大气二氧化碳贝叶斯反演中误差矩阵的构建方法研究

叶函函,王先华,吴 军,方勇华

(中国科学院安徽光学精密机械研究所 通用光学定标与表征技术重点实验室,安徽 合肥 230031)

摘 要: 针对 GOSAT(Greenhouse gases Observing Satellite)近红外波段数据进行的大气二氧化碳含量反演,提出了贝叶斯理论中误差矩阵的构建方法并进行了分析验证。首先模拟不同的初始估计值、不同的先验信息误差矩阵 $\boldsymbol{S}_a$ 和测量误差矩阵 $\boldsymbol{S}_\varepsilon$ 构建结果对 CO_2 反演的影响,然后据此挑选出 $\boldsymbol{S}_a$ 和 $\boldsymbol{S}_\varepsilon$ 的优劣两种构建结果组合分别对 2009 年塔克拉玛干沙漠地区的部分 GOSAT 观测数据进行反演验证。研究结果表明先验信息方差越大或测量误差越小,反演结果趋于一致,反之结果则较为离散,表现出明显的规律性。实际大气遥感研究中准确的误差矩阵难以获取,此研究有助于进一步选取准确值并提高反演精度。

关键词: 反演; 贝叶斯理论; 误差矩阵; GOSAT

中图分类号: TN21 **文献标志码:** A **文章编号:** 1007-2276(2014)01-0249-05

Error matrix construction method for atmospheric carbon dioxide Bayesian retrieval

Ye Hanhan, Wang Xianhua, Wu Jun, Fang Yonghua

(Key Laboratory of Optical Calibration and Characterization, Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei 230031, China)

Abstract: Based on the retrieval of atmospheric carbon dioxide using infrared spectral data detected by GOSAT, a method of building error matrixes of Bayesian theory was proposed and validated. Firstly, the effect on retrieval results by different initial guesses, different building results of priori information error matrix $\boldsymbol{S}_a$ and measurement error matrix $\boldsymbol{S}_\varepsilon$ was simulated, and then two combinations of $\boldsymbol{S}_a$ and $\boldsymbol{S}_\varepsilon$ building results were validated in the retrieval using part of GOSAT measurements of Taklimakan desert during 2009. The result shows that the retrieved results are more concentrate in the case of bigger priori information variance or smaller measurement error, and the retrieved results are dispersed under the converse circumstance. It is difficult to get real error matrixes in atmospheric remote sensing, therefore, this study will be significant for getting more accurate error matrixes and improving retrieval precision.

Key words: retrieval; Bayesian theory; error matrix; GOSAT

收稿日期: 2013-05-07; **修订日期:** 2013-06-08

基金项目: 国家自然科学基金(41175037,41071232);公益性行业(气象)科研专项(GYHY201106045-1)

作者简介: 叶函函(1987-),女,博士生,主要从事温室气体高精度反演方面的研究。Email:yehhgo@gmail.com

导师简介: 王先华(1963-),男,研究员,主要从事光学遥感机理、遥感图像分析与遥感信息提取等方面的研究。Email:xhwang@aiofm.oc.cn

0 引 言

温室气体地面观测网可以精确地观测 CO_2 浓度并及时获知 CO_2 浓度的变化，但有限的探测器数目和低空间覆盖率使其面对关键的碳循环科学问题时望而却步[1]。近年来，空基探测 CO_2 的技术和实施成为研究热点，SCIAMACHY(Scanning Imaging Absorption Spectrometer for Atmospheric Chartography)、GOSAT 和 OCO(The Orbiting Carbon Observatory)均用于探测全球温室气体，目前在轨 GOSAT 已提供相关高光谱分辨率的探测数据，有望用来对碳循环科学进行深入研究。

近红外波段富含大气二氧化碳的诊断性光谱特征，有利于其识别和反演。原理上，目标气体的信息可直接利用近红外波段的观测辐射光谱通过求解辐射传输方程来获得，然而，事实上求解这一方程是不可能的，这主要人们无法提供足够的参数，本质上是非适定问题，即其解是不唯一的，不稳定的[2]；同时由于观测误差的存在，使得直接求解公式法和逐步逼近法这些简单的单步反演法在实际应用时容易导致较大的反演误差，降低算法的抗扰能力。基于贝叶斯理论的高斯-牛顿算法，具有较少待反演参数个数的优点，使用实时探测数据替代非待反演参数，如实时温湿压廓线来自美国国家环境预报中心(National Centers for Environmental Prediction，NCEP)，再分析资料，从而将欠定问题转化为适定或者超定问题，并利用先验信息对反演结果进行约束和修正，然后采用最优化估计方法进行求解，提高反演精度和抗干扰能力。文中拟采用这一算法在大气 CO_2 反演中进行尝试。

1 反演方法

卫星观测结果可表示为：

$$y=F(x,b)+\varepsilon \tag{1}$$

式中：y 为卫星观测辐射向量，利用强吸收光谱通道与附近近似无吸收通道的比值替代绝对辐亮度 y，可减小地表反照率不确定性对反演结果的影响，并降低观测设备噪声的干扰[3]；x 为待反演的参数值向量；b 为附加的非待反演参数；$F(x,b)$为描述大气物理过程的前向模型；ε 为测量误差。

1.1 贝叶斯反演理论

贝叶斯方法是一种求解有噪声反演问题的非常好的手段。在具备目标参数先验知识和前向模型信息的基础上，可以通过贝叶斯理论将观测信息映射到目标参数状态空间上，得到待反演参数最合理的解。贝叶斯反演有多种求解方法，其中最常用的是贝叶斯最大后验解[4]，假定测量误差和先验知识误差均服从高斯分布，定义如下的代价函数$\Phi(x)$(Cost Function)，将求解辐射传输方程转化为使代价函数最小化的最优化数学问题，最小化 $\Phi(x)$即可得到待反演参数的最大后验解。

$$\Phi(x)=(x-x_a)\boldsymbol{S}_a^{-1}(x-x_a)+[y-F(x,b)]^{\mathrm{T}}\cdot \boldsymbol{S}_\varepsilon^{-1}[y-F(x,b)]\rightarrow\min \tag{2}$$

式中：$\boldsymbol{S}_a$ 是描述 x_a 不确定性的先验信息误差协方差矩阵；$\boldsymbol{S}_\varepsilon$ 是从比值观测值得到的测量噪声协方差矩阵的估计，描述 y 和前向模型不确定性的观测误差协方差矩阵。可得方程解为：

$$x=x_a+\boldsymbol{S}_a\boldsymbol{K}^{\mathrm{T}}\boldsymbol{S}_\varepsilon^{-1}[y-F(x,b)] \tag{3}$$

式中：$\boldsymbol{K}=\partial F/\partial x|_{x=\hat{x}}$为权函数矩阵。根据 $F(\hat{x})=F(x_i)+\boldsymbol{K}_i(\hat{x}-x_i)+O(\hat{x}-x_i)^2$ 对当前估计值 x_i 进行线性化，可得约束型高斯-牛顿迭代法求解的一般形式：

$$x_{i+1}=x_a+(\boldsymbol{K}_i^{\mathrm{T}}\boldsymbol{S}_\varepsilon^{-1}\boldsymbol{K}_i+\boldsymbol{S}_a^{-1})^{-1}\boldsymbol{K}_i^{\mathrm{T}}\boldsymbol{S}_\varepsilon^{-1}[y-F(x_i)+\boldsymbol{K}_i(x_i-x_a)] \tag{4}$$

式中：x_{i+1} 为 CO_2 廓线的第 i+1 次迭代的结果；x_i 为第 i 次迭代的结果；$F(x_i)$为用第 i 次迭代解通过辐射传输模式计算出来的辐射值向量；K_i 为用第 i 次迭代解计算出来的核函数；上标 T 和-1 分别表示矩阵转置和矩阵求逆[5]。

1.2 误差矩阵的构建

由于反演问题的病态性，必须使用先验信息和测量误差估计来约束反演值的范围。确定先验信息误差协方差和观测误差协方差的取值是最优化估计反演算法中最重要的部分之一，不准确的估计值将完全抹灭高光谱分辨率模拟光谱最大程度重现目标气体信息的能力[6]。初始估计值代表在历史经验的基础上对某个参数值最合适的估计，并假定初始估计值误差为高斯分布。理论上先验信息应利用多年大气历史观测资料或模型数据来构建，且选择待反演参数较真实的初始廓线信息可提高反演精度[7]，但是通常先验信息误差协方差矩阵 $\boldsymbol{S}_a$ 和测量误差协方

差矩阵 $\boldsymbol{S}_\varepsilon$ 的准确值都难以获取，特别是观测设备噪声、大气干扰和弱强度信号等误差源存在的复杂性和可变性使得测量误差难以被准确定量化。

代价函数的求解原则是在先验信息和测量结果之间寻求一种折衷，先验信息占的权重大，则反演值越接近于初始估计值，反之，则反演值越接近于测量结果所体现的目标信息。CO_2 先验值矩阵对反演结果的影响最大，据已有研究知 CO_2 全球变化约为 12 ppm[8]。文中针对基于 GOSAT 数据进行的多波段组合的大气 CO_2 反演，在将待反演参数简化为仅为大气 CO_2 含量的情况下，设先验估计的协方差矩阵即为其方差 $\boldsymbol{S}_a=\sigma_a^2$。

假定仪器各波段测量误差相互独立，且与相应波段的测量值呈一定的比例关系，比例因子为 SF，则构建测量误差协方差矩阵如下：

$$\boldsymbol{S}_\varepsilon=\sigma_y^2\boldsymbol{V}_\varepsilon \tag{5}$$

式中：$\boldsymbol{V}_\varepsilon$ 为单位矩阵；σ_y^2 为各波段的测量方差，文中定义其为 $(y/SF)^2$，的值代表在测量结果的基础上，对观测设备噪声、大气干扰和弱强度信号等导致的测量误差的量化程度。在实验室理想环境内，其物理意义类似于仪器信噪比 SNR。

2 反演试验

2.1 GOSAT 数据

GOSAT 是日本于 2009 年 1 月 23 号发射的世界第一颗从太空观测两大主要温室气体的人造卫星。GOSAT 的热红外至近红外碳观测传感器傅里叶变换光谱仪 TANSO-FTS 以天底观测模式接收经地表反射的太阳短波近红外辐射和由地表及大气发射的热红外辐射，观测波段为可见至短波近红外波段和热红外波段，光谱分辨率约 0.2 cm^{-1}。

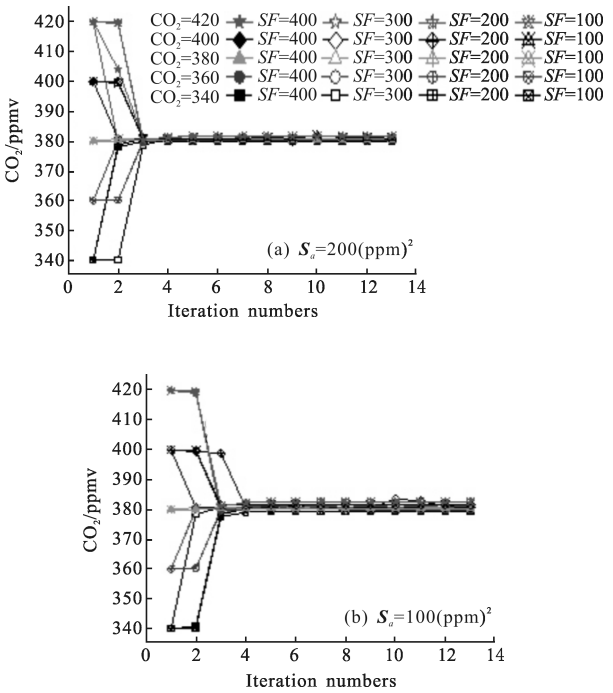
一般来说，用于气体浓度反演的波段应该具有较强的主动或被动辐射强度，从而有利于实现高信噪比信号的探测；而在气体吸收方面，其强度应适中，不会因为吸收太弱而对浓度的变化不敏感，也不会因为太强而易于饱和。因此，尽管二氧化碳在中远红外具有很强的吸收，但这却不是很好的波段，相反近红外对反演二氧化碳浓度却非常有利。注意到近红外波段的 CO_2 1.58 μm(6 300~6 400 cm^{-1})吸收带，它不仅受其它分子的干扰小，而且辐射光谱强度对温度的依赖性也较弱，它提供了近地表 CO_2 浓度的相关信息[9]，所以选择该波段内的傅里叶变换光谱仪探测的原始辐射光谱 L1B 数据来反演大气二氧化碳的浓度。

2.2 试验方案

为了验证上述 $\boldsymbol{S}_a$ 和 $\boldsymbol{S}_\varepsilon$ 构建及取值方法的可行性，利用中国典型地域塔克拉玛干沙漠的 FTS L1B 数据进行研究。首先任意选取一组塔克拉玛干沙漠的观测数据，研究 CO_2 初始估计值分别为 340 ppm、360 ppm、380 ppm、400 ppm 和 420 ppm，方差值 $\boldsymbol{S}_a$ 分别为 1(ppm)^2、25(ppm)^2、100(ppm)^2 和 200(ppm)^2，以及 SF 分别为 400、300、200 和 100 的情况下的反演值，通过比较得到较优及较差的初始估计值、$\boldsymbol{S}_a$ 和 $\boldsymbol{S}_\varepsilon$ 构建方法的组合。并选择较优的一组和较差的一组分别对 2009 年下半年沙漠的部分观测数据进行反演，根据反演值的准确性来验证参数构建方法的可行性。

2.3 结果分析

利用 2009 年 11 月 18 日的塔克拉玛干沙漠观测数据进行研究，其对应的大气 CO_2 混合比 380.5 。图 1 展示了不同的 $\boldsymbol{S}_a$ 取值情况对反演结果的影响，值大的情况下先验信息对反演值的约束能力弱，反演结果接近于实际测量中包含的大气 CO_2 信息，反



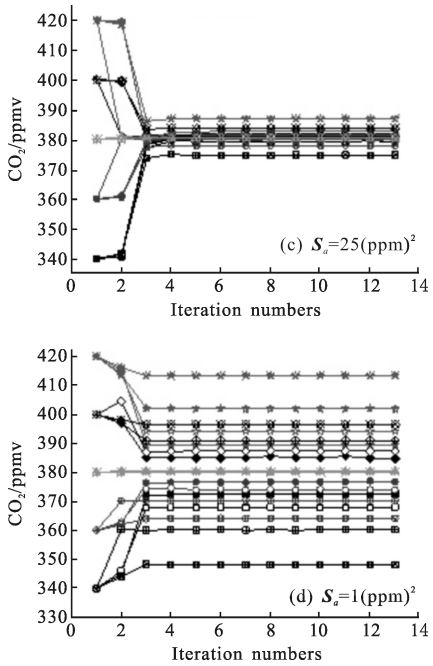


图 1 不同的 S_a 取值情况对反演结果的影响

Fig.1 Retrieved results under different S_a

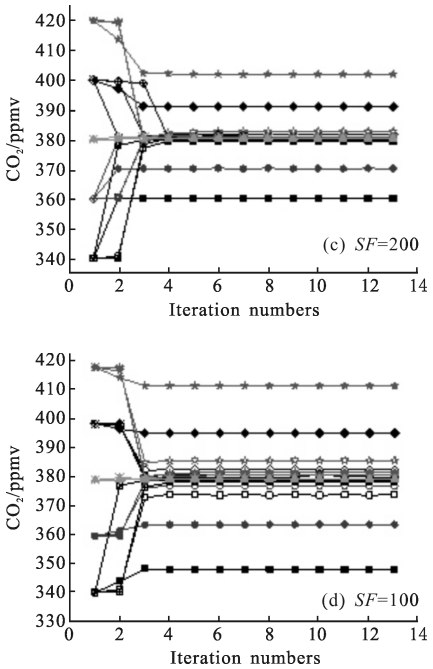
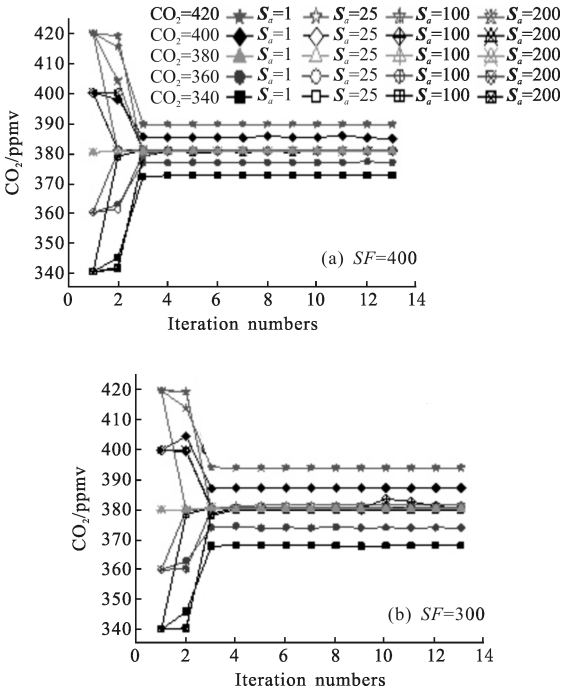


图 2 不同的 SF 取值情况对反演结果的影响

Fig.2 Retrieved results under different SF

之在 S_a 值小即约束能力强的情况下，反演结果更趋近于初值估计值。S_a 较小时结果较为离散，反演结果可靠性差。图 2 展示了不同的 SF 取值情况对反演结果的影响，SF 较大即 S_ε 较小时，反演结果受 S_a 的影响较小，反演结果较集中，可靠性好，反之 SF 非常小即 S_ε 非常大时，反演可靠性差。



2.4 验 证

由图 1 和图 2 可看出，在初始估计值为 380 ppm 的情况下，不论 S_a 和 S_ε 的取值如何，其反演结果均保持稳定良好，可见选择接近真实状况的初始估计值，对解决 S_a 和 S_ε 难以准确量化的反演问题是十分有效的。初始估计值为 380 ppm 时其在不同 S_a 和 SF 的情况下反演结果如图 3 所示，由图可见不同 和组合得到的反演结果优劣不等。S_a 和 SF 越大则先验信息约束力较弱，且测量误差较小结果也较集中；S_a 和 SF 越小则此时反演结果离散型较大且可靠性差。则据此将不同的组合分为两类，如表 1 所示。

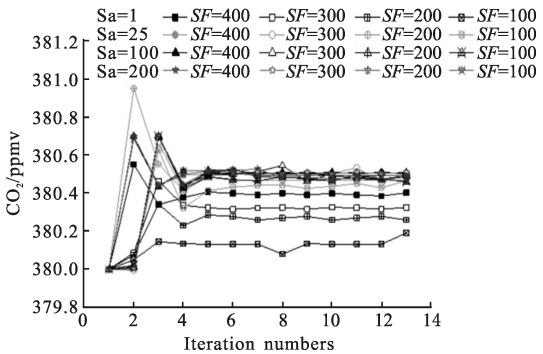


图 3 初始估计值为 380 ppm 时不同 S_a 和 SF 的情况下反演结果

Fig.3 Retrieved results under different S_a and SF in prior value 380 ppm

表 1 S_a 和 SF 的组合分类

Tab.1 Catalog of combinations of S_a and SF

Good combination		Bad combination	
S_a	SF	S_a	SF
200	100,200,300,400	1	100,200,300,400
100	100,200,300,400	25	100,200
25	300,400	–	–

从较优组合中选择出组合 1 为 S_a=200、SF=400，从较差组合中选择出组合 2 为 S_a=1、SF=100，对如表 2 所示的 8 次塔克拉玛干沙漠的 GOSAT FTS 观测数据进行反演，下表列出了这些数据的具体信息及相应的真值和反演值。由表中不同组合得到的反演值与真值的偏差可知,较大 S_a 和较大 SF 的组合 1 产生的结果确实优于较小 S_a 和较小 SF 的组合 2,S_a 和 SF 越大偏差越小且结果越集中。

表 2 两种组合得到的反演值与真值的比较

Tab.2 Comparison of retrievals of two different combinations and true values

Num	Date	Real CO_2 /ppm	Combination 1 retrieval /ppm	Error /%	Combination2 retrieval /ppm	Error /%
1	2009-06-14	382.414	382.570	0.04	380.377	-0.53
2	2009-07-11	375.63	373.179	-0.65	378.680	0.81
3	2009-10-30	376.82	379.048	0.59	379.714	0.77
4	2009-11-02	380.511	382.675	0.57	374.640	-1.54
5	2009-11-11	369.302	367.866	-0.34	376.985	2.08
6	2009-12-02	382.128	381.936	-0.05	380.578	-0.41
7	2009-12-08	373.361	377.056	0.99	378.319	1.33
8	2009-12-23	375.825	374.072	-0.47	378.319	0.66

3 结　论

构建良好的先验信息误差协方差矩阵 S_a 和测量误差协方差矩阵 S_ε 是大气 CO_2 准确反演的基础。这两个参量的构建涉及到的数据十分广泛，事实上一般只对部分参量进行假设,由此造成了不同程度的误差，这是目前制约大气 CO_2 反演精度的重要因素。依据贝叶斯理论关键误差矩阵构建的需求,利用前人有关大气 CO_2 变化的研究成果，并结合波段比值的反演方法，构建了先验信息误差协方差矩阵 S_a 和测量误差协方差矩阵 S_ε,通过对 GOSAT TANSO-FTS 数据进行的反演实验,显示了良好的反演结果。由于这种方法的简单性，使得有关大气 CO_2 的反演易于开展,并获得较好的精度。

文中是基于一个地区和较少数据进行的探索，其规律性和反演精度仍有待于进一步的研究。针对目前工作的不足,下一步需要从草地、森林、城市等不同的地理条件与不同的季节等时间条件，进行深入的规律探寻和反演精度的统计与分析。

参考文献:

[1] Mao Jianping, S Randolph Kawa. Sensitivity studies for space-based measurement of atmospheric total column carbon dioxide by reflected sunlight [J]. ***Applied Optics***, 2004, 43(4): 914-927.

[2] Li Jun, Zeng Qingcun. Infrared remote sensing of clear atmosphere and its inversion problem [J]. ***Chinese Journal of Atmospheric Sciences***, 1997, 21(1): 1-9. (in Chinese)

[3] Christi M J, Stephens G L. Retrieving profiles of atmospheric CO_2 in clear sky and in the presence of thin cloud using spectroscopy from the near and thermal infrared: a preliminary case study[J]. ***Journal of Geophysical Research***, 2004, 109(D04316): 1-11

[4] 詹海刚,施平,陈楚群. 基于贝叶斯反演理论的海水固有光学特性准分析算法[J]. 科学通报, 2006, 51(2): 204-210.

[5] Rodgers C D. Inverse Methods for Atmospheric Sounding: Theory and Practice [M]. Singapore: World Scientific, 2000.

[6] Rayner P J, O'Brien D M. The utility of remotely sensed CO_2 concentration data in surface source inversions [J]. ***Geophysical Research Letters***, 2001, 28(1): 175-178.

[7] Eguchi N, Saito R, Saeki T, et al. A priori covariance estimation for CO_2 and CH_4 retrievals [J]. ***Journal of Geophysical Research***, 2010, 115(D10215): 1-16.

[8] Dufour E, F Bréon M. Spaceborne estimate of atmospheric CO_2 column by use of the differential absorption method: error analysis[J]. ***Applied Optics***, 2003, 42(18): 3953-3609

[9] Ye Hanhan, Wang Xianhua, Wu Jun, et al. Sensitivity studies for retrieval of atmospheric column carbon dioxide with high accuracy [J]. ***Journal of Atmospheric and Environmental Optics***, 2011, 6(3): 208-214. (in Chinese)