**Pattern Analysis of Earth’s Surface temperature- A Detailed study using FFT, Spectrogram and wavelets**

**Project Members:**

* + 1. Induraj Pudhupattu Ramamurthy
    2. Kofi Dramanne
    3. Murali Krishna

Contents

[**Abstract:** 3](#_Toc6590443)

[**I.** **Introduction:** 4](#_Toc6590444)

[**II.** **Objectives of the project** 4](#_Toc6590445)

[a. General Objective 4](#_Toc6590446)

[b. Specific Objective 4](#_Toc6590447)

[**III.** **Scope of the project** 4](#_Toc6590448)

[**IV.** **Results and discussions:** 4](#_Toc6590449)

[**V.** **Future enhancement:** 16](#_Toc6590450)

[References: 17](#_Toc6590451)

[Appendix 18](#_Toc6590452)

[Matlab code 18](#_Toc6590453)

# **Abstract:**

It’s always been a foresight for the environmentalist to find if the global warming is really happening or not! Still there are lots of researches being conducted by most renowned organizations such as NASA, GISS, NOAA, BEST to find the rate at which the globe is warming up. With other independent researchers proving that the Global warming is a Hoax, we are not entering into this debate of if the global warming is a real Hoax or not, rather we focus on the finding the pattern’s in the repetition of earth’s temperature. In order, to understand the climate dynamics, it is necessary to find the patterns and trends in the temperature time series data, to reveal the patterns and trends associated we intend to use FFT, spectrogram and wavelet on the dataset by using Matlab as our programming environment.

# **Introduction:**

EISTI - As an engineering school officially recognized by the State and made competent by the CTI (French Engineering accreditation institution) the EISTI’s vocation trains future engineers in Mathematics and Computing. Having this in mind, our team choose this topic of analyzing the patterns in the global surface temperature to getting better understanding of the course work and also to find the facts related to the temperature patterns by using Matlab coding.

The earth’s surface temperature rather than exhibiting the stationary trend, it has a non-stationary trend which makes the analysis of earth’s surface temperature little more difficult! We have non stationary time series data on earth’s surface temperature for 265 years! In each year, at the sampling interval of 1 month’s data’s had been recorded, so making the total number of samples to be 3192. The dataset was made available was in kaggle.com [1] by the Berkeley earth- an independent non-profit organization!

Rather than analyzing individually data’s gathered across every major city, we consider the average mean of such temperature collected across the globe! We by the use of fast Fourier transform try to find the dominant frequency and the rate at which it is appearing! But since in fast Fourier transform we have only the frequency information, which alone might not be useful for the analysis of pattern and trend, we use the spectrogram in order to reveal the time and frequency information associated with our data.

# **Objectives of the project**

## General Objective

The general objective of our project is to analyze the global surface temperature rather than analyzing the global climatic temperature. By analyzing we intend to find any useful pattern that can help us gain better knowledge and insight on the earth’s temperature cycle if such exists.

## Specific Objective

The specific objective of our project is

* To detect the Dominant frequency that makes up our time series data
* To have better understanding of the frequency with respect to time
* To detect the patterns in the time series data by the use of the MODWT wavelet
* To analyze the pattern
* To perform useful interpretations

# **Scope of the project**

As the topic of analyzing earths global warming rate by itself is vast composing many different factors like co2 emission, rate of industrialization, earths climatic temperature, earth’s surface temperature etc., we restrict our self in analyzing only the earth’s surface temperature alone trying to find useful patterns!

# **Results and discussions:**

The 265 years of 3192 samples of data we intent to analyze is shown in the figure 1 indicating its clear non stationarity trend.

To reveal the frequency component of the signal we take fast Fourier transform on this time series data, and we find that the dominant frequency/harmonic is 3.25x10^-8 Hz as shown in figure 3. By this fact, we deduce that this dominant frequency/harmonic repeats itself every (1 / 3.25x10^-8 = 30769230.769 seconds) 🡺 11 months, 21.5 days, thereby confirming the seasonality in our data.

The power spectral density of this time series data is given in figure 4, Power spectral density function (PSD) shows the strength of the variations(energy) as a function of frequency. In other words, it shows at which frequencies variations are strong and at which frequencies variations are weak. The unit of PSD being energy per frequency(width), the plot reveals that maximum energy is concentrated on the frequency of 325nHz (3.25x10^-8 Hz). The welch PSD also ensures that it reduces noise in the estimated power spectra in exchange for reducing the frequency resolution.

A discrete wavelet transform (DWT) is wavelet transform in which the wavelets are discretely sampled. As with other wavelet transforms, a key advantage it has over Fourier transforms is temporal resolution: it captures both frequency and location information (location in time). In order to detect patterns that are not visible in the raw data, we have used the finer octave bands obtained by the wavelets for the multiresolution analysis! The patterns detected by applying MODWT (maximum overlap discrete wavelet transform) is shown in the figure 5

Rather than analyzing all temperature ranges, I focused on analyzing the patterns corresponding to the highest temperature of 15.0 degree Celsius to 15.6 degree Celsius which is mentioned in the upcoming discussions as 5.0 degree Celsius to 5.6 degree Celsius as the temperature is centered around the mean.

In-order to analyze the signal better, it was necessary to consolidate the same temperatures in groups. But while grouping, With the temperature precision of 5 decimals, I had around 1230 groups of temperature patterns, and with precision of 2 decimals the temperature group, I had 51 groups which was again difficult to analyze. So finally by rounding the temperature to one decimals, only seven groups of highest temperature range were attained which was easy to analyze the pattern.

The box plot in figure 10 shows the temperature distribution of each group of temperatures across time span. The 5.6 degree Celsius (group 7), 5.0 degree Celsius (group 1) are more sparsely distributed across the time span similar to 5.5 degree Celsius (group 6). The most densely distributed temperature recording is for 5.2-degree Celsius and 5.4 degree Celsius.

It is very clear that the temperature of 5.6 degree Celsius is having the median at 1847 with average time of occurrence to be 15.0625 years and standard deviation of 9.6779 years as depicted in figure 16. It is evident that the occurrence of 5.6 degree Celsius is distributed equally and is repeated 17 times between 1750 and 1751. Similarly, the box plot for the temperature groups namely 5.0, 5.1, 5.2, 5.3, 5.4, 5.5 degree Celsius are shown in the boxplot figures 11 to 15. With these findings of patterns across the temperature range, we can expect the next appearance of the particular temperature to occur with the average rate in the near future.

But it is to be noted that the last occurrence of the temperature 5.1 degree Celsius (15.1 degree Celsius) was in 2003 and at the mean occurrence rate of 10.913 years, it should have reappeared around 2013, But didn’t appear! Similarly, other temperature ranges 5.2, 5.3, 5.4, 5.5, 5.6 had last occurrence in 1987, 1972, 1980, 1999, 1991 with mean occurrence rate of 8.72 year, 12.235 year, 11.947 years, 10.7273 year, 15.06 years, these highest temperature ranges never reappeared in recent years despite their mean occurrence rate says that they should have appeared!

This proves the fact put forth by many renowned scientists that the earth stopped warming in this one and half decade and that the global warming is a hoax [2,3,4]!

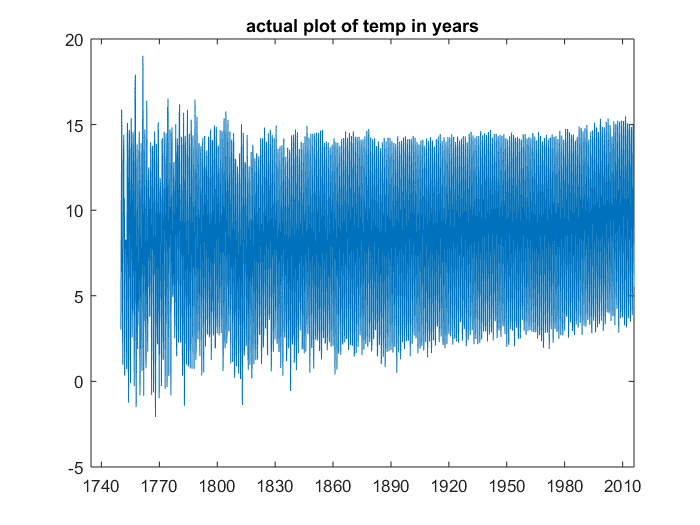


Figure 1 Actual temperature data distributed over 265 year time span

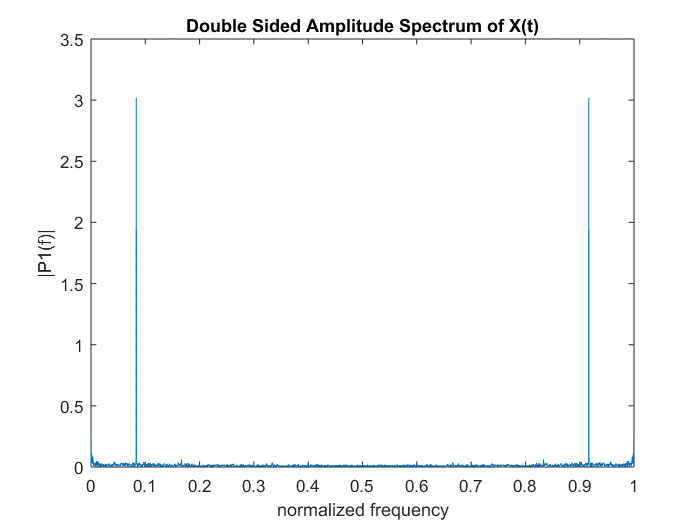


Figure 2 Fast Fourier Transform Output in Normalized Frequency

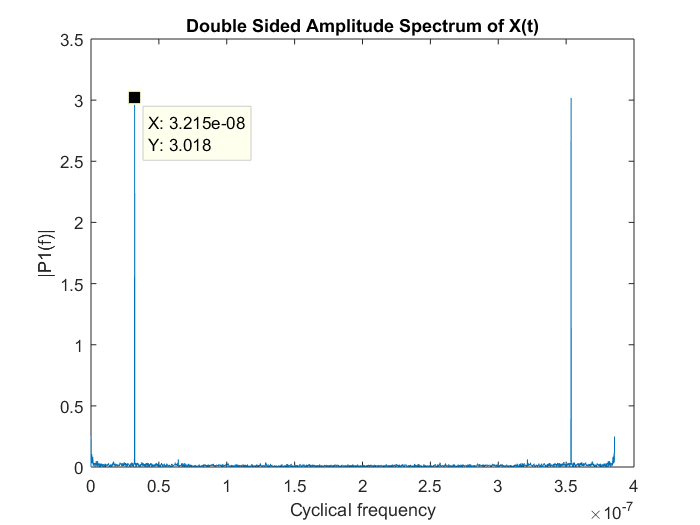


Figure 3 Fast Fourier Transform Output in cyclical frequency

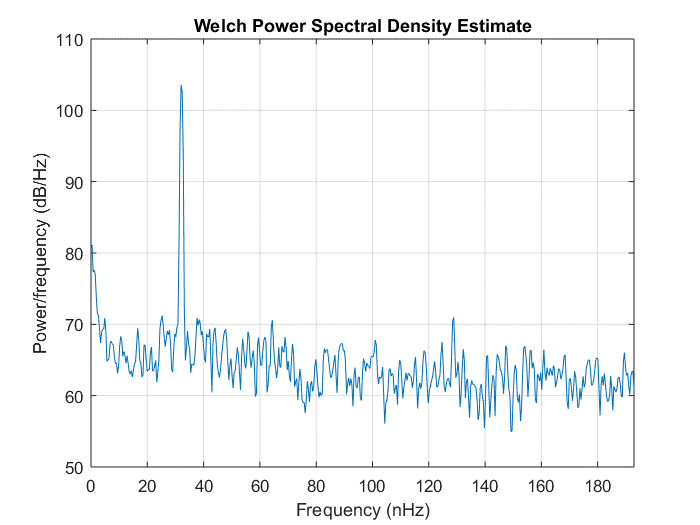


Figure 4 power spectral density estimate

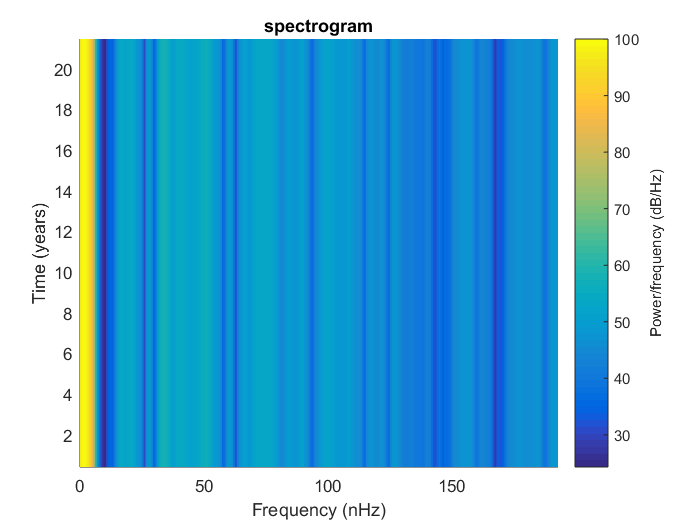


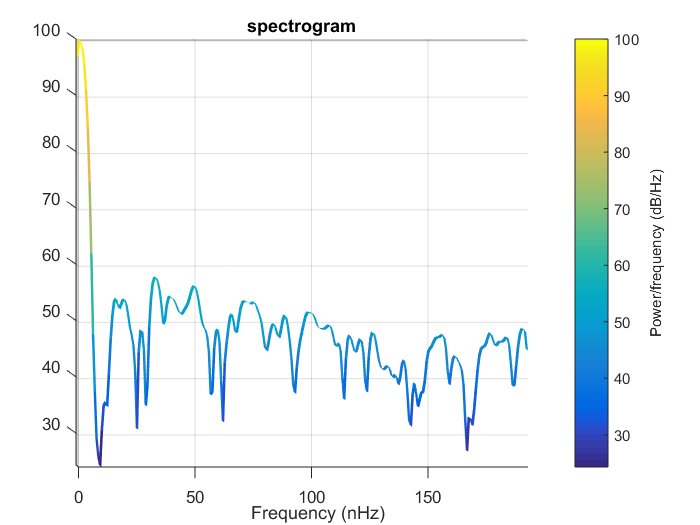
Figure 5 Spectrogram Spectrum

Figure 6 Spectrogram spectrum shown with better frequency and time resolution

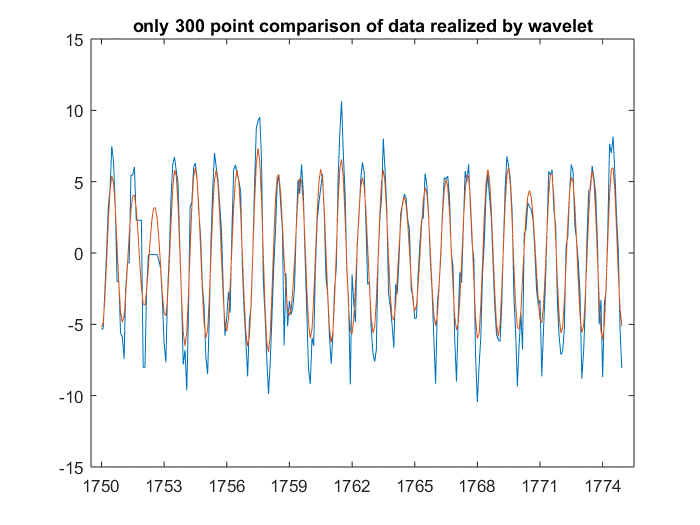


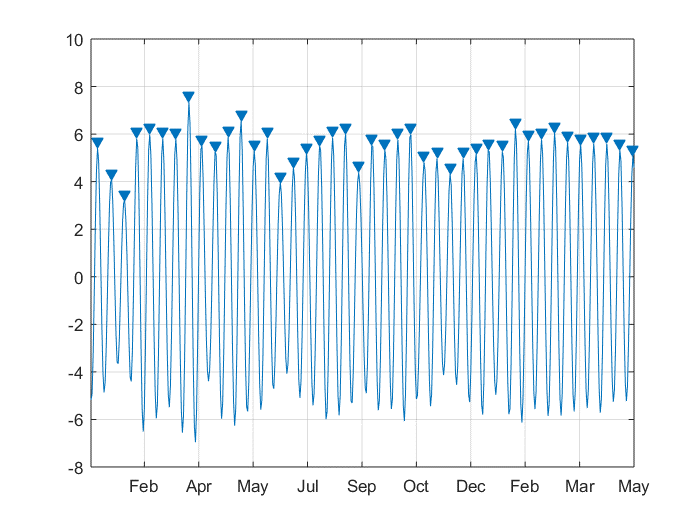
Figure 7 Modwt wavelets Coefficient compared with actual data

Figure 8 Determination of Peaks

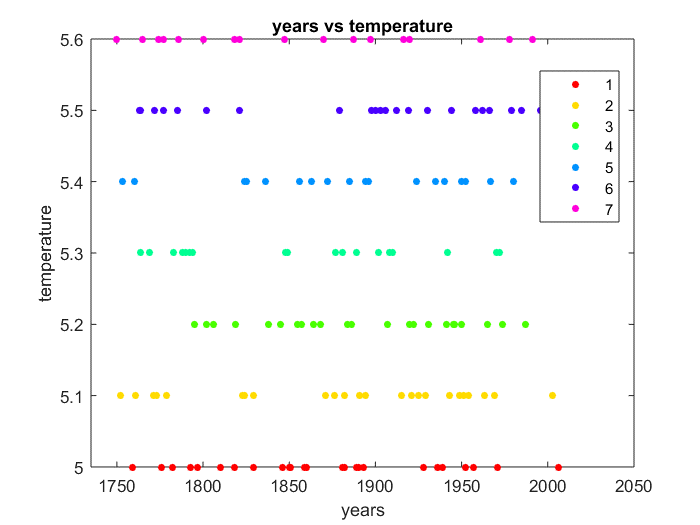


Figure 5 Repetitive Temperature Patterns found

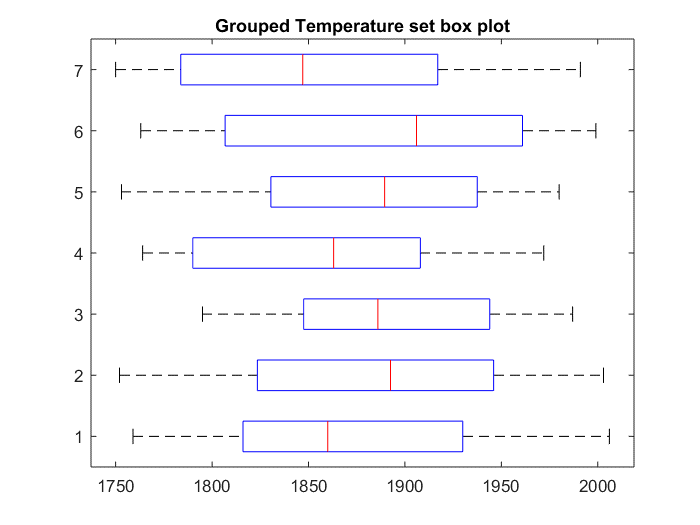


Figure 6 Box Plot of the different temperature range

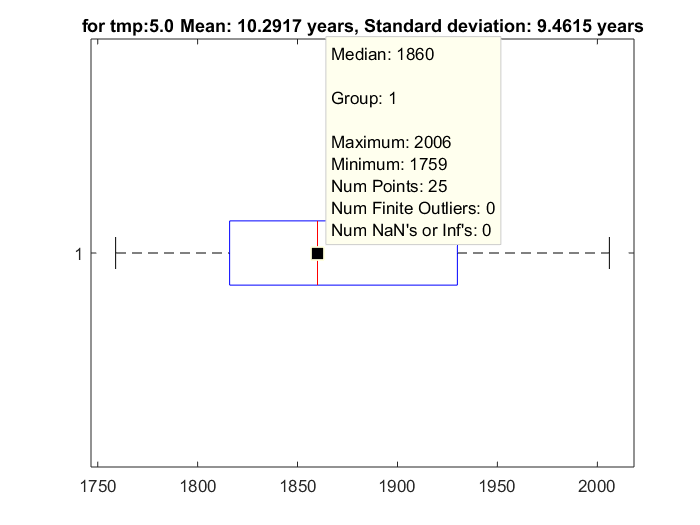


Figure 7 Box plot for Temperature of 5.1 degrees

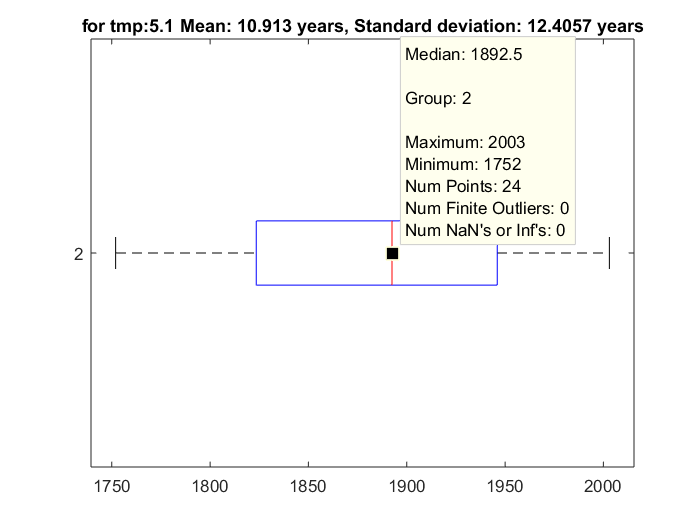


Figure 82 Box plot for Temperature of 5.1 degrees

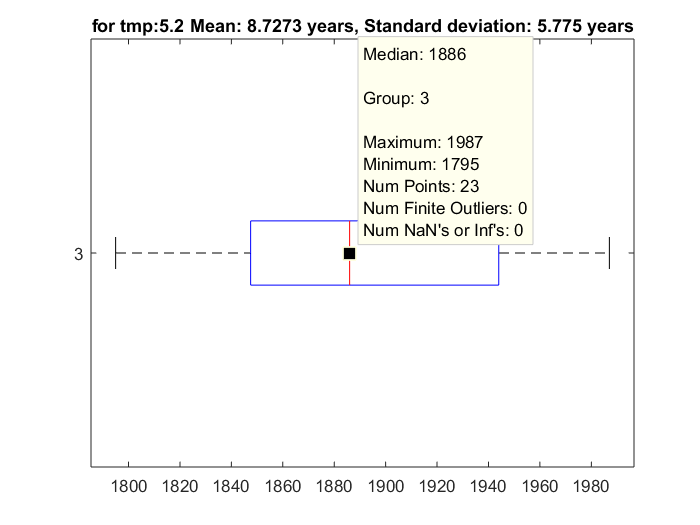


Figure 9 Box plot for Temperature of 5.2 degrees

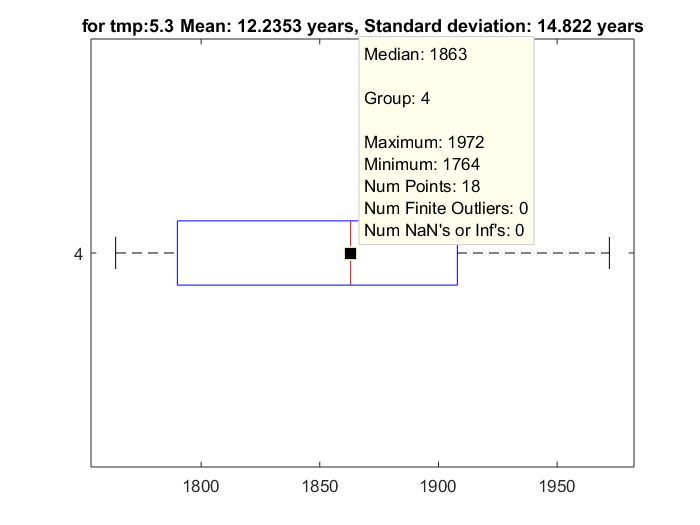


Figure 10 Box plot for Temperature of 5.3 degrees

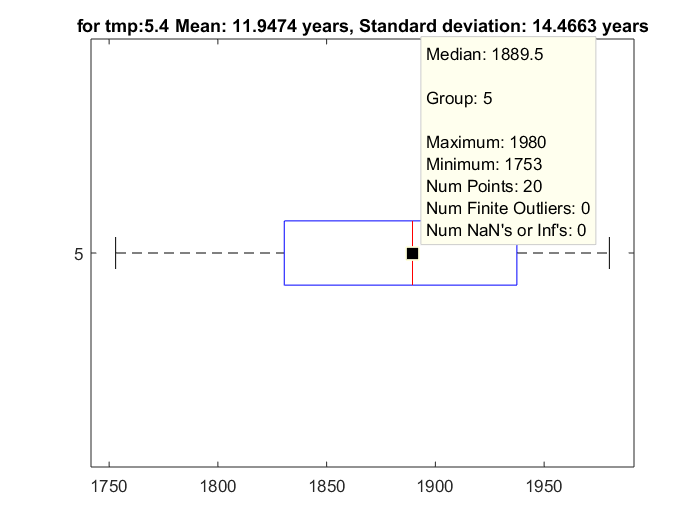


Figure 11 Box plot for Temperature of 5.4 degrees

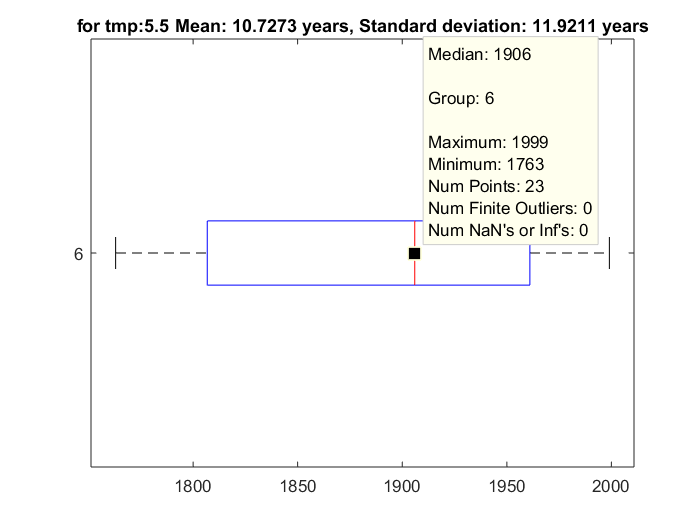


Figure 12 Box plot for Temperature of 5.5 degrees

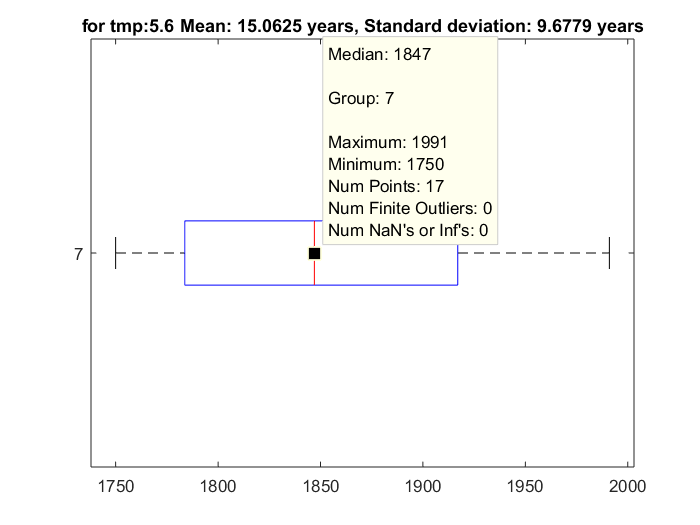


Figure 13 Box plot for Temperature of 5.6 degrees

# **Future enhancement:**

To extend the functionality of this project in terms of prediction, i.e., In-order to predict the next occurrence of highest temperature very accurately, one can use any of the prediction methodologies like regressions or by applying the inverse Fourier transform to the dominant frequency one can construct the same signal that can be used to get good insight on the future events.

Further by correlating the earth’s surface temperature along with the earths climatic temperature, one can find facts to prove if the earth is warming with sufficient level of confidence!

# References:

1. <https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data>
2. <https://www.youtube.com/watch?v=n-Z0eG1pKhA>
3. <https://www.youtube.com/watch?v=mqejXs7XgsU>
4. <https://www.wwf.org.uk/updates/10-myths-about-climate-change>
5. Analysis of monotonic greening and browning trends from global NDVI time-series , Rogierde Jong, Sytzede Bruin, Allardde Wit, Michael E.Schaepman, David L.Dent
6. How natural and anthropogenic influences alter global and regional surface temperatures: 1889 to 2006 - <https://doi.org/10.1029/2008GL034864>
7. Forecasting phenology under global warming by Inés Ibáñez , Richard B. Primack , Abraham J. Miller-Rushing , Elizabeth Ellwood , Hiroyoshi Higuchi , Sang Don Lee , Hiromi Kobori and John A. Silander

# Appendix

## Matlab code

datax=VarName1;

datay=VarName2;

%datay\_mean=datay;

datay\_mean=datay-mean(datay);

class(datax)

dataxx=datetime(datax);

class(dataxx)

%fs=1000;

figure;

plot(dataxx,datay)

datetick('x', 'keepticks','keeplimits')

title('actual plot of temp in years')

Y = fft(datay\_mean);

L = length(datay\_mean);

T = 30\*24\*60\*60

Fs = 1/(T);

f\_adj1=(0:L-1)/L;

figure;

plot(f\_adj1,abs(Y/L)) title('Double Sided Amplitude Spectrum of X(t)')

xlabel('normalized frequency')

ylabel('|P1(f)|')

f\_adj2=(0:L-1)\*Fs/L;

figure;

plot(f\_adj2,abs(Y/L))

title('Double Sided Amplitude Spectrum of X(t)')

xlabel('Cyclical frequency')

ylabel('|P1(f)|')

figure;

pwelch(datay\_mean,[],[],[],Fs) %% x which is 0.265 if multiplied by 1000 gives our actual frequency which you can verify with above fft

gmoddd=modwt(datay);

mragggg=modwtmra(gmoddd);

figure;

plot(dataxx(1:300,:),datay\_mean(1:300,:), dataxx(1:300,:),mragggg(3,1:300))

datetick('x', 'keepticks','keeplimits')

title('only 300 point comparison of data realized by wavelet')

figure;

plot(dataxx(:,:),datay\_mean(:,:), dataxx(:,:),mragggg(3,:))

datetick('x', 'keepticks','keeplimits')

title('Time series Full data realised by the use of wavelet')

figure

findpeaks(mragggg(3,1:500),'MinPeakDistance',10);

datetick('x', 'keepticks','keeplimits')

[ppp,loc]=findpeaks(mragggg(3,:),'MinPeakDistance',2);

%avgtempyearpeak= years(mean(diff(ppp))) % so the next temp rise is every 5.1956 years

avgmeantempcycle=years(mean(diff(loc)))

yearnn=year(dataxx);

yearnnn=unique(yearnn);

% for k=1:length(loc)

% i(k)=loc(k);

% end

%

% for z=1:length(loc)

% dataxnew(z)= datax(i(z));

% end

% dataxnewx=datetime(dataxnew);

%

% figure

% plot(dataxnewx,ppp);

% datetick('x', 'keepticks','keeplimits');

% title('listing the peaks height and the year at which they occured')

figure;

spectrogram(ppp,flattopwin(266),10,[],Fs);

% spectrogram(ppp,flattopwin(266),10,[],266);

title('spectrogram');

ppp\_round=round(ppp,2);

[c,ia,ic]=unique(ppp\_round);

ppp\_trans=transpose(ppp\_round);

%comb= [yearnnn,ppp\_trans,ic];

c\_tran= transpose(c);

[cc,ian,icc]=unique(ic);

for i=1:length(cc)

count=0;

for j=1:length(icc)

if(cc(i)==icc(j))

count=count+1;

appearance\_count(i)=count;

end

end

end

appearance\_count\_tran=transpose(appearance\_count);

new\_comb=[cc,c\_tran,appearance\_count\_tran];

loc\_index=1;

for date\_i=1:length(new\_comb(:,3))

if (new\_comb(date\_i,3) > 1)

repmore\_da(loc\_index)=new\_comb(date\_i,1);

repmore\_val(loc\_index)=new\_comb(date\_i,2);

repmore\_rep\_rate(loc\_index)=new\_comb(date\_i,3);

loc\_index=loc\_index+1;

end

end

repmore\_da\_tran=transpose(repmore\_da);

repmore\_val\_tran=transpose(repmore\_val);

repmore\_rep\_rate\_tran=transpose(repmore\_rep\_rate);

repmore\_comb=[repmore\_da;repmore\_val;repmore\_rep\_rate]

repmore\_comb\_tran=transpose(repmore\_comb);

datay\_rounded=round(datay,1)

jjj=1;

for zz=1:length(repmore\_da\_tran)

chk\_val=repmore\_val\_tran(zz);

kk=1;

while(kk<length(dataxx))

if (datay\_rounded(kk)==chk\_val)

new\_matched\_date(jjj)= dataxx(kk);

new\_matched\_data(jjj)= chk\_val;

jjj=jjj+1;

end

kk=kk+1;

end

end

new\_new\_matched\_date=transpose(new\_matched\_date);

class(new\_new\_matched\_date)

new\_new\_matched\_data=transpose(new\_matched\_data);

class(new\_new\_matched\_data)

conv\_date=cellstr(new\_new\_matched\_date);

dat\_num= datenum(new\_new\_matched\_date);

extracted\_year= year(new\_new\_matched\_date);

%g1\_dummy\_year\_temp\_group=[conv\_date,new\_new\_matched\_data];

%g1\_table=table(g1\_year\_temp\_group);

%g1\_sorted\_year\_temp\_g= sort(g1\_table)

%plot(g1\_year\_temp\_group(:,1),g1\_year\_temp\_group(:,2))

%scatter(extracted\_year,new\_new\_matched\_data)

% new\_new\_matched\_date1=cellstr(new\_new\_matched\_date);

group\_data=findgroups(new\_new\_matched\_data)

g1\_year\_temp\_group= [extracted\_year,new\_new\_matched\_data,group\_data];

% time= datetime(check.date,'InputFormat','mm/dd/yyyy')

check =table;

check.date=new\_new\_matched\_date;

check.data=new\_new\_matched\_data;

check.group\_data=group\_data;

figure;

gscatter(g1\_year\_temp\_group(:,1),g1\_year\_temp\_group(:,2),g1\_year\_temp\_group(:,3))

title('years vs temperature')

xlabel('years')

ylabel('temperature')

dif\_in\_year= diff(check.date);

sorted\_rows=sortrows(check);

% figure;

% plot(sorted\_rows.date, sorted\_rows.data)

[sorted\_data]=sorted\_rows(:,2);

tab\_to\_cell=table2array(sorted\_data);

figure;

boxplot(g1\_year\_temp\_group(:,1),g1\_year\_temp\_group(:,3),'orientation','horizontal')

title('Grouped Temperature set box plot');

for\_scat\_plt\_1= [g1\_year\_temp\_group(1:25,1),g1\_year\_temp\_group(1:25,3)];

for\_scat\_plt\_2= [g1\_year\_temp\_group(26:49,1),g1\_year\_temp\_group(26:49,3)];

for\_scat\_plt\_3= [g1\_year\_temp\_group(50:72,1),g1\_year\_temp\_group(50:72,3)];

for\_scat\_plt\_4= [g1\_year\_temp\_group(73:90,1),g1\_year\_temp\_group(73:90,3)];

for\_scat\_plt\_5= [g1\_year\_temp\_group(91:110,1),g1\_year\_temp\_group(91:110,3)];

for\_scat\_plt\_6= [g1\_year\_temp\_group(111:133,1),g1\_year\_temp\_group(111:133,3)];

for\_scat\_plt\_7= [g1\_year\_temp\_group(134:150,1),g1\_year\_temp\_group(134:150,3)];

intYear1 = diff(for\_scat\_plt\_1(:, 1));

intYear2 = diff(for\_scat\_plt\_2(:, 1));

intYear3 = diff(for\_scat\_plt\_3(:, 1));

intYear4 = diff(for\_scat\_plt\_4(:, 1));

intYear5 = diff(for\_scat\_plt\_5(:, 1));

intYear6 = diff(for\_scat\_plt\_6(:, 1));

intYear7 = diff(for\_scat\_plt\_7(:, 1));

figure;

boxplot(for\_scat\_plt\_1(:,1),for\_scat\_plt\_1(:,2),'orientation','horizontal')

title(['for tmp:5.0 ','Mean: ' num2str(mean(intYear1)) ' years, Standard deviation: ' num2str(std(intYear1)) ' years']);

figure;

boxplot(for\_scat\_plt\_2(:,1),for\_scat\_plt\_2(:,2),'orientation','horizontal')

title(['for tmp:5.1 ','Mean: ' num2str(mean(intYear2)) ' years, Standard deviation: ' num2str(std(intYear2)) ' years']);

figure;

boxplot(for\_scat\_plt\_3(:,1),for\_scat\_plt\_3(:,2),'orientation','horizontal')

title(['for tmp:5.2 ','Mean: ' num2str(mean(intYear3)) ' years, Standard deviation: ' num2str(std(intYear3)) ' years']);

figure;

boxplot(for\_scat\_plt\_4(:,1),for\_scat\_plt\_4(:,2),'orientation','horizontal')

title(['for tmp:5.3 ','Mean: ' num2str(mean(intYear4)) ' years, Standard deviation: ' num2str(std(intYear4)) ' years']);

figure;

boxplot(for\_scat\_plt\_5(:,1),for\_scat\_plt\_5(:,2),'orientation','horizontal')

title(['for tmp:5.4 ','Mean: ' num2str(mean(intYear5)) ' years, Standard deviation: ' num2str(std(intYear5)) ' years']);

figure;

boxplot(for\_scat\_plt\_6(:,1),for\_scat\_plt\_6(:,2),'orientation','horizontal')

title(['for tmp:5.5 ','Mean: ' num2str(mean(intYear6)) ' years, Standard deviation: ' num2str(std(intYear6)) ' years']);

figure;

boxplot(for\_scat\_plt\_7(:,1),for\_scat\_plt\_7(:,2),'orientation','horizontal')

title(['for tmp:5.6 ','Mean: ' num2str(mean(intYear7)) ' years, Standard deviation: ' num2str(std(intYear7)) ' years']);