

大气二氧化硫污染预测模型初探

杨伟民 沈建法

(华东师范大学地理系)

摘 要

考虑了影响大气扩散的主要气象因素及其它因素,提出了一个大气SO₂污染预测模型,并运用加权实时最小二乘算法对模型进行实时辨识.这一模型尤其适用于城市及城市中SO₂日均浓度的分区预测,给出了该模型在上海地区应用的结果.

关键词: 大气污染; SO₂预测模型; 加权实时最小二乘算法.

一、引 言

近年来,由于现代城市的工业和交通迅速发展,人口高度集中,工业、交通和民用因燃烧而产生的废气大量排入大气中,其中SO₂污染问题相当突出.

城市污染物排入大气后,是否造成大气污染,一方面与污染源强度有关,另一方面与气象条件有关.由于短时期内,污染源排废速率变化不大,而大气对污染物的稀释扩散能力随着气象条件的改变可发生巨大的变化.因此,气象条件是影响大气污染的最活跃和最重要的因素.

影响大气扩散能力的主要气象因素有大气稳定度和风等.如何根据这些气象及其它因子建立大气SO₂浓度预测模型,从而对大气中SO₂浓度进行预测,防止可能发生的污染危害,是污染气象面临的重要课题之一.

近年来,由于电子计算机的飞速发展,系统辨识理论的应用领域日益广泛^[1].本文运用系统辨识方法,建立了SO₂日均浓度预测模型,并编制了相应的计算机模型辨识与预测软件.在研究中采用了上海市1981—1983年SO₂实测资料和上海龙华气象台的气象资料.

二、辨识方法与预测模型

加权实时最小二乘算法是系统辨识中最具应用潜力的模型辨识方法.该方法的基本思想是:在误差函数中采用指数加权方案,以重点突出最近的数据(例如,预报明天的SO₂浓度,今天的气象数据比昨天的气象数据与之关系更为密切,同样昨天的气象数据比前天的气象数据与之关系更为密切等等);采用递推算法,按预测误差逐步校正估计参数,使参数能跟踪系统的动态变化(例如,先用气象资料作SO₂浓度的预报,其次用预报值与实际值的差值来校正下一次的模式参数,再根据当时气象数据和新的模式参数估计下一次的

SO₂ 浓度。这样模式既考虑了气象条件, 又考虑了引起误差的其它因素, 如排废速率的变化等)。因此, 这一方法尤其适用于时变参数动态系统的辨识与预测。这一算法的基本公式如下:

设待识别的系统模型为:

$$Y = \theta_1 X_1 + \theta_2 X_2 + \dots + \theta_n X_n \quad (1)$$

式中 Y 为输出; X_1, X_2, \dots, X_n 为输入; $\theta_1, \theta_2, \dots, \theta_n$ 为模型参数。

记误差函数为

$$J_m = \sum_{i=1}^m \lambda^{m-i} \varepsilon^2(i) \quad (2)$$

式中 $\varepsilon(i)$ 为第 i 次拟合误差; m 为样本总数; λ 为权数, $0 < \lambda \leq 1$ 。

运用最优化理论使误差函数 J_m 最小, 并运用矩阵求逆分解定理, 可得加权实时最小二乘算法。

$$\Gamma(m+1) = 1/[1 + \mathbf{x}^T(m+1)\mathbf{P}(m)\mathbf{x}(m+1)], \quad (3)$$

$$\hat{\theta}(m+1) = \hat{\theta}(m) + \Gamma(m+1)\mathbf{P}(m)\mathbf{x}(m+1)[y(m+1) - \mathbf{x}^T(m+1)\hat{\theta}(m)] \quad (4)$$

$$\mathbf{P}(m+1) = \frac{1}{\lambda}[\mathbf{P}(m) - \Gamma(m+1)\mathbf{P}(m)\mathbf{x}(m+1)\mathbf{x}^T(m+1)\mathbf{P}(m)]. \quad (5)$$

其中 $\hat{\theta}(m+1)$ 为第 $m+1$ 步的参数估计值

$$\hat{\theta}(m+1) = [\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_n]^T,$$

$$\mathbf{P}(m+1) = \frac{1}{\lambda} [\mathbf{x}^T(m+1)\mathbf{x}(m+1)]^{-1},$$

$$\mathbf{x}(m+1) = [x_1(m+1), \dots, x_n(m+1)]^T,$$

$$\mathbf{x}_{m+1} = \begin{bmatrix} \mathbf{x}_m \\ \mathbf{x}^T(m+1) \end{bmatrix}.$$

用式(3)一式(5)进行递推估计, 必须确定 $\hat{\theta}$ 和 \mathbf{P} 的初值, 方法是前面 m 组数据由一步估计式直接求得 $\hat{\theta}(m)$ 和 $\mathbf{P}(m)$, 从 $m+1$ 步开始向前迭代计算。

式(3)一式(5)实时估计算法具有很强的直观意义, 式(4)中 $y(m+1) - \mathbf{x}^T(m+1)\hat{\theta}(m)$ 为第 $m+1$ 次预测误差, 第 $m+1$ 步根据误差对 $\hat{\theta}(m)$ 进行校正, $\hat{\theta}(m+1)$ 取为 $\hat{\theta}(m)$ 和预测误差的加权和。

我们参考张景哲等^[2]提出的污染指数 $I_a = SP/uh$ 的形式, 提出城市 SO₂ 日平均浓度 C 的预测模式如下:

$$C = \theta(1) \frac{(s+5)^{0(2)}(C_p+5)^{0(4)}}{(u+5)^{-0(3)}} - 5 \quad (6)$$

式中 s 为稳定度等级的相对值; u 为风速的相对值(数值上取实测值); C_p 为前一天 SO₂ 日平均浓度, 可以作为污染浓度持续性的度量。

稳定度计算采用常用的帕斯廓尔-特纳尔方法。稳定度相对值见表 1。

式(6)中的常数 5 是为了保证变量能取对数而加上的。式(6)的意义是大气中 SO₂ 日平均浓度与大气稳定程度成正比; 与风速成反比; 与前一天 SO₂ 日均浓度成正比。

对式(6)线性化可得

表1 大气稳定度的相对值
Table 1 Relative value of the stability

稳定度级	A	A-B	B	B-C	C	C-D	D	D-E	E	E-F	F
相对值	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6

$$\ln(C+5) = \ln \theta(1) + \theta(2) \ln(s+5) + \theta(3) \ln(u+5) + \theta(4) \ln(C_p+5) \quad (7)$$

式(7)即为我们预测城市SO₂浓度所采用的线性化模型, 其与式(1)有类似的形式。

三、模型试验

影响大气中SO₂浓度的气象因子有大气稳定度、风、降水和混合层厚度等。但是, 这些因子之间又存在着一定的关系, 如降水时, 大气稳定度多为不稳定型或中性的。为了解影响大气中SO₂浓度的主要因子及数量, 我们运用式(3)一式(5)实时递推最小二乘算法, 用上海冬季SO₂和气象资料进行了五次估计, 每次估计所选用的变量见表2。

表2 五次计算中选用的气象变量
Table 2 Selected meteorological variables for the calculations

计算方案	气象变量
1	前一天20:00的稳定度和14:00的风速
2	前一天20:00的稳定度和20:00的风速
3	前一天20:00的稳定度和8:00—20:00的降水
4	前一天20:00的稳定度、8:00—20:00的降水、20:00的风速
5*	前一天20:00的稳定度、预报、日平均风速和前一天SO ₂ 日平均浓度

* 方案5中, 由四次定时观测风速的算术平均值为平均风速

用电子计算机对五种计算方案分别进行了估计, 检验模型优劣的统计量见表3。

表3 模型检验统计量
Table 3 Model test statistics

计算方案	RMSE(%)	S ²	AIC	λ	备注
1	33.7	0.292885	2.00046	0.85	一年资料
2	76.4	0.256564	-17.7678	1	三年资料
3	73.6	0.257929	-17.6595	1	三年资料
4	74.6	0.260018	-15.4949	1	三年资料
5	22.4	0.0513232	-7.39819	0.95	二年资料

表 3 中 $RMSE$ 为均方根模拟误差 $\left(\sqrt{\frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{m}} \times 100 \right)$; S^2 为剩余方差 $\left(\frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{m-n} \right)$; AIC 为赤池判别准则 ($AIC = m \lg S^2 + 2n$)。上述三个统计量越小, 则说明模型越好, λ 为权重。

从表 3 可见, 各方案的均方根模拟误差均小于 77%, 尤其是资料序列较短的方案 1 和方案 5, 其 $RMSE$ 仅为 33.7% 和 22.4%, 说明预测对象一个时变参数系统, 增加资料序列并不能提高预测精度, 同时也说明了采用实时递推估计算法的合理性。

此外, 还表明尽管影响大气污染的气象及其它因子很多, 但反映在模型中并非越多越好。实际计算表明, 大气稳定度、日均风速和前一天 SO_2 浓度能较好地预测大气中 SO_2 的浓度。因此, 我们选择式 (6) 作为预测模型。

四、二氧化硫实时递推预测模型的应用

以模式 (6) 为基础的城市二氧化硫实时递推预测模型必须用于连续资料的情况中。为此, 我们选用上海市区 1983 年 7 月的 SO_2 日平均资料 (7 月 11 日—25 日连续 15 天) 和对应时间的气象资料进行了模拟预测。初始参数 (第 0 次) 由 7 月 11 日—21 日的 11 组数据定出, 然后递推。计算方案的模型参数值和参数检验通过的显著性水平见表 4。

表 4 计算方案 5 的模型参数值
Table 4 Model parameters of the fifth calculation

迭代次数 \ 参数	$\ln \theta(1)$ 常数	$\theta(2)$ 稳定度	$\theta(3)$ 风速	$\theta(4)$ 前一日 SO_2 浓度
0	-3.29111	4.19571	-1.73197	0.28377
1	-4.59186	4.55815	-1.63973	0.369859
2	-3.16529	4.19016	-1.70944	0.241667
3	-2.60415	4.14099	-1.91924	0.231665
4	-1.18613	3.51295	-2.00453	0.277741
t 统计量		2.84444	-3.87982	1.43161
检验显著性水平		0.02	0.01	0.2

表 4 表明, 在各次迭代过程中, 各参数值是在逐步校正的, 但起伏不大。稳定度参数 $\theta(2)$ 为正值, 这说明 20:00 的大气越稳定, 越有利于污染物积累, 第二天 SO_2 日平均浓度就会越高, 反之则低, 风速参数 $\theta(3)$ 为负值, 亦即预报日平均风速越小, 则该日 SO_2 日均浓度越高, 反之则低。这是由于风速越小, 大气对污染物扩散稀释能力弱, 污染物不易输

送出去,故浓度高。反之风速大,浓度就低,前一天 SO_2 浓度参数 $\theta(4)$ 为正值,反映了污染源对污染浓度的持续性影响。在实际应用时,由于稳定性和前一日 SO_2 日均浓度都是预测对象的前一天值,故很易取得。而风速与预测浓度都为同时关系,为了能获得这个值,可以气象台发布的天气预报的风速值代入模式,从而作出污染浓度预报。

表5给出了本例的计算机模拟的预测值、绝对误差和相对误差。由表5可见,用模型(6)进行 SO_2 日均浓度预测精度较高。 SO_2 日均浓度的预测值与实测值较为接近,即绝对误差很小,相对误差也很小,预测结果的相对误差都在36%以下,其中有1/3小于10%,均方根模拟误差 $RMSE$ 仅为18.9%。

表5 实例的预测值、绝对误差和相对误差
Table 5 Forecasted values and its errors

序 号	预 测 值 ($\mu\text{g}/\text{m}^3$)	绝 对 误 差 ($\mu\text{g}/\text{m}^3$)	相 对 误 差 (%)
1	38.1462	9.14616	31.5385
2	19.7121	-0.287865	-1.43932
3	15.5368	-0.463224	-2.89515
4	41.8844	-5.11562	-10.8843
5	36.6538	-7.34625	-16.696
6	49.5379	2.53787	5.39972
7	55.7245	14.7245	35.9135
8	27.4086	1.40856	5.41752
9	40.4929	-7.50707	-15.6397
10	43.8078	-9.19218	-17.3437
11	45.0454	-3.95464	-8.0707
12	41.6342	-18.3659	-30.6098
13	46.6545	9.65448	26.0932
14	27.1594	3.15943	13.1643
15	30.0644	4.06444	15.6324

$$RMSE=18.9678 \quad S^2=0.0376081 \quad S=0.193928 \quad AIC=-9.37078 \quad \lambda=0.95 \quad \alpha=0$$

表6 全市 SO_2 日均浓度预测结果的相对误差(两年结果汇总)

Table 6 Relative error of the forecasted results for Shanghai area

相 对 误 差 $E\%$	频 率 (%)
$\leq 10\%$	31.4
$\leq 20\%$	67.8
$\leq 30\%$	84.7
$\leq 50\%$	96.6
$> 50\%$	3.4

为了说明用模式(6)计算结果的稳定性与可靠性(计算精度不受个别例子的影响),我们应用模式(6),按计算方案5对上海城市1982年和1983年的 SO_2 资料进行计算,得出表6。从表6可见,每百次预测中,相对误差小于10%的为31次,小于20%的为68次,小于30%的为85次,相对误差在50%以上还不足4次。由此可见,用模型(6)进行城市 SO_2 日均浓度预测精度是很高的,而且两年结果与个例情况大致相符,因而预测结果又是稳定的。

五、城市分区二氧化硫日均浓度预测

在城市内, SO_2 日均浓度并非到处均一, 各区(行政区)的平均浓度差异常常很大, 这是因为城市中各区的污染源源强差异以及风向、风速对污染物扩散方向和输送强度影响所造成。

我们运用系统辨识方法, 建立了城市中分区 SO_2 日均浓度预测模型, 作为分区预测工作的初步尝试。下面以上海杨浦区为例叙述。

1. 预测模型

我们考虑到风向、风速和源强是影响分区 SO_2 浓度的主要因素, 提出分区 SO_2 日均浓度预测模式如下:

$$C_F = \frac{\theta(1)}{(U+5)^{-0.2}} - 5 \quad (7)$$

式中 C_F 为分区 SO_2 日均浓度; U 为风向、风速和污染源组合参数。 $U = \frac{u}{F}$, 其中 u 为预报日平均风速; F 为风向-污染源加权因子(简称风向加权因子)。

杨浦区位于上海市区的东北角, 为上海主要工业区之一。该区东面和南面与川沙县仅一江(黄浦江)之隔, 污染较轻, 吹偏东风和南风时, 相对值取为 1。西面为虹口区, 该区燃料消耗比东面和南面多, 但仅及杨浦区的一半, 吹偏西风时, 相对值取 2。杨浦区工业主要分布在该区北部(监测站位于主要工业区南面), 这一方向污染最重, 吹偏北风时相对值取为 4。由于静风时风向不定, 污染物来回往返, 故相对值最大取为 6。风向加权因子相对值见表 7。

表 7 风向-污染源加权因子取值

Table 7 Weighted value of wind direction and pollution source

风 向	偏 南 风	偏 东 风	偏 西 风	偏 北 风	静 风
F_{ij}	1	1	2	4	6

表 7 中相对值越大, 表示该方向污染可能性越大。 F_{ij} 为某一时次的风向加权因子, 如 2:00 的风向加权因子记为 F_{02} 。风向规定如表 8。

表 8 各方位风向规定

Table 8 Wind direction setting of various positions

偏 北 风	偏 东 风	偏 南 风	偏 西 风
NNW N NNE	ENE E ESE	SSW S SSE	WNW W WSW

当风向介于偏西风与偏北风之间时, 如 NW 风则相对值取偏北与偏西风 F_{ij} 值的算术平均, 即取 3, 其余类推。

我们通过试验, F 值取如下形式为佳:

$$F = \frac{1}{2} [F_{02} + F_{20}] \quad (8)$$

式中 F_{02} 和 F_{20} 分别为当日 2 点和 20 点的风向加权因子, F 基本上代表了一日中的主要风向及污染源的影响。

在收集和统计资料时注意以下几点: (1) 预报日平均风速 u 可从气象台发布的天气预报中摘用; (2) 预报日风向也同样采用气象台的预报结果按 (8) 式计算, 但须注意此时 F_{02} 并不指 02:00 的风向加权因子, 而是泛指。比如, 气象台预报明天风向偏南, 午后转偏北风, 则 $F = \frac{1}{2} [F_{02} + F_{20}] = \frac{1}{2} (1 + 4) = 2.5$, 其余类推。(3) 获得 u 和 F 后再据 $U = \frac{u}{F}$ 计算出 U 值, 然后可代入模式计算。

对式 (7) 线性化可得:

$$\ln(C_F + 5) = \ln \theta(1) + \theta(2) \ln(U + 5) \quad (9)$$

上式即为我们预测城市中分区 SO_2 浓度所采用的线性化模型。

2. 分区预测实例

以模型 (7) 为基础的分区二氧化硫实时递推预测模型必须用于连续资料的情况中。为此, 我们选用上海杨浦区 1983 年 1 月的 SO_2 日均资料 (11 日到 25 日连续 15 天) 和对应时间的气象资料进行模拟预测。初始参数由 11 日到 21 日 11 组数据定出, 然后进行递推, 计算方案的模型参数值和参数检验通过的显著性水平见表 9。

该例的权重 λ 取为 0.95。

表 9 个例的模型参数值
Table 9 Model parameters in the case study

迭代次数	参数	$\ln \theta(1)$ 常数	$\theta(2)$ 风向—污染源及风速组合参数
0		6.94656	-1.27765
1		6.99916	-1.32025
2		7.14606	-1.40708
3		7.48864	-1.59129
4		7.37443	-1.52884
t 统计量			-1.68106
检验显著性水平			0.2

计算机计算结果表明, 预测浓度与实测值很接近, 绝对误差很小, 相对误差也很小 (表 10)。

由表 10 可见, 每百次预测中, 相对误差约有 1/3 数小于 10%, 86 次小于 30%, 误差大于 40% 的仅 7 次。均方根模拟误差仅 23.2%。由此可见, 该例用模型 (7) 进行分区 SO_2 日均浓度预测精度较高。

表10 杨浦区SO₂日均浓度预测结果的
相对误差(个例)

Table 10 Relative error of the
forecasted results for
Yangpu District of Shang-
hai City

相对误差 E%	频 率 (%)
≤10%	33
≤20%	66
≤30%	86
≤40%	93
>40%	7

六、结 论

1. 预测SO₂日均浓度的模型(6)适用于城市SO₂预测;模型(7)适用于城市中分区SO₂日均浓度预测。

2. 在城市SO₂预测中,我们选择预报日前一天20:00的大气稳定度,前一天SO₂日平均浓度和预报日平均风速三个因子作为模式自变量效果较好。在不同地区推广应用时,可根据实际情况挑选合适变量,变量个数可通过模型试验确定。

在分区预测中,组合因子 U 中 F (风向加权因子)尤为重要,城市中不同的区(如行政区)加权情况是不同的,可根据该区周围污染源分布(以四周燃料消耗量为依据)来确定。

3. 模式(6)、(7)中的风速和风向都与预测浓度为同时关系,在实际应用中是引用气象台的预报值,故模式预测精度在一定程度上依赖于气象部门对风向、风速预报的准确程度。建议在应用时采用计算精度也较高的计算方案1平行进行,作为参考。

4. 应用模型(6)或模型(7)时,应借助于电子计算机,预先编好和调试好程序以及模型参数。预测时,只需将当日的气象资料(或气象台的预报值)变换为相对值或实测值输入计算机运算,很快就可得到第二天的SO₂日均浓度的预报值。

5. 权重 λ 的数值,需由计算机试验确定,其大小反映了最近的气象资料对预测的影响大小。如 $\lambda=1$,则远与近的气象资料对预测影响一样大; $\lambda<1$,则反映了越近的资料对预测的影响越大。

6. 城市和分区的二氧化硫实时递推预测模型能跟踪系统参数的动态变化,预测精度较高。相对误差小于10%的约占1/3;小于20%的约占2/3,小于30%的约4/5强。在城市预测中个例与多年结果都颇为相近。

参 考 文 献

- (1) 夏天昌,系统辨识, pp. 20—24, pp. 27—29, 国防工业出版社, 1984年.
- (2) 张景哲等,环境科学, 3(6), 15(1982).

1986年9月29日收到

A FORECASTING MODEL OF SO₂ POLLUTION IN ATMOSPHERE

Yang Weimin, Shen Jianfa

(Dept. of Geography, East China Normal University)

ABSTRACT

A model for forecasting SO₂ pollution in atmosphere was developed and a weighted real-time least square algorithm was applied to the real-time identification of the model. The results of the application of the model in Shanghai area were presented.

Keywords: air pollution; SO₂; forecasting model; weighted real-time least square algorithm.

环境科学学报 7卷3期勘误

310页	错	正
18行	$r = K'' pO_2^{\beta}$	$r = K'' \{COD\}$
19行	式中 $K'' = KC^{\alpha}$	式中 $K'' = K pO_2^{\beta}$