

基于机器学习方法的悬浮泥沙遥感反演模型及敏感性分析

陈娜¹, 陈华¹, 李兰涛², 刘任莉¹, 钟毫忠³

(1.武汉大学 水资源工程与调度全国重点实验室, 湖北 武汉 430072; 2.水利部黄河水利委员会水文局, 河南 郑州 450004; 3.

中国电建集团成都勘测设计研究院有限公司, 四川 成都 610000)

摘要: 遥感反演技术为悬浮泥沙浓度 (SSC) 监测提供了高效手段, 但对于高浓度宽范围河流 SSC 的适用性亟需验证。

以黄河干流石嘴山水文站和吴堡水文站为测站, 构建了基于交叉验证递归特征消除-随机森林 (RFECV-RF) 的机器学习

模型, 通过哨兵 2 号卫星多光谱反射率信息遥感反演测站高浓度悬沙。研究结果表明, RFECV-RF 模型预测精度指标 R^2

大于 0.8, 总体上能够估算测站 (0-44.5) kg/m^3 区间的 SSC, 但对高值 SSC 的估算偏低; 高浓度宽范围 SSC 与光谱信息

呈现明显的非线性关系, 且可见光主导的特征对反演低浓度 SSC 较重要, 红光-短波红外光谱区间主导的特征对反演高

浓度 SSC 较重要; 在关键光谱特征中, B8A、B7/B5、B8-B11 波段光谱特征在 SSC (0-44.5) kg/m^3 区间保持灵敏, 而

B3/B8、B4/B8、B5/B6 随着 SSC 升高而渐趋饱和; 数据输入以及机器学习模型结构的不确定性是反演模型不确定性的

主要来源, 且当反演的 SSC 范围变宽时, 模型预测的 95%置信宽度增大且低值 SSC 的相对偏差也增大。综上, RFECV-

RF 模型可用于定量反演高浓度宽范围 SSC 并估算河段 SSC 空间分布, 为高含沙河流 SSC 自动监测方法提供技术参考。

关键词: 悬浮泥沙浓度; 遥感反演; RFECV; RF; 黄河; 光谱敏感性; 不确定性

中图分类号: TV123

引用格式: 陈娜, 陈华, 李兰涛, 刘任莉, 钟毫忠. 2025. 基于机器学习方法的悬浮泥沙遥感反演模型及敏感性分析. 遥感学报.

收稿日期: 2025-10-19; 预印本:

基金项目: 国家重点研发计划项目 (编号: 2023YFC3209101)

第一作者简介: 陈娜, 1994 年生, 女, 湖北黄石人, 博士研究生, 研究方向为智能监测、水沙数学模拟方面研究。E-mail: 2012301580256@whu.edu.cn

通信作者简介: 陈华, 1977 年生, 男, 教授, 博士生导师, 主要从事流域水文监测、模拟和智慧水利研究。E-mail: chua@whu.edu.cn

20 1 引言

21 河流悬浮泥沙反映河流泥沙输运强度,对河道演变、水库淤积、水生态健康及海岸带管理意义重大(Sun 等,
22 2020)。目前,水文测站采用的悬浮泥沙监测手段为传统的器测法,精度较高但耗时耗力,在高洪大沙的情况
23 下施测困难(应周箫等,2023)。遥感反演技术通过分析水体光谱特征与悬浮泥沙的关联性,为区域乃至全球
24 尺度的泥沙时空分布监测提供了高效手段(黄华等,2021;李子阳等,2023)。

25 目前遥感反演方法主要分为三类:理论分析方法、经验方法以及机器学习方法。理论分析法是基于光传输
26 理论,根据辐射传输方程将固有光学特征(吸收系数、散射系数)与表观光学特性(反射率、衰减系数)建立联
27 系,再推算 SSC。Yang 等(2017)采用 Sokoletsky-Aas 模型来模拟反射率,以红/绿波段比值为特征构建 SSC
28 ($<1\text{kg/m}^3$)反演模型,并结合固有光学特性与 SSC 的关系来验证反演模型精度。王雅萍等(2024)利用 Hydrolight
29 辐射传输模型,建立纯水、叶绿素 a、有色可溶性有机物(CDOM)和 SSC 四组分与固有光学特征的关系,进
30 而生成模拟反射率后再转换为卫星通道的反射率,构建出避免叶绿素 a、CDOM 干扰的 SSC ($<0.145\text{ kg/m}^3$)反
31 演模型。Jiang 等(2021)通过半分析模型分离散射与吸收系数,之后反演后向散射系数 b_{bp} ,再通过本地化校
32 准系数将 b_{bp} 转换为 SSC ($0.00009\text{-}1.4\text{ kg/m}^3$),解决 SSC 超出 1 kg/m^3 的光谱饱和问题。这种方法物理意义明
33 确,需要先获取固有光学特征数据或模拟计算光谱反射率,分析过程复杂。

34 经验方法建立光谱反射率或波段反射率组合与 SSC 之间的统计回归模型。潘洪洲等(2019)基于数据同化
35 思想,通过对 4 个指数模型的协同反演,提高珠江河口 SSC ($<1\text{ kg/m}^3$)的整体反演精度和稳定性。Marinho 等
36 (2021)耦合原位高光谱监测与 Sentinel-2 多光谱监测,通过最小二乘回归拟合实现对内格罗河极低 SSC (<0.015
37 kg/m^3)、高有色溶解物吸收情况下的 SSC 遥感反演,支撑河流泥沙动态分析。Yu 等(2023)结合 Landsat 系
38 列卫星影像,采用经验对数线性回归模型反演 1984-2021 黄河口年际 SSC ($<1\text{ kg/m}^3$)变化。而一些研究证实,
39 光谱反射与 SSC 之间存在明显的非线性关系(曾群柱等,1984),且随着 SSC 增大,反射率最强波段存在“红
40 移”现象(陈涛等,1994)。对于高浓度宽范围 SSC 的反演,由于不同浓度区间对 SSC 敏感的波段特征不相同,
41 难以采用固定的光谱特征和模式来反演河流 SSC,常常采用多种模式来提高反演精度。

42 机器学习方法挖掘光谱信息与 SSC 之间的非线性关系,而不局限于某一模式。学者采用随机森林(方馨蕊

等, 2019; Kwon 等, 2023) 和支持向量机 (Kwon 等, 2022a) 等机器学习算法挖掘光谱信息与 SSC 之间的非线性关系, 有效提升复杂光学环境下的反演精度, 但这些研究涉及的 SSC 均小于 0.7 kg/m^3 。Bazzett 和 Wang 等 (2024) 采用 5 种机器学习算法来构建室内烧杯中均匀含沙水样中 SSC ($<40 \text{ kg/m}^3$) 反演模型, 但该模型在天然高含沙河流的适用性还有待验证。鉴于在自然环境中, SSC 对光谱响应的复杂机制, Balasubramanian 等 (2020) 根据不同水色分类分别采用经验模型、物理模型、混合密度网络等多种方式, 将 SSC 动态范围 ($0\text{--}2.6 \text{ kg/m}^3$) 的反演模型应用于内陆和沿海水域, 但其 SSC 预测平均绝对误差达 1.8 kg/m^3 , 反演精度有待提升。

由于数据尺度、水体特性, 以上研究中除 Bazzett 和 Wang 等 (2024) 的室内研究, 大部分研究聚焦于 $\text{SSC} < 2.6 \text{ kg/m}^3$ 的含沙水体, 关于高浓度 ($>2.6 \text{ kg/m}^3$) 的 SSC 反演有待深入研究。在数据尺度方面, 方馨蕊等 (2019) 采用的是月尺度的 SSC、Yu 等 (2023) 分析的是年均 SSC, 均不能反映 SSC 时序过程中的量程变化。在水体特性方面, 在海洋和湖泊中的 SSC 通常较低, 而在内陆河流中在河流动力作用下 SSC 时空分布具有强非均匀性。在国内河流中, 根据 2024 年泥沙公报数据, 长江干流各水文站多年平均 SSC 为 0.392 kg/m^3 (大通) $\sim 1.44 \text{ kg/m}^3$ (向家坝), 而黄河干流各水文站多年平均 SSC 为 0.589 kg/m^3 (唐乃亥) $\sim 27.5 \text{ kg/m}^3$ (潼关); 在国外河流中, 内格罗河 (Marinho 等, 2021) 为亚马逊盆地典型的黑水河, 侵蚀速率低且泥沙来源单一, 导致泥沙总量少故而 SSC 极低。然而, 对于黄河这种宽范围 SSC 的高含沙水体, 通过遥感反演捕捉野外高含沙河流 SSC 时空变化, 可作为对传统监测方法的数据补充, 对于其流域泥沙管理和水资源调度而言十分必要 (厉玮和王向明, 2024)。

因此, 本研究采用基于 RFECV-RF 的机器学习方法构建反演模型, 以高含沙河流 (黄河干流) 为研究对象, 在可见光-短波红外波段范围内筛选出关键光谱特征, 采用组合特征构建 SSC 遥感反演模型, 并从光谱灵敏性、参数敏感性、模型不确定性和模型应用于河段空间 SSC 分布的合理性来探讨模型的适用性。

2 研究区域与数据

2.1 研究区域

本研究选取黄河干流上的石嘴山水文站 (以下简称石嘴山站) 和吴堡水文站 (以下简称吴堡站) 作为测站来构建遥感反演 SSC 模型, 并应用模型来反演河段 SSC 空间分布 (见图 1、表 1)。图 1 中的河段 1 和河段 2 反演区域分别位于黄河上游和中游, 该区域存在河流弯道和分岔, SSC 时空分布极不均匀, 点位监测具有局限

66 性。将遥感反演 SSC 模型运用到该河段，可提升对流域 SSC 时空预测的能力。



图 1 测站位置图

Fig.1 Location map of hydrological stations

表 1 水文站信息

Table 1 Information of hydrological stations

站点	观测时间	集水面积/km ²	等级	泥沙监测精度
石嘴山站	1951 年 9 月至今	309146	国家基本水文站	一类
吴堡站	1942 年 9 月至今	433514	国家基本水文站	一类

67 石嘴山站位于黄河上游干流，始建于 1951 年 9 月，处于黄河干流宁夏和内蒙古两自治区的交界河段，集水
68 面积 309146 km²，距离河口 2665 km，是国家基本站、一类精度泥沙站，也是黄河干流重要控制站和省界断面。
69 该站河床由砂质组成，泥沙主要来源于上游支流暴雨洪水以及青铜峡水库排沙（吴小爽和冯晓强，2024）。根
70 据 2024 年黄河泥沙公报，该站历史最大 SSC 为 22.4 kg/m³，历年平均 SSC 3.8 kg/m³。

71 吴堡站位于黄河中游干流，始建于 1942 年 9 月，流经广大的黄土高原地区，处于黄河粗泥沙的主要来源
72 区，集水面积 433514 km²，距离河口 1544 km，是黄河干流控制站、国家重要水文站、国家重点水情报汛站、一
73 类精度泥沙站。该站主槽为岩石，上面覆盖着砾石和泥沙，SSC 年内变化较大，汛期洪水暴涨暴落且 SSC 较高，
74 凌汛期洪水涨落缓慢且 SSC 较低（宋海松等，2019）。

81 测站的多光谱数据采用 Sentinel-2 号卫星在该测点的观测数据。河段 1 和河段 2 区域的空间多光谱影像数
82 据同样采用 Sentinel-2 号卫星观测数据，并基于改进归一化差异水体指数（MNDWI）来提取影像中的纯水体区
83 域（徐涵秋，2005）。Sentinel-2 号由 2A 和 2B 两颗卫星组成，所搭载的多光谱成像仪共有 13 个光谱带（表 2），
84 可提供 L2A 级表面反射率数据集。该数据集经过了正射校正和大气校正，可直接用作研究区域的水体表面反射
85 率，下载渠道为 GEE（<https://earthengine.google.com/>）。

86 表 2 Sentinel-2 号卫星多光谱信息

87 Table 2. Sentinel-2 satellite multispectral information

波段	Sentinel-2A	Sentinel-2B	空间分辨率/m	光谱区域
	中心波长(带宽)/nm	中心波长(带宽)/nm		
B1	443.9 (21)	442.3 (21)	60	超蓝(沿海和气溶胶)
B2	496.6 (66)	492.1 (66)	10	蓝光
B3	560.0 (36)	559.0 (36)	10	绿光
B4	664.5 (31)	665.0 (31)	10	红光
B5	703.9 (15)	703.8 (16)	20	红边 1
B6	740.2 (15)	739.1 (15)	20	红边 2
B7	782.5 (20)	779.7 (20)	20	红边 3
B8	835.1 (106)	833.0 (106)	10	近红外(NIR)
B8A	864.8 (21)	864.0 (22)	20	窄近红外(Narrow NIR)
B9	945.0 (20)	943.2 (21)	60	水蒸气
B10	1373.5 (31)	1376.9 (30)	60	短波红外(SWIR-卷云)
B11	1613.7 (91)	1610.4 (94)	20	短波红外(SWIR1)
B12	2202.4 (175)	2185.7 (185)	20	短波红外(SWIR2)

88 石嘴山站和吴堡站均为泥沙一类精度站，数据质量较高，收集 2020-2024 年经水文站整编的日均悬移质含

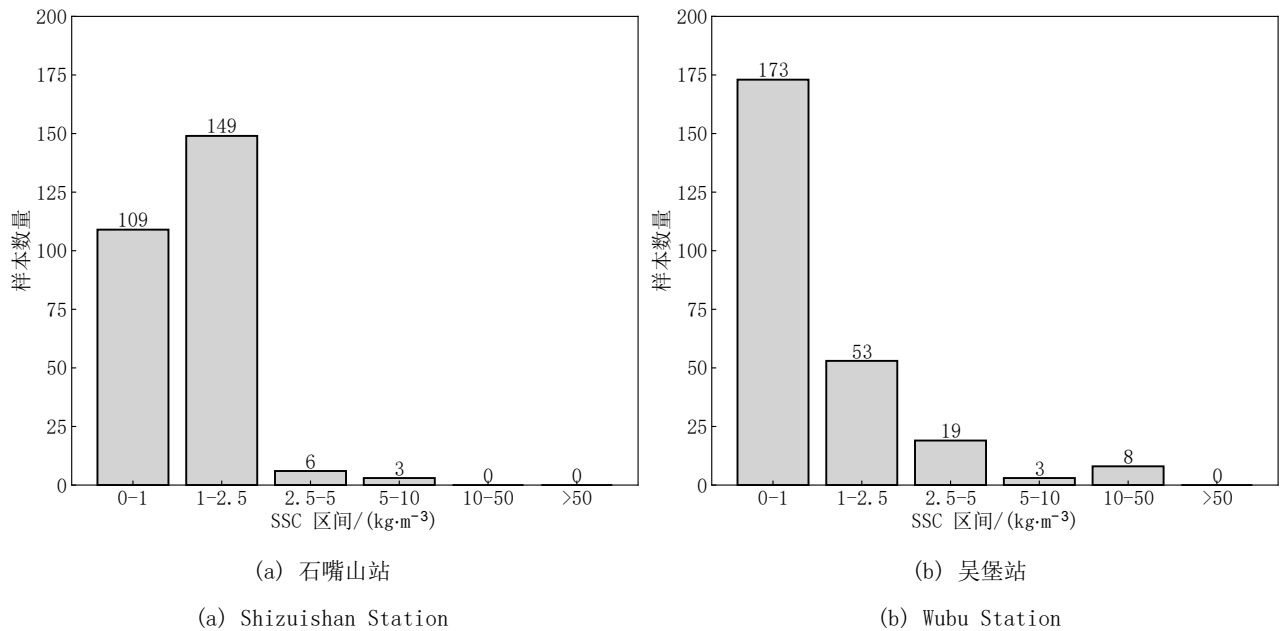
89 沙量数据作为研究数据集中的 SSC 样本标签。本研究将两个站的 SSC 数据分别与卫星多光谱数据进行匹配构造
90 数据集，研究多光谱特征与 SSC 之间的关联关系。经时空匹配，研究区域地表反射率~SSC 数据集见表 3。将
91 日均 SSC 与瞬时观测的卫星光谱反射率数据进行匹配时，由于时间偏差会在 SSC 低值区间造成较大的数据噪
92 声。因此，本研究采用孤立森林法清洗 SSC 低值区间的异常数据。孤立森林是一种高效的无监督异常检测算法，
93 其核心思想是异常数据点由于数量少且特征值与正常点差异大，因此更容易被隔离，特别适合处理高维数据。

94 经过清洗，石嘴山站、吴堡站的数据集分别为 267 条、256 条，清洗比例小于 10%（表 3）。SSC 数据呈现
95 长尾分布（图 2），石嘴山站的数据更多集中于低于 2.5kg/m³ 的范围，而吴堡站则在高于 2.5kg/m³ 的 SSC 区间
96 中具有更多数据样本且最高值达 44.5kg/m³。

97 表 3 多光谱反射率和 SSC 数据集

98 Table 3 Dataset of multispectral reflectance and SSC

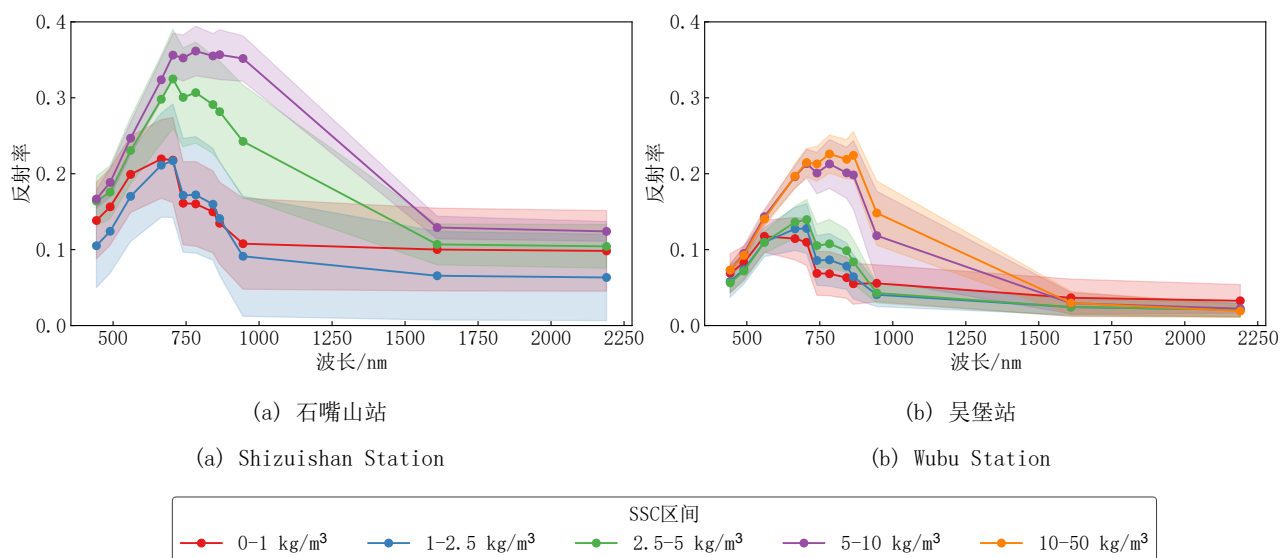
站点	数据时间	原始数据集	清洗后的数据集	SSC/(kg·m ⁻³)
石嘴山	2020/1/1-2024/12/31	289	267	0.21-8.28
吴堡	2020/1/1-2024/12/31	282	256	0.07-44.5



99 图 2 匹配清洗后的 SSC 数据分布
100 Fig.2 Distribution of matched and cleaned SSC data
101

102 不同 SSC 区间的光谱响应曲线见图 3。图中红色、蓝色、绿色、紫色和橙色的实线分别表示 SSC 在[0, 1]、
103 [2.5)、[2.5, 5)、[5, 10)、[10, 50) kg/m³ 区间内各波段的光谱反射率均值，而阴影部分表示该区间反射率的标准

104 差。由图 3 可知, 随着 SSC 区间增大, 最强反射率对应的波长呈现“红移”现象。而在相同波段以及相同 SSC
105 区间内, 吴堡站的光谱反射率平均要低于石嘴山站, 这可能与测站的泥沙特性以及当地的监测环境不同有关。



106
107 图 3 不同 SSC 区间的光谱响应曲线

108 Fig.3 Spectral response curves for different SSC intervals

109 3 水体光谱反射与 SSC 的反演

110 3.1 基于 RFECV-RF 的机器学习模型

111 交叉验证递归特征消除 (Recursive Feature Elimination with Cross-Validation, RFECV) 通过交叉验证避免过
112 拟合, 评估每折下基模型性能从而递归消除低重要性特征, 实现自动选择最优特征数量来提升模型性能。随机
113 森林 (Random Forest, RF) 通过决策树集成捕捉特征间的交互作用, 能够量化评估特征重要性, 适用于处理高
114 维、复杂或包含噪声的数据集。本研究构造基于交叉验证递归特征消除的随机森林模型 (RFECV-RF), 以 RF
115 作为 RFECV 的基模型计算特征的重要性分数, 自动选择高维数据下的最优光谱特征, 再根据 RF 实现水体光谱
116 对 SSC 的反演模拟。

117 RFECV-RF 运算流程示意图见图 4。通过 RFECV 筛选特征, 在降低特征维度的同时可不降低模型精度。将
118 数据集通过 RF 基模型运算后, 根据重要性评分, 递归剔除得到基本的重要性特征, 之后将数据集分为 k 折, 依
119 次用 $k-1$ 折数据训练基模型并计算验证每折的评分, 移除低重要性特征, 在 k 折交叉验证后选择其中评分最优
120 的特征为最优特征子集。接着, 通过 RF 模型构建最优光谱特征与 SSC 之间的关联关系。将初始数据集分层抽

样划分得到训练集和测试集之后，从训练集中有放回地随机抽样得到 n 个子集，分别输入到 n 棵决策树中训练，通过网格搜索优化决策树的结构，最后将 n 棵决策树的预测结果集成得到 RF 模型的 SSC 预测值。最后，再用测试集的数据测试以上模型的性能。

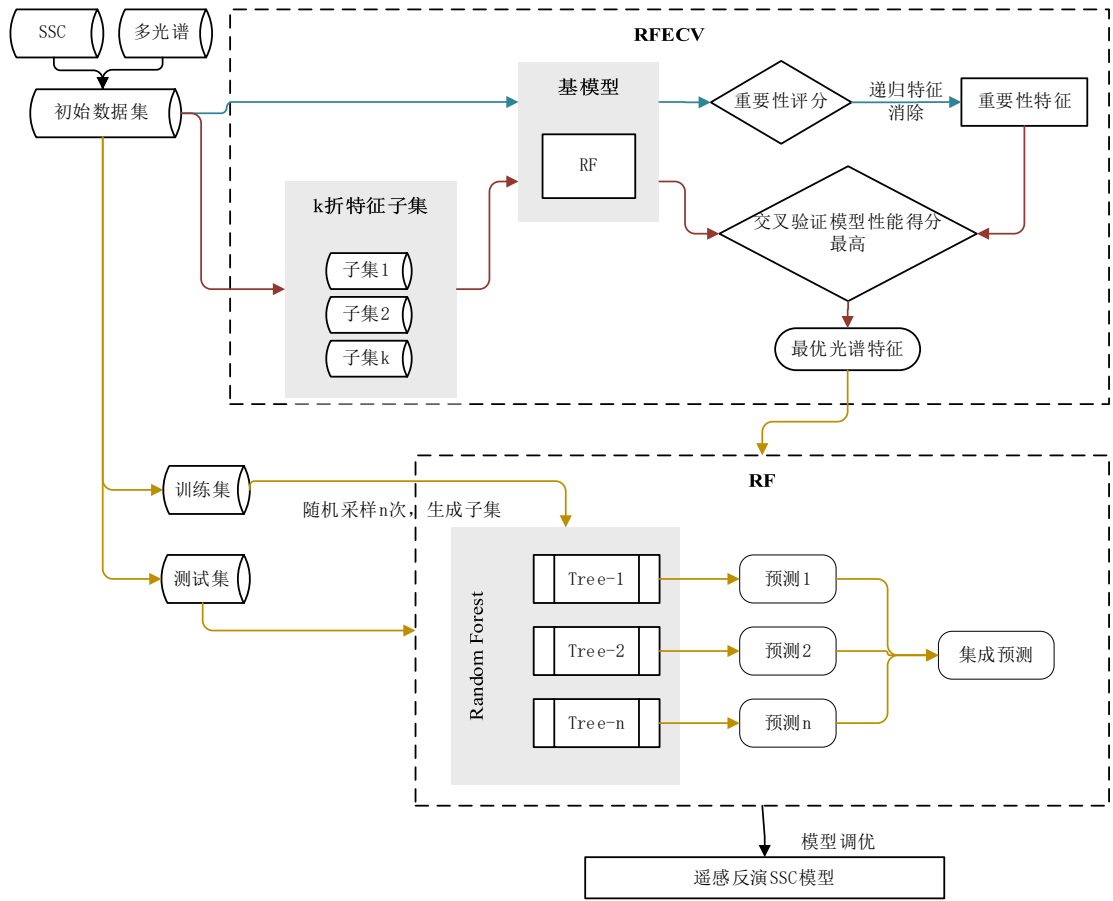


图 4 RFECV-RF 模型运算流程示意图

Fig. 4 Schematic diagram of the RFECV-RF model simulation process

3.2 光谱特征构造

反演 SSC 需要多变量共同参与，通常将核心敏感波段特征和辅助敏感波段特征进行组合形成特征库（方馨蕊等，2019；王雅萍等，2024）。在一些最大 $SSC > 1 \text{ kg/m}^3$ 的反演研究（Dogliotti 等，2015；Yao 等，2020；Jiji，2021；Hou 等，2023；Bazzett 和 Wang 等，2024；Qiu 等，2024；Fang 等，2025）中，使用较多的主要敏感特征为近红外、短波红外、近红外/可见光比值，辅助敏感特征包括蓝光、绿光、红边/可见光以及短波红外与可见光的波段差值。在一些最大 $SSC < 1 \text{ kg/m}^3$ 的反演研究（Wackerman 等，2017；Pham 等，2018；Ghorbani 等，2020；Aires 等，2022；Mohsen 等，2022；Kwon 等，2022b；Jung 等，2024；郑直等，2025）中，核心敏感特

134 征多采用绿光、红光、红边以及近红外，辅助敏感特征包括蓝光、波段比值和波段差值。由此可见，对于不同
135 浓度范围的 SSC 反演研究中，近红外波段的重要性最高，而波段组合特征常用来抵消环境干扰。

136 由于本研究的 SSC 量程大且范围宽（0.07-44.5kg/m³），从清水到浊水的敏感特征会“红移”，因此，在初
137 始光谱特征中采用了可见光、红外、红边、近红外、短波红外单波段，并在前人提出的敏感组合波段中增加了
138 一些组合特征（表 4）。其中，可见光/红边、红边/红边、可见光/近红外、近红外/红边、短波红外/短波红外等
139 比值特征在抵消环境变化干扰时，比值的数值变化也能量化悬浮泥沙引起的“红移”现象，在清水到浊水的过
140 程将发挥重要作用。而可见光-短波红外、近红外-短波红外差值特征通过波段差分消除清澈水体背景信号。

141 表 4 初始光谱特征

142 Table 4 Initial spectral features

特征类型		采用特征	备注
单波段	可见光	B2, B3, B4	反应水色，低悬浮泥沙浓度区间河流浑浊程度可用颜色判断
	红边	B5, B6, B7	对悬浮泥沙浓度的响应更稳定，有效减少水体水质、有机碳的干扰
	近红外	B8, B8A	悬浮泥沙浓度升高后的敏感波段
	短波红外	B11, B12	水体强吸收波段，反映悬浮泥沙浓度极高值
波段差值	可见光-短波红外	B2-B12	整合蓝光和短波红外响应特征，消除清澈水体光学干扰
	近红外-短波红外	B8-B11	整合近红外和短波红外响应特征，消除清澈水体光学干扰
波段比值	可见光/红边	B3/B5	反映绿波段与红边起始波段的相对强度，可识别水体颜色变化，比值可有效 区分低浊度变化
	红边/红边	B7/B5, B5/B6	红边斜率指数，反映红边区域的反射率变化速率，量化“红移”现象
	可见光/近红外	B4/B8, B3/B8,	反映红光、绿光与近红外的平衡，动态范围较大，区分清水和浊水的转变， 在极高悬浮泥沙浓度情况下，光谱饱和时趋于稳定
	近红外/红边	B8A/B6	近红外与红边的过渡带比值，捕捉从红边到近红外的反射率跃升
	短波红外/短波红外	B11/B12	短波红外区域内部对比度，体现水体在极高悬浮泥沙浓度时反射率的变化

144 RFECV-RF 模型关键参数的设置见表 5。RFECV 模型结构中，采用 RF 模型作为基模型，根据特征重要性
145 评估来递归消除不重要特征；每次删除特征数量为 5 个，加速特征计算过程；交叉验证折数为 3 折，避免小样
146 本过拟合，同时保证高 SSC 样本在每折内均匀分布；评分标准为决定性系数 R^2 ，筛选出在整个 SSC 区间整体
147 较敏感的特征。RF 模型结构中，决策树最大深度限制树的分支层数，设置为 15，防止过拟合，平衡模型对复杂
148 水体光谱特征的学习能力与泛化能力，能够捕捉波段交互作用且不会记忆噪声；叶节点最小样本数表示分裂内
149 部节点所需最小样本数，设置为 2，控制树的生长精细度，捕捉细微 SSC 变化的光谱响应；决策树数量即为随
150 机森林中树的数量，设置为 150，平衡模型集成复杂度和计算效率；样本子集采样率表示每棵树使用的训练样本
151 比例，设置为 0.9，增加树间差异，减少方差，可提高模型对少量高 SSC 样本的利用率。

152 表 5 模型参数

153 Table 5 Model parameters

模型	关键参数	设置
RFECV	模型评估	RF
	每次删除特征数量	5
	交叉验证折数	3
	评分标准	R^2
RF	决策树最大深度	15
	叶节点最小样本数	2
	决策树数量	150
	样本子集采样率	0.9

154 3.4 模型评估

155 3.4.1 模型性能评估

156 研究采用决定性系数 R^2 、均方根误差 RMSE (Root Mean Square Error)、平均绝对误差 MAE(Mean Absolute
157 Error)来评估模型可靠性。其中， R^2 反应预测值与实测值之间的整体吻合程度，该值越接近 1 则表明模型整体预

158 测效果越好。RMSE 反应预测值与实测值之间的误差但对高值误差较敏感，该值越接近 0，则表明模型精度越
159 高。MAE 反应预测值与实测值的整体平均误差，该值越接近 0，则表明模型整体误差越小。

$$160 \quad R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (1)$$

$$161 \quad \text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (2)$$

$$162 \quad \text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (3)$$

163 式中， y_i 为 SSC 的实测值， $\hat{y}_i$ 为 SSC 的预测值， $\bar{y}$ 为实测 SSC 的均值， N 为样本数。

164 3.4.2 模型参数敏感性评估

165 基于 RFECV-RF 模型的关键参数（表 5），分别设置每次删除特征数量、交叉验证折数、决策树最大深度、
166 叶节点最小样本数、决策树数量、样本子集采样率的参数变化区间为[1， 7]、[2， 7]、 [5， 35]、[2， 8]、[50，
167 300]、[0.3， 1.0]，固定其他参数不变依次扰动其中一个参数，计算模型测试集 R^2 、RMSE、MAE 的变化率，采
168 用“敏感性系数” S 来评估模型参数的敏感性。 S 越大，参数对模型性能越敏感。

$$169 \quad S = \frac{\Delta P / P_0}{\Delta \theta / \theta_0} \quad (4)$$

170 其中， θ 为模型关键参数， P 为模型性能指标（ R^2 /RMSE/MAE）， $\Delta \theta$ 、 ΔP 表示扰动后的变化量， P_0 、 θ_0 为初始
171 值。

172 3.4.1 模型不确定性评估

173 采用蒙特卡洛模拟框架，通过系统地扰动模型输入、参数和结构，量化模型反演的 SSC 预测的不确定性。

174 关于模型不确定性来源的误差量化如下：

175 （1）输入数据的不确定性：考虑遥感数据采集、预处理过程中引入的随机误差，由于卫星多光谱观测数据
176 与 SSC 数据在时间上的差异，对于低 SSC 值（ $<1 \text{ kg/m}^3$ ）的匹配误差更大，在输入数据中分别将低值样本、高
177 值样本中的光谱特征扰动 5%、3%。

178 （2）模型参数的不确定性：考虑因模型超参数选择不唯一导致预测结果的变化，分别设置决策树最大深度、
179 叶节点最小样本数、决策树数量、样本子集采样率的变化区间为 [5， 35]、[2， 8]、[50， 300]、[0.3， 1.0]，在参
180 数区间内采用均匀分布独立采样，每次蒙特卡洛迭代生成新的参数组合。

181 (3) 模型结构的不确定性：考虑因模型算法选择和固有偏差导致的不确定性，构建 RF、Gradient Boosting
182 Machine (GBM)和支持向量机 (SVR) 模型库，根据单个模型基础参数的 R^2 分数设置模型被选择的概率，设置
183 模型随机种子，并假设低值 SSC 区间误差对数正态分步、高值 SSC 区间误差正态分步来量化模型固有偏差。

184 设置 500 次蒙特卡洛模拟，每次模拟随机扰动数据输入、模型参数和模型结构，得到 500 组 SSC 反演结
185 果。采用反演结果的 95% 置信区间 CI_{95} (95% Confidence Interval)和相对标准偏差 RSD (Relative Standard
186 Deviation) 来表征总的不确定性：

$$187 \quad CI_{95} = \mu \pm 1.96\sigma \quad (5)$$

$$188 \quad RSD = \frac{\sigma}{\mu} \times 100\% \quad (6)$$

189 其中， σ 为预测结果的标准差， μ 为预测均值。

190 采用方差分解法计算各来源的贡献度 C ，不同不确定性来源的贡献度 C_i 计算如下：

$$191 \quad C_i = \frac{Var(U_i)}{Var(U_t)} \times 100\% \quad (7)$$

192 其中， U_i 为各项的不确定性（数据、模型参数、模型结构）； U_t 为总的不确定性； $Var(U_i)$ 为各类不确定性的
193 独立方差； $Var(U_t)$ 为模拟反演结果的总方差。

194 4 结果分析与讨论

195 4.1 关键特征分析

196 由 RFECV 方法筛选出的关键光谱特征见图 5。石嘴山站模型有 7 个重要光谱特征，其中重要性得分最高的
197 特征为 B8-B11，重要性得分最高的单波段特征为 B8A，其他重要的组合特征依次为 B3/B8、B4/B8、B7/B5、
198 B3/B5、B8A/B6。吴堡站模型有 5 个重要光谱特征，其中重要性得分最高的光谱特征为 B7/B5，重要性得分最高
199 的单波段特征为 B8A，其他重要的组合特征依次为 B5/B6、B11/B12、B4/B8。石嘴山站与吴堡站相同的重要特
200 征分别为近红外光 B8A 以及红光、红边主导的 B4/B8、B7/B5 特征，这是由于近红外光在高含沙河流中突出的
201 敏感性 (Yao 等, 2020)，以及红光和红边相关的波段比值可反映在河流 SSC 升高过程中对光谱敏感性的变化。

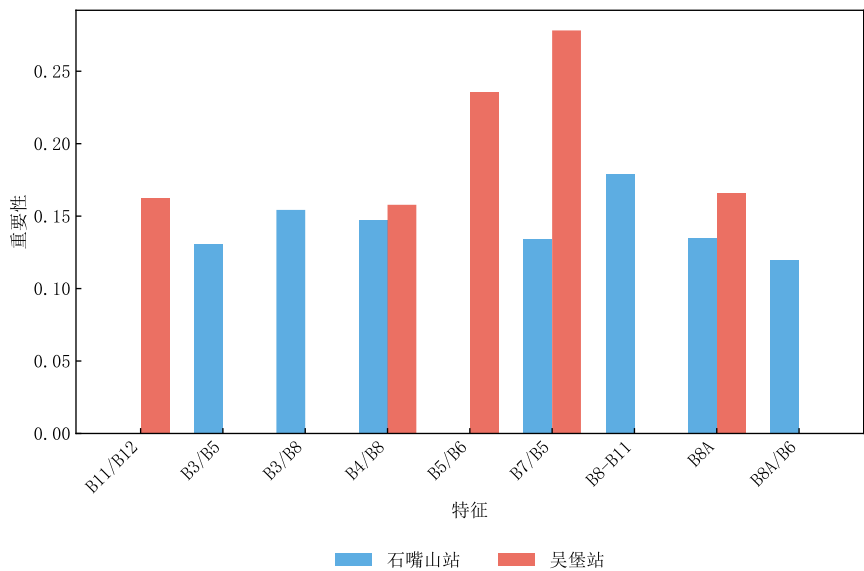
202 石嘴山站独有的关键特征为绿光主导的 B3/B5、B3/B8 以及近红外光主导的 B8-B11、B8A/B6。这是由于石
203 嘴山站的数据样本中 SSC 大多低于 2.5kg/m^3 ，B3/B5、B3/B8 特征在反应河流水色的同时还能反应出河水从清水

204 到浊水的转变; 而 B8-B11、B8A/B6 兼顾光谱在 (0.21-8.28) kg/m³ SSC 区间内的响应变化且能够抵消一些干扰。

205 吴堡站独有的关键特征分别是短波红外内部比值 B11/B12 和红边内部比值 B5/B6。这主要与吴堡站的 SSC

206 宽范围 (0.07-44.5 kg/m³) 有关。短波红外穿透水体表层的能力更强, B11/B12 特征在 SSC 极高值区间仍能发挥

207 重要作用; 而 B5/B6 能够量化“红移”现象, 对于反演宽变化范围的河流 SSC 至关重要。



208

209

图 5 关键特征重要性排序

210

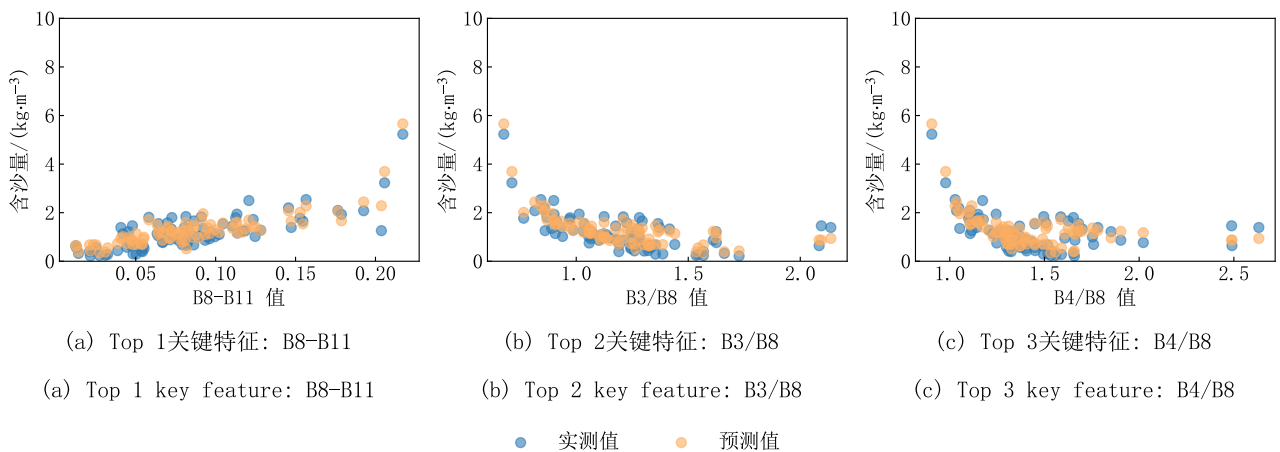
Fig. 5 Ranking of key feature importance

211

212

213

石嘴山站和吴堡站测试集中重要性排名前三的关键特征与 SSC 的散点图见图 6 与图 7。不论是测试集中 SSC (0-6) kg/m³ 区间的石嘴山站, 还是 SSC (0-44.5) kg/m³ 区间的吴堡站, 关键光谱特征与实测 SSC、预测 SSC 之间均呈现明显的非线性关系, 表明模型结构能够正确预测出这种非线性关系。



214

215

图 6 关键特征与 SSC 关系分析 (石嘴山站)

216

Fig.6 Relationship between key features and SSC (Shizuishan Station)

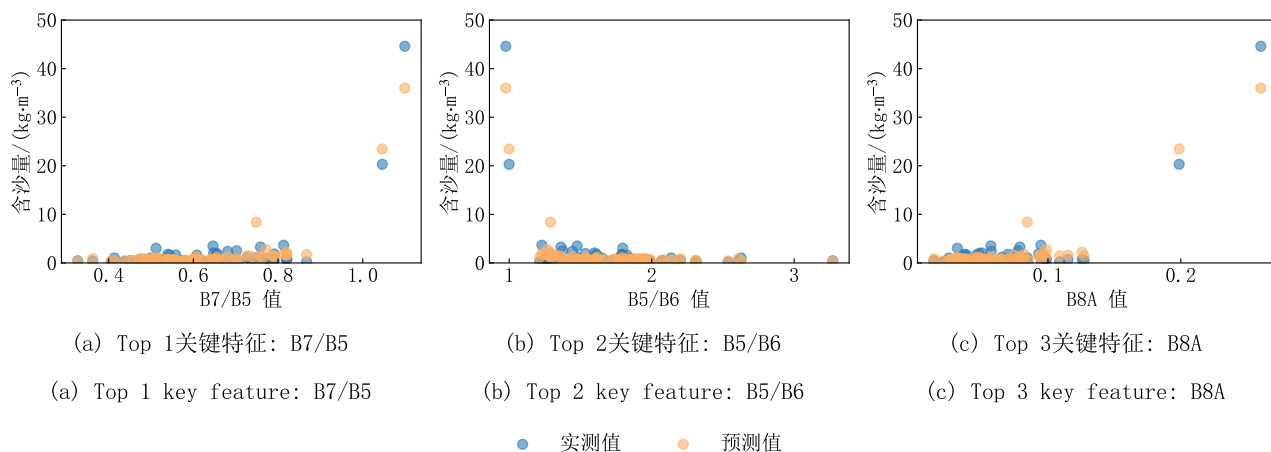


图 7 关键特征与 SSC 关系（吴堡站）

Fig.7 Relationship between key features and SSC (Wubu Station)

4.2 模型性能评估及河段反演

4.2.1 模型性能评估

由于 SSC 数据集呈现明显的长尾分布，采用分层抽样以确保各 SSC 区间的代表性，将数据集按照 7:3 的比例划分成训练集和测试集。分别将石嘴山站 267 条、吴堡站 256 条数据集输入模型中调试，模型在训练集和测试集中的表现见表 6，模型 SSC 预测值与实测值对比见图 8。

表 6 模型性能指标

Table 6. Model performance metrics

站点	数据集	R^2	RMSE/(kg·m ⁻³)	MAE/(kg·m ⁻³)
石嘴山	训练集	0.964	0.168	0.100
	测试集	0.824	0.314	0.249
吴堡	训练集	0.942	1.270	0.448
	测试集	0.918	1.567	0.841

在测试集中，模型对于石嘴山站和吴堡站 SSC 的光谱反演 R^2 均大于 0.8，RMSE 均小于 1.6 kg/m³，MAE 均小于 0.9 kg/m³，这表明模型能够基于卫星多光谱信息准确地预测在（0-44.5）kg/m³ 区间内的河流 SSC。从预测数据集与实测数据集的整体吻合程度而言，测试集中吴堡站模型的 R^2 大于 0.9，且高于石嘴山站，这表明模型充分考虑了在高值区间光谱特征与 SSC 之间的关系，适用于对高含沙宽范围河流 SSC 的反演。

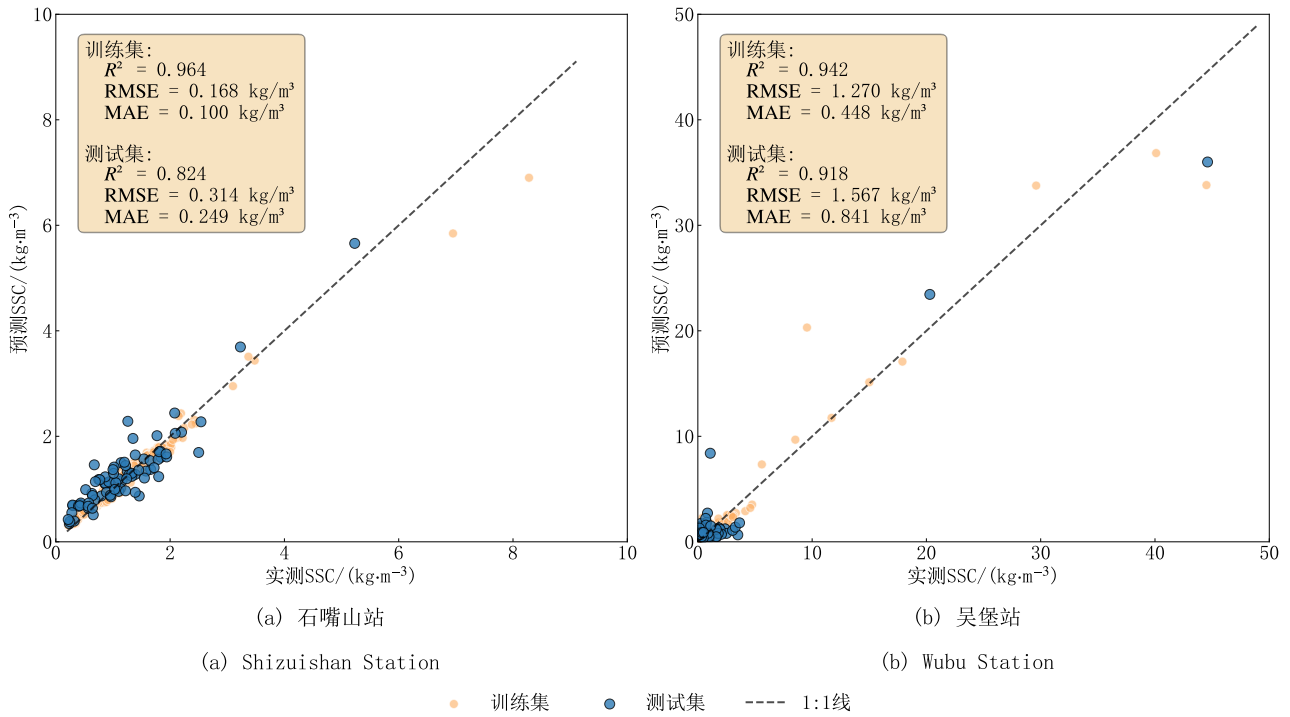


图 8 模型预测值与实测值的对比

Fig. 8 Comparison between predicted SSC and measured SSC

由图 8 (a)可知, 对石嘴山站 SSC 低于 6 kg/m³ 的预测, SSC 预测值与实测值基本相符, 但对该站 SSC 最高值 (8.28kg/m³) 的预测明显偏低。由图 8 (b)可知, 模型对低于 20kg/m³ 的 SSC 预测基本与实测值相差不大, 但对该站 SSC 最高值 (44.5kg/m³) 的预测同样偏低。这主要是由于两个站的极高值 SSC 数据都较稀疏, 模型对于极高值 SSC 情况下光谱特征的表现还有待加强。因此, 本研究所构建的 RFECV-RF 模型能够根据特定的光谱特征反演预测高含沙河流 SSC, 但会低估站点最高值 SSC, 在未来将进一步收集极高值 SSC 数据样本进行补充。

4.2.2 河段反演

基于 MNDWI 提取 Sentinel-2 中的纯水体多光谱影像之后, 分别采用石嘴山站、吴堡站的 SSC 反演模型来计算 2024 年河段 1 (黄河上游)、河段 2 (黄河中游) 反演区域的 SSC 空间分布 (图 9、图 10)。

由图 9 可知, 河段 1 SSC 最高平均值 (5.30 kg/m³) 出现在 8 月 13 日 (图 9 (c)), 尽管因卫星轨迹的原因影像存在部分缺失, 但仍能反应出河段 SSC 的空间分布, 且分岔处 SSC 不均匀的特性较明显 (图 9 (g))。河段 1 SSC 在时间变化上, 从 8 月份开始输送高含沙, 经过起伏后直至 10 月份还有一场较小的沙峰, 之后回归低沙。

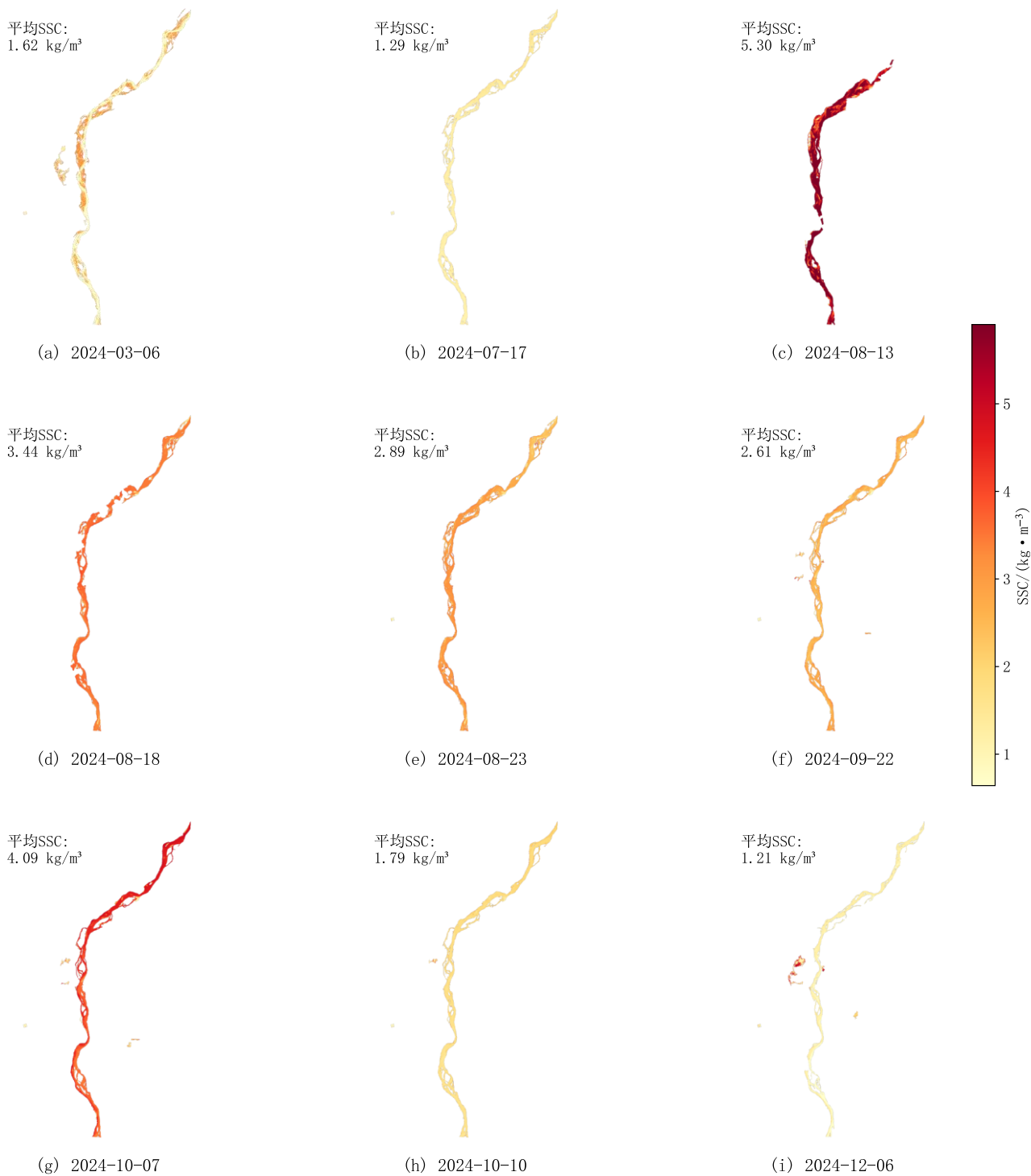


图 9 河段 1 反演区域 SSC 时空分布

Fig.9 Spatial-temporal distribution of SSC in the inversion area of the river reach 1

河段 2 SSC 在时间变化上, 8 月份的 SSC 明显高于其他月份, 9-10 月份次之, 最高为 2024 年 8 月 12 日 (平均 SSC: $25.61 \text{ kg}/\text{m}^3$), 最低为 2024 年 1 月 3 日 (平均 SSC: $2.85 \text{ kg}/\text{m}^3$), 时间趋势上呈现一个 SSC 上升后浮动再下降的过程, 并且 SSC 高值月份与王江瑜等(2025)的研究中相同, 与实际较为相符。空间上, 河段 SSC 反演能够反映出 SSC 输送来源, 例如 2024-10-1 河段 SSC 反演图 (图 10 (f)) 中, 在河段上游来沙较少的情况

下，支流的 SSC 较高，此时河段下游也呈现较高的 SSC，反映了支流泥沙汇入对主流河段 SSC 的影响。而相比河段 1，河段 2 SSC 在时间上的变化规律与河段 1 相似，但是 SSC 含量等级远高于河段 1，这主要是由于黄河中游是主要泥沙来源区域，故而中游的河段 2 平均 SSC 要高得多。

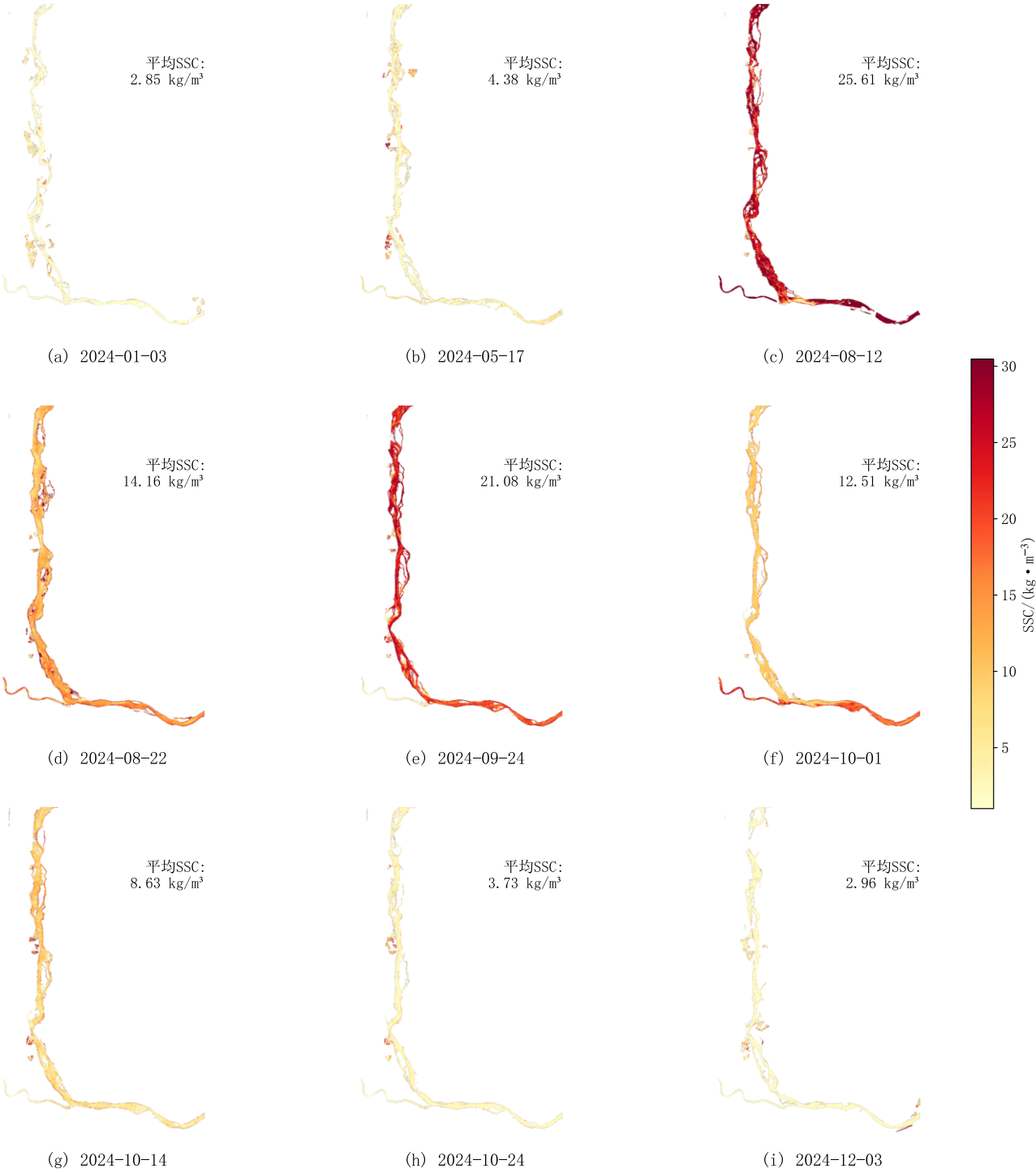


图 10 河段 2 反演区域 SSC 时空分布

Fig.10 Spatial-temporal distribution of SSC in the inversion area of the river reach 2

综上，基于卫星大范围光谱监测数据，将测站遥感反演 SSC 模型应用于高含沙河流，可以捕捉河段时空 SSC 的变化趋势，为流域泥沙输移规律的分析提供丰富的数据。

261 4.3 光谱灵敏性分析

262 由 4.1 小节可知，对于高含沙河流，光谱特征与 SSC 之间的非线性关系非常明显。而且随着 SSC 增大，光
263 谱的变化相对于泥沙含量的变化迅速降低，光谱特征灵敏性有待分析。对于 4.1 小节中筛选出的关键特征，采用
264 幂函数、指数函数、二次多项式函数的模式（表 7）来拟合关键特征与实测 SSC 之间的关系，拟合的精度用 R^2
265 来表示（图 11）。分别计算 $SSC > 2.5\text{kg/m}^3$ 和 $SSC < 2.5\text{kg/m}^3$ 区间时光谱特征变化相比 SSC 变化的斜率，用斜率
266 衰减比 SDR（特征相对高 SSC 的变化斜率与特征相对低 SSC 的变化斜率的比值）来表征光谱特征对 SSC 变化
267 的灵敏性。SDR 越接近 1 表明光谱灵敏性越高，而当 $SDR \leq 0.01$ 时，可认为光谱特征有饱和趋势。

268 表 7 光谱特征与 SSC 的拟合关系式

269 Table 7 Fitting relationship between spectral features and SSC

特征	石嘴山站	吴堡站
B8A	$SSC = 0.0016e^{21.54(B8A)} + 1.06$	$SSC = 3036.49B8A^{3.29}$
B7/B5	$SSC = 0.00000008e^{17.77(B7/B5)} + 0.97$	$SSC = 0.000011e^{13.87(B7/B5)} + 0.67$
B4/B8	$SSC = 3.10(B4/B8)^{-3.34}$	$SSC = 10.90(B4/B8)^{-7.22}$
B3/B5	$SSC = 7.18(B3/B5)^2 - 17.15B3/B5 + 10.47$	$SSC = 0.48(B3/B5)^{-7.80}$
B3/B8	$SSC = 1.59(B3/B8)^{-2.61}$	$SSC = 2.94(B3/B8)^{-4.69}$
B11/B12	$SSC = -0.23(B11/B12)^2 + 1.19B11/B12 + 0.24$	$SSC = 0.39(B11/B12)^{-7.74}$
B5/B6	$SSC = 3.98(B5/B6)^{-5.03}$	$SSC = 27.47(B5/B6)^{-13.24}$
B8-B11	$SSC = 73.58(B8 - B11)^2 - 1.15(B8 - B11) + 0.71$	$SSC = 920.08.2(B8 - B11)^2 - 64.60(B8 - B11) + 1.21$
B8A/B6	$SSC = 1.39e^{35.52(B8A/B6)} + 1.11$	$SSC = 6.30(B8A/B6)^{10.85}$

270 由图 11 可知，随着 SSC 升高至极高值，特征与 SSC 之间的斜率明显衰减，而吴堡站模型的特征灵敏性衰
271 减普遍高于石嘴山站，这是由于吴堡站 SSC 范围更宽，光谱灵敏性下降趋势更明显。而少部分特征与 SSC 之间
272 关系较为散乱，例如 B11/B12 特征与石嘴山站 SSC 相关性非常差（图 11 (f)）。一方面短波红外特征对低值 SSC
273 不敏感，另一方面由于低值样本中卫星多光谱反射率数据与水文站数据匹配的误差较大。

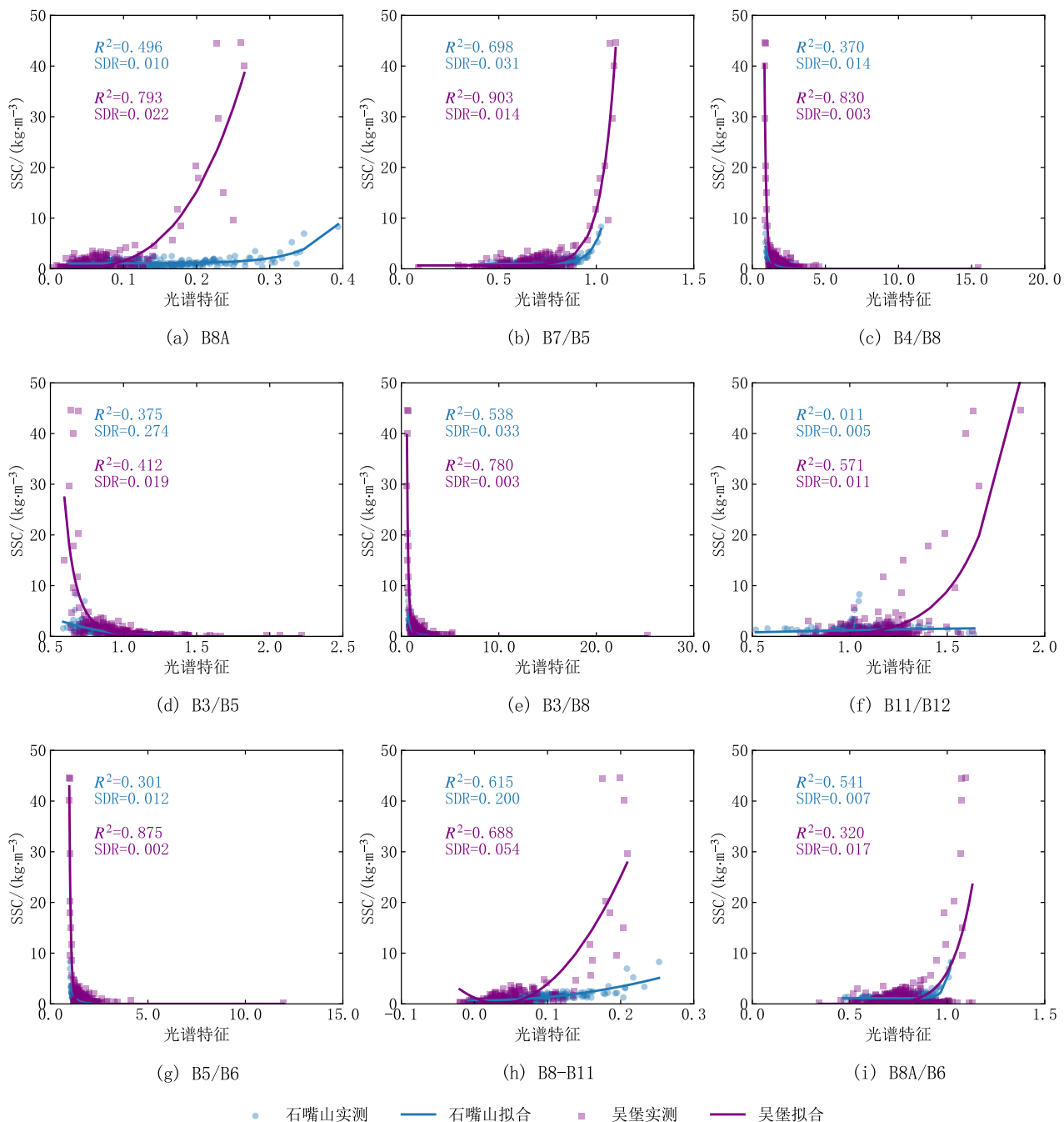


图 11 光谱特征与 SSC 的回归关系（蓝色、紫色标注分别代表石嘴山站、吴堡站拟合统计指标）

Fig.11 Regression relationship between spectral features and SSC (The blue and purple labels represent the fitting statistical indicators of the Shizuishan Station and Wubu Station, respectively)

光谱特征与 SSC 拟合的 R^2 和 SDR 对比见图 12。石嘴山站和吴堡站光谱特征与 SSC 拟合 $R^2>0.5$ 的共同特

征为 B3/B8、B7/B5、B8-B11（这些均为石嘴山站模型中的重要特征）。对比这些特征的 SDR 发现，吴堡站的

SDR 低于石嘴山站，说明这些特征随着河流 SSC 的升高灵敏性有所降低，其中 B7/B5、B8-B11 特征的 SDR 仍

高于 0.01，而 B3/B8 特征的 SDR 远小于 0.01，表明该特征已经趋于饱和。在吴堡站模型特征的斜率衰减中，

B8A 特征的 SDR 最高，说明近红外光在 SSC (0-44.5) kg/m³ 范围内始终保持较好的灵敏性。且在 $R^2>0.5$ 的光

283 谱特征中，除了 B3/B8 特征外，B4/B8、B5/B6 特征的 SDR 也明显小于 0.01，明显有饱和趋势。

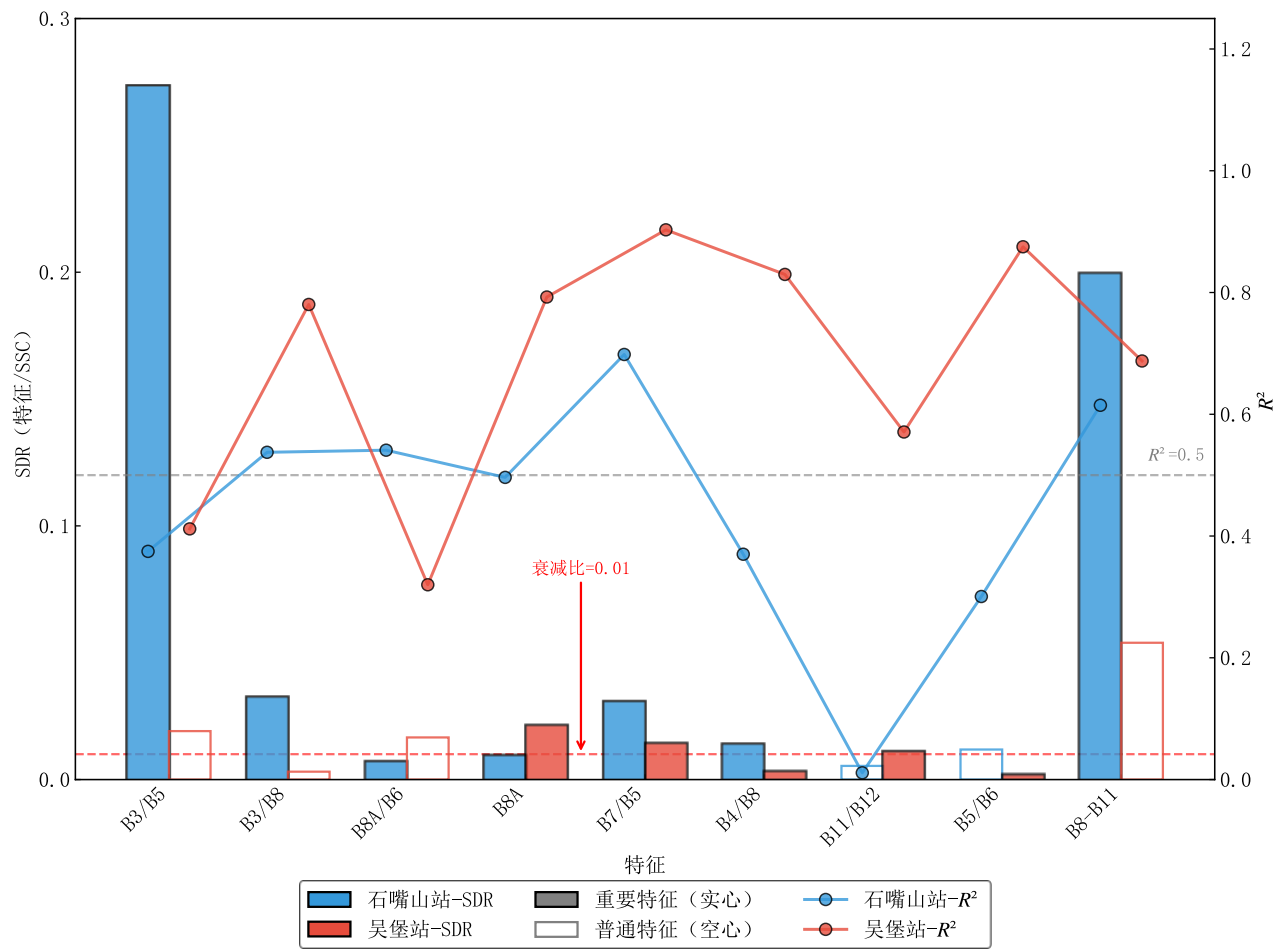


图 12 特征与 SSC 之间的 R^2 指标和斜率衰减比

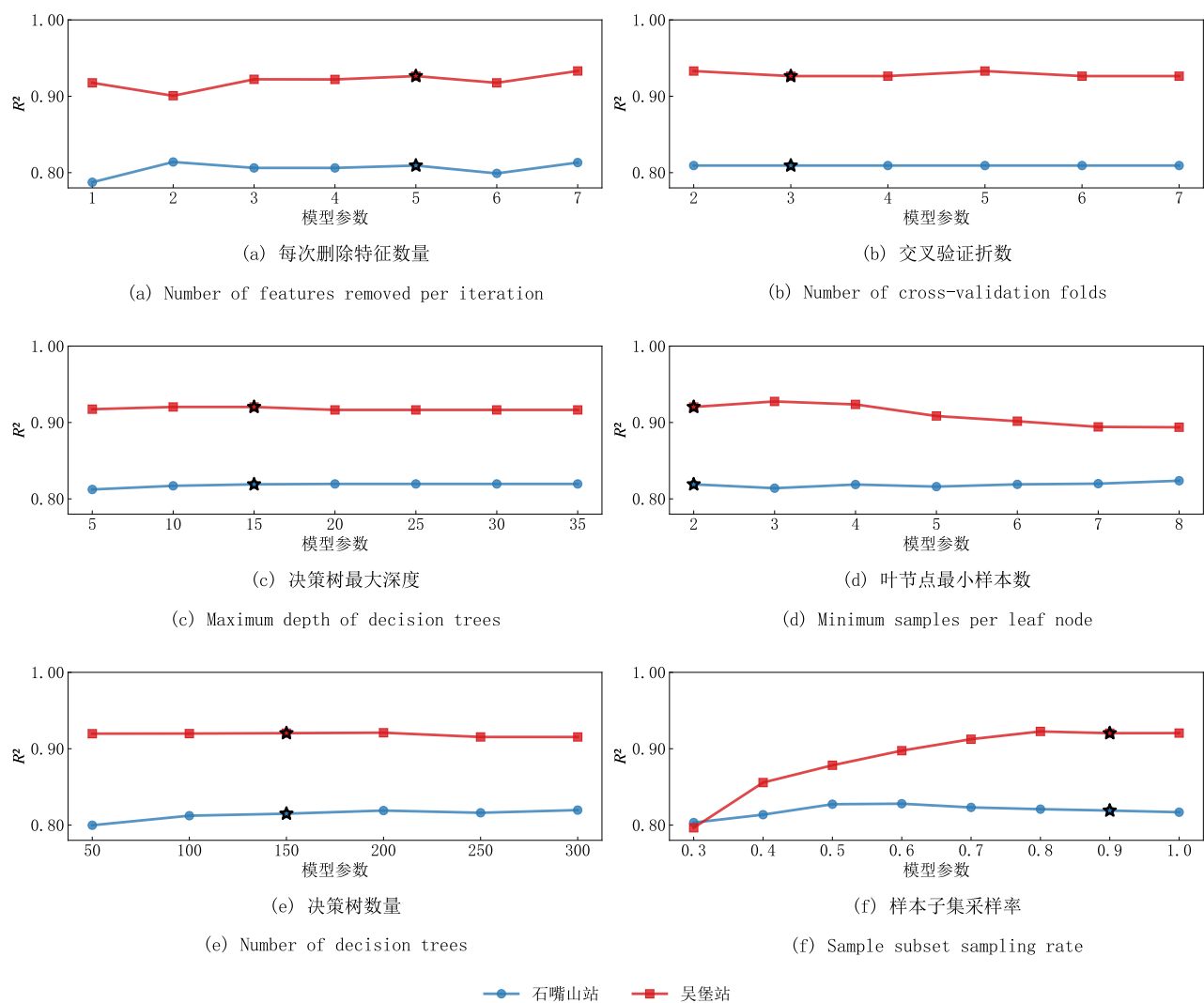
Fig.12 Metrics of R^2 and slope decay ratio between features and SSC

287 综上所述，光谱特征随着 SSC 范围变宽、极值变高而灵敏性降低，对 SSC (0-44.5) kg/m^3 区间内的光谱反
288 演研究中，B8A、B7/B5、B8-B11 始终保持灵敏，而特征 B3/B8、B4/B8、B5/B6 在 SSC 超过 2.5 kg/m^3 时具有显
289 著的饱和趋势。其他波段特征 (B3/B5、B8A/B6、B11/B12) 受到离散数据集中低值数据误差的干扰，光谱丧失
290 灵敏性时的 SSC 有待下一步分析。在宽范围 SSC 中保持灵敏的光谱特征可用于高含沙河流遥感反演。

291 4.4 模型参数敏感性分析

292 模型参数变化对模型测试集 R^2 指标的影响见图 13。模型参数变化对石嘴山站的模型性能的影响较吴堡站
293 小，对于不同的参数变化，模型 R^2 基本在 0.80 上下略微浮动。模型参数变化对吴堡站的影响稍微大一些，但对
294 于不同的参数变化，模型 R^2 同样维持在 0.80 以上，其中造成影响最大的是样本子集采样率。吴堡站模型 R^2 随

295 样本子集采样率的降低而降低，而本文采用的模型参数组合中样本子集采样率设置为 0.9，符合需求且较为合理。



296

297 图 13 模型参数变化下的 R^2 指标（五角星标记为模型参数设置值）

298 Fig.13 Metrics of R^2 under model parameters variation (pentagrams indicate the parameter settings used)

299 模型性能指标对各模型参数的敏感性系数见图 14。模型指标中， R^2 对模型参数的敏感性最低（敏感性系数
300 均小于 0.1），可以忽略不计，而 RMSE 和 MAE 对模型参数的选取较为敏感。对于石嘴山站，对 R^2 、RMSE 和
301 MAE 指标影响最大的是每次删除特征数量。每次删除特征数量关乎模型关键特征的筛选，说明石嘴山站模型对
302 于特征输入是比较敏感的。对于吴堡站，对 R^2 和 RMSE 指标影响最大的参数都是样本子集采样率，对 MAE 指
303 标影响最大的参数是每次删除特征数量，说明在宽范围高极值 SSC 反演模型中，输入数据的大小和结构对于模
304 型性能有一定影响。

305 综上，模型性能总体而言较为稳健，受模型参数的影响不大，最敏感的关键参数为样本子集采样率和每次
306 删除特征数量。

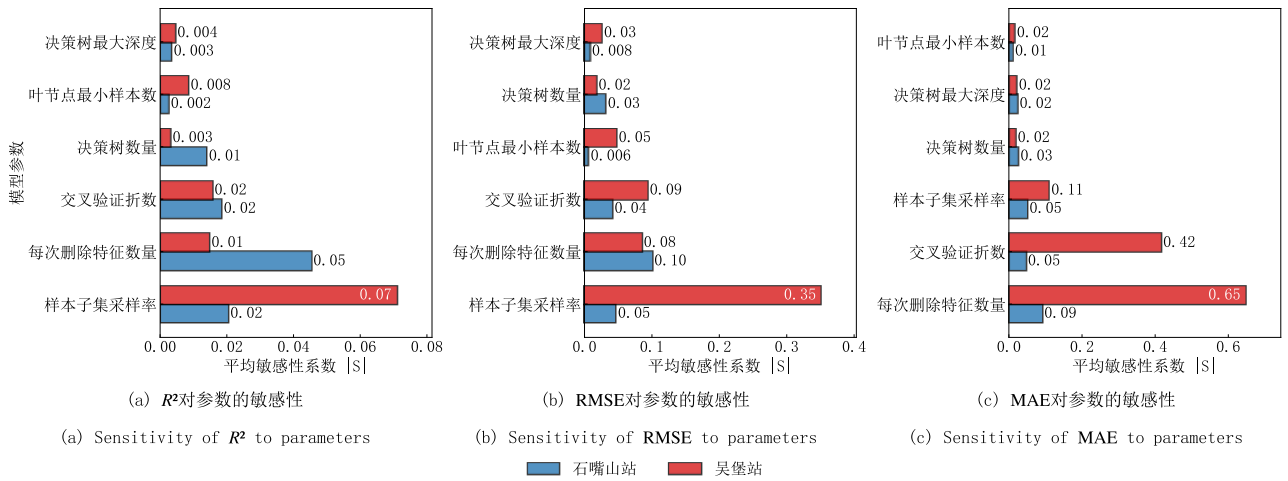


图 14 模型性能对模型参数的敏感性系数

Fig.14 Sensitivity coefficients of model performance to model parameters

4.5 模型不确定性分析

在模型结构不确定性中所设置模型库的基本性能见表 8。由表 8 可知，不同的模型在反演高含沙河流 SSC 时差异较大，GBM 模型反演石嘴山站和吴堡站 SSC 的 R^2 指标 >0.7 ，精度尚可，而 SVR 模型的 R^2 指标相较 RF 和 GBM 模型则较低，特别是在吴堡站中误差很大。根据不同模型基础参数下的精度性能设置归一化权重后，对不同模型被选的概率见表 8。而单个模型运用于遥感反演 SSC 时的不确定性，通过模型内部的随机种子来体现。

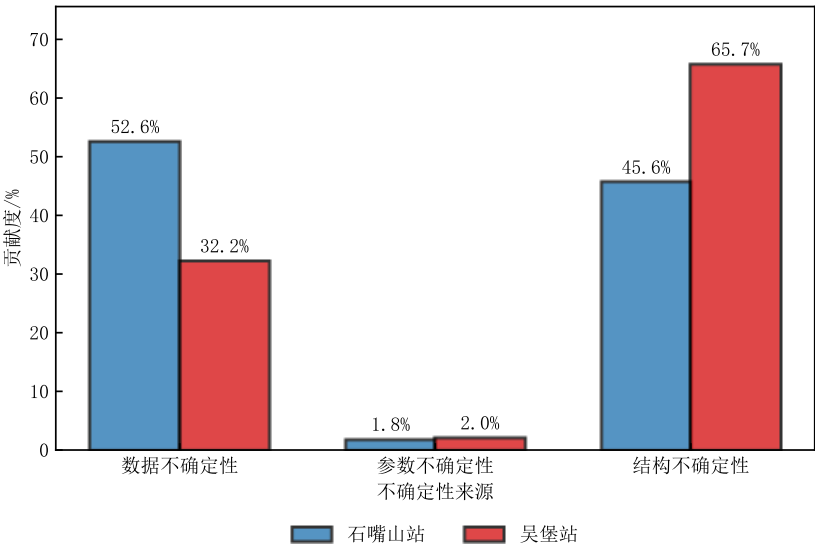
表 8 模型库的性能和被选择概率

Table 8 Performance and selection probability of the models

模型	基础参数	性能 (R^2) -石嘴山站	被选择概率-石嘴山站	性能 (R^2) -吴堡站	被选择概率-吴堡站
RF	同表 5	0.824	0.476	0.918	0.581
GBM	树的数量: 150; 最大深度: 10; 学习率: 0.1	0.731	0.283	0.895	0.377
SVR	径向基函数: rbf; 正则化参数: 1.0	0.623	0.241	0.091	0.042

经蒙特卡洛方法定量化计算，模型不确定性来源的贡献见图 15。由图 15 可知，来自于模型参数的不确定性均很小，模型不确定性来源主要是数据输入和模型结构。对于石嘴山站 SSC 反演模型，数据和结构的不确定性

319 贡献度分别为 52.6%、45.6%，可见来自数据的不确定性占主导，一方面是由于 RF、GBM 和 SVR 模型对于预
320 测该站 SSC 精度较好，另一方面是该站的低值 SSC 较多，而低值数据样本匹配时造成相对误差较大的噪声，导
321 致数据输入存在较大的不确定性。对于吴堡站 SSC 反演模型，数据和结构的不确定性贡献度分别为 32.3%、
322 65.7%，可见来自模型结构的不确定性占主导，一方面是由于模型库中 SVR 模型精度较差，预测误差较大，另
323 一方面是该站的高值 SSC 更多，而高值 SSC 数据样本在匹配时产生的相对误差较低。



324
325 图 15 模型各不确定性来源贡献度

326 Fig.15 Contributions from different model uncertainty sources

327 模型测试集中 SSC 预测值的不确定性指标见表 9。石嘴山站、吴堡站模型预测值对实测值的覆盖率分别为
328 91.4%、87.0%（表 9），模型预测的分布基本覆盖大部分的实测数据（图 16），可见模型误差的估计和量化较
329 为合理。石嘴山 95%置信区间平均宽度为 1.345kg/m³，平均 RSD 为 25.5%；吴堡站 95%置信区间平均宽度为
330 3.049kg/m³，平均 RSD 为 71.6%，吴堡站相对偏差较大且不确定性较高部分集中于低值区间（图 17）。总体而
331 言，石嘴山站模型的不确定性低于吴堡站，可见遥感反演模型在兼顾高浓度宽范围 SSC 的预测精度时，对于低
332 值 SSC 部分会产生较大的不确定性。

333 表 9 模型不确定性指标

334 Table 9 Metrics of the models uncertainty

模型	石嘴山站	吴堡站
CI ₉₅ 覆盖率	91.4%	87.0%

模型	石嘴山站	吴堡站
平均 RSD	25.5%	71.6%
CI ₉₅ 平均宽度/(kg·m ⁻³)	1.345	3.049

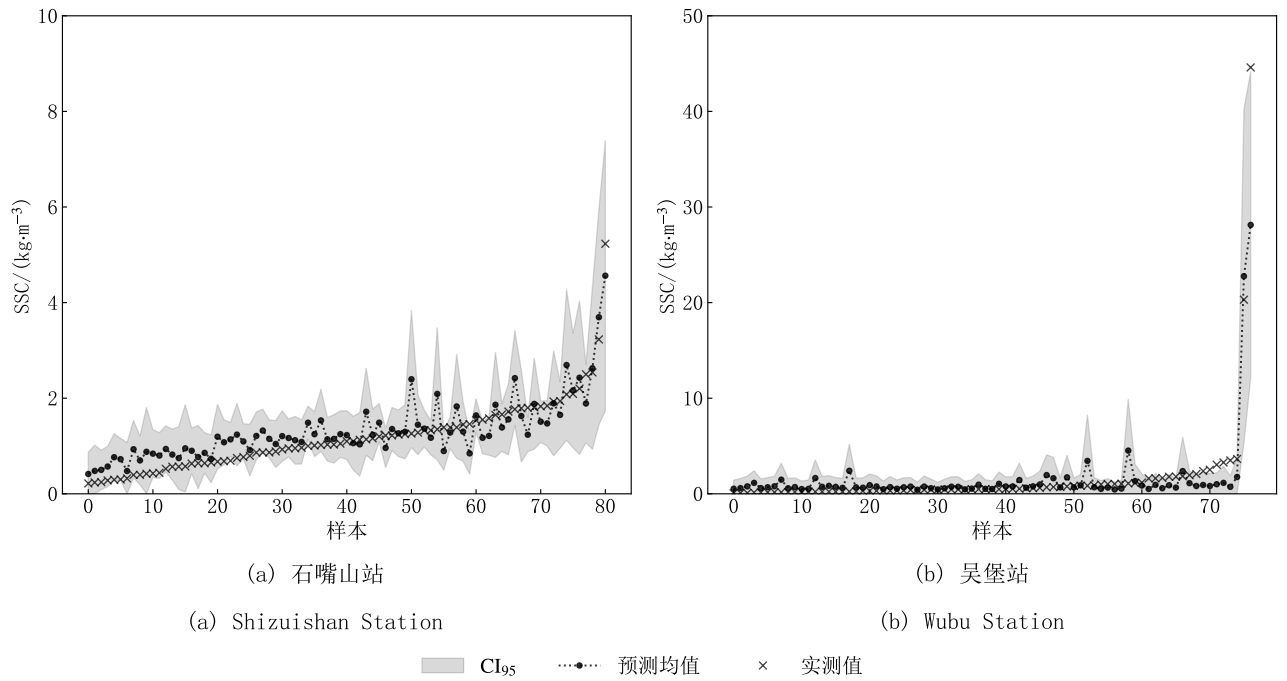


图 16 模型预测值 95%置信区间与实测值的对比

Fig.16 Comparison between the 95% confidence interval of model predictions and the measured values

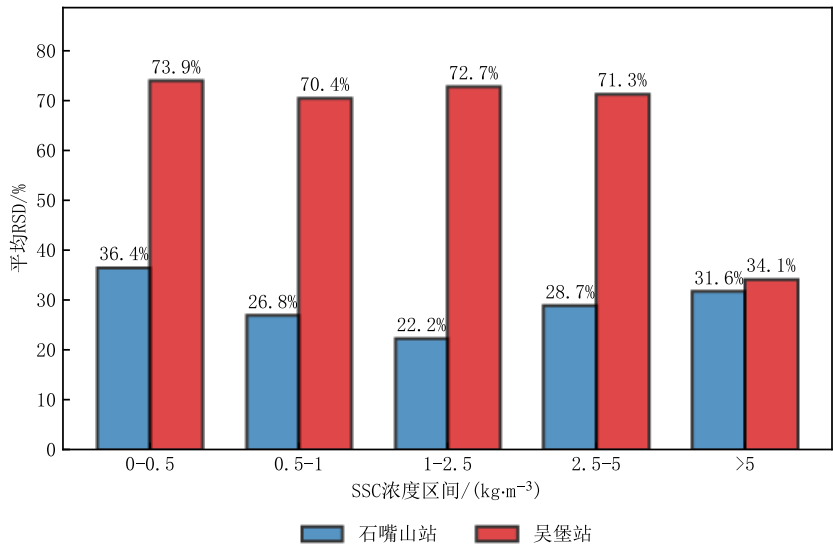


图 17 不同 SSC 区间下模型预测值的平均 RSD

Fig.17 Average RSD of model predictions across different SSC range

5 结论

342 本研究以黄河干流上的石嘴山站和吴堡站作为研究对象，采用 RFEVC-RF 框架构建基于卫星多光谱的 SSC
343 反演模型，分析高含沙河流遥感反演的关键特征及光谱灵敏性、模型精度和河段反演适用性、模型参数敏感性
344 以及模型不确定性，研究结果总结如下：

345 （1）对于高含沙河流 SSC 反演，总体而言以红光、红边、近红外光、短波红外主导的组合特征较为关键，
346 而关键光谱特征与 SSC 的关系表现出明显的非线性关系，且随 SSC 升高而不断降低灵敏性，在（0-44.5） kg/m^3
347 SSC 反演中，B8A 波段的灵敏性最高，B7/B5、B8-B11 特征同样具有灵敏性，而 B3/B8、B4/B8、B5/B6 特征已
348 经趋于饱和；

349 （2）RFECV-RF 模型总体上能够估算（0-44.5） kg/m^3 区间的河流 SSC，但对最高值 SSC 偏低估，对吴堡
350 站、石嘴山站测试集 SSC 的预测精度指标 R^2 均大于 0.8，且将模型运用于站点附近的河段反演可宏观分析河段
351 SSC 年内时序变化和空间输移特征；

352 （3）遥感反演 SSC 模型参数造成的不确定性较小（低于 2%）且模型参数对模型整体性能的影响较小，不
353 确定性主要来源于数据和模型结构（合计占比>98%），而在用于高浓度宽范围 SSC 反演时低值部分的不确定性
354 较大。

355 本研究采用可见光-短波红外范围内的单波段和波段组合特征构建 RFECV-RF 模型，在 SSC（0-44.5） kg/m^3
356 量程范围内实现了光谱特征对 SSC 的反演，分析了光谱特征灵敏性和参数敏感性，并量化了模型不确定性，为
357 高含沙河流 SSC 自动监测提供技术参考。遥感反演 SSC 模型的预测精度及不确定性大小与数据输入息息相关，
358 因此本研究下一步拟通过现场观测高含沙河流的水面高光谱与 SSC，获取光谱分辨率更精细、数据匹配更准确、
359 分层样本代表性更高的数据集，进一步研究光谱反演高浓度 SSC 的性能，为 SSC 遥测技术研究提供参考。

360 参考文献：

361 Aires U R V, Silva D D D, Filho E I F, Filho E I F, Rodrigues L N, Uliana E M, Amorim R S S, Ribeiro C B D M and Campos J A. 2022. Modeling of
362 surface sediment concentration in the Doce River Basin using satellite remote sensing. *Journal of Environmental Management*, 323: 116207 [DOI:
363 10.1016/j.jenvman.2022.116207]
364 Balasubramanian S V, Pahlevan N, Smith B, Binding C, Schalles J, Loisel H, Gurlin D, Greb S, Alikas K, Randla M, Bunkei M, Moses W, Nguyen H,

365 Lehmann M K, Donnell D O, Ondrusek M, Han T H, Fichot C G, Moore T and Boss E. 2020. Robust algorithm for estimating total suspended
366 solids (TSS) in inland and nearshore coastal waters. *Remote Sensing of Environment*, 246: 111768 [DOI: /10.1016/j.rse.2020.111768]

367 Bazzett D and Wang R Q. 2024. Can Surface Water Color Accurately Determine Sediment Concentration and Grain Size? A Hyperspectral Imaging
368 Study. *Water*, 16(15): 2184 [DOI: 10.3390/w16152184]

369 Chen T, Li W and Wu S C. 1994. Correlation between suspended sediment concentration and red-shift of peak wavelength of spectral reflectance. *Acta*
370 *Oceanologica Sinica*, 16(1): 38-43 (陈涛, 李武, 吴曙初. 1994. 悬浮泥沙浓度与光谱反射率峰值波长红移的相关关系. *海洋学报*, 16(1): 38-
371 43)

372 Dogliotti A I, Ruddick K G, Nechad B, Doxaran D and Knaeps E. 2015. A single algorithm to retrieve turbidity from remotely-sensed data in all coastal
373 and estuarine waters. *Remote Sensing of Environment*, 156: 157-168 [DOI:10.1016/j.rse.2014.09.020]

374 Fang X R, Wen Z F, Chen J L, Wu S J, Huang Y Y and Ma M H. 2019. Remote sensing estimation of suspended sediment concentration based on
375 Random Forest Regression Model. *Journal of Remote Sensing*, 23(4): 756–772 (方馨蕊, 温兆飞, 陈吉龙, 吴胜军, 黄远洋, 马茂华. 2019. 随
376 机森林回归模型的悬浮泥沙浓度遥感估算. *遥感学报*, 23(4): 756-772) [DOI:10.11834/jrs.20197498]

377 Fang X, Zhang J, Yu X, Zhang S C, Kong D L, Wang X P, Ali S, Ullah H and Xu N. 2025. Accurate estimation of suspended sediment concentration
378 integrated remote sensing information and a novel stacking machine learning model. *Stochastic Environmental Research and Risk Assessment*,
379 39(4): 1517-1535 [DOI: 10.1007/s00477-025-02930-4]

380 Ghorbani M A, Khatibi R, Singh V P, Kahya E, Ruskeepaa H, Saggi M K, Sivakumar B, Kim S, Salmasi F, Kashani M H, Samadianfard S, Shahabi M
381 and Jani R. 2020. Continuous monitoring of suspended sediment concentrations using image analytics and deriving inherent correlations by
382 machine learning. *Scientific Reports*, 10: 8589 [DOI: 10.1038/s41598-020-64707-9]

383 Hou Y, Xing Q, Zheng X, Sheng D Z and Wang F T. 2023. Monitoring suspended sediment concentration in the yellow river estuary and its vicinity
384 waters on the basis of SDGSAT-1 multispectral imager. *Water*, 15(19): 3522 [DOI: 10.3390/w15193522]

385 Huang H, Li M Y, Chen Y H, Chen G, Liu H L, Xing Q G and Cai J N. 2021. Water quality retrieval by hyperspectral for city rivers in Pearl River
386 Estuary based on partial least squares regression. *Water Resources Protection*, 37(05): 36-42 (黄华, 李茂亿, 陈吟晖, 陈耿, 刘海龙, 刑前国,
387 蔡建楠. 2021. 基于 PLSR 的珠江口城市河流水质高光谱反演. *水资源保护*, 37(05): 36-42) [DOI: 10. 3880/j.issn. 1004 6933. 2021. 05. 006]

388 Jiang D, Matsushita B, Pahlevan N, Gurlin D, Lehmann M K, Fichot C G, Schalles J, Loisel H, Bing C, Zhang Y L, Alikas K, Kangro J, Uusoue M,

389 Onedrusek M, Greb S, Moses W J, Lohrenz S and O'Donnel D. 2021. Remotely estimating total suspended solids concentration in clear to
390 extremely turbid waters using a novel semi-analytical method. *Remote Sensing of Environment*, 258: 112386 [DOI: 10.1016/j.rse.2021.112386]

391 Jiji G W. An estimation and analysis method for suspended sediment and depth map of Gulf of Mannar. 2021. *The Imaging Science Journal*, 69(1-4):
392 92-102 [DOI: 10.1080/13682199.2022.2149382]

393 Jung S H, Kwon S, Seo I W and Kin J S. 2024. Comparison between hyperspectral and multispectral retrievals of suspended sediment concentration in
394 rivers. *Water*, 16(9): 1275 [DOI: 10.3390/w16091275]

395 Kwon S, Shin J, Seo I W, Noh H, Jung S H and You H. 2022a. Measurement of suspended sediment concentration in open channel flows based on
396 hyperspectral imagery from UAVs. *Advances in Water Resources*, 159: 104076 [DOI: 10.1016/j.advwatres.2021.104076]

397 Kwon S, Seo I W, Noh H and Kim B. 2022b. Hyperspectral retrievals of suspended sediment using cluster-based machine learning regression in shallow
398 waters. *Science of the Total Environment*, 833: 155168 [DOI: 10.1016/j.scitotenv.2022.155168]

399 Kwon S, Seo I W and Lyu S. 2023. Investigating mixing patterns of suspended sediment in a river confluence using high-resolution hyspe
400 rspectral imagery. *Journal of Hydrology*, 620: 129505 [DOI: 10.1016/j.jhydrol.2023.129505]

401 Li W and Wang X M. 2024. Application research of photoelectric sediment concentration meter in Luokou Hydrological Station of the Lower Yellow
402 River // Zhejiang Hydraulic Society. *Proceedings of the 2024 (3rd) Urban Water Conservancy and Flood Control Academic Seminar*. Hangzhou:
403 2024 (3rd) Urban Water Conservancy and Flood Control Academic Seminar (厉玮, 王向明. 2024. 光电测沙仪在黄河下游冻口水文站的应用
404 研究//浙江省水利学会.2024（第三届）城市水利与洪涝防治学术研讨会论文集. 杭州:2024（第三届）城市水利与洪涝防治学术研讨会)

405 Li Z Y, Dai J Q, Huang D, Zhang W, Li H M, Wang Z and Kang R. 2023. Application and prospects of satellite remote sensing monitoring technology
406 in water conservancy projects. *Advances in Water Science*, 34(5): 798-811 (李子阳, 戴济群, 黄对, 张文, 李涵曼, 王喆, 康芮. 2023. 水利工
407 程卫星遥感监测技术应用与展望. *水科学进展*, 34(05): 798-811) [DOI: 10.14042/j.cnki.32.1309.2023.05.014]

408 Marinho R R, Harmel T, Martinez J M and Junior N P F. 2021. Spatiotemporal dynamics of suspended sediments in the Negro River, Amazon Basin,
409 from in situ and Sentinel-2 remote sensing data. *ISPRS International Journal of Geo-Information*, 10(2): 86 [DOI: 10.3390/ijgi10020086]

410 Mohsen A, Kovács F and Kiss T. 2022. Remote sensing of sediment discharge in rivers using Sentinel-2 images and machine-learning algorithms.
411 *Hydrology*, 9(5): 88 [DOI: 10.3390/hydrology9050088]

412 Pan H Z, Yang L Z, Liu C Q and Fu Q H. 2019. Muli-model collaborative retrieval of suspend sediment concentration in the estuary of Pearl River

413 based on data assimilation. *Journal of Sediment Research*, 44(4):47-53 (潘洪洲, 杨留柱, 刘超群, 扶卿华. 2019. 基于数据同化的珠江河口悬
414 沙浓度多模型协同反演. *泥沙研究*, 44(4): 47-53) [DOI: 10.16239/j.cnki.0468-155x.2019.04.008]

415 Pham Q V, Ha N T T, Pahlevan N, Oanh L T, Nguyen T B and Nguyen N T. 2018. Using Landsat-8 images for quantifying suspended sediment
416 concentration in red river (Northern Vietnam). *Remote Sensing*, 10(11): 1841 [DOI: 10.3390/rs10111841]

417 Qiu Z, Liu D, Yan N, Yang C, Chen P P, Zhang C X and Duan H T. 2024. Improving the observations of suspended sediment concentrations in rivers
418 from Landsat to Sentinel-2 imagery. *International Journal of Applied Earth Observation and Geoinformation*, 134: 104209 [DOI:
419 10.1016/j.jag.2024.104209]

420 Song H S, LI D G, Chen B H and Liu P W. 2019. Research on ADCP multi-line depth-integrating flow measurements. *Yellow River*, 41(1): 18-22 (宋
421 海松, 李德贵, 陈宝华, 刘培旺. 2019. ADCP 多线积深式流量测验技术应用研究. *人民黄河*, 41(01):18-22) [DOI: 10.3969/j.issn.1000-
422 1379.2019.01.005]

423 Sun X H, Tian L Q, Fang H W, Walling D E, Huang L, Park E, Li D R, Zheng C M and Feng L. 2025. Changes in global fluvial sediment concentrations
424 and fluxes between 1985 and 2020. *Nature Sustainability*, 1-10. [DOI:10.1038/s41893-024-01476-7]

425 Wackerman C, Hayden A and Jonik J. 2017. Deriving spatial and temporal context for point measurements of suspended-sediment concentration using
426 remote-sensing imagery in the Mekong Delta. *Continental Shelf Research*, 147: 231-245 [DOI: 10.1016/j.csr.2017.08.007]

427 Wang J Y, Zhang W, Yang Y L and Wang Z S. 2025. Study on hysteresis characteristics between flood peaks and sediment peaks and their Causes in the
428 Tongguan-Xiaolangdi reach of the middle Yellow River (2002–2024), 1-13. (王江瑜, 张为, 杨泱柳, 王治森. 2025. 2002–2024 年黄河中游潼
429 关-小浪底河段洪峰沙峰异步特性及成因研究. *工程科学与技术*, 1-13)

430 Wang Y P, Hu X K, Li J G, Jiang S, Chen X F, Zhao L M and Chen H Z. 2024. Research on suspended matter concentration retrieval of SDGSAT-1
431 Satellite in Lake Taihu based on Hydrolight model. *SPACECRAFT RECOVERY & REMOTE SENSING*, 45(1): 174-186 (王雅萍, 胡雪可, 李
432 家国, 姜晟, 陈兴峰, 赵利民, 陈洪真. 2024. 基于 Hydrolight 模型的太湖 SDGSAT-1 卫星悬浮物浓度反演研究. *航天返回与遥感*, 45(1):
433 174-186) [DOI: 10.3969/j.issn.1009-8518.2024.01.015]

434 Wu X S and Feng X Q. 2024. Analysis and research on the distribution pattern of sediment concentration at Shizuishan hydrological station on the
435 Yellow River// Zhejiang Hydraulic Society. *Proceedings of the 2024 (12th) China Water Ecology Conference*. Hangzhou: 2024 (12th) China Water
436 Ecology Conference (吴小爽和冯晓强. 2024. 黄河石嘴山水文站含沙量分布规律分析研究//浙江省水利学会. 2024 (第十二届) 中国水生

态大会论文集. 杭州:2024 年 (第十二届) 中国水生态大会)

Xu H Q. 2005. A study on information extraction of water body with the modified normalized difference water index (MNDWI). *Journal of Remote Sensing*, (5): 589-595 (徐涵秋. 2005. 利用改进的归一化差异水体指数(MNDWI)提取水体信息的研究. *遥感学报*, (5): 589-595)

Yao R, Cai L, Liu J and Zhou M R. 2020. GF-1 satellite observations of suspended sediment injection of yellow river estuary, China. *Remote Sensing*, 12(19): 3126 [DOI: 10.3390/rs12193126]

Yang X P, Sokoletsky L, Wei X D and Shen F. 2017. Suspended sediment concentration mapping based on the MODIS satellite imagery in the East China inland, estuarine, and coastal waters. *Journal of Oceanology and Limnology*, 35 (1): 39-60 [DOI: 10.1007/s00343-016-5060-y]

Ying Z X, Li N G, Guo L L and Bao J Y. 2023. Development history and prospects of sediment monitoring technology at home and abroad// Chinese Hydraulic Engineering Society. *Proceedings of the 2023 China Hydraulic Academic Conference (Part I)*. Beijing: 2023 China Hydraulic Academic Conference (应周箫, 李聂贵, 郭丽丽, 鲍金玉. 2023. 国内外泥沙监测技术的发展历程及展望//中国水利学会.2023 中国水利学术大会论文集 (第一分册). 北京: 2023 中国水利学术大会)

Yu D F, Bian X D, Yang L, Zhou Y, An D, Zhou M S, Chen S L and Pan S Q. 2023. Monitoring suspended sediment concentration in the Yellow River Estuary from 1984 to 2021 using Landsat imagery and Google Earth Engine. *International Journal of Remote Sensing*, 44 (10): 3122-3145 [DOI: 10.1080/01431161.2023.2216849]

Zeng Q Z, Cao M S, Feng X Z, Liang F X, Chen X Z and Sheng W K. A study on spectral reflection characteristics of snow, ice and water in Northwest China. 1984. *Scientia Sinica (Chemica, Biologica, Agricultural, Medica, Sinica) (Series B)*, (04): 370-377 (曾群柱, 曹梅盛, 冯学智, 梁凤仙, 陈贤章, 盛文坤. 我国西北若干种冰、雪及水体反射光谱特性的研究. 1984. *中国科学(B 辑 化学 生物学 农学 医学 地学)*, (04): 370-377)

Zheng Z, Li L W, Xue C J, Wu S Y and Wang Y. 2025. Inversion method of offshore total suspended matter concentration based on SDGSAT-1 Satellite. *Remote Sensing Information*, 40(3): 172-180 (郑直, 李连伟, 薛存金, 吴世玉, 王宇. 2025. 基于 sdgsat-1 卫星的近海总悬浮物浓度反演方法. *遥感信息*, 40(3): 172-180) [DOI: 10.20091/j.cnki.1000-3177.2025.03.020]

Machine learning-based remote sensing inversion models for suspended sediment concentration and sensitivity analysis

Chen Na¹, Chen Hua¹, Li Lantao², Liu Renli¹, Zhong Haozhong³

1. State Key Laboratory of Water Resources Engineering and Management, Wuhan University, Wuhan 430072, China;

2. Hydrological Bureau of Yellow River Conservancy Commission, Zhengzhou 450004, China;

3. Powerchina Chengdu Engineering Corporation Limited, Chengdu 610000, China)

Abstract: Remote sensing inversion offers an efficient approach for monitoring suspended sediment concentration (SSC). However, its applicability for rivers with high and wide-range SSC requires further validation. This study aims to develop a machine learning model based on Recursive Feature Elimination with Cross-Validation and Random Forest (RFECV-RF) to retrieve SSC from multispectral imagery, using the Shizuishan and Wubu Station on the main stream of the Yellow River as study sites.

Sentinel-2 multispectral reflectance data were matched with in-situ daily SSC measurements from 2020 to 2024. After outlier removal using the Isolation Forest algorithm, datasets of 267 and 256 samples were obtained for the Shizuishan and Wubu Station, respectively. Spectral features, including single bands and band combinations (ratios and differences) from visible to shortwave infrared, were constructed. The RFECV method was applied to select optimal features, and a Random Forest model was subsequently built to establish the relationship between spectral features and SSC. Model performance was evaluated using R^2 , RMSE, and MAE. Spectral sensitivity was analyzed using regression fitting and slope decay ratios. Parameter sensitivity was assessed via the perturbation method by varying key parameters and evaluating corresponding changes in performance metrics using sensitivity coefficients. Uncertainty arising from data input, model parameters, and model structure was quantified using a Monte Carlo simulation framework with 500 iterations, and variance decomposition was applied to quantify the contribution of each uncertainty source.

The RFECV-RF model achieved R^2 values greater than 0.8 for both stations, with RMSE below 1.6 kg/m³, demonstrating reliable SSC estimation within the range of 0–44.5 kg/m³, though with underestimation of

extreme high SSC values. A strong nonlinear relationship was observed between spectral features and SSC. The B8A band, along with B7/B5 and B8–B11, remained sensitive across the entire SSC range, while features such as B3/B8, B4/B8, and B5/B6 tended to saturate as SSC increased. Parameter sensitivity analysis indicated that model performance was robust, with the most sensitive parameters being the subsampling rate and the number of features removed per iteration. Uncertainty analysis revealed that data input and model structure were the dominant sources of uncertainty, contributing over 98% of total variance, with larger uncertainties observed for low SSC values when modeling wide-range SSC. When applied to river reaches, the inversion model effectively captured temporal variations and spatial transport dynamics of SSC, with higher SSC values in the middle reaches compared to the upper reaches, consistent with actual observations.

The RFECV-RF model provides a reliable and effective approach for retrieving high and wide-range SSC in sediment-laden rivers using multispectral remote sensing. The identified key spectral features and quantified uncertainties offer valuable insights for improving inversion accuracy and support the development of automated SSC monitoring methods in complex river systems.

Key words: SSC; remote sensing inversion; RFECV; RF; Yellow River; spectral sensitivity; uncertainty

Supported by the National Key Research and Development Program of China (No.2023YFC3209101)