

文章编号:1004-8227(2011)01-0090-06

Landsat7 ETM+ SLC-OFF 数据的修复 及其在武汉东湖水质反演中的应用

何报寅¹, 丁超¹, 杨小琴¹, 梁胜文²

(1. 中国科学院测量与地球物理研究所, 湖北 武汉 430077; 2. 武汉环境监测中心站, 湖北 武汉 430051)

摘要:自从 2003 年 5 月 31 日, 从陆地卫星-7 发回的 ETM+ 图像数据就存在缺陷。这是由于增强专题制图仪的扫描线校正器发生故障引起的。这些称为 SLC-OFF 数据的图像有一些黑色的不存在任何数据的扫描行。丢失的数据约占全景数据的 25%, 使它们难以正常使用。但是, 数据本身仍然保留了良好的辐射和几何性质, 如加以妥善修复, 仍可以在一些特殊领域中使用。首先介绍了如何使用自适应局部回归算法(ALR)恢复这些图像, 然后使用修复后的图像反演武汉东湖的水质参数。结果表明: ALR 算法可以对 SLC-OFF 图像进行较好的修复, 而且利用修复后的图像和东湖的地面水质监测数据, 通过多元逐步回归分析, 可以建立很好的叶绿素 a、透明度、总磷以及总氮等水质参数的经验遥感反演模型, 模型的相关系数 R^2 分别为 0.86、0.75、0.73 和 0.71。反演得到的水质参数分布与实际情况符合。这些数据有许多优点, 如空间分辨率高、存档数据非常丰富、可以从 NASA 的服务器免费下载等。在其他遥感数据不足或无法获得的情况下, 这些数据经过适当的修复, 可以作为补充或替代数据使用。

关键词: Landsat7 ETM+; 影像修复; ALR 算法; 水质反演; 武汉东湖

文献标识码: A

近年来, 卫星遥感技术被越来越多地用来监测和评价海洋和内陆水体的水质状况, 该技术具有快捷、经济、覆盖范围广和能够绘制水质空间分布图等传统地面监测所不具备的优点^[1~3]。水质遥感反演的精度依赖于卫星传感器的辐射分辨率和光谱分辨率, 而其所能反演的时空尺度则依赖于遥感图像的空间分辨率和时间分辨率^[4]。在实际工作中, 往往要考虑数据成本, 可用的遥感数据源越丰富越便宜, 则对水质的动态监测越有利。因此, 需要充分挖掘各遥感数据源的潜力。

美国陆地卫星 Landsat7 自 1999 年 4 月发射, 获取的大量数据得到了广泛应用。然而, 2003 年 5 月 31 日, 该卫星上增强专题制图仪(ETM+)的扫描线校正器(Scan Lines Corrector, 简称 SLC)发生故障, 导致获取的图像出现数据重叠和大约 25% 的数据丢失, 自此提供的图像数据上带有数据缺行, 这些缺行数据(Gap Data)很难正常应用(故障发生后获得的有缺行的图像称为 SLC-OFF 图像, 之前正常的称为 SLC-ON 图像)^[5,6]。由于该星载扫描仪

成像平台仍然正常工作, 依旧源源不断地传回大量的遥感图像, 这些 SLC-OFF 图像仍然保持了良好的几何特性和辐射特性。目前 NASA 已开放了该数据库, 用户注册后就可以免费下载。因此, 如何对这些数据进行适当的处理, 充分挖掘其应用潜力, 是摆在研究者面前的问题。

本文通过引入自适应局部回归匹配算法(Adaptive Local Regression Match, ALR)和程序, 对 landsat7+ETM 缺行影像进行自动修复, 并把修复后的影像用武汉市东湖的水质参数反演, 为 Landsat7 ETM+ SLC-OFF 数据的修复和应用提供一个实例。

1 数据源

本文采用的遥感数据为采集时间相邻的两景 Landsat7 ETM+SLC-OFF 影像, 影像获取时间分别为 2008 年 5 月 6 日和 2008 年 5 月 21 日。可以从美国地质调查局有关网站下载。

收稿日期: 2010-01-11; 修回日期: 2010-05-13

基金项目: 中国科学院知识创新工程重要方向项目(kzcx2-yw-141); 国家自然科学基金项目(51079137)

作者简介: 何报寅(1964~), 男, 广西玉林人, 研究员, 主要从事环境遥感、水资源和全球变化记录研究。E-mail: heby@whigg.ac.cn

地面水质监测数据由武汉市环境监测中心站提供,采样和测量日期为 2008 年 5 月 25 日。水质参数监测分析遵循中国国家环境保护部颁布的《地表水和污水监测技术规范》HJ/T91-2002。测量和采样点共有 22 个,遍布整个东湖,如图 1。本文用到的水质参数包括叶绿素 a 浓度(Chla)、总氮(TN)、总磷(TP)和透明度(SD),除透明度的单位为 cm 外,其它参数的单位都是 mg/L。

2 数据预处理

如前所述,ETM+ SLC-OFF 图像存在缝隙,如图 2 所示,在应用前必须进行修复,而在这之前,需要对图像进行辐射纠正。同时,为了能较好的建立水质反演模型,还必须对图像进行降低噪音等处理。由于水对太阳光的强烈吸收作用,水质信息在遥感图像中一般为弱信息,相对于其它地物而言,

更易受大气、平台姿态、仪器信号等干扰因素的影响,因此在水质遥感监测中,对遥感图像的预处理和标准化就显得尤为重要。

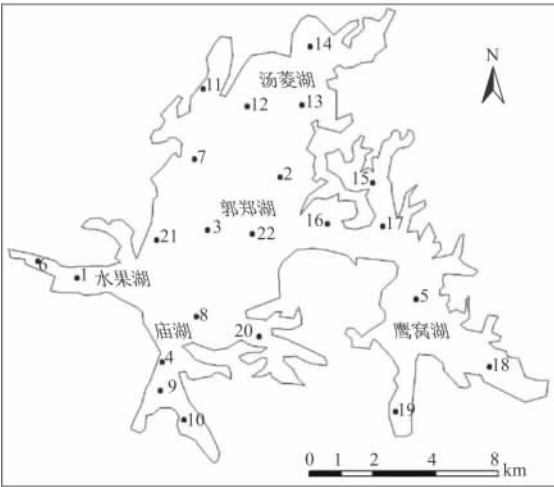


图 1 武汉东湖采样点位置

Fig. 1 Sampling Sites in East Lake, Wuhan

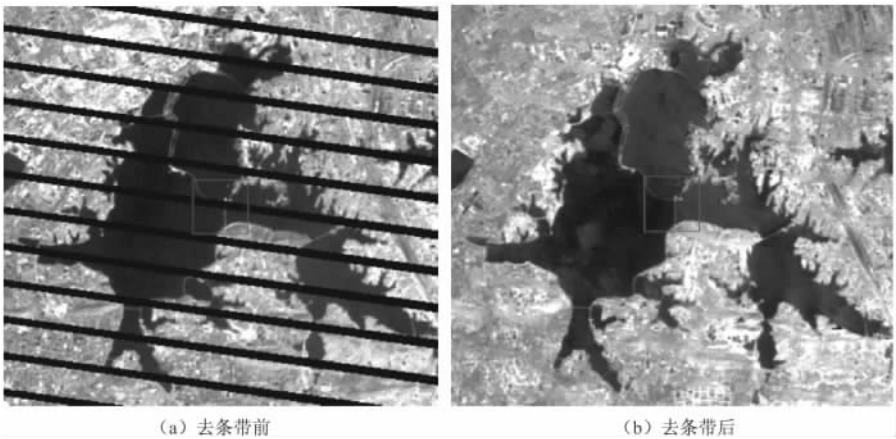


图 2 缝隙填充前后的图像

Fig. 2 Image Before and After Gaps Filling

2.1 辐射校正

利用下载图像头文件中记录的辐射校正参数,用户可方便地计算出地物在大气顶部的辐射亮度或反射率。计算式如下:

$$L=Gain\times DN+Bias$$
 (1)

$$\rho=\pi L D^2/E_{sun}\cdot\cos\theta$$
 (2)

式中:Gain、Bias 分别是该波段的定标增益和定标偏移量;L 为大气层顶进入卫星传感器的光谱辐射亮度 [W/(m²·sr·μm)];DN 是该波段图像像元的灰度值;ρ 为地物在大气顶部的辐射亮度或反射率(无量纲);π 为常量;D 为日地距离(天文单位),可以通过 $D=1-0.016\,74\times\cos(0.985\,6\times(JD-4)\times\pi/180)$ 求得,JD 为遥感成像时的儒略日;E_{sun} 为大气层顶的平均太阳光谱辐照度[W/(m²

·μm)];θ 为太阳天顶角,可以从影像头文件获得。

对 TM 影像而言,不同时期影像不同波段的辐射校正参数 Gain 和 Bias 是不同的。1~4 波段的具体参数见表 1。

表 1 定标增益和偏置取值

Tab. 1 Correcting Values of Gain and Bias

波段	1984 年 3 月 1 日~ 2003 年 5 月 4 日		2003 年 5 月 4 日之后增益	
	增益	偏置	增益	偏置
1	0.602 4	-1.52	0.762 8	-1.52
2	1.175 1	-2.84	1.442 5	-2.84
3	0.805 8	-1.17	1.039 9	-1.17
4	0.814 5	-1.51	0.872 6	-1.51

2.2 图像缝隙的修复

经过上述辐射校正的 ETM+ SLC-OFF 图像

就可以用来进行缝隙的修复,也就是想办法把图像中黑色条带部分的数据恢复出来,并且尽可能的与实际值接近。常用的修复的方法有直接插值法、全局直方图匹配法、局部直方图匹配法和自适应局部回归匹配算法(ALR)。其中,直接插值法就是利用缝隙图像中缝隙周围的有效像素进行最近邻插值、双线性插值或三次卷积插值等,但是由于缝隙的宽度达 14 个像元,故该法效果较差。全局直方图匹配法和局部直方图匹配法是利用 SLC 故障前相同季节同一景完整图像对缝隙图像进行直方图匹配恢复,全局直方图匹配法速度很快,但是误差比较大,不能很好地恢复出细节较多的图像,局部直方图匹配法对图像细节部分恢复精度较高,误差较小,但是算法复杂,不适用于遥感图像的批量处理。ALR 算法是在局部直方图匹配算法的基础上提出的一种算法,本文试验表明,该算法在恢复精度上要优于局部直方图匹配算法和全局直方图匹配法,算法复杂度介于全局直方图匹配法和局部直方图匹配法之间。有关文献的仿真实验也得出了相同的结论^[7]。

因此,本文采用 ALR 算法对 SLC-OFF 图像进行修复。用 2008 年 5 月 21 日的 SLC-OFF 图像(填充图像)去修复 2008 年 5 月 6 日的同一景图像(待修复图像)。由于不同时间获取的同一景 SLC-OFF 图像,其缝隙的位置一般不在同一位置,因此,实际上是可以对时间相邻的一对图像互相进行修复的。

ALR 算法原理:ALR 法利用回归方程建立了两幅图像 DN 值的联系,通过计算将填充图像的色调向待修复图像靠近。在对图像进行修复前,填充图像必须以待修复图像为基准进行配准,然后计算图像缺行中每个像素点的填充值。对于缺行中的每个像素点,填充 DN 值是从以它为中心的 17×17 窗口中通过回归方程算出的。如果这个窗口中,两幅图像都有值的像素点个数 $N < 144$,则窗口增大为 19×19 、 21×21 ,直到 $N > 144$ 为止。填充值的计算为:

$$DN_{ti} = a \times DN_{si} + b \quad (3)$$

$$a = \frac{\sum_{i=1}^N (DN_{ti} - \overline{DN_t})(DN_{si} - \overline{DN_s})}{\sum (DN_{si} - \overline{DN_s})^2} \quad (4)$$

$$b = \overline{DN_t} - a \times \overline{DN_s} \quad (5)$$

式中: a 为增益; b 为偏差; DN_{ti} 为修复后待修复图像缺行上第 i 点的灰度值; DN_{si} 为填充图像上第 i 点的灰度值; $\overline{DN_t}$ 为待修复图像修复前的灰度均值; $\overline{DN_s}$ 为填充图像的灰度均值。最后,将计算出

的 DN 值最终填充到被修复图像相应位置中。由于在整景图像上坏道的宽度在不同位置各有不同,待修复图像和填充图像某些地方都是坏道,无法填充修复,碰到这种情况,就要利用另一幅(或几幅)图像,遵循同一过程再修复。

2.3 噪音去除

为了有效的消除卫星传感器的噪音对单个像元的干扰,在遥感数据样本采集时,一般的做法是在地面监测点位相同坐标上采集多个像元,然后取灰度值平均值作为该点的灰度值。通过试验发现,当在采样大小达到 5×5 时,灰度平均值趋于稳定,达到一个常量。因此,这里采用 5×5 像元窗口对图像进行低通滤波处理,以消除高频随机噪音。处理后的数据用来建立反演模型。

3 反演模型和精度分析

遥感水质反演模型的建立有多种方法^[2],本文采用经验模型中的多元逐步回归分析方法。也就是用处理和修复后 2008 年 5 月 6 日的 ETM+ 影像为数据源,用东湖 22 个样点的 Chla、TN、TP、SD 这 4 个水质参数与对应监测点在影像上各波段遥感反射率 ρ 以及它们的波段组合做线性、对数、多项式等多元逐步回归分析,最后筛选出相关性高而且参数项少的最优模型。回归分析采用 OriginPro 7.5 软件完成。

在波段选择上,选择 TM/ETM1-5 波段进行建模。虽然在反演叶绿素和悬浮物时采用前 3 个波段(即可见光波段)通常已经足够,但许多研究表明,第 4、第 5 波段也是十分有用的^[2,3,9],本文也显示用 5 个波段的效果比用 3 个波段的效果有显著提高。

3.1 东湖叶绿素遥感反演回归分析

叶绿素浓度 Chla(mg/L)的反演模型为:

$$\text{Chla} = 71.804 + 0.226b1 - 0.081b3 - 25.376\ln b1 + 6.755\ln b3 - 0.209\ln b4 + 0.076\ln b5 \quad (6)$$

式中 $b1 \sim b5$ 分别表示修复后遥感影像 1~5 波段在样本点处的灰度值(取 7×7 像元平均值,下同)。方差分析显示,相关系数 $R^2 = 0.856$,校正 $R^2 = 0.798$, $F = 14.84$, $P < 0.0001$,相关关系显著(表 2)。

3.2 其它参数遥感反演回归分析

水体的反射特性虽然主要与叶绿素、黄色物质和悬浮物的含量有关,但是对于富营养化湖泊,叶绿素 a 的浓度往往与透明度、总磷、总氮浓度存在高度

表 2 Chl-a 反演模型多元逐步回归结果和方差分析表

Tab. 2 Results of Multiple Regression (a), Variance Analysis (b) and Correlationship (c) for Chl-a Retrieval Model

a 回归结果

参数	估计值	残差 <i>r</i>	<i>t</i> 值	显著性概率(> <i>t</i>)
Y-截距	71.804	21.186	3.389	0.004
<i>b</i> 1	0.226	0.076	2.980	0.009
<i>b</i> 3	-0.081	0.033	-2.448	0.027
ln <i>b</i> 1	-25.376	7.850	-3.233	0.006
ln <i>b</i> 3	6.755	2.460	2.746	0.015
ln <i>b</i> 4	-0.209	0.071	-2.948	0.010
ln <i>b</i> 5	0.076	0.045	1.679	0.114

b 方差分析表

项	自由度	平方和	均方	<i>F</i> 值
模型	6	0.028 1	0.004 7	14.84
残差	5	0.004 7	0.000 3	
总计	21	0.032 9		

c 相关关系

<i>R</i> ²	校正 <i>R</i> ²	估计的标准差	<i>P</i> 值(<i>Prob</i> > <i>F</i>)
0.856	0.798	0.018	<0.000 1

的相关性^[8],所以光学遥感技术不仅可以直接反演叶绿素、黄色物质和悬浮物的含量,而且可以间接反演透明度和总磷、总氮的浓度。与 Chla 建模过程类似,采用同样的方法对其它水质参数进行了遥感反演,得到如下的反演模型。为节省篇幅,下面仅给出各反演模型和主要参数,不再一一列出多元逐步回归结果和方差分析表。

透明度 SD(cm)反演模型:

$SD=9.558b1-6.239b3-478.416$ (7)

方差分析显示,相关系数 $R^2=0.753$,校正 $R^2=0.727$, $F=29$, $P<0.0001$,相关关系显著。

总氮 TN(mg/L)反演模型:

$TN=-0.069b1+0.240b5-6.058lnb5+21.237$ (8)

方差分析显示,相关系数 $R^2=0.732$,校正 $R^2=0.687$, $F=16.358$, $P<0.0001$,相关关系显著。

总磷 TP(mg/L)反演模型:

$lnTP=-81.528lnb2-5.299lnb4+19.728lnb5+0.876b2-0.461b5+249.270$ (9)

方差分析显示,相关系数 $R^2=0.710$,校正 $R^2=0.619$, $F=7.828$, $P<0.0007$,相关关系显著。

图 3 是上述各个模型在样本点处预测值和实测值的散点图,直观的显示了各模型的拟合效果。可见,Chl-a 和 SD 的拟合效果要比 TP 和 TN 的拟合效果好。

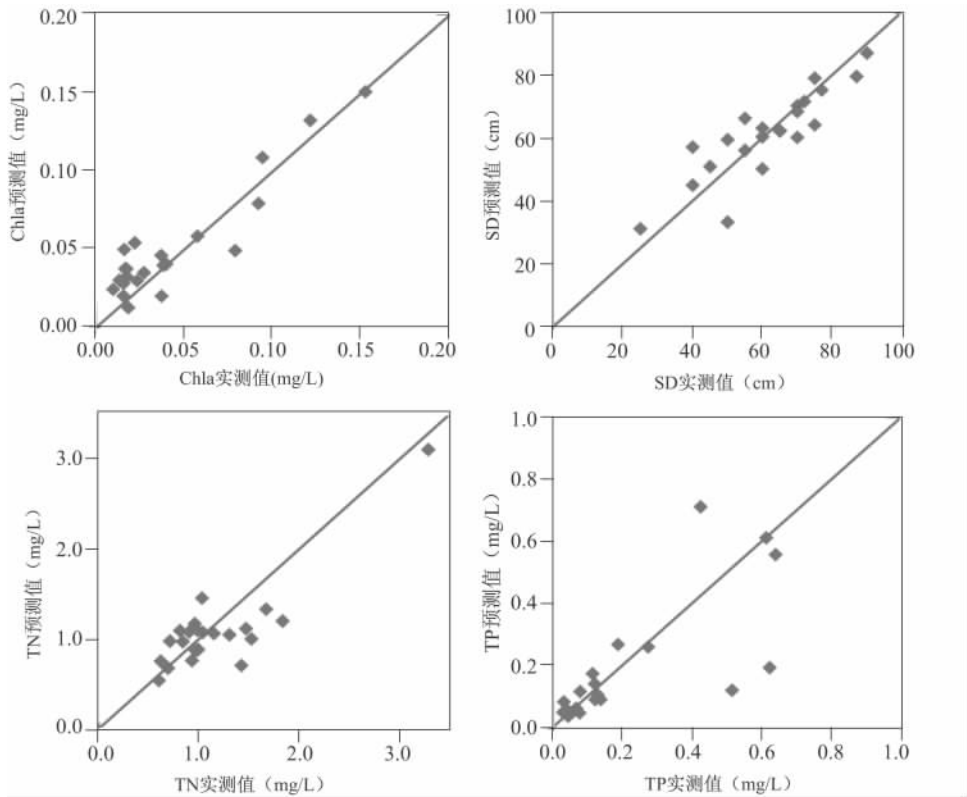


图 3 样本点各水质参数实测值与预测值对比

Fig. 3 Comparison Between Predictive Values and in Situ Monitoring Values of Water Quality Parameters at Sampling Points

4 结果和讨论

4.1 整个湖区的水质反演

将上述得到的水质反演模型用于整个东湖水体的影像进行计算、处理和等级分类,就可以得到各水质参数的空间分布图。如图 4(a)~(b)。从图 4 可看到,东湖 Chl-a、SD、TN 和 TP 的空间分布格局大体一致。这 4 个水质参数都显示,东湖水质最差的水域是西南部水果湖、庙湖,东南部的喻家湖以及北部的筲箕湖,水质相对较好的是鹰窝湖(后湖)湖中区,其次是汤林湖大部以及郭靖湖的东南部。这与东湖水质的实际情况是一致的。

从上面的模型方差分析中可看出,模型相关关系显著,利用 ETM+ 数据反演的水质参数分布图结果与东湖实际水质分布情况是符合的,这也说明采用遥感技术进行水质监测和回归分析是可行的。

4.2 影响反演精度的因素分析

上述水质参数反演模型的精度虽然已经不错,反演效果也很好,但由于一些因素的限制,其精度仍有提升的空间。首先是图像本身的问题,一是 ETM 数据是多光谱数据,波段较宽,本身是针对陆地遥感设计的,并没有专门针对水色遥感的波段;二是影像修复本身会产生一些误差。其次是样本的问题,模型的建立和验证是假定地面采样监测数据是真值,实际上这些数据本身也存在一定的误差,而且其采样时间与影像获取时间不完全一致,在空间上代表的范围也不一样,这些差异必然影响到模型的精度。其三,在所反演的 4 个水质参数中,叶绿素和透明度直接影响水体的反射光,因此是直接反演,而总磷和总氮浓度并不对太阳光的离水辐射产生影响,主要是因为它们与叶绿素和透明度有一定的相关性,所以,才能通过遥感进行反演,是间接反演,故其反演精度要比叶绿素和透明度的反演精度差一些。

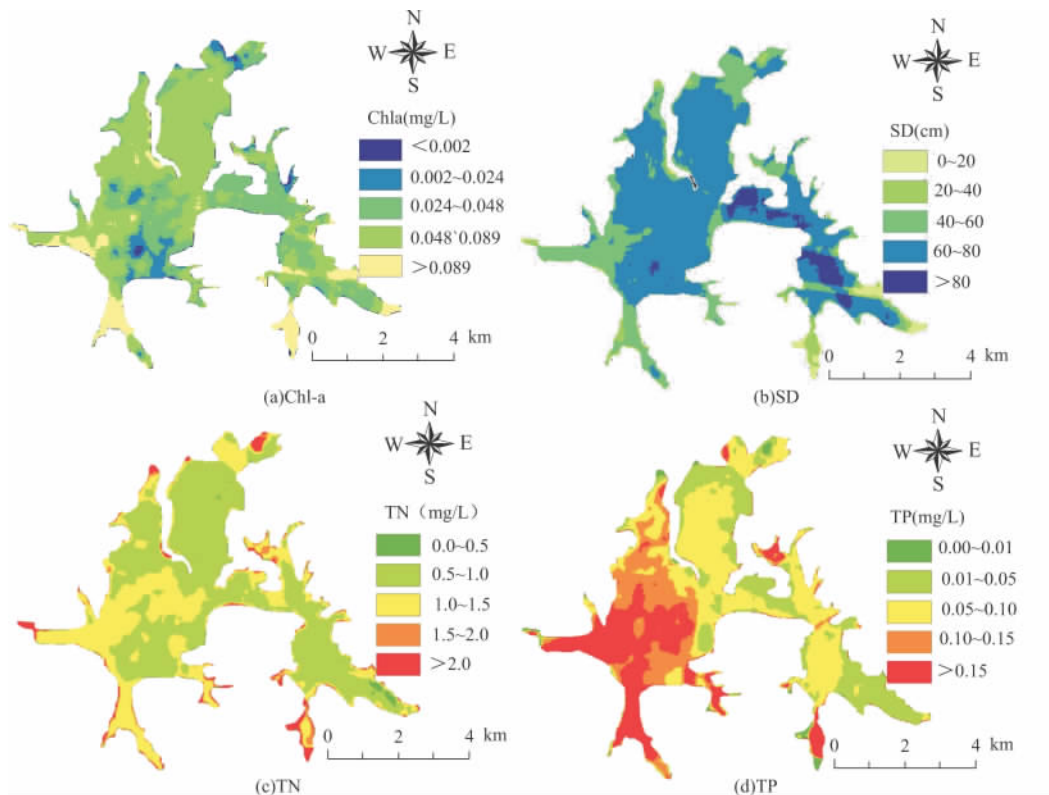


图 4 利用修复后的 Landsat7ETM+ SLC-OFF 数据反演的东湖水质参数分布图(影像获取时间为 2008 年 5 月 6 日)

Fig. 4 Maps of Water Qualities in East Lake, Wuhan Retrieving with Recovered Landsat7ETM+ SLC-OFF Data
(Date of image obtained is May 6th, 2008. (a)~(d) are distributing maps for Chl-a, SD, TN and TP respectively)

5 结论

存在缝隙的 Landsat7ETM+ SLC-OFF 卫星

遥感数据可以采用自适应局部回归匹配算法进行很好的修复。修复后的图像可以用于内陆水体的水质反演,反演模型都达到了较高的精度,可以满足实际应用的要求。由于其存档数据十分丰富,并且可以

免费下载,其空间分辨率也较高,这为低成本监测内陆水体水质提供了另外的一条途径。

东湖的遥感水质反演显示,Chla、TN、TP 和 SD 等水质参数的空间分布格局大体一致,东湖水质最差的水域是西南部水果湖、庙湖,东南部的喻家湖以及北部的筲箕湖,水质相对较好的是鹰窝湖(后湖)湖中区,其次是汤林湖大部以及郭靖湖的东南部。这个格局与城市生活污水排入东湖的分布格局是一致的。

参考文献:

- [1] 何报寅,梁胜文,杨小琴,等.利用中巴地球资源卫星数据反演武汉市湖泊营养状态指数[J].长江流域资源与环境,2009,18(12):1 181~1 186.
- [2] 刘灿德,何报寅.水质遥感监测研究进展[J].世界科技研究与发展,2005,27(5): 40~43.
- [3] ZHANG HAILIN, HE BAOYIN. Evaluation lake eutrophication with Enhanced Thematic Mapper data in Wuhan[J]. Chinese Journal of Oceanology and Limnology, 2006, 24(3): 285~290.
- [4] 张海林,何报寅,丁国平.武汉湖泊富营养化遥感调查与评价[J].长江流域资源与环境,2002,11(1): 36~39.
- [5] 寿敬文,陈雪,马建文,等.采用 ALR 算法对 Landsat-7 图像缺行修复的应用研究[J].光电子·激光,2006,17(3): 368~371.
- [6] 寿敬文.陆地卫星 7 号 ETM+ 图像数据缺行的修复与应用研究[D].哈尔滨:哈尔滨工程大学,2006.
- [7] 田晓红,林友明. Landsat-7 缝隙数据恢复的算法研究[J].计算机仿真,2007,24(12): 59~61.
- [8] 金相灿,刘树坤,章宗涉.中国湖泊环境[M].北京:海洋出版社,1995.
- [9] 王桥,张兵,韦玉春,等.太湖水体环境遥感监测实验及其软件实现[M].北京:科学出版社,2008.

RECOVERING OF LANDSAT7 ETM+ SLC-OFF DATA AND ITS APPLICATION ON WATER QUALITY RETRIEVAL IN EAST LAKE IN WUHAN

HE Bao-yin¹, DING Chao¹, YANG Xiao-qin¹, LIANG Sheng-wen²

(1. Institute of Geodesy & Geophysics, Chinese Academy of Sciences, Wuhan 430077, China;

2. Wuhan Environmental Monitoring Center, Wuhan 430051, China)

Abstract: ETM+ images from Landsat7 exists defects since May 31, 2003. This is caused by the failure of the Scan Lines Corrector on the Enhanced Thematic Mapper. These images called SLC-OFF data have some black scan lines which have no data. The lost data is about 25% of whole scene, and this makes them difficult for normal use. However, the data itself still retains some good radiometric and geometric properties; it can be used in some special purposes if they are properly repaired. This paper described how to use the Adaptive Local Regression Algorithm(ALR) to restore these images, and then used the restored image to inverse the parameters of water quality of East Lake in Wuhan, China. The results show that, SLC-OFF image could be repaired quite well by ALR algorithm, and the repaired image could be used to build quite good empirical retrieval models for water quality parameters such as Chla, SD, TP and TN in East Lake with in situ water quality monitoring data by multiple regression analysis. The square correlation coefficients(R^2) of the models were 0.86, 0.75, 0.73 and 0.71, respectively. The distributions of the retrieval water quality parameters agreed with the actual situation of the lake quite well. The author suggests that ETM+ SLC-OFF data shouldn't be neglected. They have some advantages, such as high spatial resolution, very rich in archived data and can be downloaded for free from NASA's servers. After repaired properly as the paper shows, they can be used as supplement or substitution for other remote sensing data while it is insufficient or unavailable.

Key words: Landsat7 ETM+; image recovering; ALR algorithm; water quality retrieval; East Lake in Wuhan