Bulletin of Environment, Pharmacology and Life Sciences
Bull. Env. Pharmacol. Life Sci., Vol 2 (10) September 2013: 79-83
©2013 Academy for Environment and Life Sciences, India
Online ISSN 2277-1808
Journal's URL:http://www.bepls.com

ORIGINAL ARTICLE

Forest Types Classification Using CART Algorithms and SPOT-HRG Data

Asghar Fallah[1], Syavash Kalbi*[2] and Shaban Shataee[3]
[1] Department of Forestry, Sari University of Agricultural Sciences and Natural Resources, Sari, Iran
*[2] Department of Forestry, Sari University of Agricultural Sciences and Natural Resources, Sari, Iran,
[3] Department of Forestry, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran
Email: S.kalbi63@gmail.com

ABSTRACT
Forest type mapping, is one the most necessary elements in the forest management and silviculture treatments. Traditional methods such as field surveys are almost time-consuming and cost-intensive. Improving on satellite data sources and classification methods are preparing new opportunities for obtaining the accurate forest biophysical attributes maps. This research compares performance of three non-parametric and tree-based algorithms i.e. the Classification and Regression Tree (CART), for general forest type mapping using semi-high resolution of SPOT-HRG data. Using systematic random sampling design in a small area of the Hyrcanian forests, tree and shrubs species in 150 sample plots were registered. Naming of the general forest types for sample plots were done based on frequency of dominant species. After geometric and atmospheric corrections of SOPT-HRG data, suitable image processing transformations was done on main bands to produce general vegetation indices and principal components. Results showed CART algorithms has overall accuracy 60% and kappa statistics 51%. Overall, using SPOT-HRG data is so appropriate in the studies, which the map type is considered as a base map with maximum number of existing type in the area.
***Keywords**: Forest types classification, Hyrcanian forest, SPOT-HRG.*

Received 20.07.2013 Accepted 29.08.2013

INTRODUCTION

The forest stand type map show spatial distribution of tree and shrub species as a group or stand in forest ecosystem, so the preparing a precise forest stand type map will be important for knowing of forest status. Traditional methods such as field surveys are almost time-consuming and cost-intensive. Satellite data and their potential are offering new tools for managing and mapping the forest-covered area. The advances in remote sensing technology together with improvements in estimation and classification algorithms, could prepared opportunities for improving the retrieval of on time information with increased efficiency. The remote sensing data are alternatively producing from fine to coarse spatial resolutions, which are generally grouped to low resolution (like MODIS), medium resolution (like TM or ETM+), and high resolution (like IKONOS) with different spectral wavelengths. Investigations on capabilities of these data for different applications are the main interesting for scientists and managers. To be or not be capable of these data for different subjects are relating to some factors such as forest condition i.e. structure and composition (homogenous or heterogonous) in together with topography conditions. The Hyrcanian forests have different compositions and structures that differentiate them with other forests in the world.

In other hand, in the mixed hardwood of Hyrcanian forests, the previous studies [1-4] showed that medium resolution of ETM+/TM spectral data were not accurately sufficient to classify forest types due to be species heterogeneous of Hyrcanian forests. Generally, detailed and precise mappings can be improved by enhanced spatial and spectral resolution data sources [5]. One of the semi-high resolution remote sensing data are prepared by HRG subsystems of SPOT5 satellite. The HRG subsystem, provide images with ten meters resolution in three green, red and infrared spectral wavelengths and 20 meters in middle infra red spectral wavelengths.

The conventional parametric statistical classification techniques that have been successfully used in remote sensing data analyses for over four decades are not appropriate for forest type classification [6]. In recent years, the non-parametric algorithms such as decision tree based algorithms [7] have been widely used due to their simple interpretation, high classification accuracy, and ability to characterize

complex interactions among variables [8]. Many studies have shown that non-parametric methods provide better classification results. In some studies such as Sarunas, it is demonstrated that even with small training samples, non-parametric classification algorithms provide better results than parametric ones. Classification tree analysis (CTA) is a rule-based technique that has produced highly accurate classifications using a variety of spectral and ancillary data sources [9].
Therefore, the aim of this study was mapping forest type using SPOT -HRG data and CART algorithm in the Darabkola forest, located at the Hyrcanian forest in Mazandaran province, northern Iran.

MATERIAL AND METHODS

Study area

The study area is located at the Hyrcanian forests, Mazandaran Province, district 1 of Darabkola's forests, northern of Iran (Figure 1). The Darabkola's forestry plan, with about 2500 hectare area, is a natural and mature forest with uneven age and dense to semi dense stands. Elevation is ranged from 140 to 920 meters from free sea level and general aspect of study area is northern, but with some fine different slops aspects. The forestry practices in this area are selective cutting and plantation establishment.

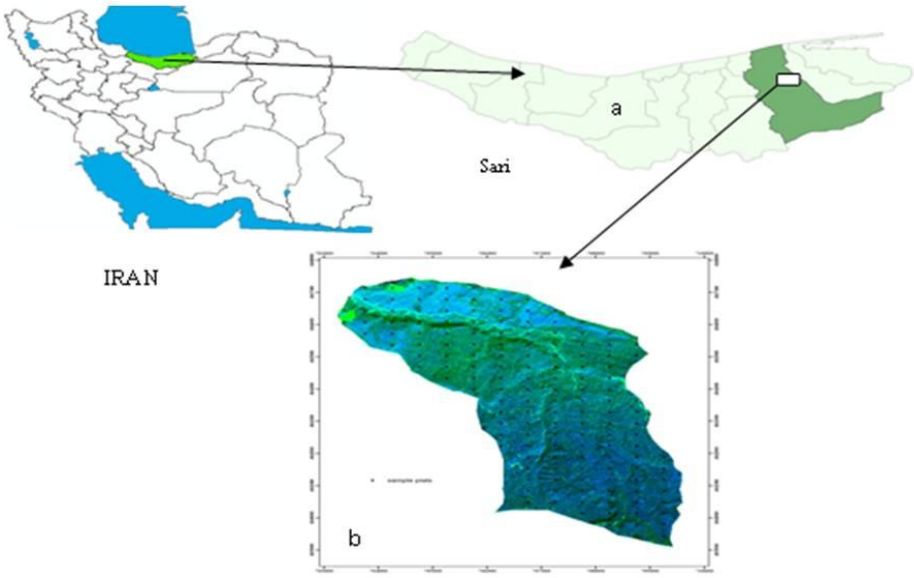

Figure 1 Location of the study area in the Mazandaran Province (a), allocation of sample plots (b) in the study area.

Field data

In summer 2010, using a systematic aligned sampling design with 350*500 m intervals, the 150 sample plots with 3600 m2 area were allocated in the study area (Figure 1). The geographical center of plots was accurately registered using high quality handy GPS and averaging methods to get accurate positions. In all samples, tree and shrub species of all trees with DBH greater than 7.5 cm were registered. Determination and naming of forest types were done based on computing frequency of dominant species in each plot. In the study area four general forest stand types including pure Fagus (PF), mixed Fagus (MF), mixed Carpinus (MC) and mixed hardwood stands (MH) were recognized.

Satellite data

Corresponding to study area, a small window of the SPOT 5 HRG XS scene acquired on1 June 2009 was used for forest type classification. The HRG subsystem provides images with the pixel sizes of 10 meters in green, red and near infrared (VNIR) bands, and 20 meters in the shortwave infrared (SWIR) band.
The SPOT-HRG data were accurately orthorectified by 10 meters spatial resolution DEM and 23 ground control points collected by handy GPS. The total root mean square errors (RMSE)
Were 0.67 for the VNIR bands and 0.5 for the SWIR band using second polynomial equation? The SWIR bands were also resized 10 meters using nearest-neighbor resampling method. The geometric precision of images was also approved using road vector layer and GPS collected control points.
Reflectance of the objects recorded by satellite sensors is generally affected by atmospheric absorption and scattering, sensor target illumination geometry and sensor calibration [5]. Processing of remote sensing data was performed by extracting different feature sets using some suitable band ratios to produce some famous vegetation indices as well as a standardized principal component analysis (PCA) transformation (Table1) due to their capabilities in exploring the forest biophysical attributes.

Table 1: Some used vegetation indices examined in this study.

Vegetation index	Formula	Reference
Stress Index (SI)	Red/ NIR	Jiang et al, 2003
Differential Vegetation Index (DVI)	NIR-RED	Tucker,1979
NDVI	NIR-Red/ NIR+Red	Rouse et al 1973
Moisture Stress Index (MSI)	SWIR/NIR	Rock et al, 1986
Simple Ratio (SR)	NIR/Red	Birth and Mcvey,1968
Normalized Difference Water Index (NDWI)	NIR-SWIR/NIR+SWIR	Gao,1996

Methods

Classification and regression tree (CART)

Classification and regression tree (hear after CART) algorithm, is a statistical procedure introduced by Breiman et al. [10], is primarily used as a classification tool, where the objective is to classify an object into two or more populations [11]. The underlying principle behind CART is to identify increasingly homogeneous configurations of predictive variables that should lead to increasingly homogenous configurations of target variables. Different types of predictive variables (categorical and continuous) can be imported into CART model [12]. The CART methodology consists of three steps including tree growing, tree pruning, and selecting the optimal tree. Initially an over fitting tree is grown by recursive partitioning of the data. In the second step that is called tree pruning, the sequence of nodes that should be eliminated to obtain a set of smaller trees is found. The last step is selection an optimal tree from the pruned trees. CART builds an overgrown tree based on the node purity criterion that is later pruned back via cross validation to avoid over fitting. In this study, measure of Gini impurity was used for categorical target variables.

Feature selection

Although, feature selection is not necessary in RF [7] and BRT [13], but in some classification algorithms such as CART, high number of independent variables may be effect on the classification results and selection of the best variables for classification can be leads to produce the better results. In addition, in some feature selection algorithms, the variables can be sorted based on their importance in classification process. The variable importance enables us to determine what set of variables is deemed important for each of the three methods and to compare them to see whether the sets are similar. The importance values were calculated by following formula:

$$I(j) = \sum_t \Delta_S(j, t) \qquad (1)$$

Where I (j) is the importance of variable x_j and $\Delta_S(j, t)$ is the reduction in mean squared error S that would be achieved if node t of the tree were split using x_j (Breiman et al., 1984).

Accuracy assessment

To evaluate performance of a classifier using methods, which are mentioned in the preceding chapter, it requires that a randomly selected set of test or unused samples (pixels) for each type class be used for computing the classification accuracy [6]. In this study, accuracy assessment was performed using 50 test samples. The classified images were then crossed with the test data to generate error matrices and calculation of the various accuracies including overall accuracy and kappa coefficient statistics.

RESULTS

As the Table 2 shows, NIR band was one of the most important variables, together with SWIR band and MSI indices.

Table 2: Variable importance for forest type mapping using CART algorithm

index	SWIR	NIR	MSI	PCI4	SVR	NDVI	PCI3
Importance	1.0000	0.987937	0.927903	0.902989	0.895310	0.830605	0.647921

Classification performances

Table 3 shows the summary of performance results including overall accuracy and kappa statistics of tree classification algorithms. Results CART method had overall accuracy of 60% and kappa coefficient of 0.51.

Table 3: Summary of accuracy results of the three classifiers

Classifier	Overall accuracy (%)	kappa statistics
CART	0.60	0.51

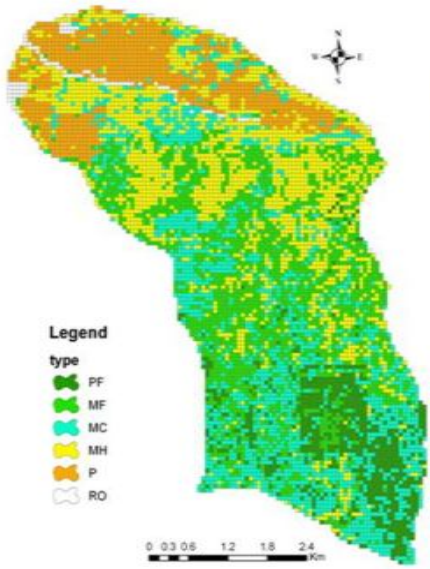

Figure 2: Classification maps obtained by CART algorithm

DISCUSSION

Obtaining detailed information about the amount of forest type's area is an important issue for practical forest management. In a comparative study, capability performance CART algorithms was investigated for forest type mapping using semi high resolution imagery of SPOT-HRG satellite data in the Darabkola's forest, as a case study area in the Hyrcanian forests. The classifications were performed using CART algorithms due to their advantages against parametric methods. Results were done using tow currently accuracy indices including overall accuracy and Kappa coefficient. The forest type maps generated by image classification will be applicable, if the classification accuracies are not known [14]. Thus, accuracy assessment is a fundamental principle in assuring the quality of thematic maps for their intended application [15].

Results of feature selection and variable importance showed that NIR band was the important variable for mapping and separating the forest type. In this spectral wavelength, the reflections of tree species are more enhanced and distinguishable. These results are similar to other studies [4] where researcher demonstrated that NIR band has high importance for segregation forest types.

In compare to studies that were done in the Hyrcanian forest, our results showed that overall accuracies obtained in this study were 60%, which are higher than previous studies that used parametric classification methods. For example, Abbasi [1] could mapped forest types with an overall accuracy of 44.6%; Shataee [2] with overall accuracy of 54.8%; Darvishsafat *et al*, [3] with overall accuracy of 51% and Rashidi *et al*, [4] with overall accuracy of 53.22%. One of the reasons to obtain the better result in our study compared to previous study is referring to use the nonparametric tree classifier methods. When we use non-parametric classifiers, it is not required to assume that the data follow a normal distribution and no statistical parameters are needed to separate image classes [16].

REFERENCE

1. Abbasi M. (2001). Investigation on the possibility of beech forest type mapping using Landsat ETM+ data (Case study: Khyroud forest), M.Sc thesis, Tehran University, 114p.
2. Shataee SH. (2003). Investigation on the possibility of beech forest type mapping using satellite data (Case study: Khyrood forest). PhD Thesis Tehran University, 155p.
3. Darvishsefat A A, Abbasi M, Marvi Mohadjer M R. (2009). Investigation on the possibility of beech forest type mapping using Landsat ETM+ data (Case study: Kheyroud forest), 2(9), 105-113.
4. Rashidi F, Babaie Kafaki S, Oladi J. (2009). Investigation on the capability of digital data of ETM+ sensor in separating of forest types (Case study: Lafoor area of Savadkooh). *Iranian Journal of Forest and Poplar Research* 17(13), 51-63.
5. Mahiny A S, Turner B J. (2003). Modeling Past change in vegetation through Remote and GIS: A Comparison of Networks and logistic Regression Methods. Geocomputation conference, Southampton, UK.
6. Richards J A, Jia X. (1999). Remote Sensing Digital Image Analysis, An Introduction, Third, Revised and Enlarged Edition. Springer, Berlin.

7. Breiman L, Cutler A. (2003). Setting up, using, and understanding Random Forests v4.0. [Online] URL: http://www.stat.berkeley. edu/users/breiman/rf.html.
8. Cutler D R, Edwards T C, Karen J, Beard H, Cutler A, Hess K T, Gibson J, Lawler J. (2007). Random forests for classification in ecology. *Ecology* 88 (11), 2783–2792.
9. Lawrence R, Burn A, Powell S, Zambon M. (2004). Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. *Remote Sensing of Environment* 90, 331-336.
10. Breiman L, Friedman, J H, Olsen R A, Stone C J. (1984). Classification and regression trees. Wadsworth and Brooks/Cole, Monterey, California, USA.
11. Tian-Shyug L, Chiu C C, Chou Y C, Lu C J. (2006). Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics & Data Analysis* 50, 1113–1130.
12. Cheng W, Zhang X, Wang K, Dai X. (2009). Integrating classification and regression tree (CART) with GIS for assessment of heavy metals pollution, *Environ Monit Assess* 158, 419–431.
13. Prasad M A, Iverson L R, Liaw A. (2006). Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems* 9, 181–199.
14. Baatuuwie N B, Leeuwen I L V. (2011). Evaluations of three classifiers in mapping forest stand types using medium resolution imagery: a case study in the Offinso Forest District, Ghana. *African Journal of Environmental Science and Technology* 5(1), 25-36.
15. Stehman S V, Czaplewski L R. (1998). Design and analysis of thematic map accuracy assessment: fundamental principles. *Remote Sensing Environment* 64, 331-344.
16. Qi J, Chehbouni A, Huete A R, Kerr Y H, Sorooshian S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment* 48, 119-126.
17. Jiang Y, Carrow R N, Duncan R R. (2003). Correlation analysis procedures for canopy spectral reflectance data of seashore paspalum under Traffic stress. *Journal of American Society. Hort* 13, 187-208.
18. Tucker C J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment* 8, 127-150.
19. Rouse J W, Haas R H, Schell J A, Deering D W. (1973). Monitoring vegetation systems in the Great Plains with ERTS. In Third Earth Resources Technology Satellite-1 Symposium, 309-317.
20. Rock B N, Vogelmann J E, Williams D L, Vogelmann A F, Hoshisaki T. (1986). Remote detection of forest damage. *Bioscience* 36, 439-445.
21. Birth GS, McVey GR. (1968). Measuring the color of growing turf with a reflectance spectrophotometer. *Agronomy Journal* 60, 640-643.

How to cite this article
Asghar Fallah, Syavash Kalbi and Shaban Shataee. Forest Types Classification Using CART Algorithms and SPOT-HRG Data. Bull. Env. Pharmacol. Life Sci., Vol 2 (10) September 2013: 79-83