

基于 landsat 卫星影像的水库水体总磷质量浓度反演研究

王云霞¹, 杨国范¹, 林茂森¹, 杨舒婷²

(1. 沈阳农业大学, 沈阳 110866; 2. 河海大学 水利水电学院, 南京 210098)

摘要:以清河水库为研究区域,利用 Landsat8 卫星 OLI 数据及实地观测数据,建立了适于清河水库水体总磷质量浓度的最小二乘支持向量机(LS-SVM)遥感反演模型,并对清河水库总磷量进行了反演分析。结果表明,LS-SVM 模型的平均相对误差为 6.06%,相比于单波段线性回归反演模型及波段组合线性回归模型,平均误差分别降低了 20.77%、12.53%,显著提高了清河水库水体总磷质量浓度反演的模型精度,达到遥感反演预测精度要求;利用 LS-SVM 反演模型对清河水库总磷量反演显示,水体中总磷质量浓度主要集中在 0.04~0.08 mg/L,水库水体总磷质量浓度总体偏高。

关键词:总磷;遥感;模型;清河水库

中图分类号:X87

文献标志码:A

doi:10.13522/j.cnki.ggps.2017.04.019

王云霞,杨国范,林茂森,等. 基于 landsat 卫星影像的水库水体总磷质量浓度反演研究[J]. 灌溉排水学报,2017,36(4):105-109.

0 引言

磷是生物体组成的必需元素,同时也是引发水体富营养化的关键因子^[1],因此,自然水体中总磷的动态监测对水环境保护和水体健康具有重要意义。传统总磷的监测方式费力、耗时、不经济,且具有一定的不合理性,不能够对整个水体进行全面的监测,存在监测时间可能受到天气等外界因素的影响,不能及时获取数据^[2-3]等问题。遥感反演总磷的监测方法可以进行实时监测,改进传统总磷采样的缺点,是今后水质监测发展的方向^[4-5]。王学军^[6]利用遥感信息和有限的实地监测数据建立了太湖水质参数模型,通过单波段、多波段因子组合以及主成分分析等手段较好地反映水质的空间分布特征。刘征等^[7]通过计算单波段和一阶微分水体光谱反射率与总磷的相关系数,确定了总磷的敏感波段,建立了总磷的遥感定量模型。王林等^[8]综合利用实测数据及环境一号卫星 CCD 数据对大洋河口海域水体的营养盐进行了反演研究。王丽艳^[9]利用遥感卫星反演总磷质量浓度并进行湖泊富营养化评价,对湖泊总磷质量浓度分布及富营养化程度进行了全面监测。Isenstein 等^[10]以 Landsat ETM+ 数据为遥感数据源,对尚普兰湖进行研究,利用多元线性回归方程建立了磷营养物质的反演方程。虽然目前许多国内外学者对总磷的遥感监测模型进行了研究^[11-13],但是仍有一定的局限性,所用模型具有地域的限制,且不同类型的水体,所用模型也不尽相同^[14-15]。大部分研究针对的是太湖、大洋河等大型湖泊河流^[16-17],针对北方内陆的小面积水体(水库等)的研究还较少。

清河水库作为辽河流域重要的大型水利枢纽,近年来,水库水体富营养化日益严重,不仅威胁水库的供水安全,也制约水库功能的正常运转。利用 Landsat8 卫星的 OLI 数据对清河水库水体总磷量进行反演研究,分析清河水库水体总磷的光谱响应特征,建立最小二乘支持向量机反演模型,提高水库的反演精度,可以对清河水库进行周期性的动态监测,为水库水质监测提供新的途径,并为建立水质生态修复方案提供依据和参考。

1 材料与方法

1.1 研究区概况

清河水库(42°31'N, 124°14'E)位于辽河中游左侧一级支流——清河干流上,水库坝址坐落在铁岭市清

收稿日期:2015-03-04

基金项目:辽宁省水利厅项目(2016JSFW-183)

作者简介:王云霞(1990-),女。硕士研究生,主要从事数字水利研究。E-mail: 1532327589@qq.com

通信作者:杨国范(1978-),男。研究员,主要从事数字水利研究。E-mail: 81041678@163.com

河区。控制流域面积为2 376 km²,设计总库容为9.71 亿m³。清河水库作为清河流域的控制性工程,是一座具有多年调节功能的大(Ⅱ)型水利枢纽,承担辽河中下游灌溉和清河区防洪、工业以及城市供水等综合任务。

1.2 影像数据的获取及预处理

应用 Landsat8 卫星(重访周期为 16 d)搭载的陆地成像仪(OLI)的 B3、B4、B5 波段,选择 2015 年 6 月 23 日的遥感数据(云量显示为 0.08%,满足数据对云量的要求),利用 ENVI 软件对 OLI 数据进行辐射定标、几何裁剪、大气校正等预处理,得到清河水库卫星影像(如图 1、图 2 所示),空间分辨率为 30 m。

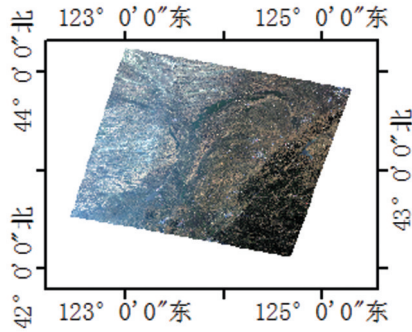


图1 清河水库卫星影像数据辐射定标结果

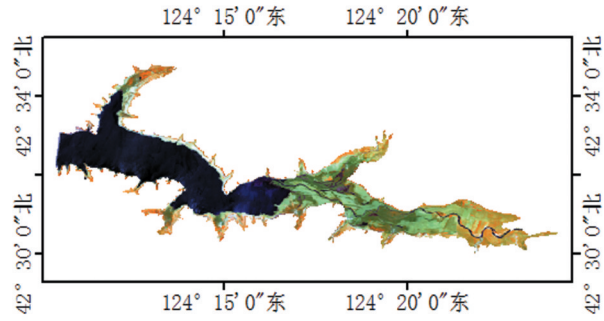


图2 清河水库卫星影像数据大气校正结果

1.3 实测数据的获取与处理

在清河水库选取 25 个采样点,采用网格法将采样点均匀分布覆盖整个库区,于 2015 年 6 月 23 日上午,对水库进行水样采集。每个样点采集 500 mL 水样,并使用手持 GPS 仪记录采样点的经纬度。在实验室应用分光光度计测定水样总磷质量浓度,计算式为:

$$c = \frac{m}{v}, \quad (1)$$

式中: c 为水样总磷浓度(mg/L); m 为水样含磷量(mg); v 为水样体积(mL)

1.4 水库水体总磷质量浓度反演模型的构建

以 Matlab 为实现平台,使用最小二乘支持向量机算法对清河水库的总磷数据进行建模反演,并选用 RBF 函数作为核函数。支持向量机由 AT & Bell 实验室的工作人员提出,其理论基础是统计学理论,采用核函数思想,把非线性空间的问题转换到线性空间,减小复杂度,能够从理论上得到全局的最优解。最小二乘支持向量机的核心思想就是用等式约束替换标准支持向量机中的不等式约束,从而降低求解的难度,具有简单性和有效性。

最小二乘支持向量机的问题表示为:

$$\min_{w,b,c} J(w,e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^l e_i^2, \quad (2)$$

约束条件为:

$$y_i = w^T \varphi(x_i) + b + e_i; \quad i=1, 2, \dots, l, \quad (3)$$

式中: γ 为正则化参数; b 为偏差量。

$$\text{其对偶问题的 Lagrange 多项式为:} \quad L(w,b,e,a) = J(w,e) - \sum_{i=1}^l a_i [w^T \varphi(x_i) + b + e_i - y_i], \quad (4)$$

式中: a_i 为 Lagrange 乘子。最优解的条件为:

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \rightarrow w \sum_{i=1}^l a_i \varphi(x_i) \\ \frac{\partial L}{\partial b} = 0 \rightarrow -\sum_{i=1}^l a_i = 0 \\ \frac{\partial L}{\partial e_i} = 0 \rightarrow a_i = \gamma e_i; i=1, 2, \dots, l \\ \frac{\partial L}{\partial a_i} = 0 \rightarrow w^T \varphi(x_i) + b + e_i - y_i = 0; i=1, 2, \dots, l \end{cases}, \quad (5)$$

消去 e 和 w ,再利用 Mercer 条件,则可得到方程组:

$$\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & \mathbf{M} + \gamma^{-1} \mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{Y} \end{bmatrix}, \quad (6)$$

式中: $\mathbf{1} = [1, \dots, 1]^T$; $\mathbf{Y} = [y_1, \dots, y_N]^T$; $\mathbf{a} = [a_1, \dots, a_N]^T$; $\mathbf{M}$ 为一个方阵,其第 i 行 j 列的元素为 $M_{ij} = \phi(x_i)^T \phi(x_j) = M(x_i, x_j)$;

$M(x,y)$ 为核函数。用最小二乘法求出 a 和 b , 由此得到预测输出:

$$y(x)=\sum_{i=1}^l\alpha_i\varphi(x)^T\varphi(x)^T\varphi(x_i)+b=\sum_{i=1}^l\alpha_iM(x,x_i)+b, \tag{7}$$

式中: $y(x)$ 为预测输出值; a_i 为拉格朗日乘子; $M(x,x_i)$ 为核函数。

2 结果与分析

2.1 水体总磷量模型精度分析

利用最小二乘支持向量机模型对清河水库总磷量进行反演,以B4波段与B5波段的比值为自变量,清河水库总磷质量浓度的实测值为因变量,选取25组数据中的17组数据为建模数据,8组数据为检验数据。最小二乘支持向量机模型所反演的总磷的反演值及其与实测值的相对误差如表1所示。由表1可知,使用最小二乘支持向量机所建立的清河水库总磷量反演模型反演的最小误差值为0.01%,最大误差值为25.07%,平均误差值为6.06%,满足反演模型对误差的要求。

表1 最小二乘支持向量机模型的反演值与实测值比较

最小二乘支持向量机 模型总磷质量浓度预测值	总磷质量浓度 实测值/(mg·L ⁻¹)	相对误差/%	最小二乘支持向量机 模型总磷质量浓度预测值	总磷质量浓度 实测值/(mg·L ⁻¹)	相对误差/%
0.040 4	0.040	1.01	0.068 9	0.070	1.59
0.042 3	0.040	5.65	0.071 8	0.070	2.57
0.043 6	0.050	12.74	0.077 8	0.080	2.76
0.048 7	0.050	2.69	0.078 3	0.080	2.13
0.053 4	0.050	6.83	0.084 9	0.080	6.17
0.053 6	0.050	7.20	0.089 5	0.080	11.89
0.062 5	0.050	25.07	0.089 7	0.080	12.10
0.062 9	0.060	4.84	0.092 5	0.110	15.95
0.064 8	0.060	7.93	0.121 3	0.120	1.08
0.065 3	0.060	8.78	0.126 9	0.130	2.35
0.065 6	0.070	6.26	0.129 4	0.130	0.48
0.068 8	0.070	1.76	0.150 0	0.150	0.01
0.068 8	0.070	1.70			

2.2 清河水库总磷质量浓度反演结果

应用最小二乘支持向量机模型反演清河水库水体总磷量,结果如图3所示(灰色为未反演区域)。由图3可知,水体中总磷质量浓度主要集中在0.04~0.08 mg/L,入库口处总磷量最高,主要集中在0.08~0.12 mg/L,但最高也达到了0.20 mg/L左右;库中较少,主要集中在0.04~0.08 mg/L;坝前处总磷量最少。清河水库总磷量总体偏高,可能的原因有农村生活的污染,包括农药、化肥污染,还有可能是分散禽畜养殖的污染等。应加强对清河水库附近的水土保持和农田径流的控制,尽可能将附近的禽畜集中养殖。

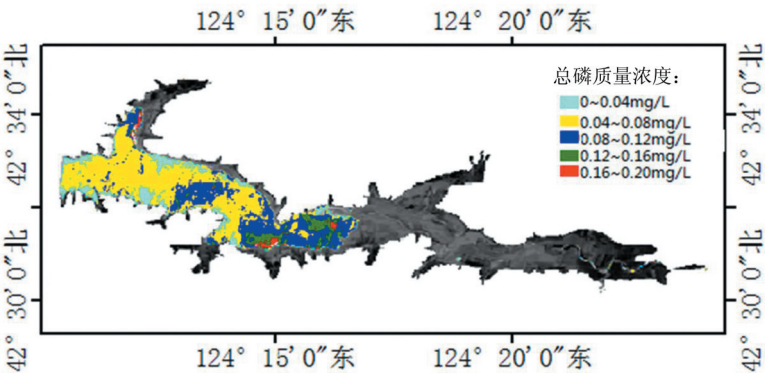


图3 清河水库总磷质量浓度分布

2.3 清河水库遥感总磷量监测模型的适用性分析

现有对水体总磷质量浓度的反演研究主要有单波段法和波段组合法^[18-19]。针对清河水库总磷质量浓度分别建立了单波段模型和波段组合模型,但结果显示均达不到反演的精度要求,因此建立最小二乘支持向量机模型反演清河水库总磷量,提高反演精度。研究表明水体总磷的敏感波段位于873 nm附近,但水体总磷的敏感波段存

在不稳定性^[7]。针对清河水库建立了单波段反演模型。首先利用 spss 软件对清河水库单波段的地表反射率的值与清河水库采样点的实测值进行 pearson 相关性分析,分析发现,数据中与总磷质量浓度相关性较高的敏感波段分别为 B3、B4、B5 波段,虽然与黄壁庄水库敏感波段不完全相同,但是由于水体水质不同等原因,会存在一定的差异,敏感波段的分布大体是相近的^[20]。分别以 B3、B4、B5 波段为自变量,建立单波段回归模型,见表 2。由表 2 可知,B3、B4、B5 波段相对应的决定系数 R^2 分别为 0.291、0.402、0.164。对单波段模型进行误差分析,其中以单波段 B4 所建模型相关性最好,最大相对误差为 73.24%,最小相对误差为 0.59%,相对误差平均值为 26.83%,较最小二乘支持向量机模型的反演相对误差高 20.77%。以单波段为自变量所建立的模型精度过低,不能对清河水库总磷量进行反演研究。

表 2 单波段回归预测模型			表 3 多波段回归预测模型		
参数波段	单波段回归预测模型	决定系数 R^2	参数波段	单波段回归预测模型	决定系数 R^2
B3	$y = 1\,381.9x + 543.3$	0.291	B3+B4	$y = 331\,6.6x + 978.5$	0.351
B4	$y = 1\,934.8x + 435.2$	0.402	B3+B4+B5	$y = 446\,5x + 142\,8$	0.284
B5	$y = 1\,148.4x + 449.48$	0.164	B4/B5	$y = 1.222\,7x + 0.993\,3$	0.604

王丽艳等^[9]在对呼伦湖水体总磷进行反演时,采用了 MODIS 数据,并使用波段组合建立了模型,发现波段组合所建立模型相关性优于单波段模型。针对清河水库建立三 3 类波段组合模型,S1=Band(a)+Band(b)、S2=Band(a)+Band(b)+Band(c)、S3=Band(a)/Band(b),S1 组合中的 B3+B4 多项式的决定系数 R^2 最高,为 0.351。S2 组合中的 B3+B4+B5 中 R^2 最高,为 0.284。S3 组合中 B4/B5 的决定系数 R^2 最高为 0.604,结果如表 3 所示。Williams 等^[21]研究认为决定系数大小可以评价模型好坏,即 $0.5 \leq R^2 \leq 0.65$ 为较差模型, R^2 在 0.66~0.81 之间为一般模型, R^2 在 0.82~0.9 之间为较好模型, R^2 在 0.91~1 之间为精准模型。则由以上数据分析波段组合所建立的反演模型中,B4/B5 的可决系数最大,为 0.604,其模型标准为较差模型。并对波段组合模型进行误差分析,结果显示以 B4/B5 波段所建立的总磷反演模型的相关性最好,最大相关误差为 68.67%,最小相对误差为 0.85%,相对误差平均值为 18.59%,虽然在相关性上有所提高,但是较最小二乘支持向量机模型的反演相对误差高 12.53%,所以该模型是遥感水质反演的较差模型,不足以准确地预测出清河水库总磷量。

本研究提出的第三种模型中,即最小二乘支持向量机模型,反演的相对误差最大值为 25.07%,最小相对误差为 0.01%,平均相对误差为 6.06%,且相对误差大于 20%的采样点只有 1 个。最小二乘支持向量机模型反演结果达到误差要求,能够较为准确地反演出清河水库总磷质量浓度。如图 4 为 3 种模型相对误差。最小二乘支持向量机模型相对误差的平均值最小,单波段模型相对误差平均值最大。多波段组合模型比单波段模型的相对误差平均值降低了 8.26%,最小二乘支持向量机模型比多波段组合模型降低了 12.53%。

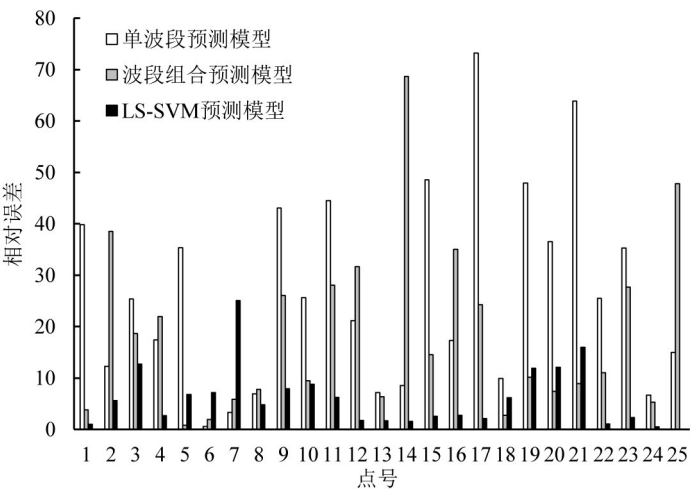


图 4 3 种模型清河水库总磷量反演相对误差比较

3 结 论

利用清河水库实测数据与 landsat8 卫星影像数据构建了适用于水库水体总磷质量浓度反演的 LS-SVM 模型。该模型所反演的总磷值与实测值的平均相对误差为 6.06%,总体优于现有研究正常用的单波段模型,多波段组合模型,满足反演模型对误差的要求,较为适用北方地区水库水体遥感反演。根据构建的 LS-SVM 模型对清河

水库水体总磷反演显示,清河水库水体中总磷质量浓度偏高,与实际情况相符,总磷来源可能为农药、化肥等的污染,应加强水库附近的水土保持和农田径流的控制与管理。但由于受到条件限制,本研究选取样点相对较少,今后研究应再加密采样点,增加模型的构建数据和模型的检验数据,可能会对模型精确度有进一步提高。

参考文献:

- [1] 王小钦,王钦敏,刘高焕,等. 水污染遥感监测[J]. 遥感技术与应用, 2002, 17(2):74-78.
- [2] 巩彩兰,尹球. 黄浦江水质指标与反射光谱特征的关系分析[J]. 遥感学报, 2006, 10(6):910-916.
- [3] 段洪涛,张柏,宋开山,等. 长春市南湖富营养化高光谱遥感监测模型[J]. 湖泊科学, 2005, 17(3):282-288.
- [4] 杨一鹏,王桥,王文杰,等. 水质遥感监测技术研究进展[J]. 地理与地理信息科学, 2004, 20(6):6-12.
- [5] SUN D Y, QIU Z F, LI Y M, et al. Detection of total phosphorus concentrations of turbid inland waters using a remote sensing method[J]. Water, Air, & Soil Pollution, 2014, 225:1 953.
- [6] 王学军,马廷. 应用遥感技术监测和评价太湖水水质状况[J]. 环境科学, 2000, 17(2):65-68.
- [7] 刘征,贺军亮,彭林,等. 黄壁庄水库总氮、总磷含量与反射光谱特征的关系[J]. 石家庄学院学报, 2009, 11(3):45-49.
- [8] 王林,赵冬至,杨建洪. 基于环境一号卫星的大洋河口海域营养盐遥感反演[J]. 中国环境科学, 2012, 32(1):136-141.
- [9] 王丽艳,李畅游,孙标. 基于MODIS数据遥感反演呼伦湖水体总磷浓度及富营养化状态评价[J]. 环境工程学报, 2014, 8(12):5 528-5 534.
- [10] ISENSTEIN E M, PARK M H. Assessment of nutrient distributions in Lake Champlain using satellite remote sensing[J]. Journal of Environmental Sciences, 2014, 26(9):1 831-1 836.
- [11] SILIÓ-CALZADA A, BRICAUD A, Gentili B. Estimates of sea surface nitrate concentrations from sea surface temperature and chlorophyll concentration in upwelling areas: A case study for the Benguela system[J]. Remote Sensing of Environment, 2008, 112(6):3 173-3 180.
- [12] WU C F, WU J P, QI J G, et al. Empirical estimation of total phosphorus concentration in the mainstream of the Qiantang River in China using Landsat TM data[J]. International Journal of Remote Sensing, 2010, 31(9):2 309-2 324.
- [13] GAO Y N, GAO J F, YIN H B, et al. Remote sensing estimation of the total phosphorus concentration in a large lake using band combinations and regional multivariate statistical modeling techniques[J]. Journal of Environmental Management, 2015, 151:33-43.
- [14] 李云梅,黄家柱,韦玉春,等. 湖泊富营养化状态的地面高光谱遥感评价[J]. 环境科学, 2006, 27(9):1 771-1 775.
- [15] CHANG N B, XUAN Z M, YANG Y J. Exploring spatiotemporal patterns of phosphorus concentrations in a coastal bay with MODIS images and machine learning models[J]. Remote Sensing of Environment, 2013, 134(4):100-110.
- [16] 雷坤,郑丙辉,王桥. 基于中巴地球资源1号卫星的太湖表层水体水质遥感[J]. 环境科学学报, 2004, 24(3):376-380.
- [17] 焦红波,查勇,李云梅,等. 基于高光谱遥感反射比的太湖水体叶绿素a含量估算模型[J]. 遥感学报, 2006, 10(2):242-248.
- [18] 王志辉. 基于MODIS的淡水湖泊富营养化状况遥感监测[D]. 武汉:武汉华中科技大学,2007.
- [19] 潘洁,张鹰,谭子辉. 基于高光谱的射阳河口悬浮泥沙浓度定量反演研究[J]. 海洋科学, 2011, 35(9):85-90.
- [20] 李国砚,张仲元,郑艳芬,等. MODIS影像的大气校正及在太湖蓝藻监测中的应用[J]. 湖泊科学, 2008, 20(2):160-166.

Calculating Total Phosphorus in Reservoirs Using the Satellite Landsat Data

WANG Yunxia¹, YANG Guofan¹, LIN Maosen¹, YANG Shuting²

(1. Shenyang Agriculture University, Shenyang 110866, China;

2. College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China)

Abstract: A least square support vector machine (LS-SVM) model was developed in this paper to establish the relationship between the data of OLI from landsat8 and the measured total phosphorus in Qinghe Reservoir. The total phosphorus (TP) in the reservoir was then inversely calculated based on the measured data from the LS-SVM. The results showed that the LS-SVM model increased the accuracy of estimated TP concentration with an average relative error of 6.06%, much lower than the 20.77% and 12.53%, the errors of the single band linear regression model and the band combination linear regression model, respectively. The results of The LS-SVM model also showed that the concentration of TP in Qinghe reservoir ranges from in 0.04~0.08 mg/L, higher than the normal range.

Key words: the total phosphorus; remote sensing; model; Qinghe reservoir

责任编辑:陆红飞