

一维河流水团示踪试验数据的模糊线性回归

康瑞龙¹, 常安定¹, 郭建青²

(1. 长安大学理学院, 陕西 西安 710064; 2. 长安大学环境科学与工程学院, 陕西 西安 710051)

摘要: 将模糊线性回归的概念引入分析河流水团示踪试验数据、确定河流水质参数的计算过程中, 建立计算河流水质参数的模糊线性回归模型, 利用该模型能计算出不同置信水平下水质参数的取值范围。算例验证结果表明, 应用模糊线性回归模型可以有效地确定河流的水质参数, 且计算过程具有一定的稳定性、可靠性和灵活性。模糊线性回归模型是分析河流水团示踪试验数据、确定河流水质参数的一种有效的途径。

关键词: 河流; 水质参数; 模糊线性回归; 环境水力学

中图分类号: P333.9 文献标志码: A 文章编号: 1006-7647(2012)04-0019-04

A fuzzy linear regression model for tracer test data of one-dimensional rivers//KANG Rui-long¹, CHANG An-ding¹, GUO Jian-qing² (1. College of Science, Chang'an University, Xi'an 710064, China; 2. School of Environmental Science & Engineering, Chang'an University, Xi'an 710051, China)

Abstract: In this study, the fuzzy linear regression (FLR) method was applied to estimate the water quality parameters by analyzing the tracer test data of rivers. A FLR model, which can be used to determine the value range of water quality parameters under different confidence levels, was established. The results indicate that this model can be used to determine the water quality parameters efficiently. The model was also found to be stable, reliable, and flexible during the calculation process. Therefore, this model is an effective way to analyze the water mass tracer test data and to determine the water quality parameters of rivers.

Key words: rivers; water quality parameter; fuzzy linear regression; environmental hydraulics

河流水质问题是目前水环境研究工作的热点问题之一, 人们在这方面进行了大量的研究工作^[1-2]。河流水质参数是反映河流综合特征、建立水质模型、预测预报河流水质状态的基础性参数。传统估算河流水质参数的方法主要有理论公式法、经验公式法和水团示踪试验法^[3]。水团示踪试验法原理简单, 易于掌握和应用, 是目前确定河流水质参数最常用也是最可靠的方法之一。除了传统的矩量法和演算法之外, 人们在分析瞬时投放示踪剂条件下的一维河流水团示踪试验数据、确定河流水质参数方面建立了数据分析方法, 其中一些方法以解析推导为基础, 包括直线图解法^[4]、演算优化算法^[5]、抛物线方程近似拟合法^[6]和相关系数极值法^[7]等; 另一些方法则引入智能优化算法, 例如神经网络法^[8]和相关系数极值法^[9]。文献^[10]对不同方法的优缺点进行了较为详细的评述。然而, 由于条件限制以及外界因素的影响, 野外试验观测数据的精度不可能很高,

而是具有一定的模糊性; 另一方面, 河流系统本身也是一种具有随机性、模糊性的复杂系统, 它自身的不确定性使得河流水质参数很难用一个精确的数值来表示, 也就是说, 河流水质参数存在着一定的不确定性。但是, 上述方法的共同特点是将试验中观测到的示踪剂浓度数据作为一组确定性的数, 而没有考虑观测数据本身存在的不确定性, 即随机性或模糊性。鉴于此, 本文根据文献^[11]的思路, 以文献^[4]中的推导过程和结果为基础, 将线性回归模型中的回归系数用正态模糊数来表示, 建立模糊线性回归模型, 利用隶属函数表示水质参数取值对河流系统的适应程度, 并用置信区间的大小来反映参数取值的优劣, 从而为分析一维河流水团示踪试验数据、确定河流水质参数提供一种新的途径。

1 基本原理

在河流流速与横断面面积沿程基本保持不变的

基金项目: 中央高校基本科研业务费专项(CHD2012TD015)

作者简介: 康瑞龙(1984—) 男, 陕西咸阳市人, 硕士研究生, 主要从事最优化理论与方法研究。E-mail: kangrl@126.com

通信作者: 常安定(1964—) 男, 陕西大荔人, 教授, 主要从事水文地质的数学方法研究。E-mail: chdanding@126.com

顺直均匀河段上进行水团示踪试验,即在该河段上游瞬时投放惰性示踪剂,设投放位置为空间坐标原点,水流方向为 x 正方向,投放时刻为零时刻,则描述河水中示踪剂质量浓度随时间变化规律的解析表达式为^[3]

$$c = \frac{M}{A\sqrt{4\pi Dt}} \exp\left[-\frac{(x-ut)^2}{4Dt} - kt\right] \quad (1)$$

式中: c 为投放示踪剂后 t 时刻距投放点距离 x 处的下游观测点的示踪剂质量浓度, g/m^3 ; M 为瞬时投放的示踪剂质量, g ; A 为河流横断面面积, m^2 ; D 为河流纵向离散系数, m^2/min ; u 为河流断面平均流速, m/min ; k 为一级反应速率常数。

由于瞬时投放的示踪剂为惰性示踪剂,故 $k=0$,从而式(1)可简化为

$$c = \frac{M}{A\sqrt{4\pi Dt}} \exp\left[-\frac{(x-ut)^2}{4Dt}\right] \quad (2)$$

一般在基本满足河流流速与横断面面积沿程不变的顺直均匀河段上,当 $x \geq x^*$ 时式(1)具有较高的精度:

$$x^* = \frac{1.8B^2u}{4Hu^*} \quad (3)$$

其中 $u^* = \sqrt{gHJ}$

式中: B 为河宽, m ; H 为水深, m ; J 为试验河段的水力坡降。

对式(2)进行变换^[4],可得

$$G = a + bT \quad (4)$$

其中 $G = t\left(1 + 2t\frac{c'}{c}\right)$ $a = \frac{x^2}{2D}$

$$b = -\frac{u^2}{2D} \quad T = t^2$$

式中: c' 为 c 关于时间 t 的一阶导。

由于河流纵向弥散系数 D ,河流断面平均流速 u 以及投放点与观测点之间的距离 x 都为模型参数,由式(4)可知,因变量 G 和自变量 T 之间存在着线性关系,因此可以建立线性回归模型。然而,由于条件的限制,河流系统模型本身就具有一定的模糊性,加之受到外界环境等因素的影响,所观测数据不可能精度很高,而是有一定的模糊性,这样就使所求得的河流纵向弥散系数以及河流断面平均流速也具有模糊性。因此,采用模糊线性回归模型来确定河流纵向弥散系数以及河流断面平均流速。

2 基于正态模糊数的模糊线性回归模型

类似于经典的线性回归模型,可建立如下模糊线性回归模型:

$$\tilde{y} = \tilde{C}_1x_1 + \tilde{C}_2x_2 + \dots + \tilde{C}_nx_n \quad (5)$$

式中: \tilde{y} 为因变量 y 的拟合值,为模糊数; $x_1, x_2, \dots,$

x_n 为自变量; $\tilde{C}_i (i=1, 2, \dots, n)$ 为各变量的系数,为模糊数。

如图1所示,定义 \tilde{C} 是正态模糊集, \tilde{C} 的隶属函数为

$$\mu_{\tilde{C}} = \exp\left[-\left(\frac{x-\alpha}{\sigma}\right)^2\right] \quad (x \in R) \quad (6)$$

式中: α, σ 为模糊集 \tilde{C} 的2个参数,分别表示 \tilde{C} 的中心和模糊幅度。可将 \tilde{C} 记为 $\tilde{C} = \tilde{C}(\alpha, \sigma)$

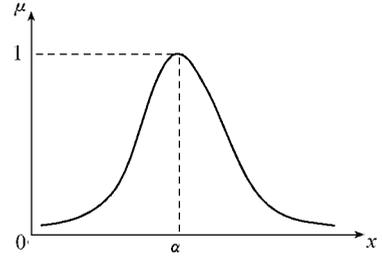


图1 正态模糊集隶属函数

为了使模型(式(5))的拟合度达到最高,在模糊线性回归中必须满足以下2个条件:①各回归系数的模糊幅度之和最小;②在一定的置信水平 $\lambda (0 \leq \lambda \leq 1)$ 下,每个线性回归观测值 y_i 对相应的模糊线性回归拟合值 \tilde{y}_i 的拟合度不小于 λ 。

将式(5)中正态模糊数作为回归系数,即 $\tilde{C}_i = \tilde{C}_i(\alpha_i, \sigma_i)$, \tilde{y}_i 可表示为

$$\tilde{y}_i = \tilde{y}_i \left(\sum_{j=1}^n x_{ij} \alpha_j, \sum_{j=1}^n |x_{ij}| \sigma_j \right) \quad (7)$$

式中: $\alpha_j (j=1, 2, \dots, n)$ 为正态模糊集 \tilde{C}_j 的中心; $\sigma_j (j=1, 2, \dots, n)$ 为正态模糊集 \tilde{C}_j 的模糊幅度。

观测值 y_i 对拟合值 \tilde{y}_i 的拟合度 λ 为

$$\lambda = \exp\left[-\left(y_i - \frac{\sum_{j=1}^n x_{ij} \alpha_j}{\sum_{j=1}^n |x_{ij}| \sigma_j}\right)^2\right] \quad (8)$$

根据条件②有

$$\begin{cases} \sum_{j=1}^n x_{ij} \alpha_j - \sqrt{-\ln \lambda} \sum_{j=1}^n |x_{ij}| \sigma_j \leq y_i \\ \sum_{j=1}^n x_{ij} \alpha_j + \sqrt{-\ln \lambda} \sum_{j=1}^n |x_{ij}| \sigma_j \geq y_i \end{cases} \quad (9)$$

由以上分析可知,模糊线性回归模型可以转化为以下线性规划问题:

$$\begin{aligned} \min N &= \sigma_1 + \sigma_2 + \dots + \sigma_n \\ \text{s.t.} &\begin{cases} \sum_{j=1}^n x_{ij} \alpha_j - \sqrt{-\ln \lambda} \sum_{j=1}^n |x_{ij}| \sigma_j \leq y_i \\ \sum_{j=1}^n x_{ij} \alpha_j + \sqrt{-\ln \lambda} \sum_{j=1}^n |x_{ij}| \sigma_j \geq y_i \end{cases} \end{aligned} \quad (10)$$

式中: N 为模糊幅度之和。

对于上述线性规划问题可用单纯形法或者调用数学软件 Matlab 中的线性规划工具箱进行求解,得到

式(10)的最优解。该最优解是以 α_i, σ_i 为分量的一个向量,即 $x = (\alpha_1, \alpha_2, \dots, \alpha_n; \sigma_1, \sigma_2, \dots, \sigma_n)$,于是便得到了一定置信水平 λ 下模糊线性回归系数 \tilde{C}_i 的值。

3 利用模糊线性回归模型确定河流水质参数

利用中心差分格式对式(4)做如下近似处理^[4]:

$$G_{i+\frac{1}{2}} = a + bT_{i+\frac{1}{2}} \quad (11)$$

其中 $G_{i+\frac{1}{2}} = \frac{1}{2}(t_{i+1} + t_i) \left(1 + 2 \frac{t_{i+1} + t_i}{t_{i+1} - t_i} \frac{c_{i+1} - c_i}{c_{i+1} + c_i} \right)$

$$T_{i+\frac{1}{2}} = \frac{1}{4}(t_{i+1} + t_i)^2$$

根据 $G_{i+\frac{1}{2}}$ 与 $T_{i+\frac{1}{2}}$ 之间的线性关系,可以建立以下模糊线性回归模型:

$$\tilde{y} = \tilde{C}_1 + \tilde{C}_2 t_{i+\frac{1}{2}}^2 \quad (12)$$

其中 $\tilde{C}_1 = \tilde{C}_1(\alpha_1, \sigma_1)$ $\tilde{C}_2 = \tilde{C}_2(\alpha_2, \sigma_2)$

对给出的数据做相应的处理之后,根据上述模糊线性回归模型调用 Matlab 线性规划工具箱进行求解,即可得线性回归系数 \tilde{C}_1, \tilde{C}_2 。根据扩张定理^[12],模糊集合及其运算性质以及 \tilde{C}_1 的隶属函数 $\mu_{\tilde{C}_1}(v) =$

$\exp\left[-\left(\frac{v - \alpha_1}{\sigma_1}\right)^2\right] (v \in R)$,可得河流纵向弥散系数 \tilde{D} 的隶属函数为

$$\mu_{\tilde{D}}(\omega) = \exp\left[-\left(\frac{\omega^2 - \alpha_1}{\sigma_1}\right)^2\right] (\omega \in R) \quad (13)$$

式中: ω 为 \tilde{D} 的支撑集; α_1 为 \tilde{C}_1 的中心; σ_1 为 \tilde{C}_1 的模糊幅度。

由回归系数 \tilde{C}_2 的隶属函数 $\mu_{\tilde{C}_2}(\eta) = \exp\left[-\left(\frac{\eta - \alpha_2}{\sigma_2}\right)^2\right] (\eta \in R)$ 及河流纵向弥散系数 \tilde{D} 的隶属函数可得河流断面平均流速 \tilde{u} 的隶属函数为

$$\mu_{\tilde{u}}(z) = \begin{cases} \sqrt{-2\eta\omega=z} \left[\mu_{\tilde{D}}(\omega) \wedge \mu_{\tilde{C}_2}(\eta) \right] & z \geq 0 \\ 0 & z < 0 \end{cases} \quad (14)$$

式中: \vee 为取大符号; \wedge 为取小符号。

利用上述模糊线性回归模型(式(12))及隶属函数(式(13)和式(14))可计算得出河流水质参数的取值范围。

4 算例验证

数据引自文献[4],本例为一数字算例,表1给出的是利用式(1)计算得到的在某均匀河段上游瞬时投放 10kg 惰性示踪剂后,距投放点下游 500m 处观测到的示踪剂质量浓度随时间的变化过程。该河段的横断面面积为 20m²,河流断面平均流速的真值为 0.5m/min,河流纵向离散系数的真值为 50.0m²/min。

根据模糊线性回归模型(式(12)),分析表1中的水团示踪试验数据,计算河流纵向离散系数和断面平均流速。

表1 不同时刻河水中示踪剂的质量浓度

t/min	c/(mg·L ⁻¹)	t/min	c/(mg·L ⁻¹)
6	0.254	16	0.642
10	0.583	20	0.552
12	0.649	24	0.444
14	0.665	36	0.197

首先,取置信水平 $\lambda = 0.6$,利用 Matlab 求得 $\alpha_1 = 42.2480, \sigma_1 = 0; \alpha_2 = -0.1477, \sigma_2 = 0.0079$,从而有 $\tilde{C}_1 = \tilde{C}_1(42.2480, 0), \tilde{C}_2 = \tilde{C}_2(-0.1477, 0.0079)$,即 \tilde{C}_1 为精确数, \tilde{C}_2 为模糊数。从而可以应用式(13)和式(14)分别计算 \tilde{D} 和 \tilde{u} 的隶属函数。然后,根据实际要求取不同的置信水平 λ ,得到 \tilde{D}, \tilde{u} 在不同置信水平下的取值范围,见表2。

表2 不同置信水平下的水质参数取值范围

λ	$\tilde{D}(\text{m}^2 \cdot \text{min}^{-1})$	$\tilde{u}(\text{m} \cdot \text{min}^{-1})$
0.2	49.3120	0.4758 ~ 0.5090
0.3	49.3120	0.4781 ~ 0.5069
0.4	49.3120	0.4800 ~ 0.5051
0.5	49.3120	0.4816 ~ 0.5035
0.6	49.3120	0.4832 ~ 0.5020
0.7	49.3120	0.4848 ~ 0.5004
0.8	49.3120	0.4864 ~ 0.4988

文献[4]采用直线图解法计算河流纵向离散系数和断面平均流速,结果分别为 51.3m²/min 和 0.49m/min。相比较而言,模糊线性回归模型求解的结果相对误差较小,且其结果为一模糊数,这也就反映了系统本身的模糊性,而且与实际情况应当是相符合的。由表2可以看出,在观测数据模糊性较弱的情况下,模糊线性回归模型计算结果与真值比较接近,而当观测数据模糊性较强时,模糊线性回归模型计算结果为一区间,与真值会有一定的偏离,但其均值与真值比较接近。

另外,当置信水平 λ 较小时,河流断面平均流速 \tilde{u} 取值区间相对较大,而随着置信水平的不断增大, \tilde{u} 的模糊度逐渐降低,其取值区间也逐渐缩小,区间均值也越来越接近真值,这一结果表明该研究方法具有一定的稳定性。同时,该方法还可以根据不同的要求,即根据在试验过程中试验数据可能存在误差的大小和其他因素的影响程度,选取适当的置信水平,由此得到相应的水质参数的可能取值范围。

5 结 语

本文建立了估计河流水质参数的正态模糊线性回归模型,并通过相关算例验证了该方法的有效性。

验证结果表明,该方法可根据不同要求估计出不同置信水平下的河流水质参数,在计算过程中具有一定的稳定性、可靠性和灵活性,是分析河流水团示踪试验数据、确定河流水质参数的一种有效途径。

参考文献:

- [1] 常文婷,王冠韩,龙喜.基于平面二维水质模型的潮汐河流水质反演[J].水资源保护,2010,26(6):5-8.
- [2] 邹海明,蒋良富,李粉茹.2004年淮河流域水质状况和聚类分析[J].水资源保护,2007,23(1):60-62.
- [3] 芮孝芳.水文学原理[M].北京:中国水利水电出版社,2004:370-372.
- [4] 郭建青,温季.示踪实验确定河流纵向弥散系数的直线图解法[J].环境科学,1990,11(2):24-27.
- [5] 顾莉,华祖林,何伟,等.河流污染物纵向弥散系数确定的演算优化法[J].水利学报,2006,38(12):1421-1425.
- [6] 郭建青,李彦,王洪胜,等.确定河流水质参数的抛物方程

近似拟和法[J].水利水电科技进展,2005,25(2):11-14.

- [7] 郭建青,王洪胜,李云峰.确定河流纵向弥散系数的相关系数极值法[J].水科学进展,2000,11(4):387-391.
- [8] 马洪波,崔柏林,常文娟.改进的BP网络模型在确定河流纵向弥散系数中的应用[J].水电能源科学,2010,28(9):19-21.
- [9] 郭建青,李彦,王洪胜,等.粒子群优化算法在确定河流水质参数中的应用[J].水利水电科技进展,2007,27(6):1-5.
- [10] 顾莉,华祖林.天然河流纵向弥散系数确定方法的研究进展[J].水利水电科技进展,2007,27(2):85-89.
- [11] BREE R M, METIN M O, GEORGE F P. Transmissivity and storage coefficient estimation by coupling the Cooper-Jacob method and modified fuzzy last-squares regression[J]. Journal of Hydrology, 2008, 353: 267-274.
- [12] 杨纶标,高英仪.模糊数学原理及应用[M].3版.广州:华南理工大学出版社,2002:162-170.

(收稿日期:2011-12-01 编辑:骆超)

(上接第9页)

- [4] 霍艾迪,康相武,张广军,等.基于MODIS数据的毛乌素沙地土壤水分模型的建立[J].干旱地区农业研究,2010,28(4):19-23.
- [5] 扎西央宗,杨秀海,边巴次仁,等.基于TVDI的西藏地区旱情遥感监测[J].气象科技,2010,38(4):495-499.
- [6] 王文种,张友静.半干旱区旱情监测指数应用分析[J].地球信息科学,2008,10(2):273-278.
- [7] 宋春桥,游松财,刘高焕,等.基于TVDI的藏北地区土壤湿度空间格局[J].地理科学进展,2011,30(5):569-576.
- [8] 康为民,罗宇翔,向红琼,等.贵州卡斯特山区的NDVI-Ts特征及其干旱监测应用研究[J].气象,2010,38(10):78-83.
- [9] 陈晨,刘媛媛,王文种,等.基于温度植被干旱指数的三花间流域旱情监测[J].东北水利水电,2008,26(2):32-34.
- [10] MORAN M S, CLARKE T R, INOUE Y. Estimating crop water deficit using the relation between surface air temperature and

spectral vegetation index[J]. Remote Sensing of Environment, 1994, 49: 246-263.

- [11] NEMANI R R, RUNNING S W. Estimation of regional surface resistance to evapotranspiration from NDVI and thermal IRVHRR data[J]. Journal of Applied Meteorology, 1989, 28(4):276-284.
- [12] GOETZ S J. Multisensor analysis of NDVI, surface temperature and biophysical variables at a mixed grassland site[J]. International Journal of Remote Sensing, 1997, 18(15):71-94.
- [13] 柳钦火,辛景峰,辛晓洲,等.基于地表温度和植被指数的农业干旱遥感监测方法[J].科技导报,2007,25(6):12-17.
- [14] 姚春生.使用MODIS数据反演土壤水分研究[D].北京:中国科学院研究生院,2003.
- [15] 范辽生,姜纪红,盛晖,等.利用温度植被干旱指数(TVDI)方法反演杭州伏旱期土壤水分[J].中国农业气象,2009,30(2):230-234.

(收稿日期:2011-12-09 编辑:熊水斌)

(上接第18页)

- [8] 赵升伟,茅泽育,罗旻,等.等宽明渠交汇水流数值计算[J].河海大学学报:自然科学版,2005,33(5):494-499.
- [9] 刘同宦,郭炜.主支汇流比对交汇区域水流脉动特性影响试验[J].水利水电科技进展,2009,29(3):6-9.
- [10] 冯镜洁,李然,王协康,等.河流交汇分离区特性研究[J].水动力学研究与进展:A辑,2009,24(3):321-325.
- [11] 吴迪,郭维东,刘卓也.复式断面河道“Y”型交汇河口水流水力特性[J].水利水电科技进展,2007,27(3):21-23.

- [12] 郭维东,梁岳,冯亚辉,等.Y型明渠交汇水流分离区的数值分析[J].水利水电科技进展,2007,27(6):50-52.
- [13] 郭维东,王晓刚,曹继文,等.“Y”型汇流口水流水力特性试验研究[J].水电能源科学,2005,23(3):53-56.
- [14] 王福军.计算流体动力学分析:CFD软件原理与应用[M].北京:清华大学出版社,2004:87-91.
- [15] WEBER L J, SCHUMATE E D, MAWER N. Experiments on flow at a 90° open-channel junction[J]. Journal of Hydraulic Engineering, 2001, 127(5):340-350.

(收稿日期:2011-12-26 编辑:骆超)