FINGERPRINTING PROJECT

BREAKDOWNS:

1) Breakdown at around 335 AM, 5/2/2018

2) Breakdown at around 145 PM, 5/17/2018

3) Breakdown at around 610 PM, 6/10/2018

4) Breakdown at around 855 AM, 7/30/2018

5) Breakdown at around 1245 PM, 12/11/2018

PROCEDURE:

Located the given break down points and store all the sensor data three days before each of the given break down points in separate data sets names y1,y2,y3,y4,y5

(I) Task 1: to compute similarity between each of the breakdowns:

1) Each of the features of the Ys was averaged to represent each breakdown event with 12 values, one for each sensor:

- 'Gear bearing temp.'
- 'Gear oil temperature'
- 'Generator bearing front temperature'
- 'Generator RPM'
- 'Generator spring temp'
- 'Grid busbar temperature'
- 'Rotor inverter temperature L1'
- 'Phase 1 temperature'
- 'Hydraulic oil temperature'
- 'Nacelle position'
- 'Nacelle temperature'
- 'Spinner temperature'

   https://stats.stackexchange.com/questions/80377/which-distance-to-use-e-g-manhattan-euclidean-bray-curtis-etc

2) Each breakdown was summed as an array of 12 values that represent the average over 3 days of the above values
3) To assess the similarity between the breakdowns , we used different similarity measures between each of these 12-dimensional arrays to compute the most similar breakdowns in descending order. The similarity measures include the following norms :
    - **Euclidean**

Absolute values are taken into consideration, relative values do not matter, scale matters, correlated features change measure

Results: Given below, in decreasing order of similarity between the breakdowns, are tuples consisting of the the two breakdowns whose similarity we are assessing and the value of the given norm

	AFTER NORMALIZATION
1	(2, 3, 0.33)
2	(1, 5, 0.45),
3	(1, 2, 0.51),
4	(3, 4, 0.52),
5	(2, 4, 0.53),
6	(1, 3, 0.56),
7	(1, 4, 0.57),
8	(4, 5, 0.64),
8	(2, 5, 0.65),
10	(3, 5, 0.79)

○

- **Manhattan**

  Results: Given below, in decreasing order of similarity between the breakdowns, are tuples consisting of the the two breakdowns whose similarity we are assessing and the value of the given norm

	AFTER NORMALIZATION
1	(2, 3, 0.9238)
2	(2, 5, 1.1694)
3	(1, 5, 1.3133)
4	(1, 3, 1.3886)
5	(3, 4, 1.4184)
6	(2, 4, 1.4222)
7	1, 2, 1.4729)
8	(1, 4, 1.4803)

8	(4, 5, 1.6741)
10	(3, 5, 2.0933)

- **Minkowski**

  Results: Given below, in decreasing order of similarity between the breakdowns, are tuples consisting of the the two breakdowns whose similarity we are assessing and the value of the given norm
  p=5 here
  As p becomes larger and larger, only the larger values in the data contribute to the distance, the smaller values are damped out
  https://math.stackexchange.com/questions/1060294/physical-meaning-of-minkowski-distance-when-p-2

	AFTER NORMALIZATION
1	(2, 3, 0.23),
2	(1,5, 0.31)
3	(1, 2, 0.36),
4	(2, 4, 0.36),
5	(3, 4, 0.36),
6	(1, 3, 0.4),
7	(1, 4, 0.4),
8	(4, 5, 0.44),
8	(2, 5, 0.61),
10	(3, 5, 0.63)

- **Braycurtis**

  Results:Given below, in decreasing order of similarity between the breakdowns, are tuples consisting of the the two breakdowns whose similarity we are assessing and the value of the given norm

  Relative magnitudes are taken into account, triangle law does not hold https://www.researchgate.net/post/What_is_the_difference_between_Bray-Curtis_Similarity_Sorensen_Distance_and_Bray-Curtis_Index

	AFTER NORMALIZATION
1	(2, 3, 0.0667)
2	(2, 5, 0.0797)
3	(1, 5, 0.0948)
4	(2, 4, 0.0969)
5	(3, 4, 0.0992)
6	(1, 3, 0.1065)
7	(1, 4, 0.1069)
8	(1, 2, 0.1099)
8	(4, 5, 0.1108)
10	(3, 5, 0.1464)

- **Chebyshev**
  Results;Given below, in decreasing order of similarity between the breakdowns, are tuples consisting of the the two breakdowns whose similarity we are assessing and the value of the given norm

  Minkowski with p = infinite

https://arxiv.org/pdf/1708.04321.pdf

	AFTER NORMALIZATION
1	(2, 3, 0.2225)
2	(1, 5, 0.2881)
3	(2, 4, 0.3121)
4	(4, 5, 0.3208)
5	(1, 2, 0.3252)
6	(1, 3, 0.3339)
7	(1, 4, 0.3456)

8	(3, 4, 1.4402)
8	(1, 2, 1.5106)
10	(3, 5, 1.8824)

- Canberra
  Results;Given below, in decreasing order of similarity between the breakdowns, are tuples consisting of the the two breakdowns whose similarity we are assessing and the value of the given norm

  https://arxiv.org/pdf/1708.04321.pdf

	AFTER NORMALIZATION
1	(2, 3, 0.7881)
2	(2, 5, 1.1062)
3	(1, 5, 1.2068)
4	(2, 4, 1.3026)
5	(4, 5, 1.3343)
6	(1, 3, 1.4203)
7	(1, 4, 1.4377)
8	(3, 4, 1.4402)
8	(1, 2, 1.5106)
10	(3, 5, 1.8824)

- Hamming distance: All are equivalent using this measure.

OBSERVATION :
After normalisation, the ranking of the breakdowns changes in all measures except for canberra (in which there is only one change)
In each of these measures, distance between 1,1 = 0

(II) Dynamic Time warping
Missing values led to dtw_distance being infinity
The missing values were replaced with preceding values (not average since it would introduce noise)
DTW is a method that tries to perform shape matching between two time series.
In this sakoe chiba diagonal (radius 4) was used so as to make the method less computationally expensive.
The radius controls the maximum no. of steps away from the diagonal at watch we can match points on the two time series.

	AFTER NORMALIZATION
1	(2, 3, 37.11)
2	(2, 5, 41.28)
3	(1, 5, 41.82),
4	(1, 2, 44.53),
5	(1, 3, 46.29),
6	(3, 4, 47.96),
7	(3, 5, 49.68),
8	(2, 4, 51.21),
8	(4, 5, 51.63),
10	(1, 4, 55.87)

(III) Symbolic aggregate approximation:
This performs rough magnitude and shape matching.
The time series is cut into strips. Each strip is assigned a letter according to which range its mean lies in.
Thus , we can represent time series as sequences. In this, we greatly reduce the complexity of the data but we lose detailed information about the shape. This can be thought of as an aggressive smoothing technique.

Once we get a sequence for each time series, we can compare the edit distance between two such sequences to get a rough estimate of the cost required to convert one time series into the other (similar to dtw)

1	(1, 4, 670)
2	(1, 2, 678)
3	(2, 4, 678)
4	(2, 3, 722)
5	(1, 3, 729)
6	(3, 4, 741)
7	(3, 5, 746)
8	(2, 5, 748)
8	(4, 5, 752)
10	(1, 5, 776)

No change in scores with normalization since scale changes appropriately to fit in the same number of letters. Distance is finally computed between the same sequence of letters

(IV)Piecewise aggregate function:

The time series are split into hourwise strips and the mean value is calculated for each strip for each sensor. Thus for one breakdown each sensor is represented by a set of 72 values. Thus each breakdown is represented by an array of 12x72.
We compare two matrices using the above distance measures to assess the similarity between the breakdowns.

BRAYCURTIS

	AFTER NORMALIZATION
1	(2, 3, 1.387)
2	(1, 5, 1.438)
3	(2, 5, 1.553)
4	(4, 5, 1.698)
5	(1, 2, 1.787)
6	(2, 4, 1.821)
7	(3, 4, 1.979)

8	(3, 5, 2.005)
8	(1, 3, 2.023)
10	(1, 4, 2.118)

CANBERRA

	AFTER NORMALIZATION
1	(2, 3, 100.27)
2	(1, 5, 110.831)
3	(2, 5, 120.327)
4	(4, 5, 130.506)
5	(1, 2, 135.837)
6	(2, 4, 144.124)
7	(3, 5, 151.819)
8	(1, 3, 152.778)
8	(3, 4, 153.675)
10	(1, 4, 162.704)

CHEBYSHEV

	AFTER NORMALIZATION
1	(2, 3, 3.736)
2	(2, 5, 3.837)
3	(3, 5, 3.864)
4	(1, 3, 4.282)
5	(3, 4, 4.477)
6	(1, 5, 4.605)
7	(4, 5, 4.755)
8	(1, 2, 4.76)
9	(2, 4, 5.008)

10	(1, 4, 5.65)

MANHATTAN

	AFTER NORMALIZATION
1	(2, 3, 95.065)
2	(1, 5, 112.973)
3	(2, 5, 121.096)
4	(1, 2, 125.206)
5	(2, 4, 141.193)
6	(1, 3, 142.173)
7	(3, 4, 142.968)
8	(4, 5, 149.508
8	(3, 5, 158.111)
10	(1, 4, 167.14)

CORRELATION DISTANCE

	AFTER NORMALIZATION
1	(1, 2, 7.277)
2	(4, 5, 9.863)
3	(1, 3, 10.259)
4	(2, 3, 10.653)
5	(3, 5, 11.022)
6	(2, 4, 11.151)
7	(1, 4, 11.409)
8	(3, 4, 11.534)
8	(2, 5, 11.942)
10	(1, 5, 12.602)

HAMMING

	AFTER NORMALIZATION
1	1, 5, 11.917)
2	(2, 3, 11.958)
3	(2, 5, 11.972)
4	(3, 5, 11.972)
5	(4, 5, 11.972)
6	(1, 3, 11.986)
7	(1, 2, 12.0)
8	(1, 4, 12.0)
8	(2, 4, 12.0)
10	(3, 4, 12.0)

Data missing from 28 april 2018 - 30th april 2018

Due to additi on of more dates for breakdown 1 , the results have changed

SL.NO.	**AVERAGE METHOD**		**STRIPWISE AVERAGE (PAA)**		**SAX**	**DTW**	
	BEFORE NORMALIZATION	**AFTER NORMALIZATION**	**BEFORE**	**AFTER**		**BEFORE**	**AFTER**
MAX SIMILARITY	-	2,3 - 6	-	2,3 - 4 1,2 - 1 1,5 - 1	1,4	-	2,3
MIN SIMILARITY	-	3,5 - 6	-	1,4 - 4 1,5 - 1 3,4 - 1	1,5	-	1,4

Hence, most similar : breakdown 2 and breakdown 3

FREQUENCY DOMAIN MEASURES:

FOURIER TRANSFORM:

For each pair of breakdowns, the 12 time series were converted to frequency domain. The coherence between corresponding pairs of sensor time series was checked.

1	(2, 4, 40.33)
2	(3, 4, 40.575)
3	(2, 3, 40.591)
4	(1, 3, 41.248)
5	(1, 4, 41.372)
6	(4, 5, 41.554)
7	(1, 2, 42.074)
8	(3, 5, 43.125)
8	(1, 5, 43.271)
10	(2, 5, 43.468)a

STRUCTURAL SIMILARITY:

Similarity between images of spectrograms was measured. This is a statistical measure that takes into account how humans perceive similarity of two images disregarding any changes in brightness, shift etc.

1	(2, 4, 0.982)
2	(1, 4, 0.979),
3	(1, 2, 0.978),
4	(4, 5, 0.976),
5	(2, 5, 0.975),
6	(3, 4, 0.974),

7	(1, 5, 0.974),
8	(2, 3, 0.971),
9	(1, 3, 0.969),
10	(3, 5, 0.968),

GREY LEVEL CO-OCCURRENCE MATRIX:

This helps us analyse the texture of the spectrogram images so as to evaluate the magnitudes of power spectral densities and frequencies occurring consecutively.

1	(2, 4, 1103201.052)
2	(3, 4, 1111852.772)
3	(1, 4, 1324024.283)
4	(3, 5, 1378010.586)
5	(1, 2, 1412948.475)
6	(2, 3, 1413449.51)
7	(4, 5, 1445831.131)
8	(1, 3, 1633071.417)
9	(2, 5, 1653474.64)
10	(1, 5, 2144381.242)

FEATURE MATCHING OF SPECTROGRAM IMAGES:

This helps us match the spectrogram images without accounting for any shift in the spectrogram.

1	(2, 4, 393)
2	(2, 3, 386)
3	(4, 5, 385)
4	(3, 5, 383)
5	(2, 5, 381)
6	(1, 5, 381)
7	(3, 4, 378)
8	(1, 2, 376)

9	(1, 4, 368)
10	(1, 3, 357)

Frequency domain analysis summary

	FFT	**SSIM**	**GLCM**	**FEATURE MATCHING**
MAX SIMILARITY	2,4	2,4	2,4	2,4
MIN SIMILARITY	2,5	3,5	1,5	1,3

LINKS AND PLOTS:

https://pdfs.semanticscholar.org/2f5a/4b8b158117928e9eee7ac6ce7da291ec9bd2.pdf

https://courses.cs.washington.edu/courses/cse455/09wi/Lects/lect6.pdf

https://www.researchgate.net/post/Is_there_any_way_to_measure_texture_similarity

https://www.tutorialspoint.com/artificial_neural_network/artificial_neural_network_kohonen_selforganizing_feature_maps.htm

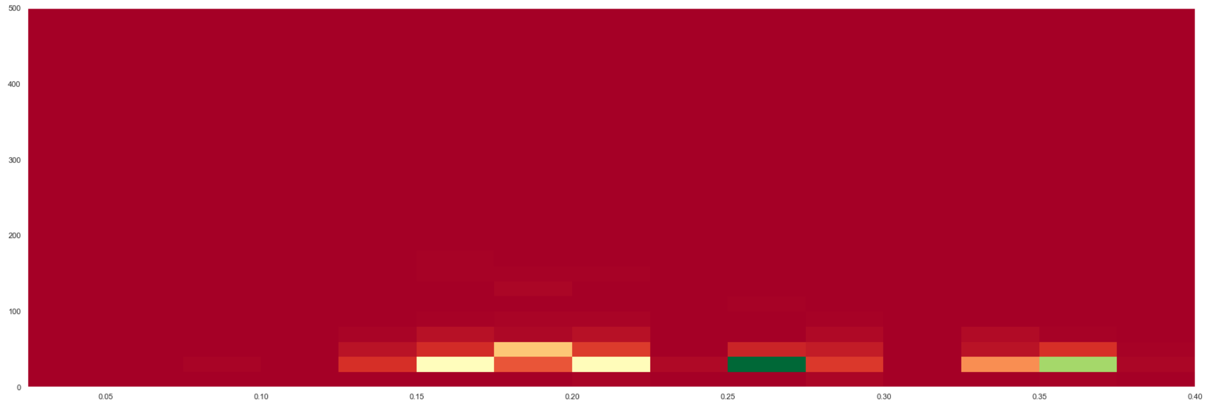
https://pdfs.semanticscholar.org/7c4b/38d3b9781f2401c999a777ed7ae6e19a7f89.pdf

https://www.tandfonline.com/doi/full/10.1080/09524622.2018.1431958 (time-warping of spectrogram)

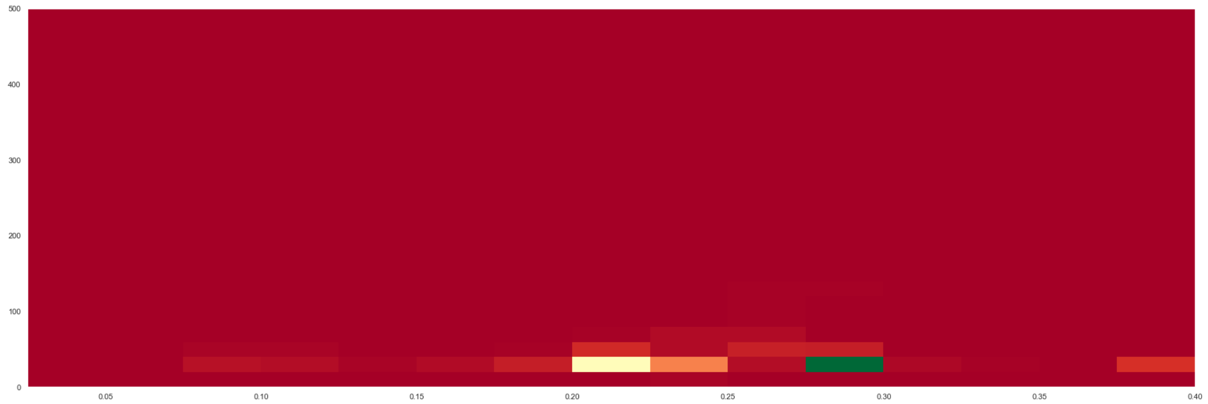
http://optics.sgu.ru/~ulianov/Bazarova/LASCA_literature/InvariantImageRecognitionZernikeMoments.pdf (Zernike moments)
[The Zernike moment of an image is the same as that of the rotated image]

SPECTROGRAM PLOTS OF HYDRAULIC OIL TEMP:

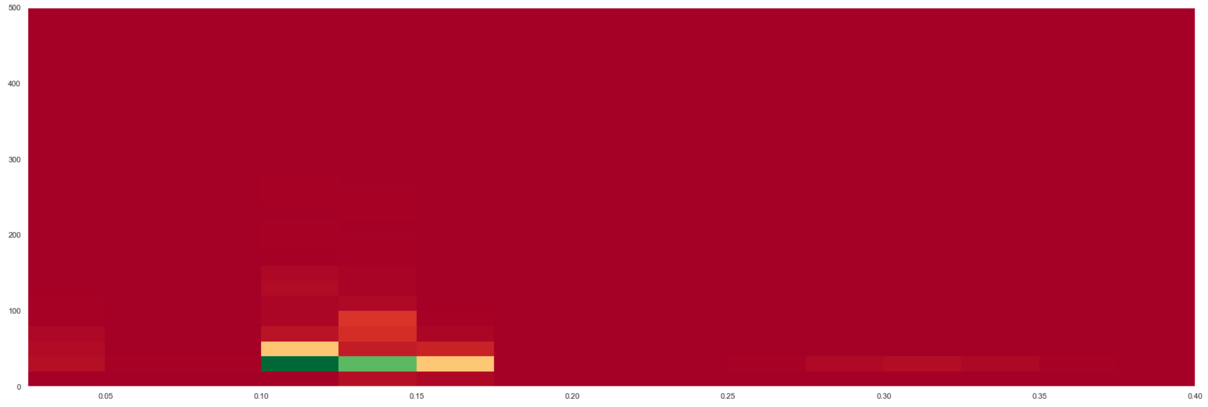
BREAKDOWN 1



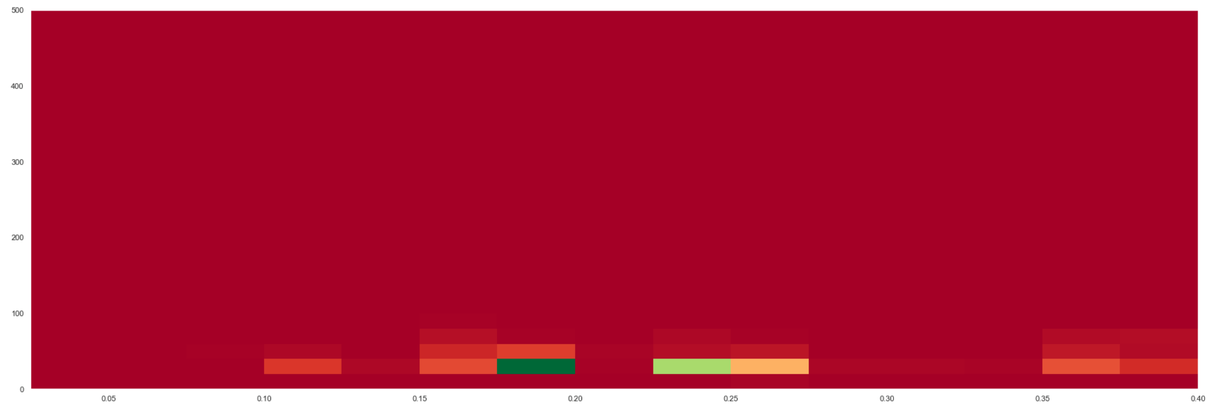
BREAKDOWN 2:



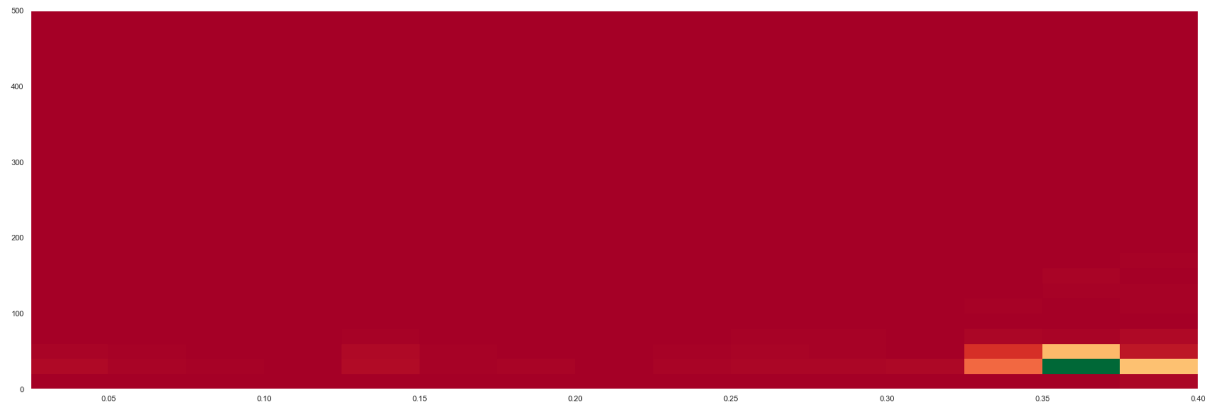
BREAKDOWN 3:



BREAKDOWN 4:

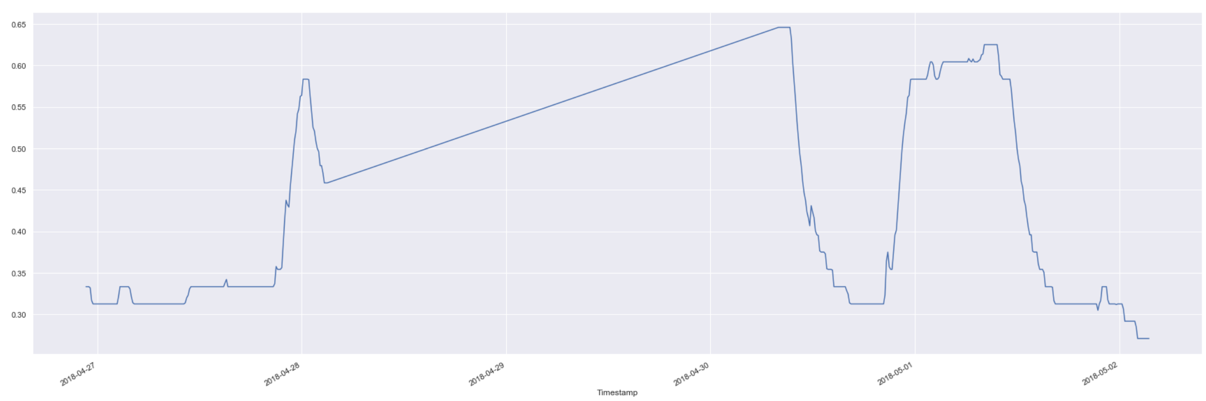


BREAKDOWN 5:

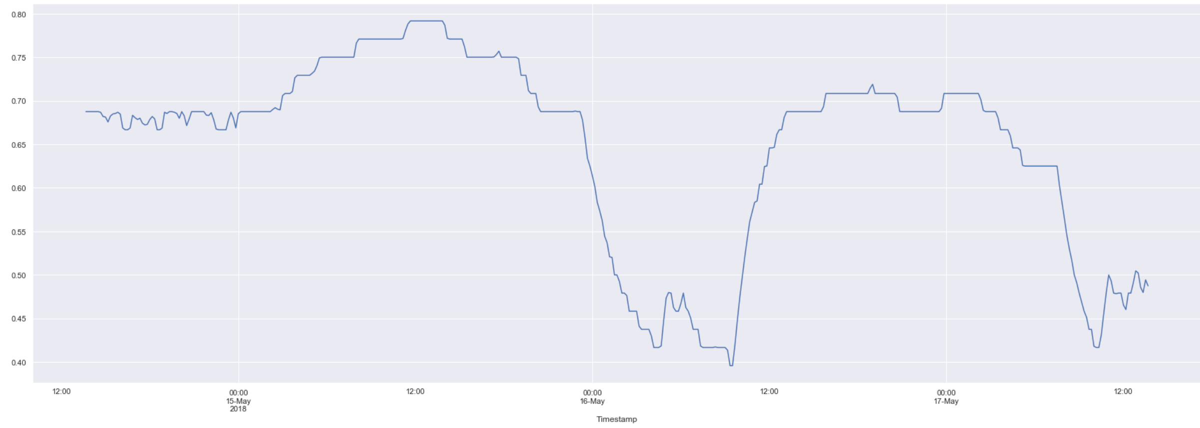


<u>TIME PLOTS OF HYDRAULIC OIL TEMPERATURE:</u>

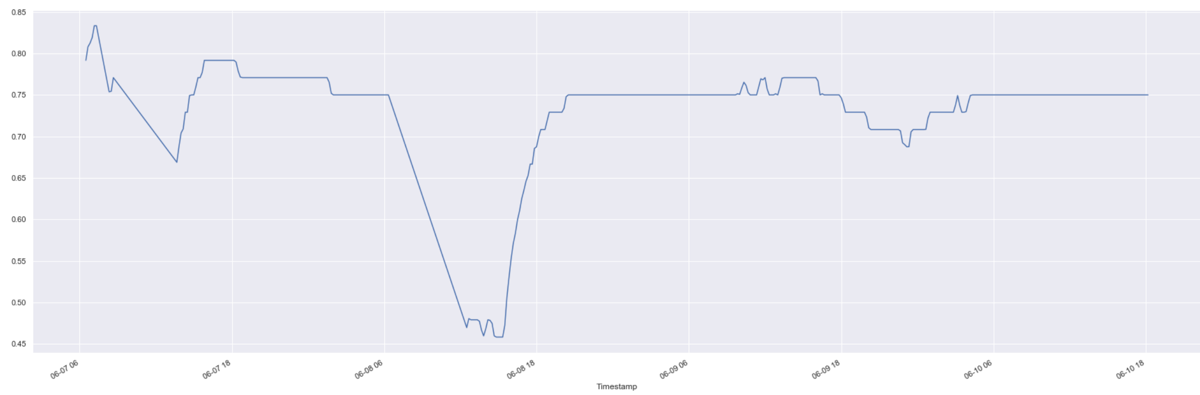
BREAKDOWN 1:



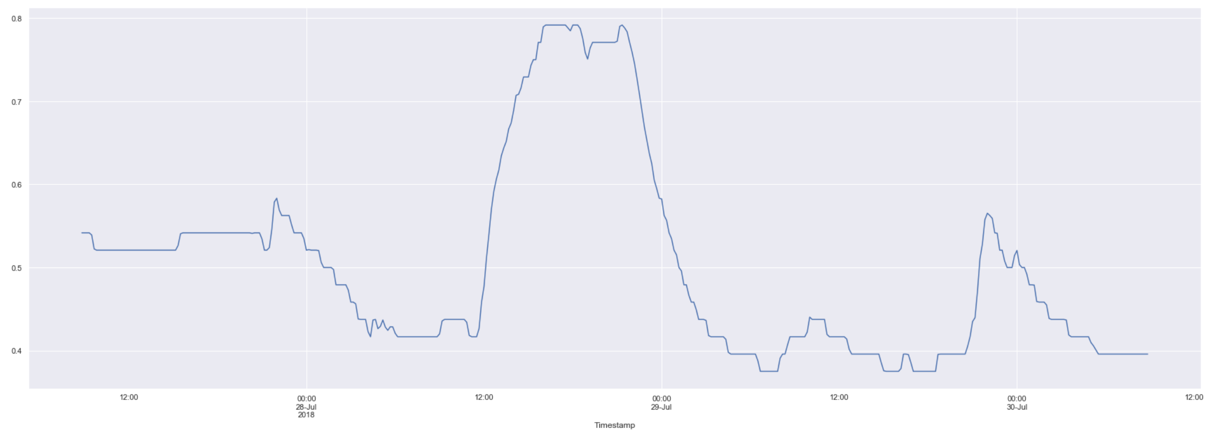
BREAKDOWN 2:



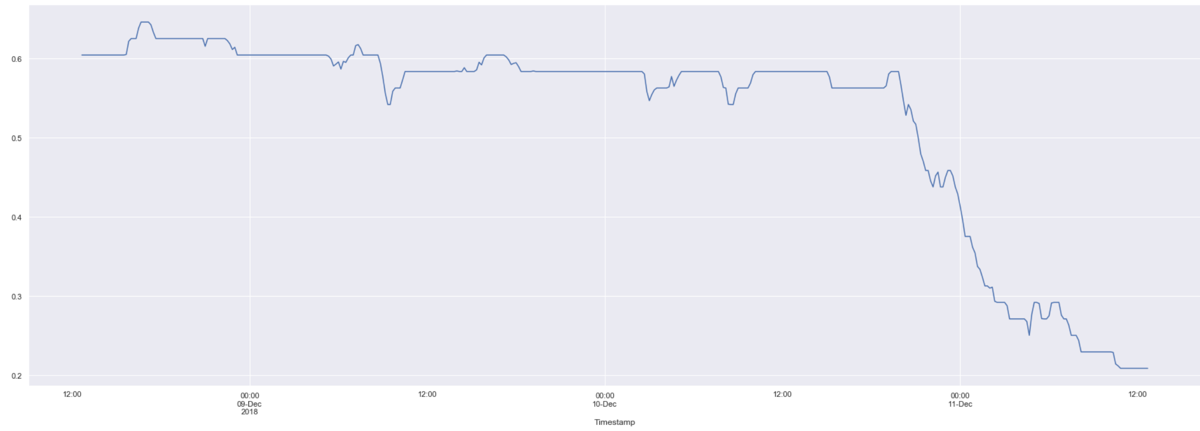
BREAKDOWN 3:



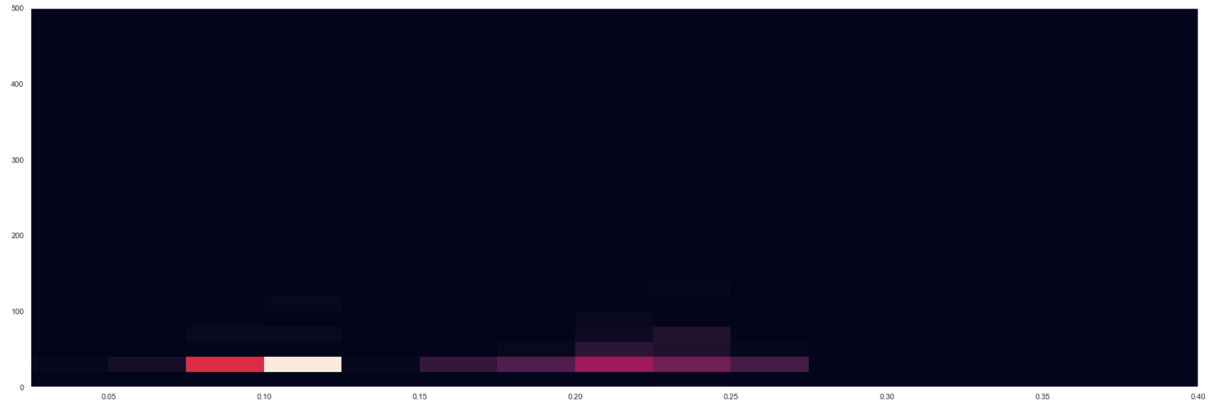
BREAKDOWN 4:

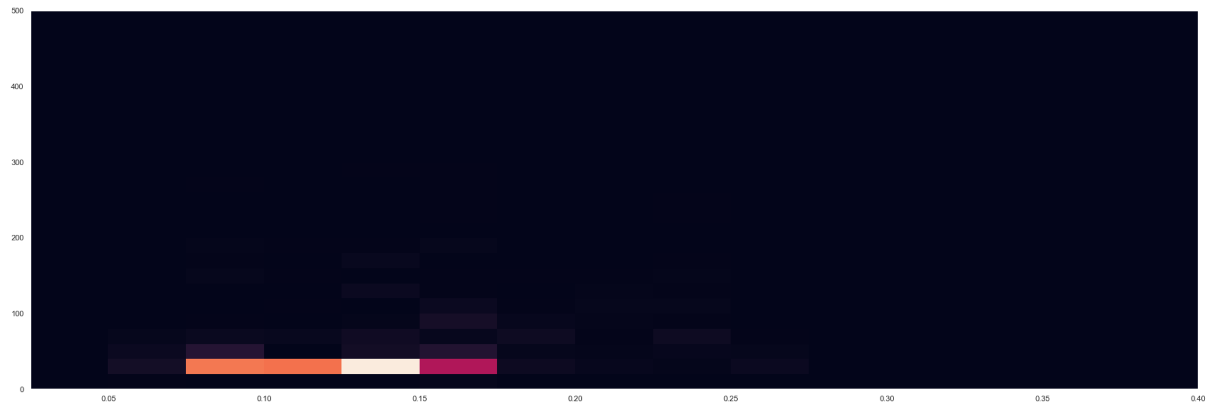


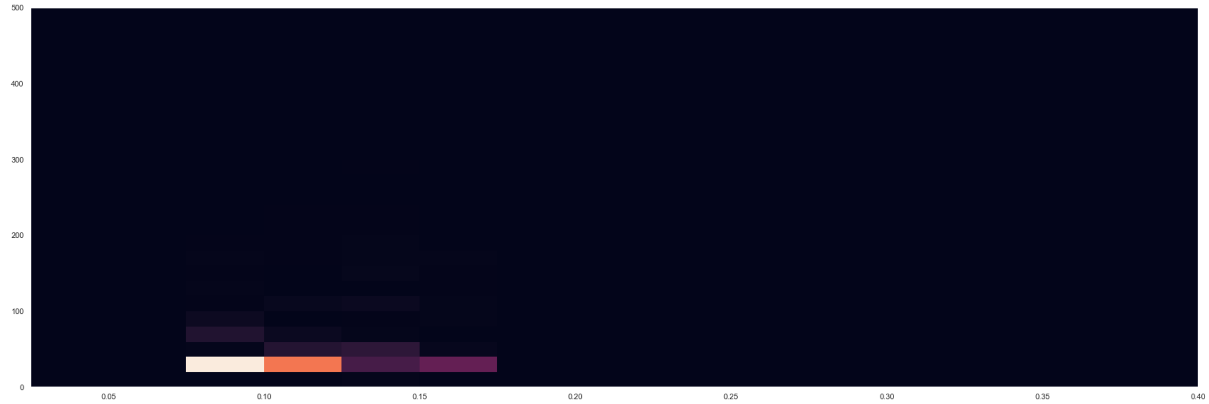
BREAKDOWN 5:

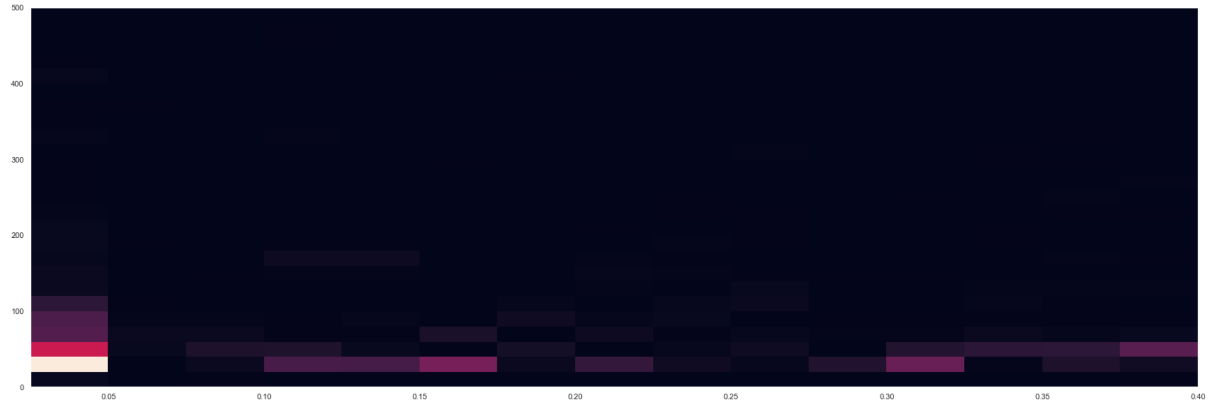


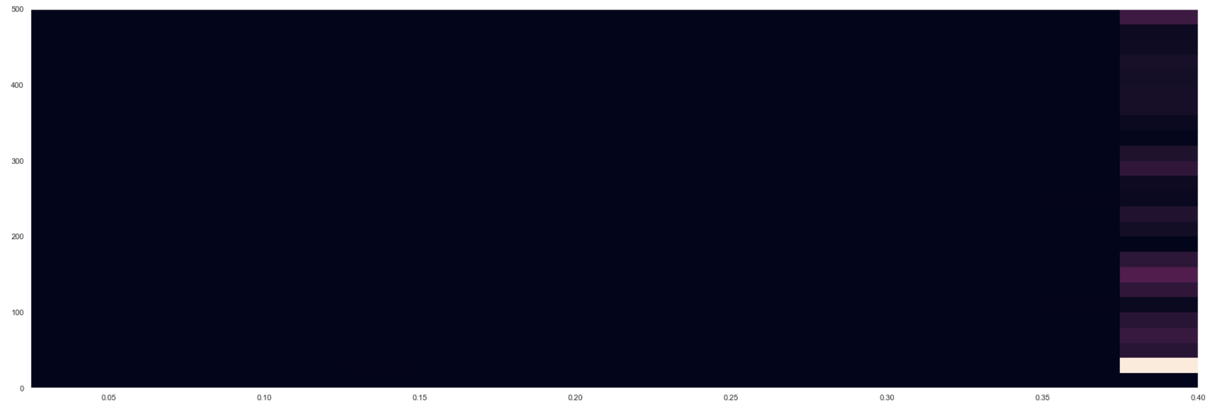
Generator rpm







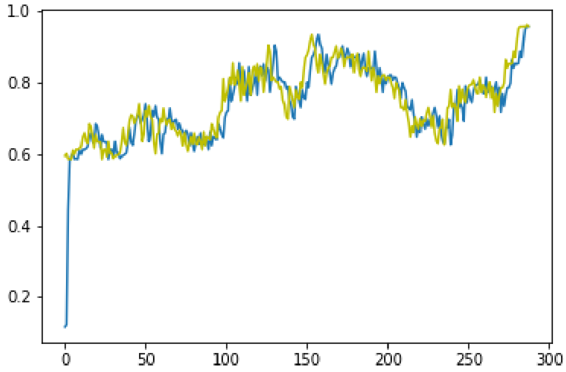


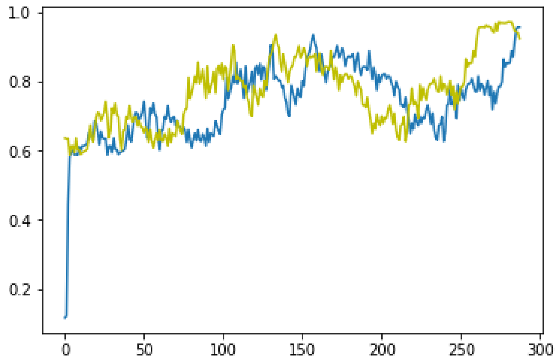


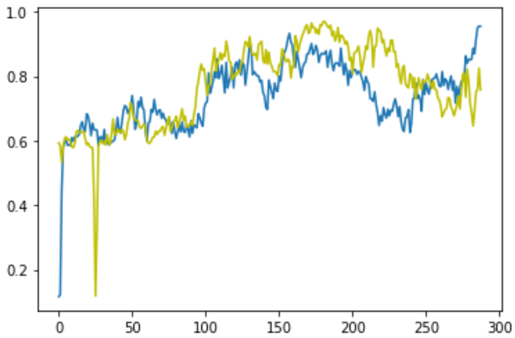
<u>OBSERVATIONS: ref = generator rpm[2000:2288]</u>
<u>PLOTS OF QUERY 1 - CANDIDATES RETURNED BY ALL DISTANCE MEASURES</u>

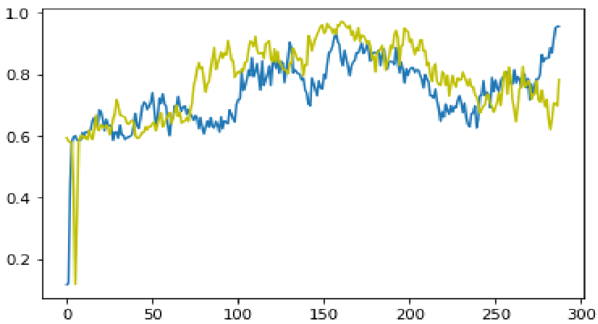
<u>DTW:</u>
Threshold = 2
No of candidates = 12

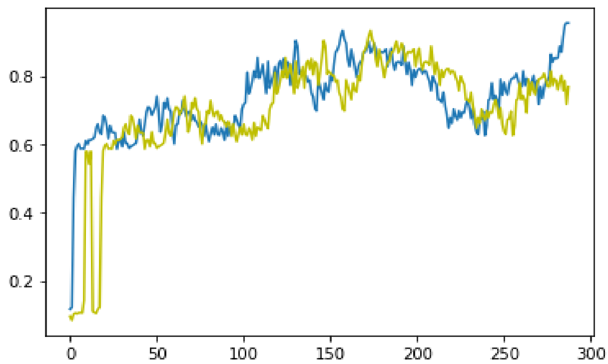
[('6/25/2018 14:40', 0.69), ('6/25/2018 18:00', 1.54), ('11/23/2018 0:20', 1.55), ('11/23/2018 3:40', 1.6), ('6/25/2018 11:20', 1.69), ('6/7/2018 13:40', 1.71), ('11/23/2018 7:00', 1.74), ('4/8/2018 9:30', 1.79), ('11/22/2018 21:00', 1.84), ('6/7/2018 17:00', 1.88), ('4/20/2018 21:20', 1.96), ('11/23/2018 10:20', 1.99)]
time for whole process is 451.29972434043884

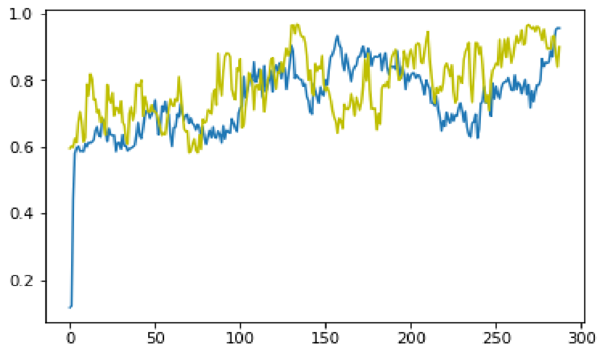






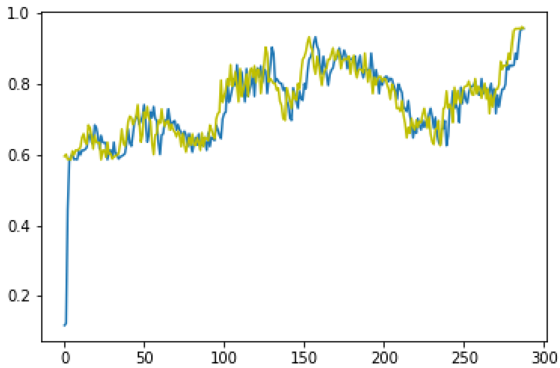


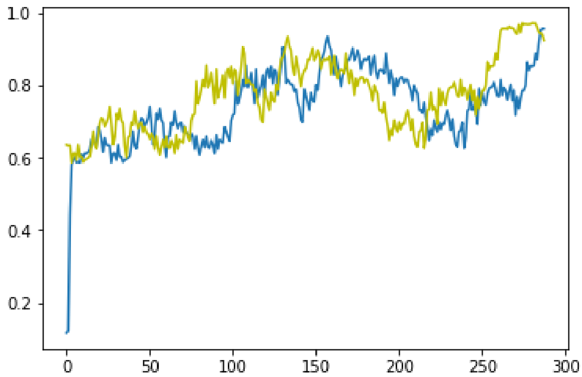


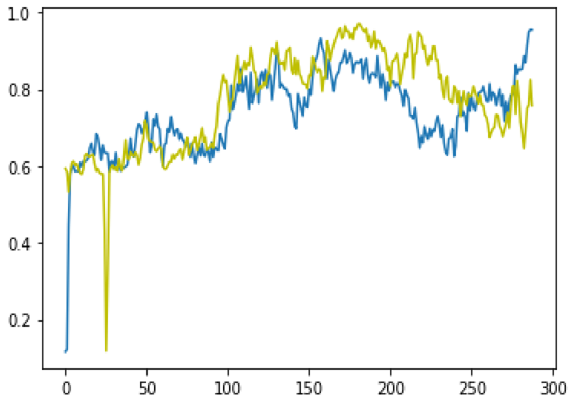


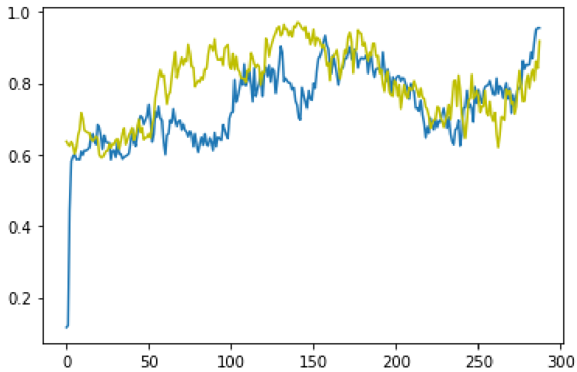
EUCLIDEAN

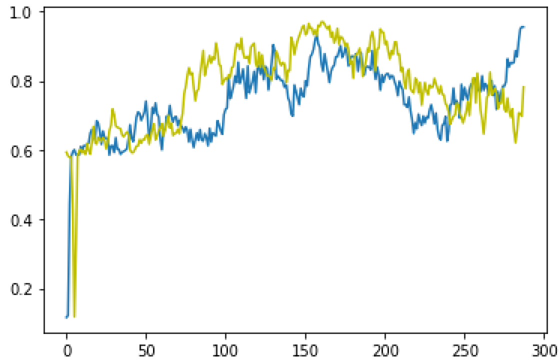
[('6/25/2018 14:40', 1.04), ('6/25/2018 18:00', 1.78), ('11/23/2018 0:20', 1.79), ('11/23/2018 3:40', 1.89), ('11/23/2018 7:00', 1.93)]
time for whole process is 0.7244181632995605
Threshold = 2
No. of candidates = 5









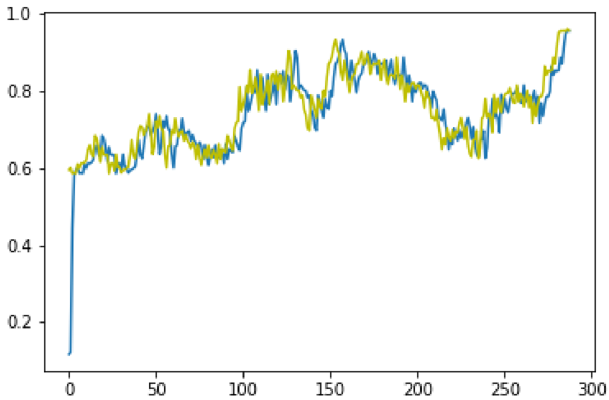


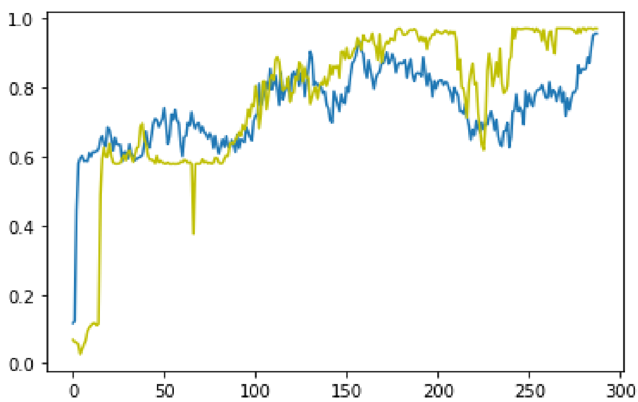
CORRELATION

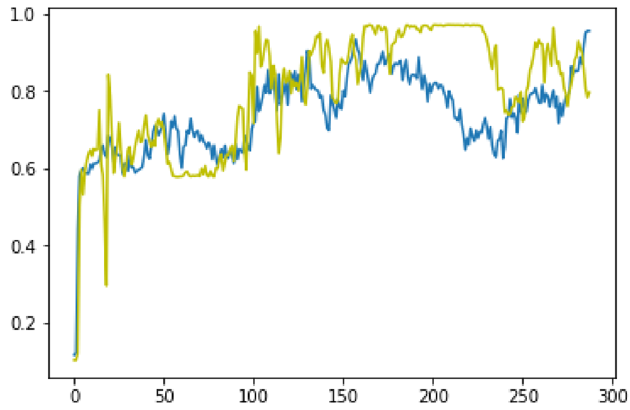
[('6/25/2018 14:40', 0.17), ('10/20/2018 6:30', 0.24), ('4/8/2018 9:30', 0.27), ('8/12/2018 22:30', 0.27), ('11/16/2018 11:50', 0.28)]

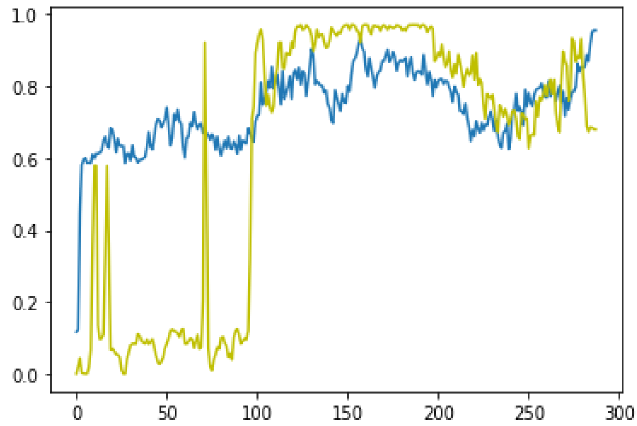
time for whole process is 0.9315962791442871

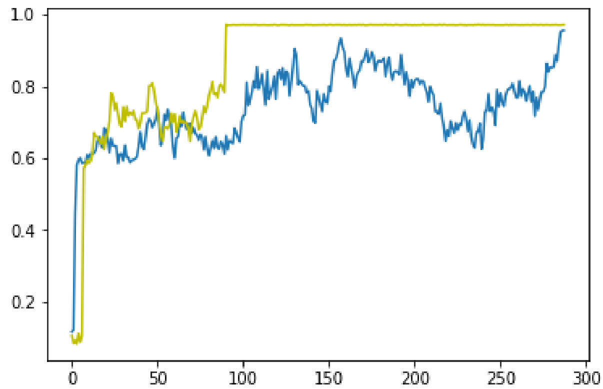
Threshold is 0.28





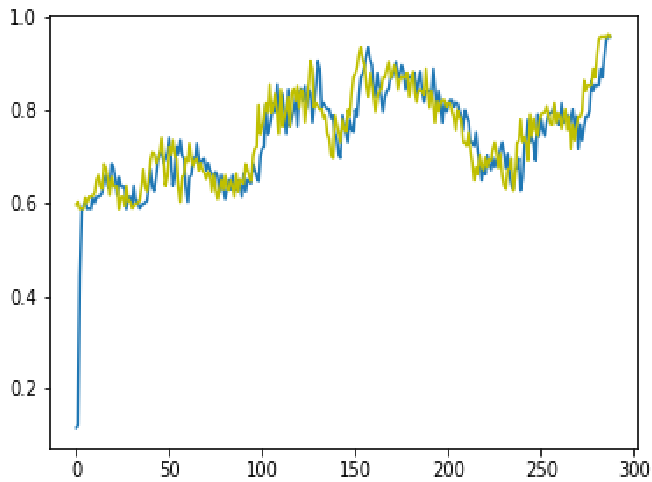


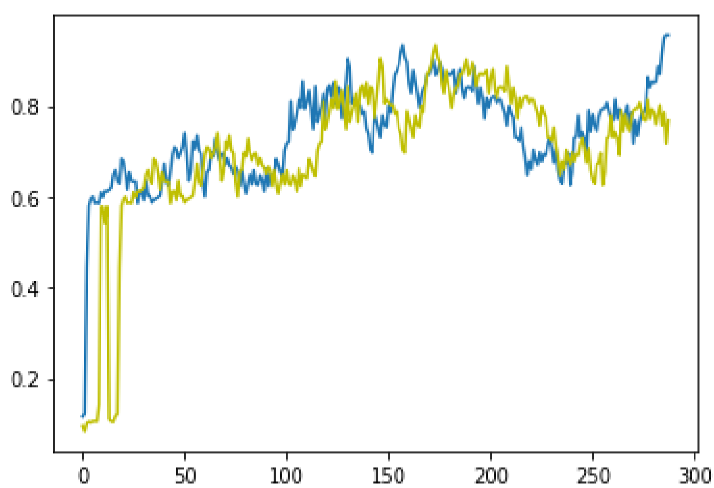


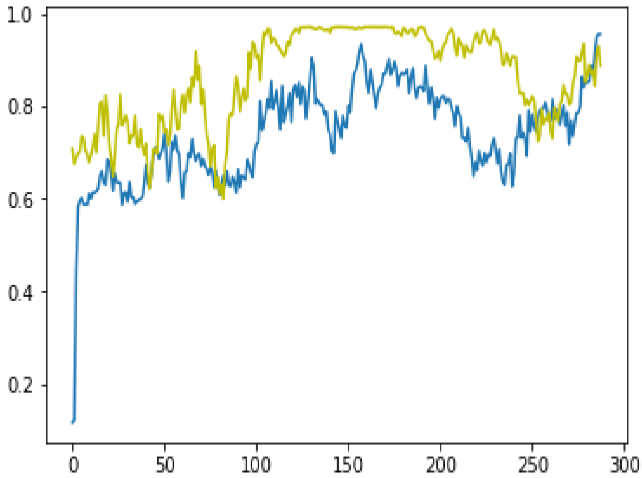


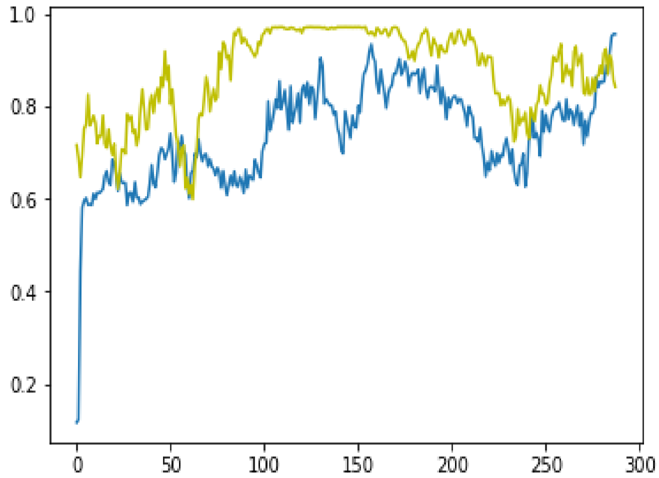
SAX - EDIT DISTANCE:

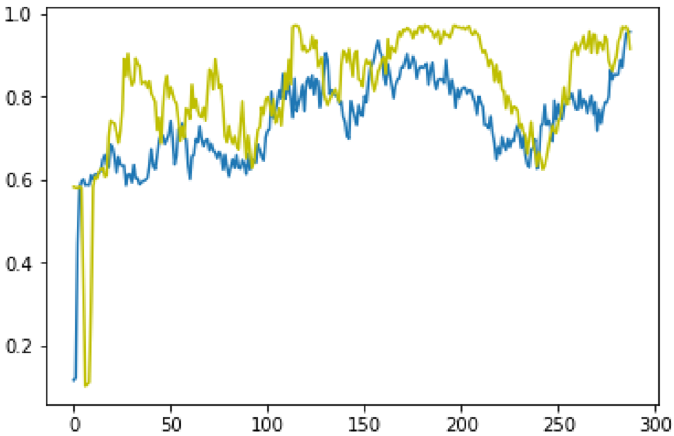
[('6/25/2018 14:40', 21), ('6/25/2018 11:20', 26), ('11/24/2018 19:40', 26), ('11/24/2018 23:00', 28), ('8/13/2018 1:50', 29), ('1/12/2018 22:40', 29), ('12/24/2018 14:00', 29), ('5/25/2018 16:50', 29), ('5/21/2018 15:40', 29), ('8/12/2018 22:30', 29), ('8/12/2018 19:10', 29), ('11/24/2018 16:20', 29)]
time for whole process is 30.173532009124756
Threshold is 29

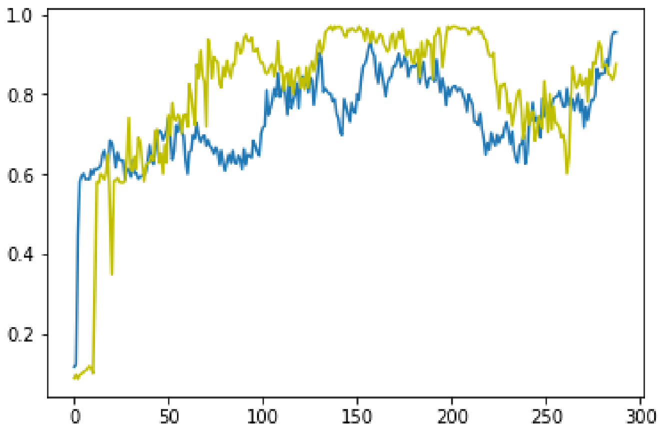


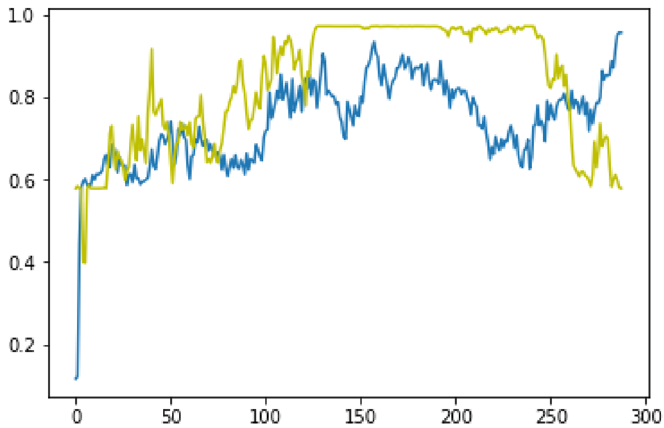


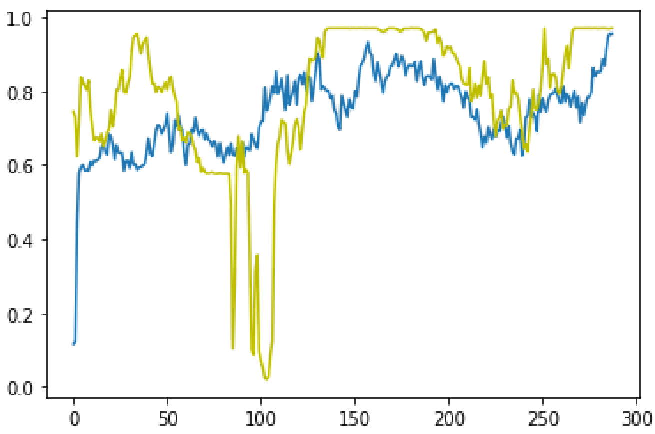

1.0
0.8
0.6
0.4
0.2
0
50
100
150
200
250
300

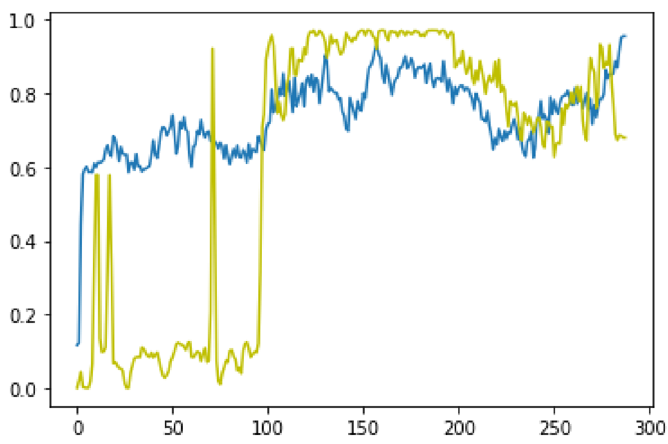

1.0
0.8
0.6
0.4
0.2
0
50
100
150
200
250
300

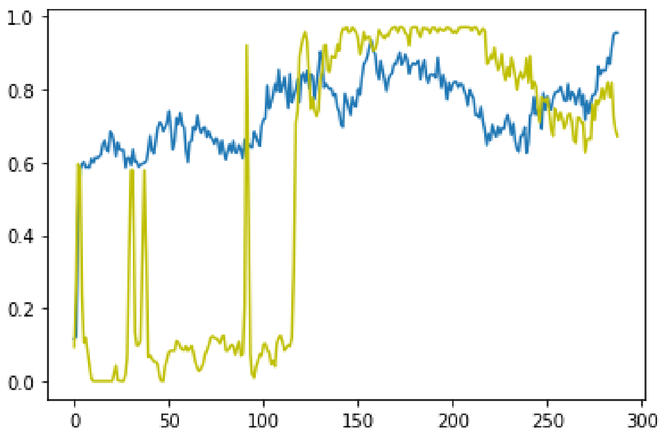

1.0
0.8
0.6
0.4
0.2
0
50
100
150
200
250
300

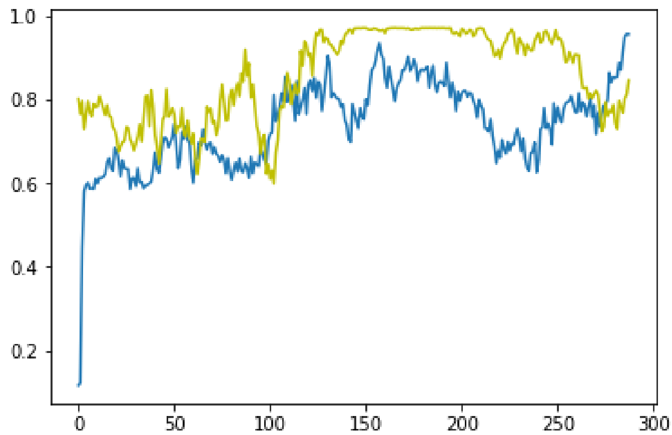

1.0
0.8
0.6
0.4
0.2
0
50
100
150
200
250
300


1.0
0.8
0.6
0.4
0.2
0
50
100
150
200
250
300





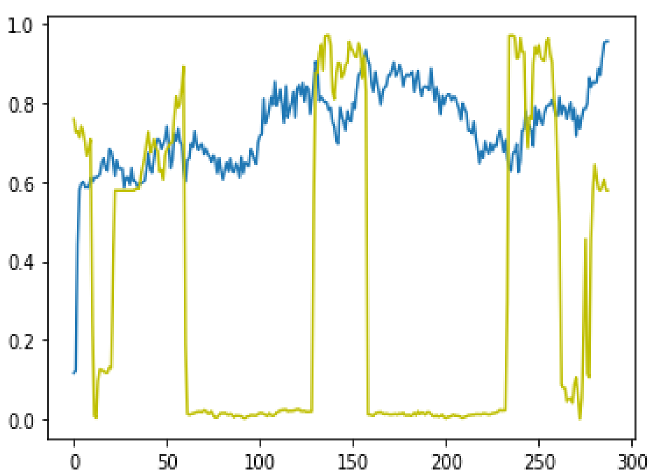


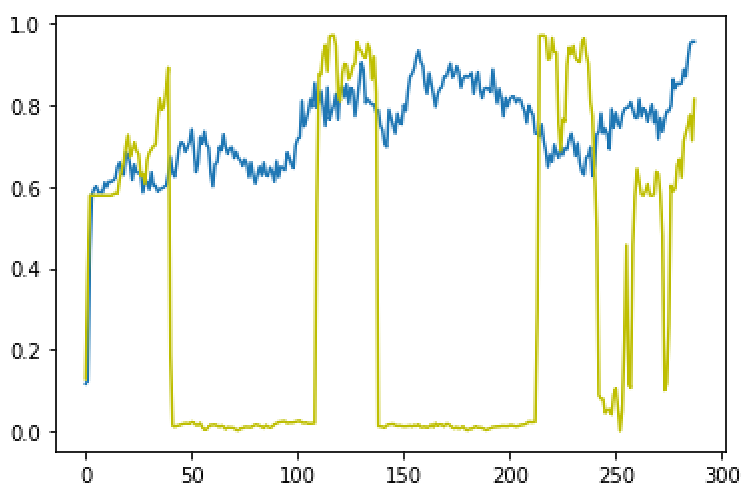


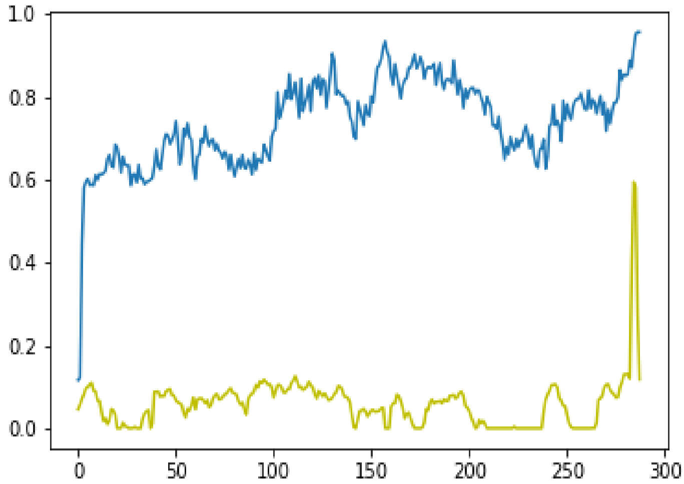
FFT - COHERENCE:

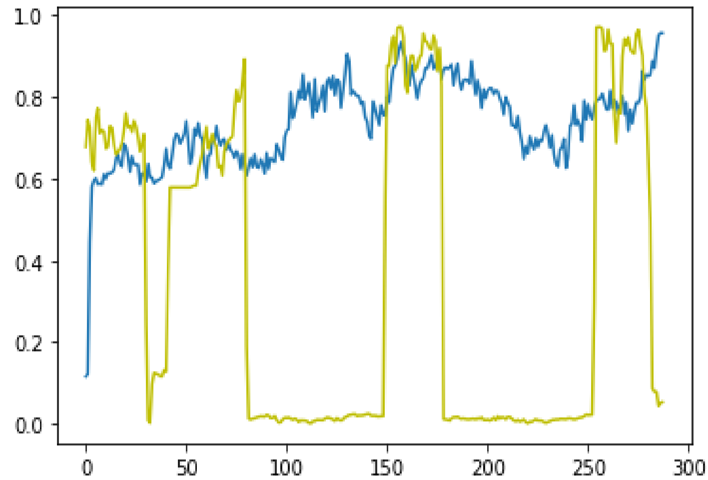
('7/26/2018 5:20', 0.64131), ('7/26/2018 10:00', 0.67235), ('8/1/2018 16:00', 0.68886), ('7/26/2018 2:00', 0.72023),('7/26/2018 13:20', 0.72446), ('7/26/2018 16:40', 0.72495), ('7/4/2018 20:40', 0.74217), ('7/4/2018 17:20', 0.7522), ('7/4/2018 14:00', 0.75404), ('7/5/2018 6:40', 0.75445)

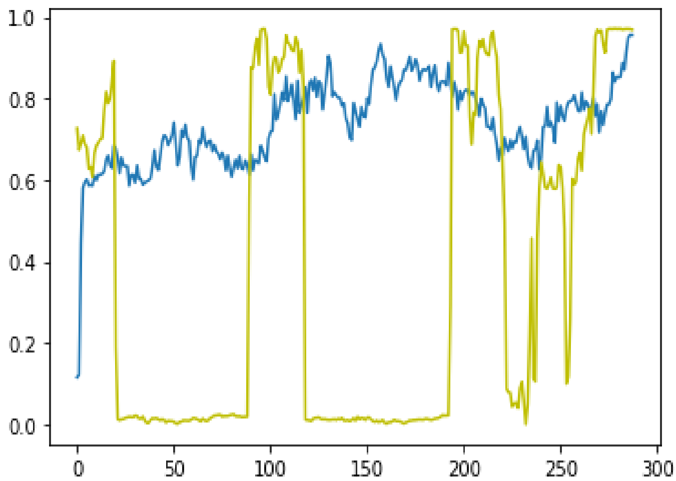
Time for the process: 19.73748803138733

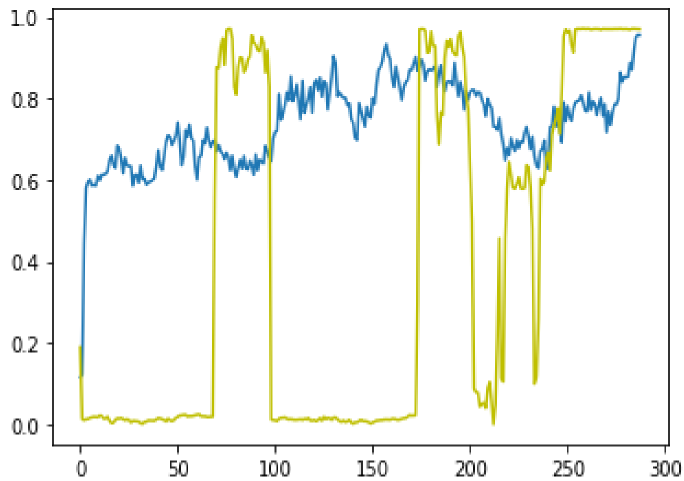




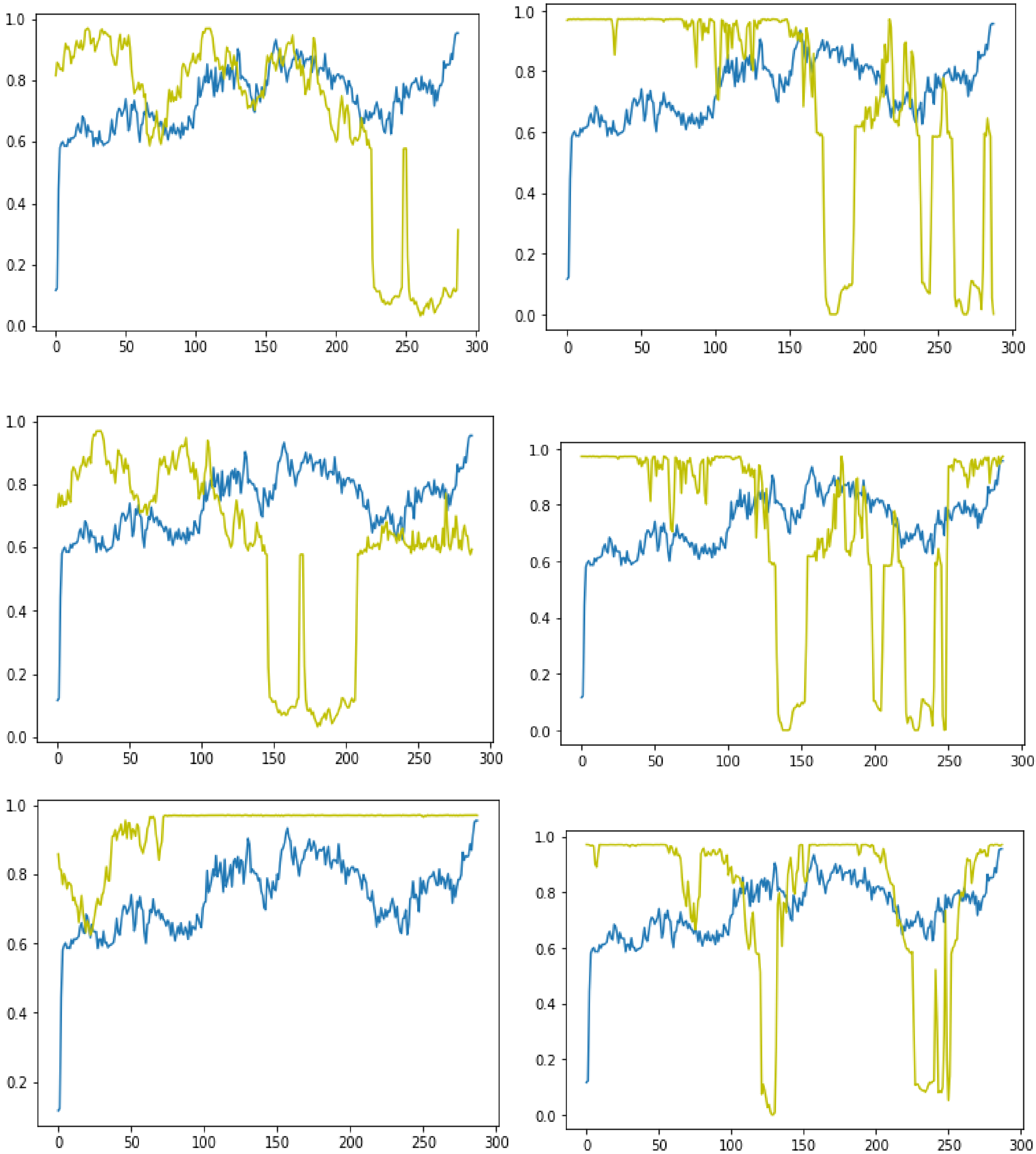


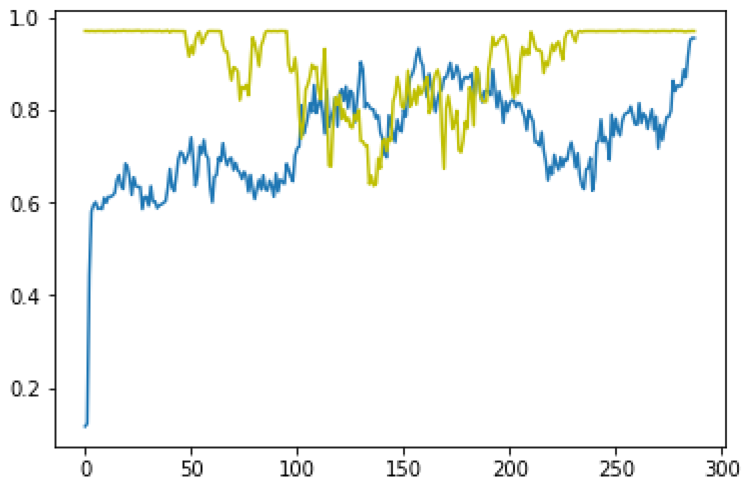






SSIM ON SPECTROGRAM IMAGES:


0.0 0.2 0.4 0.6 0.8 1.0
0 50 100 150 200 250 300



1.0
0.8
0.6
0.4
0.2
0
50
100
150
200
250
300