Int J Adv Manuf Technol (2016) 85:1565–1574
DOI 10.1007/s00170-015-8054-6

ORIGINAL ARTICLE

Principal component regression-based control charts for monitoring count data

Danilo Marcondes Filho[1] · Angelo Márcio Oliveira Sant'Anna[2]

Received: 14 July 2015 / Accepted: 27 October 2015 / Published online: 10 November 2015

Abstract Control charts based on regression models are appropriate for monitoring in which the quality characteristics of products vary depending on the behavior of predecessor variables. Its use enables monitoring the correlation structure between input variables and the response variable through residuals from the fitted model according to historical process data. However, such strategy is restricted to data from input variables which are not significantly correlated. Otherwise, colinear variables that hold substantial information on the variability of the response variable might be absent in the regression model adjustment. This paper proposes a strategy for monitoring count data combining Poisson regression and principal component analysis. In such strategy, colinear variables are turned into uncorrelated variables by principal component analysis and a Poisson regression is performed on principal component scores. A deviance residual control chart from the fitted model is then used to evaluate the process. The performance of that new approach is illustrated through a case study in a plastic plywood process with real and simulated data.

Keywords Statistical processes control · Residual control charts · PCA · Poisson regression · Count data

1 Introduction

The statistical process control (SPC) encompasses a set of techniques to analyze and evaluate the quality of industrial processes, and the control charts are used as its main tool. The traditional Shewhart control chart for monitoring means, ranges, fraction of nonconforming, and number of nonconformities (count data) assumes that successive samples taken from the process are independent and identically distributed (iid). However, in some processes, the mean varies as a function of one or more predecessor variables like in the multistage process for example, where a quality characteristic at the current step is affected by one or some of the quality characteristics at the previous steps.

The application of control charts for monitoring the normal response variable depending on predecessor variables was initially proposed by Mandel [1], and it was called regression control chart. Later investigation can be found in Zhang [2], Hawkins [3], Haworth [4], Wade and Woodall [5], Shu et al. [6], and Asadzadeh et al. [7], among others. The main idea is using the Shewhart control chart for monitoring the residuals from the adjusted model (since they are iid), built from the historical in-control data. Further, there are types of processes that generate non-normal response data such as Poisson distributed count data. Such scenario requires the development of control charts derived from a broad class of regression models called generalized linear models (GLMs). Skinner et al. [8, 9] proposed a control chart using Poisson regression to monitor count data (Poisson GLM). The approach works with the deviance residuals from the Poisson model derived from a likelihood ratio test. Additional study involving modeling of continuous non-normal response variables (using gamma distribution) is found in Jearkpaporn et al. [10, 11].

It is important to make clear the differences between the model-based control application we have cited in above works (and we are dealing in this work) and that in an emerging research area called profile monitoring. In a profile context, a response variable is monitored as a function of fixed values of input variables, set according to some experimental design

✉ Danilo Marcondes Filho
marcondes.danilo@gmail.com

[1] Department of Statistics, Federal University of Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil

[2] Department of Mechanical Engineering, Federal University of Bahia, Salvador, Bahia, Brazil

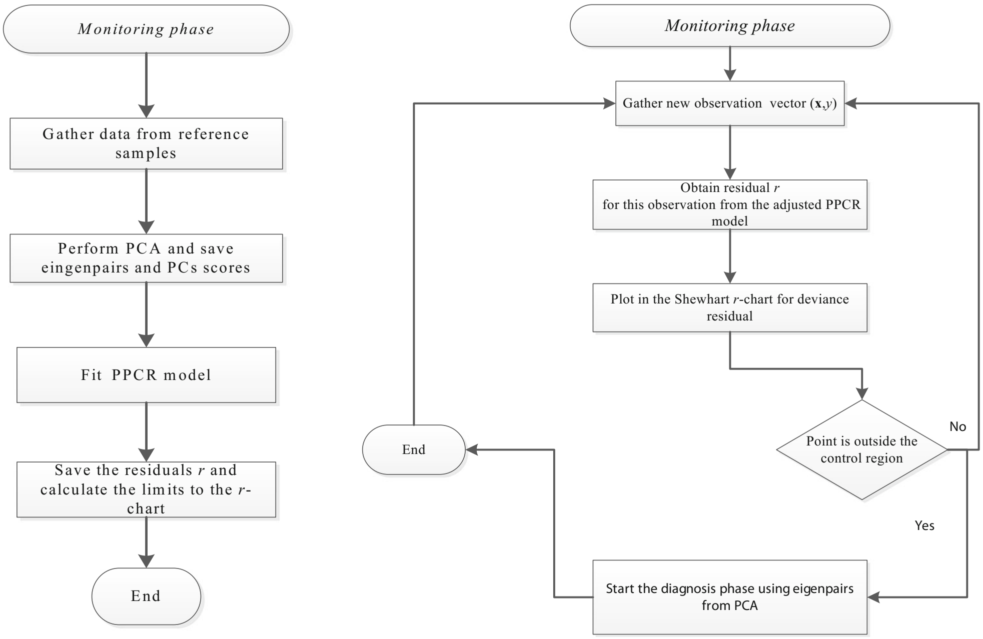
Monitoring phase
Gather data from reference samples
Perform PCA and save eingenpairs and PCs scores
Fit PPCR model
Save the residuals r and calculate the limits to the r-chart
End
Monitoring phase
Gather new observation vector $(\mathbf{x},y)$
Obtain residual r for this observation from the adjusted PPCR model
Plot in the Shewhart r-chart for deviance residual
Point is outside the control region
No
Yes
End
Start the diagnosis phase using eigenpairs from PCA

Fig. 1 The overall view of proposed approach

of interest in the process. Even the input variables have no the same values from sample to sample (i.e., different ranges of values from sample to sample), they are confined in some way in the same subinterval in each sample. In phase I, the coefficients of the fitted regression model are then estimated from a number of reference in-control samples of size n, forming a reference profile. In a monitoring phase (phase II), for each new sample with n values of response variable at the same values of input variables, the estimated coefficients of that sample profile will be compared with the coefficients from the reference profile. The profile monitoring was initially described by Kang and Albin [12], Kim et al. [13], and Mahmoud and Woodall [14] and later investigated by Mahmoud et al. [15], Noorossana et al. [16], Ayoubi et al. [17], and Amiri et al. [18], among others. The issues covered by these authors include simple and multiple (more than one input variable) linear profiles, multivariate linear profile (more than one response variable), and nonlinear profiles.

However, indeed, it is not the case here. Our proposed approach is like a multistage process, as we have discussed in the above paragraphs. In such process, in one stage, the response variable varies according to random predecessor variables from previous stage. So, we can model the correlation between them using a regression model fitted from preliminary samples of size 1, each one having one value for the response and one value for each input variable. Future samples of size 1 each are then monitored through the residual control chart, so that there is no profile monitoring here. There is a wide range of process (even with only one stage) in which a random variable response varies according to a set of random input variables like batch process (see fundamental work from Nomikos and MacGregor [19]), and the process will be shown in this work.

Since there are a few works dealing with that kind of scenario after the works cited above from the literature [8–11], much research remains to be carried out, including (i) nonnormal response data approaches, (ii) models to deal with

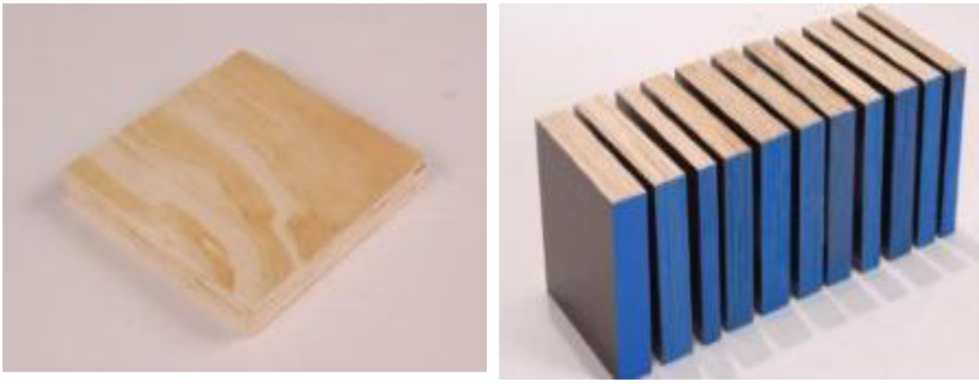

Fig. 2 View of the manufactured plywood

Table 1 Sample data from variables selected for modeling

Variable			Mean	Std. dev.	Min	Max	Unit
Input	x_1	Shrinkage	9.63	1.22	6.5	14	%
	x_2	Assembly time	14.99	1.15	12	18	min
	x_3	Wood density	0.5366	0.0156	0.5021	0.5780	g/cm^3
	x_4	Drying temper	124.43	14.68	85	165	°C
Output	y	Imperfections	14.28	44.00	0	404	–

many colinear input variables that store important information about the variability of response variables, and (iii) approaches to help the diagnosis to the unusual variation to response data.

In this paper, we present a modification of Poisson model-based control chart proposed by Skinner et al. [8] to monitor count data with multicolinearity between input variables. The new strategy that combines Poisson regression and principal component analysis (PCA) will be called *Poisson principal component regression* (PPCR). In such strategy, colinear variables are turned into uncorrelated variables by PCA and a Poisson regression is performed on PC scores. A deviance residual control chart from a fitted PPCR model was then used to evaluate the process. That approach preserves in a fitted regression model the colinear input variables which hold substantial information on the variability of the response variable. Otherwise, i.e., using the Poisson regression directly to the input variables, these could be absent in the final regression model adjustment. That is in fact the main results of the data analysis performed using the approach proposed in this paper. Another advantage over the Skinner et al. [8] approach is that the stored eigenvalues from PCA held in input variables keep important information about their correlations with a response variable that aid in the diagnosis of significant changes in the variability of the count data; i.e., it allows evaluating the effect of each input variable on the response variable. The performance of that new approach is illustrated through a case study in a plastic plywood process.

This paper is organized as follows: Section 2 presents a brief description about GLM, Poisson regression, Poisson model-based r chart (from Skinner et al. [8]), and PCA. Section 3 presents the proposed monitoring strategy. In Section 4, a case study in a plastic plywood process with real and simulated data is presented. Section 5 presents the conclusions.

2 Background

2.1 Generalized linear models

Generalized linear models (GLMs) represent a class of regression models appropriate to investigate the effect of input variables over non-normal response variables (see detailed description in McCullagh and Nelder [20]). The GLM model is based on probability distributions with unknown location parameter (θ) that belongs to the exponential family. The most important distributions in this family are normal, binomial, Poisson, gamma, and exponential. The exponential family probability density function is usually described as Eq. (1)

$$f(y;\theta,\phi) = \exp\left[a(\phi)^{-1}(y\theta - b(\theta)) + c(y,\phi)\right] \qquad (1)$$

where y is the response variable, $a(\cdot)$, $b(\cdot)$, and $c(\cdot)$ are the unknown functions, θ is the location parameter, and $\phi>0$ is the dispersion parameter.

GLMs are structured by three components: (i) random component, which defines the probability distribution of the response variable y; (ii) systematic component η, which is the linear predictor that defines the structure of the input variables; and (iii) link function, which describes the functional relationship between the systematic component and the expected

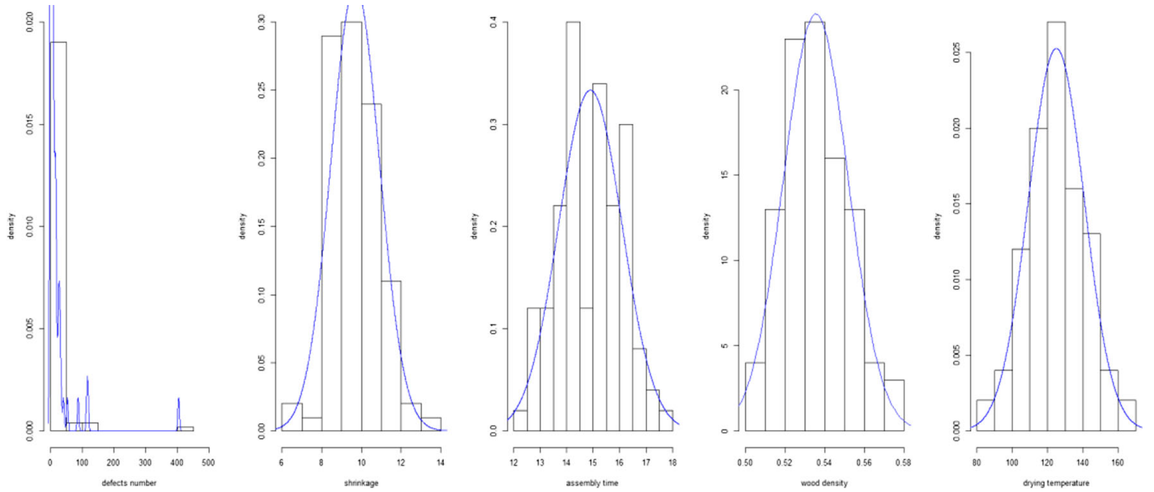

Fig. 3 Empirical distribution of input and output sample data

Table 2 Correlation matrix among variables

	y	x_1	x_2	x_3	x_4
y	1	0.8145 (0.000)	0.4223 (0.000)	0.6411 (0.000)	0.3922 (0.000)
x_1	0.8145 (0.000)	1	0.7826 (0.000)	0.0003 (0.9972)	0.0485 (0.6314)
x_2	0.4223 (0.000)	0.7826 (0.000)	1	−0.0281 (0.7809)	−0.0409 (0.6859)
x_3	0.6411 (0.000)	0.0003 (0.9972)	−0.0281 (0.7809)	1	0.8357 (0.000)
x_4	0.3922 (0.000)	0.0485 (0.6314)	−0.0409 (0.6859)	0.8357 (0.000)	1

Significant when $p<0.05$

value for the random component (i.e., the mean of response variable y). The systematic component comprising the regression model (η) is the linear combination of input variables, and it may be written by a $g(\cdot)$ function, called link function, which describes the functional relationship between the μ mean and the linear predictor (η), as in Eq. (2)

$$g(\mu) = \eta = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k \tag{2}$$

where β_{ks} are the unknown coefficients and x_{ks} are the input variables.

2.2 Poisson regression model

Consider the Poisson distributed variable y which mean λ. in witch y is a function of vector of input variables $\mathbf{x} = (x_1, x_2, \ldots, x_k)$ with unknown coefficients β_0, β_1, β_2, …, β_k. From the Poisson probability function $f(y)=\lambda^y \exp(-\lambda)(y!)^{-1}$, using Eq. (1), we can set $\theta = ln(\lambda)$, $c(y, \phi) = -ln(y)$, $b(\theta)=exp(y)$, and $a(\phi)=1$. By adopting the logarithmic link function, we define the GLM described in Eq. (2) as follows:

$$y \sim \text{Poisson}\left[\exp\left(\beta_0 + \sum_{k=1}^{K} \beta_k x_k\right)\right]. \tag{3}$$

So, $\lambda=E(y/x)$ is obtained through the inverse link function ($g^{-1}(\lambda)=\exp[\lambda]$). The coefficients β_0 and β_{ks} are estimated using the maximum likelihood method.

The degree of fit of the GLM model to the data is performed through the analysis of the deviance residuals obtained by the likelihood ratio statistic, described as

$$r = \text{sign}\left(y - \hat{\lambda}\right)\left\{2\left[y\log\left(y/\hat{\lambda}\right) - y + \hat{\lambda}\right]\right\}^{1/2} \tag{4}$$

where $\hat{\lambda}$ is the estimated conditional mean of y, calculated by means of the adjusted model. Myers et al. [21] showed that the residuals r are independent and asymptotically normal with zero mean and unit variance.

2.3 Poisson model-based control chart

Consider N as observation vectors $\mathbf{x}$, representing measures of K input variables in a process under statistical control. Each $\mathbf{x}$ observation is followed by an observation of the y count response variable. The Poisson regression model (Eq. (3)) is adjusted from historical data $\mathbf{x}$ and y, thus leading to the residuals r (Eq. (4)). Given $r \dot{\sim} N(0, 1)$, Skinner et al. [8] proposed the use of Shewhart control limits for the residuals given by

$$\mathrm{CL}_r = E(r_n) \pm w\sqrt{\mathrm{Var}(r_n)} \cong \pm w \tag{5}$$

where the constant w defines the amplitude between control limits based on the false alarm probability α.

New observations $\mathbf{x}$ and y are compared with the reference model obtained in Eq. (3), the r scores is calculated (Eq. (4)) and plotted on the r chart. The values outside the CL_r show a significant change in the correlation structure between input variables and the response variable. Skinner et al. [8] showed that the r chart is useful to monitor changes in the mean of response variable, including independent shift in a response mean, shifts caused by a change in a β_0 or one of the β_k

Table 3 Estimated coefficients of the adjusted models

Poisson regression					PPCR				
Original x	Estimated β_s	Std. error	Sig.	VIF	PC score	Estimated γ_s	Std. error	Sig.	VIF
Const.	1.454	0.050	0.000	–	Const.	1.454	0.050	0.000	–
x_1	0.279	0.673	0.678	642.97	z_1	0.515	0.024	0.000	1.234
x_2	0.702	0.657	0.285	615.24	z_2	0.854	0.018	0.000	1.162
x_3	0.343	0.312	0.271	143.17	z_3	0.069	0.092	0.456	1.119
x_4	0.676	0.354	0.056	154.17	z_4	−0.385	1.048	0.713	1.060
AIC	410.07								

Significant when $p<0.05$

Table 4 Summary of PCA results

Eigenvectors (u_i)	Eigenvalues (λ_i)	Cumulative %
(−0.155, −0.198, 0.683, 0.684)	1.457	49.53
(0.687, 0.681, 0.170, 0.184)	1.441	97.98
(0.296, −0.306, −0.650, 0.628)	0.292	99.98
(−0.644, 0.634, −0.283, 0.320)	0.030	100.00

coefficients, and shifts caused by change in a mean of at least one input variable.

2.4 Principal component analysis

The principal component analysis (PCA) is a multivariate statistical technique that seeks to summarize information about the linear correlation structure in a set of variables under analysis (see Jackson [22] and Jackson and Mudholkar [23] for a wide description of principal components, in a field of SPC).

Consider the matrix **X** with a dimension of $N\times K$ composed of N row vectors **x**, representing the measures of K variables. The linear correlation structure of data from **X** is contained within the sampling covariance correlation **S** matrix, with a dimension of $K\times K$. The PCA diagonalizes **S** in order to obtain new variables (called principal components (PCs)) which represent the projection of data from **X** in the new orthogonal directions of the variables' variability. Such projections are given by the obtained K eigenpairs $(\varphi_i, \mathbf{u}_i)$ from **S**, where $\mathbf{u}_i$, with a dimension of $K\times 1$ for $i=1,\ldots, K$, is the eigenvector which leads to the linear combination of the K variables by means of the ith PC and φ_i is the eigenvalue representing its variability. The ith PC score referring to observation **x** is given by

$$z_i = \mathbf{x}\mathbf{u}_i \tag{6}$$

where $i=1,\ldots, K$. Since PCs are not correlated, each one of them describes a unique structure of data variability.

3 Proposed monitoring strategy

This section describes the monitoring strategy that integrates Poisson regression model and PCA in order to monitor the count response variable based on colinear input variables. The strategy consists of performing the PCA in correlated input variables and then applying the Poisson regression to estimate the link between uncorrelated PCs and the count response variable. The detailed procedure will be discussed in the subsequent sections.

3.1 Combining Poisson regression and PCA

In some kinds of industrial processes, there are colinear input variables that hold relevant information about the response variable. In such case, two or more of these input variables will be excluded from the adjusted regression model, impairing the control charts' sensitivity in detecting atypical behaviors in the response variable. In some cases, all the input variables might not be significant and, therefore, they are not taken into account during response variable monitoring. The principal component regression (PCR) is a classical technique to deal with multicolinear input variables through PCA (see Rencher [24] for a wide description of PCR). In PCR, colinear input variables are turned into uncorrelated variables by PCA (as described in Eq. (6)) and a normal regression is performed on PC scores instead of original input variables. Recent studies integrating normal regression model (i.e., normal distribution response) and PCA can be found in Rajab et al. [25] and Sayadi et al. [26], among others. However, there is a lack of work that presented applications of PCR to deal with count data (i.e., Poisson distributed response) in a field of SPC.

Following the goal of this work to monitor count data in the function of colinear input variables, we perform a PCR combining PCA described in Section 2.4 and Poisson GLM model presented in Section 2.2. The new strategy will be called Poisson principal component regression (PPCR). From the historical data, we adjusted a PPCR model by rewriting Eq. (3) as follows:

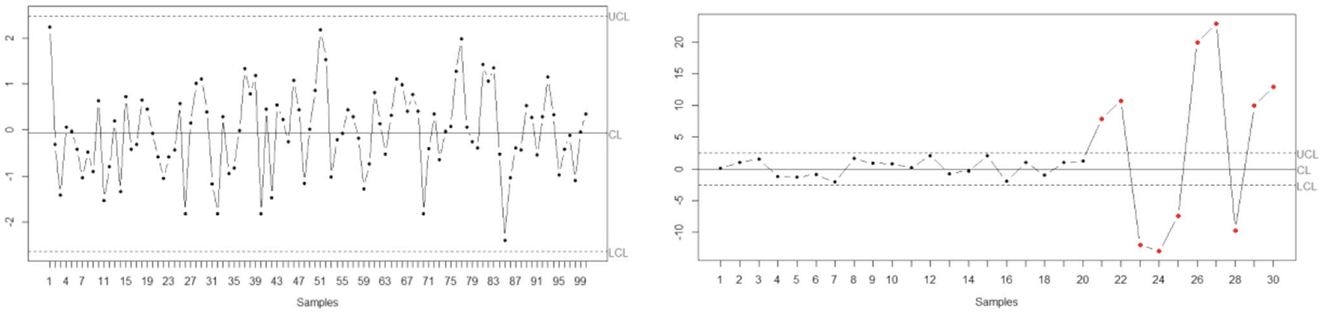

Fig. 4 Phase I with 100 historical samples for the in-control process and phase II with 30 new samples (first 20 of the in-control process and last 10 samples under bad conditions of the process)

$$y \sim \text{Poisson}\left[\exp\left(\gamma_0 + \sum_{i=1}^{K} \gamma_i(\mathbf{x}\mathbf{u}_i)\right)\right] \quad \text{or} \quad y \sim \text{Poisson}\left[\exp\left(\gamma_0 + \sum_{i=1}^{K} \gamma_i z_i\right)\right] \tag{7}$$

where $\mathbf{u}_i$ represents the eigenvector corresponding to the ith biggest eigenvalue φ_i and z_i represents the score due to the ith PC. The coefficients γ_0 and γ_{is} are estimated using the maximum likelihood method as well as the classical Poisson GLM.

When input variables show a linear correlation (as in the case of variables showing multicolinearity in regression model adjustments), a number V ($<K$) of PCs hold a substantial part of the variability structure within the input data; that is, in the SPC context, they hold relevant information about the link between input variables and the main sources of the process variability. The literature offers a wide number of criteria in order to determine the number V of PCs (see Jolliffe [27]). However, according to our approach, PCs will be selected through the Wald statistical test [28], a usual criterion for selecting input variables in regression models. The idea is selecting a number of PCs that hold relevant information about the correlation structure in the input data that affect the variability of the response count data. So, we started with all K PCs to perform PPCR described in Eq. (7).

3.2 PPCR-based control chart

The implementation of the proposed strategy was divided into two stages: (i) modeling of historical observations (phase I) and (ii) monitoring of new observations (phase II). In the first stage, based on historical data of the process under statistical control, the PPCR model is fitted and the control limits of r chart are calculated. In the second stage, the process is monitored through future samples by plotting their r scores on a r chart. The approach is detailed in the following paragraphs, and an overall view is shown in the flowchart in Fig. 1.

1. Modeling historical observations.
   (a) Collecting N observations ($\mathbf{x}$,y) referred to K input variables and the count response variable for the in-control process.
   (b) Applying PCA in observations $\mathbf{x}$ and obtaining K scores z_i (Eq. (6)).
   (c) Adjusting the PPCR from observations ($\mathbf{z}$,y), with $\mathbf{z}$=(z_1,…,z_K) (Eq. (7)).
   (d) Obtaining the deviance residuals from the PPCR model (Eq. (4)).
   (e) Setting w value and obtaining the control limits CL$_r$ of the r chart (Eq. (5)).
2. Monitoring new observations.
   (a) Collecting new observations ($\mathbf{x}$,y).
   (b) Obtaining V ($\leq K$) scores z_i, where V is the number of significant input PC scores in the PPCR model (Eq. (6)).
   (c) Obtaining deviance residuals from the new observation ($\mathbf{z}$,y), with $\mathbf{z}$=(z_1,…,z_V), in the PPCR model (Eq. (4)).
   (d) Monitoring deviance residuals through the r chart.

Table 5 Types and sizes of the imposed changes

Type	Change	Size
I	Mean of y	1σ, 2σ, 3σ
II	Mean of x_1	1σ, 2σ, 3σ
III	Mean of x_3	1σ, 2σ, 3σ

4 Application illustration

4.1 Real case study

In this section, the implementation of the proposed strategy monitoring is illustrated by a study applied to a medium-sized timber industry which manufactures laminated plastic plywood. The study consisted in evaluating the effect of input variables over the number of defects found in produced plywoods. The quality of the plywood is related to some variables, as detailed by Dermirkir et al. [29], Fang et al. [30], and Azaman et al. [31]. We are considering the number of defects per laminated plastic plywood area (y) and the following input variables: volumetric shrinkage (x_1), assembly time (x_2), wood density (x_3), and drying temperature (x_4). So, for each sample unity representing a big wooden plate with constant size (see Fig. 2), we have data of the number of imperfections accompanied by the input data of the four process variables described.

Table 6 ARL (standard error) of r chart and c chart

Size	Type	r chart	c chart
–	–	353.32 (2.70)	252.61 (2.97)
1σ	I	7.52 (0.08)	23.25 (0.31)
	II	5.61 (0.09)	45.45 (0.73)
	III	6.92 (0.08)	40.39 (0.47)
2σ	I	4.21 (0.05)	16.28 (0.18)
	II	4.49 (0.04)	35.71 (0.53)
	III	3.15 (0.04)	27.02 (0.33)
3σ	I	1.35 (0.01)	4.38 (0.07)
	II	3.37 (0.03)	25.64 (0.39)
	III	2.97 (0.02)	15.23 (0.14)

4.2 Modeling historical observations (phase I)

Table 1 shows the summary of input and response variables from reference historical data resulting from 100 observations ($\mathbf{x}$,y) (data in Appendix). We can see in Fig. 3 that the input variables are close to the normal distribution, whereas response variables show that the distribution sharply deviates from a normal shape.

Table 2 shows the sampling correlation matrix. It is possible to notice that all process variables are significantly correlated with the response variable and the input variables x_1 *and* x_2 *and* x_3 *and* x_4 *are significantly correlated to each other. Thus, we are facing a case with evidence that the count variable changes according to the colinear input variables.*

Table 3 (on the left) shows the results from the Poisson regression performed in the x_k variables (standardized to remove a scale effect). We noticed that none of them was significant on explaining the response variable y, according to the Wald test. The variance inflation factor (VIF) confirmed the presence of multicolinearity among the input variables (VIF >5). According to the proposed strategy in this study, we circumvented the multicolinearity problem using the PCA in a historical data pre-processing stage. The PCA turned the 100 vectors of observations $\mathbf{x}$ from the input variables into scores z_k. Table 3 (on the right) shows that in the fitted PPCR model, both the PCs z_1 and z_2 were significant on explaining the response variable y. We also observe that the VIF has a reduced value (VIF <5), showing that the regression model with the noncolinear PCs captures the correlation structure between the input variables x_k and the response variable y. Importantly, the significance of the first two PCs in the regression model is also expected, since they bring much of the variance-covariance structure of the original input variables (see Table 4).

In the next step, we obtained the deviance residuals that will be used in the construction of the Shewhart r control chart from the fitted regression model. The control limits are obtained using w=3 for false alarm probability α=0.0027 (area of 99.73 %, since the r scores have asymptotic normal distribution).

4.3 Monitoring new observations (phase II)

To show the performance of the Shewhart r control chart designed in last section, we used other 30 historical process data, in which 20 of them are obtained from an in-control process and the last 10 resulted in a number of defects per laminated plastic plywood area very far from the expected outcomes. The analysis of these bad samples indicates that it was caused by unusual values of the input variables combined with the low quality of raw materials. Figure 4 (chart on the left) shows the samples from the 100 data used in a phase I analysis. We can see that all the samples are well classified by the r chart, since they were indeed obtained from the in-control process. Figure 4 (chart on the right) shows the new 30 samples in a r control chart. Again, we can notice that all the samples were well classified, since the first 20 are inside the limits and the last 10 are beyond the limits.

4.4 Simulated phase II study

In order to evaluate the efficiency of our proposed strategy, we will expand the process illustrated above with simulated data including different types of disturbance imposed on the response count data. Let us keep the fitted PPCR model in the phase I analysis from the 100 historical real data (shown in Table 3, on the right) and the resulting control limits to the r chart (Fig. 4, both charts). In a phase II simulated study, we assume that the theoretical link between in-control data y and x_k follows a Poisson GLM given by Eq. (3), using coefficients with values based on those in Table 3 (on the left) giving the same weight to each input variable on the response, so that $\beta_0=1.5$ and $\beta_k=0.5$, for k=1,2,3,4, representing the standard relation between the input and output of the process. We are using the results from the Poisson regression adjusted from the sample reference data (even though none of them was significant due a multicolinearity) to estimate in a coherent way the true relation between y and x_k. Additionally, we can observe in Fig. 2 that each input variable has the empirical distribution close to normal density. In this way, we assume that each future in-control sample vector $\mathbf{x} = (x_1, x_2, x_3, x_4) \sim \mathbf{N}\,(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\rho}})$ has four-variate normal distribution with a vector of means $\hat{\boldsymbol{\mu}}$ estimated from the data in Table 1 and with a correlation estimated from the sample correlation matrix $\hat{\boldsymbol{\rho}}$, shown in Table 2. So, for each simulated data $\mathbf{x}$, the count response data y are randomly generated using a Poisson distribution with the conditional mean $\lambda=\exp(1.5+\sum_{k=1}^{k}0.5\,x_k)$.

To evaluate the power of the proposed method to monitor future observations in phase II, different disturbances were simulated in the data ($\boldsymbol{x}$,y), including three types of changes with three sizes each (totaling nine uncontrolled scenarios), as per Table 5. This shifts include (I) independent shift in a response mean, (II) shifts caused by a change in a mean of x_1, and (III) shifts caused by change in a mean of x_3. Simulations and calculations necessary for obtaining the r chart were conducted using the open source software R® [32]. In each scenario, there were 5000 replications of samples of size 1000 each.

Table 6 shows, in terms of average run length (ARL) (until a change is detected), the performance of r chart to the in-control process and to the presence of changes described in Table 5. In the first line, we can notice that ARL$_0$ (until a false alarm is detected) is close to the nominal value of 370, with w=3. Additionally, in the last column, we notice the performance of the c chart used to monitor the count data y regardless of the input variables, since, in this case study, none of them were significant in the Skinner et al. [8] approach in which Poisson regression were performed in the original input variables. So, there is no Skinner r chart in this case to compare with our PPCR-based r

chart. For that reason, we use as a benchmark the *c* chart modified with adjustments for overdispersion and asymmetry in order to equalize ARL_0.

As expected, we observed the good performance of PPCR-based *r* chart, regarding *c* chart, in detecting the three types of different uncontrolled scenarios. Additionally, at each change, it is noticed that *r* chart shows the ARL values which rapidly decrease as the change gets more intense. It demonstrates the effectiveness of using a PPCR to monitor colinear data.

It is important to highlight that, in addition to allowing the monitoring of count variable y based on the relevant colinear input variables x_k, the pre-processing of data by PCA stores the correlation structure between the process variables within the eigenvectors $\boldsymbol{u}_i$. We observed in Table 4 that the first eigenvalue is dominated by a strong correlation simulated between the variables x_1 and x_2 (shrinkage and assembly time) whereas the second eigenvalue is dominated by a strong correlation imposed on the variables x_3 and x_4 (wood density and temperature), according to the correlation structure observed in the data. Such information will be useful for diagnosing the disturbances detected by the *r* chart, by assisting in the identification of the input variables that most influenced the non-predicted change in the response variable.

5 Conclusions

The current paper presented a new strategy combining Poisson GLM regression and principal component analysis (abbreviated here as PPCR) in order to monitor a class of manufacturing processes in which the count response variable varies as a function of predecessor colinear input variables. The modified *r* chart from Skinner et al. [8] using PPCR allows monitoring the count variables, thereby preserving the relevant information about their correlation with colinear input variables. The good performance of that new approach was illustrated through a case study in a plastic plywood process with real and simulated data.

Also, with PPCR as a theoretical basis for the proposed *r* chart, the diagnosis analysis to the out-of-control points may be done. The stored eigenvalues from PCA held in input variables keep important information about their correlations with a response variable that aid in the diagnosis of significant changes in the variability of the count data.

Finally, as an additional contribution in a SPC field, we presented a new control approach. The PPCR-based *r* chart is easy to handle, which widens its applicability. Further, the proposed technique can be of a large application, since many manufacturing process have the same structure as of the process illustrated here.

Acknowledgments The authors are grateful to the National Council of Scientific and Technologic Development of Brazil (CNPq) under grants 486707/2013-0 and 441628/2014-1 (Universal) for the financial support of this work.

Compliance with ethical standards This material is the full authorship of these authors. This work did not involve any human participants and / or animals. There is no conflict of interest here. The authors are in agreement with the editorial rules of the journal and are available to provide any supporting documentation.

Conflict of interest The authors declare that they have no competing interests.

Appendix

Table 7 Reference samples of data used in plastic plywood case study

Samples	y	x_1	x_2	x_3	x_4
1	18	12.39	17.43	0.52	117.62
2	1	8.49	14.15	0.52	106.45
3	41	9.83	15.13	0.57	149.61
4	53	12.39	17.90	0.54	123.93
5	3	8.58	14.03	0.53	117.65
6	24	10.66	16.17	0.55	131.58
7	26	11.18	16.55	0.54	125.56
8	0	8.47	13.75	0.53	122.64
9	4	9.78	15.03	0.54	127.06
10	4	10.85	16.12	0.52	112.38
11	8	11.11	16.31	0.53	121.27
12	0	9.18	14.88	0.53	111.07
13	12	9.77	15.00	0.55	139.21
14	30	10.09	15.47	0.56	144.64
15	2	7.77	13.30	0.54	126.40
16	5	9.74	15.19	0.54	123.61
17	15	10.03	15.46	0.55	135.28
18	28	11.16	16.44	0.55	136.21
19	1	8.91	14.02	0.52	122.04
20	0	8.58	14.04	0.53	112.62
21	6	10.06	15.38	0.54	126.16
22	1	8.70	14.07	0.53	121.27
23	0	7.51	13.36	0.54	121.79
24	9	10.39	15.85	0.54	125.98
25	4	10.99	16.22	0.52	108.85
26	17	10.27	15.52	0.55	136.05
27	1	10.02	15.64	0.51	97.45
28	6	10.66	15.82	0.54	129.94
29	13	10.42	15.48	0.55	137.11
30	3	8.16	13.79	0.55	131.98
31	4	7.78	13.36	0.56	139.81
32	33	10.53	15.83	0.56	144.74
33	2	7.89	13.21	0.55	137.69
34	16	10.69	16.18	0.54	127.07
35	18	10.18	15.50	0.55	135.66
36	1	7.90	13.48	0.52	108.93

Table 7 (continued)

Samples	y	x_1	x_2	x_3	x_4
37	6	10.23	15.34	0.54	128.90
38	1	10.06	15.36	0.52	111.13
39	1	8.56	14.16	0.53	117.90
40	5	7.74	13.45	0.55	131.50
41	8	10.06	15.38	0.54	125.97
42	9	10.90	16.23	0.53	120.29
43	2	10.03	14.95	0.51	110.89
44	4	11.06	16.50	0.52	106.44
45	26	10.52	15.76	0.56	144.57
46	404	12.28	17.53	0.57	155.71
47	7	10.22	15.78	0.53	112.65
48	2	7.47	12.62	0.52	119.97
49	0	9.32	14.73	0.52	109.57
50	2	8.59	13.91	0.53	124.46
51	87	12.08	17.17	0.55	140.29
52	6	11.11	16.27	0.52	114.15
53	4	7.97	13.40	0.55	140.90
54	1	8.85	14.11	0.53	118.37
55	4	9.49	14.62	0.54	132.04
56	2	9.41	15.06	0.52	101.86
57	10	8.98	14.32	0.55	141.84
58	1	9.17	14.87	0.51	94.14
59	1	9.67	15.36	0.53	115.73
60	6	9.63	14.98	0.54	125.89
61	1	9.16	14.34	0.54	129.12
62	15	9.21	14.36	0.57	154.07
63	1	7.46	12.63	0.55	143.61
64	0	8.04	13.46	0.53	121.33
65	9	8.87	14.19	0.56	147.40
66	0	8.74	14.84	0.51	86.97
67	29	10.75	15.85	0.54	136.39
68	4	8.57	14.11	0.56	136.79
69	1	9.42	14.88	0.52	105.78
70	20	10.76	16.28	0.55	128.15
71	118	12.18	16.96	0.55	147.36
72	1	9.82	15.21	0.53	116.98
73	4	9.46	14.62	0.55	139.07
74	0	8.37	13.42	0.54	132.05
75	2	9.35	14.31	0.52	123.90
76	4	10.28	15.60	0.53	122.60
77	2	9.15	14.98	0.52	99.82
78	5	8.48	14.08	0.55	131.42
79	1	11.78	17.24	0.50	87.34
80	2	10.12	15.42	0.52	113.44
81	1	8.34	14.33	0.53	102.17
82	6	10.57	15.91	0.53	118.30
83	1	8.50	14.12	0.53	112.09
84	15	10.73	15.97	0.54	131.08
85	3	8.28	13.72	0.55	133.94
86	1	8.56	13.71	0.52	118.57
87	6	8.65	14.29	0.55	131.59
88	9	9.79	15.21	0.55	130.42
89	2	8.03	13.40	0.55	135.13
90	0	9.04	14.34	0.51	100.18
91	0	9.89	15.34	0.50	94.26
92	3	10.50	15.88	0.53	116.07
93	115	12.19	16.93	0.55	148.56
94	4	10.23	15.72	0.53	119.18
95	15	10.31	15.79	0.54	127.62
96	6	7.87	13.34	0.57	149.34
97	10	9.32	14.19	0.54	138.50
98	1	8.35	13.66	0.53	123.23
99	1	8.52	14.06	0.53	112.88
100	2	10.40	15.59	0.52	114.12

References

1. Mandel BJ (1969) The regression control charts. J Qual Technol 1: 1–9
2. Zhang GX (1985) Cause-selecting control charts—a new type of quality control charts. QR J 12:221–225
3. Hawkins DM (1993) Regression adjustment for variables in multivariate quality control. J Qual Technol 25:170–182
4. Haworth DA (1996) Regression control charts to manage software maintenance. Softw Maint Res Pract 8:35–48
5. Wade MR, Woodall WH (1993) A review and analysis of cause-selecting control charts. J Qual Technol 25:161–169
6. Shu L, Tsui K, Tsung F (2008) A review of regression control charts. Encycl Statist Qual Reliab 260:1–9
7. Asadzadeh S, Aghaie A, Shahriari H (2009) Monitoring dependent process steps using robust cause-selecting control charts. Qual Reab Eng Int 25:851–874
8. Skinner KR, Montgomery DC, Runger GC (2003) Process monitoring for multiple count data using generalized linear model-based control charts. Int J Prod Res 41:1167–1180
9. Skinner KR, Montgomery DC, Runger GC (2004) Generalized-linear model-based control charts for discrete semiconductor process data. Qual Reliab Eng Int 20(8):777–786
10. Jearkpaporn D, Montgomery DC, Runger GC, Borror CM (2003) Process monitoring for correlated gamma-distributed data using generalized linear model-based control charts. Qual Reliab Eng Int 19:477–491
11. Jearkpaporn D, Montgomery DC, Runger GC, Borror CM (2005) Model-based process monitoring using robust generalized linear models. Int J Prod Res 43(7):1337–1354
12. Kang L, Albin SL (2000) On-line monitoring when the process yields a linear profile. J Qual Technol 32:418–426
13. Kim K, Mahmoud MA, Woodall WH (2003) On the monitoring of linear profiles. J Qual Technol 35:317–328
14. Mahmoud MA, Woodall WH (2004) Phase I analysis of linear profiles with calibration applications. Technometrics 46:380–391

15. Mahmoud MA, Parker PA, Woodall WH, Hawkins DM (2007) A change point method for linear profile data. Qual Reliab Eng Int 23: 247–268
16. Noorossana R, Vaghefi SA, Dorri M (2011) Effect of non-normality on the monitoring of simple linear profiles. Qual Reliab Eng Int 27: 425–436
17. Ayoubi M, Kazemzadeh RB, Noorossana R (2014) Estimating multivariate linear profiles change point with a monotonic change in the mean of response variable. Int J Adv Technol 75:1537–1556
18. Amiri A, Koosha M, Azhdari A, Wang G (2015) Phase I monitoring of generalized linear model-based regression profiles. J Stat Comput Simul 85:2839–2859
19. Nomikos P, Macgregor JF (1995) Multivariate SPC charts for monitoring batch processes. Technometrics 37:41–59
20. McCullagh P, Nelder JA (1989) Generalized linear models, 2^{a}th edn. Chapman & Hall, London
21. Myers RH, Montgomery DC, Vining GG (2002) Generalized linear models with applications in engineering and the sciences. John Wiley & Sons, New York
22. Jackson JE (1991) A user's guide to principal components. John Wiley & Sons, New York
23. Jackson JE, Mudholkar GS (1979) Control procedures for residuals associated with principal component analysis. Technometrics 21(3):341–349
24. Rencher AC (2002) Methods of multivariate analysis, 2^{a}th edn. John Wiley & Sons, New York
25. Rajab JM, MatJafri MZ, Lim HS (2013) Combining multiple regression and principal component analysis for accurate predictions for column ozone in Peninsular Malaysia. Atmos Environ 71:36–43
26. Sayadi AR, Lashgari A, Paraszczak J (2012) Hard-rock LHD cost estimation using single and multiple regressions based on principal component analysis. Tunn Undergr Space Technol 27:133–141
27. Jolliffe IT (2004) Principal component analysis, 2^{a}th edn. Springer, New York
28. Neter J, Kutner MH, Nachtsheim CJ, Li W (2005) Applied linear statistical models, 5th edn. McGraw-Hill/Irwin, New York
29. Demirkir C, Özsahin S, Aydin I, Colakoglu G (2013) Optimization of some panel manufacturing parameters for the best bonding strength of plywood. Int J Adhes Adhes 46:14–20
30. Fang L, Chang L, Guo W-J, Chen Y, Wang Z (2014) Influence of silane surface modification of veneer on interfacial adhesion of wood–plastic plywood. Appl Surf Sci 288:682–689
31. Azaman MD, Sapuan SM, Sulaiman S, Zainudin ES, Khalina A (2013) Shrinkages and warpage in the processability of wood-filled polypropylene composite thin-walled parts formed by injection molding. Mater Des 52:1018–1026
32. R Development Core Team. (2014). R: a language and environment for statistical computing. R Foundation for Statistical Computing, ISBN 3-900051-07-0, 2014. Available at http://www.r-project.org.