Project 1: MATLAB Curve Fitting of Chagas Disease

Diyana Tial, ID: 16334584 (Preparing Slides, Computing,Editing,Revising)
Annalisa Berg, ID: 16339285 (Preparing slides, computing,Editing,Revising)

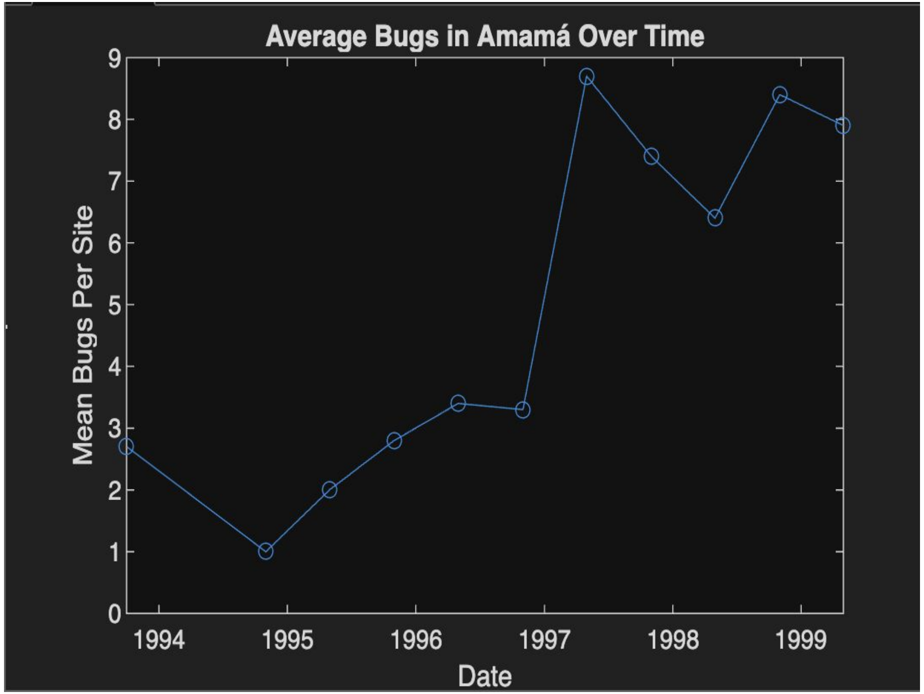
Introduction | Main Objectives

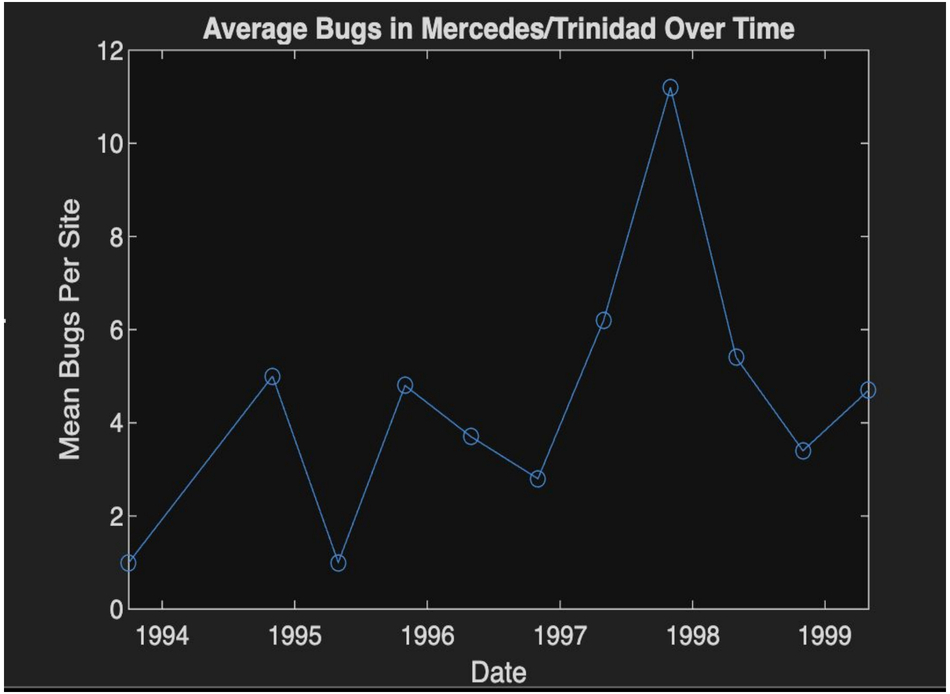
- ❑ Chagas' disease (also "kissing bug") first described in 1909 by Dr. Carlos Chagas.
- ❑ Caused by the parasite *Trypanosoma cruzi,* which is spread by blood feeding *Triatoma infestans* living in rural houses
  - ❑ Infection can occur through bites, contaminated food, blood transfusion, or from mother to fetus.

- ❑ To examine the long-term effectiveness of insecticide spraying against *Triatoma infestans* in two rural areas in Argentina (Amamá & Mercedes/Trinidad)
- ❑ To compare reinfestation patterns between Amamá & Mercedes/Trinidad.
- ❑ To analyze how the mean number of bugs per site changed over time (1993-1999)
- ❑ To understand the challenges of sustaining vector control for Chagas' disease.

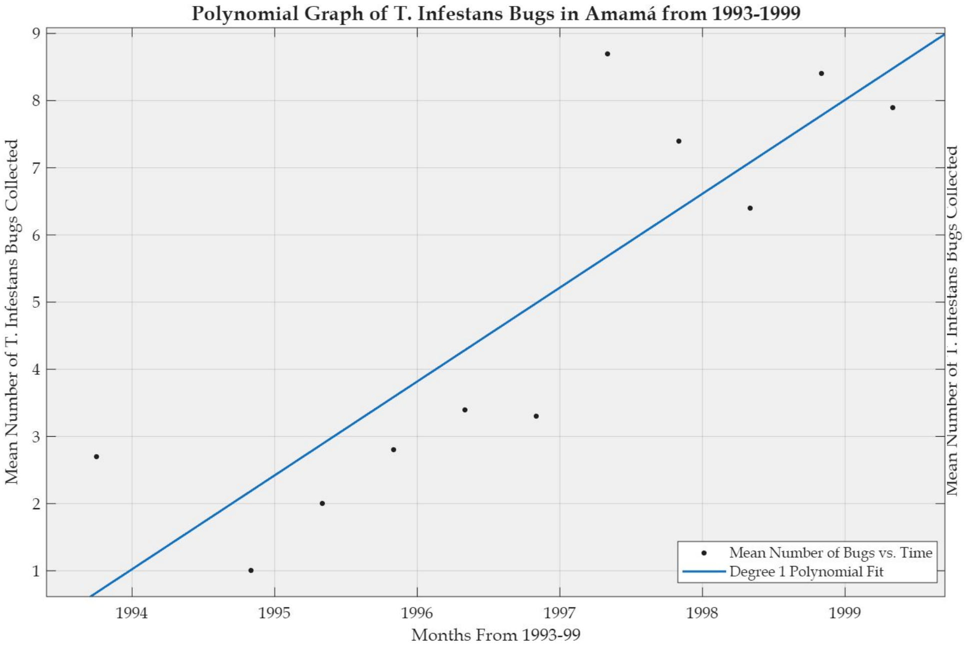
Plot of Amamá

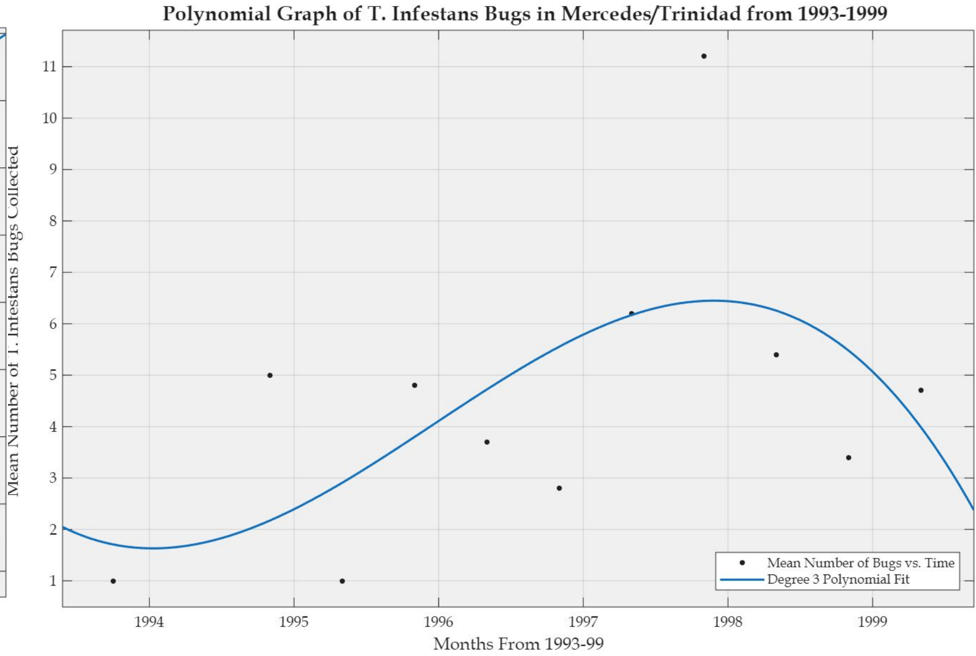
Plot of Mercedes/Trinidad



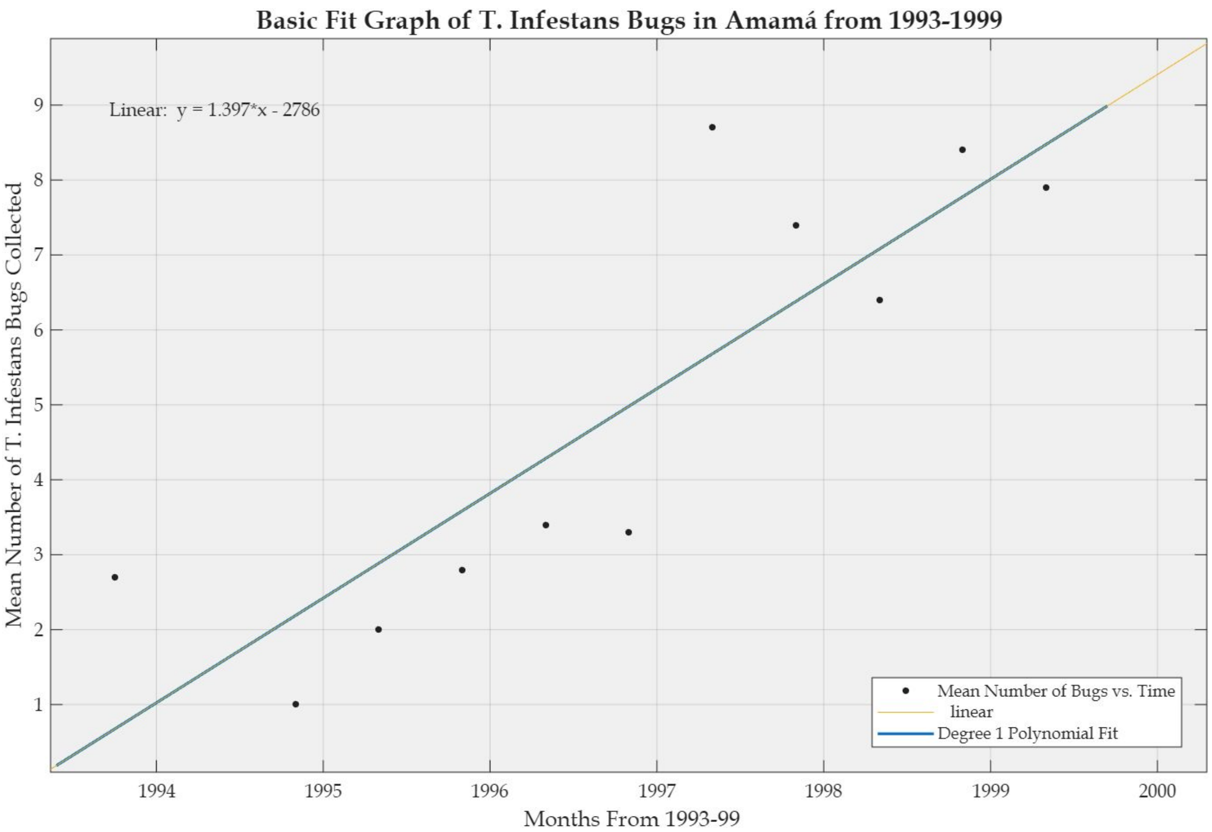


Best-fit Polynomial Graphs





Basic Fit - Amamá



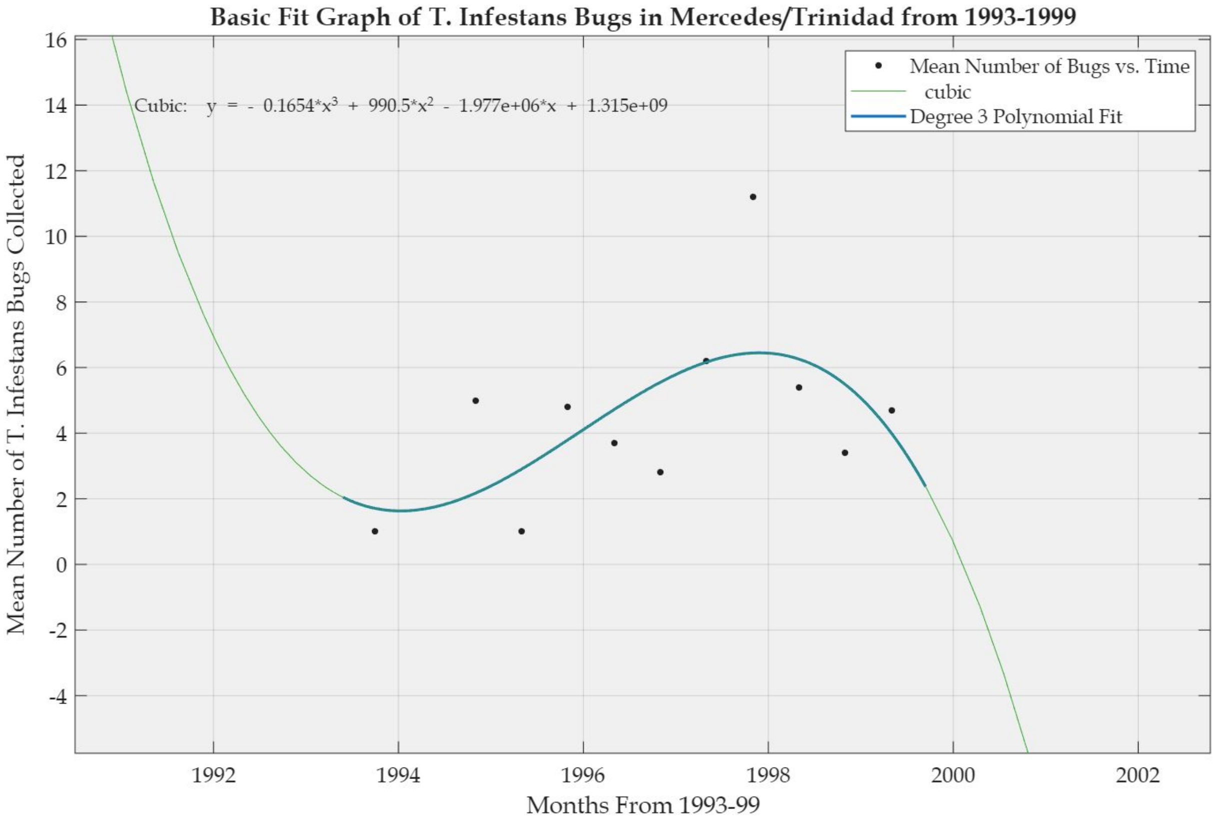
Observation of the behavior around the graph's edges indicate an upward trend in the population of T. Infestans bugs in Amamá.

This growth is likely a result of the complexity in evolution of a biological species.

T. Infestans bugs that survived the initial treatment lived on to reproduce, hence a population increase.

Additionally, the areas were not known to have been re-treated for the species, so it is not surprising that there is an increase in population.

Basic fit - Mercedes/Trinidad

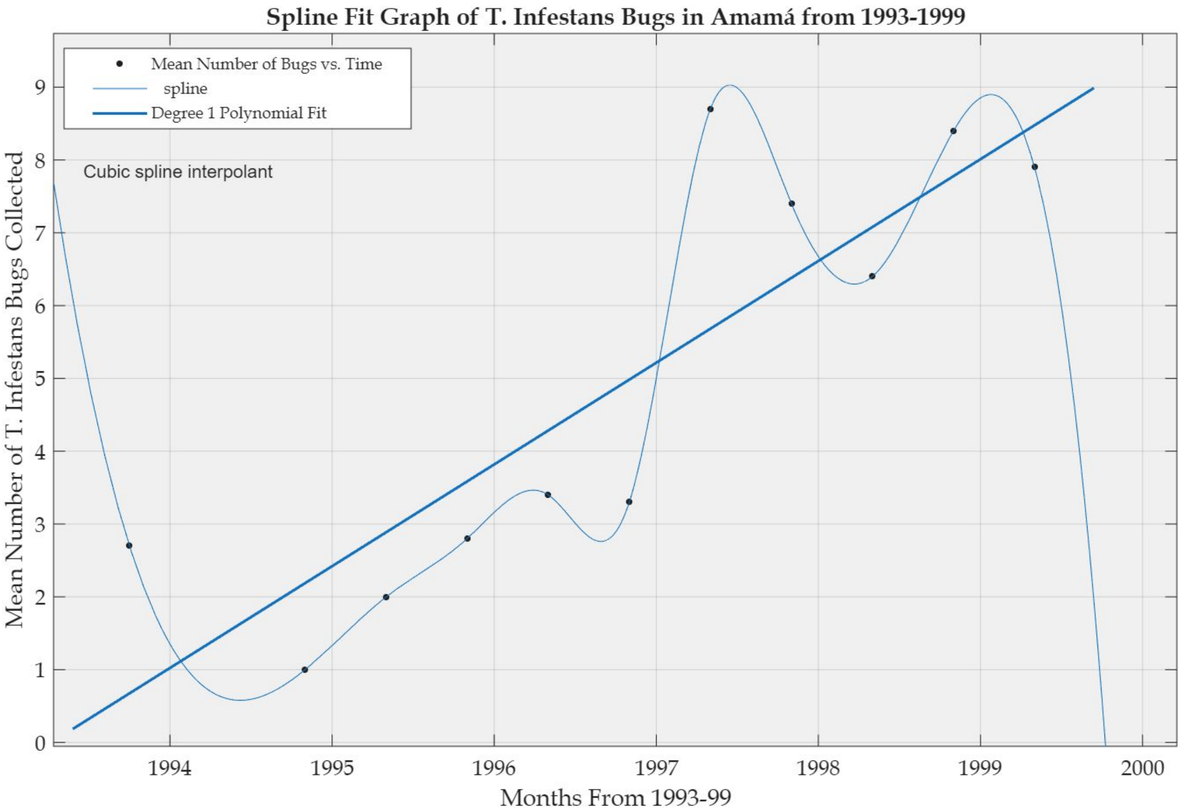


Observation of the basic-fit graph edges indicates a significant downward slope in the population of T. Infestans bugs in Mercedes/Trinidad.

This data represents a slowly growing population of the species after treatment, and a projected decline in population.

This fit is not an unlikely outcome for the population, as the data showed that in months following the treatment, the population fluctuated from periods of slow growth to periods of decline.

Spline - Amamá

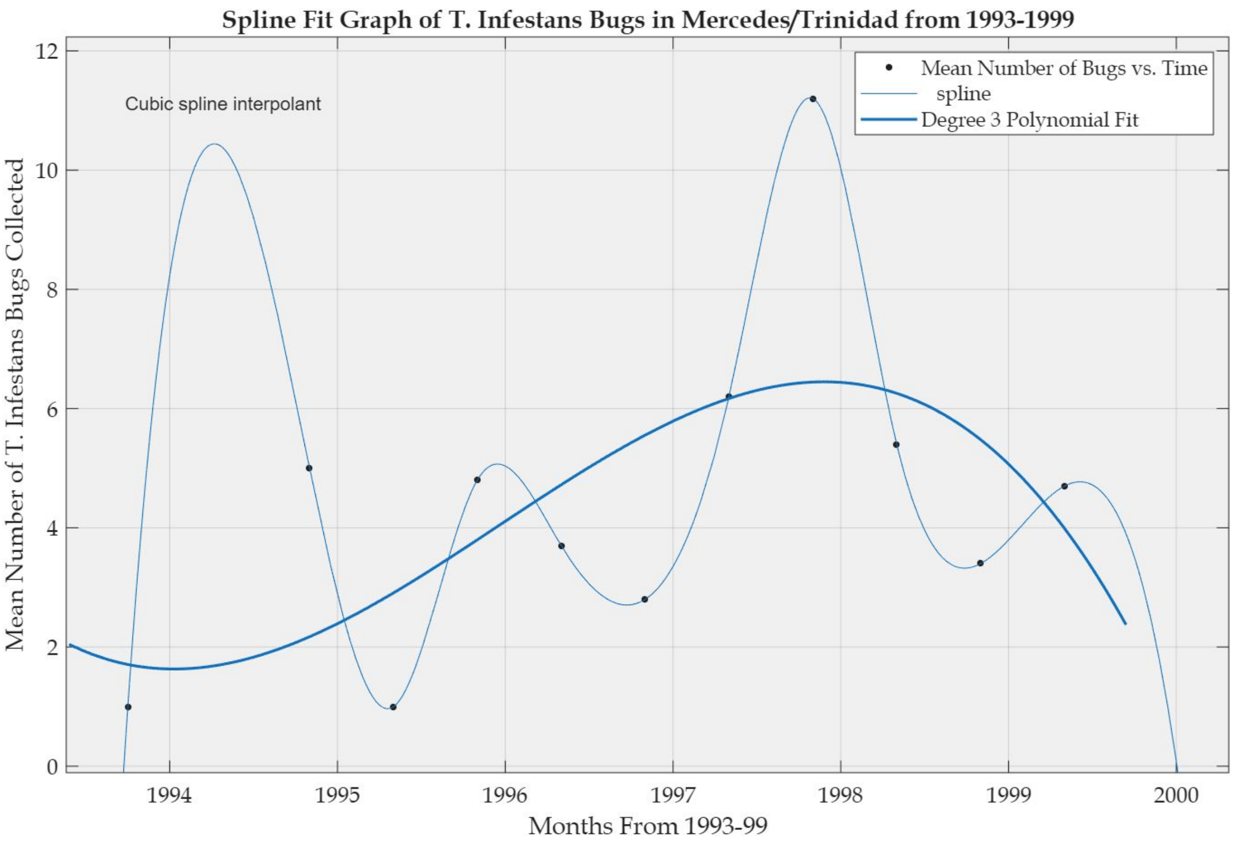


The cubic spline fit for this data does not seem to be a better fit for the data than a degree 1 polynomial.

The cubic spline displays a steep downward slope as time increases, which does not match the trend of the data points.

The beginning of the cubic spline graph, however, more likely describes the T. Infestans bug population prior to being treated with insecticide.

Spline - Mercedes/Trinidad



The cubic spline fit for the Mercedes/Trinidad T. Infestans bugs population is a better fit than a degree 3 polynomial fit.

This graph maps out the large fluctuations in population, which a polynomial function does not describe accurately.

Transformation

We believe that the data for Amamá and Mercedes/Trinidad does not need to be transformed. The reason why is because:

- The data is uniformed and stable throughout the years
- The data is not exponentially increasing

Y = Cx^a & Y = Ce^ax transformations were done on both datasets and yielded less uniform curves and lower R^2 scores.

We also took into account the standard deviation for both datasets:

- Amama: 2.8613
- Mercedes: 2.7965

This STD tells us that the variability was consistent and moderate across the both datasets. The spread is stable and does not increase exponentially

Qualitative Difference

- There is a qualitative difference between the dataset of Amamá and Mercedes/Trinidad.

- The mean population of Amamá grew over time.

- The mean population of Mercedes/Trinidad fluctuated over time.

- There was a spike in population during different seasons of the year:
  - Amamá: Between Nov. 96 & May 97 the population mean increased from 3.3 to 8.7.
  - Mercedes/Trinidad: Between May 97 & Nov. 97 the population mean increased from 6.2 to 11.2.

Redirection of the Project

Based on our professor's feedback, we did some changes to improve the R^2 score of our model for November 1999 predictions.

<u>Changes:</u>

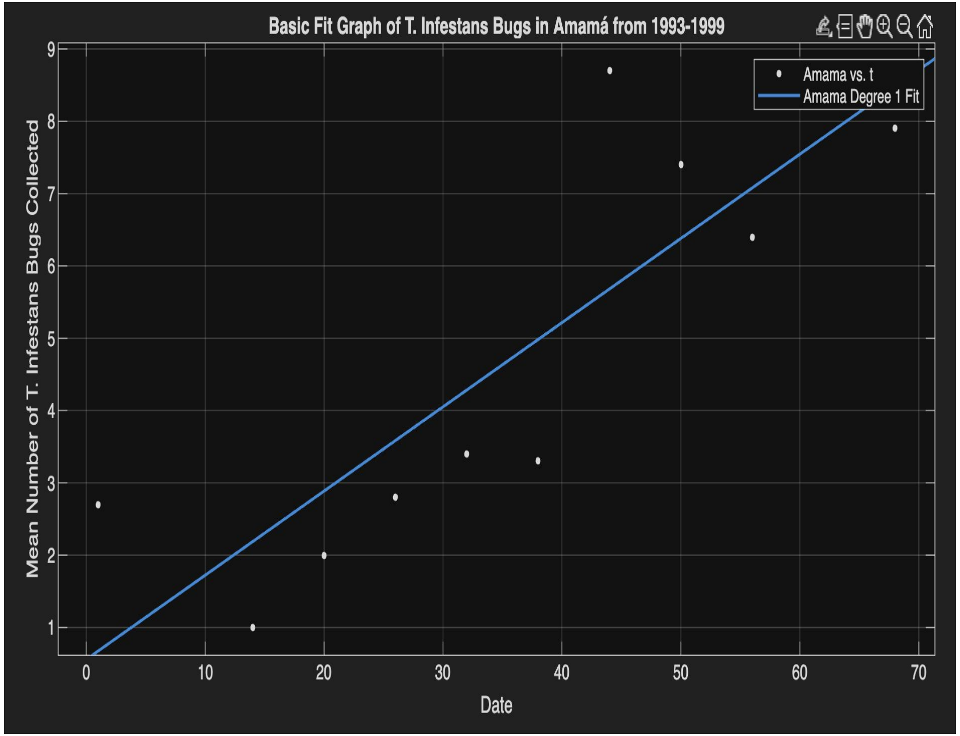
- Date (x-values) now as an int value for easier implementation for later calculation (Slide 11)

- Mercedes/Trinidad original best fit polynomial of cubic fit graph changed to polynomial degree 1 best fit for evaluating detrended value (Slide 11)
  - detrended value (Vd) = raw bug counts - best fit trend values

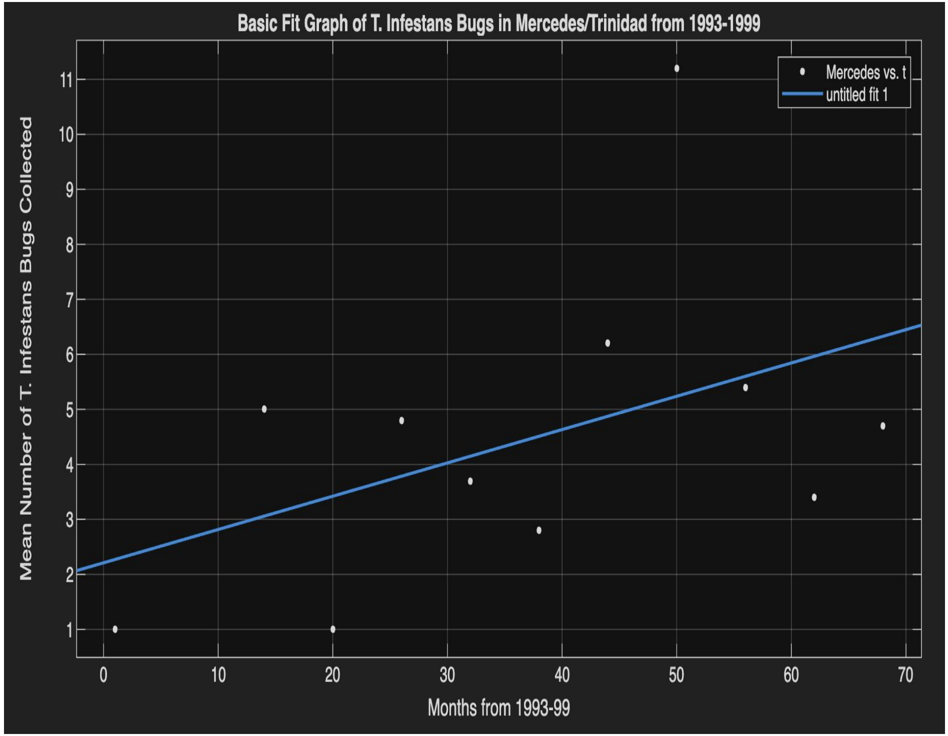
- Curve fit with t (time) as the x value and Vd as the y value and used the Sum of Sine Degree 3 for fitting which resulted in a great R^2 score (over 90) for both Amamá and Mercedes/Trinidad (Slide 14).

- We then used the linear regression line model and the Sine model for November 1999 predictions (Slide 15).

$$trend(t) = 0.1165t + 0.5579$$

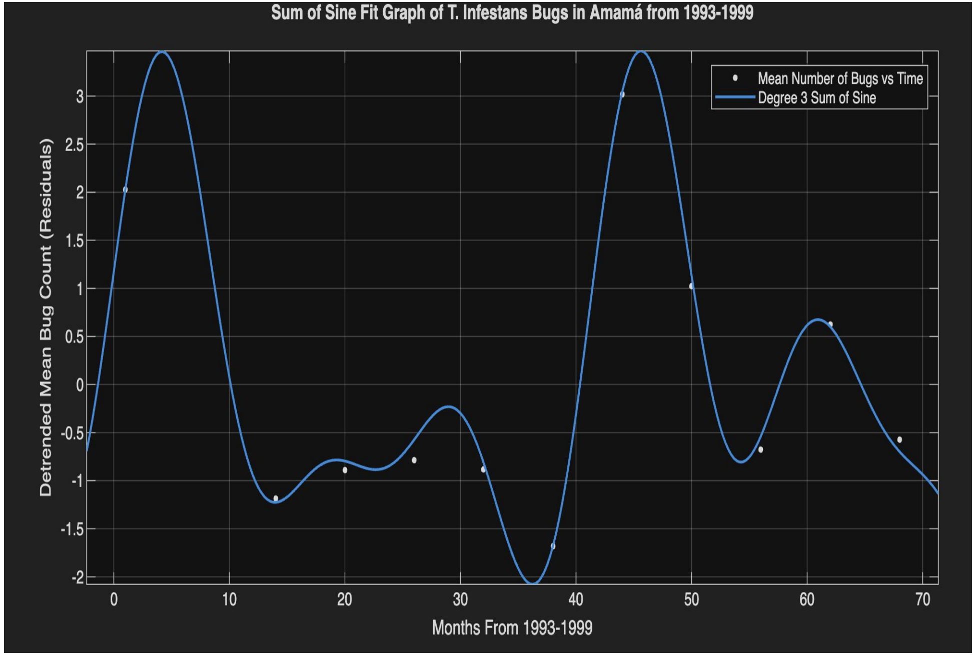


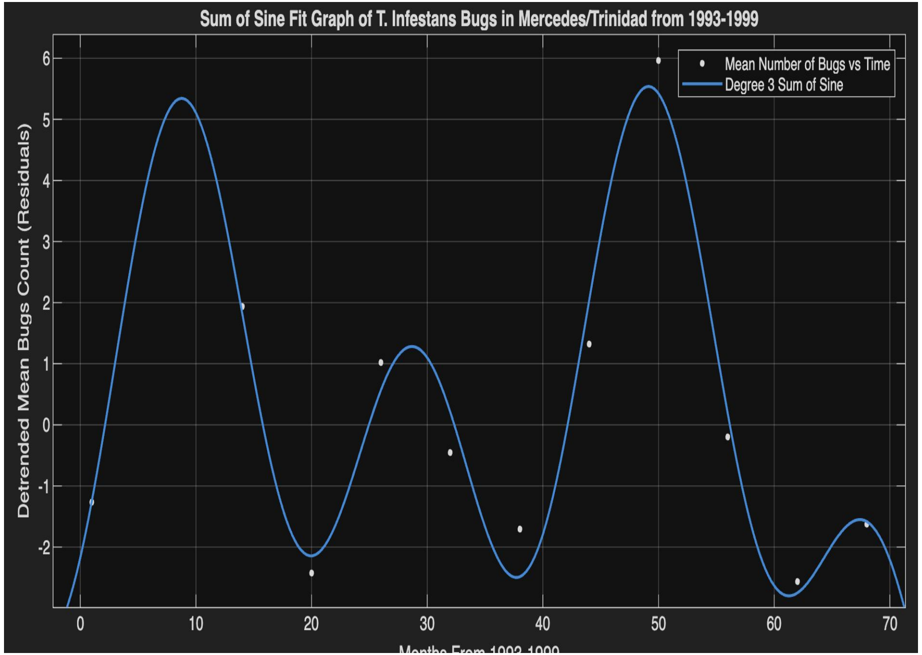
$$trend(t) = 0.06055t + 2.21$$



Detrended Sine Wave Models

f(x) = a1*sin(b1*x+c1) + a2*sin(b2*x+c2) + a3*sin(b3*x+c3)





Detrending the data allows for greater context of the data, in terms of seasons and cyclical fluctuations. The data was fit to a NEW regression fit line, then the regression line values were subtracted from the data points. Next, Sum of Sine degree 3 was used to model the data, yielding better R^2 values, with a more logical model.

Predictions for November 1999

Our code for changing dates to int values generated by ChatGPT:

```
dates = ["Oct-93","Nov-94","May-95","Nov-95","May-96","Nov-96", ...
         "May-97","Nov-97","May-98","Nov-98","May-99"];

dt = datetime(dates,'InputFormat','MMM-yy');

x_years = year(dt) + (month(dt)-1)/12; % decimal years

t0 = x_years(1); % first point (1)
t = round(12*(x_years - t0)) + 1;
```

Using this formula, we found that the value for November 1999 is 74.

Amamá model:

trend(t) = 0.1165t + 0.5579 → t = 0.1165(74) + 0.5579 = 9.1789

S(t) = a1*sin(b1*t+c1) + a2*sin(b2*t+c2) + a3*sin(b3*t+c3) using the coefficients given from MATLAB:

S(t) = -1.8053230417

Total = 9.1789 - 1.8053230417 ≈ 7.37357

Using the detrending technique, the mean population of bugs in Nov. 1999 would be: 7.37

Mercedes/Trinidad model:

trend(t) = 0.06055t + 2.21 → t = 0.06055(74) + 2.21 = 6.6907

S(t) = a1*sin(b1*t+c1) + a2*sin(b2*t+c2) + a3*sin(b3*t+c3) using the coefficients given from MATLAB:

S(t) = -1.5947636203

Total = 6.6907 - 1.5947636203 ≈ 5.0959363797

Using the detrending technique, the mean population of bugs in Nov. 1999 would be: 5.1