Exercises in attractor reconstruction

Isak Ederlöv

April 21, 2023

1 Background and objectives

In this report, measurement data from two different sources will be analysed with several statistical methods to investigate if non-linear behavior is present. The two sources are measured intensity data from an unknown laser along with monthly sea surface temperature deviation data of the El Niño / La Niña-Southern Oscillation phenomenon (ENSO).

Correlograms were created for the data sets. A correlogram is a graph that shows the autocorrelation function as a function of another metric, in this case lag time τ. The autocorrelation function is

$$C(\tau) = \frac{[x(t) - \mu][x(t + \tau) - \mu]}{\sigma^2} \tag{1}$$

The main objective is to reconstruct attractors based on peaks from the power spectral density (*PSD*). An attractor refers to a specific set of values or a region in a system's state space. It is a stable point towards which a dynamic system tends to approach or stay near over time, irrespective of the initial conditions. An attractor can be perceived as a trajectory or pattern that the system follows over time. This pattern can either be deterministic, known as a fixed point attractor, or chaotic, known as a strange attractor.

2 Methodology

Because of the similarities of the objectives regarding the two data sets, the methods used for analysis only differ marginally for the different data sets.

To begin, the time series were plotted using *matplotlib*. Following this, histograms were produced and the statistical moments were calculated using built-in functions of *numpy* and *scipy*. For the laser intensity data, a simple threshold detecting algorithm was used to extract the inter-spike intervals (*ISI*). The spikes were found by iterating through the data and extracting the index of the current position if the current position is below the threshold and the next position is above. The result is an array of indices of which the difference can be taken to produce an array of ISI.

```
thresh = -0.03
intervals = np.diff([x for x in range(len(data)-1) if data[x] < thresh and data[x+1] > thresh])
```

The x-axis series in the ENSO data was divided by twelve in Figure 11 to display the lag in years.

Return maps were produced simply by shifting the time series by a lag time τ and displaying the forward shifted series in the y-axis.

```
def return_map(data, tau):
    plt.plot(data[:-tau], data[tau:])
    plt.xlabel("$x(t)$")
    plt.ylabel("$x(t+)$")
    plt.title(f"Return map of laser intensity,  = {tau}")
    plt.show()

[return_map(data, tau) for tau in [5, 12,20,44,70,100]]
```

Different values of lags were chosen both based on frequency peaks from the PSD, but also ad libitum based on the time series plots.

The correlograms were created by calculating the autocorrelation function and plotting it against a range of lag values.

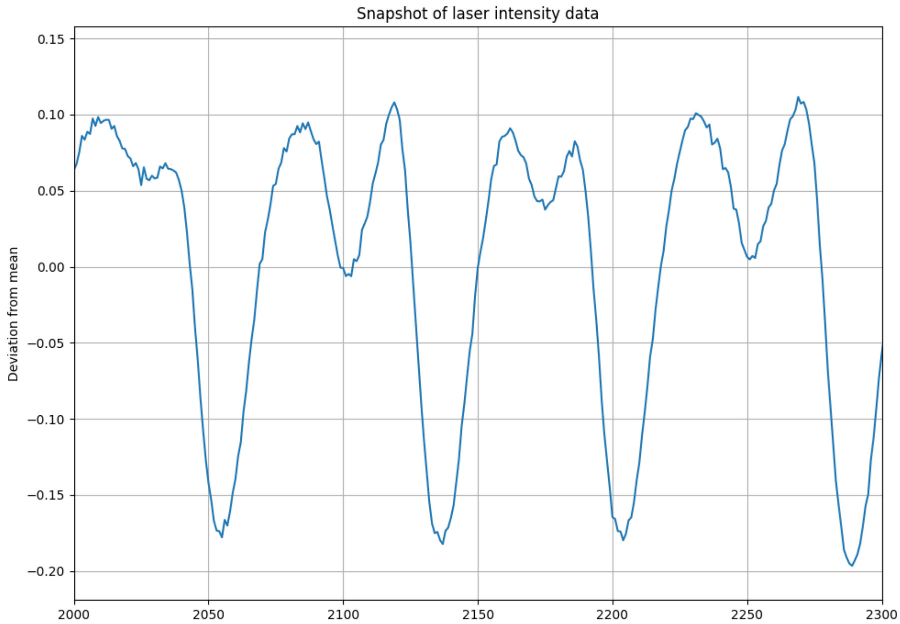


Figure 1: Shows one randomly chosen region of the laser intensity data. The characteristic two-peaked hill shape is present in the entire series.

```
def ACF(data, tau):

    N = len(data)
    mean = np.mean(data)
    std = np.std(data)

    acf = 1 / ((std**2) *(N-tau)) * sum(
        [(data[t] - mean) * (data[t+tau] - mean) for t in range(1, N-tau)])
    return acf
```

For the Gaussian surrogate data, a series with the same length, mean, and variance was generated using the *normal* function in *numpy*.

```
gaussian_noise = np.random.normal(mean, variance, size=len(flat_data))
```

3 Results

3.1 Laser intensity

The result of the in The upper and lower isolated extremes are artifacts of the thresholding algorithm for spike detection. Note that the posititve value of skewness is reflected in the "long tail"

Mean	70.33
Variance	243.3
Skewness	0.9201
Kurtosis	3.618
Coefficient of variation	0.2217

Table 1: Calculated statistical moments for the inter-spike intervals.

to the right in Figure 2.

The inverse of the main peak is 68.9, which is consistent with the mean ISI of 70.

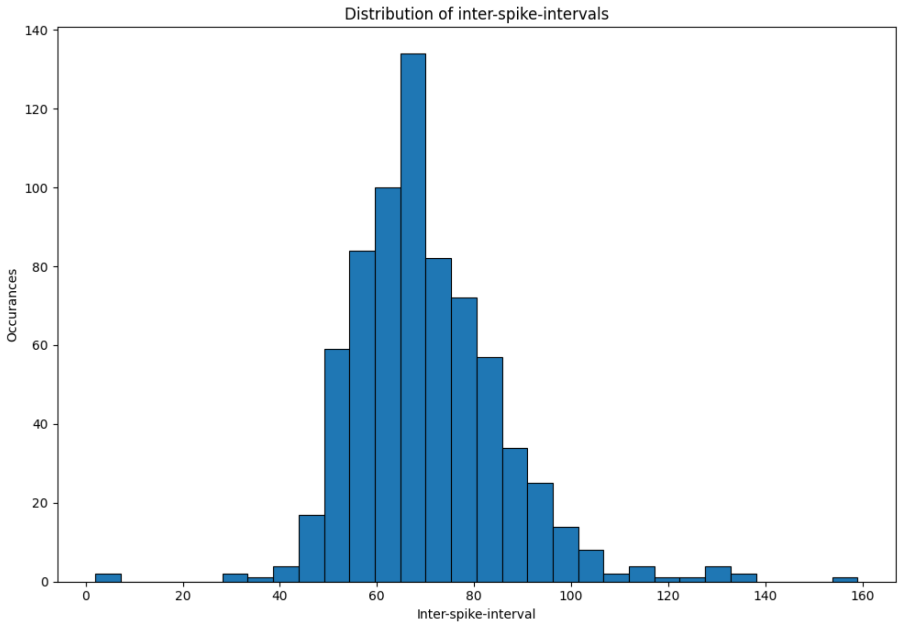


Figure 2: Distribution of the inter-spike intervals.

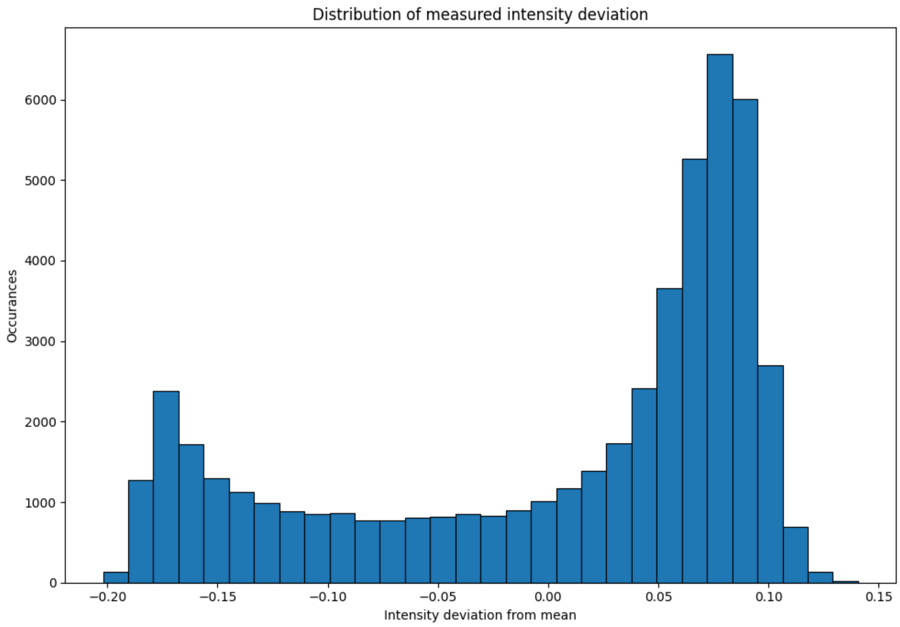


Figure 3: Shows normalised laser intensity value distribution.

Mean	3.829e-3
Variance	8.849e-3
Skewness	-0.8085
Kurtosis	-0.8553
Coefficient of variation	24.56

Table 2: Statistical moments of the intensity data.

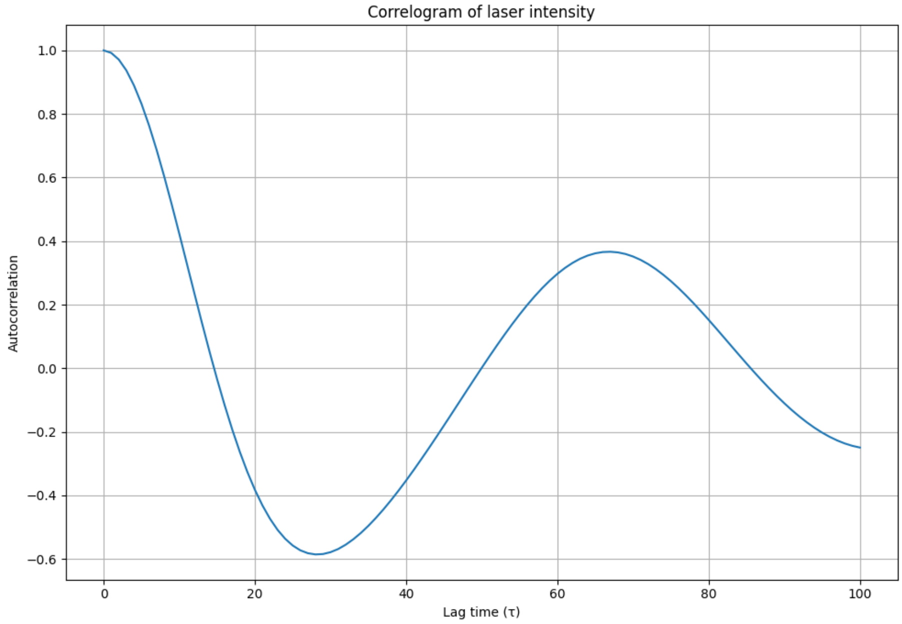
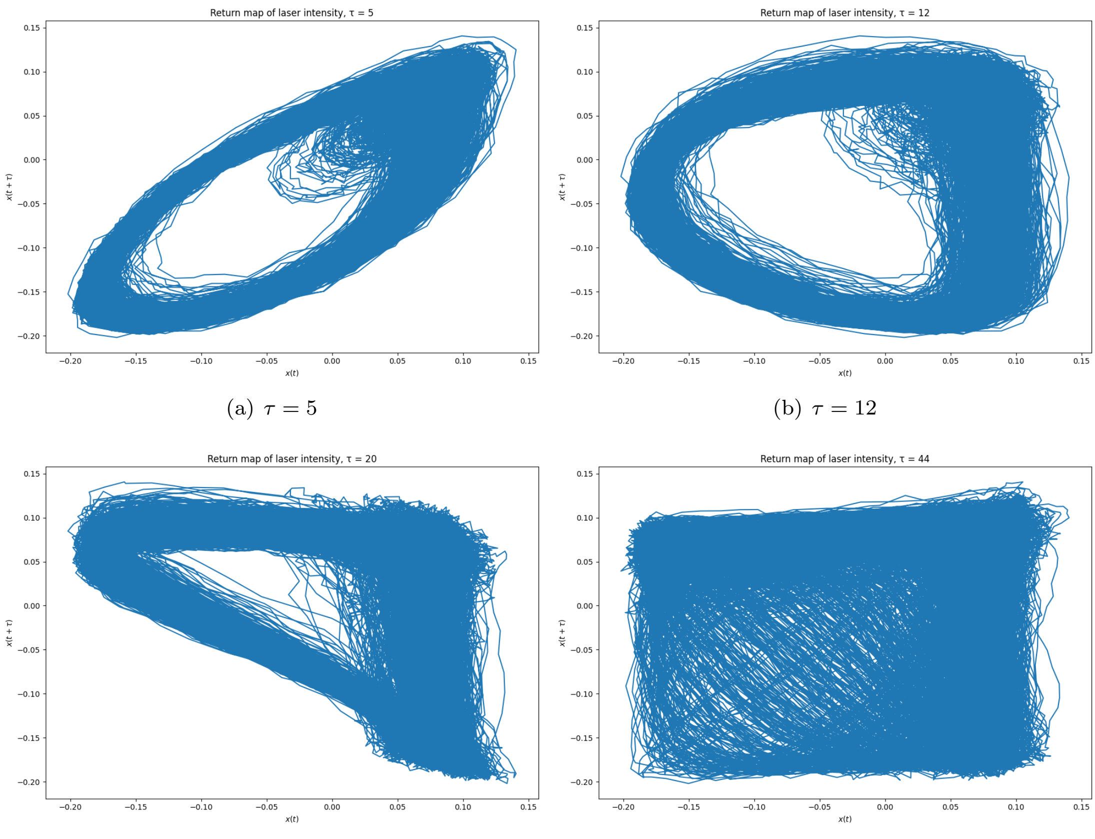


Figure 4: Shows the correlogram of the laser intensity data. Exhibits clearly periodic behavior resembling a dampened sinusoid.



(a) $\tau = 5$

(b) $\tau = 12$

(c) $\tau = 20$

(d) $\tau = 44$

Figure 5: Return maps for the laser intensity data with different values of lag time τ.

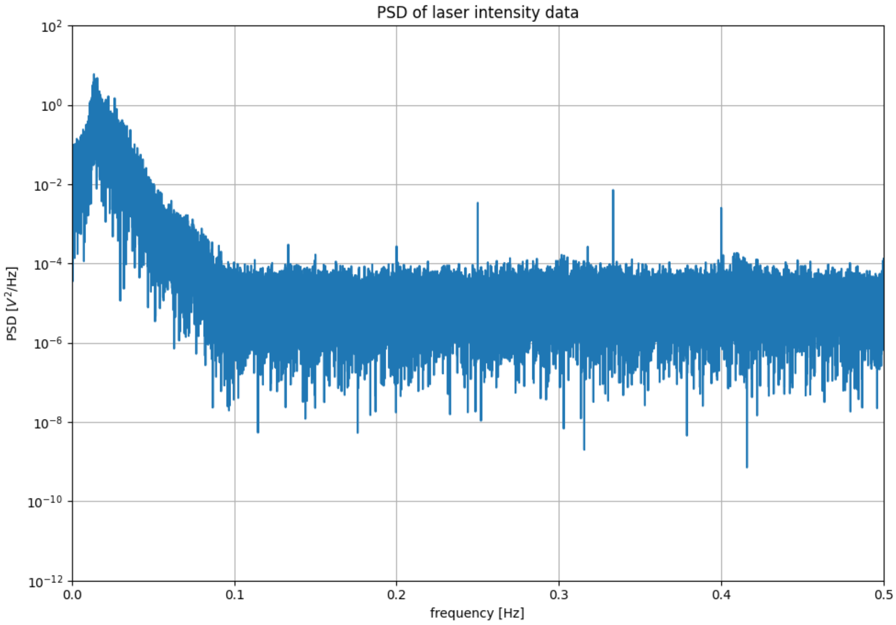


Figure 6: Power spectral density for the laser intensity data.

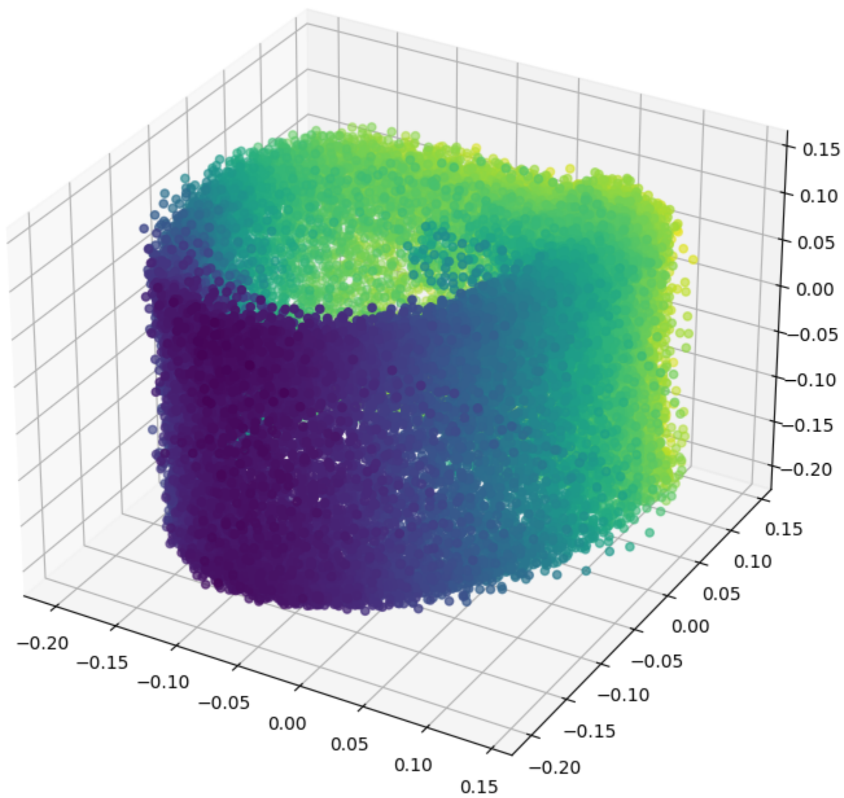


Figure 7: Reconstructed attractor based on PSD peaks.

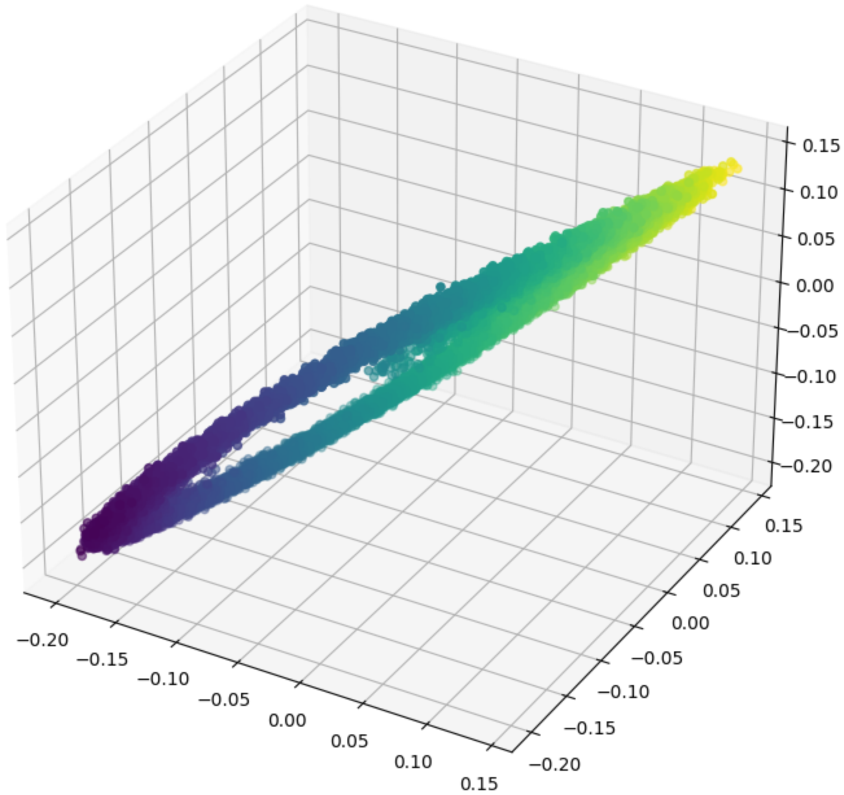


Figure 8: Reconstructed attractor based on PSD peaks from higher frequencies (local maxima).

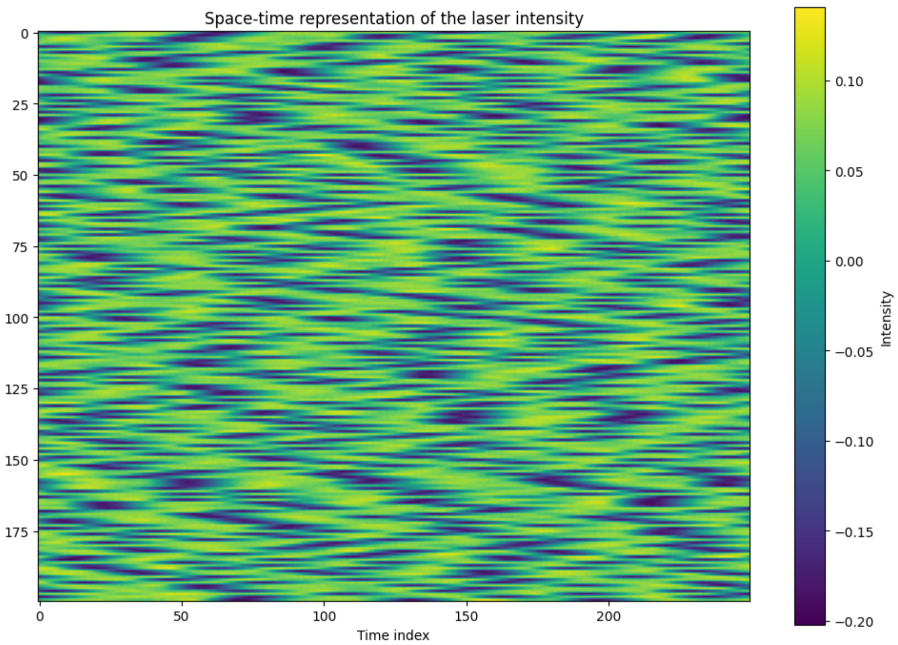


Figure 9: Space-time representation of the laser intensity data.

3.2 ENSO

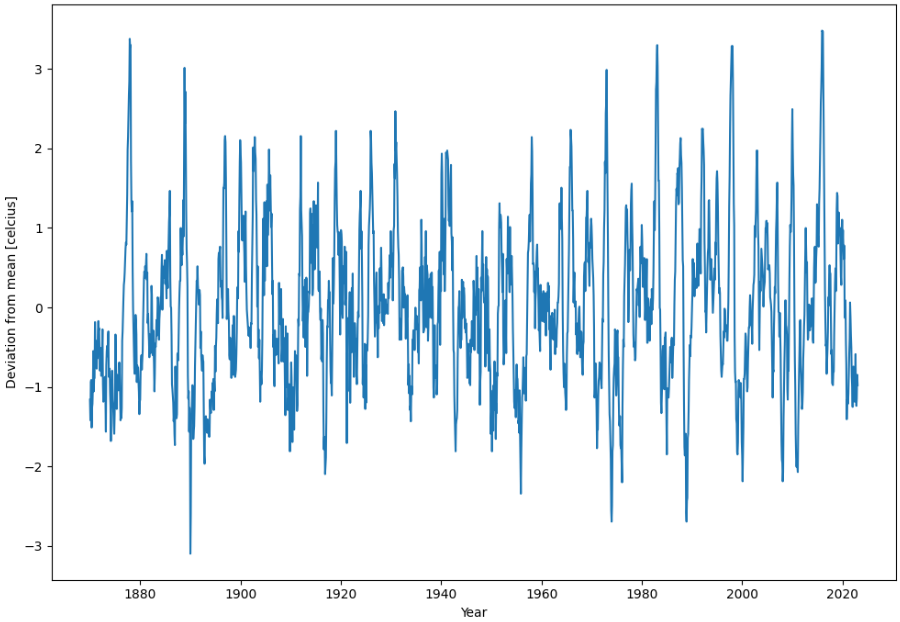


Figure 10: ENSO surface temperature deviation as a function of time

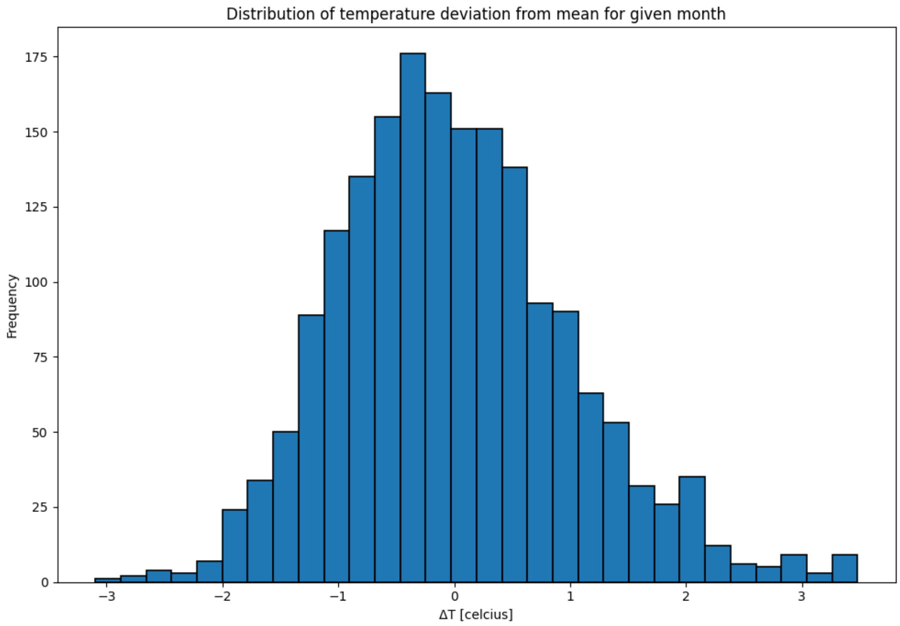


Figure 11: Histogram of the temperature data.

Mean	-0.108
Variance	0.592
Skewness	0.471
Kurtosis	0.418
Coefficient of variation	7.13

Table 3: Statistical moments of the ENSO monthly mean-temperature deviation.

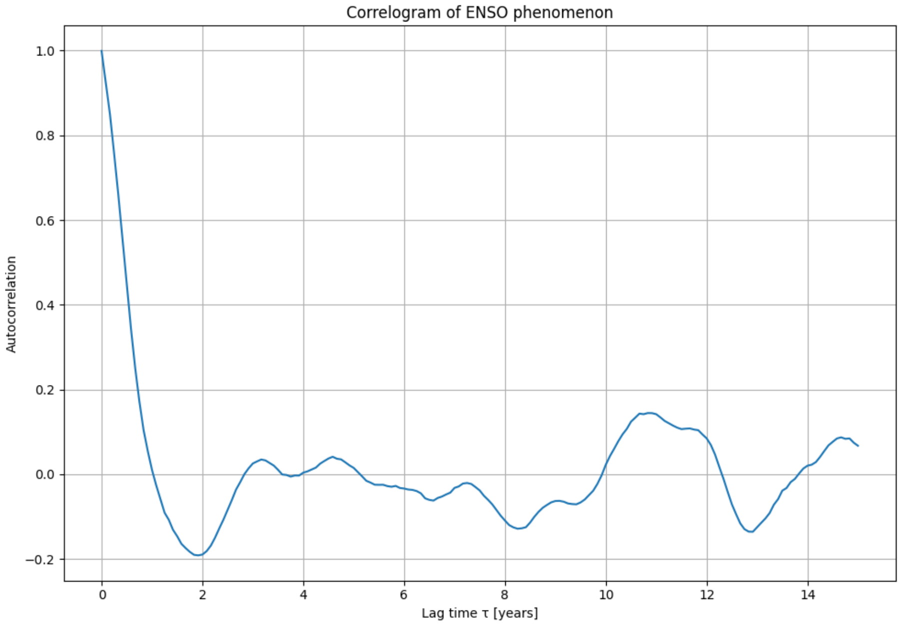


Figure 12: Autocorrelation as a function of lag time τ, here displayed in months.

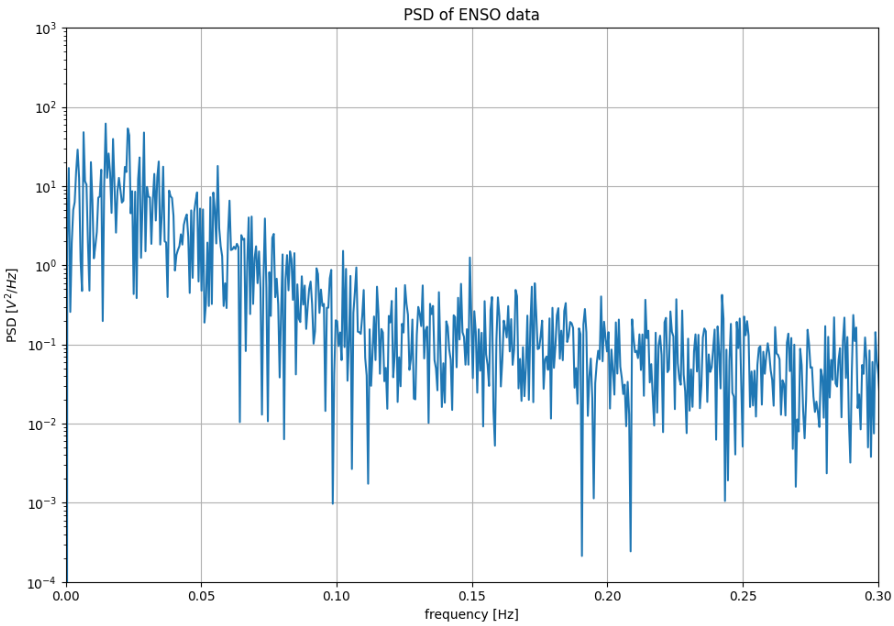
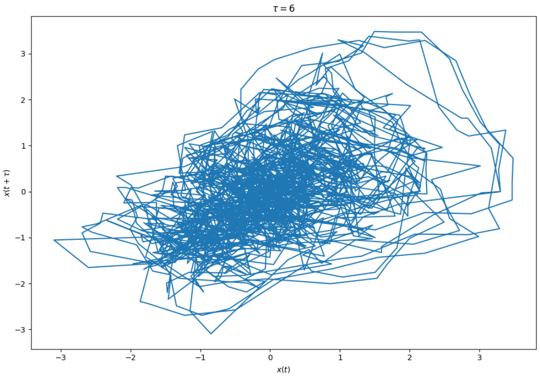
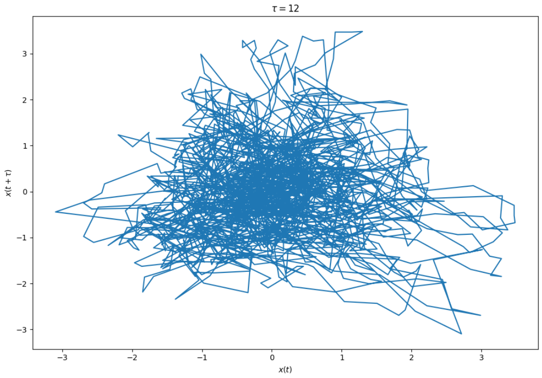


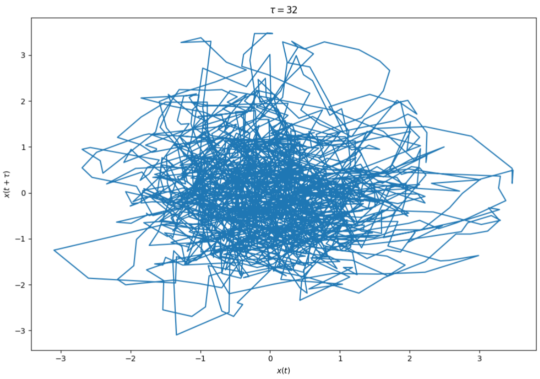
Figure 13: Power spectral density of ENSO data.



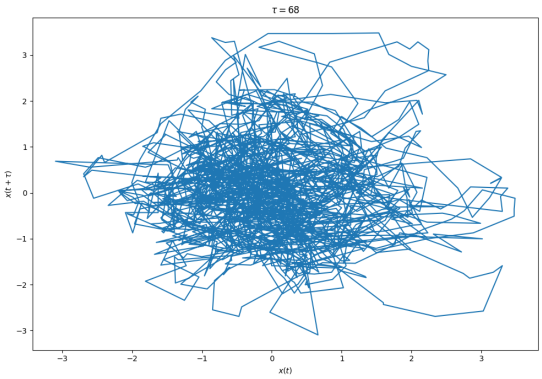
(a) $\tau = 6$



(b) $\tau = 12$



(c) $\tau = 32$



(d) $\tau = 68$

Figure 14: Shows the return map for the ENSO data with four different lags. $\tau = 1$ month.

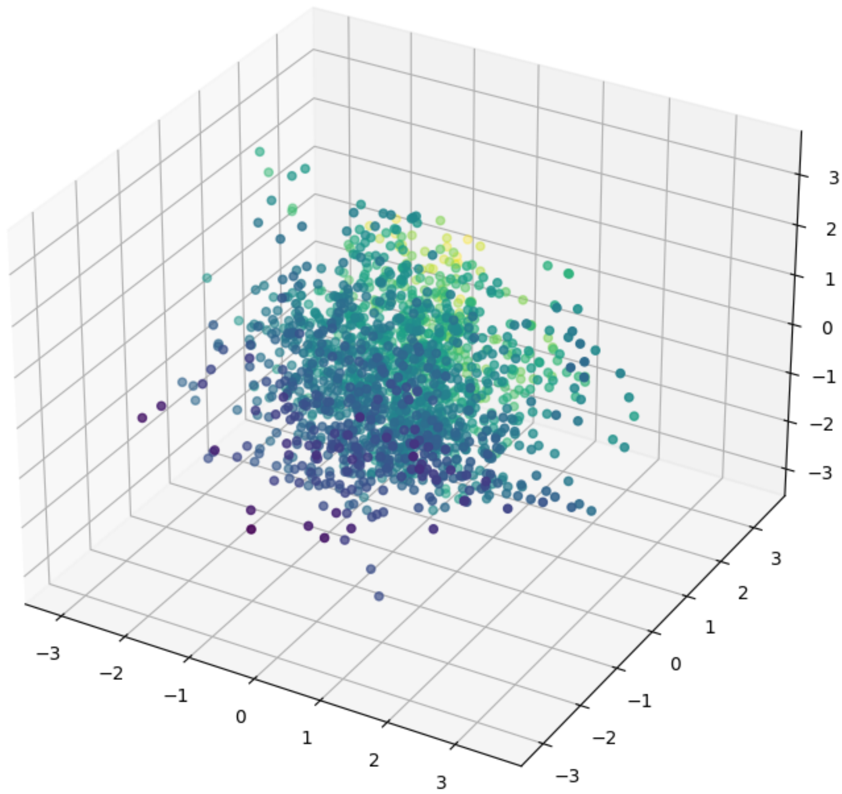


Figure 15: Reconstructed attractor based on PSD peaks.

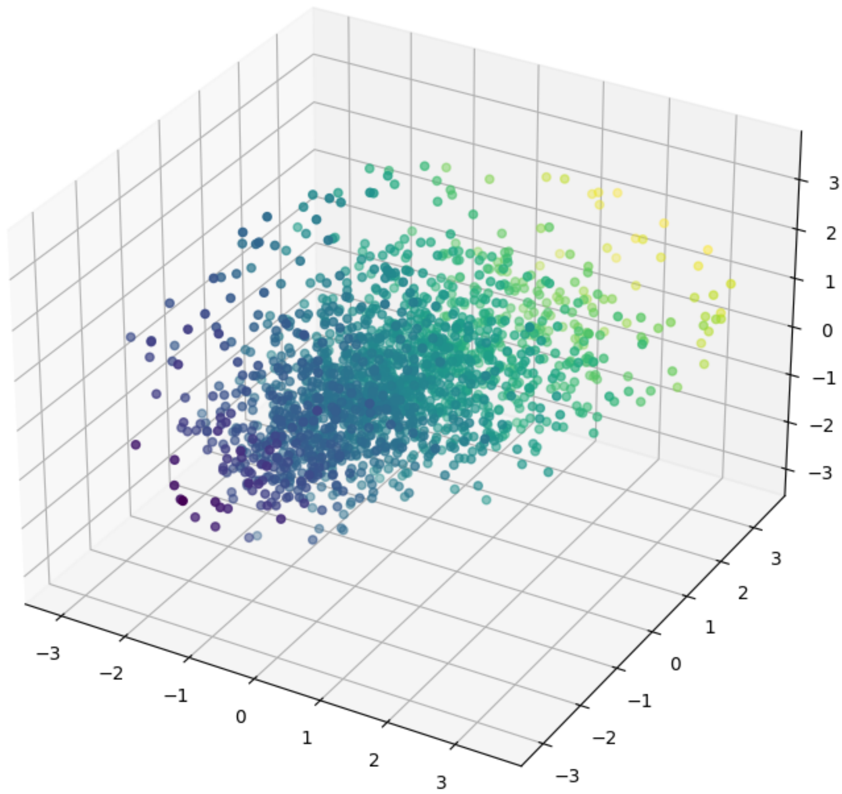


Figure 16: Reconstructed attractor based on weaker PSD peaks in the higher frequency sector (local maxima).

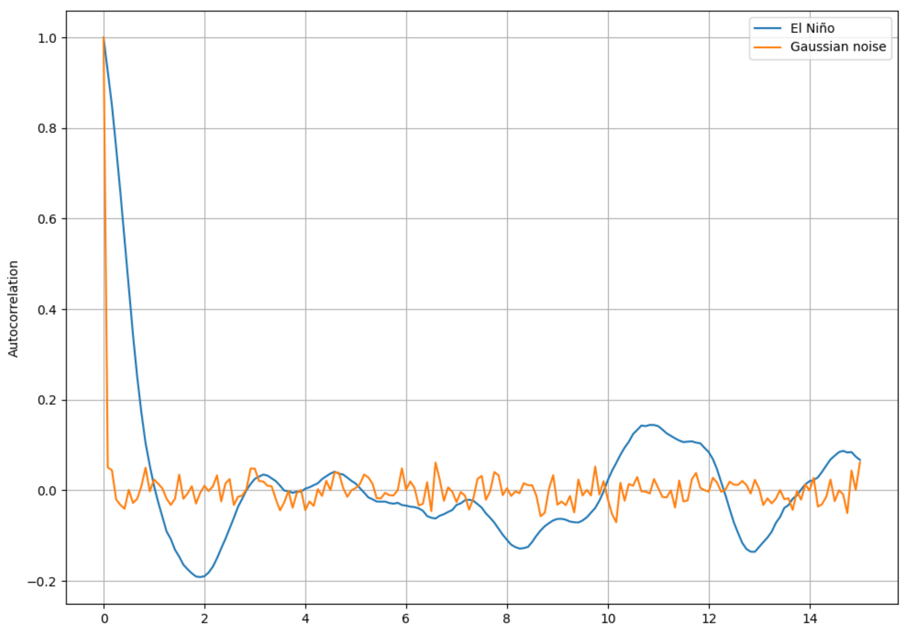


Figure 17: Shows the correlogram of the Gaussian surrogate data compared with the correlogam for the ENSO data.

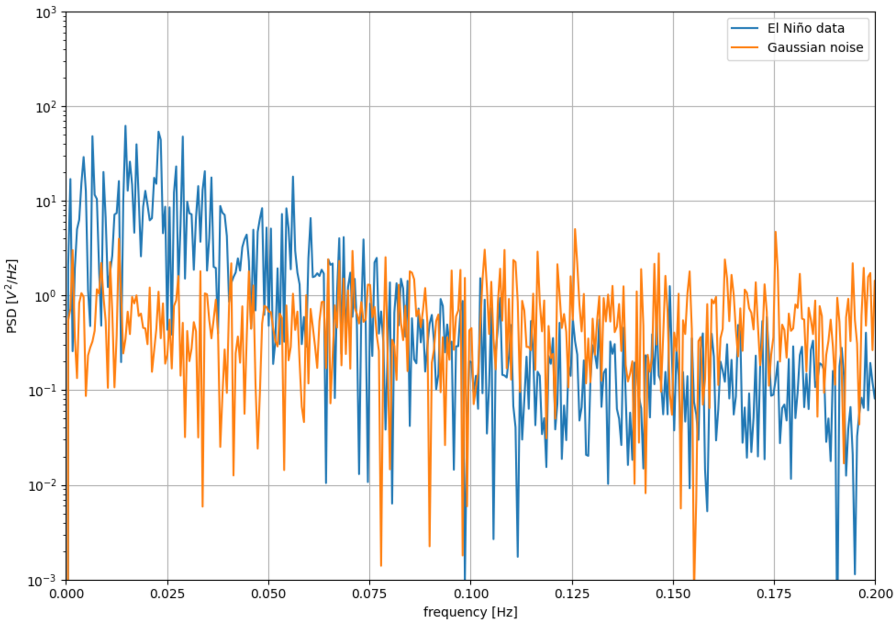


Figure 18: Shows the PSD of the Gaussian surrogate data compared with the PSD for the ENSO data.

4 Discussion

When it comes to getting a basic understanding of where the data comes from, the surface temperature data is arguably much simpler to understand, since temperature is an everyday measure that is intuitive to everyone. However, when it comes to finding interesting patterns in the data, I found the ENSO data to be much more subtle and elusive. When it comes to the attractors, the El Niño attractor seems to be either incorrectly recreated, visualised from a shielding angle or simply not that pronounced. However, it was almost hard to find lag combinations for the laser attractor that *didn't* look very strange and interesting.

When it comes to sources of error, the interval extraction algorithm produced some small anomalous intervals. Because the thresholding detects every crossing of the threshold as an interval, small fluctuations in the data around the threshold mark could count as a complete spike. A potential solution for this could have been to smooth the curve a bit, or to extend the requirement that the following t+n must also be monotonically increasing. However, this idea was discarded becuase of the minor impact these artifacts had on the distribution data. Moving the threshold from -0.03 to -0.10 only changed the amount of detected spikes by 4, (out of 710), and did not in any noticable way change the PSD.

The assumption that Gaussian white noise is not a good model for the El Niño index seems likely from the correlogram in Figure 17 and the power spectral density. However, to prove this would require some kind of hypothesis testing.