Methods of computational topology for Solar Activity forecasting

(Extended abstract)

Irina Knyazeva
Central Astronomical Observatory
of the RAS at Pulkovo
Russia, St.Petersburg
Email: iknyazeva@gmail.com

Nikolay Makarenko
Central Astronomical Observatory
of the RAS at Pulkovo
Russia, St.Petersburg
Email: ng-makar@mail.ru

Daria Malkova
Yaroslavl State University
Yaroslavl, Russia
Email: dasha.m91@gmail.com

***Abstract*—Solar activity is a space-time complex of events which produced by the Sun magnetic fields. One of the results of this activity is a huge plasma ejection which called solar flares. The solar flares occurs mainly in the areas with an especially strong magnetic fields called Active Regions (AR). Observation phenomenology indicates that significant change in the magnetic field topology precede the strong flares. We investigated changes in topology by the methods of computational topology. For this purpose the high frequency temporal sequence of AR magnetograms containing flares has been analyzed. Such data are available from the space observatory SDO. We seek distinctive patterns that could be associated with the flares through the tracking evolution of Euler characteristics and Betti numbers. These characteristics of course do not pretend on the comprehensive description of topological complexity but there are simple in construction and intuitive clear. We found that the large variations of the Betti numbers and Euler characteristics are preceded or accompanied by a large flares. These results give us hope that the approach based on computational topology could be useful in the task of monitoring of magnetic field evolution and should be developed in future.**

Keywords*—*Solar Flares, persistent homology, computational topology

I. Introduction

Large solar flares are the most dramatic result of the evolution of the magnetic fields in sunspots. Energetic flares which are occurred near the center of the Solar disc could make the disastrous damage of the terrestrial and space equipment. Large flares tend to occur in the big groups of spots so called Active Regions (AR) of the Sun. These groups may contain more of a dozen of spots with different polarities which are forming topological complex spatial configuration of the magnetic field. There is a system for the numbering of active regions. The National Oceanic and Atmospheric Administration (NOAA) numbers active regions consecutively as they are observed on the Sun.

The problem of early prediction of a period of high energy releases which accompanied by big flares is important task of solar flares. At this moment numerous approaches were published. Roughly it could be divided on two classes [1]. One is based on fundamental physical parameters and second on proxy attributes or phenomenological properties of Solar data. The list commonly used and accepted parameters could be found at [2]. Each parameter has own context and some physical meaning. At the most of investigations these parameters were applied to big statistical sample of flaring and non-flaring regions. Leka and Barnes [2] have been analyzed and compared existing approaches to solar flares prediction using the photospheric magnetic field and concluded that there is no significant differences between all of them and despite of high statistical success rate non of them could be used for robust daily flare forecast. As a conclusion the authors advice as one of further approaches is to consider the evolution of magnetic field. Such approach is very demanding to the quality of data. It should be constant-quality, high-cadence and long time series of photospheric magnetograms. Such data at first were available from the SOHO observatory and by now it is improved successor the Helioseismic and Magnetic Imager [3] of the Solar Dynamics Observatory. We present here results for NOAA AR 11158, which were the first large flare product AR tracked with the SDO laboratory.

The processes leading flare appearance and energy release are still not fully understood [1]. The initial impulse phase of the flare is generally believed to be driven by magnetic reconnection which leads changes in topology of magnetic fields. So called new emerging magnetic flux appeared before the solar flare. The reasons to believe that new emergence flux connected with the solar activity discussed in our previous work[4]. In practice it must be seen in additional critical points appearance, or emergence of thin structure in the ”old” magnetic field. In other words it must be seen in changing of topological complexity of magnetic field, but for analysis we need to introduce a formal criteria of complexity. Our approach mainly is based on the R. Adler [5] ideas of random fields topological complexity description.

In our early works [4],[6] we used computation of the Euler characteristic,and other morphological functional by excursion set of magnetogram and concluded that Euler is more informative . Now we apply methods of computational topology [7] and compute persistence homology for analysis [13]. The computation were made for more then dozen Active Regions which produced solar flares and solar quiet regions. It was found that there is specific dynamics preceding the solar flares, but there are no typical scenario for all active regions. But our work now in progress and we have many unresolved yet questions which discussed below.

This paper is structured as follows: Section 2 describes the

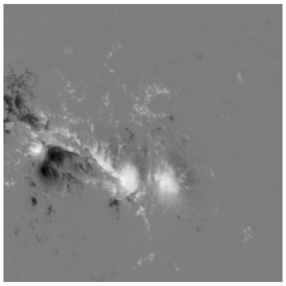

Fig. 1. Typical view line of sight HMI magnetogram, here AR 11158

solar data which we use for analysis and which could be used in future and our approach to the time evolution analysis. In Section 3 we give several examples of our estimation. And at the last section we summarizes the results and discuss our problem.

II. TOPOLOGY OF SOLAR MAGNETOGRAM

Magnetogram ,which used for numerical analysis of Solar magnetic field, is an image of Sun disc where each pixel represents a strength of magnetic field. Some magnetograph can measure only line of sight (LOS) from the observer component of magnetic field, others could measure also a transverse component from which all three components of a magnetic field can be deduced. The first magnetograph which produced continuous, constant-quality, high-cadence and long time series magnetogram was MDI (Michelson doppler imager) on the board of space observatory SOHO. It could measure only LOS component. Its successor HMI(Helioseismic and Magnetic Imager) could measure also a transverse component. A spatial resolution of MDI data $\approx 2''$/pixel or 1500 km, with 1024×1024 for full solar disc, time cadence is 96 min. SDO data have resolution $\approx 0.5''$/pixel and a time cadence is 12 min. Up to now we have worked only with the line of sight component, because it is available from both instrument. But of course in future additional component also should be analyzed. We work with a fragment of Solar disc containing a AR. Typical magnetogram of the AR is shown in Figure 1. The automated system of the Active region patches which track the location throughout lifetime developed by SDO observers team is used [3].

To describe a topological complexity of magnetic field we suggest to use methods of mathematical morphology and computational topology.

A. Mathematical morphology

The ideas of mathematical morphology were introduced by J. Serra [**?**] for binary sets. After that they were generalized for physical random fields by R. Adler [5]. For two dimensional fields the so-called excursion sets is considered. This is a set formed by the values which exceeds the specified values. Let's consider an excursion set $A_u = \{\mathbf{x} \in W : B(\mathbf{x}) \geq u\}$ of the field in a compact region W , formed by the pixels $\mathbf{x} \in W$ where the magnetic field $B(\mathbf{x})$ exceeds a specified level u. We mark these pixels black. So we translate each magnetogram into a set of black and white images, one for each selected level.

On the excursion set in Euclidean space Minkowski functionals, such as perimeter, area and Euler characteristic (EC or χ), could be estimated, see [5]. These functional have clear physical interpretation [6]. Area closely linked with unsigned flux of magnetic field. Total variation of gradient of field connected with the functional Perimeter through the co-area formula. And finally Euler characteristics according to the Morse theory [9] counts the number of critical points: $\chi(A_u) = \sharp(\text{maxima} + \text{minima}) - \sharp\text{saddle}$. That is why it could be interpret as a measure of topological complexity. As a morphological functional Euler characteristic is just the number of connected components minus the number of holes. Really at the excursion set minima and maxima correspond to connected components and holes arise only when the saddle point appeares.

In our previous work we have been computed all three morphological functionals, but now we are studying only the Euler characteristics as the most informative. The drawback of the Euler characteristics is that in this case critical points are described integrally. So increase in Euler characteristics could be caused as a change in maxima or minima and a saddle point. From the physical point of view it is important to distinct emergence of new structures in the field and changes in the topology of existent field. Moreover it is important to track the lifetime of critical points. All of this developments were made within the context of computational topology and could be find in [10].

B. Persistence homology

Recall the ideas of "Morse filtration" of excursion sets [10]. Obviously that for two excursion sets $A_u \subseteq A_v$ if$u \geq v$. Going from one level set to another, components of excursion set may merge and new components arise. Also the topology of these components may change, holes could appeared and disappeared. For classification of shapes of objects and structures formalism of algebraic topology is used, precisely notion of homology. In two dimensional space X the zeros homology $H_0(X)$ is generated by connected components of X, the homology $H_1(X)$ generated by the holes, the numbers of the components and the holes called β_0 and β_1 respectively. Then the difference between Betti numbers $\beta_0 - \beta_1$ will be the Euler characteristics. Track the changes in the homology of the sets as the function of levels sets calls persistence homology. persist. The term persistence comes from that fact that changes in homology arise only at critical points of the fields. Between them homology stay "persist". The broad description of basis of homology could be found in [7]. It is useful to describe persistence homology via notion of barcodes. A bar for each homology group, in our case components and holes, starts with the birth of component and ends with the level of component. First point will be the level of birth and the second level of death. It is useful to draw it on the plane using the beginning and the end of the barcode as point coordinates. As the result we obtain a set of points which lies above the diagonal that corresponds to barcodes of the zero length. This graph is called a persistence diagram.

It is convenient to give some simple structure at the neighborhood of the maximum — so-called simplicial structure. Without giving any formal basis we only describe an algorithm that was used. The incremental algorithm for computing homology which we used in our work is described in [11]. Modification for two dimensional matrices could be found in [13]. It consists of two sequential steps: filter construction of simplices (for two-dimensional images the simplex is a vertex, an edge or a triangle) and computing the Betti numbers on the created filtration. Let $f(x, y)$ is a value at pixel (x, y). For the filter construction we need to determine the function value for each of simplices. In order to do this we associate each pixel (x, y) of the image with the vertex. We define the value for the remaining simplices by assigning the maximum of values between their vertices. After that we iterate through all elements of the ordered sequence and add each of them to the filter. At the same time, attaching the new vertex to the filter we add all edges and all triangles that can be generated by the vertices which we already have in the filter and the new vertex. Now Betti numbers could be computed by processing the simplices in the filter and keeping track of changes in connectivity of the obtaining set. If vertices of the current edge belong to different connected components, then after merging them into a single component we suppose that the component, which appeared later than another one, disappears or "dies". In that way we can keep track of "birth" and "death" of connected components at the intensity levels. To compute the life time of holes, i. e. for the number β_1, we use the same algorithm applying it to a dual graph. (In the dual graph to each vertex corresponds the triangle of the initial graph, to each triangle corresponds the vertex in the initial graph and to each edge corresponds the dual edge). If we sum all life lengths for β_0 and for β_1 and take difference of them we receive the Euler characteristic of persistence diagram [12].

III. Results

We used a time sequence of magnetograms of the full solar disk, obtained with the help of the HMI tool. A time interval between magnetograms was 720 seconds, and the noise level does not exceed 6 gauss. A fragment of 600×600 pixels containing the AR was cut from each magnetogram. For the specified 720 seconds time gap about 700 consecutive images of the same active region passing across the solar disk were available. We used *FI* index of flare productivity to compare the variations to flare activity. Roughly speaking, it measures a weighted amount of energy produced by solar flares of various classes in the finite time interval. The flare classes *FI* were converted to numeric values. The magnitudes of C class flares were not altered, for M class flares the magnitudes were multiplied by 10, for class X were multiplied by 100, and for B class were divided by 10. We present here the results of numerical experiments for two flare-active regions AR 11520 and AR 11158.

AR 11158 appeared near the center of the solar disk as a compact β-class bipolar group on February 12, 2011. Within a day it reached δ magnetic class and on 12 February produced a flare of class M6.6. A day later M2.2 flare followed, and, finally, on 15 February X2.2 flare occurred. After this activity of this AR actually stops. The dynamics of the Euler characteristic for the high levels of magnetic field strength is shown in Figure 2. In Figure 3 represents a behavior of the persistence homology difference $B_0 - B_1$. The complexity of the field in Figure 2) is growing for the fields of north and south polarities, anticipating an increase in flare productivity. Little depression could be seen before the big flare. For comparison, Figure 3

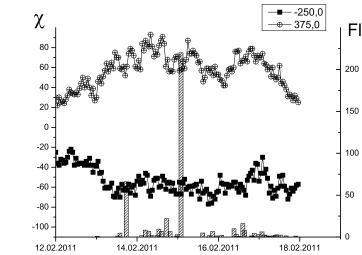

Fig. 2. The dynamics of EC for AR 11158: high levels of magnetic field strength.

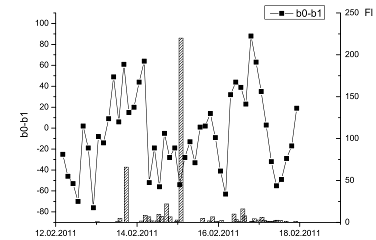

Fig. 3. The dynamics of persistent homology b0-b1

shows the behavior of the Euler characteristic obtained by the persistent homology . Here we note a depression in the EC graph preceding the phase of flare activity. The depression is the most obvious about a day before the X flare.

AR 11520. This active region appeared on the Sun at July 8, 2012. It was immediately assigned to the class of complex large groups of δ-configuration with a possible high flare productivity. Initially, the region was a single large penumbra which contained many small spots of the opposite polarity. In the course of evolution it began quickly disintegrate into several compact regions. Against all expectations, the AR 11520 produced only four flares of M class and one flare X1.4 on 12 July. The last flare approximately corresponded to the localization of the group near the center of the solar disk. After that the AR 11520 flare activity stopped. In Figure 4 dynamics of EC at high levels of magnetic field is shown, before the X flare strong depression could be seen. Figure 5 represents the evaluations of the Euler characteristic for the AR 11520 obtained by the persistent homology are shown. Again we can see well marked variations in topological complexity of the field before the X flare.

IV. Conclusion

The main purpose of the present work was to develop some topological approaches for the analysis of the dynamics of magnetic field of the Sun which are focused to the search of pre-flare scenarios. Approach is shown on an example of AR 11520 and AR 11158. For these active regions the strongest flares of the class X far from the limb of the disk were observed. Using the corresponding sequence of magnetograms we obtained time variations of the Euler characteristic (EC) at the excursion sets and the EC of persistence diagram. In case when the EC is computed for each of the excursion set,

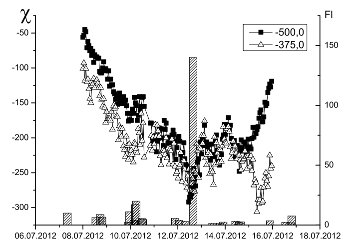

Fig. 4. The dynamics of EC for AR 11520: high levels of magnetic field strength.

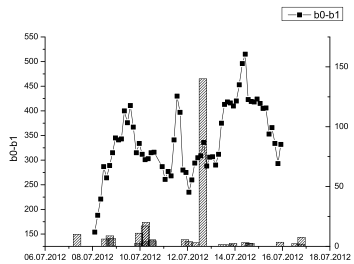

Fig. 5. The dynamics of persistent homology b0-b1

we need to specify some level of magnetic field and track evolution of chosen EC. Empirically as such level we took the rather big level of magnetic field near 500 Gs.

The AR under the study shows different dynamics which are tracked by changes in topological characteristics. Typically significant variations of the EC at the excursion sets and EC of persistence diagram often precede the flares. Note that the results presented in this paper confirm our earlier works obtained from the MDI/SOHO magnetograms. This fact slightly compensates for a lack of the adequate statistical sample restricted by the low level of the solar activity at the present time. Nevertheless, topological approaches satisfy the empirical considerations of the primary role of topological changes in the magnetic fields of active regions.

V. Problems for discussion

With a modern instrument for each active region thousands of magnetograms are available. Each magnetogram represents high variable data with the size of 600×600 px or higher. The differences between two successor magnetogram are very small. Even the expert observer couldn't tell authentically looking at the evolution of magnetograms was the AR flare active or not, and especially couldn't tell when the flare was. But it is believed that there are precursors in photospheric field. That's why it was created the system of Spaceweather HMI Active Region Patch (SHARP) (http://jsoc.stanford.edu/doc/data/hmi/sharp/sharp.htm). Each AR is tracked during crossing the face of the solar disk with this system. Data of the vector field in several projections are available online. More then twenty parameters of the field calculated by the magnetogram are also available. But the evolutional curve of these parameters rather complicated as for flare active as for flare quiet AR without obvious relation with the flares. We believe that for each magnetogram with the methods of computational topology we could fully describe topological structure of photospheric fields. But for time evolution analysis we need to extract simple characteristics from them and do not loss the information. Namely, during the analysis of EC by level sets we need to choose specific set. We take this level by the empirical way, and really there is not any grounding which level for each AR should be taken. After the computing persistence diagram we take for analysis EC of persistence diagram and loose a lot of information about structure of persistence diagram, about behavior of Betti numbers curve for each magnetogram. We need an effective measure for persistence diagram comparing. Also we work under the form of Betti number curves comparing. In addition up to now we work with only one component of magnetic field. With SDO/HMI there are available all three components, so we are waiting big progress in studding of vector magnetic field of the Sun because of high cadence date with good resolution. But it is not clear for us how to apply methods of algebraic topology for a vector field. At this moment we are working separately with all the components and after that try to extract independent characteristics. Another one problem is that magnetic field of AR is bipolar. Each polarity has its own structure of maxima and minima. But now we didn't take it into account, and it is open question how to generalize the idea of persistence in the case of a bipolar field.

References

[1] Manolis K. Georgoulis, "Toward an Efficient Prediction of Solar Flares: Which Parameters, and How?", *Entropy* vol.15, 5022-5052; doi:10.3390/e15115022, 2013.

[2] G. Barnes and K.D. Leka, "Evaluating the Performance of Solar Flare Forecasting Methods", *The Astrophysical Journal*, vol. 688, L107-L110, December 2008.

[3] J.T. Hoeksema, Y. Liu, K. Hayashi, X. Sun, J. Schou, S. Couvidat, A. Norton, M. Bobra, R. Centeno, K.D. Leka, G. Barnes, and M. Turmon, "The Helioseismic and Magnetic Imager (HMI) Vector Magnetic Field Pipeline: Overview and Performance", *Sol. Phys.*, vol. 289, pp. 3483-3530, 2014.

[4] I. Knyazeva, N. Makarenko, M. Livshits, "Detection of New Emerging Magnetic Flux from the Topology of SOHO/MDI Magnetograms", *Astronomy Reports*, vol.55, no.5, pp.463-471, 2011.

[5] R.J. Adler. *The Geometry of Random Fields*. John Wiley , N.Y. 1981.

[6] N. G. Makarenko, I. S.Knyazeva, and L.M. Karimova. Magnetic Field Topology by MDI Data: Active Region. *Astronomy Letters*, vol.38, no. 8, pp. 531-542. 2012.

[7] H. Edelsbrunner, J.L. Harer. *Computational Topology. An Introduction*. Amer. Math. Soc. Providence, Rhode Island, 2010.

[8] I. S. Knyazeva, N. G. Makarenko, "Evolution of the Magnetic Fields of Solar Flare Active Regions, from the Geometry and Topology of HMI/SDO Magnetograms",*Astronomy Reports*, vol.58, no.3, pp. 145-155, 2014.

[9] Y. Matsumoto. *An Introduction to Morse Theory*. Providence: AMS 2002.

[10] R.Adler, O. Bobrowski, M. S. Borman, El. Subag, and Sh. Weinberger. Persistent homology for random fields and complexes. *IMS Collections*,vol.6,24-143, 2010.

[11] C. A. Delfinado, H. Edelsbrunner, "An Incremental Algorithm for Betti Numbers of Simplicial Complexes", *Computer Aided Geometric Design*, vol. 12(7), pp. 771-784, 2005.

[12] O.Bobrowski, *Algebraic Topology of Random Fields and Complexes*, PhD Thesis, Israel Institute of Thechnology, 2012.

[13] N. Makarenko, D. Malkova, M. Machin, I. Knyazeva, I. Makarenko, "Methods of computational topology for the analysis of dynamics of active regions of the Sun", *Fundam. Prikl. Mat.*, vol. 18(2), pp. 79-93, 2013.