Published by SERC of Kyushu University
ISWI Newsletter
under the auspices of the United Nations
This pdf circulated in Volume 4, Number 43, On 20 April 2012.

Application of Empirical Mode Decomposition to the Treatment of SID Monitor Data

Giancarlo Fissore and Federico Belliardo

Hundreds of schools around the world, coordinated by Stanford's Solar Center, are engaged in hunting solar flares using a low cost device called a SID Monitor.[1] Our school, Liceo Scientifico-Tecnologico Vallauri in Fossano, Italy, is one of those (seen as Italy-10 in the SID Stanford database). SID stands for "sudden ionospheric disturbance" and it is such disturbances that may be detected using the SID Monitor.

As soon as we started our activity, we noticed that hunting solar flares could be quite difficult due to the interference that is always present in the signal to one degree or another.

Facing this problem, we found a method to remove background noise and other interference from SID Monitor's data by applying Empirical Mode Decomposition (EMD).

Figure 1 – A Stanford SID Monitor.

Thanks to the results obtained in our research, we were able to write a program, called SidDataAnalyzer, which decomposes the SID monitor signal, "cleans" it (removes noise), and produces a variety of graphical representations of our data superimposed on graphs of the solar X-ray flux measured by NASA satellites. We believe that our product can be of great help to all of the people who work with Stanford SID Monitors and that is why we have made our software available for everyone interested (see hyperlink below).

Background

The Sun is not a quiet star. Every now and then it produces a burst of energy called a solar flare. When this radiation reaches Earth, it disturbs the ionosphere, which in turn causes a Sudden Ionospheric Disturbance (SID) because the condition of the ionosphere effects the propagation of radio signals.

The SID Monitor is a device that records the intensity of a signal coming from a very low frequency (VLF) radio station which is normally employed in communication with submarines.

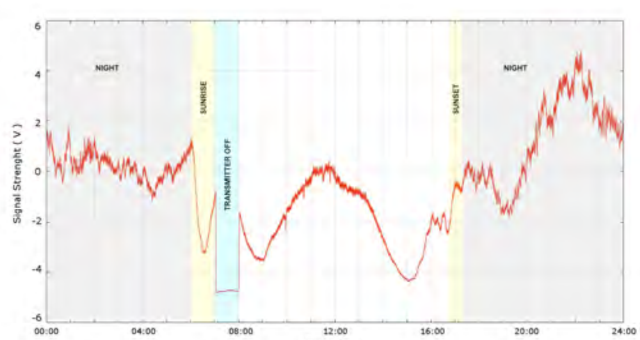


Figure 2 – Quiet day.

Because of the changing structure of the ionosphere, only the data registered during the day are consistent (during night-time the monitor goes into saturation). The grey background in Fig. 2 highlights night-time data, while the yellow background highlights the transistion between day and night (sunrise and sunset). Moreover, the light-blue background highlights the time of the day during which the transmitter is off for maintenance. No data are available for that time.

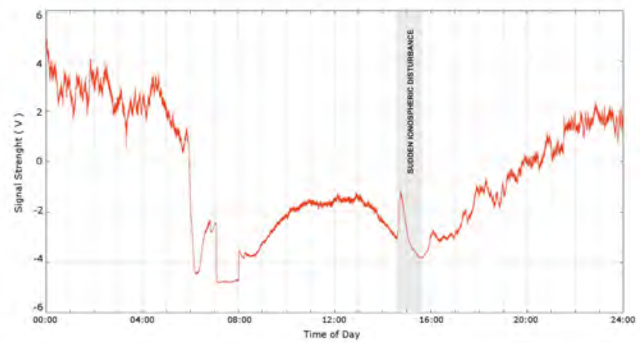


Figure 3 – Day with solar flare.

Daytime data are described by a curve that increases in the mornig, has its maximum approximately at noon, and decreases in the afternoon (white background in Fig. 2). When a solar flare occurs, a somewhat pronounced peak pops up on the graph as shown in Fig. 3.

Problem and Preliminary Research

The signals recorded by the SID monitor are often very disturbed by background noise and other interference, therefore weak flares can be difficult to identify. We wanted to make the search for solar flares easier and more reliable, so we started our research activity.

Studying the scientific literature about signal processing, we found an interesting article written by Lionel Loudet, curator of SID monitoring station A118, in which it was shown that EMD could be used to decompose SID Monitor's signals into different components.[2,3] That enabled us to devise a method to tackle our problem: if SID monitor's data are analyzed using the EMD, then the components related to the background noise and other disturbances could be separated from the signal and eliminated.

We developed our research on the basis of this hypothesis.

Introduction to Empirical Mode Decomposition

Data derived from natural phenomena, such as those generated by the SID monitors, are generally non-stationary ("stationary" means that the signal does not vary in frequency and amplitude). This characteristic does not recommend the use of Fourier transform for signal processing.

The ability to analyze non-stationary signals is the greatest strength of the Hilbert-Huang Transform, developed by Norden E. Huang, a scientist at NASA Goddard Space Flight Center. This method is based on EMD, an algorithm that breaks any data set into a finite number of functions, called "Intrinsic Mode Functions (IMFs).

These are the conditions that define an IMF:

- In the whole data set, the number of extrema and the number of zero crossings must either be equal or differ by at most by one; and,
- At any point, the mean value of the envelope defined by the local maxima and minima is zero.

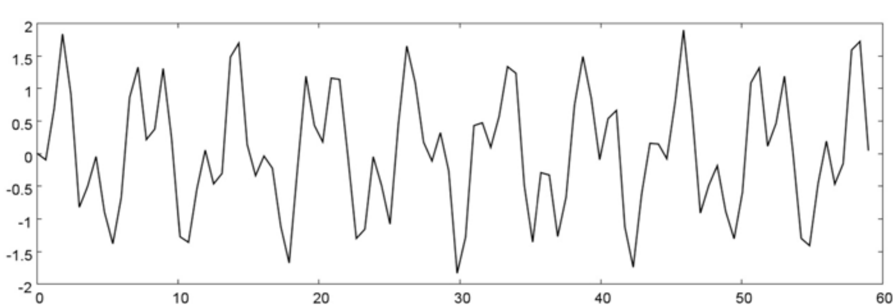


Figure 4 – Signal.

An IMF represents a simple

oscillation mode and plays the role that, in Fourier analysis, is represented by a simple harmonic function. An IMF, however, does not have constant amplitude and frequency, but can have variable amplitude and frequency depending on time. Of course, most of the signals are not IMFs; at each point, in fact, more than one mode of oscillation may be present. It is therefore necessary to introduce a method to decompose any signal into a sum of IMFs.

EMD is an empirical method of signal analysis: it is based on the assumption that any signal consists of different intrinsic oscillating modes. The basic idea for the extraction of an oscillation from these data is that, if an oscillation occurs, it must show a similar behavior to that of a sine function. A cycle of oscillation can start from a local maximum and end at the next local maximum, passing through two zeros and a local minimum.

Let $x(t)$ be an arbitrary signal (Fig. 4), which we assume to have at least one point of maximum and one minimum. The search for the IMFs takes place through a process called "sifting." The first step is to search for the local maxima and minima and generate two functions called upper and lower envelope of $x(t)$, obtained by building two cubic spline functions passing through all points of maximum and minimum, as shown in Fig. 5.

The upper and lower envelopes should contain all the

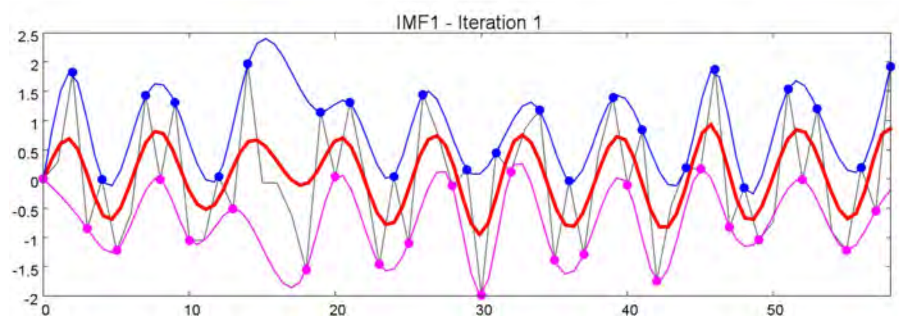


Figure 5 – First iteration of the sifting process (IMF_1).

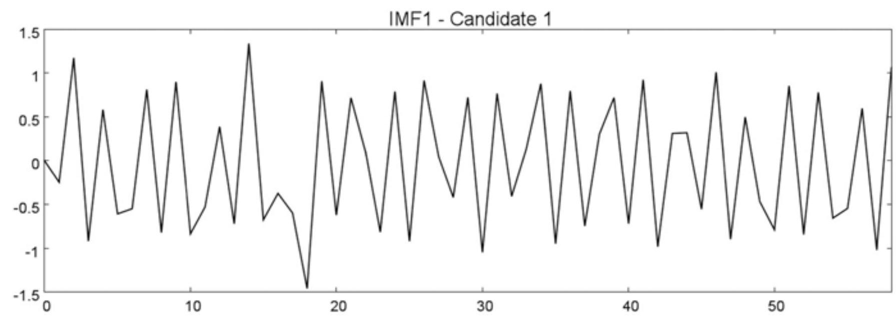


Figure 6 – First approximation of IMF_1; this component does not meet the definition of an IMF.

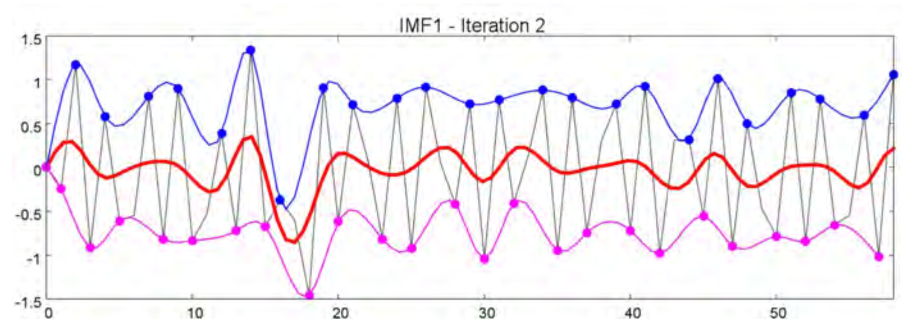


Figure 7 – Second iteration of the sifting process searching for IMF_1.

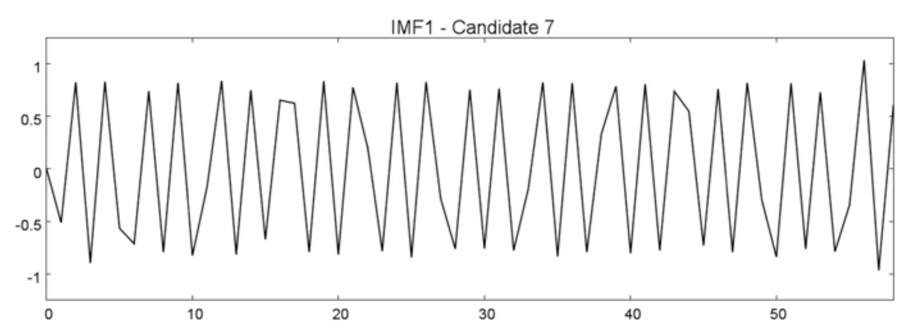


Figure 8 – After seven iterations, the function obtained meets the IMF definition; we have found IMF_1.

data range. After calculating their average and subtracting this mean envelope from the original signal, you get a high frequency oscillating component h_1 which represents the first approximation of the first IMF (IMF_1).

$$h_1 = x(t) - m_1$$

Ideally h_1 should meet the definition of IMF, since the process adopted should have made it symmetrical to a mean zero, with positive maxima and negative minima. In reality, as this is not a rigorous mathematical method, this process has to be repeated several times to obtain an IMF.

The previous procedure must be applied to the first approximation h_1 of IMF_1 (Fig. 6) again in order to get h_2 (Fig. 7).

The new component becomes a better candidate for the determination of IMF_1 and the iterative technique is then reapplied to the new approximation of IMF_1. The process ends when a predefined number of iterations is met or when the criteria that define an IMF are satisfied (that is, the number of extrema and the number of zero crossings must differ at most by one and the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero).

Let's assume that the criteria have been satisfied and the first IMF has been found (Fig. 8). When operating in this way, IMF_1 contains the finer time scale signal component; that is, the component of shortest period. It can therefore be separated from the rest of the data by subtracting it from the original signal:

$$r_1 = x(t) - \mathrm{IMF}_1$$

The residue r_1 still contains changes of lower frequency and may be subject to the same sifting process described above for IMF_1.

At the end of the process, we get the decomposition of the original signal $x(t)$ into n oscillation modes and a residue r_n. This process ends when the n^{th} component or the residue is smaller than a predefined value or when the residue becomes a monotonic function from which no more IMFs can be extracted. The residue is therefore either a mean trend or a constant.

The components of the EMD usually have physical meaning; this means that the characteristic time scales of each component are often the same as those of known physical phenomena, unlike other data analysis methods. For example, the Fourier transform decomposes a signal into an infinite number of harmonics that are, in the end, mathematical artifacts rather than physical entities.

Later in the article we describe how it is possible to identify certain known physical phenomena (in our case, the background noise and the occurrence of a solar flare) by choosing the appropriate IMFs extracted from a signal.

Application of EMD to Our Data

In Fig. 9 are shown the IMFs and the residue obtained applying EMD to the data recorded on a day during which a solar flare occurred, whose sum forms the original data.

Our goal is to find the IMFs that contain the background noise and then reconstruct the signal, adding to the residue all the IMFs that do not represent background noise, obtaining the denoised signal. Moreover, we have to make sure that the IMFs we eliminate from the signal will not alter the signal itself.

In order to reach our objective, we applied EMD to as heterogeneous data as possible (quiet days, very noisy days, days with flares, and so forth) and analyzed the results.

We also studied the effect of solar flares on the IMFs of the signal, in the context of research we presented at Google Science Fair 2011.

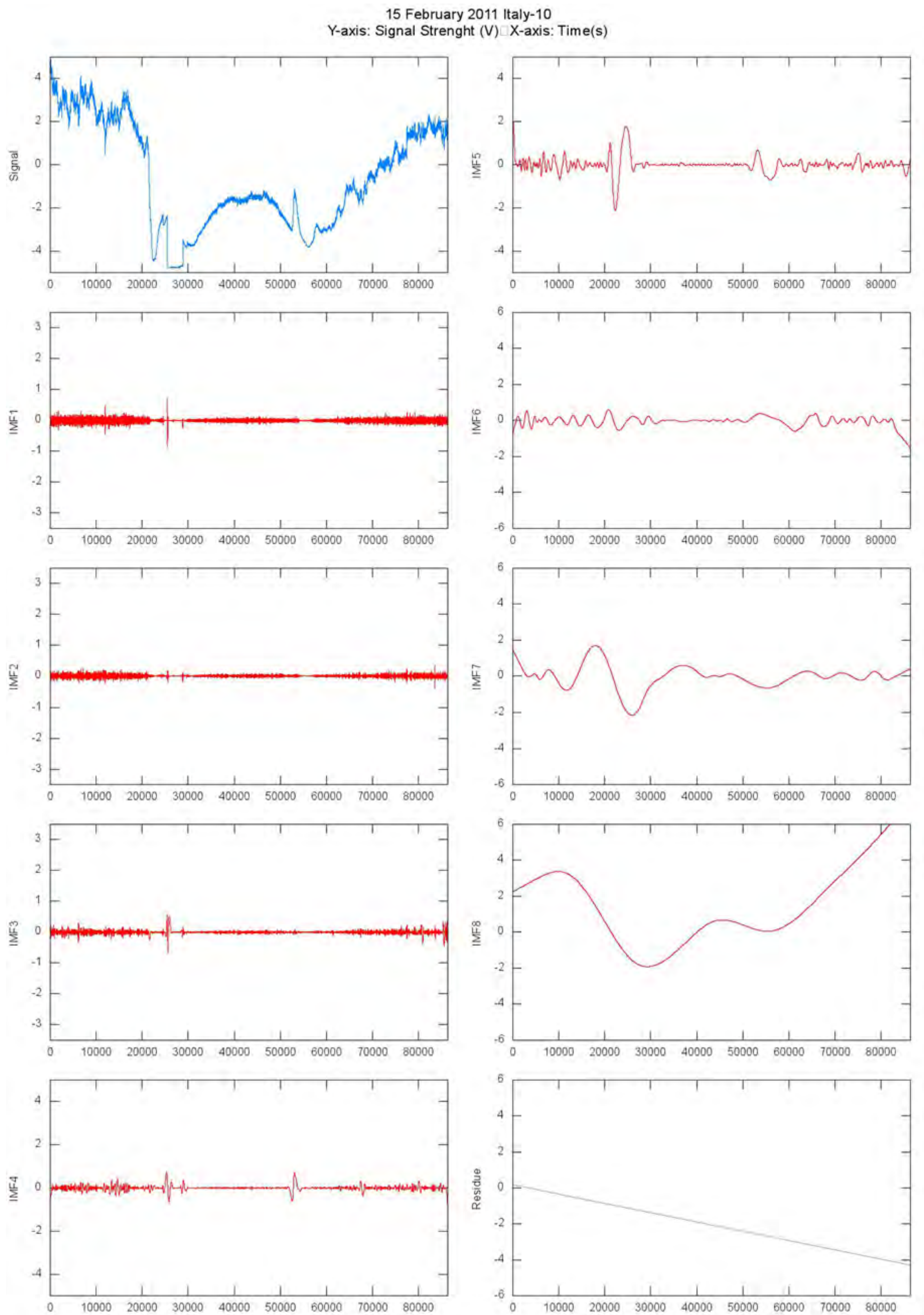


Figure 9 – Decomposition of a day with a SID

Google Science Fair: Research

In order to investigate the impact that a SID has on the various IMFs, we decided (not being able to control the Sun) to implement a procedure to add the characteristic profile of a SID to the typical Pattern of a quiet day. In doing so we tried to produce an alteration of the original signal without lowering the natural background noise.

First, we constructed the mathematical model of a solar flare, based on the analysis of the real flares captured in two years by our SID monitor. Then we developed software that allowed us to add SIDs to real data, controlling the parameters that characterize a solar flare (duration and amplitude). Applying EMD to the data obtained, we could study the effects of SIDs (and their characteristics) on the IMFs.

The whole research is available at: **http://sites.google.com/site/siddataanalyzer/**.

Google Science Fair: Results

Thanks to the results of our research, we were able to understand the relations among the various IMFs and the characteristics of the signal.

Note that the results of the signal decomposition are different when the sampling frequency is changed. The Stanford SID Monitor detects the intensity of the signal every five seconds, but this option can be altered. For example, if the data were collected every 20 seconds, one would get different IMFs in number and appearance. This is why all the comments in this section refer to a sampled signal from a SID Monitor with its default settings.

IMF_1 and IMF_2

These components only reveal very rapid signal changes, such as turning the transmitter on and off.
IMF_1 and IMF_2 do not bear traces of the daily signal variation (night, sunrise, sunset).

IMF_3

Very similar to IMF_1 and IMF_2, but weakly affected by strong flares.

IMF_4

Weakly affected by the patterns of sunrise, sunset and night-time disturbances in the ionosphere.
Whenever there is a flare, IMF_4 is always affected by this disturbance.
The IMF_4 plot is always rather flat during the day. The disturbances caused by flares emerge prominently from the rest of the signal that is always very close to zero.

IMF_5

Shows the profiles of sunrise, sunset and night disturbance of the ionosphere.
This component is always affected by the occurrence of a strong SID.
When the solar flare has a duration of the order of two hours, IMF_5 is the component that most clearly indicates the occurrence of the SID in the daily plot.
The plot of the IMF_5 is always rather wavy and it is not easy to identify a flare as it happens on IMF_4.

IMF_6 and higher

Shows the profiles of sunrise and sunset and the night disturbance of the ionosphere.
These components relate to longer period fluctuations and are associated with daily changes of the signal.
The occurrence of a flare may have effects on the layout of IMF_6, but it is hardly detectable on the subsequent IMFs.
Higher-number components contain only the information related to the passage from day to night.

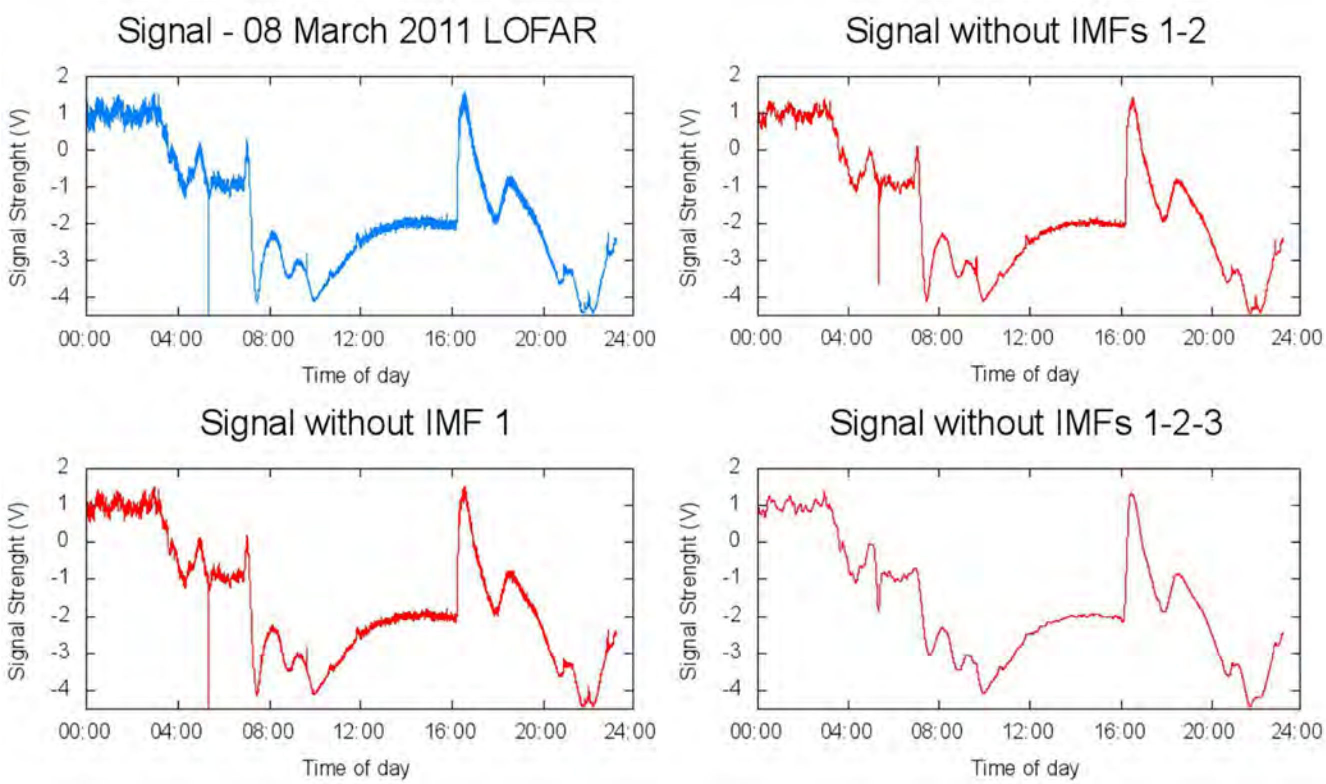


Figure 10 – Stepwise elimination of first IMFs

Denoising

The background noise present in SID monitor's data is characterized by high frequency and low amplitude. For this reason, it influences the first IMFs extracted by applying EMD (because they are the "fastest" components of the signal). Looking at the decomposition in Fig. 9, it is clear how IMF_1, IMF_2, and IMF_3 can be associated with the background noise. That gives us the suggestion that these components should be deleted in order to obtain the denoised signal.

We already noticed that IMF_1 and IMF_2 do not bear traces of the daily signal variation and are not affected by the occurrence of SIDs. The only

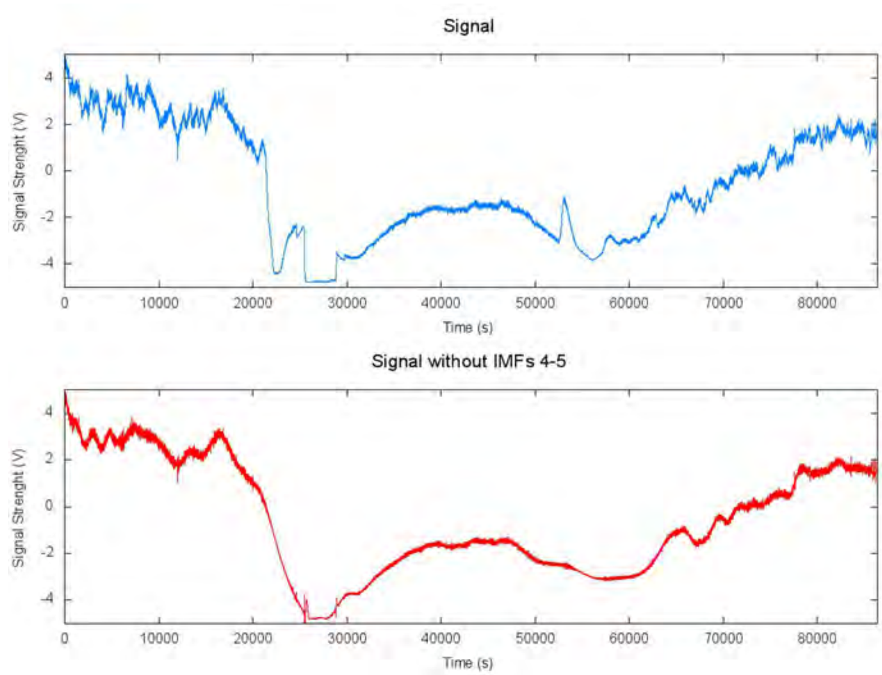


Figure 11 – Elimination of IMF_4 and IMF_5 from the signal, resulting in the deletion of the SID.

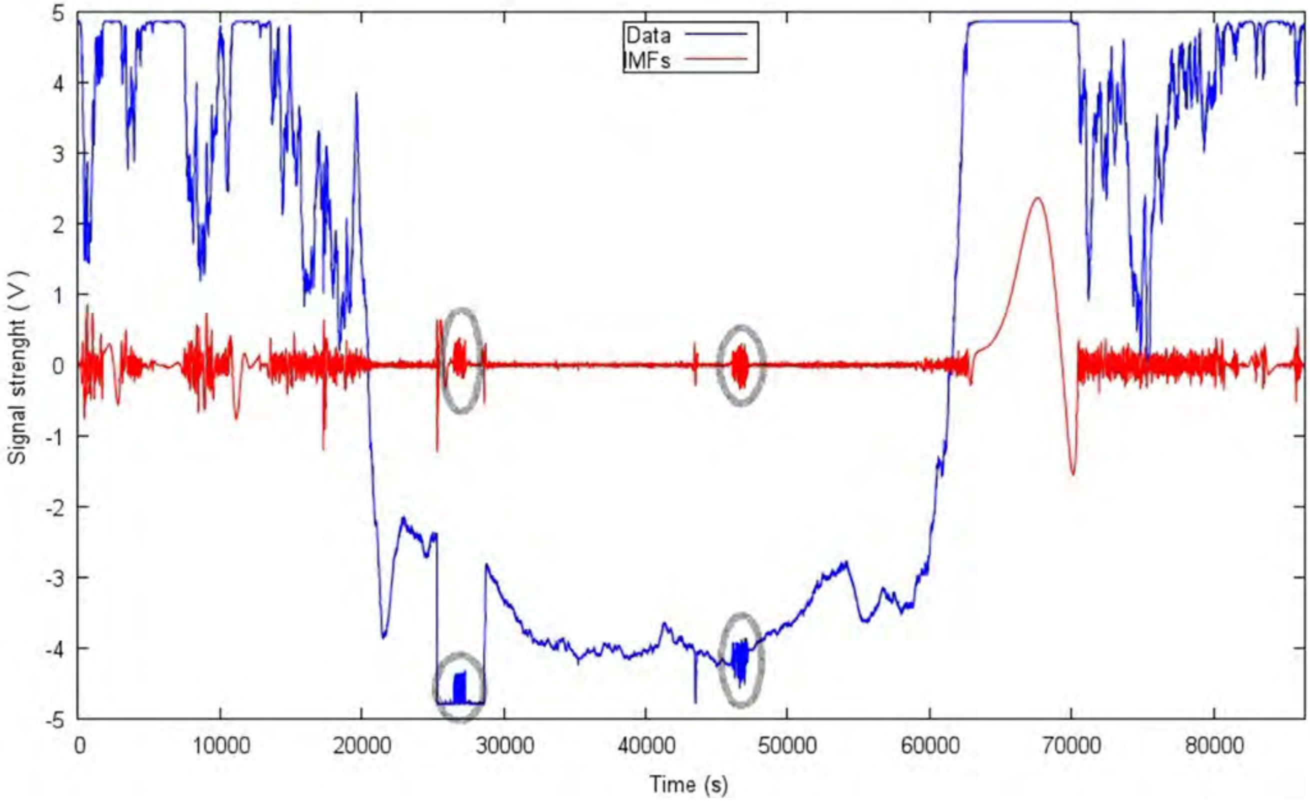


Figure 12 – Original signal (blue line) and the composition of the first three IMFs (red line). Gray circles highlight periods of interference

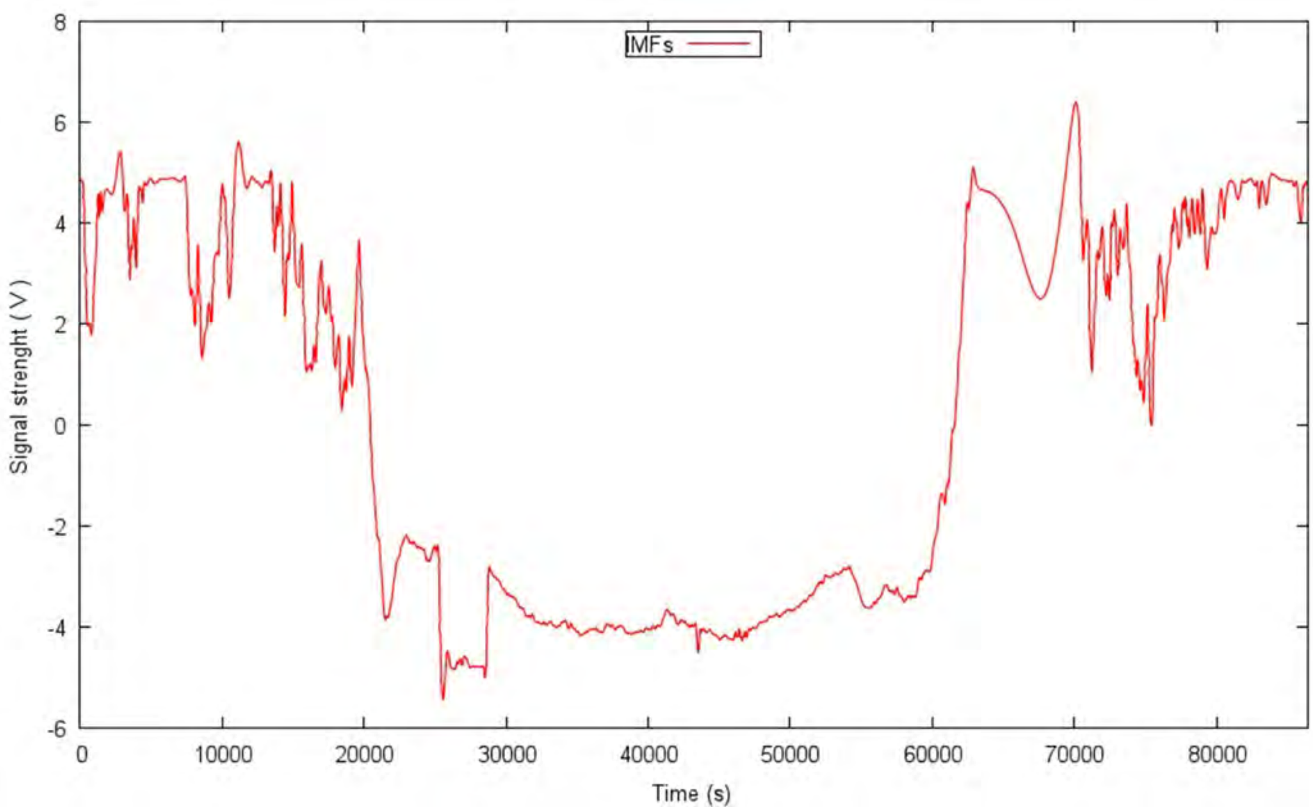


Figure 13 – De-noised signal, without interferences and background noise.

element of the signal associable with these components is the background noise, then we can eliminate them being sure to maintain the integrity of our data (Fig. 10).

Otherwise, IMF_3 can point out the occurrence of a strong flare (M to X class).[4] This kind of SID, however, is better represented in IMF_4 and IMF_5, and we can eliminate the IMF_3 from the signal and maintain the representation of the solar flare in it. That will make the "filter" more aggressive, resulting in an improved denoising.

IMF_4 and following are not influenced by background noise, as shown in Fig. 11. In that figure, the elimination of IMF_4 and IMF_5 (the components that better represent a solar flare) results in the elimination of a SID from the signal without lowering the natural background noise.

Elimination of Disturbances – Example

The United Nations office in Vienna makes use of Stanford's SID Monitor. There, the antenna for the SID Monitor is located near to the antenna of a short wave transmitter. This causes interferences when the transmitter is active, as shown in Fig. 12.

In cases like this, the denoising method based on EMD works very well. In fact, the interference is local and for this reason it doesn't affect the overall signal. In Fig. 12 it is shown how the interference is clearly contained into the first three IMFs. Fig. 13 shows how the elimination of those IMFs will result in the elimination of the interference.

The Software – SidDataAnalyzer

The application of EMD to recorded data is not easily achievable, as there are no tools that can perform this sort of analysis on SID Monitor's data files. For this reason, we developed software called SidDataAnalyzer (SDA), which is fully compatible with SID Monitor's and SuperSID's data.

We tried to keep the software as easy to use as possible. For example, we provided an option to automatically plot denoised data; in this case the complexity of data treatment is totally transparent to the end user and SDA will automatically perform the necessary elaboration.

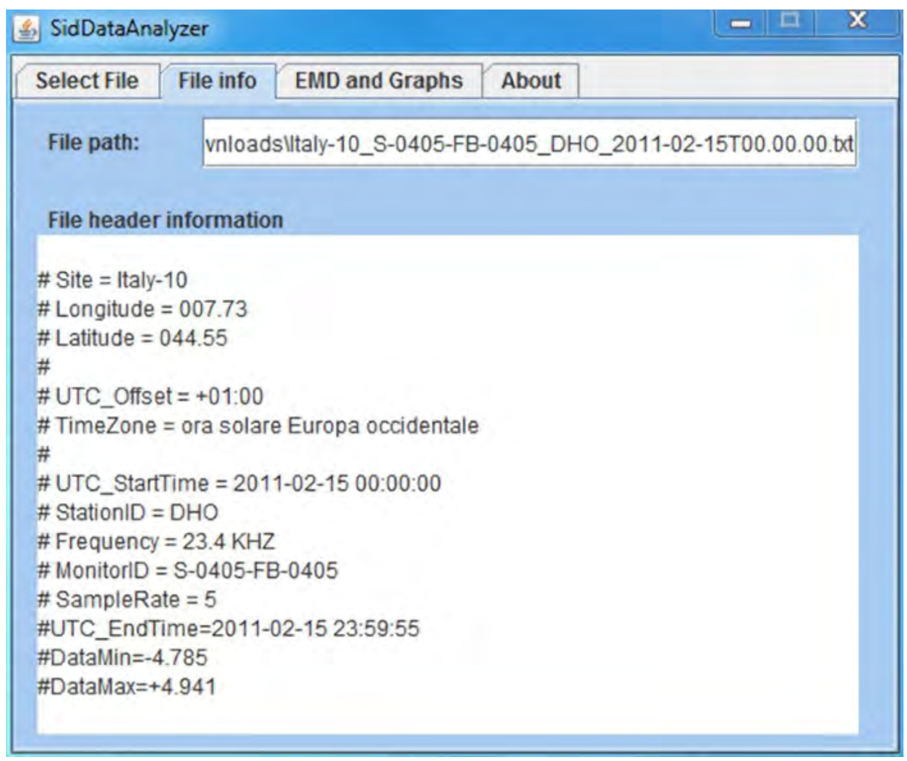


Figure 14 – The SDA Info tab shows information extracted from file headers.

Advanced users, however, can further analyze the IMFs to better understand the structure of the signal. In fact, we provide some useful options to choose and display the various IMFs. For example, it is possible to plot one or more IMFs at a time, superimposed or combined, in order to study a single component (for example IMF_4 alone, as it could contain the flare's spike) or more components together (for example IMF_1, IMF_2, and IMF_3, as they can be equally associated with the background noise).

Another functionality we provided in our software is the possibility to superimpose GOES signals to our graphs. GOES X-Ray flux data, originated from the Sun, are not affected by noise and disturbances generated in the ionosphere, as they are recorded outside our atmosphere. For this reason, GOES signals provide reliable information about strength, length, time, and shape of solar flares that can be easily compared to SIDs recorded by our SID Monitors.

Moreover, to simplify the study of each graph, it is possible to apply different levels

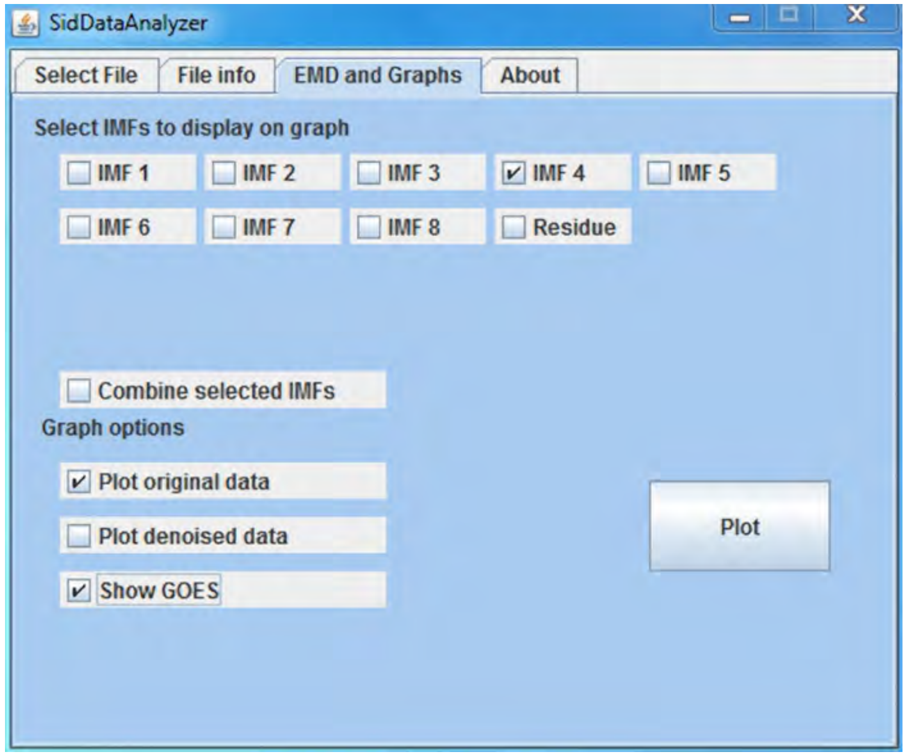


Figure 15 – The EMD and Graphs tab defines plotting options.

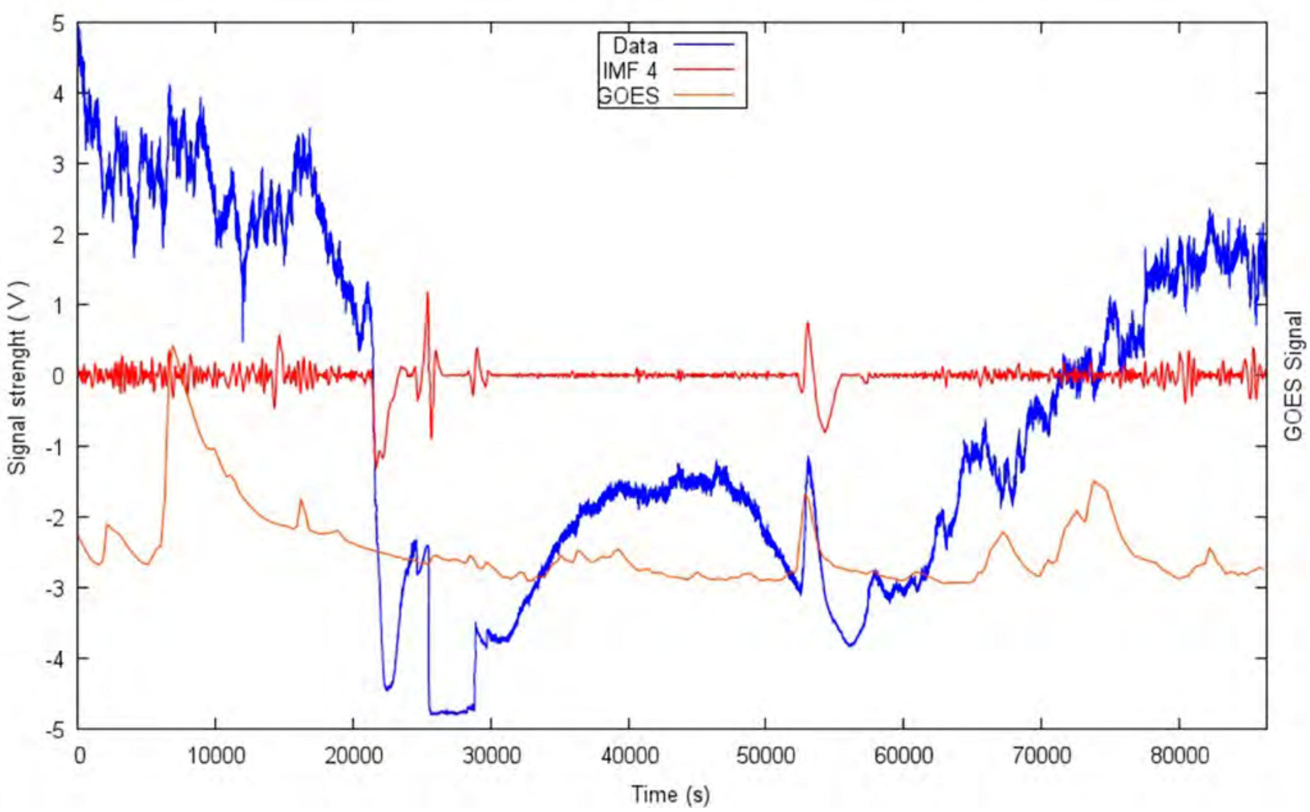


Figure 16 – The plot obtained with the options selected in Fig.15, plotting the signal, IMF_4, and GOES data.

of zoom and the graph's sizes can be easily re-scaled without losing quality.

We believe that our product can be of great help to all of the people that works with Stanford's SID monitors and this is why we released a beta version of SDA available here: **http://doubleslash.altervista.org/sda.html**. Figures 14 through 16 show several screen shots of the SDA in action.

Conclusions and Future Projects

The hypothesis for our research concerned the possibility of using EMD to separate the noise related components of the signal from those due to sudden ionospheric disturbances of solar origin.

In conclusion, we can say that our hypothesis was correct: the signal can be "cleaned up" from background noise by removing IMF_1, IMF_2, and IMF_3 without any loss of information about the overall signal.

We hope that our software, SidDataAnalyzer, will be helpful to the SID Monitor's community.

In the near future, we would like to extend our research to the SuperSID's data that are more heterogeneous than the classic SID Monitor's data. Even if SDA is already compatible with SuperSID's data files, some improvements may be possible and we would like to deepen this topic.

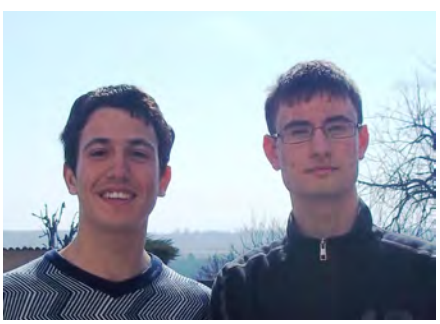
Finally, we would like to thank the Stanford Solar Center for making possible our activity and SARA for giving us the opportunity to write this article.

References

[1] Stanford's SID monitor official website: **http://solar-center.stanford.edu/SID/sidmonitor/**
[2] Application of EMD to the detection of SIDs: **http://sidstation.loudet.org/emd-en.xhtml**
[3] SID monitoring station A118: **http://sidstation.loudet.org/**
[4] Flares classification: **http://spaceweather.com/glossary/flareclasses.html**

Resources for EMD and SID Monitoring

- "SpaceWeather: News and Info about the Sun-Earth Environment": **http://www.spaceweather.com/**
- "Space Weather Monitors", Stanford Solar Center: **http://solar-center.stanford.edu/SID/**
- "On Empirical Mode Decomposition and its Algorithms", Patrick Flandrin: **http://perso.ens-lyon.fr/patrick.flandrin/NSIP03.pdf**
- "An Adaptive Data Analysis Method for nonlinear and Nonstationary Time Series: The Empirical Mode Decomposition and Hilbert Spectral Analysis", Norden E. Huang and Zhaohua Wu: **http://www.deas.harvard.edu/climate/pdf/Zhaohua.pdf**



Giancarlo Fissore (L), 19 and Federico Belliardo (R), 16.

Giancarlo Fissore (**fixormail@gmail.com**) and Federico Belliardo (**federico.belliardo@gmail.com**) are two high school students that share the same interests and passions. We both attend a scientific-technologic school in Fossano, Italy, where we met two years ago. Since then, we have been running a school research project, together with teachers and other students, in which we studied the Sun under different perspectives. In the context of this school project, we took advantage of our IT skills to develop the activity we presented in this article.

We are keeping in touch with the Astronomic Observatory of Turin, where professional researchers always encouraged us and enjoyed our work.

At the age of 17, Giancarlo got interested in the Android's world and developed an application (Guardroid) that he published on the Android Market.