Implementation copula to multivariate control chart

To cite this article: T E Lestari *et al* 2019 *J. Phys.: Conf. Ser.* **1402** 022046

Journal of Physics: Conference Series **1402** (2019) 022046 doi:10.1088/1742-6596/1402/2/022046

Implementation copula to multivariate control chart

T E Lestari*, S K Nisa, F Citra and G B Asri

Industrial Engineering, Faculty of Engineering and Technology, Sampoerna University, Jakarta, Indonesia

*tika.lestari@sampoernauniversity.ac.id

Abstract. Multivariate Control Chart is an effective tool in Statistical Process Control to identify either an out-of-control process or in-control process. Hotelling T^2 control chart is a quite popular and widely used technique in this field. However, its performance is deteriorated when the underlying distribution of the quality characteristics is not following the multivariate normal distribution. Multivariate control chart usually recommends a procedure in the phase-in Hotelling T^2 control chart although there is difficulty in interpreting the signals from multivariate control charts more work is needed on data reduction methods and graphics techniques. Basically, the multivariate control chart refers to the theory of prediction interval. Therefore, it's called predictive multivariate control charts. The purpose of this research is to construct what so-called predictive multivariate control chart both classical in this part in the phase-in Hotelling T^2 and Copula-based ones. We argue that appropriate joint distribution function may be well estimated by employing Copula. A numerical analysis is carried out to illustrate that an Application Copula-based Multivariate control chart outperforms than bivariate control chart and others.

1. Introduction

Multivariate Control Chart is defined as the application of multivariate statistical procedures which are useful for tracking large numbers of variables in the entire chart. Process monitoring of problems in which several related variables are of interest is collectively known as multivariate statistical process control [1]. The performance of the multiple parameters which are correlated to each other can be detected. This method has been adopted in the manufacturing process, patient care, and customer satisfaction. There are several charts of Multivariate Control Chart, one of them is the Hotelling T^2 Control Chart.

Hotelling T^2 rule is the most common multivariate control chart for monitoring the statistical process of data. However, Hotelling T^2 rule depends on the assumption that the controlled data is a Gaussian which rarely true in practice. Therefore, the other approach is needed to solve the multivariate case that is not underlying the normal distribution. To solve this irregular distribution of the multivariate case, copula can be one of the solutions. A copula is defined as the technique used in a statistical application of control chart assessment in order to estimate the joint distribution function that has a non-normal margin. The copula approach is a popular method for multivariate modeling applied in several fields; it defines non-parametric measures of dependence between random variables [2]. Moreover, this method is also used to estimate the random variable of the statistical data which has a dependent structure.

2. Method

Currently, in industry, there are many situations in which the industry needs to monitor or control the two or more related quality–process characteristics. Based on research, it seems reflect that most of the manufacturing companies are satisfied with labour production output, under a given wage level [3]. Statistical Process Control (SPC) is a statistical method used to achieve continuous improvement in quality [4].

It is possible to do monitoring the quality characteristic independently, but it can be very misleading. The monitoring process of problems in which several related variables are of interest is collectively known as multivariate statistical process control. This control chart has function to monitor whether or not the processes remain in control. Control charts are widely used as process monitoring tools, primarily to detect changes in the process mean or in its standard deviation, which can indicate deterioration in quality. The most useful tool for multivariate statistical process control is the quality control chart. The result of the excellent quality of the product comes from a good quality control. Many companies use specialized methods to produce good-quality products. For this reason, quality control is necessary to keep the product generated by the applicable quality standards [5].

Quality control problem appears when the process or the product with the multivariate variables is to be monitored or controlled. When these variables are correlated, the more appropriate approach would be required to monitor them at the same time. The variables will be correlated to the huge number of the correlated variables collected at the time. Consequently, multivariate control chart is needed for monitoring and diagnosing purposes in modern manufacturing systems. There are many appropriate methods of detecting and isolating process faults to utilize Multivariate Statistical Process Control (MSPC) approaches. One of the methods is Hotelling T^2.

The quality control used in research at the Bandung bakery production is carried out in an attribute that is the measurement of quality on product characteristics that cannot or are difficult to measure. The characteristics in question are good or bad product quality, in this case we have three characteristics to be analyse which defect in product or in this case bread production, the size of the bread and net or the net weight of the bread. In analysing these characteristics in this study we use the role of Hotelling T^2 and Copula in setting up the control chart. The population of this study was all types of bread produced by Bandung bakery as many as eight types of bread with a production capacity of 2,000 packs per day. Each model has generated 250 packages per day, resulting in a total population of 62,000 packages for 31 days (1 month). The sampling technique in this study was a withdrawal of bread type samples based on judgment sampling and random sampling when Bandung bakery produced eight types of bread the writers only took four types of bread products. Types of bread that the writer took was chocolate bread, peanut bread, cheese bread and green bean bread (data report collected in December 2018). Data on production results that have been obtained will be processed using Statistical Quality Control (SQC) analysis using Hotelling T^2 and Copula method.

2.1. Hotelling T^2

Hotelling's T^2 is a scalar that combines information from the dispersion and the mean of several variables [6]. As in the Univariate case, when data are grouped, the T^2 chart can be paired with a chart that displays a measure of variability within the subgroups for all the analyse characteristics. The combined T^2 and dispersion charts are thus a multivariate counterpart of the Univariate and S (or and R) charts.

The Hotelling T^2 control chart is one type of starting control for multivariate cases. This control chart is used for multivariate cases with the assumption that random variables are used with normal distribution. The lack of this control chart is the inability to interpret dependence between each random variable [7]. The stages in making the Hotelling T^2 control chart include the Start Up Stage (SUS) / Phase I and Process Control (PC) stages / Phase II.

2.1.1. Start-up stage / phase I. In the Phase I, the statistic that will be used is:

$$T_k^2 = n\left(\overline{X_k} - \bar{\bar{X}}\right)^t S^{-1}\left(\overline{X_k} - \bar{\bar{X}}\right) \quad (1)$$

$$\overline{X_k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k \tag{2}$$

$$\bar{X} = \frac{1}{m}\sum_{k=1}^{m} \bar{X} \tag{3}$$

$$S_{k=}\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i^k - \overline{X_k}\right)^t\left(X_i^k - \overline{X_k}\right) \tag{4}$$

$$\bar{S} = \frac{1}{m}\sum_{k=1}^{m} S_k \tag{5}$$

$$\text{Since } T_k^2 \sim \frac{p(m-1)(n-1)}{mn-m-p+1} \tag{6}$$

So then, the control limit of Phase I is expressed by;

$$UCL = \frac{p(m-1)(n-1)}{mn-m-p+1} F_{a,p,mn-m-p+1} \tag{7}$$

$$LCL = 0 \tag{8}$$

In the Phase I, if there are any points above the limit of UCL, then it is identified that the point is in out-of-control condition. The point above the UCL will be erased and then the UCL is recalculated without the certain point.

2.1.2. Process control phase (Phase II). Phase II is used to monitor the future production. The form of the control limit is different, regarding to the different distribution theory. In the Phase II, the future subgroups are included and they are assumed to be unconstrained of the recent set. Let $\bar{\bar{X}}$ and $\bar{S}$ respectively state grand mean sample and covariance matrix that are obtained from the Phase I. In the subgroup, the sample of $\overline{X_f}$ is mutually independent with $\bar{\bar{X}}$ and $\bar{S}$. To get the distribution of the data, the distribution $\overline{X_f} - \bar{\bar{X}}$ is needed to know. Because $\overline{X_f}$ and $\bar{\bar{X}}$ is independent with $\overline{X_f} \sim N_p\left(\bar{\mu}, \frac{\Sigma}{mn}\right)$ and $\bar{\bar{X}} \sim N_p\left(\bar{\mu}, \frac{\Sigma}{mn}\right)$ so;

$$T_f^2 = \frac{mn}{m+1} - \frac{1}{m(n-1)}\left(\overline{X_f} - \bar{\bar{X}}\right)^t S^{-1}\left(\overline{X_f} - \bar{\bar{X}}\right) \tag{9}$$

So that the equation of the Phase II in the T^2 control chart is expressed by:

$$UCL = \frac{p(m+1)(n-1)}{mn-m-p+1} F_{a,p,mn-m-p+1} \tag{10}$$

$$LCL = 0 \tag{11}$$

2.2. Copula

Copula is a function that join or couple multivariate distribution functions to their one-dimensional marginal distribution functions and as distribution functions whose one dimensional margins are uniform. But neither of these statements is a definition hence we will devote to giving a precise definition of copulas and to examining some of their elementary properties [8].

For any bivariate distribution function H (X, Y) with marginal distributions $G_1(X)$ and $G_2(Y)$, there exists a copula C: [0,1]2 → [0,1] see H (X, Y) =C($G_1(X)$, $G_2(Y)$) [9]. The remaining two copulas considered in this paper belong to a general class of symmetric copulas, named the Archimedean copulas. They are generated using a class ϕ of function $\phi : [0,1]\rightarrow[0,\infty]$, named generators, that have two continuous derivatives on (0,1) and fulfill the following conditions : $\varphi(1) = 1$. Every number of this class generates a multivariate distribution function. In this paper we consider two Archimedean copulas (see 12) and (13) with generator function (6) and (8) as follows.

- Clayton's

$$C(u,v) = (u^{-\theta} + \ v^{-\theta} - 1)^{\frac{1}{\theta}} \tag{12}$$

- Gumbel's

$$C(u,v) = \exp(-\ [(-ln\ u\ )^{\theta} +\ (-ln\ v)\ ^{\theta}\ ]\ )^{\frac{1}{\theta}} \tag{13}$$

3. Results and discussion

Product analysis in the bread production process is carried out through several stages of checking including checking data for each characteristic, namely defects, size and net. Figure 1, figure 2, and figure 3 explain about the characteristics of each data.

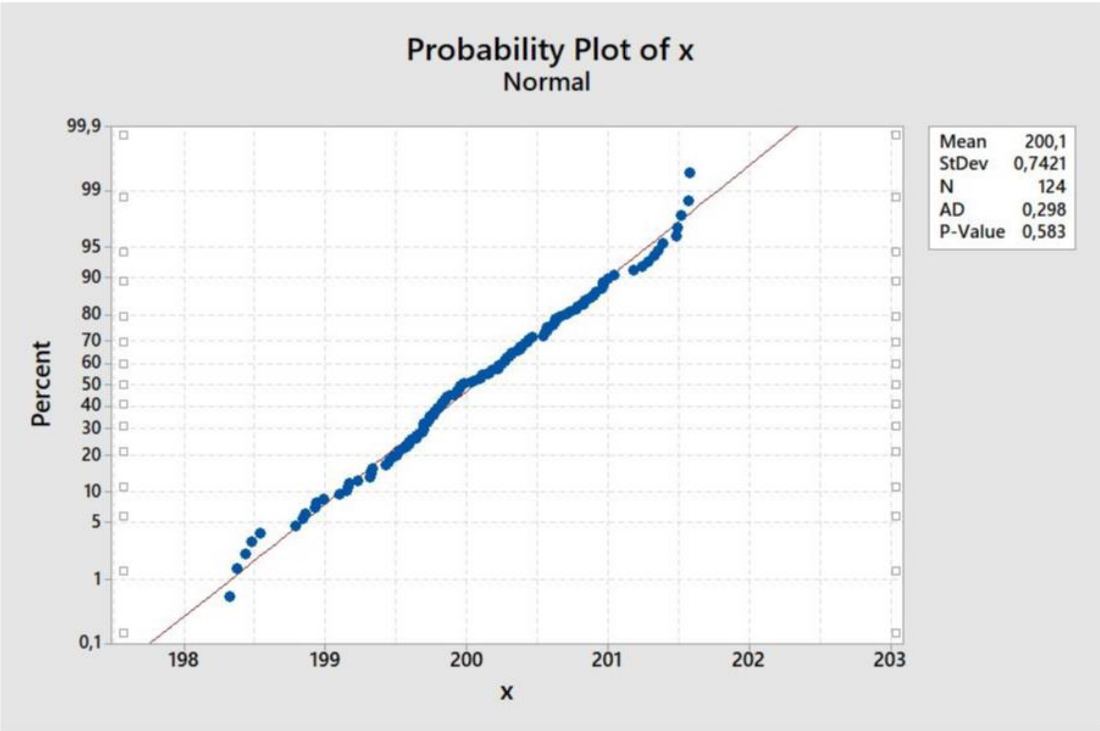

Figure 1. Normality test characteristic net.

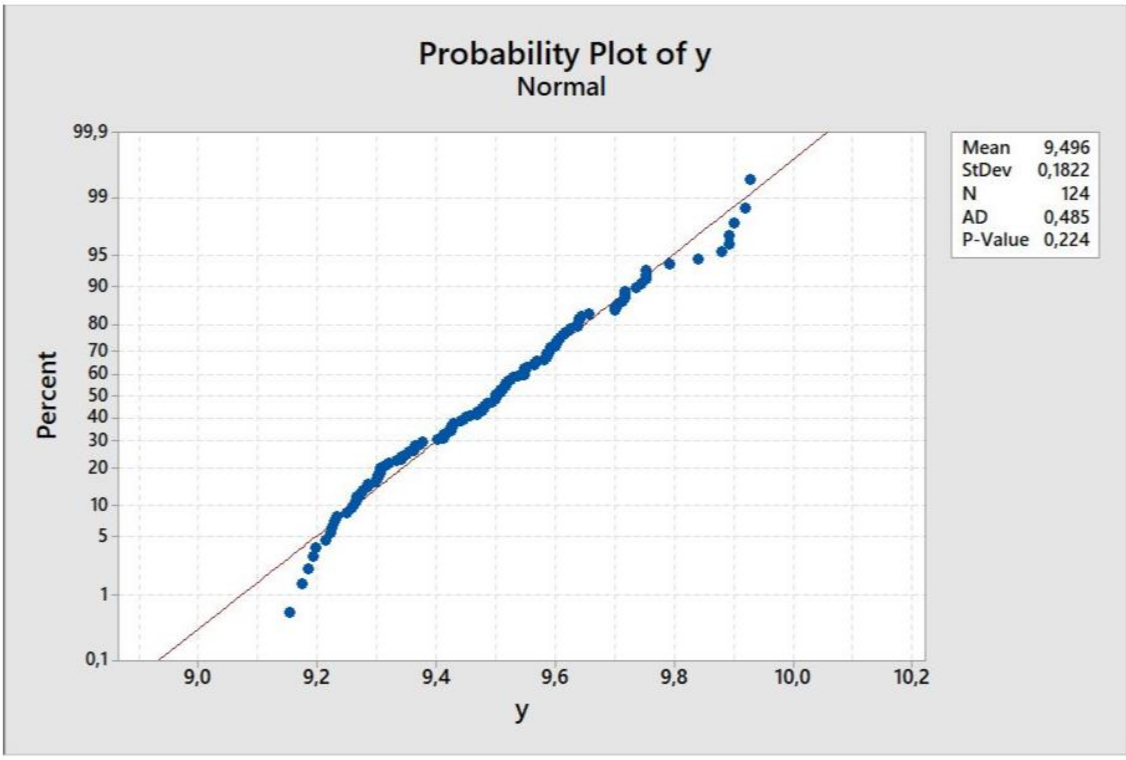

Figure 2. Normality test characteristic size.

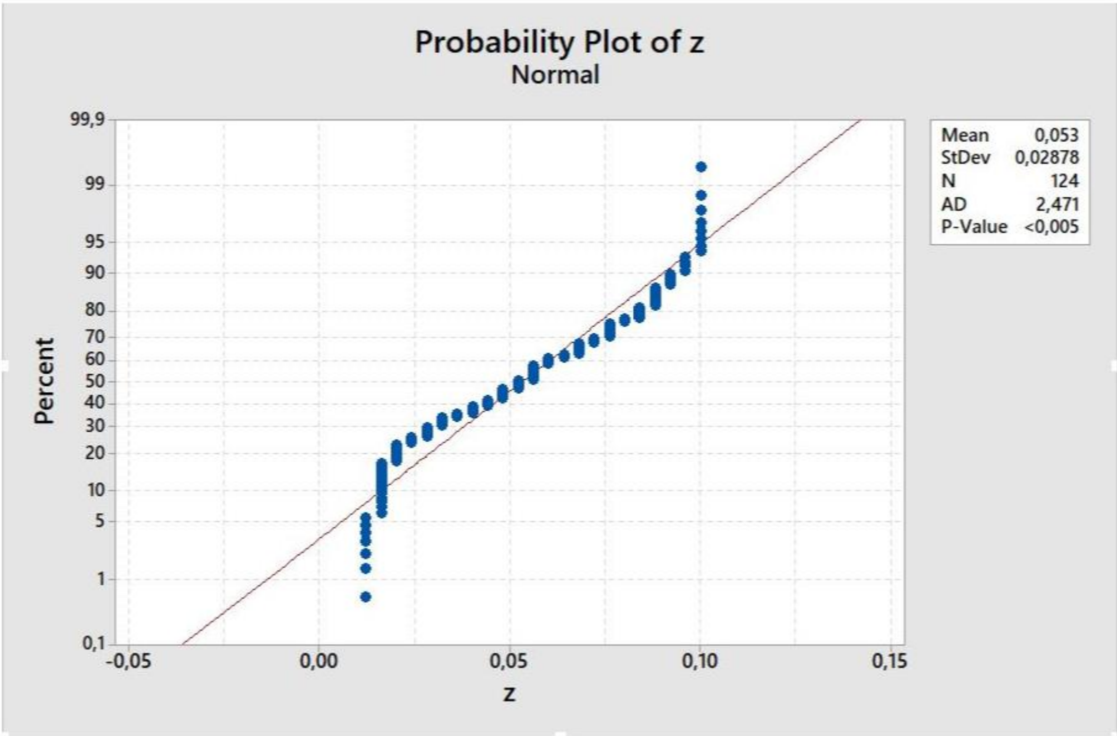

Figure 3. Normality test characteristic defect.

Figure 1 explain about how normal characteristic of net. Based on the picture characteristics of net shows that the data comes from data that are normally distributed, as well as for size characteristics. Figure 2 explain about how the normal characteristic of size where the data comes from normally distributed data is indicated by P-value 0.724, while for the next characteristic data, the defect turns out that the data does not originate from data that is not normally distributed. Figure 3 which explained how normal characteristics of defects are p-value <0.005. Based on test of normality in above pictures it shows that one of variable has a non-normal distribution, which is variable which characteristic is defect. Then, to see how quality control variables are in one process, we use hotelling T^2 as one of the methods in multivariate variable. The explanation will show in figure 4.

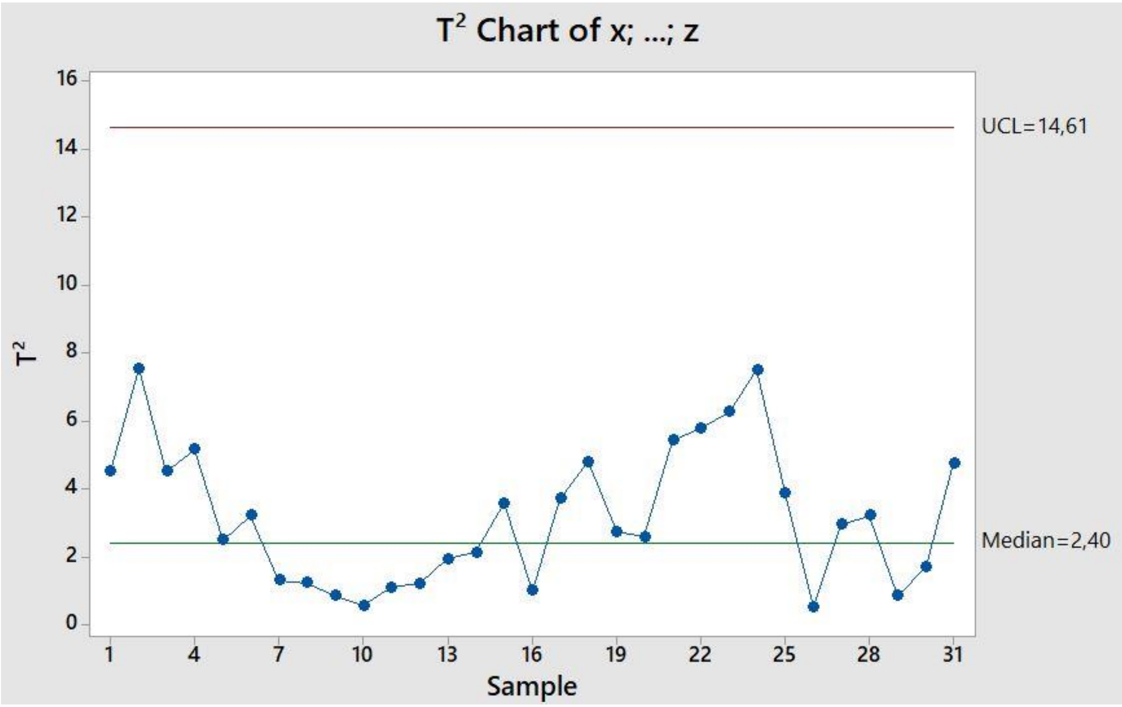

Figure 4. T^2 chart of variables.

Figure 4 explains that the bread production data in the in control condition does not have a sample out of control even though there are allegations that there are data that are not normally distributed, this is due to the copula role in establishing a distribution function control. The Hotelling T^2 probably the most used rule in industry for the problem of multivariate fault detection [10]. As explained in the previous section that copula has an important role in making control charts for cases of data that are not normally distributed, the point is that copula plays a role in combining the data so that when creating the control chart the data used becomes integrated into one normally distributed data. Based on that research we still use copula to analyse non-normal distribution observation. This can be seen when the data has been combined, the data can be seen in the Figure 5 and Figure 6.

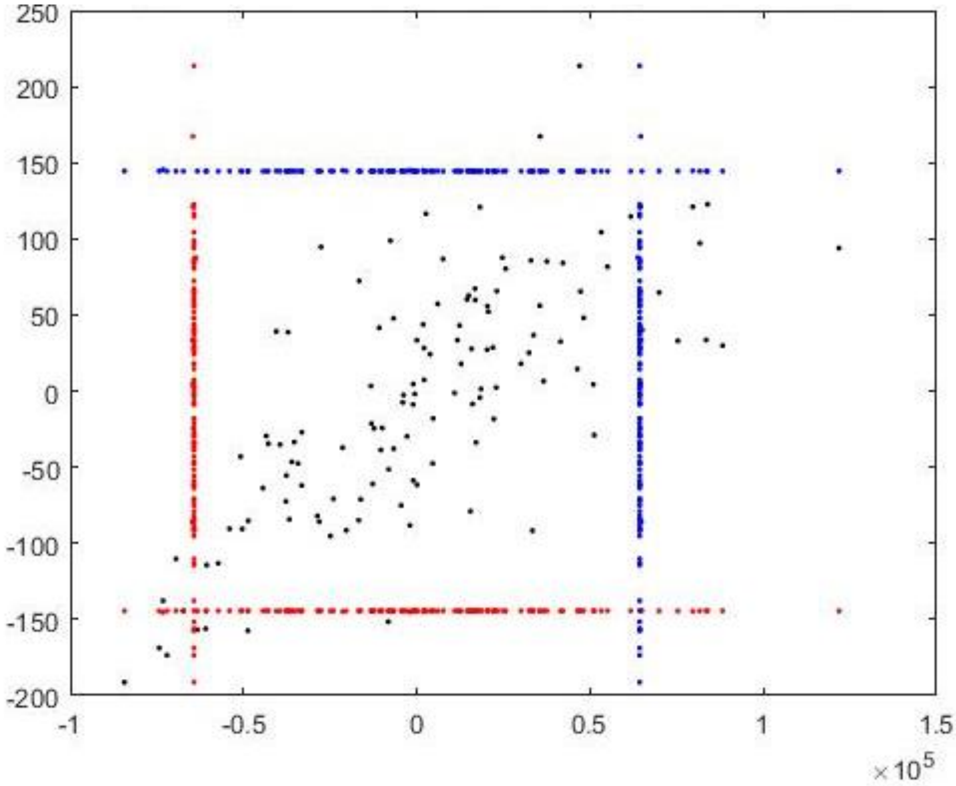

Figure 5. Control chart using copula clayton.

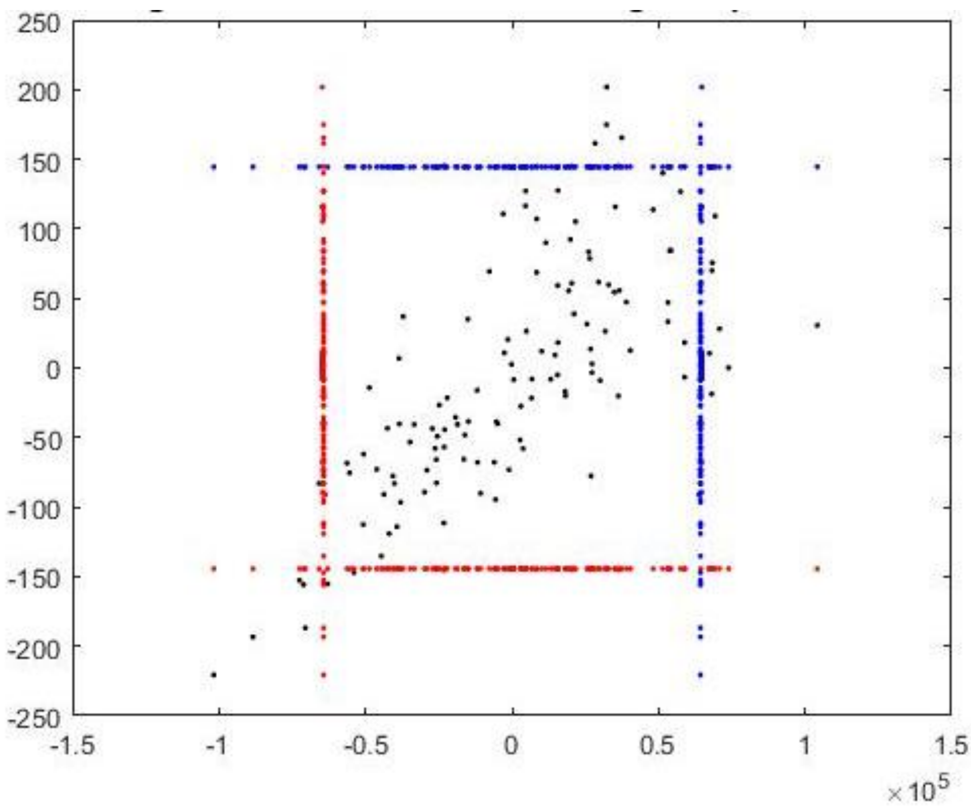

Figure 6. Control chart using copula gumbel.

4. Conclusion

In general, the control chart used for multivariate cases is a chart control of Hotelling T^2, but this control chart is not enough to be able to interpret the dependence between quality characteristics. Simulation results with studies the case shows some conclusions along with useful suggestions for further development and research. The results of making a multivariate control chart using Hotelling T^2. For simulation data and real data, it shows that the control chart is less able to interpret problems that involve more than one quality characteristic. While on the predictive control chart with Copula, the target behavior of the process by considering the existence of dependencies between visible quality characteristics.

References

[1] Durante F and Sempi C 2015 *Principles of Copula Theory,* (Florida United States: CRC Press, Taylor & Francis Group)
[2] Prins J and Mader D 1997 Multivariate Control Charts For Grouped And Individual Observations *Quality Engineering* **10** 49–57
[3] Serber G A F 2009 *Multivariate Observations* (United state of america : John Wiley & Sons)
[4] Hotelling H 1947 *Multivariate Quality Control Illustrated by Air Testing of Sample Bombsights*.In: Eisenhart, C., Hastay, M.W. and Wallis, W.A., Eds., Techniques of Statistical Analysis, McGraw Hill, New York, 111-184.
[5] Lestari T 2018 Analysis of quality controol using statistical process control (SPC) in bread production *Indonesian Journal of Fundamental Sciences* **4** 90-101
[6] Wiryawan 2019 Institutional Change and the Impact Towards Innovation Competitiveness in the Industrial Development of The Batam Free Trade Zone *Indonesian Journal of Computing, Engineering and Design* (1) 9–16.
[7] Montgomery D *2009 Introduction to Statistical Quality Control*, (United State of America : John Wiley & Sons)
[8] Lestari T 2015 Bivariate Control Chart With Copula *AIP Conference Proceedings* 1692 020017
[9] Nelsen 2005 *Introduction to Copulas* Springer
[10] Verdier G 2013 Application of Copulas to Multivariate Control Charts *Journal of Statistical Planning and Inference* **143** 2151 - 2159