

基于光谱基线校正的季节性浑浊 II 类水体叶绿素 a 浓度遥感反演*

刘朝相^{1,2,3,4}, 宫兆宁^{1,2,3,4}, 赵文吉^{1,2,3,4}, 高明亮^{1,2,3,4}, 崔天翔^{1,2,3,4}, 刘 辉^{1,2,3,4}

(1:首都师范大学资源环境与旅游学院, 北京 100048)

(2:三维信息获取与应用教育部重点实验室, 北京 100048)

(3:资源环境与地理信息系统北京市重点实验室, 北京 100048)

(4:北京市城市环境过程与数字模拟国家重点实验室培育基地, 北京 100048)

摘 要: 水体 Chl. a 浓度是水质评价的一个重要指标, 受悬浮物浓度季节性变化的影响, 如何削弱悬浮物的光谱干扰, 是实现内陆水体 Chl. a 浓度遥感高精度反演的难点之一. 基于 2011—2013 年妫水河 6 次实测水体高光谱数据和水体 Chl. a 浓度数据, 评价广泛应用的三波段模型和非线性拟合能力较好的支持向量机回归 (SVR) 模型的反演精度, 使用基线校正和一阶微分方法来削弱实测高光谱中非 Chl. a 光谱信息. 定义两种基线: 750 nm 的反射率值; 500 与 750 nm 的反射率值连线, 基线校正为光谱反射率减去基线值. 利用 2013 年 7 月的实测数据进行验证, 结果表明, SVR 模型比三波段模型更适合季节性浑浊水体的 Chl. a 浓度反演. 通过基线校正筛选后的波段反射率组合作为输入变量能够提高 SVR 模型的反演精度, 决定系数为 0.68, 均方根误差为 3.38 $\mu\text{g/L}$; 线性基线校正提高三波段 Chl. a 估算模型的反演能力有限.

关键词: 水体 Chl. a 浓度; 季节性浑浊 II 类水体; 基线校正; 三波段模型; 支持向量机; 妫水河

Remote sensing retrieval for chlorophyll-a concentration in seasonal turbid case II water based on baseline correction of spectra

LIU Chaoxiang^{1,2,3,4}, GONG Zhaoning^{1,2,3,4}, ZHAO Wenji^{1,2,3,4}, GAO Mingliang^{1,2,3,4}, CUI Tianxiang^{1,2,3,4} & LIU Hui^{1,2,3,4}

(1: College of Resource Environment and Tourism, Capital Normal University, Beijing 100048, P. R. China)

(2: Key Laboratory of 3D Information Acquisition and Application of Ministry of Education, Beijing 100048, P. R. China)

(3: Key Laboratory of Resources Environment and GIS of Beijing Municipal, Beijing 100048, P. R. China)

(4: Base of the State Laboratory of Urban Environmental Processes and Digital Modeling, Beijing 100048, P. R. China)

Abstract: Chlorophyll-a (Chl. a) concentration is an important indicator of water quality evaluation, which is affected by the seasonal changes of suspended matter concentration. How to weaken the spectral interferences of suspended matter is an important issue in the retrieval of Chl. a concentrations with remote sensing in inland water. Based on six field surveys in 2011–2013 in Guishui River, the inversion ability of Chl. a concentrations with three-band model and support vector regression model were evaluated, in which baseline correction and first order differential method were used to weaken the spectral noises. Reflectance of 750 nm was defined as a baseline, a connecting line of reflectance from 500 nm to 750 nm was defined as another baseline. Baseline correction was defined as spectral reflectance minus baseline values. The validation was performed using data of samples in July 2013, the results demonstrated that SVR model was better than the three-band model in reversion of the Chl. a concentration in seasonal turbid water. Band reflectance combinations after baseline correction screening as the input variables can improve the inversion accuracy of the SVR estimation model. The determination coefficient was 0.68 and the root mean square error was 3.38 $\mu\text{g/L}$. Linear baseline correction can slightly improve the inversion accuracy of the three-band model.

Keywords: Water chlorophyll-a concentration; seasonal turbid case II water; baseline correction; three-band model; support vec-

* 国家高技术研究发展计划“863”项目 (2012AA12A308)、国家青年科学基金项目 (41101404)、国家基础测绘项目 (2011A2001) 和北京市教委科技计划面上项目 (KM201110028013) 联合资助. 2013-12-16 收稿; 2014-02-28 修改稿. 刘朝相 (1987~), 男, 硕士研究生; E-mail: liuchaozhuqiu@163.com.

tor machine; Guishui River

Chl. a 是水生藻类的重要成分,同时也是Ⅱ类水体水色反演的主要参数之一. Chl. a 浓度是表征光能自养生物量的重要指标,经常用于估测浮游植物的生物量和初级生产力,同时也能直观反映水体的富营养化程度,因此,Chl. a 浓度是水质状况评价的一个重要指标^[1-2].

Ⅱ类水体的物质组成成分较为复杂,光学特性主要受浮游植物色素、悬浮颗粒以及黄色物质等影响,不同组分光谱之间存在复杂的耦合作用机制,给 Chl. a 浓度反演带来很大的不确定性^[3],简单的经验模型很难满足这类水体 Chl. a 浓度反演的精度要求. 随着各种遥感传感器的投入使用,如何精确地定量模拟大量遥感信息与 Chl. a 浓度之间的关系成为 Chl. a 浓度反演的关键. 研究表明:Ⅱ类水体中,水体组分之间的光学影响具有不对等性,Chl. a 的存在对提取悬浮物信息特征波段的选择影响甚微,悬浮物的存在会对提取 Chl. a 信息特征波段的选取产生显著影响;随着悬浮物浓度不断增加,Chl. a 浓度对反射率的响应不断减弱,当总悬浮物浓度达到一定水平时(10 mg/L),反射光谱对 Chl. a 浓度(0~30 mg/L)的变化完全失去响应^[3]. 国内外学者在提高 Chl. a 浓度遥感反演精度方面做了大量研究. 张渊智等利用基于叶绿素荧光特征的荧光基线高度法及其各种变换形式在珠江口 Chl. a 浓度反演时获得较好的验证精度^[4]. 波段比值法^[5-6]和三波段法是两种常用的 Chl. a 浓度反演方法,基于生物光学模型和三波段法综合考虑水体的吸收和后向散射特性,力图削弱黄色物质和非色素颗粒物的影响,在大量浑浊度较低的内陆水体 Chl. a 浓度反演应用中^[7-11],取得较高的反演精度,成为目前应用最为广泛的方法之一. Gitelson 等利用从浑浊的湖泊、水库、河口和沿海水域大范围水体环境(不同光学特性、富营养化状态和地理位置)中收集的数据建立三波段模型,在不做任何参数调整的情况下,利用 MERIS 数据对三波段模型进行验证,结果发现,Chl. a 浓度模拟值与实测值呈高度相关,三波段 Chl. a 反演模型具有一定的普适性^[7,12]. 在浑浊度较高的水体中,悬浮物浓度在近外波段的影响限制了三波段模型的应用^[13]. 周琳等基于太湖 2004 年 10 月、2007 年 6 月、2008 年 10 月与 2010 年 4 月太湖水体实测高光谱数据,构建适合于高浑浊度水体 Chl. a 浓度的遥感估算模型:四波段模型和三波段模型,并利用 2009 年 10 月巢湖水体实测数据进行了验证^[14],结果表明:在三波段的基础上引入第四波段 720~780 nm,在满足非色素颗粒在第三波段与第四波段的吸收系数与后向散射系数相等的假设条件下,四波段模型适合高浑浊水体,线性关系比三波段模型好,但假设条件的普适性有待验证. 韦玉春等将实验室内的基线校正思路引入到太湖梅梁湾水体实测高光谱数据的预处理中,以 450~750 nm 之间的反射率直线为基线,削弱悬浮泥沙对水体光谱的贡献,建立波段比值模型和三波段半分析模型,对比分析基线校正前后模型的验证精度,结果表明基线校正能够削弱悬浮泥沙对 Chl. a 的光谱贡献^[15].

波段比值、三波段、改进的三波段等方法在一定程度上提高了 Chl. a 浓度的反演精度,但并没有充分利用实测高光谱的多波段特征,忽略了可见光范围内其他光谱波段承载的 Chl. a 信息. 遥感数据各波段光谱信息与 Chl. a 浓度之间为复杂非线性映射关系,研究表明人工神经网络、支持向量机等非线性映射方法能够很好地反映二者的关系^[16-17]. 高玉蓉等比较分析波段比值、三波段模型和人工神经网络模型等多种算法的验证结果表明,人工神经网络模型精度最高,最适合应用于利用实测光谱估算千岛湖夏季 Chl. a 浓度^[18]. Sun 等利用支持向量机回归模型(SVR)和实测高光谱数据对太湖的 Chl. a 浓度进行了估算,首先经过波段分析和迭代算法识别出最优三波段,然后将它们作为输入变量建立 SVR 模型并进行 Chl. a 浓度的估算,结果表明 SVR 模型的预测精度优于典型的线性回归模型^[19]. Bazi 等利用粒子群优化算法和 SVR 技术进行基于 MERIS 和 SeaBAM 数据的海岸水体 Chl. a 浓度的估算,结果表明参数优化可以提高 SVR 的预测精度,同时,会降低训练样本大小对 SVR 模型普适性的影响^[20]. Camps-Valls 等基于卫星遥感影像数据利用支持向量机、神经网络和典型的生物光学模型进行海洋水体 Chl. a 浓度的估算,发现 SVR 模型精度最高,偏差小,尤其是训练样本有限的情况下 SVR 具有很高的实用性^[21].

妯水河是官厅水库的 3 大支流之一,监测妯水河水体 Chl. a 能间接反映库区水体的富营养化程度. 受暖温带大陆性季风气候的影响,其水体具有明显的季节性浑浊特征. 本研究通过分析妯水河不同月份的实测高光谱数据的特征,对高光谱数据进行有针对性的基线校正,建立基线校正前后的三波段 Chl. a 估算模型,以期发现适合反演 Chl. a 浓度的基线校正方法;对比分析三波段模型和 SVR 模型的估算结果,寻求遥感

反演 Chl. a 浓度的最佳反演模型。

1 数据资料及分析处理

1.1 实测数据

妫水河全长 18.5 km,流域面积 1073.6 km²,多年平均年径流量为 1.18×10^8 m³,枯水期流量为 0.3 m³/s. 研究区为妫水河下游河段,地理位置如图 1 所示. 于 2011 年 9 月 25 日、2012 年 5 月 9 日、2012 年 7 月 11 日、2012 年 9 月 29 日、2013 年 6 月 2 日、2013 年 7 月 16 日在妫水河进行 6 次野外实验,共获得 102 个样本点的水体实测高光谱数据. 图 1 代表 2013 年 7 月 16 日样本点的地理位置,其余日期的样本点位置与其相似. 光谱测量仪器选用美国 ASD 公司生产的 Field Spec 3 背挂式野外高光谱辐射仪,能够在 350 ~ 1000 nm 波长范围内连续测量,光谱分辨率为 3 nm,采样间隔为 1.4 nm,光纤前视场角为 25°. 测量光谱的时间为每天 10:00–14:00,期间天空晴朗无云,水面平静. 测量时为了避开太阳直射反射和船舶阴影对光场的破坏,现场船舶上的观测几何参照唐军武等^[22]推荐的方法,每个样本点分别测量参考白板、天空光和水体的光谱信息,各测 10 条光谱. 将光谱值较大的数据删除,剩余光谱取均值作为样本点水体的光谱数据,并将光谱数据转换为反射率. 采用手持式 GPS 对样本点进行定位,用 Hydrolab DS 5x 野外水质仪同步采集 Chl. a 浓度,每次采集数据之前都对仪器进行定标,Chl. a 浓度范围为 3.90 ~ 37.26 μg/L. 7 月部分样本点的水面有浮萍覆盖,影响水体光谱测量,掩盖 Chl. a 的真实浓度,去掉这类样本点后剩余 97 个有效样本点,用于构建模型的区域化参数.

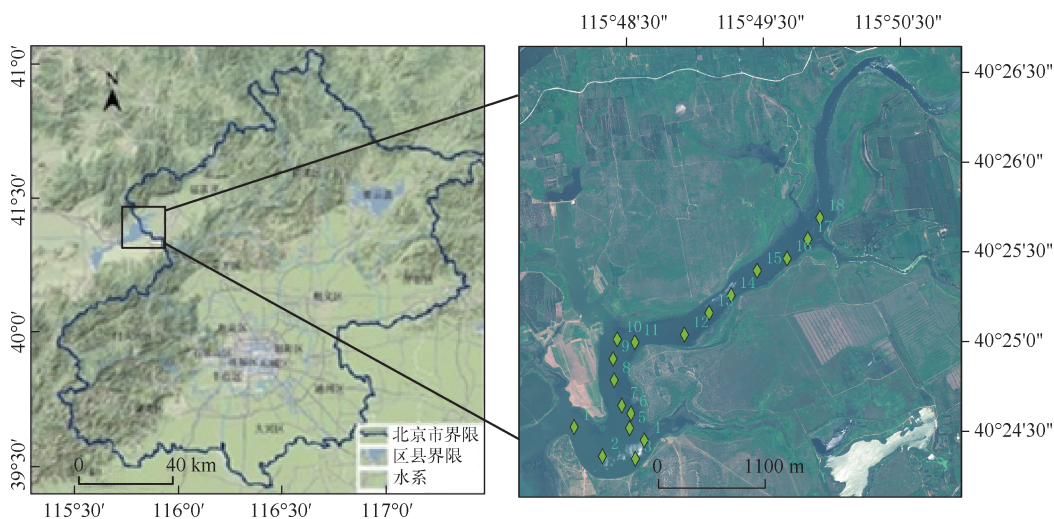


图 1 研究区地理位置

Fig. 1 Location of study area

1.2 光谱数据的分析及预处理

1.2.1 基线校正 将实验室内水体组分的光学性质引入到水体实测高光谱数据的预处理中,削弱水体其他组分对水体 Chl. a 光谱信息的贡献,有利于突出光谱中 Chl. a 信息. 基线校正为光谱反射率减去基线值,利用基线校正法去除基线,使测量结果在同一个标准上进行比较.

已有研究表明^[15],水体中悬浮物含量增加使水体反射率提高,高浓度悬浮物水体中,蓝绿光(400 ~ 500 nm)的反射率一般高于 700 nm 后的红光反射率;随着悬浮物浓度的增加,676 nm 周围的叶绿素吸收峰不明显;同时,450 nm 相对于 750 nm 的反射率均有不同程度的抬升.

水色遥感研究中,依据水体 750 nm 光谱反射率近似为 0,Althuis 等^[23]将 750 nm 的反射率作为基线,Mustapha 等^[24]将 685 nm 附近的值作为基线,Abbott 等^[25]将 Chl. a 在 680 nm 附近的第 2 个吸收峰与 750 nm 之间的连线作为基线,韦玉春等^[15]以 450 ~ 750 nm 之间的反射率直线作为基线. 以上研究结果表明线性基

线校正可以作为光谱数据处理方法,能够起到提高叶绿素 a 浓度反演精度的作用。

1.2.2 数据处理及分析 奶水河水体表面光谱曲线如图 2a、图 2b 所示。将水表面光谱数据按月平均,得到每月平均反射率光谱曲线,可为光谱数据按月分析提供概括性参考,结果如图 2c 所示。

与高悬浮物浓度水体的光谱特征相比,6 个月份的光谱曲线中,450 nm 反射率相对于 750 nm 反射率都有所抬升;2012 年 5 月 16 日、2013 年 6 月 2 日和 2013 年 7 月 16 日的光谱曲线 676 nm 左右的吸收峰和 700 nm 左右的荧光峰不明显,不同的是 Chl. a 浓度较高导致 2011 年 9 月 25 日、2012 年 7 月 11 日和 2012 年 9 月 29 日 3 个数据存在明显的吸收峰和荧光峰;同时,2012 年 5 月 16 日、2012 年 7 月 11 日和 2013 年 6 月 2 日的光谱曲线整体上相对于其他 3 个时期呈现明显抬升。反射率整体抬升,吸收峰与荧光峰不明显可以看作水体中悬浮物颗粒等后向散射的结果。因此,光谱数据存在基线校正的必要性。图中曲线 750 nm 处反射率值都不为零,据此定义第 1 种基线为 750 nm 的反射率值。

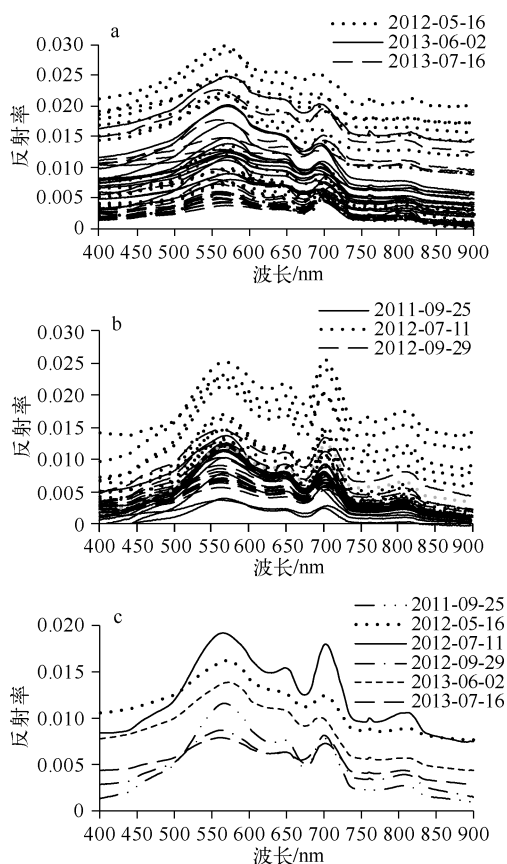


图 2 水表面实测光谱反射率
Fig. 2 Measured spectral reflectance
of water surface

水体光学性质研究表明^[26-27],悬浮物在 555 nm 的后向散射与悬浮物干重具有线性正相关关系,散射系数在 440 和 676 nm 由于藻颗粒的吸收存在低谷,散射系数从 500 ~ 750 nm 随着波长的增加而呈线性或指数减少。以此为依据定义第 2 种基线:500 与 750 nm 的反射率值连线,选择 500 nm 作为基线的起点同时可以避开可溶性有机物 (CDOM) 的影响。

基于以上分析,确定基线校正的应用条件为:(1)室内水体 750 nm 反射率接近于 0;(2)悬浮物在 500 ~ 750 nm 的后向散射系数随着波长的增加而呈线性减少。对原始光谱数据按以下 5 种方式进行基线校正,基线校正为光谱反射率减去基线值:

方式 1:以 750 nm 反射率为基线,对所有光谱进行线性基线校正。

方式 2:连接 500 和 750 nm 的反射率直线作为基线,对所有光谱曲线进行线性基线校正。

方式 3:连接 500 和 750 nm 的反射率直线作为基线,只对 5、6 和 7 月光谱数据进行基线校正。

方式 4:对 5、6 和 7 月光谱数据进行 500 ~ 750 nm 基线校正,对 9 月数据进行 750 nm 基线校正。

方式 5:先对所有光谱进行 750 nm 基线校正,然后进行 500 ~ 750 nm 基线校正。

2 研究方法

多种生物光学算法被发掘并大量应用于相关水体 Chl. a 浓度信息的提取,其中,三波段算法在许多研究中都取得了很好的结果,是简单可行、应用最广的 Chl. a 浓度估算方法。以三波段为基础的其他生物

光学算法的变形形式都存在其他附加的假设条件,普适性有待验证。SVR 算法在许多分类与回归问题中都取得优于其他算法的结果,代表着目前性能最好的神经网络算法^[28-30],在数据的非线性回归拟合应用中能很好地拟合出数据之间的相关关系。因此,本文以三波段法和支撑向量机算法为基础,进行水体 Chl. a 浓度的遥感反演。

2.1 三波段法

三波段模型的基本形式为:

$$Chl. a \propto [R_{rs}^{-1}(\lambda_1) - R_{rs}^{-1}(\lambda_2)]R_{rs}(\lambda_3) \quad (1)$$

式中, $R_{rs}(\lambda_i)$ 为波段 λ_i 的遥感反射率; λ_1 应选择对 Chl. a 吸收最敏感的区域; λ_2 是 Chl. a 荧光峰区域即 Chl. a 吸收最不敏感的区域, 且 λ_2 处的悬浮物和黄色物质的吸收系数与 λ_1 处非常接近, 因此 $R_{rs}^{-1}(\lambda_1) - R_{rs}^{-1}(\lambda_2)$ 可以排除黄色物质、非色素颗粒物吸收干扰; λ_3 波段区域内水体吸收系数远大于后向散射系数, 是各成分吸收最小的区域, 主要表现为纯水吸收, 引入 λ_3 主要是削弱悬浮物后向散射对反演 Chl. a 浓度的干扰^[31].

2.2 支持向量机

支持向量机(support vector machine, SVM)是由 Vapnik 首先提出的, 理论基础是统计学习理论的 VC 维理论和结构风险最小化原则, 是结构风险最小化的近似实现^[32]. 支持向量机的基本思想是通过非线性映射将输入向量映射到一个高维空间, 然后在高维空间构造一个最优超平面, 采用核函数避免了显示的非线性映射, 可以克服高维空间带来的计算困难, 为解决非线性回归问题提供了一种新思路. 借鉴 SVM 分类问题解决方法的思路, 先定义最优回归超平面, 进而解决线性 SVM 回归问题, 另外引入核函数, 可以解决非线性回归问题. 国内外大量研究表明支持向量机在解决小样本、非线性和高维模式识别问题中表现出许多特有的优势, 具有泛化能力强、容易训练、没有局部极小等优点. 支持向量机在解决人工神经网络“过学习”等问题方面有着较好的优势^[33]. 采用径向基函数为核函数, 算法性能的主要影响因素是核函数参数 g 和惩罚系数 c , 采用交叉验证法寻找最佳 c 和 g . SVM 被应用于回归拟合时, 输入变量的选择对回归拟合结果的优劣起决定性作用.

2.3 交叉验证(K-CV)

原始数据被随机分为 K 组(一般是均分), 将每个子集数据分别做 1 次验证集, 其余的 $K-1$ 组子集数据作为训练集, 这样得到 K 个模型, 用这 K 模型最终验证集的均方根误差的平均数作为此 K -CV 回归模型的性能指标, 及衡量模型参数 c 和 g 好坏的指标. 在 CV 意义下, 采用网格划分来寻找模型的最佳参数 c 和 g .

使用均方根误差(RMSE)和决定系数(R^2)作为三波段模型和 SVR 模型评价指标:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\hat{I}_i - I_i)^2}{N}} \quad (2)$$

$$R^2 = \frac{\sum_{i=1}^N (\hat{I}_i - \bar{I})^2}{\sum_{i=1}^N (I_i - \bar{I})^2} \quad (3)$$

式中, \hat{I}_i 、 I_i 分别表示样本点 i 对应的模拟值与检验值, \bar{I} 对应于各样本点模拟值的均值, N 为样本数. Williams 等认为如果 R^2 大于 0.91, 则模型的拟合效果表现为精确; R^2 在 0.82 ~ 0.90 之间模型表现为很好; R^2 在 0.66 ~ 0.81 之间模型是近似模拟; R^2 在 0.50 ~ 0.65 之间时, 模型的模拟效果普遍认为较差^[34].

3 模型的构建与应用

3.1 模型的构建

选择 2011 年 9 月至 2013 年 6 月的 80 个样本点数据构建模型, 2013 年 7 月的 17 个样本点用来验证模型的普适性.

一阶微分技术是处理高光谱遥感数据的一种重要方法, 通过对光谱数据进行一阶导数处理可以削弱环境中部分线性变化因素的影响. 为了分析基线校正对 Chl. a 信息提取的影响, 分别在原始光谱数据、原始光谱的一阶微分数据和原始光谱的各种基线校正数据的状态下, 分析三波段估算模型的估算精度. 总结学者大量的研究成果^[2,4,8-9,34], 本文三波段的取值范围为: $660 \text{ nm} \leq \lambda_1 \leq 690 \text{ nm}$, $690 \text{ nm} \leq \lambda_2 \leq 730 \text{ nm}$, $700 \text{ nm} \leq \lambda_3 \leq 750 \text{ nm}$. 经过三波段值与 Chl. a 浓度的迭代分析, 确定二者之间的最大相关系数和对应的三波段, 结果如表 1 所示.

表 1 三波段和三波段值与 Chl. a 浓度值的相关系数

Tab. 1 Three bands and correlation coefficients of three-band values and chlorophyll-a concentrations

类型代号	数据类型	三波段表达式	λ_1/nm	λ_2/nm	λ_3/nm	相关系数(R)
a	原始光谱数据	$[R_{rs}^{-1}(\lambda_1) - R_{rs}^{-1}(\lambda_2)]R_{rs}(\lambda_3)$	675	690	745	0.756
b	一阶微分光谱数据		682	701	705	0.816
c	基线校正方式 1 校正的数据		675	690	740	0.663
d	基线校正方式 2 校正的数据		670	690	739	0.647
e	基线校正方式 3 校正的数据		661	690	740	0.762
f	基线校正方式 4 校正的数据		680	690	744	0.659
g	基线校正方式 5 校正的数据		670	690	738	0.614

分别对表 1 中 a~g 7 种情况下的数据建立三波段模型,结果如图 3 所示. 可以发现,一阶微分三波段回归模型的 $R^2=0.73$ 最高(图 3b),只对 5、6 和 7 月的光谱进行以 500~750 nm 基线校正后的光谱三波段模型的 $R^2=0.66$ 次之(图 3e),原始光谱三波段回归模型的 $R^2(0.65)$ 排列第 3(图 3a),其余基线校正后的光谱三波段模型的 R^2 低于 0.6(图 3c、d、f、g).

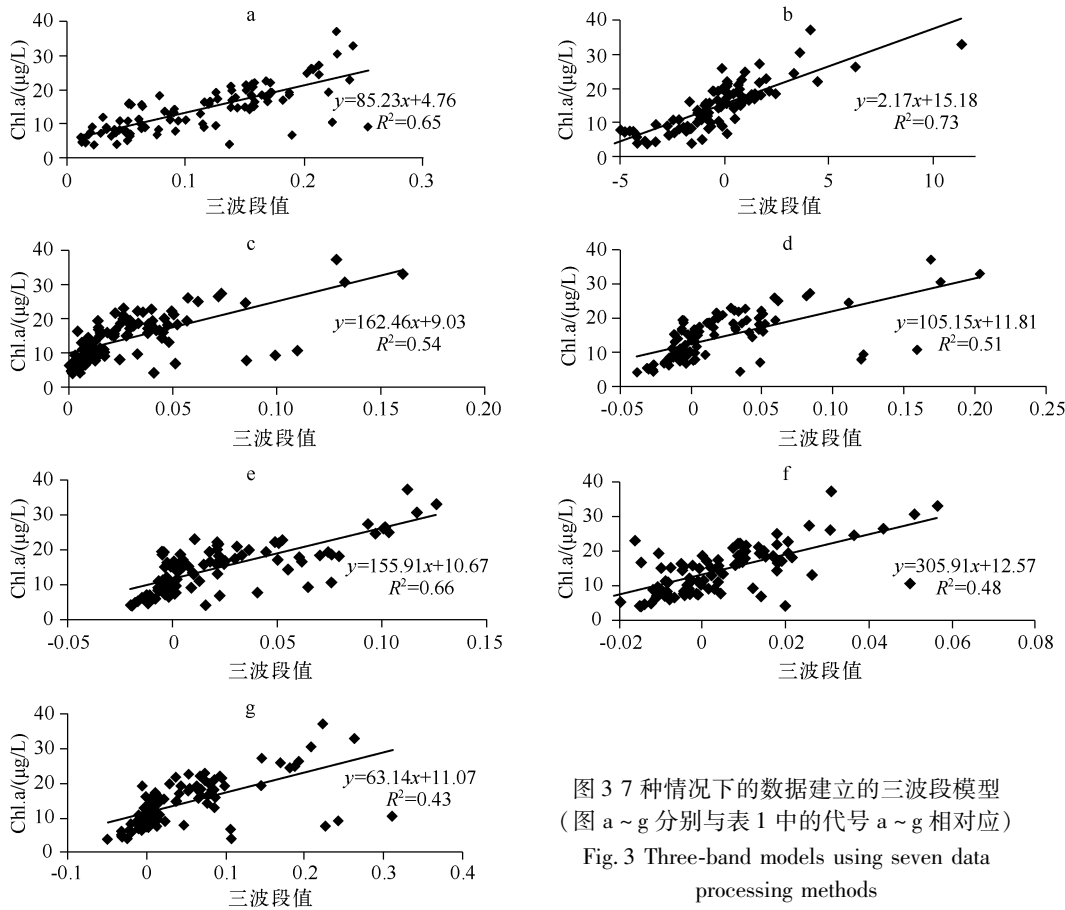


图 3 7 种情况下的数据建立的三波段模型 (图 a~g 分别与表 1 中的代号 a~g 相对应)

Fig. 3 Three-band models using seven data processing methods

原始光谱数据及其各种基线校正后的光谱数据三波段值与 Chl. a 浓度的相关系数都在 0.6 以上,三波段中 λ_1 重复出现的波段有 670 和 675 nm, λ_3 出现一个波段 740 nm 重复,除了一阶微分光谱数据的 λ_2 为 701 nm 外, λ_2 都维持在一个稳定的波段 690 nm(表 1). 相关系数表明这些三波段光谱数据里都承载着较多

的 Chl. a 信息. 因此,在此基础上筛选 SVR 模型的输入变量可以充分利用实测高光谱多个波段中蕴含的 Chl. a 信息.

输入变量的选择决定 SVR 估算模型的预测精度,首先通过光谱数据与 Chl. a 浓度值迭代分析找出 a ~ g 的三波段 λ_1 、 λ_2 、 λ_3 ,如表 1 所示,然后将所有 λ_1 、 λ_2 、 λ_3 波段值与 Chl. a 浓度做相关分析,删除相关系数小于 0.49 的 λ_1 、 λ_2 、 λ_3 波段,最终确定如表 2 所示的输入变量,在寻找最佳参数 c 跟 g 时采用 3 折交叉验证,训练好 SVR 模型,模型的训练 $R^2=0.68$.

表 2 训练三波段模型与 SVR 模型
Tab.2 Training three-band model and SVR model

类型代号	三波段模型			支持向量机	
	x	线性表达式	训练 R^2	SVR 输入变量/nm	训练 R^2
a	$[R_{rs}^{-1}(\lambda_1) - R_{rs}^{-1}(\lambda_2)]R_{rs}(\lambda_3)$	$y = 85.23 x + 4.76$	0.65		
b		$y = 2.17 x + 15.18$	0.73	682、701、705	0.68
c		$y = 162.46 x + 9.03$	0.54	675	
d		$y = 105.15 x + 11.81$	0.51	739	
e		$y = 155.91x + 10.67$	0.66	661	
f		$y = 305.91 x + 12.57$	0.48	744	
g		$y = 63.14 x + 11.07$	0.43	670、738	

3.2 模型的应用

将 2013 年 7 月的 17 个样本点数据代入训练好的三波段模型和 SVR 模型,模型的反演结果如表 3 所示. 结合表 2 和表 3 可发现:三波段模型中,一阶微分三波段模型具有最小 $RMSE(5.17 \mu g/L)$ 和最大 $R^2(0.73)$,表明一阶微分技术是最适合建立三波段模型的光谱数据预处理方法;对 5、6、7 月份的光谱数据进行 500 ~ 750 nm 基线校正后建立的三波段模型, $R^2(0.66)$ 大于原始光谱三波段模型的 $R^2(0.65)$,同时 $RMSE(5.47 \mu g/L)$ 小于后者(5.54 $\mu g/L$),表明模型的推广性较好,这种季节性基线校正有利于提高 Chl. a 浓度的估算精度.

SVR 模型的 R^2 为 0.68,略小于一阶微分后的三波段模型,但大于其他所有三波段模型,同时, $RMSE$ 为 3.38 $\mu g/L$,表明 SVR 模型明显优于三波段模型. 分析 SVR 的输入变量可发现,未经过基线校正和去噪处理的原始光谱波段未能被选中,所有入选波段均位于 660 ~ 750 nm 之间,此波段区间包含 Chl. a 的荧光峰和吸收峰,承载较多的 Chl. a 信息,同时也可能存在大量噪音信息. 研究发现将原始光谱的 661、670、675、682、701、705、738、739 和 744 nm 作为输入变量时,SVR 回归模型的 $RMSE$ 为 6.44 $\mu g/L$, R^2 为 0.52,由此发现各种基线校正有利于筛选 SVR 回归模型的输入变量,提高模型的估算精度.

表 3 三波段模型与 SVR 模型的反演结果
Tab.3 Results of three-band model and SVR model

类型代号	三波段模型		支持向量机
	线性表达式	$RMSE/(\mu g/L)$	$RMSE/(\mu g/L)$
a	$y = 85.23 x + 4.76$	5.54	3.38
b	$y = 2.17x + 15.18$	5.17	
c	$y = 162.46 x + 9.03$	6.53	
d	$y = 105.15 x + 11.81$	6.55	
e	$y = 155.91x + 10.67$	5.47	
f	$y = 305.91 x + 12.57$	5.60	
g	$y = 63.14 x + 11.07$	5.83	

从实测 Chl. a 浓度与反演 Chl. a 浓度的曲线关系(图 4)可得出,SVR 模型的整体反演效果最好,反演值

的曲线最接近实测值曲线,原始三波段模型和经过基线校正的三波段模型的反演结果都存在几个误差较大的偏离值,说明三波段模型不如 SVR 模型稳定,推广性不如 SVR 模型. 虽然一阶微分光谱三波段模型的 $RMSE$ 小、 R^2 大,但反演值太过均匀. 两个模型评价参数都表明,一阶微分光谱三波段模型适合估算 Chl. a 浓度,但反演结果不能真实反映 Chl. a 浓度的分布趋势.

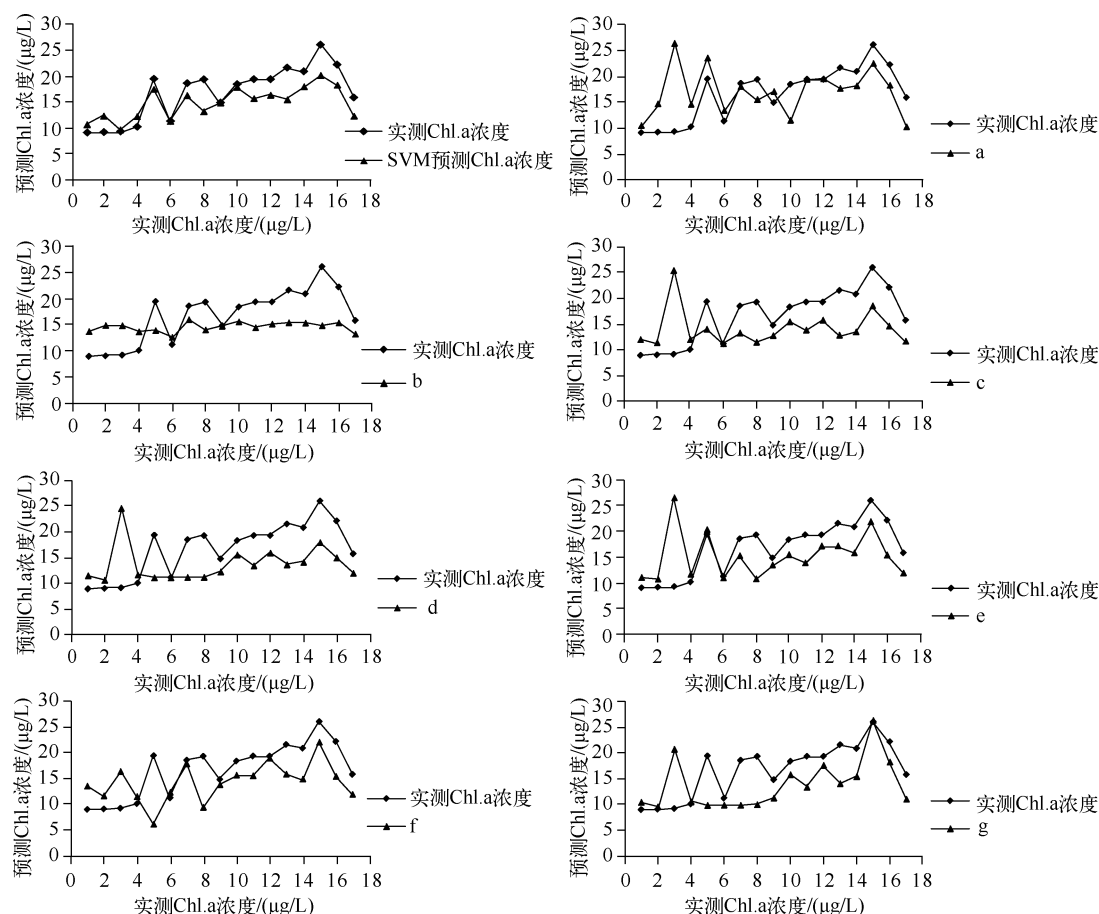


图4 实测 Chl. a 浓度与反演 Chl. a 浓度的趋势图(图中的 a~g 分别与表1中的代号 a~g 相对应)

Fig. 4 Trend charts of measured Chl. a concentrations and retrieved Chl. a concentrations

4 结论与讨论

通过分析原始光谱曲线发现5、6和7月份的光谱数据存在明显的高悬浮物浓度特征,9月份的光谱数据存在部分高悬浮物浓度特征. 为了削弱悬浮物的光谱影响,利用两种基线校正方法对不同月份的水体光谱曲线进行基线校正;同时,对原始光谱进行一阶微分去噪处理,在此基础上建立三波段模型. 为了充分利用三波段以外的其他实测高光谱波段承载的 Chl. a 信息,另外建立 SVR 模型. 对比分析各种模型的 R^2 和反演结果的 $RMSE$,得出以下结论,并对结论进行讨论:

1) 与原始数据的三波段模型相比,将 750 nm 反射率值作为基线,对所有月份光谱数据进行线性基线校正不能提高水体 Chl. a 浓度的反演精度. 导致误差较大的原因可能是 750 nm 光谱包含 Chl. a 信息,基线校正去除环境背景光谱噪音的同时也损失掉一部分 Chl. a 信息;将 500 与 750 nm 反射率值连线作为基线,对所有月份光谱数据进行线性基线校正也不利于提高水体 Chl. a 浓度的反演精度,导致误差的原因是 9 月份的光谱数据在 660~750 nm 范围存在明显的吸收峰和荧光峰,Chl. a 光谱特征信息明显,线性基线校正削弱

了部分 Chl. a 信息.

2) 与原始数据的三波段模型相比,将 500 与 750 nm 反射率值连线作为基线,只对存在明显高悬浮物浓度特征的 5、6 和 7 月份的光谱数据进行线性基线校正有利于提高水体 Chl. a 浓度的反演精度. 以此为基础的三波段模型的 R^2 为 0.66, RMSE 为 $5.47 \mu\text{g/L}$, 这种基线校正的方式要优于其他基线校正,这说明对季节性浑浊水体有针对性地实施基线校正能够提高反演精度. 一阶微分光谱技术有利于削弱光谱信息中环境信息的影响,但三波段模型的反演结果不能反映真实 Chl. a 浓度的分布趋势,不适合作为三波段建模的光谱预处理;然而,通过一阶微分可以寻找承载较大 Chl. a 信息的波段.

3) SVR 模型能够充分利用实测高光谱数据承载的 Chl. a 浓度信息,拟合出二者之间的非线性关系,适合用来反演纳水河水体 Chl. a 浓度. SVR 模型的稳定性与扩展性都要优于三波段模型,模型的 R^2 为 0.68, RMSE 为 $3.38 \mu\text{g/L}$. 两个模型评价参数都要优于所有三波段模型的对应值. 然而,SVR 模型的输入变量不容易确定,需要验证各种波段反射率组合情况,本文在基线校正和一阶微分原始光谱数据的基础上,筛选出所有与 Chl. a 浓度相关系数较大的波段,以这些波段值为因变量,能够提高 SVR 模型的拟合精度与反演精度.

综上所述,三波段模型简单可行,但应用模型估算 Chl. a 浓度时会存在偏差较大的值,影响模型的整体效果;SVR 模型的扩展性很好,估算结果不会存在个别偏差较大的值,但模型的输入变量缺少一个固定的选择方式. 以 500~750 nm 为基线,对浑浊特征明显的光谱进行基线校正能够削弱悬浮物浓度的光谱贡献,但简单的线性基线校正方法校正效果有限. 悬浮物与 Chl. a 的光谱耦合并非简单的线性关系,下一步需深入研究悬浮物与 Chl. a 的光谱耦合关系,探寻更好的基线校正方法,提高 Chl. a 浓度反演模型的反演精度.

致谢:感谢林川、赵雅丽、甄姿、李洪等在野外试验过程中给予的帮助!

5 参考文献

- [1] 李素菊,吴倩,王学军等. 巢湖浮游植物叶绿素含量与反射光谱特征的关系. 湖泊科学, 2002, **14**(3): 228-234.
- [2] 王乔,张兵,韦玉春等. 太湖水体环境遥感监测实验及其软件实现. 北京:科学出版社, 2007: 20-21.
- [3] 周冠华,田国良,陈军等. 水体组分光谱耦合效应研究. 光谱学与光谱分析, 2010, **30**(2): 470-475.
- [4] 张渊智,陈楚群,段洪涛等. 水质遥感理论、方法及应用. 北京:高等教育出版社, 2011: 4.
- [5] 段洪涛,张柏,宋开山等. 查干湖叶绿素 a 浓度高光谱定量模型研究. 环境科学, 2006, **27**(3): 503-507.
- [6] Gurlin D, Gitelson AA, Moses WJ. Remote estimation of chl-a concentration in turbid productive waters-Return to a simple two-band NIR-red model? *Remote Sensing of Environment*, 2011, **115**(12): 3479-3490.
- [7] Gitelson AA, Dall'Olmo G, Moses W. A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: Validation. *Remote Sensing of Environment*, 2008, **112**(9): 3582-3593.
- [8] 李方,徐京萍,马荣华等. 内陆水体叶绿素 a 浓度反演的区域化三波段模型研究. 遥感学报, 2011, **15**(6): 1156-1170.
- [9] 施坤,李云梅,刘忠华等. 基于不同光谱主导因子的内陆湖泊水体叶绿素浓度三波段反演模型研究. 环境科学, 2010, **31**(12): 2874-2881.
- [10] Zhou L, Roberts DA, Ma W *et al.* Estimation of higher chlorophylla concentrations using field spectral measurement and HJ-1A hyperspectral satellite data in Dianshan Lake, China. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2014, **88**: 41-47.
- [11] Gitelson AA, Schalles JF, Hladik CM. Remote chlorophyll-a retrieval in turbid, productive estuaries: Chesapeake Bay case study. *Remote Sensing of Environment*, 2007, **109**(4): 464-472.
- [12] Le C, Hu C, Cannizzaro J *et al.* Evaluation of chlorophyll-a remote sensing algorithms for an optically complex estuary. *Remote Sensing of Environment*, 2013, **129**(15): 75-89.
- [13] 王胜强,陈晋,杨伟等. Lee 生物光学模型在不同水体组分特性下的适用性. 湖泊科学, 2011, **23**(2): 217-222.
- [14] 周琳,马荣华,段洪涛. 浑浊 II 类水体叶绿素 a 浓度遥感反演(I): 模型的选择. 红外与毫米波学报, 2011, **30**(6): 531-536.
- [15] 韦玉春,王国祥,程春梅. 面向浑浊水体叶绿素 a 浓度遥感反演的光谱基线校正. 光谱学与光谱分析, 2012, **32**(9): 2546-2550.

- [16] Zhang Y, Pulliainen J. Application of an empirical neural network to surface water quality estimation in the Gulf of Finland using combined optical data and microwave data. *Remote Sensing of Environment*, 2002, **81**(2/3):327-336.
- [17] Mountrakis G, Im J, Ogole C. Support vector machine in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2011, **66**(3):247-259.
- [18] 高玉蓉, 刘明亮, 吴志旭. 应用实测光谱估算千岛湖夏季叶绿素 a 浓度. 湖泊科学, 2012, **24**(4):553-561.
- [19] Sun D, Li Y, Wang Q. A unified model for remotely estimating chlorophyll-a in Lake Taihu, China, based on SVM and *in situ* hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 2009, **47**(8):2957-2965.
- [20] Bazi Y, Melgani F. Semi-supervised PSO-SVM regression for biophysical parameter estimation. *IEEE Transactions on Geoscience and Remote Sensing*, 2007, **45**(6):1887-1895.
- [21] Camps-Valls G, Bruzzone L, Rojo-Alvarez JL *et al.* Robust support vector regression for biophysical variable estimation from remotely sensed images. *IEEE Geoscience and Remote Sensing Letters*, 2006, **3**(3):339-343.
- [22] 唐军武, 田国良, 汪小勇等. 水面光谱测量与分析: 水面以上测量法. 遥感学报, 2004, **8**(1):37-44.
- [23] Althuis IJA, Shimwell S. Modelling of remote sensing reflectance spectra for suspended matter concentration detection in coastal waters. *EARSel Advances in Remote Sensing*, 1995, **4**(1):53-59.
- [24] Mustapha B, Larouche SP. Geoscience and remote sensing symposium. IGARSS 2008. *IEEE International*, 2008, **4**(IV):914.
- [25] Abbott MR, Letelier RM. Algorithm theoretical basis document chlorophyll fluorescence, MODIS Product Number 20, NASA, 1999.
- [26] Klepper O, Vermij SG, Lingeman R. The influence of light scattering on vertical extinction in Lake Maarsseveen. *Verh. Int. Ver. Limnol*, 1984, **22**: 82-86.
- [27] Babin M, Morel A, Fournier-Sicre V *et al.* Light scattering properties of marine particles in coastal and open ocean waters as related to the particle mass concentration. *Limnology and Oceanography*, 2003, **48**(2): 843-859.
- [28] Camps-Valls G, Gómez-Chova L, Muñoz-Marí J *et al.* Retrieval of oceanic chlorophyll concentration with relevance vector machines. *Remote Sensing of Environment*, 2006, **105**(1):23-33.
- [29] 梁 栋, 管青松, 黄文江等. 基于支持向量机回归的冬小麦叶面积指数遥感反演. 农业工程学报, 2013, **29**(7): 117-123.
- [30] 蒋 刚, 肖 建, 郑永康等. 基于支持向量机的一类水域叶绿素 a 浓度反演研究. 计算机应用, 2005, **25**(10): 2398-2409.
- [31] 张渊智, 陈楚群, 段洪涛等. 水质遥感理论、方法及应用. 北京: 高等教育出版社, 2011:4.
- [32] 史 峰, 王小川, 郁 磊等. MATLAB 神经网络 30 个案例分析. 北京: 北京航空航天大学出版社, 2004:1-19.
- [33] 梁 坚. 支持向量机在水质评价及预测中的应用研究[学位论文]. 杭州: 浙江工业大学, 2011:1-73.
- [34] Song K, Li L, Tedesco LP *et al.* Remote estimation of chlorophyll-a in turbid inland waters: Three-band model versus GA-PLS model. *Remote Sensing of Environment*, 2013, **136**:342-357.