COMPARATIVE ANALYSIS OF PROCESS CONTROL CHARTS IN ORDER TO DEFINE MOST EFFICIENT SAMPLING FREQUENCY

Lappeenranta–Lahti University of Technology LUT

“International Master of Science in Engineering, Entrepreneurship and Recourses” (MSc. ENTER)

2021

Amer Bjelonja

Examiner(s): Prof. Satu – Pia Reinikainen, LUT

Dr. Thomas Leissner, TU BAF

Prof. Branko Vučijak, UNSA

MSc ENTER Engineering, Entrepreneurship and Resources

Master's thesis

for the Joint Study Programme

"International Master of Science in Engineering, Entrepreneurship and Resources"

(MSc. ENTER)

TOPIC: COMPARATIVE ANALYSIS OF PROCESS CONTROL CHARTS IN ORDER TO DEFINE MOST EFFICIENT SAMPLING FREQUENCY

edited by: Amer Bjelonja

for the purpose of obtaining one academic degree (triple degree) with three diploma certificates

Supervisor / scientific member (HU):	Prof. Branko Vučijak
Supervisor / scientific member (LUT):	Prof. Satu – Pia Reinikainen
Supervisor / scientific member (TU BAF):	Dr. Thomas Leissner

Handover of the topic: 31.03.2021

Deadline of the master's thesis: 17.12.2021

Place, date: Sarajevo 08.12.2021

…………………………………	…………………………………	…………………………………
(Prof. Branko Vučijak)	(Prof. Satu – Pia Reinikainen)	(Dr. Thomas Leissner)
Supervisor / member HU	Supervisor / member LUT	Supervisor / member TU BAF

Supported by

eit RawMaterials ACADEMY

This activity has received funding from the European Institute of Innovation and Technology (EIT), a body of the European Union, under the Horizon 2020, the EU Framework Programme for Research and Innovation

ABSTRACT

Lappeenranta–Lahti University of Technology LUT
LUT School of Engineering Science
"International Master of Science in Engineering, Entrepreneurship and Recourses" (MSc. ENTER)

Amer Bjelonja

Comparative analysis of Process Control Charts in order to define most efficient sampling frequency

Master's thesis
2021
58 pages, 27 figures, 6 tables and 2 appendices
Examiner(s): Professor Satu – Pia Reinikainen, Prof. Branko Vučijak, Dr. Thomas Leissner

Keywords: Control charts, Walter A. Shewart, sampling frequency.

This thesis conducts various comparative analysis of control charts in order to define most efficient sampling frequency. Control charts analyzed through this work are tools used in process of quality monitoring invented by Walter A. Shewart and have been used ever since in every branch, field or process that can or needs to be monitored. The paper contains theoretical as well as analytical part of creation of I, R and CuSum charts to define most efficient sampling frequency. Thesis focuses and explains process of making different control chart types as well as indicates problems and proposes solutions to problems faced during the making those charts, to create efficient monitoring system of water levels. According to conducted analysis for multiple charts to monitor water level for *i* charts real standard deviation must be used instead of approximated one and most trustful charts are created when smaller time period is monitored, one or two months. The dataset analyzed originally contains 52608 samples of water levels measured every 30 minutes through the period of 3 years. Original dataset is further categorized to make 3 different datasets, one for every year, which is even further modified to simulate sampling frequencies of 60-, and 90-minutes sampling.

THESIS ASSIGNMENT

Ak. 2020/21

University of Sarajevo
Mechanical Engineering Faculty Sarajevo
Department for Industrial Engineering and Management

Topic title:	**Comparative analysis of process control charts in order to define the most efficient sampling frequency**
Mentor/s:	**Branko Vučijak**
Scientific field:	**Industrial Engineering and Management**

Subject and explanation of topic:

To control the quality of the production process, it is necessary to constantly monitor the selected product values and check compliance with the limits of the specification. Statistical control charts are a good tool for checking whether additional and unwanted variability has occurred in the process, in order to eliminate the same and ensure the required product quality.
The paper will analyse the ability of selected process in a company, specifically will check the process variability with different sampling frequencies. Control charts will be practically fully developed for the intensive sampling frequency, and its main characteristics including warning and action lines' values and samples found to be outside of these lines will be compared with the same characteristics of the less intensive sampling frequency. Practical application of calculation of appropriate capability indices and the development of relevant control charts will provide basis for appropriate recommendations for the process improvement and the most efficient sampling frequency.

Mandatory literature:

[1] Bajrić H., Kadrić E. & Vučijak B. (2018). Six sigma - osnovni statistički alati (Six Sigma – Basic Statistical Tools), Mechanical Engineering Faculty Sarajevo
[2] Oakland, J. (2011). Statistical Process Control, Routledge, New York.
[3] Montgomery, D.C. (2005). Introduction to Statistical Quality Control, John Wiley & Sons, New York.
[4] Thompson J. R. & Koronacki J. (2002). Statistical Process Control - The Deming Paradigm and Beyond, Chapman & Hall/CRC, Washington D.C.

STATEMENT OF ORIGINALITY

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with standard referencing practices.

Place, date: Sarajevo, December 2021

Amer Bjelonja

Table of Content

List of Figures

List of Tables

List of symbols and abbreviations

CC Control Chart

SPC Statistical Process Control

CL central line

UCL upper control limit

LCL lower control limit

UWL upper warning limit

LWL lower warning limit

SH high side cumulative sum

SL low side cumulative sum

σ standard deviation, a measure of process quality

R range, the difference between the highest and lowest value of the result

1 Introduction

Control charts, also known as Shewart charts that were invented by dr. Walter A Shewart are infallible part of any quality monitoring process in every company. In today's world quality of the process is tightly related to the product quality and competitiveness among companies on the market. Quality of the given product that is offered on the market directly depends of the production process during which it is created. In order to monitor and improve production process and ultimately the quality of the product control charts are used. To gain useful control charts, monitoring of the production process is needed, including measuring key parameters. Measuring and monitoring is often conducted via various sensors, scales or people. Data collected during measuring and monitoring process are necessary inputs for creating control charts and for monitoring or improving quality of production process. (Woodall, Spitzner, Montgomery & Gupta, 2004)

Shewart Charts can be used for monitoring not only quality of production but also monitoring or even predicting any measurable processes such as production, pollution, consumption, etc. As formerly mentioned, Control charts have very wide area of application ranging from quality control in any production (auto industry, food industry, chemical industry, etc.), and all the way to the predicting chances of natural disasters to happen. They have only one clearly defined purpose, and that is to optimize the process and ensure maximum quality of the product by eliminating all possible errors during production process and reduce all unnecessary costs.

Crucial and most important part for creating high quality control charts is input data. Input data are measurements taken during production process by using sensors, scales or manually. If data are not correct, or if sampling frequency was too small or too large, this might lead to generating false alarms which might tell operator that system is not good, or the real issue of special cause of variation appearing might be missed and not noticed. This causes large problems because operator needs to trust to the control charts.

This thesis aims to analytically compare process control charts with various sampling frequency to define the most efficient one. It will analyse the ability of selected processes, specifically will check the process variability with different sampling frequencies and recognition of key warning and action points within them. Control charts will be practically fully developed for the very intensive sampling frequency, and its main characteristics including warning and action lines' values and samples found to be outside of these lines will be compared with the same characteristics of the less intensive sampling frequency. Practical application of calculation of appropriate capability indices and the development of relevant control charts will provide basis for appropriate recommendations for the process improvement and the most efficient sampling frequency.

Topicality and novelty of this work lies in application of process control charts for prediction of natural disaster as well as in the determination of right sampling frequency for most efficient and productive monitoring of water level in correlation with amount of precipitation to potentially define early warnings for floods. This topic seems very interesting since it has great potential for broad use, however there is not many scientific papers written on this topic, therefore this master thesis will contribute greatly to entire academic community.

2 Literature Review

To better understand current situation regarding the use of control charts and the role the sampling frequency plays in creation of control charts this chapter will contain short reviews of publications made on this topic.

Proof that Control Charts can be used to enhance the quality can be found the work of Nefise Gonul Sengoz (2018). Author states that these Charts are very valuable tools used for quality control and that usage of Control charts highly contributes to reducing the overall production costs. Many different Control charts are used in the practice, some of them are ranges control chart (R), means control chart $(\bar{x})$ and standard deviation control chart (s). Most important factors for the proper performance of control charts are sensitivity, sampling frequency and sample size. In order to create working control charts upper and lower control limits are required. These limits are calculated using different equations for each individual control chart. Process is stable if plotted points in the charts are between UCL and LCL, but even then, plotted patterns must be analyzed to keep the production under control. In the case of variation or the points going over or above UCL and LCL corrective measures need to be taken (Sengoz, 2018).

Maugeri and Arcidiacono (2014) any SPC practices might be undermined in case of accepting Gaussianity or neglecting correlations. In order to overcome the problems of missing data in the set or solve the problems of non-normal data Special Control Charts have been designed. They presented method to identify effective frequency and size of the sample in Shewart Control Charts (CC's). Paper does not take in concern independence of data or their distribution. Results depends on Nyquist-Shannon sampling theorem and information about average time of the signal. It shows that Nyquist criteria in relation with time scale considering the CTQ variability permits smaller sample size and collecting rate of general Shewart CC's. In the end of their work, authors claim that: *"The sampling rate lower bound is the Nyquist rate, while the sample size lower bound is the product of the Nyquist rate times the correlation time "*(Maugeri & Arcidiacono, 2014).

To prove the point that control charts can and are used in many different fields and not only production industry, next couple of literature reviews will also focus on usage of control charts in analytical laboratory, chemical and food industry, chemical plants health care as well as for interpreting environmental monitoring data and a study of the effects of climate change on precipitation patterns.

„Control charts have been found useful in the testing laboratory whenever it is desirable to compare the over-all variability of test data with the average variability of small groups of the data, and they are simpler to understand than the more complicated statistical methods of analysis of variance" is the conclusion of Walter J Murphy (1945). Although the research was conducted a long time ago, all future research was based on the fact that Murphy stated in his work. Author claims that it is necessary to control larger number of influencing parameters which effect individual testing variation in smaller groups, because they lack precision. Testing methods which are affected by uncontrolled factors often display deficiency of precision. These factors can be pointed out and put under control, however sometimes controlling, or removing those factors is not what is wanted. In case situation like that happens, changes in the small groups should be made in a way so every group is influenced by unlimited factors. Analysis of these charts shows that usual testing methods are not exact as they were considered earlier. Usually, it is shown that variations which are tested are lesser than the variations in the process which the test is designed to control. As long as differences of test are not bigger than the ones in the process, repeatable using of single samples proves to be nothing than wasteful. Murphy finishes his paper with conclusion that use of control charts in analytical laboratory is unlimited. (Murphy,1945)

Control charts are especially significant in the food sector, where product selection and monitoring are critical for meeting client expectations while maintaining a well-organized manufacturing process and services (Özdagoglu, Özdragoglu, Güler, 2018). To reach desired solutions authors used CUSUM and EWMA techniques. These techniques provided necessary visualization of the entire process and pointed out the which points or samples in this case are out of the control lines. Interesting thing is that authors used these techniques to predict future values and their position according to the upper and lower limit lines. Upon the testing first model contained high rate of relative errors, authors have proposed development of Markov chain components, since this model provides enhanced prediction. (Özdagoglu et al., 2018)

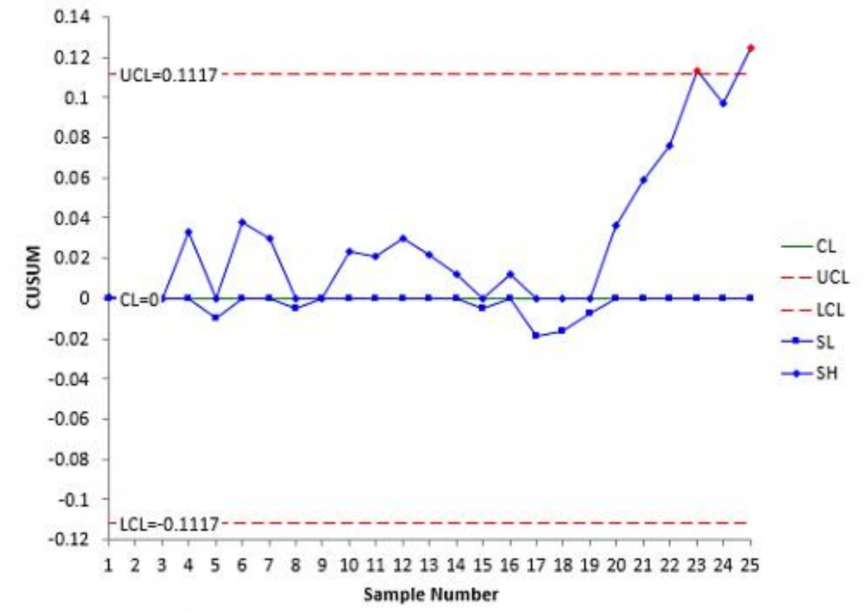


Figure 2. 1 Graphical representation of CUSUM chart

(Source: Özdagoglu et al., 2018)

CUSUM chart shown in Figure 2.1 is defined by:

- Center line (CL)
- Upper control line (UCL)
- Lower control line (LCL)
- High side cumulative sum (SH)
- Low side cumulative sum (SL)

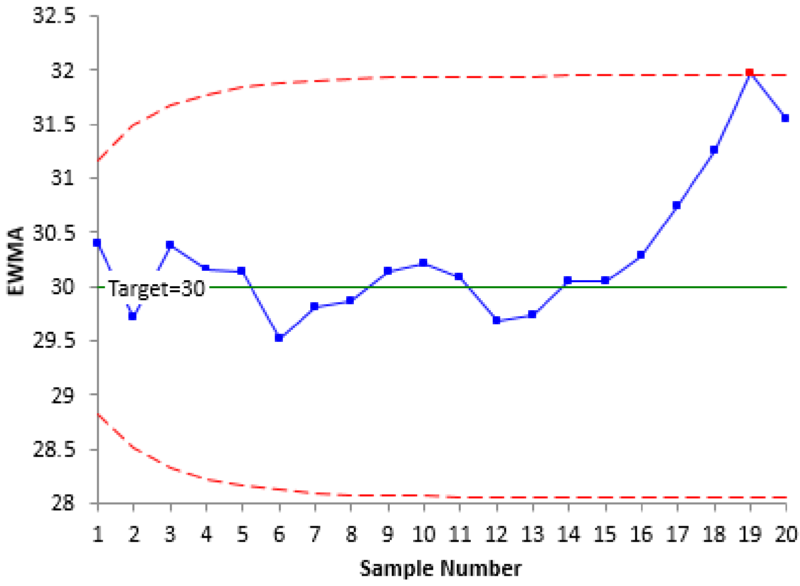


Figure 2. 2 Graphical representation of EWMA chart

(Source: spcforexcel.com)

Figure 2.2 represent EWMA – Exponentially Weighted Moving Average chart. This chart is used to recognize small differences from process average value. Red lines in the chart represent upper and lower control lines while green line represents targeted – central value. Control lines start as a curved but in time they straight out.

Toledo, Lizarelli and Santana Junior researched the use of control charts in chemical plants. Their research was conducted by introducing control charts to a chemical plant. They have recognized the necessity of using nineteen success factors which were considered most important for successful and sustained development and use of control charts. This feature needed to be in close surrounding of the operator and constantly displayed on the monitor, best case scenario would be if there could be one monitor only for control charts. Undertaken activities of this research have resulted in change of plant operator's behavior, working habits, as well as improved process which was researched and spread the idea of implementation of control charts to the other plants of the company. (Toledo, Lizarelli, Santana J, 2017)

Another extremely important domain of crucial importance for humans and for which control charts can and are used is healthcare. In domain where absence of statistical control and inability to monitor the data can have repercussions on risk management and liability, especially in critical in clinical situations, upholding constant level of patient care is essential. Many companies and health care organizations had to keep up to date with new technologies and transition from traditional to more advanced techniques of process control. Speaking about sample collecting frequency James C. Benneyan says: "*In many applications, instead of current practices of reporting data in large infrequent samples (such as quarterly), data should be collected in smaller samples much more frequently (such as weekly or monthly) and plotted on an appropriate control chart. Much of the aggregate data currently collected for various 'report cards' should be plotted and evaluated for process stability on control charts in more frequent and smaller samples."* Author, in his work, starts from a problem in which organization's Cesarean delivery rate is 15% and it never goes over that. He inspects if that 15% is continuing value of center line of a *p* control chart and if UCL is supposed to be lower than 15%. As one of the answers he proposes is that this process simply must be monitored and controlled, and that in that case center line is supposed to have same value as the standard rate, in this specific case 0.15. According to this usual rate is same as the standard but it requests operators to be educated in field of statistics. Paper points out that in case of lower rates, it is recommendable to gather statistical information and data of different variables instead of only monthly rates. This proposal comes from the fact that main reason of data collection is to understand and enhance the process, and not just evaluate and monitor singular effectiveness. Another interesting fact mentioned in his work is not only usage of Control Charts, but actually the necessity to know how to use those charts correctly. Inability to properly use and read control charts has resulted in more damage to the process than the benefit. Another problem that needs to be addressed is the errors that appear due to lack of data while making control limits, excessive application of individuals charts, wrong usage of formulas and empirical transformations, as well as use of traditional charts in case of merging data. Application of SPC calls for time and effort. It is shown that better results are achieved when focusing only couple of essential factors and after that spread the application of SPC according to the intelligence gained through the previous process. (James. C. Benneyan, 2008)

All these articles point out importance and very wide applicability of control charts for any industrial, chemical or even managerial process. Following two articles will refer to the topic of usage of control charts in interpreting and monitoring environmental data and a study of the effects of climate change on precipitation patterns since these topics are closely related to topic of this thesis.

Author Lloyd Morrison has written paper on use of Control charts to interpret environmental monitoring data in which he addresses the utilization of control charts in order to detect different types of specific patterns that could point out potential concerns with the resource in question. Paper points out simplicity of control charts usage as well as the disadvantages of control charts. Disadvantages represented in this paper mainly concern the difficulties of determining the base line of the chart. Determination of base line generally requires certain amount of experience and preciseness together with empirical data. Only limitation of control charts is the inability to display the cause of changes for monitored indicator, but this problem is solved by couple of approaches to monitoring trends do. This is done by performing couple of experiments in order to find out reason for the changes. (Morrison, 2009)

According to the formerly mentioned paper of Morrison, group of authors from Bosnia and Herzegovina, have gone one step further and did research study on using SPC methods on influence of climate changes on precipitation patterns. Their work included *i*-chart to compare the precipitation patterns from different time frames. By analyzing the charts this research proposes multiple solutions and improvements which could be suggested to the decision makers and managers. Large floods during the winter months could have been avoided if the control charts had been used and correctly analyzed. *"As the SPC method has rarely been used for the assessment of natural processes, further research will be undertaken to verify the applicability of the method for the assessment of the climate change effect."* Authors especially emphasize importance of gathering data not only for other locations but for multiple variables including sea level, water flow, temperatures, etc. (Vucijak, Kurtagic, Ceric, Kupusovic, Spago, 2013)

The application of control charts is numerous and really has a great impact on many spheres of life. Thus, for example, in addition to all the above, charts are also used in rock disintegration analysis, healthcare, and laboratory research. However, in order to get a clear picture and a large effect of control charts for a particular process, it is principal to determine the amount and frequency of data.

The following literature will review the application of control charts in health care, during the process of improving the quality of rocks disintegration, and finally highlight the impact of sample size on the performance of control charts.

Healthcare involves many people, both physicians and patients, and is therefore more likely to make mistakes or misunderstandings. For example, one patient will be examined by several doctors, which means that the process of his treatment is led by several people. Each of them records the observation they conducted, and the other one continues. If the instructions are not precise, clear and have a detailed explanation, it is very likely that mistakes or omissions will occur when the person's treatment is taken over by another doctor. Because of this, charts are highly efficient in healthcare. Although the application of the process monitoring is effective, especially people and doctors continuously monitor and improve the health of patients, unfortunately only the USA, Great Britain and Australia use this approach to a great extent. Not all types of control charts are used, but CuSum and EWMA proved to be functional. (Suman, Prajapati, 2018)

When it comes to using of control charts to rock disintegration by rotary drilling, the authors emphasize how variety of charts are used for various parts of this process. Thus, R charts are a measure of the variability of the observed technological drilling process, and for the input parameters of rotations pressure force, and output parameter vibration signal, the individual $\bar{x}$, s, $\bar{\bar{x}}$ control charts demonstrate that the technical drilling process is off-center and unstable.

For the chart analysis, in fact the process to be successful, application of corrective measures is needed, use of new measuring, remove the errors in measurings, discover systematic causes, and create most reliable experimental data. New parameters will improve rock drilling efficiency and stabilize and optimize the technological process. From the economic, environmental, and economic perspectives, there will be a substantial impact. (Flenger, Kacur, Durdan, Laciak, 2020)

In all the cited literature, control charts were used to conduct quality research, because chart is a powerful technology for monitoring processes and ensuring quality in statistical process control. Any control chart's overall performance is strongly influenced by the sample size n. The impact of n on the Shewhart control charts, which is used to keep an eye on the mean and variance of a data, is considered in work of Haridy, Maged, Kaytbay and Araby (2016). The authors conducted the study under various cases of false alarm rate and shift domain. They concluded that the best sample size of x-R and x-s control charts is $n=2$, for all cases they conducted. Also, by reducing average extra quadratic loss (AEQL), the chart will produce the least ATS – average time to signal over the full shift domain. This will result in higher capability of charts to detect large radius of changes. (Haridy, Maged, Kaytbay, Araby, 2017).

3 Control charts - Process Control tools

To keep their competitiveness on the market, during the last two decades, companies have realized that it is not enough only to invest money and ideas to the perfect product, but also that they must take one step further and improve, optimize and most importantly, to understand all processes inside their companies. To achieve these goals large number of companies have started using Control Charts as tool for process control.

3.1 Introduction to Control Charts

Control Charts are special Statistical Process Control (SPC) tools used to visualize the changes of the process over time. Data in the control charts are plotted in time order, therefore sampling frequency plays important role when using this tool.

First control chart was designed by Walter A. Shewhart during his carrier as Bell Labs employee in 1920's. Need for these charts has appeared when engineers who have worked for the company wanted to increase reliability of telephony transmission systems. Once these charts have been presented to company bosses Shewhart pointed out that it in order to manage the process and foresee the future output it is mandatory to have the entire process in state of statistical control. (Smith, 2009)

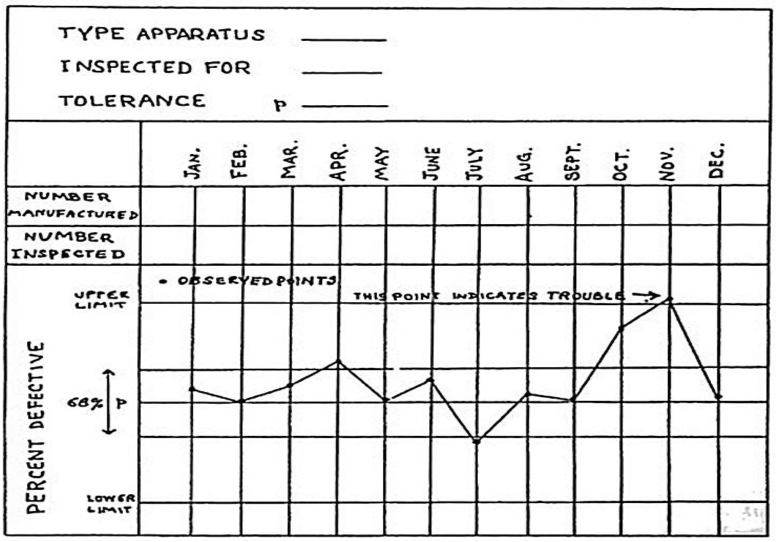


Figure 3. 1 First control chart made by W. A. Shewhart

(Source: Jones G., 2014)

Some of the basic possibilities that control charts have to offer are:

- Process quality management
- Product quality control during the manufacturing phase
- Process stabilization
- Process stability and punctuality analysis

Application of control charts are very important for discovering trends in process validation, complaint monitoring, technical support issues and product servicing. This graphical tool has proven to be highly useful and efficient technique to present and visualize data on the front lines and at the management reviews. Also use of control charts will enable enhanced efficiency in production while at the same time it can reduce flaws and malfunctioning production, improve earnings and reduce expenses.

It's critical to highlight that the control chart tracks process changes throughout time. As a result, the samples must always be the final ones generated. Variations in a process can be caused by two sorts of factors (Horvat, Edjed, Banaj, 2006):

1. General or systemic (common causes) that are inherent in the process (e.g., genotypic variations), and
2. Specific or special causes that create an excessive variation.

As a result, control charts are used to differentiate between these two types of process changes, based on an examination of the data we already have and the data we are estimating.

Control charts are widely and are most used for maintaining processes by discovering and fixing problems as they arise, estimating a process's desired results and analysing if a process is stable. There are factors that should be taken into account (in statistical control) when it comes to controlling if sample of process changes are because of special reasons (non-routine events) or common causes.

3.2 Design of the control charts

The design of the control chart is not very complicated. To start with the construction of forementioned charts, it requires taking a larger number of smaller samples, so that the connection between the basic and a certain set would be completely determined (Repic, 2015).

The following is guide to creating a control chart for desired process:

- First, choose the process to track.
- Make a sampling strategy for that process.
- Collect information (data) about process.
- Calculate the specific statistic values
- Calculate control limits.
- In the end, design control chart.

Designing a control chart begins with a midline or central line (CL) which is mean value of the quality aspect when the process is under control. After that, it is necessary to determine the upper control limit (UCL) and lower control limit (LCL) and draw them on the control map.

Process is in control if all process values are drawn inside control line values with no obvious tendency, Figure 3.2. However, if the process values are located outside the control line values or exhibit a specific tendency, the process is said to be out of control, Figure 3.3.

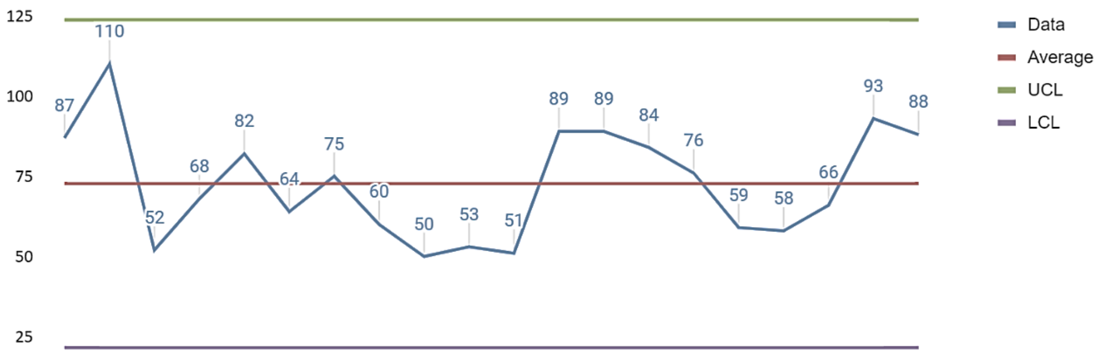


Figure 3. 2 Control Chart - In Control Process

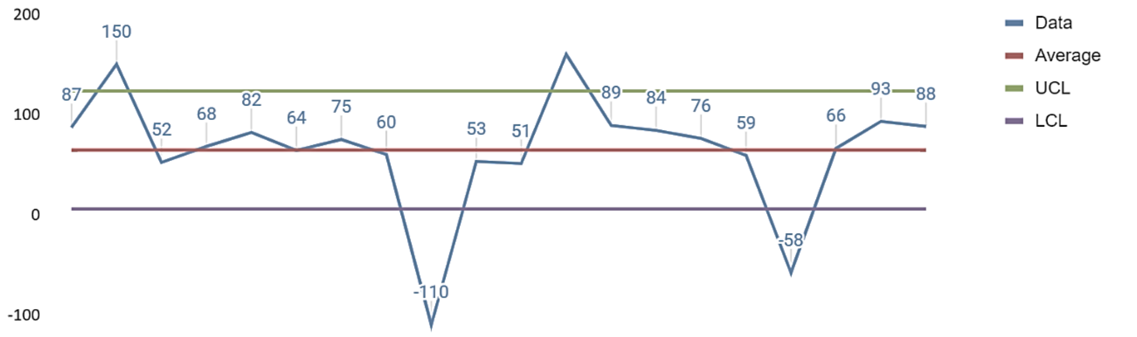


Figure 3. 3 Control Chart - Out of Control Process

The basic, simple control charts is depicted in Figures above. However, in addition to these elements, there are also upper warning limit (UWL) and lower warning limit (LWL). The upper warning limit and lower warning limit are placed between the center line and the two control limits (upper and lower), Figure 3.4. If the points on the graph are between control and warning limits, the process is still capable, but greater attention is needed because there is a chance that some values will go beyond the control limits, also known as action limits.

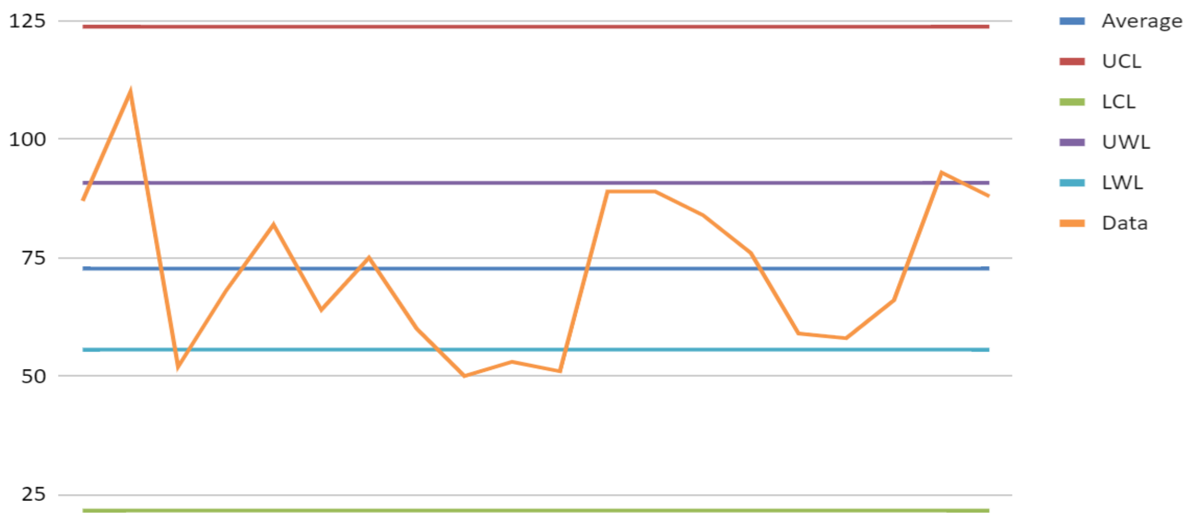


Figure 3. 4 Control chart with UWL and LWL limits

Limits are determined in one of three ways (Drenovac et al., 2013):

1. Based on the data gathered from the process that has to be controlled (and this is the most common case in practice)
2. Based on the quality feature being monitored tolerance
3. Understanding the technical process' capacities (6σ).

Control limits on control charts are usually calculated for some unknown process. Since the process is unknown, it is first necessary to study it in detail, record it and then calculate the control limits. Control limits are calculated at the scatter limits ($\pm 3\sigma$) of the statistical parameter ($\overline{x}, R, s$ etc.) calculated from the sample. (Drenovac et al., 2013)

When it comes to calculating control limits from a given tolerance, it is necessary to first plot them in a map and then monitor the measured values whether they are within the limits or not.

It is important to note that the control limits obtained by calculation (statistically), can't be linked, in fact they are not the same as the limits of requirements defined by the customer.

Although the graphs are completed after the basic components that each chart must have are drawn, it is necessary for clarity and easier interpretation, at the end, to add the name of control chart, date of monitoring the process, the person who performed the control (controller), time and shift in which the controller performed measurements, the number and basis of the machine on which the process was performed, and the number and basis of the machine on which the process was performed.

3.3 Types of control charts

Generally, all control charts can be divided into two groups, depending on whether the data being watched is "variable" or "attribute" which is showed in Figure 3.4. Different variations of statistical control charts are:

1. Control charts for variable data
2. Control charts for attribute data

Control charts for variable data are used for measurable data like money, length, weight, width, etc., and these charts often contains decimal, while control charts for attribute data are used for unmeasurable data, also known as counted data and they are always a whole number.

The type of control chart used, will depend on the type of data being worked with. Furthermore, these two types of charts can be divided into other control charts which will be mentioned below.

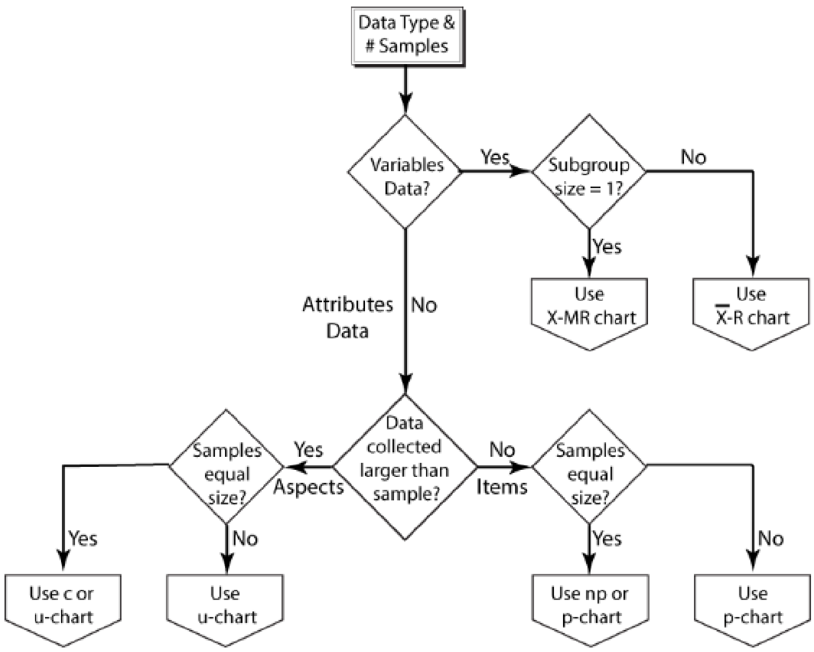


Figure 3. 5 Choice of control chart

(Source: Schenkelberg F., 2017)

3.3.1 Control charts for variable data

The mathematical basis of these control charts is the normal distribution. Since in the case of this final thesis, control charts for measurable (variable) characteristics are required, it is important to note that there are several types of them, of which the most used are:

- control charts for average values and ranges - $\bar{x}$ and R control charts
- control charts for average values and standard deviations - $\bar{x}$ and σ control charts
- control charts for individual quality monitoring

There will be more talk about these control charts below.

3.3.2 Control charts for average values and ranges – $\bar{x}$ and R control charts

Control of average value of the process, or the standard level of the quality, is usually done by applying control charts for the average values, or $\bar{x}$ control charts. (Kadric, Bajric, Vucijak, 2018).

The sample for these control charts is small, which means *n<12*, and in practice number of samples is usually *n = 5, n = 10* or between 5 and 10. To be able to subject this control chart to statistical analysis, the number of observations should be at least 20 (*k≥20*). In each observation, the controller records the individual values of x for the observed quality characteristic on the sample *n* when it examines. From these values, the average $\bar{x}$ and range R are calculated for each observation.

$$\bar{x}_i = \frac{x_1 + x_2 + x_3 \ldots + x_n}{n} \tag{3.1}$$

$$R_i = x_{i,max} - x_{i,min} \tag{3.2}$$

Based on the data thus obtained, the calculation for control chart is performed, according to the following procedure:

$$\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \cdots + \bar{x}_n}{n} \tag{3.3}$$

where:

$\bar{\bar{x}}$ – average value of all average values of samples

$$\bar{R} = \frac{R_1 + R_2 + \cdots + R_n}{n} \tag{3.4}$$

where:

$\bar{R}$ - represents the average of all ranges

The central line of the $\bar{x}$ control chart is the calculated $\bar{\bar{x}}$ value that represents the average of all average values of the samples. This value is taken because it is the best estimate of the average value of a process.

The process control limits show the qualitative movement of the process, its stability and ability. These are the limits of the possibilities of one process and represent the normal variations that can be expected. Other elements of the control chart (upper and lower action and warning lines), for $\bar{x}$ control chart is determined as follows:

$$\begin{gathered}
\textit{Upper control limit UCL} = \bar{\bar{x}} + A_2\bar{R} \\
\textit{Upper warning limit UWL} = \bar{\bar{x}} + (2/3)A_2\bar{R} \\
\textit{Central line CL} = \bar{\bar{x}} \\
\textit{Lower warning limit LWL} = \bar{\bar{x}} - (2/3)A_2\bar{R} \\
\textit{Lower control limit LCL} = \bar{\bar{x}} - A_2\bar{R}
\end{gathered} \tag{3.5}$$

where:

A_2 – the factor for calculating control limits that depend only on the sample size (n) and their values can be read in the table given in the attachment I

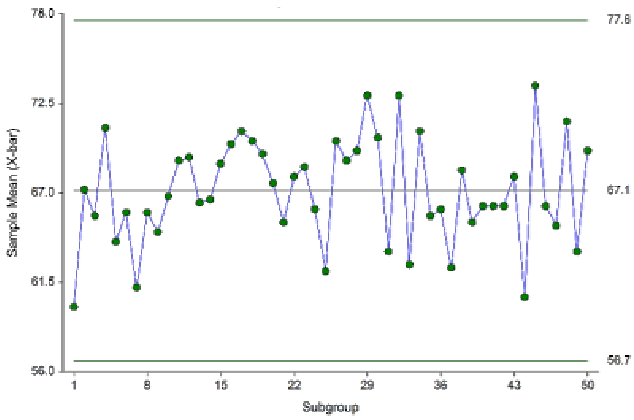


Figure 3. 6 Sample average $\bar{x}$ control chart

(Source: NCSS Statistical Software)

Process variability control, or quality level variability, is usually performed either by control charts for standard deviations, or by control charts for ranges also called R control charts (Kadric, Bajric, Vucijak, 2018).

The central line, warning limits and control line values for the *R* type control charts are

$$Upper\ control\ limit\ UCL = D'_{0,001}\bar{R}$$

$$Upper\ warning\ limit\ UWL = D'_{0,025}\bar{R}$$

$$Central\ line\ CL = \bar{R} \qquad (3.6)$$

$$Lower\ warning\ limit\ LWL = D'_{0,975}\bar{R}$$

$$Lower\ control\ limit\ LCL = D'_{0,999}\bar{R}$$

calculated using the following formulas:

where:

$D'_{0,001}$, $D'_{0,025}$, $D'_{0,975}$, $D'_{0,999}$ - constants for different sample sizes *n*, which can be read from the table given in the attachment II

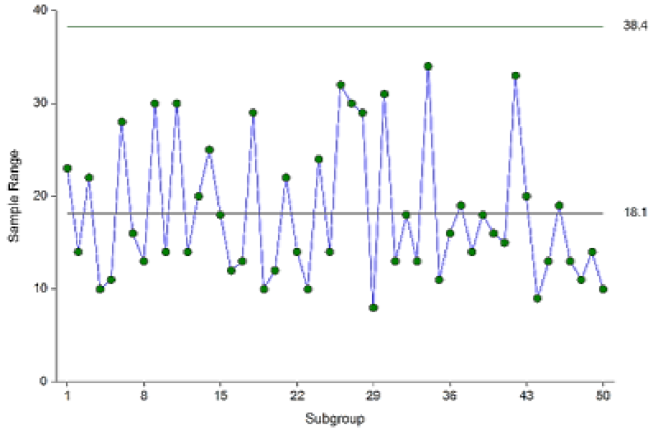


Figure 3. 7 Sample range ***R*** control chart

(Source: NCSS Statistical Software)

In case of using these charts, R charts should be created beforehand. (Bajric, Kadric, Vučijak, 2018).

3.3.3 Control charts for individual measuring and moving range charts - *i* and MR charts

I chart are simplest form of control chart for individual values. Central line CL can be set at middle of specification range, as well as at specified value or at average value of previous period of monitoring. For these charts, as well as for previous upper and lower control lines are set by using three standard deviations from control line as shown in 3.7. In case that some points are over upper control lines or under lower control lines, process is considered to be out of control, same conclusion is for the case if two out of three consecutive points are over the upper warning line or under the lower warning line. Process is also considered to be out of control if 8 consecutive points are from the same side of control line or, if 4 out of 5 consecutive points are in area over or under of one standard deviation line.

$$\text{Upper control limit } UCL = \bar{\bar{x}} + 3\sigma$$
$$\text{Upper warning limit } UWL = \bar{\bar{x}} + 2\sigma \qquad (3.7)$$

$$\text{Central line } CL = \bar{\bar{x}}$$

$$\text{Lower warning limit } LWL = \bar{\bar{x}} - 2\sigma$$

$$\text{Lower control limit } LCL = \bar{\bar{x}} - 3\sigma$$

Where $\bar{\bar{x}}$ represents the average value of the dataset and σ represents standard deviation.

These charts are considered as simple control charts, simple to create and they are used to recognize changes in process of centering or process accuracy. Downside of these charts is that it is harder to recognize changes in terms of process variability and precision.

Moving range – MR charts is one more type of charts for processes with small amount of data. These charts are used to avoid possibility of making decision regarding the process based on only one or couple of last collected data. Within these charts, two, three or even more of consecutive measuring are considered as one sample, and for sampling as such, control line is defined. In every following sample, new value replaces the oldest one in the measuring, and by doing so process repeats. First sample is created the moment when enough *n* values is collected, where *n* represents size of the sample. Sampling size of *n* should not be too large, because in that case, there is possibility of trend changes in early beginning of process monitoring. In this way individual values, which can have very high differences in values, are replaced by average values of multiple adjacent observations. Control lines for these charts are defined same as for control charts for average values.

For MR charts, it is considered that process is not in control if one point is over the upper control line or under the lower control line, if (n-1) consecutive point is between warning and control line, as well as if 2.5*n* consecutive points are on the same side of control line. (Bajric, Kadric, Vučijak, 2018).

3.3.4 Control charts for attribute data

Attributive quality features are those that are rated descriptively, such as whether something is good or bad, whether it fits or not, whether it goes or doesn't go, and so on. Attributive evaluations such as visual quality control are common. Even with attributive quality features, however, evaluation is possible with the use of appropriate measuring, control, and testing equipment. This group of control charts includes:

1. *np* control chart
2. *p* control chart
3. *c* control chart
4. *u* the control chart

The mathematical basis of these charts is the Binomial and Poisson distribution.

np control chart - directly shows the number of found scrap units in the sample. This chart is applicable when the number of measuring is constant.

p control chart - shows a graphical movement of the proportion of bad pieces in the samples. The sample sizes can be different. The mathematical basis is a binomial distribution.

c control chart - monitors the number of errors on one product or sample. The samples must be the same size.

u control chart - shows the average number of errors expressed per unit of products found in the delivery (sample). The sample sizes can be different.

3.3.5 Control charts for cumulative sum

Control charts that improve the ability to find and detect small patterns / shifts are called control charts for cumulative sum - CuSum. They are a graphical representation of statistics that includes current and previous values of data from the process. Cumulative sum charts is used to control the average of a process based on measurements taken during the process at given time intervals. (Zohuri & McDaniel, 2021).

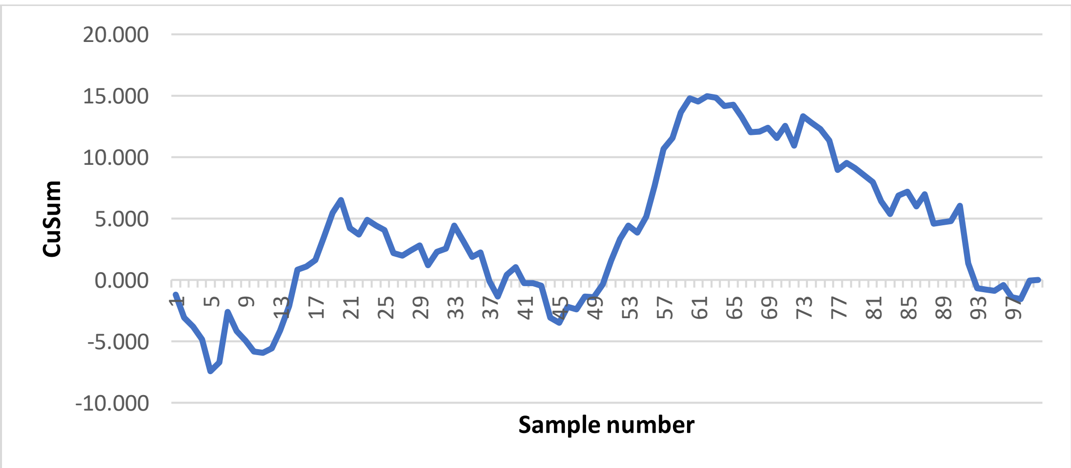


Figure 3. 8 CuSum chart

A subgroup is formed by the samples' measurements at a specific period. The CuSum chart displays the accumulation of information from present and previous samples, rather than evaluating the mean of each subgroup individually. As a result, the CuSum chart outperforms the X-bar chart when it comes to identifying tiny changes in a process's mean. The CuSum chart is based on a targetet value and a reliable standard deviation estimation.

$$\boldsymbol{C_i = \sum_{j=1}^{i} (x_j - \mu_0)} \tag{3.8}$$

where:

$\boldsymbol{x_j}$ – the average value of j-th sample

$\boldsymbol{\mu_0}$ – target value

Due to the appearance of the CuSum chart, it is very important to choose the appropriate target value. Regarding this, we can identify two cases of graphics appearance:

- The graph is decreasing - if most of the values of the samples x_j are less than the target value μ_0
- The graph is ascending - most of the sample values x_j are greater than the target value μ_0

In addition to these two cases, two other basic rules of interpretation of CuSum charts can be identified (Bajric, Kadric, Vucijak, 2018):

- The CuSum graph is horizontal if the observed values x_j are equal to the target μ_0 value
- Actual CuSum values C_i have no relative significance

3.4 Control chart analysis

When using control charts, standard deviation is used as a measure of process spoilage. The upper control limit is at a distance of $+3\sigma$, and the lower at a distance of -3σ from the central line. The space between the control boundaries can be further divided into three zones. These zones, called A, B and C, start from the central line and are used in the analysis and interpretation of control charts.

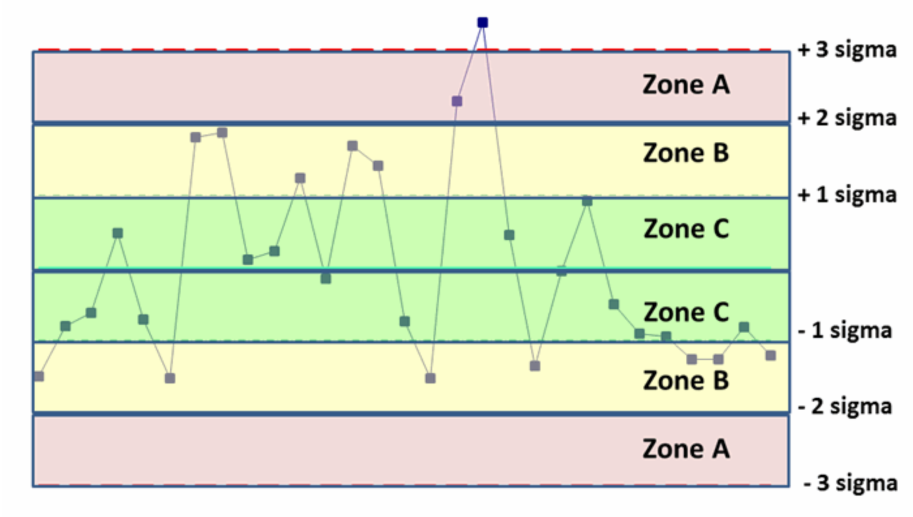


Figure 3. 9 Division of the control chart into zones

(Source: qimacros.com)

A stable process is a process that produces results a process of which all, or a satisfactory number, are within control limits. There is no completely ideal process, but we should strive for it. The International Standards Community has adopted rules for deciding on the stability of the process (Jašarević, Brdarić, Isaković, 2020). These rules are:

- Each of the last 25 dots must be within the control limits.
- Among the last 35 dots, one may go beyond the control limits.
- Of the last 100 dots, a maximum of two points may go over the control line values.

Table 3. 1 Control Chart Rules

(Source: Dr. Bill McNeese, 2016)

RULE	RULE NAME	PATTERN
1	Beyond Limits	One or more points beyond the control limits
2	Zone A	2 out of 3 consecutive points in Zone A or beyond
3	Zone B	4 out of 5 consecutive points in Zone B or beyond
4	Zone C	7 or more consecutive points on one side of the average (in Zone C or beyond)
5	Trend	7 consecutive points trending up or trending down
6	Mixture	8 consecutive points with no points in Zone C
7	Stratification	15 consecutive points in Zone C
8	Over-control	14 consecutive points alternating up and down

If the values in the chart are concentrated around the central line mostly, and none of those values is higher than values of control lines, it means that this operation is in control. However, most often, not all points are located near the central line, they vary a lot. Therefore, Dr. Bill McNeese in his publication stated 8 rules for control charts have been introduced, Table 3.1. These rules help in the analysis of the process and give indications that there are special causes that caused the variations.

Rules 1 and 2 apply to cases that happen suddenly, and only once. They are shown on the chart as sudden jumps relative to the average. Rules 3 and 4 also represent shifts, changes but much less some in rules 1 and 2, Figure 3.10. Also, here shifts are maintained over a certain period.

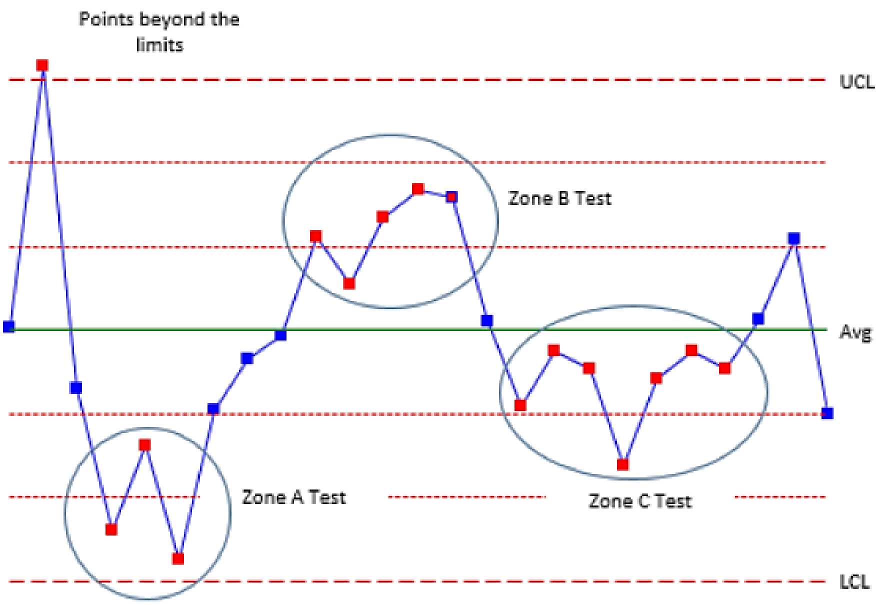


Figure 3. 10 Zone Tests (Rules 1 to 4)

(Source: Dr.Bill McNeese, 2016)

The Figure 3.11 shows the rule 5 which is a process that moves in only one direction, either up or down, and rule 6 is a mixture of different actions, which we cause ourselves.

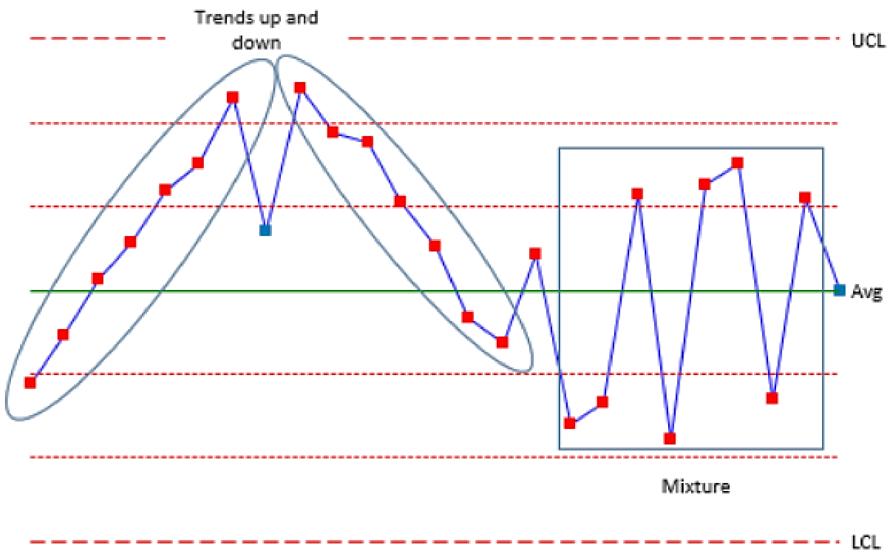


Figure 3. 11 Rules 5 and 6

(Source: Dr.Bill McNeese, 2016)

Rule 7 is similar to rule 6, except that the processes are now grouped into a subgroup. Rule 8 represents an over-adjustment of the process by the controller. It usually occurs when you want to set a certain desired value. This leads to an increase in process variations, Figure 3.12.

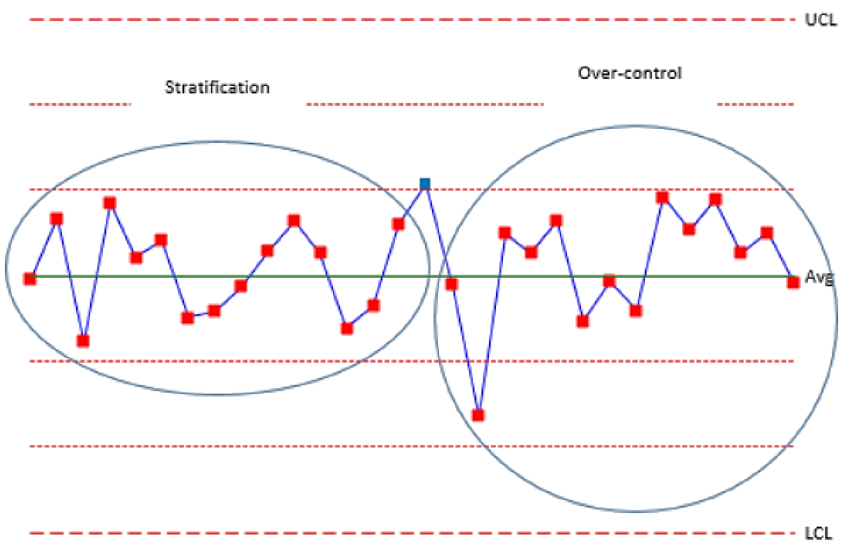


Figure 3. 12 Rules 7 and 8

(Source: Dr.Bill McNeese, 2016)

It is important to mention that not all regulations can always be applied to all types of control charts.

In addition to the precise production of control charts, their correct interpretation is also of great importance. The simplest way is to interpret control charts in which the process is "outside statistical control", the process in which the values of individual measurements are outside the control line values. This means that some of the specific causes of variation are present in the process and an adjustment needs to be made. However, the placement of all points within the control boundaries, in fact obtaining a control chart on which the process is "within statistical control", does not necessarily mean that such a process is acceptable by statisticians. When analysing control charts, a set of expressions is used to represent a specific state on the chart, which displays a process that is “under control” and refers to and explains the above rules (Juran, 1999):

RUN (flow or tendency) - occurs when seven consecutive points are located on the upper or lower side of the central line, but within control limits. This appearance of the control chart tells us that there are irregularities in the process that need to be corrected.

TREND (trend) - if there is a consecutive samples on the control chart whose values are continuously decreasing or increasing, it indicates that the process is not in control, and it is necessary to adjust the machine.

PERIODICITY - is evident when changes of the cyclic type occur within the process at the same intervals.

HUGGING - an irregularity that occurs when the measured values are located very close to the central line or control limits.

The analysis of each individual control chart should be approached very seriously and studiously. Based on the results of the control chart, it is possible to improve the production process, eliminate unwanted causes of variation, reduce production costs, and thus increase profits. The control chart, as a tool combined with the knowledge of those leading the process, replaces intuitive decision-making about the process, by making decisions on a scientific basis.

3.5 Sampling frequency role within control charts

One of the most common mistakes made when creating SPC control charts is collecting data too frequently. Too much data collection can be problematic for a variety of reasons, not only for increased sampling or measurement costs. One major concern is that it establishes overly sensitive process behavior limitations, resulting in false alerts, Figure 3.13. When we have this many false alarms, we're telling the operator that the system doesn't operate, and those alarms aren't important. Operators must have faith in the process as well as the control charts.

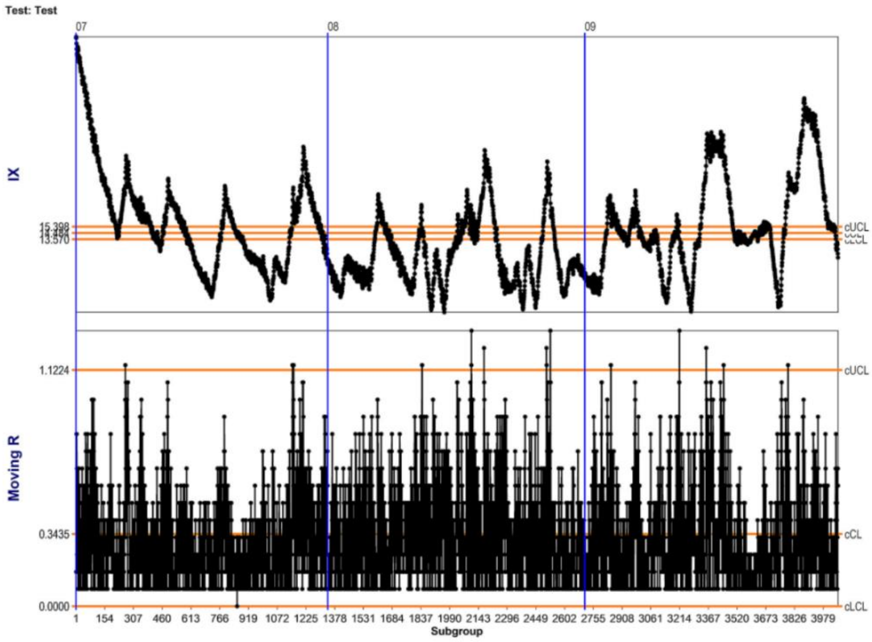


Figure 3. 13 100% data display – 1 subgroup/minute

(Source: QualityMag)

To get an accurate and good Figure of a process, it is necessary to take samples often enough to capture any expected or unexpected changes, but they must also be far enough apart from each other to be able to detect variation, Figure 3.14.

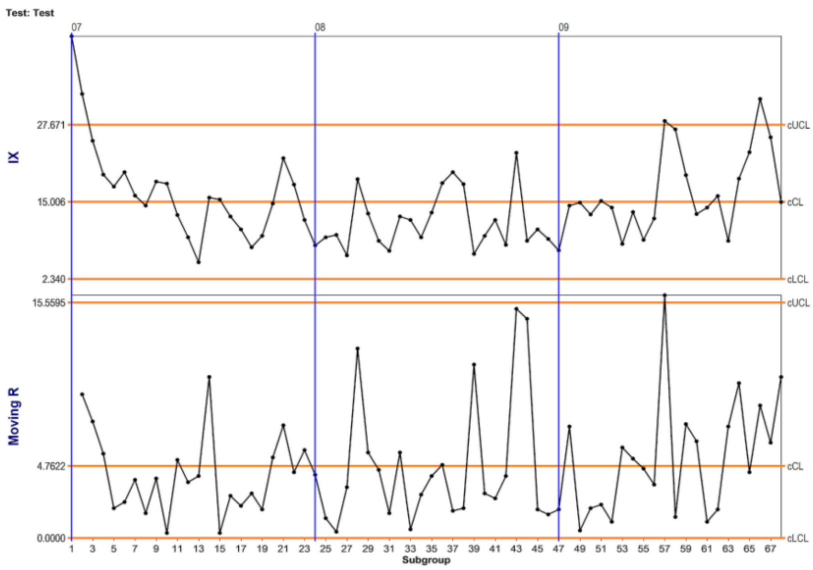


Figure 3. 14 Sampled data – 1 subgroup/hour

(Source: QualityMag)

It is crucial to understand the process in order to set rational sampling frequencies. Every process exhibits normal variation and understanding this behaviour allows us to sample the process more precisely.

To obtain an effective sampling frequency, the following needs to be done:

- Conduct a process analysis to learn about the process's typical patterns.
- Collect data as much as necessary during the process research to ensure that the process's common behaviour is understood.
- Determine the amount of time or number of items produced between process shifts by analysing the process research data trend.
- Set an SPC sampling frequency to gather two subgroups between process shifts – a three-hour process, for example, should be sampled every hour.

4 Hypothesis and expected results

Aim of master thesis is to conduct a comparative analysis of process control charts to define the most efficient sampling frequency on the gathered data. Control charts as one of the tools most often used to monitor the selected process depends on sampling frequency. The thesis analyses the applicability of different types of control charts in water level monitoring process depending on variability of sampling frequency. During the analysis process types used are:

- Comparative analysis and applicability of I control charts with different sampling frequency
- Comparative analysis of R control charts with different sampling frequency
- Comparative analysis of CuSum control charts with different sampling frequency

Dataset analysed within this master thesis is water level measured at measuring point Radobolja, near city of Mostar. Analysis focuses on defining if it is necessary to do measuring every half an hour, as it is done currently, or measurements can be conducted on more seldom period, and if that is case, weather the results differ, and if they do, how much they differ.

Basic hypothesis of this thesis is to analyse if control charts can be used, and if they can, what is most efficient sampling frequency in field of water level monitoring process. Expected results are improvement of water level monitoring process as well as design of draft sample which can be improved and later applied to different measuring stations to be used as flood predictive tools.

5 Control charts – comparative analysis to find best sampling frequency

Prior to comparative analysis of control charts from point of sampling frequency for water level dataset it is important to explain and be introduced with the process of water level measuring. Water level measuring is called hydrometry and it provides basic data and information's used in process of making engineering decisions. Hydrometry, among many other things, includes choice of place and position for measuring, setup, and equipment of hydrological stations as well as testing of new tools and development of new methods and measuring technologies. During the water level measuring process curves and constrictions should be avoided.

a)

b)

Figure 5. 1 River Radobolja a) riverbed, b) geographical location

(Source: Akta.ba)

Radobolja river is a natural river in southern part of Bosnia and Herzegovina which flows through the city of Mostar and pours into river Neretva. This river was previously used for irrigation of surrounding areas, however due to high level of pollution this is not case anymore. It flows through the area with high density population; therefore, every flood or even smaller overflows cause high material damage as well as human casualties, therefore it is of utmost priority for water levels of this river to be monitored. Previous Figure 5.1 displays picture of river Radobolja which is natural river with natural riverbed which makes the river more dangerous and unpredictable since its riverbed is full of curves and constrictions, this type of riverbed increases probability of flood occurrence.

5.1 Tool for measurement and control of water level

Through the history people have been aware of importance of water level measuring, and in that process, they have used different tools for measurement and control of water level. Some of the most used tools for measurement of water level are water level gauge, limnigraph, measuring needle, capacitive probes, ultrasound level measuring devices.

The data in this dataset is gathered by using automatic water level registrar which is called Limnigraph.

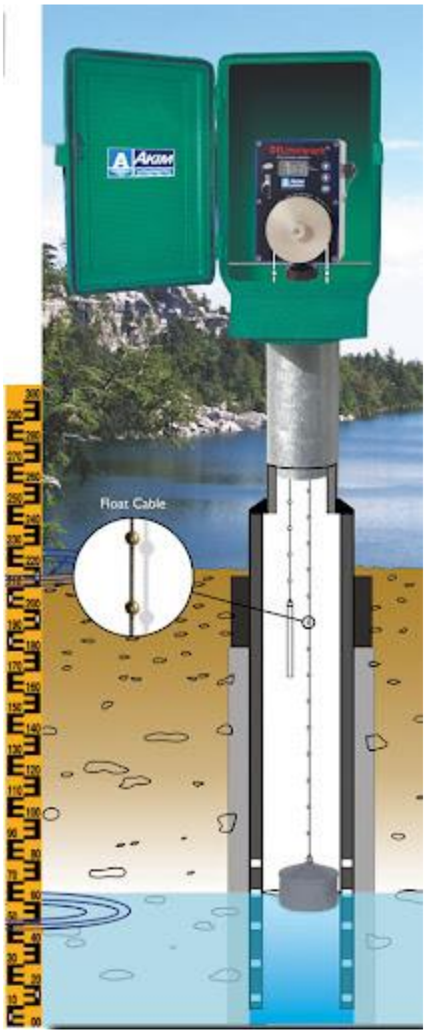


Figure 5. 2 Limnigraph

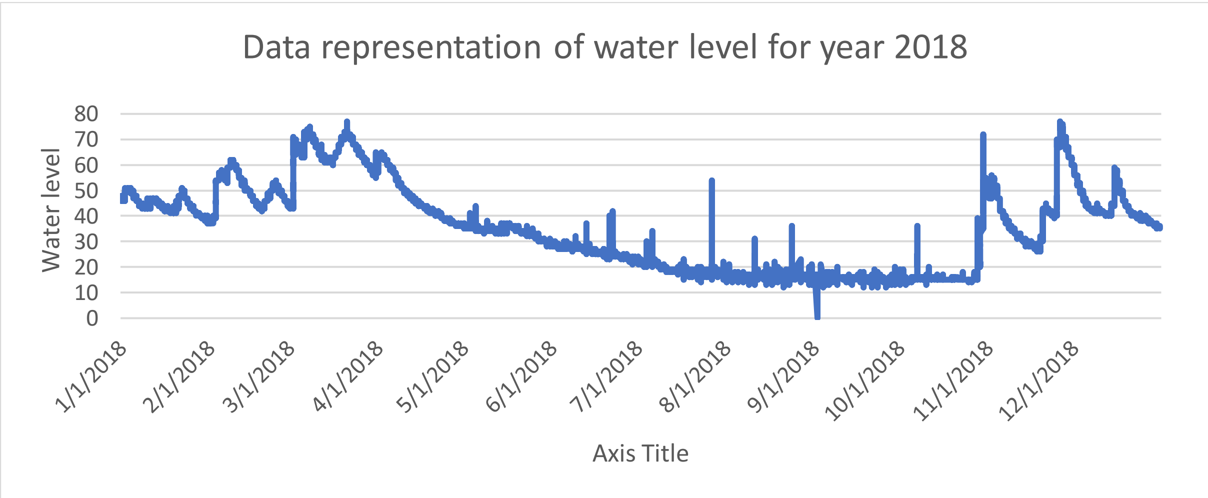
(Source: www.limnigraph.com)

Limnigraph as shown in the Figure 5.2 is used for hydrological measurements, but it can also be used for laboratory purposes. These devices are used for continuous measurement and recording of water level. Most used limnigraph design is classic type. This design contains float, counterweight, and binding system of float with feather and register with built in clock mechanism. Accuracy of this classic type depends on stability of float and amount of friction in bases of binding system.

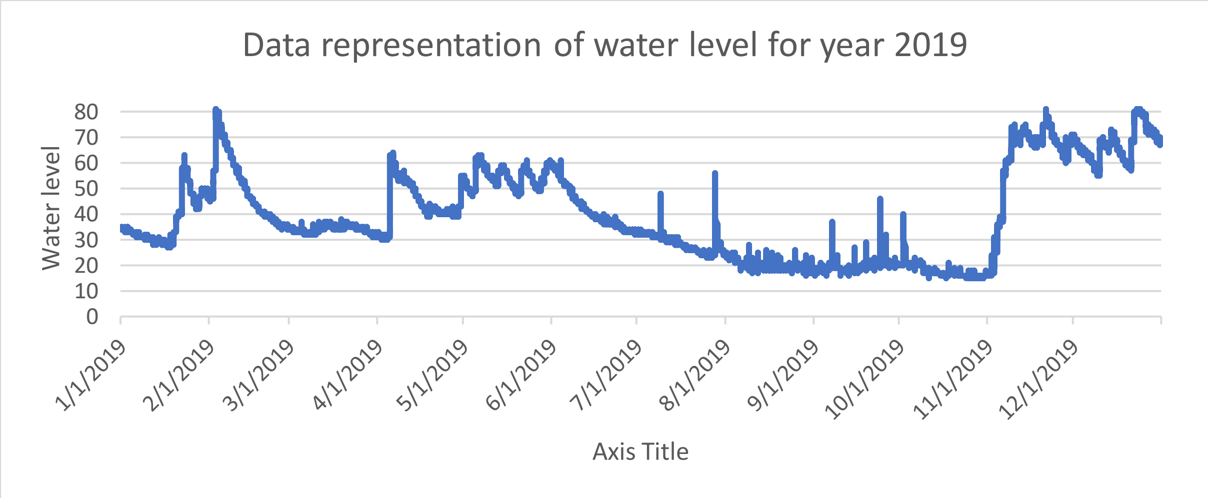
Limnigraph at measuring point Radobolja is set to collect data regarding the water level on every 30 minutes, that means that this device makes 48 measurements and delivers 48 samples of water level per day, in total of 17520 measurements per year.

5.2 Comparative analysis of control charts from point of sampling frequency

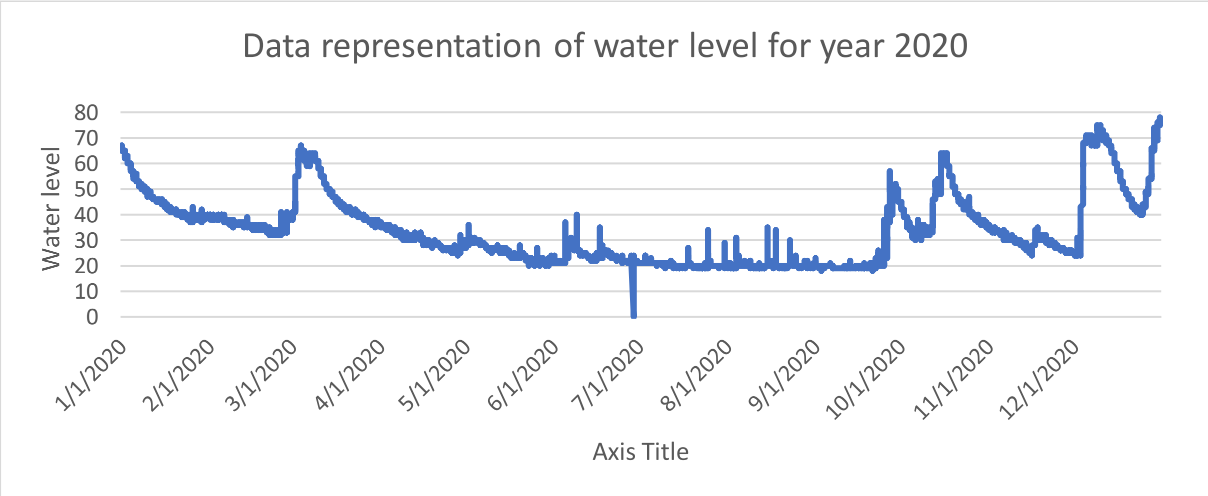
Dataset used in this thesis contains exactly 52608 data which are collected during the period of three years. Years 2019 and 2018 contain 17520 data which are measured during the period of 365 days every day, and year 2020 which contains 17568 data due to being leap year and having 366 days. Sensor measures water level 48 times per day, which means that measuring is performed every half an hour. All data collected by device are displayed as rounded numbers without decimal points. Prior to analysis all data have been separately sorted in different measuring intervals, frequencies, to define the best sampling frequency.



a) Data for 2018



b) Data for 2019



c) Data for 2020

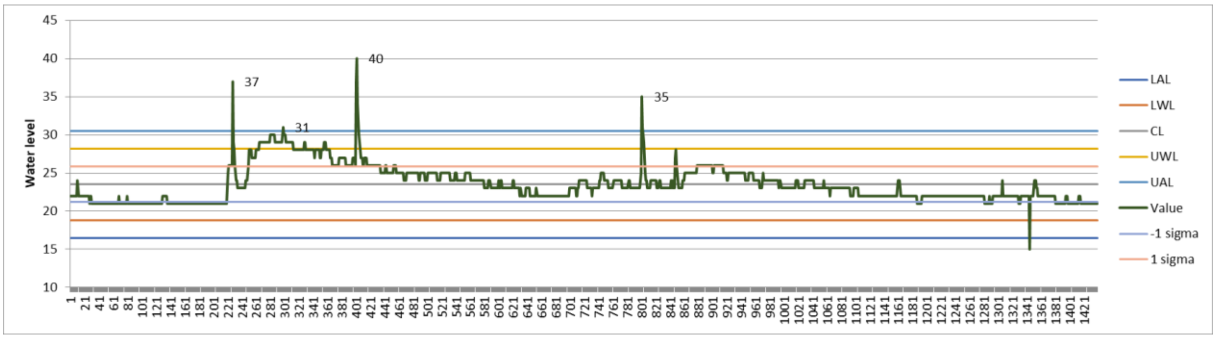
Figure 5. 3 Graphical representation of water levels in previous 3 years

To gain better and more clear insight into dataset and overall overview of river behaviour, Figure 5.3 represents data of every year. This graphical representation is good to be used as display of changes in water level throughout the year and to determine seasonality, as well as to compare water changes during the three years period. By comparing water levels of every year, it is noticeable that during the observed period every next year water levels were lower, and that period of low water levels was longer. According to Figure 5.3 plot c it is safe to determine that for year 2020, the dataset that will be further analysed, there is a period of year when water levels have similar values and are not subject to large water level variations. This period starts from 6th and lasts until the 10th month, after which water levels start to change and grow vigorously. Further in this thesis, two separate months of different periods will be analysed. First month which will be analysed is 6th month which belongs to summer period of the year. This month had small water level variations and somewhat constant water levels. The other month which will be analysed is 12th month. This month belongs to the winter period of year. Winter period, as shown in Figure 5.3 plot c is subject to large water level variations and changes caused by increased amount of precipitation, snow melting or soaked land, most and largest water level changes happen in December, and that is why this month is chosen for analysis

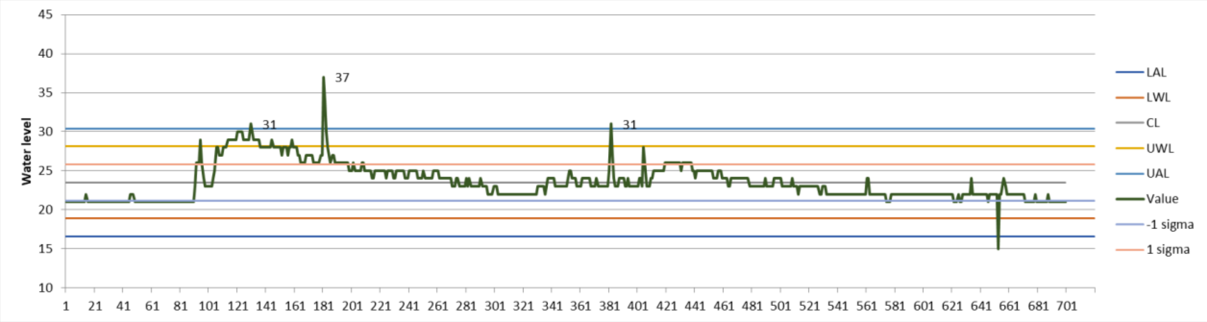
First analysis conducted is analysis of I Charts focusing on different sampling frequencies to monitor water level variations in period of one month. The data used for this chart are data of 6th month where water levels are pretty much constant. Data of 6th month of 2020 with three different variations is used for these charts. First chart represents 1440 water level samples taken every 30 minutes. Second chart represent 720 water level samples which are taken every 60 minutes in period of one month while third chart represents 480 water level samples taken every 90 minutes in same period of time (one month). Furthermore, due to the fact that measuring device is capable to display measurements as rounded numbers, values displayed on control charts are rounded to full numbers, since putting control line values with decimals is useless when all measurements are displayed as rounded numbers. Even though sampling frequency is changed, and datasets used are different regarding the number of samples, when rounded to the full numbers all three plots contain same values of control lines as shown in Table 5.1.

Table 5. 1 Control line values for I charts

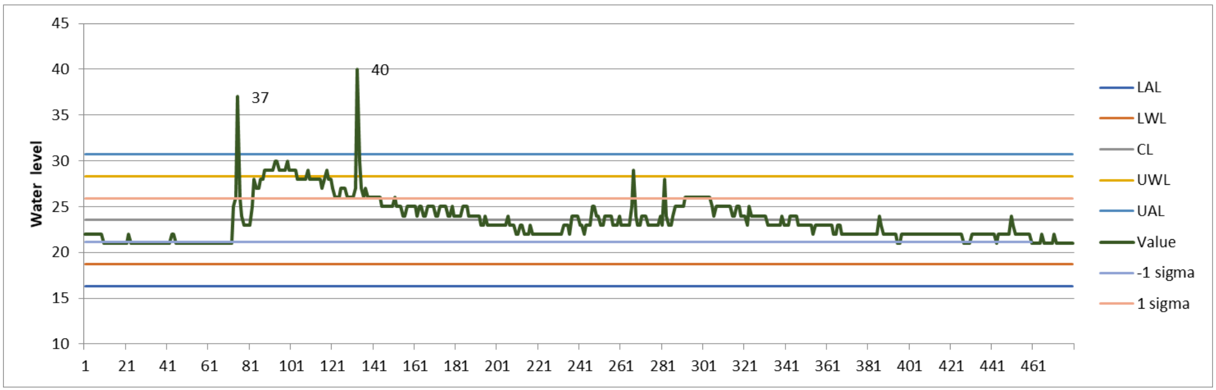
LAL	LWL	-1 sigma	CL	1 sigma	UWL	UAL
16	19	21	24	26	28	31



a) 30 minutes sampling frequency



b) 60 minutes sampling frequency



c) 90 minutes sampling frequency

Figure 5. 4 I Charts of water levels and different sampling frequencies for summer period

Figure 5.4 shows that changing the sampling frequency reflects on the outcome of the control chart. Comparing these three charts it can be noticed that plots in the charts might seem similar, but they are not. The data presented in Figure 5.4 plot a contains eight samples that exceed the upper action limit out of which four are shown (values of 37, 31, 40, 35), while in the plot b four samples go over upper action line and three of those samples are presented (31, 37, 31). By reducing sampling frequency even further the plot created using samples gathered every 90 minutes displays only two points (37, 40) over the upper action limit. Even though these differences are expected, since lesser number of samples is analysed, reducing the sampling frequency leads to losing information regarding the changes in water levels. For example, in Figure 5.4 plot a and c show that on 05.06.2020. values measured are crossing over upper action line (water level 37 cm), while plot b for that same period shows that those values are crossing only upper warning line (value 29 cm). Same thing happens for 17.06.2020 where plot a and b show that values are over the upper action line (value 35 and 31), while plot b for the same day shows that measured values barely cross upper warning line. By changing the sampling frequency, the number of samples changes as well, this causes for extremes to be even larger, if they are included, accounted and analysed by that sampling frequency, but it can also fail to provide information regarding the water behaviour, if those data are missed and not accounted due to changed sampling frequency. For example, in figure 5.4 both plot a and b on 29.06.2020. show that there is a sudden decrease of water level and measured level falls below lower action line (value 15cm), while plot c for same period does not show this measuring at all. However, since the period analysed by these charts is summer period which has low water levels and there is minimum risk of floods and all water levels are far below flood limit, it is interesting to analyse these plots from point of view focusing on draught. This river is sometimes used for irrigation of surrounding areas, and low water levels can be damaging as well. From that point of view, plots a and b provide clear, sufficient and in time information, while plot created using 90 minutes sampling frequency completely disregards one value (15) that falls under to lower action line. This comparative analysis leads to the conclusion that 30 minutes sampling frequency provides too many data which might confuse the system and the operator as well as might provide false alerts for them, while 90 minutes sampling frequency disregards some data of the water level changes and leads to unreliable monitoring process which might cause late reaction. According to this, sampling frequency of 60 minutes provides enough data to keep this process of water level monitoring in control, providing all important information in case reaction is required as well as keeps plot simple and easy to read without too many data.

Furthermore, I Charts have been created for winter period following the same principle. This period is subject to large water level variations and higher water levels which results to different control line values and high risk of floods. Dataset used for I Chart in Figure 5.5 contains 1488 samples of water levels measured every 30 minutes. For this dataset average water level value is 56 cm while maximum value of water level is 78 cm. For this chart control lines are calculated by using average (center line) value of 56 cm, and standard deviation of 14.5. When this standard deviation multiplied 2 or 3 times and added to the center line value, control lines with upper warning limit of 85 and upper action limit of 100 cm are calculated. In this case, upper action line could propose a problem since, according to the existing alert system, values over 90 cm are marked as high-risk values. Therefore, control line values on this chart might not be relative and upper action lines should be slightly modified to the maximum value of the riverbed.

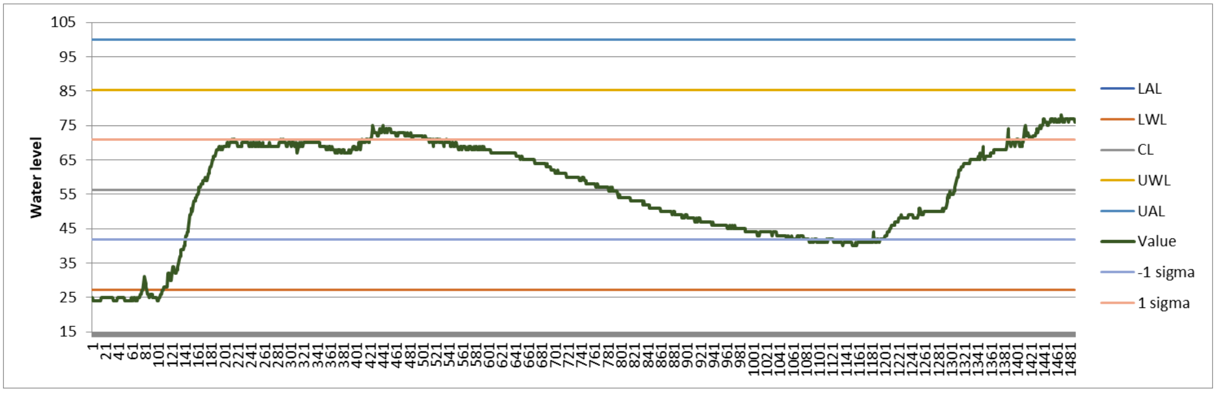


Figure 5. 5 I Chart for Winter period with 30 minutes sampling frequency

Interesting moment noticed in Figure 5.5 is period of first five days of December where water levels change extremely, rising from 25cm to 71cm in only few days. This period is analysed to show how control line values drastically change with such large water level increasements in such short period of time. After this period, water level values are increased but have similar values and values do not change that much in such short period of time. Chart in Figure 5.6 contains 120 samples gathered every 60 minutes with average value of 41.72cm.

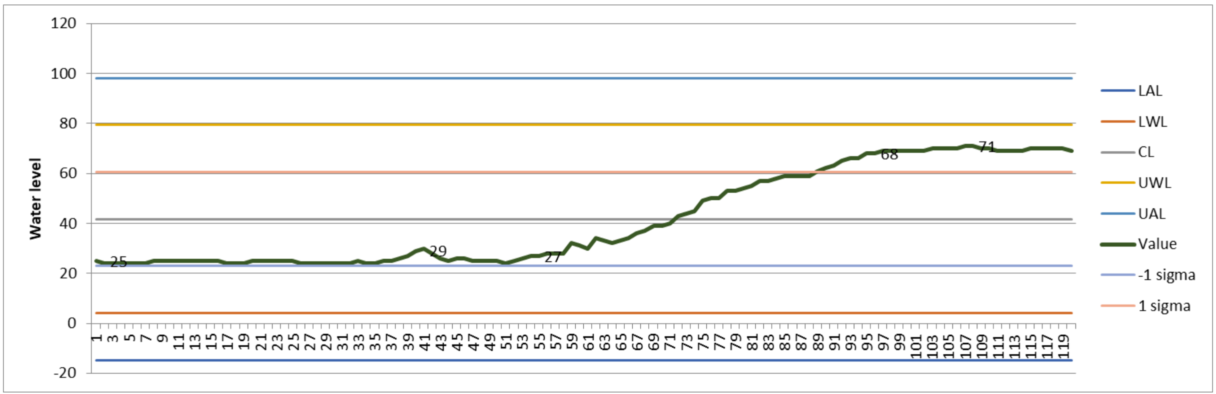


Figure 5. 6 I Chart for 5 days period with 30 minutes sampling frequency

Even though number of samples is reduced, difference between average values is more than 10cm, still upper control lines, compared to the one in Figure 5.5, have nearly same values. Upper action line for Figure 5.5 being 100 cm, and for Figure 5.6 being 98 cm. The lower average value, large water level values, and larger standard deviation value leads to this result.

Table 5. 2 Comparative overview for I Charts for different sampling frequencies

I Charts for 06.2020						
Sampling frequency	Number of samples	Number of points over UAL	Number of points over UWL	Number of points under LWL	Number of points under UAL	Maximum and Minimum value
30 minutes	1440	8	55	0	1	40 / 15
60 minutes	720	4	25	0	1	37 / 15
90 minutes	480	2	20	0	0	40 / 21
I Charts for 12.2020						
30 minutes	1488	0	0	84	0	78 / 24
60 minutes	744	0	0	32	0	77 / 24
90 minutes	498	0	0	14	0	77 / 24

Table 5.2 presents quick and easy overview of comparative analysis conducted previously and it leads us to the conclusion that reducing sampling frequency can have large effects in process of water level monitoring. By reducing the sampling frequency, number of samples is reduced and less points are plotted. First example of comparing sampling frequency of 30 and 60 minutes shows that more than half points (points over UWL) is lost, and this can have consequences for entire water level monitoring process. For summer period, sampling frequency should be 60 minutes since water levels do not show many changes and water levels are similar, however for winter period, sampling frequency of 30 minutes proves to be the best. Even though it provides many data, this period is flagged as high-risk period for floods and water levels should be monitored more frequently and provide chance for timely reaction of operator, which is what this frequency allows.

Complimentary to the I Charts, following comparative analysis of R Charts has been conducted. Within the following R charts one sample is represented by two consecutive measurements taken in different time periods. This means that samples in plot a take two consecutive measurements taken every 30 minutes, plot b takes two consecutive measurements taken every 60 minutes and plot c takes two consecutive measurements taken every 90 minutes. Since all charts use two consecutive measurements to represent one sample, following constants were used. $D_{0001} = 4.12$, $D_{0025} = 2.81$, $D_{0975} = 0.44$, $D_{0999} = 0$. R charts detect the range or changes of water levels between measured values and are one of the most reliable charts used to determine most efficient sampling frequency. Following charts are created using same datasets as previous charts, two datasets, one for 6th month, which belongs to the summer period and one for the 12th month which belongs to the winter period. The datasets of summer period are further modified to simulate different sampling frequencies, as if samples were collected and analysed every 30, 60 and 90 minutes. The chart shown in Figure 5.7 plot a is created based on dataset of summer period and contains 720 samples taken every 30 minutes, plot b contains 480 data collected every 60 minutes and plot c contains 320 samples collected every 90 minutes.

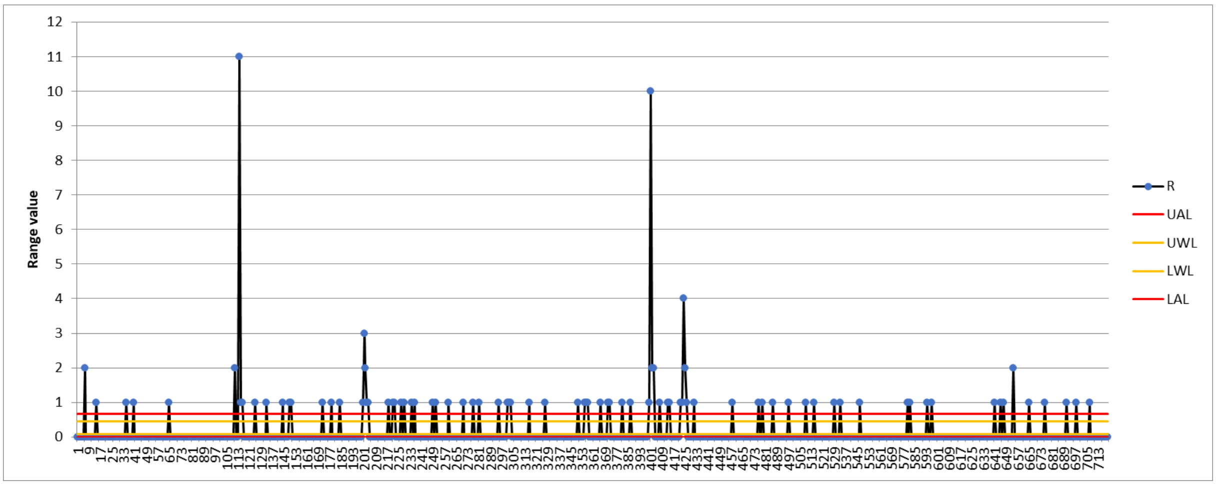
Since R Charts work on a principle comparing two consecutive measuring's and displaying difference between those measuring's, it is expected for all plots to be similar and have minimum difference between water levels, since summer period with low water level variations in analysed. However, when compared plot a with plot's b and c that for date 6.9.2020, plot a displays much lower range between consecutive values, while plots b and c are pretty much similar in this case. Table 5.3 displays values measured in that period to provide better and more understanding explanation of Figure 5.7 and to provide better insight into the dataset used.

Table 5. 3 Water level measuring's for summer period

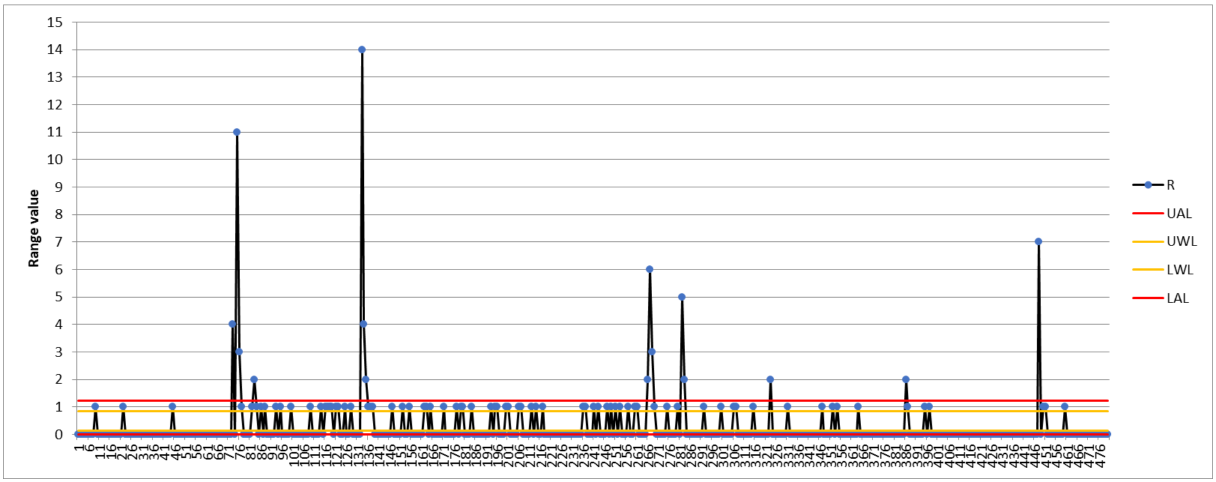
6/9/2020 6:30	27
6/9/2020 7:00	27
6/9/2020 7:30	26
6/9/2020 8:00	37
6/9/2020 8:30	40
6/9/2020 9:00	34
6/9/2020 9:30	32
6/9/2020 10:00	30
6/9/2020 10:30	29
6/9/2020 11:00	28
6/9/2020 11:30	27
6/9/2020 12:00	27
6/9/2020 12:30	27

For some reason in this period water level had a sudden increase, this also might be due to malfunction of device, summer rains, or some other reason, these things should also be considered when making control charts and should be handled carefully. Plot a which has a sampling frequency of every 30 minutes in case of measuring's taken at 7.30 and 8.00 shows difference of 11 cm, however since plot b has a sampling frequency of 60 minutes it only takes into consideration values taken at 7.30 and 8.30 which results in difference of 14cm between water levels.

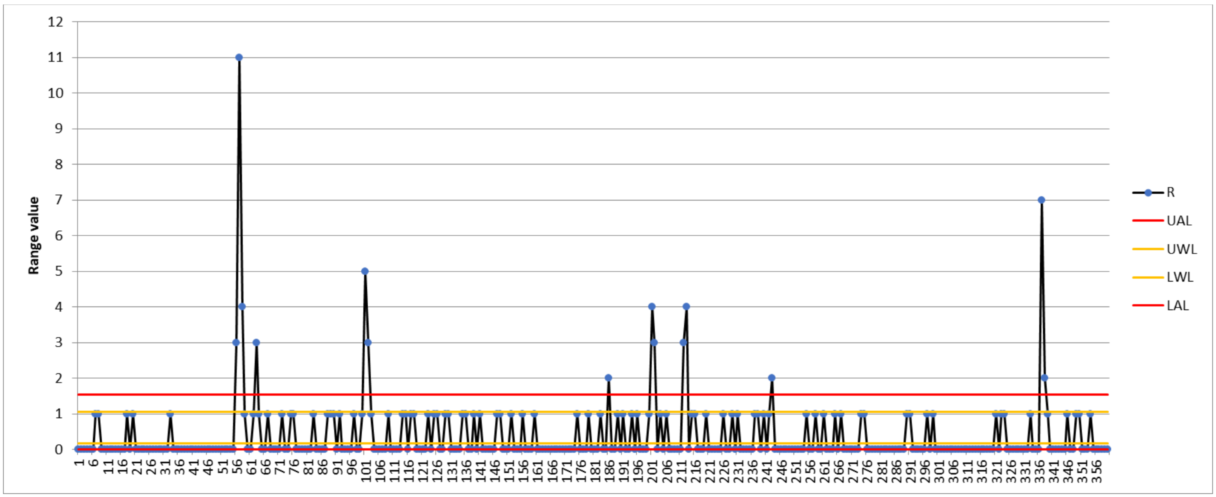
The following plots shown in Figure 5.7 representing 6th month have very narrow control line values. This is caused by the fact that there is very small difference between the measured values ranging mostly between 0 and 1 with occasional larger differences, however these differences are not common enough to cause that control line values be increased.



a) Sampling frequency every 30 minutes



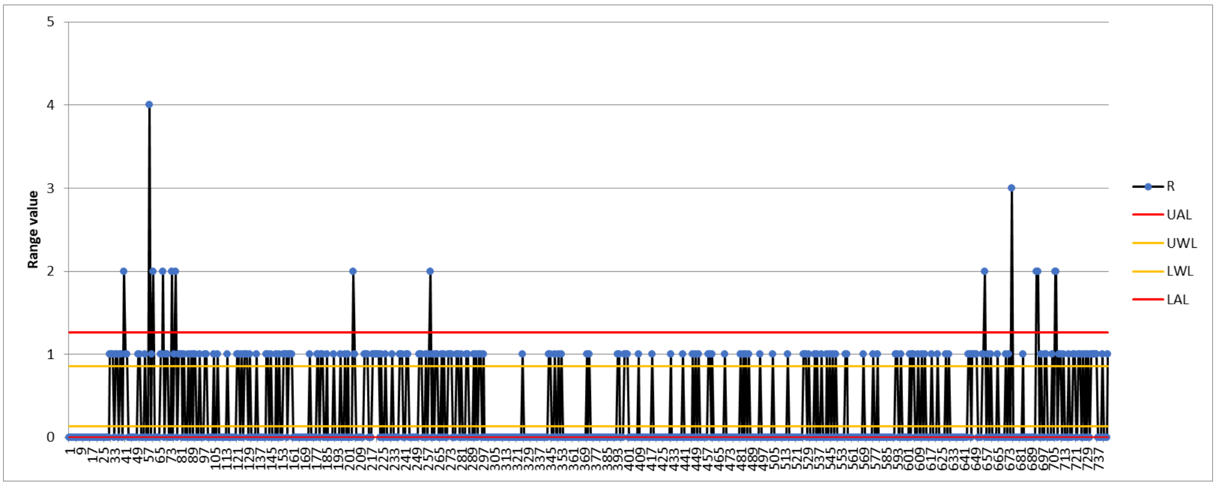
b) Sampling frequency every 60 minutes



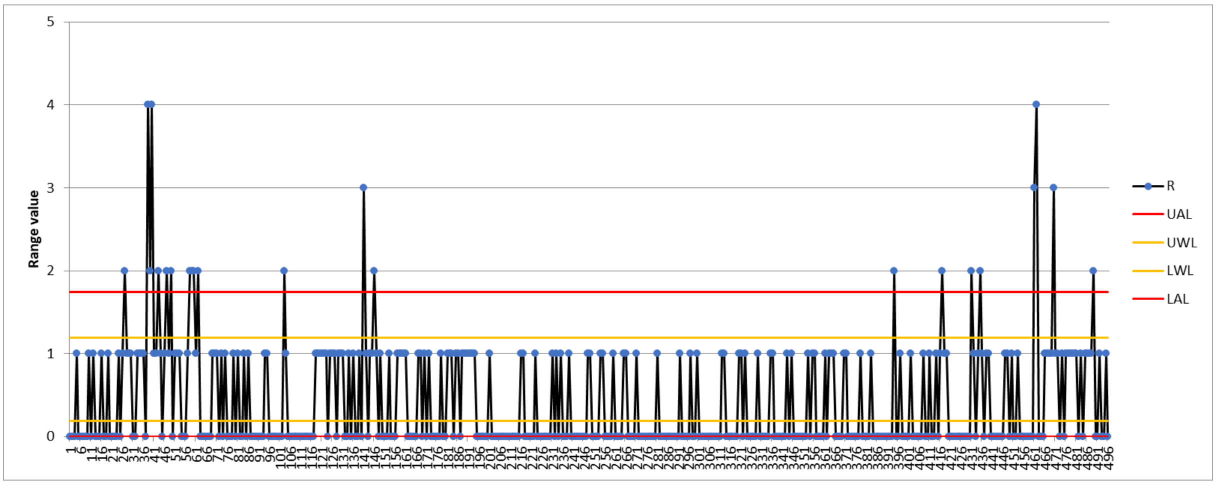
c) Sampling frequency every 90 minutes

Figure 5. 7 R Charts for different sampling frequencies in summer period

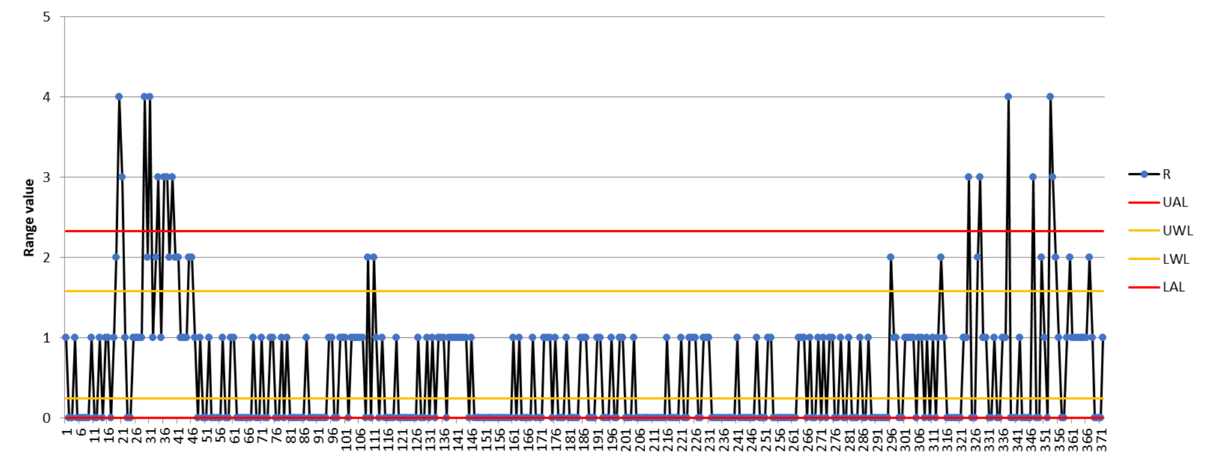
Due to low water level, as well as low range values, control charts have very low control line values, which is normal and expected for the summer period. These charts shown in Figure 5.7 can be used alone, however they might be more effective and useful if used as complimentary to the previously analysed I Charts since these charts provide closer insight to water level increasements. R charts may not be very useful during the summer period due to low and constant water levels to make them more useful and easier to read, sampling frequency for these charts should be minimum every 60 minutes, or even better, every 90 minutes since these sampling frequencies provide better control line values. Winter period where water levels are subjects to large water level changes, as shown in Figure 5.8, they prove to be very useful and effective because they give insight and show accumulated increasements of water levels.



a) 30 minutes sampling frequency



b) 60 minutes sampling frequency



d) 90 minutes sampling frequency

Figure 5. 8 R Charts for different sampling frequencies in winter period

From the point of sampling frequency, more frequent sampling does not always provide better results. When compared plots a, b, c, of a winter period shown in Figure 5.8, plot a has noticeable differences from other two plots, reason for this is because it takes data more frequently and compares them, providing real time update of information and providing in time reaction for the operator. However, if sampling frequency is too large, especially in period like this where water levels constantly increase, whole point of monitoring water levels might be lost since warning of increasement would come too late. Best example for this is comparing plots a and b or a and c of Figure 5.8. Plot a which has 30 minutes sampling frequency shows continuous growth of water levels with smaller differences, while plots b and c which have sampling frequency 60 or 90 minutes, might provide water level growth much later than it happens.

Table 5. 4 Water level values in Winter period

12/3/2020	25	25	25	24
12/3/2020	24	25	25	25
12/3/2020	26	26	27	27
12/3/2020	27	28	28	28
12/3/2020	28	28	28	32
12/3/2020	32	32	31	30
12/3/2020	30	32	34	34
12/3/2020	33	33	32	32
12/3/2020	33	33	34	35
12/3/2020	36	37	37	39
12/3/2020	39	39	39	40
12/3/2020	40	41	43	43
12/4/2020	44	44	45	47
12/4/2020	49	49	50	51
12/4/2020	50	52	53	53
12/4/2020	53	54	54	55
12/4/2020	55	56	57	57
12/4/2020	57	58	58	58
12/4/2020	59	59	59	60
12/4/2020	59	59	59	60
12/4/2020	61	61	62	63
12/4/2020	63	64	65	65
12/4/2020	66	66	66	67
12/4/2020	68	68	68	68

To provide more understanding explanation of previous statements, as well as to give better insight into the data used and water level changes during a winter period Table 5.4 represents water level values measured every 30 minutes. Water levels change for more than 40 cm in one day. Plot a in Figure 5.8 takes these values one after another, consecutive values, and compares them, providing difference between these values and plotting them in the chart, however plot b and c created by using sampling frequency of 60 and 90 minutes, take every second or third value and compares them to the previous one. To that reason plot a provides more clear and reliable information regarding water level increasements, while plot b and c skip one or two measuring's and compare them to the previous ones, displaying much larger differences between water levels, and there is also matter of time delay, especially since in these situations time is important factor. When compared all these charts together for summer and winter period it is safe to the conclude that in case of using R charts for water level monitoring process best used sampling frequency for summer period is 60 minutes because water levels are almost constant, low and don't have that many changes and control line values for this sampling frequency are adequate. According to previous figures, 60 minutes sampling frequency when compared to 30 minutes sampling frequency provides higher control line values causing that only few points cross over upper action line, as well as displays all important values in the chart. For winter periods most efficient sampling frequency is every 30 minutes since water levels change drastically in very short period and this sampling frequency provides continuous and in time update of water level behaviour. Control line values calculated with using 30 minutes frequency provide lower control line values which in this case is desirable because this is high risk period, and it is better to have closer insight and additional information on water level behaviour.

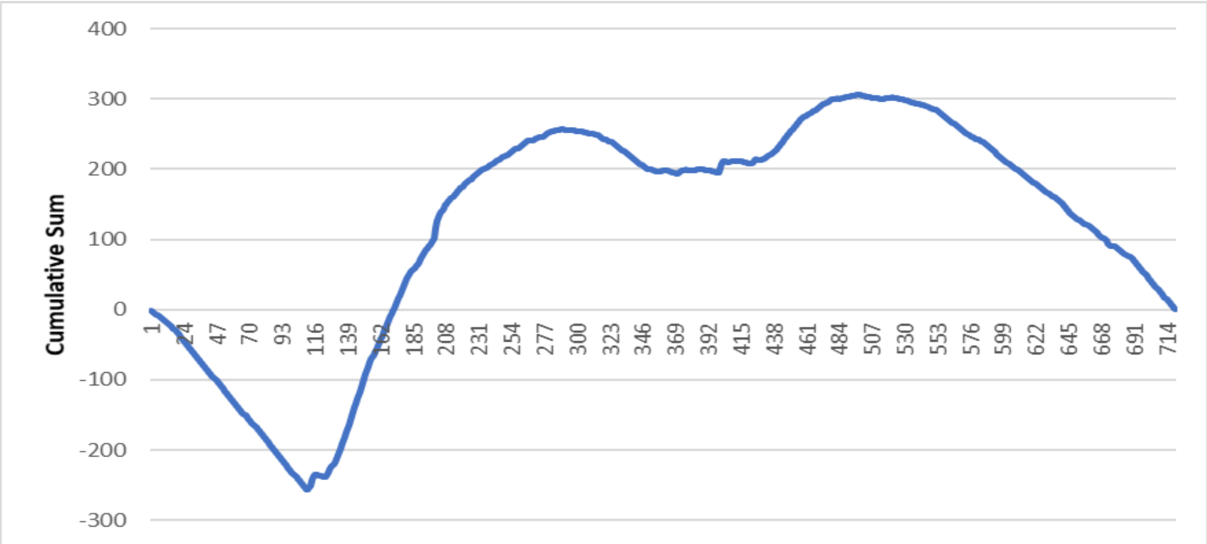
Table 5. 5 Comparative overview of R Charts control line values for different sampling frequencies

R CHARTS for 06.2020						
Sampling frequency	Number of samples	UAL	UWL	LWL	UAL	Maximum and Minimum value
30 minutes	720	0.7	0.5	0.1	0	11 / 0
60 minutes	480	1.2	0.8	0.1	0	14 / 0
90 minutes	360	1.5	1.1	0.2	0	11 / 0
R CHARTS for 12.2020						
30 minutes	744	1.3	0.9	0.1	0	4 / 0
60 minutes	496	1.7	1.2	0.2	0	4 / 0
90 minutes	372	2.3	1.6	0.2	0	4 / 0

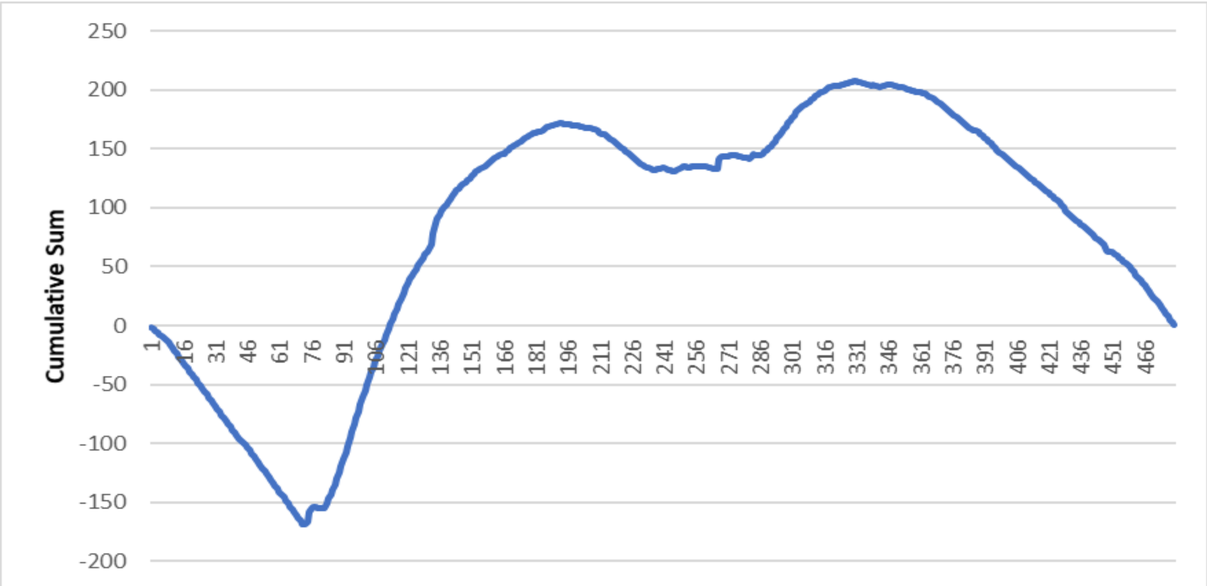
Table 5.5 represents analysis and changes of control line values of R control charts. Changing sampling frequency results in changes of control line values due to changes in range value, providing better and more readable control charts with larger values of control lines. By changing the time between the measuring water levels change and larger differences caused larger range value which has directly affected control line values to increase. For plot a in Figure 5.7 almost all values are shown to be out of the control lines, this is caused by low control line and range values. Low control line values in this case are caused by small variations of water levels in the sampling period. In plots b and c of Figure 5.7 by changing sampling frequency control line values increase, almost double than the original one, putting this process under the control. This shows that sampling frequency does play important role when using R control charts for water level monitoring process.

For summer period sampling frequency of 60 or 90 minutes proves to be the best since it shows all important increasements of water levels and water levels do not change so frequently. These sampling frequencies generate control line values which make this entire process easier to monitor, while winter period has more frequent and continuous water level changes and sampling frequency of 30 minutes should be used for this period.

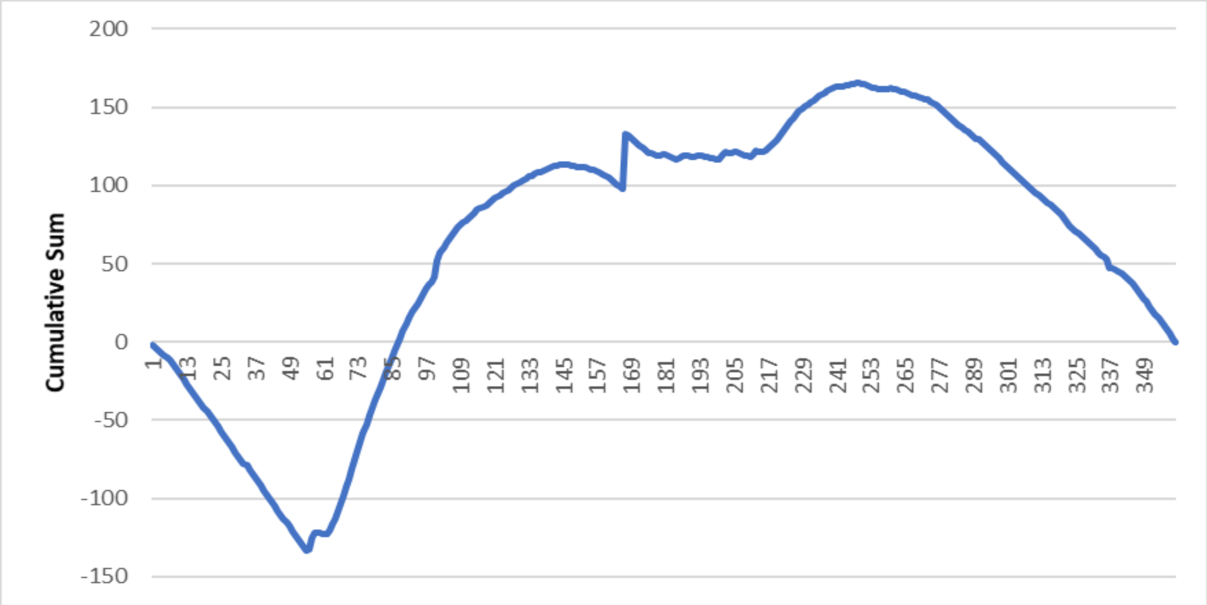
Last type of charts to be analysed and compared is Cumulative Sum or CuSum charts. These charts plot the cumulative sum of deviations from the targeted value for individual measurements and are usually applied to the production processes within companies. However, these charts have been applied to display in which direction water level changes, whether it is increasing or decreasing. Following the same principle from previous examples, CuSum charts for 06.2020 with three different sampling frequencies have been created. Chart in a plot a contains 720 samples collected every 30 minutes, while plots b and c contain 480 and 360 samples. For this period as targeted value is taken average value of all samples from 6th month of 2020, 23.5 cm and to create more sensitive chart, tolerance to increase or decrease of the water level from average water level value used is 1 cm.



a) Sampling frequency 30 minutes



b) Sampling frequency 60 minutes

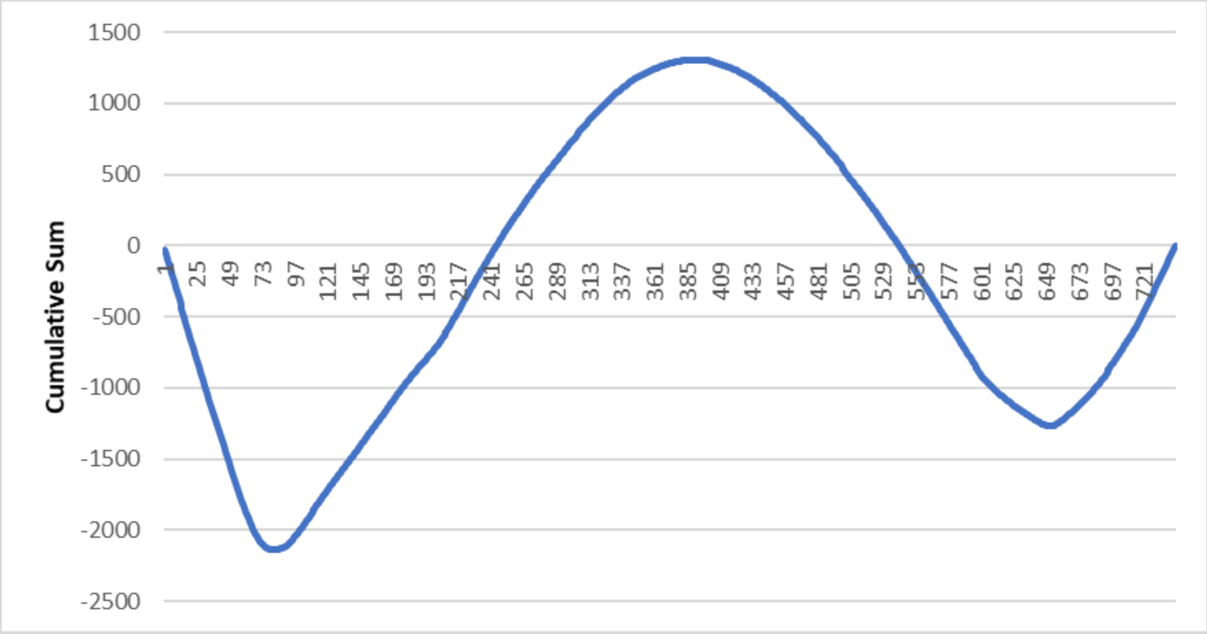


c) Sampling frequency 90 minutes

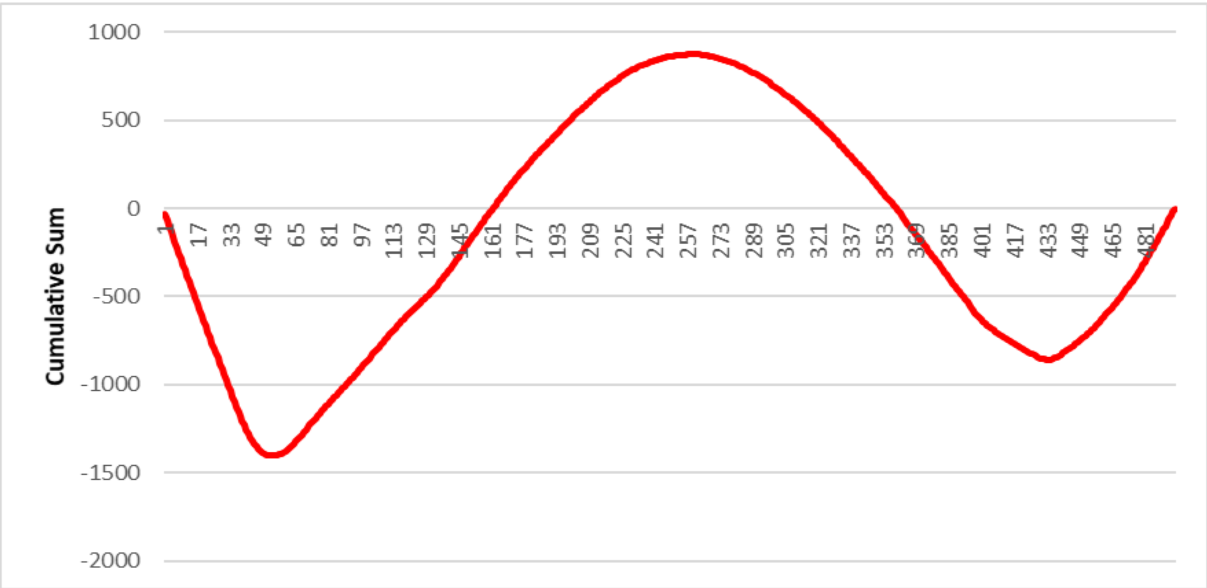
Figure 5. 9 CuSum Charts for different sampling frequencies in summer period

Figure 5.9 represent different plots of CuSum charts according to the different sampling frequencies. Although same dataset values, same target and tolerance values in all three plots are used, it is clear that by changing the sampling frequency only cumulative sum for all three plots is different but these values do not have significant role in this case. Comparing plots a and c, plots with smallest and largest sampling frequency it is noticeable that charts gives different representation of data for the date 12.06.2020 due to changed sampling frequency changes are more extreme. Changes within the maximum and minimum CuSum values caused by changing the sampling frequency are not that important since in this case only chart direction (increase or decrease) is what can be monitored.

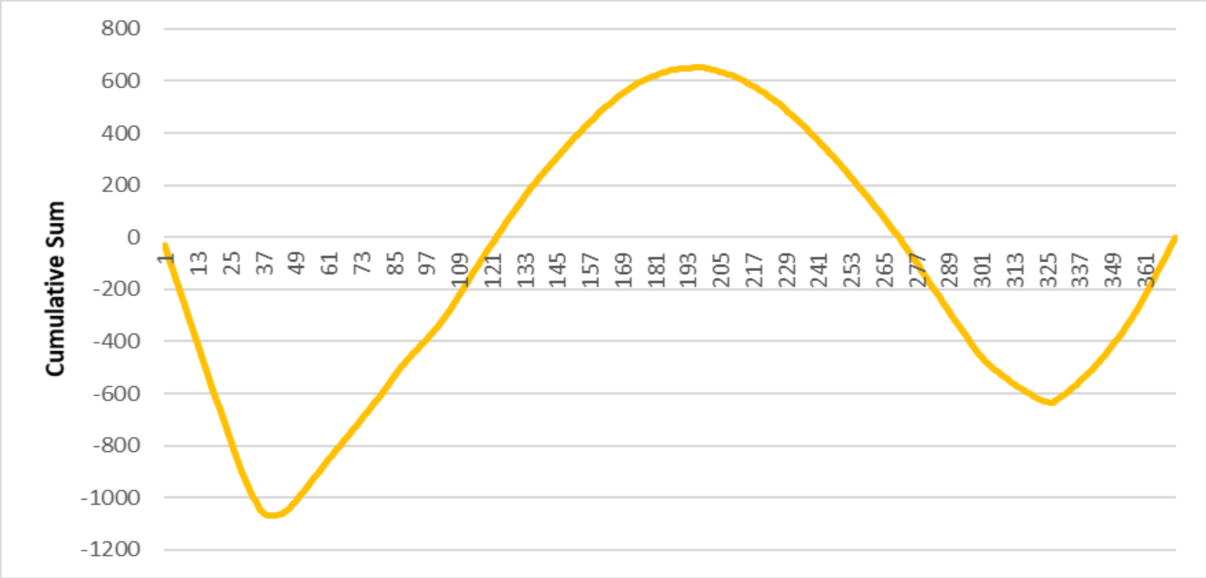
Figure 5.10 represent CuSum charts created for winter period where differences between water levels are much larger, however, to maintain consistency of the analysis, average value of all samples is taken as target value and tolerance to water level increase or decrease given is 1 cm. Plot a analyses 744 samples, plot b analyses 496 samples and plot c is created by using 372 samples.



a) Sampling frequency 30 minutes



b) Sampling frequency 60 minutes



c) Sampling frequency 90 minutes

Figure 5. 10 CuSum Charts for different sampling frequencies in winter period

Like the previous example, all three plots have different cumulative sum maximum and minimums which change by changing the sampling frequency and number of samples.

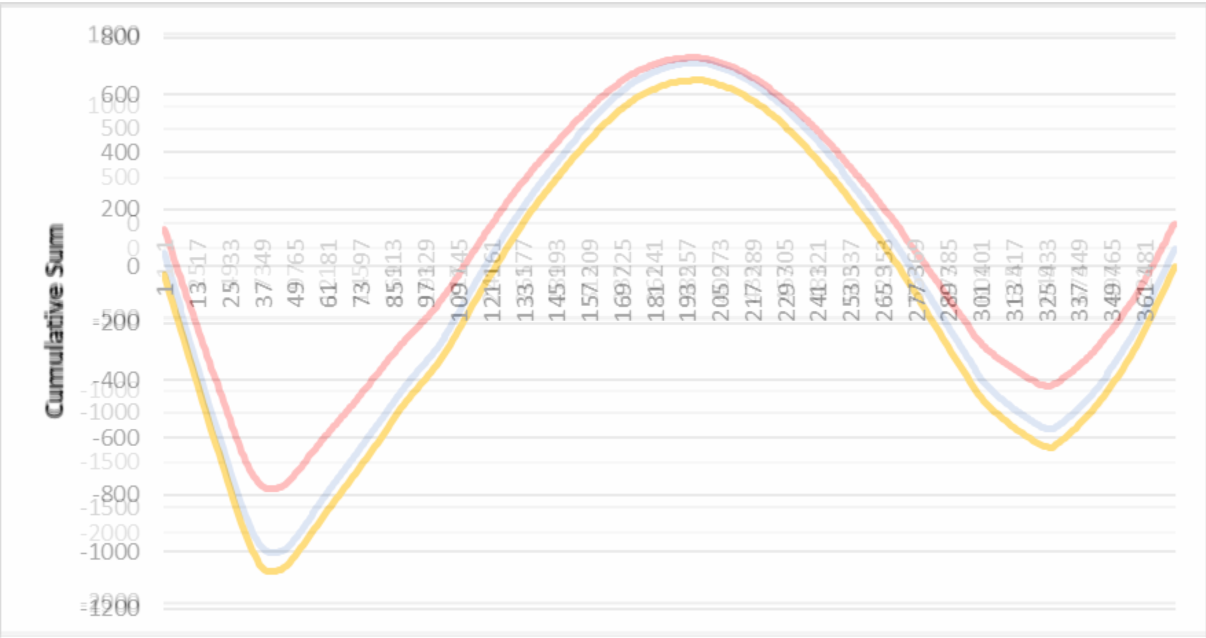


Figure 5. 11 Comparation of CuSum Charts for different sampling frequencies in winter period

With larger sampling frequency difference between measuring's increases and number of samples decrease, which causes cumulative sum to change more rapidly. Figure 5.11 shows that sampling frequency does not affect look of the CuSum chart.

CuSum charts as a tool can be used to display how water level change in time but they should always be used in combination with some other chart since these charts only provide information if water levels are increasing or decreasing without any concrete and usable values of how much water level increased or decreased.

6 Conclusion

Control charts which are mostly used in production processes or companies, can be used as well as for monitoring and keeping in control environmental processes such as water level rising and reduction from risk of floods. This thesis has focused mainly on determination of most efficient sampling frequency of water level from water level measuring device. By defining most efficient sampling frequency and preparing user friendly control charts, water level monitoring process can be done by everyone, regardless of level of education or experience.

According to data collected for previous three years seasonality can be determined, however it is also clear that every year period of low water levels is longer and water levels during the winter are lower than in previous years which shows changes in climate and is probably caused by global warming. Charts analysed through the thesis for one month from each of these two periods clearly show changes in water levels, river behaviour and control lines used for its monitoring. Comparison made for sixth month of the year represents analysis of summer period of year because water level values for period from 5^{th} to 10^{th} month are very similar and comparison for twelfth month represents analysis of winter period because this month contains largest and most often water level changes recorded in this year. By changing sampling frequency, it is shown that for all three types of charts there is changes in control line values therefore the following conclusions are derived.

- For I charts in the summer period changing sampling frequency results in small almost insignificant changes in control line values.
- With lesser number of data analysed certain points shown in chart are lost, which is expected, however changing sampling frequency can show extreme values to be larger than real ones or to disregard them completely if they are not included into dataset analysed as shown in analysis of both, I and R charts for 6^{th} month.
- For summer period where water levels are low, sampling frequency should be every 60 minutes because this sampling frequency provides good representation of water level behaviour, includes enough data, and provides time for reaction of the operator. Since this period does not propose risks of floods rather than draughts low water levels should be monitored more closely.
- For winter period, sampling frequency should be every 30 minutes because this period is characteristic for high water levels and frequent and water level changes. This

sampling frequency provides good values of control lines and enough data to monitor the process of water level changes as well as gives operator just enough time for reaction in case of sudden increasements.

- Sampling frequency has large effect on control line values of R charts. For sampling frequency of every 30 minutes in summer period R charts have very narrow and small control line values which causes too many too many points to be over the upper action line causing the system to appear out of the control as shown previously. Therefore, sampling frequency for this period should be 60 or 90 minutes, while for winter period 30 minutes sampling frequency proves as most efficient.
- One month at a time should be analysed since water level changes are too frequent for some periods which might result in unreliable control line values and unreadable control charts. Negative lower action lines in the chart should be modified to value of 0 since that is minimum level of riverbed suggesting draught.
- CuSum charts have similar looking plots because regardless of changes in sampling frequency, direction of water level changes are always same. These charts only provide information of direction of water level changes, meaning they show only if water level is increasing or decreasing and should always be combined with some other process monitoring tool.

This thesis shows new approach to the application of control charts in monitoring natural processes. These analyses were conducted based on data of water levels with simulated different sampling frequencies. However, to have clear understanding of entire process, as well as to have more reliable data and course of action in process of monitoring and protection from floods, all influencing factors such as level of precipitation, type of the land in and around the riverbed, course and depth of riverbed, tributaries of this river should be taken into account.

7 References

[1] Bajrić H., Kadrić E., Vučijak B. (2018). "Six sigma - osnovni statistički alati", šinski fakultet Sarajevo, Bosna i Hercegovina, ISBN 978-9958-601-70-5, COBISS.BH-ID 25545222

[2] Benneyan C. J. 2008. *„The design, selection, and performance of statistical control charts for healthcare process improvement",* International Journal of Six Sigma and Competitive Advantage 4(4):209-239 (https://www.researchgate.net/publication/228941929_The_design_selection_and_performance_of_statistical_control_charts_for_healthcare_process_improvement, date of access: 1.5.2021)

[3] *„Development and use of reference materials and quality control materials"* (https://www.researchgate.net/publication/236896225_Development_and_use_of_reference_materials_and_quality_control_materials, date of access: 17.5.2021)

[4] Dr. Bill McNeese, 2014. *„Keeping the Process on Target: CUSUM Charts"*, BPI Consulting (https://www.spcforexcel.com/Downloads/pdf/Keeping-the-Process-on-Target-CUSUM-Charts.pdf, date of access: 2.5.2021)

[5] Dr. Bill McNeese. 2016. „Control Chart Rules and Interpretation" (https://www.spcforexcel.com/knowledge/control-chart-basics/control-chart-rules-interpretation, date of access: 17.5.2021)

[6] Drenovac A., Drenovac B. & Drenovac D. (2013). Control cards as a statistical quality control resource. Military Technical Courier. 61. 101-122. (http://oaji.net/articles/2016/1899-1458237485.pdf , date of access: 10.5.2021)

[7] Evans C. W. *„Exponentially-Weighted Moving Average"* (https://www.geogebra.org/m/tb88mqrm, date of access: 2.5.2021)

[8] Flegner P., Kacur J., Durdan M. & Laciak M. 2020. *„Statistical Process Control Charts Applied to Rock Disintegration Quality Improvement"* Institute of Control and Informatization of Production Processes, Faculty BERG, Technical University of Košice, Němcovej 3, 042 00 Košice, Slovakia (https://www.mdpi.com/2076-3417/10/23/8343/htm, date of access: 15.5.2021)

[9] Gönül Şengöz N. 2018. *“Control Charts to Enhance Quality”*. (https://www.intechopen.com/books/quality-management-systems-a-selective-presentation-of-case-studies-showcasing-its-evolution/control-charts-to-enhance-quality, date of access: 2.5.2021)

[10] Haridy S., Maged A., Kaytbay S. & Araby S. 2017. *„Effect of sample size on the performance of Shewhart control charts“*. The International Journal of Advanced Manufacturing Technology volume 90, 1177–1185 (https://link.springer.com/article/10.1007%2Fs00170-016-9412-8, date of access: 15.5.2021)

[11] Horvat, D., Eđed, A. & Banaj, Đ. 2006. *„Statistička kontrola procesa i proizvoda u poljoprivredi“* (https://hrcak.srce.hr/5965, date of access: 7.5.2021)

[12] iSixSigma, 2002. (https://www.isixsigma.com/tools-templates/control-charts/why-control-chart-your-processes/, date of access: 28.5.2021)

[13] Jašarević S., Brdarević S., Isaković S. 2020. *“Upravljanje proizvodnjom”,* University of Zenica, Polytechnic faculty (https://suvadisakovic.ba/knjige/upravljanje_proizvodnjom.pdf , date of acces: 10.5.2021)

[14] Jones G. *„Walter Shewhart (1891 - 1967)“*. Managers-net (https://www.managers-net.com/Biography/Shewhart.html, date of access: 2.5.2021)

[15] Juran, J.M., Gryna, M.F. (1999.): Planiranje i analiza kvalitete. MATE, Zagreb: p. 664

[16] Maugeri A. & Arcidiacono G. 2014. *“Efficient Determination of Sampling Rate and Sample Size in Statistical Process Control”*. Universal Journal of Industrial and Business Management 2: 103-110 (https://www.hrpub.org/download/20140405/UJIBM2-11602141.pdf, date of access: 2.5.2021)

[17] Morrison L. 2009. *„The Use of Control Charts to Interpret Environmental Monitoring Data“,* Natural Areas Journa:66-73 (https://www.researchgate.net/publication/232678945_The_Use_of_Control_Charts_to_Interpret_Environmental_Monitoring_Data/stats, date of access: 1.5.2021)

[18] NCSS Statistical Software, „X-bar and R Charts“ (https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/X-bar_and_R_Charts.pdf, date of access: 28.5.2021)

[19] Özdagoglu A., Özdagoglu G., Güler Mehmet E. 2018. *„Developing quality control charts for the control points of a food product"*. International Journal of Contemporary Economics and Administrative Sciences, 1925-4423 (https://www.researchgate.net/publication/330542922_DEVELOPING_QUALITY_CONTROL_CHARTS_FOR_THE_CONTROL_POINTS_OF_A_FOOD_PRODUCT, date of access: 1.5.2021)

[20] QualityMag. *„Sampling Frequency in Manufacturing: Too Much of a Good Thing?"* (https://www.qualitymag.com/articles/95681-sampling-frequency-in-manufacturing-too-much-of-a-good-thing, date of access: 17.5.2021)

[21] Schenkelberg F. 2017. *„How to Select the Right Control Chart"* (https://accendoreliability.com/selecting-control-charts/, date of access: 28.5.2021)

[22] Smith L. Jim 2009 *„Remembering Walter A. Shewhart's Contribution to the Quality World"*

[23] *(https://www.qualitymag.com/articles/85973-remembering-walter-a-shewharts-contribution-to-the-quality-world* , date of access: 06.06.2021)

[24] Suman G. & Prajpati D. 2018. *„Control chart applications in healthcare: a literature review"*. Department of Mechanical Engineering Punjab Engineering College (https://www.metrology-journal.org/articles/ijmqe/full_html/2018/01/ijmqe170050/ijmqe170050.html, date of access: 01.5.2021)

[25] Toledo J. C., Lizarelli F., Santana J. M. 2017. „Success factors in the implementation of statistical process control: Action research in a chemical plant". Universidade Federal de São Carlos (https://www.researchgate.net/publication/317508426_Success_factors_in_the_implementation_of_statistical_process_control_Action_research_in_a_chemical_plant, date of access: 1.5.2021)

[26] Vucijak B., Kupusović T., Kurtagić M. S., Ceric A. 2013. *„Climate Change Assessment Using Statistical Process Control Methods"*

(https://www.researchgate.net/publication/299712748_Climate_Change_Assessment_Using_Statistical_Process_Control_Methods, date of access: 1.5.2021)

[27] Walter J Murphy.1945. *"Use of Control Charts in Analytical Laboratory"*. Grant Wernimont. Industrial & Engineering Chemistry Analytical. Edition 1946, 587-592 (https://pubs.acs.org/doi/pdf/10.1021/i560158a001the, date of access: 2.5.2021)

[28] Woodall, Spitzner, Montgomery & Gupta. 2004. Using Control Charts to Monitor Process and Product Quality Profiles. Journal of Quality Technology. 36. (https://www.researchgate.net/publication/228746999_Using_Control_Charts_to_Monitor_Process_and_Product_Quality_Profiles/stats, date of access: 25.4.2021)

8 Appendices

Attachment I - CONSTANTS FOR AVERAGE VALUE CONTROL CHARTS

Sample size n	**Hartley's constant d_2**	**Constants for average control charts**			
		Sample range		**Average standard deviation of the sample**	
		A_2	**2/3 A_2**	**A_3**	**2/3 A_3**
2	1,128	1,88	1,25	2,66	1,77
3	1,693	1,02	0,68	1,95	1,30
4	2,059	0,73	0,49	1,63	1,09
5	2,326	0,58	0,39	1,43	0,95
6	2,534	0,48	0,32	1,29	0,86
7	2,704	0,42	0,28	1,18	0,79
8	2,847	0,37	0,25	1,10	0,73
9	2,970	0,34	0,20	1,03	0,69
10	3,078	0,31	0,21	0,98	0,65
11	3,173	0,29	0,19	0,93	0,62
12	3,258	0,27	0,18	0,89	0,59

Attachment II - CONSTANTS FOR CONTROL CHARTS FOR RANGES

Sample size n	**$D'_{0,999}$**	**$D'_{0,001}$**	**$D'_{0,975}$**	**$D'_{0,025}$**	**D_2**	**D_4**
2	0,00	4,12	0,44	2,81	0	3,27
3	0,04	2,98	0,18	2,17	0	2,57
4	0,10	2,57	0,29	1,93	0	2,28
5	0,16	2,34	0,37	1,81	0	2,11
6	0,21	2,21	0,42	1,72	0	2,00
7	0,26	2,11	0,46	1,66	0,08	1,92
8	0,29	2,04	0,50	1,62	0,14	1,86
9	0,32	1,99	0,52	1,58	0,18	1,82
10	0,35	1,93	0,54	1,56	0,22	1,78
11	0,38	1,91	0,56	1,53	0,26	1,74
12	0,40	1,87	0,58	1,51	0,28	1,72