

## - Introduction

Laboratories are used to monitor the resulting data obtained from tests undertaken on quality control samples by means of the application of graphical tools known as Shewhart control charts for an analytical method; this kind of representations could be easily performed by using commercial spreadsheets.

The norm that allows to decide whether a method is under statistical control is usually derived from a certain number of well-established criteria with the aim of achieving reliable information about the daily interpretation of the control chart after each analytical run. Unfortunately, this decision might not be so practical to be implemented under every circumstance, since sometimes a large number of tests are required to detect that the method is out of statistical control, particularly when a long-term evaluation shows an out-of control situation.

**To prevent this situation, a Cusum chart graphic representation can be implemented.**

## - Cusum control charts

These charts involve cumulative calculations that take into account all the previous sample means to detect not only upwards or downwards trends but also small shifts in the mean of a process.

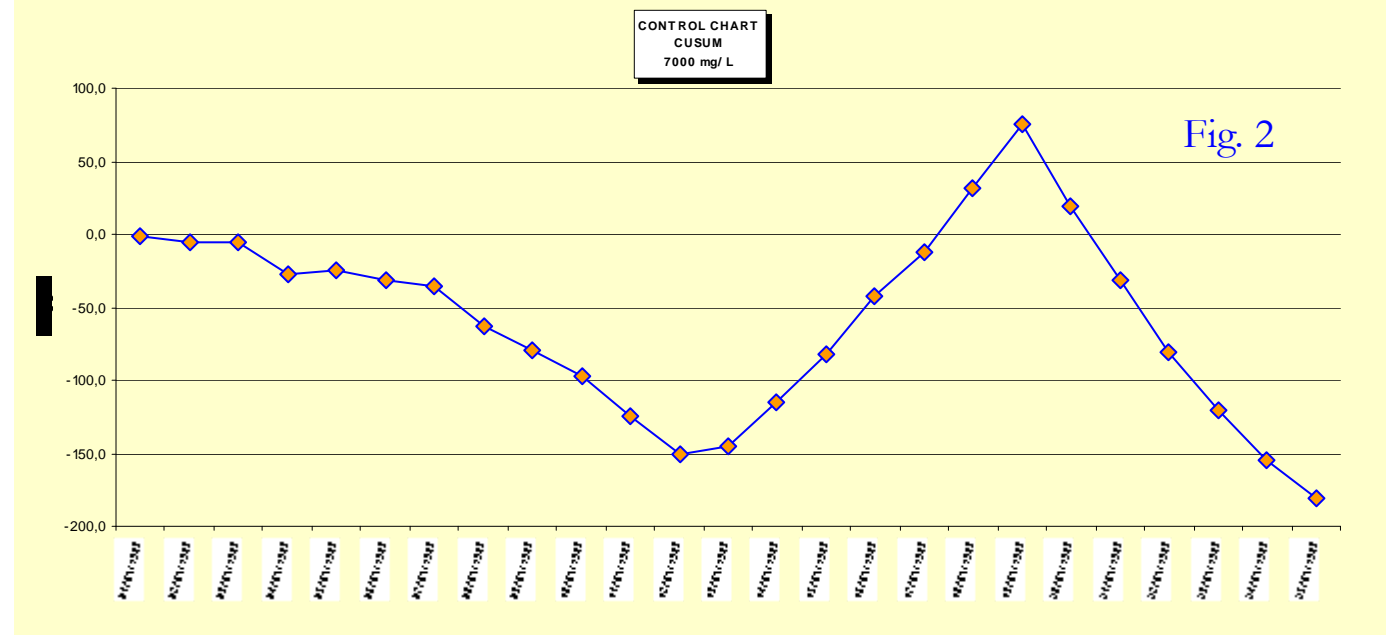
Cusum is an abbreviation for "cumulative sum", so it applies to the sum of the deviations of the sample means against the target or reference value, carried forward cumulatively from a number of consecutive results along time. If the method is under statistical control, in a quick inspection, the observed deviations between expected value and laboratory results should keep randomly distributed around zero.

## - Definition of Cusum chart

Let us suppose a typical X-chart as shown in Fig. 1 in which a series of measurements have been plotted, where the standard deviation  $s$  is calculated based on the analytical performance of the control sample, so the warning limits will be plotted in  $\pm 2s$  and the action limits in  $\pm 3s$ . In this case, just one observation fall outside the lower warning limit, but any evaluation related to the shift deviations from the reference value is achieved.

If we wish to know when the process mean has statistically varied, the construction of a Cusum chart is recommended, in which the cumulative deviations between measuring value and target value are represented in vertical axis against each observation, after every analytical series. For this purpose, a cumulative sum term is defined, being  $x_i$  the  $i$ th of the measuring value, where  $S_m$  is the total running of deviations of the  $n$  measurements from the reference value  $\mu_0$  (Fig. 2)

$$S_m = \sum_{i=1}^n (x_i - \mu_0)$$



## - Constructing a Cusum chart in a spreadsheet

The following steps have been implemented (Table 1):

- In first column (B; "sample") values corresponding to the observations sequence are filled.
- In second column (C; "ref. value") target value is introduced as an indication that the process is in statistical control; this reference could be filled just into a single cell: =C3.
- In third column (D; "result") the measuring results from each control sample in the analytical run is registered by filling each row; this value could be either an individual one or a mean value from the batch replication.
- In fourth column (E; "difference") the subtraction between measuring result and target value is calculated for each row: =D3-C3.
- In the first cell of the next column (F; "cusum"), it is included the same content of the adjacent cell in the right: =E3.
- In the remaining cells of this column (F; "cusum") the values calculated from the same row of the fourth column will be accumulated; for instance, the cumulative sum corresponding to the sample no.2 will be calculated as: =E4+F3.

Table 1: include the formulae to obtain the Cusum distribution in a commercial spreadsheet.  
Table 2: involved calculations to compute the corresponding numerical results.

To perform the Cusum chart of the analytical series, last step is to carried out the graphical representation of the observations against the cumsum values, as it is plotted in Fig.2.

- Both columns (B:sample) and (F:cusum) are selected for constructing the graph.
- The type of graph: XY Scatter Chart is chosen from the list.
- Data values from the first column are represented in the horizontal x-axis.
- Data values from the fifth column are represented in the vertical y-axis.
- Variable names, graph scales, legends and general aspect should be also considered.

## - V-mask form of the Cusum chart

Ideally the Cusum chart will show variations in a random pattern around zero, but if this were not the case, the use of a verification tool procedure, called V-mask because of its shape, will also indicate the observation from which the deviation should be considered as significant, depending on the definition of the control limits according to the selected design parameters.

Since the construction of V-mask is complicated in practice, we propose the definition not only of a visual representation but also the steps to accomplish a tabular Cusum procedure calculated in a sequential trial with a commercial spreadsheet form in which V-mask serves as control limits for the running cumulative term.

## - Definition of V-mask parameters

This tool is just a V-shaped mask which is superimposed over the Cusum chart with its axis horizontally and its vertex placed at a certain distance to the right of the last observation. The V-mask keeps moving forward with each new point plotted, so past observations are examined to see if any fall above or below the arms of the arm; as additional data are being collected, Cusum data are updated and the origin is relocated at the newest point.

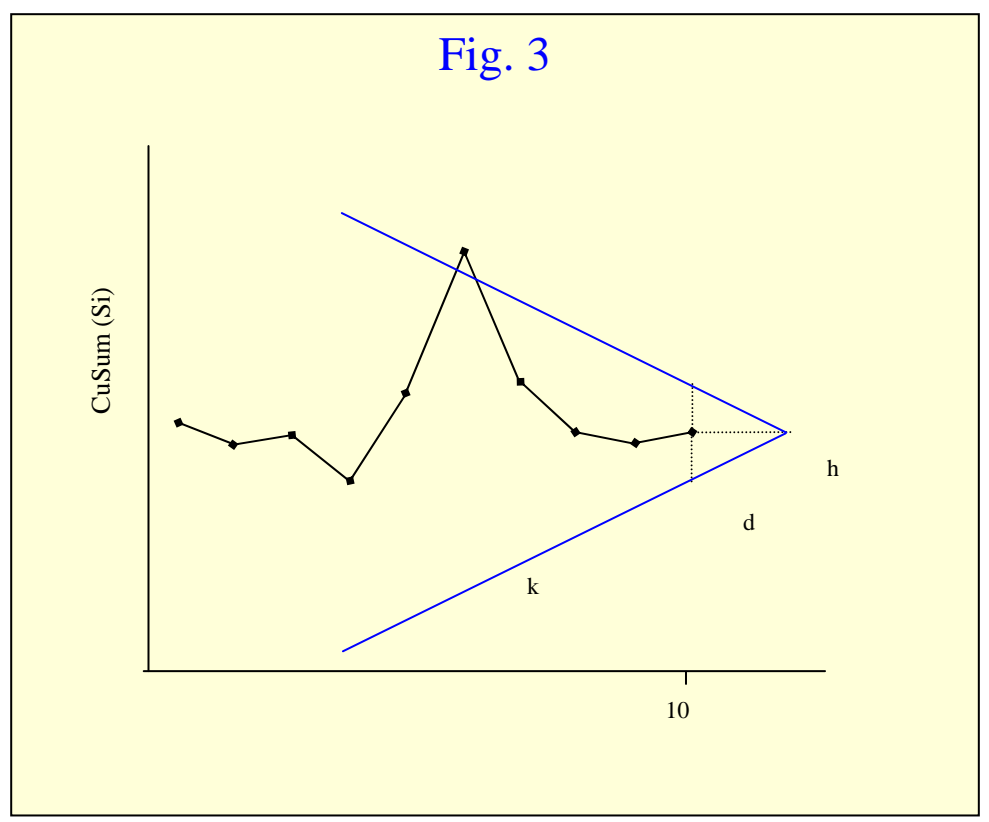
The verification that the situation is taken to be statistically under control is simple to understand:

- if all the values of the Cusum fall between the arms of the V-mask, the process is in-control.
- if one or more of the Cusum values lie outside of the mask arms, the process should be suspected of being out-of-control.

The application of V-mask requires a reliable definition of some design parameters estimated in advanced, so the performance of the mask depends largely on them. These variables need to be carefully chosen so that very few false alarms occur when the process is under control, but an important change in the process mean is quickly detected.

The V-mask design parameters are the following:

- $\mu$ : reference or target value of the mean process.
- $\sigma$ : standard deviation (sigma) settled for the process.
- $n$ : number of replicates used (subgroup size).
- $\alpha$ : probability of declaring a process out of control when in fact it is not (type I error); alpha risk determines the chance of detecting a drift that is not actually real.
- $\beta$ : probability of declaring a process in control when in fact it is not (type II error); beta risk determines the chance of having a drift that is not really detected.
- $\delta$ : amount of shift in process mean that we wish to detect (delta expressed in sigma multiples as the number of standard deviations from control mean at which a shift is detected).
- $k$ : slope of V-mask arms (two-sided scheme) in multiple units of sigma values; rise in the arm corresponding to one sampling unit.
- $h$ : vertical rise distance in the arm corresponding to distance  $d$  (between origin and the arm mask).
- $d$ : distance from the origin to the vertex of the mask..



## - Interpretation of the V-mask on the plot

The figure (Fig.3) explain how the geometric parameters are operating when the V-mask is set up.

Since these variables are interrelated you do not need to specify each one of them:

$$k = \frac{\delta \cdot \sigma}{2} \quad d = \frac{2}{\delta^2} \ln \left( \frac{1-\beta}{\alpha} \right) \quad h = d \cdot k$$

It is far more practical to construct the V-mask based in terms of probability errors, by choosing the appropriate alpha and beta values depending on the required test significance and confidence level in detecting the process shift. So, if any preceding point lies above the negatively sloped line, the process mean had shifted downward; on the contrary, if any preceding point lies below the positively sloped line, the process mean had shifted upward.

## - V-mask for a Cusum chart in a spreadsheet: tabular form

Let us suppose a real case in which the data used in the previous Cusum chart are computed in the next spreadsheets.

Data from the process (reference value, standard deviation, replicates):  
 $\mu$ : 7000 mg/l     $\sigma$ : 25 mg/l     $n$ : 1 replicate per observation

We also introduce the design parameters for the definition of V-mask:

Required data chosen from the acceptable quality level:

$\alpha$ : 0,0027 (equivalent to  $\pm 3$  sigma criteria in a Shewhart control chart)     $\beta$ : 0,01     $\delta$ : 1 (to detect a shift as large as 1 sigma)

Data calculated from the above chosen data, applying the relationships:

$k$ : 12,5     $d$ : 11,8089063     $h$ : 147,611329

## - Conclusions

The main advantage derived from the application of the V-mask form on the Cusum chart data is that by means of these control limits based in statistical parameters, a laboratory will be able to establish a decision procedure that allows determining what might have caused that a process can exhibit an anomalous pattern not justified only due to random variation, as well as the possibility of detecting unexpected shifts over time.

Some drawbacks of the proposed procedure are related to the ambiguity associated with the definition of delta, alpha and beta, apart from the difficulties on determining how far back to extend the V-mask arms, although this limitation it is also observed in other typical control charts.

Accordingly, without additional cost and with minimal effort in the daily work, the laboratory can have useful information coming from both traditional Shewart control chart and Cusum V-mask chart.

## - References

- Miller, J.C. and Miller, J.N. (1988). Statistics for Analytical Chemistry, Ellis Horwood Limited, 92-98.
- Nordtest Report TR 569 (2006). Internal Quality Control - Handbook for Chemical Laboratories.
- Woodall, W.H. and Adams, B.M. (1993). The Statistical Design of Cusum Charts, Quality Engineering, 559-570.
- Montgomery, D.C. (2000). Introduction to Statistical Quality Control, Wiley.
- César Pérez (1995). Análisis Estadístico con Statgraphics, Ed.: Ra-Ma.
- Francisco Charte Ojeda (2007). Cálculo Estadístico con Excel, Ed.: Anaya.



	A	B	C	D	E	F
1						
2		Sample	Ref. Value	Result	Difference	CuSum
3		1	7000	6998	=D3-C3	=D3-C3
4		2	7000	6997	=D4-C4	=F3+E4
5		3	7000	6999	=D5-C5	=F4+E5
6		4	7000	6978		
7		5	7000	7004		
8		6	7000	6992		
9		7	7000	6996		
10		8	7000	6973		
11		9	7000	6983		
12		10	7000	6983		
13		11	7000	6973		
14		12	7000	6974		
15		13	7000	7005		
16		14	7000	7030		
17		15	7000	7033		
18		16	7000	7039		
19		17	7000	7030		
20		18	7000	7045		
21		19	7000	7044		
22		20	7000	6943		
23		21	7000	6950		
24		22	7000	6950		
25		23	7000	6960		
26		24	7000	6966		
27		25	7000	6974		

Sample	Ref. Value	Result	Difference	CuSum
1	7000	6998	-2	-2
2	7000	6997	-3	-5
3	7000	6999	-1	-6
4	7000	6978	-22	-28
5	7000	7004	4	-24
6	7000	6992	-8	-32
7	7000	6996	-4	-36
8	7000	6973	-27	-63
9	7000	6983	-17	-80
10	7000	6983	-17	-97
11	7000	6973	-27	-124
12	7000	6974	-26	-150
13	7000	7005	5	-145
14	7000	7030	30	-115
15	7000	7033	33	-82
16	7000	7039	39	-43
17	7000	7030	30	-13
18	7000	7045	45	32
19	7000	7044	44	76
20	7000	6943	-57	19
21	7000	6950	-50	-31
22	7000	6950	-50	-81
23	7000	6960	-40	-121
24	7000	6966	-34	-155
25	7000	6974	-26	-181

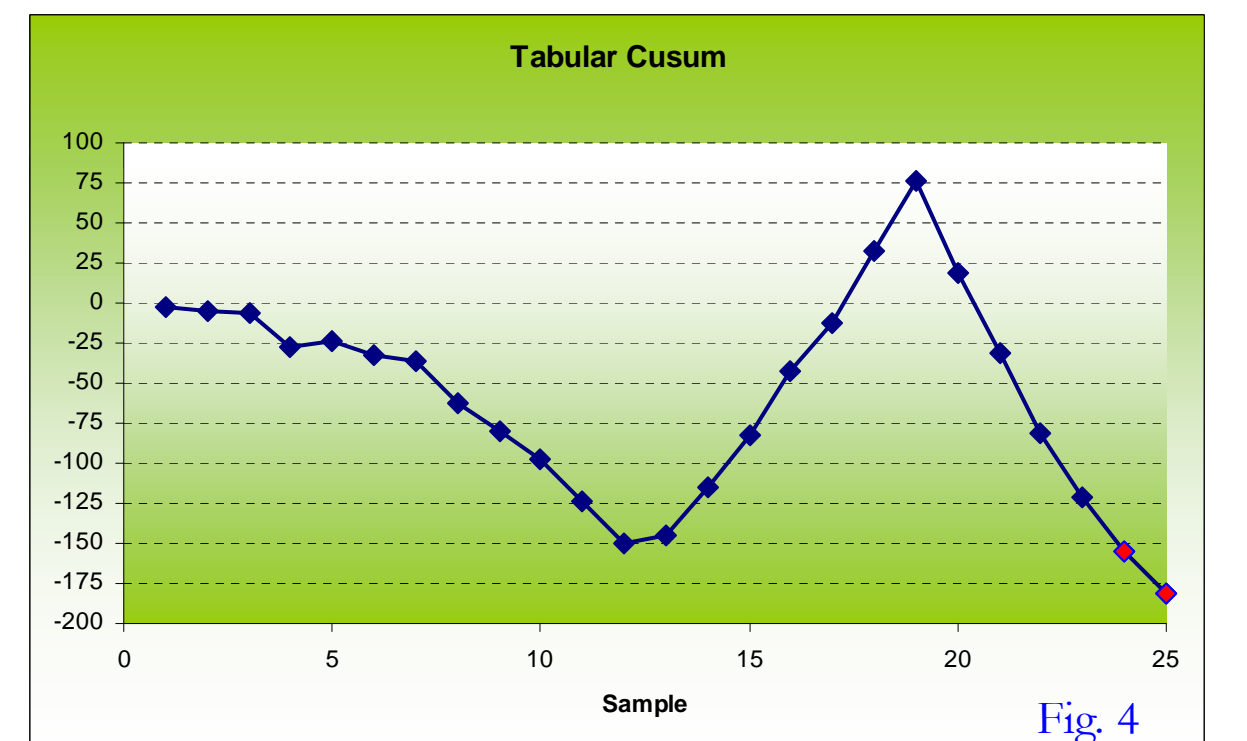
To generate the V-mask in tabular form in a spreadsheet (table 3), the values of h and k parameters calculated in the original data units are introduced in these expressions:  
 $S_{hi}(i) = \max(0, S_{hi}(i-1) + x_i - \mu - k)$      $S_{lo}(i) = \max(0, S_{lo}(i-1) + \mu - k - x_i)$

	A	B	C	D	E	F	G	H	I
1									
2		$\mu$ :	7000		$\delta$ :	1		$k$ :	=E2*B5/2
3		$\sigma$ :	25		$\alpha$ :	0,0027		$d$ :	=E2^2*(1-E4)/E3
4		$n$ :	1		$\beta$ :	0,01		$h$ :	=H3*H2
5		$s$ :	=B3/B4						
6									
7									
8	Sample	Result	Difference	CuSum	Incr.	$S_{hi}$	Decr.	$S_{lo}$	
9	1	6998	-2	-2	=C9-B\$2-\$H\$2	=0	=-C\$2-\$I\$2-D10	=0	
10	2	6997	-3	-5	=C10-B\$2-\$H\$2	=MAX(0;F9+G10)	=-B\$2-\$I\$2-C10	=MAX(0;I9+H10)	
11	3	6999	-1	-6					
12	4	6978	-22	-28					

Once performed the calculations, the following results are obtained:

	A	B	C	D	E	F	G	H	I
1									
2		$\mu$ :	7000		$\delta$ :	1		$k$ :	12,5
3		$\sigma$ :	25		$\alpha$ :	0,0027		$d$ :	11,8089063
4		$n$ :	1		$\beta$ :	0,01		$h$ :	147,611329
5		$s$ :	25						
6									
7									
8	Sample	Result	Difference	CuSum	Incr. ( $x_i - \mu - k$ )	$S_{hi}$	Decr. ( $\mu - k - x_i$ )	$S_{lo}$	
9	1	6998	-2	-2	-14,5	0	-10,5	0	
10	2	6997	-3	-5	-15,5	0,0	-9,5	0,0	
11	3	6999	-1	-6	-13,5	0,0	-11,5	0,0	
12	4	6978	-22	-28	-34,5	0,0	9,5	9,5	
13	5	7004	4	-24	-8,5	0,0	-16,5	0,0	
14	6	6992	-8	-32	-20,5	0,0	-4,5	0,0	
15	7	6996	-4	-36	-16,5	0,0	-8,5	0,0	
16	8	6973	-27	-63	-39,5	0,0	14,5	14,5	
17	9	6983	-17	-80	-29,5	0,0	4,5	19,0	
18	10	6983	-17	-97	-29,5	0,0	4,5	23,5	
19	11	6973	-27	-124	-39,5	0,0	14,5	38,0	
20	12	6974	-26	-150	-38,5	0,0	13,5	51,5	
21	13	7005	5	-145	-7,5	0,0	-17,5	34,0	
22	14	7030	30	-115	17,5	17,5	-42,5	0,0	
23	15	7033	33	-82	20,5	38,0	-45,5	0,0	
24	16	7039	39	-43	26,5	64,5	-51,5	0,0	
25	17	7030	30	-13	17,5	82,0	-42,5	0,0	
26	18	7045	45	32	32,5	114,5	-57,5	0,0	
27	19	7044	44	76	31,5	146,0	-56,5	0,0	
28	20	6943	-57	19	-69,5	76,5	44,5	44,5	
29	21	6950	-50	-31	-62,5	14,0	37,5	82,0	
30	22	6950	-50	-81	-62,5	0,0	37,5	119,5	
31	23	6960	-40	-121	-52,5	0,0	27,5	147,0	
32	24	6966	-34	-155	-46,5	0,0	21,5	168,5	
33	25	6974	-26	-181	-38,5	0,0	13,5	182,0	

In the worksheet, a conditional formatting is applied to columns (H; " $S_{hi}$ ") and (I; " $S_{lo}$ ") in which the decision rule has been calculated; any value greater than h parameter is considered as out of statistical control, so it will be displayed in red colour according to the quality requirements.  
A graphical picture of tabular form data is plotted in Fig.4, in which two out-of-control observations are identified:



## - V-mask for a Cusum chart in a spreadsheet: graphical form

To construct and plot the V-mask form of the Cusum chart, a new set of variables must be introduced in order to represent the upper and lower arm of the V-mask in a spreadsheet.

Table 5: Parameters for the V-mask graphical form.

	B	C	D	E
--	---	---	---	---