Iranian Journal of Astronomy and Astrophysics
Vol. 2, No. 1, 2015
© Available online at http:// journals.du.ac.ir

Iranian Journal of Astronomy and Astrophysics

Segmentation of Photospheric Solar Images by Using c-Means, k-Means, and FCM Algorithms

Mahdi Yousefzadeh[1] · Mohsen Javaherian[2] · Hossein Safari[2]

[1] Department of Physics, Institute for Advanced Studies in Basic Sciences (IASBS), Zanjan, P.O.Box 45195-1159, Iran

[2] Department of Physics, University of Zanjan, Zanjan, P.O.Box 45195313, Iran

Abstract. In this study, we use three kinds of clustering methods based on c-means, k-means, and fuzzy c-means (FCM) algorithms to segment solar ultra-violet (UV) images. The methods are applied on a sequence of quiet-Sun photospheric observations at 525 nm images taken by *Sunrise* on 9 June 2009. The comparison between these three algorithms represents a little bit differences in extraction of physical parameters (filling factors, brightness fluctuations, size distribution, etc.) from images. On the basis of FCM algorithm, the mean value of granule sizes is found to be about 1.8 arcsec^2 (0.85 Mm^2). Granules with sizes smaller than 2.8 arcsec^2 cover a wide range of brightness, while larger granules approaches a particular value. Granules may have lifetimes less than 10 minutes in this part of the Sun. Investigation of local fractal dimension of photospheric images shows that granulation pattern are approximately scale free in some resolutions.

Keywords: Sun: UV radiation, Sun: granulation, Techniques: image processing, Techniques: segmentation, Techniques: clustering

1 Introduction

Lots of events are stochastically appeared on the solar surface such as granules, bright points (BPs), etc. [1, 2] at times of both minimum and maximum solar activity [3]. Extracting physical properties of features happening on photosphere is important subjects of solar field. One of the features that has main role in solar activity is granulation pattern. It has been accepted the solar granulation is resulting from a convective turbulent process [4].

To optimize statistical analysis of granulation patterns, it is necessary to develop automatic detection techniques. The process of dividing data set into classes or clusters in the feature space is data clustering so that features (elements) in the same class have similarities but there is low resemblance between elements which are in various classes. One of the significant methods for granules segmentation and finding non-granular regions is using various kinds of clustering such as c-means and k-means. These algorithms are multichannel unsupervised and able to automatically segment regions fast [5]. Different kinds of clustering (e.g., spatial possibilistic clustering algorithm and fuzzy c-means clustering) have been applied on EUV solar images to segregate traditional regions into coronal hole (CH), quiet Sun (QS), and active regions (AR). The estimations of these areas characterizations are consistent with previous results [7]. We apply c-means, k-means, and fuzzy c-means (FCM) algorithms on the *Sunrise* photospheric UV images recorded on 9 June 2009 and break down this data set into regions that can be recognized as granule or non-granular region. The code is able to segment N-dimensional gray-scale images into c classes (more than two classes) by implementation of the c-means (k-means) clustering algorithm. The comparison between these two algorithms represents approximately the same results.

The paper is organized as follows: Data analysis is discussed in Section 2. Identification methods for granules and non-granular regions by using c-means, k-means, and FCM are explained in Section 3. The results are given in Section 4. The conclusions are discussed in Section 5.

2 Data Analysis

The *Sunrise* balloon-borne solar observatory was launched on 8 June 2009 [8, 9]. The high-resolution images in the UV and in the quiet Sun were begun to be recorded on 9 June 2009 [9]. *Sunrise* has a resolution of about 100 km (from a 1 m Gregory reflector telescope). This observatory is equipped with *Sunrise* filter imager [10], imaging magnetograph experiment [11], image stabilization and light distribution unit, and a correlating wave-front sensor [12]. The Imaging Magnetograph eXperiment (IMaX) data produces 936×936 pixel with resolution of 0.055 arcsec per pixel. IMaX uses a Zeeman triplet in the FeI 525.02 nm line. The *Sunrise* together with its instruments provides 4 levels of data. Level-0 is raw data. Level-1 data are fully reduced by phase-diversity reconstruction. Individual and averaged wave-front corrections are exerted on both level-2 and level-3 data, respectively. We used a sequence of IMaX (level-2) data (Figure 1A) recorded on 9 June 2009 (14:10-14:45 UT) at 525 nm to segregate granules and non-granular regions by three mentioned algorithms.

Using a subsonic filter, the global solar surface oscillations (e.g. 5-minute p-modes) are removed from data by modifying horizontal speeds above 5 km/s in momentum and frequency space.

3 Methods

3.1 *c*-Means Clustering

In c-mean clustering, called hard clustering, dataset is decomposed into certain classes where each feature belongs to exactly one cluster. For image clustering, pixels' brightness constructs our feature space [13]. Depending on the number of clusters, pixels with similar brightness set in the same class. As regions in this image set into two clusters, minimum and maximum brightness of image will be centers of our two clusters in feature space and the magnitude value of pixels brightness from centers will determine the class of each pixel. To achieve more computational efficiency, the histogram of the image brightness is calculated during the clustering process. When all of the pixels are clustered, mean of the each cluster is computed. The result of applying c-means clustering on *Sunrise* image is represented in Figure 1B.

3.2 *k*-Means Clustering

This algorithm, which is one of the unsupervised iterative non-deterministic learning algorithms, aims to divide n features into k clusters wherein each feature belongs to certain cluster with the nearest mean [13]. The purpose is minimizing an objective function, e.g. a squared error function.

Assume we have a finite collection of n features, $X = \{x_1, x_2, \dots, x_n\}$,

1- k points are randomly chosen as a center of clusters and k clusters are generated.

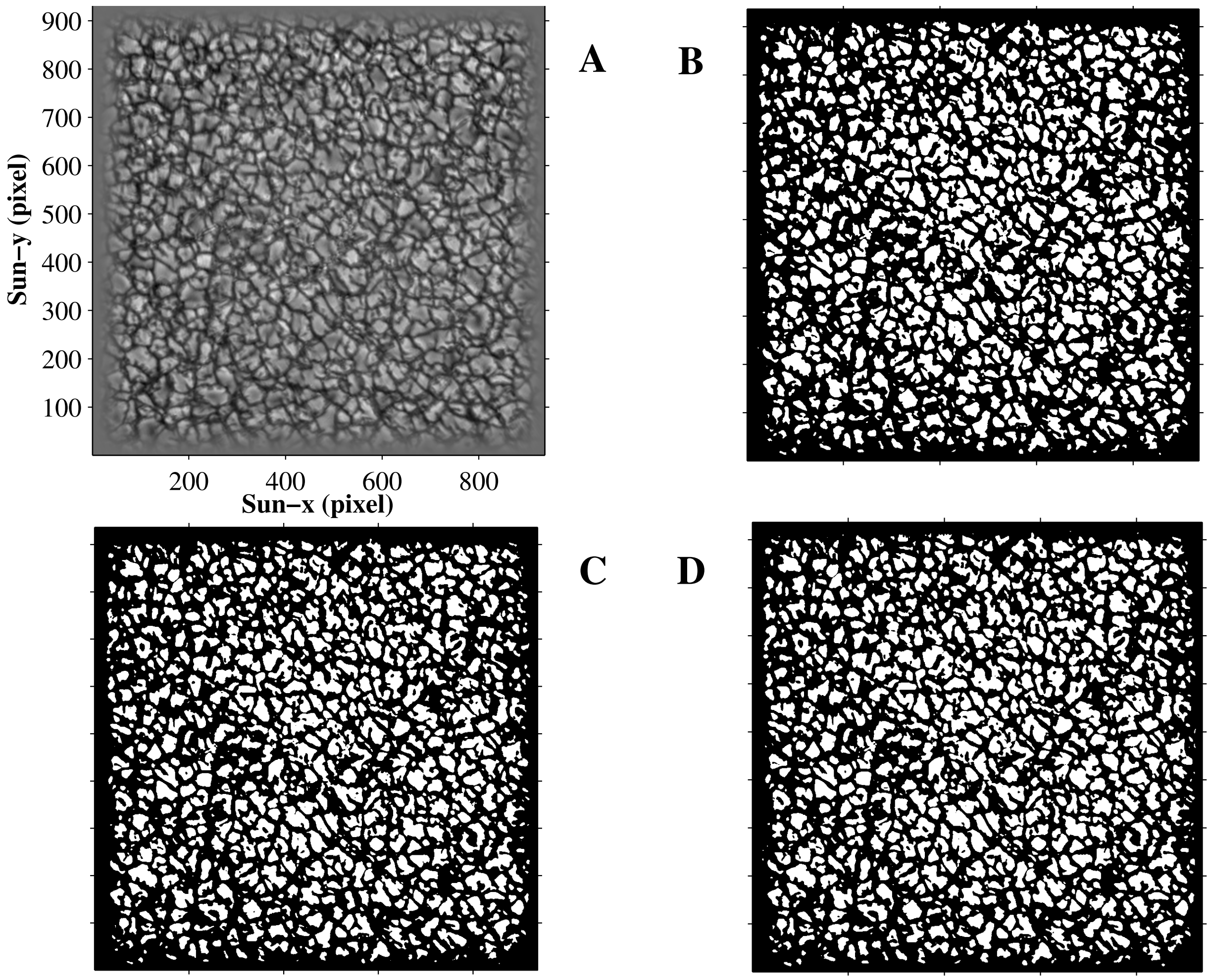

Figure 1: *Sunrise*/IMaX image (A) recorded on 9 June 2009 (14:16:00 UT). The segmented images based on c-means (B), k-means (C), and FCM procedure (D).

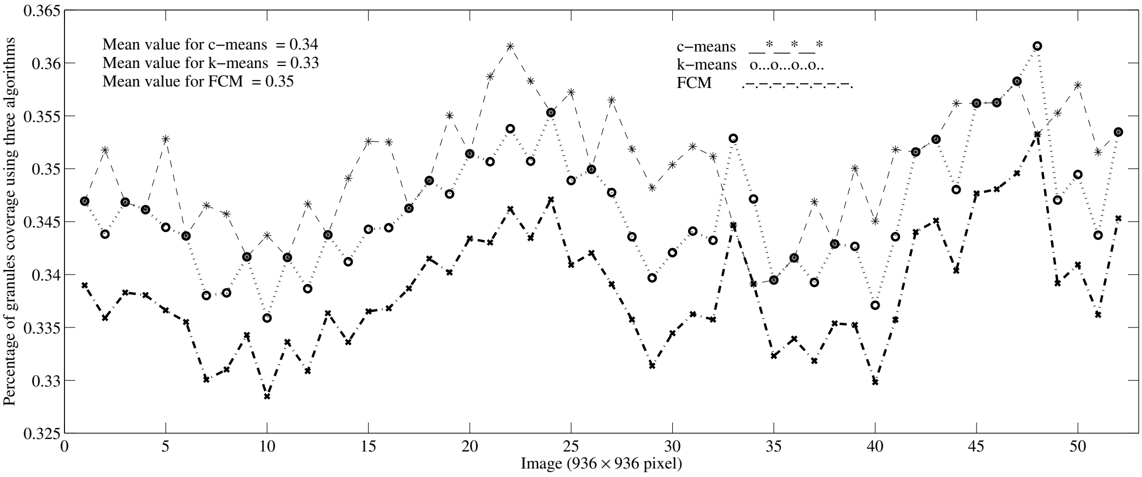

Figure 2: The filling factors of granules obtained by different clustering algorithms are presented. About one-third of each image is covered by granules.

2- Using association of each point to the nearest centroid, their belonging to a given data set is specified. Here, the objective function is introduced as

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \parallel x_i^{(j)} - c_j \parallel^2 . \tag{1}$$

3- By averaging all of the pixels, centers of clusters are recomputed.

4- Step 2 and 3 are repeated until the centers of clusters achieve to some convergency (Figure 1C).

3.3 Fuzzy c-Means (FCM) Clustering

The FCM, which is rooted in fuzzy logic, aims to minimize an objective function such as k-means algorithm. In the other words, we could label this flexible algorithm as an extension of k-means clustering. Fuzzy c-means allows data points to be assigned into more than one cluster. Each data point has a degree of membership (or probability) of belonging to each cluster. We have an objective function with an extra parameter

$$J = \sum_{i=1}^{n} \sum_{j=1}^{c} w_{ij}^m \parallel x_i^{(j)} - c_j \parallel^2, \tag{2}$$

where w_{ij} is a fuzzy membership qualification representing the membership of x_i to j cluster and m determines the degree of fuzziness:

$$w_{ij}^m = \frac{1}{\sum_{k=1}^{c} \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}. \tag{3}$$

Each element belongs to any fuzzy set with a degree of membership which takes a real value ranged from 0 to 1 [6]. Data points are given the partial degree of membership in multiple nearby clusters and final segmented image is obtained (Figure 1D).

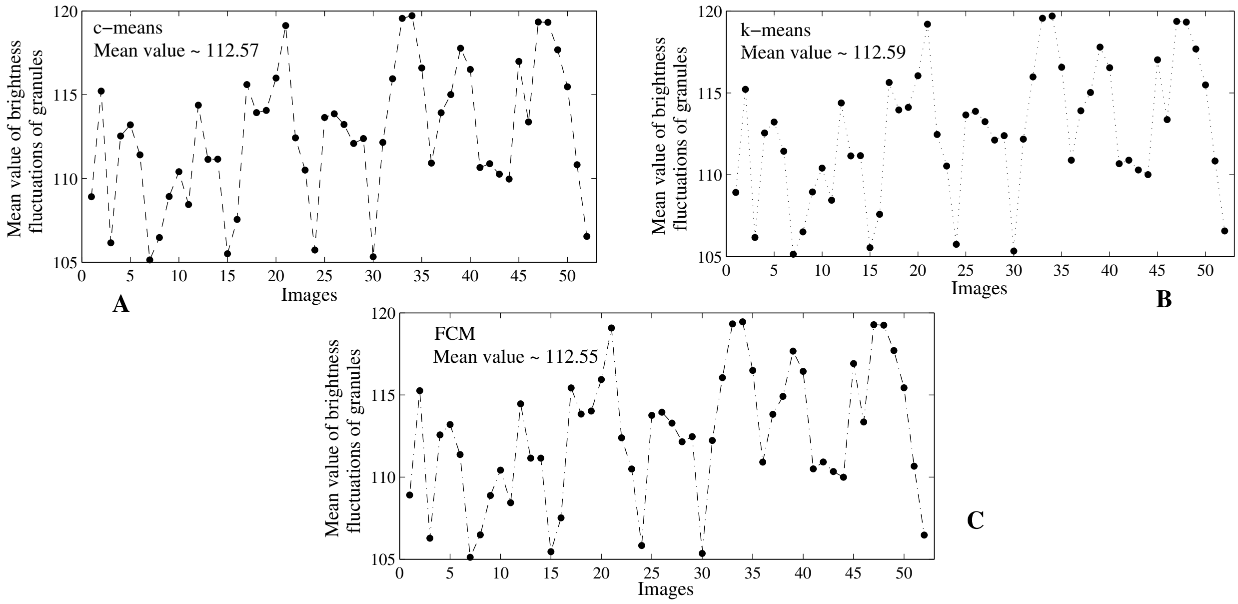

Figure 3: The brightness fluctuations of granules by using c-means (A), k-means (B), and FCM (C) algorithms are shown. The results approximately show the same behaviour of brightness fluctuations of granules extracted from algorthms.

4 Results

By applying the codes of c-means, k-means, and FCM algorithms, the sequence of photospheric images recorded by *Sunrise*/IMaX are segmented. Filling factors of granules (coverage area in each image) are calculated (Figure 2). The mean values of their filling factors derived from c-means, k-means, and FCM are about 0.34, 0.33 and 0.35, respectively. Fluctuations of filling factors obtained by algorithms are consistent and represent comparative results. Extracting the brightness fluctuations of granules from IMaX images show the same behaviour of these algorithms (Figure 3). The average of brightness fluctuations of granules is the same and equals with 112.5. The results represents that there is no preference in using these three methods to segment photospheric images. After applying FCM method on data (with membership degree of 0.5), the size distribution of granules is obtained (Figure 4). The power-law distribution is fitted with a power exponent $\alpha \approx 0.12$. The mean value of granule sizes is found to be about 1.8 arcsec^2 (0.85 Mm^2). The scatter plot of granules are shown in 5. It can be found granules follow two regimes that classifies them into two groups. Granules with sizes smaller than 2.8 arcsec^2 are more scattered in brightness, while larger ones are approaching mean value of brightness (1.3). A set of granules are followed from emergence up to disappearing in a series of images to attain lifetimes of granules with a normal size. The lifetimes are ranged from 6 to 10 minutes. The Number of granules is manually selected to compare with the results of these algorithms. It is found that false-positive detection appeared when granules are fragmented or dissolved.

The box-counting method, the technique to analyse the fractal dimension (FD) of an image, is employed to estimate changes between observations of detail and scale r by breaking our dataset. The purpose is finding the slope of the logarithmic regression line for both the number n and the size r of boxes. In this approach, by changing the resolution of box applied

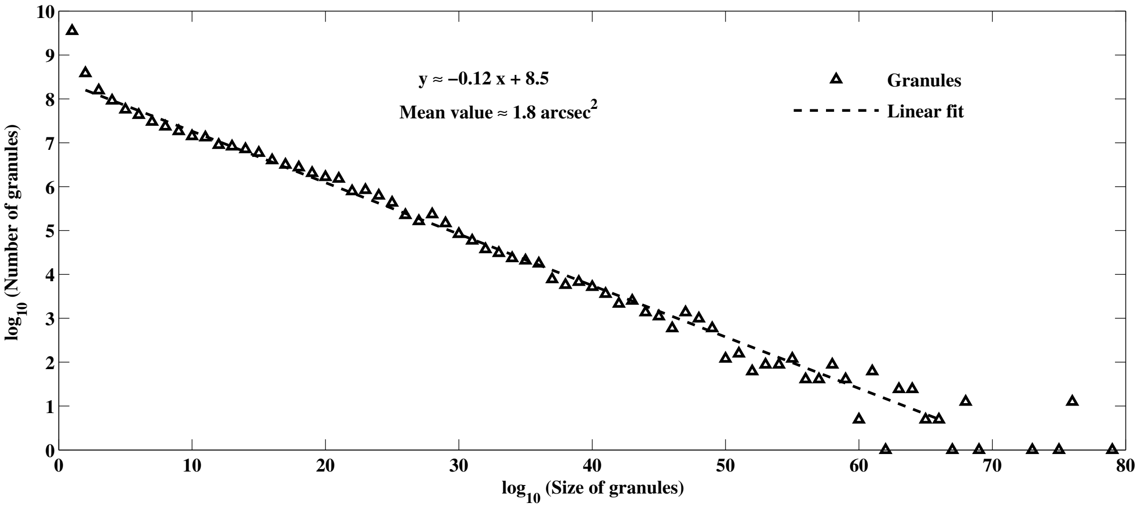

Figure 4: The size distribution of granules. The linear fit, $N \propto A^{-\alpha}$, is presented on a log scale with $\alpha \approx 0.12$.

on image pixels in each step (i.e., laying a series of grids), the size of elements that would be elongated within the box frequently varies in a form of power-low distribution [14]. For 2D (3D) images, if a power exponent is obtained with 2 (3), it shows random structure within the data set. The less values of power exponent demonstrate that how much the pattern of features captured in box is scale independence. In the case of local dimension box counting, the box for each r is centred on each pixel of interest. In box sizes ranged 10 and 100 the local FD goes to 1.8 (Figure 6). It can be said that granules appeared in grids with special box sizes represents their scale independence.

5 Conclusions

In this paper, an automated detection and characterization of photospheric granules were presented. Image processing methods and machine learning algorithms were applied to recognize and determine the physical properties of granules. The processes of three kinds of image segmentations (c-means, k-means, and FCM) were employed. The output images and the average of brightness fluctuation of granules derived from segmentations demonstrate that the detections of granules' locations are exactly correct. On the basis of Figure 2, it can be seen that one third of quiet-Sun is covered with granules. The mean value of granules size is consistent with previous studies [15, 16]. Two regimes governed on granules describe the turbulent eddies of smaller granules (< 2.8 arcsec^2) [15, 16]. The mean value of granular lifetimes is about 7 minutes in the quiet-Sun. Figure 6 indicates that it would be existed a scale range [50, 100] in which the FD estimate $D(r)$ shows the fractal characteristics of granulation. The entire computing time of procedures has taken about 5 minutes for data series in MATLAB environment. The top speed and high efficiency of unsupervised learning algorithms is the characteristics of these applications and make them easy to use.

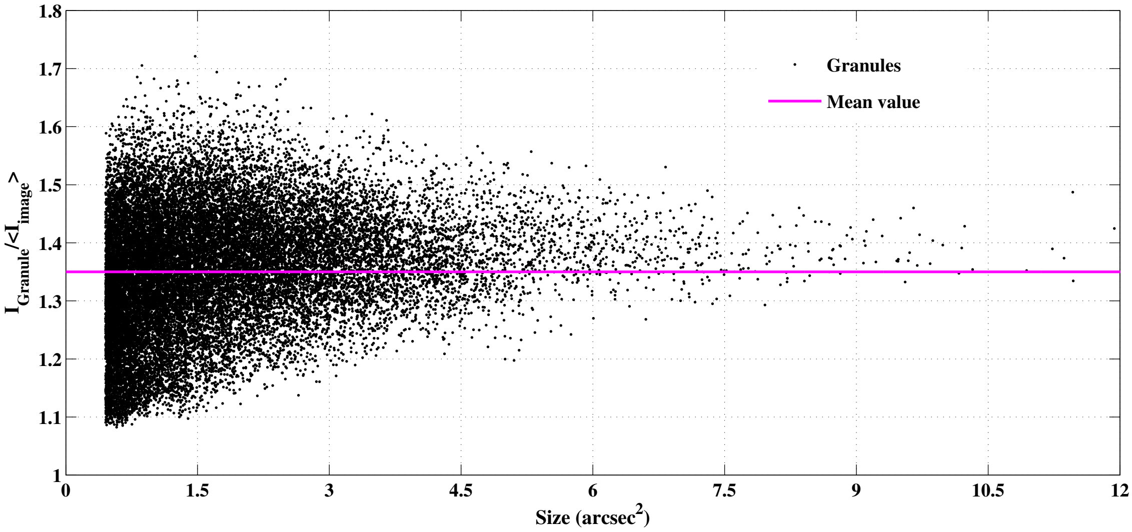

Figure 5: The scatter plot of granules brightness *vs.* size. The overall mean value of the scatter plot of the granule brightness is 1.35. For large granules, the brightness approaches a particular value (1.3).

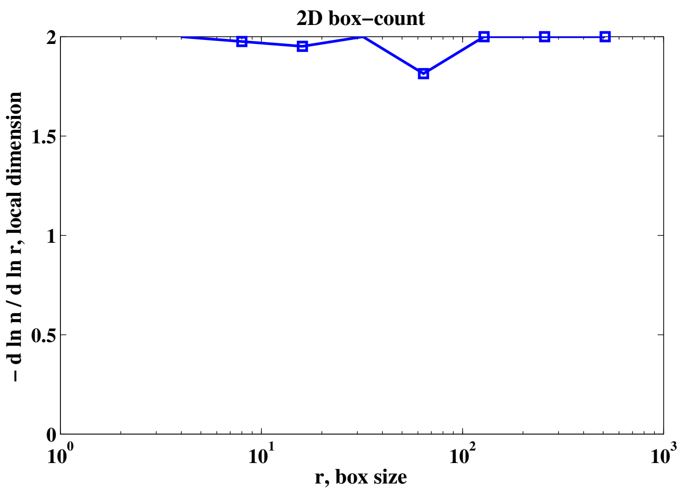

Figure 6: Local scaling exponent. If fractal characteristics are appeared over a limited range of box size r within the image, this can be demonstrated better by plotting the local exponent, $D(r) = -\frac{d \ln n}{d \ln r}$
.

Acknowledgment

The German contribution to *Sunrise* is funded by the Bundesministerium für Wirtschaft und Technologie through Deutsches Zentrum fr Luftund Raumfahrt e.V. (DLR), Grant No.50 OU 0401, and by the Innovationsfond of the President of the Max Planck Society (MPG). The Spanish contribution has been funded by the Spanish MICINN under projects ESP2006-13030-C06 and AYA2009-14105-C06 (including European FEDER funds). HAO/NCAR is sponsored by the National Science Foundation, and the HAO Contribution to *Sunrise* was partly funded through NASA grant number NNX08AH38G.

References

[1] de Wijn A.G., Stenflo J.O., Solanki S.K., Tsuneta S., 2008, Space Sci Rev 144, 275

[2] Riethmüller T.L., Solanki S.K., Martnez Pillet V., Hirzberger J., Feller A., Bonet J. A., et al., 2010, APJ Lett. 723, 169

[3] Berger T.E., Scherijver C.J., Shine R.A., Tarbell T.D., Title A.M., Scharmer G., 1995, APJ 454, 531

[4] Cloutman L.D., 1978 APJ 227, 614

[5] Barra V., Delouille V., Hochedez J.F., 2007, Adv. Spa. Res. 42, 917

[6] Bezdek J.C., Ehrlich R., Full W., 1984, Computers and Geosciences 10.2 191-203.

[7] Verbeeck C., Delouille V., Mampaey B., De Visscher R., 2013, A&A 561, A29 1

[8] Barthol P., Gandorfer A., Solanki S. K., Schüssler M., Chares B., Curdt W., et al., 2010, Solar Phys. 268, 1

[9] Solanki S.K., Barthol P., Danilovic S., Feller A., Gandorfer A., Hirzberger J., et al., 2010, APJ Lett. 723, 127

[10] Gandorfer A., Grauf B., Barthol P., Riethmüller T.L., Solanki S.K., Chares B., et al., 2010, Solar Phys. 268, 35

[11] Marínez Pillet V., del Toro Iniesta J.C., Álvarez-Herrero A., Domingo V., Bonet J.A., González Fernández L., et al., 2010, Solar Phys. 268, 57

[12] Berkefeld T., Schmidt W., Soltau D., Bell A., Doerr H.P., Feger B., et al., 2010, Solar Phys. 268, 103

[13] Jipkate B.R., Gohokar V.V., 2012, International Journal Of Computational Engineering 2 737

[14] Smith T.G., Lange G.D., Marks W.B., 1996, Journal of Neuroscience Methods 69 (2), 123

[15] Roudier T., Muller R., 1987, Solar Phys. 107, 11

[16] Javaherian M., Safari H., Amiri, A., Ziaei, S., 2014, Solar Phys. 289, 3969