**LNA Dewar Monitoring Project Progress 6**

I would like to update you with the progress of the Dewar Monitoring project so far. As I was a little late to request for an extension of contract in December, I could not get the contract for January so I have continued the work from February on wards.

In the last update, I informed you that I will be working on time domain methods. I explained that the approach is to apply some statistical tests on the residuals that are evaluated as difference between the actual Stage 2 temperature and its predicted value.

**Data Model**

My data model for prediction of Stage 2 temperature primarily consists of the pressure and environment temperature as input features.

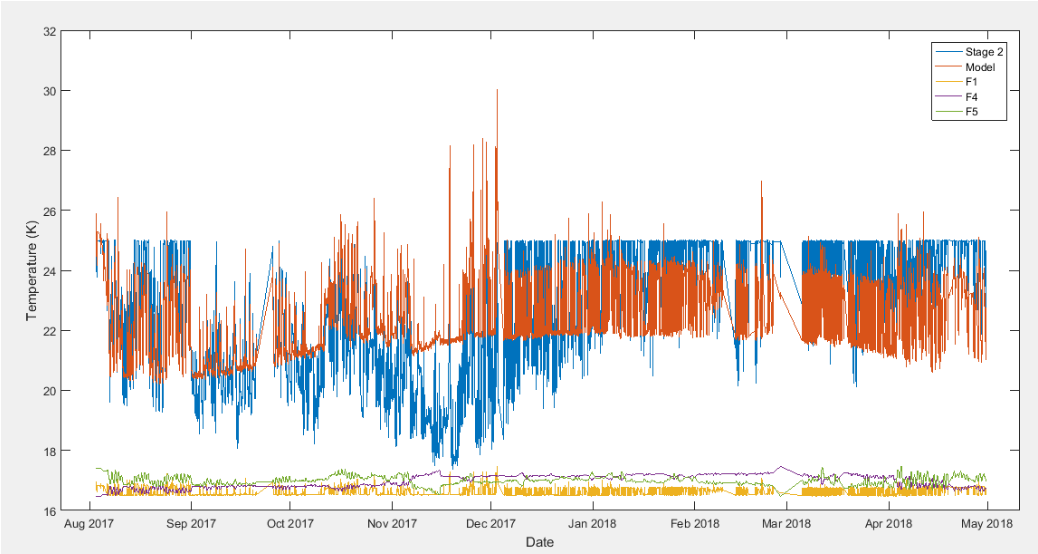
I have derived two features from the environmental temperature, namely a feature that filters the long term seasonal variation in the environment temperature and another feature that is the distance of the environment temperature from the threshold value.

```
F1 – Pressure
F4 – Distance of environment temperature from threshold
F5 – Environment temperature without seasonal variation

eqn = theta_m_sol(1,1)*f1(:,1).*(1+f4) + theta_m_sol(1,5)*f5(:,1).*(1-f4) + +
(theta_m_sol(1,4)*f4) + theta_m_sol(1,3);
```

The data model is trained on snippets of data of Stage 2 temperature where the event is not observed and for which the temperature is below the threshold (25K). The training was done to evaluate the weights `theta_m_sol` that provide the least RMS residual between the value of the data model `eqn` and the actual value of stage 2.

You can see below the results of training with a residual of 6.74e+05 between the model and stage 2 data. Overall we see that model stays within the mean value of the data and responds to short term variations. However, we also see that around the months of November, the model over predicts the temperature value.




We can expect a large residual in these months which would certainly result in false alarms. I have considered this issue by accepting only those values of data model for statistical tests for which the actual stage 2 temperature is greater than these values. I think it is reasonable to do this as we our more interested to avoid high temperatures than lower ones.

**Statistical Tests**

After the training stage, I have the data model. For each time instant hence I can find the residual between my model and the actual value. Now I can apply different statistical tests on this residual and evaluate their performance.

**I have included the plots for residuals of all of my test data sets in the last page.**

To get a metric for comparison their performance, I manually divided my data in to three classes:

- Before maintenance
- After maintenance
- No maintenance

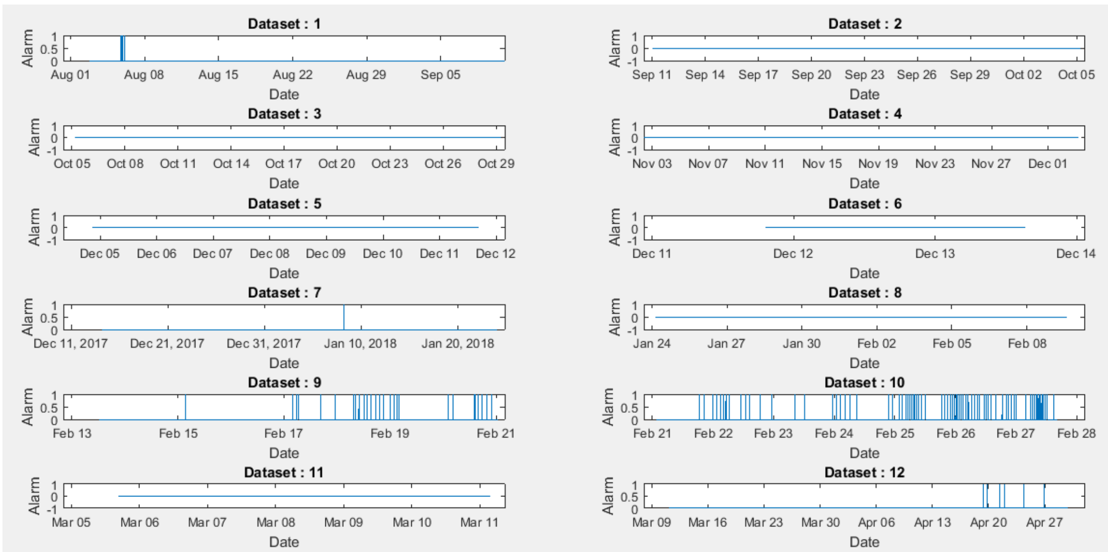
I sorted the data visually so that a given time series that come before an event is labelled as being 'Before Maintenance', the one coming after as 'After maintenance' and so on.

The tests that shall be applied will take residual of that time instance as input and will output either an Alarm or No alarm based on some statistical processing. If the test excites an alarm in a time series that was labeled as After Maintenance or No maintenance then it will be regarded as a False Alarm. If the test excites an alarm for a time series that is labeled as Before Maintenance, then it is a true alarm. I can count the number of total false alarms and divide it by the total alarms to get the probability of false alarm.

Uptil now I have applied two tests:

- CUSUM
- LS CUSUM

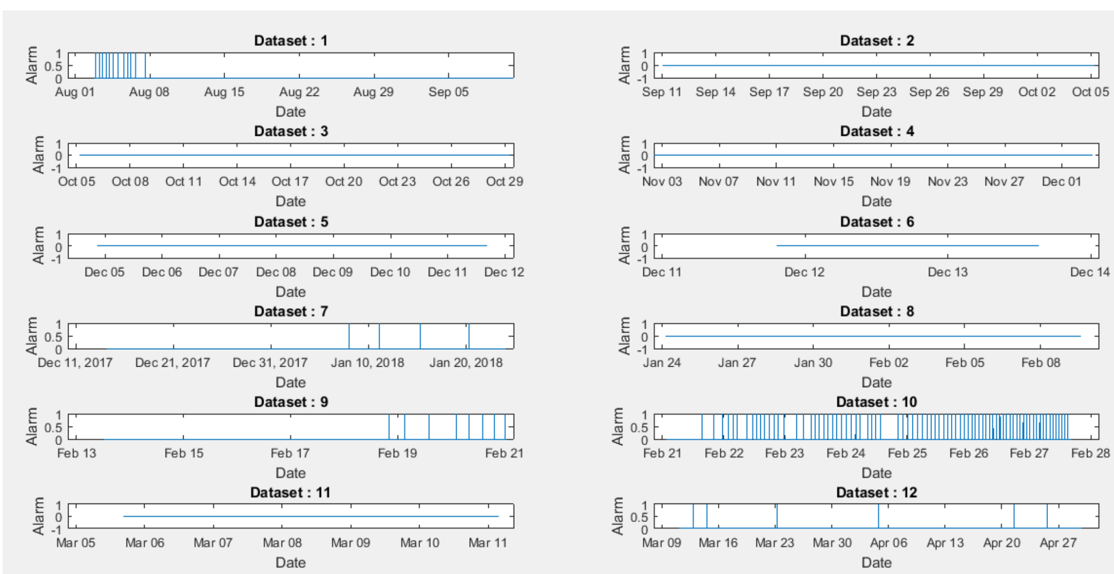
For CUSUM, following are the results of alarms:




The blue bars here represent individual alarms. The probability of false alarm here is 0.333. You can compare the alarms with the residual plots in the last page to get a better idea.

For LS CUSUM:

The probability of false alarm is here 0.2727. This test is a variation of CUSUM that takes into account the sum of previous residuals as well.



**Analysis**

A general analysis that I can make is that the approach works fine for large deviations of the stage 2 temperature over the computed model value. This is evident in the case of Data Set 10 where many alarms are produced. It should also however be noted the approach has not taken into account the missed alarms as yet. A complete comparative analysis can be made when we also take into account the probability of detection of each test.

I am now working on applying more advanced tests for better results. I am currently limited by the available data sets. You'll realize that this approach requires data of both the dewar sensors as well as external environment for each time instant.

**Residual plots**
**Red curve: data model**
**Blue curve: Stage 2 temperature**