



---

## Topographic Vertical Change Monitoring by using Tabular and Simple Cusum Statistics

Justo Claudio Eduardo <sup>\*‡</sup>, Calandra María Valeria<sup>\*</sup>, Rodríguez Franco<sup>‡</sup>

<sup>\*</sup>UIDET GAMEFI, Departamento de Ciencias Básicas, Facultad de Ingeniería, Universidad Nacional de La Plata, Calle 115 entre 47 y 48, 2do piso-La Plata (1900), Argentina

<sup>‡</sup>Departamento Agrimensura, Facultad de Ingeniería, Universidad Nacional de La Plata, Calle 47 y 116, La Plata (1900), Argentina.

---

**Abstract** For the search of significant changes in the mean of vertical topographic measurements, of the order of one or two standard deviations of the observations, the most widespread method is the comparison of means using the T Student distribution. If the expected changes to be detected are of small magnitude, the above-mentioned test is not recommended. In this case, it is advisable to use more sensitive statistical tools such as the CUSUM (or cumulative sum) statistic. This allows estimating, by accumulation of residuals, changes in the population mean for both individual observations and samples. The latter alternative is applicable as today's surveying instruments can take samples in addition to single observations. This work was carried out with sequential observations applying two CUSUM modalities: simple and tabular. We first worked with data obtained by computer simulation and then with data taken at the campus of the Faculty of Engineering of the National University of La Plata. The results showed the effectiveness of CUSUM in detecting changes in the mean within the order of accuracy of the instrumentation used and the possibility of evaluating even minor changes due to assignable causes in the measurement process or due to a change in the measurand.

**Keywords** Change monitoring, CUSUM, Statistical Test, Altimetry.

---

### Introduction

The detection of changes in processes has begun to be investigated since the 1950s [1-3]. Many of the methods originate in the discipline of quality control, where an industrial manufacturing process is studied, seeking to detect failures as early as possible [4]. Today it is applied to various types of studies, such as climatological studies, in bioinformatics applications and fluid mechanics, specifically the analysis of the fluctuation of air velocity and pressure in turbulent flows [5-7], among many others. Change Point Analysis (CPA) is a tool for determining whether an abrupt or subtle change has occurred in a time series. The first works were carried out by Girshick and Rubin [8], using a Bayesian approach. Later Page [1] [9] [10] proposed the CUSUM (cumulative sum) technique in 1954, to detect the appearance of assignable causes in time series. The purpose of the CUSUM technique is to develop control charts capable of monitoring changes in both the mean value of a process and the standard deviation. There are both parametric [11] and non-parametric [12] techniques. Two techniques were applied in this work, the Simple CUSUM and the Tabular CUSUM. The latter, developed for data that come from a Normal distribution, which belong to a series of determinations of topographic height differences. The last two techniques mentioned were applied to establish if a change in the mean has occurred, due to the appearance of an assignable cause during the measurement process. It was implemented a work scheme that could be carried out on site with a number of observations appropriate to that context. Free software has also been used in order to provide autonomy to those interested in applying the developed methods.



## Methodology

### Preliminary work

To verify the effectiveness of the algorithm, first simulated data and later real observations were used. Both for data generation and processing, the R software [13] was used. For the observed data, a bar code level of 28 magnifications was used with a kilometric standard deviation, according to the manufacturer, of 0.0015m for double leveling. This standard deviation was verified using the complete, or long, procedure of the ISO (International Organization for Standardization) 17123-2.2001 [14] standard and was propagated according to Wolf and Ghilani [15] for simple stations with 30m sights. The value obtained was  $\sigma_{\Delta h}=0.0005\text{m}$  for a simple height difference. With this standard deviation, the data of the vector  $X$  were generated in such a way that for  $X_1$  to  $X_{20}$ ,  $X$  has a Normal distribution with  $\mu=1.5000\text{m}$   $\sigma_{\Delta h}=0.0005\text{m}$  and for  $X_{20}$  to  $X_{40}$  Normal distribution with mean  $\mu+0.0005\text{m}$  and the same standard deviation. Both the arbitrary value of  $\mu=1.5\text{m}$  and the value of  $\sigma$  are assumed to be constant during the simulated experiment of 40 observations. Subsequently, to obtain the real data, the bias of 0.0005m was caused by supplementing one of the fixed marks with a probe such as those used to measure the valve clearance in automobiles. Measurements were made by stationing the instrument at each determination of each of the height differences to achieve statistical independence.

### Description of the CUSUM Statistic

Given the values  $X$  belonging to a Normal distribution whose mean  $\mu_0$  can be estimated from values taken from a process in control, a change in it can be studied from the Tabular CUSUM statistic [11] [16], that allows to show positive Eq. 1 or negative Eq. 2 accumulation of residues

$$C_i^+ = \max[0, X_i - (\mu_0 + K) + C_{i-1}^+] \quad (1)$$

$$C_i^- = \max[0, (\mu_0 - K) - X_i + C_{i-1}^-] \quad (2)$$

With initial values  $C_0^+ = C_0^- = 0$ , the statistic indicates that the process went out of control when it exceeds the decision interval determined by  $H$ . The  $K$  value called the reference value in Eq. 1 and Eq. 2 is determined by the jump in the mean to be found. Let  $\mu_1$  be the mean of the out-of-control process such that  $\mu_1 = \mu + \delta \cdot \sigma$  is defined

$$K = \frac{\delta}{2} \cdot \sigma = \frac{|\mu_1 - \mu_0|}{2} \quad (3)$$

To make the decision as to whether the process has gone out of control, the  $H$  value is established. According to Montgomery [11], it is customary to use  $H=5$ , but it is suggested to link both  $H$  and  $K$  with  $\sigma$  through  $H=h \cdot \sigma$  and  $K=k \cdot \sigma$ . For example, in this case we seek to detect a change in the mean of  $1 \sigma = 0.0005\text{m}$ . To evaluate the effectiveness of the CUSUM, the ARL cannot be directly calculated (number of experiments performed, on average, until a data out of control is found) since the values of the statistic are successively related to each other [17]. Various authors such as Lorenzen and Vance [18] and Hawkins [19] made estimates and tabulations of the ARL. We worked with the one in Table 1 calculated by Lorenzen and Vance [18]. It determines the number of average samples necessary for the statistician to detect the out-of-control state of the process as a function of both the change in the mean and the sensitivity threshold  $H$ .

**Table 1:** This table produced by Lorenzen and Vance [18] shows the ARL as a function of  $\sigma$  and  $H$

Change in $\mu$ at multiples of $\sigma$	H=4	H=5
0	168	465
0.25	74.2	139
0.50	26.6	38
0.75	13.3	17
1.00	8.4	10.4
1.50	4.75	5.75
2.00	3.34	4.01



### Study of the CUSUM with simulated data

A sample  $X_1, X_2 \dots X_{40}$  ordered in time is simulated where,  $X_i = \mu_0 + \varepsilon_i$  with  $\varepsilon_i \sim N(0, \sigma_{\Delta h}^2) \forall i = 1, \dots, 20$  being  $\sigma_{\Delta h} = 0.0005m$  and where  $X_i = (\mu_0 + k) + \varepsilon_i$  with  $\varepsilon_i \sim N(0, \sigma_{\Delta h}^2) \forall i = 21, \dots, 40$  with  $\sigma_{\Delta h} = 0.0005m$ . By applying the Tabular CUSUM statistical method, it is intended to detect the change in magnitude  $K$  in the mean of the data, that is:  $\mu_i = \mu_0 \forall i = 1, \dots, 20$  and that  $\mu_i = \mu_0 + k \forall i = 21, \dots, 40$ . The first value to take into account for its design is the magnitude of the change to be detected in the mean, as shown in Eq 4.

$$K = \frac{\delta}{2} \sigma = \frac{|\mu_1 - \mu_0|}{2} = \frac{|1.5005 - 1.5000|}{2} = \frac{\delta}{2} 0.0005$$

$$\delta = 1 \text{ y } K = 0.00025$$

According to Table 1, for  $\delta=1$  and  $h=5$ , an average of 10 samples (in this case size=1) will be required to detect the change. For  $h=4$ , 8 samples will be required on average. The qcc package [20] of the R software [13] is applied. It must be indicated the data vector to be processed, the sample size, the control mean value, the  $H$  value and the change in  $\sigma$  units. With the R command

```
>cusum(X, 1, 1.5, 0.0005, decision.interval = 5(H), se.shift = 1, plot = TRUE)
```

Where  $X$  is the vector with the 40 height differences, 1 is the size of the sample because they are individual values, 1.5m is the arbitrary value of the height difference in control, 0.0005m is the standard deviation of the process, decision.interval is the value that will be given outside control the measurement process, se.shift is the number of standard deviations I want to find. Figure 1 shows the plotting of the measured values on the abscissa axis and the successive  $C_i^+$  and  $C_i^-$  corresponding to Eq. 1 and Eq. 2 on the ordinate axis. There are 6 observations prior to the establishment of the loss of control through the accumulated sum, which coincides with what is indicated in the table. However, it must be remembered that these tabulated values are an average of the expected number of results until the loss of control of the process occurs. Next to the graph of the Tabular CUSUM, Figure 1, a Simple CUSUM was plotted, Figure 2, which consists of simply accumulating the residuals with respect to the mean of the control process. In it, you can see the moment in which the change in the mean occurs and the loss of the control state of the process. This is a simpler alternative to approach when we are only interested in determining if the process went out of control, although it does not allow us to estimate the conditions in which the loss of control should manifest itself.

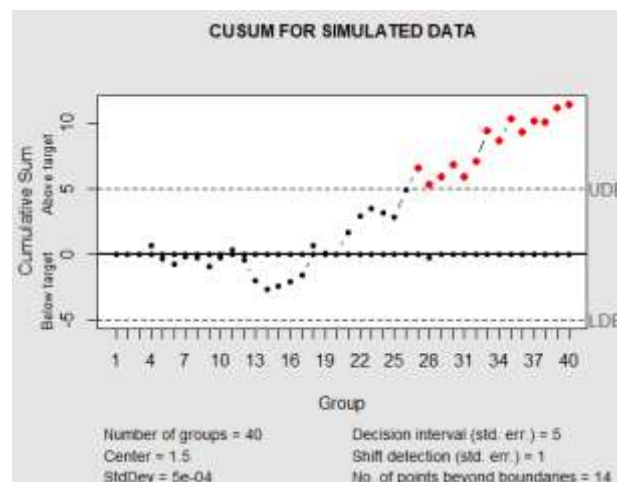


Figure 1: Tabular CUSUM



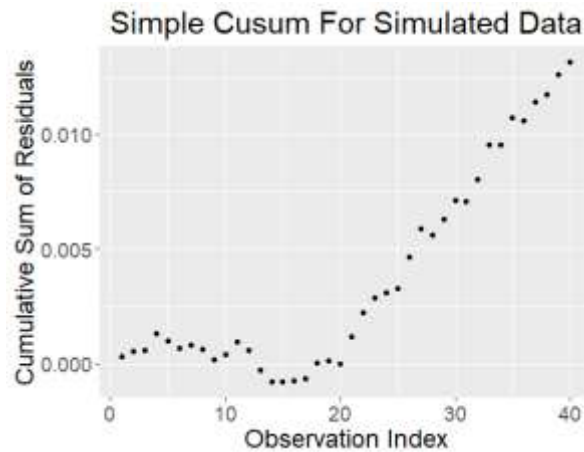


Figure 2: Simple CUSUM

The Simple CUSUM is calculated with Eq. 5:

$$C_i = \sum_{j=1}^i (\bar{X}_j - \mu_0) \quad (5)$$

It shows an accumulation of residues around the average value of the process in state of control. A quick visualization of the status of the measurement process can be seen, serving to justify or not a more detailed study through the Tabular CUSUM. Through the latter, it is possible to prepare control charts, that is, to establish limits and expected values for the process that is being carried out.

#### Study of the CUSUM with real data

40 height differences were measured individually with a level that reads barcode staff. From observation  $i=20$ , an increase in the height of the bench mark of  $1\sigma=0.0005\text{m}$  was introduced.

The Tabular CUSUM requires Normality in the data, which should be explored. For this exploration, descriptive statistics were used using tools such as the Normal QQPlot, the Shapiro-Wilk data Normality test, and a histogram. After studying the Normality of the data, the state in control of the measurement process must be determined by using Shewhart control charts for individual values which can be found in Montgomery [11] were used. The control charts for individual values and for the mobile range will indicate how the measurement process has developed in terms of its variability. This information is particularly important since the reference value  $\mu_0$  and  $\sigma^2$  must be verified or extracted from the first part of the sample. This part of the study prior to carrying out CUSUM is called Phase I. For the case study,  $\sigma^2$  can be chosen from the specifications of the manufacturer and verified through the long procedure of the ISO17123-2.2001 [14] standard. Likewise, by means of the moving range control chart, the behavior of the variability of the observations during the work can be studied. In addition to this, a Tabular CUSUM taking into account the sample standard deviation was carried out.

#### Previous studies on observations

The forty observations were plotted according to the order of measurement, Figure 3. The observations were then divided into two groups of twenty each. Phase I studies were carried out with the first twenty to establish the parameters of the process in a state of control. These studies consisted of a histogram, Figure 4, a sequential plot of the data, Figure 5, and a Normal QQplot, Figure 6. A Normality Test was also performed. Additionally, Mobile Rank and Individual Values charts were made. Once the control status of the measurement process was established, a Simple CUSUM and a Tabular CUSUM were carried out.



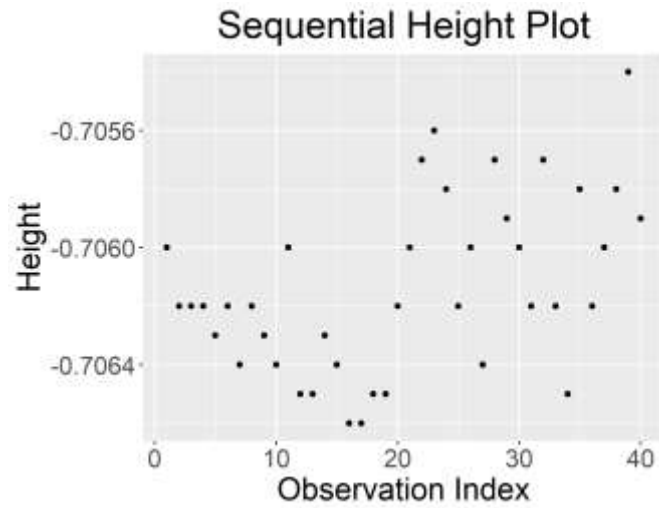


Figure 3: Sample index vs Observation

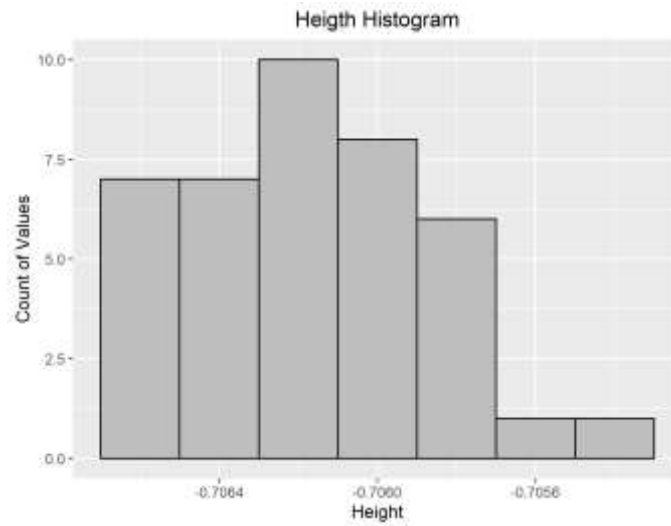


Figure 4: Height difference histogram

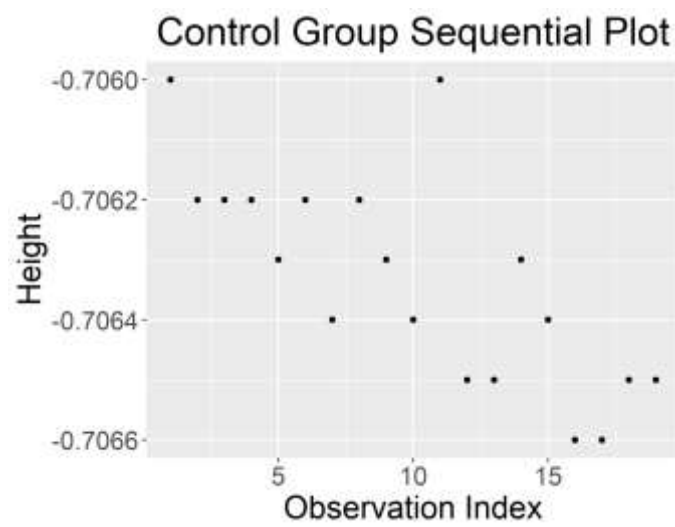


Figure 5: Control groups (n=1, ..., 20)



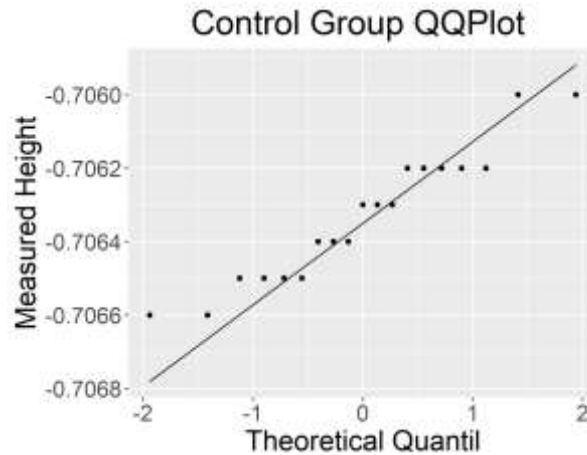


Figure 6: Control groups Normal QQ Plot

In the sequential plotting of Figure 3, is observed a decreasing trend in the values of the first 20 measurements. Subsequently, in the second 20, the process shows the horizontal band disposition, which is the expected, since it does not show a trend and it does show constancy in the variance of the process. From observation  $i=21$ , a point of change in the mean can be seen. In the Normal QQplot of Figure 6, which corresponds to the first 20 observations, it can be seen how the differences are grouped around the straight line, showing agreement between the theoretical quantiles and the sample ones. The application of the Shapiro-Wilk test to evaluate the lack of Normality of the first 20 data results in a statistic  $W = 0.93068$ , and a value  $p = 0.1591$ , so the Normality hypothesis of the data, necessary for the application, is not rejected. This condition is mandatory for the Tabular CUSUM, but not for the Simple CUSUM. The histogram in Figure 4 shows a distribution not visually associated with a normal distribution. Although histograms are recommended for around 100 data, they were included in order to show the behavior of the data with this tool. The Moving Ranges Chart, Figure 7, shows a value out of control. The individual values chart (Figure 7) shows all the values in control. For this reason, the validity of the work is not questioned. The size of the samples was chosen with a criterion of applicability in the field of an engineering work and based on the sample sizes used in the application of the ISO (International Organization for Standardization) 17123-2 [14] standard.

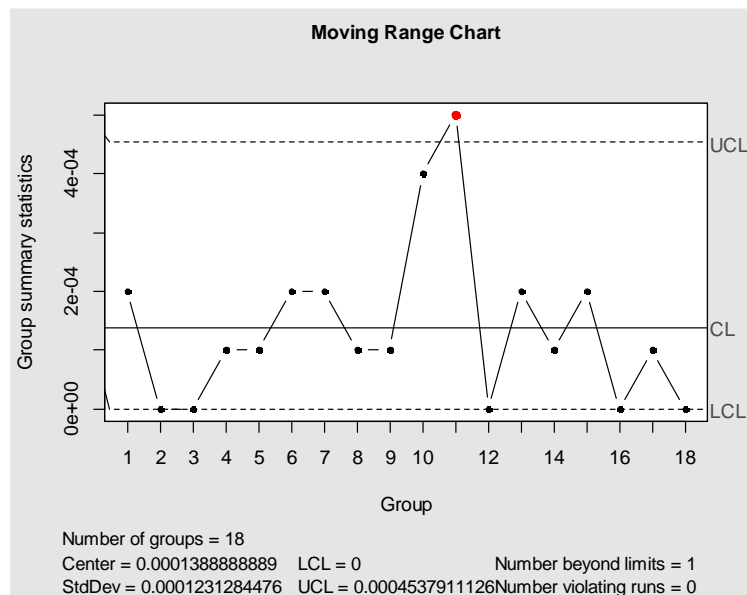


Figure 7: Mobile Rank Control Chart



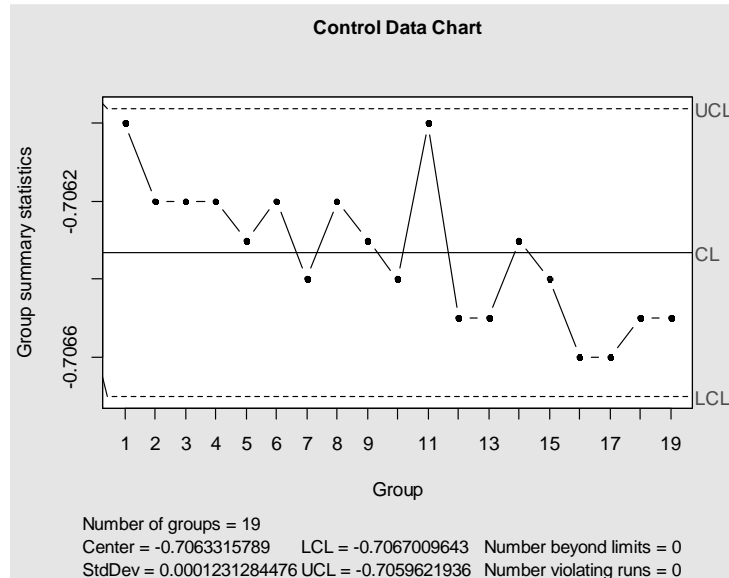


Figure 8: Individuals Control Chart

In the control chart of individual values for Phase I data, the trend of Fig.3 can be seen again, since it is basically the same information with another graphic scale. This chart shows no out-of-control values even for an estimated standard deviation even lower than the designed or reported by the instrument manufacturer. In the Simple CUSUM plot, Figure 9 and Tabular CUSUM plot of Figure 10, a change in the assignable causes of the process can be seen from observation 21. In the graph provided by the R software [13], the accumulation of significant residues with respect to the reference value can be seen and the upper and lower limits (UDB and LBD).

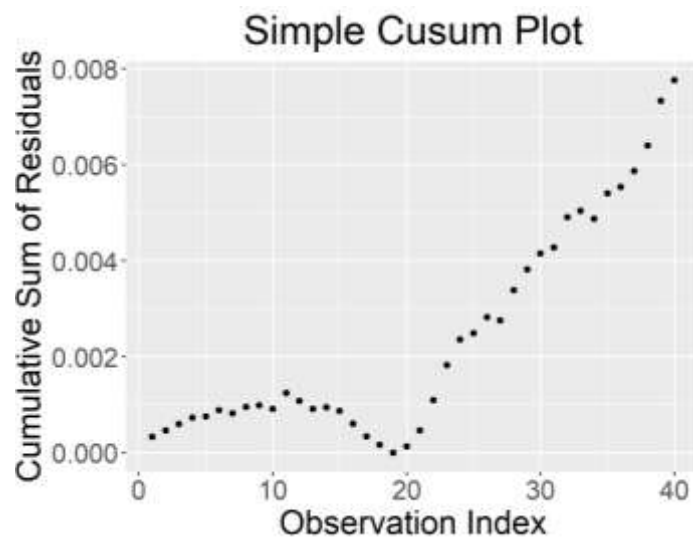


Figure 9: Simple CUSUM



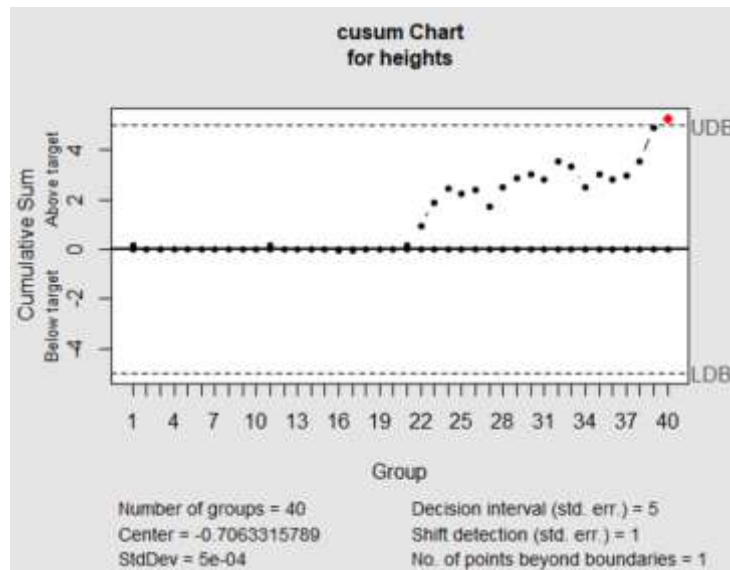


Figure 10: Tabular CUSUM

These limits indicate when to make the decision whether the process is in control or not. Only sample 40, that is, number 20 of the biased sample  $1\sigma=0.0005$ , would reach the limit that allows us to make the decision to reject the control of the process, when according to Table 1 it is expected that it will take an average of about 10 samples, which would lead to the observation 30 approximately. If the Tabular CUSUM is redesigned with Phase I sample parameters, Figure 11, both for  $\hat{\mu}$  and for  $s$ , since  $s=0.0003m$ , the desired change will then be equal to  $1.5 s$  obtaining a response closer to the expected one by Table 1 since the loss of control occurs in observation 9 of the second series when the expected value of it is for observation 6.

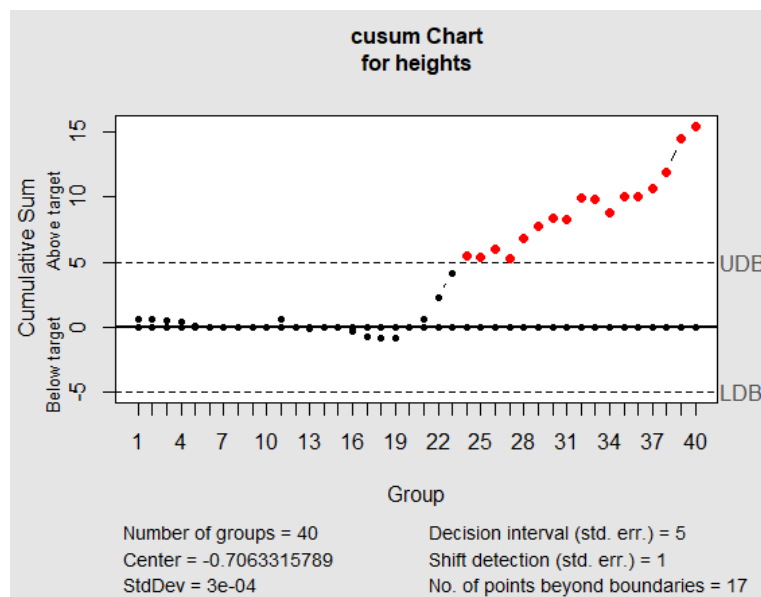


Figure 11: Phase II Tabular CUSUM

**Discussion**

In topography, to make decisions regarding whether or not two sample means belong to the same population, a comparison is used using the Student statistic [15] [21] [22]. In this work, the CUSUM statistic is presented as a tool that serves as a complement or alternative to the aforementioned one, since in the analysis of data through statistical tools, it is convenient to have several options simultaneously [23]. The study with data both simulated





and surveyed with topographical instruments shows that both the Simple CUSUM and the Tabular CUSUM [11] are sensitive to the appearance of a change immediately. This allows, eventually, having separated the samples, the use of the T Student statistic a posteriori to evaluate its significance. The small size of the samples is due to the need to present a method applicable on site and, therefore, a number of observations similar to that of ISO 17132-2 [14] is adopted with samples of size equal to 20. This presents difficulties when checking the Normality that is required for the application of the Tabular CUSUM [11] but not so to implement the Simple CUSUM. With so little data, the most appropriate tool to study Normality is the Normal QQPlot, however, it is always recommended to carry out histograms and Normality tests such as the Shapiro-Wilk test [23]. In this case, the histogram does not allow data with normal distribution to be inferred, although both the Normal QQPlot and the Shapiro-Wilk test do not allow the rejection of the Normality hypothesis. For the design of the Tabular CUSUM the Lorenzen and Vance criteria [18] was adopted, Table 1. These criteria must be seen as indicative. The design of CUSUM is in discussion [24]. What is mandatory before trying to detect changes in the mean is to carry out complete Phase I studies using Shewhart-type charts [11], since it will not only have knowledge of the status of the process under control, but we will also have information to establish the parameters in the design of the Tabular CUSUM. The analysis of these charts must be done in order to become aware of the dynamics of the process and not only evaluating the out-of-limit data but also the evolution of the process. In Phase I, for example, an undesirable decreasing trend is observed in a random sample even though there are no out-of-control points. To verify simultaneously the instruments and personnel that carry out the surveys, the application of the ISO 17132-2 [14] procedure is used. To study variability, the Moving Ranges chart is used, in which a single point is observed above the upper limit, although close to it. As soon as control of the variability of the process has been achieved, we are able to detect changes in the mean. We see as a limitation of the applied methodology that it is a single change in the mean, caused intentionally and in a controlled manner. For these reasons, we consider it is important to advance in the application of methods for detecting multiple changes in the mean and variance, both parametrically and non-parametrically.

### Conclusions

- The CUSUM method, both Tabular and Simple, showed that the change in the mean is exposed relatively quickly at the time the assignable cause appears. In our case, the assignable cause was the 0.0005m increase in the height of one of the benchmarks.
- The establishment of adequate values in Phase I will be fundamental so that the performance of the Tabular CUSUM is consistent with that predicted by Lorenzen and Vance [18], Table 1.
- Simple CUSUM despite its simplicity showed that it can be an applicable tool to the work in site for the engineer or even a technician.
- The CUSUM method also showed sensitivity to the change in variance since it was found that the variance estimated in Phase I was lower than that of the original design of the Tabular CUSUM. This sensitivity was exposed in compliance with the expected number of samples necessary to establish the loss of control of the measurement process.

### References

- [1]. Page, E.S. (1954). "Continuous Inspection Schemes". *Biometrika*, 41(1/2), 100-115. <https://www.jstor.org/stable/i315422>
- [2]. Roberts, S.W. (1959). "Control Chart Tests Based on Geometric Moving Averages", *Technometrics*, 1(3), 239-250. <https://doi.org/10.1080/00401706.1959.10489860>
- [3]. Pettitt, A.N. (1979). "A Non-Parametric Approach to the Change-Point Problem", *Journal of the Royal Statistical Society Series C - Applied Statistics*, 28(2), 126-135. <https://www.jstor.org/stable/2346729>
- [4]. Lai, T.L. (1995). "Sequential Change-point Detection in Quality Control and Dynamical Systems", *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(4), 613-658. <http://www.jstor.org/stable/2345934>
- [5]. Marañón Di Leo, J., Calandra, M.V. & Delnero, J.S. (2011). "Determinación de estructuras turbulentas mediante algoritmos numéricos", *Mecánica Computacional*, 30(6), 425- 434.
- [6]. Marañón Di Leo, J., Calandra, M.V. & Delnero, J.S. (2013). "Detección de eventos turbulentos mediante la aplicación de algoritmos de punto de cambio", *Mecánica Computacional*, 32(15), 1297-1308.



- [7]. Marañón Di Leo, J., Calandra, M.V. & Delnero, J.S. (2019). "CPM Model Applied to Turbulent Event Detections", *Journal of Scientific and Engineering Research*, 6(5), 173-182.
- [8]. Girshick, M.A. & Rubin, H. (1952). "A Bayes approach to a quality control model", *The Annals of Mathematical Statistics*, 23(1), 114-125. <http://www.jstor.org/stable/2236405>
- [9]. Page, E.S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42(3/4), 523-527. <https://doi.org/10.2307/2333401>
- [10]. Page, E.S. (1957). "On problems in which a change in a parameter occurs at an unknown point". *Biometrika*, 44(1/2), 248-252. <https://doi.org/10.2307/2333258>
- [11]. Montgomery, D.C. (2019). Introduction to Statistical Quality Control. 8th Edition. Wiley & Sons.
- [12]. Marañón Di Leo, J., Calandra, M.V. & Delnero, J.S. (2017). "Algoritmos de punto de cambio para la detección de estructuras vorticosas en campos fluidodinámicos turbulentos". *Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería*. 33(3/4), 225-234. <https://doi.org/10.1016/j.rimni.2016.04.009>
- [13]. R Core Team. (2016). R: A Language and Environment for Statistical Computing. Vienna, Austria. Recovered from: <https://www.R-project.org/>
- [14]. ISO17123-2:2001 (2001). Optics and optical instruments - Field procedures for testing geodetic and surveying instruments - Part 2: Levels. International Organization for Standardization (ISO).
- [15]. Ghilani, C.D. & Wolf, P.R. (2006). *Adjustment Computations: Spatial Data Analysis*. 4th Edition. Wiley & Sons. <https://doi.org/10.1002/9780470121498>
- [16]. ISO7870-4:2011 (2011). Control charts - Part 4: Cumulative sum charts. International Organization for Standardization (ISO).
- [17]. Prajapati, D.R. (2015). Effectiveness of Conventional CUSUM Control Chart for Correlated Observations. *International Journal of Modeling and Optimization*, 5(2), 135-139. <https://doi.org/10.7763/IJMO.2015.V5.449>
- [18]. Lorenzen, T. & Vance, L. (1986). The Economic Design of Control Charts: A Unified Approach. *Technometrics*, 28(1), 3-10. <https://doi.org/10.1080/00401706.1986.10488092>
- [19]. Hawkins, D.M. (1992). A Fast, Accurate Approximation of Average Run Lengths of CUSUM Control Charts, *Journal of Quality Technology*, 24(1), 37-43.
- [20]. Scrucca, L. (2004). qcc: an R package for quality control charting and statistical process control. *R News* 4(1), 11-17. Recovered from: [http://www.stat.unipg.it/luca/misc/Rnews\\_2004-1-pag11-17.pdf](http://www.stat.unipg.it/luca/misc/Rnews_2004-1-pag11-17.pdf)
- [21]. Ogundare, J.O. (2019). *Understanding Least Squares Estimation and Geomatics Data Analysis*. 1st Edition. John Wiley & Sons, Inc.
- [22]. ISO17123-1:2010 (2010). Optics and optical instruments - Field procedures for testing geodetic and surveying instruments - Part 1: Theory. International Organization for Standardization (ISO).
- [23]. Justo C. (2018). Tratamiento Estadístico de una Red Altimétrica Topográfica. Tesis de Maestría en Ingeniería UNLP. <http://sedici.unlp.edu.ar/handle/10915/65539>
- [24]. Argon Chen & Chen, Y.K. (2007). Design of EWMA and CUSUM control charts subject to random shift sizes and quality impacts, *IIE Transactions*, 39(12), 1127-1141. <https://doi.org/10.1080/07408170701315321>

