

# 模糊回归在太湖悬浮物反演中的应用

光洁 , 韦玉春 , 黄家柱 , 孙云峰

( 南京师范大学地理科学学院 , 江苏 南京 210097 )

[ 摘要 ] 模糊回归分析是一种能直接处理不确定性的分析方法 , 突破了传统的“观测值的不确定性就是随机性”这一基本假设。本文将此方法应用于悬浮物的遥感定量反演问题中 , 在太湖水质的采样数据和 Landsat7 ETM 遥感图像的基础上 , 应用模糊回归分析方法 , 针对悬浮物与遥感反射率数据之间的相关关系 , 建立了二者之间的模糊回归方程 , 并将该方法的计算结果与统计学中的最小二乘回归分析结果进行比较。结果表明 : 模糊回归模型比最小二乘回归模型提供了更多的信息量 , 更具优越性 ; 用模糊回归模型进行反演 , 可信度更高 , 且更具推广价值。

[ 关键词 ] 悬浮物 , 遥感 , 模糊回归

[ 中图分类号 ] TP79 [ 文献标识码 ] A [ 文章编号 ] 1001-4616( 2006 )03-0113-04

## Application of Fuzzy Regression in the Retrieval of Suspended Sediment in Taihu Lake

Guang Jie , Wei Yuchun , Huang Jiazhu , Sun Yunfeng

( School of Geographical Science , Nanjing Normal University , Nanjing 210097 , China )

**Abstract** Fuzzy regression analysis is a method directly dealing with uncertainty , which break through the basic hypothesis of “ the uncertainty of observation value is randomness ”. This paper applied this method to remote sensing quantitative analysis of suspended sediment. Aiming at the correlation between suspended sediment concentration and remote sensing reflectance , a fuzzy regression equation was built using the water quality samples and Landsat7 ETM image of Taihu Lake. Comparing the result of this method with least squares regression method in statistics , it is obviously that fuzzy regression model provides more information and is more reliable and worth popularizing.

**Key words** suspended sediment , remote sensing , fuzzy regression

## 0 引言

悬浮物( SS :Suspended Sediment )是水质监测中重要的一项 ,因为它与总初级生产力有关 ,也与重金属的变迁和微量污染物如 PCB 等相关。国内外已有多人研究过悬浮物的遥感反演 ,建立了多种理论与经验模型。最早是线性关系模型<sup>[1]</sup> ,后来又发展了对数模型<sup>[2]</sup>、双对数模型<sup>[3]</sup>、负指数模型<sup>[4]</sup>等。然而 ,这些模型中的参数估计通常采用最小二乘回归分析法 ,通过运用最小二乘法得到变量之间一种具体的表达形式 ,是一种精确的统计表达式 ,这种精确主要体现在反映变量之间关系的参数值的确定。而现实情况中 ,由于遥感数据本身的不确定性<sup>[5-6]</sup>以及测量数据中的不确定性 ,各参数的估计值的真实情况随着系统状态的变化可能是模糊的 ,表现出某种不确定性。因此用传统方法进行悬浮物的分布及浓度反演 ,有明显的局限性。模糊回归的出现为解决这个问题提供了新的方法 ,它以可能性为理论基础 ,以模糊数为研究对象 ,突破传统的“观测值的不确定性就是随机性”这一基本假设 ,能直接处理空间信息和测量数据中的不确定性 ,已

收稿日期 :2005-09-28.

基金项目 :国家环保总局 863 资助项目( 2003AA131060 )。

作者简介 :光洁 ,女 ,1982— ,硕士研究生 ,主要从事环境遥感与空间建模的学习与研究。E-mail :guangjier@163.com

通讯联系人 :韦玉春 ,1965— ,副教授 ,主要从事环境遥感与空间建模的研究与教学。E-mail :weiyuchun@njnu.edu.cn

有多方面的应用实例<sup>[7-10]</sup>. 本文以太湖水水质采样数据和 Landsat7 ETM 遥感图像为数据源 ,应用模糊回归分析方法 ,尝试在悬浮物反演方面作一新的探讨.

1 研究区概况

太湖是我国长江中下游地区著名的五大淡水湖之一 ,它位于长江三角洲的南翼坦荡的太湖平原上 ,介于北纬 30°55'40" ~ 31°32'58" ,东经 119°52'32" ~ 120°36'10" ,属北亚热带向中亚热带过渡的季风气候. 按平均水位 3. 0( 吴淞基面 )时计算 ,其湖泊面积为 2427. 8 km<sup>2</sup> ,除去湖中 51 个岛屿面积 89. 7 km<sup>2</sup> 外 ,实际水面积 2338. 1 km<sup>2</sup> ,居我国五大淡水湖的第三位. 太湖南北长 68. 5 km ,东西平均宽 34 km ,最宽处 56 km ,湖泊平均水深 1. 9 m ,最大水深 2. 6 m ,是一个典型的浅水型湖泊<sup>[11]</sup> ,湖面开阔 ,风浪对湖底表层 3 ~ 5 cm 的底泥具有扰动侵蚀作用 ,在风浪作用下底泥极易再悬浮 ,从而使湖水中悬浮物增加 ,同时也带来了底泥中营养盐的释放 ,与湖泊的内源污染息息相关<sup>[12]</sup>. 近几十年来 ,随着太湖流域经济的迅速发展和人口的快速增长 ,经济发展和环境保护之间的矛盾日益加剧 ,流域河网的有机污染和太湖的富营养化已经非常严重.

2 数据采集与处理

本文以 2001 年 1 月 15 日的 Landsat7 ETM 图像作为遥感信息源 ,选用冬季的图像是为了最大程度的减小叶绿素等因素对于悬浮物反演的干扰<sup>[13]</sup>. 太湖设有常规水质监测点. 1996 年 ~ 2002 年之间 ,湖区设置的样点数为 19 个( 图 1 ). 无锡太湖监测站对这些采样点按月进行采样分析 ,本文使用的水质采样数据即来源于此. 地面实测悬浮物浓度来自 2001 年 1 月 8 ~ 15 日无锡太湖监测站的测量数据 ,其单位为 :mg/L ,与遥感图像在时间上基本保证了同步性.

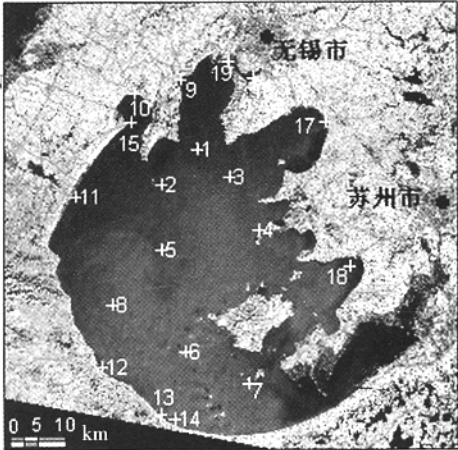


图 1 太湖水面常规监测样点位置

遥感图像的预处理主要包括以下两个步骤 ( 1 )通过差分 GPS 实测的地面坐标对遥感图像进行几何精校正 ,纠正的平均误差为 0. 50 个像素 ,最大误差为 0. 9 个像素 ,最小误差为 0. 01 个像素 ( 2 )辐射校正 :对于辐射校正的算法较多 ,例如 6S 算法 ,PCI、ERDAS 中的 ATCOR 大气校正算法 ,暗像元算法等. 在这些算法中 ,部分参数的选择具有主观性 ,使用的大气标准剖面为美国标准 ,与国内的差异较大. 此外 ,工作地区缺少同步的大气资料. 考虑到这些问题 ,我们认为 ,在没有实测数据校正的情况下 ,使用这些算法进行图像的辐射校正 ,实际上是人为的增加了数据的不确定性 ,在此基础上得到的计算结果也往往不能对比. 在本次工作中 ,考虑到上述问题和数据的对比应用 ,使用头文件和 NASA 网站提供的参数 ,将 ETM 图像各波段的灰度值转换到行星反射率值. 为了降低水面噪音 ,图像使用 5 × 5 的窗口进行了低通滤波. 按照水面采样点坐标从图像中提取各个波段的行星反射率值作为遥感数据 ,并表示为 B<sub>i</sub> ( i = 1 2 3 4 5 7 ). 在水面的常规采样点中 ,样点的设置是按照湖区的差异进行的 ,其中的部分样点 ,例如 11、18 号点靠岸很近 ,另外由于 2 号点靠近一个小岛 ,在提取波谱信息时有异 ,其数据在建立模型中则没有使用. 这样 ,建立模型使用的样点总数为 16 个.

3 模糊回归模型的建立与分析

在一般的回归方法中 ,设观察值与模型估计值之间的偏差为观察误差 ,且认为偏差是一随机变量. 建立估计值的上下界 ,界内估计值的概率表示估计的置信度. 换句话说 ,一般的回归分析就是概率分析. 但是在模糊回归中 ,设观察与估计间的偏差为系统固有的模糊 ,设给定输入的输出为可能值的一个范围 ,即输出可取这些可能值的任意值. 因此 ,模糊回归就本质上来说属于可能性问题. 此外 ,与一般回归分析具有确知系数不同 ,模糊回归分析用模糊函数表示系数. 方程 ( 1 )表示一个典型的模糊线性回归模型.

$$\hat{y} = f(x, \tilde{A}) = \tilde{A}_1x_1 + \tilde{A}_2x_2 + \dots + \tilde{A}_nx_n$$

( 1 )

式中 , $\tilde{A}_i$  是第 i 个模糊系数( 通常为一模糊数 ).

### 3.1 隶属函数定义

模糊回归用于估计可能值的范围,其范围用可能性分布表示,也称为隶属函数<sup>[14]</sup>。隶属函数是对每个估计值给予一特定的隶属值(归属程度)(图2)。本文定义模糊系数为一类对称的三角形函数,如图2所示。

每个系数的隶属函数 $\mu_{\tilde{A}}$ 表示为:

$$\mu_{\tilde{A}}(a_i) = \begin{cases} 1 - \frac{|p_i - a_i|}{c_i} & p_i - c_i \leq a_i \leq p_i + c_i \\ 0 & \text{其他} \end{cases} \quad (2)$$

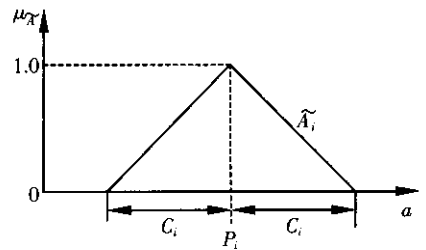


图2 模糊系数的隶属函数

模糊函数 $\tilde{A}$ 是两个分别称作中间值和跨距值的参数 $p$ 和 $c$ 的函数,跨距表征函数的模糊度。图2表示模糊数“近似 $p_i$ ”的隶属函数。

模糊参数 $\tilde{A} = (\tilde{A}_1 \dots \tilde{A}_n)$ 可记作矢量形式 $\tilde{A} = \{p \ c\}$ ,其中 $p = (p_1 \dots p_n)$ 和 $c = (c_1 \dots c_n)$ 。因此,模糊回归的方程输出可表示为:

$$\hat{y} = (p_1 \ c_1)x_1 + (p_2 \ c_2)x_2 + \dots + (p_n \ c_n)x_n \quad (3)$$

上面的方程可用于采样所得的 $m$ 个数据集。输出和输入数据可以是模糊的或非模糊的。

### 3.2 模糊回归的求解

在已有研究的基础上<sup>[15]</sup>,使用探索性数据分析方法寻找悬浮物与遥感数据之间的关系。将悬浮物浓度值(SS)作为因变量,悬浮物归一化指数NDSS =  $(B1 - B3)/(B1 + B3)$ 作为自变量,其中B1、B3表示ETM图像第一、三波段的行星反射率值。将输入数据SS、NDSS视为非模糊数,使用模糊回归方法(FR: Fuzzy Regression)求解方程(1)。

将模糊回归模型变换为线性规划问题确定其解<sup>[14]</sup>(并非是唯一的方法)。对非模糊数据来说,回归模型的目的是确定最优模糊参数 $\tilde{A}$ ,使模糊输出集具有的隶属值大于 $h$ ,如:

$$\mu_{\tilde{Y}}(y_j) \geq h, j = 1 \dots m \quad (4)$$

程度 $h$ 由用户确定,用来衡量模糊线性回归模型对样本的拟合程度,决定着模糊参数。Tanaka和Watada建议 $h$ 值的选择应当根据样本的容量<sup>[16]</sup>:当样本的容量充分大时,可选择 $h = 0$ ,随着样本容量的减小 $h$ 值随之增大。如果 $h$ 增大,输出的模糊度增大,但中心值不会变。考虑到本研究中样本容量不是很大,适当选取 $h = 0.1$ ,既提高预测范围又不至于使预测值的模糊度过大。图3表示模糊输出的隶属函数。式(4)表明模糊输出应处在图3所示的A和B之间。

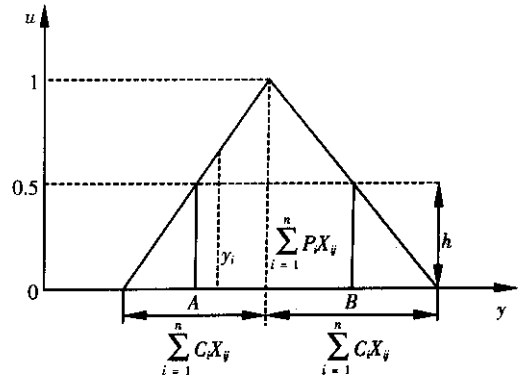


图3 模糊输出函数

回归中,我们力求寻找对所有数据集有最小模糊输出跨距的模糊系数。式(5)表示达到极小的目标函数。

$$O = \min \left\{ \sum_{j=1}^m \sum_{i=1}^n c_i x_{ij} \right\} \quad (5)$$

式(5)给出的目标函数是在两个约束条件下达到的最小值,它们是:

$$y_j \geq \sum_{i=1}^n p_i x_{ij} - (1 - h) \sum_{i=1}^n c_i x_{ij} \quad (6)$$

$$\text{和} \quad y_j \leq \sum_{i=1}^n p_i x_{ij} + (1 - h) \sum_{i=1}^n c_i x_{ij} \quad (7)$$

根据实测的16个点的悬浮物浓度值和ETM图像上提取的NDSS值,我们得到悬浮物的模糊回归方程如下:

$$\tilde{S} = \exp[(7.77 \ 0.51) - (19 \ 0) * \text{NDSS}] \quad (8)$$

其中 $\tilde{S}$ 表示模糊回归预测的悬浮物的浓度值;NDSS =  $(B1 - B3)/(B1 + B3)$ 为悬浮物归一化指数,万方数据

B1、B3 为 ETM 图像第一、三波段的行星反射率值.

4 结果与分析

为了评价模糊回归的反演效果 ,使用最小二乘回归方法( LS :Least Squares )进行对比研究. 最小二乘估计是以随机不确定性为基础的参数估计问题. 在最小二乘估计中 ,观测值和未知参数都作为随机量 ,其最优准则是观测值与其估计量之差( 即残差 )的平方和最小. 而模糊回归则是以模糊不确定性为基础 ,观测值和未知参数都作为模糊数 ,其最优准则是未知参数的模糊幅度( 即未知参数的估计值与其真值之差的绝对值 )之和最小. 尽管最小二乘估计与模糊回归异途同归 ,但二者有着本质的区别. 将模糊回归的预测结果与最小二乘回归的预测结果作一对比 ,见表 1.

表 1 模糊回归( FR )与最小二乘回归( LS )预测结果的对比分析

样点编号	实测 SS 值	LS 预测结果	FR 预测结果( 下限 )	FR 预测结果( 中心值 )	FR 预测结果( 上限 )	LS 预测的相对误差%	FR 中心值预测的相对误差%
1	32	54	32	51	80	68. 75	59. 38
3	244	159	105	166	263	34. 84	31. 97
4	244	239	164	260	411	2. 05	6. 56
5	137	146	96	151	240	6. 57	10. 22
6	244	256	177	280	443	4. 92	14. 75
7	317	217	148	234	371	31. 55	26. 18
8	146	199	134	213	337	36. 30	45. 89
9	24	26	14	22	36	8. 33	8. 33
10	18	27	15	24	38	50. 00	33. 33
12	192	156	103	163	258	18. 75	15. 10
13	150	183	123	194	308	22. 00	29. 33
14	160	210	142	225	356	31. 25	40. 63
15	28	22	12	19	31	21. 43	32. 14
16	26	27	15	24	38	3. 85	7. 69
17	108	71	43	68	108	34. 26	37. 04
19	18	14	7	11	18	22. 22	38. 89
平均相对误差						24. 82	27. 34

从表 1 中可以看出 LR、FR 中心值的平均相对误差为 24. 82% 27. 34% ,预测精度均大于 70% 以上. 但模糊回归不止给出了点估计 ,同时还给出了区间估计 ,即预测值不是一个固定的值 ,而是可以在一个可能的范围内取值 ,这一点在图 4 中同样有所体现. 图 4 是模糊回归的预测曲线和最小二乘法的预测曲线以及原始实测值. 首先 ,从图上来看 ,不论是模糊回归模型还是最小二乘回归模型 ,其预测值与实测值均很接近 ;其次 ,我们可以看到模糊回归的中心曲线对于低浓度的悬浮物的预测效果与最小二乘法回归的效果大致相同 ,但对于高浓度的悬浮物其预测效果更好 ;第三 ,所有实测点包含在模糊回归的上限曲线和下限曲线之间 ,即模糊回归能够精确预测出悬浮物浓度所在的范围. 因此 ,模糊回归模型比最小二乘回归模型提供了更多的信息 ,更具优越性.

将悬浮物的最小二乘法回归模型与模糊回归模型应用于 ETM 图像上 ,得到太湖悬浮物的浓度分布图 ,按照 60 mg/L 分为 6 个等级( 如图 5 ). 可见 ,模糊回归能够比最小二乘法回归更好地拟合出较高浓度的悬浮物值.

5 结论与讨论

本文通过模糊回归这种能直接处理不确定性的分析方法 ,将其应用于悬浮物遥感定量反演问题中 ,利用水质采样数据和 Landsat7 ETM 遥感图像 ,较好地反演出了太湖中悬浮物的浓度值. 同时 ,将该方法的计

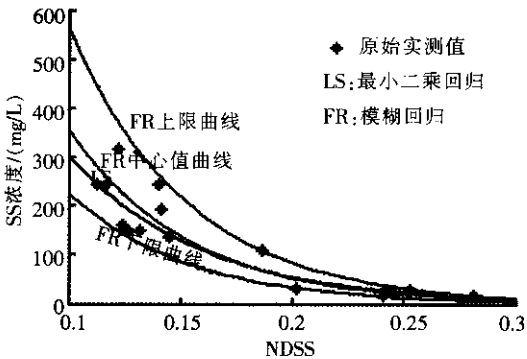


图 4 散点图以及最小二乘回归和模糊回归的拟合曲线



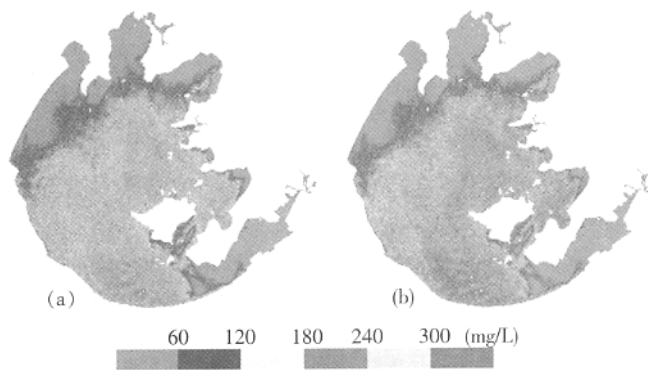


图5 太湖中悬浮物的浓度分布图  
(a) 最小二乘法回归结果;(b) 模糊回归结果

算结果与统计学中的最小二乘回归分析结果进行了比较. 结果表明:模糊回归模型比最小二乘回归模型提供了更多的信息量,不止给出了点估计,同时还给出了区间估计,更具优越性;其次,用模糊回归模型进行反演,能够体现数据中的不确定性和模糊性问题,表明模糊回归是具有弹性的,对于数据具有较好的适应性,第三可信度更高,且更具推广价值.

但同时我们也注意到随着自变量(NDSS)的减小,因变量(SS)的模糊幅度有逐渐增大的缺点,因此需要研究更稳健的模糊回归参数估计方法.同时模糊回归方程中参数的隶属度函数本文采用的是最简单的三角函数,是否有更合适的隶属度函数形式有待进一步的研究.

## [参考文献]

- [1] Williams A N. Sediment Concentration mapping in tidal estuaries[C]//. Third Earth Resources Technology Satellite - 1 Sym, Washington D C, 1973:1347 - 1386.
- [2] Jacques P, Woro H, Jean L M, et al. Remote sensing as a tool for diagnosis of water quality in Indonesian seas[J]. Ocean and Coastal Management, 1995, 27(3):197 - 215.
- [3] 王学军, 马延. 应用遥感技术监测和评价太湖水质状况[J]. 环境科学, 2000, 11(21):65 - 68.
- [4] 李京. 水域悬浮固体含量的遥感定量研究[J]. 环境科学学报, 1986, 6(2):166 - 172.
- [5] 胡圣武, 许辉, 王新洲, 等. 遥感数据的模糊不确定性及其处理方法探讨[J]. 地理与地理信息科学, 2004, 20(4):19 - 21.
- [6] 骆剑承, 周成虎, 杨艳. 具有部分监督的遥感影像模糊聚类方法研究及应用[J]. 遥感技术与应用, 1999, 14(4):37 - 43.
- [7] Berlin W, Neng F T. A new approach to fuzzy regression models with application to business cycle analysis[J]. Fuzzy Sets and Systems, 2002, 130(1):33 - 42.
- [8] Jorge A S, Antonio T G. Estimating a term structure of interest rates for fuzzy financial pricing by using fuzzy regression methods[J]. Fuzzy Sets and Systems, 2003, 139(2):313 - 331.
- [9] 王志良, 邱林, 梁川, 等. 作物需水量与气温关系的模糊回归分析[J]. 华北水利水电学院学报, 2001, 22(4):4 - 6.
- [10] 吴达, 吴柏林. 模糊回归参数估计方法及应用于景气对策信号分析[J]. 系统工程理论与实践, 2001, 21(11):61 - 67.
- [11] 秦伯强, 胡维平, 陈伟民, 等. 太湖水环境演化过程与机理[M]. 北京: 科学出版社, 2004.
- [12] 张运林, 秦伯强, 陈伟民, 等. 太湖水体中悬浮物研究[J]. 长江流域资源与环境, 2004, 5(13):266 - 271.
- [13] 戴昌达, 姜小光, 唐伶俐. 遥感图像应用处理与分析[M]. 北京: 清华大学出版社, 2004.
- [14] Timothy J R 著, 钱同惠, 沈其聪, 等译. 模糊逻辑及其工程应用[M]. 北京: 电子工业出版社, 2001.
- [15] 韦玉春, 黄家柱, 光洁. 太湖悬浮物浓度的遥感监测归一化指数模型[C]//. 第十五届全国遥感技术学术会议, 贵阳, 2005.
- [16] Tanaka H, Watada J. Possibilistic linear systems and their application to the linear regression model[J]. Fuzzy Sets and Systems, 1988, 27(3):275 - 289.

[责任编辑:陆炳新]