

多光谱结合神经网络预测模型的总磷检测研究

毛天宇¹, 蔡元学², 卞海溢^{3*}

¹交通运输部天津水运工程科学研究所, 天津

²天津科技大学, 天津

³淮阴工学院, 江苏省湖泊环境遥感技术工程实验室, 江苏 淮安

收稿日期: 2022年6月17日; 录用日期: 2022年8月30日; 发布日期: 2022年9月6日

摘要

多光谱遥感应用于水质参数检测具有良好的应用前景, 但是目前基于多光谱的水质参数反演模型大多使用经验统计模型, 通用性差。针对上述问题, 以鄱阳湖为例, 共采集了41组Sentinel-2A卫星的多光谱遥感数据, 其中35组光谱数据作为训练集, 建立了基于神经网络算法的水质总磷浓度预测模型, 该模型内部验证和测试的相关系数0.8以上。剩余6组光谱数据作为外部测试集, 预测值与真实值的相关系数为0.88, 均方根误差为0.048。实验结果表明基于神经网络算法的多光谱总磷浓度检测模型是一种可行的水质遥感检测技术。

关键词

多光谱, 神经网络, 总磷, 水质检测, 哨兵2A

Research of Total Phosphorous Detection Combined Multi-Spectra Spectroscopy with Neural Network Algorithm

Tianyu Mao¹, Yuanxue Cai², Haiyi Bian^{3*}

¹Research Institute for Water Transport Engineering, Ministry of Transport, Tianjin

²Tianjin University of Science and Technology, Tianjin

³Jiangsu Engineering Laboratory for Lake Environment Remote Sensing Technologies, Huaiyin Institute of Technology, Huai'an Jiangsu

Received: Jun. 17th, 2022; accepted: Aug. 30th, 2022; published: Sep. 6th, 2022

*通讯作者。

Abstract

The application of multispectral remote sensing to water quality parameter detection has a good application prospect, but at present, most inversion models based on multispectral water quality parameters use an empirical statistical model, which has poor universality. To solve the above problems, 41 sentinel-2A satellite multi-spectral remote sensing data were collected from Poyang Lake, 35 of which were used as training sets. A prediction model of total phosphorus concentration in water quality based on the neural network algorithm was established. The correlation coefficient of internal verification and testing of the model was above 0.8. The remaining six groups of spectral data were used as external test sets, and the correlation coefficient between predicted value and true value was 0.88, and the root mean square error was 0.048. The experimental results show that the multi-spectral total phosphorus concentration detection model based on neural network algorithm is a feasible water quality remote sensing detection technology.

Keywords

Multi-Spectra, Neural Network Algorithm, Total Phosphorous, Water Quality Detection, Sentinel-2A

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

水是生命之源，是社会经济发展的重要物质保障。但是，近几十年来，随着城市化、工业化、农业等人类活动的发展，我国水污染整体呈现恶化态势，水资源短缺及水污染已成为威胁我国人民健康安全和制约经济可持续发展的主要因素。磷是生物生长和能量输送所必需的基本营养元素，也是物质循环重要的化学元素，对水中微生物营养化有着关键性的影响。磷矿开采、化肥施用、食品生产和人类消费等过程都会释放出大量的磷，使磷的流动性大大增强。当这些磷大量流向水体时，会造成水体富营养化，导致藻类过度繁殖、水体溶解氧下降、水质变坏等问题[1]。因此，严格监测水体中的磷含量对于水生态环境的保护是十分有必要的。

传统的总磷检测为采样化验方法包括分光光度法[2]、荧光检测法[3]、红外反射光谱法[4]、色谱检测[5]、质谱检测[6]、电化学检测[7]和流动注射检测[8]等。这些方法虽然检测精度高，能够检测的水质参数多，但是需要对样品进行前处理，耗费大量的时间和成本，并且只能获得点状数据，无法全面实时地监测水质变化。

由于遥感技术具有非接触、区域化监测等优点，已成为水质监测的研究热点。常规的水质遥感参数反演方法包括回归分析法[9]、一阶微分法[10]、偏最小二乘法[11]、主成分分析法[12]和混合光谱分析法[13]等。这些方法的本质是通过线性回归实现水质参数的反演。但是实际水体环境十分复杂，因此利用线性回归模型处理光谱与水质参数间复杂的非线性问题存在一定的缺陷。

近年来，随着人工智能算法的快速发展，水质遥感结合人工智能算法成为水质参数反演的一个新热点，成为水质评价和预测的一种新方法。神经网络算法由于具有大规模并行、分布式处理、非线性

映射能力、自组织和自学习等优点，被广泛应用于语音分析、图像识别、数字水印和食品安全等众多领域[14] [15] [16]。本文以鄱阳湖为例，使用 Sentinel-2A 和实测数据，建立了神经网络模型，对鄱阳湖的总磷进行了参数反演，证明了该方法的可行性，为区域性实时监测水质，保护水文生态环境提供一种技术手段。

2. 数据来源

2.1. 地面监测数据

研究团队使用连华科技生产的多参数水质测定仪 5B-3B(V10)测定现场采集水样的总磷作为真实值。该多参数水质测定仪是基于钼酸铵分光光度法测定总磷浓度的，具体原理为：水中的含磷化合物在过硫酸钾的作用下转变为正磷酸盐。正磷酸盐在酸性介质中，可同钼酸铵和酒石酸氧锑钾反应，生成磷钼杂多酸。磷钼酸能被抗坏血酸还原，生成深色的磷钼蓝，在 700 nm 波长下，测定样品的吸光度，从而得到样品浓度。

2.2. 遥感数据

本文中鄱阳湖的光谱数据来自于 Sentinel-2A 陆地卫星。Sentinel-2A 有 13 个波段如表 1 所示。其中 Band1 为海岸/气溶胶波段，主要用于监测近岸水体和大气中的气溶胶；Band2、3、4 为可见光波段；Band5、6、7 为红光波段，对监测植被健康信息非常有效；Band8 为近红外波段(宽)；Band8A 为近红外波段(窄)；Band9 为水蒸气波段；Band10、11、12 为短波红外波段。本文通过 scihub 下载 Sentinel-2A 数据，通过 Sen2Cor 对 L1C 数据进行大气、地形和卷云校正，形成去除 Band10 的 12 个光谱数据。12 个波段相应的鄱阳湖卫星图片如图 1 所示。

Table 1. The parameters of Sentinel-2A

表 1. Sentinel-2A 参数

波段	中心波长(nm)	空间分辨率(m)
Band 1	443.9	60
Band 2	496.6	10
Band 3	560	10
Band 4	664.5	10
Band 5	703.9	20
Band 6	740.2	20
Band 7	782.5	20
Band 8	835.1	10
Band 8A	864.8	20
Band 9	945	60
Band 10	1373.5	60
Band 11	1613.7	20
Band12	2202.4	20

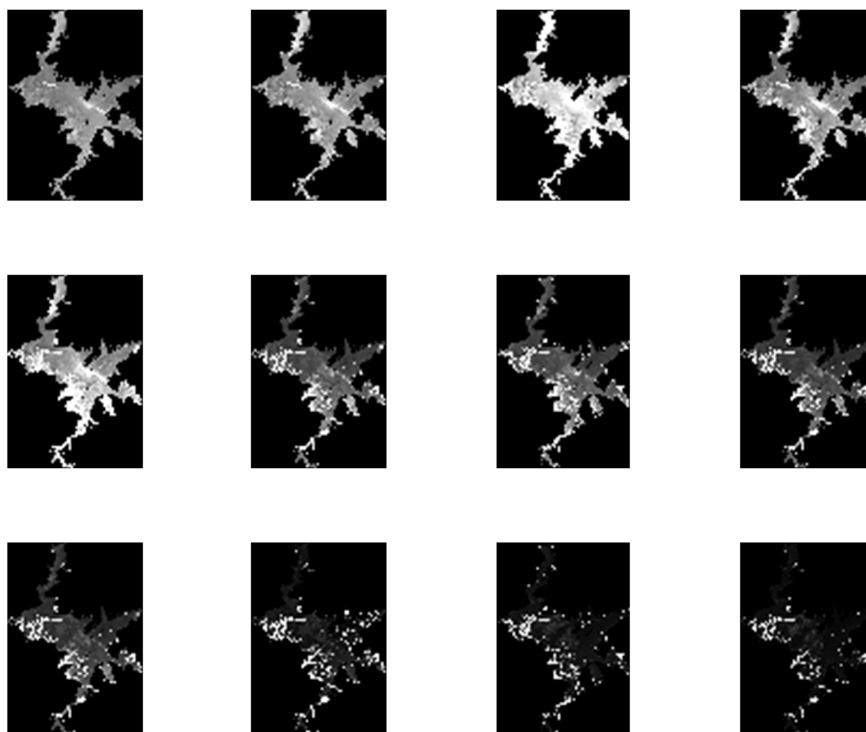


Figure 1. The image of Poyang Lake from Sentinel-2A with different bands
图 1. 不同波段的鄱阳湖影像

2.3. 数据

现场采集了鄱阳湖 41 个采样点的水样，并通过参数水质测定仪 5B-3B(V10)测定 41 个水样的总磷作为真实值。同时通过 GPS 记录采样点的位置，通过 scihub 下载相应地点的 Sentinel-2A 多光谱数据。将这 41 组归一化后的多光谱数据和相应的总磷浓度分为两类：建模集和验证集。建模集共有 35 个光谱组成：25 个光谱数据用于训练模型，5 个光谱数据用于内部验证，5 个光谱数据用于内部测试。验证集为剩余的 6 个高光谱数据。建模前需要对高光谱进行归一化处理。

3. 实验结果与讨论

本文使用 Matlab R2013b 的神经网络工具箱建立神经网络模型，如图 2 所示。模型分为输入层、隐藏层和输出层。输入层的节点数 m 为 12，对应于高光谱的光谱通道数；输出层的节点数 n 为 1，对应于所要求的水质特征数目，本文只有总磷 1 个参数。隐藏层的节点数 l 由经验公式 $l = \sqrt{(m+n)} + a$ ，其中 a 为 1~10 的常数进行试凑，最终确定 $l = 12$ 。数据分类选择随机，此时训练、验证和测试的样本比例为 70:15:15。训练算法设置为 Levenberg-Marquardt (LM) 算法，评价指标设置为均方差 MSE。网络最大训练次数设置为 1000，学习目标设置为 0，最小梯度设置为 10^{-7} ，学习速率设置为 0.001，最大学习速率设置为 1010。由图 2 可知，在 LM 算法下，仅 12 次迭代就完成了训练，出现了 5 次 Validation Checks。

图 3 是建模集中 35 个样品的训练结果，其中随机选择 25 个用来建模，5 个用来验证，5 个用来测试。图中横坐标表示水样的总磷真实浓度，纵坐标表示水样的总磷预测浓度，通过线性拟合真实值与预测值的曲线的斜率及决定系数来判断模型的有效性：斜率越接近于 1 说明模型的预测值越接近真实值，决定系数越大说明真实值与预测值之间的相关度越高。如图 3 所示建模集数据显示真实值和预测值拟合曲线的斜率大于 0.77，说明预测值接近理论值；决定系数大于 0.81，说明模型的训练效果较好，偏差较小。

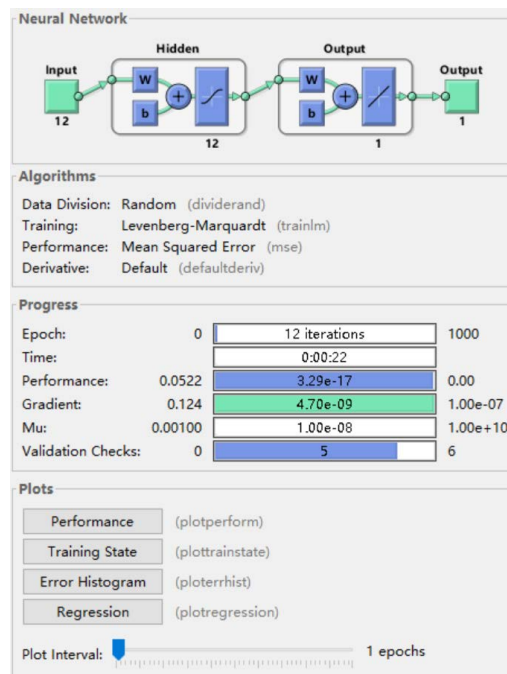


Figure 2. The model of neural network algorithm
图 2. 神经网络模型

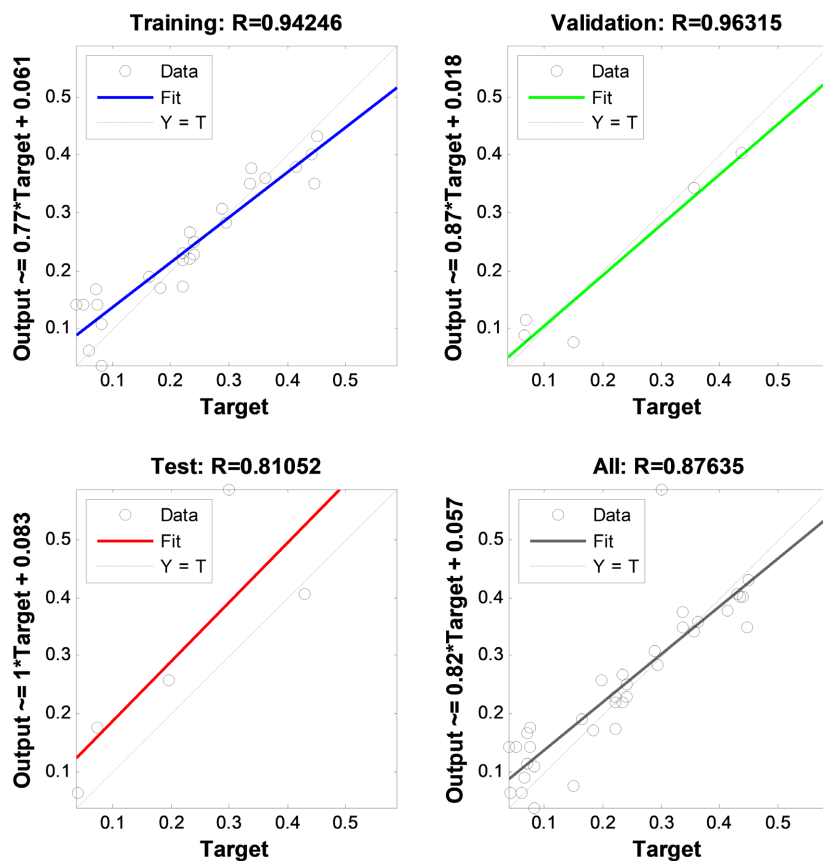


Figure 3. The linear fitting results of training data set
图 3. 训练线性回归图

图 4 是建模集 35 个数据的误差直方图，蓝色的表示训练数据、绿色的表示验证数据、红色的表示测试数据。对误差直方图进行正态分布拟合，得到误差的均值 μ 为 -0.0170506 ，均方差 σ 为 0.0668647 ，如图 5 所示。

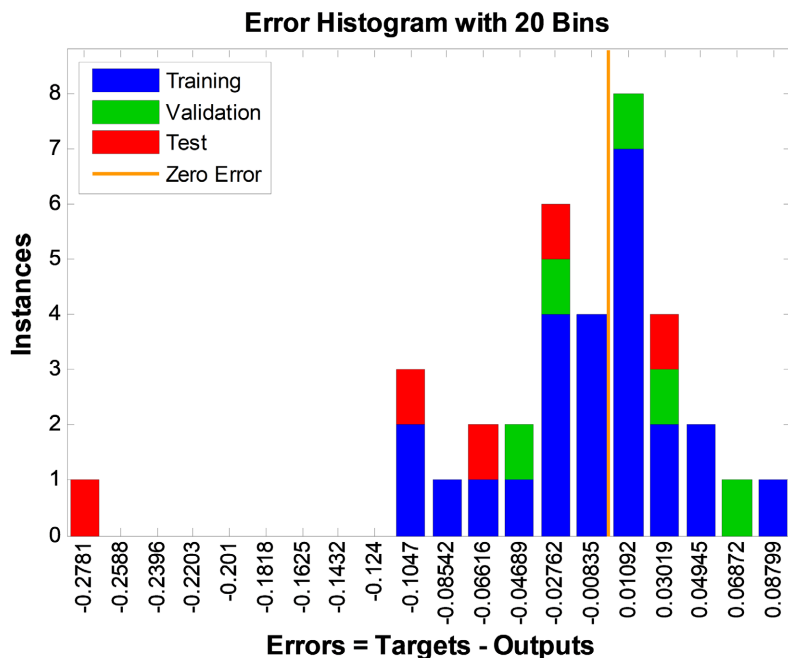


Figure 4. The error histogram of the training data set

图 4. 建模集误差直方图

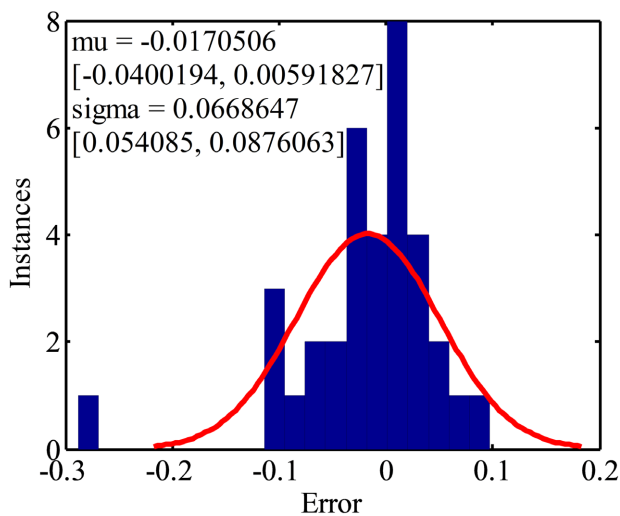


Figure 5. The normal fitting result of the error of the training data set

图 5. 建模集数据误差的正态拟合

利用该模型对验证集中的 6 个光谱数据进行了预测，模型的预测值和真实值的线性拟合曲线如图 6 所示。预测值与真实值线性拟合曲线的斜率为 0.9 决定系数为 0.85，说明模型的预测效果较好。训练集的决定系数为 0.82，均方差为 0.0668647；验证集的决定系数为 0.8592，均方差为 0.05264，训练集与验证集决定系数和均方差相差较小，说明模型没有过拟合。

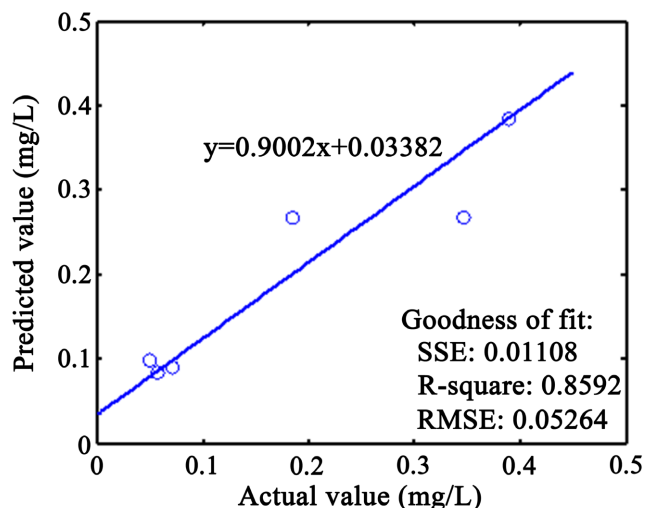


Figure 6. The linear fitting results of validation data set
图 6. 验证集数据的线性回归图

4. 结论

本文提出了一种结合神经网络和多光谱的卫星遥感水质总磷参数的检测方法。以鄱阳湖为例，获取了 41 组 Sentinel-2A 卫星的高光谱数据，利用其中的 35 组数据建立了高光谱数据与总磷浓度的神经网络预测模型，剩余 6 组数据对模型进行验证，结果证明预测值与真实值的线性回归决定系数达 0.8 以上，斜率达 0.77 以上，证明了基于多光谱的卫星遥感水质监测的可行性。

基金项目

中央级公益性科研院所基本科研业务费(TKS20210213)，天津市中新生态城 2019 年度科技型中小企业升级专项项目。

参考文献

- [1] 吴丹, 王昆, 刘英, 等. 水质总磷的检测方法研究进展[J]. 工业水处理, 2022, 42(7): 52-58.
- [2] Gentle, B.S., Ellis, P.S., Faber, P.A., *et al.* (2010) A Compact Portable Flow Analysis System for the Rapid Determination of Total Phosphorus in Estuarine and Marine Waters. *Analytica Chimica Acta*, **674**, 117-122. <https://doi.org/10.1016/j.aca.2010.06.030>
- [3] 刘保生, 王桂华, 孙汉文, 等. 吡啶橙-罗丹明 6G 能量转移荧光法测定痕量磷[J]. 分析化学, 2001, 29(1): 42-44.
- [4] 李恢宏. 近红外光谱的水质总磷检测系统[D]: [硕士学位论文]. 武汉: 中南民族大学, 2016.
- [5] Yokoyama, Y., Danno, T., Haginoya, M., *et al.* (2009) Simultaneous Determination of Silicate and Phosphate in Environmental Waters Using Pre-Column Derivatization Ion-Pair Liquid Chromatography. *Talanta*, **79**, 308-313. <https://doi.org/10.1016/j.talanta.2009.03.053>
- [6] Rocha, A.A., Miekeley, N., Bezerra, M.C.M., *et al.* (2004) An Automated System for Preconcentration/Matrix-Removal Followed by ICP-MS Determination of Organic Phosphorus in Oil Production Water. *Microchemical Journal*, **78**, 65-70. <https://doi.org/10.1016/j.microc.2004.03.012>
- [7] Kolliopoulos, A.V., Kampouris, D.K. and Banks, C.E. (2015) Rapid and Portable Electrochemical Quantification of Phosphorus. *Analytical Chemistry*, **87**, 4269-4274. <https://doi.org/10.1021/ac504602a>
- [8] Gimbert, L.J., Haygarth, P.M. and Worsfold, P.J. (2007) Determination of Nanomolar Concentrations of Phosphate in Natural Waters Using Flow Injection with a Long Path Length Liquid Waveguide Capillary Cell and Solid-State Spectrophotometric Detection. *Talanta*, **71**, 1624-1628. <https://doi.org/10.1016/j.talanta.2006.07.044>
- [9] 郭云开, 钱佳, 雷宇斌, 等. 基于 GF-1 卫星数据的水库叶绿素 a 浓度联合反演研究[J]. 测绘工程, 2021, 30(4):

14-19.

- [10] 金辉. 延边州水库水质光谱特征与营养状况研究[D]: [硕士学位论文]. 延吉: 延边大学, 2014.
- [11] 和梦雪, 刘健, 范文玉, 等. 基于偏最小二乘的絮体图像信息与水质的相关性[J]. 环境监测管理与技术, 2021, 33(6):48-51.
- [12] Nasir, M.F.M., Samsudin, M.S., Mohamad, I., *et al.* (2011) River Water Quality Modeling Using Combined Principle Component Analysis (PCA) and Multiple Linear Regressions (MLR): A Case Study at Klang River, Malaysia. *World Applied Sciences Journal*, **14**, 73-82.
- [13] De Moraes Rudoreff, C., De Moraes Novo, E.M.L. and Galvao, L.S. (2006) Spectral Mixture Analysis for Water Quality Assessment over the Amazon Floodplain Using Hyperion/EO-1 Images. *Ambiente e Agua—An Interdisciplinary Journal of Applied Science*, **1**, 65-79. <https://doi.org/10.4136/ambi-agua.13>
- [14] Bian, H.Y., Yao, H., Lin, G.H., *et al.* (2020) Multiple Kinds of Pesticides Detection Based on Back-Propagation Neural Network Analysis of Fluorescence Spectra. *IEEE Photonics Journal*, **12**, 1-9. <https://doi.org/10.1109/JPHOT.2020.2973653>
- [15] Lin, G.H., Ji, R.D., Yao, H., *et al.* (2020) Fluorescence Detection of Multiple Kinds of Pesticides with Multi Hidden Layers Neural Network Algorithm. *Optik*, **211**, Article ID: 164632. <https://doi.org/10.1016/j.ijleo.2020.164632>
- [16] Yang, W., Zhang, X. and Luo, P. (2021) Transferability of Convolutional Neural Network Models for Identifying Damaged Buildings Due to Earthquake. *Remote Sensing*, **13**, Article No. 504. <https://doi.org/10.3390/rs13030504>