

基于 GF-1 PMS 影像和 k -NN 方法的延庆区森林蓄积量估测

王海宾¹, 彭道黎¹, 高秀会², 李文芳³

(1. 北京林业大学 林学院, 北京 100083; 2. 中国空间科学技术研究院 通信卫星事业部, 北京 100094;
3. 北京市大兴区林业站, 北京 102600)

摘要: 应用高分 1 号(GF-1)全色和多光谱(PMS)影像和 k -最邻近(k -NN)方法进行县域尺度的森林蓄积量估测, 探讨 GF-1 PMS 影像以及 k -NN 方法估测森林蓄积量的适用性。以北京市延庆区森林资源二类调查数据为基础数据, 森林蓄积量为研究对象, 基于国产 GF-1 PMS 影像数据提取植被指数, 采用 k -NN 法构建森林蓄积量估测模型, 并引入偏最小二乘回归法予以比较, 选出最优估测方法对全区森林蓄积量进行反演。结果显示: 偏最小二乘回归法估测的森林蓄积量均方根误差为 $21.90 \text{ m}^3 \cdot \text{hm}^{-2}$, 相对均方根误差为 27.5%, 偏差为 $17.23 \text{ m}^3 \cdot \text{hm}^{-2}$ 。基于 k -NN 方法的森林蓄积量估测的均方根误差为 $12.80 \text{ m}^3 \cdot \text{hm}^{-2}$, 相对均方根误差为 16.0%, 偏差为 $15.02 \text{ m}^3 \cdot \text{hm}^{-2}$ 。与官方公布的全区森林蓄积量进行对比, 结果显示: 基于 k -NN 法反演的全区森林蓄积量统计结果(245.98 万 m^3 , 估测精度为 86.0%)要好于偏最小二乘回归法(266.22 万 m^3 , 估测精度为 76.6%)。最后生成了全区森林蓄积量空间分布图。图 5 表 5 参 37

关键词: 森林测计学; 森林蓄积量; 偏最小二乘回归法; k -最邻近(k -NN)法; GF-1 PMS 影像

中图分类号: S758.5 **文献标志码:** A **文章编号:** 2095-0756(2018)06-1070-09

Forest stock volume estimates in Yanqing District based on GF-1 PMS images and k -NN method

WANG Haibin¹, PENG Daoli¹, GAO Xiuhui², LI Wenfang³

(1. College of Forestry, Beijing Forestry University, Beijing 100083, China; 2. Institute of Telecommunication Satellite, China Academy of Space Technology, Beijing 100094, China; 3. Forestry Station of Daxing District, Beijing 102600, China)

Abstract: To determine the applicability of GF-1 satellite panchromatic and multispectral (PMS) images and the k -NN (k -nearest neighbor) method for estimating forest stock volume (FSV) at the county scale, FSV data of the Forest Management Inventory in Yanqing District of Beijing was taken as the object. Vegetation indices were extracted from domestic GF-1 PMS images with the FSV estimation model then constructed first using the k -NN method and next the Partial Least Squares Regression (PLSR) method to select an optimal model. Then the optimal estimation method was used to determine FSV for the whole county. Results showed that the Root Mean Square Error of FSV estimated by the partial least squares regression method was $21.90 \text{ m}^3 \cdot \text{hm}^{-2}$, and the relative RMSE was 27.5% with a bias of $17.23 \text{ m}^3 \cdot \text{hm}^{-2}$. The RMSE for the FSV estimation based on the k -NN method was $12.80 \text{ m}^3 \cdot \text{hm}^{-2}$ having a relative RMSE of 16.0% with a bias of $15.02 \text{ m}^3 \cdot \text{hm}^{-2}$. These two sets of results, compared to the official announcement of Yanqing's FSV, showed that the FSV estimation based on the k -NN method (2.459 8 million m^3 with an estimation precision of 86.0%) was better than the PLSR method

收稿日期: 2017-11-21; 修回日期: 2018-03-11

基金项目: 国家林业局引进国际先进林业科学技术项目(2015-4-32); 国家重点林业工程监测技术示范推广项目([2015]02号)

作者简介: 王海宾, 从事林业遥感与信息技术研究。E-mail: haibin-w@163.com。通信作者: 彭道黎, 教授, 博士生导师, 从事森林资源监测与评价、森林经营管理和森林资源信息管理研究。E-mail: dlpeng@bjfu.edu.cn

(2.662 2 million m^3 with an estimation precision of 76.6%). Finally, a spatial distribution map of FSV for the whole area was generated. [Ch, 5 fig. 5 tab. 37 ref.]

Key words: forest mensuration; forest stock volume; partial least squares regression; k -NN method; GF-1 PMS images

森林结构参数估测是森林可持续经营和生态环境监测的重要内容。森林蓄积量作为森林结构参数中的一个重要因子,是组成陆地植被生物量的重要成分之一,是评价森林资源数量与质量、反映森林经营管理水平的重要因子,也是评估森林固碳能力和森林碳收支的重要指标^[1-4],因此,准确地估测森林蓄积量对森林经营管理和生态环境保护建设具有重要意义^[5-6]。目前,应用遥感技术估测森林蓄积量的研究主要集中在 2 个方面:一是选取或结合不同遥感数据源(光学遥感、微波探测和激光雷达探测数据)构建估测模型,进行森林蓄积量估测;二是估测方法的多样性,由典型的参数方法(线性回归)向非参数方法转变(如人工神经网络、 k 近邻分类算法等)^[2,4,7-10]。在森林蓄积量模型构建的过程中,自变量间的复共线性问题是亟待解决的一个问题。这些在传统的线性回归估测中无法避免^[11-12]。 k -最邻近(k -nearest neighbor, k -NN)方法作为一种非线性方法,可以不考虑复共线性问题,并且具有稳定的抗逆性,在欧洲和北美地区的森林资源监测中得到了广泛的应用。国内也开展了有关 k -NN 法估测森林蓄积及生物量的研究^[10-13],但对比参数方法和非参数方法进行估测的研究不多。在以往的研究中,采用高分辨率数据进行森林蓄积量的估测多以国外数据为主。高分 1 号(GF-1)卫星是中国高分辨率对地观测卫星系统重大专项的第 1 颗卫星,它可以提供高分辨率的多光谱(8 m)和全色波段(2 m)数据,为森林资源动态监测提供了新的数据源^[14]。然而,受限于 GF-1 卫星影像的覆盖范围等原因,针对 GF-1 全色和多光谱(panchromatic and multispectral, PMS)影像用于县域尺度森林蓄积量定量估测的研究较少。GF-1 PMS 数据将高空间分辨率、高时间分辨率等优势集合于一体,挖掘并探讨 GF-1 PMS 影像数据估测森林蓄积量的潜力,对森林生理生态参数进行定量评价和 GF-1 PMS 影像数据的应用推广具有重要意义。本研究应用 GF-1 PMS 影像数据,通过挖掘影像的光谱信息,建立植被指数自变量集,并对自变量进行优选,采用 k -NN 方法构建森林蓄积量估测模型,并与偏最小二乘回归法构建的森林蓄积量估测模型进行对比,以此分析 GF-1 PMS 影像数据和 2 种估测方法的应用潜力,为县域尺度应用 GF-1 PMS 影像数据估测森林生理生态参数提供参考。

1 研究材料

1.1 研究区概况

延庆区位于北京市的西北部($40^{\circ}16' \sim 40^{\circ}47'N$, $115^{\circ}44' \sim 116^{\circ}34'E$),东邻怀柔,南接昌平,西面和北面与河北省怀来县、赤城县相接,三面环山,总面积达 $1\,993.75\text{ km}^2$ 。该区属大陆性季风气候,是暖温带与中温带,半干旱与半湿润的过渡带。年平均温度为 $8.8\text{ }^{\circ}\text{C}$,无霜期 $150 \sim 160\text{ d}$ 。年平均降水量为 467.0 mm ,降水主要集中在 6-8 月。该区内森林植被属于针阔混交林森林植被,现存植被主要为人工林以及一些次生植被类型。区内主要森林类型有油松 *Pinus tabulaeformis* 林,侧柏 *Platycladus orientalis* 林,华北落叶松 *Larix principis-rupprechtii* 林,蒙古栎 *Quercus liaotungensis* 林,刺槐 *Robinia pseudoacacia* 林,白桦 *Betula platyphylla* 林,山杨 *Populus davidiana* 林和杂木林等。据 2015 年北京市公布的森林资源数据显示,延庆区森林面积为 11.5 万 hm^2 ,森林覆盖率为 57.46% ,森林蓄积量达到 215.77 万 m^3 ,占北京市总森林蓄积量的 12.68% ,具有重要的生态系统防护作用和固碳功能。

1.2 研究数据

1.2.1 地面数据及辅助数据 地面数据是延庆区 2014 年森林资源二类调查数据(以下简称“二调数据”),以矢量地理信息数据的形式储存。全区二类调查数据中的小班个数为 2.37×10^4 个,小班面积及蓄积的统计特征见表 1。本研究所应用到的调查因子包括小班编号、地类、小班蓄积、小班面积等因子。辅助数据包括延庆区边界矢量数据和覆盖延庆区的 1:10 000 数字高程模型(DEM)数据,地面数据和辅助数据的坐标为北京 54 坐标系。

1.2.2 遥感影像 GF-1 号卫星装载有 2 m 分辨率全色和 8 m 分辨率多光谱相结合的相机(PMS1~PMS2),

多光谱相机包括蓝(0.45~0.52 μm), 绿(0.52~0.59 μm), 红(0.63~0.69 μm)和近红外(0.77~0.89 μm)4个波段, 全色相机包括1个全色波段(0.45~0.89 μm), 2台相机组合幅宽达60 km, 重访周期为4 d。本研究选取了覆盖研究区的7景GF-1 PMS影像, 由中国林业科学研究院资源信息研究所提供, 其中序号2影像由于影像缺失, 采用了2014年的影像进行补充, 并假设对估测结果未产生影响, 序号6和序号7的影像获取时间为2015年1月31号, 作为补充影像, 只覆盖了少部分区域, 因此产生的差异可以忽略不计。GF-1 PMS影像数据详情见表2。

表1 延庆区二类调查数据小班面积及蓄积基本统计量

Table 1 Descriptive statistics of small class area and volume of forest management inventory in Yanqing District

项目	小班面积/ hm^2	小班蓄积/ m^3
最大值	693.65	8 880.45
最小值	0.07	0.06
平均值	8.45	89.75

表2 覆盖延庆区的GF-1 PMS影像列表

Table 2 List of GF-1 PMS images covering Yanqing District

序号	影像传感器及轨道号	(多光谱/全色)/m	获取时间(年-月-日)
1	GF1_PMS1_E115.8_N40.3_20150505_L1A0000791191-MSS1	8/2	2015-05-05
2	GF1_PMS1_E115.9_N40.5_20140427_L1A0000212130-MSS1	8/2	2014-04-27
3	GF1_PMS1_E115.8_N40.5_20150505_L1A0000791190-MSS1	8/2	2015-05-05
4	GF1_PMS2_E116.2_N40.5_20150505_L1A0000791125-MSS2	8/2	2015-05-05
5	GF1_PMS2_E116.3_N40.7_20150505_L1A0000791124-MSS2	8/2	2015-05-05
6	GF1_PMS2_E116.4_N40.5_20150131_L1A0000624694-MSS2	8/2	2015-01-31
7	GF1_PMS2_E116.5_N40.7_20150131_L1A0000624693-MSS2	8/2	2015-01-31

1.2.3 地面数据处理 二类调查数据为区划调查数据, 可以覆盖整个延庆区。在估测森林蓄积量前, 需要计算小班内的每公顷森林蓄积量[不包括非森林地类的蓄积量(疏林地、散生木、四旁树的蓄积量), 以下统称公顷蓄积量], 采用 Arc GIS 软件中属性表中的字段计算器计算(公顷蓄积量=小班蓄积/小班面积)。本研究采用系统抽样方法在延庆区内布设抽样点, 为与 GF-1 PMS 影像 3×3 窗口大小近似匹配, 确定抽样点的面积大小为 600 m^2 , 形状为正方形。利用 Arc GIS 软件布设覆盖延庆区的格网, 格网面积大小为 600 m^2 。在布设格网时, 存在单个格网覆盖多个小班的情况。本研究采用格网内小班公顷蓄积乘以对应小班的面积权重值, 并求和获得覆盖多个小班的格网的蓄积量, 最后计算格网内公顷蓄积量的变异系数, 并取可靠水平为 90%和估计精度为 85%(二类规程要求)的控制标准, 计算需要布设的抽样点个数及间距, 并以此为基础外推估算延庆区的森林蓄积量。经计算公顷蓄积量的变异系数为 2.015 3, 确定样本数为 489 个, 抽样间距为 2.02 km。为与北京市一类清查样地布设间距相近, 在满足抽样精度的基础上, 本研究设置系统抽样间距为 2 km × 2 km(图 1)。经系统布设样地后得到延庆区的抽样点共计 492 个, 将抽样点所在的小班属性信息追加到抽样点上。处理后的样点数据包括样地号、地类、公顷蓄积量等因子, 样点数据同时用于影像的分类结果验证。提取出蓄积量大于 0 的样点用于森林蓄积量反演, 经处理后提取抽样点数据(蓄积大于 0)共计 156 个。

1.2.4 遥感影像预处理 应用 ENVI 5.1 对 7 景 GF-1 PMS 影像进行辐射定标、大气校正, 参照已校正好的 2004 年 SOPT-5 融合数据(2.5 m), 对高分影像进行正射校正, 误差控制在 1 个像元以内, 对校正好的 7 景 GF-1 PMS 影像进行镶嵌处理, 生成覆盖整个延庆区的 GF-1 PMS 影像, 并利用延庆区边界矢量数据进行裁剪, 生成延庆区范围内的 GF-1 影像(图 1), 遥感影像校正后的坐标同地面数据和辅助数据。本研究选取应用较为广泛的极大似然分类法提取森林植被信息^[12]。依据《国家森林资源连续清查主要技术规定》(2014)中的森林面积定义: 森林面积等于乔木林面积、竹林面积与特殊灌木林面积之和。由延庆区二类调查数据可知, 地类中无竹林地和特殊灌木林地, 因此本研究中的森林面积只包括乔木林面积, 森林蓄积量只包括乔木林蓄积量。按照《国家森林资源连续清查主要技术规定》(2014)地类划分标准, 结合研究区实际地类情况, 确定分类类别为乔木林地, 其他林地(包括灌木林、疏林地、未成林造林地、苗圃地、迹地、宜林地), 农地(耕地), 建筑(建设用地), 水域和其他土地(未利用地)等 6 类, 对分类结果合并归类为森林和非森林 2 类, 其中森林包括乔木林, 非森林包括除乔木林之外所有的地类, 最后

对影像分类结果进行验证。结果表明：GF-1 PMS 影像分类总精度达到 87.90%，Kappa 系数为 0.75。分类结果见图 2。

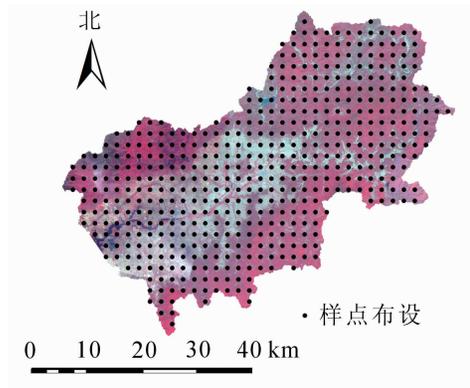


图 1 延庆区 GF-1 PMS 影像(432 假彩色)及系统
布设抽样点(2 km×2 km)

Figure 1 GF-1 PMS image of Yanqing District (432 false color)
and sampling points of system layout (2 km × 2 km)

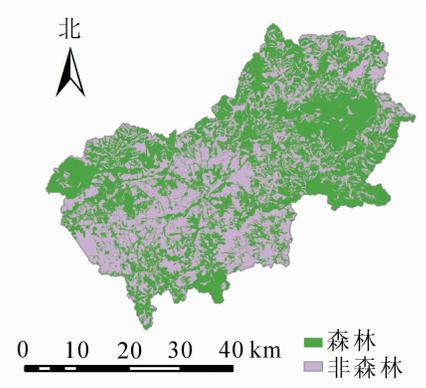


图 2 GF-1 PMS 影像分类图
Figure 2 Classification result of GF-1 PMS image

2 研究方法

2.1 遥感特征变量

遥感影像的植被信息是通过绿色植物叶片和植被冠层光谱特性及其变化差异反映的^[14]。选择合适的植被指数可以有效地增强植被信息或抑制非植被信息，减小非植被信息如阴影对模型构建的影响^[3]。因此，本研究借鉴前人的经验并结合蓄积量估测特点，选择 9 种植被指数作为建模变量(表 3)。

应用布设好的抽样点提取对应位置的 GF-1 PMS 影像光谱信息，计算样本内的光谱平均值作为样本点的光谱值，参与模型的构建。本研究中，蓄积大于 0 的抽样点均落在乔木林(即有林地)范围内，不包括散生木、四旁树和疏林地部分，因此所构建的模型为乔木林蓄积量模型，参与全区的森林蓄积量估算。

2.2 自变量筛选

在建模变量中，自变量往往存在多重共线性问题，导致所构建的模型不够稳定，尤其是在变量个数较多时，会降低模型的预测精度，因此选择适宜的方法筛选自变量用于建模具有重要作用^[24]。常用的变量筛选方法较多，有逐步回归法、Pearson 相关系数法、平均残差平方和法等^[4,24-26]。相关研究表明：平均残差平方和法在变量筛选中具有很好的适用性^[26]，因此本研究选择此方法作为自变量筛选方法，平均残差平方和方法筛选变量的实现步骤参考文献^[26]。

2.3 偏最小二乘回归法

偏最小二乘回归法是伍德(WOLD)和阿巴诺(ALBANO)于 1983 年提出的一种新型的多元统计方法。它集多元回归分析、典型相关分析和主成分分析的优点于一体，可以有效地解决多元回归分析中的变量

表 3 研究中用到的变量

Table 3 Variables used in the study

变量	计算公式	参考文献	变量	计算公式	参考文献
差值植被指数 D_{VI}	$VI = \rho_4 - \rho_3$	[15]	归一化绿波段差值植被指数 G_{NDVI}	$G_{NDVI} = \frac{\rho_4 - \rho_2}{\rho_4 + \rho_2}$	[20]
重归一化植被指数 R_{DVI}	$R_{DVI} = \sqrt{N_{DVI} \times D_{VI}}$	[16]	考虑绿波段差值植被指数 G_{INDVI}	$G_{INDVI} = \frac{\rho_4 - (\rho_1 + \rho_2)}{\rho_4 + \rho_1 + \rho_2}$	[21]
土壤植被指数 S_{AVI}	$S_{AVI} = \frac{1.5(\rho_4 - \rho_3)}{(\rho_4 + \rho_3 + 0.5)}$	[17]	比值植被指数 R_{VI}	$R_{VI} = \rho_4 / \rho_3$	[22]
优化土壤调整指数 O_{SAVI}	$O_{SAVI} = \frac{1.05(\rho_4 - \rho_3)}{(\rho_4 + \rho_3 + 1.05)}$	[18]	转换型植被指数 T_{VI}	$T_{VI} = \sqrt{N_{DVI} + 0.5}$	[23]
归一化植被指数 N_{DVI}	$N_{DVI} = \frac{\rho_4 - \rho_3}{\rho_4 + \rho_3}$	[19]			

说明： $\rho_1, \rho_2, \rho_3, \rho_4$ 分别为蓝、绿、红、近红波段

多重相关性及噪声问题^[27-29], 近年来被逐渐地应用到森林参数的估计上^[3,30-31]。

以单因变量为例阐述其基本建模思想: 设有因变量 y 和 p 个自变量 $\{x_1, x_2, \dots, x_p\}$, 样本数为 n , 构成因变量和自变量的数据表 $y = [u]_{n \times 1}$ 和 $x = [x_1, x_2, \dots, x_p]_{n \times p}$, 对数据表 x 进行主成分处理, 提取一个主成分 t_1 (t_1 是 x_1, x_2, \dots, x_p 的线性组合), 要求 t_1 尽可能多地携带 x 中的变异信息, 同时与 y 的相关性最大, 提取第 1 个主成分 t_1 后, 实施 y 和 x 对 t_1 的回归, 如果此时回归方程达到满意的精度, 则算法停止, 否则利用 x 和 y 被 t_1 解释后的残余信息进行第 2 轮的主成分提取, 如此反复, 直到能达到较为满意的精度为止。若最终对 x 提取了 m 个主成分 t_1, t_2, \dots, t_m , 偏最小二乘回归将实施 y 对 t_1, t_2, \dots, t_m 的回归, 然后表达成 y 对原变量 x 的回归方程。应用偏最小二乘回归法进行模型构建的详细过程见文献[9]。本研究应用 MATLAB 2014a 软件实现偏最小二乘回归模型的构建。

2.4 k -NN 算法基本原理

k -NN 算法是一种典型的非参数方法, 基于观测点和预测点之间的空间相似性关系进行单变量或多变量预测(如蓄积量等森林参数), 在欧洲和北美得到了广泛应用^[32-34]。 k -NN 法用于森林参数估计的最大优点在于它不仅能同时估计若干个森林参数, 而且它还能够维持参数之间的自然依赖结构, 保持参数之间的一致性, 同利用遥感自变量和森林参数建立的回归模型相比, k -NN 方法更多地考虑森林参数同自变量间的非线性依赖关系^[12,35]。其基本原理为: 记 p 为目标点, p_i 为参考点, 并且参考点的森林参数(本研究为森林蓄积量)是已知的, D_{p,p_i} 为目标点和参考点之间的光谱距离, 它是用来衡量样本参数间的相似度的。对于目标点 p , 找出其光谱空间最邻近的 k 个参考点 p_1, p_2, \dots, p_k , 其中 $D_{p,1} < D_{p,2} < \dots < D_{p,k}$ 。由于目标点 p 受其近邻的影响是不同的, 通常距离越近的参考点对其影响越大, 反之则影响越小。目标点 p 的森林参数 V_p 可以通过 k 个参考点的相应的森林参数 V_{p_i} ($i=1, 2, \dots, k$) 的加权平均法获得^[12], 其中 k 个参考点的权重 $w_{p_i,p}$ 通过光谱空间的反距离函数获得^[12]。 k -NN 实质上是一个常用于空间插值的反距离加权平均法, 当 $k=1$ 时, k -NN 即为最邻近距离法^[35]。

$$w_{p_i,p} = \frac{1}{D_{p_i,p}^t} / \sum_{i=1}^k \frac{1}{D_{p_i,p}^t}; \quad (1)$$

$$V_p = \sum_{i=1}^k w_{p_i,p} V_{p_i} \quad (2)$$

式(1)~式(2)中: t 为距离分解因子, 一般取值为 0, 1, 2。光谱距离 $D_{p_i,p}$ 可采用多种距离进行度量, 常用的有欧氏距离、马氏距离和光谱角制图等。在借鉴前人研究的基础上^[12-13], 选择 t 值为 1, 马氏距离作为距离度量标准, 在确定最优波段的基础上, 计算最优的 k 值用于全区森林蓄积量反演。本研究的 k 值是由 MATLAB 2014a 软件计算获得。

2.5 模型评价

选择留一交叉验证评价。模型估测精度采用均方根误差(E_{RMS}), 相对均方根误差(E_{RRMS}), 偏差(b_{bias}) 3 个指标来评价, 指标计算公式如下:

$$E_{\text{RMS}} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}; \quad (3)$$

$$E_{\text{RRMS}} = \frac{E_{\text{RMS}}}{\bar{y}_i}; \quad (4)$$

$$b_{\text{bias}} = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)}{n} \quad (5)$$

式(3)~式(5)中: y_i 为蓄积量实测值, \hat{y}_i 为蓄积量预测值, \bar{y}_i 为蓄积量实测值平均值, n 为样本个数。

3 研究结果

3.1 自变量筛选

利用平均残差平方和准则筛选出蓄积量估测的建模变量因子。由图 3 可知: 经过变量优选, 提取的

S_{AVI} , O_{SAVI} , N_{DVI} , G_{NDVI} , G_{BNDVI} , D_{VI} 可以得最小的平均残差平方和, 为 13.71。因此选择 S_{AVI} , O_{SAVI} , N_{DVI} , G_{NDVI} , G_{BNDVI} , D_{VI} 共 6 个因子作为蓄积量模型构建的优选变量。

3.2 蓄积量估测模型构建

3.2.1 偏最小二乘回归模型构建 将筛选的自变量构建森林蓄积量估测模型, 应用 MATLAB 计算模型各个参数, 最后得到基于偏最小二乘回归法构建的蓄积量估测模型, 其数学表达式为 $V=0.809 5+0.017 5 S_{AVI}+0.025 8 O_{SAVI}+0.025 6 N_{DVI}+2.034 4 G_{NDVI}+3.545 9 G_{BNDVI}+0.001 5 D_{VI}$ 。其中, V 为森林蓄积量, S_{AVI} , O_{SAVI} , N_{DVI} , G_{NDVI} , G_{BNDVI} 和 D_{VI} 为表 3 对应的植被指数。

3.2.2 k 值的确定 k -NN 方法的估测精度随着 k 值的不同而变化。相关研究表明^[12,23,26]: k 值取 1~10 的范围时, 对不同参数的估测效果较好。本研究将 k 的取值范围设置为 3~10, 并对不同 k 值的估测结果进行分析, 得到最优 k 值。由图 4 可知: 当 k 值为 7 时, k -NN 方法估测的均方根误差(E_{RMS})最小, 为 $10.633 4 \text{ m}^3 \cdot \text{hm}^{-2}$ 。因此, 本研究选取 7 作为蓄积量估测的最佳 k 值。

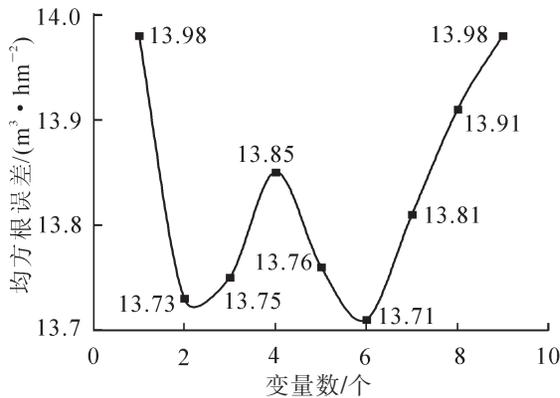


图 3 建模变量优选

Figure 3 Optimization of model variables

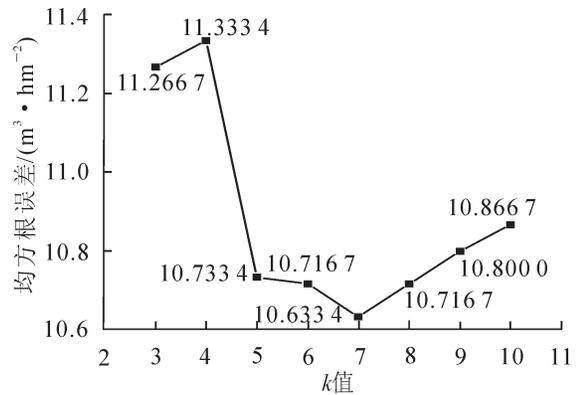


图 4 均方根误差随 k 值变化的情况

Figure 4 Situation of root mean square error with the different k values

3.2.3 模型验证 由表 4 可知: 偏最小二乘回归法估测的森林蓄积量均方根误差为 $21.90 \text{ m}^3 \cdot \text{hm}^{-2}$, 相对均方根误差为 27.5%, 偏差为 $17.23 \text{ m}^3 \cdot \text{hm}^{-2}$, 基于 k -NN 方法的森林蓄积量估测的均方根误差为 $12.80 \text{ m}^3 \cdot \text{hm}^{-2}$, 相对均方根误差为 16.0%, 偏差为 $15.02 \text{ m}^3 \cdot \text{hm}^{-2}$ 。由此可知: 基于非参数方法的 k -NN 方法估测效果要好于基于参数方法的偏最小二乘回归法, 可作为延庆区森林蓄积量估测的方法。进一步对延庆区森林总蓄积量预测结果进行精度检验。以首都园林绿化政务网 (http://www.bjyl.gov.cn/zwgk/tjxx/201604/t20160401_178533.html) 公布的 2015 年延庆区森林蓄积量数据为实际值, 对比分析基于 2 种方法反演的蓄积量预测值的差异(表 5)。由表 5 可知: 应用建立的偏最小二乘回归模型反演的森林蓄积量为 266.22 万 m^3 , 相对误差为 50.45 万 m^3 , 估测精度为 76.6%, 应用 k -NN 方法反演的森林蓄积量 245.98 万 m^3 , 相对误差为 30.21 万 m^3 , 估测精度为 86.0%。由此可知: 基于 k -NN 方法反演全区森林蓄积量的估测结果要好于偏最小二乘回归法, 估测结果可靠。

表 4 2 种估测模型的估测精度

Table 4 Estimation accuracy of two models

评价指标	均方根误差/ ($\text{m}^3 \cdot \text{hm}^{-2}$)	相对均方 根误差/%	偏差/ ($\text{m}^3 \cdot \text{hm}^{-2}$)
偏最小二乘回归法	21.90	27.5	17.23
k -NN 方法	12.80	16.0	15.02

表 5 延庆区森林蓄积量(FSV)估测误差比较

Table 5 Comparison of estimation error of forest stock volume in Yanqing District

项目	森林蓄积 量/万 m^3	相对误 差/万 m^3	误差百 分比/%
森林蓄积量统计(网站)	215.77		
偏最小二乘回归法	266.22	50.45	23.38
k -NN 法	245.98	30.21	14.00

3.3 基于 k -NN 法的森林蓄积量反演

依据遥感影像分类得到的森林边界数据, 应用 k -NN 方法建立的蓄积量估测模型对研究区森林蓄积量进行反演, 得到延庆区森林蓄积量空间分布等级图(图 5)。

4 讨论

基于偏最小二乘回归法的回归统计方法可以减少多重共线性的影响, 提高模型构建的稳定性, 保证

估测的精度。以往的研究表明^[3,27-28]：应用偏最小二乘回归法估测森林参数要好于其他常规的回归方法，可以在一定程度上提高待估参数的估测精度。 k -NN方法作为一种典型的非参数方法，它不用考虑变量间的多重共线性的关系，具有一定的应用优势。本研究比较了这2种方法的估测效果，结果表明：在研究区内，基于 k -NN方法的非参数方法的估测精度要好于基于偏最小二乘回归的参数方法，对全区森林蓄积量进行反演的估测精度达到86.0%。但本研究还存在以下问题：①研究应用的两景GF-1影像虽然只覆盖了很少的区域，但影像时相与其余影像时相存在差异，可能导致蓄积量估测结果产生偏差，但在本研究中忽略了其影响，此部分问题有待于进一步分析。②应用 k -NN法进行蓄积量估测可以减少由影像的同物异谱和同谱异物带来的随机变化^[11]，但其高值低估和低值高估现象依旧是应用 k -NN法基于像素级估测森林参数普遍存在的问题^[11-12,35]。在本研究的估测过程中，随着 k 值的增大，蓄积量估计值分布空间呈现缩小的趋势。在后续的研究中，可考虑直方图匹配方^[11]来提高 k -NN法估测变量的精度。③本研究中的均方根误差随 k 值的增加呈现不规则的变化，这与KATILA等^[37]和郑刚等^[12]的研究中的均方根误差随 k 值的增大而减小的趋势不同。向安民等^[23]应用GF影像估算森林蓄积量的研究中也出现了均方根误差随 k 值的增加出现不规则的变化趋势。这与GF影像上的椒盐噪声有很大关系，使得光谱信息与蓄积量值出现不匹配的现象，导致了均方根误差不规则的变化。

5 结论

本研究应用GF-1 PMS影像数据提取了9种植被指数作为建模变量，采用平均残差平方和法对变量进行筛选，最终提取的 S_{AVI} 、 O_{SAVI} 、 N_{DVI} 、 G_{NDVI} 、 G_{BNDVI} 和 D_{VI} 等6个变量因子可以得到最小的平均残差平方和，作为蓄积量估测模型的变量。对比了偏最小二乘回归法(参数方法)和 k -NN方法(非参数方法)估测森林蓄积量的效果。结果显示： k -NN方法的估测效果(均方根误差为 $12.80 \text{ m}^3 \cdot \text{hm}^{-2}$ ，相对均方根误差为16.0%，偏差为 $15.02 \text{ m}^3 \cdot \text{hm}^{-2}$)要好于偏最小二乘回归法(均方根误差为 $21.90 \text{ m}^3 \cdot \text{hm}^{-2}$ ，相对均方根误差为27.5%，偏差为 $17.23 \text{ m}^3 \cdot \text{hm}^{-2}$)，并与首都园林绿化网公布的延庆区森林蓄积量结果进行了对比分析，基于 k -NN方法反演的森林蓄积量估测精度达到86.0%，最后生成了延庆区森林蓄积量空间分布图。说明应用GF-1 PMS影像估测森林蓄积量是可行的。可以进一步在不同的县域范围应用GF-1 PMS影像估测森林蓄积量及相关的森林参数，对GF-1 PMS影像的估测潜力进行挖掘，这对推广应用国产GF-1 PMS影像具有重要意义。

6 致谢

感谢中国林业科学研究院资源信息研究所提供GF-1 PMS卫星影像数据。

7 参考文献

- [1] NICHOL J E, SARKER M L R. Improved biomass estimation using the texture parameters of two high-resolution optical sensors [J]. *IEEE Trans Geosci Remote Sens*, 2011, **49**(3): 930 - 948.
- [2] MENG Jinghui, LI Shiming, WANG Wei, et al. Estimation of forest structural diversity using the spectral and textural information derived from SPOT-5 satellite images [J]. *Remote Sens*, 2016, **8**(2): 125.
- [3] 张超, 彭道黎, 涂云燕, 等. 利用TM影像和偏最小二乘回归方法估测三峡库区森林蓄积量[J]. 北京林业大学学报, 2013, **35**(3): 11 - 17.
ZHANG Chao, PENG Daoli, TU Yunyan, et al. Predicting forest volume in Three Gorges Reservoir Region using TM images and partial least squares regression [J]. *J Beijing For Univ*, 2013, **35**(3): 11 - 17.
- [4] 王海宾, 彭道黎, 范应龙, 等. 基于辅助信息的森林蓄积量空间模拟[J]. 农业机械学报, 2016, **47**(6): 283 - 289.

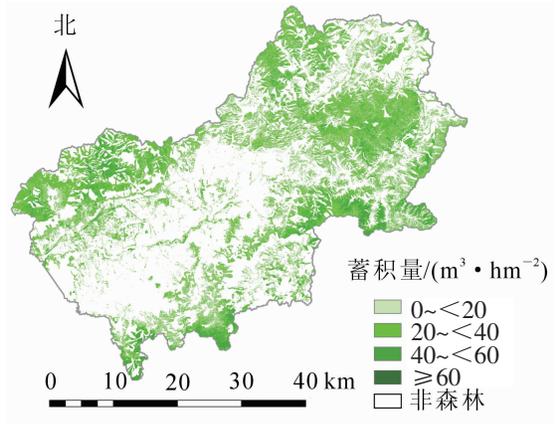


图5 延庆区森林蓄积量预测等级图
Figure 5 Predicting map of forest stock volume in Yanqing District

- WANG Haibin, PENG Daoli, FAN Yinglong, *et al.* Spatial modeling of forest stock volume based on auxiliary information [J]. *Trans Chin Soc Agric Mach*, 2016, **47**(6): 283 – 289.
- [5] 亢新刚. 森林资源经营管理[M]. 北京: 中国林业出版社, 2001.
- [6] 涂云燕. 森林蓄积量遥感估测研究[D]. 北京: 北京林业大学, 2013.
- TU Yunyan. *The Research of Estimating Forest Volume Based on Remote Sensing* [D]. Beijing: Beijing Forestry University, 2013.
- [7] CORONA P, FATTORINI L, FRANCESCHI S, *et al.* Mapping by spatial predictors exploiting remotely sensed and ground data: a comparative design-based perspective [J]. *Remote Sens Environ*, 2014, **152**(1): 29 – 37.
- [8] ZHENG Sheng, CAO Chunxiang, DANG Yongfeng, *et al.* Retrieval of forest growing stock volume by two different methods using Landsat TM images [J]. *Int J Remote Sens*, 2013, **35**(1): 29 – 43.
- [9] MORA B, WULDER M A, HOBART G W, *et al.* Forest inventory stand height estimates from very high spatial resolution satellite imagery calibrated with lidar plots [J]. *Int J Remote Sens*, 2013, **34**(12): 4406 – 4424.
- [10] 陈尔学, 李增元, 武红敢, 等. 基于 k -NN 和 Landsat 数据的小面积统计单元森林蓄积估测方法[J]. 林业科学研究, 2008, **21**(6): 745 – 750.
- CHEN Erxue, LI Zengyuan, WU Honggan, *et al.* Forest volume estimation method for small areas based on k -NN and landsat data [J]. *For Res*, 2008, **21**(6): 745 – 750.
- [11] 戚玉娇, 李凤日. 基于 k -NN 方法的大兴安岭地区森林地上碳储量遥感估算[J]. 林业科学, 2015, **51**(5): 46 – 55.
- QI Yujiao, LI Fengri. Remote sensing estimation of aboveground forest carbon storage in Daxing'an Mountains based on k -NN method [J]. *Sci Silv Sin*, 2015, **51**(5): 46 – 55.
- [12] 郑刚, 彭世揆, 戎慧, 等. 基于 K -NN 方法的森林蓄积量遥感估计和反演概述[J]. 遥感技术与应用, 2010, **25**(3): 431 – 437.
- ZHENG Gang, PENG Shikui, RONG Hui, *et al.* A general introduction to estimation and retrieval of forest volume with remote sensing based on k -NN [J]. *Remote Sens Technol Appl*, 2010, **25**(3): 431 – 437.
- [13] 曹庆先, 徐大平, 鞠洪波. 基于 TM 影像纹理与光谱特征和 k -NN 方法估算 5 种红树林群落生物量[J]. 林业科学研究, 2011, **24**(2): 144 – 150.
- CAO Qingxian, XU Daping, JU Hongbo. Biomass estimation of five kinds of mangrove community with the k -NN method based on the spectral information and textural features of TM images [J]. *For Res*, 2011, **24**(2): 144 – 150.
- [14] 孔凡婕, 李晓兵, 白云晓, 等. 基于 GF-1 WFV 影像的丰宁县植被地上碳储量估算[J]. 资源科学, 2016, **38**(6): 1054 – 1064.
- KONG Fanjie, LI Xiaobing, BAI Yunxiao, *et al.* The estimation of aboveground vegetation carbon storage based on GF-1 WFV satellite images in Fengning County [J]. *Resour Sci*, 2016, **38**(6): 1054 – 1064.
- [15] 李旭文, 季耿善, 杨静. 太湖梅梁湖湾蓝藻生物量遥感估算[J]. 国土资源遥感, 1995, **7**(2): 23 – 28.
- LI Xuwen, JI Gengshan, YANG Jing. Estimating cyanophyta biomass standing crops in Meiliang Gulf of Lake Taihu by satellite remote sensing [J]. *Remote Sens Land Resour*, 1995, **7**(2): 23 – 28.
- [16] ROUJEAN J L, BREON F M. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements [J]. *Remote Sens Environ*, 1995, **51**(3): 375 – 384.
- [17] HUETE A R. A soil-adjusted vegetation index (SAVI) [J]. *Remote Sens Environ*, 1988, **25**(3): 295 – 309.
- [18] RONDEAUX G, STEVEN M, BARET F. Optimization of soil-adjusted vegetation indices [J]. *Remote Sens Environ*, 1996, **55**: 95 – 107.
- [19] ROUSE J W, HAAS R H, SCHELL J A, *et al.* Monitoring vegetation systems in the great plains with ERTS[M]// FREDEN S C, MERCANTI E P, BECKER M A. *Third Earth Resources Technology Satellite-1 Symposium-Volume I: Technical Presentations (NASA SP-351)*. Washington D C: NASA Special Publication, 1974: 309.
- [20] GITELSON A A, KAUFMAN Y J, MERZLYAK M N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS [J]. *Remote Sens Environ*, 1996, **58**(3): 289 – 298.
- [21] ADAM E, MUTANGA O, RUGEGE D. Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: a review [J]. *Wetlands Ecol Manage*, 2009, **18**(3): 281 – 296.
- [22] PEARSON R L, MILLER L D. Remote mapping of standing crop biomass for estimation of the productivity of the

- shortgrass prairie [J]. *Remote Sens Environ*, 1972, **45**(2): 7 – 12.
- [23] 向安民, 刘凤伶, 于宝义, 等. 基于 k -NN 方法和 GF 遥感影像的森林蓄积量估测[J]. 浙江农林大学学报, 2017, **34**(3): 406 – 412.
XIANG Anmin, LIU Fengling, YU Baoyi, *et al.* Forest stock volume estimation based on the k -NN method and GF remote sensing data [J]. *J Zhejiang A&F Univ*, 2017, **34**(3): 406 – 412.
- [24] 徐小军, 周国模, 杜华强, 等. 基于 Landsat TM 数据估算雷竹林地上生物量[J]. 林业科学, 2011, **47**(9): 1 – 6.
XU Xiaojun, ZHOU Guomo, DU Huaqiang, *et al.* Estimation of aboveground biomass of *Phyllostachys praecox* forest based on landsat thematic mapper image [J]. *Sci Silv Sin*, 2011, **47**(9): 1 – 6.
- [25] 琚存勇, 邸雪颖, 蔡体久. 变量筛选方法对郁闭度遥感估测模型的影响比较[J]. 林业科学, 2007, **43**(12): 33 – 38.
JU Cunyong, DI Xueying, CAI Tijiu. Comparing impact of two selecting variables methods on canopy closure estimation [J]. *Sci Silv Sin*, 2007, **43**(12): 33 – 38.
- [26] 李崇贵, 赵宪文, 李春干. 森林蓄积量遥感估测理论与实现[M]. 北京: 科学出版社, 2006.
- [27] LI Xinchuan, ZHANG Youjing, BAO Yansong, *et al.* Exploring the best hyperspectral features for LAI estimation using partial least squares regression [J]. *Remote Sens*, 2014, **6**(7): 6221 – 6241.
- [28] JIN Xiuliang, MA Jianhang, WEN Zhidan, *et al.* Estimation of maize residue cover using Landsat-8 OLI image spectral information and textural features [J]. *Remote Sens*, 2015, **7**(11): 14559 – 14575.
- [29] 王惠文, 吴载斌, 孟洁. 偏最小二乘回归的线性与非线性方法[M]. 北京: 国防工业出版社, 2006.
- [30] DU Huaqiang, ZHOU Guomo, GE Hongli, *et al.* Satellite-based carbon stock estimation for bamboo forest with a non-linear partial least square regression technique [J]. *Int J Remote Sens*, 2012, **33**(6): 1917 – 1933.
- [31] PETER T, WOLTER A PATA, BRIAN R, *et al.* Estimation of forest structural parameters using 5 and 10 meter SPOT-5 satellite data [J]. *Remote Sens Environ*, 2009, **113**(9): 2019 – 2036.
- [32] FRANCO-LOPEZ H, EK AvR, BAUER M E. Estimation and mapping of forest stand density, volume, and cover type using the k -nearest neighbors method [J]. *Remote Sens Environ*, 2011, **77**(3): 251 – 274.
- [33] KATILA M, TOMPPO E. Selecting estimation parameters for the finnish multisource National Forest Inventory [J]. *Remote Sens Environ*, 2001, **76**(1): 16 – 32.
- [34] MCROBERTS R E, TOMPPO E O, FINLEY A O, *et al.* Estimating areal means and variances of forest attributes using the k -Nearest Neighbors technique and satellite imagery [J]. *Remote Sens Environ*, 2007, **111**(4): 466 – 480.
- [35] KAJISA T, MURAKAMI T, MIZOUE N, *et al.* Estimation of stand volumes using the k -nearest neighbors method in Kyushu, Japan [J]. *J For Res*, 2008, **13**(4): 249 – 254.
- [36] CHIRICI G, MURA M, MCINERNEY D, *et al.* A meta-analysis and review of the literature on the k -nearest neighbors technique for forestry applications that use remotely sensed data [J]. *Remote Sens Environ*, 2016, **176**(2): 282 – 294.
- [37] KATILA M, TOMPPO E. Selecting estimation parameters for the finnish multisource National Forest Inventory [J]. *Remote Sens Environ*, 2001, **76**(1): 16 – 32.