Data Processing - Non-periodic Data

N.G. Schultheiss
translated and adapted by K. Schadenberg

1 Introduction

This module is a continuation of 'Periodic data' in which the downloading and importing of data was explained. The previous module also showed how to analyse data when there is some periodic dependence. But you might also want to look at the influence of phenomena which do not have a clearly repeating pattern; the influence of weather, e.g. temperature, ambient pressure, and cloud cover, or other non-repeating effects. The temperature might roughly follow the pattern of day and night, but lightning clearly does not.

2 Change of Occurrence

Before we can determine if cosmic radiation or air showers are influenced by a certain phenomena we first need to make sure the phenomenon actually occurred. In some cases we know with 100% certainty that something happened, but in other cases we are not so sure, then we might only know the odds of it happening; the probability $P()$.

Lets start by looking at an easy example. We have been observing the weather for an entire day and recorded the times at which there were thunderstorms. The results are shown in figure 2.1.

This figure is only part of the story, we also need to look at the odds of 'no thunderstorms'. The probability of no thunderstorms occurring are shown in figure 2.2.
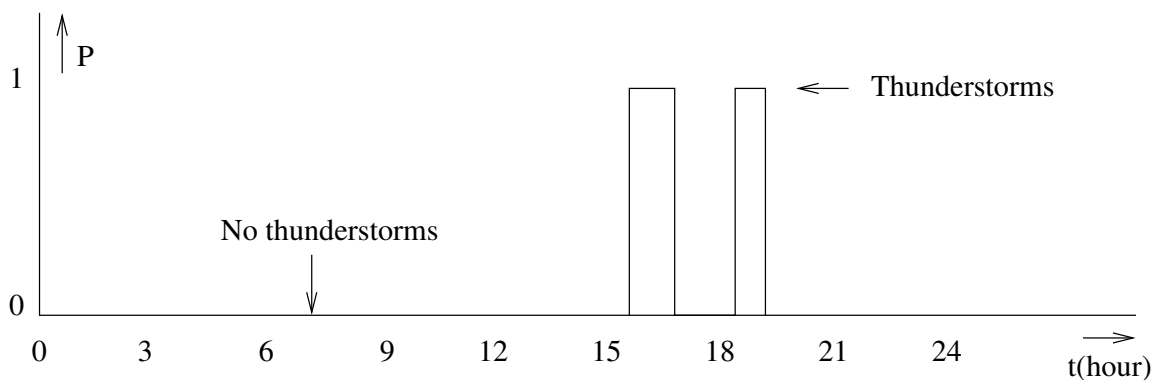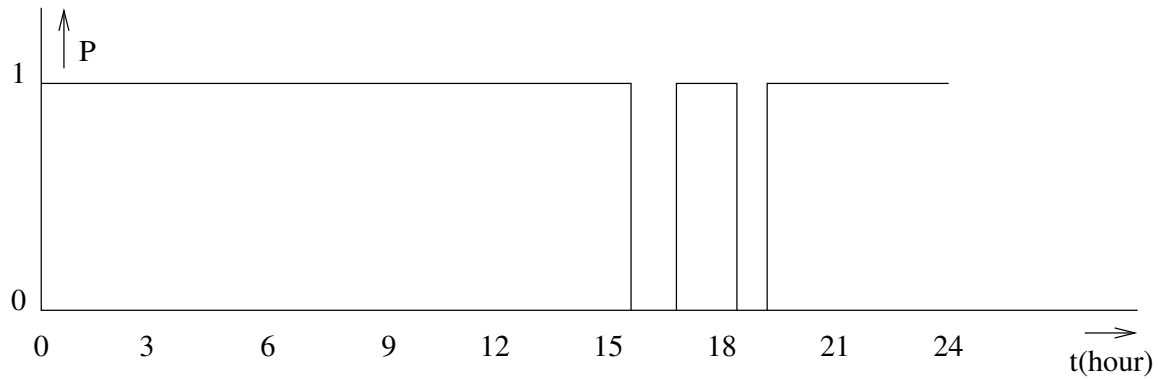

Figure 2.1: The odds of thunderstorms.
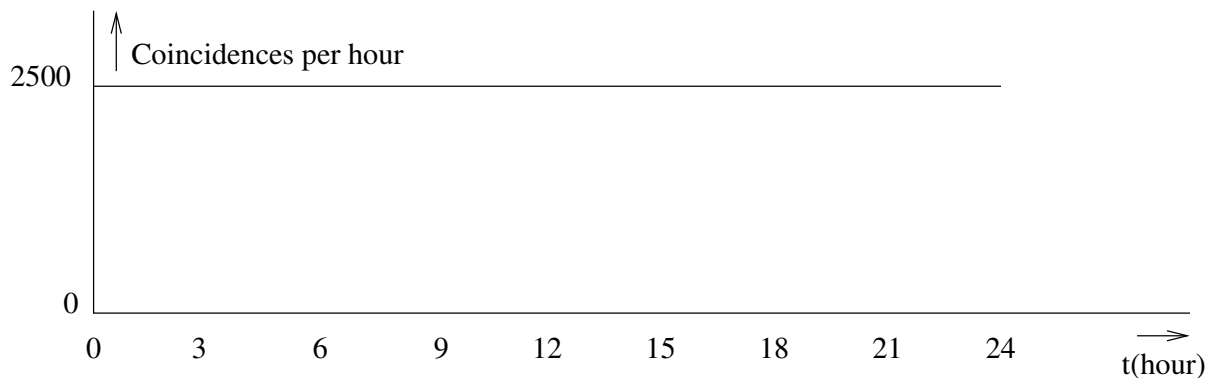
Figure 2.2: The odds of no thunderstorms.

Figure 2.3: A constant flux of cosmic radiation.

Exercise 1 : Explain why the following equality is true:

$$P(\text{Thunderstorms}) + P(\text{No thunderstorms}) = 1 \tag{2.1}$$

Lets assume cosmic radiation is not influenced by the weather. We therefore create fictitious measurement data, as shown in figure 2.3, of the day we observed the thunderstorms with no variation during the day.[1]

When investigating the influence of thunderstorms we would compare the measurements during thunder with the data taken from when there were no thunderstorms. To do this you multiply figure 2.3 with figure 2.1 and add all the values to obtain the number of coincidences for when there were thunderstorms and with figure 2.2 to obtain the data without thunderstorms. It now seems as though there is a much higher change to observe cosmic radiation in nice weather. There is one final step needed to correct this.

---

[1]We have also 'corrected' our fictitious data for any known and unknown variations other than thunderstorms.

Exercise 2 : Figure 2.1 and 2.2 show that is far more likely to have nice weather than to have thunderstorms. How does one need to use this information to determine whether or not thunderstorms influence cosmic radiation.

Hint: What does the area underneath the lines in figure 2.1 and 2.2 tell you?

3 Using a Spreadsheet

Processing non-periodic data can be done using a spreadsheet. The following steps are written for our thunderstorm example, but can be easily modified and expanded to tackle other problems.

1 Download the relevant (coincidence) data from the HiSPARC website and import it into the spreadsheet.

2 Calculate the probability of thunderstorms for every hour and place the results in a separate column 'Thunder'. If there was a thunderstorm lasting 15 minutes during an hour, the probability P(Thunderstorm) = 0.25.

3 Calculate the probability of no thunderstorms using
P(Thunderstorms) + P(No thunderstorms) = 1 and place it in a separate column labelled 'No thunder'.

4 Calculate the sum of the columns 'Thunder' and 'No thunder'.

5 For every time interval (hour) calculate #coincidences · P(Thunderstorms) and
#coincidences · P(No thunderstorms), place the results in two new separate columns.

6 Calculate the sum of #coincidences · P(Thunderstorms) and divide this by the sum of the column 'Thunder', this is 'Chance of Thunder'.

7 In a similar fashion calculate the 'Chance of No thunder' by calculating the sum of #coincidences · P(No thunderstorms) and divide this by the sum of the column 'No thunder'.

8 Compare the values for 'Chance of Thunder' and 'Chance of No thunder', the largest wins.

Exercise 3 : Some HiSPARC stations are equipped with a weather station. Temperature and ambient pressure data are sent along the with other data to the NIKHEF in Amsterdam and the results can be downloaded from the same data page. Station 501 is such a station, an example of its data page is shown in figure 3.1.

What problems arise when we try to investigate the dependence of cosmic radiation on ambient pressure when we use the data from station 501?

Barometer data
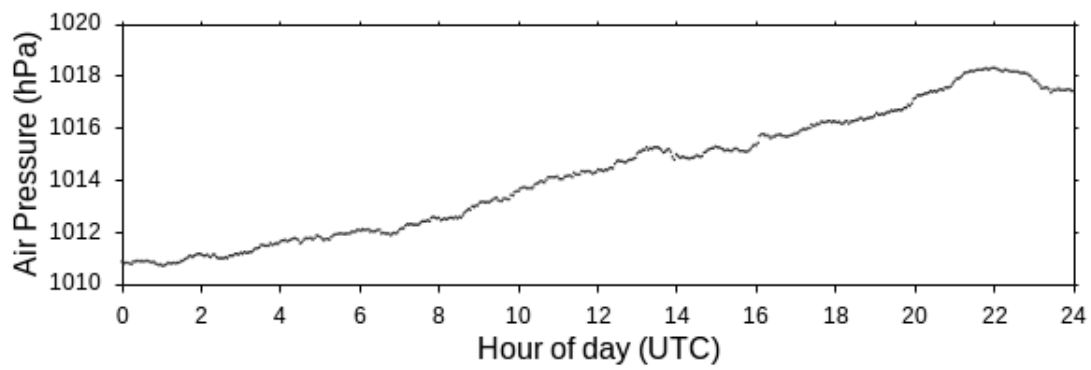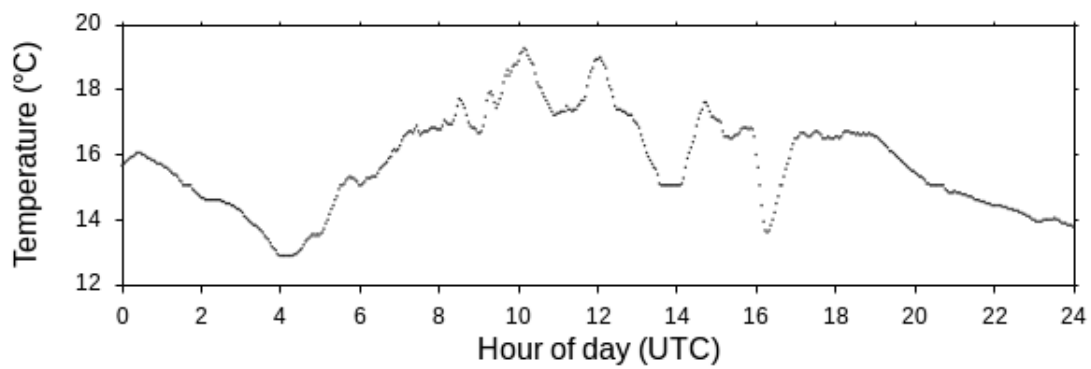
Temperature data

Figure 3.1: Weather data from station 501 on the 1st of July 2012.

Figure 3.1 shows the weather data from station 501, like the other graphs on the data page these graph have a link to the source to download the data. The results are written in a .tsv file but in a slightly different way than the detector data. The temperature and ambient pressure are measured every second. These seconds are not counted for every day, but since the 1st of January 1970. This is known as POSIX time.[2] Our coincidence is per hour, so we need to convert the weather data to match this.[3]

4 Correlation between Two Graphs

Looking for correlations between two graphs can be done using the steps described in the previous section. But there is a faster and smarter way of doing this. In the following example we will look at the possible connection between cosmic radiation and ambient pressure.

1 Download the coincidence data and import it into a spreadsheet.

2 Download the weather data for the same time interval and also import it into a spreadsheet. If needed convert the data so it can be used with the coincidence data (rebinning).

3 Determine the average pressure for the entire time interval.

4 Calculate the difference between the instantaneous pressure and the average pressure for each measurement point, place these values in a separate column.

5 Multiply the deviation from the average with the number of coincidences for each time interval. Place the result in a separate column.

6 Calculate the sum of this final column. Is the result a positive number, then an increase in pressure is associated with an increase in cosmic radiation. Is the result negative, then the reverse is true.

Exercise 4 : Compare the procedure above with the procedure from the previous section. Do they give the same result?

---

[2]Explaining all the details about POSIX or Unix time and why it does not include leap seconds for instance goes beyond the scope of this text. On the internet multiple methods can be found to convert POSIX time into something we can all understand.

[3]This is called 'rebinning', all the data was placed into bins (sort of like buckets) 1 second wide, now we want to place them into bins 3600 seconds wide.