УДК 631.171:004.94:528.8

Comparison of Supervised Classification Techniques and Gaussian Mixture Distribution Method for Estimating Land Use Type on the Landsat 7 Satellite Data

Mert Dedeoglu, Levent Başayiğit, Murat Erişoğlu

Abstract

The applications of remote sensing on the land cover classification are very important for lands planning. . For this purpose multispectral satellite data uses in agricultural areas due to practically in large-scale. Many of the studies have been carried out with supervised classification based on pixel function. However, it is not possible to determine what kind of uses are in the land where the natural pattern distribution is unknown. For this reason, some statistical approaches and conditional algorithms are trying in the classification. The most common of these is Gaussian mixture distribution analysis (GMD). Mixture distribution analysis is multivariate statistical methods that deal with the separation of objects or different clusters of observations and the re-organization of pre-defined groups. This study was conducted for classification of sugar beet and wheat with grassland by using SC and GMD on the Landsat 7 ETM+ data in 1850 ha areas of Konya region and the spatial ranges were compared with the location realities. The result of calculations has shown the SC estimated sugar beet, wheat and grassland with accuracy coefficients at 0.88, 0.90, 0.83, respectively. On the other hand, GMD predicted with accuracy coefficients at 0.94 for sugar beet, 0.96 for wheat and 0.92 for grassland. As a result, GMD can be used with high accuracy in the detection of different land cover types in unknown lands. However, similar studies should be performed on different types of land cover for creation of spectral libraries, thus GMD could determine all of the land cover types.

Keywords: Mixture distribution, remote sensing, sugar beet, supervised classification, wheat

Сравнение методов адаптивного классифицирования и Гауссового смешанного распределения для оценки типа землепользования по спутниковым данным Landsat 7

Мерт Дадаоглу, Леван Башайигит, Мурат Еришоглу

Абстракт. Применение дистанционного зондирования при адаптивном классифицировании (разбиении на классы) земельного покрова очень важно при планировании землепользования. Для данной цели многоспектральные спутниковые данные используются в сельскохозяйственных районах крупномасштабно. Многие из исследований были проведены методом попиксельного адаптивного классифицирования (SC). Однако невозможно определить виды землепользования на земельных участках в случае неизвестности распределения для естественных паттернов (образцов фрагментов изображений). По этой причине для адаптивного классифицирования разрабатываются некоторые статистические подходы и условные алгоритмы. Наиболее распространенными из них являются анализ Гауссового смешанного распределения (GMD). Анализ смешанного распределения – это многомерные статистические методы, которые касаются разделения объектов или различных кластеров наблюдений и реорганизации заранее определенных групп. Это исследование было проведено для классификации территорий землепользования, занятых сахарной свеклой и пшеницей на луговых пастбищных угодьях методами SC и GMD по данным Landsat 7 ETM + на площади 1850 га в районе Конья, а пространственные диапазоны были сопоставлены с настоящим местоположением участков. В результате расчетов была методом SC получена оценка земельных угодий, занятых сахарной свеклой, пшеницей и пастбищами с коэффициентами точности 0,88, 0,90, 0,83, соответственно. С другой стороны, GMD предсказал с коэффициентом точности 0,94 для сахарной свеклы, 0,96 для пшеницы и 0,92 для лугопастбищных угодий. В результате GMD можно использовать с высокой точностью при обнаружении различных типов растительного покрова в неизвестных землях. Однако аналогичные исследования должны проводиться на разных типах растительного покрова для создания спектральных библиотек, и тогда с использованием GMD можно определять все типы растительного покрова.

Ключевые слова: модель распределения Гаусса, дистанционное зондирование, сахарная свекла, классификация, пшеница

INTRODUCTION

Land use refers to the way in which physical features such as earth’s surface, vegetation, water, settlements etc. are used by humans, while generally emphasizing the functional role of land for economic activities (Dimyati et al., 1996). Land has a privileged status among natural resources in meeting increasing demands of basic human needs (Deniz and Turan, 2014); therefore, the most suitable designs should be planned on the choice and implementation of land use type (Rawat and Kumar, 2015). Indeed, among factors constituting these designs, determining and/or following current land use type is of capital importance beneath the definition of sustainable agricultural cultivation. The most important tool used in land management and planning in modern societies and supported with scientific research is satellite images (McRobert and Tomppo, 2007; Weih and Riggan, 2010). Remote sensing techniques are widely used in research carried out to this end (Lo and Choi, 2004). Especially in large-scale studies, thematic maps that characterize land use through satellite images can be produced fast, cheaply and properly, and can be analyzed concordantly with Geographic

Information Systems (Bisht and Kothyari, 2001; Li et al., 2018). The status of remote sensing techniques in our day as one of the most important applications in determining and following land use type has enabled image processing techniques to be developed in this field. Image processing is a method in which different algorithms are used to obtain useful information regarding what is demanded through visual data that has been relayed to digital media in numerical arrays (Chuvieco, 2009). Accuracies, advantages and shortcomings of image processing algorithms and statistical approaches used in detecting and classifying land use types are put forward (Jensen, 1983; Muller, 1988; Woodcock et al., 2001; Homer et al., 2004; Olthof et al., 2005; Sexton et al., 2013). In this study, success levels of Supervised Classification technique, a clustering algorithm, and Gaussian mixture distribution method, a decomposition analysis, in classifying land cover in satellite images of Landsat 7 ETM+ on a test land of about 1800 ha whose land use type is grassland, wheat and sugar beet have been compared and put forward.

OBJECTS and METHODS

Study area

The study was carried out on a test land of about 1800 ha between the parallels of 37° 39′ 26″ - 37° 36′ 51″ N and the meridians of 32° 45′ 37″ - 32° 47′ 51″ E, hosting three land use types, namely grassland, wheat and sugar beet, in Konya Province, Çumra District in Turkey (Figure 1).

In the study, Landsat 7 Enhanced Thematic Mapper Plus (ETM+) satellite images from June 2010 were used as data in determining land use types with image processing techniques. Landsat 7 satellite images have an 8 bit radiometric and 30 m spectral resolution; besides being classified as Raster data that have been proven to be safely useful in land characterization studies, receiving images on 7 multispectral and 1 panchromatic bands (Başayiğit et al., 2006).

Bands and wavelength detection intervals of Landsat 7 ETM+ satellite images have been shown in Table 1.

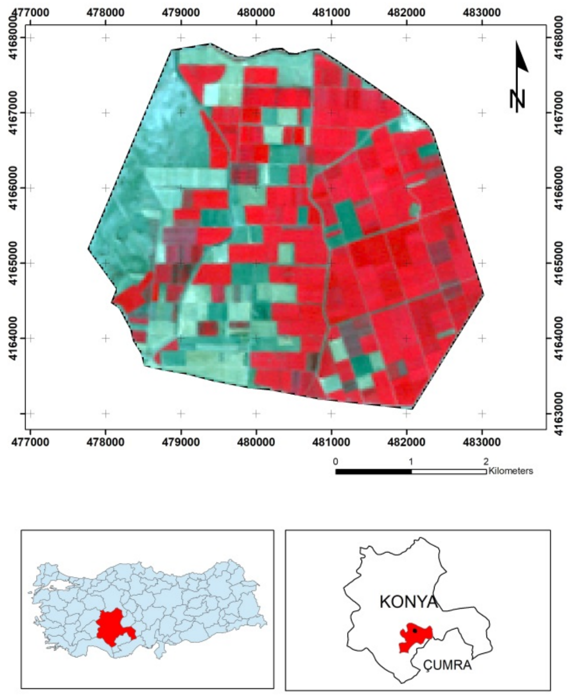

Fig 1. Landsat 4-3-2 band combinations in study area

Table 1. Band designations of Landsat 7 ETM+

Bands	Wavelength (nm)	Spatial Resolution (M)
Bant 1 - Blue	450-520	30
Bant 2 - Green	520-600	30
Bant 3 - Red	630-690	30
Bant 4 - Near Infrared (NIR)	770-900	30
Bant 5 - Shortwave Infrared (SWIR) 1	1550-1750	30
Bant 6 - Thermal	1040-1250	60
Bant 7 - Shortwave Infrared (SWIR) 2	2090-2350	30

Methods

In the study, a method was developed on the comparison of success levels of two different image classification techniques, namely supervised classification technique and Gaussian mixture distribution method (GMD), in detecting land use types. To that end, lands with a surface area of 10 x 10 m^2 were chosen as testing data for three different land use types from satellite images, and the same spectral values were used for each of two classification method. In this way, percentages of these classification techniques were obtained regarding their capacity to assign lands not included in testing data to their respective classes.

Image Processing

In the study, a method was developed on the comparison of success levels of two different image classification techniques, namely supervised classification technique and Gaussian mixture distribution method (GMD), in detecting land use types.

Of these two techniques, supervised classification technique is the most used and known one. Supervised classification is performed by introducing, on the user's part, spectral types of land cover to image processing software as pre-defined instruction samples, and by modeling other spectral types not included in this instruction algorithm with parametric or non-parametric methods (Richards and Jia, 1999). It is stated that in previous studies, supervised classification technique was reliably used in which different algorithmic approaches (nearest neighborhood, maximum likelihood algorithm etc.) could be chosen according to location facts and pixel-based thematic maps could be produced (Rogan and Chen, 2004; Başayiğit et al., 2006; Duro et al., 2012). In light of this information, Maximum Likelihood, a parametric approach, was used in *ERDAS Imagine 8.4* image processing software during supervised classification stage of the study.

Another classification technique subject to research is GMD analysis. GMD is a statistical method based on representation of one or more Gauss distributions of data as weighted components (McLachlan et al., 2000; Ok et al., 2014). GMD model is assessed as a useful method and widely used in analysis of spectral data because each spectrum of spectral data is decomposable (Manolakis et al., 2001; Deng et al., 2015). Gaussian mixed model hypothesis has been formulized as the following equation, by using the statistical software *Matlab 2016.*

$f\,(x_i; \theta)$ = the probability function of x vector belonging to θ

x_j = sample mean of the variable j

θ = vector including the unknown parameters of mixed distribution model

fj(x_i; ψj) = the multi-component mixed probability function whose average is μj and covariance matrix is Σj

π_j= mixed rate of the same component for the component *j* [j= 1, 2,….G]

Ψ = vector of all the unknown parameters in mixed probability distribution model (Reynolds, 2015)

Comparison of Classification Accuracy

On the test land, examination of vector layer parcel records showing current land uses and results from the land study carried out afterwards were produced and mapped by using the software *ArcGIS 9.3* (Figure 2). The base data showing land use types was referenced in comparing success levels of methods used in satellite images in classifying land cover accurately.

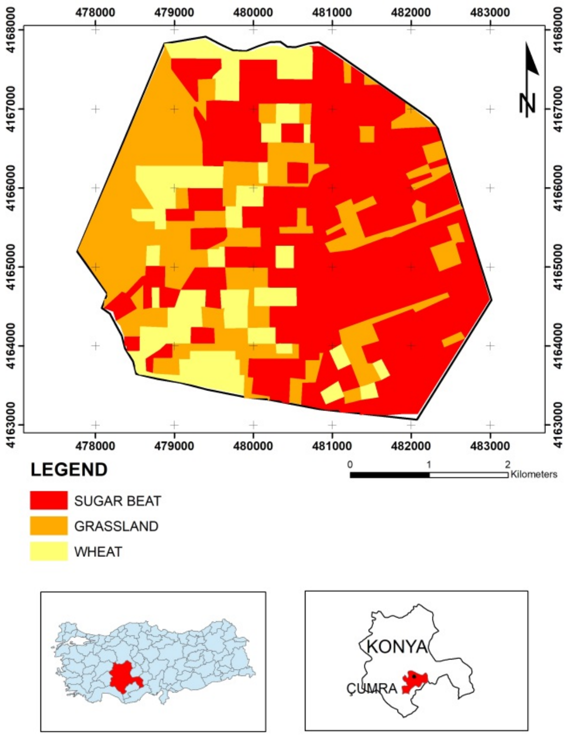

Fig 2. Land use types of study area

RESULTS AND DISCUSSION

In the study, land cover that was determined by supervised classification technique (Figure 3) and GMD model (Figure 4) was compared with the factual use. Among land use types determined in light of parcel records and land studies, grassland had a surface area of 546 ha (30%), sugar beet 1000 ha (54.5%) and wheat 290 ha (15.5%); according to maximum likelihood algorithm of supervised classification technique, the surface areas were determined as 659 ha for grassland, 882 ha for sugar beet and 323 ha for wheat. As a result of GMD model, the classifications were 581 ha for grassland, 1084 ha for sugar beet and 278 ha for wheat (Table 2).

In light of comparing land use type magnitudes determined as a result of methods used with the factual areas, GMD model was able to classify the areas of sugar beet, grassland and wheat with an accuracy of 92%, 94% and 96% respectively. The accuracy rates of supervised classification technique were 83%, 88% and 90% for the parcels of grassland, sugar beet and wheat respectively.

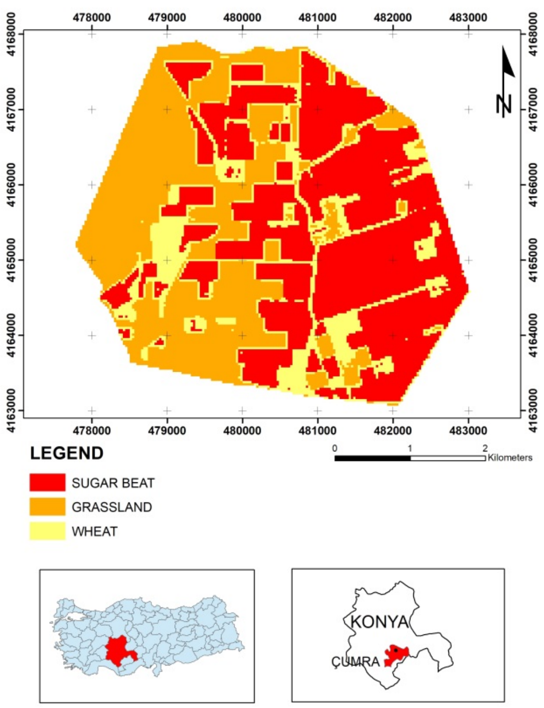

Fig 3. Map of ***Supervised*** classification

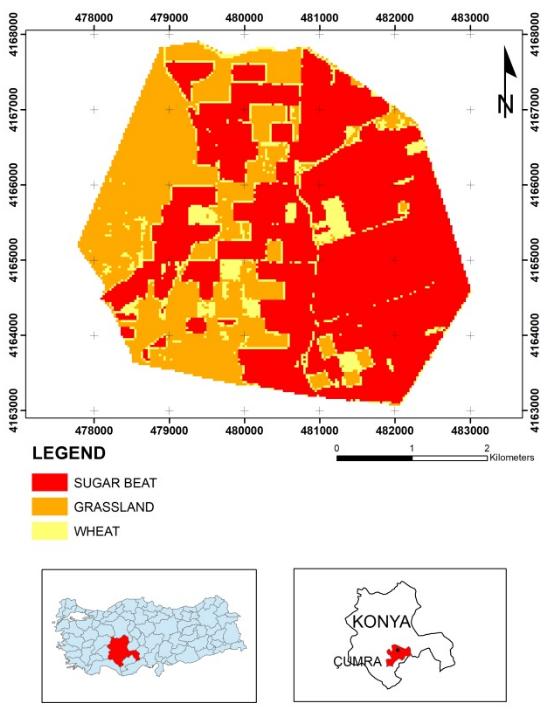

Fig 4. Map of ***GMD*** model classification

Table 2. Comparison of the areas of land use types obtained with different classification techniques

Land Use Types (ha)	Real	Supervised	R^2	GMD	R^2
Grassland	545.59	658.62	0.83	581.22	0.94
Wheat	290.39	323.10	0.90	278.63	0.96
Sugar Beet	1001.41	882.36	0.88	1083.69	0.92

The obtained findings showed that both of these classification techniques could different land use types with similar success levels. Indeed, in the studies in which supervised classification technique is used to determine land cover, accuracy rates of 70% or higher are regarded as “strong” and rendered reliable (Fitzgerald et al., 2010; Duno et al., 2012). Likewise, it has been stated that different land use types (wheat, potato, vegetable garden, citrus and bare soil) were determined with an accuracy rate of 80.22% by Gaussian mixture distribution method with raw pixel values (DNs) obtained from bands 3-4-5 Landsat 7 satellite image and that the produced model could be used with observation data in determining similar land use types (Çalış and Erol, 2012).

However, although obtained percentage results ($0.83<r^2<0.96$) were close to success levels of the methods used in determining land use types, significant differences exist in terms of magnitude. In fact, 1% area accuracy on the test land corresponds to a magnitude of approximately 185,000 m^2, which makes it necessary that methods by which land use types can be determined with the highest accuracy coefficients should be used in terms of precision agriculture implementations.

CONCLUSION

In light of the study, it has been confirmed that models produced with GMD classified all three different land use types with higher accuracy coefficients than supervised classification technique. Models produced with GMD were proven to be useful in determining different land use types with satellite images by the fact that it was highly successful in itself through creating random instruction and test data sets in areas whose sample magnitudes were known. Still, the modeling and implementation of parcel based data sets in test areas are advised in order to improve this method further, for produced models to be tested in different locations, and for production of new models according to different land use types and their use in detecting plantations with unknown land covers.

REFERANCES

1. B.S. Bisht and B.P. Kothyari, 2001. Land-Cover change analysis of garur Ganga watershed using GIS/Remote Sensing technique. *Journal of the Indian Society of Remote Sensing*. 29 (3) 137-141.
2. Başayiğit, L., Akgül, M., & Şenol, H. (2006). The relationships between the reflection values of TERRA-ASTER bands and different soil properties for mapping. I. Remote Sensing-CBS Workshop*, p*, 1-7, (in Turkish).
3. C. Homer, C. Huang, L. Yang, B. Wylie and M. Coan, 2004. Development of a 2001 National Land-cover database for the United States. *Photogrammetric Engineering and Remote Sensing.* 70 : 829–840.
4. C.E. Woodcock, S.A. Macomber, M. Pax-Lenney and W.B. Cohen, 2001. Monitoring large areas for forest change using Landsat: Generalization across space, time and Landsat sensors. *Remote Sensing of Environment*. 78: 194–203.
5. Çalış, N., & Erol, H. (2012). A new per-field classification method using mixture discriminant analysis. Journal of Applied Statistics, 39(10), 2129-2140.
6. Chuvieco, E. (2009). Fundamentals of satellite remote sensing. CRC press.
7. Deng, S., Xu, Y., Li, X., & He, Y. (2015). An infinite Gaussian mixture model with its application in hyperspectral unmixing. Expert Systems with Applications, 42(4), 1987-1997.
8. Dengiz, O., & TURAN, İ. D. (2014). Determination of Temporal Change Land Use / Land Cover Using Remote Sensing and Geographic Information System Techniques

the Central District of Samsun (1984-2011). *Turk J. Agric. Res.*, *1*(1), 78-90, (in Turkish).

9. Dimyati, M., Mizuno, K., Kitamura, T., 1996. An analysis of land use/cover change using the combination of MSS Landsat and land use map: a case study in Yogyakarta, Indonesia. Inter. J. Rem. Sen. 17, 931–944.
10. Duro D.C., Franklin S.E., Dubé M.G., 2012 - A Comparison Of Pixel-Based and Object-Based Image Analysis With Selected Machine Learning Algorithms For The Classification Of Agricultural Landscapes Using SPOT-5 HRG Imagery. Remote Sensing of Environment, 118: 259-272. doi: http://dx.doi.org/10.1016/j.rse.2011.11.020.
11. Fitzgerald, G., Rodriguez, D., & O'Leary, G. (2010). Measuring and predicting canopy nitrogen nutrition in wheat using a spectral index—The canopy chlorophyll content index (CCCI). Field Crops Research, 116(3), 318-324.
12. Franklin, S. E., He, Y., Pape, A., Guo, X., & McDermid, G. J. (2011). Landsat-comparable land cover maps using ASTER and SPOT images: a case study for large-area mapping programmes. *International journal of remote sensing*, *32*(8), 2185-2205.
13. Olthof I., C. Butson, and R. Fraser, 2005. Signature extension through space for northern landcover classification: A comparison of radiometric correction methods. *Remote Sensing of Environmen.*, 95 : 290–302.
14. Muller J.P.A.L, Key issues in image understanding in remote sensing. Philosophical Transactions of the Royal Society of London, Series A: Mathematical and Physical Sciences, 324 : 381–395, 1988.
15. Jensen J.R, 1983. Biophysical remote sensing. *Annals of the Association of American Geographers.* 73 : 111–132.
16. Li, W., MacBean, N., Ciais, P., Defourny, P., Lamarche, C., Bontemps, S., ... & Peng, S. (2018). Gross and net land cover changes in the main plant functional types derived from the annual ESA CCI land cover maps (1992–2015).
17. Lo, C.P., Choi, J., 2004. A hybrid approach to urban land use/cover mapping using Landsat 7 enhanced thematic mapper plus (ETM+) images. Inter. J. Rem. Sen. 25 (14), 2687–2700.
18. Manolakis, D., Siracusa, C., & Shaw, G., 2001. Hyperspectral subpixel target detection using the linear mixing model. IEEE Transactions on Geoscience and Remote Sensing, 39(7), 1392-1409.
19. McLachlan, G., Peel., D., 2000. Finite Mixture Models. Hoboken. NJ: John Wiley & Sons, Inc., 2000.
20. Sexton O., D.L. Urban, M.J. Donohue and C. Song, 2013. Long-term land cover dynamics by multi-temporal classification across the Landsat-5 record. *Remote Sensing of Environment.* 128 : 246–258.

21. Ok, A. O., Ok, A. O., & Schindler, K., 2014. Graph-based method based on Gaussian Mixture Modeling to classify agricultural lands. In Signal Processing and Communications Applications Conference (SIU), 2014 22nd (pp. 425-428). IEEE.
22. McRoberts R.E. and E.O. Tomppo, 2007. Remote sensing support for national forest inventories. *Remote Sensing of Environment*. 110 : 412-419.
23. Rawat, J.S., Kumar, M., 2015. Monitoring land use/cover change using remote sensing and GIS techniques: a case study of Hawalbagh block, district Almora, Uttarakhand, India. Egypt. J. Remote Sens. Space Sci. 18 (1), 77–84
24. Reynolds, D., 2015. Gaussian mixture models. Encyclopedia of biometrics, 827-832.
25. Richards J.A. and X. Jia., 1999. Remote sensing digital image analysis: An introduction (3rd ed.), Springer, Berlin.
26. Rogan, J., & Chen, D. (2004). Remote sensing technology for mapping and monitoring land-cover and land-use change. *Progress in planning*, *61*(4), 301-325.
27. Rogan, J., & Chen, D. (2004). Remote sensing technology for mapping and monitoring land-cover and land-use change. *Progress in planning*, *61*(4), 301-325.
28. Weih, R. C., & Riggan, N. D. (2010). Object-based classification vs. pixel-based classification: Comparative importance of multi-resolution imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *38*(4), C7.