

文章编号: 1007-4619 (2004) 01-0031-06

基于遗传算法的二类水体水色遥感反演

詹海刚^{1,2}, 施平¹, 陈楚群¹

(1. 中国科学院南海海洋研究所, 广东 广州 510301; 2. 青岛海洋大学, 山东 青岛 266003)

摘 要: 提出一种基于遗传算法的二类水体水色遥感反演算法。该算法以三成分(叶绿素、悬浮泥沙与黄色物质)海水光学模型作为前向模型, 以实数编码遗传算法作为优化方法, 并采用一对波段比来构造目标函数。模拟反演的结果表明, 该算法可以有效克服已有二类水体水色遥感优化反演方法在搜索策略方面存在的困难, 是一种有较高计算效率、可靠与稳健的反演算法。

关键词: 二类水体; 水色遥感; 优化反演; 遗传算法

中图分类号: TP751.1 **文献标识码:** A

1 引 言

随着新一代水色传感器陆续投入使用, 二类水体的水色遥感引起广泛关注^[1]。如何把海量的遥感数据量化为水体环境参数信息是水色遥感界面临的重要课题。其中, 反演算法是水色遥感信息量化的三个主要研究方面之一, 也是其中的一个关键。

历史上已提出多种二类水体的水色遥感反演算法。这些反演算法本质上可归结为两类。其一为建模方法, 又称显式方法(explicit solutions)^[2], 即利用实测或模拟数据, 把海水成分浓度表示成表观光学量(如离水辐射率、遥感反射率等)的函数, 这可通过统计分析、半分析或人工神经网络等方法加以实现^[3-6]; 其二为优化方法, 又称隐式方法(implicit solutions)^[2], 即先选定各水体成分的初始浓度值, 并由前向光学模型求得相应的表观光学量, 通过对一个度量计算与实测表观光学量之差距的目标函数的寻优迭代, 不断地调整各成分的浓度, 直到满足逼近条件为止, 最后所得各成分浓度值即为反演值。建模方法本质上是一种预测学习方法, 它沿袭了一类水体反演的基本思路, 即假设研究海域的反演模型可用确定性数学式子表示, 因而是一种基于海域的反演算法。实际上在二类水体海域这种假设常常并不

成立。即使成立, 受学习数据的数量和代表性的影响, 建模方法的应用仍受到很大的限制^[1]。优化方法是一种基于单个像元的局域性算法, 反演精度高但计算量比较大。随着计算机硬件技术的飞速发展, 优化方法有望成为二类水体水色遥感反演的主流算法之一。

已有优化反演方法的一个明显不足是采用局部搜索策略。二类水体水色遥感反演的目标函数是一个多波段多变量的非线性函数, 存在多个局部极值, 而已有优化反演方法所采用的搜索方法, 如基于导数的 Gauss-Newton 法^[7] 和 Levenberg-Marquart 法^[8,9], 非导数的下山单纯形算法^[10] 等, 都是局部搜索方法。对于这些局部搜索方法, 初始位置的选择对最终结果起决定性的作用, 如果初始位置选择不当, 则不可避免地会陷入局部极值点, 导致反演结果较差甚至没有物理意义^[9]。近年来发展起来的遗传算法是一种全局搜索算法, 采用遗传算法作为搜索策略, 有望可以克服优化反演的这个不足。

2 三成分海水光学模型

海洋水色遥感中, 平静海面下, 离水辐射率所携带的水体信息可由刚好在水表面下的反射比 $R(\lambda)$ 来表征。根据辐射传输方程的计算结果, $R(\lambda)$ 可用

收稿日期: 2002-07-10; 修订日期: 2002-10-08

基金项目: 本研究得到中国科学院 KZCX2-202 项目、广东省自然科学基金 032616 项目资助

作者简介: 詹海刚(1968—), 男, 副研究员, 毕业于青岛海洋大学物理海洋学专业(理学博士)。现主要从事海洋水色遥感信息提取与应用研究, 已发表论文 10 多篇。

后向散射系数 $b_b(\lambda)$ 和吸收系数 $a(\lambda)$ 近似表示。

$$R(\lambda) = 0.33 \frac{b_b(\lambda)}{a(\lambda) + b_b(\lambda)} \quad (1)$$

其中, $b_b(\lambda)$ 和 $a(\lambda)$ 是水体各成分散射和吸收的简

单线性和。若只考虑叶绿素(C)、悬浮泥沙(S)和黄色物质(Y)等 3 种物质的贡献,则 $b_b(\lambda)$ 和 $a(\lambda)$ 可表示为(忽略黄色物质的后向散射)。

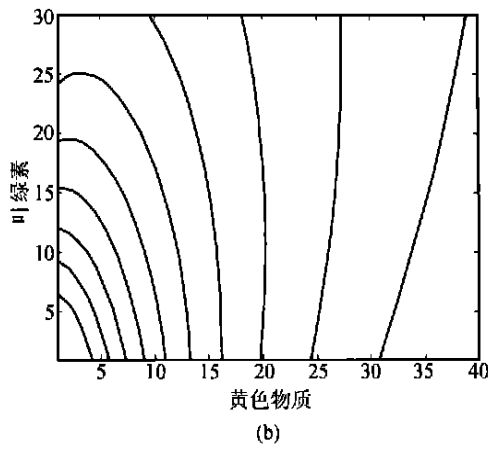
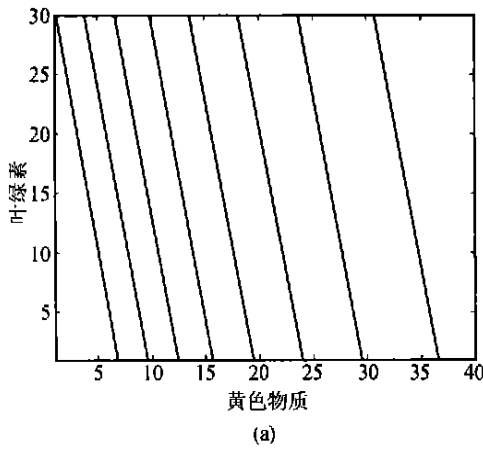
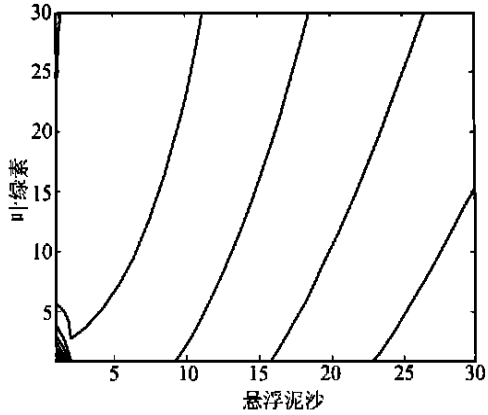
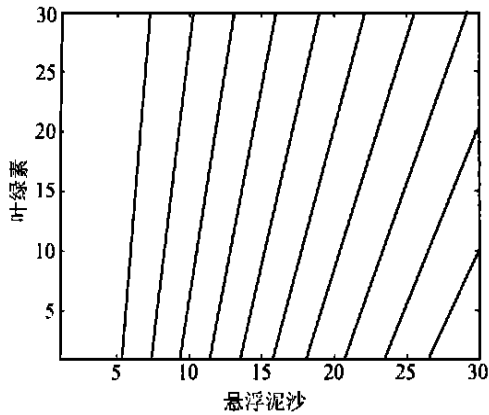


图 1 黄色物质与悬浮泥沙分别为 0.515m^{-1} 与 8.25mg/L 时目标函数的等值线图

(a)目标函数为 f_1 ; (b)目标函数为 f_2

Fig.1 Contour plot of f_1 and f_2

$$\begin{cases} b_b(\lambda) = b_{bw}b_w(\lambda) + b_{bc}b_c(\lambda) + b_{bs}b_s(\lambda) \\ a(\lambda) = a_w(\lambda) + Ca_C^*(\lambda) + Sa_S^*(\lambda) + Ya_Y^*(\lambda) \end{cases} \quad (2)$$

(2)式中, $b_w(\lambda)$, $b_c(\lambda)$ 和 $b_s(\lambda)$ 分别为纯水、叶绿素和悬浮泥沙的体散射系数, b_{bw} , b_{bc} 和 b_{bs} 分别为纯水、叶绿素和悬浮泥沙的后向散射的比例; $a_w(\lambda)$ 为纯水的吸收系数, $a_C^*(\lambda)$, $a_S^*(\lambda)$ 和 $a_Y^*(\lambda)$ 分别为叶绿素、悬浮泥沙和黄色物质的单位吸收系数, C , X 和 Y 分别为叶绿素、悬浮泥沙和黄色物质的浓度, 其中黄色物质浓度以 440nm 处的吸收系数表示。各系数的计算见文献[3, 6]。

3 目标函数与遗传算法的设计

优化反演常常采用遥感数据与计算数据的误差平方和作为目标函数,对 SeaWiFS 数据的反演而言,此目标函数为

$$f_1 = - \sum_i (R_i^m - R_i^c)^2 \quad (3)$$

式中, R 为刚好在水表面下的反射比, $i=1, \dots, 6$ 为 SeaWiFS 的前 6 个波段, 上标 m, c 分别代表测定值和计算值。图 1 是黄色物质与悬浮泥沙分别为固定值时目标函数的二维等值线图。可以看到, f_1 对叶绿素浓度的变化不敏感, 如果把它作为目标函数, 可能会影响水体成分, 特别是叶绿素浓度的反演精度。

要提取二类水体各成分浓度,关键是要分离不同水体成分对总光谱响应的贡献。对 SeaWiFS 而言,悬浮泥沙光谱贡献比较容易分离,因为与悬浮泥沙后向散射密切相关的波段 6 刚好位于叶绿素与黄色物质低吸带,受后两者的干扰较小。叶绿素与黄色物质的光谱形状比较相似,其差别主要体现在波段 2 与波段 1,尤其是波段 1,是分离叶绿素与黄色物质光谱贡献的关键波段^[6,11]。历史上的反演算法多采用一对互补的波段比来实现此两成分光谱的分离^[3,11]。本文也尝试采用波段比与波段相结合的误差平方和来定义目标函数:

$$f_2 = - \left[\left(\frac{R_2^a}{R_5^a} - \frac{R_2^c}{R_5^c} \right)^2 + \left(\frac{R_1^a}{R_3^a} - \frac{R_1^c}{R_3^c} \right)^2 + (R_4^a - R_4^c)^2 + (R_6^a - R_6^c)^2 \right] \quad (4)$$

从图 1 可以发现, f_2 对 3 种成分的变化都比较敏感。

我们采用 GAOT 遗传算法软件^[12]来实施反演算法。该软件提供多种编码方案与遗传算子。本文采用实数编码方案,即直接用浮点数串来表示成分浓度值,这种编码方案可以克服二进制编码由于 Hamming 悬崖与编码过长等引起的搜索效率低、缺乏微调功能等缺点,而且容易引入与问题相关的启发式信息以增加搜索能力。选择算子采用一种非线性排名策略,即正规化几何排名(Normalized geometric ranking)选择^[12],这种选择策略根据个体的适应值在群体中的排名来分配其选择概率,因而可以避免算法出现过早收敛和停滞现象。在杂交和变异操作中,同时使用几种算子。杂交算子包括简单杂交、算

术杂交和启发式杂交 3 个算子,变异算子包括边界变异、均匀变异、非一致性变异与多非一致性变异(multi-non-uniform mutation) 4 个算子^[12]。这些算子的搜索能力具有一定的倾向性,同时使用可以利用它们各自的长处,从而提高搜索的效率^[13]。

4 反演试验与比较

应用 3 成分海水光学模型产生一组模拟数据,方法如下:将选定水体成分浓度的区间(叶绿素浓度为 $0.5-15 \text{mg} \cdot \text{m}^{-3}$,悬浮泥沙浓度为 $1-30 \text{mg/L}$,黄色物质浓度为 $0.2-2 \text{mg} \cdot \text{m}^{-1}$) 4 等分,取各种成分浓度的所有组合,共 125 组,计算相应 6 个波段的 R 值,将其当作测定值 R_i^a 。如无特别说明,取如下参数值进行反演试验:群体规模为 100,遗传代数为 100,其它参数采用 GAOT 的默认值。

4.1 目标函数对反演结果的影响

表 1 给出了不同目标函数与遗传代数反演算法的反演浓度与输入浓度之间的均方根差(RMS)与可决系数(Rsq)。图 2 给出目标函数为 f_1 、遗传代数为 1000,以及目标函数为 f_2 、遗传代数为 100 反演浓度与输入浓度之间的散点图。可以看到,如果遗传代数足够大,两种目标函数都可获得较好的反演结果。采用 f_2 为目标函数,遗传代数为 100 时叶绿素的反演精度高于以 f_1 为目标函数,遗传代数达到 1000 时叶绿素的反演精度,表明采用 f_2 为目标函数极大地提高了反演算法的搜索效率。

表 1 目标函数与遗传代数对反演精度的影响

Table 1 Inversion accuracy affected by different objective functions and generations

目标函数	遗传代数	叶绿素		悬浮泥沙		黄色物质	
		RMS	Rsq	RMS	Rsq	RMS	Rsq
f_1	100	1.915	0.863	0.903	0.992	0.032	0.998
	500	1.029	0.963	0.418	0.998	0.018	0.999
	1000	0.433	0.994	0.149	1.000	0.009	1.000
f_2	100	0.331	0.996	0.219	1.000	0.015	1.000
	200	0.104	0.999	0.056	1.000	0.006	1.000
	300	0.055	1.000	0.024	1.000	0.002	1.000

4.2 波段组合对反演结果的影响

除 R_1/R_3 与 R_2/R_5 外, 目标函数采用其它波段比, 如 R_1/R_3 与 R_3/R_5 , R_1/R_2 与 R_2/R_5 等在遗传代

数为 100 时也有较高的反演精度(表 2)。而采用 R_1/R_3 或 R_2/R_5 等单波段比, 或 Tassan 采用的双重波段比 $(R_2/R_5)/(R_1/R_3)$ 与 $(R_5+R_6)/(R_3/R_5)$ ^[3], 则对提高算法的搜索效率并无帮助。

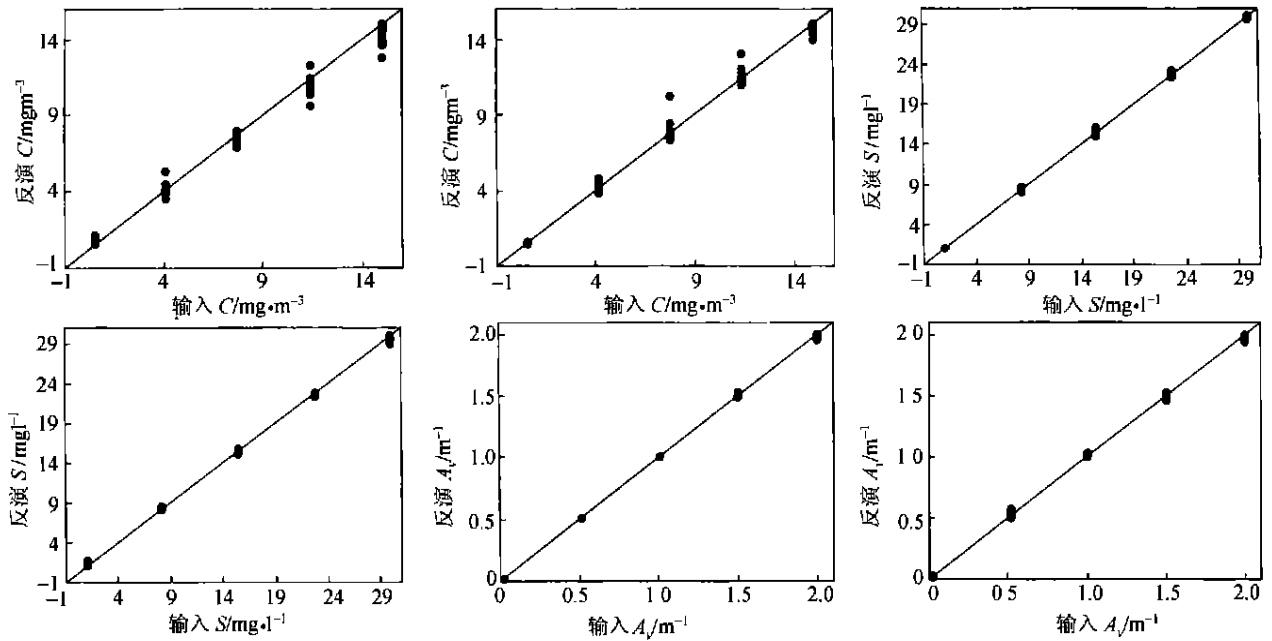


图 2 不同目标函数与遗传代数的叶绿素(C)、悬浮泥沙(S)与黄色物质(A_y)的输入值与反演值比较
(a)目标函数为 f_1 , 遗传代数为 1000; (b)目标函数为 f_2 , 遗传代数为 100

Fig. 2 Comparisons of the input and the retrieved concentrations of chlorophyll (C), suspended sediment (S) and yellow substance absorption (A_y)

表 2 波段组合对反演精度的影响

Table 2 Inversion accuracy affected by different combination of bands

波段组合	叶绿素		悬浮泥沙		黄色物质	
	RMS	Rsq	RMS	Rsq	RMS	Rsq
$R_1/R_3, R_3/R_5, R_2, R_4, R_6$	0.266	0.997	0.102	1.000	0.010	1.000
$R_1/R_2, R_2/R_5, R_3, R_4, R_6$	0.193	0.999	0.176	1.000	0.012	1.000
$(R_2/R_5)/(R_1/R_3), (R_5+R_6)/(R_3/R_5), R_4$	1.576	0.906	0.216	1.000	0.248	0.875
$R_1/R_3, R_2, R_5, R_4, R_0$	1.141	0.952	0.474	0.998	0.110	0.976
$R_2/R_5, R_1, R_3, R_4, R_6$	3.884	0.506	1.469	0.980	0.112	0.975

4.3 误差对反演结果的影响

实际海洋水色遥感中存在多种噪声源, 这些噪声源对各波段的影响基本相似^[14], 因此我们对各波段的反射比加上相同的误差项, 以考察算法对误差的敏感性。假定误差服从高斯分布 $N(0, 1)$, 分别对 $R(\lambda)$ 加上 10%, 20% 和 30% 三种大小的误差, 反演

结果列于表 3。从表 1 与表 3 可以看到, 对于悬浮泥沙和黄色物质, 误差对两个目标函数的反演算法的影响都不甚显著, 当误差加大到 30% 时, 又以 f_1 为目标函数反演算法的精度下降较快; 对于叶绿素, 以 f_2 为目标函数的反演算法对误差的敏感性较低, 算法的稳定性较好。

表 3 误差对反演精度的影响

Table 3 Inversion accuracy affected by different errors

目标函数	遗传代数	误差大小/%	叶绿素		悬浮泥沙		黄色物质	
			RMS	Rsq	RMS	Rsq	RMS	Rsq
f_1	1000	10	1.592	0.918	0.796	0.995	0.023	0.999
		20	1.713	0.892	0.953	0.992	0.093	0.982
		30	3.036	0.682	4.733	0.802	0.238	0.889
f_2	100	10	0.706	0.982	1.813	0.975	0.094	0.984
		20	0.972	0.964	2.497	0.944	0.103	0.980
		30	1.372	0.929	4.641	0.809	0.143	0.961

4.4 搜索区间对反演结果的影响

实际问题中, 往往很难对搜索区间有确知的先验知识, 因而有必要探讨搜索区间对反演结果的影响。保持搜索区间的最小值不变, 增大区间的最大值, 以 f_2 为目标函数执行反演, 结果见表 4。结合

表 1(搜索区间为 $[0.5, 1, 0.02] - [15, 30, 2]$) 与表 4 的结果, 可以看出, 设置足够的遗传代数, 不同搜索区间下算法的反演精度并不存在显著差异, 但搜索区间越小, 算法的计算开销越小。换句话说, 改变搜索区间对算法的计算效率有影响, 但在保证算法收敛的前提下, 对算法的反演精度影响不大。

表 4 搜索区间对反演结果的影响

Table 4 Inversion accuracy affected by different intervals

搜索区间	最大代数	叶绿素		悬浮泥沙		黄色物质	
		RMS	Rsq	RMS	Rsq	RMS	Rsq
$[0.5, 1, 0.02] - [30, 60, 4]$	100	1.047	0.970	0.626	0.997	0.060	0.994
	200	0.509	0.991	0.084	1.000	0.022	0.999
	300	0.409	0.994	0.057	1.000	0.016	1.000
$[0.5, 1, 0.02] - [60, 90, 6]$	100	3.409	0.809	2.250	0.967	0.203	0.949
	200	1.459	0.940	2.218	1.000	0.068	0.997
	300	0.701	0.984	0.096	1.000	0.029	0.998

4.5 与其它优化方法的比较

取各成分相邻浓度的中间值组成 4 组初始值, 采用已有二类水体优化反演使用的优化方法, 如下山单纯形^[10]、Levenberg-Marquart^[8,9] 与 Gauss-Newton^[7]等算法, 执行目标函数 f_2 的反演计算, 表 5 给出这些优化算法不能收敛到全局最小点的样本数。可以看到, 对于不同的初始值, 这些算法都存在一些反演失败的样本。

可以看到, 对于不同的初始值, 这些算法都存在一些反演失败的样本。

表 5 其它搜索方法不能收敛到全局最小点的样本数

Table 5 Number of successes out of 125 samples inverted by local optimization approaches

初始值			算法		
叶绿素	悬浮泥沙	黄色物质	单纯形	Levenberg-Marquardt	Gauss-Newton
2.3125	4.625	0.2675	9	11	13
5.9375	11.875	0.7625	2	7	12
9.5625	19.125	1.2575	3	5	8
13.1875	26.375	1.7525	2	4	1

5 结论和讨论

本文的研究表明,采用遗传算法作为二类水体水色遥感反演的搜索策略,可以获得比已有优化反演算法更为合理、可靠的结果。该算法的主要缺点是计算时间太长,尤其是采用传统方法构建目标函数时,算法的搜索效率较低,从而制约反演算法的实际应用。根据二类水体水色遥感反演的特点,利用某些波段比的信息互补性设计目标函数,则可以有效分离不同水体成分的光谱贡献,极大地提高反演算法的计算效率。充分利用先验知识的约束(如使用合适的搜索区间),采用实数编码策略并引入局部搜索能力较强的算子,也是提高反演算法计算效率的有效方法。

参考文献 (References)

- [1] IOCCG. Remote Sensing of Ocean Colour in Coastal and Other Optically-Complex Waters [A]. Sathyendranath, S. Reports of the International Ocean-Colour Coordinating Group, No. 3, IOCCG, Dartmouth, Canada, 2000.
- [2] Mobley C D. Light and Water-Radiative Transfer in Natural Waters [M]. Academic Press, 1994.
- [3] Tassan S. Local Algorithms Using SeaWiFS Data for the Retrieval of Phytoplankton Pigment, Suspended Sediment and Yellow Substance in Coastal Waters [J]. *Applied Optics*, 1994, **33**:2368—2378.
- [4] Carder K L, Chen F R, Lee Z P, et al. Semianalytic Moderate-Resolution Imaging Spectrometer Algorithms for Chlorophyll a and Absorption with Bio-optical Domains Based on Nitrate-depletion Temperatures [J]. *Journal of Geophysical Research*, 1999, **104**:5403—5421.
- [5] Zhan Haigang, Shi Ping, Chen Chuqun. Inversion of Oceanic Chlorophyll Concentrations by Neural Networks [J]. *Chinese Science Bulletin*, 2001, **46**:158—161.
- [6] Sathyendranath S, Prieur L, Morel A. A Three Component Model of Ocean Colour and its Application to Remote Sensing of Phytoplankton Pigments in Coastal Waters [J]. *International Journal of Remote Sensing*, 1989, **10**:1373—1394.
- [7] Fischer J, Doerffer R. An Inverse Technique for Remote Detection of Suspended Matter, Phytoplankton and Yellow Substance from CZCS Measurement [J]. *Adv. Space Res.*, 1987, **7**(2):21—26.
- [8] Konrdatyev K YA, Pozdnyakov D V, Pettersson L H. Water Quality Remote Sensing in the Visible Spectrum [J]. *International Journal of Remote Sensing*, 1998, **19**:957—979.
- [9] Bukata R P, Jerome J H, Konrdatyev K YA, et al. Optical Properties and Remote Sensing of Inland and Coastal Waters [M]. Boca Raton: CRC Press, 1995.
- [10] Doerffer D, Fischer J. Concentrations of Chlorophyll, Suspended Matter, and Gelbstoff in Case II Waters Derived from Satellite Coastal Zone Color Scanner Data with Inverse Modeling Methods [J]. *Journal of Geophysical Research*, 1994, **99**:7457—7466.
- [11] Carder K L, Hawes S K, Baker K A, et al. Reflectance Model for Quantifying Chlorophyll a in the Presence of Productivity Degradation Products [J]. *Journal of Geophysical Research*, 1991, **96**:20599—20611.
- [12] Houck C, Joines J, Kay M. A Genetic Algorithm for Function Optimization: A Matlab Implementation [R]. NCSU-IE, 1995. Available from <http://www.ie.ncsu.edu/mirage/GAToolBox/gaot/>.
- [13] Michalewicz Z. Genetic Algorithms + Data Structures = Evolution Programs [M]. Berlin:Springer-verlag, 1994.
- [14] Durand D, Bijaoui J, Cauneau F. Optical Remote Sensing of Shallow-water Environment Parameters: A Feasibility Study [J]. *Remote Sensing of Environment*, 2000, **73**:152—161.

A Genetic Algorithm for Retrieval of Water Constituents from Ocean Color Remote Sensed Data in Case 2 Waters

ZHAN Hai-gang^{1,2}, SHI Ping¹, CHEN Chu-qun¹

(1. South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou 510301, China;

2. Ocean University of Qingdao, Qingdao 266003, China)

Abstract: The optimization approach is one of the most promising methods for retrieval of water constituents in case 2 waters, but almost previous applications of this approach suffer from their local search techniques. In this study, a genetic algorithm is developed as a global optimization scheme to simultaneously retrieve concentrations of chlorophyll, suspended sediment and yellow substance. To separate the contributions to the radiance spectra by co-existing constituents, two reflectance ratios were embodied to the objective function, and a real-valued genetic algorithm was used to optimize it. The performance of the algorithm is demonstrated with a simulated data set. Under noise free conditions, three water constituents are estimated accurately. Tests with noisy data show that the algorithm is robust against errors in the reflectance data.

Key words: case 2 waters; ocean color remote sensing; inversion; genetic algorithm