

科学出版社
出版
中国地理学会环境遥感分会
中国科学院遥感与数字地球研究所
主办

JOURNAL OF
REMOTE
SENSING

遥感学报

2019年 Vol.23 第23卷 No.4 第4期 ISSN 1007-4619 CN11-3841 / TP CODEN YXAUAB



遥感学报

Yaogan Xuebao

第 23 卷 第 4 期 2019 年 7 月

目 次

学者观点

对遥感科学应用的一点看法..... 官鹏 (567)

综述

基于 FY-2F 数据的中国区域地表温度日变化模型评价及特征研究 孟翔晨, 刘昊, 程洁 (570)

遥感影像的海岸线自动提取方法研究进展..... 吴一全, 刘忠林 (582)

遥感数据融合研究进展与文献定量分析 (1992—2018)..... 张立福, 彭明媛, 孙雪剑, 岑奕, 童庆禧 (603)

技术方法

机载 WIDAS 地表观测的 BRDF 原型反演算法验证

..... 何丹丹, 焦子锜, 董亚冬, 张小宁, 张虎, 丁安心 (620)

融合数据内部变化信息的丰度估计算法..... 袁静, 章毓晋 (630)

多星融合的土壤湿度滚动式估算模型..... 梁月吉, 任超, 黄仪邦, 王浩宇, 卢献健, 晏红波 (648)

Ross-Li 核驱动模型热点参数化及其校正—以 POLDER 数据为例

..... 常雅轩, 焦子锜, 董亚冬, 张小宁, 何丹丹, 尹思阳, 崔磊, 丁安心 (661)

基准影像数据辅助遥感影像几何定位 潘雪琛, 姜挺, 余岸竹, 王鑫, 张一 (673)

综合多特征的极化 SAR 图像随机森林分类算法 徐乔, 张霄, 余绍淮, 陈启浩, 刘修国 (685)

Cesium 框架多源电子地图瓦片数据混搭方案设计 范俊甫, 胡桃英, 何惠馨, 秦柳, 李桂华 (695)

自适应分块加权 Wallis 并行匀色 李烁, 王慧, 王利勇, 于翔舟, 杨乐 (706)

遥感应用

Landsat 长时间序列的阳澄湖湖面围网时空变化 计璐艳, 尹丹艳, 官鹏 (717)

通过训练样本采样处理改善小宗作物遥感识别精度..... 樊东东, 李强子, 王红岩, 张源, 杜鑫, 沈宇 (730)

地基激光雷达森林近地面点云精细分类与倒木提取

..... 马振宇, 庞勇, 李增元, 卢昊, 刘鲁霞, 陈博伟 (743)

随机森林回归模型的悬浮泥沙浓度遥感估算 方馨蕊, 温兆飞, 陈吉龙, 吴胜军, 黄远洋, 马茂华 (756)

太原市黑臭水体遥感识别与地面验证 李佳琦, 李家国, 朱利, 申茜, 戴华阳, 朱云芳 (773)

2017 年“8.8”九寨沟地震滑坡自动识别与空间分布特征 李强, 张景发, 罗毅, 焦其松 (785)

热红外遥感浙江地表热环境分布研究

..... 吴文渊, 金城, 庞毓雯, 赵丽佳, 宋瑜, 胡潭高, 张登荣, 徐俊锋 (796)

JOURNAL OF REMOTE SENSING

(Vol. 23 No.4 July , 2019)

CONTENTS

Scholar's View Point

- Towards more extensive and deeper application of remote sensing
..... GONG Peng (569)

Review

- Evaluation and characteristic research in diurnal surface temperature cycle in China using FY-2F data
..... MENG Xiangchen, LIU Hao, CHENG Jie (580)
- Research progress on methods of automatic coastline extraction based on remote sensing images
..... WU Yiquan, LIU Zhonglin (602)
- Progress and bibliometric analysis of remote sensing data fusion methods (1992—2018).....
..... ZHANG Lifu, PENG Mingyuan, SUN Xuejian, CEN Yi, TONG Qingxi (619)

Technology and Methodology

- Verification of BRDF archetype inversion algorithm from surface observations of airborne WIDAS
..... HE Dandan, JIAO Ziti, DONG Yadong, ZHANG Xiaoning, ZHANG Hu, DING Anxin (628)
- Sparse and low-rank abundance estimation with internal variability YUAN Jing, ZHANG Yujin (647)
- Rolling estimation model of soil moisture based on multi-satellite fusion
..... LIANG Yueji, REN Chao, HUANG Yibang, WANG Haoyu, LU Xianjian, YAN Hongbo (660)
- Parameterization and correction of hotspot parameters of Ross-Li kernel driven models on POLDER dataset
CHANG Yaxuan, JIAO Ziti, DONG Yadong, ZHANG Xiaoning, HE Dandan, YIN Siyang, CUI Lei, DING Anxin (672)
- Geo-positioning of remote sensing images with reference image
..... PAN Xuechen, JIANG Ting, YU Anzhu, WANG Xin, ZHANG Yi (684)
- Multi-feature-based classification method using random forest and superpixels for polarimetric SAR images ...
..... XU Qiao, ZHANG Xiao, YU Shaohuai, CHENG Qihao, LIU Xiuguo (694)
- Multi-source digital map tile data mashup scheme design based on Cesium.....
..... FAN Junfu, HU Taoying, HE Huixin, QIN Liu, LI Guihua (705)
- Parallel color balancing method using adaptive block Wallis algorithm for image mosaicking
..... LI Shuo, WANG Hui, WANG Liyong, YU Xiangzhou, YANG Le (715)

Remote Sensing Applications

- Temporal-spatial study on enclosure culture area in Yangcheng Lake with long-term landsat time series
..... JI Luyan, YIN Danyan, GONG Peng (729)
- Improvement in recognition accuracy of minority crops by resampling of imbalanced training datasets of remote
sensing FAN Dongdong, LI Qiangzi, WANG Hongyan, ZHANG Yuan, DU Xin, SHEN Yu (742)
- Fine classification of near-ground point cloud based on terrestrial laser scanning and detection of forest fallen
wood MA Zhenyu, PANG Yong, LI Zengyuan, LU Hao, LIU Luxia, CHEN Bowei (755)
- Remote sensing estimation of suspended sediment concentration based on Random Forest Regression Model ...
..... FANG Xinrui, WEN Zhaofei, CHEN Jilong, WU Shengjun, HUANG Yuanyang, MA Maohua (772)
- Remote sensing identification and validation of urban black and odorous water in Taiyuan city
..... LI Jiaqi, LI Jiaguo, ZHU Li, SHEN Qian, DAI Huayang, ZHU Yunfang (784)
- Recognition of earthquake-induced landslide and spatial distribution patterns triggered by the Jiuzhaigou earthquake
in August 8, 2017 LI Qiang, ZHANG Jingfa, LUO Yi, JIAO Qisong (795)
- Distribution characteristics of surface thermal environment in Zhejiang province based on thermal infrared
remote sensing
..... WU Wenyuan, JIN Cheng, PANG Yuwen, ZHAO Lijia, SONG Yu, Hu Tangao, ZHANG Dengrong, XU Junfeng (807)

随机森林回归模型的悬浮泥沙浓度遥感估算

方馨蕊^{1,2}, 温兆飞¹, 陈吉龙¹, 吴胜军^{1,2}, 黄远洋¹, 马茂华^{1,2}

1. 中国科学院重庆绿色智能技术研究院 中国科学院水库水环境重点实验室, 重庆 400714;

2. 中国科学院大学, 北京 100049

摘要: 三峡工程蓄水以来, 清水下泄, 坝下游河段发生了长时间、长距离的沿程冲刷, 河流悬浮泥沙浓度发生改变, 给沿岸生态系统带来了不利影响。随机森林算法灵活、稳健, 已被广泛应用于各类生态环境变量的回归预测分析, 但在水体悬浮泥沙浓度估算方面的能力尚未得到充分认识。基于泥沙站点监测数据和MODIS卫星遥感反射率数据, 通过构建随机森林非参数回归预测模型, 对三峡工程坝下游宜昌至城陵矶河段在建坝前后14年间(2002年—2015年)各月的悬浮泥沙浓度进行遥感估算。研究表明: (1)基于随机森林的悬浮泥沙浓度估算模型表现较好, 模型预测值与实测值间相关性好、预测精度高, 优于其他模型(线性回归、支持向量机、人工神经网络模型)。(2)在参与模型构建的MODIS波段变量中, 红波段被认为是最重要的预测变量, 但不能单独使用它进行预测, 悬浮泥沙遥感预测需要多变量共同参与。(3)将悬浮泥沙数据按季节分类所构建的随机森林模型, 其平均误差为0.46 mg/L, 平均相对均方根误差为12.33%, 估算效果最优, 能够满足较高精度下悬浮泥沙浓度估算的需求。综上, 可以考虑以季节为划分依据, 用随机森林回归模型估算悬浮泥沙浓度, 并用于后期坝下游河道悬浮泥沙浓度时空反演。

关键词: 三峡工程, 坝下游河段, 水环境遥感, 随机森林, 遥感监测, MODIS, 泥沙反演

引用格式: 方馨蕊, 温兆飞, 陈吉龙, 吴胜军, 黄远洋, 马茂华. 2019. 随机森林回归模型的悬浮泥沙浓度遥感估算. 遥感学报, 23(4): 756–772

Fang X R, Wen Z F, Chen J L, Wu S J, Huang Y Y and Ma M H. 2019. Remote sensing estimation of suspended sediment concentration based on Random Forest Regression Model. *Journal of Remote Sensing*, 23(4): 756–772 [DOI:10.11834/jrs.20197498]

1 引言

天然水体中悬浮物主要成分为泥沙, 悬浮泥沙不仅影响水体的透明度、浊度和水色等光学性质, 还影响河道、河岸带冲淤变化过程(程天文和赵楚年, 1985)。水体中的悬浮泥沙会吸附重金属、总氮、总磷等污染物(王晓青等, 2006, 2007), 影响水生生物(特别是藻类)的繁殖和生长(谢贻发等, 2007; 龚玲等, 2006)。泥沙若得不到有效处理, 会出现大量淤积, 从而加剧流域全局性水患, 影响流域生态环境(李义天等, 2004)。

三峡工程建设规模大、周期长, 建成后在防

洪、发电、航运、抗旱和供水等方面发挥着巨大效益。由于“蓄清排浊”, 原有河流径流量的季节分配过程以及与径流相关的生态环境因子发生改变(姜刘志, 2014)。同时, 坝上游来沙减少, 坝下游河道中不饱和水流则会寻求泥沙补给持续冲刷河床(张卫军等, 2013), 从而发生自上而下、长时间长距离的沿程冲刷(刘小斌等, 2006)。这些都会对坝下游范围内的生态环境带来深远影响, 如重要生物资源的栖息水域、生活习性改变, 群落组成结构、分布特征发生变化等(姜刘志, 2014)。因此, 充分了解区域水体悬浮泥沙浓度及其时空变化规律对河道治理、生物栖息地监测和保护具有

收稿日期: 2017-11-27; 预印本: 2018-04-04

基金项目: 国家自然科学基金(编号: 51779241, 41501096)

第一作者简介: 方馨蕊, 1993年生, 男, 硕士研究生, 研究方向为环境生态。E-mail: fangxinrui15@mails.ucas.ac.cn

通信作者简介: 吴胜军, 1971年生, 男, 研究员, 研究方向为地理信息工程与环境生态。E-mail: wsj@cigit.ac.cn

重要现实意义。

传统泥沙浓度的调查方法主要采用点位采样、室内分析。该方法调查速度慢、周期长,且只能获得在时间、空间分布上都很离散的少量数据点(李洪灵等,2006)。遥感技术的发展使这一状况得到了有效改观(翟伟康,2006)。很多学者利用不同的遥感数据源和遥感途径对悬浮物进行了大量的研究,均取得了不错的效果。李四海和恽才兴(2001)采用斜率法寻找NOAA AVHRR卫星遥感数据和准同步实测表层含沙量间的关系,Miller和McKee(2004)直接探求MODIS单波段反射率数据与水体总悬浮泥沙的相关性,Bowers等(2014)利用实测数据建立了波段后向散射系数与悬浮泥沙浓度间的指数模型,林承达等(2014)则采用高斯模型构建Landsat ETM+波段反射率与悬浮泥沙浓度间的关系。目前,针对水体悬浮泥沙遥感监测多采用数理统计模型,即先将水体信息与遥感信息之间的不确定性假设成一个特定的函数关系,通过转化为多元线性函数对模型参数进行估计(张文,2012)。该方法通常要求样本总体满足特定的分布特征(如正态分布),而对于水体参数广泛存在的时空自相关性并不能满足假设条件(张文,2012)。所以,该假设在理论上就存在着缺陷(张恒喜等,2002),使得模型的预测精度不高、推广性能较差、适用范围较小。

非参数模型适用范围广、可靠性强,受到很多学者青睐,如王繁等(2009)利用人工神经网络ANN(Artificial Neural Network)方法建立了表层悬浮泥沙浓度定量遥感反演模型,李致博(2012)基于支持向量机模型SVM(Support Vector Machine)构建海洋悬浮物与遥感影像间的相关关系。尽管如此,也存在一些问题:ANN虽不需要提前假设样本浓度与遥感参数间的函数关系,但其预测精度会受到“过拟合”现象的影响;SVM在使用默认参数时模型效果不佳,在使用经验参数时预测效果却不够理想,计算量大(张文,2012)。另外,上述模型建构方法复杂,内部运行机理未知,难以有效地对变量进行合理解释,使得模型预测能力有限,依然停留在探索阶段。此外,含沙水体光谱响应不仅仅取决于悬浮泥沙浓度,还与泥沙性质(颜色、类型、颗粒大小)、矿物组成、水体中叶绿

素和有色可溶性有机物等密切相关(樊辉等,2007)。因此,悬浮泥沙反演所涉及的变量数量多且种类复杂。随机森林模型RF(Random Forest)是一种基于分类树的机器学习方法(Breiman,2001)。它灵活、稳健、实用、高效,可以用于回归、聚类、分类、预测等分析。由于模型在参数优化、变量排序以及后续变量分析解释等方面优势明显,近年来被广泛应用于悬浮泥沙模拟方面。然而,这些研究存在时相单一(江佳乐等,2014)、未有效结合遥感影像(Francke等,2008; Zimmermann等,2012; Stephens和Diesing,2015)、模型精度未知(Wang等,2016)等问题,都不能实现对水体悬浮泥沙的有效监测预报。因此,随机森林模型在水体悬浮泥沙遥感监测估算方面的能力尚未得到充分认识,相应的对比参考数据也较为缺乏。

本文针对遥感影像数据在河段水质监测方面的优势,及时开展三峡工程蓄水运用前后坝下游河段悬浮泥沙浓度变化研究。通过构建随机森林模型,对三峡工程坝下游宜昌至城陵矶河段悬浮泥沙浓度进行遥感估算,以期了解该区域水体中悬浮泥沙浓度及其时空变化规律,为区域河道治理、水库调度提供参考依据。

2 材料与方法

2.1 研究区域

选择三峡工程坝下游宜昌至城陵矶河段作为研究区域(图1)。该河段位于长江中游(29°26'28"N—30°41'42"N, 111°17'30"E—113°10'51"E),全长约480 km。根据河段特性不同,分为宜昌至枝城河段和枝城至城陵矶河段(即荆江)(高志斌和段光磊,2006)。受三峡工程蓄水直接影响,河水对河道两岸土地侵蚀严重,地区生态环境、生产生活受到影响。位于其中几个重要水文控制站沿河流方向从上往下依次是宜昌站、沙市站和城陵矶站。这些水文站所处地理位置及其水文特性,基本能够代表不同河段的水文信息(如径流、来水来沙等),其监测数据可以作为悬浮泥沙遥感反演建模和模型验证的样本数据(乔晓景等,2013)。

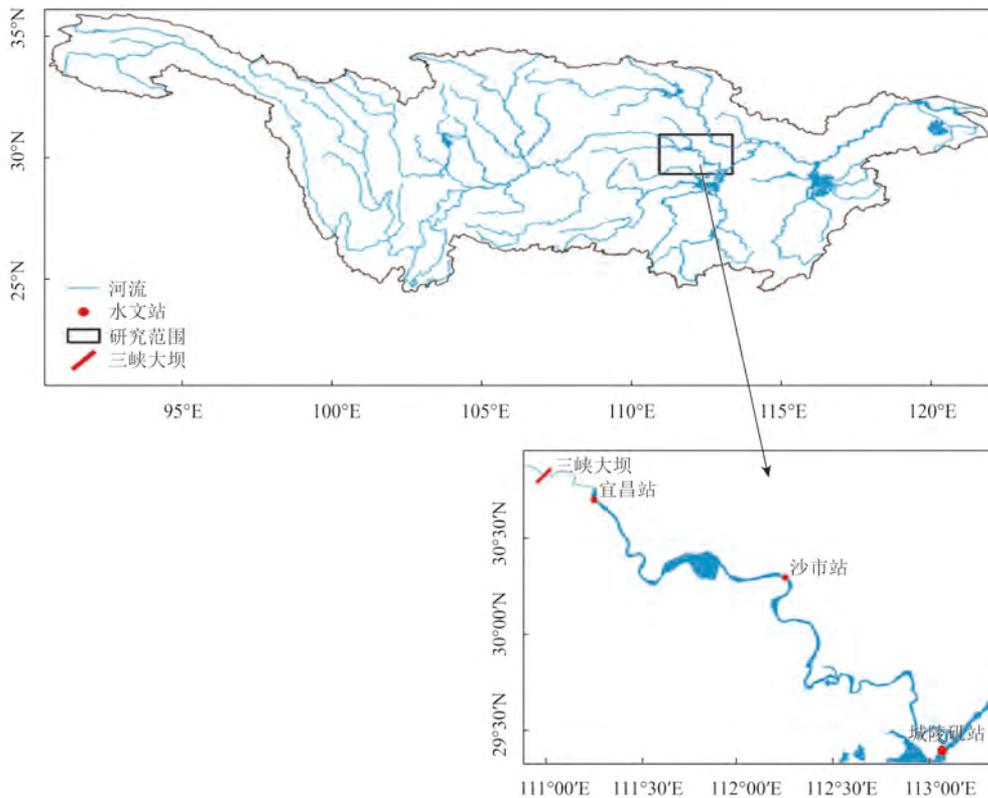


图1 研究区域

Fig. 1 The study area

2.2 研究数据

2.2.1 悬浮泥沙数据

通过收集水利部长江水利委员会发布的各年(2002年—2015年)长江泥沙公报,获得了宜昌站、沙市站、城陵矶站3个水文控制站14年所有月份的月径流量和月输沙量统计图。先利用Origin 9.0软件Digitize模块获得3个水文站各月的径流量和输沙量数值(表1),然后依据月径流量、月输沙量估算出月平均含沙量(式1)(乔晓景等,2013),一共得到504个泥沙样本浓度。

$$\rho = \frac{100m}{v} \quad (1)$$

式中, m 表示月输沙量(万吨), v 表示月径流量(亿立方米), ρ 表示月平均泥沙浓度(mg/L)。

2.2.2 遥感数据

MODIS数据具有光谱分辨率高、幅宽较大、回访周期短、时空分辨率适宜(同一地区可以得到每天至少2次白天和2次黑夜更新数据)、数据获取免费且易于批量处理等特点,能满足资源环境监测与调查的需要(丁莉东等,2006)。由于MODIS

在三峡工程坝下游地区每月都具有影像数据,能更好地对应水文站点月平均泥沙浓度。因此,MODIS是当前能进行宏观时空尺度研究的最佳数据源。

通过NASA官网下载了研究区2002年—2015年各年的MODIS Terra MOD13Q1地面反射率产品(图幅号为h27v05和h27v06)。此产品为经过大气校正的植被指数16天合成数据,投影方式采用正弦曲线投影,空间分辨率为250 m,包含蓝、红、中红外及近红外波段的地面反射率(裴雪原等,2014)。对下载的影像做如下处理:

(1)研究区域提取。利用MRT(MODIS Reprojection Tool)软件批量对每期MODIS产品中h27v05、h27v06两景影像进行镶嵌、重投影(WGS 1984 UTM Zone 49N)和裁剪,得到包含三峡工程坝下游区域在内的矩形影像(图1)。

(2)反射率信息提取。为了提取各水文站泥沙监测点对应位置的反射率,首先选择监测点周围3×3区域像元值作为潜在反射率值。由于MOD13Q1空间分辨率较低,且受不同时期河道宽度、河道中过往船只以及云雾天气的影响,区域内的9个像元中会存在陆地、云雾、船只等非水体信息,因此需要在这9个潜在像元中寻找最有可能为水体的

纯像元，并以此值作为监测样点的反射率。水体在近红外波段反射率非常低，其NDVI值为负，故某像元的NDVI值越低，则其表示水体纯像元的可能性越大。由此，先以MOD13Q1中NDVI数据为依据，确定各监测点最邻近周围3×3区域像元；然后定位9个像元中包含最低NDVI值的位置，提取该位置所有波段反射率；最后对每个月所得到的反射率值取平均，作为最终反射率值。

表1 各水文站实测月输沙量、月径流量及估算的月平均泥沙浓度(以2015年为例)

Table 1 Monthly sediment discharge, runoff and estimated sediment concentration at various hydrological stations (2015)

站点	测定时间	月输沙量/ 10 ⁴ t	月径流量/ 10 ⁸ m ³	月平均泥沙浓度/ (mg/L)
宜昌站	2015-01	5.07	175.96	2.88
	2015-02	4.10	161.81	2.54
	2015-03	7.95	216.65	3.67
	2015-04	8.91	273.51	3.26
	2015-05	17.57	350.57	5.01
	2015-06	59.61	465.15	12.82
	2015-07	179.97	557.50	32.28
	2015-08	36.52	444.94	8.21
	2015-09	43.26	536.72	8.06
	2015-10	11.90	375.97	3.16
	2015-11	7.95	247.24	3.22
	2015-12	6.03	190.10	3.17
沙市站	2015-01	28.99	175.27	16.54
	2015-02	23.10	158.35	14.59
	2015-03	46.25	211.31	21.89
	2015-04	109.61	264.28	41.48
	2015-05	92.35	324.35	28.47
	2015-06	190.43	398.89	47.74
	2015-07	357.76	485.20	73.73
	2015-08	142.37	395.95	35.96
	2015-09	257.71	468.28	55.03
	2015-10	88.62	331.46	26.74
	2015-11	40.37	223.57	18.05
	2015-12	34.48	182.87	18.86
城陵矶站	2015-01	89.80	63.41	141.63
	2015-02	38.16	72.61	52.55
	2015-03	172.44	140.00	123.17
	2015-04	232.69	176.80	131.61
	2015-05	253.64	266.57	95.15
	2015-06	322.50	455.31	70.83
	2015-07	100.99	397.86	25.38
	2015-08	103.86	189.48	54.81
	2015-09	121.94	217.08	56.17
	2015-10	112.76	171.83	65.62
	2015-11	516.74	236.48	218.52
	2015-12	405.13	229.52	176.52

2.3 模型构建

2.3.1 样本处理与潜在变量

将悬浮泥沙数据和原始波段反射率数据进行分类整理，采用迪克松准则剔除异常值(范德芹等, 2013)。为保证回归参数估计量具有良好的统计性质，提高模型精度，待求取悬浮泥沙浓度的自然对数值(以Ln(SSC)表示)后，再将它与其他波段反射率进行拟合(Smith和Croke, 2005)。为了探索不同站点、水文情势以及季节对模型的影响，除了将全部原始数据集用于建模外，还将其按照站点、水文情势(丰水期5—10月，枯水期11—次年4月)、季节(春季3—5月，夏季6—8月，秋季9—11月，冬季12—次年2月)分类，划分为不同数据子集参与建模。对于划分的各建模数据集(包括全部，丰水期、枯水期，宜昌站、沙市站、城陵矶站，春季、夏季、秋季和冬季)，分别从数据(子)集随机选择80%的样点数据作为随机森林回归模型的训练样本集，剩余的20%作为模型测试样本集，参与随机森林模型构建。

基于随机森林回归模型的悬浮泥沙遥感估算是以不同时期获得的遥感影像为主要数据源。为充分利用数据源所提供的信息，本文结合已有的相关研究结果(温兆飞, 2017)，选取了29个潜在预测变量参与随机森林回归建模(表2)。

表2 参与随机森林回归建模的29个潜在预测变量

Table 2 The 29 potential predictive variables involved in random forest regression model

变量名	获取方式(公式)	个数
Band(<i>i</i>)	B_i	4
EVI	$2.5 \times (B_4 - B_2) / (B_4 + 6.0 \times B_2 - 7.5 \times B_1 + 1)$	1
NDVI(<i>i, j</i>)	$(B_i - B_j) / (B_i + B_j)$	6
DVI(<i>i, j</i>)	$B_i - B_j$	6
RVI(<i>i, j</i>)	B_i / B_j	12

注：表中*B*表示遥感影像中波段反射率，下标*i, j*表示波段编号，对于MODIS MOD13Q1影像而言，其范围为1—4，分别对应蓝、红、中红外以及近红外波段。

29个潜在预测变量大致分为以下5类：

(1)单一波段类，比较常见，主要是利用不同遥感卫星某一波段的反射率与悬浮泥沙数据，通过构建合适的数学模型对反演情况进行研究(李四海和恽才兴, 2001；李洪灵等, 2006；乔晓景

等, 2013; 林承达等, 2014)。

(2)增强植被指数EVI(Enhanced Vegetation Index), 是Liu和Huete(1995)提出的一种参数构建方式。EVI对基础数据进行了全面的大气校正, 解决了植被指数易饱和以及与实际地表覆被缺乏线性关系等问题(王正兴等, 2003)。

(3)归一化植被指数类, 来源于归一化植被指数NDVI(Normalized Difference Vegetation Index)。NDVI是Rouse等(1974)提出的, 计算公式为: $(B_4 - B_2)/(B_4 + B_2)$ 。水体反射率在红光区高于植被, 在近红外区远低于植被。在遥感影像中, 水体的NDVI值为负, 而植被、土壤的为正。因此, 可通过选取合适的阈值增强水陆反差, 从而识别水体。

(4)差值植被指数类, 主要来源于差值植被指数DVI(Difference Vegetation Index)。DVI是Richardson和Wiegand(1977)提出的, 计算公式为 $DVI = B_4 - sB_2$, s 为土壤背景线性方程的斜率。裸地的DVI值为零, 水体为负, 植被为正。后来, Ray对DVI计算方式进行改良, 认为可以直接用近红外减去红光波段(田婷等, 2013)。DVI不仅能很好地反演植被覆盖度, 还能在一定范围内弥补大气造成的不利影响。

(5)比值植被指数类, 来源于比值植被指数RVI(Ratio Vegetation Index)。RVI由Birth和McVey(1968)提出, 计算公式为: $RVI = B_4/B_2$ 。RVI在绿色健康植被覆盖地区远大于1, 在无植被覆盖地区(裸地、水体、病虫害)趋近于1。RVI能消除土壤背景阴影, 增强土壤与植被的反射比, 可有效区分出水体与其他物体, 尤其适用于洪灾期间或薄云情况下的水体监测(王志辉和易善桢, 2007)。

2.3.2 建模过程

随机森林算法中具有两个重要的自定义参数, 分类回归树的数量 k 和分隔节点的随机变量数量 m (张雷等, 2014), 通过优化这两个参数可以提高模型预测精度(Liaw和Wiener, 2002)。本研究中, k 以50为间隔, 取值范围从10到1010; m 则以1为间隔, 取值范围涵盖1到 n (n 为预测变量的个数, 即潜在变量数29), 共同进行最优参数探索。

利用Python中scikit-learn模块进行算法建模,

并对随机森林算法得到的各变量重要性大小进行排序。变量重要性是指预测变量对预测精度贡献率的大小, 值越大表示变量越重要。在随机森林模型中, 其变量重要性为相对重要性, 总和为1。

变量按照重要性进行排序后, 需从中进行筛选, 以保证模型既有较少的预测变量又有较好的预测效果。这样不仅能使模型简单, 还有利于后续模型的解释(Ismail等, 2010)。变量选择采用后向变量选择法, 即按照重要性大小排序后, 首先将全部变量 n 都作为预测参数构建随机森林算法, 利用该模型预测测试样本集, 记录其预测精度, 此后依次减少一个变量并重复上述过程。比较不同变量组合下各模型的预测精度, 将精度最高的组合作为最优变量组, 并以此为依据, 重新构建随机森林算法模型, 检验模型预测精度。

2.3.3 模型检验

采用决定系数 R^2 、均方根误差RMSE(Root Mean Square Error)式(2)和相对均方根误差rRMSE(relative Root Mean Square Errors)式(3)(Bian等, 2013)作为评价指标, 综合评价所有回归模型的预测能力, 并绘制预测值与实测值间的1:1关系图。RMSE是最常用的衡量模型误差的统计量; rRMSE则使用百分比来表达模型误差, 其相对而言更为直观且不受原始数据取值范围的影响, 更适宜于不同数据(子)集之间的对比。

$$RMSE = \sqrt{\frac{\sum_1^n (X(i)_p - X(i))^2}{n}} \quad (2)$$

$$rRMSE = \frac{RMSE}{Mean} = \sqrt{n \sum_1^n \left(\frac{X(i)_p - X(i)}{X(i)} \right)^2} \times 100\% \quad (3)$$

式中, $X(i)$ 表示验证数据集中第 i 个($i=1, 2, 3, \dots, n$)样点的悬浮泥沙浓度的自然对数值, $X(i)_p$ 表示利用验证数据集中的第 i 个($i=1, 2, 3, \dots, n$)样点预测变量所得到的模型预测悬浮泥沙浓度的自然对数值。

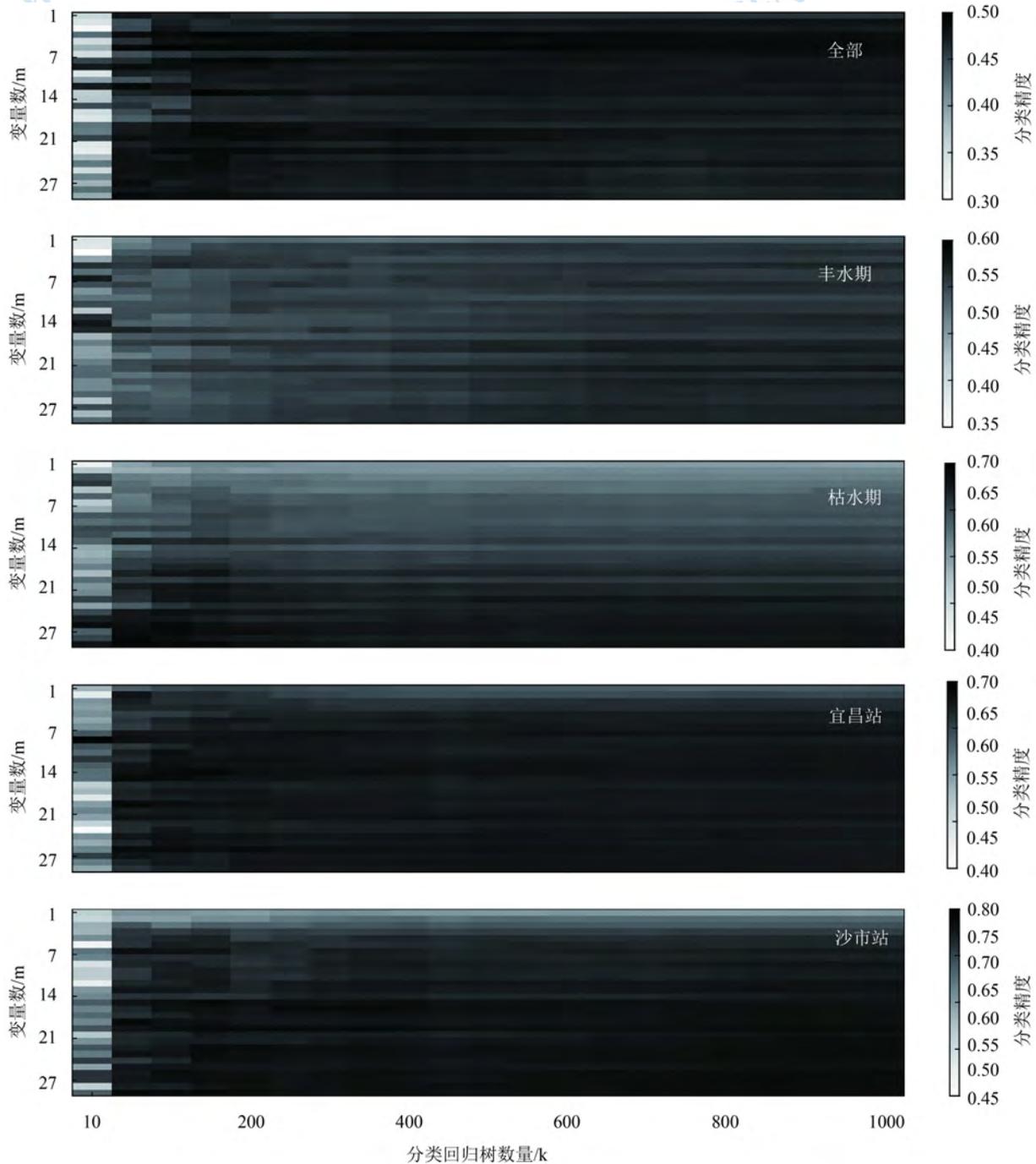
3 结果与讨论

3.1 回归模型参数优化

参与随机森林回归模型参数优化的参数是分

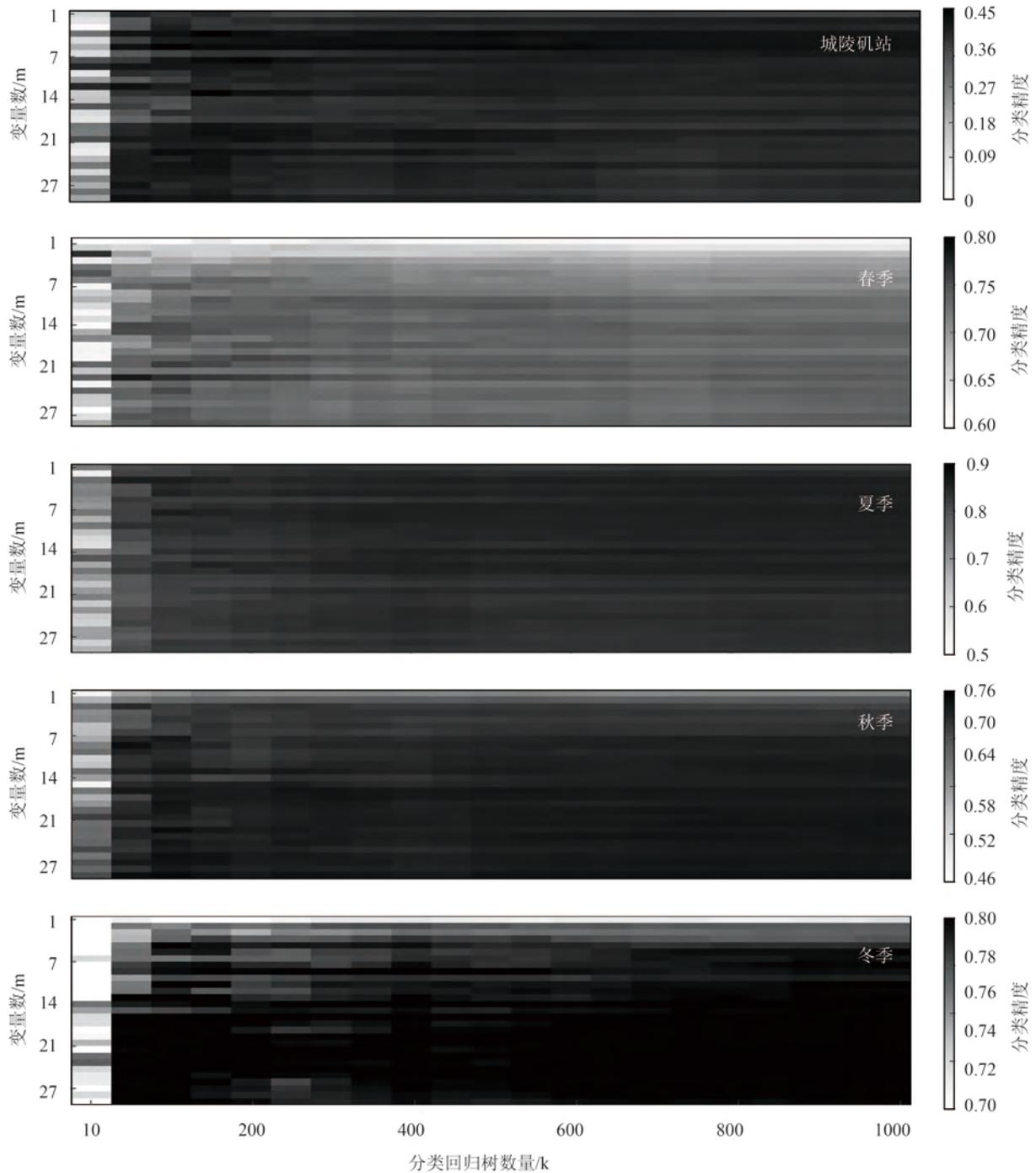
类树的数量 k 和分隔节点的随机变量数量 m ，不同参数组合所对应的分类精度如图2所示(图2只展示了当所有29个潜在预测变量参与随机森林建模运算时的参数优化结果，其中“全部”代表所有样点参与建模)。从图2中可以看出，不同的 k 和 m 组合对模型预测精度的影响较大，当 m 保持不变时， k 越

大模型的精度越高；同样，当 k 固定时， m 越大模型的精度也就越高。在图2中右下方位置可以发现较高的预测精度，综合分析数据的稳健程度，总体而言，当 $k \in [800, 1000]$ ， $m \in [15, 29]$ 时，模型的回归精度较高且变异较小。



(a) 不同数据子集下参数组合对应的分类精度 (全部、丰水期、枯水期、宜昌站、沙市站)

(a) The classification accuracy corresponding to parameter combinations under different data subsets (all, wet season, dry season, Yichang, Shashi)



(b) 不同数据子集下参数组合对应的分类精度 (城陵矶站、春季、夏季、秋季、冬季)

(b) The classification accuracy corresponding to parameter combinations under different data subsets (Chenglingji, spring, summer, autumn, winter)

图 2 随机森林回归模型不同参数下的精度评估

Fig. 2 Accuracy evaluation on random forest regression models under different parameters

3.2 回归模型变量筛选

随机森林回归模型通过评估每个变量对总体模型预测精度提高的能力, 对这些变量进行排序, 从而用于评估各变量在模型中的相对重要

性。若变量对模型预测精度提高的贡献越大, 则变量也就越重要。从图3可以发现(柱状图表示随机森林回归模型中对各预测变量的重要性排序, 折线图表示执行后向变量选择后模型的拟合

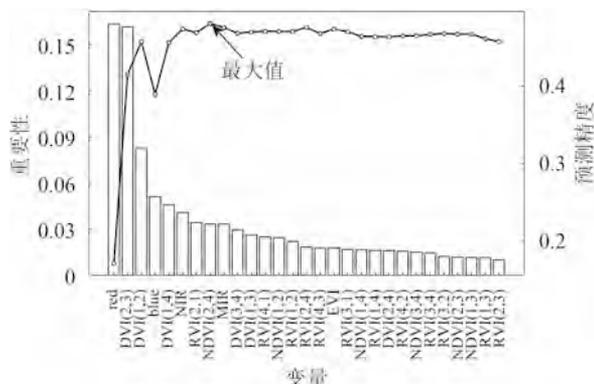
精度)：

(1)随着变量数量的变化，各子模型的预测精度会出现不同程度的波动，但最终都会逐渐归于相对稳定的状态。

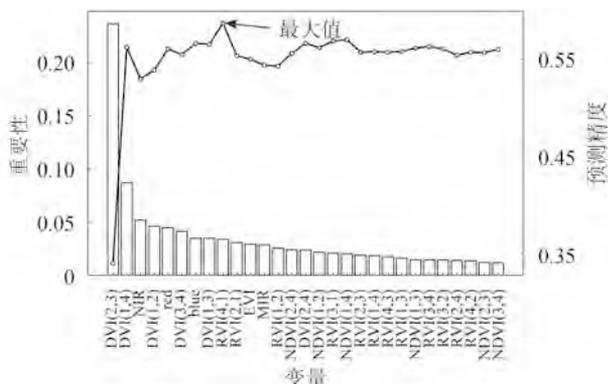
(2)当所有变量参与模型拟合时，所得预测精度不是最高的，最大模型预测精度出现在部分变量组合上；当所有悬浮泥沙样点参与建模时，其模型预测精度普遍低于其他分类模型，可以考

虑按属性(如站点、季节、水文情势)划分样点数据。

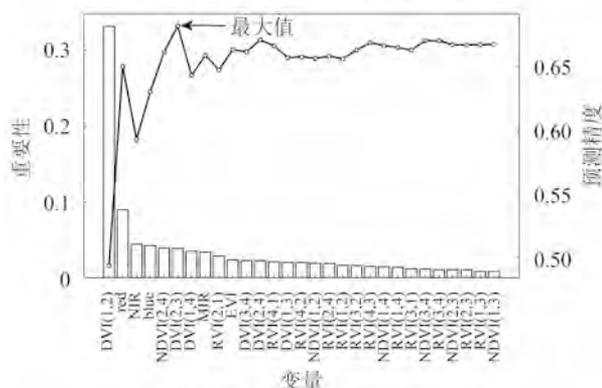
(3)从拟合效果来看，对于宜昌站、夏季、冬季数据，虽然变量数目的增加会提高模型预测精度，但提升效果不明显，而且变量过多会导致解释的困难，可见过多变量对模型精度的影响不大。因此，在重新构建模型时，以模型精度达到较大值时所具有变量组合为宜。



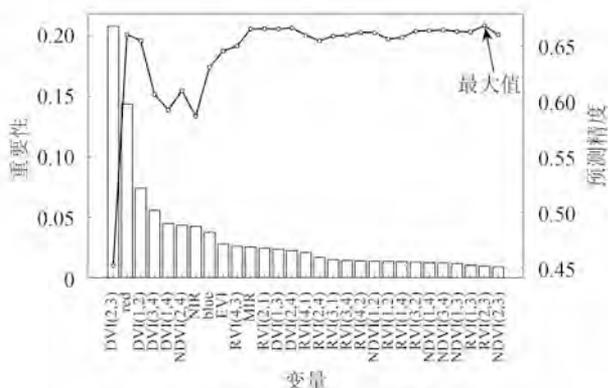
(a) Overall variable selection of suspended sediment in the downstream channel of the TGP



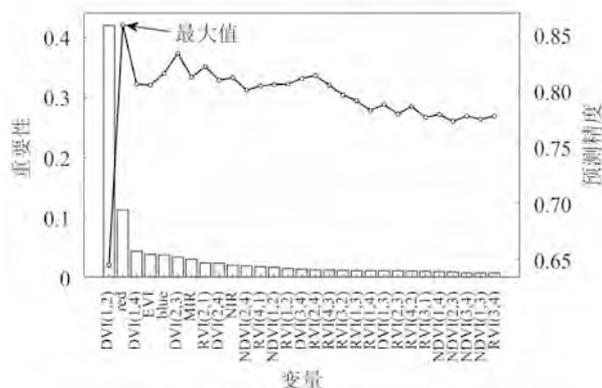
(b) Variable selection of suspended sediment in wet season



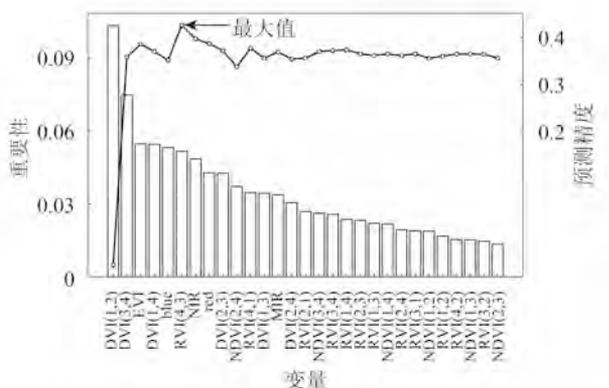
(c) Variable selection of suspended sediment in dry season



(d) Variable selection of suspended sediment in Yichang



(e) Variable selection of suspended sediment in Shashi



(f) Variable selection of suspended sediment in Chenglingji

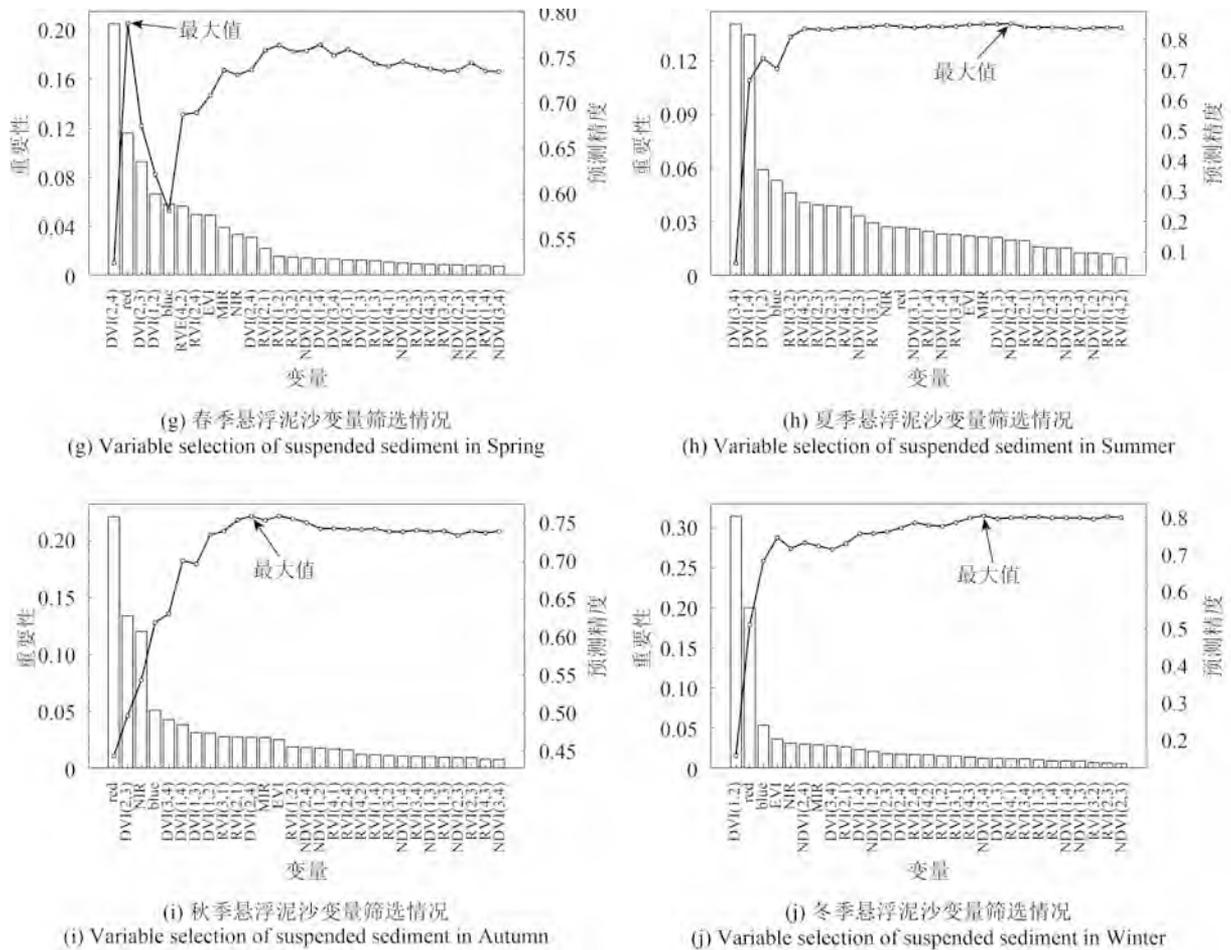


图3 随机森林回归模型变量筛选

Fig. 3 Variable selection of random forest regression model

总体上,红波段(波长范围为600—700 nm)在各模型达到最大预测精度时的变量组合中出现次数较多,可以认为是随机森林预测模型中最重要的变量,对悬浮泥沙预测精度具有重要的影响(乔晓景等, 2013; Novo等, 1991; 邬国锋等, 2009)。因为波长的改变会导致其水体反射率的变化,蓝绿波段对太阳光吸收较少,而近红外波段(波长大于600 nm)吸收能力较强,综合表现为吸收;当水体中出现泥沙颗粒时,其反射光谱曲线受泥沙散射影响会相应地发生改变,波段反射率增加,峰值在黄红波段区出现(梅安新等, 2001),这与野外采样在船中所测水体遥感反射率数据变化情况相一致(图4)。故红波段对水中悬浮泥沙反应敏感,能进行较好地识别(姜杰, 2004; 韩震, 2004)。此外,变量组合中也存有较多的复合变量,它们来源于不同波段的特定运算。表明,悬

浮泥沙估算所涉及的变量较多,会受到多种因素的影响,是一个复杂、繁琐的过程,但不能单纯使用某些特定波段(如红波段)进行估算,与乔晓景等(2013)所得出的结论并不相符。

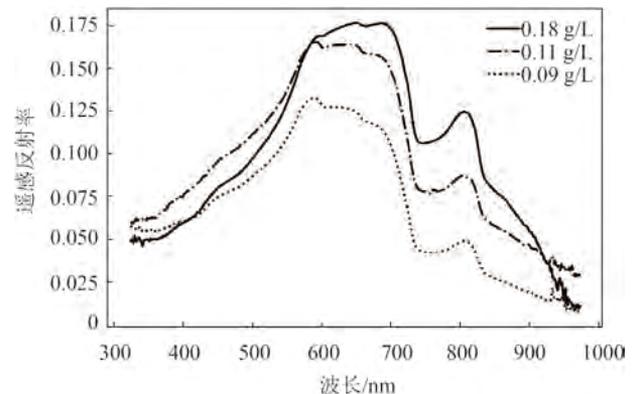


图4 不同泥沙浓度下水体反射率光谱曲线

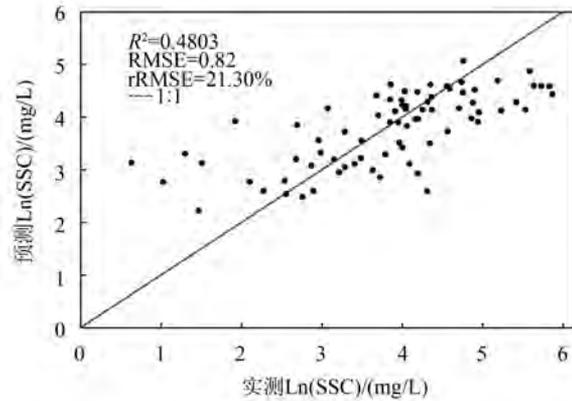
Fig. 4 Reflectance spectrum of surface waters with various SSCs

3.3 随机森林模型精度分析

以随机森林算法为基础，通过变量筛选，分别按水文情势、站点、季节建立模型训练样本集与测试样本集间的拟合关系图。图5为由随机森林算法所预测的悬浮泥沙自然对数值(预测Ln(SSC))与实际自然对数值(实测Ln(SSC))间的散

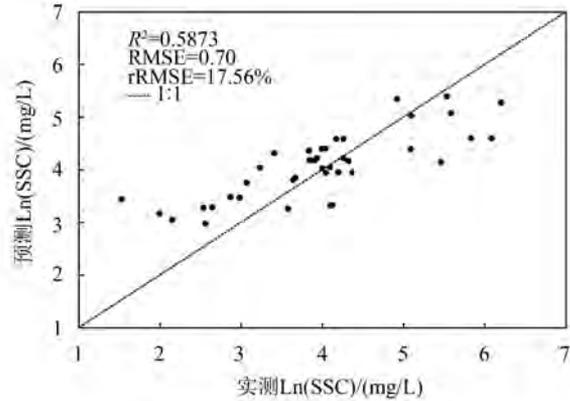
点关系图。

结果表明，由随机森林模型推算得到的预测值与实测值之间都呈比较高的相关性，RMSE和rRMSE都比较理想，说明用随机森林模型来监测悬浮泥沙浓度是可行的。但部分来看，全部数据、城陵矶站数据精度仍然较低，其决定系数不到0.5，推测原因可能包含以下几个方面：



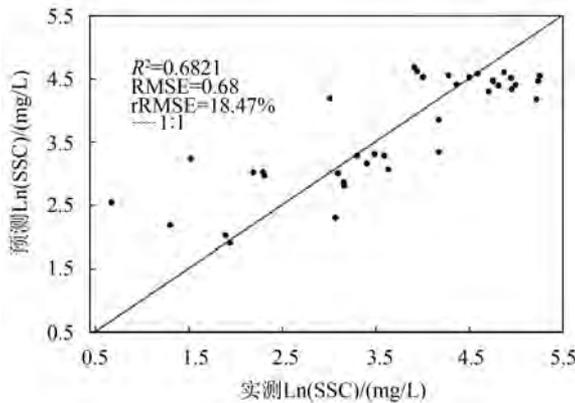
(a) 坝下游悬浮泥沙整体估算情况

(a) Estimation of suspended sediment in the downstream channel of the TGP



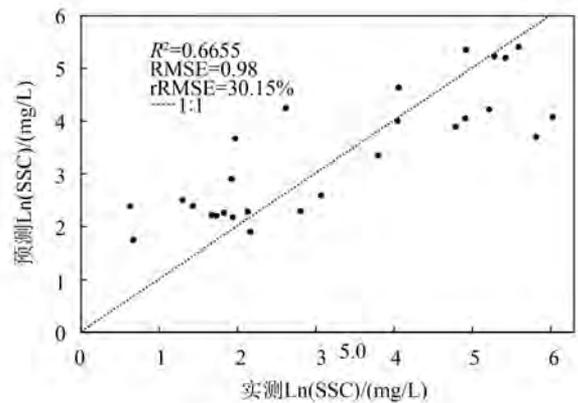
(b) 丰水期悬浮泥沙估算情况

(b) Estimation of suspended sediment in Wet season



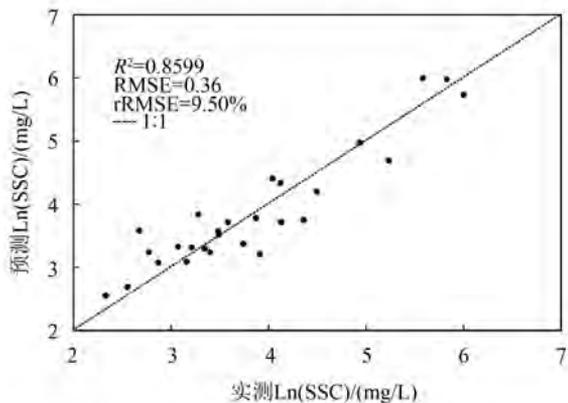
(c) 枯水期悬浮泥沙估算情况

(c) Estimation of suspended sediment in Dry season



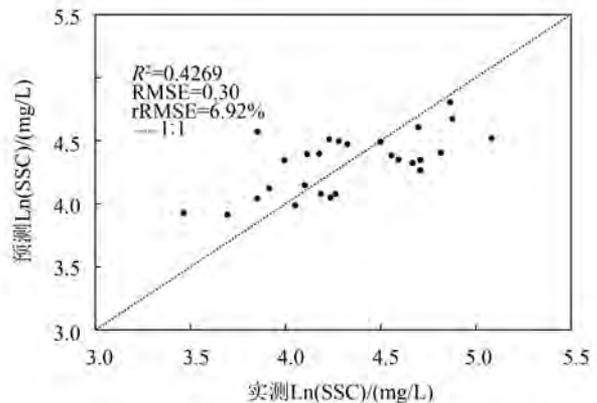
(d) 宜昌站悬浮泥沙估算情况

(d) Estimation of suspended sediment in Yichang



(e) 沙市站悬浮泥沙估算情况

(e) Estimation of suspended sediment in Shashi



(f) 城陵矶站悬浮泥沙估算情况

(f) Estimation of suspended sediment in Chenglingji

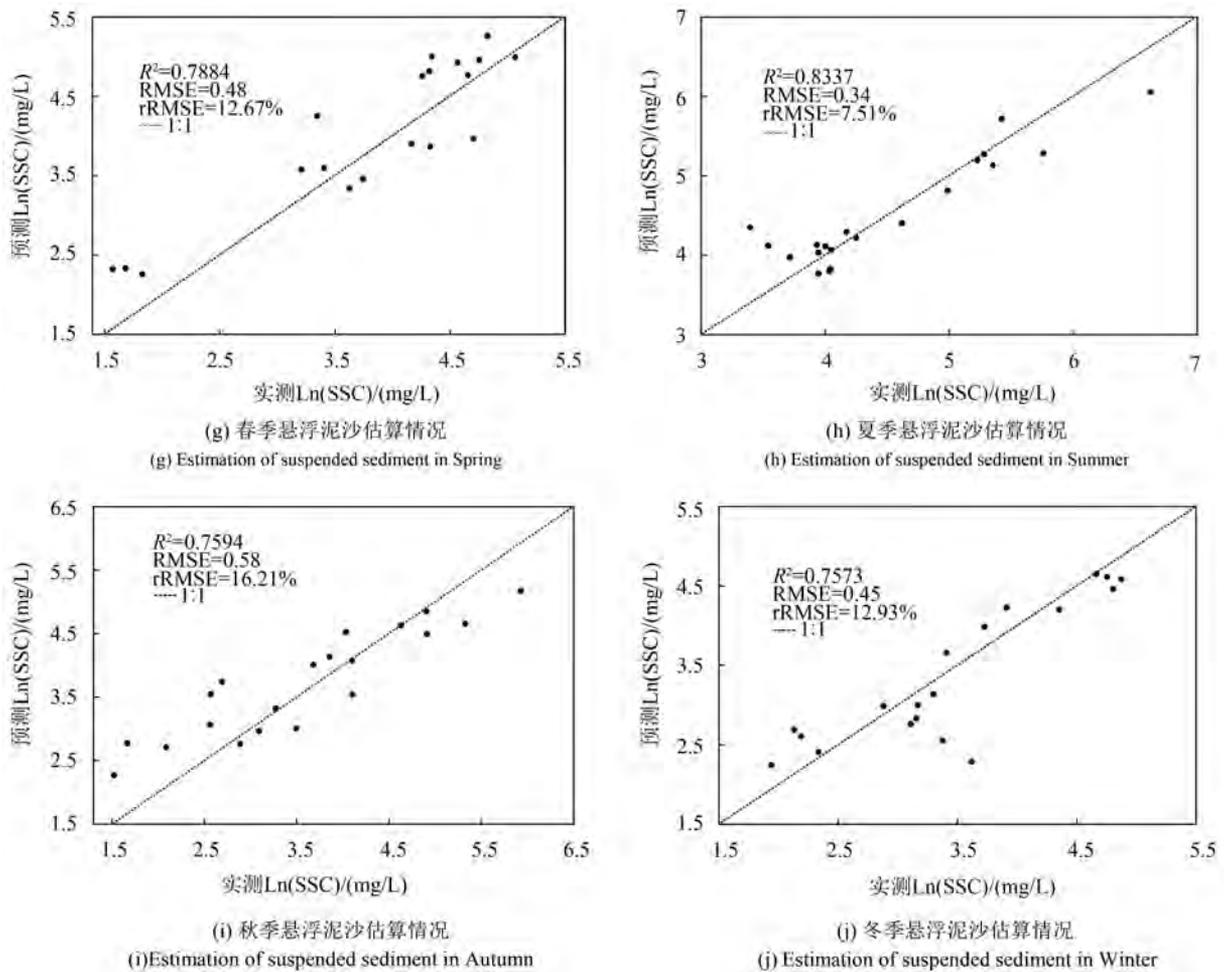


图5 随机森林回归模型精度评价

Fig. 5 Accuracy evaluation on random forest regression model

(1) 遥感卫星分辨率。MODIS卫星最高分辨率为250 m, 虽然可以反映河流水色特征信息, 但部分长江河段宽度不到1000 m, 即只有几个像元的宽度, MODIS影像在反映空间细节上的变化方面仍存在不足。为降低误差, 本研究将样点像元位置尽可能地选在距监测站点较近的河道中央, 以保证样点在较长时间序列内存在于水体当中, 但仍不能避免某些极端情况下河道出现暂时出露的可能, 也就导致了数据中存在着非河流悬浮泥沙样点。

(2) 悬浮泥沙自身性质。悬浮泥沙粒径小、不稳定, 易受河流波动影响, 区域河段悬浮泥沙浓度在短时间内也可能会有较大的变化。MODIS遥感卫星获取的则是水体瞬时信息, 它提供的波段反射信息是水体表层悬浮物的反射率, 但泥沙站点实测悬浮物浓度是各断面一定深度范围内平均的悬浮物浓度, 二者关系并不等同。

(3) 数据来源。城陵矶水文站控制着洞庭湖的总出湖水量、沙量及出湖水质状况, 其监测河段为洞庭湖出口水道, 而宜昌、沙市水文站监测河段则为长江干流三峡工程坝下游河道, 即它们所监测的悬浮泥沙在来源上就存有差异。

(4) 其他因素干扰。三峡工程坝下游河段表层悬沙分布的影响因素以径流、地形作用为主, 气象、水文等其他环境因素也会带来影响。由于三峡工程采用“冬蓄夏排”的反季节运行方案, 在它影响下的悬浮泥沙自然也会与正常季节演变时有所差异。此外, 从实地调研可知, 某些地方采砂严重, 船舶活动频繁, 不仅会影响自然水体中实际悬浮泥沙浓度, 而且也会降低由卫星所接收的水体悬浮物反射信息的质量。因此, 为使遥感监测更具科学性和客观性, 以后分析还需要综合更多的影响因素。

总体而言, 将季节作为数据分类标准时, 各

季节预测误差分别为0.48 mg/L、0.34 mg/L、0.58 mg/L、0.45 mg/L，相对均方根误差为12.67%、7.51%、16.21%、12.93%。综合比较，将悬浮泥沙数据按季节分类所划分而成的模型精度明显优于按站点、水文情势数据。它所得到的Ln(SSC)分布数据结果较为可靠，对样本点平均值附近的样本预测较好，预测的结果与实测值之间具有良好的一致性，可以作为后续悬浮泥沙时空反演的参考依据。它所包含的变量组合如表3。

表3 将季节作为分类依据时的变量组合

季节	精度最大时的变量组合
春季	NDVI(2, 4)、red
夏季	DVI(3, 4)、DVI(1, 4)、DVI(1, 2)、blue、RVI(3, 2)、RVI(4, 3)、red、DVI(2, 3)、NIR、blue、DVI(3, 4)、
秋季	DVI(1, 4)、DVI(1, 3)、DVI(1, 2)、RVI(3, 1)、RVI(2, 1)、DVI(2, 4)
冬季	DVI(1, 2)、red、blue、EVI、NIR、NDVI(2, 4)、MIR、DVI(3, 4)、RVI(2, 1)、DVI(1, 4)

3.4 随机森林模型优势分析

生态学研究，通常采用模拟响应变量与解释变量间的关系来阐述格局和过程的形成机制以及实现生态变量在空间或时间维度上的预测(张雷等, 2014)。为此，线性回归模型应运而生，但其应用范围有限，只能处理线性关系。一般情况下，各水环境参数因受复杂环境条件的影响，往往呈现复杂的、高维的非线性关系，甚至还存在共线的可能。所以，线性回归模型在预测上就存在着不足。现阶段，很多学者开始转向机器学习方面的研究，如人工神经网络模型、支持向量机模型、随机森林模型等(张雷等, 2014; 楼琇林和黄韦良, 2003; 梁栋等, 2013; 王茵茵等, 2016)。

本次研究在应用随机森林回归模型前，也尝试过其他模型，拟合效果如表4所示。在各个模型中，虽然在各属性分类中存有个别拟合精度较高的情况，但从整体效果来看，将季节作为分类属性时拟合的效果整体是最优的。以线性回归模型为例，全部数据参与模型回归时的拟合精度： R^2 为0.52，RMSE为0.73，rRMSE为18.76%(简写为(0.52, 0.73, 18.76%))，低于按水文情势(0.58, 0.74, 19.54%)(为各属性分类的平均值，下同)、站点(0.60, 0.60, 17.15%)、季节(0.70, 0.62, 15.31%)

分类。故，按季节分类是线性回归中最优的，而其他模型也满足这一发现；再比较各模型中按季节分类时的效果，随机森林(0.78, 0.62, 12.33%)、支持向量机(0.55, 0.59, 15.18%)、人工神经网络(0.63, 0.66, 17.02%)。可见，随机森林模型中，不论在决定系数，还是在拟合效果都是所有模型中最优的。

表4 各模型拟合效果

Table 4 The fitting effect of each model

模型类型	划分依据	n^*	R^2	RMSE	rRMSE	
线性回归	全部	380	0.5212	0.73	18.76%	
	丰水期	198	0.5293	0.75	18.76%	
	枯水期	189	0.6303	0.72	20.32%	
	宜昌站	133	0.6411	0.90	30.31%	
	沙市站	135	0.7506	0.53	12.50%	
	城陵矶站	132	0.4029	0.37	8.63%	
	春季	93	0.7711	0.51	12.15%	
	夏季	96	0.7423	0.67	14.85%	
	秋季	100	0.5809	0.66	16.69%	
	冬季	94	0.6717	0.64	17.55%	
	随机森林	全部	380	0.4803	0.82	21.30%
		丰水期	198	0.5873	0.70	17.56%
枯水期		189	0.6821	0.68	18.47%	
宜昌站		133	0.6655	0.98	30.15%	
沙市站		135	0.8599	0.36	9.50%	
城陵矶站		132	0.4269	0.30	6.92%	
春季		93	0.7884	0.48	12.67%	
夏季		96	0.8337	0.34	7.51%	
秋季		100	0.7594	0.58	16.21%	
冬季		94	0.7573	0.45	12.93%	
支持向量机		全部	380	0.3405	0.83	20.95%
		丰水期	198	0.3462	0.78	19.17%
	枯水期	189	0.4775	0.70	18.20%	
	宜昌站	133	0.5364	1.17	36.73%	
	沙市站	135	0.5321	0.60	14.93%	
	城陵矶站	132	0.2946	0.37	8.79%	
	春季	93	0.605	0.66	16.95%	
	夏季	96	0.6274	0.63	14.34%	
	秋季	100	0.4013	0.52	13.51%	
	冬季	94	0.5533	0.56	15.93%	
	人工神经网络	全部	380	0.4722	0.85	22.18%
		丰水期	198	0.4629	0.80	20.04%
枯水期		189	0.4361	0.88	23.30%	
宜昌站		133	0.5244	1.07	32.54%	
沙市站		135	0.6293	0.60	14.76%	
城陵矶站		132	0.3478	0.42	9.89%	
春季		93	0.6113	0.83	21.06%	
夏季		96	0.6886	0.61	13.82%	
秋季		100	0.6115	0.61	16.01%	
冬季		94	0.6217	0.60	17.20%	

注： n^* 为剔除异常值后的样点数目。

此外,支持向量机、人工神经网络模型在应用上也存在不足。由于是黑箱模型,研究者无法了解模型内部的实际运行机理,也无法直观地知道预测变量总体的重要性大小(王强等,2006)。而且,还存有诸多使用上的限制,如“过拟合”现象(金龙等,2004)、计算量大等问题。虽然随机森林回归也是黑箱模型,但它提供了其他有效方式协助解释,比如各变量对模型预测的重要性。相对于支持向量机、人工神经网络模型,随机森林更易于解释。所以,也有研究者倾向于把它定为灰箱模型(Prasad等,2006)。此外,随机森林算法中两个随机参数(k , m)的引入,使其具有较好的抗噪能力,不易陷入过度拟合。因此,在三峡工程坝下游河道悬浮泥沙浓度遥感估算中,可以考虑引入以季节为划分依据的随机森林模型。尽管如此,随机森林算法也不可能做到尽善尽美,因为它所构建的模型是非参数的。在应用方面还存在一定局限,在后期的尝试探索中还需进一步研究。

4 结论

机器学习是遥感影像处理的热点问题和核心问题。在诸多机器学习算法中,随机森林因其高效、准确而备受关注,是目前最好的机器学习算法之一,并广泛应用于各行各业。三峡工程蓄水以来,坝下游河流,特别是宜昌至城陵矶河段,泥沙问题日益严重。水体悬浮物是评价水质的重要参数,通过卫星遥感监测研究,可以为宏观了解三峡工程坝下游河段悬浮泥沙变化规律提供有利途径。此时,随机森林算法的优势日益凸显,成为估算水体悬浮泥沙浓度的有效方法之一。本文基于泥沙站点监测数据和卫星遥感反射率数据,通过构建随机森林非参数模型,结合及比较其他回归模型,对三峡工程坝下游宜昌至城陵矶河段悬浮泥沙浓度进行遥感估算,得到了如下结论:

(1)线性回归模型应用范围有限,只能处理线性关系,难以适用于复杂多变的水环境参数遥感。支持向量机、人工神经网络模型是黑箱模型,无法明确模型内部的实际运行机理,也无法直观了解预测变量总体的重要性大小,还存有“过拟合”现象、计算量大等诸多使用上的限制,在应用上存在不足。随机森林模型灵活、稳健、简单、方便,在参数优化、变量排序以及后续变量

分析解释等方面优势明显,且相关系数和预测精度明显优于线性模型、支持向量机模型、神经网络模型等其他模型,更适用于坝下游水体悬浮物监测遥感估算。

(2)由于泥沙的散射作用,波段反射率会随着波长的增加而出现变化,但峰值始终出现在黄红波段区,且能发现较为明显的区分,与野外实测数据和研究结果变化情况相一致。研究表明,MODIS中,红波段是随机森林模型中对悬浮泥沙浓度具有重要意义的预测变量。同时,在研究结果的变量组合中也存有较多的复合变量,它们来源于不同波段的特定运算。可见,悬浮泥沙估算所涉及的变量较多,会受到多种因素的影响。因此,不能单纯使用某一波段(如红波段),悬浮泥沙遥感估算需要多变量共同参与。

(3)由于数据来源、卫星分辨率、悬浮泥沙自身特性等诸多限制因素,样点全部参与随机森林时回归精度较低。将悬浮泥沙数据按季节分类所构建的随机森林模型,其平均误差为0.46 mg/L,平均相对均方根误差为12.33%,估算效果最优,能够满足较高精度下悬浮泥沙浓度估算的需求。为最大限度地保证模型精度,充分利用数据,可以考虑以季节为划分依据,通过与MODIS影像中波段反射率建立随机森林模型估算悬浮泥沙浓度,甚至反演时空范围内悬浮泥沙浓度。

参考文献(References)

- Bian M, Skidmore A K, Schlerf M, Wang T J, Liu Y F, Zeng R and Fei T. 2013. Predicting foliar biochemistry of tea (*Camellia sinensis*) using reflectance spectra measured at powder, leaf and canopy levels. *ISPRS Journal of Photogrammetry and Remote Sensing*, 78: 148–156 [DOI: 10.1016/j.isprsjprs.2013.02.002]
- Birth G S and McVey G R. 1968. Measuring the color of growing turf with a reflectance spectrophotometer. *Agronomy Journal*, 60(6): 640–643 [DOI: 10.2134/agronj1968.00021962006000060016x]
- Bowers D G, Hill P S and Braithwaite K M. 2014. The effect of particulate organic content on the remote sensing of marine suspended sediments. *Remote Sensing of Environment*, 144: 172–178 [DOI: 10.1016/j.rse.2014.01.005]
- Breiman L. 2001. Random forests. *Machine Learning*, 45(1): 5–32 [DOI: 10.1023/A:1010933404324]
- Cheng T W and Zhao C N. 1985. The runoff, sediment discharge and its impacts on the coastal waters of China's major rivers. *Acta*

- Oceanologica Sinica, 7(4): 460–471 (程天文, 赵楚年. 1985. 我国主要河流入海径流量、输沙量及对沿岸的影响. 海洋学报, 7(4): 460–471)
- Ding L D, Wu H, Wang C J, Qin Z H and Zhang Q X. 2006. Quick recognition and mapping of Lake Water information based on MODIS image. Hydrographic Surveying and Charting, 26(6): 31–34 (丁莉东, 吴昊, 王长健, 覃志豪, 章其祥. 2006. MODIS图像湖泊水体信息的快速识别与制图. 海洋测绘, 26(6): 31–34) [DOI: 10.3969/j.issn.1671-3044.2006.06.009]
- Fan D Q, Zhu W Q, Pan Y Z and Jiang N. 2013. Noise detection for NDVI time series based on Dixon's test and application in data reconstruction. Journal of Remote Sensing, 17(5): 1158–1174 (范德芹, 朱文泉, 潘耀忠, 姜楠. 2013. 基于狄克松检验的NDVI时序数据噪声检测及其在数据重建中的应用. 遥感学报, 17(5): 1158–1174) [DOI: 10.11834/jrs.20132274]
- Fan H, Huang H J and Tang J W. 2007. Spectral signature of waters in Huanghe estuary and estimation of suspended sediment concentration from remote sensing data. Geomatics and Information Science of Wuhan University, 32(7): 601–604 (樊辉, 黄海军, 唐军武. 2007. 黄河口水体光谱特性及悬沙浓度遥感估测. 武汉大学学报(信息科学版), 32(7): 601–604) [DOI: 10.3969/j.issn.1671-8860.2007.07.009]
- Francke T, López-Tarazón J A and Schröder B. 2008. Estimation of suspended sediment concentration and yield using linear models, random forests and quantile regression forests. Hydrological Processes, 22(25): 4892–4904 [DOI: 10.1002/hyp.7110]
- Gao Z B and Duan G L. 2006. Influence of boundary condition on downstream river bed evolution of TGP. Yangtze River, 37(12): 92–94 (高志斌, 段光磊. 2006. 边界条件对三峡坝下游河床演变影响. 人民长江, 37(12): 92–94) [DOI: 10.3969/j.issn.1001-4179.2006.12.032]
- Gong L, Zhong C H and Deng C G. 2006. Effect of suspended sand particles on growth of algae in aquatic system. Journal of Agro-Environment Science, 25(S2): 687–689 (龚玲, 钟成华, 邓春光. 2006. 水体中悬浮泥沙对藻类生长的影响. 农业环境科学学报, 25(S2): 687–689)
- Han Z. 2004. Remote Sensing Information Extraction and Quantitative Inversion Research of Silt Tidal Flat and Suspended Sediment of Case II Waters in Coast Zone. Shanghai: East China Normal University (韩震. 2004. 海岸带淤泥质潮滩和II类水体悬浮泥沙遥感信息提取与定量反演研究. 上海: 华东师范大学) [DOI: 10.7666/d.y609946]
- Ismail R, Mutanga O and Kumar L. 2010. Modeling the potential distribution of pine forests susceptible to *Sirex noctilio* infestations in Mpumalanga, South Africa. Transactions in GIS, 14(5): 709–726 [DOI: 10.1111/j.1467-9671.2010.01229.x]
- Jiang J. 2004. A Study on Retrieval and Change Analysis of Suspended Sediment Concentration. Nanjing: Nanjing Normal University (姜杰. 2004. 悬浮泥沙浓度遥感反演模式研究. 南京: 南京师范大学) [DOI: 10.7666/d.y642650]
- Jiang J L, Liu X N, Liu M L and Bi X Q. 2014. Remote sensing retrieval model of sea surface salinity in Hong Kong waters based on the random forest. Marine Science Bulletin, 33(3): 333–341 (江佳乐, 刘湘南, 刘美玲, 毕晓庆. 2014. 基于随机森林的香港海域海表盐度遥感反演模型. 海洋通报, 33(3): 333–341) [DOI: 10.11840/j.issn.1001-6392.2014.03.013]
- Jiang L Z. 2014. Assessment of Hydrological Alteration and its Impacts on Fish Population in the Middle-Lower Yangtze River after the Impoundment of the Three Gorges Dam. Beijing: University of Chinese Academy of Sciences (姜刘志. 2014. 三峡蓄水后长江中下游水文情势变化特征及其对鱼类的影响研究. 北京: 中国科学院大学)
- Jin L, Kuang X Y, Huang H H, Qin Z N and Wang Y H. 2004. Study on the overfitting of the artificial neural network forecasting model. Acta Meteorologica Sinica, 62(1): 62–70 (金龙, 况雪源, 黄海洪, 覃志年, 王业宏. 2004. 神经网络预报模型的过拟合研究. 气象学报, 62(1): 62–70) [DOI: 10.3321/j.issn:0577-6619.2004.01.007]
- Li H L, Zhang Y and Jiang J. 2006. Study on the inversion model for the suspended sediment concentration in remote sensing technology. Advances in Water Science, 17(2): 242–245 (李洪灵, 张鹰, 姜杰. 2006. 基于遥感方法反演悬浮泥沙分布. 水科学进展, 17(2): 242–245) [DOI: 10.3321/j.issn:1001-6791.2006.02.015]
- Li S H and Yun C X. 2001. A study on the quantitative model of the suspended sediment concentration from the meteorological satellite imagery. Journal of Remote Sensing, 5(2): 154–160 (李四海, 恽才兴. 2001. 河口表层悬浮泥沙气象卫星遥感定量模式研究. 遥感学报, 5(2): 154–160) [DOI: 10.3321/j.issn:1007-4619.2001.02.015]
- Li Y T, Sun Z H, Deng J Y and Zhang W. 2004. Variation of sediment transport and flood disaster in the middle reach of Yangtze River. Journal of Sediment Research(2): 33–39 (李义天, 孙昭华, 邓金运, 张为. 2004. 泥沙输移变化与长江中游水患. 泥沙研究(2): 33–39) [DOI: 10.3321/j.issn:0468-155X.2004.02.006]
- Li Z B. 2012. Based Support Vector Machine Retrieval Model for Ocean Suspended Solids Remote Sensing Concentration. Beijing: China University of Geosciences (Beijing) (李致博. 2012. 基于支持向量机的海洋悬浮物浓度遥感反演模型研究. 北京: 中国地质大学(北京))
- Liang D, Guan Q S, Huang W J, Huang L S and Yang G J. 2013. Remote sensing inversion of leaf area index based on support vector machine regression in winter wheat. Transactions of the Chinese

- Society of Agricultural Engineering, 29(7): 117–123 (梁栋, 管青松, 黄文江, 黄林生, 杨贵军. 2013. 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, 29(7): 117–123) [DOI: 10.3969/j.issn.1002-6819.2013.07.015]
- Liaw A and Wiener M. 2002. Classification and regression by random-forest. *R News*, 2-3: 18–22
- Lin C D, Zhou B, Ma Q and Jiang M X. 2014. Study on the inversion model of the suspended sediment in the middle Yangtze River based on remote sensing technology. *Resources and Environment in the Yangtze Basin*, 23(8): 1119–1124 (林承达, 周斌, 马琪, 姜萌薪. 2014. 基于遥感反演长江中游地区悬浮泥沙研究. 长江流域资源与环境, 23(8): 1119–1124) [DOI: 10.11870/cjlyzyyhj.201408011]
- Liu H Q and Huete A. 1995. A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE Transactions on Geoscience and Remote Sensing*, 33(2): 457–465 [DOI: 10.1109/36.377946]
- Liu X B, Lu J Y and Lin M S. 2006. The influence of the Three Gorges Project on the middle and lower reach of the Yangtze River//Proceedings of 2006 International Symposium on Hydropower. Kunming: China Dam Committee of China Hydropower Engineering Society, China Water Conservancy Society (刘小斌, 卢金友, 林木松. 2006. 三峡工程对长江中下游河道影响分析//水电2006国际研讨会论文集. 昆明: 中国水利学会中国水力发电工程学会中国大坝委员会)
- Lou X L and Huang W G. 2003. An artificial Neural Network Method for detecting red tides with NOAA AVHRR imagery. *Journal of Remote Sensing*, 7(2): 125–130 (楼琇林, 黄韦良. 2003. 基于人工神经网络的赤潮卫星遥感方法研究. 遥感学报, 7(2): 125–130) [DOI: 10.3321/j.issn:1007-4619.2003.02.008]
- Mei A X, Peng W L, Qin Q M and Liu H P. 2001. An Introduction to Remote Sensing. Beijing: Higher Education Press: 236–238 (梅安新, 彭望球, 秦其明, 刘慧平. 2001. 遥感导论. 北京: 高等教育出版社: 236–238)
- Miller R L and McKee B A. 2004. Using MODIS Terra 250 m imagery to map concentrations of total suspended matter in coastal waters. *Remote Sensing of Environment*, 93(1/2): 259–266 [DOI: 10.1016/j.rse.2004.07.012]
- Novo E M L M, Steffen C A and Braga C Z F. 1991. Results of a laboratory experiment relating spectral reflectance to total suspended solids. *Remote Sensing of Environment*, 36(1): 61–72 [DOI: 10.1016/0034-4257(91)90031-Z]
- Pei X Y, Zang S Y and Na X D. 2014. The introductions and rapid pretreatment of MODIS MOD13Q1 Vegetation product. *Natural Science Journal of Harbin Normal University*, 30(2): 65–67, 77 (裴雪原, 臧淑英, 那晓东. 2014. MODIS MOD13Q1植被产品介绍及快速预处理. 哈尔滨师范大学自然科学学报, 30(2): 65–67, 77) [DOI: 10.3969/j.issn.1000-5617.2014.02.017]
- Prasad A M, Iverson L R and Liaw A. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems*, 9(2): 181–199 [DOI: 10.1007/s10021-005-0054-1]
- Qiao X J, He B Y, Zhang W, Li Y Z and Su Z H. 2013. MODIS-based retrieval and change analysis of suspended sediment concentration in middle Yangtze River. *Resources and Environment in the Yangtze Basin*, 22(8): 1090–1095 (乔晓景, 何报寅, 张文, 李元征, 苏振华. 2013. 基于MODIS的长江中游河段悬浮泥沙浓度反演. 长江流域资源与环境, 22(8): 1090–1095)
- Richardson A J and Wiegand C L. 1977. Distinguishing vegetation from soil background information. *Photogrammetric Engineering and Remote Sensing*, 43(12): 1541–1552
- Rouse J W Jr, Haas R H, Schell J A and Deering D W. 1974. Monitoring Vegetation Systems in the Great Plains with ERTS. PAPER-A20. NASA Special Publication: 309
- Smith C and Croke B. 2005. Sources of uncertainty in estimating suspended sediment load//Proceedings of Symposium S1 Held During the Seventh IAHS Scientific Assembly. Foz do Igaussu, Brazil: IAHS-AISH Publication: 136–143
- Stephens D and Diesing M. 2015. Towards quantitative spatial models of seabed sediment composition. *PLoS One*, 10(11): e0142502 [DOI: 10.1371/journal.pone.0142502]
- Tian T, Sun C M, Liu T, Guo D D, Wang L J and Chen Y Y. 2013. Hyperspectral remote sensing and its applications in grassland and vegetation. *Journal of Anhui Agricultural Sciences*, 41(7): 3192–3195 (田婷, 孙成明, 刘涛, 郭斗斗, 王力坚, 陈瑛瑛. 2013. 高光谱遥感技术及其在草地及植被中的应用. 安徽农业科学, 41(7): 3192–3195) [DOI: 10.3969/j.issn.0517-6611.2013.07.139]
- Wang F, Ling Z Y, Zhou B, Song L S and Wang X. 2009. MODIS images monitoring short-period variation of estuary surface water suspended sediment concentration. *Journal of Zhejiang University (Engineering Science)*, 43(4): 755–759 (王繁, 凌在盈, 周斌, 宋立松, 王新. 2009. MODIS监测河口水体悬浮泥沙质量浓度的短期变异. 浙江大学学报(工学版), 43(4): 755–759) [DOI: 10.3785/j.issn.1008-973X.2009.04.028]
- Wang Q, Shen Y P and Chen Y W. 2006. Rule extraction from support vector machines. *Journal of National University of Defense Technology*, 28(2): 106–110 (王强, 沈永平, 陈英武. 2006. 支持向量机规则提取. 国防科技大学学报, 28(2): 106–110) [DOI: 10.3969/j.issn.1001-2486.2006.02.024]
- Wang W, Yang X, Liu G, Zhou H, Ma W, Yu Y and Li Z. 2016. Random forest classification of sediments on exposed intertidal flats using ALOS-2 Quad-Polarimetric SAR data. *ISPRS-International*

- Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLI-B8: 1191-1194 [DOI: 10.5194/isprsarchives-XLI-B8-1191-2016]
- Wang X Q, Li Z, Lv P Y and Guo J S. 2007. Adsorption and desorption of phosphorus on suspended particles in the Three Gorges Area. *Resources and Environment in the Yangtze Basin*, 16(1): 31-36 (王晓青, 李哲, 吕平毓, 郭劲松. 2007. 三峡库区悬移质泥沙对磷污染物的吸附解吸特性. *长江流域资源与环境*, 16(1): 31-36) [DOI: 10.3969/j.issn.1004-8227.2007.01.007]
- Wang X Q, Lv P Y and Hu C S. 2006. Influence of suspended sediment in TGP reservoir on absorption of TP, TN etc. *Yangtze River*, 37(7): 15-17 (王晓青, 吕平毓, 胡长霜. 2006. 三峡库区悬移质泥沙对TP、TN等的吸附影响. *人民长江*, 37(7): 15-17) [DOI: 10.3969/j.issn.1001-4179.2006.07.006]
- Wang Y Y, Qi Y B, Chen Y and Xie F. 2016. Prediction of soil organic matter based on multi-resolution remote sensing data and random forest algorithm. *Acta Pedologica Sinica*, 53(2): 342-354 (王茵茵, 齐雁冰, 陈洋, 解飞. 2016. 基于多分辨率遥感数据与随机森林算法的土壤有机质预测研究. *土壤学报*, 53(2): 342-354) [DOI: 10.11766/trxb201508170308]
- Wang Z H and Yi S Z. 2007. Comparison and research on the different index models used in water extraction by remote sensing. *Science Technology and Engineering*, 7(4): 534-537 (王志辉, 易善楨. 2007. 不同指数模型法在水体遥感提取中的比较研究. *科学技术与工程*, 7(4): 534-537) [DOI: 10.3969/j.issn.1671-1815.2007.04.028]
- Wang Z X, Liu C and Huete A. 2003. From AVHRR-NDVI to MODIS-EVI: advances in vegetation index research. *Acta Ecologica Sinica*, 23(5): 979-987 (王正兴, 刘闯, Huete A. 2003. 植被指数研究进展: 从AVHRR-NDVI到MODIS-EVI. *生态学报*, 23(5): 979-987) [DOI: 10.3321/j.issn.1000-0933.2003.05.020]
- Wen Z F. 2017. Spatial and Seasonal Patterns of Aboveground Net Primary Productivity and Their Responses to Environmental Factors in the Drawdown Zone of the Three Gorges Reservoir, China. Chongqing: Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences (温兆飞. 2017. 三峡水库消落带地上净初级生产力时空变化及其影响因素分析. 重庆: 中国科学院大学(中国科学院重庆绿色智能技术研究院))
- Wu G F, Cui L J and Ji W T. 2009. Time-series MODIS images-based retrieval and change analysis of suspended sediment concentration during flood period in Lake Poyang. *Journal of Lake Sciences*, 21(2): 288-297 (鄢国锋, 崔丽娟, 纪伟涛. 2009. 基于时间序列MODIS影像的鄱阳湖丰水期悬浮泥沙浓度反演及变化. *湖泊科学*, 21(2): 288-297) [DOI: 10.18307/2009.02.19]
- Xie Y F, Hu Y H, Liu Z W and Xie G S. 2007. Effects of sediment resuspension on the growth of submerged plants. *Acta Scientiae Circumstantiae*, 27(1): 18-22 (谢贻发, 胡耀辉, 刘正文, 谢贵水. 2007. 沉积物再悬浮对沉水植物生长的影响研究. *环境科学学报*, 27(1): 18-22) [DOI: 10.3321/j.issn.0253-2468.2007.01.004]
- Zhai W K. 2006. Atmospheric Correction of MODIS and Analysis Space-Time Distribution Characteristics for Ocean Color in Bohai Sea. Dalian: Dalian Maritime University (翟伟康. 2006. MODIS大气校正及渤海水色时空分布特征研究. 大连: 大连海事大学) [DOI: 10.7666/d.y855985]
- Zhang H X, Guo J L, Zhu J Y and Yu J F. 2002. *Multivariate Data Analysis Methods and Applications with Few Observations*. Xi'an: Northwestern Polytechnical University Press: 153-154 (张恒喜, 郭基联, 朱家元, 虞健飞. 2002. 小样本多元数据分析方法及应用. 西安: 西北工业大学出版社: 153-154)
- Zhang L, Wang L L, Zhang X D, Liu S R, Sun P S and Wang T L. 2014. The basic principle of random forest and its applications in ecology: a case study of *Pinus yunnanensis*. *Acta Ecologica Sinica*, 34(3): 650-659 (张雷, 王琳琳, 张旭东, 刘世荣, 孙鹏森, 王同立. 2014. 随机森林算法基本思想及其在生态学中的应用——以云南松分布模拟为例. *生态学报*, 34(3): 650-659) [DOI: 10.5846/stxb201306031292]
- Zhang W. 2012. Remote Sensing Retrieval of Suspended Matters Based on Intelligent Calculation——A Case Study of Middle Yangtze River. Beijing: University of Chinese Academy of Sciences (张文. 2012. 基于智能计算的水体悬浮物遥感反演研究——以长江中游为例. 北京: 中国科学院大学)
- Zhang W J, Wei L P and Qu G. 2013. The analysis of channel evolution of Jingjiang River in different types after operation of Three Gorges Project. *Water Conservancy Science and Technology and Economy*, 19(11): 56-59 (张卫军, 魏立鹏, 渠庚. 2013. 三峡工程运用后荆江不同河型河道演变分析. *水利科技与经济*, 19(11): 56-59) [DOI: 10.3969/j.issn.1006-7175.2013.11.021]
- Zimmermann A, Francke T and Elsenbeer H. 2012. Forests and erosion: Insights from a study of suspended-sediment dynamics in an overland flow-prone rainforest catchment. *Journal of Hydrology*, 428-429: 170-181 [DOI: 10.1016/j.jhydrol.2012.01.039]

Remote sensing estimation of suspended sediment concentration based on Random Forest Regression Model

FANG Xinrui^{1,2}, WEN Zhaofei¹, CHEN Jilong¹, WU Shengjun^{1,2}, HUANG Yuanyang¹, MA Maohua^{1,2}

1. Chongqing Institute of Green and Intelligent Technology, Key Laboratory of Reservoir Aquatic Environment, Chinese Academy of Sciences, Chongqing 400714, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: Since 2003 when the Three-Gorge Dam (TGD) was in impoundment, the dam abundantly blocks suspended sediment and cause clear water flowing through the dam, which induces scouring effect on the beds and banks of the Yangtze river below the dam. Furthermore, the altered Suspended Sediment Concentration (SSC) has adversely affected the downstream coastal environment.

In this study, the random forest model was applied for SSC estimation. The model is flexible and robust, and can be used for regression analysis of ecological environment variables. Yet, its ability in estimating SSC in aquatic environment has not been fully understood. On the basis of the monitoring data of SSC and satellite remote sensing reflectance data from 2002 to 2015, this study estimated the SSC in Yichang–Chenglingji downstream reach of the TGD by constructing a non-parametric regression model using random forest.

The results showed that: (1) the random forest model could effectively monitor SSC, and the correlation coefficient and prediction accuracy were significantly improved from those of other models (linear regression, support vector machine, and artificial neural network model). (2) the red band is a suitable predictor for SSC in the random forest model, but cannot be independently used for forecasting. SSC remote sensing prediction requires multivariate co-participation. (3) By using the random forest model, the average root mean square error of the seasonal division was 0.46 mg/L, and the average relative root mean square error was 12.33%. These values met the needs of high-precision SSC estimation.

In conclusion, this study reveals that the season shall be considered as temporal factors to estimate SSC and then prepare for the subsequent SSC spatiotemporal inversion. Which is of great help to reveal the TGD's downstream river sediment evolution, and understand the regional distribution of sediment and sediment variation process in the future.

Key words: TGD(Three Gorges Dam), downstream channel of dam, remote sensing of aquatic environment, random forest, remote sensing monitoring, MODIS, sediment inversion

Supported by National Natural Science Foundation of China (No. 51779241, 41501096)



封面说明

About the Cover

中国甘肃酒泉区域的高光谱影像

Hyperspectral image of Jiuguang Area in Gansu Province, China

封面图片为 OHS-2C (贵阳一号) 高光谱卫星于 2018 年 8 月 16 日获取的甘肃省酒泉区域的标准假彩色遥感影像。OHS-2C (贵阳一号) 属于“珠海一号”卫星星座, 是由珠海欧比特宇航科技股份有限公司发射并运营的商业遥感微纳卫星星座。星座规划由 34 颗不同种类的卫星组成, 目前已发射升空 3 颗视频星, 4 颗高光谱星, 并计划于 2019 年 8 月继续发射 5 颗卫星。每颗 OHS 高光谱卫星均配置 1 台空间分辨率优于 10 m、幅宽优于 150 km 的高光谱相机。单颗 OHS 高光谱卫星的最大成像范围是 150 km × 2500 km (即 37.5 万 km²), 4 颗卫星可实现每 5 d 左右对全球覆盖扫描一遍, 对于特定地区可以每天多次重复扫描。OHS 高光谱卫星光谱分辨率为 2.5 nm (400—1000 nm), 可以检测光谱信息的细微差异, 大大提高了区域范围内信息采集能力, 对遥感信息的定量分析和更精细化的地物识别有重大意义, 可广泛应用于农业、林业、水利、环境保护等领域。

The cover image shows the standard pseudocolor image of Jiuguang, Gansu, which was obtained by OHS-2C (Guiyang-1) hyperspectral satellite on August 16, 2018. The features of sandy land, vegetation and road are clear. OHS-2C (Guiyang-1) hyperspectral satellite belongs to "Zhuhai-1" satellite constellation, which is a commercial remote sensing micro-nano satellite constellation launched and operated by Zhuhai Orbit Aerospace Technology Co., Ltd. The Planned constellation consists of 34 different kinds of satellites. At present, three video satellites and four hyperspectral satellites have been launched, and five satellites are scheduled to be launched in August 2019. Each OHS hyperspectral satellite is equipped with a hyperspectral camera with spatial resolution more than 10 m and width more than 150 km. The maximum imaging range of a single OHS hyperspectral satellite is 150 km × 2500 km (i.e. 375,000 square kilometers). Four satellites can achieve global coverage scanning once in about five days, and multiple visits can be made daily for specific areas. OHS hyperspectral satellite has a spectral resolution of 2.5 nm (400—1000 nm). It can test the subtle differences of spectral information, which can greatly improve the ability of information acquisition in the research area. It has great significance for quantitative analysis of remote sensing information and more refined recognition of land features. It can be widely used in agriculture, forestry, water conservancy, environmental protection and other fields.

主管	中国科学院	Superintended	by	Chinese Academy of Sciences
主办	中国科学院遥感与数字地球研究所 中国地理学会环境遥感分会	Sponsored	by	Institute of Remote Sensing and Digital Earth, CAS The Associate on Environment Remote Sensing of China
主编	顾行发	Editor in Chief		GU Xingfa
编辑	《遥感学报》编委会 北京市海淀区北四环西路19号 邮编: 100190 电话: 86-10-58887052 http://www.jors.cn E-mail: jrs@radi.ac.cn	Edited	by	Editorial Board of Journal of Remote Sensing Add: P.O.Box 2702, Beijing 100190, China Tel: 86-10-58887052 http://www.jors.cn E-mail: jrs@radi.ac.cn
出版	科学出版社	Published	by	Science Press
印刷装订	北京科信印刷有限公司	Printed	by	Beijing Kexin Printing Co. Ltd.
总发行	科学出版社 北京东黄城根北街16号 国内邮发代号: 82-324 邮政编码: 100717 电话: 86-10-64017032 淘宝店名称: 中科院	Distributed	by	Science Press Add: 16 Donghuangchenggen North Street, Beijing 100717, China Tel: 86-10-64017032 Taobao: Zhongke Journal
国外发行	中国国际图书贸易总公司 北京 399 信箱 邮政编码: 100044 国外发行代号: BM 1002	Overseas distributed	by	China International Book Trading Corporation Add: P.O.Box 399, Beijing 100044, China



官网



微博



淘宝



微店

