Chapter 4

Proposed Solutions

There are many methods of looking for leaps of the mean value of a statistical process. Some of them have already been tried, some others are not appropriate for our task and some may suit the requirements we are dealing with. Following, we make an explanation of some of the most extended and we discuss their properties, advantages and drawbacks.

4.1 Hinkley detector

The Hinkley detector is a method of looking for changes of the mean of a process that are small compared to the standard deviation. In this method, deviations of the data from the mean value are integrated as far as they are higher than a specific threshold. When the process mean changes, the value of this integral is expected to grow (or decrease) statistically without end. When another threshold must be established to decide at which value of the integral the mean value is considered to have changed.

The election of these thresholds needs knowledge in advance of the magnitude of the expected step and also of the statistical distribution of the process.

A variant of this algorithm known as *Multilevel-Hinkley-detector* has already been tried on the scale in the HTWG Konstanz with some success. In the moment a patent describing the algorithm [3] is being transacted. Though, the algorithm does not use the historic data, so a new algorithm is required that enables a further offline check up of the results and, anyway, its behavior under ground vibrations needs to be improved.

4.2 Gradient test

The Gradient test consists on periodically calculating a linear regression of a certain number of preceding measuring points. The resultant line is then examined respect to its gradient. If it exceeds the gradient limit previously defined for each specific process, a leap is considered to have happened and an *alarm* is produced.

This method is typically used in biological studies to find *events* (unusual changes of the control readings). Although this method is appropriate to detect sudden changes of the process parameters, it does not give any information about the magnitude of these changes, but just about their celerity. Thus, it is not suitable for the counting scale.

4.3 Statistical test for quality control

This generic kind of test, commonly used to supervise industrial processes, consists on constantly calculating the value of a control variable named *test statistic* and checking out if it remains into a predefined control area. When the variable surpasses the control limits, an *alarm* is triggered indicating that the statistical process is, within a certain confidence level, out of control or, in other words, that the value of the tracked parameter changed. In such case like the counting scale's, those alarms should correspond the weight changes.

For a conclusion about the suitability of this kind of test to the problem of the counting scale, we need to explain some concepts related to *hypothesis tests*:

Null hypothesis. Is the assumption about the statistical process to be tested. In our case, the *null hypothesis* says that the weight (or the number of pieces) on the scale remains still. If the null hypothesis is rejected, then the *alternative hypothesis* ("the weight on the scale changed") is accepted.

Type I error. Is the error committed if the null hypothesis is rejected when it is actually true. In our case it would be the error we commit if we consider that a new unit fell onto the scale when it did not happen.

Type II error. Is the error committed if the null hypothesis is accepted when it is false.

Significance level α. Is the probability of committing type I errors. The probability of committing type II errors is designed by β.

Power of a statistical hypothesis test. Is the probability of rejecting the null hypothesis when it is actually false. That is, the probability of *not* committing type II errors. $Power = 1 - \beta$.

Statistical tests are designed looking for the least number of errors to be obtained. Thus, on the one hand, α is pretended to have a low value in order to reduce *false alarms* (i.e.: type I errors) and, on the other hand, the *power* of the test must be high (low value of β) to reduce type II errors. The problem is that the first objective is achieved increasing the width of the control area and that leads to high values of β (i.e.: low power). This means that, for a given distribution of a *test statistic* and for a given *significance level*, the power of the test is determined and cannot arbitrarily be selected.

As the Statistical analysis will show (chapter 5), in the case of the counting scale the weight increment to be detected is very small compared to the standard deviation. Precisely in these cases, the power of statistical tests tends to be too low for acceptable values of α. As a consequence, this test is not directly feasible. Though, if we use the average value of a sample that is big enough, instead applying the test directly to the reading of the sensors, we can reduce the standard deviation and the control area as much as necessary to reach the desired power for the test without increasing the significance level, α. That is exactly the fundamental of the *double-sigma test* explained next.

4.4 Double-sigma test

The double-sigma test was developed in Hamburg in the early 90's for detecting significant changes in the water quality measurements [4]. The sample mean value is used in this test as control variable. An "event" is considered to have occurred when the difference between the sample mean value and the mean expected value of the process is higher than two times his standard deviation: $|\bar{x} - \mu| > 2\sigma$.

This test could be suitable for the counting scale with the use of a large sample whose size should be calculated for warranting the required power of the test. Nevertheless, the control area size 2σ does not ensure the specified significance level.

After the statistical analysis in chapter 5, it will be proved that the interval size 2σ is not enough to fulfill the specifications, and a variation of this test with a properly sized control area will be chosen.