

CEM的波段选择方法研究及应用

陈艳拢^{1,2}, 王晓岚³, 李恩³, 宋梅萍³, 包海默^{4*}

1. 中国石油大学(华东)地球科学与技术学院, 山东 青岛 266580
2. 国家海洋环境监测中心, 辽宁 大连 116023
3. 大连海事大学信息科学技术学院, 辽宁 大连 116026
4. 大连民族大学设计学院, 辽宁 大连 116600

摘要 高光谱数据信息量丰富, 波段数量多, 能够为地物分析提供更全面的依据, 但同时也增加了数据分析的复杂性和干扰性, 尤其是水质遥感监测等低信噪比的应用领域。传统波段选择常借助相关性系数等方法, 在众多光谱波段中选择标识波段, 并在所选波段集合上进行数据分析。基于约束能量最小化(CEM)从信号检测角度提出了一种面向目标向量的波段选择算法——基于CEM的波段选择算法(CBS), 采用信号匹配滤波器从观测向量中找出与目标向量匹配度高的波段, 结合正交原理, 最大程度地选出与目标向量匹配度高且波段向量冗余度低的波段子集。以水质监测中的成分测定作为验证, 采集辽河入海口试验区的高光谱数据, 结合同步实地水样数据进行建模, 预测辽河水域氮磷含量。比较CBS算法的波段选择结果和皮尔逊相关系数(PCC)波段选择结果, 将两种方法得到的显著性波段子集作为变量进行逐步回归分析, 建立多元回归模型, 进一步对模型进行精度检验, 分析其预测值与真实值的平均相对误差。总磷浓度模型的精度检验中, 通过PCC算法选择波段得到的模型平均相对误差为20.7%, 而通过CBS算法选择波段得到的模型平均相对误差为8.17%; 总氮浓度模型的精度检验中, 通过PCC算法选择波段得到的模型平均相对误差为16.8%, 而通过CBS算法选择波段得到的模型平均相对误差为12.4%。数据分析的结果表明, CBS算法得到的波段子集, 在氮磷浓度反演的能力上, 优于传统基于相关系数的选择方法。

关键词 高光谱遥感; 波段选择; 水质监测

中图分类号: X87 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)12-3778-06

引言

高光谱遥感可以利用成像光谱仪纳米级的光谱分辨率, 获取大量窄且连续的光谱图像数据, 同步得到地物的空间、辐射和光谱信息, 在地物的细节描述和种类识别方面具有其他传感技术不可比拟的优势, 被广泛应用于土地资源利用、灾害监测、地质评估、环境保护和农林调查等领域。大量连续波段在刻画地物属性的同时, 也产生出庞大的图像数据量, 给数据分析和传输带来严重负担。波段选择技术只保留包含更多有用细节信息的波段子集, 不破坏原始数据的光谱特性和物理意义, 在相关研究中备受关注。

波段的选择原则, 通常是根据波段信息量和波段间关系进行定义, 如方差、信息熵、信息散度、相关性等, 主要以降

低数据量且保持多样性为目的^[1]; 也可以面向后续数据分析任务的特性进行定义, 如高光谱解混任务中的最大单形体体积^[2], 异常探测任务中的高阶统计量等^[3], 以加快执行速度且利于任务效果为目的。在面向任务的波段选择中, 针对监督式情况下分类任务的研究较多^[4], 而对非监督情况下目标检测、定量分析等任务的研究却相对不足^[5]。

另一方面, 水是人类赖以生存、生活、生产的基础, 然而, 随着工农业经济的高速发展, 大量的工农业废水、污水未经处理排入江、河、湖、海, 特别是与人们日常用水密切相关的内陆河、近海河口等位置。水的流动性加剧了水质的恶化, 破坏了自然生态系统, 威胁到了人类的正常用水, 水质的污染问题已经成为制约我国经济可持续发展的关键因素, 因此, 对水污染的治理以及对水体的动态监测显得极为必要与紧迫。传统的河流、湖泊水质监测主要是采用实地采

收稿日期: 2019-07-23, 修订日期: 2019-11-09

基金项目: 国家自然科学基金项目(61890964, 61971082, 61601077), 国家重点研发专项(2018YFC1407605)资助

作者简介: 陈艳拢, 1976年生, 国家海洋环境监测中心研究员, 中国石油大学(华东)地球科学与技术学院博士研究生

e-mail: ylchen@nmemc.org.cn * 通讯联系人 e-mail: bhmo@163.com

样和实验室分析等方法,需要进行现场取点采样后送至实验室,通过化学分析技术进行分析,获得局部水域的水质情况。这种监测方法虽然精确度高,但需耗费大量的人力、物力,且涵盖区域有限,不能对整体水域的水质情况进行实时监测与管理。

随着遥感技术的发展和成熟,遥感影像数据被广泛运用于水质监测中,弥补了水面采样的不足,能发现一些常规方法难以揭示的污染源的分布。国内外许多学者陆续开展了通过遥感影像数据反演并估算水质参数含量的相关研究,如王丽艳^[6]等利用 MODIS 数据反演呼伦湖水体总磷浓度并进行富营养化评价;温新龙等^[7]基于环境一号卫星数据对太湖的叶绿素 a 浓度进行反演;冯驰等^[8]利用 GOCI 影像和水体光学分类对叶绿素 a 浓度进行估算;杜成功等^[9]基于遥感数据反演太湖总磷浓度并研究其日内变化;王云霞等^[10]基于 Landsat 卫星影像研究总磷浓度反演;马驰^[11]基于遥感影像,采用回归分析的方法对松嫩平原水体的叶绿素 a 和悬浮物含量进行反演研究。随着高光谱遥感技术的不断发展,其高分辨率、多波段、图谱合一的独特优点被广泛应用于水质监测,如潘洁等^[12]通过提取高光谱遥感信息对射阳河口悬浮泥沙浓度定量反演;徐良将等^[13]利用实测光谱数据,通过微分法和波段比值法对总氮总磷浓度的反演等高光谱遥感在水质监测中的应用为水质的监测与管理提供了更多的数据支持,大大提高了水质参数的估算精度。

本研究针对待分析目标,面向数据分析任务研究有效的波段选择方法。然后,将该方法用于水质中特定成分的定量分析。使用六旋翼无人机搭载 Nano 微型机载高光谱成像仪(光谱范围为 400~1 000 nm,共有 270 个波段),以辽河口为实验区进行高光谱数据采集,结合实地采样的数据,对辽河口的水质参数含量建立反演模型,进行系统的验证与估算。

1 显著性波段选择方法

高光谱数据波段数量多,信息量丰富,但特定地物的感兴趣因素通常集中表现在有限的光谱属性中,选择对待分析目标敏感的波段组合,可以更有效地建立反演和预测模型,常用基于皮尔逊相关系数进行波段选择。

1.1 皮尔逊相关系数波段选择方法

皮尔逊相关系数(Pearson correlation coefficient, PCC),又称皮尔逊矩相关系数,是用于度量两个变量 X 和 Y 之间的相关程度,其值介于-1 与 1 之间;皮尔逊相关系数定义为两个变量之间的协方差和标准差的商,常用英文小写字母 r 代表,如式(1)

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (1)$$

传统的波段选择方法是各波段的光谱反射率信息与浓度作为两个变量,通过 PCC 度量二者之间的相关程度,选择若干具有代表性的波段进行建模。

1.2 信号匹配度的显著性波段选择方法

若将由待分析目标的定量变化值组成的向量看作目标信号,将所有光谱波段值组成的向量看作观测信号,采用信号匹配滤波器方法可以找出与目标向量匹配度高,且具有干扰抑制作用的波段集合。

约束能量最小化(constrained energy minimization, CEM)算法是有效的高光谱目标检测算法,基于该算法可以检测出与浓度向量匹配度高的波段集合。但因为波段(尤其是相邻波段)间的相关性,所检波段集合存在冗余度高的问题。为此拟结合正交原理,对候选波段集合中的观测向量进行正交投影,最大程度地选择与浓度向量匹配度高且与已有波段向量冗余度低的波段,具体算法原理和流程如下所述。

1.2.1 CEM 算法原理

CEM 算法是在已知目标而未知背景的情况下对目标进行检测的算法,其思想是先进行背景抑制然后进行匹配滤波,从而增强目标强度而抑制削弱背景或其他干扰信号,进而实现目标识别。当目标信息占总成分的比例很小的时候,CEM 算法检测效果更突出,适用于在未知的复杂背景下,对已知感兴趣目标的检测。

CEM 算法流程图如图 1 所示,基本步骤如下:

- (1) 对高光谱数据进行预处理,得到归一化后的二维数据 $\mathbf{r}(L \times N)$;
- (2) 根据高光谱数据 \mathbf{r} , 计算其自相关矩阵 \mathbf{R} ;
- (3) 确定目标向量 \mathbf{d} ;
- (4) 根据公式 $\mathbf{w}_{\text{CEM}} = \frac{\mathbf{R}^{-1}\mathbf{d}}{\mathbf{d}^T \mathbf{R}^{-1}\mathbf{d}}$, 设计 FIR 线性滤波器 $\mathbf{w} = [\omega_1, \omega_2, \dots, \omega_L]^T$;
- (5) 将归一化后的数据经过 FIR 滤波器,根据以下公式得到输出信号 y_i 。

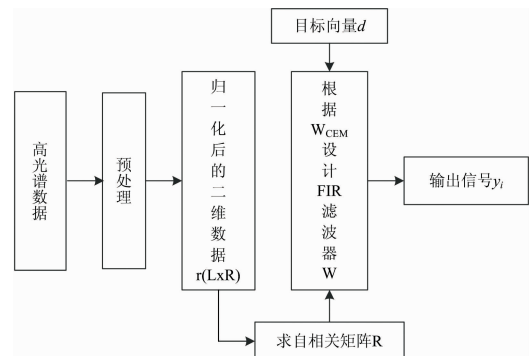


图 1 CEM 算法步骤

Fig. 1 Algorithmic steps of CEM

1.2.2 基于 CEM 的波段选择算法

本波段选择算法是利用 CEM 算法结合正交子空间投影(orthogonal subspace projection, OSP)进行波段选择。

正交子空间投影是将端元光谱矩阵 \mathbf{M} 分为两部分:感兴趣部分和非感兴趣部分,即 $\mathbf{M}=[\mathbf{P}, \mathbf{U}]$, \mathbf{P} 代表感兴趣部分, \mathbf{U} 代表非感兴趣部分,在众多数据中,为了突出感兴趣部分而对非感兴趣部分进行抑制,OSP 算法针对于非感兴趣部分 \mathbf{U} 构造正交投影算子表达形式如(2)

$$P_{\perp}^U = I - \mathbf{U}\mathbf{U}^{\#}$$

$$U^\# = (U^T U)^{-1} U^T \quad (2)$$

正交投影算子 P_U^\perp 应用于高光谱数据, 则高光谱数据将会被投影到 U 的正交投影空间中, 数据在正交空间中表达形式如(3)

$$P_U^\perp x = P_U^\perp P_s p + P_U^\perp w \quad (3)$$

由表达式可以看出, 在此空间中, 非感兴趣的部分 U 被有效抑制, 且噪声被抑制为 $P_U^\perp w$ 。

将感兴趣目标的定量属性向量作为原始信号, 即目标向量 d , 利用 CEM 算法选出匹配程度高的波段信号, 将此波段作为 OSP 中的非感兴趣目标, 通过正交投影算子 P_U^\perp 对该目标进行抑制, 目的是为了压制与当前波段类似的敏感波段, 然后在抑制后的背景中再次通过 CEM 算法选择匹配程度高的波段, 重复上述过程若干次, 选出若干个具有代表性的波段, 算法流程图如图 2。

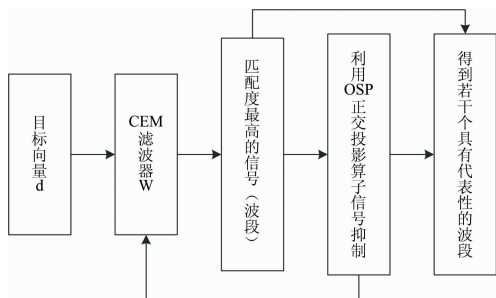


图 2 CBS 算法步骤

Fig. 2 Algorithmic steps of CBS

2 实验部分

2.1 研究区概况

辽河是中国七大河流之一, 位于中国东北地区南部, 界于东经 $117^{\circ}00' - 125^{\circ}30'$, 北纬 $40^{\circ}30' - 45^{\circ}10'$ 之间, 发源于河北省平泉县七老图山脉的光头山, 流经河北、内蒙古、吉林、辽宁, 全长 1 345 km, 注入渤海, 被称为辽宁人民的“母亲河”。但近年来由于大量非法排污, 辽河成为我国污染最严重的河流之一, 导致辽河水域生物无法存活, 无法灌溉农业, 威胁到了当地居民正常饮水。故本研究区域选在辽宁省盘锦市盘山红海岸附近, 位于辽河的入海口, 此处有红海滩景观和世界最大的芦苇荡, 是国家级自然保护区, 水质的好坏也影响到了保护区的生态环境。

2.2 数据获取及处理

2.2.1 水样采集与分析

2018 年 10 月 11 日在研究区内进行现场取样, 规划了 8 个点位进行现场水样的采集, 将取样用的采样瓶置入水中, 采样过程中严禁水中杂质进入采样瓶, 采集结束后做好密封操作, 将采样瓶置于阴凉处, 待八个点均取样结束后, 尽快将水样送回实验室进行化学分析与水质参数含量测定。

2.2.2 高光谱遥感数据的采集

2018 年 10 月 11 日(天气晴朗, 采光条件良好), 在现场水样提取的同时, 进行高光谱数据的采集。将 Nano 微型机载高光谱成像仪搭载在大疆公司的六旋翼无人机 MA-

TRICE600PRO 上, 飞行区域大小为 $1.5 \text{ km} \times 0.5 \text{ km}$, 飞行高度设置在飞行区上空 400 m 处, 飞行速度控制在 $7.7 \text{ m} \cdot \text{s}^{-1}$ 。采集结束后, 利用 Hyperspec III 高光谱数据分析软件对图像进行反射率校准和几何校准, 然后在 ENVI 遥感图像处理平台上对图像进行拼接与裁剪操作, 得到涵盖 8 个采样点的高光谱图像。

3 结果与讨论

3.1 总磷含量的反演

现场一共采取 8 个水样, 选择 5 个样点用于总磷反演模型的构建, 另外 3 个样点用来检验模型的反演精度。

3.1.1 基于 PCC 波段选择的总磷反演模型的构建

根据各个样点的经纬度信息在高光谱遥感图像上获取对应点的光谱反射率, 再与总磷浓度进行皮尔逊相关性, 如图 3 所示为总磷浓度与光谱反射率在不同波段处的相关程度。

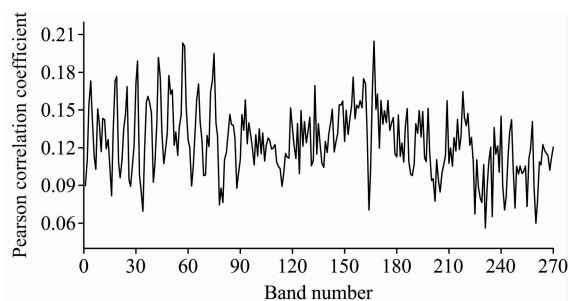


图 3 总磷浓度与反射率在不同波段的相关程度

Fig. 3 Relevance between concentration of TP and reflectivity in different bands

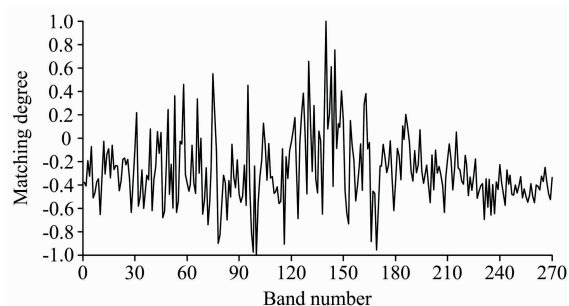


图 4 各波段与总磷浓度矩阵的匹配程度

Fig. 4 Matching degree of each band with TP concentration matrix

选择 10 个相关系数较高的波段进行模型的构建, 分别为: $R_{19}, R_{31}, R_{43}, R_{49}, R_{57}, R_{58}, R_{75}, R_{155}, R_{161}, R_{167}$, 其中, $R_i (i=1, \dots, 270)$ 代表各波段对应的反射率, 考虑到总磷浓度受多个变量的影响, 故将这些波段同时作为变量进行逐步回归分析, 建立多元线性回归模型, 其基本思想是将变量逐个引入模型, 进行 F 检验, 对选入的变量逐个进行 T 检验, 若检验表明回归效果显著, 则引入回归方程, 若回归效果不显著, 则剔除作用不显著的变量, 更新回归方程, 这一过程反复迭代, 直到没有不显著变量从回归方程中剔除, 也

没有显著变量引入为止。实验最终筛选出 3 个显著变量，即 3 个敏感波段，分别为 R_{19} , R_{31} , R_{43} ，多元回归模型为 $Y = -149\ 835R_{19} - 5\ 965.82R_{31} + 83\ 337.6R_{43} + 62.859\ 5$ 其中， Y 为浓度预测值，该模型的拟合度 R^2 为 0.975 27，均方根误差 RMSE 为 3.544 7，表明该模型拟合效果良好。

3.1.2 基于 CEM 波段选择的总磷反演模型的构建

图 4 是利用 CEM 算法得到的各波段与总磷浓度矩阵的匹配程度，选择的匹配程度最高的波段信号 R_{140} 。

最终通过显著性波段选择算法选择出 6 个具有代表性的波段，分别为 R_{140} , R_{80} , R_{95} , R_{161} , R_{94} , R_{131} ，同理将这些波段同时作为变量进行逐步回归分析，建立多元线性回归模型，最终筛选出 3 个敏感波段，分别为 R_{95} , R_{161} , R_{131} ，多元

回归模型为

$$Y = 7\ 524.18R_{95} + 7\ 960.14R_{161} - 15\ 614.8R_{131} + 106.823$$

该模型的拟合度 R^2 为 0.866 899，均方根误差 RMSE 为 8.223 54，表明该模型拟合效果较好。

3.1.3 总磷反演模型精度检验

前两小节分别利用 PCC 和基于 CEM 的波段选择算法 (CEM-based Band Selection, CBS) 两种方法对高光谱图像进行波段选择，然后利用选择好的波段进行建模，本小节对 2 种方法反演的模型进行精度检验，表 1 是 3 个检验样点在两种方法下的总磷浓度的反演值与实测值的对比，其中，相对误差为 $|(反演值 - 实测值)| / 实测值$ 。

表 1 在 2 种方法下的总磷浓度的反演值与实测值的对比

Table 1 Comparison of inversion values and measured values of TP concentration using two methods

样点 编号	实测总磷浓度 / $(\mu\text{g} \cdot \text{L}^{-1})$	PCC		CBS	
		反演总磷浓度 / $(\mu\text{g} \cdot \text{L}^{-1})$	相对误差 /%	反演总磷浓度 / $(\mu\text{g} \cdot \text{L}^{-1})$	相对误差 /%
2	92.544 8	71.643 2	22.6	95.435 9	3.1
4	94.231 44	114.032 9	21.5	104.892 0	11.3
8	100.166 58	82.021 2	18.1	90.049 7	10.1

由表 1 可知，利用 CBS 方法选择波段构造的反演的模型，虽然拟合度不高，但预测值更为准确，反演的浓度更接近实测值，相对误差最高为 11.3%，最低为 3.1%，平均相对误差为 8.17%。

3.2 总氮含量的反演

同总磷，选择 5 个样点用于总氮反演模型的构建，另外 3 个样点用来检验模型的反演精度。

3.2.1 基于 PCC 波段选择的总氮反演模型的构建

将各波段的光谱反射率与总氮浓度进行皮尔逊相关性分析，如图 5 所示为总氮浓度与光谱反射率在不同波段处的相关程度。

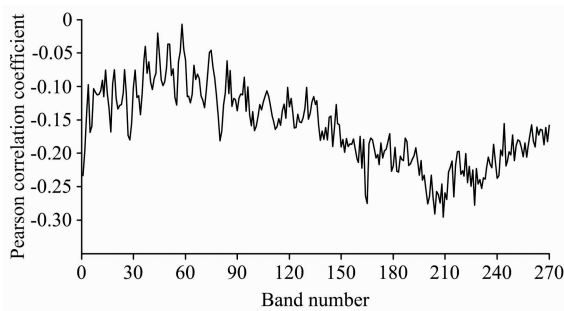


图 5 总氮浓度与反射率在不同波段的相关程度

Fig. 5 Relevance between concentration of TN and reflectivity in different bands

选择 10 个相关系数较高的波段建模，这 10 个波段分别为： R_{164} , R_{165} , R_{200} , R_{203} , R_{204} , R_{207} , R_{209} , R_{211} , R_{215} , R_{227} ，同理，这些波段同时作为变量进行逐步回归分析，建立多元线性回归模型，最终筛选出 2 个变量，即 2 个敏感波段，分别为 R_{209} , R_{215} ，多元回归模型为

$$Y = -1\ 300\ 870R_{209} + 1\ 453\ 550R_{215} + 898.555$$

该模型的拟合度 R^2 为 0.965 618，均方根误差 RMSE 为 53.550 9，表明该模型拟合效果良好。

3.2.2 基于 CEM 波段选择的总氮反演模型的构建

图 6 是利用 CEM 算法得到的各波段与总氮浓度矩阵的匹配程度，选择的匹配程度最高的波段信号为 R_{80} 。

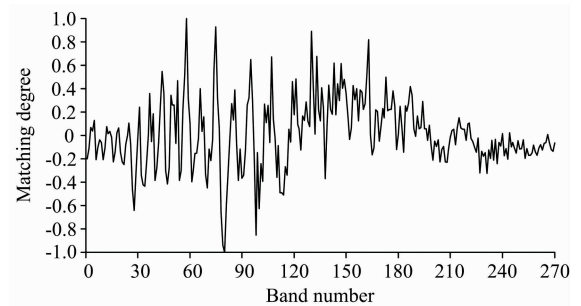


图 6 各波段与总氮浓度矩阵的匹配程度

Fig. 6 Matching degree of each band with TN concentration matrix

最终通过显著性波段选择算法选择出 6 个具有代表性的波段，分别为 R_{80} , R_{58} , R_{156} , R_{95} , R_{94} , R_{78} ，同理将这些波段同时作为变量进行逐步回归分析，建立多元线性回归模型，最终筛选出 3 个敏感波段，分别为 R_{80} , R_{58} , R_{156} ，多元回归模型为

$$Y = -158\ 206R_{80} + 115\ 568R_{58} + 98\ 478.1R_{156} + 1\ 048.31$$

该模型的拟合度 R^2 为 0.994 887，均方根误差 RMSE 为 29.206，表明该模型拟合效果较好。

3.2.3 总氮反演模型精度检验

前两小节分别利用 PCC 和 CBS 两种方法对高光谱图像

进行波段选择, 然后利用选择好的波段进行建模, 本小节对 2 种方法下的总氮浓度的反演值与实测值的对比。2 种方法反演的模型进行精度检验, 表 2 是 3 个检验样点在

表 2 在 2 种方法下的总氮浓度的反演值与实测值的对比

Table 2 Comparison of inversion values and measured values of TN concentration using two methods

样点 编号	实测总氮浓度 $/(\mu\text{g} \cdot \text{L}^{-1})$	PCC		CBS	
		反演总氮浓度 $/(\mu\text{g} \cdot \text{L}^{-1})$	相对误差 /%	反演总氮浓度 $/(\mu\text{g} \cdot \text{L}^{-1})$	相对误差 /%
2	1 738.947	1 917.0	10.2	1 815.4	4.4
4	1 793.186	1 600.7	10.7	1 622.0	9.5
5	1 997.624	1 408.8	29.5	2 465.9	23.4

由表 2 可知, 利用 CBS 方法选择波段反演的模型更为准确, 反演的浓度最接近实测值, 相对误差最高为 23.4%, 最低为 4.4%, 平均相对误差为 12.4%。

4 结 论

基于高光谱遥感数据, 研究有效波段的选择方法, 提出了一种基于 CEM 的浓度敏感波段选择方法, 深入探讨分析了不同波段选择方法对氮磷浓度反演模型精度的影响。

(1) 总磷含量的反演中, 对比了 PCC 和 CBS 两种波段选择方法下反演模型的精度, 利用 CBS 方法选择波段构造的反

演模型, 虽然拟合度不高, 但预测值更为准确, 平均相对误差为 8.17%。

(2) 总氮含量的反演中, 利用 CBS 方法选择波段构造的反演模型, 拟合效果优于 PCC, 且预测值也更为准确, 平均相对误差为 12.4%。

本工作的创新之处在于提出了基于 CEM 的浓度敏感波段选择方法, 该方法能够最大程度地选择与浓度向量匹配度高且与已有波段向量冗余度低的波段, 氮磷数据的分析比较表明了 CBS 的波段选择能力优于 PCC, 这一方法对高光谱数据的波段选择也具有一定的指导作用。

References

- [1] Yang C, Bruzzone L, Zjao H, et al. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(12): 7230.
- [2] Wang L, Zhang Y, Gu Y. Unsupervised Band Selection Method Based on Improved N-FINDR Algorithm for Spectral Unmixing. 2006 1st International Symposium on Systems and Control in Aerospace and Astronautics, Harbin, China, 2006.01.19.
- [3] Geng X, Sun K, Ji L, et al. Scientific Reports, 2015, 5: 9915.
- [4] Yu C, Song M, Chang C I. Remote Sensing, 2018, 10(1): 113.
- [5] Geng X, Sun K, Ji L. Remote Sensing Letters, 2014, 5(2): 1022.
- [6] WANG Li-yan, LI Chang-you, SUN Biao(王丽艳, 李畅游, 孙 标). Chinese Journal of Environmental Engineering(环境工程学报), 2014, 8(12): 5527.
- [7] WEN Xin-long, JING Yuan-shu, LI Ya-chun, et al(温新龙, 景元书, 李亚春, 等). Environmental Science & Technology(环境科学与技术), 2014, 37(10): 149.
- [8] FENG Chi, JIN Qi, WANG Yan-nan, et al(冯 驰, 金 琦, 王艳楠, 等). Environmental Science(环境科学), 2015, 36(5): 1557.
- [9] DU Cheng-gong, LI Yun-mei, WANG Qiao, et al(杜成功, 李云梅, 王 桥, 等). Environmental Science(环境科学), 2016, 37(3): 862.
- [10] WANG Yun-xia, YANG Guo-fan, LIN Mao-sen, et al(王云霞, 杨国范, 林茂森, 等). Journal of Irrigation and Drainage(灌溉排水学报), 2017, 36(4): 105.
- [11] MA Chi(马 驰). Wetland Science(湿地科学), 2017, 15(2): 173.
- [12] PAN Jie, ZHANG Ying, TAN Zi-hui(潘 洁, 张 鹰, 谭子辉). Marine Sciences(海洋科学), 2011, 35(9): 85.
- [13] XU Liang-jiang, HUANG Chang-chun, LI Yun-mei, et al(徐良将, 黄昌春, 李云梅, 等). Remote Sensing Technology and Application(遥感技术与应用), 2013, 28(4): 681.

Research and Application of Band Selection Method Based on CEM

CHEN Yan-long^{1,2}, WANG Xiao-lan³, LI En³, SONG Mei-ping³, BAO Hai-mo^{4*}

1. College of Geosciences and Technology, China University of Petroleum (East China), Qingdao 266580, China

2. National Marine Environment Monitoring Center, Dalian 116023, China

3. College of Information Science and Technology, Dalian Maritime University, Dalian 116026, China

4. College of Design, Dalian Minzu University, Dalian 116600, China

Abstract Hyperspectral data is rich in information and bands, which can provide a more comprehensive basis for geophysical analysis, but at the same time, it also increases the complexity and interference of data analysis, especially in low signal-to-noise ratio applications such as remote sensing monitoring of water quality. Traditional band selection often uses correlation coefficient and other methods to select the identification band in many spectral bands and to analyze the data on the selected band set. In this paper, based on the constrained energy minimization (CEM), a target-oriented band selection algorithm is proposed, which is called CEM-based band selection (CBS). The signal matching filter is used to find the band with a high matching degree with the target vector from the observation vector, and then combined with the orthogonal principle to maximize the selection of a subset of bands that have a high degree of matching with the target vector and low redundancy of the band vector. Based on the determination of the components in the water quality monitoring, the hyperspectral data of the Liaohe estuary test area was collected and combined with the synchronous field water sample data to predict the nitrogen and phosphorus content in the Liaohe waters. Comparing the band selection results of the CBS algorithm with the band selection results of the Pearson correlation coefficient (PCC), the significant band subsets obtained by the two methods are used as variables to carry out stepwise regression analysis, and multiple regression models are established to further test the accuracy of the model and analyze the average relative error between the predicted value and the true value. In the accuracy test of the total phosphorus concentration model, the average relative error of the model obtained by the PCC algorithm is 20.7%, and the average relative error of the model obtained by the CBS algorithm is 8.17%. In the accuracy test of the total nitrogen concentration model, the average relative error of the model obtained by the PCC algorithm is 16.8%, and the average relative error of the model obtained by the CBS algorithm is 12.4%. The results of the data analysis show that the band subset obtained by the CBS algorithm is superior to the traditional selection method based correlation coefficient in the ability of nitrogen and phosphorus concentration inversion.

Keywords Hyperspectral remote sensing; Band Selection; Water quality monitoring

(Received Jul. 23, 2019; accepted Nov. 9, 2019)

* Corresponding author