

Extracting Fractal Dimensions from Uneven Time Series



Daichi Hiramatsu

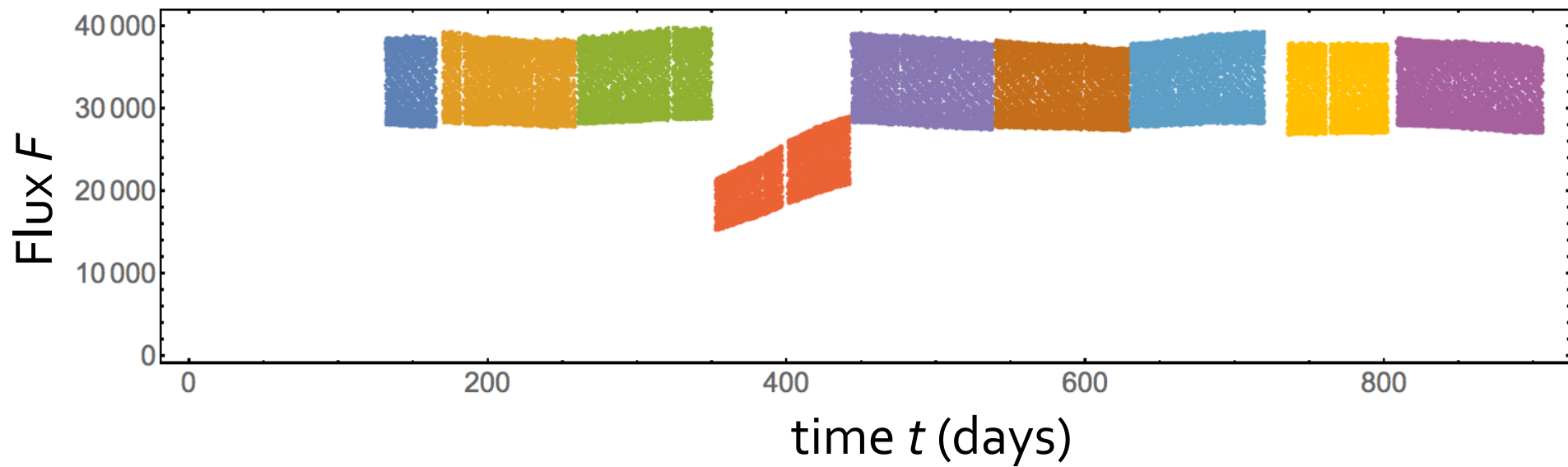
University of Hawai'i at Mānoa

Chaos Among the Stars, August 18, 2015.

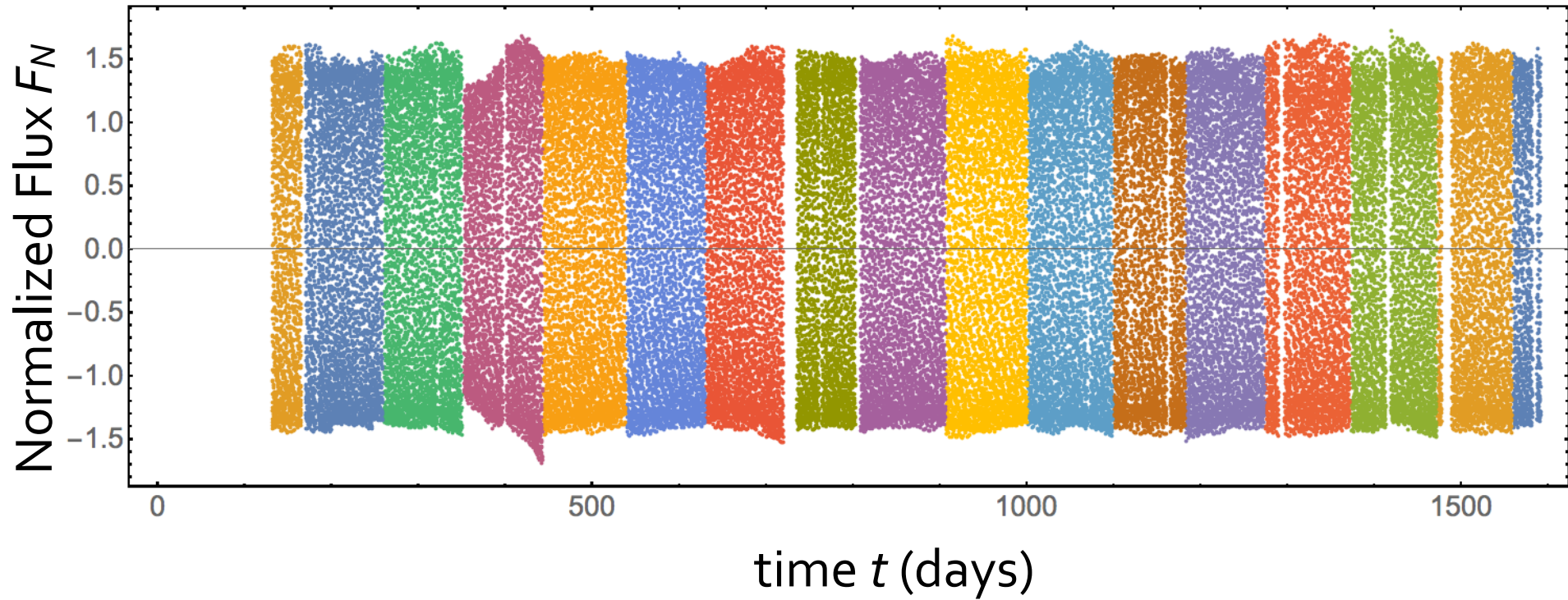
Introduction



Raw Flux

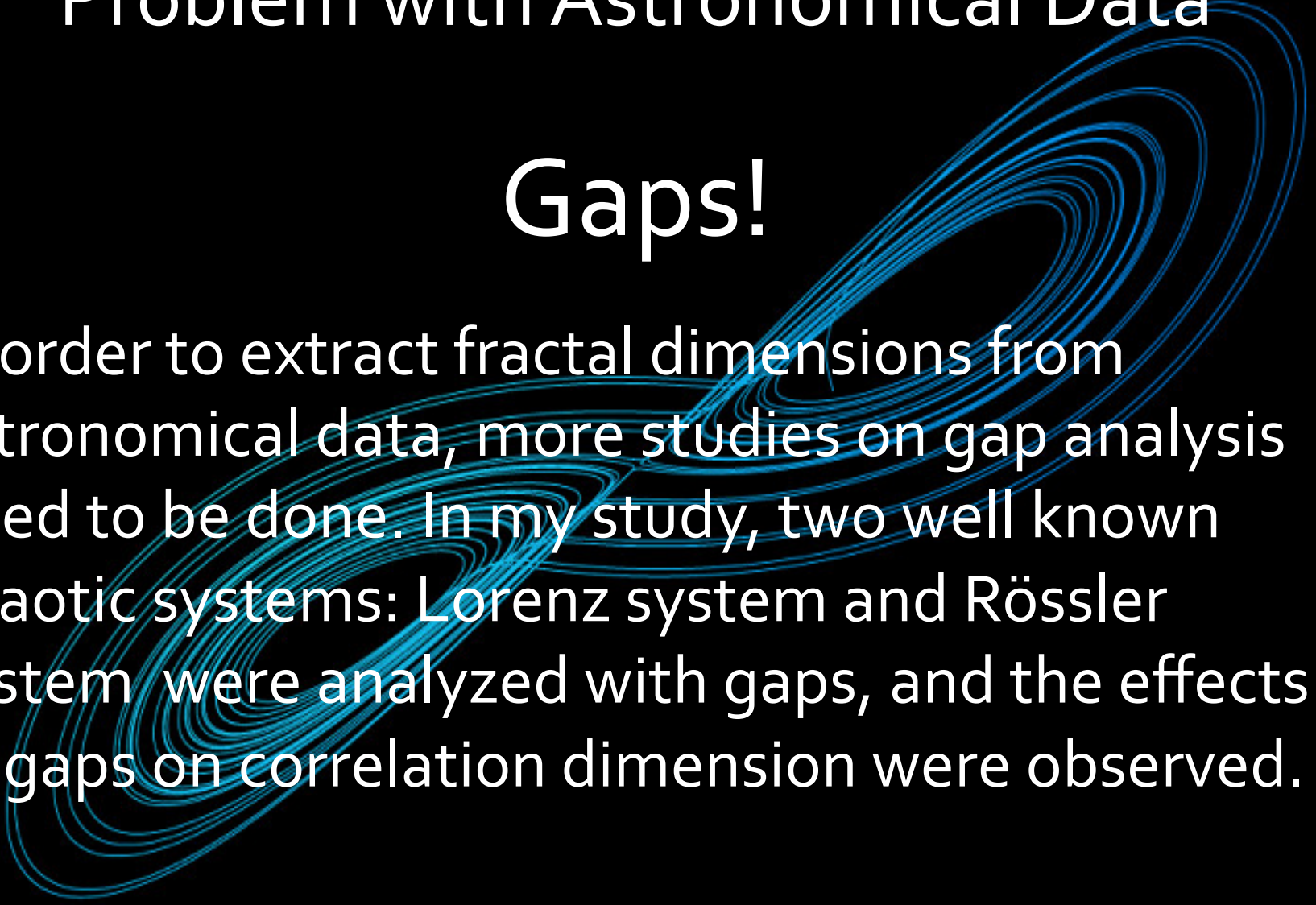


Detrend & Rescale



Problem with Astronomical Data

Gaps!



In order to extract fractal dimensions from astronomical data, more studies on gap analysis need to be done. In my study, two well known chaotic systems: Lorenz system and Rössler system were analyzed with gaps, and the effects of gaps on correlation dimension were observed.

Lorenz System & Rössler System

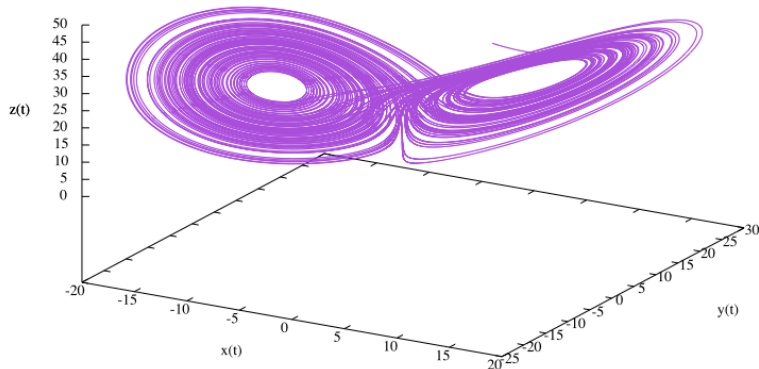
$$\frac{dx}{dt} = \sigma(y - x), \quad \frac{dy}{dt} = x(r - z) - y, \quad \frac{dz}{dt} = xy - bz$$

$$\sigma = 10, \quad r = 28, \quad b = 8/3$$

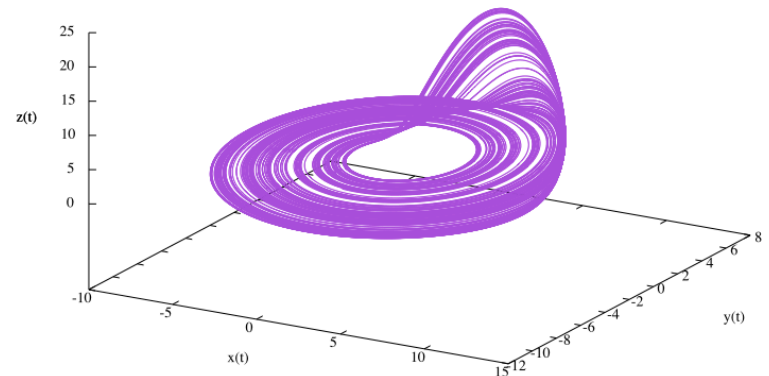
$$\frac{dx}{dt} = -y - z, \quad \frac{dy}{dt} = x + ay, \quad \frac{dz}{dt} = b + z(x - c)$$

$$a = 0.2, \quad b = 0.2, \quad c = 5.7$$

Lorenz Attractor



Rössler Attractor



Data Analysis

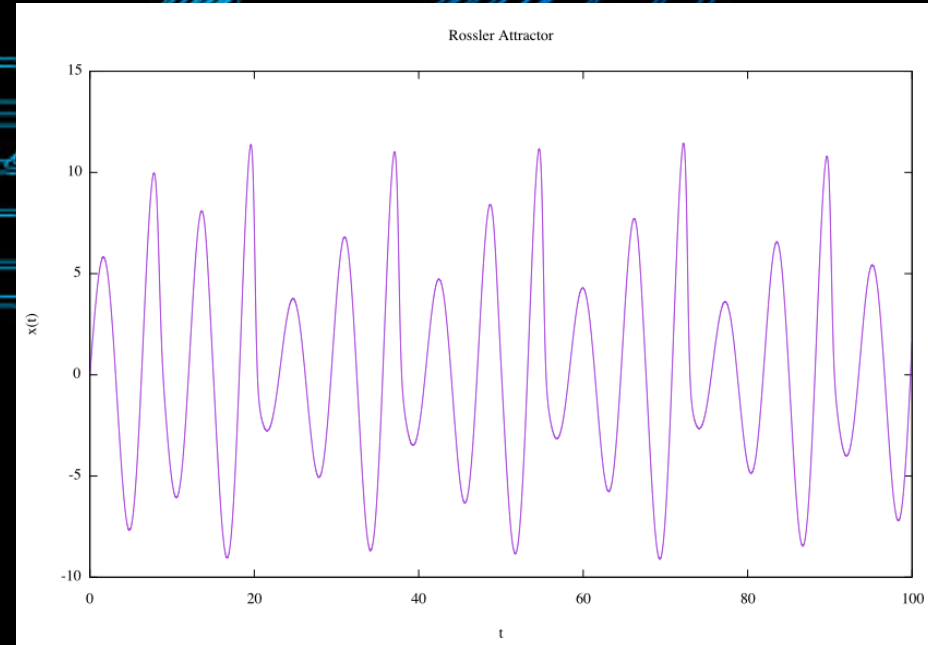
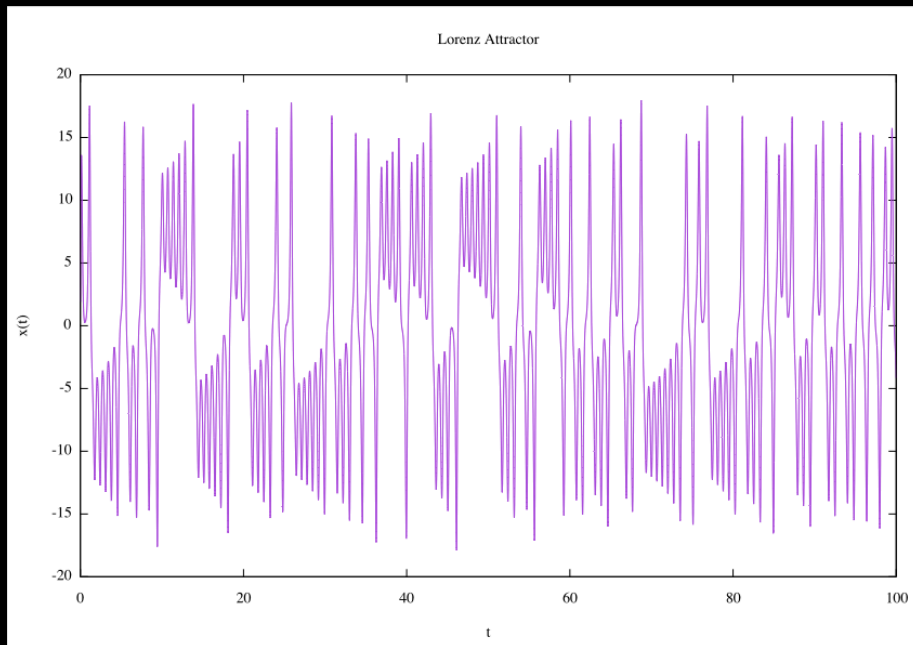


An Engineering Hat...



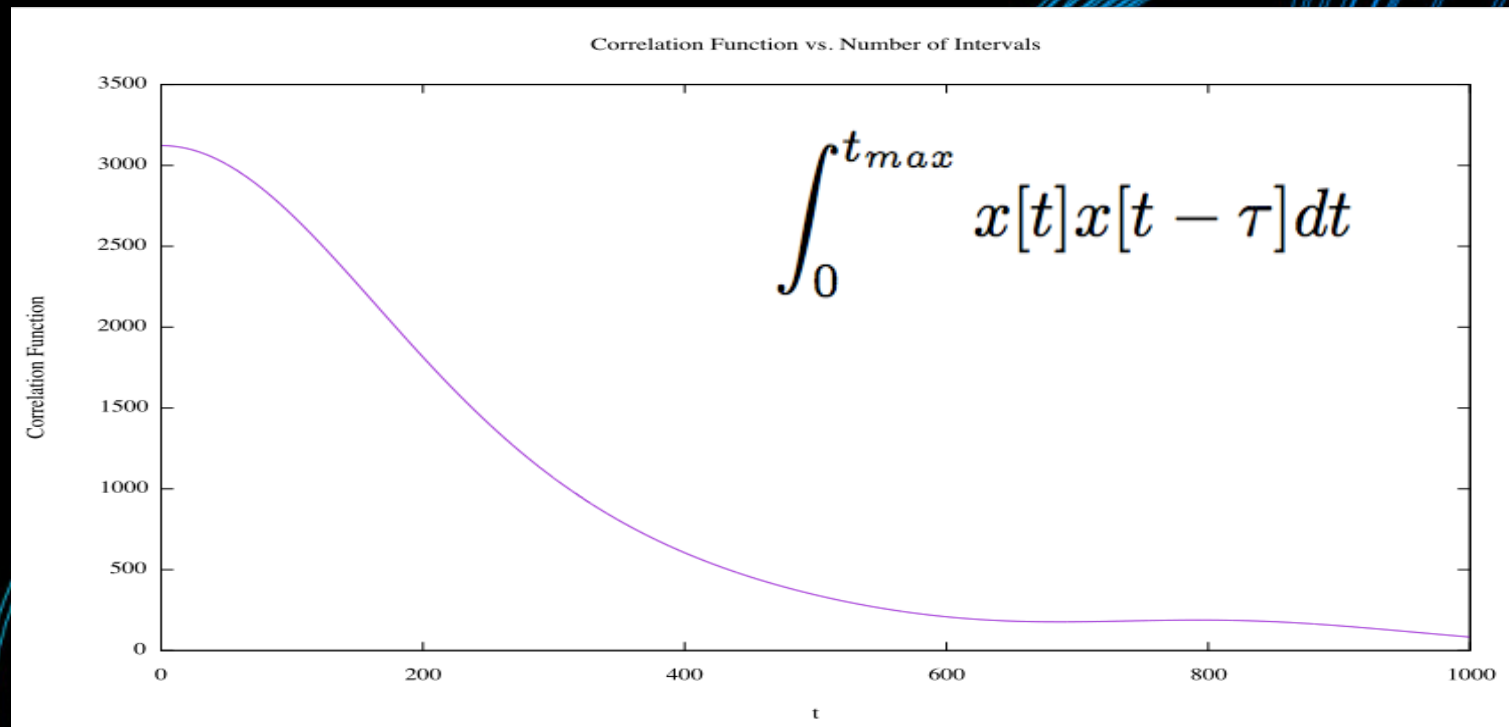
Astronomical Data

Astronomical data only has one dimensional data with time;
Intensity as a function of time!



Basic Flow of Data Analysis

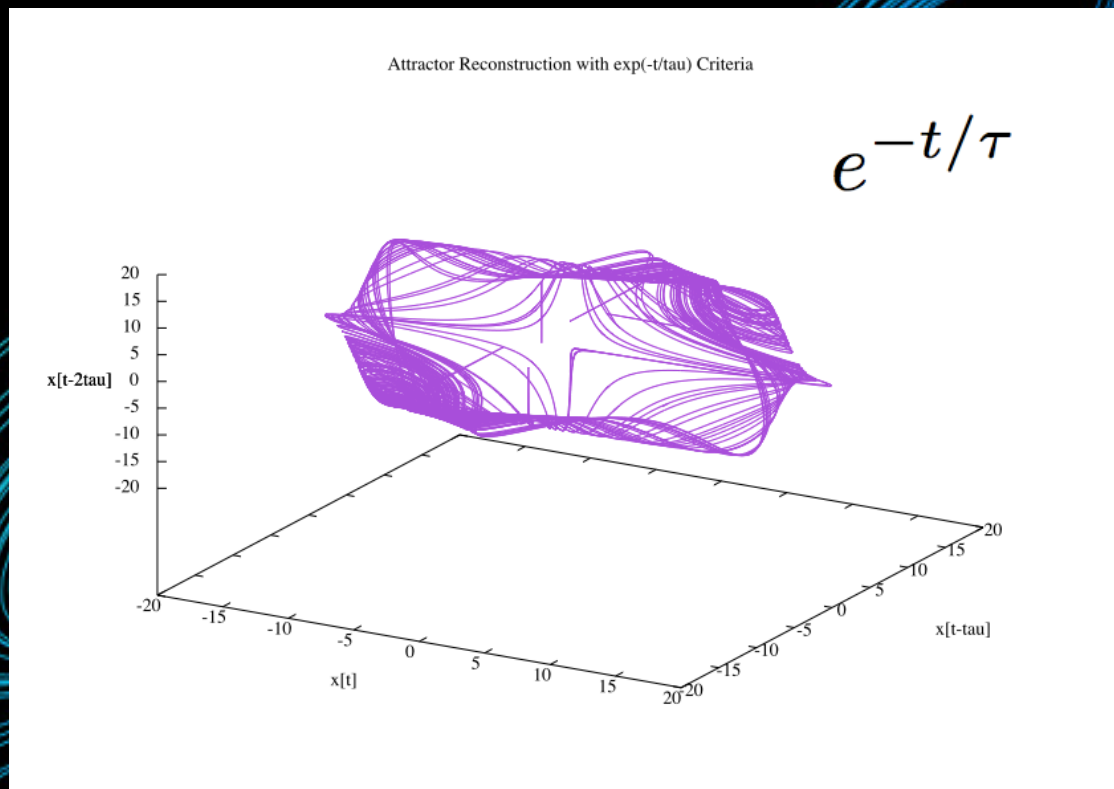
Delay Time Embedding



One well used way to find the value of τ is using 1/e criteria

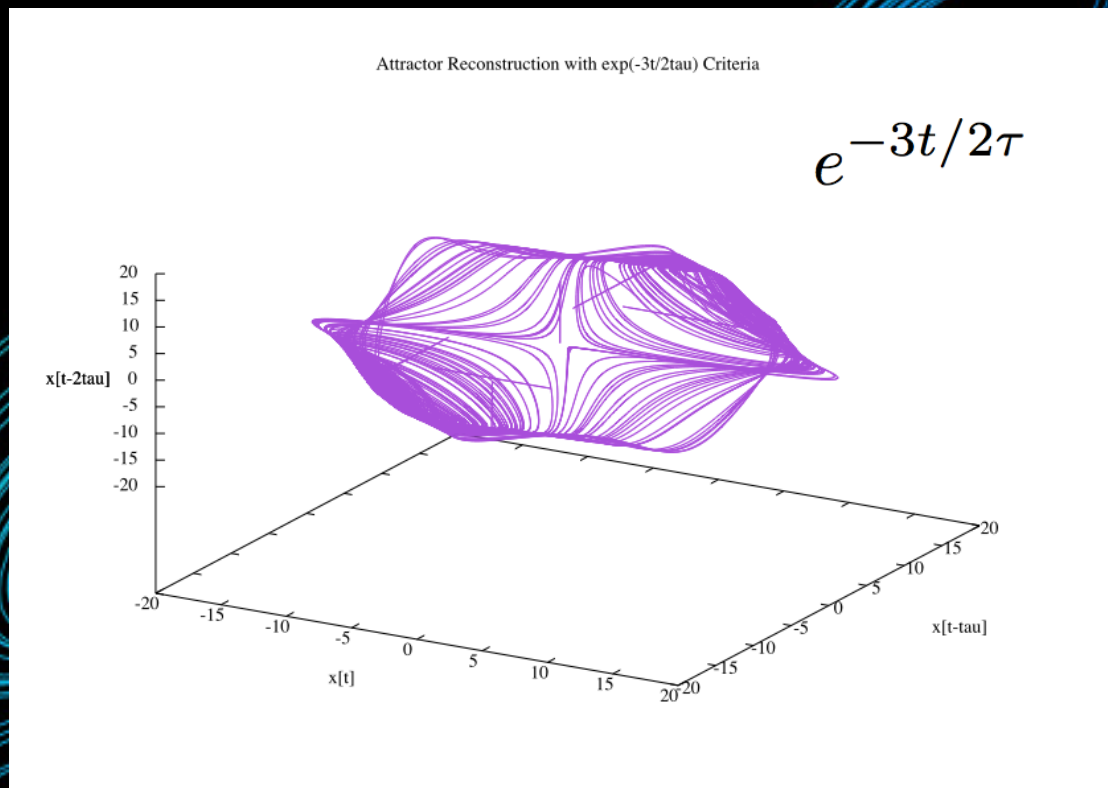
Basic Flow of Data Analysis

Delay Time Embedding



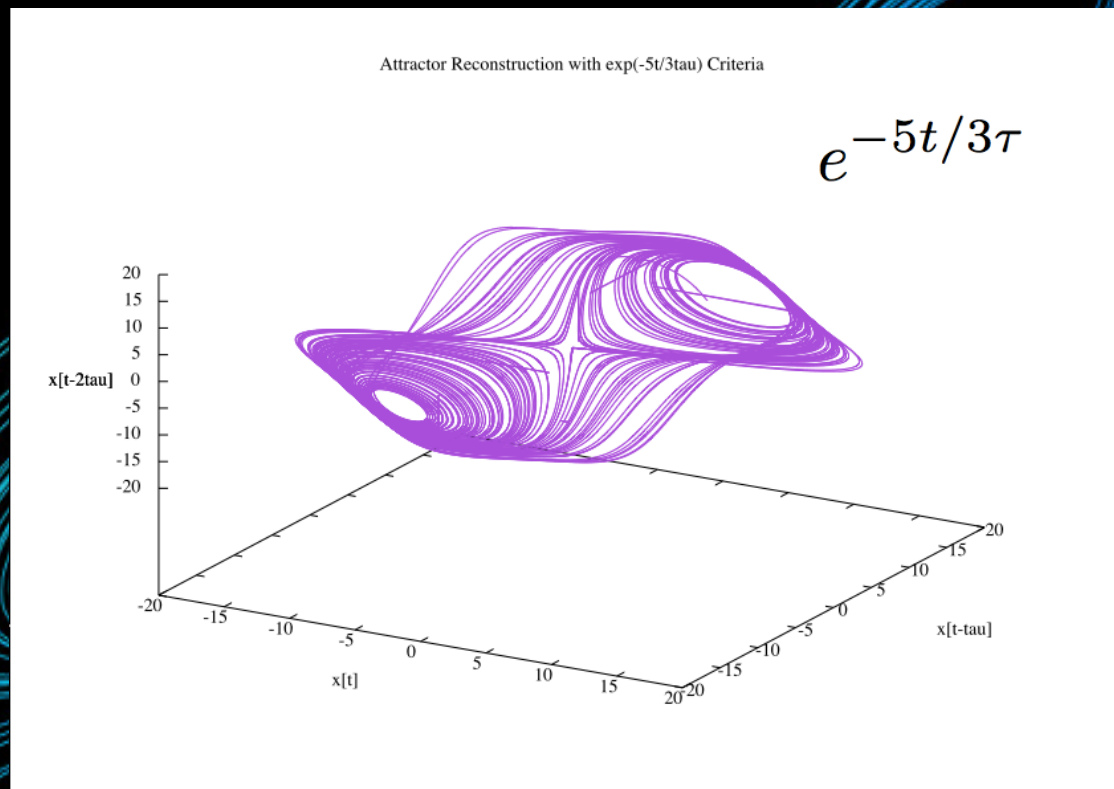
Basic Flow of Data Analysis

Delay Time Embedding



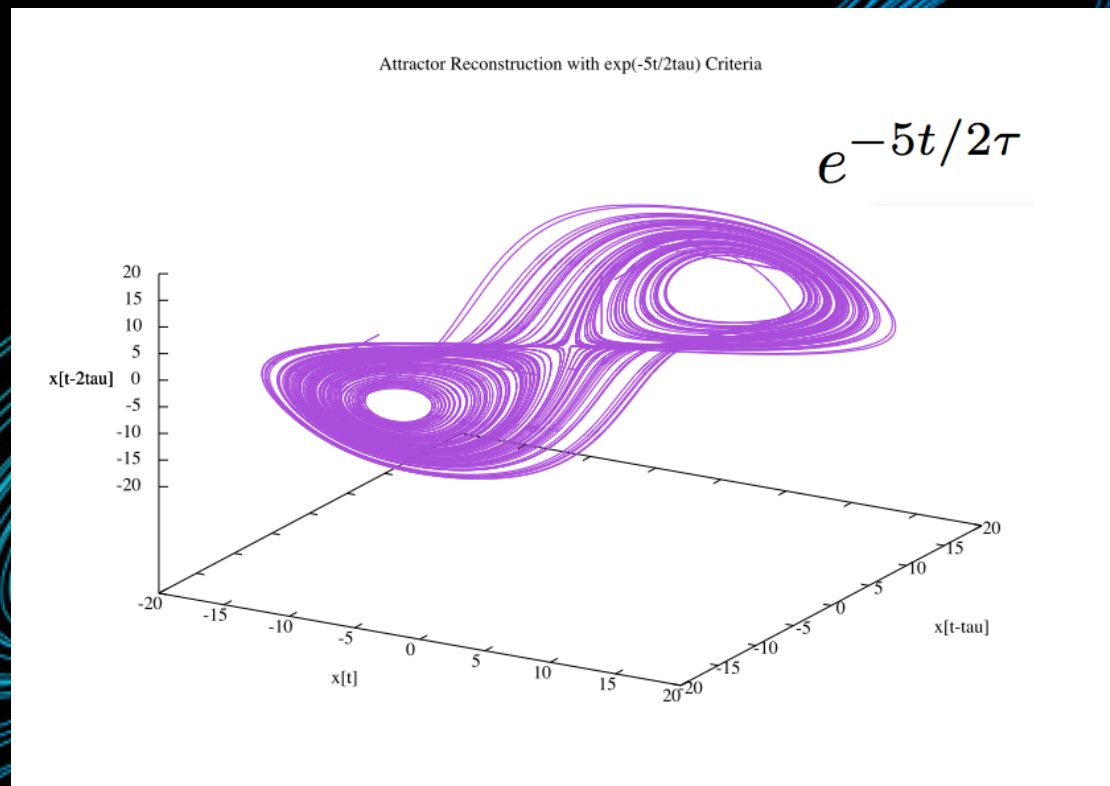
Basic Flow of Data Analysis

Delay Time Embedding



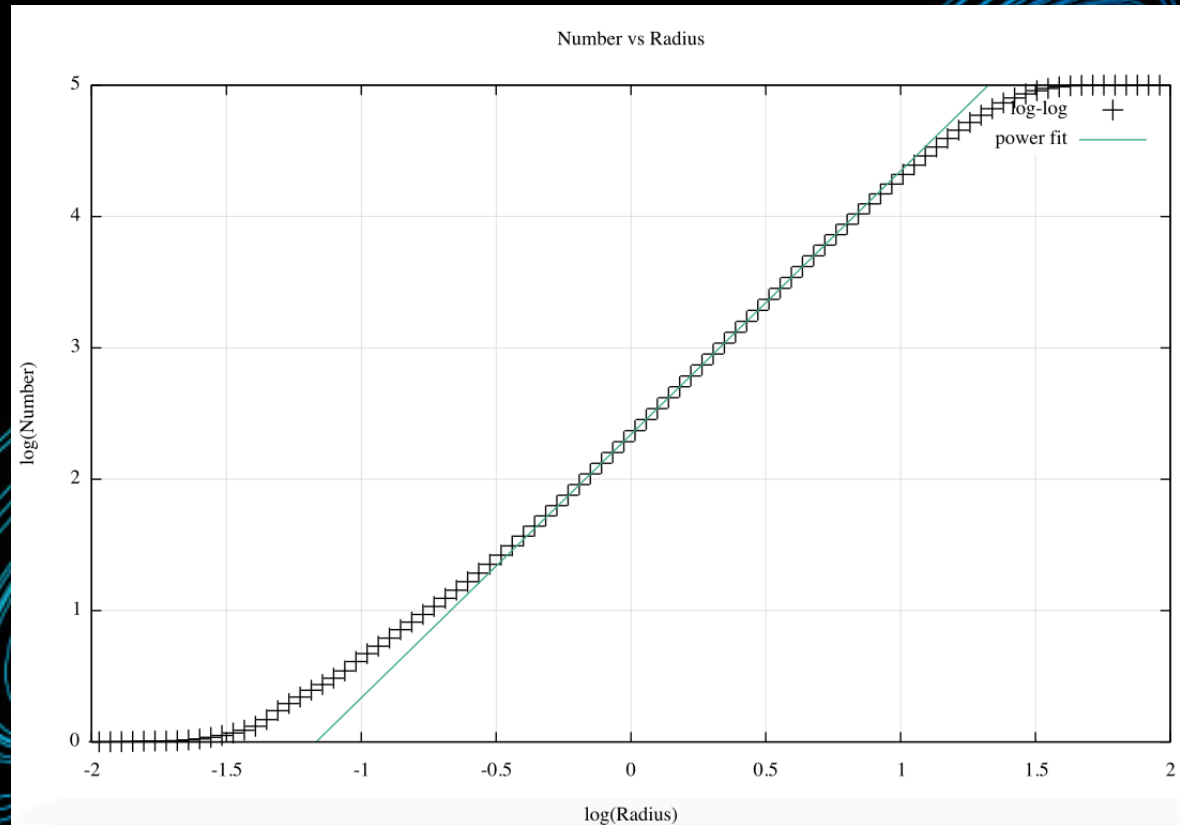
Basic Flow of Data Analysis

Delay Time Embedding



Basic Flow of Data Analysis

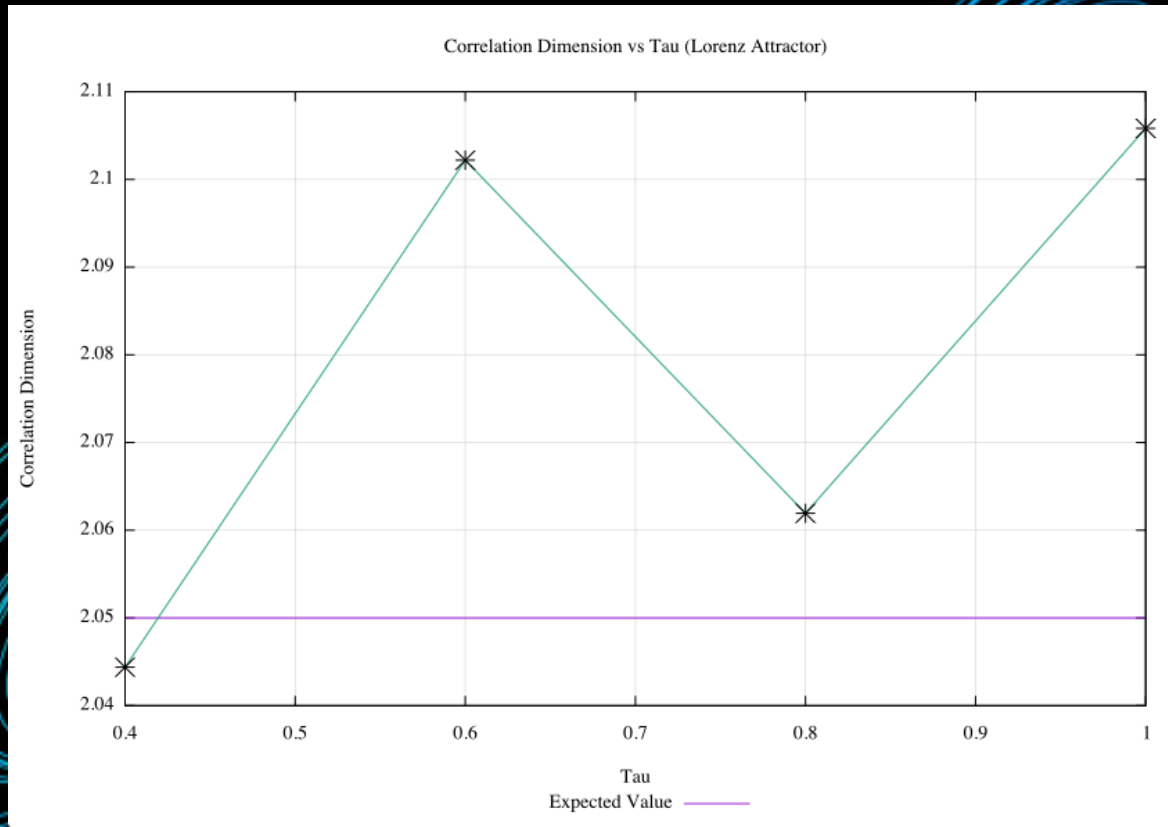
Correlation Dimension



Number of points as a function of the radius of a sphere.
Non integer values of correlation dimension indicate fractal geometry.

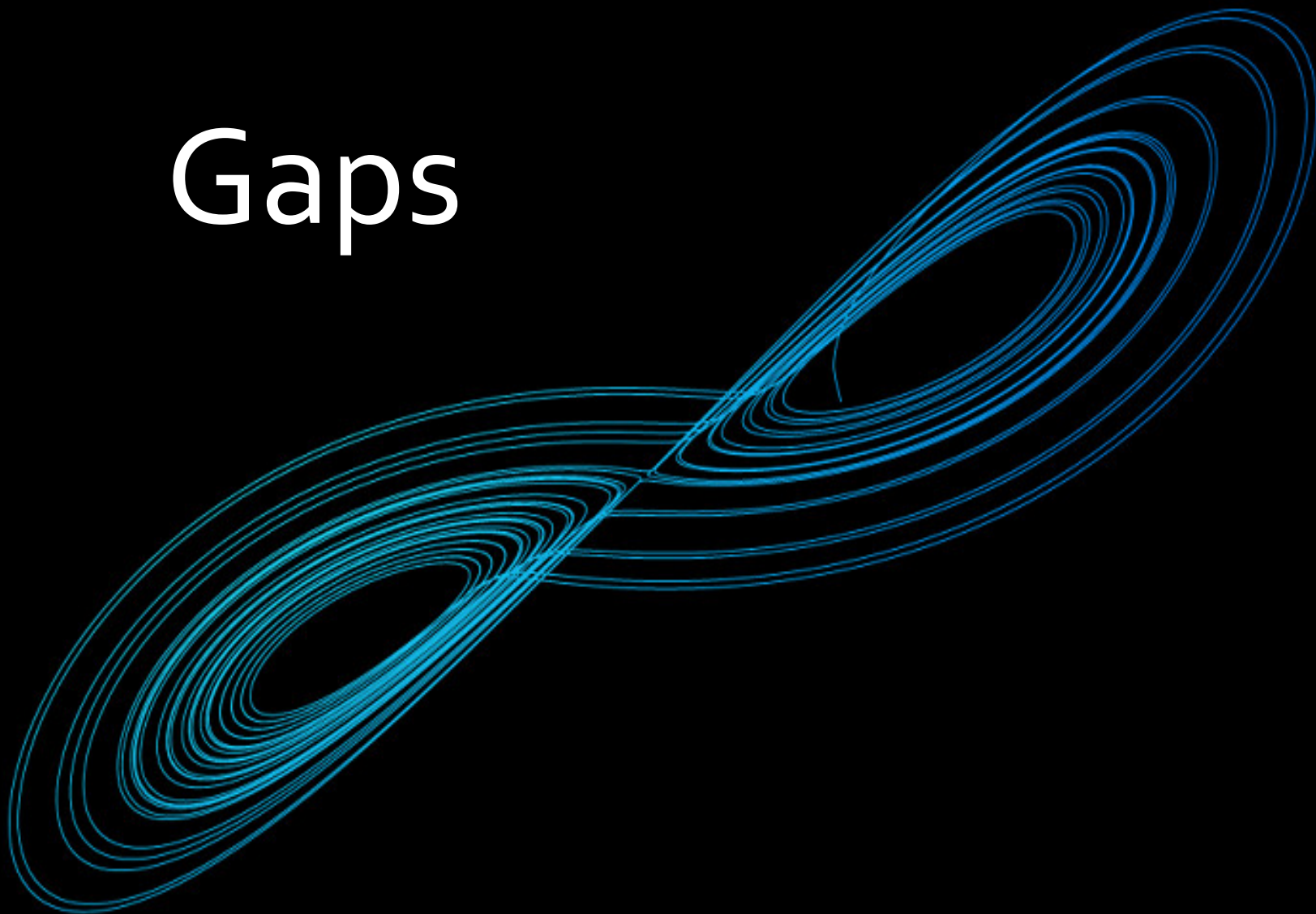
Basic Flow of Data Analysis

Delay Time Embedding & Correlation Dimension

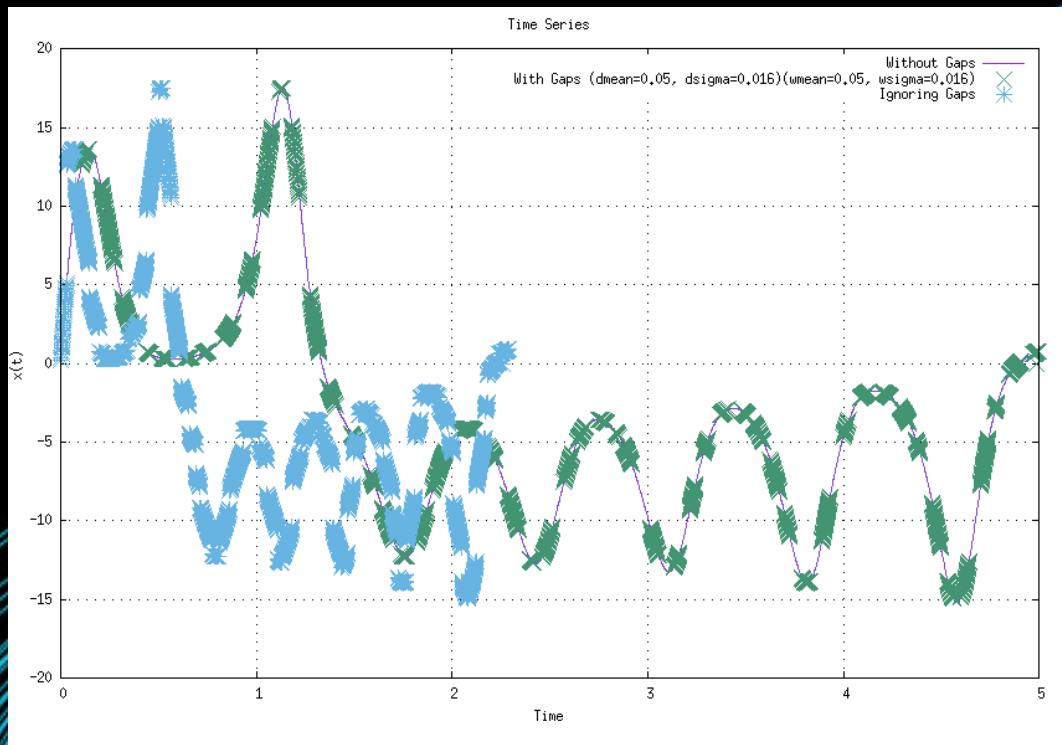


As expected, choice of tau does not affect values of correlation dimension significantly. However, $\exp(-5t/2\tau)$ criteria yielded the best estimation value in this case. It was the case for the Rössler system as well, so $\exp(-5t/2\tau)$ criteria was used to the following gap analysis.

Gaps



Introducing Gaps



George et al. (2014, 2015)

Gap widths and distributions follow two independent Gaussian distributions. The mean value of gap widths (w-mean) and the mean value of data point distribution (d-mean) are the two parameters controlling gaps. No interpolation was used.

Introducing Gaps

Two cases were considered in this study.

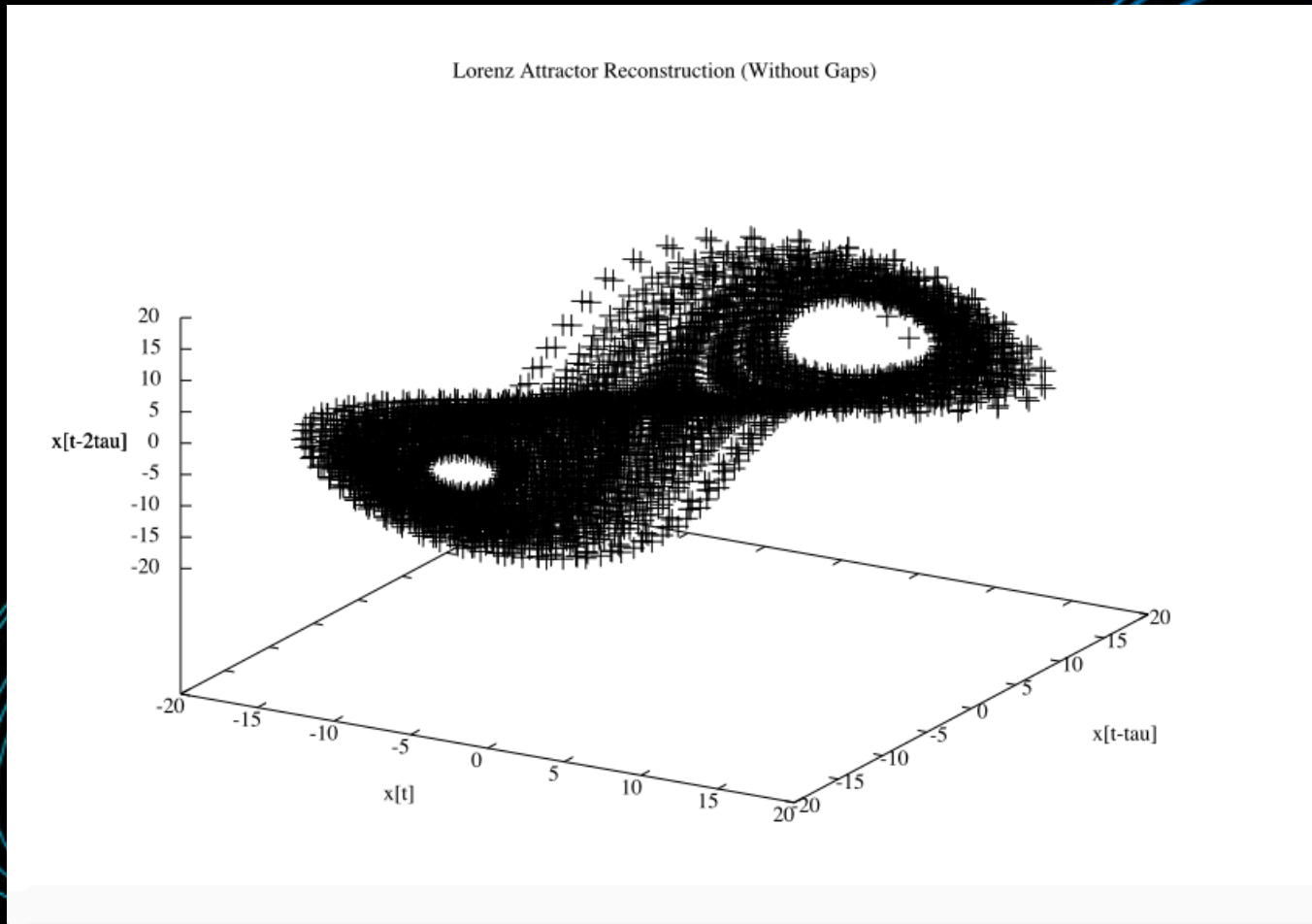
1. Fixed Duration of Sampling Time

With the fixed duration of sampling time, the number of data points decreases as the number and width of gaps increase.

2. Fixed Number of Data Points

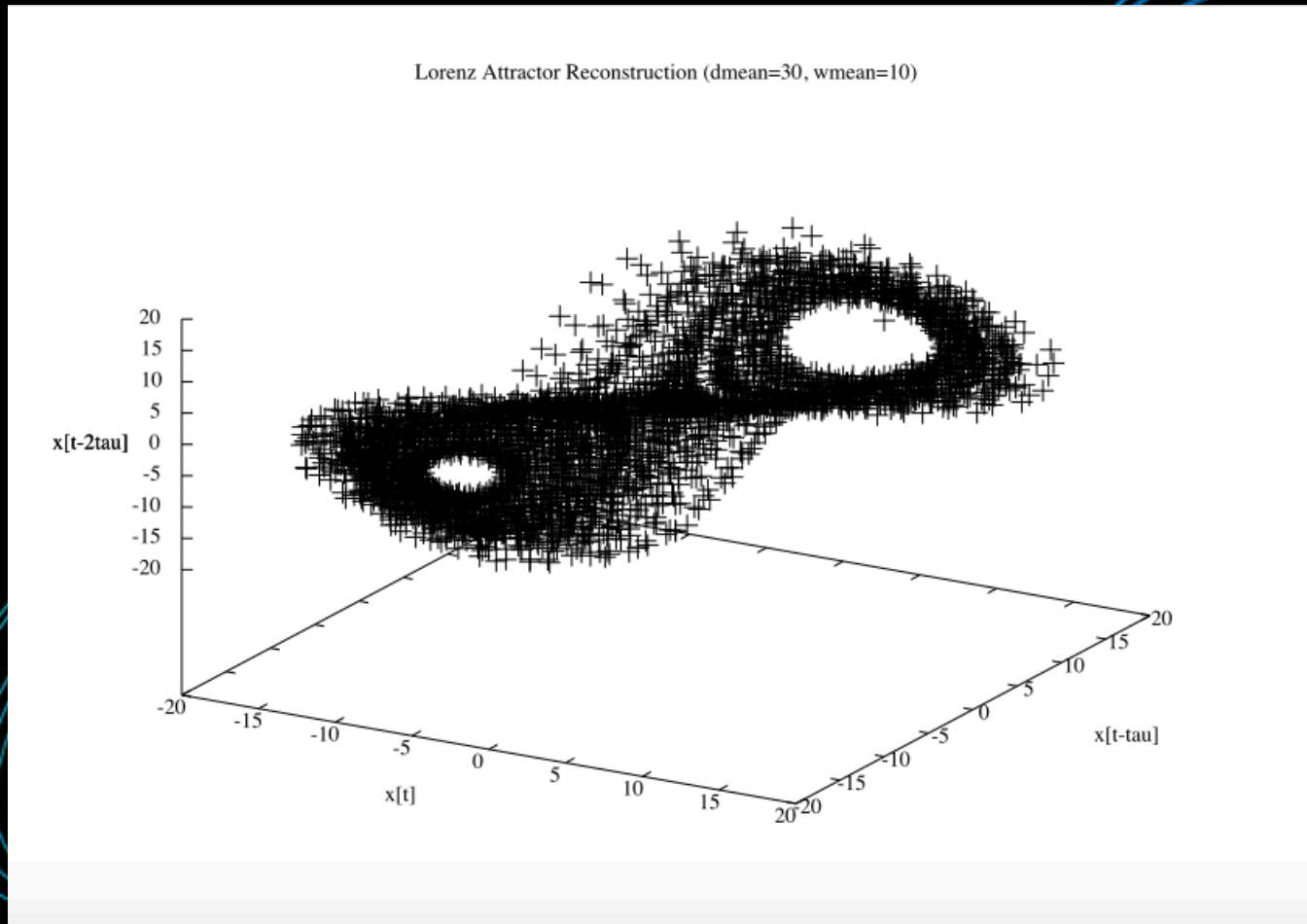
With the fixed number of data points, the duration of sampling time increases as the number and width of gaps increase.

1. Fixed Duration of Sampling Time



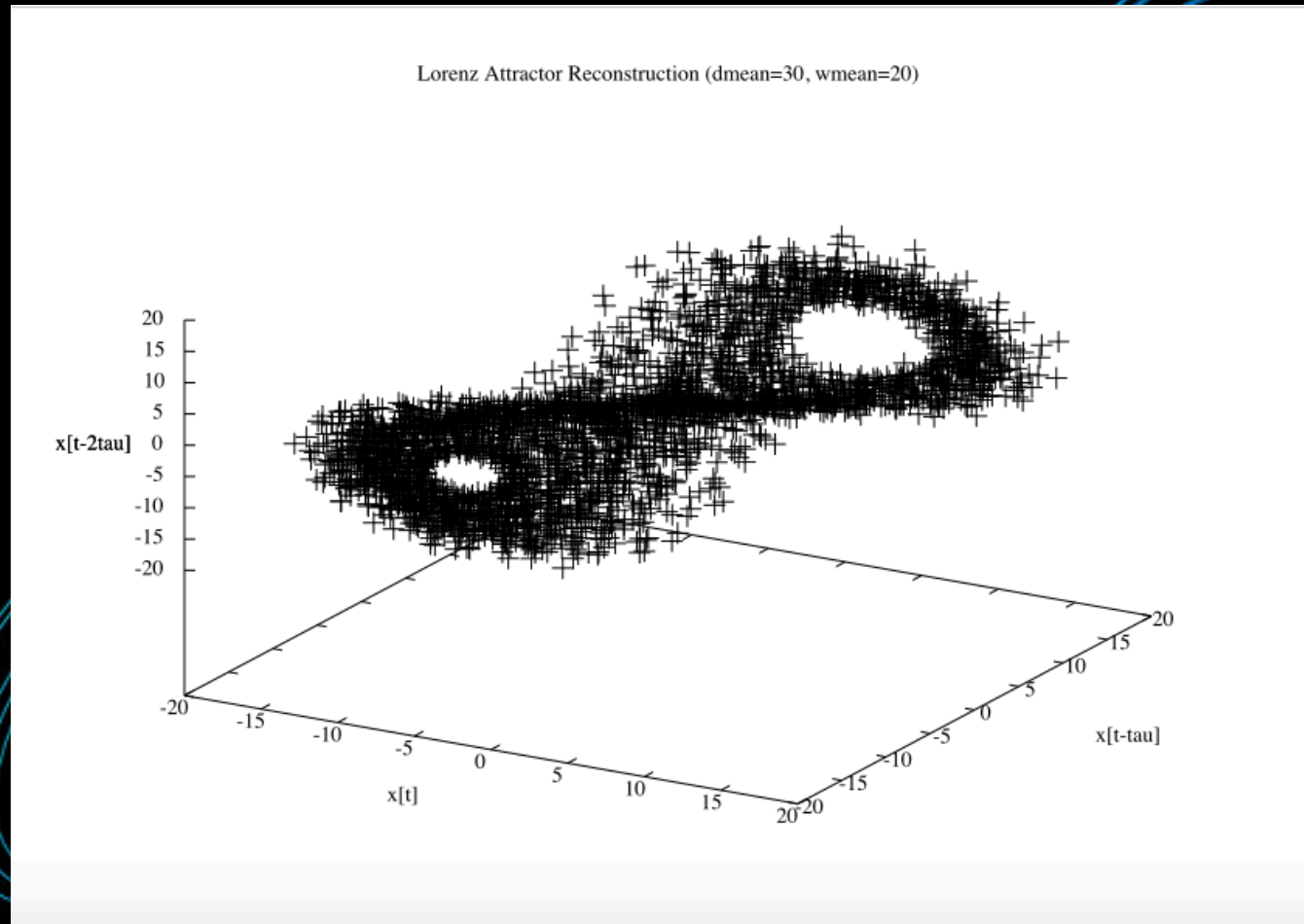
Without Gaps

1. Fixed Duration of Sampling Time



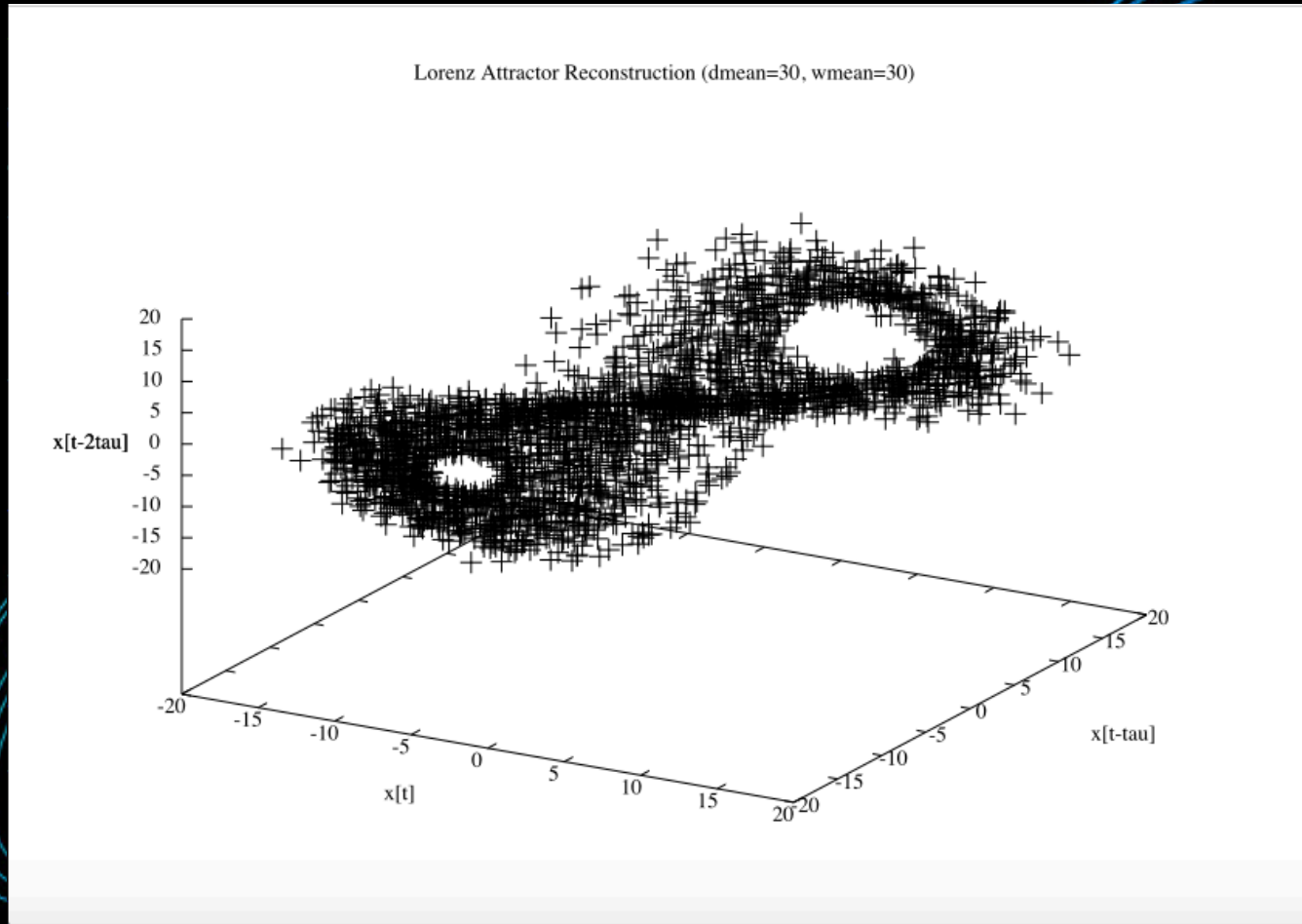
d-mean = 30, w-mean = 10

1. Fixed Duration of Sampling Time



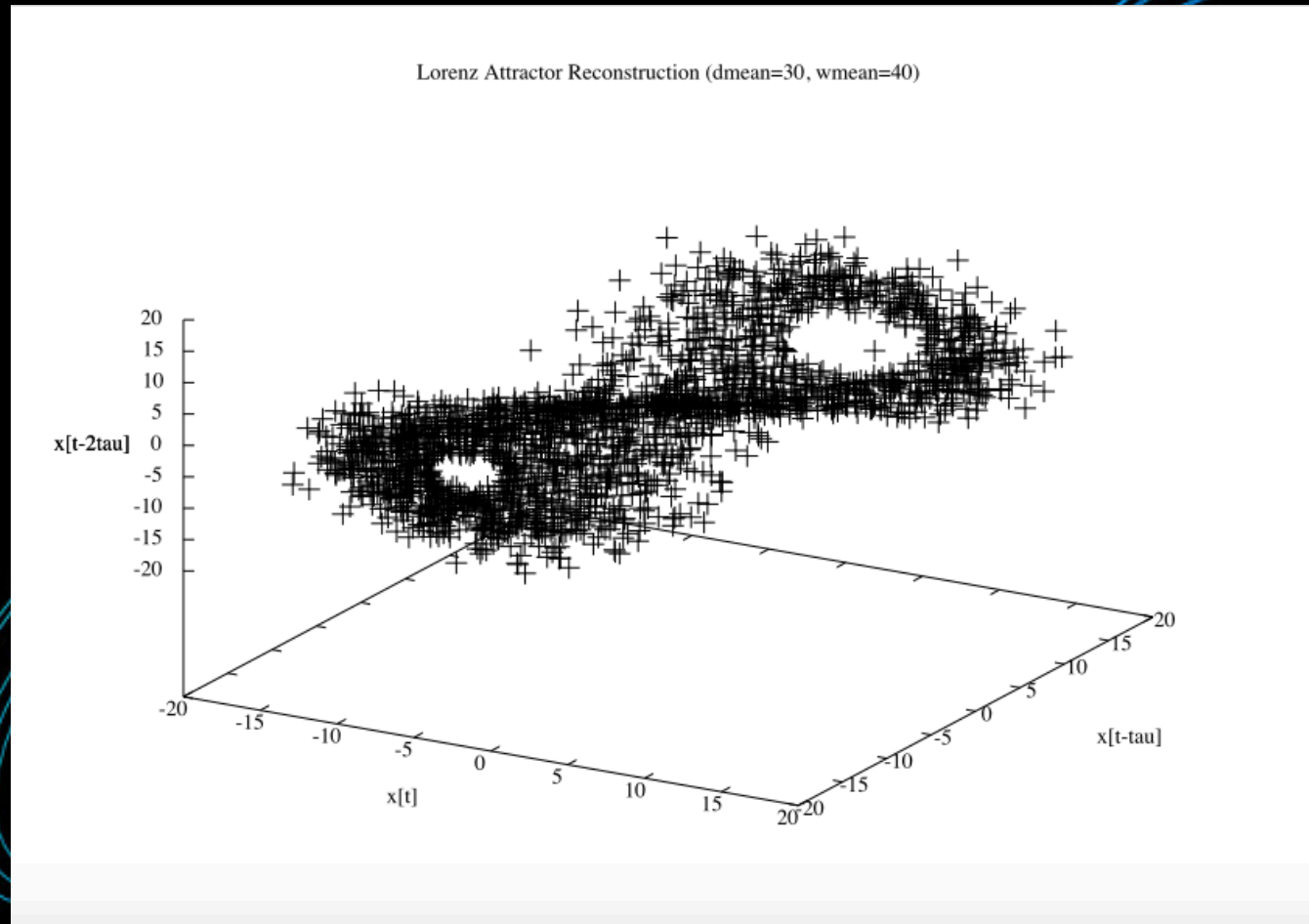
d-mean = 30, w-mean = 20

1. Fixed Duration of Sampling Time



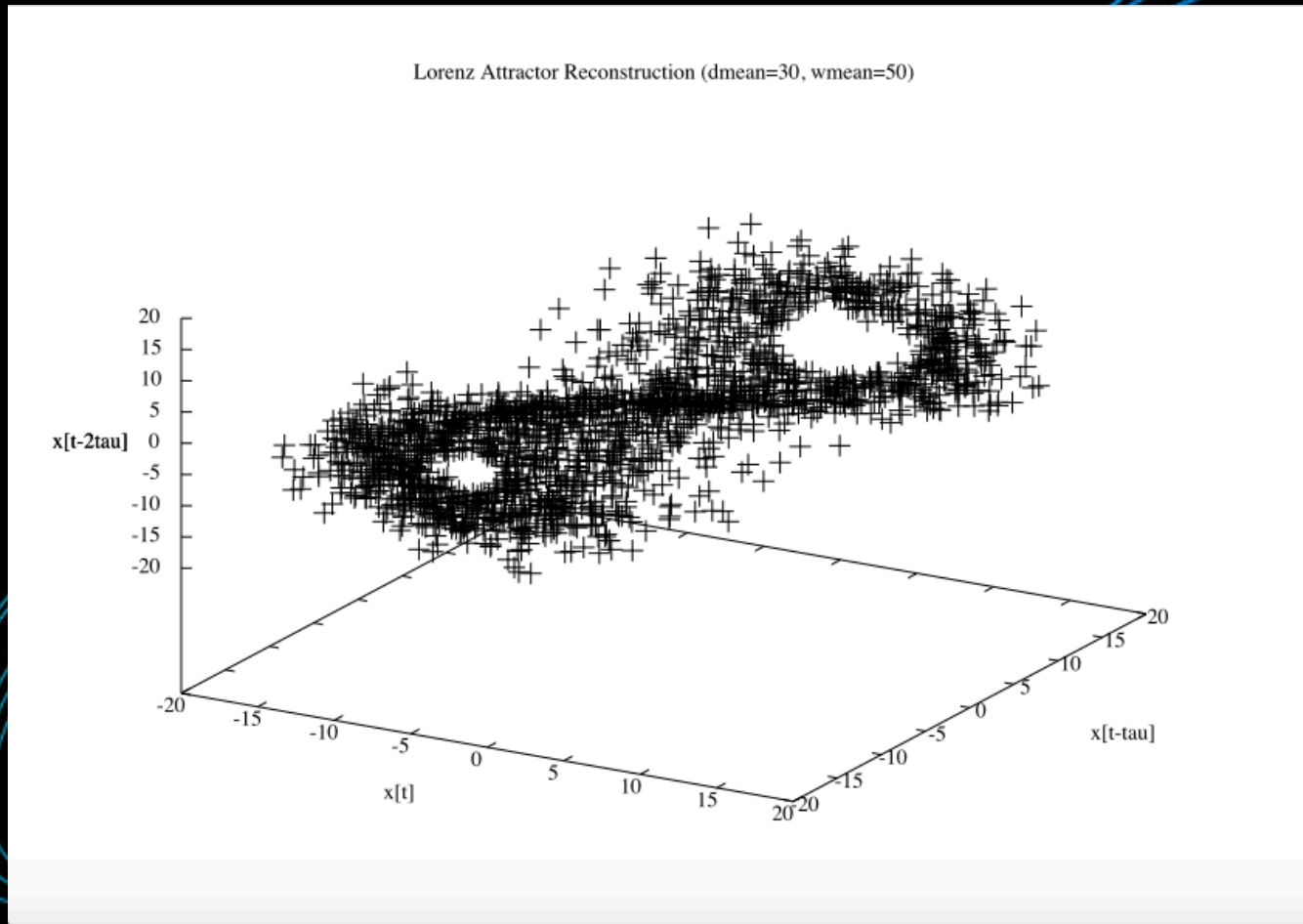
d-mean = 30, w-mean = 30

1. Fixed Duration of Sampling Time



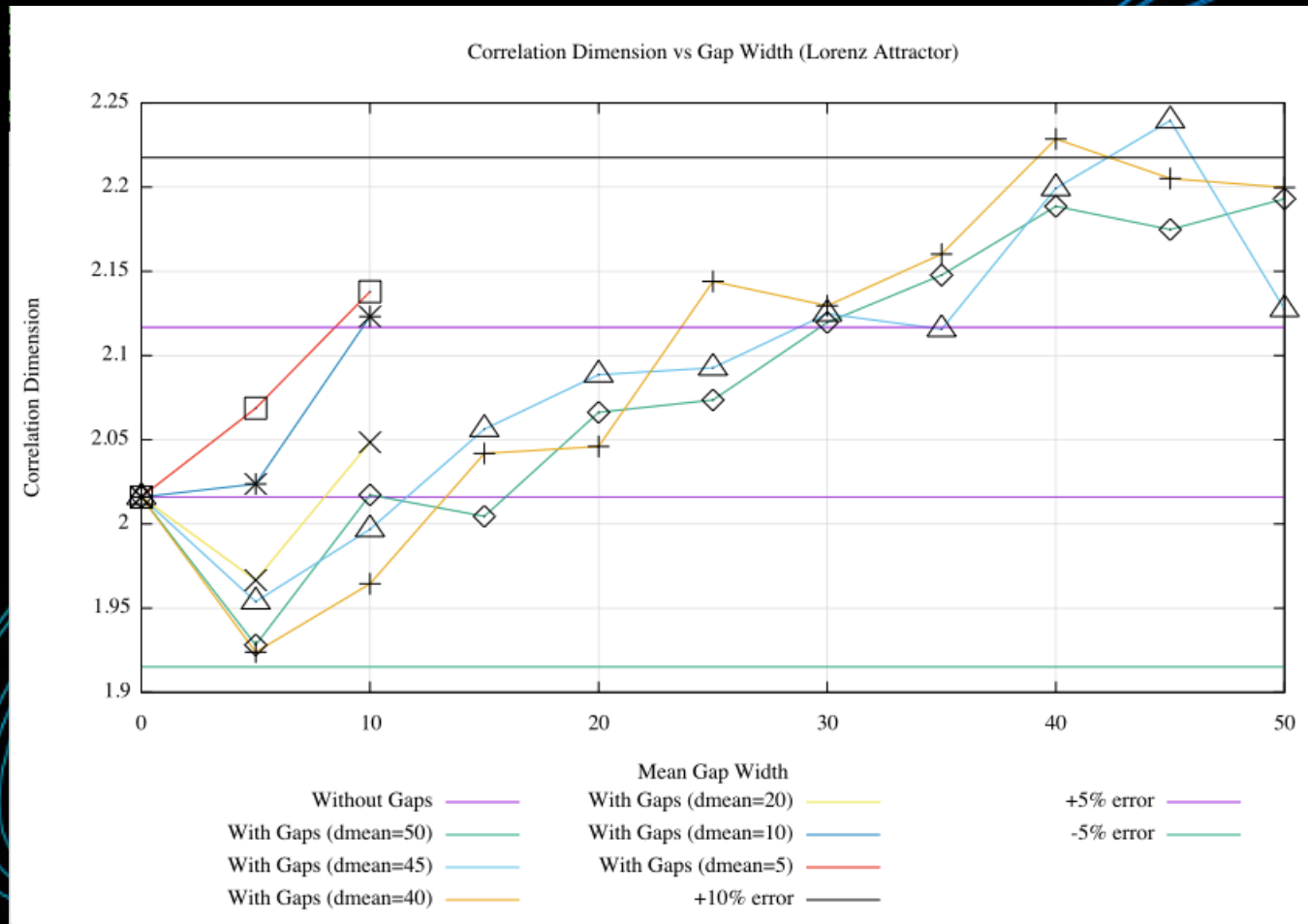
d-mean = 30, w-mean = 40

1. Fixed Duration of Sampling Time



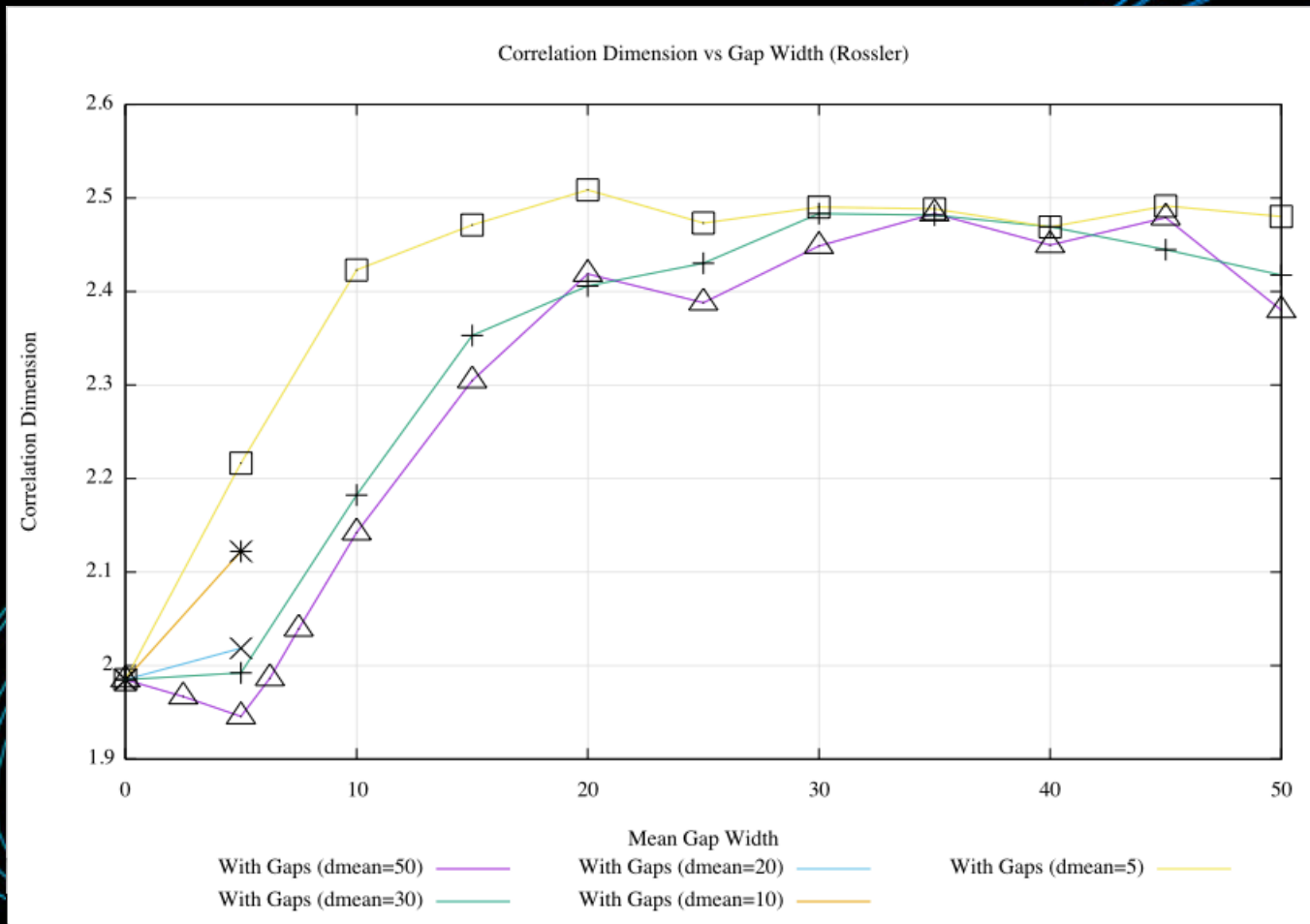
d-mean = 30, w-mean = 50

1. Fixed Duration of Sampling Time



With gaps, correlation dimension tends to be overestimated. This is because as gaps increase, attractors get fuzzier, and some points sit inside unexpected counting spheres.

1. Fixed Duration of Sampling Time

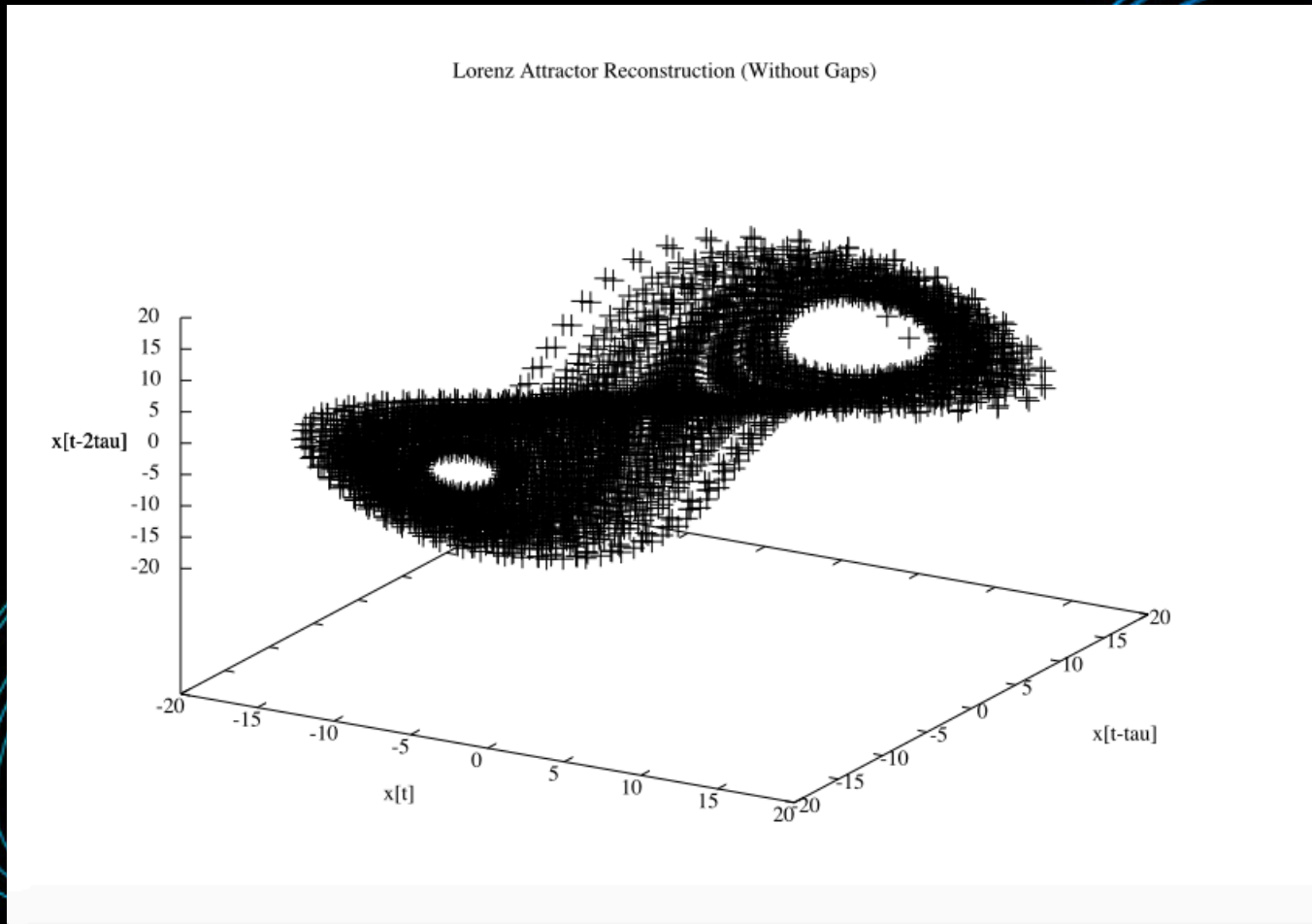


With gaps, correlation dimension tends to be overestimated. This is because as gaps increase, attractors get fuzzier, and some points sit inside unexpected counting spheres.

1. Fixed Duration of Sampling Time

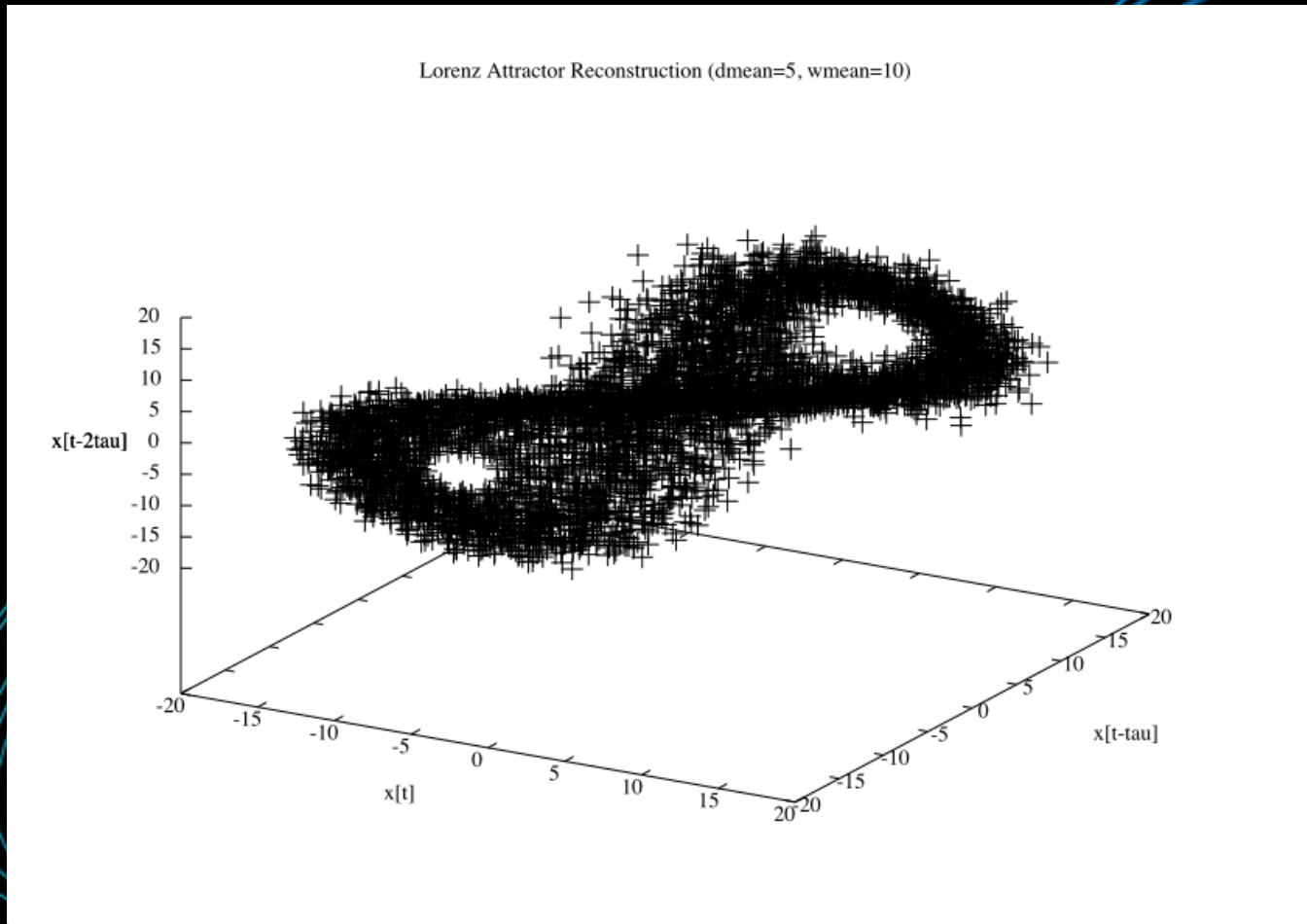
According to the data, the value of correlation dimension saturates at some certain amount of gaps. This is because, even though the attractors get fuzzier, the number of data points decrease, and so correlation dimension cannot be higher than some certain value. For both the Lorenz and Rössler systems, the saturation occurs when about more than 40% of the original data points is lost due to gaps.

2. Fixed Number of Data Points (100,000)



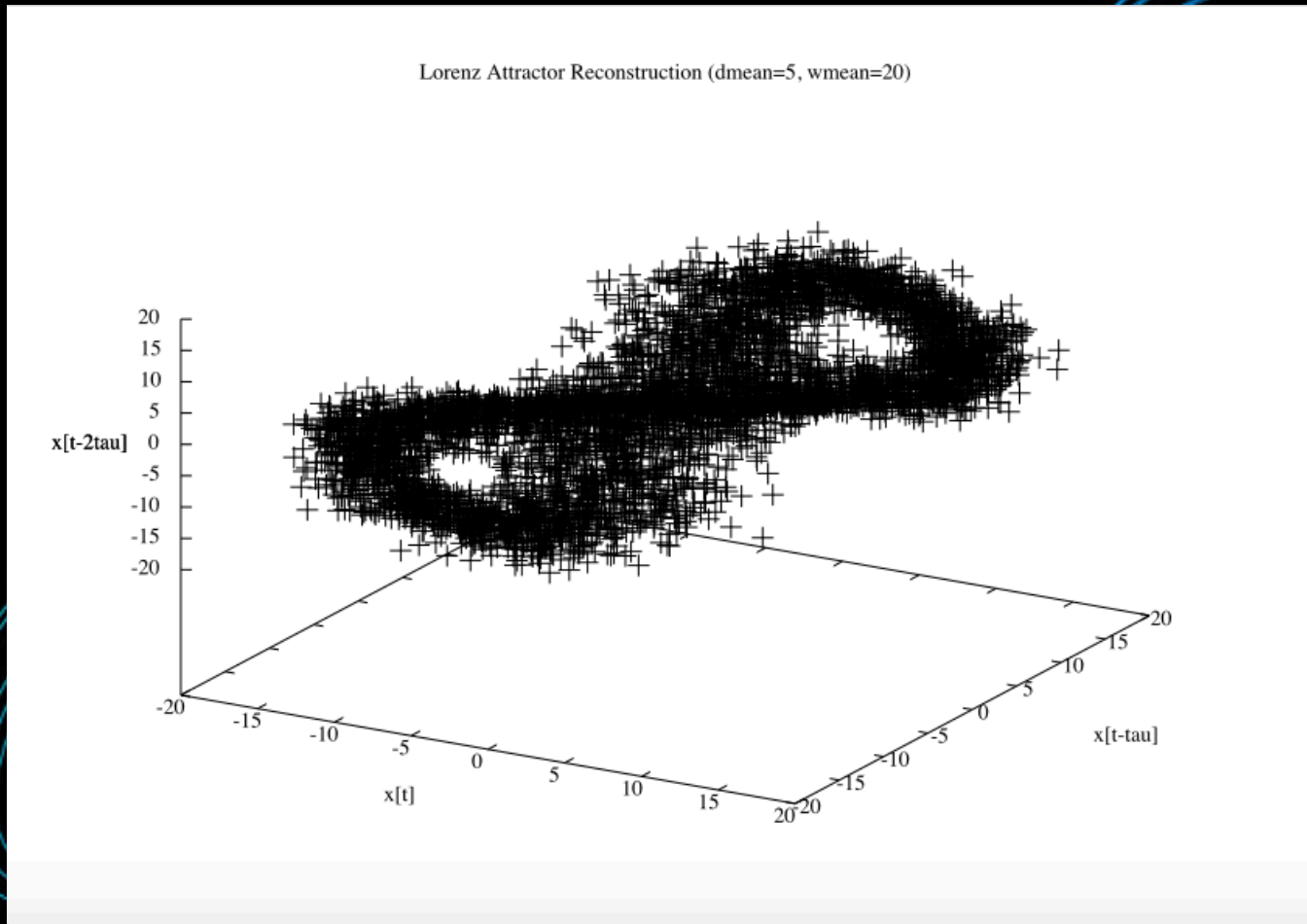
Without Gaps

2. Fixed Number of Data Points (100,000)



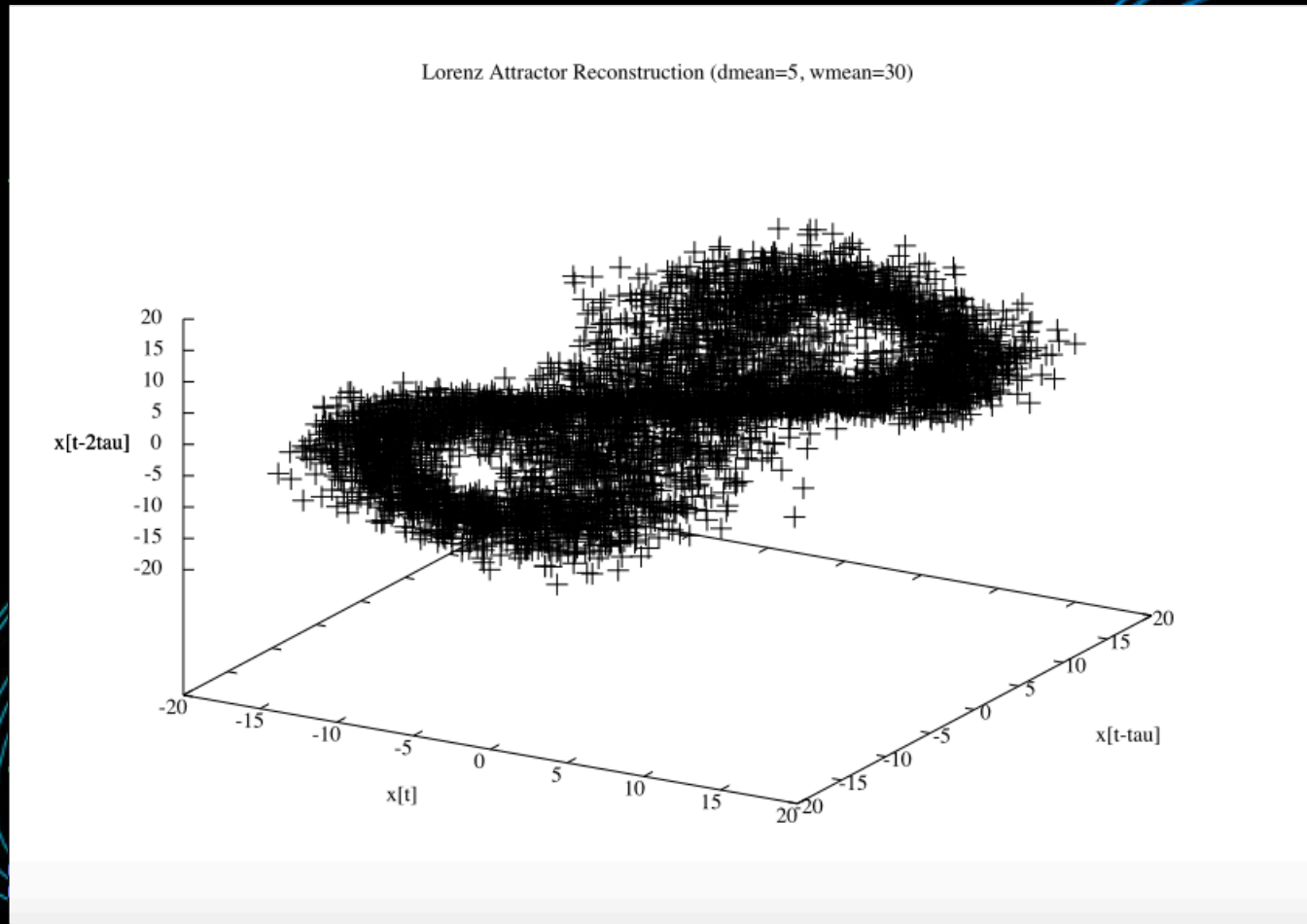
d-mean = 5, w-mean = 10

2. Fixed Number of Data Points (100,000)



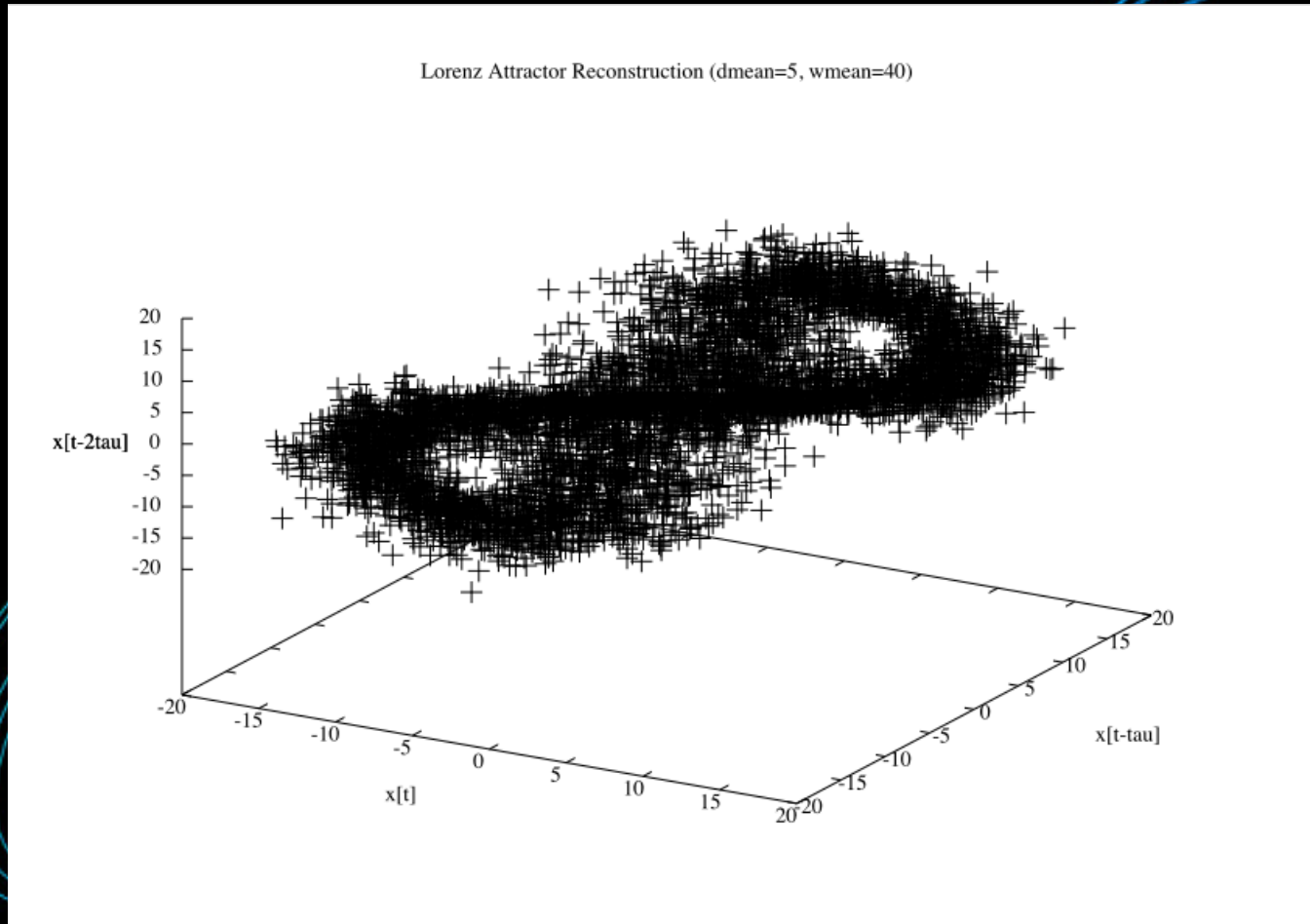
d-mean = 5, w-mean = 20

2. Fixed Number of Data Points (100,000)



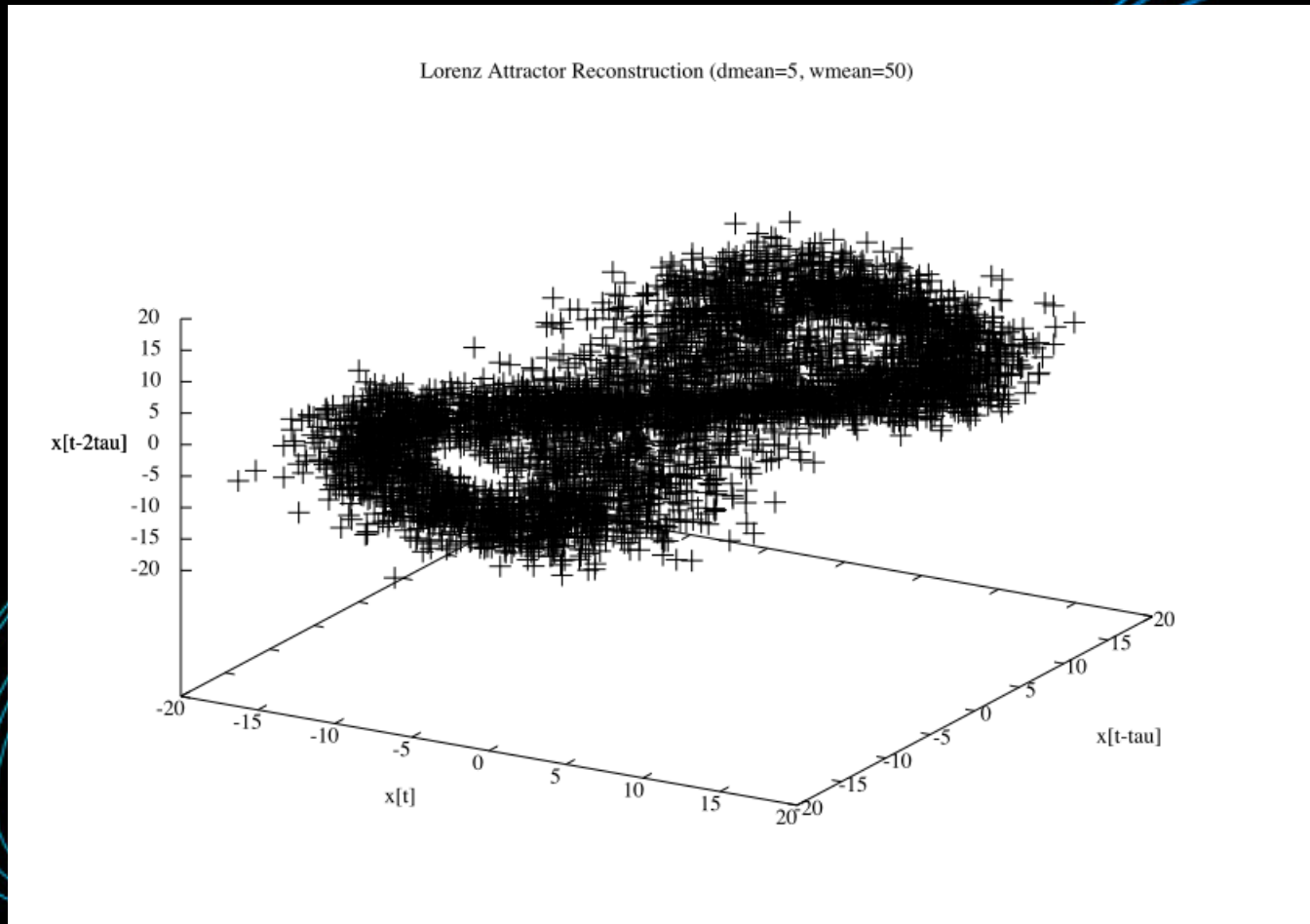
d-mean = 5, w-mean = 30

2. Fixed Number of Data Points (100,000)



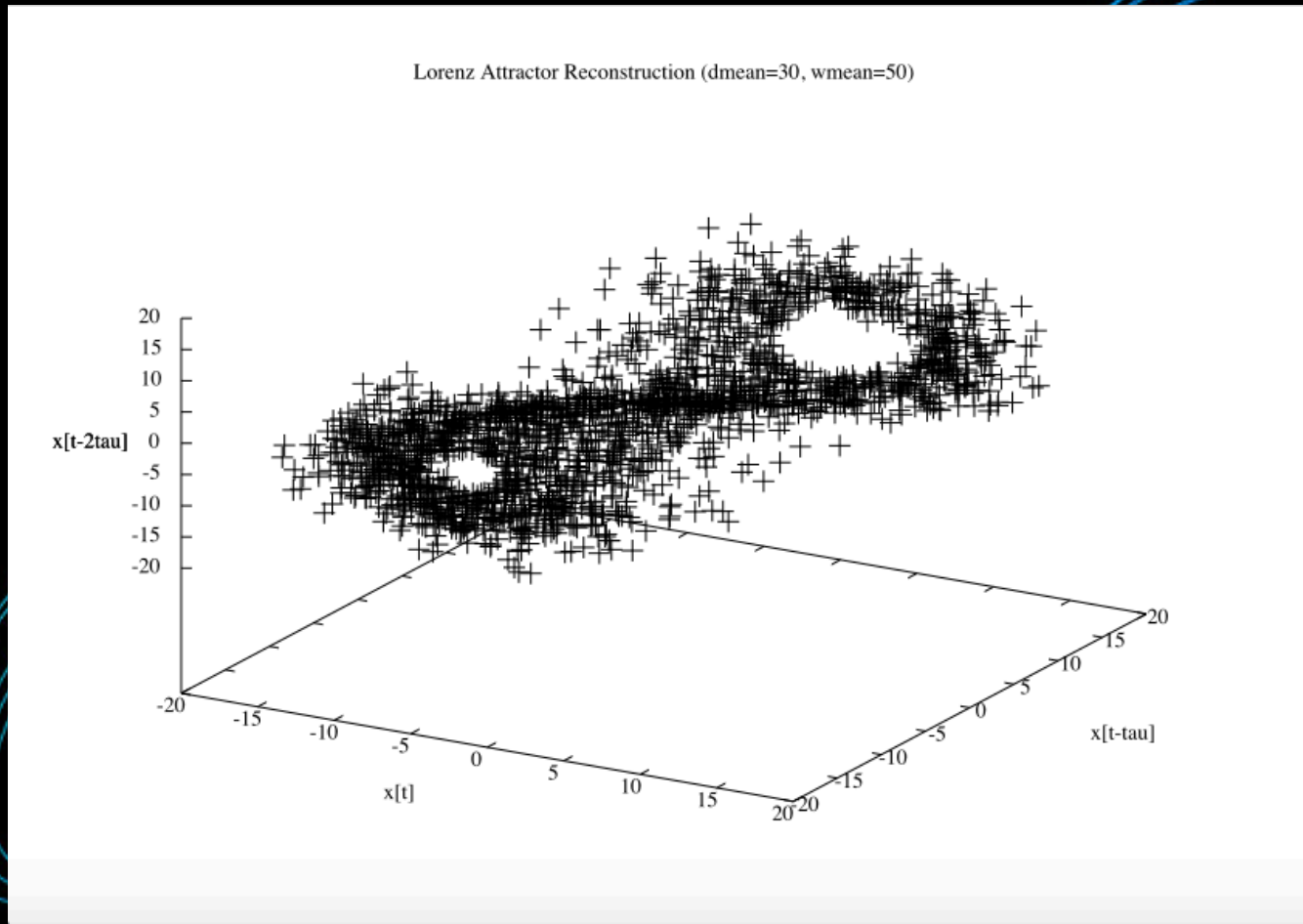
d-mean = 5, w-mean = 40

2. Fixed Number of Data Points (100,000)



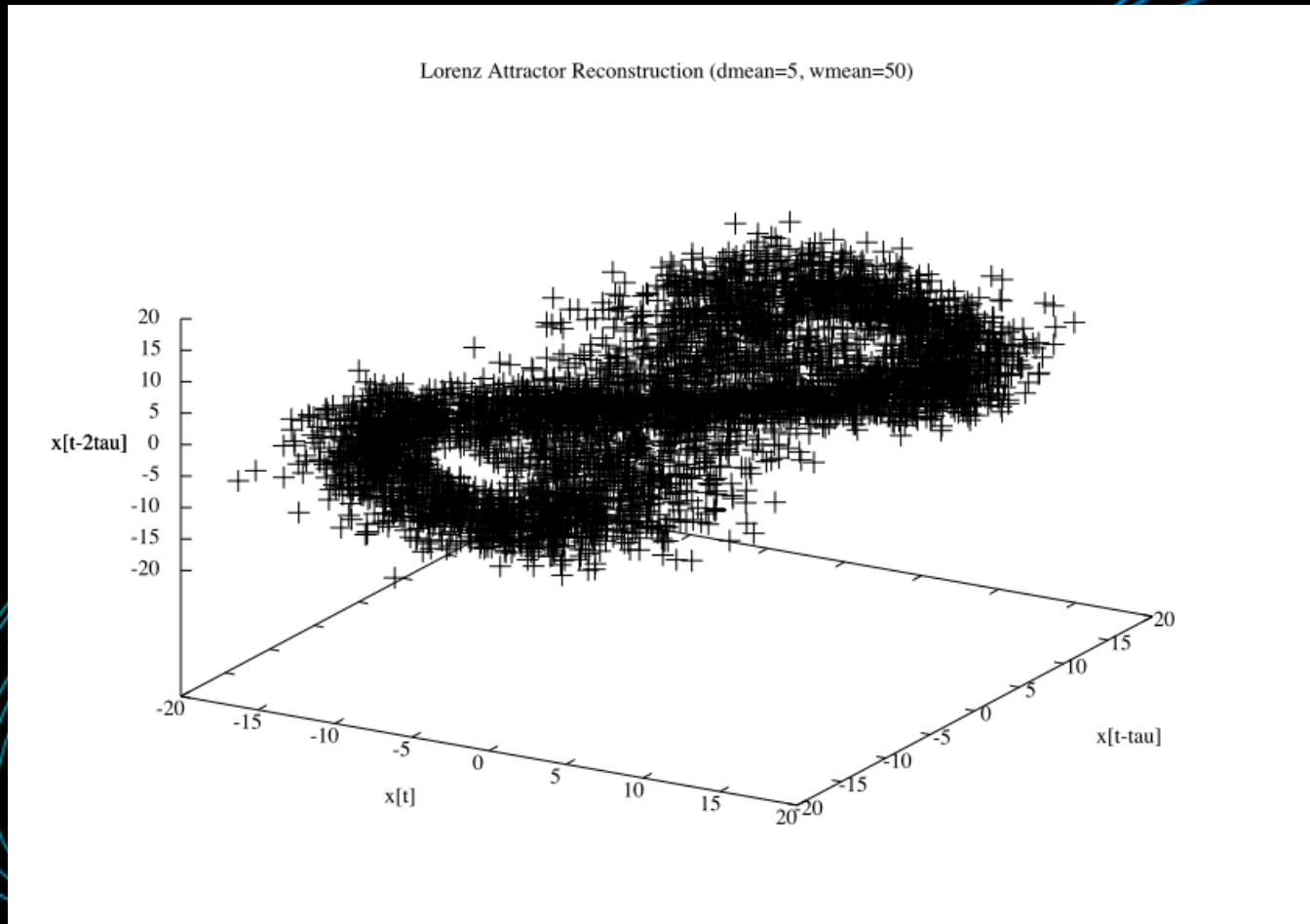
d-mean = 5, w-mean = 50

1. Fixed Duration of Sampling Time



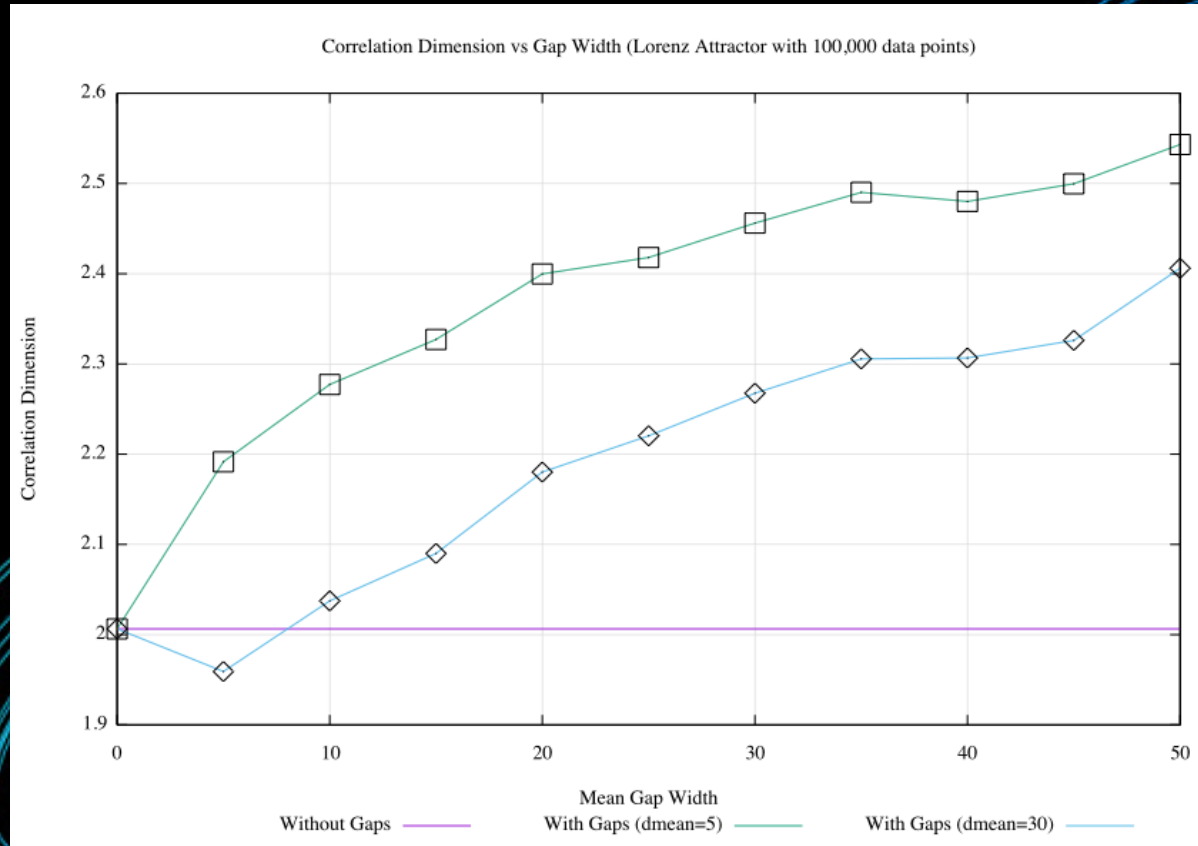
d-mean = 30, w-mean = 50

2. Fixed Number of Data Points (100,000)



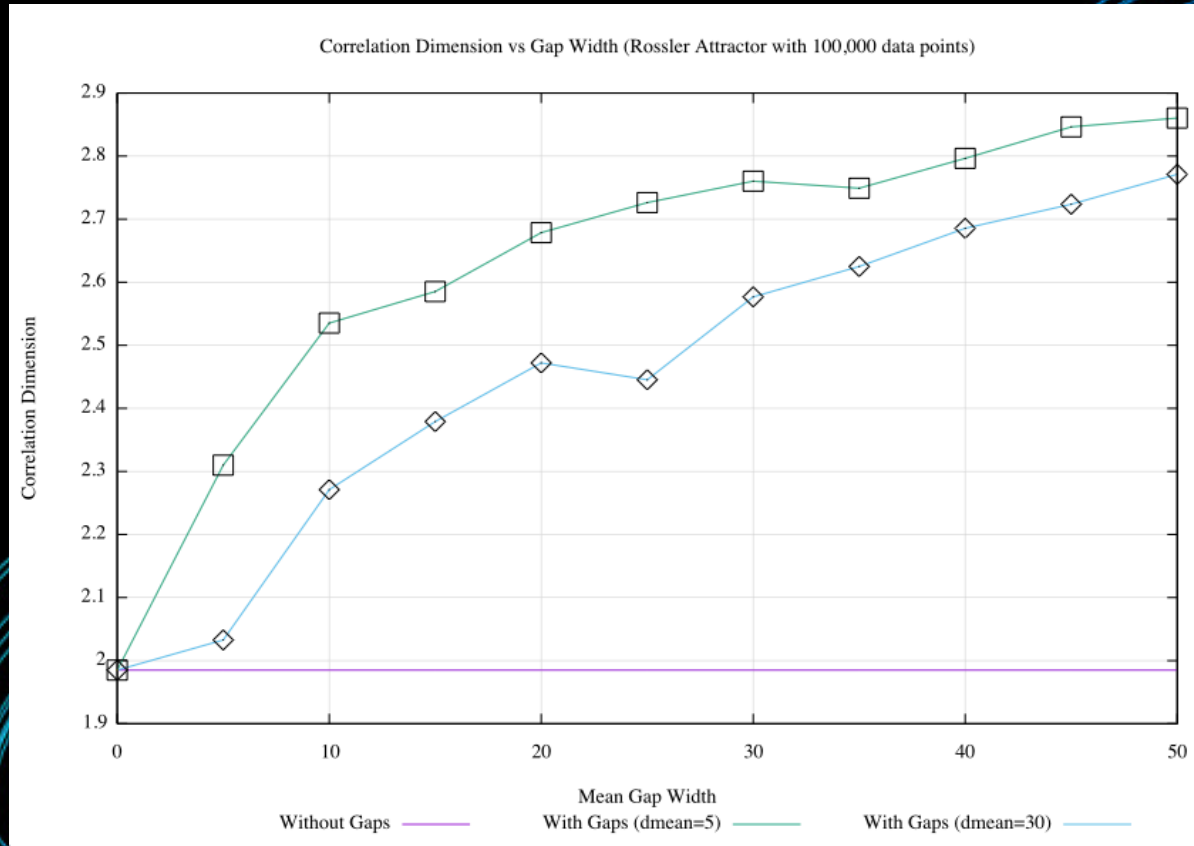
d-mean = 5, w-mean = 50

2. Fixed Number of Data Points (100,000)



In this case, again, correlation dimension tends to be overestimated for the same reason as case 1. However, in this case, there is no saturation of correlation dimension observed because attractors get fuzzier with constant number of points (almost covered fixed points).

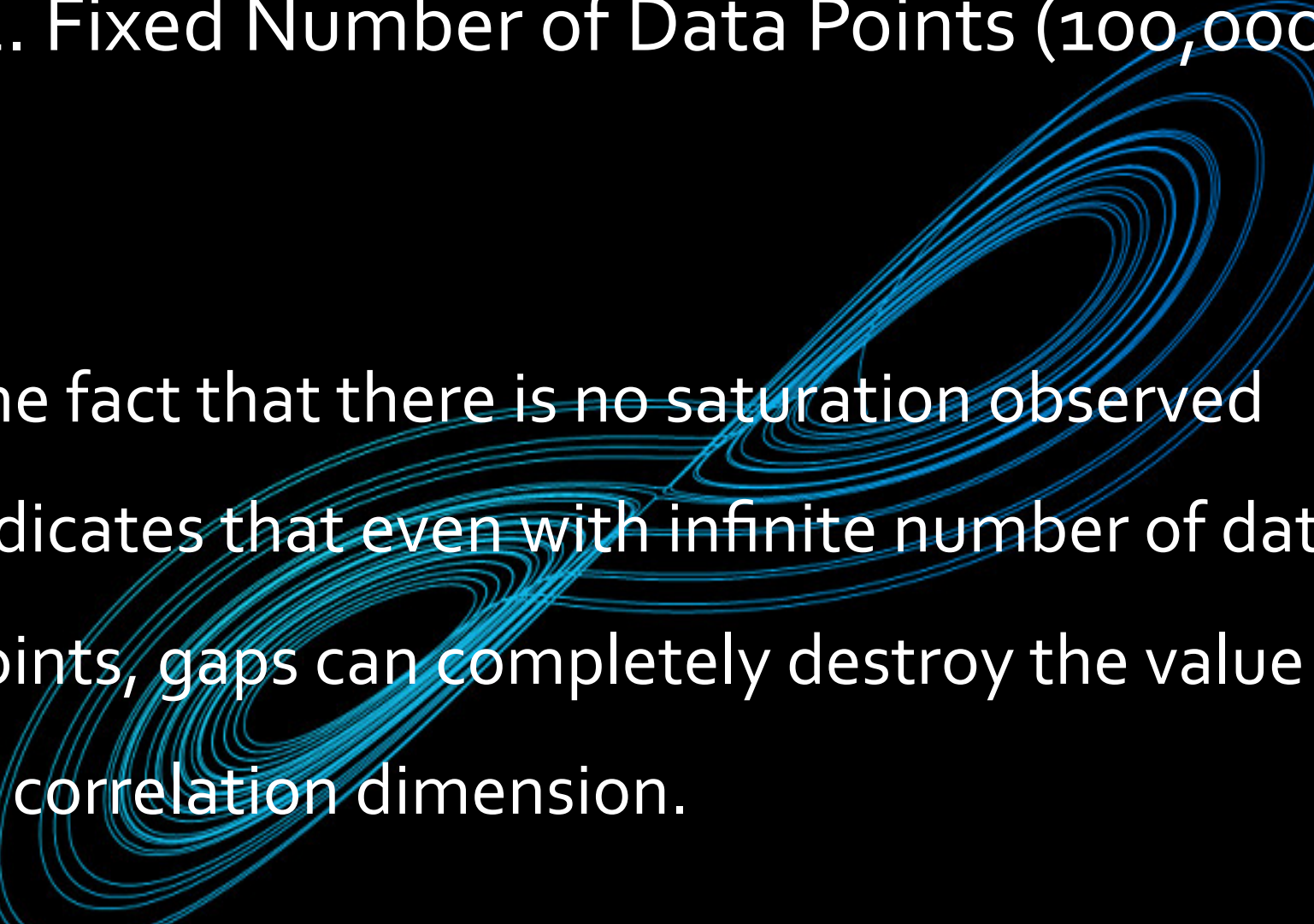
2. Fixed Number of Data Points (100,000)



In this case, again, correlation dimension tends to be overestimated for the same reason as case 1. However, in this case, there is no saturation of correlation dimension observed because attractors get fuzzier with constant number of points (almost covered fixed points).

2. Fixed Number of Data Points (100,000)

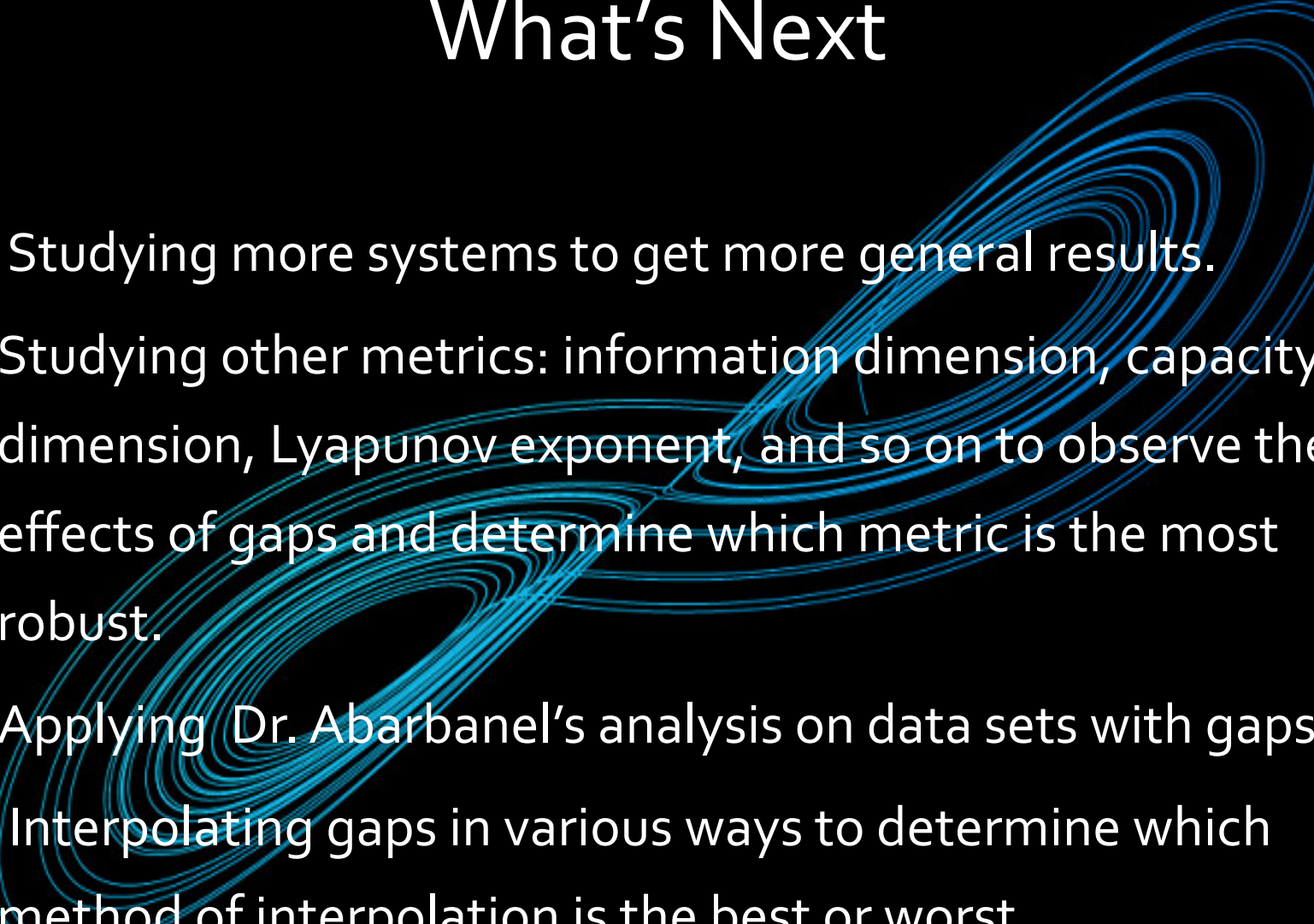
The fact that there is no saturation observed indicates that even with infinite number of data points, gaps can completely destroy the value of correlation dimension.



Summary

- For the delay time embedding, $\exp(-5t/2\tau)$ criteria can be used as a way of determining the value of τ .
- With gaps, the value of correlation dimension tends to be overestimated.
 - For the case with the fixed duration of sampling time, the saturation of correlation dimension was observed.
 - For the case with the fixed number of data points, no saturation of correlation dimension was observed, indicating that infinite number of data points does not help. Instead, it makes worse.

What's Next

- 
- A decorative graphic consisting of numerous concentric, slightly irregular blue oval lines. These lines are centered around the middle of the slide, overlapping the text of the second and third bullet points. They create a sense of depth and movement, resembling a stylized orbit or a series of ripples.
- Studying more systems to get more general results.
 - Studying other metrics: information dimension, capacity dimension, Lyapunov exponent, and so on to observe the effects of gaps and determine which metric is the most robust.
 - Applying Dr. Abarbanel's analysis on data sets with gaps.
 - Interpolating gaps in various ways to determine which method of interpolation is the best or worst.

References

- S. V. George, G. Ambika, & R. Misra, “Effect of uneven sampling on correlation dimension computed from time series data”, arXiv:1410.4454v1 [nlin.CD] (2014).
- S. V. George, G. Ambika, & R. Misra, “Effect of data gaps on correlation dimension computed from light curves of variable stars”, arXiv:1410.4454v2 [nlin.CD] (2015).