

WANG Yu, ZHANG Xianzhong, WU Tong, ZHANG Yijian, SUN Yue, LI Shijie, LI Xinqi, ZHONG Kai, YAN Zhaoai, XU Degang, YAO Jianquan.
Research on Atmospheric Temperature Retrieval Based on Rayleigh Lidar Using Optimal Estimation Method (in Chinese). *Chinese Journal of Space Science*, 2023, **43**(4): 627–639. DOI:10.11728/cjss2023.04.2022–0035

基于最优估计法的瑞利激光雷达 反演大气温度研究*

王 煜^{1,2} 张献中^{1,2} 吴 同^{1,2} 张益健^{1,2}
孙 悦^{1,2} 李世杰^{1,2} 李欣岐^{1,2} 钟 凯^{1,2}
闫召爱^{3,4} 徐德刚^{1,2} 姚建铨^{1,2}

1(天津大学精密仪器与光电子工程学院 天津 300072)

2(天津大学光电信息技术教育部重点实验室 天津 300072)

3(中国科学院国家空间科学中心 北京 100190)

4(中国科学院大学 北京 100049)

摘 要 基于瑞利激光雷达的回波光子信号对中层大气进行探测,结合最优估计法,对大气温度进行反演。本文基于瑞利散射激光雷达方程建立正向模型,选择大气模型的温度廓线作为先验状态信息,构建用于最优化处理的成本函数,利用 Levenberg-Marquardt 最优化算法对成本函数执行最优化处理,得到大气温度的反演结果,对反演结果的不确定度分析的同时利用平均核矩阵对反演结果中真实信息的贡献进行评估。利用瑞利激光雷达方程产生的模拟回波信号进行了大气温度的反演处理与分析,对中国科学院国家空间中心提供的瑞利激光雷达实测数据进行大气温度的最优估计反演。结果表明,90 km 以下的反演不确定度在 10 K 以内,且相较于 CH 方法,最优估计法具有反演有效范围高的优势;在回波光子信噪比较高的区域,反演不确定度较小,且真实信息对反演结果的贡献占主导地位。

关键词 中层大气,瑞利激光雷达,最优估计法,平均核矩阵

中图分类号 P356

Research on Atmospheric Temperature Retrieval Based on Rayleigh Lidar Using Optimal Estimation Method

WANG Yu^{1,2} ZHANG Xianzhong^{1,2} WU Tong^{1,2} ZHANG Yijian^{1,2}
SUN Yue^{1,2} LI Shijie^{1,2} LI Xinqi^{1,2} ZHONG Kai^{1,2}

* 国家自然科学基金项目资助(62175184)

2022-07-28 收到原稿,2022-11-20 收到修定稿

E-mail: 2020202055@tju.edu.cn. 通信作者 钟凯, E-mail: zhongkai@tju.edu.cn; 徐德刚, E-mail: xudegang@tju.edu.cn

©The Author(s) 2023. This is an open access article under the CC-BY 4.0 License
(<https://creativecommons.org/licenses/by/4.0/>)

YAN Zhaoai^{3,4} XU Degang^{1,2} YAO Jianquan^{1,2}¹(School of Precision Instrument and Optoelectronic Engineering, Tianjin University, Tianjin 300072)²(Key Laboratory of Optoelectronic Information Science and Technology (Ministry of Education),
Tianjin University, Tianjin 300072)³(National Space Science Center, Chinese Academy of Sciences, Beijing 100190)⁴(University of Chinese Academy of Sciences, Beijing 100049)

Abstract The middle atmosphere is detected based on the echo photon signal of Rayleigh lidar, combined with the optimal estimation method, the atmospheric temperature is retrieved. In this paper, the forward model is established based on the Rayleigh scattering lidar equation, the temperature profile of the atmospheric model is selected as the prior state information, and the cost function for optimization is constructed, and the cost function is optimized using the Levenberg-Marquardt optimization algorithm, get the inversion results of atmospheric temperature, and use the average kernel matrix to evaluate the contribution of the real information in the inversion results while analyzing the uncertainty of the inversion results. Using the simulated echo signal generated by the Rayleigh lidar equation, the inversion processing and analysis of the atmospheric temperature is carried out, and the optimal estimation inversion of the atmospheric temperature is carried out on the Rayleigh lidar measured data provided by the National Space Center of the Chinese Academy of Sciences. The results show that the inversion uncertainty below 90 km is within 10 K, and compared with the CH method, the optimal estimation method has the advantage of a higher inversion effective range; the inversion uncertainty is small, and the contribution of real information to the inversion results is dominant.

Key words Middle atmosphere, Rayleigh lidar, Optimal estimation method, Average kernel matrix

0 引言

中层大气的温度探测对于了解中层大气温度变化与低层大气相应变化的联系、监测地球气候变化有着重要意义。目前对于中层大气温度探测的主要手段有星载仪器、探空仪器和激光雷达等。星载仪器可以在相对较短的时间内对不同经纬度处的大气进行探测,但其探测成本高,时空分辨率低,无法实现连续观测。探空仪器的原理成熟,发射设施相对简单,成本低,时间弹性强,但观测时间有限。激光雷达是一种主动遥感探测仪器,其基本原理和普通微波雷达相似,具有时空分辨率高、测量精度高、能够连续观测等优点,具有很好的大气探测应用前景。瑞利散射激光雷达发射一定波长的脉冲激光至大气中,与大气分子碰撞发生瑞利散射,产生回波信号,通过接收并分析回波强度,可以反演得到大气相对密度、温度等参

数。利用瑞利散射激光雷达开展中层大气温度探测与研究已经取得了一定的进展。

Hauchecorne 等^[1]于 1980 年首次提出了基于瑞利散射激光雷达的大气温度反演方法,即 CH 方法。瑞利散射激光雷达的测量准确度与其使用的反演方法有一定的关系。多数瑞利散射激光雷达采用 CH 方法进行温度反演。CH 方法利用流体静力平衡假设、理想气体定律确定温度反演公式,并从探测范围顶部选取一个参考点向下进行积分。顶部温度值自身的不确定度会影响温度反演结果的准确性,距离参考点越远,反演温度受到的影响越小。因此,为了保证反演结果的准确度,研究通常舍弃观测范围顶部 10~15 km 的反演结果。2019 年 Liu 等^[2]采用由均匀搜索生成温度初值的方法,改进了 CH 方法,提高了温度反演的高度上限。本文利用最优估计法对大气温度进行反演。最优估计法 (Optimal Estima-

tion Method, OEM) 最初是 Rodgers^[3] 提出的用于大气遥感的反演方法, 该方法利用非线性回归来求解反演问题, 用以获得大气状态空间中的状态信息。其基于贝叶斯定理, 将测量数据、先验信息与相应的不确定度结合在一起, 找到大气状态的最优解^[4]。2014 年 Povey 等^[5] 将最优估计法应用于激光雷达领域, 并对对流层气溶胶后向散射系数和消光系数进行了反演, 初步验证了最优估计法应用于激光雷达领域的可行性。此后, Sica 等^[6,7] 对最优估计法在激光雷达领域的应用进行了进一步探索。

本文基于瑞利散射激光雷达给出了最优估计法中的贝叶斯基本原理、正向模型与成本函数的具体含义, 阐述了最优估计法在瑞利散射激光雷达中的应用。对最优估计法的应用进行了模拟仿真和验证, 并基于最优估计法对瑞利散射激光雷达实测数据进行了处理, 对大气温度反演结果进行了分析。

1 最优估计法的理论基础

1.1 瑞利散射激光雷达方程

瑞利散射激光雷达发射脉冲激光至大气中, 入射激光会与大气粒子发生作用, 产生瑞利散射光。激光雷达通过收集后向散射光将散射光转换为数字或模拟信号, 最终生成回波廓线。瑞利散射激光雷达方程将激光雷达观测到的后向散射光强度与激光雷达的系统参数和观测范围的大气有关特性等联系起来, 其方程如下^[8]:

$$N_R(z) = \left[\frac{P(\lambda) \Delta t}{hc/\lambda} \right] [\sigma_R(\lambda) n(z) \Delta z] \cdot \left[\frac{A}{(z - z_0)^2} \right] [\eta(\lambda) T^2(z) G(z)] + N_B. \quad (1)$$

式(1)由两部分组成, 其中第一部分由四项(中括号部分)因式组成。第一项中, $P(\lambda)$ 激光雷达的发射功率, Δt 为进行观测时总的积累时间, h 为普朗克常数, c 为光速, λ 为激光雷达发射波长, 该项表示的是激光雷达观测时发射的总光子数。第二项中, $\sigma_R(\lambda)$ 为瑞利后向散射截面, $n(z)$ 为相应高度的大气分子密度, Δz 为激光雷达观测时的空间分辨率, 该项表征的是激光入射到大气后与大气介质发生后向散射的概率。第三项中, A 为激光雷达接收望远镜的有效面积, $z - z_0$ 为目标高度 z 和激光雷达所在高度 z_0 的距离, 第

三项整体表示后向散射的光子被激光雷达接收到的立体角比率。第四项中, $\eta(\lambda)$ 为激光雷达系统的光学探测效率, $T^2(z)$ 为激光信号的双程大气透过率, $G(z)$ 为激光雷达观测的几何因子, 该项给出了光信号由于激光雷达系统本身和大气传输导致的衰减。 N_B 为激光雷达探测时的背景光子数。

式(1)中的双程大气透过率 $T^2(z) = e^{-2\tau(z)}$, 其中 $\tau(z)$ 为光学厚度, 有

$$\tau(z) = \int_0^z n(s) \cdot \sigma(s) ds. \quad (2)$$

其中, $n(s)$ 为大气密度, $\sigma(s)$ 为消光截面。在建模瑞利散射激光雷达方程时, 双程大气透过率的值接近于 1, 对模型结果影响不大。在实际的激光雷达观测时, 通常需要考虑两个方面的因素, 大气中臭氧层对光信号的吸收作用和发生于激光信号和大气分子的瑞利消光作用。

此外, 瑞利激光雷达探测到的回波光子数廓线的信噪比会对反演过程产生很大影响, 反演的精度主要是由回波光子数廓线的统计误差造成的。通常认为回波光子数在统计上服从泊松分布, 即

$$\sigma_N^2 = N, \sigma_N = \sqrt{N}. \quad (3)$$

因此, 回波光子数廓线的信噪比表达式为

$$S/N = 10 \lg \left(\frac{N_R(z) - N_B}{\sqrt{N_R(z)}} \right). \quad (4)$$

由式(4)可知, 增加回波光子数的数值, 例如提升激光光源的发射功率, 增大接收望远镜的有效接收面积, 延长积累回波信号的时间, 即可提升回波光子数廓线的信噪比, 进而提升反演的精度。

1.2 最优估计法理论

1.2.1 最优估计法基本原理

最优估计法是一种利用贝叶斯估计来约束的解空间的非线性回归形式, 通过建立正向模型并对成本函数进行最优化处理以得到反演结果^[3], 正向模型表征如下:

$$\mathbf{y} = \mathbf{F}(\mathbf{x}, \mathbf{b}) + \varepsilon. \quad (5)$$

式中, $\mathbf{x}$ 为状态向量, $\mathbf{y}$ 为测量向量; $\mathbf{F}$ 为正向模型, 其将状态空间的值映射到测量空间; $\mathbf{b}$ 为正向模型已知的系统参数; ε 为测量误差, 表示测量物理过程存在不确定性, 该不确定性由高斯分布下可能的值 ε 来表征。对于激光雷达而言, 测量向量描述的是激光雷达

通道探测到的每个高度的光子数,状态向量描述的是大气介质的某种特性,正向模型描述的是激光雷达发射光如何与大气介质相互作用以产生测量结果的物理过程。

实际应用中,即使式(5)中的正向模型是明确已知的,在存在测量噪声的情况下,状态空间的某点映射到测量空间后得到的是某一区域而非一个精确的点,反之亦然。此外,在进行测量之前,可能已经存在某些先验信息对状态向量进行了约束。因此,需要利用概率密度函数来量化这些存在不确定度的物理量,进而利用贝叶斯定理来表示先验信息与后验信息之间的关系,即

$$P(x|y) = \frac{P(y|x) \cdot P(x)}{P(y)}. \quad (6)$$

其中, $P(x|y)$ 是后验项,表示给定测量值 y 时的随机变量 x 的概率值,表征已有测量信息下的目标参数的状态, $P(x)$ 为测量未发生时随机变量 x 的概率值,称为先验概率, $P(y|x)$ 表示随机变量 x 为真时,得到测量值 y 的概率值, $P(y)$ 为得到测量值为 y 的概率值,为归一化因子,通常不予考虑。利用贝叶斯定理,可以根据探测系统的测量值及其不确定度以及先验信息找到最可能的状态信息。

贝叶斯定理的应用需要假设测量过程中的物理量服从高斯分布,以此得到成本函数。向量形式下的高斯概率密度分布如下:

$$P(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^k |\mathbf{S}_x|}} \exp\left(-\frac{1}{2}[\mathbf{x} - \bar{\mathbf{x}}]^T \mathbf{S}_x^{-1} [\mathbf{x} - \bar{\mathbf{x}}]\right). \quad (7)$$

式中, $\mathbf{S}_x$ 为随机向量 $\mathbf{x}$ 的协方差矩阵,且其必须为可逆矩阵。由式(7)可以得到

$$\begin{cases} -2 \ln P(\mathbf{y}|\mathbf{x}) = [\mathbf{y} - \mathbf{F}(\mathbf{x}, \mathbf{b})]^T \mathbf{S}_\epsilon^{-1} [\mathbf{y} - \mathbf{F}(\mathbf{x}, \mathbf{b})] + c_1, \\ -2 \ln P(\mathbf{x}) = [\mathbf{x} - \mathbf{x}_a]^T \mathbf{S}_a^{-1} [\mathbf{x} - \mathbf{x}_a] + c_2. \end{cases} \quad (8)$$

其中, $\mathbf{S}_\epsilon$ 为测量误差协方差矩阵, $\mathbf{x}_a$ 为先验状态向量, $\mathbf{S}_a$ 为先验协方差矩阵, c_1 和 c_2 为常数。将式(8)中的两项代入式(6)中,且将归一化因子 $P(\mathbf{y})$ 和常数省去,可以得到成本函数,即

$$\chi^2 = [\mathbf{y} - \mathbf{F}(\mathbf{x}, \mathbf{b})]^T \mathbf{S}_\epsilon^{-1} [\mathbf{y} - \mathbf{F}(\mathbf{x}, \mathbf{b})] + [\mathbf{x} - \mathbf{x}_a]^T \mathbf{S}_a^{-1} [\mathbf{x} - \mathbf{x}_a]. \quad (9)$$

实际上,在式(9)中, $\chi^2 = -2 \ln P(\mathbf{x}|\mathbf{y})$ 。为了使

得 $P(\mathbf{x}|\mathbf{y})$ 后验概率最大,要对式(9)进行最优化处理,得到最优解 $\hat{\mathbf{x}}$ (即反演结果)。在本文中,由于正向模型属于非线性模型,所以使用 Levenberg-Marquardt 算法(后称 LM 算法)进行对上式进行最优化处理^[9,10]。由于式(9)中两项均是高斯分布情形下的,得到关于最优解 $\hat{\mathbf{x}}$ 的关系式为

$$-2 \ln P(\mathbf{x}|\mathbf{y}) = [\mathbf{x} - \hat{\mathbf{x}}]^T \hat{\mathbf{S}}^{-1} [\mathbf{x} - \hat{\mathbf{x}}]. \quad (10)$$

结合式(9)和式(10)可以得到,最优解 $\hat{\mathbf{x}}$ 应满足

$$\hat{\mathbf{x}} = \mathbf{x}_a + [\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} + \mathbf{S}_a^{-1}]^{-1} \mathbf{K}^T \mathbf{S}_\epsilon^{-1} [\mathbf{y} - \mathbf{F}(\mathbf{x}_a)]. \quad (11)$$

其中, $\mathbf{K}$ 是正向模型测量向量对状态向量的导数矩阵^[11],也称为 Jacobian 矩阵。为了在状态空间中寻找 $\hat{\mathbf{x}}$ 的相对结果,由式(11)可以得到增益矩阵为

$$\mathbf{G} = \frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{y}} = [\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} + \mathbf{S}_a^{-1}]^{-1} \mathbf{K}^T \mathbf{S}_\epsilon^{-1}. \quad (12)$$

结合式(11)可以得到

$$\hat{\mathbf{x}} = \mathbf{x}_a + \mathbf{G} [\mathbf{y} - \mathbf{F}(\mathbf{x}_a)] = [\mathbf{I} - \mathbf{A}] \mathbf{x}_a + \mathbf{A} \mathbf{x} + \mathbf{G} \epsilon. \quad (13)$$

其中, $\mathbf{A} = \mathbf{G} \mathbf{K}$ 称为平均核矩阵^[3], $\mathbf{I}$ 为单位矩阵。平均核矩阵描述了真实状态向量和先验状态向量对于反演解的贡献占比。一个理想的反演过程,其对应的平均核矩阵应为一个单位矩阵,表明反演状态全部来自于真实状态而非先验状态,否则,先验状态对反演结果有贡献。在实际应用中,平均核矩阵的每一行均为峰值曲线,表征反演解的某个元素如何受到其他各元素的影响,其半高全宽称为反演的垂直分辨率。此外,平均核矩阵的迹称为反演的自由度理想情形下,自由度应该等于状态向量的元素个数。实际应用中,自由度通常小于状态向量元素的个数。

1.2.2 最优估计法的不确定度

最优估计法可以基于每条廓线来全面地进行不确定度的评估^[3],其中包括由于系统参数引起的系统不确定度。测量噪声引起的反演不确定度为

$$\mathbf{S}_m = \mathbf{G} \mathbf{S}_\epsilon \mathbf{G}^T. \quad (14)$$

由于探测系统对状态向量的各元素缺乏敏感性会导致平滑误差,由此引起的平滑不确定度为

$$\mathbf{S}_s = [\mathbf{A} - \mathbf{I}] \mathbf{S}_\epsilon [\mathbf{A} - \mathbf{I}]^T. \quad (15)$$

其中, $\mathbf{S}_\epsilon$ 为真实状态向量的协方差矩阵。要对平滑不确定度进行分析,需要明确知道状态向量的实际统计结果,而这通常难以实现。因此,利用合理的先验协

方差矩阵作为替代,以对平滑不确定度尝试分析。此外,正向模型中相关参数本身存在误差,所以由此产生的不确定度称为模型参数不确定度,有

$$\mathbf{S}_f = \mathbf{G}\mathbf{K}_b\mathbf{S}_b\mathbf{K}_b^T\mathbf{G}^T. \quad (16)$$

其中, $\mathbf{K}_b$ 为正向模型参数的 Jacobian 矩阵, $\mathbf{S}_b$ 为模型参数 $\mathbf{b}$ 的协方差矩阵。实际上,由于正向模型只是为了解决问题进行的数值近似,因此建模产生正向模型的本身与真实物理过程可能也存在一定的误差。但是,正向模型所表征的物理过程往往是极其复杂的,所以正向模型误差也难以估计。

1.3 最优估计法反演大气温度

将最优估计法应用于瑞利散射激光雷达,首先需要明确瑞利散射激光雷达探测大气参数的真实物理过程,以构建正向模型,即真实物理过程的数值近似。然后,利用激光雷达的测量结果作为测量向量,选定合适的测量协方差矩阵,选择合适先验温度状态廓线和先验协方差矩阵,最后对式(9)中的成本函数进行最优化。

在本文中,利用瑞利散射激光雷达对大气温度廓线进行探测,因此大气温度廓线为状态向量,激光雷达回波光子数廓线为测量向量,瑞利散射激光雷达方程为正向模型。正向模型中认为目标大气服从理想气体定律 $PV = n_{\text{mol}}RT$, 再结合流体静力平衡 $dP = -\rho g dz$, 以及瑞利散射激光雷达方程(1),可以得到如下正向模型:

$$\begin{aligned} N_R(z) &= \frac{C}{(z - z_0)^2} \cdot \frac{P(z)}{kT(z)} + N_B, \\ C &= \left[\frac{P(\lambda) \Delta t}{hc/\lambda} \right] [\sigma_R(\lambda) \Delta z A] [\eta(\lambda) T^2(z) G(z)], \\ P(z) &= P_{\text{top}} \exp \left(\frac{1}{R} \int_z^{z_{\text{top}}} \frac{M(z') g(z')}{T(z')} dz' \right). \end{aligned} \quad (17)$$

式中, C 为激光雷达的系统参数, P 为 z 高度处的压强, P_{top} 为探测高度范围顶部的参考压强值, k 为玻尔兹曼常数, T 为 z 高度处的大气温度, R 为理想气体常数, M 为大气分子的摩尔质量, g 为重力加速度。式(17)中自向下积分得到探测高度范围内的大气压强值。

由于海拔高度大于 80 km 时,瑞利散射截面不是定值,因此通过 NRLMSISE-00 大气模型输出的主要大气成分占比,结合各成分的瑞利散射截面来计算总体瑞利散射截面^[12]。本文中 P_{top} 的确定通过 NRLM-

SISE-00 大气模型输出的温度和密度利用理想气体定律来计算。不同高度的大气分子的平均摩尔质量 $M(z)$ 通过 NRLMSISE-00 大气模型输出的各主要成分的权重来计算^[13]。不同高度下的重力加速度参考 1976 年美国标准大气^[14] 来计算。

式(9)为进行最优化处理的成本函数,需要确定其中的相关参数。由于激光雷达探测系统的回波光子计数服从泊松分布,因此测量向量的方差即为相应高度处的回波光子数。不同高度的回波光子数之间没有相关性,因此测量误差协方差矩阵的非对角线元素均为零。利用 CIRA-86 大气模型中纬度地区(或 1976 年美国标准大气模型)的温度廓线作为先验状态廓线^[15],且选择其温度不确定度为 35 K,符合美国标准大气中所描述的其他探测手段相对于 1976 年美国标准大气温度的系统变化范围。在进行模拟反演时,将先验温度不确定度设定为 15 K。此外利用三角形相关函数来计算其不同高度之间的相关性^[16]。

本文中瑞利散射激光雷达系统的正向模型是非线性模型的,利用 LM 算法对成本函数进行最优化处理。采用查尔姆斯理工大学的 Eriksson 等开发的开源程序 ARTS(The Atmospheric Radiative Transfer Simulator)^[17] 下的 Matlab 工具包 Atmlab^[18] 来进行迭代处理。Atmlab 本身用于与 ARTS 进行交互,包括米散射与吸收、大气反演、生成大气场景等。LM 算法是解决非线性最小二乘问题的最优化算法,改进于 Gauss-Newton 算法。利用 LM 算法对成本函数最优化处理时,迭代公式如下:

$$\begin{aligned} \mathbf{x}_{i+1} &= \mathbf{x}_i + [(1 + \gamma_i) \mathbf{S}_a^{-1} + \mathbf{K}_i^T \mathbf{S}_\epsilon^{-1} \mathbf{K}_i]^{-1} \cdot \\ &\quad \{ \mathbf{K}_i^T \mathbf{S}_\epsilon^{-1} [\mathbf{y} - \mathbf{F}(\mathbf{x}_i, \mathbf{b})] - \mathbf{S}_a^{-1} [\mathbf{x}_i - \mathbf{x}_a] \}. \end{aligned} \quad (18)$$

其中, γ_i 为迭代因子,随着迭代结果的增大和减小有着不同的改变。首次迭代时,状态向量的迭代初始值选定为先验状态向量,但这并不是初始值的唯一选择。在本文中迭代 10 次以下即可反演得到状态向量的最优估计解。

2 反演结果

以瑞利散射激光雷达方程作为正向模型,对一定海拔高度范围的大气温度进行最优估计反演,得到观测范围内的大气温度廓线,并对反演结果进行不确定

度分析,验证最优估计法对于瑞利散射激光雷达反演大气温度的可行性和有效性。利用平均核矩阵对反演过程中测量信息对于反演结果的贡献进行评估。

2.1 基于仿真数据的最优估计法验证

在仿真模拟阶段,测量数据是由瑞利散射激光雷达方程本身生成的。本文选择 2018 年 9 月 3 日的 NRLMSISE-00 大气模型的输出数据来生成正向模型的测量结果,经纬度选择中国科学院国家空间科学中心北京站点(116.68°E, 40.33°N)。所采用瑞利散射激光雷达系统详细参数见表 1。

设定激光雷达为地基雷达,由于低层大气中气溶胶含量较高,此时激光信号和大气发生的主要是米散射,所以瑞利散射信号的有效探测高度范围设定为海拔 30~120 km 的大气空间,积分时间为 1 h。模拟得到的回波光子数廓线,加入泊松统计不确定度后的结果如图 1 所示,其空间分辨率为 100 m。模拟回波光子数廓线信噪比结果如图 2 所示。

由图 1 可以看出,在观测范围顶部区域,由于海

表 1 模拟激光雷达系统参数
Table 1 Simulation lidar system parameters

系统参数	参数选值
脉冲能量/ mJ	40
重复频率/ Hz	50
激光波长/ nm	532
望远镜口径/ mm	350
系统效率(unitless)	0.191

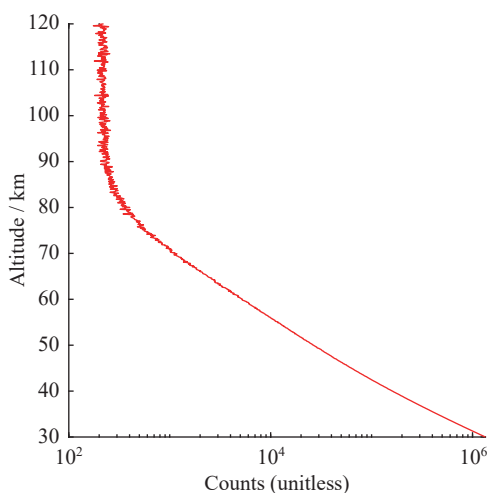


图 1 模拟回波光子数廓线

Fig. 1 Simulated echo photon count profile

拔较高,有效的大气瑞利散射信号很弱,激光雷达回波信号趋于恒定值,且信噪比很低。由于测量向量产生于 NRLMSISE-00 大气模型的大气温度数据,将当天的 NRLMSISE-00 大气模型大气温度廓线视为真实大气温度廓线。以式(17)的瑞利散射激光雷达方程作为正向模型,将图 1 中的模拟回波光子数廓线作为正向模型的测量向量 y ,各高度的回波光子数作为成本函数中测量协方差矩阵 S_e 的对角线元素。为了区分 NRLMSISE-00 大气模型的真实大气温度廓线,选择 CIRA-86 大气模型的 9 月份的大气温度廓线作为先验温度信息,各个高度的温度不确定度设定为 15 K。利用相关长度为 5 km 的三角形相关函数来建立先验状态的协方差矩阵 S_a 。采用 LM 算法进行最优化迭代, γ 因子的初始值设定为 100,迭代成功则 γ 缩小 2 倍,否则 γ 扩大 5 倍。在迭代次数达到 4 次后,成功达到收敛标准,反演得到大气温度廓线如图 3 所示。

图 3 中,红色曲线为反演得到的大气温度廓线,蓝色曲线为 CIRA-86 模型先验温度廓线,绿色曲线为 NRLMSISE-00 模型真实温度廓线,红色半透明区域是反演温度廓线的不确定度,包含测量引起的不确定度式(14)和平滑不确定度式(15)。可以看出,在观测范围的底部区域,反演温度廓线和真实温度廓线吻合度较高;随着海拔的升高,反演温度廓线与真实温度廓线的偏差逐渐增大,而且不确定度也随着海拔的升高而逐渐增大。模拟数据反演温度廓线的误差与

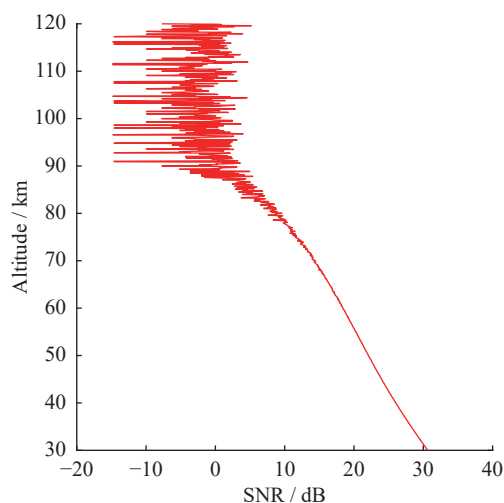


图 2 模拟回波信噪比廓线

Fig. 2 Signal-to-noise ratio of simulated echo photon count profile

不确定度结果如图 4 所示。

图 2 的信噪比廓线在 80 km 处约为 10 dB, 随着海拔升高, 有效瑞利散射信号变弱, 信噪比降低, 反演温度廓线的误差和不确定度均逐渐增大。在图 4 中, 反演结果和真实大气温度在 30~80 km 的范围内的反演误差在 5 K 以内, 在 80~90 km 的范围反演误差在 10 K 以内, 90 km 以上范围的反演误差大于 10 K。在 30~80 km 的海拔范围内, 反演温度的不确定度在 10 K 以内; 在海拔 80 km 以上, 不确定度逐渐

趋近于先验温度不确定度 15 K。由图 4 还可以看出, 反演过程的误差和不确定度并非同一数值。在实际应用中, 有时难以获得待测量的真实值, 所以测量结果和真实值的误差也不得而知, 更多地使用测量值的不确定度来衡量结果的置信程度, 不确定度小则说明测量结果的置信程度高。

式(13)中, 平均核矩阵表征测量信息对于反演结果的贡献。将图 3 中反演温度廓线的平均核矩阵 $\mathbf{A}$ 和垂直分辨率绘制如图 5 所示。在图 5(a) 中, 不同

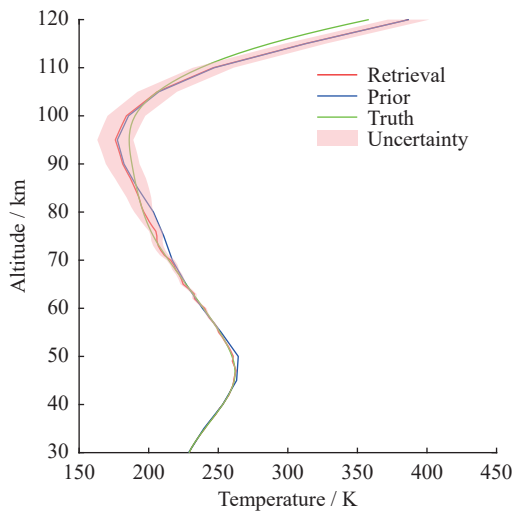


图 3 模拟数据的反演温度廓线

Fig. 3 Retrieved temperature profiles from simulated data

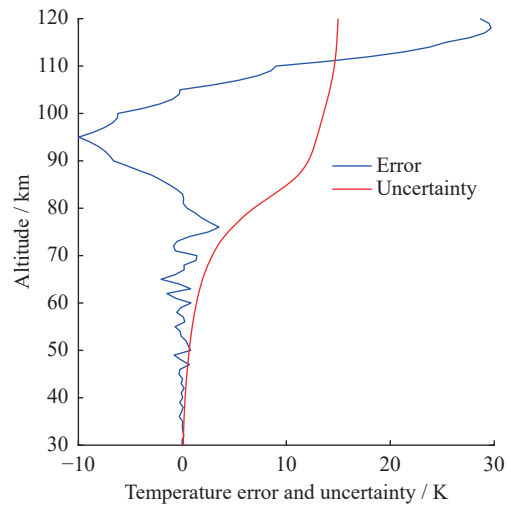


图 4 模拟数据反演温度廓线的误差和不确定度

Fig. 4 Errors and uncertainties of inversion of temperature profiles from simulated data

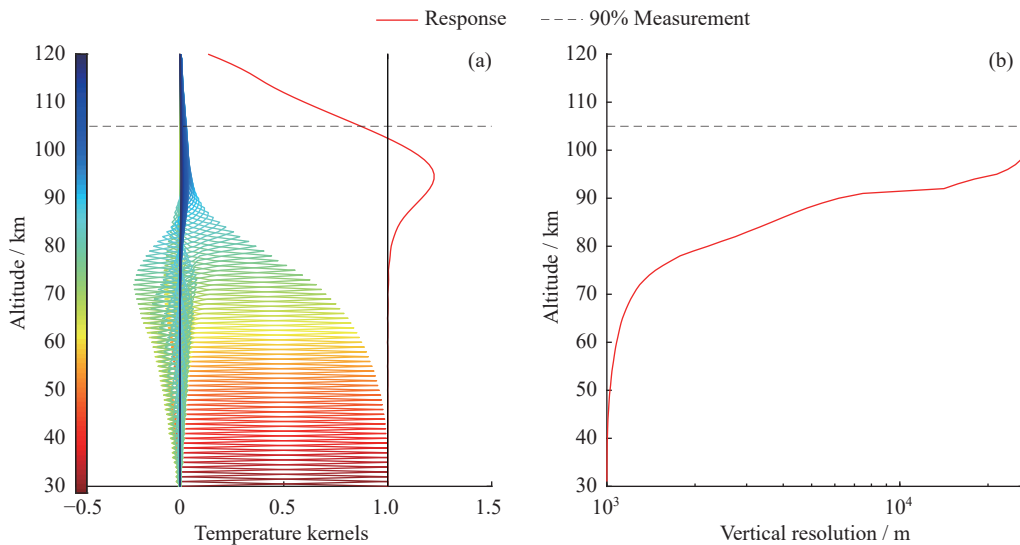


图 5 模拟反演温度廓线的平均核矩阵 (a) 和垂直分辨率 (b)

Fig. 5 Average kernel matrix (a) and vertical resolution (b) of simulated inversion temperature profile

高度的平均核绘制在相应的高度,并以特定的颜色标注。红色曲线称为响应,是平均核矩阵每行的和,即 $\mathbf{A}\mathbf{u}$, 其中 $\mathbf{u}$ 为单位列向量。黑色水平虚线高度处,响应值下降到 0.9, 这表明在该高度处测量信息对于反演结果的贡献达 90%, 剩余 10% 来自先验信息。图 5(b) 中反演温度廓线的垂直分辨率, 即平均核矩阵 $\mathbf{A}$ 的各行半高全宽, 是观测系统空间分辨率的量度。

根据图 5(a) 可以看出, 在大于 100 km 的空间区域, 平均核矩阵的响应值逐渐低于 0.9, 这时认为测量信息不足以对反演结果产生足够的贡献。在 90 km 以上的空间区域, 各高度的平均核峰值接近于 0, 这可以解释图 3 中反演廓线在探测范围顶部逐渐远离真实廓线, 即将趋近于先验廓线的现象, 同时反演温度的不确定度也逐步趋近于先验温度的不确定度设定值 15 K。图 5(b) 为反演温度廓线的垂直分辨率, 可以看出, 30~80 km 的垂直分辨率约为 1~2 km, 这和温度状态向量的空间分辨率相当, 随着海拔高度继续升高, 垂直分辨率迅速降低。

综上所述, 利用 NRLMSISE-00 大气模型作为真实状态信息, 通过瑞利散射激光雷达方程产生回波光光子数廓线作为测量向量, 以 CIRA-86 大气模型的温度廓线作为先验温度廓线, 对成本函数进行最优化处理, 得到了大气温度的反演结果。这在数据仿真上验证了最优估计法应用于瑞利散射激光雷达探测大气温度的可行性。

2.2 CH 方法与 OEM 方法的反演对比

CH 方法中, 在进行大气参数反演时, 需要选择某一个高度 z_r 的大气数密度作为反演过程的参考值, 得

到相对大气密度廓线, 即

$$n(z) = \frac{[N_R(z) - N_B](z - z_0)^2}{[N_R(z_r) - N_B](z_r - z_0)^2} \cdot n(z_r). \quad (19)$$

结合理想气体方程和流体静力平衡, 有

$$\begin{cases} PV = n_{\text{mol}}RT, \\ dP = -\rho g dz. \end{cases} \quad (20)$$

可以得到如下大气温度廓线的反演公式:

$$T(z) = \frac{T(z_r)\rho(z_r)}{\rho(z)} + \frac{M}{R} \int_z^{z_r} \frac{g(z')\rho(z')}{\rho(z)} dz'. \quad (21)$$

式中大气温度的参考点为探测范围的顶部, 可以选择探测范围顶部大气模型的值作为参考温度。然而, 顶部区域的参考值往往具有较大的不确定度, 会对反演温度廓线产生影响。向下积分时, 式(21)的第二项会逐渐增大占主导位置, 参考温度不确定度的影响会逐渐变小。因此, 实际反演得到的大气温度廓线通常应该舍去顶部 10~15 km 的反演结果。

按照表 1 的激光雷达系统参数, 产生模拟回波信号, 进行 CH 反演, 得到如图 6 所示的回波光光子数廓线与信噪比结果。

根据图 6 可以看出, 在 80 km 处, 信噪比约为 10 dB, 在 80 km 以上, 信噪比逐渐降低, 因此选择 80 km 作为大气温度的参考高度, 利用 NRLMSISE-00 模型的温度数据作为参考值, 参考温度中加入 $\pm 20\%$, $\pm 10\%$, $\pm 5\%$, $\pm 2\%$ 的不确定度, 依据式(21)进行多次 CH 反演, 反演结果如图 7 所示。

根据图 7 可以看出, 由于在 80 km 处为参考温度引入了不确定度, 即使此处回波光光子数廓线的信噪比

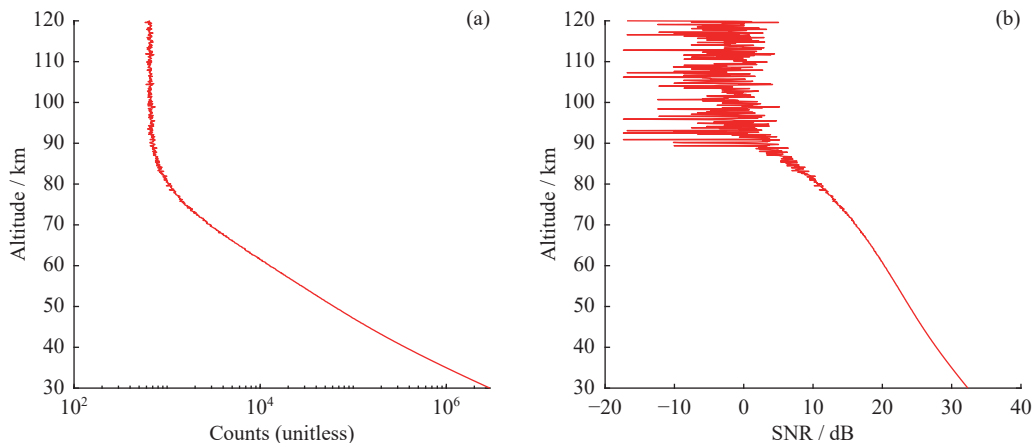


图 6 CH 反演模拟回波光光子数廓线 (a) 与信噪比 (b)

Fig. 6 CH inversion simulated echo photon number profile (a) and signal-to-noise ratio (b)

较高,在该高度以下 10 km 左右,反演温度廓线和真实温度廓线间也存在较大的误差。在实际反演过程中,观测范围顶部区域的真实参考温度往往难以得知,因此,CH 方法不得不舍弃观测范围顶部 10~15 km 顶部的数据。同样在正向模型式(17)中的参考压强引入不确定度,进行多次最优估计法反演,得到的结果如图 8 所示。

从图 8(a)中可以看出,引入参考不确定度后,多次反演得到的大气温度廓线基本吻合。从图 8(b)中可以看出,80 km 以下各条反演温度廓线与真实模型大气温度廓线的误差均在 5 K 以内。因此,相较于

CH 反演,OEM 反演克服了前者在 80 km 以下 10~15 km 范围内的反演数据不得不舍弃的弊端。进行反演时,OEM 进行大约 4~5 次迭代过程即可得到收敛。在运算速度上,两种方法计算用时相近。

2.3 最优估计法对实测数据的处理

利用最优估计法对瑞利散射激光雷达的实测数据进行反演与分析。回波光光子廓线数据来源于中国科学院国家空间科学中心,该激光雷达的所在经纬度为 116.68°E, 40.33°N,激光雷达系统参数列于表 2。

2018 年 9 月 3 日 17:30 UT 回波光光子数如图 9 所示。观测回波光光子数廓线的有效高度范围在

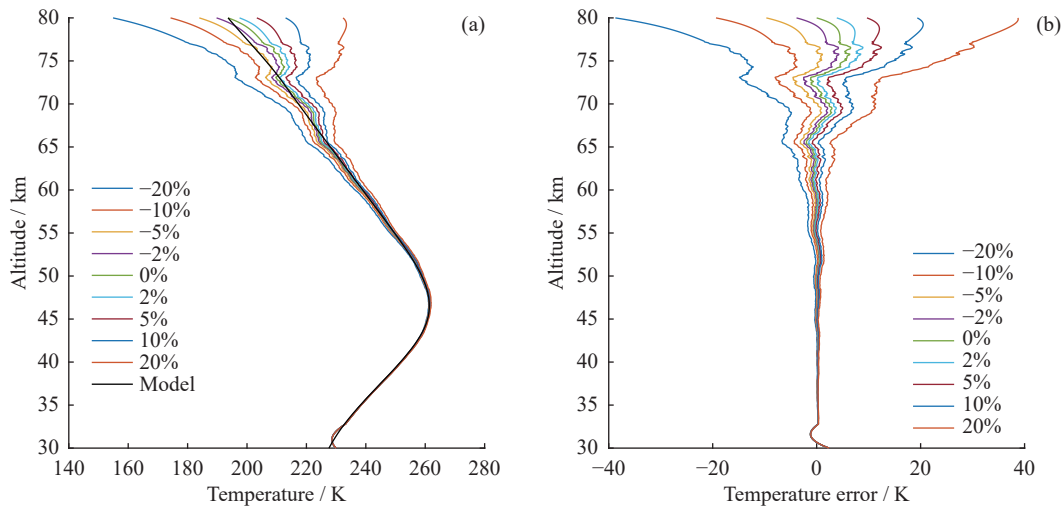


图 7 参考温度引入不确定度的 CH 反演结果

Fig. 7 CH inversion results with reference temperature introducing uncertainty

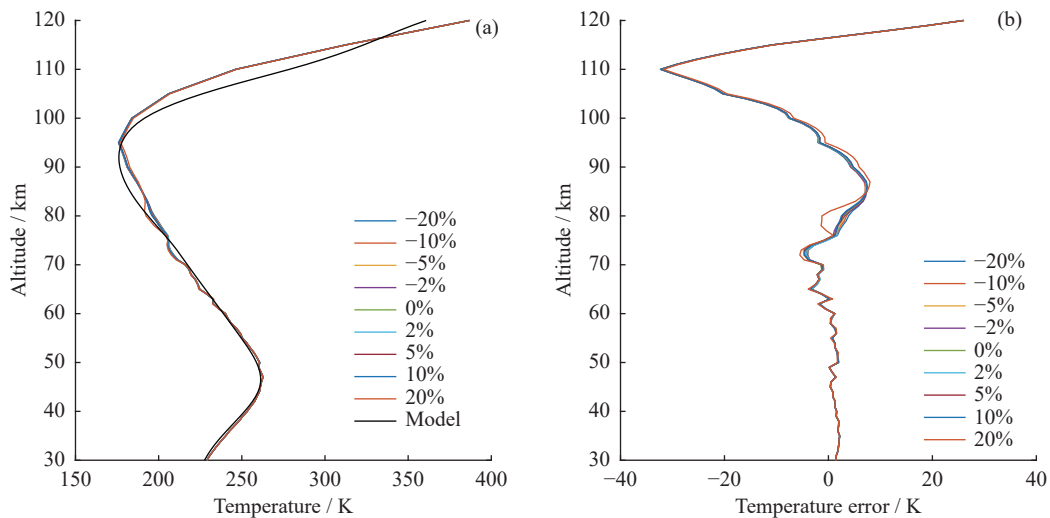


图 8 参考压强引入不确定度的 OEM 反演结果

Fig. 8 OEM inversion results after introducing uncertainty into reference pressure

40 km 以上, 光子数廓线的空间分辨率约为 123 m。可以看出, 90 km 以上的回波信号的信噪比较低, 此时回波信号中瑞利散射信号很小, 通常认为观测范围顶部附近的回波信号均为背景信号。该回波信号的信噪比廓线如图 10 所示。

最优估计法属于第一原则反演, 因此无需对图 9 的实测数据进行去噪等操作, 避免人为引入难以估计的不确定度。图 9 中的光子数廓线是正向模型的测量向量 y , 由于光子计数统计服从泊松分布, 所以测量向量的方差(S_y 的对角线元素)为回波光子计数, 忽略不同高度回波光子数之间的相关性。同样地, 选择 CIRA-86 大气模型的温度数据作为先验温度廓线, 先验温度的不确定度设定为 35 K。利用 LM 算法对成本函数最优化处理, 迭代达 8 次后成功收敛, 得到反演大气温度廓线(见图 11)。

图 12 给出了最优估计反演结果的不确定度。图 10 的信噪比廓线中, 80 km 处的信噪比约为

表 2 实测数据的瑞利散射激光雷达系统参数
Table 2 National space center rayleigh lidar system parameters

系统参数	参数选值
脉冲能量/mJ	500
脉冲宽度/ns	8
重复频率/Hz	30
激光波长/nm	532
望远镜口径/mm	1000
系统效率(unitless)	0.02

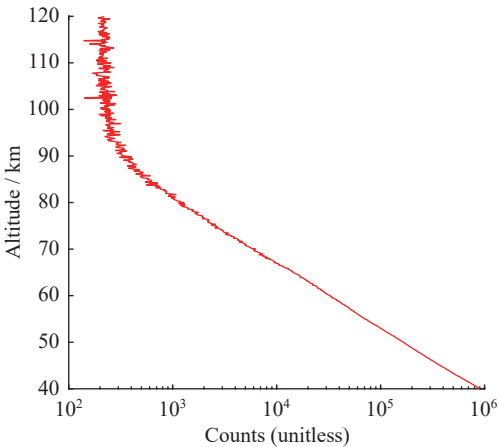


图 9 瑞利散射激光雷达实测回波光子数廓线
Fig. 9 Rayleigh lidar measured echo photon profile

15 dB, 90 km 信噪比约为 7 dB。在 90 km 以下, 反演温度的不确定度在 10 K 以内, 在 80 km 以下的温度不确定度在 5 K 以内, 在 90 km 以上, 信噪比低于 7 dB, 反演温度的不确定度逐步达到 30 K 左右。

对于合理的数据反演, 正向模型和测量值之间的差异(即所谓的残差), 应该在探测系统测量值的不确定度之内。将反演结果作为状态向量, 通过正向模型得到测量向量, 并将其与实际测量值进行比较, 结果如图 13 所示。

图 13 表明, 本文所使用的正向模型正确地描述了瑞利散射激光雷达对于大气物理特性的探测过程。利用平均核矩阵 A 评估反演过程中测量信息对

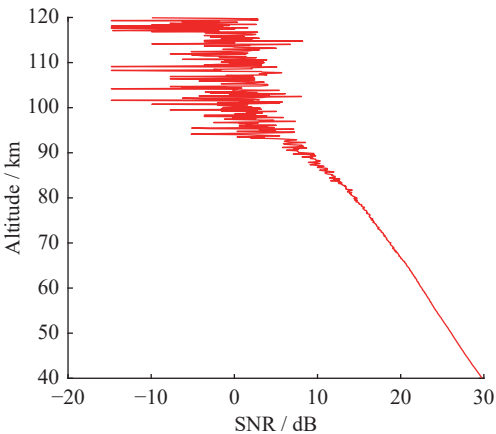


图 10 实测回波信号的信噪比廓线
Fig. 10 Signal-to-noise ratio profile of the measured echo photon profile

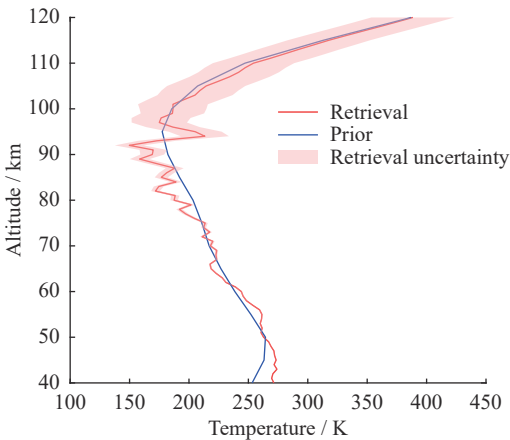


图 11 瑞利散射激光雷达实测数据的反演温度廓线
Fig. 11 Retrieved temperature profiles from rayleigh lidar measured data

于反演得到状态向量的贡献占比,将平均核矩阵和垂直分辨率绘制如图 14 所示。

图 14(a)给出了实测数据反演结果的平均核矩阵,其中红色曲线是响应曲线,黑色水平曲线是测量信息对反演结果贡献占比为 90% 的高度。由图 14(a)可见,在 80 km 以下,响应曲线的值基本为 1,意味着相应高度处的测量信息对反演结果的贡献占比接近 100%,这是由于 80 km 下的瑞利回波信号的信噪比较高,达到 15 dB 以上。图 14(b)中,反演温度廓线

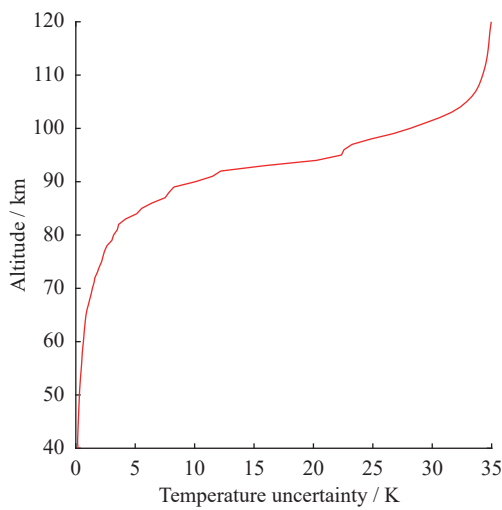


图 12 最优估计反演结果的不确定度

Fig. 12 Uncertainty of optimal estimation inversion result

的垂直分辨率在 80 km 以下为 1 km,这与起初设定的大气温度状态向量的空间分辨率相同,超过 80 km,垂直分辨率随着海拔高度提升而逐步降低。

为了验证不同先验温度廓线对于反演温度廓线

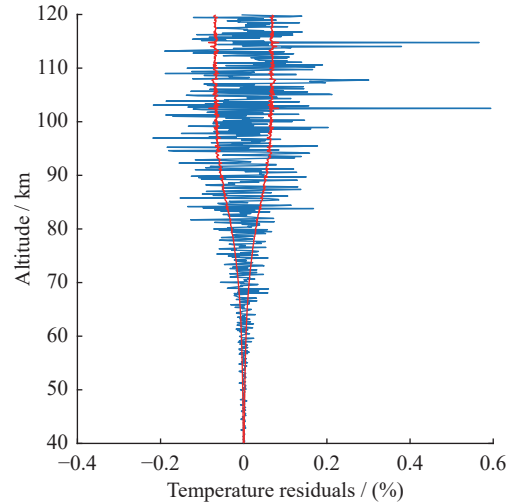


图 13 正向模型与测量值的残差 (蓝色曲线是反演状态的正向模型测量值与实际测量值的百分比偏差,红色曲线是测量不确定度占测量值得百分比权重)

Fig. 13 Residual error between the forward model and the measured value (The blue curve is the percentage deviation between the forward model measured value of the inversion state and the actual measured value, and the red curve is the percentage weight of the measurement uncertainty in the measured value)

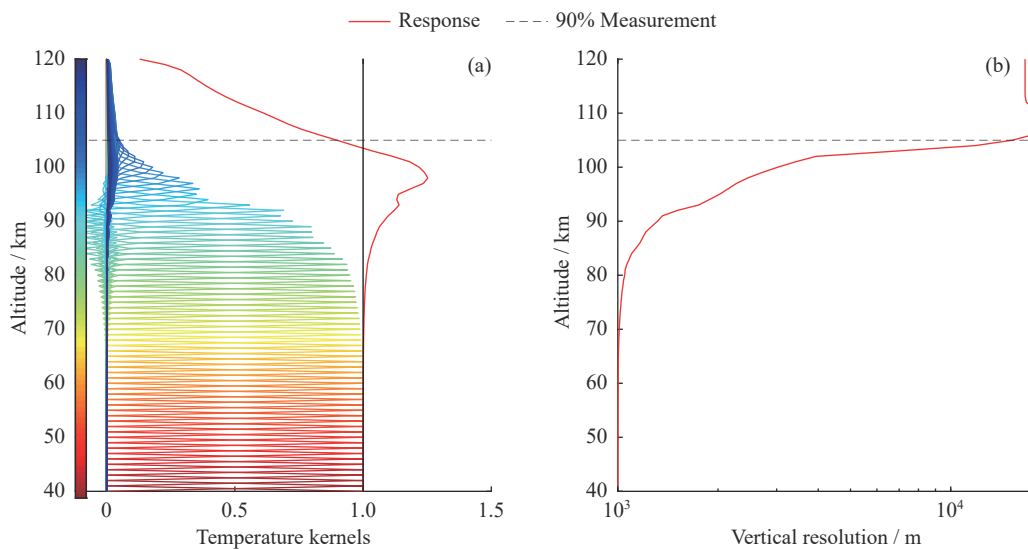


图 14 实测数据反演温度廓线的平均核矩阵 (a) 和垂直分辨率 (b)

Fig. 14 Average kernel matrix (a) and vertical resolution (b) of the measured data inversion temperature profile

是否存在影响,针对同一激光雷达测量到的回波光子信号,分别选择 CIRA-86 大气模型和 NRLMSISE-00 大气模型在相同条件下的大气温度廓线作为先验状态信息,对成本函数执行最优估计反演,绘制反演结果如图 15 所示。

图 15 给出了两种大气模型作为先验温度的反演温度廓线及其差异值。在 90 km 以下,图 15(a)中两种大气模型反演得到的反演温度廓线基本吻合,而图 15(b)中二者反演结果的温度差异值小于 5 K,这证明了此时先验信息的选择并不会影响反演结果,图 14(a)中的平均核矩阵也说明了这一点,中低海拔区域测量信息对反演结果占主要贡献。

利用中国科学院国家空间科学中心的瑞利散射激光雷达实测数据,通过构建瑞利散射激光雷达正向模型,以大气模型的温度廓线作为先验信息,对大气温度廓线进行了反演。在回波光子信号足够强的高度范围内,测量信息对反演结果占主要贡献,先验温度的选取对反演结果的影响很小。在 90 km 以下,反演温度廓线的不确定度在 10 K 以内;在 80 km 以下,不确定度在 5 K 以内,且测量信息对反演结果的贡献占比接近于 1。

3 结论

基于瑞利散射激光雷达正向模型,应用最优估计法来反演中层大气的温度廓线。从模拟仿真和实测

数据反演两个方面,验证了最优估计法应用于瑞利散射激光雷达反演大气温度的可行性。

最优估计法避免了传统 CH 方法中观测范围顶部的大气温度或压强自身的不确定度导致顶部 10~15 km 的数据不得不舍去的弊端,一定程度上对激光雷达的观测数据实现了更为充分的利用。此前,研究人员致力于在计算温度之前对测量信号进行去噪、修正等处理,而在最优估计法中,第一原则反演使得无需对回波光子数据进行去噪等处理,在反演计算中使用 0 级测量数据,避免了某些难以评估的不确定度的引入。在 0~90 km 的高度范围内反演温度的不确定度在 10 K 以内。相对于 CH 方法对不确定度的考察方式,最优估计法能够利用自带的评估手段对不确定度进行评估。

此外,本文利用平均核矩阵来评估测量信息和先验信息对反演温度廓线的贡献占比,同时也能确定反演垂直分辨率随着高度的变化廓线,在大气回波信号信噪比较高的区域,测量信息对于反演结果通常占主要贡献,而且相应的反演垂直分辨率也高。最优估计法对反演不确定度的考量分为四种,分别是反演误差、平滑误差、参数误差和模型误差。本文中的提到的不确定度,包括了反演误差和平滑误差两项之和。对于激光雷达系统的参数误差和模型误差,本文中尚未考量,有待后续研究中进一步分析。

在进行最优估计反演时,选定参数的不确定度并不是不可更改的,目前只是基于已有信息的合理选

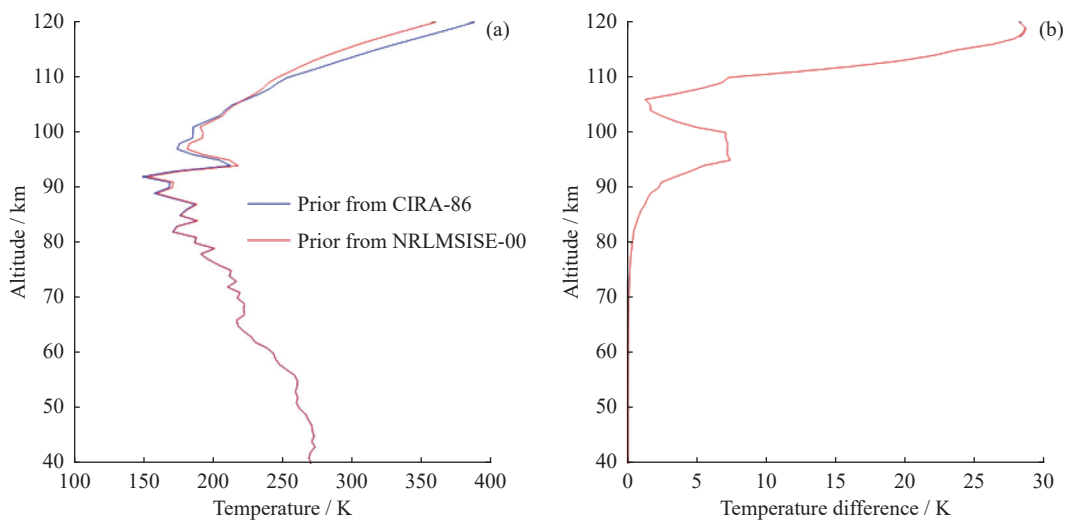


图 15 不同先验温度廓线对反演结果的影响

Fig. 15 Influence of different prior temperature profiles on the inversion results

择,在后续的研究中可以开发更为有益的选择手段。在更为长期的瑞利散射激光雷达观测活动中,可以利用最优估计法来处理更多的激光雷达测量数据并进行不确定度评估,用以生成更加完善的大气温度数据库,为中层大气的热力学结构提供一定的数据基础。此外,凭借最优估计法的优良特性,未来可以将其应用于其他主动探测仪器的观测中,例如对激光雷达测量大气气溶胶、水蒸气的数据进行最优估计反演。

致谢 本文使用的瑞利散射激光雷达实测回波光光子数廓线数据由中国科学院国家空间科学中心提供。

参考文献

- [1] HAUCHECORNE A, CHANIN M L. Density and temperature profiles obtained by lidar between 35 and 70 km[J]. *Geophysical Research Letters*, 1980, **7**(8): 565-568
- [2] LIU Xing, YANG Guotao, WANG Jihong, *et al.* Retrieval algorithm of middle atmospheric temperature using rayleigh lidar[J]. *Chinese Journal of Space Science*, 2019, **39**(2): 186-190 (刘星, 杨国韬, 王继红, 等. 瑞利激光雷达反演中层大气温度算法[J]. *空间科学学报*, 2019, **39**(2): 186-190)
- [3] RODGERS C D. Inverse Methods for Atmospheric Sounding: Theory and Practice[M]. Singapore: World Scientific, 2000
- [4] MAAHN M, TURNER D D, LÖHNERT U, *et al.* Optimal estimation retrievals and their uncertainties: what every atmospheric scientist should know[J]. *Bulletin of the American Meteorological Society*, 2020, **101**(9): E1512-E1523
- [5] POVEY A C, GRAINGER R G, PETERS D M, *et al.* Retrieval of aerosol backscatter, extinction, and lidar ratio from Raman lidar with optimal estimation[J]. *Atmospheric Measurement Techniques*, 2014, **7**(3): 757-776
- [6] SICA R J, HAEFELE A. Retrieval of temperature from a multiple-channel Rayleigh-scatter lidar using an optimal estimation method[J]. *Applied Optics*, 2015, **54**(8): 1872-1889
- [7] SICA R J, HAEFELE A. Retrieval of water vapor mixing ratio from a multiple channel Raman-scatter lidar using an optimal estimation method[J]. *Applied Optics*, 2016, **55**(4): 763-777
- [8] WEITKAMP C. Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere[M]. New York: Springer, 2005
- [9] LEVENBERG K. A method for the solution of certain non-linear problems in least squares[J]. *Quarterly of Applied Mathematics*, 1944, **2**(2): 164-168
- [10] MARQUARDT D W. An algorithm for least-squares estimation of nonlinear parameters[J]. *Journal of the Society for Industrial and Applied Mathematics*, 1963, **11**(2): 431-441
- [11] SICA R J, HAEFELE A, JALALI A, *et al.* How to apply the optimal estimation method to your lidar measurements for improved retrievals of temperature and composition[J]. *EPJ Web of Conferences*, 2018, **176**: 01025
- [12] ARGALL P S. Upper altitude limit for Rayleigh lidar[J]. *Annales Geophysicae*, 2007, **25**(1): 19-25
- [13] PICONE J M, HEDIN A E, DROB D P, *et al.* NRLMSISE-00 empirical model of the atmosphere: statistical comparisons and scientific issues[J]. *Journal of Geophysical Research: Space Physics*, 2002, **107**(A12): 1468
- [14] National Geophysical Data Center. U. S. standard atmosphere (1976)[J]. *Planetary and Space Science*, 1992, **40**(4): 553-554
- [15] CHANDRA S, FLEMING E L, SCHOEBERL M R, *et al.* Monthly mean global climatology of temperature, wind, geopotential height and pressure for 0-120 km[J]. *Advances in Space Research*, 1990, **10**(6): 3-12
- [16] ERIKSSON P, JIMÉNEZ C, BUEHLER S A. Qpack, a general tool for instrument simulation and retrieval work[J]. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 2005, **91**(1): 47-64
- [17] BUEHLER S A, MENDROK J, ERIKSSON P, *et al.* ARTS, the Atmospheric Radiative Transfer Simulator - version 2.2, the planetary toolbox edition[J]. *Geoscientific Model Development*, 2018, **11**(4): 1537-1556
- [18] BERENS P. CircStat: A MATLAB toolbox for circular statistics[J]. *Journal of Statistical Software*, 2009, **31**(10): 1-21