



文章栏目：环境监测技术

DOI 10.12030/j.cjee.202009182

中图分类号 X832; X524

文献标识码 A

崔嘉宇, 郭蓉, 张悦, 等. 基于极限学习机算法对太湖东部湖区水源地总磷数据的自动监测分析[J]. 环境工程学报, 2021, 15(6): 2165-2173.

CUI Jiayu, GUO Rong, ZHANG Yue, et al. Analysis of automatic monitoring data of total phosphorus in drinking water source in east Taihu Lake based on improved extreme learning machine algorithm[J]. Chinese Journal of Environmental Engineering, 2021, 15(6): 2165-2173.

## 基于极限学习机算法对太湖东部湖区水源地总磷数据的自动监测分析

崔嘉宇<sup>1,2</sup>, 郭蓉<sup>1,\*</sup>, 张悦<sup>1,2</sup>, 徐亮<sup>1</sup>, 钟声<sup>1</sup>, 董圆媛<sup>1</sup>, 李心雨<sup>3</sup>

1. 江苏省环境监测中心, 南京 210019

2. 江苏省苏力环境科技有限责任公司, 南京 210036

3. 南京大学环境学院, 南京 210093

第一作者: 崔嘉宇(1988—), 男, 硕士, 工程师。研究方向: 水环境监测、分析。E-mail: [cjy19881029@163.com](mailto:cjy19881029@163.com)

\*通信作者: 郭蓉(1985—), 女, 硕士, 高级工程师。研究方向: 水环境监测与污染防治。E-mail: [guorong@jshb.gov.cn](mailto:guorong@jshb.gov.cn)

**摘要** 对太湖东部湖区金墅港和渔洋山水源地总磷自动监测值与实验室监测值进行比对分析, 基于相关性分析, 选取对自动监测总磷浓度影响较大的浊度、蓝绿藻和叶绿素等指标, 提出可选择极限学习机隐层节点的改进极限学习机模型(IELM)修正自动监测总磷数据。结果表明, 总磷自动监测值相对实验室监测值偏差较大, 绝对误差范围在  $0.05\sim 0.112\text{ mg}\cdot\text{L}^{-1}$ , 平均绝对误差为  $0.017\text{ mg}\cdot\text{L}^{-1}$ 。若以相对误差小于或等于 30% 作为比对合格的标准值, 总磷自动监测数据合格率仅为 52.9%。经过 IELM 算法模型训练和测试, 训练均方误差为 0.000 073 5, 测试均方误差为 0.000 103。经过模型修正后的自动监测总磷数据更接近实验室监测值, 其平均绝对误差降低  $0.026\text{ mg}\cdot\text{L}^{-1}$ , 平均相对误差降低 45%。按比对误差 30% 计算, 模型修正后的总磷自动监测数据合格率为 92.0%, 可应用于自动监测总磷数据的修正。

**关键词** 总磷; 自动监测; 改进极限学习机; 隐层节点; 修正数据

总磷 (total phosphorus, TP) 是表征水体富营养化程度的重要污染特征因子, 是评价水质的重要指标之一。磷是促进生物生长的重要元素, 水体环境中磷浓度过高会导致藻类过度繁殖, 造成水体富营养化<sup>[1]</sup>。目前, 我国国标规定的 TP 测定方法为钼酸铵分光光度法 (GB 11893-1989)。传统的 TP 监测具有准确性高、抗干扰性强、质控体系完善等优点, 但需要占用大量时间、人力和物力, 难以实时反映水质变化情况, 故无法满足水源地水质安全监控需求。水中 TP 的自动监测可实现实时在线监测, 已经成为环保、水利等部门水质监测的主要手段<sup>[2]</sup>。然而, 传统 TP 自动监测方法的抗干扰性弱, 易受水体色度与浊度 (Turb) 的变化及样品处理条件等干扰<sup>[3]</sup>, 不确定性较高, 且监测结果与同步实验室的结果相比存在一定差异, 故无法精准反映水体的实际情况。

收稿日期: 2020-09-26; 录用日期: 2021-04-26

基金项目: 国家水体污染控制与治理科技重大专项 (2017ZX07302-003); 2018 年江苏省环保科研课题 (重大技术攻关类)(2018008); 江苏省环境监测科研基金项目 (2003)

近年来, 人工神经网络广泛应用于水质研究。传统的神经网络通过误差反馈不断调整权重可解决一些实际问题, 但梯度下降算法容易导致学习陷入局部最优状态、学习过程耗时长等问题。针对上述问题, HUANG 等<sup>[4-5]</sup>提出了极限学习机算法(extreme learning machine, ELM)。该算法由广义逆直接求得输出层权重, 在确保精度的同时提升了算法的学习速度。ELM 已被探索应用于环境领域的研究中。张颖等<sup>[6]</sup>基于粒子群算法优化的极限学习机对淮河水质进行类别判定; 崔东文<sup>[7]</sup>构建了基于 ELM 的湖库 TP、TN 模型, 具备参数选择简便、训练速度快、不会陷入局部最优值等优点; 边冰等<sup>[8]</sup>验证了深度极限学习机在水质预测方面具备合理性和可行性。

本研究基于江苏省环境监测中心在太湖东部湖区金墅港和渔洋山水源地建设的 2 个水质自动监测站 TP 数据, 分析 TP 自动监测与实验室监测的差距及其影响因素。采用改进的极限学习机算法, 建立包含 TP 自动监测数据、相关影响因子及 TP 实验室监测数据的学习机模型, 修正 TP 自动监测数据, 以便进一步缩小与实验室数据的差异, 更好地表征水源地 TP 的真实状况, 为太湖环境综合治理提供参考。

## 1 材料与方 法

### 1.1 研究区域概况

太湖富营养化问题突出。多年来, 大多研究集中在蓝藻水华问题形势严峻的太湖西部和北部区域<sup>[9]</sup>, 而对太湖东部湖区的研究相对较少。根据江苏省环境监测中心长期水质及卫星遥感监测结果, 2016 年前, 太湖东部湖区水质较好, TP 较低, 未曾出现蓝藻水华现象; 而 2016 年以后, 太湖东部湖区开始出现蓝藻水华, 同时 TP 逐渐升高。一是由于入湖的磷污染物大幅增加, 并通过湖体水动力变化, 逐年从西向东迁移<sup>[10]</sup>; 二是由于东太湖水生植被的大量减少, 沉水植物对磷元素的吸收相应减少, 增加了风浪对底泥扰动导致的内源性营养盐的释放<sup>[11]</sup>。此外, TP 变化还与风向、风速等因素密切相关<sup>[12]</sup>。2019 年个别月份, 太湖东部湖区金墅港和渔洋山水源地的 TP 甚至超过地表水环境质量标准(GB 3838-2002)集中式生活饮用水源地二级保护区Ⅲ类标准值, 直接威胁饮用水水源地水质安全。

### 1.2 数据处理

太湖东部湖区金墅港、渔洋山水源地(位置如图 1 所示)的 TP 自动监测数据频次为每 2 h 一次, 较高的数据量可实现 TP 的有效监测。目前, 这 2 处水源地的 TP 自动监测仪器是日本岛津 TNP4110, 测定原理是采用碱性过硫酸钾消解-紫外分光光度法分析水样中的 TP, 仪器稳定性好, 分析方法符合国家环境保护行业标准 HJ 103-2003 的要求。然而, 若出现较大风浪导致水体浑浊, 以及水体中出现过多蓝藻颗粒等情况时, 会干扰 TP 消解和比色过程, 此时 TP 的测定浓度并不能准确反映 TP 的实际情况<sup>[13]</sup>。本研究选用金墅港、渔洋山水源地 2016 年 1 月—2020 年 4 月的 TP 自动监测数据, 与实验室监测数据进行比对分析。实验室监测数据频次为每月一次, 方式为手工采样。采样点位于自动监测站取水口, 样品的采集、运输、保存均执行地表水采样技术规范(HJ 91-2002、HJ 493-2009)。实验室分析采用钼酸铵分光光度法(GB 11893-1989)。实验室监测数据能准确反映水源

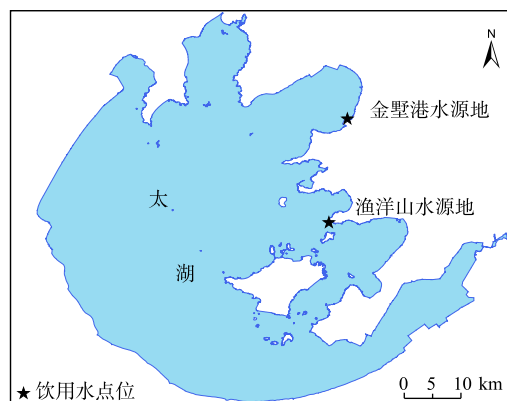


图 1 太湖饮用水水源地监测点位分布图  
Fig. 1 Monitoring sites in drinking water source in east Taihu Lake

地 TP, 可作为 TP 自动监测数据的真值。比对分析时, 自动监测选用在采样时间上与实验室手工采样时间最为临近的一组数据, 时间差为 0~59 min。

### 1.3 研究方法

极限学习机 (ELM) 是一个单隐层的神经网络, 是在 Moore-Penrose 矩阵理论上提出的快速学习算法<sup>[14-15]</sup>。相较传统的神经网络具有训练速度快、参数选择简单、不易陷入局部最优等特点, 可运用于水质评价和水质预测等方面。已有研究<sup>[16]</sup>表明, ELM 网络模型已在多输入、多输出的水质预测中取得了较好的效果。

根据极限学习机理论<sup>[4-5]</sup>, 对于一个有  $L$  个节点的单隐层神经网络可表示为式 (1)。

$$\sum_{i=1}^L \beta_i g(W_i \cdot X_j + b_i) = t_j, j = 1, \dots, N \quad (1)$$

式中:  $g(x)$  为激活函数,  $W_i = [W_{i1}, W_{i2}, \dots, W_{im}]^T$  为连接第  $i$  个隐层节点的输入权值;  $b_i$  为第  $i$  个隐层节点的偏置;  $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$  为第  $i$  个隐层节点的输出权值。令  $H = g(W_i \cdot X_j + b_i)$ , 则式 (1) 可换算为式 (2)。

$$H\beta = T \quad (2)$$

$$\text{式中: } H(W_1, W_2, \dots, W_L, b_1, b_2, \dots, b_L, x_1, x_2, \dots, x_N) = \begin{bmatrix} g(W_1 \cdot X_1 + b_1) & \cdots & g(W_L \cdot X_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(W_1 \cdot X_N + b_1) & \cdots & g(W_L \cdot X_N + b_L) \end{bmatrix}; \quad \beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times M}$$

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times M}。$$

为达到训练效果, 需要得到  $W_i$ 、 $\beta_i$ 、 $b_i$  使式 (2) 无限接近于最小化, 得到式 (3)。

$$\|H\beta - T\| = \min \|H\beta - T\| \quad (3)$$

由此得到等价的最小化损失函数, 见式 (4)。

$$E = \sum_{j=1}^N \left( \sum_{i=1}^L (\beta_i g(W_i \cdot X_j + b_i)) - t_j \right) \quad (4)$$

在 ELM 算法中, 一旦输入权重  $W_i$  及隐层的偏置  $\beta_i$  被随机确定, 隐层的输出矩阵  $H$  就被确定。训练单隐层神经网络可转化为求解一个线性系统  $H\beta = T$ , 并且输出权重  $\beta$  的范数最小且唯一, 计算公式见式 (5)。

$$\beta = H^{-1}T \quad (5)$$

式中:  $H^{-1}$  是矩阵  $H$  的 Moore-Penrose 广义逆。

根据 ELM 原理, 隐层节点数越多, 训练误差会越小。当隐层节点数等于样本数时, 训练误差会趋于 0; 但会由此造成算法隐层节点的冗余, 且一些隐含层节点并无意义<sup>[16-17]</sup>, 故有学者在水质评价时通过粒子群算法来优化极限学习机的输入权值和隐含层偏置<sup>[6]</sup>。然而, 粒子群算法对离散的优化问题处理不佳, 容易产生早熟收敛。本研究提出可选择极限学习机隐层节点的改进极限学习机模型 (IELM, Improved ELM), 采用多次试验取平均值的方法确定隐层节点数。多次试验求平均可减少随机产生的权值和阈值对模型稳定性的干扰, 并增强学习机的稳定性<sup>[18-19]</sup>。通过试触法对每个节点数所建立的模型进行测试, 分析测试误差的变化趋势, 以此来确定合适的隐层节点数, 可大大提升模型的预测精度, 并有效降低随机值的产生对模型的干扰。

## 2 结果分析与讨论

### 2.1 自动监测值与实验室监测值的比对

图2为2016年以来金墅港水源地和渔洋山水源地的TP自动监测和实验室监测值的对比图。由图2可见,TP自动监测值和实验室监测值相差值为 $-0.05\sim 0.112\text{ mg}\cdot\text{L}^{-1}$ ,计算差值绝对值的平均值为 $0.017\text{ mg}\cdot\text{L}^{-1}$ ,相对于湖库型TP标准为II类( $0.01\sim 0.025\text{ mg}\cdot\text{L}^{-1}$ );而最大值 $0.112\text{ mg}\cdot\text{L}^{-1}$ ,相对于湖库型TP浓度标准为V类( $0.1\sim 0.2\text{ mg}\cdot\text{L}^{-1}$ )。自动站每隔一段时间会将自动监测值与实测值进行人工比对,以确定自动站是否正常运行。根据《江苏省水质自动监测质量管理暂行规定(苏环监〔2012〕6号)》,将自动值和手工值之差与手工值的比值定为比对误差。比对误差在 $\pm 30\%$ 以内,TP自动监测数据才能有效反应实际情况。由此计算得到TP自动监测数据的合格率仅为52.9%。

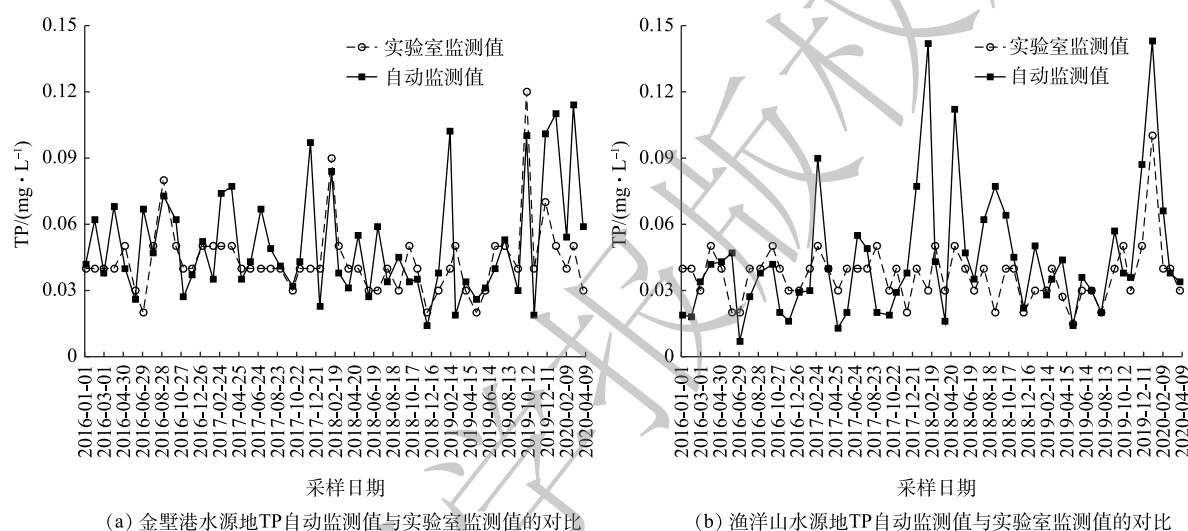


图2 自动监测和实验室监测TP浓度变化

Fig. 2 Variation of total phosphorus concentrations by automatic water monitoring compared with laboratory method

造成数据偏差较大的原因有如下几点。1) 悬浮颗粒物是氮磷等营养元素在水中迁移转化的重要载体<sup>[20]</sup>,同时也是微生物附着生长的重要载体<sup>[11]</sup>,风浪等其他因素搅动底泥致使浊度增大,水中悬浮颗粒物增多,故实测数据偏大。2) 悬浮物和蓝绿藻携带的TP经过消解转换为可检测的无机磷导致检测值较水体实际值高,另外蓝绿藻色度会干扰TP显色反应、浑浊的水质会干扰比色透光性亦会导致检测值和实际值产生偏差。3) 对于实验室TP分析。根据《国家地表水环境质量监测网监测任务作业指导书(试行)(环办监测函〔2017〕249)》,TP人工采样静置及虹吸方式有利于粒径较大的颗粒物沉淀,可在一定程度上降低浊度对测定的干扰<sup>[3,21]</sup>。在对太湖进行的实际采样过程中,若水样藻类较多,会采用 $63\ \mu\text{m}$ 筛过滤掉部分藻,并在分析时采用浊度及色度补偿以消除实验室的比色干扰,以进一步降低对TP测定结果的影响。样品从采样、运输,到实验分析、数据报出都有完整的全过程质量控制,以确保人工监测数据准确性更高。

### 2.2 对TP影响因子的分析

选取金墅港和渔洋山水源地水质自动监测站2016年1月—2020年4月的pH、溶解氧(DO)、浊度(Turb)、高锰酸盐指数( $\text{COD}_{\text{Mn}}$ )、氨氮( $\text{NH}_4\text{-N}$ )、叶绿素质量浓度(Chl-a)、蓝绿藻密度(藻密度)等指标与TP进行相关性分析,自动监测频率为每2h一次。结果表明:Turb与TP的相关性较高,相关系数为0.442;其次为藻密度,与TP的相关系数为0.207;相关性最小的为pH,与TP的相关系数为 $-0.015$ 。各指标与TP的相关性分析结果见表1。

表 1 自动监测 TP 数据与其他监测指标的相关性分析

Table 1 Spearman correlation analysis between total phosphorus and other water quality parameters by automatic water monitoring

监测指标	pH	DO	Turb	COD <sub>Mn</sub>	NH <sub>4</sub> <sup>+</sup> -N	蓝绿藻	Chl-a
TP	-0.015	-0.048	0.442*	0.088	0.050	0.207	0.123
pH		0.531**	-0.489*	-0.175	0.126*	-0.077*	-0.115*
DO			-0.173**	-0.189**	0.198	0.150	0.038
Turb				0.731*	0.064	0.139	-0.004
COD <sub>Mn</sub>					0.161	0.477*	0.317*
NH <sub>4</sub> <sup>+</sup> -N						0.037	0.108
藻密度							0.075

注：\*表示在0.05水平下的显著水平，\*\*表示在0.01水平上显著相关；监测指标中，除pH无单位、藻密度单位为10<sup>4</sup>个·L<sup>-1</sup>外，其余单位为mg·L<sup>-1</sup>。

7个指标中，Chl-a、藻密度与TP的相关性较弱，分别为0.123、0.207。然而，也有学者发现太湖水质中的Chl-a、藻密度和TP呈显著正相关关系<sup>[11, 22-23]</sup>；对与TP相关系数较大的Turb、藻密度和Chl-a等3项指标进行自相关分析发现，Turb与藻密度的相关性为0.139，具有一定相关性。以上结果的主要原因为颗粒态藻类会导致水体浊度增加，藻密度对水体色度会干扰显色反应，影响测定结果。其余指标之间相关系数均小于0.1，无相关性。

### 2.3 IELM 模型构建

根据TP的影响因素分析，将TP自动监测值及其影响较大的Turb、藻密度和Chl-a等指标数据作为模型的输入数据集，TP的实验室监测值作为模型测试数据集，构建ELM模型。

选取2个水源地2016年1月—2019年10月共92组数据为模型训练数据集，即训练样本数 $N=92$ ；选取2个水源地2019年11月—2020年4月共12组数据作为模型测试数据集。为避免指标单位不统一对模型的影响，将Turb、藻密度和Chl-a进行归一化处理。正则化后的数据与TP自动监测值组合成数据输入集。

将92组训练样本数据单独分为训练样本和测试样本：前80组数据作为训练样本；后12组数据作为测试样本。以1~80作为隐层节点的个数，分别进行10次、50次、100次、200次、500次和1000次的训练，分析不同训练次数与隐层节点数的平均均方误差(mean-mse, M-MSE)的变化规律，结果如图3所示。

由图3可见，重复训练10、50和100次的M-MSE随隐层节点数增加的变化趋势不稳定，说明训练次数较少不能很好地降低随机值干扰。训练200次以上，随机值干扰明显降低。到了500、1000次的时候，随机值的干扰很小。两者在最优节点数的M-MSE相差仅0.000005，可忽略不计。训练的M-MSE在1~80个隐层节点总体呈逐级下降趋势。1~4个隐层节点对应的M-MSE呈指数级递减，4~80个隐层节点对应的M-MSE递减趋势逐步放缓。测试的M-MSE在1~4个隐层节点对应M-MSE变化趋势与训练的M-MSE走势一致；而在5个隐层节点以后对应M-MSE的变化幅度较小；在21个隐层节点左右，对应的M-MSE降至最小；在40个隐层节点以后对应的M-MSE快速增加。

综合以上规律，同时考虑内存和计算成本等因素，确定该模型最佳隐层节点数为21个，最佳重复训练次数为500次。

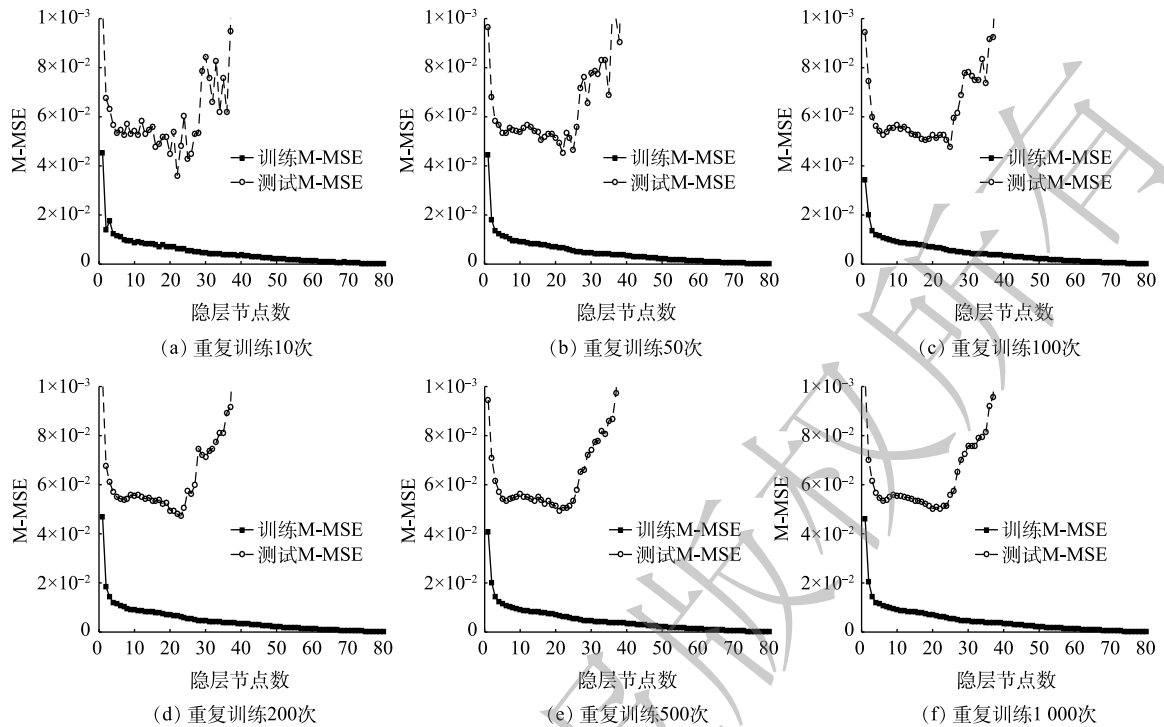


图3 不同隐层节点数的 M-MSE 变化情况

Fig. 3 Variation of M-MSE under different numbers of hidden layer nodes

## 2.4 模型测试

将前 92 组数据的 TP 自动监测值、Turb、藻密度、Chl-a 作为模型的训练输入数据集，TP 实验室监测值作为训练输出数据集，隐层节点数设为 21，建立模型，重复训练 500 次。将后 12 组数据作为测试模型。经过模型训练和测试，试验用时较短为 4.4 s。模型训练的 M-MSE 为 0.000 073 5，模型测试的 M-MSE 为 0.000 103。在  $P$  值小于 0.01 时，相关系数为 0.881，具有显著相关性。模型训练的结果见图 4，模型测试的结果见表 2。由模型测试值和实验室监测值的差值来看，其绝对误差为  $-0.017\sim 0.012\text{ mg}\cdot\text{L}^{-1}$ ，相较自动监测值和实验室监测值的绝对误差为  $-0.021\sim 0.084\text{ mg}\cdot\text{L}^{-1}$ ，绝对误差降幅较大。相对误差在 30% 以内的数据有 11 个，超过 30% 的数据有 1 个，为 40%，但相较测试前的相对误差有大幅降低。比较模型测试前后相对误差的绝对值，除第 12 个数据相对误差略有升高外，其余相对误差的降幅为 4%~103%，相对误差总体降幅较大。因此，经过模型修正的 TP 自动监测数据已大大降低了其他因素的干扰，模型测试值更接近实验室监测值。

模型测试中第 8 组和第 11 组数据误差  $\geq 30\%$ 。这 2 组数据中，Turb 分别为 23 NTU 和 26 NTU，已有研究<sup>[24]</sup>表明水质浊度值大于 80 NTU 时才会对 TNP 分析仪(日本岛津)的监测结果产生明显干扰；藻密度分别为  $368\times 10^4$  和  $387\times 10^4$  个 $\cdot\text{L}^{-1}$ ，相较于 2019 年度太湖蓝藻预警期间湖体藻密度均值 ( $1\ 218\times 10^4$  个 $\cdot\text{L}^{-1}$ )，水源地的藻密度相对较低。因此，Turb 和藻密度并不是导致这 2 组数据偏差较大的原因。TP 的测定

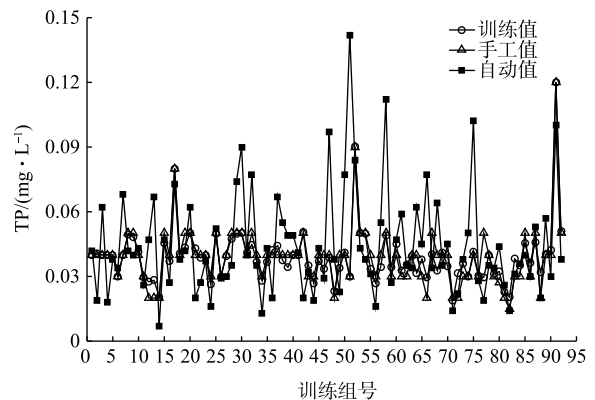


图4 TP 模型训练结果比较

Fig. 4 Training results and performance comparison of total phosphorus concentrations

表2 TP模型测试值与实测值对比

Table 2 Comparison of model test values and measured values of total phosphorus concentrations

序号	TP/(mg·L <sup>-1</sup> )				相对误差		
	实验室监测值	自动监测值	模型测试值	自动监测与实验室监测差值	实验室监测与模型计算差值	自动监测与实验室监测	模型计算与实验室监测
1	0.040	0.019	0.030	-0.021	-0.010	-53%	-24%
2	0.030	0.036	0.034	0.006	0.004	20%	13%
3	0.070	0.117	0.055	0.047	-0.015	67%	-21%
4	0.050	0.087	0.043	0.037	-0.007	74%	-13%
5	0.050	0.110	0.041	0.060	-0.009	120%	-19%
6	0.100	0.184	0.083	0.084	-0.017	84%	-17%
7	0.040	0.054	0.043	0.014	0.003	35%	6%
8	0.040	0.066	0.052	0.026	0.012	65%	30%
9	0.050	0.114	0.037	0.064	-0.013	128%	-25%
10	0.040	0.038	0.040	-0.002	0.000	-5%	-1%
11	0.030	0.059	0.042	0.029	0.012	97%	40%
12	0.030	0.034	0.034	0.004	0.004	13%	15%

可能受到其他因素的影响,例如TP分析中磷钼蓝络合物的强吸附能力会导致实验用的玻璃器皿及比色皿必须充分酸泡,否则容易产生干扰;另外,实验中大多数玻璃器皿中含有的少量硼硅酸盐会在显色时产生痕量硅钼蓝,引入正误差<sup>[25]</sup>。

### 3 结论

1)太湖东部湖区水源地自动监测TP数据和实验室TP数据偏差较大,受Turb、藻密度、Chl-a等影响,TP自动监测数据的合格率仅为52.9%,严重影响TP水质自动监控预警系统的发挥。

2)通过试触法确定ELM隐层节点数和多次训练求均值的方法可减少随机产生输入权值和阈值所导致的模型误差,提高了模型的准确度和可重复性。经过IELM修正的TP自动监测数据与实验室数据的偏差明显降低。因此,将极限学习机算法嵌入到自动监测系统中修正TP自动监测数据,可有效提升数据的准确性,使其真实较为地反映TP实际情况,为管理决策及水质预警提供可靠数据支撑。

### 参考文献

- [1] 陈小华,李小平,程曦,等.太湖流域典型中小型湖泊富营养化演变分析(1991-2010)[J].湖泊科学,2013,25(6):846-853.
- [2] 李军,陈程.总氮总磷在线自动监测仪器的现状与问题[J].中国环境监测,2013,29(2):156-158.
- [3] 娄保锋,藏小平,洪一平,等.水样不同处理方式对总磷监测值的影响[J].环境科学学报,2006,26(8):1393-1399.
- [4] HUANG G B, ZHU Q Y, SIEW C K, et al. Extreme learning machine: theory and applications[J]. Neurocomputing, 2006, 70(1/2/3): 489-501.
- [5] HUANG G B, CHEN L, SIEW C K, et al. Universal approximation using incremental feed forward networks with arbitrary input weights[J]. Neural Networks, 2006, 17(4): 879-892.

- [6] 张颖, 李梅. 基于粒子群优化极限学习机的水质评价新模型[J]. 环境科学与技术, 2016, 39(5): 135-139.
- [7] 崔东文. 极限学习机在湖库总磷、总氮浓度预测中的应用[J]. 水资源保护, 2013, 29(2): 61-66.
- [8] 边冰, 赵明政. 基于深度极限学习机的水质预测研究[J]. 华北理工大学学报(自然科学版), 2020, 42(1): 51-57.
- [9] 冯驰, 金琦, 王艳楠, 等. 基于GOCI影像和水体光学分类的内陆湖泊叶绿素a浓度遥感估算[J]. 环境科学, 2015, 36(5): 1557-1564.
- [10] 张玲玲, 郭兴芳, 顾森, 等. 大型调水对受水湖体富营养化关键因子的影响[J]. 中国给水排水, 2020, 36(1): 58-62.
- [11] 王华, 陈华鑫, 徐兆安, 等. 2010-2017年太湖总磷浓度变化趋势分析及成因探讨[J]. 湖泊科学, 2019, 31(4): 919-929.
- [12] 杜成功, 李云梅, 王桥, 等. 面向GOCI数据的太湖总磷浓度反演及其日内变化研究[J]. 环境科学, 2016, 37(3): 862-872.
- [13] 林舒. 水中总磷测定各影响因素的探讨及方法优化[J]. 中国无机分析化学, 2019, 9(6): 1-5.
- [14] FENG G, HUANG G B, LIN Q, et al. Error minimized extreme learning machine with growth of hidden nodes and incremental learning[J]. *IEEE Transactions on Neural Networks*, 2009, 20(8): 1352-1357.
- [15] 徐睿, 梁循, 齐金山, 等. 极限学习机前沿进展与趋势[J]. 计算机学报, 2019, 42(7): 1640-1670.
- [16] 于海洋. 在线预测的极限学习机方法研究[D]. 长春: 吉林大学, 2019.
- [17] DENG W Y, BAI Z, HUANG G B, et al. A fast SVD-hidden-nodes based extreme learning machine for large-scale data analytics[J]. *Neural Networks*, 2016, 77(1): 14.
- [18] 徐勇, 王东, 张慧. 基于自适应在线极限学习机模型的预测方法[J]. 统计研究, 2016, 33(7): 103-109.
- [19] 陆慧娟, 安春霖, 马小平, 等. 基于输出不一致测度的极限学习机集成的基因表达数据分类[J]. 计算机学报, 2013, 36(2): 341-348.
- [20] 高湘, 李妍, 何怡. 湖泊底泥磷释放及磷形态变化[J]. 环境工程学报, 2015, 9(7): 3350-3354.
- [21] 吴丹, 金小伟, 朱红霞, 等. 地表水总磷测定中现场前处理方式及分析方法比对研究[J]. 中国环境监测, 2020, 36(4): 123-131.
- [22] 朱广伟, 秦伯强, 张运林, 等. 2005-2017年北部太湖水体叶绿素a和营养盐变化及影响因素[J]. 湖泊科学, 2018, 30(2): 279-295.
- [23] 于洋, 彭福利, 孙聪, 等. 典型湖泊水华特征及相关影响因素分析[J]. 中国环境监测, 2017, 33(2): 88-94.
- [24] 夏文文, 钟声, 郁建桥. 浊度对不同型号总磷自动监测分析仪的影响[J]. 环境监控与预警, 2015, 7(4): 28-30.
- [25] 金筱请, 周慧. 过硫酸钾-钼锑抗分光光度法测定地表水中总磷的若干影响因素[J]. 环境监测管理与技术, 2002, 14(1): 38.
- (责任编辑: 靳炜)



## Analysis of automatic monitoring data of total phosphorus in drinking water source in east Taihu Lake based on improved extreme learning machine algorithm

CUI Jiayu<sup>1,2</sup>, GUO Rong<sup>1,\*</sup>, ZHANG Yue<sup>1,2</sup>, XU Liang<sup>1</sup>, ZHONG Sheng<sup>1</sup>, DONG Yuanyuan<sup>1</sup>, LI Xinyu<sup>3</sup>

1. Jiangsu Environmental Monitoring Center, Nanjing 210019, China

2. Jiangsu Suli Environmental Science and Technology Co., Ltd, Nanjing 210036, China

3. School of the Environment, Nanjing University, Nanjing 210093, China

\*Corresponding author, E-mail: guorong@jshb.gov.cn

**Abstract** A comparative analysis was performed on the concentration of total phosphorus measured in the Jinshugang and Yuyangshan drinking water source areas in East Taihu Lake by automatic water monitoring and the laboratory method. Three key factors were found to influence the automatic monitoring of total phosphorus according to a correlation analysis, i.e. water turbidity, algal density and chlorophyll-a. The three parameters were introduced into the improved extreme learning machine model (IELM) for correction of automatic water monitoring data of total phosphorus. Compared with laboratory method, the total phosphorus concentrations measured by automatic water monitoring were shown to have relatively large errors with absolute error ranging between  $0.05 \text{ mg}\cdot\text{L}^{-1}$  and  $0.112 \text{ mg}\cdot\text{L}^{-1}$  and the mean absolute error being  $0.017 \text{ mg}\cdot\text{L}^{-1}$ . The training error was 0.0000735 and test error was 0.000103 after the training and testing of the IELM model. The measurement results by the application of IELM showed better performance. With 30% of relative error rate as judgment criteria, the eligible rates increased from 52.9% to 92.0%, the absolute error decreased by 0.026 mg/L on average, and the relative error rate dropped by 45%. The results in this study show the promise of application of IELM model in correction of automatic monitoring of total phosphorus in the field.

**Keywords** total phosphorus; automatic monitoring; improved extreme learning machine algorithm; hidden node; corrected data