

基于机器学习的海陆过渡环境中陆源有机质分布预测模型

——从沉积模拟实验到地质应用

邵俊哲¹, 李阳², 汪勇³, 吴楠¹, 魏薇¹, 刘海²

1. 长江大学地球科学学院, 武汉 430100

2. 油气地球化学与环境湖北省重点实验室(长江大学资源与环境学院), 武汉 430100

3. 中国石化股份有限公司胜利油田分公司西部油气勘探项目部, 山东东营 257099

摘要 【目的】陆源海相烃源岩是我国多个近海盆地的主力烃源岩, 其差异性展布特征制约着烃源岩的分布预测精度和油气勘探成效。而陆源有机质的搬运沉积过程决定着陆源海相烃源岩质量和分布。【方法】运用水槽沉积模拟结合三维激光扫描技术, 从“正演”的角度动态记录并定量表征不同水体盐度条件下陆源分散有机质的搬运过程并运用机器学习算法建立总有机碳(Total Organic Carbon, TOC)预测模型。【结果】结果表明, 海陆过渡环境中的陆源有机质主要富集在三角洲前缘和前三角洲沉积相带内, 随着搬运距离的增大, 海陆过渡环境中的陆源有机质丰度呈现先增大后减小的趋势。在盐絮凝作用影响下, 咸水环境中陆源有机质搬运距离更靠近物源区, 沉积厚度更大。基于三种深度学习算法建立了实验条件下的TOC预测模型, 最终优选出基于随机森林算法的预测模型为最优模型。【结论】将实验条件下的TOC预测模型与地质条件相结合, 完成了崖南凹陷崖城组三角洲内烃源岩的TOC预测。结果显示, 崖南凹陷陆源有机质搬运距离可达50 km, 在距离物源区31 km左右处有机质富集程度最高。

关键词 海陆过渡相; 陆源有机质; 机器学习; 沉积模拟; 预测模型

第一作者简介 邵俊哲, 男, 2001年出生, 硕士研究生, 烃源岩评价与油气成藏机理, E-mail: shao413958@163.com

通信作者 李阳, 男, 讲师, 烃源岩沉积机制与分布规律, E-mail: lyugly@163.com

中图分类号 P512.2 文献标志码 A

0 引言

海陆过渡相是发育于滨海地区的海陆过渡或海陆混合沉积环境中的沉积相, 具体包括三角洲相、潟湖相、障壁岛相、潮坪相、河口湾相等。海陆过渡环境中发育海相烃源岩、煤系烃源岩和陆源海相烃源岩。在三角洲—浅海沉积体系中, 主要发育煤系烃源岩和陆源海相烃源岩两种类型, 具体的岩性以煤、泥岩、页岩为主^[1]。近年来的勘探实践表明, 陆源海相烃源岩是近海海域含油气盆地的主力烃源岩, 尤其是大气田发育的重要条件^[2-3]。全球的河流—三角洲—海湾体系与富含油气盆地的分布密切相关, 与三角洲相关的烃源岩和储层一直是学者们研究的重点^[4-7]。

无论是煤系烃源岩还是陆源海相烃源岩,其有机质均主要为源于湖泊水生生物和高等植物的陆源有机质^[8-9],陆源有机质的搬运特征和聚集规律决定了海陆过渡相烃源岩的空间展布。邓运华等^[1]研究南海周边三角洲的分布与烃源岩发育的关系指出三角洲控制了南海盆地海陆过渡相烃源岩的形成与分布,也影响着不同盆地的含油气类型和资源潜力。但海陆过渡相烃源岩陆源有机质的搬运沉积过程尚不明确,制约着海洋油气勘探成效。

鉴于海洋油气勘探“高风险、高科技、高投入”的三高特点^[10],要想取得海洋油气资源的更大突破,迫切需要明晰海域烃源岩形成机理和展布规律,建立高精度烃源岩分布预测模型。本论文通过开展沉积物理模拟实验,采取“正演”的手段模拟沉积物的搬运沉积过程。完成对沉积物理模拟实验的动态记录和定量表征是获取可靠结论的前提。应用三维激光扫描仪可在几分钟内创建环境或物件精确全面的三维图像,同时可获取实验过程中全方位点云数据。结合高清录像设备和 MATLAB 程序可实现沉积物理模拟实验的全程动态记录和精准定量表征,以期从陆源有机质搬运沉积角度探究海陆过渡相烃源岩发育机制。结合岩心实测数据和地震测井资料,通过支持向量机(Support Vector Machine, SVM)、随机森林(Random Forest, RF)和人工神经网络(Artificial Neural Network, ANN)三种机器学习算法可建立整个沉积体系范围内的烃源岩分布预测模型,可提高海域内优质烃源岩识别和预测精度,对降低海洋油气勘探风险和投入具有重要指导意义。

1 沉积模拟实验条件

1.1 装备与材料

本文涉及的沉积模拟实验均在长江大学沉积模拟实验室进行,主体实验装置为一个长 8.0 m 宽 6.0 m 的水槽,并将该水槽沿长轴方向均分为两个对比实验小水槽,除去分隔墙面厚度,每个水槽规格为长 8.0 m,宽 2.6 m(图 1)。

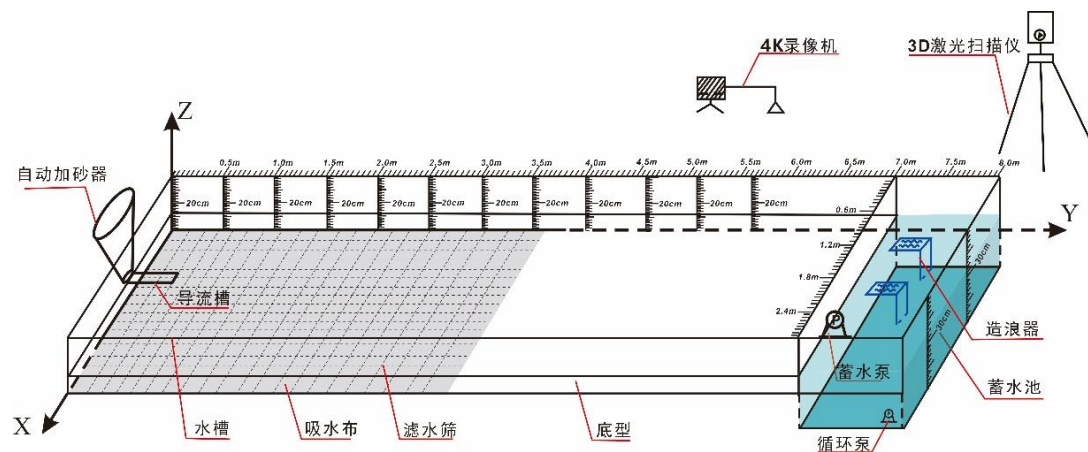


图 1 沉积模拟实验主体装置示意图

Fig.1 Schematic diagram of the main device for the sedimentary simulation experiment.

单物源条件下，物源口设置在水槽宽轴方向的中心位置，导流槽采用不锈钢材质，长度为 0.5 m；自动加砂器与导流槽相连，通过档位调节将来自储砂槽的物源定量且匀速地加入导流槽；水槽底部铺设一层 PVC 材质的方格滤水筛，可加快沉积模拟实验结束后的排水晾干进程；滤水筛上部铺设一层海绵吸水布，其作用是保证切片平整和加速排水；水槽边缘墙体上用油漆标注长度和深度便于读取岸线位置、三角洲长度及水深等数据；在 7.0~8.0 m 处通过活动底板下沉建立一个蓄水池，亦可模拟深海区；蓄水和水位的调整通过功率 600 W，13 m 扬程的蓄水泵实现；物源处水流供给和水动力控制通过功率 38W，3 m 扬程的循环泵实现；水槽顶部吊装 4K 高清录像机以记录沉积模拟过程；3D 激光扫描仪定期进行微地形扫描以获取沉积体厚度变化数据。造浪器为 CX-W3 型造浪泵，其磁吸于深水区，可根据需要调整波浪形式和强度，与物源供给方向正交造浪。

1.2 实验条件设定与取样测试

珠江口盆地白云凹陷恩平组海陆过渡相烃源岩母质来源主要为陆源有机质，孢粉相分析显示其陆源有机质主要在搬运和分选作用影响下分布于三角洲—浅海沉积体系^[11-12]。因此，白云凹陷恩平组三角洲的地质条件是进行陆源有机质沉积模拟的原型地质模型的可靠参考依据。

通过开展白云凹陷典型钻井岩心观察工作，通过岩心观察与刻画获取了白云凹陷恩平组地层岩性、岩石相、粒度特征等实际地质资料^[13-15]。结果表明，白云凹陷储层砂体以中细砂为主，南部地区发育部分中—粗砂岩^[15]，烃源岩以灰色—黑色泥岩为主，据此特征设计沉积模拟实验的加砂组成。考虑到在物源供给适中条件下，水动力越强其搬运粗粒能力越大，同时参考前人沉积模拟实验加砂组成特征^[16]，将不同水动力强度下加砂组成配置如表 1 所示。

表 1 沉积模拟实验沉积期次和加砂组成数据表

Table 1 Sedimentary periods of the sedimentary simulation experiment and composition data values

来水特征	加砂组成/%											
	第一沉积期				第二沉积期				第三沉积期			
	粗砂	中细砂	粉砂	泥	粗砂	中细砂	粉砂	泥	粗砂	中细砂	粉砂	泥
洪水期	25	40	27	8	25	40	27	8	33	37	23	7
中水期	20	42	28	10	20	42	28	10	30	35	30	5
枯水期	10	46	32	12	10	46	32	12	25	34	30	11

调研白云凹陷恩平组层序地层特征发现恩平组可划分为 3 个三级层序（SQ1、SQ2 和 SQ3），SQ1 和 SQ2 沉积时期主要以湖相三角洲沉积为主，SQ3 沉积期主要为半封闭的局限海沉积环境，在海侵影响下水体盐度升高且波浪对陆源有机质的搬运沉积产生影响^[13-14]。因此，基于白云凹陷恩平组层序地层划分结果和沉积环境特征，将部分沉积模拟实验的沉积期次设定为 3 期，第一沉积期为水退条件下的三角洲进积，不加有机质，沉积结束进行激光扫描；第二沉积期为一个完整的沉积旋回，第一沉积期结束后，缓慢水进进行三角洲退积，

岸线恢复至第一沉积期初始位置时,铺红色彩砂建立分隔层;第二沉积期时期定期定量加入相同有机质,实验半程和结束进行两次激光扫描以明确水体盐度对有机质的影响,沉积结束后在沉积体表面铺撒橘色彩砂建立分隔层;第三沉积期同样为一个完整的旋回,但第三沉积期咸水水槽全程造浪,先模拟波浪影响下的水进退积过程,再模拟波浪影响下的水退快速进积过程。第三沉积期两个水槽均定期定量加入有机质,实验半程和结束进行两次激光扫描以明确波浪在淡水咸水条件下对有机质分布影响。三维激光扫描仪对于部分没有砂体沉积的三角洲前缘、前三三角洲、浅海陆棚以及大陆坡区域中的陆源有机质分布特征和聚集规律的表征可发挥重大作用。无砂体沉积部位的沉积物几乎全部是泥质和陆源有机质,这些部位的三维激光扫描仪的扫描结果即完全反映了陆源有机质和泥质的分布特征,其差异性分布在平面上和剖面上均可做到精准定量表征。淡咸水条件下沉积模拟的具体实验条件参数如表 2 所示。

表 2 淡/咸水沉积模拟实验条件数据表
Table 2 Freshwater/saline data for sedimentary simulation experiment conditions

水体盐度/‰	实验期次	来水过程	历时/min	流量/(L/s)	有机质添加速率/(L/h)
0.7 (淡水)	第一沉积期	枯水	600	0.05	1.2
	第二沉积期	中水—洪水—中水—枯水	1 000	0.1~0.15~0.1~0.05	2.4~3.6~2.4~1.2
	第三沉积期	中水—洪水—中水—枯水	1 000	0.1~0.35~0.1~0.05	2.4~8.4~2.4~1.2
54 (咸水)	第一沉积期	枯水	600	0.06	1.2
	第二沉积期	中水—洪水—中水—枯水	1 000	0.1~0.15~0.1~0.05	2.4~3.6~2.4~1.2
	第三沉积期	中水—洪水—中水—枯水	1 000	0.1~0.35~0.1~0.05	2.4~8.4~2.4~1.2

根据模拟区地质条件,底型设计出河口区、三角洲沉积区、浅海陆架区、大陆坡、深海区 5 个主要沉积区域,模拟比例为 1: 10 000。河口区设计坡度为 3°, 三角洲沉积区范围为 0.5~3.0 m, 坡度由河口区逐渐过渡为 1°。陆架区范围为 3.0~6.0 m, 坡度为 3°, 大陆坡范围为 6.0~7.0 m, 坡度 12°。上述底型如图 2 所示。

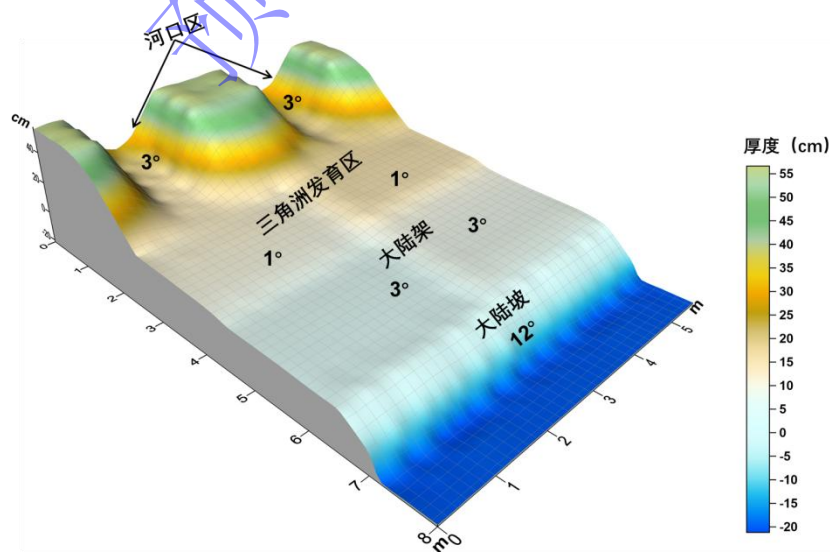


图 2 沉积模拟实验底型 3D 示意图

Fig.2 Three dimensional model of the bottom shape for the sedimentary simulation experiment
沉积模拟实验完成后,首先进行排水晾干工作,待排水完成后对沉积体进行横、纵剖面

的切片、取样和拍照工作。在三角洲沉积区以 0.3 m 为间距在宽轴方向上切取至多 8 条横剖面，以 0.6 m 为间距切取 3 条纵剖面（图 3a）。切片过程中，用高清单反相机拍摄剖面照片，记录不同部位的陆源有机质的沉积模式。制作厚 2 mm，高 15 cm，分度值为 0.5 mm 的透明玻璃取样管进行取样（图 3b），应用此取样管既能记录剖面沉积特征，又能读取有机质沉积厚度。浅海陆架区沉积的全部为泥质和有机质，以 0.2 m 为间距在宽轴方向上切取横剖面，以 0.6 m 为间距切取 3 条纵剖面，此部位主要读取有机质厚度和取样测试。实验过程中，详细测量并记录不同沉积期的底型坡度、加砂组成、加砂速率、水流流量、沉积时间、有机质加入速率、岸线位置以及三角洲长度和宽度等数据。

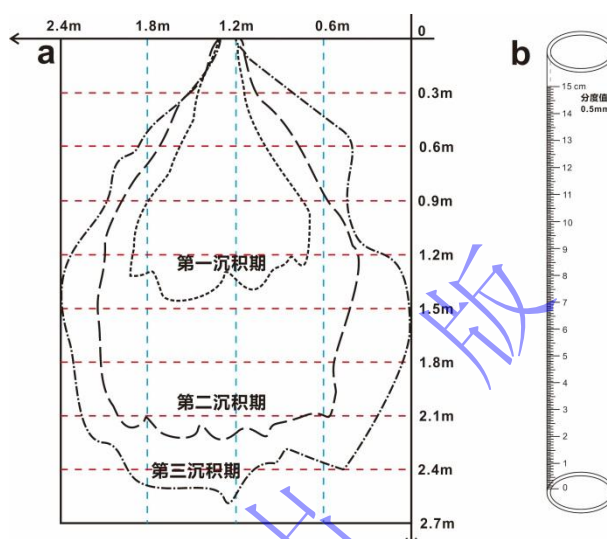


图 3 三角洲沉积区切片方案 (a) 与取样管 (b)
Fig.3 Slicing scheme (a) and sampling tube (b) for the delta sedimentary area

2 沉积模拟实验结果

笔者团队已经通过沉积模拟实验研究了不同沉积相带陆源有机质的分布模式，发现在三角洲平原内陆源有机质以互层型、透镜型、断续型分布；在三角洲前缘—前三角洲，则以厚度较大的条带型分布为主；坡度在三角洲演化的早期对陆源有机质分布影响较大，坡度越小，搬运距离越大；波浪影响下，在三角洲前缘陆源有机质向前搬运受阻，可在三角洲平原发育局部富集区^[17]。本文主要探讨水体盐度对陆源有机质搬运沉积的影响。实验记录显示，在相同的水动力和坡度条件下，淡水水体中有机质搬运距离较咸水水体更远，其扩散速度更快（图 4）。

从沉积体中央纵剖面上来看，有机质在不同沉积部位呈以下四种分布模式：①断续型，颜色较浅，呈断续状分布，厚度平均值在 2 mm，有机碳含量比较低，是有机质在水动力较弱时，在低地势部位沉积形成，主要分布在三角洲平原；②透镜型，颜色浅，形状呈现不对称透镜状，由河道摆动和切割作用导致，分布不规则，主要集中在三角洲平原漫溢部位。③

互层型，数量多，间距小，近似平行分布，有间断，厚度薄，颜色浅，集中分布在三角洲平原部位；④条带型，数量较少，间距大，常见斜交条带，厚度大，颜色深，有机碳含量高，主要分布在三角洲前缘，陆源有机质搬运到三角洲前缘与湖水交互，水动力迅速下降，在前缘部位快速沉积，呈现条带状分布。在同一沉积期的三角洲沉积相带内，随着搬运距离的增加，有机质的沉积模式呈现由“断续型”向“互层型”再向“条带型”转变的趋势（图5）。上述分布模式与陡/缓坡实验条件下的沉积模拟实验具有相似特征^[17]。

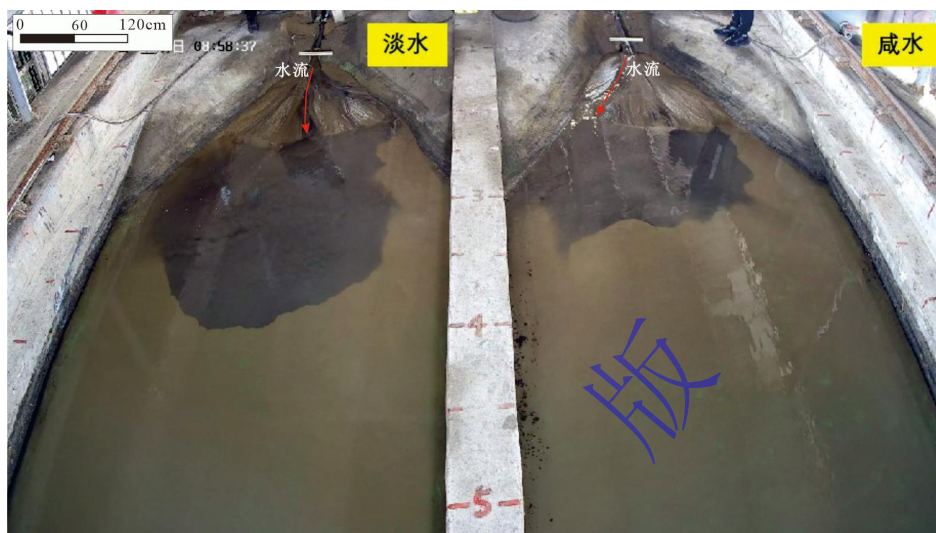


图4 淡咸水条件下有机质搬运情况对比

Fig.4 Comparison of organic matter transport under fresh/saline water conditions

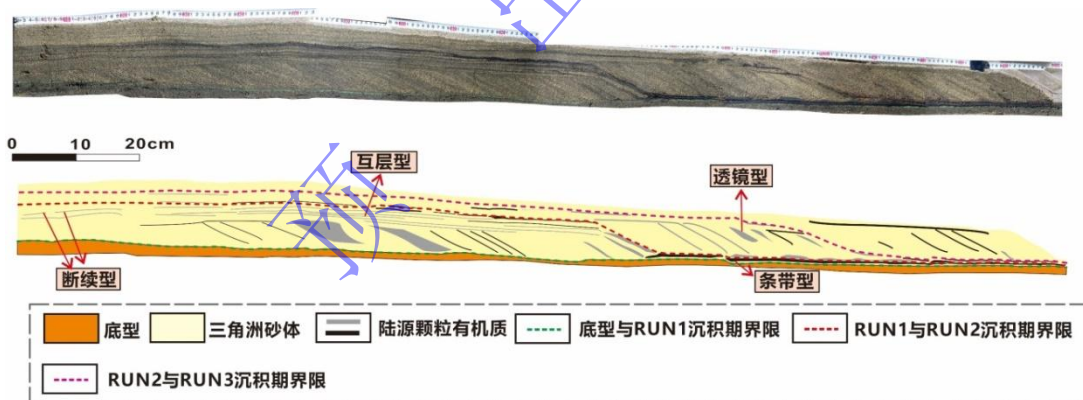


图5 淡水条件下三角洲纵剖面刻画图

Fig.5 Vertical profile characterization of the delta under freshwater conditions

3 基于实验结果的陆源有机质分布预测模型构建

3.1 陆源有机质分布的影响因素

尽管大量研究已经证明中国近海盆地诸多凹陷内的油气主要来源于陆源海相烃源岩^[1,3,5,9,11]，但因盆地地质条件复杂多样，陆源有机质搬运路径和沉积特征不明确，含油气盆地中陆源海相烃源岩的发育机制和分布规律一直未能形成统一认识，导致近年来我国近海盆地

未能实现油气勘探的重大突破。从古生产力和保存条件方面探讨有机质富集机制的同时，我们应重视烃源岩主要母质来源的物理搬运和沉积过程对烃源岩分布的影响。存在陆源有机质的搬运和沉积是海陆过渡相烃源岩发育的必要前提。因此，研究陆源有机质的搬运沉积过程，明确陆源有机质的机械沉积分异作用和沉积分布模式是实现海陆过渡相烃源岩分布预测的必要步骤。

陆源有机质机械沉积分异作用受到沉积物粒度、沉积速率、水动力条件、物源供给、水体盐度等诸多因素的控制^[18-26]。目前，已有相关研究对地形坡度、水动力强度、波浪、物源粒度等搬运条件对陆源有机质分布的影响做了系统研究^[9,17]，但水体盐度对陆源有机质搬运沉积的影响尚未有讨论。我国近海盆地往往受到海侵的影响，海侵作用会导致水体盐度发生变化进而影响陆源有机质的絮凝沉降速率和搬运距离，这将一定程度上影响陆源海相烃源岩的分布。因此，本文开展了不同水体盐度条件下的陆源分散有机质沉积模拟实验，探讨水体盐度对陆源有机质搬运沉积过程的影响。

基于三维激光扫描获取的数据，对第二沉积期的水进和水退两个过程的有机质沉积区的沉积增量进行三维定量表征。结果显示，淡水条件下，有机质展布范围大，沉积厚度薄；咸水条件下，有机质展布范围小，沉积厚度大（图6）。

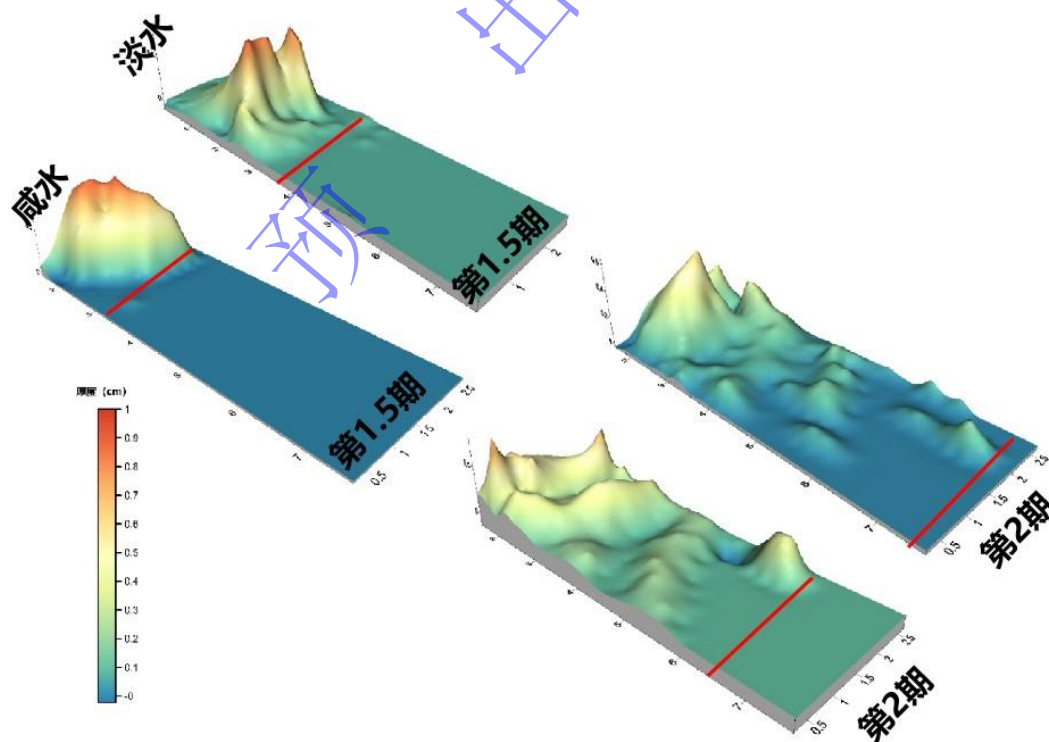


图6 淡咸水条件下有机质沉积厚度分布

Fig.6 Distribution of organic matter thickness under fresh/saline water conditions

从 TOC 的平面展布来看，在三角洲规模相同，坡度和水动力条件一致情况下咸水三角

洲的有机质主要富集区距离物源更近（图7）。综上所述，在水深较浅，水动力较弱的情况下，随着盐度的增加，有机质搬运距离变小。推测在咸水条件下，黏土矿物和有机质吸附的大量阴离子与水中的阳离子结合，形成颗粒的物理絮凝也可称之为电凝或盐絮凝，使沉积物形成团块并克服浮力而快速沉降^[27-28]。

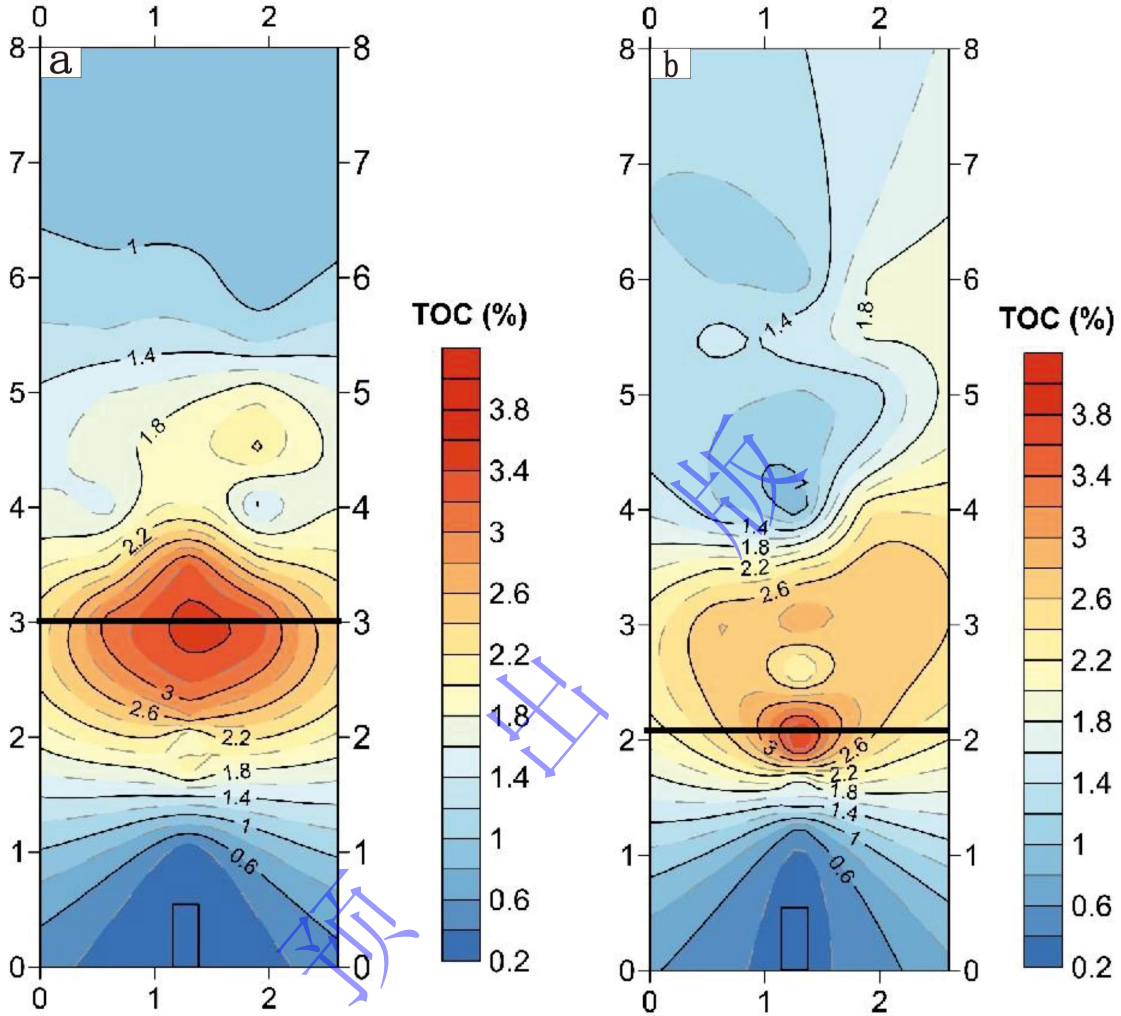


图7 淡咸水条件下 TOC 平面分布
(a) 淡水条件； (b) 咸水条件

Fig.7 Plane distribution of total organic carbon (TOC) under fresh/saline water conditions

3.2 机器学习算法选择与模型构建

前人及本文研究表明，陆源有机质的富集程度和搬运距离并非线性关系，常用的基于统计学分析的线性回归预测存在一定的局限性。机器学习在非线性相关数据处理和数据规律总结方面具有显著的优势，应用机器学习建立的预测模型通常具有较高的精度^[28-33]。基于沉积模拟实验和机器学习算法建立陆源有机质分布预测模型将为海洋油气勘探中的有利区带优选和资源评价工作提供重要依据。

结合实验过程中的地形坡度、水动力条件、沉积物粒度、水体盐度等多种地质因素的准

确定量表征^[17]，通过机器学习算法可建立整个沉积体系范围内的烃源岩分布预测模型。实现从点到线，从线到面，从面到体的空间范围内的烃源岩分布预测，其在油气资源评价和指导油气资源勘探领域具有广阔的应用空间。

本文选用沉积模拟实验中实测 TOC 作为标签变量，结合取样时对样品信息和实时实验条件的准确记录，采用支持向量机（SVM）、随机森林（RF）和人工神经网络（ANN）三种机器学习算法进行 TOC 预测模型的建立，具体的模型建立流程如图 8 所示。

本模型选取搬运距离、沉积相带、水动力条件（水流流量和造浪与否）、坡度、三角洲规模、物源粒度（以不同粒径颗粒占比表征）、水体盐度、水位变化率、有机质供给速率 9 个参数进行训练。选取所有数据中的 75% 进行训练，剩余的 25% 进行测试。通过不断调参，选取每种算法中精度最高的模型作为该算法的最优模型。

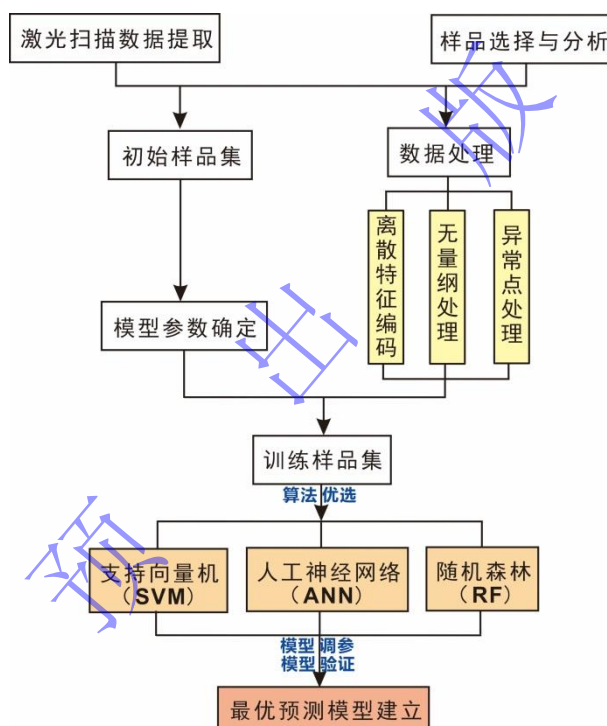


图 8 有机质分布预测模型建立流程图

Fig.8 Process for establishing a prediction model for TOC distribution

每种算法的输入特征可选择离散特征编码与基于经验赋值的数据，标准化和不标准化的数据以及去异常值和不去异常值的数据进行训练。离散特征编码的数据处理工作主要用于区分样品所属沉积相带的划分。由于不同的沉积相带中陆源有机质富集程度差异较大，在进行预测模型建立时，要对不同的沉积相带进行编码以适用于软件平台。本文采取独热编码（one-hot）的方式完成沉积相带的划分。对每一个样品而言，只对其所在沉积相带赋值为 1，其余沉积相属性均赋值为 0。这样所有样品的沉积相将通过 0 和 1 表示。这种方法避免了人

为经验赋值的主观性误差,可有效对沉积相带参数进行区分。基于经验的赋值是基于传统地质理论,采用专家赋值的方法给不同的沉积相带赋值,值的高低以其可能的有机质富集程度表征,最大值为4,最小值为1。如前三角洲通常粒度较细,有机质相对富集,可以基于经验赋值为4,而三角洲平原则以储层为主,烃源岩不发育,有机质富集程度不高,基于经验赋值为1。

数据的无量纲处理即将不同类型参数进行归一化处理,这样可有效避免不同类型参数数量级差异较大带来的不便。在参数数值归一化处理时会出现以下两种情况:

第一种情况是参数值与 TOC 存在正相关关系,此时在数据归一化时应选用以下公式:

$$Y(i, j) = \frac{X(i, j) - \min X(i, j)}{\max X(i, j) - \min X(i, j)} \quad (i = 0, 1, 2, \dots, m; j = 1, 2, \dots, n) \quad (1)$$

第二种情况是参数值与 TOC 存在负相关关系,此时在数据归一化时应选用以下公式:

$$Y(i, j) = \frac{\max X(i, j) - X(i, j)}{\max X(i, j) - \min X(i, j)} \quad (i = 0, 1, 2, \dots, m; j = 1, 2, \dots, n) \quad (2)$$

上述两式中: $Y(i, j)$ 是对应的归一化数值; $X(i, j)$ 是样本的参数值; $\max X(i, j)$ 是样本参数的最大值; $\min X(i, j)$ 是样本参数的最小值。

异常值检验和剔除是为了避免样品分析测试误差较大对模型精度的影响,在模型训练前先将异常值筛选出来并剔除。本项目拟选用箱型图检验离群值的方法完成异常值处理。将样品数据进行统计分析绘制数据箱型图,通过四分位检测可快速识别异常值。

支持向量机算法的参数调整范围设置为①C 调整, γ 默认; ② γ 调整, C 默认; ③C, γ 调整三种; 神经网络参数调整范围设置为① α 调整, hidden_layer 默认; ② hidden_layer 调整, α 默认; ③ α , hidden_layer 调整三种; 随机森林算法参数调整范围设置为① max_depth 调整, minsamplesleaf 默认; ② minsamplesleaf 调整, max_depth 默认; ③ max_depth , minsamplesleaf 调整三种。这样,每种算法的输入特征有 18 种方式,从中优选精度最高的模型即可。本文选用 10 折交叉验证的方法(k fold cross validation)。将训练集分成不相交的 10 个子集,每次从分好的子集里面,选择 9 个作为训练集,剩下的 1 个作为验证集。每次把训练好的模型放在验证集上,得到准确率,计算 10 次求得平均值,作为该模型的最终精度。预测精度的评价指标为 RMS 和 R^2 。其计算公式为:

$$RMS = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{\sum_{i=1}^m (y_i - \bar{y}_i)^2} \quad (4)$$

3.3 优选模型与验证

本文从 525 个样品数据中选取 420 个样品数据进行训练，105 个数据进行训练测试。每种算法设置了 16 种变量组合形式，设置 3 种调参模式，共计 144 个模型，其中每种算法 48 种模型，从每种算法中选取精度最高的一种模型。最终，SVM 算法的最优模型精度为 72.94%，ANN 算法的最优模型精度为 72.2%，随机森林算法的最优模型精度为 78.23%。每种算法的最优模型的输入特征见表 3。

最优模型为基于随机森林算法的以去异常值和基于经验的相带赋值为输入特征的预测模型，其预测精度为 78.22%，同时确定，对有机质富集程度影响较大的因素为沉积相带、有机质供给速率和搬运距离（图 9）。

表 3 三种算法最优模型参数和输入特征

Table 3 Optimal model parameters and input features of the three algorithms

模型名称	模型参数	模型精度	输入特征
随机森林 (RF)	max_depth=11 n_estimators:500 alpha:[1],	0.7822	去异常值+基于经验的相带参数
神经网络 (ANN)	'solver':['lbfgs'], 'max_iter':[10000]	0.7467	相带独热编码+标准化
支持向量机 (SVM)	C:[10] 'max_iter':[10000]	0.7283	相带独热编码+标准化

注：max_depth 为决策树最大深度；n_estimators 为随机森林算法中森林中的树木的数量；alpha 为正则化参数，alpha 它的目的是防止过拟合；solver 为优化器，用来调整群众参数；max_iter 为迭代次数，C 为惩罚系数。

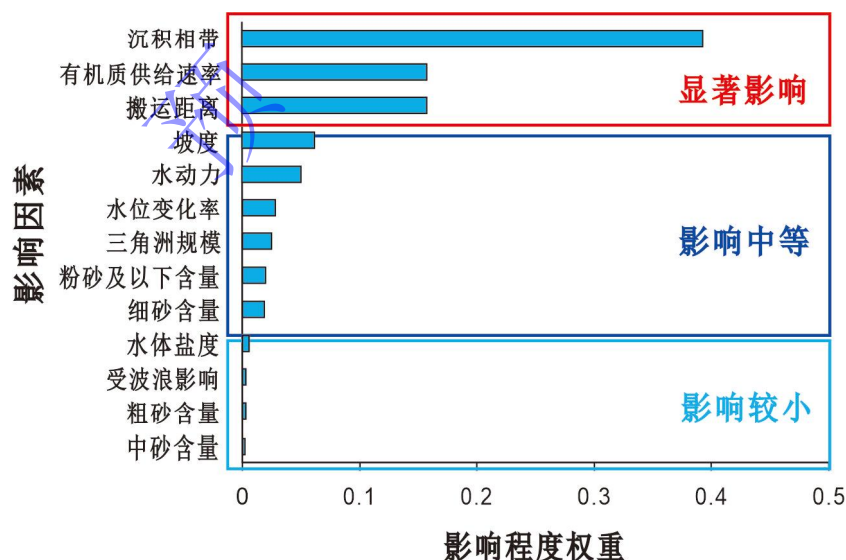


图 9 不同参数对 TOC 分布的影响程度

Fig.9 Impact of different parameters on TOC distribution

4 基于地质条件的模型建立与应用

基于实验条件下的 TOC 预测模型建立后并验证预测精度良好, 我们选取实验条件下的不同算法的最优模型输入特征开展了地质条件下的 TOC 预测模型的建立与应用工作。由于地质条件下的预测模型建立需要足够的实测数据作为训练集, 白云凹陷恩平组目前难以收集到足够的训练数据, 因此选定与白云凹陷三角洲沉积环境相近, 同样以陆源海相烃源岩为主的琼东南盆地崖南凹陷崖城组三段烃源岩为预测层位。

以崖南凹陷崖城组沉积相平面展布图为底图, 以 A 井为原点, 向东和向南设定为 X 轴和 Y 轴, 基于比例尺确定每口井在研究区的相对位置坐标。然后根据收集到的 12 口钻井分析测试数据、井位所在区域古地形坡度和单井沉积相确定 TOC 预测模型建立所需的沉积相、岩性、岩石颜色、TOC、深度和地形坡度等参数。由于崖南凹陷同层位中不同颜色(如灰色和黑色)烃源岩实测 TOC 差异较大, 模型训练时将岩石颜色作为一个输入参数可有效避免不同颜色烃源岩预测结果的混淆, 也可提高预测精度。最终选定搬运距离(位置 X 和位置 Y)、沉积相带、岩性、岩石颜色、坡度、TOC 为标签变量, 通过支持向量机(SVM)、随机森林(RF)和人工神经网络(ANN)三种算法进行机器学习, 并确定精度最高的预测模型。最优模型确定后, 选取部分数据对三种算法进行了模型精度验证, 验证结果如表 4 所示。

结果显示, ANN 算法的预测准确度(预测值和实测值之差与实测值的比值)范围为 -68.91%~98.05%, 平均预测准确度为 50.69%, 准确度大于 70%占比为 45.45%, 准确度大于 90%占比为 22.73%, 异常值占比为 18.18%; SVM 算法的预测准确度介于 7.96%~99.77%, 平均预测准确度为 76%, 准确度大于 70%占比为 72.73%, 准确度大于 90%占比为 45.45%, 异常值占比为 4%; RF 算法的预测准确度介于 40.77%~97.24%, 平均预测准确度为 77%, 准确度大于 70%占比 77.27%, 准确度大于 90%占比为 27.27%, 异常值占比为 0%。综合分析, 最优算法为随机森林算法。

应用地震资料可换算出该层位古地形坡度^[34-35], 确定预测区坐标位置, 便可上述预测模型预测海域内少井/无井区某一位置不同沉积相带内烃源岩的 TOC (图 10a)。本文以崖南凹陷崖城组三段烃源岩为例, 先完成单条剖面的 TOC 预测, 然后在研究区网格选取多条地震剖面开展 TOC 预测, 最终完成了崖南凹陷崖城组烃源岩 TOC 平面预测(图 10b)。

从主三角洲平面展布图中拉取一条剖面(图 10b), 观察沿着搬运方向上其 TOC 的分布特征发现, 三角洲前缘和前三三角洲为分散有机质主要富集区, 分散有机质沿着物源方向的搬运距离可达 50 km, 在距离物源区 31 km 左右处有机质富集程度最高(图 11)。

表 4 预测模型验证数据表

Table 4 Data for prediction model validation

井号	深度/m	X 坐标	Y 坐标	岩性	岩石颜色	坡度	沉积相	实测 TOC/%	TOC_RF/%	误差	准确度/%	TOC_SVM (%)	误差/%	准确度/%	TOC_ANN/%	误差	准确度/%
B	3 718.56	2.14	8.82	粉砂质泥岩	灰色	1.74	分流河道	0.34	0.37	0.025 3	92.56	0.44	0.10	70.53	0.40	0.063 5	81.32
B	3 733.00	2.14	8.82	粉砂质泥岩	灰色	1.74	泛滥平原	0.65	0.69	0.043 4	93.32	0.66	0.01	98.66	1.38	0.725 3	-11.58
C	3 924.50	4.57	9.12	泥岩	灰色	1.74	分流河道	0.43	0.18	0.254 7	40.77	0.34	0.09	80.05	0.40	0.026 5	93.84
C	3 930.00	4.57	9.12	煤层	黑色	1.74	泛滥平原	30.97	38.21	7.241 3	76.62	31.86	0.89	97.12	33.30	2.334 7	92.46
C	4 034.50	4.57	9.12	碳质泥岩	黑色	1.74	泛滥平原	19.06	13.79	5.270 5	72.35	11.87	7.19	62.28	8.00	11.055 2	42.00
C	4 049.50	4.57	9.12	碳质泥岩	黑色	1.74	分流间湾	5.01	5.63	0.615 1	87.72	6.03	1.02	79.64	6.97	1.955 3	60.97
C	4 067.56	4.57	9.12	煤层	黑色	1.74	分流间湾	43.50	54.42	10.91 9	74.90	43.40	0.10	99.77	63.48	19.975 4	54.08
C	4 070.50	4.57	9.12	泥岩	黑色	1.74	分流间湾	2.06	2.81	0.748 8	63.65	3.96	1.90	7.96	4.73	2.665 5	-29.39
C	4 070.60	4.57	9.12	碳质泥岩	黑色	1.74	分流间湾	6.39	5.83	0.563 8	91.18	6.14	0.25	96.15	7.02	0.631 3	90.12
C	4 137.50	4.57	9.12	泥岩	灰色	1.74	泛滥平原	1.96	1.52	0.441 3	77.48	1.86	0.10	94.88	2.51	0.545 8	72.15
D	3 820.97	7.86	12.50	泥岩	灰色	1.50	分流间湾	0.83	0.99	0.156 6	81.13	1.10	0.27	67.52	0.58	0.252 9	69.53
D	3 821.30	7.86	12.50	煤层	黑色	1.50	分流间湾	52.10	67.19	15.09 2	71.03	43.86	8.24	84.19	63.73	11.629 2	77.68
D	3 822.20	7.86	12.50	碳质泥岩	黑色	1.50	分流间湾	6.68	6.14	0.545 0	91.84	6.78	0.10	98.50	7.26	0.583 1	91.27
D	3 830.12	7.86	12.50	泥岩	灰色	1.50	分流河道	0.18	0.24	0.060 6	66.33	0.28	0.10	44.00	0.40	0.223 5	-24.17
D	3 946.86	7.86	12.50	泥岩	灰色	1.50	泛滥平原	0.74	1.01	0.269 5	63.58	0.93	0.19	74.28	1.99	1.249 9	-68.91
E	4 126.00	6.86	9.26	碳质泥岩	黑色	1.70	分流间湾	6.54	6.36	0.180 2	97.24	6.64	0.10	98.47	9.84	3.304 3	49.48
E	4 126.99	6.86	9.26	泥岩	灰色	1.70	分流间湾	1.62	1.37	0.245 4	84.85	1.47	0.15	90.94	1.59	0.031 6	98.05
E	4 127.00	6.86	9.26	泥岩	黑色	1.70	分流间湾	4.23	4.52	0.286 1	93.24	4.37	0.14	96.69	5.72	1.489 8	64.78
F	39 70.02	9.00	13.24	煤层	黑色	1.40	分流间湾	45.21	56.01	10.797 0	76.12	46.91	1.70	96.23	65.57	20.361 1	54.96
F	39 97.45	4.57	9.12	煤层	黑色	1.74	泛滥平原	28.16	38.27	10.105 0	64.11	32.52	4.36	84.51	33.68	5.520 0	80.40
F	38 07.60	9.00	13.24	泥岩	灰色	1.40	分流间湾	0.79	1.00	0.206 6	73.85	1.29	0.50	36.54	0.63	0.164 0	79.24

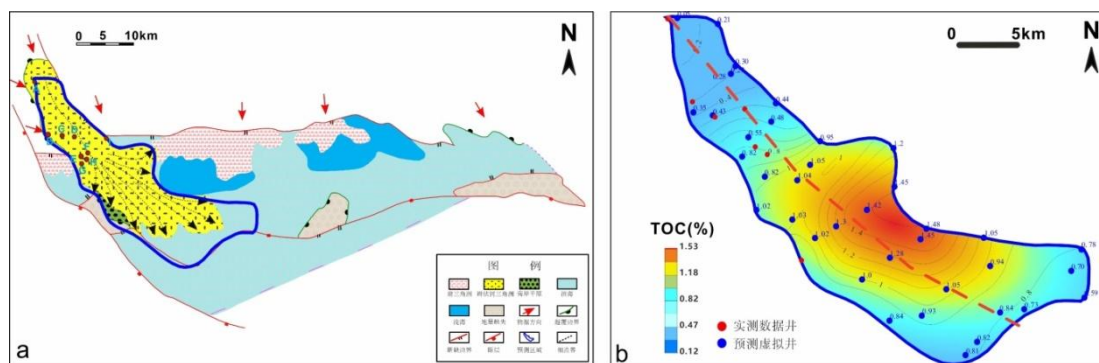


图 10 崖南凹陷主三角洲泥岩 TOC 分布预测及平面展布特征
(a) TOC 分布预测; (b) TOC 平面展布

Fig.10 Prediction of TOC distribution and planar distribution characteristics of mudstone in the main delta of the Yanan depression

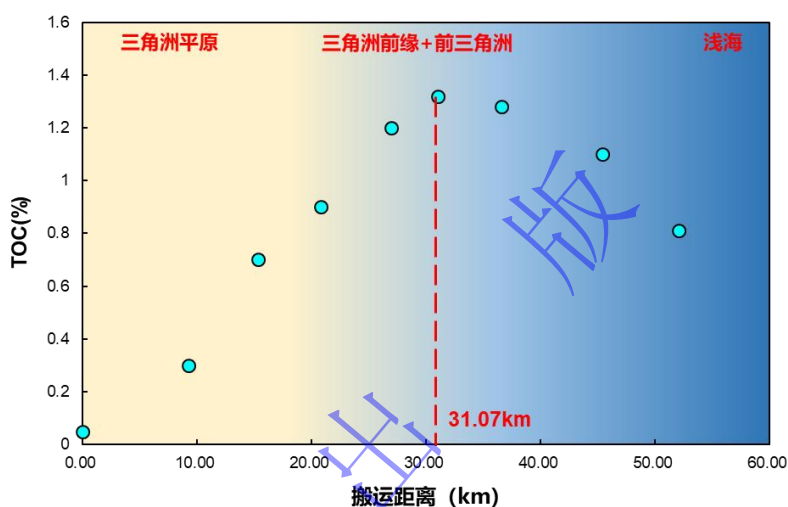


图 11 崖南凹陷主三角洲泥岩 TOC 与陆源有机质搬运距离的关系

Fig.11 The relationship between TOC of mudstone in the main delta of Yanan Depression and the distance of terrestrial organic matter transport

5 结论

(1) 在不同水体盐度的沉积模拟实验条件下, 咸水条件时, 盐絮凝作用加快了有机质的沉降速率, 使得咸水环境下陆源有机质搬运距离更靠近物源区, 其沉积厚度更大。淡水条件下, 有机质展布范围大, 沉积厚度薄。陆源在剖面上以断续型、互层型、透镜型和条带型四种分布模式存在, 断续型、互层型、透镜型主要分布于三角洲平原, 而条带型有机质主要分布于三角洲前缘和前三角洲。在同一沉积期的三角洲沉积相带内, 随着搬运距离的增加, 有机质的沉积模式呈现由“断续型”向“互层型”再向“条带型”转变的趋势。

(2) 基于沉积模拟实验建立的 TOC 预测模型显示, 对有机质富集程度影响较大的因素为沉积相带、有机质供给速率和搬运距离。陆源有机质主要富集在三角洲前缘和前三角洲沉积相带内。富含陆源有机质的物源供给速率越大, 烃源岩中有机质丰度越高。三角洲—浅海

沉积体系中，随着搬运距离的增大，陆源有机质丰度呈现先增大后减小的趋势。

(3) 基于随机森林、神经网络和支持向量机三种深度学习算法建立了实验条件下的 TOC 预测模型，优选出基于随机森林算法的预测模型具有最高的预测准确度和稳定性。基于实验条件下最优模型的输入特征，建立了地质条件下的 TOC 预测模型。地质预测模型预测了琼东南盆地崖南凹陷崖城组无井、少井区烃源岩 TOC 平面分布，取得了较好的效果。结果显示，崖南凹陷崖三段沉积期陆源有机质搬运距离可达 50 km，在距离物源区 31 km 左右处有机质富集程度最高。

致谢 在实验进行得到中海油专家李友川老师、李威老师等专家的指导，在数据获取及三维激光扫描技术应用方面得到冯文杰老师、高飞翔同学指导，在实验过程中得到罗顺社老师、刘忠保老师、胡光明老师、吕奇奇老师指导，在此一并表示衷心的感谢！

参考文献 (References)

- [1] 邓运华, 兰蕾, 李友川, 等. 论三角洲对南海海相油气田分布的控制作用[J]. 石油学报, 2019, 40(增刊2): 1-12. [Deng Yunhua, Lan Lei, Li Youchuan, et al. On the control effect of deltas on the distribution of marine oil and gas fields in the South China Sea[J]. Acta Petrolei Sinica, 2019, 40(Suppl.2): 1-12.]
- [2] 李友川, 邓运华, 张功成. 中国近海海域烃源岩和油气的分带性[J]. 中国海上油气, 2012, 24(1): 6-12. [Li Youchuan, Deng Yunhua, Zhang Gongcheng. Zoned distribution of source rocks and hydrocarbon offshore China[J]. China Offshore Oil and Gas, 2012, 24(1): 6-12.]
- [3] 张功成, 王琪, 苗顺德, 等. 中国近海海陆过渡相烃源岩二元分布模式: 以珠江口盆地白云凹陷为例[J]. 天然气地球科学, 2014, 25(9): 1299-1308. [Zhang Gongcheng, Wang Qi, Miao Shunde, et al. The duality distribution pattern of marine-continental transitional hydrocarbon source rocks: A case study from Baiyun Sag in Pearl River Mouth Basin, China Offshore[J]. Natural Gas Geoscience, 2014, 25(9): 1299-1308.]
- [4] 李燕, 邓运华, 李友川. 河流—三角洲体系微相控烃及机理研究: 以珠江口盆地恩平组煤系烃源岩为例[J]. 现代地质, 2021, 35(4): 1065-1077. [Li Yan, Deng Yunhua, Li Youchuan. Characteristics and mechanism of source rock development controlled by sedimentary microfacies in river-delta system: Case study of coal-measure source rock in Enping Formation in Pearl River estuary Basin[J]. Geoscience, 2021, 35(4): 1065-1077.]
- [5] Wang D D, Zhang G C, Li Z X, et al. The development characteristics and distribution predictions of the Paleogene coal - measure source rock in the Qiongdongnan Basin, northern South China Sea[J]. Acta Geologica Sinica (English Edition), 2021, 95(1): 105-120.
- [6] El-Gendy N H, Radwan A E, Waziry M A, et al. An integrated sedimentological, rock typing, image logs, and artificial neural networks analysis for reservoir quality assessment of the heterogeneous fluvial-deltaic Messinian Abu Madi reservoirs, Salma field, onshore East Nile Delta, Egypt[J]. Marine and Petroleum Geology, 2022, 145: 105910.
- [7] El Diasty W S, Moldowan J M, Peters K E, et al. Organic geochemistry of possible Middle Miocene–Pliocene source rocks in the west and northwest Nile Delta, Egypt[J]. Journal of Petroleum Science and Engineering, 2022, 208: 109357.
- [8] 李丹, 杨香华, 常吟善, 等. 澳大利亚北卡那封盆地中上三叠统 Mungaroo 三角洲陆源有机质分布特征[J]. 古地理学报, 2014, 16(2): 193-204. [Li Dan, Yang Xianghua, Chang Yinshan, et al. Distribution of the Upper Triassic terrigenous organic matter in Mungaroo delta of North Carnarvon Basin, Australia[J]. Journal of Palaeogeography, 2014, 16(2): 193-204.]
- [9] 屈童, 高岗, 徐新德, 等. 三角洲—浅海沉积体系泥质沉积规律模拟实验: 以琼东南盆地崖南凹陷为例[J]. 岩性油气藏, 2022, 34(1): 24-33. [Qu Tong, Gao Gang, Xu Xinde, et al. Simulation experiment of argillaceous sedimentary law of

- delta-shallow sea sedimentary system: A case study of Yanan Sag, Qiongdongnan Basin[J]. *Lithologic Reservoirs*, 2022, 34(1): 24-33.]
- [10] 谢玉洪. 中国海洋石油总公司油气勘探新进展及展望[J]. *中国石油勘探*, 2018, 23(1): 26-35. [Xie Yuhong. New progress and prospect of oil and gas exploration of China National Offshore Oil Corporation[J]. *China Petroleum Exploration*, 2018, 23(1): 26-35.]
- [11] 朱伟林. 中国近海新生代含油气盆地古湖泊学与烃源条件[M]. 北京: 地质出版社, 2009. [Zhu Weilin. Paleolimnology and source rock studies of Cenozoic hydrocarbon-bearing offshore basins in China[M]. Beijing: Geology Press, 2009.]
- [12] Tyson R V, Follows B. Palynofacies prediction of distance from sediment source: A case study from the Upper Cretaceous of the Pyrenees[J]. *Geology*, 2000, 28(6): 569-571.
- [13] 曾智伟, 杨香华, 朱红涛, 等. 白云凹陷恩平组沉积晚期大型三角洲发育特征及其意义[J]. *地球科学*, 2017, 42(1): 78-92. [Zeng Zhiwei, Yang Xianghua, Zhu Hongtao, et al. Development characteristics and significance of Large Delta of Upper Enping Formation, Baiyun Sag[J]. *Earth Science*, 2017, 42(1): 78-92.]
- [14] 曾智伟, 朱红涛, 杨香华, 等. 珠江口盆地白云凹陷恩平组物源转换及沉积充填演化[J]. *地球科学*, 2017, 42(11): 1936-1954. [Zeng Zhiwei, Zhu Hongtao, Yang Xianghua, et al. Provenance transformation and sedimentary evolution of Enping Formation, Baiyun Sag, Pearl River Mouth Basin[J]. *Earth Science*, 2017, 42(11): 1936-1954.]
- [15] 郭帅, 杨海长, 曾清波, 等. 白云凹陷恩平组南部物源研究及其油气地质意义[J]. *海洋地质前沿*, 2020, 36(6): 56-63. [Guo Shuai, Yang Haizhang, Zeng Qingbo, et al. Provenance of Enping Formation in the southern Baiyun Sag, Pearl River Mouth Basin and its implications for petroleum geology[J]. *Marine Geology Frontiers*, 2020, 36(6): 56-63.]
- [16] 张春生, 刘忠保. 现代河湖沉积与模拟实验[M]. 北京: 地质出版社, 1997. [Zhang Chunsheng, Liu Zhongbao. Modern river and Lake deposits and simulated experiment[M]. Beijing: Geological Publishing House, 1997.]
- [17] 刘海, 徐耀辉, 李阳, 等. 海陆过渡相三角洲沉积体系陆源有机质沉积特征及其影响因素的模拟实验研究[J]. *沉积学报*, 2024, 42(1): 251-265. [Liu Hai, Xu Yaohui, Li Yang, et al. Experimental simulation of terrigenous organic matter sedimentary characteristics and its influencing factors in transitional facies delta depositional system[J]. *Acta Sedimentologica Sinica*, 2024, 42(1): 251-265.]
- [18] Ramaswamy V, Gaye B, Shirodkar P V, et al. Distribution and sources of organic carbon, nitrogen and their isotopic signatures in sediments from the Ayeyarwady (Irrawaddy) continental shelf, northern Andaman Sea[J]. *Marine Chemistry*, 2008, 111(3/4): 137-150.
- [19] 李中乔, 吴莹, 李珍, 等. 越南红河水下三角洲表层沉积物中有机物分布及来源分析[J]. *海洋与湖沼*, 2013, 44(3): 577-583. [Li Zhongqiao, Wu Ying, Li Zhen, et al. Distribution and source of organic matter in the surface sediments from the red river subaquatic delta, Vietnam[J]. *Oceanologia et Limnologia Sinica*, 2013, 44(3): 577-583.]
- [20] Burd A B, Frey S, Cabre A, et al. Terrestrial and marine perspectives on modeling organic matter degradation pathways[J]. *Global Change Biology*, 2016, 22(1): 121-136.
- [21] Li Z Q, Wu Y, Liu S M, et al. An 800-year record of terrestrial organic matter from the East China Sea shelf break: Links to climate change and human activity in the Changjiang Basin[J]. *Deep Sea Research Part II: Topical Studies in Oceanography*, 2016, 124: 64-73.
- [22] 蔡进功, 包于进, 杨守业, 等. 泥质沉积物和泥岩中有机质的赋存形式与富集机制[J]. *中国科学(D辑): 地球科学*, 2007, 37(2): 234-243. [Cai Jingong, Bao Yujin, Yang Shouye, et al. Research on preservation and enrichment mechanisms of organic matter in muddy sediment and mudstone[J]. *Science China (Seri. D): Earth Sciences*, 2007, 37(2): 234-243.]
- [23] Hu L M, Shi X F, Bai Y Z, et al. Recent organic carbon sequestration in the shelf sediments of the Bohai Sea and Yellow Sea, China[J]. *Journal of Marine Systems*, 2016, 155: 50-58.
- [24] 陈诚, 朱怡翔, 石军辉, 等. 断陷湖盆浅水三角洲的形成过程与发育模式: 以苏丹 Muglad 盆地 Fula 凹陷 Jake 地区 AG 组为例[J]. *石油学报*, 2016, 37(12): 1508-1517. [Chen Cheng, Zhu Yixiang, Shi Junhui, et al. The forming process and development pattern of shallow water delta in fault Depression lacustrine Basin: A case study of AG Formation in the Jake area in Fula Sag, Muglad Basin, Sudan[J]. *Acta Petrolei Sinica*, 2016, 37(12): 1508-1517.]

- [25] 魏康强, 张元福, 李媛, 等. 利用水槽实验对不同流速下三角洲发育区别的探究[J]. 复杂油气藏, 2017, 10(3): 6-11. [Wei Kangqiang, Zhang Yuanfu, Li Yuan, et al. Study on difference of delta development under different flow velocities by flume experiment[J]. Complex Hydrocarbon Reservoirs, 2017, 10(3): 6-11.]
- [26] 郭超, 何青. 黏性泥沙絮凝研究综述与展望[J]. 泥沙研究, 2021, 46(2): 66-73. [Guo Chao, He Qing. Review of the research on cohesive sediment flocculation[J]. Journal of Sediment Research, 2021, 46(2): 66-73.]
- [27] 李胜利, 李顺利, 付超. 长试管静置沉降实验结果对湖盆细粒沉积纹层成因的启示[J]. 古地理学报, 2022, 24(3): 405-414. [Li Shengli, Li Shunli, Fu Chao. Implication of the large-tube settling experiment results on genesis of fine-grained deposition lamination of lacustrine Basin[J]. Journal of Palaeogeography (Chinese Edition), 2022, 24(3): 405-414.]
- [28] 程乐峰, 余涛, 张孝顺, 等. 机器学习在能源与电力系统领域的应用和展望[J]. 电力系统自动化, 2019, 43(1): 15-31. [Cheng Lefeng, Yu Tao, Zhang Xiaoshun, et al. Machine learning for energy and electric power systems: State of the art and prospects[J]. Automation of Electric Power Systems, 2019, 43(1): 15-31.]
- [29] Niu X X, Suen C Y. A novel hybrid CNN-SVM classifier for recognizing handwritten digits[J]. Pattern Recognition, 2012, 45(4): 1318-1325.
- [30] 林景栋, 吴欣怡, 柴毅, 等. 卷积神经网络结构优化综述[J]. 自动化学报, 2020, 46(1): 24-37. (Lin Jingdong, Wu Xinyi, Chai Yi, et al. Structure optimization of convolutional neural networks: A survey[J]. Acta Automatica Sinica, 2020, 46(1): 24-37.)
- [31] 周永章, 王俊, 左仁广, 等. 地质领域机器学习、深度学习及实现语言[J]. 岩石学报, 2018, 34(11): 3173-3178. [Zhou Yongzhang, Wang Jun, Zuo Renguang, et al. Machine learning, deep learning and Python language in field of geology[J]. Acta Petrologica Sinica, 2018, 34(11): 3173-3178.]
- [32] 苏恺明, 徐耀辉, 徐旺林, 等. 鄂尔多斯盆地延长组多油源贡献比例与分布规律: 基于机器学习与可解释性研究[J/OL]. 地学前缘:1-17[2024-05-13]. [SU Kaiming, XU Yaohui, XU Wanglin, et al. Contribution ratio and distribution of numerous oil sources in the Yanchang Formation of the Ordos Basin: Based on machine learning and interpretability studies[J/OL]. earth science Frontiers: 1-17[2024-05-13].]
- [33] 匡立春, 刘合, 任义丽, 等. 人工智能在石油勘探开发领域的应用现状与发展趋势[J]. 石油勘探与开发, 2021, 48(1): 1-11. [Kuang Lichun, Liu He, Ren Yili, et al. Application and development trend of artificial intelligence in petroleum exploration and development[J]. Petroleum Exploration and Development, 2021, 48(1): 1-11.]
- [34] 李增学, 张功成, 王东东, 等. 确定控制陆源分散有机质分布三角洲体系外缘坡度的方法: 中国, 108629130B[P]. 2019-09-24. [Li Zengxue, Zhang Gongcheng, Wang Dongdong, et al. A method to control the slope of the outer edge of the terrigenous dispersed organic matter distribution delta system is determined: CN, 108629130B[P]. 2019-09-24.]
- [35] 李增学, 刘莹, 李晓静, 等. 琼东南盆地古近纪泥炭沼泽破坏与重建作用对煤型源岩物质形成的控制[J]. 石油与天然气地质, 2022, 43(6): 1309-1320. [Li Zengxue, Liu Ying, Li Xiaojing, et al. The control of Paleogene peat swamp destruction and reconstruction on the Formation of coal-type source material in the Qiongdongnan Basin[J]. Oil & Gas Geology, 2022, 43(6): 1309-1320.]

Machine Learning-Based Model for Predicting the Distribution of Terrestrial Organic Matter in a Marine-Terrestrial Transitional Environment: From sedimentary simulation experiments to geological applications

SHAO Junzhe¹, LI Yang², WANG Yong³, WU Nan¹, WEI Wei¹, LIU Hai²

1. School of Geosciences, Yangtze University, Wuhan 430100, China

2. Hubei Key Laboratory of Oil and Gas Geochemistry and Environment (School of Resources and Environment, Yangtze University), Wuhan 430100, China

3. Oil and Gas Exploration and Management Center of SINOPEC Shengli Oilfield Branch, Dongying Shandong 257099, China

Abstract: [Objective] Marine-terrestrial transitional source rocks are the main source rocks in several offshore basins in China, and their differential distribution characteristics restrict the prediction accuracy of source rocks and the effectiveness of oil and gas exploration. In addition, the transport and sedimentation process of terrestrial organic matter determines the quality and distribution of source rocks in a marine-terrestrial transitional environment. [Methods] Using a combination of flume sedimentary simulation and three-dimensional (3D) laser scanning technology, the dynamic recording and quantitative characterization of the transport process of terrestrial dispersed organic matter under different water salinity conditions were conducted using forward modeling. Machine learning algorithms were used to establish a total organic carbon (TOC) prediction model. [Results] The results show that terrestrial organic matter in the marine-terrestrial transitional environment is mainly enriched in the delta front and pro-delta. As the transportation distance increases, the abundance of terrestrial organic matter first increased and then decreased. Under the influence of salt flocculation, the transportation distance of terrestrial organic matter in the saltwater environment is closer to the source area, and the sediment thickness is larger. A TOC prediction model was established under experimental conditions based on three deep learning algorithms and, ultimately, the prediction model based on random forest algorithm with outlier removal and experience based sedimentary facies assignment as input features was selected as the optimal model. [Conclusion] The TOC prediction model under experimental conditions is combined with geological conditions to complete the TOC prediction of source rocks in the Yacheng Formation of the Yanan depression. The results show that the transportation distance of terrestrial organic matter in the Yanan Depression can reach 50 km, and the highest degree of organic matter enrichment occurs at a distance of approximately 31 km from the source area.

Key words: marine-terrestrial transitional environment; terrestrial organic matter; machine learning; sedimentary simulation; prediction model