LNA Dewar Monitoring Project Progress 4

In the previous report, I briefly compared the results of applying different transforms on the Dewar sensor data and concluded that the Hilbert Huang Transform (HHT) provides the best spectral resolution. From there on, I mainly worked on feature extraction from the HHT result.

Feature Extraction

As discussed earlier, the HHT is applied over a sliding window of size 500 samples over the downsampled sensor data. The downsampling factor is 100.

The result of applying HHT on this window is a time-frequency matrix of order N x I where i^{th} column represent the i^{th} time sample of the sliding window and the n^{th} row represent the magnitude of the n^{th} frequency component. Currently N and I are both set to 400.

Taking the individual integrals of all the rows over the complete length of columns provides a vector of order N x 1 where the n^{th} row represents the energy contained in the n^{th} frequency component over the complete sliding window. Now this vector in a sense provides the energy-frequency profile of the system at a particular position of the sliding time window. When the sliding window is moved forward by one step, we get a new energy-frequency profile. By saving the new energy-frequency profiles in a separate vector, we obtain the time varying signals of the energy content of each frequency over the complete data set. We name this vector as freqTrend.

The sliding window can be moved along the considered data set until the data set terminates so the number of steps that the window moves depends on the length of data set. In total I have been able to identify and extract 7 data sets that contain a maintenance event.

The approach for feature extraction was as the following:

1. Visually identify the time-step of the sliding window just before the maintenance event appears and note it down. We name this time step as : eventStart

   Also visually identify the time-step of the sliding window when the maintenance event passes away in the sliding window and note it down. We name this time step as: eventStop

2. Find the difference in integral of freqTrend for samples from (eventStart–30: eventStart) and (eventStop:end) for all frequencies. We name this difference in integral as: diffInFreqTrend.

   diffInFreqTrend is again a vector which gives the difference between the energy content of a short window of 30 samples before the event and for all samples after the event for all frequencies.

3. Sort the diffInFreqTrend vector in descending order so the frequencies with the highest difference come to the top.

4. Repeat 1 to 3 for the three training data sets and then compare the sorted results of diffInFreqTrend.

5. Select a range of frequencies that is amongst the top in the sorted result and which are also common for the results of the three data sets.
6. Select the integral of the HHT result over this identified range as the feature.

The basic approach is that we are expecting some frequencies to have a higher magnitude just before the event and less magnitude after the event is finished. To quantitatively identify these frequencies, I took the approach of finding the difference between area under the curve (AUC) just before the event and AUC after the event and then sorted them.

Results

Following are the results for Stage 1 Temperature signal of the process as mentioned above.

S1 Temp				S2 Temperature				Pressure		
Data Set 1	Data Set 2	Data Set 3		Data Set 1	Data Set 2	Data Set 3		Data Set 1	Data Set 2	Data Set 3
94	29	11		95	46	7		121	26	17
95	19	12		94	4	8		122	25	18
96	28	10		96	22	6		108	27	16
93	20	9		97	3	5		279	24	19
97	21	13		98	45	4		259	28	13
98	30	14		93	23	52		271	23	15
99	18	8		99	17	51		278	29	12
100	31	7		80	18	12		220	22	14
101	32	6		79	5	53		272	30	11
92	27	5		81	47	54		260	31	9
102	33	4		78	21	50		280	32	23
86	22	3		77	19	3		277	33	20
103	26	93		76	20	11		345	34	22
88	23	81		83	24	55		357	35	21
87	34	94		84	16	49		346	36	24
91	25	2		75	6	2		358	89	8
89	24	82		82	25	56		400	37	10
104	93	98		100	2	60		347	38	25

105	17	15		85	10	58		344	88	6
90	35	95		92	7	63		359	21	28
85	92	92		91	9	57		327	90	5
106	91	99		86	44	59		397	39	27
107	90	97		101	8	86		360	91	31
108	94	96		74	11	92		348	87	30
109	36	109		102	26	91		364	92	7
110	89	106		90	49	62		391	40	26
183	95	80		153	15	87		396	86	29
174	86	100		69	50	61		301	41	4
212	87	110		137	48	85		399	84	3
286	37	91		88	27	93		328	85	33

The values in the rows correspond to the index of the respective frequency in freqTrend. Only the first 30 out of 400 rows of the sorted diffInFreqTrend are shown above.

For S1:

It can be seen that although the sorted result is not the same for all the datasets, frequencies having index from 90 – 100 are dominant as well as are common between the data sets. We therefore select 90-100 as feature range for S1.

For S2:

It can be seen that 90-100 range is still dominant in data set 1 and 3. However, in data set 2 this range is not very dominant. When we go down till row 60, we notice that a range 80-95 appears in data set 2. As this range is already dominant in data set 1 and 3, we select 80-95 as feature range for S2.

For P:

For pressure, we cannot see any significant commonality between the data sets. The dominant ranges are very different for each data set. Therefore, without much justification, we select a range of 120:200 .

Transformation to Feature Space

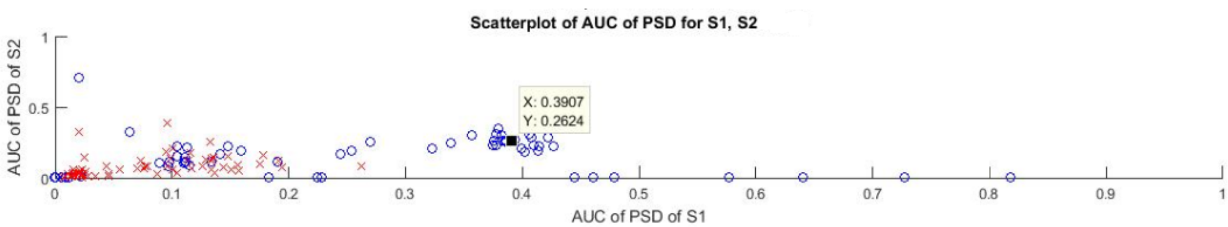
After selecting our features, we run the HHT over the sliding window over the complete data set again. As defined, our features will be the integral of the HHT result over the selected frequency ranges.

The features were normalized before plotting by dividing them with their maximum values as observed in data set 1 for all the data sets.

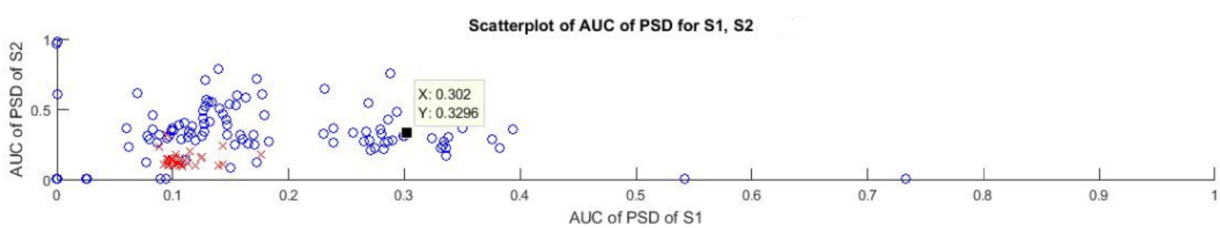
As there was no significant trend identification in the pressure time series, we only plot a scatter plot for the S1 and S2 features.

Following are the scatter plots:

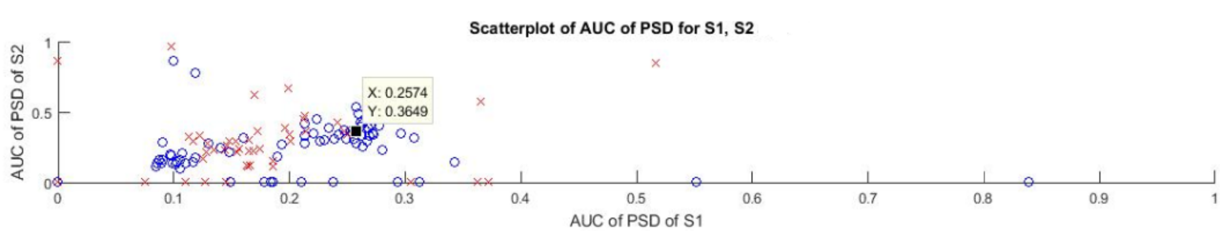
Data Set 1



Data Set 2



Data Set 3



The blue points are plotted before the event appears and during when the event is visible in the sliding window. As the event passes, the color of points is changed to red.

By intuition, we expect different clusters to form for points just before the event and for points after the event.

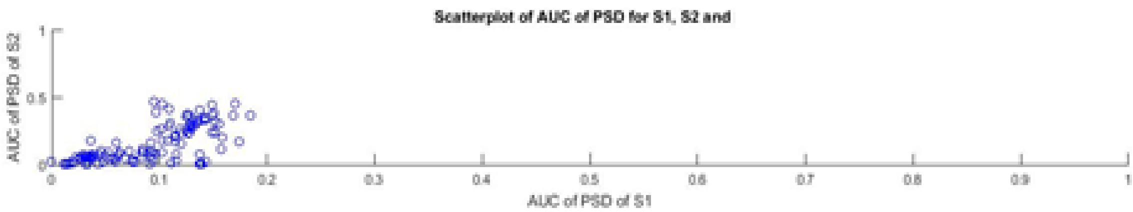
Analysis

The scatter plots of the three data sets seem to maintain cluster formation with cluster spacing best in data set 1. Overall, we can say that the clusters can be divided by a line drawn at AUC of PSD of S1 = 0.2. Indeed, we observe no significant differentiation along the y-axis.

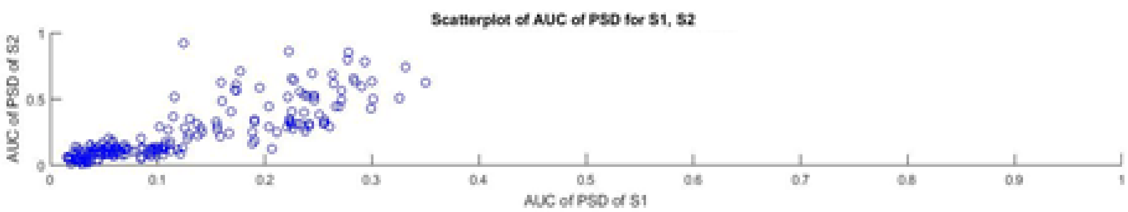
Before applying a training algorithm for a classification boundary, I also plotted the scatter plot for some other data sets.

For the following data sets, there were no maintenance events at all:

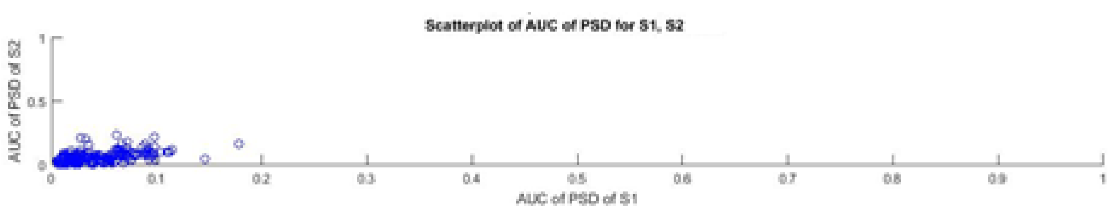
Data Set 4



Data Set 5



Data Set 6

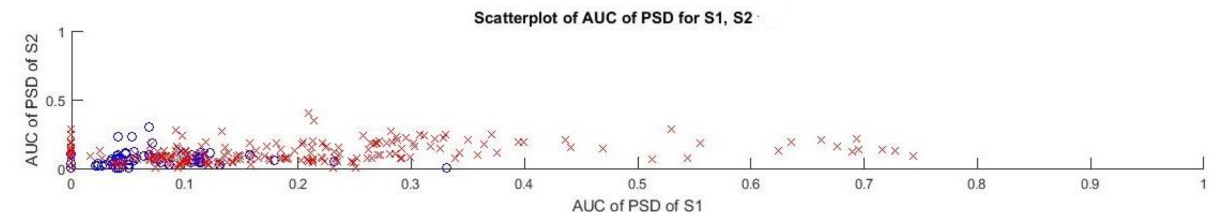


For data set 3 and 6, we notice a clustering before AUC of PSD of S1 = 0.2. This meets our expectation as with no maintenance event, we expect the energy content of frequency range 90-100 to be very less throughout the data set.

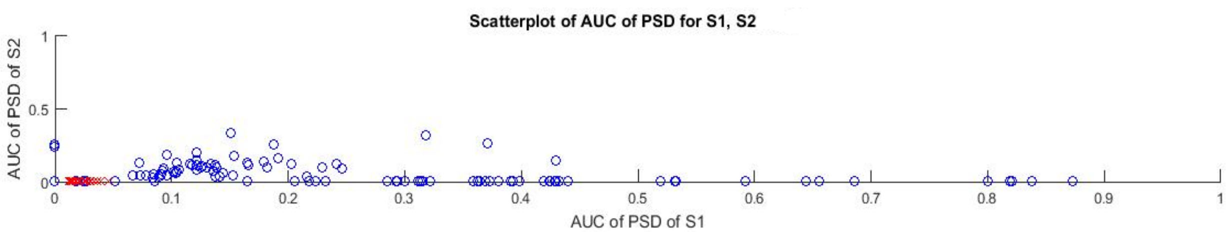
However, for data set 5, we notice the points to go beyond the 0.2 threshold. This observation do go against our expectation.

Some more data sets containing maintenance events were also considered. Following are their scatter plots:

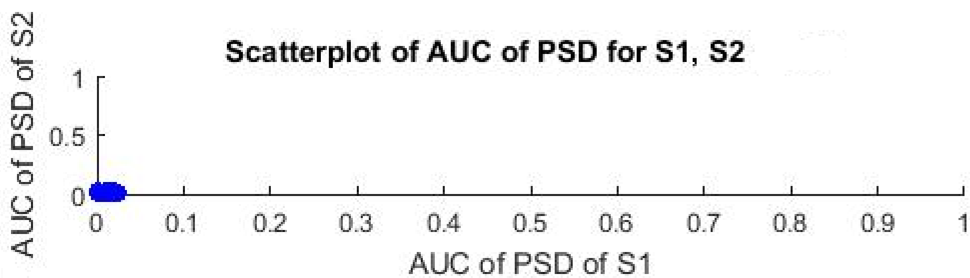
Data Set 7



Data Set 8



Data Set 9



For these data sets 7,8 and 9 no reliable cluster formation has taken place. This implies that the identified frequency range is not dominant in these data sets.

Overall, the results from these features indicate that although the identified frequency range is suitable for some data sets, it does not hold true for all the data sets. Therefore these features are unstable and in order to have good repeatability of results, more stable features have to be identified. Currently, I am working on finding stable features from the HHT result.