

DOI :10.3876/j.issn.1000-1980.2008.06.003

利用地面光谱数据反演悬浮泥沙质量浓度

程立刚^{1,2}, 张 鹰¹

(1. 南京师范大学地理科学学院, 江苏 南京 210097; 2. 河海大学水文水资源学院, 江苏 南京 210098)

摘要: 为研究用高光谱数据反演悬浮泥沙质量浓度的方法, 以近海悬浮泥沙为实验材料, 配制了不同质量浓度悬浮泥沙样品, 利用 2 种方法对样品进行建模: 一种是将可见光和近红外光谱进行小波变换并将小波低频系数以偏最小二乘回归建模, 另一种为波段组合法. 利用交叉检验方法, 分析了交叉检验结果, 并将 2 种方法建立的模型进行了比较. 研究表明: 波段组合法中, TM4 与 TM1 波段反射率比值的指数模型效果最好; 在预测能力上, 小波偏最小二乘模型总体上比波段组合模型精度高并且稳定性好, 适用于悬浮泥沙光谱定量分析.

关键词: 悬浮泥沙; 小波系数; 偏最小二乘回归; 光谱反射率

中图分类号: TV141⁺.1 文献标识码: A 文章编号: 1000-1980(2008)06-0738-05

通过遥感方法获取水体表层悬浮泥沙信息具有速度快、周期短、成本低、范围大等优势, 有着广阔的应用前景. 遥感数据和悬浮泥沙含量之间的定量关系可分为水下的含沙量-水面光谱反射率关系和大气中的水面-遥感器光谱反射率关系 2 个部分^[1]. 因此, 对含沙量-水面光谱反射率关系的研究成为悬浮泥沙遥感定量分析的基础. 目前普遍认为, 悬浮泥沙定量模式包括 5 种关系式: 线性关系式、对数关系式、Gordon 关系式、负指数关系式和统一关系式, 其中, 对数关系式和指数关系式是近年来国内学者最多使用的模式, 适合我国含沙量高的实际情况. 但由于缺乏足够可靠的同步资料, 至今尚无统一的定量模式或可靠的模型参数, 即使在同一地区也是如此^[2]. 在应用上述关系式时通常的做法是: 首先选择合适的若干波段或波长(一般不超过 4 个), 然后将选中的波长进行非线性变换得到关于波长的表达式, 主要变换方法有波段比值^[3-4]、波段多项式^[5]、指数变换^[6]等, 最后将波长表达式与悬浮泥沙质量浓度建立回归方程. 本文根据国内最常用的波段组合方案设计了 9 个传统回归模型, 并进行了比较.

偏最小二乘回归(partial least-squares regression, PLSR)是 20 世纪 80 年代中期出现的一种多元统计数据分析方法^[7]. PLSR 方法较好地解决了变量之间多重相关问题, 尤其适合样本点数量远小于数据维数的高光谱情况. 采用这种方法不需要进行波段组合, 所有波段都可参加反演, 适合悬浮泥沙高光谱反演. 小波分析是 20 世纪 80 年代出现的信号分析工具, 常应用于信号压缩、噪声消除等方面^[8]. 留一法(leave-one-out)是一种交叉检验方法, 就是根据 $n-1$ 个样本建立的模型进行预测, 利用未参与建模的样本考察模型的预测精度^[7]. 留一法是常用的考察模型预测精度的方法. 为了消除噪声对模型的干扰和降低数据的维数, 本文首先对 ASD 光谱仪数据进行了小波变换, 然后将变换结果用于 PLSR, 建立了悬浮泥沙质量浓度和光谱反射率反演模型, 最后用留一法评价模型的预测精度.

1 实验简介

1.1 实验方法

实验的悬浮泥沙样本采自江苏省海安县老坝港镇新北凌闸附近潮沟中的海水. 使用 ASD 公司的 FieldSpec HandHeld 便携式手持光谱仪测量悬浮泥沙水样反射光谱. 实验设备和样品制备参照傅克忭等^[9]的设计, 盛放悬浮泥沙水样的水槽长宽高均为 0.8 m, 为消除阳光反射的影响, 水槽内壁及底部敷设黑色不反光塑料布. 沙样经过清水清洗, 以消除非悬浮泥沙物质的干扰. 清洗后将沙样烘干称重分成若干份, 分别放入定

收稿日期: 2007-12-10

基金项目: 国家自然科学基金(40606044)

作者简介: 程立刚(1971—)男, 山西祁县人, 讲师, 硕士, 主要从事悬浮泥沙遥感研究.

容过的器皿,并溶解、浸泡数日,备用.实验时清水中加入沙样,搅拌均匀并保持泥沙的悬浮状态.配制得到悬浮泥沙质量浓度从 0.010 ~ 3.699 g/L,基本覆盖了江苏海域的含沙量范围.反射率测量遵循 NASA 规范,测量时间选择在 11 ~ 13 时,天气晴好无风,日面无云.光谱仪距水面 0.2 m,观测平面与太阳入射平面夹角 $\varphi_v = 135^\circ$,与水平法线方向的夹角 $\theta_v = 40^\circ$,标准板采用反射率为 30% 的灰板.

1.2 数据特征分析

通过实验,得到 12 个悬浮泥沙水样光谱数据,如图 1 所示(图 1 水样 1 ~ 12 的悬浮泥沙质量浓度分别为 0.010 g/L,0.049 g/L,0.078 g/L,0.096 g/L,0.190 g/L,0.377 g/L,0.598 g/L,0.796 g/L,0.991 g/L,1.527 g/L,1.964 g/L,3.699 g/L).为清晰起见,图 1 的曲线作了平滑处理.图 1 显示出光谱曲线具有双峰特征以及反射峰红移现象,这些与大型实验^[10]得到的结果相同.悬浮泥沙光谱曲线之间基本不相交.随着悬浮泥沙质量浓度增加,几乎所有波长的反射率都有所增加,但增加规律各不相同(图 2),因此,建模波段的选择及其组合方式存在经验性.PLSR 模型可以使用全部波段,但需要对数据进行预处理,降低数据维数,滤去噪声,以提高计算速度和反演精度.

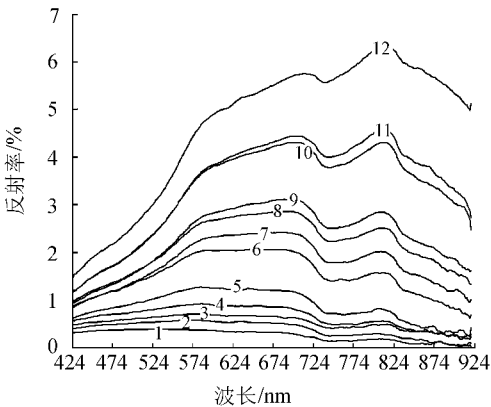


图 1 悬浮泥沙反射率

Fig.1 Spectral reflectance curves of suspended sediment samples

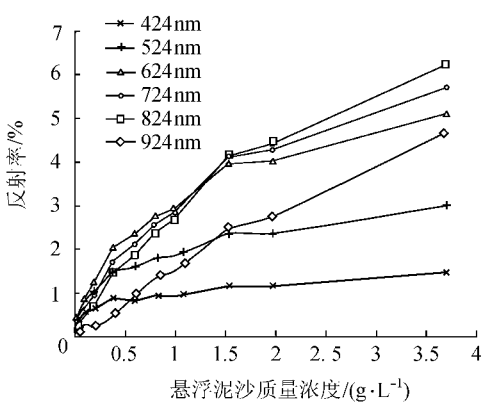


图 2 部分波长的悬浮泥沙质量浓度-反射率关系

Fig.2 Relationship between concentration and reflectance of suspend sediment (partial data)

2 建立反演模型

2.1 建立小波-偏最小二乘法模型

本文综合利用小波和偏最小二乘法建立了小波-偏最小二乘法模型.

第一步 利用小波变换对光谱数据进行预处理.许多小波基都可用于数据压缩和降噪,其中 Daubechie 小波(简称 db 小波)具有有限紧支集、可正交分解、数据冗余小等优良性质,常常用作数据降噪和压缩的小波基.故本文采用 db 小波.通常光谱的主要信息包含在低频信号中,高频信号往往是噪声成分,因此,本文主要对光谱信号的低频小波系数进行研究,以期达到减少数据噪声和降低数据规模的作用.数据预处理采用 Matlab7.0 所提供的小波工具箱编程实现.

第二步 对预处理后的光谱数据进行偏最小二乘回归分析.通过计算样本值与回归估计值的误差平方和可确定偏最小二乘回归最佳成分数量.具体步骤为 (a)计算 h 个成分下误差平方和 S_h : $S_h = \sum_{i=1}^n (y_i - \hat{y}_i)^2$,其中 $i = 1, 2, \dots, n$, $\hat{y}_i$ 为回归估计值, y_i 为样本值, n 为样本数量, h 为成分数量 (b)比较不同成分数量下的 S_h 大小,可得到最小取值 $\min S_h$ (c)选取 $\min S_h$ 所对应的 h 值作为最佳成分数量.根据文献 [7],偏最小二乘法采用 Matlab 编程实现,并且根据该文献中所提供的示例数据对程序正确性进行了验证.

在模型计算中,还需要确定 db 小波的阶数和分解层数.一般应用中,db 小波的阶数不大于 45,并且对于不同阶数的 db 小波存在相应的合理分解层数.通常低阶 db 小波的合理分解层数比高阶多,但一般不超过 8 层.由于 1 层分解达不到降低数据量的效果,因此 db 小波分解层数可确定为 2 ~ 8 层.据此,本文对 db1 ~ db45 小波(即 1 阶 db 小波 ~ 45 阶 db 小波)的 2 ~ 8 层分解进行了比较(表 1),选取 12 个样本数据作为训练数据,通过交叉检验的方法确定模型的最佳 db 小波参数.具体步骤如下 (a)将每个样本进行 db 小波变换,可

以得到不同阶数和分解层数的 db 小波的低频系数 (b) 分别将这些小波系数与悬浮泥沙质量浓度进行偏最小二乘回归建模 (c) 利用交叉检验法计算得到不同小波系数的偏最小二乘回归模型对应的平均相对误差绝对值 (表 1) 以及相对误差绝对值的方差 (表 2) . 为清晰起见 , 表 1 省略了误差较大的 db 小波 , 表 2 列出了平均相对误差绝对值最小的 6 种 db 小波分解的方差 . 容易看出 , db7 小波 3 层分解的相对误差绝对值的方差最小 , 即预测误差的稳定性最好 . 由于表 2 所列各分解方式之间的平均相对误差绝对值差异性较小 (< 14 %) 相对误差绝对值的方差差异较大 (最大值是最小值的 2.6 倍) , 因此 , 确定将相对误差绝对值的方差作为选择小波分解方式的主要依据 , 确定 db7 小波 3 层分解为此次建模所用的小波分解方式 . 通过对原始光谱样本进行 db7 小波 3 层分解可建立关于小波低频系数与悬浮泥沙质量浓度的最小二乘回归模型 , 其交叉检验结果见表 4 .

表 1 不同 db 小波分解下交叉检验的平均相对误差绝对值

Table 1 MAPE of cross-validation under decomposition of different db wavelets

阶数	不同分解层数的平均相对误差绝对值						
	2	3	4	5	6	7	8
1	93.980	42.467	66.464	97.483	388.310	75.332	441.050
2	108.950	108.340	190.420	52.703	68.034	63.440	
3	77.647	70.130	102.010	284.450	43.041		
4	82.415	88.645	253.440	82.895	298.320		
5	57.898	56.002	171.680	134.520			
6	56.930	59.522	42.594	111.500			
7	37.279	35.543	70.518	196.380			
8	44.822	54.304	144.080	130.060			
45	90.740						

表 2 交叉检验平均相对误差及其方差

Table 2 MAPE and VAPE of cross-validation under decomposition of different db wavelet

阶数	分解层数	平均相对误差绝对值	相对误差绝对值的方差
7	2	37.279	2055.6
7	3	35.543	1380.6
13	2	35.147	3049.5
14	2	39.066	3013.7
17	2	39.767	3086.5
21	2	39.236	3603.8

2.2 建立波段组合模型

国内的悬浮泥沙遥感反演对 TM 图像应用研究较多 , 取得了不少典型成果^[11-16] . 本文选择悬浮泥沙反演最常用的 TM1 和 TM4 波段 , 利用本实验的样本数据 , 建立了关于悬浮泥沙质量浓度的指数模型、对数模型和线性模型 . 模型中分别以 TM1 波段中心波长 (485 nm) 和 TM4 波段中心波长 (830 nm) 的反射率代表 TM1 和 TM4 波段的反射率 . 计算结果表明 (表 3) 模型 9 的残差均方根和残差平方和比其他模型要小得多 , 相关系数平方 $R^2 = 0.9880$, 非常接近 1 , 说明模型对数据的解释能力非常强 . 因此 , 对于本实验的悬浮泥沙样本 , 模型 9 拟合结果最好 .

表 3 波段组合模型对比分析

Table 3 Comparison among band composite models

模型编号	关系式	残差平方和	相关系数平方 R^2	残差均方根	模型编号	关系式	残差平方和	相关系数平方 R^2	残差均方根
1	$S = a \frac{R_{TM4} - R_{TM1}}{R_{TM4} + R_{TM1}} + b$	4.578	0.649	0.677	6	$S = a \exp \left(\frac{R_{TM1}}{R_{TM4}} b \right)$	0.686	0.947	0.262
2	$\ln S = a \frac{R_{TM4} - R_{TM1}}{R_{TM4} + R_{TM1}} + b$	1.403	0.958	0.374	7	$S = a \frac{R_{TM4}}{R_{TM1}} + b$	2.248	0.828	0.474
3	$S = a \exp \left(\frac{R_{TM4} - R_{TM1}}{R_{TM4} + R_{TM1}} b \right)$	0.461	0.965	0.214	8	$\ln S = a \frac{R_{TM4}}{R_{TM1}} + b$	3.994	0.880	0.632
4	$S = a \frac{R_{TM1}}{R_{TM4}} + b$	7.025	0.462	0.838	9	$S = a \exp \left(\frac{R_{TM4}}{R_{TM1}} b \right)$	0.143	0.988	0.119
5	$\ln S = a \frac{R_{TM1}}{R_{TM4}} + b$	1.753	0.947	0.418					

注 : S 为悬浮泥沙质量浓度 ; R_{TM1} , R_{TM4} 分别为 TM1 和 TM4 波段中心波长的反射率 ; a 和 b 为常数 .

2.3 模型预测能力的分析与比较

模型最重要的能力是它的预测能力 , 而交叉检验方法能较好地体现模型的预测精度 , 因此 , 可通过交叉检验来考察模型的预测能力 .

对于小波-偏最小二乘法模型 , 小波阶数及分解层数对预测精度有着重要的影响 . 从表 1 可以看出 : 在合理分解层数范围内 , 精度较差的小波分解集中在 4 ~ 8 层分解 , 小波分解层数越多 , 回归精度越差 ; 而 db 小波阶数的增长不会显著降低预测精度 . 推断是由于分解层数增多 , 小波系数数量减少 , 小波系数所代表的光谱

信息随之减少,从而导致回归精度降低。

通过表 4 的交叉检验结果可以看出,模型 9 在低质量浓度的 1~4 号样本相对误差较大,尤其是 1 号样本,在中等质量浓度(样本 5~9)除 5 号样本的相对误差大于波段组合法,其余样本的相对误差均比传统模型小,在高质量浓度(样本 10~12)波段组合法的预测精度略高于小波-偏最小二乘法。总体上,小波-偏最小二乘法总的平均相对误差为 34.834,约为波段组合法的 1/3,相对误差绝对值的方差约为波段组方法的 1/4,小波-偏最小二乘法预测精度比波段组合法高,并且更为稳定。

表 4 小波-偏最小二乘法与波段组合法交叉检验结果对比

Table 4 Comparison of cross validation results between wavelet-PLRS model and band composite model

样本 编号	实测值/(g·L ⁻¹)	小波-偏最小二乘法		波段组合法(模型 9)		两种模型预测 相对误差比较
		预测值/(g·L ⁻¹)	预测相对误差/%	预测值/(g·L ⁻¹)	预测相对误差/%	
1	0.010	0.006	39.181	0.096	863.010	-
2	0.049	0.089	83.032	0.108	120.460	-
3	0.078	0.151	94.596	0.156	100.250	-
4	0.096	0.095	0.641	0.147	53.519	-
5	0.190	0.008	95.750	0.159	16.097	+
6	0.377	0.434	15.099	0.280	25.472	-
7	0.598	0.553	7.484	0.463	22.497	-
8	0.796	0.844	6.039	0.577	27.392	-
9	0.991	0.911	8.031	0.893	9.869	-
10	1.527	1.839	20.482	1.704	11.648	+
11	1.964	1.670	14.950	2.200	12.025	+
12	3.699	2.488	32.725	3.101	16.164	+
平均值			34.834		106.533	-
方差			1 280.400		58 075.000	-

注:两种模型预测相对误差比较中,+/-表示小波-偏最小二乘法误差大于/小于波段组合法。

无论小波-偏最小二乘法模型还是传统模型,低质量浓度区域的相对误差较大,主要是由于低质量浓度区域悬浮泥沙质量浓度绝对值较小,对相对误差比较敏感所致。小波-偏最小二乘法在高质量浓度部分误差略大,可能是这样的原因:PLSR 是一个线性变换,而反射率与质量浓度关系在高质量浓度部分包含了一些非线性的趋势,为了使整体误差最小化,必然造成对实测值不敏感的高质量浓度样本预测误差偏大,模型 9 本身是一个非线性表达式,因此在高质量浓度部分的预测误差远小于中低质量浓度部分。

3 结 论

在模型选用上,对于波段组合法,模型 9 的性能最好,小波-偏最小二乘法模型反演总体上优于波段组合法,也较适合悬浮泥沙反演的模型,但分解层数不宜过多。波段组合法的优势在于,只需要通过几个波段的计算就可以得到较好的结果,消耗计算资源少,并且容易推广到遥感图像反演,缺点是波段组合形式的选择需要较多的经验。小波-偏最小二乘法模型的优点在于全部波段都参与建模,不需要考虑波段的选取与组合,并且预测能力总体较好也较为稳定。小波-偏最小二乘法模型缺点是,所有波段参与建模的方式有可能引入噪声,降低模型预测能力,确定小波分解层数上还需要依赖一定的经验;由于数据量大,消耗计算资源较多。2 种模型在低质量浓度区域的预测精度的相对误差较大,可考虑将高中低质量浓度数据分别建模,以提高模型预测能力。本文尝试使用新的反演方法和模型评价方法研究高光谱数据反演悬浮泥沙质量浓度问题,但是由于实验条件所限,悬浮泥沙样本的数量和类型还需要进一步增加和丰富,以增强模型的代表性和通用性。

参考文献:

[1] 李炎,李京.基于海面-遥感器光谱反射率斜率传递现象的悬浮泥沙遥感算法[J].科学通报,1999,44(17):1892-1897.
[2] 李四海,辉才兴.河口表层悬浮泥沙气象卫星遥感定量模式研究[J].遥感学报,2001,5(2):154-160.
[3] DOXARAN D,FROIDEFOND J M,CASTAING P.A reflectance band ratio used to estimate suspended matter concentrations in sediment-dominated coastal waters[J].International Journal of Remote Sensing,2002,23(23):5079-5085.
[4] GIN K Y-H,KOH S T,LIN I-I.Spectral irradiance profiles of suspended marine clay for the estimation of suspended sediment concentration in tropical waters[J].International Journal of Remote Sensing,2003,24(16):3235-3245.

- [5] BINDING C E ,BOWERS D G ,MITCHELSON-JACON E G. An algorithm for the retrieval of suspended sediment concentrations in the Irish Sea from SeaWiFS ocean colour satellite imagery[J]. International Journal of Remote Sensing 2003 24(19): 3791-3806.
- [6] MYINT S W ,WALKER N D. Quantification of surface suspended sediments along a river dominated coast with NOAA AVHRR and SeaWiFS measurements : Louisiana ,USA[J]. International Journal of Remote Sensing 2002 23(16): 3229-3249.
- [7] 王惠文. 偏最小二乘回归方法及其应用[M]. 北京 :国防工业出版社 ,1999.
- [8] 陈基明. 小波分析基础[M]. 上海 :上海大学出版社 2002.
- [9] 傅克忭, 荒川久幸, 曾宪模. 悬沙水体不同波段反射比的分布特征及悬沙量估算实验研究[J]. 海洋学报 ,1999 21(3): 134-140.
- [10] 韩震, 辉才兴, 蒋雪中. 悬浮泥沙反射光谱特性实验研究[J]. 水利学报 2003(12): 118-122.
- [11] 黄海军, 李成治, 郭建军. 黄河口海域悬沙光谱特征的研究[J]. 海洋科学 ,1994(5): 40-44
- [12] 何青, 辉才兴, 时伟荣. 长江口表层水体悬沙浓度场遥感分析[J]. 自然科学进展 ,1999 9(2): 160-164.
- [13] 温令平. 伶仃洋悬浮泥沙遥感定量分析[J]. 水运工程 2001(9): 9-17.
- [14] MA Rong-hua ,DAI Jin-fang. Investigation of chlorophyll-a and total suspended matter concentrations using landsat ETM and field spectral measurement in Taihu Lake ,China[J]. International Journal of Remote Sensing 2005 26(13): 2779-2795
- [15] 李文达, 王心源, 周迎秋, 等. 利用 TM 影像分析巢湖悬沙分布及其原因[J]. 水土保持研究 2006 13(2): 179-181.
- [16] 李洪灵, 张鹰, 姜杰. 基于遥感方法反演悬浮泥沙分布[J]. 水科学进展 2006 17(2): 242-245.

Suspended sediment concentration estimation retrieved by field spectral data

CHENG Li-gang^{1,2}, ZHANG Ying¹

(1. School of Geography Science , Nanjing Normal University , Nanjing 210097 , China ;

2. College of Hydrology and Water Resources , Hohai University , Nanjing 210098 , China)

Abstract : In order to develop the models to retrieve the suspended sediment concentration information by use of the spectral data , the experiments were carried out on suspended sediment samples (made of materials from coastal suspend sediment) with different concentrations. Two types of models were established to predict the suspend sediment concentration by spectral data. One was the partial least-squares regression (PLSR) model through the wavelet transformation of visible light and near infrared light spectra by use of the wavelet low frequency coefficient. The other was the band composite model. The cross-validation results of the two models were analyzed and compared. The analysis shows that as for the band composite model , the TM4/TM1 reflectance exponential model is the best , while with regard do the prediction capacity , the wavelet-PLS model is of better precision and stability than the band composite model in general , and it is suitable for the quantitative analysis of spectra for the suspend sediment samples.

Key words : suspended sediment ; wavelet coefficient ; partial least-squares regression ; spectral reflectance