

大气廓线统计反演方法的改进

范天锡

(卫星气象中心)

提要

目前,在由 TOVS 资料反演大气温度和水汽廓线的业务处理中,使用统计方法。本文探讨了改进现行统计反演方法的途径,并通过数值模拟试验证明,改进的方法与现业务方法相比,反演精度可以有相当大的改进。

一、前言

用TIROS-N业务垂直探测器(TOVS)的数据反演大气廓线的方法有两种,一种是物理方法,一种是统计方法。物理方法从理论或发展前景来看应更为理想;而统计方法简单稳定,不需要大气透过率数据,不受仪器定标精度和性能变化引入系统误差的影响,也有其优点。

目前,美国国家地球卫星和资料信息服务局(NESDIS)的业务处理系统仍采用统计反演方法,即Smith于1976年提出的协方差矩阵特征向量法^[1]。其反演精度,温度大约为2.0—2.5°C,相对湿度约为30%^[2]。这种反演精度尚不能满足使用要求,特别是湿度反演精度较差。那么,反演精度能不能通过反演方法的改进而进一步提高呢?就特征向量法本身而言,它已是一种相当完备的统计方法。近年来大量工作涌向物理方法的研究,取得相当进展。然而对于业务处理来讲,物理方法的反演精度目前仍未能明显超过统计方法。对于统计方法,当然也有一些研究工作。

作者认为并通过数值模拟试验证实,统计方法仍有相当大的改进余地,它可以通过以下途径实现。

1. 使用统计方法时,产生反演系数的统计样本应反映待反演的大气廓线的特性。NESDIS通过把全球分成5个纬度带每周更新一次反演系数来解决这一问题。然而即使在一个纬度带和一个不长的时期内,仍然存在不同类型的大气廓线。因此,对大气廓线分类,分别计算其反演系数,然后对探测数据加以自动判识,分不同类型进行反演,应能得到更好的效果。

2. 目前,特征向量法最后给出的是一个

多重线性回归方程,要求参与反演的因子和被反演因子之间具有线性关系。然而有的并非如此,正如NESDIS用通道亮温值反演水汽混合比,二者之间就很难说有线性关系,因此必将带来较大误差,这是应予改进的。

3. 参与反演的参数中,应充分包含被反演因子的信息,这一点可以通过适当选择和参与反演的参数加以处理来实现。

4. 对初始反演结果进行修正。

此外,对表面特性、地形和常规探空资料与卫星资料匹配中时间差别的影响加以修正,也是提高反演精度应考虑的问题。

本文通过数值模拟试验,探讨了改进统计反演方法的一些途径,并显示出很好的效果。

二、反演方法及其改进

1. 基本方法仍采用协方差矩阵特征向量法。对于m个通道的亮温值 $\hat{x} = \hat{x}_1, \dots, \hat{x}_m$ 和n层大气温度或湿度值 $\hat{y} = \hat{y}_1, \dots, \hat{y}_n$,其N个样本的平均值分别为 $\bar{x}_1, \dots, \bar{x}_m$ 和 $\bar{y}_1, \dots, \bar{y}_n$, \hat{x} 和 \hat{y} 与其平均值的差值分别为 $x = x_1, \dots, x_m$ 和 $y = y_1, \dots, y_n$ 。我们有

$$x = E_x a \quad (1)$$

$$y = E_y b \quad (2)$$

式中, E_x 和 E_y 分别是 \hat{x} 和 \hat{y} 的协方差矩阵的特征向量,向量a和b则分别是x和y的变换系数。由于 E_x 、 E_y 的正交性,则

$$a = E_x^T x \quad (3)$$

$$b = E_y^T y \quad (4)$$

E_x^T 和 E_y^T 分别为 E_x 和 E_y 的转置矩阵。

由于采用的特征向量数可比原参数少很多,并可起抑制观测中随机噪声的作用,由a反演b比直接由x反演y更稳定。

$$b = Ca \quad (5)$$

式中C为变换系数。将(3)式代入(5)

式, 再将 (5) 式代入 (2) 式则得到:

$$y = (E_y C E_x^T) x \quad (6)$$

C可用最小二乘法确定。令A、B、X、Y分别是a、b、x、y的样本矩阵, 则

$$B = CA$$

于是

$$C = BA^T (AA^T)^{-1}$$

其中

$$A = E_x^T X$$

$$B = E_y^T Y$$

2. 对y进行分类反演。由于a、b的各分量之间是互不相关的, 前n个分量集中了x、y的主要特征, 所以可用a或b的前n个主分量代替x或y进行分类, 从而使分类简便易行。我们采用聚类方法分类, 具体步骤如下。

(1) 形成初始聚类中心 首先计算N个样本的 a_i 的平均值 \bar{a}_i 和标准差 s_i , 并计算

$$g_i \pm = \bar{a}_i \pm s_i$$

如果采用a的前3个分量进行分类, 通过 $g_i \pm$ 相互搭配, 可得到8个初始聚类中心。

(2) 分类 求出每个样本的 a_i 值, 并求出其与第j类聚类中心的距离

$$D_{ij}^2 = \sum_{i=1}^3 [(a_i - g_{ij}) / s_i]^2$$

比较所有的 D_{ij} , 样本属于 D_{ij} 最小的那一类。

(3) 重新聚类 计算每一类样本的平均变换系数 \bar{a}_{ij} 和标准差 s_{ij} 。以 \bar{a}_{ij} 为新的聚类中心, 并计算各聚类中心的距离

$$D_{jk}^2 = \sum_{i=1}^3 [(\bar{a}_{ij} - \bar{a}_{ik}) / s_i]^2$$

如果某两聚类中心的距离 $D_{jk} < Q$ (Q为阈值), 则将其合并。如果某一类的样本数 $M < I$ (I为阈值), 则将其与距离最近的一类合并。合并后再算出样本的平均值 \bar{a}_{ij} 作为新的聚类中心。

如果某一类的 $s_{ij} > T$ (T为阈值), 则将其分为两类, 新的聚类中心为

$$g_{ij1} = \bar{a}_{ij} - s_{ij}$$

$$g_{ij2} = \bar{a}_{ij} + s_{ij}$$

(4) 迭代 新的聚类中心形成后, 再重新进行分类, 反复多次, 直至其稳定或进行到一定迭代次数后停止。

用上述步骤分类后, 再用协方差矩阵特征向量法求出每一类的反演系数。反演时, 首先用a判别x属于何类, 再用该类反演系数进行反演。

3. 湿度反演中处理方法的改进。现在业务处理系统中使用的回归方程为

$$\hat{q}_j = C_{0j} + \sum_{i=1}^m C_{ij} T_{Bi} \quad (7)$$

式中, \hat{q}_j 是大气第j层的水汽混合比, C_{0j} 和 C_{ij} 是回归系数, T_{Bi} 是第i通道的亮温值。

如果被反演的因子用露点温度代替水汽混合比, 则有

$$\hat{T}_{Dj} = D_{0j} + \sum_{i=1}^m D_{ij} T_{Bi} \quad (8)$$

式中, \hat{T}_{Dj} 是大气第j层的露点温度, D_{0j} 和 D_{ij} 是回归系数。

由于在目前业务处理中, 实际上水汽含量与大气温度的强相关性是水汽廓线反演的物理基础^[3,4], 即 T_{Bi} 中以测温通道为主体, \hat{q}_j 或 \hat{T}_{Dj} 主要反映与大气温度廓线相适应的水汽廓线。如果实际露点温度为 T_{Dj} , 则有偏差值

$$\Delta T_{Dj} = \hat{T}_{Dj} - T_{Dj}$$

如果我们用测温通道的亮温值统计回归3个水汽通道的亮温值,

$$\hat{T}_{Bk} = B_{0k} + \sum_{i=1}^m B_{ik} T_{Bi} \quad (9)$$

式中, \hat{T}_{Bk} 是第k水汽通道的亮温度, B_{0k} 和 B_{ik} 是回归系数。 \hat{T}_{Bk} 应反映与大气温度廓线相适应的水汽廓线所具有的水汽通道亮温值, 它与实际水汽通道亮温值 T_{Bk} 亦有偏差值

$$\Delta T_{Bk} = \hat{T}_{Bk} - T_{Bk} \quad (10)$$

在 ΔT_{Bk} 和 ΔT_{Dj} 之间建立回归方程

$$\Delta \hat{T}_{Dj} = A_{0j} + \sum_{k=1}^3 A_{kj} \Delta T_{Bk} \quad (11)$$

式中 A_{0j} 和 A_{kj} 是回归系数。

当我们把(8)式得到的 \hat{T}_{Dj} 看成初始反演结果, 于是可由 ΔT_{Bk} 去求 $\Delta \hat{T}_{Dj}$, 用 $\Delta \hat{T}_{Dj}$ 去修正 \hat{T}_{Dj} , 从而得到修正后的反演结果

$$\tilde{T}_{Dj} = \hat{T}_{Dj} - \Delta \hat{T}_{Dj} \quad (12)$$

(9)和(11)式中的回归系数亦可用特征向量法求出。

二、数值模拟试验

取反映中纬度(30—60°N)各种天气状况的400条大气廓线, 每条由0.1—1000hPa分为40层。用大气辐射传递方程, 根据大气透过率计算出与这400条大气廓线对应的NOAA卫星高分辨率红外辐射探测

表1 分类和不分类方法的温度和湿度反演误差

层 序 号		26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	平均
气压值(hPa)		300	350	400	430	475	500	570	620	670	700	780	850	920	950	1000	
温度反演误差 (°C)	不 分 类	2.70	2.07	2.14	1.80	1.63	1.71	1.01	0.95	1.24	1.49	1.38	2.35	1.17	0.69	0.11	1.50
	分 类	1.95	1.52	1.70	1.42	1.33	1.44	0.86	0.83	1.11	1.33	1.07	1.96	0.98	0.58	0.07	1.21
湿度反演误差 (相对湿度%,)	不 分 类	20.5	26.7	24.2	24.7	25.2	24.7	27.2	27.8	27.7	27.4	28.1	30.4	30.8	31.8	32.9	27.3
	分 类	15.7	20.5	19.0	18.1	18.6	18.2	19.0	20.5	21.3	21.4	19.7	22.3	20.3	20.5	22.3	19.8

器(HIRS/2)的19个红外通道的亮温值,于是构成了模拟匹配数据集。在匹配数据集中取一部分为统计样本,另一部分为检验样本。

在分类反演的数值试验中,温度反演取17个通道(1—17)的数据,湿度反演取13个通道(4—16)的数据,15层温度和湿度的反演精度列于表1。在聚类中取 $T=0.5$, $I=20$, $Q=0.8$ 。

在试验(7)和(8)式的湿度反演中,采用4种通道组合方式,(I)第1—17通道;(II)第4—16通道;(III)第4—8、13—16通道;(IV)第8、10—12通道。15层反演结果的平均误差列于表2。

表2 (7)和(8)式的湿度反演误差

反演公式		通道组合方式			
		I	II	III	IV
(7)式	混合比误差	0.765	0.760	0.839	1.055
	相对湿度误差(%)	23.4	24.1	27.2	33.2
(8)式	露点温度误差(°C)	6.57	6.55	8.26	8.90
	相对湿度误差(%)	17.1	17.0	21.0	21.8

对于(8)式取得的初始反演结果,可以通过(9)一(12)式作进一步修正,在数值试验中采用了通道组合(II)和(III),15层湿度反演结果的平均误差列于表3。

表3 在(II)和(III)通道组合下(8)式及修正后的湿度反演误差

反演误差	通道组合	(8)式反演误差		修正后反演误差	
		II	III	II	III
露点温度误差(°C)		6.55	8.26	6.01	5.97
相对湿度误差(%)		17.0	21.0	15.3	15.4

从表1可以看到,不分类的15层温度和相对湿度的平均反演精度分别为1.5°C和

27.3%。通过分类反演,温度反演精度平均提高了0.29°C,湿度提高了7.5%,二者都大约提高了20%。在我们的试验中,选用的400条大气廓线代表了各种天气状况,很难清楚地划分为12个大气廓线类型,这种状况反映在聚类过程中,很难稳定地聚成几类,如果在实际业务中应用,只处理某一地区和某一时期中的资料,大气廓线的类型易于确定,聚类可以按大气实际情况分类,应能取得更好的效果。

从表2可以看到,(8)式的湿度反演精度比(7)式有很大提高,平均提高了29%,这说明露点温度与通道亮温值的线性关系比水汽混合比要好。

4种通道组合的区别是:(I)采用全部通道,(II)去掉高层大气测温通道,(III)只含测温通道,(IV)只含水汽和表面温度通道。反演结果是,(II)和(I)反演精度大致一样,说明高层测温通道在反演中不起作用;(III)比(IV)好,说明水汽含量与温度的强相关性,水汽信息主要含于测温通道中;(II)比(III)好,说明水汽通道含有测温通道中所没有的水汽信息,对提高反演精度有明显的作用。

从表3可以看到,对(8)式的反演结果进行修正,精度可以进一步提高,对于通道组合(III),反演精度平均提高了27%,对于通道组合(II),虽然初始反演中已经使用了水汽通道,修正后的精度仍大约提高了10%,不论初始反演采用何种通道结合,修正后的反演结果的精度基本一致。

四、结论

数值模拟试验证实,本文中所采用的一些处理方法,特别是通过对数据的妥善运用,是可以将统计反演方法改进的。在我们的试验中,分类方法对温、湿反演精度都提高了20%,用露点温度代替水汽混合比

作为被反演因子,湿度反演精度可提高10%,三者结合使用,如能得到分别使用时的效果,则湿度反演精度可提高49%。

上述反演精度提高之大是令人吃惊的。当然,我们的试验是在一些理想条件下进行的,目的在于对不同方法的反演精度相对比较。在实际资料反演时,由于观测误差、云污染、表面作用、资料匹配误差等等多种因素的影响,可能是达不到这样大幅度的改进,然而一定程度的改进显然是可能的。

参考文献

[1] Smith W. L. and H. M. Woolf, The use of eigenvectors of statistical covariance matrices for interpreting satellite sounding radiome-

ter observations, J. Atmo. Sci., Vol. 33, No. 7, p. 1127, 1976.

[2] Dismachek D. C., A. C. Booth and J. A. Leese, National environmental service catalog of products, Third edition, NOAA Technical Memorandum NESS 109, 1980.

[3] McMillin L. M. and A. Sanyal, The use of brightness temperature ratios to retrieve water vapor, Proceedings of the fifth conference on atmospheric radiation, 51-53, 1983.

[4] Sanyal A., C. A. Dean, J. S. Prasad and L. M. McMillin, An evaluation of the use of atmospheric radiances for water vapor retrieval in a global retrieval system, Proceedings of the fifth conference on atmospheric radiation 76-79, 1983.

An improved statistical method to retrieve the atmospheric profiles from satellite data

Fan Tianxi

(Satellite Meteorological Center)

Abstract

Based on the statistical method used to operational processing an improved method has been developed for retrieving atmospheric temperature and water vapor profiles from TOVS data. The accuracy of this method is compared with that of operational method by numerical simulation study. The present method is more accurate than operational method.