International Journal of Architectural, Civil and Construction Sciences
ISSN: 2415-1734
Vol:12, No:9, 2018

Spectral Mixture Model Applied to Cannabis Parcel Determination

Levent Basayigit, Sinan Demir, Yusuf Ucar, Burhan Kara

***Abstract*—**Many research projects require accurate delineation of the different land cover type of the agricultural area. Especially it is critically important for the definition of specific plants like cannabis. However, the complexity of vegetation stands structure, abundant vegetation species, and the smooth transition between different seconder section stages make vegetation classification difficult when using traditional approaches such as the maximum likelihood classifier. Most of the time, classification distinguishes only between trees/annual or grain. It has been difficult to accurately determine the cannabis mixed with other plants. In this paper, a mixed distribution models approach is applied to classify pure and mix cannabis parcels using Worldview-2 imagery in the Lakes region of Turkey. Five different land use types (i.e. sunflower, maize, bare soil, and cannabis) were identified in the image. A constrained Gaussian mixture discriminant analysis (GMDA) was used to unmix the image. In the study, 255 reflectance ratios derived from spectral signatures of seven bands (Blue-Green-Yellow-Red-Rededge-NIR1-NIR2) were randomly arranged as 80% for training and 20% for test data. Gaussian mixed distribution model approach is proved to be an effective and convenient way to combine very high spatial resolution imagery for distinguishing cannabis vegetation. Based on the overall accuracies of the classification, the Gaussian mixed distribution model was found to be very successful to achieve image classification tasks. This approach is sensitive to capture the illegal cannabis planting areas in the large plain. This approach can also be used for monitoring and determination with spectral reflections in illegal cannabis planting areas.

***Keywords*—**Gaussian mixture discriminant analysis, spectral mixture model, World View-2, land parcels.

I. Introduction

ONE of the most widely used areas of remote sensing is the mapping of land cover and land use. Information about area estimates and crop yields during harvest periods are very important for government agencies responsible for security and policy. Thus, these government agencies require complete and accurate estimates of area.

When remotely sensed data have been used for monitoring the landscape, there is a suggestion that an acceptable accuracy limit for land cover maps is to be 85% [1]. But, there are many small parcels in the fragmented arable land. The traditional land use patterns of these parcels are the mixed plant growing. This is a problem for monitoring specific plant that is cultivated under the control of the governments. In the world, cultivation of cannabis is permitted by United Nations. Cultivation is under the control of governments. But, it is reported that the cannabis cultivation has dramatically increased for illegal purposes [2]. Therefore, the requirement of data for monitoring and detecting in large areas of cannabis plant is increasing [3]. However, common classification approaches is a more useful technique in the very large agricultural areas and monoculture crops. In small parcels and mixed land use patterns, the high spectral resolution images are required [4]. Thus, there is a need for high-resolution satellite images to determine the cannabis in which mixed with sunflower and maize planting.

Spectral mixture analysis has been widely used in remote sensing for sophisticated applications such as object discrimination, detection, and classification [5]. Spectral mixture analysis has been used as a technique for analyzing the mixture of components in remotely sensed images for three decades [6]. During that time, it has been used to estimate the soil properties and vegetation cover [7]-[9]. The technique is based on the assumptions that several primitive classes of interest can be selected, each of these primitive classes has a unique spectral signature (a so-called endmember or pure pixel), which can be identified and the mixing between these classes can be adequately modeled as a linear combination of the spectral signatures.

Linear mixture models are fairly simple and generally result in poor mixture analysis accuracy. Neural network models can achieve much higher accuracy but typically lack interpretability. Another model is the mixture discriminant analysis (MDA) model. It was reported that the mixtures method of MDA model with subclasses of Gaussian distributions was successfully combined the performance parameters of more complex neural network models, with the interpretation simplicity for linear mixture models. In this situation, researchers were convinced that MDA models made a useful way for solving the mixture modeling problem in remote sensing [10]. Therefore, mixed distribution model assumed an attractive alternative for addressing the land use type in this study. The mixed distribution models' approach was applied for distinguishing cannabis vegetation in the mixed cultivated area using World View-2 images.

L. Basayigit is with the Applied Science University of Isparta, Faculty of Agricultural Sciences and Technologies, Soil Science and Plant Nutrition Department, Isparta Turkey (phone: +90-246-2118589; fax: +90-246-2118696; e-mail: leventbasayigit@hotmail.com).

S. Demir is with the Applied Science University of Isparta, Faculty of Agricultural Sciences and Technologies, Soil Science and Plant Nutrition Department, Isparta Turkey.

B. Kara is with the Applied Science University of Isparta, Faculty of Agricultural Sciences and Technologies, Field Crop Department, Isparta Turkey.

Y. Ucar is with the Applied Science University of Isparta, Faculty of Agricultural Sciences and Technologies, Agricultural Structures and Irrigation Department, Isparta Turkey.

II. Materials and Methods

A. Study Area

The study was conducted in Bucak-Burdur, Turkey in 2016. The location of the study area was shown in Fig. 1. Bucak-Burdur is located in western Anatolia region of Turkey.

The coordinates of the study area are spanning from 37°26′42″ to 37°26′27″ N, and from 30°33′12″ to 30°30′07″ E.

Fig. 1 The location of the study area

B. Experiment Design

In this study, cannabis, poppy, maize, and sunflower were used as experiment plants. While the cannabis was basis plant for the purpose of the study, others were used to hidden of cannabis. The experiment was set up according to the randomized complete-block design. Experiment design is showed in Fig. 2.

The plot size was 15 m x 60m = 900 m^2 and consisted of 22 rows, and the experiment was set up in the last week of April. The experiment consisted of total 20 parcels. In the study, the area of 2.7 hectares was under the cultivation. Seeds were sown at 70 cm x 30 cm row spaces. Regular cultural practices were kept for all treatments. The irrigation was applied using a drip system.

C. Image Classification

Four Worldview-2 images were used in the vegetation period. These images were obtained at 06/01/2016, 06/20/2016, 07/02/2016, 08/02/2016. The False Color images were shown Fig. 3.

D. Converting to Top of Atmosphere Radiance

WorldView satellite quantifies the vegetation's photosynthetic response to red radiation absorption and near-infrared reflectance. Spectral patterns of Cannabis were investigated through the analysis of spectral reflectances derived from multi-temporal series in four different periods of pixel values. Twenty-eight layers were produced using pixel values of seven bands of WorldView2 images in four periods. These bands were blue (447.5 – 508.3 nm), green (511.3 – 581.1 nm), yellow (588.5 – 627.0 nm), red (629.2 – 688.5 nm), red-edge (703.8 – 743.6 nm), near-infrared 1 (772.4 – 890.2 nm) and near-infrared 2 (861.7 – 954.2 nm).

The pixel values of each band of the satellite images were converted into reflectance values according to the process of the study. For this purpose, the atmospheric correction was done and the pixel values are transformed into reflection ratio [11], [12].

Converting WorldView DN to Top of Atmosphere Radiance was described in Radiometric Use of WorldView-2 Imagery, Technical Note [12]. To account for sensor characteristics, the images were converted from DN's to LTOA spectral radiance using ArcMap 9.3 Raster Calculator [13] according to (1):

$$L_{\lambda_{\text{Pixel Band}}} = \frac{K_{\text{Band}} \cdot q_{\text{Pixel Band}}}{\Delta\lambda_{\text{Band}}} \quad (1)$$

where: $L_{\lambda Pixel,band}$ represents TOA spectral radiance image pixels (W-m^{-2}-sr^{-1}-µm^{-1}); K_{Band} is the absolute radiometric calibration factor (W-m^{-2}-sr^{-1}-count^{-1}) for a given band; $Q_{Pixel,Band}$ represents the radiometrically corrected image pixels (DN); and $\Delta_{\lambda Band}$ is the effective bandwidth (µm) for a given band. The absolute calibration (K_{Band}) and effective bandwidth ($\Delta_{\lambda Band}$) parameters for each band are obtained from the metadata supplied with the imagery.

1. Converting to Top of Atmosphere Reflectance

Converting to Top of Atmosphere Reflectance were done by using ArcMap 9.3 Raster Calculator [13] with the following (2):

$$\rho_{\lambda_{\text{Pixel Band}}} = \frac{L_{\lambda_{\text{Pixel Band}}} \cdot d_{ES}^{2} . \pi}{\text{Esun}\ \lambda_{\text{Band}} \cdot \cos(\theta_s)} \quad (2)$$

where $P_{\lambda Pixel,Band}$ is reflectance rate for each pixel (%); Π is mathematical constant (3.14159265358); L_λ represents TOA radiance derived from conversion of WorldView DNs; d_{ES}^2 is Earth-Sun distance in astronomical units, interpolated values; Esun is mean solar exoatmospheric irradiance (s) (W/m^2/µm), interpolated values which were prepared for each band of WorldView 2 by Digital Globe (Table I); and *cos (Θs)* is solar zenith angle from the image acquisition's metadata.

E. Gaussian Mixture Discriminant Analysis

MDA method was used to model training data set and to classify test data effectively when the structure of the training groups was heterogeneous. Discriminant analysis can be used as a model-based classification method for per-field classification of remotely sensed multispectral image data of an agricultural area. In this type of classification method, a cluster of observations taken from unknown groups or test groups is assigned to one of the known pre-defined groups or training groups, indicating that it is a supervised classification method [14]-[16].

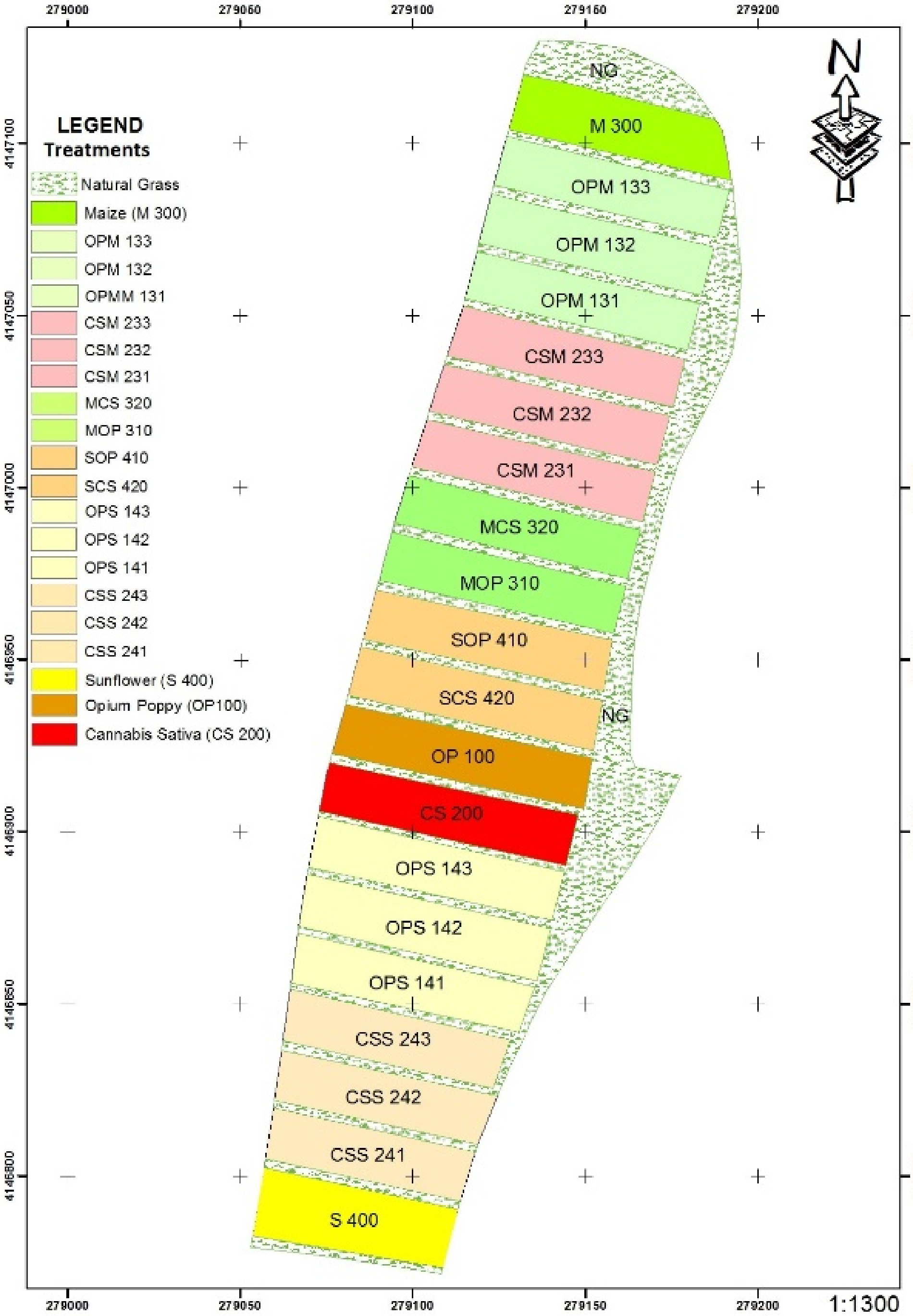

Fig. 2 Experiment design

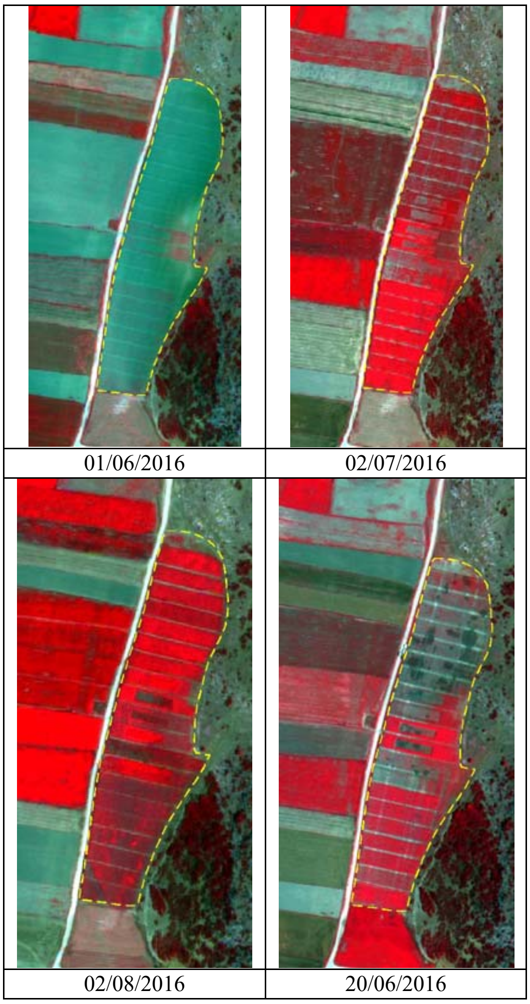

Fig. 3 False Colour images with bands of 7 (near-IR), 5 (red) and 3 (green)

TABLE I
MEAN SOLAR EXOATMOSPHERIC IRRADIANCE FOR WORLDVIEW 2

Spectral Band	Spectral Irradiance [W-m^{-2}- πm^{-1}]
Panchromatic	1580.8140
Coastal	1758.2229
Blue	1974.2416
Green	1856.4104
Yellow	1738.4791
Red	1559.4555
Red Edge	1342.0695
NIR-1	1069.7302
NIR-2	861.2866

GMDA based on Multi-Variable Normal Distribution was applied using the Matlab 2016 statistical program [17]. In this respect, the analysis of spectral data was the useful method because of discrimination capability of each spectrum.

The Gaussian mixture model is a statistical method based on the Gaussian distribution which was described as weighted components. The assumption of the Gaussian mixture model is made for p-dimensional data with n observations and G groups. Thus, the probability density function f (x; θ) for the multispectral image data has the following form:

$$f(xi;\theta) = \sum_{j=1}^{G} \pi_j f_j\,(xj;\Psi j),\ i = 1,2,\ldots,n, \qquad (3)$$

where πj is the mixture proportion of the jth component for j = 1, 2, . . . ,G such that 0 < πj < 1 and _Gj =1 πj = 1. *fj(xi;_j)* is the component probability density function which has a multivariate normal distribution with mean $\boldsymbol{\mu} j$ and covariance matrix_j for the jth component.

The setting of GMDA models for five different land use types in study area includes following operation steps;

I. Production of Random Numbers corresponding to reflection ratio (total 255),
II. Sorting of reflection ratios according to random numbers,
III. Selecting the first 80% of the randomly ranked reflection ratios as training data set (total 205),
IV. Dimensioning training data to subgroups,
V. Modeling of detailed data with GMDA.

III. Results and Discussion

Remotely sensed the multispectral image of the experiment area was classified with the proposed supervised classification method based on GMDA. All classes in the multispectral image were determined to apply an algorithm and using the parcels map of this area. Four multitemporal images were separated all bands, and reflection ratio of twenty-eight bands was prepared for each land use classes. So that the model for separating cannabis parcels from other land use patterns such as poppy, corn, sunflower, and grass etc. in the area was developed (Fig. 4).

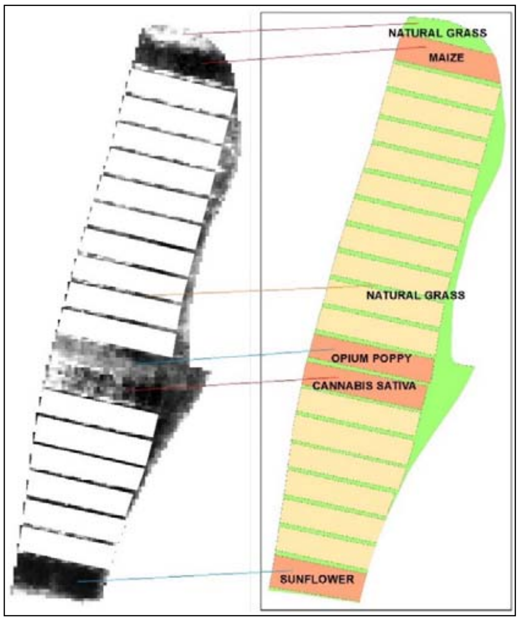

Fig. 4 The location of plants in the experiment design

The reflectance values of each class are produced, and the reflection graphs of mean values are presented in Fig 5. The different number of components in the mixture of distribution model was determined with the Bayes classification of randomized training data.

Fraley and Raftery [18] proposed to select the parameterization of Rcj and the number Mc for each class separately by the Bayesian information criterion (BIC) [19], [20] in order to reduce the computation complexity [21].

In accordance with this information, the reflectance values obtained from all bands of July image were classified with Bayes rule. It was found that 205 reflection ratios in cannabis parcels were divided into 2 component.

The produced graphs according to Bayes rule showed that classification of the training data into 2 subgroups is sufficient to explain the effect levels of the data. Mean (μ) and covariance matrices (Σ) of the two subgroups of dimensioned training data and the probability of these values belong to the classes (Pi, k) and the component mixture ratios (π) were produced and showed in Table II. The generated values were modeled using MDA based on Multivariate Normal Distribution. Then encoding (programming) was performed by Matlab software so that test data could be parametrically integrated into the model.

It was a very difficult operation that the determination of the probability of belongs to class k (Pi, k) of xi label vector used in the model. It was time-consuming the calculation of the parameters such as averages of subgroups (μk), component covariance matrices (Σk) and component mixing ratios (π) manually. These processes have also high error margin.

In order to overcome these problems, an algorithm has been developed in the Matlab software with high detail. However, this algorithm was produced for the project supported a Scientific and Technical Research Council and the project has not yet been finalized. So, it is not presented in detail here for this reason.

As a result of the mixed distribution analysis, the equation of the approach that produced for the five different land use classes and function digit were presented below:

$$f(x_i;\ \Psi) = \sum_{k=1}^{g} \pi_k \Phi_k(x_i;\ \mu_k, \Sigma_k) \tag{4}$$

where f(xi; Ψ): Component probability density function, Φk(xi; μk, Σk) : Multivariable normal distribution probability density function, k =1,2……, g and i =1,2,……,n such that Φk(xi; μk, Σk) function mean vector μi, covariance matrix Σi and,

$$\Phi_k\big(\ x_j;\ \mu_k, \Sigma_k\big) = (2\pi)^{-\frac{p}{2}} |\Sigma_k|^{-\frac{1}{2}}\ e^{\left\{-\frac{1}{2}(x_j-\mu_k)^{I} \Sigma_k^{-1}\ (x_j-\ \mu_k)\right\}} \tag{5}$$

is a multivariate normal distribution function. Where, unknown parameters of the mixture model were denoted by Ψ = (π1, ..., πg-1, ξ) T.

In equation ξ, the component mean vectors which parameter of component probability densities function in the mixture distribution model were denoted by μ = (μ1, μ2,….. μg) and covariance matrix were denoted by Σ = (Σ1, Σ2,….., Σg).

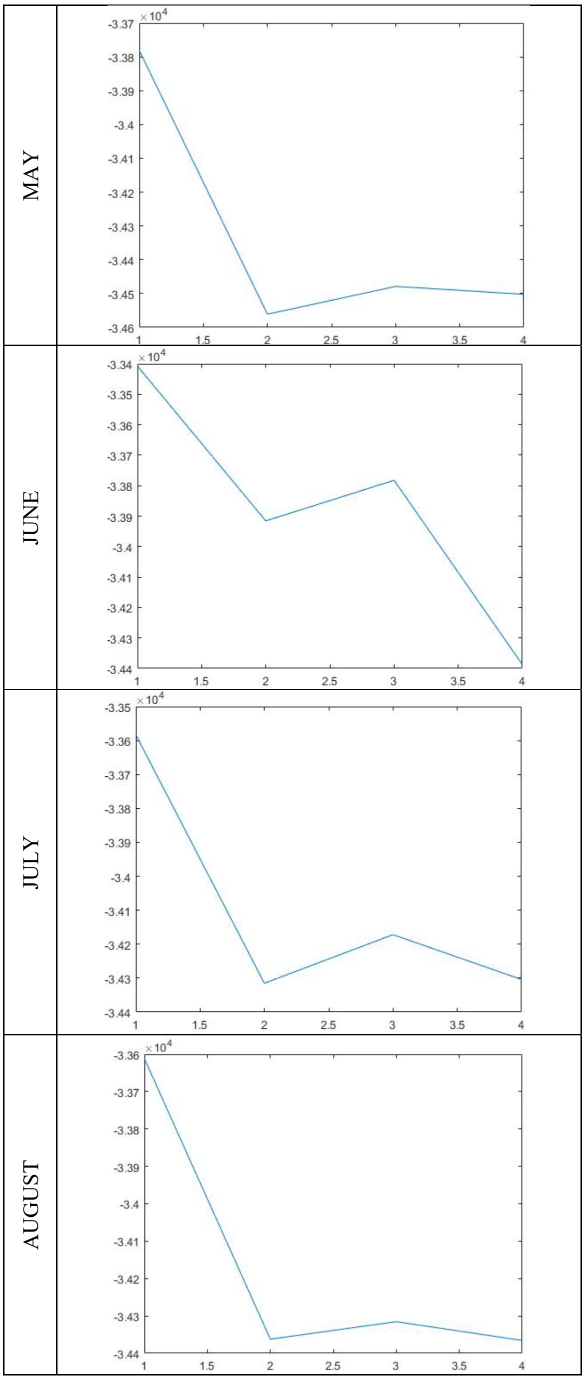

Fig. 5 Subgroup graphs of Bayes rule

TABLE II
THE SPECTRAL VALUE OF COMPONENT

		Blue	Green	Yellow	Red	Red Edge	NIR-1	NIR-2
MAY	μ_1	0,035502	0,039333	0,034383	0,054606	0,041671	0,070064	0,058480
	μ_2	0,147990	0,163957	0,143327	0,227625	0,173703	0,292059	0,243772
JUNE	μ	0,037885	0,041811	0,043817	0,048472	0,052417	0,072007	0,060682
	μ_2	0,202227	0,223183	0,233886	0,258737	0,279796	0,384364	0,323909
JULY	μ_1	0,172149	0,176282	0,180758	0,208073	0,234406	0,438818	0,354860
	μ_2	0,046418	0,047533	0,048740	0,056105	0,063205	0,118323	0,095685
AUGUST	μ_1	0,135002	0,134985	0,138904	0,148144	0,192192	0,368953	0,294764
	μ_2	0,046022	0,046016	0,047352	0,050502	0,065517	0,125774	0,100484

The parameters that should be used in the model and suitable for this study xi = reflectance values of all bands for plant pattern, μk = Mean value of subgroups obtained from Bayes rule, Σk = The covariance matrix of reflection ratios for 5 classes and 7 bands in 4 different periods, p = Belonging probability that any reflection ratio one of the 5 classes (k class) relative to the covariance matrix (xi, label vector), πk = The class k (any of one in the 5 classes) was described as the mixture distribution ratio for xi label (covariance matrix of reflection ratio).

This study indicates that the Gaussian mixed distribution models' approach is a promising method for distinguishing cannabis vegetation in the mixed cultivated area using World View-2 images. The use of this model, all test data assigned to real classes with 98 % accuracy. Thus, a new approach is sensitive to capture the studies of monitoring and determination with spectral reflections in illegal cannabis planting areas on a large scale.

IV. CONCLUSION

The results of the model which using 205 randomly selected training data in the Normal Distribution based MDA was found successful for five class. As a matter of fact, it was determined that the label vector X (reflectance ratios of different classes) could be correctly assigned to one of the possible G groups (5 classes). In particular, selected test data (randomly 51 reflectance ratios) included in the function in 5 different models and assigned to all classes were assigned to the belonging classes. This result was the very important achievement for the study's goal. The most striking result of the study was those test parameters, not known the classes, were randomly arranged and 1785 reflectance values (51 reflectance values and 5 classes and 7 bands = 1785 values) were accurately assigned to their classes in the produced models.

Finally, when the spectral data obtained from WV-2 images (digital numbers converted to reflectance values) were modeled by GMDA, it was given an opportunity to determine the cannabis parcels with high accuracy in the large plain.

ACKNOWLEDGMENT

We would like to acknowledge the financial support of The Scientific and Technical Research Council of Turkey (TÜBİTAK). Agriculture. Forestry & Veterinary Research Grant Committee (TOVAG) Project No 215 O 618.

REFERENCES

[1] J. R. Anderson, E. E. Hardy, J. T. Roach, and R. E.Witmer, 1976. A land use and land cover classification system for use with remotely sensed data, *U.S. Geol. Survey*, Professional Paper 964, Reston, VA, 1976.
[2] EMCDDA, 2011. Europan Monitoring Center for Drugs and Drug Addiction. *National Report*.
[3] United National Office on Drugs and Crime. 2009. World Drugs Report 2009. Vi- enna. url: https://www.unodc.org/ documents/wdr/WDR_2009/WDR2009_eng_ web.pdf.
[4] L. Basayigit S. Demir B. Kara and Y. Ucar, 2017. Monitoring Cannabis Cultivation with High-Resolution Images. *ICGRS 2017: 19th International Conference on Geoscience and Remote Sensing.* Rome Italy Sep 18-19, 2017, 19 (9) Part XI.
[5] C. Heinz Daniel, 2011. Fully Constrained Least Squares Linear Spectral Mixture Analysis Method for Material Quantification in Hyperspectral Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 39, No. 3.
[6] H. M. Horwitz, R. F. Nalepka, P. D. Hyde, and J. P. Morganstern, 1971. "Estimating the proportion of objects within a single resolution element of a multispectral scanner," *Tech. Rep. NASA* Contract NAS-9-9784, Univ. Michigan, Ann Arbor, MI.
[7] F. J. Garcia-Haro, M. A. Gilabert, and J. Melia, 1996. "Linear spectral mixture modelling to estimate vegetation amount from optical spectral data," *Int. J. Remote Sensing*, vol. 17, no. 17, pp. 3373–3400, 1996.
[8] J. J. Settle and N. A. Drake, 1993. Linear mixing and the estimation of ground cover proportions, *Int. J. Remote Sensing*, vol. 14, no. 6, pp. 1159–1177, 1993.
[9] E. J. van Koowijk, H. van der Voet, and J. J. M. Berdowski, 1995. Estimation of ground cover composition per pixel after matching image and ground data with subpixel accuracy, *Int. J. Remote Sensing*, vol. 16, no. 1, pp. 97–111, 1995.
[10] J. Junchang Eric D Kolaczykb. Sucharita G. 2003. Gaussian mixture discriminant analysis and sub-pixel land cover characterization in remote sensing. *Remote Sensing of Environment* 84, 550–560.
[11] J. A. Richards and J. Xiuping, 1999. Remote sensing digital image analysis. Springer.
[12] Digital Globe, 2013. Digitalglobe Constellation Worldview-2. WorldView-2 spacecraft information and specifications. www.digitalglobe.com.
[13] ESRI, 2010. ArcMap 9.3 Redlands, CA: Environmental Systems Resource Institute.
[14] T. Hastie and R. Tibshirani, 1996. *Discriminant analysis by Gaussian mixtures*, J. R. Stat. Soc. 58, pp. 155–176.
[15] N. Çalış and H.Erol, 2012. A new per-field classification method using mixture discriminant analysis, *Journal of Applied Statistics* Vol. 39, No. 10, October 2012, 2129–2140.
[16] S. Deng, X. Yifei, L. Xiaoli, He. Yong, 2015. An infinite Gaussian mixture model with its application in hyperspectral unmixing, *Expert Systems with Applications* 42 (2015) 1987–1997.
[17] Reynolds D, 2015. Gaussian mixture models. *Encyclopedia of biometrics*, 827-832.
[18] Fraley and Raftery (2002) Fraley, C., & Raftery, A. E. (2002). Model based clustering, discriminant analysis and density estimation. Journal of the *American Statistical Association*, 97, 611 – 631.
[19] Schwartz, G., 1979. Estimating the dimension of a model. *Ann. Statist.* 6, 461–464.
[20] Kass, R. E., Raftery, A. E., 1995. Bayes factors. *J. Amer. Statist. Assoc*. 90, 773–795.
[21] Halbe Z, and Aladjem M, 2007. Regularized mixture discriminant

analysis. *Pattern Recognition Letters* 28 (2007) 2104–2115.