

文章编号 1004-924X(2016)06-1529-11

## 基于快速局域线性回归的 IRAS/FY-3B 大气温湿廓线反演

王忠一, 蒋耿明\*

(复旦大学 电磁波信息科学教育部重点实验室, 上海 200433)

**摘要:**提出一种快速的局域线性回归(Fast Locally Linear Regression, FLLR)算法,用于从搭载在风云三号 B 星(FY-3B)上的红外大气探测仪(IRAS)红外观测数据反演大气温湿廓线。算法所需的观测样本为 IRAS/FY-3B L1 数据红外观测值与 AIRX2RET V5 产品的时空匹配数据,以 2011 年为例,在  $180^{\circ}\text{W}\sim 180^{\circ}\text{E}$ 、 $60^{\circ}\text{N}\sim 60^{\circ}\text{S}$  的研究区域内按照观测时间绝对差小于 15 min 和观测角度绝对差小于  $2^{\circ}$  的条件获取观测样本,并对样本进行了评价。在匹配观测样本的基础上比较分析了 FLLR 算法与 LLR 算法、D 矩阵算法和非线性的神经网络算法,然后采用 FLLR 算法从 IRAS/FY-3B L1 数据反演得到 2011 年全年的大气温湿廓线,并外推反演得到 2012 年第 1 季度的大气温湿廓线。最后,利用相应的 ECMWF 再分析数据和 RAOB 探空观测对 2011 年的反演结果进行了精度验证,采用 AIRX2RET V5 产品对 2012 年第 1 季度的外推反演结果进行了验证。结果显示:与 D 矩阵算法相比,FLLR 算法反演大气温度和湿度廓线的均方根误差分别减小  $\sim 0.8\text{ K}$  和  $\sim 0.5\text{ g/kg}$ ,其精度与非线性的神经网络算法相当;相对于 ECMWF 再分析数据,本文大气温度和湿度廓线反演结果的均方根误差分别小于  $2.5\text{ K}$  和  $2.3\text{ g/kg}$ ;而相对于 RAOB 数据,其均方根误差分别小于  $3.5\text{ K}$  和  $2.0\text{ g/kg}$ ;2012 年第一季度外推反演结果的均方根误差分别小于  $2.5\text{ K}$  和  $1.6\text{ g/kg}$ ,与算法精度基本一致。IRAS/FY-3B 大气温湿廓线的反演精度与 MOD07 V5 大气廓线产品相当。

**关键词:**风云三号 B 星;红外大气探测仪;快速局域线性回归算法;大气温湿廓线反演;精度验证

**中图分类号:** TP394.1; TH691.9 **文献标识码:** A **doi:** 10.3788/OPE.20162406.1529

## Inversion of IRAS/FY-3B atmospheric temperature and humidity profiles based on fast locally linear regression

WANG Zhong-yi, JIANG Geng-ming\*

(Key Laboratory for Information Science of Electromagnetic Waves  
(Ministry of Education), Fudan University, Shanghai 200433, China)

\* Corresponding author, E-mail: jianggm@fudan.edu.cn

**Abstract:** The FLLR algorithm was developed to retrieve atmospheric temperature and humidity profiles from IRAS/FY-3B data. The observation samples required by the algorithm are IRAS /FY-3B L1 observation value and time-spacing matching data of AIRX2RET V5. Take the year of 2011 for an example, the observation sample was obtained in the survey region of  $180^{\circ}\text{W}\sim 180^{\circ}\text{E}$  and  $60^{\circ}\text{N}\sim 60^{\circ}\text{S}$  under the condition that absolute difference of the observation time is less than 15min and absolute

收稿日期:2016-02-19;修订日期:2016-04-15.

基金项目:国家自然科学基金项目(No. 41271012, No. 61331020)

difference of the observation angle is lower than  $2^\circ$ , and then an evaluation on the sample was conducted. Based on the observation sample matching, a comparison and analysis was carried out on FLLR algorithm and LLR algorithm, D Matrix algorithm and non-linear neural network algorithm. Then the year-round atmospheric temperature and humidity profiles in 2011 were obtained from IRAS/FY-3B L1 data inversion by using the FLLR algorithm, and the profile in the first quarter of 2012 can be achieved through extrapolation inversion. The last step was a precision test on the inversion results of 2011 by means of corresponding ECMWF reanalysis data and RAOB radiosonde observation, as well as verification on the extrapolation inversion results in the first quarter of 2012. It demonstrates that compared with D Matrix algorithm, the root mean square error in the inversion of the atmospheric temperature and humidity profile by using the FLLR algorithm has respectively reduced by  $\sim 0.8\text{K}$  and  $\sim 0.5\text{g/kg}$ , while its precision is quite the same as the non-linear neural network algorithm. With respect to ECMWF reanalysis data, the root mean square error of the atmospheric temperature and humidity profile inversion is respectively lower than  $2.5\text{K}$  and  $2.3\text{g/kg}$ , and for RAOB data, the root mean square error is lower than  $3.5\text{K}$  and  $2.0\text{g/kg}$  respectively. The root mean square error of the extrapolation inversion in the first quarter of 2012 is respectively lower than  $2.5\text{K}$  and  $1.6\text{g/kg}$ , and the precision is almost the same in different algorithms. The inversion precision of IRAS/FY-3B atmospheric temperature and humidity profile is quite the same as the precision of MOD07 V5 atmospheric profile.

**Key words:** FY-3B; IRAS; fast locally linear regression algorithm; atmospheric temperature and humidity profile inversion; accuracy validation

## 1 引 言

大气温湿廓线是描述大气热力和动力状态必不可少的参数。探测全球大气温湿度分布及其变化,对天气预报和气象保障工作,特别是数值天气预报和气候变化研究具有重要意义。然而,由于时空限制,常规的无线电探空观测难以满足业务和科研工作需求,而卫星大气探测能弥补地面探空资料的不足,是获得全球大气温湿廓线的最佳选择<sup>[1]</sup>。

为了从卫星传感器观测到的辐射值中得到大气温湿廓线,需要一种快速有效的反演方法。目前,大气温湿廓线反演方法可以分为两类:物理方法和统计方法。1)物理方法是通过求解辐射传输方程,试图直接由卫星观测辐射值推导出大气温湿廓线<sup>[2-6]</sup>。由于反演过程的不稳定性,物理方法通常利用迭代和优化代价函数的方法反演大气廓线。另外,物理反演方法需要较好地估计初始廓线和快速精确的辐射传输模式,同时地表温度和比辐射率等参数也要作为反演输入,它们对大气窗区通道的卫星模拟观测有较大的影响,然而这些参数很难精确获得。2)统计方法,利用已知的

大气温湿廓线和其对应的卫星观测辐射,直接建立二者之间的回归方程。如线性回归<sup>[7]</sup>、特征向量法<sup>[8]</sup>、和人工神经网络方法<sup>[9-11]</sup>等。与物理反演方法相比,统计方法的优点在于算法简单,计算速度快,计算稳定性强,便于处理大量数据,从而得到了广泛应用。

中国风云三号(FengYun 3, FY-3)是我国第二代极轨气象卫星,红外大气探测仪(InfraRed Atmospheric Sounder, IRAS)是 FY-3 系列卫星载荷中观测地球天气和环境的主要仪器,其覆盖的光谱范围为  $0.69\sim 15.5\ \mu\text{m}$ ,它可提供从可见光到红外波长的多波段辐射观测,用于全球大气温湿廓线、臭氧含量和云参数的反演等。IRAS 前 20 个波段位于  $3.7\sim 15\ \mu\text{m}$  的红外区间,后 6 个波段位于  $0.69\sim 1.64\ \mu\text{m}$  的可见光和近红外区间。红外波段中,第 8、9、19 和 20 波段为大气窗区波段,可用于地表探测;第 11、12、13 波段位于水汽吸收带;第 6 和 7 波段位于水汽和  $\text{CO}_2$  混合吸收带;第 10 波段位于  $\text{O}_3$  吸收带,其他波段位于  $\text{CO}_2$  或  $\text{NO}_2$  吸收带<sup>[12]</sup>。根据波段参数,IRAS/FY-3 红外观测非常适合反演大气温度和湿度廓线。

本文尝试发展一种快速的局域线性回归算

法,用于从 IRAS/FY-3B L1 数据反演大气温度和湿度廓线。本文主要包含大气温湿廓线反演方法的发展、反演应用和精度验证、结论等内容。

## 2 大气温湿廓线反演方法

无论是温度廓线还是湿度廓线均与卫星观测亮温之间存在非线性关系,故使用全局线性回归算法(例如:D 矩阵算法)反演大气温湿廓线时存在较大误差<sup>[13-14]</sup>。局域线性回归(Locally Linear Regression, LLR)是非参数回归理论中回归函数估计的一种有效方法,能够较好地描述变量之间的非线性关系,并且克服了核估计在内点处收敛速度慢的问题,在经济学、医学、生物学、气象等领域得到一定的应用<sup>[15-16]</sup>。但是,LLR 算法非常耗时。为了提高其计算效率,本文对 LLR 算法进行了改进,以便用于从 IRAS/FY-3B 红外观测亮温反演大气温湿廓线。

### 2.1 快速局域线性回归

给定  $M$  组观测值  $(y_i, \mathbf{x}_i)$ , 其中  $i=1, 2, \dots, M$ ,  $y_i$  是大气某一层温度或湿度观测值,  $\mathbf{x}_i$  是由卫星观测值组成的  $N$  维向量,  $\mathbf{x}_i$  和  $y_i$  满足:

$$y_i = r(\mathbf{x}_i) + \epsilon_i, \quad (1)$$

其中:  $r$  是回归函数,  $\epsilon_i$  是一个零均值的误差项。对于任何一个观测向量  $\mathbf{x}$ , 假设  $\mathbf{z}$  在  $\mathbf{x}$  的邻域, 对  $r(\mathbf{z})$  以  $\mathbf{x}$  为中心进行一阶泰勒级数展开, 可以得到:

$$r(\mathbf{z}) = r(\mathbf{x}) + \sum_{j=1}^N \frac{\partial r(\mathbf{x})}{\partial x_j} (z_j - x_j), \quad (2)$$

令  $\beta_0 = r(\mathbf{x})$ ,  $\beta_j = \partial r(\mathbf{x}) / \partial x_j$ , ( $j=1, 2, \dots, N$ ), 那么  $r(\mathbf{z})$  可以表示为:

$$r(\mathbf{z}) = \beta_0 + \sum_{j=1}^N \beta_j (z_j - x_j). \quad (3)$$

可以通过加权最小二乘方法找到一组  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_N)^T$ , 使得估计值  $r(\mathbf{z})$  和观测值  $y$  之间的加权均方误差  $E$  最小:

$$E = \sum_{i=1}^M [y_i - \beta_0 - \sum_{j=1}^N \beta_j (x_{ij} - x_j)]^2 K_H(\mathbf{x}_i - \mathbf{x}), \quad (4)$$

式中  $K_H(\mathbf{x}_i - \mathbf{x})$  是加权函数或核函数, 一般可选用如下的高斯核函数:

$$K_H(\mathbf{x}_i - \mathbf{x}) = \frac{1}{2\pi h_1 h_2 \dots h_N} \exp\left[-\frac{1}{2}(\mathbf{x}_i - \mathbf{x})^T \mathbf{H}^{-1}(\mathbf{x}_i - \mathbf{x})\right], \quad (5)$$

其中  $\mathbf{H} = \text{diag}\{h_1, h_2, \dots, h_N\}$ ,  $h_1, h_2, \dots, h_N$  为各预报因子的带宽, 可用下式计算<sup>[17]</sup>:

$$h_i = \left\{ \frac{4}{(N+2)M} \right\}^{1/(N+4)} \times \sigma_i, \quad (6)$$

其中  $\sigma_i$  为每个输入参数的标准方差。

$$\text{令 } \mathbf{X}_D = \begin{bmatrix} 1 & x_{11} - x_1 & x_{12} - x_2 & \dots & x_{1N} - x_N \\ 1 & x_{21} - x_1 & x_{22} - x_2 & \dots & x_{2N} - x_N \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{M1} - x_1 & x_{M2} - x_2 & \dots & x_{MN} - x_N \end{bmatrix},$$

$\mathbf{Y}^T = [y_1, y_2, \dots, y_M]$ ,  $\mathbf{W} = \text{diag}\{K_H(x_1 - x), K_H(x_2 - x), \dots, K_H(x_M - x)\}$  是一个  $M \times M$  维的对角矩阵, 则式(4)可用矩阵的形式表示为:

$$E = (\mathbf{Y} - \mathbf{X}_D \boldsymbol{\beta})^T \mathbf{W} (\mathbf{Y} - \mathbf{X}_D \boldsymbol{\beta}). \quad (7)$$

对式(7)最小化, 可得  $\boldsymbol{\beta}$  的加权最小二乘估计:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}_D^T \mathbf{W} \mathbf{X}_D)^{-1} \mathbf{X}_D^T \mathbf{W} \mathbf{Y}. \quad (8)$$

由上可知,  $r(\mathbf{x})$  的局部线性回归估计为  $\hat{\boldsymbol{\beta}}$ , 而  $\hat{\boldsymbol{\beta}}$  的其它分量是函数  $r(\mathbf{x})$  的一阶偏导数, 即

$$\hat{y} = \hat{\boldsymbol{\beta}}_0 = \mathbf{e}_1^T (\mathbf{X}_D^T \mathbf{W} \mathbf{X}_D)^{-1} \mathbf{X}_D^T \mathbf{W} \mathbf{Y}, \quad (9)$$

其中,  $\mathbf{e}_1$  是一个  $(N+1) \times 1$  的向量, 第一个元素值为 1, 其余的等于 0, 即:  $[\mathbf{e}_1]^T = [1, 0, \dots, 0]$ 。

令  $\mathbf{L}(\mathbf{x}) = \mathbf{e}_1^T (\mathbf{X}_D^T \mathbf{W} \mathbf{X}_D)^{-1} \mathbf{X}_D^T \mathbf{W}$ , 是一个  $1 \times M$  维的行向量, 称为权函数, 则(9)可以改写成:

$$\hat{y} = \mathbf{L}(\mathbf{x}) \mathbf{Y}, \quad (10)$$

上式即为 LLR 算法<sup>[18]</sup>。LLR 算法的关键是通过先验观测样本  $[x_1, x_2, \dots, x_M]$  和待估计观测样本  $x$  求得权函数  $\mathbf{L}(\mathbf{x})$ 。分析算法可知, 计算  $\mathbf{L}(\mathbf{x})$  的时间复杂度为  $O(M^2)$ , 即计算代价是先验样本数目的平方倍。因此, 当  $M$  较大时, LLR 算法的计算量则较大, 严重制约了其实际应用。

如果用 LLR 算法反演大气温湿廓线, 则有必要对其进行改进, 以减少计算代价。一种可能的改进方法是对式(4)直接进行最小化, 而不是对矩阵形式的式(7)进行最小化。对于式(4), 令  $\partial E / \partial \beta_j = 0$ ,  $j=0, 1, 2, \dots, N$ , 可以得到一个方程组:

$$\begin{cases} a_{11} \beta_0 + a_{12} \beta_1 + \dots + a_{1m} \beta_N = b_1 \\ a_{21} \beta_0 + a_{22} \beta_1 + \dots + a_{2m} \beta_N = b_2 \\ \vdots \\ a_{m1} \beta_0 + a_{m2} \beta_1 + \dots + a_{mn} \beta_N = b_m \end{cases}, \quad (11)$$

其中:  $m = N+1$ ,  $b_1, b_2, \dots, b_m$  和  $a_{jk}$  可以表示为:

$$\begin{cases} b_j = \sum_{i=1}^M y_i x_{Dij} K_H(\mathbf{x} - \mathbf{x}_i) \\ a_{jk} = \sum_{i=1}^M x_{Dij} x_{Dik} K_H(\mathbf{x} - \mathbf{x}_i) \end{cases}, \quad (12)$$

其中  $x_{Dij}$  表示设计矩阵  $\mathbf{x}_D$  的第  $i$  行第  $j$  列的元素。

$$\text{令 } \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{bmatrix}, \text{ 方程组可以表}$$

示为  $\mathbf{A}\mathbf{b} = \mathbf{b}$ , 计算  $\mathbf{A}$  的伪逆可得 LLR 的估计:

$$\hat{\mathbf{y}} = \hat{\mathbf{b}}_0 = \mathbf{e}_1^T \cdot \text{pinv}(\mathbf{A}) \cdot \mathbf{b}. \quad (13)$$

令  $\mathbf{Q} = \mathbf{e}_1^T \cdot \text{pinv}(\mathbf{A}) = [q_1, q_2, \dots, q_m]$ , 那么

$$\begin{aligned} \hat{\mathbf{y}} = \hat{\mathbf{b}}_0 &= q_1 \cdot b_1 + q_2 \cdot b_2 + \cdots + q_m \cdot b_m = \\ &= \sum_{i=1}^M [q_1 + q_2(x_{i1} - x_1) + \cdots + q_m(x_{iN} - \\ & x_N)] K_H(\mathbf{x}_i - \mathbf{x}) \cdot y_i = \sum_{i=1}^M l_i(\mathbf{x}) \cdot y_i, \end{aligned} \quad (14)$$

其中  $l_i(\mathbf{x})$  为权值函数, 可表示为:

$$l_i(\mathbf{x}) = [q_1 + q_2(x_{i1} - x_1) + \cdots + q_m(x_{iN} - x_N)] K_H(\mathbf{x}_i - \mathbf{x}). \quad (15)$$

对式(12)分析可知, 计算  $\mathbf{A}$  和  $\mathbf{b}$  分别需要  $2m^2 \times M$  和  $2m \times M$  次乘法运算, 对于确定的卫星传感器,  $m$  是固定值, 且通常远远小于  $M$ 。因此, 改进后的 LLR 算法时间复杂度为  $O(M)$ , 仅为改进前的  $1/M$ , 极大地缩短了计算时间。故称改进后的 LLR 算法为快速局域线性回归 (Fast LLR, FLLR) 算法。

## 2.2 样本获取和评价

若要发展 FLLR 算法, 则需要大量具有代表性, 且在时空上与 IRAS/FY-3B 观测匹配的大气廓线。气象站点的无线电探空可以提供大气温湿廓线数据, 例如 RAOB 数据 (<http://www.esrl.noaa.gov/raobs>), 但站点在全球分布稀疏且不均匀, 一般只提供 0 点和 12 点 (UTC 时间) 的数据; 无线电探空数据与 IRAS/FY-3B 观测在时空上匹配点非常少, 时空代表性差, 难以满足发展算法的要求。另一种可能的选择是数值预报场或卫星反演的大气温湿廓线产品。相对于探空数据, 该类数据时空均匀性好, 容易获得较多的匹配数据。常用的 AIRX2RET V5 产品是用物理方法从 AIRS 和 AMSU 数据反演得到的, 包含 50~1 000 hPa 内 14 个气压层 (1000, 925, 850, 700,

600, 500, 400, 300, 250, 200, 150, 100, 70, 50 hPa) 的大气温度和湿度廓线, 并具有较高的精度<sup>[19]</sup>。因此, 以 2011 年为例, 本研究选择 AIRX2RET V5 产品作为真值廓线, 选择  $180^\circ\text{W} \sim 180^\circ\text{E}$ 、 $60^\circ\text{N} \sim 60^\circ\text{S}$  的范围为研究区域。

IRAS/FY-3B 数据 (<http://fy3.satellite.cma.gov.cn>) 的星下点分辨率为 17 km, 而 AIRX2RET V5 数据 (<http://reverb.echo.nasa.gov>) 的星下点分辨率约为 50 km。为了实现两种数据空间上的配准, 本文用面积加权方法将 IRAS/FY-3B L1 数据重采样到 AIRX2RET 数据空间, 即:

$$V = \sum_{i=1}^N V_i S_i / \sum_{i=1}^N S_i, \quad (16)$$

其中:  $V$  为输出值,  $V_i$  为输入值,  $S_i$  为 IRAS/FY-3B 像元与 AIRX2RET 像元交集的面积,  $N$  为输入值的个数。除了 IRAS/FY-3B 观测亮温外, 重采样还包括观测角度和观测时间等参数。

经过重采样后, IRAS/FY-3B 数据与 AIRX2RET 数据在空间位置上一致。在保证足够数量匹配点对和精度的前提下, 本文的匹配条件是同地点、观测时间绝对差小于 15 min ( $|\Delta\text{time}| < 15'$ ) 和观测角度绝对差小于  $2^\circ$  ( $|\Delta\theta| \leq 2^\circ$ )。根据文献[4]方法和 AIRX2RET 廓线相对湿度小于 90% 的要求去除有云的观测。另外, 为了减轻大气不均匀的影响, 在一个重采样像元中, 要求参与重采样的 IRAS/FY-3B 观测亮温相对于加权均值的标准方差小于 5.0 K。

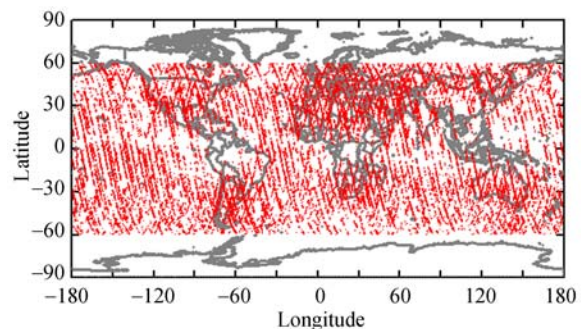


图 1 IRAS/FY-3B 观测值与 AIRX2RET V5 产品匹配样本的空间分布

Fig. 1 Spatial distribution of matching samples between IRAS/FY-3B measurements and AIRX2RET V5 product

IRAS/FY-3B 属于交叉轨道扫描探空仪, 在

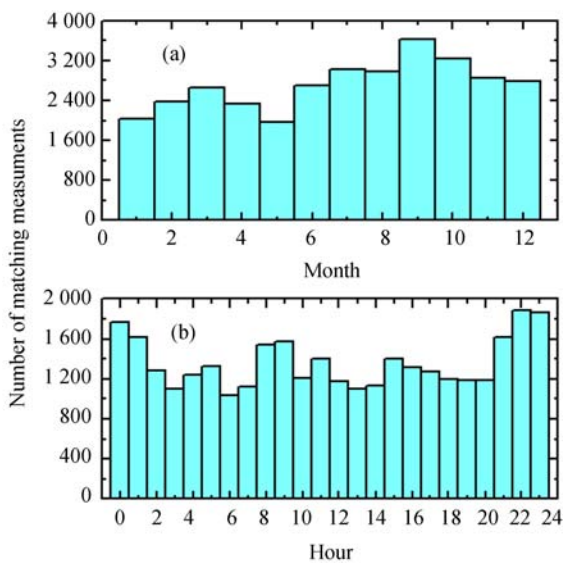


图2 IRAS/FY-3B 观测值与 AIRX2RET 廓线匹配样本的时间分布 (a) 月分布 (b) 时分布

Fig. 2 Temporal distribution of matching samples between IRAS/FY-3B measurements and AIRX2RET V5 product (a) with month (b) with hour

每条扫描线上,从星下点至两边,扫描点的观测天顶角逐渐增加,相应的光学路径变长。光学路径对红外测量值具有较大的影响<sup>[13]</sup>。如果把不同角度的观测值放在一起发展算法会降低廓线反演的精度,因此有必要根据角度进行分类讨论。根据 AIRX2RET V5 产品的 15 个观测角度,将匹配数据分为 15 组,每组数据分别发展反演算法。以第一组匹配数据为例(观测角度为  $1.8^\circ$ ),共获得 10 486 组陆地匹配样本和 22 095 组海洋匹配样本,其空间和时间分布分别显示在图 1 和图 2 中。由图可见,匹配样本在空间和时间上的分布较为均匀。

大气温湿廓线反演的输入参数选择非常关键<sup>[20]</sup>。在本文的研究中,IRAS/FY-3B 第 1~20 波段的观测亮温无疑是主要的输入参数。IRAS/FY-3B 第 19 和 20 波段位于  $3\sim 4\ \mu\text{m}$  的中红外区域,白天受太阳光和 aerosol 影响较大。气溶胶吸收辐射后释放能量,对大气加热,影响地区和全球辐射平衡,因此受气溶胶影响大的波段不适合用于红外大气温湿廓线反演<sup>[21-22]</sup>。根据交叉辐射定标方法<sup>[23-25]</sup>,IRAS/FY-3B 第 1 波段的观测值相对于 IASI/Metop-A 的观测值波动较大,且受太阳照射的影响,也不适合用于大气温湿廓线反

演。另外,增加纬度和海拔参数可以提高大气温湿廓线的反演精度<sup>[20,26]</sup>。因此,本文选择的输入参数包括 IRAS/FY-3B L1 数据给出的第 2~18 波段观测亮温、纬度和海拔。

为了评价匹配样本的代表性和反演算法的泛化能力,按均匀分布将每组匹配样本分为互补的 10 个子样本集,每次选 9 个子集用于建立反演算法,剩下的 1 个子样本集用作算法精度评价,这样可以建立 10 个反演算法,保证了所有的数据都用于算法建立和算法验证,且验证数据和算法发展数据是相对独立的。

同时,为了对比说明本文所发展的 FLLR 算法的优势,用同样的分组匹配样本发展了 LLR 算法、D 矩阵算法和 3 层 BP 神经网络算法,其余几种算法的输入和输出参数与 FLLR 算法一致。根据文献[27]中的计算公式和本文的实验结果可知,神经网络隐含层节点数为 31。神经网络的权值和偏置用 Nguyen - Widrow 方法初始化<sup>[9]</sup>。隐含层和输出层的激励函数都用双曲正切函数。利用量化共轭梯度方法(Scaled Conjugate Gradient)训练神经网络,最大训练次数设为 10 000 次,在训练过程中检测 10% 评价样本的误差,当评价样本的误差不再减小时,停止训练。

相对于验证样本集,FLLR 算法、LLR 算法、D 矩阵算法和 NN 算法反演大气温度和湿度廓线的误差分别绘制在图 3 和图 4 中。结果显示,4 种算法反演温度和湿度廓线的偏差分别小于 0.2 K 和 0.2 g/kg,这说明验证样本可以用其余的样本较好地表示,本文所获取的样本集具有较好的广泛性和代表性,在其基础上发展出来的算法也将具有较好的泛化能力。

另外,由实验数据还可得知,FLLR 算法和 LLR 算法的反演误差完全相同,但随着验证样本数  $M$  的增加,FLLR 算法的计算时间优势越来越明显。如表 1 所示,当  $M$  等于 1 000 时,FLLR 算法耗时约为 LLR 算法的 38.7%,而当  $M$  等于 10 000 时,FLLR 算法耗时仅为 LLR 算法的 3.7%。FLLR 算法在反演大气温度廓线时,所有层的偏差在  $\pm 0.2\ \text{K}$  以内,陆地和海洋的均方根误差分别小于 2.1 K 和 1.8 K,其中,陆地上空 200~800 hPa 区间的均方根误差在 1.0~1.5 K 之间变化,海洋上空 300~1 000 hPa 区间的均方根误差约为 1.0 K。FLLR 算法在反演大气湿度

廓线时,所有层的偏差在 $\pm 0.2 \text{ g/kg}$ 之内,陆地和海洋的均方根误差分别小于 $1.6 \text{ g/kg}$ 和 $1.4 \text{ g/kg}$ ,并随气压的降低而减少,直至 $250 \text{ hPa}$ 时均方根误差接近零。海洋上空反演误差小的主要原因是海洋表面起伏小,辐射特性相对均匀。无论是反演大气温度廓线还是反演大气湿度廓线,

FLLR算法与NN算法的反演精度均相当,但与D矩阵算法相比,FLLR算法反演大气温度廓线的精度提高了 $0.3\sim 0.8 \text{ K}$ ,反演大气湿度廓线的精度最大提高了 $0.5 \text{ g/kg}$ 。此结果再次验证了大气温湿廓线与卫星观测亮温之间存在非线性关系,而FLLR算法能较好地描述这种非线性关系。

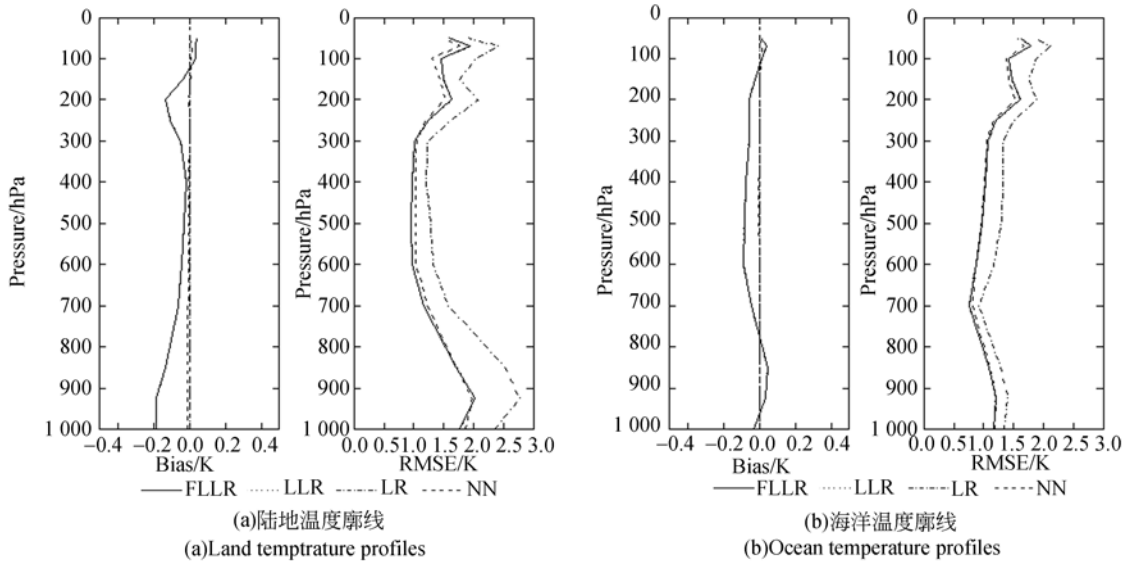


图 3 大气温度廓线反演误差比较

Fig. 3 Comparison of temperature retrieval biases and RMSEs for different methods

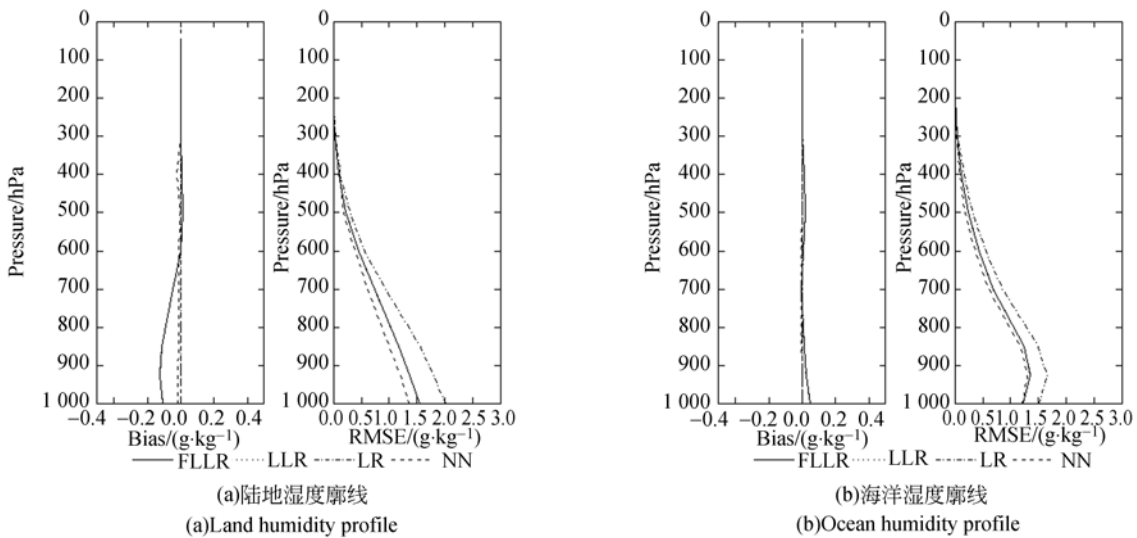


图 4 大气湿度廓线反演误差比较

Fig. 4 Comparison of retrieval errors for atmospheric humidity profile between different methods

表 1 快速局域线性回归与局域线性回归的计算时间比较

Tab. 1 Comparison of computation times between FLLR and LLR (s)

算法	样本数=1 000	样本数=5 000	样本数=10 000
局域线性回归	0.62	12.63	45.05
快速局域线性回归	0.24	0.84	1.68

\* 计算机配置: Intel 酷睿 I5-3470, 3.2 GHz 主频, 8 GB 内存, Windows7 操作系统

### 3 大气温湿廓线反演应用和精度验证

由上节分析可知,在反演大气温湿廓线性能上,本文发展的 FLLR 算法优于 D 矩阵算法,与神经网络算法相当。因此,基于 FLLR 算法和匹配样本数据,从 2011 年全年的 IRAS/FY-3B L1 红外数据反演得到整个研究区域内的大气温湿廓线,其中根据文献[4]所述方法去除有云的 IRAS/FY-3B 红外观测数据。

为评价 IRAS/FY-3B L1 数据反演得到的大气温湿廓线的精度,本文分别用 ECMWF (European Center for Medium-Range Weather Forecasts)再分析数据和 RAOB 探空数据对反演结果进行精度验证。

#### 3.1 相对于 ECMWF 再分析数据的精度验证

欧洲中尺度天气预报中心生产的 ECMWF 再分析数据可提供 0、6、12 和 18 点共 4 个格林威治时刻的气压、大气温度、大气相对湿度和重力势

等大气参数,具有较高的精度 (ECMWF report, 1995)。本文购买了研究区域内 2011 年 3、6、9 和 12 月共 4 个月份、空间分辨率为  $0.25^{\circ} \times 0.25^{\circ}$  的 ECMWF 再分析数据用于验证本研究的反演精度。首先,从 ECMWF 再分析数据中提取气压、大气温度和相对湿度,并把相对湿度转换成以 g/kg 为单位的水汽混合比<sup>[28]</sup>;然后,将本文从 IRAS/FY-3B L1 数据反演得到的大气温湿廓线按面积加权的方式重采样至 ECMWF 再分析数据网格空间;最后,根据同地点和观测时间绝对差小于 10 min 两个条件提取星下点大气温湿廓线匹配数据。为了进一步去除云的影响,本文对匹配结果按下述条件进行筛选:1) IRAS/FY-3B 第 9 波道观测亮温大于 270 K;2) ECMWF 廓线每一层的大气相对湿度不大于 90%。按上述匹配条件,总共获得 5 719 对陆地上空匹配数据和 8 268 对海洋上空匹配数据。本文的反演结果相对于 ECMWF 再分析数据的误差绘制在图 5 中。对于大气温度廓线,陆地上空廓线的偏差为  $-0.2 \sim 1.4$  K,海洋上空廓线的偏差为  $\pm 0.6$  K,均方根误差整体小于 2.6 K,其中在 700~1 000 hPa 区间海洋上空廓线的均方根误差比陆地上空廓线减小了  $\sim 0.5$  K。对于大气湿度廓线,其偏差为  $-0.05 \sim 1.0$  g/kg,主要呈现正偏差,说明反演值高于 ECMWF 数据,而均方根误差整体小于 2.3 g/kg。

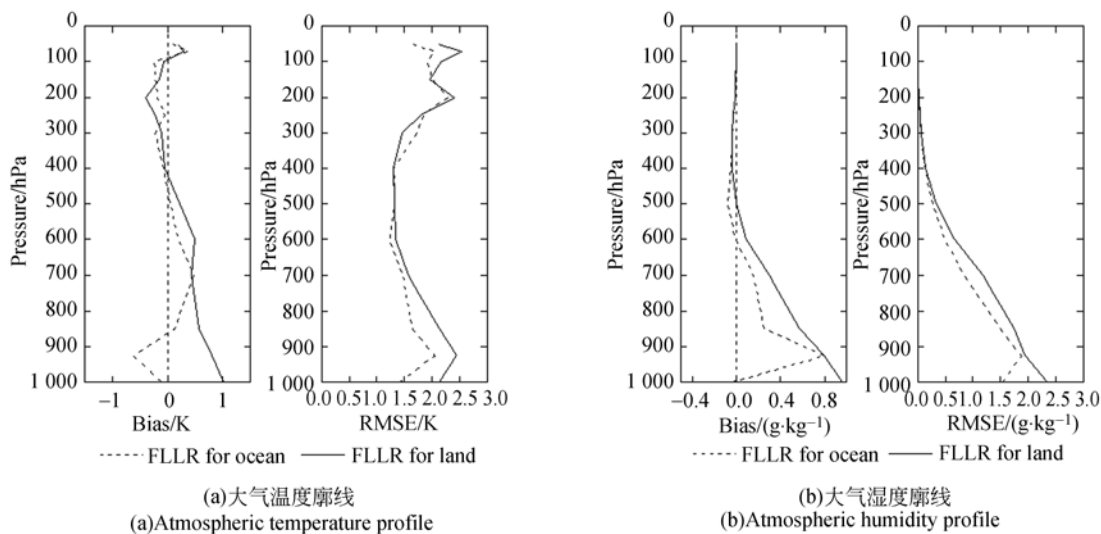


图 5 IRAS/FY-3B 反演大气温湿廓线相对于 ECMWF 再分析数据的误差

Fig. 5 Errors of atmospheric temperature and humidity profiles retrieved from IRAS/FY-3B measurements in contrast to ECMWF reanalysis data

### 3.2 相对于 RAOB 探空数据的精度验证

RAOB 数据为全球气象站点探空气球实测数据,本文下载了全球 2011 年的 RAOB 数据用于验证本文的大气温湿廓线的反演精度。匹配条件为:1)RAOB 探空站点的地理位置(经纬度)落入 IRAS/FY-3B 的观测视场内;2)RAOB 探空数据的观测时间与 IRAS/FY-3B 的观测时间之差在 $[-0.2 \text{ h}, 0.8 \text{ h}]$ 区间。为了进一步去除云的影响,要求 RAOB 探空数据所有层的相对湿度小于 90%。由于满足条件的匹配数据太少,本文对观测角度的限制放大至  $20^\circ$ 。根据以上条件获得 490 组匹配数据,其中陆地 476 组,海洋 14 组。把满足匹配条件的 RAOB 探空数据用线性插值方法插值到本文反演廓线的气压层。

图 6 显示的是本文反演结果相对于 RAOB 探空数据的误差。对于大气温度廓线,偏差在  $\pm 1.0 \text{ K}$  之间;陆地上空廓线均方根误差整体小于  $3.5 \text{ K}$ ,其中在  $300 \text{ hPa}$  以下的空廓线小于  $2.8 \text{ K}$ ,海洋上空廓线均方根误差小于  $2.5 \text{ K}$ 。对于大气湿度廓线,陆地上空廓线偏差在  $\pm 0.3 \text{ g/kg}$  之间,而海洋上空廓线的偏差较大,在  $-0.1 \sim 1.2$

$\text{g/kg}$  之间,这可能与海洋上空的匹配数据太少有关;陆地和海洋上空大气湿度廓线的均方根误差均小于  $2.0 \text{ g/kg}$ 。本文的反演精度与 MOD07 大气廓线产品的精度相当<sup>[26]</sup>。

对于大气温度廓线反演结果在对流层底层误差较大,主要有两个原因:1)大气红外波段辐射穿透力较弱,容易被上层大气吸收;2)对流层底层大气易受地表辐射影响,不同地表类型之间的辐射特性差异较大,相比于陆地表面,海洋表面的辐射特性更为单一,其底层大气温度反演的误差也更小。另外大气温度廓线反演结果在对流层上层误差较大,主要原因如下:1)对流层顶上下大气差异较大,且对流层顶高度随季节和纬度变化较大,通常,利用卫星遥感得到的大气温度廓线的误差在该层附近较大;2)IRAS/FY-3B 仪器各波段对该层大气辐射的响应较弱;3)所选用的真值廓线 AIRX2RET V5 在这些层的精度不够。

大气湿度廓线反演的误差与湿度本身的垂直分布有关,湿度在  $250 \text{ hPa}$  以上的误差接近 0 的原因是高层大气湿度本身就接近于 0。

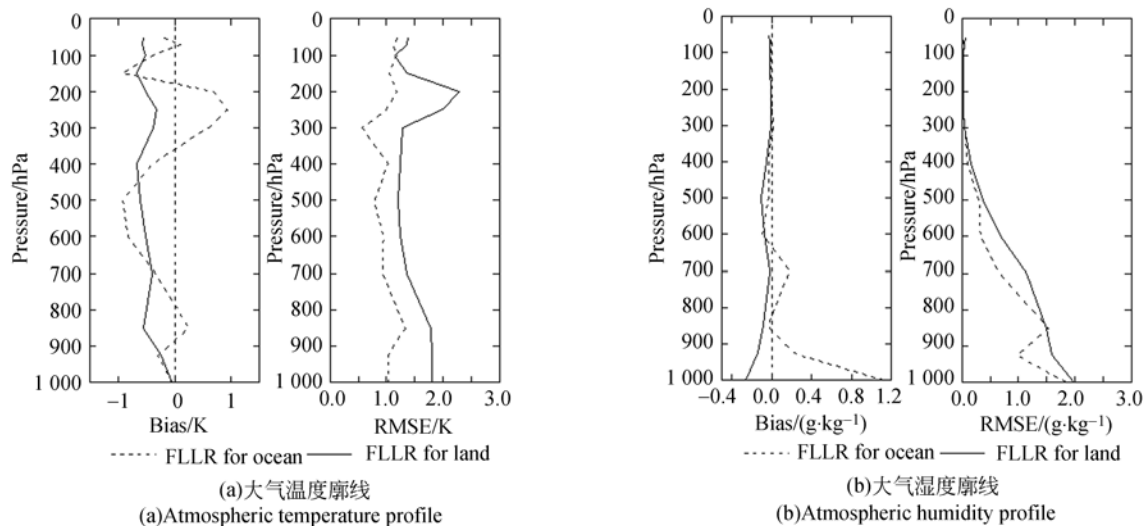


图 6 IRAS/FY-3B 反演结果相对于 RAOB 探空数据的误差

Fig. 6 Errors of atmospheric temperature and humidity profiles retrieved from IRAS/FY-3B measurements in contrast to RAOB radiosonde data

### 3.3 外推反演和精度验证

如第 2 部分所述,本文使用 2011 年的 IRAS/FY-3B 与 AIRX2RET V5 产品的匹配数据发展出 IRAS/FY-3B 大气温湿廓线反演算法。本文尝试把该算法应用到 2012 年第 1 季度的大气温

湿廓线外推反演,数据处理和匹配过程与第 2 部分所述完全一致。相对于 AIRX2RET V5 产品,2012 年第 1 季度 IRAS/FY-3B 反演温湿廓线的误差绘制在图 7 中。与图 3 和图 4 所示的算法验证误差相比,图 7 中外推反演结果的均方根误差

基本保持一致,而偏差有所增加。对于大气温度廓线反演结果,当气压小于 200 hPa 时,偏差为  $\pm 0.5$  K,在 200 hPa 以上,偏差为  $\pm 1.0$  K;对于大气湿度廓线反演结果,偏差为  $\pm 0.2$  g/kg,均方根

误差全部小于 1.6 g/kg。此结果说明本文发展的算法具有较好的代表性,短期的外推反演结果具有较高的精度。

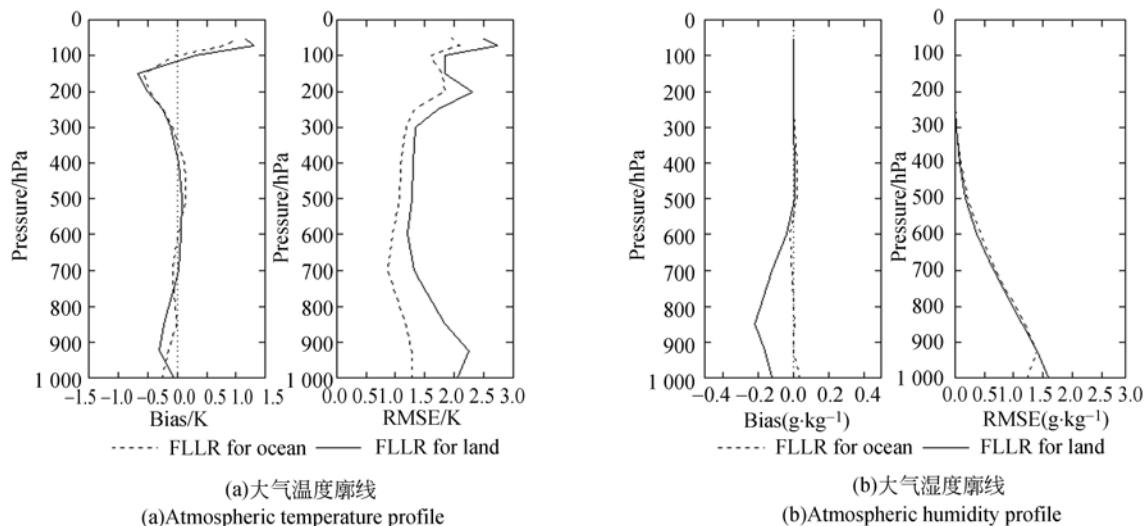


图7 2012年第1季度 IRAS/FY-3B 外推反演结果相对于 AIRX2RET 产品的误差

Fig. 7 Errors of atmospheric temperature and humidity profiles retrieved from IRAS/FY-3B infrared measurements relative to AIRX2RET product in the first quarter of 2012

## 4 结 论

本文发展出一种快速的局域线性回归算法用于从 IRAS/FY-3B 红外观测亮温反演大气温湿廓线,算法所需的观测样本为 IRAS/FY-3B 红外观测值与 AIRX2RET V5 产品的匹配数据。以 2011 年为例,在  $180^{\circ}\text{W}\sim 180^{\circ}\text{E}$ 、 $60^{\circ}\text{S}\sim 60^{\circ}\text{N}$  的研究区域内,本文用 IRAS/FY-3B 观测值与 AIRX2RET 产品的匹配观测样本对 FLLR 进行了分析,并与 D 矩阵算法和神经网络算法进行了比较。利用该算法从 IRAS/FY-3B L1 数据反演得到 2011 年的大气温湿廓线和外推反演得到 2012 年第一季度的大气温湿廓线,用 ECMWF 再分析数据和 RAOB 探空数据对 2011 年的反演结

果进行了精度验证,用 AIRX2RET V5 产品对 2012 年第一季度的外推反演结果进行了精度验证。结果表明:1) 本文发展的 FLLR 算法不仅拟合精度高而且计算速度快。与 D 矩阵算法相比,本文所发展的温度和湿度廓线反演算法的均方根误差分别减小了  $\sim 0.8$  K 和  $\sim 0.5$  g/kg,其精度与非线性的神经网络算法相当。2) 对于 ECMWF 再分析数据,从 2011 年 IRAS/FY-3B L1 数据反演得到的大气温度和湿度廓线的均方根误差分别小于 2.5 K 和 2.3 g/kg。而相对于 RAOB 数据,其均方根误差分别小于 3.5 K 和 2.0 g/kg。实际反演精度与 MOD07 V5 大气廓线产品相当。3) 2012 年第一季度外推反演结果的均方根误差分别小于 2.5 K 和 1.6 g/kg,与算法精度基本一致。

## 参考文献:

- [1] 王颖,黄勇,黄思源. 大气温湿廓线反演问题的研究[J]. 国土资源遥感,2008(1):23-26.  
WANG Y, HUANG Y, HUANG S Y. A preliminary study of the retrieval methods for atmosphere and humidity profiles [J]. *Remote Sensing For Land*

& Resources, 2008(1):23-26. (in Chinese)

- [2] CHAHINE M T. Inverse problems in radiative transfer: Determination of atmospheric parameters [J]. *Journal of the Atmospheric Sciences*, 1970, 27(6): 960-967.  
[3] RODGERS C D. Retrieval of atmospheric temperature and composition from remote measurements of

- thermal radiation [J]. *Reviews of Geophysics*, 1976, 14(4): 609-624.
- [4] LI J, WOLF W W, MENZEL W P, *et al.*. Global soundings of the atmosphere from ATOVS measurements: The algorithm and validation [J]. *Journal of Applied Meteorology*, 2000, 39(8): 1248-1268.
- [5] AUGUST T, KLASE D, SCHLUSSEL P, *et al.*. IASI on metop-A: operational level 2 retrievals after five years in orbit [J]. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 2012, 113(11): 1340-1371.
- [6] LIU C Y, LIU G R, LIN T H, *et al.*. Using surface stations to improve sounding retrievals from hyperspectral infrared instruments [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, 52(11): 6957-6963.
- [7] ROSENKRANZ P W, KOMICHAK M J, STAELIN P H. A method for estimation of atmospheric water vapor profiles by microwave radiometry [J]. *Journal of Applied Meteorology*, 1982, 21(9): 1364-1370.
- [8] SMITH W L, WOOLF H M. The use of eigenvectors of statistical covariance matrices for interpreting satellite sounding radiometer observations [J]. *Journal of the Atmospheric Sciences*, 1976, 33(7): 1127-1140.
- [9] BLACKWELL W J. A neural-network technique for the retrieval of atmospheric temperature and moisture profiles from high spectral resolution sounding data [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2005, 43(11): 2535-2546.
- [10] BLACKWELL W J, MILSTEIN A B. A neural network retrieval technique for high-resolution profiling of cloudy atmospheres [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014, 7(4): 1260-1270.
- [11] GANGWAR R K, MATHUR A K, GOHIL B S, *et al.*. Neural Network based retrieval of Atmospheric Temperature Profile using AMSU-A observations [J]. *International Journal of Atmospheric Sciences*, 2014, 2014:1-8.
- [12] QI C, CHEN Y, LIU H, *et al.*. Calibration and validation of the InfraRed Atmospheric Sounder onboard the FY3B satellite [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2012, 50(12): 4903-4914.
- [13] YAO Z G, CHEN H B, LIN L F. Retrieving atmospheric temperature profiles from AMSU-A data with neural networks [J]. *Advances in Atmospheric Sciences*, 2005, 22(4): 606-616.
- [14] GOLDBERG M D. Generation of retrieval products from AMSU-A: Methodology and validation [C]. *Tech. Proc. 10th Int. TOVS Study Conf.*, 1999: 215-219.
- [15] 薛留根. 现代非参数统计 [M]. 北京: 科学出版社, 2015.
- XUE L G. *Modern Nonparametric Statistics* [M]. Beijing: Science Press, 2015. (in Chinese)
- [16] 胡邦辉, 张惠君, 杨修群, 等. 基于非参数回归模型的局部线性估计云量预报方法研究 [J]. *南京大学学报: 自然科学版*, 2009, 45(1): 89-97.
- HU B H, ZHANG H J, YANG X Q, *et al.*. A cloud amount forecasting method based on local linear estimation of non-parametric regression model [J]. *Journal of Nanjing University (Natural Sciences)*, 2009, 45(1): 89-97. (in Chinese)
- [17] BOWMAN A W, AZZALINI A. *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations: The Kernel Approach with S-Plus Illustrations* [M]. New York: Oxford University Press, 1997.
- [18] WASSERMAN L. *All of nonparametric statistics* [M]. New York: Springer Science & Business Media, 2006.
- [19] FETZER E J, OLSEN E T. Validation of AIRS/AMSU/HSB retrieved products [J]. *Proceedings of SPIE*, 2003, 5(5): 211-219.
- [20] 董超华, 李俊, 张鹏, 等. 卫星高光谱红外大气遥感原理和应用 [M]. 北京: 科学出版社, 2013, 92-93, 98-103.
- DONG CH H, LI J, ZHANG P, *et al.*. *Principles and Applications of Satellite Hyperspectral Infrared Atmospheric Remote Sensing* [M]. Beijing: Science Press, 2013, 92-93, 98-103. (in Chinese)
- [21] 邵士勇, 梅海平, 黄印博, 等. 大气气溶胶等效吸收的研究 [J]. *红外与激光工程*, 2014, 43(4): 1057-1061.
- SHAO SH Y, MEI H P, HUANG Y B, *et al.*. Analysis of effective absorption for atmospheric aerosol [J]. *Infrared and Laser Engineering*,

- 2014, 43(4): 1057-1061. (in Chinese)
- [22] 徐梦春,徐青山. 气溶胶粒子特性和垂直分布对辐射的影响[J]. 红外与激光工程, 2016, 45(2): 211002-0211002(7).
- XU M CH, X Q SH. Effect of aerosol particle characteristic and vertical distribution on radiation [J]. *Infrared and Laser Engineering*, 2016, 45(2): 211002-0211002(7). (in Chinese)
- [23] 徐寒列,胡秀清,徐娜,等. FY-3C/可见光红外扫描辐射计中红外通道太阳污染的识别和修正[J]. 光学精密工程, 2015, 23(7): 1874-1879.
- XU H L, HU X Q, XU N, *et al.*. Discrimination and correction for solar contamination on min-infrared band of FY-3C/VIRR [J]. *Opt. Precision Eng.*, 2015, 23(7):1874-1879. (in Chinese)
- [24] 王玲,胡秀清,陈林. 基于多种亮度稳定目标的FY-3C/中分辨率光谱成像仪的反射太阳波段辐射定标[J]. 光学精密工程, 2015, 23(7): 1911-1920.
- WANG L, HU X Q, CHEN L. FY-3C/MERSI calibration for solar band using multi-reflectance stable targets [J]. *Opt. Precision Eng.* 2015, 23(7):1911-1920. (in Chinese)
- [25] 周蔚,蒋耿明. VIRR/FY-3A 分裂窗波段与 AIRS/AQUA 和 IASI/METOP-A 数据的交叉辐射定标 [J]. 光学精密工程, 2015, 23(7): 1892-1902.
- ZHOU W, JIANG G M. Cross-calibration of VIRR/FY-3A split-window channels with AIRS/AQUA and IASI/METOP-A data [J]. *Opt. Precision Eng.*, 2015, 23(7): 1892-1902. (in Chinese)
- [26] BORBAS E E, LI J, MENZEL W P, *et al.*. MODIS atmospheric profile retrieval algorithm theoretical basis document [EB/OL]. [http://eosps.nasa.gov/sites/default/files/atbd/atbd\\_mod07.pdf](http://eosps.nasa.gov/sites/default/files/atbd/atbd_mod07.pdf). 2006.
- [27] 张雪慧,官莉,王振会,等. 利用人工神经网络方法反演大气温度廓线[J]. 气象, 2009, 35(11): 137-142.
- ZHANG X H, GUAN L, WANG ZH H, *et al.*. Retrieving atmospheric temperature profiles using artificial neural network approach [J]. *Meteorological Monthly*, 2009, 35(11): 137-142. (in Chinese)
- [28] ALDYCHOY O A, ESKRIDGE R E. Improved Magnus form approximation of saturation vapor pressure [J]. *Journal of Applied Meteorology*, 1996, 35(4): 601-609.

#### 作者简介:



王忠一(1989—),男,云南保山人,2011年于云南大学获得学士学位,现为复旦大学信息科学与工程学院在读研究生,主要从事红外大气参数反演方面的研究。E-mail: zhongyiwang13@fudan.edu.cn

#### 导师简介:



蒋耿明(1975—),男,湖南道县人,博士,副教授,1997年于武汉测绘科技大学获得学士学位,2000年于中国科学院遥感应用研究所获得硕士学位,2007年于法国斯特拉斯堡第一大学获得博士学位,主要从事定量遥感的教学与科研工作。E-mail: jianggm@fudan.edu.cn

(本栏目编辑:李自乐)

(版权所有 未经许可 不得转载)