



Control chart based on residues: Is a good methodology to detect outliers?

Jean Paulo Guarnieri¹ · Adriano Mendonça Souza¹ · Luciane Flores Jacobi¹ · Bianca Reichert¹ · Claudimar Pereira da Veiga² 

Received: 30 June 2018 / Accepted: 20 July 2019 / Published online: 27 July 2019
© The Author(s) 2019

Abstract

The purpose of this article is to evaluate the application of forecasting models along with the use of residual control charts to assess production processes whose samples have autocorrelation characteristics. The main objective is to determine the efficiency of control charts for individual observations (CCIO) and exponentially weighted moving average (EWMA) charts when they are applied to residuals of models of AR(1) or MA(1) to detect outlier in autocorrelated processes. Considering autocorrelation strength and sign in the data series and the outlier range, the series were simulated accomplishing 640,000 sets. The series were contaminated by anomalous observations at 100th position, an AR(1) or MA(1) model were fitted, and the residuals were evaluated by CCIO and EWMA control charts; the points correctly detected as an autocorrelation were recorded. For the parameters investigated (autocorrelation and outlier range), a detection rate was generated in each chart, and nonparametric comparison tests were applied. The result of the tests showed the superiority ($p < 0.05$) of the CCIO chart for both models. The study of the influence of the sign and magnitude of the autocorrelation parameter showed no significant ($p > 0.05$) for either AR(1) or MA(1) charts and models. In this context, control charts for individual observations (CCIO) were confirmed to effectively detect outliers through residuals in industrial autocorrelated processes originated in first-order AR and MA models.

Keywords Quality control · Residual control charts · Outliers · Efficiency of control charts · Residual control charts

Introduction

The statistical process control (SPC) methodology is used by companies and industries to evaluate the specification standards pattern and the process capability based on control charts, in order to promote the continuous improvement (He and Wang 2018; Woodall 2000; Peres and Fogliatto 2018). The justification and importance of studying this theme is revealed due to its incipient growth in terms of articles deposited in the Web of Science journal repository but not

less important to industries. The growth of the scientific production of the researched subject is shown in Fig. 1.

Figure 1 shows that we find only 188 works in the most diverse areas of knowledge. Just 28 works are from the industrial engineering area, published from 1992 to 2017. This shows that the theme “statistical process control” can still be explored, existing is a literature gap. For Montgomery (2004), statistical quality control techniques can be applied in industrial and non-industrial processes, in order to detect and correct the anomalies in the process to meet the specified target of a given product. For Duarte and Saraiva (2008), Toledo et al. (2008) and Bouslah et al. (2018), the SPC is an important management tool for improving quality and increasing the productivity of industrial processes, used to reduce the variability process.

SPC uses different evaluation techniques applied to autocorrelated data, such as time series analysis, regression analysis and artificial neural networks and support vector regression (Du and Lv 2013; Du and Zhang 2016). We notice that

✉ Claudimar Pereira da Veiga
claudimar.veiga@gmail.com

¹ Universidade Federal de Santa Maria, Avenida Roraima, 1000, Statistics Department Building 13, Office 1205 C - CCNE/UFSM, Santa Maria, RS, Brazil

² Department of General and Applied Administration, Federal University of Parana (UFPR), Lothário Meissner Ave, 632, Curitiba, PR 80210-170, Brazil

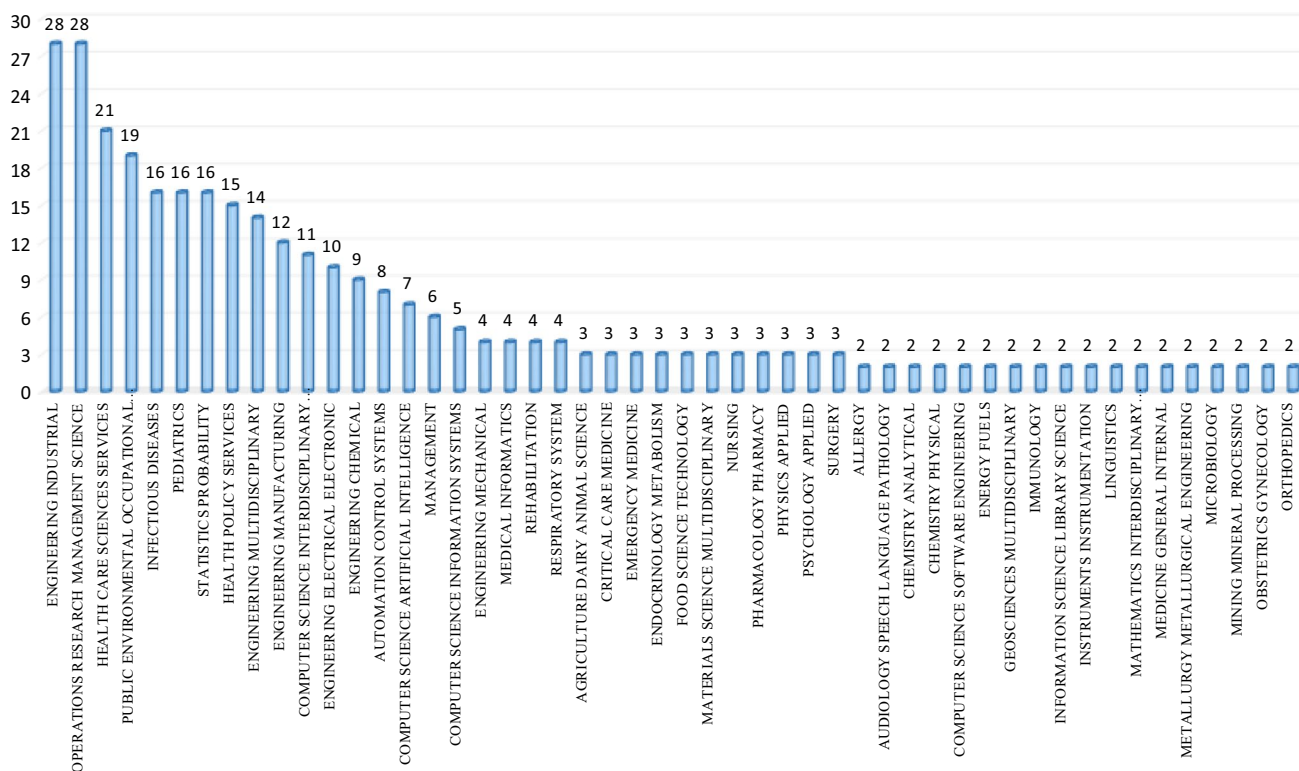


Fig. 1 Number of publications on the topic statistical process control in the Web of Science

research in the case of outlier detection in correlated processes using neural networks is rare in the literature.

Lalor and Zhang (2001) presented a multivariate case to detect three types of outliers: range, spatial and relationship outliers. They showed that neural networks are a free and effective modelling to detect irregular data. Guo and Xue (2012) used statistical methods to detect outliers and treat these discrepancies to later carry out training and extrapolations through neural networks. The conclusion was that statistical methods and neural networks promote analysis and prediction more assertively.

In order to evaluate water quality control, methodologies such as artificial neural networks (ANN), principal component analysis (PCA) and universal process modelling (UPM) were used by Cancilla and Fang (1996) to characterize quality variables of the Niagara River. In this research, the methodologies such as PCA and UPM were able to capture outliers and ANN was used to carry out the predictions.

The basic quality control activities relay on reducing the quality variations, so the variance magnitude can be estimated and the potential variability can be identified, resulting in the minimization of potential losses (Christino et al. 2010). In relation to the control charts, these were created with the purpose of monitoring the process, by observing the average and standard deviation as highlighted by McCracken and Chakraborti (2013) and Trafimow et al.

(2018). However, the relevant literature in statistical shows that other control charts could be used to control quality and productivity (Costa and Machado 2007; Abbas 2018).

Statistical process control charts were used as process monitoring tool, in order to evaluate the effect of the regulation of the rotary separation system on the losses observed during the mechanized harvest. Cunha et al. (2014) observed that the control charts were effective tools in detecting the total losses of the industrial tomato crop, because it was detected that the process was out of control and that the adoption of corrections in the process would provide a better harvesting efficiency (Cunha et al. 2014). Voltarelli et al. (2015) used the control charts as an indicator of quality in the monitoring of losses in mechanized harvesting of sugarcane in the Triângulo Mineiro region, in the State of Minas Gerais, Brazil. The control charts for individual values and for exponentially weighted moving averages were used. However, the control chart for individual values was one that had a good response in the monitoring of losses in mechanized harvesting of sugarcane.

The simultaneous monitoring process using a single nonparametric control chart makes it easy for process controllers to use, because it allows to identify the presence of special causes in the process and does not hurt the assumption of normality present in the traditional control charts (McCracken and Chakraborti 2013).

Dumičić and Žmuk (2015) used exponentially weighted moving average (EWMA) and cumulative sum (CUSUM) charts to support decision making in the stock market trading process. The EWMA and CUSUM control charts were used for the purpose of pointing out stock purchase and sale signals. The control charts did not perform satisfactorily with the presence of anomalies such as non-normality and autocorrelation, which impacted the performance of the control charts. The importance of control charts is also highlighted by Aparisi et al. (2018), demonstrating that the quality process has improved a lot in relation to previous decades, and the majority of the samples collected in high-quality processes do not defective units.

Silva et al. (2017) have developed a strategy for continuous monitoring from multivariate statistical process control in the Consigua™25 tablet manufacturing line. This strategy was based on the technique of principal component analysis, while the impact of the deviations imposed on the continuity of the process was evaluated from the residues through the Hotelling T^2 control charts. The results evidenced that the residues imposed were detected in the control chart, thus showing effectiveness in its application, mainly in monitoring the temperature control of the granulator drum.

Durmusoglu (2018) proposed an approach to detect abnormal deviations by updating the methodology of time series forecasting models using control charts. Its purpose was to monitor the residues (difference between actual and expected values of the interest variable) in a continuous way using the control charts, in order to ascertain whether such residues were outside the established limits, which would be an alert or a process change.

Control charts are usually planned and evaluated assuming that consecutive observations of the process are independent and identically distributed (*i.i.d.*). The observations must also meet the assumptions of normality and homoscedasticity (constant variance), and they cannot show autocorrelation characteristics. However, Montgomery (2004), Montgomery and Runger (2003) and Claro et al. (2007) emphasize that although the independence among the observations is the most important assumption, it is often violated in practice. This is due to the fact that, in general, manufacturing processes are governed by inertial elements, and when the interval between observations become small compared to these forces, they become correlated over time.

Failure to meet these assumptions when applying control charts can result in a significant increase in false alarms, an unwanted factor that not only increases control costs but also leads to wrong conclusions and causes the operator to lose credibility as a consequence (Costa et al. 2004; Del Castillo 2002). So, the alternative used in an autocorrelated process is to fit an ARIMA model and use the residues produced for this model to evaluate the process (Veiga et al. 2016; Kalavani et al. 2019).

However, it should be noted that the residuals produced by the models must be approximately normal and independent with zero mean and constant variance, fully satisfying the assumptions on the proper use of control charts. This procedure is one of the major alternatives to avoid problems caused by the violation of assumptions of no correlation among the observations (Del Castillo 2002).

The aim of this research is to use the statistical process control (SPC) technique applied to univariate time series in order to determine the efficiency of control charts for individual observations (CCIO) and the exponentially weighted moving average (EWMA) chart to detect *outliers* in autocorrelated processes (Miranda 2001; Santos and Barreto 2018). The charts will be applied in residuals originated from an autoregressive (AR) or moving average (MA) process. In addition to demonstrating the influence of the autocorrelation strength of the process, this study also verified the detection power of the charts in relation to the magnitude of the anomalous observation.

It is evident that the detection of outliers and their treatment is important to obtain a better fit and a smaller prediction error, because the forecast models are based on the assumption of outliers pretreatment (Veiga et al. 2010; Bashiri and Moslemi 2013; Puchalski et al. 2018). Ghomi and Sogandi (2018) showed that many actual production processes are contaminated by a continuous stream in correlated data. So, it is important to distinguish trends, seasonalities and outliers.

These outliers can often go unnoticed in residual control charts which originate in autocorrelated processes, because, according to Chang (1982), the mathematical model used to remove the serial correlation can incorporate the behaviour of the outlier in its structure, reducing its effect in the residual series and thus hindering outlier detection. It should also be noted that the presence of outliers enhances the control limits because the variability of the process is increased.

Outliers and structural changes are often found in time series analysis, and they may be associated with unexpected or uncontrollable events. Such different observations may compromise the usual methods of time series analysis (Miranda 2001). The presence of an outlier can seriously bias the least squares estimates of the parameters of an ARMA model. Palma (1999), and Rounaghi and Zadeh (2016) explain that studies on outliers in time series are relatively scarce when compared to studies in the field of linear regression. This is due to the multiplicity of the ARIMA models—AR(p), MA(q), ARMA(p , q), ARIMA(p , d , q)—which requires various detection mechanisms (Palma (1999).

This research proposes a methodology to join the statistical process control and engineering process to identify outliers in autocorrelated data, using control charts applied to residuals from an autoregressive and moving average models.

Methodology

The methodological steps suggested below were used to test the efficiency of residual control charts in outlier detection in autocorrelated processes, as well as the variables that influence such efficiency, represented in this research by outlier range and the autocorrelation coefficient.

Database

To accomplish the main purpose of this research, autoregressive (AR) and moving average (MA) processes are generated under the following restrictions, and eight autocorrelation parameters (± 0.5 , ± 0.6 , ± 0.7 and ± 0.8) and eight outlier ranges were combined in each model. They were inserted at position 100th of the series of simulated data, considering the following ranges of deviations (1σ ; 1.5σ ; 2σ ; 2.5σ ; 3σ ; 3.5σ ; 4σ ; and 4.5σ), yielding a total of 64 possible combinations. In total, 10,000 series were simulated for each combination of autocorrelation parameters and outlier range in order to make the performance of the control chart more robust and, thus, determine the percentage of outlier detection. Hence, 640,000 series were simulated for each model—AR(1) and MA(1).

Methodological steps

Figure 2 clearly shows a flow chart comprising the twelve steps that were followed in this research and the decisions that were taken.

The following (Chart 1) are the twelve steps (S1–S12):

Chart 1 Twelve steps (S1–S12)

S1	Simulate a time series with 200 observations with $\mu_0=0$ and constant variance, through a data generating process, such as AR(1) or MA(1). The parameters (φ or θ) must have positive and negative autocorrelations, with a magnitude of 0.5, 0.6, 0.7 and 0.8	S3	The residual are checked whether they meet the assumptions required to be used by control charts through diagnostic tests (Box et al. 1994; Werner and Ribeiro 2003)
S2	The series through autoregressive AR(1) and MA(1) are fitted using the Box and Jenkins (1970) methodology to obtain <i>i.i.d.</i> residuals, and achieve the assumptions of normality, homoscedasticity (constant variance) and do not show autocorrelation (Morettin and Toloï 2004; Enders 2003; Hamilton 1994; Veiga et al. 2014)	S4	The residuals will be validated by means of a control chart for individual observations (CCIO) and a moving range (MR) chart (Montgomery 2004; Montgomery and Runger 2003; Claro et al. 2007; Costa et al. 2004), using residuals obtained from the AR and MA processes. If the residual series does not have any points outside the control limits, or if no particular pattern is identified, the simulated series will be considered stable and used for research purposes. If the residuals are not statistically under control, the simulated series will be discarded and considered not valid for the study
		S5	After validation step, an outlier was introduced at position 100th of the simulated data series. The outlier range should vary in the following magnitudes: 1σ ; 1.5σ ; 2σ ; 2.5σ ; 3σ ; 3.5σ ; 4σ ; 4.5σ where σ represents the standard deviation obtained from the original data series. To avoid a biased analysis, it should be noted that if the outlier has a positive value in the observation of the original series to be introduced, the value of the outlier to be inserted should be also positive, keeping the same movement of the series. If the original observation is negative, the outlier to be introduced will be also negative
		S6	Introduced the <i>outlier</i> , the series again will be fitted in order to eliminate the autocorrelation effect (Fava 2000) and obtain the residuals that meet the conditions to apply control charts

S7	Construct CCIO and EWMA control charts with the residuals of the original series contaminated by the outlier. In the EWMA chart, parameters $\lambda=0.2$ and $L=2.86$ must be used because, according to Montgomery (2004), these parameters represent values of average run length (ARL) equal to 370, a similar value to the ARL of the CCIO chart. Thus, an effective comparison can be made between the two types of control charts; Box and Luceño (1997) also suggest such parameters
S8	Check whether the observation marked with the outlier is outside the control limits, which is detected by the control charts
S9	Record outlier detection by the control chart in order to compute the detection rate
S10	Develop steps S1 to S9 until 10,000 series are obtained for each combination of autocorrelation parameter (φ or θ) and outlier range
S11	After obtaining the 10,000 series for each combination, determine the detection rates and present the results for each control chart.
S12	Answer the following questions based on the simulations and the detection rate of the charts (i) What residual control chart—CCIO or EWMA—is more efficient in detecting an outlier with variable range? (ii) Is there a significant difference between positive and negative autocorrelation parameters for each chart? (iii) Is there any significant influence of the autocorrelation value on the detection power of the charts for each outlier range?

The series were simulated by means of free statistical software *R-Project* from steps S1 to S10, and the filters $AR(p)$ and $MA(q)$ were fitted by the *Statistica 7.0* package as well as the CC and nonparametric tests.

Results and discussion

This section is divided into two subsections: the first one will discuss the autoregressive models and the second subsection will introduce the moving average model, for both models.

Autoregressive model AR(1)

The series with different magnitudes of outliers inserted in each predetermined position were fitted again, and the residuals coming from the model were evaluated applying the control charts. Thus, the residual series analysis of the model verified whether the charts were effective in detecting the outliers previously inserted in the original series.

Table 1 shows the rate of outlier detection in an AR(1) process by the CCIO control chart. The values were arranged as a function of the variation of the autocorrelation parameter φ and the outlier range. The data in Table 1 indicate, for example, that for a parameter whose autocorrelation is 0.5 and an outlier with range of 1σ , the detection power is 0.0449; that is, the outlier was detected in 449 of the 10,000 simulated series for the control chart—CCIO, in this case.

Data visual analysis on Table 1 shows that the autocorrelation has affected the detection of the outlier for the CCIO control chart except for small outlier ranges, 1σ and 1.5σ , where the detection rate did not have significant variation with the autocorrelation magnitude. The reduction in detection power as the autocorrelation parameter value increases is best viewed in Fig. 3, which displays a chart of detection efficiency as a function of the autocorrelation parameter for each outlier range.

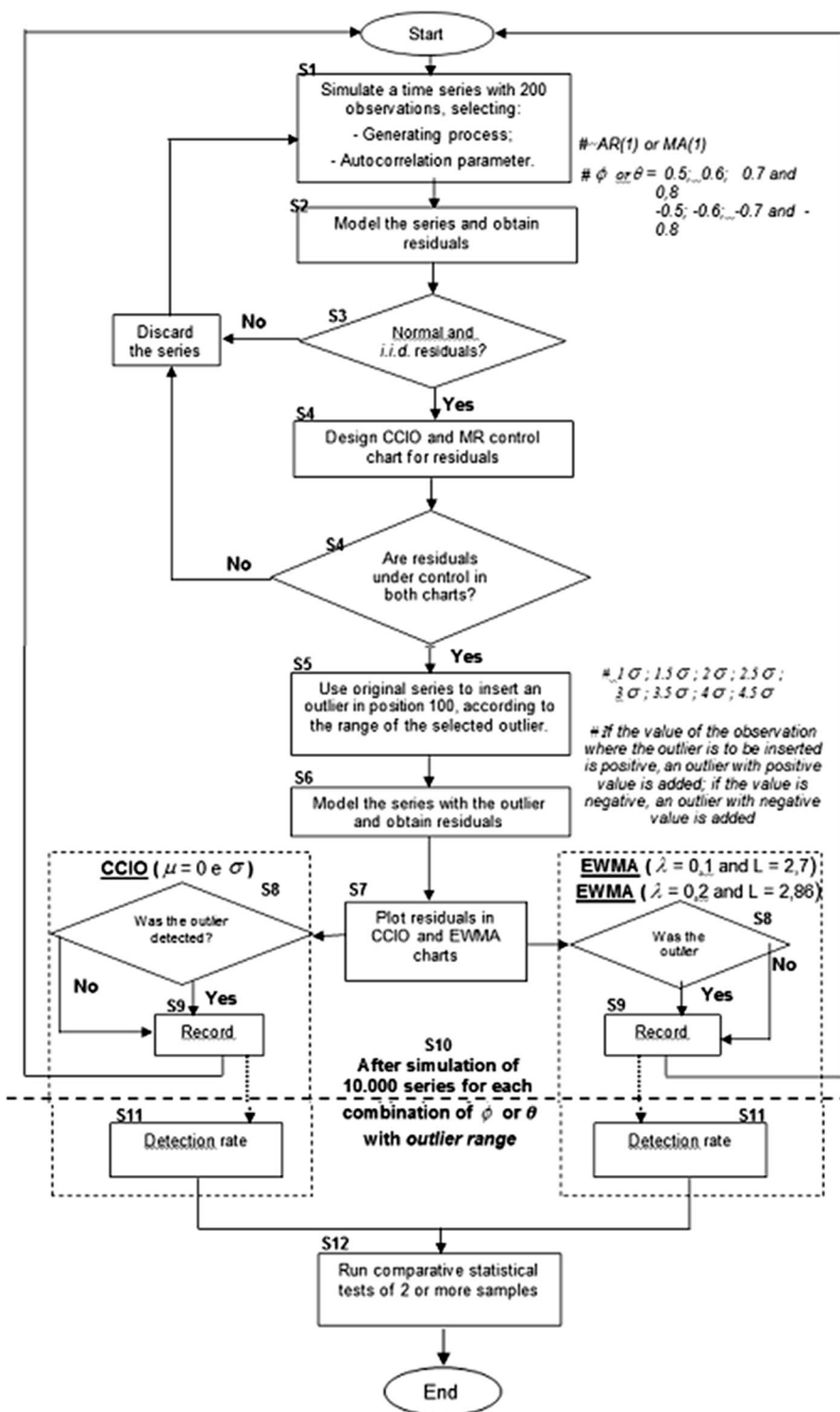
For moderate positive autocorrelation whose order is $\varphi=0.5$, CCIO charts were observed to have a higher detection efficiency, with a decrease in value as the autocorrelation changed from moderate to strong. The same occurred in negative autocorrelations, where for moderate values ($\varphi=-0.5$), efficiency was higher in comparison with the value of $\varphi=-0.8$, which represents a strong autocorrelation.

This behaviour can be explained by the fact that, after the outlier is inserted in the simulated series, it is fitted again to obtain the residuals. Thus, the residual corresponds to the difference between the original observation and the past observation multiplied by the autocorrelation parameter. The larger the parameter of autocorrelation, the smaller the difference and consequently the smaller the residual, with a residual of lesser magnitude being hardly detected by the control charts.

Comparing the detection rate between the positive and negative autocorrelations, it can be seen that the CCIO chart is more efficient for positive autocorrelation values, since the detection rates in these cases are slightly higher compared to the negative parameters for each outlier range.

Table 2 shows outlier detection rate in an AR(1) process by the EWMA control chart, with parameters $\lambda=0.2/L=2.86$. The values were arranged as a function of the variation of the autocorrelation parameter φ and the outlier range. The data in Table 2, with $\lambda=0.2$ and $L=2.86$, are illustrated in Fig. 4, where an efficiency detection chart was

Fig. 2 Flow chart of the methodological steps of the study



designed for each range as a function of the autocorrelation parameter.

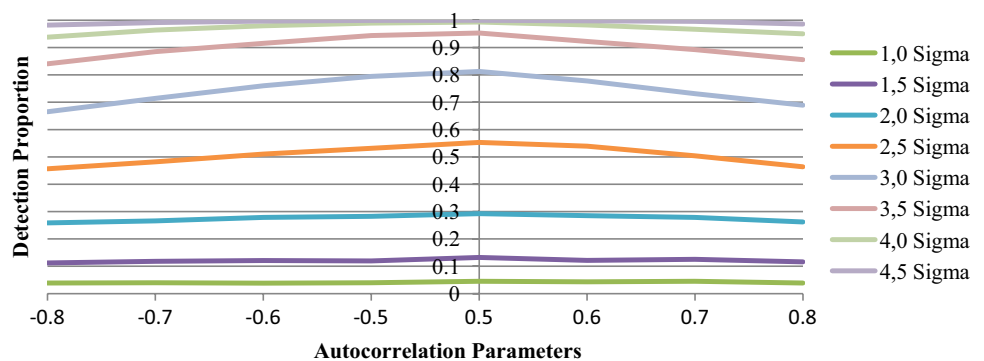
Figure 4 shows a distinct behaviour in outlier detection by the EWMA control chart, when the parameters of positive and negative autocorrelations are compared. It can be seen that for small outlier ranges, 1σ and 1.5σ , the detection rate

is not significant in relation to the magnitude of the autocorrelation. This behaviour is similar to what happened in the CCIO chart. For the other ranges with positive autocorrelations, an increased detection power is observed as the correlation changes from moderate to strong. This behaviour does

Table 1 Outlier detection rate in autoregressive processes AR(1) through the control chart for individual observations (CCIO)

Autocorrelation parameter	Outlier range							
	1σ	1.5σ	2σ	2.5σ	3σ	3.5σ	4σ	4.5σ
-0.8	0.0389	0.1124	0.2587	0.4563	0.6654	0.8406	0.9381	0.9822
-0.7	0.0398	0.1183	0.266	0.4819	0.7143	0.8851	0.9638	0.9916
-0.6	0.038	0.1212	0.2787	0.5103	0.7601	0.9156	0.9797	0.9967
-0.5	0.0397	0.1197	0.2826	0.5315	0.7944	0.9439	0.9896	0.9992
0.5	0.0449	0.1324	0.2925	0.5531	0.8124	0.9533	0.9932	0.9996
0.6	0.043	0.1213	0.2851	0.4563	0.7778	0.9221	0.9826	0.9979
0.7	0.0452	0.1256	0.2786	0.504	0.7315	0.8925	0.9671	0.9957
0.8	0.0386	0.116	0.2622	0.4641	0.6891	0.8557	0.95	0.9861

Fig. 3 Efficiency of detection, on the y axis, as a function of the autocorrelation parameter, on the x axis, for data coming from an AR(1) process using the CCIO control chart



not occur in the CCIO chart, as seen previously. Moreover, the behaviour is reversed for negative autocorrelations; that is, as the strength of the autocorrelation increases, the detection power decreases.

The data in Table 4 and the scales of the charts displayed in Fig. 4 show a great difference in the detection rate between the positive and negative parameters for the same range of differences. The same effect was seen in the CCIO chart, but the magnitude of the difference had a much lower value.

Moving average model: MA(1)

The methodological steps used in the moving average model are similar to those used in the autoregressive models to obtain the detection rates of the CCIO and EWMA control charts, when varying the autocorrelation parameter and the outlier range. Table 3 displays outlier detection rate in MA(1) processes by the control chart CCIO. The values were arranged as a function of the variation of the autocorrelation parameter and the outlier range. We can see in Table 3 that for an autocorrelation parameter of 0.5 and an outlier with range of 1σ, the detection power was 0.0407; that is, the outlier was detected by the control charts in 407 of the 10,000 simulated series.

Table 2 Outlier detection rate in autoregressive processes AR(1) through the EWMA control chart with parameters λ=0.2 and L=2.86

Autocorrelation parameter	Outlier range							
	1σ	1.5σ	2σ	2.5σ	3σ	3.5σ	4σ	4.5σ
-0.8	0.0113	0.0262	0.0442	0.0819	0.1272	0.1918	0.2739	0.3598
-0.7	0.0143	0.0275	0.0475	0.0813	0.1338	0.2006	0.2867	0.3703
-0.6	0.012	0.0274	0.0512	0.0929	0.1429	0.2151	0.3044	0.4084
-0.5	0.0124	0.0295	0.0554	0.0959	0.1521	0.2293	0.3166	0.4303
0.5	0.0186	0.0397	0.0731	0.138	0.2359	0.3526	0.504	0.6313
0.6	0.0187	0.0412	0.0789	0.1463	0.2407	0.3673	0.5153	0.6614
0.7	0.0212	0.0405	0.0826	0.1521	0.2488	0.3711	0.5272	0.6851
0.8	0.0192	0.0449	0.0819	0.1559	0.2497	0.3805	0.5366	0.6985

Fig. 4 Detection efficiency, on the y axis, as a function of the autocorrelation parameter, on the x axis, for data coming from an AR(1) process using the EWMA control chart ($\lambda=0.2$ and $L=2.86$)

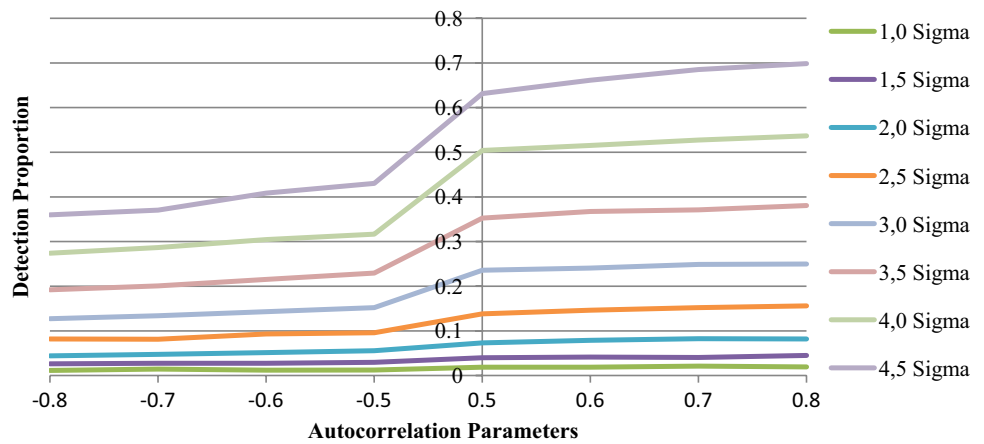


Table 3 Outlier detection rate in moving average MA(1) processes through the control chart for individual observations (CCIO)

Autocorrelation parameter	Outlier range							
	1σ	1.5σ	2σ	2.5σ	3σ	3.5σ	4σ	4.5σ
-0.8	0.0383	0.108	0.2476	0.4611	0.7095	0.8795	0.9726	0.9950
-0.7	0.0423	0.1148	0.2646	0.4924	0.7409	0.9126	0.9838	0.9980
-0.6	0.0403	0.115	0.2729	0.5204	0.7728	0.9307	0.9888	0.9992
-0.5	0.0419	0.1157	0.289	0.5395	0.8131	0.9568	0.9948	0.9996
0.5	0.0407	0.1234	0.2917	0.5543	0.8227	0.96	0.9948	0.9999
0.6	0.043	0.12	0.2877	0.5482	0.8013	0.9431	0.9907	0.9990
0.7	0.0403	0.1185	0.2856	0.5216	0.7787	0.9317	0.9857	0.9985
0.8	0.0426	0.1185	0.2792	0.5102	0.7576	0.9139	0.9821	0.9975

Table 3 shows that the autocorrelation affects the outlier degree detection by the CCIO control chart, except for small outlier ranges (1σ and 1.5σ), where the detection rate shows hardly any variation with the magnitude of the autocorrelation.

For moderate positive autocorrelation whose order is $\theta=0.5$, CCIO charts are considered to be more efficient in the detection when the outlier range varies between 2.5σ and 4.5σ , with a reduction in its value as the strength of the autocorrelation increases. The same occurs in negative autocorrelations, where for moderate values ($\theta=-0.5$), efficiency is higher in comparison with the value $\theta=-0.8$, which represents a strongly negative autocorrelated process. This behaviour is shown in Fig. 5, which illustrates a chart of detection efficiency as a function of the autocorrelation parameter for each outlier range.

The comparison of the detection rate between the positive and negative autocorrelations discloses that the CCIO chart is slightly more efficient for positive autocorrelation values, except for the range of 1σ . In this magnitude, a significant difference in the detection rate between the positive and negative autocorrelations cannot be said to exist without a detailed statistical analysis.

Table 4 displays outlier detection rate in MA(1) processes by the EWMA control chart, with parameters

$\lambda=0.2/L=2.86$. The values were arranged as a function of the variation of the autocorrelation parameter θ and the outlier range.

The data in Table 4 are displayed in Fig. 6, where an efficiency detection chart was designed for each outlier range, as a function of the autocorrelation parameter. The analysis of Fig. 6 shows that for the positive autocorrelations there is not a clear trend of the detection rate behaviour as the strength of the autocorrelation is increased. As mentioned above for the AR model, a possible cause may be the heterogeneity of the residuals, which affects the design of the EWMA control chart. For all the outlier ranges, the detection percentage for the negative autocorrelations had a decline when the autocorrelation ranged from -0.5 to -0.8 .

The observation of the data in Table 4 and the scales of the charts in Fig. 6 enable the perception of a great difference in the detection rate between the positive and negative autocorrelation parameters for the same ranges of differences. The same effect was seen in the CCIO chart, but the magnitude of the difference had a much lower value.

Fig. 5 Detection efficiency, on the y axis, as a function of the autocorrelation parameter, on the x axis, for data coming from a MA(1) process using the CCIO control chart

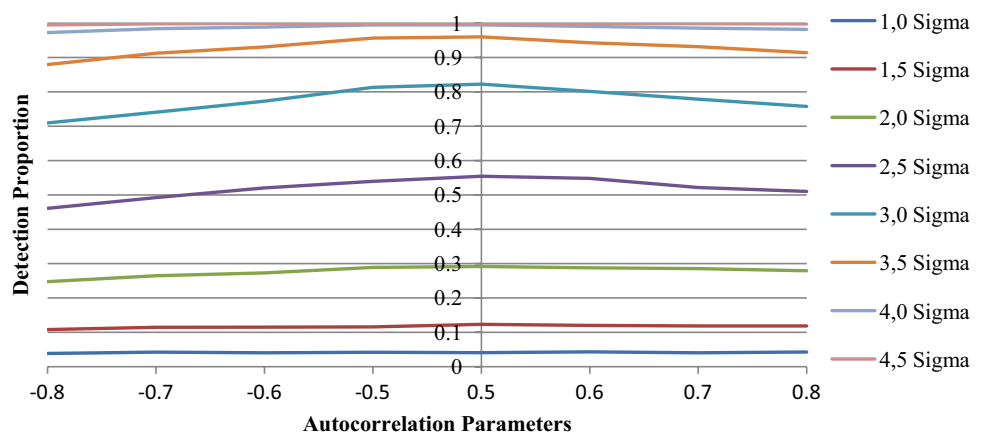
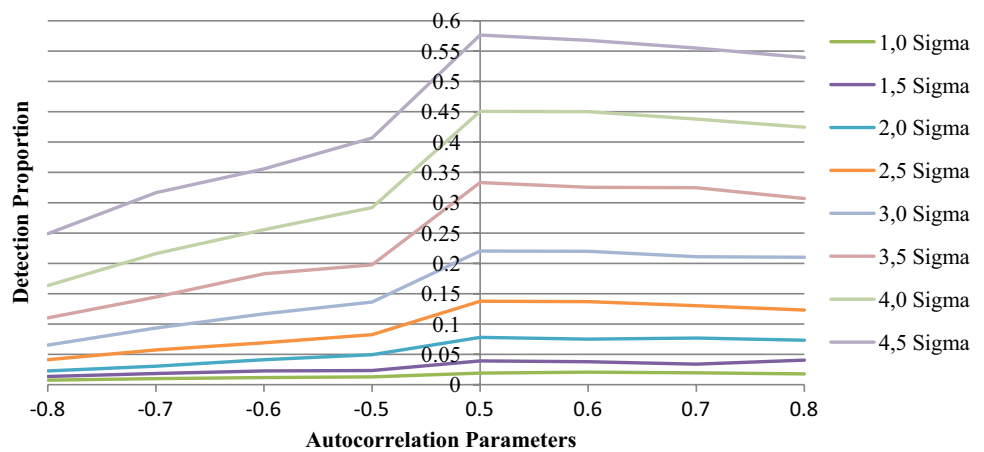


Table 4 Outlier detection rate in MA(1) processes through the EWMA control chart with parameters $\lambda=0.2$ and $L=2.86$

Autocorrelation parameter	Outlier range							
	1σ	1.5σ	2σ	2.5σ	3σ	3.5σ	4σ	4.5σ
-0.8	0.0076	0.0135	0.0226	0.0413	0.0653	0.1101	0.1635	0.2489
-0.7	0.0100	0.0185	0.0304	0.0572	0.0934	0.1444	0.2162	0.3165
-0.6	0.0118	0.0226	0.0413	0.0691	0.1168	0.1829	0.2556	0.3557
-0.5	0.0129	0.0233	0.0495	0.0825	0.1362	0.1975	0.2919	0.4066
0.5	0.0191	0.0391	0.0779	0.1376	0.2203	0.3332	0.4506	0.5764
0.6	0.0206	0.0379	0.0750	0.1369	0.2199	0.3254	0.4501	0.5677
0.7	0.0195	0.0339	0.0768	0.1303	0.2111	0.3247	0.4380	0.5550
0.8	0.0178	0.0405	0.0733	0.1231	0.2100	0.3069	0.4245	0.5394

Fig. 6 Detection efficiency, on the y axis, as a function of the autocorrelation parameter, on the x axis, for data coming from an MA(1) process using the EWMA control chart ($\lambda=0.2$ and $L=2.86$)



Discussion

In this subsection, the data obtained in Tables 1, 2, 3 and 4 will be analysed by means of nonparametric statistics in order to provide a statistical basis to answer the questions listed below and to prove some behaviours that were visually observed in the data tables. This analysis aims to answer the questions made in step S12, indicating which chart is more efficient considering all restrictions imposed in this study. To

answer the questions asked in step S12, Table 5 shows the results of nonparametric statistical tests applied.

A comparison of the results obtained for the AR(1) and MA(1) models shows that the behaviour concerning the efficiency in detecting an outlier previously inserted in these models is the same. When the CCIO and the EWMA charts were compared to check whether there is a significant difference in their detection power, the null hypothesis of equality between the two samples was rejected at a significance level of 5% in both models. Therefore, there is a significant

Table 5 Results of the nonparametric statistical tests

Comparison	Used test	Chart	AR(1)	MA(1)
Difference in the detection power between the CCIO and EWMA charts	Mann–Whitney U test	–	Significant ($p < 0.05$)	Significant ($p < 0.05$)
Difference in the detection power between positive and negative autocorrelation parameters	Mann–Whitney U test	CCIO	Not significant ($p > 0.05$)	Not significant ($p > 0.05$)
		EWMA	Not significant ($p > 0.05$)	Not significant ($p > 0.05$)
Difference in the detection power between the values of autocorrelation parameters	Kruskal–Wallis test	CCIO	Not significant ($p > 0.05$)	Not significant ($p > 0.05$)
		EWMA	Not significant ($p > 0.05$)	Not significant ($p > 0.05$)

(+) Positive autocorrelations, (–) negative autocorrelations; the value in parentheses represents the p value. All the tests are significant for p value < 0.05

difference between the two charts regarding the power of outlier detection; the CCIO chart has greater efficiency than the EWMA chart.

A comparison as to whether there is a difference between positive and negative autocorrelation parameters for both charts (CCIO and EWMA) and both models (AR and MA) shows that the behaviour between the positive and negative autocorrelation parameters is different in each chart, with the EWMA chart having a greater difference in detection. The value of weighting constant λ influences the performance of the CICO and EWMA charts, as greater as the value of the λ the chart is able to capture large discrepancies, and the EWMA chart behave as a CICO chart. Smaller values to the constant λ are able to capture short discrepancies. So, a compromise between the number of standard deviations from the central line (L) and the weighting constant (λ) must be chosen as a way to compare the average runs length in the CICO and EWMA charts. When a comparison was made as to whether there is a difference in the detection power between the different autocorrelation values on both models and charts, the null hypothesis of equality of the samples was not rejected at a significance level of 5%. Thus, there is no difference in outlier detection when there is variation in the correlation parameter value.

Conclusions

This article aimed to evaluate the application of forecasting models along with the use of residual control charts to assess production processes whose samples have autocorrelation characteristics. The main objective was to determine the efficiency of control charts for individual observations (CCIO) and exponentially weighted moving average (EWMA) charts when they are applied to residuals of models of AR(1) or MA(1) to detect outlier in autocorrelated processes. Results showed that for both the AR and MA models, the CCIO

control chart is more efficient than the EWMA chart in the detection of outliers, even for low ranges. As shown by Montgomery (2004), the EWMA control chart is more effective in detecting small permanent process changes ranging from 1.5σ to 2σ , while the CCIO chart is more efficient in detecting major changes at the process level. However, as observed in this study, when the intention is to detect an outlier or make an abrupt change to a process (represented by the change of only a sample of such process), the CCIO chart is more efficient for small and large ranges. One possible reason for the poor performance of the EWMA chart in detecting an outlier by means of the residuals lies in the low weight given to the current residual by the weighting constant λ . Thus, the effect of the outlier is “masked” when developing EWMA statistics to be plotted on the control chart. A spike in the series of original data will result in a disruption in the EWMA statistics, but in the subsequent periods, the interference of the outlier disappears. However, when there is a change in the average of the process, the EWMA statistics tends to increase in the subsequent periods until it extrapolates the control limits. For this reason, the EWMA control chart is more appropriate to detect changes on average, but not recommended to detect outliers.

When an evaluation was made to check whether there is a significant difference in detection between autocorrelations with positive and negative parameters for models AR(1) and MA(1), both CCIO and EWMA control charts showed different behaviours when the parameters of positive and negative autocorrelations were different in each chart, and there is a greater difference in the EWMA chart, since its p value is closer to the rejection region of the null hypothesis of the test. Although the weights of the autocorrelations do not significantly affect detection rates, the visual analysis of the data enables the observation of a small variation in the CCIO charts and a more significant change in the EWMA charts. Another observed behaviour is related to the detection power of the EWMA chart, when positive parameters

are used in AR(1) processes. In all the studied cases, for both positive and negative autocorrelations, detection decreases as the autocorrelation strength increases from 0.5 to 0.8. However, in the case mentioned, the behaviour is the opposite, with an increase in the detection efficiency when the autocorrelation parameters vary from 0.5 to 0.8.

As suggestion for future research is adopting the mixed model, autoregressive integrated moving average (ARIMA) is used to extract the residuals of an autocorrelated process and, thus, study the performance of the control charts applied in this research, and test whether heteroscedasticity can affect the design of the EWMA control chart.

Acknowledgements We thank CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), the Brazilian Federal Agency for the Support and Evaluation of Graduate Education, for providing financial support for the development of this research, which resulted in the master's thesis of the first author. The authors would like to thank the Editor Sadigh Raissi and anonymous reviewers of the Journal of Industrial Engineering International for their significant contribution to the final version of this article.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Abbas N (2018) Homogeneously weighted moving average control chart with an application in substrate manufacturing process. *Comput Ind Eng* 120:460–470
- Aparisi F, Epprecht EK, Mosquera J (2018) Statistical process control based on optimum gages. *Qual. Reliab Eng Int* 34:2–14
- Bashiri M, Moslemi A (2013) The analysis of residuals variation and outliers to obtain robust response surface. *J Ind Eng Int* 9(2):1–10
- Bouslah B, Gharbi A, Pellerin R (2018) Joint production, quality and maintenance control of a two-machine line subject to operation-dependent and quality-dependent failures. *Int J Prod Econ* 195:210–226
- Box GE, Jenkins GM (1970) *Time series analysis, forecasting and control*. Holden Day, San Francisco, p 1970
- Box G, Luceño A (1997) *Statistical control by monitoring and feedback adjustment*. Wiley, New York
- Box GEP, Jenkins GM, Reinsel GC (1994) *Time series analysis: forecasting and control*. Holden-Day, San Francisco
- Cancilla DA, Fang XC (1996) Evaluation and quality control of environmental analytical data from the Niagara River using multiple chemometric methods. *J Great Lakes Res* 22(2):241–253
- Chang IH (1982) *Outliers in time series*. University of Wisconsin-Madison, Dept. of Statistics
- Christino EM, Bonduelle GM, Iwakiri S (2010) Application of control charts in the production process of tauari (*Couratari oblongifolia*) wood flooring. *Cerne* 16(3):299–304
- Claro FAE, Costa AFB, Machado MAG (2007) Gráficos de controle de EWMA e de Xbar para monitoramento de processos autocorrelacionados. *Produção* 17(3):536–546
- Costa AFB, Machado MAG (2007) Gráficos de controle de EWMA e de X-barra para monitoramento de processos autocorrelacionados. *Revista Produção* 17(3):536–546
- Costa AFB, Epprecht EK, Carpinetti LCR (2004) *Controle estatístico da qualidade*. Atlas, São Paulo
- Cunha JPB, de A. Machado T, Santos FL, Coelho LM (2014) Perdas na colheita de tomate industrial em função da regulação da colhedora. *Pesq Agropec Trop Goiânia* 44(4):363–369
- Del Castillo E (2002) *Statistical control adjustment for quality control*. Wiley, New York
- Du S, Lv J (2013) Minimal Euclidean distance chart based on support vector regression for monitoring mean shifts of autocorrelated processes. *Int J Prod Econ* 141(1):377–387
- Du S, Zhang R (2016) Modelling and joint monitoring of input and output of systems with arbitrary order autoregressive disturbance. *Int J Prod Res* 54(6):1822–1838
- Duarte BPM, Saraiva PM (2008) An optimization based approach for designing attribute acceptance sampling plan. *Int J Qual Reliab* 25(2):824–841. <https://doi.org/10.1108/02656710810898630>
- Dumičić K, Žmuk B (2015) Statistical control charts: performances of short term stock trading in Croatia. *Bus Syst Res* 6(1):22–35
- Enders W (2003) *Applied econometric time series*. Wiley series in probability and mathematical statistics. Wiley, New York
- Fava VL (2000) In: Vasconcelos MAS, Alves D (eds) *Manual de econometria*. Editora Atlas, São Paulo
- Ghomi SMTF, Sogandi F (2018) A two-sided Bernoulli-based CUSUM control chart with autocorrelated observations. *J Ind Eng Int*. <https://doi.org/10.1007/s40092-018-0290-6>
- Guo WW, Xue H (2012) An incorporative statistic and neural approach for crop yield modelling and forecasting. *Neural Comput Appl* 21(1):109–177
- Hamilton JD (1994) *Time series analysis*. University Press, Princeton
- He QP, Wang J (2018) Statistical process monitoring as a big data analytics tool for smart manufacturing. *J Process Control* 67:35–43
- Kalavani F, Mohammadi-Ivatloo B, Zare K (2019) Optimal stochastic scheduling of cryogenic energy storage with wind power in the presence of a demand response program. *Renew Energy* 130:268–280
- Lalor GC, Zhang CS (2001) Multivariate outlier detection and remediation in geochemical databases. *Sci Total Environ* 281:99–109
- McCracken AK, Chakraborti S (2013) Control chart for joint monitoring of mean and variance: an overview. *Qual Technol Quant Manage* 10(1):17–36
- Miranda CF (2001) *Modelação linear de series temporais na presença de outliers*. Dissertação, Faculdade de Ciências da Universidade do Porto, PT
- Montgomery DC (2004) *Introdução ao controle estatístico da qualidade*. LTC, Rio de Janeiro
- Montgomery DC, Runger GC (2003) *Estatística aplicada e probabilidade para engenheiros*. LTC, Rio de Janeiro
- Morettin PA, Toloi CMC (2004) *Análise de séries temporais*. Edgard Blücher, São Paulo
- Palma JA (1999) *Estimação dos Parâmetros de Séries Temporais em Presença de Outliers*. Comunicação no âmbito da conferência EST/10 anos, Escola superior de tecnologia de Setúbal
- Peres FAP, Fogliatto FS (2018) Variable selection methods in multivariate statistical process control: a systematic literature review. *Comput Ind Eng* 115:603–619
- Puchalski W, Ribeiro GT, Da Veiga CP, Freire RZ, Coelho LS (2018) Agribusiness time series forecasting using Wavelet neural networks and metaheuristic optimization: an analysis of the soybean sack price and perishable products demand. *Int J Prod Econ* 203:174–189
- Rounaghi MM, Zadeh FN (2016) Investigation of market efficiency and financial stability between S&P 500 and London stock exchange:

- monthly and yearly forecasting of time series stock returns using ARMA model. *Physica A* 456(15):10–21
- Santos JDA, Barreto GA (2018) Novel sparse LSSVR models in primal weight space for robust system identification with outliers. *J Process Control* 67:129–140
- Toledo A et al (2008) Caracterização das perdas e distribuição de cobertura vegetal em colheita mecanizada de soja. *Engenharia Agrícola Jaboticabal* 28(4):710–719
- Trafimow D, Wang T, Wang C (2018) Means and standard deviations, or locations and scales? That is the question! *New Ideas Psychol* 50:34–37
- Veiga CR, Veiga CP, Duclós LC (2010) The accuracy of demand forecast models as a critical factor in the financial performance of the food industry. *Future Stud Res J* 2:81–104. <https://doi.org/10.24023/FutureJournal/2175-5825/2010.v2i2.58>
- Veiga CP, Veiga CRP, Catapan A, Silva WV, Torato U (2014) Demand forecasting in food retail: a comparison between the Holt-Winters and ARIMA models. *WSEAS Trans Bus Econ* 11:608–661
- Veiga CP, Veiga CRP, Puchalski W, Coelho LS, Tortato U (2016) Demand forecasting based on natural computing approaches applied to the foodstuff retail segment. *J Retail Consum Serv* 31:174–181
- Voltarelli MA, Silva RP, Zerbato C, Paixão CSS, de O. Tavares T (2015) Monitoring of mechanical sugarcane harvesting through control charts. *Engenharia Agrícola* 35(6):1079–1092
- Werner L, Ribeiro JLD (2003) Previsão de demanda: Uma aplicação dos modelos Box-Jenkins na área de assistência técnica de computadores pessoais. *Gestão & Produção* 10(1):47–67
- Woodall WH (2000) Controversies and contradictions in statistical process control. *J Qual Technol* 32(4):341–350