CS2035 - Assignment 4 - 2017

London Weather Analysis

Out: Wednesday, March 22th, 2017
In: Friday, April 7th, 2017 at 11:55pm via Owl

Introduction

This assignment requires you to do simple (and **primitive**) analysis on London weather data from 1941 to 2013. Is the seasonal weather in London becoming warmer? Is there more or less seasonal snowfall, snow on the ground, precipitation and rain? You are given the weather data, constructed mostly from a combination of data measurements made at the London International Airport (1941-2005), from an automated weather station (2006-2013) and from a volunteer (2003-2006). This data is contained in a mat file `london_weather_1941_2013.mat`, available on the course webpage.

This file has the following data: year, month, element and data for up to 31 days (data element 32 specifies the number of days), saved in a 4D array, named **climate**. The data is for years 1941 to 2013, 12 months a year (29 days in February in leap years), 7 data elements (types of data) followed by the data for the specified element and month for some year. There are 7 data elements stored for most days:

1. max daily temperature degrees (0.1C)
2. min daily temperature degrees (0.1C)
3. ave daily temperature degrees (0.1C)
4. total rainfall (0.1mm)
5. total snowfall (0.1mm)
6. total precipitation (0.1mm)
7. snow on the ground (1.0cm)

Sometimes data elements are missing (not recorded or lost). In this case, `nan` (not a number) was recorded as the value. Note that you cannot use `nan` numbers in your calculations (any calculation with `nan` is `nan`). Thus, to read and compute the mean of the temperature and snowfall data for a particular `month` and `year`, we would use:

```
% read the number of days in the month for year
LAST=climate(year,month,3,32) % or LAST=climate(year,month,5,32)
% Read 1:LAST elements of data.
temperature_data(1:LAST)=squeeze(climate(year,month,3,1:LAST));
snowfall_data(1:LAST)=squeeze(climate(year,month,5,1:LAST));
% copy only non-nan corresponding temperature and snowfall
% data to the same variables
temperature_data=temperature_data(~isnan(temperature_data) & ~isnam(snowfall));
snowfall_data=snowfall_data(~isnan(temperature_data) & ~isnam(snowfall));
% average the non-nan data and save on array average_data.
% The temperature element is 1 in this array
monthly_mean_data(year,month,1)=mean(temperature_data(:));
monthly_mean_data(year,month,3)=mean(snowfall_data(:));
```

Note that the condition `~isnan(temperature_data) & ~isnam(snowfall)` guarantees that temperature and snowfall data for array elements where both elements are not `nan` are copied: thus, if `temperature(6)` is `nan` and `snowfall(6)` is 20, then neither element is copied. In this way, `temperature_data` and `snowfall_data` are guaranteed to be the same size, with corresponding data in each element.

Perform the following two tasks for this assignment:

1. Task 1: Compute the mean temperature, rainfall, precipitation, snowfall and snow on the ground values (elements 3, 4, 5, 6 and 7). Plot all the average data for the 12 months and 1942 to 2013 years using **surf**. Use `X` and `Y` as `meshgrid(1:12,1942:2013)`. Thus `surf(X,Y,squeeze(monthly_mean_data(year,month,1)))` and `surf(X,Y,squeeze (monthly_mean_data(year,month,3)))` produces 3D surface plots shown in Figures 1 and 2. `colorbar` was used to display the colour bar and smoothing shading was accomplished using `shading interp`.

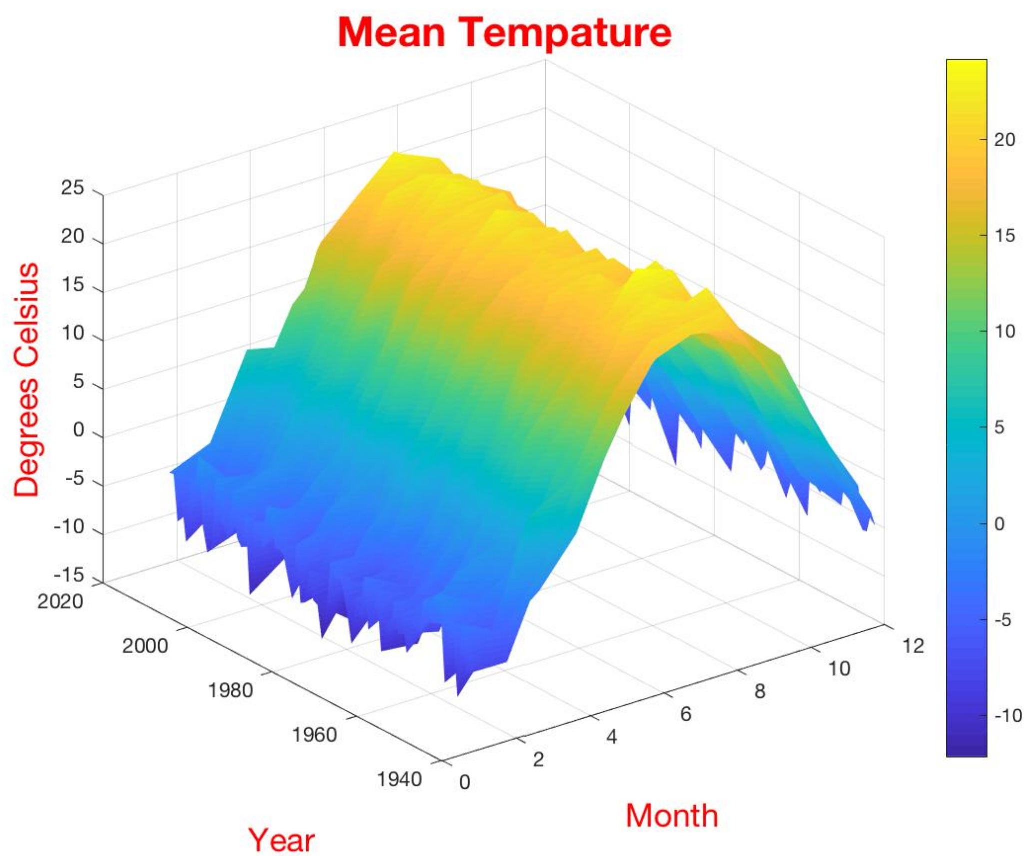

Figure 1: Mean temperature for each month for the years 1942 to 2013.

Critically analyze the plots to make sure they make sense. Figure 1 shows the mean temperature. We can see the middle months are the largest at up to 25 degrees Celsius. The months are the ends of the surface for months 1-2 and 11-12 are the coldest. For example, January is a cold month while July is a hot month. Figure 2 show the mean snowfall. Note that there is no snowfall for June to September and little snow for May and October. The titles and labels are displayed in red.

2. Task 2: Compute the Pearson's correlation coefficient, r, for vectors `x` and `y`:

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2) + \epsilon}}$$

Here n is the number of non-`nan` elements in `x` and `y`. ϵ is `eps` (machine epsilon). It

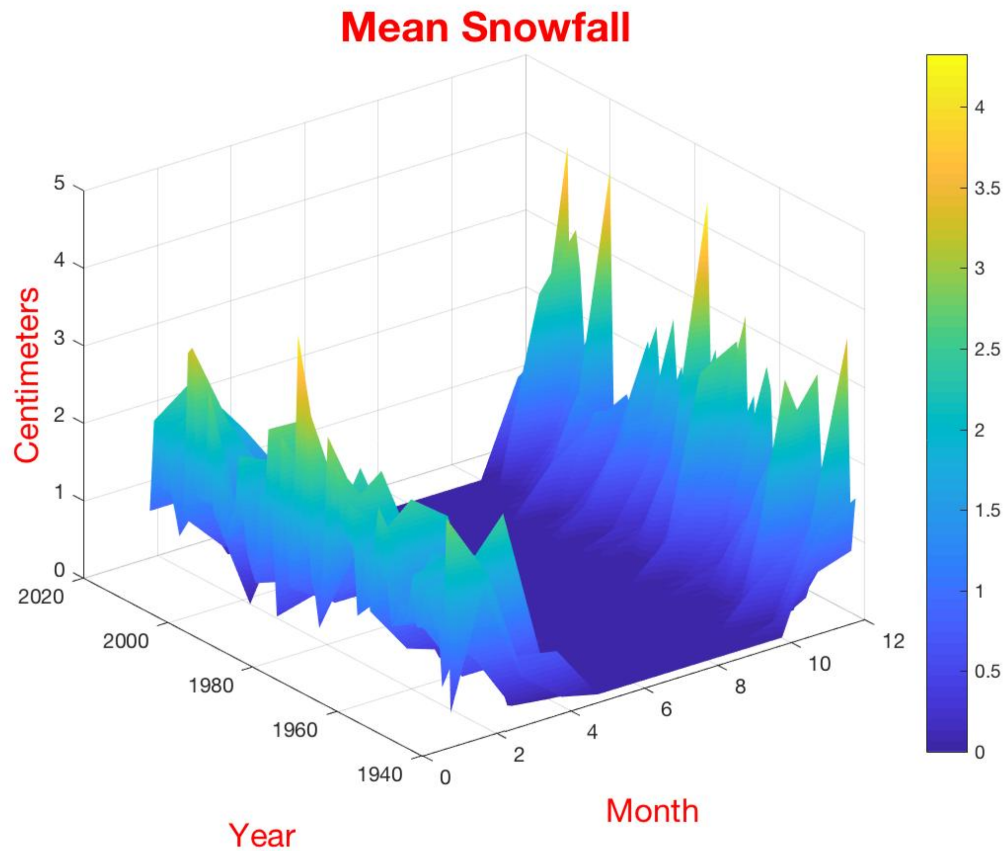

Figure 2: Mean snowfall for each month for the years 1942 to 2013.

prevents zero division if all elements in either or both `x` and `y` are zero. $r \in [-1, 1]$. When r is around 1, `x` and `y` are highly correlated, i.e. there is a linear relationship between `x` and `y` with positive slope. When r is around zero, `x` and `y` are not correlated. When r is around -1, `x` and `y` are highly anti-correlated, i.e. there is a linear relationship between `x` and `y` with negative slope.

Consider the correlation of 6 sets of monthly data:

(a) temperature versus precipitation

(b) temperature versus snowfall

(c) temperature versus snow on the ground

(d) precipitation versus snowfall

(e) rainfall versus snowfall

(f) snow on the ground versus snowfall

For `x` and `y` being the non-`nan` values, compute Pearson's correlation coefficient for each month and plot these values as surface graphs.

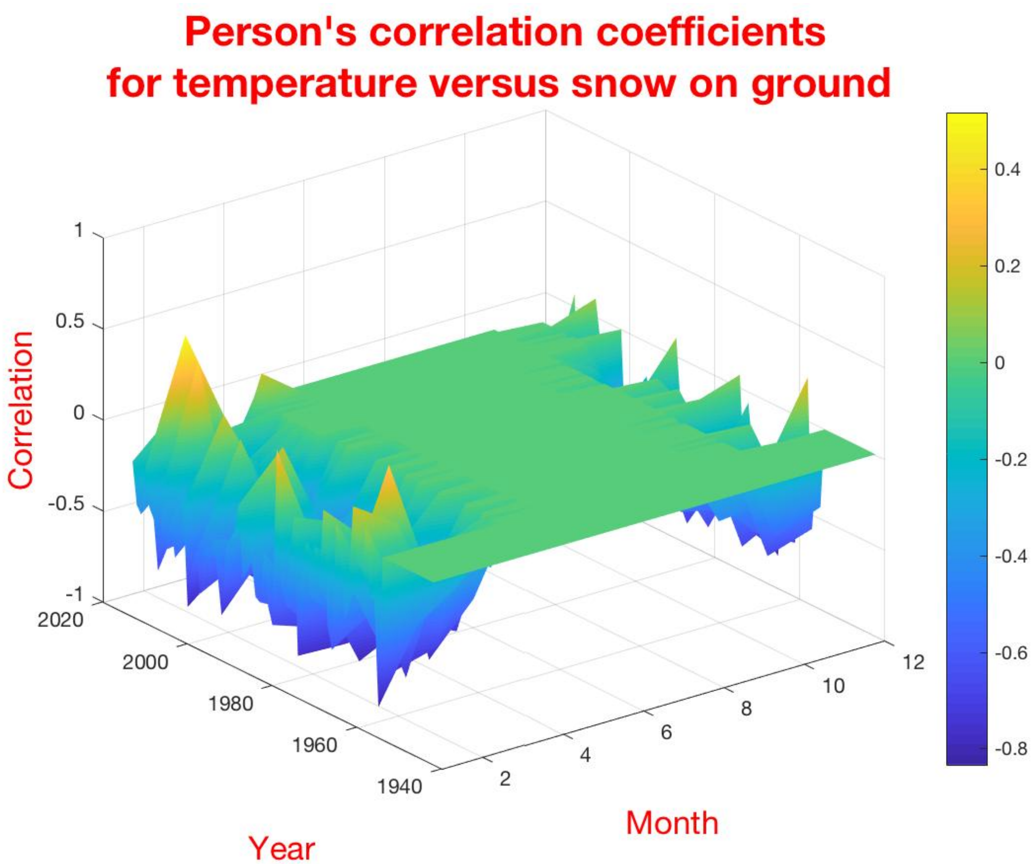

Figure 3: Pearson's correlation coefficients for temperature versus snow on the ground for each month for all years 1942 to 2013.

Again, critically analyze your plots to make sure they make sense. `colorbar` produced the colour bars, `shading interp` produced the smooth shadings and `axis` was used to control the tick labellings. Figure 3 shows the Pearson's correlation coefficients for temperature versus snow on the ground for the middle months are 0 (no correlation) because there is no snow in the summer. On the other hand, when the average temperatures are below zero (for

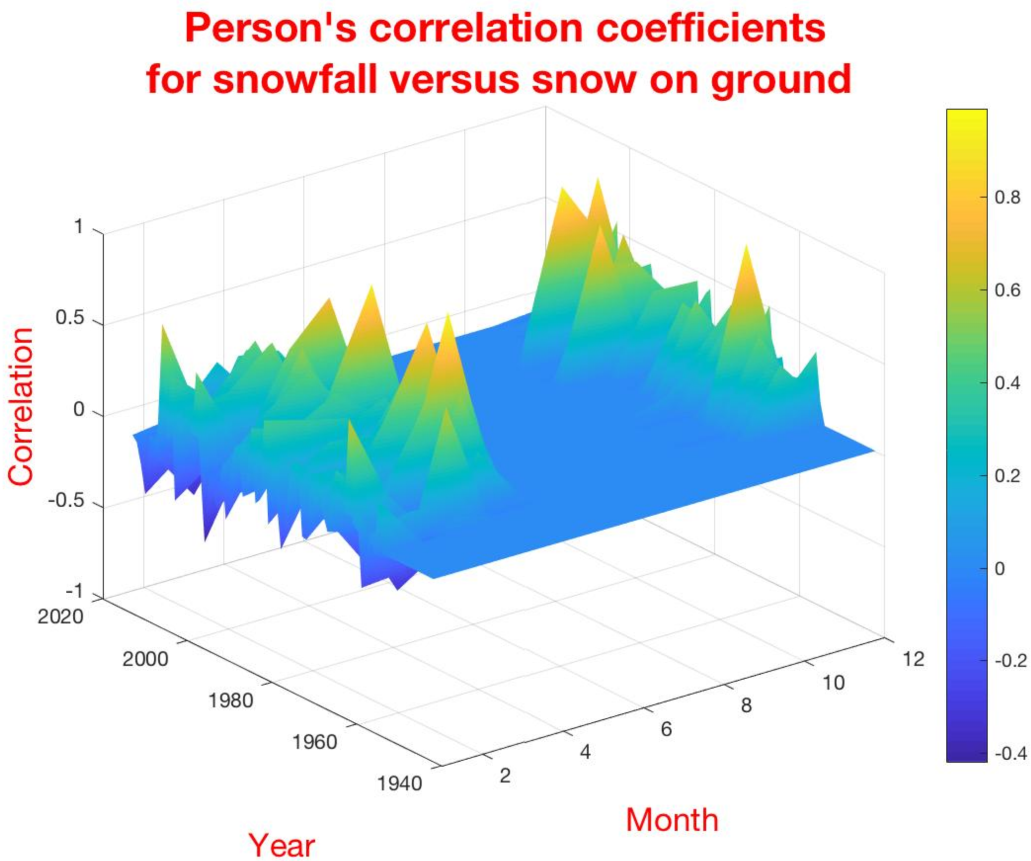

Figure 4: Pearson's correlation coefficients for snowfall versus snow on the ground for each month for all years 1942 to 2013.

example, in December, January or February) there may be snow on the ground. Figure 4 shows that the Pearson's correlation coefficients for snowfall versus snow on the ground for the middle months are zero as there is no snow or snow on the ground. Again, not the labels and titles are displayed in red.

You are provided with a shell of the program you must write, `ass4_shell_2017.m`, on the course webpage. This program has many useful predefined variables (which may or may not be useful) and it reads the initial climate data for you. Fill in your code in the places indicated to complete this assignment.