

遥感影像的辐射定标是确保多时相、多传感器数据一致性与可比性的基础，其目的是消除传感器响应非一致性所带来的定量测值偏差。传统的定标方法多依赖于预发射实验室标定与地基绝对定标，在成本与可推广性方面存在明显局限。相对定标技术以公共地面目标为参照，通过统计回归模型对不同时期或不同传感器影像进行批量校正，成为当前定量遥感领域的研究热点。

Chander 等（2009）^[1]基于多年份 Landsat MSS、TM、ETM+与 EO-1 ALI 公共样本，系统总结了大气顶反射率及辐射率的线性标定系数为波段级增益—偏移参数表，为后续批量化定标提供了权威基准；Slater 等（2001）^[2]进一步提出了统一的太阳反射波段定标方案，将预发射、在轨太阳漫反射板与实际地基反演校正结合，实现了从太阳辐照度尺度到现场反射率尺度的无缝衔接。

在线性回归框架下，定标模型假设同一地物目标在不同观测条件下的 DN 与目标真实辐射值满足线性关系，通过最小二乘拟合获得波段增益与偏移，从而实现波段级映射。该方法具有模型可解释性强、参数可追溯、计算效率高等优点，已被广泛应用于同传感器内部归一化及跨传感器映射研究。然而，当影像数据存在光谱响应非线性差异或地物光谱曲线显著变化时，线性模型的拟合精度将受到限制。

为克服线性模型的局限，近年研究者开始将非参数机器学习方法引入定量遥感定标。随机森林作为一种集成学习算法，通过生成多棵决策树并进行分类或回归，在处理高维、多变量、非线性映射问题时表现出优异的性能和对异常值的不敏感性。Breiman（2001）^[3]提出了随机森林的数学原理，证明了其随着树数增加泛化误差收敛的特性，并引入了基于置乱的变量重要性度量；Belgiu 与 Drăguț（2016）^[4]综述了随机森林在遥感分类、回归、特征选择等多领域的应用与未来发展方向，为多波段联合回归提供了方法学指导。在本研究中，我们利用多输出回归器将随机森林扩展至多输出波段回归，实现了同传感器内部归一化与跨传感器映射的统一建模框架，显著提升了定标精度与适用性。

本研究以 Landsat TM（Thematic Mapper）系列与 Landsat OLI（Operational Land Imager）系列影像为对象，利用公共地面样本点开展两种定标方案：基于波段级线性回归的经典定标与基于随机森林多输出回归的非参数定标。以下从数据准备、同传感器内部归一化、跨传感器映射与多输出回归四个方面，系统阐述所用方法与核心数学原理。

（1）数据准备

选取 1998、2003、2008 年 Landsat TM 与 2013、2018 年 Landsat OLI 影像中相同地面目标的公共样本点，提取各波段数字号(DN)或地表反射率值。TM 传感器波段包括蓝(Blue)、绿(Green)、红(Red)、近红外 1(NIR1)、近红外 2(NIR2)及短波红外(MIR)；OLI 传感器波段则为近海气溶胶(Coastal Aerosol)、蓝、绿、红、NIR1、短波红外 1(SWIR1)与短波红外 2(SWIR2)。为保证不同年份、不同传感器数据的可对比性，本研究仅保留各传感器中与目标定标最相关的公共波段(TM: Blue、Green、Red、NIR1、SWIR1←MIR；OLI: Blue、Green、Red、NIR1、SWIR1)，并以“点编号”(point_id)作为索引对齐样本。

（2）基于线性回归的同传感器内部归一化

针对同一传感器不同年份影像的响应差异，选择 2003 年 TM 与 2013 年 OLI 分别作为参考年，采用简单线性回归模型将其他年份波段值归一到参考年水平。以某传感器的第 i 个波段为例，令 $x = [x_1, \dots, x_n]^T$ 为待校正年份的样本点 DN 值向量， $y = [y_1, \dots, y_n]^T$ 为参考年份同一波段的对应值，假设二者满足线性关系：

$$y_j = ax_j + b + \varepsilon_j, j = 1, \dots, n,$$

其中 a 为增益系数、 b 为偏移项、 ε_j 为残差。通过最小二乘法求解：

$$\hat{a}, \hat{b} = \arg \min_{a,b} \sum_{j=1}^n (y_j - ax_j - b)^2$$

归一化后, 得到校正值 $\hat{y}_j = \hat{a}x_j + \hat{b}$ 。对所有公共波段分别拟合后, 可计算归一化前后均方根误差 (RMSE) 和拟合优度 R^2 以评估模型性能。

(3) 基于线性回归的跨传感器映射

完成同传感器内部归一化后, 利用 2003 年 TM 与 2013 年 OLI 样本点的配对数据进行跨传感器映射。对每一公共波段再应用线性模型:

$$\rho_i^{\text{OLI}} = a_i \rho_i^{\text{TM}} + b_i + \varepsilon_i, \quad i \in \{\text{Blue, Green, Red, NIR1, SWIR1}\}$$

其中 ρ_i^{TM} 与 ρ_i^{OLI} 分别为归一化后 TM 与 OLI 在该波段的反射率。通过同样的最小二乘拟合, 获得跨传感器映射系数 (a_i, b_i) , 并以此模型将所有归一化后的 TM 数据转换至 OLI 波段空间。

(4) 基于随机森林多输出回归的定标

为捕捉地物光谱与传感器响应间的非线性关系, 本研究进一步采用随机森林结合多输出回归器统一完成内部归一化与跨传感器映射。给定输入矩阵 $X \in \mathbb{R}^{n \times p}$ (各波段的待校正值) 与输出矩阵 $Y \in \mathbb{R}^{n \times q}$ (参考年或目标传感器的对应波段值), 构建模型

$$Y = f(X) + E$$

其中 f 由多棵回归树的集合定义, 单树 t_k 的预测为 $t_k(x)$, 随机森林的输出为各树预测值的平均:

$$f(x) = \frac{1}{K} \sum_{k=1}^K t_k(x)$$

利用训练集 $(X_{\text{train}}, Y_{\text{train}})$ 拟合模型, 并在测试集上计算各波段的 RMSE 与 R^2 。对于内部归一化, 分别以 TM 与 OLI 不同年份样本为输入、参考年样本为输出; 对于跨传感器映射, 则以归一化后的 TM 样本为输入、OLI 样本公共波段为输出。该方法无需预设函数形式, 能够在高维、多变量条件下自动学习最优映射规则, 并通过袋外误差估计 (out-of-bag error) 辅助模型调参。