

高光谱卫星遥感监测土壤有机质技术规范

Technical specification for monitoring soil organic matter by hyperspectral satellite
remote sensing

(点击此处添加与国际标准一致性程度的标识)

(征求意见稿)

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX - XX - XX 发布

XXXX - XX - XX 实施

目 次

前言	III
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	1
5 总体要求	1
5.1 空间参考	1
5.2 时间系统	2
5.3 监测时间	2
6 总体流程	2
7 资料收集与整理	3
7.1 资料收集	3
7.2 资料整理	3
8 数据获取与预处理	3
8.1 遥感数据选择	3
8.2 遥感数据预处理方法	3
8.3 裸土分布信息提取	3
9 土壤样品采集与测试分析	4
9.1 样点布设	4
9.2 样品采集	4
9.3 样品处理	5
9.4 样品有机质测定	5
10 模型构建与模型应用	5
10.1 模型构建	5
10.2 模型应用	5
10.3 精度验证	5
11 成果整理与资料归档	6
11.1 成果整理	6
11.2 资料归档	7
附录 A(资料性)最大类间方差法	8
附录 B(资料性)样品采样记录表示例	9
附录 C(资料性)常用的光谱特征计算	10
附录 D(资料性)常用的统计预测模型	12

附录 E(资料性) 监测成果图示例 13
参考文献 14

前 言

本文件按照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。

本文件由中华人民共和国自然资源部提出。

本文件由全国地理信息标准化技术委员会卫星应用分技术委员会（SAC/TC230/SC3）归口。

本文件起草单位：自然资源部国土卫星遥感应用中心、自然资源部国土整治中心、中国科学院南京土壤研究所、中国科学院空天信息创新研究院、中国农业科学院农业资源与农业区划研究所。

本文件主要起草人：尚坤、肖晨超、李少帅、王昌昆、张霞、刘佳。

高光谱卫星遥感监测土壤有机质技术规范

1 范围

本文件确立了高光谱卫星遥感监测土壤有机质的总体要求和监测流程，并规定了资料收集与整理、数据获取与预处理、土壤样品采集与有机质测定、模型构建与模型应用、精度验证、成果整理与资料归档等内容。

本文件适用于空间分辨率不低于30m、包含可见光至短波红外波段范围高光谱卫星遥感数据的耕地土壤有机质监测。

2 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 30115—2013 卫星遥感影像植被指数产品规范

GB/T 41475 1: 25 000~1: 500 000 土壤养分图用色与图例规范

CH/T 3019—2018 1: 25 000 1: 50 000 光学遥感测绘卫星影像产品生产技术规范

NY/T 1121.1—2006 土壤检测 第1部分：土壤样品的采集、处理和贮存

NY/T 1121.6—2006 土壤检测 第6部分：土壤有机质的测定

3 术语和定义

下列术语和定义适用于本文件。

3.1

高光谱卫星遥感 hyperspectral satellite remote sensing

在电磁波谱的可见光、近红外等波段范围内，以人造地球卫星作为遥感平台，利用高光谱分辨率影像进行地物目标分析的技术。

3.2

耕地 cultivated land

种植农作物的土地。

3.3

土壤有机质 soil organic matter

土壤中形成的和外加入的所有动植物残体不同阶段的各种分解产物和合成产物的总称，包括高度腐解的腐殖物质、解剖结构尚可辨认的有机残体和各种微生物体。

[来源：GB/T 33469—2016, 3.9]

4 缩略语

下列缩略语适用于本文件。

CGCS2000：2000国家大地坐标系(China Geodetic Coordinate System 2000)

CST：中国国家标准时间(China Standard Time)

5 总体要求

5.1 空间参考

5.1.1 坐标系应采用 CGCS2000。必要时，可采用经批准的其他坐标系。

5.1.2 高程基准宜采用 1985 国家高程基准。若无法实现，可采用数据源的高程基准或已公开的其他高

程基准。

5.2 时间系统

采用CST。

5.3 监测时间

应根据农作物物候期、当地农事等因素，选择最佳的遥感监测时间。一般应选择耕地翻耕后实施水肥措施前或作物收割后，土壤裸露且状态较为稳定的时间段。

6 总体流程

高光谱卫星遥感监测土壤有机质的流程主要包括资料收集与整理、遥感数据获取与预处理、土壤样品采集与有机质测定、模型构建与模型应用、成果整理等，高光谱卫星遥感监测土壤有机质技术流程见图1。

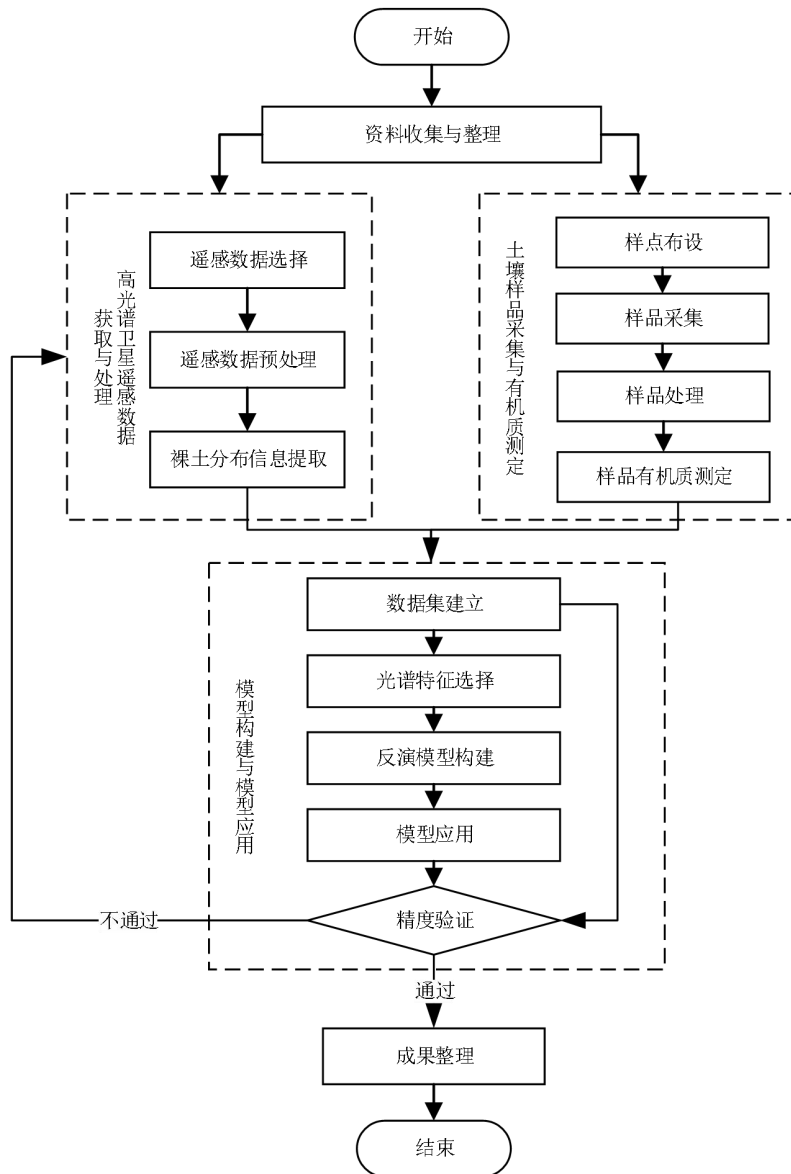


图1 高光谱卫星遥感监测土壤有机质技术流程

7 资料收集与整理

7.1 资料收集

资料收集时间上宜从新到老，比例尺应从大到小。应收集以下资料：

- a) 土壤类型分布图、土壤类型数据库、土壤类型调查数据等土壤类型资料；
- b) 通过遥感解译、野外调查等方式获取的土地利用、地表覆盖或耕地分布资料；
- c) 区域内长期定位观测站点的土壤调查数据及图件等土壤参量分布资料；
- d) 数字高程模型（DEM）数据；
- e) 行政区划数据。

7.2 资料整理

资料整理要求如下：

- a) 查看收集资料的数据时相、数据来源、空间参考、比例尺、成果精度等，比较各类资料的地理分布特性及监测区覆盖情况，评估资料的准确性和可靠性；
- b) 纸介质图件资料应转换成栅格数据或矢量数据，具有不同量纲的数据应进行归一化处理，各种资料应配准到统一的坐标系上。如有必要应对栅格数据进行重采样，以匹配遥感数据分辨率；
- c) 对收集的资料进行分析，明确区域内土壤类型、土地利用/土地覆盖现状、耕地分布情况以及土壤质地、肥力状况，确定土壤有机质的空间分异情况。

8 数据获取与预处理

8.1 遥感数据选择

遥感数据选择应符合下列要求：

- a) 遥感数据具有可见光波段、近红外波段、短波红外波段；
- b) 遥感数据的空间分辨率应不低于30m，光谱分辨率为（5-20）nm；
- c) 遥感数据云覆盖面积总和应不超过影像覆盖面积的15%；
- d) 遥感数据全部或部分覆盖监测区内耕地；
- e) 遥感数据图面应清晰，无数据丢失，无明显条纹、点状或块状噪声，无严重辐射和几何畸变。

8.2 遥感数据预处理方法

8.2.1 辐射定标

对选择的遥感数据应进行辐射定标，将量化数字灰度值转换为对应视场表观辐亮度值，得到辐亮度数据。辐射定标按照GB/T 30115-2013中7.3的规定执行。

8.2.2 大气校正

为消除或减弱遥感数据获取时因大气传输中的吸收或散射作用引起的辐射畸变，应对辐射定标得到的辐亮度数据进行大气校正，得到地表反射率数据。大气校正按照GB/T 30115-2013中7.7的规定执行。

8.2.3 几何纠正

应对地表反射率数据进行几何纠正以消除几何畸变带来的误差。几何纠正包括几何精纠正与正射纠正，几何精纠正应按照CH/T 3019—2018中8.5的规定执行，正射纠正应按照CH/T 3019—2018中9.5的规定执行，处理得到几何纠正后的地表反射率数据。

几何精度应符合CH/T 3019—2018中4.5的规定。

8.3 裸土分布信息提取

8.3.1 提取方法

8.3.1.1 裸土指数计算

利用8.2.3几何纠正后的地表反射率数据，按照公式（1）计算裸土指数。

$$BI = \frac{(S1+R)-(N+B)}{(S1+R)+(N+B)} \quad (1)$$

式中：

- BI*——裸土指数；
- SI*——短波红外波段（一般为1500nm~1700nm或2100nm~2300nm）；
- R*——红波段（600nm~700nm）；
- N*——近红外波段（700nm~1000nm）；
- B*——蓝波段（400nm~500nm）。

8.3.1.2 裸土像元提取

按照下列步骤提取裸土像元：

- a) 统计耕地范围内像元在裸土指数中的数值，并绘制频率分布直方图；
- b) 确定裸土像元与非裸土像元的划分阈值，宜采用最大类间方差法确定阈值，具体方法见附录A；
- c) 将裸土指数中大于阈值的像元判定为裸土像元，赋值为1，其他像元（包括非耕地像元）判定为非裸土像元，赋值为0，生成裸土提取结果；
- d) 在提取的裸土像元中随机生成不少于100个像元作为验证集，以目视判读结果作为真值，按照公式（2）计算裸土提取精度。

$$P = \frac{TP}{TP+FP} \quad (2)$$

式中：

- P*——裸土提取精度；
- TP*——验证集中目视判读为裸土的像元数；
- FP*——验证集中目视判读为非裸土的像元数。

8.3.2 技术要求

裸土分布信息提取应符合下列要求：

- a) 裸土提取结果应以栅格二值化形式存储；
- b) 裸土提取精度应不低于90%。

9 土壤样品采集与有机质测定

9.1 样点布设

样点布设要求和方法如下：

- a) 样点布设应能反映区域内土壤有机质的空间变异性，应覆盖不同有机质含量等级的地块序列；
- b) 样点布设宜采用历史土壤有机质空间分布数据辅助分层抽样方式，将土壤有机质数值范围划分为（5~8）级，每级应布设不少于5个样点，如果监测区域无历史土壤有机质数据，应结合土壤类型和/或土地利用信息分层布设土壤样点；
- c) 样点应布设在地块中心位置，样点周边（3×3）个像元范围内的土壤类型、耕地利用类型、农田管理措施应一致；
- d) 在远郊区，样点布设应选择代表区域内主要土壤类型、土地利用类型和地理地貌的地块；
- e) 在近郊区，样点布设应选择种植面积相对较大的粮食作物、蔬菜基地以及特色农产品种植的代表性地块；
- f) 样点应远离城市垃圾堆放点、工业及生活排污口、交通、餐饮、住宅、沟渠、粪坑、坟墓等污染源影响显著的地区300m以上，并避免在水土流失严重或表土被破坏处布设样点。

9.2 样品采集

9.2.1 样品采集要求

样品采集要求如下：

- a) 样品采集前，应准备好采样工具、工作底图和采样记录表等，样品采样记录表示例见附录B；
- b) 样品采集时间宜在高光谱遥感数据拍摄日期前后10日内完成，样品采集与遥感数据拍摄时土壤表面状况未发生明显变化；
- c) 每个样点的土壤样品应为混合样，即样点所在一定半径范围内若干子样经均匀混合后的土壤样品，混合样的半径范围应小于高光谱卫星遥感数据空间分辨率的一半，每个混合样由（3~5）个子样组成；

- d) 子样宜采用“X”形布点法、“S”形布点法、梅花形布点法或棋盘形布点法，每个子样采集深度为（5~20）cm耕作层土壤，采土部位应一致；
- e) 土壤样品应装入塑料密封袋内并外套布袋，写好便签、做好采样记录，标签与采样记录应按照NY/T 1121.1—2006中2.2 f)的规定执行。

9.2.2 样品采集方法

样品采集方法应按照NY/T 1121.1—2006中2.2 c)~e)的规定执行。

9.3 样品处理

样品处理应按照NY/T 1121.1—2006中3.2的规定执行。

9.4 样品有机质测定

样品有机质测定应按照NY/T 1121.6—2006的规定执行。

10 模型构建与模型应用

10.1 模型构建

10.1.1 数据集建立

应根据土壤样品有机质含量数值分布情况，采用分层随机方法划分训练集与验证集，训练集与验证集的样品比例应在2:1至3:1的范围内。训练集用于光谱特征选择与反演模型构建，验证集用于精度验证。

10.1.2 光谱特征选择

光谱特征选择包括光谱特征计算和土壤有机质敏感特征筛选：

- a) 利用8.2.3几何纠正后的地表反射率数据，根据训练集中样品坐标信息提取样品对应的像元反射率光谱，计算光谱特征。常用的光谱特征包括光谱斜率、光谱吸收位置、光谱吸收深度、光谱吸收宽度等参量化光谱特征，以及经过数学变换后的光谱，如倒数光谱、对数光谱、导数光谱、积分光谱等，常用的光谱特征计算方法见附录C；
- b) 土壤有机质敏感特征筛选宜采用基于相关分析的特征选择方法，按照公式（3）计算训练集中样品的各光谱特征与有机质含量间的皮尔逊相关系数 ρ_j ， ρ_j 大于0.4可通过筛选。

$$\rho_j = \frac{\sum_{i=1}^m (x_{ij} - \bar{x}_j)(y_i - \bar{y})}{\sqrt{\sum_{i=1}^m (x_{ij} - \bar{x}_j)^2 \sum_{i=1}^m (y_i - \bar{y})^2}} \quad (3)$$

式中：

- ρ_j ——第j个光谱特征与有机质含量间的皮尔逊相关系数；
- x_{ij} ——第i个样品的第j个光谱特征的值；
- \bar{x}_j ——所有样品第j个光谱特征的平均值；
- y_i ——第i个样品的有机质含量值；
- \bar{y} ——所有样品有机质含量的平均值；
- m ——样品的数量。

10.1.3 反演模型构建

根据10.1.1划定的训练集样品数据，以10.1.2筛选的土壤有机质敏感特征为自变量、以有机质含量为因变量，作为模型的输入参数训练模型。宜采用偏最小二乘回归、随机森林回归和高斯过程回归等统计预测模型进行反演模型构建。常用的统计预测模型见附录D。

10.2 模型应用

监测区土壤有机质含量分布结果估算方法如下：

- a) 利用8.2.3几何纠正后的整个监测区地表反射率数据，根据8.3提取的裸土像元，计算10.1.2中筛选的土壤有机质敏感特征，计算方法见附录C；
- b) 应用10.1.3构建的模型进行估算，得到监测区土壤有机质含量分布结果。

10.3 精度验证

10.3.1 验证方法

根据10.2估算的土壤有机质含量分布结果，采用以下两种方式进行精度验证。

a) 利用验证集样品数据，按照公式（4）计算皮尔逊相关系数；

$$\rho = \frac{\text{cov}(Y, Y')}{\sigma_y \sigma_{y'}} \quad (4)$$

式中：

ρ ——皮尔逊相关系数；

Y ——验证集土壤样品有机质的实测值；

Y' ——验证集土壤样品有机质的遥感估算值；

$\text{cov}(Y, Y')$ —— Y 与 Y' 的协方差；

σ_y —— Y 的标准差；

$\sigma_{y'}$ —— Y' 的标准差。

b) 利用验证集样品数据，根据公式（5）计算随机均方根误差。

$$r = \sqrt{\frac{1}{n-1} \times \sum_{i=1}^n (Y_i - Y'_i)^2} \quad (5)$$

式中：

r ——随机均方根误差；

n ——验证集样品数量；

i ——验证集中第 i 个土壤样品；

Y_i ——验证集中第 i 个土壤样品有机质的实测值；

Y'_i ——验证集中第 i 个土壤样品有机质的遥感估算值。

10.3.2 合格判定

土壤有机质估算结果精度合格判定如下：

a) 根据公式（4）和公式（5）的计算结果，如果 $\rho \geq 0.6$ 且 $r \leq 10 \text{ g/kg}$ ，则判定土壤有机质估算结果精度合格，否则为不合格；

b) 精度判定为不合格，应逐个检查并优化数据获取与预处理、土壤样品采集与测试分析、模型构建、模型应用中各个环节，直至精度合格。

11 成果整理与资料归档

11.1 成果整理

11.1.1 监测成果图

监测成果图制作要求如下。

a) 监测成果图坐标系应采用CGCS2000；

b) 监测成果图应包括图名、图例、比例尺、指北针、监测时间、制图单位等，监测成果图示例见附录E；

c) 应利用土地利用数据对监测成果图中非耕地区域进行掩膜；

d) 应根据8.3中裸土提取结果对非裸土区域进行掩膜；

e) 对耕地土壤有机质反演结果进行彩色渲染或分级设色，用色与图例按照GB/T 41475的规定执行。

11.1.2 监测成果报告

监测成果报告宜包括但不限于如下内容：

a) 任务来源、工作目标和具体指标要求等资料与信息；

b) 区域内土壤类型、土地利用/土地覆盖现状、耕地分布情况，以及土壤有机质、质地、肥力等土壤指标的空间分布情况；

c) 高光谱卫星遥感数据获取情况与处理结果；

d) 土壤样品采集与测试分析结果；

e) 土壤有机质光谱特征选择与反演模型构建；

f) 区域内土壤有机质遥感监测结果，宜采用统计表格和图片等，统计表格应包括土壤有机质含量、监测精度评价等信息。图片应包括样点实际分布图、样点处实景照片、监测成果图等；

g) 土壤有机质遥感监测结果精度。

11.2 资料归档

归档介质为光盘、磁带或硬盘等，归档资料包括：

- a) 监测成果图；
- b) 监测成果报告；
- c) 其他。

附录 A
(资料性)
最大类间方差法

A.1 概述

最大类间方差法是一种确定图像二值化分割阈值的优化算法，将图像的像元值分为两类，使两类之间的方差最大，同一类内的方差最小，通过计算类间方差，可以找到一个最优的阈值，以最大化两个类别之间的差异。

A.2 计算直方图

统计图像的灰度直方图，显示每个灰度级在图像中出现的频率。

A.3 灰度级划分

假设图像的灰度级范围是 0 到 $L-1$ （例如，对于 8 位图像， $L=256$ ）。对于每一个可能的阈值 T ($0 \leq T \leq L-1$)，可以将图像分为两类：

- a) 类 0：包含所有灰度级小于 T 的像元；
- b) 类 1：包含所有灰度级大于或等于 T 的像元。

A.4 计算权重和均值

对于每个阈值 T ，计算两个类的权重（即像元数量占总数量的比例）和均值：

- a) 类 0 的权重 $w_0(T)$ 是灰度级小于 T 的像元数量除以总像元数量；
- b) 类 1 的权重 $w_1(T)$ 是灰度级大于或等于 T 的像元数量除以总像元数量；
- c) 类 0 的均值 $\mu_0(T)$ 是类 0 中所有像元灰度值的和除以类 0 的像元数量；
- d) 类 1 的均值 $\mu_1(T)$ 是类 1 中所有像元灰度值的和除以类 1 的像元数量；
- e) 全局均值 μ 是所有像元灰度值的和除以总像元数量。

A.5 计算类间方差

对于每个阈值 T ，计算类间方差 $\sigma_B^2(T)$ ，计算方法见公式 (A.1)。

$$\sigma_B^2(T) = w_0(T) \times [\mu_0(T) - \mu]^2 + w_1(T) \times [\mu_1(T) - \mu]^2 \quad (\text{A.1})$$

式中：

- $\sigma_B^2(T)$ ——类间方差；
- $w_0(T)$ ——类 0 的权重；
- $\mu_0(T)$ ——类 0 的均值；
- μ ——全局均值；
- $w_1(T)$ ——类 1 的权重；
- $\mu_1(T)$ ——类 1 的均值。

A.6 选择最优阈值

遍历所有可能的阈值 T ，找到使类间方差 $\sigma_B^2(T)$ 最大的那个阈值，即最优的分割阈值。

A.7 应用阈值

应用确定的最优阈值将图像二值化。所有小于阈值的像元被设置为 0（或某个其他指定的值），而所有大于或等于阈值的像元被设置为 255（或其他指定的值）。

附录 B
(资料性)
样品采样记录表示例

样品采样记录表示例见表 A.1。

表 A.1 样品采样记录表示例

序号	记录项	内容示例
1	样点编号	JL-DAN-001
2	位置	吉林省大安市
3	采集日期	2023.04.27
4	经度(°)	123.389626E
5	纬度(°)	45.406029N
6	海拔(m)	129.9
7	地形	平原下坡底部
8	照片编号	IMG 001-008
9	土地利用类型	耕地
10	耕地利用类型	水田
11	作物或秸秆类型	水稻秸秆
12	作物或秸秆覆盖比例	小于5%
13	土壤颜色	黄棕色
14	土壤湿度	较干
15	地表侵蚀状况	轻度风蚀
16	地表盐碱状况	少量盐碱斑

附录 C
(资料性)
常用的光谱特征计算

C.1 参量化光谱特征

C.1.1 光谱斜率

在某一个波长区间, 连接光谱曲线起始波段与终止波段的光谱反射率形成一条线段, 该线段的斜率即为光谱斜率, 计算方法见公式 (C.1)。

$$S = \frac{R(\lambda_2) - R(\lambda_1)}{\lambda_2 - \lambda_1} \quad (\text{C.1})$$

式中:

- S ——待求波长区间的光谱斜率;
- λ_1 ——起始波段波长;
- λ_2 ——终止波段波长;
- $R(\lambda_1)$ ——起始波段的光谱反射率;
- $R(\lambda_2)$ ——终止波段的光谱反射率。

C.1.2 光谱吸收位置

在光谱吸收谷中, 反射率最低处的波长被定义为光谱吸收位置, 计算方法见公式 (C.2)。

$$\lambda_a = \operatorname{argmin}_\lambda R(\lambda) \quad (\text{C.2})$$

式中:

- λ_a ——光谱吸收位置;
- $R(\lambda)$ ——波长为 λ 处的反射率;
- $\operatorname{argmin}_\lambda$ ——使 $R(\lambda)$ 最小的波长。

C.1.3 光谱吸收深度

在某一光谱吸收范围, 反射率最低点到归一化包络线的距离被定义为光谱吸收深度, 计算方法见公式 (C.3)。

$$AD = 1 - \frac{R(\lambda_a)}{R_{env}} \quad (\text{C.3})$$

式中:

- AD ——光谱吸收深度;
- $R(\lambda_a)$ ——光谱吸收位置 λ_a 处的反射率;
- R_{env} ——光谱吸收位置 λ_a 处归一化包络线上的反射率。

C.1.4 光谱吸收宽度

在光谱吸收谷中, 最大吸收深度一半处的光谱带宽FWHM (Full Width at Half the Maximum Depth) 被定义为光谱吸收宽度, 计算方法见公式 (C.4)。

$$AW = \lambda_2 - \lambda_1 \quad (\text{C.4})$$

式中:

- AW ——光谱吸收宽度;
- λ_1 、 λ_2 ——分别为达到最大吸收深度一半处的两个波长, 其中 $\lambda_2 > \lambda_1$ 。

C.2 光谱数学变换

C.2.1 倒数光谱

倒数光谱指光谱曲线在某一波段处的倒数值。

C.2.2 对数光谱

对数光谱指光谱曲线在某一波段处以10或自然常数为底的对数值。

C.2.3 导数光谱

导数光谱也叫微分光谱，指光谱曲线在某一波段处的导数值。第*i*个波段的一阶导数光谱计算方法见公式（C.5）。

$$R'(\lambda_i) = \frac{R(\lambda_{i+1}) - R(\lambda_{i-1})}{\lambda_{i+1} - \lambda_{i-1}} \dots \dots \dots (C.5)$$

式中：

$R'(\lambda_i)$ ——第*i*个波段的光谱导数；

λ_i ——第*i*个波段的波长；

$R(\lambda_{i+1})$ ——第*i+1*个波段的光谱反射率；

$R(\lambda_{i-1})$ ——第*i-1*个波段的光谱反射率；

λ_{i+1} ——第*i+1*个波段的波长；

λ_{i-1} ——第*i-1*个波段的波长。

C.2.4 积分光谱

积分光谱指光谱曲线在某一波长范围内的下覆面积，计算方法见公式（C.6）。

$$\phi = \int_{\lambda_1}^{\lambda_2} f(\lambda) d\lambda \dots \dots \dots (C.6)$$

式中：

ϕ ——积分光谱；

$f(\lambda)$ ——待求积分的光谱曲线；

λ_1 ——积分的起始波段波长；

λ_2 ——积分的终止波段波长。

附录 D
(资料性)
常用的统计预测模型

D.1 偏最小二乘回归

偏最小二乘回归 (Partial Least Squares Regression, PLSR) 是一种用于解决多元线性回归中自变量间高度相关的问题的统计建模方法。其基本思想是, 将原始自变量 X 和因变量 Y 分别转换为一组新的综合变量 (称为主成分或潜在变量), 然后再使用这些主成分进行回归分析, 从而减少自变量之间的共线性, 并且提高模型的稳定性和预测能力。

假设有一个新的自变量向量 x^* , 预测其对应的因变量 y^* 主要包含以下步骤:

- a) 对 x^* 进行标准化处理, 使其均值为 0, 方差为 1;
- b) 按照公式 (D.1) 计算的 x^* 主成分 t_i^* :

$$t_i^* = x^* w_i \quad (\text{D.1})$$

式中:

t_i^* —— x^* 的第 i 个主成分;

x^* —— 新的自变量向量;

w_i —— X 的第 i 个主成分的权重向量。

- c) 按照公式 (D.2) 计算 y^* 的预测值 \hat{y}^* 。

$$\hat{y}^* = \sum_{i=1}^k b_i t_i^* c_i \quad (\text{D.2})$$

式中:

\hat{y}^* —— 因变量 y^* 的预测值;

k —— 主成分的数量;

b_i —— X 与 Y 的第 i 个主成分间相关系数;

c_i —— Y 的第 i 个主成分的权重向量。

D.2 随机森林回归

随机森林回归 (Random Forest Regression) 是一种基于决策树的集成学习算法, 通过构建多个决策树并对其预测结果进行平均或投票来提高预测精度和稳定性。最终预测结果的计算方法见公式 (D.3):

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N \hat{y}_i \quad (\text{D.3})$$

式中:

\hat{y} —— 随机森林模型的最终预测结果;

N —— 树的数量;

\hat{y}_i —— 第 i 棵树的预测结果。

D.3 高斯过程回归

高斯过程回归 (Gaussian Process Regression, GPR) 是一种基于贝叶斯非参数统计学的回归分析方法。其基本原理是利用高斯过程作为先验知识, 对输入数据之间的相似性进行建模, 并通过高斯分布的均值和方差进行估计, 得到输出值的概率分布。最终预测结果的计算方法见公式 (D.4):

$$\hat{y}^* = K(X^*, X)[K(X, X) + \sigma_n^2 I]^{-1} y \quad (\text{D.4})$$

式中,

K —— 核函数, 是高斯过程回归的核心, 用于计算不同点之间的相似度或协方差。常用的核函数包括径向基函数、线性核函数等;

X, y —— 分别表示训练数据集的输入和输出;

X^*, \hat{y}^* —— 分别表示测试数据集的输入和预测输出;

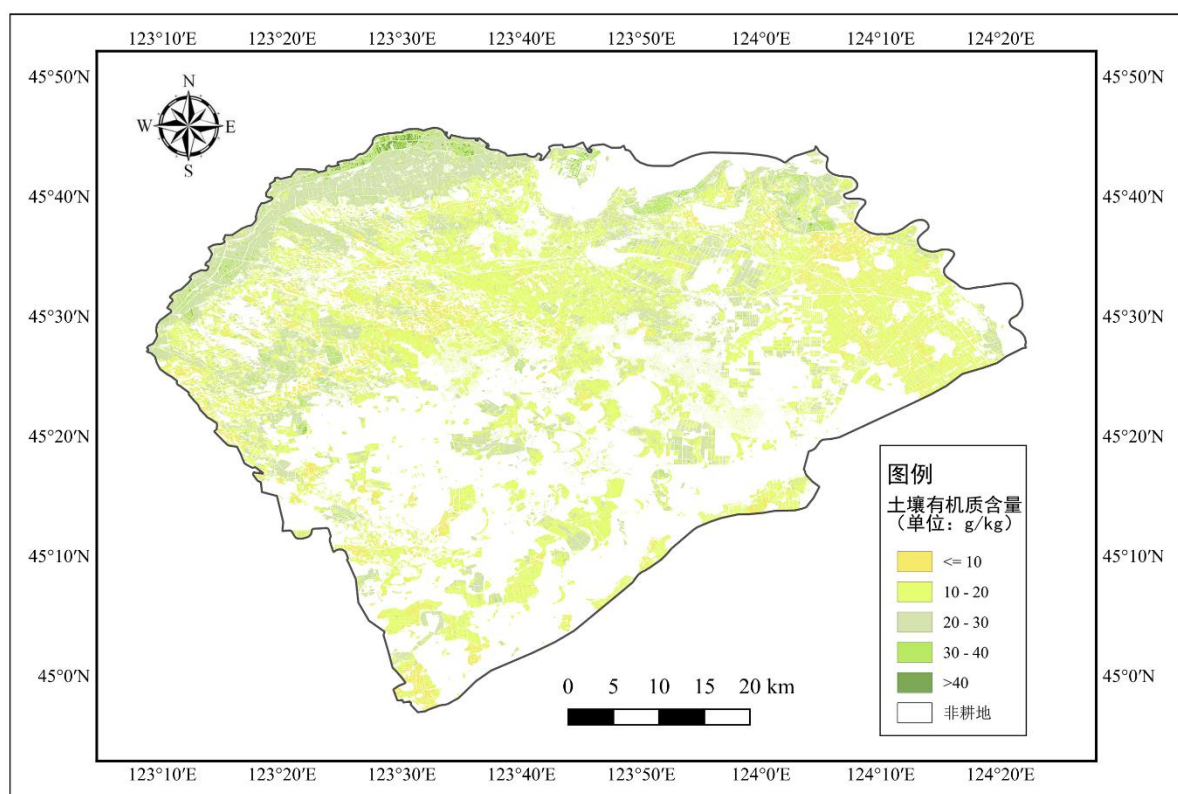
σ_n^2 —— 观测噪声的方差, 表示数据中的噪声水平;

I —— 单位矩阵, 用于确保数学表达式的完整性。

附录 E
(资料性)
监测成果图示例

监测成果图示例见图E. 1。

吉林省大安市耕地土壤有机质高光谱卫星监测成果图



监测时间: 2021年4月
制作人员: XXX

制图单位: 自然资源部国土卫星遥感应用中心
审核: XXX

图 E. 1 监测成果图示例

参 考 文 献

- [1] GB/T 14950—2009 摄影测量与遥感术语
 - [2] GB/T 17296—2009 中国土壤分类与代码
 - [3] GB/T 18834—2002 土壤质量 词汇
 - [4] GB/T 21010—2017 土地利用现状分类
 - [5] GB/T 30115—2013 卫星遥感影像植被指数产品规范
 - [6] GB/T 32722—2016 土壤质量 土壤样品长期和短期保存指南
 - [7] GB/T 32726—2016 土壤质量 野外土壤描述
 - [8] GB/T 33469—2016 耕地质量等级
 - [9] GB 35650—2017 国家基本比例尺地图测绘基本技术规定
 - [10] GB/T 36197—2018 土壤质量 土壤采样技术指南
 - [11] GB/T 36199—2018 土壤质量 土壤采样程序设计指南
 - [12] GB/T 36393—2018 土壤质量 自然、近自然及耕作土壤调查程序指南
 - [13] NY/T 1119—2006 土壤监测规程
 - [14] NY/T 1121.1—2006 土壤检测 第1部分：土壤样品的采集、处理和贮存
 - [15] NY/T 1121.6—2006 土壤检测 第6部分：土壤有机质的测定
 - [16] 童庆禧，张兵，郑兰芬. 高光谱遥感[M]. 北京：高等教育出版社，2006
 - [17] 城乡规划学名词审定委员会. 城乡规划学名词[M]. 北京：科学出版社，2021
-

《高光谱卫星遥感监测土壤有机质技术规范》

编制说明

行业标准项目名称： 高光谱卫星遥感监测土壤有机质技术规范

行业标准项目编号： 202233004

送审行业标准名称： _____

（此栏送审时填写）

报批行业标准名称： _____

（此栏报批时填写）

承担单位： 自然资源部国土卫星遥感应用中心

当前阶段： 征求意见 送审稿审查 报批稿报批

编制时间： 二〇二四年一月

高光谱卫星遥感监测土壤有机质技术规范 编制说明

一、概况

1.1 任务来源

2022年9月6日，自然资源部下达了《自然资源部办公厅关于印发2022年度自然资源标准制修订工作计划的通知》（自然资办发〔2022〕39号），本标准是自然资源部发布的2022年自然资源卫星应用行业标准计划项目之一，项目编号：202233004，标准计划名称《高光谱卫星遥感监测土壤有机质技术规范》。本标准由全国地理信息标准化技术委员会卫星应用分技术委员会归口，由自然资源部国土卫星遥感应用中心牵头起草。计划周期：24个月。

1.2 目的意义

随着自然资源高光谱业务卫星的陆续发射，卫星应用技术与处理系统的不断发展，高光谱卫星遥感已逐渐成为自然资源要素质量监测的重要手段。近年来，党和国家高度重视粮食安全，而耕地作为粮食生产的依托，其质量监测更是自然资源监测业务的重中之重。2021年的中央一号文件更是将耕地保护列为保障国家粮食安全的两个“要害”之一。

土壤有机质一直是耕地质量调查评价的关键指标，也是高光谱卫星土壤参量反演应用研究的热点，相关技术较为成熟。但由于涉及高光谱卫星数据预处理、野外采样与样品测试化验、反演模型构建与模型参数优化等诸多环节，各行业、各区域、各单位以分散作业为主，并未采用统一的、规范化的技术，为监测成果的横向对比与集成带来了很大难度。

十三五以来，在耕地质量保护相关规划文件中，已多次提出要开展“遥感

动态监测”，但由于缺少统一的技术规范，高光谱卫星数据土壤质量监测方面的优势并未完全发挥。随着第三次全国土地调查（“国土三调”）、第三次全国土壤普查（“土壤三普”）等工作的陆续开展，迫切需要制定统一的高光谱卫星遥感监测土壤有机质技术规范，对包括高光谱卫星数据预处理、野外土壤样品采集，到土壤有机质反演模型构建与应用，以及监测成果质量控制、成果整理等全链路的技术方法进行规范，促进高光谱卫星数据在国土三调、土壤三普以及全国耕地质量等别调查与评定等业务中的统一化和规范化应用。

1.3 主要起草人及工作分工

编制任务下达后，自然资源部国土卫星遥感应用中心牵头，由自然资源部国土整治中心、中国科学院南京土壤研究所、中国科学院空天信息创新研究院、中国农业科学院农业资源与农业区划研究所等作为参编单位共同完成。编制组成员包括总体技术负责人和长期从事高光谱卫星土壤质量监测等专业领域的专业技术人员和专家分工合作开展标准各章节的编写，编制组主要人员组成及分工见表 1。

表 1 编制组人员分工

序号	姓名	单位	任务分工	备注
1	尚坤	自然资源部国土卫星遥感应用中心	项目负责人，负责编制标准草案、征求意见稿、送审稿、报批稿等初稿的编制，组织专家评审	
2	肖晨超	自然资源部国土卫星遥感应用中心	负责卫星遥感数据、模型应用与精度评价、成果整理与质量控制相关内容	
3	李少帅	自然资源部国土整治中心	负责土壤样品、模型应用与精度评价、成果整理与质量控制相关内容	
4	王昌昆	中国科学院南京土壤研究所	负责土壤样品、光谱测量、反演模型构建、模型应用与精度评价等相关内容	
5	张霞	中国科学院空天信息创新研究院	负责高光谱数据处理、反演模型构建、模型应用与精度评价相关内容	
6	刘佳	中国农业科学院农业资源与农业区划研究所	负责卫星遥感数据、模型应用与精度评价、成果整理与质量控制等相关内容	

1.4 主要工作过程

1.4.1 征求意见稿阶段

2022年6月-2022年10月，编制组开展了大量的调研工作，包括国内外有关现有标准，以及高光谱卫星土壤有机质含量监测业务的实际实施情况，编制组开始起草标准草案。

2022年11月-2024年1月，以标准草案为基础，编制组又以电话、社交软件、电子邮件和视频会议的形式与高光谱卫星土壤质量监测领域生产作业单位、大学、科研院所的多位技术专家和生产专家进行多次交流探讨，并根据专家意见对标准草案进行修改完善，于2024年1月完成了标准征求意见稿和编制说明。

二、标准编制原则和确定标准主要内容的依据

2.1 标准编制原则

(1) 全面性

高光谱卫星遥感监测土壤有机质是自然资源卫星土壤质量监测中的一项重要工作与技术，遵循测绘标准体系，确定本标准的定位、内容以及与其它标准的关系。标准编制过程中充分考虑到行业领域技术发展现状和实际工作需求，依托已有的技术优势和前期研究基础，不断总结经验，凝练高光谱卫星遥感监测土壤有机质技术中规范流程。旨在保证标准内容规范、对实际工作有指导作用，为形成统一而全面的行业技术规范，指导后续作业生产，提高生产效率，并为未来技术留有发展空间。

(2) 适用性

本标准与现有的国标行标等相关标准相互协调，保持标准内容间的一致性，避免新制定标准同已经颁布实施或正在报批的相关标准之间的冲突和矛盾。本

标准规定了高光谱卫星遥感监测土壤有机质的程序、方法、监测结果的精度评价、成果表达等内容，对完善高光谱卫星应用系列标准体系起到积极作用。

（3）可操作性

高光谱卫星遥感监测土壤有机质涉及较多新技术，实现的技术路线也差异较大，因此需要明确最基本、最普遍适用的技术指标和要求以进行规范和约定，使之既可控制整个高光谱卫星遥感监测土壤有机质全过程的质量，又可以充分发挥不同技术、软件的特点。该技术规范明确清晰地介绍了各流程，对于技术人员而言具有较强的可操作性。

（4）先进性

作为指导和规范生产作业的技术标准，应具有先进性与可继承性。国内区域或旧的指标和要求有的必须调整以适应新技术的发展和境外作业的特点，有的必须保留和继承以保持测绘生产的延续性以及测绘成果的可靠性。

2.2 国内外调研情况

国外在高光谱卫星遥感监测土壤有机质方面开展了较多的研究工作，在算法模型方面，以运用偏最小二乘回归模型得到土壤参量为主。Reis 等基于高光谱数据，通过收集 8 个不同土壤深度的样品，采用主成分分析和线性判别分析方法对不同深度光谱曲线进行聚类，结合偏最小二乘模型预测土壤有机质含量， R^2 达到 0.75，RPD 达到 2.1，表明高光谱数据可用于获取不同深度的土壤有机质含量。Castaldi 等比较了 7 种多光谱或高光谱卫星数据估算土壤有机质的能力，根据不同卫星光谱响应函数获得模拟反射率，分析重采样间隔对精度的影响，验证了高光谱卫星对土壤有机质估算的潜力。

随着我国高光谱卫星的接连发射，国内学者针对国产高光谱卫星土壤有机质含量反演也开展了大量研究。杨雅煜等利用 ZY1-02D 卫星高光谱数据，通过

建立土壤有机质光谱指数，实现长三角地区宿迁市的土壤有机质含量反演。孙伟超等基于高分五号高光谱卫星数据，采用遗传算法和特征投影重要性综合打分对光谱谱段进行特征选择，利用有机质与黏土矿物组合光谱带估算土壤有机质含量，与利用全谱段数据有机质含量相比， R^2 从0.34提高到0.76，模型精度显著提高。孟祥添等人利用运用不同去噪方法的GF-5卫星高光谱数据预测土壤有机质含量，充分挖掘了卫星高光谱数据在监测土壤有机质方面的潜力。颜祥照基于国产星载高分五号高光谱卫星数据估算耕地土壤有机质含量，利用光谱反射率数据构建偏最小二乘回归模型，同时评估了不同土壤类型对土壤有机质预测精度的影响，应用于黑土区后模型验证精度R达到0.8，在沼泽土内模型验证精度R达到0.6。

目前国内外在高光谱卫星遥感监测土壤有机质方面已经取得了显著的进展。研究结果充分论证了利用高光谱卫星监测土壤有机质的可行性，并显示高光谱卫星的监测效果明显优于多光谱数据。随着高光谱卫星技术的不断成熟，越来越多的高光谱卫星数据产品将投入业务化应用。但是，高光谱卫星遥感监测土壤有机质涉及多项关键技术，在数据处理及技术流程上，仍未形成统一的指标，不利于标准化的产品生产与批量应用，无法应对未来国产高光谱数据大规模、业务化、规范化处理，可能会滞后高光谱数据的应用与推广，影响我国卫星产业的发展。因此，迫切需要开展高光谱卫星遥感监测土壤有机质规范研究，推动我国高光谱卫星业务化运行。

本标准在制定过程中，广泛收集了相关国家标准和行业标准，作为本标准制定的参考和借鉴依据，主要包括：

GB/T 14950-2009 摄影测量与遥感术语

GB/T 17296-2009 中国土壤分类与代码

GB/T 18834-2002 土壤质量 词汇

GB/T 21010-2017 土地利用现状分类

GB/T 30115-2013 卫星遥感影像植被指数产品规范

GB/T 32722-2016 土壤质量 土壤样品长期和短期保存指南

GB/T 32726-2016 土壤质量 野外土壤描述

GB/T 33469-2016 耕地质量等级

GB/T 36197-2018 土壤质量 土壤采样技术指南

GB/T 36199-2018 土壤质量 土壤采样程序设计指南

GB/T 36393-2018 土壤质量 自然、近自然及耕作土壤调查程序指南

GB/T 41475—2022 1: 25 000~1: 500 000 土壤养分图用色与图例规范

GB/T 42363-2023 土壤质量 土壤理化分析样品的预处理

CH/T 3019-2018 1: 2 5000 1: 500 000 光学遥感测绘卫星影像产品生产技术规范

NY/T 1119-2006 土壤监测规程

NY/T 1121.1-2006 土壤检测 第1部分：土壤样品的采集、处理和贮存

NY/T 1121.6-2006 土壤检测 第6部分：土壤有机质的测定

本标准充分借鉴以上技术标准，但是具有明显的差别。GB/T 14950-2009 是基本术语要求，是很多测绘、遥感标准的基础，用于解释本标准中的部分专业术语。GB/T 17296-2009 确定了中国土壤分类与代码，用于土壤类型的划分和标注。GB/T 18834-2002 规定了土壤质量词汇，适用于对土壤保护、土壤监测、土壤治理等方面活动中用的名词术语及定义的有关内容。GB/T 21010-2017 规定了土地利用现状的总则、分类与编码，明确了土地利用各类型的含义。CH/T 3019—2018 用于规定 1: 2 5000 1: 500 000 光学遥感测绘卫星影像产品的生产，

其中对于卫星数据的几何校正处理要求适用于高光谱卫星遥感数据处理，具有一定参考价值。GB/T 30115-2013 明确了卫星遥感影像植被指数产品规范，对于裸土提取中的指数计算具有一定参考价值。GB/T 32722-2016、GB/T 32726-2016、GB/T 36197-2018、GB/T 36199-2018、GB/T 36393-2018、GB/T 41475-2022、GB/T 42363-2023 用于规定和指导土壤样品保存、野外土壤描述、土壤采样、土壤调查、土壤理化分析样品的预处理，主要是针对传统的野外样品采集，其中部分内容适用于高光谱卫星遥感监测土壤有机质，具有一定参考价值。GB/T 33469-2016 用于规定耕地质量等级以及不同区域的划分标准。NY/T 1119-2006、NY/T 1121.1-2006、NY/T 1121.6-2006 用于规定传统采样为主的土壤监测相关技术环节，和土壤检测中土壤样品的采集、处理和贮存以及土壤有机质的测定。

通过分析，可以发现现有标准难以满足高光谱卫星遥感监测土壤有机质的技术要求，需要开展针对性的标准制定工作，以更好地指导相关作业。标准化的试验流程是国产高光谱卫星监测土壤质量走向业务化前提条件，因此，迫切需要尽快填补该领域标准规范的缺失，构建完善的高光谱卫星遥感监测土壤有机质技术规范，为业界提供明确的指导。

2.3 主要技术内容的说明

(1) 标准的定位

本文件规定了高光谱卫星遥感监测土壤有机质的程序、方法、监测结果的精度评价、成果表达等内容。本文件适用于空间分辨率不低于 30 米、包含可见光、近红外与短波红外波段范围高光谱卫星遥感数据的耕地土壤有机质监测。

(2) 确定标准主要内容的依据

本标准的任务是规范并指导资料收集与整理、遥感数据获取与预处理、土壤样品采集与有机质测定、模型构建与模型应用、成果整理等环节，用于进行

基于星载高光谱数据的土壤有机质监测。为提高标准的实用性，编制组在标准制定前期进行了大量的调研、资料收集，以及技术试验等工作，在编制过程中与自然资源部国土整治中心、中国科学院南京土壤研究所、中国科学院空天信息创新研究院、中国农业科学院农业资源与农业区划研究所、中国科学院地理科学与资源研究所、中国科学院东北地理与农业生态研究所等单位的专家进行了多次的交流和探讨。同时充分利用主编单位的有利条件，总结凝练利用高光谱卫星遥感监测黑土地土壤有机质含量的实践经验，开展了大量技术试验。这些都是本标准主要内容编制的重要依据。

（3）已发射的国产高光谱卫星

从2018年起，我国已成功发射高分五号、资源一号02D、高分五号卫星、高分五号B星、高分五号01A星等5颗高光谱观测卫星，高光谱对地观测当前已迈入星载时代。2020和2021年自然资源部作为业主相继发射的资源一号02D、资源一号02E两颗高光谱卫星，组成了全球首个民用高光谱业务卫星星座。该星座不仅具有大幅宽、高光谱分辨率的特点，保持国际同类载荷最高的信噪比，并具备同时刻、同角度、同大气条件下，高分辨率多光谱成像以及大气辐射反演能力，可为我国高光谱应用深化研发与规模化应用持续提供星载数据方面的有力支持和保障。

（4）资料收集与整理

利用高光谱遥感卫星开展土壤有机质监测，首先需要进行资料收集与整理。资料收集在时间上一般宜从新到老，比例尺从大到小。收集的资料应包括土壤类型分布资料、土地利用/地表覆盖/耕地分布资料、土壤有机质等参量分布资料、数字高程模型数据以及行政区划数据等。在资料整理时，需要首先查看资料的数据时相、数据来源、空间参考、比例尺、成果精度等，综合分析各类型

资料的分布以及监测区覆盖情况，评估资料的准确性和可靠性。然后将介质图片资料转换成栅格数据或矢量数据，具有不同量纲的数据应进行归一化处理，各种资料应配准到统一的坐标系上，如有必要应对栅格数据进行重采样，以匹配遥感数据分辨率。最后，对收集的资料进行分析，明确区域内土壤类型、土地利用/土地覆盖现状、耕地分布情况，以及土壤有机质、质地、肥力等土壤指标的空间分布情况，确定土壤有机质的空间分异情况。资料收集与整理一方面用于为样点布设的依据，另一方面可与土壤有机质监测结果进行对照，分析结果的合理性与科学性为样点布设提供依据。

(5) 数据获取与预处理

数据获取与预处理包含遥感数据选择、遥感数据预处理、裸土分布信息提取三个环节。

遥感数据选择时，需要重点考虑谱段范围与设置、空间分辨率是否可满足土壤有机质反演的指标要求，影像的云量覆盖、耕地覆盖以及数据质量是否满足要求等。为此，在遥感数据选择中从上述五个方面分别提出了具体要求：a) 遥感数据具有可见光波段、近红外波段、短波红外波段；b) 遥感数据的空间分辨率应不低于 30m，光谱分辨率为 (5-20) nm；c) 遥感数据云覆盖面积总和应不超过影像覆盖面积的 15%；d) 遥感数据全部或部分覆盖监测区内耕地；e) 遥感数据图面应清晰，无数据丢失，无明显条纹、点状或块状噪声，无严重辐射和几何畸变。

高光谱卫星遥感数据的预处理一般包括辐射定标、大气校正和几何纠正三部分。辐射定标一般是根据传感器各波段定标系数利用定标公式对影像各波段数据进行辐射定标，将量化数字灰度值转换为对应视场表观辐亮度值，得到辐亮度数据。通过对现有标准规范的查阅比较，本文件中确定高光谱卫星遥感监

测土壤有机质中的辐射定标具体计算方法按照 GB/T 30115—2013 中 7.3 的规定执行。为消除或减弱遥感数据获取时因大气传输中的吸收或散射作用引起的辐射畸变，一般应对辐射定标得到的辐亮度数据进行大气校正，得到地表反射率数据。大气校正方法应根据传感器特性、地表覆盖和气象等条件选择适宜的方法，具体按照 GB/T 30115—2013 中 7.7 的规定执行。为消除几何畸变带来的误差，应对地表反射率数据进行几何纠正。一般的几何纠正包括几何精纠正和正射纠正。几何精纠正的方法按照 CH/T 3019—2018 中 8.5 的规定执行，正射纠正的方法按照 CH/T 3019—2018 中 9.5 的规定执行。几何精度应符合 CH/T 3019—2018 中 4.5 的规定。

裸土分布信息提取是高光谱卫星遥感监测土壤有机质中的关键预处理环节。在前面获得几何纠正后的地表反射率基础上，首先按照公式计算裸土指数，然后统计耕地范围内像元在裸土指数中的数值，并绘制频率分布直方图，确定裸土像元与非裸土像元的划分阈值，宜采用最大类间方差法确定阈值，将裸土指数中大于阈值的像元判定为裸土像元，赋值为 1，其他像元（包括非耕地像元）判定为非裸土像元，赋值为 0，生成裸土提取结果。在提取的裸土像元中随机生成不少于 100 个像元作为验证集，以目视判读结果作为真值计算裸土提取精度。为确保后续结果的精度，提取的裸土分布结果一般应以栅格二值化形式存储，且提取精度应不低于 90%。

（6）土壤样品采集与有机质测定

土壤样品采集与有机质测定具体包含四方面的内容：样点布设、样品采集、样品处理、样品有机质测定。

样点布设时应重点考虑样点是否能综合反映区域内土壤有机质的空间变异性、应覆盖不同有机质含量等级的地块序列。其次，样点布设宜采用历史土壤

有机质空间分布数据辅助分层抽样方式，将土壤有机质数值范围划分为（5~8）级，每级应布设不少于 5 个样点，如果监测区域无历史土壤有机质数据，应结合土壤类型和/或土地利用信息分层布设土壤样点。样点应布设于平坦地块中心位置，样点周边（3×3）个像元范围内的土壤类型、耕地利用类型、农田管理措施应一致。在远郊区，样点布设应选择代表区域内主要土壤类型、土地利用类型和地理地貌的地块；在近郊区，样点布设应选择种植面积相对较大的粮食作物、蔬菜基地及特色农产品种植的代表性地块。此外，样点应远离城市垃圾堆放点、工业及生活排污口、交通、餐饮、住宅、沟渠、粪坑、坟墓等污染源影响显著的地区 300m 以上，并避免在水土流失严重或表土被破坏处布设样点。

样品采集前，应准备好采样工具、工作底图和采样记录表等。样品采集时间应综合考虑监测时间需求与卫星数据获取计划确定，一般应在高光谱影像数据拍摄日期前后 10 日内完成，样品采集与影像拍摄时土壤表面理化性状未发生明显变化。样点采集方法按照 NY/T 1121.1—2006 2.2 中 c)-e) 的规定执行。每个样点的土壤样品应为混合样，即采样点所在一定半径范围内若干子样经均匀混合后的土壤样品。混合样的半径范围应小于高光谱卫星影像分辨率的一半。每个混合样由 3-5 个子样组成。子样宜采用“X”形布点法、“S”形布点法、梅花形布点法或棋盘形布点法，每个子样采集深度为 5-20cm 耕作层土壤，采土部位应一致。土壤样品应装入塑料密封袋内并外套布袋，写好便签、做好采样记录。标签与采样记录按照 NY/T 1121.1—2006 2.2 中 f) 的规定执行。

样品的处理按照 NY/T 1121.1—2006 中 3.2 的规定执行，样品有机质测定按照 NY/T 1121.6—2006 的规定执行。

（7）模型构建与模型应用

模型构建与模型应用部分具体包括模型构建、模型应用和精度验证三方面

内容。

在模型构建中，具体包含数据集建立、光谱特征选择、反演模型构建三个环节。数据集的建立应根据土壤样品有机质含量数值分布情况，采用分层随机方法划分训练集与验证集，训练集与验证集样品比例在 2:1 至 3:1 的范围内。训练集用于光谱特征选择与反演模型构建，验证集用于精度验证。光谱特征选择包括光谱特征计算和土壤有机质敏感特征筛选。常用的光谱特征包括光谱斜率、光谱吸收位置、光谱吸收深度、光谱吸收宽度等，以及经过数学变换后的光谱，如倒数光谱、对数光谱、导数光谱、积分光谱等。在土壤有机质敏感特征筛选时，宜采用基于相关分析的特征选择方法，以皮尔逊相关系数的绝对值大于 0.4 作为标准，筛选与土壤有机质含量相关性高的光谱特征组合。根据划定的训练集样品数据，以筛选的土壤有机质敏感特征为自变量、以有机质含量为因变量，作为模型的输入参数训练模型，采用偏最小二乘回归、随机森林回归和高斯过程回归等统计预测模型进行反演模型构建。

在应用模型估算土壤有机质含量分布结果时，利用几何纠正后的地表反射率数据，根据提取的裸土像元，计算筛选的土壤有机质敏感特征，然后应用构建的模型进行估算，得到土壤有机质含量分布结果。

精度验证应根据估算的土壤有机质含量分布结果，采用以下两种方式进行判定：1) 利用验证集样品数据，计算土壤样品有机质实测值 Y 与遥感监测土壤有机质含量 Y' 间的皮尔逊相关系数 ρ ；2) 利用验证集样品数据，计算土壤样品有机质实测值 Y 与遥感监测土壤有机质含量 Y' 间的均方根误差 r 。满足 $\rho \geq 0.6$ 且 $r \leq 10$ g/kg 则判定为精度合格，精度判定不合格的，应逐个检查并优化数据获取与预处理、土壤样品采集与有机质测定、模型构建、模型应用中各个环节，直至精度合格。

（8）成果整理与资料归档

成果整理包括监测成果图与监测成果报告。土壤有机质遥感监测成果图应利用土地利用数据对监测成果图中非耕地区域进行掩膜；对耕地土壤有机质反演结果进行彩色渲染或分级设色，用色与图例按照 GB/T 41475 的规定执行；监测成果图坐标系应采用 CGCS2000；监测成果图应包括图名、图例、比例尺、指北针、监测时间、制图单位等。

监测成果报告方面，宜包括但不限于如下内容：任务来源、工作目标和具体指标要求等资料与信息；区域内土壤类型、土地利用/土地覆盖现状、耕地分布情况，以及土壤有机质、质地、肥力等土壤指标的空间分布情况；高光谱卫星遥感数据获取情况与处理结果；土壤样品采集与有机质测定结果；土壤有机质光谱特征选择与反演模型构建；区域内土壤有机质遥感监测结果，宜采用统计表格和图片等，统计表格应包括土壤有机质含量、监测精度评价等信息。图片应包括样点实际分布图、样点处实景照片、监测成果图等；土壤有机质遥感监测结果精度。

资料归档介质主要为光盘、磁带或硬盘等，归档资料包括：监测成果图、监测成果报告、其他。

三、 验证试验的情况和结果

在验证试验方面，编制组首次利用星载高光谱数据实现了中国黑土地耕地土壤有机质全覆盖制图，包括黑龙江省、吉林省、辽宁省和内蒙古自治区东部的赤峰市、通辽市、兴安盟、呼伦贝尔市，区域总面积约 126.54 万平方公里。监测区域覆盖黑土地全域耕地，具体包含水田、旱地、水浇地。本次监测使用自然资源部作为业主的我国首颗民用高光谱卫星（资源一号 02D）。

在试验中，针对中国耕地复杂地理环境与土壤类型、农业分区特点，研发了一套地形因子与目标光谱特征协同的土壤参量反演技术，首次提出了基于多指标评价的变量重要性因子-竞争性自组织选择-随机蛙三算法联合特征选择技术（VIP-CARS-Frog），实现高质量、稳定的特征选择。该反演技术现已成功应用于 GF5、5 米光学业务卫星、5 米光学 02 星等国产高光谱卫星，利用该技术构建了东北黑土区尺度土壤有机质反演模型。

整个过程分为以下部分：1) 对高光谱数据进行预处理得到地表反射率数据，基于实测点数据提取影像中点位光谱反射率；2) 基于点位光谱计算光谱指数、光谱反射率（包含不同变换后的光谱）；基于高程数据提取坡度、坡向等地形因子；基于实测点数据提取对应点位地形因子；利用光谱反射率、光谱指数、地形因子构建特征集；3) 按照一定比例将土壤有机质和特征集划分为建模集和验证集；4) 基于随机蛙算法（Random Frog, RF）对不同特征集组合进行降维，构建高斯过程回归模型（Gaussian Process Regression, GPR），通过比较精度对不同特征集组合进行评价，选择出最优特征集组合；5) 基于 RF、竞争性自适应重加权抽样（Competitive Adaptive Reweighted Sampling, CARS）、投影变量重要性（Variable Importance in Projection, VIP）三种特征选择方法对最优特征集组合进行降维，筛选出各自优选的特征变量，构建集成提升树模型（Least-squares boosting, LSBoost），基于验证集比较每种方法中不同特征对决定系数（Adj-R²）、均方根误差（RMSE）、相对分析误差（RPD）的贡献，以此确定最优特征变量和特征数量；6) 基于最优特征变量构建不同反演模型，通过比较精度选择最优反演模型，应用该模型并得到土壤有机质。

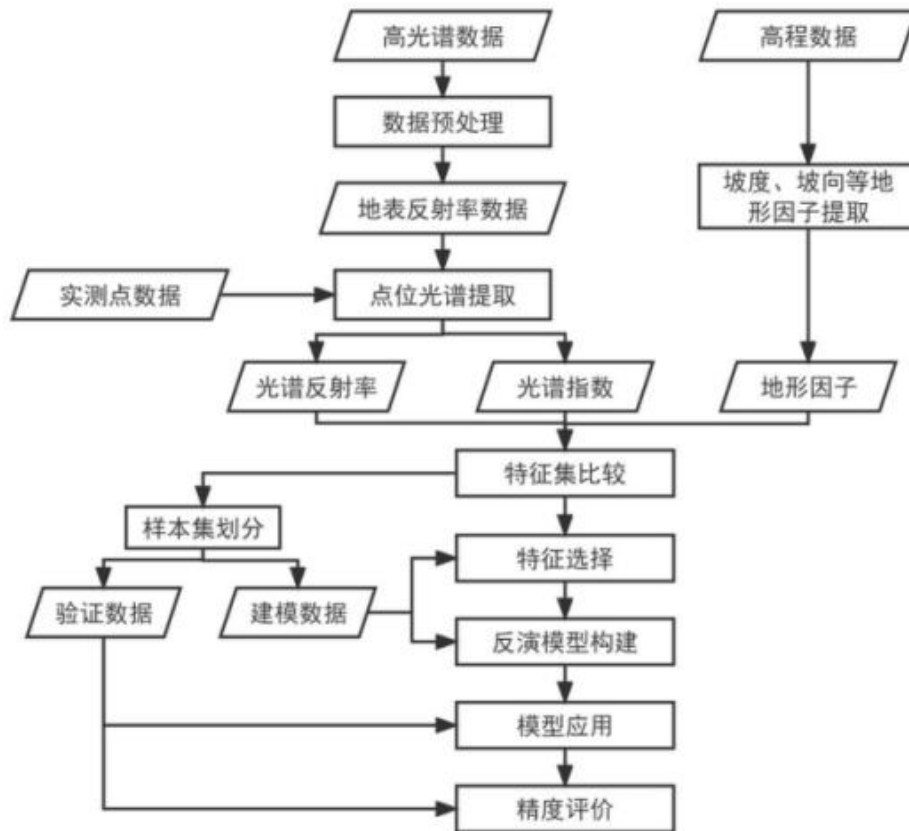


图 1 技术流程图

1) 样本集划分

为减小参量反演模型的预测偏差，考虑到土壤有机质含量差异，将研究区的 163 个样本分成建模集和验证集 2 组，建模集用于特征选择与模型构建，验证集用于评价模型精度。

2) 特征集比较

考虑到光谱反射率、光谱指数、地形因子的特征集数量较大，为提高所选特征的代表性，为此，本研究进行特征集比较，通过计算不同组合的精度，选择最优特征集组合。首先，为解决光谱反射率波段数量多、冗余性较大的问题，分析实测点土壤有机质与光谱反射率（包括点位光谱与基于点位光谱计算不同变换后的光谱）之间的相关性，其相关性的绝对值如下所示，可见原始光谱与土壤有机质之间的相关性最高，倒数、平方根、对数、微分与土壤有机质之间

的相关性较低，由此，选择原始光谱作为光谱反射率特征集。

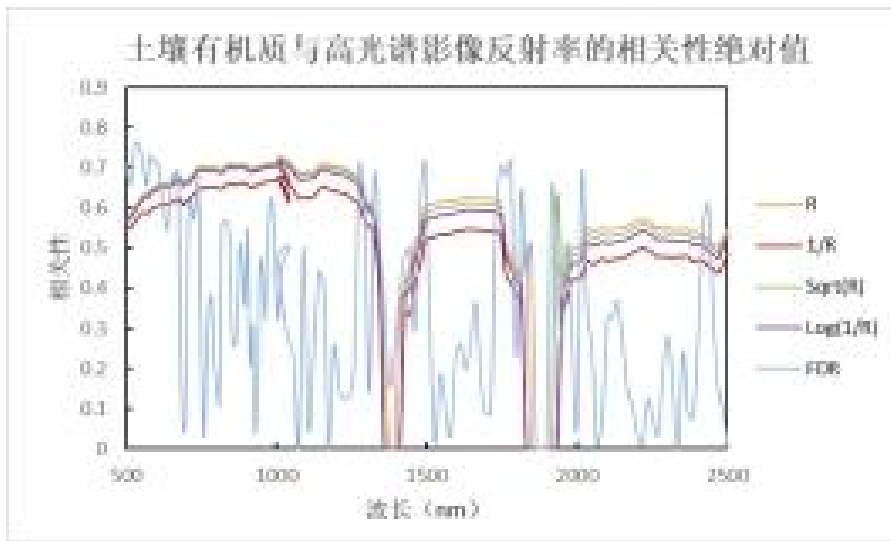


图 2 土壤有机质与光谱反射率间相关性

其次，基于随机蛙算法对不同特征集组合进行降维，组合分为光谱反射率 (R)、光谱指数 (SI)、光谱反射率+地形因子 (R+DEM)、光谱指数+地形因子 (SI+DEM)、光谱指数+光谱反射率 (SI+R)、光谱指数+光谱反射率+地形因子 (SI+R+DEM)，分别选择每组的特征变量，利用这些特征变量建立 GPR 反演模型，通过对不同特征集组合进行精度评价，选择光谱指数+地形因子 (SI+DEM) 为最优特征集组合。

3) 特征选择

为筛选出对模型构建有帮助的特征，基于光谱指数+地形因子的组合，比较了 RF、CARS、VIP 三种特征选择方法在 LSBoost 反演模型下的精度，由此，我们确定最终选用的特征变量与特征数量。

综合比较，我们选择 7 个用于反演土壤有机质含量的特征变量，这些在每组别中均有一定贡献，分别是 elevation、mean_650_750、Slp620_500、SOC_C600、DIOR580735、Slp820_500、DIOR740525。

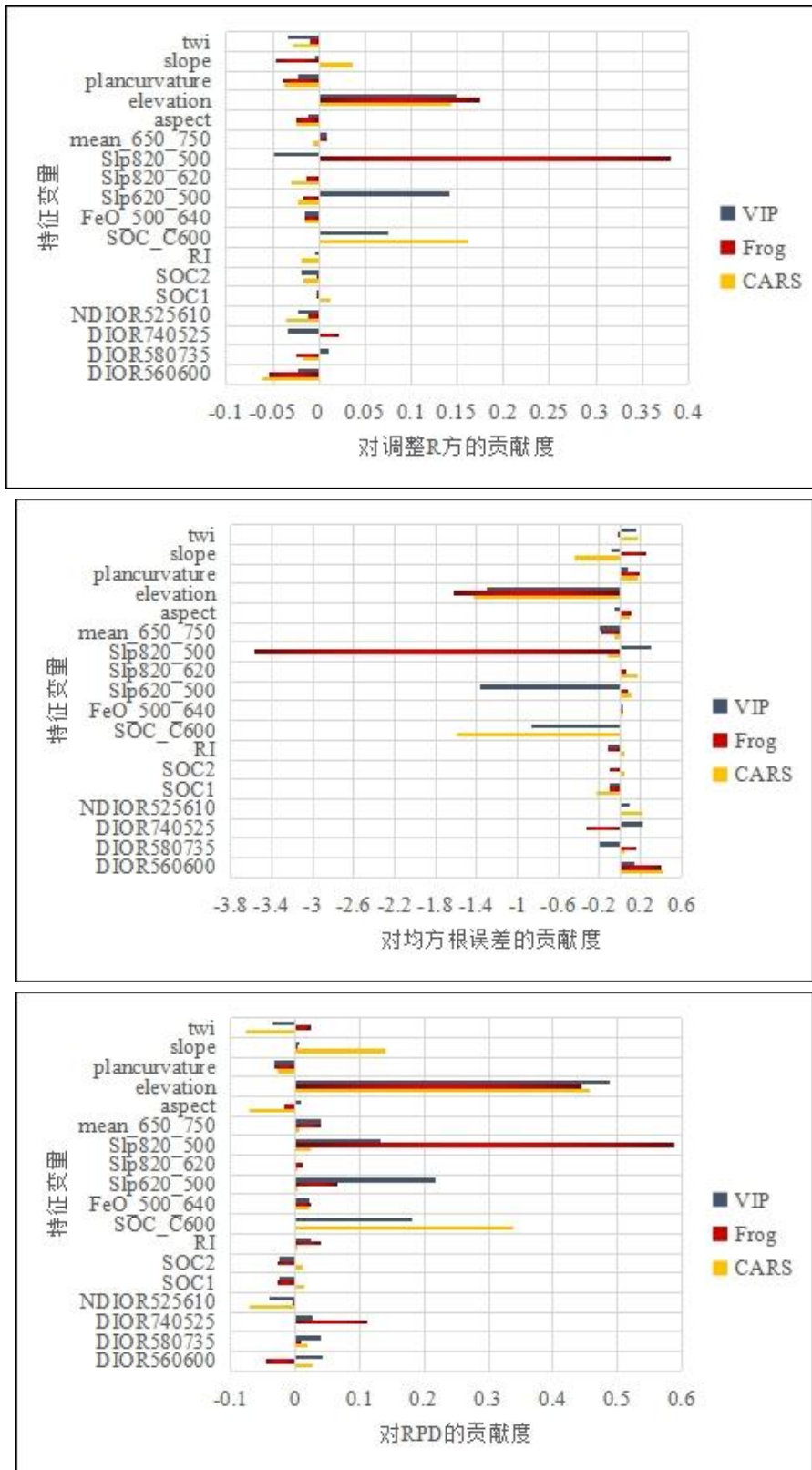


图 3 不同特征选择方法下各特征的贡献度

4) 反演模型构建

基于最优特征变量比较不同反演模型的精度，利用 10 折交叉验证方法进行模型参数优化，分别计算基于交叉验证数据的 R²、RMSE 和基于验证集数据的 R²、RMSE，由此，确定最终的土壤有机质反演模型。下表为不同反演模型的精度对比，从表中可得知，与其他模型相比，GPR 模型交叉验证 R 方、RMSE 最高，分别为 0.75、7.0588；LSBoost 模型验证集 R 方、RMSE 最高，分别为 0.8353、5.1579。GPR 与 LSBoost 模型相比，LSBoost 精度更优。由此，选择 LSBoost（集成学习周期数为 30，最小叶子节点为 8，学习率为 0.1）作为反演模型。

表 2 不同土壤有机质含量反演模型精度对比

	交叉验证 R 方	交叉验证 RMSE	R 方	RMSE
高斯过程回归	0.75	7.0588	0.7618	6.2020
回归树	0.66	8.2681	0.6975	6.9893
逐步线性回归	0.64	8.4768	0.7165	6.7660
集成提升树	0.66	8.1835	0.8353	5.1579

5) 精度评价

基于土壤有机质反演所需特征，分别运用最优反演模型进行应用，利用建模数据和验证数据分别计算各个模型精度。基于建模数据计算精度，土壤有机质含量反演模型 R 方、RMSE 分别为 0.9102、4.1801，实际土壤有机质含量与模型预测值之间的散点图如下图所示。

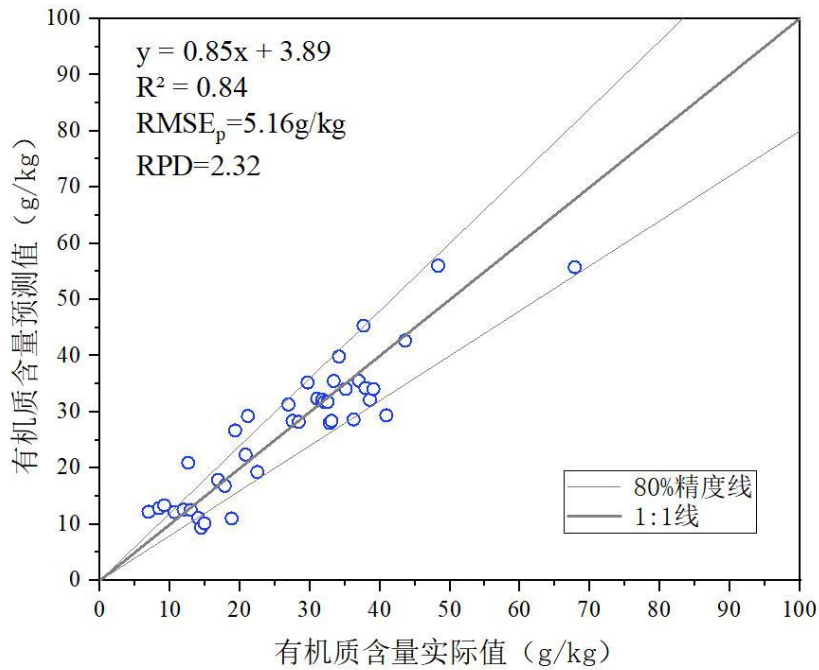


图 4 实际土壤参量含量与模型预测值对比

基于本技术，构建的土壤参量模型反演精度高、质量稳定且多星适配，其中，有机质与砂粒反演精度 R^2 分别高达 0.84 和 0.9。该技术支撑我国率先具备了高光谱卫星耕地土壤质量高精度、规模化、多星协同业务监测能力，大幅度提高了国产星载高光谱数据耕地土壤质量监测的精度与频次。

土壤有机质反演图是基于高光谱卫星的直接反演成果，其空间分辨率为 30 米，可以清晰的反映各参量的宏观分布格局，同时，具有最为丰富的细节信息，可展示土壤有机质在耕地地块内部的分布情况。

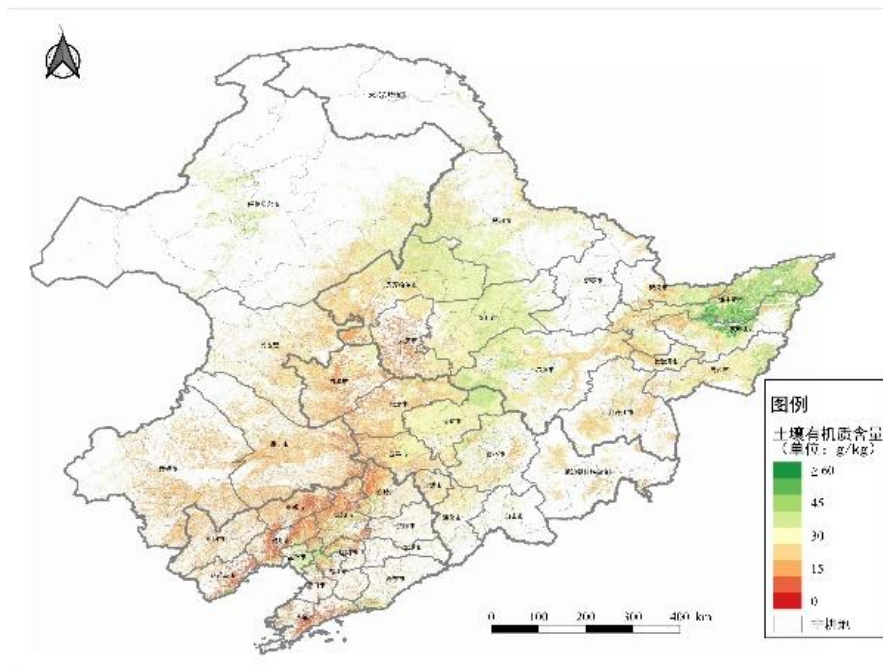


图 5 土壤有机质含量反演图

四、 采用国际标准和国外先进标准的程度，以及与国际、国外同类标准水平的对比情况

经国家标准共享服务平台检索，尚未有已经颁布执行的相关国际标准、他国国家标准记录情况。

五、 与现行法规、标准的关系

本技术规定与现行法律、法规和强制性标准无不协调。

六、 重大分歧意见的处理经过和依据

无。

七、 废止现行有关标准的建议

无。

八、 实施标准的要求和措施建议

本标准颁布实施后，编制组将根据全国地理信息标准化技术委员会及其卫星应用分技术委员会的安排，积极做好标准的宣贯培训等工作。为发挥本标准作为高光谱卫星遥感监测土壤有机质技术规范的作用，建议今后有关国家标准、行业标准制修订时，认真做好与本标准之间的协调。

九、 其他应予说明的事项

无。