

# 土壤有机碳空间预测与驱动解释研究进展

曾国安

长江大学资源与环境学院, 湖北 武汉

收稿日期: 2026年9月16日; 录用日期: 2026年10月17日; 发布日期: 2026年10月27日

## 摘要

土壤有机碳空间预测是土壤资源管理和碳储量评估的基础, 其应用价值取决于预测精度、区域适用性及解释依据。本文采用叙述性综述方式, 梳理数字土壤制图的模型方法、环境协变量与尺度效应, 重点讨论验证策略、过程融合和管理信息的利用。以改进RothC与随机森林融合的已发表研究为例, 分析模拟样本传递、尺度匹配和校准验证的技术环节; 结合分类、重复观测与层级数据, 比较管理信息的建模路径。在此基础上, 提出连接应用目标、数据条件、模型选择和验证方案的决策流程。现有证据尚不足以支持统一的算法优劣排序, 空间分块验证与基于概率抽样的地图精度评价也不能相互替代。对于碳储量变化和水土保持应用, 需要统一土层与质量核算口径, 检验预测区间和适用域, 并以重复观测或试验设计区分预测关联与管理效应。

## 关键词

土壤有机碳, 数字土壤制图, 空间验证, 预测不确定性, 模型适用域, 可解释机器学习

# Advances in Spatial Prediction and Driver Interpretation of Soil Organic Carbon

Guo'an Zeng

School of Resources and Environment, Yangtze University, Wuhan Hubei

Received: September 16, 2026; accepted: October 17, 2026; published: October 27, 2026

## Abstract

Spatial prediction of soil organic carbon (SOC) supports soil resource management and carbon stock assessment. Its value depends on predictive accuracy, regional applicability and the evidence underlying interpretation. This narrative review examines modelling approaches, environmental co-variables and scale effects, with emphasis on validation, process-data integration and management information. A published study combining a modified RothC model with random forest illustrates the

transfer of simulated training samples, scale matching, calibration and validation. Methods for categorical, repeated-measurement and hierarchical management data are compared. An application-oriented workflow then links study objectives and data availability to model selection and validation. Existing evidence does not support a universal ranking of algorithms. Spatially blocked validation and probability-based map accuracy assessment address different inferential targets and are not interchangeable. Applications to carbon stock change and soil and water conservation require consistent soil depths and mass accounting, evaluation of prediction intervals and areas of applicability, and repeated observations or experimental designs that distinguish predictive associations from management effects.

## Keywords

Soil Organic Carbon, Digital Soil Mapping, Spatial Validation, Prediction Uncertainty, Area of Applicability, Interpretable Machine Learning

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

土壤有机碳(soil organic carbon, SOC)参与土壤结构形成、养分供给和水分保持,也是陆地碳循环研究的重要对象。SOC 的积累与持久性取决于有机物输入、微生物转化及其所处的矿物和水热环境,不能仅由有机物本身的化学组成解释[1]。在土壤调查和农业管理中,了解 SOC 在何处富集、何处偏低,有助于识别土壤退化风险和安排监测样点;若进一步估算碳储量及其变化,还需将含量与容重、土层厚度和粗颗粒比例结合起来。

野外采样和实验室测定提供了 SOC 的直接证据,调查成本却限制了样点的空间覆盖。数字土壤制图(digital soil mapping, DSM)利用土壤属性与环境信息的统计联系,将离散观测扩展为空间连续的预测结果。McBratney 等提出的 SCORPAN 框架为协变量组织奠定了基础[2];国内相关综述进一步从环境变量、采样设计、推理方法和验证等方面梳理了 DSM 技术体系[3]。GlobalSoilMap 则将这一思路拓展至全球细分分辨率土壤属性网格,并强调不确定性的表达[4]。这些研究共同表明,算法只是制图过程中的一个环节,数据来源与推断方式同样影响地图质量。

机器学习增强了 SOC 与环境因素之间非线性关系的表达能力,也使数据质量和模型解释问题更加突出。已有综述分别从 SOC 制图算法与协变量[5]、机器学习在 DSM 中的应用条件[6]以及大尺度 GlobalSoilMap 属性制图[7]展开评述,为方法选择提供了基础。在此基础上,仍有必要将预测精度、跨区域适用性和解释可靠性放在同一研究任务中考察,厘清不同评价结果各自能够支持的结论。

本文围绕 SOC 空间预测与驱动解释,比较不同模型处理空间结构的方式,分析协变量选择和尺度变化对结果的影响,并讨论验证设计、预测区间与模型解释的联系。对于侵蚀和土地管理应用,重点辨析空间关联与过程证据的差别,以明确 SOC 地图在措施评价中的用途和局限。

## 2. 评述范围与比较思路

本文采用叙述性综述方式,围绕 SOC 及相关土壤属性的数字制图研究,组织具有代表性的理论、方法和应用文献。文献线索由主题关键词、既有综述的前后向引文及期刊官网检索获得,检索词围绕土壤有机碳、数字土壤制图、空间验证、过程-机器学习融合和管理信息建模展开;本次修订对代表性研究

的题名、作者、年份、卷页及在线状态核对至 2026 年 9 月 22 日。所引文献发表于 2003~2025 年，评述对象包括 SOC 含量、单位面积碳储量及剖面预测。实验室光谱研究用于说明土壤信息获取方式，不与区域制图精度直接比较。本文不作穷尽性文献计量或效应量合并。

文献比较主要考察研究范围、样点布局、土层深度、协变量时空口径及验证方式。模型原理以方法论文为依据，SOC 制图效果对应具体应用研究，侵蚀与碳稳定性分析则结合过程研究。由于响应变量、取值范围和验证设计存在差异，不同论文报告的决定系数或误差值仅在可比条件下用于评价算法。方法论文中的模拟结果及其他土壤属性案例，主要用于说明方法特征，不作为 SOC 制图效果的直接证据。为使比较落到证据条件上，本文另将代表性应用的数据支持、验证方式、主要结果和适用边界并列整理(表 2)，并据此提出面向应用目标的模型选择流程。

### 3. SOC 数字土壤制图的理论基础

#### 3.1. 数字土壤制图的基本框架

DSM 通常包括样点整理、协变量准备、模型拟合、空间预测和结果验证[2] [3]。样点代表性决定模型可以学习的环境组合，协变量反映可纳入预测的环境差异，验证设计则界定误差估计所对应的应用情景。这些环节相互制约：计算效率提高有助于重复制图，观测不足却仍会限制推断范围；输出网格变细，也未必带来同等尺度上的精度提升。

#### 3.2. SCORPAN 成土因素框架

SCORPAN 将已有土壤信息、气候、生物、地形、母质、年龄或时间以及空间位置纳入土壤预测框架[2]，可概括为：

$$S = f(s, c, o, r, p, a, n) + \varepsilon$$

式中， $S$  为待预测土壤属性； $s$  为其他土壤属性或已有土壤图； $c$  为气候； $o$  为生物及相关土地利用信息； $r$  为地形； $p$  为母质； $a$  为年龄或时间； $n$  为空间位置； $\varepsilon$  为模型未解释部分。是否对  $\varepsilon$  建立空间相关结构，取决于具体模型及残差诊断。土壤质地属于可观测土壤属性，也能间接反映母质影响，但两者不能视为同一概念。

SCORPAN 为协变量选择提供了分类框架，具体变量仍需依据数据条件和预测目标确定。多年气候平均值适合描述背景环境，采样时间则用于识别观测所对应的时期，两者承担不同作用。坐标可吸收部分空间变化，也可能增强模型对训练区位置关系的依赖，其取舍因而需要结合跨区验证和解释目的[8]。

#### 3.3. SOC 含量与碳储量的口径

SOC 含量通常表示单位质量细土中的有机碳质量，常用  $\text{g}\cdot\text{kg}^{-1}$ 。体积碳密度表示单位体积土壤中的碳质量，常用  $\text{kg}\cdot\text{m}^{-3}$ ；给定土层的单位面积碳储量常用  $\text{kg}\cdot\text{m}^{-2}$  或  $\text{Mg}\cdot\text{ha}^{-1}$ ，中文文献也常称其为“有机碳密度”。区域碳总量则需进一步对面积积分，常用  $\text{Mg}$  或  $\text{Pg}$ 。使用“密度”一词时，应同时说明体积或面积口径。

若 SOC 含量  $C$  以  $\text{g}\cdot\text{kg}^{-1}$  表示，细土基质容重  $\rho$  以  $\text{Mg}\cdot\text{m}^{-3}$  表示，土层厚度  $d$  以  $m$  表示，粗颗粒体积分数为  $v$ ，则该土层单位面积 SOC 储量可写为：

$$M = 10C\rho d(1-v)$$

式中， $M$  的单位为  $\text{Mg}\cdot\text{ha}^{-1}$ 。该式要求  $\rho$  按扣除粗颗粒后的细土基质体积定义， $v$  为 0 至 1 之间的体积分数；若粗颗粒以体积百分数给出，需先除以 100。若容重已表示单位整体土体体积内的细土质量，则不应

再次乘以 $(1-v)$ 。多土层储量应逐层计算后相加,区域总量则由各空间单元的单位面积储量乘以相应面积后汇总。 $1\text{ kg}\cdot\text{m}^{-2}$ 等于  $10\text{ Mg}\cdot\text{ha}^{-1}$ ,  $1\text{ Mg}$  等于  $1\text{ t}$ 。SoilGrids 等产品分别提供 SOC、容重和粗颗粒信息,换算前需核对单位、缩放系数和土层定义[9] [10]。

固定深度碳储量用于描述给定土层的碳量,但容重不同的土体在相同深度内并不包含相同土壤质量。Wendt 和 Hauser 给出了适用于多土层的等质量土壤法(equivalent soil mass, ESM)计算方案,以共同的土壤质量基准比较不同处理或时期的碳储量[11]。用于变化监测时,应说明参考质量、分层质量的累计方式,以及底层截取或补足的处理规则。侵蚀与沉积还会改变土体来源和碳的侧向通量,质量基准的一致并不能代替对这些过程的核算。

## 4. SOC 空间预测方法的演进

地统计、机器学习、空间混合模型和过程融合方法在 SOC 制图中长期并存,各自处理的数据结构和预测问题有所侧重[12]。方法选择取决于目标变量、观测密度与协变量质量,也取决于预测区和训练区的环境差异。

### 4.1. 地统计与回归克里金

普通克里金通过变异函数或协方差模型描述空间相关性,适用于能够从样点中较稳定地估计空间结构的情形。相关假设针对具体随机场及其均值、协方差结构,各向同性是可以采用的一种简化,并非所有克里金方法的必需条件。样点稀疏、空间趋势明显或区域非平稳性较强时,变异函数估计和局部预测都可能不稳定[2]。

回归克里金先利用协变量拟合趋势,再对残差作空间预测。趋势部分可以采用线性模型,也可由机器学习拟合,因此其非线性表达能力取决于所用趋势模型。克里金方差是在给定模型及其参数条件下计算的预测不确定性,不能自动涵盖测量误差、参数估计误差和全部协变量误差。在模型比较中,回归克里金仍可作为检验空间残差信息是否有用的参照[5] [6]。

### 4.2. 机器学习模型

随机森林(random forest, RF)、梯度提升树、支持向量回归和神经网络可拟合 SOC 与协变量之间的非线性关系,其中 RF 在 SOC 制图中应用较广[5]。这类方法减少了对响应分布和函数形式的预设,预测能力仍依赖训练样本与应用场景的可比性。模型复杂度、缺失值处理方式和异常值敏感性各不相同,因而需要在具体数据条件下比较。

普通表格型机器学习通常不显式建立残差空间协方差,但气候、地形和植被变量中的空间格局已能为预测提供信息。坐标又补充了位置关系,其带来的精度变化与环境机制的解释是两个不同问题。Meyer 等指出,空间预测变量的选择会影响模型从再现已有数据到预测新位置的能力[8]。比较算法时,统一样本、输入信息和验证划分,并在训练数据内完成变量筛选与调参,有助于区分算法差异和评价程序的影响[13]。

### 4.3. 空间机器学习与混合模型

空间混合模型通过残差校正、局部加权或邻域特征利用位置关系。RF 与残差克里金结合,可进一步利用机器学习拟合后保留的空间结构。周琪清等在江西上高县耕地研究中,将物候、极端气候信息与 RF 残差校正结合,展示了协变量扩展与混合建模的应用[14]。这一类研究同时改变了输入信息和模型结构,其增益来源需要在一致的验证条件下分别比较。

地理加权机器学习通过邻域内拟合表达局部关系,性能受空间非平稳性、带宽及样点布局影响。Lee

等在田块与流域尺度比较了地理加权模型和常规模型，并在田块尺度设置均匀与分层随机样点布局。地理加权模型在田块尺度总体表现较好，在流域尺度则与常规模型接近；样点密度降低和分布不规则限制了局地异质性的识别[15]。这一差异表明，局部建模的收益具有尺度依赖性，也需要足够的邻域样本支持。

卷积神经网络可利用栅格邻域信息，其表现同时受训练样本数量和区域代表性制约。与表格型模型、空间基线模型比较时，训练样点及影像邻域的重叠情况尤其值得关注，因为共享邻域信息可能影响验证结果[6]。网络结构的选择需要与数据的空间组织方式相适应。

#### 4.4. 过程与数据融合模型

过程模型描述 SOC 输入、转化和损失，统计模型从观测中学习预测关系，两者可以在参数估计、数据同化、模拟样本扩充或残差订正等环节结合。Tao 等综合微生物过程模型、数据同化与深度学习分析全球 SOC 储存，展示了多类信息联合利用的可能性[16]。判断融合是否有效，首先应说明两个模型之间传递的对象，以及这些信息是否提供了可检验的约束。

Zhang 等提出的过程-机器学习融合框架提供了较清晰的实例[17]。该研究使用中国东部耕地 428 个地点在 1980 年和 2000 年的 856 份表层样本，以改进 RothC 模拟 1981~1999 年的碳储量，将模拟值作为附加响应样本，与实测样本共同训练随机森林。实测误差与模拟误差分别归一化并赋权，以免数量较多的模拟样本支配拟合；同一地点的记录在交叉验证中归入同一折。该研究说明的是特定资料条件下的融合可行性，并未通过逐年实测验证全部中间年份的变化轨迹。

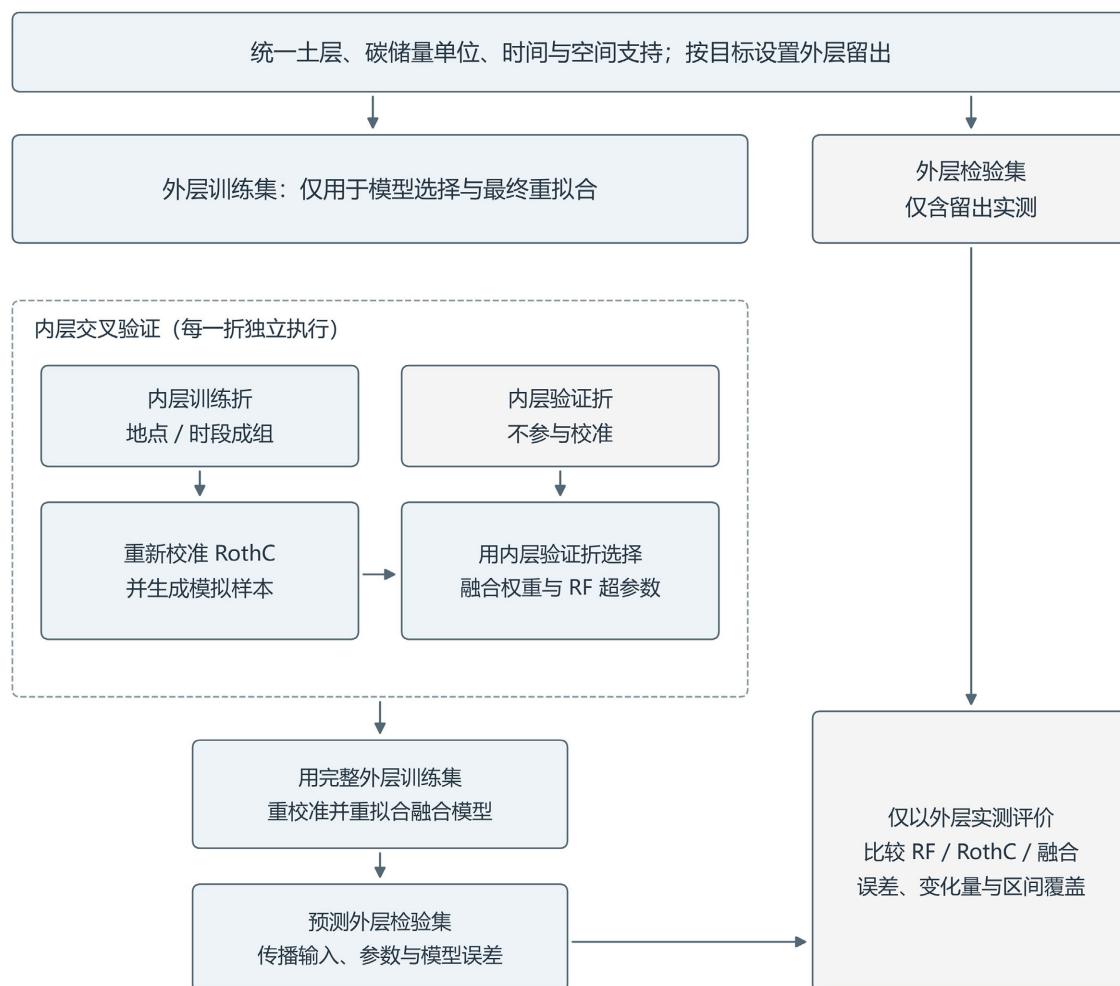
参数与状态的传递需要先于模型训练确定。标准 RothC 以月为步长，使用气候、黏粒、植被覆盖和有机碳输入等信息描述不同碳库的周转；初始碳库与碳输入并不是可由 SOC 总量唯一确定的量[18]。上述案例在 RothC 中加入微生物调节项，并在训练折内校准相关参数[17]。在类似应用中，宜列明固定参数、待估参数、初始状态和输入数据，先做敏感性分析，再用重复 SOC 观测及可获得的碳组分信息约束少量关键参数。若由机器学习估算残体碳输入，传递给 RothC 的应是统一单位的月碳输入及其误差，而不能把施肥量、NDVI 或模型重要性直接视为碳通量。资料不足时，多组参数得到相近拟合并不意味着各参数都已被可靠识别。

尺度匹配至少包括土层、时间和空间支持三个方面。观测 SOC 含量应先按一致土层、容重及粗砾口径换算为碳储量，再与过程模型状态比较；管理导致容重变化时，还应检查等质量土壤比较的必要性[11]。月尺度气候和碳输入用于过程模拟，输出则应对齐实测月份或预先定义的年度时点。粗分辨率气候、县域管理统计与点位土壤数据不能因为重采样到同一网格，就被视为具有相同的信息精度。本文建议保留各数据的原始支持范围，以区域聚合、分层建模或分辨率敏感性试验评估尺度转换的影响，而不是用更细的输出像元替代这一检验。

校准与结果检验应在流程上隔离。Zhang 等采用地点成组的五折交叉验证，并只用留出折中的实测值评价过程模型、RF 与融合模型[17]。若融合权重、RF 超参数或过程参数还要通过数据选择，宜进一步采用嵌套验证(图 1)：先按地点或时段划分外层训练集与外层检验集；在每个内层训练折中重新校准过程参数并生成模拟样本，仅用相应的内层验证折选择融合权重和 RF 超参数；选择完成后，再以完整外层训练集重校准、重生成并重拟合整条流程，最后只在外层留出的实测值上报告性能[13][19]。这样可以避免内层验证信息经过程序校准或模拟样本回流到训练环节。普通 RF、单独 RothC 与融合模型应共用外层检验数据，分别比较碳储量水平、可观测变化量和区间覆盖。模拟样本只增加了来自过程假设的信息，不能计入独立实测样本量；训练误差降低也不能替代外部验证。

过程约束还须服从模型边界。标准 RothC 并非为永久淹水土壤设计，扩展到水稻土时应说明所采用的改进和验证依据[18]；常规土壤周转模拟也不自动包含坡面侵蚀、搬运与沉积。若研究水土保持碳效应，

应另行描述侧向碳通量及其边界,再讨论区域净变化,不能把局部碳储量增加直接解释为大气净碳汇[20][21]。碳量非负和质量收支可以作为一致性检查,而“SOC 必定随深度递减”等经验关系不宜作为普遍硬约束。



注: 据 Zhang 等的融合思路绘制[17]。原研究采用地点成组的五折交叉验证,并在留出折实测值上评价;图中“内层各折重新校准-重新生成模拟样本-选择权重与超参数-外层检验”的完整嵌套流程,是本文为避免选择偏倚提出的实施建议,并非原研究全部操作的逐项复现。模拟样本不是新增实测,数据划分方式须与空间预测或时间外推任务相对应。

**Figure 1.** Conceptual workflow for RothC-machine learning integration and independent validation

**图 1.** RothC-机器学习融合及独立验证的概念流程

#### 4.5. 方法选择的比较依据

表 1 概括了各类方法的用途与评价重点。样点布局、环境梯度和预测任务会改变模型表现,因此表中侧重适用条件,而非给出统一排序。对于不确定性,还需区分重复拟合结果的离散程度与未知观测值的预测区间。

方法的适用性还取决于研究证据如何产生。表 2 选取全球产品、区域制图、尺度比较、过程融合和管理动态建模五类研究,列出其数据支持、验证方式与可解释边界。表中数值用于说明证据条件,不用于跨研究排列算法优劣。

**Table 1.** Approaches to SOC spatial prediction and priorities for evaluation**表 1.** SOC 空间预测方法及其评价重点

| 方法      | 空间信息的表达         | 不确定性表达         | 适用条件与评价重点           |
|---------|-----------------|----------------|---------------------|
| 普通克里金   | 变异函数或协方差结构      | 条件于所用模型的克里金方差  | 检查趋势、空间结构可估计性及采样密度  |
| 回归克里金   | 协变量趋势与残差空间结构    | 需同时考虑趋势和残差部分   | 检验残差空间校正是否改善独立预测    |
| RF 及提升树 | 通过协变量表达；可加入空间特征 | 可配分位数模型、校准或重采样 | 统一输入与调参预算；评估相关变量和外推 |
| 地理加权模型  | 邻域权重与局部拟合       | 需专门估计并验证区间     | 检查带宽、邻域样本量与地理迁移     |
| 卷积神经网络  | 提取影像或地形邻域特征     | 需额外量化与校准       | 检查邻域重叠及额外数据带来的收益    |
| 过程融合模型  | 依赖过程模型与空间化方案    | 考虑数据、过程参数及结构误差 | 检验过程约束适用性和质量平衡      |

注：据相关方法与综述文献归纳[2] [5] [6] [22]。本表比较方法特点，不代表在统一数据集上的性能排序。

**Table 2.** Evidence conditions and applicability limits of representative SOC studies**表 2.** 代表性 SOC 研究的证据条件与适用边界

| 研究任务与数据                                            | 模型与输入                           | 验证或比较                                 | 主要证据与边界                                                                                                                       |
|----------------------------------------------------|---------------------------------|---------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|
| 全球土壤属性制图；约 24 万剖面、92 万余土层，6 个标准深度，250 m [10]       | 分位数回归森林；400 余个全球环境协变量           | 空间平衡的 10 折交叉验证；同一剖面入同一折；检验 90% 预测区间覆盖 | 提供全球格局与空间不确定性。遗留样点分布不均，交叉验证指标不能替代概率抽样意义上的目标区域地图精度。                                                                            |
| 江西上高县耕地 SOC 区域制图[14]                               | 物候、极端气候等协变量；RF 及残差空间校正          | 比较协变量组合与模型结果                          | 说明物候和极端气候信息可补充常规协变量。输入信息与模型结构同时改变时，精度增益不能全部归因于单一变量。                                                                           |
| 田块与流域尺度 SOC、137Cs 及土壤再分配率预测[15]                    | 地理加权机器学习与常规模型；田块尺度比较均匀与分层随机样点布局 | 跨尺度模型比较及残差空间分析                        | 地理加权模型在田块尺度总体较优，在流域尺度与常规模型接近；低密度、不规则样点限制局地异质性识别。结论依赖所研究尺度与样点布局。                                                               |
| 中国东部耕地 428 个地点、1980 和 2000 年共 856 份 0~20 cm 样本[17] | 改进 RothC、RF 及模拟样本加权融合 (POML)    | 地点成组的五折交叉验证；留出折仅用实测值评价                | 融合模型 RMSE 为 0.29 kg·m <sup>-2</sup> ，低于 RF 的 0.36 和 RothC 的 0.53 kg·m <sup>-2</sup> ；相对 RF 改善 19%。中间年份轨迹主要受过程模拟约束，未由逐年实测独立验证。 |
| 新西兰长期耕作试验；42 个地块、2000~2010 年、多类型观测[26]             | CQUESST 贝叶斯层级动态模型；6 个碳库、月时间步长   | 在试验设计下比较作物、耕作和季节处理，并表达参数、过程和观测不确定性    | 支持试验地块的碳周转与处理比较；尚不能据此证明模型已具备区域时空外推能力。                                                                                         |

注：样点规模、土层、验证方式和响应变量均会影响误差数值。表中研究并非同一数据集上的统一基准测试；其中 Zhang 等的性能数值只适用于其研究区与交叉验证设计[17]。

## 4.6. 面向管理信息的建模方法

管理信息进入模型前, 应先明确其数据结构。耕作方式、轮作制度属于分类变量, 施肥和还田量属于强度变量, 实施年份、持续年限与历史投入则构成时序信息; 同一农场或地块的重复采样还带来组内相关。本文建议以“地块-观测时间”为基本记录单元, 保留农场或试验点标识、措施起止时间、剂量及单位, 并区分“未实施”“记录缺失”和“无法核实”。土地利用类型可作为背景协变量, 但不能替代管理历史。

对于以分类变量和连续环境变量为主的数据, CatBoost 可作为提升树候选。其有序目标统计与有序提升用于缓解类别编码及训练过程中的目标泄漏[23]。例如, 可将耕作方式、作物类型与黏粒、气候及措施持续年限共同用于 SOC 预测, 但不能把无序的耕作类别编码为具有大小关系的整数。该算法并不自动处理同一地块重复观测的依赖, 也不意味着时间预测中可以使用未来记录; 类别处理及调参仍须限制在训练折内, 并与空间或时间留出设计配合。

对于地块或农场内有重复观测的数据, 混合效应随机森林(mixed-effects random forest, MERF)将 RF 估计的总体非线性关系与组层随机效应结合[24]。在一个假设的多农场 SOC 监测设计中, 可由 RF 拟合气候、土壤和管理年限的关系, 再用农场随机截距描述未观测的组间差异; 随机斜率则需要足够的组内观测支持。对已观测农场可利用其估计的随机效应, 对全新农场则主要依赖总体部分, 二者不能混为一种预测任务。原始 MERF 并不自动刻画任意的时间自相关; 农场数量很少或组内观测极稀疏时, 宜先与较简单的混合效应模型比较。

若非线性响应之外还存在组间差异或空间相关, GPBoost 提供了另一种组合方式: 用提升树学习均值结构, 用分组随机效应或高斯过程描述相关结构, 并联合估计协方差参数[25]。例如, 在有坐标与农场标识的监测资料中, 可以检验加入农场效应或空间残差后, 跨农场预测及区间表现是否改善。它与 CatBoost 的主要区别不是类别数量, 而是是否显式建模观测依赖; 它也不能直接等同于对所有未知量进行完整后验推断的贝叶斯模型。

对于管理措施随时间变化、且需要刻画潜在碳库轨迹的研究, 贝叶斯层级状态空间模型可分别描述真实过程、测量误差和参数先验。Pagendam 等的 CQUESST 将随机碳周转模型用于新西兰长期耕作试验, 利用层级结构评估处理差异并表达碳动态的不确定性, 为管理信息进入过程参数提供了实例[26]。该实例属于试验地块的动态建模, 而非已经完成区域时空外推。若用于区域监测, 还需根据采样结构增加空间相关项、时间演化项或地区层级, 并检查先验敏感性、参数可辨识性、采样收敛及留出预测; 计算开销和数据需求通常也高于普通树模型。

因此, 分类信息处理可从 CatBoost 与常规树模型比较起步, 重复观测资料可考虑 MERF, 兼具非线性和相关结构时可比较 GPBoost, 过程轨迹与不确定性分解则更适合层级动态模型。上述农场例子是应用设计示例, 不是这些方法已在同一 SOC 数据集上完成的性能比较。对于秸秆还田效果, 可先以农场留出检验迁移能力, 再以时间前推检验后期预测; 若要估计措施效应, 还必须处理基线差异、措施选择和同期气候变化等混杂。加入管理变量或得到较高的 SHAP 贡献, 均不能替代对照试验或相应的因果识别设计[19][27]。

## 5. 环境协变量体系及其尺度效应

协变量既是预测信息的来源, 也是解释模型结果的依据。变量是否可获得、能否对应采样时间与土层, 以及在目标区域是否有完整覆盖, 通常比简单增加变量数量更重要[5][7]。

### 5.1. 气候因素

温度和降水通过植物输入、微生物活动及土壤水分影响 SOC, 对碳输入和分解的作用可能方向相反。

年均温、年降水量描述背景梯度,季节性指标和极端事件变量补充年均值未能反映的变化[1] [14]。气候梯度在大区域研究中通常更易识别,而田块或小流域内气候变幅有限,模型中的气候变量重要性也会随之受到样本范围的限制。

不同气候变量常彼此相关,也可能与海拔、植被类型共同变化。因此,模型中的温度或降水响应应结合其他变量的联合分布解释。若样点跨越多年,而协变量仅取某一年,所得关系还可能混合空间差异与时间差异。

## 5.2. 地形与侵蚀沉积过程

高程、坡度、坡向、曲率、地形湿度指数及汇水面积可用于表征水热条件和物质迁移环境。坡面侵蚀改变表土厚度并搬运富含有机碳的细颗粒,沉积有可能使碳埋藏或改变其分解环境。Kirkels 等比较了侵蚀、运输和沉积过程中 SOC 去向的不同认识[20],Doetterl 等则分析了侵蚀景观中的碳组分与稳定性[21]。这些研究说明,坡位间 SOC 差异需要同时考虑搬运、输入和转化过程。

因此,坡脚 SOC 较高可与沉积富集相一致,但不能单凭这一空间格局认定景观形成净碳汇。地形指标也不是侵蚀通量的直接测量。将侵蚀产品引入 SOC 模型时,应核查其计算是否已经使用 SOC 或有机质含量,避免再以该变量的重要性作为侵蚀独立作用的证据。

## 5.3. 植被与遥感信息

归一化植被指数(NDVI)、增强型植被指数和物候参数可间接表征植被状态及有机物输入。时序遥感能够补充单期影像难以辨认的生长信息,采用何种时间窗口则取决于采样时期与管理历史。周琪清等将物候和极端气候变量用于耕地 SOC 制图[14],说明遥感信息的价值不仅在于影像数量,还在于能否描述与土壤观测相关的生长过程。

遥感与 SOC 的联系还取决于地表覆盖状态。植被覆盖时,传感器主要观测冠层,模型利用的是植被与土壤的间接联系;裸土条件下,光谱对土壤表层性质更敏感,但水分、粗糙度和残茬也会影响信号。Viscarra Rossel 等建立的全球土壤光谱库说明了光谱信息用于土壤表征的潜力[28],其测量条件不能直接等同于卫星混合像元。

## 5.4. 土壤与母质信息

质地、矿物组成、pH 和水分状况影响 SOC 转化及有机质与矿物的相互作用。黏粒含量可反映部分稳定化条件,但不同矿物表面性质差异较大,不能将黏粒比例视为矿物保护作用的完整度量[1]。地质图与土壤图能够补充母质和土类背景,使用时需注意其制图比例尺、边界精度和图斑概化。

栅格土壤产品本身也是模型预测结果,其误差和数据来源会传递到后续分析。如果产品训练数据包含待验证样点,或某项属性与目标 SOC 存在直接计算关系,将其作为协变量就可能造成信息泄漏[6] [10]。因此,使用已有土壤图层时,对数据来源的追溯与精度信息的说明同样重要。

## 5.5. 人类活动与管理因素

耕作、施肥、秸秆还田、灌溉和土地利用历史会改变 SOC 输入与损失条件,管理信息却往往难以按地块和年份获得。土地利用分类可以区分部分管理背景,但同为耕地的地块,其种植制度和投入水平仍可能差异很大[3] [5]。管理信息缺失使模型更多依赖地形或植被代理,也限制了对管理措施的解释。

管理数据的补充需要兼顾调查口径和空间匹配。县域平均施肥量难以反映地块差异,直接赋值还会引入行政边界;遥感识别的耕作或作物制度则带有分类误差。农田与自然植被样本之间的预测精度差异可能受多种因素影响,管理数据缺失只是其中一种待检验的解释。

5.6. 空间尺度与土层深度

多尺度分析涉及研究范围、栅格粒度、邻域窗口和观测空间支持等不同要素。扩大范围通常增加环境梯度，增大像元则平滑局地差异，两者改变变量重要性的原因并不相同[7] [15]。区分这些尺度要素，有助于判断精度变化来自数据聚合、样本组成还是潜在的过程差异。

土层深度也会改变协变量与 SOC 的联系。表层受新鲜有机物输入和地表管理的影响较直接，深层则涉及根系输入、溶解性有机物迁移、矿物环境和埋藏历史[1]。SoilGrids 2.0 已提供至 200 cm 的标准深度层信息[10]，深层观测的数量与代表性仍是制图质量的制约。分层建模和以深度为协变量的联合建模都可用于三维表达，两者的差别主要在于深度信息的组织方式及剖面独立验证下的表现。

6. 模型验证不确定性与空间迁移

6.1. 验证目标与样本划分

随机交叉验证、空间分块交叉验证与独立样点验证所回答的问题并不相同。若预期应用位置与训练样点在地理和环境空间中接近，随机划分可近似这一预测情景；若目标是应用于远离训练区的区域，则空间隔离更能体现地理迁移任务。样点存在聚集或同一剖面重复观测时，应避免训练与验证数据分享近乎相同的信息[19] [29]。

Roberts 等主张依据预期预测任务，处理数据的空间、时间和层级结构并划分验证集[19]。Wadoux 等则从抽样推断出发，强调采用适当的概率抽样和设计型估计评价目标区域的地图总体精度[30]。前者关注特定迁移条件下的预测误差，后者关注目标总体上的地图误差，区别在于推断对象。跨区域评价中的分块大小、缓冲距离、各折样本量和环境覆盖，决定了验证所模拟的迁移条件；地图总体精度评价中的抽样框、入样概率及权重，则关系到估计的区域代表性。

数据来源独立与样本具有代表性是两个条件。外部样点若集中于易到达的区域，仍可能偏离目标区域的环境分布。为保持模型比较的一致性，可固定外层验证样点，在外层训练数据中进行变量筛选、缺失值处理和参数调优，调参时再设置内层划分。同一剖面的不同土层或同一位置的重复记录按预测任务成组处理，可减少共享信息对误差估计的干扰。最终测试数据独立于上述选择过程[13][19]，常见任务及其评价重点见表 3。

Table 3. Prediction tasks and corresponding validation designs  
表 3. 预测任务与验证设计的对应关系

| 应用任务       | 验证设计                      | 主要解释范围            |
|------------|---------------------------|-------------------|
| 已采样区域内的预测  | 按应用时样本与预测位置关系设置留出；处理重复与聚集 | 误差对应所模拟的样本和位置分布   |
| 向未采样区域迁移   | 空间分块、距离隔离或留一区域验证          | 反映特定隔离距离和环境差异下的性能 |
| 目标区域地图精度   | 独立概率抽样并采用相应设计型估计          | 支持目标区域的面积或总体精度推断  |
| 未来时间或新剖面预测 | 按时间或整剖面留出                 | 检验时间迁移或剖面外推广能力    |

注：综合相关空间验证研究[13] [19] [30]。验证名称本身不保证独立性，需同时报告抽样和划分细节。

6.2. 评价指标及其可比性

决定系数  $R^2$ 、均方根误差(RMSE)和平均绝对误差(MAE)分别描述相对于基准的预测表现和不同形式的误差大小。评价时应说明  $R^2$  的计算定义，不能把实测值与预测值的相关系数平方直接当作预测  $R^2$ ；后

者在预测较差时可以为负。MAE 相较 RMSE 对大误差的惩罚较弱,但并非不受异常值影响。平均误差可揭示系统高估或低估,必要时再辅以一致性指标[6] [22]。

评价指标的含义与目标变量的单位、变异范围及抽样设计有关。不同土层或地区的 SOC 变异程度不同, $R^2$  较高并不必然意味着更能满足管理需求;相近的总体误差也可能掩盖不同的局地偏差。结合主要土类、土地利用和稀疏采样区检查误差分布,可以补充总体指标未能反映的信息。

### 6.3. 预测区间与不确定性表达

SOC 预测的不确定性可来自测量和定位误差、历史数据口径差异、协变量误差,以及模型结构和参数估计。重复拟合所得地图的标准差主要反映所采用重采样与建模程序下的变动,不一定包含未知观测值的全部变异。分位数回归森林估计条件分位数,为构造预测区间提供了途径[31]; SoilGrids 2.0 则展示了全球土壤信息与空间不确定性一并发布的实践[10]。

预测区间的评价需要同时考察名义覆盖水平、实际覆盖率和区间宽度,并关注它们随区域、土类及 SOC 水平的变化。覆盖率与宽度之间存在权衡:宽区间可能包含更多观测值,却降低了决策中的分辨能力。保形预测提供了分布无关的区间构建思路,其常规有限样本覆盖保证仍依赖可交换性等条件。空间依赖和分布迁移可能破坏这些条件,边际覆盖也不意味着各地点具有相同的条件覆盖[32]。

### 6.4. 空间迁移与模型适用域

跨区域应用的难点之一是协变量联合分布发生变化。即使单个变量均处于训练范围内,新的变量组合也可能缺乏样本支持。Meyer 和 Pebesma 将环境差异与交叉验证设计相联系,提出识别模型适用域的方法[33]。适用域图标示了模型误差估计较有依据的区域,表达的是样本支持程度,而非每个像元的精度保证。

迁移学习、目标区域补采样和模型更新可用于缓解分布差异,效果仍取决于目标区域的独立检验。补充当地样点时,环境覆盖与评价数据的独立性需要同时考虑。对外推区而言,连续的图面或看似合理的空间格局只能提供有限的判断依据,样本支持和外推范围因而也是制图结果的重要组成部分[8] [33]。

## 7. 驱动因素识别与可解释机器学习

### 7.1. 变量重要性的含义

变量重要性回答的是模型依赖哪些输入信息。RF 回归模型中的不纯度重要性通常由节点分裂带来的平方误差或方差减少量累计得到,不能直接称为分类模型中的“基尼重要性”。置换重要性考察打乱某一变量后预测表现的变化,结果取决于评价数据、损失函数及变量之间的相关性[6]。

高度相关的变量可能分摊重要性,置换也可能产生原数据中少见的变量组合。因此,重要性排名反映的是特定模型与数据条件下的信息使用情况,难以直接对应过程的强弱。对地形、气候和植被等相关变量组,分组分析及跨折稳定性比较有助于辨别排名对变量表达方式的依赖。

### 7.2. 响应曲线与空间异质性

偏依赖图和个体条件期望曲线描述模型在改变部分输入时的预测响应。SHAP (SHapley additive explanations)将单次预测相对于基准值的差异分配给各特征,可用于分析局部贡献及其全局分布[34]。解释时需交代模型输出尺度、背景数据和所用解释方法。例如,若模型预测的是对数变换后的 SOC,SHAP 值也对应这一输出尺度,不能直接标注为原始 SOC 含量单位;其正负表示相对于基准预测的贡献方向,不等同于增加该环境因素后的因果效应。

响应曲线拐点的解释取决于样本密度和拟合稳定性。结合横轴上的样本分布,比较不同训练折或空间留出方案下的曲线,可以辨别转折是否仅出现在稀疏样本区,或随数据划分而移动。相关变量较多时,背景样本和变量组合也会影响拐点位置。稳定的转折可作为后续检验的候选区间,其生态含义仍需独立样点、过程观测或处理对照支持。跨地区、土层和模型的比较,则需要可比的输出尺度与背景条件。

多尺度地理加权回归(MGWR)允许不同关系采用不同带宽,用于描述空间非平稳性[35];地理探测器可检验空间分层异质性及其与解释变量分层的一致性[36] [37]。局地回归系数和 SHAP 贡献可在各自定义下表达正负方向,地理探测器  $q$  值衡量分层对空间分异的解释程度,本身不表示作用方向。三类结果不能相互替代;即使识别出相近的重要因素,也不能据此视为因果证据。

### 7.3. 从模型解释到过程判断

SOC 驱动解释通常始于识别模型依赖的信息,再分析预测响应的方向、形态及空间差异,最后与土壤过程观测相对照。重要性和响应分析提供的是模型层面的证据;从这些模式进一步解释机制,关键在于能否获得独立证据支持[6] [34]。

例如,NDVI 与 SOC 的正向关联可能反映植被输入增加,也可能与共同气候背景、管理投入或土壤肥力对植被的反向影响有关。重复观测、处理与对照设计,以及明确因果假设下的混杂控制,有助于区分这些解释。Tao 等结合过程模型和多类资料分析微生物碳利用效率[16],展示了过程研究对证据整合的要求,也说明单纯的模型贡献分解难以承担同样的解释任务。

水力侵蚀中的碳效应涉及坡面搬运、沉积后的埋藏与分解,以及植被补充的新碳输入。SOC 沿坡位的差异可用于提出过程假设,景观净碳收支则需要进一步核算相关通量[20] [21]。在一致的时空口径下,泥沙输出量与泥沙有机碳含量的乘积估计的是相应边界上的颗粒有机碳输出,并未涵盖所有溶解态碳通量,也不等于进入大气的  $\text{CO}_2$  量。空间调查能够识别与侵蚀沉积相一致的格局,碳的后续去向仍需过程资料补充。

## 8. 现存问题与研究方向

### 8.1. 样点代表性与数据口径

历史样点增加并不一定扩大环境覆盖。偏向道路、农业区或特定土类的采样,可能在部分环境中形成重复信息,而在其他环境中留下空缺。黄思华等梳理了面向 DSM 的采样设计方法[38]。这些方法的选择与采样用途有关:建模样点侧重覆盖环境组合,验证样点则需满足目标区域精度推断的要求。

相较继续增加同类样点,补齐采样日期、深度、测定方法和位置精度,往往更有助于判断多源数据能否合并。基于不确定性补采样时,还应检查高不确定性来自观测不足还是输入产品失真;若原因不同,补采样和改进协变量的优先顺序也会不同。

### 8.2. 深层 SOC 与剖面制图

深层 SOC 制图同时受到剖面观测不足和深度信息表达方式的制约。将深度纳入模型可建立层间联系,但同一剖面的土层若被随机分入训练集与验证集,共享信息仍可能高估对新剖面的预测能力。采用整剖面留出,比较分层方法和联合深度模型的误差及区间覆盖,更能反映其在新剖面上的表现[10] [19]。

深层 SOC 的过程分析依赖根系、矿物、水文及埋藏资料,其与表层 NDVI 或管理变量的关系可能不同。对不规则土层进行标准化时,插值和储量积分会影响剖面形态与碳量估计。保留可能的埋藏富碳层,有助于避免平滑处理掩盖具有过程意义的局部变化。

### 8.3. 时间预测与变化识别

SOC 动态制图的前提是观测与协变量的时间对应。用某一年 SOC 样点训练的空间模型替换其他年份的遥感输入，可以生成情景预测，但时间变化的可信度仍需重复采样或独立历史资料检验。两期 SOC 地图相减也会叠加两期误差，其差值能否识别实际变化，取决于预测误差及其跨期相关性[6]。

区分空间差异和时间变化，是 SOC 动态制图中的主要困难。不同年份、不同地点的样本同时进入模型时，估计的时间关系可能夹杂地点差异。重复采样、测定方法的一致性和跨时间验证能够为变化识别提供更直接的依据，也是进一步讨论管理与气候情景的基础。

### 8.4. 管理数据与遥感协变量

管理信息整合的首要任务不是继续增加代理变量，而是建立可核查的地块管理历史。物候可反映生长时序，却不能代替还田量、施肥量或措施持续时间[14]。后续监测宜将地块位置、措施起止时间、投入强度和土壤预测一一对应，并保存调查误差与遥感分类不确定性；第 4.6 节的候选方法只有在记录口径可比、层级标识可信且时间顺序明确时才具有应用基础。

### 8.5. 过程约束的适用边界

过程融合的后续研究应把关注点从单次精度提升转向增益来源与失效条件。第 4.4 节所述模拟样本扩充、参数传递和过程约束提供了不同的信息，宜在相同实测数据与检验集下分别比较，说明改善究竟来自额外驱动信息、训练样本构造还是结构限制。

对资料稀疏区，应优先检验关键参数能否识别，以及碳输入、初始碳库和过程结构的误差能否被观测约束。对新区域、新管理制度或变化的气候条件，应把适用性检验置于常规制图之前；过程名称本身不构成外推可靠性的证明。

### 8.6. 管理应用的证据缺口

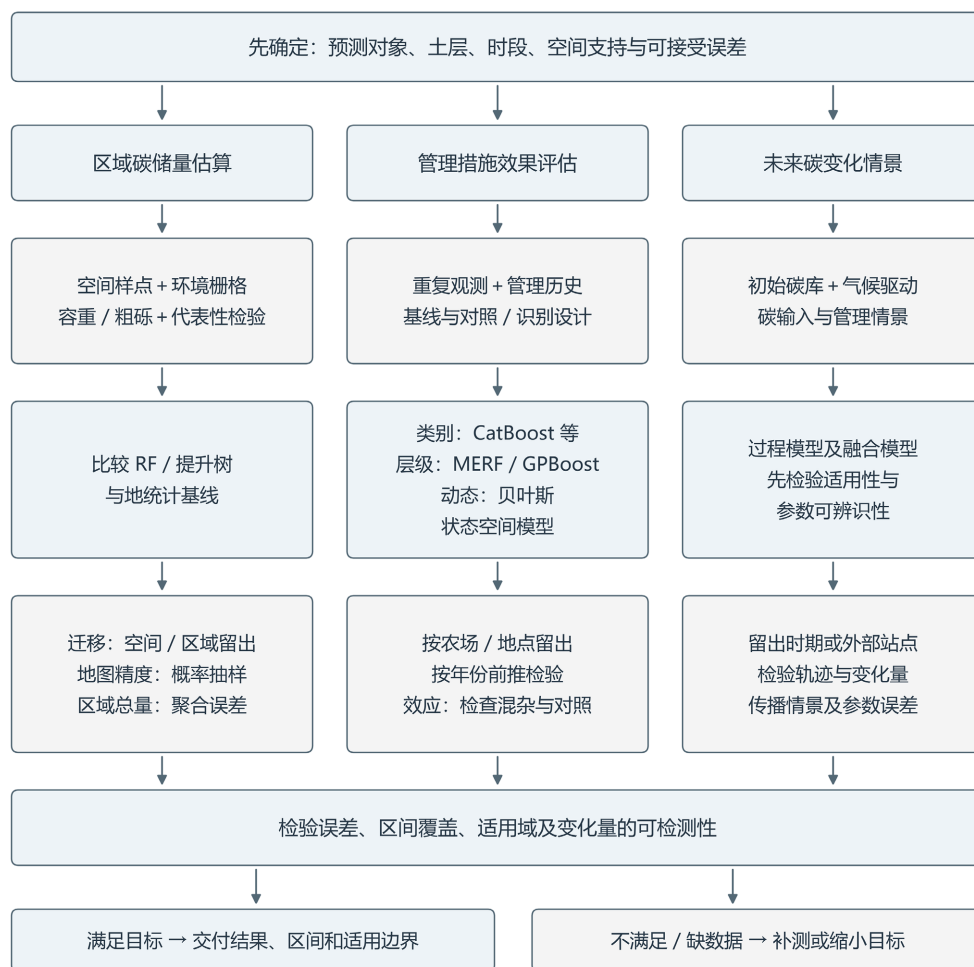
面向土壤改良、侵蚀防控和碳储量监测，当前短板并不只是地图精度，而是预测尺度、决策尺度与证据尺度常不一致。DSM 输出应同时给出预测值、区间和适用域，并说明相邻地块之间的差异是否超过可检出的误差[3] [10] [33]。区域平均表现良好的模型，未必能够支持地块级措施排序。

水土保持措施的碳效应还涉及基线、可比对照、重复观测和侧向碳迁移。固定土层储量、等质量土壤储量与景观碳收支回答不同问题；缺少实施前资料或对照时，地块间 SOC 差异容易混入土壤、坡位和历史管理差异[11] [20] [21]。因此，现阶段 SOC 地图可用于确定监测重点和形成假设；只有在质量口径一致、处理分配与混杂得到控制，并核算侵蚀 - 沉积边界后，才宜进一步估计措施净效应。第 9 节据此给出目标驱动的操作流程，不再把模型精度直接等同于管理证据。

## 9. 面向应用的模型与验证决策流程

模型选择应从拟解决的问题出发，而不是从算法清单出发。图 2 将本文讨论归纳为三条应用路径：区域碳储量估算、管理措施效果评估和未来变化情景分析。三者可以共享部分数据与算法，但预测对象、最低数据需求和验证证据并不相同。实施前应约定土层、时间窗口、空间支持和允许误差，并检查资料是否足以回答该问题。

区域估算以代表性的 SOC 样点和环境协变量为基础；若目标是碳储量，还需容重、粗砾与土层信息，不能只将含量图直接相加。可以比较地统计、RF 及提升树等基线；空间或区域留出用于检验迁移，基于概率抽样的独立样本用于目标区域地图精度评价。估算区域总量时，还须考虑面积权重及预测误差的空间相关，不能把每个像元视为独立重复[27] [30]。



注：本文归纳的应用框架，并非算法性能排序。三条路径均以满足相应数据条件为前提；缺少容重、管理历史或独立观测时，应先补充资料或调整研究目标。管理效应的因果识别需要另外设计，不能由预测验证替代。

**Figure 2.** Decision workflow for model selection and validation across SOC applications

**图 2.** 面向不同 SOC 应用的模型选择与验证决策流程

管理效果评估需要措施历史、基线及重复 SOC 观测，并以对照或明确的识别假设支持效应解释。模型可依据数据结构在第 4.6 节的方法中选择，但 CatBoost、MERF 和 GPBoost 在本文中是与数据结构相匹配的候选方法，并非已经在同一 SOC 数据集上证明优于常规模型；CQUESST 提供的是试验地块层面的动态应用证据[26]。没有管理记录或可比对照时，应将目标限定为关联分析或探索性情景。未来情景则需要初始状态、气候和碳输入序列，先检验过程模型适用性，再用留出时期或外部站点检查碳库水平与变化轨迹[18][27]。

最后一道判断是误差是否满足用途，而非某个通用  $R^2$  阈值。应结合 RMSE、偏差、预测区间覆盖和适用域评价结果；对管理引起的变化，还要判断采样和测量精度是否足以检出预期变化。未达到要求时，应优先补充关键观测、缩小适用范围或延长监测，而不是仅更换算法[27][33]。水土保持情景还需纳入侵蚀与沉积的侧向转移边界；在未核算区域碳收支和相关排放之前，模型给出的储量差异应保留为土壤碳变化估计。

## 10. 结论

SOC 空间预测方法的适用性取决于样点布局、协变量质量和预测任务，现有证据尚不足以支持固定

的算法排序。过程-机器学习融合应明确参数、状态或模拟样本的传递方式，并在相同实测检验集上辨别新增信息的贡献；管理数据建模则应匹配分类、重复观测与层级结构，而非只增加代理变量。

空间分块验证与目标区域地图精度评价服务于不同的推断目标，前者考察地理隔离条件下的预测能力，后者依赖相应的抽样设计和估计方法。预测区间与适用域能够补充误差和外推风险信息，其可信度仍需独立检验及样本代表性支撑。

面向区域估算、管理评价和未来情景的研究，应沿“目标-数据-模型-验证”的顺序作出选择。模型贡献、统计响应与过程证据仍需区分。对于管理和水力侵蚀应用，措施历史、重复观测、可比对照及一致的碳质量核算，是由空间关联走向措施效应评价的基础；精度或证据不足时，应限定结论，而不以模型复杂度补足观测缺口。

## 参考文献

- [1] Schmidt, M.W.I., Torn, M.S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I.A., *et al.* (2011) Persistence of Soil Organic Matter as an Ecosystem Property. *Nature*, **478**, 49-56. <https://doi.org/10.1038/nature10386>
- [2] McBratney, A.B., Mendonça Santos, M.L. and Minasny, B. (2003) On Digital Soil Mapping. *Geoderma*, **117**, 3-52. [https://doi.org/10.1016/s0016-7061\(03\)00223-4](https://doi.org/10.1016/s0016-7061(03)00223-4)
- [3] 朱阿兴, 杨琳, 樊乃卿, 等. 数字土壤制图研究综述与展望[J]. 地理科学进展, 2018, 37(1): 66-78.
- [4] Arrouays, D., Grundy, M.G., Hartemink, A.E., *et al.* (2014) GlobalSoilMap: Toward a Fine-Resolution Global Grid of Soil Properties. In: *Advances in Agronomy Vol. 125*, Academic Press, 93-134.
- [5] Lamichhane, S., Kumar, L. and Wilson, B. (2019) Digital Soil Mapping Algorithms and Covariates for Soil Organic Carbon Mapping and Their Implications: A Review. *Geoderma*, **352**, 395-413. <https://doi.org/10.1016/j.geoderma.2019.05.031>
- [6] Wadoux, A.M.J.C., Minasny, B. and McBratney, A.B. (2020) Machine Learning for Digital Soil Mapping: Applications, Challenges and Suggested Solutions. *Earth-Science Reviews*, **210**, Article 103359. <https://doi.org/10.1016/j.earscirev.2020.103359>
- [7] Chen, S., Arrouays, D., Leatitia Mulder, V., Poggio, L., Minasny, B., Roudier, P., *et al.* (2022) Digital Mapping of GlobalSoilMap Soil Properties at a Broad Scale: A Review. *Geoderma*, **409**, Article 115567. <https://doi.org/10.1016/j.geoderma.2021.115567>
- [8] Meyer, H., Reudenbach, C., Wöllauer, S. and Nauss, T. (2019) Importance of Spatial Predictor Variable Selection in Machine Learning Applications—Moving from Data Reproduction to Spatial Prediction. *Ecological Modelling*, **411**, Article 108815. <https://doi.org/10.1016/j.ecolmodel.2019.108815>
- [9] Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., *et al.* (2017) SoilGrids250m: Global Gridded Soil Information Based on Machine Learning. *PLOS ONE*, **12**, e0169748. <https://doi.org/10.1371/journal.pone.0169748>
- [10] Poggio, L., de Sousa, L.M., Batjes, N.H., Heuvelink, G.B.M., Kempen, B., Ribeiro, E., *et al.* (2021) SoilGrids 2.0: Producing Soil Information for the Globe with Quantified Spatial Uncertainty. *Soil*, **7**, 217-240. <https://doi.org/10.5194/soil-7-217-2021>
- [11] Wendt, J.W. and Hauser, S. (2013) An Equivalent Soil Mass Procedure for Monitoring Soil Organic Carbon in Multiple Soil Layers. *European Journal of Soil Science*, **64**, 58-65. <https://doi.org/10.1111/ejss.12002>
- [12] Minasny, B. and McBratney, A.B. (2016) Digital Soil Mapping: A Brief History and Some Lessons. *Geoderma*, **264**, 301-311. <https://doi.org/10.1016/j.geoderma.2015.07.017>
- [13] Schratz, P., Muenchow, J., Iturriza, E., Richter, J. and Brenning, A. (2019) Hyperparameter Tuning and Performance Assessment of Statistical and Machine-Learning Algorithms Using Spatial Data. *Ecological Modelling*, **406**, 109-120. <https://doi.org/10.1016/j.ecolmodel.2019.06.002>
- [14] 周琪清, 赵小敏, 郭熙, 等. 基于物候与极端气候信息的耕地土壤有机碳空间分布预测研究[J]. 土壤学报, 2024, 61(3): 648-661.
- [15] Lee, Y., Jeong, H., Du, L., Li, X., Zhang, X., Lee, B., *et al.* (2025) Impacts of Spatial Scales and Data Distribution Patterns on Geographically Weighted Machine Learning Models for the Spatial Prediction of Soil Organic Carbon. *Pedosphere*. (In Press) <https://doi.org/10.1016/j.pedsph.2025.11.010>
- [16] Tao, F., Huang, Y., Hungate, B.A., Manzoni, S., Frey, S.D., Schmidt, M.W.I., *et al.* (2023) Microbial Carbon Use Efficiency Promotes Global Soil Carbon Storage. *Nature*, **618**, 981-985. <https://doi.org/10.1038/s41586-023-06042-3>

- [17] Zhang, L., Heuvelink, G.B.M., Mulder, V.L., Chen, S., Deng, X. and Yang, L. (2024) Using Process-Oriented Model Output to Enhance Machine Learning-Based Soil Organic Carbon Prediction in Space and Time. *Science of the Total Environment*, **922**, Article 170778. <https://doi.org/10.1016/j.scitotenv.2024.170778>
- [18] Coleman, K. and Jenkinson, D.S. (2014) RothC—A Model for the Turnover of Carbon in Soil: Model Description and Users Guide. Rothamsted Research. <https://www.rothamsted.ac.uk/media/8/download>
- [19] Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., *et al.* (2017) Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure. *Ecography*, **40**, 913-929. <https://doi.org/10.1111/ecog.02881>
- [20] Kirkels, F.M.S.A., Cammeraat, L.H. and Kuhn, N.J. (2014) The Fate of Soil Organic Carbon upon Erosion, Transport and Deposition in Agricultural Landscapes—A Review of Different Concepts. *Geomorphology*, **226**, 94-105. <https://doi.org/10.1016/j.geomorph.2014.07.023>
- [21] Doetterl, S., Six, J., Van Wesemael, B. and Van Oost, K. (2012) Carbon Cycling in Eroding Landscapes: Geomorphic Controls on Soil Organic C Pool Composition and C Stabilization. *Global Change Biology*, **18**, 2218-2232. <https://doi.org/10.1111/j.1365-2486.2012.02680.x>
- [22] Nussbaum, M., Spiess, K., Baltensweiler, A., Grob, U., Keller, A., Greiner, L., *et al.* (2018) Evaluation of Digital Soil Mapping Approaches with Large Sets of Environmental Covariates. *Soil*, **4**, 1-22. <https://doi.org/10.5194/soil-4-1-2018>
- [23] Prokhorenkova, L., Gusev, G., Vorobev, A., *et al.* (2018) CatBoost: Unbiased Boosting with Categorical Features. *Advances in Neural Information Processing Systems* 31, Montréal, 2-8 December 2018, 6638-6648.
- [24] Hajjem, A., Bellavance, F. and Larocque, D. (2014) Mixed-Effects Random Forest for Clustered Data. *Journal of Statistical Computation and Simulation*, **84**, 1313-1328. <https://doi.org/10.1080/00949655.2012.741599>
- [25] Sigrist, F. (2022) Gaussian Process Boosting. *Journal of Machine Learning Research*, **23**, 1-46.
- [26] Pagendam, D., Baldock, J., Clifford, D., Farquharson, R., Murray, L., Beare, M., *et al.* (2025) CQUESST: A Dynamical Stochastic Framework for Predicting Soil-Carbon Sequestration. *The Annals of Applied Statistics*, **19**, 2005-2026. <https://doi.org/10.1214/25-aos2057>
- [27] Smith, P., Soussana, J., Angers, D., Schipper, L., Chenu, C., Rasse, D.P., *et al.* (2020) How to Measure, Report and Verify Soil Carbon Change to Realize the Potential of Soil Carbon Sequestration for Atmospheric Greenhouse Gas Removal. *Global Change Biology*, **26**, 219-241. <https://doi.org/10.1111/gcb.14815>
- [28] Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Brown, D.J., Demattê, J.A.M., Shepherd, K.D., *et al.* (2016) A Global Spectral Library to Characterize the World's Soil. *Earth-Science Reviews*, **155**, 198-230. <https://doi.org/10.1016/j.earscirev.2016.01.012>
- [29] Brenning, A. (2012). Spatial Cross-Validation and Bootstrap for the Assessment of Prediction Rules in Remote Sensing: The R Package Sperrorest. 2012 *IEEE International Geoscience and Remote Sensing Symposium*, Munich, 22-27 July 2012, 5372-5375. <https://doi.org/10.1109/igarss.2012.6352393>
- [30] Wadoux, A.M.J.C., Heuvelink, G.B.M., de Bruin, S. and Brus, D.J. (2021) Spatial Cross-Validation Is Not the Right Way to Evaluate Map Accuracy. *Ecological Modelling*, **457**, Article 109692. <https://doi.org/10.1016/j.ecolmodel.2021.109692>
- [31] Meinshausen, N. (2006) Quantile Regression Forests. *Journal of Machine Learning Research*, **7**, 983-999.
- [32] Angelopoulos, A.N. and Bates, S. (2023) Conformal Prediction: A Gentle Introduction. *Foundations and Trends in Machine Learning*, **16**, 494-591. <https://doi.org/10.1561/22000000101>
- [33] Meyer, H. and Pebesma, E. (2021) Predicting into Unknown Space? Estimating the Area of Applicability of Spatial Prediction Models. *Methods in Ecology and Evolution*, **12**, 1620-1633. <https://doi.org/10.1111/2041-210x.13650>
- [34] Lundberg, S.M. and Lee, S.I. (2017) A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems* 30, Long Beach, 4-9 December 2017, 4765-4774.
- [35] Fotheringham, A.S., Yang, W. and Kang, W. (2017) Multiscale Geographically Weighted Regression (MGWR). *Annals of the American Association of Geographers*, **107**, 1247-1265. <https://doi.org/10.1080/24694452.2017.1352480>
- [36] Wang, J., Li, X., Christakos, G., Liao, Y., Zhang, T., Gu, X., *et al.* (2010) Geographical Detectors-Based Health Risk Assessment and Its Application in the Neural Tube Defects Study of the Heshun Region, China. *International Journal of Geographical Information Science*, **24**, 107-127. <https://doi.org/10.1080/13658810802443457>
- [37] 王劲峰, 徐成东. 地理探测器: 原理与展望[J]. 地理学报, 2017, 72(1): 116-134.
- [38] 黄思华, 濮励杰, 解雪峰, 等. 面向数字土壤制图的土壤采样设计研究进展与展望[J]. 土壤学报, 2020, 57(2): 259-272.