

Statistical Model for the Forecast of Hydropower Production in Ecuador

Mónica Mite-León*, Julio Barzola-Monteses*‡

*Department of Scientific Research, Faculty of Mathematics and Physical Sciences, University of Guayaquil, P.B. EC090514,
Guayaquil, Ecuador

(monica.miteleon@ug.edu.ec, jbarzola@ieee.org)

‡ Corresponding Author; Julio Barzola-Monteses, P. B. EC090514, jbarzola@ieee.org

Received: 06.09.2017 Accepted: 01.11.2017

Abstract- The main sources of electricity generation are hydroelectric and thermo-fossil type in Ecuador. Gross hydroelectric production have sometimes exceeded 50% of national production. However, this type of hydroelectric is often threatened by droughts and their effect on the water reservoirs which our country has experienced on some occasions, such as in 2009. On the other hand, the National Plan for Good Living 2013-2017, in conjunction with the National Master Plan for Electrification 2013-2022, has planned new hydroelectric generation projects, which are expected to exceed 90% of the national balance. The objective of this article is to model the monthly production of hydroelectric energy for prediction purposes, by implementing five stochastic process models on a historical series of monthly hydroelectric energy production in Ecuador, during the period 2000-2015. The results show that the model that best fit the data of this time series is the ARIMA model $(1, 1, 1) \times (0, 0, 1)_{12}$ with seasonality. This model shows that the energy monthly production can be forecasted to one and twelve months. The range used was from 2000 to 2014 and it was validated with data from January to December of 2015. With this model, the forecast is made for the year 2020, proving an increase of monthly production. The real values are in the confidence interval of the predicted values of the ARIMA model with annual seasonality. This model will help to describe and predict hydroelectric energy generation of Ecuador. In other words, it could be used in future planning studies of the electric sector.

Keywords- Electric Power; Hydroelectric; Monthly Production, Stochastics Models, ARIMA.

1. Introduction

At the global level, non-renewable energies play an important role in the production of electric energy. In 1973, fossil sources generated 75% of electric power, and in 2009, these accounted for 67% of the production. The main reasons for this variation not to be so significant are the increase of the world population and per capita energy consumption [1].

However, some studies have been developed to take advantage of renewable energies alone or combined with non-renewable sources of electricity generation [2-6]. A study by [7] analyzed the feasibility of a solar hybrid - wind - diesel system in a locality of Ecuador. Another study by [8] used an optimization approach to 100% renewable systems such as hydroelectric and wind power using algorithms and techniques of artificial intelligence in an African locality. Other studies focussed specifically on the hydroelectric

systems in which the optimization of a hydroelectric generator is analyzed such as the one shown in [9]. There are other studies related to the effects of greenhouse gases, climate change and the environmental impact of these hydroelectric systems [10-12].

It should be noted that though fossil fuels have always existed in Ecuador, renewable sources of hydroelectric generation have already had a historical contribution in our electricity demand that has been related to certain indices such as population growth and per capita energy consumption [1]. A study by [13] shows that the hydroelectric potential density in Ecuador is 0.62 GWh/km² and ranks sixth out of twenty countries in the region.

On the other hand, within the National Plan for Good Living 2013-2017, one of its objectives states: "Promote the transformation of the productive matrix," and among the sectors prioritized is the industry referred to Renewable

Energy [14]. Hence, in the National Electrification Master Plan 2013-2022 [15], a major boost has been planned to hydroelectric generation, which is expected to exceed 90% of the total electricity production of the national balance.

However, this type of energy has often been threatened by the baseflow caused by droughts. For example, in 2009, a national emergency was raised due to the water level reduction in the Paute Molino Power plant which feeds the Paute Hydroelectric Plant. With a power output of 1100 MW, it generated the most energy of the country at the time. The annual rainfall cycle of the Paute river basin is constituted by a rainy season from April to September and a dry season from October to March. Generally, the most severe droughts occur between December and January [16].

In general, the energy balance of a country depends on several factors such as climate, season, population, economy, energy substitutes and indices of the electric sector. Thus, each factor is related to relevant parameters such as river inputs, sea surface temperature, reservoir level, months of increase or decrease in demand, among others. Further details were reviewed by [17].

By the end of 2016, the installed hydroelectric capacity in Ecuador represented 58% of the total (7587 MW). The country's 10 largest hydroelectric power stations represent 85% of hydroelectric capacity [12]. However, as already mentioned, this type of electricity generation has its limitation due to the great meteorological dependence, since the rainfall raises the levels of the flows in the hydrographic basins where these electrical systems are located.

The objective of this research is to model the historical data of the monthly production of hydroelectric energy in Ecuador between 2000 and 2015, performing a detailed analysis of the data to reduce the forecast error. This will allow us to make a stochastic model to forecast the monthly production of hydroelectric energy. The resulting model will help predict the country's hydroelectricity generation and can be used for energy planning from different sources of electricity production.

Series can be modelled by deterministic methods using mathematical functions or by stochastic processes when near-time observations are correlated. One of the most used methods for the adjustment of time series is the methodology of Box-Jenkins who in 1970 [18] developed a method for the analysis of time series known as Autoregressive Integrated Moving Average (ARIMA). The study of the time series serves to make forecasts in a specific time horizon, in different areas of knowledge, such as economics, demography, physics, electricity, etc.

According to the patterns of the data, the series can present components of trend, cyclicity, seasonality, randomness and stationary. For this reason, the entire process is then detailed to identify the best model to fit Monthly Gross Production (MGP) series data.

2. Methodology

2.1. Collecting data

The data for this research was obtained from the official site of the Electricity Regulation and Control Agency (ARCONEL) corresponding to the years 2000 to 2015. Monthly reports of the energy production were registered about the hydroelectric plants of Ecuador. The study for the modelling in this research considered data between the years 2000 to 2015. The data of the year 2015 was used for the comparison of the predicted data.

2.2. Series analysis

As the data was measured over time, and uniformly spaced, we considered using the Box-Jenkins methodology [19]. The modelling of the data is done using integrated autoregressive and moving average models. These models are regression models with delays in the dependent variable X_t and delays with respect to the error term.

In the ARIMA models (p, d, q), the parameters p, d and q must be identified, where the parameter p is the autoregressive value of the dependent variable, d is the finite difference transformation, and q is the delay of the error term or the moving average value of these stochastic models. To find these values, the stationarity of the time series data was analyzed in detail. The single Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are correlograms functions help to determine the degree of correlation between two consecutive values of the series and give an idea of the possible parameters of the ARIMA models [20].

2.3. Series transformation

The Box-Jenkins methodology evaluates series that present stationarity to be modelled. The MGP series does not present this characteristic because the mean and variance are not constant over time and the covariance is time variant. For this reason, the series was transformed to eliminate variability and trend over time.

Logarithms were obtained from the MGP series data to eliminate the variability, and the first difference was obtained to decrease the trend. Because the graphs of the correlation functions are similar, the differentiated series for this work has been used because the different models to be developed are less complex.

2.4. Stationarity evaluation

Initially, the single ACF and PACF functions of the transformed series were plotted.

The first-order autocorrelation is formally tested with the Durbin-Watson statistic which measures the linear association between adjacent residuals. The null hypothesis for this case is that the series presents autocorrelation.

In addition, a higher power method such as the Augmented Dickey Fuller (ADF) test or unit root test was used, which was based on assuming that the series can approximate an autoregressive of order 1. This test assumes three variants such as a random walk with null mean, random walk with drift and random walk with drift and linear trend. The null hypothesis in this test is that the MGP series having a unit root.

2.5. Model identification

In ARIMA models (p, d, q) [21] the series differs d times to obtain a stationary series. These stationary models present the following equation 1:

$$X_t^d = c + \overbrace{\phi_1 X_{t-1}^d + \dots + \phi_p X_{t-p}^d}^{AR(p)} + \overbrace{\theta_1 \epsilon_{t-1}^d + \theta_2 \epsilon_{t-2}^d + \dots + \theta_q \epsilon_{t-q}^d}^{MA(q)} + \epsilon_t \tag{1}$$

Where

X_t^d is the series with differences of order d

ϵ_t represents the process of white noise with normal distribution $N(0, \sigma_2)$ being independent and identically (i.i.d) and c, $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ are the model parameters.

As the MGP presents strong seasonality [22], models have been used where ARIMA models are combined with seasonal terms. This new combined model has two components: a component with regular structure ARIMA (p,d,q) that models the non-independence associated with the data and the other component with ARIMA structure (P,D,Q) that models the seasonality component, where P is the autoregressive seasonal term, D seasonal term of difference and Q seasonal term of moving average.

The equation 2 of the general mathematical model for this type of model, also called SARIMA [23] is:

$$X_t = c + \overbrace{\phi_1 X_{t-1} + \dots + \phi_p X_{t-p}}^{AR(p)} + \overbrace{\theta_1 X_{t-s} + \dots + \theta_p X_{t-ps}}^{SAR(P)} + \underbrace{\epsilon_t - \phi_1 \epsilon_{t-1} - \dots - \phi_q \epsilon_{t-q}}_{MA(q)} - \underbrace{\vartheta_1 \epsilon_{t-q} - \dots - \vartheta_Q \epsilon_{t-Qs}}_{SMA(Q)} \tag{2}$$

Where c, $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \vartheta_1, \dots, \vartheta_Q$ are the model parameters to estimate.

2.6. Model estimation

Five models were designed and implemented. Then, for the selection of the best model we used adjustment statistics such as Mean Absolute Percentage Error (MAPE), Coefficient of Determination (R^2), Akaike Information Criterion (AIC) [24] and Bayesian Information Criterion (BIC) [25].

2.7. Model validation

The selected model proceeds to determine if it fulfils all the hypothesis of validation of the residues, although the estimation error is first plotted to analyze if there is a presence of atypical errors, which indicate a presence of intervention. Among the validation tests of the model, we have the Box-Ljung test [20], with the null hypothesis being that the MGP series is uncorrelated. For the constant variance of the residuals we use the Box-Ljung test with squared residues with the null hypothesis that the MGP series has constant variance, and finally the Jarque-Bera test for the normality of the residues being the null hypothesis that the residues are normal.

2.8. Model forecast

Once the ARIMA combined model has been validated, energy production is forecasted for a 60 months horizon, corresponding for the period 2016 a 2020. The predicted values have been estimated with a 95% confidence interval.

3. Results and Discussion

Figure 1 shows the time series of the MGP of the data, in which it is observed that the MGP series presents an upward trend over time, annual cyclicity with peaks of energy production from March to August and strong seasonality. At the end of 2011 and until September 2012, there was a higher energy production at high levels due to the operation of new hydroelectric plants, an effect that is also observed in the following years of the series. The monthly MGP series presents mean and variance not constant over time, producing variability over time [26]. The MGP series as non-stationary in mean, variance and covariance not consistent over time. In the period 2010-2015, the hydroelectric power production increased due to the potential nominal and effective potential increase of the new hydropower plant in Ecuador.

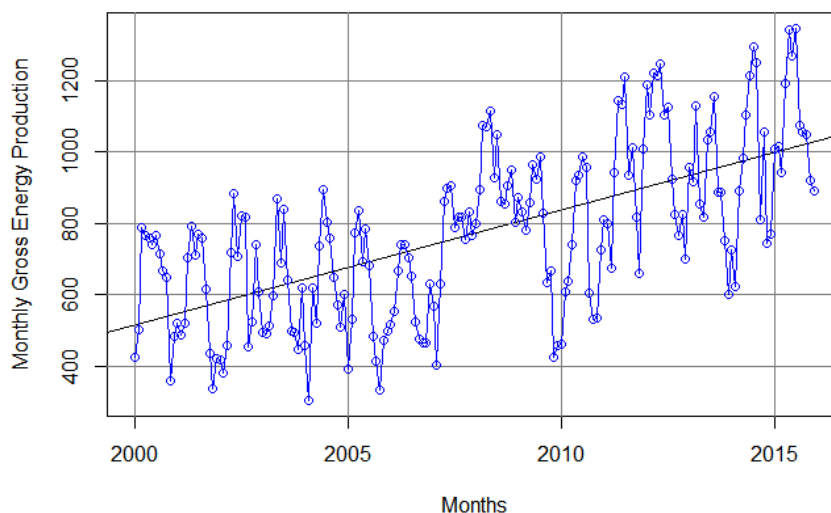


Fig. 1. Series of gross monthly hydroelectric power production in Ecuador in the period 2000-2015, with a trend line

Table 1 presents some measures of central trend and dispersion of data per year. The statistics calculated for the energy production data show the variability in the form of peaks of ups and downs characteristic of the time series.

Table 1. Measures of central trend and dispersion.

Year	Annual average production (GWH)	Standard Desviation	Minimum	Maximum	Median
2000	634.14	152.14	356.42	789.11	690.86
2001	580.95	152.21	334.29	792.88	520.90
2002	611.23	177.68	379.25	802.07	609.33
2003	599.21	134.22	444.40	866.99	595.89
2004	617.64	163.38	302.34	892.99	608.81
2005	573.52	171.23	331.14	834.60	512.99
2006	594.12	106.08	464.79	739.00	590.99
2007	753.14	148.51	403.60	906.04	801.75
2008	941.11	109.93	797.68	1114.69	915.65
2009	768.54	185.52	424.10	984.95	803.03
2010	719.70	187.06	460.45	984.97	681.27
2011	927.76	181.47	659.52	1209.74	937.13
2012	1019.80	199.09	697.90	1246.16	1102.66
2013	919.90	159.13	599.22	1153.91	902.04
2014	954.83	227.12	622.53	1294.40	935.86
2015	1091.00	158.77	891.70	1347.00	1051.00

In 2008 and since 2011, the annual average of energy production is higher than 900 GWH.

Figure 2 presents the autocorrelation functions ACF and PACF. Both functions decay exponentially in a delay or lag of 5, which are significant with period seasonal frequencies suggested by the SARIMA model.

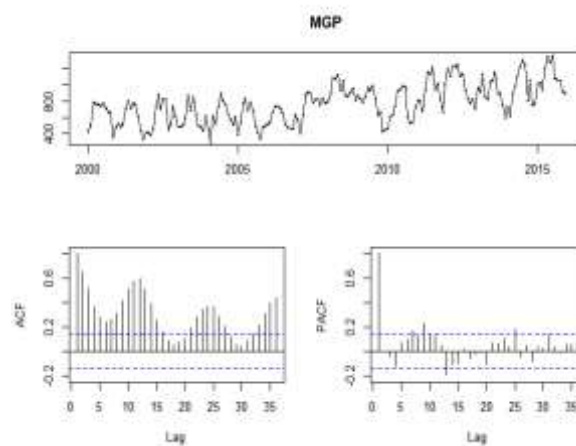


Fig. 2 Correlogram functions of the MGP series

Figure 3 shows the series with first differences and autocorrelation functions. It is graphically shown that the series is stationary over time. The value of the Durbin-Watson statistic obtained is 1.98, so the series presents evidence of weak positive autocorrelation.

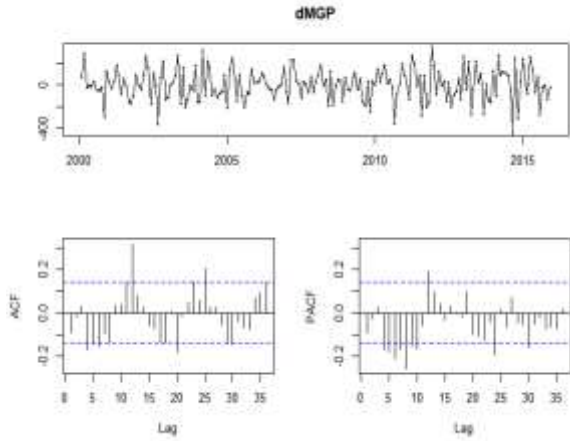


Fig. 3 Differenced MGP series and the ACF and PACF autocorrelation functions.

For the analysis of the stationarity of the series, we obtained the DFA or unit root statistics, whose value was $t\text{-DFA} = -8.77$ with a p-value of 0.01. If we compare it to a significance level of 5%, the $t = -2.87$ statistics, the t-DFA falls into the rejection zone so there is significant evidence that the series is stationary and has no unit roots, with a maximum delay of 13. To determine the variants, the regression test was used to derive the DFA test which resulted to be significant with a value $t = -15.06$ at 5% level of significance, which means that the series shows a random walk with mean zero.

As the differentiated MGP series is stationary, we proceed to identify the parameters of the combination of the SARIMA model. The model selected was ARIMA (1,1,1)x(0,0,1)₁₂ with random walk. The model is presented in Figure 4.

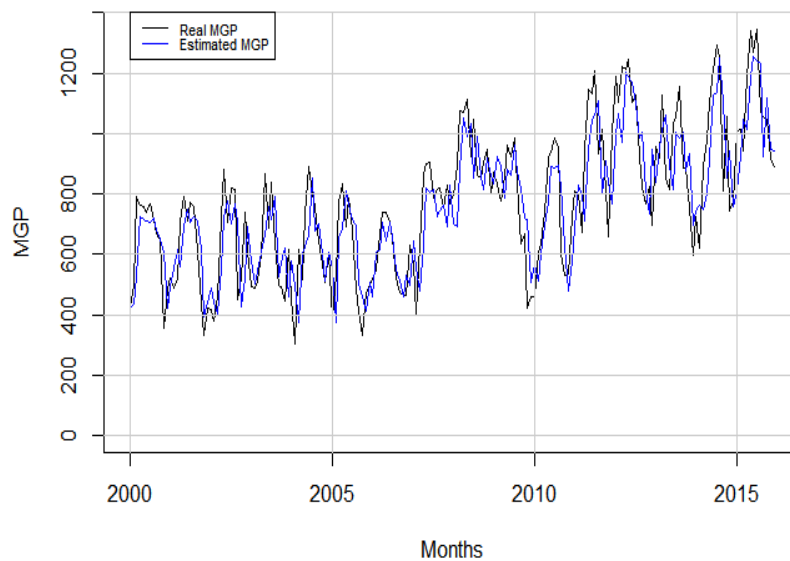


Fig. 4. ARIMA model (1,1,1)x(0,0,1)₁₂ with random walk

Table 2 shows the coefficients of the model: ARIMA (1,1,1)(0,0,1)₁₂.

Table 2. Model coefficients.

	AR(1)	MA(1)	SAR(1)
Coefficients	0.65	-0.98	0.38
Standard desviation	0.06	0.02	0.07

The equation of the model is:

$$X_t = 0.65X_{t-1} + \epsilon_t - 0.98 \epsilon_{t-1} + 0.38 X_{t-12} \tag{3}$$

The estimation of the model was performed through the adjustment statistics, presented in Table 3.

Table 3. Adjustment Statistics.

Statistic	MAPE	R ²	AIC	BIC
	14.32%	72.39%	2392.85	2405.86

The MAPE represents the forecast error statistic. R², the percentage explained by the estimated model in the MGP series, is quite good. The AIC and BIC are the lowest values of all the models designed.

We analyze if the selected model presents intervention and then determine the validation of the model residuals. Figure 5 shows the residuals of the model are represented with more or less three standard deviations. It was observed that it was not necessary to apply an intervention variable because there are no atypical values.

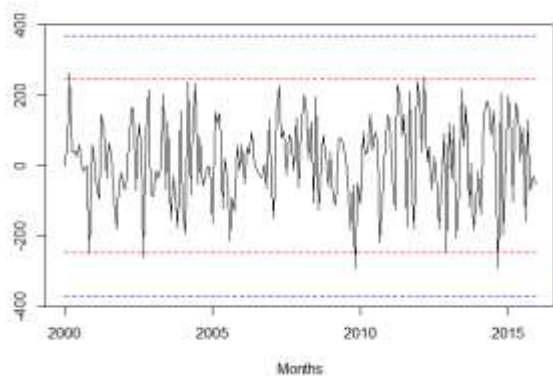


Fig. 5 Estimation error of the selected model of the MGP series.

For the validation of the model, the validation hypotheses of the residuals are tested. First, the Box-Ljung statistic was used to determine if the residuals of the model are uncorrelated. The p-value obtained was 0.735, so there is statistical evidence that the differentiated MGP series are uncorrelated. The variance of the residuals is tested with the Box-Ljung statistic in which the squared residuals were used. In this case, the null hypothesis is that the residuals have constant variance. The p-value was 0.957, so we fail to reject the null hypothesis. To analyze the normality of the residues, the Jarque-Bera test is used. The null hypothesis is that the residuals are normal. The resulting p-value is 0.215 so that the residuals normality is not rejected.

In order to do a model fit, a full-cross validation was performed. Data from 2000 to 2014 was used to estimate the year 2015 and the results were very close to the actual values of 2015. With this validation, a forecast was made over a

period of 60 months.

Table 4 shows the predicted values. It is observed that the actual values and the predicted values are within the designated confidence interval.

Table 4. Comparison of actual values and predicted values.

Actual Value 2015	Predicted value 2015	Lower bound CI 95%	Upper bound CI 95%
1006,74	827,51	585,23	1069,78
1014,29	790,99	494,74	1087,24
941,68	901,53	581,05	1222,01
1190,49	939,32	606,49	1272,16
1340,09	989,12	649,40	1378,39
1269,00	1034,49	690,59	1378,39
1346,73	1067,86	721,21	1414,51
1074,23	1050,10	701,48	1398,72
1054,70	874,79	524,66	1224,93
1048,23	973,22	621,84	1324,60
918,34	848,30	495,84	1200,76
891,68	859,17	505,73	1212,60

Figure 6 shows the actual MGP series, the adjusted series, and the predicted horizon 60 months projection.

Table 5 shows the predicted values. It is observed that the actual values and the predicted values are within the designated confidence interval.

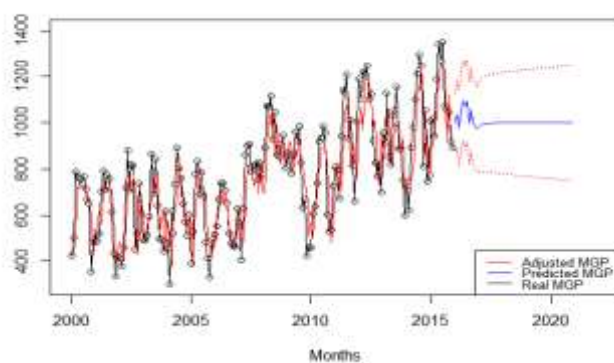


Fig. 6. 60 Months Forecast of Energy Production in Ecuador using ARIMA (1,1,1)x(0,0,1)₁₂.

Table 5. Forecast 2016-2020.

Months	Years	2016	2017	2018	2019	2020
Jan		1002.89	982.71	999.09	999.21	999.21
Feb		1035.70	988.32	999.13	999.21	999.21
Mar		976.83	992.02	999.16	999.21	999.21
Apr		1054.40	994.46	999.18	999.21	999.21
May		1097.70	996.07	999.19	999.21	999.21
Jun		1071.04	997.14	999.19	999.21	999.21
Jul		1092.67	997.84	999.20	999.21	999.21
Aug		997.48	998.31	999.20	999.21	999.21
Sep		1052.74	998.61	999.20	999.21	999.21
Oct		1007.19	998.81	999.20	999.21	999.21
Nov		993.61	998.95	999.21	999.21	999.21
Dec		974.22	999.04	999.21	999.21	999.21

Conclusions

The ARIMA (1,1,1)(0,0,1)₁₂ model with selected random walk of five the models constructed in this research reflects the trend of gross monthly energy production. This random walk mean zero model with has estimated that the occurrence of the production is for the parameter $p = 1$ and a suitable moving average of $q = 1$ for the regular component. For the seasonal component, it is $Q = 1$ and the strong seasonality of $S = 12$. The result suggests that the model adequately adjusts the data of the series, although the series presents periods with greater production of energy. The value of the standard absolute deviation MAPE presented 14.32% and R^2 of 72.39 % concluding that the estimation was quite good.

It should be emphasized that the model estimated the monthly production over 60 months from 2016 to 2020 based on the times series of 192 months. The forecast in this time horizon exhibits a good fit in the first 24 months but after that, the values become constant. This model predicts that the MGP is on the rise in the coming years with an annual rate of 1.95%. Based on this, we can conclude that these models can be used for the better understanding of the hydroelectric energy production trend and thus, can help with the future planning in the energy sector.

However, in the coming years, it is foreseen that there will be a great growth of energy production of this type, due to the operation of new hydroelectric plants of emblematic projects [27]. Hence, it would be advisable in the future to develop new models incorporating representative new variables with respect to this activity, to construct an economical model and to make predictions that allow the governmental authorities in the energy area to make decisions in the production of energy.

References

- [1] F. Buñay and F. Perez, “Comparación de costos de producción de energía eléctrica para diferentes tecnologías en el Ecuador,” Universidad de Cuenca, 2012.
- [2] D. Matamoros, J. Barzola, and M. Espinoza, “Sizing Photovoltaic Irrigation Systems using Meteorological Data,” in *8th International Conference on Energy Planning, Energy Saving, Environmental Education*, 2015, pp. 220–227.
- [3] C. Pavón, J. Barzola, F. Cabrera, C. Briones, and M. Espinoza, “Fuentes de Energías Renovables como potencial de producción eléctrica en zonas rurales del Ecuador,” in *Proceedings of the Thirteenth Latin American and Caribbean Conference for Engineering and Technology*, 2015, pp. 1–2.
- [4] J. Barzola, M. Espinoza, C. Pavón, and F. Cabrera, “Solar-Wind Renewable Energy System for Off-Grid Rural Electrification in Ecuador,” in *14 LACCEI International Multi-Conference for Engineering, Education, and Tecnology*, 2016, pp. 1–7.
- [5] C. Pavón and J. Barzola, “Estimación de la demanda energética mensual mediante encuesta aplicada en la Provincia de Santa Elena,” *Rev. Cient. Yachana*, vol. 4, no. 2, pp. 1–12, 2015.
- [6] J. Barzola and L. Rubini, “Análisis técnico y financiero de Grid Parity residencial con fuente de energía solar,” *Rev. Cient. Yachana*, vol. 4, no. 1, pp. 11–18, 2015.
- [7] J. Barzola, M. Espinoza, and F. Cabrera, “Analysis of Hybrid Solar / Wind / Diesel Renewable Energy System for off-grid Rural Electrification,” *Int. J. Renew. Energy Res.*, vol. 6, no. 3, pp. 1146–1152, 2016.
- [8] A. Acakpovi, “Performance Analysis of Particle Swarm Optimization Approach for Optimizing Electricity Cost from a Hybrid Solar, Wind and Hydropower,” *Int. J. Renew. Energy Res.*, vol. 6, no. 1, pp. 323–334, 2016.
- [9] L. Jasa, I. P. Ardana, A. Priyadi, and M. H. Purnomo, “Investigate Curvature Angle of the Blade of Banki’s Water Turbine Model for Improving Efficiency by Means Particle Swarm Optimization,” *Int. J. Renew. Energy Res.*, vol. 7, no. 1, pp. 170–177, 2017.
- [10] E. Duque, J. Patiño, and L. Velez, “Implementation of the ACM0002 methodology in small hydropower plants in Colombia under the Clean Development Mechanism,” *Int. J. Renew. Energy Res.*, vol. 6, no. 1, pp. 21–33, 2016.
- [11] A. Kumar, M. P. Sharma, and A. Kumar, “Green House Gas emissions from Hydropower Reservoirs: Policy and Challenges,” *Int. J. Renew. Energy Res.*, vol. 6, no. 2, pp. 472–476, 2016.
- [12] P. E. Carvajal, G. Anandarajah, Y. Mulugetta, and O. Dessens, “Assessing uncertainty of climate change

- impacts on long-term hydropower generation using the CMIP5 ensemble—the case of Ecuador,” *Clim. Change*, pp. 1–14, 2017.
- [13] F. Posso, J. L. Espinoza, J. Sánchez, and J. Zalamea, “Hydrogen from hydropower in Ecuador: Use and impacts in the transport sector,” *Int. J. Hydrogen Energy*, vol. 40, no. 45, pp. 15432–15447, 2015.
- [14] Secretaría Nacional de Planificación y Desarrollo, “Buen Vivir: Plan Nacional 2013-2017,” 2013. [Online]. Retrieved July 2, 2017, from: <https://goo.gl/nakmoe>.
- [15] ARCONEL, “Plan Maestro de Electrificación 2013 – 2022,” 2013. [Online]. Retrieved July 10, 2017, from: <https://goo.gl/PF4KPG>.
- [16] M. Mera and S. Flores, “Resultado de las Acciones Ejecutadas Durante la Crisis Energética 2009,” *Cicyt Espol*, no. 1, 2010.
- [17] S. Medina and J. García, “Predicción de demanda de energía en Colombia mediante un sistema de inferencia difuso neuronal,” *Energética 33*, pp. 15–24, 2005.
- [18] G. E. P. Box, G. Jenkins, and G. Reinsel, *Time series analysis*. New Jersey: Prentice Hall, 1994.
- [19] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications With R Examples*, 3th editio. New York: Springer, 2011.
- [20] G. E. P. ; Box and A. Pierce, “Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models,” *Am. Stat. Assoc.*, vol. 65, no. 332, pp. 1509–1526, 1970.
- [21] K. Chan, *Time series analysis with applications in R*. New York: Springer Science, 2008.
- [22] F. Canova and B. E. Hansen, “Are seasonal patterns constant over time? A test for seasonal stability,” *J. Bus. {&} Econ. Stat.*, vol. 13, no. 3, pp. 237–252, 1995.
- [23] N. Sugiura, “Further analysts of the data by akaike’ s information criterion and the finite corrections,” *Commun. Stat. - Theory Methods*, vol. 7, no. 1, pp. 13–26, 1978.
- [24] F. X. Diebolt, *Elements of Forecasting*. South-Western College Publ., 1998.
- [25] B. J. Abraham and R. A. Ledolter, *Introduction to Time Series and Forecasting*. New York: John Wiley and Sons, 2000.
- [26] M. B. Priestley, *Non-linear and Non-stationary time series analysis*. London: Academic Press, 1988.
- [27] MEER, “Proyectos de Generación Eléctrica,” 2013. [Online]. Retrieved July 10, 2017, from: <https://goo.gl/vvceV5>.