www.revistageintec.net
ISSN: 2237-0722
revista GEINTEC
Gestão, Inovação e Tecnologias

DEFECT MONITORING IN IRON CASTING USING RESIDUES OF AUTOREGRESSIVE MODELS

MONITORAMENTO DOS DEFEITOS EM FERRO FUNDIDO USANDO RESÍDUOS DE MODELOS AUTOREGRESSIVOS

Vanusa Andrea Casarin[1]; Adriano Mendonça Souza[2]; Jaime Alvares Spim[3] *(in memorian)*

[1]Regional Integrated University of High Uruguay and Missions (URI), Santo Ângelo, RS
School of Engineering, Technology Center (UFRGS), Porto Alegre, RS
vanusa.casarin@gmail.com
[2]Department of Statistics, Department of Industrial Engineering (UFSM), Santa Maria, RS
amsouza.sm@gmail.com
[3]School of Engineering, Technology Center, UFRGS, RS

Abstract

The purpose of this study is to monitor the index of general waste irons forecasting nodular and gray using the residues originated from the methodology Box & Jenkins by means of X-bar and R control charts. Search is to find a general class of model ARIMA (p, d, q) but as data have autocorrelation is found to the number of residues which allowed the application of charts. The found model was the model SARIMA (0,1,1)(0,1,1)$_{12}$*. In step of checking the stability of the model was found that some comments are out of control due to temperature and chemical composition.*

Key-words: control charts; time series; residues; Box & Jenkins methodology; index of general waste.

1 INTRODUCTION

Process statistic control techniques are widely used in productive processes for the monitoring and improvement of the product quality and services. When the quality control charts

are applied, one must take into consideration that the observation process is independent and normally distributed with constant mean and variance (MONTGOMERY, 1997). Process observations are often autocorrelated which affects control charts performances, resulting in many points out of control limits (TSAI; CHIANG; WU, 2004).

Thus it is extremely important to identify the independence of the observations or their autocorrelation, before applying control charts for the process monitoring since an inadequate control charts that produces false alarms in excess, will end up being discarded or will only be kept for the company's formal demands, being that the alarms will simply be ignored by the people involved in the process (COSTA; EPPRECHT; CARPINETTI, 2005; MONTGOMERY; MASTRANGELO, 1991). Therefore, if the process observations are autocorrelated a model of time series can be adjusted for the data of the process under study and then, use the forecasting residues to develop the control charts (ALWAN; ROBERTS, 1988; HARRIS; ROSS, 1991).

At this research, the forecasting and monitoring of the gray and nodular cast iron dross general rate manufactured by a foundry industry in Santa Rosa for firms in the region, will be done. This study aims at monitoring and forecasting the dross and the general rate of the nodular and cast iron using the residues of a forecasting model derived from Box & Jenkins methodology, through X-bar control charts for the mean and R for the amplitude.

The forecasting of iron dross amount becomes extremelly important for the company, because besides the quality the final product must have, it is a way for the company to know how much material, power, labor and equipment is being wasted. It is through forecasting that measures like doing business, stock control, and the aquisition of raw material among other factors can be planned and optimized. According to Makridakis, Wheelwright and Hyndman (1998), accomplishing the forecasting of needs is important in order to help determine the necessary resources for activities developed by the company. Based on the given data the strategic planning will be done and the position in the market in which it fits will also by defined.

The gray iron is the one used to manufacture materials with low mechanical resistance and in the company it is used to manufacture pulleys for harvesters, pulley for elevators in grain transportation and stoker among others. The nodular iron is superior in mechanical features to the gray iron and inferior to steel. It has several applications, in the company being, studied, as in the hydraulics part, lifting of cargos specially the tractor ones, as well as in the railroad, highway and agricultural lines.

The paper is organized as follows. Section 2 presents a description of the Box & Jenkins methodology, X-bar ($\overline{X}$) and R Control Charts and Defects in processes of casting for nodular and gray iron. In Section 3 is made to apply the methods described in Section 2 and is the best model to predict where the experimental results are shown and, finally, in Section 4 the major conclusions are drawn.

2 MATERIALS AND METHODS

The employed methodology to develop this research was initially the one to find a model of the general class ARIMA (p,d,q) which represented the series under study in order to obtain the values for the monitoring of the general rates of the cost and nodular iron dross. However, as the data in analysis are auto correlated, the direct use of these values in control charts are not possible, once basic postulations of independence are violated. Then, after the best representative model is found for the general rate series of gray and nodular iron dross, used by the company under study, the residue series was found in which data show up non correlated, with mean zero and constant variance enabling this way, the application of X-bar and R control charts for the monitoring of dross rates which provide information about a certain process based on samples periodically collected from the process.

The series general rate of dross daily control in the foundry refers to the period – July 2004 to May 2008. After ARIMA model is found, the residues of this model will be used for later application of control charts, once this way non-correlated variables are obtained with the use of the statistics program “Statistica 7.0”. Foundry defects are mainly related to labor problems with consequence in process deviation in moulding and trimming off, machinery, soldering temperature and chemical composition.

The raw material to be reused due to defects totalizes 5% for each amount of kg controlled during the month, in an approximate value of 8,700 kg.

Box & Jenkins Methodology

A methodology used a lot in the analysis of parametric models is known as Box and Jenkins (1970) approach. This methodology consists in adjusting autoregressive integrated moving average ARIMA (p,d,q) to a set of data where p represents the autoregressive part, d represents the number

of differences given in the series to stabilize it and q represents the moving average part. Box & Jenkins models are mathematical models that capture the behavior of seriate correlation or autocorrelation, in order to accomplish future forecasting which once well-modeled will provide good forecasting for the series under study. The goal is to identify and estimate an statistical model that can be interpreted as having generated the sample data, if this estimated model is used for forecasting, one must suppose that this model features are constant in time and mainly in the future period (GUJARATI, 2000).

According to Morettin and Toloi (2004), the building of Box & Jenkins models is based on an iterative cycle, in which the model choice is done based on the data themselves. Iterative cycle stages are represented in Figure 1.

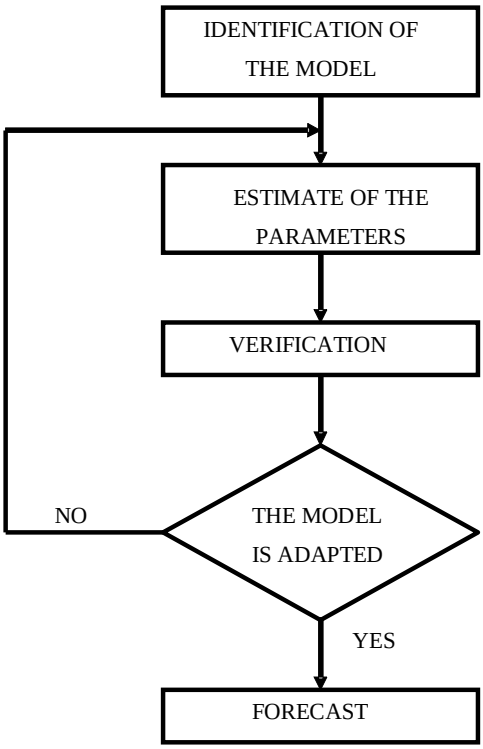

Figure 1 - Flowchart of iterative cycle of Box and Jenkins
Source: Adapted of Pereira e Requeijo (2008)

Box & Jenkins models are based on stochastic processes and must present some features so that they can be estimated:

- The observations will have to be spaced in equal intervals in time and the series will have to have al least 50 observations;

- Datum series will have to be stabilized, that is stochastics features along the series are kept the same, like for example, the same means, variance and covariance. This way, the means of the parts must not be meaninfully different from the mean of the whole sample;

- The series can also be homogeneous non-stabilized by differentiation or integration process. A differentiated variable is usually represented by the symbol 1; this way the variable y with the first difference will be $\Delta^2 y = y_t - y_{t-1}$, that is, y differentiated is equal to the values of y in the anterior period (t-1). Its order can be represented by the letter d. In that case, for the first difference, $d = 1$, for the second difference we would have $d = 2$ and $\Delta^2 y = y_{t-1} - y_{t-2}$.

The general formula of an ARIMA model, considering that the series needed to be differentiated once to stabilize it, can be represented the following way:

$$W_t = \Delta y = y_t - y_{t-1}. \quad (1)$$

$$W_t = \phi_1 W_{t-1} +,...,+\phi_p W_{t-p} + \xi_t - \theta_1 \xi_{t-1} -,...,-\theta_q \xi_{t-q}. \quad (2)$$

This way, the model can be expressed by: ARIMA (*p*,*d*,*q*), where *p* represents the autoregressive part order, *q* represents the moving average order and *d* represents the number of differentiations accomplished in order to stabilize them. When the models do not present differentiation in their series, they can still be called ARMA (*p*,*q*); when they only present the autoregressive part, as AR(*p*) or when they present only the moving average parts, MA(*q*).

Some advantages of using Box & Jenkins models are:

a) The model is chosen among a big variety of types.

b) The choice is based on the behavior of the series observation itself.

c) The model is parsimonious considering the number of parameters.

d) Obtained observations are more precise.

Among the disadvantages, the most relevant one is the determination of the forecasting model, once it has been the biggest obstacle to its use, depending on the user's sagacity and experience.

Next some steps that should be followed when the model is identified are shown, that is the determination of the orders *p* (autoregressive part), *d* (differentiation quantity) and *q* (moving average part). In practice a parsimony concept which depends on the researcher's sensitiveness is used, but that invariably does not allow assuming values of *p*, *d* and *q* higher than 2. Evidently in some cases the values can be higher mainly *p* and *q* values, so, the steps to be followed are:

- Determining if the series is stabilized: the function of autocorrelation chart (FAC) and of its confidence interval is used. If FAC presents an accute decrease in its first values, it means the series is stabilized and $d = 0$. If the decrease is mild until it reaches zero, it means that it is not stabilized and it will have the value d determined by the differentiation number. At each differentiation a new stabilizing test is done. If after some successive differentiations it does not become stabilized, the series will not be homogeneous and ARIMA models will not be applicable.

- Determining the autoregressive part order (p value) through the partial autocorrelated function chart, (PACP). FACP value numbers that are above the confidence interval will be considered the maximum value that p can reach. The parsimony criterion must always be observed. FAC and FACP are easily obtained by software commands specialized in temporal series, as for example, the Statistica one.

- Determining the moving average part order (*q* value) through the function of autocorrelation (FAC). As in the autoregressive part, the number of FAC values that are above their confidence intervals represent the maximum value that *q* will be able to reach.

- Steeming the possible model combinations using the maximums for *p,d,q* found in prior stages. This way, there is a set of models that must be chosen by some statistical criteria together with the *p* value, which will indicate, in case $p<0,05$, that the proposed model is meaningful.

- After steeming the probable models, it has to be checked which the best model is, for that, the model diagnosis must be done through the following criteria:

- Checking the presence of white noise. Through the residue FAC and FAC chart it is checked if they are within the confidence interval, meaning there is white noise, that is, there is a series of independent random variables identically distributed (IID) with mean zero and constant variance. Whenever a temporal series is being modeled there is the search for white noise presence.

- Applying adjusting criteria (Statistical ones) of esteemed models Akaike (AIC – Akaike Information Criteria), and Bayes (BIC – Bayesian Information Criteria), criteria are equivalent and measure the likelihood of the model. These criteria are calculated through the following formulas:

AKAIKE Criterion (AIC) → $AIC = \ell n\ \hat{\sigma}_e^2 + \frac{2(p+q)}{n}$, (3)

Bayes Criterion (BIC) → $BIC = \ell n\ \hat{\sigma}_e^2 + \frac{(p+q)\ell nn}{n}$, (4)

Where $\hat{\sigma}_e^2$ is the stemmed variance of errors; n is the sample size and p and q are parameter values. For that, these criteria are use together with the white noise checking, in order to choose the best model.

X-bar ($\overline{X}$) and R Control Charts

Control charts of the mean, X-bar and the amplitude chart, R, are, usually used when the interest quality feature is expressed by a number in continuous scale of measure. X-bar chart is used with the purpose of controlling the process mean, while the R chart is employed to control the variability of a variable along the time (MONTGOMERY, 1997).

For the construction of X-bar and R control chart it is necessary, first, to choose the variable to be analyzed. Next, it is also defined the sampling method, if it is instantaneous or periodical, and the size of the sample to be used. After, datum collection is performed and the central value is determined for the mean and amplitude respectively (MONTGOMERY, 1997; MITRA, 1998).

$$\overline{\overline{x}} = \frac{\overline{x}_1 + \overline{x}_2 + ... + \overline{x}_m}{m} = \frac{1}{m}\sum_{i=1}^{m}\overline{x}_i \,, \tag{5}$$

with $i = 1,2,...,m$, that sample is the average of i-ésima:

$$\overline{x}_i = \frac{x_{i1} + x_{i2} + ... + x_{in}}{n},$$

$$\overline{R} = \frac{R_1 + R_2 + ... + R_m}{m} = \frac{1}{m}\sum_{i=1}^{m} R_i \,, \tag{6}$$

with $i = 1,2,...,m$ that sample is the amplitude of i-ésima.

Having the value of $\overline{\overline{x}}$ and $\overline{R}$, the superior limit of control (LSC), the inferior limit of control (LIC) and the mean limit (LM) can be calculated for the variable under study, where the chart for the process mean is represented by equations (7), (8) and (9) and for the process amplitude by equations (10), (11) and (12).

$$LSC\overline{x} = \overline{\overline{x}} + A_2\overline{R}, \tag{7}$$

$$LM\overline{x} = \overline{\overline{x}}, \tag{8}$$

$$LIC\overline{x} = \overline{\overline{x}} - A_2\overline{R}, \tag{9}$$

$$LSCR = D_4\overline{R}, \tag{10}$$

$$LMR = \overline{R}, \tag{11}$$

$$LICR = D_3 \overline{R}. \quad (12)$$

In equations (7), (8), (9), (10), (11) and (12), the chart control limits are represented. Constants A_2, D_4 and D_3 are the factors that very according to the sub-group size.

In order to evaluate if a process is under control, the observations $\overline{x}_i$ and R_i, i = 1,2,...m are represented in the corresponding charts and the disposition of the point is analyzed at these charts. If all the points are within the control limits and no special cause is evident, it can be said that the process was under control when preliminary samples were taken. This way, experimental control limits are appropriate to control future and present production. They can be adapted by the ones responsible for the process control, once the condition of control reached is suitable for the process, having in mind technical and economic considerations (Montgomery, 1997).

Defects in processes of casting for nodular and gray iron

According to Chiaverini (2005), by knowing the diagram Fe-C one can define casting iron as an iron-carbon-silicon alloy, where carbon has higher proportions than 2%, in superior amount to the one that is retained in solid solution in the austenite so that it results in partially free carbon in the form of veins or graphite lamellas.

In the general denomination of casting iron there are a few kinds of alloy:

Gray casting iron: microstructurally it presents itself with carbon in the free form, graphite in morphology of lamellas, carbon in the combined form and Cementite (Fe_3C). In this case the main alloy elements are the carbon and the silicon. Its fracture surface presents a dark gray color.

Nodular gray iron: In the microstructural form it presents itself as free carbon in the module morphology which gives this kind of iron superior mechanical features to those of the soft casting iron. It is obtained by chemical changes in material composition in liquid state. Its fracture surface presents a silvery color.

3 RESULTS AND DISCUSSION

The analyzed data represent 888 observations referring the dross rate of the casting iron produced by a Brazilian industry from July 2004 to May 2008.

Initially the serial dependence structure between the observations through autocorrelation and partial autocorrelation functions was checked, because in case it exists, the direct application of

control charts is not possible, once in order for them to be applied they have to show observation independence and normal distribution. This way an alternative for the variable evaluation in study is its modeling through ARIMA models (p,d,q) and the use of the series residues which will reflect datum behavior and that present the necessary features for the application of control charts, after that it is only finding a model that is white noise.

The models competing for the dross general rate that present white noise are described in Table 1, being that the best model is chosen, the one that presents the best values for the penalty criteria AIC and BIC.

Table 1 - ARIMA models for the variable dross general rate, with their respective parameters and AIC and BIC criteria.

Models	Parameters	AIC	BIC
SARIMA (0,1,1)(0,1,1)$_{12}$	**q(1)=0,818503** **Qs(1)= 0,875114**	**-5,335863482**	**-5,338115734**
SARIMA (0,1,1)(2,1,0)$_{12}$	q(1)=0,873637 Ps(1)=-0,689257 Ps(2)= -0,331738	-5,172517958	-5,17477021
SARIMA (0,1,2)(2,1,0)$_{12}$	q(1)=0,814637 q(2)=0,069063 Ps(1)=-0,682470 Ps(2)= -0,329207	-5,170265705	-5,17477021
SARIMA (1,1,1)(2,1,0)$_{12}$	p(1)=0,085584 q(1)=0,895084 Ps(1)=-0,681730 Ps(2)= -0,329118	-5,175895	-5,1804

Source: Author own

The best model was a SARIMA (0,1,1)(0,1,1)$_{12}$ where the residues of this model are going to be used to outline X-bar and R charts, as shown in Figure 2 and 3 respectively, with the purpose of verifying process stability.

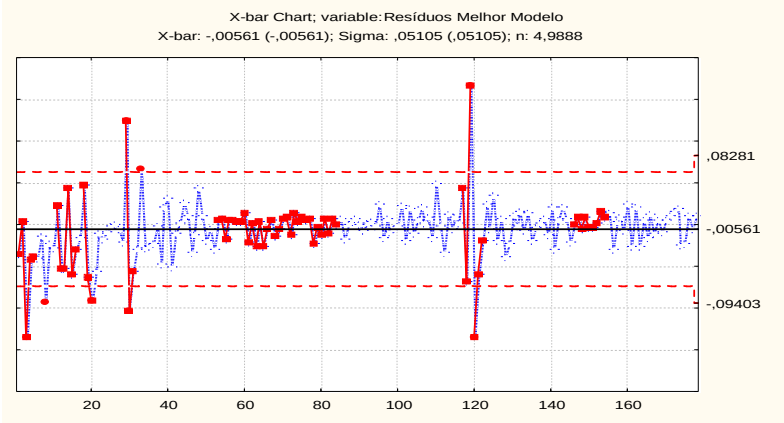

Figure 2 - X-bar chart carried through with the residues of the SARIMA
Source: Author own

According to the residues of the best model found for X-bar chart, three zones are presented in which their drawn points in the chart suggest process instability in the casting of the nodular and gray irons, using the run tests. Zone A with sigma 3,000, zone B with sigma 2,000 and zone C with sigma 1,000. It was detected that 9 samples are in the same side of the center going from 146 to 154 and 6 samples are in decreasing lines; 14 samples alternate up and down going from 71 to 84 samples; 2 to 3 samples located in zone A or out of it vary from 11 samples to 122; 4 to 5 samples in zone B or out of it varying from 1 sample up to 16 and 15 samples in zone C varying 53 to 82 samples.

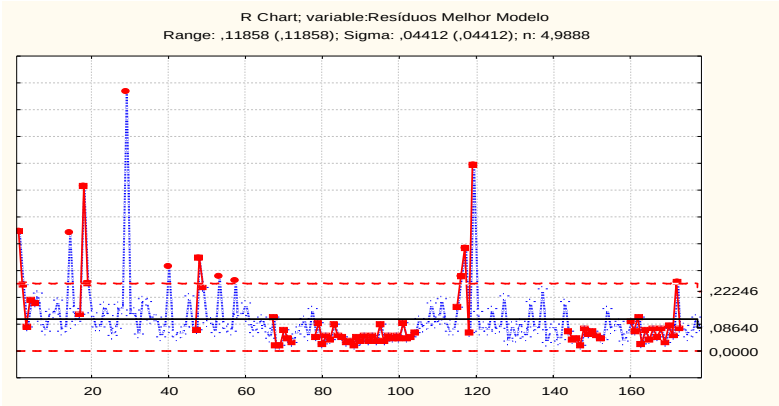

Figure 3 - R chart of the residues of the SARIMA
Source: Author own

Observing that the process has few points out of control, it is still valid to forecast for 12 days ahead, as table 2.

Table 2 - ARIMA models for variable index of general waste, with its respective parameters and criteria AIC and BIC.

Period	Forecast	Inferior 90%	Superior 90%	Standard error
8/5/2008	0,051190	-0,063385	0,165766	0,069584
9/5/2008	0,048245	-0,068202	0,164692	0,070721
10/5/2008	0,043278	-0,075011	0,161567	0,071840
11/5/2008	0,040020	-0,080083	0,160123	0,072941
12/5/2008	0,065406	-0,056485	0,187296	0,074027
13/5/2008	0,049726	-0,073925	0,173378	0,075096
14/5/2008	0,060045	-0,065343	0,185433	0,076151
15/5/2008	0,050560	-0,076541	0,177660	0,077191
16/5/2008	0,040701	-0,088090	0,169491	0,078217
17/5/2008	0,046115	-0,084344	0,176573	0,079230
18/5/2008	0,045218	-0,086888	0,177323	0,080230
19/5/2008	0,058041	-0,075691	0,191773	0,081218

Source: Author own

It is important to remember that the forecasting are in percentage, thus in order to obtain the correct forecasting value it would be necessary to multiply each forecasted value by 100. Also, in order to get the esteemed damage it would be good to multiply this dross percentage by the costs it generated to produce it.

4 CONCLUSION

For the monitoring of the dross general rate forecasting to be feasible, it was necessary to treat data and then control them through X-bar and R charts, once they presented autocorrelation. This way Box and Jenkins methodology was used to find SARIMA model $(0,1,1)(0,1,1)_{12}$ where the parameters were all statistically meaningful and residue analysis showed that the adjusted model was suitable. Thus model residues were used in order to accomplish the process monitoring once using original data would be impossible to obtain the correct interpretation give that the data were auto correlated.

Acknowledgements

To Indústria de Fundição Metalúrgica Candeia Ltda, for providing the data and to CNPq for the partial financial support for the development of this research.

REFERENCES

ALWAN, L.C.; ROBERTS, H.V. Time Series Modeling for Statistical Process Control. **Journal of Business & Economic Statistics**, n. 6, p. 87-95, 1998.

BOX, G.E.P.; JENKINS, G.M. **Time Series Analysis**: Forecasting and Control. San Francisco, Holden-Day (Revised Edition), 1970.

CHIAVERINI, V. **Aços e ferros fundidos**. ABM, 2005.

COSTA, A.F.B.; EPPRECHT, E.K.; CARPINETTI, L.C.R. **Controle Estatístico de Qualidade**. 2. ed. São Paulo: Atlas, 2005.

GUJARATI, D.N. **Econometria Básica**. 3. ed. São Paulo: Makron Books, 2000.

HARRIS, T.J.; ROSS, W.H. Statistical process control procedures for correlated observations. **The Canadian Journal of Chemical Engineering**, v. 69, p. 48-57, 1991.

MAKRIDAKIS, S.; WHEELWRIGHT, S.; HYNDMAN, R.J. **Forecasting methods and applications**. 3. ed. New York: John Wiley & Sons, 1998.

MITRA, A. **Fundamentals of quality control and improvement**. 2. ed. Prentice-Hall, 1998.

MONTGOMERY, D.C. **Introduction to statistical quality control**. 3. ed. New York: John Wiley & Sons, Inc., 1997.

MONTGOMERY, D.C.; MASTRANGELO, C.M. Some Statistical Process Control Methods for Autocorrelated Data. **Journal of Quality Technology**, n. 23, p. 179-204 (with discussion), 1991.

MORETTIN, P.A.; TOLOI, C.M.C. **Análise de séries temporais**. 1. ed. Edgard Blücher, 2004.

PEREIRA, Z. L.; REQUEIJO, J. G. **Planejamento e controlo estatístico de processos**. Faculdade de Ciências e Tecnologia – FET. Universidade Nova de Lisboa, 2008.

TSAI, T.; CHIANG, Y., WU, S. Conditional maximum likelihood estimation for control charts in the presence of correlation. **Brazilian Journal of Probability and Statistics**, december, v. 18, n. 2, p. 151-162, 2004.