

Data Processing - Periodic Data

N.G. Schultheiss

translated and adapted by K. Schadenberg

1 Introduction

Obtaining data from a single HiSPARC detector is very easy. All data is available from <http://data.hisparc.nl/>. This module explains how to import the data into a spreadsheet application and perform a number of analyses.

2 Getting the data

Data can be downloaded via <http://data.hisparc.nl/>. This page shows all registered HiSPARC detectors, active detectors are shown as links. Clicking on a detector link shows the last day with available data. In figure 2.1 the data from detector 13001, placed at the University of Bristol, from the 1st of November 2012 is shown as an example.

This page shows the measurement data on the left and the detector settings on the right. Lets start at the top right of the page. The first thing we see is a small calendar. With this calendar we can browse to specific dates and look at the results from that day. The link ‘Station List’ brings us back to the overview of detector stations.

Below the calendar is the position of the detector with a the location on a map.

Below the location some hardware parameters are listed. Station 13001 has four detector plates and thus uses two electronics boxes, one master and one slave. Here you can see the hardware number, its firmware version, and the voltage supplied to the different photomultiplier tubes.

On the left hand side of the page the actual data is shown. There are three graphs:

Event Histogram shows the number of registered coincidences per hour.

Pulse height histogram shows the height of the registered coincidences.

Pulse integral histogram shows the integral of the registered coincidences.

The data that is used to generate these graphs can be directly downloaded from this site. In the top right corner of every graph there is a link to the source data ‘Source’. The data is saved in a tab-separated values (.tsv) file. The contents of the file should look something like figure 2.2.¹ The top of the file gives a short description of the file contents. In the example of figure 2.2 the data of 17th of June 2012 by station 13001 is used. It registered 792 coincidences between 2 a.m. and 3 a.m.

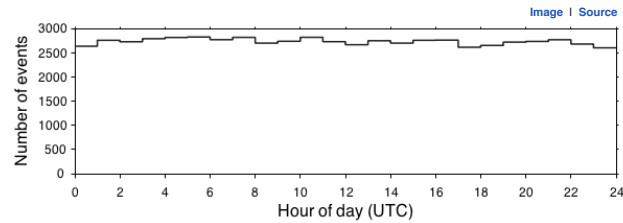
¹Windows systems do not always correctly recognise the file settings of the .tsv file. If this is the case you should use the second import method for Excel described in this module.



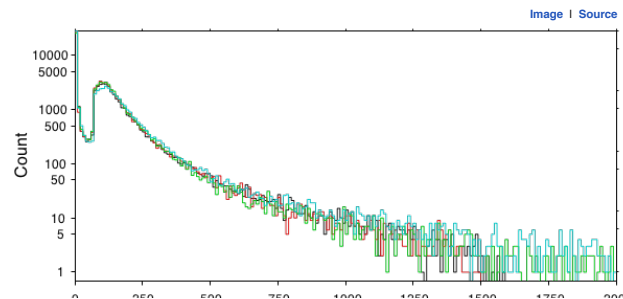
Station: 13001 — Bristol University

Thu, 1 Nov. 2012

Event histogram (number of events per hour)



Pulseheight histogram



Stations [List](#) [Map](#)

Station [Data](#) [Status](#) [Config](#)

← November 2012 →

Mon	Tue	Wed	Thu	Fri	Sat	Sun
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30		

Jan Feb Mar Apr May Jun
Jul Aug Sep Oct Nov Dec

2012 2013

Master

Version Hardware: 176 FPGA: 15
HV Ch1 818 V
HV Ch2 738 V

Slave

Version Hardware: 14 FPGA: 15
HV Ch3 841 V

Figure 2.1: Information available via <http://data.hisparc.nl/>.

```
eventtime-13001-2012-6-17.csv ✕
# "HiSPARC eventtime histogram source"
# "Please note: the 'bin' is the left bin edge. The width of the bin is 1"
# "hour. So bin 0 means between 0:00 and 1:00. Value means the number of"
# "events which were measured during 1 hour."
# bin value
0 763
1 808
2 792
3 769
4 753
5 768
6 760
7 752
8 671
9 705
10 681
11 681
```

Figure 2.2: Part of the .tsv file.

3 Importing the data

The process of importing data is described for two different spreadsheet applications: Microsoft Excel and LibreOffice Calc².

3.1 Microsoft Excel

If the .tsv file is correctly recognised by Windows then the .tsv file can be directly opened by Excel and the bin number and measurement value will have been placed in separate columns. If this is not the case then the .tsv data needs to be manually imported into Excel. To do this follow these steps:

1. Start Excel.
2. Select 'Data' in the Ribbon (top menu bar).
3. Click on 'From text' and select the file you want to import.
4. Excel will now start a Wizard to import the file.
5. Most probably Excel will recognise the correct settings. In step one of the wizard choose 'Delimited' as file type. In step two make sure that only the option 'Tab' is selected. In step three the 'Column data format' should be 'General'.
6. After clicking 'Finish' you can specify where you want to place the imported data.

3.2 LibreOffice Calc

When trying to open the .tsv file with Calc the Text Import Wizard should automatically start. Make sure to select 'Tab' under 'Separated by' and deselect 'Space' (this should be the default setting). Under 'Other options' the first option 'Quoted field as text' should be selected. Click 'OK' and you are done.

4 Simple data processing

Now that we have our data we can look patterns and relationships. We could for instance examine if there are more coincidences during the day time, or if the moon influences our measurements. Let's start simple with the determination of the average number of coincidences during one week.

Exercise 1 : Download the coincidence data of an entire week, make sure the detector was up and running correctly the entire week.

Exercise 2 : Import the data from the first day and try to recreate the graph from the HiSPARC data site. Remember: a graph always has a title and clear descriptions on the axes.

²There are a few differences between LibreOffice and OpenOffice are, but this manual should be applicable to both.

Exercise 3 : Import the data from the other six days. Copy all the data into a new sheet and place the different days in separate columns. Now calculate the average of every hour. Make a graph of the averages. Are there differences between this graph and the previous graph you made? Can you see a pattern?

Exercise 4 : How would you compare data from two different detector stations? What information about the different detectors is available that could help you explain any differences?

Exercise 5 : If you want to look at possible effects of the moon, what is the minimum amount of time you would have to take data from?

5 More advanced data processing

In the previous section we did some easy data analysis, let's take it a step further. Let's see if we can locate possible sources of cosmic radiation.

In theory the HiSPARC detectors could detect particles coming from any direction. However, because of the geometry of the detectors they are most sensitive to particles coming straight from above. Luckily this is also where most of the showers come from (see module 'Airshowers'). Let's assume that all particles we measure at 12 o'clock come from directly above our detector, that would be the direction of the Sun. An hour later our detector has turned away from the Sun and is now looking at a slightly different part of the Sky.

The Earth rotates both around its own axis and around the Sun. That means that, although at 12 o'clock we will always look at the Sun, the stars will have a slightly different position each day. To take this into account in our calculations we need to use a different time system; celestial or sidereal time (see the module 'The Sky' for a bit of background information).

Every day the Earth completes one revolution around its axis of rotation and every year it makes one revolution around the Sun. If we take the stars as reference point however it seems as though the Earth makes one extra revolution around its axis every year. That is why the *rotation period* as used by astronomers differs from the solar day, it includes that part of the rotation that is needed to accommodate the portion of the *orbital period* during one day.

Exercise 6 : Calculate the duration of a sidereal day in solar hours.

Using this new time system we can try to figure out where cosmic radiation comes from. To do this we will divide the sky into four quadrants and calculate the amount of radiation coming from each of the quadrants using vector decomposition, see figure 5.1.

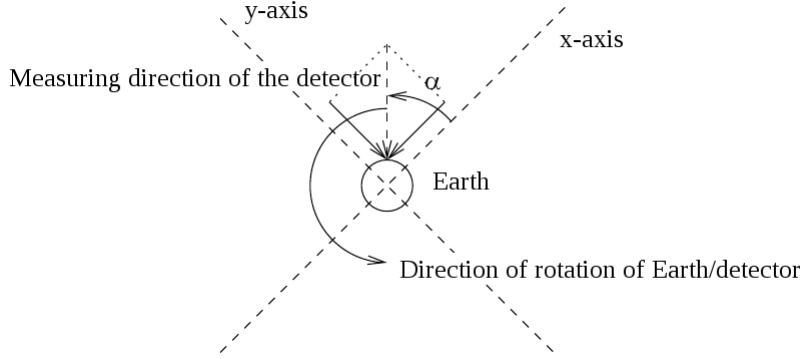


Figure 5.1: The sky is divided into quadrants and the data is assigned to a quadrant according to the measurement angle α .

Time represents a certain angle with respect to the stars. We can switch between time and angle using the following equation:

$$\alpha = 2\pi \frac{t}{T} \quad (5.1)$$

t is the time we want to convert into an angle and T is the duration of an entire revolution. Try to figure out for yourself what time we need to use in this equation, solar or sidereal.

Assigning the data from every hour to the quadrant of figure 5.1 can be done using the following equations:

$$Data_y = Data \cdot \sin(\alpha) \quad (5.2)$$

$$Data_x = Data \cdot \cos(\alpha) \quad (5.3)$$

Or, if we combine these equations with equation 5.1:

$$Data_y = Data \cdot \sin\left(2\pi \frac{t}{T}\right) \quad (5.4)$$

$$Data_x = Data \cdot \cos\left(2\pi \frac{t}{T}\right) \quad (5.5)$$

In these equations $Data$ is the total amount of coincidences recorded during a solar hour. $Data_x$ and $Data_y$ represent the data assigned to the quadrants, there is a positive and negative direction making four quadrants.

If the measured cosmic radiation is uniformly distributed over all directions than the sum of $Data_x$ and $Data_y$ of one entire period will be zero. If however there is more radiation coming from a particular direction than at least one of the sums will not be zero. This enables us to point towards a direction of a possible source of cosmic radiation.

This analysis involves a lot of calculations, luckily the spreadsheet programs can do these calculations for us. But first we need to enter the correct formulas and obtain enough data because there is one

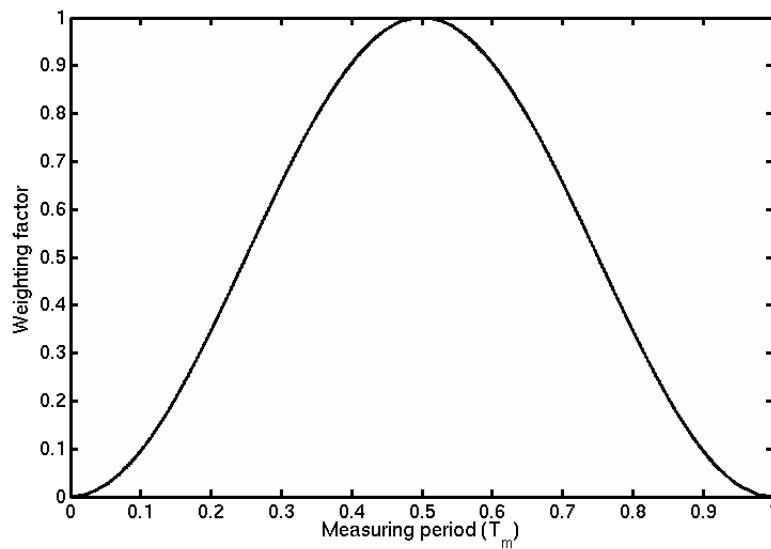


Figure 5.2: The weighting curve used on the data, a raised cosine.

small caveat: the calculations described above only work with a whole number of periods of data.

Exercise 7 : How many solar hours data are needed to obtain a whole number of rotation periods?

As the last question showed we need a lot of data to make the calculations work. Not only does this involve importing large amount of data (a lot of work) but our detector has to be up and running without problems for the same period of time. To make life a bit easier we will use a shorter period of time. To negate the problems in our analysis caused by the incomplete number of periods used we will assign weighting factors to the data. Data at the start and end of our data set will count less towards the final answer while the middle part will be treated normally.

The weighting function should be a pure Gauss-curve. However this curve is infinitely long and we already had the problem of insufficient data. Therefore we will use an approximation which can be easily calculated using a spreadsheet program, a raised cosine:

$$\text{weighting factor} = \frac{1}{2} \left(1 - \cos \left(2\pi \frac{t}{T_m} \right) \right) \quad (5.6)$$

The entire procedure now becomes:

1. Collect the data from a single detector station.
2. Check the validity of the data: Was the detector always on? Are there no strange peaks? Etc.
3. Calculate the duration of a sidereal day.
4. Multiply the data with the raised cosine function to apply weighting.

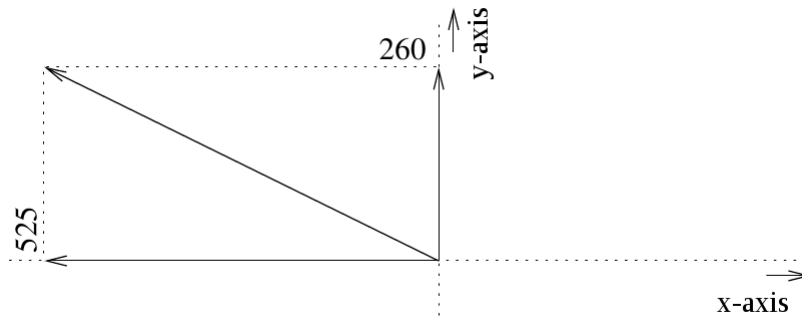


Figure 5.3: The result of the data analysis, the direction of excess radiation is represented as a line.

5. Calculate $Data_x$ and $Data_y$.
6. Determine the sum of $Data_x$ and the sum of $Data_y$.

Are the sums 0? Then there is no direction from where more radiation is detected. Is the sum not 0, then we can take our analysis further. We can use our sums to reconstruct the angle from which the excess of radiation was detected. Lets say we obtained the following sums $\sum Data_x = -525$ and $\sum Data_y = -260$. We can represent this as a line in a coordinate system such as figure 5.3.

Does this nice line mean there is actually something interesting producing cosmic radiation in the direction of the line? We cannot be sure because charged particles are deflected by magnetic fields. But before we start looking up at the sky, let us first look at the amount of excess radiation. To do this we will compare the amount of radiation coming from the found direction with the total amount of radiation using:

$$\text{deviation} = \frac{\sqrt{\sum Data_x^2 + \sum Data_y^2}}{\sum Data} \cdot 100\% \quad (5.7)$$

Exercise 8 : Use the analysis described above to analyse one week worth of data from your own detector station.