Combining RapidEye Satellite Images and Forest Inventory Data for Assesment of Forest Biomass

Viktor Myroniuk[1], Andrii Bilous[2]

Forest Mensuration and Forest Inventory Department, National University of Life and Environmental Sciences of Ukraine, UKRAINE, Kyiv, Heroyiv Oborony street 15, E-mail: victor.myroniuk@nubip.edu.ua

Forest Mensuration and Forest Inventory Department, National University of Life and Environmental Sciences of Ukraine, UKRAINE, Kyiv, Heroyiv Oborony street 15, E-mail: bilous@nubip.edu.ua

Abstract – The paper presents the results of estimation of growing stock volume and live biomass in forest stands using combination of forest inventory measurements, multispectral satellite images RapidEye and digital elevation model (DEM). In a context of classification of remote sensing data we considered two nonparametric methods – k-Nearest Neighbors (k-NN) and Random Forest (RF). We concluded that RF outperforms kNN method nevertheless both of them provide quite accurate estimation of mean value of growing volume in a range of ±5 $m^3 \cdot ha^{-1}$, different components of aboveground biomass - ±1–2 $t \cdot ha^{-1}$.

Key words – forest, biomass, RapidEye, *k*-NN imputation, random forest.

I. Introduction and Methods

Remote sensing data serve as support for forest inventory therefore development of modern techniques based on satellite imagery plays the dominant role in forest inventory. Since early 90-th satellite imagery widely used in sample-based national forest inventory. Basic principles of classification of satellite images in context of estimation of forest parameters were developed by finnish scientists [7]. Among the advantages of their method, known as k-Nearest Neighbors (k-NN), is that it allows combining field measurements with remote sensing data, different topographic and climatic maps (digital elevation models (DEM), soil maps etc.). One of the successful example of application of k-NN is the Finish multi-source forest inventory (MS-NFI) carried out in Finland since 1989. Similar multi-source tachniques have also been tested in different conditions in a number of other countries (USA, Canada, Norway, Sweden, Germany, Italy etc.).

According to R. E. McRoberts the role of k-NN method in forest inventory is fourfold: 1) imputation of missing values in forest databases; 2) mapping; 3) small area estimation; 4) model-based inference of forest attributes.

The main idea of k-NN method is that value of some attribute that corresponds to relatively small forest area can be obtained as a response variable in classification model of satellite images. It is proved that pixel-level prediction of forest attributes is sutable for representing of forests in continuous form. Estimated values of certain characteristics in scale of separate pixel could be easily grouped to form descrete objects, such as forest stands. Produced maps are used widely in national forest inventory as well as for operational forest managment.

Also other classification methods that do not have strict requirements for distribution of study parameters in population and are known in computer science as mashine learning algorithms should be mentioned (Random Forest (RF), Support Vector Machine (SVM), Neural Networks (NN)) More successful in forest inventory is method RF [2].

II. Input datasets

The research was conducted on a territory of Snov district of Chernigov region where we established test poligon with a total area of 45 km. The forest area was divided into forest compartments following instructions of forest inventory act. According to this 14 forest blocks were organised. After stand-wise inventory detailed forest characteristics of each compartment were estimated and forest database was filled.

Forest stands cover 41.8 % (1881,0 ha) of total area of test polygon. The main forest forming species are as following: pine (44.7 %), birch (39.8 %), black alder (13.1 %), aspen (0.7 %), oak (0.6 %), ash (0.5 %), black lockust (0.3 %), spruce 90.1 %), maple (0.1 %), linden and poplar (0.1 %).

A georeferenced data set comprised 5-bands RapidEye image (date acquired – 2016.06.07, spatial resolution – 5 m) and DEM (spatial resolution – 10 m). All data we reprojected to WGS 84 / UTM zone 36N referencing system and resampled to 5 meters resolution.

The classification model were developed using attribute information from forest database. We used RF classifier to create forest mask that included such landuse categories: forest areas covered by forest, agricultural fields, grasslands, wetlands, water bodies.

We used random sampling to locate training dataset over the territory. Inclusion of forest stand to sample depended on species composition. For this purpose we used only pure forest stands with at least 80 % presence of main species. According to structure of forest fund for classification we used 5 tree species (Pinus sylvestris L., Betula pendula Roth, Alnus glutinosa L., Populus tremula L., Quercus robur L.). In general, training dataset for forest masking included 5150 cases, but for tree species classification – arround 2600. We used mathematical models [1] for modelling of aboveground biomass of forest stands formed by birch, black alder and aspen, equations for pine and oak forest were taken from (Shvidenko, Schepaschenko, Nilsson, & Bouloui, 2007) [6].

For mapping of spatial distribution of growing stock volume and aboveground biomass of forest stands using k-NN method separate training dataset was created. Usually for this task forest mensuration data is used collected on fixed- or variable-radius (relascope) plots. Besides pixel-level classification model, stand-level models and mapping strategy could be used. Some authors [4] are pointing out the benefits of forming training dataset using mean values of pixels aggregated within stand. Therefore this we estimated mean values of spectral features for 90 pure forest stands formed by birch, black alder, aspen and pine. In addition to

5 spectral bands of RapidEye image we included elevation from DEM into training dataset as independent variable.

III. Results

One of the initial step of imputation of forest attributes using k-NN method is selecting optimal number of nearest neighbors and estimation of distance between them. The simplest measure of distance are Euqlidean (EUC) or Mahalanobis (MAL) distance between spectral characteristics of candidate and refference pixels. But in the practise different measures are also possible, such as most similar neighbor (MSN), gradient nearest neighbors (GNN) or neighborhood according to independent component analysis (ICA). Crookston N. L. and A.O. Finley [3] in {yaimpute} package for R proposed using RF algorithm to find nearest neighbor. Authors proposed to use this algorithm for modeling continuous variables in case of large number of variables in training dataset.

We investigated performance of different classification models (EUC, MAL, MSN, GNN, ICA, RF) and different number of nearest neighbors (from 1 to 20). The precision of approaches we tested by means of scaled root mean square differences (RMSD) between observed and predicted (imputed) data. We concluded that RF model was more precise followed by MSN. Thus RF model we used only for prediction and mapping of growing stock volume.

Classification model were calculated in R using {yaImpute} and {randomForest} package.

To develop classification model MSN for prediction of growing stock volumes and aboveground biomass we investigated how RMSD value was changing if number of neighbors was increasing. The smallest values of RMSD were obtained when k = 4, in other cases this values are significantly greater.

In general MSN model provided adequate estimation of growing stock volumes, but for forest stands that have less than 50 $m^3 \cdot ha^{-1}$ prediction will be overestimated. The situation of estimating aboveground biomass very similar. Aplication of RF impoved accuracy. Application of RF improved accuracy of estimation, almost all observation were located near 1 : 1 line.

Estimation of mean values of forest attributes for test area was very close to those obtained from forest database. The results proved the general peculiarities of k-NN method: the most precise results were obtained for whole test area, for smaller units such as forests stands accuracy was decreased.

The best results in estimation were obtained for softwoods (table 1). Deviations in estimation of growing stock volumes and biomass for oak forests are explained by small area of such stands (0.6 %). In case of pine stands overestimation of forest parameters we can explain by large areas of naturally regrown forests on former agricultural fields that have non-homogeneuos spatial structure.

TABLE 1

MEAN VALUES OF GROWING STOCK VOLUME OBTAINED FROM FOREST DATABESE AND CLASSIFICATION

Tree species	Growing Stock Volume, $m^3 \cdot ha^{-1}$	
	database	RF
Birch	140	136
Black alder	119	119
Oak	234	120
Aspen	179	176
Pinus	235	292
Other	140	94
Mean	162	158

Conclusion

Combination of RS data, DEM and other georeferenced datasets with limited number of field measurement shoul be considered as one of possible approaches of forest inventory. Application of k-NN and RF methods for classification of RapidEye images provides unbiased estimation of mean values of growing stock volumes and aboveground biomass of forest stands for large geographical regions. Besides estimation of forest attributes both methods could be used for mapping and production of forest related cartographical materials.

References

[1] Bilous A. M. Biological productivity and ecosystem functions of softwood deciduous forests in the Ukrainian Polissya : The Manuscript : 06.03.02 , 06.03.03 / Bilous Andrii – Kyiv, 2016. – 423 p.

[2] Breiman L. Random Forest / L. Breiman // Machine Learning. – 2001. – Vol. 45. – № 1. – P. 5–32.

[3] Crookston N. L. yaImpute: An R Package for ¬kNN Imputation / N. L. Crookston, A. O. Finley // Journal of Statistical Software. – 2008. – Vol. 23. – Issue 10. – 1–16.

[4] Imputing forest structure attributes from stand inventory and remote sensed data in Western Oregon, USA / A. T. Hudak, A. T. Haren, N. L. Crookston et al. // Forest Science. – 2014. – Vol. 60. – Issue. 2.– P. 253–269.

[5] McRoberts R. E. Estimation forest attribute parameters for small areas using nearest neighbors techniques / R. E. McRoberts // Forest Ecology and Management. – 2012. – Vol. 272. – P. 3–12.

[6] Tables and models of growth and productivity of forest of forming species of Northern Eurasia (standard and reference materials) – M.: 2006. – 803 p.

[7] Tomppo E. Satellite image-based National Forest Inventory of Finland / E. Tompo // International Archives of Photogrammetry and Remote Sensing. – 1991. – Vol. 28: 1–7. – P. 419–424.