Forecasting and Controlling Key Performance Indicators in Call Centers

Negin Mehrbod (✉ n.mehrbod@campus.fct.unl.pt)

Faculty of Science and Technology: Universidade Nova de Lisboa Faculdade de Ciencias e Tecnologia

https://orcid.org/0000-0002-2774-5040

Izunildo Cabral

Faculty of Science and Technology: Universidade Nova de Lisboa Faculdade de Ciencias e Tecnologia

José Requeijo

Faculty of Science and Technology: Universidade Nova de Lisboa Faculdade de Ciencias e Tecnologia

Antonio Grilo

Faculty of Science and Technology: Universidade Nova de Lisboa Faculdade de Ciencias e Tecnologia

---

Forecasting and Controlling Key Performance Indicators in Call Centers

Negin Mehrbod[11]

n.mehrbod@campus.fct.unl.pt

Izunildo Cabarel[2]

i.cabral@fct.unl.pt

Jose G. Requeijo[3]

jfgr@fct.unl.pt

Antonio Grilo[4]

acbg@fct.unl.pt

[1,2,3,4]UNIDEMI, Departamento de Engenharia Mecânica e Industrial, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, 2829-516, Caparica, Portugal

---

[1] corresponding author

Abstract

This paper proposes a methodology for modeling and controlling the performance of call centers. Most call centers use *CRM* (Customer Relationship Management) systems to record data of all contacts between agents and clients. These data may be autocorrelated. To model autocorrelated processes effectively, the proposed methodology integrates in a logical way *ARIMA* (Autoregressive Integrated Moving Average) modeling and *SPC* (Statistical Process Control) tools. *ARIMA* is used to model the process and identify the model that best fits the time series. The fitted model is used to compute residuals, predict future values for the quality variable(s) being monitored and determine the prediction errors. To achieve these goals, the Box-Jenkins methodology is employed. These outputs are then used to apply *SPC*, in this case the Shewhart control charts for autocorrelated data. First, the computed residuals are used to build the control charts in Phase I of *SPC*, verify the process stability and estimate the process parameters. Then, these parameters are used to establish the control limits of the charts used in Phase II of *SPC* to monitor and control the prediction errors. The proposed methodology is tested in a case study of a large call center in Portugal. The results of the case study suggest that *ARIMA* modeling and *SPC*, when properly integrated, provide a set of effective tools for monitoring call center performance when autocorrelated data are available. This paper has important implications for both theory and practice.

1.Introduction

Performance measurement is a research topic that has attracted considerable academic and practitioner attention, in both the manufacturing and service industries. In service industries, the evaluation of call center performance has been extensively investigated, given its importance in the decision-making process of managers of call centers. Performance measurement can help managers know whether the interactions between agents and clients are conducted effectively. Managers can learn from these experiences and use the evidence systematically to improve the quality of the service they offer. Call centers can be broadly defined as facilities designed to support the delivery of an interactive service via telephone communication (Excoffier et al., 2016).

To evaluate and control call center performance in a reliable way, it is important to secure on-line real-time measurements of Key Performance Indicators (*KPI*s). To achieve this, call centers have been replacing traditional information systems with Customer Relationship Management (*CRM*) systems. These systems can collect and record, in real-time, all calls and log all agents' and customers' activities in a single system, producing a massive amount of data. These data can be collected from different call center units, which creates new challenges for the way call center managers control the performance of their agents. Collecting and recording massive amounts of data is not enough, on its own, to inform policies and decision making in any organization. One challenge call center managers face is how to analyze and transform these data into information that adds value and supports their decision-making in real-time and in an effective way. Another

challenge is how to use historical activities and customer experiences to predict future events, and take action to avoid future problems and bottlenecks.

A third challenge is more technical than strategic. As is known, automated systems, such as *CRM,* record all measurements of call center *KPI*s. As a result, the probability of producing autocorrelated data is higher. Thus, these data must be properly modeled to ensure reliable analysis and prediction of the way the performance of call center agents evolves over time. This can be achieved by applying advanced statistical tools that are effective in the modeling of autocorrelated data. Autoregressive Integrated Moving Average (*ARIMA*) modeling is a popular and flexible class of prediction techniques that use historical data to forecast time series that are autocorrelated(Khashei et al., 2012, Villalobos et al., 2021).

After predicting the trend of *KPI*s in call centers, managers need to monitor the quality of service in order to ensure customers are satisfied and identify potential areas for improvement. Attention to the quality of customer service is a basic condition for attracting new customers and retaining current customers. Statistical Process Control (*SPC*) is an approach that has been widely applied in many industries to improve quality(Özdemir, 2021, Yang and Zhou, 2015). It is a powerful collection of problem-solving tools that is useful in achieving process stability and improving capability through the reduction of variability. It involves the collection, analysis, and interpretation of data from quality monitoring activities (Montgomery, 2020). Control charts are one of the powerful SPC tools and be used to monitor the evolution of one or more quality characteristics, i.e., to monitor whether a process is in control or not(Kaya and Turgut, 2021, Zan et al., 2020). They can provide information about a process mean and variance and detect the presence of special causes of variation, enabling quick intervention to eliminate these causes and bring the process into a state of control where the average proportion of out-of-spec units is below some specified level (Montgomery, 2020, Quesenberry, 1997).

An analysis of the current literature reveals that, although performance measurement of call centers has been extensively investigated (see Table 1), there are no studies that apply *ARIMA* modeling and *SPC* to predict and control call center KPIs in situations where autocorrelated data are generated. Those studies that provide guidelines to evaluate call center performance do not explain how to monitor the performance of call centers over time, or how to deal with autocorrelated data (e.g. (Baraka et al., 2013, Baraka et al., 2015, Ma et al., 2011)). This paper proposes an innovative methodology to forecast and monitor the performance of call centers' agents, and quickly detect pattern changes due to abnormal situations. The proposed methodology integrates *ARIMA* modeling and *SPC* tools to address the issue of autocorrelated data explicitly.

The remainder of this paper is organized as follows. In Section 2 we provide the theoretical background to call center performance, *ARIMA* modeling and *SPC*, with emphasis on Shewhart control charts. Section 3 presents the proposed methodology and explains the steps necessary to implement it. In Section 4, the applicability of the proposed methodology is tested in a case study of a large call center in Portugal. This section presents the data, and discusses the results. Lastly, in Section 5, conclusions and suggestions for future work are presented.

2. Theoretical background

2.1 Call center performance measurement

High competition in the market has made performance improvement one of the main concerns of all businesses including call centers. *KPI*s are critical indicators to measure the performance of an organization or a special process. *KPI*s have been used for a long time in call centers and have been described in several research studies (Mehrbod et al., 2017). *KPI*s should only be changed if a company's primary objectives also change. However, it is important to use *KPI*s which are completely relevant and clearly show the results, to know if the call center is on the right track.

Over the years, several studies of performance measurement in call centers have been conducted. These studies have identified several metrics that are widely used to evaluate call center performance in different domains and from different perspectives, such as service quality, agent performance, process efficiency, emotional labor, working conditions, job satisfaction, and so on. These *KPI*s and their definitions are summarized in Table 1.

Table 1: KPIs frequently used in call center performance measurement research

Reference	Objective	Analyzed/Identified KPIs	Type of analysis
(Soliman et al., 2012)	To propose a methodology for evaluating call center performance using queueing models	Waiting time and abandonment rate	Queue modelling
(Jose et al., 2017)	To identify the key factors affecting the operational performance of a telecom contact center and the potential role of big data analytics in overcoming performance bottlenecks	First call resolution, average answer time, transfer rate and customer satisfaction	Big data analytics
(van der Westhuizen and Bezuidenhout, 2017)	To investigate age and tenure effects on the relationship between job demands and resources, and job-related burnout and work engagement	Burnout, work engagement, job demands and job resources, work-related sense of coherence	Multiple hierarchical regression analyses and moderation analyses
(Hsu et al., 2016)	To develop a neural network model to predict call center performance	Service quality, test score, personal attendance, total calls per hour, first call resolution, survey success rate, and customer satisfaction	Neural networks
(Baraka et al., 2015)	To introduce a technique that can support decision makers in the call center industry to evaluate, and analyze the performance of call centers	Success index and gap index	Systematic Analytical Approach
(Phung-Duc and Kawanishi, 2014)	To provide a stationary analysis of a detailed model taking into account the most important human behaviors in call centers.	Waiting time, after-call work, and average number of retrials per customer	Queueing system – Markov chain
(Valle et al., 2012)	To propose an approach to predict the performance of sales agents in a call center dedicated exclusively to sales and telemarketing activities	Logged hours, talked hours, effective contacts, finished records	Naïve Bayesian classifier
(Garcia et al., 2012)	To investigate customers' satisfaction with telephone waiting time	Waiting time and customer satisfaction	Discriminant analysis

(Chen et al., 2011)	To analyze first call resolution performance in order to produce decision rules	First call resolution	Rough set theory
(Cheong et al., 2008)	To identify the KPIs affecting customer satisfaction with a call center's services	Service level, average speed of answer, average time in queue, average answer time, percentage of calls closed on first contact, average abandoned rate, average talk time, adherence to schedule, average after-call work time, percentage of calls blocked, employment turnover rate	Regression analysis
(Jaiswal, 2008)	To examine customer satisfaction and service quality measurement practices followed by call Indian call centers	Average speed answer, abandonment rate, total calls, longest delay, average talk time, average work time after-call, average handle time, service level, queue time, first-call resolution, percentage of calls blocked, calls per agent, adherence, and agent turnover	In-depth interviews to senior managers
(Laureani et al., 2010)	To explore the application of lean six sigma in call centers	First-call resolution ratio, operator turnover and streamlining of processes	Lean Six Sigma methodology
(Robinson and Morley, 2006)	To investigate call center management from the perspective of managers	Customer satisfaction index, level of service, staff turnover rate, abandonment rate, wrap-up time (post call work), call duration or average handling time, occupancy rate and number of calls per agent	Survey of call center managers, followed by in-depth interviews
(Tuten and Neidermeyer, 2004)	To investigate the role of optimism and its effect on stress in call centers	Personal orientation towards optimism, perceptions of job stress, work/non-work conflict, absenteeism and intent to turnover	Survey approach
(Lewig and Dollard, 2003)	To assess the relationship between the emotional demands associated with call center work and call centre worker well-being	Emotional demands, psychosocial demands, rewards, autonomy, social support, emotional exhaustion, job satisfaction	Survey of call center workers
(Grebner et al., 2003)	To analyze working conditions, well-being, and job-related attitudes among call center agents.	Working conditions, well-being, and job-related attitudes	Hypothesis testing (multivariate and univariate analysis of variance)

Table 1 shows that there is no shortage of performance indicators that can be used in the analysis of call center performance. Since more comprehensive and reliable analysis of call center performance is strategic for large call centers, it is important to evaluate which *KPI*s are more relevant, and more or less efficient for the company. Some of the most important *KPI*s used by call center managers are provided in Table 2, together with their definitions.

Table 2: KPIs frequently used by call center managers

KPIs	Description	References

Average Waiting Time	Average time a customer has to wait before contacting an agent; The time a caller waits on the line	(Parmenter, 2015, Ma et al., 2011)
First Call Resolution	The percentage of calls received by the call center in which the customer had their needs met during the first contact without having to transfer, escalate or return the call; The percentage of calls closed on the first connect	(Ma et al., 2011, Jouini et al., 2013)
Average Answer Time	The time between answering a call by agent and terminating the call	(Feinberg et al., 2000)
Abandonment Rate	The proportion of people who disconnect during the operation, either before they are answered by an agent or during a call, usually because they have to wait too long during the transfers; The percentage of callers who disconnect prior to being answered	(Feinberg et al., 2000, Ma et al., 2011)
Average Hold Time	The average time that the caller stays on hold by agent for searching in the system/web or getting help from supervisor	(Parmenter, 2015)
Hold Abandonment Rate	The proportion of customers who were on hold but left the call because of the delay in service	(Mehrbod et al., 2017)
After Call Work	The average time it takes each agent to terminate the process after disconnecting the contact	(Feinberg et al., 2000)
Number of Answered Calls	The number of calls answered by agents in a period of time such as per hour or per day	(Parmenter, 2015)
Customer Satisfaction	The number of customers, or percentage of customers, whose reported experience with a firm, its products or its services (rating) exceeds specified satisfaction goals.	(Farris et al., 2010)

2.2 Shewhart control charts

A control chart is a graph of the values of a statistic of a given quality characteristic that has been measured or computed from a sample plotted against the sample number or time(Montgomery, 2020). The general form of control charts was first developed by Dr. Walter A. Shewhart at Western Electric's Bell Telephone Laboratories beginning in about 1924. A classical Shewhart control chart consists of a center line (*CL*) and two control limits, one on either side of the *CL*, called the upper control limit (*UCL*) and lower control limit (*LCL*). The *CL* represents the average value of the quality characteristic corresponding to the in-control state, and the other two lines are chosen to ensure that, if the process is in-control, nearly all of the sample points will fall between them (Kaya et al., 2017) and exhibit a random pattern. That is, only chance causes are present. One of the fundamental assumptions in the development of the classical Shewhart control chart is the Normality of the underlying distribution of the quality characteristic. Thus, if ω is the sample statistic that measures some quality characteristic of interest with mean μ_ω and standard deviation σ_ω, then the *CL*, the *UCL* and the *LCL* are given by Equation 1 (see(Montgomery, 2020)).

$$UCL = \mu_\omega + k\sigma_\omega$$
$$CL = \mu_\omega \qquad (1)$$
$$LCL = \mu_\omega - k\sigma_\omega$$

where *k* is the distance of the control limits from the centerline, expressed in standard deviation units.

When ω is a Normal random variable, and the process is in-control, the probability that it will fall above the upper control limits is 0.00135 and is the same that it will fall below the lower control limit, resulting in significance level α of 0.27%. Thus, it is possible to conclude that $Z_{\alpha/2} = Z_{0.00135} = 3$. As a result, the control limits provided in Equation 1 are written as shown in Equation 2.

$$
\begin{aligned}
UCL &= \mu_\omega + 3\sigma_\omega \\
CL &= \mu_\omega \\
LCL &= \mu_\omega - 3\sigma_\omega
\end{aligned}
\qquad (2)
$$

Classical Shewhart control chart implementation involves two phases, with different objectives. In Phase I, the main objectives are to estimate the process parameters and to establish the control limits to be used in Phase II. In this phase, the control limits are determined based on the *m* subgroups and the data plotted on the control charts. Points that are outside the control limits are investigated and excluded, since, when a process is in-control, the control limits reflect only common causes of variation. This procedure is iteratively repeated because sometimes this type of retrospective analysis will require several cycles until achieving control limits that represent an in-control process performance. Then, the control chart designed with the control limits of Phase I is used to monitor the process by comparing the sample statistic for each successive sample as it is drawn from the process to the control limits(Montgomery, 2020). Unlike the action taken in Phase I, when a point falls outside the control limits in Phase II, it should not be eliminated and new limits should not be computed. Instead, the cause(s) of the out-of-control situation(s) should be investigated as soon as possible in an effort implement corrective actions and bring the process to the in-control state. This approach avoids the production of non-conforming units in the near future (Requeijo et al., 2014).

Another important aspect in the application of classical Shewhart control charts is the validation of two underlying assumptions. One of them is the Normality assumption, mentioned above. This assumption is easily satisfied when the control charts are developed using sample statistics from subgroups of four or more units. According to the Central Limit Theory, for most probability distributions, the distribution of the sample means is approximately normal for sample sizes of four or more units, even when the individual measurements do not follow a normal distribution. The second assumption, which is more difficult to satisfy and may contribute to a poor performance of the Shewhart control charts, is the one of independence. According to (Montgomery, 2020), the existence of quality characteristics that exhibit even low levels of correlation over time is the main reason that classical Shewhart control charts do not work well. (Montgomery, 2020) also emphasizes that even in situations where the normality assumption is violated to a slight or moderate degree, the control charts will still work reasonably well if the independence assumption is satisfied. Unfortunately, the assumption of independent observations is usually not satisfied in situations where consecutive measurements of quality variables are performed (e.g. chemical processes) or in situations where measurements are performed using automated information systems that collect and record all generated data (*e.g.* call centers). The implementation of classical Shewhart control charts when the independence assumption is not satisfied will give misleading results, with too many false alarms being produced. In other words, the control chart will indicate many out-of-control situations when the process is in an in-control state. One approach that can help to overcome this issue, is to integrate *ARIMA* modeling to remove the autocorrelation from the data and then apply control charts (see Section 3.2).

Classical Shewhart control charts can be grouped into two categories, depending on the type of quality characteristic to be controlled. Quality characteristics that are measured on a continuous scale (e.g. times, volumes, weights, diameters, dimensions) are called variables, and are controlled through control charts for variables. When dealing with this type of quality characteristic, since the process dispersion does not only depend on the central tendency (Pereira and Requeijo, 2012), it is a standard practice to implement two control charts, one to control the mean value of the quality characteristic and other to control its variability. The process average is typically controlled using a control chart for the mean, called the $\acute{X}$ chart. The process variability is usually controlled using a control chart for the standard deviation (S chart) or a control chart for the range (R chart). In situations where the sample used for process control consists of an individual unit ($n = 1$), a control chart for individual measurements (X chart) is used for controlling the process average . In this situation, the process variability is controlled through a control chart for the moving range(MR chart). It is a standard practice to use the moving range of two successive observations to estimate the process variability, where each moving range is defined as $MR_i = |x_i - x_{i-1}|$ (Montgomery, 2020).

On the other hand, characteristics that cannot be conveniently represented in continuous scale are called attributes (e.g. number of non-conforming units, number of defects per unit), and are controlled through control charts for attributes. These include control charts for fraction nonconforming (*p-chart*), control charts for number of nonconformities (*np chart*), control charts for number of defects (*c-chart*) and control charts for number of defects per unit of product (*u-chart*) (for details, see (Montgomery, 2020, Pereira and Requeijo, 2012)).

In this paper, the quality variable to be controlled is the Average Hold Time (*AVGHT*), which is continuous. Measurements of this variable are performed automatically for all calls. As a result, X and MR charts are the rational choice, but applied to residuals and prediction errors (see Section 4.3).

2.3 Control charts for autocorrelated data

As explained in Section 2.2, classical Shewhart control charts are applicable when the measurements of the quality characteristic are independent. If the process data exhibit autocorrelation, at least three approaches are possible (Pereira and Requeijo, 2012): (1) use the classical Shewhart control charts, *CUSUM* or *EWMA*, but with modified control limits to take account of autocorrelation in the process, (2) determine the mathematical model that best fits the autocorrelated data and build the control charts for residuals or prediction errors (Shewhart, *CUSUM* or *EWMA*), and (3) implement specific charts such as the Moving Centre-line *EWMA* (*MCEWMA*) chart or the *EWMAST* (*EWMA* for stationary processes).

In this research, the second approach is chosen, namely the application of Shewhart control charts to residuals and prediction errors, based on individual measurements. This approach, which is based on time series models, has proved useful in dealing with autocorrelated data (Requeijo and Cordeiro, 2013). It consists of modeling the correlative structure of the time series and use that model to remove the autocorrelation from the data, and apply control charts to the residuals and prediction errors (Montgomery, 2020, Pereira and Requeijo, 2012). The random variables represented in the control charts are not the individual measurements X but the residuals (Phase

I) and the prediction errors (Phase II) (Montgomery, 2020, Pereira and Requeijo, 2012). To study the time series of the collected data and determine a mathematical model that best fits this series, we use *ARIMA* modeling (see Section 3.1), using the method described by (Box et al., 2015), which is briefly described in Section 3.3. In most cases, after removing the autocorrelation from the data, the corresponding residuals will be normally and independently distributed, satisfying the two assumptions underlying the implementation of Shewhart control charts. This forms the rationale for choosing the second approach to deal with autocorrelated data.

3. The proposed methodology

Call centers' KPIs have been modeled using different techniques, as shown in Table 1. Despite some of these techniques can be used for predictions, the main advantage of the methodology proposed in this paper is the ability to deal with autocorrelated data. By applying ARIMA modeling, the autocorrelation in the dataset can be effectively modeled and removed. Another advantage is the ability to predict and monitor call centers' KPIS, and also to detect pattern changes. In the following sections, the proposed methodology is explained in detail.

3.1 ARIMA Modeling

As mentioned in Section 2.3, one way to remove inherent autocorrelation in the process data and apply Shewhart control charts in a reliable way is to fit time series models to the quality characteristics to be controlled, using the method described by (Box et al., 2015). *ARIMA* modeling is a widely used and flexible technique that is used to specify a model for time series data that contain autoregressive (*AR*), differencing, and moving average (*MA*) components (see (Box et al., 2015)), *i.e.* time series that are autocorrelated. In short, a time series is autocorrelated if the value of variable *X* at time *t* depends on the past values lagged by one or more periods. In such situations, *ARIMA* is a powerful technique to model the process and determine the mathematical model that best fits the time series under analysis. The specified mathematical model is then used to extract the inherent autocorrelation and then use residuals for forecasting.

A general expression for an *ARIMA* model is

$$\Phi_p(B)\nabla^d X_t = \Theta_q(B)\varepsilon_t \tag{3}$$

where,

$$\Phi_p(B) = \left(1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_p B^p\right) \tag{4}$$

$$\Theta_q(B) = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right) \tag{5}$$

$$B = \frac{X_{t-1}}{X_t} \tag{6}$$

$$\nabla = \frac{X_t - X_{t-1}}{X_t} = 1 - B \tag{7}$$

In Equations 3 – 7, B is the backward shift operator, ∇ is the backward difference operator, d is the differencing order needed to make the process stationary, X_t is the measurement at time t, ε_t is the residual or error at time t, $\Phi_p(B)$ is the autoregressive polynomial of order p and $\Theta_q(B)$ is the moving average polynomial of order q.

To model a process using *ARIMA*, we must determine the *ARIMA (p, d, q)* model that best fits the time series. A standard approach to achieve this goal is to compare the Estimated Autocorrelation Function (*EACF*) with the theoretical Autocorrelation Function (*ACF*) and the Estimated Partial Autocorrelation Function (*EPACF*) with the theoretical Partial Autocorrelation Function (*PACF*). A practical way to perform these comparisons is to design their correlograms and analyze the behavior of the lags. This procedure enables the researcher to identify an appropriate model of the correlative structure of the time series. One important requirement is that the series must be made stationary before we can usefully interpret the correlogram (Chatfield, 2000). When this requirement is satisfied, the process can be modeled as an *AR(p)*, an *MA(q)* or an *ARMA (p,q)*. Table 3**Error! Reference source not found.** summarizes the different behaviors of correlograms for each type of *ARIMA* model.

Table 3: Behaviors of correlograms of stationary processes

Model	Behavior		Mathematical model
	ACF	*PACF*	
AR(p)	Tails off, *i.e.*, decays exponentially or as a damped sine wave to zero	Cutoff after lag p, *i.e.* sudden decays to zero, from lag p + 1	$X_t = \xi + \emptyset_1 X_{t-1} + \emptyset_2 X_{t-2} + \cdots + \emptyset_p X_{t-p} + \varepsilon_t$, with $\emptyset_p \neq 0$
MA(q)	Cutoff after lag p, *i.e.* sudden decays to zero, from lag q + 1	Tails off, *i.e.*, decays exponentially or as a damped sine wave to zero	$X_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$, with $\theta_q \neq 0$
ARMA(p,q)	Tails off, *i.e.*, decays exponentially or as a damped sine wave to zero		$X_t = \xi + \emptyset_1 X_{t-1} + \cdots + \emptyset_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$

In the equations provided in **Error! Reference source not found.**, ξ is the parameter that determines the process mean, $\emptyset_j$ the order parameter j of the *AR* or *ARMA* model, θ_j the order parameter j of the *MA* or *ARMA* model, and σ_ε^2 the variance of residuals.

Once the model that best fits the time series is identified, it is possible to compute the residual values e_t at each time t, as the difference between the value of the quality variable at time t (X_t) and the expected value at the same time t ($\hat{X}_t$). It is also possible to estimate the values of the process parameters, which correspond to Phase I of *SPC*. In Phase II, as the *ARIMA* model that best fits the process data is known, it is possible to predict the future values of the quality variable under analysis. Using the results of these predictions and the new data, the researcher can determine the prediction errors and use them to build the control chart of prediction errors (see Section 3.2).

3.2 Shewhart control charts for residuals and prediction errors

After modeling a process with autocorrelated data, determining the residuals and verifying their independence, the Shewhart control charts can be designed, for Phase I. In this case, the random variable is the residuals instead of the quality variable *X*. The control charts are designed depending on the nature of the collected data. If samples are collected, the sample of size *n*, collected at time *t*, is constituted by the measurements of the quality variable *X* in that time t,

given by X_{t1}, X_{t2}, …, X_{tm}. This sample, in turn, gives rise at the same time t, to a sample constituted by the n residuals e_{t1}, e_{t2}, …, e_{tm}. Based on this sample of n residuals, statistics such as the sample mean $\acute{e}_t$, the sample range R_t and the sample standard deviation S_t can be computed. On the other hand, when individual measurements are collected, the statistics to be considered at time t to control the process mean and the process dispersion are the residual e_t and the moving average MR_t, respectively. As mentioned in Section 2.2, a standard practice for computing the moving range is to use two successive observations, in this case, two successive residuals, as shown in Equation 8.

$$MR_t = |e_t - e_{t-1}| \qquad (8)$$

Once the sample statistics are computed, $\acute{e}_t$ and S_t for mean and standard deviation charts, $\acute{e}_t$ and R_t for mean and range charts respectively, or e_t and MR_t for individual measurements and moving range charts, the corresponding average values $\acute{e}$, $\acute{S}$, $\acute{R}$, $\acute{e}$, and $\acute{MR}$ are determined and used to establish the control limits for Phase I of *SPC*, as illustrated in Table 4**Error! Reference source not found.**. Table 4**Error! Reference source not found.** also provides the estimator for the standard deviation of the residuals, $\hat{\sigma}_\varepsilon$.

Table 4: Control limits for Phase I of residuals control charts and estimators $\hat{\sigma}_\varepsilon$

Residuals charts		LCL	CL	UCL	$\hat{\sigma}_\varepsilon$
$\acute{e}$ and S charts	$\acute{e}$	$-A_3\acute{S}$	0	$A_3\acute{S}$	
	S	$B_4\acute{S}$	$\acute{S}$	$B_3\acute{S}$	$\dfrac{\acute{S}}{c_4}$
$\acute{e}$ and R charts	$\acute{e}$	$-A_2\acute{R}$	0	$A_2\acute{R}$	
	R	$D_4\acute{R}$	$\acute{R}$	$D_3\acute{R}$	$\dfrac{\acute{R}}{d_2}$
e and MR charts	e	$\dfrac{-3\acute{MR}}{d_2}$	0	$\dfrac{3\acute{MR}}{d_2}$	
	MR	$D_4\acute{MR}$	$\acute{MR}$	$D_3\acute{MR}$	$\dfrac{\acute{MR}}{d_2}$

The process parameters and the corresponding control limits for Phase I of *SPC* can only be estimated if the process is under control. If this requirement is not satisfied, the control limits should be reviewed. In contrast to the classical Shewhart control charts, where points corresponding to special causes of variation are eliminated from the time series, when control charts for residuals are applied, such points should not be eliminated. Instead, their values should be replaced by the expected values at the same time t, giving rise to a new time series. The expected values are computed based on the fitted ARIMA model, using the equations provided in Table 5**Error! Reference source not found.**. The process parameters μ and ξ, used in those equations are computed according to the equations provided in Table 5**Error! Reference source not found.**, where, ρ_j is the correlation coefficient of *lag j* and γ_j is the auto-covariance of lag *j*.

Table 5: Expected values of the process parameters (mean and variance) for the quality variable X

Model	Mean	Variance

$AR(p)$	$E(X) = \mu = \dfrac{\xi}{1 - \sum_{j=1}^{p} \emptyset_j}$	$Var(X) = \dfrac{\sigma_\varepsilon^2}{1 - \sum_{j=1}^{p} \rho_j \cdot \emptyset_j}$	
$MA(q)$	$E(X) = \mu$	$Var(X) = \sigma_\varepsilon^2 \cdot \displaystyle\sum_{j=0}^{q} \theta_j^2, \theta_0 = 1$	
$ARMA(p,q)$	$E(X) = \mu = \dfrac{\xi}{1 - \sum_{j=1}^{p} \emptyset_j}$	$Var(X) = \displaystyle\sum_{j=1}^{p} \emptyset_j \cdot \gamma_j - \theta_1 \cdot \gamma_{X_\varepsilon}(-1) - \cdots - \theta_q \cdot \gamma_{X_\varepsilon}(-q) + \sigma_\varepsilon^2$	

After replacing the points corresponding to special causes of variation, it is necessary to compute the new residuals, and update the control charts limits. In situations where a significant number of out of control points is found, it is necessary to investigate the root causes and implement appropriate corrective actions to achieve the process stability. Even if process stability is achieved, before starting the Phase II of *SPC*, capability analysis must be performed in order to check if the process can produce according to the technical specifications. If this requirement is not achieved, the root causes must be identified, and corrective action taken.

Once the process is stable and is shown to be capable, Phase II of *SPC* can be initiated. The objective of Phase II is to monitor the evolution of the quality variable to quickly detect the occurrence of abnormal patterns and take appropriate action. The method to monitor future values of the quality variable, in autocorrelated processes, consists of implementing Shewhart control charts applied to prediction errors. In these control charts, the statistics represented in the charts are the prediction errors, defined by Equation 9.

$$e_\tau(T) = X_{T+\tau} - \mathring{X}_{T+\tau}(T) \qquad (9)$$

The distribution of prediction errors is characterized by an expected value of zero and a variance defined by Equation 10.

$$Var(e|\tau(T)) = \sigma_\varepsilon^2\left(1 + \sum_{j=1}^{\tau-1} \psi_j^2\right) \qquad (10)$$

In Equations 9 and 10, $e_\tau(T)$ is the prediction error for time $T + \tau$, $X_{T+\tau}$ is the value of X at time $T + \tau$, $\mathring{X}_{T+\tau}(T)$ is the prediction for time $T + \tau$, performed at time t, and ψ_j are coefficients determined from $X_t = \psi(B)\varepsilon_t$.

Table 6**Error! Reference source not found.** provides the control limits of prediction errors, where the standard deviation of the prediction errors (σ_{ep}) is given by $\sqrt{Var(e|\tau(T))}$.

Table 6: Control limits for Phase II of prediction error control charts

Prediction error charts		*LCL*	*CL*	*UCL*
$\acute{e}$ and S charts	$\acute{e}$	$-A\sigma_{ep}$	0	$A\sigma_{ep}$
	S	$B_6\sigma_{ep}$	$c_4\sigma_{ep}$	$B_5\sigma_{ep}$
$\acute{e}$ and R charts	$\acute{e}$	$-A\sigma_{ep}$	0	$A\sigma_{ep}$
	R	$D_1\sigma_{ep}$	$d_2\sigma_{ep}$	$D_2\sigma_{ep}$
e and MR charts	e	$-3\sigma_{ep}$	0	$3\sigma_{ep}$
	MR	$D_1\sigma_{ep}$	$d_2\sigma_{ep}$	$D_2\sigma_{ep}$

3.3 Box-Jenkins methodology to build an ARIMA model for call centers KPIs

To identify the mathematical model that best fits the autocorrelated times series investigated in this paper and build the control charts for residuals or prediction errors, the (Box et al., 2015)' methodology has been applied. This classical approach has been successfully employed in different fields to model autocorrelated time series. Building an *ARIMA* model using this method is usually best achieved by a three-step iterative procedure based on identification, estimation and diagnostic checking, as shown in Fig 1. This three-step iterative procedure enables the researcher to identify the mathematical *ARIMA* model that best fits the time series, eliminate the inherent autocorrelation in the time series and then use the model to predict future values of the quality variable, and determine the prediction errors.
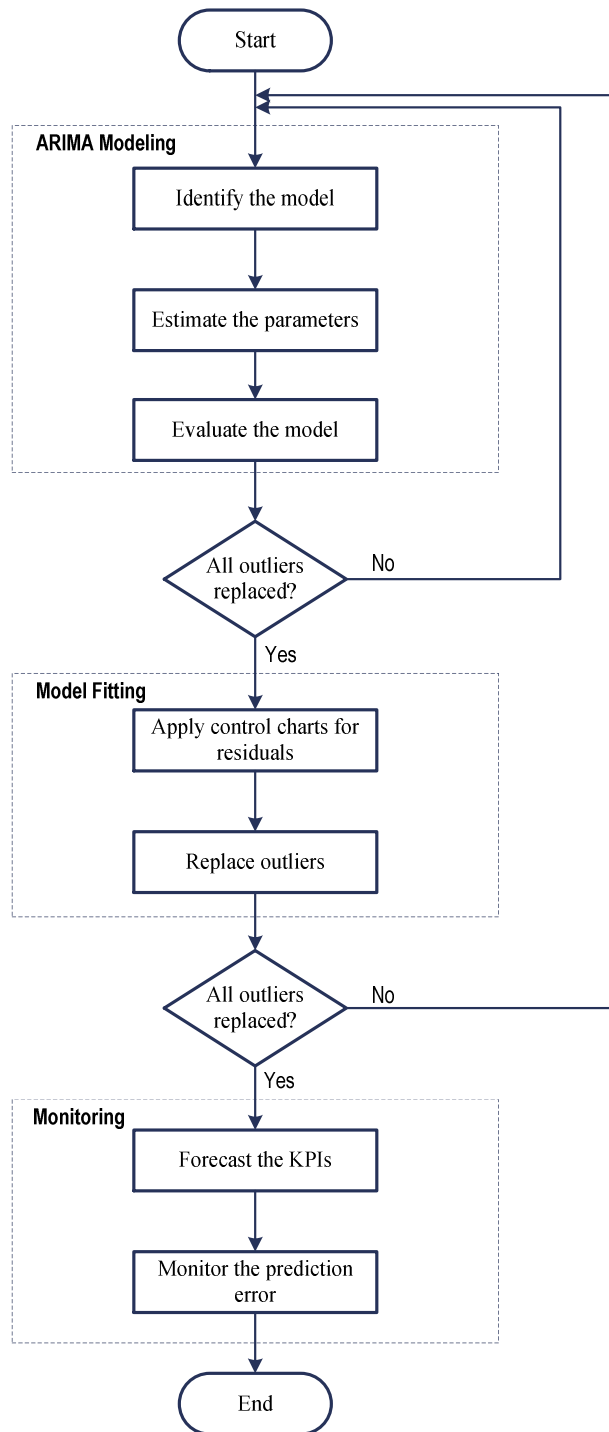
Fig 1. Box-Jenkins methodology to build an ARIMA model for call centers KPIs

In the following paragraphs, each phase of Fig 1 is briefly described:

1. **ARIMA modeling**: In this phase, the first task is to check if the time series is stationary, (see the rationale in Section 3.1). This goal is achieved by plotting and analyzing the time series (t, X_t), which may reveal non-stationary behavior of the process parameters, i.e., mean, variance or both. The second task is to determine the $ARMA(p,q)$ model that best fits the time series. As mentioned in Section 3.1, the best $ARIMA$ model can be identified by comparing the behaviors of the $EACF$ with the ACF and the $EPACF$ with the $PACF$.

The potential behaviors of the *ACF* and *PACF* are summarized in Table 3**Error! Reference source not found.**. The third task is to estimate the parameters $\emptyset_j$, θ_j, ξ, and σ_ε^2. After estimating these parameters, it is necessary to examine their adequacy, *i.e.*, to check if they meet a set of requirements, namely those related to the normality and independence of residuals (see Section 2.2). It is important to test if the model is stationary and invertible, if the estimated coefficients are statistically significant and are independent of each other.

2. **Model fitting**: It is the nature of real-time series to have noisy data and outliers. In fact, when an *ARIMA* model is fitted to a time series, the residuals of the outlier observations are usually higher. To find these outliers, control charts are helpful. By applying control charts to the residuals of the fitted model, it becomes easy to identify data points that do not belong to the distribution of residuals and any other abnormal patterns in the evolution of the residuals. If the control chart shows a significant number of out of control situations, this indicates that the identified model was not well-fitted to the time series. After finding the outliers and/or the out of control points, they must be "removed" from the time series in order to have a clean dataset. Then, a new model that properly fits the cleaned dataset must be identified, since replacing some data points probably changes the model fitted previously. Then a new model is fitted to the new time series. This iterative process continues until the pattern of the residual's evolution exhibits an in-control state.

3. **Monitoring**: After ensuring that the previous requirements are satisfied, we can conclude that the identified model is the most appropriate to model the process under study. The last phase in building an *ARIMA* model is to monitor the prediction errors, which implies the prediction of future values of the quality characteristic and the determination of the corresponding prediction errors (see Section 3.2).

4. Results and discussion

In this section, the methodology based on control charts for autocorrelated data, described in Section 3.3, is applied to forecast and control a call center critical-to-quality (*CTQ*) variable, namely the *AVGHT*. This *KPI* has been chosen because it is considered of great importance to improve the customers satisfaction. In the following section, we describe the steps conducted to build the best *ARIMA* model for forecasting and controlling this *KPI*.

4.1 Data Collection and Pre-processing

The data used in this paper were provided by an inbound call center of a big company in Portugal. They were automatically collected and stored by a *CRM* system. This database includes interactions for one month of call center activity. The data contains more than 175 thousand records of inbound calls recorded call by call. The database stores data on 4 important call center *KPI*s, namely Average Answer Time, Average Hold Time, After Call Work and Customer Satisfaction for each call. "Customer Satisfaction" is an attribute data with binary feature, "Good" or "Bad". This feature specifies the customer view about their conversation, which is solicited by the call center agent at the end of the call. However, in many cases customers refuse to give feedback at the end of the conversation, resulting in missing required data on the "Customer Satisfaction". In the first step of data preparation, the records of the calls with missing data were removed from the dataset. Since a significant number of clients that refused to give their feedback was found in the dataset, the *KPI* "Customer Satisfaction" has not been modeled in this research.

Moreover, the methodology proposed in this paper is not applicable to attribute data such as "Customer Satisfaction".

Regarding the *KPI* "After Call Work", we analyzed the dataset and concluded that more than 95% of calls lasted less than 16 seconds and, approximately 98% of them lasted less than 24 seconds. Therefore, we consider that the potential improvements for this *KPI* would not be significant. Consequently, we excluded this *KPI* from the study. "Answer Time" and "Hold Time" are not only dependent on the agent's performance but also on the type of clients and their questions. In fact, a preliminary descriptive statistics analysis of the individual measurements revealed a significant variation in the dataset of these two KPIs. This suggests that these *KPI*s have a variation that is very difficult to predict and perhaps their measurements do not follow Normal distributions. To confirm this suspicion, we applied the Kolmogorov-Smirnov (K-S) test and, for a significance level $\alpha = 5\%$, we rejected the null hypotheses in both cases. Therefore, we concluded that observations regarding these two *KPIs* do not follow Normal distributions. To achieve a Normal distribution for these two *KPI*s, we computed a new statistic based on their average. It is common practice in call centers to track and monitor the KPIs based on their average. Due to the multiplicity of agents in the call centers, building a model for each agent is time-consuming and may not be feasible in terms of resource management. Consequently, call centers' managers usually track the averages and then, if any abnormal situation is reported, they go future into detail to identify the root causes and take corrective actions. In this research, each average was based on samples data of one working hour. Table 7 shows the descriptive statistics of Hold Time and Answer Time, before and after averaging. As can be seen in Table 7, the standard deviation of both variables decreases significantly after averaging.

Table 7: Hold Time and Answer Time distribution

	STDV	Mean	Min	Max
Hold Time	134.65	99.16	0	3360
Answer Time	235.32	338.28	0	5585
AVG Hold Time/hour	22.04	89.2	0	191.46
AVG Answer Time/hour	36.49	334.74	123.71	544.85

4.2 ARIMA modeling for Average Hold Time

In this section, *ARIMA* modeling is used to predict and control the behavior of *AVGHT*, as mentioned in Section 4. The model is built using a sample of 260 working hours. To predict the future values of this *KPI*, 40 working hours have been used. All the steps to apply *ARIMA* modeling to the *AVGHT* were performed in *STATISTICA* 12. Finally, the prediction results were compared with the real data to evaluate the model adequacy.

As explained in Section 3.3, in the first step we plotted the time series of the 260 observations of the *AVGHT* to confirm that the series was stationary. An analysis of *Fig 2* reveals that the resulting time series is stationary (differentiation level is zero). As a consequence, an *AR (p)*, *MA(q)* or *ARMA(p,q)* can be applied to model the time series (Shumway and Stoffer, 2017) and remove the autocorrelation from the data.
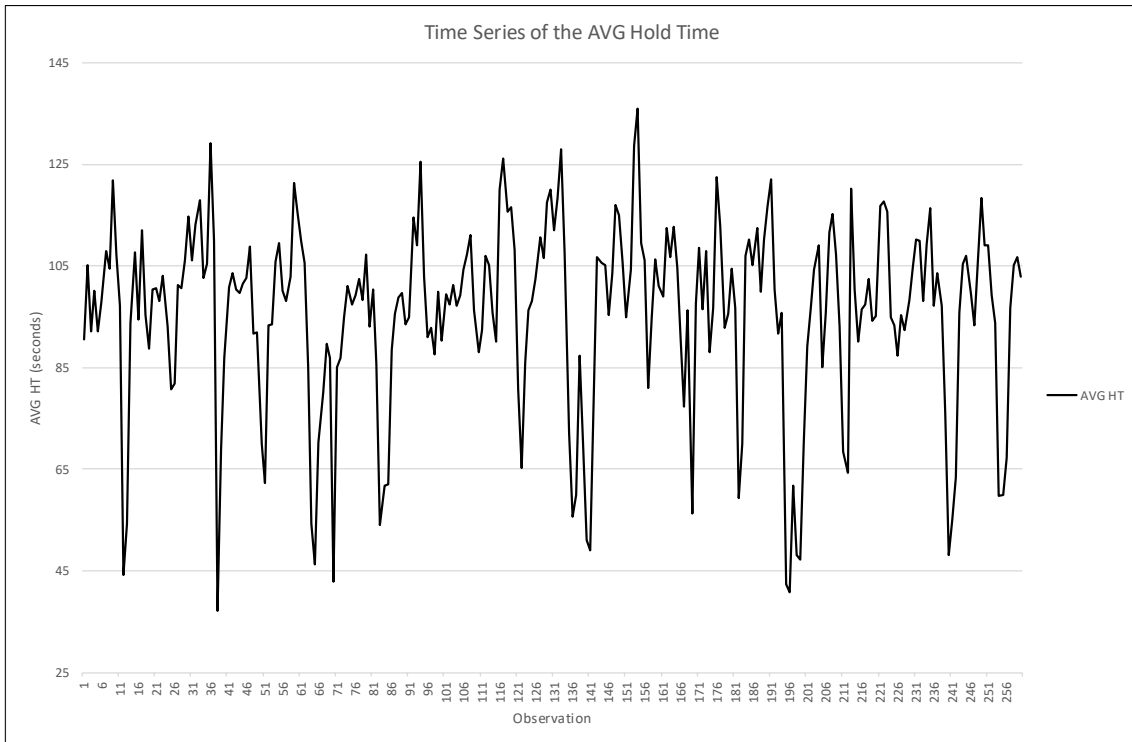
Fig 2. Evolution of the Average Hold Time

In the following step, we examined the autocorrelation by designing and analyzing the *EACF* and the *EPACF* correlograms of the *AVGHT*, shown in Fig 3.
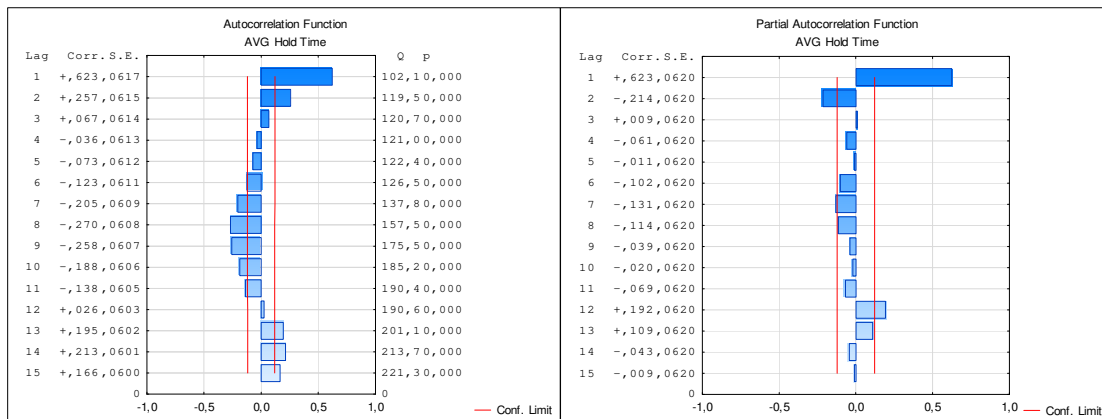

Fig 3. EACF and EPACF of the Average Hold Time

The analysis of the *EACF* and *EPACF* provided in Fig 3 reveals that there are significant correlations in the time series and therefore the classical Shewhart control charts could not be applied directly. It also reveals that *ACF* decays as a damped sine wave to zero and the *PACF* cuts off, both after *lag 2*. As a result, the time series under analysis can reasonably be modeled using an *ARIMA* (2,0,0) or simply a *AR*(2) (see the rationale in Table 3**Error! Reference source not found.**). The estimators of the *AR*(2) parameters were obtained from *Statistica*, and their values are provided in Table 8.

Table 8: Estimated values for the AR(2) parameters

Parameter	Estimated value
μ	95.476
ϕ_1	0.75644
ϕ_2	-0.21394

The parameters provided in Table 8 are used to determine the expected values of *X*, at each time *t*, which are then used to compute the residuals used to fit the *AR*(2) model identified previously.

4.3 ARIMA model fitting

After modeling the process, the identified *ARIMA* model must be fitted. As explained in Section 3.3, real-time series usually include noisy data and outliers. These outliers must be removed from the time series since they do not belong to the distribution that characterizes the time series. A standard approach to find out these outliers and/or any other unusual patterns is to apply control charts. Since the time series under analysis exhibits significant autocorrelation, classical Shewhart control charts cannot be directly applied, as concluded in Section 4.2. In this situation, control charts for residuals should be built, provided that they are independently and normally distributed, that is, they are white noise. The corresponding 260 residuals where obtained from *Statistica*, and have been computed based on the *AR*(2) model identified in the first iteration. With these values, we built the *X* and *MR* charts for residuals, presented in Fig 4.
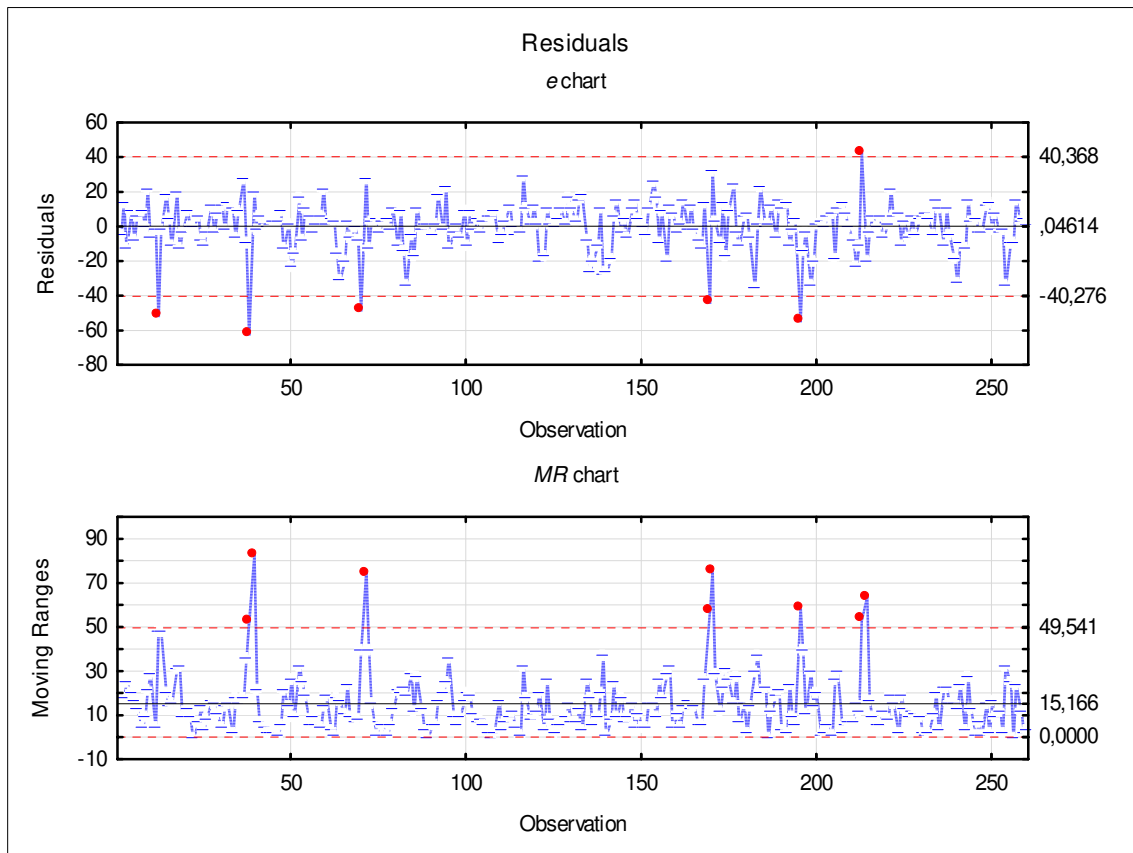

Fig 4. X and MR charts for residuals, Iteration 1

An analysis of the patterns of the *X* and *MR* charts provided in Fig 4 reveals the occurrence of some special causes of variation in both the *X* and *MR* charts. These out of control points indicate that the parameters $\emptyset_1$ and $\emptyset_2$ had changed, implying that the quality variable *X* was out of control. As a result, the process was out of control. As explained in Section 3.2, these out of control points have been replaced by their expected value, determined by the *AR*(2) model identified in Iteration 1. Table 9 summarizes the observations of residuals that indicate out of control situations, their values, and the observed and the expected values of the *AVGHT* characteristic.

Table 9: Control limits for Phase I of residuals control charts and estimators $\hat{\sigma}_\varepsilon$

t	$Residuals$	$AVGHT_t$	$A V \hat{G} H T_t$
12	-50.0386	44.222	94.261
38	62.1473	37.059	99.207
70	-47.4596	42.857	90.317
169	-43.8151	56.200	100.015
195	-54.2707	42.333	96.604
213	42.6320	120.318	77.686

After replacing the out of control observations provided in Table 9 by their expected values, a new time series is obtained. As a result, a second iteration is needed to identify the *ARIMA* model that best fits the new time series. Following the same steps carried out in Section 4.2, we identified the new *ARIMA* model, which is still an *AR*(2), but with different parameters. Based on this model we computed the new residuals and built their *X* and *MR* charts, presented in Fig 5.
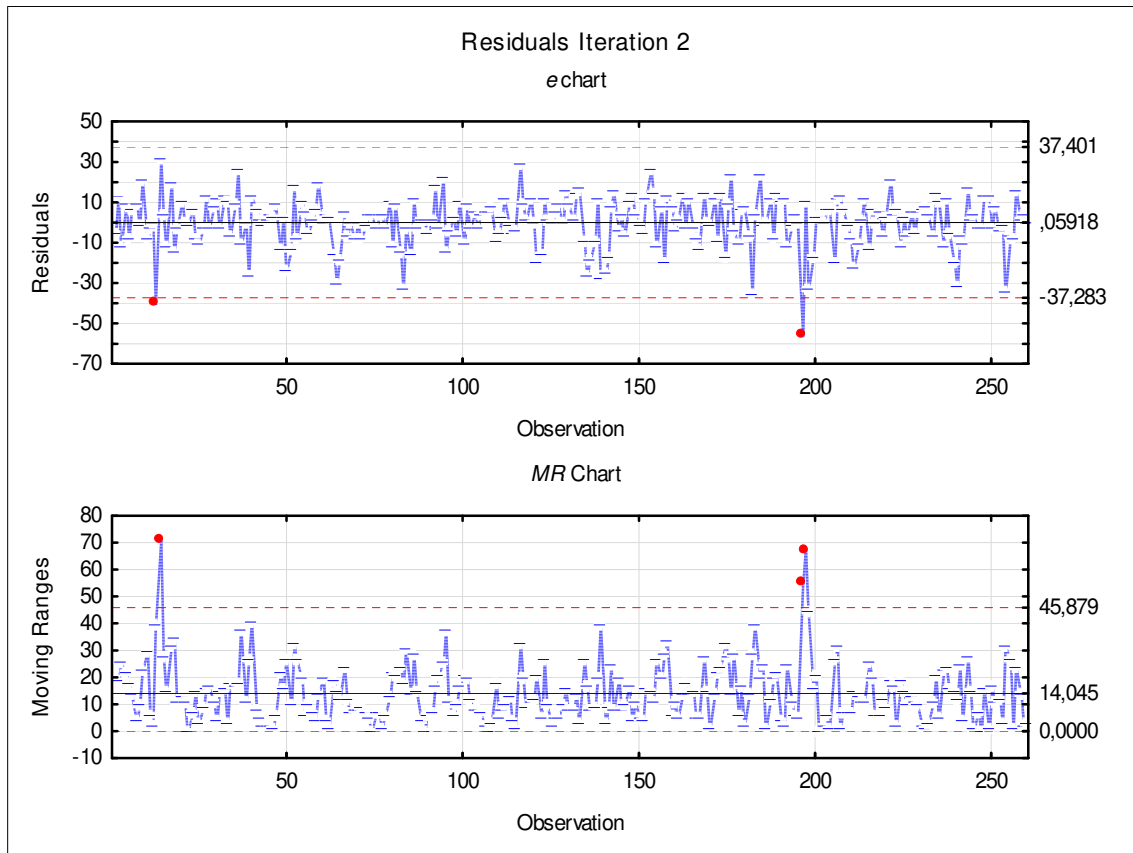

Fig 5. X and MR charts for residuals, Iteration 2

An analysis of the patterns in the *X* and *MR* charts provided in Fig 5 reveals that both charts still exhibit out of control points. Therefore, we performed a third iteration, in which we obtained the residuals plotted in the *X* and *MR* charts, provided in Fig 6.
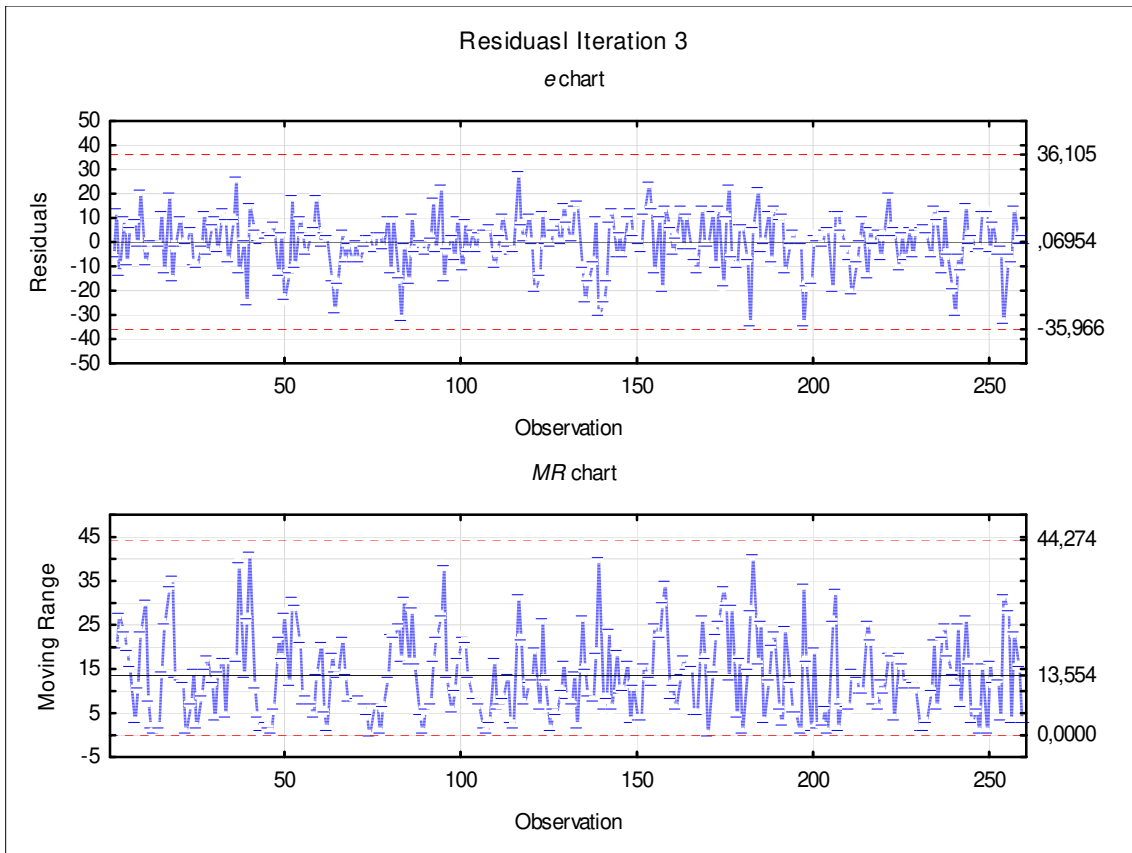

Fig 6. X and MR charts for residuals, Iteration 3

Analyzing the patterns of the charts provided in Fig 6, we can conclude that the process is under statistical control. Once this requirement is achieved, the model that best fits the time series is still an *AR*(2), as can be seen in Fig 7.
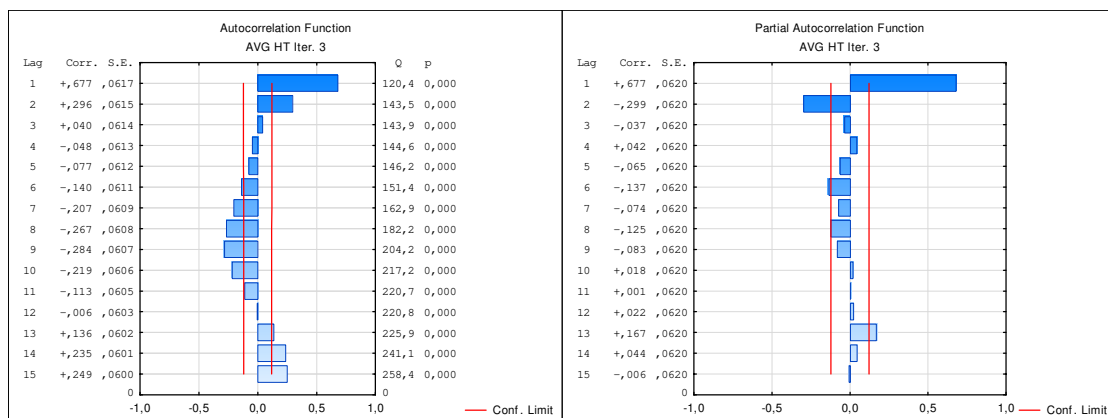

Fig 7. EACF and EPACF of the fitted AR(2) model

Before estimating the parameters of the fitted *AR*(2) model, it must be validated. Achieving an in-control state is not sufficient to validate the fitted *ARIMA* model. As explained in Section 2.3, the

residuals must be independently and normally distributed, with mean zero and constant variance σ_ε^2. Analyzing the X chart provided in Fig 6, it is possible to confirm that the average value of the residuals is approximately zero. Note that the average value of the residuals corresponds to the center line of the X chart. The variance of the residuals is estimated using the estimator $\acute{M}R$, as follow:

$$\sigma_\varepsilon = \frac{\acute{M}R}{d_2} = \frac{13.554}{1.128} = 12.016 \leq \sigma_\varepsilon^2 = 144.384$$

where the value of the constant $d_2 = 1.128$ is obtained from the table of constants for control charts for variables (see (Montgomery, 2020), Appendix VI), considering $n = 2$, which is the number of consecutive observations used to determine each MR_i.

To test the independence of the residuals from the fitted $AR(2)$ model, we designed the correlograms of the $EACF$ and the $EPACF$, shown in Fig 8. The analysis of the correlograms presented in Fig 8 indicates that Lag 9 is slightly outside the confidence interval, in the ACF and the $PACF$. However, we consider that these autocorrelations imply an error that is not significant. Therefore, the hypothesis of residuals independence has not been rejected. Therefore, we consider that the independence assumption is reasonably satisfied.
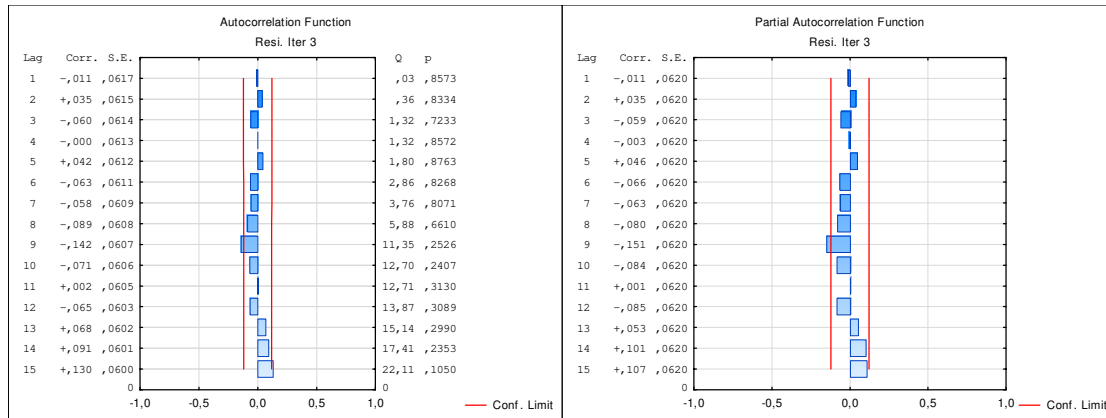

Fig 8. Verification of the Independence of the fitted model residuals

To test the normality of the residuals' distribution, we applied the Kolmogorov-Smirnov (K–S) test. Analyzing the histogram in Fig 9, we find that the K-S test statistic value is $d = 0,02828$. As the total number of observations $N = 260$ is greater than 30 (see (Lilliefors, 1967)), considering a significance level $\alpha = 5\%$, the critical value of the statistic is $Dcrit = \frac{0.886}{\sqrt{N}} = 0.0549$. As $d < Dcrit$, the null hypothesis is not rejected, concluding that there are no evidences that the residuals are not normally distributed.
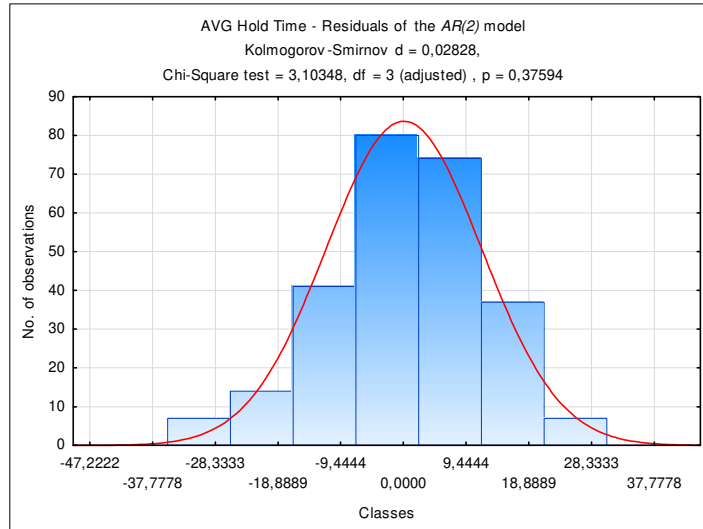
Fig 9. Verification of the normality of the fitted model residuals

Since the requirements for process stability, residual independence and residual normality, are satisfied, we are able to state that the fitted $AR(2)$ model is valid and its parameters are provided in Table 10.

Table 10: Estimated parameters of the fitted AR(2) model

Parameter	Estimated value
μ	96.600
ϕ_1	0.8789
ϕ_2	-0.2986

Mathematically, the fitted $AR(2)$ model is defined as follow:

$$X_t = \xi + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t \leq X_t = 40.543 + 0.8789 X_{t-1} - 0.2986 X_{t-2} + \varepsilon_t$$

where the value of the parameter to determine the process mean, $\xi = 40.543$, was computed based on the $E(X)$ equation for an $AR(p)$ model, provided in Table 5**Error! Reference source not found.**. Using the mathematical model provided above, we computed the predicted values for the next 40 observations on the *AVGHT*, taking into account X_{t-2} and X_{t-1}, the last two observations from the sample of 260 observations used to fit the *ARIMA* model. Having on hand these predicted values, we computed the prediction errors (*e*) and proceeded to process monitoring, Phase II (see following Section).

4.4 Monitoring

In this phase, we applied the Shewhart control charts to the prediction errors, namely the *e* and *MR* charts, as explained in Section 3.2. Table 11 provides the real values of the next 40 observations used to compute the prediction errors.

Table 11: Sample data used to compute and monitor prediction errors

t	Real *AVGHT*	Predicted *AVGHT*	Prediction error
261	90.99497	99.067	-8.072
262	102.3908	96.909	5.482
263	101.5072	96.135	5.372

264	125.8271	96.099	29.728
265	106.7971	96.299	10.499
266	102.288	96.485	5.803
267	115.0254	96.589	18.437
268	38.44828	96.624	-58.176
269	163.5833	96.625	66.958
270	48.44444	96.615	-48.170
…	…	…	…
300	106.40299	96.600	9.803

To monitor these prediction errors provided in Table 11, we computed the control limits, using the equations for the e and MR charts, provided in Table 5**Error! Reference source not found.**. The standard deviation of the prediction errors, σ_{ep}, is computed using Equation 10, which led us to determine the coefficients ψ_j of the polynomial $\psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + \cdots$

As the $AR(2)$ model is defined by $\phi_2(B)X_t = \varepsilon_t$ and, considering that $X_t = \psi(B)\varepsilon_t$, we obtain $\phi_2(B)\psi(B) = 1$. Developing this equation, we obtain

$$(1 - \emptyset_1 B - \emptyset_2 B^2)(\psi_0 + \psi_1 B + \psi_2 B^2 + \cdots) = 1$$

$$\psi_0 + \psi_1 B + \psi_2 B^2 + \cdots - \emptyset_1 \psi_0 B - \emptyset_1 \psi_1 B^2 - \emptyset_1 \psi_2 B^3 + \cdots - \emptyset_2 \psi_0 B^2 - \emptyset_2 \psi_1 B^3 - \emptyset_2 \psi_2 B^4 + \cdots = 1$$

$$\psi_0 + (\psi|1 - \emptyset_1 \psi_0)B + (\psi|2 - \emptyset_1 \psi_1 - \emptyset_2 \psi_0)B^2 - (\emptyset_1 \psi_2 + \emptyset_2 \psi_1)B^3 - \emptyset_2 \psi_2 B^4 + \cdots = 1$$

The values of the coefficients ψ_j are then, given by

$$\psi_0 = 1$$

$$\psi_1 - \emptyset_1 \times 1 = 0 \leq \psi_1 = \emptyset_1 = 0.87862$$

$$\psi_2 - \emptyset_1 \emptyset_1 - \emptyset_2 \times 1 = 0 \leq \psi_2 = \emptyset_1 \psi_1 - \emptyset_2 \psi_0 \leq \psi_2 = \emptyset_1^2 + \emptyset_2 = 0.87862^2 - 0.2984 = 0.4736$$

$$\psi_j = \emptyset_1 \psi_{j-1} - \emptyset_2 \psi_{j-2} \quad (11)$$

Having determined the coefficients ψ_j and the σ_{ep}, it is possible to compute the control limits of the e and MR charts, applying the corresponding equations provided in Table 5**Error! Reference source not found.**. The values of these control limits, the coefficients ψ_j, and σ_{ep}, at each time t, are summarized in Table 12.

Table 12: Summary of the data used to build the prediction errors' chart

t	ψ	σ_{ep}	LCL_e	CL_e	UCL_e	LCL_{MR}	CL_{MR}	UCL_{MR}
$\tau = 260$	1	−	−	−	−	−	−	−
261	0.8789	10.65229	-31.9569	0	31.9569	0	12.0158	39.2643
262	0.4739	14.18182	-42.5455	0	42.5455	0	15.9971	52.2742
263	0.1540	15.05336	-45.1601	0	45.1601	0	16.9802	55.4867
264	-0.0061	15.14253	-45.4276	0	45.4276	0	17.0808	55.8154
265	-0.0514	15.14267	-45.4280	0	45.4280	0	17.0809	55.8159

266	-0.0433	15.15255	-45.4577	0	45.4577	0	17.0921	55.8523
267	-0.0227	15.15958	-45.4787	0	45.4787	0	17.1000	55.8782
268	-0.0070	15.16151	-45.4845	0	45.4845	0	17.1022	55.8853
269	0.0006	15.16170	-45.4851	0	45.4851	0	17.1024	55.8860
270	0.0026	15.16170	-45.4851	0	45.4851	0	17.1024	55.8860
271	0.0021	15.16173	-45.4852	0	45.4852	0	17.1024	55.8861
272	0.0011	15.16174	-45.4852	0	45.4852	0	17.1024	55.8862
273	0.0003	15.16175	-45.4852	0	45.4852	0	17.1025	55.8862
…	…	…	…	…	…	…	…	…
300	0.0000	15.16175	-45.4852	0	45.4852	0	17.1025	55.8862

Using the data provided in Table 12, we built the control chart shown in Fig 10, which will be used to monitor the prediction errors provided in Table 12.
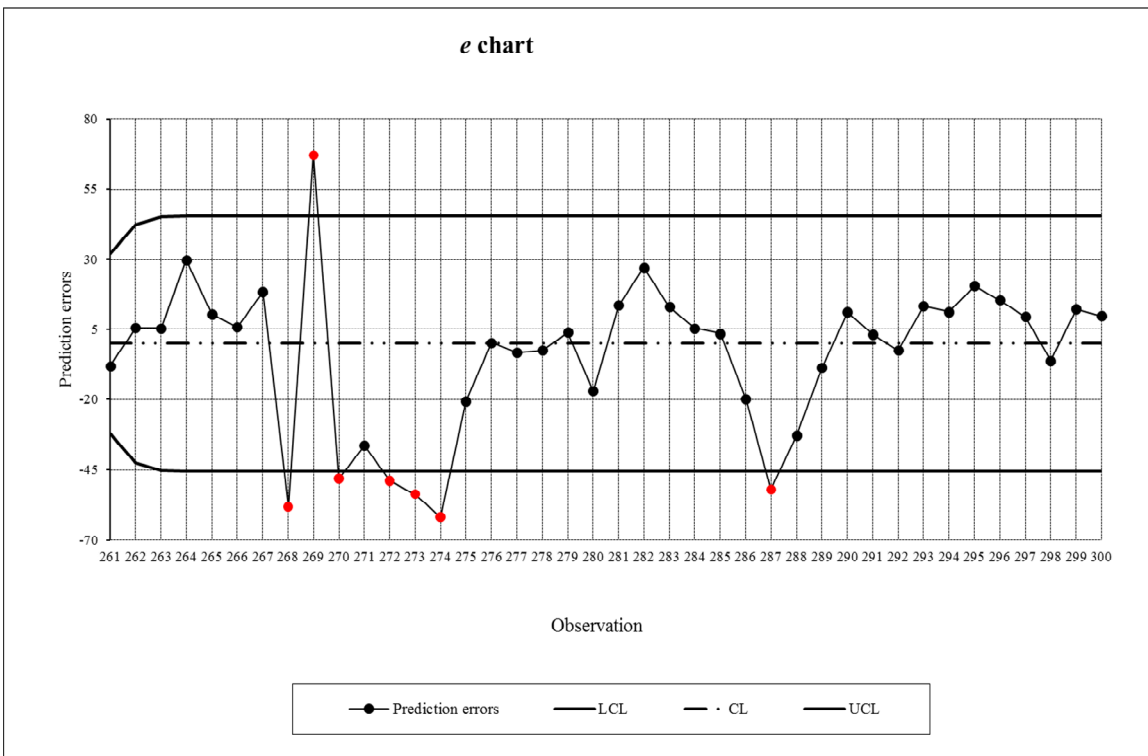

Fig 10. e chart for monitoring prediction errors

4.5 Discussion

The analysis of the patterns of the control chart illustrated in Fig 10 reveals the existence of some special causes of variation. However, apart from observation 269, the out of control points under the *LCL* suggest that the performance of the call center has improved. Prediction errors are computed from the difference between the observed value and the predicted value for a given time *t*, implying that, in the context of this study, prediction error is a "*small-is-better*" characteristic. In other words, the lower the value of prediction error, the higher is the performance. According to this rationale, the analysis of the chart provided in Fig 10 suggests that something desirable has happen at time 268, 270, 272, 273, 274 and 287, which resulted in observed values that are significantly less than the expected ones (short Hold Time is desirable to increase the customer satisfaction). From this perspective, the appropriate approach that the call

center manager should adopt is to investigate the root causes of such performance improvement and try to replicate them in the future, in order to achieve higher performance. On the other hand, the data point over the *UCL* suggests that the observed Hold Time at time 269 was significantly higher than the expected value, which is undesirable. In that situation, the call center manager should also investigate the root causes of such bad performance and take corrective actions in order to improve performance and bring the process to an in-control condition. This may have been the approach that was actually adopted by the call center manager. In fact, the run from time 270 to 274 suggests that after the occurrence of an unacceptable value at time 269, effective action was taken that resulted in a significant decrease in the Hold Time and consequently improved performance. If this is, in fact, what happened, such improvement action should be implemented for similar situations in the future.

5.   Conclusions

This paper presents a methodology for modeling and controlling the performance of call centers, which integrates ARIMA modeling and SPC. Process modeling through ARIMA models, in addition to being able to model the pattern of a certain characteristic of a process, in conjunction with SPC make it possible to monitor of the random behavior (residual) of that characteristic. In this example the characteristic is AVGHT. The integration of ARIMA and SPC is an approach that has not been widely used, particularly in processes where autocorrelated data are generated, as is the case for call centers. However, as suggested by the case study results reported here, the joint approach ARIMA/SPC provides a set of effective tools that can be used to model autocorrelated processes and monitor their behavior, and support decision making in the daily operations of call centers. The control chart for prediction errors proves to be an effective way to routinely monitor the evolution of the *AVGHT* characteristic and to detect pattern changes due to abnormal situations. Another important conclusion is that, for the sample data used in Phase II, the call center under analysis performed very well. However, the monitoring of the *AVGHT* characteristic must be an ongoing activity in order to test if this good performance was something temporary or if it is standard. This ongoing activity is also necessary to identify any undesirable sources of pattern changes and continuously decrease the *AVGHT*.

Although the proposed methodology is intended to forecast and control call center KPIs, it can easily be extrapolated to other industries where autocorrelated data are generated (e.g. chemical industry). As mentioned in Section 1, nowadays a massive amount of data has been generated in most industries. Therefore, effective tools are needed to extract knowledge from those massive datasets. However, there are also some difficulties that may difficulty its implementation, namely:

1. The complexity of ARIMA modeling;
2. The process operation data are often dynamic. As a result, periodically, the analyst should test if the model must be adjusted. In other words, it is necessary to test if: (2.1) the model remains the same or has changed (e.g., it can change from an *AR(1)* to an *AR(2)*, or eventually to a *MA(q)* or an *ARMA(p,q)*; (2.2) the values of the model parameters have changed significantly (e.g. $\emptyset_1, \emptyset_2, \ldots$);
3. Deciding how frequently the model should be adjusted is not trivial. We suggest that it should only be adjusted when significant changes in the process parameters (mean and variance) are detected;
4. There is some complexity inherent to the calculation of prediction errors (see Section 4.4).

In terms of relevance to theory, this is the first time that the joint approach ARIMA/SPC has been used to model call center performance. The main contribution of this joint approach is that it incorporates the monitoring phase (Phase II – see Section 4.4), something that has not been done in previous studies. In addition, this paper provides detailed guidelines on how to model processes where autocorrelated data are generated, how to identify the model that best fits the data, and how to compute and monitor prediction errors. With regard to the practical contribution, the proposed methodology is intended to inform policies and decision-making in call centers, i.e. to monitor the performance of their agents and quickly detect pattern changes due to abnormal situations. By quickly detecting such abnormal situations, managers can take prompt and appropriate action to correct the process and, at the same time, prevent the occurrence of such undesirable situations in the future. On the other hand, when a pattern change has a positive impact on the performance of the call center (e.g. decreasing trend in the AVGHT), managers can quickly investigate the reasons and try to replicate them in the future. This will help to improve the performance of call centers and increase customer satisfaction.

As is known, nowadays companies are recording data related to every activity and consequently producing massive amounts of data. One of the biggest challenges is the possibility of having autocorrelation in the dataset. The methodology proposed in this paper can be a valuable tool, not only to deal with such potential autocorrelation in the dataset but also to effectively monitor and quickly detect the changing pattern behaviors. This will support managers in predicting critical situations before happening and take appropriate actions to maintain the desired level of process performance.

Despite its potential theoretical and managerial contributions, this paper has an important limitation. Namely, a single case study was conducted and only one call center KPI was modeled. For these reasons, the conclusions may not generalized. Future research should apply and test the proposed methodology in more call centers and other services/industries where autocorrelated data are common to improve the generalization. In such future studies, more KPIs must also be modeled.

References

BARAKA, H. A., BARAKA, H. A. & EL-GAMILY, I. H. 2013. Assessing call centers' success: A validation of the DeLone and McLean model for information system. *Egyptian Informatics Journal,* 14**,** 99-108.

BARAKA, H. A., BARAKA, H. A. & EL-GAMILY, I. H. 2015. Information systems performance evaluation, introducing a two-level technique: case study call centers. *Egyptian Informatics Journal,* 16**,** 9-22.

BOX, G. E., JENKINS, G. M., REINSEL, G. C. & LJUNG, G. M. 2015. *Time series analysis: forecasting and control*, John Wiley & Sons.

CHATFIELD, C. 2000. *Time-series forecasting*, CRC press.

CHEN, R.-R., CHIANG, Y.-I., CHONG, P. P., LIN, Y.-H. & CHANG, H.-K. 2011. Rough set analysis on call center metrics. *Applied Soft Computing,* 11**,** 3804-3811.

CHEONG, K., KIM, J. & SO, S. 2008. A study of strategic call center management: relationship between key performance indicators and customer satisfaction. *European Journal of Social Sciences,* 6**,** 268-276.

EXCOFFIER, M., GICQUEL, C. & JOUINI, O. 2016. A joint chance-constrained programming approach for call center workforce scheduling under uncertain call arrival forecasts. *Computers & Industrial Engineering,* 96**,** 16-30.

FARRIS, P. W., BENDLE, N., PFEIFER, P. & REIBSTEIN, D. 2010. *Marketing metrics: The definitive guide to measuring marketing performance*, Pearson Education.

FEINBERG, R. A., KIM, I.-S., HOKAMA, L., DE RUYTER, K. & KEEN, C. 2000. Operational determinants of caller satisfaction in the call center. *International Journal of Service Industry Management,* 11**,** 131-141.

GARCIA, D., ARCHER, T., MORADI, S. & GHIABI, B. 2012. Waiting in vain: managing time and customer satisfaction at call centers. *Psychology,* 3**,** 213.

GREBNER, S., SEMMER, N., FASO, L. L., GUT, S., KÄLIN, W. & ELFERING, A. 2003. Working conditions, well-being, and job-related attitudes among call centre agents. *European Journal of Work and Organizational Psychology,* 12**,** 341-365.

HSU, H.-H., CHEN, T.-C., CHAN, W.-T. & CHANG, J.-K. Performance Evaluation of Call Center Agents by Neural Networks. 2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA), 2016. IEEE, 964-968.

JAISWAL, A. K. 2008. Customer satisfaction and service quality measurement in Indian call centres. *Managing Service Quality: An International Journal.*

JOSE, B., RAMANAN, T. R. & KUMAR, S. M. Big data provenance and analytics in telecom contact centers. TENCON 2017-2017 IEEE Region 10 Conference, 2017. IEEE, 1573-1578.

JOUINI, O., KOOLE, G. & ROUBOS, A. 2013. Performance Indicators for Call Centers with Impatience. *IIE Transactions,* 45**,** 359-372.

KAYA, I., ERDOĞAN, M. & YıLDıZ, C. 2017. Analysis and control of variability by using fuzzy individual control charts. *Applied Soft Computing,* 51**,** 370-381.

KAYA, I. & TURGUT, A. 2021. Design of variable control charts based on type-2 fuzzy sets with a real case study. *Soft Computing,* 25**,** 613-633.

KHASHEI, M., BIJARI, M. & HEJAZI, S. R. 2012. Combining seasonal ARIMA models with computational intelligence techniques for time series forecasting. *Soft computing,* 16**,** 1091-1105.

LAUREANI, A., ANTONY, J. & DOUGLAS, A. 2010. Lean six sigma in a call centre: a case study. *International journal of productivity and performance management.*

LEWIG, K. A. & DOLLARD, M. F. 2003. Emotional dissonance, emotional exhaustion and job satisfaction in call centre workers. *European journal of work and organizational psychology,* 12**,** 366-392.

LILLIEFORS, H. W. 1967. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American statistical Association,* 62**,** 399-402.

MA, J., KIM, N. & ROTHROCK, L. 2011. Performance assessment in an interactive call center workforce simulation. *Simulation Modelling Practice and Theory,* 19**,** 227-238.

MEHRBOD, N., MEHRBOD, A., GRILO, A., VASCONCELOS, C. & SILVA, J. L. 2017. A call-center KPIs analysis: Do hours of the day affect the call center performance? *Proceedings of International Conference on Computers and Industrial Engineering, CIE47; Costa da Caparica, Lisboa, Portugal, 1113th October, 2017*

MONTGOMERY, D. C. 2020. *Introduction to statistical quality control*, John Wiley & Sons.

ÖZDEMIR, A. 2021. Development of fuzzy $\bar{X}$-S $\bar{X}$-S control charts with unbalanced fuzzy data. *Soft Computing,* 25**,** 4015-4025.

PARMENTER, D. 2015. *Key performance indicators: developing, implementing, and using winning KPIs*, John Wiley & Sons.

PEREIRA, Z. & REQUEIJO, J. 2012. Qualidade: Planeamento e Controlo Estatístico de Processos (Quality: Statistical Process Control and Planning). *FCT/UNL Foundation Editor, Lisboa.*

PHUNG-DUC, T. & KAWANISHI, K. I. 2014. Performance analysis of call centers with abandonment, retrial and after-call work. *Performance Evaluation,* 80**,** 43-62.

QUESENBERRY, C. P. 1997. *SPC methods for quality improvement*, Wiley.

REQUEIJO, J. G. & CORDEIRO, J. 2013. Implementation of the statistical process control with autocorrelated data in an automotive manufacturer. *International Journal of Industrial and Systems Engineering,* 13**,** 325-344.

REQUEIJO, J. G., PUGA-LEAL, R. & MATOS, A. S. Road Map to the Statistical Process Control. Proceedings of the Eighth International Conference on Management Science and Engineering Management, 2014. Springer, 949-957.

ROBINSON, G. & MORLEY, C. 2006. Call centre management: responsibilities and performance. *International Journal of Service Industry Management*.

SHUMWAY, R. H. & STOFFER, D. S. 2017. *Time series analysis and its applications: with R examples*, Springer.

SOLIMAN, O. S., ELATIF, M. M. A. & GHAREEB, E. T. Studying the impact of abandonment on call centers system performance. 2012 8th International Conference on Informatics and Systems (INFOS), 2012. IEEE, BIO-105-BIO-114.

TUTEN, T. L. & NEIDERMEYER, P. E. 2004. Performance, satisfaction and turnover in call centers: The effects of stress and optimism. *Journal of business research,* 57**,** 26-34.

VALLE, M. A., VARAS, S. & RUZ, G. A. 2012. Job performance prediction in a call center using a naive Bayes classifier. *Expert Systems with Applications,* 39**,** 9939-9945.

VAN DER WESTHUIZEN, S. & BEZUIDENHOUT, L. 2017. Work-related well-being in a call centre environment: The moderating role of age and tenure. *Journal of Psychology in Africa,* 27**,** 216-220.

VILLALOBOS, K., SUYKENS, J. & ILLARRAMENDI, A. 2021. A flexible alarm prediction system for smart manufacturing scenarios following a forecaster–analyzer approach. *Journal of Intelligent Manufacturing,* 32**,** 1323-1344.

YANG, W.-A. & ZHOU, W. 2015. Autoregressive coefficient-invariant control chart pattern recognition in autocorrelated manufacturing processes using neural network ensemble. *Journal of Intelligent Manufacturing,* 26**,** 1161-1180.

ZAN, T., LIU, Z., WANG, H., WANG, M. & GAO, X. 2020. Control chart pattern recognition using the convolutional neural network. *Journal of Intelligent Manufacturing,* 31**,** 703-716.