

地基微波遥感大气水汽总量的 普适性回归反演*

吕达仁 魏 重 忻妙新**

(中国科学院大气物理研究所, 北京 100029)

冯士雍 王晓刚

(中国科学院系统科学研究所)

提 要

本文利用不同季节、经纬度和海拔高度的历史探空资料为样本, 用回归分析和岭回归分析相结合建立了适用于全球各地的双波长地基微波辐射计遥感晴空大气水汽总量的回归方程。多种数值检验表明, 所提出的反演关系式的精度较高能提供实际应用。同时, 将 20.6 和 31.65GHz 及 22.2 和 35.0GHz 两组频率对的结果作了比较。

关键词: 微波遥感; 水汽遥感; 反演算法; 岭回归; 回归分析。

一、引 言

大气水汽作为大气系统的潜热和水循环中的重要一环, 是天气和气候系统中具有本质重要性的关键因子之一, 其总量和垂直分布的全球和区域性测量一直受到十分的重视, 对它的精确快速测量也是大气探测的关注重点之一。现有的探空仪测量一般认为有 10% — 15% 的误差, 即使如此也由于施放条件的限制而无法全球推广。卫星红外波段的水汽遥感(如 NOAA/ TOVS)往往具有更大的误差而至今较少实用。相对而言, 地基和空基微波遥感已被认为是较适用的水汽遥感手段, 并已在世界范围内得到一定的发展。如卫星微波水汽遥感的 DMSP/SSMI 试验^[1] 以及地基双波长微波水汽和云水遥感试验^[2-4]。由于地基微波遥感在时间上的极高分辨率以及操作的方便性, 它已经成为各类研究和业务应用中的重要手段之一。由于它在遥感水汽总量中的精度已在许多情况下与探空相当, 因此一些空间遥感的对比试验中作为“实况”(ground truth)加以利用^[5]。

当前微波遥感水汽总量和廓线的反演方案有物理和统计方法两类。但无论卫星或是地面遥感, 真正能投入应用的基本上是统计方法。将统计方法建立于更好的物理考虑基础上是当前反演算法的发展方向。一般的统计反演方法都是采用局地(地区代表性)探空资料作为统计样本进行回归分析, 以求得实测辐射亮温和大气水汽总量的较高精度的统

1991年2月22日收到, 1992年7月20日收到修改稿。

* 国家自然科学基金支持项目。

** 陈英同志参加了工作。

计关系。局地样本所包含的地区代表性优点同时也成为应用局限性的缺点,因为对于一个无探空资料的地区,如何获得统计关系就成为问题。

为了充分发挥地基微波遥感水汽在时间连续性及设备易移动和操作的优越性,本文提出一种普适性回归反演方案,企图使建立的回归方程不仅适用于全球不同经纬度、不同海拔高度也适用于不同季节。

二、基 本 分 析

我们分析水汽总量 Q 和双波长辐射计垂直方向的探测亮温 $T_b(\lambda)$ (λ 为辐射计波长) 之间的关系

$$Q = \int_0^{\infty} \rho_{H_2O}(z) dz, \quad (1)$$

$$T_b(\lambda) = T_{\infty} \exp \left(- \int_0^{\infty} (\alpha_{H_2O}(\lambda, z) + \alpha_r(\lambda, z)) dz + \int_0^{\infty} T(z) (\alpha_{H_2O}(\lambda, z) + \alpha_r(\lambda, z)) \right. \\ \left. \cdot \exp \left(- \int_0^z (\alpha_{H_2O}(\lambda, z') + \alpha_r(\lambda, z')) dz' \right) \right), \quad (2)$$

其中 T_{∞} 为宇宙背景亮温 2.7K, $\rho_{H_2O}(z)$ 代表 z 高度的水汽含量, $\alpha_{H_2O}(\lambda, z)$ 和 $\alpha_r(\lambda, z)$ 分别代表 z 高度处的水汽吸收系数和水汽以外(主要是氧气)的吸收系数。很明显, 所测 $T_b(\lambda)$ 与 Q 之间的关系绝不是单一的。从定性分析可看出 $T_b(\lambda)$ 与 $T(z)$, $\rho_{H_2O}(z)$ 和 $p_r(z)$ 均有关系, 而这些量又直接或间接地与大气的温、湿、压廓线有关。由此推断我们要建立的关系必定是一种多元回归关系。温、湿、压廓线的探测较复杂, 从实用出发在回归关系中的自变量应是易探测量, 因而很自然想到地面的常规气象参数, 以及它们的某些组合。我们的目的在于建立全球范围, 不同季节与时间都适用的水汽总量回归方程, 这就要尽量多地考虑到对 Q 有影响的因子, 以使方程具有较好的普适性。出于这样的考虑, 我们选了以下 10 个因子参与回归关系的建立: T_{b_1} (单位 K)——第一个波长的亮温, T_{b_2} (K)——第二个波长的亮温, T_s (K)——地面气温, e_s (g/m^3) 地面绝对湿度, p_s (hPa)——地面气压, h (km)——测站海拔高度, T_{b_1}/T_s , T_{b_2}/T_s , e_s/p_s 和 $T_s \times e_s$ ($273.15 \times p_s$)。以上这些因子可根据探空资料所提供的信息, 或进一步将探空代入(1)和(2)式计算均可获得。所以进一步的问题归结为选取有足够代表性的探空样本, 它们应尽量包括高原、平原、海洋和大陆的特征, 同时也顾及季节和时间的代表性。表 1 列出了这些样本来源的情况, 全部样本总量为 2740。

问题的另一方面是, 我们应从众多的自变量中挑选那些有重要影响的因子, 而剔除那些影响不大的因子, 以使回归方程稳定、简便和实用。

从当前国内外双波长微波辐射计实际采用的频率来看大致分为两组: 一组是 22.2 和 35.0GHz (下面称为第一组频率对), 另一组是 20.6 和 31.65GHz (第二组频率对)。对这两组频率我们都作了讨论。

表 1 回归样本采集范围和数量

站 名	经 度	纬 度	海 拔(m)	样 本 量		
				夏 季	冬 季	总 计
关岛	144° 50' E	13° 33' N	111	114	338	452
·雅浦	138° 05' E	09° 39' N	77	98	209	307
广州	113° 19' E	23° 08' N	7	141	213	354
北京	116° 17' E	39° 56' N	55	46	226	272
拉萨	91° 08' E	29° 40' N	3650	176	200	376
那曲	92° 04' E	31° 29' N	4508	147	198	345
张掖	100° 26' E	38° 56' N	1483	113	200	313
丽江	100° 26' E	26° 52' N	2394	125	196	321

三、关于水汽总量的逐步回归分析(最小二乘估计)

1. 多元回归模型和逐步回归思想

记 Y 为因变量, $x_1, x_2, \dots, x_m$ 是所考察的回归因子, 记 n 组观测数据为

$$(x_{1t}, x_{2t}, \dots, x_{mt}; Y_t), t = 1, 2, \dots, n$$
 (3)

Y 对于所有回归因子的线性回归模型的矩阵形式为

$$Y = X\beta + \delta, \tag{4}$$

其中

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & X_{11} & X_{12} \cdots X_{1m} \\ 1 & X_{21} & X_{22} \cdots X_{2m} \\ \vdots & \vdots & \vdots \quad \vdots \\ 1 & X_{n1} & X_{n2} \cdots X_{nm} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}. \tag{5}$$

对于上述回归模型, 通常用最小二乘法对回归系数 $\beta' = (\beta_0, \beta_1, \dots, \beta_m)$ 进行估计, 估计量 $\hat{b} = \hat{\beta}$ 满足以下正规方程

$$(X'X)\hat{\beta} = X'Y. \tag{6}$$

从而 $\hat{b}' = (b_0, b_1, \dots, b_m)$ 为

$$\hat{b} = \hat{\beta} = (X'X)^{-1}X'Y. \tag{7}$$

此时所得(经验)回归方程为

$$Y = b_0 + b_1x_1 + b_2x_2 + \dots + b_mx_m. \tag{8}$$

当误差 ε 是无偏且诸分量 ε_i 不相关时, $\hat{\beta}$ 是 β 的无偏估计, 且它的方差——协方差阵为

$$D(\hat{\beta}) = \sigma^2(X'X)^{-1}, \tag{9}$$

其中 σ^2 为 ε_i 的公共方差.

对样本数据而言, 回归方程(8)的拟合效果可用以下标准残差 S_e 衡量, S_e 愈小, 回归方程的拟合效果愈好,

$$S_e = \sqrt{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 / (n - m - 1)}.$$

另一个衡量拟合效果好坏的量是复相关系数 R , R 愈接近 1 表明回归方程拟合效果愈好.

在可考虑的回归因子数目较多的情况下, 回归方程的选择是十分重要的. 一般采用逐步回归分析方法只将那些对因变量有显著作用的因子逐个引入回归方程, 在每个因子引入方程后, 对原先已引入方程的因子重新作检验以剔除因新因子引入而变得不显著的因子. 这样可保证最后得到的回归方程只包含重要因子.

2. 逐步回归分析的结果

对所有回归样本的水汽总量, 按二组频率对, 分别用前面所说的 10 个因子进行逐步回归. 无论对第一组还是对第二组频率对中的 Y , 回归方程的效果均十分显著. 事实上, 当取 $F_1 = F_2 = 2.5$ 时, 第一组频率对(22.2, 35 GHz), 除 e_s 外的 9 个因子均入选, 复相关系数 R 高达 0.999801, 标准残差 $S_e = 0.0406533$ cm. 第二组频率对(20.6, 31.65GHz), 除 T_{b_1} 外的 9 个因子都入选, $R = 0.999794$, $S_e = 0.041299$ cm.

从应用的角度而言, 包括太多的回归因子的方程使用不便, 而且稳定性可能较差. 因此, 我们仅选用只包含 2—4 个最显著且使用方便因子的回归方程. 表 2 与表 3 列出了这些方程的主要结果.

表 2 22.2 和 35GHz 频率的水汽总量回归方程

变量个数	回 归 系 数					R	S_e
	常数项	T_{b_1}	T_{b_2}	P_s	e_s/P_s		
2	-0.798088	0.0387127	0.0685678	/	/	0.997665	0.139307
3	0.135305	0.0268420	0.116971	-0.00157068	/	0.999378	0.071819
4	0.200101	0.029090	0.123548	-0.00170427	-17.059720	0.999639	0.0546346

从表 2 与表 3 可以看出:

1) 对于第一组频率对(表 2), 在晴空大气条件下探测水汽总量起决定性作用的因素是两个波长的亮温, 地面气压及地面水汽分压与气压之比. 其他因子, 比如台站海拔高度、地面温度, 或是本来就对水汽总量影响不大, 或是它们的影响已由引入方程的其他量表达出来, 因此就相对地显得不太重要了.

2) 第二组频率对(表 3)情况与第一组有所不同, 首先, 两个波长中只有 20.6GHz 的亮温 T_{b_1} 起主要作用, 31.65GHz 亮温(T_{b_2})在晴空探测时作用不太显著, 我们没有引入表 3. 其次, 表 3 中一个变量回归方程的标准残差 S_e (0.105235) 已比表 2 中二个变量回归方程的标准残差(0.139307)要小. 依此类推, 表 3 中二个变量方程的 S_e 比表 2 中三个变量方程的 S_e 也小. 第三, 表 3 中随着变量数增加(即辅助探测量增加)回归方程的效果改善程度(S_e 减少的幅度)不如表 2 中显著. 换言之, 对于 20.6 和 31.65GHz 这一组频率来说, T_{b_1} 的作用占了最显著的位置, 以至于不需要 T_{b_2} 来作补充订正, 就可达到较高的精度, 同时地面其他辅助量引入后方方程精度改善也就不如第一组频率那么

表3 20.6 和 31.65GHz 频率的水汽总量回归方程

变量个数	回 归 系 数					R	S _e
	常数项	T _{b1}	P _S	e _S /P _S	T _{b1} /T _S		
1	-0.510496	0.101052	/	/	/	0.998661	0.105235
2	-0.0570359	0.104439	-0.000642427	/	/	0.999496	0.064688
3	-0.00346922	0.109980	-0.000720553	-13.232850	/	0.999660	0.053180
4	+0.0364442	0.140528	-0.000725351	-15.442320	-9.049497	0.999688	0.050947

重要。所有这些现象都说明 20.6GHz 比 22.2GHz 测晴空水汽总量优越，这是由于 20.6GHz 处于 22.235GHz 吸收峰的侧翼，该波段亮温对水汽廓线分布的敏感程度较后者低，因而与水汽总量的线性相关程度更好的缘故。但是，若要反演水汽廓线分布，则采用 22.2GHz 波段更有利。还应指出，如果把云天情况也考虑进去的话，31.65GHz 的亮温值 T_{b_2} 就不再是可有可无，而是至关重要。这里，作为初步研究晴空探测的问题，我们着重对表 3 中的一些方程作深入的讨论，对云天情况将在另文讨论。

四、水汽总量的岭回归分析

1. 回归系数的岭估计

本文所考虑的回归因子之间存在着高度的相关性。以第一组频率为例，4 个变量 T_{b_1} 、 T_{b_2} 、 p_s 和 e_s/p_s 之间的相关系数中，有 3 个超过 90%。系数矩阵的特征根分别为 $\lambda_1 = 2.41170$ ， $\lambda_2 = 1.14841$ ， $\lambda_3 = 0.41835$ ， $\lambda_4 = 0.02155$ ，最小特征根 λ_4 相当小，因此正规方程系数矩阵 $X'X$ 的行列式也接近于 0，我们称这样的 X 是病态的，按最小二乘估计的回归系数的方差(根据(9)式与 $(X'X)^{-1}$ 有关)就会变得很大，从而大大影响回归方程的稳定性。为此，我们对样本数据作岭回归分析。

岭回归分析^[6-8]是克服 X 矩阵是病态的一种方法，其方法是在估计回归系数 β 时将(7)式中正规方程的系数矩阵 $X'X$ 的对角线上的每个元素上增加一个相对较小的正数 k ，则回归系数的岭估计为

$$\hat{\beta}(k) = (X'X + kI)^{-1} X'Y, \tag{10}$$

其中 k 称为岭参数。与 β 的最小二乘估计 $\hat{\beta}$ (也即 $\hat{\beta}(0)$) 相比，岭回归方程对样本数据的拟合稍差，即 S_e 变大。但理论证明，由于 $\hat{\beta}(k)$ 的方差较小，故适当地选取 k 值，有可能使 $\hat{\beta}(k)$ 的总的均方误差比 $\hat{\beta}(0)$ 的均方误差更小，从而岭回归方程对新数据的预测效果较好。

选择什么样的 k 是岭回归分析中的重要问题。我们以 0.005 为步长，将 $\hat{\beta}(k)$ 的值(这里采用的是归一化的值，以便可以标在一张图上)描点， $\hat{\beta}(k)$ 各分量的这种数值变化曲线称为岭迹图。图 1 中给出了两组频率对不同变量回归方程的岭迹图。从图中可见，各岭回归系数 $\hat{\beta}_i(k)$ ，当 k 从 0 变大时，刚开始变化较大，在 $k = 0.01 - 0.02$ 左右开始趋于平稳，因此根据岭迹图我们可以大致确定 k 的范围，下面再根据对岭回归方程的效果检验最终确定 k 的选取。

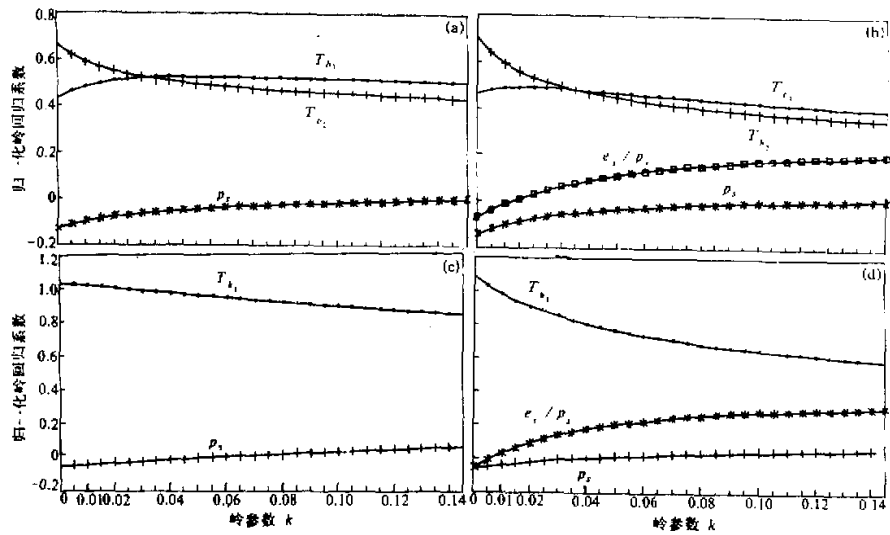


图 1 两组频率对的岭迹图

(a) 22.2 35.0GHz 三变量, (b) 四变量, (c) 20.6 31.65GHz 两变量, (d) 三变量

2. 岭回归分析的主要结果

表 4 和表 5 中分别给出了二组频率对岭回归分析的主要数值结果, 其中 $k=0$ 即是最小二乘估计, 表中包括各变量的回归系数与标准残差 S_e , 其中回归系数已变换回原单位.

表 4 22.2 和 35.0GHz 水汽总量的岭回归分析主要结果

	k	b_0	$b_1(T_{b1})$	$b_2(T_{b2})$	$b_3(p_s)$	$b_4(e_s/p_s)$	S_e
三 变 量	0	0.135305	0.026842	0.116971	-0.001571		0.071819
	0.005	0.011529	0.028929	0.108455	-0.001342		0.074193
	0.010	-0.074334	0.030211	0.102854	-0.001181		0.078548
	0.015	-0.139112	0.031052	0.098859	-0.001058		0.083290
	0.020	-0.190864	0.031624	0.095842	-0.000959		0.088060
四 变 量	0	0.200101	0.029090	0.123548	-0.001704	-17.059720	0.054634
	0.005	0.044987	0.030244	0.111973	-0.001411	-9.650537	0.061297
	0.010	-0.063641	0.030683	0.104008	-0.001203	-3.363843	0.073559
	0.015	-0.145722	0.030774	0.098096	-0.001044	2.060932	0.086401
	0.020	-0.210960	0.030683	0.093479	-0.000916	6.797159	0.098576

五、回归方程的检验

对表 4、5 中得到的回归方程 (包括最小二乘估计及岭估计两种方法得到的方程), 我们用三种样本从总体上检验这些方程的效果, 在此基础上进一步选定岭估计的 k 值,

表 5 20.6 和 31.65GHz 的水汽总量岭回归分析结果

	k	b_0	$b_1 (T_{d1})$	$b_2 (p_s)$	$b_3 (e_s/p_s)$	S_e
一 变 量	0	-0.057036	0.104439	-0.000642		0.064688
	0.005	-0.091952	0.103548	-0.000572		0.066157
	0.010	-0.125528	0.102677	-0.000503		0.070252
	0.015	-0.157821	0.101826	-0.000437		0.076388
	0.020	-0.188884	0.100995	-0.000372		0.083981
二 变 量	0	-0.003469	0.109980	-0.000721	-13.232850	0.053180
	0.005	-0.085305	0.104257	-0.000581	-1.709275	0.063663
	0.010	-0.154963	0.099435	-0.000461	7.880942	0.083589
	0.015	-0.215109	0.095312	-0.000357	15.972060	0.104305
	0.020	-0.267674	0.091743	-0.000264	22.877920	0.123635

最后给出在选定的 k 值下分台站(和季节)的普适性检验结果。

三类检验样本为：1) 参与建立回归方程的原始样本数据，即表 1 中的 2740 份个例，总称为自检验样本。2) 与自检验样本有着相同台站但年份和(或)季节不同的新数据样本，共 706 份，称为时间外推检验样本。3) 自检验样本之外的 7 个新选台站的数据样本，共 312 份，称之为空间外推检验样本。新选台站照顾到他们的地理位置、海拔高度、气候、季节的代表性，详细情况见表 9。

每组检验样本都求出其均方根绝对误差 $p(k)$ 和平均相对误差 $j(k)$ ，其中

$$p(k) = \sqrt{\frac{1}{n} \sum_{t=1}^n [Y_t - \hat{Y}_t(k)]^2}, \quad (t=1, 2, \dots, n)$$
$$j(k) = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t(k)|}{Y_t},$$

式中 Y_t 是实际观测值， $\hat{Y}_t(k)$ 是按岭参数为 k 的回归方程的预报值。 $k=0$ 时即是最小二乘估计的回归方程预测结果。应该指出，在研究普适性问题时，平均相对精度的引入很有必要，因为这时检验样本的水汽总量离散范围很大，最大可达 $7.3\text{g}/\text{cm}^2$ ，最小可到 $0.041\text{g}/\text{cm}^2$ 。这时，均方根误差主要反映那些水汽总量值大的样本情况，而难于真实反映 Y_t 很小的个例情况。因此，普适性讨论时，相对误差能较全面反映情况。当然，相对误差检验对水汽总量小的样本的探测精度提出了更高的要求。

表 6 给出了三组样本的检验结果，从中可以得出以下结论：

1) 对第一组频率，对原测量台站，用最小二乘估计的回归方程效果较好，但用 $k=0.005$ 和 0.010 的方程的预测结果也令人满意，对新台站，用岭回归方程的预测效果优于用最小二乘回归方程的效果。比较起来，当 $k=0.005$ 时方程能满足各种情况，普适性较好。

2) 对第二组频率，当 k 小于等于 0.015 时都有较好的预测效果。其中，对原测量台站以 $k=0.005$ 或 0.010 最好，对新台站，以 $k=0.010, 0.015$ 最好。综合考虑，选 $k=0.010$ 能满足各种情况。

3) 随 k 的增加，变量个数相对多的方程与变量个数相对少的方程的检验效果逐渐接近，也就是说，最小二乘估计所体现的增加变量对回归精度改善较大的优越性已不太

表 6 岭回归方程的绝对精度与相对精度检验

样 本			自检验(2740例)		时间外推检验(706)		空间外推检验(312)	
			$p(k)$ (cm)	$j(k)$ (%)	$p(k)$ (cm)	$j(k)$ (%)	$p(k)$ (cm)	$j(k)$ (%)
第一组 频率对	三 变 量	$k=0$	0.071703	7.463	0.061073	2.358	0.076687	4.569
		0.005	0.074055	8.839	0.070416	2.546	0.069212	3.116
		0.010	0.078403	9.653	0.079019	2.853	0.072117	2.446
		0.015	0.083104	10.829	0.086313	3.197	0.077034	2.350
		0.020	0.087855	11.938	0.092525	3.480	0.083229	2.563
	四 变 量	0	0.054583	5.663	0.053376	2.123	0.058232	3.607
		0.005	0.061242	6.706	0.064106	2.462	0.059812	2.767
		0.010	0.073493	8.972	0.076539	2.824	0.068580	2.370
		0.015	0.086321	11.188	0.088148	3.220	0.079326	2.370
		0.020	0.098484	12.999	0.098771	3.569	0.090365	2.600
第二组 频率对	二 变 量	0	0.064654	6.1841	0.054881	1.869	0.077422	4.180
		0.005	0.066120	4.592	0.058260	1.720	0.077999	3.366
		0.010	0.070214	3.029	0.064427	1.754	0.081808	2.636
		0.015	0.076349	3.224	0.072337	1.973	0.087795	3.151
		0.020	0.083936	4.035	0.081320	2.272	0.095857	4.291
	三 变 量	0	0.053143	4.084	0.052888	1.756	0.074473	3.892
		0.005	0.063618	4.079	0.057100	1.691	0.076983	3.296
		0.010	0.083529	5.141	0.072297	1.977	0.089311	2.874
		0.015	0.104231	6.540	0.090006	2.446	0.105091	2.937
		0.020	0.123549	8.107	0.107109	2.969	0.120989	3.298

显著, 这种情况对于第二组频率尤为突出, 因此, 我们只选用变量个数少的二变量方程作为第二组频率的普适性方程。

由上面的分析, 我们最后选以下三个方程作为反演晴空水汽总量 Q 的普适性回归方程, 对频率组 22.2 和 35.0GHz, 岭参数选为 $k=0.005$, 这时三变量和四变量的回归方程分别是

$$Q=0.011529+0.028929T_{b_1}+0.108455T_{b_2}-0.001342p_s, \quad (11)$$

和

$$Q=0.044987+0.030244T_{b_1}+0.111973T_{b_2}-0.001411p_s-9.650537e_s/p_s, \quad (12)$$

对频率组 20.6 和 31.65GHz, 选岭参数 $k=0.010$ 所对应的含二个变量的方程为

$$Q=-0.125528+0.102677T_{b_1}-0.000503p_s, \quad (13)$$

其中水汽总量 Q 的单位为 g/cm^2 , 两频率亮温 T_{b_1} 和 T_{b_2} 的单位为 K, 地面气压单位为 hPa, 地面绝对湿度 e_s 以 g/m^3 为单位。

表 7—9 分别给出了普适性回归方程(11)—(13)的分台站检验效果, 由这三个表可以看出, 表中共包括 15 个台站的不同年份和季节的 29 组数据, 每组包含少至几十多至几百个样本, 共计为 3758 个样本, 除了第一组频率所对应的表 7 中那曲和拉萨以及表 9 中乌鲁木齐冬季三个资料组的检验结果相对误差较大, 超过了 10% 以外, 其余所有样本组检验结果的相对误差都在 10% 以内, 大部分不到 5%, 考虑到我们检验样本在时间、季节、地理位置和海拔高度几方面都有很大跨度, 以及探空测水汽本身有

表7 原始回归样本的分台站自检验

站 名		关岛	雅浦	广州	北京	拉萨	那曲	张掖	丽江	
年 份		79, 80, 81 82	79, 80, 82	79, 80	79, 80	79, 80	79, 80	79, 80	79, 80	
季 节		冬 夏	冬 夏	冬 夏	冬 夏	冬 夏	冬 夏	冬 夏	冬 夏	
样 本 数		452	307	354	272	376	345	313	321	
第 一 组 频 率	二 变 量	$p(0.005)$ (cm)	0.0956	0.1024	0.1037	0.0303	0.0642	0.0595	0.0327	0.0399
		$j(0.005)$ (%)	2.341	1.856	1.938	7.537	23.773	21.01	6.367	5.227
	四 变 量	$p(0.005)$ (cm)	0.0691	0.0752	0.1012	0.0253	0.0534	0.0460	0.0308	0.0390
		$j(0.005)$ (%)	1.495	1.201	2.446	5.802	19.22	13.521	4.563	4.884
第 二 组 频 率	二 变 量	$p(0.01)$ (cm)	0.0986	0.1088	0.1130	0.0269	0.0331	0.0108	0.0220	0.0217
		$j(0.01)$ (%)	2.322	1.747	2.090	1.777	6.689	5.939	1.889	1.047

表8 原测量台站时间外推分台站检验

站 名			关岛	雅浦	广州	北京		拉萨	那曲	张掖	丽江	
年 份			81	81	81	81		81	81	81	81	
季 节			夏	夏	冬	春	秋	夏	夏	夏	春	秋
样 本 数			108	89	115	28	28	101	89	96	29	23
第一组频率	三	$p(0.005)$ (cm)	0.0789	0.0855	0.0456	0.0340	0.0329	0.0894	0.1059	0.0248	0.0110	0.0200
	变	$j(0.005)$ (%)	1.058	1.167	1.655	2.155	2.990	4.424	7.713	0.723	0.451	1.286
	量											
	四	$p(0.005)$ (cm)	0.0595	0.0660	0.0747	0.0388	0.0402	0.0744	0.0879	0.0367	0.0130	0.0303
	变	$j(0.005)$ (%)	0.799	0.818	2.896	2.361	3.094	3.616	6.324	1.070	0.564	1.991
第二组频率	变	$p(0.01)$ (%)	0.0967	0.1017	0.0553	0.0595	0.0270	0.0396	0.0135	0.0388	0.0573	0.0702
	量	$j(0.01)$ (%)	1.304	1.354	2.354	3.099	1.730	1.898	0.919	1.184	2.925	4.312

3%—10%的误差,还考虑到上述三组相对误差较大的样本或是大陆性气候的冬季样本组,或是高原气候中冬季样本占的比重较大的样本组,因此它们的绝对误差并不大,我们认为,回归方程的普适性效果是很不错的,这样的精度已能满足绝大多数实际应用的需。

表9 新台站空间外推分站检验

站 名			西 安		乌 鲁 木 齐		上 海		昆 明		釜 岛	丹 佛	华 盛 顿
经 度			108° 56' E		87° 37' E		121° 26' E		102° 41' E		142° 11' E	104° 52' W	77° 02' W
纬 度			34° 18' N		43° 47' N		31° 10' N		25° 01' N		27° 05' N	39° 45' N	38° 51' W
海拔高度 (m)			398		918		5		1891		4.1	1625	84
年 份			1980		1980		1980		1980		1980	1961	1961
季 节			冬	夏	冬	夏	冬	夏	冬	夏	冬	夏	夏
样 本 数			30	28	28	27	26	30	28	28	30	31	26
第 一 组 频 率	二 变 量	$p(0.005)$	0.0242	0.0932	0.0522	0.0154	0.0239	0.1584	0.0319	0.0208	0.0933	0.0213	0.0512
		(cm)											
	四 变 量	$j(0.005)$	3.706	1.379	13.98	0.797	2.784	2.195	3.134	0.466	3.364	0.970	1.522
		(%)											
第 二 组 频 率	二 变 量	$p(0.005)$	0.0191	0.0627	0.0395	0.0195	0.0352	0.1207	0.0335	0.0269	0.1053	0.0345	0.0444
		(cm)											
	四 变 量	$j(0.005)$	2.790	0.917	10.464	0.952	3.236	1.671	3.197	0.658	3.755	1.492	1.310
		(%)											
第 三 组 频 率	二 变 量	$p(0.01)$	0.0103	0.0845	0.0133	0.0762	0.0163	0.1623	0.0320	0.0652	0.0595	0.1174	0.0970
		(cm)											
	四 变 量	$j(0.01)$	1.085	1.653	2.886	4.214	1.176	2.241	2.872	1.743	2.201	5.786	2.951
		(%)											

六、结论和讨论

本文利用具有气候、海拔和地表等不同代表性的探空样本建立了由地基双频微波辐射计反演水汽总量的回归方程(表4中方程11, 12和13)。这些方程包括2到4个变量, 它们至少可适用于北半球绝大部分地区, 满足必要的水汽总量反演精度。由于未对南半球台站资料进行检验, 不能提出确切的结论。但由于我们在样本选取中已考虑了热带海洋地区以及温带、副热带台站, 我们推测所得回归方程时南北球也是适用的。

单独从水汽总量反演看, 20.6和31.65GHz这组频率更为优越, 而22.2和35.0GHz一组频率也具有相当好的精度。如果考虑用扫角法来反演水汽廓线, 则后一组有其优越性。

本文的讨论未计及仪器的观测误差, 在实际应用时, 反演精度受仪器误差的影响, 要比这里的讨论精度低。在某些情况下, 如何提高仪器的精度成为解决问题的关键, 这里不再多加讨论。

作为本工作的下一步, 将建立双频联合遥感水汽总量和云中含水量的普适性回归反演方案。

参 考 文 献

- [1] Alishouse J. C., S. A. Snyder, J. Vongsathoia, and R. R. Ferraro, 1990, Determination of Oceanic Total Precipitable Water from the SSM/S, *IEEE Transactions on Geoscience and Remote Sensing*, 28 (5), 811—816.

- [2] Hogg, D. C., F. O. Guiraud, J. B. Snider, M. T. Decker and E. R. Westwater, 1983, A steerable dual-channel microwave radiometer for measurement of water vapor and liquid in the troposphere, *J. Clim. & Appl. Meteor.*, **22**(5), 789—806.
- [3] Mark Heggli, Robert M. Rauter and J. B. Snider, 1987, Field evaluation of a dual-channel microwave radiometer designed for measurements of integrated water vapor and cloud liquid water in the atmosphere, *Jour. of Atmos. and Oceanic Tech.*, **4**(1), 204—213.
- [4] 魏重、忻妙新、王普才等, 1992, 西太平洋赤道海域上空可降水和云液态水的遥感分析, *大气科学*, **16**(1), 1—10.
- [5] Alishouse J. C., et al., 1990, Determination of cloud liquid water content using the SSM/1, *IEEE Transactions on Geoscience and Remote Sensing*, **28**(5), 817—822.
- [6] Hoel, A. E. & R. W. Kennard, 1970, Ridge regression biased estimation for non-orthogonal problems, *Technometrics*, **12**, 55—67.
- [7] Hoel, A. E. & R. W. Kennard, 1970, Ridge regression application to non-orthogonal problems, *Technometrics*, **12**, 69—82.
- [8] 陈希孺, 王松桂, 1989, 近代回归分析, 安徽教育出版社.

A Universal Regressive Relationship for Ground-Based Microwave Remote Sensing of Atmospheric Precipitable Water Vapor

Lu Daren, Wei Chong and Xin Miaoxin

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

Feng Shiyong Wang Xiaogang

(Institute of Systematical Sciences, Chinese Academy of Sciences)

Abstract

Several universal equations to apply to all local places all over the world are obtained for ground-based dual-frequency microwave radiometer remote sensing the precipitable water in clear atmosphere by combining regression analysis and ridge regression techniques based on a set of a priori radiosonde data in typical seasons and at some stations typical of climate and sea level elevation. Many kinds of numerical testing for the resultant equations are conducted and the results show that the equations have much higher accuracies and can be put in practice. In addition, the comparison of the results for the frequency-pair of 20.6 and 31.65 GHz to those of 22.2 and 35.0 GHz is given.

Key words: Microwave remote sensing; Precipitable remote sensing; Retrieval method; Ridge regression; Regression analysis.