

基于神经网络模型的海水硝酸盐测量方法研究

侯耀斌^{1,2}, 冯巍巍^{1,2*}, 蔡宗岐¹, 王焕卿¹, 刘增东³

1. 中国科学院海岸带环境过程与生态修复重点实验室(烟台海岸带研究所), 山东 烟台 264003

2. 中国科学院大学, 北京 100049

3. 山东省烟台生态环境监测中心, 山东 烟台 264000

摘要 硝酸盐浓度是海洋生态系统研究的重要指标。光学法硝酸盐原位传感器具有测量速度快、无需化学试剂的优点,在长时间连续监测方面,优于镉柱还原法等实验室化学方法。在计算模型方面目前国内外硝酸盐光学传感器多使用偏最小二乘法(PLS)对紫外吸收光谱进行光谱分析建模,模型优化难度较大且泛化能力较低,而神经网络模型理论上能够以任意精度逼近任何非线性连续函数,样本充足的情况下精度较高,泛化能力强。利用自主研发的海水硝酸盐原位传感器,研究了硝酸盐浓度范围为 $30\sim 750\ \mu\text{g}\cdot\text{L}^{-1}$ 的人工海水的紫外吸收光谱,建立神经网络模型,定量计算水中的硝酸盐浓度。对比研究了单隐藏层和双隐藏层神经网络模型对硝酸盐浓度测量的性能,确定采用双隐藏层结构,模型的输入层为 $200\sim 275\ \text{nm}$ 波段的吸收光谱数据,输出层为硝酸盐浓度测量值,使用sigmoid函数作为激活函数。采用梯度下降法更新每一层神经网络的权值参数,学习率为0.26,迭代55000次进行网络训练。通过8组随机验证数据进行模型盲测验证,得到的双隐层神经网络模型的硝酸盐浓度预测值和实际值的线性相关度较高($R^2=0.997$),均方根误差为10.864,平均绝对误差为 $8.442\ \mu\text{g}\cdot\text{L}^{-1}$,平均相对误差为2.8%,模型的精度较高,比单隐层神经网络模型的平均相对误差降低了4.92%,而利用同样的光谱数据建立的偏最小二乘算法的均方根误差为11.470,平均相对误差为4.58%,说明神经网络模型在一定条件下优于PLS模型的精度,验证了神经网络模型应用于紫外吸收光谱法硝酸盐浓度测量的优越性。搭载自然资源部“环监01”监测船环渤海航次进行了实际应用测试,在11个站位与实验室方法进行了比对测试,两种方法测量结果基本一致,进一步证明了该方法的可靠性和实用性。

关键词 紫外吸收光谱;海水;硝酸盐;神经网络;光谱分析;原位传感器

中图分类号: X853 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)10-3211-06

引言

硝酸盐浓度是衡量水质状况的重要指标,水中的硝酸盐含量过高会导致水体富营养化,严重时会引起赤潮,破坏海洋生态系统的平衡^[1]。而近年来由于人类的活动,越来越多的硝酸盐被排放到海洋中,使赤潮等灾害的发生愈加频繁,监测海水中的硝酸盐浓度有利于海洋灾害预测及环境保护。

目前海水硝酸盐浓度测量的方法多用化学方法,如镉柱还原法^[2],该方法虽然精度较高,但是操作复杂,无法实现快速测量,并且测量过程中产生的化学物质容易造成二次污

染,不能适应硝酸盐浓度连续长时间监测的目的。而基于紫外吸收光谱特性制成的原位传感器具有不需水样预处理、测量耗时短、无需化学试剂等优点,能够实现长时间连续监测,具有独特的优势^[3]。目前,国内外很多学者也对硝酸盐的光学测量方法进行了研究,Jonhson等研发了专门进行海洋硝酸盐浓度测量的ISUS原位传感器^[4],硝酸盐浓度测量精度可达 $7\ \mu\text{g}\cdot\text{L}^{-1}$,可部署在浮标、拖船、水下探测器等设备上^[5-6],并且还针对温度、压力等对测量的影响进行了校准^[7],而基于ISUS传感器优化的Deep SUNA传感器性能更加稳定的同时,能适应2000 m深海工作。国内杨鹏程等进行了原位硝酸盐传感器的研制^[8],李丹等完成了硝酸盐在

收稿日期:2019-09-06,修订日期:2020-01-10

基金项目:国家重点研发计划项目(2016YFC1400601),中国科学院关键技术人才项目(科发人函字[2016]65号),烟台市重点研发计划项目(2016ZH071),中国科学院STS项目(KFJ-STZ-ZDTP-055)资助

作者简介:侯耀斌,1995年生,中国科学院烟台海岸带研究所硕士研究生 e-mail: ybhou@yic.ac.cn

* 通讯联系人 e-mail: wwfeng@yic.ac.cn

线监测系统,但没有很好地处理盐度、温度等干扰因素,使用的模型也较为单一,相对国外还不是很完善。

硝酸盐原位传感器多是基于偏最小二乘法(partial least squares, PLS)进行测量模型的建模,该方法在应对复杂多变的水体环境时,容易受到未知干扰物的影响,模型泛化能力不高。而神经网络模型具有自学习和自适应特点,容错能力和泛化能力较高^[9],在一定条件下优于偏最小二乘法。本文在实验室环境下,利用自主研发的硝酸盐传感器,基于配置的人工海水的紫外吸收光谱,建立了基于神经网络的硝酸盐浓度反演模型,具有较高的精度,能够应用于原位传感器中进行海水硝酸盐浓度测量。

1 实验部分

1.1 原位硝酸盐传感器

基于紫外吸收光谱法,课题组研发了硝酸盐原位传感系统并进行水体吸光度的测量。该传感器主要由光谱采集模块、脉冲氙灯光源、数据处理模块等组成,测量光程为 100 mm。光谱采集模块光谱分辨率为 2 nm,光谱范围为 180~750 nm。传感器可与配套的水质监测软件结合使用,可以直接测得海水中的硝酸盐浓度,实时显示所测水体的光谱图并对光谱数据进行保存。

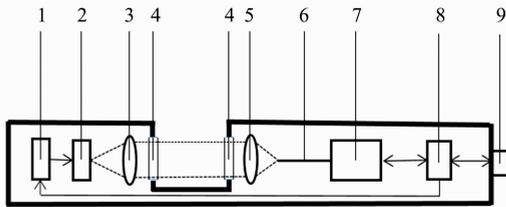


图 1 硝酸盐传感器结构示意图

1: 光源控制板; 2: 光源; 3: 透镜 1; 4: 光窗; 5: 透镜 2; 6: 光纤; 7: 微型光谱仪; 8: 信号控制与处理器; 9: 接口

Fig. 1 Schematic diagram of the nitrate sensor

1: Light source control panel; 2: Light source; 3: Lens 1; 4: Light windows; 5: Lens 2; 6: Optical fiber; 7: Microspectrometer; 8: Control and signal processing module; 9: Interface

1.2 人工海水配置

大洋海水中硝酸根离子(NO_3^- -N)浓度一般低于 $600 \mu\text{g} \cdot \text{L}^{-1}$,部分海域(如我国长江入海口)硝酸盐浓度高,能够达到 $900 \mu\text{g} \cdot \text{L}^{-1}$ ^[10]。本实验利用配置的人工海水^[11],添加硝酸钾固体,配置 23 组具有硝酸盐浓度梯度样品,其硝酸盐浓度范围为 $30 \sim 750 \mu\text{g} \cdot \text{L}^{-1}$ 。根据硝酸盐浓度,从中抽取 15 组作为训练集进行神经网络建模,剩余 8 组样本作为测试集来检验模型的精度。

1.3 BP 神经网络原理

BP 神经网络(back propagation neural network, BPNN),即基于误差逆向传播算法的多层前馈神经网络,在理论上可以以任意精度逼近任意非线性连续函数,对于非线性函数的拟合能力强,具有自学习和自适应的特点。神经网络模型通过训练,抗干扰能力较强,能够适应复杂海水环境

中硝酸盐浓度测量的工作。BP 神经网络模型分为三层:输入层、隐藏层、输出层^[12],如图 2 所示。

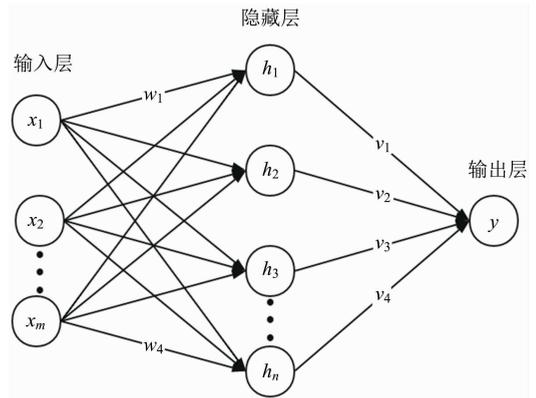


图 2 BP 神经网络结构示意图

Fig. 2 Structure chart of BPNN

神经网络模型建立的过程分为前向传递过程和后向传播过程。

信号的前向传递过程:将样本数据由输入层输入,信号经各层神经元处理,逐层向前传递至输出层输出。

$$y = F(\mathbf{W}x + \mathbf{b}) \quad (1)$$

其中: y 是模型预测值; x 为输入的吸收光谱数据; \mathbf{W} 为权重矩阵, \mathbf{b} 为阈值矩阵, $F(\cdot)$ 为激活函数,使用 Sigmoid 函数作为激活函数。

后向传播过程:根据输出层输出的预测结果和样本的实际值,求得模型的预测误差,将误差逆向传播来进行权值和阈值的调整。权值的更新通过梯度下降法来进行。

$$E = y - y_{\text{pre}} \quad (2)$$

$$\Delta\omega = -\eta \frac{\partial E}{\partial \omega} \quad (3)$$

$$\omega \leftarrow \omega + \Delta\omega \quad (4)$$

其中: E 为误差, y 为期望输出; y_{pre} 为实际输出; $\Delta\omega$ 为权值的变化值, η 为学习率。

1.4 模型的评价方法

选择相关系数(R^2)、预测均方根误差(root mean square error of prediction, RMSEP)、平均绝对误差(mean absolute error, MAE)、平均相对误差(mean relative error, MRE)作为评价模型性能的指标, RMSEP, MAE 和 MRE 值越低, R^2 值越接近 1,表示模型的预测精度越高,稳定性越好。

$$\text{RMSEP} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_{i, \text{pre}})^2} \quad (5)$$

2 神经网络模型建立

2.1 数据归一化处理

样本数据在输入神经网络模型之前,对原始数据进行归一化处理有利于避免一些不必要的数值问题,同时使神经网络快速收敛,保证模型预测值量^[13]。本实验使用离差标准化(min-max normalization)对原始数据进行线性变换,将原始数据值映射到 $[0-1]$ 之间。

2.2 神经网络结构的确定

实验神经网络模型的输入层数据为 200~275 nm 波段的吸收光谱数据,共 34 个波段,故输入层节点数为 34。硝酸盐浓度作为模型输出的唯一目标,故输出层节点数为 1。

隐藏层节点数由以下经验公式确定,取尽量多的隐藏层节点数,确定隐藏层节点数为 16 个。

$$h = \sqrt{m+n} + a \quad (6)$$

其中: h 为隐含层节点数; m 为输入层节点数; n 为输出层节点数; a 为 1~10 之间的调节常数。

通常采用单隐层神经网络便能得到精度较为不错的模型,为了进一步提升模型的精度,可以增加隐藏层的数量,但同时也增加了模型的复杂度,使其训练开销变大,因此隐藏层个数的确定需要综合实际需求来进行选择。本实验分别用单隐藏层神经网络和双隐藏层神经网络建模并进行对比,择优选取含有 2 个隐藏层的神经网络作为最终模型。

2.3 训练参数的设置

神经网络模型建立的过程中需要涉及到很多参数的设置,但由于神经网络模型的不可解释性,很多参数的选择多是基于经验的。本文针对神经网络模型,采用试凑法来进行模型参数的设置,最终模型训练的学习率设置为 0.26,精度要求为 1×10^{-6} ,最大迭代次数为 2×10^5 。随着训练次数的增加,模型训练均方误差(mean square error, MSE)的变化如图 3 所示,从图中可以看到,随着训练次数的增加,模型的误差降低,精度提高,训练 55 000 次后收敛,因此模型训练次数设置为 55 000 次。

3 结果与讨论

3.1 光谱及分析

传感器内置光谱预处理程序,能够对测得的光谱进行平滑去噪处理,根据测得光谱信息计算得到实验样品的吸收光谱曲线(如图 4 所示),溶液浓度较小时,吸光度较小,噪声影响严重导致 200~210 nm 处吸光度分布杂乱。可以看出样品溶液在紫外波段具有较强的吸光性,吸光波段主要在 200~240 nm 处,于 220 nm 波段处有一强吸收峰,并且吸光强度随着溶液中硝酸盐浓度的增加而不断上升,具有一定的梯度,证明了硝酸盐在紫外波段具有较强的吸光性,且其吸光

度与浓度成正比关系。可以根据硝酸盐的这一吸光特性建立模型来反演海水中的硝酸盐浓度。

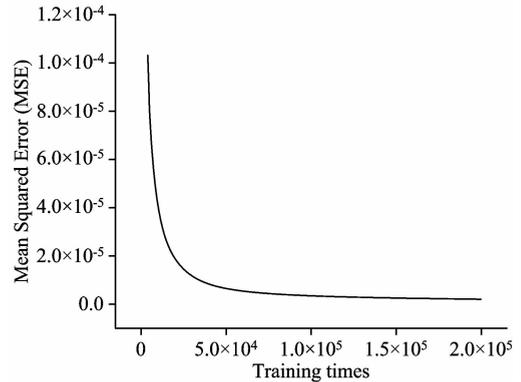


图 3 网络训练过程

Fig. 3 Network training process

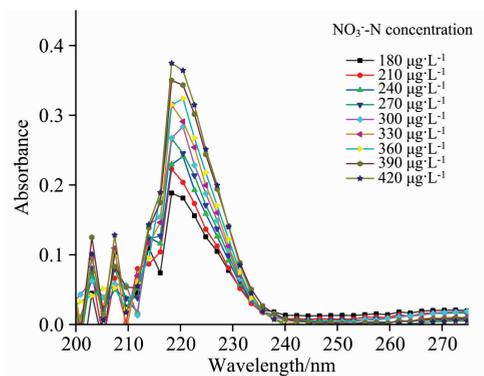


图 4 不同硝酸盐浓度样品的紫外吸收光谱图

Fig. 4 UV absorption spectra of different nitrate concentrations

3.2 预测结果及精度评价

使用建立的两层结构的神经网络模型和偏最小二乘模型对配置的人工海水测试样本进行硝酸盐浓度预测,测试样本的硝酸盐浓度实际值和模型的预测结果如表 1 所示,可以看出硝酸盐浓度预测值与实际值相近,说明三种模型都能成功测量人工海水溶液中的硝酸盐浓度。

表 1 模型预测结果

Table 1 Model prediction results

Serial number	Actual Value / ($\mu\text{g} \cdot \text{L}^{-1}$)	1 Hidden Layer BPNN Predictive Value / ($\mu\text{g} \cdot \text{L}^{-1}$)	2 Hidden Layer BPNN Predictive Value / ($\mu\text{g} \cdot \text{L}^{-1}$)	PLS Predictive Value / ($\mu\text{g} \cdot \text{L}^{-1}$)
1	90.000	104.788	93.374	108.898
2	180.000	192.742	169.137	168.596
3	240.000	252.648	229.105	241.797
4	330.000	371.021	332.520	332.716
5	360.000	386.348	353.995	365.818
6	420.000	420.715	415.918	427.090
7	630.000	649.778	654.724	639.038
8	660.000	594.306	654.930	679.720

图 5 为三种模型对不同硝酸盐浓度的测试样本溶液的预测结果的相对误差图,从图中可以看出当溶液中的硝酸盐浓度较低为 $90 \mu\text{g} \cdot \text{L}^{-1}$ 时,单隐层神经网络模型和偏最小二乘模型预测结果的相对误差较大,都超过了 15%,而双隐层神经网络模型预测结果的相对误差较小,说明双隐层神经网络模型在测量低硝酸盐溶液浓度时精度更高。从 8 个样品的预测结果的相对误差图来看,双隐层神经网络模型波动范围较小,相对误差值较低,模型最为稳定。

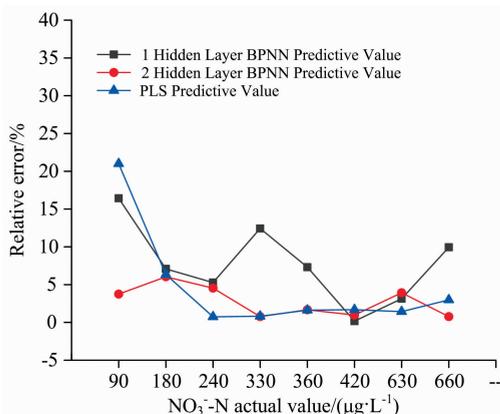


图 5 模型预测结果相对误差

Fig. 5 Relative error of model prediction results

分别用神经网络模型和偏最小二乘法模型对测试样本进行预测后得到的硝酸盐浓度预测值和实际值进行线性拟合分析,得到线性拟合结果如图 6 所示,可以看出三种模型对人工海水中的硝酸盐浓度预测值和其实际值具有良好的线性相关性,图 6(a)中单隐藏层神经网络模型相关系数 r 为 0.975,图 6(b)双隐层神经网络模型和图 6(c)偏最小二乘法模型相关系数 r 都为 0.997,表明两种模型都能较好地预测海水中的硝酸盐浓度。

分别用神经网络模型和偏最小二乘法模型对测试样本进行预测后各样本评价指标如表 2 所示,根据表中所列参数评价模型的精度和稳定性,从表中可以看出单隐藏层神经网络模型的稳定性和精度低于偏最小二乘模型,而双隐藏层神经网络模型的平均绝对误差为 $8.442 \mu\text{g} \cdot \text{L}^{-1}$,平均相对误差为 2.8%,预测均方根误差为 10.864,稳定性和精度都高于偏最小二乘模型,说明增加隐藏层的数目能大大提升 BP 神经网络模型的精度,并且证明神经网络模型在一定条件下优于偏最小二乘模型,模型能够较为准确地预测海水中的硝酸盐浓度。

3.3 实际应用与对比测试

将该模型应用于硝酸盐原位监测工程样机,2018 年 8 月在自然资源部“环监 01”监测船环渤海航次进行了实际应用测试,在 11 个站位与实验室方法(镉柱还原法)进行了对比测试,实验室海水水样硝酸盐浓度数据由烟台市环境监测中心站提供。11 处测试站位分布如图 7 所示,主要分布在烟台近海水域。图 8 为相应采样点实验室方法测量结果和基于神经网络模型的硝酸盐传感器测量结果,从图中可以看出,两种方法的硝酸盐浓度测量结果虽然有所偏差,但变化趋势基

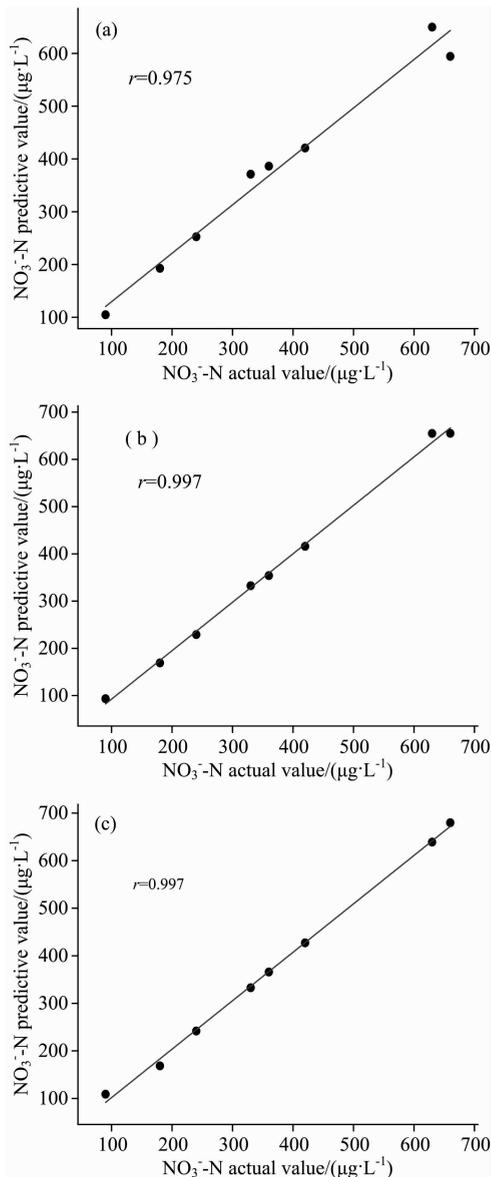


图 6 不同模型预测值和实际值的线性相关性

(a): 单隐层神经网络预测值 vs 实际值; (b): 双隐层神经网络预测值 vs 实际值; (c): PLS 预测值 vs 实际值

Fig. 6 Correlation between predicted and actual values in different models

(a): 1 hidden layer BPNN Predictive value vs Actual value;
(b): 2 hidden layer BPNN Predictive value vs Actual value;
(c): PLS Predictive value vs Actual value

表 2 三种方法预测硝酸盐浓度的评价
Table 2 Evaluation of the nitrate concentration by the three methods

Model	r	MAE/ ($\mu\text{g} \cdot \text{L}^{-1}$)	MRE/%	RMSEP
1h-BP	0.975	24.217	7.7	30.873
2h-BP	0.997	8.442	2.8	10.864
PLS	0.997	9.560	4.6	11.470

本一致,证明了基于该神经网络模型的硝酸盐传感器的可靠性。光学法测量硝酸盐浓度的与实验室化学方法采用的原理不同,化学方法的氧化效率与水质状况有很大的关系,这是对比测试中两者测量结果产生偏差的原因,但是由于光学传感器测量速度快,且无二次污染,在现场快速监测及长时间连续监测的情况下相对实验室化学方法具有明显的优势。

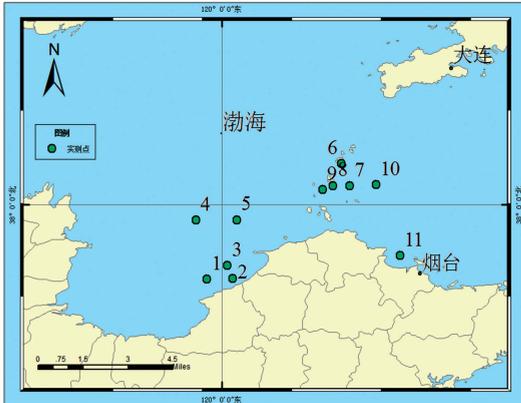


图 7 航次采样站点
Fig. 7 Sampling point

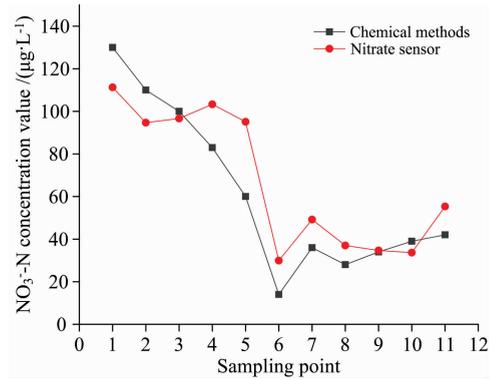


图 8 原位传感器与化学方法对比
Fig. 8 Nitrate sensor vs chemical method

4 结 论

通过建立的双隐藏层神经网络模型代替偏最小二乘模型,应用于硝酸盐原位监测工程样机,实现对人工海水中的硝酸盐浓度快速定量分析。经过实验室精度评价,建立的双

隐藏层神经网络模型对浓度为 $30 \sim 750 \mu\text{g} \cdot \text{L}^{-1}$ 硝酸盐浓度预测的平均绝对误差为 $8.442 \mu\text{g} \cdot \text{L}^{-1}$, 平均绝对误差为 2.8% , 预测均方根误差为 10.864 , 预测结果和实际浓度值之间的线性相关系数 $R^2 = 0.997$ 。在实际海水现场测试应用中,与实验室方法(国标方法)进行了对比测试,结果相似,表明神经网络模型应用于原位硝酸盐传感器能够实现硝酸盐浓度快速测量,可为海洋环境监测提供大量的硝酸盐含量监测数据。另外,实际海水中我们考虑了海水浊度和盐度梯度对测量的影响,在实际工程应用中,并不是将神经网络简单用于吸光度数据的直接处理,而是对吸收光谱进行复杂的预处理,本文讨论的重点在于对神经网络模型建立及评价。如何与温盐深仪(CTD)联用,进一步开展实时盐度校准,是后续工程应用需要解决的问题。

References

- [1] Meyer D, Prien R, et al. *Frontiers in Marine Science*, 2018, 5: 431.
- [2] Singh P, Singh M K, Beg Y R, et al. *Talanta*, 2019, 191(3): 364.
- [3] FENG Wei-wei, LI Ling-wei, LI Wei-ran, et al(冯巍巍, 李玲伟, 李未然, 等). *Acta Photonica Sinica(光子学报)*, 2012, 41(8): 883.
- [4] Johnson K S, Coletti L J. *Deep Sea Research Part I: Oceanographic Research Papers*, 2002, 49(7): 1291.
- [5] Sakamoto C M, Johnson K S, Coletti L J. *Limnology and Oceanography-Methods*, 2009, 7(1): 132.
- [6] Johnson K S, Coletti L J, et al. *Journal of Atmospheric and Oceanic Technology*, 2013, 30(8): 1854.
- [7] Sakamoto C M, Johnson K S, Coletti L J, et al. *Limnology and Oceanography: Methods*, 2017, 15(10): 897.
- [8] CHENG Chang-kuo, SONG Jia-ju, DU Jun-lan, et al(程长阔, 宋家驹, 杜军兰, 等). *Transducer and Microsystem Technologies(传感器与微系统)*, 2017, 36(12): 75.
- [9] CHEN Shan-lin, HUANG Chun-hui(陈珊琳, 黄春晖). *Chinese Journal of Quantum Electronics(量子电子学报)*, 2017, 34(4): 465.
- [10] ZHANG Zheng-bin(张正斌). *Marine Chemistry(海洋化学)*. Qingdao: China Ocean University Press(青岛: 中国海洋大学出版社), 2004. 121.
- [11] Shan L, Wang Y X, Li J L, et al. *Surface & Coatings Technology*, 2013, 226: 40.
- [12] WANG Zhi-fang, WANG Shu-tao, WANG Gui-chuan(王志芳, 王书涛, 王贵川). *Acta Photonica Sinica(光子学报)*, 2019, 48(4): 0412004.
- [13] Ren C, An N, Wang J Z, et al. *Knowledge-Based Systems*, 2014, 56(1): 226.

Nitrate Measurement in the Ocean Based on Neural Network Model

HOU Yao-bin^{1,2}, FENG Wei-wei^{1,2*}, CAI Zong-qi¹, WANG Huan-qing¹, LIU Zeng-dong³

1. CAS Key Laboratory of Coastal Environmental Processes and Ecological Remediation, Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai 264003, China
2. University of Chinese Academy of Sciences, Beijing 100049, China
3. Shandong Provincial Yantai Eco-environmental Monitoring Center, Yantai 264000, China

Abstract Nitrate concentration is an important indicator for the marine ecosystem. Compared with laboratory chemical methods such as Cadmium-Reduction method, in-situ nitrate optical sensor is much faster and reagent-free in a long time and continuous monitoring. Partial Least Squares (PLS) method is often used in ultraviolet absorption spectrum modeling, which is difficult to optimize and has low generalization ability. The neural network can compel any non-linear function by any precision, which has high generalization ability in the modeling. A neural network model is established in the in-situ nitrate sensor to measure the nitrate concentration in seawater in which the nitrate concentration range is $30 \sim 750 \mu\text{g} \cdot \text{L}^{-1}$. Double-hidden layer neural network model is determined to adopt by contrasting performance of single-hidden layer and double-hidden layer to measure nitrate concentration, the input layer is absorption spectrum from 200 to 275 nm, the output layer is nitrate concentration, and sigmoid function is used as the activation function. Gradient descent method is used to update weighting parameters for the neural network of each layer, after 55 000 times iteration, network training is conducted based on the learning rate of 0.26. After validation for the blind test of the model through 8-group randomized validation data, the nitrate concentration using double-hidden layer neural network model is higher in linear correlation to its actual concentration ($R^2 = 0.997$) in which the Root Mean Squared Error is 10.864, average absolute error is $8.442 \mu\text{g} \cdot \text{L}^{-1}$, average the relative error is 2.8%. Compared with single-hidden layer neural network model, the double-hidden layer neural network model has higher accuracy in which the average relative error is reduced by 4.92%, the Root Mean Squared Error of PLS is 4.58% using the same spectral data, while the mean relative error is 11.470. The result shows that the neural network model is much better than the Partial Least Squares model under certain conditions. It verifies the superiority of the neural network model applied to the nitrate concentration measurement by ultraviolet absorption spectrometry. The application test was carried out on the "Environmental Monitoring 01" monitoring vessel of the Ministry of Natural Resources, the measurement results are basically identical with the laboratory method in 11 stations, which is further proved from the reliability and practicality.

Keywords Ultraviolet; Absorption spectroscopy; Seawater; Nitrate; Neural network; Spectral analysis

(Received Sep. 6, 2019; accepted Jan. 10, 2020)

* Corresponding author