The forest composition monitoring system using k-means algorithms on satellite imagery. Case study - Indepedenta Forest.

Gabriel Murariu
European Center of Excellence for the Environment
"Dunărea de Jos" University of Galaţi
Galaţi, Romania
Gabriel.Murariu@ugal.ro

Adrian Murariu
Department of Physics, Imperial College London, Kensington, London SW7 2AZ, UK
adrianmurariu@yahoo.com

Mihai Daniel Dragu
European Center of Excellence for the Environment
"Dunărea de Jos" University of Galaţi
Galaţi, Romania
dragumihaidaniel@gmail.com

Dan Munteanu
Department of Computer and Information Technology
"Dunărea de Jos" University of Galaţi
Galaţi, Romania
dan.munteanu@ugal.ro

Ionel Popa
the National Institute of Research and Development in Forestry "Marin Drăcea" – Campulung Moldovenesc, 73, Calea Bucovinei Street, RO-725100, Suceava,
popaicas@gmail.com

Lucian Puiu Georgescu
European Center of Excellence for the Environment
"Dunărea de Jos" University of Galaţi
Galaţi, Romania
Lucian.Georgescu@ugal.ro

Valentin Hahuie
National Forestry Directorate - Romsilva, County Galati ,
valiforest@yahoo.com

Catalina Iticescu
European Center of Excellence for the Environment
"Dunărea de Jos" University of Galaţi
Galaţi, Romania
catalina.iticescu@ugal.ro

***Abstract*— The increasing use of satellite faraway sensing for civilian use has proved to be the most cost effective means of mapping and monitoring environmental changes. From this point, these new tools are now essential in monitoring operations for vegetation and non-renewable resources, especially in developing countries. Data can be gotten as frequently as required to deliver information for determination of quantitative and qualitative changes in terrain. In reaching the aim of studying the forest composition and the vegetation dynamics during the time using satelitar imagery a k-means algorithm of chromatic analysis is proposed. In this way, by using comparisons between chronologically ordered records, could be described the vegetation changes and its developments in a given area. The exposed method in the present work suggests the joining two successive phases: the first step were used resolution suitable satellite images in order to succeed in building a model. The study was conducted between 1990 and 2017 by using LANDSAT data set. In the second phase, are applied chromatic investigation algorithm in order to succeed in reaching a map of vegetation composition. The case study is that of Independenta Forest from Galati County. For discussion are presented three cases of September 2011, 2016 and 2017 Preliminary results on the composition evaluations are promising and the research is ongoing.**

Keywords*— *chromatic analysis, satellite imagery, k-means cluster analysis, inhomogeneous media

I. INTRODUCTION

The Environmental monitoring and especially prevention of degradation is currently a major challenge. Soil degradation, reduction of forested areas, global warming, etc., all generate effects difficult to assess in the future. In this respect, any effort to identify changes in environmental monitoring and ground is welcome.

In the last years, the natural disasters' damages recording action has become a priority concern of governments and international organizations. This worry is justified by the fact that the recorded damages losses sometimes exceed existing budgets and national plans for economic and social development.

Indeed, the European Commission's official statistics (http://ec.europa.eu/) show explosive growth of the costs of the damages from natural hazards in the last decades. Although for hundreds of years of climate phenomena that cause catastrophic landslides or floods occur with approximately the same frequency and intensity, the damages increase in price.

For instance, over the past five decades, the values of damages by natural disasters in the world, especially those caused by earthquakes, landslides and floods, have increased exponentially [1]. For example, the long-term statistics show in the decade from 1950 to 1959 there were 20 catastrophes with losses of $ 42.2 billion. In the decade from 1990 to 1999 the number increases to 89 with losses over $652.3 billion [2]. Only in 2001 there had been 700 natural disasters around the world in the category of which had been also included two major earthquakes: one in El Salvador on 13th January 2001 with 845 dead in 10,000 landslides and $ 1.5 billion loss, and the second in India on 26th January 2001c with 14,000 deaths and $ 4.5 billion loss.

For all these reasons, it appears the necessity of adopting proactive measures against climate change which increasingly affects our life. Among these measures, the forest protection appears to be a highly effective measure [1-2]. This is neither new nor innovative.

Starting with antique times, when forests were the shelter and ending with our days, when forests provide raw materials and means mitigation of climate variability, man has always been in close contact with the forest [3, 4].

From this point of view, mankind has always been interested in the supervision and care of forests [5, 6, and 7].

By protecting forests, many helpful effects were recorded over time, starting with soil stabilization against landslides and ending of course with the incomes against desertification and soil degradation [8, 9].

In this respect, a former study proposes to define a new investigation method that uses only the visible RGB spectra at different time moments [9, 10 and 11]. The starting hypothesis was that by comparisons between chronologically ordered RGB spectra records, could describe the vegetation changes and its developments in a given area. Seeking a minimal cost price, we used images that contain [10, 11 and 12] only RGB visible spectrum. Such images are certainly less purchased from specific services [13, 14]. Images that are obtained in extended spectral ranges are not necessarily more expensive. Basing on these chronological multi spectral imagery it could be build a spatial model. The model is then verified and validated through referenced data obtained from the ground measurements.

This paper presents only the results on the composition identifying and its dynamics. Other additional obtained results obtained basing on this method will be the subject of future papers.

In this spirit, the present paper examines a critical insight into one of the contemporary used method in monitoring vegetation over large areas using satellite LANDSAT photos type. In this respect, in the next section is described our specific technique that has been developed in our Statistical Analysis and Cadastral Applications Laboratory (LASAC) from our European Centre of Excellence for the Environment – ECEE (https://erris.gov.ro/European-Centre-of-Excellenc-1). The method has been tested on a series of forestlands from Galati County and in this paper is presented the Bălăbăneşti Forest study case results.

The first innovative feature comes from the fact that we use georeferenced LANDSAT satellite images in seven spectral bands. In this way we aimed to investigate the possibility of using images with different spectral bands in a recursive way.

The second innovative characteristic lies in combining advanced analysis methods of LANDSAT spectral imagery photos with high precision monitoring data from the ground.

II. Theoretical Aspects

As a definition, given a set of observations (x_1, x_2, …, x_n), where each observation is a d-dimensional real vector, k-means clustering method aims to partition the n observations into m (≤ n) sets S = {S1, S2, …, Sm} so as to minimize the within-cluster variance [2].
In our study case, working with satelitar imagery, the k-means problem is defined in a different way.
Using satelitar imagery for LANDSAT or SENTINEL missions, there are actually a series of photos in different spectra for the same study area [1, 2].
Each of these images is in fact, a colour tone table with 256 levels of shade. In our approach, for each physical point A of physical coordinates (z, y, z) we will find for each spectral image obtained a numerical value corresponding to the respective tone. Thus, if for the LANDSAT system the images can be in 11 spectral bands, the SENTINEL system is offered in 7 spectral bands.
In this mode, we can build in representative multi-dimensional space, and on each axis we associate as the coordinate, the numerical value found in the satellite photo. (Fig. 1). Note that the domain obtained is necessarily bounded.

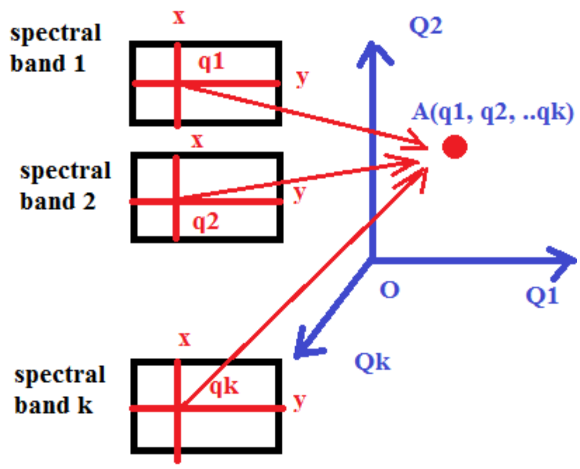

Fig. 1. The process of constructing the representative space of spectral images

From the mathematical topology, some results of reference can be mentioned [15]:
Lemma 1 [15]
In any multi-dimensional finite space, any bounded allocation is a compact one.

Fig. 2. Picture of the Bălăbăneşti forest - Galati County

Definition [15]
An allocation in a topological space is complete if any Cauchy sequence of numbers from this allocation is convergent and has the limit included in the set.
Definition [15]
An allocation in a topological space is pre-compact if it can be covered with a finite number of open allocations.
Lemma 2 [15]
In a finite multi-dimensional space, any compact set is pre-compact and complete.For this reason, considering a certain geographic area observed and photographed in the LANDSAT or SENTINEL missions, by building representative spots in the representative colour space, we will obtain a bounded allocation. For example, for the Bălăbăneşti Forest (Fig. 2) study case [16], for the RGB imagery analysis, the representative space cloud is presented in figure 3.

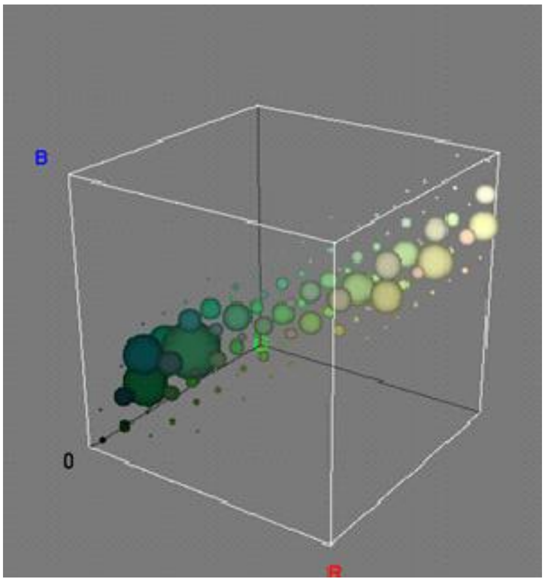

Fig. 3. Graphic representation in the RGB space of the satellite image of the Bălăbăneşti forest - Galati County

These topology theorems allow us to construct a process in which representative points are grouped into disjoint clusters that cover the entire set.

Below we briefly describe the algorithm that allows recursively grouping of representative points to some "condensation" centres.
The algorithm considered was applied recursively starting with the first three spectral images. After investigating and obtaining the clusters following the process described below, a set of conventional colours is chosen and the map is coloured using these conventional hues.
Preliminary processing:
a) Rows are observations (individuals) and columns are variables
b) Any missing value in the data must be removed or estimated.
c) The data must be standardized (i.e., scaled) to make variables comparable. Recall that, standardization consists of transforming the variables such that they have mean zero and standard deviation one
d) The classification of observations into groups requires some methods for computing the distance. The choice of distance measures is a critical step in clustering. It defines how the similarity of two elements (x, y) is calculated and it will influence the shape of the clusters. The choice of distance measures is very important, as it has a strong influence on clustering results.
Step 0: - Choose a function of metric kind or the representative space variables - in our case - the Euclidian metric (1) the Euclidian or more general (2):

$$d_{euclid}(P_1, P_2) = sqrt\left(\sum_{k=1}^{N}\left(x_{p1} - x_{p2}\right)^2\right) \qquad (1)$$

In the same step should be set the desired number of clusters. The minimum number is of course equal to unity and the maximum should be set to double to number of species.
Step 1: construct the set of representative points (in the figurative space of spectral bands - RGB, or NIR-RGB or IR-NIR-RGB etc.).
Step 2: The symmetric matrix of the relative distances between the representative points in the set of unpowered points
Step 3: Sort the distances in ascending order and identify pairs of points with minimal distances
Step 4: Point pairs with minimal distances are replaced by a single point - the centre of gravity and then, in the set of representative points, the related points are removed with the calculated centre of gravity and then step 1.
For each calculated centre of gravity, a list of points for these centres is built.

When the number of clusters - centres of gravity - reaches the minimum set value, the "condensation" or clustering process stops. The algorithm includes in a recurrent mode an increasing number of spectral images.

$$d\ (P_1, P_2) = sqrt\left(\sum_{k=1}^{N}\left(x_{p1} - x_{p2}\right)^{\alpha}\right) \qquad (2)$$

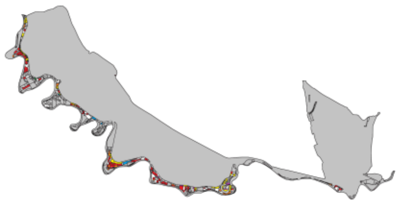

Fig. 4. Image of the Independenta Hanu Conachi Forest structure- Galati County

The LANDSAT bands allocation was the following (https://landsat.usgs.gov/):

TABLE I. TABLE FOR THE BANDS ALLOCATION FOR LANDSAT 1-5 MISSIONS

Landsat 1-3	Landsat 4-5	Wavelength (micrometers)	Resolution (meters)
Band 4 - Green	Band 1 - Green	0.5-0.6	60*
Band 5 - Red	Band 2 - Red	0.6-0.7	60*
Band 6 - Near Infrared (NIR)	Band 3 - Near Infrared (NIR)	0.7-0.8	60*
Band 7 - Near Infrared (NIR)	Band 4 - Near Infrared (NIR)	0.8-1.1	60*

TABLE II. TABLE FOR THE BANDS ALLOCATION FOR LANDSAT 6 MISSIONS

Bands	Wavelength (micrometers)	Resolution (meters)
Band 1 - Blue	0.45-0.52	30
Band 2 - Green	0.52-0.60	30
Band 3 - Red	0.63-0.69	30
Band 4 - Near Infrared (NIR)	0.76-0.90	30
Band 5 - Shortwave Infrared (SWIR) 1	1.55-1.75	30
Band 6 - Thermal	10.40-12.50	120* (30)
Band 7 - Shortwave Infrared (SWIR) 2	2.08-2.35	30

I. EXPERIMENTAL RIG

In this paper, I only used images taken from the LANDSAT portal. The case studies considered have covered several forest areas in the Galati County. In this paper we will present the case study for the Independenta Hanu Conachi Forest (Fig. 4). Also, as long as the data taken from the LANDSAT service is between 1990 and 2017, in this paper we will make a presentation covering the years 2011-2016. In order to perform a similar analysis, we considered the LANDSAT 7 missions and the analysis comprised all 7 spectral images.

TABLE III. THE BANDS ALLOCATION FOR LANDSAT 7 MISSIONS

Bands	Wavelength (micrometers)	Resolution (meters)
Band 1 - Blue	0.45-0.52	30
Band 2 - Green	0.52-0.60	30
Band 3 - Red	0.63-0.69	30
Band 4 - Near Infrared (NIR)	0.77-0.90	30
Band 5 - Shortwave Infrared (SWIR) 1	1.55-1.75	30
Band 6 - Thermal	10.40-12.50	60 * (30)
Band 7 - Shortwave Infrared (SWIR) 2	2.09-2.35	30
Band 8 - Panchromatic	.52-.90	15

TABLE IV THE BANDS ALLOC. FOR LANDSAT 8 MISSIONS

Bands	Wavelength (micrometers)	Resolution (meters)
Band 1 - Ultra Blue (coastal/aerosol)	0.435 - 0.451	30
Band 2 - Blue	0.452 - 0.512	30
Band 3 - Green	0.533 - 0.590	30
Band 4 - Red	0.636 - 0.673	30
Band 5 - Near Infrared (NIR)	0.851 - 0.879	30
Band 6 - Shortwave Infrared (SWIR) 1	1.566 - 1.651	30
Band 7 - Shortwave Infrared (SWIR) 2	2.107 - 2.294	30
Band 8 - Panchromatic	0.503 - 0.676	15
Band 9 - Cirrus	1.363 - 1.384	30
Band 10 - Thermal Infrared (TIRS) 1	10.60 - 11.19	100 * (30)
Band 11 - Thermal Infrared (TIRS) 2	11.50 - 12.51	100 * (30

II. EXPERIMENTAL RESULTS

The algorithm considered was applied recursively starting with the first three spectral images. After investigating and obtaining the clusters following the process described below, a set of conventional colours is chosen and the map is coloured using these conventional colours.

Next we present the results obtained in chronological order in the case of the forests photographed in 2011, 2016 and 2017 respectively. For a more detailed comparison, we will present the case in which 3 spectral bands, 4, 5 and 6 spectral bands were combined. In the last section we will present the results in comparison with the forestry situation taken from the ground from the national services (Fig. 5).

Fig. 5. Picture of the Independenta Hanu Conachi Forest - Galati County – composition structure

In figure 5 is the structure on the plots of the Independenta Hanu Conachi Forest. This is a benchmark against which we can evaluate the accuracy of the results obtained on the basis of the described algorithm

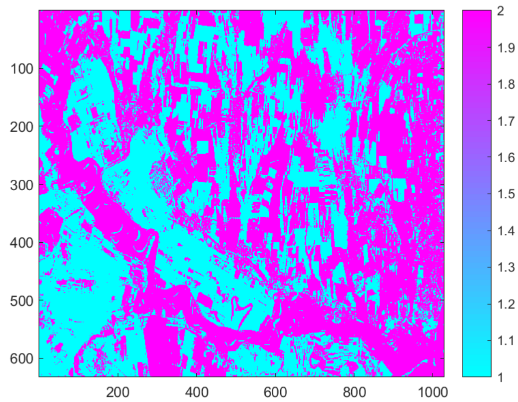

Fig. 6. Picture of the Independenta Hanu Conachi Forest - Galati County – year 2011 – composition structure - colour map in conventional colours after compiling 3 spectral bands and considering a number of 2 clusters

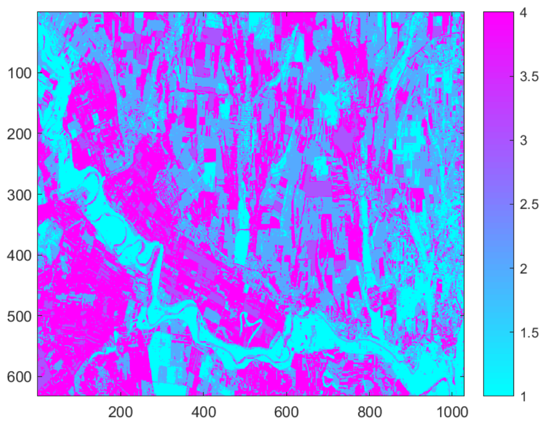

Fig. 7. Picture of the Independenta Hanu Conachi Forest - Galati County – year 2011 – composition structure - colour map in conventional colours after compiling 3 spectral bands and considering a number of 5 clusters

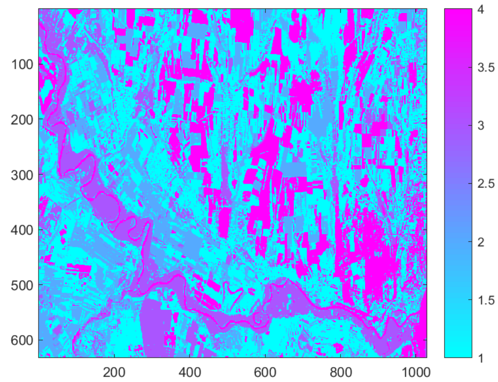

Fig. 8. Picture of the Independenta Hanu Conachi Forest - Galati County – year 2011 – composition structure - colour map in conventional colours after compiling 4 spectral bands and considering a number of 4 clusters

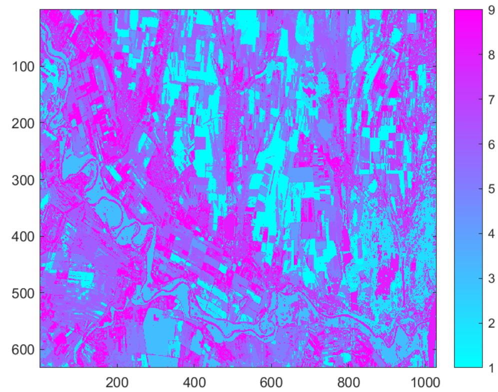

Fig. 9. Picture of the Independenta Hanu Conachi Forest - Galati County – year 2011 – composition structure - colour map in conventional colours after compiling 5 spectral bands and considering a number of 9 clusters

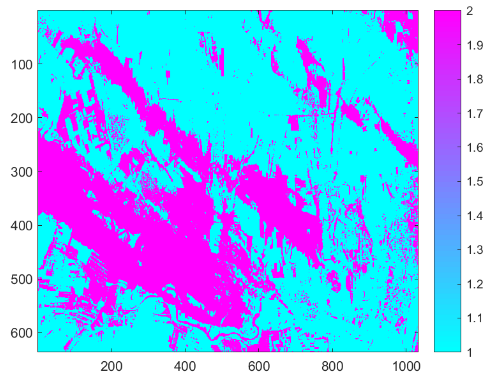

Fig. 10. Picture of the Independenta Hanu Conachi Forest - Galati County – year 2016 – composition structure - colour map in conventional colours after compiling 3 spectral bands and considering a number of 2 clusters

Figure 7 shows the result of the algorithm running using 3 bands and considering 3 clusters. The coincidence value of the coincidence map with Figure 4 is relatively small - approx. 5%.

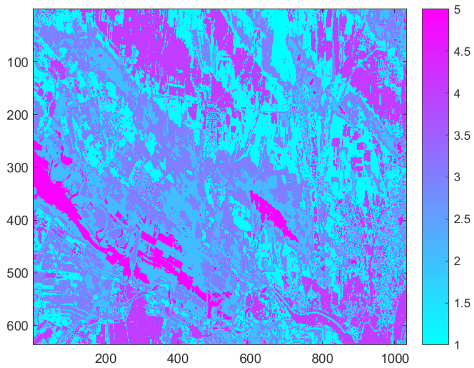

Fig. 11. Picture of the Independenta Hanu Conachi Forest - Galati County – year 2013 – composition structure - colour map in conventional colours after compiling 3 spectral bands and considering a number of 5 clusters

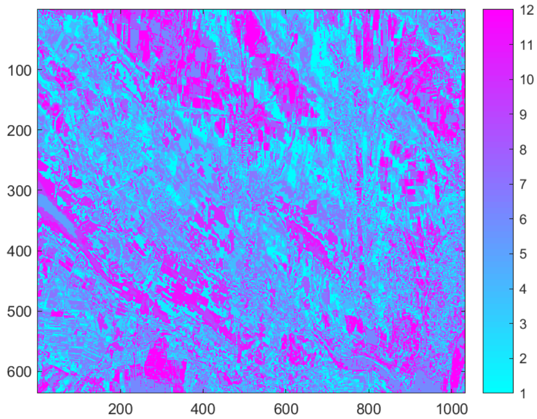

Fig. 12. Picture of the Independenta Hanu Conachi Forest - Galati County – year 2013 – composition structure - colour map in conventional colours after compiling 3 spectral bands and considering a number of 12 clusters

TABLE V THE EUCLIDIAN OBTAINED DISTANCES

No of band	No of clusters	Sum of squares – total distance
3	5	296.643
4	5	183.584
5	5	118.031
6	5	84.4257
7	5	67.0822

TABLE VI THE TOTAL EUCLIDIAN DISTANCES

No of band	No of clusters	Sum of squares – total distance
3	5	312.643
4	5	153.581
5	5	137.042
6	5	94.4257
7	5	56.890

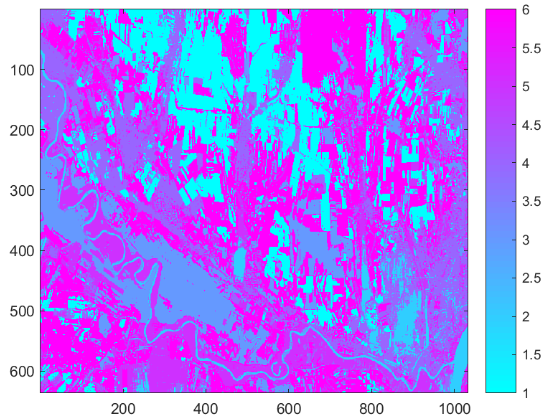

Fig. 13. Picture of the Independenta Hanu Conachi Forest - Galati County – year 2013– composition structure - colour map in conventional colours after compiling 6 spectral bands and considering a number of 12 clusters

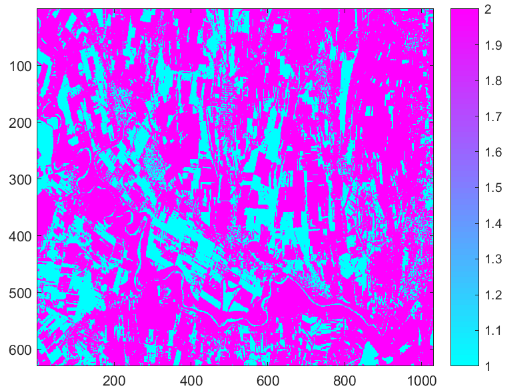

Fig. 14. Picture of the Independenta Hanu Conachi Forest - Galati County – year 2017 – composition structure - colour map in conventional colours after compiling 6 spectral bands and considering a number of 2 clusters

Figures 8 and 9 show the results of the algorithm run using 4 bands and 5 spectral bands respectively. The cluster numbers are 4 or 9. The coincidence value of the coincidence map with Figure 4 has increased but is still relatively low - approx. 11%.

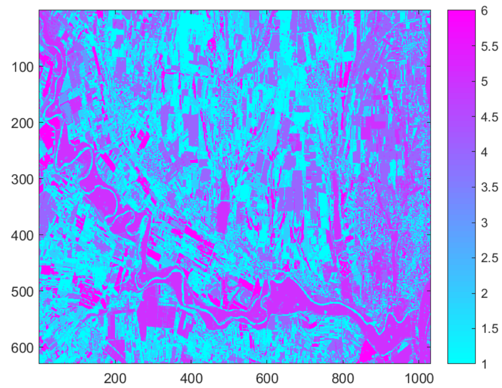

Fig. 15. Picture of the Independenta Hanu Conachi Forest - Galati County – year 2017 – composition structure - colour map in conventional colours after compiling 5 spectral bands and considering a number of 6 clusters

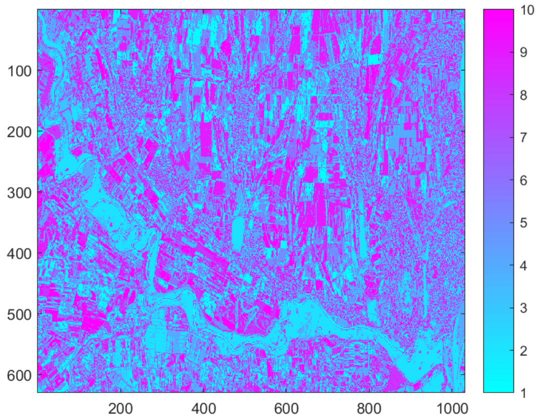

Fig. 16. Picture of the Independenta Hanu Conachi Forest - Galati County – year 2017 – composition structure - colour map in conventional colours after compiling 3 spectral bands and considering a number of 10 clusters

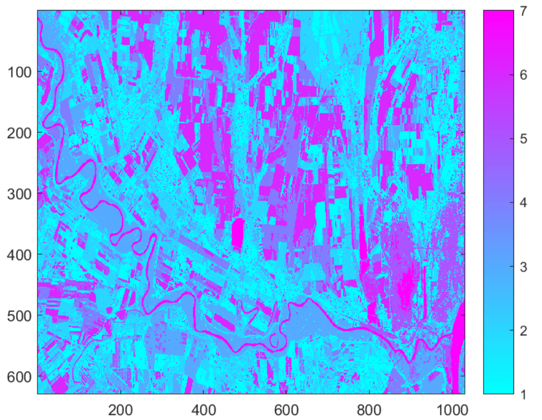

Fig. 17. Picture of the Independenta Hanu Conachi Forest - Galati County – year 2017– composition structure - colour map in conventional colours after compiling 6 spectral bands and considering a number of 12 clusters

Figures 10, 11 and 12 show how, by increasing the number of clusters in the representative space, image finesse increases. It also increases, but also slowly, the index of coincidence.

TABLE VII THE TOTAL EUCLIDIAN DISTANCES

No of band	No of clusters	Sum of squares – total distance
3	5	312.643
4	5	183.352
5	5	123.046
6	5	94.424
7	5	73.0831

The best results are obtained by combining a maximum number of spectral bands. Thus, by considering the 6 bands, we obtain an image with a structure close to that in Figure 4. Figures 13, 14 show the results obtained by combining the 6 bands.In Figures 15, 16 and 17, the combination of 2017 spectral images is shown by comparison. It is observed that by increasing the number of clusters in the representative space, the fineness of the image obtained increases. the index of coincidence with the figure in Figure 4 remains at values that cannot exceed 40%.

Compared to the detailed picture presented by ground monitoring services, we can see that there is a share of the identification of different tree parcels using this method. If for a small number of clusters, the correlation factor between the recorded and ground values is relatively small, in the case of images with a finer structure, the correlation factor has increased up to about 40%. The results are encouraging and the studies will continue.

III. CONCLUSIONS

In this paper we present the results of implementing a k-mean cluster algorithm to identify the composition of the forests using satellite imagery. The results obtained in comparison to ground data are encouraging. Considering the poor resolution of the images taken and considering the possibilities of improving the algorithm, the research should be continued.

ACKNOWLEDGMENT

This work was supported by Ministry of Research and Innovation project number 6PS, within Sectorial Program: Research to support the modernization of the national forestry monitoring system using remote sensing techniques and UAV systems.

REFERENCES

[1] Hahuie V., Georgescu L., Iticescu C., Murariu G., 2015. Investigation on Satellitar and UAV Cadastral Results. Case Study - Independenta Forest Areas, Annals of “Dunarea de Jos” University of Galati, Mathematics, Physics, Theoretical Mechanics, Fascicle II Year VII, No. 1, 26-32

[2] Hahuie V., E. Dănilă, A. Dănilă, L. Georgescu, 2015. Annals of “Dunarea de Jos” University of Galati, Mathematics, Physics, Theoretical Mechanics, Fascicle II Year VII, No. 1, 26-32,

[3] Coops N.C., Hilker T., Wulder M.A., St-Onge B., Newnham G., Siggins A., Trofymow J.T., 2007. Estimating canopy structure of Douglas-fir forest stands from discrete-return LiDAR. Trees 21(3): 295-310. DOI: 10.1007/s00468-006-0119-6.

[4] Eid T., Gobakken T., Næsset E., 2004. Comparing stand inventories for large areas based on photo-interpretation and laser scanning by means of cost-plus-loss analyses. Scandinavian Journal of Forest Research 19(6): 512- 523. DOI: 10.1080/02827580410019463.

[5] Falkowski M.J., Smith A.M.S., Hudak A.T., Gessler P.E., Vierling L.A., Crookston N.L., 2006. Automated estimation of individual conifer tree height and crown diameter via two-dimensional spatial wavelet analysis of LIDAR data. Canadian Journal of Remote Sensing 32(2): 153-161. DOI: 10.5589/m06-005.

[6] Ferdinent J.J., Padmanaban R.C., 2013. Development of a methodology to estimate biomass from tree height using airborne digital image. International Journal of Advanced Remote Sensing and GIS 2(1): 49-58.

[7] Gagnon P.A., Agnard J.P., Nolette C., 1993. Evaluation of a soft-copy photogrammetry system for tree-plot measurements. Canadian Journal of Remote Sensing 23(9): 1781-1785. DOI: 10.1139/x93-225.

[8] Gakis S.F., Orfanoudakis M., Papaioannou A., Mantzanas K., Papanastasis V. P., Seilopoulos D. G., Alifragis D.A., Kostakis S.N., 2014. Long term evolution of tree growth, understorey vegetation and soil properties in a silvo-pastoral system of Northern Greece, Annals of Forest Research 57(2):247-265

[9] Gruber M., Ponticellia M., Bernögger S., Leberl L., 2008.Ultracamx, the Large Format Digital Aerial Camera System by Vexcel Imaging/Microsoft. In: Chen J., Jiang J., Baudoin A. (eds.). Proceedings of ISPRS XXI st Congress "Silk Road for Information from Imagery", 3-11,

[10] Heurich M., 2008. Automatic recognition and measurement of single trees based on data from airborne laser scanning over the richly structured natural forests of the Bavarian Forest National Park. Forest Ecology and Management 255(7): 2416-2433. DOI: 10.1016/ j.foreco. 2008.01.022.

[11] Anttila P., 2005. Assessment of manual and automated methods for updating stand-level forest inventories based on aerial photography. PhD thesis, University of Joensuu, Faculty of Forestry, Joensuu, 42 p. Web: http:// www.metla.fi/ dissertationes/ df9.pdf.. Accessed: 2013.

[12] Arcangeli C., Klopf M., Hale S.E., Jenkins T.A.R., Hasenauer H., 2013. The uniform height curve method for height–diameter modelling: an application to Sitka spruce in Britain. Forestry 87: 177-186. DOI: 10.1093/ forestry/cpt041.

[13] Balenović I., Seletković A., Pernar R., Marjanović H., Vuletić D., Paladinić E., Kolić J., Benko M., 2011. Digital photogrammetry – State of the art and potential for application in forest management in Croatia. South- east European forestry 2 (2): 81-93. DOI: 10.15177/ seefor.11-09.

[14] Balenović I., Seletković A., Pernar R., Marjanović H., Vuletić D., Benko M., 2012. Comparison of classical terrestrial and photogrammetric method in creating management division. In Pentek T., Poršinsky T., Šporčić M. (eds) ''Forest Engineering - Concern, Knowledge and Accountability in Today's Environment'', 8-12 October

[15] S. Radulescu, M. Radulescu, Theorems and Problems of Mathematical Analysis, Editura: Didactica si Pedagogica, Bucharest, 1982.

[16] Benavides R., Douglas G.B., Osoro K., 2009. Silvopastoralism in New Zealand: review of effects of evergreen and deciduous trees on pasture dynamics. Agroforestry Systems 76: 327-350. DOI: 10.1007/s10457-008-9186-2009..