Final Project

Martin Waites
Peicen Liu
Yichen Liu
Alexander Knyazev

MNXB01 - Introduction to Programming and
Computing for Scientists
Lund University
November 2019

1 Introduction

The purpose of the project is to demonstrate some of the things we have learned about data processing and analysis throughout the course. To this end we have been given data about temperatures measured in different locations around Sweden. This data comes from the Swedish Meteorological and Hydrological Institute (SMHI) where in some cases measurements have been going on for hundreds of years. From this data we produced five different types of results named oneDay, WarmColdDay, everyDay, LattDiff and getTemperature named after the names of the main functions called to produce the results.

2 Function Descriptions

2.1 Data Cleaning

The raw data files were in the CSV form. Some small manual modifications were done first, such as taking away the description lines at the beginning of the files, and the rest cleaning process was done by a Python function. The input arguments of the function are the directory of where the original data file and the directory to store the clean data file. The function converts the CSV file into a .dat file which has the format [year, month, day, temperature] in each line. Some of the subsequent analysis functions are created for the clean data files.

2.2 oneDay

The oneDay function is consisted of three parts, or three sub-functions. The first part is called oneDayMDI which takes two integers (month and date) as inputs and outputs either a histogram showing the temperature of the input day in every year, or an array that contains the mean and the standard deviation of the input day. The output is decided by the argument plot=0 for producing the array and plot=1 for generating the histogram. The second part is called oneDay and takes one input argument of the day number (an integer between 1 and 365). The output is the same as in part one. However, it was decided to skip the 29th of February in each leap year for the second part, and therefore the day number only goes up to 365. An example histogram is shown in Figure 1, and it shows the temperature on the Christmas day in Uppsala from 1722 to 2013.

The third part is different from the previous two parts. The function for the third part is called oneDayProb which takes input arguments of month, day, a given temperature and an error range. The function computes the probability of getting the given temperature, for a specific date in a region, within plus/minus half of the given error range. The function uses Gaussian distribution for the probabilistic modelling which is the most reasonable distribution in this case. For example, oneDayProb(”uppsala.dat”, 12, 25, -1,

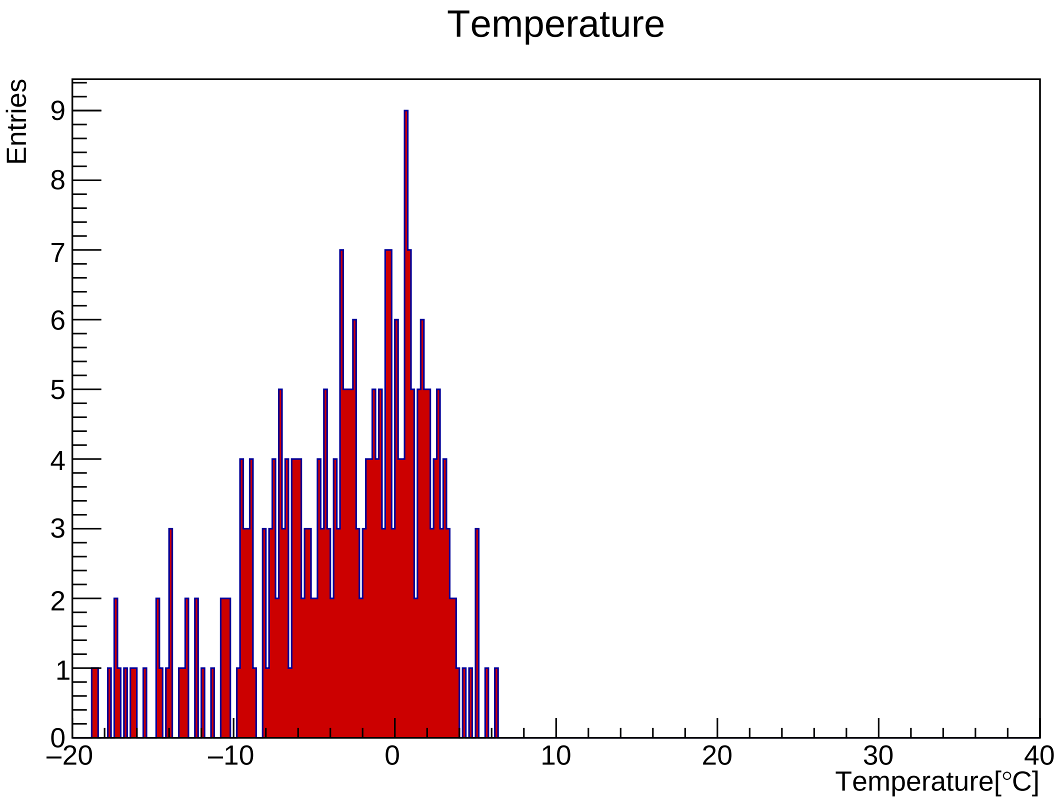

Figure 1: Christmas day temperature in Uppsala from 1722 to 2013

2) calculates the probability of getting a temperature between -2 and 0 (-1 plus/minus 1) on the Christmas day in Uppsala.

2.3 WarmColdDay

The WarmColdDay function demonstrates the days of warmest and coldest days occurred at and the output shown by histogram allows us to visualise when the warmest and coldest days are most likely to occur during the year. A Gaussian function is plotted to the histogram to fit the distribution.

The input of this function is the name of the location, which are lund, visby, umeå, luleå and uppsala. Figure 2 shows one of the outputs, where the blue and red histogram indicates the coldest and warmest days respectively, while the blue and red gaussian line indicates the distribution of warmest and coldest days respectively. The outputs for the other four locations are also plotted and the value of mean value and standard deviation are shown in Table 1, we can see that the error values are really large and this is due to lack of data in the dataset provided for the four locations.

2.4 everyDay and LattDiff

The everyDay function produces a plot showing the mean and standard deviation of the temperature of each day for every day of the year. This allows us to visualise how the temperature changes over the year and see when the seasons occur. Figure 3 shows the output of this function.

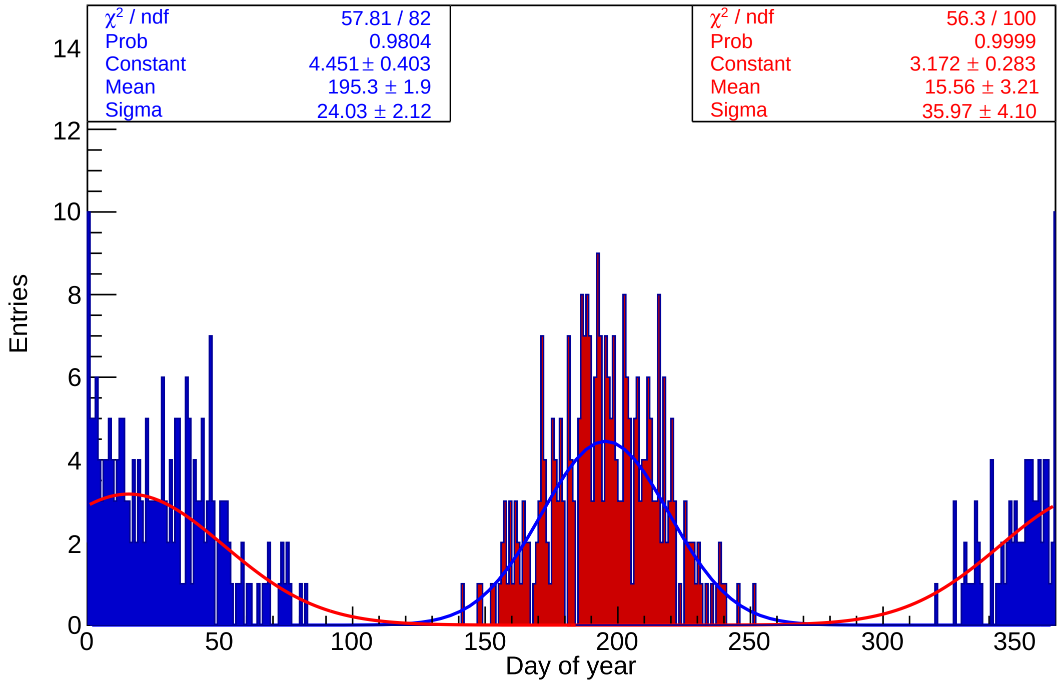

Figure 2: Warmest and coldest day in Uppsala

Location	Mean Value		Standard Deviation	
	Cold	Warm	Cold	Warm
Uppsala	15.6 ± 3.2	195.3 ± 1.9	36.0 ± 4.1	24.0 ± 2.1
Luleå	20 ± 28	178 ± 50	71 ± 54	80 ± 149
Umeå	23 ± 44	187 ± 91	82 ± 95	122 ± 143
Visby	29 ± 55	211 ± 73	144 ± 212	92 ± 166
Lund	25 ± 60	184 ± 8	105 ± 188	81 ± 84

Table 1: Mean Value and Standard Deviation

The LattDiff function produces two plots, these plots show the temperature of each day for every day of the year similarly to the everyDay function however for five locations of different Latitudes. This allows us to show the effect of latitude on the temperature distribution over the year. Figure 4 shows this graph while Figure 6 in the Results shows only the mean without the standard deviation in order to get a clearer image of the change with Latitude.

2.5 getTemperature

The `getTemperature` function reads an initial data file for a location from a predefined list, which includes Boras, Falsterbo, Falun, Karlstad, Lulea, Lund, Soderarm, Umea, Uppsala and Visby. The function plots the data and makes a linear fit of the data, and save result as a `.pdf` document. The time axis on the plot is in seconds, to include the leap years. The fit results can be interpreted as the average temperature in the location (`p0`) in °C and

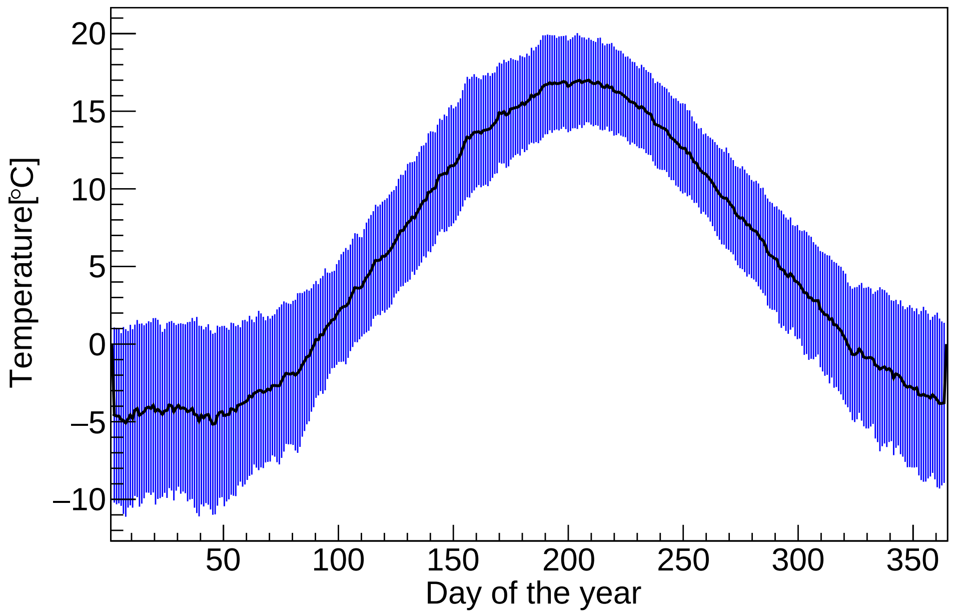

Figure 3: histogram plot produced by the everyDay function

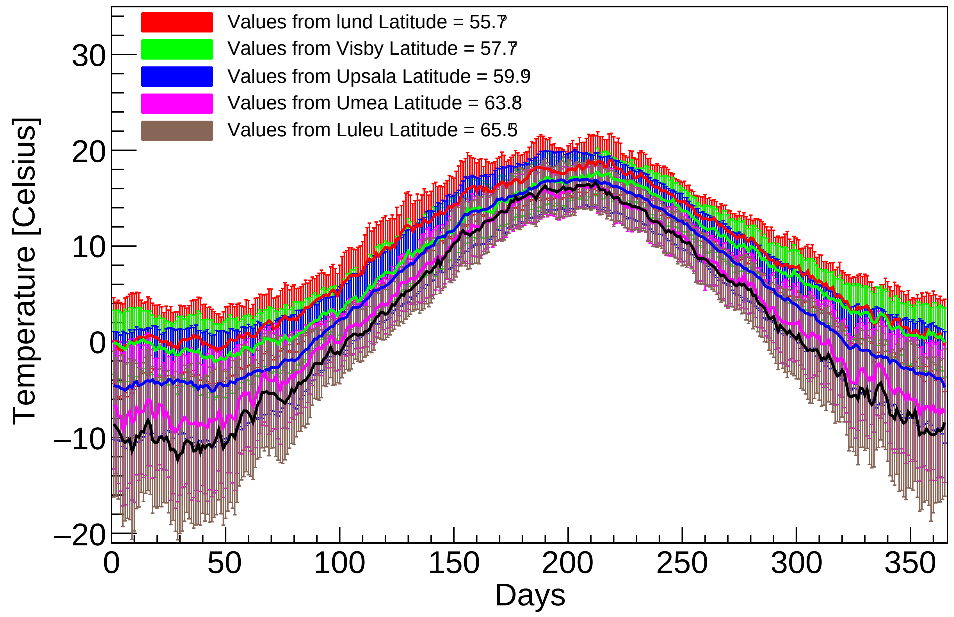

Figure 4: Plot produced by the LattDiff function

the change of average temperature over time (p1) in °C/s.

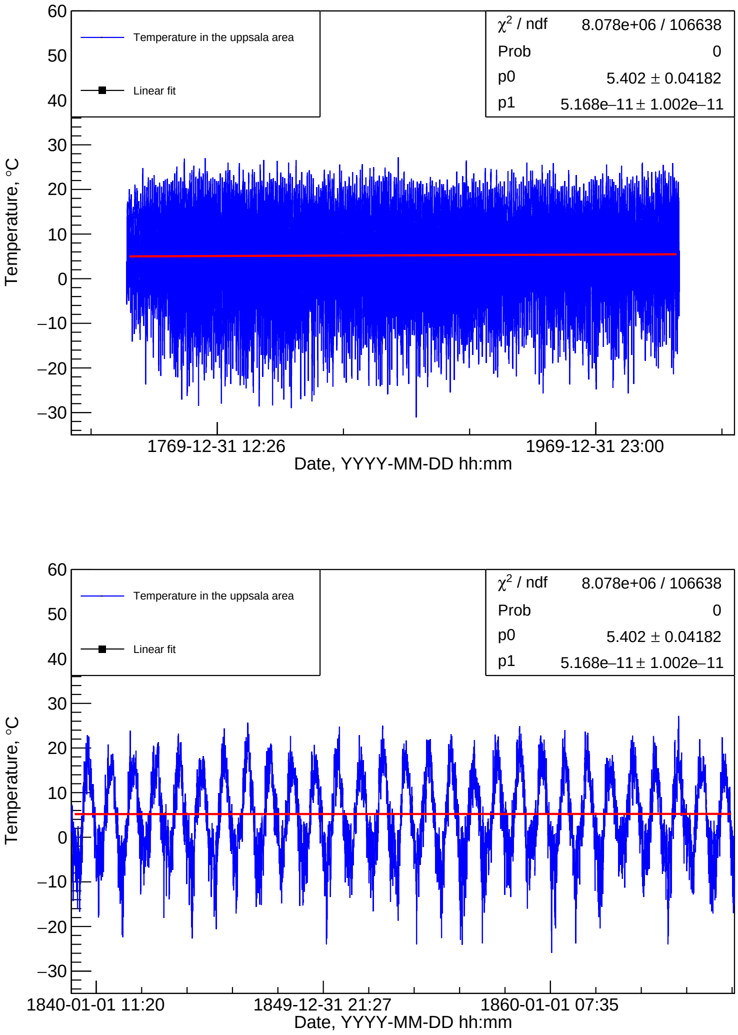

Figure 5: An example of output of `getTemperature` function. Top panel: the full range of data. Bottom panel: a reduced sample of 10% of the data, from the middle of the data set.

3 Results

Figure 3 In the everyDay and LattDiff section shows what we expect, there is a clear region for each of the seasons

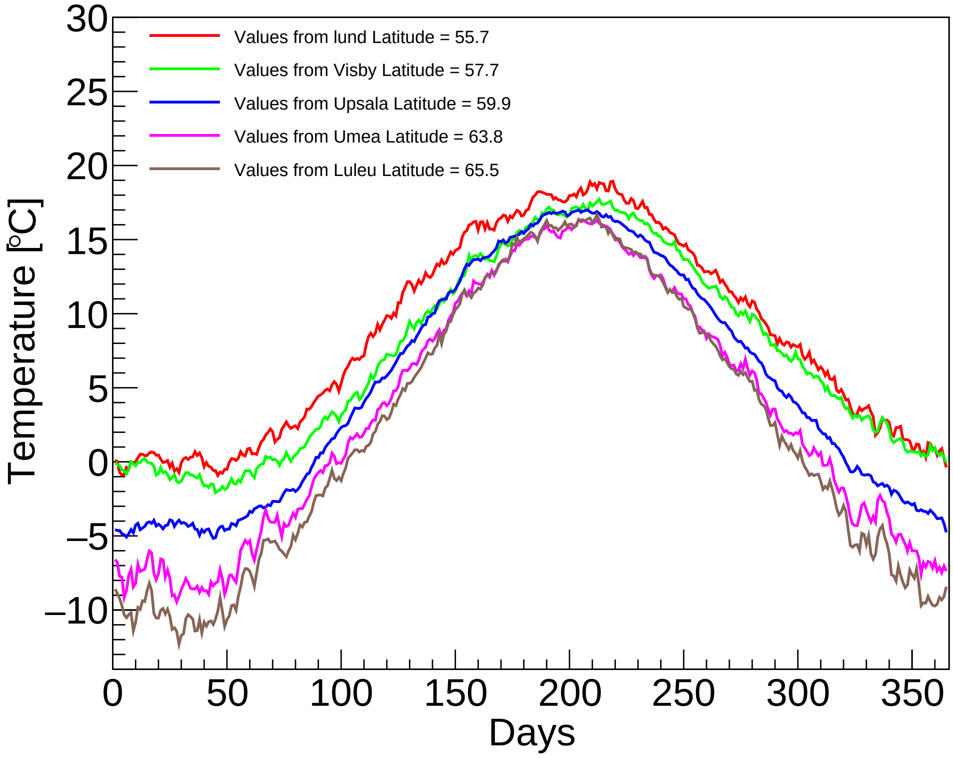

Figure 6: Plot produced by the LattDiff function showing only the mean

From Figure 6 we see that as expected the overall temperature tends to increase as the latitude decreases.

The analysis of the data as to the average temperature and increase of temperature overtime is presented in Table 2. It was found that the data for first 10 years of measurements for Falun location was corrupted (Figure 7), this data was removed from the analysis.

From the results presented in Table 2 one can notice, that for each data sample, a positive change of average temperature is observed. This change is especially noticeable for the data gathered in 20 century. This fact could indicate a connection between the temperature and increase of power consumption by society in the recent 100 years. Another noticeable fact is that the change of average temperature is the highest in the locations with lowest temperature in the analysed data.

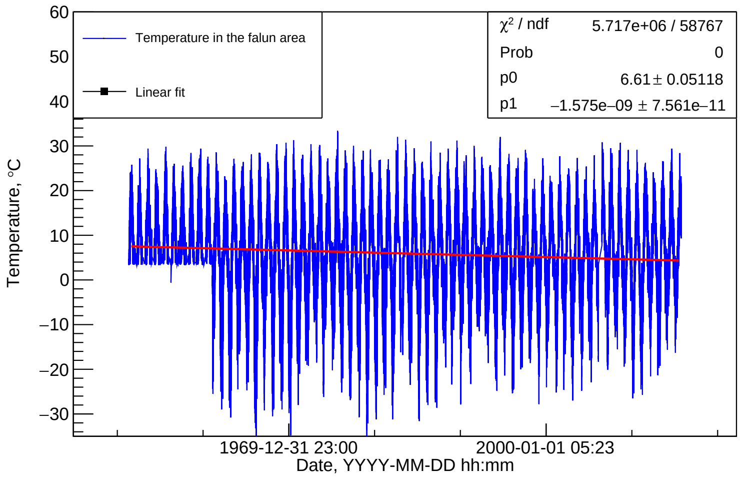

Figure 7: The data set for the Falund location, note the data corruption for the first 10 years. This corrupted data was excluded from the analysis.

Location	Average Temperature, °C	Change of average temperature, °C/year
Boras	6.59 ± 0.05	0.015 ± 0.002
Falsterbo	8.13 ± 0.02	0.030 ± 0.001
Falun	4.85 ± 0.08	0.020 ± 0.003
Karlstad	6.31 ± 0.02	0.002 ± 0.001
Lulea	1.91 ± 0.02	0.024 ± 0.001
Lund	8.36 ± 0.05	0.011 ± 0.002
Soderarm	5.61 ± 0.02	0.038 ± 0.001
Umea	2.42 ± 0.02	0.041 ± 0.001
Uppsala	5.40 ± 0.04	0.002 ± 0.000
Visby	6.93 ± 0.01	0.022 ± 0.001

Table 2: Average temperature and change of average temperature over time for the selected 10 data sets. Details in the text.