

区域平均温度计算的最优权重方法*

—— I. 理论分析

S.S.P. Shen

(Dept. of Math., Canada Univ. of Alberta)

王晓春 梁幼林 李荣凤

(中国科学院大气物理研究所, 北京 100080)

关键词 最优权重方法 区域平均温度 算术区域平均

资料空间覆盖不均匀给全球(或半球)及区域平均温度的计算造成了很大困难. 在求全球(或半球)平均温度时, Jones^[1,2], Hansen^[3], Vinnikov^[4]等分别用了不同的方法来解决这一问题. Vinnikov等用最优权重方法, 即在误差最小的前提下, 求出不同纬度带上每一站的权重, 而后再用加权平均方法来计算全球(或半球)平均温度. Shen等^[5]在用最优权重方法求全球平均温度方面又做了一些工作, 结果表明, 全球只用60多个分布合适的台站便可得到精度甚高的平均温度.

本文将分析资料空间覆盖不均匀对区域平均温度计算的影响. 首先推导区域平均温度计算的最优权重方法, 随后分析最优权重方法的几种特殊情形, 限于篇幅, 第三及第四节给出部分计算结果及结论. 最优权重方法与一般算术区域平均方法的详细比较, 我们将另文给出.

1 理论推导

将区域 C 内任一点的位置记为矢量 r , r 点上 t 时刻的温度记为 $\Theta(r, t)$, r 点上 t 时刻 τ 时间尺度温度的平均记为 $\Theta_\tau(r, t)$, 依定义有

$$\Theta_\tau(r, t) = \frac{1}{\tau} \int_{t-\frac{\tau}{2}}^{t+\frac{\tau}{2}} \Theta(r, t') dt' \quad (1)$$

在实际问题中 τ 可为月、季或年.

理想情形下, 对于空间上连续分布的温度场在区域 C 内相应时间尺度 τ 的平均可表示为如下积分

$$\Theta(t) = \frac{1}{A} \int_C d\Omega \Theta_\tau(r, t) \quad (2)$$

其中 A 为区域 C 的面积, $d\Omega$ 为面积元. 另外要注意上式与时间尺度 τ 有关.

在实际问题中, 需要根据有限的台站资料来估计 $\Theta(t)$, 即

1994-07-30 收稿, 1994-12-08 收修改稿

* 中国科学院《灾害气候预测及其对农业年景和水资源调配的影响》资助项目

$$\hat{\Theta}(t) = \sum_{i=1}^n w_i \Theta_i(r_i, t) \quad (3)$$

或

$$\hat{\Theta}(t) = \frac{1}{A} \int_C d\Omega W \Theta_\tau(r, t), \quad (4)$$

而

$$W = A \sum_{i=1}^n w_i \delta(r - r_i), \quad (5)$$

其中 n 为区域内台站数量, r_i 为站点 i 的位置, $\Theta_i(r_i, t)$ 为 i 站点 t 时刻 τ 时间尺度上温度的平均值, w_i 为 i 站点的权重, $\delta(r - r_i)$ 为 δ 函数. 并且 w_i 满足

$$\sum_{i=1}^n w_i = 1. \quad (6)$$

用 $\hat{\Theta}(t)$ 代替 $\Theta(t)$ 所产生的误差为

$$\varepsilon^2 = \langle (\Theta(t) - \hat{\Theta}(t))^2 \rangle, \quad (7)$$

其中 $\langle \cdot \rangle$ 为总体平均. 将上式展开, 得

$$\varepsilon^2 = \frac{1}{A^2} \int_C d\Omega \int_C d\Omega' \rho(r, r') - \frac{2}{A} \sum_{i=1}^n w_i \int_C \rho(r, r_i) d\Omega + \sum_{i,j=1}^n w_i w_j \rho(r_i, r_j), \quad (8)$$

其中

$$\rho(r_i, r_j) = \langle \Theta_i(r_i, t) \Theta_j(r_j, t) \rangle, \quad (9)$$

在这个表达式中, 我们假设 $\Theta_\tau(r_i, t)$ 是平稳序列.

我们的问题是找一组权重 $w_i, i=1, \dots, n$, 使在满足 (6) 的条件下 ε^2 达到最小. 这样的一组权重在本文中称为最优权重.

利用拉格朗日乘子法, 为以后推导方便取乘子为 -2Λ , 记

$$F = \varepsilon^2 - 2\Lambda \left(\sum_{i=1}^n w_i - 1 \right). \quad (10)$$

类似于 Shen 等^[9]、Vinnikov 等^[4] 的处理方法, 由

$$\frac{\partial F}{\partial w_i} = 0 \quad \text{和} \quad \frac{\partial F}{\partial \Lambda} = 0, \quad (11)$$

则最优权重实际上是下面方程组的解

$$\begin{cases} \sum_{j=1}^n w_j \rho(r_i, r_j) - \Lambda = \bar{\rho}(r_i) & i=1, \dots, n, \\ \sum_{i=1}^n w_i = 1, \end{cases} \quad (12)$$

其中

$$\bar{\rho}(r_i) = \frac{1}{A} \int_C d\Omega \rho(r, r_i). \quad (13)$$

若记矩阵 $(\rho_{ij} = \rho(r_i, r_j))$ 的逆矩阵为 (b_{ij}) , 则方程组 (12) 的解为

$$\begin{cases} w_i = \sum_{j=1}^n b_{ij}(\Lambda + \bar{\rho}(r_j)), \\ \Lambda = \frac{1 - \sum_{i=1}^n \sum_{j=1}^n b_{ij} \bar{\rho}(r_j)}{\sum_{i=1}^n \sum_{j=1}^n b_{ij}}. \end{cases} \quad (14)$$

当将温度资料标准化后, $\bar{\rho}(r_i, r_j)$ 则为两站间的相关系数. 因而 $\rho(r_i, r_j)$ 可视为 i 站与 j 站联系的紧密程度的一种度量, 由 $\bar{\rho}(r_i)$ 的表达式(13)可以看出, $\bar{\rho}(r_i)$ 表示了 i 站点在求区域 C 的平均温度时的重要程度.

考虑到要解决的问题是求出误差最小的区域平均温度, 而 $\bar{\rho}(r_i)$ 本身也以区域积分的形式表达, 那么如何求得精确的 $\bar{\rho}(r_i)$ 呢? 要解决的问题化成了一个同类性质的新问题.

2 讨论

上述推导为求取误差最小的区域平均温度指出了一条理论上可行的途径. 但若 $\bar{\rho}(r_i)$ 计算的不好, 仍然得不到好的效果. 以下分析几种特殊的情形来揭示最优权重解对 $(\bar{\rho}(r_i))$ 计算方法的依赖.

(1) 用实际相关系数场来计算 $\bar{\rho}(r_i)$. 若只用 n 个台站间的相关系数场来计算 $\bar{\rho}(r_i)$, 即将式(13)离散化, 可得

$$\bar{\rho}(r_i) = \frac{1}{A} \int_C d\Omega \rho(r, r_i) = \sum_{j=1}^n \frac{\Delta A_j}{A} \rho(r_j, r_i), \quad (15)$$

这里 ΔA_j 为 j 站点所能代表的面积. 由 (12) 可直接得到最优权重解为

$$\begin{cases} w_i = \frac{\Delta A_i}{A}, \\ \Lambda = 0, \end{cases} \quad (16)$$

这相当于以面积比(站点代表面积与总面积之比)为权重. 特殊情形下, 若假定各站代表面积都相等, 则最优权重解为

$$\begin{cases} w_i = \frac{1}{n}, \\ \Lambda = 0, \end{cases} \quad (17)$$

这时最优权重区域平均便退化为一般的算术区域平均.

(2) 用预先指定的相关系数场来计算 $\bar{\rho}(r_i)$. 若能对相关系数场预先指定一个理论上的分布, 则 $\bar{\rho}(r_i)$ 的计算可能会更加精确. 在分析全球平均温度时, Vinnikov 等人每隔 30 纬度指定了相关系数随距离的分布, 对于 $30^\circ \sim 60^\circ \text{N}$, 他所指定的表达式为

$$\rho(s) = \exp(-0.21s^{0.893})J_0(0.852s), \quad (18)$$

其中 ρ 为相关系数, s 为以 10^3 km 为单位的站间距离, J_0 为 Bessel 函数. Vinnikov 等的方法在求全球平均温度时得到了好的效果. 但温度相关系数随距离的分布随地理位置不同有很大差别, 并且当时间尺度不同时, 这一关系还要发生变化^[9].

当考虑区域平均温度时, 我们的计算结果表明, 相关系数随距离的变化基本满足 e 指数距离平方衰减的规律. 在本文的试算中, 我们假定以 r_i 为中心的相关系数场 $\rho(r, r_i)$ 满足

$$\rho(r, r_i) = a_i \exp\left(-\frac{|r-r_i|^2}{d_i^2}\right), \quad (19)$$

其中 a_i 及 d_i^2 为与台站 i 有关的常数.

按照如上假定,先根据实际的相关系数场及站间距离计算每一站的 a_i 及 d_i^2 ,然后在给定区域内计算 $\bar{\rho}(r)$,最后由式(14)便可得到最优权重及相应的区域平均温度.

当然,比之一般的算术区域平均方法,最优权重方法在计算上复杂一些,但考虑到可以用大量的近期观测较为精确地确定以每一站为中心相关系数场的分布参数,而这些参数可以用来得到观测台站较少时期的区域平均问题,这种计算上的复杂性还是值得的.另外要指出的是这些参数依赖时间尺度 τ .

3 试算结果

我们用中国东北区 23 站 1961~1990 年 30 年的温度资料对最优权重方法进行了试算.计算步骤为:首先将 23 站月平均温度资料减去多年月平均以去掉季节变化,然后将温度距平资料标准化,求出相关系数矩阵,而后用最小二乘法估计每一站的 a_i 及 d_i^2 ,根据以上讨论的方法来计算 $\bar{\rho}(r)$ 及每一站的权重.最后比较了最优权重区域平均温度距平(OTM)及算术区域温度距平(TM).这里的距平指标标准化距平.

图 1 给出月温度资料相关系数随距离的分布,从图中可以注意到,这时相关系数随距离的分布与 e 指数距离平方衰减的关系较为接近,而与 Vinnikov 等所指定的关系(式 18)则相差较大.

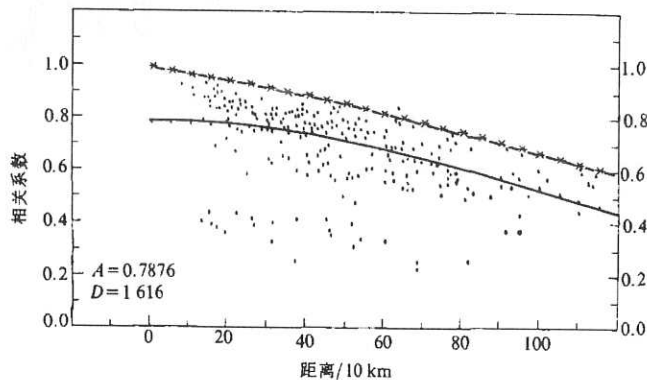


图 1 相关系数随距离的分布(黑点)

图中两条曲线分别为相关系数随距离指数平方衰减模型及 Vinnikov 模型.——代表 $A \exp\left(-\frac{r^2}{D}\right)$, $A = \frac{1}{n} \sum_{i=1}^n a_i$,

$D = \frac{1}{n} \sum_{i=1}^n d_i^2$, r 为站间距离 ($A=0.7876$, $D=1616 \text{ km}^2$), a_i , d_i^2 的意义详见本文; * 代表 $\exp(-0.21s^{0.893})J_0(0.852s)$,

s 为以 10^3 km 为单位的站间距离, J_0 为 Bessel 函数

为清楚起见,图 2 下部给出了最优权重区域平均温度距平与算术区域平均温度距平的逐年变化,上部给出了两者之差的逐月变化.从图 2 可以看出两者非常一致,事实上两者的均方差别

$$\text{RMSD} = \left(\frac{1}{m} \sum_{i=1}^m (\text{OTM}_i - \text{TM}_i)^2 \right)^{1/2}$$

(其中 m 为总月数, OTM_i , TM_i 分别为第 i 月的 OTM 及 TM) 为 TM 均方差的 9.0%. 上述的比较表明,最优权重方法若用本文所设计的算法可得到合理的区域平均.

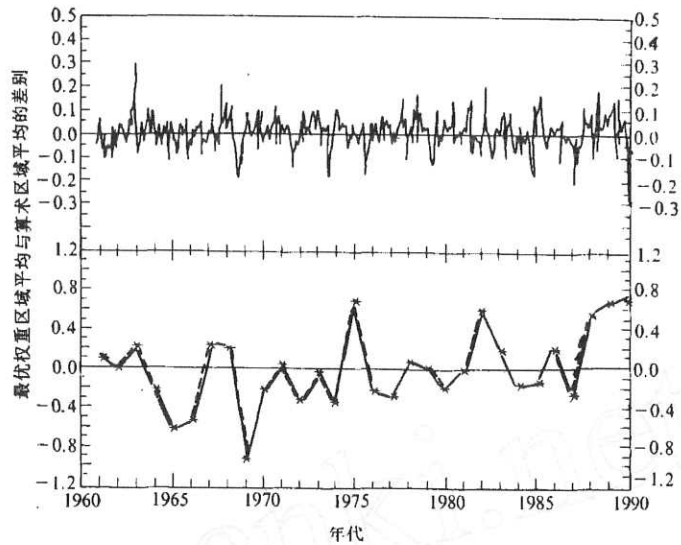


图 2 最优权重区域平均温度距平、算术区域平均温度距平的逐年变化(下部图中虚线及实线)及两者之差的逐月变化(上部图中实线)

本文推导了计算区域平均温度的最优权重方法,对推导结果的分析表明,若 $\bar{\rho}(r_i)$ 计算得好,则可得到误差小的区域平均。

另外,本文根据区域平均温度计算的特点设计了一种算法,并用我国东北区 23 站的温度资料进行了试算,结果表明,这种算法可以得到合理的区域平均。

致谢 胡增臻博士对初稿提出了许多宝贵的建议,特此表示衷心感谢。

参 考 文 献

- 1 Jones P D, Raper S C B, Bradley R S *et al.* Northern hemisphere surface air temperature variations: 1851~1984. *J Clim Appl Meteorol*, 1986, 25(2): 161~179
- 2 Jones P D, Raper S C B, Bradley R S *et al.* Southern hemisphere surface air temperature variations: 1851~1984. *J Clim Appl Meteorol*, 1986, 25(9): 1213~1228
- 3 Hansen J S, Lebedeff, global trends of measured surface air temperature. *J Geophys Res*, 1987, 92(D11): 13345~13372
- 4 Vinnikov K Ya, Groisman P Ya, Lugina K M. Empirical data on contemporary global climate changes (temperature and precipitation). *J Climate*, 1990, 3(6): 662~677
- 5 Shen S S P, North G R, Kim K Y. Spectral approach to optimal estimation of the global average temperature. *J Climate*, 1994 (in press)
- 6 Kim K Y, North G R. Surface temperature fluctuations in a stochastic climate model. *J Geophys Res*, 1991, 96(D10): 18573~18580