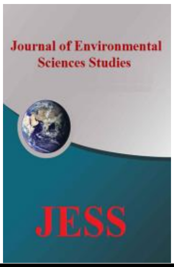

Journal of Environmental Sciences Studies

Journal home page: www.jess.ir

Comparison of Parametric and Non-Parametric Techniques to Accurate Classification of Forest Attributes on Satellite Image Data

Roya Abedi[a*]
a[*]. Assistant Professor, Department of Forestry, Ahar Faculty of Agriculture and Natural Resources, University of Tabriz, Iran
[*]Email Address : royaabedi@tabrizu.ac.ir

Received: (June -27- 2020)
Accepted: (October - 04 - 2020)

Abstract

Satellite images classification techniques obtain large information in forest areas. In this study, comparisons between k Nearest Neighbor (kNN) non-parametric method and Maximum Likelihood Classification (MLC) parametric method was performed in forest attributes consists of volume, basal area, density, and tree cover type estimation in the north forest of Iran. Results showed that kNN non-parametric method produces an accurate classification map in comparison to the MLC parametric method and the accuracy of kNN has the most amount in all attributes. Kappa coefficient estimation showed that the kNN method had the most amount of this coefficient in all attributes. Accordingly, the kNN non-parametric technique was identified as a feasible classification technique to produce forest attributes thematic maps.

Keywords: k nearest neighbour Maximum likelihood Non-Parametric method Parametric method

1. Introduction

Remote sensing data obtain large information in large forest areas easily by techniques that could provide great potential to improve the efficiency of forest information data (Barth et al., 2009; Huiyan et al., 2006). Environment and biodiversity monitoring, mapping forest attributes and many other practical applications using satellite images are some important applications of imagery classification (Li et al., 2017; Torres and Qiu, 2014; Schlund et al. 2017). In recent years, All of image classification algorithms and classification methods (object or pixel base, hard or soft and parametric or non-parametric) have been developed for improving the efficiency of methods to integrate remote sensing data and forest inventory data (Samaniego et al., 2007; Torabzadeh et al. 2014). Among various techniques, k Nearest Neighbors (kNN) as a non-parametric and MLC (Maximum Likelihood Classification) as a parametric technique have popularity in predict forest attributes by integrating the inventory field data and satellite images (McRoberts, 2012). These algorithms (both parametric and non-parametric) have mainly been used as per-pixel classifiers in remote sensing image analysis (Wieland and Pittore, 2014). The kNN technique is a simple technique that identifies the category of unknown data points based on its nearest neighbors whose class is known. In kNN, data is classified into training and sample data points. Distance is evaluated from all training points to sample point and the point with the lowest distance is called nearest neighbor. This method is very easy to implement but the value of k and distance metric affects the results (Dhanabal and Chandramathi, 2011). The Maximum Likelihood Classification is based on the probability density function and is a probabilistic statistical classifier. MLC classifies impute a pixel to a class that has the highest (maximum) probability. This method assumes that the statistical distribution of the features for each class is normal, and calculates the probability that a given pixel belongs to a specific class of the particular training site. Then, the pixels are assigned to the selected class that has the highest weight of probability. (Shataee Joibary et al., 2007; Tapakis and Charalambides, 2013). Parametric classifiers assume a normally distributed dataset and statistical parameters are directly inferred from training data. A parametric classifier, which has been widely used in remote sensing applications, is the MLC. Non-parametric classifiers do not assume a specific data distribution to separate a multi-dimensional feature space into classes (Wieland and Pittore, 2014).Both parametric and non-parametric estimation methods have been tested and applied in forest inventory applications in many cases. kNN and MLC were used in many studies and showed reliable results. Numerous studies of the relationship between spectral response and forest attributes have been conducted by kNN method during the past several decades such as mapping of various forest attributes (Zhu and Liu, 2014; Wilson et al., 2012), estimation tree biomass (Nystrom et al., 2012), tree volume (Chirici et al., 2008, Shataee, 2013), basal area and density (Franco-Lopez et al., 2001), improving forest mapping efficiency and accuracy (Finley et al., 2003; Gjertsen, 2007, Shataee, 2013) and land cover classification (Budreski et al., 2007; Noi and Kappas, 2017) . Also, many studies applied MLC to forest attributes classification (Lu et al., 2014; Chaplin and Brabyn, 2013; Kaliraj and Chandrasekar, 2012; Lu and Weng, 2007; Joshi et al., 2006; Lennartz and Congalton, 2004). The object of this study was the comparisons between non-parametric (kNN) and parametric (MLC) approaches by combining field inventory and remote sensing data in forest attributes estimation, classification, and mapping in a plantation in the north forest of Iran.

2. Materials and Methods

2.1. Study area

The study area was a plantation located in the west of Guilan province in Hyrcanian forests in the north of Iran. This plantation has a total area of 1850.44 ha and located in 37˚ 32΄ 40″ to 37˚ 36΄ 20″ N latitude and 49˚ 2΄ 20″ to 49˚ 7΄ 20″ E longitude (Figure 1) and covered a smooth topography. The study area characterized by even-aged pure conifers species consists of Pinus taeda and Cryptomeriya japonica (Pinus taeda was dominant) and pure broadleaf species stands consist of Alnus subcordata, Alnus glutinosa, and Populous spp. and uneven-aged species stand consist of mixture of broadleaf species consist of Quercus castaneaefolia, Acer velotonium, Pterocarya fraxinifolia, Populus caspica, and Carpinus betulus (Figure 1).

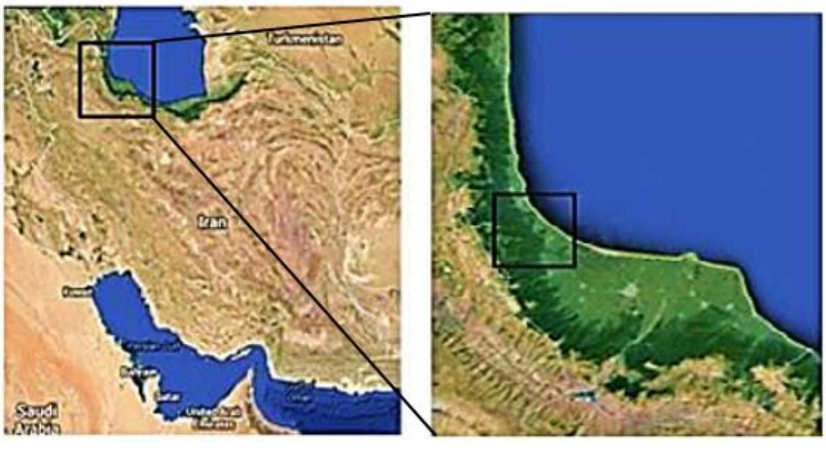

Figure 1. Location of the study area

2.2. Field and satellite data

408 circular sampling plots were distributed according to a systematic random design in a network grid with 150m×200m spacing. Each sampling plot was 1000 m2 area, and measurements of diameter at the breast height (DBH) of all trees species with a DBH ≥ 7.5 cm was measured in sampling plots. Locations of these plots were registered and navigated with handheld GPS (Global Positioning System). The operational inventory provided reference field data to estimate forest variables. We assumed that the field data to be a measurement of the ground truth.The 4 bands of the IRS P6-LISS III satellite images were used. The Image was rectified to the UTM national coordinate system by fitting a polynomial mathematical model to the coordinate of 20 control points. The geometric precision of images was also verified using the road vector layer and field-collected GPS control points.

2.3. Data analysis

The variables were calculated from field inventory data including volume, basal area, density, and tree cover type. Average digital numbers (i.e. pixel value) in the center of sampling plots as satellite images parameter extracted and combined with field data. The optimization of the number of nearest neighbors and distance metrics was done for the kNN method. This spectral distance was derived using various distance metrics including Euclidean, Mahalanobis, and Fuzzy distances which are the most common in the literatures. The accuracy of the number of neighbors (k) and forest attributes estimations was evaluated using RMSE percentage. kNN imputation was performed using the YaImpute package (Crookston and 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 (Clarks Labs, 2012). All preliminary analysis of ortho-rectifying, MLC parametric classification, and post-classification accuracy assessment was performed using PCI Geomatica version 9.1. As supervised classification methods were using MLC and kNN algorithms and Classification results were validated by means of the overall accuracy parameter and the Kappa coefficient.

3. Results & Discussion

The optimization results of the number of k and distance metrics in the kNN method for each forest attribute as a preliminary analysis were showed in Table 2. This optimized kNN was used in classification and mapping later.

Table 2. Optimization of kNN method for each forest attributes.

Satellite image	Attributes	Optimum No. k	Distance metric	rRMSE (%)
IRS P6-LISSIII	Volume	5	Fuzzy distance	57.6
	Basal Area	6	Mahalanobis distance	47.6
	Density	6	Euclidean distance	78.9
	Tree Cover Type	6	Euclidean distance	39.2

Results of mapping accuracy assessments showed that kNN non-parametric method produces accurate classification in comparison to the MLC parametric method. Figure 2 showed that the accuracy of the kNN non-parametric approach has the highest amount in all attributes. The highest accuracy was in tree cover type (89.90%) and followed by volume, density, and basal area (88.89%, 85.19%, and 80.00%, respectively) in the kNN method. kNN non-parametric approach improved the accuracy of attributes estimation 25.93% for volume, 23.23% for tree cover type, 22.86% for basal area, and 19.97% for density. Therefore, the highest improvement was shown in volume (Table 2). In the MLC method, the highest accuracy was in tree cover type (66.67%) and followed by density, volume, and basal area (65.22%, 62.96%, and 57.14%, respectively).

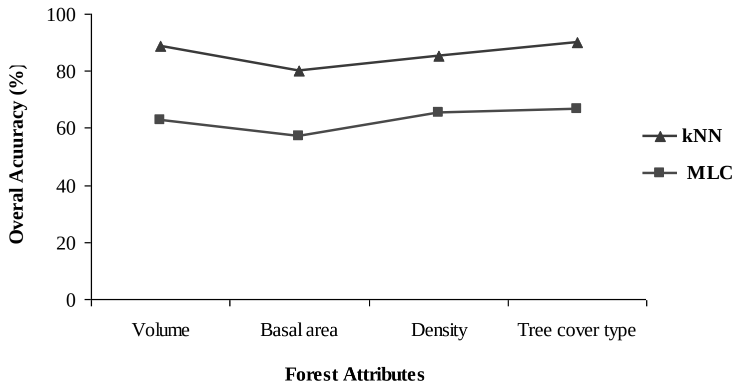

Figure 2. Accuracy assessment of kNN and MLC methods.

Kappa coefficient estimation showed that the volume estimation map has the highest Kappa coefficient in the kNN method (0.718) and the highest amount of it in the MLC method was accrued in tree cover type (0.584). Figure 3 showed that the kNN method had the highest amount of this coefficient in all attributes. The tree cover type had a similar kappa coefficient in two methods (0.593 for kNN and .584 by MLC). The highest difference of kappa coefficients between two methods was showed in volume (0.271) and followed by basal area (0.126), density (0.071), and tree cover type (0.009) (Table 3).

Table 3. Percentage of overall accuracy and Kappa coefficient in kNN and MLC methods

	kNN			MLC	
Attributes	**OA (%)**	**Kappa**		**OA (%)**	**Kappa**
Volume	88.89	0.72		62.96	0.45
Basal area	80.00	0.57		57.14	0.44
Density	85.19	0.56		65.22	0.49
Tree cover type	89.90	0.59		66.67	0.58

Note: OA: Overall Accuracy (%), Kappa: Kappa coefficient

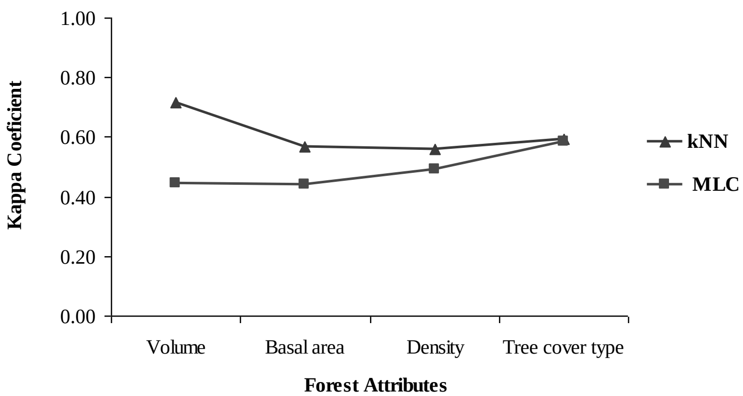

Figure 3. Kappa coefficient of kNN and MLC methods

The purpose of this study was to produce highly accurate forest attributes classification estimation. Therefore, both parametric and non-parametric estimation methods as two more relatively successful approaches have been tested in this study. Results showed that kNN non-parametric method had a higher amount of accuracy comparison of the MLC parametric method in all attributes. This is probably due to the features of the kNN method. The first step in kNN method application is the calibration of the dataset to obtain the most appropriate distance and the number of neighbors (k) for each variable (Pereira, 2006, Shataee, 2013; Noi and Kappas, 2017). The main criterion for classification by kNN is the minimum spectral distance between target pixels and reference pixels based on the similarity of spectral properties. Therefore, the integration of forest inventory plot data with remotely sensed data provides a powerful tool for estimating forest attributes (Ohmann et al., 2014). Therefore, we computed the optimum numbers of neighbors (k) and optimum distance metric in the best set for estimating attributes. Consequently, the forest attributes map was prepared using optimum values for every attribute separately in our study area. This optimization phase could improve kNN precision in comparison to MLC. This result confirms the results of Shataee (2013) that it showed the kNN could produce the best estimation in volume and basal area if it calibrates for k and distance method. kNN has several advantages including in the kNN method, the study area was initially classified in forest and non-forest areas then, forest area (as forest mask) were used for further analysis and classification. Also, Non-parametric methods (e.g. kNN) do not make assumptions about data distribution (Lemay and Temesgen, 2005). This ability is useful for mixed-species forests such as north forest of Iran that ground inventory data often have not a normal distribution. The results showed that the kNN technique could improve results considerably in comparison with MLC. Therefore, the high accuracy of our study may be due to the strength of both the kNN approach and IRS-P6 LISS III image data. We concluded that forest attributes maps could easily produce by the kNN approach with accuracy estimation and level of prediction error (Pereira, 2006). kNN method was well studied in recent years to integrate ground information and remote sensing data. In this study, Results indicated that the kNN method could estimate forest attributes by suitable accurate classification and later produce feasible thematic maps. All attributes were accordingly classified using the kNN and MLC method, having the overall accuracy above 80 and 58 percent respectively, and Kappa index of nearly above 0.5 and 0.4. Nonetheless, the need for accuracy measures depends on the user’s need (Tomppo et al., 2009).

4. Conclusions

Consequently, we identified the kNN non-parametric method as a suitable and feasible classification technique to combine remote sensing and terrestrial sample-based inventory data in the estimation procedure and production forest attributes thematic maps in this study area. Field sampling improvement or other remotely sensed image data could improve the ability of this method.

List of Abbreviations

kNN: k Nearest Neighbors; MLC: Maximum Likelihood Classification; OA: Overall Accuracy; DBH: diameter at the breast height; GPS: Global Positioning System; UTM: Universal Transverse Mercator coordinate system; IRS: Indian Remote Sensing.

References

- Barth, A., Wallerman, J., Stahl, G. 2009. Spatially consistent nearest neighbor imputation of forest stand data. Remote Sensing of Environment 113: 546-553.
- Budreski, K.A., Wynne, R., Browder, J.O., Campbell J.B. 2007. Comparison of Segment and Pixel-based Non-parametric Land Cover Classification in the Brazilian Amazon Using Multitemporal Landsat TM/ETM+ Imagery, Photogrammetric Engineering and Remote Sensing 73 (7): 813-827.

- Chaplin, J., Brabyn, L. 2013. Using remote sensing and GIS to investigate the impacts of tourism on forest cover in the Annapurna Conservation Area. Nepal Applied Geography, 43:159-168.
- Chirici, G., Corona, P., Marchetti, M., Mastronardi, A., Maselli, F., Bottai, L., Travaglini, D. 2012. K-NN FOREST: a software for the non-parametric prediction and mapping of environmental variables by the k-Nearest Neighbors algorithm. European Journal of Remote Sensing, 45: 433-442.
- CLARKS LABS. 2012. Idrisi Selva. Clark University. USA. Available on-line at: www.clarklabs.org (accessed 11 October 2013).
- Crookston, N.L., Finley, A.O. 2008. yaImpute: An R package for kNN imputation. Journal of Statistical Software, 23: 1–16.
- Dhanabal, S., Chandramathi, S. 2011. A Review of various k-Nearest Neighbor Query Processing Techniques International Journal of Computer Applications, 31 (7): 14-22.
- Finley, A.O., Ek, A.R., Bai, Y., Bauer, M.E. 2003. K-Nearest Neighbor estimation of forest attributes: improving mapping efficiency. In: Preceding of 5th Annual Forest Inventory and Analysis Symposium, 18-20 Novomber 2003, New Orleans, LA., pp. 62-68.
- Franco-Lopez, H., Ek, A.R., Bauer. M.E. 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.
- Gjertsen, A.K. 2007. Accuracy of forest mapping based on Landsat TM data and a KNN-based method. Remote Sensing of Environment, 110: 420–430.
- Huiyan, G., Limin, D., Gang, W., Dong, X., Shunzhong, W., Hui, W. 2006. Estimation of forest volumes by integrating Landsat TM imagery and forest inventory data. Science in China: Series E Technological Sciences, 49: 54-62.
- Joshi, Ch., De Leeuw, J., van Andel, J., Skidmore, A.K., Lekhak, H.D., van Duren, I.C., Norbu, N. 2006. Indirect remote sensing of a cryptic forest understorey invasive species Forest Ecology and Management 225: 245-256.
- Kaliraj, S., Chandrasekar, N. 2012. Spectral recognition techniques and MLC of IRS P6 LISS III image for coastal landforms extraction along south west coast of Tamilnadu, India. Bonfring International Journal of Advances in Image Processing, 2(3): 1-7.
- Lemay, V., Temesgen, H. 2005. Comparison of Nearest Neighbor methods for estimating basal area and stems per hectare using aerial auxiliary variables. Forest Science, 51: 109-119.
- Lennartz, S.P., Congalton, R.G. 2004. Indirect remote sensing of a cryptic forest understorey invasive species. Classifying and mapping forest cover types using IKONOS imagery in the northeastern United States. In: Proceedings of ASPRS Annual Conference May 2004, Denver, Colorado.
- Lu, Z., Im, J., Rhee, J., Hodgsond, M. 2014 Building type classification using spatial and landscape attributes derived from LiDAR remote sensing data. Landscape and Urban Planning, 130: 134–148.
- Lu, D., Weng, Q. 2007. A survey of image classification methods and techniques for improving classification performance. International Journal of Remote Sensing, 28:5: 823-870.
- Mcroberts, R.E. 2012. Estimating forest attribute parameters for small areas using nearest neighbors techniques. Forest Ecology and Management, 272: 3–12.
- Nystrom, M., Holmgren, J., Olsson, H. 2012. Prediction of tree biomass in the forest–tundra ecotone using airborne laser scanning. Remote sensing of environment, 123: 271–279.
- Noi, P.T., Kappas, M. 2018. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. Sensors, 18(1): 1-20.
- Ohmann, J.L., Gregory, M.J., Roberts, H.M. 2014. Scale considerations for

integrating forest inventory plot data and satellite image data for regional forest mapping. Remote sensing of environment, 151: 3-15.
- Pereira, C.R. 2006. Estimating and mapping forest inventory variables using the k-nn method: [23] Mocuba district case study–Mozambique. PhD dissertation, university of Tuscia, Italy. 86 pp.
- R Development Core Team. 2010. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing 3-900051-07-0 Available online at: www.R-project.org (accessed 22 June 2014).
- Samaniego, L., Bardossy, A., Schulz, K. 2007. Supervised Classification of Remotely Sensed Imagery using a Modified k-NN Technique. Transactions on Geoscience and Remote Sensing, 46: 2112-2125.
- Shataee, Sh. 2013. Forest attributes estimation using aerial laser scanner and TM data, Forest Systems, 22(3): 484-496.
- Shataee Joibary, Sh., darvishsefat, A.A., Kellenberger, T.W. 2007. Forest type mapping using incorporation of special models and ETM+ data. Pakistan journal of biological science, 10 (14): 2292-2299.
- Tapakis, R., Charalambides, A.G. 2013. Equipment and methodologies for cloud detection and classification: A review. Solar Energy, 95:392-430.
- Torabzadeh, H., Morsdorf, F., Schaepman, M.E. 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.
- Tomppo, E.O., Gagliano, C., De Natale, F., Katila, M., Mcroberts, R.E. 2009. Predicting categorical forest variables using an improved k-Nearest Neighbour estimator and Landsat imagery. Remote Sensing of Environment, 113: 500-517.
- Wieland, M., Pittore, M. 2014. Performance Evaluation of Machine Learning Algorithms for Urban Pattern Recognition from Multi-spectral Satellite Images. Remote Sensing, 6: 2912-2939.
- Wilson, B.T., Lister, A.J., Riemann, R.I. 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.
- Zhu, X., Liu, D. 2014. Accurate mapping of forest types using dense seasonal Landsat time-series. ISPRS journal of photogrammetry and remote sensing, 96: 1–11.