

Temperature Comparisons Between Lund, Uppsala, and Luleå

Emil Boman,^{*} Henrik Olsson,[†] and Pavel Oshchepkov[‡]
(Dated: November 5, 2023)

INTRODUCTION

In this project we explore a number of long term temperature comparisons between Lund, Uppsala, and Luleå. The first of these is the number of winter days per year. This both includes studying the change in absolute numbers over time, and the relative change between the cities. The second is the average monthly temperature difference between Lund and Luleå, and in order to obtain this the average monthly temperature for both cities was determined. The third comparison is the hottest and coldest day of the year, where the number of times each day of the year has been the hottest or coldest is shown for each of the three cities.

The datasets, in the form of comma separated values (CSV) files, used for these comparisons were taken from SMHI open data [1]. The Luleå dataset contains hourly measurements going back to 1949. For the same time period the Uppsala dataset also contains hourly measurements. In contrast, the Lund dataset varies between two, three, and five measurements per day during the time period. Table I shows the time of day measurements where taken in the Lund dataset from 1949 to 2023.

Time Period	Time of Measurements
1949→ 1974-05-31	06:00, 12:00, and 18:00
1974-06-01→1980-02-29	06:00, 09:00, 12:00, 15:00 and 18:00
1980-03-01→1996-06-30	06:00, 12:00, and 18:00
1996-07-01→	06:00 and 18:00

TABLE I: The time of day measurements were made for different time periods in the Lund dataset.

Concerning the git-policy used for this project, all developers were allowed to directly interact with the main repository. The reason for this choice was that developers were working at different parts of the project, and had differing working hours, making it the most convenient policy for a short-term project.

The structure of the remainder of the report is as follows. First, the data cleaning procedure common to all subprojects will be described. After which, the subprojects will be presented one by one. For each of these, data preparation, processing, and plotting will be described, before the results are shown and discussed. Lastly, conclusions and outlook are given.

DATA CLEANING

The datasets, in the form of csv-files, treated in this project contain meta data that need to be removed from the versions given as input to the data processing functions. In order to accomplish this a simple bash script `cleanData.sh` was used. It first uses the `cut` command to separate the first 3 columns in the file, which contain data on the format `year;time;temperature`. It then uses the `tail` command to remove metadata at the top of the file.

NUMBER OF WINTER DAYS

The aim of this part of the project was to compare the number of winter days, i.e. the number of days with negative average temperature, per year in Lund, Uppsala, and Luleå, during the time period 1964-2022. To accomplish this the number of winter days were calculated and plotted, together with the corresponding centralized moving average over five-year intervals. In addition the relative number of winter days for Lund and Uppsala, using Luleå as the base line, was also determined. The data processing consisted of both a bash script and C++ functions, and the plotting was done using the ROOT framework. In this section begins with a description of the pre-C++ data processing, followed by the C++ data processing and plotting, after which the results are described and analyzed.

Pre-C++ Data Processing

The first step of the data treatment, after the cleaning process, is the extraction of days with negative temperature, which is done using the bash script `negTempDays.sh`. This script also introduces the cut off year 1964. This step reduces the size of the datasets to be treated significantly. Similarly to `cleanData.sh`, the `tail` command is utilized to do the line cut off. The `grep` and `sort -u` commands are then used to track the days with negative temperatures, which are lastly extracted, again using `grep`.

C++ Data Processing and Plotting

The data processing using C++ is done using the functions `winterday_map` and `winterday_hist_fill` in `winterdays`. The first of these take a csv-file on the format `date,time,temperature` and determines the num-

ber of winter days for each year by calculating the number of days where the sum of the temperature measurements are negative, which corresponds to days with negative average temperatures. The number of winter days for each year is then given as output inside a `map<int year,int temperature>`. The second function, `winterday_his_fill`, is then used to insert the data in these maps into the corresponding histogram, and to calculate the centralized moving average over five-year intervals, which is also stored in a histogram.

The plotting of the number of winter days histograms is treated by `winterday_his_draw`, which draws histograms and moving averages, provided in two vectors, on a canvas and then saves the canvas as a pdf. For the plotting of relative number of winter day histograms, `rel_winterday_his_gen` is used. This function takes two maps containing years with corresponding number of winter days and inserts the quotient of the number of days into a histogram, using the years as bins. The histogram is then drawn on a canvas, which is saved to a pdf.

Results

The number of winter days per year recorded in Lund, Uppsala, and Luleå, from 1964 to 2022, is shown as histograms in figure 1. Included in the figure is also the centralized moving average over five-year intervals for each histogram.

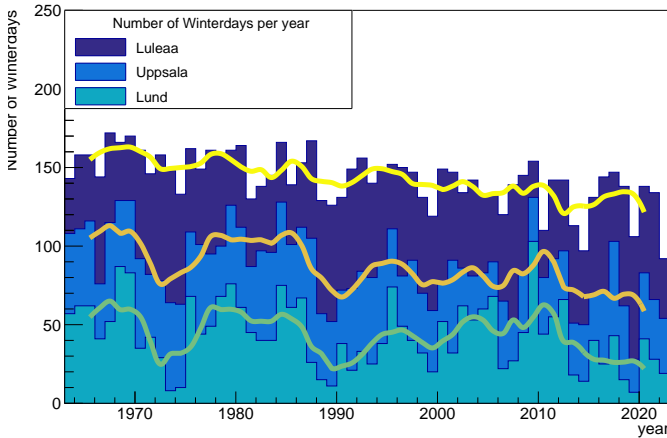
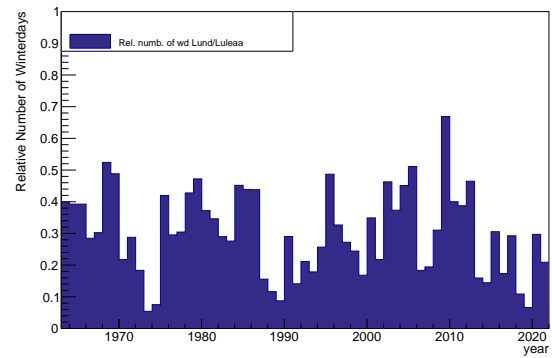


FIG. 1: The number of winter days from 1964 to 2022 for Lund, Uppsala, and Luleå. The lines are the centralized running average over five-year intervals for the corresponding histogram.

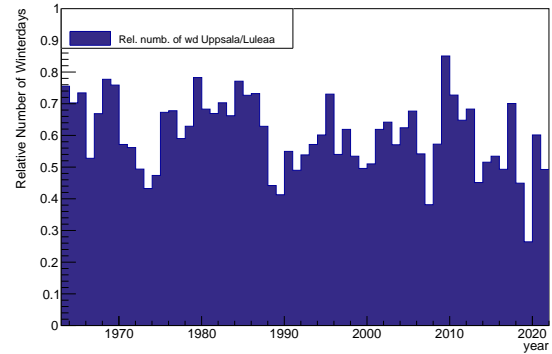
As expected the further north a city is located, the more winter days it tends to have. In addition, the relative variation in number of winter days decreases, and the periods of relatively few winter days shortens, for cities further north. Considering the moving average for Uppsala and Luleå, they show a clear decrease over time, indicating the increase in average temperatures due to global warming. This decrease is not as clear for Lund.

In Lund and Uppsala there was a significant spike in the number of winter days in 2009, after which there has been a period of relatively few winter days, particularly for Lund. It should be noted that Uppsala has had two smaller spikes during this period. Luleå, on the other hand, does not display a similar period.

The relative number of winter days between Lund and Luleå, from 1964 to 2022, is shown in figure 2a, and the counterpart for Uppsala is shown in figure 2b. The 2009 spikes are also noticeable in the relative plots, particularly for Lund. Periods of fewer winter days shown in figure 1 have corresponding periods in the relative number of winter days as well, clarifying the higher variation in number of winter days for cities further south. When comparing figures 2a and 2b, the higher variation for Lund is also clear.



(a) The relative number of winter days between Lund and Luleå from 1964 to 2022.



(b) The relative number of winter days between Uppsala and Luleå from 1964 to 2022.

FIG. 2: The relative number of winter days with Luleå as a reference.

Considering the quality of the data, it is important to note that while Uppsala and Luleå have temperature measurements taken every hour Lund has measurements with varying intervals during the considered time period (2, 3, and 5 times per day). With fewer measurements the uncertainties in the number of winter days increases, since the calculations of average daily temperatures become much less precise.

AVERAGE MONTHLY TEMPERATURE DIFFERENCE BETWEEN LUND AND LULEÅ

The main goal of this sub-project was to look into the monthly temperature difference between Lund and Luleå over a the same time period. Throughout the data analysis plots were prepared for average monthly temperature throughout the common time period for which the data was present which is 1949-01-01 to 2022-12-31. If the temperatures stayed consistent over this years the data can compiled into average temperature difference for every month.

Data Preparation

The data was first cleaned using the shell script for cleaning the data. The a new script `cut_to_size.sh` was written in order the cut the data to same date range. The script takes in data multiple data sets and ask for a start and end date. Produces a new file with the header from the original file and with the data in between the provided start and end date. Both of the data sets were cut down to common range.

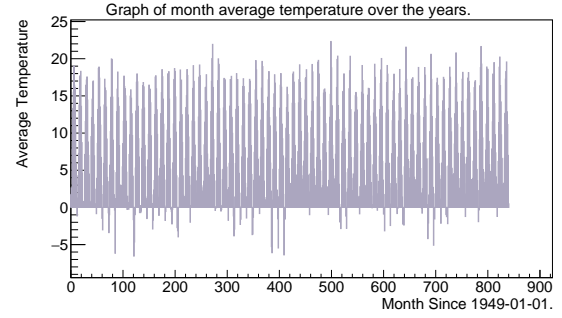
Data Analysis Using C++

All of the data analysis and plotting using C++ was is done in the `tempDelta.cxx` script. During the data analysis the data was first organized into two maps of `std::string` dates and double total temperatures in `totalTemperaturePerDay(std::string file)` method and another map of `std::string` dates and int measurement amount in the `measurementsPerDay` method. Those maps were then used to find the average temperature each day from multiple measurements for each day this was again organized into a map and then formatted to use `date::year_month_day` as the key value. That was done in the `averageTemperaturePerDay` and `averageTemperaturePerDayFormatted` methods. This was used to calculate the monthly average temperature in the `averageTemperaturePerMonth` method. The output map was used to prepare the first set of plots for average monthly temperature in Luleå and Lund in the `PlotData` method. Then the temperature difference was calculated in the `GetTemperatureDelta` method and plotted in `PlotDelta` delta. Lastly the data was compiled to find the average temperature difference for each month in `GetAverageTotalDeltaByMonth` method and plotted in `PlotDeltaByMonth`.

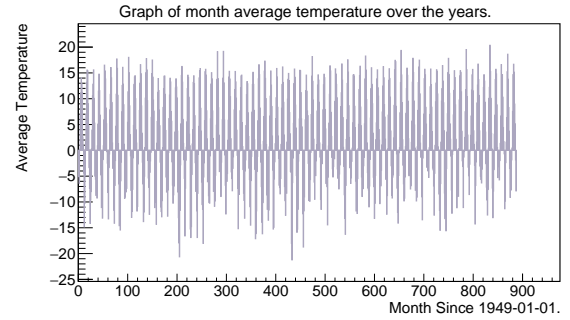
Results

The first results are just the average monthly temperatures over the years for both cities. Looking at this it can be clearly seen that the average temperature in Lund is pretty consistently above 0 throughout the year while the

temperature in Luleå varies significantly from summer to winter.



(a) Monthly average temperature from 1949-01-01 in Lund.



(b) Monthly average temperature from 1949-01-01 in Luleå.

FIG. 3: Monthly average temperatures in Lund and Luleå.

The second result was from looking at the temperature difference between the two cities as $Lund - Luleå$.

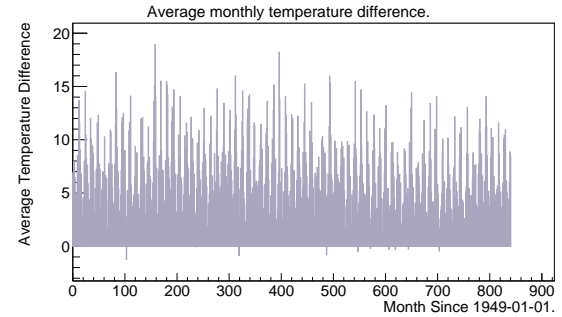


FIG. 4: Temperature delta Lund - Luleå from 1949-01-01.

The final result of the sub-project is the average temperature delta between the cities for each month. This result shows that the temperature difference is greatest in the cold month and goes to its minimum during July. The reason for this is that as already discussed before the temperature in Lund stays relatively consistently positive throughout the year while Luleå changes drastically during the year meaning that the temperature difference is

smallest during the warm month of the year.

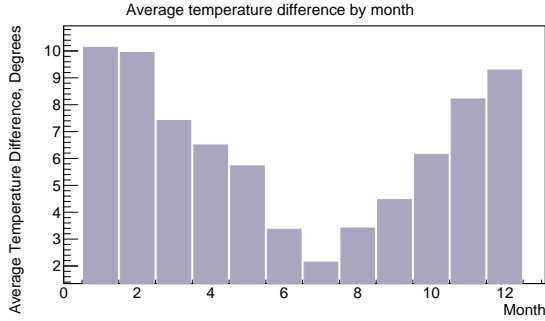


FIG. 5: Monthly Temperature Delta.

HOTTEST AND COLDEST DAYS OVER THE YEARS

This sub projects goal was to try to find out the day which was likeliest to be the hottest or coldest day by analyzing data provided by SMHI. Specifically the data regarding Lund, Uppsala and Luleå during the time period 1949-2023. The approach to this goal was to take out the days which had the hottest/coldest temperature from the data and then process it so that it could be entered into a plot after which one could analyze the results. The processing was done with both a bash script and and c++ functions. The plotting was done using the ROOT framework. Similar to previous sections there will be three sections, the first one discussing the data preparation, the second one discussing the C++ script and the ROOT plotting and the last one discussing the results.

Data pre processing

The data was, after being cleaned using bash script mentioned above, run through another bash script called MaxMinPerYear.sh. This script picks goes through every year after 1949 and picks out the hottest and coldest days from these years. It also removes the year from the data as only the month and day matters for our purpose. Since bash doesn't handle numbers the script picks out temperatures for the relevant year and sorts them which allows us to bypass any need to compare numbers as the bottom and top result is the highest and lowest temperatures.

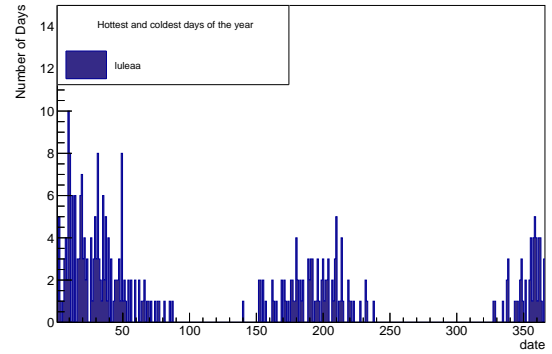
C++ and plotting

The data processing using C++ is done using the functions MinMax map and MinMax hist fill in MinMax-Temp. The first function takes a csv file cleaned using aforementioned method as input. The function then converts the dates into one of the numbers between 1-366 corresponding to the date and returns it as an `jint intj` map. The second function takes the map from the first

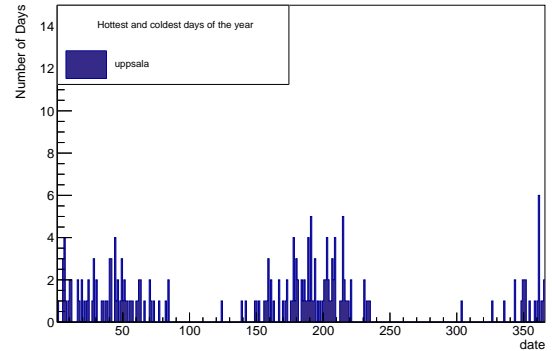
function and adds the data into a histogram. The plotting is then done by MinMax hist draw and the plot is saved as a pdf called MinMax.pdf.

Results

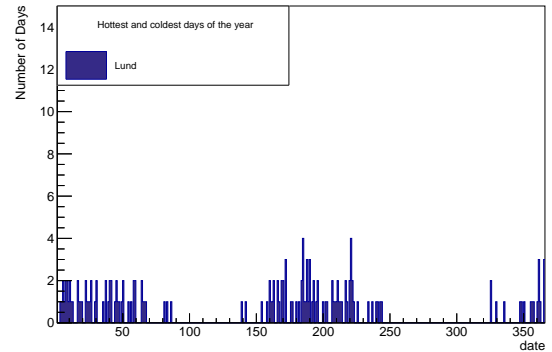
The resulting plots are histograms over the warmest and coldest days over the years since 1949 in Luleå, Uppsala and Lund. Looking at the histograms one can see that for luleå the hottest day is about 27:th of July and the coldest is the 8:th of January. On the histogram for uppsala the hottest day is the 9:th of July and the 1:st of August and the coldest day 29:th of december. For the Lund histogram one can see that the hottest day is first of July and 7:th of August and the coldest day the 28:th and 31:st of December.



(a) Maximum and Minimum temperature Luleå



(b) Maximum and Minimum temperature Uppsala



(c) Maximum and Minimum temperature Lund

As one can see from the figures the farther south you go the harder it gets to predict the hottest and coldest days which is interesting. The one consistent pattern that can be spotted is that the end of July and beginning of August is the likeliest to contain the days that are the hottest and the whole month of January is the likeliest to contain the coldest day by far.

CONCLUSIONS AND OUTLOOK

Number of Winter Days

The key findings for the number of winter days sub-project were the decreasing moving averages, which are an indication of global warming. This feature was clear in the Uppsala and Luleå datasets, but much harder to discern in the Lund dataset. Concerning the relative number of winter days, using Luleå as a reference, Lund showed higher variations than Uppsala. 2009 was the year with the highest number of winter days, during the time period 1964-2022, in both the Lund and Uppsala datasets. This year also had the highest relative number of winter days when comparing Lund and Uppsala to Luleå. Possible improvements to the data analysis process include replacing the maps with a new class containing the year and number of winter days as variables. This would make the code far more transparent for any potential readers. Another improvement would be to not have a hard-coded time period, which would make the program far less case specific.

Average monthly temperature

The results for the average monthly temperature part of the project turned out nicely. The main takeaway in terms of the analysis was the fact that the temperature difference between Lund and Luleå is lowest during July, the hottest month of the year and increases during the colder month. The data analysis tooling developed for this part of the experiment can work on any two data sets that were formatted in the same way and have a common time period that could be used for analysis. As a possible

improvement for the tooling the performance of the C++ program could likely be improved significantly if, instead of a large amount of functions traversing through the maps and passing them to one another, there would be a single function that does the calculations and averaging straight away.

Hottest and coldest days

While finding a date where the temperature was likely the hottest or the coldest was not successful it was still narrowed down to a small area of dates that were more likely to be the hottest. Some interesting patterns were found though. All of the datasets had a hot peak near late June and early August and the coldest day was more likely to be close to new years eve. Another very interesting pattern is that the further south we go the harder it gets to predict peak and bottom temperature. This could be due to Lund not having as many measurements as the other datasets though. There are some improvements to be made such as highlighting which part is the hottest and which is the coldest in the histograms. Another improvement could be to add markers to more easily see the peaks on the histogram and also making the date easier to read. It is currently hard to say a date with precision without digging in the data. The code in itself could be improved by changing the if statements to switch cases and removing legacy code which would improve legibility. It might also be interesting to create a function that could create a histogram that collects data from several datasets creating a combined map in order to get a highest and lowest temperature over the entire country.

* Emil Boman: nat14ebo@student.lu.se

† Henrik Olsson: he4511ol-s@student.lu.se

‡ Pavel Oshchepkov: paul.oshchepkov@icloud.com

[1] Sveriges meteorologiska och hydrologiska institut, *Ladda ner meteorologiska observationer*, SMHI's öppna data, URL: <https://www.smhi.se/data/meteorologi/ladda-ner-meteorologiska-observationer/#param=airtemperature-Instant,stations=core> .