

基于环境一号卫星超光谱数据的多元回归克里格模型 反演湖泊总氮浓度的研究

潘邦龙^{1,2}, 易维宁¹, 王先华¹, 秦慧平¹, 王家成¹, 乔延利¹

1. 中国科学院安徽光学精密机械研究所, 安徽 合肥 230031

2. 安徽建筑工业学院环能学院, 安徽 合肥 230601

摘要 水体中总氮含量是表征湖泊水质的主要指标, 利用遥感技术对其动态定量监测可以及时掌握湖泊污染状况。文章以巢湖为例, 利用 HJ1A 卫星 HSI 超光谱遥感数据, 通过分析总氮与叶绿素 a、悬浮物的相关性, 采用回归克里格方法建立总氮浓度的定量反演模型, 实现了对巢湖水体总氮浓度的反演。结果显示, 波段 B72, B79 和 B97 的多元线性组合与总氮浓度的相关系数 R^2 为 0.76, 而利用多元回归克里格模型, 相关系数 R^2 提高到 0.83。通过使用这种对常规回归模型残差优化的方法, 能有效提高反演的精度, 为建立通用的湖泊总氮浓度定量反演模型提供了有益的探索。

关键词 回归克里格; 总氮; 环境一号卫星; 超光谱

中图分类号: TP791; X87 文献标识码: A DOI: 10.3964/j.issn.1000-0593(2011)07-1884-05

引言

总氮等营养盐严重超标是导致湖泊污染和富营养化的关键因素。目前对总氮的监测手段较多, 其中利用遥感技术监测, 具有监测范围广、速度快、成本低和便于进行长期动态监测的优势, 但由于受湖泊水体总氮的遥感机理、反演方法、技术条件等因素的限制, 目前国内外关于对总氮等营养盐的定量遥感反演面临着信息量不足、难以建立有效反演模型的困难。从已有的遥感监测总氮浓度的研究来看, 常用的反演方法为半经验分析法: (1) 根据总氮与叶绿素含量之间密切的相关性, 建立总氮的遥感信息模型。如张穗等^[1]通过对水体叶绿素光谱特征的分析和水体富营养化指标的研究, 选取叶绿素浓度解译方法, 利用总氮与叶绿素的相关特征得出适合河口特征的富营养化评价方法; (2) 利用高光谱数据和同步水面监测数据, 寻找水体遥感反射率与总氮浓度之间的最大相关波段, 然后运用特征波段比值法建立线性、对数函数和指数函数模型^[2,3]。这种用来描述总氮浓度与水体遥感光谱反射率之间关系的回归模型拟合带来的误差和原始数据自身的误差相互作用, 在一定程度上制约了反演的精度^[4]。

因此, 本文根据水体空间分布特性及其遥感的光谱特

性, 利用环境一号卫星超光谱数据和同步水面实测数据, 通过分析总氮与叶绿素 a、悬浮物之间的关系, 提出了一种改进的回归模型方法, 即基于空间信息和光谱信息的多元回归克里格模型, 对巢湖水体的总氮浓度进行反演。该模型通过对回归的残差数据进行克里格插值, 有效降低了回归模型反演的误差, 从而为分析巢湖总氮浓度时空变化特征提供依据。

1 数据与方法

1.1 研究区概况

巢湖地处长江中下游, 位于安徽省中部, 距省会合肥约 15 km, 是安徽省第一大湖泊。全流域面积 9 173 km², 涉及七县二市, 地貌总体呈南北高、中间低; 湖水多年平均水位 8.37 m, 对应的湖泊面积为 789 km², 蓄水量为 40 亿 m³, 岸线长达 155.7 km。有杭埠河、丰乐河、上派河、南淝河、柘皋河等注入, 湖水经裕溪河入长江, 有蓄水、灌溉之利。巢湖以中庙—姥山—齐头咀为界, 分东西两大湖区, 整个湖底由西向东倾斜, 湖水浑浊, 呈黄褐色, 透明度较差。从 2008 年 1 月—2009 年 6 月对巢湖湖区和九条主要环湖河流水质监测数据来看, 巢湖西半湖区水质重度污染, 东半湖区水质轻度污染; 东半湖水体呈中营养状态、西半湖水体呈轻度富

收稿日期: 2010-09-29, 修订日期: 2010-12-14

基金项目: 国家自然科学基金项目(41071232)和安徽省高校省级自然科学研究重点项目(KJ2009A021Z)资助

作者简介: 潘邦龙, 1976 年生, 中国科学院安徽光学精密机械研究所博士研究生 e-mail: panbanglong@163.com

营养状态。主要环湖河流中, 杭埠河、裕溪河水质优, 白石天河、兆河和柘皋河水质良好, 南淝河、十五里河、派河和双桥河水质重度污染^[5]。

1.2 水面采样数据

2009年6月13—14日在巢湖水域进行了同步水质采样实验, 共选取24个样点, 每个样点用标准采样器采集水面下约20 cm处表层水约2 000 mL, 置于10% HCl和去离子水洗净的塑料桶中, 并用GPS记录样点经纬度坐标。取样后的水质指标数据由巢湖市环境监测站化验给出, 主要包括叶绿素a、悬浮物、总氮、透明度等。然后根据各采样点经纬度, 经ArcGIS软件投影转换为样点地理坐标数据, 并生成相应的采样点平面分布图, 如图1。由于在数据分析时发现11号点位数据异常, 去除后剩余的23个点用于建模和测试。

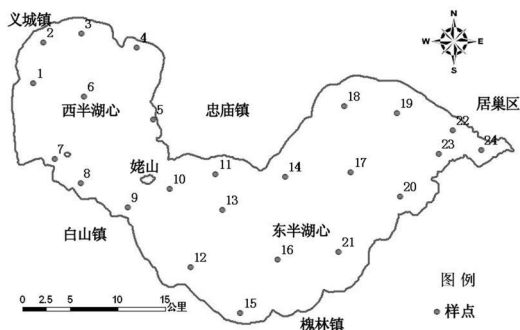


Fig 1 Distribution of water quality sampling points in Chaohu

1.3 遥感影像数据

HJ-1A/B卫星于2008年9月6日成功发射。HJ-1A星搭载了CCD相机和超光谱成像仪(HSI), HJ-1B星搭载了CCD相机和红外相机(IRS)。HSI所形成的图像幅宽大于50 km、地面像元分辨率为100 m, 具有±30°侧视能力和星上定标功能。工作波段为0.45~0.95 μm, 波段数为110~128, 平均光谱分辨率为5 nm^[6]。本文采用的HSI超光谱数据成像日期为2009年6月13日, 当日巢湖上空晴朗无云, 湖面风力很小, 基本处于静风状态。对获取的遥感数据分别进行几何校正和辐射校正。其中几何校正利用1:50 000的地形图, 以高斯克吕格投影6度分带, 运用二次多项式模型校正, 用最邻近法进行重采样, 校正均方根误差小于一个像素。辐射校正采用2009年HJ-1A星HSI相机绝对辐射定标系数对数据进行绝对辐射定标, 再利用6S对部分波段数据进行大气校正, 以获得水体遥感反射率。

1.4 分析方法

克里格法是一种最佳局部估计方法, 它以最小的估计方差给出区域平均值的无偏线性估计量。本文首先通过回归模型初步反演研究区总氮浓度的空间分布, 然后利用克里格法对总氮浓度回归模型的残差值进行修正。具体的克里格原理如下:

设位于点 x_0 处的真实值为 $Z(x_0)$, 它的预测点 $Z_1(x_0)$ 是周围已知点的观测值 $Z(x_i)$ 的线性组合, 即

$$Z_1(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (1)$$

求解的原则是在保证这个预测值 $Z_1(x_0)$ 无偏性, 且在估算方差最小的前提下, 求出 n 个权系数 λ_i 。由于 $Z_1(x_0)$ 的数学期望是未知的, 若要使 $Z_1(x_0)$ 是 $Z(x_0)$ 的无偏估计量, 即要求

$$E\{Z_1(x_0) - Z(x_0)\} = 0 \quad (2)$$

寻求优化估算变成下列在约束条件下求极值的问题

$$\begin{aligned} \min \sigma_E^2 &= E\{[Z_1(x_0) - Z(x_0)]^2\} \\ &= 2 \sum_{i=1}^n \lambda_i \gamma(h_{i0}) - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(h_{ij}) \end{aligned} \quad (3)$$

约束条件

$$\sum_{i=1}^n \lambda_i = 1 \quad (4)$$

式中: σ_E^2 为估计方差; E 为数学期望; $\gamma(h_{i0})$ 和 $\gamma(h_{ij})$ 分别表示预测点 x_0 与观测点 x_i 之间以及不同观测点 x_i 与 x_j 之间的变差函数值, 它们均可由区域化变量空间结构分析获得。通过求解上述最优化问题, 就可求出普通克里格权系数和方差。

回归克里格法是回归模型和克里格相结合的统计方法, 它通过目标变量与辅助变量之间的回归和回归残差的克里格技术来预测空间的特征。设预测变量为 $Z(x_0)$, $Z(x_1)$, ..., $Z(x_n)$, x_i 为观测点, n 为观测点的数量, 则预测点 x_i 的值 $Z(x_i)$ 可以用下式进行预测

$$Z(x_i) = Z_1(x_i) + \varepsilon(x_i) \quad (5)$$

式中, 趋势项 $Z_1(x_i)$ 通常用回归模型进行拟合, 残差项 $\varepsilon(x_i)$ 利用克里格法进行预测。可分为两步, 趋势项计算和残差项计算。本文在趋势项计算中, 回归克里格的基础数据为HJ-1A卫星波段组合数据, 结合采样点的TN浓度建立多元线性回归方程, 并利用matlab软件计算趋势项预测结果。残差项计算中, 将各采样点的TN浓度观测值减去多元线性回归方法趋势项预测数据得到残差值, 在ArcGIS软件中对残差进行普通克里格预测; 最后, 各自的趋势项数据与残差预测数据相加得到全区域TN浓度的预测结果^[7,8]。

2 结果与分析

2.1 总氮浓度与叶绿素a、悬浮物浓度间的相关性分析

通过对采样点总氮与叶绿素a、悬浮物浓度的数据分析可知, 三者之间具有较好的相关性, 其中总氮浓度与叶绿素a的相关系数 R^2 为0.68, $p < 0.001$, 总氮浓度与悬浮物浓度的相关系数 R^2 为0.50, $p < 0.001$, 如图2所示。结果表明总氮并没有因研究区内浮游植物大量消耗吸收而降低, 这部分由于雨水的补给带来较高的无机氮盐, 随径流入湖, 使得总氮含量升高, 即补充量大于消耗量所致^[9]。

2.2 总氮样点浓度与水质参数变量的回归模型

由于总氮浓度与叶绿素a、悬浮物浓度间具有较好的相关性, 并且通过已有的巢湖水体光谱特性研究发现^[6,7,10,11], 水体光谱反射率与叶绿素a的最大相关性位于680和706 nm附近, 与悬浮物的最大相关系数位于816 nm附近。基于

此, 本文选取总氮浓度与超光谱可见光波段 B79 和 B72、近红外波段 B97 三波段组合进行相关性分析, 结果显示 B79/B72 和 B97 的组合, 相关系数最高, R^2 为 0.76, $p < 0.001$ 。故本文选取最佳组合建立巢湖总氮浓度反演的回归模型, 并

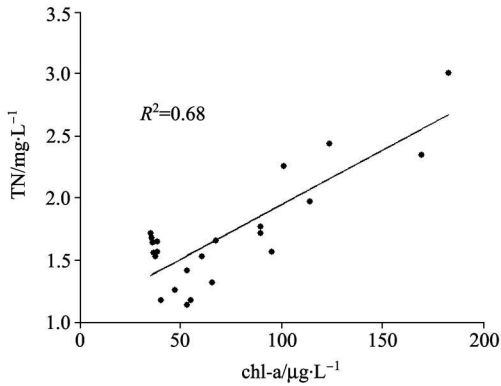


Fig 2 Analysis of the correlation between the concentration of total nitrogen and chlorophyll a, suspended solids

将 23 个样点分为两组, 其中 16 个用于建模, 7 个用于模型检验。为便于比对, 本文分别建立了 B79/B72 一次、二次回归模型, B79/B72 和 B97 组合的多元线性回归模型, 结果如表 1 所示。

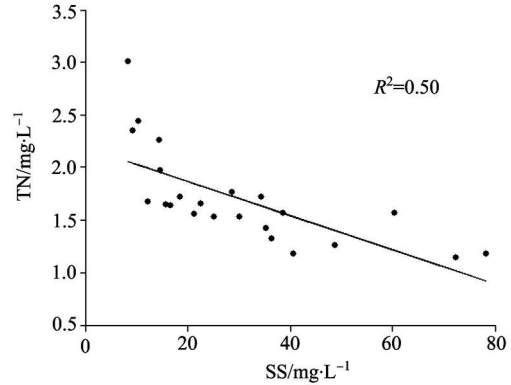


Table 1 Regression inversion model of total nitrogen concentration

模型	观测点	R^2	Adjusted R^2	回归方程	F	Sig F	MS	RM SE
线性	23	0.70	0.69	$1.162 \times B79/B72 + 0.505$	60.18	0.00	0.055	0.235
二次多项式	23	0.73	0.72	$0.886 \times (B79/B72)^2 - 1.042 \times B79/B72 + 1.741$	36.1	0.00	0.039	0.189
多元线性	23	0.76	0.76	$0.295 \times B79/B72 + 78.842 \times B97 + 1.037$	78.22	0.00	0.025	0.159

其中 B72、B79 和 B97 为遥感相应波段的水体遥感反射率, RMSE 为均方根误差, MS 为平均标准差。

2.3 回归残差的普通克里金分析

从表 1 可知, 多元线性回归模型的相关度高、均方根误差小, 可以作为克里格的回归模型。通过比较采样点总氮浓度模型的反演值和观测值, 获得回归残差值。为了利用有限的采样点的回归残差数据获取整个研究区总氮反演误差的空间分布, 采用理论模型对各样点回归残差数据应用变异函数曲线进行拟合, 其中最常用的克里格模型包括普通克里格和简单克里格, 变异函数模型包括球状模型 (Spherical)、指数模型 (Exponential) 及高斯模型 (Gaussian) 等, 模拟的结果参数如表 2 所示。

利用普通克里格方法, 分别运用上述 3 种变异函数模型进行插值, 得出整个研究区总氮浓度反演误差的空间分布。并根据插值结果选择一组最优结果作为残差项分析结果, 参与预测点的浓度计算, 从而获取整个研究区总氮浓度的空间分布情况。

为了评估插值结果, 在进行普通克里格插值计算之后, 利用已有样点的残差数据和插值结果进行交叉校验, 交叉检验采用的评价指标包括平均误差、平均标准差、均方差和标准均方差, 以反映插值结果的无偏性、准确性和可靠性。插值分析结果见表 3。从表 3 可以看出, 3 种模型虽结果相近, 但指数模型预测的均方差、平均标准差和标准均方差均最小, 因此预测效果相对最好^[8, 13]。

Table 2 Parameter values of different Kriging model

模型	变异模型	块金	基台值	变程
普通克里格	Spherical	0.0207	0.0016	1.7435
	Exponential	0.0223	0.0016	1.8375
	Gaussian	0.0223	0.00002	1.7435
简单克里格	Spherical	0	0.0209	1.697
	Exponential	0	0.0209	1.0136
	Gaussian	0.00002	0.0209	1.3223

从表 2 可以看出, 普通克里格模型较简单克里格的变程大, 块金高, 基台值低, 参数估值效果要优。对于三种不同的变异拟合函数模型, 结果基本相同。虽然在个别样点的拟合结果中基台值和块金值有所不同, 但是从基台值和变程总体来看, 拟合结果基本一致^[12]。

2.4 插值结果

Table 3 Interpolation analysis of ordinary Kriging model

模型	变异模型	回归函数	平均误差	均方差	平均标准差	标准均方差
普通克里格	Spherical	$-0.031x - 0.026 - 0.0133$	0.1564	0.1636	0.966	0.966
	Exponential	$-0.060x - 0.001$	0.0157	0.1565	0.996	0.996
	Gaussian	$-0.033x - 0.026 - 0.0133$	0.1568	0.1637	0.968	0.968

最终运用 ArcGIS 对多元线性回归的趋势项和普通克里格插值的残差项进行叠加分析, 获得超光谱遥感的总氮浓度反演的空间分布图, 如图 3 所示。

从反演结果来看, 巢湖西半湖北部及沿岸延伸至忠庙南部, 总氮浓度较高。从实际现场调查来看, 该区域已出现部分浮游藻类, 由于丰富的营养盐的不断聚集, 藻类有扩大的趋势。西半湖南部和东半湖总氮浓度较低, 且分布较均一。

对比同步观测分析发现, 由多元回归克里格模型提取的巢湖湖区总氮分布图与实际情况基本吻合, 能有效地表征湖区的实际情况。

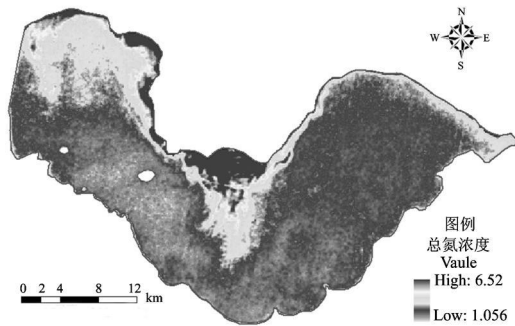


Fig 3 Spatial distribution of inverting the total nitrogen concentration by regression kriging

2.5 反演精度的准确性检验

通过对比总氮实测值与预测值, 7 个检验样点和 16 个建模样点基本上都在 95% 的置信区间内(如图 4 所示), 并且从表 4 来看, 回归克里格与多元线性回归模型相比, 相关系数从 0.76 提高到 0.83, RMSE 从 0.159 降到 0.137, 反演精度有了明显的改进, 这表明多元回归克里格预测模型能够更准确地反映研究区总氮浓度的空间分布特征。

3 结论与讨论

本文利用 HJ1A 卫星 HSI 超光谱遥感数据, 通过对总氮与叶绿素 a、悬浮物相关性分析, 确定总氮反演的最佳波段, 然后利用回归克里格方法建立总氮浓度的定量反演模

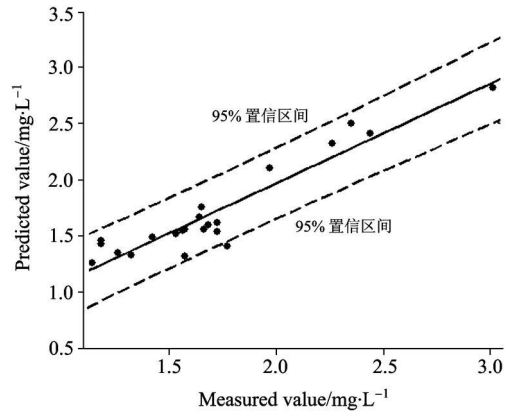


Fig 4 Comparison between nitrogen measured and predicted values

Table 4 Analysis of multiple linear regression model and regression kriging

参数	R^2	F	Sig F	MS	RMSE
多元线性回归	0.76	78.22	0.000	0.0250	0.159
回归克里格	0.83	162.07	0.000	0.0206	0.137

型, 并对各种回归克里格组合模型进行了误差分析与检验。实验结果表明, 多元线性回归与普通克里格组合, 优化了常规回归模型的残差, 有效提高了反演的精度, 为建立湖泊总氮浓度定量反演模型进行了探索。然而由于本文的研究是以总氮与叶绿素 a、悬浮物的相关性为基础, 没有对总氮的表层水体的辐射光谱特性进行研究, 缺乏一定的物理意义, 并且由于不同季节的湖泊水体, 总氮与叶绿素 a、悬浮物的相关性变化较大, 模型的推广有待进一步研究^[14, 15]。

References

[1] ZHANG Sui, HE Bao-yin (张 穗, 何报寅). Journal of Yangtze River Scientific Research Institute(长江科学院院报), 2004, 21(3): 29.

[2] LI Xiao-bin, CHEN Chu-qun, SHI Ping, et al(李小斌, 陈楚群, 施 平, 等). Acta Scientiae Circumstantiae(环境科学学报), 2007, 27(2): 315.

[3] LEI Kun, ZHENG Bing-hui, WANG Qiao(雷 坤, 郑丙辉, 王 桥). Acta Scientiae Circumstantiae(环境科学学报), 2004, 24(3): 376.

[4] WANG Jian-ping, CHENG Sheng-tong, JIA Hai-feng, et al(王建平, 程声通, 贾海峰, 等). Environmental Science(环境科学), 2003, 24(2): 73.

[5] LI Su-ju, WU Qian, WANG Xue-jun, et al(李素菊, 吴 倩, 王学军, 等). Journal of Lake Sciences(湖泊科学), 2002, 14(3): 230.

[6] WANG Qiao, YANG Yu, WU Chuan-qing(王 桥, 杨 煜, 吴传庆). Spacecraft Engineering(航天器工程), 2009, 11(6): 135.

[7] Park No-Wook, Chi Kwang-Hoon, Kwon Byung-Doo. Geosciences Journal, 2003, 7(4): 337.

[8] Ludovisi Alessandro, Minozzo Marco, Pandolfi Piera, et al. Ecological Modelling, 2005, 181: 253.

[9] GONG Caifan, YIN Qiu, KUANG Ding-bo(巩彩兰, 尹 球, 匡定波). Journal of Remote Sensing(遥感学报), 2006, 10(6): 913.

[10] Wu Chunfa, Wu Jiaping, Qi Jiao. International Journal of Remote Sensing, 2010, 31(9): 2315.

[11] DU Wei-jing, LI Shu-min, LI Hong, et al(杜为静, 李淑敏, 李 红, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2010, 30(3): 757.

[12] Rivero R G, Grunwald S, Bruland G L. Geodema, 2007, 140: 435.

[13] Thiemann S, Kaufmann H. Remote Sensing of Environment, 2000, 73(2): 228.

[14] Baban Serwan M J. International Journal of Remote Sensing, 1993, 14(7): 1257.

[15] Atkinson P M, Curler M E J, Lewis H. International Journal of Remote Sensing, 1997, 4: 928.

Inversion of the Lake Total Nitrogen Concentration by Multiple Regression Kriging Model Based on Hyperspectral Data of HJ-1A

PAN Bang-long^{1,2}, YI Wei-ning¹, WANG Xia-hua¹, QIN Hui-ping¹, WANG Jia-cheng¹, QIAO Yan-li¹

1. Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei 230031, China

2. Environment Engineering Department, Anhui University of Architecture, Hefei 230601, China

Abstract The content of total nitrogen in the waters is an important index to measure lake water quality, and the technique of remote sensing plays a large role in quantitatively monitoring the dynamic change and timely grasping the status of lake pollution. Taking Chaohu as an example, quantitative inversion models of total nitrogen were established by multivariable regression Kriging under analyzing of an correlation between total nitrogen and chlorophyll-a or suspended solids by HIS hyperspectral remote sensing data of HJ-1A satellite. The result shows that the correlation of 0.76 was discovered between total nitrogen and the multiple combination with band 72, band 79 and band 97, while the correlation could be increased to 0.83 by applying combined model of multiple linear regression and ordinary Kriging. The optimization of the residuals of the conventional regression model can improve the accuracy of the inversion effectively. These results also provide useful exploration for further establishing a common model of quantitative inversion of lake total nitrogen concentration.

Keywords Regression kriging; Total nitrogen; HJ-1A satellite; Hyperspectral

(Received Sep. 29, 2010; accepted Dec. 14, 2010)