

污染物浓度预测的 PPR 模型*

李祚泳

(成都气象学院大气电子工程系, 成都 610041)

摘要 采用投影寻踪回归技术的审视数据-模拟-预测新思路, 对环境监测数据进行探索性分析, 建立了污染物浓度预测的 PPR 模型. 该模型用于水污染物 BOD 和 DO 浓度预测结果, 其建模样本拟合合格率为 100%, 预留检验样本预测合格率达 83.3%.

关键词 环境污染, 污染物, 预测, 投影寻踪, 回归模型.

传统的数据分析预测法的一个共同点是采用假定-拟合-预测这样一种证实性数据分析思维方法(简称 CDA)^[1,2]. 这种方法用于高维、非正态分布数据建模有很大困难. 因此, 有必要探索新的污染物浓度预测新方法.

70 年代国际统计界发展了一种处理和析多维观测数据的投影寻踪新技术(简称 PP)^[3]. 它不需要对数据结构作任何事先假定, 而是采用审视数据-模拟-预测这样一条探索性数据分析新途径(简记为 EDA). 它对于高维、非正态分布数据的处理特别有效, 已被成功地用于诸多领域^[4-6]. 但在环境科学领域的应用未见报道.

环境污染预测问题多数具有维数高、样本容量少和非正态特征, 用传统的 CDA 分析法预测的结果往往不够理想. 因此, 本文试采用投影寻踪回归技术(projection pursuit regressive, PPR)建立污染物浓度预测模型.

1 PPR 原理及算法简介

PPR 的基本思想是将多维数据投影到低维空间上, 通过极小化某个投影指标, 寻找出反映多维数据结构或特征的投影, 在低维空间上对数据结构进行分析, 以达到研究和分析高维数据的目的. PPR 的模型如下:

设 $\bar{X} = (X_1, X_2, \dots, X_P)$ 是 P 维随机向量, $y = f(x)$ 是一维随机变量. 为了避免线性回归不能反映实际非线性情况的矛盾, PPR 采用

一系列岭函数 $G_m(Z)$ 的“和”来逼近回归函数, 即

$$f(x) \sim \sum_{m=1}^M \beta_m G_m(Z) = \sum_{m=1}^M \beta_m G_m(\bar{\alpha} \cdot \bar{X}) \quad (1)$$

式中, $G_m(Z)$ 表示第 m 个岭函数(数值函数). $Z = \bar{\alpha} \cdot \bar{X}$ 为岭函数的自变量, 它是 P 维随机向量 $\bar{X}$ 在 $\bar{\alpha}$ 方向上的投影. $\bar{\alpha}$ 也为某方向的 P 维向量, M 为岭函数的个数.

显然, 当 $M = 1$, $G_m(\bar{\alpha} \cdot \bar{X}) = \sum_{j=1}^P \alpha_m X_j$ (2) 公式(1)就成了一般多元线性回归模型.

Friedman 和 Stuetzle 提出了实现 PPR 的多重平滑回归计算软件 SMART^[7]. 其思路如下: 考虑有模型

$$y \sim \sum_{m=1}^M \beta_m G_m \left(\sum_{j=1}^P \alpha_m X_j \right) \quad (3)$$

式中, $\sum_{j=1}^P \alpha_m^2 = 1$, 系数 $\alpha_m, \beta_m, G_m(Z)$ 和最优组合项数 M_u 都是待确定的参数. 模型的核心和关键是采用分层、分组叠代交替优化方法最终估计出这些参数. PPR 模型的极小化判别仍采用最小二乘法准则: 选择适当的参数 $\alpha_m, \beta_m, G_m(Z)$ 和 M_u 组合, 使式

$$L_2 = E \left[y - \sum_{m=1}^{M_u} \beta_m G_m \left(\sum_{j=1}^P \alpha_m X_j \right) \right]^2 = \min \quad (4)$$

* 国家自然科学基金资助项目(编号: 49471048)
收稿日期: 1996-12-11

具体实现法是把全体参数分成几组,除其中一组外,都给定一初值,然后对留下的一组参数求最优.求得结果后,把这一组参数的极值点当作初值,另选一组参数在这一初值下求优,多次重复直到参数值收敛为止.即将 $\alpha_m(j=1,2,\cdots,P)$ 、 β_m 及岭函数 G_m 划入一组, $m=1,2,\cdots,M$,共有 M 组.固定其中 $M-1$ 组,而对第一组 α_m 、 β_m 和 G_m 优化求解.此时,又将其分成3个子组,分别固定其中2个子组,对第3个子组寻优,然后重复这一过程,直到 L_2 不再减小为止.

使用 SMART 计算软件进行 PPR 计算时,共有4个参数需要在运行时指定:①光滑系数 S , S 的取值范围为 $0<S<1$.它决定模型的灵敏度, S 愈小,模型愈灵敏; S 愈大,模型愈迟钝,一般初始计算时,可取中值 $S=0.5$.②模型计算样本容量 N ,其值最大取值可达500,屏显值与指定值 N 之差由程序自动视为预留检验样本数.预留数据不参与建模计算,仅作预留检验之用.③岭函数的最多个数 M 及最优个数

M_u . SMART 软件对 M 和 M_u 作了限定,要求 $M_u\leq M\leq 9$,这2个参数决定着模型寻找数据内在结构的精细程度.一般 M 、 M_u 可取5,3;6,5;6,4;9,6这几组参数进行调试,其最优值最终由计算结果分析确定.PPR的理论及算法详见文献[7,8].

2 PPR 用于 BOD 和 DO 的浓度预测

此处,将 PPR 用于洛河某河段河水污染预测.预测指标为 Y_1 :BOD 浓度(mg/L)和 Y_2 :DO 浓度(mg/L).经调查和初步分析,与 Y_1 和 Y_2 有关的7个因素为: X_1 :初始断面的BOD 浓度 L_0 (mg/L); X_2 :初始断面的氧亏浓度 C (mg/L); X_3 :水温 $T(^{\circ}\text{C})$; X_4 :河流流量 $Q(\text{m}^3)$; X_5 :排污口污水流量 $q(\text{m}^3)$; X_6 :污水中BOD 浓度 $I(\text{mg/L})$; X_7 :流过该河段所需时间 $t(\text{s})$.

该河段上共有15组监测数据,数据引自文献[9],见表1.

因为与预测量BOD 和 DO 有关的共有7

表 1 洛河污染浓度与相关因素实测值

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	Y_1	Y_2
1	6.88	-0.25	27.0	6.75	1.12	4.77	0.083	9.35	-2.66
2	6.08	-2.21	27.5	4.78	1.12	1.93	0.083	12.30	-4.02
3	2.14	-3.04	26.0	4.78	1.12	4.04	0.083	15.60	-4.59
4	5.02	-0.73	26.0	8.56	1.12	3.63	0.073	5.88	-3.96
5	7.89	-2.26	26.0	8.56	1.12	3.63	0.069	6.34	-3.02
6	2.38	-1.65	15.0	1.49	1.56	4.28	0.104	4.00	-1.74
7	1.86	-1.35	15.8	1.49	1.56	4.28	0.104	3.76	-1.47
8	1.02	-2.12	17.1	1.49	1.38	4.28	0.104	3.98	-2.33
9	1.22	-1.92	17.5	1.49	1.38	4.28	0.104	3.98	-2.19
10	0.90	-0.27	17.0	3.63	0.99	2.02	0.104	2.78	0.33
11	1.58	-0.09	17.0	3.63	0.99	2.02	0.104	1.83	0.23
12	2.78	-1.17	13.5	3.27	0.99	1.14	0.104	2.56	-0.74
13	2.10	-1.30	13.5	3.27	0.99	1.14	0.104	2.72	-0.80
14	2.32	-0.60	14.5	3.65	0.86	0.57	0.104	1.64	-0.62
15	1.96	-0.60	14.5	3.65	0.86	0.57	0.104	2.36	-0.32

个因子,而且从表1中可以看出,这些因子监测数据与BOD 和 DO 监测数据之间的关系并不呈线性关系,而是具有高维非线性,样本数又较少的特点.因此,它可用PP 技术建立BOD 和 DO 预测模型.对 Y_1 和 Y_2 分别进行PPR 分析

计算.用表1中的前12组数据建模,后3组数据预留检验.将表1中全部数据输入PPR 的计算软件 SMART,在固定因子数 $P=7$ 和建模样本数 $N=12$ 情况下,分别选择2个模型中的3个参数 S 、 M 和 M_u 的不同组合,使每个模型

计算输出的拟合效果达到最佳. 2 个模型各自参数的最佳组合见表 2. BOD 和 DO 浓度的拟合和预留预测效果分别见表 3 和表 4. 在 PPR 分析计算过程中, 还可给出各因素对 BOD 和 DO 的贡献大小的相对权值见表 5. 从表 5 可以看出因子 X_7 虽然对 BOD 影响很小, 可以忽略, 但它对 DO 的影响则最大, 不可忽略; 此外

因子 X_6 虽然对 DO 的影响较小, 但它对 BOD 的影响还是不应忽略. 因此, 在选择与 BOD 和 DO 有关的共同因子时, 这 7 个因子都应选上.

表 2 2 个模型参数的最优值最终组合

参数	N	P	S	M	M_u
$Y_1(\text{BOD})$	12	7	0.4	6	5
$Y_2(\text{DO})$	12	7	0.5	5	3

表 3 BOD 和 DO 的 PPR 拟合效果

BOD				DO			
实测值	拟合值	绝对误差	相对误差/ %	实测值	拟合值	绝对误差	相对误差/ %
9.350	9.352	0.002	0.0	-2.660	-2.666	-0.006	0.2
12.300	12.290	-0.010	-0.1	-4.020	-4.010	0.001	0.0
15.600	15.612	0.012	0.1	-4.590	-4.605	-0.015	0.3
5.880	5.886	0.006	0.1	-3.960	-3.958	0.002	-0.1
6.340	6.364	0.024	0.4	-3.020	-3.008	0.012	-0.4
4.000	3.972	-0.028	-0.7	-1.740	-1.796	-0.056	3.2
3.760	3.811	0.051	1.4	-1.470	-1.438	0.032	-2.2
3.980	3.966	-0.014	-0.3	-2.330	-2.334	-0.004	0.2
3.980	4.001	0.021	0.5	-2.190	-2.154	0.036	-1.6
2.780	2.696	-0.084	-3.0	0.330	0.324	-0.006	-1.9
1.880	1.897	0.017	0.9	0.230	0.230	0.000	0.0
2.560	2.562	0.002	0.1	-0.740	-0.736	0.004	-0.6
合格项数	12	合格率	100%	合格项数	12	合格率	100%

表 4 BOD 和 DO 的 PPR 模型预留检验结果

BOD				DO			
实测值	拟合值	绝对误差	相对误差/ %	实测值	拟合值	绝对误差	相对误差/ %
2.720	2.726	0.006	0.2	-0.800	-0.746	0.054	-6.7
1.640	1.930	0.290	17.7	-0.620	-0.526	0.094	-15.2
2.360	1.952	-0.408	-17.3	-0.320	-0.448	-0.128	39.9

表 5 各因素的相对权值

权序	$Y_1(\text{BOD})$		$Y_2(\text{DO})$	
	因素	相对权值	因素	相对权值
1	X_3	1.00000	X_7	1.00000
2	X_4	0.49887	X_4	0.56134
3	X_5	0.33300	X_3	0.38851
4	X_1	0.24886	X_1	0.38707
5	X_2	0.14523	X_2	0.26288
6	X_6	0.11085	X_5	0.11107
7	X_7	0.01021	X_6	0.06632

从表 3 可见, BOD 和 DO 浓度的预测模型拟合的相对误差均有 $|\delta K| < 4\%$, 全部合格. 从表 4 可见每个模型预留的 3 个检验样本的相对误差, 除 DO 最后一个样本预测相对误差 $|\delta K| > 20\%$ 属不合格外, 其余 5 个均有 $|\delta K| < 20\%$, 因此总的预留检验合格率为 83.3%.

3 结论

(1) PPR 用于污染物预测建模, 只需要原始监测数据, 避免了人为干预, 客观性好.

(下转第 44 页)

表 5 化合物的 BOD₅/ COD 比值

化合物	BOD ₅ / COD 比值
硝基苯	~0
苯	< 0.39
苯 胺	0.56
苯 酚	0.84

2.3 基质浓度对生物降解速率的影响

有机物的浓度对降解速率也有影响. 实验结果表明, 难降解的有机物硝基苯只有在保持很低的浓度下才能被降解, 且浓度越大, 越难降解. 从表 3 中可看出, 浓度从 0.1mg/L 逐渐增大到 10mg/L, 而 K 值则从 $0.085d^{-1}$ 逐渐减少到 0. 这恰好可用化学结构的影响来说明, 由于硝基苯生态学方面的毒性及硝基的存在对细菌酶的抑制, 使得浓度越大, 其毒性及抑制作用越强. 而苯的生物降解速率随浓度的变化关系则与硝基苯刚好相反, 浓度从 0.1mg/L 逐步增大到 10mg/L, K 值亦由 $0.021d^{-1}$ 逐步增大到 $0.068d^{-1}$. 苯胺、苯酚在浓度为 1mg/L 以下时, 符合硝基苯的变化规律.

2.4 溶解度对生物降解速率的影响

苯酚、苯胺的溶解度要大于苯和硝基苯的溶解度^[3], 而实验得到的降解速率结果亦是前者大于后者. 说明溶解度小, 生物降解速率也小, 难溶化合物一般要比易溶化合物保留的时

间长. 这可能是因为化合物很难进入微生物细胞, 增加了在惰性物质中的吸附或夹带, 使其难以进行反应.

3 结论

(1) 应用方程 $Y(BOD) = L_0(1 - 10^{-Kt})$ 及其相应的参数 K 和 L_0 模拟计算在天然海水中有有机污染物 BOD 值随时间变化的规律, 所得计算结果与实测值较好地吻合.

(2) 生物降解速率 $K_{\text{苯酚}} > K_{\text{苯胺}} > K_{\text{苯}} > K_{\text{硝基苯}}$. 难降解的有机物硝基苯只有保持在很低的浓度下才能被降解, 浓度越大, 越难降解. 苯胺、苯酚在浓度低于 1mg/L 时, 符合硝基苯的变化规律. 而苯则随浓度增大, 降解速率亦增大.

(3) 有机污染物溶解度小, 则在海水中生物降解速率也小.

参 考 文 献

1 R·S·拉马尔奥著, 严忠琪, 王凤石译. 废水处理概论. 中国建筑工业出版社, 1986: 56
2 美)W·J·莱曼等著, 许志宏等译. 化学性质估算方法手册. 北京: 化学工业出版社, 1991: 354—357
3 金相灿等编. 有机化合物污染化学——有毒有机物污染化学. 北京: 清华大学出版社, 1990: 36

(上接第 40 页)

(2) 由于 PPR 能较好地利用信息, 它无论对高维或低维数据, 正态或非正态, 线性或非线性, 独立或非独立分布数据都能有效处理. 因此它适用于分析和处理环境污染中高维非正态非线性的有关问题.

(3) PPR 分析还同时给出以相对权值表示的因子对预测量的贡献大小. 因此这种方法还可用于优选环境因子, 为环境管理和污染防治决策提供依据.

(4) 由于 PPR 采用审视数据-模拟-预测这样一条探索性数据分析新途径, 建立的模型稳健性和抗干扰性好, 因而预测结果具有较高的精度.

(5) 直接使用已编辑好的 PPR 计算软件

SMART 预测污染物浓度时, 只有 S 、 M 和 M_u 很少几个参数需要指定和反复调试, 因此使用方便.

参 考 文 献

1 吴聿明编著. 环境统计学. 北京: 中国环境科学出版社, 1991: 287—373
2 王煜. 系统工程. 1996, 14(1): 60
3 Friedman J H & Tukey J W. IEEE Trans. Computers C-23, 1974: 881
4 郑祖国. 八一农学院学报. 1993, 16(2): 1
5 刘大秀等. 数理统计与管理. 1995, 14(1): 47
6 徐伟, 朱燕堂. 数理统计与管理. 1994, 13(2): 24
7 Friedman J H & Stuetzle W. J. Amer. Statis. Assoc., 1981, 76: 817
8 Huber P J. Ann. Statistics, 1985, 13: 435
9 卢崇飞等编著. 环境数理统计学应用及程序. 北京: 高等教育出版社, 1990: 75—77

respectively. The COD_{Cr} removal efficiency is more than 80% and COD_{Cr} of the final effluent is decreased to less than 200mg/L which meet the requirement of the National Standard.

Key words: dye intermediate; J-acid waste liquor, Tobias acid waste liquor, chemical oxidation; Fenton reagent.

Projection Pursuit Regressive Model on Concentration Forecast of Pollutant. Li Zuoyong (Dept. of Atmospheric Electronics Eng., Chengdu Institute of Meteorology, Chengdu 610041): *Chin. J. Environ. Sci.*, **18**(4), 1997, pp. 38—40

A projection pursuit regressive model on concentration forecast of pollutant was investigated by analysing environmental monitoring data. Use is being made of the new idea of examination of data-simulation-forecast of projection pursuit regressive techniques. Concentration forecast of water pollutants BOD and DO were carried out, respectively, using this model. The results of the forecasts showed that the qualified rates of fitting and forecasting of the model acquire 100% and 83.3%, respectively.

Key words: environmental pollution, pollutant, forecast, projection pursuit regressive model.

Kinetic Study on the Biodegradation of Organic Pollutants in Coastal Seawater of Dalian.

Tong Zhiyou, Zhou Jiti, Chen Yuchen (Dept. of Environ. Eng., Dalian University of Technology, 116012): *Chin. J. Environ. Sci.*, **18**(4), 1997, pp. 41—44

In this paper, studies were carried out to determine the biological degradation kinetic characteristics of four kinds of organic pollutants in coastal seawater of Dalian. By adding benzene, nitro-benzene, phenol and aniline to natural sea water, respectively, the biochemical oxygen demand (BOD) of the organic compound was tested under different concentration groups of 0.1mg/L, 0.3mg/L, 0.5mg/L, 1mg/L etc. Through data analysis, the kinetic

parameters K and L_0 of the compound were obtained (K represents first-order biodegradation rate constant, L_0 represents ultimate BOD). The mathematic model $y = L_0(1 - 10^{-Kt})$ of the BOD, which was expressed by first-order reaction kinetics, well demonstrated the variation of BOD with time. The results calculated according to the model were similar to the experiment values. Some influence factors on the biodegradation rate of the compound were also analyzed and discussed in this paper. The conclusion was that biodegradation was mainly affected by molecular structure.

Key words biodegradation, benzene, nitro-benzene, aniline, phenol, coastal seawater, biochemical oxygen demand.

Study on Additional Experiment of Chemisorption Catalytic Oxidation for SO₂ from the Smelting Waste Gas. Ning Ping, Sun Peishi, Song Wenbiao (Dept. of Environ. Eng., Kunming University of Science and Technology, Kunming 650093): *Chin. J. Environ. Sci.*, **18**(4), 1997, pp. 45—48

Additional experiments of catalytic oxidation of SO₂ in aqueous solution have been carried out at a smelting plant in a foam tower containing one serve-plate, 16% (wt) H₂SO₄ has been obtained while the absorption efficiency of SO₂ is above 50%. Producing rate of H₂SO₄ is about 1.8% (wt) / h. The best ratio of solution and gases is 5L/Nm³. Addition of Al₂(SO₄)₃ and O₂ in the solution improves the absorption process. The change of SO₂ concentration has small effect on the absorption efficiency.

Key words: sulfur dioxide, catalytic oxidation, foam tower, smelting waste gas.

Global Reaction Rate of NO Formation from Fuel Nitrogen. Fan Yaoguo, Xu Minghou et al. (National Laboratory of Coal Combustion, HUST, Wuhan 430074): *Chin. J. Environ. Sci.*, **18**(4), 1997, pp. 49—52

Based on the detailed mechanism of nitrogen chemistry, the reaction characteristics of fuel