

doi: 10.3969/j.issn.1003-2029.2026.01.002

海水硝酸盐原位传感器小型化紫外检测波长优化与仿真研究

苗鑫, 关银龙, 李丛, 曹守启

(上海海洋大学信息学院, 上海 201306)

摘要: 面向海水检测行业对小型化、低功耗海水硝酸盐原位传感器的迫切需求, 解决传统紫外光谱宽谱扫描方案难以兼顾微型化与检测精度的核心问题。本文构建集成硝酸盐吸收、溴离子等干扰效应及仪器噪声的高保真仿真模型, 提出有限最优波长检测策略, 结合随机森林算法实现波长智能筛选与非线性建模。根据仿真结果筛选出 229 nm、230 nm、290 nm 三个波长组合, 所建模型在 0.5~1 000 $\mu\text{mol/L}$ 浓度范围内决定系数 (R^2) 达到 0.965, 残差相对误差约为 2.87%; 确定仪器信噪比 14 dB 为核心设计门槛。该策略显著降低传感器光源与分光系统复杂度, 为基于紫外发光二极管的小型化低功耗传感器奠定了坚实基础, 尤其适用于寡营养海域低浓度硝酸盐检测。

关键词: 硝酸盐检测; 紫外光谱; 传感器小型化; 有限波长; 随机森林; 低功耗

中图分类号: P716.5; TH744.1; X834 **文献标识码:** A **文章编号:** 1003-2029 (2026) 01-0014-08

海洋作为地球生命的摇篮和全球气候系统的重要调节器, 其水体的化学组成变化直接关系到生态系统的平衡与人类的生存发展。其中, 溶解无机氮, 特别是硝酸盐的浓度, 是评估海水营养水平、理解碳氮生物地球化学循环及监测赤潮等生态灾害现象的核心指标之一^[1]。传统海水硝酸盐的测定主要依赖于水样采集后实验室内的化学分析方法^[2], 如镉柱还原-重氮偶氮分光光度法等^[3]。尽管这些方法具有较高的精度, 但其操作流程繁琐、耗时长、成本高昂, 且无法实现时间与空间上的高分辨率连续观测, 极大地限制了对海洋动力过程快速变化的认知能力。因此, 研发能够实现原位、实时、连续监测的海水硝酸盐传感器, 已成为物理海洋学、海洋生态学以及环境监测领域的迫切需求。

在众多原位检测技术中, 紫外吸收光谱法^[4]展现出独特的应用潜力。该方法基于硝酸根离子在紫外光区, 特别是在 200 nm 至 300 nm 波段存在特征

吸收这一物理现象, 通过测量特定波长下光强的衰减, 依据朗伯-比尔定律反算出硝酸盐的浓度。与化学方法相比, 紫外光谱法最大的优势在于无试剂测量, 避免了化学试剂的消耗、储存与废液处理问题。紫外光谱法极大地降低了维护需求, 使得传感器能够长期布放在浮标、剖面仪等观测平台上, 获取宝贵的长期序列数据。此外, 紫外光谱响应速度快, 有助于捕捉海洋中发生的分钟级甚至秒级的快速生化过程。

然而, 将紫外光谱法成功应用于复杂多变的海水基质中, 面临着严峻的挑战。海水中并非仅有硝酸盐吸收紫外光^[5], 其他成分通常被称为干扰物, 会共同贡献于总吸光度, 导致测量值偏离真实值。在这些干扰物中, 溴离子的影响最为显著。海水中的溴离子浓度与盐度密切相关且相对稳定, 其在 200 nm 至 240 nm 波段具有连续的吸收带, 与硝酸盐的特征吸收峰严重重叠。有色溶解有机物主要来

收稿日期: 2025-09-29

基金项目: 国家重点基础研究发展计划资助项目 (D-8011-20-0115)

作者简介: 苗鑫 (1992—), 男, 硕士, 工程师, 主要从事海洋智能装备研究。E-mail: xmiao@shou.edu.cn

通信作者: 曹守启 (1973—), 男, 博士, 教授, 主要从事海洋智能装备研究。E-mail: sqcao@shou.edu.cn

源于动植物残体分解产生的黄色物质,在紫外区也有强烈的吸收,且其吸收强度随波长增加呈指数衰减,干扰模式复杂。另外,水体中的悬浮颗粒物所引起的浊度效应不仅会造成光散射损失,其本身也可能产生吸收,尤其是在较短波长处。这些干扰物的浓度在不同海域、不同季节、不同水层中变化很大,使得建立一个普适性强、抗干扰能力高的检测模型变得异常困难^[6-7]。

为解决上述挑战,本文摒弃了传统的全光谱扫描思路,转而探索一种基于有限最优特征波长的检测新范式。其核心科学假设是:通过精心筛选一组数量有限但信息量最大的特定波长,并结合先进的人工智能(Artificial Intelligence, AI)学习算法进行信号解析,可以在不损失精度的前提下,极大地简化光学系统结构。这种“有限波长”策略允许使用多个独立的窄带紫外发光二极管(Light Emitting Diode, LED)作为光源和光电二极管作为探测器,替代笨重的宽谱光源与光谱仪,从而为传感器微型化、低功耗化和坚固化设计开辟一条全新的技术路径。

为此,本文开展面向小型化传感器的海水硝酸盐紫外检测波长优化与仿真研究。本文建立一套高保真的数字仿真系统,精确模拟硝酸盐紫外吸收、关键干扰物效应及仪器噪声;重点在于利用随机森林等机器学习算法,从全光谱数据中智能筛选出针对性强、抗干扰能力高的最优有限波长组合;最终目标是为下一代小型化、低功耗海水硝酸盐原位传感器的光学系统设计与智能解算算法提供理论依据和优化方案。

1 仿真系统的理论基础与模型构建

1.1 核心理论基础

构建一个可信的海水硝酸盐紫外光谱检测仿真系统,必须建立在坚实的物理化学原理之上。仿真系统的核心是光的吸收定律。朗伯-比尔定律描述了单色光通过均匀吸收介质时,吸光度与吸光物质浓度及光程之间的线性关系^[8],其数学表达式如下。

$$A = \varepsilon(\lambda) \cdot c \cdot l \quad (1)$$

式中, A 为吸光度,是一个无量纲量; $\varepsilon(\lambda)$ 为

波长 λ 处的摩尔吸光系数; c 为吸光物质的浓度; l 为光通过溶液的光程长度。对于硝酸盐而言,其摩尔吸光系数 $\varepsilon(\lambda)$ 随波长变化而剧烈变化,在 210 nm 附近达到最大值,之后随着波长向长波方向移动而逐渐降低,至 300 nm 时已降至较低水平。本仿真系统精确地采用了文献中报道的硝酸盐在不同波长下的标准 ε 值,为后续的光谱模拟提供了准确的基准。对于微型化传感器,每个所选波长必须承载最大化、最独特的浓度信息或干扰校正信息。其反演问题可表述为寻找一个函数 f , 使得 $c = f(A_{\lambda_1}, A_{\lambda_2}, \dots, A_{\lambda_n})$ 。该函数 f 通常是非线性的,这正是引入机器学习算法的原因。

然而,海水是一个多组分共存的复杂体系。总吸光度并非仅来源于硝酸盐,而是所有吸光组分贡献的线性叠加^[9]。

$$A_{\text{total}}(\lambda) = A_{\text{NO}_3}(\lambda) + A_{\text{Br}}(\lambda) + A_{\text{CDOM}}(\lambda) + A_{\text{turb}}(\lambda) + \dots \quad (2)$$

式中, $A_{\text{total}}(\lambda)$ 为海水在波长 λ 处的总吸光度; $A_{\text{NO}_3}(\lambda)$ 为硝酸盐在波长 λ 处的吸光度; $A_{\text{Br}}(\lambda)$ 为溴离子在波长 λ 处的吸光度; $A_{\text{CDOM}}(\lambda)$ 为有色溶解有机物(Colored Dissolved Organic Matter, CDOM)在波长 λ 处的吸光度; $A_{\text{turb}}(\lambda)$ 为海水浊度相关的吸光度。

因此,精确模拟干扰物的吸收光谱是提高仿真逼真度的关键。对于溴离子的吸收,仿真模型考虑到了其与温度和盐度的依赖关系。溴离子的吸收系数可以表述为与温度相关的线性项与指数衰减项的乘积,并乘以盐度的归一化因子^[8]。这种建模方式能够反映在不同温盐条件下溴离子干扰强度的变化,更贴近海洋环境的实际情况^[9-10]。有色溶解有机物的吸收光谱采用指数衰减模型,其表达式如下。

$$A_{\text{CDOM}}(\lambda) = a \cdot \exp[-S \cdot (\lambda - \lambda_0)] \quad (3)$$

式中,参数 a 与 CDOM 的浓度成正比; S 为表征光谱斜率的参数; λ_0 为参考波长。该模型很好地描述了 CDOM 吸收随波长增加而迅速减弱的特性。

浊度的影响,模型将其简化为由散射和吸收共同造成的吸光度增加,并假设其与浊度值成正比,且与波长的负幂次方相关^[11-12]。

$$A_{\text{turb}}(\lambda) \propto \lambda^{-n} \quad (4)$$

1.2 仪器噪声

模拟在获得了由硝酸盐和各类干扰物构成的“理想”吸光度矩阵后，仿真系统引入了仪器噪声模块以模拟真实检测器的性能局限。采用加性高斯白噪声来模拟电子学噪声和光子散粒噪声的综合效应。噪声的强度与信号本身的强度相关联，并通过用户设定的信噪比 (Signal-to-Noise Ratio, SNR) 进行控制。首先计算每个波长下总吸光度信号的均方根 (Root Mean Square, RMS)，然后根据目标信噪比计算出对应的噪声标准差 σ_{noise} 。

$$\sigma_{\text{noise}} = \frac{R_{\text{signal}}}{10^{A/20}} \quad (5)$$

式中， R_{signal} 为总吸光度信号的 RMS； A 为 SNR。最后将符合该标准差的高斯分布随机数叠加到纯净信号上。

1.3 仿真实现方法

(1) 数据生成

采用分层抽样的策略来生成硝酸盐浓度数据。考虑到自然海水中硝酸盐浓度分布的不均匀性，仿真程序生成了 5 000 个样本，其中 40% 的样本浓度分布在 0.5~20 $\mu\text{mol/L}$ 的低浓度区间，以代表寡营养海域；50% 的样本分布在 20~1 000 $\mu\text{mol/L}$ 的中等浓度区间，代表大部分沿岸和上升流区域；剩余 10% 的样本则分布在 1 000~2 000 $\mu\text{mol/L}$ 的高浓度区间，用以模拟受人类活动影响严重或某些特殊生化过程导致的富营养化水体。这种浓度分布设计确保了所建立的校正模型能够在较宽的动态范围内保持良好的预测能力，避免了对某一浓度区间的过拟合。所有浓度数据在经过物理模型转换为光谱数据后，还会加入前述的干扰和噪声，最终形成一个包含光谱矩阵和浓度向量的完整仿真数据集，为下一阶段的波长优化算法提供输入^⑥。环境参数 (温度、盐度) 参考东海、南海典型海域观测值 (温度 10~30 $^{\circ}\text{C}$ 、盐度 30~35 psu)，通过 Python 生成溴离子 (与盐度线性关联：盐度每增加 1 psu，溴离子浓度增加 0.06 mmol/L)、CDOM (光谱斜率 $S = 0.018 \text{ nm}^{-1}$) 及浊度的浓度矩阵，确保干扰物分布符合海洋实际情况。

(2) 保真度验证设计

为确保模型能复现真实海水紫外吸收特性，本

文设计了可落地的验证方案：①数据输入接口：模型支持导入外部实测光谱数据 (如航次观测数据、实验室配制标准海水样本光谱)，需将实测数据整理为“样本数 \times 波长数”的二维数组格式，并同步输入实测样本的温度、盐度等环境参数；②相关性验证流程：基于实测环境参数生成对应仿真光谱后，计算各波长下实测光谱与仿真光谱的皮尔逊相关系数，以“平均相关系数 ≥ 0.95 ”作为模型保真度达标阈值；③验证结果记录：代码预留验证结果输出接口，可将相关系数、实测-仿真偏差等指标自动写入分析报告，为后续实验验证提供数据记录支持。

2 基于随机森林的有限波长优化与性能评估方法

2.1 随机森林算法原理及其适用性分析

随机森林^[13-15]是由多棵决策树构成的集成模型，其核心思想是“集思广益”，通过构建并组合多棵弱学习器 (决策树) 来形成一个强学习器。其工作流程主要分为两步：一是采用自助采样法训练单棵决策树；二是聚合所有树的预测结果以输出最终结论。

2.1.1 决策树构建

对于给定的训练数据集 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ ，其中 x_i 为第 i 个样本的光谱特征向量， y_i 为其硝酸盐浓度。随机森林采用自助采样法为每棵决策树生成一个训练子集 D_i ，即从原始数据集 D 中有放回地随机抽取 n 个样本，每个样本未被抽中的概率为 $(1 - 1/n)^n \approx e^{-1} \approx 36.8\%$ ，这部分数据被称为袋外数据，可用于模型验证。

在决策树的每个节点进行分裂时，算法并非从所有 m 个特征中寻找最佳分裂特征，而是随机选择 m_{try} 个特征构成一个特征子集，然后从这个子集中选择最优分裂点。这一步引入了随机性，确保了森林中树木的多样性，有效防止过拟合。

2.1.2 预测聚合

在回归问题中，随机森林的最终预测值 \hat{y} 是所有 T 棵决策树预测值的简单平均。

$$\hat{y} = \frac{1}{T} \sum_{i=1}^T f_i(x) \quad (6)$$

式中, $f_t(x)$ 为第 t 棵决策树对于输入特征向量 x 的预测输出。随机森林算法对于每棵树, 首先计算其袋外数据的预测误差, 记为基准误差; 然后随机打乱袋外数据中某个特定特征的数值, 破坏该特征与目标值之间的关联后再次计算预测误差。将该特征在所有树中打乱前后误差增加量的平均值作为其重要性得分, 得分越高表明该特征对预测硝酸盐浓度的贡献越大。因此, 随机森林算法能够定量评估每个特征的重要性, 这是其被选作波长筛选算法的重要原因。

同时, 通过自助采样和特征随机选择引入的随机性, 使模型具备强大的抗过拟合能力, 即使数据特征较多也能保持泛化性能。此外, 其均值聚合策略对异常值和测量噪声不敏感, 增强了模型的稳健性, 加之其训练效率远高于深度学习等复杂模型, 使其非常适用于仿真研究中需要快速迭代和验证的场景。

2.2 两阶段波长优化策略

获得高保真的仿真光谱数据集后, 本文的核心任务是从数十个候选波长中, 智能地筛选出数量有限但信息量最大的最优波长子集。这一过程被称为特征选择, 在多元校正中至关重要。本文设计并实现了一种结合了物理先验知识与数据驱动算法的两阶段波长优化策略。

第一阶段为基于物理知识的强制选择。根据硝酸盐的紫外吸收光谱特性, 其吸收峰值位于 210 nm 至 230 nm 之间, 这一波段包含了最丰富的硝酸盐浓度信息。因此, 算法首先将搜索范围锁定在这一“特征峰区”。它计算该区间内每个波长处的吸光度与硝酸盐真实浓度之间的相关系数绝对值, 然后选取相关性最高的两个波长作为必选波长。这一步骤保证了模型的基础灵敏度, 确保能够捕捉到硝酸盐最本质的光谱特征。这是一种将领域知识嵌入算法流程的有效做法, 避免了纯数据驱动方法可能漏掉关键物理特征的风险。

第二阶段为基于统计学习的优化扩充。在确保了核心特征波长之后, 算法转向更长波长区域。这些波长区域硝酸盐的直接吸收已经较弱, 但它们可能包含了关于干扰物的宝贵信息。算法采用随机森林这一强大的集成学习算法作为评估工具。初始模

型仅使用第一阶段选出的两个特征峰波长进行训练, 并记录其基线性能。随后, 算法逐一评估将其其他候选波长加入模型后带来的性能提升。只有当一个新波长的加入能够显著提升模型的预测精度时, 该波长才会被纳入最终的特征集。这一过程持续进行, 直到达到预设的最大波长数量。通过这种方式, 算法能够自动识别出那些虽不直接反映硝酸盐浓度, 但能有效校正背景干扰的补偿波长。

2.3 模型性能评估体系

本仿真系统采用了多种统计指标, 从不同角度综合评价模型的优劣。

决定系数 R^2 是最常用的指标, 它反映了模型预测值能够解释真实值方差的比例, 其值越接近 1, 说明模型的拟合优度越高。

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2} \quad (7)$$

式中, y_i 为真实浓度; \hat{y}_i 为预测浓度; \bar{y}_i 为平均浓度。

残差的标准差 σ_{res} 直观地表示了预测值围绕真实值的波动范围, 是衡量测量精度的直接指标。

$$\sigma_{\text{res}} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{(n - k)}} \quad (8)$$

式中, n 为样本数; k 为模型参数个数。 σ_{res} 越小, 检测精度越高。

平均绝对误差 (Mean Absolute Error, MAE) 则给出了误差的平均幅度, 对异常值不如均方误差敏感, 更能反映模型的典型表现。

$$E_{\text{MAE}} = \sum |y_i - \hat{y}_i| / n \quad (9)$$

式中, E_{MAE} 为平均绝对误差的表征参数; n 为样本数。更重要的是, 由于仿真数据覆盖了从 0.5~2 000 $\mu\text{mol/L}$ 低到高的宽浓度范围, 本文对模型进行了分段性能评估, 即分别计算低浓度区间和高浓度区间的 R^2 和残差标准差。这有助于检验模型在不同浓度水平下的稳定性和可靠性, 避免出现整体性能良好, 局部预测偏差大的问题。所有这些评估均在模型未见过的测试集上进行, 以确保评估结果的客观性和泛化能力。

3 仿真结果与分析

波长相关性分析 (图 1) 揭示了不同波段所承

载的信息差异。正如理论基础所预测，在 210 nm 至 230 nm 的硝酸盐特征吸收峰区内，吸光度与浓度的相关系数普遍高达 0.85 以上，尤其在 215 nm 和 225 nm 附近出现了两个峰值，印证了将这些波长作为模型基石的合理性。超出此范围，相关系数随波长增加而逐渐下降，但在 250 nm 至 300 nm 区

间，仍维持在一定水平 (0.6~0.7)。随机森林算法的筛选过程表明，从这些长波区域中精心选择的少数波长，如 254 nm、275 nm 和 300 nm，虽然其单独携带的硝酸盐信息有限，但当它们与特征峰波长组合在一起时，能够显著提升模型的整体性能。

基于上述三个优选波长 (229 nm、230 nm、

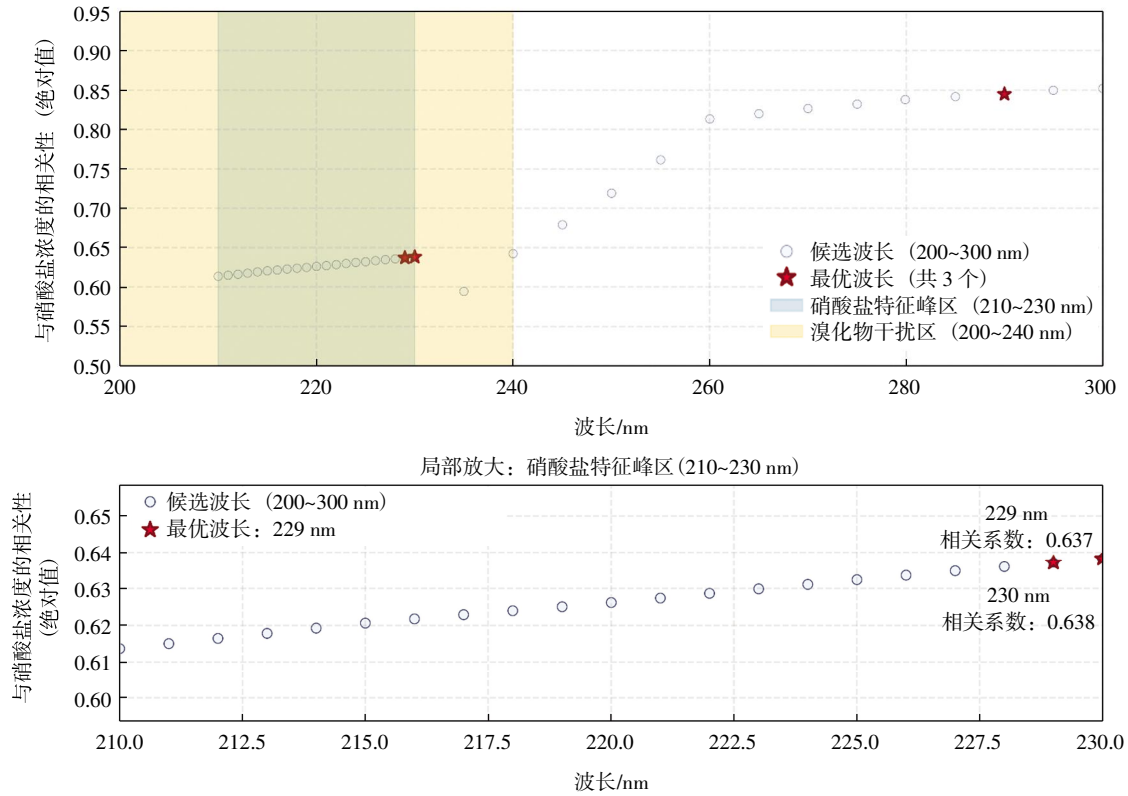


图 1 波长相关性分布 (最佳 $R^2 = 0.966$)

290 nm) 所建立的随机森林回归模型，在测试集上表现出了优异的整体预测线性度，模型的决定系数 R^2 达到了 0.965。然而，对预测结果进行分层解析

揭示了一个关键特征，模型性能呈现出显著的浓度依赖性 (图 2)。

在浓度低于 $20 \mu\text{mol/L}$ 的清洁水体中，模型展

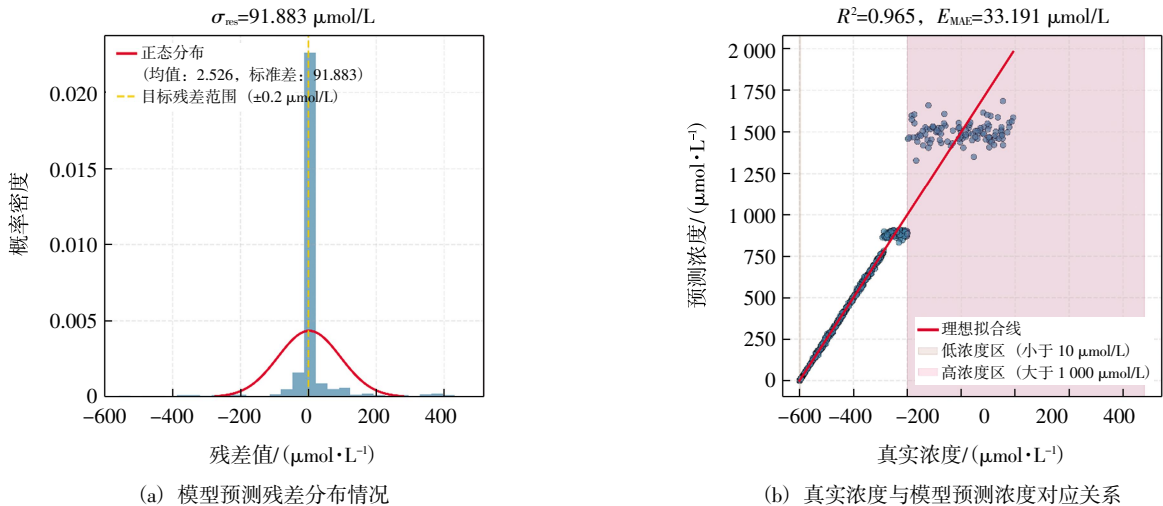


图 2 模型性能分析 (最优波长 229 nm、230 nm、290 nm)

现了近乎完美的拟合优度, R^2 值高达 0.998, 这表明在该浓度区间内, 三波长组合能够极其精确地分离出硝酸盐信号, 为解决紫外光谱法在寡营养海域的应用难点提供了卓越方案。与之形成鲜明对比的是, 在高于 $1\ 000\ \mu\text{mol/L}$ 的高浓度区间, 模型预测能力出现显著下降, R^2 值为 -0.005 , 表明模型在此区间未能建立起有效的预测关系。这一现象清晰界定了该波长组合方案的优势应用边界, 即其特别适用于中低浓度水体的高精度检测。虽然高浓度区间 ($1\ 000\sim 2\ 000\ \mu\text{mol/L}$) 模型 R^2 很小, 但该区间对应河口、养殖区等受人类活动影响显著的海域, 纳入样本可确保模型对全海洋环境的覆盖性, 为后续高浓度校正算法开发提供数据基础。实际应用中通过浓度阈值判断, 当检测值大于 $1\ 000\ \mu\text{mol/L}$ 时, 自动切换至“高浓度补偿模式”, 后续研究将针对性地优化该区间算法。

模型的整体残差标准差为 $91.883\ \mu\text{mol/L}$, 平均绝对误差为 $33.191\ \mu\text{mol/L}$ 。这一残差水平需结合其宽广的检测范围 ($0.5\sim 2\ 000\ \mu\text{mol/L}$) 来评估, 它反映了模型在高低浓度区间性能不均衡的综合结果, 进一步强调了将该方案应用于其优势浓度区间的必要性。

为了评估仪器硬件性能对检测结果的影响, 本文进行了信噪比敏感性分析 (图 3)。仿真结果表明, 模型性能随信噪比提升而改善, 并存在一个关键的性能阈值。当信噪比低于 $14\ \text{dB}$ 时, 模型的决定系数 R^2 无法达到 0.9 的最低目标要求。随着信噪比提高, 模型性能稳步提升。分析确认, 当信噪比达到 $14\ \text{dB}$ 后, 模型整体 R^2 即可稳定在 0.9 以上, 满足预设的检测精度标准。在实际传感器设计中, 只需将光学和电子系统的信噪比设计在 $14\ \text{dB}$ 这一达标门槛之上, 即可在保证核心检测精度的同时, 最优地平衡系统的成本、复杂度与功耗, 这对于资

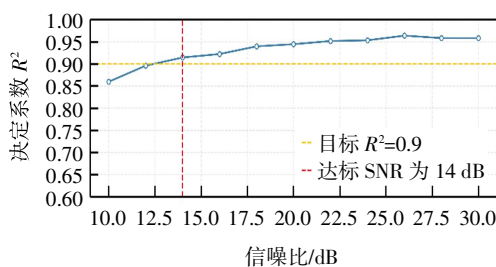
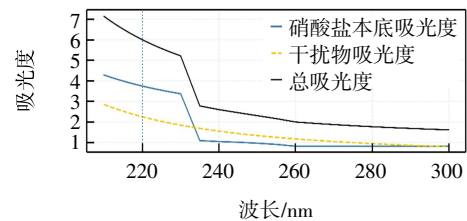


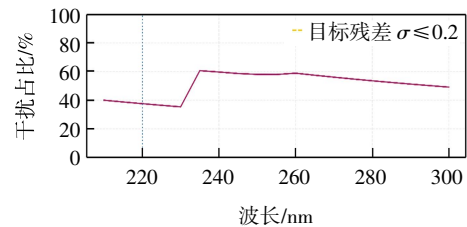
图 3 信噪比对硝酸盐浓度反演模型决定系数的影响

源严格受限的小型化传感器设计至关重要。

此外, 仿真结果还定量分析了各干扰物的贡献比例 (图 4)。在 $220\ \text{nm}$ 处, 溴离子干扰可占总吸光度的 30% 以上, 而在 $250\ \text{nm}$ 以后, CDOM 则成为主要干扰源。这一定量分析为理解复杂背景下的信号构成提供了直观依据。



(a) 各波长吸光度组成



(b) 干扰物吸光度占比

图 4 干扰因素对各波长的影响

4 讨论与展望

本文所证实的多波长优化策略, 相较于传统的单波长或双波长法, 在抗复杂背景干扰方面展现出巨大优势。双波长法通常选择一个测量波长和一个参比波长, 通过差值法来抵消基线漂移。这种方法在干扰物光谱形状固定且简单时较为有效。但对于海水这种包含多种光谱特性各异干扰物的复杂体系, 简单的双波长扣除效果有限。而本文采用的多波长模型, 实质上构建了一个“数字滤波器”, 它能够同时利用多个波长点的信息, 更精确地估计出干扰背景的整体轮廓, 并进行针对性更强的减法运算。

从工程实现的角度看, 仿真结果直接转化为具体的设计参数。首先, 最优波长组合 ($229\ \text{nm}$ 、 $230\ \text{nm}$ 、 $290\ \text{nm}$) 为光源的选择提供了明确目标。这意味着可以选用发射波长与此匹配的紫外发光 LED, 而不是使用宽谱光源加单色仪的笨重方案, 这极有利于传感器的小型化、低功耗和坚固性设计。整体光学模块体积控制在 $20\ \text{mm} \times 15\ \text{mm} \times$

10 mm。采用低功耗微控制单元 (Microcontroller Unit, MCU) (如 STM32L4 系列) 实现信号采集与算法运算, 集成电源管理模块 (休眠电流 $\leq 10 \mu\text{A}$), 通过脉冲驱动 LED 减少能耗。采用防水密封设计 (IP68 等级), 光学窗口选用石英玻璃 (紫外透射率 $\geq 90\%$), 整体传感器重量控制在 50 g 以内, 适配浮标、剖面仪等原位观测平台。集成温盐传感器实时采集环境参数, 建立模型系数动态校正表, 通过线性插值调整算法参数, 适应不同海域环境变化。用低功耗蓝牙或 LoRa 模块实现数据传输。

尽管本仿真系统已经较为完善, 但仍有进一步拓展的空间。未来的研究工作可以从以下几个方向展开: 其一, 将更多的潜在干扰物纳入模型, 如亚硝酸盐、硫化物等, 使仿真环境更接近真实海水的极端复杂性; 其二, 开发自适应校准算法, 使模型能够根据实时测量的环境参数 (如温度、盐度) 或光谱特征, 动态微调模型系数, 从而适应不同海区、不同季节的水体性质变化, 提升传感器的长期稳定性和普适性; 其三, 将先进的机器学习算法, 如深度学习网络, 引入到光谱解析中, 探索其从复杂光谱中提取微弱特征的能力, 或许能进一步挖掘紫外光谱法的潜力。最后, 将本仿真系统的结论应用于实际传感器的研制, 并通过大规模现场比测试验进行验证与反馈, 形成仿真指导设计到实验验证仿真的闭环优化流程, 将是推动该技术走向成熟应用的必由之路。

参考文献:

- [1] 董晶, 唐军武, 吴国俊. 紫外光谱法海水硝酸盐原位检测技术研究进展[J]. 海洋技术学报, 2022, 41(5): 19-34.
- [2] BEATON A D, WADHAM J L, HAWKINGS J, et al. High-resolution in situ measurement of nitrate in runoff from the Greenland Ice Sheet[J]. Environmental Science & Technology, 2017, 51(21): 12518-12527.
- [3] BLAEN P J, KHAMIS K, LLOYD C E M, et al. Real-time monitoring of nutrients and dissolved organic matter in rivers: Capturing event dynamics, technological opportunities and future directions[J]. Science of the Total Environment, 2016, 569: 647-660.
- [4] SINGH P, SINGH M K, BEG Y R, et al. A review on spectroscopic methods for determination of nitrite and nitrate in environmental samples[J]. Talanta, 2019, 191: 364-381.
- [5] LANGERGRABER G, FLEISCHMANN N, HOFSTAEDTER F. A multivariate calibration procedure for UV/VIS spectrometric quantification of organic matter and nitrate in wastewater[J]. Water Science and Technology, 2003, 47(2): 63-71.
- [6] JOHNSON K S, COLETTI L J. In situ ultraviolet spectrophotometry for high resolution and long-term monitoring of nitrate, bromide and bisulfide in the ocean[J]. Deep Sea Research Part I: Oceanographic Research Papers, 2002, 49(7): 1291-1305.
- [7] ZHU X, YU K, ZHU X, et al. An improved algorithm for measuring nitrate concentrations in seawater based on deep-ultraviolet spectrophotometry: A case study of the Aoshan Bay seawater and Western Pacific seawater[J]. Sensors, 2021, 21(3): 965.

5 结 论

本文成功构建了一个高保真的海水硝酸盐紫外光谱检测仿真系统, 该系统综合集成了硝酸盐吸收、多种干扰物影响及仪器噪声等关键物理过程。通过将物理先验与数据驱动算法相结合的两阶段波长优化策略, 系统地筛选出了由 229 nm、230 nm、290 nm 构成的极具简洁性的最优三波长组合。基于该组合建立的随机森林校正模型, 在宽浓度范围内表现出了优异的整体预测线性度, 其决定系数 R^2 达到 0.965。深入分析表明, 该模型在低浓度区间 (小于等于 $20 \mu\text{mol/L}$) 具有极高的精度 ($R^2 = 0.998$), 明确了其在大洋寡营养盐区等低浓度水域应用的突出优势。研究同时量化了仪器信噪比对性能的影响, 确定了 14 dB 为实现目标检测精度 ($R^2 \geq 0.9$) 的关键设计门槛。

综上所述, 本文不仅深化了对紫外光谱法检测海水硝酸盐过程的理解, 更通过有限波长策略为传感器小型化与低功耗设计提供了创新路径, 为开发面向特定应用场景的高性能海水硝酸盐原位传感器奠定了坚实的理论基础, 对推动我国海洋环境监测技术的进步具有重要的理论意义和实际应用价值。

- [8] SAKAMOTO C M, JOHNSON K S, COLETTI L J. Improved algorithm for the computation of nitrate concentrations in seawater using an in situ ultraviolet spectrophotometer[J]. *Limnology and Oceanography: Methods*, 2009, 7(1): 132–143.
- [9] SAKAMOTO C M, JOHNSON K S, COLETTI L J, et al. Pressure correction for the computation of nitrate concentrations in seawater using an in situ ultraviolet spectrophotometer[J]. *Limnology and Oceanography: Methods*, 2017, 15(10): 897–902.
- [10] BLEYEN N, ALBRECHT A, DE CANNIÈRE P, et al. Non-destructive on-line and long-term monitoring of in situ nitrate and nitrite reactivity in a clay environment at increasing turbidity[J]. *Applied Geochemistry*, 2019, 100: 131–142.
- [11] 陈颖, 何磊, 崔行宁, 等. 基于紫外光谱的水体硝酸盐氮浊度补偿研究[J]. *光谱学与光谱分析*, 2020, 40(2): 472–477.
- [12] 陈晓伟, 殷高方, 赵南京, 等. 浊度干扰下硝酸盐浓度紫外导数光谱检测方法研究[J]. *光谱学与光谱分析*, 2019, 39(9): 2912–2916.
- [13] BREIMAN L. Random forests[J]. *Machine Learning*, 2001, 45(1): 5–32.
- [14] BREITENBACH M, NIELSEN R, GRUDIC G Z. Probabilistic random forests: Predicting data point specific misclassification probabilities[R]. Boulder: University of Colorado, 2003.
- [15] LIN Y, JEON Y. Random forests and adaptive nearest neighbors[J]. *Journal of the American Statistical Association*, 2006, 101(474): 578–590.

UV Detection Wavelength Optimization and Simulation for Miniaturized In-Situ Seawater Nitrate Sensors

MIAO Xin, GUAN Yinlong, LI Cong, CAO Shouqi

(College of Information Technology, Shanghai Ocean University, Shanghai 201306, China)

Abstract: To address the urgent demand for miniaturized and low-power in situ seawater nitrate sensors in the marine monitoring industry, this study tackles the core challenge that traditional broadband UV spectral-scanning approaches struggle to simultaneously achieve compactness and high detection accuracy. We develop a high-fidelity simulation model that integrates nitrate absorption, bromide interference effects, and instrument noise, and propose a “finite optimal wavelength” detection strategy. By combining this strategy with a random forest algorithm, we enable intelligent wavelength selection and nonlinear modeling. Based on simulation results, the wavelength combination of 229 nm, 230 nm, and 290 nm was selected. The resulting model achieved an R^2 of 0.965 across a concentration range of 0.5–1 000 $\mu\text{mol/L}$, with a residual relative error of approximately 2.87%. An instrument signal-to-noise ratio (SNR) of 14 dB was identified as the critical design threshold. This strategy significantly reduces the complexity of the light source and spectral components, providing a solid foundation for compact, low-power sensors based on UV LEDs. It is particularly well suited for detecting low nitrate concentrations in oligotrophic marine environments.

Keywords: nitrate detection; ultraviolet spectroscopy; sensor miniaturization; limited wavelengths; Random Forest; low-power consumption