Original Research

Robust Cluster-Based Method for Monitoring Generalized Linear Profiles in Phase I

Davood Saremian[1] . Rassoul Noorossana[2*] . Sadigh Raissi[1] . Paria Soleimani[1]

Received: 2021-01-16 / Accepted: 2021-07-06 / Published online: 2021-08-21

* Corresponding Author, rassoul@iust.ac.ir
1- Department of Industrial Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran
2- Department of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran

Abstract
Profile monitoring is one of the new statistical quality control methods used to evaluate the functional relationship between the descriptive and response variables to measure the process quality. Most of the studies in this field concern processes whose response variables follow the normal distribution function, but in many industries and services, this assumption is not true. The presence of outliers in the historical data set could have a deleterious effect on phase I parameter estimation. Therefore, in this paper, we propose a robust cluster-based method for estimating the parameters of generalized linear profiles in phase I. In this method, the effect of data contamination on estimating the generalized linear model parameters is reduced and as a result, the performance of $\boldsymbol{T^2}$ control charts is improved. The performance of this method has been evaluated for two specific modes of generalized linear profiles, including logistic and Poisson profiles, based on a step shift. The simulation results indicate the superiority of this cluster-based method in comparison to the non-clustering method and provide a more accurate estimation of the parameters.

Keywords - Generalized Linear Models; Phase I; Hotelling T^2; Clustering; Robust Technique.

1. INTRODUCTION

Monitoring the quality and stability of products and services in many manufacturing and service industries is done by examining the relationship between a response variable and one or more descriptive variables. This relationship is called a profile in the literature of statistical process control. Profile monitoring such as conventional methods is typically carried out in two phases, and each of them pursues different goals. The main purpose of Phase I is to check the stability of the process and estimating of the parameters of control charts based on in-control profiles. In Phase II, the changes in process parameters have to be identified quickly and accurately using the estimated parameters in Phase I. Profile monitoring studies are classified into linear and non-linear profiles.

Linear profiles were studied by researchers such as Soleymani et al. (2013), Kang and Albin (2000), Noorossana et al. (2010), Chen and Nambahard (2011), Aly et al. (2015), and nonlinear profiles were studied by other researchers such as Walker and Wright (2002), Nikoo and Noorossana (2013), Li et al.(2018) and pan et al.(2019). Most of these studies assume a normal distribution function and a continuous response variable. Preliminary study on the monitoring of generalized linear profiles have been performed by Yeh et al. (2009) using five Hotelling T^2 control charts. Amiri and Koosha (2011) studied the performance of the T^2 control chart for monitoring logistic profiles in the presence of autocorrelation. Koosha and Amiri (2012) presented two methods including a generalized linear mixed model and modification of upper control limit to consider autocorrelation in logistic regression. Payneber et al. (2012) proposed an LRT-based change point approach for monitoring binary profiles in phase I. Shadman et al. (2014, 2015) proposed a control chart based on the change point for monitoring generalized linear profiles in phases I and II. Amiri et al. (2014) proposed three methods including SLRT, F, T_I^2 for monitoring GLM profiles. Izadbakhsh et al. (2018) proposed four methods for monitoring multinomial logistics profiles in phase I. If there are any outliers in the historical data set, these data are likely to affect the accuracy of the control chart parameters. Bahiraee (2014) proposed an economic design of Hotelling's T^2 control chart on the presence of fixed sampling rate and exponentially assignable causes. To overcome this problem, various methods based on robust estimation have been proposed. Hakimi et al. (2017) proposed three methods including RM, WRM, and WMLE to reduce the effect of outlier data on the estimation of logistic regression parameters. Moheghi et al. (2020) used the C-R robust estimation method proposed by Cantoni and Ronchetti (2001) to estimate the parameters of GLM profiles. Therefore, in this paper, a cluster-based method for monitoring of GLM profiles in phase I is developed. The purpose is to provide a more accurate estimate of control chart parameters for these types of profiles in the presence of outliers. The paper is structured as follows. In Section 2, the GLM parameter estimation is briefly reviewed. In Section 3, two control charts for monitoring logistic and Poisson profiles are presented. In Section 4, the proposed cluster-based method for monitoring GLM profiles is presented. In Section 5, the performance of the cluster-based method versus the non-cluster method is evaluated based on Six criteria. The estimation of logistic and Poisson profile parameters is presented based on the maximum likelihood estimation method and the cluster-based maximum likelihood estimation method. Conclusions and suggestions for future research are provided in Section 6.

1. GENERALIZED LINEAR MODELS PARAMETER ESTIMATION

Generalized linear models provide an integrated approach to modeling a variety of continuous and discrete response variables such as normal, binary, nominal, proportion, count, and ordinal, and therefore include exponential distributions such as normal, Poisson, gamma, and binomial distributions. According to Olsson(2002), in these models, instead of directly modeling μ = E (y) as a function of the linear predictor variables Xβ, certain functions g(μ) are used to relate to the predictor variables as follows:

$$g(\mu) = \eta = X\beta \tag{1}$$

The g (·) function is called the link function, and X is a vector of predictor variables. There is a number of link functions, some of the most common are logit, probit, and log-log. Parameter estimation of GLM is mostly performed using the maximum likelihood method as follows:

$$\iota = \log[(\theta,\phi;y)] = \frac{y\theta-b(\theta)}{a(\phi)} + c(y,\phi) \tag{2}$$

where $a(\cdot)$, $b(\cdot)$, and $c(\cdot)$ are specific functions. The parameter θ is a location parameter and ϕ is a dispersion parameter of the distribution.

The model parameters, β is a vector p × 1 of the regression coefficients, which is a function of θ. Estimation of model coefficients is obtained by deriving of ι with respect to β as Eq. 3.

$$\frac{\partial\iota}{\partial\beta_j} = \frac{\partial\iota}{\partial\theta}\frac{\partial\theta}{d\mu}\frac{d\mu}{d\eta}\frac{d\eta}{\partial\beta_j} \tag{3}$$

The value of $\frac{\partial\mu}{\partial\theta} = V$, based on the linear relationship $\frac{\partial\mu}{\partial\beta_j} = x_j$ and $W^{-1} = (\frac{d\mu}{d\eta})^2$V, Therefore

$$\frac{\partial\iota}{\partial\beta_j} = \frac{(y-\mu)}{a(\phi)}\frac{1}{V}\frac{d\mu}{d\eta}x_j = \frac{W}{a(\phi)}(y-\mu)\frac{d\eta}{d\mu}x_j \tag{4}$$

So far, the likelihood function has been provided for an observation. By summing over the observations, the probability function for the parameter β_j is as follows:

$$\sum_i \frac{W_i}{a(\phi)}(y_i-\mu_i)\frac{d\eta_i}{d\mu_i}x_{ij}=0 \tag{5}$$

Numerical methods such as the iteratively reweighted least squares suggested by Mc Cullagh and Nelder in 1989 is used to solve this equation.

CONTROL CHARTS FOR MONITORING LOGISTIC AND POISSON PROFILES

In this paper, we used two control charts, including T_I^2 and T_R^2 from the five control charts introduced by Yeh et al. (2009) that performed better than the others for monitoring the logistic profiles. These charts are described as follows:

T^2 based on sample mean and covariance matrix

The control chart statistic and mean and covariance matrix are calculated as follows:

$$T_{I,t}^2 = (\hat{\beta}_t - \bar{\beta})^T S_I^{-1}(\hat{\beta}_t - \bar{\beta}) \tag{6}$$

$\bar{\beta} = 1/k\sum_{t=1}^{k}\widehat{\beta}_t$ and $S_I = \frac{1}{k}\sum_{t=1}^{k} var(\widehat{\beta}_t) = \frac{1}{k}\sum_{t=1}^{k}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}$

T^2 based on sample mean and difference between successive observations

The control chart statistic and mean and covariance matrix are calculated as follows:

$$T_{R,t}^2 = (\hat{\beta}_t - \bar{\beta})^T S_R^{-1}(\hat{\beta}_t - \bar{\beta}) \tag{7}$$

$S_R = \frac{1}{2(k-1)}\sum_{t=1}^{k-1}(\hat{\beta}_{t+1} - \hat{\beta}_t)(\hat{\beta}_{t+1} - \hat{\beta}_t)^T$ and $\bar{\beta} = 1/k\sum_{t=1}^{k}\widehat{\beta}_t$

In this control chart, the covariance matrix is calculated on the basis of the difference between the successive observations.

1. THE PROPOSED PROCEDURE TO DESIGN THE CLUSTER-BASED CONTROL CHART

The cluster-based control charts approach for monitoring of GLM profiles in Phase I is presented in this section. The flow chart of the proposed method is shown in Figure 1.

In first step, according to the type of data patterns, an appropriate regression model is fitted to each profile and its parameters ($\widehat{\mathrm{B}}$) are estimated. Based on the estimated parameters, the variance-covariance matrix($\hat{V}_D$) and subsequently, the similarity matrix is calculated as Eq 8.

$$S_{ij} = (\hat{\beta}_i - \hat{\beta}_j)^T\hat{V}_D^{-1}(\hat{\beta}_i - \hat{\beta}_j) \tag{8}$$

Where, $\hat{\beta}_i$ and $\hat{\beta}_j$ are different rows of the estimated parameter vector.

Profiles are clustered according to the calculated similarity matrix and using a suitable link function. The clustering continues until at least half of the profiles are in one cluster and set this cluster as the main cluster. T^2 statistic for the profiles outside the main cluster is as follows:

$$T_i^2 = (\hat{\beta}_i\_\hat{\beta}_{MC})^T \hat{V}_D^{-1}(\hat{\beta}_i\_\hat{\beta}_{MC}) \quad (9)$$

Where $\hat{\beta}_{MC}$ is the estimated parameter of the profile in the main cluster and $\hat{\beta}_i$ is the i^th profile that is outside the main cluster.

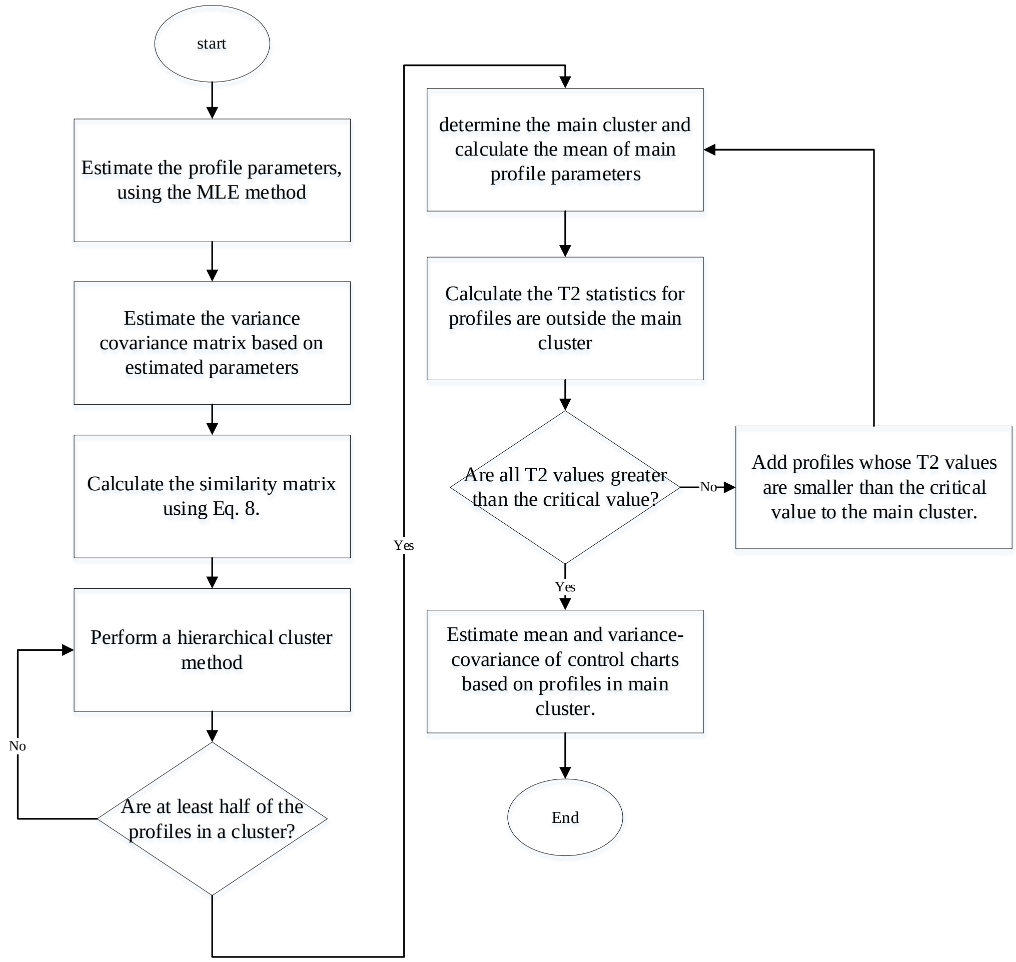

FIGURE 1.

FLOW CHART OF THE CLUSTER-BASED CONTROL CHART

If the calculated T^2 value is smaller than the critical value, the profile is added to the main cluster and the mean vector and the variance-covariance matrix of the new main cluster is updated. Then, the T^2 value is recalculated for the profiles outside the new main cluster and is compared with the critical value. The described process is repeated until it is no possible to add a profile to the main cluster. After this, the profiles in the main cluster are considered as in-control profiles and the profiles that are not in this cluster are considered as out-of-control profiles. Finally, based on the profiles in the main cluster, the mean vector and the variance-covariance matrix is calculated. Shiau and Sun (2009) and Chen et al. (2015) proposed a similar algorithm for multivariate control charts and linear mixed models, respectively.

Case study

We considered a process, adapted from Montgomery et al. (2006), to investigate the compressive stress of the alloy fastener used in the construction of aircraft structures. A number of these fasteners are tested at 10 pressure levels in the range of 2500 to 4300 psi. The response variable is the sum of fasteners that are broken at each given pressure level. Accordingly, the logistic regression model is fitted to these data and the parameters are estimated as β = (-42.1110, 5.1772). Based on estimated parameters, we simulated 20 logistic profiles that 4 of them experienced a step shift($\beta_0 + \delta_1\sigma_1$ and $\beta_1 + \delta_2\sigma_2$) at samples 4, 8, 12, and 16 where $\delta_1 = 0.16$ and $\delta_2 = 0$. The variance-covariance matrix is equal to $\begin{pmatrix} 18.5689 & -2.2833 \\ -2.2833 & 0.2809 \end{pmatrix}$.

The steps for implementing the algorithm are presented as follows:

STEP 1: The logistic regression is fitted to each of these 20 profiles and the estimated parameters by using Eq.5 are provided in Table 1.

STEP 2: The variance-covariance matrix corresponding to the estimated parameters is estimated as $\hat{V}_D = \begin{pmatrix} 42.3065 & -5.2038 \\ -5.2038 & 0.6404 \end{pmatrix}$

TABLE 1.
ESTIMATED PARAMETERS OF 20 LOGISTICS PROFILES

Profile	β_0	β_1	Profile	β_0	β_1
1	-47.787	5.866	11	-33.348	4.109
2	-32.099	3.957	12	-42.780	5.321
3	-37.843	4.670	13	-37.937	4.639
4	-53.665	6.690	14	-48.981	6.038
5	-41.093	5.062	15	-39.944	4.896
6	-38.473	4.731	16	-42.813	5.366
7	-33.072	4.063	17	-42.882	5.251
8	-45.507	5.682	18	-27.854	3.401
9	-32.318	3.970	19	-37.290	4.611
10	-44.726	5.508	20	-39.729	4.896

STEP 3: Based on the variance-covariance matrix, the similarity matrix (Eq.8) is calculated and presented in Table 2.

STEP 4: According to the calculated similarity matrix and using a complete linkage function, the profiles are clustered.

STEP 5: As shown in the dendrogram (Figure 1), profiles 1, 2, 3, 5, 7, 9, 10, 11, 14, 19, and 20 are in a cluster hence, we stopped clustering and considered these profiles to be the primary main cluster. Therefore, the calculated mean vector is equal to (-38.8966, 4.7901). As shown in Table 3, the T^2 values are calculated for the profiles outside the main cluster.

TABLE 3.
THE T^2 VALUES FOR PROFILES ARE OUTSIDE THE PRIMARY MAIN CLUSTER

Profile	4	8	12	13	15	16	17	18
T^2	30.332	23.2531	10.5723	3.8944	1.8558	32.2817	3.5854	6.2509

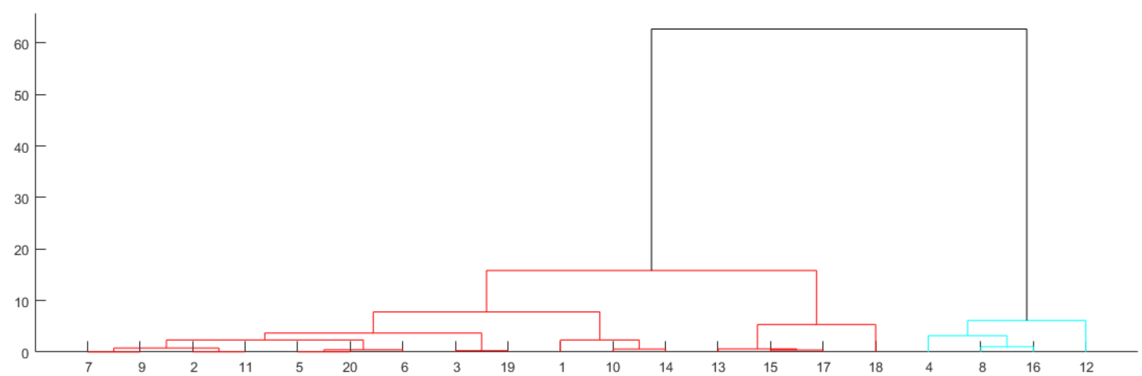

FIGURE 1.
CLUSTERING DENDROGRAM OF LOGISTICS PROFILES

These values compared to the upper control limit, 10.2132. Because the T^2 values for profiles 13, 15, 16, 17, and 18 are smaller than UCL, they are added to the main cluster. According to the new main cluster, the mean of the main cluster is updated and T^2 is recalculated for the profiles outside the new main cluster.

TABLE 4
THE T^2 VALUES FOR PROFILES ARE OUTSIDE THE SECONDARY MAIN CLUSTER

Profile	4	8	12	16
T^2	35.187	27.676	13.621	37.462

Since all the T^2 values are greater than the upper control limit, the algorithm is stopped and the profile cannot be added to the main cluster. Profiles in the main cluster are considered as in control profiles and the profiles 4, 8, 12, and 16 that are not in this cluster, hence are considered as out-of-control profiles. Results show that the cluster-based approach has performed properly in classifying in-control and out-of-control profiles.

2. PERFORMANCE EVALUATION AND COMPARISON OF CONTROL CHARTS

Comparison of the cluster-based method versus the non-cluster based method

In this section, using Monte Carlo simulations, the performance of CB and NCB methods are evaluated based on the probability of signal (POS), Fraction correctly classified (FCC), sensitivity, specificity, False positive rate (FPR), False negative rate (FNR) for logistic and Poisson profiles in phase I. Due to the complicated computation of the distribution of the T^2 statistic, control limits are calculated by 100,000 simulations according to an overall false alarm probability of 0.05 (Vargas (2003), Williams et al. (2006), Jensen et al.(2008)). T^2 statistics for each of k independent samples with m replications were calculated and selected maximum T^2 in each sample as T^2_{max}. The 95th quantile of these simulated T^2_{max} was considered as upper control limit estimation.

logistic profiles

We adopt the same simulation settings used in Yeh et al. (2009) and therefore assume that $p = 2$ and the in-control parameter $\boldsymbol{\beta}_0 = (3,2)^T$, and the design matrix is equal to:

$$X = \begin{pmatrix} 1 & 2 & \cdots & 1 & 1 \\ \log(0.1) & \log(0.2) & \cdots & \log(0.8) & \log(0.9) \end{pmatrix}^T$$

The simulation study is performed for the different combinations of sample size (m = 30, 50,100) and replication (k = 30, 60, 90). The upper control limits for CB and NCB methods are presented in Table 5.

Poisson profiles

In the case of Poisson profiles, we adopt the same simulation settings used in Amiri et al. (2014) and therefore assume that $p = 2$ and the in-control parameter $\boldsymbol{\beta}_0 = (3,2)^T$, and the design matrix is equal to:

$$X = \begin{pmatrix} 1 & 2 & \cdots & 1 & 1 \\ 0.1 & 0.2 & \cdots & 0.8 & 0.9 \end{pmatrix}^T$$

For each control chart, k = 30, 60, 90 independent samples are considered and the upper control limits for CB and NCB methods are presented in Table 5.

These results show that the estimated upper control limits for CB are smaller than the NCB methods, and these control limits increase when the sample size increases.

Step shifts

In this section, the effect of creating step shifts on the parameters of logistic and Poisson profiles is studied. The β_0 and β_1 parameters are changed to $\beta_0 + \delta_1\sigma_1$ and $\beta_1 + \delta_2\sigma_2$ and the variance-covariance matrix is calculated as $\boldsymbol{\Sigma}_0 = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$. The variance-covariance matrix in logistic and Poisson profiles is equal to $\begin{pmatrix} 0.0141 & -0.0196 \\ -0.0196 & 0.0314 \end{pmatrix}$ and $\begin{pmatrix} 0.0651 & 0.0449 \\ 0.0449 & 0.0412 \end{pmatrix}$, respectively. We assume that the step shift occurs in the last third of the profiles. Performance of T_I^2 and T_R^2 in CB and NCB was evaluated on basis of six criteria. Some of the results, including the logistic profiles with m = 30 and k = 60 and Poisson profiles with k = 60, are depicted in Figures 2 to 13. Other simulations are compatible with the reported results and are available with the authors.

- *TABLE 5.*
- *THE SIMULATED UPPER CONTROL LIMIT FOR THE CB AND NCB CONTROL CHARTS*

Sample size (k)	Replication(m)	Logistic				Poisson			
		T_R^2		T_I^2		T_R^2		T_I^2	
		NCB	CB	NCB	CB	NCB	CB	NCB	CB
30	30	13.495	12.747	15.080	12.652	12.499	12.210	12.664	10.998
	50	12.998	12.544	13.935	11.882				
	100	12.642	12.355	13.113	11.302				
60	30	16.216	15.450	17.977	15.481	14.119	13.867	14.384	12.843
	50	15.263	14.607	16.313	14.346				
	100	14.565	14.042	15.106	13.488				
90	30	18.052	17.109	19.614	17.096	15.057	14.865	15.399	14.066
	50	16.665	16.056	17.626	15.675				
	100	15.688	15.230	16.244	14.677				

Evaluation of control charts based on the probability of signal (POS) index

POS of CB and NCB for T_I^2 and T_R^2 are computed under different step shifts and some of the results are shown in Figures 2 and 3. The simulation results indicate that the performance of both control charts is better in the CB than the NCB method and the CB for T_I^2 performs better than the other three control charts.

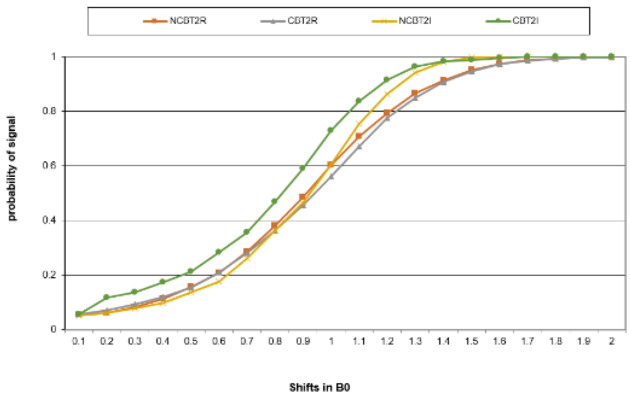

FIGURE 2.
POS FOR STEP SHIFT IN LOGISTICS PROFILES

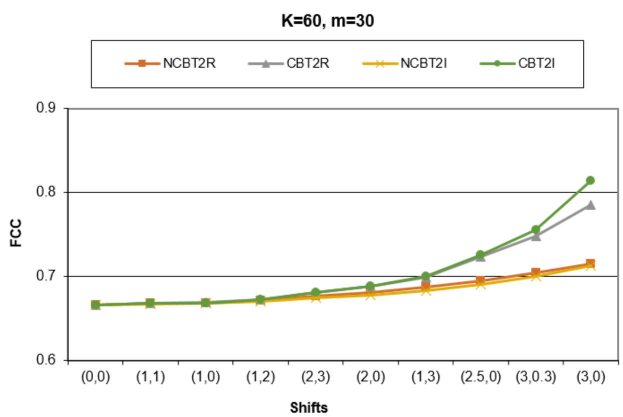

FIGURE 3.
POS FOR STEP SHIFT IN POISSON PROFILES

Evaluation of the control charts based on fraction correctly classified (FCC) index

FCC displays the ratio of properly detected the out-of-control and the in-control profiles to the total number of profiles. The simulation results show that CB for T_R^2 performs much better than NCB for T_R^2. This index has almost the same results for the small to medium shifts for the T_I^2 control chart in both clustering and non-clustering scenarios, but in medium to large shifts, CB performs better and in general, CB for T_I^2 performs better than the other control charts.

Evaluation of the control charts based on the sensitivity index

- The ratio of profiles identified correctly as the out-of-control to the total number of out-of-control profiles is measured by this index. The Simulation results show that the values of this index for T_R^2 and T_I^2 are increased by increasing the number of shifts, and in all the simulation results, CB is superior to NCB control charts

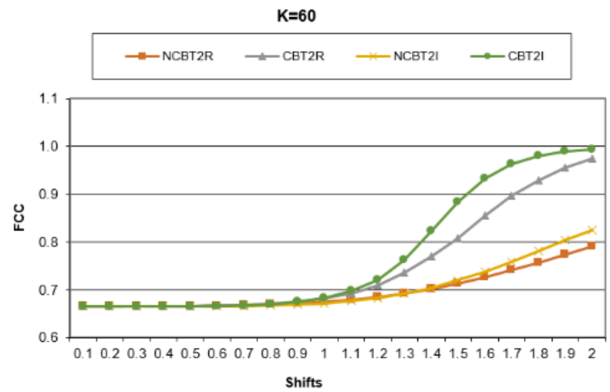

FIGURE 4.
FCC FOR STEP SHIFT IN LOGISTICS PROFILES

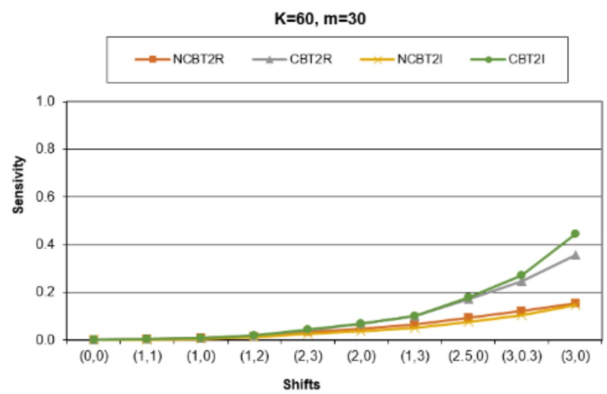

FIGURE 5.
FCC FOR STEP SHIFT IN POISSON PROFILES

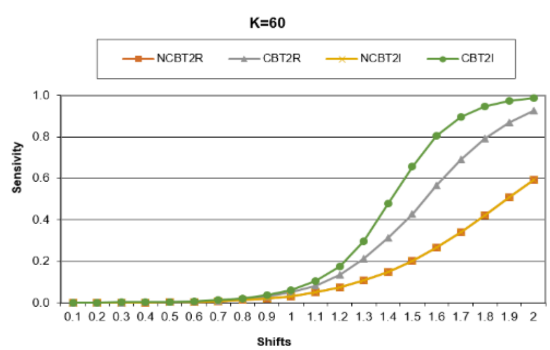

FIGURE 7.
SENSITIVITY FOR STEP SHIFT IN POISSON PROFILES

Evaluation of the control charts based on the specificity index

This index indicates the ratio between the correctly identified in-control profiles and the total number of in-control profiles. The simulation results indicate that CB and NCB for T_R^2 and T_I^2 have proper performance, but CB performs better than NCB control charts in medium to large shifts.

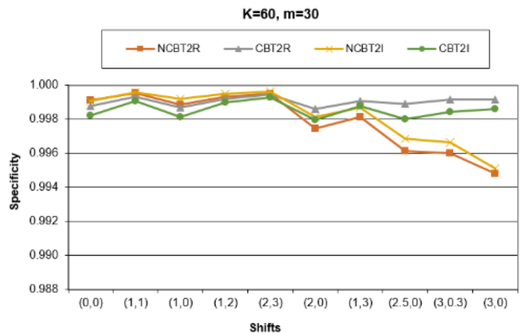

FIGURE 8.
SPECIFICITY FOR STEP SHIFT IN LOGISTICS PROFILES

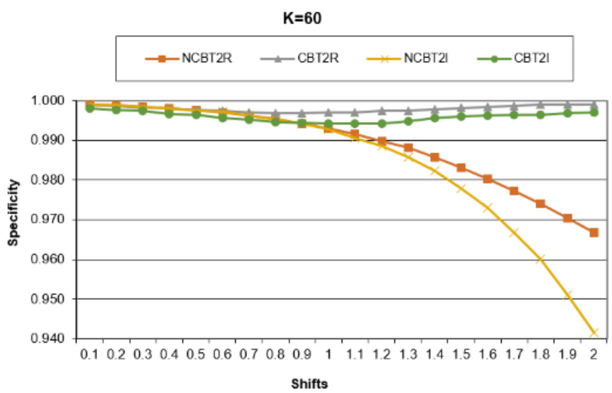

FIGURE 9.
SPECIFICITY FOR STEP SHIFT IN POISSON PROFILES

Evaluation of control charts based on the false-positive rate (FPR) index

This index indicates the ratio of profiles that were misidentified as out of control to the total number of out-of-control profiles. Simulation results show that with increasing the amount of shifts, this index has a downward trend in both CB and NCB control charts. In general, CB for T_I^2 has an appropriate performance compared to the other control charts.

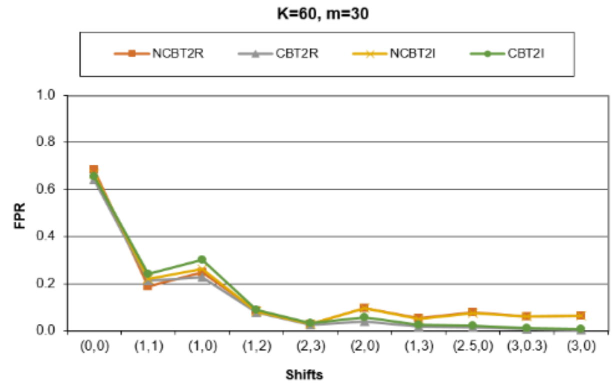

FIGURE 10.
FPR FOR STEP SHIFT IN LOGISTICS PROFILES

•

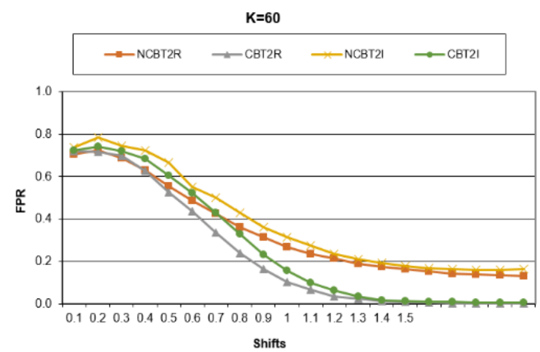

FIGURE 11.
FPR FOR STEP SHIFT IN POISSON PROFILES

Evaluation of control charts based on the false-negative rate (FNR) index

This index shows the ratio of false identified in-control profiles to the total number of in-control profiles. The simulation results show a downward trend in both CB and NCB control charts. In general, CB control charts perform better than the NCB control charts.

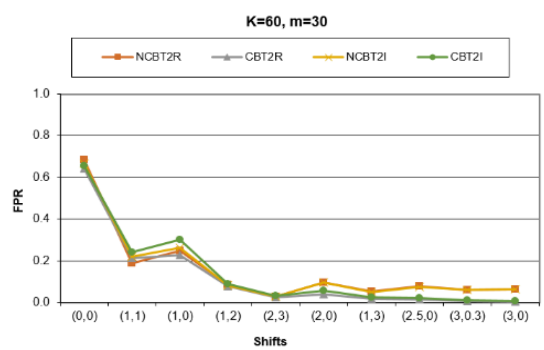

FIGURE 12.
FNR FOR STEP SHIFT IN LOGISTICS PROFILES

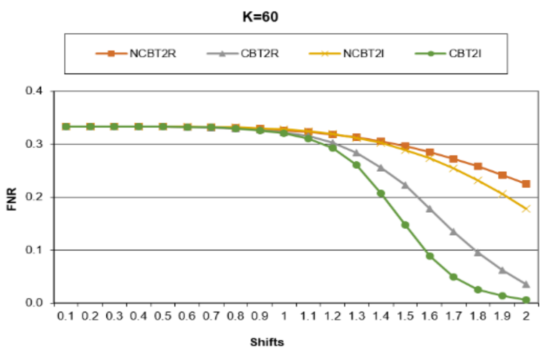

FIGURE 13.
FNR FOR STEP SHIFT IN POISSON PROFILES

Comparison of parameter estimation in CB and NCB methods

In this section, the parameters of logistic and Poisson profiles are estimated based on maximum likelihood estimation (MLE) and cluster-based maximum likelihood estimation (CB MLE). Simulations are calculated based on creating the step shift in the last third of profiles and 10,000 runs. The results in Tables 6 to 9 show that by increasing the amount of shift to 0.5 and 0.6 in the β_0 and β_1 of logistic and Poisson profiles, in the CB MLE, the estimates of parameters and standard deviation has an ascending and then a descending trend. However, in the case of the NCB MLE, with increasing the shift value to 1 and 0.6 for the β_0 and β_1,respectively, the estimation of the parameters has an ascending and then descending trend, but their standard deviation increases with increasing the shift value. These results show that the estimated parameters based on both methods are somewhat biased and the bias of the MLE is higher than the CB MLE. The main reason for this is properly identifying the in control profiles in CB MLE than the MLE. In general, the results indicate that the CB MLE provides a more accurate estimate of the parameters of logistic and Poisson profiles than the NCB MLE concerning the applied step shifts.

TABLE 6.
ESTIMATION OF PARAMETER β_0 AND ITS STANDARD DEVIATION IN LOGISTIC PROFILE FOR TWO METHODS
- *BASED ON CLUSTERING AND NON-CLUSTERING*

Shift	Method	β_1	SD β_1	Shift	Method	β_1	SD β_1
0	MLE	3.040	0.337	1.5	MLE	3.139	0.510
	CB MLE	**3.039**	**0.336**		CB MLE	**3.040**	**0.337**
0.5	MLE	3.	0.358	2	MLE	3.059	0.490
	CB MLE	**3.189**	**0.352**		CB MLE	**3.040**	**0.336**
1	MLE	3.247	0.410	2.5	MLE	3.042	0.490
	CB MLE	**3.059**	**0.339**		CB MLE	**3.041**	**0.336**

TABLE 7.
ESTIMATION OF PARAMETER β_1 AND ITS STANDARD DEVIATION IN LOGISTIC PROFILE FOR TWO METHODS
- *BASED ON CLUSTERING AND NON-CLUSTERING*

Shift	Method	β_1	SD β_1	Shift	Method	β_1	SD β_1
0	MLE	2.031	0.268	0.9	MLE	2.188	0.330
	CB MLE	**2.030**	**0.268**		CB MLE	**2.037**	**0.268**
0.3	MLE	2.130	0.278	1.2	MLE	2.117	0.376
	CB MLE	**2.052**	**0.276**		CB MLE	**2.032**	**0.268**
0.6	MLE	2.198	0.296	1.5	MLE	2.060	0.421
	CB MLE	**2.141**	**0.278**		CB MLE	**2.032**	**0.268**

TABLE 8.
ESTIMATION OF PARAMETER β_0 AND ITS STANDARD DEVIATION IN POISSON PROFILE FOR TWO METHODS
- *BASED ON CLUSTERING AND NON-CLUSTERING*

Shift	Method	β_1	SD β_1	Shift	Method	β_1	SD β_1
0	MLE	3.003	0.206	1.5	MLE	3.115	0.334
	CB MLE	**3.002**	**0.205**		CB MLE	**3.002**	**0.205**
0.5	MLE	3.166	**0.194**	2	MLE	3.125	0.334
	CB MLE	**3.096**	0.197		CB MLE	**3.003**	**0.205**
1	MLE	3.040	0.241	2.5	MLE	3.105	0.334
	CB MLE	**3.003**	**0.205**		CB MLE	**3.002**	**0.205**

TABLE 9.
ESTIMATION OF PARAMETER β_1 AND ITS STANDARD DEVIATION IN POISSON PROFILE FOR TWO METHODS
BASED ON CLUSTERING AND NON-CLUSTERING

Shift	Method	β_1	SD β_1	Shift	Method	β_1	SD β_1
0	MLE	2.052	0.415	0.9	MLE	2.330	0.507
	CB MLE	**2.048**	**0.413**		CB MLE	**2.285**	**0.476**
0.3	MLE	2.154	0.440	1.2	MLE	2.400	0.550
	CB MLE	**2.146**	**0.436**		CB MLE	**2.303**	**0.485**
0.6	MLE	2.247	0.470	1.5	MLE	2.458	0.598
	CB MLE	**2.229**	**0.458**		CB MLE	**2.276**	**0.481**

3. Conclusion

In this paper, a cluster-based method was developed for monitoring GLM profiles in Phase I. This method is based on combining the maximum likelihood estimation with the hierarchical clustering method. In this method, at first, the estimated parameters are clustered and divided into two main and non-main clusters. The T^2 value for the profiles outside the main cluster is calculated and compared with the critical value. If the value of T^2 statistics is less than the critical value, these profiles are added to the main cluster, and this algorithm is repeated until it is no possible to add a profile to the main cluster. The profiles in the main cluster are identified as the in-control profiles and the profiles outside this cluster are identified as the out-of-control profiles. Based on the parameters in the main cluster, control charts are created. The proposed method provides a robust estimate for GLM profiles and reduces the effects of outgoing data on the estimating control chart parameters accurately. The performance of this CB and NCB methods under step shift was evaluated based on the six criteria including POS, FCC, sensitivity, specificity, FPR, FNR. The simulation results show that T_I^2 and T_R^2 have proper performance in the clustering method for monitoring Poisson and logistic profiles. Nevertheless, the CB of T_I^2 control chart performs better than the other methods. The simulation results also show that the parameters that have been estimated by CB MLE are more accurate than the usual MLE. For future research, it is suggested to evaluate the performance of the proposed method for the other shifts such as drift shift, etc., as well as to evaluate the effect of these estimations in Phase II of the generalized linear models.

References

[1] Amiri A, Koosha, M., Azhdari & Aand Wang G. (2014). Phase I monitoring of generalized linear model-based regression profiles. Journal of Statistical Computation and Simulation, 85(14), 2839-2859. doi:10.1080/00949655.2014.942864

[2] Aly, A. A., Mahmoud, M. A., & Woodall, W. H. (2014). A Comparison of the Performance of Phase II Simple Linear Profile Control Charts when Parameters are estimated. *Communications in Statistics - Simulation and Computation*, 44(6), 1432–1440. doi:10.1080/03610918.2013.821484

[3] Bahiraee Ehsan, Raissi Sadigh, (2014), Economic design of Hotelling's T^2 control chart on the presence of fixed sampling rate and exponentially assignable causes, Journal of Industrial Engineering International, a Springer Journal, December 2014, Volume 10, Issue 4, pp 229–238, ISSN: 1735-5702.

[4] Cantoni, E., & Ronchetti, E. (2001). Robust inference for generalized linear models. *Journal of the American Statistical Association, 96*(455), 1022-1030.

[5] Change Point Method for Monitoring Generalized Linear Profiles in Phase I. *Quality and Reliability Engineering International*, 31(8), 1367-1381. doi:10.1002/qre.1671

[6] Chen, S., & Nembhard, H. B. (2010). A high-dimensional control chart for profile monitoring. *Quality and Reliability Engineering International*, 27(4), 451–464. doi:10.1002/qre.1140

[7] Chen, Y., Birch, J. B., & Woodall, W. H. (2015). Cluster-Based Profile monitoring in phase 1. *Journal of Quality Technology*, 47(1), 14-29. doi:10.1080/00224065.2015.11918103

[8] Cheung, J. M. Y., Bartlett, D. J., Armour, C. L., Laba, T. L., & Saini, B. (2018). To drug or not to drug: A qualitative study of patients' decision-making processes for managing insomnia. *Behavioral Sleep Medicine, 16*(1), 1-26. doi:10.1080/15402002.2016.1163702

[9] Hakimi, A., Amiri, A., & Kamranrad, R. (2017). Robust approaches for monitoring logistic regression profiles under outliers. *International Journal of Quality & Reliability Management*, 34(4), 494-507. doi:10.1108/ijqrm-04-2015-0053

[10] Izadbakhsh, H., Noorossana, R., & Niaki, S. T. A. (2018). Monitoring multinomial logistic profiles in Phase I using log-linear models. *International Journal of Quality & Reliability Management*, 35(3), 678-689. doi:10.1108/ijqrm-04-2017-0068

[11] Jensen WA, Birch JB, & Woodall WH. Monitoring Correlation Within Linear Profiles Using Mixed Models. *Journal of Quality Technology* 2008, 40(2), 167–183. doi:10.1080/00224065.2008.11917723.

[12] Kang, L., & Albin, S. L. (2000). On-Line Monitoring When the Process Yields a Linear Profile. *Journal of Quality Technology*, 32(4), 418 426. doi:10.1080/00224065.2000. 11980027

[13] Koosha, A. A. a. M. (2011). The Effect of Neglecting Autocorrelation on the Performance of T2 Control Charts in Monitoring of Logistic Profiles. Proceedings of the 2011 IEEE ICQR.

[14] Koosha, M., & Amiri, A. (2012). Generalized linear mixed model for monitoring autocorrelated logistic regression profiles. *The International Journal of Advanced Manufacturing Technology*, 64(1-4), 487-495. doi:10.1007/s00170-012-4018-2

[15] Li, C.-I., Pan, J.-N., & Liao, C.-H. (2019). Monitoring nonlinear profile data using support vector regression method. *Quality and Reliability Engineering International*, 35(1), 127-135. doi:https://doi.org/ 10.1002/qre.2385

[16] Moheghi, H. R., Noorossana, R., & Ahmadi, O. (2020). GLM profile monitoring using robust estimators. *Quality and Reliability Engineering International*, 1– 17.doi:https://doi.org/ 10.1002/qre.2755

[17] McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*, second edition, Chapman & Hall, London, UK.

[18] Nikoo, M., & Noorossana, R. (2012). Phase II Monitoring of Nonlinear Profile Variance Using Wavelet. *Quality and Reliability Engineering International*, 29(7), 1081-1089. doi:10.1002/qre.1460

[19] Noorossana, R., Eyvazian, M., Amiri, A., & Mahmoud, M. A. (2010). Statistical monitoring of multivariate multiple linear regression profiles in phase I with calibration application. *Quality and Reliability Engineering International*, 26(3), 291–303. doi:10.1002/qre.1066

[20] Olsson, U. (2002). *Generalized Linear Models: An Applied Approach*: Lightning Source.

[21] Pan, J.-N., Li, C.-I., & Lu, M. Z. (2019). Detecting the process changes for multivariate nonlinear profile data. *Quality and Reliability Engineering International*, 35(6), 1890-1910. doi:https://doi.org/10.1002/qre.2482

[22] Paynabar, K., Jin, J., & Pacella, M. (2013). Monitoring and diagnosis of multichannel nonlinear profile variations using uncorrelated multilinear principal component analysis. *IIE Transactions*, 45(11), 1235-1247. doi:10.1080/0740817x. 2013.770187

[23] Shadman, A., Zou, C., Mahlooji, H., & Yeh, A. B. (2014). A change point method for Phase II monitoring of generalized linear profiles. *Communications in Statistics - Simulation and Computation*, 46(1), 559-578. doi:10.1080/03610918.2014. 970698

[24] Shadman, A., Mahlooji, H., Yeh, A. B., & Zou, C. (2015). A change point method for monitoring generalized linear profiles in phase I. Quality and Reliability Engineering International 31 (8), 1367-1381. https://doi.org/10.1002/qre.1671

[25] Shiau, J.-J. H., & Sun, J.-H. (2009). A new strategy for Phase I analysis in SPC. *Quality and Reliability Engineering International*, 26(5), 475-486. doi:10.1002/qre.1075

[26] Soleimani Paria, Narvand Ali, Raissi Sadigh, (2013), Online Monitoring of Auto-correlated Linear Profiles via Mixed Model, International Journal of Manufacturing Technology and Management, 27(4/5/6), 238-250.

[27] Vargas, N. J. A. (2003). Robust Estimation in Multivariate Control Charts for Individual Observations. *Journal of Quality Technology*, 35(4), 367–376. doi:10.1080/00224065.2003. 11980234.

[28] Walker, E., & Wright, S. P. (2018). Comparing Curves Using Additive Models. *Journal of Quality Technology*, 34(1), 118-129. doi:10.1080/00224065.2002.11980134

[29] Williams JD, Woodall WH, Birch, JB, & Sullivan JH (2006). Distribution of Hotelling's T2 Statistic Based on the Successive Differences Estimator. *Journal of Quality Technology* 2006, 38(3), 217–229. doi:10.1080/00224065. 2006.11918611.

[30] YEH, A. B. (2009). profile monitoring for a binary response. *IIE Transactions* 931–941. doi:10.1080/0740817090273540

TABLE 2

.THE SIMILARITY MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0	7.42	5.08	37.74	2.34	2.44	5.28	33.45	5.82	1.48	6.24	18.70	3.12	2.35	1.54	45.59	1.11	9.99	7.31	3.14
2	7.42	0	0.93	34.12	1.93	1.37	0.77	24.54	0.75	3.79	0.05	11.62	5.54	6.80	3.89	32.32	6.75	4.60	1.45	1.38
3	5.08	0.93	0	25.47	0.52	1.07	2.11	18.35	2.28	1.40	0.74	7.34	6.57	2.95	3.90	26.13	6.30	8.27	0.27	0.23
4	37.74	34.12	25.47	0	28.17	35.19	42.23	1.67	42.94	26.45	34.08	6.12	54.64	21.27	45.03	3.18	49.11	62.71	21.61	27.73
5	2.34	1.93	0.52	28.17	0	0.42	2.06	21.99	2.35	0.31	1.42	9.81	4.41	1.62	2.07	31.16	3.56	7.79	1.41	0.06
6	2.44	1.37	1.07	35.19	0.42	0	0.74	27.68	0.94	1.17	0.89	13.62	2.36	3.42	0.89	37.47	2.30	4.63	2.42	0.45
7	5.28	0.77	2.11	42.23	2.06	0.74	0	32.51	0.01	3.73	0.55	17.08	2.28	7.24	1.60	42.02	3.56	2.03	3.54	1.80
8	33.45	24.54	18.35	1.67	21.99	27.68	32.51	0	32.92	21.66	24.92	2.48	46.00	18.36	37.53	1.05	42.48	50.26	14.59	21.09
9	5.82	0.75	2.28	42.94	2.35	0.94	0.01	32.92	0	4.15	0.57	17.41	2.47	7.81	1.86	42.34	3.93	1.87	3.69	2.04
10	1.48	3.79	1.40	26.45	0.31	1.17	3.73	21.66	4.15	0	3.06	9.91	5.21	0.59	2.54	31.31	3.52	10.33	2.40	0.61
11	6.24	0.05	0.74	34.08	1.42	0.89	0.55	24.92	0.57	3.06	0	11.78	4.71	5.92	3.10	33.08	5.67	4.40	1.42	0.98
12	18.70	11.62	7.34	6.12	9.81	13.62	17.08	2.48	17.41	9.91	11.78	0	27.22	8.38	20.88	6.02	24.88	30.58	5.07	9.15
13	3.12	5.54	6.57	54.64	4.41	2.36	2.28	46.00	2.47	5.21	4.71	27.22	0	8.79	0.47	58.59	0.61	2.42	9.48	4.81
14	2.35	6.80	2.95	21.27	1.62	3.42	7.24	18.36	7.81	0.59	5.92	8.38	8.79	0	5.23	27.80	5.96	15.83	3.65	2.09
15	1.54	3.89	3.90	45.03	2.07	0.89	1.60	37.53	1.86	2.54	3.10	20.88	0.47	5.23	0	49.23	0.40	3.69	6.23	2.44
16	45.59	32.32	26.13	3.18	31.16	37.47	42.02	1.05	42.34	31.31	33.08	6.02	58.59	27.80	49.23	0	55.37	61.32	21.34	29.84
17	1.11	6.75	6.30	49.11	3.56	2.30	3.56	42.48	3.93	3.52	5.67	24.88	0.61	5.96	0.40	55.37	0	5.34	9.15	4.23
18	9.99	4.60	8.27	62.71	7.79	4.63	2.03	50.26	1.87	10.33	4.40	30.58	2.42	15.83	3.69	61.32	5.34	0	10.78	7.51
19	7.31	1.45	0.27	21.61	1.41	2.42	3.54	14.59	3.69	2.40	1.42	5.07	9.48	3.65	6.23	21.34	9.15	10.78	0	0.95
20	3.14	1.38	0.23	27.73	0.06	0.45	1.80	21.09	2.04	0.61	0.98	9.15	4.81	2.09	2.44	29.84	4.23	7.51	0.95	0