Lund University

MNXB01

Introduction to Programming and Computing for Scientists

Temperature analysis

Authors:
Albin Hedse
Marcus Sköld
Harald Havir
Karolis Sulinskas

Supervisor:
Oxana Smirnova

November 9, 2017

1 Introduction

We were provided with several data sets of measured temperatures in different locations around Sweden. The data sets included the date of measurement and the time during the day when the measurements were taken. The continued analysis made here was done using the Lund and Visby data sets.

In the Lund dataset, every measurement taken was confirmed to be reliable, whereas the quality of the data from Visby was mostly unverified until recent years (near 2010- onwards).

2 Preparation

2.1 Reading the data set

First off, we needed a way to read the datasets provided in a proper format. First we created a class that would contain every function that were going to be used. The first lines were discarded since they did not contain useful data. After this, each line was divided using a delimiter and put into new text files with all relevant data from the datasets.

In order to read the data in an efficient manner, we wanted to restructure the data delivery method. Originally, the data was written in a spreadsheet format with the data being separated by the delimiter ";". We created a function which read the data using the stringstream data type. We used the partitions as markers in order to identify the useful information and wrote the information we wanted into a .dat file with empty space partitioning the information. We could then use standard iostream file processing methods to access the data more easily and quickly.

In order to get rid of redundant information at the start of the data files, we identified the format definition line to be identical and in the same position in each document. We then used the getline function to reach the last line in the document before the temperature measurements. After reaching this point, the getline function was used to read off the temperature data.

Because the way our functions were written, we only cared about average temperature measured for each day, this further shortened down our .dat files since we could omit between two and 24 data points for each day, in favor of the mean value and the standard deviation of each day's temperature.

This was done for both Lund and Visby where the data was then saved for later use and further analysis. It should be noted that the data set had more than just one measurement on all days.

2.2 Mean value and Standard deviation

The data is read from the files by a function that progressively checks each line of the data file in a while loop. The temperature and date for each measurement is saved. Then, every time the date changes, the temperatures measured from

the previous day are averaged to give the mean temperature of the day, the standard deviation of the mean temperature is also calculated as well.

We implemented a function calculating the standard deviation for each day in the data set

$$\sigma = \sqrt{\frac{\sum(x_i - X)^2}{N}} \tag{1}$$

Where the x_i is each measured temperature over one day, X is the mean temperature of that day and N is the number of measured temperatures on that day. The code would leap through the data set and give the standard deviation of each day.

3 Temperature of a given day each year

3.1 Approach

For starters we wanted to view how the mean temperature distribution throughout the years look like. With our new generated data file we implemented a function in the class that first read the file using dummy variables as long as the format is the one provided in the code. However the class function needs an input for what day it should watch for and runs thought the data set. To make sure that it worked properly we also output the information of that given day each year in the data set.

Some statements were included later on that prevents the user from entering an input that is not correct and asks for another. Once this was corrected for, a histogram was implemented to present that data in a more intuitive way. Showing how many days had what temperature.

3.2 Results

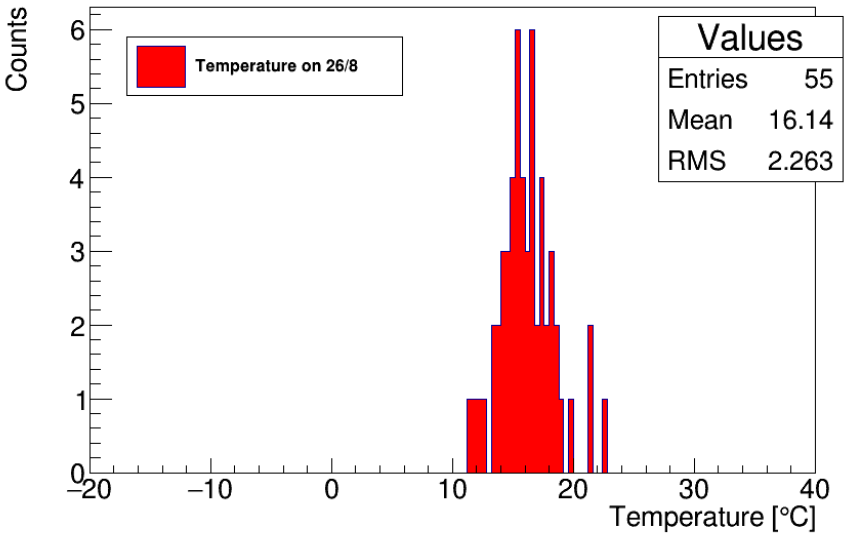

Fig. 1: Histogram for the mean temperature during 26/8 each year in Lund

3.3 Discussion

This is of course only one of many different histograms that one could generate with the code written and given two data sets. As expected for each histogram it had a Gaussian distribution, centering around a specific temperature with not much variance to speak off. If we would take every data point instead of the mean the result would be the same however the histogram would be more spread out over temperature values.

Given more time we wanted to compare the Y and G values inside the Visby data to see how and if their distributions would differ.

4 Temperature of a single year each day

4.1 Approach

The purpose of this function was to read the average temperature of each day during a single year and its standard deviation and output this data into a graph. This was implemented by asking the user for the desired year to be read. The function then reads through the simplified data file and only reads the temperatures and corresponding standard deviations from the file belonging to the year that has been input by the user. This is done by use of a while loop and storing of variables in a similar fashion to the readfile function above. A graph is then drawn using the temperature and standard deviation resulting in the below graphs.

4.2 Results

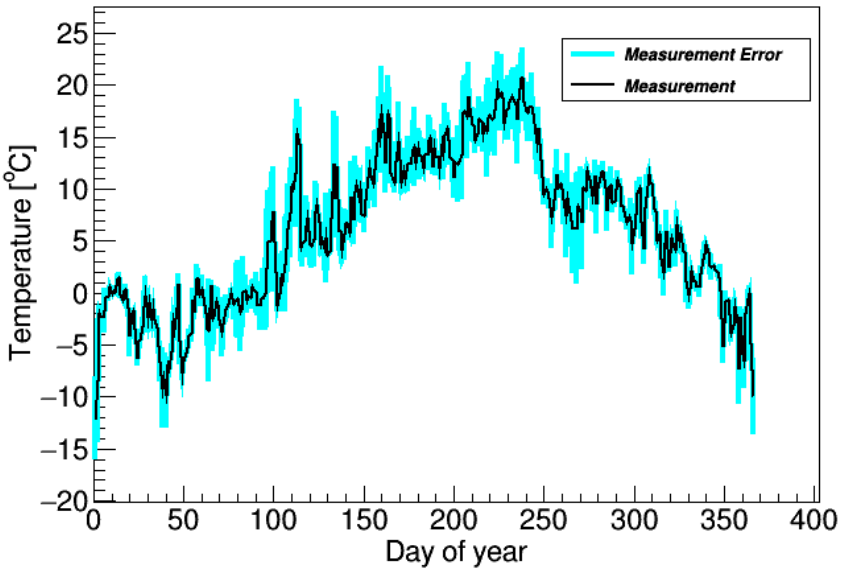

Fig. 2: Graph over the measured temperatures over the whole year 1996 in Visby

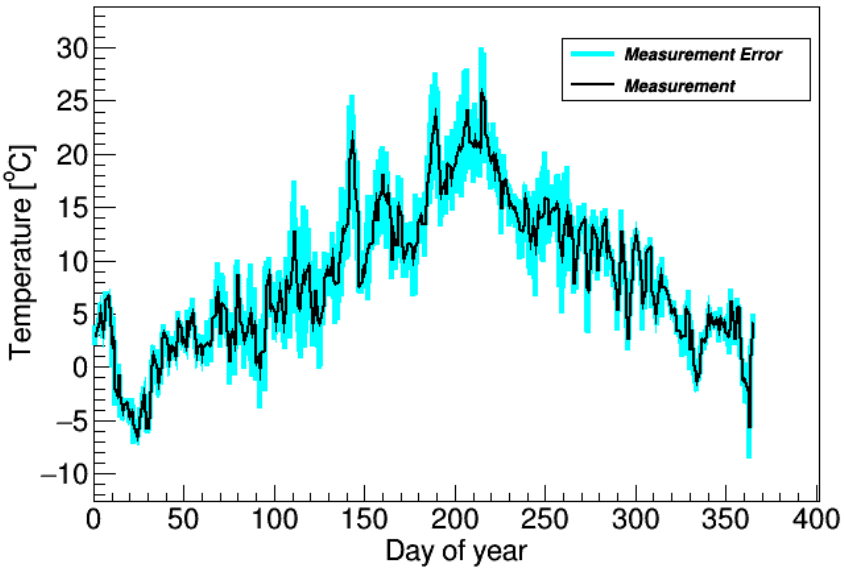

Fig. 3: Graph over the measured temperatures over the whole year 2014 in Visby

4.3 Discussion

These results give a rather good indication of the change in temperature over the course of a year. This gives an indication of how temperature changes over the course of a year, and in what range the temperature stays at during the year. A function that takes data from every year and calculates a standard deviation to make a kind of "mean year" calculation over the last 60 years could be implemented to reduce the impact of variance during a single year. This would however make it impossible to distinguish trends characteristic to certain years/groups of years, so it might not be a perfect solution.

5 Temperature difference between Lund and Visby

5.1 Approach

We wanted to see the temperature difference between the two selected data sets. For this analysis, we re-used the core method of the temperature of each day code, but made some changes to adapt it to two sets of measurements. We performed this analysis in hopes of noticing temperature differences between a coastal city (Visby) and a mainland city (Lund). The expected results would be that Lund has more extreme temperatures whereas the temperatures in Visby fluctuate less. This would show up in the comparison graph as a smaller difference in the winter, and a larger difference in the summer. We also implemented a new standard deviation for the comparison using the following formula

$$\sigma_i = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \tag{2}$$

where σ_i is the standard deviation for the difference of the mean temperature of the same day i in our data sets, running through every day in each year. The different indices 1 and 2 just show that it's the standard deviation σ and number of entries of n, that are measured during a given day in each data set.

5.2 Results

The results of this analysis can be seen in figures 4 and 5. We can in figure 4 see some hints of a confirmation of the hypothesis, where the temperature difference seems to be greater for about 100 days of the year. In figure 5, however, we can not clearly see evidence of this trend.

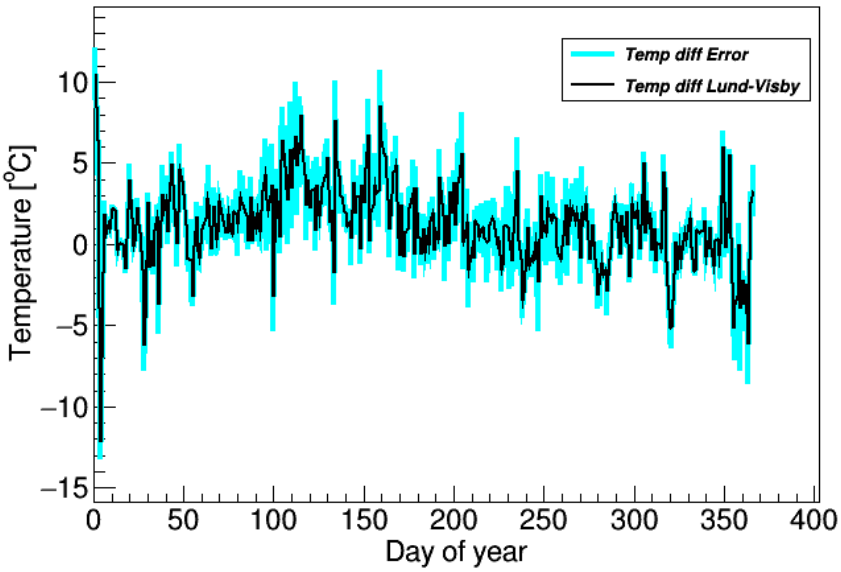

Fig. 4: Graph comparing the two data sets of Lund and Visby during the year of 1996

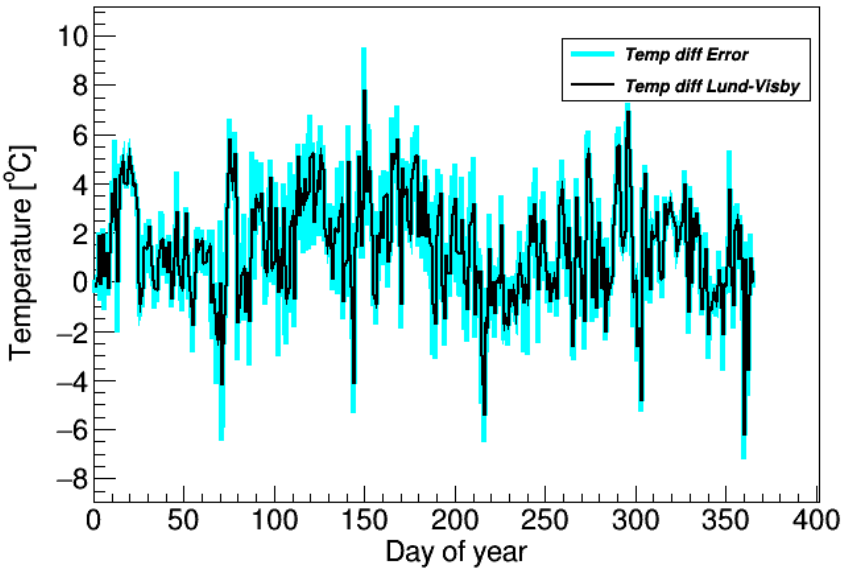

Fig. 5: Graph comparing the two data sets of Lund and Visby during the year of 2014

5.3 Discussion

The results could probably have been easier to analyze if we also plotted the mean temperature of each week / every two weeks. This would have given a line with lower resolution and remove a lot of the noise from the measurement data. In order to get a proper conclusion, we should have made this analysis over every available year, to include as many measurements as possible in one data set. As it is now, we have too large variance in each individual resulting graph to draw any concrete conclusions.