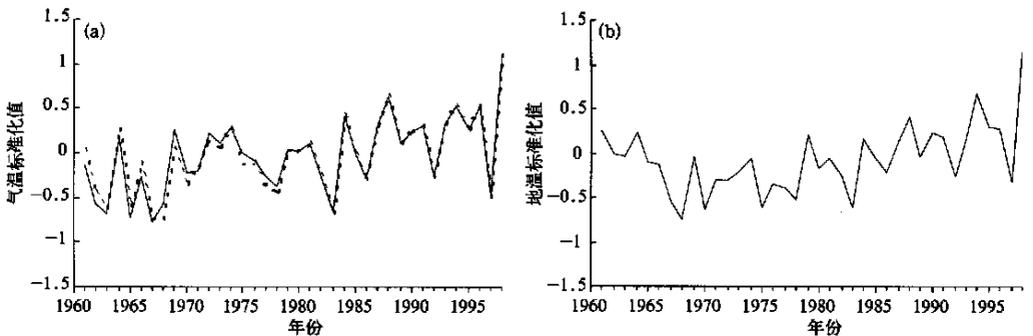


图 4 (a) 实线为先将各站气温资料标准化,再用 Barnes 方法得到的高原平均气温序列,虚线为将缺测处理后的气温序列标准化,再用 Barnes 方法得到的高原平均气温序列,
(b) 为先将各站地温资料标准化,再用 Barnes 方法得到的高原平均地温序列

年不一致,资料连续性差,台站分布地区差异较大给客观分析再求平均造成的困难。

鉴于以上分析,对各站“地温”观测资料做标准化处理,再用 Barnes 方法将其插值到格点上,再求高原全区平均值,得到图 4b 中的实线,可见地温其实是从 1960 年代末开始逐步回升,1980 年代中期出现迅速增温趋势,这一结论与图 2b 很不一致,而由以上分析知道图 4b 的结论较图 2b 可信。

为了使以上论述更为清楚明了,本文举一个简单的例子来进行说明。表 1 中给出了 3 个带有“数值增大线性倾向”的序列,这 3 个序列是任意给出的,但具有不同的均值。图 5a 是这 3 个序列的曲线图,图 5b 是其平均值序列的曲线图。可以看出各个序列及其平均值序列均为线性增大倾向。表 1 给出的 3 个序列都有完整的记录,如果假设“序列 C”前 3 次缺测,从第 4 次开始才有观测,得到表 2 中的 A、B、C 三个序列。做客观分析求平均时,短序列是被不断引入的,即前 3 次的平均值序列是“由 A 和 B 两个序列”求平均得到,从第 4 次开始,平均值序列是“由 A 和 B 和 C 三个序列”求平均得到(假设这 3 个序列

表 1 任意给出的 3 个具有“数值增大线性倾向”的序列

	1	2	3	4	5	6	7	8	9	10	11	12	...
序列 A	5.0	5.1	5.3	5.2	5.6	5.7	5.9	6.0	6.3	6.0	6.4	6.5	...
序列 B	3.2	3.3	3.0	3.6	3.5	3.7	3.8	3.9	3.7	4.0	4.2	3.8	...
序列 C	1.3	1.5	1.6	2.0	1.7	1.5	1.8	2.0	2.3	2.4	2.6	2.5	...
平均值	3.2	3.3	3.3	3.6	3.6	3.6	3.8	4.0	4.1	4.1	4.4	4.3	

表 2 在假设表 1 中的“序列 C”前 3 次缺测情况下得到的新序列

	1	2	3	4	5	6	7	8	9	10	11	12	...
序列 A	5.0	5.1	5.3	5.2	5.6	5.7	5.9	6.0	6.3	6.0	6.4	6.5	...
序列 B	3.2	3.3	3.0	3.6	3.5	3.7	3.8	3.9	3.7	4.0	4.2	3.8	...
序列 C				2.0	1.7	1.5	1.8	2.0	2.3	2.4	2.6	2.5	...
平均值	4.1	4.2	4.2	3.6	3.6	3.6	3.8	4.0	4.1	4.1	4.4	4.3	

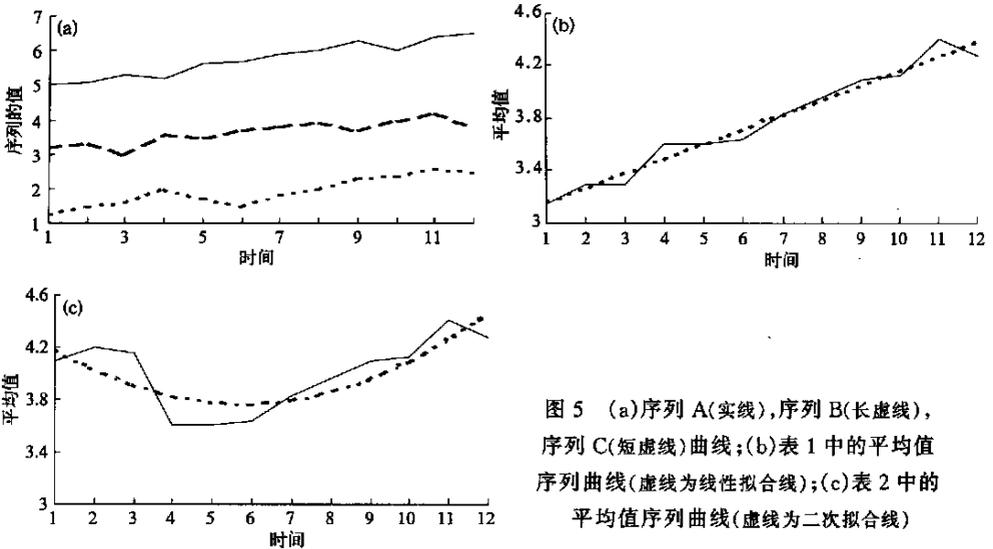


图 5 (a)序列 A(实线),序列 B(长虚线),序列 C(短虚线)曲线;(b)表 1 中的平均值序列曲线(虚线为线性拟合线);(c)表 2 中的平均值序列曲线(虚线为二次拟合线)

所占的权重都一样),这样得到表 2 中的平均值序列。用表 2 的平均值序列做图,得到图 5c,可以看出这条平均值曲线与图 5b 中的平均值曲线差别很大,出现了明显的“减小再增大”倾向,而 A、B、C 三个序列中没有一个具有这种倾向,因此,这种倾向是不真实的。造成这一现象出现的原因从表 2 中很明显可以看出:即序列 C 的均值明显低于序列 A 或 B 的均值,前 3 次求平均时,由于序列 C 没有观测值,平均值序列是“由均值较高的 A 和 B 两个序列”求平均得到的,从第 4 次开始,平均值序列是“由 A 和 B 和 C 三个序列”求平均得到的,也就是说从第 4 次开始,均值明显偏低的 C 序列参加了平均,所以在第 3 和第 4 个平均值之间造成了显著的下降,这一下降完全是由于前后样本量及样本特征不一样造成的,而并非这 3 个序列本身具有的特征。

如果细想“客观分析求平均”的计算过程,可以知道与这里的 A、B、C 三个序列求平均有很相似的地方。如果把 A、B、C 设想为高原上的“3 类”地温观测站,就很容易理解前面对地温序列的分析。序列 C 相当于那些海拔高度较高,地温观测开始较晚的台站,序列 A 和 B 相当于那些海拔高度居中或较低,地温观测开始较早的台站。假设把高原划分成 $1^\circ \times 1^\circ$ 的网格点,对应于中低海拔地区的网格点的地温值可由其周围的台站按距离的远近作为权重,求平均得到,这些网格点的地温值应具有“相对较高的均值”。对应于较高海拔地区的网格点的地温值也可由其周围的台站按距离的远近作为权重,求平均得到,这些网格点的地温值应具有“相对较低的均值”。再将这些网格点的地温值求平均得到高原的区域平均。20 世纪 60 年代初,高原上只有 A、B 类台站有地温观测,经客观分析后与其对应的网格点可以得到地温值,C 类台站没有观测,在一定扫描半径范围内(高原地形复杂,扫描半径不宜取得过大),某些与 C 类台站对应的网格点得不到地温值,这种情况类似于表 2 的前 3 个记录。后来,高原地温观测逐渐增多,尤其是高海拔地区增加了不少地温观测,即 C 类台站逐渐有了地温观测,经客观分析后与高、中、低海拔地区对应的网格点均可以得到地温值,这种情况类似于表 2 的后 9 个记录,区别在于实际地温计算中短序列被不断引入,而表 2 中仅仅引入了一个短序列。得到了不同时间的格点值(插值场)后,在对格点求平均得到高原区域平均值时,由于 60 年代初以均值较高的格点为主,以后不断有均值较低的格点也参加了平均,就造成了一个较大的下降趋势,这与表 2 的问题非常类似,只是比表 2 的计算更复杂。不同时间参加平均的格点不一样,就造成了平均值的偏移,无疑这种平均值序列不能真实反映高原地温的实际变化情况,这就犹如表 2 的平均值序列不能真实反映 A、B、C 三个序列的实际变化情况一样。

因此,利用这三幅图和两张表很简明地说明了“客观分析再求平均”中存在的问题。

现在将表 2 中的 A、B、C 序列分别做标准化处理,可对应得到表 3 中的 A、B、C 序列。再对表 3 中的 A、B、C 序列求平均,得到表 3 中的平均值序列(注:表 3 中的平均值序列不

表 3 将表 1 中的序列标准化后得到的新序列

	1	2	3	4	5	6	7	8	9	10	11	12
序列 A	-1.35	-1.25	-0.86	-1.06	-0.29	-0.10	0.29	0.48	1.06	0.48	1.25	1.45
序列 B	-1.28	-0.99	-1.86	-0.12	-0.41	0.17	0.46	0.75	0.17	1.04	1.62	0.46
序列 C				-0.24	-1.02	-1.54	-0.76	-0.24	0.55	0.81	1.34	1.10
平均值	-1.32	-1.12	-1.36	-0.47	-0.57	-0.49	0.00	0.33	0.59	0.78	1.40	1.00

是由表 2 中的平均值序列标准化而来), 图 6 为该序列的曲线图, 可见其变化与图 5b 变化基本一致, 而与图 5c 变化不一致。这说明: “标准化处理”在一定程度上克服了“长短不一, 均值差异较大”的资料在做客观分析求平均时存在的困难。众所周知, 标准化处理不会改变序列的变化规律和趋势, 即经标准化处理后 A、B、C 序列的变化规律和趋势都不会发生根本变化, 但是平均值序列的变化规律和趋势却发生了根本变化(对比图 5c 和图 6), 这主要是由于标准化处理后的序列都具有“均值为 0, 方差为 1”的特点(即消除了地区差异), 其中某一序列缺测, 不会造成“平均值序列的偏移”, 从而基本克服了表 2 中出现的问题。通过以上分析, 不难看出在求高原平均地温序列时, 如果首先将单站地温资料标准化, 其次通过客观分析将站点资料(已标准化)插值到格点上, 再求区域平均值序列, 那么即使有部分资料缺测, 也不会造成平均值序列的偏移, 从而得到较为可信的平均值序列。

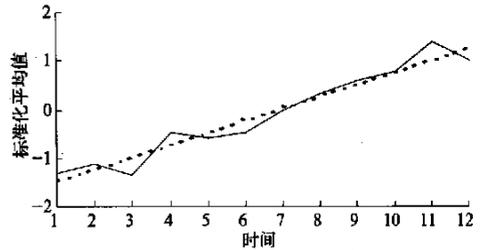


图 6 表 3 中的平均值序列曲线

4 气温和地温的线性趋势分布

图 7a 和图 7b 分别给出了地温和气温的线性趋势分布, 大于零表示该站总体呈增温趋势, 小于零表示该站总体呈降温趋势。从图 7a 可以看出, 仅在川、青、藏交界地区存在较大面积的地温下降区, 高原大部分地区以地温增加为主, 这就从各个台站的地温变化情况上证明了图 4b 的结论较图 2b 可信。图 7b 表明, 高原绝大部分地区以气温增加为主, 这也与图 2a 和图 4a 的结论一致。

当然, 资料缺测造成的信息流失是无法完全克服和弥补的, 只能尽量做到最大可能地合理利用这些宝贵资料, 这也是本文研究的目的。

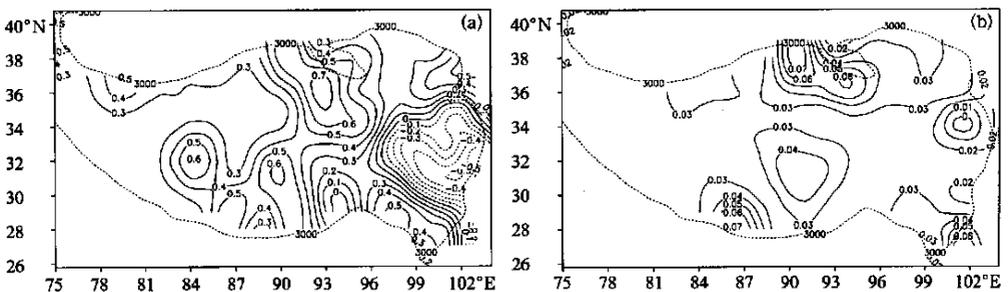


图 7 各站 0 cm 地温(a)、气温(b)的线性趋势分布

5 结论

(1) 对于那些长短不一, 缺测较多的台站观测资料, 如果直接用“Baners 方法做客观

分析再求平均”或用“先区域化,再面积加权求平均”的方法,都有可能得到与原事实并不相符的结论。因此在做客观分析时,应对台站的分布情况、资料的连续情况首先有个了解。对于那些长短较一致、连续性好的资料,其客观分析再求平均的结论是可信的;对于那些长短不一致、连续性差的资料,其客观分析再求平均的结论不一定可信,尤其是那些台站分布地区差异较大的资料更应注意。

(2) 对台站观测资料进行标准化处理,在一定程度上克服了观测的起始年不一致,资料连续性差,台站分布地区差异较大给“客观分析再求平均”和“区域化求平均”造成的困难,可以较合理地利用这些宝贵资料。

参 考 文 献

- 1 Mitchell J M. Recent secular changes of global temperature. *Ann. N. Y. Acad. Sci.*, 1961, **95**: 235 ~ 250.
- 2 Hansen J, Lebedeff S. Global trend of measured surface air temperature. *J. Geophys. Res.*, 1987, **92**(D11): 13345 ~ 13372.
- 3 Jones P D. The influence of ENSO on global temperatures. *Climate Monitor*, 1988, **17**: 80 ~ 89.
- 4 Vinnikov K Ya, et al. Empirical on contemporary global climate changes (temperature and precipitation). *J. Climate*, 1990, **3**: 662 ~ 677.
- 5 严中伟,季劲钧,叶笃正. 60年代北半球夏季气候跃变. *中国科学 B辑*, 1990, **1**: 97 ~ 103.
- 6 王绍武,叶瑾琳. 近百年全球气候变暖的分析. *大气科学*, 1995, **19**(5): 545 ~ 553.
- 7 翟盘茂,任福民. 中国近四十年最高、最低温度变化. *气象学报*, 1997, **55**(4): 418 ~ 429.
- 8 Barnes S L. A technique for maximizing details in numerical weather map analysis. *J. Appl. Met.*, **3**: 396 ~ 409.

A QUESTION WHEN USING OBSERVATION DATA TO GET REGIONAL AVERAGED VALUE

Niu Tao Liu Hongli Chen Longxun

(*Chinese Academy of Meteorological Sciences, Beijing 100081*)

Abstract

In order to get a time series that represent the total characteristics of some regions, Objective analysis and regional-averaged method are widely used to handle observation data. Barnes objective analysis method is a convergent weighted-averaging interpolation scheme which is most generally used in meteorological and ocean field. The characteristics of the elements such as soil temperature, air temperature (include max and min temperature), air pressure etc. on the Tibetan-Qinghai Plateau were analyzed by using these two methods. It is found that when data are not good in continuity, the influence caused by the different region of observe station must be removed in order to get a reliable conclusion, otherwise an erroneous conclusion is got, so this question should be pay attention.

Key words: Objective analysis method Territory difference Data discontinuity