哨兵2卫星综合水质指标的河流水质遥感监测方法

王歆晖[1,2],田　华[3],季铁梅[3],巩彩兰[2,4],胡　勇[2,4],李　澜[1,2],何志杰[1,2]
(1.中国科学院大学 物理科学学院,北京 100049;2.中国科学院 上海技术物理研究所,上海 200083;
3.上海市水文总站,上海 200232;4.中国科学院 红外成像光谱重点实验室,上海 200083)

摘　要: 上海市河流水污染监测是当前研究的热点问题。针对上海市河流水质监测的需求,提出了一种基于因子分析的河流综合水质遥感反演的方法。以欧洲航空局发射的哨兵-2A(Sentinel-2A)卫星搭载的多光谱成像仪获取的遥感影像,以及准同步获取的河流断面实测水质参数数据为例,研究对象为上海市青浦区和松江区部分河道。选取高锰酸盐指数、氨氮等5项主要水质参数为监测指标,在因子分析的基础上建立水质遥感反演的模型,得到综合水质指标用于确定水质类别。实验结果表明:综合水质指标值越小,内陆河流水质状况越良好,水体更清洁。该方法能够应用于内陆河流水质遥感监测,为水环境管理部门提供参考信息。

关键词: 综合水质; 波段筛选; 因子分析; 反演模型; Sentinel-2A 卫星
中图分类号: TP 79　　**文献标志码:** A　　**DOI:** 10.19328/j.cnki.1006-1630.2020.05.014

Remote Sensing Monitoring Method for Comprehensive Water Quality Index in Rivers Based on Sentinel-2 Satellite

WANG Xinhui[1,2], TIAN Hua[3], JI Tiemei[3], GONG Cailan[2,4], HU Yong[2,4], LI Lan[1,2], HE Zhijie[1,2]
(1.School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China; 2.Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China; 3.Shanghai Hydrological General Station, Shanghai 200232, China; 4.Key Laboratory of Infrared System Detection and Imaging Technology, Chinese Academy of Sciences, Shanghai 200083, China;)

Abstract: River water pollution monitoring in Shanghai is a hot research issue. Aiming at the demand of river water quality monitoring, a water quality remote sensing inversion method for rivers is proposed based on factor analysis. The images captured from a multispectral imager mounted on the Sentinel-2A satellite launched by the European Aviation Administration and the water quality parameters of river sections obtained in a quasi-synchronous manner are taken as examples. The research object is part of rivers of Qingpu and Songjiang in Shanghai. Five main water quality parameters such as the permanganate index and the ammonia nitrogen index are selected as the monitoring indices. Based on factor analysis, a water quality remote sensing inversion model is established, and the comprehensive water quality index is obtained. The obtained index is then used to determine the water quality grade. The experimental results show that the smaller the comprehensive water quality index, the better the water quality of the inland rivers and the cleaner the water. The method can be applied to the remote sensing monitoring of inland river water quality and provide reference information for water environment management departments.

Key words: comprehensive water quality; band selection; factor analysis; inversion model; Sentinel-2A satellite

收稿日期: 2019-08-16; **修回日期:** 2019-10-13
基金项目: 上海市水务局科研项目(沪水科 2018-07);上海市科委科技创新项目(18511102400);国家重点研发计划(2017YFC0602103)
作者简介: 王歆晖(1995—),女,硕士,主要研究方向为遥感应用、数字图像处理。
通信作者: 巩彩兰(1974—),女,博士,研究员,主要研究方向为遥感图像处理与应用。

0 引言

内陆河流作为重要的资源和载体,是生态系统的绿色生命线,关系到人类的生存和发展。然而,河流水环境极易受到污染,特别是城市河道,随着现代城市的飞速发展与扩张,人口密度增加,给城市河流水环境带来了挑战。从2017年起,上海市就推行“河长制”,目的在于加强对上海市市管河道的监管,对水污染进行防治,保护水资源。因此,保障居民的正常生产生活及推进可持续发展,对城市河流水质快速、准确的监测工作提出了高要求。目前,河流水质监测主要采用人工实地采样与实验室分析的方法,该方法结果准确,但只能获取河流截面采样点数据[1-2]。与其相比,内陆河流遥感监测具有以下优势:1)遥感技术具有快速、大范围和周期性的特点,可以弥补常规定时、定点监测的不足,节省大量人力、物力和财力[3];2)随着对水体光谱特性的深入研究以及水质参数反演模型的改进,遥感图像能够较为准确地获取河流水质信息,与地面实测水文参数、地理位置信息等结合,可有效发现污染水体以及水质变化趋势,作为水质管理的参考。

随着新的卫星传感器不断升空,时间分辨率和空间分辨率不断提高,水质遥感监测从定性发展到定量,可实现遥感反演的水质指标也由叶绿素a、悬浮物,发展到总氮、总磷、高锰酸盐指数等[4-7]。当前,利用卫星遥感技术对水质监测的研究主要集中在区域跨度较大的研究对象上,例如大型湖泊水库和河口海岸的监测,对小尺度和微型态环境问题研究相对较少。MATTEWS[8]利用MERIS数据发展了改进的最大峰值算法,对内陆和近岸水体叶绿素a浓度进行探测,研究了非洲南部50多个湖库水体从2002到2012年的富营养化及其变化状况。HOU等[9]利用MODIS地表反射率数据645 nm波段反演了长江中下游102个大型湖泊的水体总悬浮浓度,并分析了其变化趋势。国内学者臧友华[10]将遗传算法和支持向量机组合,提出一种新的反演模型GA-SVM模型,采用ETM+数据对渭河水质参数叶绿素a、透明度、总磷进行了反演。巩彩兰等[11]对黄浦江进行了水体光谱测量和同步水质采样,分析了各水质参数与反射率及相关反射率特征的相关性,建立了常规水质参数与反射率之间的关系模型。以上研究主要是针对部分具有光学特性的遥感水质参数或者是国家《地表水环境质量标准》中规定的单一水质参数,虽然简单直观地反映了单因子污染状况,但是缺乏对河流水质的综合研究。因此,本文利用5项主要水质参数建立河流综合水质遥感模型,以综合水质指标来表征河流水质状况。

本文采用Sentinel-2A卫星遥感影像数据和同步地面河流断面实测水质参数数据,对5项水质参数溶解氧、高锰酸盐指数、五日生化需氧量、氨氮以及总磷进行因子分析,在此基础上建立水质遥感反演的模型,得到综合水质指标用于确定水质类别。

1 实验数据及预处理

1.1 实验数据

本文采用欧洲航空局发射的Sentinel-2A卫星搭载的多光谱成像仪获取的遥感影像,影像成像日期为2018年4月9日,区域为上海市,具体研究选择上海市青浦区和松江区的部分河流流域,研究区域如图1所示。影像覆盖13个光谱波段,幅宽达290 km,重访周期10 d。从可见光和近红外到短波红外,具有不同的空间分辨率。

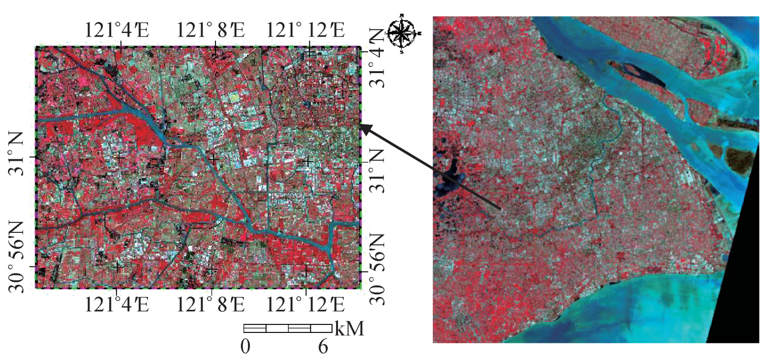

图 1　研究区域卫星影像 band 8、4、3 假彩色合成图

Fig.1　False color composite images of bands 8, 4, and 3 in the study area

考虑空间分辨率,本研究用到的波段为 band 2、band 3、band 4、band 8(见表 1)。

表 1 Sentinel-2A 卫星部分波段信息

Tab.1 Partial band information of Sentinel-2A satellite

波段	中心波长/μm	空间分辨率/m
band 2-蓝光	0.490	10
band 3-绿光	0.560	10
band 4-红光	0.665	10
band 8-近红外	0.842	10

水质参数采样数据获取的是卫星成像同一天的上海市河流断面实地采样分析数据,按照《地表水环境质量标准(GB 3838—2002)》对上海市主要河流 94 个断面采样点溶解氧(DO)、高锰酸盐指数(COD_{Mn})、五日生化需氧量(BOD_5)、氨氮(NH_3-N)、总磷(TP)共 5 项指标进行评价,如图 2 所示。总体上,溶解氧、高锰酸盐指数、总磷评价等级较好,基本达到 3 类水的标准。超标的主要水质参数是氨氮,通常来源于饲料、水生动物的排泄物、肥料及动植物尸体分解,未处理或处理过的城市生活和工业废水、各种浸滤液和地表径流是氨氮含量升高的主要原因[12]。

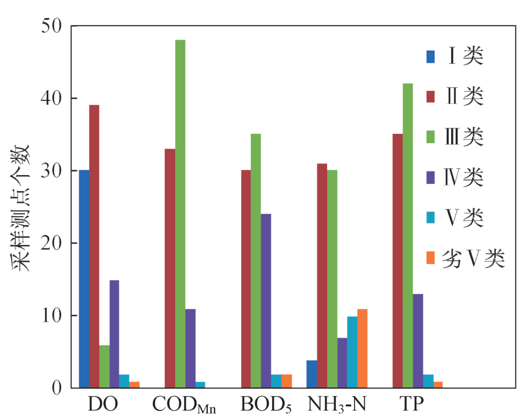

图 2 各水质参数所属水体类型分布情况图

Fig.2 Water grade summary of each water quality parameter

1.2 实验数据预处理

对于经过辐射定标等预处理后的反射率影像,运用归一化差异水体指数(NDWI)对 2018 年 4 月 9 日的实验影像研究区进行河流水体提取。对 NDWI 图像设置掩模模板提取水体,阈值往往设为 0,即 NDWI 图像大于 0 的像素被认为是水体[13]:

$$\mathrm{NDWI}=\frac{\mathrm{band_{green}}-\mathrm{band_{NIR}}}{\mathrm{band_{green}}+\mathrm{band_{NIR}}} \tag{1}$$

式中:$\mathrm{band_{green}}$ 为第 3 波段(绿光)反射率;$\mathrm{band_{NIR}}$ 为第 8 波段(近红外)反射率。

将影像与水体提取掩模模板叠加对比,掩模结果如图 3 所示,其中白色部分为识别的水体,而黑色部分为非水体。目视解译结果表明,归一化差异水体指数 NDWI 能够较好地识别出河流。

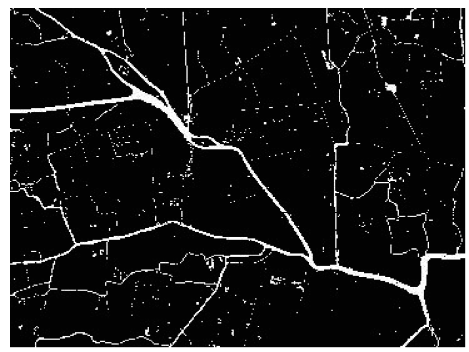

图 3 2018 年 4 月研究区水体 NDWI 提取掩模结果

Fig.3 Mask results of water extraction with NDWI in the study area in April 2018

2 基于因子分析的水质遥感反演方法

2.1 数据波段筛选

在正式进行建模前对波段进行了筛选,结合单波段与各主要水质参数的相关性以及 OIF 指数,来选择对水质参数变化较为敏感的特征波段或者波段组合。本文分析了各单波段与各水质参数的相关性,整体相关性较低,Pearson 系数整体不超过 0.4,所以 Sentinel-2 单波段并不适合进行水质遥感反演,需要考虑其他波段组合。研究区主要超标的水质参数为氨氮,各波段与氨氮相关系数按照从大到小顺序排列,波段顺序为 band 3、band 4、band 2、band 8。OIF 指数[14]是进行波段筛选的依据之一,通常波段数据的标准差越大,包含的信息量越多;而波段之间的相关系数越小,则波段之间独立性越高,数学表达式为

$$\mathrm{OIF}=\frac{\sum_{i=1}^{3}S_i}{\sum_{i=1}^{3}\left|R_{ij}\right|},\quad j>i \tag{2}$$

式中:S_i 为第 i 个波段的标准差;R_{ij} 为第 i、j 两波段的相关系数。

不同三波段组合的OIF指数见表2,按照从大到小的顺序排列来选择最优组合方案,理想组合顺序为:1) band 2、band 4、band 8;2) band 3、band 4、band 8;3) band 2、band 3、band 8。因组合1和组合2的OIF指数值差异并不明显,结合研究区主要超标的水质参数氨氮与波段的相关性,最终确定选用组合2波段3、4、8组合。

表2　不同三波段组合OIF指数表

Tab.2　OIF index table of different three-band combinations

不同三波段组合			OIF指数
2	4	8	0.030 732
3	4	8	0.030 409
2	3	8	0.029 797
2	3	4	0.029 248

2.2　综合因子分析理论

因子分析是一种处理多元变量数据的统计方法,研究众多变量之间的内在关系,对可观测变量进行概括和抽象,用较少的因子变量来最大程度上解释原始的观测变量[15-16],其主要表达形式为

$$\begin{cases} X_1 = a_{11}F_1 + a_{12}F_2 + \cdots + a_{1m}F_m + \varepsilon_1 \\ X_2 = a_{21}F_1 + a_{22}F_2 + \cdots + a_{2m}F_m + \varepsilon_2 \\ \quad\vdots \\ X_p = a_{p1}F_1 + a_{p2}F_2 + \cdots + a_{pm}F_m + \varepsilon_p \end{cases} \tag{3}$$

式中:$X_1, X_2, \cdots, X_p$为p个存在相关关系的可观测变量;$F_1, F_2, \cdots, F_m$为m个互相独立的因子,又称为公共因子,是不可观测的变量;$\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_p$为特殊因子,是无法包含在公共因子中的部分;$a_{ij}(1 \leqslant i \leqslant p, 1 \leqslant j \leqslant m)$为第$i$个变量与第$j$个因子的相关系数。

基于因子分析建立遥感反演模型,首先计算原始变量的相关系数矩阵,然后利用相关系数矩阵结合反演模型和先验知识提取公共因子,通常只取方差大于1或者特征值大于1的因子,或者按照因子的累计方差贡献率来确定,一般要求至少达到70%。因子F_j的方差贡献是式(3)中对应列相关系数的平方和,根据如下公式计算方差贡献率,方差贡献率进行累加即为累计方差贡献率:

$$\omega_j = \frac{g_j^2}{\sum_{j=1}^{m} g_j^2}, g_j^2 = \sum_{i=1}^{p} a_{ij}^2 \tag{4}$$

式中:g_j^2为方差贡献。

接着进行因子旋转,其目的是通过坐标变换使因子实际意义更易解释,最后根据如下公式计算综合因子得分:

$$F = \frac{(\omega_1 F_1 + \omega_2 F_2 + \cdots + \omega_n F_n)}{\omega_1 + \omega_2 + \cdots + \omega_n} \tag{5}$$

式中:$\omega_j(1 \leqslant j \leqslant n)$为对应因子$F_j$的方差贡献率。

2.3　基于综合水质因子的遥感反演模型

由于水质参数本身数量级的不一致,为了避免对模型造成影响,对数据进行最大最小化标准化,将其映射到−1至1范围内。采用因子分析法对研究区水质参数进行分析,首先KMO值为0.763,大于阈值0.5,说明了变量之间是存在相关性的,符合要求;然后是Bartlett球形检验的结果,Sig检验概率值为0.000,小于阈值0.05,说明结果显著,表明数据可以进行因子分析。

河流水质参数因子分析的主要因子成分如图4所示。主要因子1的特征值为2.8,方差贡献率为56.74%;主要因子2的特征值为0.95,方差贡献率为18.93%;前3个主要因子的方差贡献率之和为87.5%。因此,水质参数监测数据原始5个变量的信息能够用变换后的3个主要因子来表达。

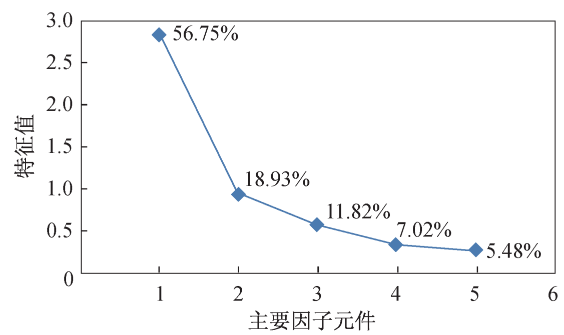

图4　监测数据的主要因子成分

Fig.4　Major factor components of the monitoring data

由表3可知,河流水质参数因子分析的旋转载荷矩阵和贡献率,主要抽取了方差贡献率较大的前3个主要因子成分,因子成分F_1主要受高锰酸盐指数和五日生化需氧量控制,系数分别为0.928和0.832,这两项指标越大,反映水体受有机物的污染越严重。因子成分F_2主要由总磷和氨氮控制,系数分别为0.921和0.649。通常是城市生活洗涤污水、垃圾等包含的无机盐类,与水体的富营养化有关。因子成分F_3则主要由溶解氧控制,系数值最大为0.970。

表3 因子旋转载荷矩阵与方差贡献率

Tab.3 Factor loading matrix and variance contribution rate

	成分		
	F_1	F_2	F_3
溶解氧	−0.120	−0.212	0.970
高锰酸盐指数	0.832	0.331	−0.156
五日生化需氧量	0.928	0.092	−0.051
氨氮	0.564	0.649	−0.164
总磷	0.146	0.921	−0.204
贡献率/%	56.748	18.930	11.822
累计贡献率/%	56.748	75.678	87.501

5项主要水质参数作为式(1)中各可观测变量,经过因子旋转后借助表3的因子旋转载荷矩阵,可以得到如下3个主要因子成分的表达式:

$$\begin{aligned} F_1 &= -0.120\times \mathrm{DO} + 0.832\times \mathrm{COM_{mn}} + 0.928\times \mathrm{BOD_5} + 0.564\times \mathrm{NH_3\text{-}N} + 0.146\times \mathrm{TP} \\ F_2 &= -0.212\times \mathrm{DO} + 0.331\times \mathrm{COM_{mn}} + 0.092\times \mathrm{BOD_5} + 0.649\times \mathrm{NH_3\text{-}N} + 0.921\times \mathrm{TP} \\ F_3 &= 0.970\times \mathrm{DO} - 0.156\times \mathrm{COM_{mn}} - 0.051\times \mathrm{BOD_5} - 0.164\times \mathrm{NH_3\text{-}N} - 0.204\times \mathrm{TP} \end{aligned} \tag{6}$$

将各个公共因子的方差贡献率等参数代入式(5),得到水质综合因子得分模型为

$$F = \frac{F_1\times 56.748 + F_2\times 18.930 + 11.822}{87.501} \tag{7}$$

为了得到最终水质遥感反演模型,需要建立各个水质参数和遥感影像波段3、4、8之间的关系。以影像波段反射率数据作为自变量,水质参数数据作为因变量,采用最小二乘法确定多元线性回归模型的待定参数,当实测值与回归模型得到的拟合值的差值平方和达到最小时的模型参数即为所求待定参数。根据研究数据得到的回归模型结果见表4。由回归分析结果可见,溶解氧、高锰酸盐与波段之间的关系相对较密切。

表4 不同水质参数的线性回归模型

Tab.4 Regression models of different water quality parameters

水质参数	线性回归模型	R
溶解氧	$\mathrm{DO}=0.638+b_3\times 2.764-2.988\times b_4-0.036\times b_8$	0.499
高锰酸盐指数	$\mathrm{COD_{Mn}}=0.660-b_3\times 2.574+1.375\times b_4+0.816\times b_8$	0.481
五日生化需氧量	$\mathrm{BOD_5}=0.543-b_3\times 2.297+1.480\times b_4+0.488\times b_8$	0.391
氨氮	$\mathrm{NH_3\text{-}N}=0.414-b_3\times 2.810+1.846\times b_4+0.526\times b_8$	0.440
总磷	$\mathrm{TP}=0.353-b_3\times 1.844+1.164\times b_4+0.391\times b_8$	0.359

根据得出的综合因子得分模型以及水质参数和波段的回归分析结果,计算得到最终水质遥感反演模型为

$$F = 0.644\,2 + 0.195\,7\times b_4 - 1.018\,2\times b_3 + 0.441\,7\times b_8 \tag{8}$$

3 水质遥感反演结果与精度分析

对预处理[17]后的2019年4月9日的研究区影像,应用综合水质因子反演模型进行反演,得到研究区的综合水质图,如图5所示。一般情况下,综合水质F值越小,代表水体越洁净,否则水体则存在水质参数超标,或者受到污染的可能性。从结果图来看,主干河流的值基本低于支流,是符合实际水质情况的。

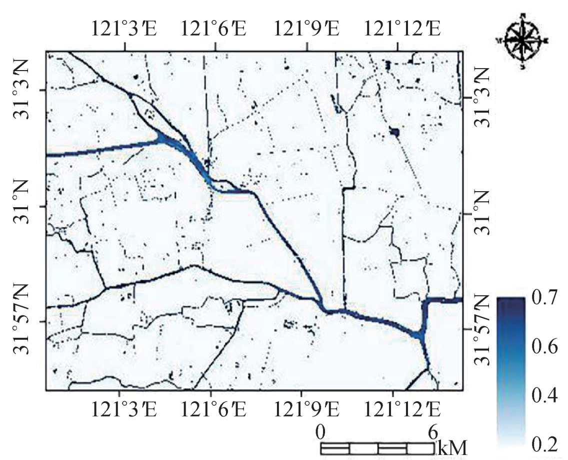

图5 2019年4月研究区综合水质图

Fig.5 Comprehensive water quality map of the study area in April 2019

在综合水质图的基础上，据不同水质等级的情况分别设定阈值，最终确定不同情况下综合水质F对应的水质等级，绘制综合水质等级图，如图6所示。因卫星遥感影像分辨率的限制，图中小河流存在断裂、破碎等状况。从结果图看Ⅲ类和Ⅳ类水主要分布在主干河流，相对而言，Ⅴ类水体在支流较多，整体水质状况良好，劣5类水体只出现在局部区域。

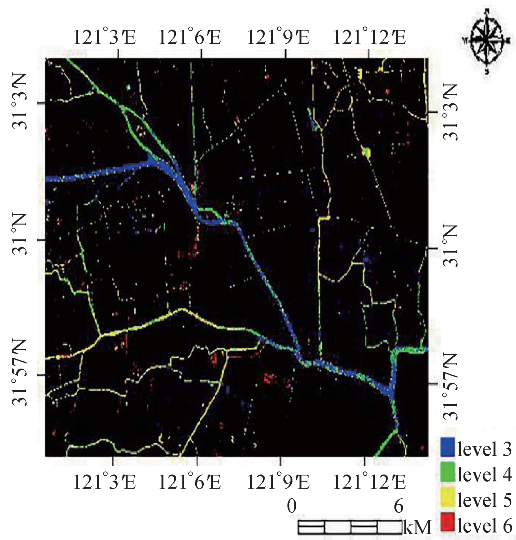

图6　2019年4月研究区综合水质等级图

Fig.6　Comprehensive water quality grade map of the study area in April 2019

随机选取30个样本数据，将反演的综合水质等级结果与实际测量采样的结果进行对比验证，如图7所示，水质类别2、3、4、5、6分别对应标准水质类别Ⅱ类、Ⅲ类、Ⅳ类、Ⅴ类以及劣Ⅴ类。

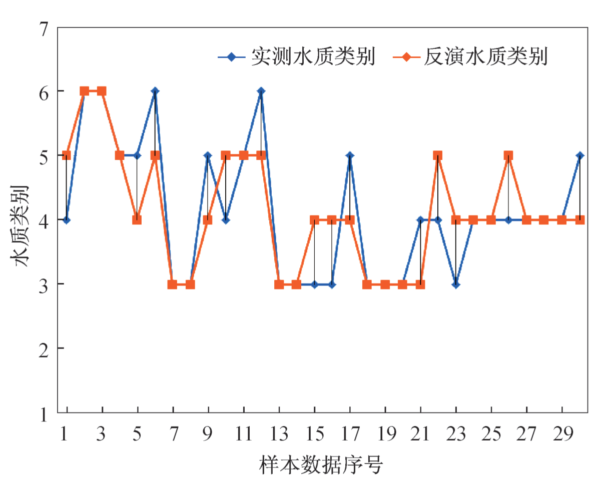

图7　实测水质与反演综合水质类别对比图

Fig.7　Comparison of the measured water quality and the inversion comprehensive water quality

在大河道区域准确度能够达到71.2%，但是细小的支流准确度只能够达到42.8%，误差的主要来源如下：1）地面实测采样水质数据与卫星影像时间和空间的不一致性；2）仅以5项主要水质参数表征河流水质状况，仍存在其他因素的干扰；3）卫星数据空间分辨率的限制会带来的小河流混合像元问题。

4　结束语

利用2018年4月Sentinel-2A遥感影像以及同步获取的河流断面实测水质参数数据，结合5项主要水质参数溶解氧、高锰酸盐指数、五日生化需氧量、氨氮以及总磷，在因子分析的基础上建立水质遥感反演的模型，得到综合水质指标用于确定水质类别。实验结果表明：1）上海市主要超标的水质参数指标为氨氮，水污染治理应着重处理难降解的有机物以及含氮的可溶性无机物等；2）综合水质指标值越小，内陆河流水质状况越良好，水体更清洁；3）本文方法对于大河流的水质等级判定的正确率达到了71.2%。基于因子分析的水质遥感反演模型能够应用于内陆河流水质遥感监测，为水环境部门提供参考信息。但是本文仅考虑了5项水质参数，对于总体水质状况的表达还略有不足，需要考虑其他水质遥感监测指标，辅助综合水质类别判定；另外，该模型仅适用于一个季节，获取长时间序列的影像以及同步水质参数数据，能够对模型进行改进，并且利用多时相数据对内陆河流水质状况进行动态变化研究。

参考文献

[1] PALMER S C J, KUTSER T, HUNTER P D. Remote sensing of inland waters: challenges, progress and future directions [J]. Remote Sensing of Environment, 2015, 157:1-8.

[2] 段洪涛，罗菊花，曹志刚，等. 流域水环境遥感研究进展与思考[J]. 地理科学进展，2019，38(8)：1182-1195.

[3] 马荣华，段洪涛，唐军武，等. 湖泊水环境遥感[M]. 北京：科学出版社，2010: 1-2.

[4] HAJIGHOLIZADEH M, MELESSE A M. Assortment and spatiotemporal analysis of surface water quality using cluster and discriminant analyses [J]. Catena, 2016, 151: 247-258.

（下转第104页）