ECOLOGIA BALKANICA
2015, Vol. 7, Issue 1 June 2015 pp. 93-102

Estimation and Mapping Forest Attributes Using "k Nearest Neighbor" Method on IRS-P6 LISS III Satellite Image Data

Roya Abedi**, *Amir Eslam Bonyad

Department of Forestry, Faculty of Natural Resources, University of Guilan, IRAN.
* Corresponding author: abedi.roya@yahoo.com

Abstract. In this study, we explored the utility of k Nearest Neighbor (kNN) algorithm to integrate IRS-P6 LISS III satellite imagery data and ground inventory data for application in forest attributes (DBH, trees height, volume, basal area, density and forest cover type) estimation and mapping. The ground inventory data was based on a systematic-random sampling grid and the numbers of sampling plots were 408 circular plots in a plantation in Guilan province, north of Iran. We concluded that kNN method was useful tool for mapping at a fine accuracy between 80% and 93.94%. Values of k between 5 and 8 seemed appropriate. The best distance metrics were found Euclidean, Fuzzy and Mahalanobis. Results showed that kNN was accurate enough for practical applicability for mapping forest areas.

Keywords: Forest attributes, IRS, k Nearest Neighbor (kNN).

Introduction

Monitoring forest attributes as an important tool to sustainable forest management need up-to-date information on forest resources (LABRECQUE *et al.*, 2006; MOHAMMADI *et al.*, 2011). Ground field inventory is the most frequently way to get forest cover information but it is not the optimal quick way (JIA *et al.*, 2014). Therefore, nowadays methods of estimations are used by remote sensing data (MOHAMMADI *et al.*, 2011). Remote sensing refers to indirect measurement of emitted electromagnetic energy using sensors (AHAMED *et al.*, 2011). Remote sensing has been efficient and effective means to obtain forest cover information, mapping and monitoring forest cover in recent decades due to its large scale information collection ability and synoptic and repeated coverage especially if combined with field data (REESE *et al.*, 2002; JIA *et al.*, 2014). There are a wide range of methods to integrate remote sensing data with ground inventory data (LATIFI *et al.*, 2010; CHERNETSKIY *et al.*, 2011).

Classification of satellite data for mapping forest attributes is the most frequent use of them (IVERSON *et al.*, 1989). A classification procedure may be supervised (training samples are labeled), semi supervised (only some of training samples are labeled) or non-supervised (the training samples are not labeled and these classes are not known a priori) (LIU *et al.*, 2011; SOUZA *et al.*, 2014). In recent years, hardwares and softwares have improved to process digital satellite data (IVERSON *et al.*, 1989), mapping forest species, illustrating distribution different vegetation species in a study area and providing essential indices for forest inventory and management. These maps are mostly generated with image classification methods. All of classification methods are object or pixel base, hard or soft classifier, parametric or non-parametric (TORABZADEH *et al.*, 2014).

One of the simplest and also more sophisticated nonparametric techniques that are used to link between field inventory data and remote sensing data is k nearest neighbor (kNN) classification method. kNN finds a group of k objects in the training set that are closest to the test object, and bases the assignment of a label on the predominance of a particular class in this neighborhood. Therefore, there are three key elements of this approach: 1) a set of labeled objects (stored records such as ground sampling plots data), 2) a distance or similarity metric to compute distance between objects and 3) the value of k (number of nearest neighbors). To classify an unlabeled object such as pixels of satellite image data, the distance of that to the labeled object is computed, its k nearest neighbors are identified and the dominant class of these nearest neighbors are used to determine the class label of the object (WU *et al.*, 2008). kNN expands terrestrial point observations to spatially explicit coverage by utilizing similarities in the spectral image space of remote sensing data (STUMER *et al.*, 2010). It is a simple classifier to use the k training samples that are closest to the test sample to classify it (XU *et al.*, 2013).

Satellite remote sensing data and kNN have played a key role in forest studies and have several applications in the field of forest inventory (MUINONEN *et al.*, 2001; OHMANN *et al.*, 2014). JUNG *et al.* (2013) improved the accuracy of estimation with moderate spatial resolution satellite imagery and the kNN algorithm for forest carbon stock estimation due to its simplicity and feasibility. ZHU & LIU (2014) studied the accurate mapping of forest types using kNN method. OHMAN *et al.* (2014) studied integrating forest inventory plot data and satellite image data for regional forest mapping using kNN method. WILSON *et al.* (2012) used kNN imputation method to mapping tree species over large areas using forest inventory plots and moderate resolution raster data. Also, kNN was used to estimate tree biomass (NYSTROM *et al.*, 2012), tree volume (CHIRICI *et al.*, 2008; KAJISA *et al.*, 2008), basal area and density (BAFFETTA *et al.*, 2009).

This study was investigated estimation forest attributes and accurate mapping using kNN method with IRS (Indian Remote-Sensing Satellite) imagery data. In this study, we evaluated the potential of the IRS-P6 LISS III satellite data and performance of kNN method as a non-parametric supervised classifier to create forest attributes mapping.

Material and Methods

The study area was a plantation that located in west of Guilan province in the north of Iran in 37° 32' to 37° 36' N latitude and 49° 2' to 49° 7' E longitude. It has total area of 1850.44 hectare and covered a smooth topography (Fig. 1.). This plantation characterized by even aged pure conifers stand (*Pinus taeda* (loblolly pine) as dominant coniferous species, pure deciduous stand (*Alnus subcordata* (caucasian alder), *Alnus glutinosa* (common alder) and *Populous* spp. (poplar)) and uneven aged mixed deciduous stand (*Acer velotonium* (persian maple), *Pterocarya fraxinifolia* (Caucasian wing nut), *Populus caspica* (caspian poplar) and *Carpinus betulus* (hornbeam)). Therefore, forest types were determined based on dominant tree species in sampling plots and five forest cover types were assigned to classification consist of 0: non-Forest, 1: *Pinus taeda*, 2: *Alnus glutinosa*, 3: *Populous* spp. and 4: mixed deciduous species.

408 circular sampling plots with 1000 m^2 (0.1 ha) area were distributed according to a systematic-random design and a network grid with 150m×200m spacing. Sampling plots were measured in the September to November of 2012. In each sampling plot diameter at the breast height (DBH) of all trees species with a DBH ≥ 7.5 cm was measured. Two trees height for one the nearest tree to the sampling plot center and another one that have the largest DBH in the sampling plot were recorded.

An IRS-P6-LISS satellite image data collected on July 18 2008 (Path 67; Row 43) was orthorectified with a digital elevation model (created from elevation of the 1:25000 national topographic database) and georeferenced by 20 ground control points (UTM, WGS 84, zone 39 N) (Table 1).

We used kNN imputation in the yaImpute package (Crookston & Finley, 2008) in statistical software R (R Development Core Team, 2010) and the kNN-forest software (Chirici *et al.*, 2012) in the Idrisi Selva 17.0 to data analyses consist of forest inventory and satellite images data combination, search optimum number of k and distance metric calculation. The dependent variables were calculated from field inventory data including volume, basal area, density and forest cove types and independent variables were defined by satellite images band's average digital numbers (i.e. pixel value) in the center of sampling plots. Spectral distance was derived using three common different distance metrics Euclidean, Mahalanobis and Fuzzy distances (Fig. 2.).

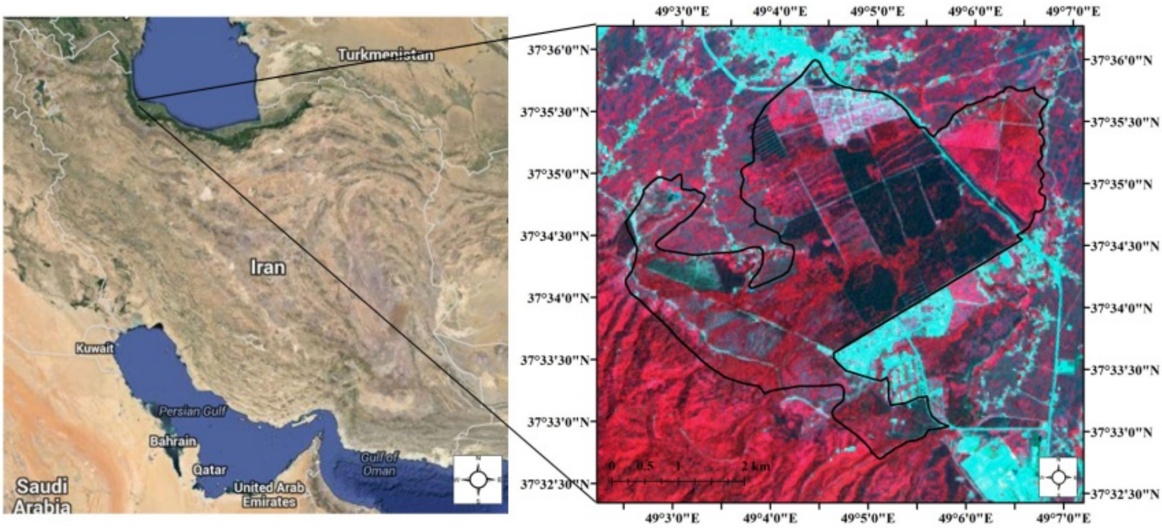

Fig. 1. Location of the study area.

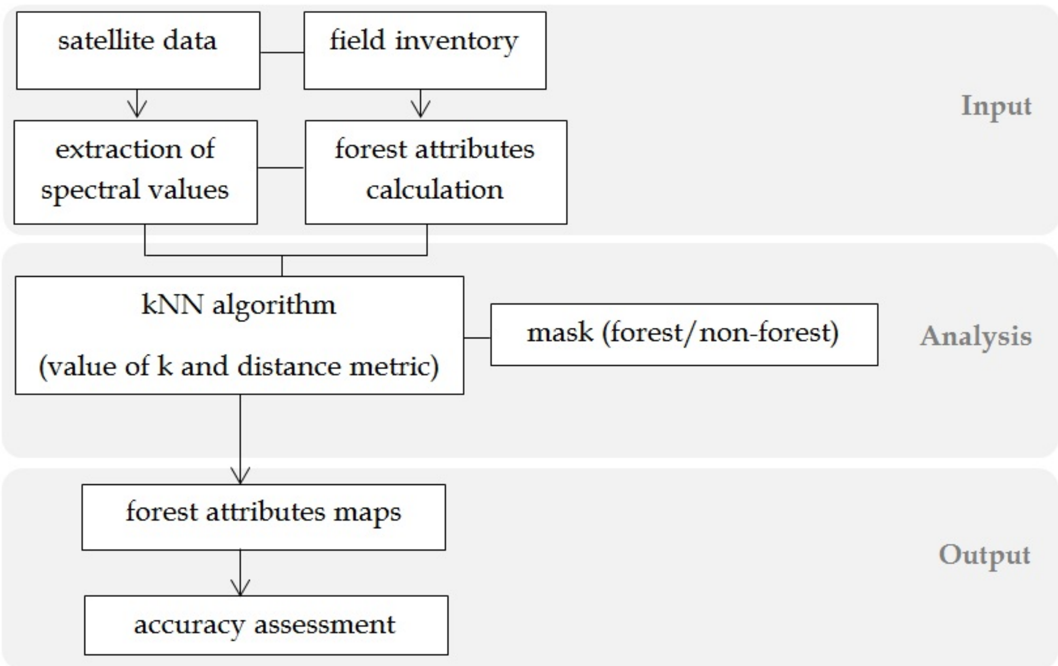

Fig. 2. Flowchart of the procedure used in kNN method.

Table 1. Detailed information of satellite image used in this study.

Satellite	Sensor	Bands	Wavelength (μm)	Spatial resolution (m)	Radiometric resolution (bit)
IRS-P6 (Resourcesat)	LISS-III (Linear Imaging Self-Scanning Sensor-III)	2 (G) 3 (R) 4 (NIR) 5 (mid IR)	0.520-0.590 0.620-0.680 0.770-0.860 1.550-1.700	23.5	7

The accuracy of the number of neighbors (k), distance metrics and forest attributes estimations were evaluated using Leave One Out Cross Validation that define as omitting training sample units one by one. For each omission, apply the kNN prediction to the remaining sample and summarize the error. In totality, it is applied the prediction rule n times and predicts the outcome for n units and estimate of error called Root Mean Square Error (RMSE) (Equation 1) (FRANCO-LOPEZ *et al.*, 2001). We used leave one out cross validation test for different values of k (the numbers of nearest neighbors) from 1 to 25.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \qquad (1)$$

where $\hat{y}_i$ is the estimated forest attribute and y_i is the forest attribute value measured in field sampling plots.

All preliminary analysis was performed in PCI Geomatica 9.1 and ARCGIS 10 softwares.

Results

Results showed that DBH and tree height have strongly correlation (r=0.743) with each other, and DBH has moderate correlation with volume and basal area (0.594 and 0.531, respectively). Tree height has moderately correlation to volume and basal area (0.629 and 0.591, respectively) and volume has strongly correlation to basal area (r=0.805).

Correlation analysis among spectral bands information and forests attributes showed a weak correlation in all cases except band 2 (Green) and 4 (NIR). Green band has the highest correlation for height, volume and DBH, respectively. NIR band was the band that shows the highest correlation for DBH, height, volume and basal area, respectively. Table 2 illustrates the Pearson correlation between the various bands and forest attributes for all sampling plots data.

The results of the optimum number of neighbors (k) and distance metrics have showed in Fig. 3. (a to f). The lowest value of cross validation RMSE as the best obtained in the estimation of Euclidean distance metric was in k=6 for DBH (Fig. 3., a), Euclidean distance metric in k=8 for tree height (Fig. 3., b), Fuzzy distance metric in k=5 for volume (Fig. 3., c), Mahalanobi distance metric in k=6 for basal area (Fig. 3., d), Euclidean distance metric in k=6 for both density and forest cover type (Fig. 3. e and f, respectively).

In all cases, there was a steep decrease in RMSE in the first numbers of neighbors then fixed the value of error.

kNN classification map of DBH, tree height, volume, basal area, density and forest cover types was produced and showed in Fig. 4. (a to f). The results of accuracy assessments including overall accuracy and kappa coefficient have showed in Table 3. Results showed that the amount of accuracy was high in all maps (from 80% to 93.94%) and map of density have the highest amount of accuracy (93.94%).

Discussion

The purpose of this study was to investigate the possibility of kNN method to estimate main forest attributes including trees DBH, height, volume, basal area, density and forest cover type based on existing forest inventory field sampling data as reference data to produce highly accurate

Table 2. Pearson correlation coefficient between attributes and mean digital numbers of sampling plots.

Attributes	DBH	H	V	BA	D	T	Mean digital number by band			
							B2	B3	B4	B5
DBH	1									
H	0.743	1								
V	0.594	0.629	1							
BA	0.531	0.591	0.805	1						
D	-0.364	-0.076	0.205	0.384	1					
T	-0.323	-0.215	-0.187	-0.164	0.074	1				
B2	-0.374	-0.478	-0.389	-0.391	-0.076	0.088	1			
B3	-0.036	0.006	-0.020	-0.016	0.017	0.044	0.043	1		
B4	-0.503	-0.417	-0.391	-0.332	0.167	0.196	0.226	-0.006	1	
B5	-0.066	-0.106	-0.078	-0.085	-0.034	-0.030	0.024	0.000	-0.007	1

Note: DBH: diameter at the breast height (cm), H: tree height (m), V: volume (m³/ha), BA: basal area (m²/ha), D: density (tree count/ha), T: forest cover type, B2: band 2, B3: band 3, B4: band 4, B5: band 5.

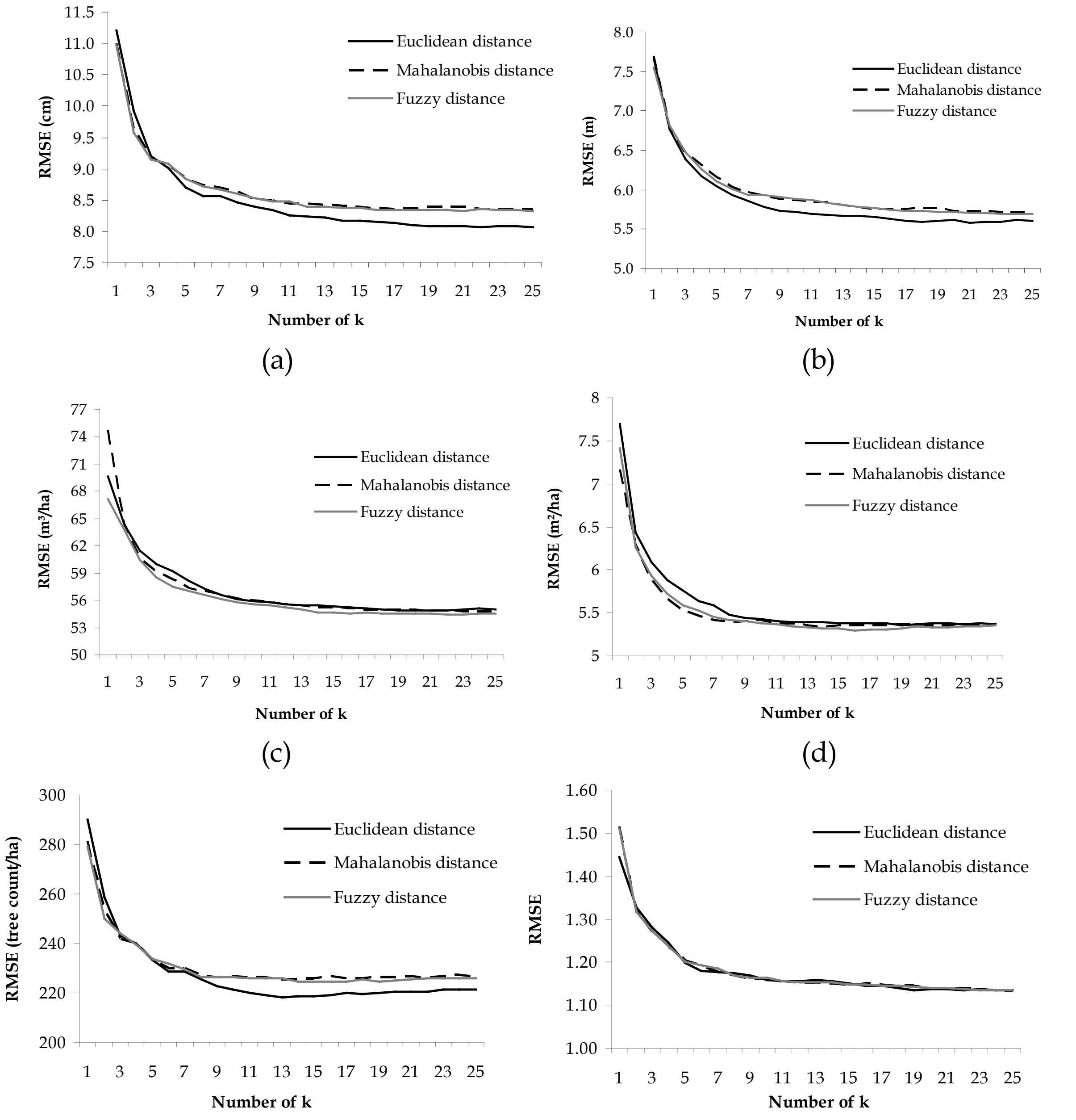

Fig. 3. Relation between k and RMSE for DBH (a), tree height (b), volume (c), basal area (d), density (e) and forest cover types (f) estimation.

Fig. 4. kNN classification maps of DBH (a), tree height (b), volume (c), basal area (d), density (e) and forest cover types (f).

Table 3. Accuracy assessment of mapping forest attributes.

	DBH	Height	Volume	Basal area	Density	Type
Overall accuracy (%)	87.88	80.0	88.89	80.0	93.94	89.90
Kappa coefficient (%)	0.61	0.61	0.72	0.57	0.79	0.59

classification maps. The visualization of forest variables is an important tool and easy way for decision making (PEREIRA, 2006, Italy – pers. comm.). Land cover map as a visual tool has basic information to management of natural resources (TRIEPKE *et al.*, 2008).

The first step in kNN method application is the calibration of kNN parameters to obtain the most appropriated distance and number of neighbors (k) for each variable that to be predicted (PEREIRA, 2006, Italy – pers. comm.). Therefore, we computed the optimum numbers of neighbors (k) and optimum distance metric in the best set for estimating and mapping. Consequently, forest attributes map was prepared using optimum values for every attributes separately in our study area. According to the same author the application of kNN method to other forest areas should always be done with prior kNN parameters calibration (k, distance) to local data.

We found that k less than 10 was the best in our study area. This result is similar to other researches in different forest conditions (RAHMAN, 2006; WILSON *et al.*, 2012; JUNG *et al.*, 2013). Increasing the number of neighbors produces an undesirable reduction of variability in predicted values, especially if the goal of using kNN is map production. Clearly, this problem is reduced with use of fewer neighbors (FRANCO-LOPEZ *et al.*, 2001). In all cases, there was a steep decrease in RMSE in the first numbers of neighbors then fixed the value of error. This behavior was reported by several studies. OHMANN *et al.* (2014) reported that maps assessed with k=5 were much more accurate than when assessed with k=1. It is clear that there was a rapid early gain in overall precision with the addition of neighbors (FRANCO-LOPEZ *et al.*, 2001).

Another essential in kNN method is satellite image correctly rectified and coordinates of the field plots were accurately determined until satellite rectification error and sampling error have the lowest amount and accuracy of created maps have been acceptable (JUNG *et al.*, 2013; OHMANN *et al.*, 2014). The main criterion to classification by kNN is the minimum spectral distance between target pixels and reference pixels. Therefore, the integration of forest inventory plot data with remotely sensed data provide a powerful tool for mapping of forest attributes (OHMANN *et al.*, 2014). Suitable methods to combine remote sensing and terrestrial sample based inventory data is the main difficulties in the estimation procedure (RAHMAN, 2006). The use of different remotely sensed data and various inventory sampling plots procedure showed that kNN method is a feasible classification technique. This method has the potential for estimating various attributes by different image data. kNN method was well studied in recent years to integrate ground information and remote sensing data (SOUZA *et al.*, 2014; ZHU & LIU, 2014; KAJISA *et al.*, 2008; FRANCO-LOPEZ *et al.*, 2001).

Error of kNN estimation was determined by RMSE (root mean square error). Almost, all studies on kNN method were analyzed by this criterion. RMSE is a diagnostic measure of difference between observed and predicted values, which is most commonly utilized for model evaluation. This measure has a relatively high weight to large errors because the errors are squared before being averaged. Thus, RMSE is most useful when large errors are particularly undesirable. RMSE= {[Σ(P−O) 2]/N}0.5; P: predicted value; O: observed value (HAMADA *et al.*, 2011). But, some researchers believed that there was not a reliable and sufficient method of error evaluation in the kNN estimation and also, further studies should be carried out to overcome this limitation (RAHMAN, 2006). However, this algorithm is being improved in terms of accuracy and efficiency.

Results of correlation between spectral band information and forest variables showed weak correlation due to heterogeneous conditions of forest stands (PEREIRA, 2006, Italy – pers. comm.).

Our results showed that kNN has good results in pure stands of conifers and deciduous species and also the mixture of deciduous species in our study area. This result is consistent with CAVARAVDICH (2007, Germany – pers. comm.). Therefore, high accuracy of our study maps may be due to the kNN approach strength and IRS-P6 LISS III image data strength. We concluded that forest attributes maps can be easily produced with kNN with accuracy estimation and level of prediction error (PEREIRA, 2006, Italy – pers. comm.). kNN method was well studied in recent years to integrate ground information and remote sensing data. In general, kNN classification method applied to estimate and mapping forest attributes in order to assess accurate estimations in this study. Results indicated that kNN method could estimate and mapping forest attributes by suitable accurate classification and produce feasible thematic maps.

References

AHAMED T., L. TIAN, Y. ZHANG, K.C. TIKNG. 2011. A review of remote sensing methods for biomass feedstock production. - *Biomass and Bioenergy*, 35(7): 2455-2469. DOI:10.1016/j.biombioe.2011.02.028.

BAFFETTA F., L. FAKTTORINI, S. FRANCESCHI, P. CORONA. 2009. Design-based approach to k-nearest neighbours technique for coupling field and remotely sensed data in forest surveys. - *Remote sensing of environment*, 113(3): 463–475. DOI:10.1016/J.RSE.2008.06.014.

CHERNETSKIY M., I. PASKO, A. SHEVYRNOGOV, N. SLYUSAR, A. KHODYAYEV. 2011. Study of forest vegetation dynamics in the south of the Krasnoyarskii Krai in spring. - *Advances in space research*, 48: 819-825. DOI:10.1016/j.asr.2011.04.032.

CHIRICI G., A. BARBATI, P. CORONA, M. MARCHETTI, D. TRAKVAGLINI, F. MASELLI, R. BERTINI. 2008. Non-parametric and parametric methods using satellite images for estimating growing stock volume in alpine and Mediterranean forest ecosystems. - *Remote sensing of environment*, 112: 2686-2700. DOI:10.1016/j.rse.2008.01.002.

CHIRICI G., P. CORONA, M. MARCHETTI, A. MASTRONARDI, F. MASELLI, L. BOTTAI, D. TRAVAGLINI. 2012. *K-NN FOREST* (Data analysis software system), User's guide and tutorial. Computer software. Available at: [http://www.forestlab.net].

CROOKSTON N.L., A.O. FINLEY. 2008. yaImpute: An R package for kNN imputation. - *Journal of Statistical Software*, 23: 1–16.

FRANCO-LOPEZ H., A.R. EK, M.E. BAUER. 2001. Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. - *Remote sensing of environment*, 77: 251– 274. DOI:10.1016/S0034-4257(01)00209-7.

HAMADA Y., D.A. STOW, D.A. ROBERTS. 2011. Estimating life-form cover fractions in California sage scrub communities using multispectral remote sensing. - *Remote sensing of environment*, 115: 3056–3068. DOI:10.1016/j.rse.2011.06.008.

IVERSON L.R., R.L. GRAHAM, E.A. COOK. 1989. Applications of satellite remote sensing to forested ecosystems. - *Landscape Ecology*, 3: 131-143.

JIA K., S. LIANG, L. ZHANG, X. WEI, Y. YAO, X. XIE. 2014. Forest cover classification using Landsat ETM+ data and time series MODIS NDVI data. - *International Journal of Applied Earth Observation and Geoinformation*, 33: 32-38. DOI:10.1016/j.jag.2014.04.015.

JUNG J., S. KIM, S. HONG, K. KIM, E. KIM, J. IM, J. HEO. 2013. Effects of national forest inventory plot location error on forest carbon stock estimation using k-nearest neighbor algorithm. - *ISPRS Journal of Photogrammetry and Remote Sensing*, 81: 82–92. DOI:10.1016/j.isprsjprs.2013.04.008.

KAJISA T., T. MURAKAMI, N. MIZOUE, F. KITAHARA, S. YOSHIDA. 2008. Estimation of stand volumes using the k-nearest neighbor method in Kyushu, Japan. - *Journal of forest research*, 13(4): 249–254. DOI:10.1007/s10310-008-0077-5.

LABRECQUE S., R.A. FOURNIER, J.E. LUTHER, D. PIERCEY. 2006. A comparison of four methods to map biomass from Landsat-TM and inventory data in western Newfoundland. - *Forest ecology and management*, 226: 129-144. DOI:10.1016/j.foreco.2006.01.030.

LATIFI H., A. NOTHDURFT, B. KOCH. 2010. Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: application of multiple optical/LiDAR-derived predictors. - *Forestry*, 83(4):395-407. DOI: 10.1093/forestry/cpq022.

LIU K., W. SHI, H. ZHANGA. 2011. A fuzzy topology-based maximum likelihood classification. - *ISPRS journal of photogrammetry and remote sensing*, 66(1): 103-114. DOI:10.1016/j.isprsjprs.2010.09.007.

MOHAMMADI J., S. SHATAEE, M. BABANEZHAD. 2011. Estimation of forest stand volume, tree density and biodiversity using Landsat ETM+ Data, comparison of linear and regression tree Analyses. - In: *Procedia Environmental Sciences, 1st Conference on Spatial Statistics*. Netherlands. 23.-25. March 2011. Enschede, Netherlands, pp. 299–304.

MUINONEN E., M. MALTAMO, H. HYPPANEN, V. VAINIKAINEN. 2001. Forest stand CHARACTERISTICS estimation using a most similar neighbor approach and image spatial structure information. - *Remote sensing of environment*, 78(3): 223– 228. DOI:10.1016/S0034-4257(01)00220-6.

NYSTROM M., J. HOLMGREN, H. OLSSON. 2012. Prediction of tree biomass in the forest–tundra ecotone using airborne laser scanning. - *Remote sensing of environment*, 123: 271–279. DOI:10.1016/j.rse.2012.03.008.

OHMANN J.L., M.J. GREGORY, H.M. ROBERTS. 2014. Scale considerations for integrating forest INVENTORY plot data and satellite image data for regional forest mapping. - *Remote sensing of environment*, 151: 3-15. DOI:10.1016/j.rse.2013.08.048.

R CORE TEAM. 2010. R: *A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, [www.R-project.org]

RAHMAN M.M. 2006. Tropical forest biomass estimation and mapping using k-nearest neighbour (KNN) method. - In: S. Nayak, S.K. Pathan. J.K. Garg (Eds.): *Proceedings of Symposium of ISPRS Commission IV.* India. 25.-30. September 2006, Goa, India, pp. 860-865.

REESE H., M. NILSSON, P. SANDSTROM, H. OLSSON. 2002. Applications using estimates of forest parameters derived from satellite and forest inventory data. - *Computers and electronics in agriculture*, 37(1-3): 37-55. DOI:10.1016/S0168-1699(02)00118-7.

SOUZA R., L. RITTNER, R. LOTUFO. 2014. A comparison between k-Optimum Path Forest and k-Nearest Neighbors supervised classifiers. - *Pattern recognition letters*, 39: 2–10. DOI:10.1016/j.patrec.2013.08.030.

STUMER W., B. KENTER, M. KOHL. 2010. Spatial interpolation of in situ data by self-organizing map algorithms (neural networks) for the assessment of carbon stocks in European forests. - *Forest ecology and management*, 260: 287–293. DOI:10.1016/j.foreco.2010.04.008.

TORABZADEH H., F. MORSDORF, M.E. SCHAEPKMAN. 2014. Fusion of imaging spectroscopy and airborne laser scanning data for characterization of forest ecosystems – A review. - *ISPRS Journal of Photogrammetry and Remote Sensing*, 97: 25–35. DOI:10.1016/j.isprsjprs.2014.08.001.

TRIEPKE F.J., C.K. BREWER, D.M. LEAVELL, S.J. NKOVAK. 2008. Mapping forest alliances and associations using fuzzy systems and nearest neighbor classifiers. - *Remote sensing of environment*, 112(3): 1037–1050. DOI:10.1016/j.rse.2007.07.014.

WILSON B.T., A.J. LISTER, R.I. RIEMANN. 2012. A nearest-neighbor imputation approach to mapping tree species over large areas using forest inventory plots and moderate resolution raster data. - *Forest ecology and management*, 271: 182–198. DOI:10.1016/j.foreco.2012.02.002.

WU X., V. KUMAR, J.R. QUINLAN, J. GHOSH, Q. YANG, H. MOTODA, G.J. MCLACHLAN, A. NG, · B. LIU, P.S. YU, Z.H. ZHOU, M. STEINBACH, D.J. HAND, D. STEINBERG. 2008. Top 10 algorithms in data mining. *Knowledge and information systems*, 14: 1–37. DOI:10.1007/s10115-007-0114-2.

XU Y., Q. ZHU, Z. FAN, M. QIU, Y. CHEN, H. LIU. 2013. Coarse to fine K nearest neighbor classifier. - *Pattern recognition letters*, 34(9): 980–986. DOI:10.1016/j.patrec.2013.01.028.

ZHU X., D. LIU. 2014. Accurate mapping of forest types using dense seasonal Landsat time-series. - *ISPRS Journal of Photogrammetry and Remote Sensing*, 96: 1–11. DOI:10.1016/j.isprsjprs.2014.06.012.

Received: 15.04.2015
Accepted: 27.05.2015