

DOI:10.13869/j.cnki.rswc.2024.05.035.

嵇晶晶, 白立影, 余冬立, 等. 基于 GF-1 卫星遥感反演排水河沟水体溶存 N_2O 浓度模型对比研究[J]. 水土保持研究, 2024, 31(5): 257-264.

Ji Jingjing, Bai Liying, She Dongli, et al. Comparative Study on Retrieval Model of Dissolved N_2O Concentration in Drainage Ditch Based on GF-1 Satellite Remote Sensing[J]. Research of Soil and Water Conservation, 2024, 31(5): 257-264.

基于 GF-1 卫星遥感反演排水河沟水体溶存 N_2O 浓度模型对比研究

嵇晶晶¹, 白立影², 余冬立¹, 管伟³, 阿力木·阿布来提¹, 潘永春¹

(1.河海大学 农业科学与工程学院, 南京 211100;

2.江苏省水文水资源勘测局镇江分局, 江苏 镇江 212008; 3.江苏省临海农场有限公司, 江苏 射阳 224353)

摘要: [目的]探究利用 GF-1 卫星数据反演水体溶存氧化亚氮(N_2O)浓度的可行性,为实现低成本、高效率的水质实时监测提供有效途径。[方法]以宁夏青铜峡灌区第 1 和第 5 排水河沟为研究对象,选取与排水河沟水体溶存 N_2O 浓度相关性高的 GF-1 卫星影像波段反射率和水质参数作为自变量,通过最优子集筛选法确定最优自变量组合,分别构建了多元线性回归、BP 神经网络和支持向量机模型,对水体溶存 N_2O 浓度进行预测对比。[结果]水温(T)、溶解性有机碳(DOC)等是影响水体溶存 N_2O 浓度的主要因素,同时近红外(NIR)等卫星波段与水体溶存 N_2O 浓度变化趋势显著相关。当自变量包括 T , NIR 等 7 个因素时,模型预测效果最佳。在 3 种模型中, BP 神经网络模型验证结果 R^2 为 0.64,具有最高预测精度。[结论]GF-1 卫星数据以及水质参数与水体溶存 N_2O 浓度存在复杂的相关性关系,且 BP 神经网络能够实现利用 GF-1 卫星数据较高精度地反演水体溶存 N_2O 浓度。

关键词: N_2O 浓度; GF-1 卫星; 反演; 机器学习

中图分类号: X87

文献标识码: A

文章编号: 1005-3409(2024)05-0257-08

Comparative Study on Retrieval Model of Dissolved N_2O Concentration in Drainage Ditch Based on GF-1 Satellite Remote Sensing

Ji Jingjing¹, Bai Liying², She Dongli¹, Guan Wei³, Ali mu · Abu laiti¹, Pan Yongchun¹

(1.College of Agricultural Science and Engineering, Hohai University, Nanjing 211100,

China; 2.Zhenjiang Branch of Jiangsu Provincial Hydrology and Water Resources Survey Bureau,

Zhenjiang, Jiangsu 21208, China; 3.Jiangsu Linhai Farm Co., Ltd., Sheyang, Jiangsu 224353, China)

Abstract: [Objective] The aims of this study are to explore the feasibility of using GF-1 satellite data to retrieve the concentration of dissolved nitrous oxide (N_2O) in water, so as to provide an effective way to realize low-cost and high-efficiency real-time monitoring of water quality. [Methods] The 1st and 5th drainage ditches in Qingtongxia Irrigation District of Ningxia were selected as the research objects, and the reflectance and water quality parameters of GF-1 satellite image band, which were highly correlated with the concentration of dissolved N_2O in the drainage ditches, were selected as independent variables, and the optimal combination of independent variables was determined by optimal subset screening method. Multiple linear regression, BP neural network and support vector machine models were respectively constructed to predict and compare the concentration of dissolved N_2O in water. [Results] The water temperature (T) and dissolved organic carbon (DOC) were the main factors affecting the concentration of dissolved N_2O in water, and satellite bands such as near infrared (NIR) were significantly correlated with the variation trend of

收稿日期: 2023-09-14

修回日期: 2023-10-20

资助项目: 国家自然科学基金“海涂盐碱地农业汇水区反硝化和厌氧氨氧化脱氮机制与过程模拟”(42177393)

第一作者: 嵇晶晶(2000—), 女, 江苏宿迁人, 硕士研究生, 主要从事农业水土工程研究。E-mail: jijingjing836@163.com

通信作者: 余冬立(1980—), 男, 湖南长沙人, 博士, 教授, 博士生导师, 主要从事农业生态与水土过程模拟研究。E-mail: shedongli@hhu.edu.cn

<http://stbcyj.paperonce.org>

dissolved N_2O concentration in water. When the independent variable including 7 factors such as T and NIR, the model had the best prediction effect. Among the three models, the R^2 of BP neural network model was 0.64, which had the highest prediction accuracy. [Conclusion] There is a complex correlation between GF-1 satellite data and water quality parameters and dissolved N_2O concentration in water bodies, and BP neural network can use GF-1 satellite data to retrieve dissolved N_2O concentration in water bodies with high accuracy.

Keywords: N_2O concentration; GF-1 satellite; inversion; machine learning

随着全球变暖不断加剧,温室气体排放研究已成为热点问题,其中氧化亚氮(N_2O)对温室效应的贡献率约为 6%^[1]。河沟是一个庞大的水体和溶解物质的运输网络,由于农业施肥和污水排放,大量氮素进入河流,导致水体中氮素含量不断增加^[2]。目前的研究主要通过在水域内布设大量监测点,利用气相色谱法^[3],结合 N_2O 排放扩散模型和相关因素,来估算水体溶存 N_2O 浓度和排放通量。然而,对于较大面积的水域,传统的人工监测方法很难考虑到多种因素对水域 N_2O 排放的影响,并且需要大量的人力和物力投入。因此,研究一种广泛适用、监测便利且成本低廉的 N_2O 监测方法变得尤为重要。

遥感技术能够利用不同物体对电磁波的反射特性,远距离获取地表信息,实现对水体水质参数的定性和定量分析。根据前人研究^[4-5],多光谱技术广泛应用于水质监测,包括利用 GF-1 卫星数据对水质的时空变化进行监测^[6-7],并取得了良好的效果。鉴于内陆水体环境的复杂性,遥感获取数据难以直接评估水质参数,而 BP 神经网络、支持向量机等机器学习算法可以通过自动学习优化,探索这些参数与遥感数据的潜在联系,从而有效地反演或评价这些参数^[8-10]。Xiao 等^[11]利用环境卫星 HJ1 A/1 B CCD 多光谱数据,分别建立了多元线性回归模型和神经网络模型,对水体总氮(Total Nitrogen, TN)和总磷(Total Phosphorous, TP)进行反演,通过对比分析反演结果,发现神经网络模型能够较好地反演水质参数的分布,并具有较强的通用性。刘梅等^[12]利用 GaiaSky-mini 高光谱数据,分别采用线性模型和偏最小二乘法(Partial Least Squares, PLS)、径向基函数(Radial Basis Function, RBF)机器学习算法,以单波段和波段指数对可溶性固体(Total Soluble Solid, TSS)、TN 等水质参数进行反演。研究结果表明,使用波段指数作为自变量能够更好地反演水质参数。然而,以往的研究主要侧重于对水中叶绿素 a 、悬浮物、有色可溶性有机物以及少数非光敏参数的反演研究,而对非光敏特性的水体溶存气体 N_2O 浓度的研究较少。此外,大多数研究直接利用遥感数据对水质参数进行反演,未充分利用不同数据源的信息以提高

模型精度。综上,本研究以宁夏青铜峡灌区第 1 和第 5 排水河沟作为研究对象,结合 GF-1 卫星数据和相对容易测定的水质参数进行分析,通过最优子集筛选最优自变量组合,以预测河沟水体的溶存 N_2O 浓度,并探究使用 GF-1 卫星反演水体溶存 N_2O 浓度的可行性,以期为实现低成本、高效率监测水体溶存 N_2O 浓度提供科学依据。

1 材料与方法

1.1 研究区概况

青铜峡灌区($37^{\circ}49' - 39^{\circ}23' \text{N}$, $105^{\circ}37' - 106^{\circ}39' \text{E}$)位于黄河冲积平原,黄河贯穿其中,灌区涉及青铜峡市、银川市等 4 个县(市)的引黄灌溉部分,占宁夏回族自治区土地总面积的 53%^[13]。本研究以青铜峡灌区第 1 排水河沟和第 5 排水河沟为研究对象,两条排水河沟贯穿整个灌区,如图 1 所示。第 1 排水河沟和第 5 排水河沟分别位于研究区北部和南部,第 1 排水河沟主要位于银川市,流经青铜峡市和永宁县境内,总长度约 26.4 km;第 5 排水河沟主要位于石嘴山市,流经贺兰县、平罗县和惠农县,总长约 68.7 km。

1.2 地面数据采集

本研究采用定点监测方法获取了青铜峡灌区第 1 和第 5 排水河沟的地面数据。在第 1 和第 5 排水河沟分别布置了 12, 13 个监测点(图 1)。于 2021 年 7 月 15 日、9 月 25 日、10 月 11 日和 11 月 28 日,以及 2022 年 5 月 8 日和 7 月 14 日分别采集了各监测点的水样,共计 150 个样品,其中包括 100 个 2021 年水样和 50 个 2022 年水样。对水样进行了分析化验,测定了水体硝态氮(nitrate nitrogen, $\text{NO}_3^- - \text{N}$)、水温(T)、有机碳(dissolved organic carbon, DOC)、电导率(electrical Conductivity, EC)、pH 以及水体溶存 N_2O 浓度。使用无菌采样瓶采集了表层水样(0~0.5 m),并将其存储于低温环境中,随后立即带回实验室,使用 $0.7 \mu\text{m}$ 孔径的 Whatman GF/F 滤膜进行过滤;水体的 DOC 浓度采用有机碳分析仪(Shimadzu TOC-V CPN, 日本岛津)测定;水体的 $\text{NO}_3^- - \text{N}$ 浓度采用连续流动式分析仪(SKALAR SAN++, Skalar, 荷兰)测定,使用温度计现场计对 T 进行了测定,水体 EC

和 pH 使用便携式水质测定仪 (SL1000, HACH, 美国) 测定。

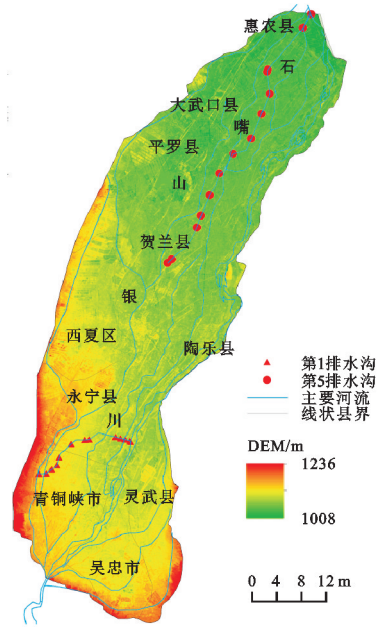


图1 研究区概况

Fig. 1 Overview map of the study area

水体溶存 N_2O 浓度 (C_{obs}) 的测定使用配备电子捕获检测器 (ECD) 的气相色谱仪 (Agilent-7890B) 测定^[14]。为此, 使用了两根型号分别为 $1\text{ m} \times 2\text{ mm}$ 和 $3\text{ m} \times 2\text{ mm}$ 填充有 Poropak Q (80/100) 的色谱柱。ECD 温度为 $330\text{ }^\circ\text{C}$, 柱温度为 $55\text{ }^\circ\text{C}$ 。ECD 中使用高纯度 Ar 和 CH_4 ($95.0\%\text{ Ar}$, $5.0\%\text{ CH}_4$) 作为载气 (35 cm , 3 min)。进样时使用装有三通阀的针筒 (5 ml), 每次进样 5 ml , 计算公式^[15]为:

$$C_{obs} = C_0 \left(\frac{\beta RT}{22.4} + \frac{V_0}{V_1} \right) \quad (1)$$

式中: C_0 为注入氦气的平衡器内气—水平衡后测定的气相 N_2O 浓度 ($\mu\text{mol/L}$); V_0 为平衡器室的气体体积 (ml); V_1 为水样体积 (ml); 22.4 是气体摩尔体积 (mol/L); R 为理想气体状态常数 [$0.082\text{ L} \cdot \text{atm}/(\text{mol} \cdot \text{L})$]; T 为平衡过程中的实际水温 (K); β 为用体积分子表示的气体溶解热 [$\text{L}/(\text{L} \cdot \text{atm})$]。根据 Henry 定律, β 值可由以下公式得出:

$$\beta = 2.4 \times 10^{-6} \times K_H \times P_g \quad (2)$$

式中: P_g 为 N_2O 的标准大气压; K_H 为标准大气压条件下的气体 Henry 常数 [$\mu\text{mol}/(\text{L} \cdot \text{atm})$], K_H 的值可由以下公式得出:

$$K_H = K_H^\theta \times e^{K_T \left(\frac{1}{T_K} - \frac{1}{T_\theta} \right)} \quad (3)$$

式中: K_H^θ 为标准大气压条件下的 Henry 常数 [$24\text{ }714\text{ } \mu\text{mol}/(\text{m}^3 \cdot \text{Pa})$]; K_T 表示 Henry 常数随温度而变化的参数, N_2O 取值为 $1\text{ }750$; T_K 为实际绝对温度 (K); $T_\theta = 298\text{ K}$ 。

1.3 卫星数据及其处理

本研究使用了中国资源卫星应用中心 (<https://data.cresda.cn>) 提供的 GF-1 卫星遥感影像作为数据源, 该影像的分辨率为 16 m , 由蓝色 (Blue, B1)、绿色 (Green, B2)、红色 (Red, B3) 和近红外 (Near IR, NIR) 4 个波段组成^[16]。卫星成像时间与实测日期同步。在 ENVI 5.3 环境下对光谱数据进行大气校正和辐射定标, 并利用 ArcGIS 10.4 提取监测点波段信息。

1.4 最优子集筛选

最优子集筛选是一种基于预测决定系数 R^2 的方法, 用于筛选最优自变量组合^[17]。具体步骤如下: 首先设定自变量个数 N 为 $2 \sim 8$, 对所有自变量进行排列组合, 得到不同数量和不同组合的自变量子集。基于各自变量子集建立回归模型, 然后根据回归模型的预测 R^2 , PRESS (预测误差平方和)、Mallows-Cp (马洛斯 C_p 值) 和 S (模型响应度) 的结果对回归模型进行评价。其中, 预测 R^2 反映了模型对新数据的预测能力, PRESS 表示拟合值与观测值之间的误差大小, Mallows-Cp 描述了模型的精确度和偏倚程度^[18], 而 S 反映了数据值与拟合值之间的离散程度。最后, 选择综合表现最优的自变量子集作为输入模型的最优自变量组合。这种方法既考虑了自变量个数对模型复杂度和过拟合风险的影响, 又考虑了多种评价指标对模型性能和稳健性的影响, 从而有效提高模型预测精度。

1.5 模型构建

1.5.1 模型构建方法

(1) 多元线性回归。多元线性回归模型是处理多个自变量与因变量之间的线性关系的常用方法, 其数学模型为

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_N X_N + \epsilon$$

式中: Y_i 为因变量; X_1, X_2, \dots, X_N 为 N 个自变量; $\beta_0, \beta_1, \dots, \beta_N$ 为方程的回归常数; ϵ 为随机误差, 主要是用来解释随机变量引起 Y_i 变化的部分。

(2) BP 神经网络。BP (Back Propagation) 神经网络是一种利用梯度下降法来最小化误差的多层前馈神经网络^[19]。它由输入层 (Input layer)、隐含层 (Hidden layer) 和输出层 (Output layer) 构成, 并通过反向传播算法来调整网络参数。本研究首先将给定的训练样本自动划分为训练集、测试集和验证集。然后使用 mapminmax 函数对所有样本的输入值进行归一化处理, 使其落在 $[0, 1]$ 区间内。利用训练集对 BP 神经网络进行训练, 并使用验证集来监测模型的泛化能力。当模型达到预设的精度或者达到最大迭代次数时, 停止训练, 并利用测试集来评估模型的预测性能。在输出预测值时, 使用 mapminmax 函数对输出值进行反归一化处理, 并与实测值进行比较。如

果预测误差(Error)过大,则使用反向传播算法对网络参数进行修正,使预测值更贴近实测值。BP 神经网络的原理示意图如图 2 所示。设置最小误差为 $1\text{E}-6$ 。

隐含层节点数是神经网络模型性能的重要参数之一。如果节点数过少,BP 神经网络可能过于简单,导致数据欠拟合。然而,如果节点数过多,则可能会发生过拟合,导致测试数据集误差增加。除了隐含层节点数以外,学习率也是影响模型性能的重要参数之一。学习率过大可能会使函数直接越过误差最小点,进而导致误差增大;如果学习率过小,就会大大增加训练时间,并且容易导致函数陷入局部最小值。

(3) 支持向量机。SVR 支持向量机回归是一种基于 SVM(Support Vector Machine, SVM)支持向量机的回归方法,通过在特征空间中构造最大间隔超平面来拟合数据,并引入容忍偏差 ε [20],允许部分数据点超出间隔带,并定义这些数据点(即支持向量)与间隔带边界的距离为残差 ξ 。如图 3 所示,SVR 的目

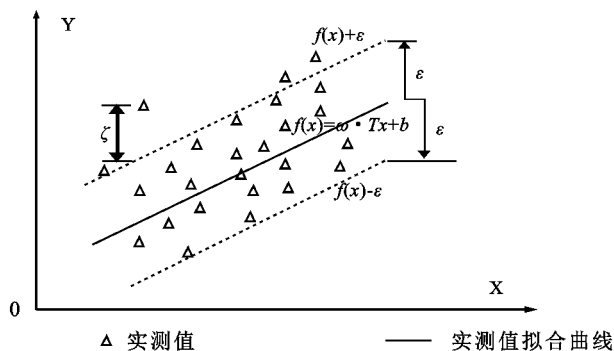


图 3 SVR 原理图

Fig. 3 SVR schematic diagram

1.5.2 模型精度评价指标 本研究采用决定系数 R^2 和均方误差 RMSE 作为模型精度的评价指标。决定系数 R^2 反映了模型对数据变异的解释程度,而均方误差 RMSE 描述了模型预测值与实测值之间的偏差大小。通常情况下, R^2 越大, RMSE 越小,表明模型拟合效果越好,预测能力越强。

2 结果与分析

2.1 水体溶存 N_2O 浓度与各因子间相关性分析

通过对选定的排水河沟水体理化指标和 N_2O 溶存浓度进行监测,结果如表 1 所示。排水河沟中各水质参数的最大值和最小值相差较大,其中 $\text{NO}_3^- - \text{N}$, DOC, T 和 EC 的变异系数(CV)在 $10.0\% \sim 100.0\%$,属于中等变异水平,表明这些参数受季节影响较大,而 pH 的变异系数最低,较为稳定。此外,水体溶存 N_2O 浓度变化范围为 $0.01 \sim 2.42 \mu\text{mol/L}$,平均值为 $0.49 \mu\text{mol/L}$,变异系数为 8.0% 。

以 GF-1 卫星遥感影像 4 个波段反射率和归一化水体指数(normalized Difference Water Index,

标是找到一个使残差最小化的超平面,并对间隔带之外的数据点进行回归。为提高模型的性能、稳定性以及训练精度,SVR 需要优化惩罚系数和核函数参数 gamma 以控制模型的复杂度和泛化能力,同时还需要对输入数据进行归一化处理,使其在 $[0,1]$ 区间内分布。

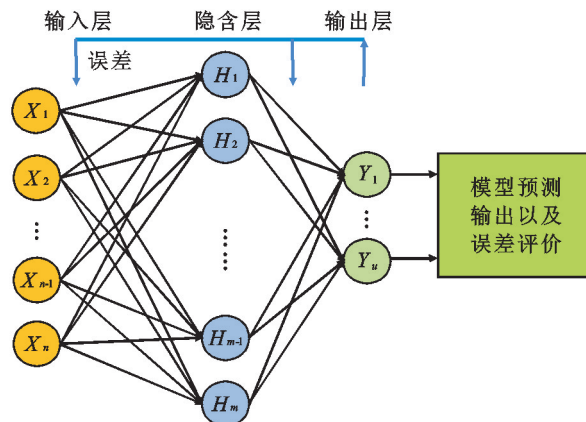
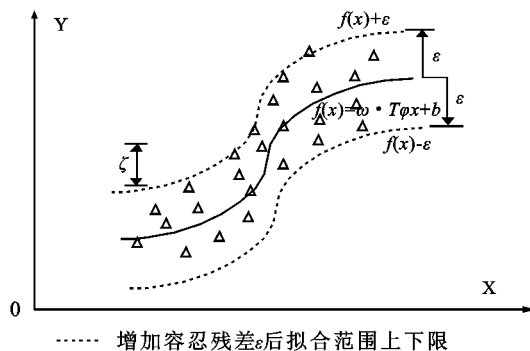


图 2 BP 神经网络原理图

Fig. 2 BP neural network schematic diagram



NDWI)以及沟道水体水质观测数据作为自变量,进行了水体溶存 N_2O 浓度的 Pearson 相关性分析,水体溶存 N_2O 浓度与 T 呈显著正相关关系($p < 0.01$),而与水体 DOC 浓度呈显著负相关关系($p < 0.05$)。水体溶存 N_2O 浓度与 GF-1 卫星遥感影像 4 个波段反射率(B1, B2, B3 和 NIR)呈显著正相关关系($p < 0.01$),而与 NDWI 呈显著负相关关系($p < 0.01$)。

2.2 基于最优集筛选的最优自变量组合

基于河沟地面监测数据样本,采用 minitab 18 软件对遥感波段参数和河沟水体水质参数进行最优自变量组合筛选,并计算相应的预测 R^2 , PRESS, Mallows-Cp 和 S 等评价指标(表 2)。随着 N 增加,预测 R^2 呈上升趋势,但增长速率逐渐减缓,表明模型对数据的拟合程度随着 N 的增加而提高,但增长幅度逐渐降低。PRESS 呈下降趋势,但下降速率也逐渐减缓,表明模型对新数据的预测误差随着 N 的增加而减小,但减大幅度逐渐降低。Mallows-Cp 随着 N 的增加而接近于 N 的值,但在 $N = 8$ 时比 $N = 7$

时略有上升;S 随着 N 的增加而下降,但在 $N=7$ 和 $N=8$ 时保持不变。综合考虑各个评价指标,可以得出结论:当 N 较小时,增加自变量可以显著改善模型精度和稳健性;当 N 达到一定水平($N=7$)时,模型精度和稳健性趋于稳定;当 N 过大时,模型可能出现过拟合现象。因此,在本研究中,最优的自变量组合包含 T^{**} 、 DOC^* 、 pH 、 B_2^{**} 、 B_3^{**} 、 NIR^{**} 和 NDWI^{**} 因子,其中带有 * 号的为显著影响因子。

表 1 光谱波段以及地面参数分析以及与水体溶存 N₂O 浓度的 Pearson 相关系数

Table 1 Analysis of spectral bands and ground parameters and Pearson correlation coefficient with dissolved N₂O concentration in water

| 变量 类型 | 参数 名称 | 平均值 | 变异 系数/% | 水质因子 | | | | | | 波段反射率 | | | | |
|----------|---------------------------------|---------|------------|---------|---------------------------------|---------|-------|-------|-------|----------------|----------------|----------------|---------|------|
| | | | | Cobs | NO ₃ ⁻ -N | T | DOC | EC | pH | B ₁ | B ₂ | B ₃ | NIR | NDWI |
| 水质因子 | Cobs | 0.49 | 8.0 | 1 | | | | | | | | | | |
| | NO ₃ ⁻ -N | 2.27 | 98.0 | 0.16 | 1 | | | | | | | | | |
| | T | 15.75 | 47.4 | 0.34** | -0.16 | 1 | | | | | | | | |
| | DOC | 4.47 | 53.1 | -0.21* | 0.29** | 0.01 | 1 | | | | | | | |
| | EC | 1256.16 | 32.2 | 0.08 | 0.30** | -0.03 | 0.23* | 1 | | | | | | |
| | pH | 8.31 | 3.0 | 0.16 | -0.17* | 0.27** | 0.18* | 0.20* | 1 | | | | | |
| 波段反射率 | B ₁ | 0.11 | 20.4 | 0.23** | -0.08 | 0.42** | 0.06 | 0.03 | 0.17 | 1 | | | | |
| | B ₂ | 0.24 | 32.7 | 0.23** | -0.08 | 0.38** | 0.07 | 0.03 | 0.08 | 0.98** | 1 | | | |
| | B ₃ | 0.25 | 36.2 | 0.25** | -0.07 | 0.21* | 0.06 | -0.01 | -0.01 | 0.90** | 0.95** | 1 | | |
| | NIR | 0.27 | 44.3 | 0.44** | -0.68** | 0.72** | -0.06 | -0.04 | 0.20* | 0.63** | 0.62** | 0.48** | 1 | |
| | NDWI | -0.06 | 23.7 | -0.30** | 0.01 | -0.54** | 0.15 | -0.10 | 0.13 | 0.13 | 0.17 | 0.30* | -0.66** | 1 |

注:**表示达到 0.01 的显著水平;*表示达到 0.05 显著水平。

表 2 基于最优子集筛选的最优自变量子集

Table 2 Optimal independent variable subset based on optimal subset screening

| 变量数 N | 最优自变量子集 | 预测 R^2 | PRESS | Mallows-Cp | S |
|---------|---|----------|-------|------------|------|
| 2 | T^{**} 、 B_4^{**} | 21.07 | 39.96 | 39.12 | 0.54 |
| 3 | B_2^{**} 、 B_3^{**} 、 B_4^{**} | 25.71 | 38.65 | 30.11 | 0.53 |
| 4 | T^{**} 、 B_2^{**} 、 B_3^{**} 、 B_4^{**} | 30.12 | 37.00 | 21.71 | 0.51 |
| 5 | DOC^* 、 T^{**} 、 B_2^{**} 、 B_3^{**} 、 B_4^{**} | 33.33 | 35.38 | 15.94 | 0.50 |
| 6 | DOC^* 、 T^{**} 、 B_2^{**} 、 B_3^{**} 、 B_4^{**} 、 NDWI^{**} | 36.01 | 33.91 | 11.35 | 0.49 |
| 7 | T^{**} 、 DOC^* 、 pH 、 B_2^{**} 、 B_3^{**} 、 B_4^{**} 、 NDWI^{**} | 37.76 | 32.99 | 8.76 | 0.48 |
| 8 | $\text{NO}_3^- \text{-N}$ 、 T^{**} 、 DOC^* 、 pH 、 B_2^{**} 、 B_3^{**} 、 B_4^{**} 、 NDWI^{**} | 38.23 | 32.95 | 8.80 | 0.48 |

注:**表示达到 0.01 的显著水平;*表示达到 0.05 显著水平。

2.3 水体溶存 N₂O 浓度反演模型评价

在建立模型之前,由于神经网络参数较多,采用 Matlab 2022b 选择最佳隐含层节点数。以 1~20 步长为 1 的值,随机选择 50 组随机初始权重和阈值,建立了 9 000 个模型,以探究不同隐含层节点数对模型性能的影响。同样地,我们将学习率设置为 0.01~0.1步长为 0.01 的值,建立了 4 500 个模型,以选择最佳学习率,结果见图 4。随着隐含层节点数和学习率的改变,RMSE 也在不断变化。当隐含层节点数为 9 时,RMSE 明显低于其他节点数的误差;同样地,当学习率为 0.02 时,RMSE 达到最小值。因此,将 BP 神经网络的隐含层节点数设置为 9,学习率设置为 0.02。随后,基于最优自变量组合,分别构建多元线性回归、BP 神经网络和支持向量机模型,以水体溶存

N₂O 浓度为因变量,采用 2021 年数据进行训练验证,并以 2022 年数据作为测试集。图 5A 显示了 3 种模型在训练集和测试集上的 R^2 值,可以看出,BP 神经网络和支持向量机模型在训练集上的 R^2 值均超过 0.61,明显优于多元线性回归模型 R^2 值(0.41)。在测试集上,BP 神经网络模型的 R^2 值为 0.64,高于支持向量机模型的 0.52,而多元线性回归模型的 R^2 值仅为 0.39。图 5B 显示了 3 种模型在训练集和测试集上的 RMSE 值。可以看出,多元线性回归在两个数据集上的 RMSE 值均最大,而 BP 神经网络在训练集上的 RMSE 值为 0.32,明显低于支持向量机模型(0.39)。在测试集上,BP 神经网络模型的 RMSE 值为 0.37,低于支持向量机的 0.43,表明 BP 神经网络模型具有更高预测精度。综合考虑 R^2 和 RMSE

两个指标, BP 神经网络模型更适合反演河沟水体溶存 N_2O 浓度的变化。

3 讨论

通过对基于 GF-1 卫星数据及河沟水体水质参数的全子集进行筛选, 得到了一个包括 T^{**} 、 B_2^{**} 、 B_3^{**} 等 7 个变量的最优自变量组合。随着自变量的增加, 预测 R^2 逐渐提高, 但当自变量达到一定阈值后, 增加幅度变缓, 这表明最优子集筛选的效率减弱^[21]。张智韬等^[22]的相关研究通过对最优子集筛选获得了最佳自变量组

合, 并利用筛选后的变量进行建模, 结果明显优于未筛选的模型。本研究使用 BP 神经网络、多元线性回归模型以及支持向量机回归模型, 来预测 2022 年河沟水体溶存 N_2O 浓度。多元线性回归模型通常只能处理线性问题, 而水中溶存 N_2O 浓度与其影响因素之间通常存在非线性关系^[23], 而支持向量机需要使用核函数将数据映射到高维空间^[24]。当出现水体溶存 N_2O 浓度数据样本线性不可分或核函数选择不当时, 可能导致支持向量机模型无法找到合适的超平面来分割数据, 从而造成较差的预测效果。

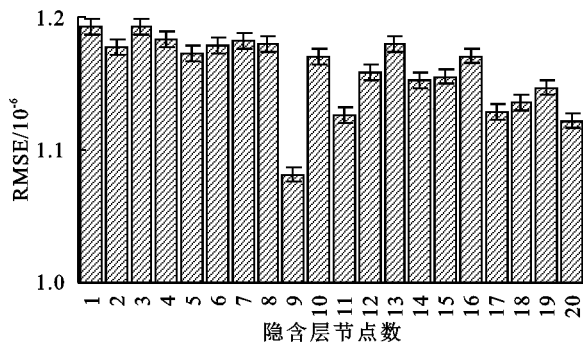


图 4 RMSE 随隐含层节点数和学习率变化情况

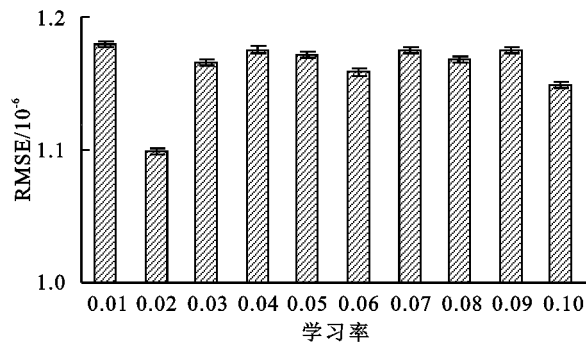


Fig. 4 RMSE changes with the number of nodes and learning rate of the hidden layer

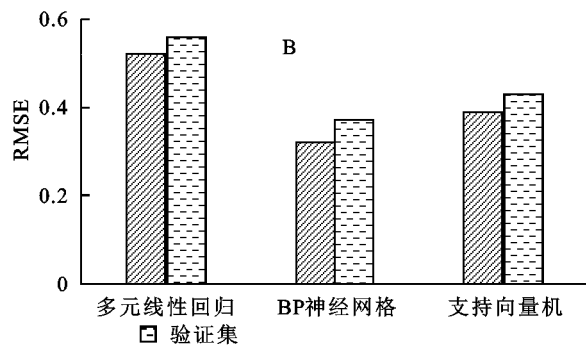
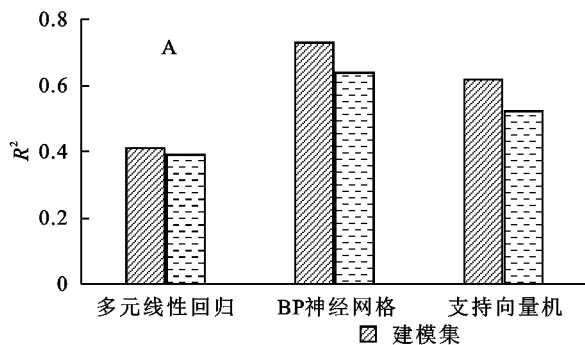


图 5 河沟水体溶存 N_2O 浓度反演模型对比

Fig. 5 Comparison of inversion models of dissolved N_2O concentration in drainage ditch water

与其他方法相比, BP 神经网络具备更强的适应性和泛化能力, 能够灵活地拟合何种复杂的非线性函数, 而且无需进行特征转换或降维操作^[25]。与多元线性回归和支持向量机模型相比, BP 神经网络能够更准确地预测水体中 N_2O 的变化情况, 因此被认为是提高水质监测准确性的强大工具^[26]。然而, 在使用 BP 神经网络模型进行预测时, 如果水体环境发生突发的排污事件, 可能会导致影响水体溶存 N_2O 浓度的因素发生变化, 从而使预测结果偏离正常值。因此, 建议在 BP 神经网络模型中增加更多样本以提升模型的拟合能力, 并对模型进行优化以提高精度。

4 结论

(1) 水体溶存 N_2O 浓度与 DOC 呈现显著的负相关关系、而与 T 和 $\text{NO}_3^- - \text{N}$ 呈现显著的正相关关

系, 卫星遥感数据中的各波段都与水体溶存 N_2O 呈现极显著正相关关系, 其中 NIR 相关性最强。

(2) 通过最优子集筛选法确定了不同数量的最优子集, 比较各个指标后发现, 当自变量数量为 7 时 (T 、DOC、pH、 B_2 、 B_3 、NIR、NDWI), 可以得到最优的自变量组合, 避免过度拟合并获得最佳的预测效果。

(3) 由模型的反演结果可知, 模型对水中溶存 N_2O 浓度反演能力由强到弱依次为: BP 神经网络、支持向量机、多元线性回归模型。与多元线性回归和支持向量机模型相比, BP 神经网络模型对水体溶存 N_2O 浓度的预测更可靠, 预测结果更贴合实际情况。

参考文献 (References):

- [1] 谭永洁, 王东启, 周立旻, 等. 河流氧化亚氮产生和排放研究综述[J]. 地球与环境, 2015, 43(1): 123-132.
Tan Y J, Wang D Q, Zhou L M, et al. Review on the

- production and emission of nitrous oxide from rivers[J]. *Earth and Environment*, 2015,43(1):123-132.
- [2] 王安吉,马文林,吴建繁,等.2015—2019 年北京市农业氧化亚氮排放变化与分析[J].*中国农业资源与区划*, 2023,44(1):47-56.
- Wang A J, Ma W L, Wu J F, et al. Changes and analysis of agricultural nitrous oxide emission in Beijing from 2015 to 2019 [J]. *Chinese Journal of Agricultural Resources and Regional Planning*, 2023,44(1):47-56.
- [3] 赵松,侯毅凯,张安兵,等.基于无人机多光谱影像的滏阳河水质遥感反演[J].*测绘与空间地理信息*, 2022,45(3):81-85.
- Zhao S, Hou Y K, Zhang A B, et al. Remote sensing inversion of water quality of Fuyang River based on UAV multi-spectral images[J]. *Geomatics & Spatial Information Technology*, 2022,45(3):81-85.
- [4] 赵琳.基于长时序多光谱遥感技术的白洋淀水质演变监测分析[J].*现代信息科技*, 2023,7(5):141-144.
- Zhao L. Monitoring and analysis of water quality evolution in Baiyang lake based on long time series multispectral remote sensing technology[J]. *Modern Information Technology*, 2023,7(5):141-144.
- [5] 黄昕晰,应晗婷,夏凯,等.基于无人机多光谱影像和 OPT-MPP 算法的水质参数反演[J].*环境科学*, 2020,41(8):3591-3600.
- Huang X X, Ying H T, Xia K, et al. Inversion of water quality parameters based on UAV multispectral images and the OPT-MPP algorithm [J]. *Environmental Science*, 2020,41(8):3591-3600.
- [6] 祁兰兰,王金亮,农兰萍,等.基于 GF-1 卫星数据的洱海干季水质时空变化监测[J].*人民长江*, 2021,52(9):24-31.
- Qi L L, Wang J L, Nong L P, et al. Temporal and spatial monitoring on water quality of Erhai lake in dry season based on GF-1 satellite data[J]. *Yangtze River*, 2021,52(9):24-31.
- [7] 彭保发,陈哲夫,李建辉,等.基于 GF-1 影像的洞庭湖区水体水质遥感监测[J].*地理研究*, 2018,37(9):1683-1691.
- Peng B F, Chen Z F, Li J H, et al. Monitoring water quality of Dongting lake region based on GF-1 image[J]. *Geographical Research*, 2018,37(9):1683-1691.
- [8] Lin S, Novitski L N, Qi J, et al. Landsat TM/ETM+ and machine-learning algorithms for limnological studies and algal bloom management of inland lakes[J]. *Journal of Applied Remote Sensing*, 2018,12(2):1-17.
- [9] Chang N B, Xuan Z M, Yang Y J. Exploring spatiotemporal patterns of phosphorus concentrations in a coastal bay with MODIS images and machine learning models [J]. *Remote Sensing of Environment*, 2013,134:100-110.
- [10] Zheng Z B, Li Y M, Guo Y L, et al. Landsat-based long-term monitoring of total suspended matter concentration pattern change in the wet season for Dongting lake, China[J]. *Remote Sensing*, 2015,7:13975-13999.
- [11] Xiao X, Xu J, Hu C F, et al. Assessment of water quality in natural river based on HJ1A/1B CCD multispectral remote sensing data [C] // *Multispectral, Hyperspectral, and Ultraspectral Remote Sensing Technology, Techniques and Applications V*. Beijing, China. Spie, 2014,9263:270-283.
- [12] 刘梅,马启良,原居林,等.基于无人机高光谱遥感技术对内陆养殖池塘水质监测的研究[J].*海洋与湖沼*, 2022,53(1):195-205.
- Liu M, Ma Q L, Yuan J L, et al. Water quality monitoring of inland aquaculture ponds based on UAV hyperspectral remote sensing technology[J]. *Oceanologia Et Limnologia*, 2022,53(1):195-205.
- [13] 魏艳芳.青铜峡灌区排水沟农业非点源污染物分布特征及贡献率研究[D].银川:宁夏大学,2019.
- Wei Y F. Study on conventional pollution transport in drainage ditch of Qingtongxia irrigation area and its contribution rate to drainage ditch [D]. Yinchuan: Ningxia University, 2019.
- [14] Xia Y Q, Li Y F, Ti C P, et al. Is indirect N_2O emission a significant contributor to the agricultural greenhouse gas budget: A case study of a rice paddy-dominated agricultural watershed in Eastern China [J]. *Atmospheric Environment*, 2013,77:943-950.
- [15] Miyashiro R, Takano Y. Mixed integer second-order cone programming formulations for variable selection in linear regression[J]. *European Journal of Operational Research*, 2015,247(3):721-731.
- [16] Liu G, Wu M Q, Niu Z, et al. Investigation method for crop area using remote sensing sampling based on GF-1 satellite data[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2015,31(5):160-166.
- [17] 姚一飞,王爽,张珺锐,等.基于 GF-1 卫星遥感的河套灌区土壤含水率反演模型研究[J].*农业机械学报*, 2022,53(9):239-251.
- Yao Y F, Wang S, Zhang J R, et al. Inversion model of soil moisture in Hetao irrigation district based on GF-1 satellite remote sensing[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022,53(9):239-251.
- [18] Qin X H, Zheng L L. Mathematical models for analy-

- sis of water pollution[J]. *Applied Mechanics and Materials*, 2012, 209: 1941-1947.
- [19] Liu K L, Lin T, Zhong T T, et al. New methods based on a genetic algorithm back propagation(GABP) neural network and general regression neural network (GRNN) for predicting the occurrence of trihalomethanes in tap water[J]. *Science of the Total Environment*, 2023, 870: 161976.
- [20] 冯汉中, 杨淑群, 刘波. 支持向量机(SVM)方法在气象预报中的个例试验[J]. *四川气象*, 2005, 25(2): 9-12.
Feng H Z, Yang S Q, Liu B. A case study of support vector machine(SVM) method in meteorological forecast[J]. *Journal of Sichuan Meteorology*, 2005, 25(2): 9-12.
- [21] Okcu D, Pektas A O, Uyumaz A. Creating a non-linear total sediment load formula using polynomial best subset regression model[J]. *Journal of Hydrology*, 2016, 539: 662-673.
- [22] 张智韬, 韩佳, 王新涛, 等. 基于全子集-分位数回归的土壤含盐量反演研究[J]. *农业机械学报*, 2019, 50(10): 142-152.
Zhang Z T, Han J, Wang X T, et al. Soil salinity inversion based on best subsets-quantile regression model[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2019, 50(10): 142-152.
- [23] 祝令亚. 湖泊水质遥感监测与评价方法研究[D]. 北京: 中国科学院遥感应用研究所, 2006.
Zhu L Y. Remote sensing monitoring and assessment of water quality for lakes[D]. Beijing: Institute of Remote Sensing Applications, Chinese Academy of Sciences, 2006.
- [24] 徐红敏. 基于支持向量机理论的水环境质量预测与评价方法研究[D]. 长春: 吉林大学, 2007.
Xu H M. Study on the prediction and assessment methods of water environment quality based on support vector machines theory[D]. Changchun: Jilin University, 2007.
- [25] 徐敏. 基于复杂性理论的河湖环境系统模型研究[D]. 长沙: 湖南大学, 2007.
Xu M. Research of water environmental system models based on complexity theory[D]. Changsha: Hunan University, 2007.
- [26] 王渤权, 金传鑫, 周论, 等. 基于长短期记忆网络的西丽水库水质预测[J]. *长江科学院院报*, 2023, 40(6): 64-70.
Wang B Q, Jin C X, Zhou L, et al. Water quality prediction for Xili reservoir based on long-short term memory[J]. *Journal of Changjiang River Scientific Research Institute*, 2023, 40(6): 64-70.
- ~~~~~
- (上接第 256 页)
- [24] Schaap M G, Leij F J, van Genuchten M T. Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions[J]. *Journal of Hydrology*, 2001, 251(3/4): 163-176.
- [25] 罗雅曦, 刘任涛, 张静, 等. 腾格里沙漠草方格固沙林土壤颗粒组成、分形维数及其对土壤性质的影响[J]. *应用生态学报*, 2019, 30(2): 525-535.
Luo Y X, Liu R T, Zhang J, et al. Soil particle composition, fractal dimension and their effects on soil properties following sand-binding revegetation within straw checkerboard in Tengger Desert, China[J]. *Chinese Journal of Applied Ecology*, 2019, 30(2): 525-535.
- [26] 单桂梅, 张春平, 刘霞, 等. 沂蒙山区小流域坡耕地土壤颗粒结构与养分退化特征[J]. *中国水土保持科学*, 2013, 11(5): 76-82.
Shan G M, Zhang C P, Liu X, et al. Grain structure and nutrient degradation characteristics of soil from small watershed's sloping lands in Yimeng mountainous areas[J]. *Science of Soil and Water Conservation*, 2013, 11(5): 76-82.
- [27] Tyler S W, Wheatcraft S W. Application of fractal mathematics to soil water retention estimation[J]. *Soil Science Society of America Journal*, 1989, 53(4): 987-996.
- [28] 黄冠华, 詹卫华. 土壤颗粒的分形特征及其应用[J]. *土壤学报*, 2002, 39(4): 490-497.
Huang G H, Zhan W H. Fractal property of soil particle size distribution and its application[J]. *Acta Pedologica Sinica*, 2002, 39(4): 490-497.
- [29] 党亚爱, 李世清, 王国栋, 等. 黄土高原典型土壤剖面土壤颗粒组成分形特征[J]. *农业工程学报*, 2009, 25(9): 74-78.
Dang Y A, Li S Q, Wang G D, et al. Fractal characteristics of soil particle composition for typical types of soil profile on Loess Plateau[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2009, 25(9): 74-78.