LNA Dewar Monitoring Project Progress 5

In the previous report, I discussed on the approach of extracting features from the Hilbert Huang Transform results computed for sliding windows over the sensor data series. It was evident that although the extracted features worked well for certain data sets, they gave poor results for other data sets. As discussed in the last meeting, I am now investigating more about time domain methods that can be applied over the time series of sensor data for change detection.

Basic Time Domain Approach

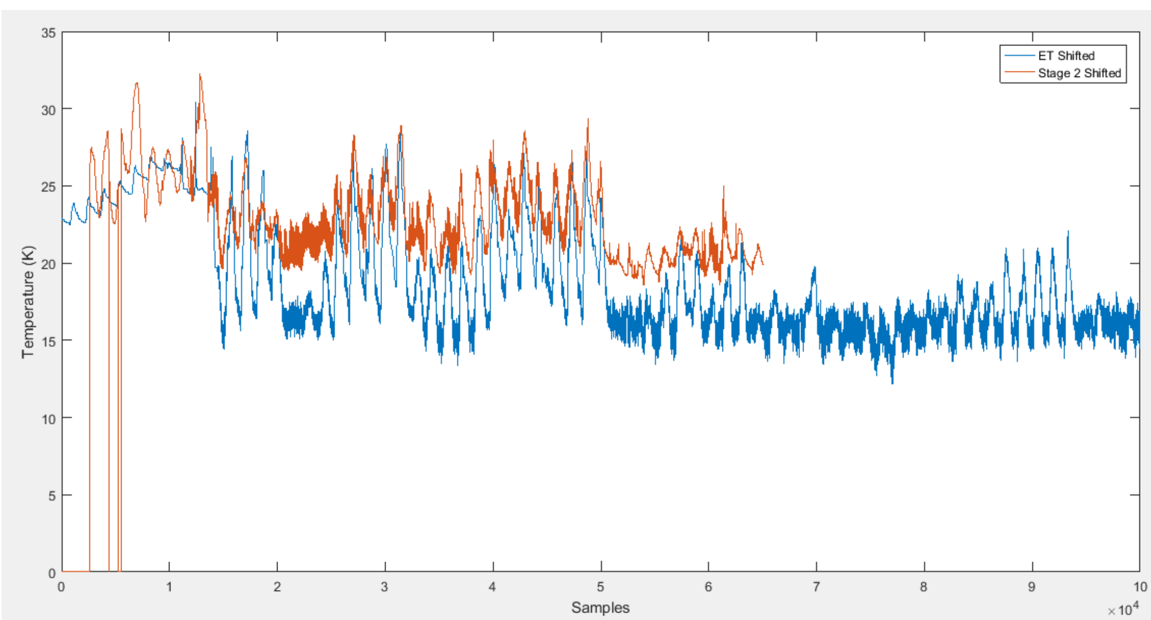
In the general time domain approach, one formulates a data model for the quantity that is to be measured and then finds the residuals between the predicted values and actual measurements. Then after, a statistical test is applied over the evaluated residuals and based on the thresholding results of such a test, change detection in the time-series is sought.

After the last meeting, I focused my analysis on Stage 2 (20K) Temperature and Pressure (P) sensor readings. I also considered the Environment Temperature (ET) sensor readings. It is evident that the Stage 2 readings are influenced by both daily variations of environment data and periodic variations of the pressure data.

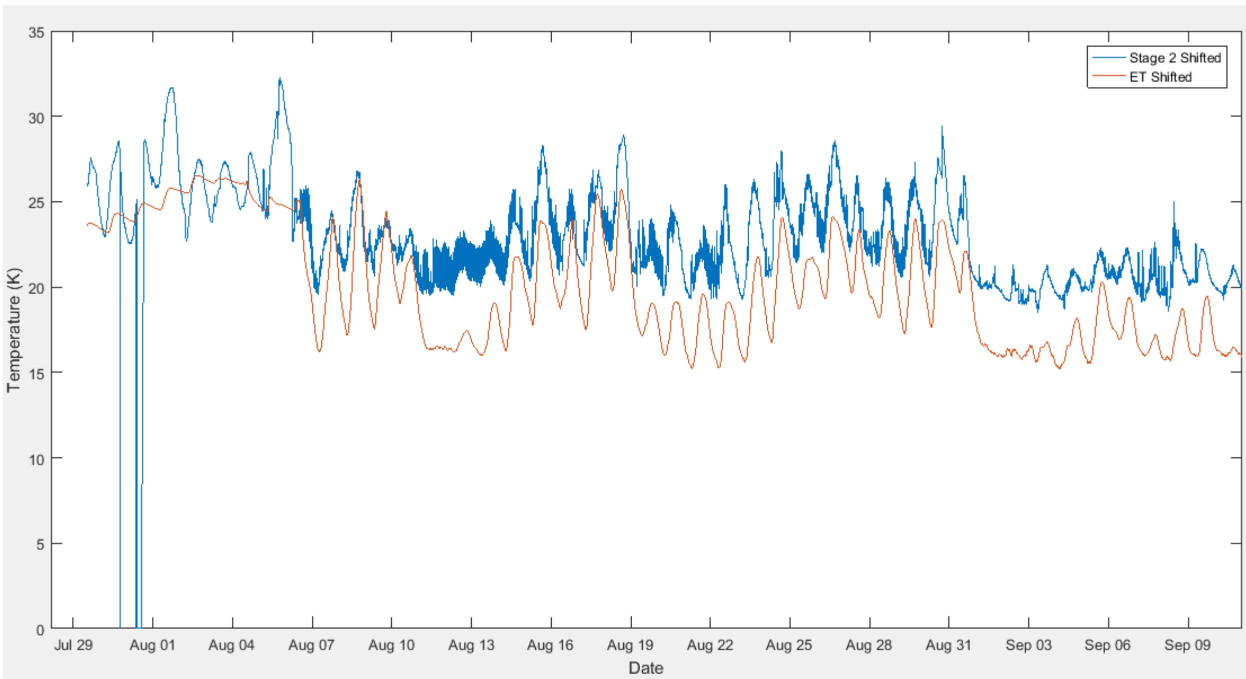
Thus, an intuitive basic model for Stage 2 can be derived by simply considering Stage 2 to be a sum of the scaled ET and scaled P measurements shifted by an offset. One can then find the residual between the output of this model for a given time instant and the actual Stage 2 measurement and then apply a statistical test over it.

Stage 2 and Environmental Temperature

The Stage 2 has a sampling period of 15 secs while the ET is sampled every 60 secs. Therefore, in order to align both the measurements, I down sampled Stage 2 by 4. Also, as the divisions of the data log files cover different time intervals for Stage 2 and ET, it was required to shift Stage 2 and ET with respect to each other to align with each other as shown below:



- The shifting was done based on the timestamps available in both the time series.
- It is to note that the ET measurements contain noise of its own. Therefore, any linear combination of it with Pressure would introduce further noise. To avoid this, the ET measurements were Kalman filtered to provide smooth daily variations.
- The sampling period of the Stage 2 temperature can be seen to deviate to 14 secs or 16 secs over time before it is again corrected. Therefore, if the whole Stage 2 acquisition is selected at once and aligned with ET, then there will be some offsets remaining between the peaks of both. To avoid this, the aligning operation was moved into a user defined function and to this functions, small intervals of the Stage 2 were passed to align with ET. This then gave better alignment of both the data sets.



Model Training

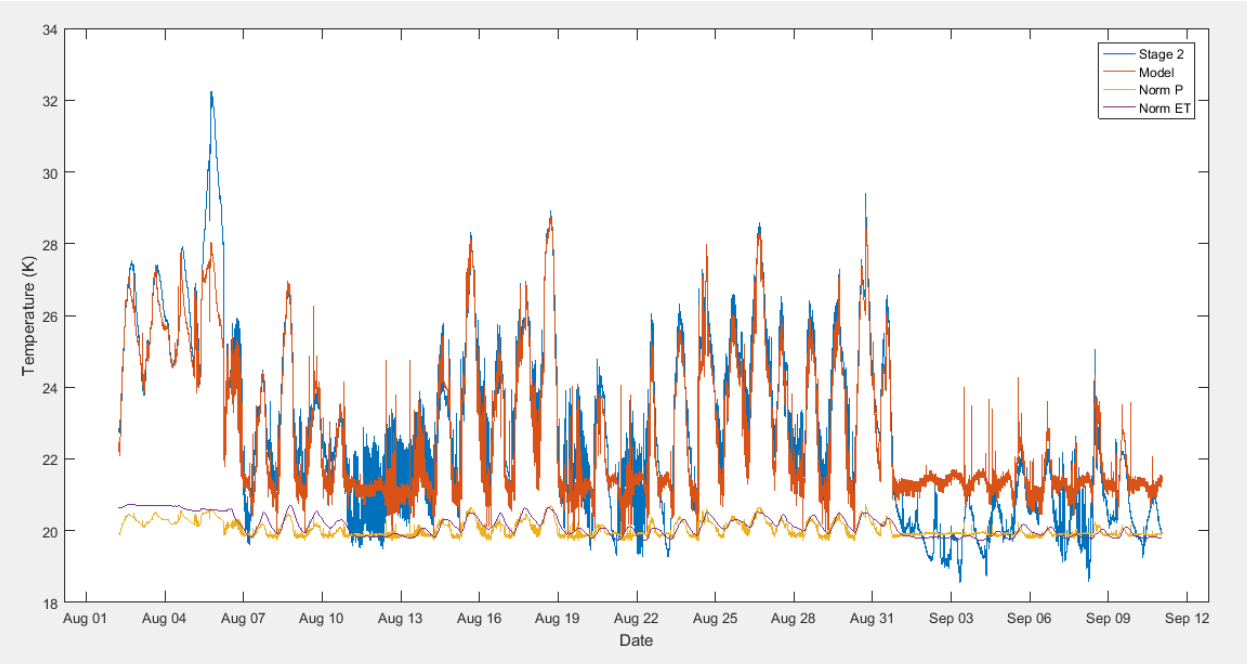
- The basic model for Stage 2 was defined as:

  Stage 2 = (theta1 * f1) + (theta2 * f2) + theta3

- where f1 is normalized Pressure and f2 is normalized ET.
- The Pressure exhibits a similar peak pattern as in Stage 2 but the difference between the peak maximas and non-peak values is much higher in Pressure measurements. Therefore, Pressure was scaled using a log function and was then normalized.
- A gradient descent algorithm was applied to determine the values of the weight coefficients (thetas) which using the Least Squared Adjustment.
- The values obtained for one train data set were:

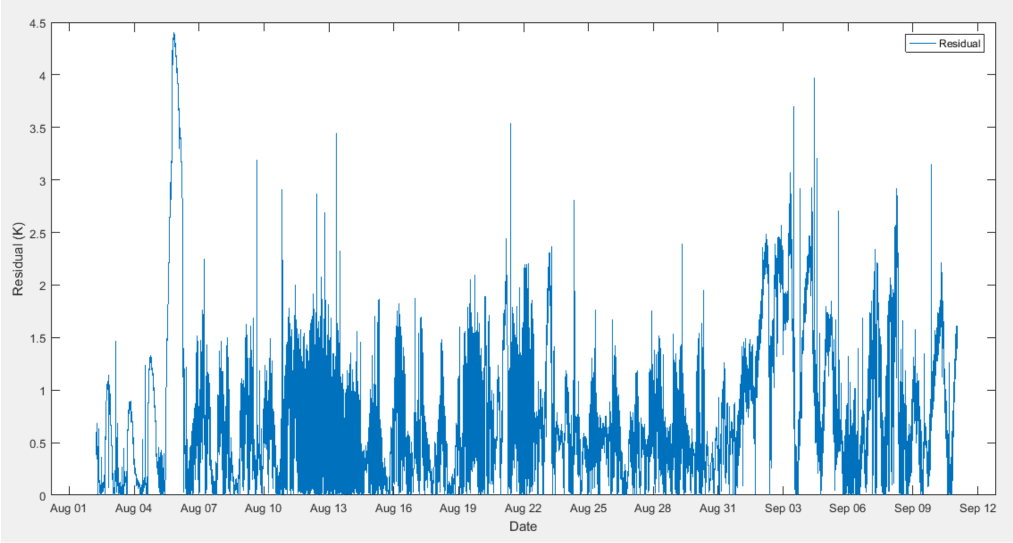
theta1 = 5.51770608945191
theta2 = 1.17235548090479
theta3 = 19.7271796846091

- The following plot show the normalized Pressure and ET shifted by theta3 along with the result of trained model and the original Stage 2 measurements.

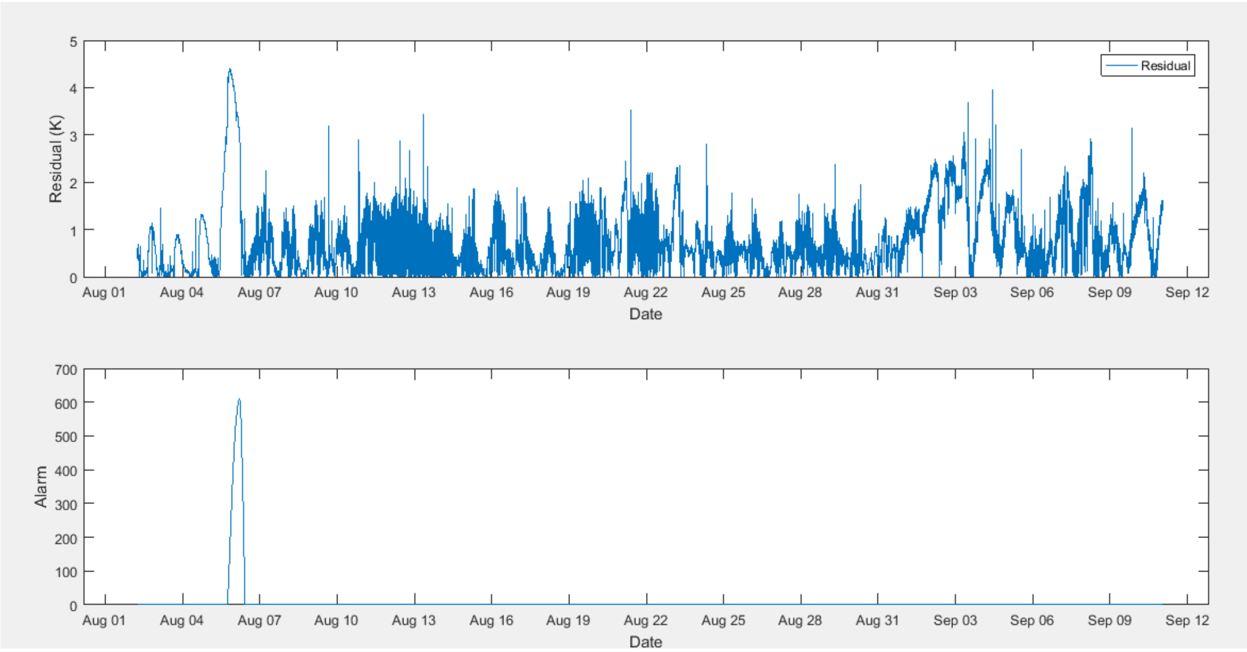


Residuals and CUSUM Test

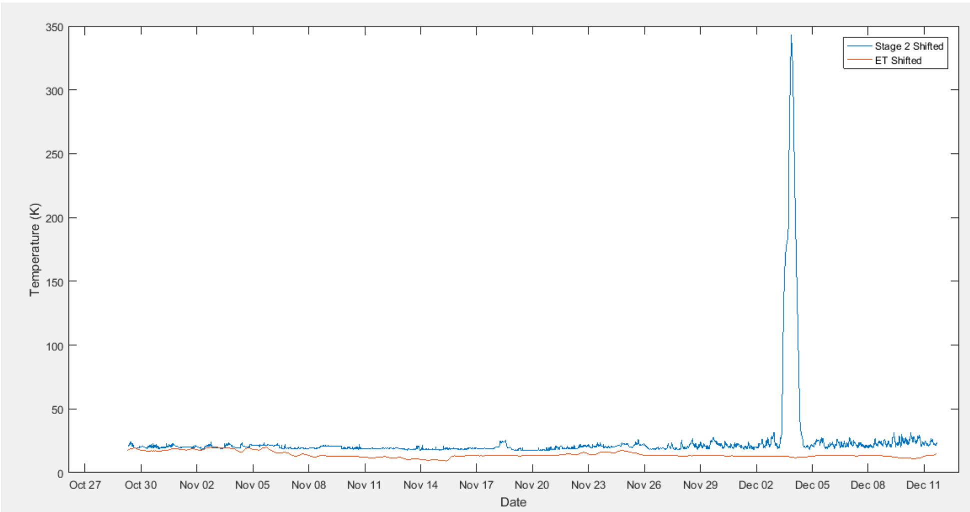
The following plot show the calculated residuals between model and actual Stage 2 measurements. The residuals are actually the absolute difference between model and Stage 2 at each time instant.



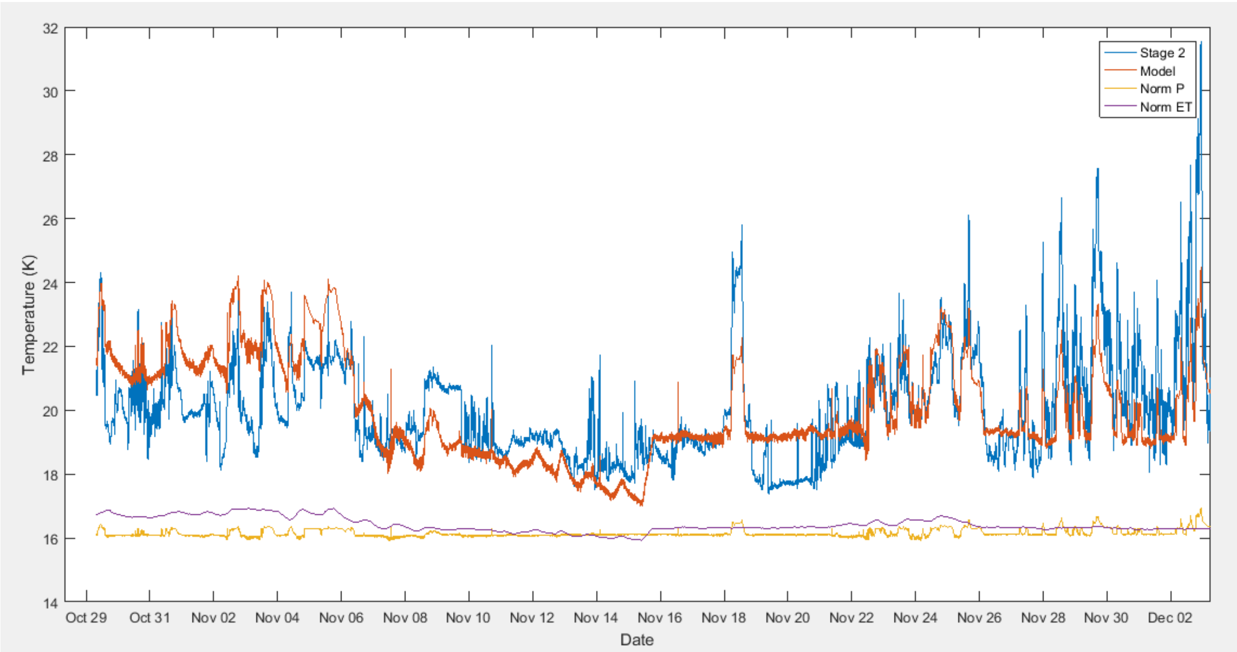
- It can be observed that the maximum residual in this case reach around 4.5 K. In fact, we can set a threshold over the maximum residual beyond which it can be inferred that the system is undergoing a change.
- A test that can be applied over the evaluated residuals is the CUSUM test that sums the residuals over time and checks if this sum does not exceed a specified bound. The test includes two tuning parameters: Threshold and Forgetting Factor. The Threshold specifies the maximum allowable bound for the summed residuals and the Forgetting Factor specifies how much the sum should depend on past values.
- For the following case a value of three was chosen both for Threshold and Forgetting Factor. The plot below indicates one alarm when the summed residual exceeds beyond 3.



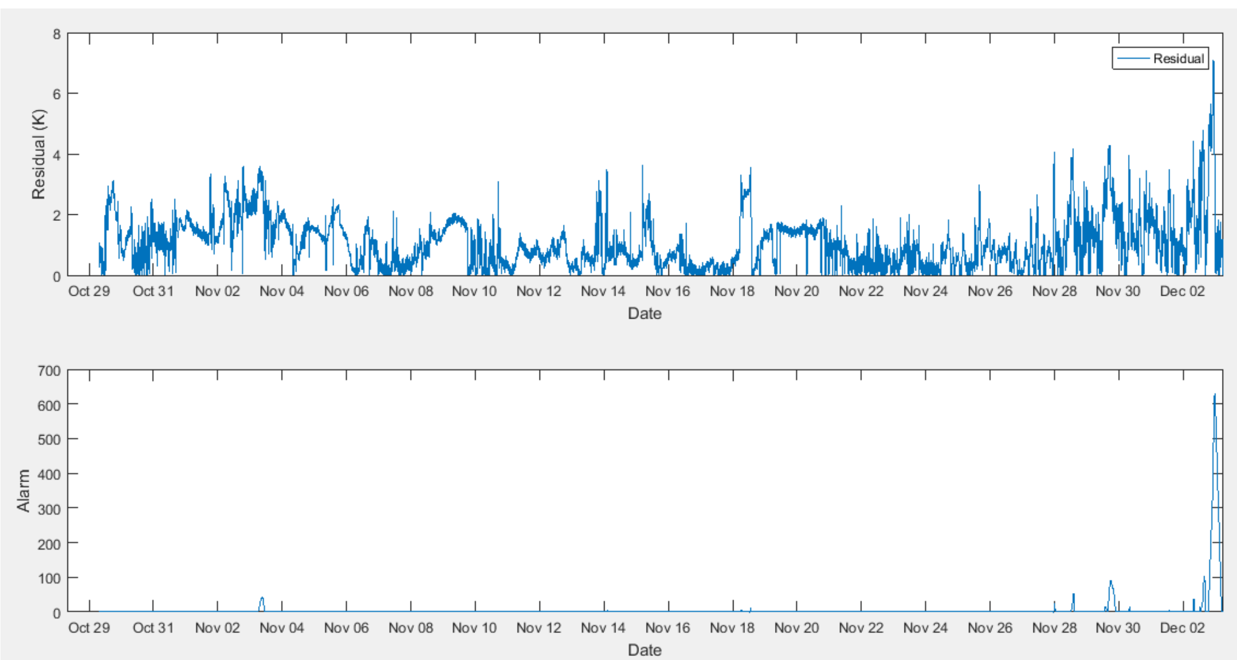
Considering the following data set that contains a maintenance event:



- To avoid the anomaly occurring at the instance of maintenance event, I selected the aligned data sets till just before the maintenance event arrives. Doing so, yields the following model vs measurement plot:



- And provides with the following residual and alarm plots:



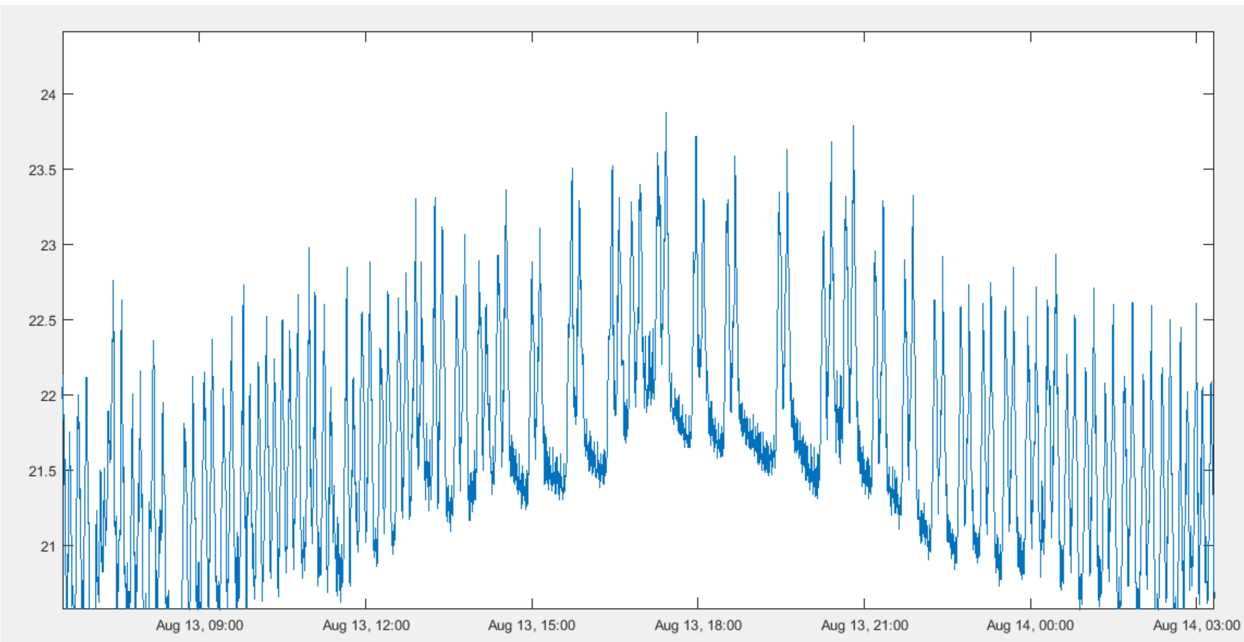
- It can be observed that we obtain alarms just before the arrival of maintenance event using such a simple data model only. It is evident that we would obtain better results if the model is more accurate.

The aim is therefore to make the data model as accurate to the expected value of Stage 2 as possible and to reduce the number of false alarms.

Discussion on an Ideal Data Model

It is worthwhile to analyze the short-term variations in Stage 2. While the daily variations are intuitively explained by the change in the surrounding environment temperature, the short term variations can be inferred to be more associated with the physical degradation of the system. To have a very reliable data model for Stage 2, one should investigate more about the physical reason behind the short-term variations. As discussed in the last meeting, the natural reason for Dewar system degradation is leak of vacuum over time which allows ice to form over the head which then sublimes and then formulates a gaseous atmosphere inside the chamber.

The Stage 2 temperature under such conditions is seen to exhibit short period temperature peaks whose period can vary from 15 mins to 45 mins and are shown in the figures below:



These peaks definitely seem to follow a pattern which change over time. These short term peaks are also visible in the pressure data. In fact, one can attribute the peaks in Stage 2 to the sudden rise and falls in pressure over short durations by considering the temperature / pressure relationship in an isothermal chamber. Therefore, one has to investigate about the physical model for change in pressure. The intuitive physical principle that shall explain this temperature behavior is convection occurring in the gaseous atmosphere that forms around the head. If one can solve for such convection cycles over the physical structure of Dewar, then one might obtain a correlation between the simulation results and the pressure measurements. This should in principle give a much more reliable data model for the overall measurement of Stage 2.

Way Forward

I am now working to analyze the approach further and to formulate more accurate data model for Stage 2. In parallel, I also look into convection models whose results provide a resemblance to the periodic peaks we observe in the pressure data series.

Also, I am considering other statistical tests than CUSUM that can be applied over the residuals for more reliable results.

It would be helpful if I get the latest environment and the corresponding sensor data set so I can run this approach on more maintenance events.