Forest Type Classification Improvement Using Spatial Predictive Distribution Models

Shaban Shataee Joibary[1] Ali .A. Darvishsefat[2]

[1]Assistance professor of Forestry, Gorgan University, 49165, Gorgan, Iran – shataee@gau.ac.ir
Associate professor of natural resources faculty, Tehran University, Karaj, Iran, adarvish@chamran.ut.ac.ir

Abstract-The last experiences showed that spectral data have not sufficient to classify forest types in the mountainous area. In order to clear abilities of spatial models to classify forest types and improve results, an investigation was planned in a case study in the northern forests of Iran by ETM+ data. The spatial models based on aspect, elevation, incorporated aspect-height and homogenous units constructed for each type individually. Probability occurrence rates of types were extracted in the each class. Classification was accomplished with the best spectral data sets by maximum likelihood classifier using only spectral data and with spatial models separately. The accuracy of results was assessed with a sample ground truth map. The results showed that spatial models could improve considerably results in compare with only spectral data (14%). This study exposed that spatial models based on the homogenous units in compare to other models could better improve classification.

Keywords: Forest Type, Classification, Spatial models, Improvement, Maximum Likelihood.

1. INTRUDUCTION

The Caspian zone forests, which also called the Hyrcanian forests, are the most valuable forests in Iran. They cover the northern slops and foothills of the Alborz Mountains. Iran has forests with an area of nearly 12.4 mil. ha, it has various geographic conditions, producing different forests of various tree and shrub species and production capacity in different edapho- climatic conditions (FAO, 2002). The species such as Beech (Fagus Orientalis), hornbeam (Carpinus Betelus), Alder (Alnus Glutinousa), Oak (Quercus castaneafolia), maple (Acer Velotonia), Ironwood (Parotia Persica) are main trees. These forests are very complex and mixed due to specific topographic, edaphic, climatic and ecological conditions. Mapping forest variables such as types and stands is fundamental for forest management. However, forest types mapping trough current fielding ways, is time consuming and cost-intensive. Using satellite Imagery and its potentials are posed tools for mapping forest-covered area. Forest extent mapping is almost possible by satellite data in the northern mountainous forests of Iran (Darvishsefat and Shataee, 1997). The next expectation for whom was feasibility investigation on forest types classification and producing executive forest types maps. The last results were exposed that discrimination of forest types that are compound only with one species, as pure types are very successful when uses satellite data (Walsh, 1980). While a forest type is comprised with two or many species in the forests such as study area, it will be difficult to separate them each other's (Shataee, et al., 2004). Yet spectral signatures used in supervised classifications may overlap considerably, making effective discrimination unachievable based on spectral reflectance characteristics alone. Make an attempt on the improvement of classification result was main objective for who were interested forest type mapping using satellite data. Many attempts were performed using different techniques such as rule-based classification (Bolstad and Lillesand, 1992), incorporated domain knowledge and using ancillary data (Hutchinson, 1982). Information from ancillary data sources has been widely shown to aid discrimination of classes that are difficult to classify using remote sensing data (Apisit, et al., 2000; Hopkins et al, 1988; Hutchinson, 1982; Strahler, 1980). In theses cases, ancillary data sources and expert knowledge's related to spatial distribution of types can provide useful information to help distinguish between inseparable classes. Using ancillary data related with forest could improve results (Brockhaus, et al., 1992; Franklin, 2001; Hopkins, et al., 1988). Determination and delineation of environmental factors, which have effective role on the spatial distribution of types or groups of homogeneous species, is first step to incorporate this non-spectral data with spectral data. Finding of suitable technique to integrate and incorporate these data is the last step. The distribution of forest types can be affected by general landscape characteristics, such as soil and micro climatic, as well as specific terrain related features such as elevation, slope, and aspect. These elements can be considered as indicators of species composition and distribution. Hence, the variables may be incorporated into prediction models to estimate likelihood of type's occurrence. These spatial predictive distribution models can help more accurate make decision to belong a class to a pixel by algorithm based on accurate location and distribution range of forest types. One of the ways to incorporate ancillary information is using prior probabilities of classes' memberships. Also, the Maximum Likelihood classifier is based on the probability density function associated with a particular training site signature. Pixels are assigned to the most likely class based on a comparison of the posterior probability that it belongs to each of the signatures being considered. The Maximum Likelihood is also known as Bayesian classifier, since it has the ability to incorporate prior knowledge using Bayes' Theorem (Richards, 1993):

$$P\ (h/e) = p\ (e/h)*\ p\ (h)\ /\ \sum P\ (e/h\ *\ p\ (h)$$

Where:
P (h|e) = the probability of the hypothesis being true given the evidence (posterior probability)
P (e|h) = the probability of finding that evidence given the hypothesis being true (derived from training data)
P (h) = the probability of the hypothesis being true regardless of the evidence (prior probability)
Bayesian model uses Bayes theorem to combine the information in the data with additional, independently available information (prior) to produce a full probability distribution (posterior distribution) for all parameters (Congdon 2001, Gelman et al. 1995, Carlin and Louis 2000).

When no knowledge exists about the prior probabilities with which each class can occur, it assumes that prior probabilities should be considered as equal for each class (Lo and Watson 1998). If there is reasonable knowledge of the expected proportional area of each class over the image as a whole, it can specify a prior probability value for each class (Pedley and Curran, 1991, Maselli, et al., 1995). In addition, prior probabilities can also be entered as a separate real number image (values between 0-1) for each class. This allows incorporating spatial models to determine prior probabilities of each class. In the mountainous forests and complex forest sites like this study area, forest types commonly are related with many variables. These variables will prepare specific ecological conditions for each type, which, are expressed as homogenous units. Determination of occurrence rate of each type in homogenous units and construction of this spatial model to incorporate with remote sensing data may be improve classification results more than use of each variable individually. Thus, specification of type characters in exchange for species characters can better help to construct spatial models. This paper presents application of integrating different spatial models, created with topographic parameters to improve the classification results. Investigation on how the spatial models incorporate with spectral data and how much can improve the classification results in comparison with only using spectral data were other objectives in this study.

2. METHODS

2.1 Study Area

The study area is located at research forest of Tehran University in the north of Iran between 51°33'12"E and 51°39'56" E longitude and 36°32'08" N and 36°36'45 5" N latitude. The study has performed on three districts that are about 3000 hectares (figure 1).

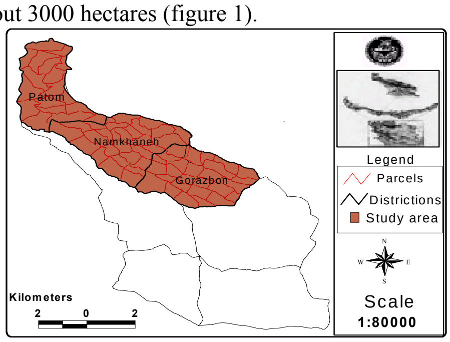

Figure 1: Location of study area in the north of Iran.

2.2 Data

In order to investigate ETM+ data potential for forest types mapping, a small window on 164-35 Scene from 2nd August 2000 was selected. In addition, some ancillary data extracted DEM such as aspect and elevation maps were resized to spatial resolution of satellite data.

2.3 Ground Truth

For accuracy assessment, a sample ground truth of forest types was designed and generated. The one-hectare area sample plots were distributed systematically throughout study area. In the each plot, diameter and kind of species all trees were measured. Finally, 193 plots were measured in the study area. Based on experiences, trough computing of 100 thick trees and percent of species frequency, kind of forest type has been determined in the each plot. In addition to plantation area, six forest types were recognized by this method.

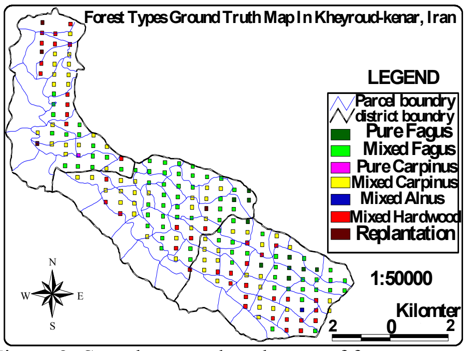

Figure 2: Sample ground truth map of forest types

2.4 Pre Processing and Processing

The ETM+ bands were geo-referenced in two steps. First, panchromatic band was geo-referenced by a digital 1:25000 map and ground control points. Then, other images have been geo-referenced using image to image technique. The final RMSe was found about 0.54 pixels. All images corresponding to ground truth map were resized to 10 meters resolution by second order transformation and cubic convolution re-sampling method. Some suitable processing image analyses were applied to create the new artificial bands. The Tasseled Cap calculation were accomplished to calculate brightness and greenness component, Principal Component analysis to extract the components which have more information and some suitable ratioing transformations which can be reduce the topographic effects. These bands were used together with ETM+ images as spectral data.

3. RESULTS

3.1 Classification with Spectral Data

In order to comparison of using spectral data only with integration of spectral and ancillary data, classification of ETM+ bands and some artificial bands was accomplished to extraction of forest types. For supervised classification, some pixels were selected randomly as training area for each type. Finally, the best bands set were selected based on spectral properties of training area by Bhathacharya index (table 1).

Table 1: The ETM+, Artificial Bands

ETM+ Bands	Artificial Bands
1, 2, 3, 4, 5,7 and Pan	PCA1, PCA2, PCA3, Brightness, Greenness, Ratio(NIR-G), Ratio(NIR/G), Ratio(NIR/R+G), Ratio(NIR-MIR/NIR+MIR), Ratio(NIR -R/ NIR +R)

Since, the maximum likelihood classifier has been reported as a suitable classifier (Hopkins, et al., 1988; Williams, 1992; darvishsefat, 1994; Shataee et. al, 2004), it was applied to classify forest types.

3.2 Determination of Parameters Related to Forest Types

To construct models, it should be first specified which parameters are more effective on the spatial distribution of types. This information can be extracted from different ways such as sample plots. Since a considerable sample ground truth has been prepared at this study, they have been extracted from these sample plots. In addition, some of forest researchers have expressed that topographic parameters have strength correlation with forest types (Asadollahi, 1987). Based on, a digital elevation model (DEM) was generated using 1:25000 map. The elevation, slop and aspect maps with giving classes were extracted from DEM. These maps were crossed with ground truth map to specify correlation between parameters and types. The primary results showed that elevation are more effective on forest types compared with other topographic parameters (table 2).

Table 2: Occurrence Range of Forest Types in Study Area

Forest types	Elevation (m)	Aspect	Slop
Pure Fagus	1100-1350	All aspect	0-40 %
Mixed Fagus	400-1300	All aspect	0-40 %
Pure Carpinus	700-800	Southern	0-40 %
Mixed Carpinus	400-1350	All aspect	0-60 %
Mixed Alnus	1100-1300	West southern	7-40 %
Mixed	0-1300	All aspect	0-100 %

Based on table, it has been certainly made that the distribution of forest types has not correlated with only one parameter and it has a complex relation with other parameters. At result, it seems that using of one-parameter spatial models with spectral data may not improve classification results more than two or multi- parameters models. With these concepts, it was concluded that multi-parameters topography spatial models like those constructed by homogenous units may be improve the results. Regard to these information, it was investigated that which of the spatial models constructed by each of this parameters can be improve the results in comparison with using spectral data only. Based on this, the spatial models were constructed for each parameter separately and or incorporated as multi parameters.

3.3 Classification with Aspect Spatial Model

In the natural forests, distribution of forest types also is correlated with aspect. In this study was tried to compute the prior probability of forest types based on aspect and was investigated to improve classification by using of spatial distribution models. An aspect classes map was extracted from DEM. The occurrences rates of forest types in the each aspect classes was computed as same as elevation (table 4).

Table 4: Occurrences Rates of Types in the each Aspect Class

Forest types/ Aspect classes	Replant area	Mixed	Mixed Alnus	Mixed Carpin	Pure Carpin	Mixed Fagus	Pure Fagus	Total
N	0.06	0.24	0	0.13	0	0.53	0.04	1
EN	0.02	0.12	0	0.11	0	0.61	0.14	1
E	0.07	0.32	0	0.15	0	0.46	0	1
ES	0.04	0.32	0	0.47	0.02	0.08	0.07	1
S	0	0.25	0.01	0.43	0.02	0.23	0.06	1
WS	0.01	0.22	0.01	0.38	0	0.36	0.02	1
W	0.02	0.23	0	0.26	0	0.46	0.03	1
WN	0.02	0.13	0	0.29	0	0.47	0.09	1
F	0	0	0	0.43	0	0.23	0.34	1

Based on these prior probabilities, for each forest type a spatial predictive model was created as image. Classification of forest types has been accomplished using integration of the best band set and spatial predictive imagery models.

3.4 Classification with Elevation Spatial Model

To determine the occurrence prior probability rates for each type in each height class (table 3), they were computed by:

$$P\ (f/h) = N\ (f/h) / \sum N\ (fi/h)$$

Which:
P (f/h) is probability of type A in height classes
N (f/h) is number pixel of type A in height classes
∑ N (fi/h) is total pixels number of type A in each height class

Table 5: Occurrences Rates of Types in the each height Class

Forest types/ Height classes	Replant Area	Mixed	Mixed Alnus	Mixed Carpinus	Pure Carpinu	Mixed Fagus	Pure Fagus	Total
0-100	0	1	0	0	0	0	0	1
100-200	0	0.63	0	0.37	0	0	0	1
200-300	0	0.83	0	0.17	0	0	0	1
300-400	0	0.37	0	0.63	0	0	0	1
400-500	0.12	0.27	0.05	0.35	0	0.21	0	1
500-600	0	0.46	0	0	0	0.54	0	1
600-700	0.03	0.36	0	0.28	0	0.33	0	1
700-800	0.02	0.13	0	0.42	0.05	0.38	0	1
800-900	0.03	0.12	0	0.57	0	0.28	0	1
900-1000	0	0.15	0	0.33	0	0.52	0	1
1000-1100	0	0.25	0	0.29	0	0.46	0	1
1100-1200	0.02	0.18	0	0.29	0	0.3	0.12	1
1200-1300	0	0.18	0.02	0.19	0	0.51	0.1	1
1300-1400	0	0.08	0	0.55	0	0.13	0.24	1

For this reason, the digital elevation model was classified to 100-meters classes. Resultantly, a spatial predictive model was created for each forest type (figure 3). These images had values, which showed prior probabilities rates for each forest type in the height classes. Classification has been accomplished using integration of the best band set and spatial predictive imagery models.

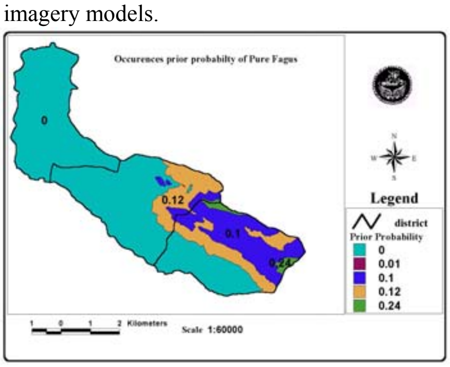

Figure 3: One of the height spatial models for pure fagus

3.5 Classification with Spatial Model Based on Integration of Aspect and Elevation

As it expressed before, distribution of forest types is note be related with only one parameter. It means that forest types can not be separable only using one index or parameter. At result, it seems that using multi-parameters spatial model may be improves better the classification results when it is integrated to spectral data. With this assumption, a spatial model was constructed by incorporating of aspect and elevation parameters. This spatial model was built as follow:

1) Adding aspect and height occurrence image that subdivided two created aspect-height occurrence images.
2) Height probability image was built on condition where each forest type is occurring will be “1” and other places will be “0”.
3) Spatial distribution model was obtained by multiplication of 1and (2) step images.

Therefore, aspect-height spatial models were created for each type separately. Classification has been accomplished using integration of the best bands and these spatial models.

3.6 Classification with Homogenous Units Spatial Model

Homogenous units are places where have equal conditions regarded to some variables such as aspect, height or slope. In the mountainous lands, some different aspects can be finding in the each height class. Regard to impact of aspect on the distribution of forest types, the occurrences rates of forest types can be also different in a given height class. With these reasons, the homogenous units were created by crossing of height and aspect class maps. Corresponding to last ways, the occurrence rates of types have been extracted. Then, homogenous spatial models were built for each type. Classification has been accomplished using integration of the best band set and homogenous unit spatial predictive imagery. The accuracy assessment of results was done with sample ground truth and was compared with result of classification using only spectral data (table 5).

TABLE 5: Results of Accuracy Assessment

Method/ Accuracy (%)	Only spectral data	Aspect spatial models	Height spatial models	Height-aspect spatial models	homogenous units models
Overall accuracy	49.68	56.28	57.65	58.34	60.87
Overall Kappa	27.5	34.08	34.78	36.56	41.22

4. DISCUSSION AND CONCLUSION

In compliance with others research results, this study are also showed that using only spectral data is not advantageous to classify classes in the mountainous places, where separating species are difficult. Resultantly, they should be integrated with non-spectral data or should be used other techniques. The to be low overall accuracy of 49.68 % or kappa coefficient of 0.275 for using only spectral data is confirmed these documents.

This study confirmed that using topographic data related with classes could improve results as it before reported by other researchers (Janssen et al., 1990). The primary results of showed that elevation parameter is more effective on the forest types distribution in comparison with other topographic parameters i.e. aspect and slop. Results showed that slop parameter is not a parameter which can stratify forest types.

As results are exposed, using ancillary data as prior probability imagery could imported into classification processes. Compared with spectral data, these spatial predictive models could improved classification results and increased the overall accuracy from 6.5 to 11 percent and kappa coefficient from 6.5 to 14 percent.

Using only spatial predictive model based on aspect could improve the overall accuracy about 8 percent that it was a significant increment in accuracy compared with spectral data. This increment generally refers to role of aspect parameter on accurate addressing of some types and increment of occurrence probability for them.

With constructing of spatial predictive model based on height parameter and integration this model with spectral data, it was specified that increment of overall accuracy was very poor, about 1 (%) more than aspect and there is no significant improvement compare to aspect. This result exposed that aspect and height have almost equal impact on the distribution of forest types. This refers to be equal impact on the formation of forest types or grouping establishment of species that comprise a forest type.

Incorporated spatial predictive model based on aspect and height considerably could not improve classification results. Although an about two percent increment have been found in the overall kappa compared with using of aspect and height spatial models but, results was not attractive.

Creating homogenous units based on aspect and height and using as spatial predictive model with spectral data showed that classification results could be significantly improved about 11 percent in overall accuracy and 14 percent in overall kappa. From these results can conclude that each spatial model, which can accurately specialized spatial addresses of forest types occurrences or determine the distribution of forest types on the appreciate related parameters, could have ability to improve results when integrated with spectral data.

In despite of the overall accuracy of both spectral data and integration of spectral data and spatial data results are generally low and insufficient to illustrate for application, but results emphasize on the considerable improvement. These hopeful results encourage us to investigate other techniques and methods that may improve classification results so that it would be feasible to apply for forest management.

Due to not to be access other parameters which, are related with spatial distribution of forest types such as soil information, future research can be performed with these kind of information or geographical knowledge to integrate with spectral data. Using other methods such as rule-based classifier or expert system should be investigated to increase results until an executive method can be obtained using satellite data.

REFERENCES:

-Apisit E. and Shrestha R. P., 2000, Application of DEM Data to Landsat Image Classification: Evaluation in a Tropical Wet-Dry Landscape of Thailand, Photogrametric Engineering & Remote Sensing, Vol. 66, NO 3.

-Asadollahi, 1987, Study on vegetation geographic and association in the Hyrcanian northwest forests, forestry politics of northern forest congress, 39p.

-Bolstad, P.V., and Lillesand, T.M., 1992, Improved classification of forest vegetation in northern Wisconsin through a rule-based combination of soils, terrain, and landsat thematic mapper data, Forest Scinece, 38(1): 5-20.

-Brockhaus, J. A, Khorram S., Bruck, R. I., Campbell, M. V., 1992, A comparison of Landsat TM and SPOT HRV data for use in the development of forest defoliation models, Int. J. Remote Sensing, Vol. 13,NO. 16: 3235-3240.
-Carlin,B.P., and T. A. Louis, 2000. Bayes and Empirical Bayes Methods for Data Analysis, Second edition. Chapman and Hall, New York.
-Congdon, P. 2001. Bayesian Statistical Modelling. John Wiley and Sons, Chichester.
-Darvishsefat, A. A., 1994, Einsatz und Fusion von multisensoralen satellitenbildddaten zur erfassung von waldinventuren, Ph.D. Thesis, Zurich University.
-Darvishsefat, A., Shataee SH., 1997, Digital Forest Mapping using of landsat-TM data, Iranian natural resources journal, Vol. 50, No.
-FAO, 2002, www.fao.org/forestry/site/23747/en/irn
-Franklin, S.E., 2001, Remote Sensing for Sustainable Forest Management, Lewis Publishers,
-Gelman, A., J. Carlin, B and Rubin.D. B,1995. Bayesian Data Analysis. CRC Press, Boca Raton.
-Janssen, L. F., Jaarsma, M. N., and Erike T.M., 1990, Integration of topographic data with Remote Sensing for Land cover classification, phtogrammetric Engineering $ Remote Sensing, Vol. 56, No. 11: 1503-1509.
-Hopkins P. F., Maclean A. L., Lillesand T. M., 1988, Assessment of Thematic Mapper Imagery for Forestry Application under Lake States Conditions, Photogrameteric Engineering and Remote sensing, Vol. 54, No. 1, 61-68.
-Hutchinson Charles F., 1982, Techniques for combining Landsat and ancillary data for digital classification improvement, PE & RS, Vol. 48, No. 1: 123-130
-Lo, C. P., Watson, L. J., 1998, The influence of Geographic Sampling Methods on Vegetation Map Accuracy Evaluation in a Swampy Environment, Photogrametric Engineering and Remote Sensing, Vol. 64, No. 12: 1189-1200.
-Maselli, F., Cones, C., Fillips, T. D., Romani, M., 1995, Integration of ancillary data into a maximum likelihood classifier with nonparametric priors, ISPRS Journal of photogrammetry and remote sensing, 50(2): 2-11.
-Pedley M. I., Curan P. J., 1991, Pre-field classification: an example using SPOT HRV imagery, International Journal of Remote Sensing, Vol. 12, No. 11: 2181-2192
-Richards, J.A., (1986), Remote Sensing Digital Image Analysis: an introduction. Berlin: Springer-Verlag.
-Shataee, SH., Mohajer, M., 2002, Forest Classification based on Thick Trees (Case Study: Research Forest of Faculty of Natural Resources in Kheyroudkenar), Iranian Journal of Natural Resources, Volume 55, No.3.
-Shataee, SH., Kelenberger, T., Darvishsefat, A., 2004, Forest Types Classification Using ETM+ Data in the North of Iran/ Comparison of Object-Oriented With Pixel-Based Classification Techniques, XXth ISPRS congress, Turkey
-Strahler, A.H., 1980, The use of prior probabilities in maximum likelihood classification of remotely sensed data. Remote sensing of Environment, 10: 135-136.
-Walsh, S.J., 1980, Coniferous tree species mapping using LANDSAT data, Remote sensing of environment, 9:11-26.
-Williams, J.A. 1992, "Vegetation Classification Using Landsat TM and SPOT-HRV Imagery in Mountainous Terrain, Kananaskis Country, S.W. Alberta". Alberta recreation and Parks, Kananaskis Country Operations Branch.