Retrieving Method of Differential Optical Absorption Spectroscopy Based on M-estimator Robust Regression*

MA Jin-ji[1,3], LI Su-wen[2,3]

(1 *College of Physics and Electronic Information, Anhui Normal University, Wuhu, Anhui* 241000, *China*)

(2 *College of Physics and Electronic Information, Huaibei Coal Industry Teachers College, Huaibei, Anhui* 235000, *China*)

(3 *Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei* 230031, *China*)

Abstract: With differential optical absorption spectroscopy (DOAS), trace gases in the atmosphere can be measured and its concentrations can be retrieved based on least squares regression. Under complicated atmospheric conditions, there are outliers and the error distribution is not normal in DOAS differential spectra, which resulted in misestimate of least-squares regression. The retrieving model of robust regression based on M-estimator was developed to evaluate concentrations of trace gases in DOAS system. The evaluation procedure and effects using M-estimator robust regression were studied. The normal spectrum and abnormal spectrum were retrieved basing on two methods. Experimental results show that reliability is improved with method of robust regression in DOAS evaluation.

Key words: DOAS; Rrobust regression; M-estimator; Least squares regression

CLCN: O433 **Document Code**: A **Article ID**: 1004-4213(2009)08-2035-5

0 Introduction

Differential optical absorption spectroscopy (DOAS) has become a widely used method to research and measure trace gases in the atmosphere. DOAS uses the narrow molecular absorption bands to identify trace gases and their differential absorption intensity to retrieve trace gases concentration based on least squares regression[1].

DOAS is the integration of optical, mechanical and electronic technologies. The trace gas measurement is continuously and automatically performed in DOAS system[2-4]. The outliers and heteroscedasticity will appear in DOAS measurement under complicated atmospheric environment. The retrieving model of robust regression based on M-estimator was developed to process the absorbing spectra in DOAS evaluation.

1 Experiment

Differential optical absorption spectroscopy permits sensitive, specific, and temporally well-resolved in situ measurements. Fig. 1 shows the typical experimental setup of a DOAS system to measure tropospheric trace-gas concentrations. DOAS setup consists of light source, the transmitting and receiving telescope, a spectrograph, retro-reflector arrays, an optical fiber, a detector and the computer.

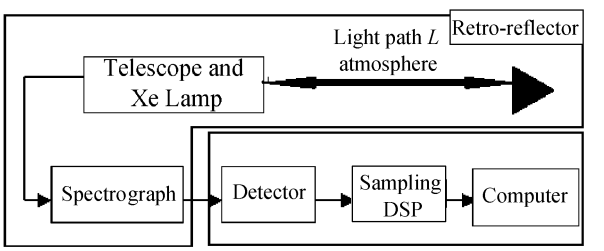

Fig. 1 Schematic view of a DOAS system

1.1 Basic principle

Lambert-Beer's law is the basic principle behind the DOAS technique[1-2]

$$I(\lambda)=I_0(\lambda)\exp(-L\sigma(\lambda)c) \quad (1)$$

Where $I_0(\lambda)$ and $I(\lambda)$ denote the initial emitted light intensity and received light intensity, respectively. λ denotes wavelength and the absorption cross section $\sigma(\lambda)$, c denotes the average concentration of trace gases at optical path L[3-4].

The basic idea of DOAS is the separation of the cross section $\sigma(\lambda)=\sigma_i(\lambda)+\sigma_i'(\lambda)$ in a part $\sigma_i(\lambda)$ that represents narrow spectral structures and $\sigma_i'(\lambda)$ representing the broad spectral features[5-6]. The Rayleigh extinction and Mie extinction by aerosols is described by $\varepsilon_R(\lambda)$ and $\varepsilon_M(\lambda)$, and $B(\lambda)$

*Supported by National Natural Science Foundation of China(40701132), the Key Project of Chinese Ministry of Education (209057), Anhui Provincial Natural Science Foundation (090412028), and Foundation of Anhui Educational Committee(KJ2008A114)

Tel: 0553-3883563 Email: jinji_ma@yanhoo.com.cn

Received date: 2008-03-29 Revised date: 2008-05-08

is the noise depending on $I(\lambda)$.

$$I(\lambda)=I_0(\lambda)\exp\{\sum_{i=1}^{n}[(-\sigma_i(\lambda)-\sigma_i'(\lambda))c_i-\varepsilon_R(\lambda)-\varepsilon_M(\lambda)]L\}+B(\lambda) \quad (2)$$

The logarithm of the ratio of the measured $I(\lambda)$ and $I_0(\lambda)$ is expressed by

$$\ln[I_0(\lambda)/I(\lambda)=\sum_{i=1}^{m}[(-\sigma_i(\lambda)-\sigma_i'(\lambda))c_i-\varepsilon_R(\lambda)-\varepsilon_M(\lambda)]L+B'(\lambda) \quad (3)$$

A polynomial of a specified degree is used to filter the 'slow' variation. So the differential optical density is defined as

$$\ln(I_0/I)=\sum_{i=1}^{m}\sigma_i c_i L+B'(\lambda) \quad (4)$$

Then the concentration of trace gases is retrieved by fitting reference absorption spectra to the differential optical density[5-6], which used a stochastic model based on least-squares regression.

The stochastic model may be written in matrix notation as

$$U=V+\Gamma \quad (5)$$

$$Z=Vk+\eta \quad (6)$$

where $v_{ij}=L_j\sigma_j'(\lambda_i)c_{j,\mathrm{ref}}$, $v_{i,0}=1$, $i=1,2,3\cdots\cdots n$, $z_i=\ln(I_0/I)$, V and Γ are $n\times(m+1)$ matrices, Z and η are $n\times 1$ vectors and k is a $(m+1)\times 1$vector. Assuming that $\{\eta_i\}$ is not normally distributed and$E(\eta_i)=0$, $\mathrm{Var}(\eta_i)=\omega_\eta^2$. It is straightforward to get an estimate of the coefficient vector k in this model if we assume that $\Gamma=0$. $\hat{k}$ is the value of estimation, the model is as follow

$$Z=V\hat{k}+e_i \quad (7)$$

Least squares estimation minimizes $\sum_{i=1}^{n}e_i^2$. The fitting coefficients based on least squares will be wrong if $\{\eta_i\}$ is not normally distributed, in particular there is heteroscedastic.

For temporal fluctuation of Xe lamp, the resolution of detector and the interference caused by different absorptive species existing under complicated atmospheric environment, the outliers and heteroscedasticity appear in DOAS system. Therefore, robust regression based on M-estimator was developed to retrieve the concentrations of trace gases in DOAS system.

The M-estimator minimizes the objective function[7-8]

$$\sum_{i=1}^{n}\rho(e_i)=\sum_{i=1}^{n}\rho(z_i-v_{ij}k) \quad (8)$$

Here, the function ρ gives the contribution of each residual to the objective function. Assume $\varphi(k)=\rho'(k)$, $\hat{k}$ should meet with: $\sum_{i=1}^{n}\varphi(\eta_i)v_{ij}=0$. Based on simulated experiments in DOAS system, objective function φ is the Andrews function

$$\varphi(k)=\begin{cases}\sin(k/c) & |k|\leqslant c\pi\\ 0 & |k|>c\pi\end{cases}\quad c>0 \quad (9)$$

Define the weight function $w(e)=\varphi(e)/e$, and let $w_i=w(e_i)$, then the estimating equations may be written as

$$\sum_{i=1}^{n}w_i(z_i-v_i'k)v_i'=0 \quad (10)$$

Solving the estimating equations is a weighted least squares problem, minimizing $\sum w_i^2e_i^2$. The weights, however, depend upon the residuals, the residuals depend upon the estimated coefficients and estimated coefficients depend upon the weights[9-10]. An iterative solution is therefore required

1)Select initial estimates $k^0=\hat{k}$, such as the least squares estimates.

2)At each iteration t, calculate residuals $e_i^{(t-1)}$ and associated weights $w_i^{(t-1)}=w[e_i^{(t-1)}]$ from the previous iteration.

$$e^t=e^{t-1}+\Delta e^{t-1}\quad w^{(t-1)}=w[e^{t-1}] \quad (11)$$

3) Solve for new weight least squares estimates.

$$k^t=[V^TW^{(t-1)}V]^{-1}V^TW^{(t-1)}Z \quad (12)$$

Step 2 and 3 are repeated until the estimated coefficients converge. The asymptotic covariance matrix of k is (13)

$$v(k)=\frac{E(\varphi^2)}{[E(\varphi')]^2}(X'X)^{-1} \quad (13)$$

Using $\sum[\varphi(e_i)]^2$ to estimate $E(\varphi^2)$, and $[\sum\varphi'(e_i)/n]^2$ to estimate $[E(\varphi')]^2$ produces the estimated asymptotic covariance matrix.

Using $\sum[\varphi(e_i)]^2$ to estimate $E(\varphi^2)$, and $[\sum\varphi'(e_i)/n]^2$ to estimate $[E(\varphi')]^2$ produces the estimated asymptotic covariance matrix.

1.2 Example

In order to exemplify the results of presented procedures based on the M-estimator, the two types of spectra, which attained by DOAS system, were retrieved. The one was normal spectrum (Fig. 2(a)), the other was abnormal spectrum (Fig. 2(b)). After they were corrected offset, dark current and a polynomial of a specified degree was used to filter the 'slow' variation, the differential optical densities were obtained and shown in Fig. 3, in which (a) and (b) were normal and abnormal spectra, respectively. The abnormal values were still existence in analyzing spectra.

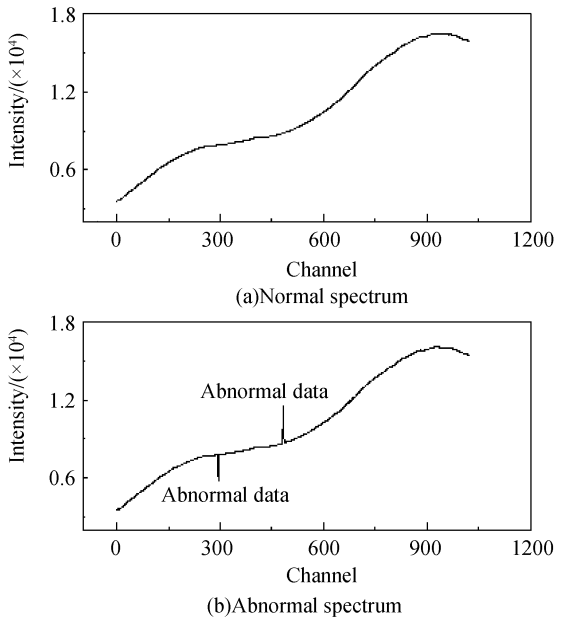

Fig. 2 Raw atmospheric spectrum

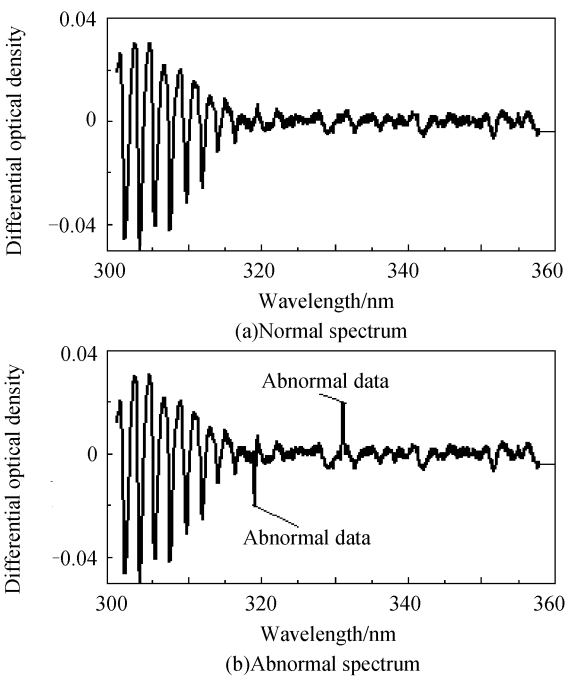

Fig. 3 Differential optical density

Fig. 4 was fitting examples of abnormal spectra, which was overlay of all of trace gases absorption. Least squares estimation approached the abnormal points in order to minimize $\sum_{i=1}^{n} e_i^2$. The weights for robust estimator decline when $|e|>k$, therefore fitting results were accurate. In the differential absorption, we saw the absorptions of O_3, NO_2, SO_2, HCHO and Lamp. The reference spectra of O_3, NO_2, SO_2, HCHO and Lamp were shown in Fig. 5. Wavelength 308.5 nm～322.6 nm was used to retrieved O_3 since it had stronger absorption feathers, 332.3～359.3 nm for NO_2, 299.6～306.6 nm for SO_2, 313.07～325.9 nm for HCHO. Lamp reference was used to remove its

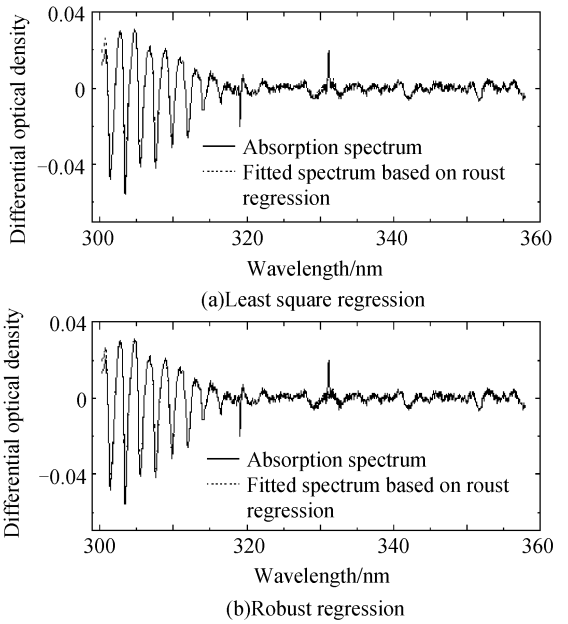

Fig. 4 Fitting examples of abnormal spectra based on least square and robust regressions

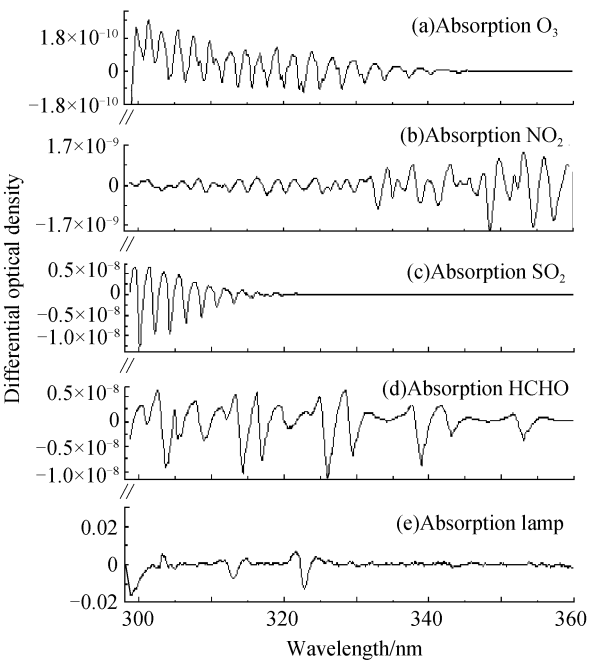

Fig. 5 Differential absorption spectrum

influence on evaluation of trace gases.

2 Analysis

In order to compare robust regression based on M-estimator with the least squares method, the simulated measurement spectra were composed of SO_2, HCHO, NO_2, O_3, lamp and residuals. The fitted results and relative errors were shown in Tab. 1 and Tab. 2. When there were not abnormal values in measuring spectra, both of the least squares fitting and robust regression based on M-estimator can give accurate values, and relative errors were lower (seen in Tab. 1). In case abnormal values appeared (seen in Fig. 3(b)), the

roust regression based on M-estimator attained right values and relative errors were below 10%. But least squares method gave the wrong values. HCHO error reached up 14%, especially O_3 error was 100% (seen Tab. 2). For O_3 cross section is the smallest of all other gases[1-2], O_3 concentration was biggest in measuring spectra. The influence on the trace gases of the smaller cross section is significant. Therefore, there are outliers, and the distribution of random errors is not Gaussian in DOAS measurement, robust regression based on M-estimator must be applied to process DOAS spectral.

Table 1 Results of normal spectra fitting and relative errors

Sample		Least squares regression		Robust regression	
Molecules	cell values (ppb)	fitting values	relative errors	fitting values	relative errors
SO_2	28	27.5	1.8%	27.4	1.8%
NO_2	45	45.7	1.5%	44.8	0.4%
HCHO	15	15.5	3.3%	14.3	5.3%
O_3	80	84.2	5.3%	83.8	4.8%

Table 2 Results of abnormal spectra fitting and relative errors

Sample		Least squares regression		Robust regression	
Molecules	cell values (ppb)	fitting values	relative errors	fitting values	relative errors
SO_2	28	29.1	3.9%	27.3	2.5%
NO_2	45	48.2	7.1%	42.5	5.6%
HCHO	15	17.1	14%	14.0	6.7%
O_3	80	-0.08	100%	73.1	8.6%

3 Conclusions

The presented robust regression based on M-estimator for DOAS spectra analysis shows that it is necessary to use the new method if there are outliners in the measuring spectra, and errors distribution is not abnormal in DOAS system. Test results show that reliability is improved with method of robust regression, especially for the trace gases of the smaller cross section. In principle, the new methods are not confined to DOAS and may also appropriate for application to other spectroscopic techniques such as tuneable diode laser spectroscopy and Fourier transform infrared spectroscopy.

References

[1] STUTZ J, PLATT U. Numerical analysis and estimation of the statistical error of differential optical absorption spectroscopy measurement with least-squares methods [J]. *Appl Opt*, 1996, **35**(30): 6041-6053.

[2] LI Su-wen, XIE Pin-hua, LI Yu-jin, *et al*. A method of data processing based on wavelet transform in differential optical absorption spectroscopy [J]. *Acta Optica Sinica*, 2006, **26**(11): 1601-1604.

[3] LI Su-wen, XIE Pin-hua, LIU Wen-qing, *et al*. A study of applicability of light emitting diodes in differential optical absorption spectroscopy measurements [J]. *Acta Phys Sinica*, 2008, **57**(3): 646-650.

[4] PENG Fu-min, XIE Pin-hua, ZHANG Ying-hua, *et al*. Effect of spectral resolution on measurement of trace gases in atmosphere by differential optical absorption spectroscopy [J]. *Acta Optica Sinica*, 2008, **28**(9): 1643-1648.

[5] LI Ang, XIE Pin-hua, LIU Wen-qing, *et al*. Monitoring of total emission volume from pollution sources based on passive differential optical absorption spectroscopy [J]. *Acta Optica Sinica*, 2007, **27**(9): 1537-1542.

[6] LI Su-wen, LIU Wen-qing, XIE Pin-hua, *et al*. Effect of water vapor absorption on measurements of atmospheric nitrate radical by LP-DOAS [J]. *Chinese Journal Of Chemical Physics*, 2008, **21**(5): 433-437.

[7] POOR H V. Signal detection in the presence of weakly dependent noise-part ii: robust detection [J]. *IEEE Transactions on Information Theory*, 1982, **28**(5): 744-752.

[8] HO W H, TSAI J T, CHOU J H. Robust-stable and quadratic-optimal control for TS-fuzzy-model-based control systems with elemental parametric uncertainties [J]. *IET Control Theory Appl*, 2007, **1**(3): 731 - 742.

[9] WELSH A H. Robust estimation of smooth regression and spread functions and their derivatives [J]. *Statist Sinica*, 1996, **6**: 347-366.

[10] MAEVSKII V V, KHARIN YU S. Robust regressive forecasting under functional distortions in a model [J]. *Automation and Remote Control*, 2002, **63**(11): 1803-1820.

基于稳健回归 M 估计的差分吸收光谱反演方法

麻金继[1,3] 李素文[2,3]

(1 安徽师范大学 物理与电子信息学院,安徽 芜湖 241000)

(2 淮北煤炭师范学院 物理与电子信息学院,安徽 淮北 235000)

(3 中国科学院安徽光学精密机械研究所 合肥 230031)

摘 要:基于最小二乘回归,差分吸收光谱技术(DOAS)可以获得痕量气体的大气浓度.鉴于在复杂大气环境下,测量结果可能出现异常值以及误差的非正态分布,导致最小二乘回归估计偏差较大.针对这一情况,本文研究了利用稳健回归 M 估计来反演 DOAS 测量光谱数据的方法,讨论了估计过程和效果,并对正常谱和异常谱进行两者回归方法比较.研究结果表明基于稳健回归 M 估计方法收到了良好的效果,提高了回归可靠性.

关键词:DAOS;稳健回归;M 估计;最小二乘回归

MA Jin-ji was born in 1975. Dr. Ma is active in research pertaining to optical remote sensing foundation and technology. The current areas of research include quantitative of optical remote sensing, remote sensing image analysis and information extraction.