

# 光谱二阶微分波动指数的河道水质类型识别方法研究

李 澜<sup>1, 2, 3</sup>, 田 华<sup>4</sup>, 季铁梅<sup>4</sup>, 巩彩兰<sup>1, 2\*</sup>, 胡 勇<sup>1, 2</sup>, 王歆晖<sup>1, 2, 3</sup>, 何志杰<sup>1, 2, 3</sup>

1. 中国科学院上海技术物理研究所, 上海 200083
2. 中国科学院红外探测与成像技术重点实验室, 上海 200083
3. 中国科学院大学, 北京 100049
4. 上海市水文总站, 上海 200232

**摘 要** 城市河道水资源是重要的生态资源, 近年来城市工业的不断发展, 导致河道水污染问题日益突出。治理水污染的前提条件是及时掌握河道水质状况(即水质类型), 传统采样化验的监测方法精度高, 但较费时费力, 该研究提出一种基于光谱二阶微分波动指数的水质类型快速识别方法, 可实现城市河道水质类型的快速排摸。首先使用光谱差分计算得到光谱二阶微分曲线, 对曲线进行平滑处理消除噪声和其他干扰; 之后使用滑动窗口提取曲线上的局部极大值和极小值点, 并设置最小距离阈值逐步去除虚假极值点, 再使用三次样条插值法得到包含光谱二阶微分曲线的双包络线, 最后利用上下包络线计算得到光谱二阶微分的波动指数曲线。通过对各类样本的波动曲线分析后发现, 在 720~740, 750~770 以及 820~840 nm 处各类水体的波动指数差异较大, 随后统计了各类水体光谱在这三波段内的平均波动指数的均值、标准差等统计特征, 发现平均波动指数与水质级别具有正向相关关系, 水质级别越高水质状况越差, 平均波动指数也越大。为验证光谱二阶微分波动指数可用于城市河道水质的快速识别, 将各类水体光谱样本随机划分为训练集和测试集, 结合 LSSVM(least-squares support vector machine)构建水质类型识别模型, 将上述三个特征波段的平均波动指数作为特征输入, 经测试, 各类样本的平均识别准确率达 80.65%, 误差不超过一类的识别精度超过 95%。基于光谱二阶微分波动指数的河道水质高光谱识别方法识别精度较高, 可作为城市河道水质类型快速排摸的一种辅助技术手段。

**关键词** 河道水质; 高光谱; 光谱二阶微分; 遥感识别

**中图分类号:** O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)05-1645-05

## 引 言

城市河道是城市生态环境的重要组成部分, 改革开放以来我国城市经济快速发展, 但也造成河道出现不同程度的污染<sup>[1-2]</sup>。城市河道水污染治理的前提和关键是水质状况的实时监测; 目前水质监测方法以传统采样监测为主遥感监测为辅, 传统的水质监测通过实验室化验确定水质指标浓度再根据相关标准确定水质级别; 遥感监测方法是通过卫星、飞机、无人机等平台搭载的传感器获取河道区域的遥感影像, 再经过一系列处理分析建立水质参数与光谱反射率之间的定量关系, 进而确定河道的水质级别。传统采样监测技术已较为成熟, 但存在费时、费力、只能获得断面点位水质等不足。

随着遥感和高光谱技术的发展, 可以准确、快速地获取各种地物的各波段的光谱信息, 国内外学者提出了众多基于遥感反演水质参数进而判断水质类型的研究方法<sup>[3-6]</sup>。王丽艳等使用 MODIS 数据反演了呼伦湖水体 COD 浓度并确定水质级别<sup>[7]</sup>。Wang 等提出了一种综合水质评价方法, 并利用该方法反演了多个水质参数浓度<sup>[8]</sup>。但河道水质类别的划分指标如溶解氧(DO)、总磷(TP)、氨氮(NH<sub>3</sub>-N)等不具有明显的光学特性, 通过遥感反演各水质指标进而判断水质类型这类方法的局限性在于最终级别的识别精度受各个水质指标反演精度共同决定, 单一指标反演精度不高必然整体识别精度低下, 模型应用局限性大、泛化能力不强, 适用性不足。

水体反射率是诸多水质参数综合作用的结果, 水体的光谱特征与水质类型之间存在某种映射关系。提出一种基于

收稿日期: 2019-05-05, 修订日期: 2019-09-21

基金项目: 国家重点研发计划项目(2017YFC0602100, 2017YFC0602103), 上海市水务局科研项目(沪水科 2018-07)资助

作者简介: 李 澜, 1994 年生, 中国科学院上海技术物理研究所博士研究生 e-mail: llan@mail.ustc.edu.cn

\* 通讯联系人 e-mail: gcl@mail.sitp.ac.cn

光谱二阶微分波动指数的光谱分析方法,能有效反映各波段光谱二阶微分的波动情况,在此基础上可实现城市河道水质类型的快速识别,可推广应用于利用其他遥感平台获得的高光谱数据监测内陆水体的水质类型。

## 1 实验部分

基于光谱二阶微分波动特性的水质光谱特征提取及分类方法,其基本思想在于利用上下包络线来提取光谱二阶微分曲线的振动范围,具体的数据处理流程如图 1。

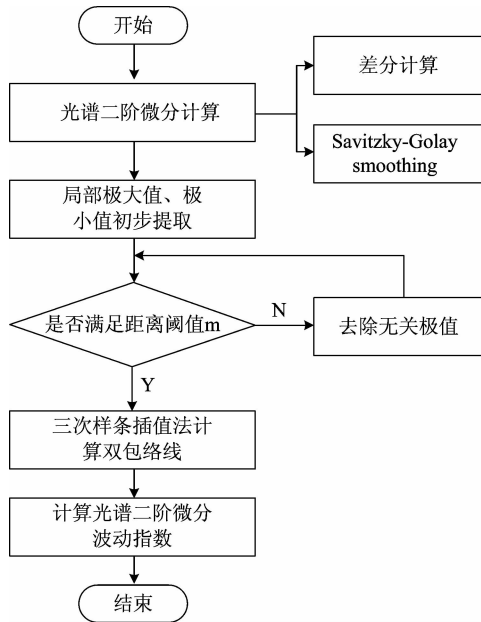


图 1 基于光谱二阶微分波动指数计算流程图

Fig. 1 Flow chart of second-order differential fluctuation index calculation

(1)光谱二阶微分计算。光谱二阶微分计算包括差分计算和平滑两部分,首先利用光谱差分计算原始光谱的二阶微分,随后使用 Savitzky-Golay Smoothing 法<sup>[9]</sup>对原始二阶光谱进行平滑处理,消除噪点和其他干扰。

(2)局部极大值极小值提取。曲线的波峰和波谷通常对应局部极大值和极小值,而波峰波谷的值可以反映曲线局部的变化范围,使用  $3 \times 3$  的滑动窗口逐点提取局部极值,满足式(1)或(2)的即为局部极大值或极小值点。

$$R''_{\lambda} > R''_{\lambda-1} \text{ 且 } R''_{\lambda} > R''_{\lambda+1} \quad (1)$$

$$R''_{\lambda} < R''_{\lambda-1} \text{ 且 } R''_{\lambda} < R''_{\lambda+1} \quad (2)$$

式(1)和式(2)当中,  $R''_{\lambda}$  表示窗口中心点的光谱二阶微分,  $R''_{\lambda+1}$  和  $R''_{\lambda-1}$  则表示中心点相邻点的光谱二阶微分。

(3)无关极值点去除。曲线上的波峰波谷对应着局部极值点,通常光谱二阶微分曲线某波段附近只会对应一个峰值,由于仪器探测性能限制或其他原因的影响,峰值周围会出现一系列无关极值点,这会对后续处理产生干扰,为此可设定一个距离阈值  $m$ ,采用循环迭代方式去除无关极值点,下面以极大值为例介绍具体实现步骤。①寻找当前所有极大

值中最大值以及对应的波长位置;②去除最大值周围距离小于阈值  $m$  的所有极大值点;③弹出并保存最大值和波长位置。重复①②③,直到所有的无关极大值点均消除。无关极小值的去除与上述步骤类似,只需每次弹出最小值和位置即可。

(4)双包络线生成。在上一步处理后可得到光谱二阶微分曲线上的各波峰点和波谷点,使用三次样条插值法得到光谱二阶微分曲线的上下两条包络线。图 2 是某光谱二阶微分及双包络线示意图。

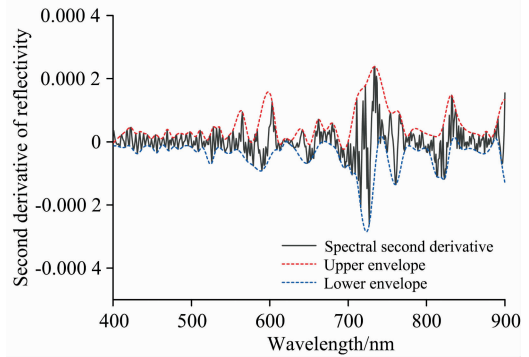


图 2 光谱二阶微分曲线与双包络线

Fig. 2 Second-order differential curve and double envelope

(5)光谱二阶微分波动指数计算。上下包络线能表示光谱二阶微分的振动范围,而某一波段处的对应上下包络线函数值之间的差值能表示该波段附近光谱二阶微分变化的剧烈程度。因此可定义一种光谱二阶微分波动特性指数来描述这一特征,具体计算公式如式(3)

$$F(\lambda) = e^{A(f(\lambda)-g(\lambda))} \quad (3)$$

式(3)中,  $F(\lambda)$  表示波长为  $\lambda$  处附近的变化特征,  $f(\lambda)$  和  $g(\lambda)$  分别对应该波段上下包络线的函数值,  $A$  是固定的放大系数用于增加区分度。

### 1.1 水体光谱采集

先后多次在上海市嘉定区采集了各水质类型水体的光谱数据,光谱数据获取采用 ASD 公司生产的 FieldSpec3 便携式地物光谱仪,波长范围选择为 400~900 nm。测量和处理参照国家标准《水体可见光-短波红外光谱反射率测量 GB/T 36540—2018》进行。图 3 是获取的水体反射光谱曲线图。

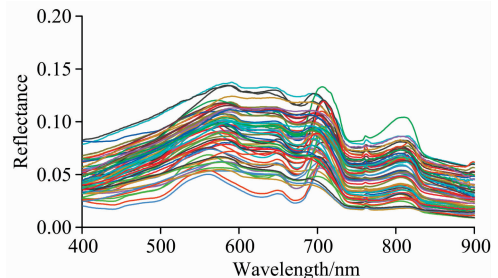


图 3 反射光谱曲线图

Fig. 3 Reflectance spectral curves

### 1.2 水质采样与处理

采集光谱数据的同时,还需要获取相应的水质数据。将

现场采样的水体样本密封保存带回实验室化验获得各水质指标,再根据《地表水环境质量评价办法(试行)》等标准划分类别,共得到 I 和 II 类样本 11 个(由于 I 类样本稀少,将其与 II 类样本合并)、III 类 14 个、IV 类 20 个、V 类 23,劣 V 类 20 个。

## 2 结果与讨论

### 2.1 特征波段选择

采用上述的数据处理方法计算各类水体的光谱二阶微分波动系数,合适的距离阈值和放大系数可以有效去除虚假极值点,增加不同类别之间的区分度,经多次试验最小距离阈值  $m$  设置为 10,放大系数  $A$  设为 500 效果较好,图 4 是计算出的各类水体的平均光谱二阶微分波动指数曲线图。

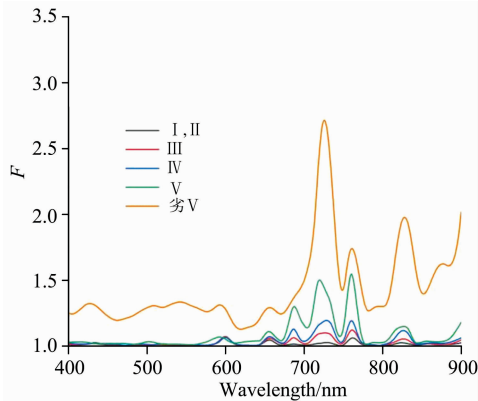


图 4 各类型水体的平均波动指数

Fig. 4 Average fluctuation index of each type of water body

从图 4 可以看出,各类水体波动指数在 720~740, 750~770 以及 820~840 nm 处均有明显的峰值,且不同水质类型水体峰值高度不同,随后统计了各类样本中这三个波段内的平均波动指数的均值、标准差等特征,结果见表 1—表 3。

上述统计结果表明,水体在 720~740, 750~770 和 820~840 nm 内的光谱二阶微分平均波动系数与水质级别具有正向相关关系,随着水质级别的上升,这三个波段内的平均波动指数的平均值和中位数也随之上升,且水质级别越高上

表 1 各类型水体 720~740 nm 平均波动指数统计

Table 1 Average fluctuation index from 720~740 nm

类别	最大值	最小值	平均值	中位数	标准差
II	1.045 4	1.011 6	1.027 1	1.025 9	0.012 4
III	1.233 9	1.019 4	1.091 1	1.087 2	0.052 8
IV	1.298 4	1.137 0	1.170 5	1.173 7	0.079 9
V	1.545 1	1.253 5	1.467 7	1.482 5	0.132 9
劣 V	3.130 8	1.369 8	2.535 3	2.471 9	1.036 2

表 2 各类型水体 750~770 nm 平均波动指数统计

Table 2 Average fluctuation index from 750~770 nm

类别	最大值	最小值	平均值	中位数	标准差
II	1.063 6	1.026 7	1.046 5	1.052 2	0.013 6
III	1.187 5	1.059 8	1.081 8	1.087 2	0.034 7
IV	1.277 1	1.091 6	1.131 6	1.123 5	0.085 5
V	1.600 4	1.151 2	1.499 8	1.474 2	0.145 7
劣 V	3.831 5	1.352 7	1.614 5	1.652 2	0.411 1

表 3 各类型水体 820~840 nm 平均波动指数统计

Table 3 Average fluctuation index from 820~840 nm

类别	最大值	最小值	平均值	中位数	标准差
II	1.032 1	1.011 7	1.023 8	1.024 0	0.008 1
III	1.068 0	1.033 0	1.047 1	1.047 5	0.010 2
IV	1.110 1	1.078 3	1.093 9	1.097 8	0.014 0
V	1.192 3	1.121 7	1.151 4	1.160 9	0.039 4
劣 V	2.387 4	1.288 0	1.920 5	1.934 6	0.488 7

升幅度越大,变化越剧烈,同样平均波动指数的标准差也在增加。因此可利用这一特性实现河道水质类型的自动识别分类。

### 2.2 水质类型识别模型建立及验证

首先将各类样本按照 2 : 1 的比例随机划分为测试集和训练集,随后采用本方法计算训练集内各样本的二阶微分波动指数曲线,得到三个特征波段内的平均波动指数,最后以这三个特征波段的平均波动指数为输入,使用 LSSVM<sup>[10]</sup> 构建水质类型识别模型,训练完成后进行测试,取重复 10 次实验的平均值作为最终结果,具体结果见图 5(a)。

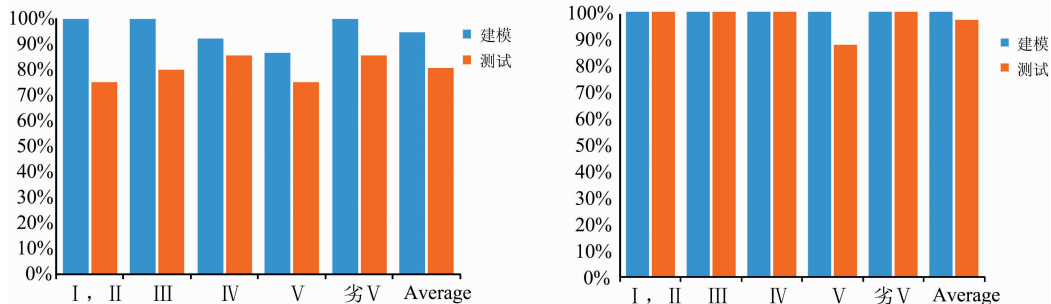


图 5 (a)模型识别精度; (b)误差不超过 1 类的模型识别精度

Fig. 5 (a)Model recognition accuracy; (b)Model recognition accuracy of no more than 1 class

图 5(a)显示平均识别准确度达 80.65%，而通过对误分类的样本进行分析后发现，误分类样本多集中在两类样本特征边界处，故单独测试了分类误差不超过 1 类的识别准确度，重复测试 10 次，如图 5(b)所示。结果表明误差不超过 1 类的平均识别准确度达 96.77%，在 31 个测试样本中，仅 1 个样本分类误差超过 1 类，因此光谱二阶微分波动系数结合机器学习模型可实现城市河道水质的快速识别，且识别误差不超过 1 类。

### 3 结 论

针对目前城市水体大面积水质类型调查的需求，提出了

基于光谱二阶微分波动指数的高光谱数据处理和分析方法，在对各类型水体的波动指数分析后得到如下结论：(1)720~740, 750~770 和 820~840 nm 波段内的平均波动指数与水质级别具有正相关关系，随着水质级别的上升，平均波动指数数值升高，对应波段附近的光谱二阶微分振幅越大，这表明水质状况越差的水体，其光谱二阶微分波动越大。(2)以三个特征波段的平均波动指数作为输入特征，采用 LSSVM 模型构建水质类型识别模型，经测试，模型的平均识别准确度达 80.65%，误差不超过一类的识别准确度达 96.77%，能较准确地识别城市河道水质类型。

### References

- [ 1 ] SUN Yu-ping, YU Heng-guo, ZHOU Qin, et al(孙玉平, 于恒国, 周 钦, 等). Journal of Lake Sciences(湖泊科学), 2018, 30(2): 375.
- [ 2 ] ZHANG Yuan-yuan, ZHANG Wei-jia(张媛媛, 张维佳). Environmental Sciences and Management(环境科学与管理), 2007, 32(12): 66.
- [ 3 ] ZHANG Bing, SHEN Qian, LI Jun-sheng, et al(张 兵, 申 茜, 李俊生, 等). Journal of Lake Sciences(湖泊科学), 2009, 21(2): 182.
- [ 4 ] YANG Shi-zhi(杨世植). Journal of Atmospheric and Environmental Optics (大气与环境光学学报), 2002, 15(1): 1.
- [ 5 ] Tan J, Cherkauer K A, Chaubey I. International Journal of Remote Sensing, 2015, 36(21): 5466.
- [ 6 ] Giardino C, Brando V E, Dekker A G, et al. Remote Sensing of Environment, 2007, 109(2): 183.
- [ 7 ] 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.
- [ 8 ] Wang K, He P H, Dong Y, et al. Procedia Environmental Sciences, 2011, 10: 2511.
- [ 9 ] Kitamura K, Hozumi K. Analytica Chimica Acta, 1987, 201: 301.
- [10] Suykens J A K, Lukas L, Vandewalle J. Sparse Approximation Using Least Squares Support Vector Machines. IEEE International Symposium on Circuits and Systems ISCAS'2000.

## Study on River Water Quality Type Identification Method Based on Fluctuation Index of Second-Order Differential Spectra

LI Lan<sup>1, 2, 3</sup>, TIAN Hua<sup>4</sup>, JI Tie-mei<sup>4</sup>, GONG Cai-lan<sup>1, 2\*</sup>, HU Yong<sup>1, 2</sup>, WANG Xin-hui<sup>1, 2, 3</sup>, HE Zhi-jie<sup>1, 2, 3</sup>

1. Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China

2. Key Laboratory of Infrared System Detection and Imaging Technology, Chinese Academy of Sciences, Shanghai 200083, China

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

4. Shanghai Hydrological Station, Shanghai 200232, China

**Abstract** Urban river water resources are important ecological resources. In recent years, the continuous development of urban industry has led to the increasingly prominent problem of river water pollution. The traditional sampling and testing methods have high precision, but it is time-consuming and laborious. This study proposes a rapid identification of water quality types based on the second-order differential fluctuation index of the spectrum. The method can realize rapid displacement of water quality types in urban rivers. The method first uses the spectral difference calculation to obtain the second-order differential curve of the spectrum, and smoothes the curve to eliminate noise and other disturbances; then uses the sliding window to extract the local maximum and minimum values on the curve, and sets the minimum distance threshold to gradually remove The extreme false point is obtained by using the cubic spline interpolation method to obtain the double envelope of the second-order differential curve of the spectrum. Finally, the fluctuation index curve of the second-order differential of the spectrum is calculated by using

the upper and lower envelopes. After analyzing the fluctuation curves of various samples, it is found that the fluctuation indexes of various water bodies are different at 720~740, 750~770 and 820~840 nm, and then the average fluctuations of various water bodies in these three bands are counted. Statistical characteristics such as the mean value and standard deviation of the index show that the average fluctuation index has a positive correlation with the water quality level. The higher the water quality level, the worse the water quality condition and the larger the average fluctuation index. In order to verify the second-order differential fluctuation index of the spectrum, it can be used for the rapid identification of urban river water quality. The water body spectral samples are randomly divided into training sets and test sets, and LSSVM is used to construct the water quality type, identification model. The average fluctuation index is used as the character input. After testing, the average recognition accuracy of each type of sample is 80.65%, and the recognition accuracy of no more than one type exceeds 95%. The high-spectral identification method of river water quality based on spectral second-order differential fluctuation index proposed by this study has high recognition accuracy and can be used as an auxiliary technical means for rapid detection of urban river water quality types.

**Keywords** River water quality identification; Hyperspectral; Second order differential of spectrum; Remote sensing recognition

(Received May 5, 2019; accepted Sep. 21, 2019)

\* Corresponding author