

大气边界层观测资料的质量控制 方案研究及其应用

陈 明 洪钟祥

(中国科学院大气物理研究所, 北京 100029)

提 要

利用北京 325m 气象观测塔的历年数字化观测资料, 本文设计了一种对边界层实测资料进行质量控制的方案. 该方案从气象学及气候学的一般特性出发, 利用逐点、分段、加权拟合法, 对单站资料的空间点列依次进行气候极值诊断、天气学诊断、异点拟合诊断、一致性拟合诊断, 从而达到判断和修正资料中异点的目的. 分析结果表明: 该方案不仅具有较强的诊断力, 而且在改变控制参数时, 也能应用到其它气象资料的质量控制中去, 这使该方案同时具有一定的普适性和实用价值.

关键词: 异点; 数字权重; 拟合精度.

一、引 言

在大气边界层特别是近地层的研究中, 目前也许还没有比气象观测塔更能提供丰富、翔实资料的探测手段. 仅就资料的连续性一项指标, 就很难用其它定时和应用观测相比拟. 因此, 研究塔层资料的质量控制问题, 对实现边界层实测资料的质量控制具有相当的代表性和实用价值.

长期以来, 特别是近二十几年来, 伴随着新技术的出现以及电子计算机的更新, 实验工作者使感应元件变得更加灵敏、精度更高. 然而, 利用塔层进行测量所带来的一些问题至今依然存在. 例如, 塔结构对气流的阻挡、恶劣天气使测量元件的性能变异等. 有时, 这些纯技术上的问题, 可以使资料的准确性降到令人难以置信的程度. 如果不修正或剔除这些异常值, 在随后的理论分析中必然会出现许多问题. 例如, 一个达到数字化装置最大值的异常值, 可使谱分析的整个噪声水平增大; 而两个相距很近的这种异常值将在谱分析中产生许多虚假的频率. 因此, 在对原始资料进行理论分析前, 有必要修正或剔除这些异常值, 这就是所谓的质量控制. 尽管质量控制工作是非常基础性的, 然而却是十分复杂的. 对于边界层资料, 这种复杂性尤为突出. 首先, 判据难定, 即难以选择一个精确的判断标准. 整个边界层物理学, 除了均匀、定常、中性的近地层有比较精确的半经验公式可以描述外, 其它天气形势和下垫面情况下的非定常边界层的整体研究, 目前依然停留在定性的天气气候描述和数值模拟上. 很难

1992年2月24日收到, 8月10日收到修改稿.

用统一的理论函数去描述边界层特性要素的时空分布,特别是其中的湍流结构特性。其次,边界层通常以湍流运动为特征,这种运动具有非常小的内尺度,重要的标量,如温度、湿度、痕量气体等都受湍流运动影响,因而也具有小尺度的脉动,还有一些中尺度的因子,如局地地形都可能对不同层次的数据产生不等量的偏差。在尺度谱如此宽的范围,要确定一个数据的正误是必需十分小心的。

对于塔层资料的质量控制,人们还将遇到的问题是,缺少精确的对比观测资料。空间对比资料的匮乏是可以想象的,然而,不断更变的下垫面又改变了时间序列的一致性(局地气候变化),而且,全球气候变迁也可能给资料一致性的检验带来许多不便。

当然,正如人们所熟知的那样:规律的产生是一种长期、反复、大量统计的结果,而边界层运动既然是大气环流的组成部分,那么,它必然满足大气运动的一般规律——气象学和气候学的。谱域如此宽的边界层运动,使我们有可能研究它在不同尺度中的一般特性。较为成熟的气象学和气候学的先进理论为这种研究提供了非常有益的帮助。在本文的质量控制方案中,正是应用了气象学和气候学中许多成熟的结果作为方案设计的依据,修正那些主要由于技术性问题,或者简单地称为系统原因所造成的异常值(我们称异常值对应的空间点为异点)。这种从一般特性出发,利用有效数学工具处理问题的方法,不仅避免了具体过程的复杂性,而且避免了由于理论不完备而造成的对实际问题的错误判断,从而使我们在保证资料原始性的基础上,最大限度地诊断出那些真正的异点。

二、北京气象塔资料的采集与提取

1. 塔和观测系统的简介

北京 325m 气象观测塔位于北纬 $39^{\circ}58'$, 东经 $116^{\circ}22'$, 海拔高度 49m。除正南面北三环路两侧有零星的超过 50m 的建筑外,以塔为中心约 1km 的范围内无高度超过 20m 的建筑物。

塔体采用通风良好的钢管格构式结构(通风系数 0.25),塔上共装 15 层观测平台,观测高度分别为 9, 15, 32, 49, 65, 80, 103, 120, 140, 160, 180, 200, 240, 280 和 320m。100m 以下各层基本按对数分布设置,这给近地层研究带来了很大方便。塔的 15 层观测平台上,分别按盛行风向装有二个活动伸臂,一个指向东南,一个指向西北。测量的基本要素包括:水平风向、风速、温度、相对湿度。塔上仪器检测到的所有信号,均通过屏蔽电缆传输到地面计算机房进行数据处理。

2. 资料的提取及初步整理

所有资料来源于 1986 年 5 月—1988 年 6 月份的 2 分钟存盘资料。采用 Basic 语言编程,随机文件形式存取,使用时非常简单便利。考虑到数学物理上的要求以及大气边界层研究和其它相关学科的需要,在进行质量控制前,我们对资料进行了 30 分钟正点前等权平均,其中风向资料的平均采用了气象上通用的 16 方位“最大频率风向

法”, 即对风向从 $0 \sim 360^\circ$ 划分为 16 个等角方位, 然后取频率最高的平均风向为实际风向, 平均后的资料仍然以随机文件的形式保存待用。

三、质量控制方案的设计及调试

1. 质量控制方案的数学物理方法

如果我们能知道一个物理量的时空函数, 即使是函数形式, 那么, 我们就能很快发现并剔除那些异常的实测数据。然而, 这种定量或定性的函数关系在自然中是很少的, 行星边界层中的非线性物理过程以及复杂的边界条件, 都使这种函数关系成为水中月。传统的数学拟合方法为寻找这种函数关系的近似表达——经验公式提供了可能。然而用拟合法求经验公式本身会遇到两个问题: (1) 当样本中出现异常值, 特别是当这种异常值大至歪曲整个系统的物理特性时, 这种拟合是没有任何意义的。(2) 拟合中的可取函数类通常为有理函数, 譬如多项式等, 如果拟合点数太多, 函数阶数太大时, 在拟合点之间势必会产生激烈振荡, 因而使拟合的精度大大降低。这两个问题使得用拟合经验公式来判断异常值的可能性几乎成为泡影。然而, 如果能利用“权重”这样一个数学参量去减少可能异点的信息量, 那么, 我们就有可能避开异常值的干扰, 从而得到精度较高的拟合公式。再者, 将样本点合理分组, 同时根据经验, 选用优化的可取函数类, 就能避开拟合振荡带来的影响。关于样本分组以及权重因子和可取函数类的合理选择, 我们将在下一小节作详细讨论, 这里先来分析一下加权的拟合问题。

给定样本 S :

$$S = (x_i, y_i), \quad i = 0, 1, 2, \dots, m, \quad (1)$$

要找连续变量 X 的一个函数 $y = g(x)$, 对样本 S 拟合得最好。所谓拟合的好坏, 是指向量 $\vec{\delta} = (\delta_0, \delta_1, \dots, \delta_m)^T$ 是大是小, 其中

$$\delta_i = g(x_i) - y_i, \quad i = 0, 1, 2, \dots, m. \quad (2)$$

一般常取 $\vec{\delta}$ 的范数 $\|\vec{\delta}\|_p$ ($p = 2, \infty, 1$) 作为 $\vec{\delta}$ 大小的衡量标准, 如果 $p = 2$, 就构成通常的最小二乘问题。

拟合中最重要的还有 $g(x)$ 的选择问题, 即在什么函数空间 Φ (可取函数类) 内进行比较的问题, 通常总是把含若干参量的形式函数作为可取函数类, 这样, 对于含 $n+1$ 个参数 $a_0, a_1, \dots, a_n$ 的族是

$$\sum_{j=0}^n a_j \varphi_j = a_0 \varphi_0(x) + a_1 \varphi_1(x) + \dots + a_n \varphi_n(x), \quad (3)$$

其中 $\varphi_j(x)$ 是给定的函数。

暂定 $g(x) = g(x; a_0, a_1, \dots, a_n)$ 是 Φ 中的任意一元, 现要选择 $a_0^*, a_1^*, \dots, a_n^*$ 使以 $g(x) = g(x; a_0^*, a_1^*, \dots, a_n^*)$ 作为拟合曲线时误差向量 $\|\vec{\delta}\|_p$ 的范数最小。

选择 $p = 2$, 设对于样本 S , 要用形如 (3) 式的函数按最小二乘原理挑选拟合函数,

$$f_{ij} = \sum_{i=0}^m P_i \varphi_i \varphi_{ij}, \quad (j=0, 1, 2, \dots, n)$$

$$g_l = \sum_{i=0}^m P_i y_i \varphi_{il}, \quad (l=0, 1, 2, \dots, n)$$

则(6)式变为

$$\sum_{j=0}^n a_j f_{ij} = g_l, \quad (l=0, 1, 2, \dots, n)$$

这是一个关于变量 $a_j, j=0, 1, \dots, n$ 的线性方程组, 通过 Gauss-Seidel 迭代法得到相应 a_j 值. 再由(3)和(2)式对应点的拟合值 $g(x_j)$ 以及偏差 δ_i , 实际应用中, 如果结合具体的物理过程, 给定这种拟合偏差的最大允许范围, 我们称之为误差极值标准, 那么, 偏差大于标准的点, 一般可定为异点, 相反为常点. 这种通过拟合而诊断出异点的方法, 我们称为拟合诊断法, 由于异点的“怪异”程度不一, 在拟合中引进的权重不一样, 因而又分异点拟合诊断和一致性拟合诊断.

2. 可取函数类的选择

上面给出了诊断异点的质量控制方程, 但并没有给出 $\varphi_j(x), j=0, 1, \dots, n$ 的具体形式, $\varphi_j(x)$ 的选择与具体的物理过程和经验有很大关系. 在对单站资料的空间点列进行五点一组的分组后, 我们选择二次多项式即 $\varphi_j(x) = x^j, 0 \leq j \leq 2$ 作为拟合行星边界层内廓线的可取函数类, 这不仅从简单实用的角度考虑, 更重要的是由大气运动本身的特性决定的. 一般情况下, 自由大气可以看作准地转和近绝热的, 即使考虑热成风和局地地形的影响, 实际风速与地转风的偏差也不会很大, 即

$$|\vec{U}| \approx |\vec{U}_g|, \quad |\Delta \vec{U}|/|\vec{U}_g| < O(1)$$

其中 $|\Delta \vec{U}|$ 是自由大气中上下两层风速矢量差特征值, $|\vec{U}_g|$ 是地转风. 在行星边界层内, 尽管受地面影响, 然而其上部的 Ekman 层内这种分析是不会引起太大偏差的. 就是存在低空急流时, 上面的关系也能近似成立. 在近地层, 由于受下垫面的影响较大, 各气象要素在垂直方向上呈较大的梯度分布, 然而湍流的混合效应又大大减弱了这种梯度分布, 梯度越大, 湍流交换越强, 各种要素越趋于一致, 而塔上设置平台时, 100m 以下的近地层是按对数分布设置的, 使得越近地面, 越可进行 Taylor 级数展开, 这就保证了风速分段拟合的数学合理性. 同时, 准绝热的条件又使温度在较大范围(除近地层)内呈近似线性分布, 这就避免了由于可取函数类的错误选择而造成的诊断失误(称误诊). 正因为如此, 用二次式拟合边界层廓线是不会出现太大问题的. 所有质量控制结果也证明了这种方法的可行性. 只是在大风和弱稳定时, 如果误差极值标准定得不合理也有误诊的情况. 但只要适时地修正标准, 那么就可以避免误诊或漏诊的发生.

3. 权重因子的确定

正如质量控制是一种带有相当程度经验化的问题一样, 权重因子的确定也同样必须针对具体情况才能合理地定出. 特别是在异点拟合诊断中, 权重因子即使不是在数字化

问题中也不可能是某个连续权重函数的逐点函数值，它只能是一个离散化的数列，称数字权重。然而，无论什么具体问题，都必须遵循以下两个原则：

(1) 必须使准确度越高的点权重越大，尽量减少可能异点的权重。

(2) 必须适时加大被修正后点的权重。

在设计质量控制方案时，我们将所有点都进行两次拟合诊断，两次拟合的目的是不一样的，我们分别称之为异点拟合诊断和一致性拟合诊断，表 1 给出了调试后的结果。

表 1 权重向量 $\vec{P} = (p_{i-2}, p_{i-1}, p_i, p_{i+1}, p_{i+2})$

权重向量 要素	类别	有连续异点	
	没有连续异点	第 i 点	第 $i+1$ 点
风速	(15, 20, 2, 1, 1)	(15, 20, 1, 1, 15)	(15, 20, 1, 15, 1)
温度	(15, 20, 2, 1, 1)	(15, 20, 1, 1, 15)	(15, 20, 1, 15, 1)
湿度	(30, 25, 2, 1, 1)	(30, 35, 1, 1, 30)	(30, 35, 1, 30, 1)

1) 异点拟合诊断

异点拟合诊断以诊断出不符合一般气候特征以及少数明显不合大尺度运动特性的点(大锯齿点)为目的。在这样的拟合中，由于异点的干扰非常大，如果权重选择不当，将因为拟合精度的问题而使“常点”被误诊。为了“孤立”异点，从而使异点以外的其它各点都落在拟合线附近，第一次拟合中，我们采用“灰箱”法：即不能确定准确度的点，不管正确与否，其权重相对降低，相反则加大权重。对北京 325 m 气象塔，我们具体的办法是：预先控制好最低两点的准确度，再在递进的拟合中使拟合点以下两点的权重较拟合点及其以上两点权重大。但相对权重不能太大，否则会造成仅由较低两点决定的准直线。一个简单的关系可作为决定相对权重的参考：

$$\frac{p_i^0}{p_i^1} = \frac{\delta_i^1}{\delta_i^0} \approx O(Q),$$

式中 p_i^0 , δ_i^0 为常点权重及偏差， p_i^1 , δ_i^1 为异点的权重及偏差， Q 为气象要素的特征值。

2) 一致性拟合诊断

为了使廓线最大限度地光滑，从而使廓线上下各点连续一致，有时有必要再加一个对称数字窗，对各点再进行一次拟合诊断，以便将异点拟合中未能诊断出的偏差虽小，但又不合连续性要求的“小锯齿”光滑一下。但这种光滑作用有时会将边界层内的一些特征层，如低空急流区、热力内边界层等光滑掉，因此，如果没有绝对根据，不要轻易作这一步光滑。我们的办法是：(1) 加大被诊断点权重(不能太大，否则诊断不出)；(2) 对有连续两点被定为“异点”时，将它们恢复至原始值，再从第 $i+3$ 点开始诊断，直到不出现连续异点为止。因为特征层附近各点特性相差较大，极容易都被误诊，用回复原始值的办法就可保留下来。

4. 诊断步骤及误差极值标准

质量控制的主要目的是为了修正或剔除异点并使数据之间尽可能地协调一致。因此在质量控制时，我们注意了以下几点：

- (1) 尽量对错误或缺测资料作订正，使资料在垂直方向保持完整性。
- (2) 尽量使修正或剔除的资料最少，保持资料的原始性。
- (3) 尽量使修正后的资料保持一致性，即使得廓线尽可能地光滑。

针对以上几点，我们按次序对资料进行了气候极值诊断、大尺度天气特性诊断、异点拟合诊断、一致性拟合诊断等四步诊断与修正。各步的误差极值标准见表 2。

表 2 误差极值标准

标准 要素	步骤	气候极值	大尺度天气特征	异点拟合		一致性拟合	
				$2 < i < 15$	$i \leq 2$	$2 < i < 15$	$i \leq 2$
风速 (m/s)		40	$ g_i < \frac{ g_{i-1} + g_{i+1} }{10}$	2	3	1.5	3
温度 (℃)		最低 -40， 最高 +55		2	3	1	3
湿度 (%)		最低 0， 最高 100		25	25	15	25

气候极值是根据历年气象资料统计出的各气象要素的最大值和最小值。如果一个数据不符合一般的气候统计特性，超出该地区的气候极值，一般可以定为异点。表 2 中的气候极值诊断标准参考了北京地区 1951 — 1980 年地面气候要素的平均与极端值。

大尺度天气特征诊断以平滑廓线中的“锯齿”锋为目的，在满足以下条件的情况下，将预先用线性内插加以平滑。假定以 g_i 表示第 i 点的气象要素值，如果

$$|g_i| < (|g_{i-1}| + |g_{i+1}|) / 10$$

则定为异点。这主要考虑到异点拟合时权重的设定。

关于异点拟合诊断及一致性拟合诊断已在前面作了讨论，在此不再复述。

四、结 论

利用分段、逐点、加权拟合的控制方案，我们对 1986 年 6 月 — 1988 年 5 月的资料进行了质量控制，结果表明：该方案不仅能较好地完成边界层内简单结构的质量控制，对于有特征层存在时，也同样具有很高的诊断率。图 1 给出了 1987 年 1 月 12 日温度、风速的原始资料时空演变图（等值线），图 2 给出了相应的质量控制后的演变图，从两图可以看出，质量控制主要修正了那些由于仪器僵死而造成的“僵值”，从而使本身零乱不堪的演变图变得有规可循，譬如温度场的逆温结构和风速场中的大风区形成很好的对应关系等等。方案在保持此类特征结构原始性的基础上又力求对资料进行最

大限度质量控制的例子很多, 由于篇幅有限, 在此不一一列举. 总之, 只要经过气候极值诊断、大尺度天气学特征诊断, 然后合理地选择权重因子和误差极值标准, 用分段二次拟合诊断法对边界层资料进行质量控制, 不仅能准确地诊断出异点, 而且能合理地给出异点平滑值, 这对目前研究甚少的资料质量控制方案也许是一个好的参考. 不仅如此, 由于该方案从普遍的数学物理方法出发, 只是在误差极值标准和权重确定时才涉及专业知识, 因此, 如果结合具体的物理过程, 确定最优的可取函数类和合理的误差极值标准, 它也能应用到别的资料控制中去.

由于该方案具体实施中存在误差极值标准和保持资料原始性的矛盾, 因此在误差极值标准给得太大时, 会有“漏诊”现象, 但从保持原始性的角度上还是值得的.

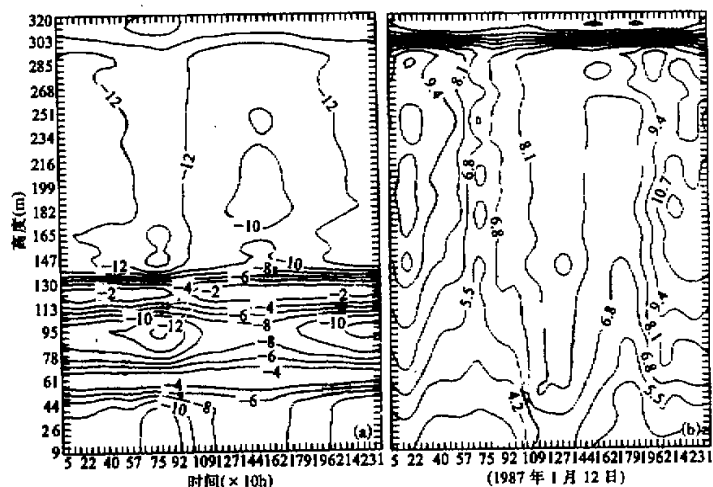


图1 修正前的风温时空演变图 (1987-01-12)

(a) 温度场, (b) 风场.

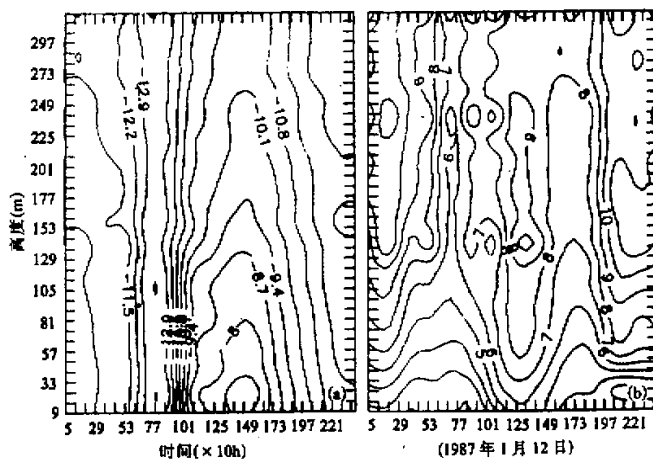


图2 修正后的风温时空演变图 (1987-01-12)

(a) 温度场, (b) 风场.

参 考 文 献

- [1] Lorenc A. C. and O. Hammon. 1988, *Objective Quality Control of Observations Using Bayesian Methods Theory, and a practical Implementation*, *Q. J. R. Meteor. SOC.*, **114**, 515 - 543.
- [2] Leys, Gandin, 1988, *Complex Quality Control of Meteorological Observations*, *Mon. Wea. Rev.*, **116**, 1137 - 1156.
- [3] 德罗兹多夫, 1963, 气象观测的气候学整理方法, 科学出版社 (中译本).

The Quality Controlling Method and Application of the Original Data of the Boundary Layer

Chen Ming Hong Zhongxiang

(*Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029*)

Abstract

Based on the original data of the Beijing 325-meter Meteorological Tower, a method has been devised here to control the quality of the original data. According to the general meteorological property and using the Weighting Stagewise Regression Procedure, this method can automatically single out and modify almost all the "odd-datum". Each of the processes goes through the whole spatial sequence point-by-point and includes five steps: climatologic maximum diagnosis, synoptic diagnosis, "odd-point" regression diagnosis, consistence regression diagnosis.

Key words: Odd-point; Digital weight; Regression precision.