

PAPER • OPEN ACCESS

Dynamic and sequential update for time series forecasting

To cite this article: H J Gallardo Pérez *et al* 2020 *J. Phys.: Conf. Ser.* **1587** 012016

View the [article online](#) for updates and enhancements.

You may also like

- [The impact of digitalization on the client-oriented approach in the provision of public services to business entities](#)
Svetlana Nyashenko, Anna Zvereva, Ekaterina Golubtsova et al.
- [Modelling and Forecasting of SO₂ Concentration in Atmospheric Air – A Case Study of the City of Krakow](#)
Zbigniew Skrobacki, Lidia Ewa Dabek and Piotr Lagowski
- [Instruments for the measurement of solar radiation](#)
Robert S Whipple



The Electrochemical Society
Advancing solid state & electrochemical science & technology

241st ECS Meeting

May 29 – June 2, 2022 Vancouver • BC • Canada

Abstract submission deadline: Dec 3, 2021

Connect. Engage. Champion. Empower. Accelerate.
We move science forward



Submit your abstract



Dynamic and sequential update for time series forecasting

H J Gallardo Pérez¹, M Vergel Ortega¹, and J P Rojas-Suárez¹

¹ Universidad Francisco de Paula Santander, San José de Cúcuta, Colombia

E-mail: henrygallardo@ufps.edu.co

Abstract. Two different sciences, physics and statistics, have worked, from the foundations of each, on the explanation and modelling of stochastic processes characterized by the succession of random variables whose realizations at each instant of time give rise to time series. From Physics we have worked with the Fourier transform to explain the dynamics of time series, a similar case occurs from statistics where dynamic models of time series are worked to explain the variations of the series and, in both cases, to make reliable forecasts. The main objective of this research is to adjust a model, using the methodology framed in the sequential update procedure of the forecast, to a time series of coal production observed quarterly during the years 2007 to 2011, in order to disaggregate quarterly the annual production for the years 2012 to 2018. Once the process has been carried out and validated, a quarterly production model is estimated which allows valid and reliable forecasts to be made for each quarter in subsequent years.

1. Introduction

Coal production is one of the pillars of the productive sector of the department of Norte de Santander, located in the northeast region of the Republic of Colombia; however, the information recorded on the production of this mineral is disaggregated quarterly from 2007 to 2011 and from then on only the annual production is recorded, making it necessary to build reliable estimates of quarterly production of coal in recent years [1,2]. For this purpose, the use of time series models is used, in which the analysis of variables requires the use of statistical methods to extract information from the observations made that do not obey any sampling plan or any experimental design, so that a model can be built that allows reliable forecasts to be made.

There are different methods for time series analysis, among which are: the classical method, which explains the behavior of the series with a structural model as a function of time, and the Box and Jenkins methodology, which describes the behavior of the series as a function of values observed in the past and their random variability. This methodology has proven to be a highly efficient technique for making predictions in situations where the pattern inherent in the series is very complex and difficult to unravel [3,4]. Both methods can be complemented in the analysis of the structure of the chronological series that allows, among others, to disaggregate the values obtained annually to a series in which the values of the production are related in quarterly form [5]. A time series, also called a time series or time series, consists of a set of data, derived from the outputs of a random variable observed successively over time. Its analysis involves the use of mathematical, physical and statistical methods to fit models to explain their behavior and to make forecasts. However, a time series can be understood as the aggregate of many waves, which is why the Fourier transform is an alternative for its decomposition and analysis, since it allows it to be broken down into those simple waves that make it up, in which case the amplitude, frequency and phase shift of each one must be identified, constituting the spectrum of the series; for this



reason it is said that the Fourier transform passes from space-time to space-frequency in order to estimate the signal or model of the time series [6].

This research works with the methodology developed by Box and Jenkins [4] which allows an autoregressive integrated moving average model (ARIMA) to be adjusted to a time series. The final objective is to adjust a model of this type to the part of the series observed quarterly, then forecasts are made that are updated sequentially and evaluated year by year, by comparing the estimated annual aggregate with the value observed for that year. In this way, estimators of quarterly production are obtained for the years in which this variable was not recorded [7,8]. Specifically, a model is adjusted to the quarterly coal production series for the years 2007 to 2011 and the iterative method of sequential updating of the forecast is applied in order to disaggregate the annual coal production for the years 2012 to 2018 on a quarterly basis. The added value of the research is that it allows for the establishment of a methodology to carry out this type of estimates disaggregated into different time series for production, climate and ecological variables, among others, when for some reason the complete information has not been recorded.

2. Method

This research is framed in the quantitative paradigm, part of information obtained from the free databases of the Colombian mining information system on coal production in the department of Norte de Santander and based on it adjusts statistical models associated with the time series.

The time series is composed of a set of realizations of a random variable Z , each of them observed in a period of time t , this is a realization of a stochastic process in discrete time: $\{Z_t\}_T = \{Z_1, Z_2, Z_3, \dots, Z_t, \dots\}$. The observed value of the variable in period t is noted Z_t , in each period of time is observed a single realization of the random variable. It is assumed that there is equiespaciamento between the observations and that these correspond to discrete points in the time, so that the collected data can be considered as finite successions of accomplishments of stochastic variables [5,9]. Time series analysis, which consists of using sample data for inference purposes (estimation, decision making and prediction), is complex from a functional point of view. However, it can be identified more as an art than as a science, although most of its procedures are based on results from mathematical statistics that have theoretical validity or have been empirically validated [7,10].

In particular, it should be borne in mind that, even though the variable being observed is the same in each period of time, it has a different probability distribution in each of those periods and, on observing its implementation, what is being observed is the value of a sample of size one in each of the periods. Thus, the construction of a time series model corresponds to an estimate of its parameters based on a sample size one; however, it has been demonstrated that the different time series models are estimated using methods that are highly reliable [5,8].

The observations of the phenomenon under study by means of time series are frequently correlated, with a correlation that increases as the time interval between each pair of observations decreases. In this work, two different time intervals are considered for each series, one registered quarterly and another registered annually; the purpose is to join the two series to construct a single one that corresponds to quarterly registers and to leave the methodology to continue developing this process in the following years.

The structural model, to be used in this research, Equation (1), assumes that the observed value of the series, Z_t , in time period t is the result of the interaction of four components: T_t trend component, C_t cyclic component, S_t seasonal component and I_t residual or random component. The model ARIMA explains the value of the series as a function of the combination of two polynomials: the autoregressive polynomial and the moving average [4,11]. Adjustment of the ARIMA model is performed based on the exploration of the functions of autocorrelation, FAC, and partial autocorrelation, FPAC, which are obtained once adjusted to a stationary form through differentiations to stabilize the mean and transformations to stabilize variance. [4,7,12].

$$Z_t = \mu + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}. \quad (1)$$

On the other hand, a stochastic process is said to be seasonal if its averaging function presents the behavior of a wave; in these processes the FAC and FPAC functions reflect the correlation between consecutive periods and the correlation between seasonal periods. The proposed model is constructed from two points of view: intra-and inter-stations, obtaining a composite model $ARIMA(p,d,q) \times (P,D,Q)_s$ in which p and q represent the orders of the autoregressive and moving average polynomials of the non-seasonal component and P and Q the corresponding orders of the autoregressive and moving average seasonal polynomials, respectively. The level is stabilized by means of non-seasonal differentiations and seasonal differentiations. The length of seasonality is represented by “S” [13-15].

3. Results

Figure 1 shows the quarterly production in tons of coal in the department of Norte de Santander, Colombia, for the years 2007 to 2011 [1]. A seasonal ARIMA model with autoregressive delay of seasonal type of order one is adjusted for this time series. The random variable W_t is then defined, depending on the delay operator B [$B^k Z_t = Z_{t-k}$], in such a way that it involves the differentiation of seasonal delay 4: $W_t = (1 - B^4)Z_t$. The model proposed for the variable W_t is of type $ARIMA(0,0,0) \times (1,1,0)_4$ which is autoregressive in the delay 4; the series does not present atypical data; the model does not include constant, since this is significantly equal to zero: $\hat{W}_t = 0.356\hat{W}_{t-4}$.

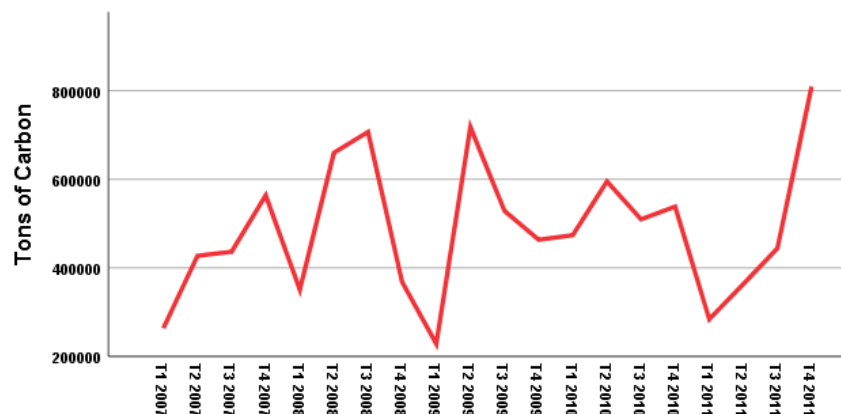


Figure 1. Quarterly coal production in Norte de Santander, Colombia.

Consequently, after performing the corresponding replacements and simplifications, the equation governing the model associated with coal production in Norte de Santander, Colombia, is: $\hat{Z}_t = 1.356\hat{Z}_{t-4} - 0.356\hat{Z}_{t-8}$. The estimated model is used to make the production forecast for the four quarters of 2012 and is adjusted proportionally by the value of the total production for that year. The series shown in Figure 2 is then obtained. The new adjusted model is $\hat{W}_t = -0.404\hat{W}_{t-4}$. After making the pertinent adjustments, the new model associated with the quarterly production of coal in Norte de Santander, Colombia, years 2007 to 2012, is: $\hat{Z}_t = 0.596\hat{Z}_{t-4} + 0.404\hat{Z}_{t-8}$.

The procedure described is followed iteratively, adjusting the forecast sequentially each year and weighting by the annual value to improve its estimation, thus obtaining a quarterly disaggregated series of coal production in the department of Norte de Santander, Colombia. After performing the corresponding iterations year by year, the following ARIMA estimated model is obtained for the quarterly coal production series in Norte de Santander, Colombia, with the sequentially updated forecasts: $\hat{Z}_t = 0.645\hat{Z}_{t-4} + 0.355\hat{Z}_{t-8}$. Figure 3 shows the values obtained in the whole process and, therefore, the series of production disaggregated quarterly.

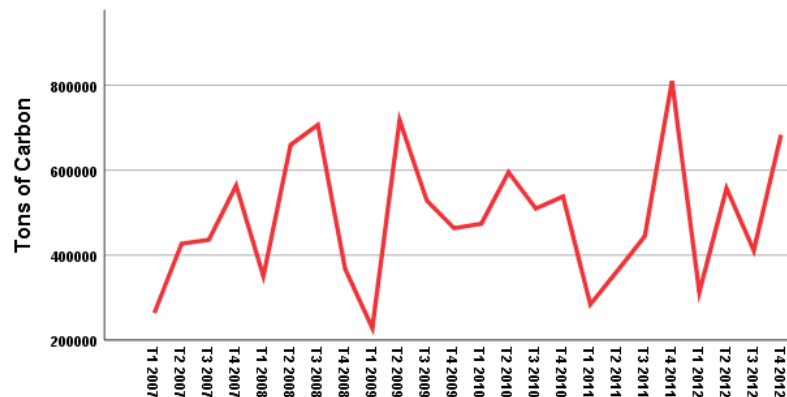


Figure 2. Quarterly coal production in Norte de Santander, Colombia, with forecast for 2012.

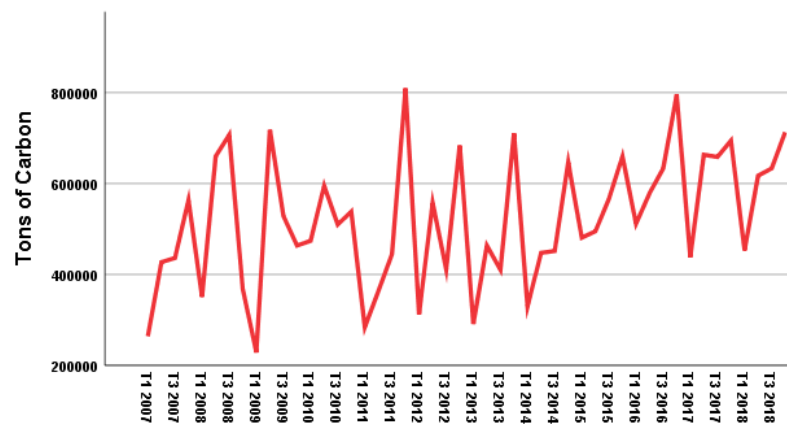


Figure 3. Quarterly coal production in Norte de Santander, Colombia, years 2007 to 2018.

4. Discussion

Stochastic variables in time produce different phenomena that can be studied from physics or statistics. Both sciences developed theories about this, physics tries to find the signal it describes with the properties of a wave and the residual effect is known as noise, it uses Fourier analysis and wave theory [16]. Statistics develops methods based on parameter estimation, from conditional hope, maximum likelihood and Bayesian estimation, among others, for the construction of a model, equivalent to the signal, with a residual effect or noise. Mathematics contributes to both sciences in the formalization of the model and theoretical demonstration of the properties but leaves the empirical work to physics and statistics; even, new theories involve fractal geometry to estimate the periodicity of the cycles that the series may present [5].

In both cases we work with the observed values of the variable, which are realizations of the variable under study in each time interval, that is, we work with size one samples in each instant of time [17]. The values observed in a time series are unrepeatably and if the value is not taken or an error is made in its measurement, there is no way to go back in time to repeat the observation. Methods have been developed to estimate missing values in a series [18], but the case presented in this paper refers to the application of statistical methods to estimate unobserved value vectors due to changes in the periodicity of information collection. It should be noted that this periodicity is a discrete deterministic variable and obeys exogenous criteria such as policies, customs, physical or financial, which, even when there is no knowledge of the variable being carried out in those periods of time, differs from the case of missing data presented for more random reasons, *i.e.*, it does not predict when the information will no longer be taken.

This is not the case presented in the paper, in this case, for some external reason, in the database of the Colombian mining and energy information system, in 2012 the periodicity in the collection of information is changed; until 2011, coal production was recorded quarterly, but from 2012 it is recorded annually. The relevance of the research lies in the presentation of the iterative method of updated sequential forecast estimation to disaggregate that annual production into quarterly production, following the pattern of the series and involving the expected random variability of the data.

There are other methods for making this estimation, for example, a relationship can be established, through a linear regression model, between the variable under study and other related variables over the time in which all the information is available; a multivariate time series model is then estimated and, based on the observed values of the other explanatory variables, the data vectors of the variable of interest are estimated [19]. This method implies a greater search for information and the validation of non-cointegration between the selected variables and, although it may give more reliable results, it is not within the objective of the present work and its analysis in another phase of the investigation is not discarded.

In the analyzed series, reliable forecasts are made, however, better results are obtained if the series initially observed is longer in terms of time than that which is required to disaggregate according to the forecast. The fundamental purpose of this phase of the investigation is the incorporation of the sequential forecast to the development of the analysis of temporal series, the variable analyzed here, and related variables, constitute a contribution not only to the production and the economy, but also contribute significantly to the line of investigation in this field.

5. Conclusions

The fundamental purpose of this work was the application of univariate time series statistical methods to the estimation of missing data vectors due to predetermined external causes, as is the case in which the database has registered quarterly information on the performance of a temporary random variable, but which from a moment in time is changed to an annual report. The developed iterative procedure of updated sequential forecast estimation is convenient for application when only data of one variable is available.

There are multivariate methods, in which the proposal developed here can be generalized after fulfilling the assumptions of non-Cointegration of the variables and the estimation of a multivariate time series model; the estimation of the unknown data will correspond to the estimated values of the variable of interest exposed as a dependent variable in the multivariate model. This work will be the continuation of the research presented. Another procedure to be developed is the use of the Fourier transform to estimate the identification of the time series signal and to evaluate the estimation of the missing intertemporal data.

References

- [1] Unidad de Planeación Minero-Energética 2019 *Boletín Estadístico de Minas y Energía* (Bogotá: Ministerio de Minas y Energía)
- [2] Unidad de Planeación Minero-Energética 2019 *Sistema de Información Minero Energético Colombiano* (Bogotá: Ministerio de Minas y Energía)
- [3] Peña D 1990 *Estadística Modelos y Métodos: 2. Modelos Lineales y Series Temporales* (Madrid: Alianza Editorial)
- [4] Box G, Jenkins G 1969 *Time Series Analysis, Forecasting and Control* (San Francisco: Holden-Day)
- [5] Gallardo H, Gallardo O, Rojas J 2019 Estimation of models and cycles in time series applying fractal geometry *Journal of Physics: Conference Series* **1329** 012018:1
- [6] Saavedra V, Fernández T, Harmony T, Castro V 2006 Ondeletras en ingeniería, principios y aplicaciones *Ingeniería, Investigación y Tecnología* **7(3)** 185
- [7] Guerrero V 1991 *Análisis Estadístico de Series de Tiempo Económicas* (México: Universidad Autónoma Metropolitana)
- [8] Peña D 2010 *Análisis de Series Temporales* (Barcelona: Alianza Editorial)
- [9] Gao J, Cao Y, Tung W, Hu J 2007 *Multiscale Analysis of Complex Time Series: Integration of Chaos and Random Fractal Theory, and Beyond* (New Jersey: John Wiley & Sons)

- [10] Gallardo H Rojas J, Gallardo O 2019 *Modelación de Series Temporales en el Sector Productivo del Norte de Santander* (Bogotá: ECOE)
- [11] Madrigal S 2014 Modelos de regresión para el pronóstico de series temporales con estacionalidad creciente *Computación y Sistemas* **18(4)** 821
- [12] Render B Stair R, Hanna M 2006 *Métodos Cuantitativos para los Negocios* (México: Pearson)
- [13] Brockwell P, Davis R 2002 *Introduction to Time Series and Forecasting* (New York: Springer)
- [14] Frances P 1998 *Time Series Models for Business and Economic Forecasting* (Cambridge: University Press)
- [15] Shumway R, Stoffer D 2017 *Time Series Analysis and its Applications* (Gewerbestrasse: Springer)
- [16] White G 2010 *Introducción al Análisis de Vibraciones* (New York: Azima DLI)
- [17] Nores M, Díaz M 2005 Construcción de modelos gee para variables con distribución simétrica *Revista de la Sociedad Argentina de Estadística* **9** 43
- [18] Medina R Montoya E, Jaramillo A 2008 Estimación estadística de valores faltantes en series históricas de lluvia *Cenicafé* **59(3)** 260
- [19] Mauricio J 2007 *Introducción al Análisis de Series Multivariadas* (Madrid: Universidad Complutense de Madrid)