

# 基于聚类分析算法的浮游植物吸收系数非线性模型

杨超宇<sup>1</sup> 杨顶田<sup>2\*</sup>

<sup>1</sup> 国家海洋局南海预报中心, 广东 广州 510310

<sup>2</sup> 中国科学院南海海洋研究所, 广东 广州 510301

**摘要** 浮游植物吸收系数是海洋光学研究中重要的参数。该参数随着海区以及时间不同呈现较大的差异。为了更好地表征该参数, 探讨了聚类法在浮游植物吸收系数非线性模型中的应用, 修正了非线性模型中浮游植物吸收特性随区域性变化带来的误差。通过计算不同波段的反射率二阶导光谱和吸收系数光谱的自适应指数(ARI)发现 555~681 nm 波段的反射率二阶导光谱可以较好地表征吸收系数光谱的光谱特征。因此, 以 555~681 nm 的反射率二阶导光谱为聚类分析样本对数组进行聚类分析。通过实测数据对聚类法进行了验证, 结果证明该方法可以较好地表征浮游植物吸收系数(均方根误差大于 0.79)。该研究是应用光学遥感方法提取浮游植物信息及海水光学特征必不可少的基础研究。

**关键词** 遥感; 浮游植物; 吸收系数

中图分类号 P733.3+1 文献标识码 A

doi: 10.3788/CJL201542.s114001

## A Non-Linear Model of Phytoplankton Absorption Based on Cluster Analysis

Yang Chaoyu<sup>1</sup> Yang Dingtian<sup>2</sup>

<sup>1</sup> South China Sea Marine Prediction Center, State Oceanic Administration, Guangzhou, Guangdong 510310, China

<sup>2</sup> South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, Guangdong 510301, China

**Abstract** Phytoplankton absorption coefficient is an important parameter for ocean optics. The parameter varies widely with different locations and time. In order to model the phytoplankton absorption coefficient, application of the cluster analysis on the retrieval of phytoplankton absorption coefficient based on a non-linear model of phytoplankton absorption coefficient is discussed. The method reduces the error due to different optical properties of phytoplankton in different areas. By calculating adjusted Rand index (ARI) between the second derivative reflectance and absorption spectra at different wavelengths, the spectral range of 555~681 nm is found to be more sensitive than other bands. The spectral band of 555~681 nm is applied to analyze the second derivative reflectance. *In situ* data are used to validate the cluster analysis method, and it is found that the cluster analysis is effective for phytoplankton absorption model (root mean square error is higher than 0.79). The study is a fundamental research for application of optical remote sensing to retrieving phytoplankton information and ocean component.

**Key words** remote sensing; phytoplankton; absorption coefficient

**OCIS codes** 280.1415; 100.3005; 100.3008; 100.3010

## 1 引 言

浮游植物居于海洋生物食物链的底层, 是整个海洋能量循环的重要元素。因而浮游植物的光学特性对

收稿日期: 2015-01-09; 收到修改稿日期: 2015-04-10

基金项目: 气候变化重大基础研究项目(2013CB956503)、国家自然科学基金(U1405234, 41176161)、国家海洋局南海分局海洋科学技术局长基金(1338)、海洋公益性行业科研专项(2015418014)

作者简介: 杨超宇(1985-), 女, 博士, 工程师, 主要从事浮游植物遥感信息提取和赤潮预报等方面的研究。

E-mail: ycy@scsio.ac.cn

\* 通信联系人。E-mail: dtyang@scsio.ac.cn

于海洋生物资源调查以及海洋生态健康状况调查有着重要的作用。浮游植物吸收系数随着浮游植物粒径增大而减小<sup>[1]</sup>。一般光照充足的富营养水体,叶绿素浓度相对较高且以大型浮游植物为主<sup>[2]</sup>。这种典型案例下,大型浮游植物的打包效应较强,浮游植物吸收系数也相应较小。因此,在应用遥感数据提取浮游植物的光学特性时需考虑其粒径大小。Bricaud 等<sup>[3]</sup>针对一类水体提出了浮游植物吸收系数反演模型[(1)式]。Maritorena 等<sup>[4]</sup>针对两类水体建立了浮游植物反演模型。Matsuoka 等<sup>[5]</sup>针对高纬度海域研究了浮游植物非线性反演回归参数。由此可见,针对光学特征明显不同的海域,浮游植物反演算法具有一定的差异性。为了简化不同海域的浮游植物吸收系数反演模型算法,提出了一个可以应用于不同海域的模型。该模型根据反射率光谱特征提取相应的区域性模型参数,通过不同参数选取进一步修正了不同海域浮游植物吸收特征差异所引起的误差。

$$a_{ph} = A_p C^{B_p}, \quad (1)$$

式中  $A_p, B_p$  为根据叶绿素浓度  $C$  反演浮游植物吸收系数的模型参数(表 1),  $a_{ph}$  为浮游植物吸收系数( $m^{-1}$ )。

$$a_{ph} = A_p a_{ph}^{B_{ph}}, \quad (2)$$

式中  $A_{ph}, B_{ph}$  为 440 nm 处浮游植物吸收系数反演浮游植物吸收系数模型的参数。

表 1 浮游植物吸收系数反演模型参数

Table 1 Parameters in the non-linear model of phytoplankton absorption

Wavelength/nm	411	443	490	510	555	620	665	681	709
$A_p$	0.0323	0.039	0.0274	0.018	0.0068	0.0065	0.0144	0.0182	NaN
$B_p$	0.286	0.348	0.361	0.26	0.027	0.064	0.131	0.155	NaN
$a_{ph}/(m^2/mg)$	0.00665	0.05582	0.02055	0.0191	0.01015	0.00975	0.01424	0.02215	0.00161
$A_{ph}$	0.9318	0.8955	0.6031	0.4922	0.2071	0.1928	0.3009	0.4898	NaN
$B_{ph}$	1.008	0.997	0.981	0.999	1.008	1.041	1.093	1.070	NaN

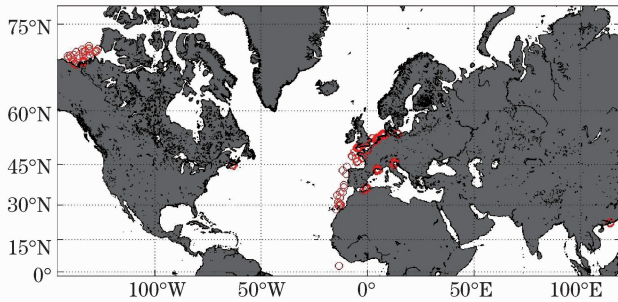


图 1 实验海区的站点分布图

Fig. 1 Location of research stations

## 2 数据与方法

### 2.1 现场实验

数据来自 1997 年 2 月到 1998 年 9 月在欧洲沿海 COASTLOOC 项目( $30^{\circ} \sim 53^{\circ} N, -15^{\circ} \sim 15^{\circ} W$ ), 2009 年 7 月 30 日到 8 月 27 日 MALINA 项目( $69^{\circ} \sim 72^{\circ} N, 125^{\circ} \sim 145^{\circ} W$ ), 2004 年 1 月 1 日到 11 月 15 日在加拿大哈利法克斯 Bedford Basin ( $44.683^{\circ} N, 63.63^{\circ} W$ ) 每周调查实验,以及 2012 年 7 月 5 日珠江口现场实验调查( $22^{\circ} \sim 22.80^{\circ} N, 113.30^{\circ} \sim 114^{\circ} E$ )。该研究的光学参数测量方法参考 SeaWIFS 调查规范进行。水下光谱测量采用 SPMR(Satlantic Inc. Halifax, 加拿大)在水下 0.08~0.016 m 测量了 13 个波段。表面法采用光谱仪 Avaspec-2048FT-3-DT。光谱仪波段为 350~1000 nm;传感器探头 FOV 为  $10^{\circ}$ 。为了测量固有光学参数 IOPs, 采集海水表面水样,并采用 25 mm Whatman GF/F 滤膜过滤。光学密度值(OD)采用 Perkin Elmer Lambda 19 测量<sup>[6]</sup>。滤膜采用次氯酸钠漂白,并测量漂白后的滤膜在 380~750 nm 的 OD 值<sup>[7]</sup>。漂白前与漂白后的吸收系数差异为浮游植物吸收系数<sup>[8]</sup>。色素测量采用高效液相色谱法(HPLC)<sup>[9]</sup>。

### 2.2 研究方法

由于比吸收系数随区域、时间变化,可以通过聚类分析分组反射率二阶导数  $R''$ ,用阈值选取适合该组的

吸收系数反演模型。该方法可分 4 步实现(图 2):1)假设可以用  $N$  个簇总结所有  $a_{ph}$  特征,并满足(1)式;2)计算  $a_{ph}$  和  $R''$  的自适应指数(ARI),比选最优波段( $\lambda_1, \lambda_2, \dots, \lambda_m$ )并构建阈值选取条件;3)聚类分析比选波段的  $R''$ ;4)用每簇  $R''$  对应的  $a_{ph}$  数组构建相应的  $a_{ph}$  经验反演模型,并与每簇  $R''$  形成映射关系。

$$a_{ph}(\lambda) = A_{ph}^*(i, \lambda), \tag{3}$$

式中  $A_{ph}^*$  为第  $i$  个案例的浮游植物吸收系数反演参数( $m^2/mg$ ),  $i = 1, 2, \dots, N$ 。

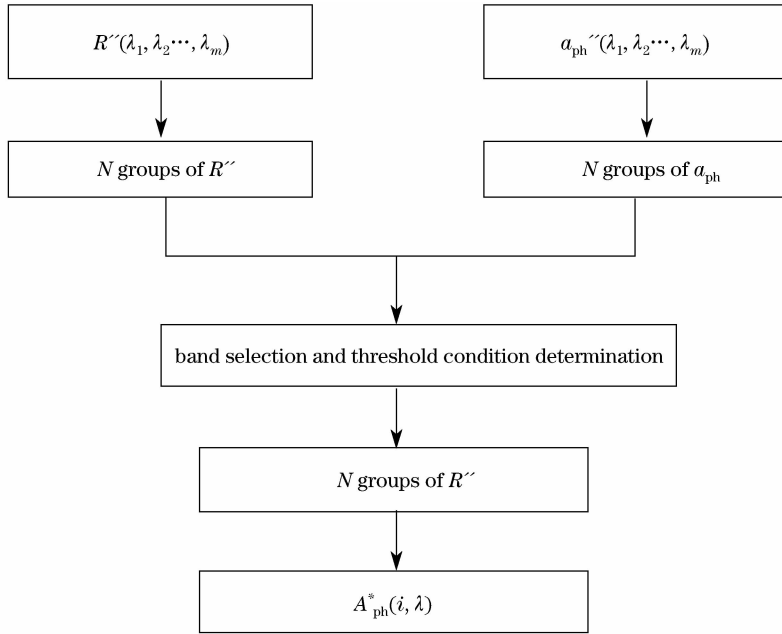


图 2 聚类分析反演  $a_{ph}$

Fig. 2 Retrieval of  $a_{ph}$  based on cluster analysis

### 3 测量分析和讨论

为了确定聚类分析样本波段,首先计算了不同波段的反射率二阶导光谱和吸收系数光谱的自适应指数(表 2)。ARI 指数范围为 0~1。假设某个波段  $\lambda$  的反射率二阶导光谱和吸收系数光谱的类的 ARI 越大,则表示这两个类越相似。从结果中可以看出,555~681 nm 波段的反射率二阶导光谱和吸收系数光谱的分类更为相似。这就意味着当采用 555~681 nm 波段的反射率二阶导光谱为样本进行聚类分析时更能代表吸收系数光谱的光谱特征和分类情况,为聚类分析最优波段。因此,选定 555~681 nm 的反射率二阶导光谱作为聚类分析样本。

表 2 自适应指数

Table 2 Adjusted Rand index

Wavelength /nm	412	443	490	510	555	620	665
490	0.6784	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
510	0.6795	0.6936	0.0000	0.0000	0.0000	0.0000	0.0000
555	0.6918	0.6941	0.6675	0.0000	0.0000	0.0000	0.0000
620	0.6764	0.7014	0.6968	0.6795	0.0000	0.0000	0.0000
665	0.6880	0.6892	0.6891	0.6879	0.7047	0.0000	0.0000
681	0.7002	0.6890	0.6941	0.6906	0.7137	0.6662	0.0000
705	0.6919	0.6684	0.6702	0.6846	0.6981	0.6622	0.6739

为了确定  $N$  的个数,分别采用快速聚类函数 K-means、层次聚类分析等方法进行聚类分析。从图 3 可以看出,当  $N$  从 1 增加到 3,K-means 的组内平方和(SSW)明显减小。但是,当  $N$  大于 3 时,SSW 减小趋势逐渐减弱。为了更好地说明三组模型足够表征所有案例情况,进一步做了聚类分析(图 4)。图 4 更为直观、清晰地展示了三个不同的区域较为集中地聚集了三组样本。此外,层次聚类分析的结果(图 5)也证明了这

一结论。采用 dist 函数计算样本对象的空间距离。假设每个光谱样本为一个类(样本个数为 334),计算所有样本的距离,生成  $334 \times 334$  维度的矩阵,找出矩阵中最小的元素,并将对应的两个类合并为一类。重新计算新的距离,生成新的聚类矩阵,重复计算与合并。该研究在计算距离时采用欧氏距离,并将欧氏距离代入分层聚类函数 hclust 进行聚类。从层次聚类树形图可以看出,若  $N$  小于 3,类间的最大距离过大,相似程度太低,不足以合并一类。若  $N$  大于 3,类间距离也不会再明显减小。基于以上结果分析,选择三类进行聚类分析。

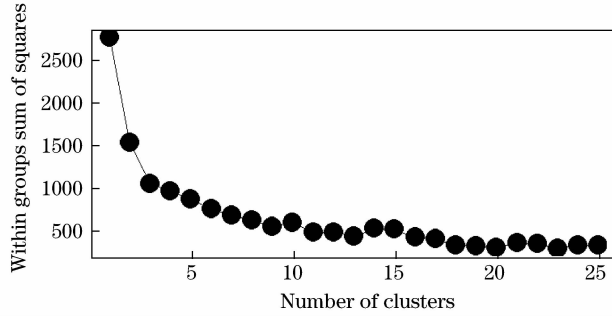


图 3 K-mean 聚类结果图

Fig. 3 Results of K-mean cluster analysis

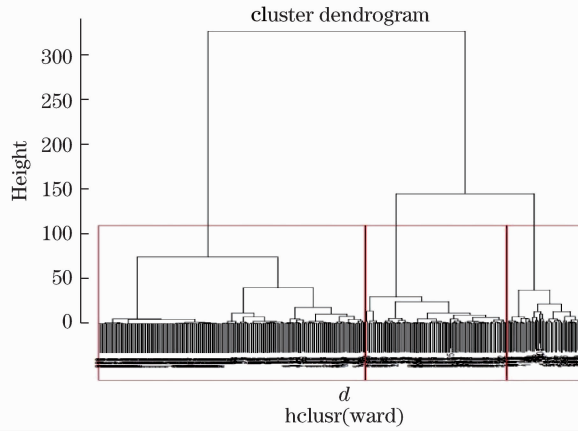


图 4 聚类树形图

Fig. 4 Hierarchical tree plot of cluster analysis

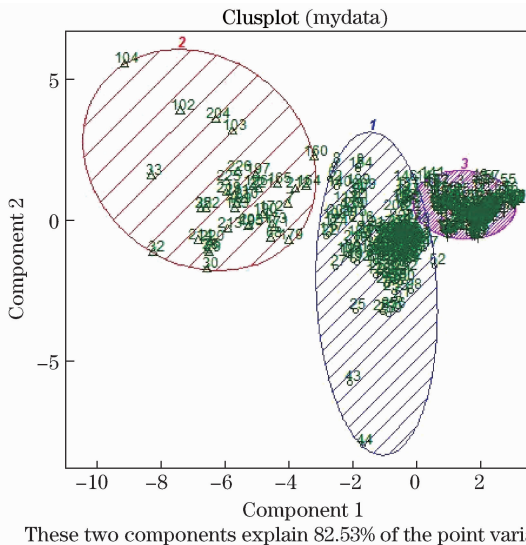


图 5 对  $R''$  进行 Partition Around Medoids 分析,采用 Clusplot() 函数在二维空间绘制聚类结果

Fig. 5 Analysis of  $R''$  using Partition Around Medoids, and plot it by function Clusplot()

确定  $N$  值、确定聚类分析样本最优波段( $\lambda_1, \lambda_2, \dots, \lambda_m$ )后,可以对( $\lambda_1, \lambda_2, \dots, \lambda_m$ )的二阶导数反射率样本进行分类(图 6)。采用 K-means 函数将二阶导数光谱分为三组,查找对应的浮游植物吸收系数。将这三组导数光谱构成一个导数光谱数据库,可以应用这个数据库反演浮游植物吸收系数反演模型参数。为了验证聚类法反演浮游植物比吸收系数的误差,采用与所建立数据库不相关的独立数据进行验证(图 7)。计算验证数据的二阶导数光谱在这三组数据库的最小距离,比选最优,从而匹配该组的浮游植物比吸收系数。图 7 为实测浮游植物吸收系数以及应用聚类匹配最优浮游植物比吸收系数估算的浮游植物吸收系数的比对。通过计算不同波段的均方根误差(RMSE)[(4)式],发现在 555 nm 处反演结果的偏差相对较大,大于 0.79。这可能是由于溶解有机物(CDOM)在该波段的荧光影响造成了一定的光谱解读偏差。而其他波段均大于 0.8。结果表明采用聚类法分析浮游植物比吸收系数是有效的。

$$f_{\text{RMSE}} = \sqrt{\frac{\sum_{i=1}^N (\lg X_{\text{in-situ}} - \lg X_{\text{est}})^2}{N - 2}}, \quad (4)$$

式中  $X_{\text{in-situ}}$  为实测值,  $X_{\text{est}}$  为模型模拟结果,  $N$  为总样本个数。

### 4 结 论

该研究探讨了聚类法分析在浮游植物吸收系数反演中的应用。通过计算不同波段反射率二阶导光谱和吸收系数光谱的 ARI 发现 555~681 nm 波段的反射率二阶导光谱和吸收系数光谱的 ARI 指数更大,分类结果也更为相近。因此,以 555~681 nm 的反射率二阶导光谱为聚类分析样本对数组进行聚类分析。通过计算 K-means 的 SSW, dist 函数的空间距离等方法,确定采用三组数据可以表征所有数据集。通过实测数据对该聚类法进行了验证,发现该方法可以较好地表征浮游植物吸收系数( $f_{\text{RMSE}} > 0.79$ )。该算法的验证海域包括了一类水体,二类水体,甚至是高纬度水体(MALINA 项目;69°~72° N, 125°~145° W)。结果证

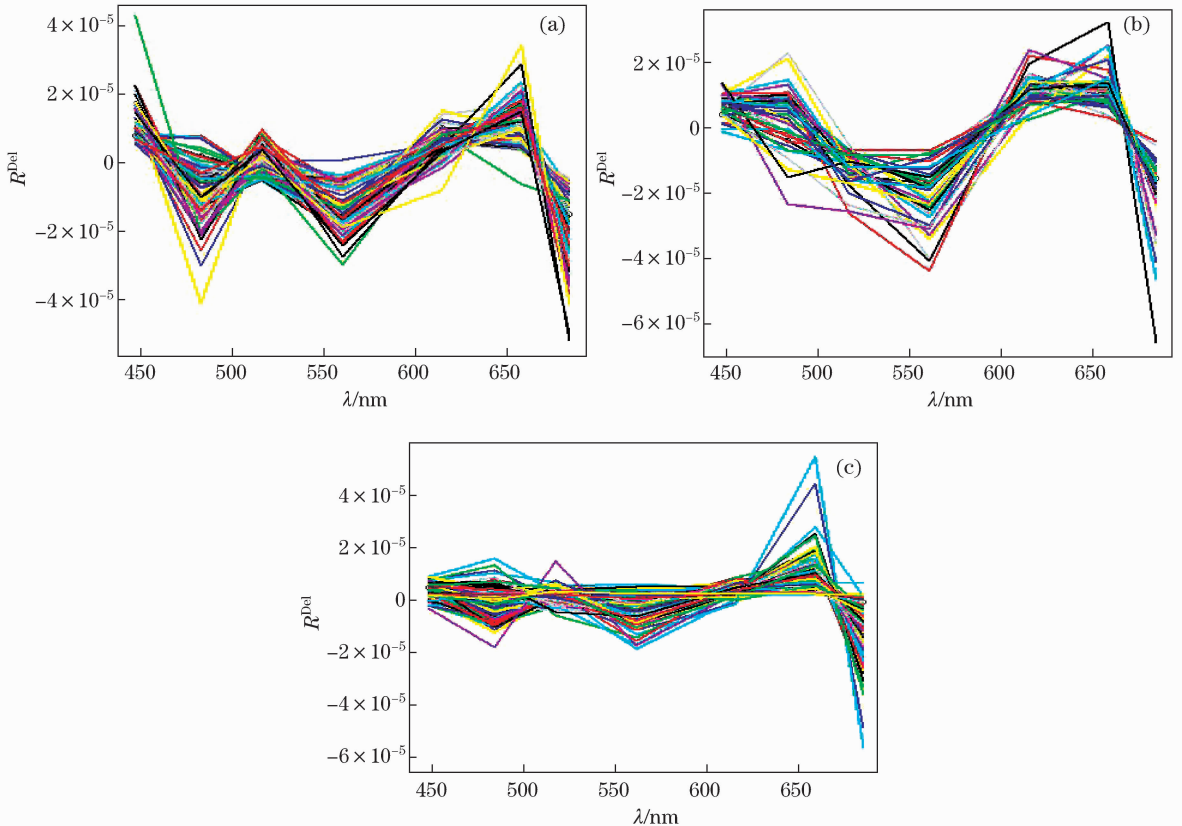


图 6 Clustplot()在二维空间绘制聚类结果。(a)~(c)三簇数据集

Fig. 6 Result of cluster analysis using function Clustplot(). (a)~(c) Three groups of dataset

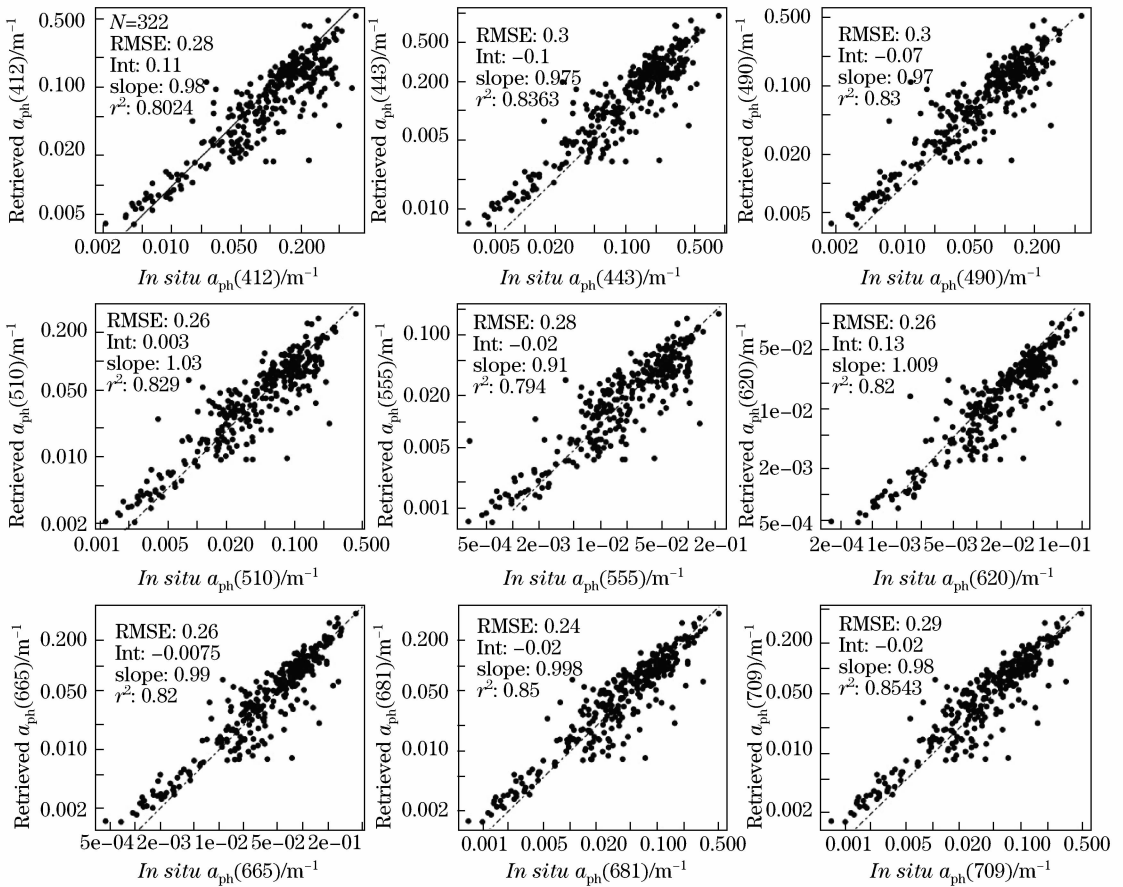


图 7 实测浮游植物吸收系数与应用聚类分析反演结果的对比

Fig. 7 Comparison of *in situ* measured phytoplankton absorption and modeled values

明,该反演算法较好地表征了所有区域的浮游植物吸收系数。反演数据 RMSE 均大于 0.79,且模拟数据与实测数据较好地分布在 1:1 的趋势线周围。因此,可以认为聚类分析算法是能够有效提取浮游植物吸收系数的方法。

### 参 考 文 献

- Babin M, Stramski D, Ferrari G M, *et al.*. Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around Europe [J]. *Journal of Geophysical Research-Oceans*, 2003, 108(C7): 3211.
- Devred E, Sathyendranath S, Platt T. Decadal changes in ecological provinces of the northwest Atlantic Ocean revealed by satellite observations [J]. *Geophysical Research Letters*, 2009, 36(19): L19607.
- Bricaud A, Morel A, Babin M, *et al.*. Variations of light absorption by suspended particles with chlorophyll a concentration in oceanic (case 1) waters: Analysis and implications for bio-optical models [J]. *Journal of Geophysical Research-Oceans*, 1998, 103(C13): 31033-31044.
- Maritorena S, Siegel D A, Peterson A R. Optimization of a semianalytical ocean color model for global-scale applications [J]. *Applied Optics*, 2002, 41(15): 2705-2714.
- Matsuoka A, Bricaud A, Benner R, *et al.*. Tracing the transport of colored dissolved organic matter in water masses of the southern Beaufort Sea: Relationship with hydrographic characteristics [J]. *Biogeosciences*, 2012, 9: 925-940.
- Ferrari G M, Bo F G, Babin M. Geo-chemical and optical characterizations of suspended matter in European coastal waters [J]. *Estuarine, Coastal and Shelf Science*, 2003, 57(1-2): 17-24.
- Tassan S, Ferrari G M. Proposal for the measurement of backward and total scattering by mineral particles suspended in water [J]. *Applied Optics*, 1995, 34(36): 8345-8353.
- Tassan S, Ferrari G M. A sensitivity analysis of the 'transmittance-reflectance' method for measuring light absorption by aquatic particles [J]. *Journal of Plankton Research*, 2002, 24(8): 757-774.
- Vidussi F, Claustre H, Bustillos-Guzman J, *et al.*. Determination of chlorophylls and carotenoids of marine phytoplankton: Separation of chlorophyll a from divinyl-chlorophyll a and zeaxanthin from lutein [J]. *Journal of Plankton Research*, 1996, 18(12): 2377-2382.

栏目编辑: 吴秀娟