Using remote sensing to map in-field variability: Clustering Approaches

M. Verőné Wojtaszek
Institute of Geoinformatics, Alba Regia Faculty of Engineering, Óbuda University, Székesfehérvár, Hungary
wojtaszek.malgorzata@amk.uni-obuda.hu

Abstract— **Precision Farming concept is spreading rapidly worldwide as a tool to enable farmers for profitable production while fulfilling environmental and food safety conditions. The introduction and application of precision technologies in agriculture has been motivated by the high degree of variability of agro-ecological conditions within fields. With increased use of precision agriculture techniques information concerning within field soil and crop variability is becoming increasingly important for effective crop management. As a result of new technological advances, more and more data is available. The safe and reliable transfer of different kinds of data into detailed information for management purposes is also of increasing importance. In this paper an overview is given on the use of remote sensing data in site-specific farming for assessing crop growth and yield variability. Many studies have been carried out to find an appropriate method to classify the high resolution remote sensing data within a field. According to the results of my research two classification approach - pixel-based and object-based (OBIA) - are presented. Image processing techniques including vegetation indices, segmentation, and classification were used in this research. The models presented in the article are suggested to be used for crop monitoring and supporting decision making.**

I. DATA GATHERING

Precision agriculture technology is a farm management system, which relies on various measurements, data collections and analysis, as well as decision making. Measurements include soil chemical and physical characteristics determination, grain yield and quality measurements, and several remotely sensed property determination. The base of precision agriculture is information on the spatial and temporal variability of soil and crop characteristics as well as the environment and external conditions [7, 10]. The data on the properties of soils, plants, occurrence of pests, obtained yield and meteorological parameters are needed to optimize field management [1, 2]. How to measure and map the spatially and temporally varied field parameters accurately and efficiently has been the focus of the researches in the field. In many studies remote sensing data are used as a source of information and GIS technology as tool for data analysis. Geographical information systems (GIS) are systems for the storage, analysis and presentation of spatial data. GIS can also support translation of the research findings into operational systems for use at farm level by providing a good platform for storage of base data, simple modelling, presentation of results, development of a user interface and, in combination with a global positioning system, controlling the navigation of farm vehicles. On the basis of GIS a decision support system could be developed for operational application of precision agriculture at farm level [6, 8].

Remote sensing technologies are being used more and more often in agriculture providing data for monitoring within and between-field. In parallel with development of spatial, temporal and spectral resolution in the last decade the quantity and quality of data is constantly increasing as well. Results of research in the field from all over the world have provided a fundamental information relating spectral properties of soils and crops to their agronomic and biophysical characteristics. This knowledge has facilitated the development and use of various remote sensing methods to detect spatially and temporally varied environmental stresses which limit crop productivity [1, 2, 10]. This can make significant contribution in optimizing crop management as sowing, irrigation, fertilization and harvest. However, gathering, accessing, and processing of remote sensing images from different satellites require high technical skills and agricultural knowledge as well. Besides expertise, a computer (software and hardware) background is also required. Processing large amounts of data can be time consuming. The lack of comprehensive software platforms to extract useful spatially and temporally varied information from remote sensing and other sources, and the lack of knowledge can hindered the wide application of technology to support precision farming.

II. DATA GATHERING

The introduction of precision technologies in agriculture has been motivated by the high degree of variability of agro-ecological conditions within fields. One of the criterion for introducing precision agricultural technologies is the development of an up-to-date arable crop information system that provides information on soils, crop land cultivation, plant status, etc. [1, 5, 11, 12]. This information can be used as starting data for cultivation and predicting yield estimate. In order to set up such an information system, it is essential to use modern data gathering and analysis technologies. Remote sensing is the most effective tool for surveying the Earth's surface and tracking its changes.

With increased use of precision agriculture techniques, information on crop yield variability within-field is becoming increasingly important for effective crop management. The yield map integrates the effects of various spatial variables so that it provide also information on soil properties, topography, plant population, nutrient replacement practices and applied agrotechnology. A yield map can therefore been indispensable input for site-specific operations either by itself or in combination with other spatial data. Yield data can be obtained at harvest by harvester-mounted yield monitors. The first yield measuring devices were applied in the early '80s, since that they have started to become standard equipment on combines. With technological advances of harvester-mounted yield monitors and Global Position Systems (GPSs) farmers are able to collect intensive and accurate yield data during harvest. Yield maps can be generated immediately following data collection to show yield patterns within fields [3, 13]. The result of analysis can be used for after-season management. The yield monitor accuracy depending on the type and brand of yield monitor, calibration regime, flow rate and conditions at harvest. Yield monitor calibration plays a key role in obtaining the best possible accuracy from the yield monitor [3].

Despite the commercial availability of yield monitors, many crop harvesters are not equipped with them. Furthermore, additional information is needed for effective analysis of the map e.g. information on plant stress during the growing season. Yield monitor data cannot be used to detect problems and generate maps within season.

Within season estimates of relative yield variation and stress detection can more useful address by remote sensing. Satellite imagery obtained during the growing season can be used to generate yield maps for both within season and after season management (Figure 1.).

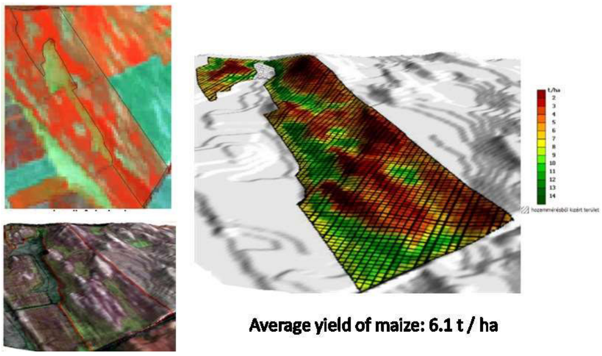

Figure1. Cell yield of maize created from Sentinel data: input data and output). Source: [13]

III. VEGETATION INDECISES

By measuring the reflectance of the plants at various wavelengths, it is possible to collect a lot of information about the status of the plants. The reflectance of light spectra from plants depends on plant type, water content within tissues, and other intrinsic factors [4]. The reflectance of vegetation is low in the blue and red regions of the visible spectrum, due to absorption by chlorophyll for photosynthesis. It has a peak at the green region which gives rise to the green colour of vegetation. In the near infrared (NIR) region, the reflectance is much higher than that in the visible band due to the cellular structure in the leaves. In the mid infrared there are more water absorption regions (Figure 2.) [14].

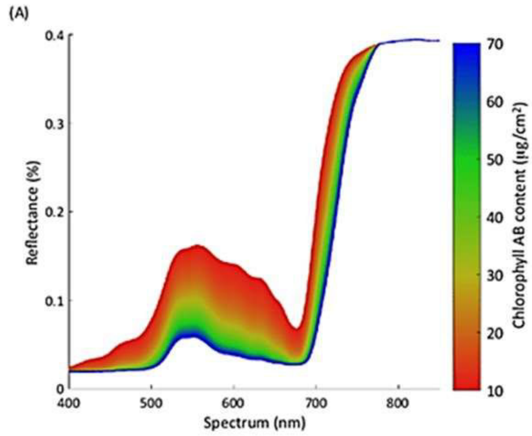

Figure2. Reflectance pattern of a typical crop as a function of the chlorophyll AB content Source: [16]

During the quantitative interpretation of remote sensing information from vegetation can be created by extracting vegetation information using individual light spectra bands or a group of single bands for data analysis. The construction of VI algorithms are effective tools to measure vegetation status. Vegetation information from remote sensed images is mainly interpreted by differences and changes of the green leaves from plants and canopy spectral characteristics. The data from near infrared (0.7–1.1 m) and red (0.6–0.7 m) or other bands are combined in different ways according to their specific objectives [4]. From previous research, there are known relationships between the indices using those two regions of the spectrum and the amount of vegetation. Different indices have been developed to better model the actual amount of vegetation on the ground. A lot of research has been done to derive relationships between a vegetation index of a crop, measured at a particular time, and the final crop yield. Low index values usually indicate little healthy vegetation while high values indicate much healthy vegetation [11, 12].

IV. DATA ANALYSIS BY UNSUPERVISED CLASSIFICATION

There are various approaches and quantitative methods for using remote sensing data to discriminate different types of habitat cover. In this study pixel-based and object-based approaches were applied for mapping spatial variability within a field.

In an unsupervised classification, the analyst does not predefine the land cover or habitat types. The image processing software divides the image into a certain number of classes, based entirely on the spectral data and with no knowledge of what cover types are present in the image. The user can define limits to the number of output classes and spectral variance within each class. The resulting classes are identified by different numbers, and the analyst must then assign names to these classes with the support of field knowledge and an understanding of how different habitats should appear in these images. There are several mathematical strategies to represent the clusters of data in spectral space. For example: IsoData

Clustering (Iterative Self Organising Data Analysis Techniques). It repeatedly performs an entire classification and recalculates the statistics. The procedure begins with a set of arbitrarily defined cluster means, usually located evenly through the spectral space. After each iteration new means are calculated and the process is repeated until there is some difference between iterations. This method produces good result for the data that are not normally distributed and is also not biased by any section of the image. The other one is Sequential Clustering. In this method the pixels are analysed one at a time pixel by pixel and line by line. The spectral distance between each analysed pixel and previously defined cluster means are calculated. If the distance is greater than some threshold value, the pixel begins a new cluster otherwise it contributes to the nearest existing clusters in which case cluster mean is recalculated. Clusters are merged if too many of them are formed by adjusting the threshold value of the cluster means.

In this study to delineate classes the IDRISI software was used. So called ‘CLUSTER’ algorithm was used for classification. It provides an unsupervised classification of input images using a histogram peak technique. CLUSTER uses a histogram peak technique of cluster analysis. This is equivalent to looking for the peaks in a one-dimensional histogram, where a peak is defined as a value with a greater frequency than its neighbors on either side. Once the peaks have been identified, all possible values are assigned to the nearest peak and the divisions between classes fall at the midpoints between peaks. Here a one to seven-dimensional histogram is used to find the peaks. A peak is thus a class where the frequency is higher than all of its cardinal neighbors. The diagonal neighbors are omitted because of the correlation between bands.

CLUSTER does not use the histogram from the original images to define peaks. Instead, it stretches each input image into a given number of grey levels before calculating their histograms. There are three stages involved in this process:

1. The histogram of each input image is obtained by going through the original input image’s digital numbers. From this histogram, the numeric cumulative proportion at each grey level is found by adding up the frequency of each grey level from low to high frequencies. The cutoff points are then determined according to the saturation percentage specified. For example, to saturate an image by 5% in each tail, the 5% and 95% cumulative proportion cutoff points are determined by checking the accumulative proportion from the histogram. To derive the histogram for each input image, the gray levels of the input images are linearly stretched into 256 levels, ranging from 0 – 255.
2. Each image is linearly stretched into the given grey levels by applying the cut off points. In the stretch process, a pixel’s new value is evaluated using the following procedure:

Assuming $DN_{original}$ is the original value in the input image, the two cutoff points obtained in the previous stage are $Cutoff_{left}$ and $Cutoff_{right}$, the number of grey levels specified is GreyLevel, the new value of each pixel Val_{new} is evaluated using the following logic:

If $DN_{original} <= Cutoff_{left}$ then $Val_{new} = 0$;
Else
If $DN_{original} >= Cutoff_{right}$ then $Val_{new} = GreyLevel-1$;
Else
$Val_{new}= (DN_{original}-Cutoff_{left})/(Cutoff_{right}-Cutoff_{left})*(GreyLevel-1)$

3. Finally, histograms are developed from each newly stretched image. The histograms derived from the newly stretched images are used to find peaks to perform the cluster analysis [15].

The Figure 3 presents the work flow of mapping spatial variability by unsupervised classification methods and the results of classification.

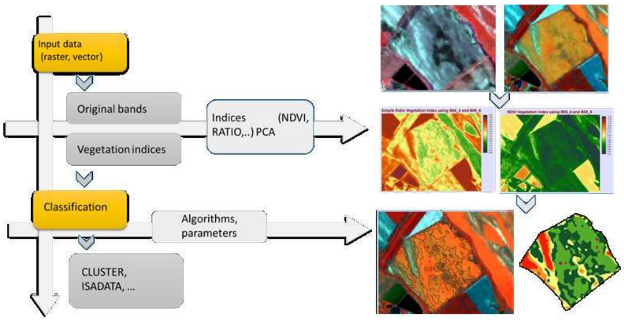

Figure3. Work flow of mapping spatial variability by unsupervised classification methods.

V. Conclusion

Remote sensing data can be useful as data sources for estimating and mapping within field crop yield variability for precision agriculture. Imagery taken during the growing season can be used to monitor crop growing conditions and identify potential problems that could be addressed within the growing season. The multitemporal imagery can also be used to generate yield maps to record the spatial variation in yield. Vegetation index is a useful tool in crop monitoring and crop estimation: identification of problems (lack of sowing, cultivation problems, pests, ..), mapping stressed plants, mapping variable requirements for irrigation within a field, identifying fertilization and pesticide requirements, and mapping potential management zones within fields [6, 7, 11]. However, the analysis should take into account that the reliability of such surveys depend on more factors like as the type of index, the time of data gathering and the stage of plant. In this article, one of the way of performing clustering was discussed. The results of the study show that method of unsupervised classification is suitable for mapping differences within a field. This type of research cannot be considered complete as more and more satellite and airborne imagery are becoming available, more research is needed to develop right data analysis techniques for yield estimation and other precision agriculture applications.

References

[1] Chenghai Yang, James H. Everitt, Qian Du, Bin Luo, Jocelyn Chanussot: Using High-Resolution Airborne and Satellite Imagery to Assess Crop Growth and Yield Variability for Precision Agriculture. Publisher: IEEE, https://ieeexplore.ieee.org/document/6236221

[2] David J. Mulla, Xuxin Miao: Precision farming, https://www.researchgate.net/publication/286381915_Precision_Farming

[3] Colvin T.S., Arslan S.: Yield monitoring accuracy, http://www.ipni.net/publication/ssmg.nsf/0/00BE56FD07426288852579E5007681D4/$FILE/SSMG-09.pdf

[4] Jinru Xue and Baofeng Su, 2017: Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications, Journal of Sensors Volume 2017, Article ID 1353691, 17 pages, https://doi.org/10.1155/2017/1353691

[5] Marshall Burkea, David B. Lobella, 2017: Sustainability Science, Agricultural Sciences Satellite-based assessment of yield variation and its determinants in smallholder African systems. DOI: 10.1073/pnas.1616919114, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5338538/#s3

[6] Milics G., Smuk N.,Virág I., Néményi M.: Precision Agriculture Technical Development for Sustainable Agriculture. Nemzti Tankönyvkiadó Budapest. pp. 231-240. ISBN 978-963-19-7352-5

[7] Németh T. et al, 2007, A precíziós mezőgazdaság módszertana, edited by JATEPress-MTA TAKI, Szeged

[8] Neményi M.,.Mesterházi P.Á, Pecze Zs.,.Stépán Zs: The role of GIS and GPS in precision farming, https://doi.org/10.1016/S0168-1699(03)00010-3

[9] Pierce, F.J., Novak, P., 1999, Aspects of Precision Agriculture. Advances in Agronomy. 67, 1-85

[10] Schimmelpfennig, David, and Robert Ebel. On the Doorstep of the Informa-tion Age: Recent Adoption of Precision Agriculture, EIB-80, U.S. Dept. of Agriculture, Economic Research

[11] Verőné Wojtaszek M.,Körmendy E., 2015 Távérzékelés szerepe a precíziós mezőgazdaságban, különös tekintettel az UAV felvételekre. Remote Sensing. 5(2), pp.426-430

[12] Verőné Wojtaszek M. 2018, Remote sensing based detection of spatial variability within a field. Journal of Geographical Society of Uzbekistan. Special volume. Tashkent pp.169-175.

[13] Csornai G., 2019: Precíziós és tábla-hozammérés muhold-felvételekkel és ezek alkalmazásai szántóterületeken. http://www.gisopen.hu/eloadasok/2019/sz21.pdf

[14] https://www.researchgate.net/figure/Assessment-of-Nutrient-Status-A-Reflectance-pattern-of-a-typical-crop-as-a-function-of_fig2_329679450

[15] https://clarklabs.org/wp-content/uploads/2016/10/IDRISI-Selva-Spanish-Manual.pdf