The University of Novi Sad
The Faculty of Technical Sciences
Department: Geodesy and Geomatics

Research Paper:

The Classification of Crops and the Yield Evaluation Using Vegetation Indices

Students: Miroslav Petrovic
Denis Papovic
Veljko Latinovic

SUMMARY:

Due to lagging behind the development of modern technologies, Serbia, as a country with the economy considerably depending on agriculture, is facing with the real problem of inability of observing the development of agricultural plants and crops so as to react timely and potentially influence on the improvement of the same. Adequate agrotechnical measures and modern technologies for observing the crops at macro level, and especially at the micro level, can highly contribute to the increase of productivity and competitiveness of the domestic agricultural production.

Remote sensing is a technology created by development of satellites. It has been used for the last several decades already, but its use for civil purposes has only recently become possible and economically justified. Its potential in agriculture is tremendous – from monitoring meteorological changes to observing the development level of crops (plants), monitoring plant diseases, floods, forest fires, classification of crops, air and water pollution, etc. The technology itself is very available, and a large quantity of geospatial data can be obtained for a relatively low price, the data which had been collected in the past directly in the field and afterwards processed in the laboratories.

In the further analysis conducted within the project, the emphasis is placed on discovering the optimal method for classification of crops in certain stages of growth, and on the attempt of determining the connection between vegetation indices and crops yield, as well as their interdependence in certain weather and climate conditions. The final aim of this project is a possibility to apply it on the rough estimate of the expected yield on the basis of vegetation indices measured at the height of plant growth, that is, at the moment when vegetation is the most luxuriant.

Address Offices in Brussels : Rue du Nord 76, BE – 1000 Bruxelles. Tel +32/2/217.39.72 Fax +32/2/219.31.47
E-mail: jean-yves.pirlot@clge.eu - www.clge.eu EC-Register of interest representatives 29077535421-69

1. INTRODUCTION

The software used for the data processing and analyzing is ERDAS Imagine 9.1. The analysis has been conducted on the Landsat (5 TM) images of spatial resolution of 30m, where, depending on the vegetation index, infrared and visible spectrum ranges have figured. The lots suitable for analysis on the territory of agricultural land property of Stari Tamiš, in the northern Serbia, have been taken as the observation area. Before the images were selected, it had been determined which plants were the most common on the lots which were suitable for analysis due to the low image resolution. By studying the chosen agricultural plants, depending on their sowing-time, the culmination period of crops growth is determined. This period is most favourable for defining plant features and for the prediction of yield quantity. On the basis of those researches, the decision has been made on using the images whose temporal feature corresponds to the previously determined period in the plant growth. In this project, the research has been conducted on the images from 2000 (May, July) and 2006 (July), which represent the marginal values of yield quantity in the indicated period (2000-2006).

The classification itself, based on the selection of the favourable vegetation indices, has been conducted by the engineering method, whereas the confirmation has been carried out by the *supervised* classification. The entire research is facing a series of limitations (image resolution, reliability of the data from the agricultural cooperative which refer to the size of the lot and the yield quantity, the lack of information on the time of the sowing), and thus the result itself is liable to checkings and slight variations for each image, especially if it is known that the values of the vegetation indices differ for each image due to different environmental conditions and the period when the images were made. The average accuracy of classification regarding to all methods after the application of some of the statistical correction methods and spatial models (mask) is satisfactory.

Address Offices in Brussels : Rue du Nord 76, BE – 1000 Bruxelles. Tel +32/2/217.39.72 Fax +32/2/219.31.47
E-mail: jean-yves.pirlot@clge.eu - www.clge.eu EC-Register of interest representatives 29077535421-69

2. VEGETATION INDICES

Vegetation indices are created to increase the vegetation signal by combining the information from different wavelength ranges, thus providing a piece of information which could not be available within neither of the single ranges. Those are mostly the combinations of the near-infrared (infrared range of short wavelengths has got the best reflexion from green plants) and some of the channels of the visible spectrum (mainly the red one). Mostly used indices for the analysis of vegetation are the following: RVI (*Ratio Vegetation Index*), NDVI (*Normalized Difference Vegetation Index*), GNDVI (*Green NDVI*), TNDVI (*Transform NDVI*), RNDVI (*Ratio NDVI*), EVI & EVI2 (*Enhanced Vegetation Index*), SAVI (*Soil Adjusted Vegetation Index*), OSAVI (*Optimal Soil Adjusted Vegetation Index*), CL (*Chlorophyll Index*), BI (*Brightness Index*).

Depending on the requirements of our research, we have taken into account the NDVI (Rouse, 1973), which in fact represents a relation between the red and the near-infrared spectrum, normalized to the values in the range of -1 and 1. The healthier plants (with higher concentration of chlorophyll) absorb high quantity of the red spectrum, and just as well reflect the near-infrared light. If the NDVI of some plant is lower than the average for that plant, it means that the observed plant does not perform the process of photosynthesis so well, which indicates the lower development level of the plant or some disease, that is, it requires additional nursing. The problems which appear with this index are:

- the value for the vegetations with similar biophysical features can be different, depending on the reflexion of the soil (Bausch, 1993),
- it is sensitive to the atmospheric effects which cause scattering and absorption of energy by aerosols, especially the red range, because the shorter wavelengths are more liable to the influence of aerosols (Ben-Ze'ev et al., 2006),
- the limitation when the luxuriance of the vegetation with the average or high LAI (*leaf area index)* is determined because the higher LAI values bring to the saturation of the NDVI, and almost linear interdependence between NDVI and LAI ceases being in effect (Gitelson, 2004).

EVI is an index created to bridge the shortcomings of the NDVI. It provides better sensitivity within the ecosystems with higher LAI (Boegh et al., 2002), simultaneously decreasing the influence of the reflexion of the soil itself and the atmospheric scattering. EVI2 (Jiang, 2007) is used because of the data from older satellites which do not have the blue range (only the red one and NIR), and that is the reason why it is used more often. It provides similar results as EVI, and therefore it was taken into account during the process of analysis.

Address Offices in Brussels : Rue du Nord 76, BE – 1000 Bruxelles. Tel +32/2/217.39.72 Fax +32/2/219.31.47
E-mail: jean-yves.pirlot@clge.eu - www.clge.eu EC-Register of interest representatives 29077535421-69

In case of very low vegetations, the soil itself has an influence on measuring the spectral reflexion, creating interference whose reflexion is interfering with the reflexion of the plant. This is especially problematic as regards to the comparison of the same plant on different types of soil (which reflect different quantities of light in NIR and RED spectrum). Due to this interference, one plant can be wrongly recognized as two different ones. The best way to correct that error is the use of SAVI, which in fact represents NDVI with the implemented correction that depends on the soil brightness. The values of this correction are ranging within the interval of 0 to 1. Very thick vegetation has the negligible reflexion from the soil itself, and thus the correction value equals 0 (NDVI), whereas the value concerning the soil with the negligible vegetation is 1. Determining this correction is not simple due to the dynamics of soil changes, so SAVI with the optimal correction factor – OSAVI (Rondeaux, 1996) is the most commonly used.

Gitelson et al. (1996) present *Green Normalized Difference Vegetation Index* (GNDVI) using the green channel instead of the red one for calculating the vegetation index. Taking into account that in some cases GNDVI is more sensitive to the chlorophyll concentration in the leaves than NDVI, it was taken into consideration within this research.

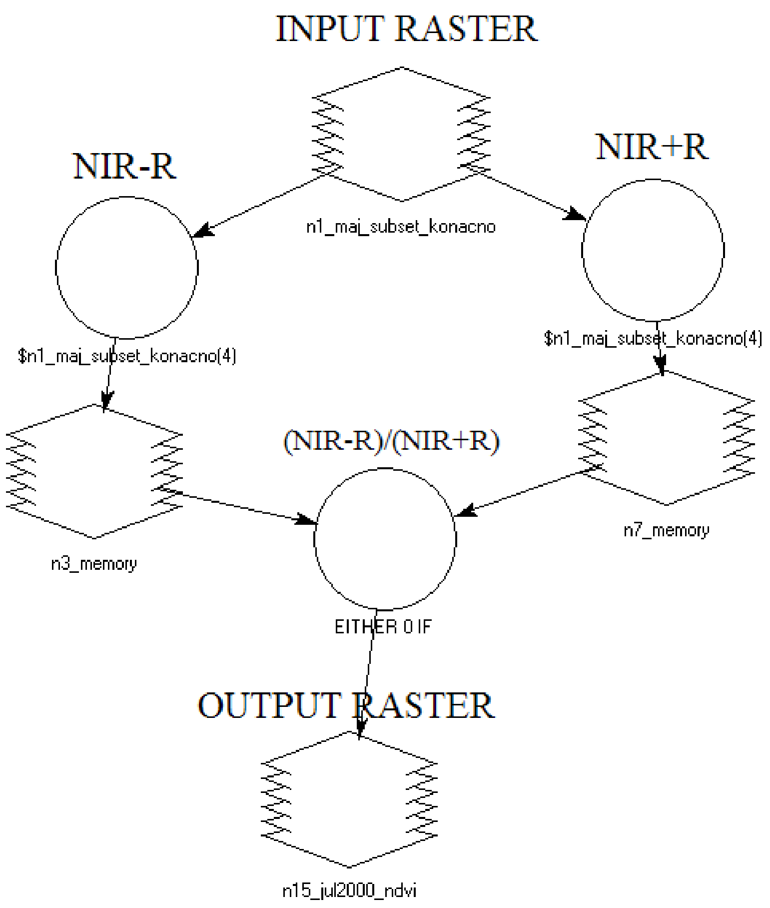
Gitelson et al. (2005) officially announced *chlorophyll index* (Cl), which is strictly specialized for the estimation of the total chlorophyll concentration in the plant cover, because of which it is useful in the final estimation of the yield of tested crops. Just like GNDVI, it uses the green channel instead of the red one for calculating the index value.

The formulae for calculating the abovementioned indices are:

$$NDVI = (NIR-R)/(NIR+R)$$
$$EVI2 = 2.5(NIR - red)/(NIR + red + 1)$$
$$OSAVI = (NIR-R)/(NIR+R+0.16)$$
$$GNDVI = (NIR - G)/(NIR + G)$$
$$Cl = (NIR/G) - 1$$

Address Offices in Brussels : Rue du Nord 76, BE – 1000 Bruxelles. Tel +32/2/217.39.72 Fax +32/2/219.31.47
E-mail: jean-yves.pirlot@clge.eu - www.clge.eu EC-Register of interest representatives 29077535421-69

Normalized Difference Vegetation Index



Example of a *spatial model* for calculating NDVI

Address Offices in Brussels : Rue du Nord 76, BE – 1000 Bruxelles. Tel +32/2/217.39.72 Fax +32/2/219.31.47
E-mail: jean-yves.pirlot@clge.eu - www.clge.eu EC-Register of interest representatives 29077535421-69

3. Crop classification on the basis of vegetation indices

Throughout the entire process, from finding the adequate areas on the Landsat's image, forming vegetation indices for those areas, to detecting plants on the basis of the VI values, as well as the analyses and accuracy verifications of the given results, software package Erdas Imagine 9.1 was used.

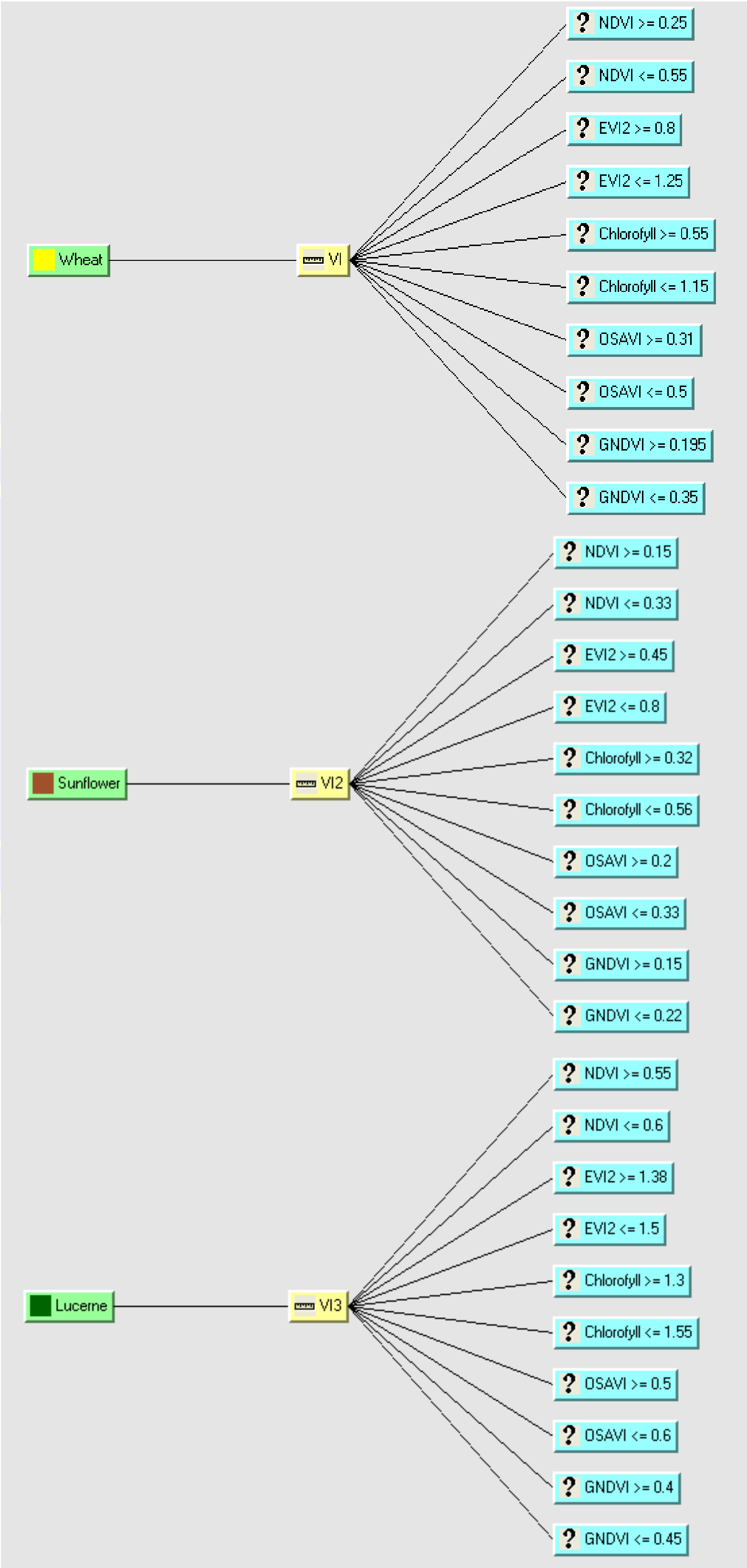
The first step in classifying denotes the limitation of the image to an area for which the information on the sowed crops and the yield exist, on the basis of which the sought relation between the crops and the VI will be established, as well as between the VI and the yield.

On the basis of the statistical analysis of the lots under certain crops and the VI value related to the same lots, the value interval of certain vegetation indices, which is characteristic for the observed type of crops, is being determined. Marginal values of vegetation indices for each plant were used within the engineering classification process. All vegetation indices were used together while classifying each plant, which implies the condition that one pixel would be classified as some of the plants is suchlike that the values of all vegetation indices within this pixel lie within the defined limits.

The image dating from the 25th May 2000 is convenient for the observation of wheat, sunflower and lucerne, concerning the sown plants and the growth stage.

The image of the adequate area dating back from May, 2000 (true colours)

Engineering classification has been conducted as follows:

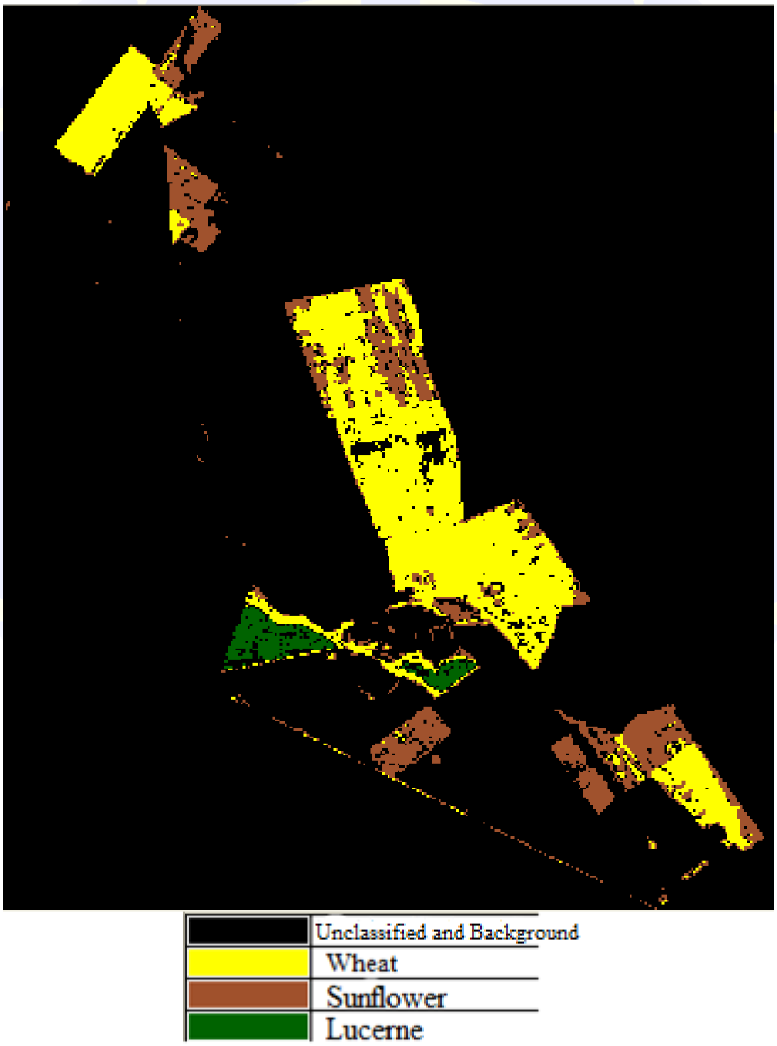


Example of classification using *knowledge engineer* tools

Address Offices in Brussels : Rue du Nord 76, BE – 1000 Bruxelles. Tel +32/2/217.39.72 Fax +32/2/219.31.47
E-mail: jean-yves.pirlot@clge.eu - www.clge.eu EC-Register of interest representatives 29077535421-69

The result of the classification are the areas under abovementioned plants. The error made while classifying, that is, the difference between the area under certain plants detected by the classification and the real area of those plants, expressed in percents, is following:

- Wheat – 22%
- Sunflower – 33%
- Lucerne – 36%



the Result of the Engineering Classification (May, 2000)

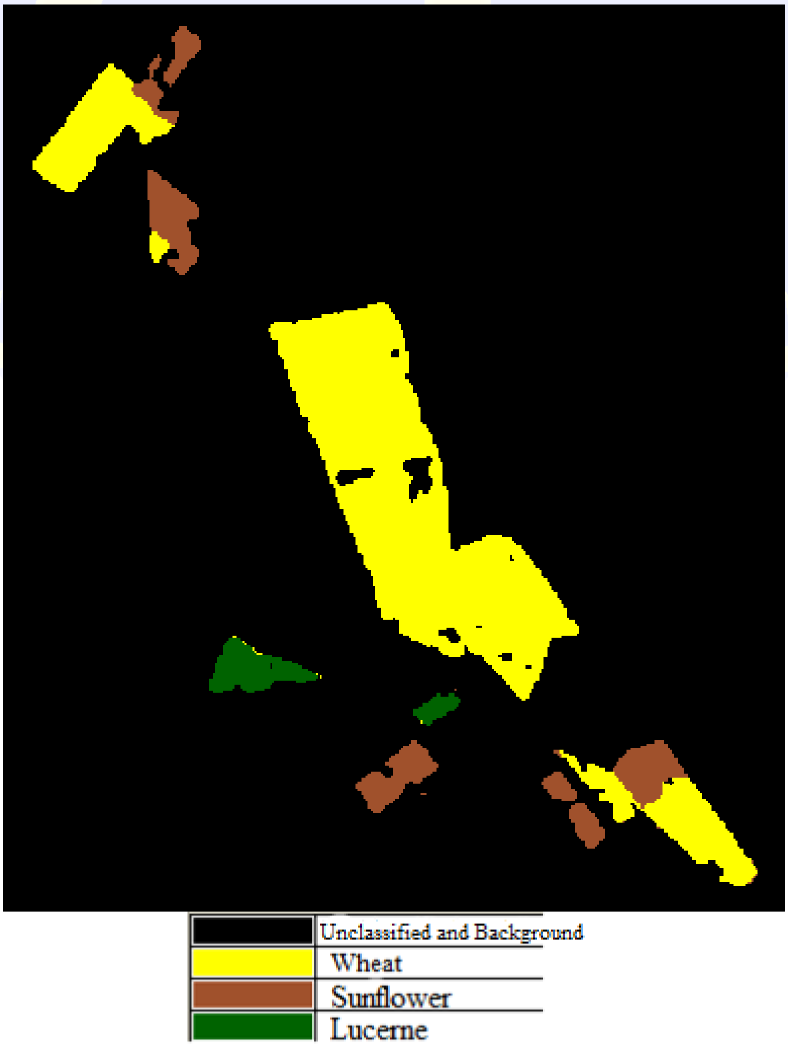
Of course, it is possible to apply certain statistical corrections so as to increase the accuracy. In this case, statistical filters integrated in the applied software package were used – *Majority function* with the window size of 3x3, which divides the image on the windows of 3x3 pixels and assigns them the values of pixels which predominate in the window. Similar purpose is a feature of the used function *area fill (majority)*, which converts all pixels within the defined area into a predominant one.

Address Offices in Brussels : Rue du Nord 76, BE – 1000 Bruxelles. Tel +32/2/217.39.72 Fax +32/2/219.31.47
E-mail: jean-yves.pirlot@clge.eu - www.clge.eu EC-Register of interest representatives 29077535421-69

These adjustments have a justified role, for it can be stated that only one type of plant is sown at each lot. A spatial model called ”*mask*” was used as an additional mode of classification correction. Namely, it is based on the use of multitemporal images. In this case, there was a mutual use of May and July images from the same year. On the basis of a logical assumption which ensues from the knowledge of the agricultural laws, on the territories within which certain plant exists in May, they will be naked soil or under scarce vegetation in July, and therefore that statement is used as a filter of the attained results.

After the previously mentioned corrections were applied, the error in detecting certain plants is equal to:

- Wheat – 0%
- Sunflower – 28%
- Lucerne – 26%



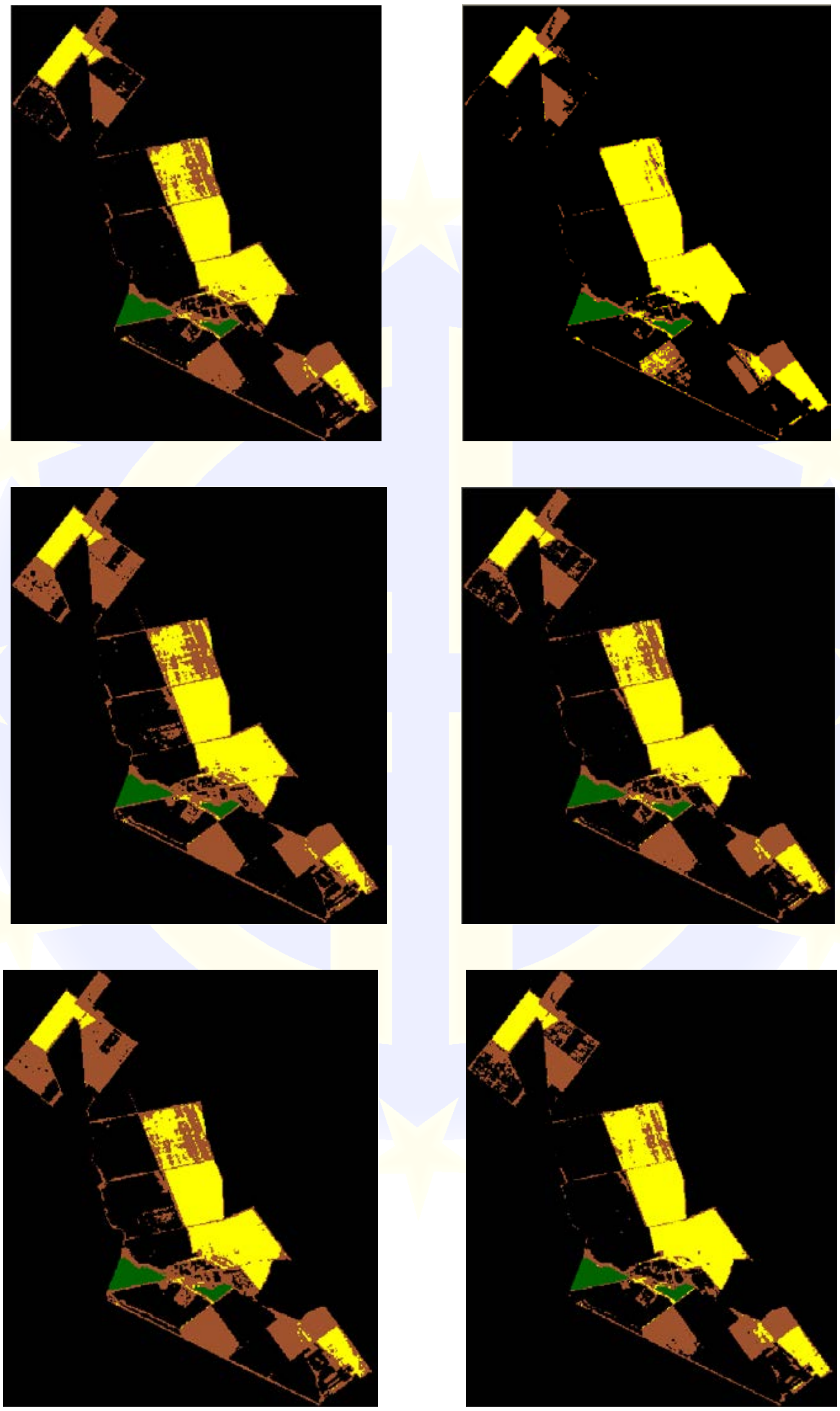
the Result of the Engineering Classification with the subsequent corrections (May, 2000)

A comparative classification has been conducted by the *supervised* method as a confirmation of detecting the crops using the VI. Each plant is represented by one training sample whose parameters vary so that it could be shown empirically which parameters give the best final results. The changing parameters are the set area expressed in number of pixels, and the spectral distance. The success in those classifications is listed in the following table:

area	**spectral distance**	**error (without corrections)**			**error (with corrections)**		
		wheat	sunflower	lucerne	wheat	sunflower	lucerne
500	15	20%	3%	7%	14%	1%	3%
500	20	24%	2%	7%	16%	0%	2%
500	25	23%	1%	6%	15%	1%	3%
1000	15	4%	11%	6%	0%	7%	3%
1000	20	18%	1%	6%	10%	~~4%~~	3%
1000	25	11%	2%	6%	8%	2%	3%

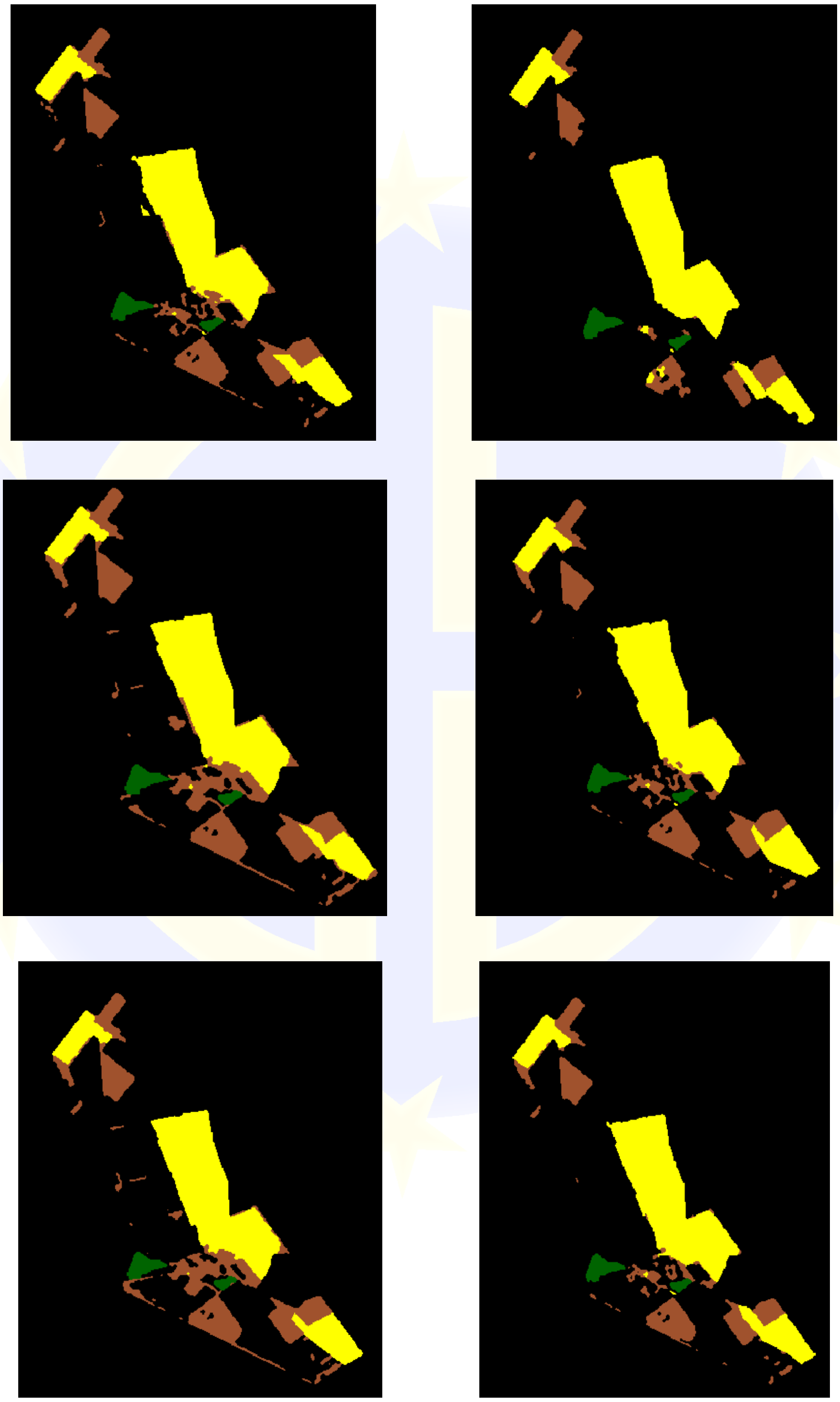
There are cases when the application of corrections degrades the results. The reason is the incorrect identification of some other crops or weeds as a plant of interest, which occurs mostly due to poor spatial image resolution (inability of discerning the borders of the lot, which results in adding area beside certain plant to the plant itself).

Address Offices in Brussels : Rue du Nord 76, BE – 1000 Bruxelles. Tel +32/2/217.39.72 Fax +32/2/219.31.47
E-mail: jean-yves.pirlot@clge.eu - www.clge.eu EC-Register of interest representatives 29077535421-69



the Results of the *supervised* classification (without corrections):
left (500px area) – from the top 15, 20, 25
right (1000px area) – from the top 15, 20, 25

Address Offices in Brussels : Rue du Nord 76, BE – 1000 Bruxelles. Tel +32/2/217.39.72 Fax +32/2/219.31.47
E-mail: jean-yves.pirlot@clge.eu - www.clge.eu EC-Register of interest representatives 29077535421-69



the Results of the *supervised* classification (with corrections):
left (500px area) – from the top 15, 20, 25
right (1000px area) – from the top 15, 20, 25

Address Offices in Brussels : Rue du Nord 76, BE – 1000 Bruxelles. Tel +32/2/217.39.72 Fax +32/2/219.31.47
E-mail: jean-yves.pirlot@clge.eu - www.clge.eu EC-Register of interest representatives 29077535421-69

The image from the 28th July 2000 is suitable for the analysis of maize and sugar-beet.

The Image of the suitable area dating back from July, 2000 (true colours)

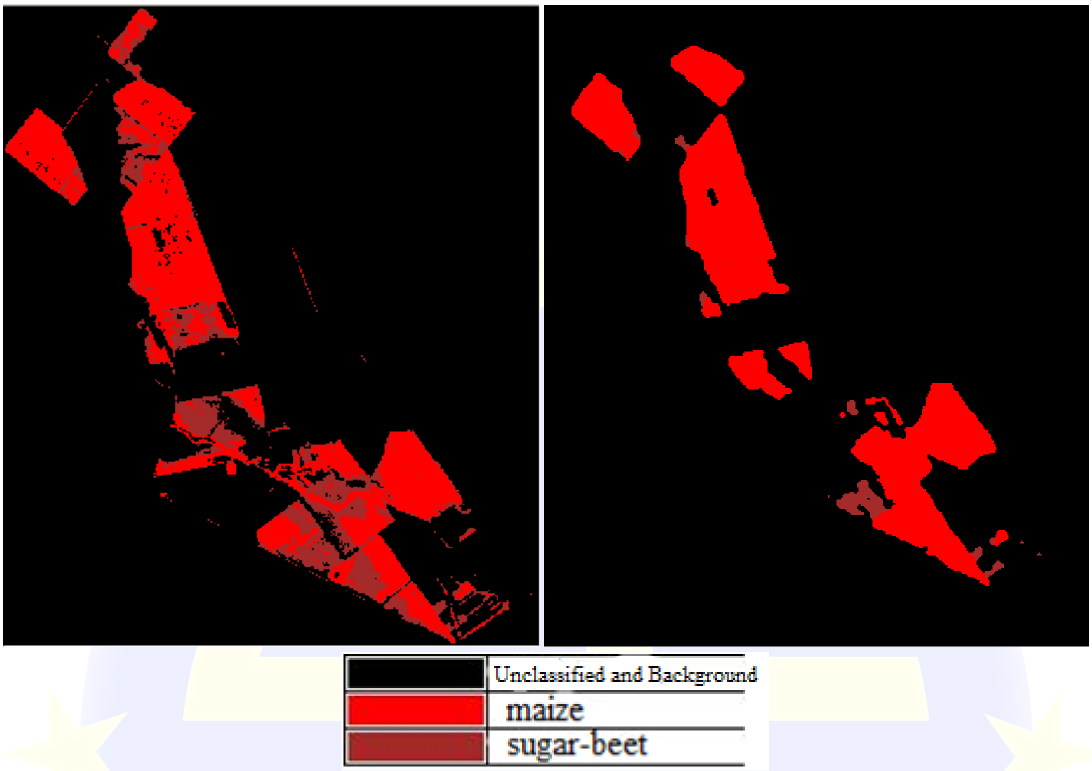
The classification was conducted with the error of:

- Maize – 22%
- Sugar-beet – 43%

After apllying corrections, the error was lowered to:

- Maize – 3%
- Sugar-beet – 1%

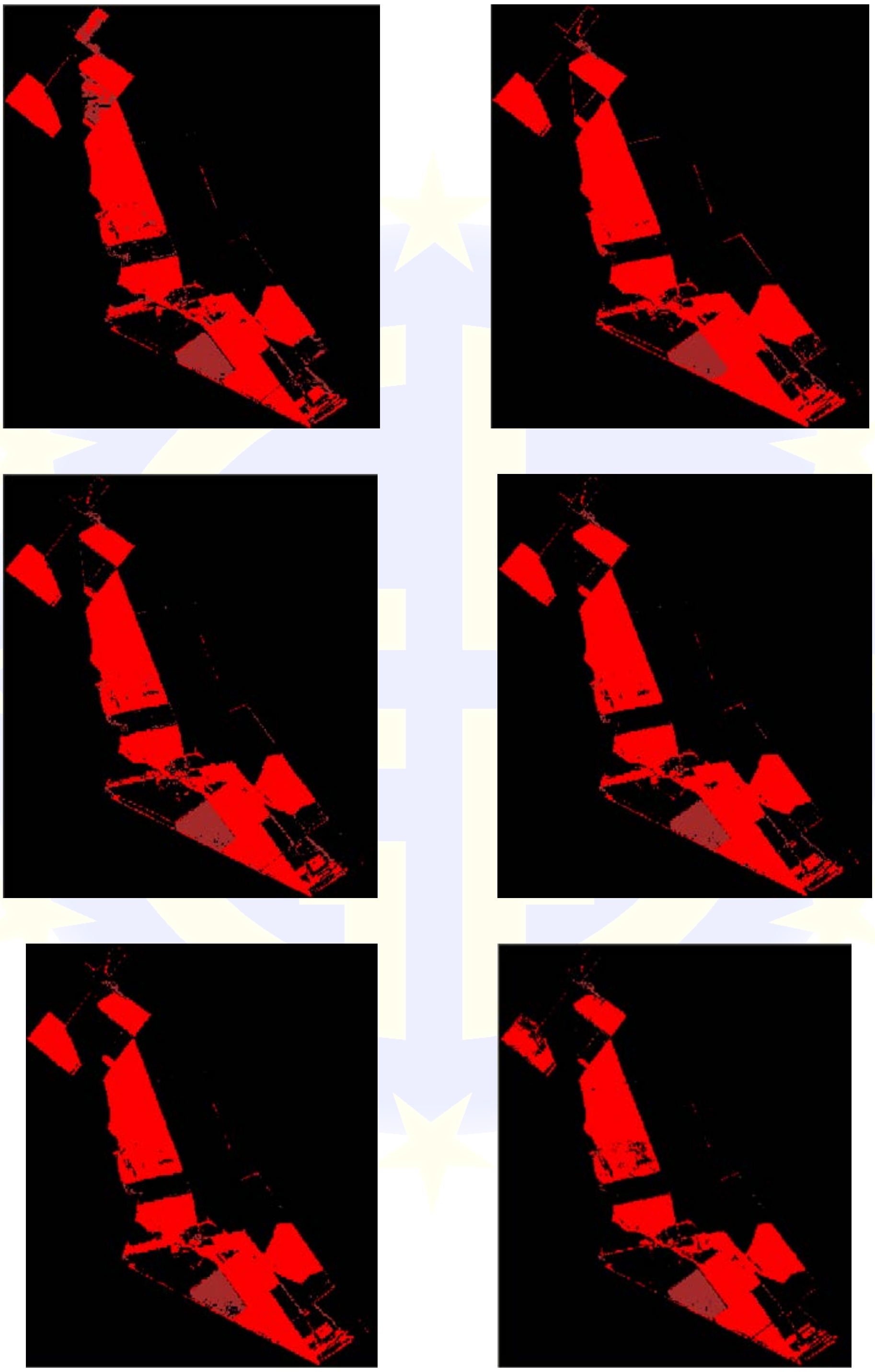
Address Offices in Brussels : Rue du Nord 76, BE – 1000 Bruxelles. Tel +32/2/217.39.72 Fax +32/2/219.31.47
E-mail: jean-yves.pirlot@clge.eu - www.clge.eu EC-Register of interest representatives 29077535421-69



The result of engineering classification in the July, 2000 image (on the right – with corrections)

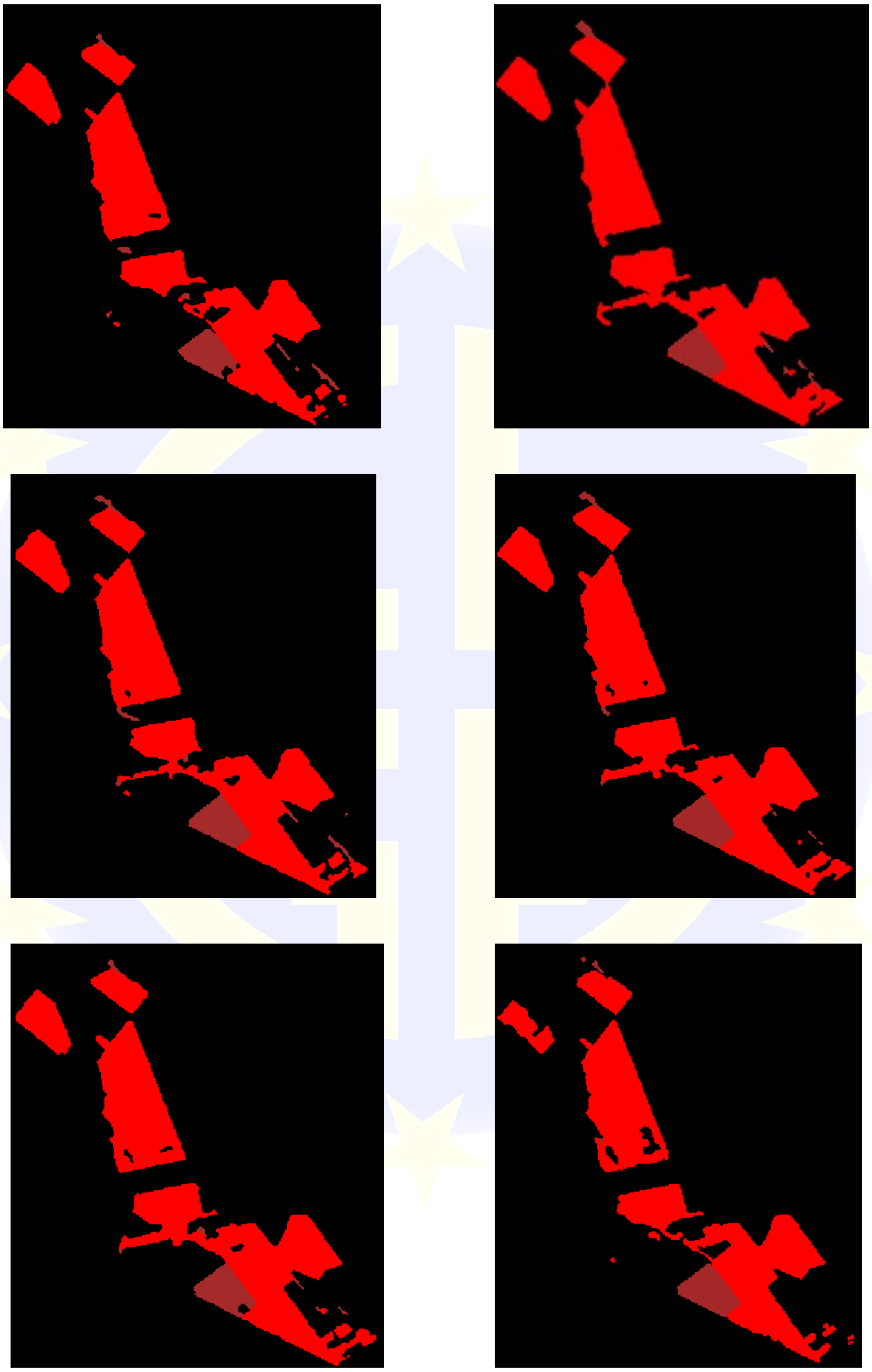
Similarly to the May image, the *supervised* classification was being conducted simultaneously. The attained results were as following:

area	**spectral distance**	**error (without corrections)**		**error (with corrections)**	
		maize	sugar-beet	maize	sugar-beet
500	15	3%	16%	4%	11%
500	20	2%	3%	5%	1%
500	25	1%	12%	7%	10%
1000	15	1%	10%	7%	5%
1000	20	1%	6%	6%	1%
1000	25	2%	7%	0%	0%



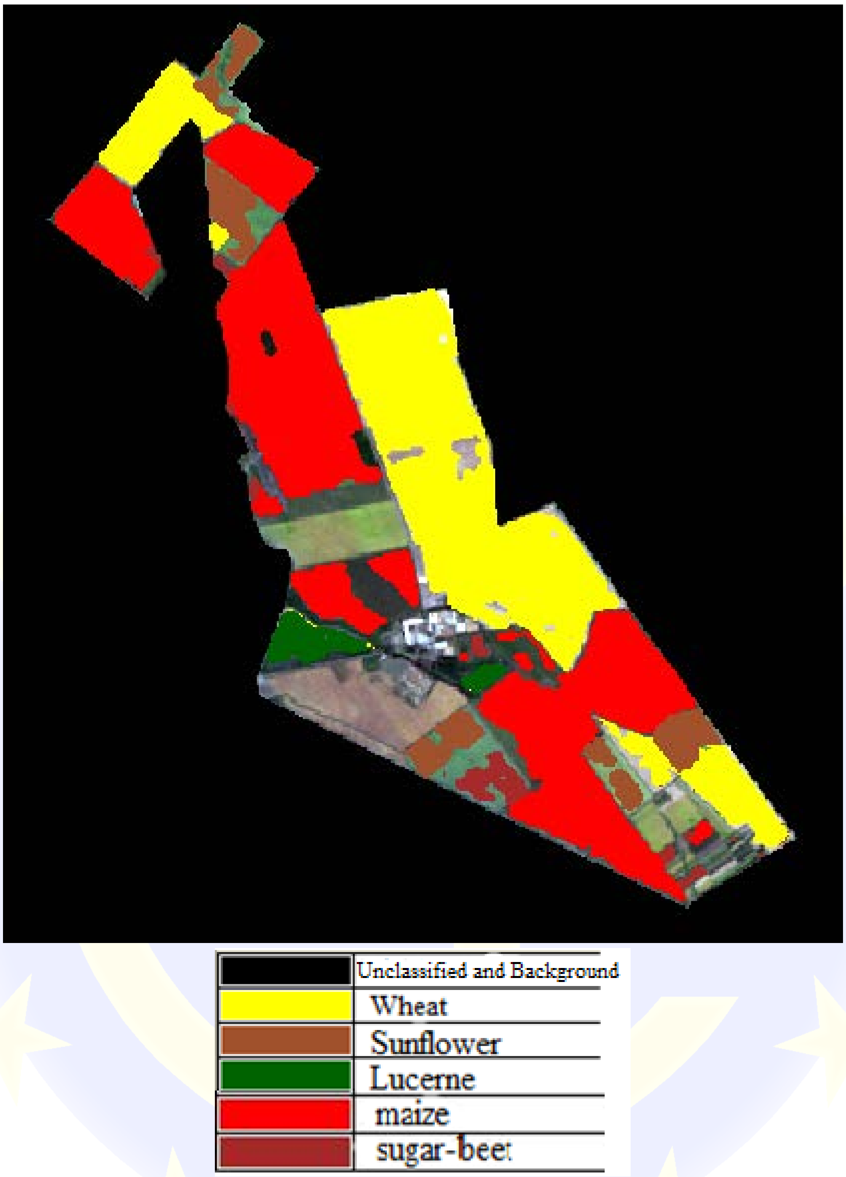
the results of the *supervised* classification (without corrections):
left (500px area) – from the top 15, 20, 25
right (1000px area) – from the top 15, 20, 25

Address Offices in Brussels : Rue du Nord 76, BE – 1000 Bruxelles. Tel +32/2/217.39.72 Fax +32/2/219.31.47
E-mail: jean-yves.pirlot@clge.eu - www.clge.eu EC-Register of interest representatives 29077535421-69



the results of the *supervised* classification (with corrections)
left (500px area) – from the top 15, 20, 25
right (1000px area) – from the top 15, 20, 25

Address Offices in Brussels : Rue du Nord 76, BE – 1000 Bruxelles. Tel +32/2/217.39.72 Fax +32/2/219.31.47
E-mail: jean-yves.pirlot@clge.eu - www.clge.eu EC-Register of interest representatives 29077535421-69



overlapping of the crops classified by the engineering method over the *true colour* image

After the same VI values were applied it the images dating from different years, it has appeared that the VI values are unique for each image due to different enviromental conditions, and thus it cannot be confidently stated that the model, as a result for one image, may be used for the classification of crops with the other images, made in some other time interval or on some other territory. With every new analysis, it is necessary to correct the parameters, that is, the VI, and to check the real situation in the field so as to establish an adequate link between the VI and the plants themselves.

4. Estimation of yield using the vegetation indices

Address Offices in Brussels : Rue du Nord 76, BE – 1000 Bruxelles. Tel +32/2/217.39.72 Fax +32/2/219.31.47
E-mail: jean-yves.pirlot@clge.eu - www.clge.eu EC-Register of interest representatives 29077535421-69

The estimation of yield is based on the assumption that the yield quantity mostly depends on the vegetation luxuriance (the higher concentration of chlorophyll, the healthier the plant) at the height of the vegetation development period of the plant (Groten, 1993). In consideration of the fact that the vegetation index Cl is the most sensitive in estimations of the chlorophyll concentration in the vegetation, it had the priority in the conducted research.

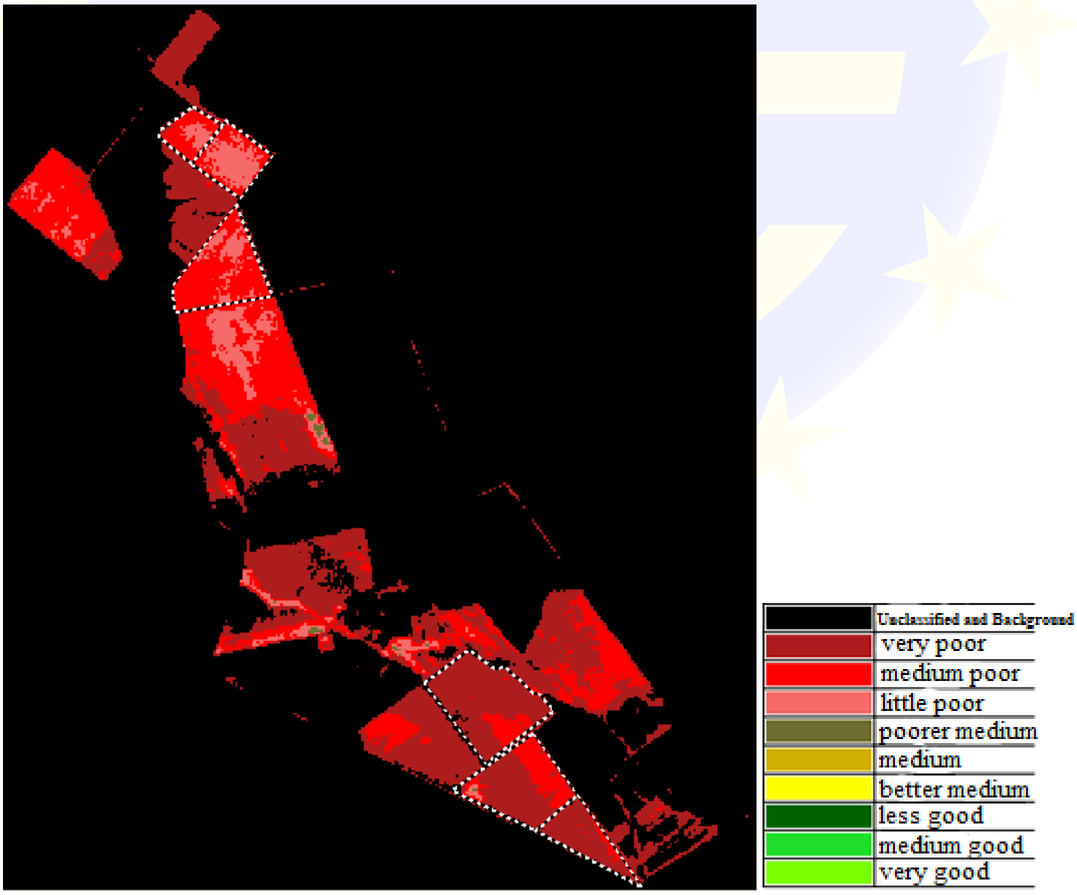
Due to the existing differences in the yield quantity over the years, the images from 2000 and 2006 were taken into account, because, according to the data given by the agricultural cooperative ”Stari Tamiš”, those were the years of the worst and the best yield, respectively. Since the only available image in 2006 was from July, the yield estimate refers to the plant which had previously been analyzed in the same month in 2000 – maize (there are no conducted researches for sugar-beet due to the small area of the lots under this plant).

The lots under this agricultural plant in both years are marked on the following image:

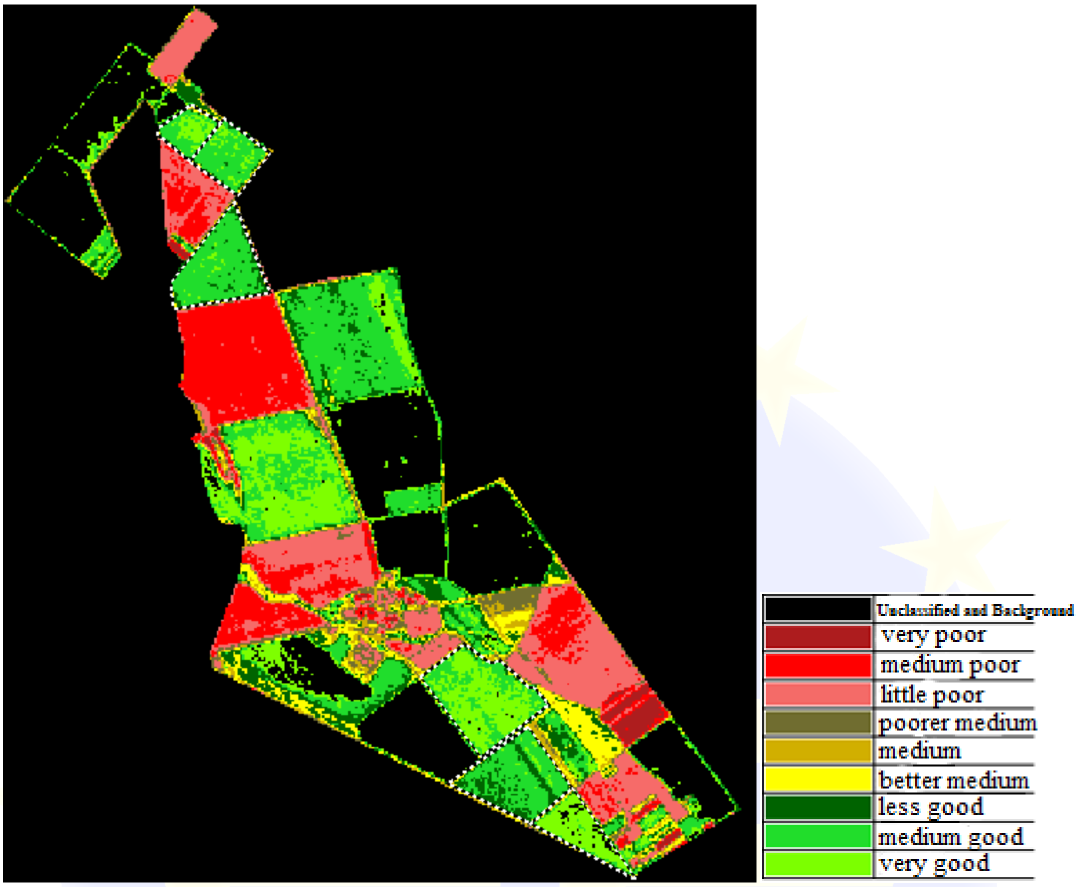
Maize yield quantities on those lots are listed in the following table:

Maize				
year 2000		year 2006		
lot number	yield (*t/ha*)	lot number	yield (*t/ha*)	area (*ha*)
15	7.18	15	10.83	93
20	4.02	20	11.81	120
21	4.7	21	11.12	96
22	4.12	22	10.83	50
44	7.18	44	12.8	49
45	7.18	45	11.7	30

The data clearly show big difference between the yield quantities in these two years. According to the width of the yield value interval, the yield quantity in the research is divided into nine categories. Consequently, the values of the *chlorophyll index* are divided into nine corresponding categories. The results obtained by the engineering classification confirm the predicted proportional correlation between the vegetation indices and the yield quantity.



rough estimate of maize yield (July, 2000),
observed common lots for two different years are edged with a dashed line
(it should be noted that on the surrounding lots are presented estimates of yield for different cultures, which do not need to apply quality classes defined for maize yield)

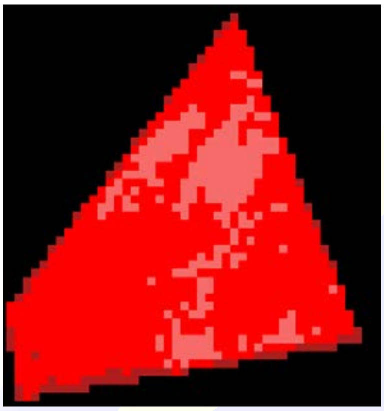
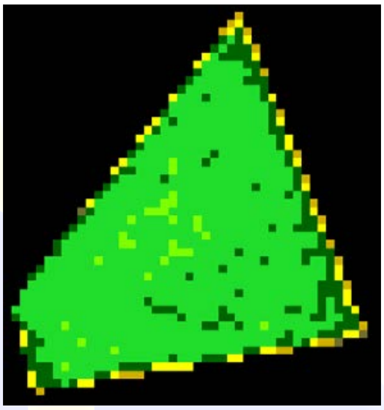


rough estimate of maize yield (July, 2006)
observed common lots for two different years are edged with a dashed line
(it should be noted that on the surrounding lots are presented estimates of yield for different cultures, which do not need to apply quality classes defined for maize yield)

The problem, which has appeared during the analysis, is interference of the VI values of maize with the VI values of some other crops. The problem has arisen, first of all, due to big difference concerning luxuriance and health of not only maize, but the other plants as well, which led to big differences between Cl index values of the same plants between the two images. The index values of the very poor maize in the image from the year 2000 are identical to those of some other plants that otherwise contain less chlorophyll, but in the image from 2006 achieved a great ”success”.

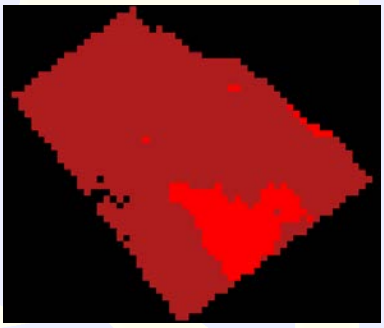
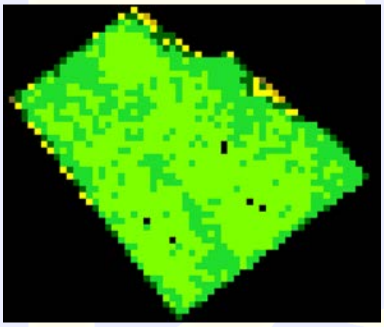
Comparative review of common lots:

- lot number 15:

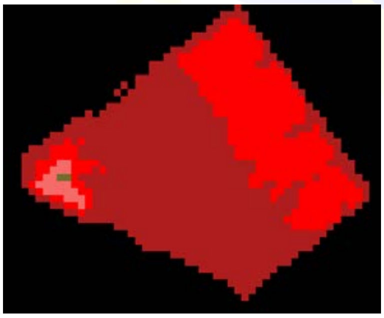
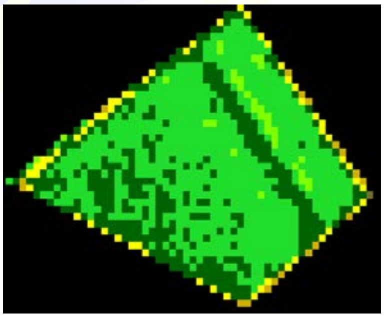
Year 2000 (left) – medium poor yield predominates
Year 2006 (right) – medium good yield predominates

- lot number 20:

Year 2000 (left) – very poor yield predominates
Year 2006 (right) – very good yield predominates

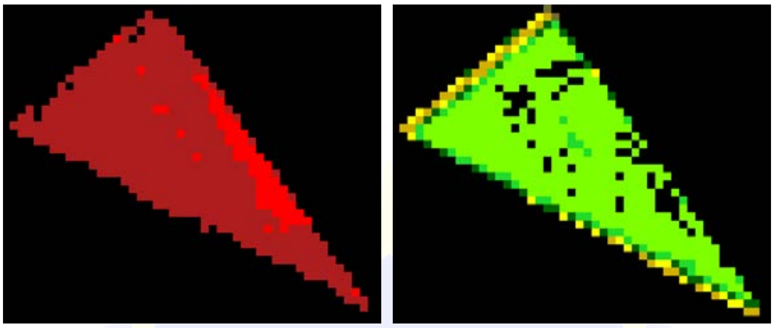
- lot number 21:

Year 2000 (left) – very poor yield predominates
Year 2006 (right) – medium good yield predominates

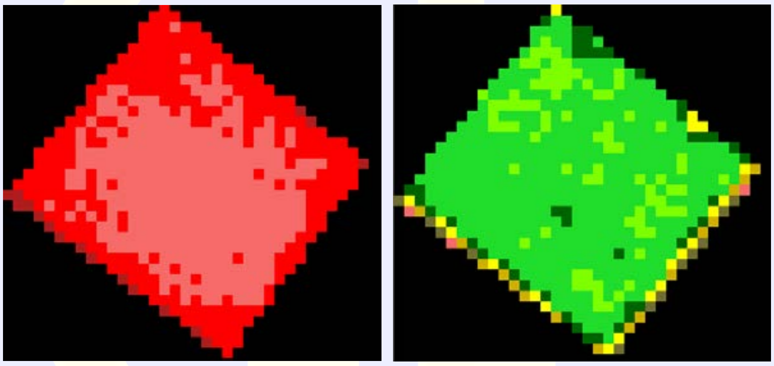
Address Offices in Brussels : Rue du Nord 76, BE – 1000 Bruxelles. Tel +32/2/217.39.72 Fax +32/2/219.31.47
E-mail: jean-yves.pirlot@clge.eu - www.clge.eu EC-Register of interest representatives 29077535421-69

- lot number 22:



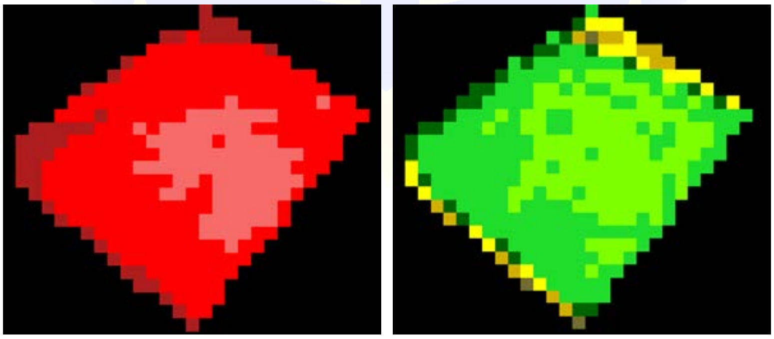
Year 2000 (left) – very poor yield predominates
Year 2006 (right) – very good yield predominates

- lot number 44:



Year 2000 (left) – little poor yield predominates
Year 2006 (right) – medium good yield predominates

- lot number 45:



Year 2000 (left) – medium poor yield predominates
Year 2006 (right) – medium good yield predominates

5. Conclusion of the Research

Address Offices in Brussels : Rue du Nord 76, BE – 1000 Bruxelles. Tel +32/2/217.39.72 Fax +32/2/219.31.47
E-mail: jean-yves.pirlot@clge.eu - www.clge.eu EC-Register of interest representatives 29077535421-69

The research has been burdened from the start by the limitation regarding poor funds, which is why free images of poor spatial resolution where used. Also, there was a lack of information concerning the time interval of sowing, which is the reason why the plants suitable for analysis have been chosen on the basis of the assumption and the previous experiences.

Yield estimation is possible when the accurate data are obtained on when and where each plant is sown, because thusly potential interferences in classification of different plants can be avoided (e.g. matching of index values of very poor maize with sunflower of good quality). All this leads to a conclusion that the same VI values during the classification cannot be applied in the images made in different conditions (first of all, due to different progression of a plant or a sowing moment, which results in different chlorophyll concentration in the plant), but a rough estimation on yield quantity can be made if there is a reliable datum that some plant exists on certain territory, for which an exact stage of development is known.

Research has shown that the yield estimates in this way is possible and applicable method if it meets the minimum requirements, such as details about the type of crop on a particular lot and its date of sowing. Of course, it is necessary to make categorization of vegetation index values, which define the amount of yield, based on the analysis of yield in previous years.

References:

Bausch, W. C. (1993). Soil background effects on reflectance-based crop coefficients for corn. Remote Sensing of Environment, 46, 213−222,

Ben-Ze'ev, E., Karnieli, A., Agam, N., Kaufman, Y., & Holben, B. (2006). Assessing vegetation condition in the presence of biomass burning smoke by applying the aerosol-free vegetation index (AFRI) on MODIS. International Journal of Remote Sensing, 27, 3203-3221,

Boegh, E., Soegaard, H., Broge, N., Hasager, C. B., Jensen, N. O., Schelde, K., et al. (2002). Airborne multispectral data for quantifying leaf area index, nitrogen concentration, and photosynthetic efficiency in agriculture. Remote Sensing of Environment, 81, 179−193,

Gitelson, A. A. (2004). Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. Journal of Plant Physiology, 161, 165−173,

Gitelson, A. A., Vina, A., Ciganda, V., Rundquist, C. D., & Arkebauer, J. T. (2005). Remote estimation of canopy chlorophyll content in crops. Geophysical Research Letters, 32, L08403. doi:10.1029/2005GL022688,

Gitelson, A.A., Y.J. Kaufman, and M.N. Merzlyak, 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS, Remote Sens. Environ. 58:289–298,

Groten, S. M. E. (1993). NDVI - crop monitoring and early yield assessment of Burkina Faso. International Journal of Remote Sensing, 14(8), 1495-1515,

Jiang, Zhangyan, Alfredo R. Huete, Youngwook Kim, and Kamel Didan, 2007. 2-band enhanced vegetation index without a blue band and its application to AVHRR data, Proc. SPIE, Remote Sensing and Modeling Theory, Techniques, And Applications I, Vol. 6679, 667905,

Rondeaux, G., M. Steven and F. Baret, 1996. Optimization of soil-adjusted vegetation indices, *Remote Sensing of Environment*, 55, pp.95–107,

Rouse, J. W., R. H. Haas, J. A. Schell, and D. W. Deering, 1973. Monitoring vegetation systems in the Great Plains with ERTS, Third ERTS Symposium, NASA SP-351 I, pp. 309-317.

.

Address Offices in Brussels : Rue du Nord 76, BE – 1000 Bruxelles. Tel +32/2/217.39.72 Fax +32/2/219.31.47
E-mail: jean-yves.pirlot@clge.eu - www.clge.eu EC-Register of interest representatives 29077535421-69