Three Decades of Monthly Temperatures in Virginia

In this project you will develop and fit models of monthly temperatures over three different decades for the state of Virginia. The files d7383.txt, d9303.txt and d1323.txt each contain 120 numbers each, the monthly temps between October 1973 to September 1983, October 1993 to September 2003, and October 2013 to September 2023, respectively. The data was collected from the National Centers for Environmental Information, a research dissemination arm of NOAA (https://www.ncei.noaa.gov/). You will use least squares to fit some models to monthly temperature measurements Because of the fact that the data has an obvious annual oscillation (warm in the summer, cold in the winter), a straight average is a poor fit to the data. We will investigate simple models and compare results with more complicated models, that try to take account of the annual oscillation.

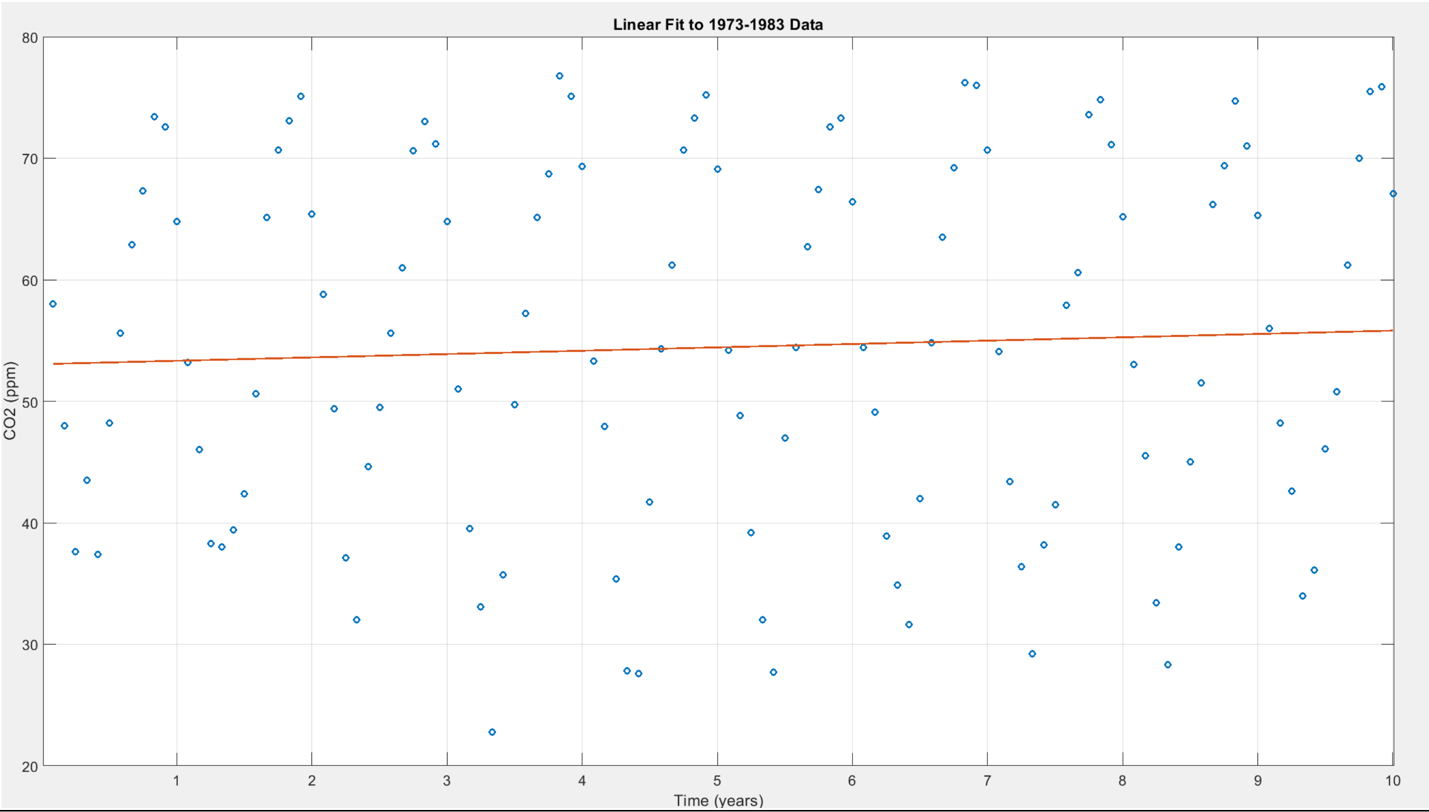
PART 1

1. Fit each of the three-monthly time series with the model y = c1 + c2t, where y denotes the monthly average temperature and t represents time in years, $0 \le t \le 10$. The MATLAB command t = (10*(1:120)/120)’; will define the time variable with units of years. The command load filename will load the data into your MATLAB window. Plot each of the three-time series along with the corresponding linear model, and compute the root mean squared error (RMSE) of each of the fits. Use the grid command after the plot to see grid markings.

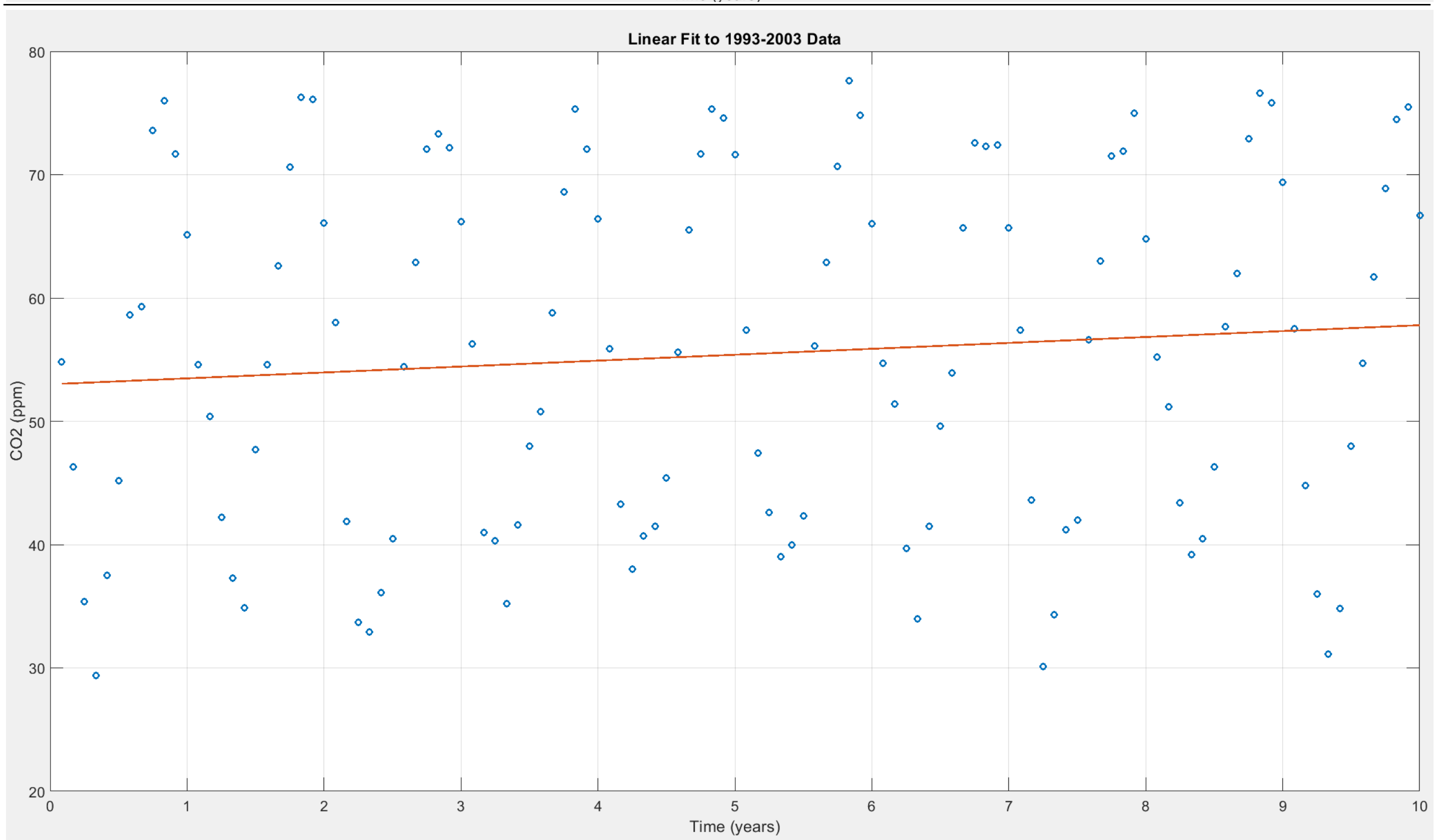
Solution:

1973-1983: c1 = 53.055350, c2 = 0.275798, **RMSE = 14.762034**
1993-2003: c1 = 53.001807, c2 = 0.479146, **RMSE = 14.065907**
2013-2023: c1 = 53.976092, c2 = 0.544246, **RMSE = 14.152635**
(Answers in bold)

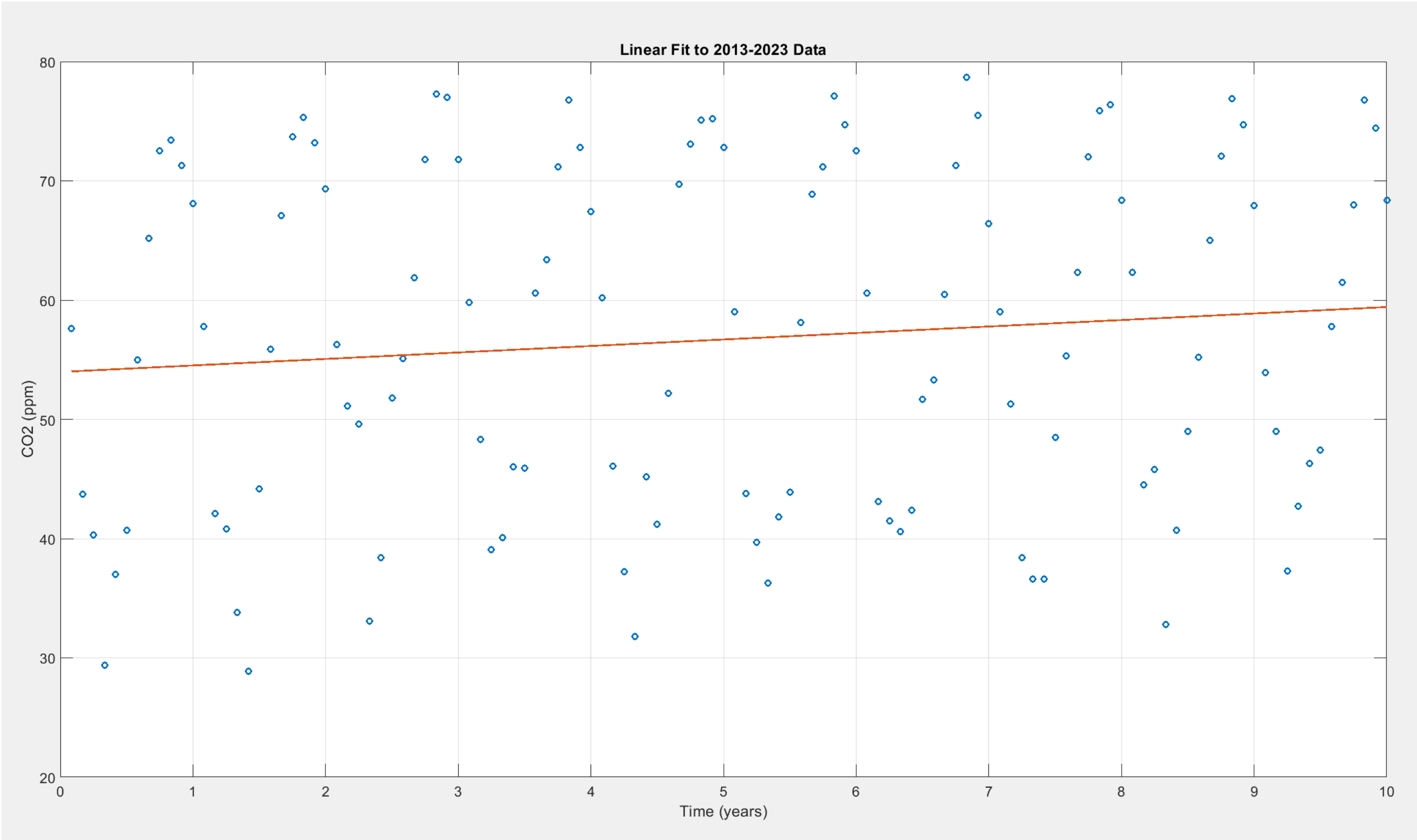
Plots:



Linear Fit to 1973-1983 Data
CO2 (ppm)
Time (years)
80
70
60
50
40
30
20
1
2
3
4
5
6
7
8
9
10



Linear Fit to 1993-2003 Data
CO2 (ppm)
Time (years)
80
70
60
50
40
30
20
0
1
2
3
4
5
6
7
8
9
10



Code:

```
format long
% Load the data
data1 = load('d7383.txt'); % 1973-1983 data
data2 = load('d9303.txt'); % 1993-2003 data
data3 = load('d1323.txt'); % 2013-2023 data

% Define the time variable
t = (10*(1:120)/120)'; % time in years

aa1=linear(t,data1,'1973-1983');
aa2=linear(t,data2,'1993-2003');
aa3=linear(t,data3,'2013-2023');

% Function to fit the linear model and plot
function fittedValues = linear(t, b, plotTitle)
   n = length(b);
   a = [ones(n,1) t];

   % Compute coefficients 'c' for the linear model using least squares
   c = (a'*a)\(a'*b);
```

```
    % Generate a finer time grid for plotting the linear model
    t1 = (10*(1:120)/120)';

    % Calculate the fitted values using the linear model
    fittedValues = c(1) + c(2)*t;

    % Calculate the RMSE
    RMSE = sqrt(mean((b - fittedValues).^2));

    % Plot the original data and the linear fit
    plot(t, b, 'o', t1, c(1) + c(2)*t1, 'LineWidth', 2);
    grid; set(gca, 'FontSize', 16); xlabel('Time (years)'); ylabel('Temperature');
    title(['Linear Fit to ', plotTitle, ' Data']);

    % Display coefficients and RMSE
    fprintf('%s: c1 = %f, c2 = %f, RMSE = %f\n', plotTitle, c(1), c(2), RMSE);
end
```

PART II:

It may be difficult to see the annual periodicity in the data from the plot in Step 1. To fit the annual cycle, repeat Step 1 with the trigonometric model $y = c1 + c2t + c3 \cos 2\pi t + c4 \sin 2\pi t$. Show the three plots, and analyze the new c2 and RMSE – how did they change with the more complex model? Do you consider the new c2 to be more or less accurate than the one in Step 1? (Asking for your opinion – I don't know what the right answer is.)

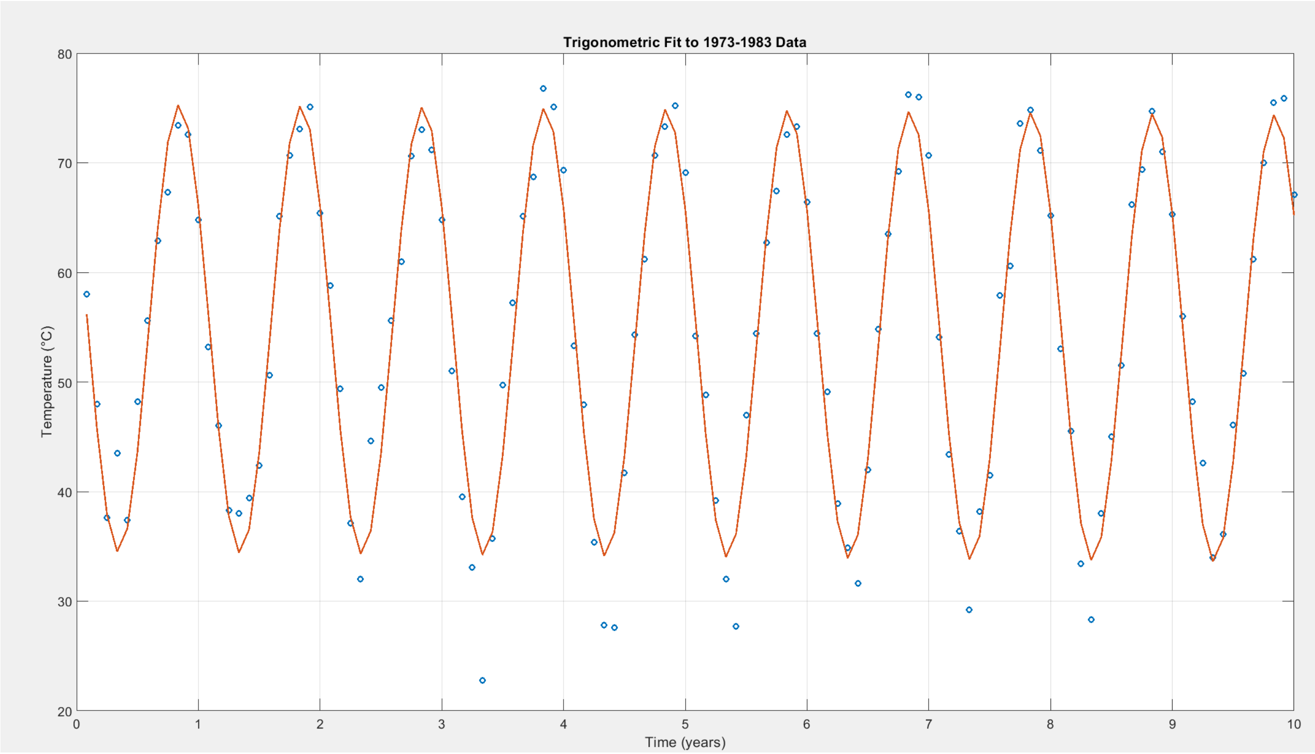
Solution:

(Answers in bold)

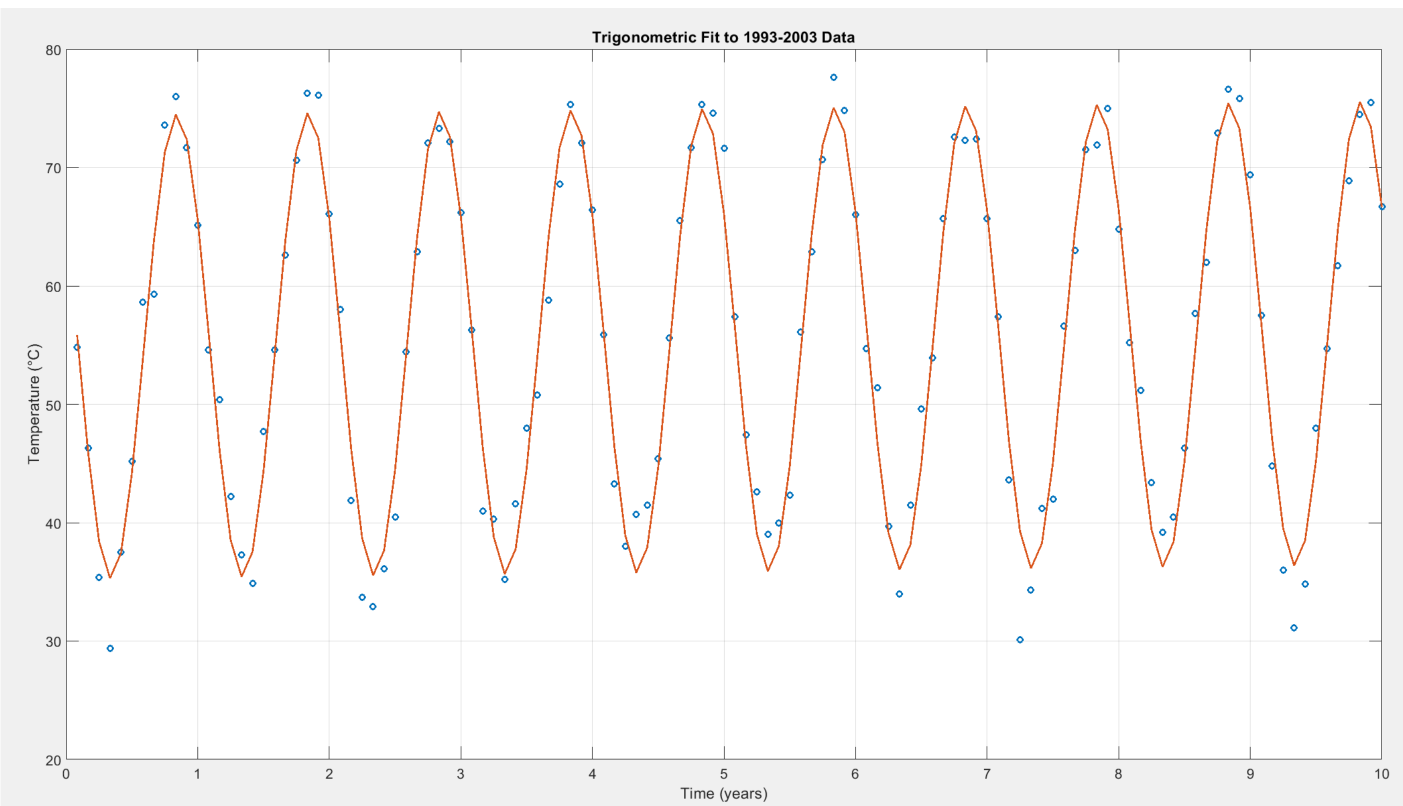
1973-1983: c1 = 54.942320, **c2 = -0.098477,** c3 = 11.270219, c4 = -17.036101**, RMSE = 3.234446**
1993-2003: c1 = 54.814798, **c2 = 0.119544,** c3 = 10.668363, c4 = -16.411068, **RMSE = 2.712694**
2013-2023: c1 = 55.775029, **c2 = 0.187432,** c3 = 10.998427, c4 = -16.173250, **RMSE = 3.176201**

The RMSE and C2 from this trigonometric model (part2) and linear model from part 1 are different. This model provides a much lower RMSE suggesting that this model is a better fit for this data. C2 is also much lower for the 3 data sets. These two suggest this model is more accurate to describe this type of data set. Which is expected because the data is based on temperature reading of each month for 3 decades. it makes sense that a cyclical model will be a much better fit as we have 4 different 4 seasons in VA (Spring, Summer, Fall and winter)

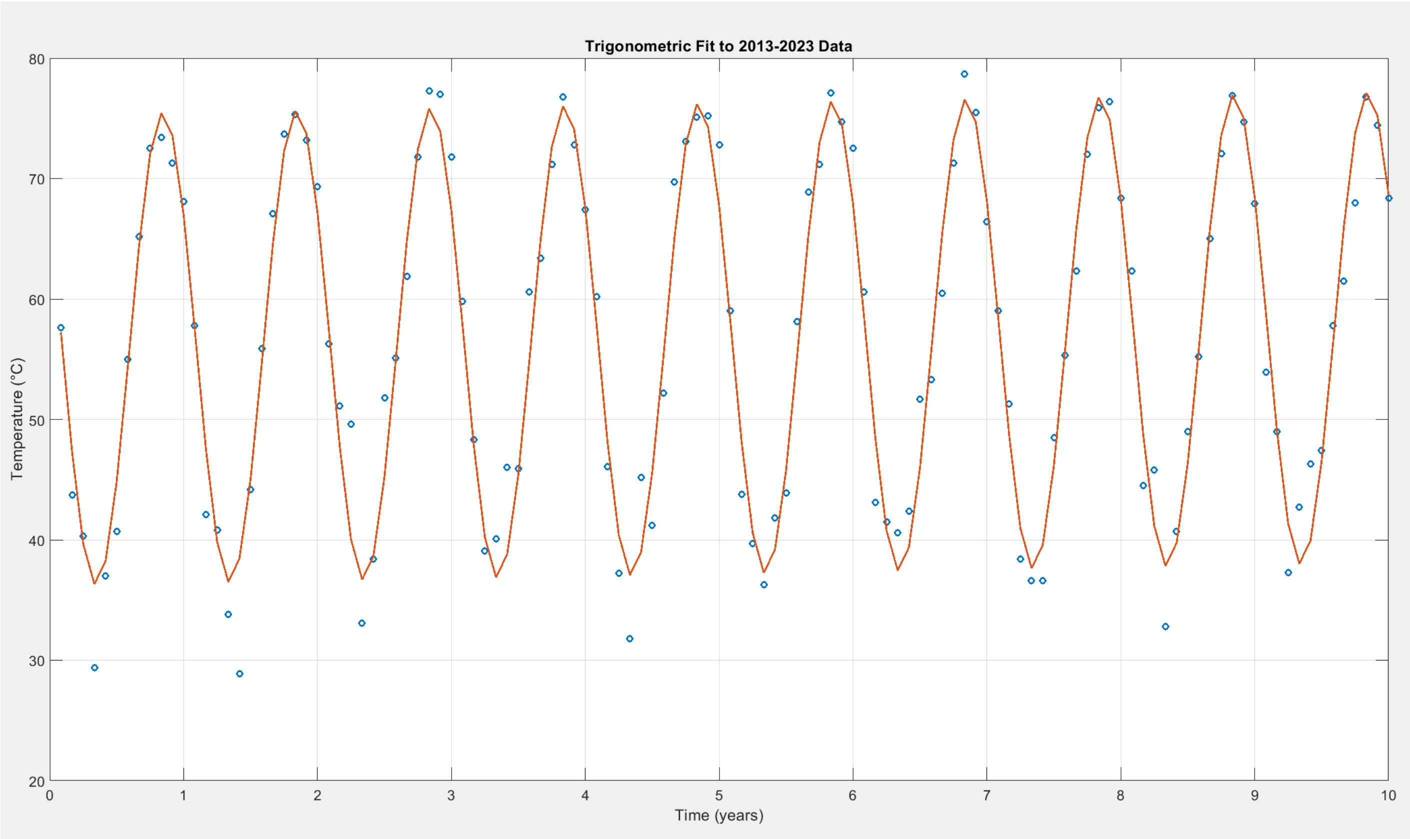
Plots:



Trigonometric Fit to 1973-1983 Data
Temperature (°C)
Time (years)
80
70
60
50
40
30
20
0
1
2
3
4
5
6
7
8
9
10



Trigonometric Fit to 1993-2003 Data
Temperature (°C)
Time (years)
80
70
60
50
40
30
20
0
1
2
3
4
5
6
7
8
9
10



Code:

```
data1 = load('d7383.txt'); % 1973-1983 data
data2 = load('d9303.txt'); % 1993-2003 data
data3 = load('d1323.txt'); % 2013-2023 data

% Define the time variable
t = (10*(1:120)/120)'; % time in years

% Fit and plot for each dataset
[coeffs1, RMSE1] = fitAndPlot(data1, t, 1);
[coeffs2, RMSE2] = fitAndPlot(data2, t, 2);
[coeffs3, RMSE3] = fitAndPlot(data3, t, 3);

% Display results
fprintf('1973-1983: c1 = %f, c2 = %f, c3 = %f, c4 = %f, RMSE = %f\n', coeffs1, RMSE1);
fprintf('1993-2003: c1 = %f, c2 = %f, c3 = %f, c4 = %f, RMSE = %f\n', coeffs2, RMSE2);
fprintf('2013-2023: c1 = %f, c2 = %f, c3 = %f, c4 = %f, RMSE = %f\n', coeffs3, RMSE3);
% Function to fit the model and plot
function [coeffs, RMSE] = fitAndPlot(data, t, figureNum)
  % Create new figure
  figure(figureNum);

  % Fit the model
  [fitresult, gof] = fit(t, data, 'a + b*x + c*cos(2*pi*x) + d*sin(2*pi*x)', 'StartPoint', [1, 1, 1, 1]);
```

```
    % Plot
    plot(t, data, 'o'); % original data
    hold on;
    plot(t, fitresult(t), '-r'); % fitted model
    title(sprintf('Temperature Data Fitting for Dataset %d', figureNum));
    xlabel('Time (years)');
    ylabel('Temperature');
    grid on;
    hold off;

    % Calculate RMSE
    RMSE = gof.rmse;

    % Extract coefficients
    coeffs = coeffvalues(fitresult);
end
```

PART III:

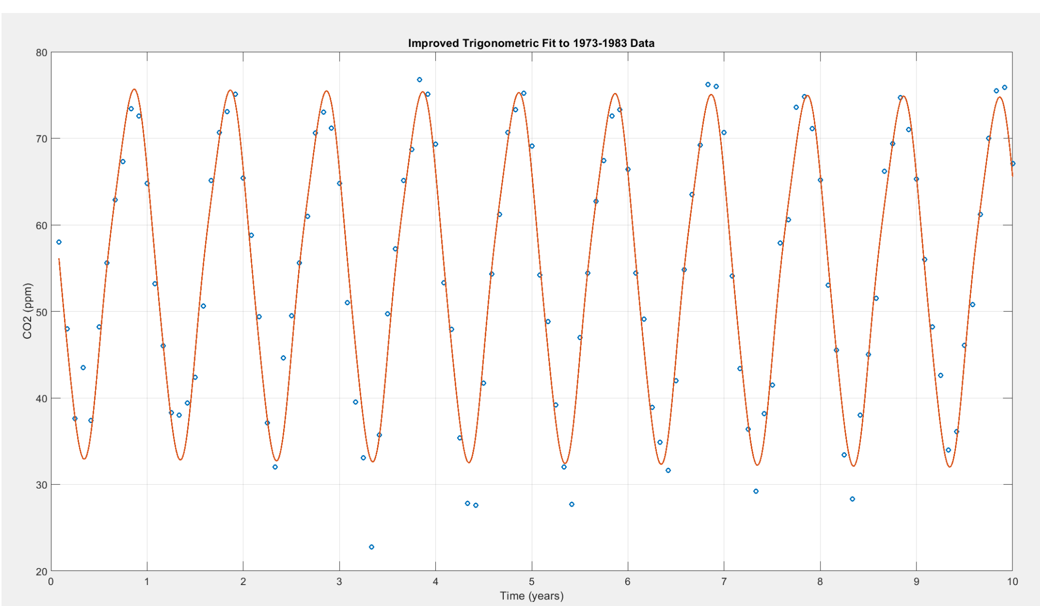
Repeat Step 2, but with the improved trig model y = c1 + c2t + c3 cos 2πt + c4 sin 2πt + c5 cos 4πt + c6 sin 4πt + c7 cos 6πt + c8 sin 6πt.

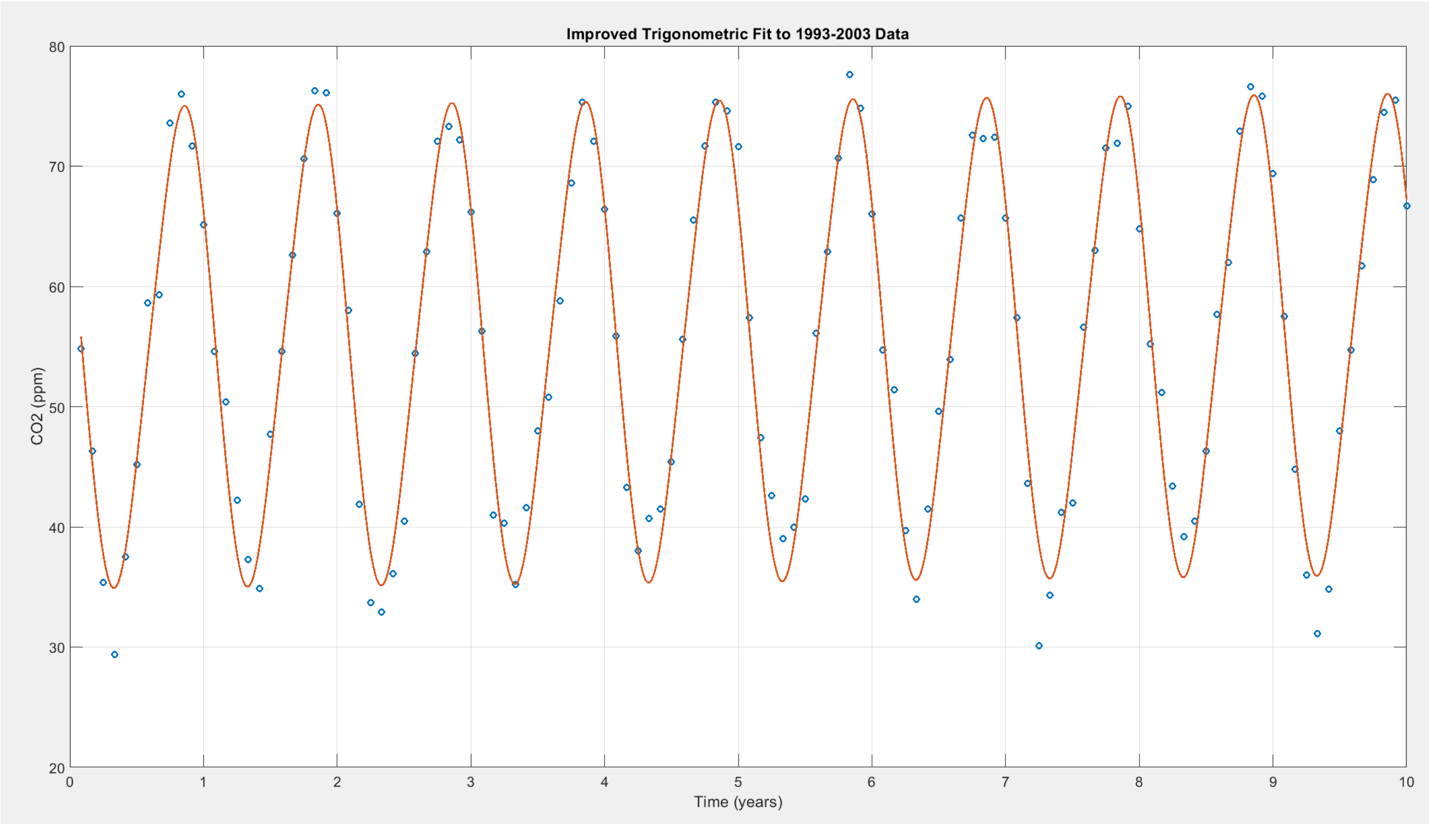
Solution:

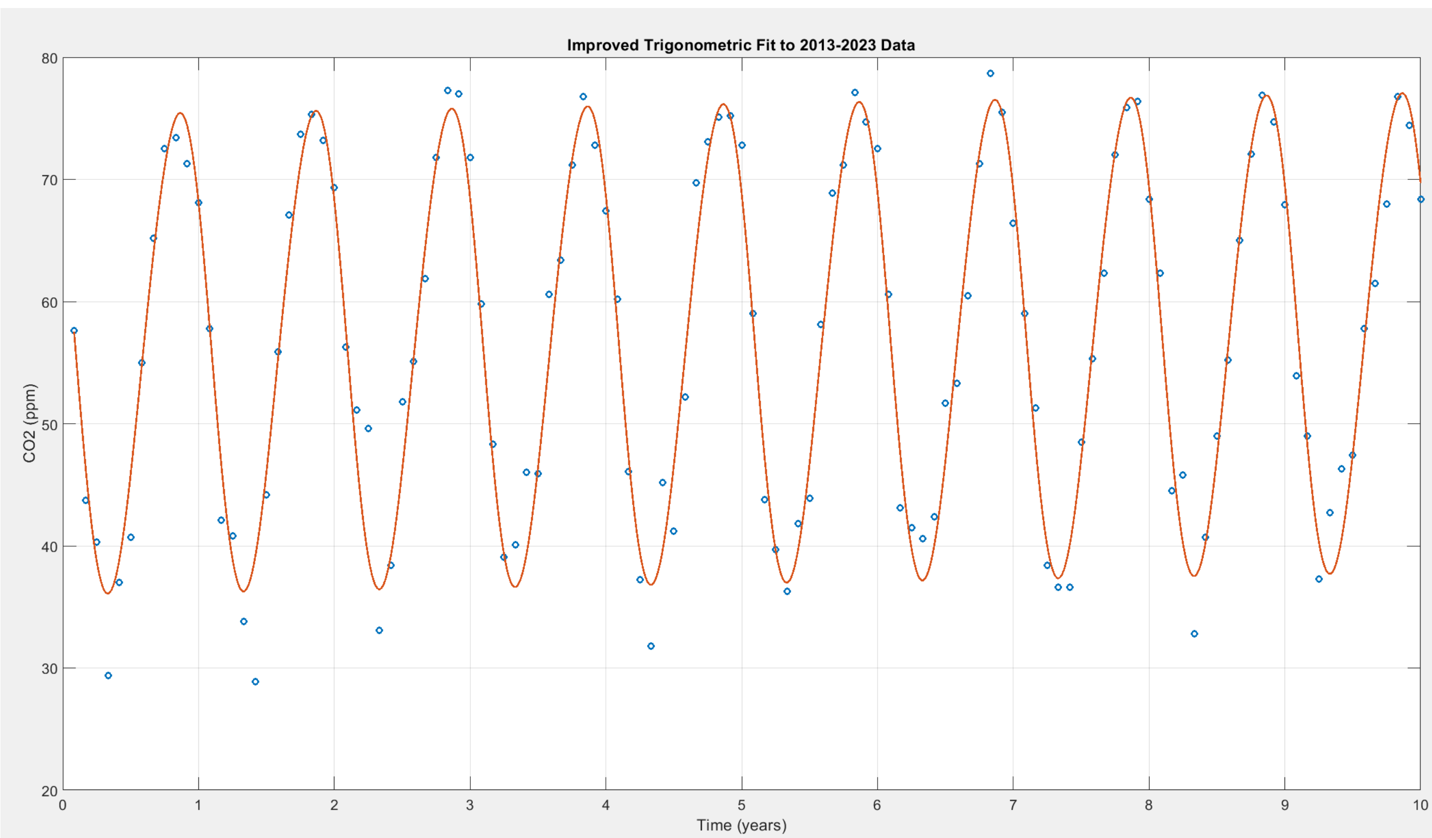
1973-1983: c2 = -0.101491, RMSE = 3.041010
1993-2003: c2 = 0.111935, RMSE = 2.598847
2013-2023: c2 = 0.179666, RMSE = 3.086630

The RMSE and C2 from this improved trigonometric model (part3) and linear model from part 1 and less complex trig model in part2 are different. This model provides a slightly lower RMSE than in part2 suggesting that this model improved. The addition of higher harmonic terms slightly improves its ability to fit the data. C2 is also much lower for the 3 data sets than in parts 1 and 2 of this report.

Plots:




Improved Trigonometric Fit to 1993-2003 Data
CO2 (ppm)
Time (years)
80
70
60
50
40
30
20
0
1
2
3
4
5
6
7
8
9
10


Improved Trigonometric Fit to 2013-2023 Data
CO2 (ppm)
Time (years)
80
70
60
50
40
30
20
0
1
2
3
4
5
6
7
8
9
10

Code:

```
format long
% Load the data
data1 = load('d7383.txt'); % 1973-1983 data
data2 = load('d9303.txt'); % 1993-2003 data
data3 = load('d1323.txt'); % 2013-2023 data

% Define the time variable
t = (10*(1:120)/120)'; % time in years

[coeffs, rmse] = trigModelImproved(t, data1, '1973-1983');
[coeffs2, rmse2] = trigModelImproved(t, data2, '1993-2003');
[coeffs3, rmse3] = trigModelImproved(t, data3, '2013-2023');

function [coefficients, RMSE] = trigModelImproved(t, b, plotTitle)
    n = length(b);

    % Create the design matrix for the trigonometric model
    a = [ones(n, 1), t, cos(2*pi*t), sin(2*pi*t), ...
         cos(4*pi*t), sin(4*pi*t), cos(6*pi*t), sin(6*pi*t)];

    % Compute coefficients 'c' for the trigonometric model using least squares
    c = (a'*a)\(a'*b);

    % Generate a finer time grid for plotting the trigonometric model
    t1 = linspace(min(t), max(t), 1000);

    % Calculate the fitted values using the trigonometric model
    fittedValues = c(1) + c(2)*t1 + c(3)*cos(2*pi*t1) + c(4)*sin(2*pi*t1) + ...
                   c(5)*cos(4*pi*t1) + c(6)*sin(4*pi*t1) + ...
                   c(7)*cos(6*pi*t1) + c(8)*sin(6*pi*t1);

    % Calculate the RMSE
    modelValues = c(1) + c(2)*t + c(3)*cos(2*pi*t) + c(4)*sin(2*pi*t) + ...
                  c(5)*cos(4*pi*t) + c(6)*sin(4*pi*t) + ...
                  c(7)*cos(6*pi*t) + c(8)*sin(6*pi*t);
   coefficients = c;

    RMSE = sqrt(mean((b - modelValues).^2));

    % Plot the original data and the trigonometric fit
    plot(t, b, 'o', t1, fittedValues, 'LineWidth', 2);
    grid; set(gca, 'FontSize', 16); xlabel('Time (years)'); ylabel('CO2 (ppm)');
    title(['Trigonometric Fit to ', plotTitle, ' Data']);
   fprintf('%s:   c2 = %f,  RMSE = %f\n', plotTitle,   c(2),   RMSE);

end
```

PART IV:

Summarize your conclusions. What happened to the RMSE as you increased the complexity of the model? For each of the three decades, is the temperature increasing or decreasing over that decade? Does that answer depend on the model you are using, or do the different models agree?

RMSE:

In part 1, using a simple linear model, the RMSE values are relatively high with a range from 14.15 to 14.76, then part 2 RMSE decreased to a range of 2.71-3.23 with the trigonometric model, and lastly even a better RSME was achieved with the improved trigonometric model which harmonics yielding a range 2.598847 to 3.09. The analysis shows that cyclical models fit the three decades of temperature data better than a linear model. This is expected, as monthly temperatures naturally exhibit seasonal cycles. Hence, models with cyclical components, like the trigonometric models, more accurately represent these patterns, offering a superior fit over linear model.

Temperature:

- Using the data from 1973-1983: The linear model (part 1) suggests a slight increase in temperature (c2 = 0.275798). In part 2, the coefficient c2 changes sign to negative (-0.098477), which might suggest a decrease; however, this could also be a result of the model capturing the oscillatory nature of the data rather than a realistic trend. In part 3, the coefficients become more complex, making it hard to spot a trend.
- Using the data from 1993-2003: The linear model indicates an increase (c2 = 0.479146). This increasing trend is consistent in part 2 (c2 = 0.119544). Part 3 again introduces complexity, but the linear trend component still suggests an increase.
- Using the data from 2013-2023: Similar to the other decades, both parts 1 and 2 indicate an increasing trend (c2 = 0.544246 and 0.1847, respectively). Part 3 continues to follow this pattern at 0.1796.

In conclusion, the lineal model's coefficient c2, which represents the trend is positive for all decades suggesting a temperature increase for each decade. Similarly, the two trigonometric models return a c2 that suggested the same trend as the linear model suggesting an increase of temperature. For the 1973-1983 and 1993-2003 periods, c2 values are small and change sign, indicating a less clear linear trend. While for 2013-2023 decade, c2 is positive in both trig models suggesting a temperature increase.

Does that answer depend on the model you are using, or do the different models agree?

The Linear model suggested temperature across all 3 decades. While trigonometric models provided a small and varying c2 for decades 73-83 and 83-93 which made it hard to determine if the temperature was rising or not. For the 13-23 decade, these models agree with the linear model in showing an increasing trend. It seems than in fact the answer depends on the chosen model, which highlights the importance of choosing the right model.