# ISEN 614 PROJECT


# ADVANCED QUALITY CONTROL


## Submitted by

Abhijith Narayanan (827008921)

Malik Nabeel Aamer (827000011)

Rohit Birhade (427009769)

**Executive Summary**

An important aspect in a manufacturing process is to maintain quality of the product being produced. Any thriving business backbone is built upon Quality Control principles that enables it to meet the expectations of a customer. In high production environments even, slight deviations in control parameters can lead to poor quality and expensive losses for a company. It is therefore highly imperative that such possible changes are detected in a timely and efficient manner.

The data set provided from a manufacturing process has sample size, number of observations and quantity of attributes equal to 1, 552 and 209 respectively. We carried out Phase 1 analysis from the in-control data to estimate the population mean and covariance matrix. The process involved mapping control charts from the given data and removing any out of control points that are found outside the control limits. Remaining data points are used to form an updated control chart in the next iteration. The procedure is repeated over as many iterations needed until all the remaining points are in-control. This would mark the end of Phase 1.

Certain charts like Hotelling T^2 control chart is good for detecting large spikes. Whereas m-EWMA and m-CUSUM are better equipped at identifying small sustained mean shift in data. While utilizing two charts at once like T^2 coupled with m-EWMA chart can help with overcoming each other's shortcomings. In this Project we performed Phase 1 analysis using five different methodologies namely T^2, m-CUSUM, m-EWMA, T^2 m-CUSUM combination and T^2 m-EWMA. We saw little encouragement in using combination of m-CUSUM and m-WMA control chart since both these methods have similar strengths and weaknesses when it comes to detecting out of control data points.

Since the number of attributes were large in number, 209 to be precise we decided to use Principal Component Analysis that enabled us to perform dimension reduction. Selection of the number of Principal Components was made with the help of Scree and Pareto Plot that are also included in this report.

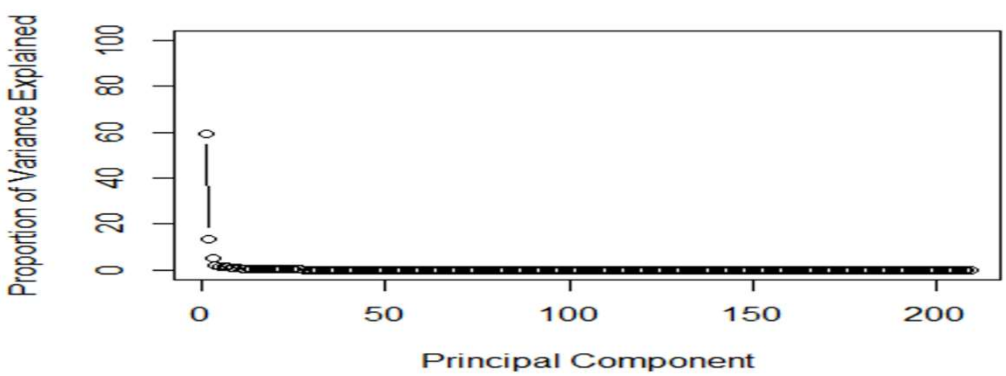
Overall the project gave us practical exposure to limitations and strengths of different detection methodologies when it comes to dealing with real world data. There wasn't any single straightforward solution and hence multiple options were explored. R Language and pre built packages were utilized to perform Phase 1 Analysis. In the end we came away with better understanding of the material covered in the class.

**Principal Component Analysis (PCA)**

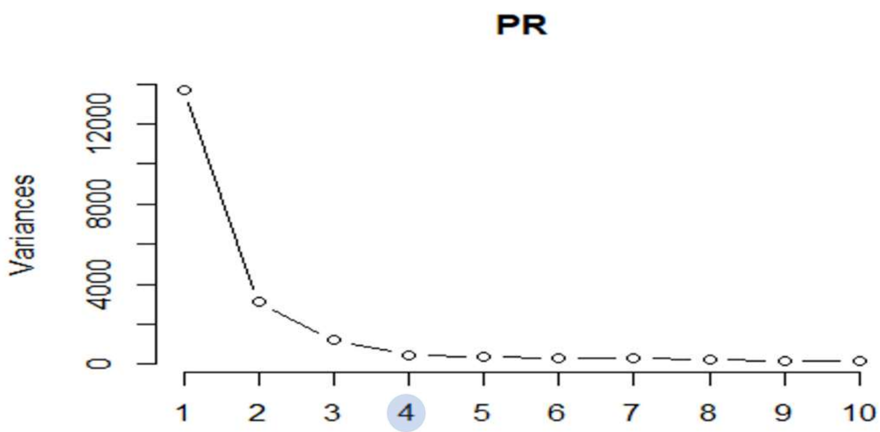
In higher dimension aggregated noise effect can adversely affect the signal to noise ratio making the detection process more complicated. Principal Component Analysis is applied on data of high dimension providing us resulting components where "vital few" can be selected to explain the variation of the data. This aids in the detection process as it is now performed on a significantly reduced dimension.

Scree and Pareto Plots were used instead of Minimum Description Length (MDL) because of the known fact that MDL retains too many eigen values. The following control charts were obtained as a result of the R simulation.

**Scree Plot**



**Pareto Plot**




With the help of both these plots we were able to pick four Principal Components that accounted for maximum variance (roughly 80%).

**<u>T^2 Hotelling Control Chart</u>**

It is one of the most popular detection methods for multivariate cases as its easy to set up and very effective at detecting anomalies such as large spikes. By declaring type "t2" in the mult.chart function from "MSQC" library in R, Phase 1 Analysis was performed for p =4 and alpha = 0.005.

**Procedure:**

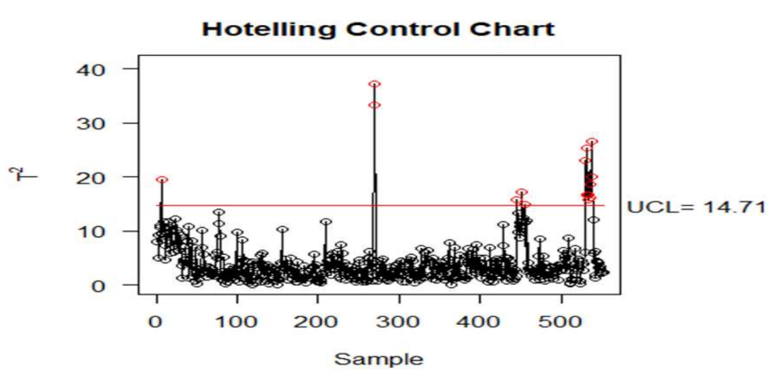
1) T^2 chart for the provided dataset having four principal components was plotted.

2) Out of Control points that exceeded the calculated UCL were subsequently removed.

3) Steps 1 and 2 will be repeated as long as we obtain a control chart with no points that can be labelled as being out of control.

The results of the T^2 Hotelling Control Chart method are shown below: -

| **Iteration 1** |
|---|

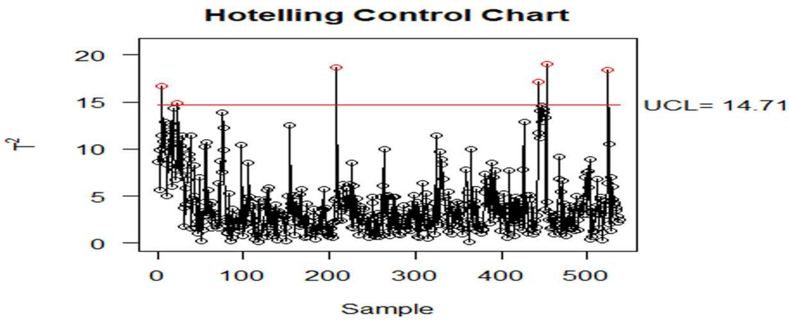



UCL = 14.71

LCL = 0

Out of Control Points = 15

**Iteration 2**

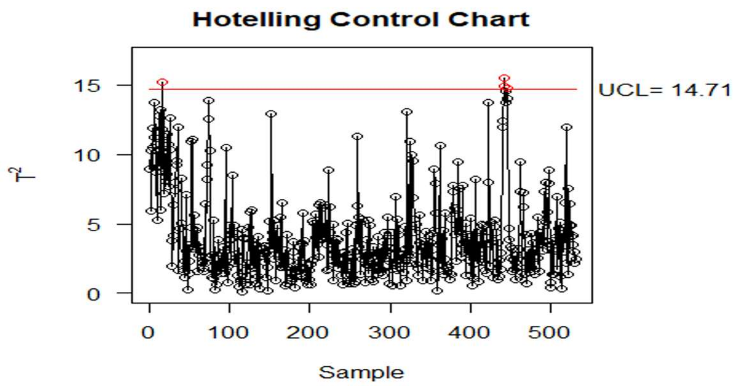



UCL = 14.71

LCL = 0

Out of Control Points = 6

**Iteration 3**

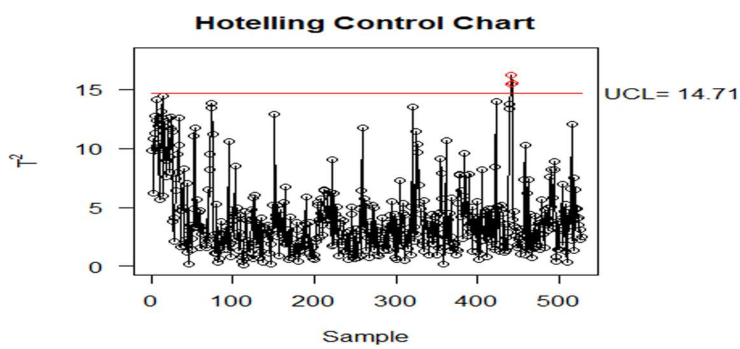



UCL = 14.71

LCL = 0

Out of Control Points = 4

**Iteration 4**

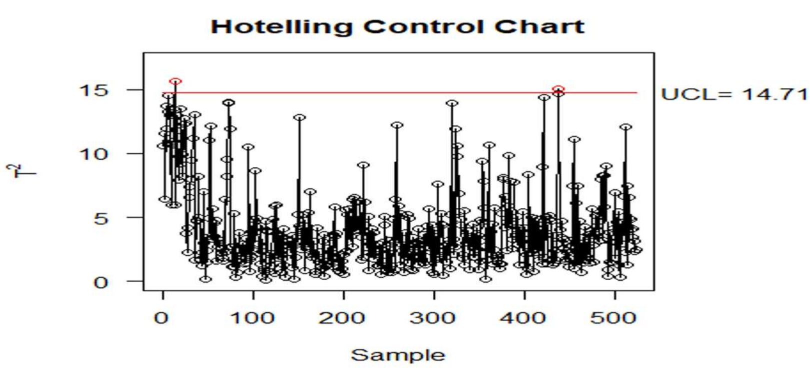



UCL = 14.71

LCL = 0

Out of Control Points = 3

**Iteration 5**

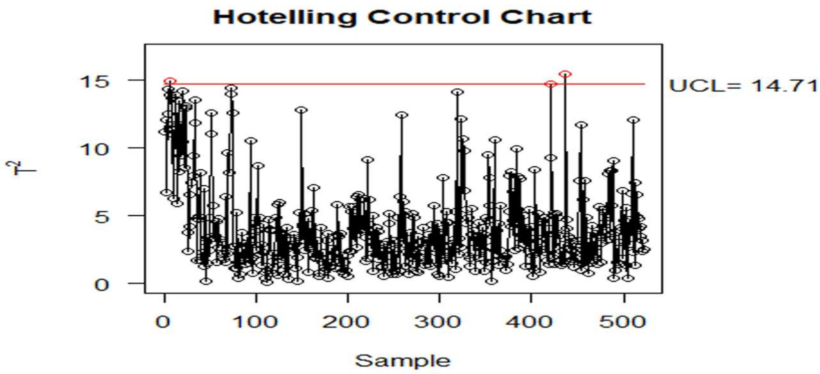



UCL = 14.71

LCL = 0

Out of Control Points = 2

**Iteration 6**

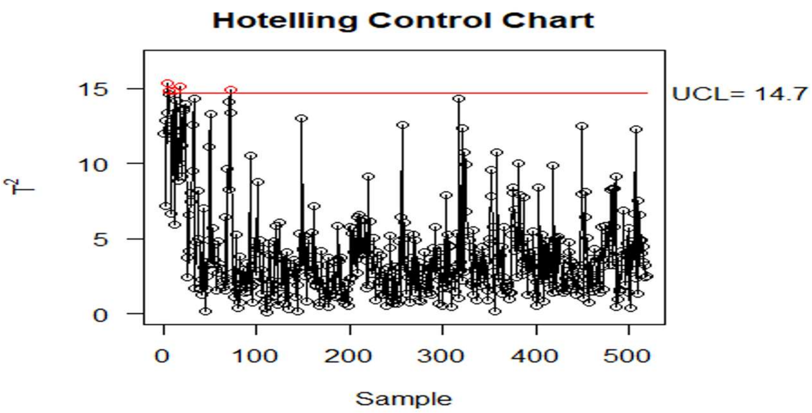



UCL = 14.71

LCL = 0

Out of Control Points = 3

**Iteration 7**

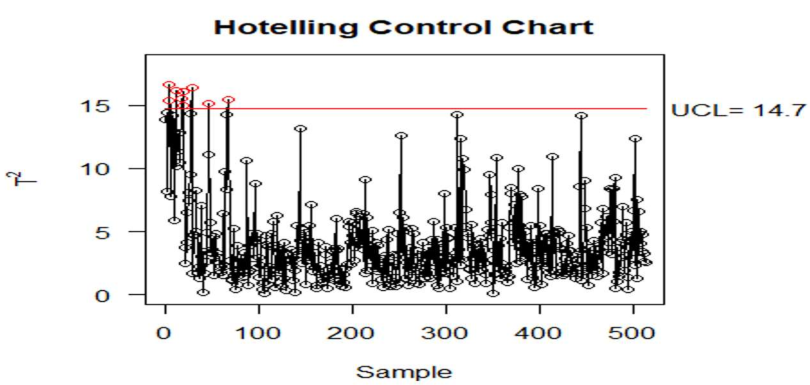



UCL = 14.70

LCL = 0

Out of Control Points = 5

**Iteration 8**

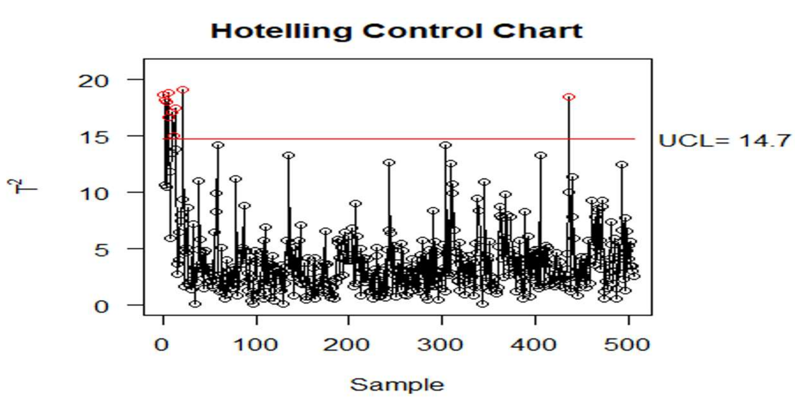



UCL = 14.70

LCL = 0

Out of Control Points = 9

**Iteration 9**

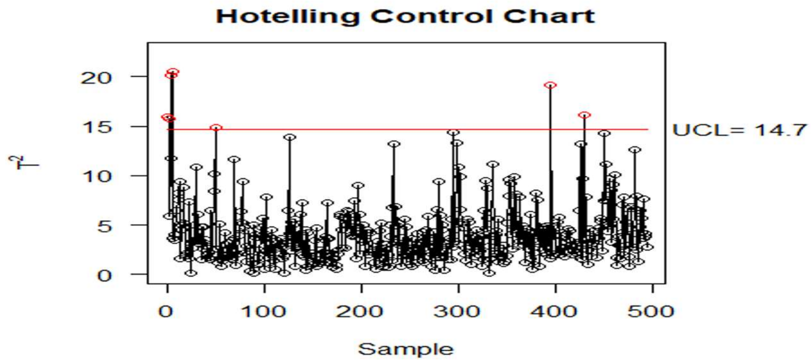



UCL = 14.70

LCL = 0

Out of Control Points = 10

**Iteration 10**

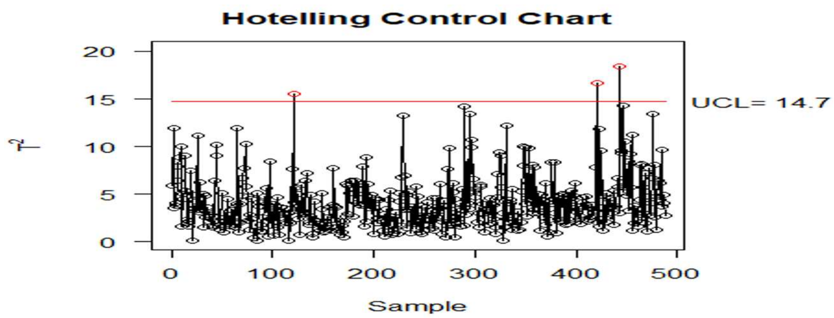



UCL = 14.70

LCL = 0

Out of Control Points = 7

**Iteration 11**

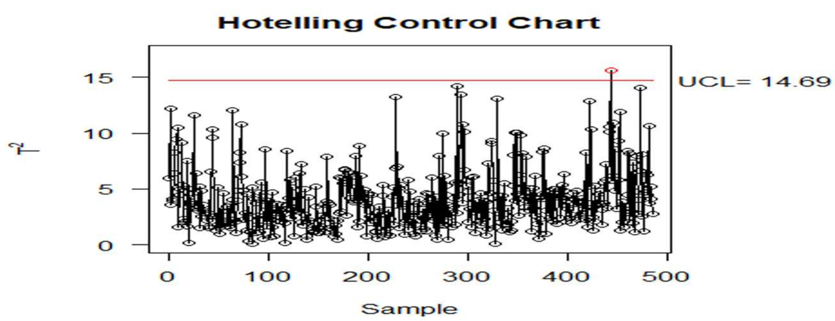



UCL = 14.70

LCL = 0

Out of Control Points = 3

**Iteration 12**

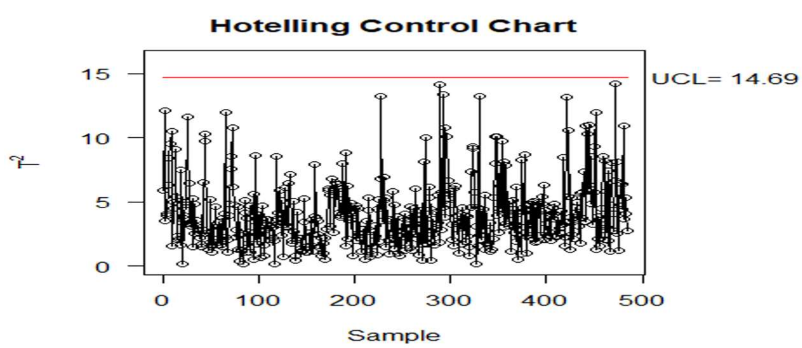



UCL = 14.69

LCL = 0

Out of Control Points = 1

**Iteration 13**




UCL = 14.69

LCL = 0

Out of Control Points = 0

After 13 iterations we find all remaining points to be in-control. Hence, we stop here. In total 68 Out of Control data points were removed and 484 data points remain.

**m-EWMA Chart**

Multivariate Exponential Weighted Moving Average(m-EWMA) Chart is commonly deployed to identify small consistent mean shifts in data being processed. By declaring type "mewma" in the mult.chart function from “MSQC” library in R, we were able to perform Phase 1 analysis for the following set critical values :-

alpha = 0.005, L0 = 200, lambda = 0.5, p = 4 and l = 0.5.

The Upper Control Limit (UCL) was calculated in R through mewma.crit function present in “spc” package.

**Procedure:**

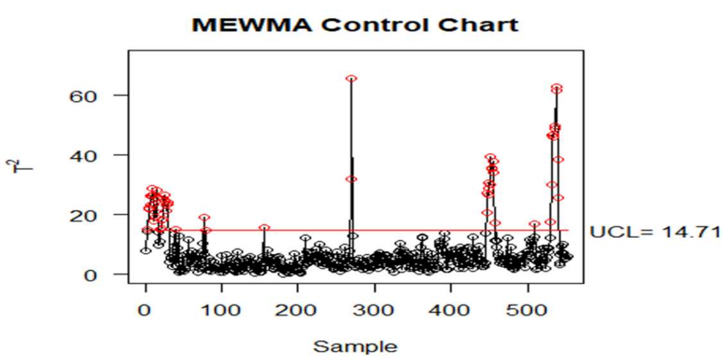
1) M-EWMA chart for the provided dataset having four principal components was plotted with lambda set as 0.5

2) Out of Control points that exceeded the calculated UCL were subsequently removed.

3) Steps 1 and 2 will be repeated as long as we obtain a control chart with no points that can be labelled as being out of control.

The results of the m-EWMA method are shown below: -

**Iteration 1**

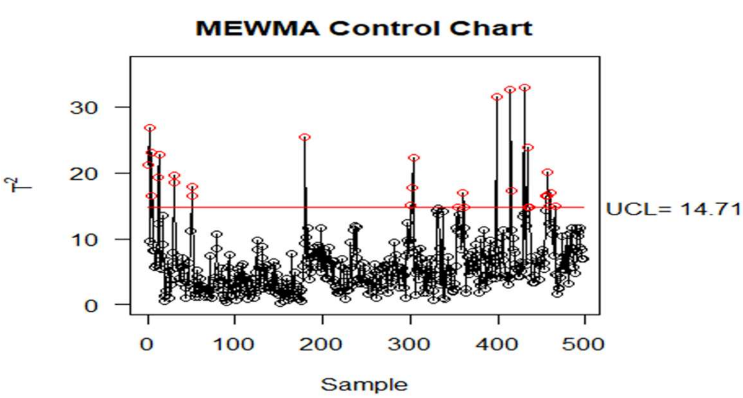



UCL = 14.71

LCL = 0

Out of Control Points = 55

**Iteration 2**

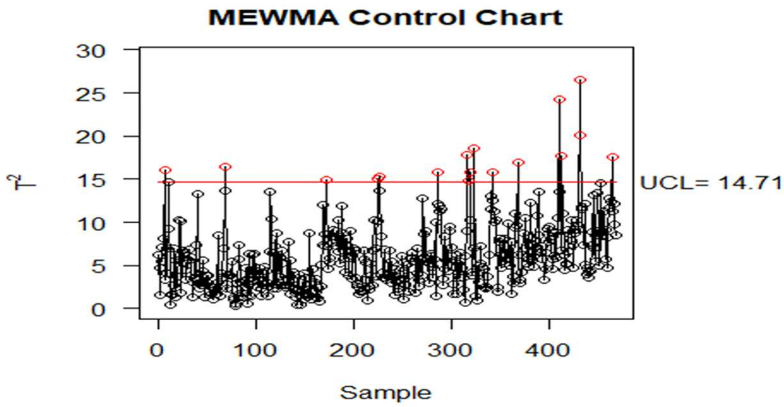



UCL = 14.71

LCL = 0

Out of Control Points = 30

**Iteration 3**

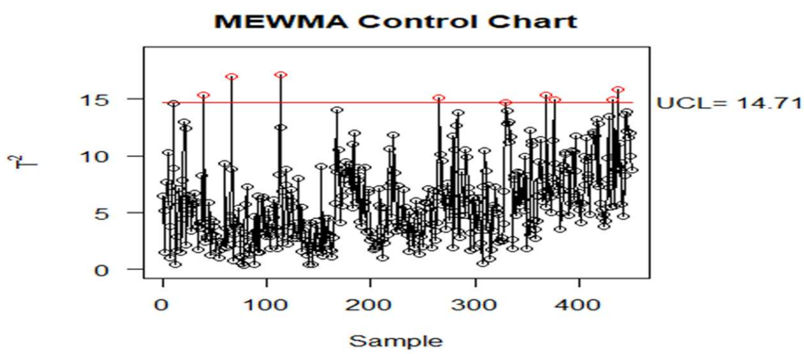



UCL = 14.71

LCL = 0

Out of Control Points = 17

**Iteration 4**

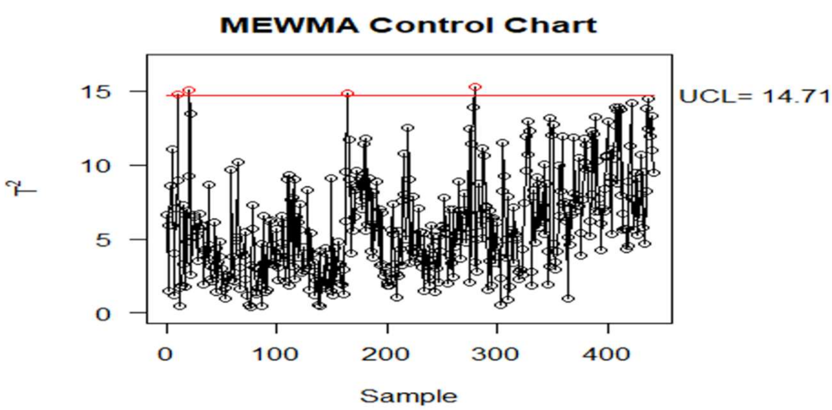



UCL = 14.71

LCL = 0

Out of Control Points = 9

**Iteration 5**

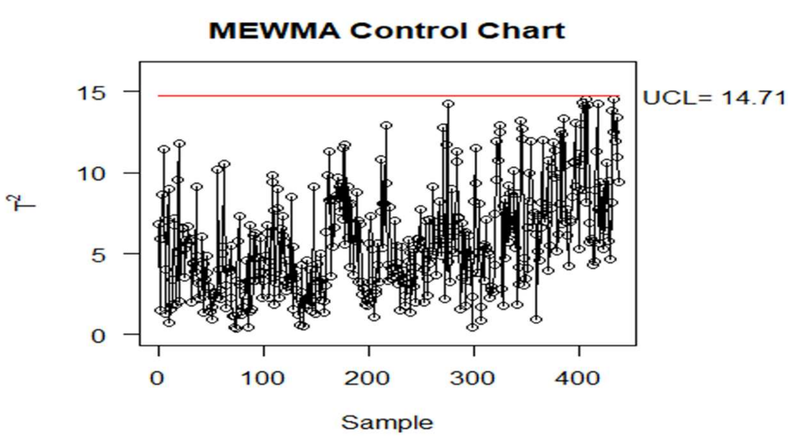



UCL = 14.71

LCL = 0

Out of Control Points = 4

**Iteration 6**




UCL = 14.71

LCL = 0

Out of Control Points = 0

We stop the process at iteration 6. In total 115 out of control data points were removed and now 437 data points remain at the end of Phase 1.

**m-CUSUM Chart**

Like m-EWMA, Multivariate Exponential Weighted Moving Average(m-CUSUM) Chart is also frequently used in practical applications due to its good ability to detect mean shifts of small size in a process of interest. By declaring type "mcusum"in the mult.chart function from “MSQC” library in R, we were able to perform Phase 1 analysis for the following set critical values :-

k = 1.5 and h = 6. The Upper Control Limit (UCL) is basically value of h which is 6 in this case.

**Procedure:**

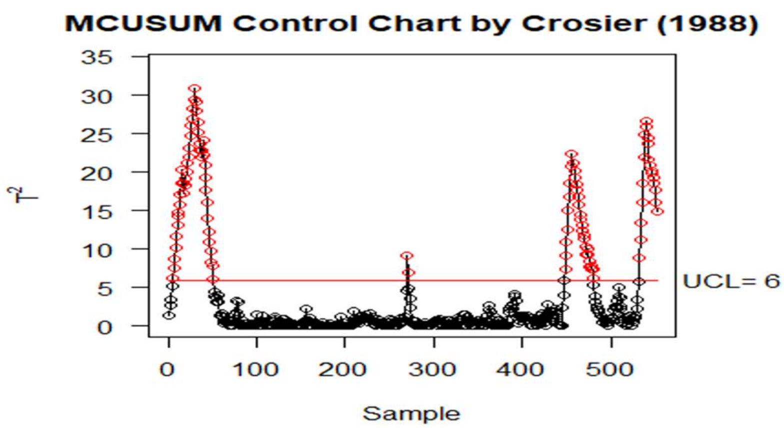
1) m-CUSUM chart for the provided dataset having four principal components was plotted with k=1.5, h =6.

2) Out of Control points that exceeded the UCL were subsequently removed.

3) Steps 1 and 2 will be repeated as long as we obtain a control chart with no points that can be labelled as being out of control.

The results of the m-CUSUM method are shown below: -

**Iteration 1**

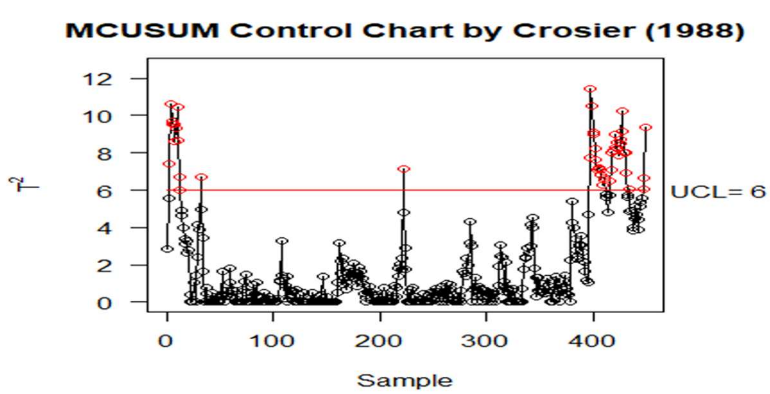



UCL = 6.0

LCL = 0

Out of Control Points = 103

**Iteration 2**

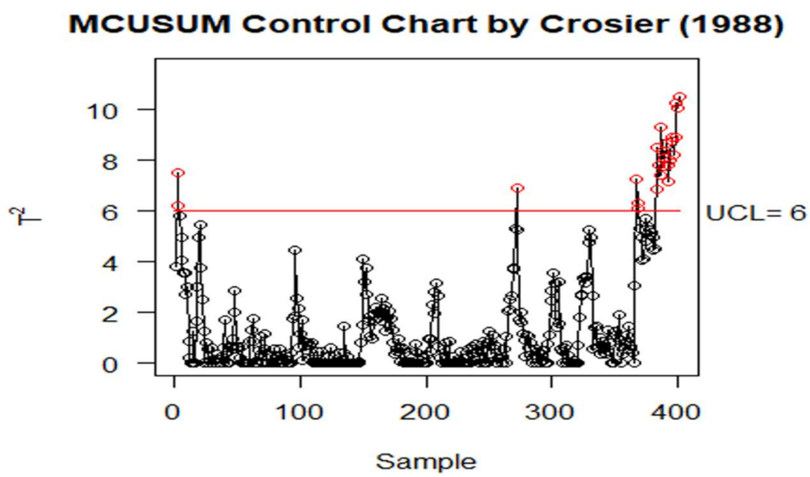



UCL = 6.0

LCL = 0

Out of Control Points = 48

**Iteration 3**

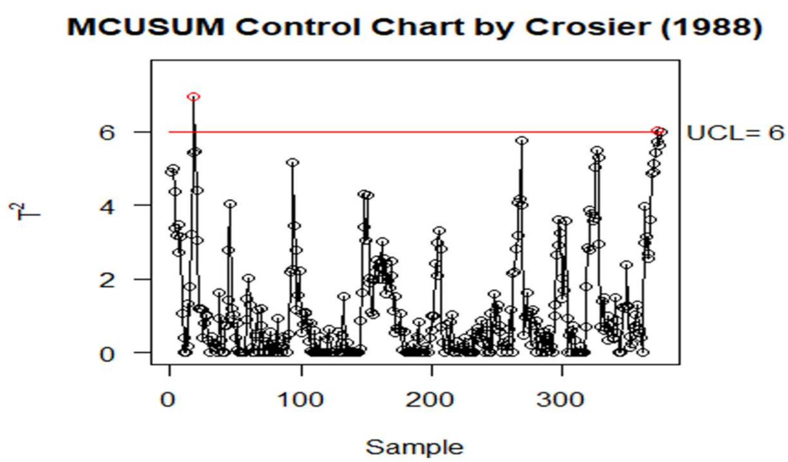



UCL = 6.0

LCL = 0

Out of Control Points = 25

**Iteration 4**

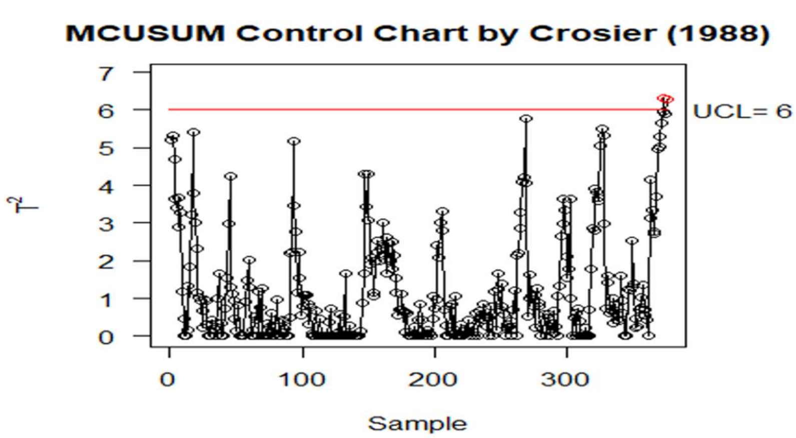



UCL = 6.0

LCL = 0

Out of Control Points = 1

**Iteration 5**

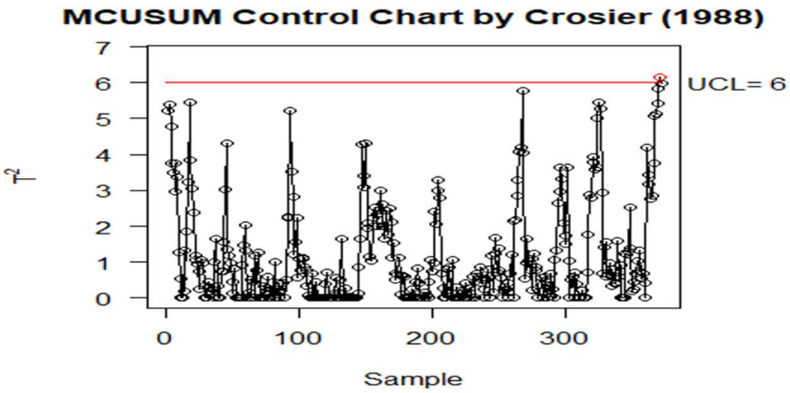



UCL = 6.0

LCL = 0

Out of Control Points = 2

**Iteration 6**

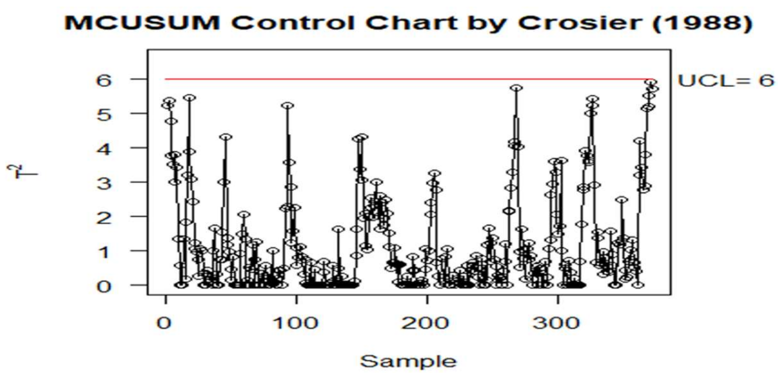



UCL = 6.0

LCL = 0

Out of Control Points = 1

**Iteration 7**




UCL = 6.0

LCL = 0

Out of Control Points = 0

We stop at iteration 7. After removal of 180 points, in total 372 points remain.

**<u>T^2 and m-EWMA Combination</u>**

Combining the T^2 and m-EWMA control charts for detection purposes in order to overcome their individual limitations. T^2 is good for detecting large spikes whereas m-EWMA is popular for detecting small consistent mean shifts in data.

**Procedure:**

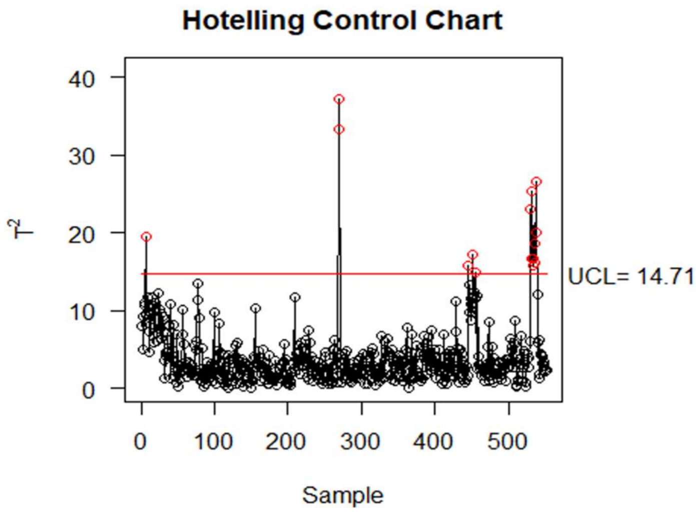
1) T^2 and m-EWMA charts for the provided dataset having four principal components were plotted.

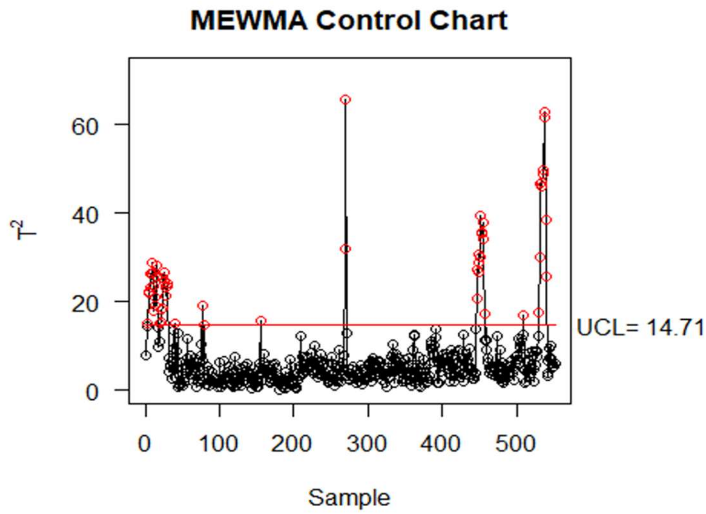
2) Out of Control points that exceeded the UCL in either of the control charts were subsequently removed.

3) Steps 1 and 2 will be repeated as long as we obtain a control chart with no points that can be labelled as being out of control.

The results are shown below: -

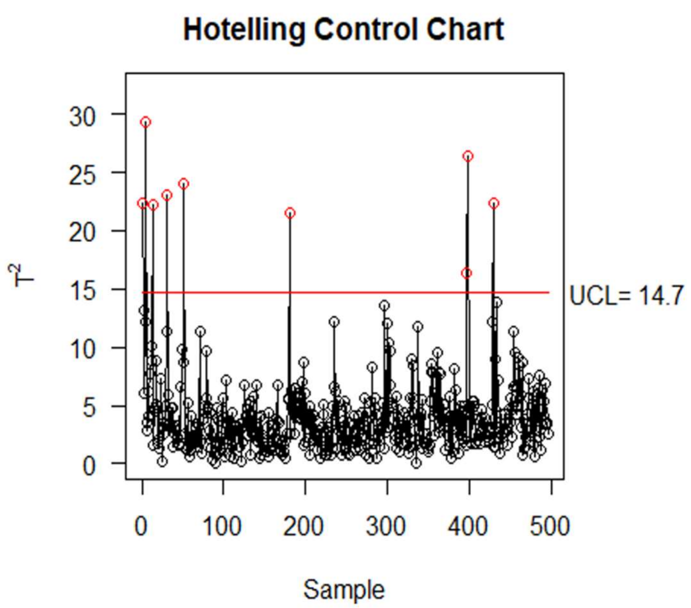
**Iteration 1**

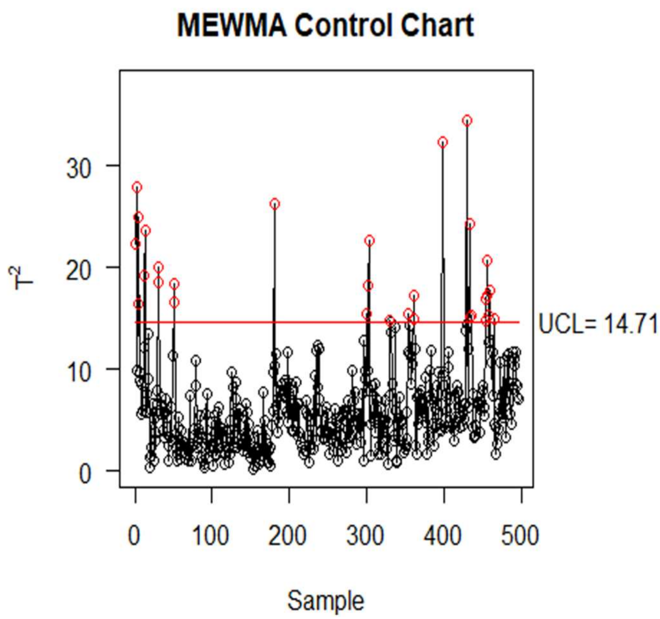



T^2 UCL = 14.71, m-EWMA UCL = 14.71, LCL = 0, Out of Control Points = 56

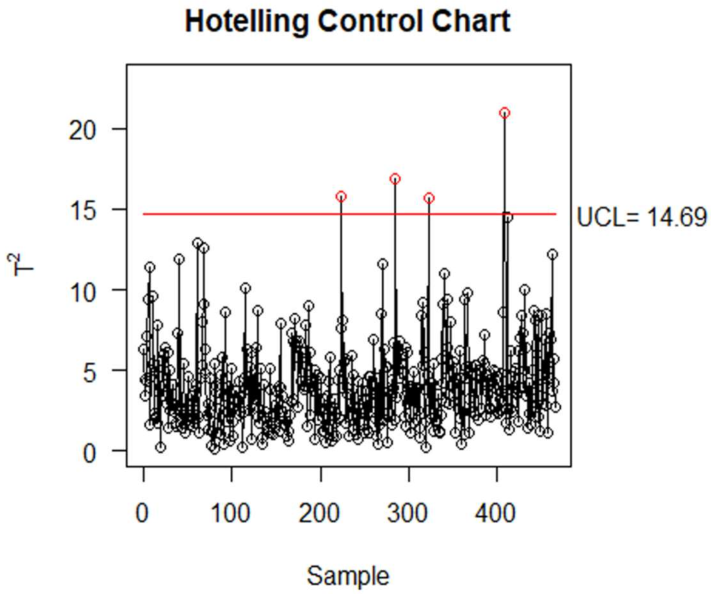
**Iteration 2**

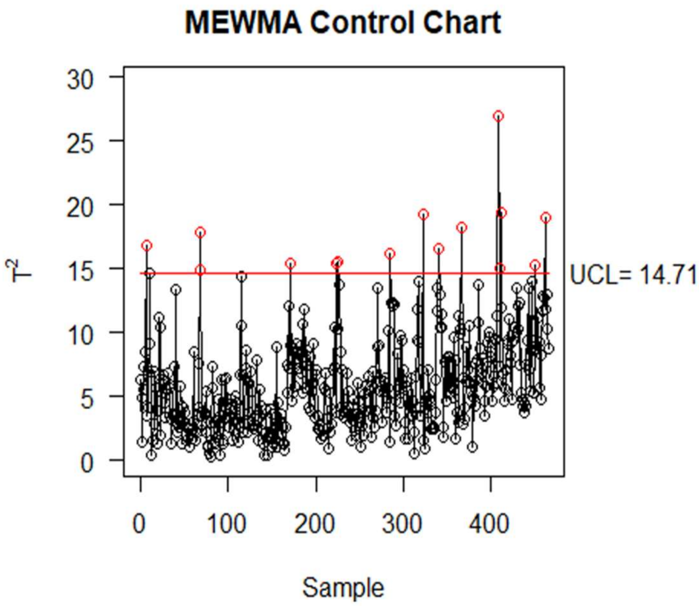



T^2 UCL = 14.70, m-EWMA UCL = 14.71, LCL = 0, Out of Control Points = 31

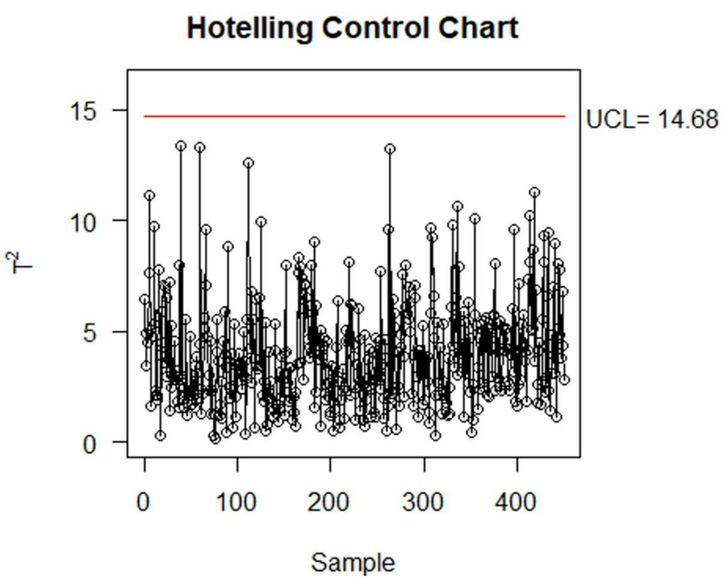
**Iteration 3**

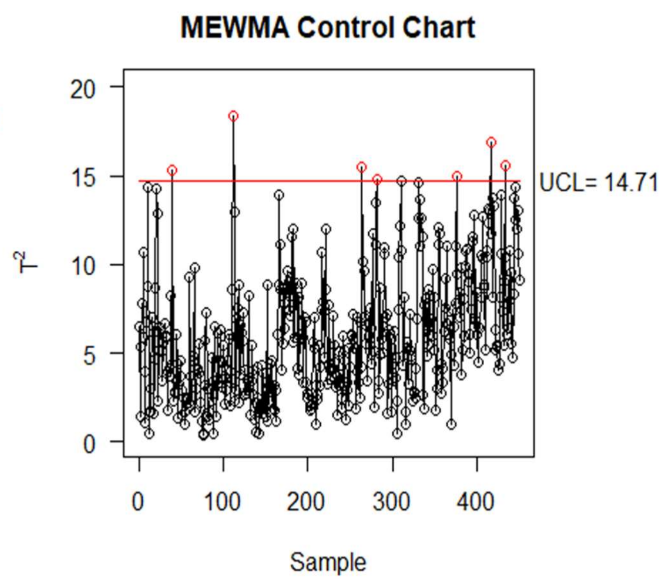



T^2 UCL = 14.69, m-EWMA UCL = 14.71, LCL = 0, Out of Control Points = 15

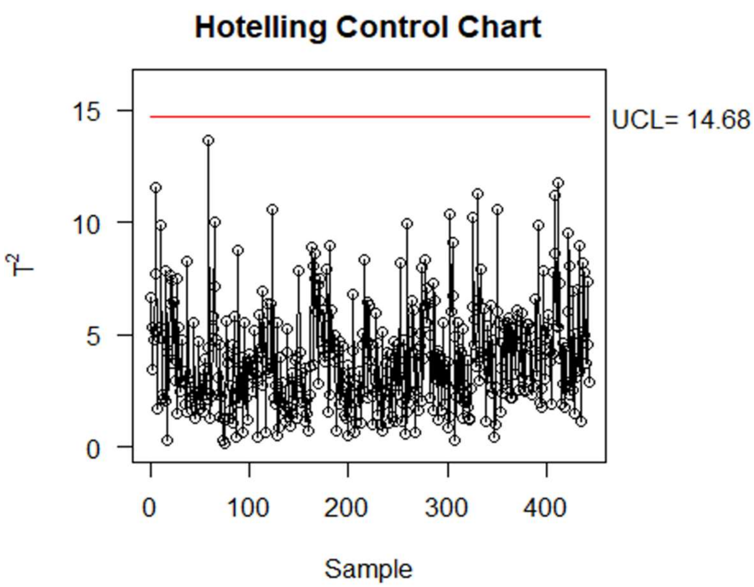
**Iteration 4**

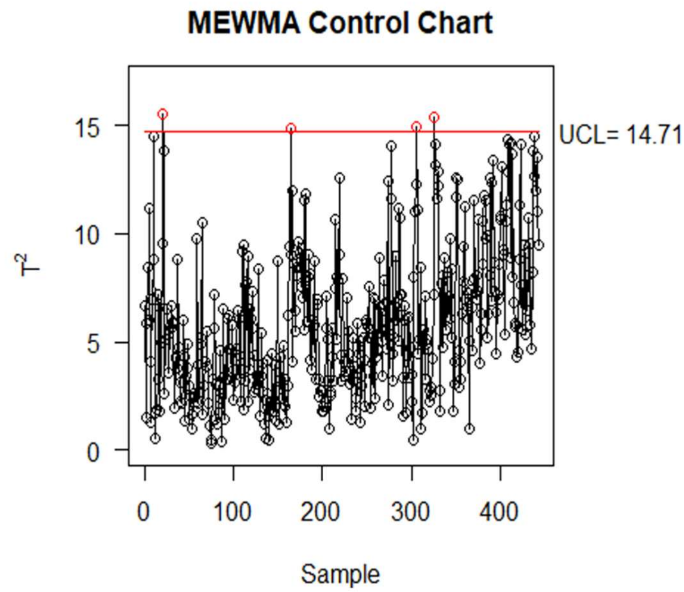



T^2 UCL = 14.68, m-EWMA UCL = 14.71, LCL = 0, Out of Control Points = 7

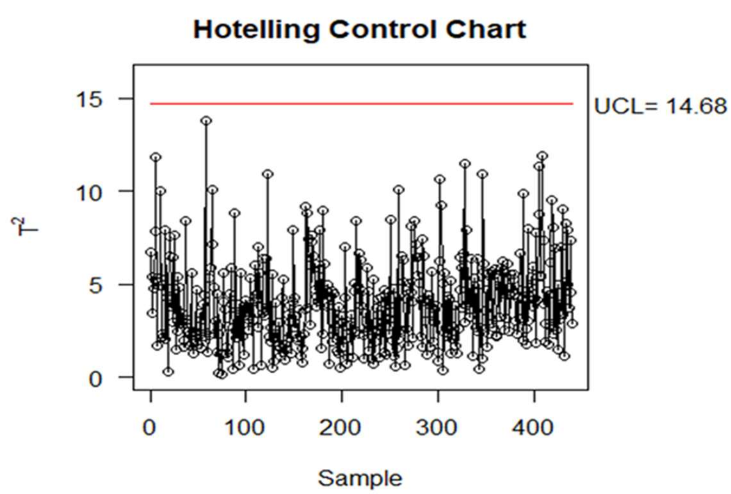
**Iteration 5**

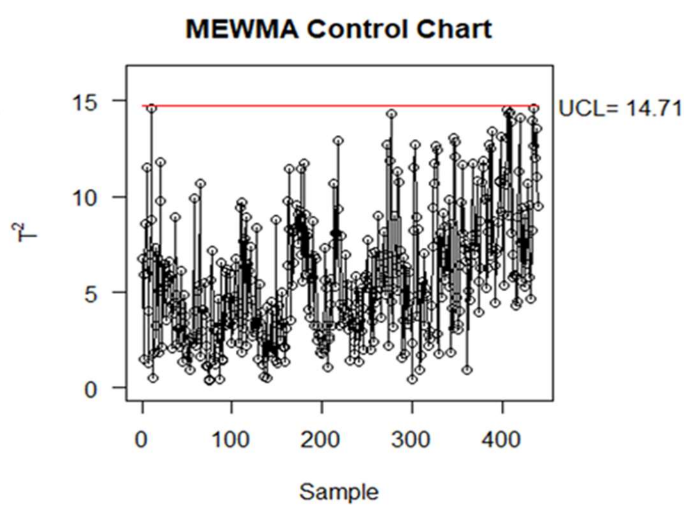



| T^2 UCL = 14.68, m-EWMA UCL = 14.71, LCL = 0, Out of Control Points = 4 |
|---|

| **Iteration 6** |
|---|




| T^2 UCL = 14.68, m-EWMA UCL = 14.71, LCL = 0, Out of Control Points = 0 |
|---|

We stop at iteration 6. In total 113 points were removed and 439 points remain.

**<u>T^2 and m-CUSUM Combination</u>**

Combining the T^2 and m-CUSUM control charts for detection purposes in order to overcome their individual limitations. T^2 is good for detecting large spikes whereas m-EWMA is popular for detecting small consistent mean shifts in data.

**Procedure:**

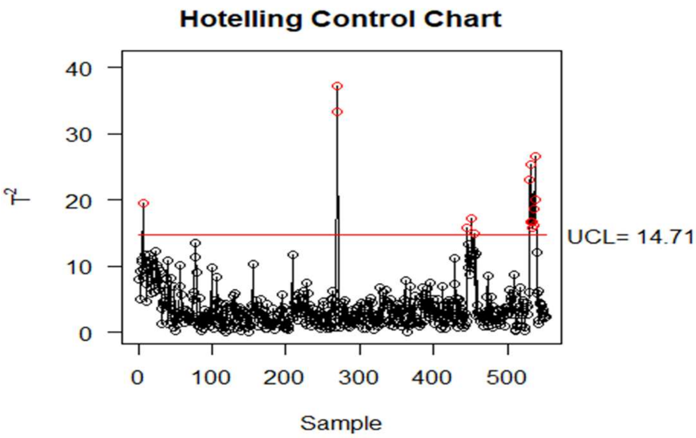
1) T^2 and m-CUSUM charts for the provided dataset having four principal components were plotted.

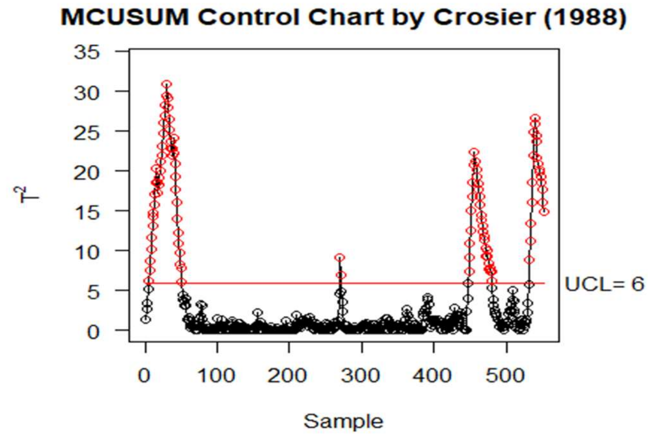
2) Out of Control points that exceeded the UCL in either of the control charts were subsequently removed.

3) Steps 1 and 2 will be repeated as long as we obtain a control chart with no points that can be labelled as being out of control.

The results are shown below: -

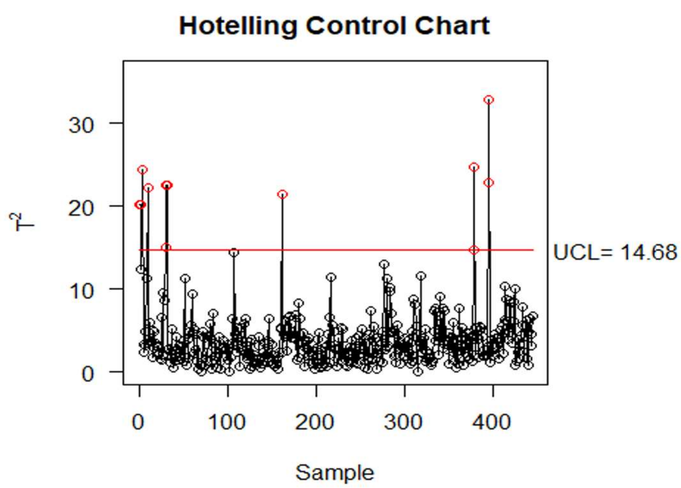
**Iteration 1**

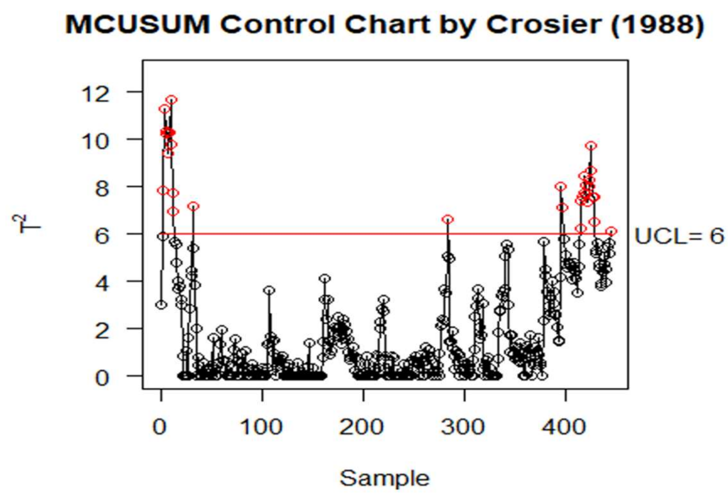



T^2 UCL = 14.71, m-CUSUM UCL = 6.0, LCL = 0, Out of Control Points = 107

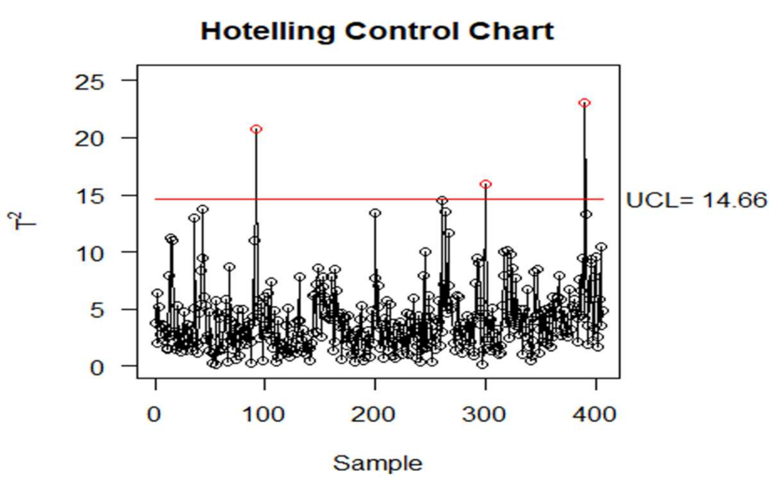
**Iteration 2**

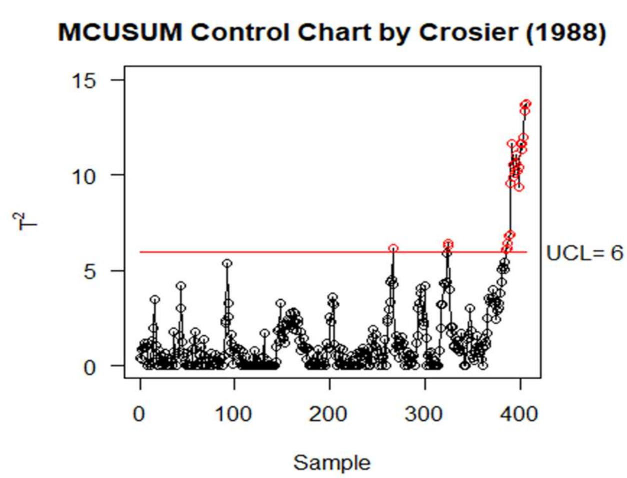



T^2 UCL = 14.68, m-CUSUM UCL = 6.0, LCL = 0, Out of Control Points = 39

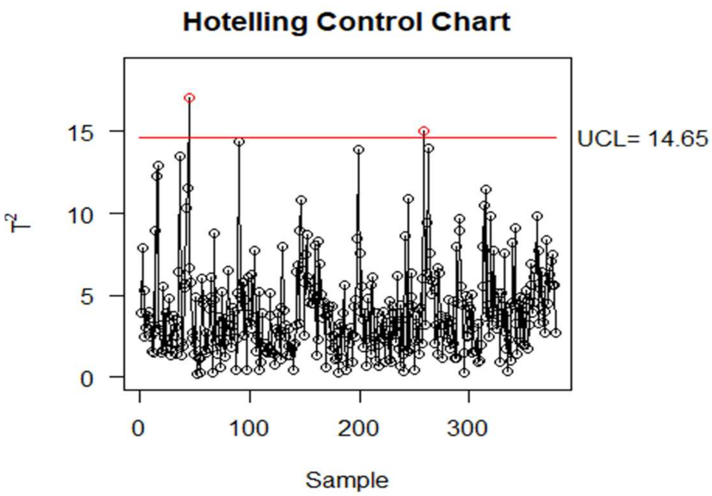
**Iteration 3**

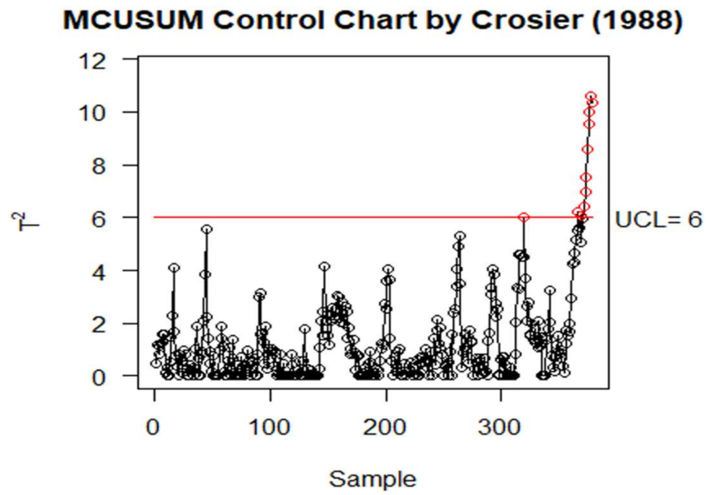


T^2 UCL = 14.66, m-CUSUM UCL = 6.0, LCL = 0, Out of Control Points = 27

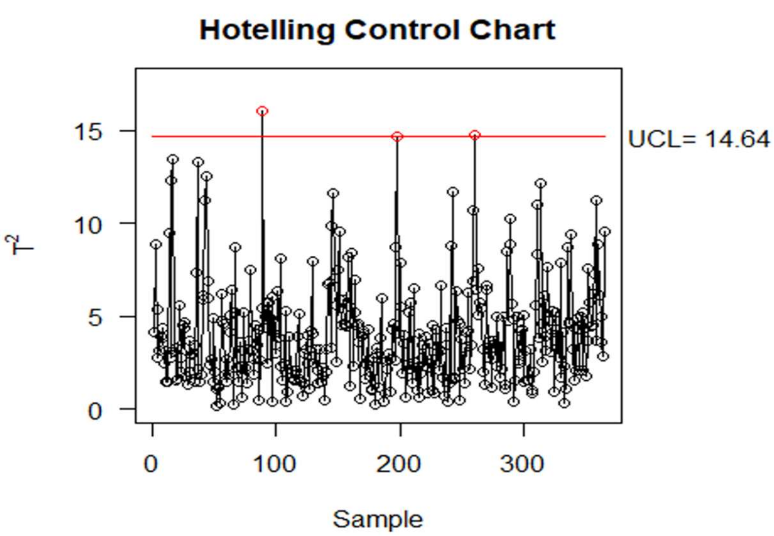
**Iteration 4**

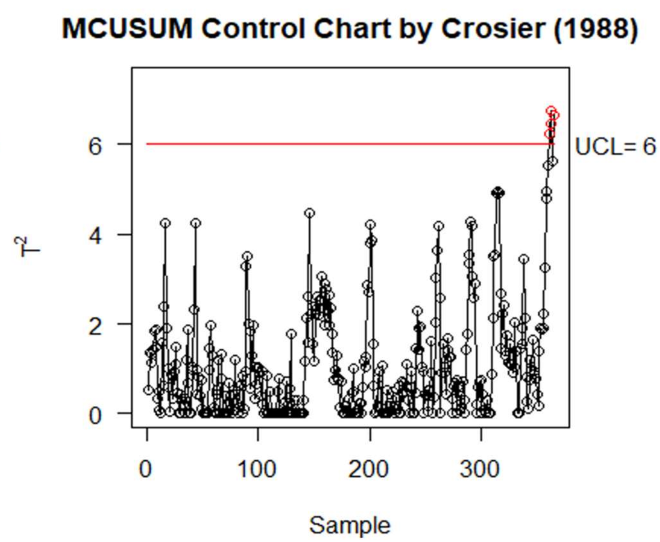



T^2 UCL = 14.65, m-CUSUM UCL = 6.0, LCL = 0, Out of Control Points = 13

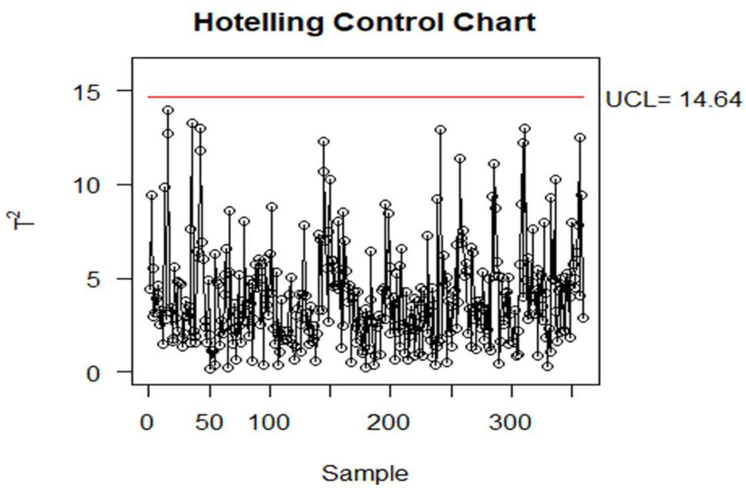
**Iteration 5**

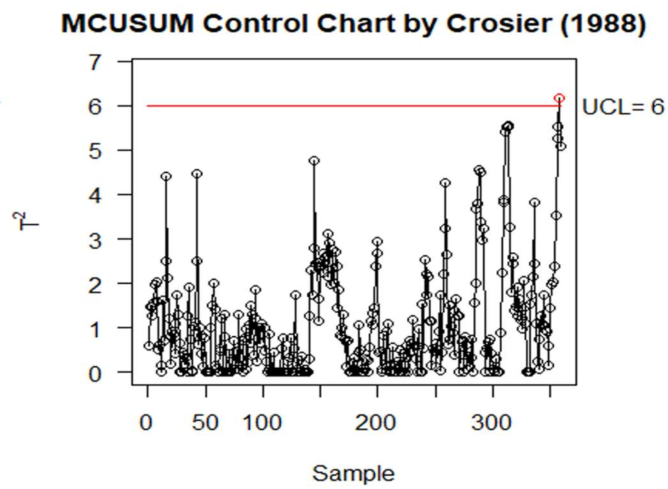



T^2 UCL = 14.64, m-CUSUM UCL = 6.0, LCL = 0, Out of Control Points = 7

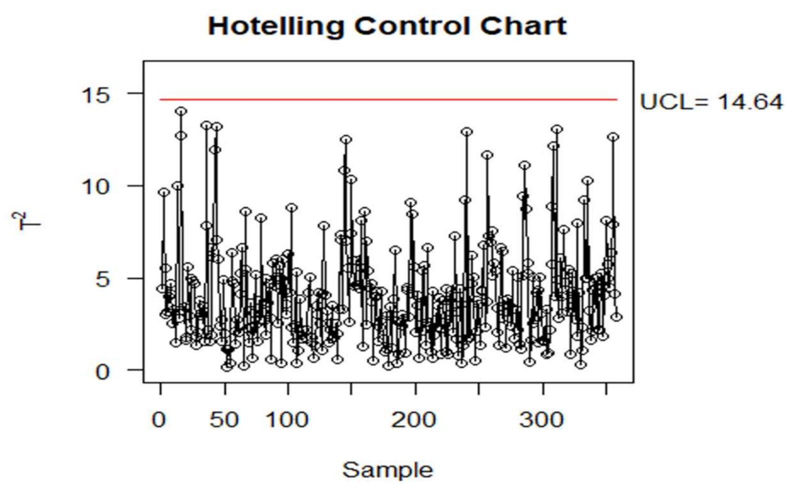
**Iteration 6**

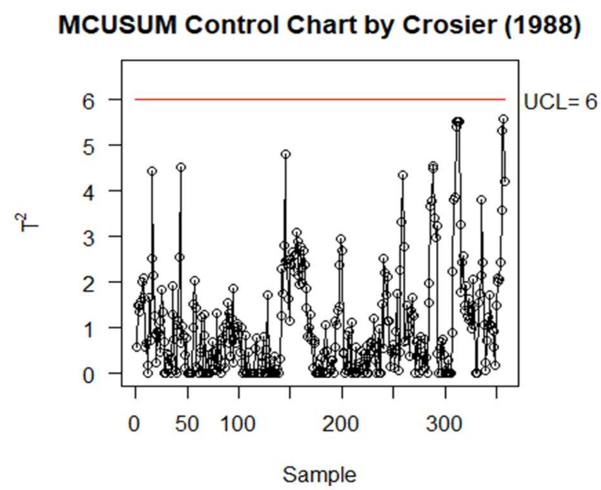



T^2 UCL = 14.64, m-CUSUM UCL = 6.0, LCL = 0, Out of Control Points = 1

**Iteration 7**




T^2 UCL = 14.64, m-CUSUM UCL = 6.0, LCL = 0, Out of Control Points = 0

We stop at iteration 7. In total 194 points were removed and 358 points remain.

**Concluding Remarks**

| Control Chart Type | ARL0 | No. of Iterations | In-control data points |
|---|---|---|---|
| Hotelling T^2 | 200 | 13 | 484 |
| m-CUSUM | 200 | 7 | 372 |
| m-EWMA | 200 | 6 | 437 |
| Combined T^2 & m-EWMA | 200 | 6 | 439 |
| Combined T^2 & m-CUSUM | 200 | 7 | 358 |

From the above provided table we can conclude that Hotelling T^2 control chart method had the most in-control data points after conclusion of Phase 1 analysis (484) whereas the Combined T^2 and CUSUM had the lowest (358). T^2 took the largest number of iterations (13) to reach its conclusion whereas both m-EWMA and Combined T^2 m-EWMA took the shortest (6 iterations each). It can be deduced that the T^2 might have been unable to detect small consistent mean shifts present in the given data records. That's why we decided to use Combination of T^2 with m-EWMA and m-CUSUM that would be able to detect small consistent mean shift along with large spikes in data. Combination of T^2 and m-EWMA had 439 in-control points whereas Combination of T^2 with m-CUSUM resulted in 358 in-control data points.