•争鸣与探索•

曲线拟合对空气污染物浓度的预测

卢云鹤[1], 梁　鸿[2], 黄锡坚[2], 乔妙杰[3]

(1. 深圳市南山区环境监测站, 广东　深圳　518052;
2. 深圳市环境保护监测站, 广东　深圳　518008;
3. 深圳市气象局, 广东　深圳　518001)

摘　要: 通过对深圳市 1991 年—2000 年环境空气中 SO_2 浓度月均值、年均值变化规律的分析, 得出了其数学表达式所需的函数特征, 建立了曲线拟合方程, 并通过实例计算及一系列统计回归检验, 表明 SO_2 浓度的月均值、年均值实际变化与拟合曲线结果基本一致。

关键词: 曲线拟合; SO_2; 空气污染物; 浓度预测

中图分类号: X830.3　　文献标识码: B　　文章编号: 1006- 2009(2002) 02- 0034- 03

Prediction of Air Pollutants' Concentration with Fitting Curve

LU Yun-he[1], LIANG Hong[2], HUANG Xi-jian[2], QIAO Miao-jie[3]

(1. *Nansan Environmental Monitoring Station of Shenzhen, Shenzhen, Guangdong* 518052, *China;*
2. *Shenzhen Environmental Monitoring Station, Shenzhen, Guangdong* 518008, *China;*
3. *Shenzhen Weather Bureau, Shenzhen, Guangdong* 518001, *China)*

Abstract: The monthly average value and annual average value of SO_2 in air in Shenzhen from 1991 to 2000 were studied, and its mathematical expression was reached. For case study and serial regression analysis, the results indicated that the change situation was coordinated with the change of fitting curve.

Key words: Fitting curve; SO_2; Air pollutant; Prediction of concentration

统计回归分析是进行环境空气污染物浓度预测的一种有效方法, 但是由于污染物浓度的变化十分复杂, 所得到的监测数据, 往往是多种因素共同作用的结果, 难免会发生各种各样的偏差, 得不到令人满意的预测效果。现以空气污染物浓度的月均值为研究对象, 应用曲线拟合的方法, 结合统计回归分析, 试图建立适用于污染物月均值预测的曲线拟合方程。

1　建立曲线拟合方程的步骤

1.1　曲线特征分析

分析污染物浓度月均值要依据时间变化规律和浓度变化原因找出其曲线变化特征。环境空气中各类污染物浓度随时间变化错综复杂, 归纳起来主要有两个方面影响: (1) 污染源的排放状况; (2) 气候气象的变化情况。污染物浓度的年均值变化主要与当年污染源的排放量密切相关, 而污染物浓度的月均值变化除了与当月污染源的排放状况有关外, 还与当地的气候和季节变化有关。

1.2　数学表达式

根据上述变化规律寻找有类似变化特征的数学函数, 通过一个或多个函数的复合和叠加结果进行曲线拟合, 提出初步的数学表达式。此步骤的关键在于对各类数学函数的变化特征要十分熟悉, 同时还需要掌握函数的复合和叠加的知识。

1.3　建立曲线拟合方程

应用回归分析中的最小二乘法原理, 确定上述表达式中的相关参数, 建立曲线拟合方程。观察拟

收稿日期: 2001- 09- 29; 修订日期: 2002- 02- 10

作者简介: 卢云鹤(1964—), 男, 广西玉林人, 助理工程师, 大专, 从事环境监测工作。

合曲线与实际曲线的吻合程度, 当拟合曲线出现明显违背常理的变化时, 应设置一些常数对方程中的各函数作适当的调整。

2　实例研究

现以深圳市 1991 年—2000 年全市 SO_2 月均值为例, 进行曲线拟合预测。

2.1　曲线特征分析

深圳市 1991 年—2000 年 SO_2 浓度月均值变化曲线见图 1, 年均值变化曲线见图 2。

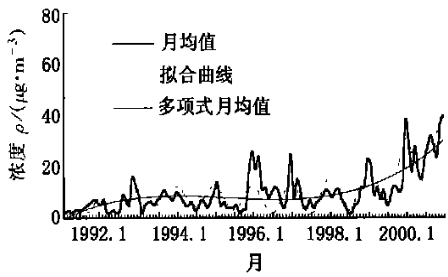

图 1　深圳市 1991 年—2000 年 SO_2 月均值

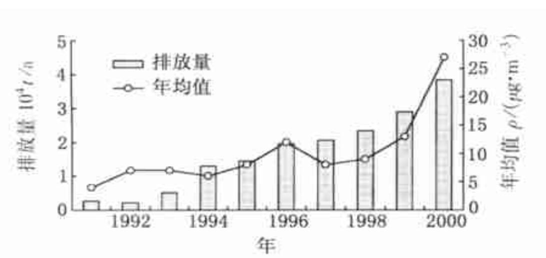

图 2　1991 年—2000 年 SO_2 年均值与排放量关系

由图 1、图 2 分析, 可以发现以下特征:

(1) 1991 年—2000 年 SO_2 浓度年均值逐年上升, 与全市 SO_2 的排放量密切相关。

(2) 在年度内, 通常表现为冬季(1 月、12 月) 浓度较高, 夏季(6 月 —8 月) 浓度较低, 月均值呈连续的周期性的季节波动。

(3) 波动的幅度逐年增大。

(4) 每年春节长假所在月份的月均值大多有一个相对低点, 这与深圳是独特的移民城市有关。

2.2　数学表达式[1]

通过对上述前 3 个变化特征的分析, 可以初步形成这样一个判断: 月均值围绕一条上升曲线作周期性的波动, 波动周期为一年。结合数学函数的特征, 得知余弦函数恰好是这样一个周期函数, 即在波动周期内具有两头高, 中间低的特点。可以将月均值围绕的这条上升曲线看做是均值变化函数, 月均值的变化曲线则可以视为是均值变化函数与周期变化函数——余弦函数的叠加。因此, 初步提出的数学表达式为:

$$f(x) = af_1(x) + bf_2(x)\cos(kx) \tag{1}$$

式中: $f(x)$ ——SO_2 月均值随月份 (x) 的变化函数;

$f_1(x)$ ——均值变化函数;

$f_2(x)$ ——振幅变化函数;

a, b ——常数。

从 1991 年 1 月开始, 依次取 $x = 1, 2, 3, \ldots, 120$(月)。由于 SO_2 月均值大多在 12 月份达到最大, 取 $k = \frac{\pi}{6}$。

2.3　建立曲线拟合方程

设均值变化函数 $f_1(x)$ 为三次多项式, 可以通过统计回归的方法求得:

$$f_1(x) = 8.441\,54 \times 10^{-5}x^3 - 0.013\,136x^2 + 0.619\,17x - 0.619\,337\,5 \tag{2}$$

考查月均值年度振幅 $A(y)$ 与年度 (y) 的对应关系, $A(y)$ 与 y 近似呈指数变化。应用最小二乘法, 可以得到:

$$A(y) = 6.124\,7e^{0.157\,2y} \tag{3}$$

从 1991 年开始, 依次取 $y = 1, 2, \ldots, 10$(年)。由于 $x = 12y$, 所以:

$$f_2(x) = 0.5 \times 6.124\,7e^{0.157\,2x/12} \tag{4}$$

将方程(2), (4) 代入方程(1), 取 $a = 1$, $b = 0.75$, 曲线拟合方程为:

$$f(x) = f_1(x) + 0.75f_2(x)\cos(\frac{\pi}{6}x) = 8.441\,54 \times 10^{-5}x^3 - 0.013\,136x^2 + 0.619\,17x - 0.619\,337\,5 + 0.75 \times 0.5 \times 6.124\,7e^{0.157\,2x/12}\cos(\frac{\pi}{6}x) \tag{5}$$

拟合方程曲线见图 1。

3　方程检验

3.1　F 检验[2]

计算月均值的实际值样本方差 $S_1^2(n = 120)$ 和拟合值样本方差 $S_2^2(n = 120)$, 得出:

$S_1^2 = 67.01$, $S_2^2 = 58.83$, $F = S_1^2/S_2^2 = 1.14$。

给定显著性水平 α= 0.01, 查 F 分布表得出临界值:

$$F_{0.005(119,119)} > F_{0.005(120,120)} = 1.61$$

$F < F_{0.005(120,120)} < F_{0.005(119,119)}$, 可以认为拟合值与实际值的样本方差无显著差异。

3.2　t 检验[3]

计算实际值平均值 $(\bar{x}_1)$ 和拟合值平均值 $(\bar{x}_2)$, 得: $\bar{x}_1 = 10.075$, $\bar{x}_2 = 10.115\,16$。

$$|t| = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{S_1^2 + S_2^2}}\sqrt{n} = 0.039$$

选择显著性水平 α= 0.02, 自由度 f 为 238, 查 t 分布表得:

$$t_{0.01(238)} > t_{0.01(\infty)} = 2.576$$

$t < t_{0.01(\infty)} < t_{0.01(238)}$, 可以认为拟合值与实际值的平均值无显著差异。

3.3　方差分析

对曲线拟合月均值与实际月均值进行方差分析, 结果见表 1。表 1 中 $r = 2$, $n = 240$。给定显著性水平 α= 0.005, 查 F 分布表得临界值:

$$F_{0.005(1,238)} > F_{0.005(1,\infty)} = 7.88$$

$F < F_{0.005(1,\infty)} < F_{0.005(1,238)}$, 可以认为曲线拟合月均值与实际月均值无显著差异。

表 1　方差分析

方差来源	离差平方和	自由度	方差	F 值
组间	$Q_A = 0.097$	$r-1=1$	$S_A^2 = 0.097$	$S_A^2/S_E^2 =$
组内	$Q_E = 14\,974.81$	$n-r=238$	$S_E^2 = 62.92$	0.001 5
总和	$Q = 14\,974.9$	$n-1=239$		

3.4　相关系数检验[4]

计算拟合值与实际值的相关系数 $r = 0.805$, 计算 t 的数值 $(n = 120)$:

$$t = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2} = 24.807$$

选择显著性水平

α= 0.005, 自由度 $n = n - 2 = 118$

查 t 分布表得: $t_{0.005(118)} < t_{0.005(30)} = 2.75$

$t > t_{0.005(30)} > t_{0.005(118)}$, 表明曲线拟合月均值与实际月均值相关性非常显著。

4　讨论

4.1　方程(1)中的均值变化函数 $f_1(x)$ 和振幅变化函数 $f_2(x)$ 与污染物排放量密切相关, 因此, 在污染物排放量的变化趋势基本维持不变的情况下, 拟合方程的预测是适用的。当预计未来污染物排放量的变化趋势发生改变时, 该拟合方程不适用。此时应对 $f_1(x)$ 和 $f_2(x)$ 的数学表达式作相应的调整。

4.2　一个地区的气候气象条件一般不会发生突变, 因此在相当长的预测时间内, 反映污染物浓度季节变化的周期函数将保持不变。

4.3　均值变化函数项反映了污染物排放量对污染物浓度变化的影响, 周期变化函数项则反映了气候气象条件对污染物浓度季节变化的影响。

4.4　对于特殊月份污染物浓度值的长期偏差, 如深圳市每年春节长假所在月份的月均值普遍偏低, 在预测时, 需要对拟合方程(5)修正。

4.5　从图 1 曲线的拟合情况看, 当 SO_2 月均值较低时, 如 1991 年—1992 年, 方程(5)拟合的情况不理想。当 SO_2 月均值较高时, 如 1998 年—2000 年, 拟合曲线的变化值与月均值的实际变化基本吻合。

4.6　当仅仅依靠统计回归分析建立预测方程时, 方程(1)和方程(5)仅剩下均值变化函数项 $f_1(x)$, 方程预测的准确性将会降低。

4.7　当同时建立几个曲线拟合方程时, 应比较拟合值与实际值之间的均方差, 从中选取均方差最小的一个作为最佳的曲线拟合方程。

5　结论

曲线拟合方程是运用数学方法对实际变化曲线最大程度的逼近, 它综合考虑了污染物排放量和气候气象条件对污染物浓度的影响, 在污染物排放量的变化趋势和气候气象条件基本维持不变的情况下, 能够得到较好的预测结果。

[参考文献]

[1]　北京大学数学系. 高等数学[M]. 北京: 北京大学出版社, 1984.

[2]　奚元福. 实用环境统计学[M]. 成都: 四川科学技术出版社, 1992.

[3]　陶永德, 杨庆霄. 数理统计[M]. 沈阳: 辽宁科学技术出版社, 1991.

[4]　李茂年, 周兆麟. 数理统计学[M]. 天津: 天津人民出版社, 1982.