利用神经网络反演海水叶绿素浓度

詹海刚[①②] 施 平[①] 陈楚群[①]

(①中国科学院南海海洋研究所，广州 510301；②中国水产科学研究院南海水产研究所，广州 510300.
Email: zhanhaigang@yahoo.com)

摘要 利用美国国家航空与航天局 SeaBAM 小组收集的海水叶绿素浓度和辐射同步实测资料，建立由遥感反射率反演海水叶绿素的神经网络模型，并与 SeaBAM 统计算法的计算结果进行比较. 神经网络模型的反演值在叶绿素浓度略小于 0.03 mg/m^3 的很小范围内有些偏高，在叶绿素浓度的其他区间精度较高，在二类海水区域则明显好于 SeaBAM 统计算法. 总体而言，神经网络模型的反演结果优于 SeaBAM 统计算法.

关键词 神经网络 叶绿素浓度 遥感反射率 反演 二类海水

自 20 世纪 70 年代第 1 代海洋水色传感器——海岸带水色扫描仪 CZCS(Coastal Zone Color Scanner)投入使用至今，已发展起多种提取海水叶绿素浓度的经验、半分析和基于辐射传递模型的理论算法[1,2]. 以“比值波段组合”为基础的经验统计算法因其应用简便和在一类海水中较高的反演精度而成为业务上的主要算法. 为配合第 2 代海洋水色传感器——海视宽视野传感器 SeaWiFS(Sea-viewing Wide Field-of-view Sensor)的投入使用，美国国家航空与航天局的 SeaBAM(SeaWiFS Bio-optical Algorithm Mini-Workshop)小组收集全球范围内海水叶绿素浓度与辐射的同步观测数据，并应用此数据对 15 个海洋水色经验模型和两个半分析模型作出评价，提出适用于 SeaWiFS 的海水叶绿素浓度统计算法[2].

海水叶绿素浓度的反演从本质上说是不确定性问题[3]，在含有二类海水的情况下反演模型是局域性质的[4]，因此其统计算法不具备全局性，回归模型形式常常难以事先确定，需要进行大量的试验，而且容易丢失有效信息. 神经网络作为一种有效的非线性逼近方法，近年来在海洋水色反演中已有应用[5~7]，但这些反演所使用的数据或数量太少或是模拟光谱的结果，其结论的合理性未得到真实情况的检验. 本文基于 SeaBAM 数据，利用神经网络反演海水叶绿素浓度，在保持 SeaBAM 算法优化性能的同时，反演精度得到提高.

1 资料与方法

(1) 资料. SeaBAM 从其收集的数据中经质量控制与校正处理选出 919 组海水叶绿素浓度和与 SeaWiFS 所设波段一致的辐射同步实测资料，所有辐射数据均转化为水面遥感反射率. 叶绿素浓度值介于 0.019 与 32.79 mg/m^3 之间，平均浓度为 0.915 mg/m^3，大部分为一类非极地水，有 19 组数据可确定为沿岸二类海水，来自中高营养盐海水的数据偏多，但全部数据仍具一定的全球代表性.

(2) 反演模型. 采用 3 层 BP 神经网络(Backpropagation NN)模型，以遥感反射率及相应的叶绿素浓度为网络的输入与输出，以双曲正切函数作为神经元的激励函数. 原始数据经对数变换后用标准化方法作正归化处理. 采用 Levenberg-Marquardt 快速学习算法[8]训练网络，即网络的权值与偏置按以下方式更新:

$$x_{k+1} = x_k - [\boldsymbol{J}^{\mathrm{T}}\boldsymbol{J} + m\boldsymbol{I}]^{-1}\boldsymbol{J}^{\mathrm{T}}\boldsymbol{e},$$

式中 J 为 Jacobi 矩阵，由 BP 算法求得，e 为方向矢量，参数m 决定了训练的进程. 为避免过拟合，将 MacKey 的 Bayes 正则化方法应用于 Levenberg-Marquardt 学习算法[9]，从而改善网络的推广性能. 本研究的多次数值试验显示，与其他训练策略和正则化方法相比，该模型在运算速度与推广能力方面具有一定的优越性.

2 结果与分析

将 SeaBAM 数据按叶绿素浓度值升序进行排列，取其奇数序 460 组数据为训练集，其偶数序 459 组数据作为测试集，两个数据集有相似的频率分布，并分别包含 10 组与 9 组二类海水的数据. 随机组合波段作为网络输入，用训练集的数据对网络进行训练，以测试集的反演结果与实测叶绿素浓度的均方根差(RMS)的变化来选定最佳隐结点数目. 结果显示对所有波段组合输入，隐层小于 7 个结点即可获得最佳反演精度. 表 1 给出反演精度较高的几种神经网络模型对测试集的反演对数值与叶绿素浓度实测对数值之间的可决系数(Rsq)和均方根差，作为比较，也给出相应波段组合的 SeaBAM 统计算法对相同测试集数据的计算结果. 可以看出，各种输入神经网络模型的反演结果较相应的统计算法均有一定程度的改进，采用 443，555 nm 与 490，555 nm 二波段输入神经网络模型即可达到 SeaBAM 统计算法中精度最高的需 4 个波段遥感反射率数据 OC4 的反演精度.

表 1 神经网络与统计算法的反演结果比较

输入波段/nm	神经网络			统计回归		
	隐单元数	Rsq	RMS	算法	Rsq	RMS
412, 555	4	0.913	0.175	OC2a	0.897	0.193
443, 555	4	0.932	0.155	OC2b	0.919	0.171
490, 555	4	0.927	0.160	OC2	0.919	0.172
最大值(443, 490), 555	5	0.938	0.148	OC3d	0.927	0.163
最大值(443, 490, 510), 555	5	0.938	0.148	OC4	0.929	0.161
443, 490, 555	6	0.937	0.149			
443, 490, 510, 555	7	0.948	0.136			
412, 443, 490, 510, 555	4	0.947	0.136			

图 1 是 SeaBAM 统计算法 OC2 与 OC4 及相应神经网络模型 NN2(输入波段为 490 和 555 nm)与 NN4 (输入波段为 443，490，510 和 555 nm)对测试集的反演对数值与叶绿素浓度实测对数值的散点、q-q 和相对频率图. 如图 1 所示，与 SeaBAM 统计算法相比，神经网络模型的 q-q 图仍保持较好的线性与连续性；结合叶绿素浓度不同频率分部的偏倚(Bias)及 RMS 的分布(图 2)可以看到，除叶绿素浓度值约小于 0.03 mg/m^3 的很小范围内反演值有点偏高外，NN2 与 NN4 在叶绿素浓度大部分范围内的反演结果较好，尤其在叶绿素浓度中高值区比 SeaBAM 统计算法有较明显的改进.

造成神经网络模型在低值区的反演值偏高主要有两个因素，其一为此区间训练集的数据量相对较少；其二为神经网络模型所使用的误差准则函数，对固定的误差准则函数，叶绿素实测值过大的范围可以导致低值区较大的相对误差[10]. 为考察叶绿素浓度值小于 0.031 mg/m^3 的 17 个测试集样本的反演结果随训练集叶绿素实测值范围的变动而变化的状况，对训练集按

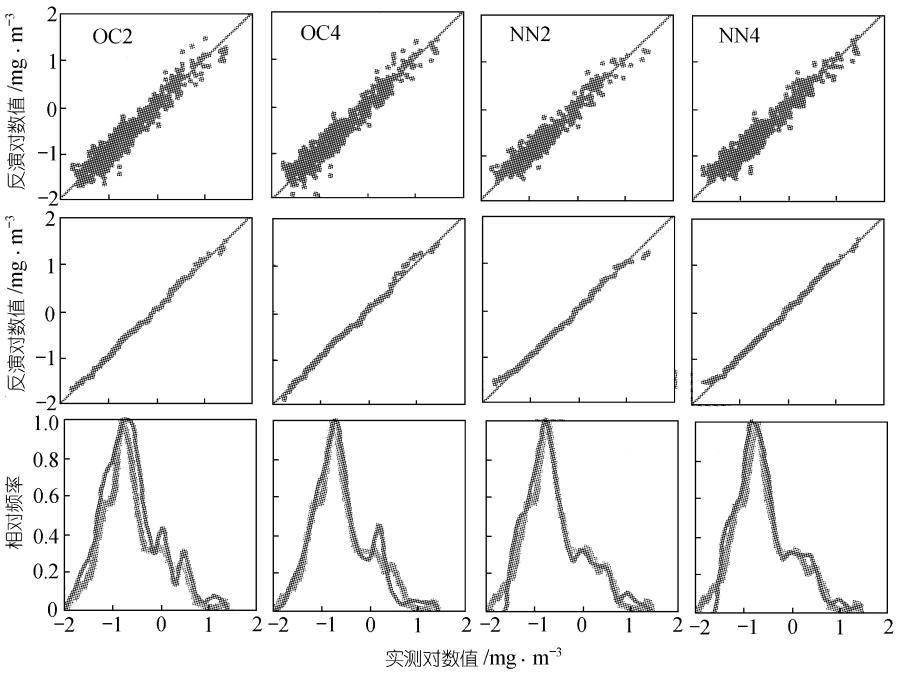


图 1 叶绿素浓度反演对数值与实测对数值的比较

从上至下分别为散点、q-q 和相对频率图. 黑线为反演值, 灰线为实测值

叶绿素浓度从大到小逐个剔除样本后重新训练网络. 表 2 列出部分结果, 其中 Bias 与 RMS1 为 17 个低值区测试样本反演结果, RMS2 为全体测试样本的反演结果. 可以发现, 剔除叶绿素浓度较大值样本, 可在一定程度上改善 NN2 在低值区的反演, 但反演精度仍比统计方法差, 而且 NN4 在低值区的反演精度进一步降低, 整个测试集的反演结果也受到较大影响, 说明训练集数据经对数变换后叶绿素浓度实测值的范围已得到有效减小, 叶绿素浓度低值区反演精度的提高主要还有赖于此区间训练集样本数量的增加.

表 2 低值区测试样本与全体测试样本反演结果随叶绿素浓度区间的变化

叶绿素区间/mg·m^{-3}	训练集样本数	NN2			NN4		
		Bias	RMS1	RMS2	Bias	RMS1	RMS2
[0.019, 32.787]	460	0.170	0.209	0.160	0.142	0.181	0.136
[0.019, 24.300]	458	0.166	0.206	0.161	0.157	0.191	0.143
[0.019, 17.010]	456	0.163	0.204	0.163	0.153	0.186	0.143
[0.019, 9.9300]	454	0.151	0.199	0.171	0.159	0.190	0.145
[0.019, 8.2600]	452	0.147	0.197	0.174	0.149	0.186	0.194
[0.019, 6.4080]	450	0.150	0.199	0.172	0.132	0.171	0.159
[0.019, 5.8250]	448	0.156	0.201	0.171	0.153	0.187	0.175
[0.019, 4.8730]	446	0.158	0.203	0.173	0.156	0.187	0.186
[0.019, 3.6800]	444	0.161	0.204	0.176	0.153	0.184	0.185
[0.019, 3.4100]	442	0.166	0.207	0.178	0.149	0.183	0.192

在对不同波段组合的响应方面, 神经网络与 SeaBAM 统计算法基本相似, 即以性能较好的 SeaBAM 统计算法所采用的波段作为输入的神经网络的反演精度也较高, 其中以四波段输入的 NN4 的结果最好, 对 NN4 再加入波段 412 nm 对提高反演精度并无帮助. 从表 1 还可看

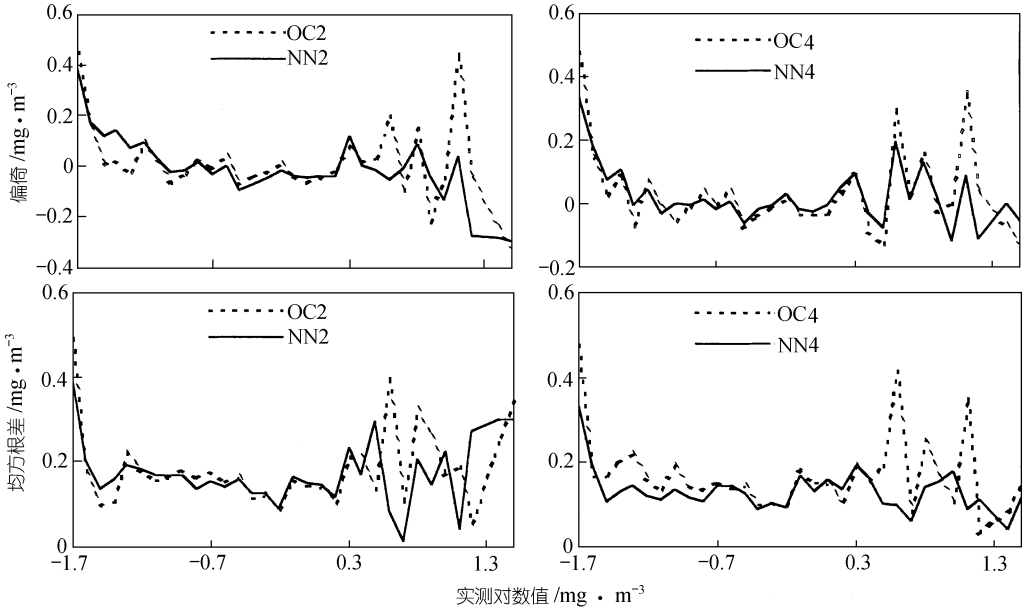


图 2　不同频率分部的叶绿素浓度反演对数值与实测对数值的比较

上图为 Bias，下图为 RMS

到，对神经网络模型而言，直接使用多波段代替最大波段值作为输入可明显提高反演精度；而在 SeaBAM 统计算法中，基于最大波段比值 MBR (Maximum Band Ratio)的算法 OC4 的反演结果优于 OC2，但直接使用多波段比值如(Rrs443 + Rrs490 + Rrs510)/Rrs555 的反演结果较之 OC2 并无明显改善. 这说明虽然 MBR 因其在保留传感器信息方面有潜在的优越性而被 SeaBAM 认为是建立海洋水色经验算法的新手段[2]，但波段转换仍然可能丢失部分有用的信息，在缺乏先验知识的情况下，如何将这些信息包含到回归模型里来是统计方法提高反演精度面临的主要困难.

二类海水水色算法一直是困扰海洋水色遥感应用的主要问题之一. 由于其他物质成分对叶绿素光谱特性的影响，经验统计算法对二类海水叶绿素浓度的反演值常常严重偏高. 表 3 给出 SeaBAM 统计算法与神经网络模型对测试集中 9 个二类海水样本的反演对数值与实测对数值的比较结果,图 3 则给出其散点图.可以看出, SeaBAM 统计算法对多数样本的反演值偏高，

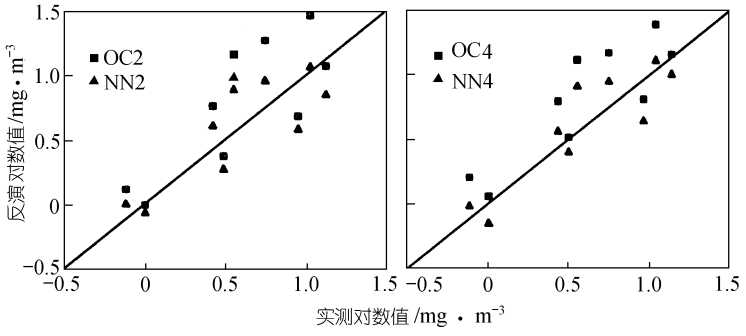


图 3　二类海水叶绿素浓度反演对数值与实测对数值的散点图

表 3 神经网络与统计算法对二类海水的反演结果比较

	OC2	NN2	OC4	NN4
RMS	0.352	0.229	0.311	0.195
Rsq	0.637	0.703	0.747	0.797

反演精度较低; 而神经网络模型的反演结果则比较合理, 反演精度远高于 SeaBAM 统计算法. 由表 4 所示的二类海水反演结果随叶绿素浓度区间的变化情况可以发现, 适当减少叶绿素浓度低值区的训练样本, 可以提高 NN4 的反演精度, 但对提高 NN2 的反演精度作用不大, 与 NN2 和 NN4 对 17 个叶绿素浓度低值区的反演情况正好相反. 这种对于叶绿素浓度低值区与二类海水等特定测试样本集, 不同波段输入的神经网络模型对变化训练集的不同响应, 可能与各波段在反演叶绿素浓度不同区间所起的作用不同以及受训练数据中噪声与不精确度的影响有关.

表 4 二类海水测试样本反演结果随叶绿素浓度区间的变化

叶绿素区间/mg·m^{-3}	训练集样本数	NN2		NN4	
		Rsq	RMS	Rsq	RMS
[0.023, 32.787]	457	0.702	0.230	0.801	0.194
[0.028, 32.787]	451	0.750	0.267	0.814	0.184
[0.034, 32.787]	443	0.759	0.267	0.629	0.269
[0.041, 32.787]	438	0.766	0.265	0.776	0.218
[0.051, 32.787]	424	0.712	0.351	0.671	0.254
[0.064, 32.787]	407	0.714	0.355	0.638	0.267
[0.080, 32.787]	384	0.787	0.260	0.656	0.259
[0.101, 32.787]	349	0.707	0.229	0.897	0.158
[0.128, 32.787]	325	0.783	0.263	0.887	0.181
[0.160, 32.787]	283	0.785	0.263	0.630	0.268
[0.200, 32.787]	210	0.781	0.260	0.865	0.181
[0.259, 32.787]	191	0.764	0.257	0.845	0.191

3 结论

本文的研究结果表明, 神经网络模型是一种切实可行的叶绿素浓度反演算法. 与经验统计算法相比, 神经网络模型最大的优点是可以直接有效地利用各波段辐射所携带的海洋水色信息, 可以在包括二类海水区域等各种环境条件下取得较好的反演结果, 是一种具有普遍意义的叶绿素浓度反演方法.

致谢 感谢 SeaBAM 小组提供资料. 本工作为国家专项(编号: 97-926-05-03)、中国科学院重点(编号: KZ952-J1-404)和广东省自然科学基金(编号: 990311)资助项目.

参 考 文 献

1 Gordon H R, Brown O B, Evans R H, et al. A semianalytic radiance model of ocean color. J Geophys Res, 1988, 93(D9): 10909~10924

2 O'Reilly J E, Maritorena S, Mitchell B G, et al. Ocean color chlorophyll algorithms for SeaWiFS. J Geophys Res, 1998, 103(C11): 24937~24953

3 Verstraete M M, Pinty B, Myneni B. Potential and limitation of extraction on the terrestrial biosphere from satellite remote sensing. Remote Sensing Environment, 1996, 58: 201~214

4 Tassan S. An improved in-water algorithms for the determination from Thematic Mapper data in coastal waters. Int J Remote Sens, 1993, 14: 1221~1229
5 Keiner L E, Yan X H. A neural network model for estimating sea surface chlorophyll and sediments from Thematic Mapper imagery. Remote Sensing Environment, 1998, 66: 153~165
6 Schiller H, Doerffer R. Neural network for estimating of an inverse model-operational derivation of Case Ⅱ water properties from MERIS data. Int J Remote Sens, 1999, 20(9): 1735~1746
7 Buckton D, O' Mongain E. The use of neural networks for the estimation of oceanic constituents based on MERIS instrument. Int J Remote Sens, 1999, 20(9): 1841~1851
8 Hagan M T, Menhaj M. Training feedforward networks with the Marquardt algorithm. IEEE Transactions on Neural Networks, 1994, 5(6): 989~993
9 Demuth H, Beale M. Neural Network Toolbox User's Guide. Natick: The Math Works, 1997. 137~194
10 Krasnopolsky V M, Breaker L C, Gemmill W H. A neural network as a nonlinear transfer function model for retrieving surface wind speeds from the special sensor microwave imager. J Geophys Res, 1995, 100(C6): 11033~11045

(2000-03-21 **收稿**, 2000-05-18 **收修改稿**)