

Characterization of multispectral aerial images of sugarcane

A Solano¹, G Schneider¹, A Kemerer² A Hadad¹

¹Facultad de Ingeniería, Universidad Nacional de Entre Ríos, Argentina

Grupo de Investigación en Inteligencia Artificial, Oro Verde, Entre Ríos, Argentina

²Grupo de Recursos Naturales y Factores Abióticos de la EEA Paraná del INTA

Cátedras de Climatología Agrícola y SIG de la FCA UNER Oro Verde, Entre Ríos, Argentina

E-mail: aguchosolano@gmail.com, gerschneider@gmail.com,
akemerer@parana.inta.gov.ar, hadad@santafe-conicet.gov.ar

Abstract. In this paper we present a implementation and characterization the status of sugarcane plantations based on the analysis of multispectral aerial images. Currently there are no precise techniques to estimate objectively the cane area fall or overturned, and this causes significant losses in crop productivity and industrialization. For the realization of this work was made a dataset benchmark images, and implemented a software from which were obtained indicators related to agronomic phenomenon, and analyzes of the data generated. In addition was used Principal Component Analysis to visualize and integrate statistical texture features. The results indicate the statistical features used characterize partially the sugar cane phenomena and suggest include another texture focus to complement this feature set, previous to built an cane identification process.

1. Introduction

The cultivation of sugar cane in Argentina is found principally localized in the northwest region of the country (99 %), where represents, in agreement with [1], one of the productive more important activities, and in scarce extension in the littoral region (1 %). The productive systems present different technological levels according to the system of crop used, with systems from manuals to totally mechanized, although, at present more than 70% of the total volume it harvests of way mechanized. One of the more critical aspects of the process of crop of the sugar cane are the losses that it is produced and have decided as passable levels about 2,5%. The presence of fallen cane is one of the factors with bigger influence. Incidence of superior overturning to 20% decides increments in the losses that go from the 4 to 6,4%, according to express [2].

The presence of fallen sugar cane (CC) immediately of the crop brings at important losses in the collection and in the sugar mill. In the phenomenon of CC for bankrupt person take part in random aspects and unpredictable. It is normal that the sugar cane plantation in a productive square reach a height mediate, but presents variations of this variable in different sectors of the field. This must be because of the differential conditions of the earth that presents different levels of fertility, humidity, compressing, etcetera. For other side, each plant of the sugar cane plantation is constituted in a different way, for which will manifest in unique form the different external stimulus.

When the plant is young and has developed under the best conditions, reaches certain height and for its own weight it begins to warp, moving close to break once its breakage tension is surpassed product of the wind or environmental phenomena as the hail. The CC appears with a great space heterogeneity.



Once produces the phenomenon, the earth can short remain, or on the contrary can be produced a product with new green color of the appearance of new green sheets, with phototropic growth. This supreme heterogeneity in the presence of an external observation. The above-mentioned factors make evident the great complexity in the phenomenon in studies, which does it difficult of analyzing and quantize.

The strange matter, called trash, when enters to the sugar mill represents an important loss of efficiency in the process of manufacture of the sugar, assure [3]. The challenge of industrial engineering takes root in exploring the alternatives to control the harmful effects that these tailings or trash and the present dust in the cane have in the equipments, processes and operative performances of the factory. That is to say that the quality of the cane affects in direct form the industrial performance and the quality of the obtained sugar [4]. In turn the quality of the raw material (cane) can be affected for changing aspects of the agriculture of the sugar cane, such as the introduction of new cultivars, climatic variations, the use of ripening chemists, changes in the cultural practices and in the systems of crop or the appearance in the cultivations of illnesses or infections. As for the types of crop, it observes that the contents of trash and dust in cane harvested in form mechanized are appreciably lower that in the semi-mechanized, being more than a 2 to a 5,7 %. At the same time, the values of fiber (vegetable remainder) goes up from 14 to 18 % approximately, due to the presence of major quantity of vegetable parts for a less efficient crop [5]. This means that the industries become calm to the producers a significant part of the weight of the trash of the entered raw material for scales to cane value, as long as the form of muestreo and analysis of this parameter is generally inadequate.

To this you must incorporate the other additional costs that produces this strange product in the factory properly happiness, in any way whatever:

- 1 Wear in the mills for abrasion and a progressive loss in the capacitance of extraction of the cane juice.
- 2 Increase significant of the amount of sloth and with it of the losses of sugar in the same thing.
- 3 Increase of the consumption of chemicals for the treatment of the juices.
- 4 Increase of the color in the juice and subsequent in the sugar with it who to avoid it reduce the cycles of crystallization by decreasing the industrial efficiency.
- 5 Wear of equipments, pipeses and dazed.
- 6 Diminution of the can calorific of the waste pulp and consequently the efficiency in the burnt forest of the same, with an increase of the consumption of and of the emissions of boiler.
- 7 Other.

To decrease to the minimum possible the percentage of trash in the cane, several mills installed washing tables of cane, where they remove the content of dust by means of big quantities of water (approximate report of 3 m³ of water by ton. of cane). This supposes saving of money, but adds problems like bigger costs for the required power, the necessary water and the additional manual labor, join to the new discussions with the producers since some studies confirm that in the wash of the cane loses a percentage of the performance for extracting sugar. Moreover, the technique requires availability for service of terrain to install settling tanks of sand and blunders to re-use the water and does not return the so to cause them by increasing the contamination. A promising strategy to achieve the reduction of this impact take root in achieving a bigger operative efficiency in the tasks of crop of the cane in the field, previous stage to the entrance to the factories, and fundamentally in sectors with broken cane that increase notably the percentage of trash. This is an area of intensive manual labor where for its particular properties, considers [5], have not achieved to introduce technological solutions that reduce its negative impact in the production.

At present, the quantization of the fall square productive sugar cane is an expensive process and inefficient. It requires the sending of evaluators to the terrain and they must explore big surfaces to obtain a representative sampling that permit an estimation. The displacement around the field produce damage, require many man hours of work and generate estimations with a high degree of uncertainty. In turn, the presence of space variations in the productivity of the cane is very wide [6] decided variations of until 50 tn/ha to level of lot of production. It is spectable that this owner, as well as variations in the quality of the cane observe to diverse scales, in a similar way to which has demonstrated for other variables they fall in the production, as the properties of earth [7], which raises objections establish a distance of unique sampling and reverberates negatively in the costs.

Other strategy is carrying out flights over the cultivations with observant specialists who estimates the percentage of fallen cane according to their experience of it observed from the air. This strategy is subjective and dependent of the available specialists in each area of interest.

By keeping in mind these antecedents it is made evident that has not observed report of some methodology with certainly measurable certainty to decide the affected area, in order to help to plan of better way the process of crop and the associates costs of the productive involucrate actors.

In this sense, the national institute of agricultural technology of Argentina, INTA, has approached the problematic, by carrying out works as of aerial photography of high resolution [8,9]. This data source works as a whole to advance with new technological strategies to approach the problem. In this work it presents the information developed substructure and the analysis of information and agronomic results.

From the information viewpoint mounted an application to carry out the labelling of images, as of which develops a pipeline type structure with capacitances of analysis of variables and classification of regions.

2. Methodology

It counted on a set of images obtained for the national institute of agricultural technology (INTA) of Argentina from a Sky Arrow 650 airplane TCNS ERA, whose system of capture is composed of a Global Positioning System, a multi-spectral chamber Geospatial MS4100 and a system of control and storage of data. The chamber employed permits the acquisition of photograms in three bands of the electromagnetic specter: green(530-580 nms.), red (650-685 nms.) and near infrared (770-830 nms.) with a resolution of image of 1920 x 1075 pixels. The GPS provides the position, attitude and height quickly in synchronism with the acquisition of each photogram.

The flight carried out on 5th May 2008, moment of the cycle of the cultivation where the presence of fall of the cane is clearly evident. The flights were carried out in the solar midday when all is said and done with a homogeneous illumination in the surface. The fly it is designed for an altitude of 1200 m. by resulting a size of pixel of 0.7 ms. There were obtained 540 photograms that were assembled by means of the reconnaissance anchoring knitted in consecutive photograms generating a mosaic.

With the supervision of an agricultural engineer it was carried out the labelling of the images in the classes of interest, using the application Label Me [10] of the Massachusetts Institute of Technology

(M.I.T.), who was in a local server to facilitate the access to the data and the incorporation to a pipeline of processing. The figure 1 shows the process of labelling who consisted in delimiting the images with polygonal continents of representative portions of the cane classes: standing cane (CP), fallen in form of patch (CCP), fallen cane in big areas (CCA), fallen cane with phototropism (CCF) and in excess or areas of exposition earth (C).

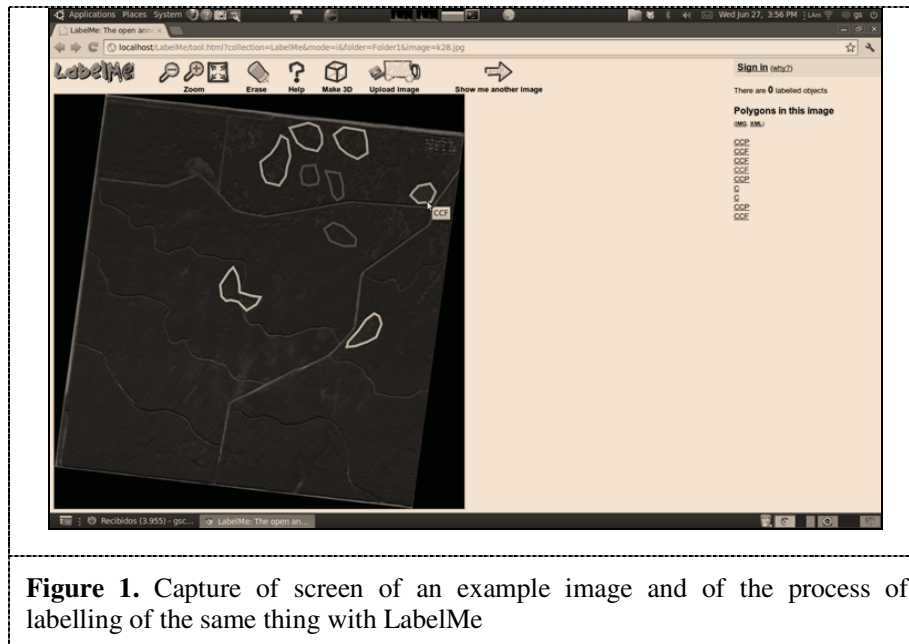


Figure 1. Capture of screen of an example image and of the process of labelling of the same thing with LabelMe

It was developed in Java a software capable of acquire the information with form of metadata generated in the labelling, and of cut and process automatically the portions of obtained images from the dataset, in order to obtain representative examples of each class. It incorporates in this stage a pre-processing of normalization in brightness of the images that is carried out by calculating the brightness mediantes by canal of all polygonal labellings, and by fixing arbitrarily as parameter for the normalization a 120 % of this value, which enlarged the dynamic range without producing saturation in the histogram.

Given the characteristics of the agronomic phenomenon, carried out a boarding from the viewpoint of the textures of the images, by defining the following characteristics or indicators of interest:

- Standard deviation of the histogram of the sample by canal
- Entropy mediantes of the histogram of the sample by canal
- Value mediantes of the pixels of the sample by canal (mentioned as spectral signature)
- Index of vegetation of difference normalized (NDVI) [11]

The extraction of characteristics was carried out in an iterative process of the Batch type. The volume of information generated was stored in a Postgre database, of mode of can access concurrent immediately of the analysis to the whole information and it obtains the by using the potentialities of a language of SQL consultation.

By attending to the problem of the complexity of the data, minimized negative effects of the desbalance [12] by means of the generation of sets with similar number of samples for class, and obtained random of the dataset as of the SQL query:

```

DROP VIEW rocked;
CREATE VIEW balanced AS
( SELECT e.id FROM tablaAnálisisEspectro WHERE e.tamVentana=tamVent AND e
e.clase= "CP" ORDER BY RANDOM () LIMIT CantMuestras ) UNION
SELECT e.id FROM tablaAnálisisEspectro WHERE e.tamVentana=tamVent AND e
e.clase= "CCA" ORDER BY RANDOM () LIMIT CantMuestras ) UNION
( SELECT e.id FROM tablaAnálisisEspectro WHERE e.tamVentana=tamVent AND e
e.clase= "CCP" ORDER BY RANDOM () LIMIT CantMuestras ) UNION
( SELECT e.id FROM tablaAnálisisEspectro WHERE e.tamVentana=tamVent AND e
e.clase= "CCF" ORDER BY RANDOM () LIMIT CantMuestras ) UNION
( SELECT e.id FROM tablaAnálisisEspectro WHERE e.tamVentana=tamVent AND e
e.clase= ORDER BY RANDOM "c" () LIMIT CantMuestras );

```

Figure 2. SQL balanced query

To lead with dimensional problem was used Principal Component Analysis (PCA). PCA is a statistical technique whose purpose is to condense the information from a large set of correlated variables in fewer variables, without losing most of the variability present in the data set [13].

This method consists of transforming the original attributes or variables x_1, x_2, \dots, x_m of the examples into another set of attributes f_1, f_2, \dots, f_p , where $p < m$. This process can be geometrically seen as a change of shafts in the representation (projection).

The new attributes are derived as a linear combination of the variables in the data set, with "weights" chosen in such a way that the major components are not mutually correlated. Besides, the first f_i are those containing most of the variability present in the original data.

Mathematically, if we assume $m = p$, the transformation process can be posed as follows.

$$\begin{bmatrix} \overline{f_{(mx1)}} \\ \overline{a_1} \\ \overline{a_2} \\ \vdots \\ \overline{a_m} \end{bmatrix}^T \cdot \overline{x_{(mx1)}} = A^T_{(mxm)} \cdot \overline{x_{(mx1)}}$$

To find the vectors of matrix A, so that multiplying A by each example vector results in a new vector in the new dimensions space, variance is assumed to be higher for the first attributes than for the last ones. It can be shown that the result is obtained by first calculating the following estimate of the S covariance matrix:

$$S_{(mxm)} = \frac{1}{n-1} \sum_{i=1}^n (u_i)_{(mx1)} (u_i)^T_{(1xm)}$$

Where n is the number of examples in the data set, m is the number of attributes and u_i difference vectors are calculated by subtracting each instance x_i with the average attributes of all examples.

Once S covariance matrix is built, vectors of matrix A are obtained by calculating the eigenvectors, ordered from highest to lowest according to the extent of the eigenvalues associated with each eigenvector.

3. Results

The indicators of interest evaluated resulted, in general, promissory for the process of characterization of the fallen cane, to exception of the NDVI. This index presented low values in the case of the fallen cane where the cultivation reaches more biomass and the value must be bigger [14-16]. This happens because the band IR wins the variations in structure of the cultivation become a partner of the biomass and the index of area to folio. When the cultivation falls the structure seems modified, decreases, that does go down the IR and then the NDVI takes more low values. Also there is a confused effect of the contribution of reflectance of the earth in the fallen cane that decreases the NDVI [17,18]. Therefore, this index was not incorporate in the later analysis.

In the case of the values observed of deviation of the histogram it is observed for the cane class not fall (CP) a running towards valuing inferiors in the three canals analyzed (figures 3). It considers oneself to the standard deviation of the histogram as an indicating candidate of the cane class standing.

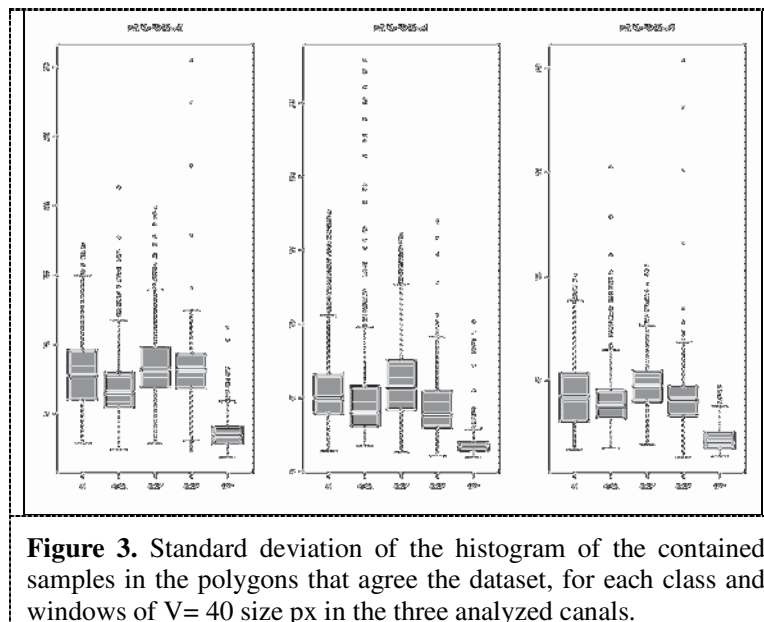


Figure 3. Standard deviation of the histogram of the contained samples in the polygons that agree the dataset, for each class and windows of $V=40$ size px in the three analyzed canals.

The distribution of the entropy presented more high values for the classes of fallen cane (CCA, CCF and CCP) and exposition earth concern to the cane class standing, with valuing different media between classes. Therefore, it considers the entropy of the histogram a promissory indicator for the identification of cane standing. As for the spectral signature, although cavity is observed in the distributions between classes, for the IR canal a displacement is observed towards valuing superiors for cane standing. The spectral signature for the IR canal appears as promissory indicator to identify cane standing, for which will be considerate as one of the variables to incorporate.

Given that the proposed variables have similar behavior and overlap between classes, there was joined by PCA the indicators and reduced the dimensionality of the feature set in order to get a lesser number of possible descriptors.

By applying PCA was obtained that 96.6% of the variability in the data was explained by the first 3 principal components. However it was observed that in the space generated by the main components only three classes are distinguished: CP, C and CC. The latter groups the types of cane CCA, CCP and CCF. Said distribution is shown in Figure 4.

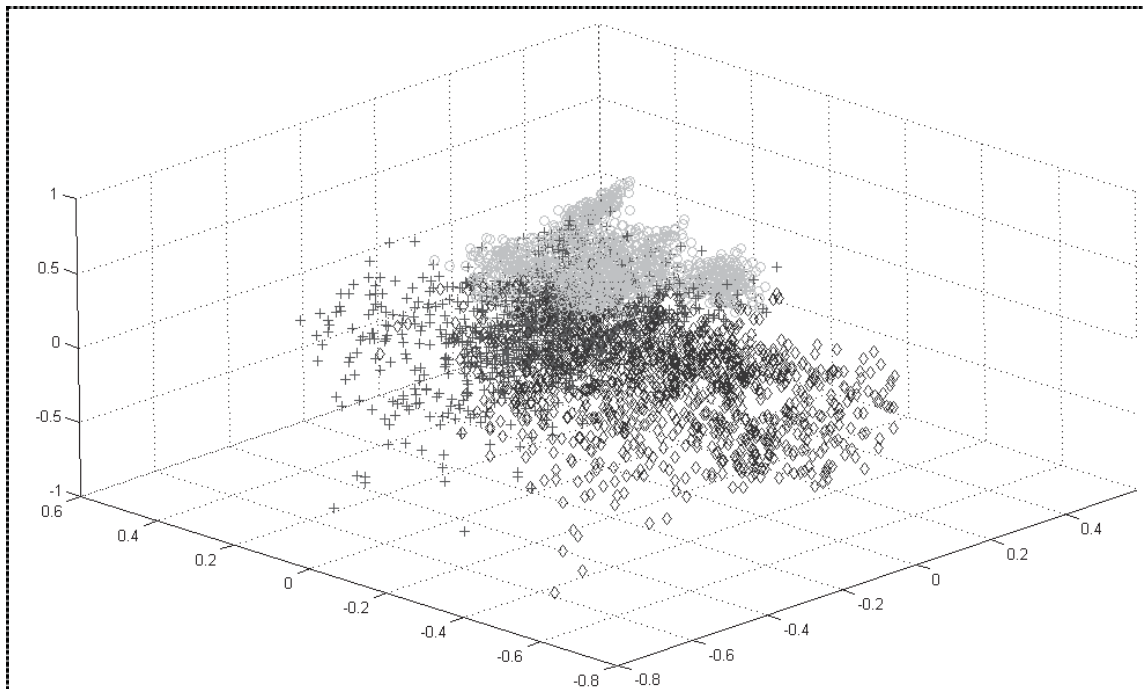


Figure 4. Distribution of the types of Cane in the Principal Components space.

(+)CC; (o)CP ; (diamond)C

4. Conclusions

It was compiled a reference dataset and a database for the study of the cultivation of sugar cane in its different states, from which the software created could provide information on the usefulness of the proposed indicators and to estimate the surface of sugarcane fall.

As to characterization of the cane types there is a need to combine several statistics associated with the texture of the image to address the problem of identification of sugar cane. In turn this combination by PCA can distinguish between walking cane and rod drop, resulting insufficient to differentiate most of the reeds falls. The only CC that can be identified is referred to as C (run) which corresponds to the zone of lower fertility. While it is a partial result has its value from the agricultural point of view have identified these areas of lower performance.

Furthermore it is noted that the approach of textures from the statistical viewpoint is insufficient for complete detection of canes falls, so it must include spectral or multiscale approach to the problem. As future work is planned to incorporate new descriptors in order to build a classifier oriented segmentation of such imagery.

5. References

- [1] Pérez D, Fandos C, Mazzone L, Soria F, Scandalaris P and Scandalaris J. 2005 *Caña de azúcar en Tucumán y Argentina: evolución de algunos aspectos económicos y productivos en la campaña 2004* [on line]. Reporte agroindustrial. Estadísticas y márgenes de cultivos tucumanos. Vol. 2, Boletín N° 6, (ene.). Tucumán (Argentina): Estación Experimental Agroindustrial Obispo Colombres, EEAOC. p 1-6. <<http://www.eeaoc.org.ar/upload/publicaciones/archivos/189/20120316170830000000.pdf>> [consulted: 02/04/2013]

- [2] Rodríguez R.A, Sopena, R.A, Saleme P.M and Vicini L.E 2010 *Pérdidas durante la Cosecha del Cultivo de Caña de Azúcar: Evaluaciones 2009 en la Provincia de Tucumán – Argentina* Informes Técnicos del proyecto Precop, No. 1. Famaillá (Tucumán, Argentina): INTA EEA Famaillá, N°1, 18 p. ISSN: 1852-9399.
- [3] Tonatto J, Romero E.R., Leggio Neme M.F., Scandalaris J, Alonso J, Digonzelli P, Alonso L and Casen S 2005 *Importancia de la calidad de la materia prima en la productividad de la agroindustria azucarera* Gacetilla Agroindustrial de la EEAOC, No. 67, Tucumán (Argentina): EEAOC. 13 p.
- [4] Larrahondo J.E. 1995 *Calidad en la Caña de Azúcar* CENICAÑA. El cultivo de la caña en la zona azucarera de Colombia, Cali, CENICAÑA. p.349-350.
- [5] Sustaita G 2005 *Modelo estratégico para la industria azucarera regional*. Tesis Final MBA, San Miguel de Tucumán (Tucumán, Argentina): Universidad Católica del Norte Santo Tomás de Aquino, Fundación del Tucumán y Pontificia Universidad Católica de Valparaíso.
- [6] Vicini L 2007 *Determinación de la variación de rendimiento cultural de caña de azúcar en Tucumán, Argentina, posicionada con GPS* Actualización Técnica N° 8 (jun.). Manfredi (Córdoba, Argentina): Proyecto Nacional Agricultura de Precisión, INTA, 18 p.
- [7] Anderson D.L., Portier K.M., Obreza T.A., Collins M.E. and Pitts D.J. 1999 *Tree regression analysis to determine effects of soil variability on sugarcane yields* Soil Science Society of America Journal, Vol. 63, No. 3, Madison (WI, USA), Soil Science Society of America, p. 592–600, ISSN: 0361-5995.
- [8] Kemerer A., Melchiori A, Albarenque S and Melchiori R 2010 Utilización de fotografías aéreas multiespectrales para caracterizar la variabilidad espacial en la producción de caña de azúcar 9º Curso de Agricultura de Precisión y 4ª Expo de Máquinas Precisas (14-16/07/2010), Manfredi (Córdoba, Argentina): INTA - EEA Manfredi. Material del 9º Curso de Agricultura de Precisión, p. 177-182, 2010
- [9] Melchiori A, Kemerer A, Melchiori R and Bellomo M 2009 *Utilización de fotografías aéreas multiespectrales de alta resolución para la cuantificación de caña de azúcar caída* I Congreso Argentino de Agroinformática (24-25/08/2009), Mar del Plata (Argentina): Sociedad Argentina de Informática y el Instituto Nacional de Tecnología Agropecuaria, INTA.
- [10] Russel B.C., Torralba A, Murphy K.P. and Freeman W.T. 2008 *LabelMe: a database and web-based tool for image annotation* International Journal of Computer Vision, Vol. 77, No. 1-3 (may.). New York (USA): Springer US. p. 157-173, ISSN: 0920-5691.
- [11] Rouse J.W., Haas R.H., Schell J.A. and Deering D.W. 1973 *Monitoring vegetation systems in the great plains with ERTS* Third ERTS Symposium (10-14/12/1973), Washington D.C. (USA): NASA. Proceedings of Third ERTS Symposium Vol. 1., p. 309-317.
- [12] Toribio P, Rodriguez B.G. and Alejo R 2009 *Complejidad de los datos en las Redes Neuronales Artificiales: Estado de la cuestión* 7º Congreso Internacional de Cómputo en Optimización y Software, CICOS 2009 (17-20/11/2009), México (México): UAEM. Marco Antonio Cruz-Chávez, M.A. & Zavala-Díaz, J.C. (eds.). Memorias del 7mo Congreso de Cómputo CICOS 2009. ISBN(e) 978-607-00-1970-8, México (México): UAEM. p. 229-235. <<http://campusv.uaem.mx/cicos/imagenes/memorias/7mocicos2009/Articulos/24%20%20Complejidad%20de%20los%20Datos%20en%20Redes.pdf>> [consulted: 02/04/2013]

- [13] Castells F, Laguna P, Sornmo L, Bollmann A and Roig JM 2007 *Principal Component Analysis in ECG Signal Processing* EURASIP Journal on Advances in Signal Processing. Volume 2007, Article ID 74580, 21 pages. doi:10.1155/2007/74580
- [14] Flowers M.W.R., Heiniger R, Tarleton B and Meijer A 2003 *Field Validation of a Remote Sensing Technique for Early Nitrogen Application Decisions in Wheat* Agronomy Journal, Vol. 95, No. 1. Madison (WI, USA): American Society of Agronomy, p 167–176. ISSN: 0002-1962.
- [15] Raun W.R., Solie J.B., Stone M.L., Martin K.L., Freeman K.W. Mullen R.W., Zhang H., Chepers J.S. and Johnson G.V. 2005 *Optical Sensor-Based Algorithm for Crop Nitrogen Fertilization* [online] Communications in Soil Science and Plant Analysis, Vol. 36, No. 19-20. Johannesburg (South Africa): Taylor & Francis, Inc. p. 2759-2781. ISSN: 0010-3624, EISSN: 1532-2416.
- [16] Melchiori R.J.M., Caviglia O.P., Bianchini A.A., Faccendini N, Albarenque S and Raun W 2006 *Wheat Yield Prediction by Using an Active Sensor in the Northern Argentinean Pampas* ASA-CSSA-SSSA Annual meeting (12-16/11/2006), Indianapolis (Indiana, USA): American Society of Agronomy, ASA - Crop Science Society of America, CSSA - Soil Science Society of America, SSSA. Proceedings of ASA-CSSA-SSSA Annual Meeting 2006. Madison (WI, USA): ASA-CSSA-SSSA.
- [17] Qi J, Chehbouni A, Huete A, Ferry Y and Sorooshian S 1994 *A modified soil adjusted vegetation index* Remote Sensing of Environment, Vol. 48, No. 2 (may.). Philadelphia (PA, USA): Elsevier Inc. p. 119-126, ISSN: 0034-4257. <<http://www.sciencedirect.com/science/article/pii/0034425794901341>> [consult: 01/04/2013]
- [18] Rondeaux G, Steven M. and Baret F 1996 *Optimization of soil-adjusted vegetation indices* Remote Sensing of Environment, Vol. 55, No. 2 (feb.). Philadelphia (PA, USA): Elsevier Inc. p. 95-107. ISSN: 0034-4257.