

# 对差分光学吸收光谱法的监测 数据进行实时预测研究\*

齐 锋 刘文清 周 斌 李振壁 崔延军

(中国科学院安徽光学精密机械研究所, 合肥 230031)

(2002 年 4 月 7 日收到, 2002 年 8 月 23 日收到修改稿)

差分光学吸收光谱法(DOAS)已经成为测量大气中微量气体浓度广泛应用的方法. 通过对大气差分吸收光谱的分析可以得到它们的浓度. 但在恶劣气候条件下, DOAS 系统不能获得连续的实时监测数据, 因此不能满足我国环境监测条例对环境监测子站的要求. 文中提出了一种利用逐步回归分析的方法, 对 DOAS 系统的监测数据进行实时预测. 通过大量的对比实验表明, 利用该方法得到的预测结果与实际测量结果一致, 目前该技术已经被应用于 DOAS 监测仪器中.

关键词: 差分光学吸收光谱法(DOAS), 逐步回归分析, 预测

PACC: 8670L, 0260, 0365D

## 1. 问题的提出

随着全球工业的迅猛发展, 对环境的污染越来越严重, 尤其在发展中国家, 环境监测及治理手段还都比较落后. 我国已经把环境保护确立为一项长期坚持的基本国策, 确定了环境与经济社会协调和可持续发展的战略.

对环境监测目前普遍采用化学方法和光谱学方法. 差分光学吸收光谱法(DOAS)便是光谱方法的典型代表.

### 1.1. DOAS 方法

DOAS 方法是通过窄带分子吸收波段来区分微量气体和它们的吸收强度来得到对流层或平流层中微量气体的浓度. 一种典型的测量平流层中气体浓度的 DOAS 系统的安装如图 1 所示<sup>[1,2]</sup>. 强度为  $I_0(\lambda)$  的光, 由一个合适的光源发出, 通过大气, 最后被望远镜接收. 在光通过大气的路径中, 光经历了不同种类微量气体的吸收和气体分子与雾颗粒的散射. 根据 Lambert-Beer 公式, 在光路末端的光强  $I(\lambda, L)$

表示为

$$I(\lambda, L) = I_0(\lambda) \exp \left\{ \int_0^L \sum_j [ -\sigma_j^{\text{ABS}}(\lambda, p, T) \times \rho_j(l) ] - \epsilon_R(\lambda, l) - \epsilon_M(\lambda, l) dl \right\} + N(\lambda). \quad (1)$$

第  $j$  种气体的吸收由它的吸收截面  $\sigma_j^{\text{ABS}}(\lambda, p, T)$  来标识, 它依赖于波长、压力和温度, 以及它在光路上  $l$  处的密度  $\rho_j(l)$ . 悬浮颗粒的瑞利散射和米氏散射通过  $\epsilon_R(\lambda, l)$  和  $\epsilon_M(\lambda, l)$  描述.  $N(\lambda)$  是光子的噪声, 它依赖于  $I(\lambda, L)$ .

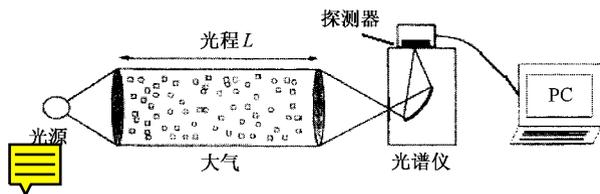


图 1 典型的 DOAS 系统组成图

DOAS 的基本思想是将吸收截面分成两部分  $\sigma_j^{\text{ABS}} = \sigma_j^{\text{B}} + \sigma_j'$ , 其中  $\sigma_j^{\text{B}}$  代表宽带光谱特征, 而差分吸收截面  $\sigma_j'$  代表了窄带光谱特征. 当仅仅考虑  $\sigma_j'$  时, 将会避免瑞利散射和米氏散射的影响<sup>[3,4]</sup>.

\* 中国科学院知识创新工程(批准号: KGCX2-403)资助的课题.

### 1.2. 存在问题

由(1)式可知,DOAS方法在分析过程中,光强度是至关重要的参数,尤其是望远镜接收的光强 $K(\lambda, L)$ ,但是 $K(\lambda, L)$ 的大小除了与污染气体的吸收有关外,很大程度受气候条件的制约,如在大雨或者大雾天气时, $K(\lambda, L)$ 会变得很低,以至于无法满足计算的需要,使得不能正确分析出微量气体的浓度.为了使DOAS系统能满足我国环境监测条例要求,应用到环境监测子站中,必须解决恶劣条件下,DOAS系统能正常给出数据.本文提出了利用逐步回归分析,对DOAS系统监测的历史数据进行分析,找出影响监测结果的显著因素,对污染气体浓度进行实时预报的方法.

## 2. 逐步回归分析模型

回归模型是传统的数据处理方法,通过给定模型方程式的结构,利用已知数据确定公式系数,在自然规律的预测预报中,已得到广泛的应用,具有成熟的理论和应用基础.

### 2.1. 逐步回归分析模型

逐步回归的通用模型为

$$y = b_0x_0 + b_1x_1 + \dots + b_{l-1}x_{l-1} + b_l, \quad (2)$$

其中 $b_0, b_1, \dots, b_l$ 是未知参数, $x_0, x_1, \dots, x_{l-1}$ 是 $l$ 个一般变量,它们的值是可以实验精确测量的.为了建立回归方程(2),估计回归系数 $b_0, b_1, \dots, b_l$ ,可以进行 $n$ 次观察,得到 $n$ 组观察数据 $(x_{i,0}, x_{i,1}, \dots, x_{i,l-1}, y_i), i=1, 2, \dots, n (n > l)$ ,将 $n$ 组观测值代入方程(2),可得到如下回归方程组:

$$\begin{cases} y_1 = b_0x_{1,0} + b_1x_{1,1} + \dots + b_{l-1}x_{1,l-1} + b_l, \\ y_2 = b_0x_{2,0} + b_1x_{2,1} + \dots + b_{l-1}x_{2,l-1} + b_l, \\ \dots\dots\dots \\ y_n = b_0x_{n,0} + b_1x_{n,1} + \dots + b_{l-1}x_{n,l-1} + b_l. \end{cases} \quad (3)$$

(3)式是一个超越方程(因为 $n > l$ ),回归系数 $b_0, b_1, \dots, b_l$ 可通过最小二乘法进行估计<sup>[5]</sup>.

(2)式是逐步回归分析的通用模型,在用该模型对微量气体浓度进行实时预报时,首先要定性的确定影响微量气体浓度的因素,也就是(2)式中的自变量 $x_i (i=0, 1, \dots, l-1)$ ,然后利用历史数据,通过逐步回归方法,找出各自变量的系数 $b_i (i=0, 1, \dots, l-1)$ 和常数项 $b_l$ .自变量及其系数确定后,便可通

过(2)式对微量气体浓度进行预报.逐步回归分析的优点在于它分析过程中,不断的进行自变量(影响因子)的显著性检验,如果判定自变量不是影响因变量的显著性因子,则自动剔除.

### 2.2. 影响微量气体浓度因素定性分析

通过长期的实践发现,大气中微量气体浓度的变化是有规律的,这就使得对其进行实时预报成为可能.

在确定对微量气体浓度的影响因素上,必须保证在预报的时刻这些影响因素是能够得到的.中科院安徽光学精密机械研究所研制的环境监测子站系统包括:DOAS21C-X(利用DOAS原理,监测微量气体浓度),气象5参数(温度,湿度,气压,风速,风向)和PM10(空气中直径小于 $10\mu\text{m}$ 的颗粒物)在线监测.故对于该系统来说,能够得到比较重要的气象5参数,这5个参数和大气中微量气体的浓度可能存在一定的关系.

**温度** 在实际监测中,我们发现微量气体的浓度每天都有一定的趋势,同时,对于一个地区来说,每天的温度变化也是有一定规律的,所以可以认为它们之间存在着一定的关系,故温度可以作为一个影响因素.

**湿度** 湿度的大小在一定程度上反映了当天的气候,如雨天湿度大,晴天湿度小,而空气的微量气体浓度显然是和晴天、雨天有关系的,故湿度也作为一个影响因素.

**气压** 由于气压是海拔的函数,而对应一个点来说气压基本上是不变的,故尽管微量气体的浓度可能与气压有关,但对于同一个点来说,这个常量可以不考虑.

**风速和风向** 在实践中我们发现,当刮某个方向的风时,微量气体的浓度变得较高,而其他方向的风与无风时微量气体的浓度相比变化不大或者很小,这说明风速和风向也是影响污染物浓度的因素.这是很容易理解的,如在监测点的北方有污染源,则显然风向来自北方时监测到的污染物浓度就会变高.

除以上几个影响因素外,我们还知道污染物的浓度在两个时刻不会发生突变(因为时间间隔比较短,对应DOAS21C-2约为3min),其变化是一个相对缓慢的趋势,同时我们发现,每天污染物的浓度的变化趋势有一定的相似性,如图2是一个浓度变化

趋势的对比图.图中  $L$  为 2001 年 9 月 24 日 4:00 到 2001 年 10 月 2 日 17:56 在桂林市某监测子站监测的  $\text{SO}_2$  浓度曲线, $O$  为  $L$  曲线向左移动 24h 后的曲线,这样,对应图中的某一时刻, $L$  图和  $O$  图分别是

该时刻和昨天这个时刻的  $\text{SO}_2$  浓度值.从图中可以发现,在大部分时间里,这两条曲线有很大的相似性.因此我们将前一时刻的污染物浓度和对应该时刻的前一天的浓度作为两个影响因素.

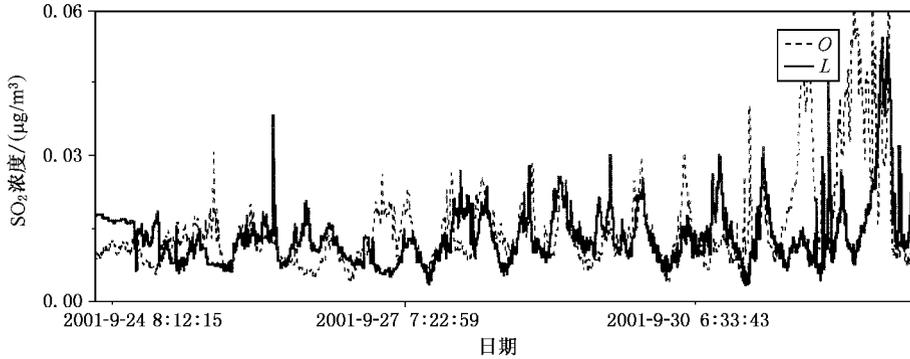


图 2 不同日期同一时刻的  $\text{SO}_2$  浓度对比图 曲线  $L$  为 2001 年 9 月 24 日 4:00 至 2001 年 10 月 2 日 17:56  $\text{SO}_2$  的浓度曲线,曲线  $O$  为将  $L$  数据向左平移 24 h 后的浓度曲线

经过以上的初步分析(2)式可以转换为

$$c = b_1 T + b_2 H + b_3 W_s + b_4 W_d + b_5 c_p + b_6 c_1 + b_7 \quad (4)$$

(4)式即为微量气体浓度的实时预报模型.式中  $c$  为要预测时刻的浓度值; $T, H, W_s, W_d$  分别为要预测时刻的温度,湿度,风速和风向; $c_p, c_1$  分别为要预测时刻的上一时刻的值和前一天对应该时刻的值; $b_i (i = 1, 2, \dots, 7)$  为回归系数.

(4)式中  $T, H, W_s, W_d, c_p, c_1$  对应(2)式中的自变量  $x_i$ ; $b_i (i = 1, 2, \dots, 7)$  的含义同(2)式.

如假设仪器正常工作时 3min 产生一组数据,如果要预测 2002 年 5 月 12 日 12:40 的数据,则  $T, H, W_s, W_d$  分别为从气象参数仪得到的 2002 年 5 月 12 日 12:40 的值, $c_p$  的值为 2002 年 5 月 12 日 12:37 微量气体的浓度测量值, $c_1$  的值为 2002 年 5 月 11 日 12:40 该微量气体浓度的测量值.

### 2.3. 逐步回归分析过程及结果

逐步回归分析是对多元线性回归进行因子筛选,最后给出一定显著性水平下各因子均为显著的回归方程中的诸回归系数.逐步回归分析的算法是一个筛选的过程,其具体步骤如下<sup>[6]</sup>.

1)首先作出  $(l+1) \times (l+1)$  的规格化的初始相关系数矩阵.

$$R = \begin{bmatrix} r_{0,0} & r_{0,1} & \dots & r_{0,l-1} & r_{0,y} \\ r_{1,0} & r_{1,1} & \dots & r_{1,l-1} & r_{1,y} \\ \dots & \dots & \dots & \dots & \dots \\ r_{l-1,0} & r_{l-1,1} & \dots & r_{l-1,l-1} & r_{l-1,y} \\ r_{y,0} & r_{y,1} & \dots & r_{y,l-1} & r_{y,y} \end{bmatrix}$$

矩阵中的各元素为

$$r_{i,j} = \frac{d_{i,j}}{d_i d_j} = \frac{\sum_{t=0}^{n-1} (x_{t,i} - \bar{x}_i)(x_{t,j} - \bar{x}_j)}{\sqrt{\sum_{t=0}^{n-1} (x_{t,i} - \bar{x}_i)^2} \cdot \sqrt{\sum_{t=0}^{n-1} (x_{t,j} - \bar{x}_j)^2}} \quad i, j = 0, 1, \dots, l-1, l.$$

自变量的个数为  $l, n$  为观测的样本数.

$r_{i,j} (i, j = 0, 1, \dots, l-1)$  为自变量  $x_i$  与  $x_j$  的相关系数, $r_{y,j} (i, j = 0, 1, \dots, l-1)$  为因变量  $y$  与自变量的相关系数, $r_{y,y}$  为因变量的自相关系数. $d_{i,j}$  为自变量  $x_i$  与  $x_j$  的协方差, $d_i$  和  $d_j$  分别为  $x_i$  和  $x_j$  的方差.式中

$$\bar{x}_i = \sum_{t=0}^{n-1} x_{t,i} / n, \quad i = 0, 1, \dots, l-1.$$

当  $i = l$  时,有

$$\bar{x}_l = \bar{y} = \sum_{t=0}^{n-1} y_t / n.$$

2)计算偏回归平方和

$$V_i = \frac{r_{i,y} r_{y,i}}{r_{i,i}}, \quad i = 0, 1, \dots, l-1.$$

3) 若  $V_i < 0$ , 则对应的  $x_i$  是已被选入回归方程的因子.

从所有的  $V_i < 0$  的  $V_i$  中选出  $V_{\min} = \min |V_i|$ , 其对应的因子为  $x_{\min}$ . 然后检验因子  $x_{\min}$  的显著性. 若

$$\frac{\varphi V_{\min}}{r_{yy}} < F_2,$$

则剔除因子  $x_{\min}$ , 并对系数相关阵  $R$  进行该因子的消元变换 (转 2).

4) 若  $V_i > 0$ , 则对应的  $x_i$  为尚待选入回归方程的因子.

从所有的  $V_i > 0$  的  $V_i$  中选出  $V_{\max} = \max\{V_i\}$ , 其对应的因子为  $x_{\max}$ . 然后检验因子  $x_{\max}$  的显著性. 若

$$\frac{(\varphi - 1)V_{\max}}{r_{yy} - V_{\max}} \geq F_1,$$

则剔除因子  $x_{\max}$ , 并对系数相关阵  $R$  进行该因子的消元变换 (转 2).

上述过程一直进行到无因子可剔可选为止.

在以上步骤中,  $\varphi$  为相应的残差平方和的自由度.  $F_1$  及  $F_2$  均为  $F$  (分布值), 它们取决于观测点数、已选入的因子数及取舍显著性水平  $\alpha$ . 通常取  $F_1 > F_2$ , 当选入单个因子的显著性水平取为  $\alpha$  时, 则可以从  $F$  分布表中取  $m = 1$ , 观测点数为  $n$  时的  $F_\alpha$  为  $F_2$ , 而取  $m = 1$ , 观测点数为  $n - 1$  时的  $F_\alpha$  为  $F_1$ .

当要剔除或选入某个因子  $x_k$  时, 均需对系数相关阵  $R$  进行该因子的消元变换, 其算法为

$$r_{i,j} = r_{i,j} - \frac{r_{i,j} r_{i,k}}{r_{k,k}} \quad i, j = 0, 1, \dots, l; \quad i, j \neq k.$$

$$r_{k,j} = r_{k,j} / r_{k,k} \quad j = 0, 1, \dots, l; \quad j \neq k.$$

$$r_{i,k} = -r_{i,k} / r_{k,k} \quad i = 0, 1, \dots, l; \quad i \neq k; \quad r_{k,k} = 1 / r_{k,k}.$$

当筛选结束时, 就可得出规格化回归方程的各回归系数  $b_i (i = 1, 2, \dots, 7)$ , 其中值为 0 的系数表示对应的自变量可剔除.

利用逐步回归方法, 我们对广西桂林市某监测子站从 2001 年 9 月 24 日到 2001 年 10 月 2 日监测到的 3000 组历史数据进行了分析, 最后得到  $SO_2$  回归方程如下:

$$\begin{aligned} c = & -7.80 \times 10^{-5} T + 1.58 \times 10^{-5} H \\ & -2.56 \times 10^{-4} W_s + 7.58 \times 10^{-6} W_d \\ & + 0.92 c_p + 1.93 \times 10^{-3} c_1 - 0.0015. \end{aligned} \quad (5)$$

图 3 为利用得到的回归方程, 对桂林市某子站的 2001 年 10 月 3 日和 4 日的 800 组数据进行实时

预测和真实测量的结果对照图, 曲线 A 为利用 DOAS 原理对测得的光谱进行分析得到的结果, 曲线 B 为利用回归方程计算的结果, 表 1 是预测结果与实测结果的比较.

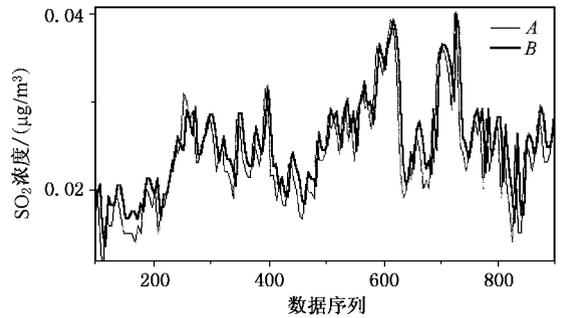


图 3 DOAS 系统的实测数据与逐步回归方法预测数据对比图  
曲线 A 为利用 DOAS 方法的实测数据, 曲线 B 为由逐步回归方法得到的预测数据

表 1 预测结果与实际监测结果比较

样本数	误差均值	最大误差	最小误差	标准方差
800	$2.84812 \times 10^{-5}$	0.00385	$3.28655 \times 10^{-11}$	$7.20855 \times 10^{-6}$

从图 3 和表 1 能够清楚的看出, 经过逐步回归分析得到的回归方程对 DOAS 系统测量的微量气体浓度的预测与实际测得结果非常相近. 在恶劣气候下, DOAS 方法无法正常工作时, 完全可以相信利用回归方程预测的结果.

### 3. 回归方法的实际应用与效果

在 DOAS 系统中加入逐步回归分析模型, 对 DOAS 的数据进行实时预测的目的就是克服 DOAS 的自身缺陷, 使 DOAS 系统能够更加适合我国环境监测子站的要求. 图 4 是回归模型嵌入软件系统中后, 软件的数据处理流程. 其工作过程, 首先通过光谱获取系统得到光谱数据, 然后通过气象参数仪得到气象 5 参数, 气象 5 参数也是环境监测子站常规监测的项目之一, 同时在 DOAS 系统中, 测量的结果要根据当时的气象 5 参数转换成标况 (1 个大气压, 273.15K) 下的结果, 另外在回归模型中也要用到气象 5 参数. 得到气象参数后, 软件系统要做两件事, 其一是判断信号强度是否足够大, 使得利用 DOAS 方法进行分析能够得到可靠的结果, 同时它要根据气象参数以及系统测量的前一组数据和昨天同时刻的数据对测量结果进行实时预测. 接下来的工作取决于信号强度, 如果信号强度能够满足 DOAS 方法

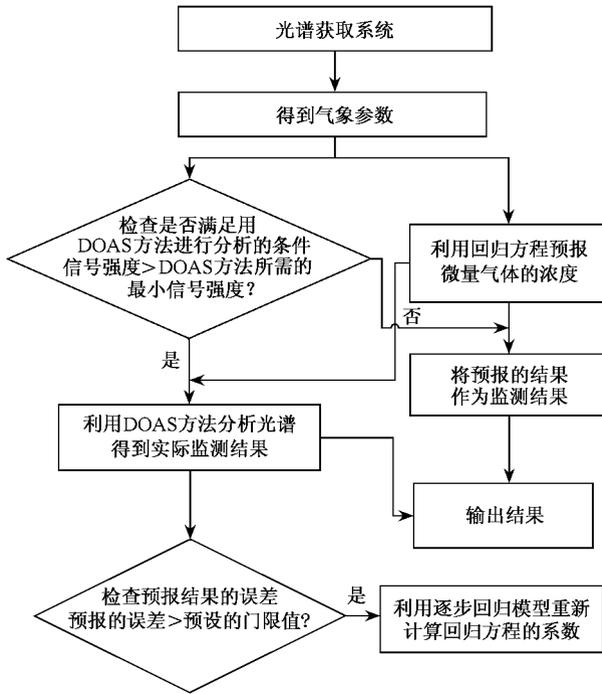


图 4 回归模型嵌入软件系统后的数据处理流程

的要求,则最终的输出结果取 DOAS 的实际测量值,同时,将比较实际测量值与实时预测值的误差,如果在设定的范围内,则进入下一个循环测量周期,否则要将新得到的实际测量数据融入历史数据,重新调用逐步回归模型算法,得到回归方程的系数。如果信号强度很弱,则输出预测的结果作为本次测量的结果。

利用逐步回归分析方法对 DOAS 系统的监测数据进行实时预测结果与实际的监测结果具有极大的相似性,并且在出现偏差较大时,软件系统自动重新进行学习校准,在大量的实验中获得了良好的结果,该技术被应用于 DOAS21C-2 的软件系统 CNDOAS 中,DOAS21C 空气质量监测系统目前已经安装在广西的两个国家重点监测城市桂林和北海,并已经成功的向国家环境监测总站上报数据近一年,目前这两个监测站已经纳入了国家常规监测网,监测数据由中央电视台向全国播报。经过长时间的实践,在恶劣的气候条件下,DOAS21C-2 系统仍能预报出与其他不受气候条件影响的仪器相一致的监测结果,克服了 DOAS 系统自身的缺陷,大大拓宽了 DOAS 系统的应用领域。

- [ 1 ] Platt U and Perner D 1979 *Geophys. Res.* **84** 6329  
 [ 2 ] Stutz J and Platt U 1996 *Applied Optics* **35** 6041  
 [ 3 ] Zhou B, Liu W Q and Qi F *et al* 2001 *Acta Phys. Sin.* **50** 1818 ( in Chinese ) 周 斌、刘文清、齐 锋等 2001 *物理学报* **50** 1818 ]  
 [ 4 ] Zhou B, Liu W Q and Qi F *et al* 2001 *Research of Environment Sciences* **14** 23 ( in Chinese ) 周 斌、刘文清、齐 锋等 2001 *环境科学研究* **14** 23 ]

- [ 5 ] Fang K T, Quan H and Chen Q Y 1988 *Applied Regression Analysis* ( Beijing Science Press ) p205 ( in Chinese ) 方开泰、全 辉、陈庆云 1988 *实用回归分析*(北京 科学出版社)第 205 页 ]  
 [ 6 ] Xu S L 1996 *C program sets for frequently used Algorithm* 2nd Edition ( Beijing Tsinghua university press ) p236 ( in Chinese ) 徐士良 1996 *C 常用算法程序集第二版*(北京 清华大学出版社)第 236 页 ]

# Forecasting of the real-time data monitored by differential optical absorption spectroscopy method<sup>\*</sup>

Qi Feng Liu Wen-Qing Zhou Bin Li Zhen-Bi Cui Yan-Jun

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

(Received 7 April 2002; revised manuscript received 23 August 2002)

## Abstract

Differential optical absorption spectroscopy (DOAS) has become a widely used method to measure trace gases in the atmosphere. Their concentrations can be retrieved by a numerical analysis of the atmospheric absorption spectra. But it has been found that DOAS could not gain continuous data under a bad weather condition. Thus it could not fully fit for the demand of the environment monitoring statute. A prediction technology has been developed, based on the step-wise regression analysis method. It is shown that the forecasting data using this method is consistent with the real data after doing a great deal of experiments. This technology has been used in practice.

**Keywords** : differential optical absorption spectroscopy (DOAS), step-wise regression analysis, prediction

**PACC** : 8670L, 0260, 0365D

---

<sup>\*</sup> Project supported by the Knowledge Innovation Engineering Program of Chinese Academy of Sciences, China (Grant No. KGCX2-403).