

用 TM 影像进行湖泊水色反演研究的人工神经网络模型

王建平¹,程声通¹,贾海峰¹,王志石²,邓宇华²(1. 清华大学环境科学与工程系,北京 100084;2. 澳门大学科技学院,澳门)

摘要:利用人工神经网络技术进行了湖泊水色遥感的反演研究,在同步实验的基础上构造了包含一个隐含层的 BP 神经网络模型,利用 TM 卫星影像反演悬浮物、COD_{Mn}、溶解氧、总磷、总氮和叶绿素浓度.反演精度较高,相对误差基本在 25% 以下,同时分析了该人工神经网络反演模型的误差来源,改进措施以及应用前景.研究表明,在进行小规模同步监测的基础上,此模型可用于湖泊水质调查、分析和评价.

关键词:人工神经网络;环境遥感;水色遥感;反演;湖泊

中图分类号:P237;X87 文献标识码:A 文章编号:0250-3301(2003)02-04-0073

An Artificial Neural Network Model for Lake Color Inversion Using TM Imagery

Wang Jianping¹, Cheng Shengtong¹, Jia Haifeng¹, Wang Zhishi², Tang U Wa²(1. Department of Environmental Science and Engineering, Tsinghua University, Beijing 100084, China; 2. Faculty of Science and Technology, University of Macao, China)

Abstract: The technology of artificial neural network was used for inverting water quality parameters from TM imagery data in the paper in order to study water quality and eutrophic status of lake. On the basis of satellite synchronous monitoring experiment, a BP neural network model was constructed, in which concentrations of SS, COD_{Mn}, DO, T-N, T-P and chl-a were inverted from Landsat TM data and the accuracy of which was good, the relative error of which could be controlled below 25%. Moreover, the reasons of simulating error, ways of improving model and applications of the model were also analyzed in detail. The results of this research told that based on a small scale of satellite synchronous experiment, the model could be applied successfully in investigation, analysis and estimation of lake water quality.

Keywords: artificial neural network; environmental remote sensing; water color remote sensing; inversion; lake

长期以来,湖泊水质数据采集都依赖于常规监测,耗时费力,且难以及时、迅速地获取数据.随着环境信息技术的不断进步,水色遥感技术由于其大范围、同步性和相对成本较低等优点,在大洋水体、沿海以及湖泊水库的水质监测中都有广泛的应用^[1~3].但是,由于水色遥感卫星如 SeaWiFS、MODIS 的空间分辨率在几百米甚至上千米,相对适合于大洋开阔水体的研究,湖泊、水库遥感反演研究一般只能用空间分辨率较高的陆地遥感卫星进行,为之付出的代价是光谱分辨率降低,这样更加剧了水体中悬浮物、叶绿素、黄色物质识别的难度,一定程度上也限制了水色遥感在陆地水体中的应用.

人工神经网络是 80 年代迅速发展起来的一门非线性科学,它力图模拟人脑的一些基本特性,如自适应性、自组织性和容错性能,已用

于模式识别、系统模拟等领域.针对水色反演的难点及神经网络的特点,本文构造了一个人工神经网络模型——“反向传播”模型,进行了湖泊水色反演研究.

1 数据采集和处理

1.1 同步监测

2001-07-08 在江西鄱阳湖进行了 TM 卫星同步水质监测,采样点位置如图 1 所示.

采样点为 10 个,每个点均测量了悬砂、叶绿素 a、TN、TP、COD_{Mn}、DO、温度、pH 等指标.其中,DO、温度、PH 是在现场实测,其余指标在实验室测定.另外,各点均使用 GPS 接收仪确

基金项目:国家自然科学基金资助项目(49971058);国家杰出青年基金资助项目(49525102)

作者简介:王建平(1977~),男,博士研究生,环境系统工程.

收稿日期:2002-03-28;修订日期:2002-06-04

定经度和纬度.

1.2 卫星数据

卫星数据为同步的 Landsat 7 ETM + 数据,轨道号为 121/40. 图像清晰,无云彩.

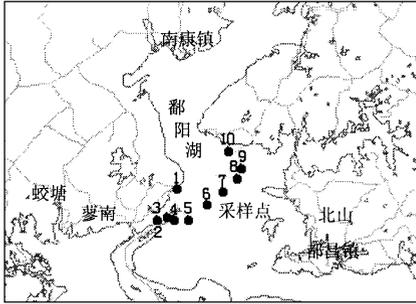


图 1 采样点分布图

Fig. 1 Distribution of sampling points

(1) 遥感数据预处理 在开发模型前需要对遥感数据进行预处理,包括辐射校正、几何校正和大气校正. 一般购买到的影像已经做了辐射校正和几何初校正,用户需要做的工作是几何精校正和大气校正. 本文几何精校正采样地面控制点 (GCPs) 完成,精度优于 0.5 个像素. 着重介绍一下大气校正过程.

(2) 大气校正 PCI 提供了一套针对陆地传感器 (TM/MSS/SPOT) 的大气校正命令 ATCOR0、ATCOR1 和 ATCOR2,流程如图 2.

《原始数据: TM/MSS 或 SPOT1/2》

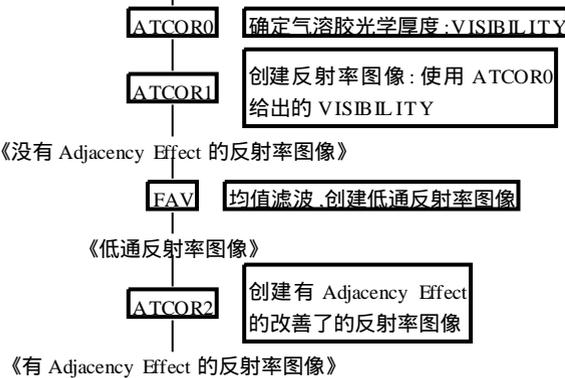


图 2 PCI 提供的大气校正流程图

Fig. 2 Flow chart of atmospheric correction provided by PCI

此方法的理论依据由里克特 (Richter) 给出^[4,5]. 其中心思想是:按照标准大气的分类,计算不同气溶胶类型、太阳天顶角、地面海拔高

度、大气能见度下的大气散射,并存放为一个类似查找表的目录. 在实际应用时按照查找表进行相应的大气校正. 里克特使用的分类依据来自中分辨率大气透射模型 MODTRAN. 他的算法还考虑了地面反射的邻接效应并进行了校正,最后输出地面辐射率.

2 神经网络反演模型

理论上已经证明:具有偏差和至少一个 S 型隐含层加上一个线性输出层的网络,能够逼近任何有理函数^[6]. 水色遥感理论研究表明*,去除大气影响后,遥感数据直接与水体组分及其浓度有关,但是由于水体组分的相互影响,常规反演方法如建立波段与组分之间的函数关系很难奏效,神经网络正适合模拟这种错综复杂的关系,因此理论上,利用神经网络进行水色遥感反演研究是可行的.

2.1 模型结构

图 3 表示了本文采用的人工神经网络结构的概念图. 每个输入节点代表一个 TM 可见光波段,输入层的值分发到隐藏层的每个节点,并在此进行运算. 隐藏层的输出值再次成为输出层的输入,并再次进行计算,输出层的输出将是最终感兴趣的参数值.

具体地说,本文采用的人工神经网络是 2 层网络.

$p_i (i = 1 \dots r), r = 5$, TM 影像数据 5 个波段输入 (band1 ~ band5); $iw_{1,1}$ 为输入与第一层神经元之间的权重, $lw_{2,1}$ 为第一层输出和第二层神经元之间的权重; b^1, b^2 分别是第一层和第二层神经元的偏差; f^1, f^2 为转移函数. BP 网络中 f^1 一般采样 S 型转移函数,本模型 f^1 采样双曲正切 S 型转移函数, f^2 一般采样线性转移函数,理论上已经证明:具有偏差和至少一个 S 型隐含层加上一个线性输出层的网络,能够逼近任何有理函数^[6]; a^1, a^2 分别是第一层和第二层神经元的输出,由图 3 所示:

$$a^1 = f^1(IW_{1,1} \times p + b^1),$$

* 况昶. 水色遥感理论模型的识别与求解途径探讨. 清华大学博士学位论文. 1999. 6: 15 ~ 86.

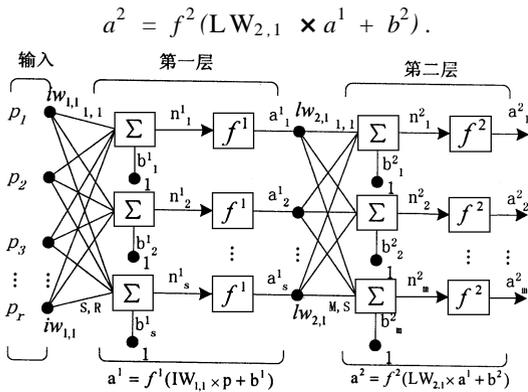


图 3 人工神经网络结构

Fig. 3 The structure of artificial neural network

第二层神经元的数目 m 由神经网络模型目标数目决定 本文希望通过神经网络模型模拟悬浮物、COD_{mn}、DO、TP、TN 和叶绿素浓

度,故 $m = 6$;

第一层神经元数目 s 一般在网络训练过程中确定,根据网络训练结果 $s = 10$.

2.2 训练结果及分析

2.2.1 模型训练和校验数据

需要指出的是,为了使神经网络训练效率更高、精度更好,一般输入数据要进行归一化^[7],表 1 为神经网络的输入和目标.

网络训练过程中,逐渐调整隐含层神经元数目,即 s 的大小,通过校验样本的校验精度、训练时间和迭代次数对比,最终确定 $s = 10$,即隐含层有 10 个神经元.

2.2.2 模型校验

人工神经网络模型经过训练,代入校验样本数据,校验结果如表 2.

表 1 神经网络模型原始数据/ mg L^{-1}

Table 1 Original data of the artificial neural network model

项目		训练样本							校验样本		
		1	2	3	4	5	6	7	8	9	10
输入	band-1	1.212	-0.736	-0.885	-0.587	-1.290	-0.141	1.672	1.131	-0.317	-0.060
	band-2	1.153	-0.628	-0.824	-0.677	-1.333	-0.176	1.637	1.235	-0.357	-0.029
	band-3	1.082	-0.313	-0.736	-0.684	-1.082	-0.456	1.825	1.258	-0.574	-0.319
	band-4	0.476	1.606	-0.714	-0.654	-1.071	-0.595	1.546	0.714	-0.654	-0.654
	band-5	-0.887	2.070	0.222	0.222	-0.795	-0.979	0.961	-0.979	0.499	-0.333
目标	SS	2.05	3.65	15.65	19.65	20.65	25.15	35	2.4	15.15	21.55
	COD _{mn}	1.7	2.05	2.3	1.85	2	2	1.95	1.7	1.8	1.9
	DO	5.3	4.76	5.68	5.15	5.8	6.5	5.95	4.78	6.2	6.38
	T-P	0.055	0.03	0.04	0.035	0.03	0.03	0.055	0.05	0.04	0.025
	T-N	0.855	0.605	0.915	0.715	0.855	0.765	0.655	0.79	0.79	0.875
chl-a/ $\mu\text{g L}^{-1}$	5	5	5	4	3	5	4	5	6	5	

从表 2 可以看出,除第二个验证数据的 SS 和 chl-a 的反演误差超过 30%,其它指标的反演误差基本能控制在 25%以下,结果较满意.

2.2.3 结果分析

结果表明,人工神经网络反演模型能较好地通过遥感影像实现湖泊水色反演.

(1) 反演效果较好 如表 2 所示,除第 2 个验证数据的 SS 和 chl-a 的反演误差超过 30%,其它指标的反演误差基本能控制在 25%以下,如果在数据采集、处理和分析中采取相应措施(见后面分析),反演精度会进一步提高.

(2) 模拟复杂关系能力强 第二类水体(包

表 2 神经网络模型反演验证结果

Table 2 The results of inversion and validation of the model

水质参数	SS	COD _{Mn}	DO	TP	TN	chl-a
反演计	2.09	1.66	5.23	5.68	8.43	4.91
算结果	20.29	1.64	4.98	3.59	6.21	4.11
	26.78	2.01	6.52	3.14	6.69	4.88
同步监	2.40	1.70	4.78	5.00	7.90	5.00
测结果	15.15	1.80	6.20	4.00	7.90	6.00
	21.55	1.90	6.38	2.50	8.75	5.00
反演误差	-13.01	-2.15	9.33	13.52	6.69	-1.84
	33.95	-8.81	-19.67	-10.26	-21.44	-31.46
	24.29	5.79	2.26	25.63	-23.50	-2.35

括沿海、河口、湖泊和水库)中悬浮物、叶绿素、黄色物质等相互影响,导致其对遥感卫星各波段的响应不灵敏*,这就要求进行水色遥感时要有较高的光谱分辨率(一般 $<20\text{nm}$),但是光谱分辨率的提高必然要损失空间分辨率,而湖泊研究又需要较高的空间分辨率,这种矛盾的解决导致遥感影像中各波段数据难以突出某一组分的影响,所以常规统计反演方法来提取水体特定组分,很难达到实用的程度.而人工神经网络通过神经节点权重和偏差的调整,能很好地实现从输入到输出状态空间的非线性映射.结果表明,模拟效果不错.人工神经网络这种模拟复杂输入输出关系的能力理论上已经证明^[6].

(3)实现了多水质参数的同时反演.通过人工神经网络训练,利用 TM 影像 5 波段数据即可对湖泊各点悬浮物、 COD_{Mn} 、DO、TP、TN 和叶绿素浓度进行反演.

(4)十分强的自适应、自学习和容错功能.人工神经网络模型对训练样本数要求不高,通过很强的自学习能力、自适应能力和容错能力,有效地利用训练数据进行最佳模拟,这对一个模拟模型是非常重要的,可以大大降低常规监测强度,提高工作效率.

造成个别参数误差较高的原因主要有这么几方面:同步监测带来的试验误差,如浅水区船舶搅动、实验误差等;采样点较少,水质参数在其浓度范围内采样数不均匀,这些都可能带来神经网络训练误差;遥感数据处理误差,尤其是大气校正误差;模型本身具有的结构误差.

研究表明,第 1 项误差通过适当措施可以避免或减小,如采样在船舶停稳一段时间后进行、增加平行采样数等措施;第 2 项误差可采取增加采样点数、合理布置采样点等措施予以消除;第 3 项误差是较难解决的,如果可能的话,在卫星同步监测同时,进行一定光学试验,这样可有效地降低大气校正误差;第 4 项误差是不可避免的,但可通过不同网络结构、不同神经元数目对比分析,误差可以适当降低.

模型可应用在以下几方面:人工神经网络

经过训练,整个 TM 影像数据就可通过神经网络计算.当整个 TM 图像作为输入时,神经网络输出图像的每个象元点代表的将是水体考察因素指标.因为模型同时对湖泊的悬浮物、 COD_{Mn} 、DO、TP、TN 和叶绿素浓度进行反演,这样就可了解整个湖泊的富营养状况.有助于定期了解湖泊的水质状况,此外通过一个时间序列的遥感影像数据可进行湖泊富营养演化过程模拟和了解.

3 结论

理论研究和实验分析均表明,人工神经网络技术用于水色遥感反演研究是可行的,而且对二类水体中各组分相互影响的复杂关系有较强的模拟能力.

在同步实验的基础上构造了一个 2 层的 BP 神经网络模型,该模型可利用 TM 卫星影像反演悬浮物、 COD_{Mn} 、DO、TP、TN 和叶绿素浓度.模型反演精度较高,相对误差基本在 25% 以下,随后分析了该人工神经网络反演模型的主要误差来源,讨论了提高模型反演精度的措施,同时探讨了该水色反演人工神经网络模型的应用前景等.研究表明,在进行小规模的同时监测的基础上,此模型可用于湖泊水质调查、分析和评价.

参考文献:

- 1 陈楚群,施平,毛庆文.应用 TM 数据估算沿岸海水表层叶绿素浓度模型研究.环境遥感,1996,11(3):168~176.
- 2 Claudia Giardino, Monica Pepe. Detecting chlorophyll, Secchi disk depth and surface temperature in a sub-alpine lake using Landsat imagery. The Science of the Total Environment, 2001, 268: 19~29.
- 3 赵碧云,贺彬等.滇池水体中总悬浮物含量的遥感定量模型.环境科学与技术,2000,3(94):16~18.
- 4 Richter R A. Fast atmospheric correction algorithm applied to Landsat TM Images. Int. J. Remote Sens., 1990, 11: 159~166.
- 5 Richter R A. Spatially adaptive fast atmospheric correction algorithm. Int. J. Remote Sens., 1996, 17: 1201~1214.
- 6 丛爽.面向 MATLAB 工具箱的神经网络理论与应用.合肥:中国科学技术大学出版社,1998. 11:45~60.
- 7 詹海刚,施平等.利用神经网络反演海水叶绿素浓度.科学通报,2000,45(17):1779~1884.