A Study of The Relationship Among Parameters of M/X Solar Flares via Association Rules

Jéssica de Farias Pereira *jessica.pereira1764@gmail.com*
School of Technology
University of Campinas
Limeira, 13484-332, Brazil

Ana Estela Antunes da Silva *aeasilva@ft.unicamp.br*
School of Technology
University of Campinas
Limeira, 13484-332, Brazil

André Leon Sampaio Gradvohl *gradvohl@ft.unicamp.br*
School of Technology
University of Campinas
Limeira, 13484-332, Brazil

Guilherme Palermo Coelho *guilherme@ft.unicamp.br*
School of Technology
University of Campinas
Limeira, 13484-332, Brazil

José Roberto Cecatto *jr.cecatto@inpe.br*
Astrophysics Division
National Institute of Space Research
São José dos Campos, 12227-010, Brazil

Abstract

This paper introduces a method to study the relation among parameters that can cause the origin of M/X solar flares. Solar flares, especially flares of types M and X, make the Earth's atmosphere more ionized and have an effect on radio signals, which can cause disruptions in wireless communications. This situation points out to the need for better identification of the parameters involved in M/X solar flares. The method is based on four categorical parameters and their relations. Relations are demonstrated by association rules which were extracted by the APRIORI algorithm and the most promising rules were filtered by support and confidence metrics. Results of the most promising rules had been compared by application to different periods of the 23rd and the 24th solar cycles.

Keywords: Solar Flares, Data Mining, Association Rules.

1. INTRODUCTION

A basic characteristic of the Sun is that its activities are described by a cycle of eleven years, on average. During this period, activities range from the minimum through the ascending phase of the cycle towards the maximum. From then on, there is a decay phase of the cycle towards the next minimum. Activities follow this typical behavior to the next cycle. What is observed during a cycle, for instance, is a gradual increase in the number of sunspots on the visible solar disk, from the minimum towards the maximum of the cycle, when that number reaches a top level along a few years, followed by a gradual decrease along the decaying phase towards the next minimum. Similar behavior is observed in the total solar irradiance and frequency of energetic solar phenomena.

Among the energetic phenomena observed during those solar active periods is the flare, which is a sudden release of the magnetic energy hoarded within an active region of the solar atmosphere. It is recorded as a strong glare of radiation observed in a wide spectrum from radio waves up to X-rays or gamma-rays that lasts from a few minutes to a few hours, in case of the most energetic flares. The flares can be classified into five classes (A, B, C, M, and X), according to the X-ray flux level recorded in Watt per square meter on a logarithmic scale, being A the lowest and X the highest intensity flares.

This work focuses on the study of M/X solar flares and the relationship among their parameters. Solar flares, when intense, can trigger events that may affect the Earth and our daily activities [1]. The effects caused by these phenomena interfere in technological systems, such as mobile signals, satellites and global positioning devices [2].

A great challenge regarding the solar flare is its appropriate prediction. Once a solar flare can be predicted with due anticipation, it is possible to minimize the damages it causes. Since the 1960s, in the 20th century, many studies have tried to predict the occurrence of solar flares. However, to adequately predict it, there is still the need to understand which parameters are the most important for its occurrence and how they influence flares' triggering. Such parameters include: radio emission, X-ray emission, the number and area of sunspots, and many others [3-6].

We propose the use of data mining [7] to study some of those most important parameters and their relation in solar flares of classes M and X. Data mining is an area of computing that intends to find patterns in datasets instances. Patterns are occurrences of data that have similar characteristics or behave similarly during a specific event. Three main data mining tasks can be performed in datasets: association, classification, and regression. The problem of finding associations among data can be solved by the association task. As its basic premise, the association task has to find elements that imply the presence of other elements in the same transaction in a dataset. This kind of relationship among elements is called association rule. Association rules are presented, in datasets, under the form: IF X THEN Y, where, one set of attributes in X implies another set of attributes in Y.

The main idea of the paper is to apply an association rule algorithm to find out the association rules that can point out the relationship among parameters present in solar flare of classes M and X. Datasets containing data from 1997 to 2016 were considered as a whole. Besides, datasets from the 23th and 24th solar cycles were also used separately. Datasets and their attributes were chosen according to the main parameters used to predict solar flares [8].

This study aimed to: (i) choose a set of parameters on solar flares; (ii) pre-treat the dataset to allow the application of association rules; (iii) apply the APRIORI algorithm of association rules [9] and ; (iv) choose rules according to the best values of support and confidence which were the metrics used to analyze the rules.

Our main contributions are: (i) discovery of relations throughout association rules among some parameters of M/X solar flares; (ii) comparison of association rules found in different periods of solar cycles.

The paper is divided into six parts: Introduction which presented the motivation of this work; Background, in which the supporting concepts are described; a third section which presents some studies on the relationship among solar activities' parameters; Methodology, in which the methods used in the experiments are described; Results and Discussion, in which the experiments are explained and discussed and Conclusion and Future Works, in which the main points of the method and future studies are mentioned.

2. BACKGROUND

This section presents the main concepts which were used to create our methodology. As an interdisciplinary study, concepts of both areas – solar flares and association rules – are presented.

2.1 Solar Flares

The activities that occur in the Sun, including solar flares, are distributed into periods called solar cycles. Each cycle has, on average, eleven years of duration [10]. A solar flare is a large burst of electromagnetic radiation originated in an active region of the solar atmosphere that can attain the Earth about eight minutes later. It stems from the release of a large amount of magnetic energy that was previously hoarded inside an active region.

An active region of the Sun is an area with strong magnetic fields. Most solar events, like solar flares, are originated in these active regions, as the magnetic fields in active regions are stronger than the average magnetic field of the Sun. Active regions are more common during the peak of the solar cycle.

Sunspots are dark regions observed in the solar photosphere, usually surrounded by brighter areas, which indicates the presence of a strong magnetic field in the form of a spots-group (Figure 1a). Sunspots are indicators of active regions and are usually surrounded by lighter-shaded areas of magnetic disturbance. They seem dark, because they are almost two thousand degrees colder than the surrounding photosphere (5800 K), and constitute the feet of arch-like magnetic structures observed in the solar atmosphere.

Figure 1 shows the flare which has the number region NOAA 9077, and was named the Bastille day flare, observed on July 14th, 2000, at four distinct wavelengths (white light, 1600 Ångströms (Å), 195 Å, and 171 Å) [11]. The white light (WL) image (Figure 1a) was recorded at 10:28, Universal Time (UT), four minutes after peak time of the flare (10:24 UT), while the three ultraviolet (UV) wavelengths were recorded 20 minutes later. In Figure 1a, the white light image, it is possible to see the dark regions according to the sunspot pattern of the active region with identification number NOAA 9077.

The structure of the flare can be observed in Figures 1b, 1c and 1d, by the two-ribbon structure which forms the flare and limits the bright flaring arcades.

According to Qahwaji and Colak [12], the arrival of solar X-rays which are emitted by solar flares and travel at the speed of light can disrupt point-to-point high frequency radio communications. Besides, concurrent with X-ray emission, solar activity often emits radio noise that can interfere with communications and radar systems and can shorten the life of satellites. This decrease in the useful life of satellites is caused by the emission of high-energy particles [13].

Sello [14] describes the study of the activities of solar cycles, particularly, their prediction, as a challenging task due to the high frequency of solar radiation, the lack of a theoretical quantitative model of the magnetic cycle of the Sun, the noise in obtaining the data and the high variability in the phase and amplitude of cycles. Besides, the temporal aspect of solar activities involves the consideration of a set of time ranges in different kinds of indices of solar activity.

Because of these complex characteristics of solar activities, it is important to study the relations among parameters of solar flares, which is the focus of this research paper, in order to better understand solar phenomena.

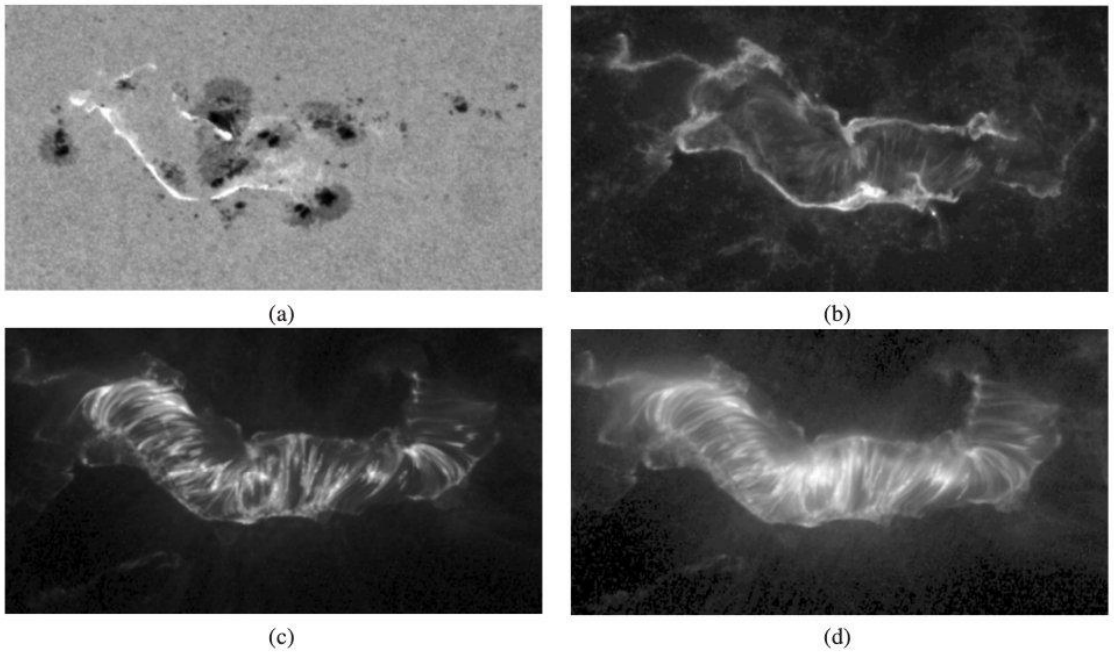

FIGURE 1: Multi-wavelength trace observations of the Bastille day flare occurred in the active region NOAA 9077 on July 14th of 2000. (a) WL image taken at 10:28 UT; (b) 1600 Å image taken at 10:48:24 UT; (c) 195 Å image taken at 10:48:38 UT; (d) 171 Å image taken at 10:48:32 UT [11].

2.2 Parameters of Solar Flares

Solar flares are regularly observed and recorded by various indexes coded in a numerical form as, for instance: area, magnetic classification of sunspots and the integrated fluxes measured in X-ray [8]. These various indexes, whose values can be found in the databases of the Space Weather Prediction Center [15], represent, in different ways, the characteristics of solar activities.

One of the sunspots classifications is the called Mt. Wilson, which is based on the distribution of magnetic polarities within specific groups, and which has eight sunspot classification classes: alpha, beta, beta-gamma, beta-delta, gamma, delta, beta-gamma delta, and gamma-delta [16]. There are two other sunspot classifications: the Zurich classification [17] and the McIntosh classification. In this study, the Mt. Wilson classification was adopted. The X-ray flux emitted by the Sun is another solar parameter regularly registered, corresponding to the background level of solar X-ray emission. As mentioned before, the X-ray flux, in Watt per square meter, is measured in the wavelength band of 1 to 8 Ångström and classified into five orders of magnitude on a logarithmic scale, as shown in Table 1 [18].

X-ray flux (1 to 8 Ångströms)	**Solar Flare Class**
flux < 10^{-7}	A
10^{-7} < flux < 10^{-6}	B
10^{-6} < flux < 10^{-5}	C
10^{-5} < flux < 10^{-4}	M
flux > 10^{-4}	X

TABLE 1: Classification of Flares In Terms of X-ray Flux Level.

2.3 Association Rules For Solar Flares

The exploration of frequent patterns refers to the search for recurring relationships in a given dataset. The idea of exploring these patterns is to discover interesting associations among attributes and the frequency with which these associations appear in the dataset [7]. A typical example of the exploitation of these patterns is the market basket transaction. Market basket transactions reflect the purchases made by certain customers in a store [19]. Each transaction is unique and contains the items that were purchased. An example of a market basket transaction could be a set of items purchased by a customer in a supermarket in a specific date. The basket could contain the products: milk, bread, butter and beer, meaning that these products were purchased at the same time by this customer. In this case, the set of items is considered as a transaction, and an identification number, known as transaction identification, would index each transaction.

An association rule represents the relationship found among items in a transaction, using an implication relation. Formally, a set of transactions $T = \{t_1, t_2, \ldots, t_n\}$, and a set of items $I = \{i_1, i_2, \ldots, i_n\}$, where a single transaction t_n contains a subset of items I called item-sets, are the basic concepts of association rules. Therefore, based on these concepts, it is possible to define an association rule as an $X \rightarrow Y$ rule, where X and Y are subsets of I and $X \cap Y = \varnothing$ [20].

By bringing the idea of market basket transactions to the solar flare context, we would have transactions whose items would be parameters of solar flares, as shown by the association rule: {area_huge, beta_gamma_delta, radio_veryhigh} → {Xray_X}.

From the example of the supermarket purchase, an association rule could assume the form of: {bread, butter} → {milk, beer}. In this case, X = {bread, butter} and Y = {milk, beer}. The rule can be described as: a customer who bought bread and butter also bought milk and beer.

For the example of the solar flare context, X = {area_huge, beta_gamma_delta, radio_veryhigh} and Y = {Xray_X}. This means that the combination of these parameters in the antecedent part of the rule results in the Xray_X parameter in the consequent part, which means that this rule represents a class-X solar flare.

According to Han, Kamber and Pei [7], to filter association rules from a dataset, the item-sets of the respective transactions must be selected. The minimum frequency of these item-sets in the general dataset is called support. For such item-set to become an interesting association rule, this minimum frequency number must be established and considered in rule creation, i.e., the item-set only becomes a rule when it reaches a minimum support value. The support of a rule is measured by the ratio of the number of transactions that contain a particular set of items and the total number of transactions. Formally, the support can be described by the formula given by Equation 1.

$$support\ (X \rightarrow Y) = \frac{supp(X \cup Y)}{|N|} \qquad (1)$$

Where: X is the set of items presented in the antecedent of the rule – items in the left side of the rule; Y is the set of items presented in the consequent of the rule – items in the right side of the rule; N is total number of transactions and *supp*(X ∪ Y) is the number of transactions in which item-sets of sets X and Y appear together [19].

In addition to the support, to filter an association rule, it is necessary to establish a minimum confidence level. While the support value can be used to filter out rules that are interesting because they impact the entire database, the confidence value provides the estimated probability of Y given X, i.e., P(Y|X). It shows the probability of, given Y as the consequent of the rule, X is the consequent [19]. Equation 2 shows how to calculate the confidence of a rule that contains X in the antecedent and Y in the consequent.

$$confidence\ (X \rightarrow Y) = \frac{supp(X \cup Y)}{supp(X)} \qquad (2)$$

It is important to explain that an association rule does not imply a cause-effect relation, but it indicates a strong relationship among the items present in the antecedent and consequent parts of the rule.

The objective of this study is to find out which parameters appear together in rules that result into solar flares of the types X and M. To do that, the consequent of the desired rules should contain item-sets that indicate the presence of M/X solar flares.

3. WORKS ON THE RELATIONSHIP AMONG SOLAR ACTIVITIES PARAMETERS

Studies that used association rules as a way of studying the parameters of solar flares were not found in the literature. Most of them presented statistical correlations between parameters of solar flares.

It is important to highlight that, unlike the statistical correlation that measures the relationship between two variables as, for example, the linear correlation, the discovery of an association rule indicates the probability of occurrence of a set of items given that some items also occur. Therefore, the idea of reporting the following studies is not about performing a numerical comparison to our work, but to show the importance of some parameters of solar activity and the variety of studies developed to analyze the relationship among these parameters.

Another point about this section is that most of the works cited here used different periods of the solar cycles in their studies, which also prevents a direct comparison between these works and ours. Following are some works which were taken into account because of the relevance of the solar parameters cited by them and the relation cited among such parameters.

According to [8], most studies involving solar activities and their impact on Earth have used the number of sunspots as the main parameter. Later on, the X-ray flux also began to be used. In recent years, many other parameters have been used to study the relationship among parameters in solar activities. Authors describe a correlative study using the monthly data of the following solar activity parameters: sunspot numbers, solar flux, grouped solar flares, coronal index and tilt angle for the solar cycles from 19 to 23.

Feng, Yu and Yang [21] investigated the relationship between grouped solar flares and sunspot numbers during the time interval from Jan-1965 to Jun-2008. Grouped Solar Flares is another type of solar index. The term grouped means that the solar flares observations were made in different locations by different solar observatories and then considered as one [5]. It was found that the relationship between grouped solar flares and sunspot numbers is not only time-dependent but also frequency dependent. According to the authors, the relationship between grouped solar flares and sunspot numbers is a complex nonlinear relationship, despite being highly correlated with each other.

In the study of Meera, Munika and Shrivastava [5], several parameters involved in solar flares were considered to study the relationship among them. The parameters considered were: sunspot numbers, solar flux (10.7 cm flux), grouped solar flare, solar flare index, and coronal index for the last five 11-year solar cycle (20 to 24). According to the authors, the sunspot number is the most famous and popular index for studying the solar activity, being the most analyzed time series in the solar physics. The sunspot number was the solar activity parameter that showed the highest degree of correlation with other parameters. The highest degree of correlation was between sunspot numbers and solar flux.

In the study by Shen, Dun, Zhang, and Jiang [22], the Principal Component Analysis (PCA) method was used to analyze the main parameters of the solar active regions with solar proton events. Using the PCA method, the main parameters of solar active regions were analyzed to get the scores of the principal components of the active regions involved in solar activities. According to the authors, the main parameters to represent the solar activity strength are the relative sunspot number and 10.7 cm radio flux.

The objective of current study is to discover association rules among parameters that are involved in solar flares of classes M and X. As described by the previous studies, some of the most studied solar flare parameters are: X-ray flux, radio flux and number of sunspots. In addition, the Grouped Solar Flares index showed an association with the sunspot number, reinforcing the importance of the sunspot number during solar flares. The parameters X-ray flux, radio flux, number of sunspots and the Mt. Wilson classification were used in this work.

4. METHODOLOGY

This section presents the preparation of the transactional dataset, the choice of the period of the solar cycle, and the used tools and algorithms.

4.1 Preparation of The Transactional Dataset

A transactional dataset contains a transaction as a register of the dataset. This kind of dataset is used when the association task is applied to data to discover association rules. The transaction contains its identification and all its attributes. In the case of solar flares, a transaction can be explained as an occurrence of a solar flare in one specific solar region.

Two main problems were found when creating a transactional dataset for solar region events: one problem regarding data storage and another regarding data format. With respect to the storage problem, it was necessary to consider that the solar data are distributed into several datasets provided by the SWPC [15,23]. Data used in the current work were extracted from two of these datasets: Solar Region Summary (SRS) and Solar and Geophysical Activity Summary (SGAS). Information of how to download data from SRS and SGAS datasets can be found in [24,25].

From the SRS data collection, the following attributes were used: region identification, the sunspot area, and Mt. Wilson magnetic classification. From the SGAS data collection, attributes used were: X-ray data flux and radio flux.

From these two datasets - SGAS and SRS -, it was necessary to merge all chosen attributes into a single dataset. It is important to point out that for one region in the SRS dataset, there can be several related events in the SGAS dataset. This can be noticed in the second and third lines of Table 2, which shows the merged dataset. Transactions 2 and 3 relate to the same region, but have different X-ray classifications. Therefore, these two events were considered as different transactions in the resulting dataset.

Trans.ID	Date (Y M D)			Reg. ID	Spot Area (Millionths of the Solar Disk Area)	Magnetic Classification (Mt. Wilson)	X-ray flux (W/m^2) (in the wavelength range of 1 to 8 Ångströms)	Radio flux (10.7cm)
1	2000	3	3	8898	10	Alfa	$3.7.10^{-5}$	220
2	2000	3	4	8882	850	beta-gamma	$3.8.10^{-5}$	240
3	2000	3	4	8882	850	beta-gamma	$6.1.10^{-6}$	200

TABLE 2: Example of transactions resulting from the merging of SGAS and SRS datasets.

Besides the data storage problem, another problem faced when creating the transactional dataset was the data format, as solar flares data are presented into diverse formats. Because the APRIORI algorithm demands categorical data as input, it was necessary to preprocess all numerical data, transforming them into categorical data.

As it can be seen in Table 2, the parameters used in the experiments were: sunspot area, magnetic classification of the event, X-ray flux and radio flux. The only categorical parameter is the magnetic classification. The other three parameters (spot area, X-ray and radio fluxes) are numerical. To use them in the APRIORI algorithm, it was necessary to have them transformed into categorical attributes. The attributes were discretized according to the classes showed in Tables 1, 3 and 4. For the X-ray flux, in Watt per square meter, which is measured from the wavelength band of 1 to 8 Ångström, the categorization presented in Table 1 was used [18].

For the background radio flux, a categorization was created according to the flux values observed during solar cycles 23 and 24 [15]. Being quantified in “solar flow unit” (sfu), F10.7 radio flux can range from 50 sfu–300 sfu over a solar cycle. It is considered that the annual mean value of 10.7 cm radio flux larger than 180 sfu corresponds to the solar maximum year, while the annual mean value of 10.7 cm radio flux smaller than 90 sfu corresponds to the solar minimum year [22]. Considering these values, four levels of flux were established: “low”, “medium”, “high”, and “very high” as shown in Table 3.

Radio flux (10.7 cm frequency)	**Category**
< 80	Low
80 a 120	Medium
120 a 160	High
> 160	Very high

TABLE 3: Radio Flux Categorization.

For the sunspot area, a categorization was created according to Table 4, following the same way of dividing data into four classes.

Spot area (Millionths of the Solar Disk Area - MADS)	**Category**
0 a 200	Small
200 a 500	Medium
500 a 1000	Large
> 1000	Huge

TABLE 4: Categorization of the sunspot area.

4.2 Selection of The Solar Cycles’ Periods

As explained in Section 1, the solar cycle has an average period of eleven years. This eleven-year period has a maximum phase when most of the flares normally occur. Cycles usually have similar behavior, however, there may be differences in events’ behavior between cycles. Considering this fact, three experiments were carried out according to three different periods in order to compare the resulting rules.

For the first experiment, all the data from 1997 to 2016 were used. The main goal of this experiment was to evaluate the association rules that would be generated considering two whole solar cycles (23rd and 24th solar cycles), independently of the period of the solar cycle.

The second and third experiments used specific periods of the 23rd and the 24th solar cycles, respectively. To do so, the most active parts of the cycles were selected. The most active parts are those which show the highest number of flares of classes M and X and are, generally, concentrated between the year of the maximum and the following years [26]. For the 23rd solar cycle, the period from the year 2000 to the year 2003 was used. For the 24th solar cycle, the period from the year 2012 to the year 2015 was considered. Those periods presented the highest levels of solar events, therefore presenting the highest number of M and X solar flares. The goal of the last two experiments was to compare the two sets of rules between the two cycles.

4.3 Tools, Algorithms and Metrics

The following tools, algorithms and metrics were applied to the transactional dataset showed in Table 2:

- The APRIORI algorithm was applied to all datasets used in the experiments of Section 5, and libraries of the R framework – a free software environment for statistical computing and graphics creation [27] – were used to apply the APRIORI algorithm;

- For data pre-processing and dataset joining, functions were developed using JAVA language;

- The association rules generated in each set of experiments were analyzed according to the support and confidence metrics.

5. RESULTS AND DISCUSSION

As explained in Section 4.2, three experiments were performed according to different periods. The attributes used in all experiments were the same. The attributes used, as explained in Section 4.1, were:

- Classification of the spot area in Millionths of the Solar Disk Area;
- Magnetic classification of the spot by the Mt. Wilson classification;
- Classification of solar flares by the measured flux in X-rays in the band from 1 to 8 Ångströms;
- Classification of solar flares by the radio flux emission at a wavelength of 10.7 cm.

The results of the experiments were described based on the resulting association rules. The rules that presented X or M X-ray classification in their consequent part were selected for analysis. Consider, for example, the rule: {area_huge,beta_gamma_delta,radio_veryhigh}→{Xray_X}. In this example, the antecedent of the rule is a set of three parameters and the consequent, a set of one parameter. The rule can be read as: when a solar explosion of intensity X occurs (X_ray_X), the value of the parameter area is bigger than 1000 MADS (area_huge), the sunspot classification is beta_gamma_delta and radio emission is higher than 160 sfu (radio_veryhigh) (categorizations of the parameters were explained in Section 4.1).

The support and confidence metrics were used to show how often this association occurs and the probability of its occurrence (support and confidence were explained in Section 2.3). The three experiments are explained in the following sections.

5.1 Experiment I: Data From 1997 to 2016

For the first experiment, all the data provided by the SWPC from 1997 to 2016 were used. This data period was chosen to assess the kind of association rules that would be generated considering the solar cycles 23 and 24, independently of the period of the solar cycle. Following is the description of the experiment.

- Period: 1997 to 2016;
- Number of considered regions: 1,414;
- Support and Confidence: we selected rules of classes M and X with the highest values of support and confidence. The minimum support was 1%.

Table 5 shows the rules and their support and confidence. The first rule presents the highest support for the rules containing the class-X flares in the consequent part of the rule. From rule number 1 it is possible to conclude that in almost 10% of the transactions of the dataset, the parameters radio-veryhigh and Xray_X appear together in the transactions. The confidence of rule number 1 shows that in about 26% of the rules in which Xray_X is the consequent of the rule, the radio_veryhigh is the antecedent. Rule number 2 presents the highest confidence for the rules containing the class-X flares in the consequent part of the rule.

Rule number 3 presents the highest support for the rules containing class-M flares in the consequent part of the rule. Rule number 4 presents the highest confidence for class-M flares in the consequent part of the rule.

#	Support	Confidence	Association rule
1	9.83%	25.78%	*{radio_veryhigh} → {Xray_X}*
2	2.26%	43.24%	*{area_huge, beta_gamma_delta, radio_veryhigh} → {Xray_X}*
3	23.97%	62.89%	*{radio_veryhigh} → {Xray_M}*
4	2.47%	83.33%	*{beta_gamma_delta, radio_high} → {Xray_M}*

TABLE 5: Selected rules for M/X flares for the 1997–2016 period.

5.2 Experiment II: Data From The 23rd Solar Cycle

For the second experiment, data from the 23rd solar cycle were considered. The goal was to assess the rules generated by events that occurred in the period between 2000 and 2003 which presented the highest number of solar activities in the23rd solar cycle. The second experiment can be described as follows.

- Period: 2000 to 2003;
- Number of considered regions: 665;
- Support and Confidence: we selected rules of classes M and X with the highest values of support and confidence. The minimum support was 1%.

Table 6 shows the rules and their support and confidence. Similarly to Table 5, in Table 6, the first rule presents the highest support for rules containing class-X flares in the consequent part of the rule. From rule number 1, it is possible to conclude that in almost 8.5% of the transactions of the dataset, the attributes radio_veryhigh and Xray_X appear together in the transactions. The confidence of rule number 1 shows that in almost 26% of the rules in which Xray_X is the consequent of the rule, the radio_veryhigh is the antecedent.

#	Support	Confidence	Association rule
1	8.42%	25.45%	*{ radio_veryhigh} → {Xray_X}*
2	1.95%	43.33%	*{area_huge, beta_gamma_delta, radio_veryhigh} → {Xray_X}*
3	27.21%	60.53%	*{radio_low} → {Xray_M}*
4	1.20%	100%	*{area_medium, beta_gamma_delta, radio_medium} → {Xray_M}*

TABLE 6: Selected rules for flares of classes M and X for the 2000–2003 period.

Rule number 2 presents the highest confidence for the all rules which present class-X flares in the consequent part of the rule. Rule number 3 presents the highest support for the rules containing class-M flares in the consequent part of the rule. Rule number 4 presents the highest confidence for the rules containing the M type of flare in the consequent part of the rule.

5.3 Experiment III: Data From The 24th Solar Cycle

For the third experiment, data from the 24th solar cycle were considered. The objective was to assess the rules generated by events that occurred between 2012 and 2015. This period presented the highest number of solar activities for the 24rd solar cycle. Following is the description of the experiment.

- Period: 2012 to 2015;
- Number of regions considered: 239;
- Support and Confidence: we selected rules of classes M and X with the highest values of support and confidence. The minimum support was 1%.

Table 7 shows the rules and measures of support and confidence for flares of classes M and X.

#	Support	Confidence	Association rule
1	13.80%	27.96%	*{ radio_veryhigh} → {Xray_X}*
2	2.09%	62.50%	*{beta_delta, radio_veryhigh} → {Xray_X}*
3	35.14%	72.41%	*{beta_gamma_delta} → {Xray_M}*
4	2.09%	100%	*{area_huge, beta_gamma_delta, radio_low} → {Xray_M}*

TABLE 7: Selected rules for flares M and X for the 2012–2015 period.

Similarly to Table 6, in Table 7 the first rule presents the highest support for the rules containing the class-X flare in the consequent part of the rule. From rule number 1, it is possible to conclude that in almost 14% of the transactions of the dataset, the parameters radio_veryhigh and Xray_X appear together in the transactions.The confidence value of rule number 1 shows that in almost 28% of the rules that have Xray_X as the consequent of the rule, radio_veryhigh is the antecedent.

Rule number 2 presents the highest confidence for the rules containing the X type of flare in the consequent part of the rule. Rule number 3 presents the highest support for the rules containing the class-M flare in the consequent part of the rule, while rule number 4 presents the highest confidence for the rules containing the class-M flare in the consequent part of the rule.

5.4 Comparison of Results

Tables 8 and 9 show a comparison among the generated rules for the three experiments. Table 8 shows all the rules generated with Xray_X in the consequent. Table 9 shows all the rules generated with Xray_M in the consequent.

In Table 8, rule 1 {radio_veryhigh}→{Xray_X} appears in all three experiments as the rule with the highest support level, indicating that, among all the other rules with Xray_X in the consequent, this rule is the most frequent one. For rules 1, 3, and 5, the antecedent is the radio_veryhigh, indicating that these two attributes have the highest association in the dataset. Although their support values are the highest, their confidence remained at around 26%, which indicates that in 26% of the cases, when Xray_X appears in the consequent, radio_veryhigh appears in the antecedent.

Rules 2 and 4 agree that when the X_rayX is the consequent, there is a probability of around 43% that the antecedent part of the rule is formed by the attributes: area_huge, beta_gamma_delta and radio_veryhigh.

Despite not being very frequent, rule 6 shows the highest confidence level, indicating that the probability of an X solar flare is 62.5% when beta_delta and radio_veryhigh appear together. In all rules the attribute radio_veryhigh is present. The area attribute in two rules always assumes huge value. The attribute radio_very_high is present in all rules.

#	Period	Supp.	Conf.	Association rule
1	1997–2016	9.83%	25.78%	*{radio_veryhigh}→{Xray_X}*
2	1997–2016	2.26%	43.24%	*{area_huge, beta_gamma_delta, radio_veryhigh}→{Xray_X}*
3	2000–2003	8.42%	25.45%	*{radio_veryhigh}→{Xray_X}*
4	2000–2003	1.95%	43.33%	*{area_huge, beta_gamma_delta, radio_veryhigh}→{Xray_X}*
5	2012–2015	13.80%	27.96%	*{radio_veryhigh}→{Xray_X}*
6	2012–2015	2.09%	62.5%	*{beta_delta, radio_veryhigh}→{Xray_X}*

TABLE 8: Comparison of rules with *Xray_X* in the consequent.

Table 9 presents rules that contain Xray_M in the consequent. These rules present an interesting behavior, showing relatively low support, but high confidence, particularly rules 2, 4 and 6. This kind of rule can represent an interesting relation for not being so frequent, but very reliable. This means that when it occurs, there is a high possibility of the association among the attributes of the rule. Another characteristic of these set of rules is that most of them (2, 4, 5 and 6) contain the attribute beta_gamma_delta.

#	Period	Supp.	Conf.	Association rule
1	1997–2016	23.97%	62.9%	*{radio_veryhigh}→{Xray_M}*
2	1997–2016	2.47%	83.3%	*{beta_gamma_delta, radio_high}→{Xray_M}*
3	2000–2003	27.21%	60.5%	*{radio_low}→{Xray_M}*
4	2000–2003	1.20%	100%	*{area_medium, beta_gamma_delta, radio_medium}→{Xray_M}*
5	2012–2015	35.14%	72.4%	*{beta_gamma_delta}→{Xray_M}*
6	2012–2015	2.09%	100%	*{area_huge, beta_gamma_delta, radio_low}→{Xray_M}*

TABLE 9: Rules with *Xray_M* in the consequent.

If a confidence higher than 80% was considered here, only the rules 2, 4 and 6 would be present. This would imply the presence of the attribute beta_gamma_delta in all rules. If we had a confidence of 100%, only rules 4 and 6 would be considered. For these two rules, the radio attribute assumes medium and low values, respectively. This is an interesting information to be further investigated as, it is well known that, the beta_gamma_delta classification is, generally, present in solar flares of type X. However, when the value of the radio emission is medium or low, even in the presence of a beta_gamma_delta sunspot, the solar flare was of M class.

6. CONCLUSIONS AND FUTURE WORKS

This article focused on the study of M/X solar flares and the relationship among their parameters. The parameters here analyzed were: classification of the sunspot area; magnetic classification of the sunspot; classification of solar flares by the measured flux in X-rays, and classification of solar flares by the radio flux emission. Support and confidence metrics were used to analyze all experiments.

Three periods were chosen: 1997–2016; 2000–2003 and 2012–2015. The rules that presented M or X classification in their consequent part were selected for analysis. Among those rules, the ones that presented higher values of support and confidence were filtered and compared for the three-period experiments.

For all experiments, regardless the period of the solar cycle, authors found association rules

relating high radio flux emission (> 160 sfu) with class- X solar flare. This indicates the importance of high radio emission in the occurrence of class-X solar flares.

Regarding the sunspot classification, only the betta_gamma_delta and beta_delta classifications were present in the rules indicating class-X solar flares. Those rules presented the highest values of confidence, indicating the importance of those sunspots' classifications in class-X solar flares. For the spot area, only huge areas were found in the most promising rules of class-X solar flares. However, the rules presented low confidence. Hence, it is only possible to point out that spots with huge areas appeared in class-X solar flare association rules.

For the rules with class-M solar flares in the consequent, all filtered rules presented a level of confidence much higher than rules of class-X solar flares, including two rules with confidence of 100%. One of the differences among rules of M and X solar flares was the radio flux emission. While in the class-X solar flares rules only high emissions of radio appeared, in the class-M solar flare rules, all values of radio flux emissions appeared (low, medium, high and very high). This indicates that it was not possible to point out a value for radio flux and class-M solar flares. However, it is clear that when class-M solar flares occur, radio emission is always found.

In four of the five rules of class-M solar flares, the beta-gamma-delta sunspot classification was found, which reinforces the importance of this sunspot classification for the analysis of class-M solar flares.

The sunspot area parameter was found in two of the five rules with different values (huge and medium). Therefore, it is not possible to observe a very consistent behavior of the spot area parameter for class-M solar flares association rules.

Considering all results, it was possible to establish some relations among the proposed parameters and solar flares of classes M and X, regardless the selected period. It was noticeable that, by using association rules, it was possible to synthesize the relations of frequently used parameters of solar flares of classes M and X.

As mentioned in Section 3, works which used association rules in order to evaluate relations among solar flare parameters were not found. However, a recent work [28], mentions the use of a spatiotemporal association rule mining algorithm in solar images. This kind of association rule involves, besides the parameters´ values of solar images, their time and space properties, which means that, these are association rules which contain parameters´ values as well as the period and the solar region where their occurred. Authors obtained solar images and from them extracted parameters which were submitted to the spatiotemporal association rule mining algorithm. Results showed that the tool would help the understanding of solar properties by the extraction of solar images and the use of temporal association rules in order to evaluate the evolution of the images.

In future studies, authors intend to create a method to study the temporal behavior of the solar cycle. The intention is to use an association rule algorithm which also considered time as an attribute of the rule. The idea is to show, for each part of the solar cycle, the most relevant association rules which could show relations among parameters that resulted in M and X solar flares. This study is similar to [28] in terms of the use of temporal association rules. However, instead of considering solar images and their related descriptions, the magnetic parameters of the active regions of the Sun would be considered. An example of the use of magnetic parameters of active regions in order to predict solar flares can be found in [29].

7. ACKNOWLEDGEMENTS

The authors would like to acknowledge the financial support for this research through the grant #2017/14836-4 received from the São Paulo Research Foundation (FAPESP).

The authors thank Espaço da Escrita – Pró-Reitoria de Pesquisa, UNICAMP – for the provided language services.

8. REFERENCES

[1] T. Colak, R. Qahwaji. “Automated Solar Activity Prediction: A hybrid computer platform using machine learning and solar imaging for automated prediction of solar flares”. Space Weather. vol. 7, pp 1-12, Jun. 2009.

[2] K. Fox. “Impacts of Strong Solar Flares. 2013”. Internet: https://www.nasa.gov/mission_pages/sunearth/news/flare-impacts.html#.WAGF2uArLIV. [Sep. 1, 2019].

[3] M. Dierckxsens, K. Tziotziou, S. Dalla, I. Patsou, M. S. Marsh, N. B. Crosby, O. Malandraki, G. Tsiropoula. “Relationship between Solar Energetic Particles and Properties of Flares and CMEs: Statistical Analysis of Solar Cycle 23 Events”. Solar Physics. vol. 290(3), pp.841-874, Mar. 2015.

[4] C. Yanmei, W. Huaning. “Correlation between solar flare productivity and photospheric vector magnetic fields”. Advances in Space Research. vol. 42 (9), pp. 1475–1479, Nov. 2018.

[5] G. Meera, R. Munika, A. K. Shrivastava. “Various Solar Activity Parameters and their Interrelationship from Solar cycles 20 to 24”. International Research Journal of Science and Engineering. vol. 5 (5), pp 59-69. May. 2017.

[6] K. D. Leka, G. Barnes. “Solar Flare Forecasting: Present Methods and Challenges” in Extreme Events in Geospace, 1st ed., vol. 1. N. Buzulukova, Ed. Elsevier, 2017. pp 65-98.

[7] J. Han, M. Kamber, J. Pei. “Data Mining: Concepts and Techniques”. 3rd. ed. Morgan Kaufmann, 2012.

[8] G. Meera, V. K. Mishra, A. P. Mishra. “Solar activity parameters and their interrelationship: Continuous decrease in flare activity from solar cycles 20 to 23”. Journal of Geophysical Research, vol. 112. pp 1-10. May, 2007.

[9] R. Agrawal, R. Srikant. “Fast algorithms for mining association rules in large databases,” in Proc VLDB, 1994, pp 487-499.

[10] A. Loskutov, I. A. Istomin, K. M. Kuzanyan, O. L. Kotlyarov. “Testing and forecasting the time series of the solar activity by singular spectrum analysis.” arXiv preprint – ArXiv, nlin/0010027. [Online] Available: https://arxiv.org/abs/nlin/0010027 [Nov. 24, 2019].

[11] F. Zuccarello. “Multi-spectral observations of flares.” Astronomische Nachrichten, vol.337(10), pp.1070. Nov. 2016.

[12] R. Qahwaji, T. Colak. “Automatic Prediction of Solar Flares using Machine Learning: Practical Study on the Halloween Storm,” in Proc. RAST 2007. pp. 739-742.

[13] A. Ajabshirizadeh, N. M. Jouzdani, S. Abbassi. “Neural network prediction of solar cycle 24”. Research in Astronomy and Astrophysics, vol. 11(4), p. 491-496, 2011.

[14] S. Sello, “Time Series Forecasting: A Nonlinear Dynamics Approach. Termo-Fluid Dynamics Research Center.” arXiv preprint – ArXiv, physics/9906035. [Online] Available: https://arxiv.org/abs/physics/9906035 [Nov. 24, 2019].

[15] Space Weather Prediction Center: NOAA / NWS Space Weather Prediction Center. 2015. Internet: http://www.swpc.noaa.gov. [Aug. 17, 2019].

[16] R. Qahwaji, T. Colak. "Automated Prediction of Solar Flares Using Neural Networks and Sunspots Associations". Advances in Soft Computing. vol 39. pp. 316 – 324. Jun. 2007.

[17] T. T. Nguyen, C. P. Willis, D. J. Paddon, H. S. Nguyen. "On Learning of Sunspot Classification," in Intelligent Information Processing and Web Mining, 1st ed., vol. 25. M. A. Kłopotek, S. T. Wierzchoń, K. Trojanowski (eds). Berlin, Heidelberg: Springer. 2004. pp 59-68.

[18] C. Guo B. Xue, Z. Lin. "Study on the Prediction Method of Characteristic Parameters of Solar X-ray Flares". Chinese Astronomy and Astrophysics. vol 37(3). pp. 255–265, Jul. 2013;

[19] P. Tan, M. Steinbach, V. Kumar, V. "Association Analysis: Basic Concepts and Algorithms," in Introduction to Data Mining. P. Tan, M. Steinbach, V. Kumar (eds). London:Pearson, 2005. pp. 328-414.

[20] C. Basu, M. Padmanaban, S. Guillon, L. Cauchon, M. DeMontigny, I. Kamwa. "Association rule mining to understand GMDs and their effects on power systems," in Proc. IEEE Power and Energy Society General Meeting, 2016, pp.1-6.

[21] S. Feng, L; Yu, Y. Yang. "The relationship between grouped solar flares and sunspot activity," in Proc. Bull. Astr. Soc. India, 2013, pp. 237–246.

[22] L. Shen, J. Dun, X. Zhang, Y. Jiang. " Analysis of the Major Parameters in Solar Active Regions Based on the PCA Method". Chinese Astronomy and Astrophysics. vol. 39. pp 212 – 224. Jul. 2015.

[23] SWPC. Data Access: SWPC Data Service. 2017. Internet: http://www.swpc.noaa.gov/content/data-access. [Sep. 10, 2019].

[24] SRS Solar Region Summary. Internet: ftp://ftp.swpc.noaa.gov/pub/forecasts/SRS/README. [Aug. 10, 2010].

[25] SGAS Solar and Geophysical Activity Summary. Internet: ftp://ftp.swpc.noaa.gov/pub/forecasts/SGAS/README. [Aug. 10, 2019].

[26] X. Yang, G. Le, C. Zhang, Z. Yin, W. Zhao. "A Statistical Analysis of X1- and X2- or Higher-Class Flares during Solar Cycles 21∼23". Chinese Astronomy and Astrophysics. vol. 38, pp. 92 –99. Jun. 2014.

[27] THE R FOUNDATION. The R Project for Statistical Computing. Internet: https://www.r-project.org. [Aug. 12, 2019].

[28] Silveira C.R., Cecatto J.R., Santos M.T.P., Ribeiro M.X. (2018) Thematic Spatiotemporal Association Rules to Track the Evolving of Visual Features and Their Meaning in Satellite Image Time Series. In: Latifi S. (eds) Information Technology - New Generations. Advances in Intelligent Systems and Computing, vol 738. Springer, Cham.

[29] Liu, Chang, Deng, Na, Wang, Haimin, and Wang, Jason T. L. "Predicting Solar Flares Using SDO /HMI Vector Magnetic Data Products and the Random Forest Algorithm". United States: N. p., 2017.