Sentinel-2卫星落叶松林龄信息反演

唐少飞,田庆久,徐凯健,徐念旭,岳继博

- 1. 南京大学 国际地球系统科学研究所, 南京 210023; 2. 江苏省地理信息技术重点实验室, 南京 210023
- 摘 要:林龄结构信息能够有效反映区域森林群落不同生长阶段的固碳能力,对于评估森林生态系统的健康状况具有重要意义。本研究以中国温带典型优势树种落叶松林为研究对象,分别选择其芽萌动期、展叶期和落叶期时段的 Sentinel-2影像,采用多元线性回归(MLR)、随机森林(RF)、支持向量机回归(SVR)、前馈反向传播神经网络(BP)以及多元自适应回归样条(MARS)等 5 种方法依次构建落叶松林龄反演模型。通过相关性分析首先确定最佳遥感反演物候期,并在此基础上根据相关性差异筛选出 5 个最优特征变量用于模型反演,分别为冠层含水量(CWC),归一化水体指数(NDWI),叶面积指数(LAI),光合有效辐射吸收率(FAPAR)和植被覆盖度(FVC)。研究结果表明,展叶期为落叶松林最佳遥感反演物候期。除植被衰减指数(PSRI)以及落叶期的NDVI、RVI外,落叶松林龄与各指标之间均呈负相关关系,其中与冠层含水量(CWC)的相关性最高,pearson相关系数达到-0.74(p<0.01)。此外,不同模型反演结果表明,随机森林模型(RF)为最佳落叶松林龄估测模型,其平均决定系数 R^2 和平均均方根误差 RMSE 分别为 0.89 和 2.91 a;多元线性回归模型(MLR)的林龄估测结果最差,其平均决定系数 R^2 和平均均方根误差 RMSE 仅为 0.57 和 5.69 a,非线性模型能更好的解释林龄与建模变量之间的关系。

关键词:遥感, Sentinel-2, 落叶松, 林龄反演, 生物物理参数, 随机森林

引用格式: 唐少飞,田庆久,徐凯健,徐念旭,岳继博.2020.Sentinel-2卫星落叶松林龄信息反演.遥感学报,24(12): 1511-1524

Tang S F, Tian Q J, Xu K J, Xu N X and Yue J B. 2020. Age information retrieval of *Larix gmelinii* forest using Sentinel-2 data. Journal of Remote Sensing(Chinese), 24(12):1511-1524[DOI:10.11834/jrs.20208500]

1 引 言

森林通过吸收大气中的CO₂将碳固定于林分植被和土壤中,使整个森林生态系统碳储量约占到全球陆地生态系统总碳储量的33%—46%(Watson等,2000)。因此,森林植被对调节全球生态系统碳平衡时空分布格局和减缓全球气候变暖具有重要作用(Lal,2008; Dai等,2013)。林龄结构作为评价森林生态系统碳循环的关键因子,森林的碳交换强度具有明显的林龄效应(王兴昌等,2008)。林龄效应与表征森林碳汇潜力的指标密切相关,其中包括碳储量、净植被生产力(NPP)和森林蓄积量等(Haywood和Stone,2017;巨文珍等,2011; He等,2012; Yu等,2013)。以中国

乔木林碳储量为例,碳储量从幼龄林到中龄林阶段不断增长,在中龄林达到最大值后,又逐渐显现出不断下降的趋势(李奇等,2018)。林龄结构的动态变化在时间和空间上影响着森林生态系统的稳定性。在空间上,合理的林龄结构空间分布对于制定合适的森林经营措施,最大化提高森林生产率具有重要意义;在时间上,林龄结构能够有效反映森林干扰及恢复状况,是特别重要的森林参数(沈文娟和李明诗,2017; Fraser和Li,2002; Litvak等,2003)。

林分年龄指树木自种子萌发后生长的年数, 有两种表示方法:一种是林分中占优势部分树木 的平均年龄,另一种是全部林木的平均年龄(孟 宪宇,2006;江伟钰和陈方林,2005),本文使用

收稿日期: 2019-01-04; 预印本: 2019-03-13

基金项目: 国家重点研发计划(编号:2017YFD0600903);国家自然科学基金(编号:41771370);国家科技重大专项(编号:03-Y20A04-9001-17/18,30-Y20A07-9003-17/18);民用航天技术预先研究项目(编号:Y7K00100KJ)

第一作者简介: 唐少飞,1994年生,男,硕士研究生,研究方向为区域森林信息遥感定量反演。E-mail: tangsf1120@163.com 通信作者简介: 田庆久,1964年生,男,教授,研究方向为高光谱遥感与遥感信息定量化。E-mail: tianqj@nju.edu.cn

的是优势树种的平均年龄。目前, 林龄信息的获 取手段主要包括森林样地清查和遥感监测。森林 样地清查主要是通过历史记录资料或者利用生长 锥等方法直接获取林龄信息(吴斡宁和万涛等, 2013)。其中,生长锥取样受到成本高、破坏树木 以及调查周期长等条件限制, 当前研究普遍采用 间接方式获取,即根据较易测量的林分立地指标 (胸径、树高等)来建立林龄生长模型。例如,周 凤艳(2017)利用胸径,枝下高等指标建立章古 台地区樟子松的林龄生长模型;李凡等(2018) 结合树高属性构建帽儿山针叶混交林、阔叶混交 林、针阔混交林的林龄的一元线性模型。二者研 究结果表明立地指标与林龄呈极显著相关关系 (R>0.85; p<0.01), 同时模型反演精度能够满足林 龄实际需求。但是,受实测数据空间分布代表性 以及样地选择客观性等限制,森林样地清查方法 难以用于区域尺度上的信息监测, 而遥感数据具 备较高的空间异质性与时效性特征,这在估算森 林林龄的空间格局具有不可替代的优势(戴铭等, 2011)。目前,遥感反演林龄基本原理是:植物在 物候周期中叶片叶绿素以及内部结构的动态变化 会对其光谱响应模式产生重要影响, 其中成熟林 冠层叶片往往会出现更高的近红外与红光反射率 对比度特征 (Grant, 1987; Knipling, 1970)。基 于此, 现有的遥感林龄模型建模参量多考虑红波 段,近红外波段以及由二者构成的植被指数,在 反演模型方面,大多数的研究者采用人工神经网 络和多元线性回归模型进行林龄模型反演研究。 如Jensen等(1999)使用TM数据的单波段(Red、 NIR和SWIR) 以及植被指数(NDVI和RVI) 构建 了火炬松多元线性回归和人工神经网络林龄反演 模型,结果证实近红外(NIR)与林龄信息具有较 强的相关性 (R=0.89),同时非线性模型能更好的 解释林龄与模型参数之间的关系。Sivanpillai等 (2006) 使用和Jensen等(1999) 研究中相同的数 据源和方法,并在模型参量上增加缨帽变换的3个 参量,包括Brightness,Greenness和Wetness,结 果表明采用植被指数 (R=0.88) 相对于单波段 (R=0.83) 具有更高的反演精度。Chen等 (2012) 在Jensen 以及Sivanpillai 等研究的基础上,进一步 结合人工橡胶林的时相信息采用相同方法建立林 龄反演模型,结果表明在不同时相下橡胶林林龄 与光谱信息表现存在差异性, 其中近红外波段与 林龄在 10 月份的相关性 (R=-0.40) 远高于 6 月份的相关性 (R=-0.13)。此外,对比多项研究结果发现,非线性林龄模型精度优于线性模型 (Wulder等,2004;Champion等,2008)。

截至目前,相关区域林龄遥感反演相关研究 中所涉及的优势树种多局限于热带或亚热带地区 的橡胶林或火炬松等,而针对温带地区典型树种 (如落叶松等)的林龄反演研究十分缺乏;同时, 相关林龄研究在模型参量选择上,基本选择传感 器单波段(如红波段,近红外波段等)以及几种 常见的植被指数(如NDVI, RVI等),对于植被生 物物理参数等相关特征缺乏考量;此外,现有研 究结果未考虑不同物候条件下林龄光谱差异性引 起的林龄反演物候期问题, 使研究结果存在不确 定性。基于此,本文以中国北方典型优势树种落 叶松为研究对象,采用6景不同物候期的Sentinel-2 影像为数据源,在2类传统遥感反演方法(多元线 性回归模型和人工神经网络模型)的基础上增加 3类机器学习算法(多元自适应回归样条模型、随 机森林模型、支持向量机回归模型) 构建落叶松 林龄反演模型,同时在林龄反演模型构造参量上 引入生物物理参数,以深入探讨影响区域林龄反 演的主要因素,并优化反演模型与结果精度,为 相关研究提供参考和依据。

2 研究区与数据

2.1 研究区概况

研究区位于内蒙古自治区赤峰市喀喇沁旗旺业甸林场,地理位置为118°09′E—118°30′E,41°21′N—41°39′N(图1)。研究区占地面积551 km²,其中森林覆盖面积约为463 km²。研究区最高海拔达1890 m,最低海拔为800 m,平均海拔在1150 m左右。地形以山地为主,地势东低西高,天然形成3种不同的地貌类型(中北部低山丘陵、西南部中山以及东南部河谷平川)。主要优势树种类型为落叶松(Larix gmelinii)、油松(Pinus tabuliformis)、白桦(Betula platyphylla)、山杨(Populus davidiana)、蒙古栎(Quercus mongolica)等(徐凯健等,2018)。其中,落叶松主要分布于海拔300—1200 m地带,喜光性强,对水分要求较高,在各种不同环境条件下(如湿润而土壤富腐殖质的阴坡及干燥的阳坡、湿润的河谷及山顶等)均能较好生长。

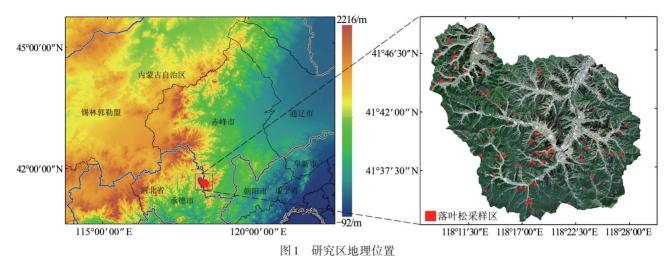


Fig.1 Location of study area

2.2 数据与处理

2.2.1 Sentinel-2数据获取

Sentinel-2环境监测卫星是欧盟委员会(EC) 和欧洲航天局(ESA)共同倡议的全球环境与安全 监测系统"哥白尼计划"中的第二颗卫星,数据 可通过ESA的数据共享网站进行下载(https:// scihub. copernicus. eu/dhus/#/home [2019-01-04]) 。 根据中国东北地区4种落叶松物候信息显示,落 叶松的芽萌动期为3月下旬—4月下旬,展叶期为 5月上旬-8月上旬,落叶期为9月中旬-10月下 旬(张志伟, 2013)。因此,选择研究区6景云量 较少、质量较高的L1C级别遥感影像,获取时间 分别为 2018-04-25, 2017-06-14 和 2017-09-22。 获取的Sentinel-2 L1C级数据已经过几何重采样、 地理配准、辐射定标和大气表观反射率TOA (Top-Of-Atmosphere) 计算。为了降低大气影响, 使用由 Python 语言编写的 Sen2cor 模型进一步将 L1C级的Sentinel-2数据处理成L2A级产品,同时 使用SNAP软件将L2A级数据重采样为10m,以保 证 L2A 级数据各波段空间分辨率一致(Louis, 2016; Main-Knorn 等, 2017)。

本文的模型参数包括植被指数和生物物理参数,具体计算方法见表1。其中,植被指数包括叶绿素红边指数(CIre)、植被衰减指数(PSRI)、归一化水体指数(NDWI)、归一化植被指数(NDVI)和比值植被指数(RVI);生物物理参数包括:叶面积指数(LAI)、光合有效辐射吸收率(FAPAR)、叶绿素含量(CCL)、冠层含水量

(CWC) 和植被覆盖度 (FVC)。上述参数能够全面、定量描述植被生长状态信息 (刘振元等,2017; 杨斌等,2017)。其中,生物物理参数是根据 L2A 级数据和 SNAP 软件中的生物物理量处理器模块进行提取,其原理是利用 PROSPECT+SAIL 辐射传输模型对研究区内植被特性进行数据训练,然后采用神经网络模型分析并计算每个栅格像元的冠层特征,并从宏观区域上最大程度精细计算出各像元植被特征与冠层表观发射率之间的关系,最后通过模型自验证功能保证最佳的遥感反演模拟精度 (Muller-Wilm等,2016; Weiss和 Baret,2016)。

2.2.2 林龄结构数据

落叶松林龄信息来源于2017-09和2018-06两 次野外实地调查,以及2017年旺业甸林场林业小 班数据。后者包含的相关属性信息有优势树种, 平均林龄,坡向,平均胸径等。为了进一步保证 数据的准确性,本研究从面积大于1 km²且郁闭度 大于0.7的林业班块中选择10×10像元的矩形光谱 信息均值作为林龄样本数据, 此样本选取方法能 有效避免植被边界地物(道路等)以及林下背景 对于冠层反射率的影响。最终,筛选出满足条件 的落叶松建模样本数量共71个,林龄跨度包括幼 龄林、中龄林、近熟林和成熟林, 划分依据为 2014年的《国家森林资源连续清查主要技术规 定》。将整个样本数据按7:3比例随机分成训练样 本(50)和验证样本(21),具体信息见表2。进 行分析前,对所有数据进行基于最大最小值的归 一化处理,以消除指标之间的量纲影响。

编号	建模参数	计算方法	参考文献
1	叶绿素红边指数(CIre)	$ ho_{ m red_edge3}/ ho_{ m red_edge1}$ – 1	Gitelson等(2003)
2	植被衰减指数(PSRI)	$\left(ho_{ m red}^{} - ho_{ m green}^{} ight)\!\!\left/ ho_{ m red_edge2}^{}$	Merzlyak 等(1999)
3	归一化水体指数(NDWI)	$\left(ho_{ m nir} - ho_{ m swri2}\right) / \left(ho_{ m nir} + ho_{ m swri2}\right)$	Gao(1996)
4	归一化植被指数(NDVI)	$\left(ho_{ m nir} - ho_{ m red} ight) \! / \! \left(ho_{ m nir} + ho_{ m red} ight)$	Tucker(1979)
5	比值植被指数(RVI)	$ ho_{ m nir}/ ho_{ m red}$	Birth和McVey(1968)
6	叶面积指数(LAI)	PROSPECT+SAIL辐射传输模型	Weiss和Baret(2016)
7	光合有效辐射吸收率(FAPAR)	PROSPECT+SAIL辐射传输模型	Weiss和Baret(2016)
8	叶绿素含量(CCL)	PROSPECT+SAIL辐射传输模型	Weiss和Baret(2016)
9	冠层含水量(CWC)	PROSPECT+SAIL辐射传输模型	Weiss和Baret(2016)
10	植被覆盖度(FVC)	PROSPECT+SAIL辐射传输模型	Weiss和Baret(2016)

表 1 林分年龄预测模型参数

Table 1 Parameter variables used to develop stand age prediction models

表 2 落叶松林龄信息统计

Table 2 Summary statistics for Larix gmelinii stand age

编号	龄组	林龄范围 / a	数量
1	幼龄林	<20	9
2	中龄林	21—30	15
3	近熟林	31—40	28
4	成熟林	41—60	19

3 研究方法

3.1 相关性分析

相关性分析是指对两个或多个具备相关性的变量元素进行分析,从而衡量两个变量因子的相关程度,本文采用皮尔逊相关系数(Pearson correlation coefficient)分析林龄与建模指标之间的相关程度,其值一般介于-1与1之间(Eggers等,2003)。

3.2 模型介绍与实现

3.2.1 多元线性回归模型 MLR

多元线性回归模型是用两个或两个以上的变量解释因变量的一种模型。本研究以落叶松林龄为因变量和所筛选的建模指标为自变量建立多元线性回归模型,并采用R语言中的lm包实现多元线性回归模型。

3.2.2 随机森林模型 RF

随机森林 RF(RandomForest)是由 Cutler 等

(2011) 提出的一种基于分类与回归决策树 CART (Classification and Regression Tree) 的组合算法,随机森林实现原理是在构建过程中采用了自助法 (bootstrap) 重采样技术,它是从原始训练样本集 N中有放回地重复随机抽取 k个样本生成新的训练样本集合,然后根据自助样本集生成 k个分类树组成随机森林,新数据的分类结果按分类树多少形成的分数决定(Cutler等,2011)。随机森林模型实现最重要的参数包括 ntree 和 mtry,ntree 指随机森林中所包含的决策树数目,mtry表示节点中用于二叉树的变量个数,默认情况下为数据集变量个数的1/3(李欣海,2013),具体实现采用R语言中的 Random Forest 包。

3.2.3 前馈反向传播神经网络模型BP

神经网络是一种模仿生物神经网络的结构和功能的数学模型,由大量的节点(或称"神经元")之间相互联接构成。每个节点代表一种特定的输出函数,称为激励函数。每两个节点间的连接都代表一个对于通过该连接信号的加权值,称为权重。网络的输出则根据网络的连接方式,权重值和激励函数的不同而不同。BP(Back Propagation)网络是1986年由Rumelhart和McCelland为首的科学家小组提出,是一种按误差逆传播算法训练的多层前馈网络,是目前应用最广泛的神经网络模型之一,它由信息的正向传播和误差的反向传播两个过程组成,是一个包含输入层、隐藏层和输出层的三层网络(焦李成,1990)。本研究利用

R语言中的nnet包实现BP神经网络算法,它只有 一层隐藏层, 训练方法使用包含动量的最速下降 法, 传递函数使用 sigmoid 函数。模型运行参数包 括: 隐层结点数 (size), 权值衰减参数 (decay), 初始随机数权值 (rang), 最大迭代次数 (maxit), 线性输出单元开关 (linout)。

3.2.4 支持向量回归模型 SVR

支持向量机 SVM(Support Vector Machine)是 建立在统计学习理论的VC维理论和结构风险最小 原理(SRM)基础上的,根据有限的样本信息在 模型的复杂性和学习能力之间寻求最佳折衷,以 期获得最好的推广能力(刘莉, 2005)。支持向量 回归SVR是由SVM发展出来的回归方法,与SVM 的区别在于回归所求超平面是使所有样本点到超 平面的距离为最小。对于非线性回归问题,可通 过核函数变换将样本映射到一个高维特征空间中 用线性回归来解决(Gunn, 1998)。支持向量机回 归方法能较好的解决小样本、过拟合和非线性等 问题, 泛化推广能力优异, 这些优势使得支持向 量机在遥感方面得到广泛应用。例如,黎锐等 (2009) 利用支持向量回归的方法估测了不同时相 下冬小麦的产量, 研究结果显示该方法能够保证 预测结果的稳健性,有效提高模型的估测精度。 本研究利用R语言中的e1071包实现支持向量回归 模型,其中运行的核函数选择默认的高斯核函数, 其中cost和gamma参数根据e1071包中超参数调优 的tune函数选择最优参数(Tauxe等, 2013)。

3.2.5 多元自适应回归样条模型 MARS

多元自适应回归样条 MARS (MultIvariate 首次提出,是一种泛化能力很强的专门针对高维 数据的非线性回归方法。算法原理是利用样条函 数处理复杂的非线性变量关系,并将整个非线性 模型分为若干个区域, 在这些区域内采用一段线 性回归模型进行拟合。MARS模型构建包括前向过 程、后向剪枝和模型选择3个过程,其中前向过程 对输入的样本数据进行划分,每个划分的点就是 结点,对每个划分的小区间用样条函数拟合得到 新的基函数,产生一个过拟合的模型,它由数据 自动完成,不需要人为设定;后向剪枝过程就是 去掉多余基函数的过程,同时保障模型的精确度, 它是基于广义交叉验证 (GCV) 标准进行, 当 GCV 达到最小时,对应的预测模型为最佳模型 (陈立宇, 2015)。MARS整个算法运算快捷且模 型具有较好的解释能力,该方法已应用于遥感领 域包括道路信息提取 (Miao 等, 2013; Liu 等, 2016) 以及估测生物量等方面(Filippi等, 2014; Güneralp等, 2014)。具体实现采用R语言中的 earth包。

3.3 模型精度评价

采用决定系数 (R2) 和均方根误差 (RMSE) 反映模型的准确性和稳定性。计算公式如下:

$$R^{2} = 1 - \frac{\sum_{i=1}^{n} (\widehat{y_{i}} - y_{i})^{2}}{\sum_{i=1}^{n} (y_{i} - \overline{y})^{2}}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_{i} - \widehat{y_{i}})^{2}}{n}}$$
(2)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \widehat{y_i})^2}{n}}$$
 (2)

式中, γ ,代表第i个落叶松的林龄实际值, $\hat{\gamma}$,代表 第i个落叶松的林龄预测值, \bar{v} 代表落叶松林龄的 平均值, n代表落叶松样本数量。

4 结果与分析

4.1 不同物候期参量相关性分析

不同物候期落叶松林龄与建模指标之间的相 关性结果表明(表3),落叶松林龄与所选建模指 标基本呈负相关关系,其中,林龄与红波段、近 红外波段和植被指数呈显著负相关关系,这与多 数学者的相关研究结果一致 (Jensen 等, 1999; Chen等, 2012; Fiorella和Ripple, 1993)。例如, Jensen等(1999)研究表明火炬松林龄与NDVI (R=-0.62; p=0.001) 和 RVI (R=-0.68; p=0.001)呈显著负相关关系, Fiorella 和 Ripple (1993) 研 究结果表明道格拉斯冷杉林龄与TM数据的红波段 呈显著负相关关系 (R=-0.86; p<0.01)。然而,各 物候期下的PSRI和落叶期的NDVI、RVI与林龄之 间呈正相关(数值微小)。其中,不同物候期PSRI 均与落叶松林龄呈正相关关系,从高到低依次为: 芽萌动期>展叶期>落叶期,这主要是因为PSRI常 用于检测森林的枯枝落叶层的状态特性, 在幼龄 林阶段, PSRI呈现低值(枝干少), 而成熟林枝干 多且枯叶多年积累, PSRI则呈现高值。因此, 在 整个物候期内落叶松林龄与PSRI呈正相关关系。

此外, 在不同物候期条件下, 生物物理参数

的平均相关系数绝对值|R|均高于植被指数,这表 明林龄与生物物理参数相关性高于植被指数,生 物物理参数更适用于林龄反演。在各个物候期,冠 层含水量(CWC)和归一化水体指数(NDWI)与 落叶松林龄均呈现出显著的负相关关系(p<0.05), 其中展叶期二者相关性达到最大值,分别为 0.74 (p<0.01) 和 0.67 (p<0.01), 表明水分指标与 林龄的相关性密切,后续研究可将水分指标作为 林龄反演切入点。综合各物候期中林龄与建模参 数相关性表现,可发现展叶期为最佳林龄反演物 候期。表4结果表明参量的共线问题比较突出,其 中各生物物理指标之间均呈显著正相关关系,例 如,叶面积指数(LAI)与光合有效辐射吸收率 (FAPAR) 和植被覆盖度(FVC) 相关性达到0.99, 0.96。但各指标从不同角度表征植被生长状态特 性,因此从展叶期中选择 pearson 相关系数大于 0.5 的参量作为模型的输入参数,最终构建模型的参 量为:冠层含水量(CWC),归一化水体指数(NDWI),叶面积指数(LAI),光合有效辐射吸收率(FAPAR)和植被覆盖度(FVC)。

表 3 落叶松林龄与建模指标的相关系数 R

Table 3 Correlation coefficient R of Larix gmelinii stand
age versus modeling index

	_		_	
编号	建模参量	芽萌动期	展叶期	落叶期
1	CIre	-0.19	-0.27*	-0.10
2	RVI	-0.25*	-0.27*	0.05
3	PSRI	0.25*	0.01	0.06
4	NDVI	-0.27*	-0.33**	0.02
5	NDWI	-0.29*	-0.67**	-0.24*
6	LAI	-0.31**	-0.57**	-0.20
7	CCL	-0.13	-0.45**	-0.01
8	CWC	-0.32**	-0.74**	-0.30**
9	FVC	-0.33**	-0.55**	-0.23
10	FAPAR	-0.31**	-0.58**	-0.21

注:*,**分别代表显著性水平p<0.05,p<0.01。

表4 展叶期各建模指标的相关系数R

Table 4 Correlation coefficient R of each modeling index in elongating period of leaf

	CIre	RVI	PSRI	NDVI	NDWI	LAI	CCL	CWC	FVC	FAPAR
CIre	1									
RVI	0.62**	1								
PSRI	0.69**	0.12	1							
NDVI	0.62**	0.93**	0.04	1						
NDWI	0.68**	0.48**	0.29^{*}	0.56**	1					
LAI	0.83**	0.62**	0.52**	0.64**	0.90**	1				
CCL	0.92**	0.60**	0.67**	0.61**	0.82**	0.97**	1			
CWC	0.44**	0.39**	0.06	0.47**	0.93**	0.74**	0.61**	1		
FVC	0.70**	0.50**	0.43**	0.55**	0.88**	0.96**	0.90**	0.72**	1	
FAPAR	0.78**	0.55**	0.49**	0.60**	0.91**	0.99**	0.95**	0.74**	0.98**	1

注:*,**分别代表显著性水平p<0.05,p<0.01。

4.2 林龄模型构建与检验

4.2.1 模型参数选择

不同模型需要设置不同参数的搭配组合以获取最优反演结果,而最优参数设置能够反映数据的普遍状态,体现数据间内在逻辑关系。随机森林模型实现最重要的参数包括 ntree 和 mtry,根据自变量贡献率以及建模数量,ntree 选择 3000,mtry 选择系统默认参数 2; 支持向量机回归模型运行最重要参数为核函数,选择普遍使用的高斯核函数,其中 cost 和 gamma 参数根据最优参数估算结果分别设置为 10 和 0.01; BP 神经网络模型运行

参数包括隐蔽单元个数,初始随机数权值,权值 衰减参数,最大迭代次数,线性输出单元,经过 不断实验模型参数,不断进行调参,各个参数分 别设置为5,0.1,0.01,1000,TRUE;多元自适 应回归样条中广义交叉验证(GCV)达到最小时, 预测模型即为最佳林龄模型。

4.2.2 林龄模型对比

本研究对样本数据进行10次交叉采样,并对模型预测值进行取整运算,以降低由样本选择的随意性导致的林龄误差结果差异,并使用平均 R^2 和RMSE用于评价各模型的稳定性。各模型的对比结果如表5所示。

T	able 5 C	ompariso	on of regre	ession mo	dels fitted to	Larix gm	elinii stai	nd age		
评价指标	训练样本(N=50)						验证样本(N=21)			
计价值体	MLR	BP	SVR	RF	MARS	MLR	BP	SVR	RF	MARS
平均决定系数(R2)	0.57	0.67	0.62	0.89	0.68	0.50	0.62	0.58	0.52	0.60
平均均方根误差(RMSF)	5 69	4 99	5 31	2 91	4 92	6.64	5.83	6.10	6.51	5.88

表 5 落叶松林龄回归模型精度比较

注: MLR, BP, SVR, RF, MARS分别代表多元线性回归模型,前馈反向传播神经网络模型,支持向量机回归模型,随机森林模型,多元自适应回归样条模型,N代表样本数量。

通过对落叶松林分年龄训练样本和验证样本 的模型分析发现 (表5): 针对训练样本, 最佳的 林龄估算模型为随机森林模型 (RF), 多元线性回 归模型(MLR)反演精度最低;针对验证样本, 最佳的林龄估算模型为前馈反向传播神经网络模 型 (BP), 多元线性回归模型 (MLR) 反演精度最 低。除随机森林模型(RF)外,各模型在训练样 本和验证样本中林龄估算精度相似, 随机森林模 型(RF)在验证样本中的表现差于训练样本。将 训练样本作为评价标准,各模型预测林龄精度顺 序从高到低依次为: RF>MARS>BP>SVR>MLR, 模型误差范围为 2.91-5.69 a, 优于 Suratman 等 (2004) 在马来西亚针对橡胶林构建的树龄模型 (误差范围 6.4—8.2 a), 略低于 Sivanpillai 等 (2006) 在美国德克萨斯州东部区域构建的火炬松林龄模 型(误差范围 2.89—3.44 a)精度。与上述研究差 异产生的原因是研究区位于山地地区, 地形复杂 性带来同林龄落叶松光谱信息的差异性。以展叶 期的冠层含水量为例(图2),可发现不同坡面下 冠层含水量信息存在差异, 从幼龄林到近熟林阶 段, 阴面冠层含水量高于阳面且变化趋势更快, 这种差异在落叶松林龄反演中会带来更大误差。

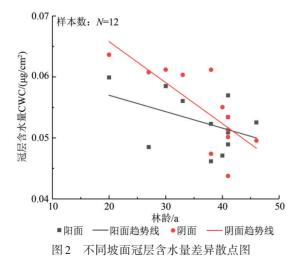


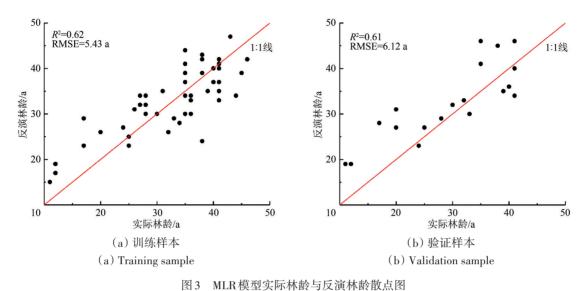
Fig.2 Scatter plots of difference of water content of canopy on different slopes

4.2.3 模型验证与分析

为了进一步分析 5 类模型的林龄误差分布情况,并探究随机森林回归算法在不同样本中结果存在差异的原因。图 3—图 7显示了不同模型中 R^2 最大的一组数据估算值与实测值之间的关系。

图 3-图 7结果表明,5类模型估算结果均存 在误差。其中,各模型均出现高估幼龄林、低估 成熟林的情况,而中龄林和近熟林相对其他龄组 估测误差较小。随机森林回归模型在不同样本集 中结果存在差异的原因是随机森林模型倾向于观 测值较多的类别,在不同样本集中幼龄林样本数 量相对较少目验证样本集中不同龄组样本数量也 低于训练样本,导致验证样本林龄估算误差高于 训练样本。为了进一步分析误差产生原因以及探 讨各模型预测结果的误差分布, 本研究将模型预 测结果与实际结果作差值运算,并对结果取绝对 值,以避免极值对于模型预测精度的影响。根据 训练样本和验证样本的预测误差绝对值分布结果 发现(表6和表7),对于训练样本,各模型预测 误差绝对值 $|\varepsilon| \leq 5$ 的累积百分比由高到低依次为: RF>BP=MARS>SVR>MLR, 随机森林模型 (RF) 预测误差绝对值 |ε|≤5表现明显优于其他反演模 型;多元线性回归模型预测误差绝对值 $|\varepsilon| > 5$ 高达 18个, 预测误差绝对值与其他模型相比结果偏大。 对于验证样本,各模型预测误差绝对值 $|\varepsilon| \leq 5$ 的累 积百分比由高到低依次为: SVR>RF=MARS>BP> MLR, 支持向量机模型 (SVR) 预测误差绝对值 $|\varepsilon|$ ≤ 5 的累积百分比为 0.76,与训练样本进行比 较,可发现支持向量机回归模型 (SVR) 在训练样 本中预测误差绝对值 | € | > 5 高达 16 个, 训练样本 与验证样本表现存在差异,模型稳定性不足,而 随机森林模型(RF)的预测误差绝对值 $|\varepsilon| \leq 5$ 的 累积百分比为0.71, 低于SVR的0.76, 但是预测误 差绝对值 $|\varepsilon| = 0$ 有5个且预测误差绝对值 $|\varepsilon| > 5$ 的 数量仅比SVR多1个,随机森林模型(RF)稳定 性高于支持向量机回归模型 (SVR); 多元线性回归模型 (MLR) 预测误差与训练样本表现类似, 预测误差偏大。综上分析, 虽然随机森林模型在验证样本中预测幼龄林误差高于其他非线性模型

的预测误差,但是预测误差绝对值 |e| ≤ 5 的累积百分比在不同样本集中表现相似,预测误差偏小。因此在不同样本集中,随机森林模型为最佳林龄反演模型。



因 5 MLR(疾至关的怀取 可及便怀取 取点因

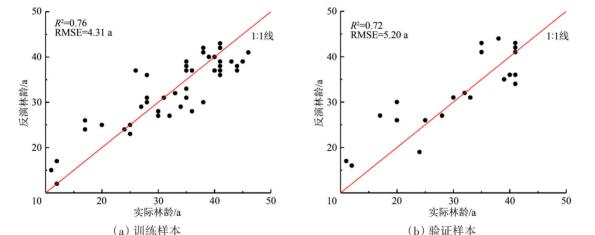


Fig.3 Scatter plots of observed age versus retrieval age using MLR model

图4 BP模型实际林龄与反演林龄散点图

Fig.4 Scatter plots of observed age versus retrieval age using BP model

对于模型中出现高估幼龄林、低估成熟林的情况,根据落叶松整个生长阶段的抚育措施,分析造成该结果的原因主要有以下两点。首先,是由于种植密度差异,落叶松在幼龄林阶段生长速度快且种植密度高于中龄林,导致幼龄林光谱承载信息与中龄林相似。另一方面,是由于冠层结构饱和效应,随着林龄不断增长,虽然间伐等抚育措施会导致种植密度不断下降,但在近熟林和成熟林阶段,落叶松冠层已完全郁闭,并且进入成熟期后落叶松生长缓慢,此时成熟林光谱信息与近熟林相似(Zhang等, 2004)。

(a) Training sample

图 8 显示了基于随机森林模型 (RF) 估算的 研究区落叶松龄组空间分布结果 (Van der Linden 等,2015),其中落叶松分布范围参考了研究区林相图以及徐凯健等 (2018)针对该区落叶松分布的提取结果,其中制图精度 (PA),用户精度 (UA)分别达到89.6%和92.9%。统计各龄组的分布面积,其中幼龄林面积为348.38 m²,中龄林面积为1128.03 m²,近熟林面积为5900.50 m²,成熟林面积为3476.11 m²,落叶松近熟林和成熟林占据整个龄组分布面积的86%。不同龄组分布范围主要为:幼龄林主要集中于研究区中部区域,呈散

(b) Validation sample

状分布,中龄林主要分布于研究区西部地区,近 熟林在整个研究区内均匀分布,而成熟林主要集 中于研究区北部地区。

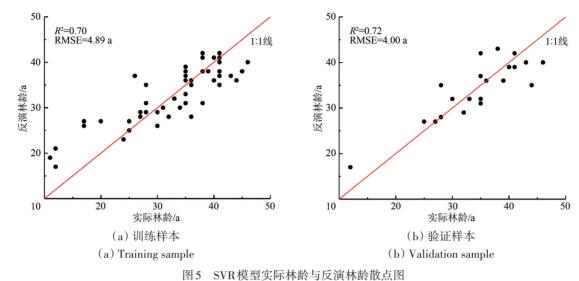


Fig.5 Scatter plots of observed age versus retrieval age using SVR model

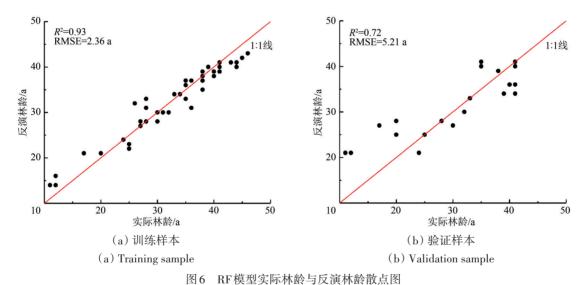


Fig.6 Scatter plots of observed age versus retrieval age using RF model

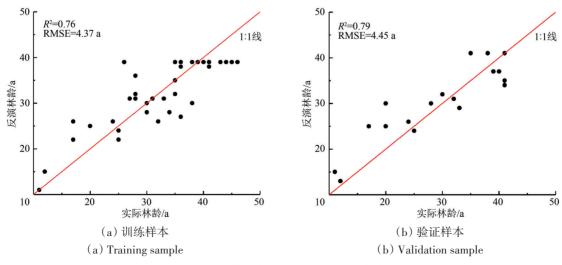


图7 MARS模型实际林龄与反演林龄散点图

Fig.7 Scatter plots of observed age versus retrieval age using MARS model

		error for training samples				
Table 6	6 Distribution table of absolute value of predi					
	表6	训练样本预测误差绝对值分布表				

预测	预测误差绝对值(N=50)									
模型	0 a	1 a	2 a	3 a	4 a	5 a	> 5 a	$ \varepsilon \le 5$		
MLR	6	4	4	3	10	5	18	0.64		
BP	5	6	11	6	7	6	9	0.82		
SVR	4	13	4	4	8	1	16	0.72		
RF	9	14	13	7	4	2	1	0.98		
MARS	5	6	11	7	8	4	9	0.82		

注:N代表样本数量, ε 代表误差。

表 7 验证样本预测误差绝对值分布表

Table 7 Distribution table of absolute value of prediction error for validation samples

预测	预测误差绝对值(N=21)									
模型	0 a	1 a	2 a	3 a	4 a	5 a	> 5 a	$ \varepsilon \le 5$		
MLR	0	5	2	1	1	2	10	0.52		
BP	2	4	2	0	3	2	8	0.62		
SVR	3	3	4	3	1	2	5	0.76		
RF	5	2	1	2	1	4	6	0.71		
MARS	3	3	4	2	2	1	6	0.71		

注:N代表样本数量, ε 代表误差。

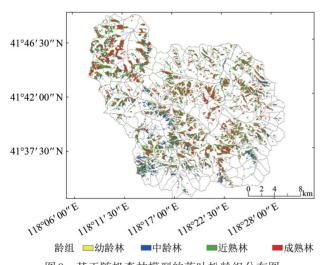


图 8 基于随机森林模型的落叶松龄组分布图 Fig.8 Distribution map of *Larix gmelinii* forest age group using randomForest model

5 结 论

5.1 讨论

本研究采用不同物候期的Sentinel-2卫星数据

构建了5种落叶松林龄遥感反演模型,相对于使用 传统的 Landsat 系列卫星数据构建林龄模型, 其增 加了能反映植被生长状态的红边信息,从而为林 龄模型反演提供了优势。在建模方面,相关研究 较多的采用神经网络模型和多元线性回归模型进 行林龄的估测(Kimes等, 1999; Chen等, 2012), 而本研究增加了对目前较为成熟的深度学习算法 的应用,包括支持向量回归模型(SVR),多元自 适应回归样条模型 (MARS) 以及随机森林模型 (RF)。上述模型应用于林龄方面的研究较少,本 文的研究结果表明随机森林模型为最佳的落叶松 估测林龄模型, 比BP神经网络模型和多元线性回 归模型平均决定系数 R^2 分别提高0.22和0.32,能 够为后期相关针对林龄模型的研究应用提供有效 借鉴。此外,本文构建的BP神经网络模型预测林 龄结果精度较好 (R2=0.67; RMSE=4.99 a), 表明 神经网络模型在预测林龄方面存在一定优越性。

针对模型建模参量,本研究增加了针对不同 物候期林分及其相关生物物理参数的考虑,结果 表明在不同物候期下生物物理参数相关性均高于 植被指数,其中水分指标(冠层含水量)与林龄 相关性最强 (R=-0.74), 该研究结果与 Fiorella 等 (1993) 使用缨帽变换计算的湿度分量 (Wetness) 与美国花旗松林龄相关性程度(R=0.86)类似, 同时 Li 等(2014)的研究结果也证实湿度分量 (Wetness) 与东北白桦林龄具有显著相关性 (R=0.74),后续研究将进一步考虑水分指标对于 林龄反演的影响。此外,在不同的物候期,不同 龄组的林分冠层光谱存在明显差异, 因此针对不 同树种, 林龄模型反演存在最佳反演物候期问题。 本研究选取了落叶松林的全年3个主要物候期(芽 萌动期、展叶期、落叶期)进行对比分析,结果 表明不同物候期建模参量与林龄相关性存在差异 性,其中展叶期的相关性结果最好。因此,相关 林龄模型反演研究需要考虑树种的物候期问题, 尽可能选择目标树种的最佳物候期, 以提高相关 模型的反演精度。

由于基础林龄数据获取困难等条件限制,当 前研究仅针对中国东北地区的落叶松林龄信息进 行了相关遥感反演研究,而对于其他树种以及不 同的地区生长条件的差异性是否会导致相关结果 发生变化,即模型的普适性问题,仍需要做进一 步研究。此外,本研究仅分析了相关植被指数和生物物理参数在估算林龄方面的优劣,后期还将进一步挖掘 Sentinel-2 数据红边波段及其动态信息,并引入与林龄相关的特征参数以提高区域林龄信息反演精度。

5.2 结论

以中国北方典型优势树种落叶松为研究对象,构建最佳物候期下5类落叶松林龄与植被指数以及生物物理参数之间的反演模型,提出最佳林龄的遥感估算方法,提取研究区落叶松龄组的空间分布。主要研究结论如下:

- (1)最佳林龄估算模型为随机森林模型,平均决定系数 R²以及平均均方根误差 RMSE 分别为 0.89 和 2.91 a。各模型估算结果精度从高到低依次为: RF>MARS>BP>SVR>MLR。不同模型整体上均存在高估幼龄林且低估成熟林的情况,这主要是由于种植密度差异以及冠层结构饱和效应引起的。
- (2) 研究区落叶松林龄反演的最佳反演物候期从高到低顺序依次为: 展叶期>芽萌动期>落叶期。在不同物候期, 冠层含水量(CWC)和归一化水体指数(NDWI)与落叶松林龄均呈现出显著的负相关关系(p<0.05), 在展叶期二者相关性达到最大值, 分别为0.74(p<0.01)和0.67(p<0.01),使用水分指标能更好的解释落叶松林龄生长规律。
- (3)本研究采用10×10像元的中心均值林龄样本获取方法,一方面可保证像元的纯净程度且降低复杂地形对于林龄反演的影响,另一方面可确保采样区域的差异不会对反演结果产生影响。

参考文献(References)

- Birth G S and McVey G R. 1968. Measuring the color of growing turf with a reflectance spectrophotometer. Agronomy Journal, 60(6): 640-643 [DOI: 10.2134/agronj1968.00021962006000060016x]
- Champion I, Dubois-Fernandez P, Guyon D and Cottrel M. 2008. Radar image texture as a function of forest stand age. International Journal of Remote Sensing, 29(6): 1795-1800 [DOI: 10.1080/01431160701730128]
- Chen B Q, Cao J H, Wang J K, Wu Z X, Tao Z L, Chen J M, Yang C and Xie G S. 2012. Estimation of rubber stand age in typhoon and chilling injury afflicted area with Landsat TM data: a case study in Hainan Island, China. Forest Ecology and Management, 274:

- 222-230 [DOI: 10.1016/j.foreco.2012.01.033]
- Chen L Y. 2015. Adaptive Regression Model and Application of Multivariate Smoothing Splines. Tangshan: North China University of Science and Technology (陈立宇. 2015. 多元光滑样条自适应回归模型及其应用. 唐山: 华北理工大学)
- Cutler A, Cutler D R and Stevens J R. 2011. Random forests. Machine Learning, 45(1): 157-176
- Dai L M, Jia J, Yu D P, Lewis B J, Zhou L, Zhou W M, Zhao W and Jiang L H. 2013. Effects of climate change on biomass carbon sequestration in old-growth forest ecosystems on Changbai Mountain in Northeast China. Forest Ecology and Management, 300: 106-116 [DOI: 10.1016/j.foreco.2012.06.046]
- Dai M, Zhou T, Yang L L and Jia G S. 2011. Spatial pattern of forest ages in China retrieved from national-level inventory and remote sensing imageries. Geographical Research, 30(1): 172-184 (戴铭, 周涛, 杨玲玲, 贾根锁. 2011. 基于森林详查与遥感数据降尺度 技术估算中国林龄的空间分布. 地理研究, 30(1): 172-184) [DOI: 10.11821/yj2011010017]
- Eggers J J, Bauml R, Tzschoppe R and Girod B. 2003. Scalar costa scheme for information embedding. IEEE Transactions on Signal Processing, 51(4): 1003-1019 [DOI: 10.1109/TSP.2003.809366]
- Filippi A M, Güneralp İ and Randall J. 2014. Hyperspectral remote sensing of aboveground biomass on a river meander Bend using multivariate adaptive regression splines and stochastic gradient boosting. Remote Sensing Letters, 5(5): 432-441 [DOI: 10.1080/2150704X.2014.915070]
- Fiorella M and Ripple W J. 1993. Analysis of conifer forest regeneration using Landsat Thematic Mapper data. Photogrammetric Engineering and Remote Sensing, 59(9): 1383-1388
- Fraser R H and Li Z. 2002. Estimating fire-related parameters in boreal forest using SPOT VEGETATION. Remote Sensing of Environment, 82(1): 95-110 [DOI: 10.1016/S0034-4257(02)00027-5]
- Friedman J H. 1991. Multivariate adaptive regression splines. The Annals of Statistics, 19(1): 1-67 [DOI: 10.1214/aos/1176347963]
- Gao B C. 1996. NDWI-a normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment, 58(3): 257-266 [DOI: 10.1016/S0034-4257(96)00067-3]
- Gitelson A A, Gritz Y and Merzlyak M N. 2003. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. Journal of Plant Physiology, 160(3): 271-282 [DOI: 10. 1078/0176-1617-00887]
- Grant L. 1987. Diffuse and specular characteristics of leaf reflectance.

 Remote Sensing of Environment, 22(2): 309-322 [DOI: 10.1016/0034-4257(87)90064-2]
- Güneralp İ, Filippi A M and Randall J. 2014. Estimation of floodplain aboveground biomass using multispectral remote sensing and non-parametric modeling. International Journal of Applied Earth Observation and Geoinformation, 33: 119-126 [DOI: 10.1016/j.jag.

- 2014.05.004]
- Gunn S R. 1998. Support Vector Machines for Classification and Regression. University of Southampton: 1-28 [DOI: 10.1039/b918 972f]
- Haywood A and Stone C. 2017. Estimating large area forest carbon stocks- a pragmatic design based strategy. Forests, 8(4): 99 [DOI: 10.3390/f8040099]
- He L M, Chen J M, Pan Y D, Birdsey R and Kattge J. 2012. Relationships between net primary productivity and forest stand age in U. S. forests. Global Biogeochemical Cycles, 26(3): (GB3009 [DOI: 10.1029/2010gb003942])
- Jensen J R, Qiu F and Ji M H. 1999. Predictive modelling of coniferous forest age using statistical and artificial neural network approaches applied to remote sensor data. International Journal of Remote Sensing, 20(14): 2805-2822(18) [DOI: 10.1080/0143116 99211804]
- Jiang W Y and Chen F L. 2005. Dictionary of Resources and Environment Law. Beijing: China Legal Publishing House (江伟钰, 陈方林. 2005. 资源环境法词典. 北京: 中国法制出版社)
- Jiao L C. 1990. Neural Network System Theory. Xi'an: Xidian University Press (焦李成. 1990. 神经网络系统理论. 西安: 西安电子科技大学出版社)
- Ju W Z, Wang X J and Sun Y J. 2011. Age structure effects on stand biomass and carbon storage distribution of *Larix olgensis* plantation. Acta Ecologica Sinica, 31(4): 1139-1148 (巨文珍, 王新杰, 孙玉军. 2011. 长白落叶松林龄序列上的生物量及碳储量分配规律. 生态学报, 31(4): 1139-1148)
- Kimes D S, Nelson R F, Salas W A and Skole D L. 1999. Mapping secondary tropical forest and forest age from SPOT HRV data. International Journal of Remote Sensing, 20(18): 3625-3640 [DOI: 10. 1080/014311699211246]
- Knipling E B. 1970. Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. Remote Sensing of Environment, 1(3): 155-159 [DOI: 10.1016/S0034-4257(70)80021-9]
- Lal R. 2008. Carbon sequestration. Philosophical Transactions of the Royal Society B: Biological Sciences, 363(1492): 815-830 [DOI: 10.1098/rstb.2007.2185]
- Li D Q, Ju W M, Fan W Y and Gu Z J. 2014. Estimating the age of deciduous forests in northeast China with Enhanced Thematic Mapper Plus data acquired in different phenological seasons. Journal of Applied Remote Sensing, 8(1): 083670 [DOI: 10.1117/1.JRS.8. 083670]
- Li F, Li M Z, Shi Z L, Jiang H Y and An J P. 2018. Estimates stand age distribution based on forest survey and remote sensing data. Forest Engineering, 34(2): 30-34 (李凡, 李明泽, 史泽林, 姜宏宇, 安金鹏. 2018. 基于遥感数据与森林详查估算林龄的空间分布.森林工程, 34(2): 30-34) [DOI: 10.3969/j. issn. 1006-8023.2018. 02.007]

- Li Q, Zhu J H, Feng Y and Xiao W F. 2018. Carbon storage and carbon sequestration potential of the forest in China. Climate Change Research, 14(3): 287-294 (李奇, 朱建华, 冯源, 肖文发. 2018. 中国森林乔木林碳储量及其固碳潜力预测. 气候变化研究进展, 14 (3): 287-294) [DOI: 10.12006/j.issn.1673-1719.2017.102]
- Li R, Li C J, Xu X G, Wang J H, Yang X D, Huang W J and Pan Y C. 2009. Winter wheat yield estimation based on support vector machine regression and multi-temporal remote sensing data. Transactions of the Chinese Society of Agricultural Engineering, 25(7): 114-117 (黎锐, 李存军, 徐新刚, 王纪华, 杨小冬, 黄文江, 潘瑜春. 2009. 基于支持向量回归(SVR)和多时相遥感数据的冬小麦估产. 农业工程学报, 25(7): 114-117) [DOI: 10.3969/j.issn.1002-6819.2009.07.021]
- Li X H. 2013. Using "random forest" for classification and regression.

 Chinese Bulletin of Entomology, 50(4): 1190-1197 (李欣海.

 2013. 随机森林模型在分类与回归分析中的应用. 应用昆虫学报, 50(4): 1190-1197) [DOI: 10.7679/j.issn.2095-1353.2013.163]
- Litvak M, Miller S, Wofsy S C and Goulden M. 2003. Effect of stand age on whole ecosystem CO₂ exchange in the Canadian boreal forest. Journal of Geophysical Research: Atmospheres, 108(D3): 8225 [DOI: 10.1029/2001JD000854]
- Liu L. 2005. Support Vector Machine and Its Application in Remote Sensing Image Processing. Hefei: University of Science and Technology of China (刘莉. 2005. 支持向量机及其在遥感图像处理中的应用. 合肥: 中国科学技术大学) [DOI: 10.7666/d.y730870]
- Liu R, Miao Q G, Huang B M, Song J F and Debayle J. 2016. Improved road centerlines extraction in high-resolution remote sensing images using shear transform, directional morphological filtering and enhanced broken lines connection. Journal of Visual Communication and Image Representation, 40: 300-311 [DOI: 10.1016/j.jvcir.2016.06.024]
- Liu Z Y, Zhang J and Chen L. 2017. The latest change in the Qinghai-Tibetan Plateau vegetation index and its relationship with climate factors. Climatic and Environmental Research, 22(3): 289-300 (刘 振元, 张杰, 陈立. 2017. 青藏高原植被指数最新变化特征及其 与气候因子的关系. 气候与环境研究, 22(3): 289-300) [DOI: 10. 3878/j.issn.1006-9585.2017.14247]
- Louis J. 2016. Sentinel 2 MSI-level 2A Product Definition. Paris: European Space Agency: 4
- Main-Knorn M, Pflug B, Louis J, Debaecker V, Müller-Wilm U and Gascon F. 2017. Sen2Cor for Sentinel-2//Proceedings of the SPIE 10427 Image and Signal Processing for Remote Sensing XXIII. Warsaw: SPIE: 3 [DOI: 10.1117/12.2278218]
- Meng X Y. 2006. Holzmesslehre. Beijing: China Forestry Publishing House, 2006 (孟宪宇. 2006. 测树学. 北京: 中国林业出版社)
- Merzlyak M N, Gitelson A A, Chivkunova O B and Rakitin V Y U. 1999. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. Physiologia Plantarum, 106 (1): 135-141 [DOI: 10.1034/j.1399-3054.1999.106119.x]

- Miao Z L, Shi W Z, Zhang H and Wang X X. 2013. Road centerline extraction from high-resolution imagery based on shape features and multivariate adaptive regression splines. IEEE Geoscience and Remote Sensing Letters, 10(3): 583-587 [DOI: 10.1109/ LGRS.2012.2214761]
- Muller-Wilm U, Devignot O and Pessiot L. 2016. Sen2Cor Configuration and User Manual. S2-PDGS-MPC-L2A-SUM-V2.3 Issue: 01. European Space Agency
- Shen W J and Li M S. 2017. Mapping disturbance and recovery of plantation forests in southern China using yearly Landsat time series observations. Acta Ecologica Sinica, 37(5): 1438-1449 (沈文娟, 李明诗. 2017. 基于长时间序列 Landsat 影像的南方人工林干扰与恢复制图分析. 生态学报, 37(5): 1438-1449) [DOI: 10.5846/stxb201510142074]
- Sivanpillai R, Smith C T, Srinivasan R, Messina M G and Wu X B. 2006. Estimation of managed loblolly pine stand age and density with Landsat ETM+ data. Forest Ecology and Management, 223 (1/3): 247-254 [DOI: 10.1016/j.foreco.2005.11.013]
- Suratman M N, Bull G Q, Leckie D G, Lemay V M, Marshall P L and Mispan M R. 2004. Prediction models for estimating the area, volume, and age of rubber (*Hevea brasiliensis*) plantations in Malaysia using Landsat TM data. International Forestry Review, 6(1): 1-12 [DOI: 10.1505/ifor.6.1.1.32055]
- Tauxe G M, MacWilliam D, Boyle S M, Guda T and Ray A. 2013. Targeting a dual detector of skin and CO₂ to modify mosquito host seeking. Cell, 155(6): 1365-1379 [DOI: 10.1016/j. cell. 2013. 11.013]
- Tucker C J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment, 8(2): 127-150 [DOI: 10.1016/0034-4257(79)90013-0]
- Van der Linden S, Rabe A, Held M, Jakimow B, Leitão P J, Okujeni A, Schwieder M, Suess S and Hostert P. 2015. The EnMAP-Box—a toolbox and application programming interface for EnMAP data processing. Remote Sensing, 7(9): 11249-11266 [DOI: 10.3390/ rs70911249]
- Wang X C, Wang C K and Yu G R. 2008. Spatio-temporal patterns of forest carbon dioxide exchange based on global eddy covariance measurements. Science in China Series D: Earth Sciences, 51(8): 1129-1143 (王兴昌, 王传宽, 于贵瑞. 2008. 基于全球涡度相关的森林碳交换的时空格局. 中国科学 D 辑: 地球科学, 38(9): 1092-1102)
- Watson R T, Noble I R, Bolin B, Ravindranath N H, Verardo D J and Dokken D J. 2000. Land Use, Land-Use Change, and Forestry. Cambridge University Press: 375
- Weiss M and Baret F. 2016. S2 Toolbox Level 2 Products: LAI,

- FAPAR, FCOVER. INRA [http://step. esa. int/docs/extra/ATBD_S2ToolBox L2B V1.1.pdf]
- Wu W N and Wan T. 2013. Progress of dating methods of tree age.

 Journal of Green Science and Technology, (7): 152-155 (吴斡宁, 万涛. 2013. 浅谈树木年龄测定方法的研究进展. 绿色科技, (7): 152-155) [DOI: 10.3969/j.issn.1674-9944.2013.07.072]
- Wulder M A, Skakun R S, Kurz W A and White J C. 2004. Estimating time since forest harvest using segmented Landsat ETM+ imagery. Remote Sensing of Environment, 93(1/2): 179-187 [DOI: 10. 1016/j.rse.2004.07.009]
- Xu K J, Tian Q J, Yue J B and Tang S F. 2018. Forest tree species identification and its response to spatial scale based on multispectral and multi-resolution remotely sensed data. Chinese Journal of Applied Ecology, 29(12): 3986-3994 (徐凯健, 田庆久, 岳继博, 唐少飞. 2018. 基于多光谱影像的森林树种识别及其空间尺度响应.应用生态学报, 29(12): 3986-3994) [DOI: 10.13287/j.1001-9332. 201812.011]
- Yang B, Li D, Wang L and Chen C. 2017. Retrieval of surface vegetation biomass information and analysis of vegetation feature based on Sentinel-2A in the upper of Minjiang River. Science and Technology Review, 35(21): 74-80 (杨斌, 李丹, 王磊, 陈财. 2017. 基于 Sentine1-2A 岷江上游地表生物量反演与植被特征分析. 科技导报, 35(21): 74-80) [DOI: 10.3981/j. issn. 1000-7857.2017.
- Yu H N, Lee W K, Son Y, Kwak D, Nam K, Kim M, Byun J, Lee S and Kwon T. 2013. Estimating carbon stocks in Korean forests between 2010 and 2110: a prediction based on forest volume-age relationships. Forest Science and Technology, 9(2): 105-110 [DOI: 10.1080/21580103.2013.801174]
- Zhang Q F, Pavlic G, Chen W J, Latifovic R, Fraser R and Cihlar J. 2004. Deriving stand age distribution in boreal forests using SPOT VEGETATION and NOAA AVHRR imagery. Remote Sensing of Environment, 91(3/4): 405-418 [DOI: 10.1016/j.rse.2004. 04.004]
- Zhang Z W. 2013. Investigation and observation of growth and annual cycle phenology of four Larix species. Anhui Agricultural Science Bulletin, 19(15): 106-107 (张志伟. 2013. 4种落叶松的生长与年周期物候的调查观测. 安徽农学通报, 19(15): 106-107) [DOI: 10.3969/j.issn.1007-7731.2013.15.057]
- Zhou F Y. 2017. Relationships among DBH, crown width and stand age of *Pinus sylvestris* var. mongolica plantation in sand land. Protection Forest Science and Technology, (2): 19-21 (周凤艳. 2017.沙地樟子松人工林林木胸径、冠幅等生长指标与林龄相关性研究. 防护林科技, (2): 19-21) [DOI: 10.13601/j.issn.1005-5215. 2017.02.007]

Age information retrieval of Larix gmelinii forest using Sentinel-2 data

TANG Shaofei, TIAN Qingjiu, XU Kaijian, XU Nianxu, YUE Jibo

 International Institute for Earth System Science, Nanjing University, Nanjing 210023, China;
 Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Nanjing University, Nanjing 210023, China

Abstract: The information of forest age structure can effectively reflect the carbon sequestration capacity of regional forest communities at different growth stages. This way is important for assessing the health status of forest ecosystems. In this study, the typical dominant tree species Larix gmelinii forest in temperate zone of China is selected as the object, and Sentinel-2 images of its bud germination period, elongating period of leaf, and defoliation period are selected. The retrieval model of Larix gmelinii stand age is constructed using Multiple Linear Regression (MLR), Random Forest (RF), support vector regression, feedforward back propagation neural network, and multiple adaptive regression spline. The optimal phenophase of remote sensing retrieval is first determined through correlation analysis. On this basis, five optimal characteristic variables, namely, Canopy Water Content (CWC), normalized difference water index, leaf area index, fraction of absorbed photosynthetically active radiatio, and fractional vegetation cover, are selected for model retrieval according to the difference in correlation. Results show that the elongating period of leaf is the optimal remote sensing retrieval phenophase. Except for the plant senescence reflectance index and NDVI and RVI in defoliation period, a negative correlation exists between the stand age of Larix gmelinii and each index, among which the correlation between the stand age and (CWC is the closest, and the correlation coefficient of Pearson reaches -0.74 (p<0.01). The results of different model retrievals indicate that RF model is the best model for estimating the age of Larix gmelinii, and its average coefficient of determination (R^2) and mean Root Mean Square Error (RMSE) are 0.89 and 2.91 a, respectively. MLR is the worst for estimating Larix gmelinii forest age, and its average R^2 and RMSE are 0.57 and 5.69 a, respectively. Nonlinear models can better explain the relationship between stand age and modeling variables.

Key words: remote sensing, Sentinel-2, *Larix gmelinii*, stand age retrieval, biophysical parameters, Random Forest **Supported by** National Key Research and Development Program of China (No. 2017YFD0600903); National Natural Science Foundation of China (No. 41771370); National Science and Technology Major Project (No. 03-Y20A04-9001-17/18, 30-Y20A07-9003-17/18); Civil Aerospace Technology Advance Research Project (No.Y7K00100KJ)