Master production program based on the modelling and analysis of a time series as a statistical technique for forecasting demand

Henry Omar Paipa Álvarez
Department of Industrial Processes
Universidad Francisco de Paula Santander, Cúcuta, Norte de Santander, Colombia

Wlamyr Palacios Alvarado
Department of Industrial Processes
Universidad Francisco de Paula Santander, Cúcuta, Norte de Santander, Colombia

Byron Medina Delgado
Department of Electrical and Electronics
Universidad Francisco de Paula Santander, Cúcuta, Norte de Santander, Colombia

Abstract- This article presents the modelling and analysis of a time series of the demand for a product with a high contribution margin in terms of profits for the company, with the objective of designing a master production program based on input data such as resource availability, production capacity, quantities to be manufactured, production rate, costs, inventories and the forecast of demand in a given planning horizon. For this purpose, the sales of product P during a sample of 24 months were considered, making a descriptive analysis of the behavior of this variable, and identifying the components of the demand and the quantitative method to be used. Thus, a regression analysis was applied by the method of least squares to estimate the linear relation between the variables, the seasonality index was calculated and season variations were discounted from the original data. Finally, it was possible to generate a projection of pessimistic forecast and a projection of optimistic forecast, with the objective of providing a forecast in terms of interval and the real value can oscillate between such intervals. This forecast served as the main input and from it the master production program of product P was elaborated.

Keywords – Production plan, time series, statistical forecasting, linear regression, production engineering.

I. INTRODUCTION

The basic concepts inherent in a material requirement program system have been known for many years, even before they were actually used. Prior to the emergence of the Material Requirements Program (MRP) almost all companies employed reorder point variations, where inventory was allowed to be reduced to a specified quantity, considered the minimum allowable point before ordering replenishment of a standard number of units [1]. The MRP system is the longest developed and explicitly addresses dependent demand. It excels when dependent demand is "more irregular", meaning that it occurs sporadically and production is in large batches [2]. As a complement to this process there is capacity planning. This process has a hierarchical structure, from demand estimates to MRP [3]. It is therefore necessary to project and know this demand. In this study demand planning is developed through the application of the analysis of a time series, from this data a production plan or program is elaborated. An adequate programming proposal is presented to manage an inventory efficiently, that is, not to have excess or deficit what is needed in order to satisfy the demand and thus be able to give a better service to the client. Within the results of the present investigation a proposal of production planning is generated for the company and its product P., for this projection two scenarios are presented, one pessimistic and the other optimistic, intervals within which the possibility of reality is found, being a mathematical approximation of the data, since, the impulse in sales and others, is due to factors like the development of marketing activities.

II. MATERIALS AND METHODS

The process begins with the exploratory analysis of the time series described as a set of historical data on the demand for product P, a product that presents a contribution margin in terms of high profits from a family of products that the company manages. These data start from the month of October 2013 to September 2015, provided by the General

Management, responsible for planning inventories, demand, and materials, among others. Figure 1. Shows the demand behavior of the monthly product P for the period described above.

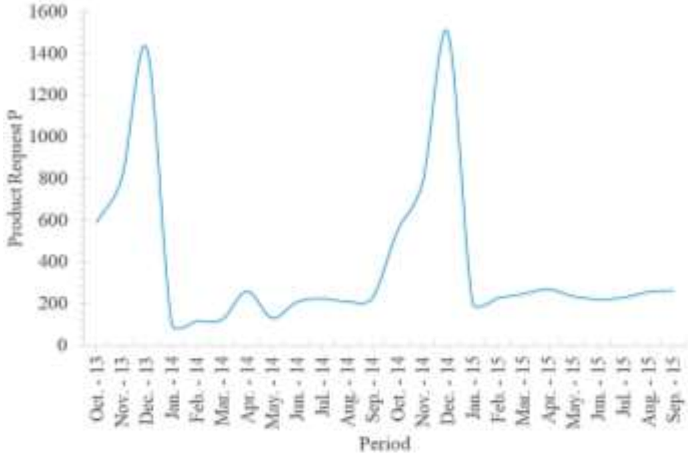

Figure 1. Demand behavior of monthly product P for the period October 2013 to September 2015. For the development of the master program, input data such as inventory, costs, workday hours and others were considered, which were shown in the results and discussion.

III. Experiment and Result

In order to carry out a complete study of the time series, the analyses were grouped into four stages as follows: Descriptive analysis, modelling and estimation of the model, diagnosis and validation, production program. Each one with a defined objective, from which are derived the mathematical, statistical and analytical understanding of the series. Next, the processes considered in each of the stages of the study are presented.

3.1. First stage. Descriptive analysis –

In this first phase, the qualitative description of the series was carried out, identifying the components of the series such as trend and seasonality. The descriptive analysis of the series allows us to know the road to travel to determine the model and its estimates.

3.1.1 Description of the series –

The series presented in Figure 1, clearly shows the existence of a seasonal component. The repetitive behavior in the last quarter of each year respectively represented in peaks, so it can be stated that the series has seasonal component of period every four quarters (October, November, December). From Figure 1, it can be deduced that in the months of October, November and December 2013 and 2014, i.e. the fourth quarters of each year, the highest sales are recorded, as well as in the months of April 2013 and 2014, due to the celebration of the International Children's Day in Colombia, It can be noted that in 2014 sales were in the order of 200 units, however in the year 2015, have decreased markedly and the units sold oscillate below 200 units compared to the previous year, it could be said that it is due to the economic situation that is going through the city of Cucuta [4].

The management of the company uses qualitative methods and the experience of thirty years of work in the sector to predict the quantity of products to be sold and manufactured, however, in the present study a quantitative method was used to try to predict a demand interval that allows us to know how much to manufacture for the 2015 season, knowing this data will answer questions such as. How much raw material to ask for? Through a plan of material requirements and an adequate inventory management, the demand forecasts are the support to support the answers to the questions previously mentioned [5].

3.1.2. Identification of the components of the series –

In order to determine the seasonal component, a first approximation was made using moving averages of order 12, resulting in an additive seasonal scheme that requires deseasonalization of the data. To define the function of simple correlation or correlogram, it was practiced the methodology of [6] who expresses that, the coefficient of linear correlation between the variables Z_t and Z_{t-1}, is called coefficient of linear autocorrelation of order 1 (r_1) whose calculation needs to obtain the time series offset by a time unit. The calculation of the value of the coefficient of linear autocorrelation was obtained by means of Equation (1). Sample correlation.

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \quad (1)$$

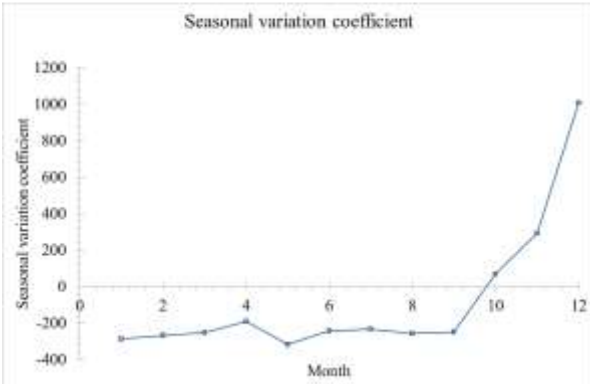

Figure 2. Component of the trend.

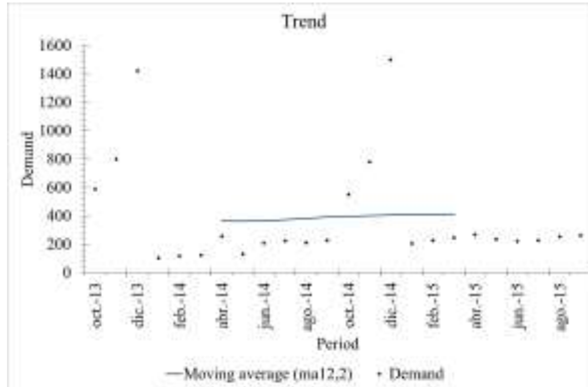

Figure 3. Seasonal component of the time series.

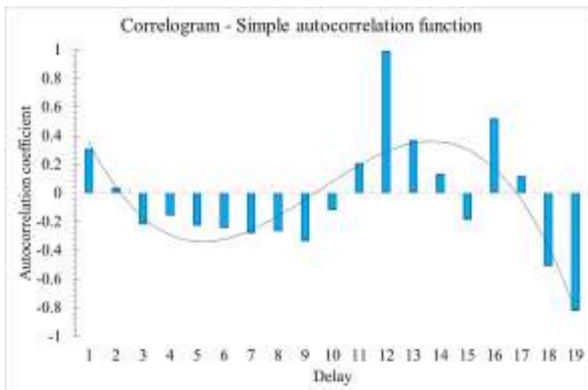

Figure 4. Linear autocorrelation component of order 19 (r_{19}).

Applying Equation (1) it was obtained that, component of the trend. According to Figure 2, the series does not present a clear trend, and also maintains seasonal effects. Therefore, a deterministic (linear) trend does not seem suitable for the Z_t series. Analyzing the behavior of the data, it becomes evident that, seasonal component of the time series. According to Figure 3, the time series presents a pattern of seasonal behavior for the months of October, November and December, in other words, for the last quarter of the year. The seasonal variation coefficients were calculated, presenting positive values for the mentioned quarter. This can be interpreted as the quarter with the greatest need or demand for product P in the market.

According to Figure 4, the Z_t series has been decalibrated up to 19 time units. With the help of Microsoft Excel ® the autocorrelation coefficients of order 1 (r_1) between the Z_t and Z_{t-1} variables have been obtained, up to order 19 between the Z_t and Z_{t-19} variables. Where it is observed according to Figure 4, the correlogram shows that the autocorrelation structure decreases and grows in instants of time in a cyclic manner, which means that the dependence of the series with time is less or greater according to seasonal behavior.

3.1.3. Quantitative method to be used. –

For the determination of the quantitative method, the guide presented by [7] was used, where the method of linear regression by least squares is selected according to the amount of historical data, the pattern of data, and the forecast horizon.

3.2. Second stage. Modelling and estimation of the model –

The following is a brief description of the modelling and estimation process of the model, starting with the calculation of the seasonality index, the discounting of seasonal variations, and the design of the linear regression model from non-stationalized data, *i.e.* those from which seasonal variations have been subtracted. The results of the calculations can be seen in Table 1. Non-seasonal demand [8].

Due to the fact that the data present a seasonal behavior, and that in addition the data presented are in months, a quarterly conversion was made, as can be seen in Table 1. This conversion is necessary to better understand the pattern of seasonality presented in the data, where quarter I is comprised of the months of January, February and march, in quarter II; April, May, June, quarter III; July, august and September, quarter IV October, November and December. To convert the months to quarters, the demand for each category was added.

From the results, the seasonal component is evident in quarters IV, which require a greater demand of product P. Once demand is seasonally adjusted, there is a softening of the peaks that generate seasonality in the series; on the other hand, a linear trend behavior is observed that is directly proportional among the variables. Similarly, Equation (5) was solved for ($\widehat{y_i}$) in each period, the behavior of the seasonally adjusted forecast is similar to the real one [9]. So, we proceeded to calculate the seasonality index and to subtract seasonal variations from the original data, as can be seen in Table 1.

Table 1. Results of the calculations for the deseasonalization of demand.

Period	Year	Quarter	Actual demand (Z_t)	Average demand per season	Seasonal index	Non-stationalized demand (Y)	Forecast data with linear regression	Seasonalised forecast
1	2013	IV-2013	2810	2820	2.40	1170	984	2364
2	2014	I-2014	344	511	0.44	790	1038	452
3	2014	II-2014	598	662	0.56	1061	1093	616
4	2014	III-2014	662	704	0.60	1105	1147	687
5	2014	IV-2014	2830	2820	2.40	1178	1201	2885
6	2015	I-2015	678	511	0.44	1558	1255	546
7	2015	II-2015	725	662	0.56	1287	1310	738
8	2015	III-2015	745	704	0.60	1243	1364	817
9	2015	IV-2015	N/A	2820	2.40	N/A	1418	3407
10	2016	I - 2016	N/A	N/A	0.44	N/A	1472	641
11	2016	II - 2016	N/A	N/A	0.56	N/A	1527	860
12	2016	III - 2016	N/A	N/A	0.60	N/A	1581	947

When we relate seasonally adjusted demand (y), we can observe the behavior of the trend pattern, behavior that was generated once the seasonality was partially eliminated in the time series. It is evident that the trend is of the directly proportional type, with a pronounced ascent in the demand, therefore we proceeded to perform the calculations with the method of linear regression by method of least squares according to [10] determine the linear equation, as well, least squares estimators for the simple linear regression model, Equation (2) indicates the linear base mathematical model, Equation (3) indicates how the estimators are calculated, and Equation (4) is the union of the base equation and the estimators.

$$\hat{y}_i = \widehat{\beta_0} + \hat{\beta}_1 x_i \tag{2}$$

$$\widehat{\beta_1} = \frac{S_{xy}}{S_{xx}}, \text{Where } S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \text{ y } S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{3}$$

$$\widehat{\beta_0} = \bar{y} + \hat{\beta}_1 \bar{x} \tag{4}$$

The estimates of least square for the slope and points of intersection of the adjusted line were calculated, being then that, the linear regression equation is obtained, with $\widehat{\beta_1} = 54.266$ and $\widehat{\beta_0} = 929.804$, therefore, Equation (5).

$$\hat{y}_i = 929.804 + 54.266 x_i \tag{5}$$

Our purpose is to forecast the period quarter IV of the year 2015, the first thing is to solve the Equation (7) for $\hat{y}_i$ in each period, the results are shown in Table 1. Due to the variability presented by the data, two scenarios are proposed; one optimistic and the other pessimistic, where they are compared with the real demand, to give an idea of the interval in which the data can alternate their value, this allows us to observe the relation between the variation of the forecast and its error with respect to the real demand. For this reason, it is necessary to evaluate the predictive capacity of Equation (5), by calculating the standard error of estimation and the coefficient of determination, in order to be able to estimate the intervals of confidence and prediction.

3.3. Diagnosis and validation –

Significance test of the parameter called slope of the line. The regression Equation (5) is then analyzed by a hypothesis test to see if the slope of the regression line is other than zero. If this is possible, then it can be concluded that using the regression equation increases the ability to predict or predict the dependent variable based on the dependent variable. Equation (8) indicates the null hypothesis, Equation (9) indicates the non-null hypothesis, Equation (10) is test t student, where your hypothesis parameters are represented by Equation (11) indicating that it is less than or equal to zero, and Equation (12) indicates the parameter greater than zero. According to the procedure expressed by [11] hypothesis test the null hypothesis and alternative are:

$$H_0: \beta_1 = 0 \tag{6}$$

$$H_1: \beta_1 \neq 0 \tag{7}$$

$$t = \frac{\beta_1 - 0}{S_b} \tag{8}$$

$$H_0: \beta_1 \leq 0 \tag{9}$$

$$H_1: \beta_1 > 0 \tag{10}$$

Knowing that, Equation (8) should be applied with $n - 2$ degrees of freedom. The distribution t is the test statistic, there are 6 degrees of freedom, determined by $n - 2 = 8 - 2 = 6$. The decision rule is to reject the null hypothesis if the value calculated with Equation (8) is greater than 1.440. The calculated value of 1.899 exceeds the critical value of 1.440, so the null hypothesis is rejected and the alternative hypothesis is accepted, in conclusion, the slope of the line is greater than zero. The independent variable, which refers to time, is useful to obtain a better estimate of the demand for product P.

3.3.1. Standard error of estimation and coefficient of determination –

In order to determine how inaccurate an estimate can be, we refer to the standard error of estimation, given by Equation (11), the standard error of estimation measures the dispersion with respect to the regression line for a given value of x. The standard error of estimation measures the dispersion with respect to the regression line for a given value of x. On the other hand, the determination coefficient [12] provides the total variation of the dependent variable, which is explained or accounted for, by the variation of the dependent variable. Equation (11) indicates the standard estimation value, and Equation (12) indicates the determination coefficient.

$$S_{yx} = \sqrt{\left[\frac{\sum(y - \hat{y})^2}{n - 2}\right]} = \sqrt{\left[\frac{SSE}{n - 2}\right]} \tag{11}$$

$$r^2 = \left[1 - \left(\frac{SSE}{SS\ Total}\right)\right] \tag{12}$$

Obtaining as a standard error of estimation a value of 185.185, which indicates that the data are relatively not so close to the regression line, that is, they are scattered with respect to the line. For the determination coefficient, it indicates that 62% of the total variation of the dependent variable is a residual variation.

3.3.2. Confidence intervals and prediction –

When a regression equation is used, two different predictions can be made for a selected value of the independent variable, these differences tend to be subtle reserving great importance [13-14]. For the present study, confidence intervals in linear regression given by Figure 5, the upper yellow line indicates the optimistic prediction interval and the lower yellow line indicates the pessimistic. For the case of the upper red line, this indicates the optimistic confidence interval and the lower indicates the pessimist. For the present study, the confidence intervals and the prediction interval, given by equations (13) and (14), were determined as follows:

Confidence interval
$$\hat{y} \pm tS_{yx}\sqrt{\frac{1}{n} + \frac{(x - \bar{X})^2}{\sum(x - \bar{X})^2}} \tag{13}$$

Prediction interval
$$\hat{y} \pm tS_{yx}\sqrt{1 + \frac{1}{n} + \frac{(x - \bar{X})^2}{\sum(x - \bar{X})^2}} \tag{14}$$

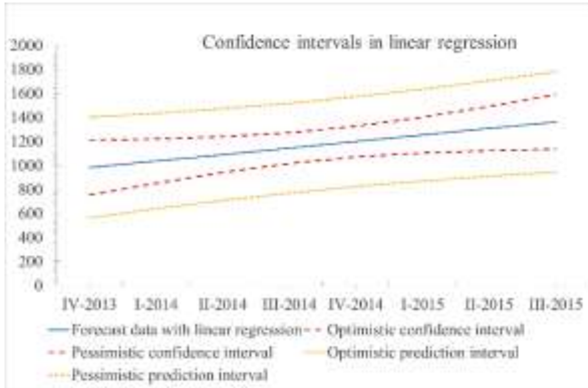

Figure 5. Confidence intervals in linear regression.

3.3.3. Evaluation of waste through statistical testing –

A sample of eight residues was obtained from the linear regression model fit by the least squares method applied to the time series. By means of the statistical software R the tests of Shapiro - Wilk, Kolmogorov - Smirnov and Jarque - Bera were carried out obtaining the information shown in Table 3. Where, Figure 6, shows the autocorrelation of waste and the Figure 7, shows the waste QQ-plot. The contrast between the three cases was as follows:

H_0: The waste follows a normal model.
H_1: The waste does not follow a normal model.

Table 2. Test contrasts referred to a significance level of 5%.

Test	Statistical value	P – value
Shapiro – Wilk normality test	W = 0.7315	0.8180
One – simple Kolmogorov – Smirnov test	D = 0.1745	0.6520
Jarque Bera test	JB = 4.5475	0.1029

The rule decision to resolve the contrast is that if the p-value is greater than the set level of significance (α =0.05) the hypothesis of normal waste cannot be rejected. In this case, as the "p – value" are greater than 0.05, we can conclude that the wastes are normal or that they do not differ from a normal distribution.

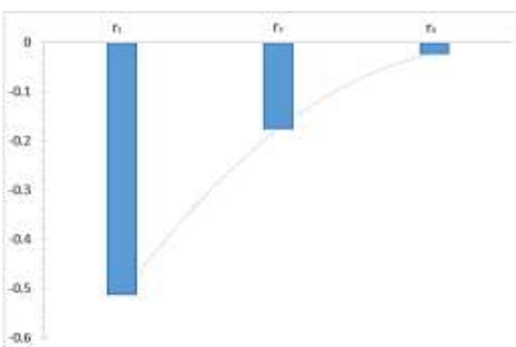

Figure 6. Autocorrelation of waste.

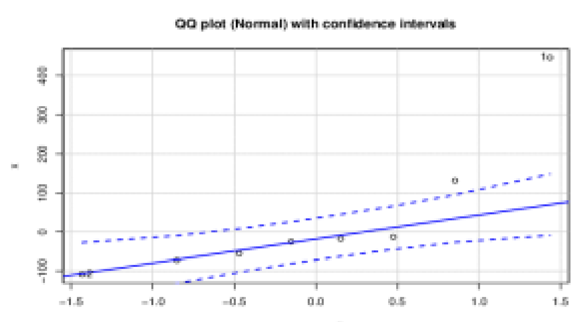

Figure 7. Waste QQ-plot.

According the Figure 6, Autocorrelation of waste. The autocorrelation of the residues was analysed by offsetting the series in three (3) time units, evidencing an exponential increase towards zero, which indicates that the dependence of the series with time is less and less. And, to Figure 7, Waste QQ-plot. It was observed that there are no indications of non-normality in the waste as the graph presents the typical behaviour of data from a normal distribution.

3.4. 3.4. Production program –

For the elaboration of the production master plan, the demand, working days, production rate, hours per day, number of workers, initial and final inventory, production capacity and quantity to be manufactured per week within the IV quarter of 2015 were taken into account, for this purpose the production capacity was balanced by means of a capacity plan, generating a balanced production in the Table 3.

Finally, to alleviate the workload of the month of December, half of the production of the month of December is scheduled for the months of October and November, in order to meet the demand of the last month, reserving units in warehouse until week 9, where they begin to be used to reduce the level of production in December.

Table 3. Master production program.

Product	P											
Current inventory	0											
	Week											
	1	2	3	4	5	6	7	8	9	10	11	12
Forecast	213	213	213	213	213	213	213	213	426	426	426	426
Initial inventory	107	214	321	428	535	642	749	856	0	0	0	0
Ending inventory									214	214	214	214
Master program	320	320	320	320	320	320	320	320	212	212	212	212
Demand or orders	320	320	320	320	320	320	320	320	212	212	212	212

IV. CONCLUSION

According to the significance test and the evaluation of the predictive capacity of the linear regression, the applied parameters and adjustments, we have that, the independent variable, that refers to time, and the relation with the demand, obtain a good estimation and turns out to be useful for the forecast of later periods in the short term, on the other hand, an error of 185.185 was obtained which indicates that the data relatively not so close to the straight presents dispersion, and that 62% of the total variation of the dependent variable is a residual variation.

REFERENCES

[1] C Stephen N 2006 Planificación y control de la producción, 1ª Edición (Ciudad de México: Pearson Educación) p 55

[2] L J Krajweski y L P Ritzman 2000 Administración de operaciones. Estrategia y análisis, 5ª Edición (Ciudad de México: Pearson Educación) p 491

[3] D Sippper y R L Bulfin Jr 1999 Planeación y control de la producción, 1ª Edición (Ciudad México: McGraw Hill Editores) p 96

[4] R B Chase y F R Jacobs 2014 Administración de operaciones, producción y logística, 13ª Edición (Ciudad de México: McGraw Hill Editores) p 485

[5] R Prada Nuñez y C A Hernández Suárez 2015 Análisis de una serie de tiempo utilizando diseño de experimentos como herramienta de calibración Ecomatemático Vol 6 N°1

[6] D A Lind, W G Marchal y S A Wathen 2015 Estadística aplicada a los negocios y la economía, 16ª Edición (Ciudad de México: McGrawHill Editores) p 250

[7] A M Alonso y C García Martos 2012 Análisis de series de tiempo (Madrid: Universidad Carlos III de Madrid) p 20

[8] A G Pérez 2015 Guía metodológica para la presentación de anteproyectos de investigación, 1ª Edición (Caracas: Universidad Pedagógica Experimental Libertador) p 23

[9] J C García Díaz 2011 Series temporales, análisis, predicción. Ejercicios prácticos, 1ª Edición (Madrid: Universitat Politécnica de València) p 105

[10] D D Wackerly, W Mendenhall III y R L Scheaffer 2008 Estadística matemática con aplicaciones, 7ª Edición (Ciudad de México: Cengage Learning Editores S.A.) p 302

[11] H J Gallardo, J P Rojas, y O A Gallardo. 2019 Modelación de series temporales en el sector productivo del Norte de Santander, 1ª Edición (Bogotá: Ecoe Ediciones Ltda.) p 7

[12] M Ferrán Aranaz y L Escot 2019 Una propuesta metodológica para el análisis gráfico de series temporales regionales: una aplicación a las tasas de paro provinciales en España Investigaciones Regionales Journal of Regional Research p 43 57

[13] C Faloutsos, J Gasthaus, T Januschowski & Y Wang 2018 Forecasting big time series: Old and new Proceedings of the VLDB Endowment Vol 11N° 12.

[14] González, Y. C. O., & Gaitán, I. M. G. 2018. Control estadístico de procesos en organizaciones del sector servicios. Respuestas, Vol 23 N° 1, pp 42-49.