

基于 EMD 与神经网络的油膜高光谱数据特征提取

董士伟¹, 周子勇¹, 文百红²

(1. 中国石油大学(北京)油气资源与探测国家重点实验室, 北京 102200;

2. 中国石油勘探开发研究院, 北京 100083)

摘要:针对油膜和本底海水的光谱特征,提出了一种基于经验模态分解的高光谱数据特征提取方法,并通过 BP 神经网络构建分类器对油膜和本底海水进行分类识别。该方法首先利用经验模态分解把原始信号在 425~2 390 nm 谱段范围内分解为若干个固有模态函数(IMF-Intrinsic Mode Function),然后计算出每一个 IMF 的能量,选取若干个包含主要特征信息的 IMF 分量的能量特征参数作为 BP 神经网络的输入参数来识别海洋表面油膜信息。研究结果表明,该方法能准确、有效地识别出海洋表面微薄的油膜信息。

关 键 词:经验模态分解;神经网络;油膜;高光谱;特征提取

中图分类号:TP 70 **文献标志码:**A **文章编号:**1004-0323(2010)02-0221-06

1 引 言

随着高光谱遥感影像空间分辨率和光谱分辨率的不断提高,通过高光谱遥感方法识别油气渗漏位置,进而在陆上开展直接油气勘探成为可能,近年的研究也有很多成功的经验^[1,2]。同时也有文献报道通过高光谱油气勘探进行海上油膜的识别,进而把高光谱遥感方法用于海洋油气勘探^[3-6]。然而由于油膜与海水的反射波谱特征信号微弱,如何提取油膜中微弱的反射波谱信息,就成为目前高光谱海洋油气勘探研究中的一个热点问题。

与多光谱遥感一样,在高光谱遥感数据解释中,特征变量的提取是一个关键内容。由于高光谱遥感影像中的每一像元都可以得到一条连续的反射光谱曲线,使得高光谱遥感影像所提供的信息比多光谱影像要丰富得多;但另一方面,高光谱影像特征的提取也要复杂得多。对于高光谱影像特征提取,目前方法有很多,总结起来,可以大致分为三大类。一是以地物反射波谱为基础,设法寻求不同地物特定波段所对应的吸收特征或漂移(红移或蓝移)特征,或通过最佳波段组合、比值等来突出有用信息;二是对反射波谱曲线进行求导、积分或二值化等处理选取特征值;三是从统计学或信号分析的角度出发对高

光谱曲线进行不同的数学处理并选取合适的特征值。如目前用得较多的主成分分析方法以及改进的最小噪声分离(MNF)方法,另外还有分形分析方法、小波分析方法等。

由于油气微渗漏在地表能形成一系列特有异常,因此对于陆上高光谱油气遥感,上述的前两种特征提取方法,常常可以取得较好的效果。但用这些方法提取海洋高光谱油气异常则比较困难。原因主要有两个,一是海洋油膜信息比较微弱,受噪声影响大,因此如何去除干扰信号并把油膜的信息突显出来,就成为了油膜信息提取的一个重要内容。另外就是如何提取油膜特征值。去除干扰的方法有很多,近年来在信号处理中使用相当多的小波方法对于压制干扰有一定的效果。由于小波变换可将信号在不同的尺度上展开,并可提取信号在不同频带上的特征,且同时保留信号在各个尺度上的时域特征,即小波变换可在时域和频域内同时得到较高的分辨率,因此小波方法广泛用于信号去噪以及通过小波分解提取高频(局部)信息。但小波变换的基础是傅里叶变换,适合于非平稳的线性信号的处理,对于非线性信号处理效果并不理想,同时小波变换分析方法作为一种非局部适应方法会把一些有用的局部信息平滑掉,造成局部信息缺失。20 世纪 90 年代末,

美籍华人 Huang^[7] 等在对瞬时频率的概念进行了深入研究后, 创立了一种新的时频分析方法——经验模态分解 (EMD-Empirical Mode Decomposition)。这是一种以瞬时频率为核心概念的方法, 理论上能精确给出信号中频率随时间变化的规律, 避免虚假频率等冗余现象^[8], 该方法基于信号的局部特征时间尺度, 可把信号分解为若干个固有模态函数 (IMF) 之和, 分解出的各个 IMF 分量突出了原始信号的局部特征。

与其它传统油气勘探方法相比, 海洋高光谱油气遥感探测尚处于探索阶段, 而海面油膜的高光谱特征提取方法一直是一个难点。海底油气藏烃类渗漏、海上溢油和海洋生物活动等均可在海面形成油膜, 该油膜改变了海洋表层的物理和化学性质及生物特征^[9], 而且有研究指出海面微薄油膜存在平行多光束干涉现象, 入射到薄膜内的光由于多次反射和折射导致海面可见光/近红外的光谱反射特征发生改变^[6]。因此从理论上讲, 本底海水和有油膜的海水表面高光谱特征是不同的。要了解其不同特征, 可选的方法有两种, 一是在海面上寻找有油膜的表面及洁净海水 (无油膜) 表面进行高光谱测定, 并进行对比分析; 另一种方法则是选择某一洁净的海面, 在海面上滴入原油, 并对比原油滴入前后海面光谱特征的变化。相对来说, 第二种方法可控, 操作相对简单, 而且滴入的原油与实际油气藏烃类渗漏物质具有一定的可比性, 其结果对于从高光谱影像中提取油膜信息具有重要的参考意义。为此本文采用第二种实验方法进行对比实验, 而如何提取出两种海水表面的高光谱特征则是本文研究的重点。

根据 EMD 原理, 如果对高光谱曲线进行经验模态分解, 分解后的 IMF 分量在整个波长范围内突出了原始信号的局部 (即不同地物) 特征。基于这一思想, 本文尝试通过 EMD 方法对野外实测的油膜和本底海水高光谱曲线进行分解, 并得到不同的 IMF。由于油膜和本底海水的 IMF 存在局部差异但不具有统计规律, 而每阶 IMF 能量中包含了丰富的信息, 因而可以依据各阶 IMF 的能量特征参数来识别油膜和本底海水, 并通过人工神经网络方法构建 BP 神经网络分类器, 对实测的油膜和本底海水样本进行分类。

2 EMD 方法的基本原理

2.1 EMD 分解过程

EMD 分解方法可以将原始信号分解成一系列

IMF, 其中 IMF 必须满足以下两个条件^[7]:

(1) 整个数据中, 零点数与极点数相等或至多相差 1。

(2) 信号上任意一点, 由局部极大值点确定的包络线和由局部极小值点确定的包络线的均值均为 0。

EMD 的分解过程^[7,8] 如下: 对任一待分析信号 $s(t)$, 首先将其所有极大值点和所有极小值点分别用两条三次样条曲线进行拟合, 使两条曲线间包含所有信号数据, 从而得到 $s(t)$ 的上、下两条包络线。用 $m(t)$ 记为包络线的平均值, 令 $h(t) = s(t) - m(t)$, 则 $h(t)$ 为一个近似 IMF。将 $h(t)$ 视为新的 $s(t)$, 重复以上操作, 直到当 $h(t)$ 满足 IMF 的条件时, 就得到 $s(t)$ 的第一阶 IMF $c_1(t)$, 即:

$$c_1(t) = h(t) \quad (1)$$

再令:

$$r(t) = s(t) - c_1(t) \quad (2)$$

将 $r(t)$ 视为新的 $s(t)$, 重复以上过程可依次得到第二阶 IMF $c_2(t)$, 第三阶 IMF $c_3(t)$, ...。最终得到分解式:

$$s(t) = \sum_{k=1}^n c_k(t) + r(t) \quad (3)$$

其中: $r(t)$ 称为残余函数, 代表信号的平均趋势。

2.2 IMF 能量特征提取

经 EMD 分解后的各 IMF 分量代表了一组特征尺度下的平稳信号, 更加突出了原始信号的局部特征。由于不同地物的局部特征的差异使得分解后各 IMF 分量的能量分布存在较大差异, 因此可以选取各 IMF 分量的能量作为输入网络的特征向量来识别不同地物。基于 EMD 的能量特征提取如下:

计算各 IMF 的能量 E_i ($i = 1, 2, \dots, n$), n 为 IMF 的个数, 则有:

$$E_i = \int |c_i(t)|^2 dt = \sum_{j=1}^m |c_{ij}|^2 \quad (4)$$

($i = 1, 2, \dots, n; j = 1, 2, \dots, m$), 其中 c_{ij} 为 IMF 的离散点的幅值, m 为信号采样点的数目。

构造特征向量 T :

$$T = [E_1, E_2, \dots, E_n] \quad (5)$$

对特征向量进行归一化处理:

$$E = (\sum_{i=1}^n |E_i|^2)^{1/2} \quad (6)$$

则有:

$$T' = [E_1/E, E_2/E, \dots, E_n/E] \quad (7)$$

向量 T' 即为归一化后的向量。

3 数据处理与结果分析

3.1 实验方法

实验步骤如图 1 所示。其中的关键是 EMD 分解及 BP 神经网络拓扑结构的设计。

3.2 数据来源与预处理

实验所用的数据是利用 ASD 全谱段地物光谱仪(FieldSpec-FR)所采集的野外实验数据。实验场地为辽东湾盘锦海域某处。实验数据采集基本过程

如下:先测出实验场地本底海水的光谱反射曲线,然后再在同一点滴入原油,原油样品来自于辽河油田,待其扩散静止后再测量油膜光谱反射率。光谱测量范围为 350~2 500 nm,每组数据获取 2 151 个数据点。将该数据在 425~2 390 nm 谱段范围内进行 EMD 分解,以便使分解后的 IMF 分量的局部特征能与高光谱影像数据波段一一对应,有利于下一步的光谱分析和特征提取等。图 2 是某测量点油膜和本底海水的高光谱曲线。

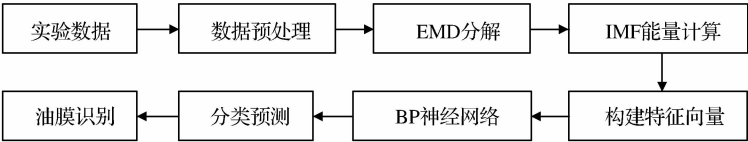


图 1 基于 EMD 与神经网络的油膜特征提取流程

Fig. 1 Flow chart of feature extraction of oil slick based on EMD and neural network

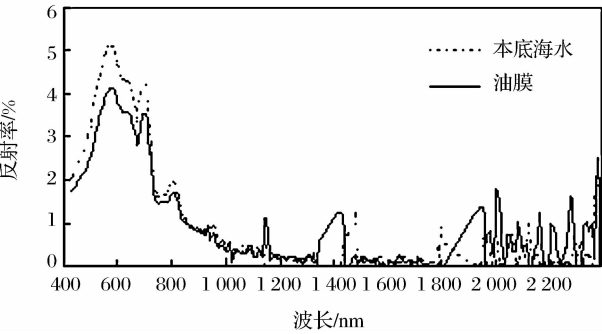


图 2 油膜和本底海水的光谱特征

Fig. 2 Reflectance spectrum of oil slick and background seawater

3.3 EMD 分解和 IMF 能量计算

EMD 分解和 IMF 能量计算程序采用可视化交互数据语言 ENVI/IDL6.3 编写^[10]。利用 EMD 分解程序对油膜和本底海水在 425~2 390 nm 谱段范围进行分解,计算结果表明,不同样本点,或者同一测量点油膜和海水样本高光谱曲线分解得到的 IMF 个数不完全一样,油膜的 IMF 个数为 8~12 个,而本底海水的 IMF 个数为 8~14 个。图 3 和图 4 分别是某个油膜样本和本底海水样本的 EMD 分解结果。

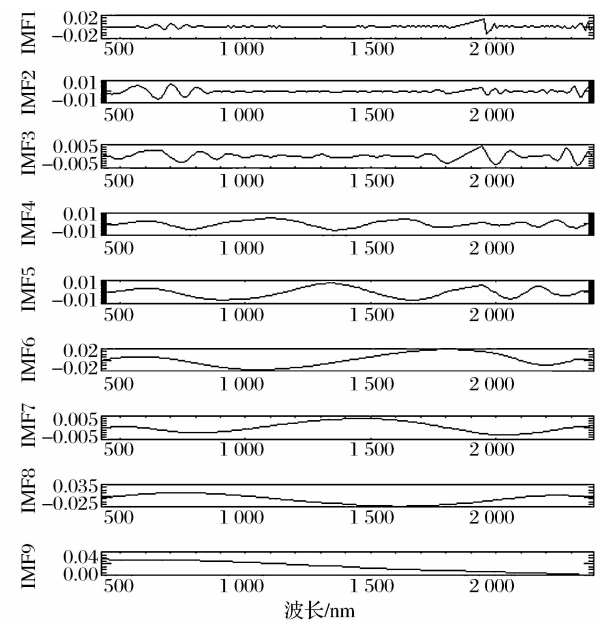


图 3 油膜的 EMD 分解结果

Fig. 3 The EMD results of oil slick

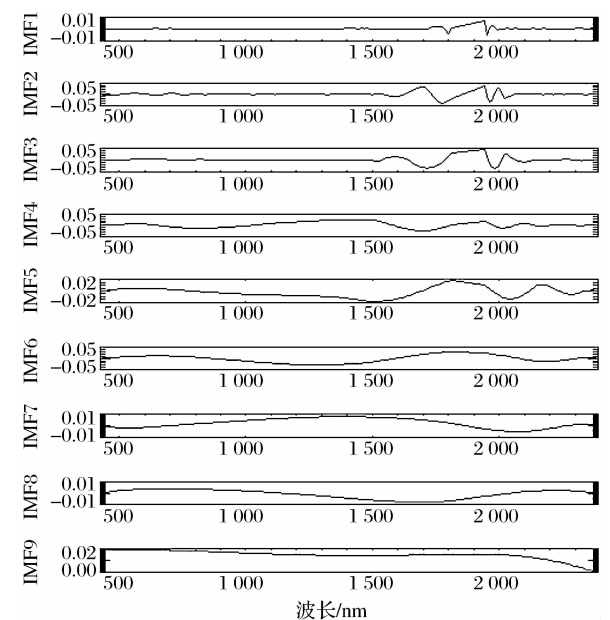


图 4 本底海水的 EMD 分解结果

Fig. 4 The EMD results of background seawater

一般来说,最先得到的 IMF,如图 3、图 4 中 IMF1、IMF2 等可能与噪声有关,而最后的 IMF,如图 3、图 4 中的 IMF9 等,主要反映的是信号的总趋势,中间的 IMF 主要反映的是信号中不同信号源的信息。而不同地物各 IMF 所反映的能量是不同的,因此依据不同 IMF 的能量特征的差异,就有可能对不同的地物进行识别与分类。为此随机选取 5 个油膜样本和 5 个本底海水样本,并计算两者各阶 IMF 平均能量,图 5 为油膜和本底海水的 1~8 阶 IMF 平均能量分布。从图 5 可以看到,两者在第 1、2 和 3 阶能量差异不大,而在第 4、5、6、7 和 8 阶差异明显;两者在前 3 阶能量非常小,表明前 3 阶所含信号的特征信息较少,噪声信息较多。因此选取第 4、5、6、7 和 8 阶 IMF 分量的能量构建特征向量,同时去除与噪声有关的前 3 阶 IMF 分量,更加保证了后续数据分析的可靠性。

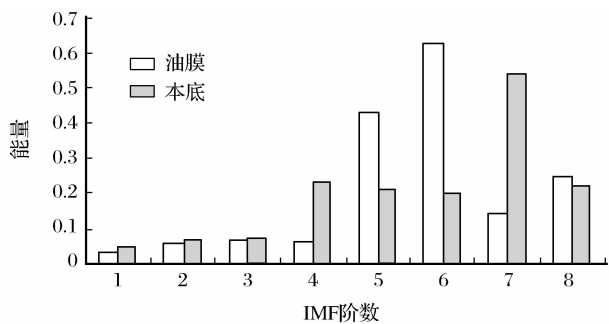


图 5 油膜和本底海水的 IMF 能量分布图

Fig.5 The IMF's energy distribution of oil slick and background seawater

3.4 神经网络构建与结果分析

BP 神经网络是目前应用最广泛、发展最成熟的一种神经网络模型,是对非线性可微分函数进行权

值训练的多层网络,具有良好的自学习和分类能力。它由输入层、隐含层(可以有多层)和输出层组成。

在进行网络设计时,重要的是确定网络的总层数和各层单元数。输入层单元数由特征参数个数确定;输出层单元数由状态数和状态编码方式确定;隐含层单元数的选取目前尚无理论依据,可根据经验和使用情况确定^[11]。然后用油膜和本底海水组成训练样本集进行网络训练,使得网络的期望平均均方差(Mean Square Error,MSE)最小。

实验采用的 BP 神经网络结构为 3 层,输入参数为前面构建的特征向量,即 4、5、6、7 和 8 阶 IMF 分量的能量;隐含层的传递函数采用双曲正切 S 型传递函数 tansig,输出层为一个神经元,输出层的传递函数采用 S 型的对数函数 logsig,该函数将输出量的值域限制在[0,1] 之间,目标输出值为 0 时表示油膜,为 1 时表示本底海水。网络结构如图 6 所示。结合实际情况,随机选取 26 个油膜和本底海水样本进行 BP 神经网络训练,达到截止误差 0.0001 时训练结束。部分训练样本如表 1 所示。

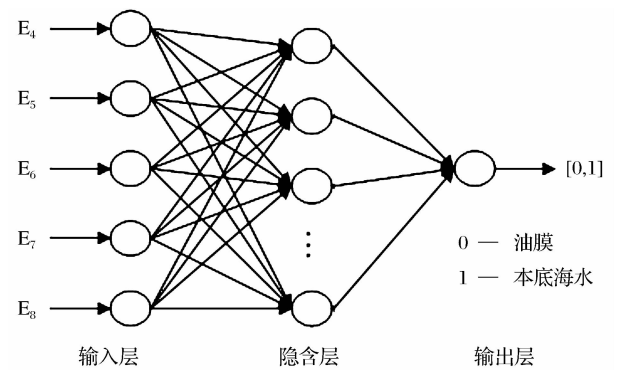


图 6 神经网络拓扑结构图

Fig.6 Topology of designed BP neural network

表 1 部分训练样本

Table 1 Part of training samples

序号	类型	E_4	E_5	E_6	E_7	E_8	编码
1	油膜	0.055808	0.106999	0.901754	0.030451	0.403520	0
2	油膜	0.215108	0.310445	0.506155	0.659289	0.036854	0
3	油膜	0.000206	0.719200	0.694682	0.012948	0.000260	0
4	油膜	0.301014	0.517959	0.070112	0.237499	0.096626	0
5	油膜	0.007614	0.772084	0.635475	0.000619	1.57E-05	0
6	本底	0.000183	0.000236	0.963955	0.154742	0.216439	1
7	本底	0.002002	0.009038	0.805594	0.592352	0.006270	1
8	本底	0.728606	0.684897	0.006949	0.000576	0.000185	1
9	本底	0.006024	0.360974	0.932030	0.029345	0.006251	1
10	本底	0.000325	0.004232	0.458298	0.888654	0.015464	1

为了验证 BP 网络的准确性,用 12 个测试样本进行了测试,测试结果如表 2 所示。从表 2 可以看出,网络的实际输出与期望输出非常吻合,错误识别

样本 1 例,诊断准确率约为 91.67%。实验结果表明该神经网络模型可靠,能准确地对油膜和本底海水进行区分。

表 2 网络模型测试结果

Table 2 Test results of network model

序号	类型	E_4	E_5	E_6	E_7	E_8	期望输出	网络输出
1	油膜	9.42E-05	0.714154	0.699975	0.001287	0.004137	0	0.002228
2	油膜	0.039229	0.093977	0.600289	0.025492	0.791776	0	7.46E-17
3	油膜	0.002018	0.727296	0.686310	0.000169	0.000440	0	4.78E-06
4	本底	0.421225	0.215884	0.713573	0.067342	0.046271	1	1
5	本底	4.06E-06	0.005077	0.679899	0.733288	0.000154	1	1
6	本底	0.043321	0.726254	0.684841	0.001339	0.000597	1	0.980160
7	本底	0.297764	0.219737	0.505742	0.742154	0.177759	1	0.998660
8	油膜	0.001217	0.716021	0.698076	0.000118	0.000296	0	0.011413
9	油膜	0.632331	0.774651	0.008400	0.001057	0.000361	0	0.001931
10	本底	0.084500	0.151918	0.909276	0.038267	0.009736	1	1
11	本底	3.17E-08	2.79E-07	0.179506	0.926103	0.331830	1	0.999650
12*	本底	0.335282	0.459327	0.015320	0.013919	0.007991	1	0.053239

* 注:错误识别样本

4 结 语

本文把 EMD 分解和 BP 神经网络方法用于本底海水和油膜海水表面的高光谱实验数据特征提取与分类研究,得到以下几点初步认识:

(1) 通过 EMD 分解后的 IMF 分量突出了原始信号的局部特征信息。结果表明,IMF 能量特征可以清楚地揭示油膜和本底海水的局部特征差异,而通过神经网络方法对两种样本进行分类的精度较高,表明了方法的有效性。

(2) 由于实验条件所限,本实验通过在海面滴入原油来模拟实际烃渗漏在海面形成的油膜并对其高光谱特征进行研究,具有一定的可比性和参考价值,但海面不同成因油膜高光谱特征的差异性需要进一步的研究。

(3) 高光谱海洋油气遥感的最终目的是希望从高光谱遥感影像中获得烃渗漏在海面所形成油膜的空间分布规律,进而可以进行油气勘探。然而高光谱影像中像元的曲线特征与海水表面实测的高光谱数据还存在一定的差距,如何对高光谱遥感影像中各像元曲线进行 EMD 分解并根据分解结果进行分类,还需要进一步研究;而且通过 EMD 分解得到的各阶 IMF 的物理意义目前还不很清楚,这对于基于 EMD 分解的高光谱遥感影像的处理有着重要的

意义。

尽管存在以上一些问题,本实验研究结果对于海洋高光谱油气勘探还是具有一定的借鉴意义和参考价值。由于现阶段实施物探和化探采样分析等传统油气勘探技术方法及手段,均受到成本高、周期长等诸多因素的影响和限制,使得人们往往无法获取大面积、动态、连续的信息,亦制约了海洋油气勘探的进程与发展。而利用各类遥感数据特别是高光谱遥感数据进行海洋油气探测,由于成本低,可动态观测,特别是随着高光谱遥感影像的光谱分辨率和空间分辨率的提高,基于高光谱进行海洋油气勘探有可能成为海洋油气勘探的一种重要手段而越来越受到重视,本文方法为以后高光谱遥感影像进行油膜特征提取提供了参考依据。

参考文献:

[1] Horig B,Kuhn F,Oschutz F,*et al.* Hymap Hyperspectral Remote Sensing to Detect Hydrocarbons[J]. International Journal of Remote Sensing,2001,22(8):1413-1422.

[2] Saunders D F,Burson K R,Thompson C K. Model for Hydrocarbon Microseepage and Related Near-surface Alterations [J]. AAPG BULL,1999,83(1):170-185.

[3] Hese S,Schmullius C. High Spatial Resolution Image Object Classification for Terrestrial Oil Spill Contamination Mapping in West Siberia[J]. International Journal of Applied Earth Observation and Geoinformation,2009,11(2):130-141.

- [4] Salem F M F. Hyperspectral Remote Sensing: A New Approach for Oil Spill Detection and Analysis[D]. George Mason University, 2003, 49-85.
- [5] Lu Y C, Tian Q J, Wang J J, *et al.* Experimental Study on Spectral Responses of Offshore Oil Slick[J]. Chinese Science Bulletin, 2008, 53(24): 3937-3941.
- [6] Lu Yingcheng, Tian Qingjiu, Qi Xiaoping, *et al.* Spectral Response Analysis of Offshore Thin Oil Slicks[J]. Spectroscopy and Spectral Analysis, 2009, 29(4): 986-989. [陆应诚, 田庆久, 齐小平, 等. 海面甚薄油膜光谱响应研究与分析[J]. 光谱学与光谱分析, 2009, 29(4): 986-989.]
- [7] Huang N E, Shen Z, Long S R, *et al.* The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Nonstationary Time Series Analysis[C]//Proceedings of the Royal Society of London Series A-Mathematical Physical and Engineer Sciences, 1998, 454: 903-995.
- [8] Zhong Youming, Qin Shuren. Research on the Uniform Theoretical Basis for Hilbert-huang Transform[J]. Journal of Vibration and Shock, 2006, 25(3): 40-43. [钟佑明, 秦树人. 希尔伯特—黄变换的统一理论依据研究[J]. 振动与冲击, 2006, 25(3): 40-43.]
- [9] Chen Yu, Xu Ruisong, Cai Rui, *et al.* Remote Sensing Detection Method Research of the Oil and Gas Reservoirs Distribution in North Part of South China Sea[J]. Remote Sensing Technology and Application, 2009, 24(1): 120-125. [陈彧, 徐瑞松, 蔡睿, 等. 南海北部油气分布的遥感探测方法研究[J]. 遥感技术与应用, 2009, 24(1): 120-125.]
- [10] Han Peiyu. Visualization Analysis and Application of IDL[M]. Xi'an: Northwestern Polytechnical University Press, 2006, 24-112. [韩培友. IDL 可视化分析与应用[M]. 西安: 西北工业大学出版社, 2006, 24-112.]
- [11] Yang Yu, Yu Dejie, Cheng Junsheng. Roller Bearing Fault Diagnosis Method Based on EMD[J]. China Mechanical Engineering, 2004, 15(10): 908-911. [杨宇, 于德介, 程军圣. 基于经验模态分解的滚动轴承故障诊断方法[J]. 中国机械工程, 2004, 15(10): 908-911.]

Feature Extraction of Offshore Oil Slick from Hyperspectral Data Based on EMD and Neural Network

DONG Shi-wei¹, ZHOU Zi-yong¹, WEN Bai-hong²

(1. State Key Laboratory of Petroleum Resource and Prospecting,
China University of Petroleum, Beijing 102200, China;

2. Research Institute of Petroleum Exploration and Development, CNPC, Beijing 100083, China)

Abstract: Based on the reflectance spectrum of oil slick and background seawater, an EMD based approach to feature extraction from hyperspectral data is presented in the work, and a BP neural network is constructed to classify the oil slick and seawater samples. Firstly, the EMD is adopted to decompose the hyperspectral data from 425 nm to 2 390 nm into a series of Intrinsic Mode Functions (IMFs), and the energy of each IMFs is computed, then based on the computed energy, some IMFs energy are selected as inputted parameter of neural networks to classify the oil slick and seawater. The results show that the EMD based feature extraction approach can identify oil slick accurately and effectively.

Key words: Empirical mode decomposition; Neural network; Oil slick; Hyperspectral; Feature extraction