

基于人工合成样本和随机森林回归模型的长时序中国寒温带森林类型和树种覆盖度反演

王梦雨^{1,2}, 赵峰², 庞勇^{3,4}, 孟冉⁵, 英文^{3,4}, 岳超^{1,6}

1. 西北农林科技大学 资源环境学院, 杨凌 712100;

2. 华中师范大学 城市与环境科学学院, 武汉 430079;

3. 中国林业科学研究院资源信息研究所, 北京 100091;

4. 国家林业和草原局林业遥感与信息技术重点实验室, 北京 100091;

5. 华中农业大学 资源与环境学院, 武汉 430070;

6. 中国科学院水利部水土保持研究所, 杨凌 712100

摘要: 寒温带森林是陆地上分布面积最广的森林生态系统, 具有重要的生态和社会经济价值。定量描述长时序寒温带森林类型和树种覆盖信息对于量化其生态系统服务功能以及制定森林管理政策具有重要意义。然而受实测覆盖度数据缺乏和多光谱影像光谱信息有限的限制, 现有研究较少探究中分辨率多光谱星载数据 (如 Landsat 卫星) 对中国寒温带森林类型覆盖度和树种覆盖度进行长时序反演的可行性, 并且对于遥感影像获取时间频率 (单时相、多时相) 对反演精度的影响仍缺乏定量评估。为此, 本研究利用人工合成样本和随机森林回归模型对黑龙江省孟家岗林场的森林类型和树种覆盖度分别进行了反演, 并将模型应用于 1986 年—2020 年的 Landsat 影像, 估算孟家岗林场阔叶林和针叶林 35 年的覆盖度。结果表明: (1) 对于森林类型覆盖度反演, 基于生长季 Landsat 波段和植被指数 (归一化耕作指数以及缨帽变换系数) 的中值特征估算的精度最高, 阔叶林覆盖度估算 $\bar{R}^2=0.76$, 针叶林覆盖度估算 $\bar{R}^2=0.71$; (2) 对于树种覆盖度反演, 基于多时相 Landsat 波段和植被指数的精度最高, 落叶松覆盖度估算 $\bar{R}^2=0.40$, 红松覆盖度估算 $\bar{R}^2=0.23$, 樟子松覆盖度估算 $\bar{R}^2=0.61$; (3) 增加影像获取时间密度对于森林类型覆盖度反演精度的提高没有显著贡献 (阔叶林 $\Delta\bar{R}^2=0.01$, 针叶林 $\Delta\bar{R}^2=-0.03$), 但对提高树种覆盖度的反演精度帮助较大 (落叶松 $\Delta\bar{R}^2=0.04$, 红松 $\Delta\bar{R}^2=0.07$, 樟子松 $\Delta\bar{R}^2=0.27$)。本研究可为中国北方森林以及全球寒温带森林类型和树种覆盖度的大尺度长时序估算提供了思路。

关键词: 遥感, 森林类型覆盖度, 树种覆盖度, 人工合成样本, 长时间序列, 机器学习

中图分类号: TP79/P2

引用格式: 王梦雨, 赵峰, 庞勇, 孟冉, 英文, 岳超. 2025. 基于人工合成样本和随机森林回归模型的长时序中国寒温带森林类型和树种覆盖度反演. 遥感学报, 29(1): 118–133

Wang M Y, Zhao F, Pang Y, Meng R, Jia W and Yue C. 2025. Mapping forest type and tree species fractions in China's cold-temperate forests based on synthetically mixed data and random forest regression. National Remote Sensing Bulletin, 29(1): 118–133 [DOI: 10.11834/jrs.20233103]

1 引言

寒温带森林约占全球森林总面积的三分之一 (FAO, 2020), 是陆地上分布面积最广的森林生态系统, 对于气候调节 (Bonan 等, 1992)、碳循环 (Jiang 等, 2002) 和社会经济的可持续发展 (Chen

等, 2016) 等方面都具有重要意义。寒温带森林为人类带来了巨大的生态、经济、社会和健康效益 (Brinck 等, 2017; Drever 等, 2021; Girardin 等, 2021; Heilmayr 等, 2020)。森林类型或树种覆盖度表示的是单位面积内 (例如一个像元), 某一森林类型或树种垂直投影面积的百分比 (陈晋

收稿日期: 2023-04-04; 预印本: 2023-08-03

基金项目: 国家自然科学基金 (编号: 41901382, 42101403); 遥感国家重点实验室开放基金 (编号: OFSLRSS201917); 国家重点研发计划 (编号: 2017YFD0600404)

第一作者简介: 王梦雨, 研究方向为森林碳循环监测。E-mail: wangmy1002@gmail.com

通信作者简介: 赵峰, 研究方向为森林生态系统动态变化及碳循环模拟。E-mail: fzhaoc@cnu.edu.cn

等, 2001)。在气候变化和人类活动的双重影响下, 寒温带森林的干扰频率逐渐增加。因此, 动态准确估算长时序寒温带森林类型和树种覆盖度百分比, 对于了解森林生态系统的结构和功能、监测森林变化以及制定有效的森林保护和管理策略至关重要 (Zhang 等, 2000)。

遥感技术的迅速发展有利于准确反演森林定量参数, 尤其是长时序卫星遥感影像 (例如 Landsat 系列) 的不断发展, 有望对森林类型和树种覆盖度进行大尺度动态估算。对于 Landsat 等中等分辨率卫星数据, 在森林区域存在大量混合像元 (Savage 等, 2015), 因此需要在亚像元尺度反演像元内各类森林类型和树种的丰度 (即覆盖度) (Deng 和 Zhu, 2020)。目前利用遥感数据进行植被覆盖度估算的方法主要有回归模型法和混合像元分解模型法两类。回归分析在像元尺度提供连续的输出, 能够有效估算亚像元尺度上的植被覆盖度 (Okujeni 等, 2014)。通过建立实测植被覆盖度和遥感数据特征 (包括光谱特征, 植被指数以及雷达数据回波等) 之间关系的参数模型, 进行植被覆盖度估算是最常见的方法之一 (Guerschman 等, 2009; Morsdorf 等, 2006)。参数模型需要指定某一特定的函数形式 (例如线性关系, 比值关系, 指数关系等), 能够应对较为简单的植被、林地覆盖度的反演。基于机器学习的回归算法, 可以结合多种遥感特征, 建立高精度稳定的植被覆盖度反演模型 (Hansen 等, 2013; Huang 和 Townshend, 2003; Kattenborn 等, 2020)。然而, 机器学习回归模型的训练, 往往需要大量、全面、准确的样本。通过目视解译或实地采集的手段获取训练模型所需的样本成本高、难度大。

混合像元分解模型物理意义明确, 且不需要大量样本, 能够有效估算植被覆盖度。根据光谱混合函数的性质可将混合像元分解模型分为线性和非线性模型。其中线性光谱混合模型最为常用 (Adams 等, 1995; Xiao 和 Moody, 2005; 姬翠翠等, 2016), 像元二分模型是应用最为广泛的线性光谱混合模型, NDVI 是像元二分模型中最常见的指数 (陈晋等, 2016; 姬翠翠等, 2016; 刘垚焱等, 2021; 王思等, 2022; 王曦和张怡雯, 2021)。为估算不同地形的植被覆盖度, 有学者 (龙爽等, 2020; 吴志杰等, 2016) 利用其他指数 (例如 NDMVI, EVI) 建立像元二分模型, 并取得了较好

的效果。为提高线性光谱混合模型的精度, 现有研究改进了光谱混合模型 (Michishita 等, 2012; Powell 等, 2007; 皮新宇等, 2021) 和端元提取算法 (Li 等, 2015; 胡姝婧等, 2010)。像元二分模型依赖两个关键参数: 纯净植被和裸地对应的植被指数值。然而由于植被和裸土的光谱反射具有区域性, 像元二分模型只适合特定区域的植被覆盖度计算, 在时间和空间上的迁移性较差。并且当混合像元分解模型面对多种端元光谱之间存在相关关系时, 分解精度较低 (陈晋等, 2016)。对于光谱特征十分相近的不同森林类型及树种, 利用混合像元分解模型难以实现其覆盖度的反演。

Okujeni 等 (2013) 结合机器学习回归模型和线性光谱混合模型的优势, 创造性提出了利用人工合成样本和回归模型进行不同地物类型覆盖度分数估算的方法。该方法提取不同地物纯净像元的光谱特征, 并基于线性光谱混合原理生成混合光谱, 构建人工合成样本库。以此作为经验回归模型的训练样本, 估算不同地物类型, 例如不透水面 (Okujeni 等, 2013, 2014, 2015, 2017)、植被 (Okujeni 等, 2021)、裸地 (Okujeni 等, 2015) 等的覆盖度分数。基于线性光谱混合原理人工合成的模拟数据解决了覆盖度反演研究中实测数据获取困难且花费较大的问题 (张明等, 2012)。王光镇等 (2018) 利用实测纯净端元光谱曲线 (包括光合植被, 非光合植被和裸地) 模拟得到混合场景下 66 条 Landsat 8 混合光谱曲线, 成功反演了非光合植被和光合植被的覆盖度。并验证了 Landsat-DFI 指数 (Dead Fuel Index) 对于非光合植被覆盖度反演的有效性。基于人工合成样本和机器学习回归模型的地物覆盖度分数估算方法也被成功应用至不同生态系统中森林覆盖度的估算 (Okujeni 等, 2013; Schug 等, 2020; Senf 等, 2020)。然而, 该方法在物候变化大的欧亚大陆寒温带森林类型覆盖度反演的表现以及对光谱信息更为接近的树种 (例如: 落叶松, 红松以及樟子松等) 覆盖度反演的适用性仍有待探索。

近年来, 时序光谱信息在树种分类研究中得到广泛关注, 并取得了较好的效果 (Hemmerling 等, 2021)。现有研究 (Wang 等, 2020; 雷光斌等, 2016) 指出基于多时相遥感影像光谱特征建立的树种分类模型能够实现更高的分类精度。Persson 等 (2018) 利用 4 个时相的哨兵二号数据

对瑞典一处成熟林进行分类, 结果表明利用多时相特征的模型比利用单一时相特征的模型分类精度高 8.2%。长时间序列的星载多光谱卫星影像 (例如 Landsat 系列) 提供了丰富的时序光谱信息, 能够捕捉到各类树种不同的物候信息, 有效提高了树种分类的精度 (Pasquarella 等, 2018)。然而时序光谱信息对于森林类型和树种覆盖度估算的作用仍缺乏定量评估。基于时间序列影像建立不同时间频率的人工合成样本, 并将其作为回归模型的输入, 可充分利用寒温带森林的丰富物候信息, 为提高森林类型和树种覆盖度反演精度提供了新的可能。对比不同时间频率 (单时相、多时相和时序光谱) 的影像特征对森林类型和树种覆盖度反演精度的影响, 有利于定量评价长时序卫星遥感影像在寒温带森林类型和树种覆盖度反演研究中的潜力, 揭示森林类型和树种对不同生长阶段的光谱响应差异, 以及不同时间频率光谱特征对森林类型和树种覆盖度反演的贡献。从而确定基于中低分辨率长时序卫星数据的森林类型和树种覆盖度反演的最佳光谱特征组合, 为中国寒温带森林乃至整个欧亚地区寒温带森林类型和树种覆盖度反演提供科学依据和方法支持。准确估算长时序寒温带森林类型和树种覆盖度及其变化, 有利于定量评估和监测该生态系统在气候变化下的健康状况和生态功能稳定性, 对于制定和实施寒温带森林可持续管理以及积极应对气候变化具有重要意义 (Zhao 等, 2018)。此外, 长时序寒温带森林类型和树种覆盖度的精确评估, 也为探究气候变化对寒温带森林碳水循环过程影响等相关研究提供可靠的基础数据和参考 (Venäläinen 等, 2020)。

因此, 本研究基于 Landsat 长时序卫星数据, 利用人工合成样本和随机森林回归模型估算了孟家岗林场 1986 年—2020 年森林类型和树种的覆盖度, 探索中国北方森林类型和树种覆盖度反演的最优模型, 并定量评估不同时间频率影像特征对覆盖度反演精度的影响, 以期为更大尺度的中国北方森林以及全球寒温带森林类型覆盖度估算提供思路。

2 研究区概况与数据来源

2.1 研究区概况

本研究选择黑龙江省佳木斯市桦南县东北部

的孟家岗林场为研究区 (图 1), 该区域冬季漫长且严寒, 夏季短暂且凉爽, 年平均气温 2.7°C 。年平均降水量为 550 mm, 且降水主要发生在夏季 (Peng 等, 2019)。孟家岗林场是黑龙江省规模最大的国有人工林林场之一。林场经营面积 155 km^2 , 林业用地 134.43 km^2 , 森林覆盖率为 86.73%。

孟家岗林场的树种组成在中国北方森林中具有一定代表性, 包括了北方寒温带生境中常见的关键树种, 如兴安落叶松 (*Larix gmelini*)、樟子松 (*Pinus sylvestris* var. *mongolica*)、红松 (*Pinus koraiensis*)、柞树 (*Quercus mongolica*) 等。林场除了主要树种的纯林, 还包括针叶混交林, 阔叶混交林以及针阔混交林等混交林。

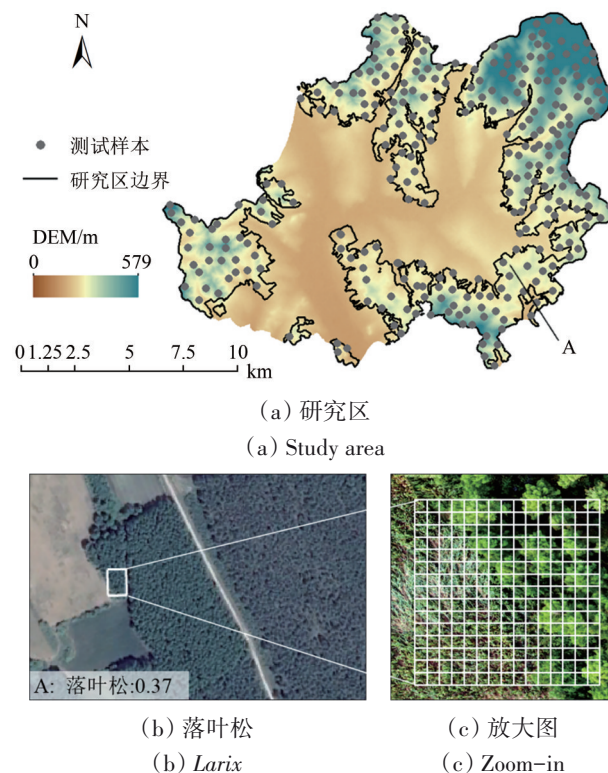


图 1 研究区概况及落叶松覆盖度估算示意图

Fig. 1 Study area and example of *Larix gmelini* cover fraction

2.2 数据及预处理

本研究使用了 1986 年—2020 年 Landsat 系列卫星 (包括 Landsat 5、Landsat 7 和 Landsat 8) 的生长季清晰无云的影像 (1985 年数据无法完全覆盖研究区)。并通过目视的方法, 挑选了 2016 年—2020 年生长季和落叶期能够完整覆盖研究区且无云覆盖的 Landsat 8 地表反射率数据 SR (Surface Reflectance) (表 1) 参与模型的训练。由于 2020 年

生长季缺少一景能够完全覆盖研究区的高质量影像，因此本研究对2020年7月和8月研究区所有影像的可见光波段、近红外波段、短波红外波段以及本研究所使用的植被指数（式（2）—（5））进行中值合成，作为2020年生长季的特征影像。Landsat系列数据空间分辨率为30 m，时间分辨率为16 d，本研究所使用的波段为可见光、近红外和短波红外共计6个波段。本研究涉及的影像均来自GEE（Google Earth Engine），Landsat数据为经过几何校正和辐射校正的地表反射率数据，只需要对遥感影像进行去云，研究区拼接裁剪等预处理。采用的去云算法为GEE提供的CFmask算法（Foga等，2017）。中值影像合成采用的是GEE提供的ImageCollection.median算法。本研究采用MODIS产品MCD12Q2确定研究区生长季的开始与终止时间，确定孟家岗林场生长季为每年6月20日至9月20日。

表1 单时相和多时相所使用的Landsat 8影像信息

Table 1 Landsat 8 images used for single- and multi-date

物候期	传感器	条带号	时间
生长季	Landsat 8	116028	2016-07-06
		115028	2017-08-19
		115028	2019-07-23
		115028, 116028	2020-07-08 中值影像
落叶期	Landsat 8	116028	2016-10-10
		115028	2017-10-22
		115028	2019-10-27
		115028	2020-10-30

验证样本来自中国林业科学研究院资源信息研究所基于机载遥感系统CAF-LiCHy（The Chinese Academy of Forestry’s LiDAR CCD and Hyperspectral airborne observation system）生成的孟家岗林场硬分类产品（2017年）（Jia和Pang，2023），该产品空间分辨率为2 m，使用的分类方法为随机森林，分类精度可达到91.0%以上。验证样本数量共计773个。

3 研究方法

在构建覆盖度反演模型之前首先建立包含不同时间频率特征（单时相光谱特征、多时相光谱特征、时序光谱特征）的纯净光谱库。其中，单时相特征为孟家岗林场生长季影像特征，多时相

光谱特征包括生长季和落叶期两个关键物候期的影像特征；时序光谱特征指的是基于孟家岗林场生长季的合成影像（包括极大值、极小值、中值和均值）所提取的多个光谱指标。之后基于线性光谱混合模型原理（式（1））和上述纯净光谱库，生成混合光谱M及其对应的丰度 f_k （即不同森林类型和树种对应的覆盖度）。将纯净光谱库和混合光谱库合并，得到人工合成光谱库，并将其作为随机森林回归模型的输入。建立阔叶林、针叶林、落叶松、红松和樟子松5个目标类的覆盖度反演模型。将训练好的模型应用至孟家岗林场，生成5个目标类对应的覆盖度结果。最后利用基于机载激光雷达和高光谱数据生成的孟家岗林场高空间分辨率树种分类产品（2 m）作为参考数据，评价上述覆盖度反演模型的精度。将效果最好的模型应用至1986年—2020年，生成长时序森林类型/树种覆盖度结果。研究技术路线如图2所示。

$$M = \sum_{k=1}^n f_k R_k + \varepsilon \tag{1}$$

式中，M向量为混合光谱， f_k 为第k类端元的丰度比例， R_k 向量为第k类端元的光谱， ε 为残差项。

3.1 构建人工合成光谱库

利用谷歌地球提供的高空间分辨率影像和孟家岗小班数据，采集纯净样本，详见表2；并提取对应像元的光谱和植被指数特征，生成纯净光谱库。

本研究采取了分层的思路去反演森林类型和中国北方主要针叶树种的覆盖度。森林类型分为阔叶林和针叶林。根据中国寒温带优势树种类别，将树种类型设定为落叶松、樟子松和红松。非林地只作为背景类型参与人工合成光谱库过程，并不进行覆盖度反演。孟家岗林场阔叶树种类型多样，但优势树种主要为柞树以及以柞树为优势树种的阔叶混交林（柞树林面积占阔叶林面积的62.39%），因此不进行阔叶林具体树种的覆盖度反演。

为保证采集的纯净样本所对应的Landsat像元为纯净像元，选取面积>1600 m²的同质区域作为纯净样本。对于纯净样本的定义为：在同质区域内，林地样本必须为单一树种，且该类森林（阔叶林、针叶林、落叶松、红松或樟子松）的覆盖度需>85%；非林地样本可以混合多种地物类型，只要不

包含森林即可。本研究共采集了纯净像元476个。对于非林地，为减少纯净光谱库中非林地样本的光谱异质性，并未包括发生森林干扰（火灾、风倒、病虫害等）的样本。为使回归模型能够适应由于物候和年际的变化而带来的树种光谱的变化，本研究从谷歌地球提供的多年（2016年、2019年和2020年）高分辨率影像中采集纯净样本，并提取样本对应的Landsat纯净光谱。

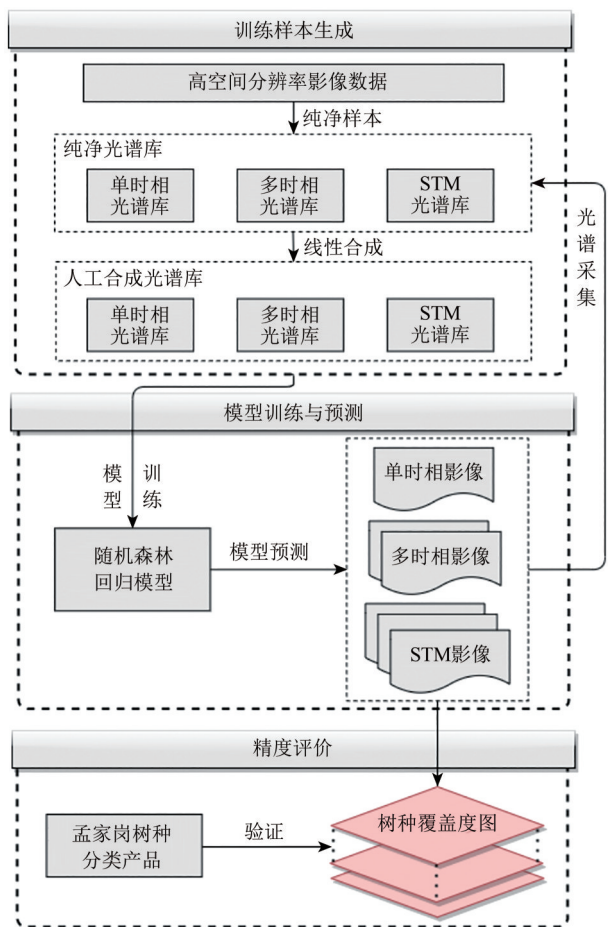


图2 技术路线图(STM表示时序光谱特征)

Fig. 2 Technology roadmap(STM=Spectral Temporal Metrics)

随后，基于线性光谱混合模型的假设，人工合成了训练回归模型所需的大量样本，其基本原理如图3所示。图3表示了目标类分别为阔叶林和针叶林的四种混合情况，目标类为阔叶林，2端元混合的背景类为非林地，混合比例分数分别为50%和50%；3端元混合的背景类为非林地和针叶林，混合比例分数分别为20%、40%和40%。与以针叶林为目标类的合成与阔叶林原理相同。该过程可描述为以下3个具体步骤：

步骤1：将纯净光谱库划分为两个类别：目标

类和背景类。回归模型反演结果表示目标类的覆盖度；背景类为除目标类外的其他类别。

步骤2：确定人工合成混合光谱的参数：（1）混合复杂度，即混合像元内所含端元的数量。由于北方人工林组成结构简单，因此设置混合复杂度为2端元和3端元；（2）混合间隔，混合间隔指的是目标类光谱和背景类光谱混合时，0—100%的给定分数范围内的中间数。本研究通过随机生成0—100%之间的数作为混合间隔；（3）人工合成样本的数量，本研究对于每个目标类，生成1000个混合光谱。包括2端元混合光谱800个，3端元混合光谱200个。

步骤3：将纯净光谱库和人工合成的混合光谱库合并，得到用于回归模型训练的样本库，训练样本数量共计为1476。

表2 影像光谱库的层次结构以及纯净像元的数量

Table 2 Hierarchical structure of the image spectral library and corresponding number of image spectra

森林类型	森林树种	端元的数量		
		2016年	2019年	2020年
针叶林	落叶松	10	19	50
	红松	14	15	34
	樟子松	15	18	35
阔叶林	柞树	7	0	50
	杨树	4	0	15
	其他阔叶树种	10	10	20
非林地	裸土	3	3	4
	作物	25	25	30
	灌木	4	5	10
	不透水面	—	—	23
	水体	—	—	18

为测试人工合成样本和随机森林回归模型对森林类型和树种覆盖度反演的适用性以及探究基于不同特征人工合成光谱库所建立的覆盖度反演模型的精度，本研究建立了3组纯净光谱库，分别包含生长季单时相光谱特征、生长季+落叶期多时相光谱特征和生长季时序光谱指标（包括生长季所有可用观测的极大值、极小值、均值、中值和标准差）（图2）。为了提高回归模型的精度，本研究不仅使用了光谱波段，还提取了区分北方人工林关键的植被指数：归一化耕地指数 NDTI (Normalized Difference Tillage Index)、缨帽变换亮

度指数 TCB (Tasseled Cap Brightness)、缨帽变换绿度指数 TCG (Tasseled Cap Greenness) 和缨帽变换湿度指数 TCW (Tasseled Cap Wetness) (Wang 等, 2022)。NDTI 通过计算短波红外波段之间的差异能够有效区分北方人工林的不同树种。缨帽变换系数通过增强不同树种光谱特征之间的差异, 从而提高树种之间的光谱可分性。上述指标的具体计算公式分别如下:

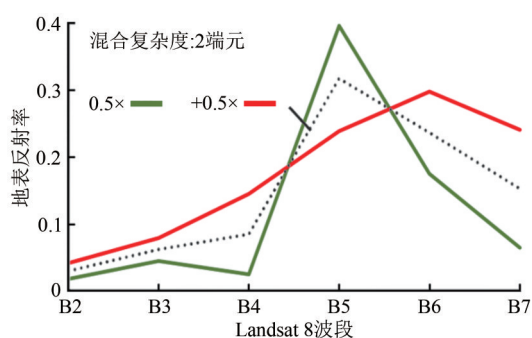
$$\text{NDTI} = \frac{\text{SWIR1} - \text{SWIR2}}{\text{SWIR1} + \text{SWIR2}} \quad (2)$$

$$\text{TCB} = 0.3029 \cdot \text{Blue} + 0.2786 \cdot \text{Green} + 0.4733 \cdot \text{Red} + 0.5599 \cdot \text{NIR} + 0.5080 \cdot \text{SWIR1} + 0.1872 \cdot \text{SWIR2} \quad (3)$$

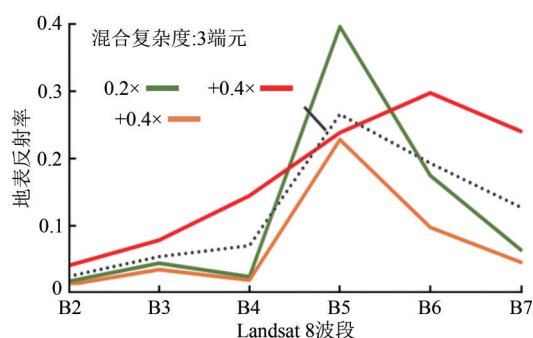
$$\text{TCG} = -0.2848 \cdot \text{Blue} - 0.2435 \cdot \text{Green} - 0.5426 \cdot \text{Red} + 0.7243 \cdot \text{NIR} + 0.0840 \cdot \text{SWIR1} - 0.1800 \cdot \text{SWIR2} \quad (4)$$

$$\text{TCW} = 0.1511 \cdot \text{Blue} + 0.1973 \cdot \text{Green} + 0.3283 \cdot \text{Red} + 0.3407 \cdot \text{NIR} - 0.7117 \cdot \text{SWIR1} - 0.4559 \cdot \text{SWIR2} \quad (5)$$

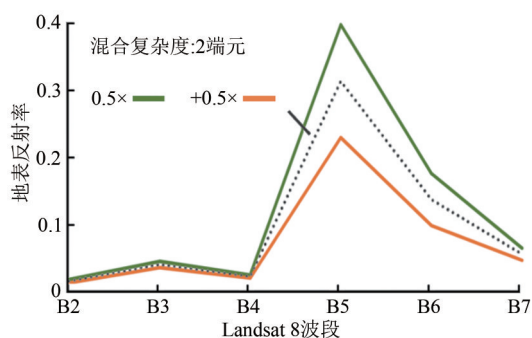
式中, Blue、Green、Red、NIR、SWIR1 和 SWIR2 分别为 Landsat 8 的蓝光、绿光、红光、近红外、以及两个短波红外波段。



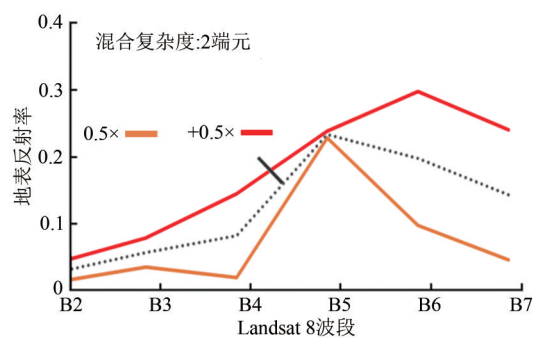
(a) 阔叶林覆盖度 50% (非林地和阔叶林混合)
(a) Broadleaf forest cover fraction 50%
(Broadleaf forest + non-forest)



(b) 阔叶林覆盖度 50% (针叶林, 非林地和阔叶林混合)
(b) Broadleaf forest cover fraction 50% (Broadleaf forest + needleleaf forest + non-forest)



(c) 针叶林覆盖度 50% (阔叶林和针叶林混合)
(c) Needleleaf forest cover fraction 50%
(Needleleaf forest + broadleaf forest)



(d) 针叶林覆盖度 50% (非林地和针叶林混合)
(d) Needleleaf forest cover fraction 50%
(Needleleaf forest + non-forest)

— 阔叶林 (柞树) — 非林地 (不透水面) — 针叶林 (樟子松)

图3 混合光谱生成示意图

Fig. 3 Examples of synthetic mixture spectra libraries

3.2 构建集成随机森林回归模型

随机森林 RF (Random Forest) 是一种结合了 Bagging 集成学习理论与随机特征选择方法, 基于多棵决策树集成的监督式机器学习算法, 广泛应用于分类和回归分析 (Breiman, 2001)。随机森林因实现简单, 训练速度快, 抗过拟合能力强等优

点被广泛使用本研究通过调用 Python “sklearn” 库中的 RandomForestRegressor 实现随机森林回归模型的构建, 设置 ntree (随机树的棵数) 为 500, mtry (每个节点上特征的数量) 为 2 到 k , k 表示的是所有特征的总数量。将人工合成光谱库划分为训练样本集和测试样本集, 两个数据集的大小分别为

样本库的 70% 和 30%。为提高模型的预测精度和稳定性, 本文基于 Bagging 的集成策略, 对单个随机森林回归模型进行集成, 生成最终预测模型, 见图 4。

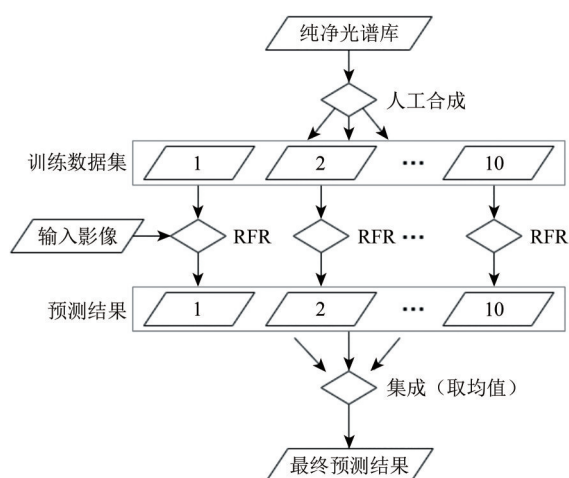


图4 集成随机森林回归模型示意图

Fig. 4 Schematic diagram of ensemble random forest regression model

3.3 覆盖度反演精度评估

首先利用基于机载激光雷达和高光谱数据生成的空间分辨率为 2 m 的孟家岗林场分类产品生成 30 m×30 m 网格单元 (共计 773 个), 并分别计算这些网格单元中阔叶林、针叶林、落叶松、红松和樟子松的覆盖度, 以此作为模型精度验证的真实值。本研究选择校正决定系数 (adjusted R^2 , 后续表示为 $\bar{R}^2$) 和均方根误差 RMSE (Root Mean Square Error) 两个评价指标衡量覆盖度反演结果。具体计算公式如下:

$$R^2 = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (6)$$

$$\bar{R}^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - p - 1} \quad (7)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (y_i - \bar{y})^2}{n}} \quad (8)$$

式中, $\hat{y}_i$ 为预测值; y_i 为真实值; n 为样本个数; p 为模型中解释变量的总数; $\bar{y}$ 为所有样本真实值的平均值。

4 结果与分析

4.1 森林类型及树种覆盖度反演最优回归模型

图 5 显示了森林不同分类层次在生长季 (2020 年 6 月) 和落叶期 (2020 年 10 月) 的光谱特征。可见: 森林类型的光谱曲线在生长季可分性大, 在落叶期光谱可分性降低; 树种光谱可分性与森林类型相反, 生长季各针叶林树种光谱曲线相似。其中, 落叶期光谱可分性提高, 尤其是落叶松和其他针叶林的光谱可分性提高。对于针叶林和阔叶林, 在生长季, 两种森林类型光谱特征稳定, 且存在较大区别。然而在落叶期, 落叶松表现出与阔叶林相似的光谱特征, 针叶林内部光谱异质性提高, 降低了阔叶林和针叶林的光谱可分性。对于孟家岗主要的针叶树种, 生长季针叶林光谱特征相似; 落叶期, 落叶松光谱特征明显区别于其他常绿针叶林 (红松和樟子松)。

由表 3 可见: 仅基于生长季单时相光谱特征, 森林类型覆盖度反演模型可以达到较高的精度, 其中阔叶林覆盖度反演模型的 $\bar{R}^2=0.69$, 针叶林 $\bar{R}^2=0.43$, 增加影像时间频率后, 建模精度没有显著提高; 对于中国北方常见的针叶树种, 利用多时相光谱特征, 覆盖度反演精度有明显提高, 尤其是樟子松覆盖度, 其模型 $\bar{R}^2$ 可达到 0.61; 然而利用生长季时序光谱特征没有提高森林类型或树种覆盖度反演模型的精度; 生长季植被的光谱特征稳定, 各个时序光谱指标 (极大值、极小值、中值和均值) 较为相近, 多个时序光谱指标对于森林类型和树种覆盖度的反演没有提供更多的光谱信息。Srisa-An (2021) 研究发现并且生长季的各个指标之间呈现高度的相关性, 回归模型或受到特征共线性的影响, 覆盖度模型的反演效果有所下降。

考虑到生长季单时相特征在建模时取得的较高精度, 以及难以通过目视方法筛选出长时间序列上的生长季单时相高质量影像。本研究对生长季时序光谱特征的单个指标进行了测试, 结果如表 4 所示。对于森林类型覆盖度反演效果最佳的模型为生长季中值模型, 基于该特征建立的森林类型覆盖度模型平均 $\bar{R}^2$ 为 0.74, 比单时相模型高 0.18。其中, 阔叶林覆盖度模型 $\bar{R}^2$ 为 0.76, 针叶林

覆盖度 $\bar{R}^2$ 可达到0.71。相较于单时相模型，RMSE也有不同程度的降低，阔叶林和针叶林分别降低了0.02和0.11。为获得一景完整覆盖研究区的生长季数据，本研究规定生长季单时相影像的采集时间为6月20日到9月20日的所有影像。生长季

单时相影像采集时间的差异会导致预测精度的下降（Senf等，2020）。利用生长季光谱时序的中值特征，能够有效降低影像获取时间不同对模型反演效果的影响，进而提高森林类型覆盖度的反演精度。

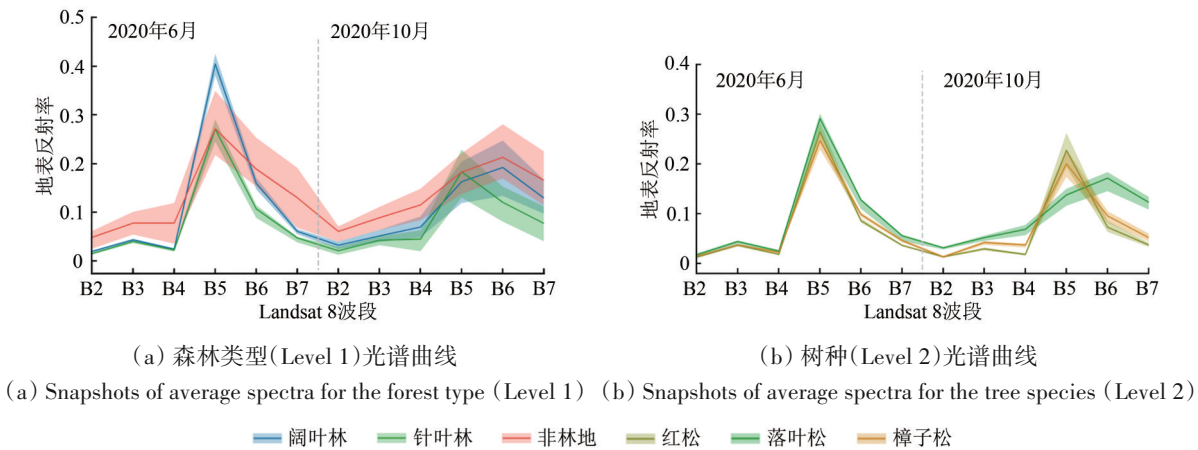


图5 Landsat 8 光谱曲线(实线表示的是光谱均值,上下界限分别表示的是该类纯净像元光谱值的25%和75%)

Fig. 5 Snapshots of average Landsat 8 spectra

表3 基于不同时间频率特征的森林类型和树种覆盖度验证精度

森林类型及树种		单时相		多时相		生长季STM		多时相	生长季STM
		$\bar{R}^2$	RMSE	$\bar{R}^2$	RMSE	$\bar{R}^2$	RMSE		
森林类型	阔叶林	0.69	0.24	0.70	0.24	0.21	0.41	0.01	-0.48
	针叶林	0.43	0.34	0.40	0.23	0.27	0.51	-0.03	-0.16
	均值	0.56	0.29	0.55	0.24	0.24	0.46	-0.01	-0.32
森林树种	落叶松	0.36	0.32	0.40	0.29	0.14	0.23	0.04	-0.22
	红松	0.16	0.46	0.23	0.29	0.04	0.30	0.07	-0.12
	樟子松	0.34	0.30	0.61	0.23	0.10	0.32	0.27	-0.24
	均值	0.29	0.36	0.41	0.27	0.09	0.28	0.13	-0.19

表4 基于生长季单个指标的森林类型和树种覆盖度验证精度

森林类型及树种		生长季STM极大值		生长季STM极小值		生长季STM均值		生长季STM中值		极大值	极小值	均值	中值
		$\bar{R}^2$	RMSE	$\bar{R}^2$	RMSE	$\bar{R}^2$	RMSE	$\bar{R}^2$	RMSE	$\Delta\bar{R}^2$ to 单时相			
森林类型	阔叶林	0.73	0.22	0.54	0.33	0.73	0.24	0.76	0.22	0.04	-0.15	0.04	0.07
	针叶林	0.68	0.24	0.54	0.30	0.69	0.24	0.71	0.23	0.25	0.11	0.26	0.28
	均值	0.71	0.23	0.54	0.32	0.71	0.24	0.74	0.23	0.15	-0.02	0.15	0.18
森林树种	落叶松	0.31	0.30	0.18	0.32	0.35	0.28	0.39	0.28	-0.05	-0.18	-0.01	0.03
	红松	0.14	0.31	0.13	0.29	0.15	0.28	0.13	0.29	-0.02	-0.03	-0.01	-0.03
	樟子松	0.49	0.26	0.45	0.26	0.44	0.26	0.53	0.23	0.15	0.11	0.10	0.22
	均值	0.31	0.29	0.25	0.29	0.31	0.27	0.35	0.26	0.02	-0.04	0.02	0.07

对于北方常见针叶林树种覆盖度的反演，基于生长季中值和均值，模型反演精度相较于生长

季单时相有不同程度的提高。但是其精度仍低于多时相光谱特征模型。

4.2 森林类型覆盖度反演结果

森林类型的覆盖度估算结果与空间分布见图6。由图6(b)(c)可见：森林类型(针叶林和阔叶林)覆盖度反演最优模型为基于生长季中值指标建立的随机森林回归模型，将该模型应用至孟家岗林场，得到针叶林和阔叶林的覆盖度。由图6(a)可见，森林类型的覆盖度反演结果整体上与林相图的空间格局相似。林场从东北到西南有明显的从阔叶林到针叶林的过渡。由图6(b)可见，对于阔叶林覆盖度的估算效果最好，尤其是在阔叶林覆盖度高的区域。阔叶林纯林的光谱特征与针叶林和非林地的光谱特征区分度较大，在其纯林的区域覆盖度反演效果好。然而根据图6(b)放大

图，在阔叶林覆盖度较低的区域，预测值存在一定的高估。孟家岗林场阔叶林多以萌发的天然林为主，阔叶林覆盖度较低区域地物类型多为灌木、裸土以及阔叶林的混交。由于灌木的光谱表现与阔叶林存在较高的相似性(Qian等, 2019)，因此在阔叶林稀疏区域，其覆盖度的估算值高于实测值。根据图6(c)和图7(b)可知，针叶林覆盖度的估算效果略差于阔叶林，尤其是在针叶混交林区域，其覆盖度的预测值低于实际值。在针叶林混交林区域，其光谱异质性高于针叶林纯林区域，对于该区域针叶林覆盖度的估算结果存在一定偏差，预测值低于实际值。

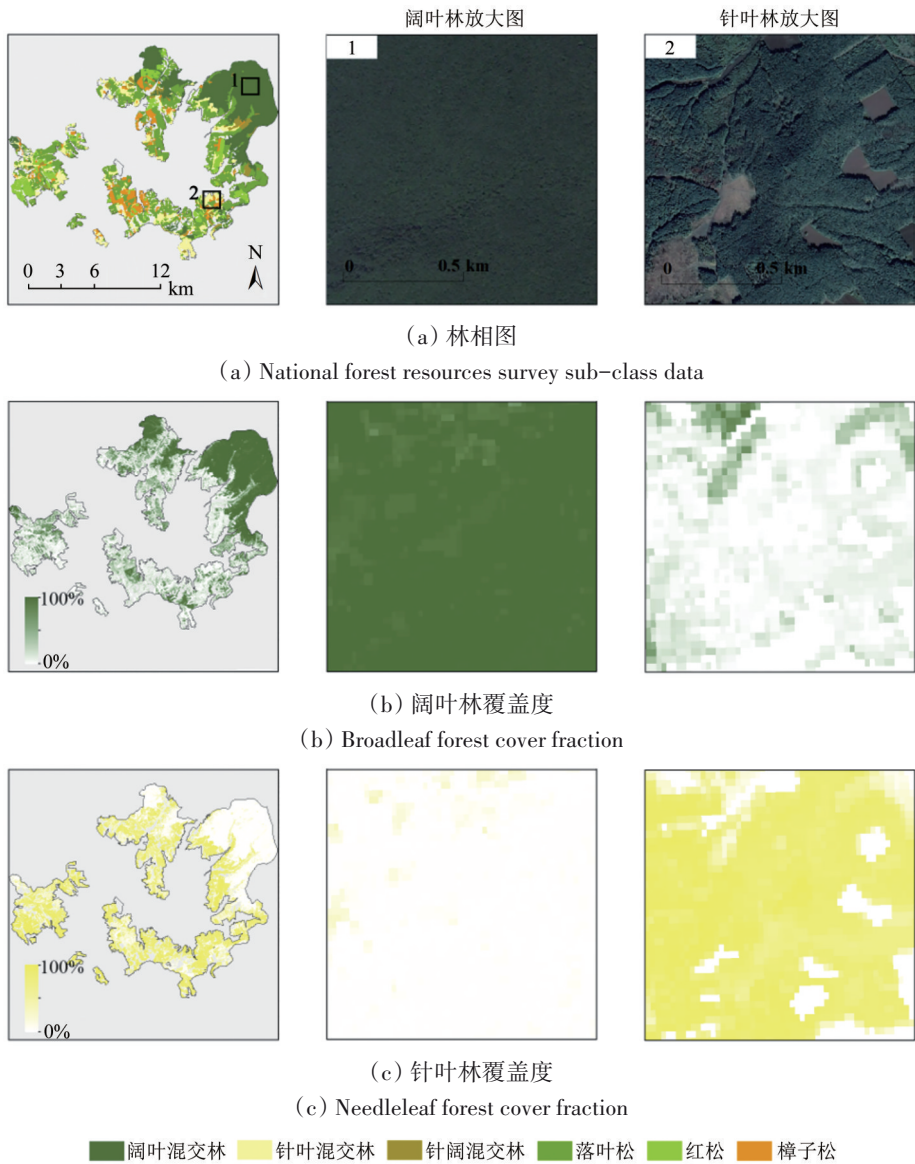


图6 森林类型覆盖度估算结果(高空间分辨率数据来自于谷歌地球)

Fig. 6 The estimated results of forest type cover fraction (The very-high resolution images are from Google Earth)

由图 7 (a) (b) 可知，对于森林类型覆盖度反演结果，其验证样本参考比例和预测比例的散点图显示了实测覆盖度和预测覆盖度之间较高的相关性：验证样本点主要分布在 1 : 1 线附近，阔叶林覆盖度和针叶林覆盖度对应的回归直线斜率

分别为 0.81 和 0.76。点距离 1 : 1 的偏离度小，即残差普遍较低。由于在阔叶林覆盖度较低区域预测结果的高估，导致了阔叶林预测残差高于 0。与之相反，由于针叶林覆盖度高值区的低估，使得针叶林覆盖度的整体预测残差低于 0。

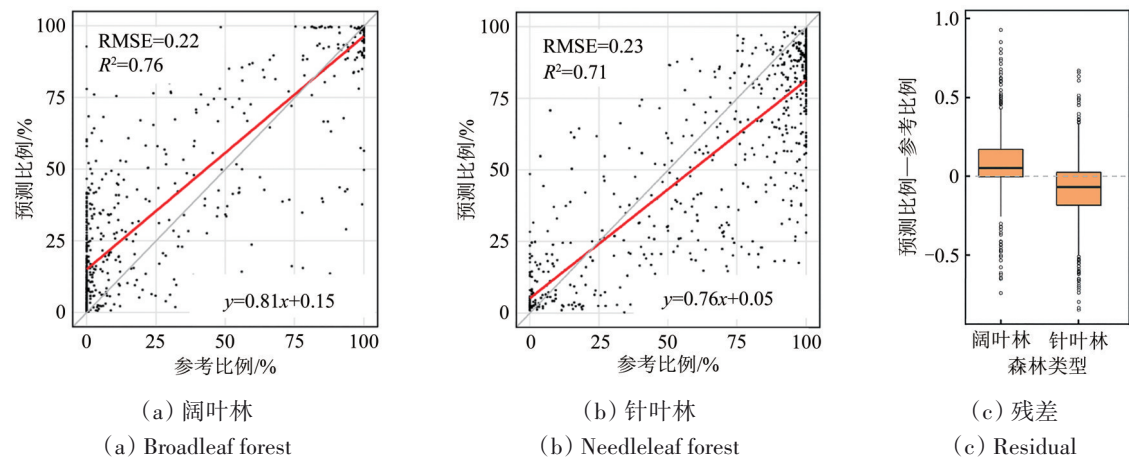


图 7 森林类型估算覆盖度与真实值的相关性
Fig. 7 Correlation between estimated forest type cover fraction and true value

由图 5 (a) 可知，在生长季，植被的光谱特征稳定，且中国北方阔叶林（主要树种包括柞树、水曲柳、东北杨）和针叶林（主要树种包括落叶松、红松、樟子松）的光谱区分度较大，仅使用生长季光谱就能较好区分针叶林和阔叶林。增加落叶期的光谱特征后，在此阶段落叶松光谱曲线与阔叶林类似，针叶林内部光谱异质性提高，对于针叶林和阔叶林的区分没有提供更多有效信息。

4.3 森林树种覆盖度反演结果

表 3 为森林树种覆盖度的反演结果，由表 3 可见，基于多时相 Landsat 波段和植被指数特征建立的森林树种覆盖度模型反演效果最好。根据图 8 (d) 可知，樟子松覆盖度的估算效果最好，不论是在樟子松纯林区域还是针叶混交林区域，樟子松覆盖度估算模型均表现出了较好的效果。落叶松覆盖度反演效果次之，图 8 (b) 显示，在其主导区域反演效果最好。孟家岗针叶林的优势树种为落叶松和樟子松，且多为成熟林，光谱特征稳定，基于光谱特征对其覆盖度反演精度高。然而红松覆盖度反演结果相对较差，在其他树种主导区域，红松覆盖度估算结果存在明显的高估。孟家岗林场红松种植面积低，且多为中龄林和幼龄林，光谱变异性较强，因此基于人工合成样本和随机森

林回归模型对其覆盖度反演效果较差。

提高影像获取的时间密度（单时相—多时相），有效提高了森林树种覆盖度反演模型的精度。对于针叶林各类树种覆盖度的反演，增加落叶期的光谱信息后，其覆盖度反演精度明显提高，尤其是两种常绿针叶林（红松和樟子松）。根据图 5 (b) 显示，在落叶期，两种常绿针叶林的光谱特征明显区别于阔叶林和落叶松。然而随着分类层次的细致化（森林类型—树种），覆盖度反演效果明显降低。由图 5 可见，随着分类层次的递进，类与类之间的光谱可分性降低。不论是森林的分类研究还是森林覆盖度的反演研究，都会随着森林分类系统的细化，精度有所降低（Li 等，2021；Wang 等，2022）。

4.4 森林类型覆盖度长时序制图结果

本研究利用 1986 年—2020 年 Landsat 数据以及森林类型覆盖度反演最优模型（生长季 Landsat 波段和植被指数中值模型）生成了孟家岗林场 35 年的森林类型覆盖度结果。长时序 Landsat 特征影像为孟家岗林场生长季（每年 6 月 20 日—9 月 20 日）的波段和植被指数中值合成影像，生长季中值影像合成方法参考 2.2 节。上述得到的孟家岗林场 1986 年—2020 年生长季中值数据用于后续针叶林和阔叶林覆盖度预测。

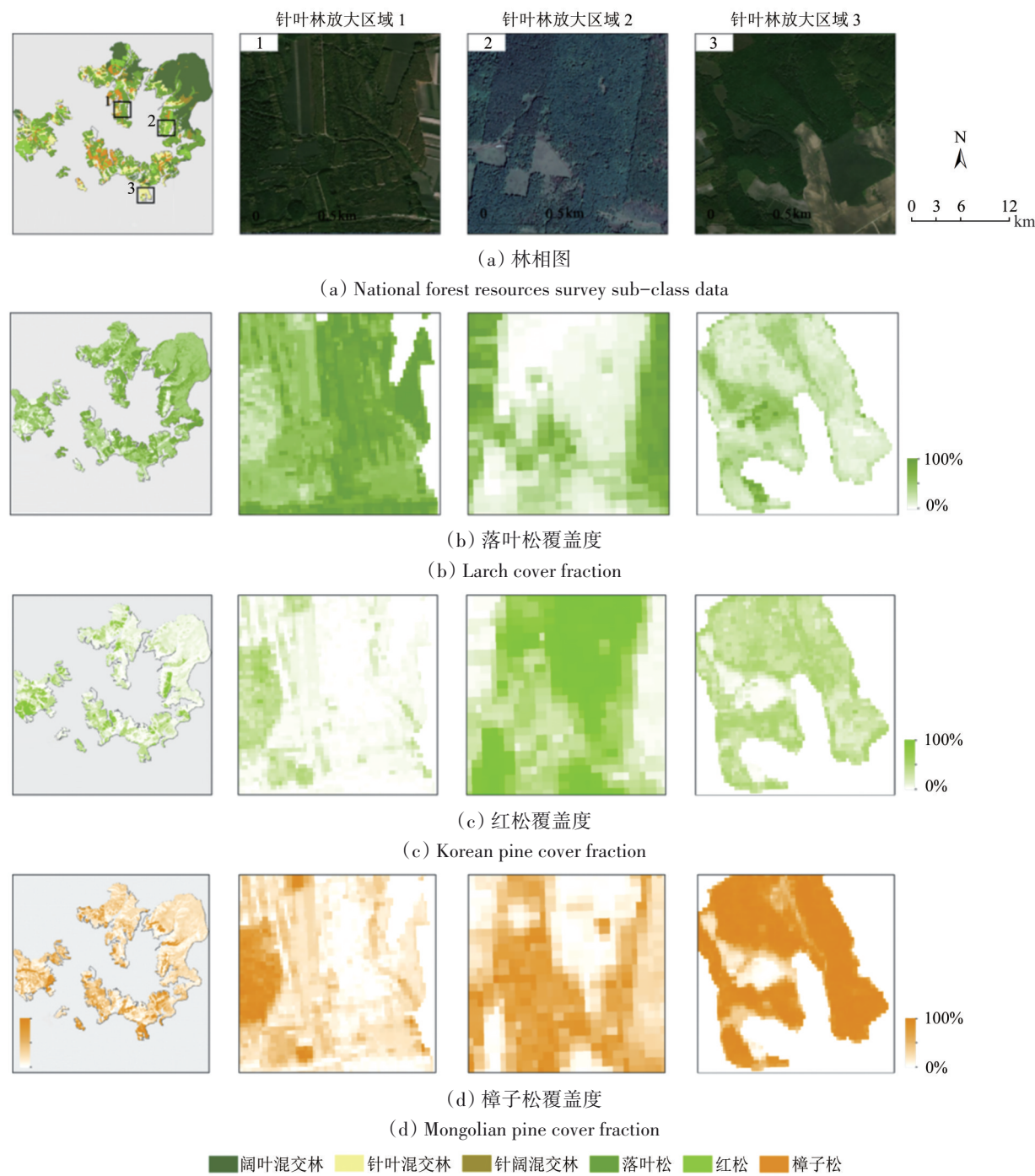


图8 森林树种覆盖度估算结果(高空间分辨率数据来自于谷歌地球)

Fig. 8 The estimated results of tree species cover fraction (The very-high resolution images are from Google Earth)

本研究分析了稳定阔叶林，稳定针叶林以及发生采伐的针叶林区域的覆盖度变化。分别随机筛选了100个稳定阔叶林和针叶林纯林像元以及在2015年发生采伐的像元（46个），图9为上述样本1986年—2020年的覆盖度。两类稳定的森林类型纯林区域的覆盖度较高，阔叶林和针叶林的平均覆盖度分别为79.4%和74.0%。受光学影像的质量和森林生长的影响，阔叶林（72.6%—92.1%）和

针叶林（40.9%—85.7%）的覆盖度在年际间呈现波动上升现象。相较于在像元尺度对森林类型进行分类（即每个像元被分为某一森林类型），在亚像元尺度计算不同森林类型的覆盖度，更能捕捉到森林的生长变化情况（Senf等，2020）。随着时间的变化，稳定森林区域的冠层不断生长，阔叶林和针叶林覆盖度逐渐增大。基于人工合成样本和随机森林回归模型预测森林覆盖度，能够较好

识别森林干扰。2015 年，孟家岗林场对部分针叶林进行了采伐。采伐前，该区域为稳定的针叶林，平均覆盖度为 78.9%。2015 年该区域发生采伐后，

针叶林覆盖度迅速下降至大约 10%，随后 5 年，该针叶林区域经历了缓慢的恢复，覆盖度逐渐提高。

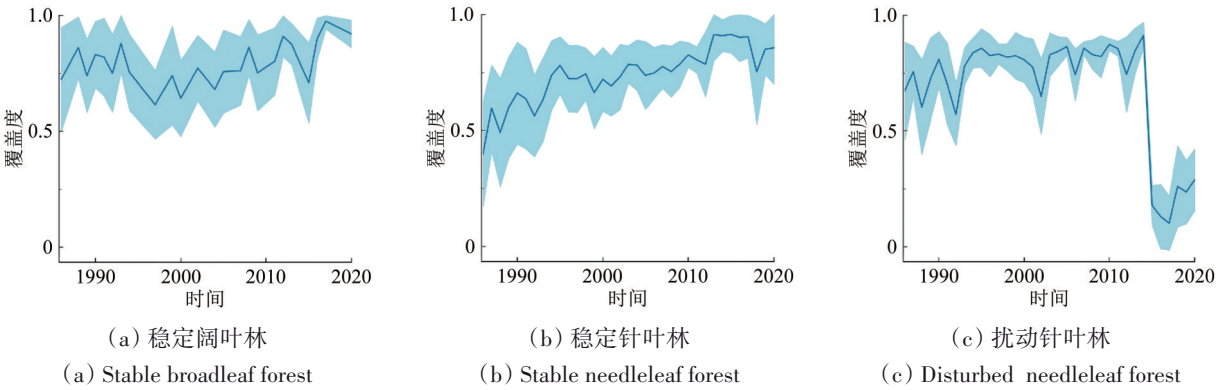


图9 不同森林类型覆盖度随时间变化
Fig. 9 Change of forest cover for different forest type

5 结 论

本研究以孟家岗林场为研究区，分析了基于人工合成样本和随机森林回归模型在森林类型和树种覆盖度反演问题中的可行性，并定量评估了不同影像获取时间频率（单时相、多时相、时序光谱）对于森林类型和树种覆盖度反演精度的影响。主要结论如下：

（1）对于森林类型覆盖度的反演，最优的估算模型为基于生长季 Landsat 波段和植被指数中值所建立的回归模型，阔叶林覆盖度反演模型的 R^2 为 0.76，RMSE 为 0.22；针叶林覆盖度反演模型的 R^2 为 0.71，RMSE 为 0.23。

（2）对于森林树种覆盖度的反演，最优的估算模型为基于多时相 Landsat 波段和植被指数所建立的回归模型，落叶松覆盖度反演模型的 R^2 为 0.40，RMSE 为 0.29；红松覆盖度反演模型的 R^2 为 0.23，RMSE 为 0.29；樟子松覆盖度反演模型的 R^2 为 0.61，RMSE 为 0.23。

（3）增加影像获取的时间密度，针叶林各树种覆盖度模型的 R^2 有明显提高，增加范围为 0.04—0.27 之间。阔叶林和针叶林覆盖度的反演精度无明显提高。

本研究对孟家岗林场森林类型（阔叶林和针叶林）的覆盖度进行了高精度的估算。基于生长季 Landsat 波段和植被指数中值所建立的随机森林回归模型可实现对中国寒温带森林类型覆盖度的

较高精度反演，该结果为更大区域的寒温带森林类型覆盖度反演提供了科学依据和方法支持。后续研究可以对黑龙江以及中国寒温带森林生态系统的阔叶林和针叶林覆盖度进行估算，为中国东北地区森林碳储量估算等研究提供基础数据。此外，为进一步提高覆盖度模型的反演精度，在后续研究中可引入多源数据，例如星载雷达数据。结合森林的光谱信息和结构信息，有望实现森林类型覆盖度反演精度的提高。

1986 年—2020 年孟家岗林场长时序阔叶林和针叶林覆盖度估算结果表明：亚像元尺度上的森林类型覆盖度估算更能捕捉到森林的生长和扰动情况。该模型在长时序上的成功应用，为中国北方寒温带森林类型的长时序覆盖度反演研究提供了思路。

本研究在预测针叶林树种覆盖度时，使用多时相多光谱特征效果最好，然而对于树种覆盖度的估算精度依旧较低。由于各类针叶林树种之间的光谱区分度低，仅使用多光谱特征很难将其区分开来。因此，在后续研究中，引入高光谱卫星（例如高分五号卫星），提取多时相高光谱特征，有望进一步提高树种覆盖度估算精度。

参考文献(References)

Adams J B, Sabol D E, Kapos V, Filho R A, Roberts D A, Smith M O and Gillespie A R. 1995. Classification of multispectral images based on fractions of endmembers: application to land-cover

- change in the Brazilian Amazon. *Remote Sensing of Environment*, 52(2): 137-154 [DOI: 10.1016/0034-4257(94)00098-8]
- Bonan G B, Pollard D and Thompson S L. 1992. Effects of boreal forest vegetation on global climate. *Nature*, 359(6397): 716-718 [DOI: 10.1038/359716a0]
- Breiman L. 2001. Random forests. *Machine learning*, 45 (2001): 5-32 [DOI: 10.1023/A:1010933404324]
- Brinck K, Fischer R, Groeneveld J, Lehmann S, Dantas De Paula M, Putz S, Sexton J O, Song D X and Huth A. 2017. High resolution analysis of tropical forest fragmentation and its impact on the global carbon cycle. *Nature Communications*, 8: 14855 [DOI: 10.1038/ncomms14855]
- Chen J, Chen Y H, He C Y and Shi P J. 2001. Sub-pixel model for vegetation fraction estimation based on land cover classification. *Journal of Remote Sensing (in Chinese)*, 5(6): 416-422 (陈晋, 陈云浩, 何春阳, 史培军. 2001. 基于土地覆盖分类的植被覆盖率估算亚像元模型与应用. *遥感学报*, 5(6): 416-422) [DOI: 10.11834/jrs.20010603]
- Chen J, Ma L, Chen X H and Rao Y H. 2016. Research progress of spectral mixture analysis. *Journal of Remote Sensing (in Chinese)*, 20(5): 1102-1109 (陈晋, 马磊, 陈学泓, 饶玉晗. 2016. 混合像元分解技术及其进展. *遥感学报*, 20(5): 1102-1109) [DOI: 10.11834/jrs.20166169]
- Chen S, Shahi C and Chen H H. 2016. Economic and ecological trade-off analysis of forest ecosystems: options for boreal forests. *Environmental Reviews*, 24(3): 348-361 [DOI: 10.1139/er-2015-0090]
- Deng C B and Zhu Z. 2020. Continuous subpixel monitoring of urban impervious surface using Landsat time series. *Remote Sensing of Environment*, 238: 110929 [DOI: 10.1016/j.rse.2018.10.011]
- Drever C R, Cook-Patton S C, Akhter F, Badiou P H, Chmura G L, Davidson S J, Desjardins R L, Dyk A, Fargione J E, Fellows M, Filewod B, Hessing-Lewis M, Jayasundara S, Keeton W S, Kroeger T, Lark T J, Le E, Leavitt S M, Leclerc M E, Lemprière T C, Metsaranta J, Mcconkey B, Neilson E, St-Laurent G P, Puric-Mladenovic D, Rodrigue S, Soolanayakanahally R Y, Spawn S A, Strack M, Smyth C, Thevathasan N, Voicu M, Williams C A, Woodbury P B, Worth D E, Xu Z, Yeo S and Kurz W A. 2021. Natural climate solutions for Canada. *Science Advances*, 7(23): eabd6034 [doi: 10.1126/sciadv.abd6034]
- FAO. 2020. Global Forest Resources Assessment 2020. Main Report. FAO
- Foga S, Scaramuzza P L, Guo S, Zhu Z, Dilley Jr R D, Beckmann T, Schmidt G L, Dwyer J L, Joseph Hughes M and Laue B. 2017. Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote Sensing of Environment*, 194: 379-390 [DOI: 10.1016/j.rse.2017.03.026]
- Girardin C A, Jenkins S, Seddon N, Allen M, Lewis S L, Wheeler C E, Griscom B W and Malhi Y. 2021. Nature-based solutions can help cool the planet—if we act now[J]. *Nature*, 593(7858): 191-194 [DOI: 10.1038/d41586-021-01241-2]
- Guerschman J P, Hill M J, Renzullo L J, Barrett D J, Marks A S and Botha E J. 2009. Estimating fractional cover of photosynthetic vegetation, non-photosynthetic vegetation and bare soil in the Australian tropical savanna region upscaling the EO-1 Hyperion and MODIS sensors. *Remote Sensing of Environment*, 113(5): 928-945 [DOI: 10.1016/j.rse.2009.01.006]
- Hansen M C, Potapov P V, Moore R, Hancher M, Turubanova S A, Tyukavina A, Thau D, Stehman S V, Goetz S J, Loveland T R, Kommareddy A, Egorov A, Chini L, Justice C O and Townshend J R G. 2013. High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160): 850-853 [DOI: 10.1126/science.1244693]
- Heilmayr R, Echeverría C and Lambin E F. 2020. Impacts of Chilean forest subsidies on forest cover, carbon and biodiversity. *Nature Sustainability*, 3(9): 701-709 [DOI: 10.1038/s41893-020-0547-0]
- Hemmerling J, Pflugmacher D and Hostert P. 2021. Mapping temperate forest tree species using dense Sentinel-2 time series. *Remote Sensing of Environment*, 267: 112743 [DOI: 10.1016/j.rse.2021.112743]
- Hu S J, Hu D Y and Zhao W J. 2010. Extract urban vegetation coverage based on LSMM and improved FCM: a case study in Haidian District. *Acta Ecologica Sinica*, 30(4): 1018-1024 (胡姝婧, 胡德勇, 赵文吉. 2010. 基于LSMM和改进的FCM提取城市植被覆盖度——以北京市海淀区为例. *生态学报*, 30(4): 1018-1024)
- Huang C and Townshend J R G. 2003. A stepwise regression tree for nonlinear approximation: applications to estimating subpixel land cover. *International Journal of Remote Sensing*, 24(1): 75-90 [DOI: 10.1080/01431160305001]
- Ji C C, Jia Y H, Li X S and Wang J Y. 2016. Research on linear and nonlinear spectral mixture models for estimating vegetation fractional cover of nitraria bushes. *Journal of Remote Sensing (in Chinese)*, 20(6): 1402-1412 (姬翠翠, 贾永红, 李晓松, 王金英. 2016. 线性/非线性光谱混合模型估算白刺灌丛植被覆盖度. *遥感学报*, 20(6): 1402-1412) [DOI: 10.11834/jrs.20166020]
- Jia W and Pang Y. 2023. Tree species classification in an extensive forest area using airborne hyperspectral data under varying light conditions. *Journal of Forestry Research*, 34(5): 1359-1377 [DOI: 10.1007/s11676-022-01593-z]
- Jiang H, Apps M J, Peng C H, Zhang Y L and Liu J X. 2002. Modeling the influence of harvesting on Chinese boreal forest carbon dynamics. *Forest Ecology and Management*, 169(1/2): 65-82 [DOI: 10.1016/S0378-1127(02)00299-2]
- Kattenborn T, Eichel J, Wiser S, Burrows L, Fassnacht F E, Schmidtlein S. 2020. Convolutional Neural Networks accurately predict cover fractions of plant species and communities in Unmanned Aerial Vehicle imagery. *Remote Sensing in Ecology and Conservation*, 6(4): 472-486 [DOI: 10.1002/rse2.146]
- Lei G B, Li A N, Tan J B, Zhang Z J, Bian J H, Jin H A, Zhao W and Cao X M. 2016. Forest types mapping in mountainous area using multi-source and multi-temporal satellite images and decision tree models. *Remote Sensing Technology and Application*, 31(1): 31-41 (雷光斌, 李爱农, 谭剑波, 张正健, 边金虎, 靳华安, 赵伟, 曹小敏. 2016. 基于多源多时相遥感影像的山地森林分类决策树模型研究. *遥感技术与应用*, 31(1): 31-41) [DOI: 10.11873/j.issn.1004-0323.2016.1.0031]

- Li Q S, Wong F K K and Fung T. 2021. Mapping multi-layered mangroves from multispectral, hyperspectral, and LiDAR data. *Remote Sensing of Environment*, 258: 112403 [DOI: 10.1016/j.rse.2021.112403]
- Li Y, Wang H and Li X B. 2015. Fractional vegetation cover estimation based on an improved selective endmember spectral mixture model. *PLoS One*, 10(4): e0124608 [DOI: 10.1371/journal.pone.0124608]
- Liu Y Y, Zeng P, Zhang R, Sun F Y and Che Y. 2021. Vegetation coverage change of the demonstration area of ecologically friendly development in the Yangtze River Delta, China based on GEE and BRT during 1984-2019. *Chinese Journal of Applied Ecology*, 32(3): 1033-1044 (刘焱焱, 曾鹏, 张然, 孙凤云, 车越. 2021. 基于GEE和BRT的1984年—2019年长三角生态绿色一体化发展示范区植被覆盖度变化. *应用生态学报*, 32(3): 1033-1044) [DOI: 10.13287/j.1001-9332.202103.011]
- Long S, Guo Z F, Xu L, Zhou H Z, Fang W H and Xu Y J. 2020. Spatiotemporal variations of fractional vegetation coverage in China based on Google Earth Engine. *Remote Sensing Technology and Application*, 35(2): 326-334 (龙爽, 郭正飞, 徐粒, 周华真, 方伟华, 许映军. 2020. 基于Google Earth Engine的中国植被覆盖度时空变化特征分析. *遥感技术与应用*, 35(2): 326-334) [doi: 10.11873/j.issn.1004-0323.2020.2.0326]
- Michishita R, Jiang Z B, Gong P and Xu B. 2012. Bi-scale analysis of multitemporal land cover fractions for wetland vegetation mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 72: 1-15 [DOI: 10.1016/j.isprsjprs.2012.04.006]
- Morsdorf F, Kötz B, Meier E, Itten K I and Allgöwer B. 2006. Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction. *Remote Sensing of Environment*, 104(1): 50-61 [DOI: 10.1016/j.rse.2006.04.019]
- Okujeni A, Jänicke C, Cooper S, Frantz D, Hostert P, Clark M, Segl K and van der Linden S. 2021. Multi-season unmixing of vegetation class fractions across diverse Californian ecoregions using simulated spaceborne imaging spectroscopy data. *Remote Sensing of Environment*, 264: 112558 [DOI: 10.1016/j.rse.2021.112558]
- Okujeni A, van der Linden S and Hostert P. 2015. Extending the vegetation-impervious-soil model using simulated EnMAP data and machine learning. *Remote Sensing of Environment*, 158: 69-80 [DOI: 10.1016/j.rse.2014.11.009]
- Okujeni A, van der Linden S, Jakimow B, Rabe A, Verrelst J and Hostert P. 2014. A comparison of advanced regression algorithms for quantifying urban land cover. *Remote Sensing*, 6(7): 6324-6346 [DOI: 10.3390/rs6076324]
- Okujeni A, van der Linden S, Suess S and Hostert P. 2017. Ensemble learning from synthetically mixed training data for quantifying urban land cover with support vector regression. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(4): 1640-1650 [DOI: 10.1109/JSTARS.2016.2634859]
- Okujeni A, van der Linden S, Tits L, Somers B and Hostert P. 2013. Support vector regression and synthetically mixed training data for quantifying urban land cover. *Remote Sensing of Environment*, 137: 184-197 [DOI: 10.1016/j.rse.2013.06.007]
- Pasquarella V J, Holden C E and Woodcock C E. 2018. Improved mapping of forest type using spectral-temporal Landsat features. *Remote Sensing of Environment*, 210: 193-207 [DOI: 10.1016/j.rse.2018.02.064]
- Peng S Z, Ding Y X, Liu W Z and Li Z. 2019. 1 km monthly temperature and precipitation dataset for China from 1901 to 2017. *Earth System Science Data*, 11(4): 1931-1946 [DOI: 10.5194/essd-11-1931-2019]
- Persson M, Lindberg E and Reese H. 2018. Tree species classification with multi-temporal Sentinel-2 data. *Remote Sensing*, 10(11): 1794 [DOI: 10.3390/rs10111794]
- Pi X Y, Zeng Y N and He C Q. 2021. High-resolution urban vegetation coverage estimation based on multi-source remote sensing data fusion. *Journal of Remote Sensing (in Chinese)*, 25(6): 1216-1226 (皮新宇, 曾永年, 贺城墙. 2021. 融合多源遥感数据的高分辨率城市植被覆盖度估算. *遥感学报*, 25(6): 1216-1226) [DOI: 10.11834/jrs.20219178]
- Powell R L, Roberts D A, Dennison P E and Hess L L. 2007. Sub-pixel mapping of urban land cover using multiple endmember spectral mixture analysis: Manaus, Brazil. *Remote Sensing of Environment*, 106(2): 253-267 [DOI: 10.1016/j.rse.2006.09.005]
- Qian X J, Zhang Y J, Liu L Y and Du S S. 2019. Exploring the potential of leaf reflectance spectra for retrieving the leaf maximum carboxylation rate. *International Journal of Remote Sensing*, 40(14): 5411-5428 [DOI: 10.1080/01431161.2019.1579940]
- Savage S L, Lawrence R L and Squires J R. 2015. Predicting relative species composition within mixed conifer forest pixels using zero-inflated models and Landsat imagery. *Remote Sensing of Environment*, 171: 326-336 [DOI: 10.1016/j.rse.2015.10.013]
- Schug F, Frantz D, Okujeni A, van der Linden S and Hostert P. 2020. Mapping urban-rural gradients of settlements and vegetation at national scale using Sentinel-2 spectral-temporal metrics and regression-based unmixing with synthetic training data. *Remote Sensing of Environment*, 246: 111810 [DOI: 10.1016/j.rse.2020.111810]
- Senf C, Laštovička J, Okujeni A, Heurich M and van der Linden S. 2020. A generalized regression-based unmixing model for mapping forest cover fractions throughout three decades of Landsat data. *Remote Sensing of Environment*, 240: 111691 [DOI: 10.1016/j.rse.2020.111691]
- Srisa-An C. 2021. Guideline of collinearity - avoidable regression models on time-series analysis//Proceedings of the 2021 2nd International Conference on Big Data Analytics and Practices. Bangkok, Thailand: IEEE: 28-32 [DOI: 10.1109/IBDAP52511.2021.9552165]
- Venäläinen A, Lehtonen I, Laapas M, Ruosteenoja K, Tikkanen O P, Viiri H, Ikonen V P and Peltola H. 2020. Climate change induces multiple risks to boreal forests and forestry in Finland: a literature review. *Global Change Biology*, 26(8): 4178-4196 [DOI: 10.1111/gcb.15183]
- Wang G Z, Wang J P, Han L, Chai G Q and Wang Z L. 2018. Estimating fractional cover of non-photosynthetic vegetation using field spectral to simulate Landsat-8 OLI. *Journal of Geo-Information Science*, 20(11): 1667-1678 (王光镇, 王静璞, 韩柳, 柴国奇, 王

- 周龙. 2018. 基于实测光谱模拟 Landsat-8 OLI 数据估算非光合植被覆盖度. 地球信息科学学报, 20(11): 1667-1678 [DOI: 10.12082/dqxxkx.2018.180196]
- Wang J, Xiao X M, Liu L, Wu X C, Qin Y W, Steiner J L and Dong J W. 2020. Mapping sugarcane plantation dynamics in Guangxi, China, by time series Sentinel-1, Sentinel-2 and Landsat images. *Remote Sensing of Environment*, 247: 111951 [DOI: 10.1016/j.rse.2020.111951]
- Wang M Y, Zheng Y, Huang C Q, Meng R, Pang Y, Jia W, Zhou J, Huang Z H, Fang L C and Zhao F. 2022. Assessing Landsat-8 and Sentinel-2 spectral-temporal features for mapping tree species of northern plantation forests in Heilongjiang Province, China. *Forest Ecosystems*, 9: 100032 [DOI: 10.1016/j.fecs.2022.100032]
- Wang S, Zhang L L, Lin W B, Huang Q S, Song Y X and Ye M. 2022. Study on vegetation coverage and land-use change of Guangdong Province based on MODIS-NDVI. *Acta Ecologica Sinica*, 42(6): 2149-2163 (王思, 张路路, 林伟彪, 黄秋森, 宋亦心, 叶脉. 2022. 基于 MODIS-归一化植被指数的广东省植被覆盖与土地利用变化研究. *生态学报*, 42(6): 2149-2163) [DOI: 10.5846/stxb202104261100]
- Wang X and Zhang Y W. 2021. An analysis of change trend of fractional vegetation cover in Beijing based on Landsat imagery. *Remote Sensing Technology and Application*, 36(6): 1388-1397 (王曦, 张怡雯. 2021. 基于 Landsat 影像的北京植被覆盖度变化趋势分析. *遥感技术与应用*, 36(6): 1388-1397) [DOI: 10.11873/j.issn.1004-0323.2021.6.1388]
- Wu Z J, He G J, Wang M M, Fu J F and Zou D. 2016. Estimation of fractional vegetation cover and its spatial-temporal variation in hilly area of Southeastern China: Yongding County, Fujian Province. *Remote Sensing Technology and Application*, 31(6): 1201-1208 (吴志杰, 何国金, 王猛猛, 傅娇凤, 邹丹. 2016. 南方丘陵区植被覆盖度遥感估算与时空变化研究——以福建省永定县为例. *遥感技术与应用*, 31(6): 1201-1208) [DOI: 10.11873/j.issn.1004-0323.2016.6.1201]
- Xiao J F and Moody A. 2005. A comparison of methods for estimating fractional green vegetation cover within a desert-to-upland transition zone in central New Mexico, USA. *Remote Sensing of Environment*, 98(2/3): 237-250 [DOI: 10.1016/j.rse.2005.07.011]
- Zhang M, Zhang X, Zhao D, Huang C P and Zhang L F. 2012. An approach to generate synthetic hyperspectral data used for evaluating linear spectral unmixing algorithms. *Remote Sensing Technology and Application*, 27(5): 680-685 (张明, 张霞, 赵冬, 黄长平, 张立福. 2012. 一种用于线性光谱解混算法验证的模拟数据生成方法. *遥感技术与应用*, 27(5): 680-685) [DOI: 10.11873/j.issn.1004-0323.2012.5.680]
- Zhang P C, Shao G F, Zhao G, Le Master D C, Parker G R, Dunning Jr J B and Li Q L. 2000. China's forest policy for the 21st century. *Science*, 288(5474): 2135-2136 [DOI: 10.1126/science.288.5474.2135]
- Zhao Y J, Zeng Y, Zheng Z J, Dong W X, Zhao D, Wu B F and Zhao Q J. 2018. Forest species diversity mapping using airborne LiDAR and hyperspectral data in a subtropical forest in China. *Remote Sensing of Environment*, 213: 104-114 [DOI: 10.1016/j.rse.2018.05.014]

Mapping forest type and tree species fractions in China's cold-temperate forests based on synthetically mixed data and random forest regression

WANG Mengyu^{1,2}, ZHAO Feng², PANG Yong^{3,4}, MENG Ran⁵, JIA Wen^{3,4}, YUE Chao^{1,6}

1. College of Natural Resources and Environment, Northwest A&F University, Yangling 712100, China;

2. College of Urban and Environmental Science, Central China Normal University, Wuhan 430079, China;

3. Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Beijing 100091, China;

4. Key Laboratory of Forestry Remote Sensing and Information Technology, National Forestry and Grassland Administration, Beijing 100091, China;

5. College of Resources & Environment, Huazhong Agricultural University, Wuhan 430070, China;

6. Institute of Soil and Water Conservation, Chinese Academy of Sciences and Ministry of Water Resources, Yangling 712100, China

Abstract: Cold-temperate forests, recognized as the most extensive terrestrial ecosystems, cover vast areas around the globe and hold important ecological and social values. Accurate mapping of forest type and tree species cover fraction in these forests across space and time is crucial for quantifying ecosystem services and formulating effective forest management policies to ensure their sustainable conservation. However, despite the increasing development of remote sensing technologies, studies exploring the feasibility of inverting forest type and tree species cover fraction using medium-resolution multispectral satellite-based data, such as Landsat, in China's cold-temperate forests are limited. This limitation is primarily attributed to the scarcity of reference data and the restricted spectral information available in multispectral images. Moreover, the quantitative impact of the temporal frequency of data acquisitions (e.g., single-date, multirate) on mapping forest type and tree species cover fraction remains largely unexplored. The timing and frequency of satellite data acquisition can significantly influence the detection and characterization of dynamic changes in forests, which in turn affects the accuracy of mapping forest attributes. To address these gaps, our study aims to map the forest type and tree species cover fraction in Mengjiagang Forest, Heilongjiang

Province by employing synthetically mixed data and a random forest regression model. We extend our analysis to three decades (from 1986 to 2020) of Landsat data, mapping the cover fractions of broadleaf and needleleaf forests in Mengjiagang Forest by using an optimal broadleaf and needleleaf random forest regression model. The results of our study reveal the following key findings: (1) For forest type cover fraction inversion, the random forest regression model based on the growing season median index (including spectral bands, NDTI, and TCT indices) is the optimal model (achieving $\bar{R}^2=0.76$ for broadleaf and $\bar{R}^2=0.71$ for needleleaf). (2) For tree species cover fraction inversion, the random forest regression model based on multivariate spectral features (including the spectral bands, NDTI, and TCT indices of growth and leaf off seasons) is the optimal model (achieving $\bar{R}^2=0.40$ for Larch, $\bar{R}^2=0.23$ for Korean pine, and $\bar{R}^2=0.61$ for Mongolian pine). (3) Increasing the temporal frequency of data acquisition can enhance tree species cover fraction inversion accuracy (achieving $\Delta\bar{R}^2=0.04$ for Larch, $\Delta\bar{R}^2=0.07$ for Korean pine, and $\Delta\bar{R}^2=0.27$ for Mongolian pine), whereas its effect on improving forest type cover fraction inversion accuracy is limited. By effectively combining the advantages of synthetically trained data and random forest regression, we have successfully mapped the forest type and tree species cover fraction of Mengjiagang Forest. Moreover, our study provides a comprehensive analysis that accurately quantifies the influence of temporal data acquisition frequency on mapping forest type and tree species cover fraction. This study offers valuable insights into the future mapping of forest type and tree species cover fraction across space and time, particularly for regions with similar species composition. The outcomes of this research will make a significant contribution to the understanding and management of cold-temperate forests, thereby supporting their conservation and sustainable use.

Key words: remote sensing, forest type coverage, tree species coverage, synthetically training data, long time series, machine learning

Supported by National Natural Science Foundation of China (No. 41901382, 42101403); Open Fund of State Key Laboratory of Remote Sensing Science (No. OFSLRSS201917); National Key Research and Development Program of China (No. 2017YFD0600404)