|  |  |
| --- | --- |
| ICS | 65.020.01 |
| CCS | |  | | --- | | D:\000000部门项目\09标准化插件开发\程序源代码\StandardEditor_ShanDongKeXieYuan\团标首页面字母T.pngD:\000000部门项目\09标准化插件开发\程序源代码\StandardEditor_ShanDongKeXieYuan\团标首页面字母T后面的反斜杠.png GXAS |   B 05 |

团体标准

T/GXAS XXXX—XXXX

森林土壤有机碳GF-5高光谱影像反演技术规程

Technical code of practice for GF-5 hyperspectral image inversion of forest soil organic carbon

XXXX - XX - XX发布

XXXX - XX - XX实施

广西标准化协会  发布

目次

[前言 II](#_Toc207868562)

[1 范围 1](#_Toc207868563)

[2 规范性引用文件 1](#_Toc207868564)

[3 术语和定义 1](#_Toc207868565)

[4 缩略语 1](#_Toc207868566)

[5 反演流程 2](#_Toc207868567)

[6 数据准备 2](#_Toc207868568)

[6.1 GF-5高光谱影像 2](#_Toc207868569)

[6.2 土壤实测数据 2](#_Toc207868570)

[7 GF-5高光谱影像预处理 3](#_Toc207868571)

[7.1 影像获取 3](#_Toc207868572)

[7.2 辐射定标 3](#_Toc207868573)

[7.3 大气校正 3](#_Toc207868574)

[7.4 几何校正 4](#_Toc207868575)

[7.5 影像裁剪 4](#_Toc207868576)

[8 光谱数据采集与预处理 4](#_Toc207868577)

[8.1 光谱仪选择 4](#_Toc207868578)

[8.2 光谱采集 4](#_Toc207868579)

[8.3 光谱预处理 5](#_Toc207868580)

[9 光谱特征波段筛选 5](#_Toc207868581)

[9.1 相关性分析 5](#_Toc207868582)

[9.2 算法筛选 5](#_Toc207868583)

[10 反演模型构建 6](#_Toc207868584)

[10.1 数据集划分 6](#_Toc207868585)

[10.2 模型建立 6](#_Toc207868586)

[10.3 模型参数优化 6](#_Toc207868587)

[11 反演模型验证 7](#_Toc207868588)

[12 反演结果输出 7](#_Toc207868589)

[12.1 反演影像生成 7](#_Toc207868590)

[12.2 成果报告 7](#_Toc207868591)

[参考文献 8](#_Toc207868592)

1. 前言

本文件参照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由广西壮族自治区林业科学研究院提出和宣贯。

本文件由广西标准化协会归口。

本文件起草单位：广西壮族自治区林业科学研究院、桂林理工大学、广西华沃特集团股份有限公司、南宁师范大学、四川省林业科学研究院、中国林业科学研究院亚热带林业研究所、中南林业科技大学、广西大学、广西森林资源与生态环境监测中心、广西壮族自治区林业勘测设计院、中国林业科学研究院热带林业实验中心。

本文件主要起草人：石媛媛、唐健、赵隽宇、潘桂颖、邓昀、吴立潮、宋贤冲、覃祚玉、柯琴、蒋湖波、刘毅。

森林土壤有机碳GF-5高光谱影像反演技术规程

* 1. 范围

本文件界定了森林土壤有机碳高分五号卫星（GF-5）高光谱影像反演涉及的术语和定义、以及缩略语，确立了利用GF-5高光谱影像反演森林土壤有机碳含量的程序，规定了数据准备、GF-5高光谱影像预处理、光谱数据采集与预处理、光谱特征波段筛选、反演模型构建、反演模型验证、反演结果输出等要求。

本文件适用于利用GF-5高光谱影像定量反演森林土壤有机碳含量的操作。

* 1. 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 14950 摄影测量与遥感术语

GB/T 36301 航天高光谱成像数据预处理产品分级

HJ 695 土壤 有机碳的测定 燃烧氧化-非分散红外法

NY/T 1121.1 土壤检测 第1部分：土壤样品的采集、处理和贮存

* 1. 术语和定义

GB/T 14950、GB/T 36301界定的以及下列术语和定义适用于本文件。

GF-5高光谱影像 GF-5 hyperspectral imagery

由高分五号（GF-5）卫星搭载的高光谱成像仪（AHSI）获取的包含连续窄波段的光学遥感影像，覆盖可见光-近红外-短波红外波段（400nm～2500nm），空间分辨率为30m。

高光谱反演 hyperspectral inversion

基于高光谱影像的光谱特征与土壤属性之间的定量关系，通过数学模型估算目标区域土壤属性值的过程。

土壤有机碳 soil organic carbon

土壤矿质土和有机土（包括泥炭土、沙砾层）中的有机碳储量。

* 1. 缩略语

下列缩略语适用于本文件。

DN：数字灰度值（digital number）

FLAASH：快速光谱高光谱立方体的大气分析（fast line-of-sight atmospheric analysis of spectral hypercubes）

LeakyReLU：泄漏修正线性单元（leaky rectified linear unit）

ReLU：修正线性单元（rectified linear unit）

RMSE：均方根误差（root mean squared error）

* 1. 反演流程

见图1。

反演结果输出

数据准备

GF-5高光谱影像预处理

光谱数据采集与预处理

光谱特征波段筛选

反演模型构建

反演模型验证

1. 反演流程图
   1. 数据准备
      1. GF-5高光谱影像
         1. 数据源

选用GF-5卫星高光谱成像仪传感器获取的Level 2级（辐射定标后）或Level 3级（辐射定标+大气校正后）产品。波段范围覆盖（400nm～2500nm），波段数≥330个，空间分辨率30m。时相选择应结合区域特征：

1. 热带、亚热带地区：优先选择植被生长季前期（3～4月）或枯落物分解较快的秋季（10～11月），避开雨季（6～8月）；
2. 温带地区：选择春季（4～5月）植被萌发前或秋季（9～10月）枯落物覆盖较少时段，避开冬季积雪期（12～2月）。
   * + 1. 数据质量

宜选择重叠缓冲区≥500m包含目标森林且边界完整的影响区域；影像获取时间与地面采样时间间隔≤3个月，且影像无严重条带、噪声或云覆盖（云量≤10％）。

* + - 1. 数据获取

通过国家高分辨率对地观测系统广西数据与应用中心、中国资源卫星应用中心等官方平台获取，获取时需同步下载影像对应的元数据文件（包含辐射定标系数、几何定位参数、成像时间等信息）。

* + 1. 土壤实测数据
       1. 采样点布设

采用网格化布点与典型抽样相结合的分层布设方法，覆盖目标森林土壤类型、植被类型及地形梯度：

1. 基础网格布设：利用ArcGIS等地理信息系统（GIS）软件，基于研究区数字高程模型（DEM）和土壤类型图，划分1km×1km基础网格，若研究区地形复杂（坡度＞25°），网格大小可调整为0.5km×0.5km；
2. 典型抽样补充：在每个基础网格内，根据森林类型（针叶林、阔叶林、混交林）、林分年龄（幼龄林、中龄林、成熟林）、土壤质地（砂土、壤土、黏土）设置典型样地，形成采样单元；
3. 样本数量要求：总采样点数量≥100个，且满足“样本数/模型参数个数＞5”，若采用复杂模型（如卷积神经网络），样本数宜≥200个。
   * + 1. 样品采集

按NY/T 1121.1的规定执行。

* + - 1. 样品制备

将采集到土壤自然风干研磨，过2mm筛去除粗颗粒，部分样品进一步过0.149mm筛，用于土壤有机碳含量测定和地面土壤光谱数据采集。

* + - 1. 土壤有机碳含量测定

按照HJ 695的规定执行。

* 1. GF-5高光谱影像预处理
     1. 影像获取

使用ENVI 5.6、ArcGIS 10.8等专业软件对下载的影像压缩包进行解压，获取研究区GF-5高光谱影像数据，影像时间应与地面采样时间相近。影像应包含可见光-近红外波段400nm～2500nm。

* + 1. 辐射定标

将DN值转换为表观反射率（或辐亮度）见公式（1）：

()

式中：

——波段；

的数字化量值；

卫星提供的增益与偏置系数（来自元数据文件），单位分别为μW/(cm²・sr・nm)、μW/(cm²・sr・nm)。

* + 1. 大气校正

根据影像数据级别和研究区精度需求选择合适的校正方法，如下。

1. FLAASH方法，适用于Level 2级影像，满足常规精度需求：
   1. 大气模型：根据影像获取时间和经纬度确定，热带地区纬度＜23.5°选择Tropical模型，中纬度夏季（6～8月，纬度23.5°～66.5°）选择Mid-Latitude Summer模型，中纬度冬季（12～2月）选择Mid-Latitude Winter模型；
   2. 气溶胶模型：森林地区以自然气溶胶（如灰尘、花粉）为主，宜选择Rural(乡村)模型；若研究区靠近城市或工业区（距离≤10km），受人为气溶胶（如工业粉尘、汽车尾气）影响较大，选择Urban（城市）模型；
   3. 气溶胶反演：采用2-Band(K-T)方法，使用1240nm和550nm波段，反演气溶胶光学厚度；
   4. 能见度设置：初始值设为40km，根据研究区空气质量调整，污染严重地区（如工业密集区）可降至20km～30km，清洁地区（如原始林区）可提高至50km～80km。
   5. 水汽反演：利用1135 nm波段的水汽吸收特征反演大气柱水汽含量；若研究区有同步（影像获取前后1小时内）地面气象站水汽观测数据，可将其作为参考值，验证并调整反演结果。
2. 6S模型方法，适用于精度要求高的场景：
   1. 大气模式：根据研究区季节和纬度从6S模型提供的6种大气模式中选择，热带模式（纬度＜23.5°，全年适用）、中纬度夏季模式（纬度 23.5°～66.5°，6～8月）、中纬度冬季模式（纬度23.5°～66.5°，12～2月）、副极地夏季模式（纬度＞66.5°，6～8月）、副极地冬季模式（纬度＞66.5°，12～2月）、平流层模式（高海拔地区，海拔＞3000m）；
   2. 气溶胶类型：森林区域默认选择大陆型气溶胶（模型内置类型2）；若研究区为干旱半干旱森林（如西北部分林区，年降水量＜400mm），受沙尘影响较大，选择沙漠型气溶胶（模型内置类型3）；若研究区临近沿海（距离海岸线＜50km），选择海洋型气溶胶（模型内置类型1）；
   3. 气溶胶光学厚度：宜优先使用同步气溶胶机器人网络（AERONET）站点实测数据或中分辨率成像光谱仪（MODIS）气溶胶产品；
   4. 目标高度：从研究区数字高程模型（DEM，分辨率≤30m）中提取平均海拔高度，精确至0.1km，作为目标高度输入模型；
   5. 太阳与观测几何参数输入：从GF-5影像元数据中提取成像时的太阳高度角、太阳方位角、传感器方位角，精确至0.1°，按6S模型输入格式要求填写；
   6. 校正验证：使用同步地面实测反射率数据或伪不变特征点(PIFs)验证校正结果；若相对误差＞15％，需重新调整气溶胶光学厚度、水汽含量等参数，再次进行校正。
      1. 几何校正

采用二次多项式等方法对影像进行几何校正，符合高精度地形图或DOM，校正误差控制≤1个像元，具体要求如下。

1. 参考数据选择：采用研究区1:10000或1:50000高精度地形图（国家测绘地理信息局发布，坐标系为2000国家大地坐标系）或数字正射影像图（DOM，分辨率≤2.5m，坐标系与目标坐标系一致）作为参考数据。
2. 控制点选择：在影像和参考数据中均匀选取20个以上控制点，控制点应满足以下条件：
   1. 地物特征明显且稳定（如道路交叉点、河流拐点、山顶、桥梁端点、标志性建筑物角点等）；
   2. 避开植被覆盖密集区、云阴影区、水体区域；
   3. 在影像边缘和中心区域均有分布；
   4. 使用ENVI或ArcGIS的控制点工具，记录控制点在影像和参考数据中的坐标，定位精度≤0.5个像元。
3. 校正方法与重采样：采用二次多项式变换方法建立影像坐标与地理坐标的映射关系，多项式阶数设为2阶；重采样方法选择双线性内插法；若定量反演科研场景等对像元值保真度要求极高的，可选择三次卷积内插法。
4. 误差控制：校正后计算每个控制点的RMSE，整体RMSE需≤1个像元（30m）；若单个控制点RMSE＞2个像元，删除该控制点并重新选取；若整体RMSE＞1个像元，需增加控制点数量（增至25个以上）或更换参考数据，重新进行校正。
5. 地形校正（可选）：若研究区地形复杂（坡度＞15°，如山地林区），需进行地形校正，消除地形阴影对光谱的影响；推荐采用SCS+C校正模型（Sun-Canopy-Sensor+C校正）方法。
   * 1. 影像裁剪

根据研究区边界矢量数据，裁剪影像至目标森林区域（保留≥500m缓冲区）。

* 1. 光谱数据采集与预处理
     1. 光谱仪选择

宜采用ASD FieldSpec 4地物光谱仪，性能应满足以下要求：

1. 光谱范围应覆盖350nm～2500nm，与GF-5高光谱影像波段范围匹配；
2. 光谱分辨率满足350nm～1000nm波段≤3nm，1000nm～2 500nm波段≤10nm；
3. 采样间隔满足350nm～1000 nm波段≤1nm，1000nm～2500nm波段≤2nm；
4. 仪器信噪比≥500:1（在700nm波段，带宽10nm条件下）。
   * 1. 光谱采集

在暗室内进行光谱数据采集，具体步骤如下：

1. 采集环境准备：暗室温度控制在20℃～25℃，相对湿度≤60％；不应气流直接吹拂样品和光纤探头；将土壤样品（2mm筛下样品）装入聚四氟乙烯样品皿，样品厚度≥1cm，表面压平；
2. 光源与探头设置：将卤素灯固定在样品皿正上方，光源入射角设为45°（与样品表面夹角），光源距离样品表面50cm～60cm；将光纤探头固定在样品皿正上方，探头距离样品表面5cm～10cm，探头轴线与样品表面垂直，视场角覆盖整个样品皿表面；
3. 白板校正：每次采集样品前，将标准白板放置在样品皿位置，运行光谱仪采集程序，进行白板校正，获取参考光谱；校正完成后，检查参考光谱反射率是否在99％左右，若偏差＞2％，需重新清洁白板并校正；每间隔30min或采集10个样品后，重新进行白板校正；
4. 样品光谱采集：将装有土壤样品的样品皿放置在光源与探头之间，运行光谱仪采集程序，每个样品采集10次光谱反射率；采集过程中，避免触碰仪器和样品皿；采集完成后，光谱仪自动计算10次反射率的平均值，作为该样品的最终光谱数据。
   * 1. 光谱预处理
        1. 噪声去除

去除光谱两端（＜400nm和＞2 400 nm）噪声较大的波段。

* + - 1. 平滑处理

采用离散小波变换降噪（DWT）、多元散射校正（MSC）、平滑去噪（SG）等方法对光谱数据进行平滑处理。

* + - 1. 数学变换

根据GF-5高光谱数据特性和土壤有机碳光谱响应特征，选择以下1～2种数学变换方法增强光谱特征。

1. 一阶导数(FDR)：
   1. 窗口大小：宜使用9-15点窗口(对应GF-5的10nm光谱分辨率)；
   2. 平滑方法：Savitzky-Golay滤波结合；
   3. 适用波段：宜选用土壤有机碳在1450 nm～1625nm（羟基吸收）、2200nm～2300 nm（碳氢键吸收）波段的特征吸收区。
2. 对数一阶导数(FDL)：
   1. 先对反射率取自然对数，再进行一阶导数计算；
   2. 宜用于弱吸收特征增强效果。
3. 标准正态变量变换(SNV)：
   1. 逐像元计算：对每个像元的光谱曲线进行中心化和标准化；
   2. 适用场景：适用于不同采样点的光谱数据标准化。
4. 连续统去除：
   1. 关键参数：凸包拟合节点数设为10～15个，节点选择在光谱反射率峰值处；
   2. 适用波段：适用于620nm～810nm有机碳敏感波段
   3. 光谱特征波段筛选
      1. 相关性分析

计算土壤有机碳含量与预处理后的光谱反射率及其变换形式的相关系数，筛选出相关性显著（p＜0.05）的波段作为候选特征波段。

* + 1. 算法筛选

采用递归特征消除（RFE）、随机森林特征（Ranger）算法等方法对候选特征波段进一步筛选，确定最优特征波段组合，算法如下。

1. 递归特征消除（RFE）：
   1. 基础模型选择：以偏最小二乘回归（PLSR）为基础模型，处理波段间共线性；
   2. 筛选步骤：将候选特征波段作为输入变量，土壤有机碳含量作为输出变量，构建RFE-PLSR模型；通过递归删除对模型贡献最小的波段（根据变量重要性系数VIP＜1筛选），每次删除5％～10％的波段，重新训练模型并计算交叉验证均方根误差（RMSEcv）；
   3. 最优波段确定：当交叉验证均方根误差（RMSEcv）达到最小值且不再显著下降时，停止筛选，保留的波段即为最优特征波段组合；记录最优波段数量（30～50个，避免过拟合）和对应的波长区间。
2. 随机森林特征选择（Ranger）算法：
   1. 模型参数设置：构建随机森林模型，决策树数量设为200～300棵（通过OOB误差确定，OOB误差稳定时的树数量为最优），最大特征数设为候选波段数的平方根，最小叶节点样本数设为5～10；
   2. 特征重要性计算：模型训练完成后，计算每个候选波段的特征重要性得分（基于节点不纯度减少量）；
   3. 波段筛选：按特征重要性得分从高到低排序，选取前30％～50％的波段作为最优特征波段；同时，通过变量置换检验（Permutation Test）验证特征重要性的可靠性，置换次数设为100次。
   4. 反演模型构建
      1. 数据集划分

将土壤有机碳含量数据与对应的光谱数据按照7:3或8:2的比例随机划分为训练集和验证集。

* + 1. 模型建立

以土壤有机碳含量为因变量，原始光谱反射率（R）、反射率一阶导数（FDR）、对数一阶导数（FDL）的全波段（400nm～2400nm）及显著相关波段（R：600nm～1150nm；FDR：500nm～540nm、630nm～680nm；FDL：980nm～1000nm、1250nm～1350nm、1550nm）数据为自变量，根据GF-5高光谱数据特性，选择以下4类适配模型。

1. 正交偏最小二乘法(OPLS)，适用于高维、小样本且线性特征明显的数据：
   1. 数据预处理：输入的数据应先经标准正态变量变换（SNV）和一阶导数组合预处理；
   2. 潜变量数：采用交叉验证确定，以交叉验证均方根误差（RMSEcv）最小化为目标，确定最优潜变量数，取值3～5个；
   3. 变量筛选：通过变量重要性投影（VIP）值筛选特征，保留VIP＞1的波段，剔除冗余变量；建模时以筛选后的波段为输入，土壤有机碳含量为输出，构建OPLS回归模型。
2. 随机森林(RF)，适用于非线性关系建模，抗噪声能力强，用于特征与目标变量间关系复杂的数据：
   1. 核心参数设置：决策树数量初始取值100～500，当OOB误差稳定（变化＜2％）时的树数量为最优；最大特征数设为特征波段总数的平方根；最小叶节点样本数设为5～10；
   2. 模型训练流程：将训练集数据随机bootstrap抽样（有放回抽样）生成多组子数据集，每组子数据集构建1棵决策树；每棵树生长时，在随机选取的特征波段中选择最优分裂节点；所有决策树训练完成后，通过投票（分类问题）或均值（回归问题）输出模型预测结果；
   3. 特征重要性分析：训练完成后，输出各特征波段的重要性得分，得分前10％的波段作为后续模型优化的核心变量。
3. 支持向量回归(SVR)，适用于小样本、非线性场景，通过核函数映射非线性特征，建立线性回归关系：
   1. 核函数选择：宜选择径向基函数（RBF）；
   2. 参数优化：采用网格搜索法结合交叉验证优化关键参数，惩罚系数C（取值范围10⁻²～10³）、核函数参数γ（取值范围10⁻³～10¹），优化目标为交叉验证均方根误差（RMSEcv）最小化，最终参数组合满足模型在训练集上的决定系数（R²）≥0.75；
   3. 数据标准化：输入的光谱特征波段数据应进行标准化处理（均值为0，标准差为1），土壤有机碳含量数据不用标准化，可作为输出变量。
4. 卷积神经网络(CNN),适用于复杂光谱特征提取，样本量≥200个的场景：
   1. 网络结构设计：宜选择3～5层一维卷积层，每层卷积核数量16～64个（从下到上递增），卷积核大小3～5；
   2. 训练参数设置：初始学习率设为0.001，损失函数采用均方误差；
   3. 激活函数：ReLU或LeakyReLU。
      1. 模型参数优化

采用交叉验证的确定系数（R²）和交叉验证的RMSE评价模型精度。R²越接近1，模型拟合度越好；RMSE越小，模型估算能力越强。同时，计算预测值与观测值的相对分析误差（RPD），RPD＞2时，表明模型具有较好的预测能力。

* 1. 反演模型验证

采用验证集数据对构建的模型进行验证，计算验证集的决定系数（R²）和RMSE。当R²≥0.7且RMSE较小时，模型可用于森林土壤有机碳含量反演。

* 1. 反演结果输出
     1. 反演影像生成

基于最优模型，将GF-5高光谱影像的筛选特征波段（或原始光谱）输入模型，逐像元计算土壤有机碳含量，生成与原始影像空间分辨率一致的SOC反演结果图。

* + 1. 成果报告

成果报告应包括以下内容：

1. 研究区概况：地理位置、地形地貌、土壤类型、森林类型等；
2. 数据处理过程：样品采集与处理、影像预处理、光谱预处理、特征波段筛选、模型构建与验证等；
3. 反演结果分析：有机碳含量的空间分布特征、平均值、变异系数等统计信息；
4. 精度评价：模型验证的R²和RMSE等指标；
5. 结论与建议：总结反演结果，提出森林土壤管理建议。

参考文献

[1] LY/T 2988-2018 森林生态系统碳储量计量指南

[2] 栾福明,张小雷,熊黑钢,等.基于不同模型的土壤有机质含量高光谱反演比较分析[J].光谱学与光谱分析,2013,33(1):196-200.

[3] 周伟,谢利娟,杨晗,等.基于高光谱的三江源区土壤有机质含量反演[J].土壤通报,2021,52(3):564-574.

[4] 于雷,洪永胜,耿雷,等.基于偏最小二乘回归的土壤有机质含量高光谱估算[J].农业工程学报,2015,31(14):103-109.

