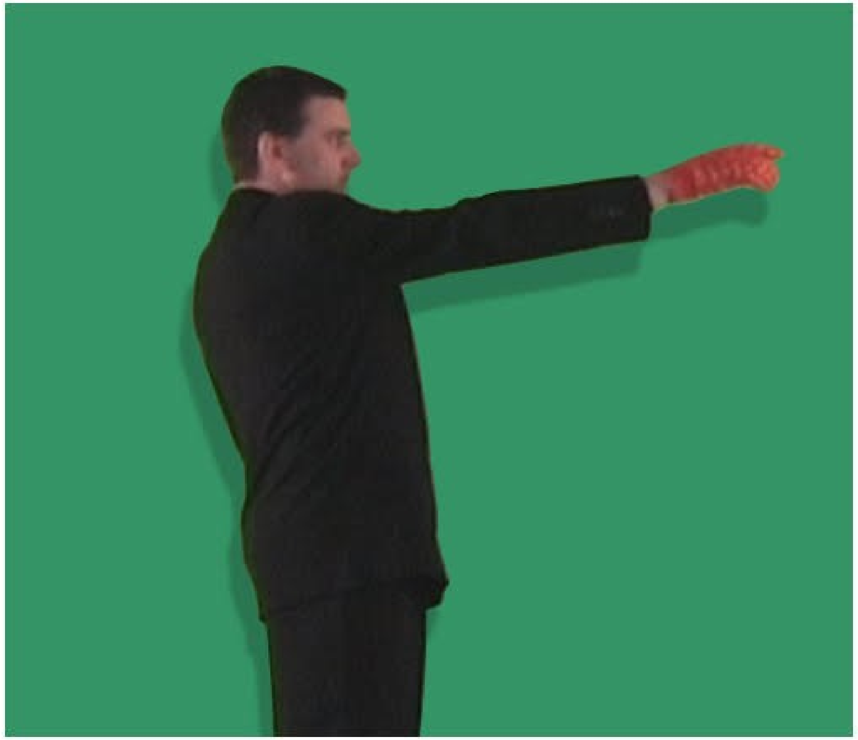



Time Series Data Mining: A Unifying View



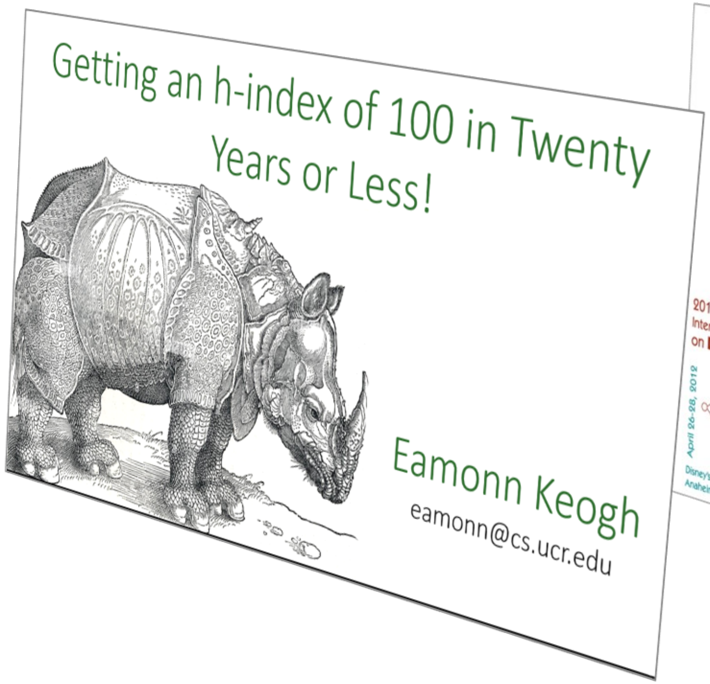

Eamonn Keogh

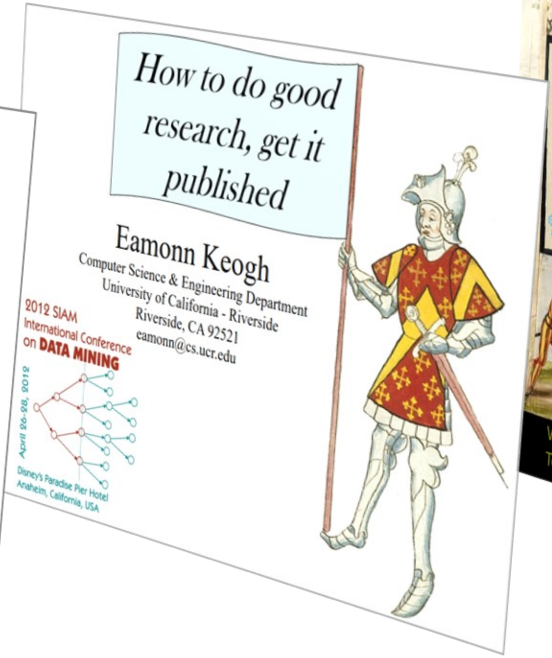
UCR

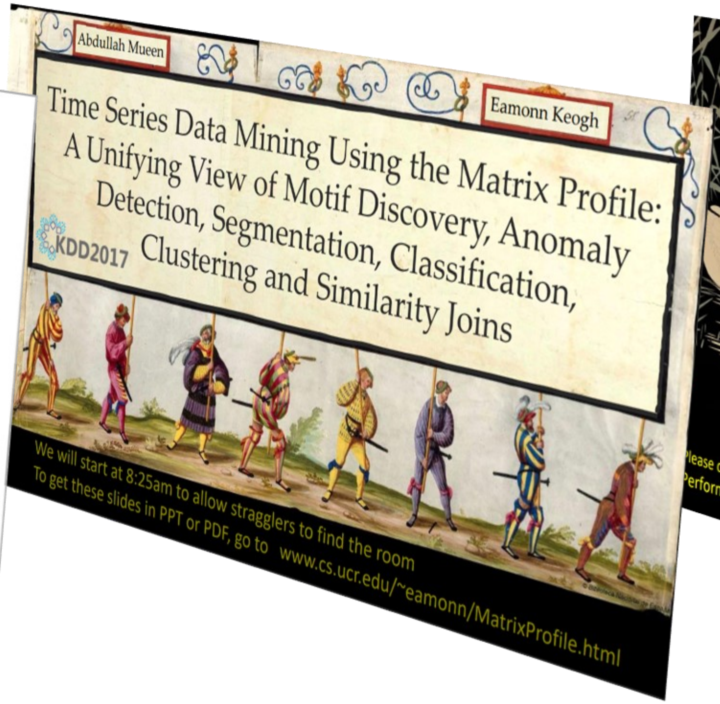
eamonn@cs.ucr.edu

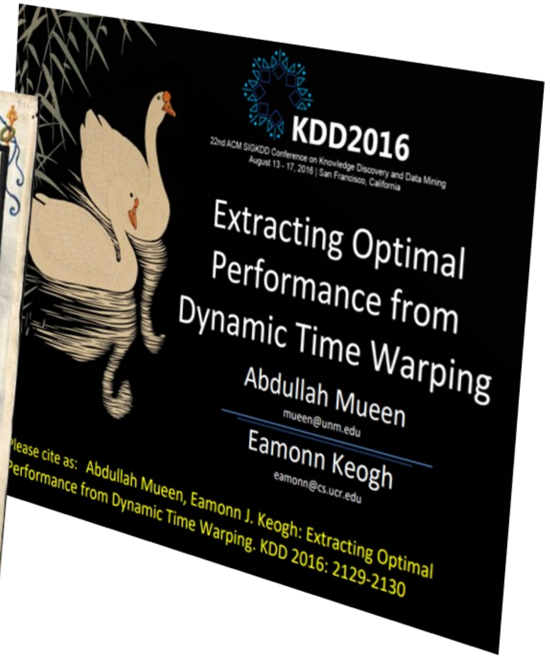
Welcome

- If you like this tutorial, you may enjoy my others ;-)









Note

- If you are reading a PDF, you are missing a handful of animations available in the PowerPoint version.

Disclaimer:

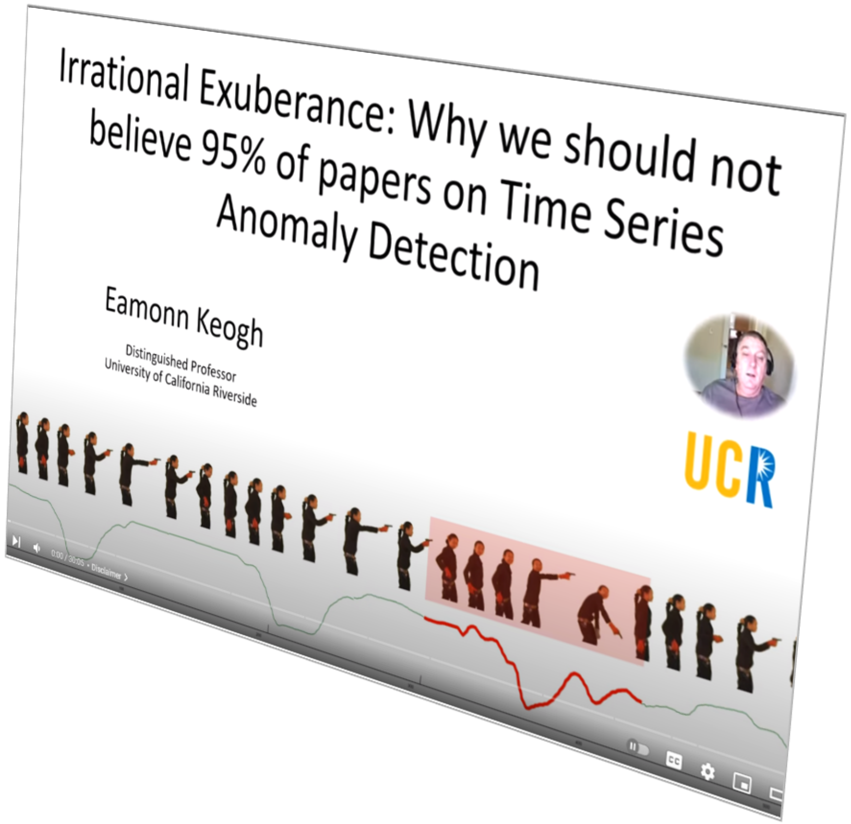
- This is not a disinterested survey of work in time series
- This pragmatic tutorial focus mostly on work from my lab, and from the lab of a former student, Mueen.
- However, I will point out other individuals/groups doing interesting and useful work.

Format

- This tutorial has little math or code.
- I want to give you the *intuition* behind the SOTA time series data mining.
- I will illustrate with short case studies.

Addressing the elephant in the room

- There is no deep learning in the tutorial!
- Contrary to 100's of paper's claims, there is little evidence that deep learning really helps for time series problems.
- I am not going to spend time explaining this, if you are interested, see this talk…



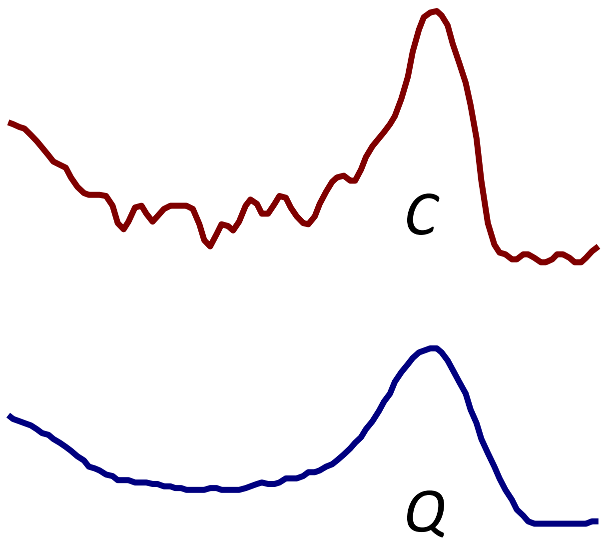
www.youtube.com/watch?v=Vg1p3DouX8w&ab_channel=EamonnKeogh

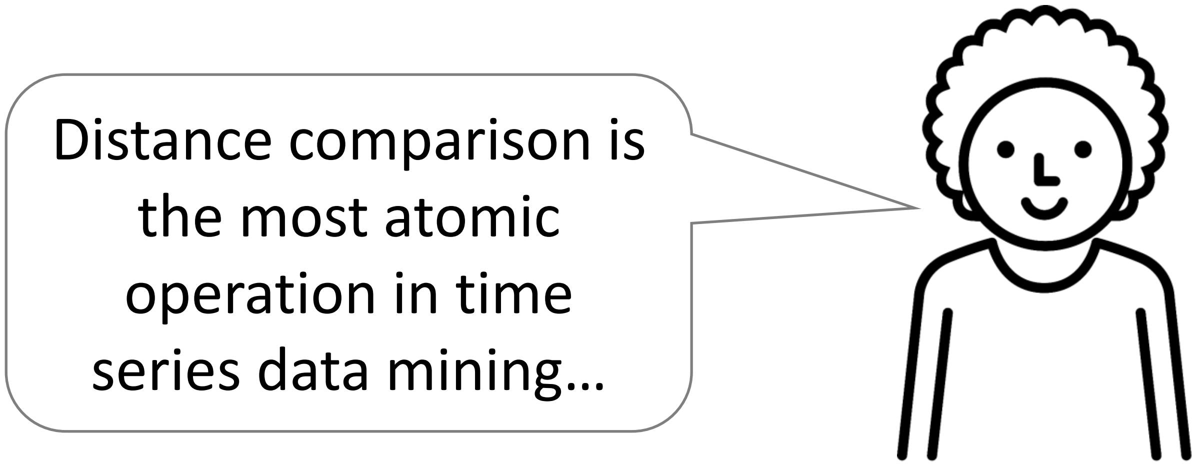
Preview

- What do you want to do with time series data?
  - *Everything!* Similarity search, classification, clustering, summarization, visualization, segmentation, anomaly discovery, repeated pattern discovery, rule discovery, emerging behavior discovery etc.
  - Amazingly, with just two very simple tools, MASS and the *Matrix Profile*, you can basically do all the above and more!

Comparing two time series

How similar are these two-time series?





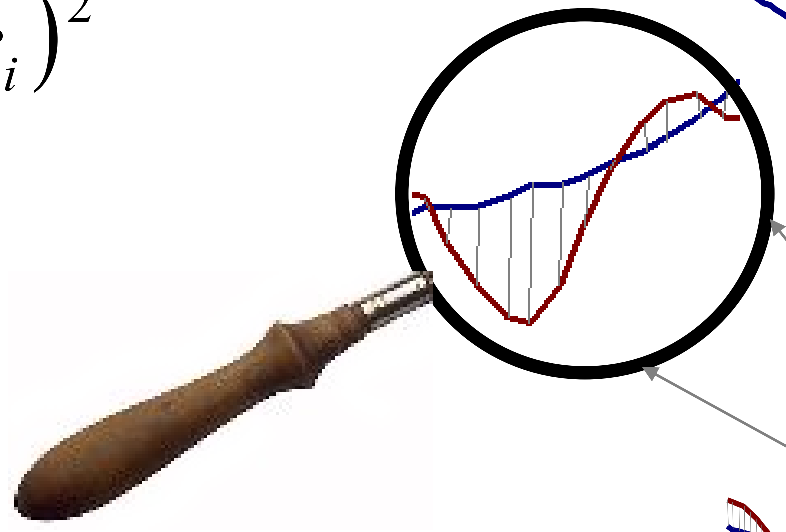
Euclidean Distance Metric (ED)

Given two time series:

$Q = q_1 \ldots q_n$

$C = c_1 \ldots c_n$

$$D(Q,C) \equiv \sqrt{\sum_{i=1}^{n} (q_i - c_i)^2}$$



We always z-normalize the data before computing the ED

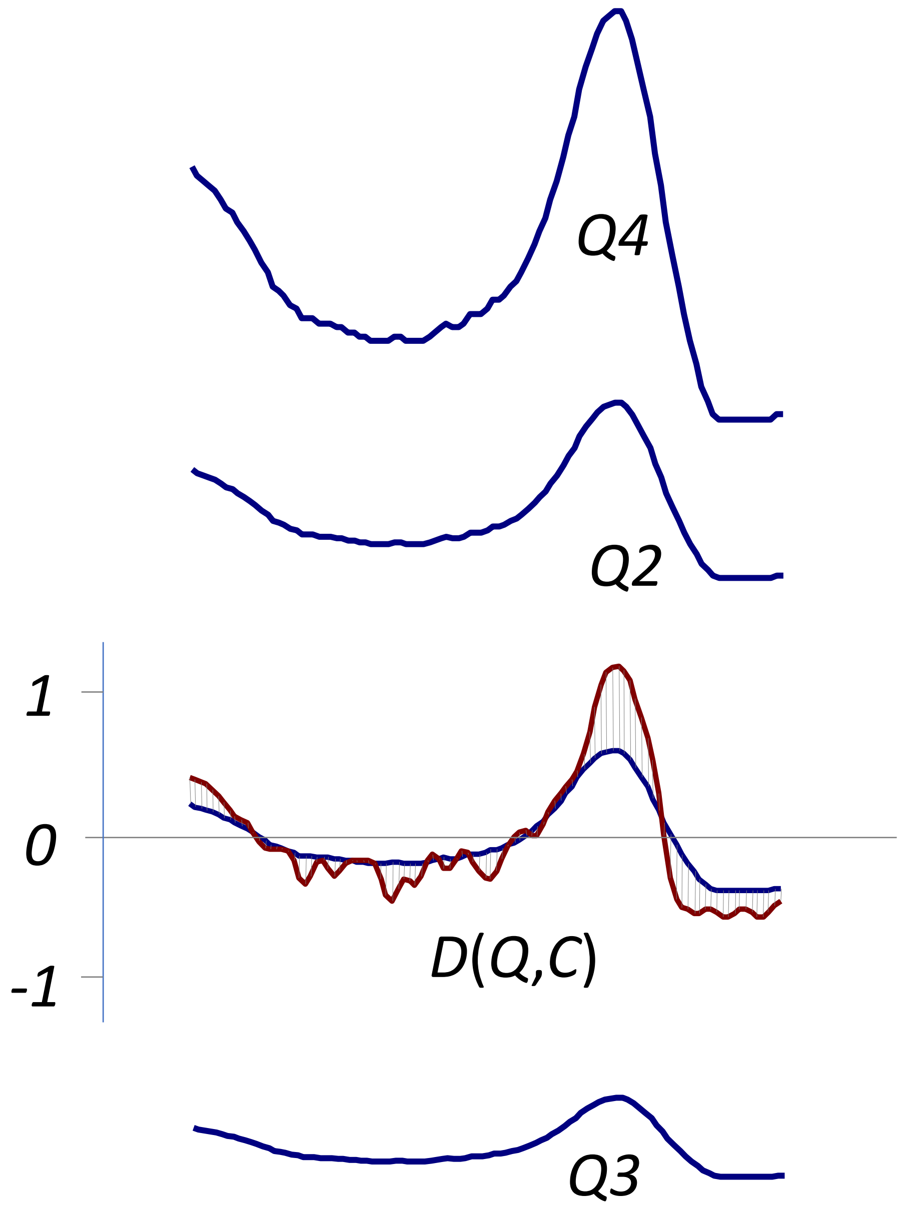
This is logically equivalent to Pearson's Correlation

Euclidean Distance Metric (ED)

Z-normalization means we are ignoring mean and standard deviation of the data.

At least 99.9% of the time, that is the right thing to do.

So here $D(Q,C) = D(Q2,C) = D(Q3,C) = D(Q4,C)$



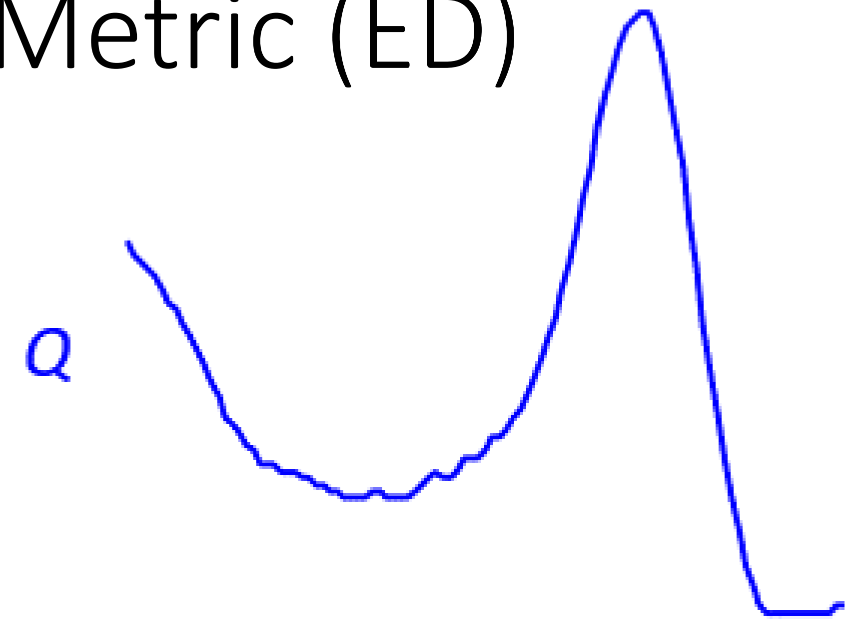
Euclidean Distance Metric (ED)

Z-normalization means we are ignoring the mean and standard deviation of the data.

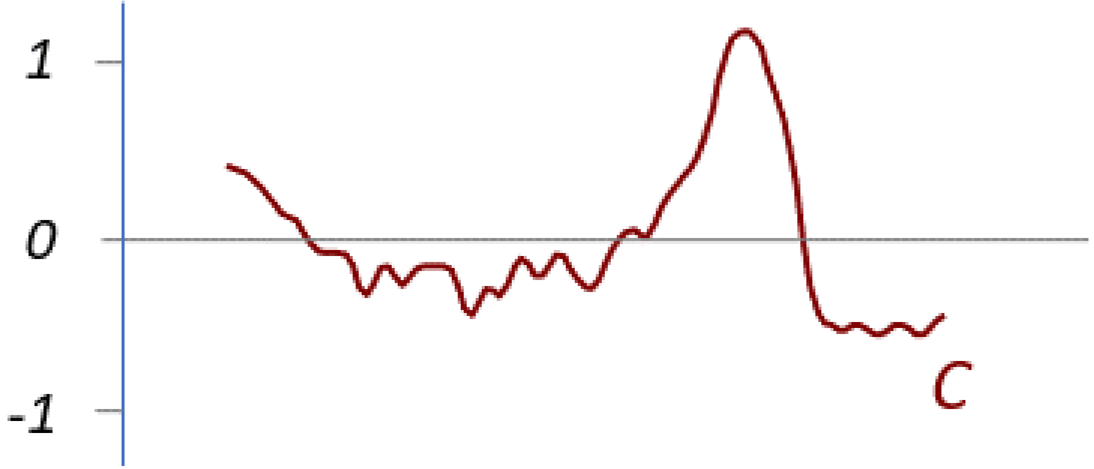
At least 99.9% of the time, that is the right thing to do.

In Matlab

```
>> Q = zscore(Q);
```
Or
```
>> Q = (Q-mean(Q))/std(Q);
```



D(Q,C)



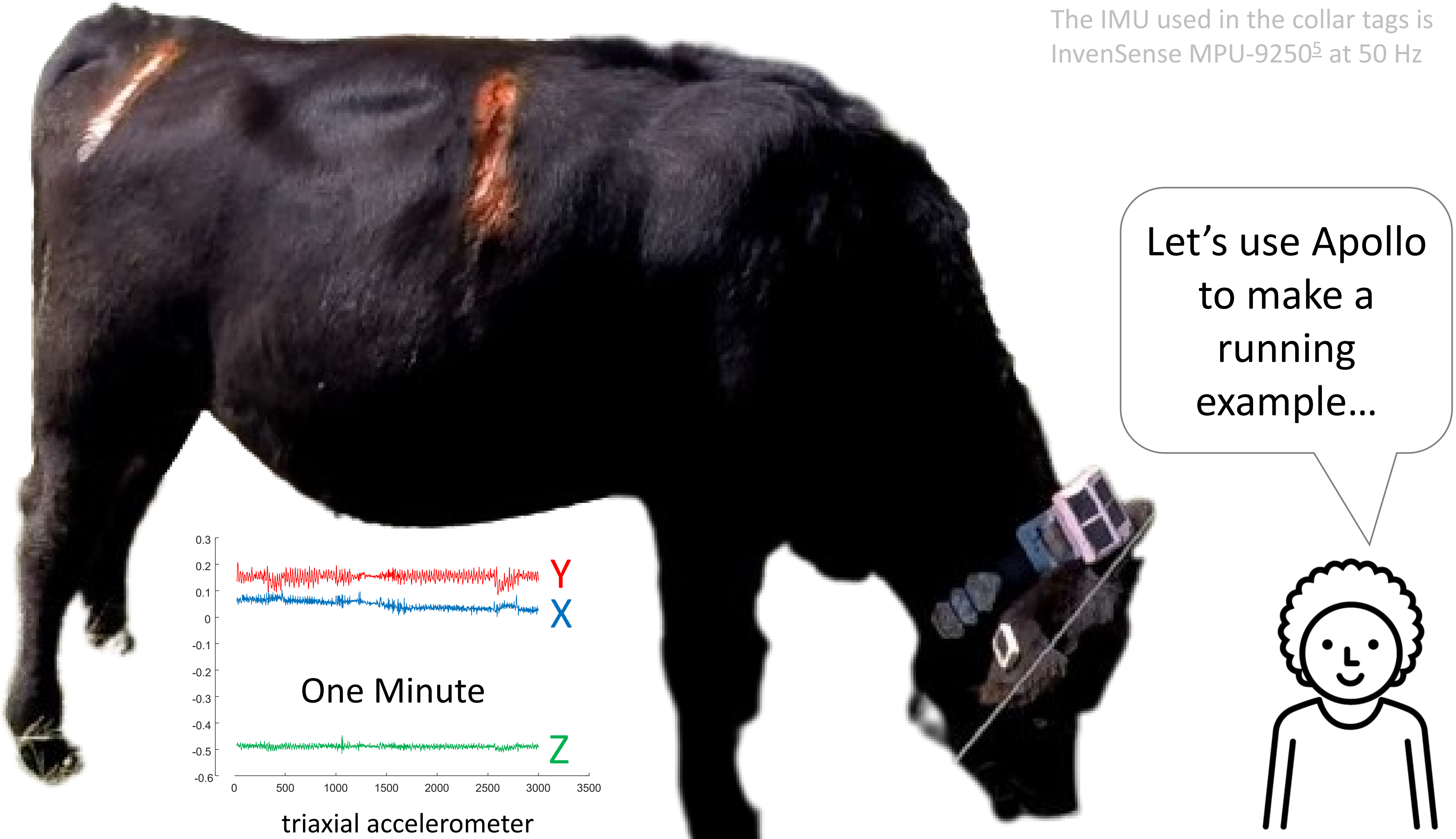
I have done " *A Large Comparison of Normalization Methods on Time Series*"

Vinícius de Souza

doi.org/10.1016/j.bdr.2023.100407

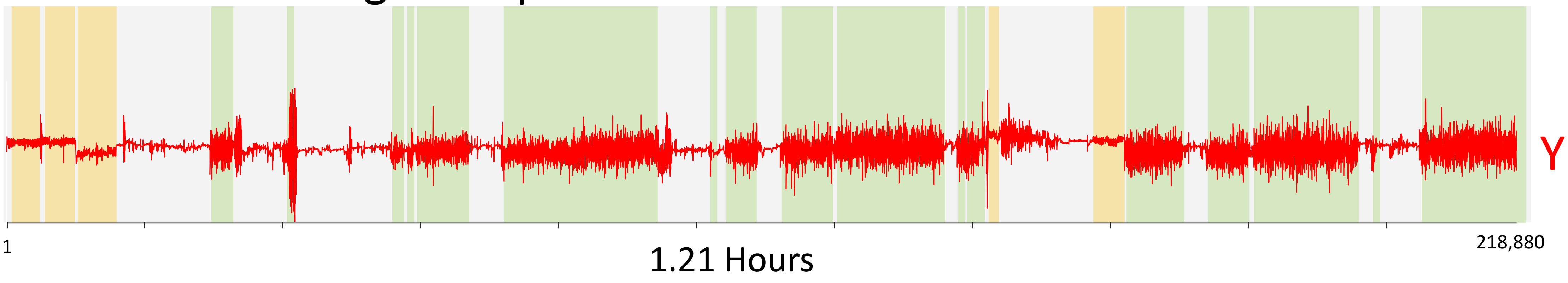

triaxial
accelerometer

The IMU used in the collar tags is InvenSense MPU-9250[5] at 50 Hz

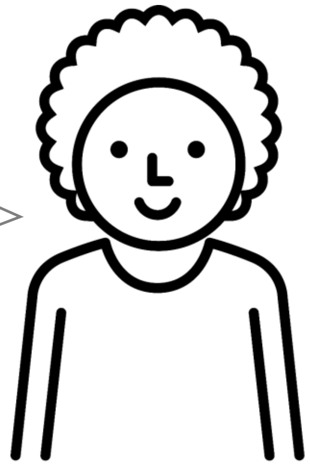


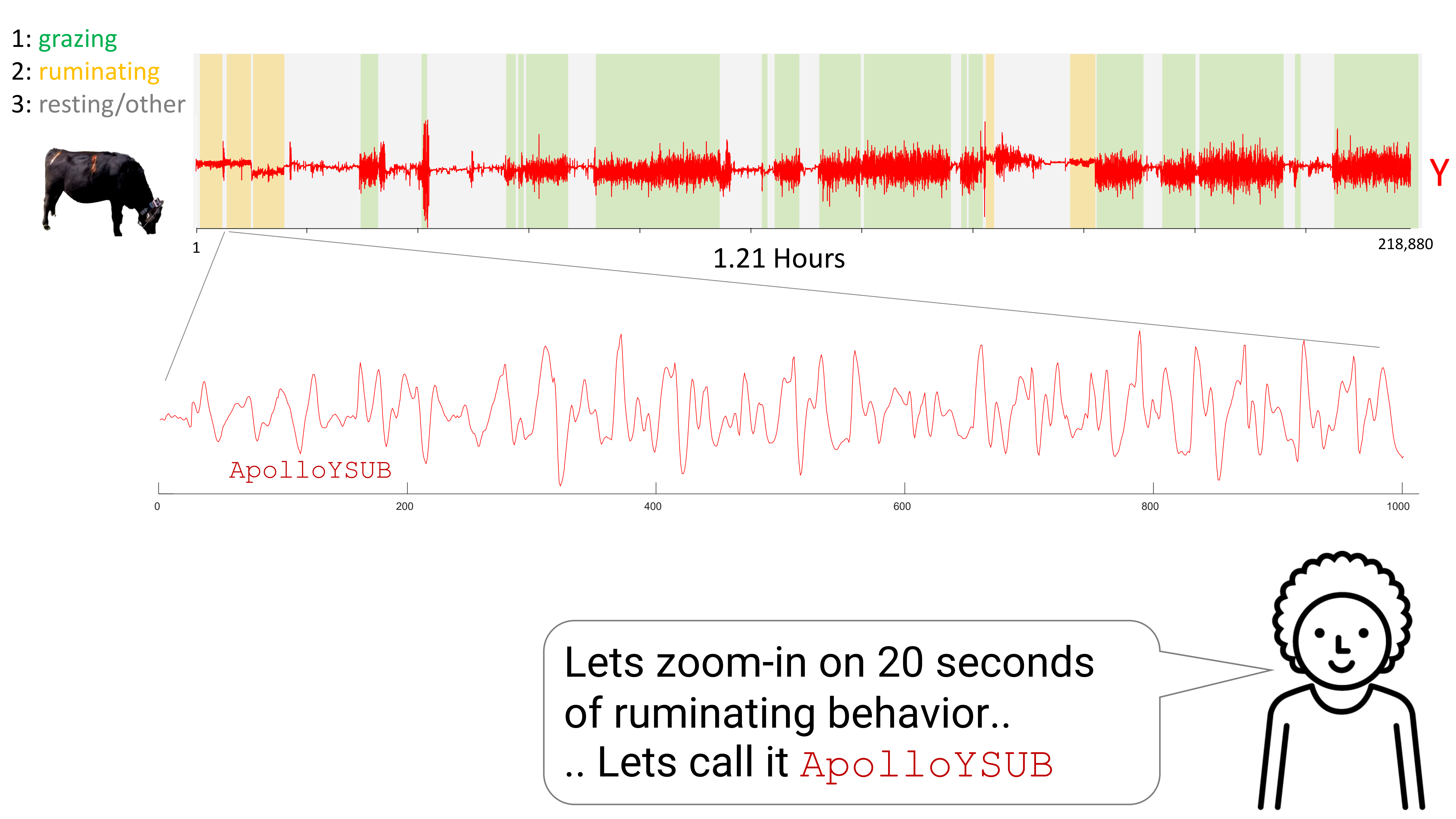
A Running Example

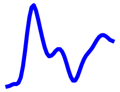
1: grazing
2: ruminating
3: resting/other



Here is about 72 minutes of Y-axis acceleration of Apollo. Someone labeled this data by looking at accompanying video

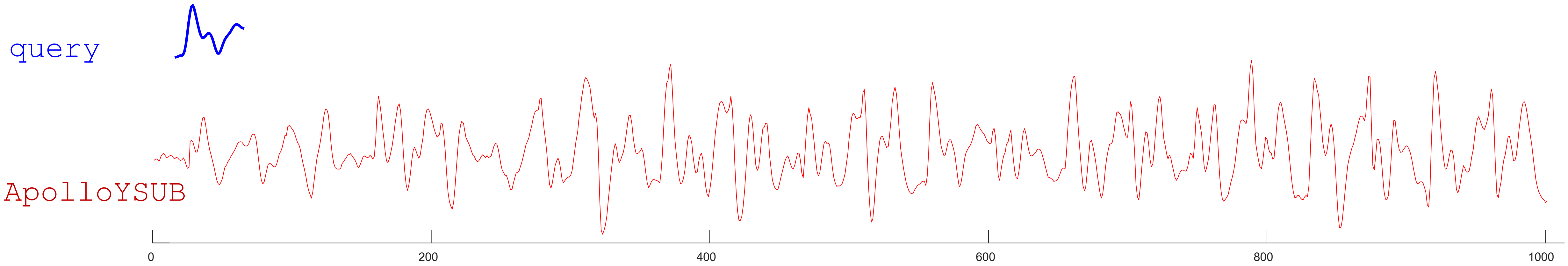


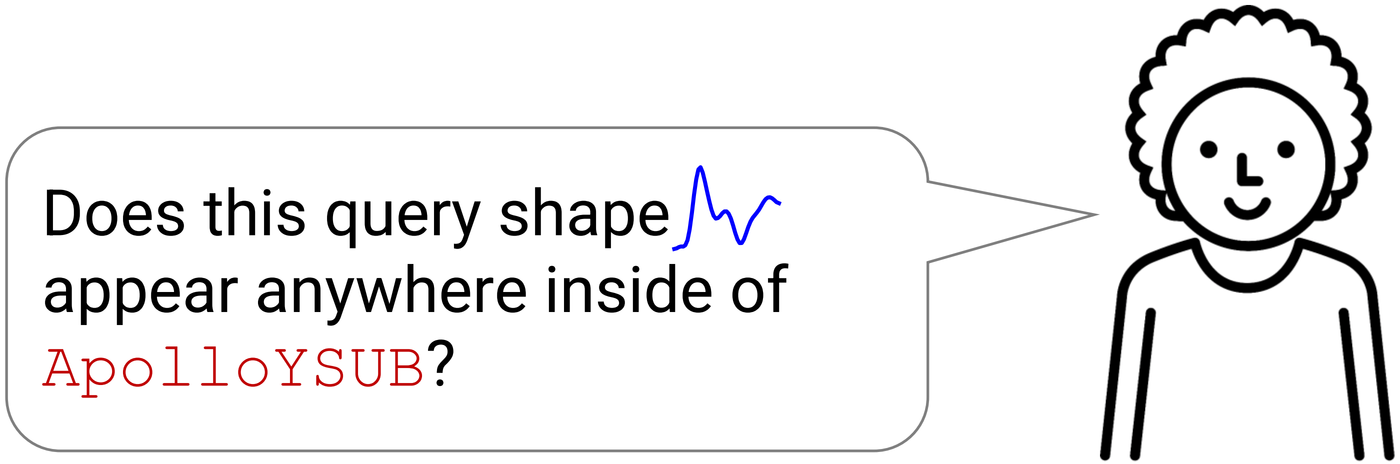

1: grazing
2: ruminating
3: resting/other
Y
1
218,880
1.21 Hours
ApolloYSUB
0
200
400
600
800
1000
Lets zoom-in on 20 seconds
of ruminating behavior..
.. Lets call it ApolloYSUB

I have this one-second-long behavior  ,I am going to call this query.

I have reason to think that it is indicative of Bovine spongiform encephalopathy (BSE)

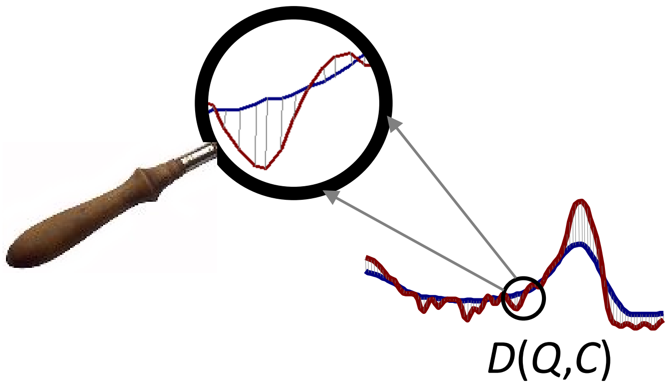
Does this query behavior exist in Apollo? To find out, we will build a *distance profile*.

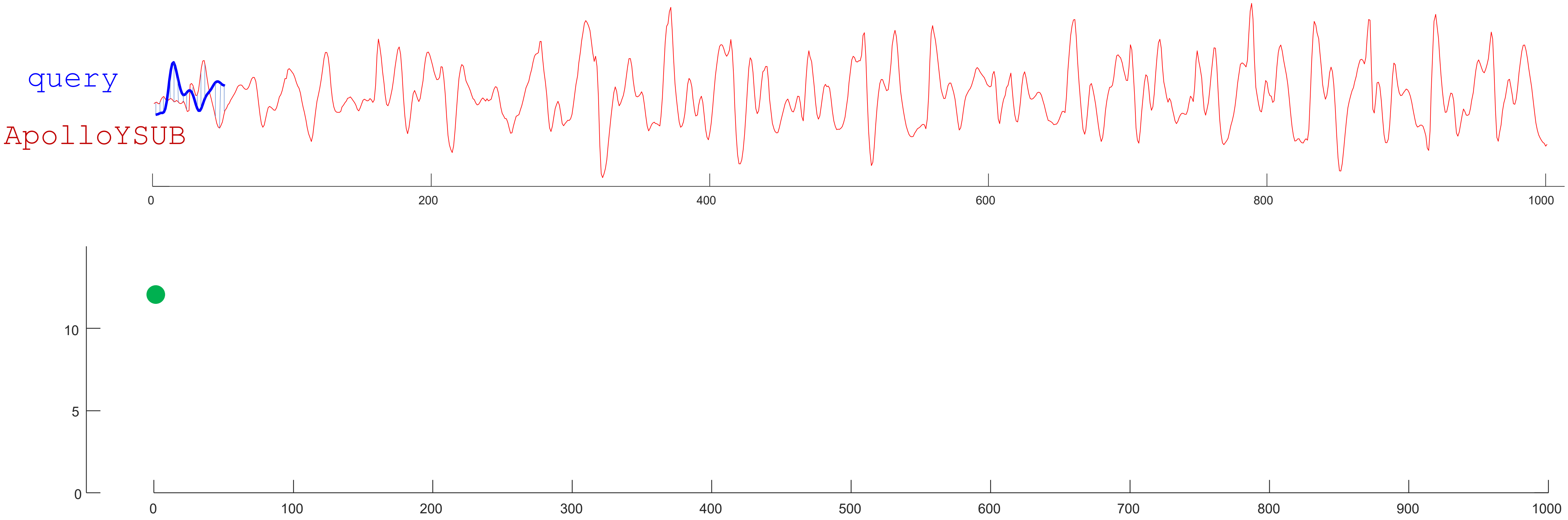




To find out, we will build a *distance profile*.

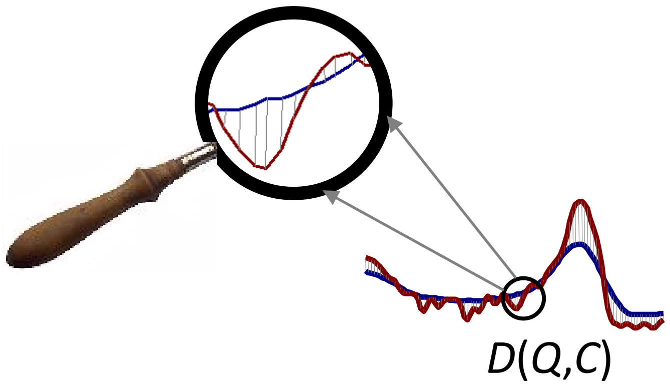
This is simply the z-normalized Euclidean distance between the query and every subsequence in the longer time series…

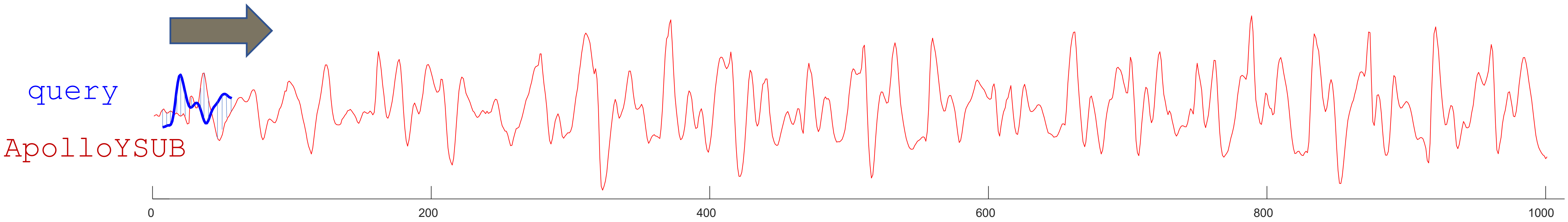


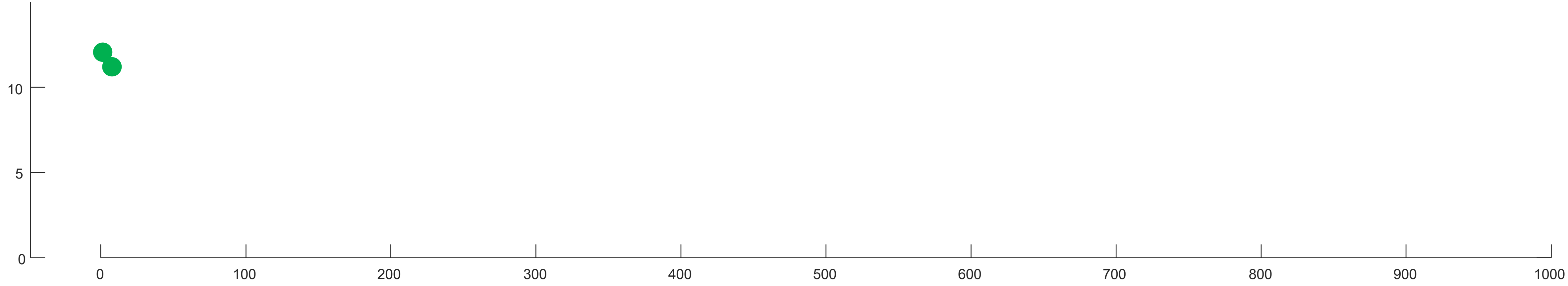


To find out, we will build a *distance profile*.

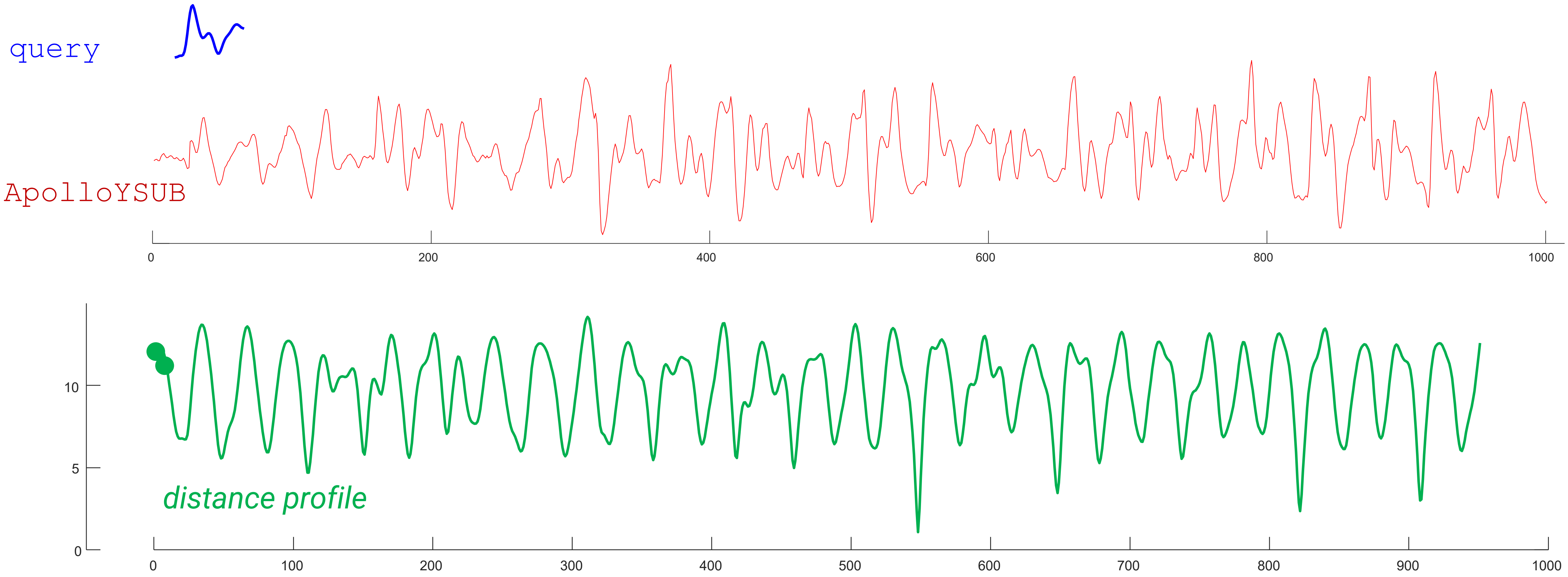
This is simply the z-normalized Euclidean distance between the query and every subsequence in the longer time series…

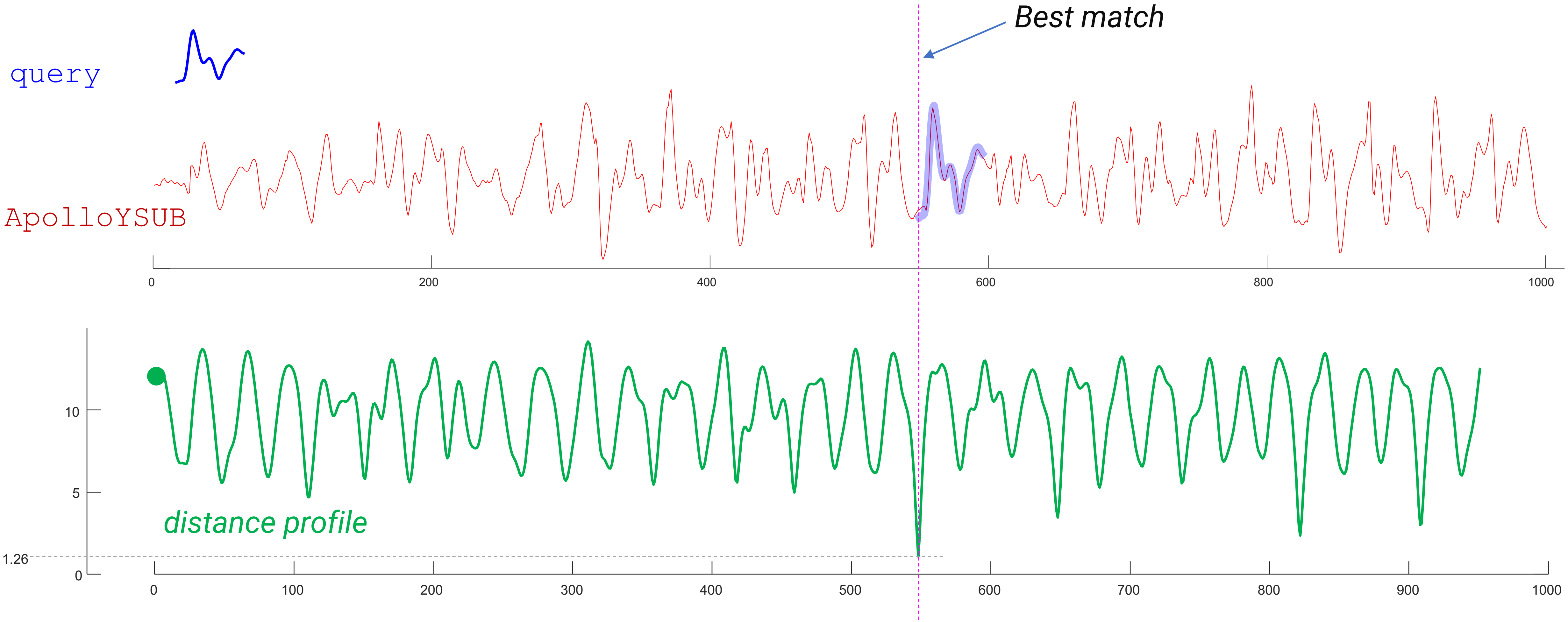






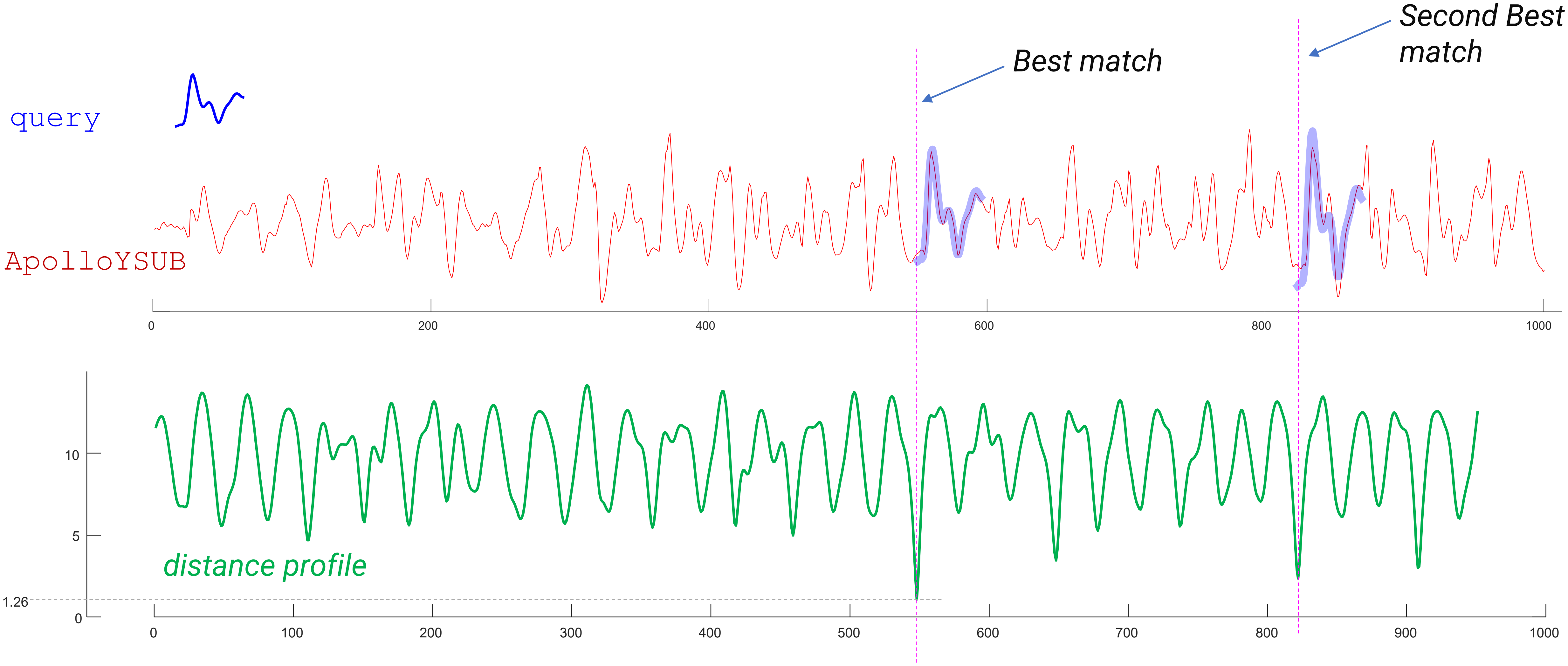
```
>> dist_profile = real(MASS(ApolloYSUB,query));
>> plot(dist_profile)
```

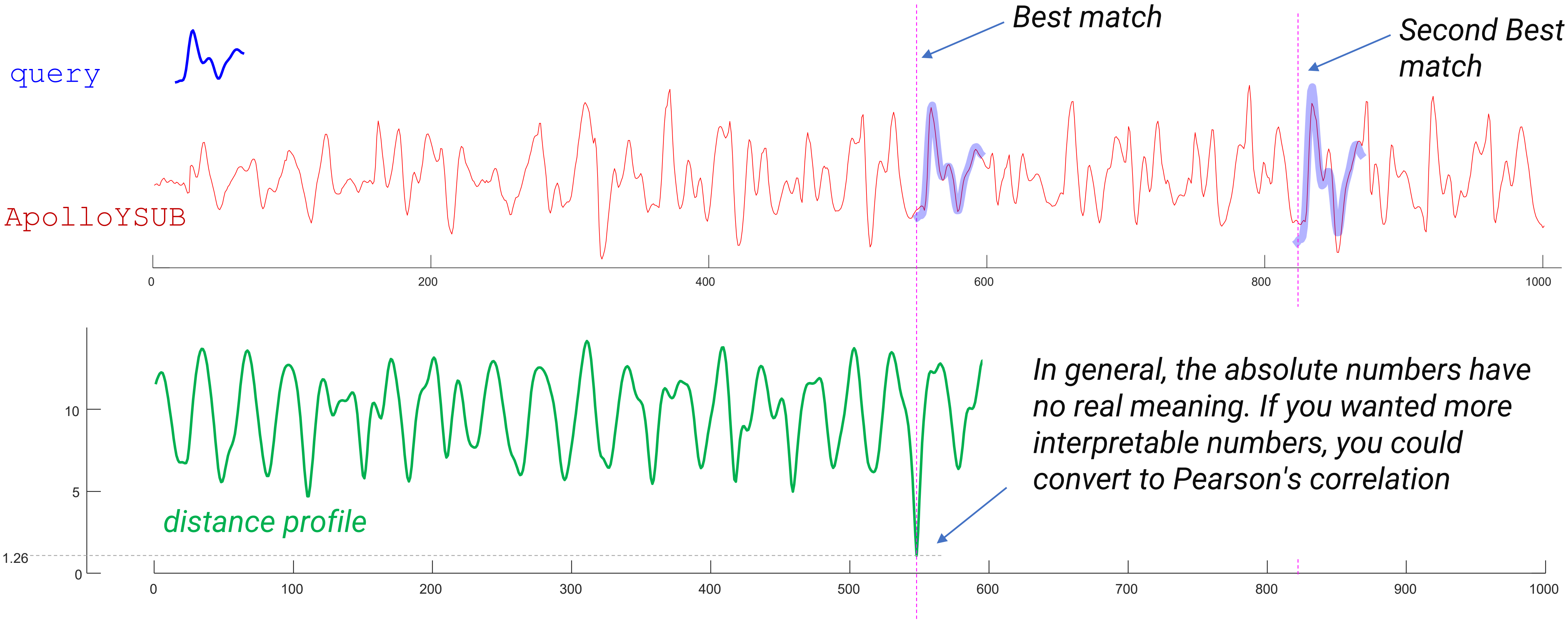




query
Best match
ApolloYSUB
0
200
400
600
800
1000
10
5
1.26
0
distance profile
0
100
200
300
400
500
600
700
800
900
1000

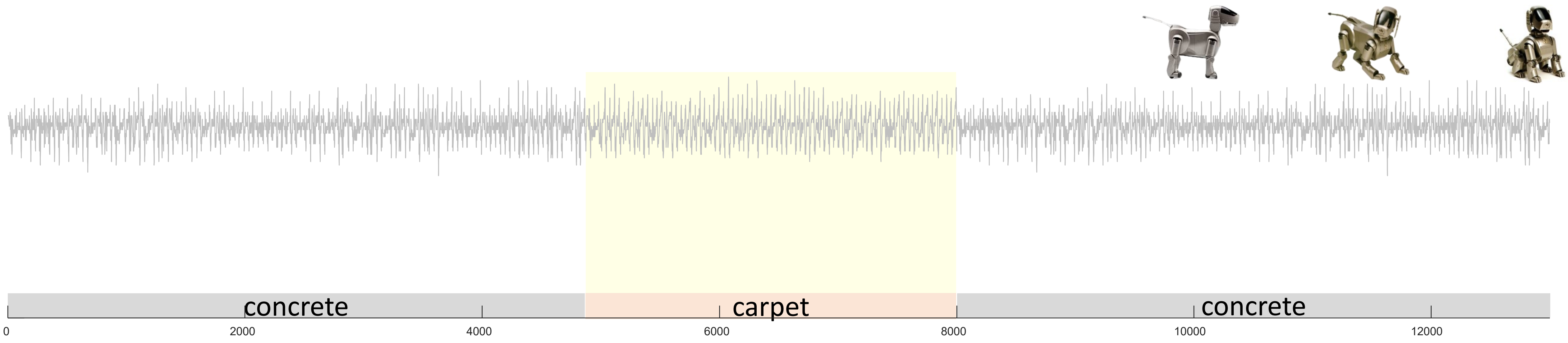
(define trivial match)

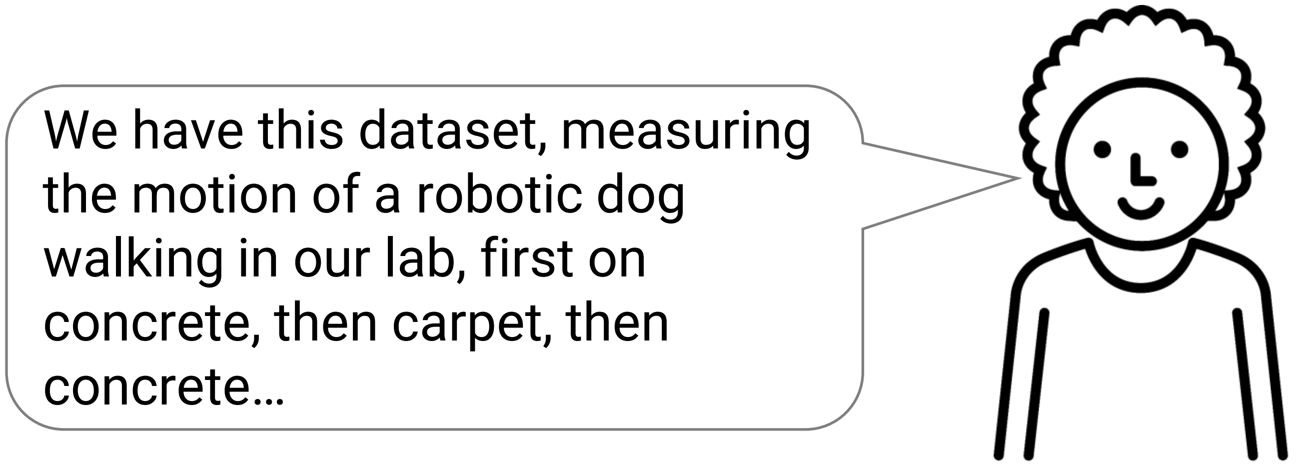




In general, the absolute numbers have no real meaning. If you wanted more interpretable numbers, you could convert to Pearson's correlation

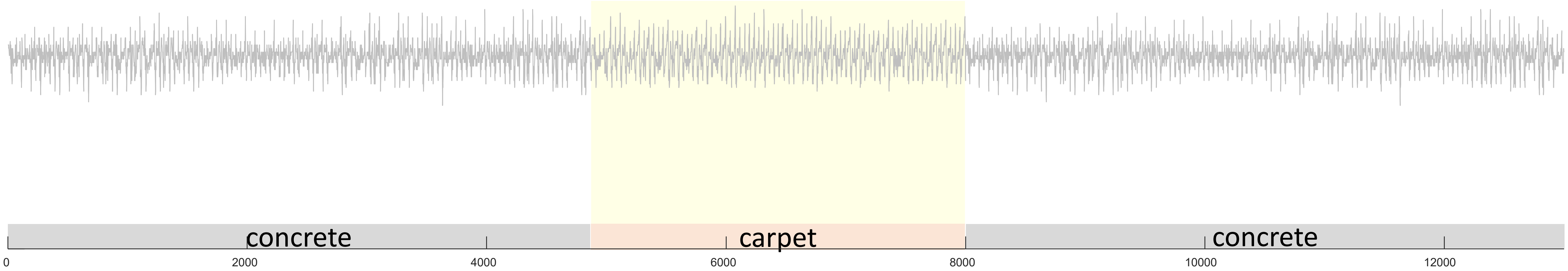
Lets do a full worked example





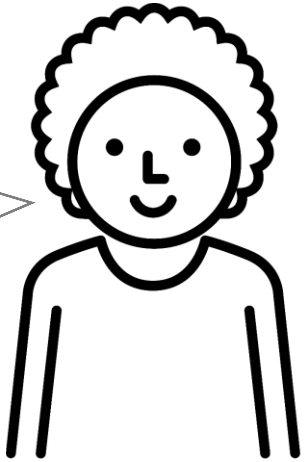
Lets do a full worked example

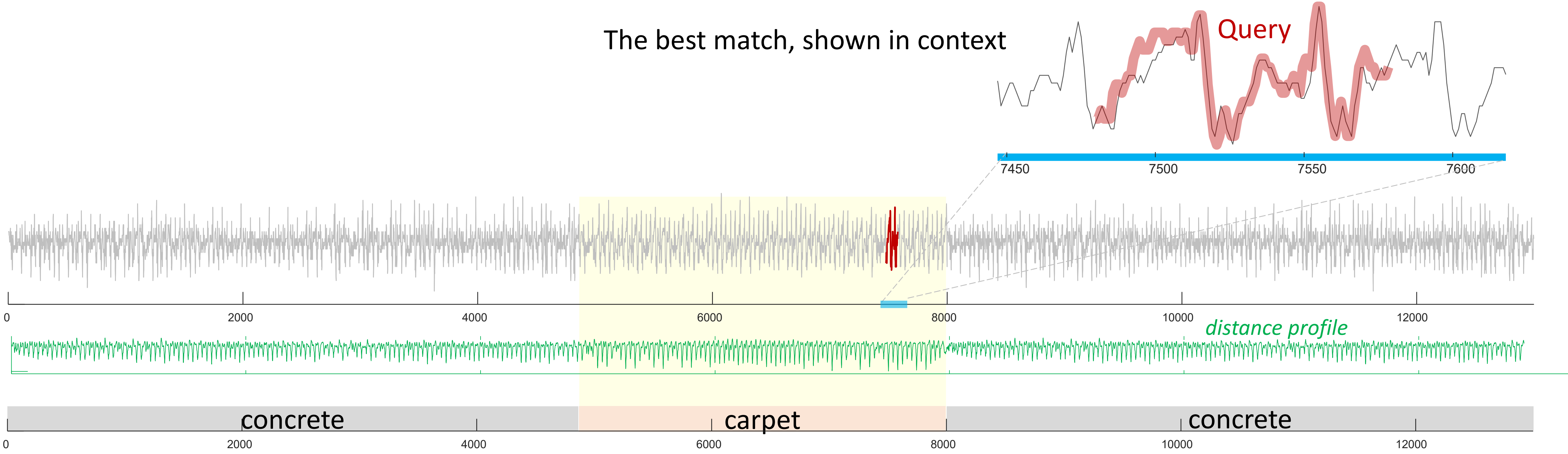
Not to scale with below



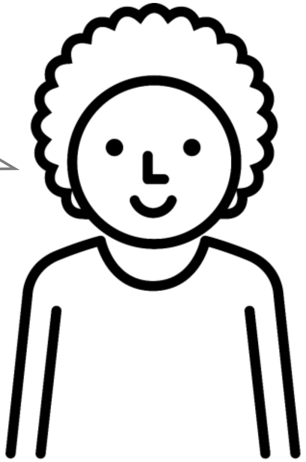
Do you think the query is from when the dog was walking on carpet or concrete or something else…

.. we also have a query, for the same dog walking, on a different day..





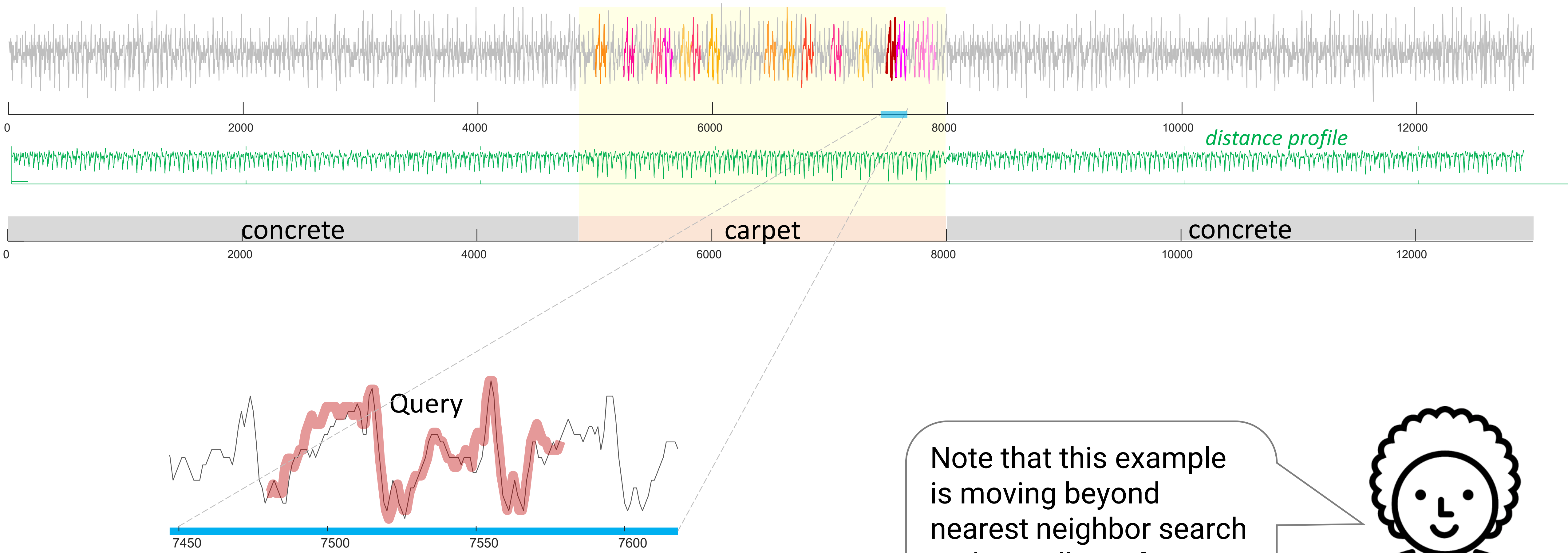
The best match to the query, is in the carpet walking session.
So, the query is probably also carpet-walking.



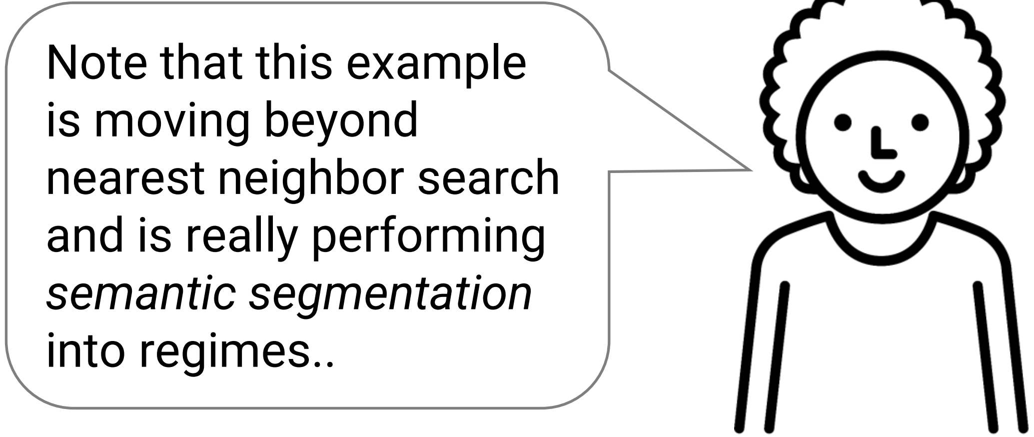
This task is trivial with MASS code…

```
>> dist = MASS(dog ,carpet_query );          % compute a distance profile
>> [val loc] = min(dist);                    % find location of match
>> disp(['The best matching subsequence starts at ',num2str(loc)])
The best matching subsequence starts at 7479
```

Below we plot the *16* best matches. Note that they all occur during the carpet walking period. This entire process takes about 1/10,000th of a second.



The best match, shown in context



Mini-Review I: The *distance profile* MASS

- This is a simple idea called: query-by-content/similarity-search/nearest-neighbor search etc.
- You can use the distance profile to find the K-nearest neighbors to any query.
- This is an *incredibly* powerful and useful tool, limited only by your imagination.
- It is by far, the most important subroutine in all of time series data mining.

--

- Suppose you have multidimensional data? You can just compute the individual distance profiles, sum them, then find the lowest values as the multidimensional nearest neighbor!
- I gave you a fully worked example in **Appendix Multidimensional MASS** (Penguin example)

--

- The computation of the distance profile is *incredibly* fast (details later)
- You can get MASS in most computer languages/platforms.

*https://www.cs.unm.edu/~mueen/FastestSimilaritySearch.html

Mini-Review II: The *distance profile*   MASS

Since we are at VLDB, where we like to index things…

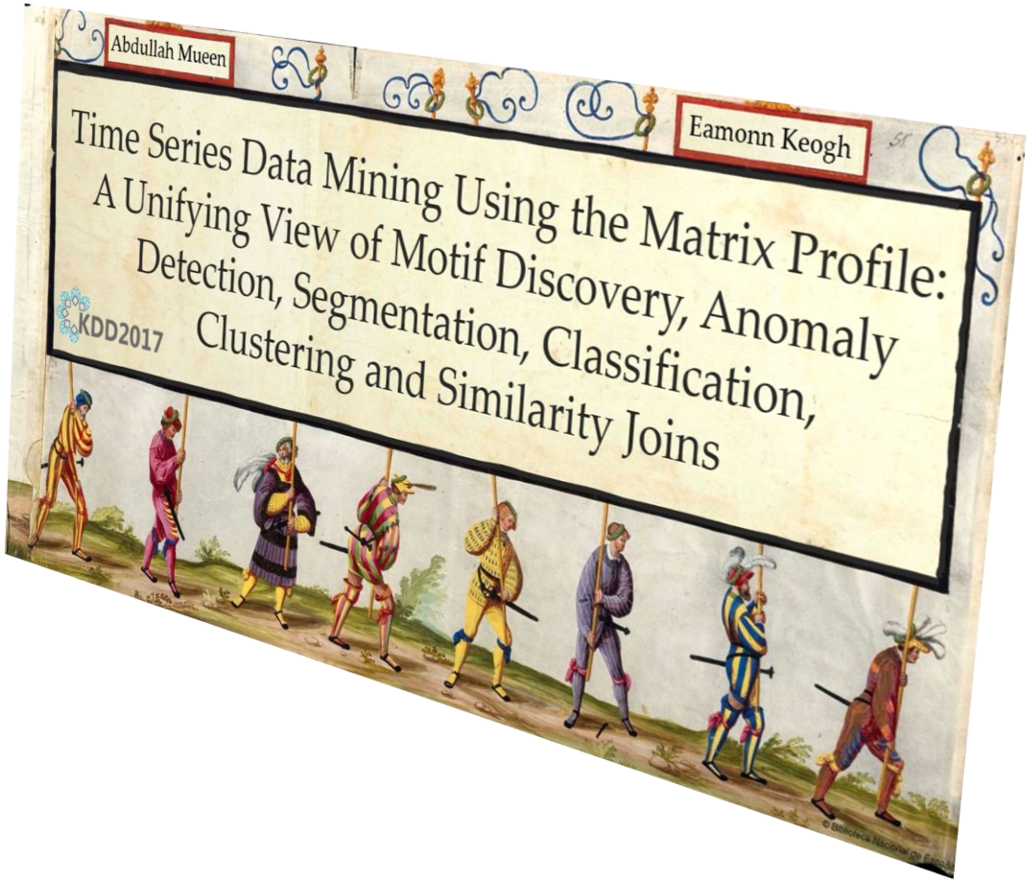
I believe that *Indexing of Time Series is not useful and not possible in the general sense.* See **Appendix: Why no indexing for Time Series**

Other advantages of MASS

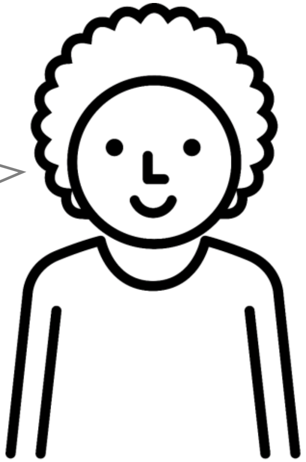
- For free, you get:
  - The top-*K* nearest neighbors, for any *K*.
  - Range queries, for any range *r*.
  - The top-*K* furthest neighbors, for any *K*.
- MASS supports GPUs, a billion datapoints in a second!
- MASS supports *weighted* queries.
- Predictable time. Almost any index will be fast for some queries, but very slow for others. But MASS takes the same time for *any* query.

Preview: The *Matrix Profile*

- We are about to learn about the Matrix Profile*
- It is a very simple idea.
- My claim is that once you have the Matrix Profile, almost all time series problems are trivial.
- Longer tutorials are online..



The Matrix Profile is the best idea in time series data mining in the last decade ;-)



*https://www.cs.ucr.edu/~eamonn/MatrixProfile.html

For the Cassini spacecraft in orbit around Saturn.. ..the best-performing method was the Matrix Profile.. Wagstaf et. al. NASA

(for an industrial IoT problem) *Matrix Profiles perform well with almost no parameterisation needed.* Anton et al ICDM 2018.

While there will never be a mathematical silver bullet, we have discovered that the Matrix Profile, a novel algorithm developed by the Keogh research group at UC-Riverside, is a powerful tool. Andrew Van Benschoten, lead engineer at Target.

If anybody has ever asked you to analyze time series data and to look for new insights then (the Matrix Profile) is definitely the open source tool that you'll want to add to your arsenal Sean Law, Ameritrade.

(for) intrusion detection in industrial network traffic, theMatrix Profiles rises significantly during the attacks. ..as a result, time series-based anomaly detection methods are capable of detecting deviations and anomalies. Schotten (2019).

The MatrixProfile technique is the state-of-the-art anomaly detection technique for continuous time series. Bart Goethals et. al. (ECML-PKDD 2019).

Based on the concept of Matrix Profile ..without relying on time series synchronization.. the Railway Technologies Laboratory of Virginia Tech has been developing an automated onboard data analysis for the maintenance track system Ahmadian et. al. JRC2019

Matrix Profile is the state-of-the-art similarity-based outlier detection method. Christian Jensen et. al. IJCAI-19

we use the exact method based on the Matrix Profile (to assess the effectiveness of therapy) Funkner et al Procedia 2019.

Recently, a research group from UCR have proposed a powerful tool - the Matrix Profile (MP) as a primitive...(we use it for) *fault detection* Jing Zhang et al. ICPHM 2019

Inspecting both graphs one can see that the matrix-profile algorithm was able to identify regions where there is a change on the power level over the observed band. F Lobao 2019.

RAMP builds upon an existing time series data analysis technique called Matrix Profile to detect anomalous distances...collected from scientific workflows in an online manner. Herath et. al. IEEE Big Data 2019

Based on obtained results for the considered data set, matrix profiles turned out to be most suitable for the task of anomaly detection Lohfink et al. VISSEC2019

The computation speed and exactness of the Matrix Profile make it a powerful tool and (our) results back this. Barry & Crane AICS 2019

(examining) *manufacturing batches considering raw amperage* (we found that the) *Matrix Profile highlights anomalies* Hillion & O'Connell of TIBCO Data Science. re:Invent 2019.

we use the exact method based on the matrix profile to search for motifs can be used to monitor the patient's condition, to assess the effectiveness of therapy or to assess the physician's actions. Funknera et al.

(The Matrix Profile is a) *similarity join to measure the similarity between two given sequences. we opt for the median of the profile array as the representative distance* (3D Dancing Move Synthesis from Music)" Anh et al. IEEE Robotics and Automation Letters

We were amazed by the power of MP and seek to incorporate it into our framework Ye and Ageno.

..adopting the concept of (the) Matrix Profile, we conduct the first attempt to.. J. Zuo et. al. Big Data 20019

The accuracies obtained ...indicate that the Matrix Profile is useful for the task at hand instead of using the CNN features directly Dhruv Batheja

To speed up online bad PMU data detection a fast discovery strategy is introduced based on (the Matrix Profile) Zhu and Hill.

Specifically, ALDI uses the matrix profile method to quantify the similarities of daily subsequences in time series meter data, Zoltan Nagy, Energy & Buildings (2020)

Our two-fold approach first leverages the Matrix Profile technique for time series data mining.. Nichiforov 2020.

*the class of matrix profile algorithms.. ..is a promising approach, as it allows simplified post-processing and analysis steps by examining the resulting matrix profile structure*A. Raoofy et al.

We only require information about the time of several critical incidents to train our methods, as previously. To this end, we employ the Matrix Profle.. Bellas. et al.

a matrix-profile based algorithm applied across all trajectory data against a validation set revealed four significant motifs which we defined as motif A, B, C and D.. Fernandez Alvarez 2020.

The main building block of this (game analytics) algorithm is the matrix profile, Saadat and Sukthankar AAAI2020

We leverage the Matrix Profile (MP), ... to create a micro-service-based machinery monitoring solution Naskos et al 2021

SLMAD uses statistical-learning and employs a robust box-plot algorithm and Matrix Profile (MP) to detect anomalies Team from Huawei/UCD.

We found that all these similarity or randomness measures can be estimated with variants of the highly efficient Matrix Profile (MP) algorithm.

it is important to be able to effectively extract the sales correlation between products from retail datasets. In this work, we use the fast matrix profile (MP) to discover Li et al Journal of Retailing and Consumer Services. 2023

(after apply the Matrix Profile to three diverse domains) The matrix profile has proved to be a useful tool in terms of making most time series data mining tasks intuitive and to require less effort compared to other methods.. Hehir and Smeaton 2023.

this study demonstrates that matrix profile data can enhance attention-based models.. Butler and Crane. 2023

One of the core concepts behind (our work) LS-USS is the matrix profile. Strommen et al. 2023.

(Finding) repetitions in ship roll time simulation data becomes rapidly very demanding computationally. For this reason, it is necessary to employ some optimization of the searching algorithm (of the Matrix Profile) Tsoumpelis and Spyrou 2023

We introduce an anomaly detection paradigm for healthcare. The proposed paradigm inherits the features from two state-of-the-art algorithms: STAMP and Scalable STOMP. Razaque 2022

applying the Matrix Profile technique and the pattern (motif) recognition methods we also observe the practicality and effectiveness of these methods for performance anomaly detection. Wallace, Ombuki-Berman and Ezzati-Jivan (2023)

for anomaly detection of axial piston pumps the Matrix Profile is chosen Dong et. al. IEEE Transactions on Instrumentation and Measurement. 2023

(we observe) matrix profile based methods perform well on different datasets... He et al. 2023

Using the matrix profile to compare subsequences of the time series to the most recent pattern, I discovered that the level of the late 1990s was the most similar to the current situation. This is interesting as there were similar extreme flooding events during this period. Paul dos Santos (2023)

The underlying principles behind the Matrix Profile are delightfully simple yet powerful. Markus Heidt (2022).

The Matrix profile approach long serves as a fundamental data mining method to investigate time series, ...is widely accepted in the data science community and has been successfully applied to various application domains.. Ju et al. IPDPS 2022.

For our purposes, the information 'source runs stably or not' can be extracted by observing the statistical properties of the matrix-profile data-array Geithner et al. IPAC2021

(for) intrusion detection in industrial network traffic, distances as calculated with Matrix Profiles rises significantly during the attacks. ..as a result, time series-based anomaly detection methods are capable of detecting deviations and anomalies. Schotten (2019).

The results in the Matric Profile provide insights into the stability of the patterns presented in the players. Guglielmo et. al.

Based on the concept of Matrix Profile ..without relying on time series synchronization.. the Railway Technologies Laboratory of Virginia Tech has been developing an automated onboard data analysis for the maintenance track system Ahmadian et. al. JRC2019

Matrix Profile is the state-of-the-art similarity-based outlier detection method. Christian Jensen et. al. IJCAI-19

we use the exact method based on the Matrix Profile (to assess the effectiveness of therapy) Funkner et al Procedia 2019.

Recently, a research group from UCR have proposed a powerful tool - the Matrix Profile (MP) as a primitive...(we use it for) *fault detection* Jing Zhang et al. ICPHM 2019

Inspecting both graphs one can see that the matrix-profile algorithm was able to identify regions where there is a change on the power level over the observed band. F Lobao 2019.

RAMP builds upon an existing time series data analysis technique called Matrix Profile to detect anomalous distances...collected from scientific workflows in an online manner. Herath et. al. IEEE Big Data 2019

Based on obtained results for the considered data set, matrix profiles turned out to be most suitable for the task of anomaly detection Lohfink et al. VISSEC2019

The computation speed and exactness of the Matrix Profile make it a powerful tool and (our) results back this. Barry & Crane AICS 2019

(examining) *manufacturing batches considering raw amperage* (we found that the) *Matrix Profile highlights anomalies* Hillion & O'Connell of TIBCO Data Science. re:Invent 2019. [video]

we use the exact method based on the matrix profile to search for motifs (that) *can be used to monitor the patient's condition, to assess the effectiveness of therapy or to assess the physician's actions*. Funknera et al. YSCCS 2019

(The Matrix Profile is a) *similarity join to measure the similarity between two given sequences. we opt for the median of the profile array as the representative distance* (3D Dancing Move Synthesis from Music)" Anh et al. IEEE Robotics and Automation Letters (video)

We were amazed by the power of MP and seek to incorporate it into our framework Ye and Ageno.

..adopting the concept of (the) Matrix Profile, we conduct the first attempt to.. J. Zuo et. al. Big Data 20019

The accuracies obtained ...indicate that the Matrix Profile is useful for the task at hand instead of using the CNN features directly Dhruv Batheja

To speed up online bad PMU data detection a fast discovery strategy is introduced based on (the Matrix Profile) Zhu and Hill.

Specifically, ALDI uses the matrix profile method to quantify the similarities of daily subsequences in time series meter data, Zoltan Nagy, Energy & Buildings (2020)

Our two-fold approach first leverages the Matrix Profile technique for time series data mining.. Nichiforov 2020.

*the class of matrix profile algorithms.. ..is a promising approach, as it allows simplified post-processing and analysis steps by examining the resulting matrix profile structure*A. Raoofy et al.

We only require information about the time of several critical incidents to train our methods, as previously. To this end, we employ the Matrix Profle.. Bellas. et al.

a matrix-profile based algorithm applied across all trajectory data against a validation set revealed four significant motifs which we defined as motif A, B, C and D.. Fernandez Alvarez 2020.

The main building block of this (game analytics) algorithm is the matrix profile, Saadat and Sukthankar AAAI2020

We call the custom Matrix Profile and box-plot approach as STUMPI-BP and employ it for all types of time series that.. Huawei Ireland Research Centre

The state-of-the-art exact anytime method for motif and discord discovery is matrix profile. Onur Mutlu et al.

Matrix Profile requires less training time, data and parameter tuning compared to other data mining methods.. Stamatescu et al.

(our) approach is based on matrix profile, which is a method for time series analysis that is robust, domain-agnostic, and computationally efficient. Van Hoecke et al 2020

(we conducted) five cases, which demonstrate how the matrix profile based pattern recognition approach can be used in the power boiler environment. In general (this) resulted in promising outcomes. Liisa Nokelainen 2020.

To investigate the physiological reaction, we used Matrix Profile.. Lalanne et al. 2020

Our time series mining uses the Matrix Profile approach.. Poerwawinata 2020

To this end, by adopting the concept Shapelet and Matrix Profile, we conduct the first attempt to.. Zeitouni et al.

Three anomalies were found near bridges and were indicated as low priority track issues. Two anomalies were... The Matrix Profiler found all these anomalies Steenwinckel et al 2020

Matrix Profile is a dimension reduction approach which requires less training time, data and parameter tuning compared to other data mining methods. Nichiforov et al.

Our solution will rely on Matrix Profile to allocate these discords in order to learn the anomalies. Alshaer et al.

The state-of-the-art exact anytime method for motif and discord discovery is matrix profile. Fernandez et. al.

Out of these methods, matrix profile is chosen in this paper due to its scalability and computational efficiency, while being also robust and parameter-free. Herrera et al 2021

The Matrix Profile (MP) algorithm has the potential to revolutionise many areas of data analysis.. Cartwright et al Engineering Proceedings

..we detect and locate the appearance of repeating patterns based on the Matrix Profile.. Li et al. IEEE MASS

We leverage the Matrix Profile (MP) to create a micro-service-based machinery monitoring solution Naskos et al 2021

SLMAD uses statistical-learning and employs a robust box-plot algorithm and Matrix Profile (MP) to detect anomalies. Team from Huawei/UCD.

We found that all these similarity or randomness measures can be estimated with variants of the highly efficient Matrix Profile (MP) algorithm. Austin Health/ University of Melbourne

We present, to the best of our knowledge, the first FPGA implementation of a Matrix Profile algorithm. Future Generation Computer Systems 2021

Out of these methods, matrix profile is chosen in this article due to its scalability and computational efficiency, while being also robust doing so increased accuracy and computational efficiency for detecting where and when anomalies and pattern irregularities occur. PARLIKAD et al 2021

Our approach exploits the synergism of the matrix profile data analysis technique. Nichiforov et al. Energies 2021, 14, 7465.

Our version of time series analysis is based on Matrix Profile. Onur Mutlu et al. 2022

(to diagnose) Decubitus Prophylaxis we use the Matrix Profile. Richter et. al. PerCom 2022

We successfully detected a deauthentication attack on a smart farm architecture using MERLIN, a (discord) distance-based anomaly detection algorithm. Sina Sontowski 2022

we detect and locate the appearance of repeating patterns based on the Matrix Profile. Lit et al. 2021 IEEE Mobile Ad Hoc and Smart Systems.

Another use case for the metamodel is the processing of more than 200,000 time series with a total of more than 30 million individual data points. ..for detecting anomalies and finding regime change locations, Matrix Profile algorithms are used.. Latapie et al. Front. Psychol., 20 May 2022

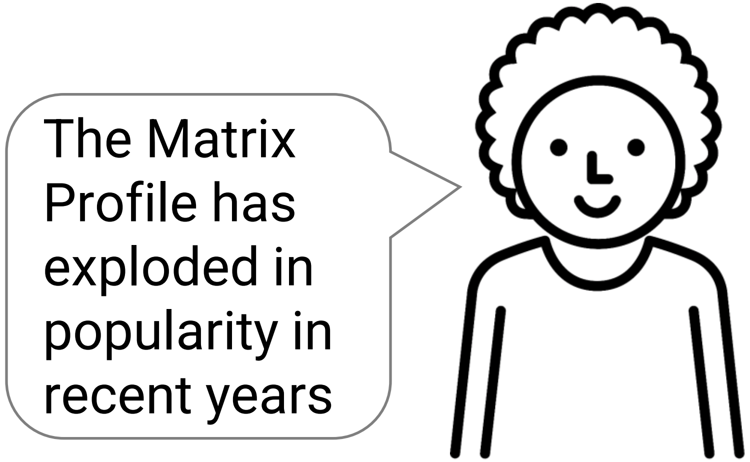
For detecting Decubitus Prophylaxis we use the Matrix Profile. Grunerbel et. al. 2022 IEEE Pervasive Computing and Communications.

As the most complex and largest dataset, the BETH dataset presents challenges to any anomaly detection technique. The matrix profile anomaly detector presented here performed well, not missing any of the evil outliers. Alexander Beattie 2022

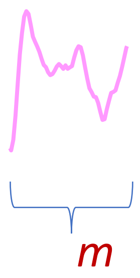
..our new recovery method that uses the matrix profile to extract the top-k repeated patterns from sensor nodes within the same group and.. Berjab et al 2022

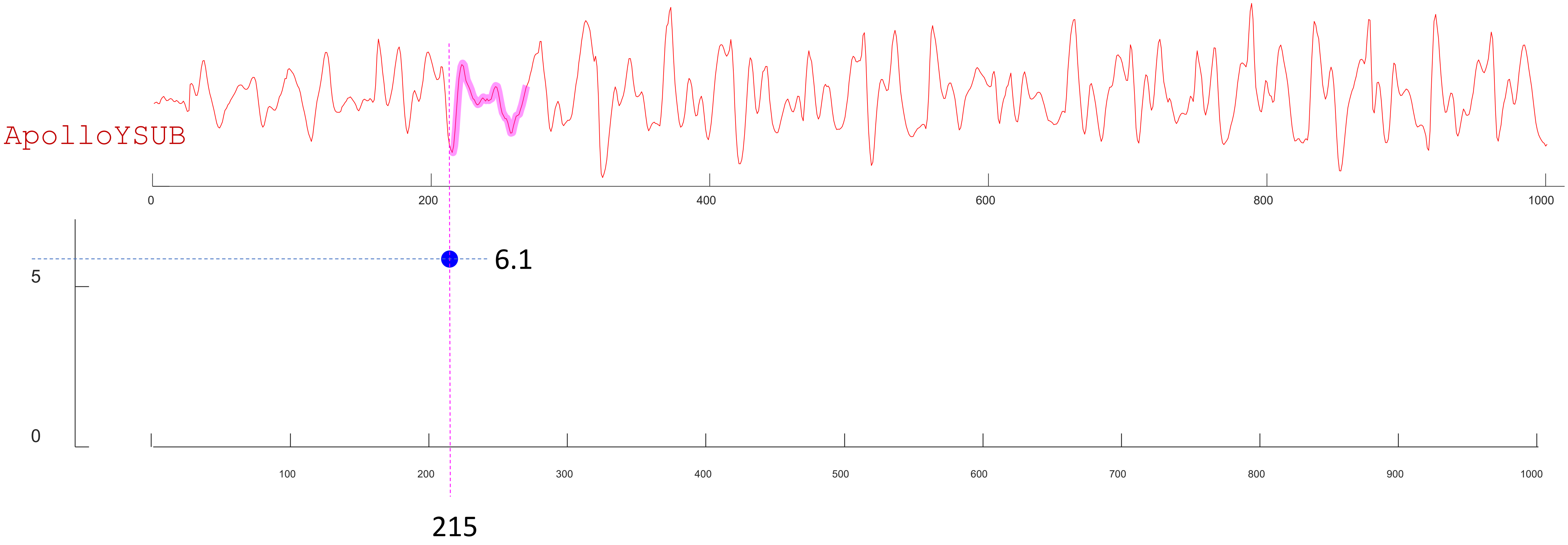
The matrix profile based framework yielded a cumulative average recall value of 0.92. This is proof that this algorithm is working remarkably well. Gandra 2022, Anomaly Detection for Smart Infrastructure: An Unsupervised Approach for Time Series Comparison

This *is* self-indulgent, but..

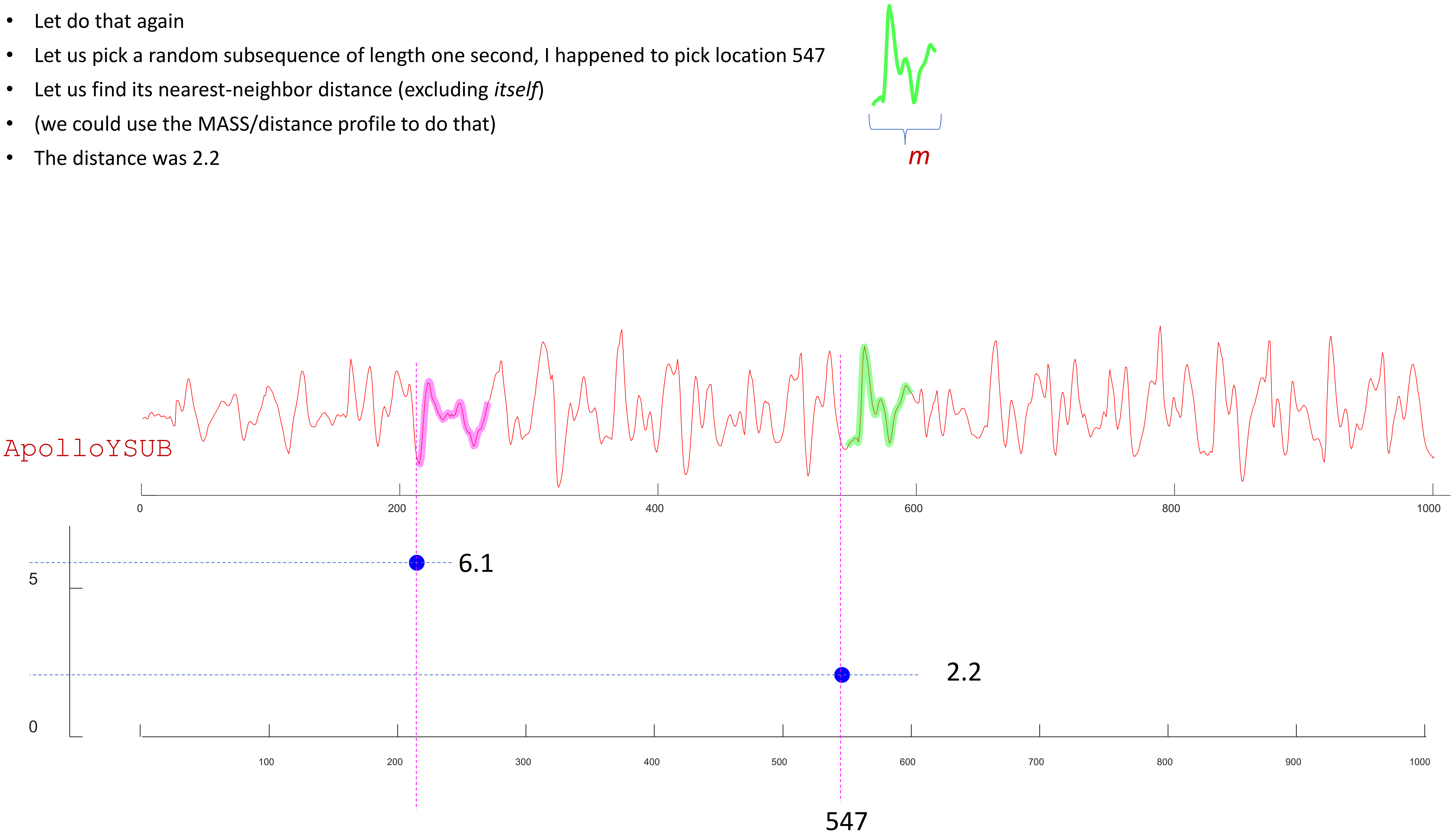


- Let us return to our small bovine example.
- Let us pick a random subsequence of length one second, I happened to pick location 215
- Let us find its nearest-neighbor distance (excluding *itself*) to anywhere else in the time series.
- (we could use the MASS/distance profile to do that)
- The distance was 6.1

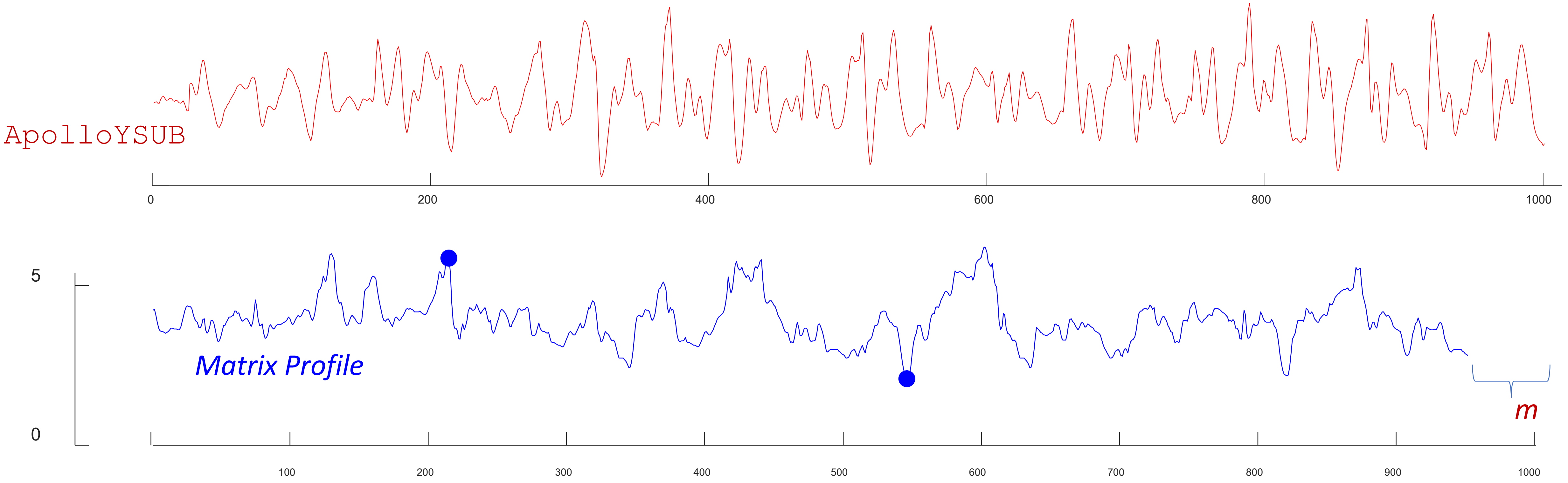




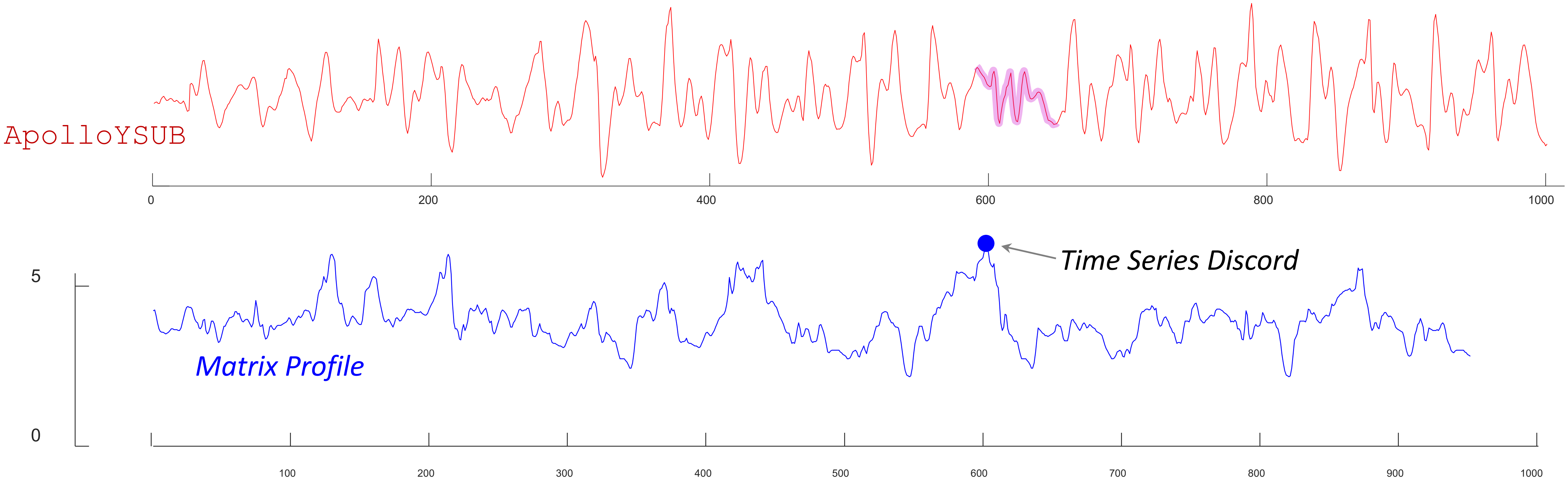
- Let do that again
- Let us pick a random subsequence of length one second, I happened to pick location 547
- Let us find its nearest-neighbor distance (excluding *itself*)
- (we could use the MASS/distance profile to do that)
- The distance was 2.2



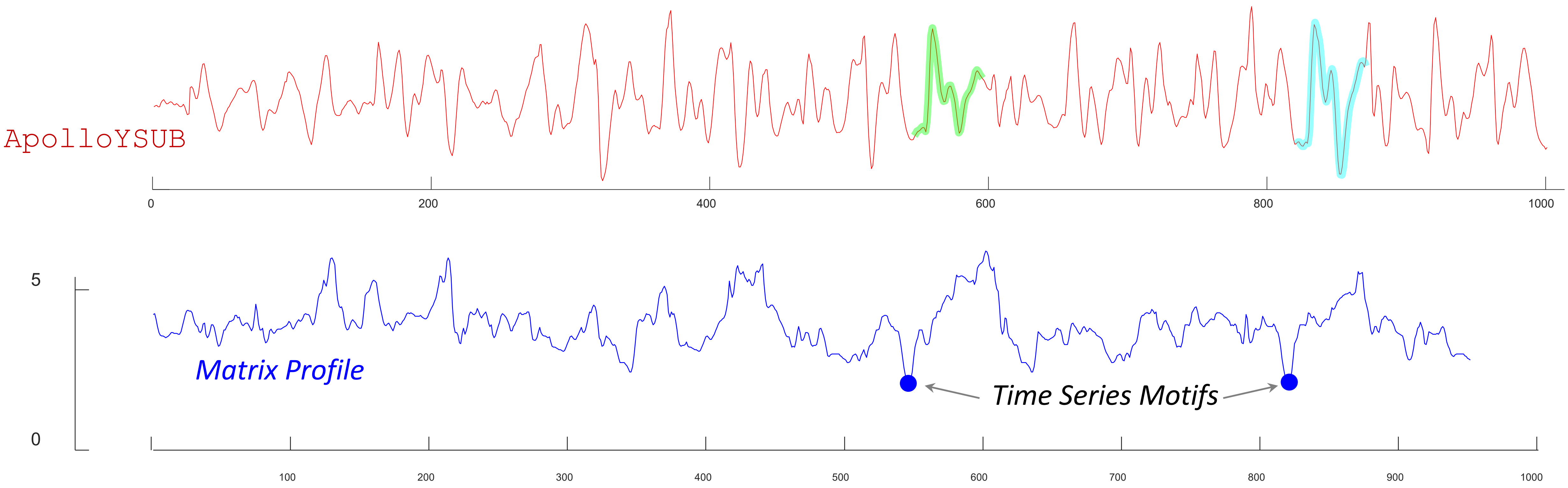
- Let us do this for every location!
- The resulting curve is called the *Matrix Profile*



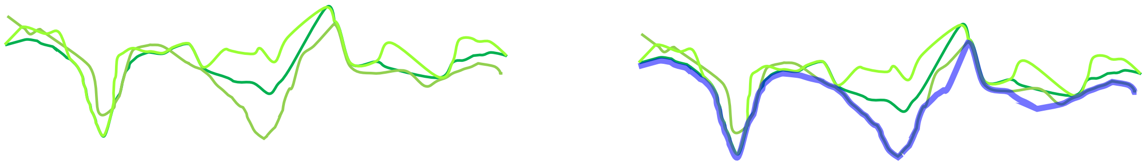
- There are some parts of the Matrix Profile that have special names
- The highest location is called the *Time Series Discord*



- There are some parts of the Matrix Profile that have special names
- The highest location is called the *Time Series Discord*
- The lowest locations (there will always be a tie) is called the *Time Series Motif* pair



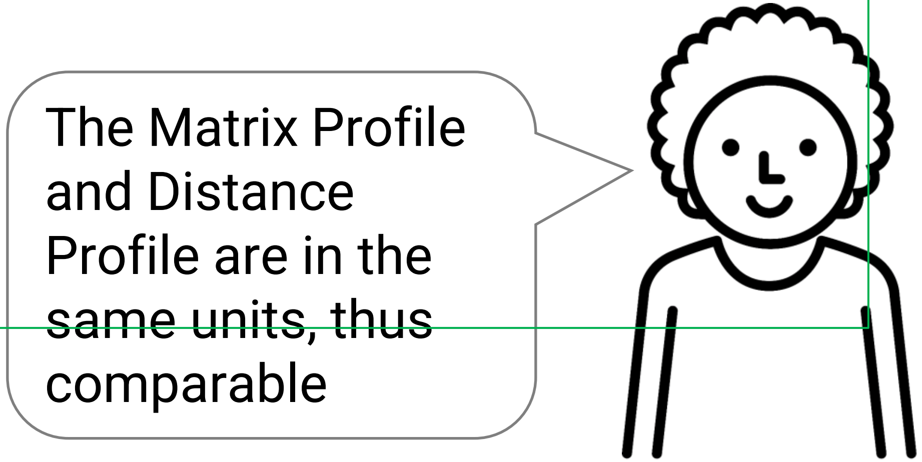
A Quick Aside



- What is the relationship between the *Distance Profile* and the *Matrix Profile?*
- If you..
  - Use every subsequence of a time series to create a set of distance profiles (setting the trivial matches to infinity)
  - Then for every location i, take the minimum value from the i^{th} location in the set of distance profiles
  - That would produce the Matrix Profile
- This is the *logical* relationship, there are faster and more space efficient ways to make the Matrix Profile.
- Nevertheless, it means that both are “in the same units”, which, as we will see, is useful.

A Useful Key

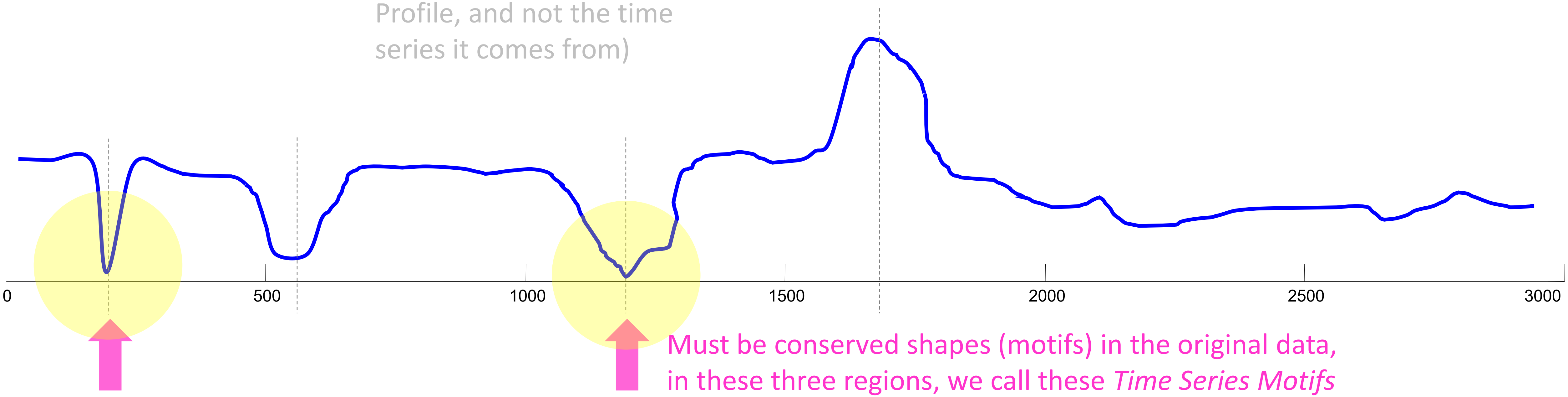
- Distance Profile is “one to many”, you provide a query
- Matrix Profile is “*many to many*”, you provide only the length



Reading a Matrix Profile

Where you see relatively low values, you know that the subsequence in the original time series must have (at least one) relatively similar subsequence elsewhere in the data (such regions are “motifs” or reoccurring patterns)

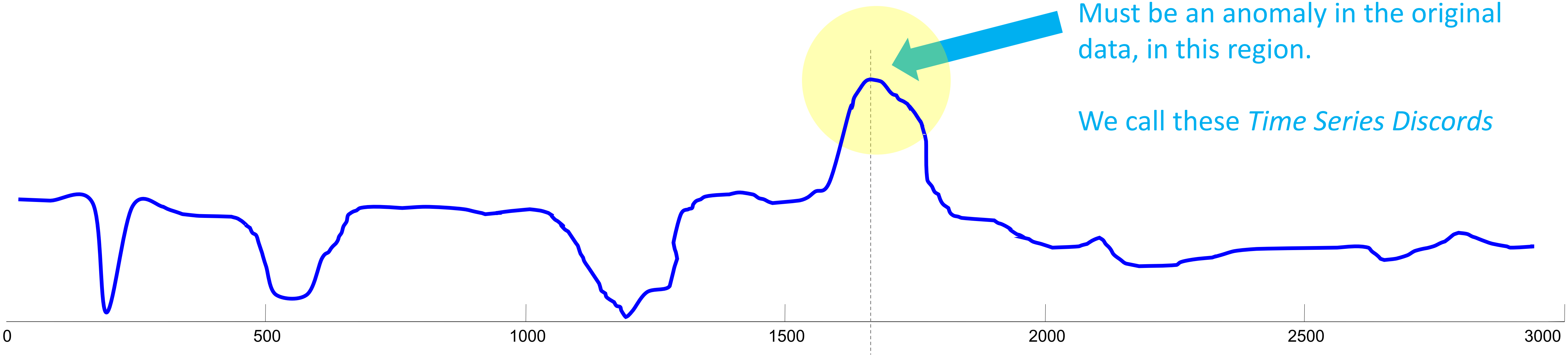
(here I am deliberately only showing you a Matrix Profile, and not the time series it comes from)



Must be conserved shapes (motifs) in the original data, in these three regions, we call these *Time Series Motifs*

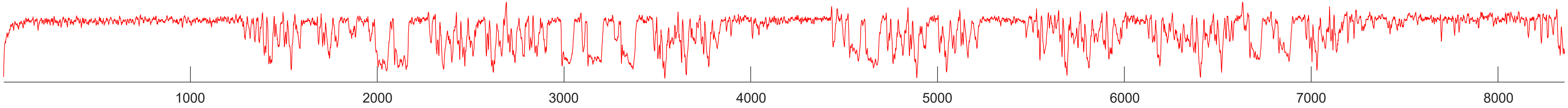
Reading the Matrix Profile

Where you see relatively high values, you know that the subsequence in the original time series must be unique in its shape (such areas are “discords” or anomalies).

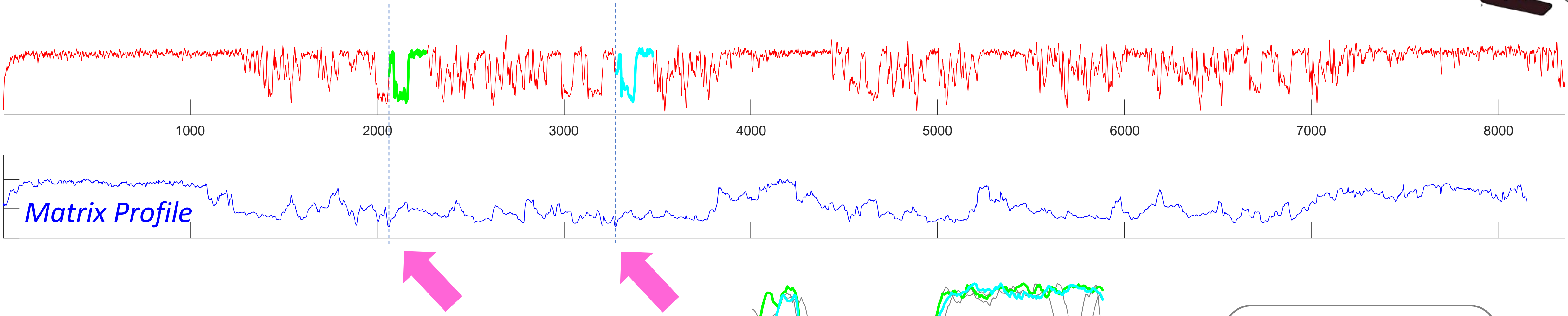


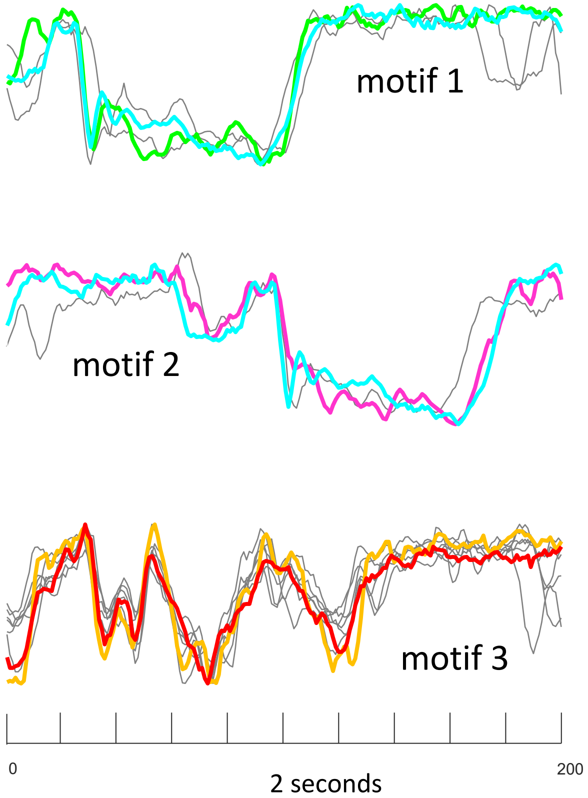
Zebra Finch

(Zebra Finch Vocalizations in MFCC, 100 day old male)

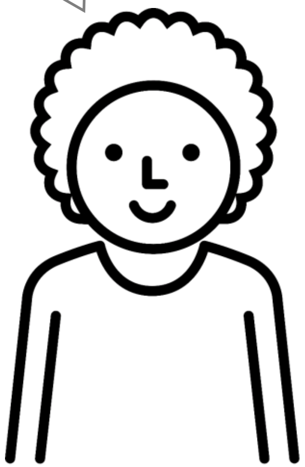


Zebra Finch

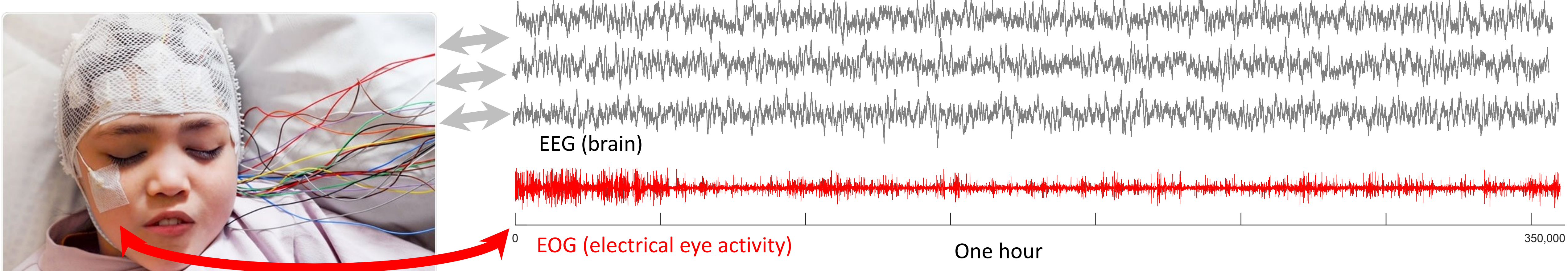




Motif discovery can often surprise you



While it is clear that this time series is not random, we did not expect the motifs to be so well conserved or repeated so many times. There is evidence of a *vocabulary*, and maybe even a *grammar*…



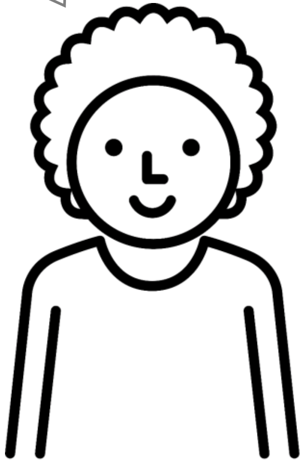
A quick case study

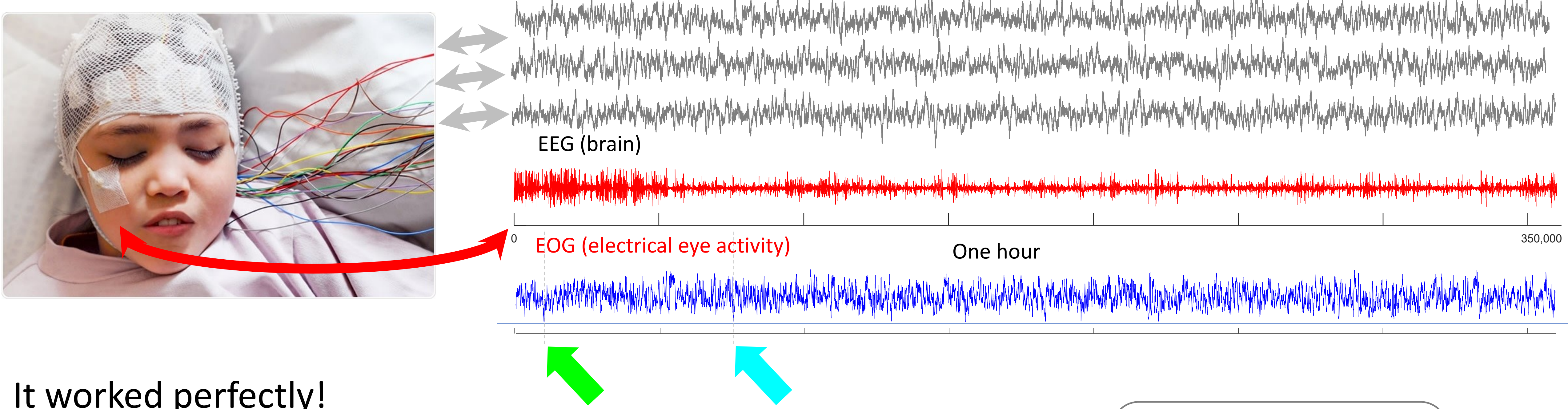
We have one hour of data, 64 EEG traces and one EOG trace.
The neuroscientists want to do some analysis on EEGs.
However, they don't want to consider any time periods in which there are eye blinks.
So, we need to search the EOG time series for any eye-blinks and record their locations.

Problem

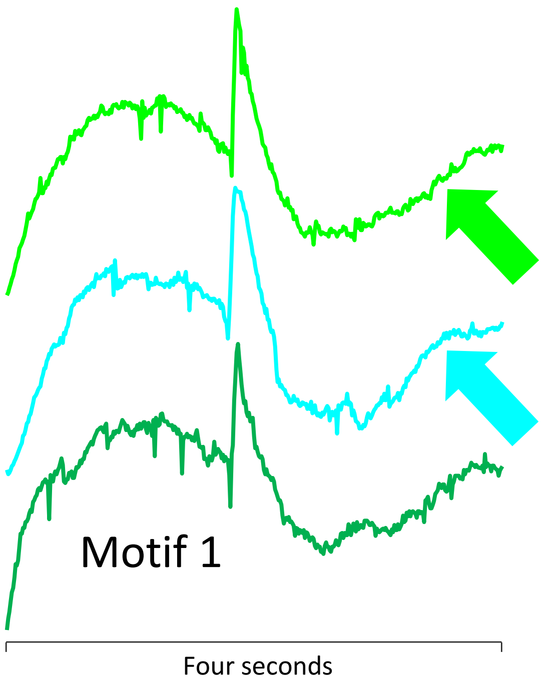
- Eye-blinks can vary from person to person
- Even for a single person, the placement of the sensor will change what the blinks look like, so you can't directly compare shapes day to day
- For most people, the shape of the blink can be polymorphic

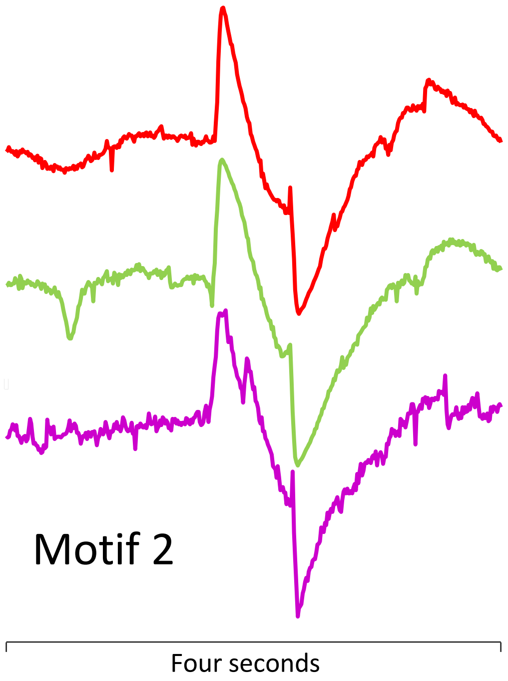
Let's solve a medically important problem with motifs



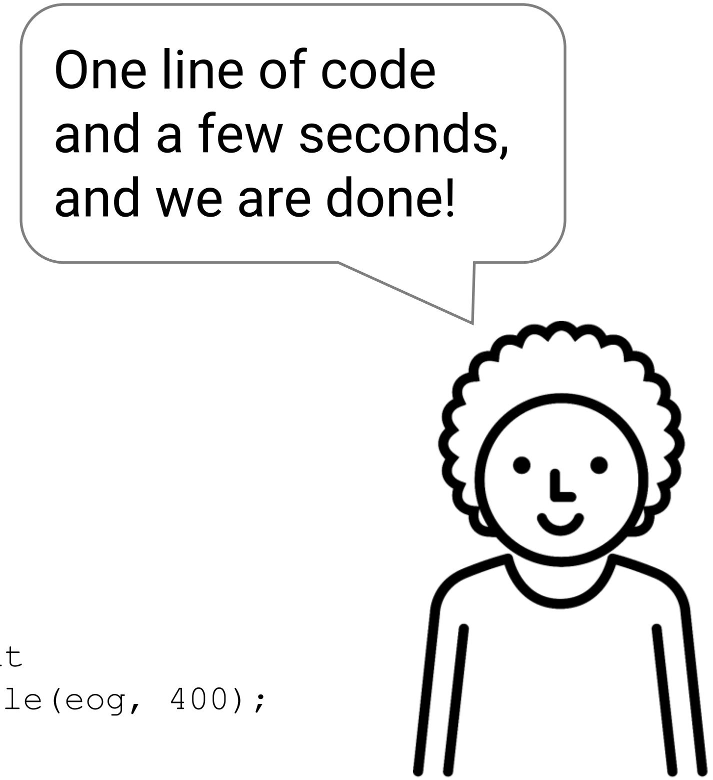


It worked perfectly!

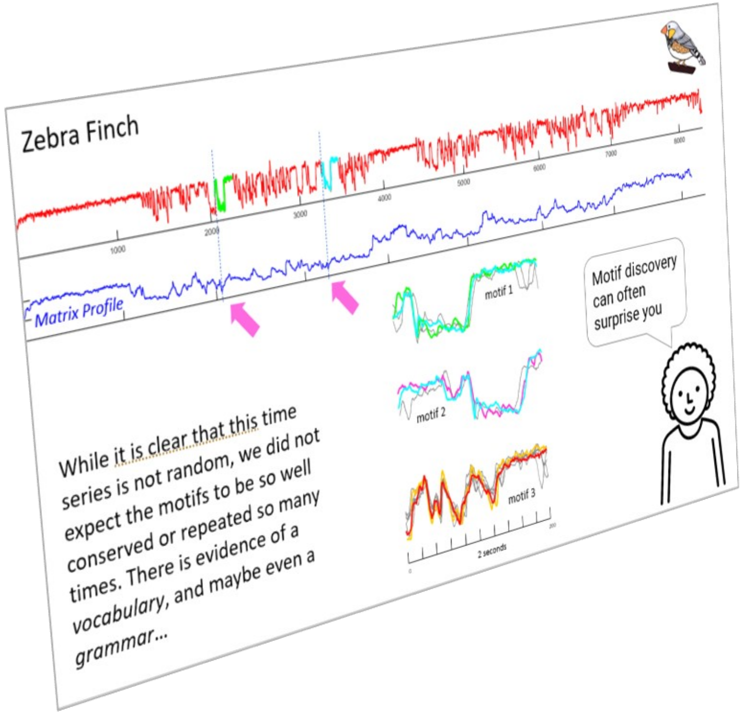


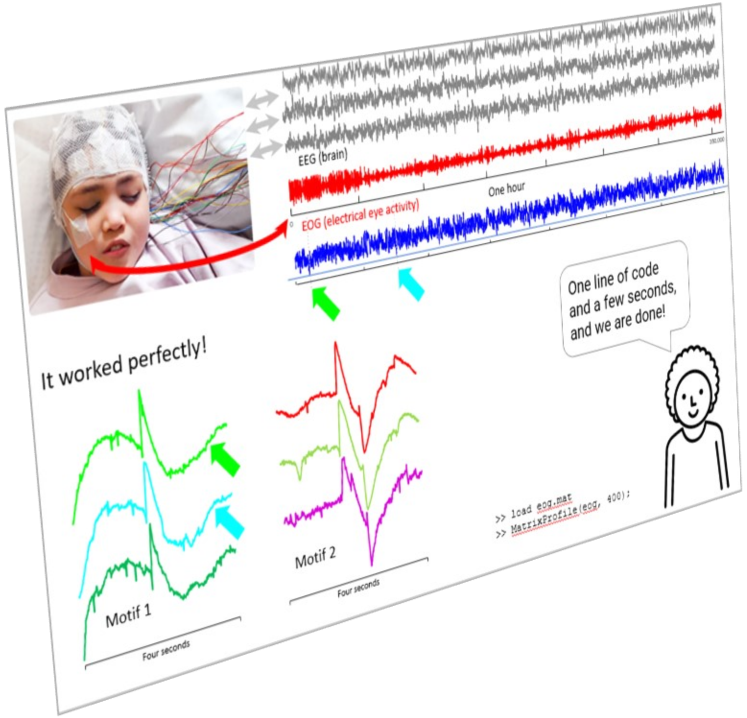



```
>> load eog.mat
>> MatrixProfile(eog, 400);
```



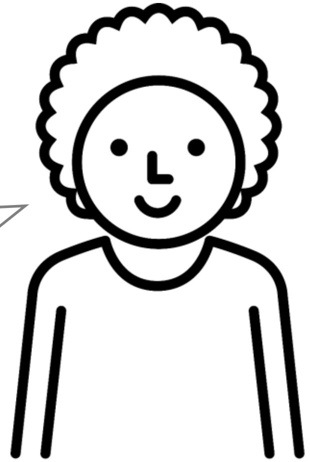
- For Zebra Finch and EOG, we were looking at the *lowest* values in the Matrix Profile, the *motifs*.
- In the next example (Taxi), we will look for the *highest* values in the Matrix Profile, the *discords*.





Just making sure you see this distinction

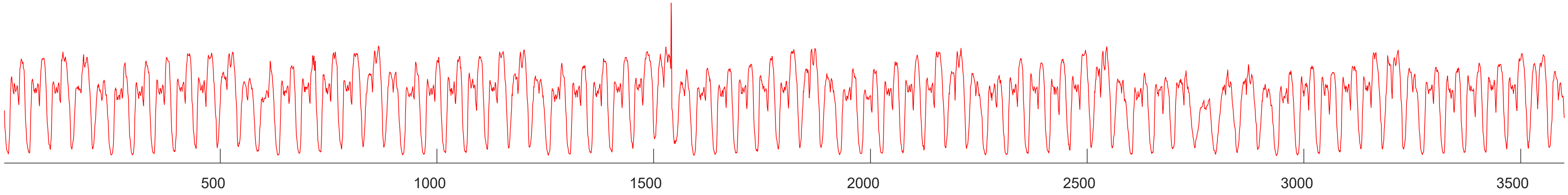
We will toggle between *motifs* and *discords* in this tutorial

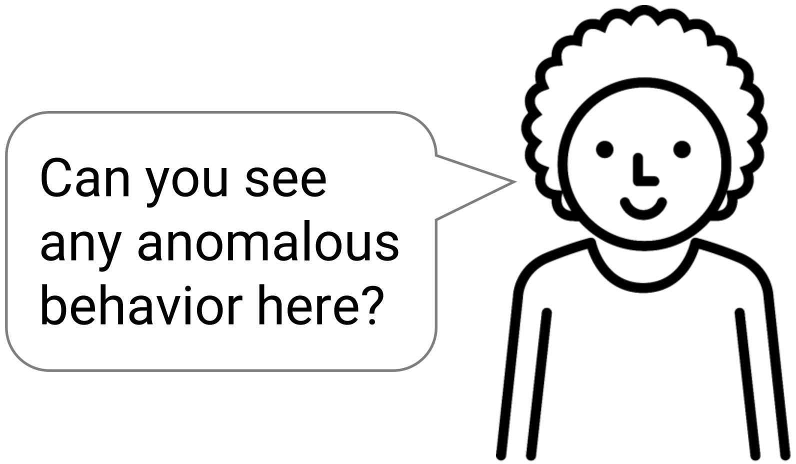


Taxi Example: Part I

Below is the hourly average of the number of NYC taxi passengers over 75 days in Fall of 2014.

Lets compute the Matrix Profile for it, we choose a subsequence length corresponding to two days…. (next slide)

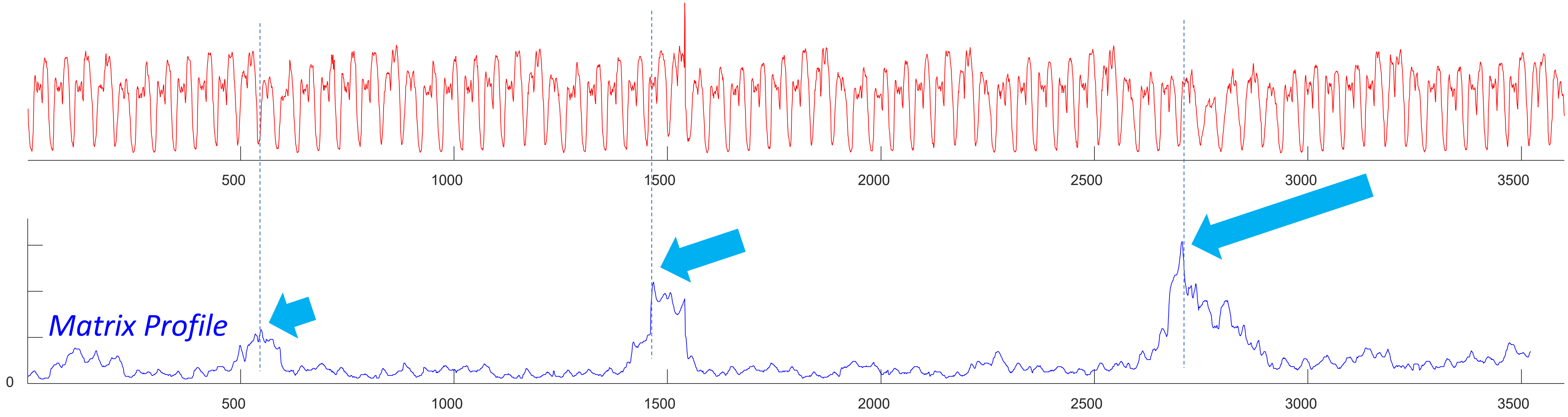




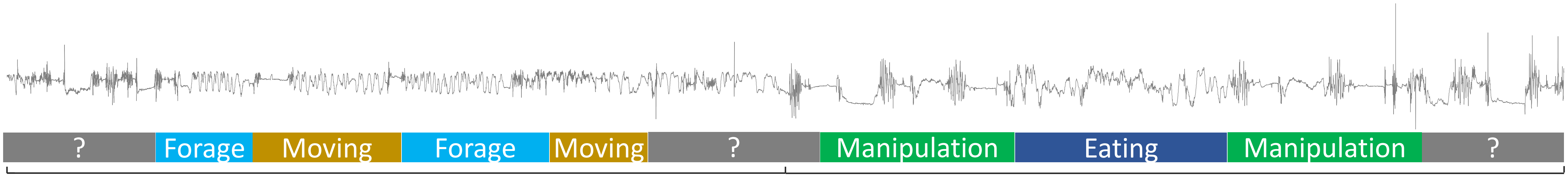
[a] http://futuredata.stanford.edu/ASAP/extended.pdf

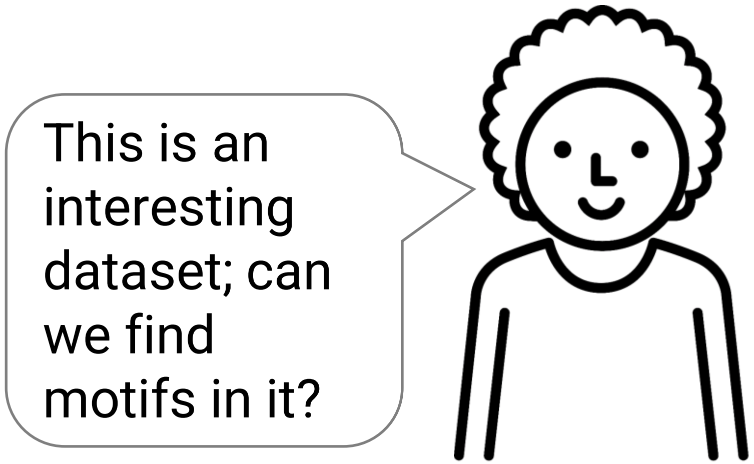
Taxi Example: Part II

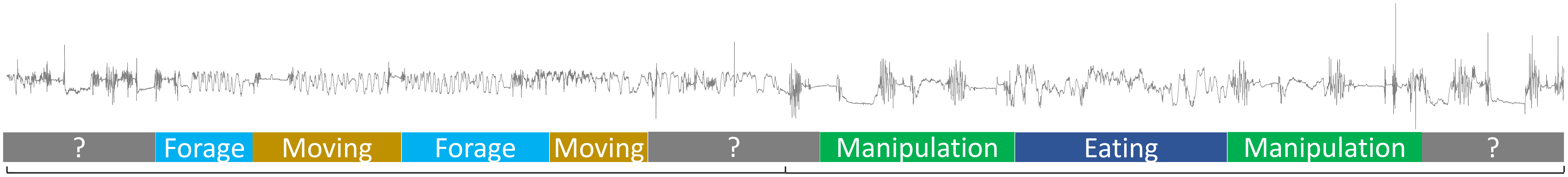
- The highest value corresponds to Thanksgiving
- We find a secondary peak around Nov 6th, what could it be? Daylight Saving Time! The clock going backwards one hour, gives an *apparent* doubling of taxi load.
- We find a tertiary peak around Oct 13th, what could it be? Columbus Day! Columbus Day is largely ignored in much of America, but still a big deal in NY, with its large Italian American community.



Australian Fur Seal (*Arctocephalus pusillus*)


?
Forage
Moving
Forage
Moving
?
Manipulation
Eating
Manipulation
?


This is an interesting dataset; can we find motifs in it?



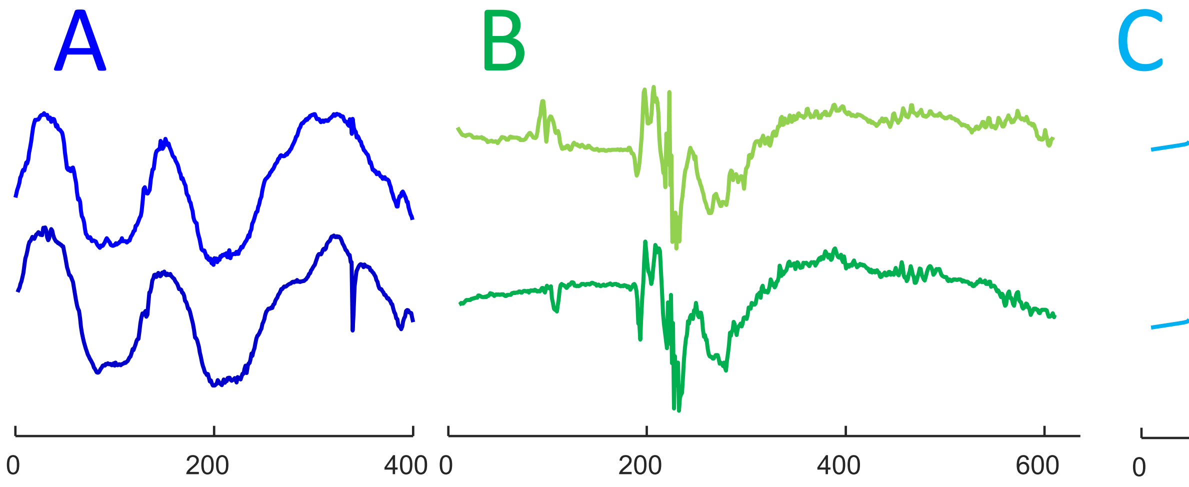
We can find *many* interesting motifs in this data.

Suppose we label them A, B, C , D etc.

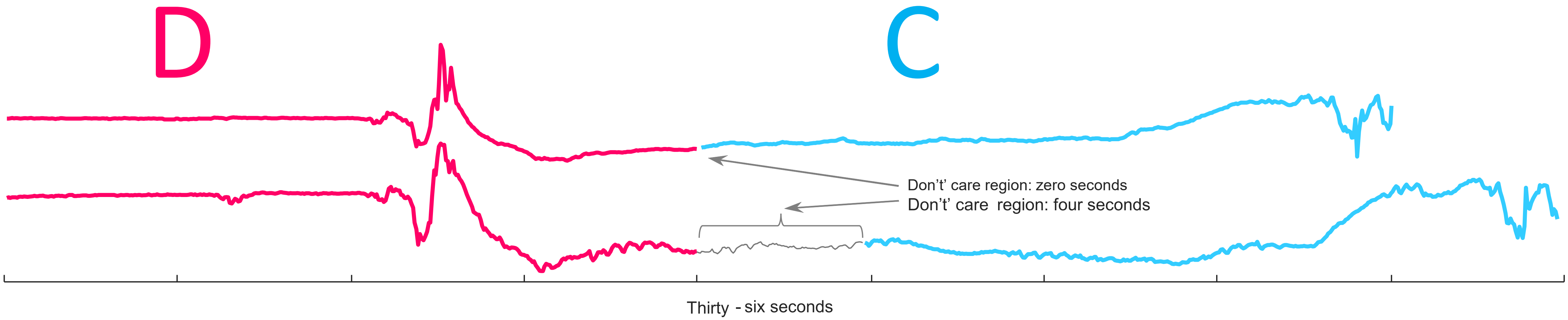
We can then ask the question, is there any patterns in the occurrence of these motifs?

In fact, there is, when we see D, we almost always see C within a few seconds…

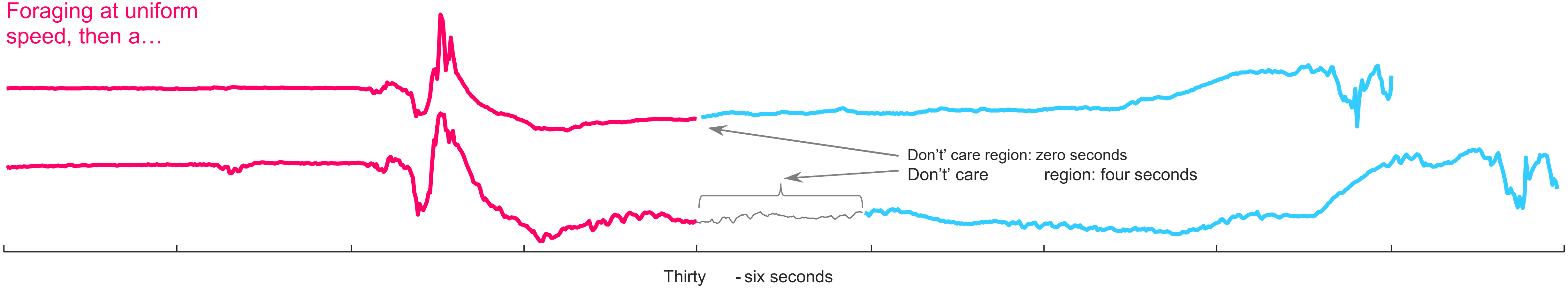
If D Then C

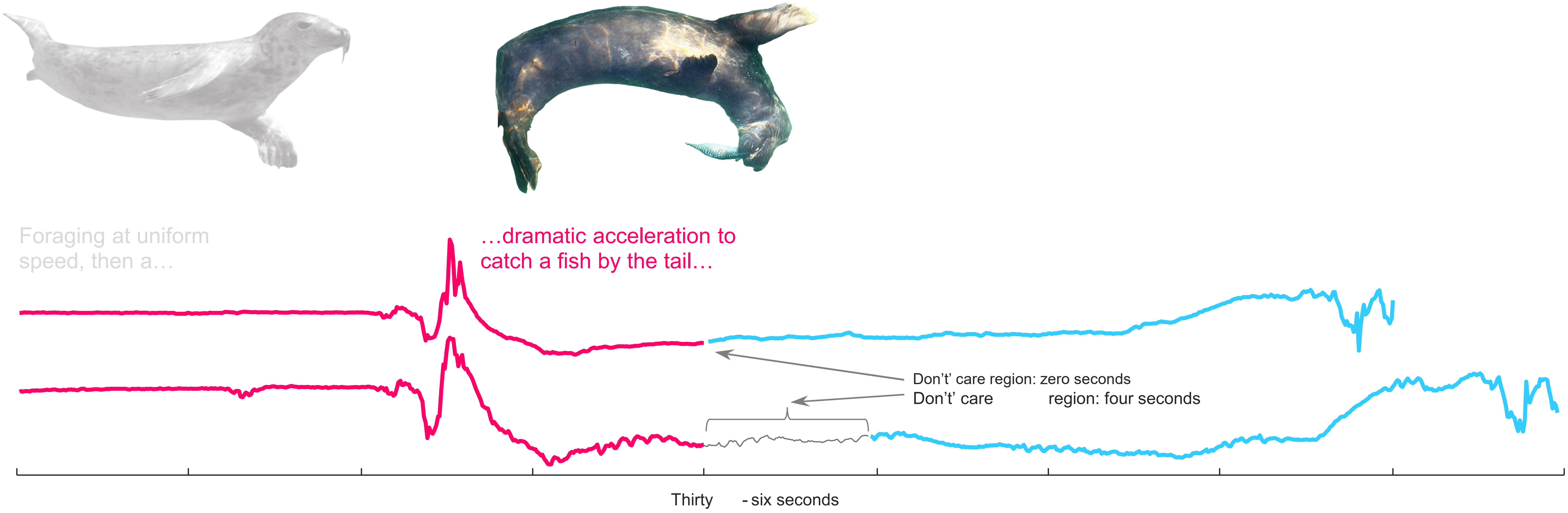


What does this mean?

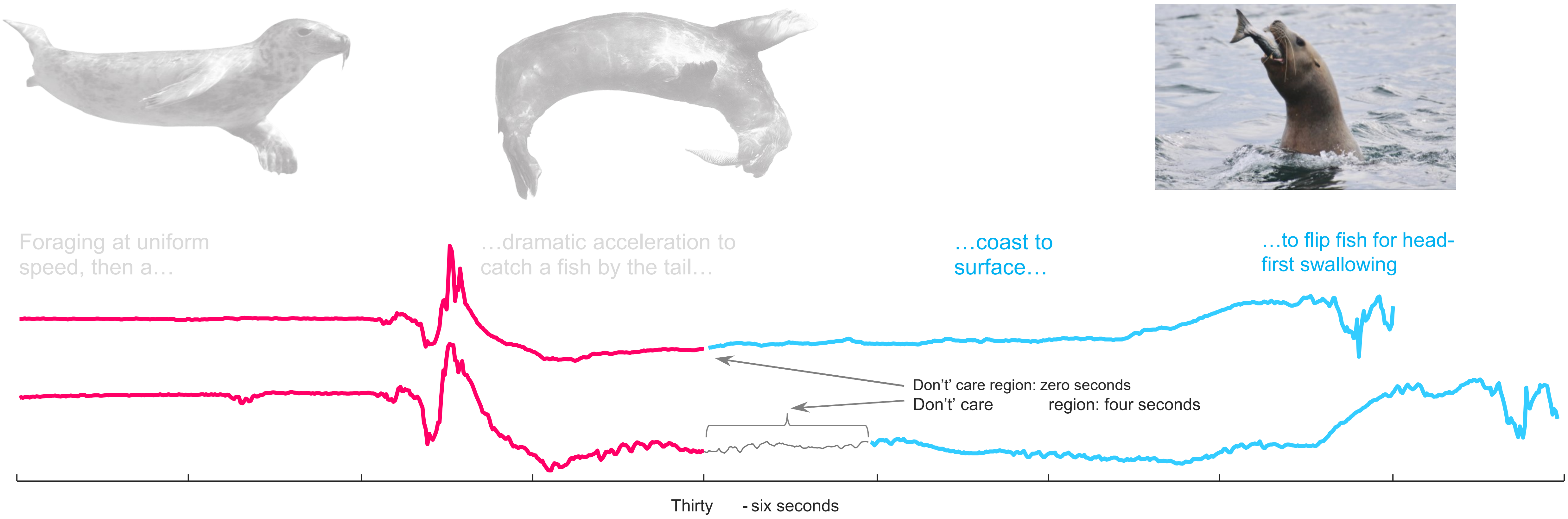


Foraging at uniform
speed, then a…





Foraging at uniform
speed, then a…
…dramatic acceleration to
catch a fish by the tail…
Don't' care region: zero seconds
Don't' care region: four seconds
Thirty - six seconds

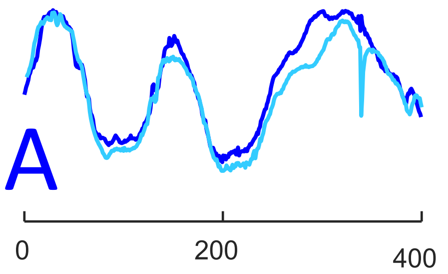

Foraging at uniform speed, then a…
…dramatic acceleration to catch a fish by the tail…
…coast to surface…
…to flip fish for head-first swallowing
Don't' care region: zero seconds
Don't' care region: four seconds
Thirty - six seconds

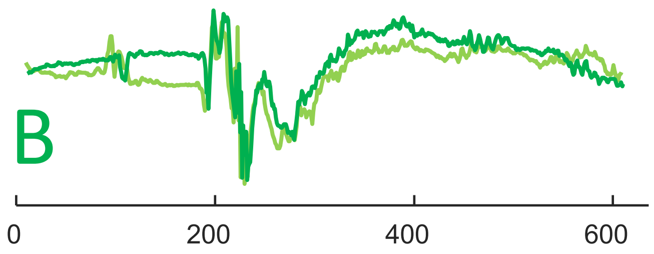
This idea is very general

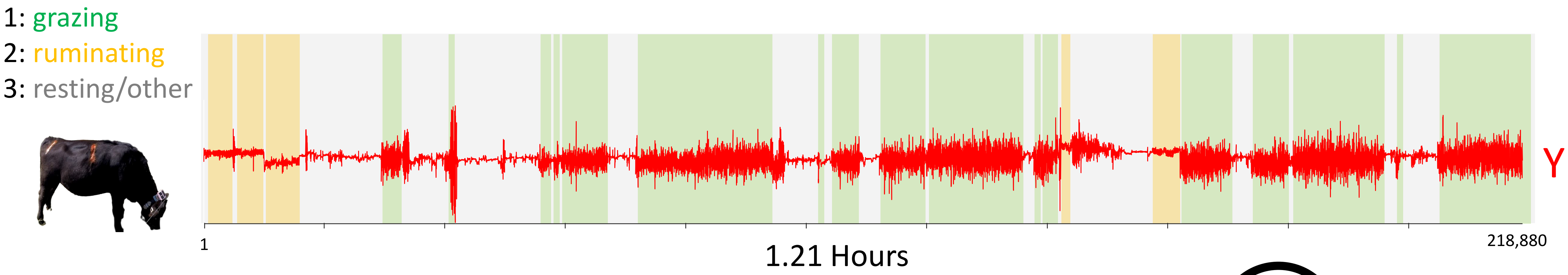
Find motifs, label them A, B, C , D , E etc.

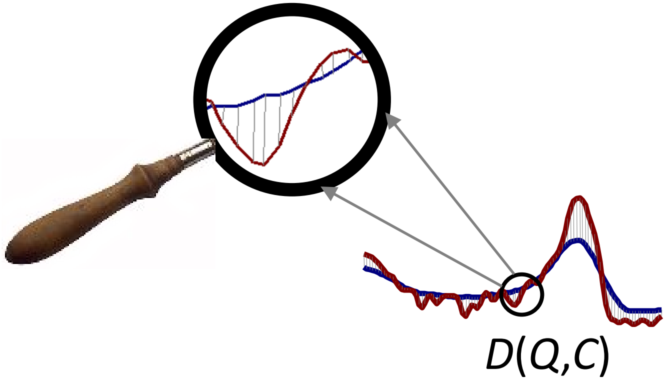
Now you can ask lots of interesting questions in the symbolic space, that are not defined in the real-valued space

- We see about 10 motif B's per hour in males, but only 1 or 2 motif B's per hour in females, why?
- If I give my patient the drug *warfarin*, does it change the frequencies of any of the motifs?
- Which motifs (if any) are associated with machine breakdowns?
- Are any of the motifs dependent on the outside temperature?
- It seems like motif D is much more common in good drivers. So, let me change these two things under my control, to see if it increases the frequency of the D motif.









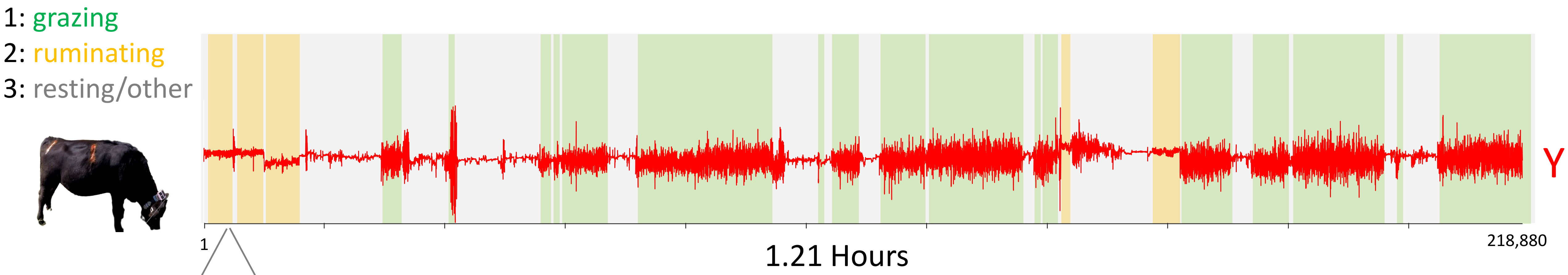
Let us return to Apollo

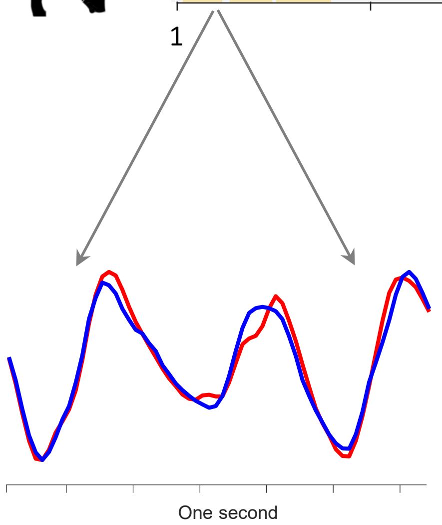
Let's compute the Matrix Profile for this dataset

To do so, we have to compute the Euclidean Distance between the approximately 24 billion possible pairwise combinations of subsequences.

Because of an amazing algorithm called SCRIMP++, we can do this in seconds!

There are other MP algorithms: STAMP, STAMPI, STOMP, SWAMP, DAMP, SCAMP, GPU-STOMP, TranSCRIMP, TranSCAMP and TranSCAMPfpga…





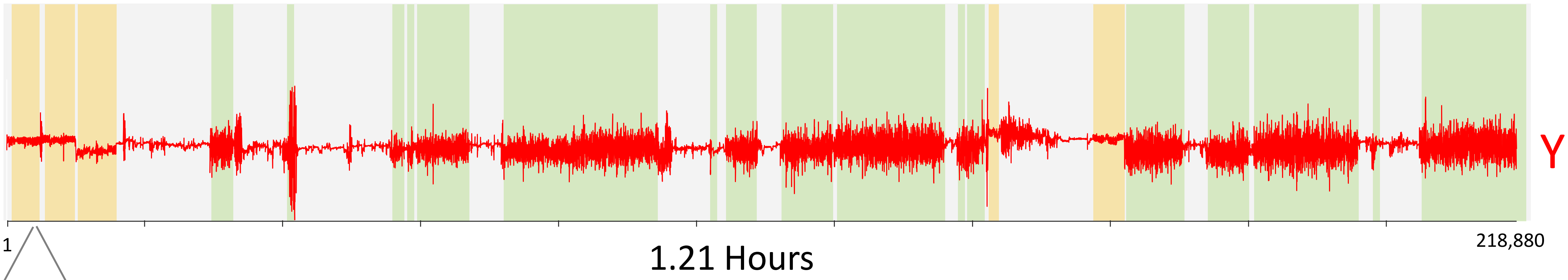
Here is the best motif.

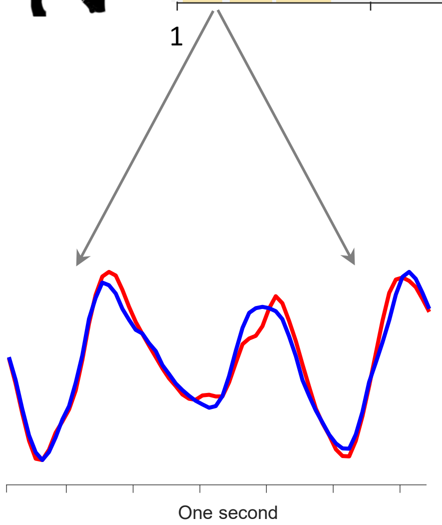
Note that both occurrences happen during **ruminating**

They are 1.29 units apart

Here 3 is a magic constant that could be tuned, to change precision/recall

1: grazing
2: ruminating
3: resting/other



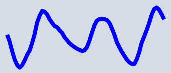


Here is the best motif.

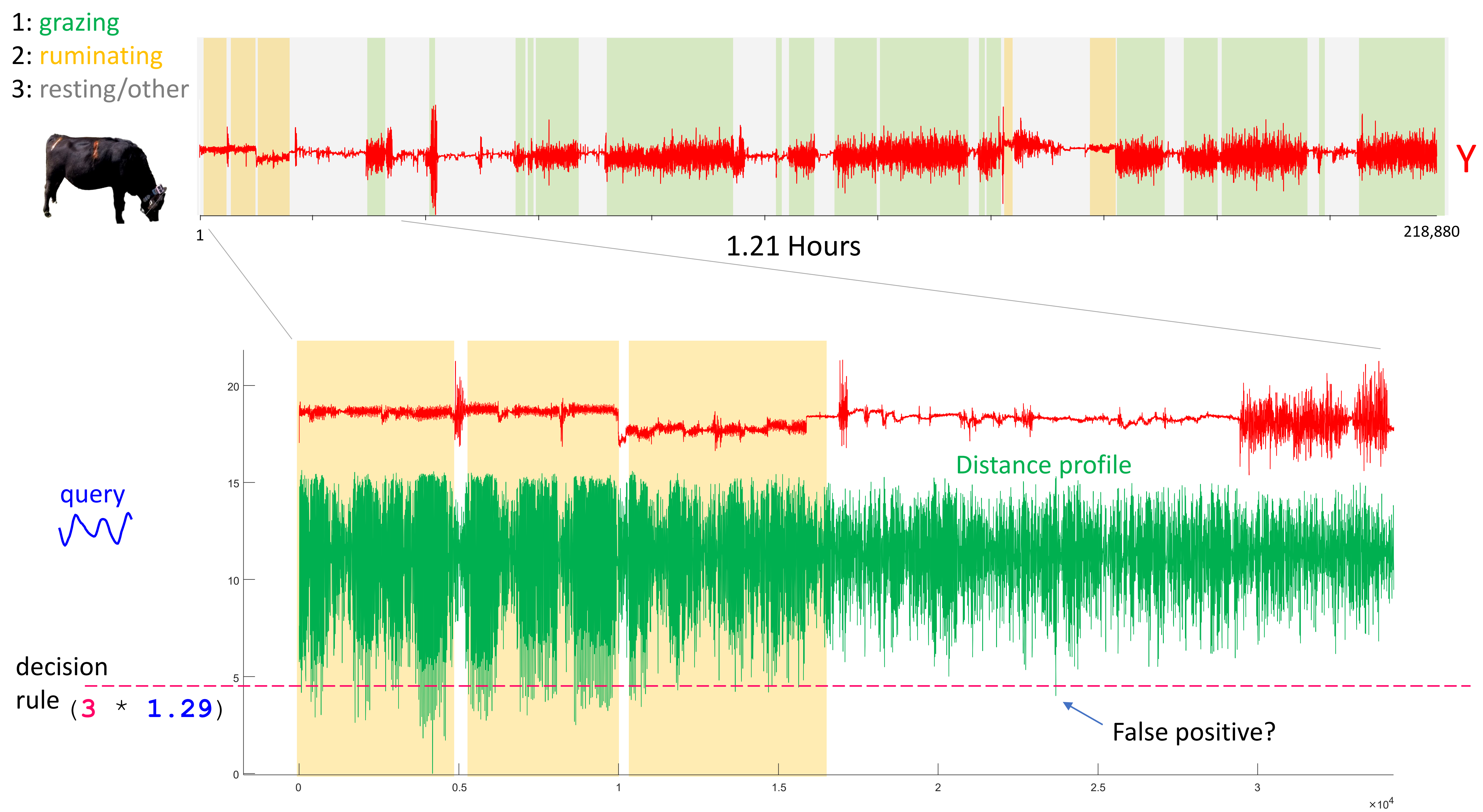
Note that both occurrences happen during **ruminating**

They are 1.29 units apart

This suggest that we may have found a *decision rule*

If you see a subsequence(X) such that
D(X, ) < (**3** * **1.29**)
Then Print('Bovine is **ruminating!**')

Here 3 is a magic constant that could be tuned, to change precision/recall



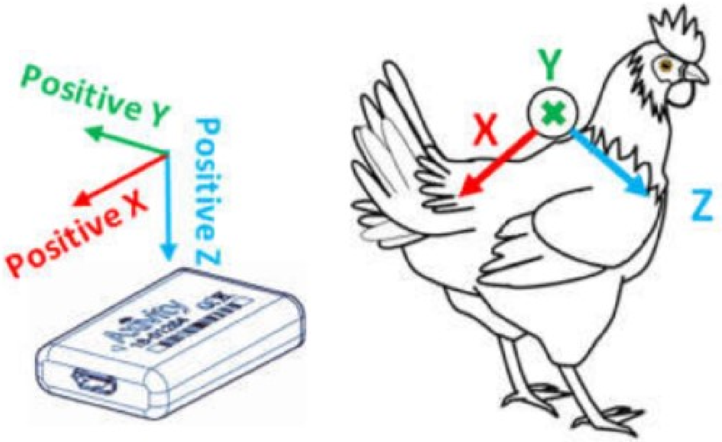
1: grazing
2: ruminating
3: resting/other
Y
1
218,880
1.21 Hours
Distance profile
query
decision
rule (3 * 1.29)
False positive?
20
15
10
5
0
0
0.5
1
1.5
2
2.5
3
×10^4

Mini Review I

- What I have shown you is *amazing* if you think about it.
  - Knowing *nothing* about bovine behavior…
  - Using less than eight lines of code.
  - Using one minute of brain power, and a few seconds of CPU power.
  - I built a tool to correctly annotate complex behavior in a bovine.

Chicken Example

- What I have shown you is *amazing* if you think about it.
  - Knowing *nothing* about bovine behavior…
  - Using less than eight lines of code in total.
  - Using one minute of brain power, and a few seconds of CPU power.
  - I built a tool to correctly annotate complex behavior in a bovine.

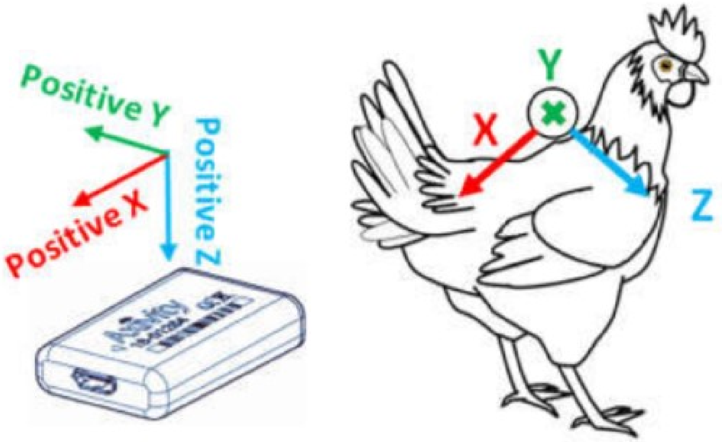


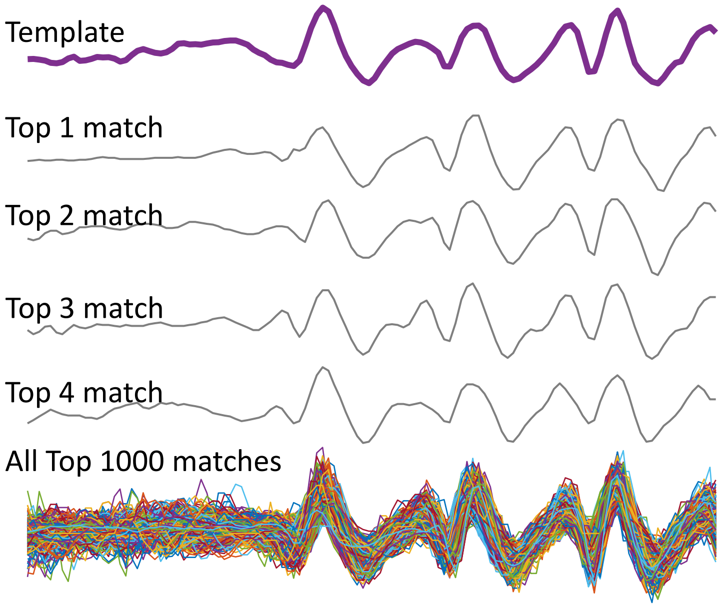
- Was this a fluke?
- Let quickly do it again!
- This time with chickens.
- I have *four full years* of chicken data!
- (many different chickens in parallel, over months)

Chicken Example

- Here is the template that motif discovery found

- It was associated with *Dustbathing*
- I used it to search a 12,679,054,727 datapoint archive of chicken behavior for the one thousand best matches. i.e. the 1,000 nearest neighbors

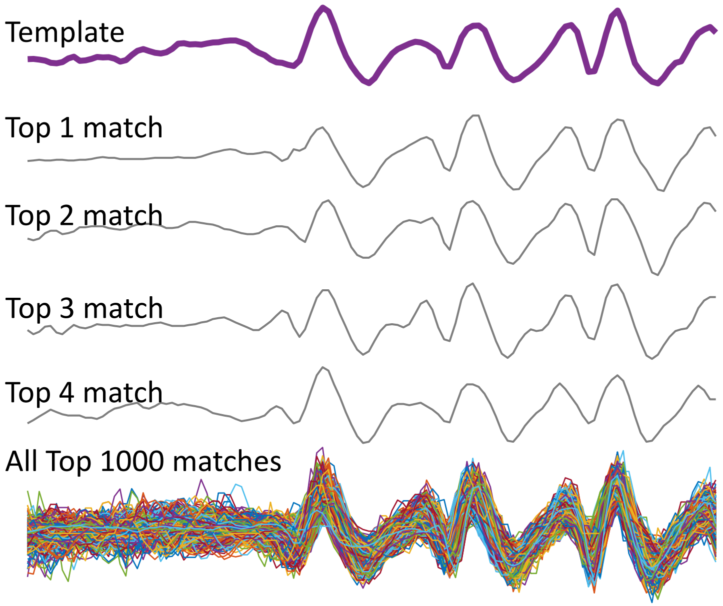




Chicken Example

- Here is the template that motif discovery found
- It was associated with *Dustbathing*
- I used it to search a 12,679,054,727 datapoint (four full years) archive of chicken behavior for the one thousand best matches. i.e. the 1,000 nearest neighbors

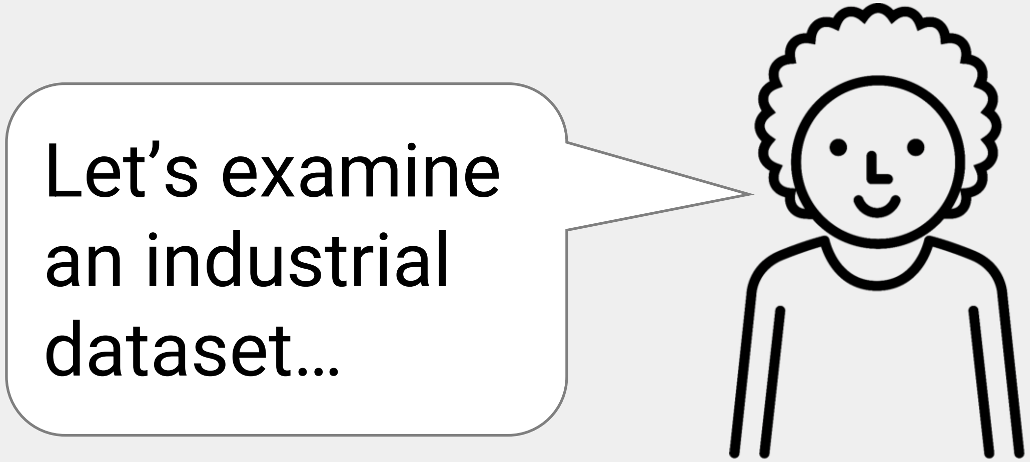
- It gets better.
- My data is annotated with two types of chicken: `has-mites|no-mites`.
- I can see that *Dustbathing* is more common in the `has-mites` class, so I have learned that chickens with mites will dustbath more.



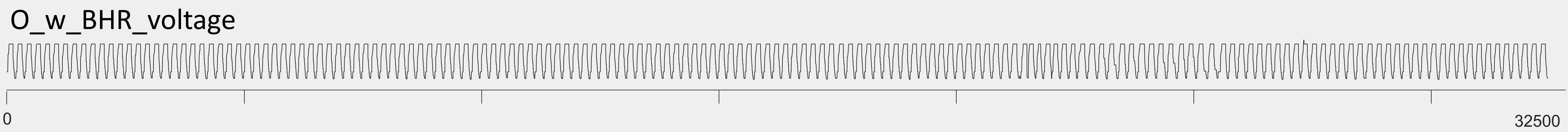
We have seen the utility of the Matrix Profile for *Motifs*..
Let's now see utility of the Matrix Profile for *Discords*..

The Matrix Profile as an Anomaly Detector

- Time Series Anomaly Detection (TSAD) is hot!
- I claim that *time series discords* are SOTA
- Let's see an example of their utility

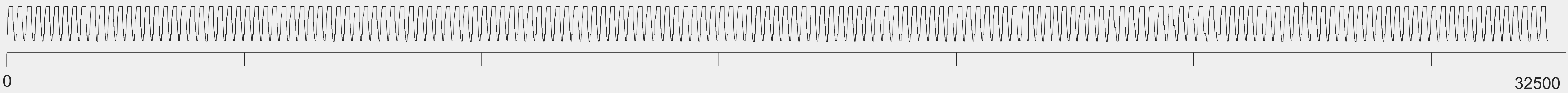


High Storage System Data for Energy Optimization

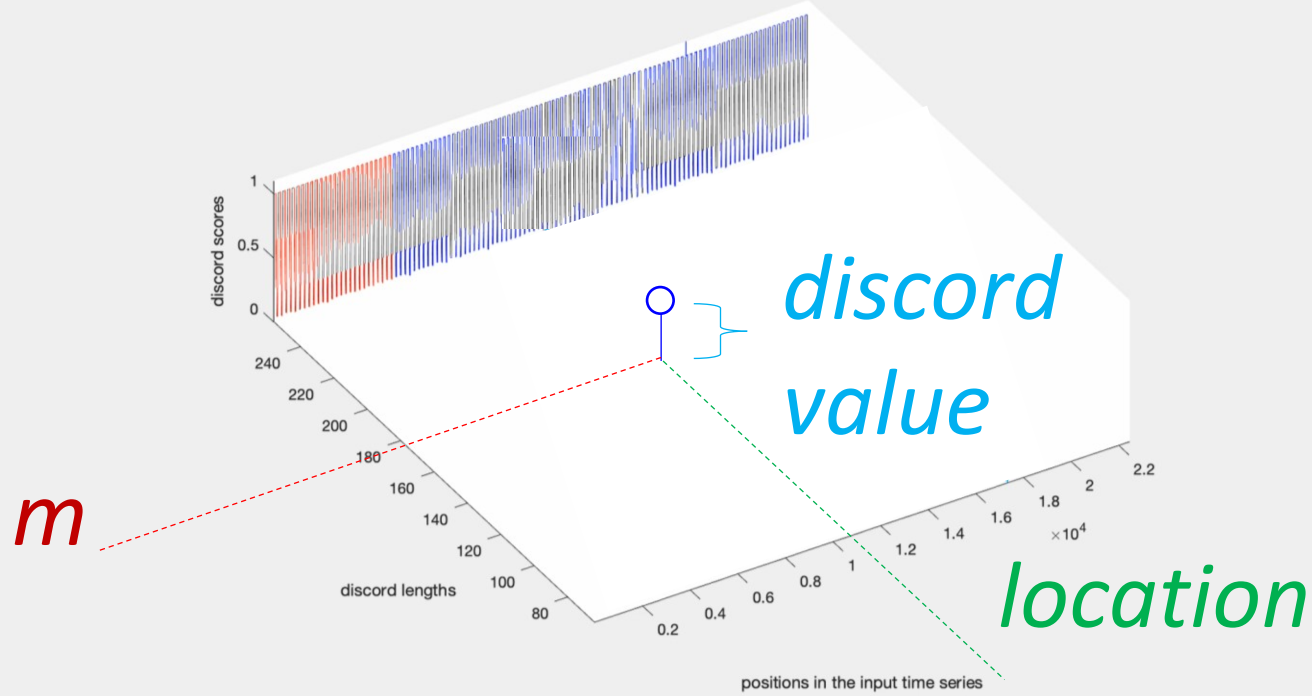


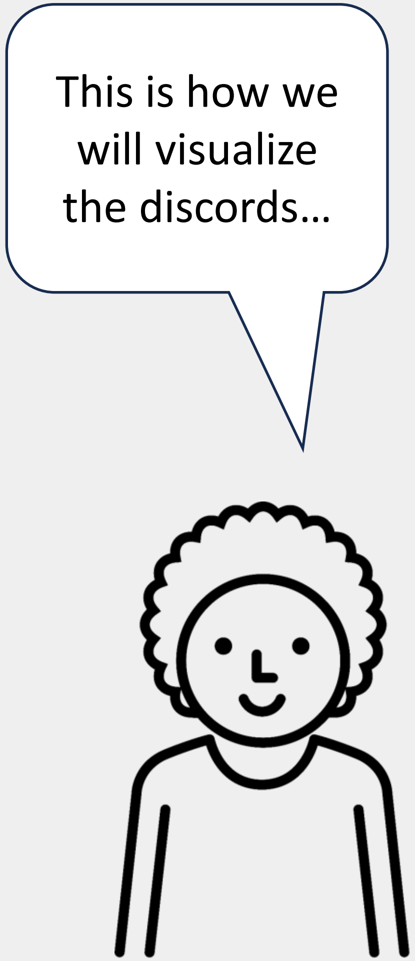
High Storage System Data for Energy Optimization

O_w_BHR_voltage



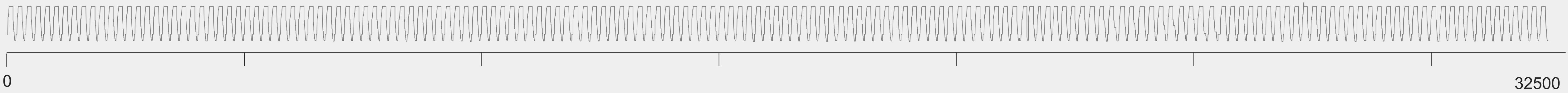
We only plot the top-1 discord
For a subsequence length *m*
We show the location it occurs
With a stem proportional to the discord value

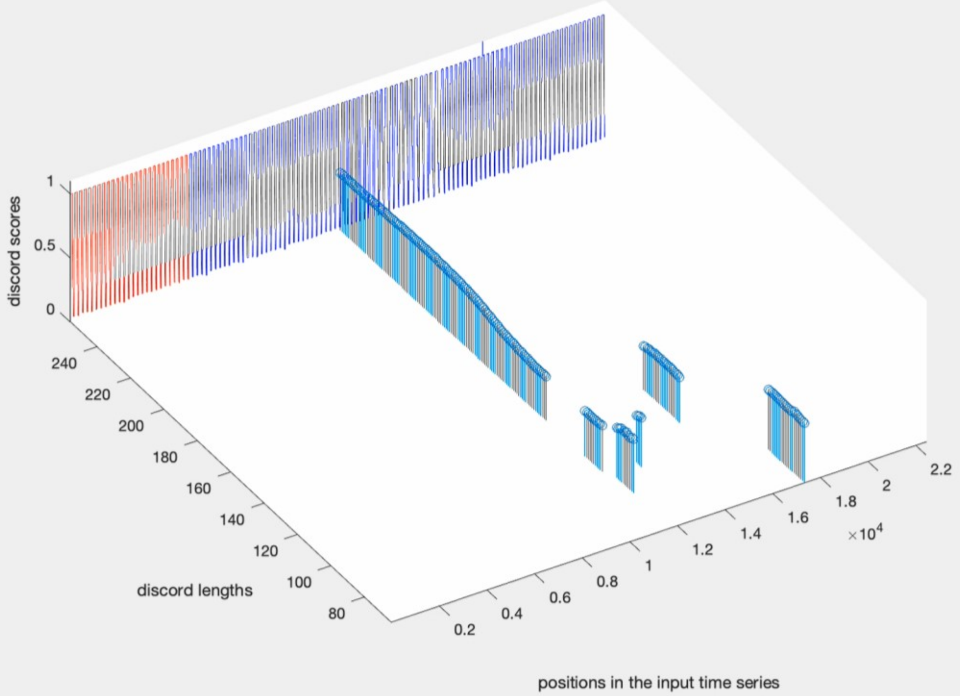




High Storage System Data for Energy Optimization

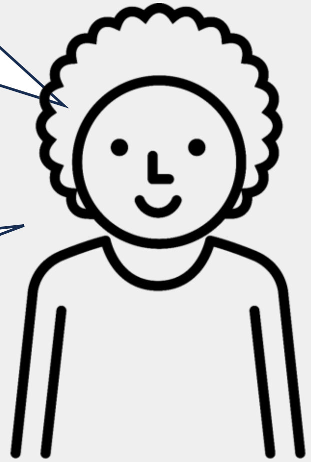
O_w_BHR_voltage



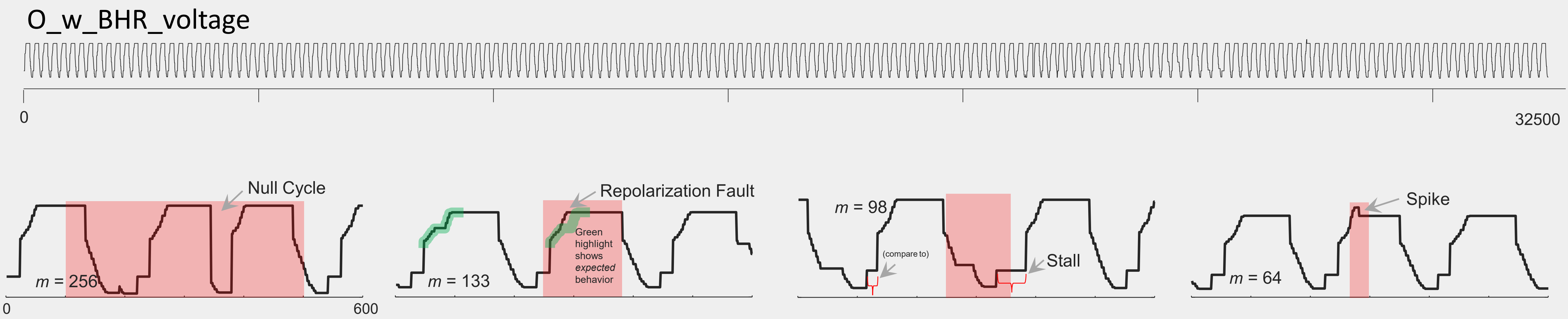


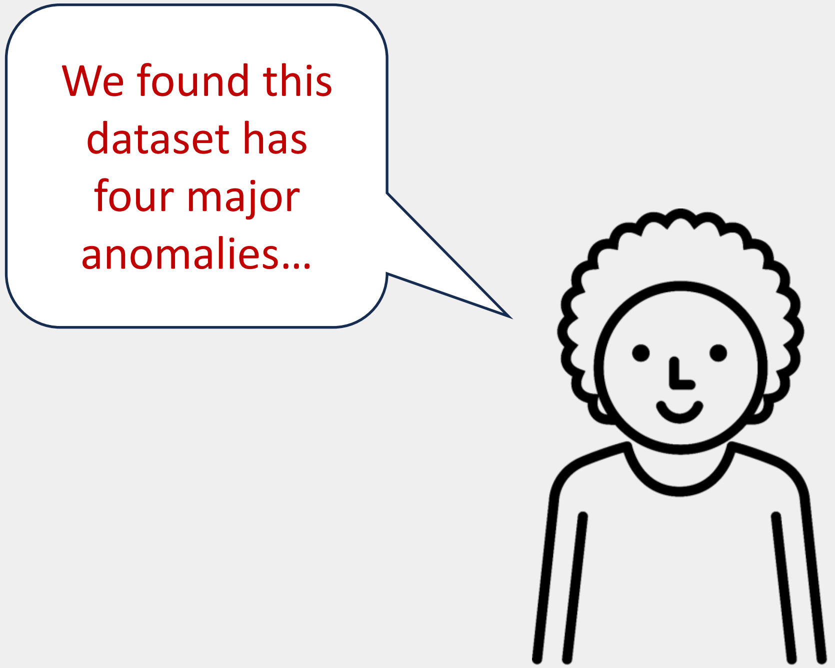
However, the Matrix Profile is so fast, we will do this for *every* subsequence length *m* from 64 to 256

There appear to be four anomalies..

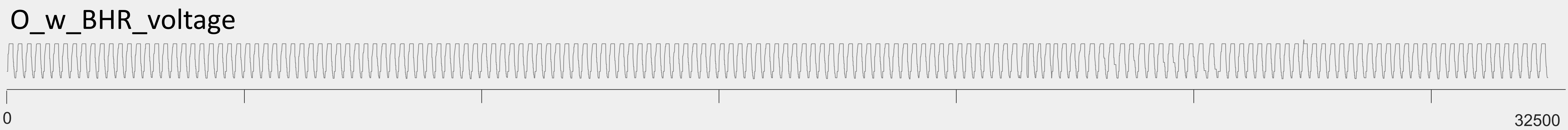


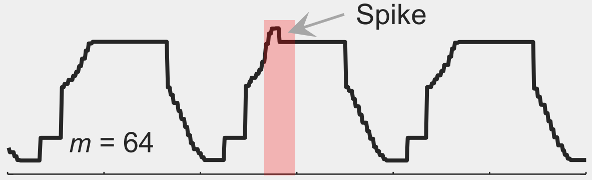
High Storage System Data for Energy Optimization

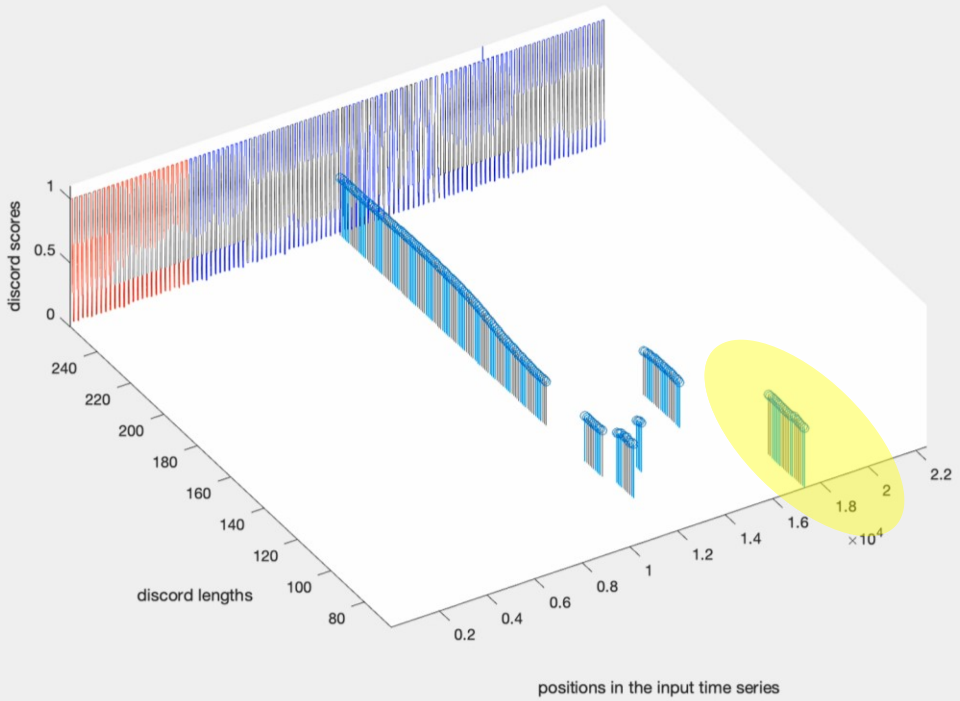




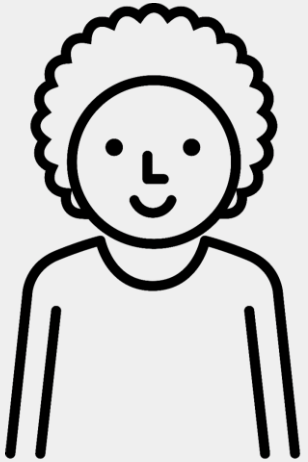
High Storage System Data for Energy Optimization



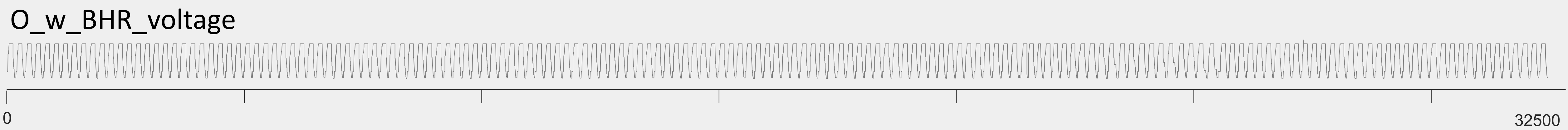


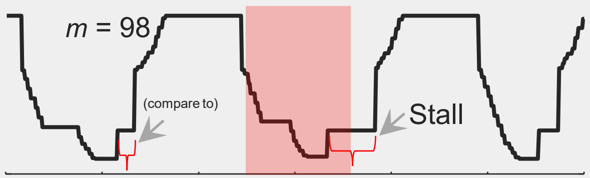


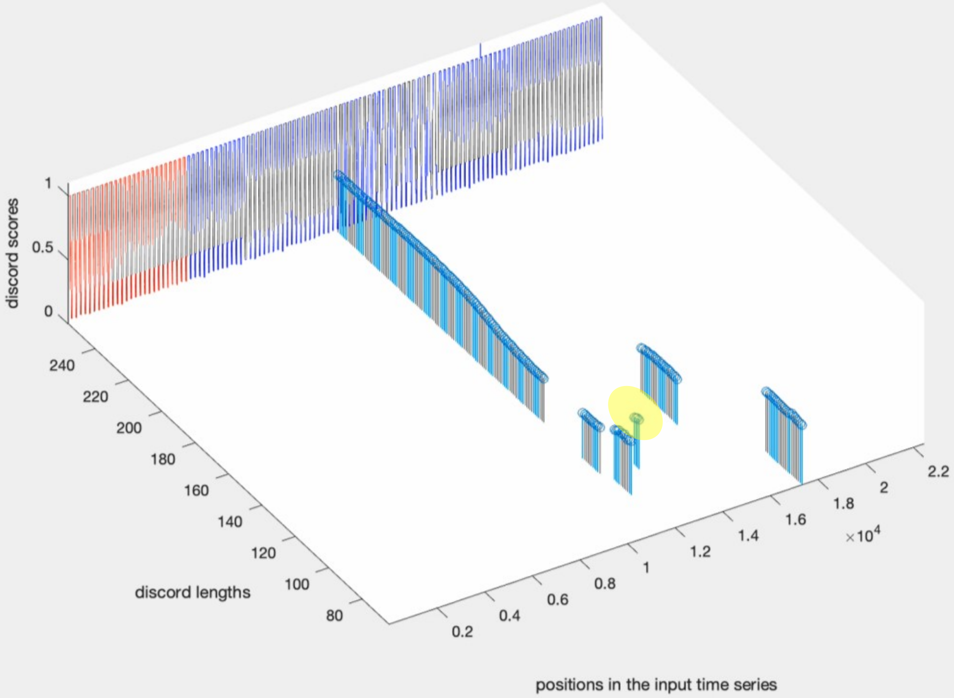
A “spike” shows up at short lengths…



High Storage System Data for Energy Optimization

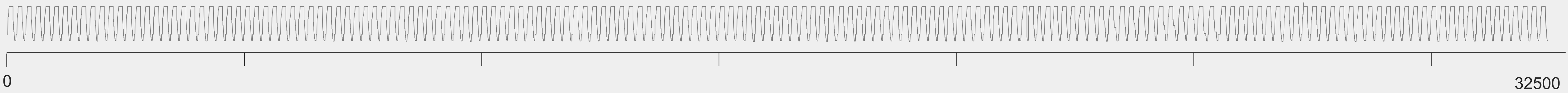


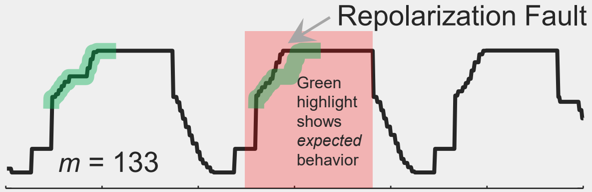


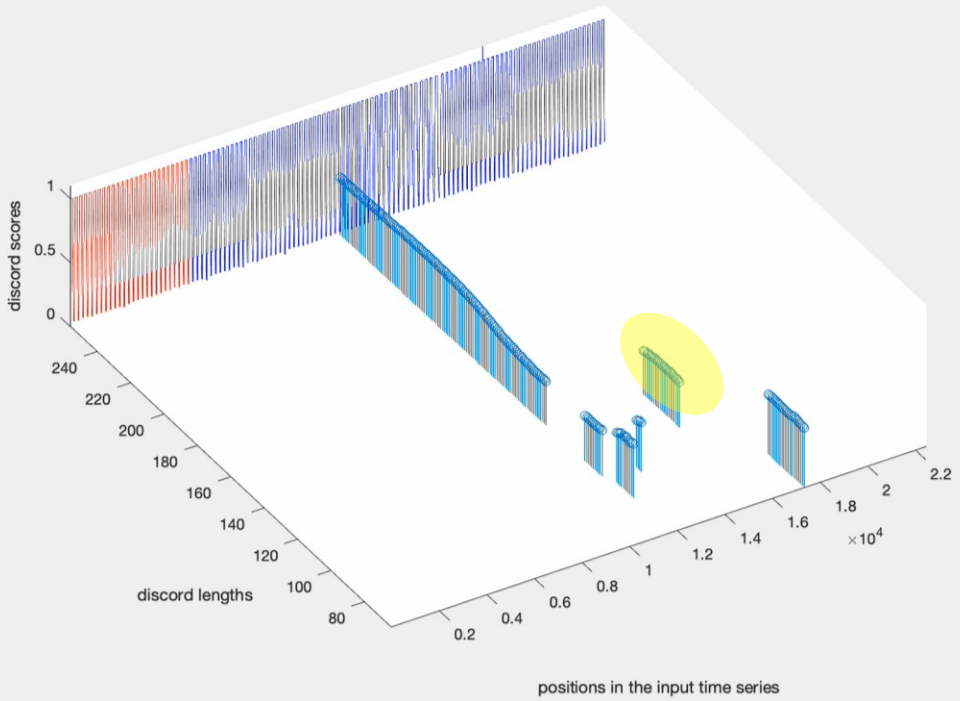


High Storage System Data for Energy Optimization

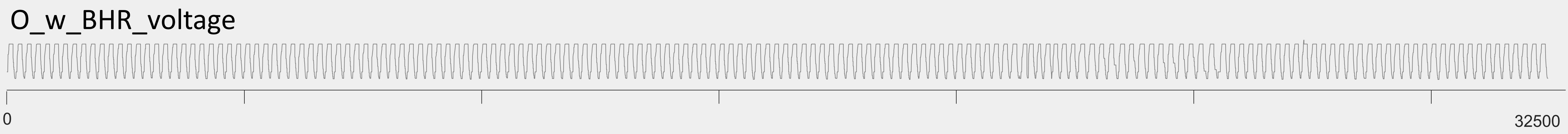
O_w_BHR_voltage

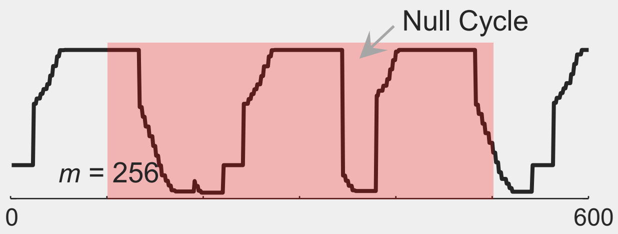


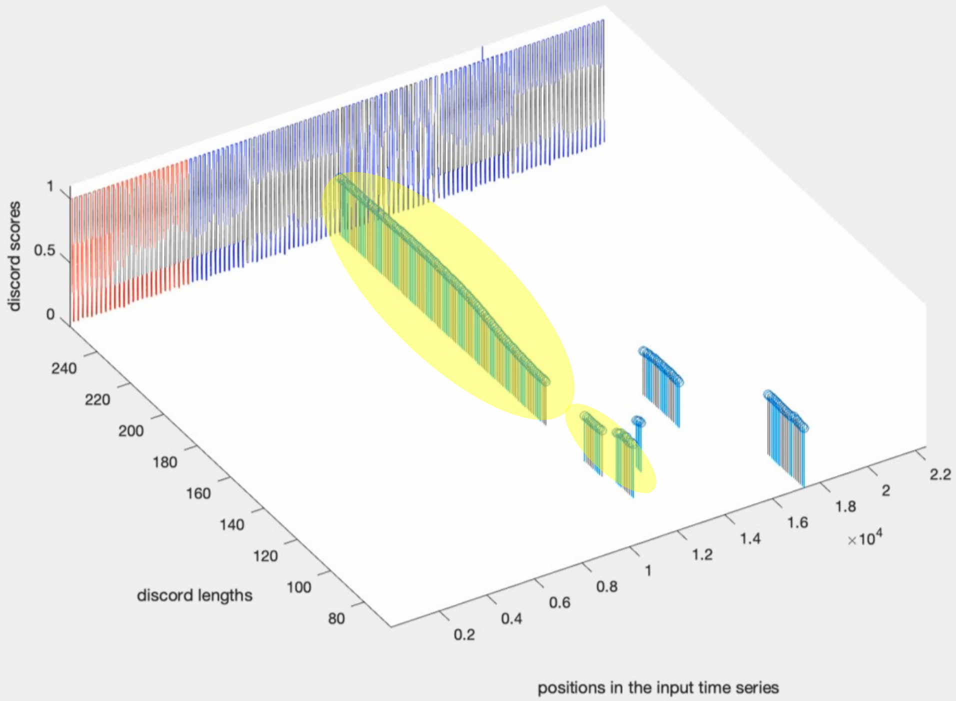


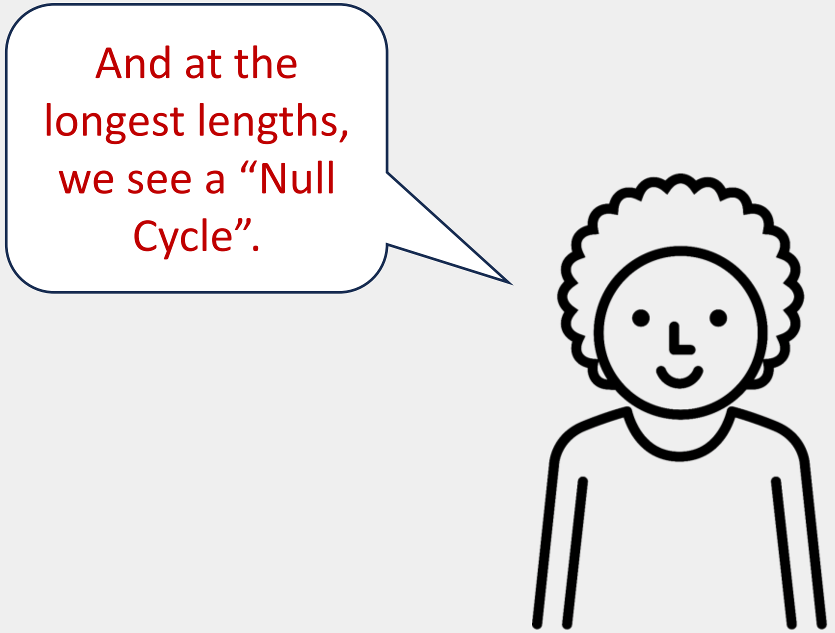


High Storage System Data for Energy Optimization

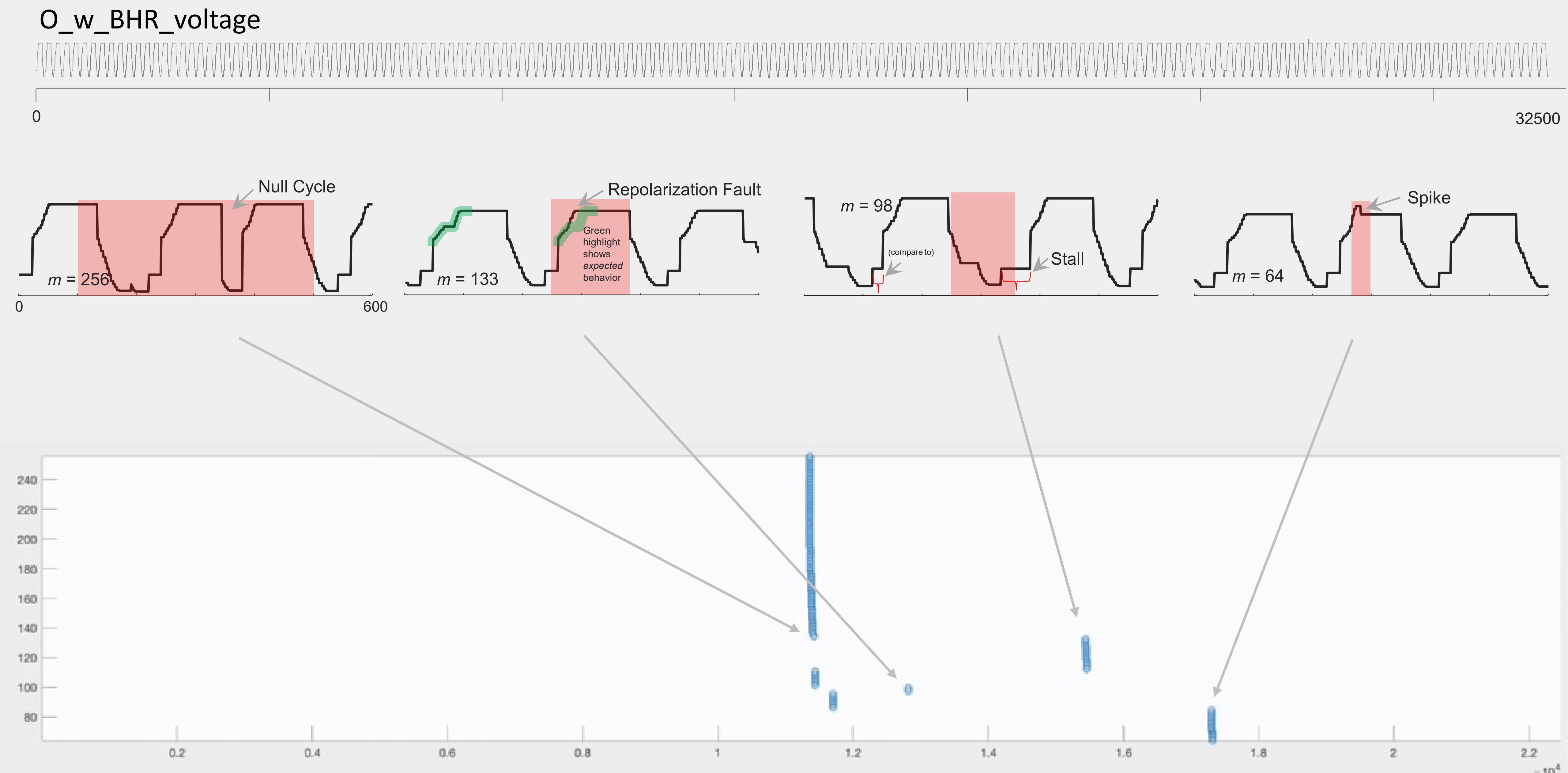




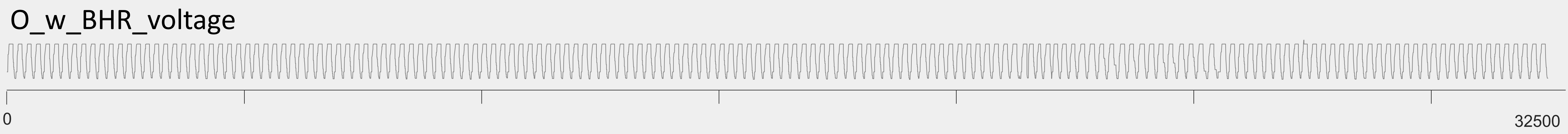


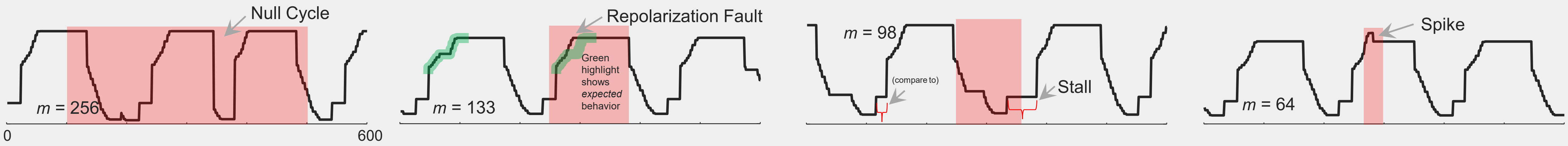


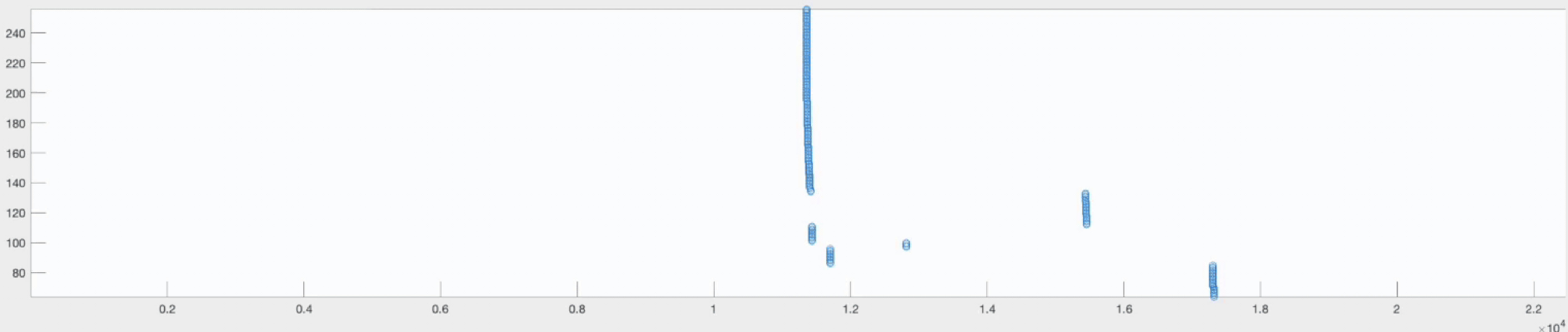
High Storage System Data for Energy Optimization



High Storage System Data for Energy Optimization







Amjid Chohan • 3rd+
Engineering Leader @ Amazon AWS - SDM
2w •

In this post we present a streaming time series anomaly detection algorithm on matrix profiles and left-discords with Apache Flink. We provide a working example that will run on our managed Apache Flink service using Amazon K Data Analytics (KDA).

aws

AWS BIG DATA BLOG

Real-time time series anomaly detection for streaming applications on Amazon Kinesis Data Analytics

Hugo Latapie · 1st
Principal Engineer at Cisco
Temecula, California, United States · Contact info

BTW we've managed to do some amazing work on large scale multivariate time series (millions) using matrix profile and a form of

Introduction

As one of the world's largest strea experience steady growth. We like that are essential to our success.

It's the reason Swhy the Tubi Data Science Team developed a new alerting system that detects outliers and trends in our Key Performance Indicators (KPIs). To detect them in Tubi KPIs, we utilized the Matrix Profiling (MP) algorithm.

Physiological Measurement

PAPER

Extraction of cardiac-related signals sensor during sleep

Luca Cerina[1,*], Gabriele B Papini[1,2], Pedro Fonseca[1,2] Johannes P van Dijk[1,3] and Rik Vullings[1]

[1] Electrical Engineering,Technische Universiteit Eindhoven, Eindhoven,
[2] Philips Research, Eindhoven, Noord Brabant, The Netherlands
[3] Center for Sleep Medicine, Kempenhaeghe Foundation, Heeze, Noord
[*] Author to whom any correspondence should be addressed.

E-mail: l.cerina@tue.nl, gabriele.papini@philips.com, pedro.fonseca@p kempenhaeghe.nl and r.vullings@tue.nl

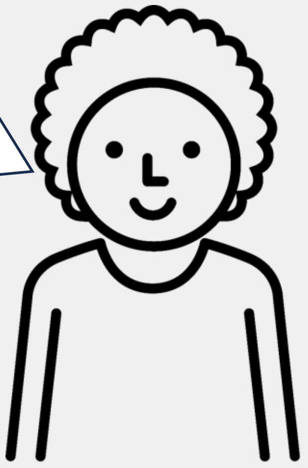
(we can) estimate the presence of an artifact (here called *discords*) using from a suprasternal pressure sensor...

Nin
Play
21h

I recently sta
#python wit

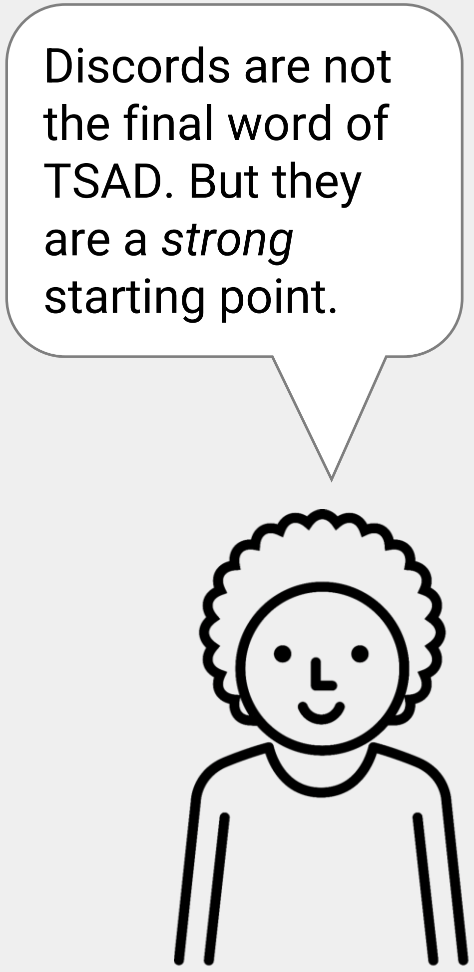
DAMP (Disc
#anomalyde
Keogh and
processing #
Processing)

While my implementation is still work-in-progress and there are rooms to enhance the code further, I can see how great DAMP is. In an attempt to reproduce one of



Discords as an Anomaly Detector *Summarized*

- Time Series Anomaly Detection (TSAD) is hot!
- I claimed that time series discords are SOTA
- Because:
  - They have zero or one parameter
  - They are blindingly fast (million, billion, trillion*)
  - Batch or Online with zero-lag
  - Can work with or without training data
  - Trivial to have *golden batch* or *amnesic versions**
  - The only TSAD algorithm to have been used by at least 100 teams to solve a real problem
  - You can incorporate domain knowledge^
  - Invariant to concept drift
  - Allows contract algorithms (explained below)
  - etc.



*www.cs.ucr.edu/~eamonn/DAMP_long_version.pdf
^www.cs.ucr.edu/~eamonn/guided-motif-KDD17-new-format-10-pages-v005.pdf

Mini Review II

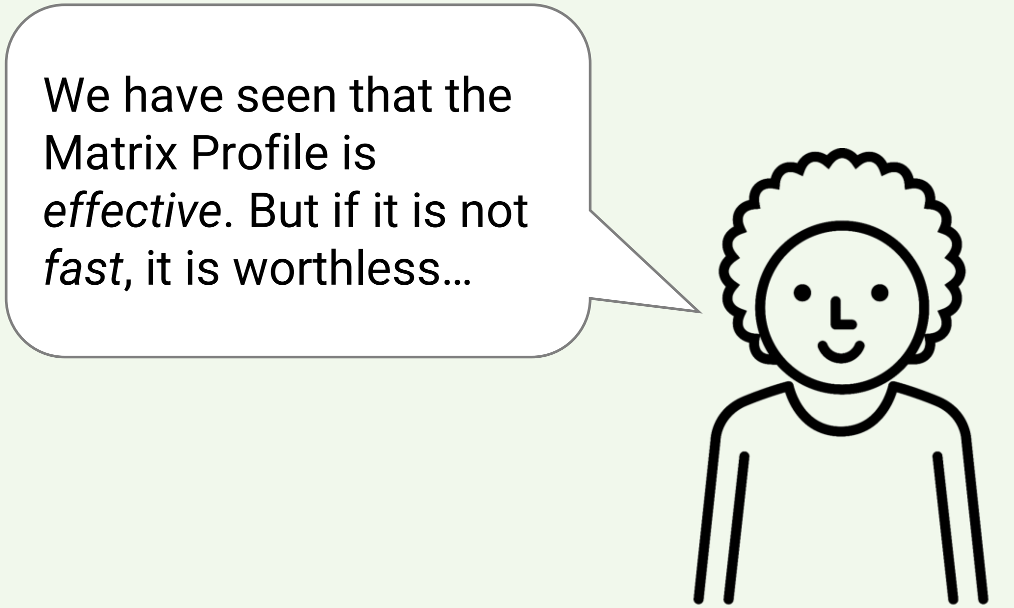
- With just the *Distance Profile* and the *Matrix Profile*, you can solve many (most/all) problems in time series data mining.
- Once installed on your machine, these are both one line of code!
- There is a large and growing community of Matrix Profile users, so these tools exist in most languages/platforms.

The *Matrix Profile* generalizes to multiple time series (i.e. joins) so you can ask questions that *compare* and *contrast* behaviors: (will not show this today)

- What patterns occur in *Males* **but not** *Females* (join *discord*)
- What patterns occur in *Japan* **and** *Ireland* (join *motif*)
- What patterns occur in Queen **and** *Vanilla Ice* (join *discord*) See Appendix

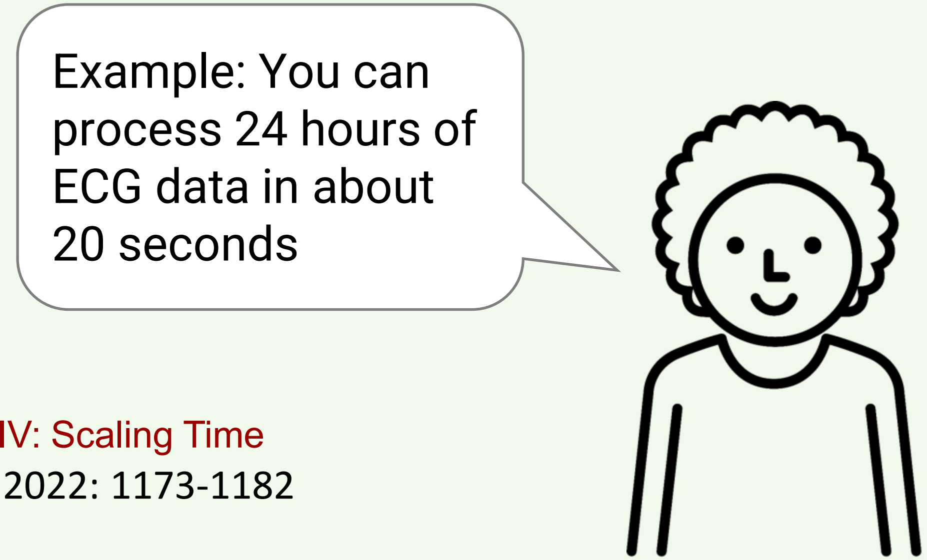
The *Matrix Profile* generalizes to other useful primitives, *Chains*, *Novelets*, *Shapelets*, *Platos, Snippets, FLOSS*…. (will not show these today)

Lets Talk about the *Speed* of the Matrix Profile I



Lets Talk about the *Speed* of the Matrix Profile I

- If you only want the *discords* (not motifs), then there is an ultrafast algorithm called DAMP that can process the data at over 100,000 Hz



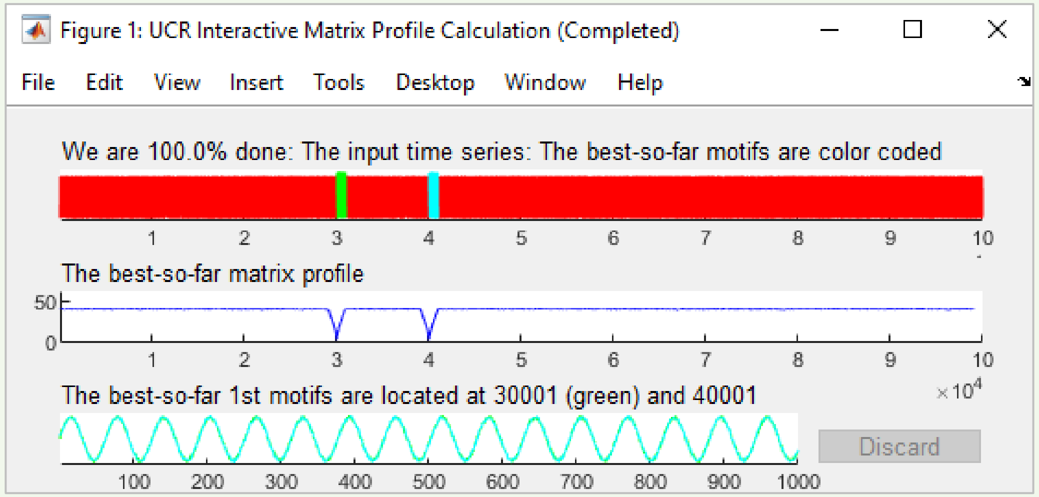
Yue Lu, Renjie Wu, Abdullah Mueen, Maria A. Zuluaga, Eamonn J. Keogh: Matrix Profile XXIV: Scaling Time Series Anomaly Detection to Trillions of Datapoints and Ultra-fast Arriving Data Streams. KDD 2022: 1173-1182
www.cs.ucr.edu/~eamonn/DAMP_long_version.pdf

Lets Talk about the *Speed* of the Matrix Profile II

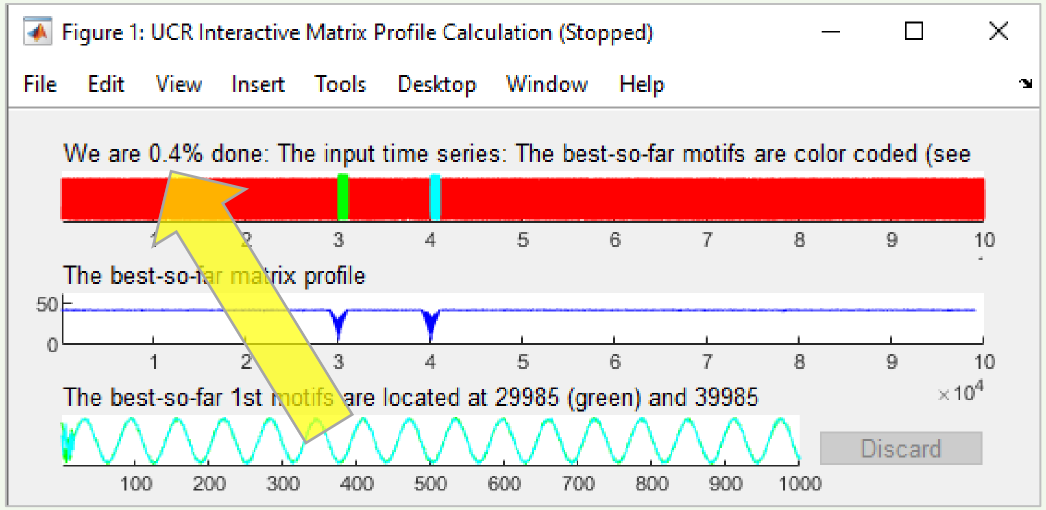
- If you also want the motifs…
- SCRIMP++ is a strongly *anytime* Matrix Profile algorithm.
- That means you can stop it anytime and get a very good approximation.

In the below, we stopped the algorithm as fast as we could click the mouse, and got the right result!

Full Convergence: SCRIMP++ takes 18.12 seconds



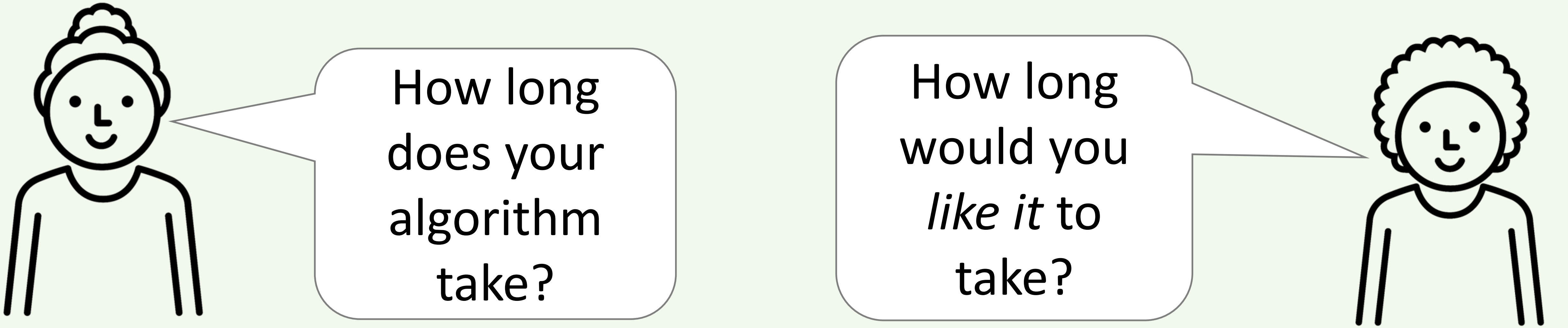
Anytime after about 0.078 seconds , or 0.4%



The time series (in red) is so noisy it looks solid, but it does have a subtle motif at 30,000 and 40,000, discovered by SCRIMP++

Let's Talk about the *Speed* of the Matrix Profile III

- Both MASS and the Matrix Profile allow *perfect* contract algorithms.
- They are essentially unique in offering this feature.
- This is one of the most useful, yet underused features of these algorithms, so let's spend a few minutes to explore this.



Contract Algorithms

- A contract algorithm is an algorithm you can instruct to finish in an arbitrary amount of time.

  (related to, but distinct from *anytime* algorithms).

- Thus, you can tell the MP something like "*Process this massive dataset, but you have to give me your best answer in 45 seconds*".
- The algorithm will then give you its best answer in *exactly* 45 seconds.

- Note that you *could* just use the anytime algorithm property of the MP to stop at 45 seconds. But the contract algorithm's result will be better. You get a "reward" for letting the algorithm know in advance when it must stop.
- Also note that the contract versions are *still* anytime algorithms!!

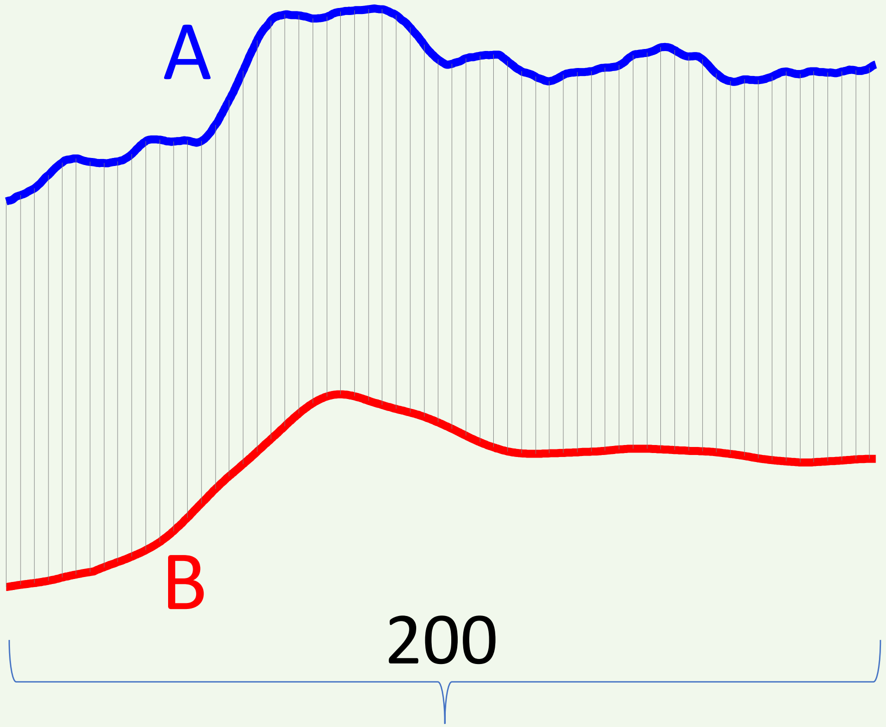
Contract Algorithms

- **Background Fact:** The time needed for MASS or MP is independent of the data and can be predicted perfectly ahead of time.

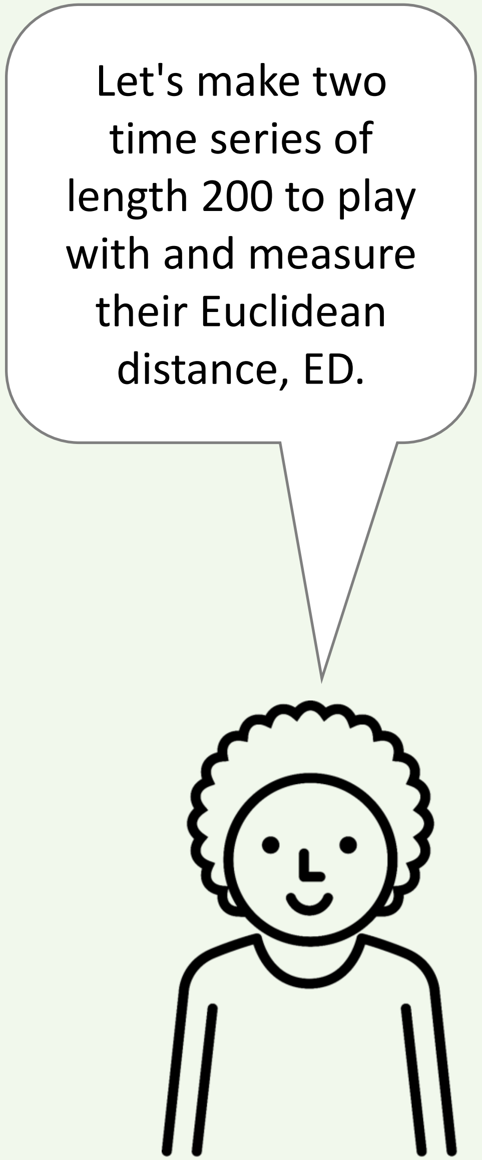
    (As an aside that is a very rare property for algorithms in this space)

- To predict the time needed on given hardware, you just need to have done one calibration run (on any length dataset).

ED(A,B)= 55.8

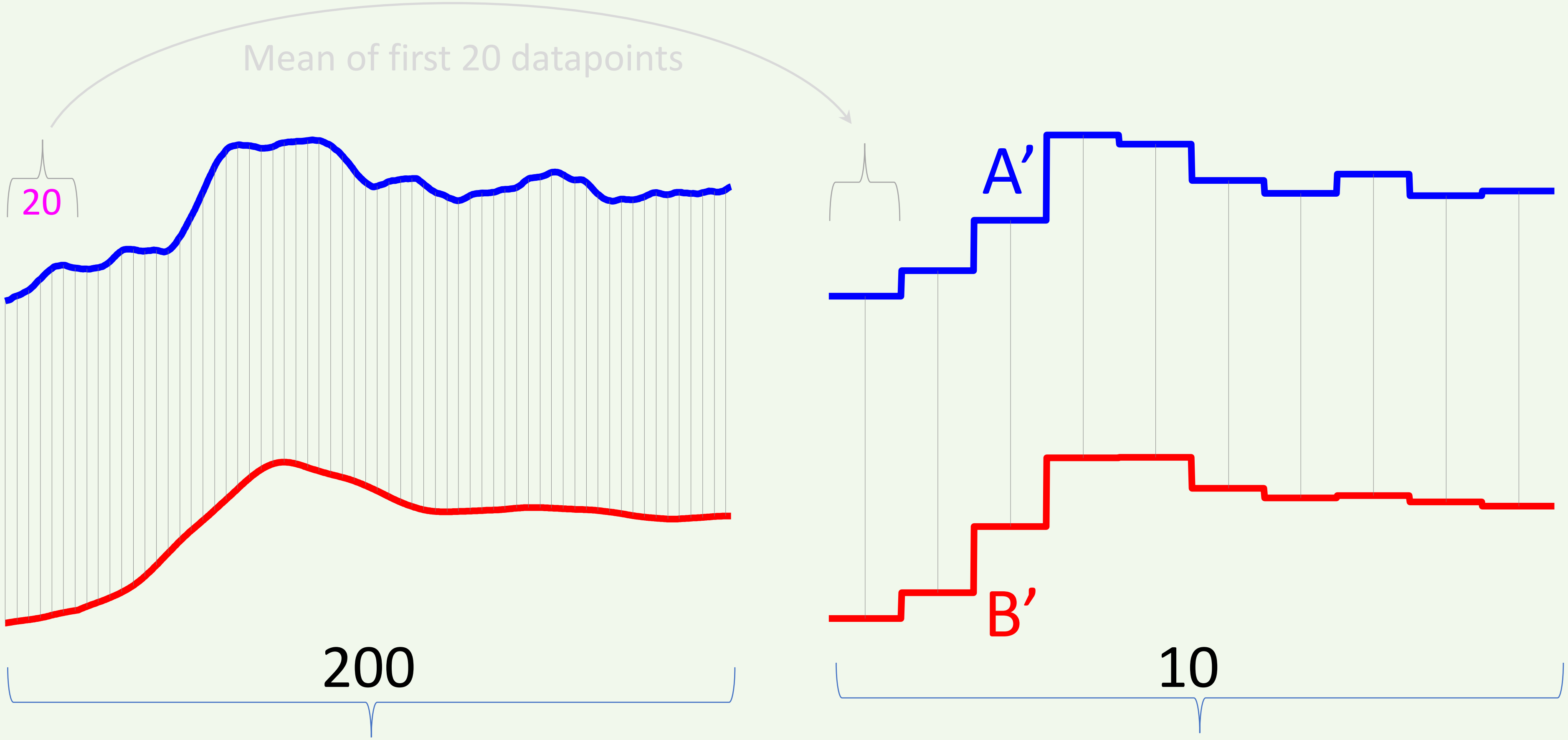


I am showing these datasets shifted apart in the Y-axis for clarity, but they are z-normalized before they are compared.

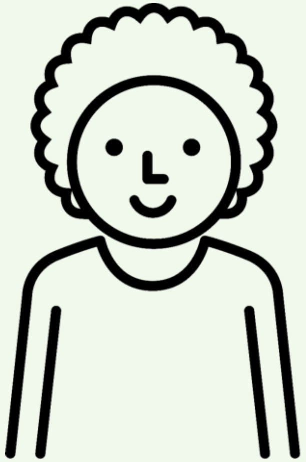


Compress to 1 in 20 using PAA

ED(A,B)= 55.8



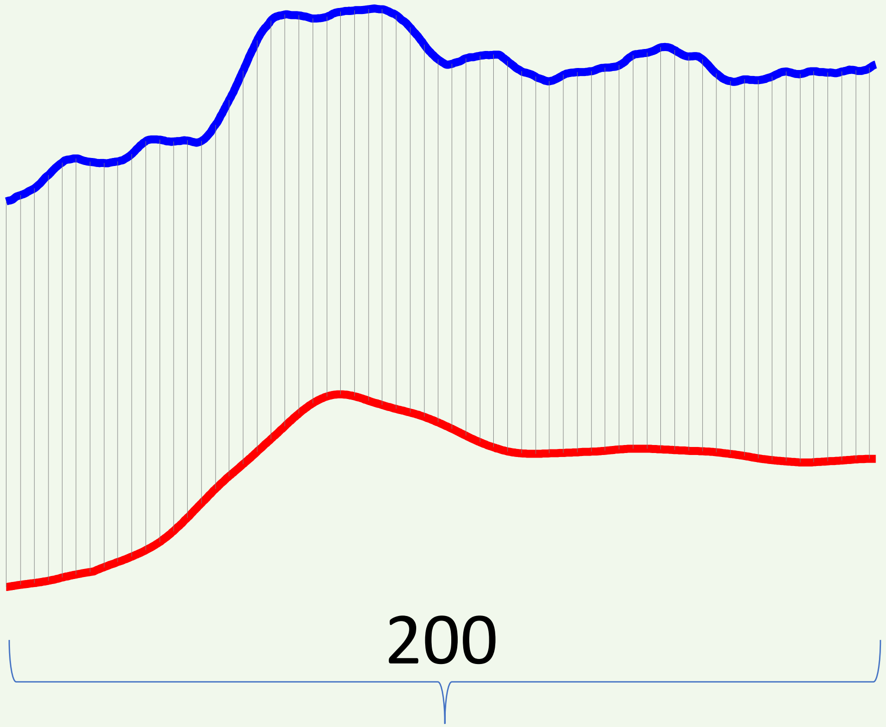
We can make downsampled versions of our time series, using PAA.
This just replaces regions of time series (here of size 20) with their mean value.

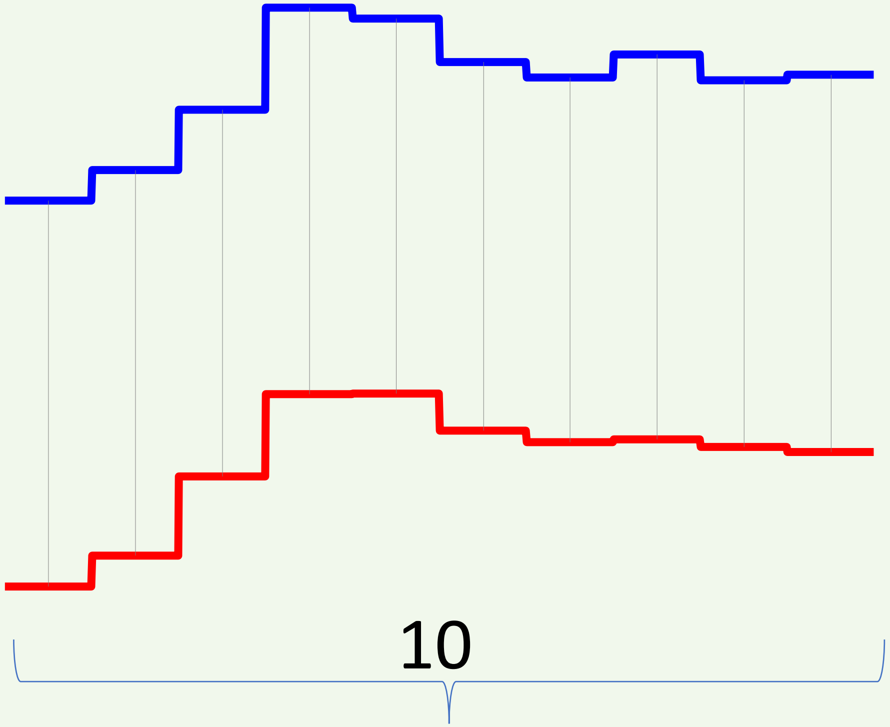


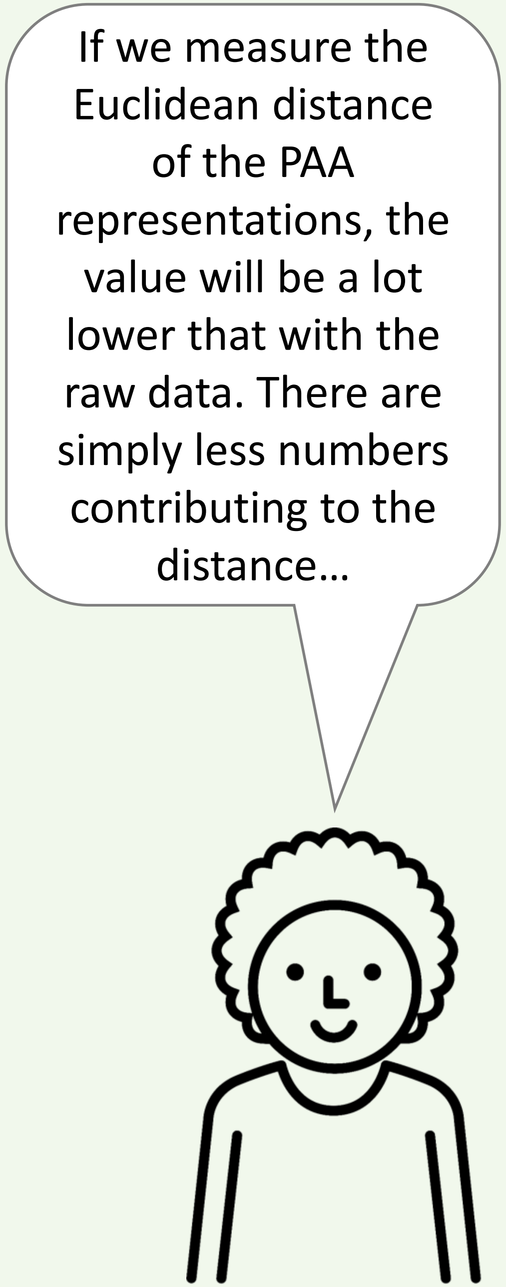
Compress to 1 in 20 using PAA

ED(A',B') = 12.21

ED(A,B)= 55.8





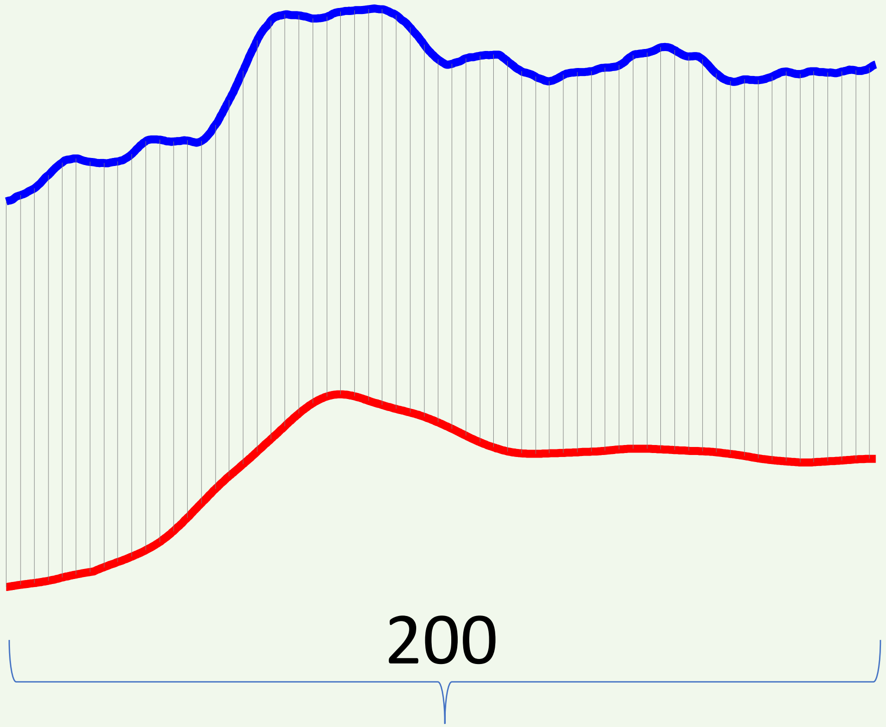


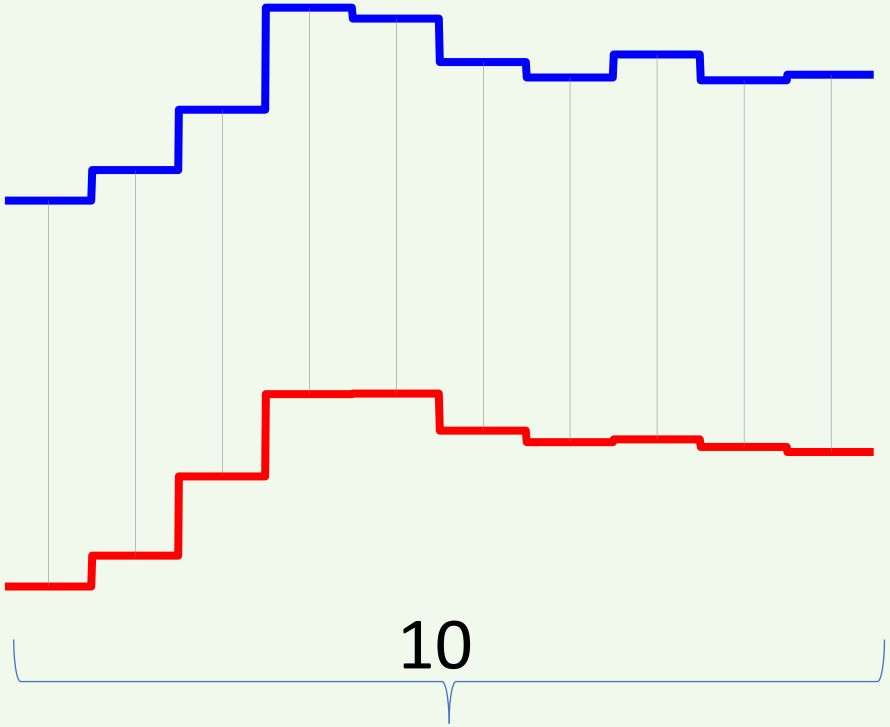
Compress to 1 in 20 using PAA

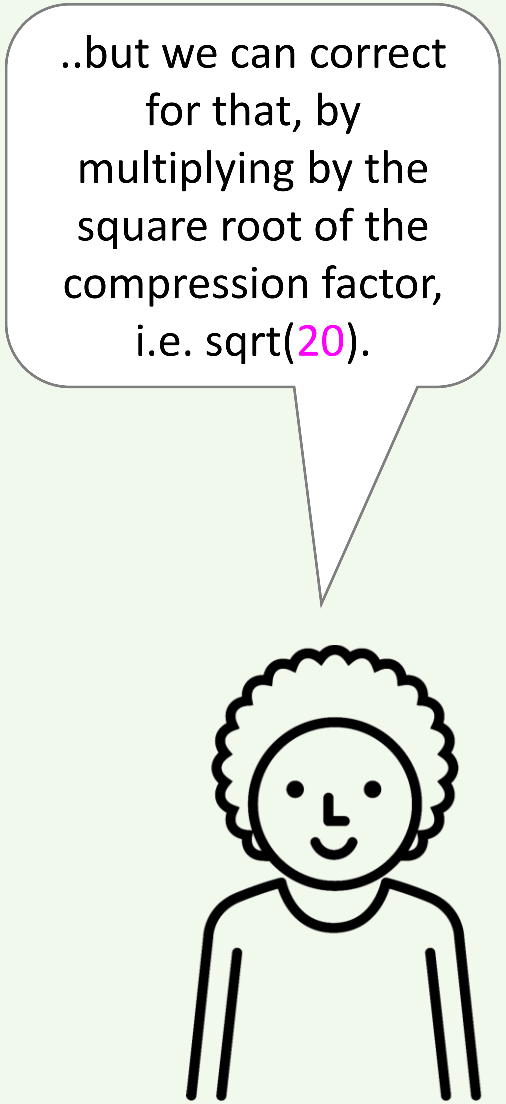
ED(A',B') = 12.21

ED(A,B)= 55.8

ED(A',B') * sqrt(20) = 54.5







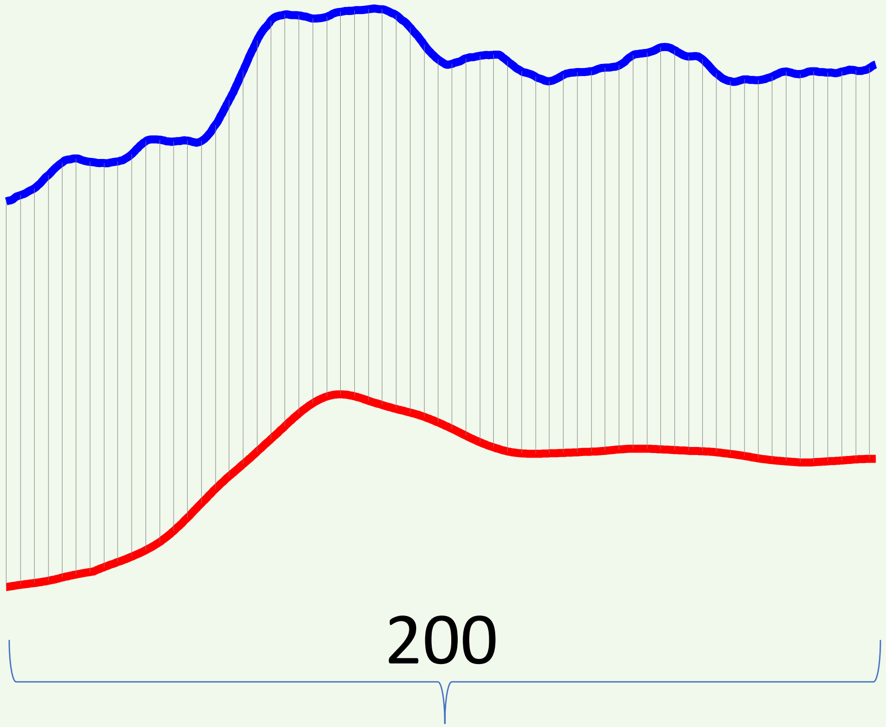
Compress to 1 in 20 using PAA

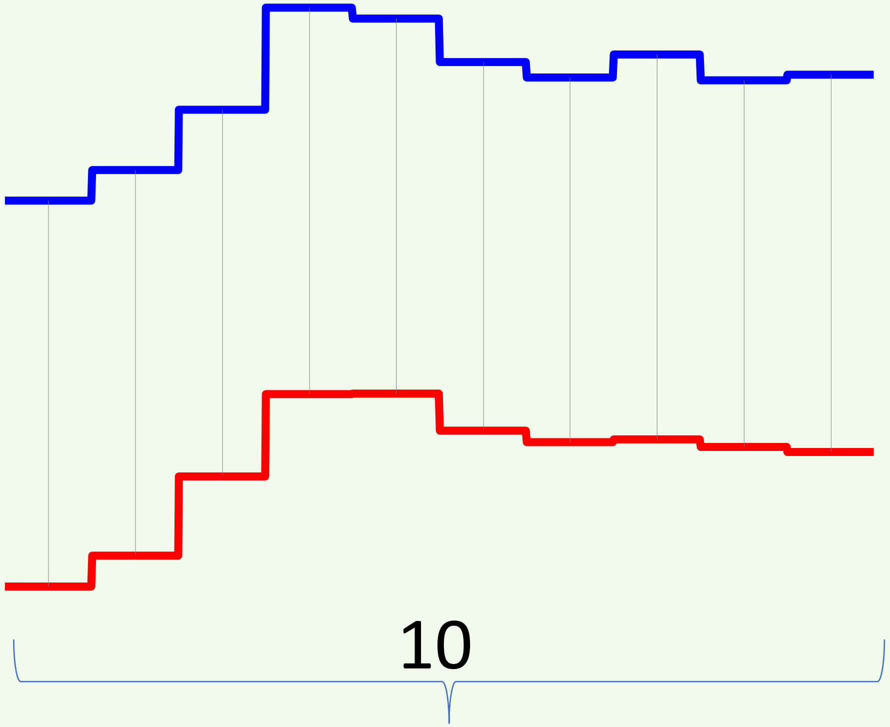
ED(A’,B’) = 12.21

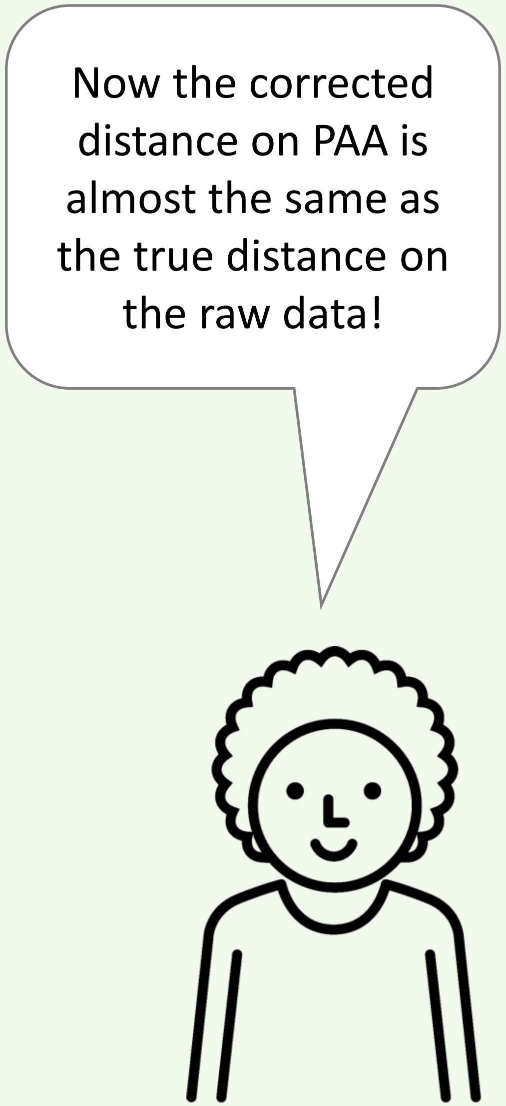
ED(A,B)= 55.8

ED(A’,B’) * sqrt(20) = 54.5

Note that we have 55.8 ≈ 54.5







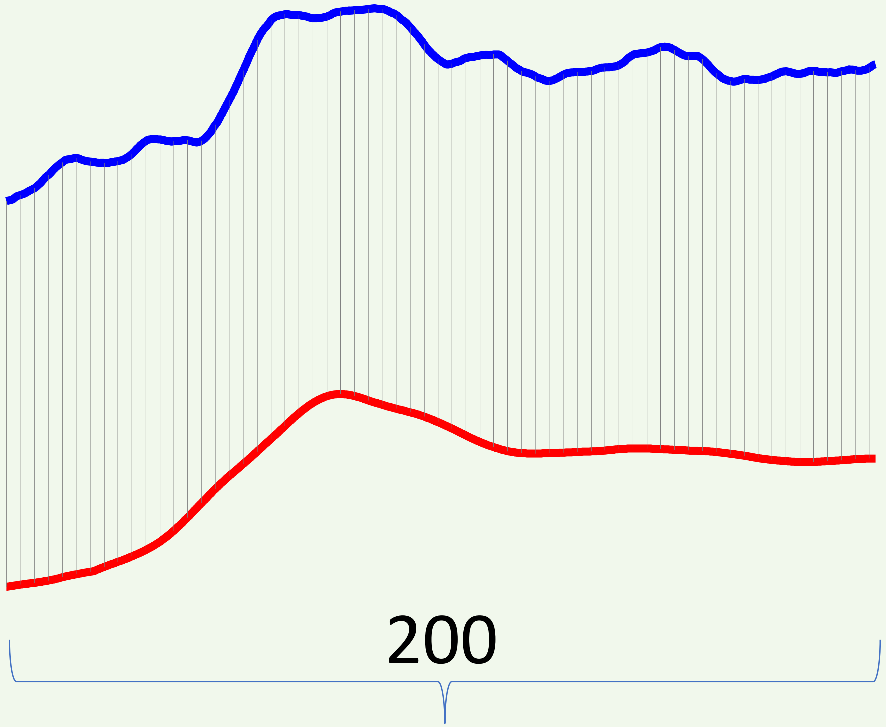
Compress to 1 in 20 using PAA

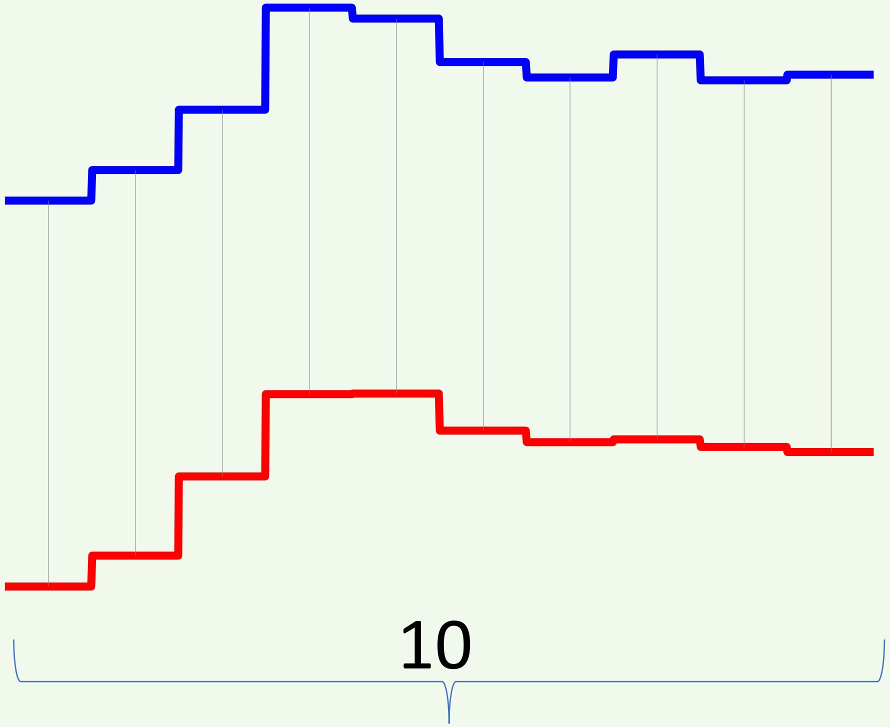
ED(A’,B’) = 12.21

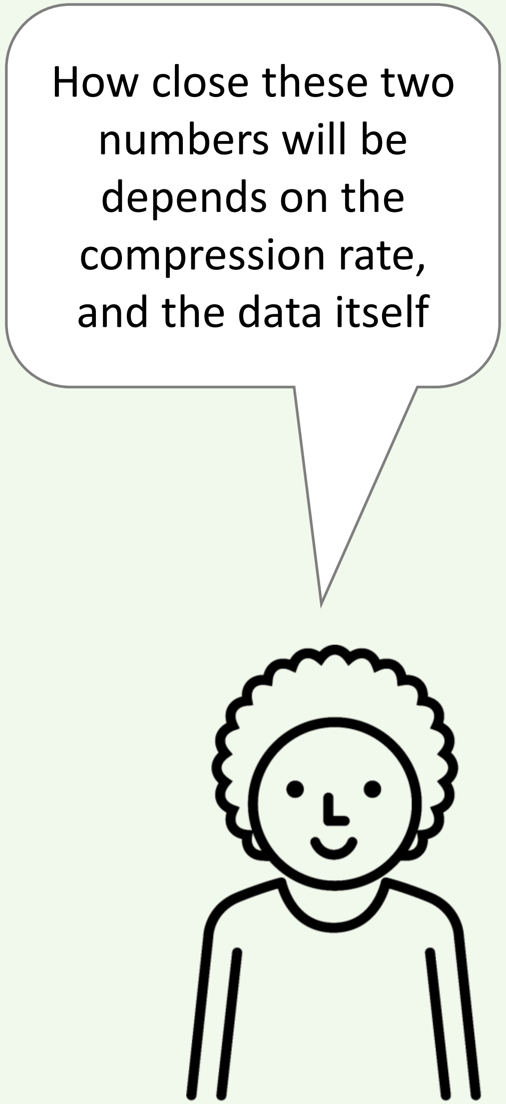
ED(A,B)= 55.8

ED(A’,B’) * sqrt(20) = 54.5

Note that we have 55.8 ≈ 54.5







Compress to 1 in C using PAA

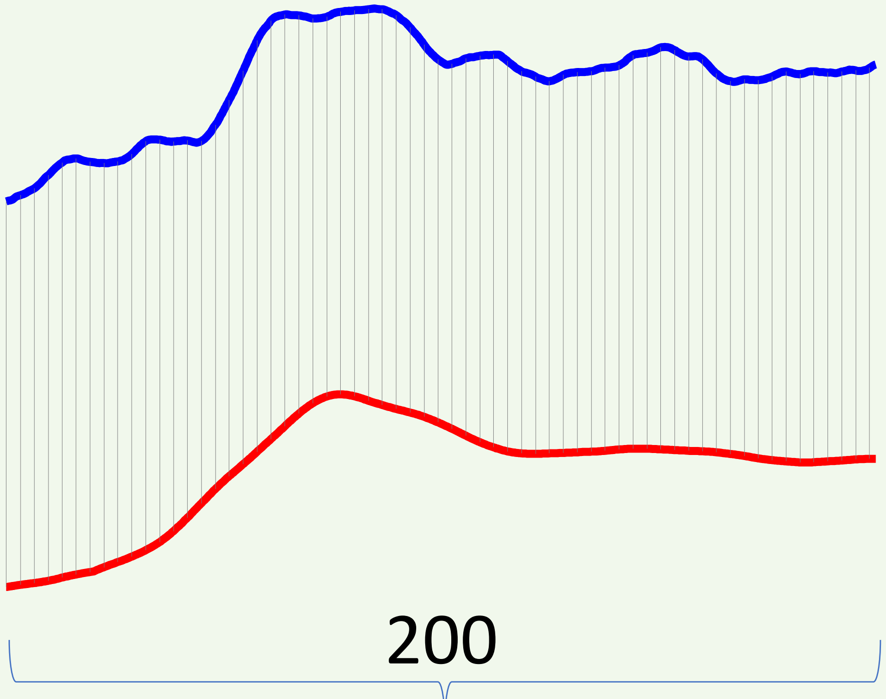
ED(A',B') = 12.21

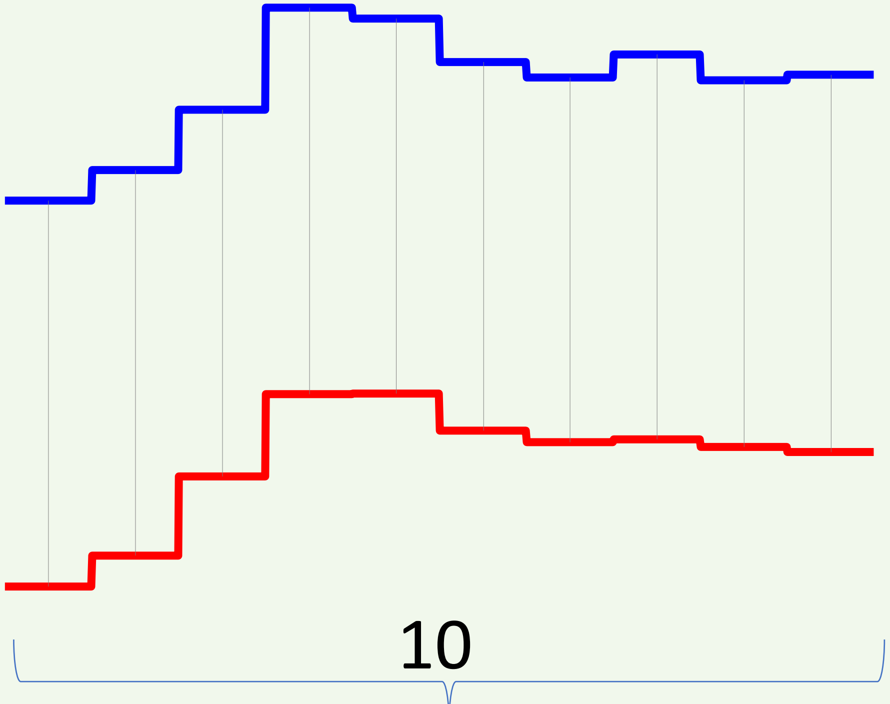
ED(A,B)= 55.8

ED(A',B') * sqrt(20) = 54.5

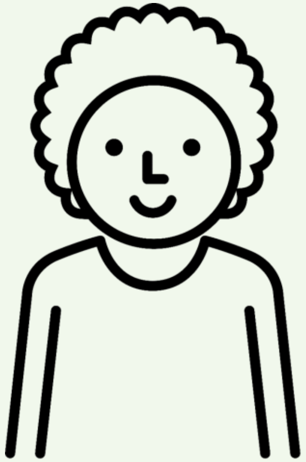
Note that we have 55.8 ≈ 54.5

Actually, we have: ∀A,B [ED(A',B') * sqrt(C)] ≤ ED(A,B)



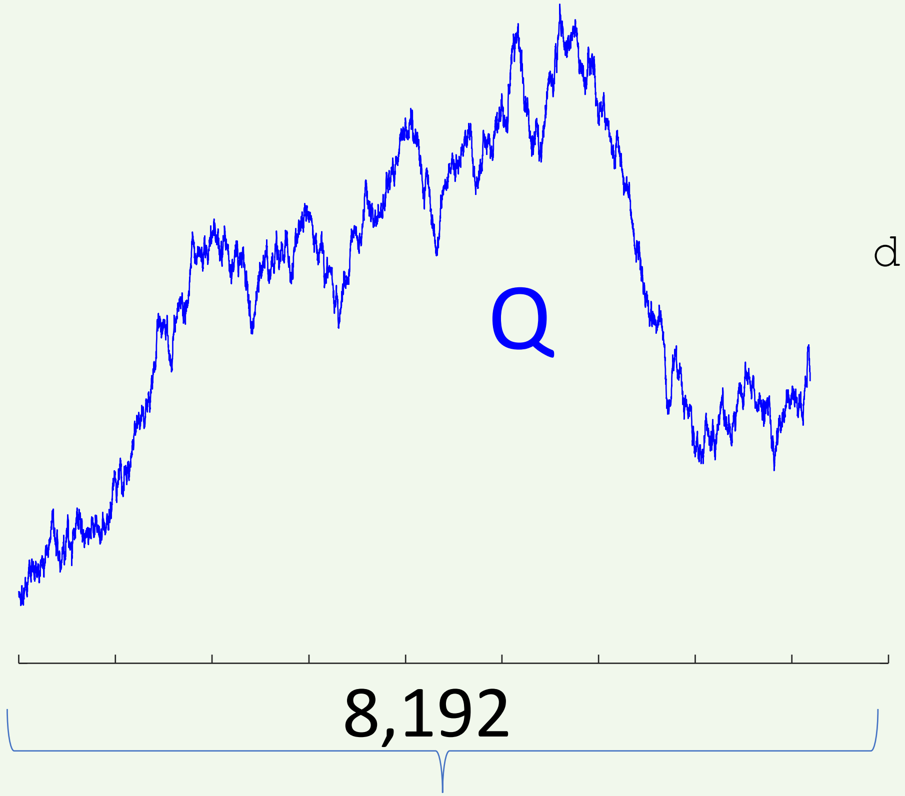


In fact, we have this nice **lower bounding** property*, which can be *very* useful, but we will not exploit it today.

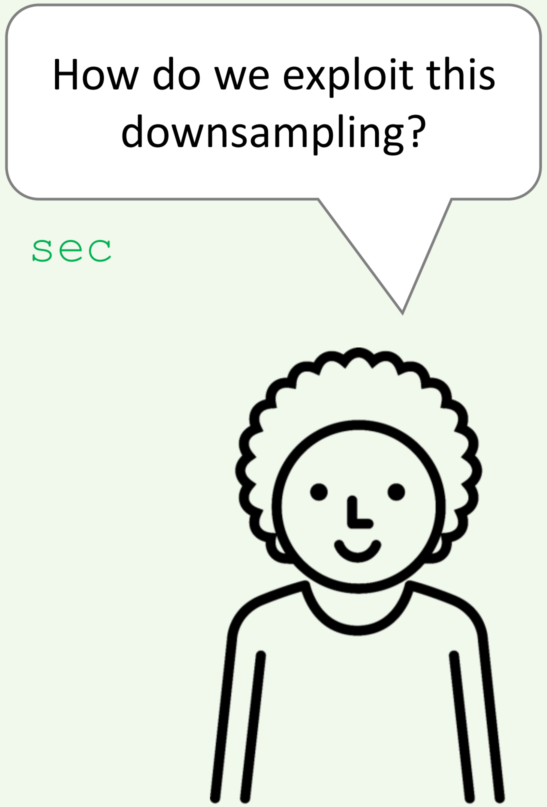


*Keogh, et al. Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases. Knowl. Inf. Syst. 3(3): 263-286 (2001)

Suppose we want to search for Q in a very long dataset D of 67,108,864
We can predict it will take 5.1 seconds.
However, I don't have that much time, I only have 0.1 sec.




```
dist = MASS(D,Q); % This will take 5.1 sec
```



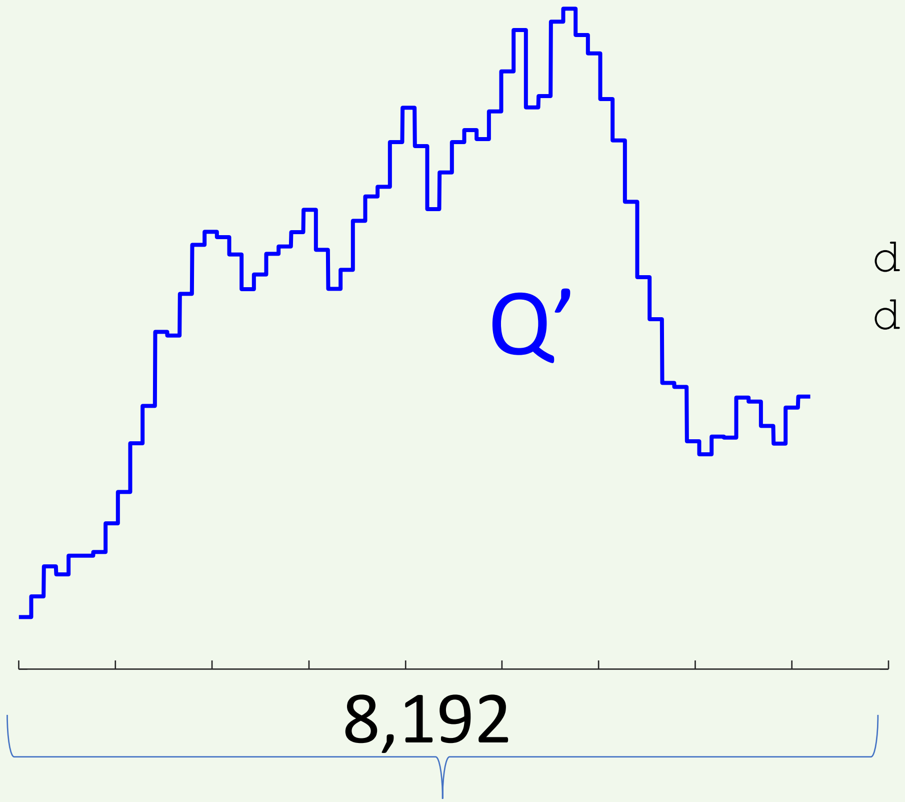
Suppose we want to search for Q in a very long dataset D of 67,108,864

We can predict it will take 5.1 seconds.

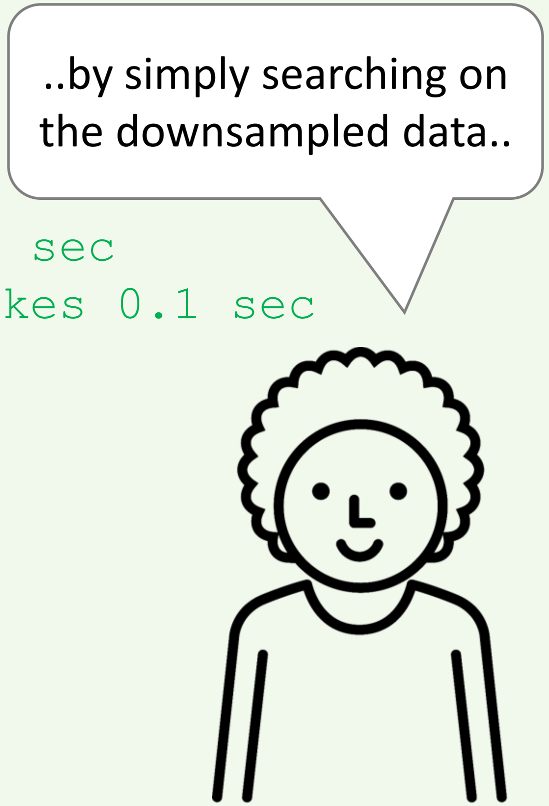
However, I don't have that much time, I only have 0.1 sec.

I calculate that if I compress to 1 in 64 using PAA, it *will* take 0.1 sec.

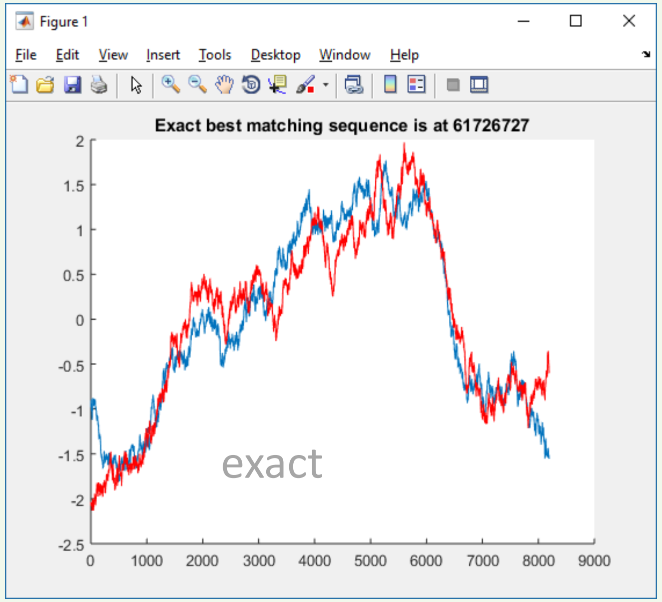
So, I downsample both Q and D by 1 in 64 using PAA, and run my search.



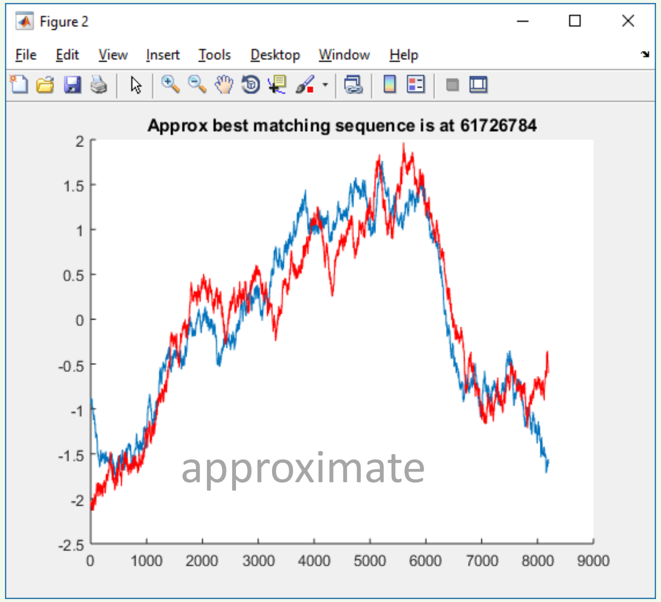

```
dist = MASS(D,Q); % This will take 5.1 sec
dist = MASS(PAA(D,64),PAA(Q,64)); % Takes 0.1 sec
```



How well did this work?



If we had do the full search, it takes 5.2 seconds, and here is the *exact* answer.



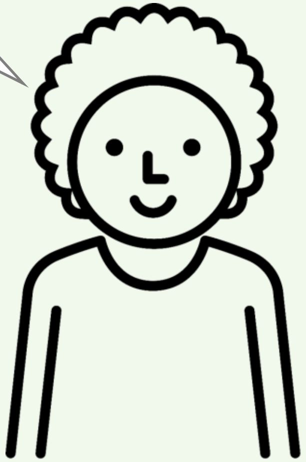
If we do a contract algorithm search, it takes 0.1009 seconds, and here is the answer, after we project the downsampled result back to the original raw data.

In this case we get almost exactly the same answer!! The result is just shifted a few datapoints.

And our prediction of *how long* it would take was within 1%

How well this works depends on your dataset, but most datasets are highly oversampled.

For example: Many ECG datasets are sampled at 4096 Hz. But if you downsample them to 128 to 256 Hz, it make almost no difference.



This is cool!

Suppose you had to process the ECGs of 500 patients in exactly one minute…

You could finish in *exactly* one minute and use *all* of that minute's CPU resource!

- You could prioritize arbitrarily, for example:
  - **Out-of-Band**: Spend three times as much CPU time on people over the age of 60.
  - **In-Band:** Spend 0.1 seconds on all the patients. Then, look at the results and then spend one second each on the ten worse patients (by some measure).
  - **In-Band:** Spend X seconds per hour-of-ECG recording (Because spending 0.1 seconds on each patients might penalize patients with longer recordings. It might be fairer to spend time proportion to the ECGs length)
  - **Or..**

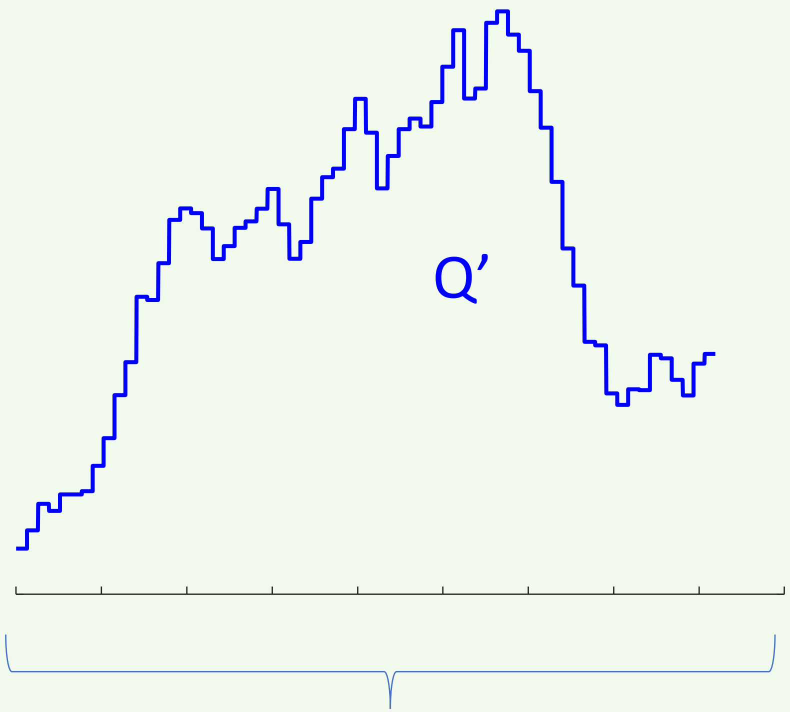
Moving from MASS to Matrix Profile

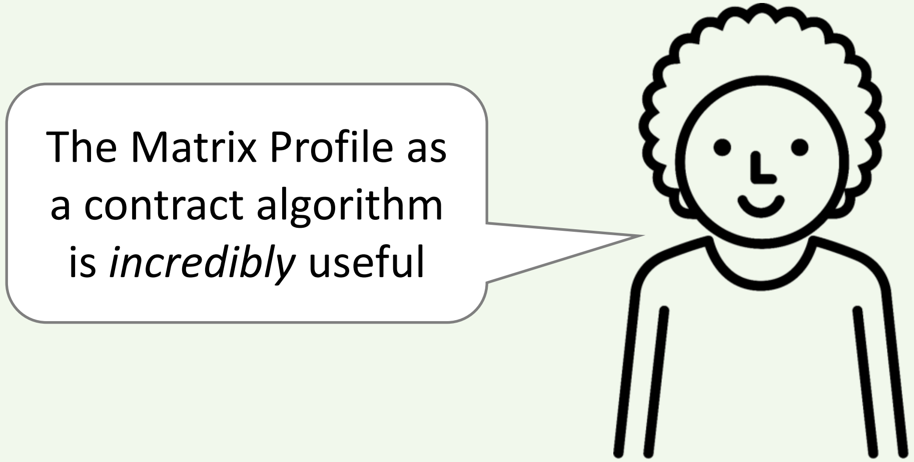
The MASS algorithm is so fast, that we rarely need to us the contract algorithm version.

It is with the Matrix Profile that the contract algorithm version becomes incredibly useful.

With the Matrix Profile, the speedup obtained is *quadratic* in the compression,

So, for example, if we downsampled a dataset 1 in 64 like the example below, we would get a speedup of about 4,096 times!!





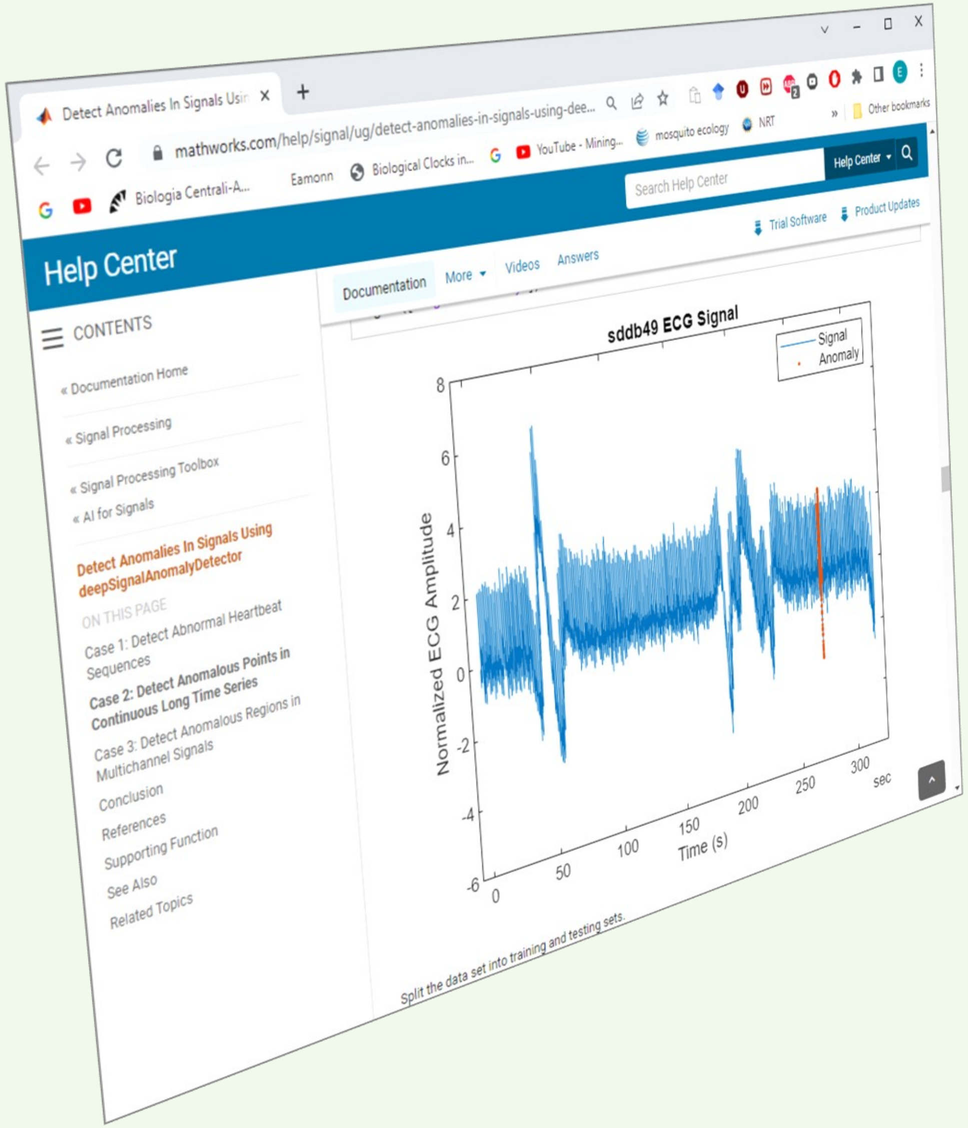
This is really cool!

Suppose you had to search eight massive datasets, for the one that has the best pair of motifs, in one minute.

You could finish in *exactly* one minute and use *all* of that minute's CPU resource!

- You could search with some clever algorithm you invent, for example:
  - Spend 2 seconds on all datasets, then spend 4 seconds on the best half, then spend 8 seconds on the best half of *that* subset, then spend the last 12 seconds on the best of the final two. (2*8) + (4*4) + (8 *2) + 12 = 60 seconds
  - Just give each dataset 60/8 seconds.
  - Give each 1 second, but then based on…

Contract Matrix Profile on MATLAB's ECG Example



Let's find the top discord on this dataset.

We are not sure how long the subsequence should be, so let's try *every* length from 128 (1/2 second) to 1024 (4 seconds)!

Testing all 896 possible lengths might be very slow..

https://www.mathworks.com/help/signal/ug/detect-anomalies-in-signals-using-deep-learning.html

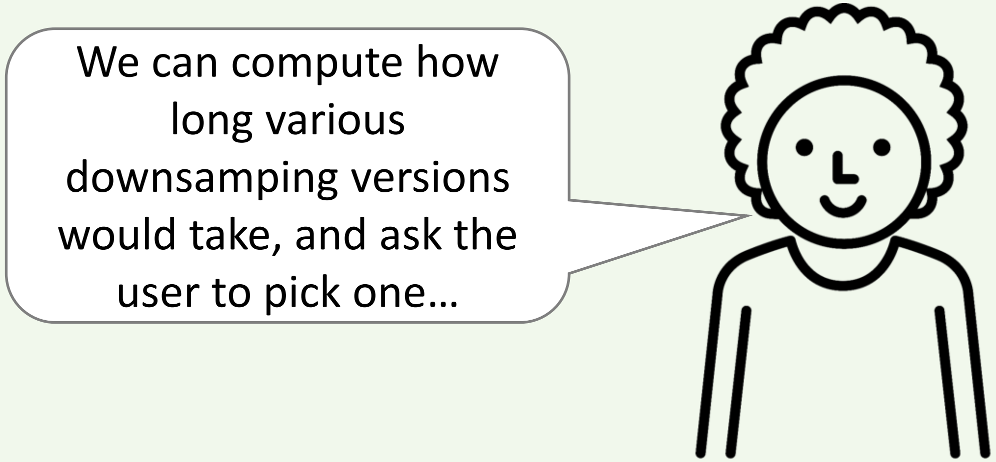
Contract Matrix Profile on MATLAB's ECG Example

```
>> HowLongWillItTakeMP(MATLAB_ECG,128,1024)

   1) Predicted execution time for 1 to 16: 3.91 seconds
   2) Predicted execution time for 1 to 8:  32.2 seconds
   3) Predicted execution time for 1 to 4:  258 seconds
   4) Predicted execution time for 1 to 2:  2062 seconds
   5) Predicted execution time for 1 to 1:  16502 seconds
   6) Quit
   Please pick an option:
```

I show *powers of two* compression factors, but they don't have to be *powers of two,* in fact, they don't even have to be integers.

The absolute time looks slow, but..
That is for 896 runs of the Matrix Profile
It is on my cheap old laptop ;-)

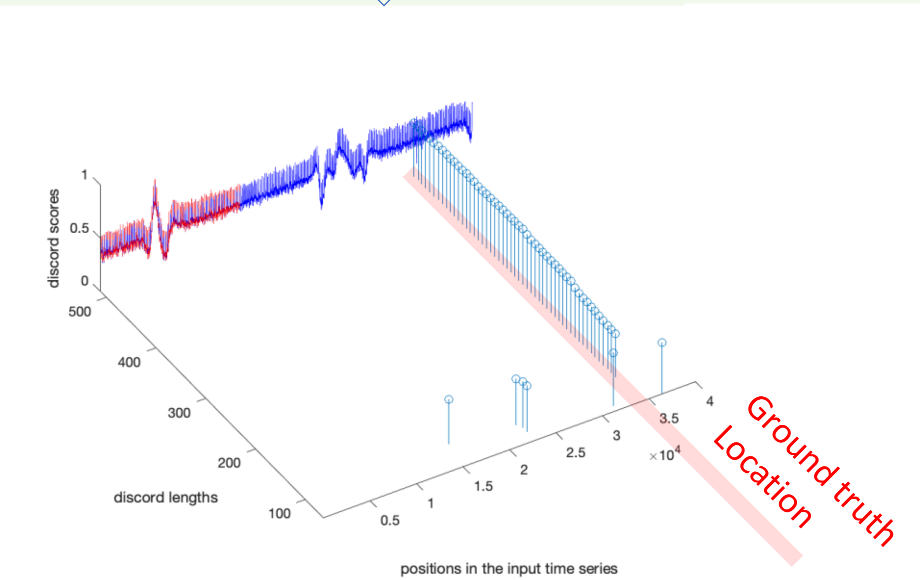


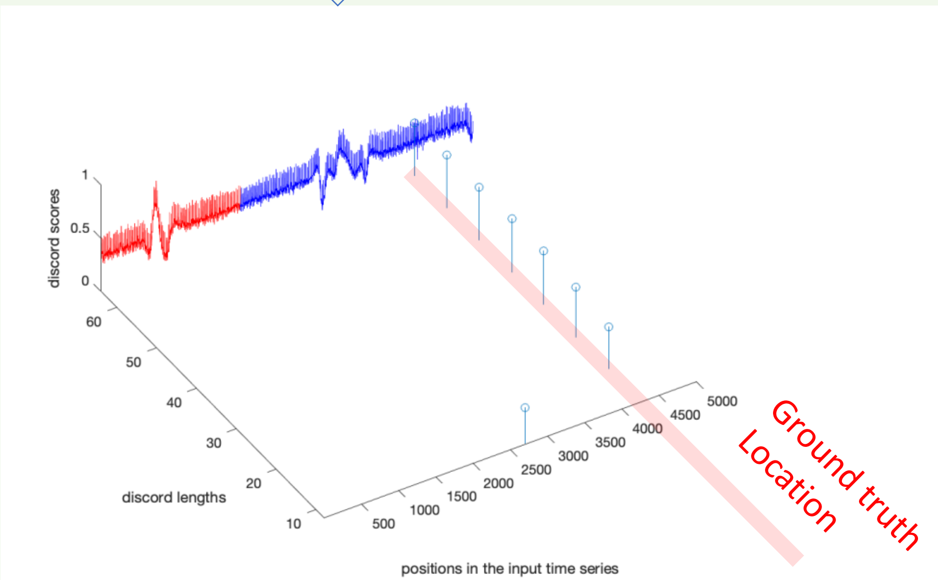
Contract Matrix Profile on MATLAB's ECG Example

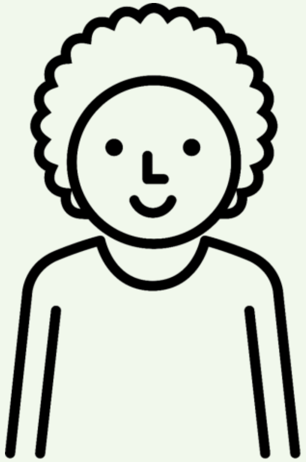
If the user chooses 1 to 2, then in 2065 seconds…

If the user chooses 1 to 16, then in 3.90 seconds…

If I was working on this dataset I would probably start with 1 in 16.
In this case, I would probably end there.
However, If the results were more ambiguous, I would go to a finer resolution, say 1 in 8…







Since TSAD is such a hot topic

- I don't want to end the tutorial on what could be perceived as a negative note ;-)

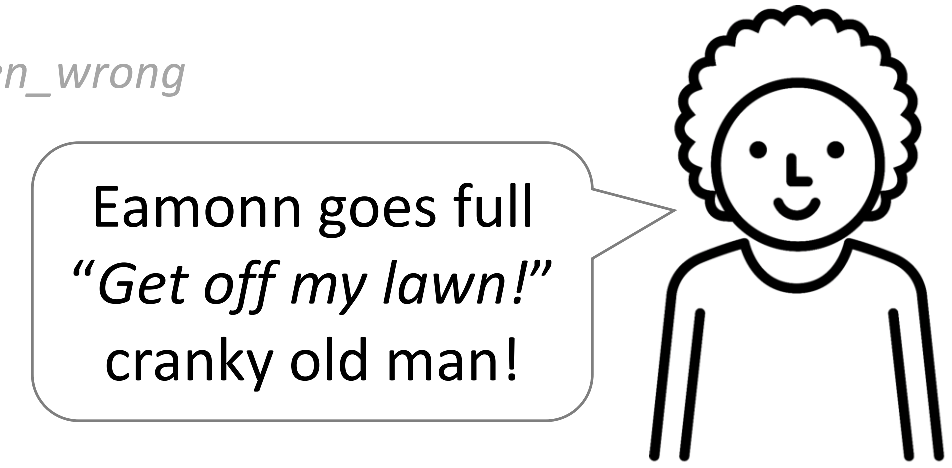
- However, TSAD is a hot topic right now.

- There are dozens of papers on TSAD at the top conferences each year..

- But virtually all of them are simply *wrong*.

- Some are *not even wrong* *en.wikipedia.org/wiki/Not_even_wrong*

- I think is worth seeing why that is the case..



Problems with Time Series Anomaly Detection

- In many cases, it is *not* Time Series Anomaly Detection!

Even if we ignore the above..

- In most datasets considered, a seventy-year-old algorithm is superior.

Even if we ignore the above..

- In most cases, the scoring functions used greatly inflate performance.

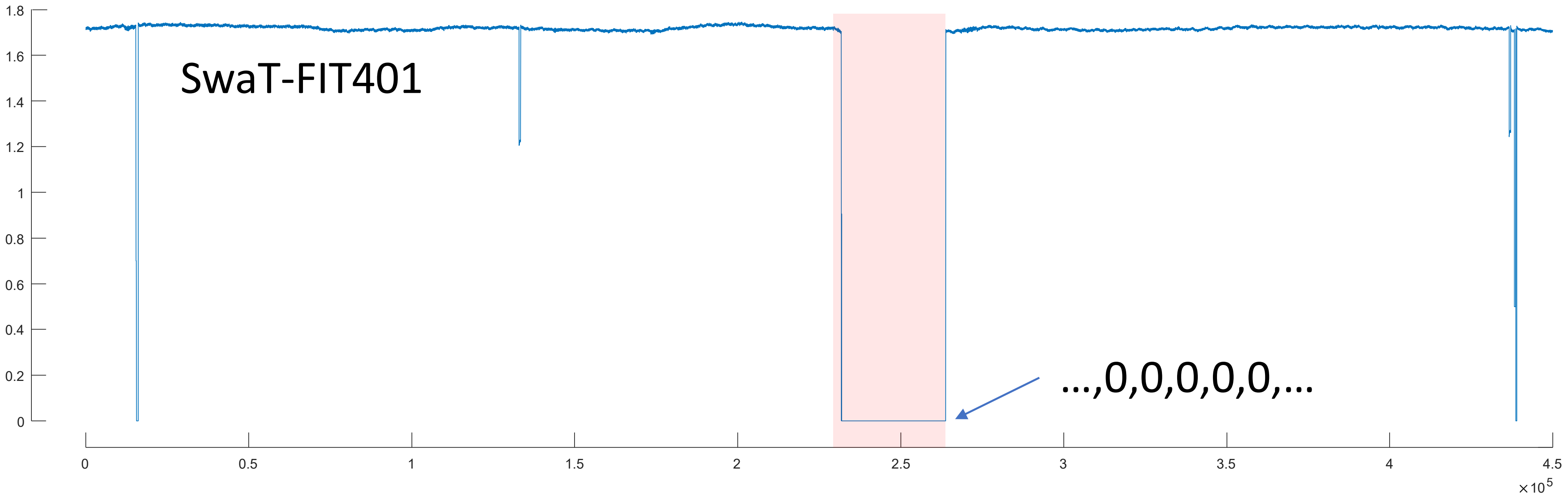
Even if we ignore the above..

- In most cases, any high precision claims are *necessarily* wrong.

Let's see all of this with an example from one of the most cited datasets, SWaT, which has 400+ citations.

A dataset to support research in the design of secure water treatment systems. J Goh, S Adepu, KN Junejo, A Mathur - Critical Information Infrastructures 2017

It is **not** Time Series Anomaly Detection



Most domains have a special character for invalid data. For example:

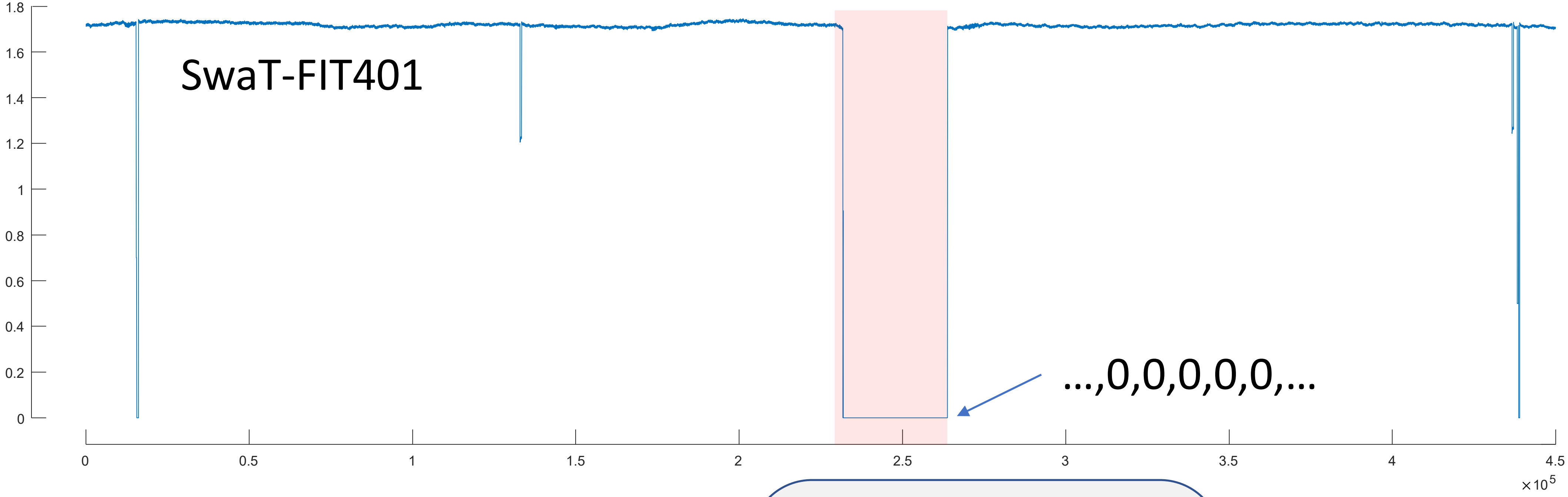
- `…,NaN, NaN, NaN,...`
- `…,-9999,-9999,-9999,…` (AspenTech)
- `…,0,0,0,0,…` (where zero is a physically impossible value)
- `…,inf,inf,inf,…`

But you do not need to do any anomaly detection/machine learning to detect these!

You *know* about these *before* you see any data. About 70% of the anomalies in SWaT are of this type!

This is *not* time series anomaly detection, this is basic data processing.

It is **not** Time Series Anomaly Detection



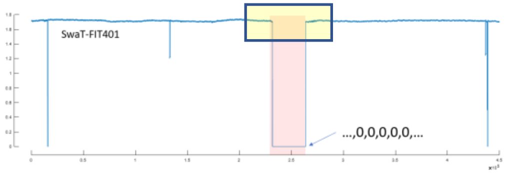
No one working in the real world would every consider this an anomaly. It is a trivial "pipeline" issue that can deal with using a "simple rule"* that you can define before you see the data.

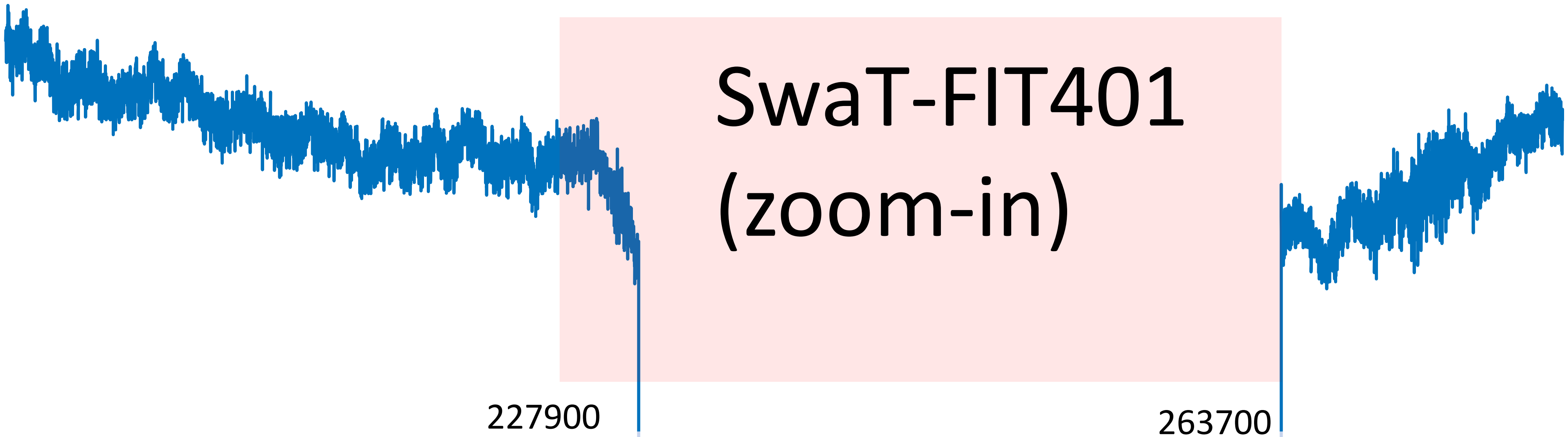
.. Varying signal becomes flat line.. ..we can set a simple rule to detect to detect the "anomaly"

Dhaval Patel
IBM TSAD expert

*Toolkit for Time Series Anomaly Detection. Patel et al. SIGKDD 2022

A seventy-year-old algorithm is often superior



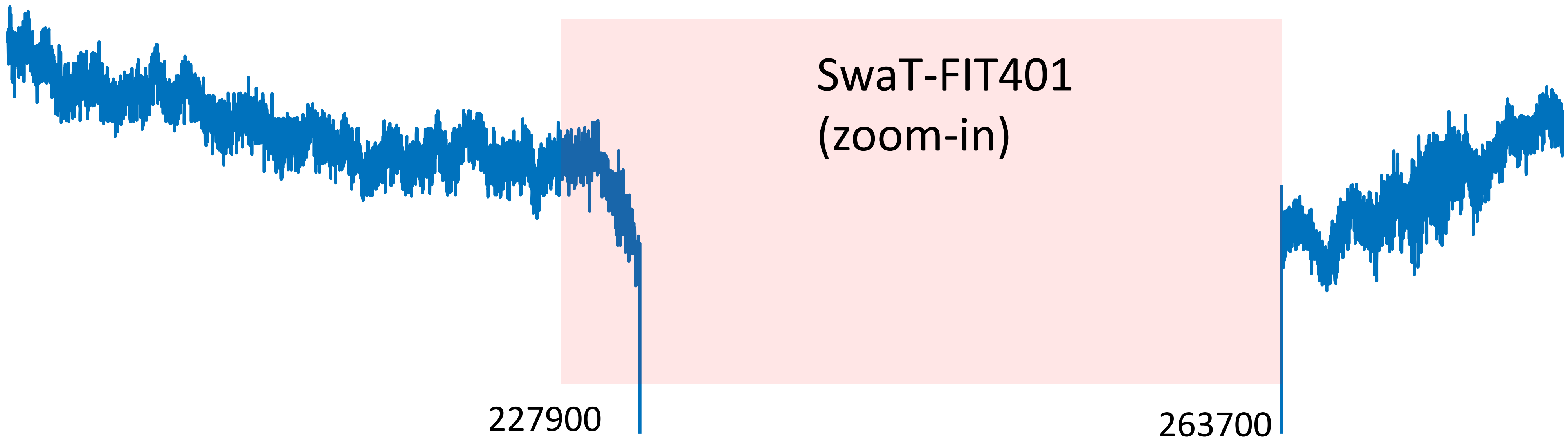


Let's bend over backwards, and *pretend* it is a TSDA problem.

A classic Statistical Process Control (SPC) algorithm says: "*If the next value is more than <u>three Standard Deviations</u> away from data in a short history of points, flag as anomaly*".

This is ***tens*** of *Standard Deviations*, so we can solve this with a simple 70-year-old algorithm.

The scoring functions used greatly inflate performance

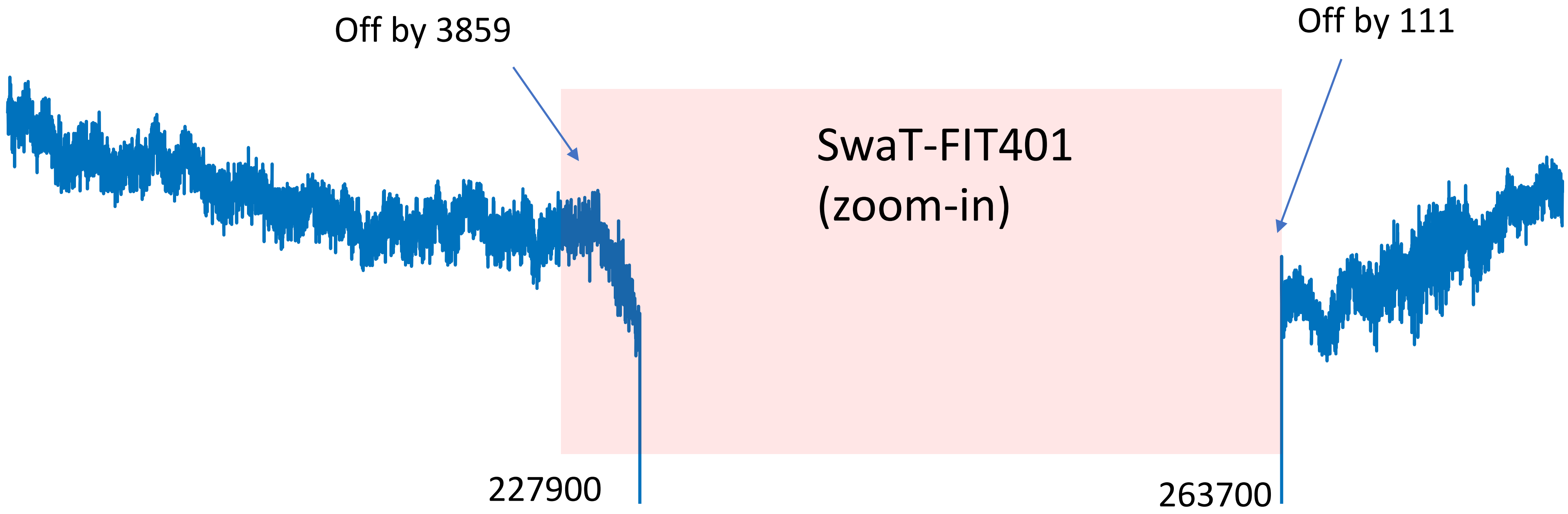


Finding this trivial anomaly should be scored as 1-out-of-1, failing to find it scored as 0-out-of-1.

However, the anomaly is of length 35,800, so most papers report a score like 35,701 out of 35,800, or "*Our accuracy is 0.99723*".

But these are *not* 35,800 independent events! This is a *single* event.

Any high precision claims are necessarily wrong

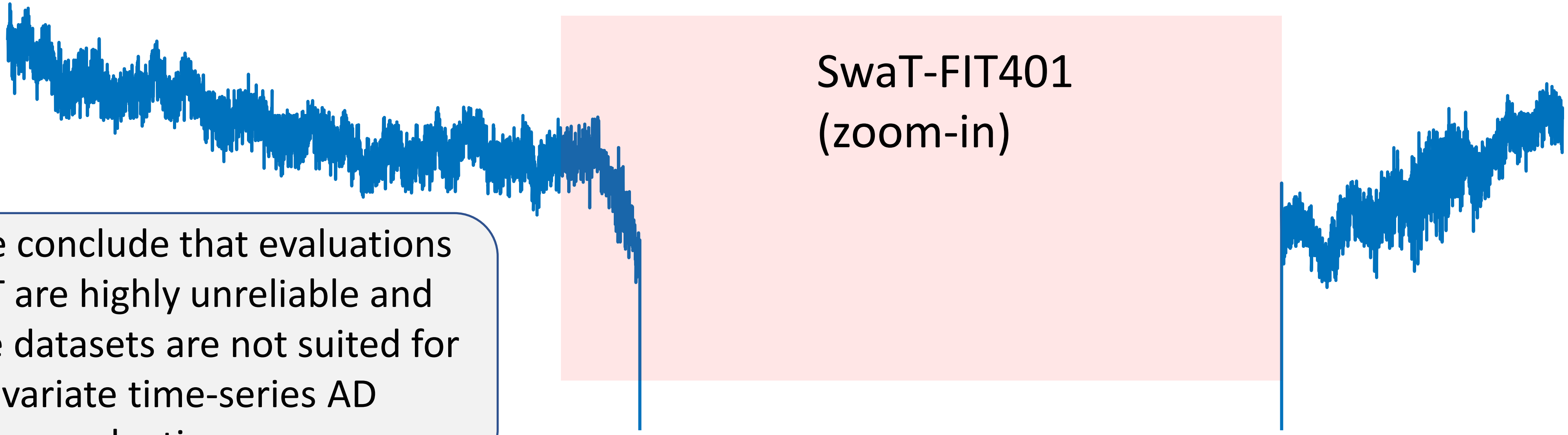


The original labels are almost certainly “sloppy” and “rough”

The fact that they are rounded to the nearest 100 would suggest that, however the error is *greater* than 100.

I have argued elsewhere, that it is essentially *never* possible or meaningful to have high temporal precision claims about exactly *when* an anomaly occurs.

Others in the community are beginning to get it…



…Thus we conclude that evaluations on SWaT are highly unreliable and that these datasets are not suited for multivariate time-series AD evaluation.

Maja Rudolph, Bosch AI

Dennis Wagner, Tobias Michels, Florian C. F. Schulz, Arjun Nair, Maja Rudolph, Marius Kloft: TimeSeAD: Benchmarking Deep Multivariate Time-Series Anomaly Detection. Trans. Mach. Learn. Res. 2023 (2023)

https://ml.informatik.uni-kl.de/publications/2023/TimeSeAD.pdf

Why does this matter?

There are hundreds of papers, containing many hundreds of tables, comparing different algorithms on SWaT.

In most cases, these papers report 3, 4 or even 5 significant digits!

It is all meaningless!

TABLE 7. Comparison between the different detection methods on the second log of the SWaT dataset.

Method	F1-score	Precision	Recall
NN [12]	0.812	0.976	0.696
SVM [15]	0.796	0.925	0.699
1D-CNN [14]	0.860	0.867	0.854
RNN [15]	0.802	0.982	0.678
TABOR [18]	0.823	0.862	0.788
KNN [17]	0.350	0.348	0.348
FB [17]	0.360	0.358	0.358
AE [17]	0.520	0.516	0.516
EGAN [17]	0.510	0.406	0.677
GAN [17]	0.810	0.700	0.954
DIF	0.882	0.935	0.835

TABLE IV: Average Fc_1 score, with the Gauss-D scoring function (except those denoted with *) and top-k threshold. See SI Table S6 for standard deviation.

	DMDS	MSL	SKAB	SMAP	SMD	SWaT	WADI	Mean	Rank
Raw Signal	0.4927	0.2453	0.5349	0.2707	0.5151	0.3796	0.4094	0.4068	9.3
PCA	0.5339	0.4067	0.5524	0.3793	0.5344	0.5314	0.3747	0.4733	5.6
UAE	**0.6378**	**0.5111**	**0.5550**	**0.4793**	0.5501	**0.5713**	0.5105	**0.5450**	**1.6**
FC AE	0.6047	0.4514	0.5408	0.3788	0.5395	0.4478	0.5639	0.5038	4.7
LSTM AE	0.5999	0.4481	0.5418	0.4536	0.5271	0.5163	0.4265	0.5019	4.7
TCN AE	0.5989	0.4354	0.5488	0.3873	**0.5800**	0.4732	0.5126	0.5052	3.9
LSTM VAE	0.5939	0.3910	0.5439	0.2988	0.5427	0.4456	**0.5758**	0.4845	6.0
BeatGAN	0.5391	0.4531	0.5437	0.3732	0.5479	0.4777	0.4908	0.4894	5.0
MSCRED	0.2906	0.3944	0.5526	0.3724	0.4145	0.4315	0.3253	0.3973	8.1
NASA LSTM	0.1284	0.4715	0.5339	0.4280	0.3879	0.1398	0.1058	0.3136	8.9
DAGMM*	0.0000	0.1360	0.0000	0.1681	0.0187	0.0000	0.0256	0.0498	12.9
OmniAnomaly*	0.1425	0.4120	0.4561	0.3767	0.5002	0.1466	0.2443	0.3255	9.4
OCAN*	0.2532	0.3009	0.4369	0.2787	0.4614	0.1547	0.0000	0.2694	11.0

Table 5. Performance of anomaly detection models for the SWaT dataset.

Model	F1-Score	Precision	Recall	FNR	FPR	TP	FN	FP
InterFusion	**90.7%** (92.8% [13])	91.1%	**90.3%**	**9.7%**	1.2%	**49,309**	**5312**	4799
RANSynCoder	82.7% (84.0% [14])	96.6%	72.3%	27.7%	0.4%	39,511	15,110	1380
GDN	80.6% (81.0% [15])	97.8%	68.5%	31.5%	0.2%	37,403	17,218	836
LSTM-ED	81.2% (76.0% [64])	**98.9%**	68.8%	31.2%	**0.1%**	37,586	17,035	**410**
USAD	75.0% (79.1% [12])	91.6%	63.6%	36.4%	0.8%	34,856	19,940	3208

Another Quick Example

There are dozens of papers that use the dataset below for TSAD experiments.

They offer 2, 3 or even 4 significant digits of accuracy measurements.

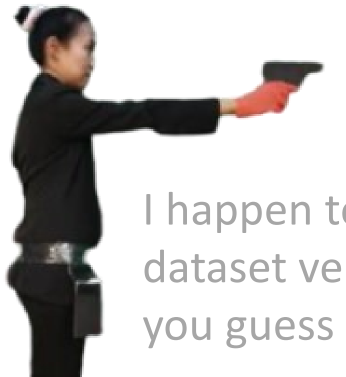
However, this is all just nonsense!

At best, you can get a “**1-out-of-1**” type measurement on this dataset.

NeurIPS 2020

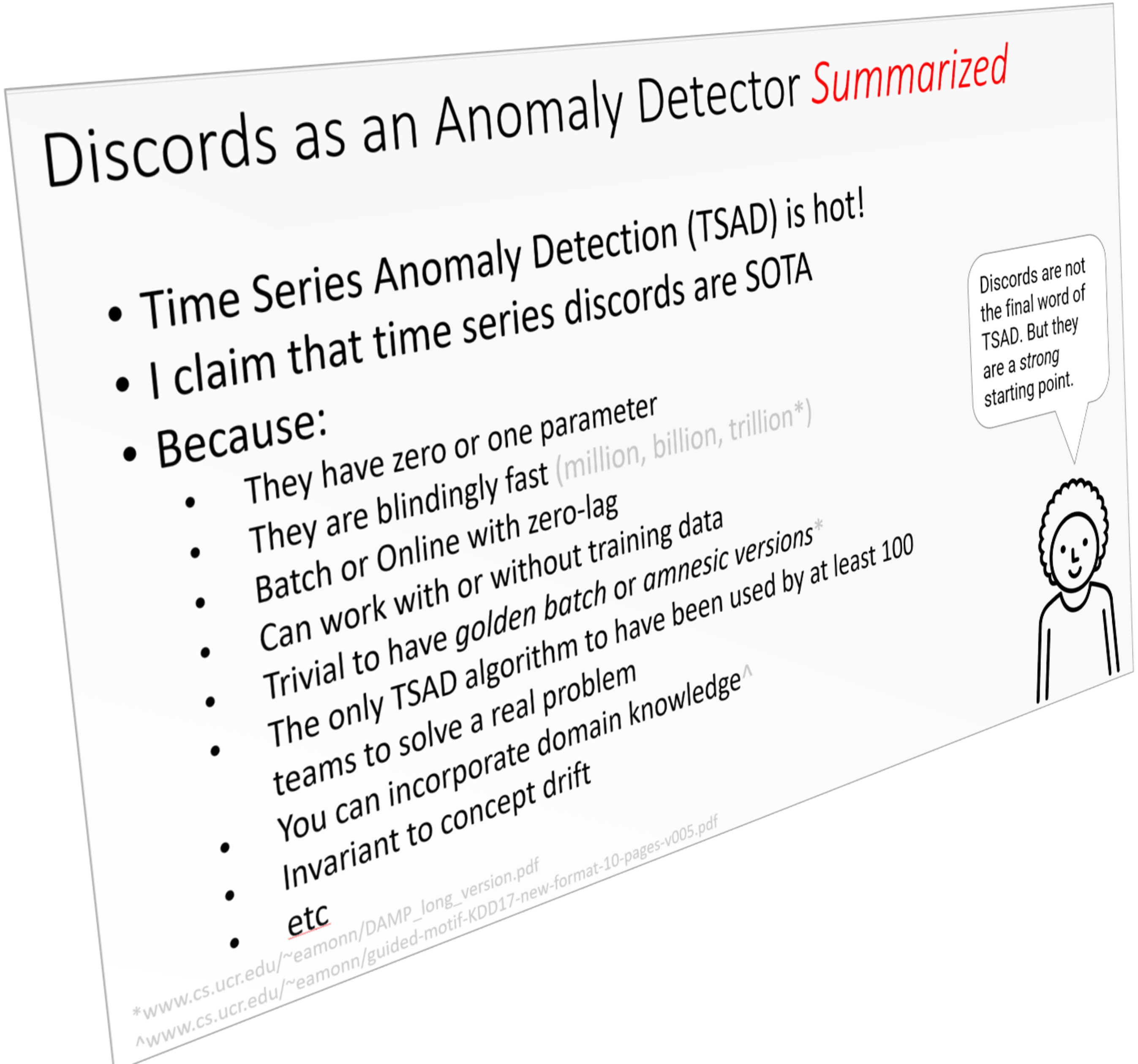
	2D-gesture		
	prec	rec	F1
LOF	27.82	87.21	42.18 (8)
OC-SVM	65.50	25.57	36.78 (14)
iso forest	28.54	68.04	40.22 (10)
deep SVDD	26.26	64.53	37.32 (13)
AnoGAN	57.85	46.50	51.55 (4)
DAGMM	25.66	80.47	38.91 (12)
EncDec-AD	24.88	100.0	39.85 (11)
LSTM-VAE	36.62	67.76	47.54 (6)
MadGAN	29.41	76.40	42.47 (7)
BeatGAN	55.11	45.33	49.74 (5)
OmniAnomaly	27.70	79.67	41.11 (9)
MSCRED	61.26	59.11	60.17 (2.5)
CVDD	56.05	64.95	60.17 (2.5)
THOC	54.78	75.00	**63.31** (1)

Timeseries Anomaly Detection using Temporal

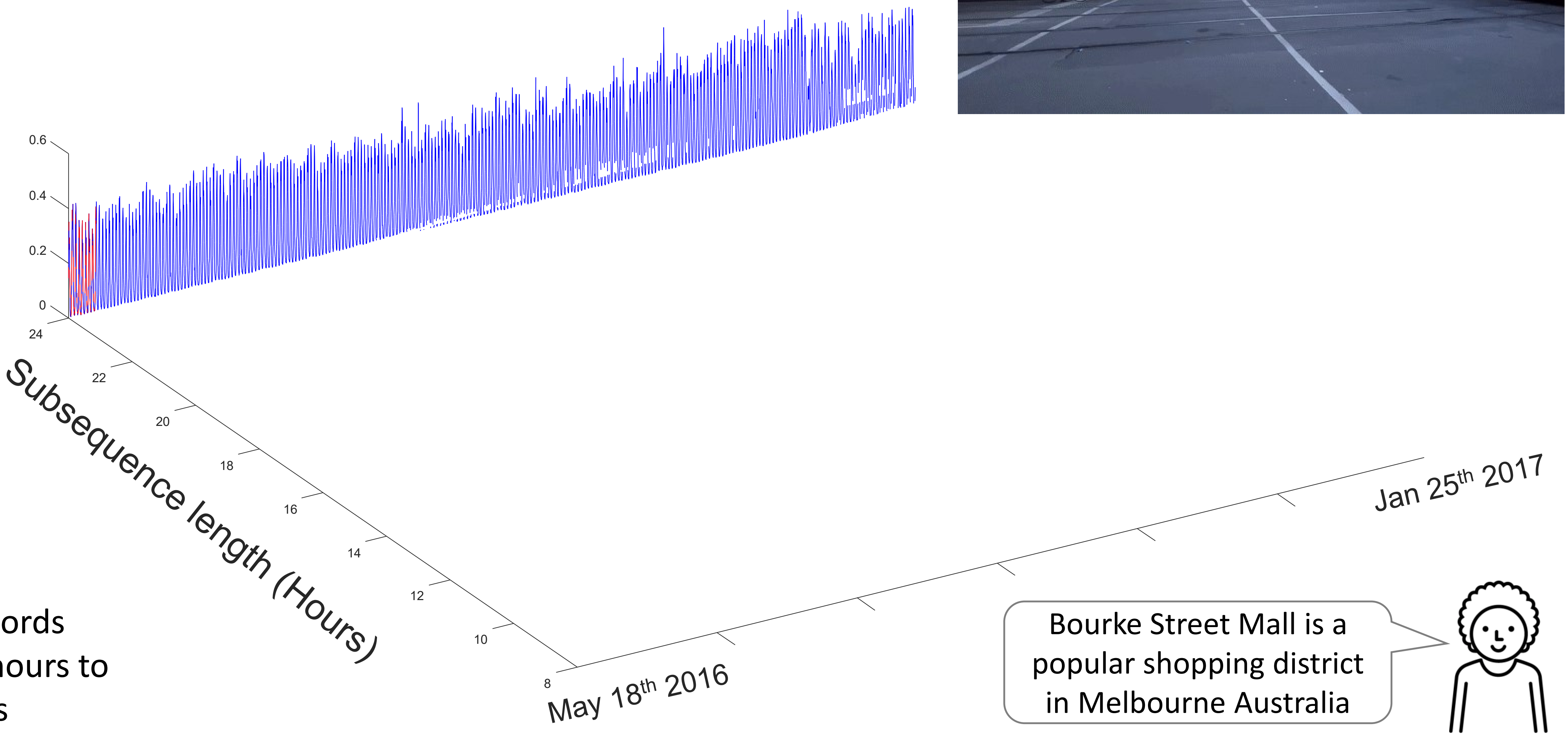
I happen to know this dataset very well; can you guess why?


Time Series Discords are a Pretty Good Anomaly Detector



Let's remind ourselves how good discords are, and then finish up for the day…

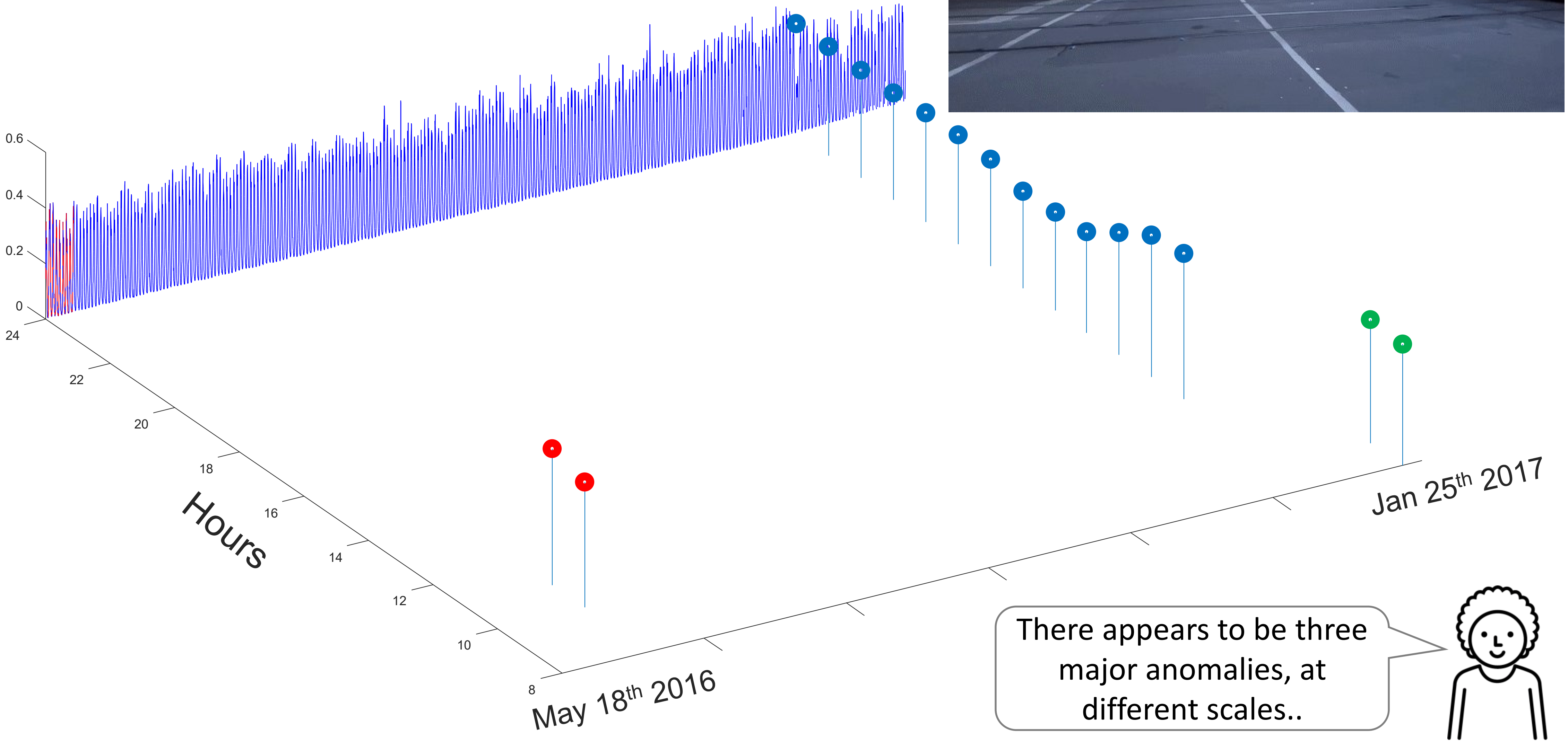
Bourke Street Mall (North) Pedestrian Traffic



left-Discords from 8 hours to 24 hours

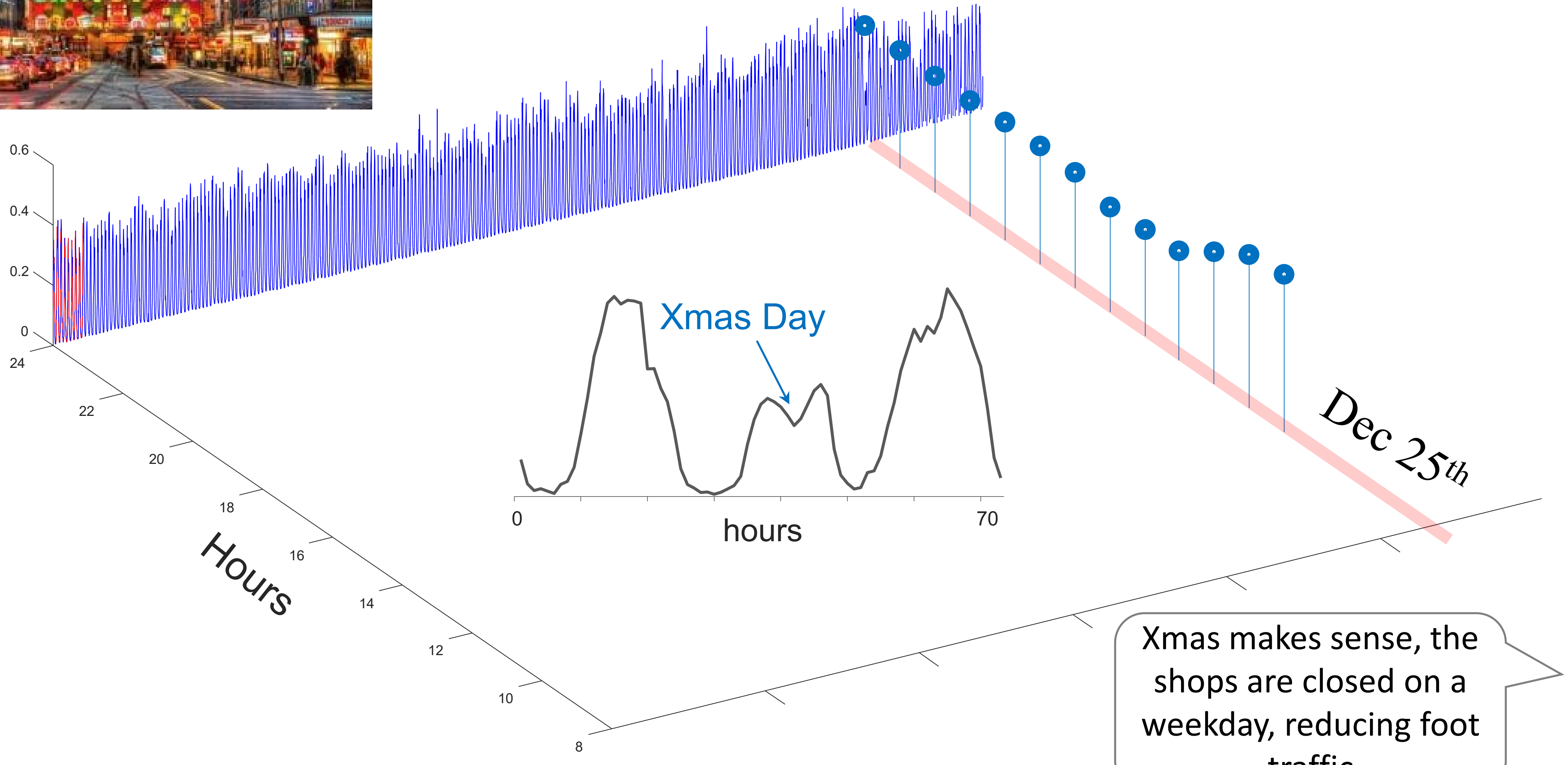
Bourke Street Mall is a popular shopping district in Melbourne Australia

Bourke Street Mall (North) Pedestrian Traffic



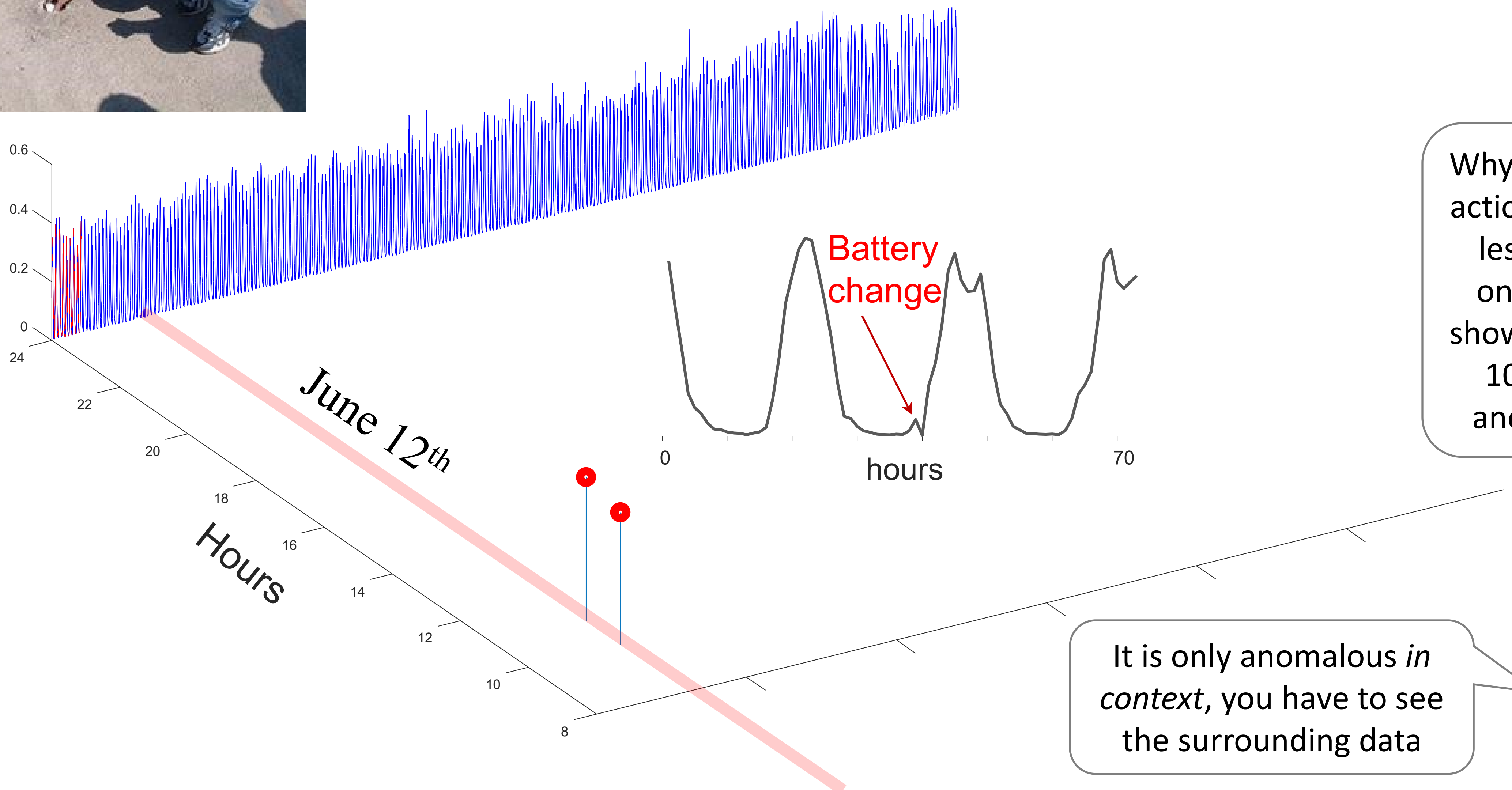
There appears to be three major anomalies, at different scales..

Bourke Street Mall (North) Pedestrian Traffic




Xmas makes sense, the shops are closed on a weekday, reducing foot traffic

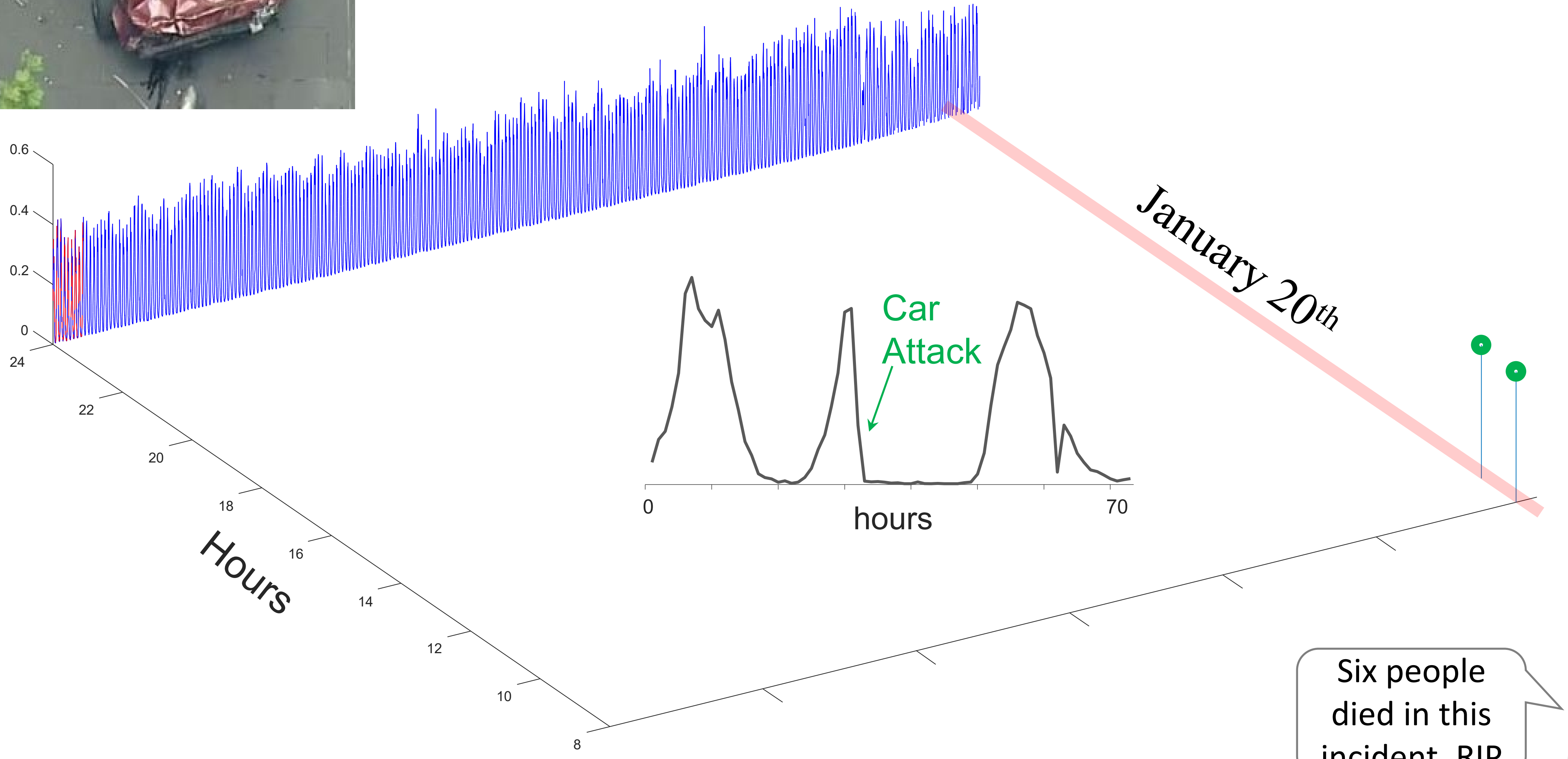
Bourke Street Mall (North) Pedestrian Traffic

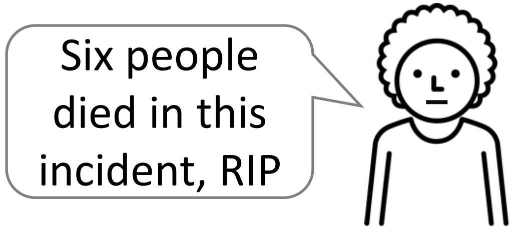


Why does an action taking less than one hour show up as a 10-hour anomaly?

It is only anomalous *in context*, you have to see the surrounding data

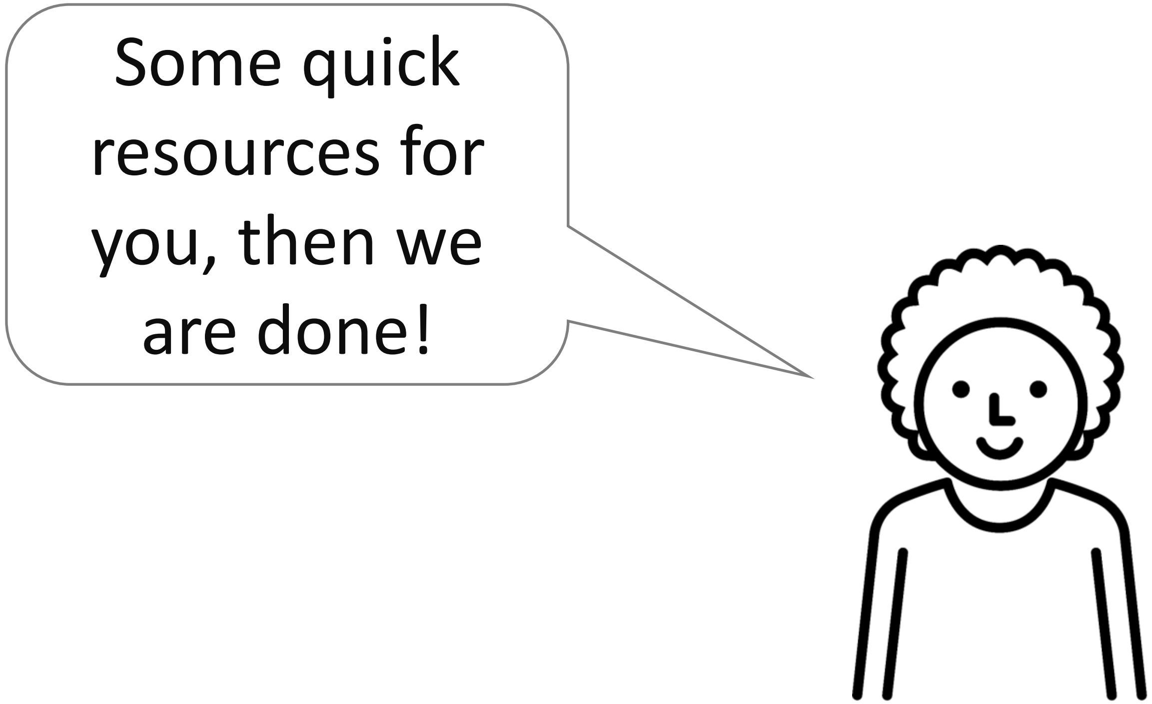
Bourke Street Mall (North) Pedestrian Traffic





Mini Review for TSAD

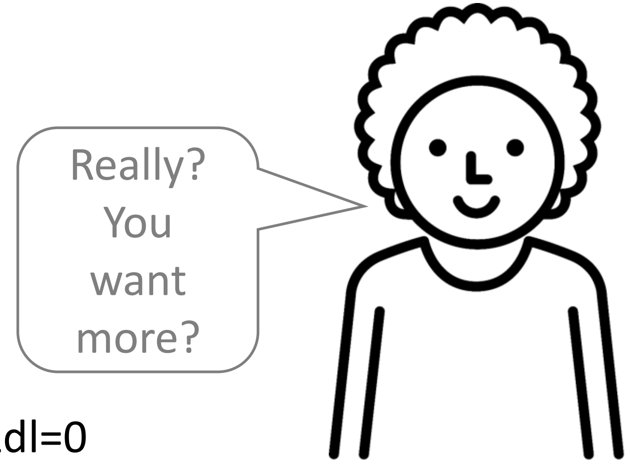
- Most TSAD research is wrong
  - **Bad datasets**
  - Bad evaluation metrics
  - Deliberate crippling of rivals
  - Ignoring rivals
  - A real lack of introspection (this is the real problem)
  - etc.
- This is a great opportunity. There is a need for a *great* paper on TSAD
- Better datasets would help *a lot*. Ideally, a *committee*, not an *individual*, would create such datasets.
- While not the perfect or only solution, *time series discords* are very competitive.


Some quick resources for you, then we are done!

Want more slides?

I have made bonus slides! All at

- Review of (almost) all available TSAD datasets *Problems with Time Series Anomaly Detection.pptx*
- Tutorial on DTW (co-authored with Mueen) *DTW_Tutorial.pptx*
- Advice of doing time series research (using special data to test if you are finding the structures you claim to be finding) *Sanity Checks for TS Research.pptx*
- Why Indexing is needed (and generally, not *possible*) for time series data. *Why_no_indexing_for_time_series.pptx*
- A gallery of interesting and beautiful time series motifs *Gallery of Cool Time Series Motifs.pptx*
- Long tutorial on the Matrix Profile *Matrix_Profile_Tutorial_Part1.pptx* and *Matrix_Profile_Tutorial_Part2.pptx*
- Why most time series anomaly detection papers are wrong (extended) Irrational *Exuberance_why_most_TSAD_is_wrong.pptx*
- Why “Hashing Motifs” is a dead-end
- Multidimensional MASS *Multidimensional similarity search with MASS (Penguin example).pptx*
- Advice on making figures *MakingGoodFigures.pptx*
- **META** Increasing your h-index *Getting an h-index of 100 in Twenty Years or Less.pptx*
- **META** How to do good research *How_to_do_Research_Keogh.ppt*



https://www.dropbox.com/scl/fo/7n23jk9te8kto64q4qnba/h?rlkey=yzl9ct3bi0c9xei5t19gxlztw&dl=0

Want Datasets?

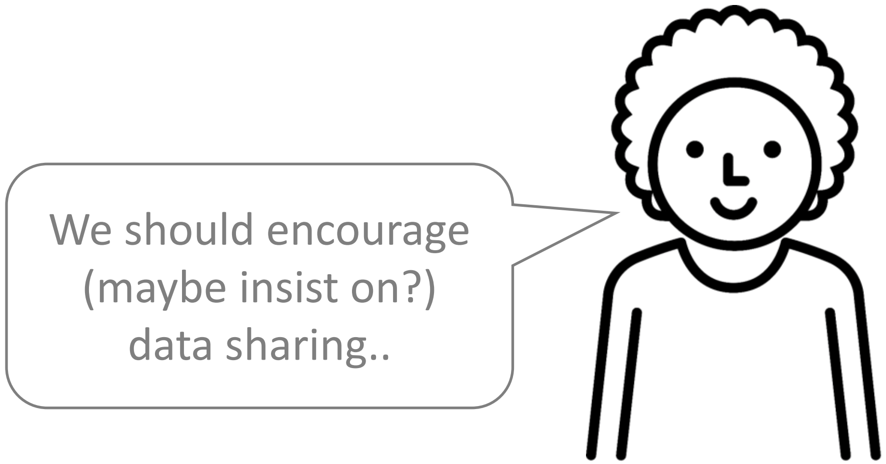
- Time Series Classification
  - *https://www.cs.ucr.edu/~eamonn/time_series_data_2018/*

- Time Series Anomaly Detection The **Hexagon ML**/UCR Time Series Anomaly Detection Datasets
  - *www.cs.ucr.edu/~eamonn/time_series_data_2018/UCR_TimeSeriesAnomalyDatasets2021.zip*

- General source of great data (but it helps to work with a medical doctor)
- PhysioBank ATM *archive.physionet.org/cgi-bin/atm/ATM*

- General source DRYAD *datadryad.org/stash*

- Misc: I have huge collection of datasets, ask me



Who is doing original and/or useful research in TS

Ben Fulcher, Sydney University. **The** greatest expert of time series *features*. www.benfulcher.com

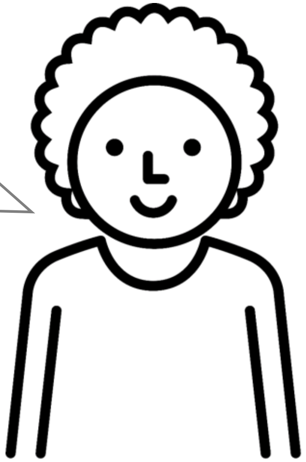
Abdullah Mueen, University of New Mexico. **The** expert on time series similarity search, DTW, and computing time series correlation in various settings. Inventor of MASS.

Anthony (Tony) Bagnall, University of East Anglia. Expert on *time series classification*. Also really understands how to compare multiple machine learning algorithms (do a “bakeoff”) to find small, but significantly significant differences.

Geoff Webb, Monash University - Expert on *time series classification*, *time series distance measures, large scale empirical comparisons*, and *optimizing DTW lower bounds*.

Rob Hyndman, Monash University - Expert on *time series forecasting*.

Eamonn’s subjective view, apologies for any errors/omissions, feel free to suggest names for updated slides



Summary

We have seen two simple tools, MASS and the Matrix Profile

- These two tools let you find matches to *queries*, *motifs*, and *discords*.
- This allows you to do *so much* useful time series data mining.
- We did not explore:
  - Some additional useful primitives, *Chains*, *Consensus Motifs*, *Novelets*, *Shapelets*, *Platos, Snippets, FLOSS*…
  - Generalizing the self-joins to joins between multiple time series…
- These ideas are expounded in other tutorials and papers.

We have seen that MASS and the Matrix Profile are fast enough for almost anyone

- And, if needed, can be made faster by:
  - GPU and FPGA versions
  - Using the *anytime* algorithm property
  - Using the *contract* algorithm property

Thanks for listening!

Time Series Data Mining: A Unifying View

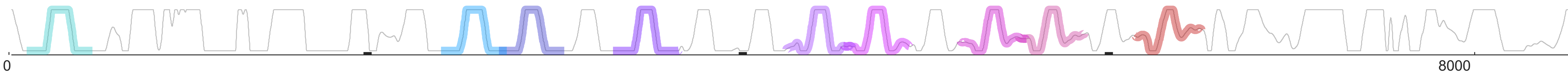
Questions or Comments?

Eamonn Keogh
UCR
eamonn@cs.ucr.edu

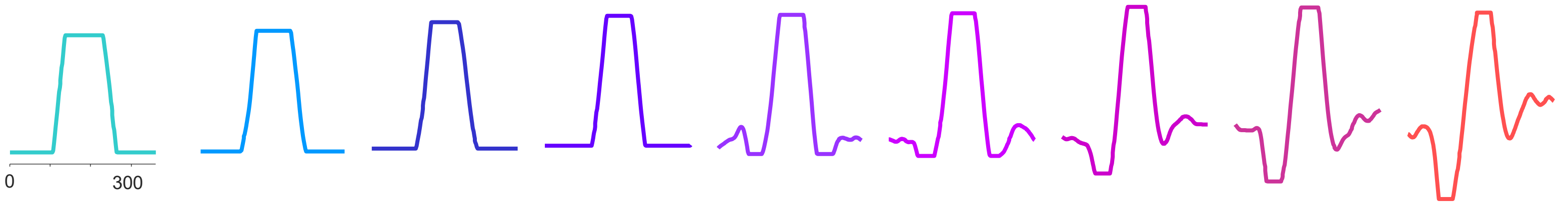
Appendices/ Backup Slides

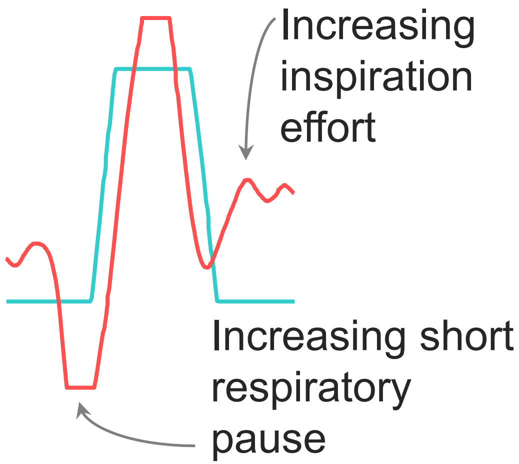
Time Series Chains

Dr. Greg Mason an expert on cardiopulmonary interactions, says that the chain appears to be "*attempts to inspire against an obstruction coming the back of the tongue*".

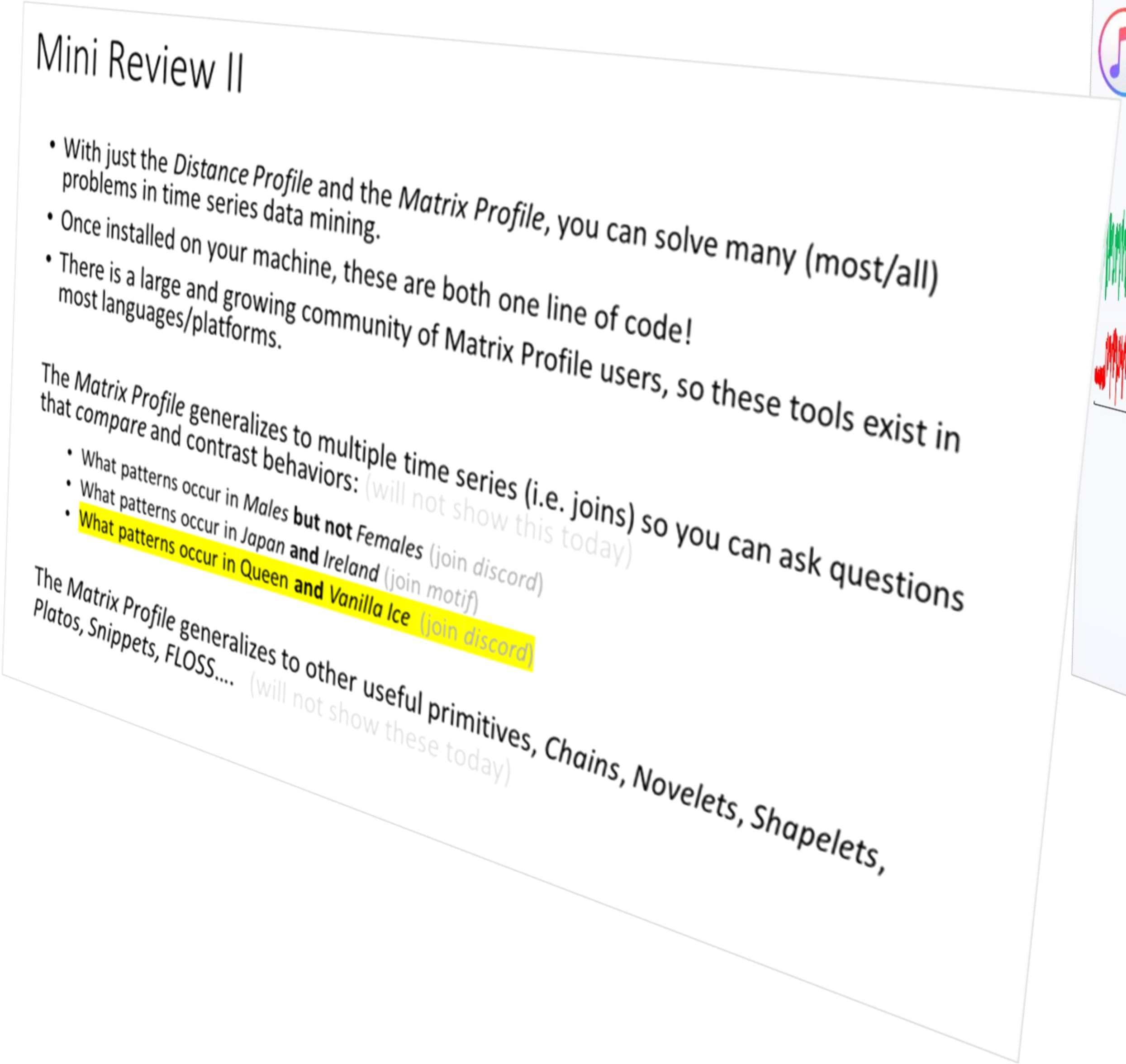
- 

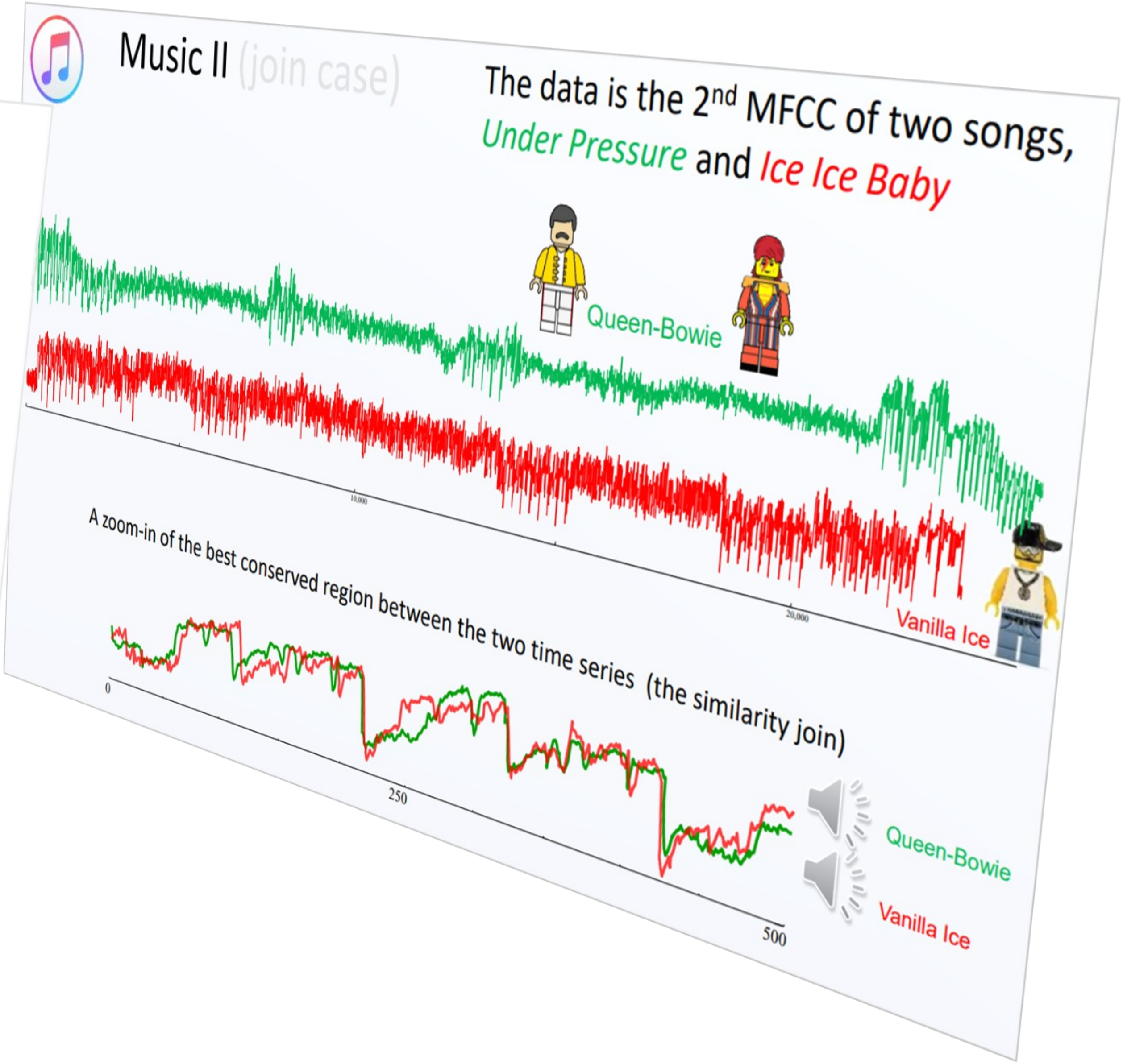
- Chains often have a physical interpretation, including this one.





Sigh…. The Queen/Vanilla Ice example is *my* original example, it dates back to 2015.





The Twin Freak “Problem” is Philosophically Meaningless

Decimus Junius Juvenalis

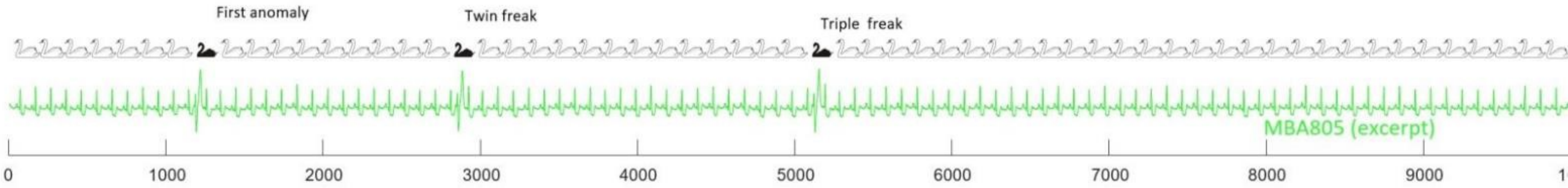
Rara avis in terris nigroque simillima cygno

When Willem de Vlamingh became the first European to see a black swan on the 10th January 1697, it *was* an anomaly. By the next day, when he had seen hundreds, black swans were no longer an anomaly.

Consider the dataset below used to justify twin freak invariant algorithms. Perhaps the first arrythmia can be considered an anomaly. But there are 132 more such beats. Is the 133rd such arrythmia still an “anomaly”?

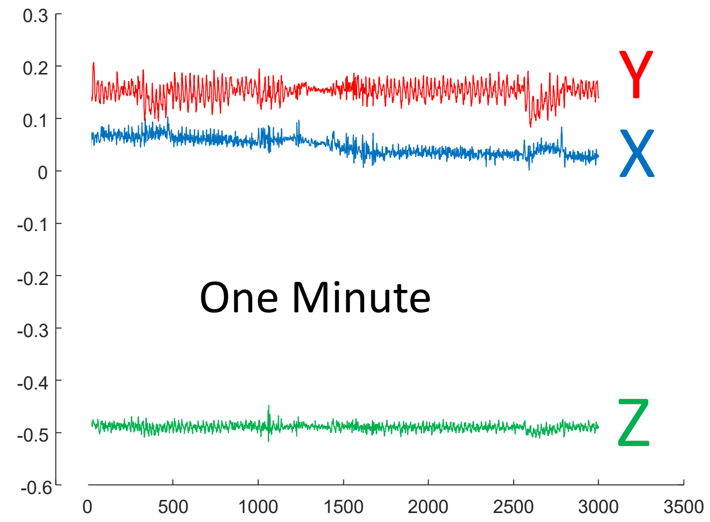
No! In fact, you could argue the opposite. If we had ten seconds *without* such an arrythmia, that would be a stunning anomaly.

This misguided focus on this trivially solvable pseudo-problem has wasted much of the energy of the community[1].

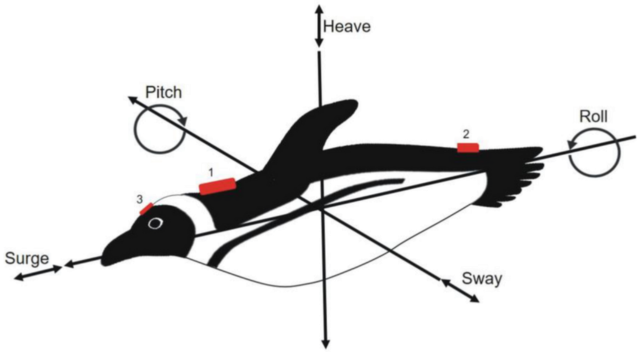


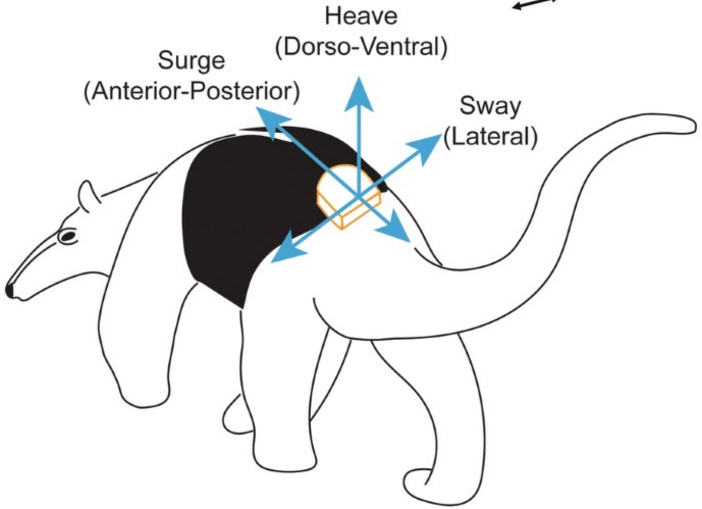
To be clear, X, Y and Z do not have any standard meaning for animal work.
Surge, *Heave* and *Sway* are the biological terms for acceleration directions.
(roll, pitch, yaw for rotation, are measures with a gyroscope).

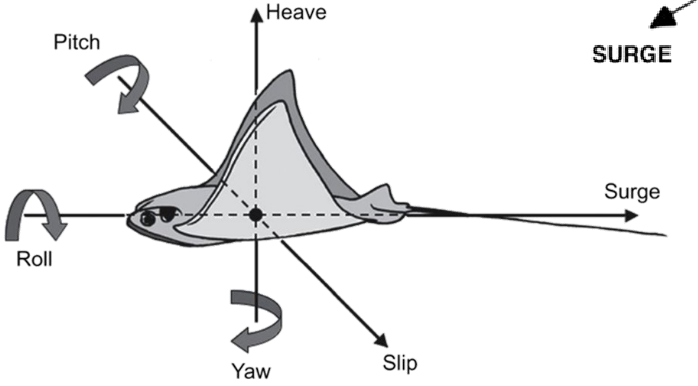
It is rare that X, Y and Z can be made to exactly map to Surge, Heave and Sway, because of the orientation of the sensors against the animal's body.

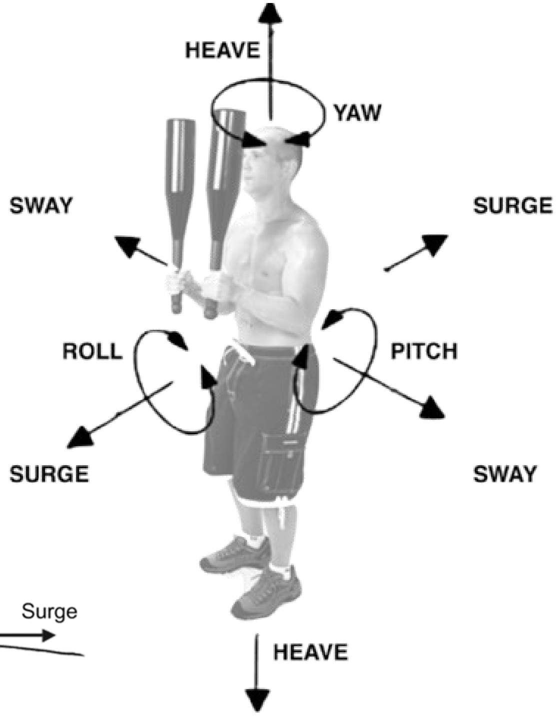


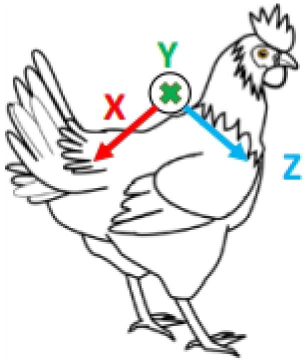
triaxial accelerometer

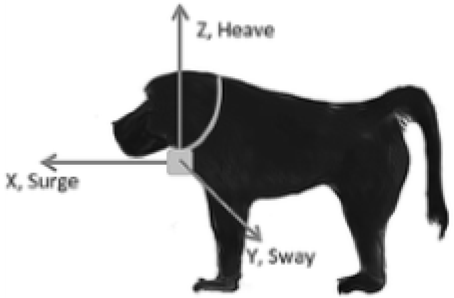












- **Background Fact:** MASS has its fastest case when the data's length is a power-of-two, it is worth padding to the next power of two.

```
• data= cumsum(randn(1,67000000));  % Suppose you had this dataset..
• dist = MASS_V2(data ,query );     % This search is pretty fast..
• % ..however, if I pad the data being searched up to the next power of two..
• data = [data zeros(1,2^nextpow2(67000000) -67000000)]; % ..it is even faster
```