

# Analysis of hydropower ratio from total energy production in Romania

Angela Neagoe<sup>1</sup>, Eliza Isabela Tică<sup>1</sup>, Mihaela Amalia Diminescu<sup>1</sup>, Bogdan Popa<sup>1\*</sup>, and Suzana Carmen Cismaș<sup>2</sup>

<sup>1</sup>University POLITEHNICA of Bucharest, Faculty of Power Engineering, 313 Spl. Independentei, 060042, Bucharest, Romania

<sup>2</sup>University of Agronomic Sciences and Veterinary Medicine, Faculty of Management and Business Engineering, 59 Mărăști Boulevard, 011464, Bucharest, Romania

**Abstract.** It is important to know the structure of electricity production of a country and to establish the place of each source within the total energy mix. The aim of this paper is to analyze statistically the percentage of the hydropower production related to the total electrical energy produced in Romania. The time series data analysis was done to determine the trends of average and standard deviation of hydro energy data produced in Romania over eleven years' period (2006-2016).

## 1 Introduction

Generally, recorded data from any process can be used to create models based on statistical analysis and machine learning. In order to be able to process such series, it is important to determine the main features that allow further predictions. By analyzing the known data, the tendency of their variation can be shown by identifying the nature of the represented phenomenon.

Numerous activities that present some uncertainties can be defined through time series of observations that are time-dependent, such as: monthly energy production, the temperatures recorded for a given region in the last 10 years, the evolution of the exchange rate, the annual evolution of a country's population, etc.

Based on the obtained model with statistical analysis of the temporal series, it is possible to predict the future evolution of the process from the last recorded values of the variable of interest.

In the energy sector, analysis of time series was applied to forecast the electricity consumption [1], precipitations for hydroelectric power management [2], monthly electricity peak-load [3], [4], and also electricity demand, hourly prices, [5], [6], etc.

Paper [7] presents the main characteristics of the energy data and existing models for energy research. Mainly, the analysis of time series admits that four components can be identified: tendency, periodicity, cyclicity and stochastic. For example, in different countries, a comparative analysis of energy consumption was made by using an integrated data characteristic approach [7]. Hence, formulating a forecasting model based on time series analysis requires the determination of these four components [8].

Various authors propose an improvement in forecasts based on time series analysis (in particular for non-stationary series) using the artificial intelligence [9-11].

For Romania the evolution of electricity consumption and production of hydropower plants in time, since 1884, was presented in [12]. The Romanian hydropower sector recorded an important increase in hydroelectric power between 1950 and 1990, when the most important and biggest hydropower developments (HPDs) were made (e.g. Izvorul Muntelui Bicaz HPD, Vidraru HPD, Lotru HPD, Iron Gates HPD, etc.). From 1 MW installed capacity in 1884, Romania reached an installed capacity of 6761 MW on January 1<sup>st</sup>, 2018. In 2017, the electricity sector in Romania was dominated by coal and hydrocarbons, 25.1%, hydro, 14.5%, nuclear, 10.6%, and renewable energy sources (wind, solar and biomass), 9.6%, [14].

The analysis of hydropower production variation was also related to the construction and modernization of hydropower plants. Furthermore, the economy of a country can influence the variation in time of produced and consumed energy [15].

## 2 Data analyzed as time series

With the implementation of European Directives, over the past ten years, an increase in electricity production from renewable energy sources can be noticed. For example, wind energy accounted for approximately 0.02% of the total electricity production in Romania in 2008, reaching 10.1% in 2016.

In this context, it becomes necessary to observe the evolution over time of the percentage of energy from hydropower in the total energy production in Romania by performing a statistical analysis.

Chronological records of the hydropower production and the total electricity production, monthly outputs, in Romania, between 2006 and 2016, define the time series analyzed in this paper. We mentioned that data from the first two years of the series are approximated from the

\*Corresponding author: [bogdan.popa@upb.ro](mailto:bogdan.popa@upb.ro)

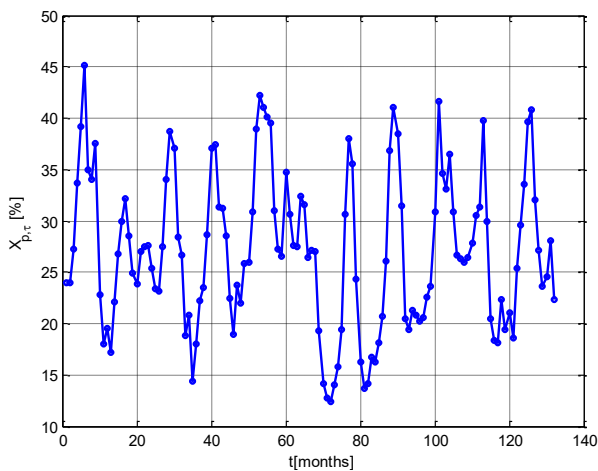
existing graphs, and precision may not be accurate. The rest of the values used in the analysis are expressed as numerical values in the annual reports of Hidroelectrica available on-line [16].

For the coefficient defined as the hydropower production divided by the total electricity production in Romania, hereafter referred to as the random variable  $X$ , the analyzed values are recorded with yearly time-step  $p=1,2,\dots,N$  and monthly time-step  $\tau=1,2,\dots,\omega$  and notation is  $X_{p,\tau}$ .

Since the analysis period is not long enough, only tendency, periodicity and stochastic components were determined; it was not the case for the cyclicity component.

### 3 Determination of time series characteristics

Time variation of the random variable  $X_{p,\tau}$  is presented in figure 1 (hydropower production ratio from the total electricity production in Romania), monthly average values for 11 years (2006-2016), 132 values. It is clear that the registered process shows periodicity, which leads to the conclusion that the analyzed series is non-stationary. Values are within a maximum range of 45% and minimum of 12%, with an average of 27.11%.



**Fig. 1.** Variation of the original random variable.

If tendency, periodicity and stochastic components are identified and removed from the original random variable, what remains is a variable that behaves as a dependent stationary process.

#### 3.1 Modeling the tendency component on average

To analyze the tendency of the monthly average values, a linear regression equation can be determined using the average values of the variable  $X(T_{mp})$  for each year:

$$T_m(t) = A_m + B_m \cdot t, \quad (1)$$

where  $A_m, B_m$  are coefficients of the polynomial function and  $t$  is the year index.

To change the time from years to months the following relation is used:

$$t = (p-1)\omega + \tau, \quad (2)$$

a regression equation for each step in the original series can be found ( $T_{mp,\tau}$ ).

Removing the average trend from the original variable leads to the generation of a new variable:

$$Y_{p,\tau} = X_{p,\tau} - T_{mp,\tau}. \quad (3)$$

If the tendency in average is important, this new variable will be different from the original one.

#### 3.2 Modeling the tendency component on variation

The string of annual averages of variable  $Y_{p,\tau}$  is determined using the relation:

$$\bar{Y}_p = \frac{1}{\omega} \sum_{\tau=1}^{\omega} Y_{p,\tau}, \text{ for } p=1,2,\dots,N, \quad (4)$$

and then the standard deviation values for each year can be computed with the formula:

$$T_{sp} = \sqrt{\frac{1}{\omega-1} \sum_{\tau=1}^{\omega} (Y_{p,\tau} - \bar{Y}_p)^2}, \text{ } p=1,2,\dots,N. \quad (5)$$

For this variable, a linear regression equation can be determined, of the form:

$$T_s(t) = A_s + B_s \cdot t,$$

(6)  $A_s, B_s$  are the first order polynomial coefficients.

Eliminating the tendency in variance in the process  $Y_{p,\tau}$  leads to a new variable, noted  $Z_{p,\tau}$ , with no tendency in average and variance:

$$Z_{p,\tau} = \frac{Y_{p,\tau}}{T_{sp,\tau}}. \quad (7)$$

#### 3.3 Determination of the periodic component

In order to determine the periodicity of this series, the method used is the autocorrelation.

To determine the existence of periodicity of the time series, the Anderson test [17] is used, according to which the boundaries of the confidence interval for the process correlation are given by the equation:

$$CL(k) = \frac{-1 \pm Z_c \sqrt{N-k-2}}{N-k-1}, \quad (8)$$

where  $N$  is the length of the data series string,  $k$  is the correlations' order and  $Z_c$  - the standard normal

distribution quantile for which the distribution function is equal to  $(1-\alpha/2)$ , for a chosen threshold of significance  $\alpha$ .

For  $k \geq 1$ , if the correlogram exceeds the limits specified by (8), then the process is not purely stochastic, but also has a component of periodicity.

In this case, variable  $Z$  can be expressed in the form:

$$Z_{p,\tau} = \mu_\tau + \sigma_\tau \varepsilon_{p,\tau}, \quad (9)$$

with  $\mu_\tau$  and  $\sigma_\tau$  the average and the standard deviation - both varying with  $\tau$  within the period  $T$ , having  $\omega$  steps  $\Delta t$ , and  $\varepsilon_{p,\tau}$  is a dependent stochastic component, which may be stationary or not.

Periodic terms  $\mu_\tau$  and  $\sigma_\tau$  are admitted to the form of Fourier series:

$$\begin{aligned} \mu_\tau &= m_Z + \sum_{j=1}^n \left[ A_j \cos\left(2\pi \frac{j}{\omega} \tau\right) + B_j \sin\left(2\pi \frac{j}{\omega} \tau\right) \right], \\ \sigma_\tau &= s_Z + \sum_{j=1}^n \left[ A'_j \cos\left(2\pi \frac{j}{\omega} \tau\right) + B'_j \sin\left(2\pi \frac{j}{\omega} \tau\right) \right], \end{aligned} \quad (10)$$

for  $\tau = 1, 2, \dots, \omega$ ;  $m_Z$  and  $s_Z$  are the average and the standard deviation of  $Z_{p,\tau}$  values ( $p = 1, 2, \dots, N$ ,  $\tau = 1, 2, \dots, \omega$ );  $n$  is the number of significant harmonics taken into account, and  $A_j, B_j, A'_j, B'_j$  are the Fourier coefficients of the harmonics  $j$ , obtained with relations:

$$\begin{aligned} A_j &= \frac{2}{\omega} \sum_{\tau=1}^{\omega} (m_\tau - m_Z) \cos\left(2\pi \frac{j}{\omega} \tau\right); \\ B_j &= \frac{2}{\omega} \sum_{\tau=1}^{\omega} (m_\tau - m_Z) \sin\left(2\pi \frac{j}{\omega} \tau\right); \\ A'_j &= \frac{2}{\omega} \sum_{\tau=1}^{\omega} (s_\tau - s_Z) \cos\left(2\pi \frac{j}{\omega} \tau\right); \\ B'_j &= \frac{2}{\omega} \sum_{\tau=1}^{\omega} (s_\tau - s_Z) \sin\left(2\pi \frac{j}{\omega} \tau\right). \end{aligned} \quad (11)$$

The averages and standard deviations in these relationships will be given by:

$$\begin{aligned} m_\tau &= \frac{1}{N} \sum_{p=1}^N Z_{p,\tau} \\ s_\tau &= \sqrt{\frac{1}{N} \sum_{p=1}^N (Z_{p,\tau} - m_\tau)^2} \\ m_Z &= \frac{1}{\omega} \sum_{\tau=1}^{\omega} m_\tau \\ s_Z &= \frac{1}{\omega} \sum_{\tau=1}^{\omega} s_\tau \end{aligned} \quad (12)$$

According to the mean  $\mu_\tau$  and the standard deviation  $\sigma_\tau$ , variable  $Z$  can be written:

$$Z_{p,\tau} = \mu_\tau + \sigma_\tau \varepsilon_{p,\tau}, \quad (13)$$

where  $\varepsilon_{p,\tau}$  is a time-dependent stochastic component, the current value being dependent on previous values.

This component is simulated by a second-order autoregressive model:

$$\varepsilon_{p,\tau} = \alpha_{1,\tau-1} \varepsilon_{p,\tau-1} + \alpha_{2,\tau-2} \varepsilon_{p,\tau-2} + \sqrt{1 - (\alpha_{1,\tau-1}^2 + \alpha_{2,\tau-2}^2 + 2\alpha_{1,\tau-1} \alpha_{2,\tau-2} \rho_{1,\tau-1}^2)} \cdot \xi_{p,\tau} \quad (14)$$

$$\begin{aligned} \alpha_{1,\tau-1} &= \frac{\rho_{1,\tau-1} - \rho_{1,\tau-2} \rho_{2,\tau-2}}{1 - \rho_{1,\tau-2}^2} \\ \alpha_{2,\tau-2} &= \frac{\rho_{2,\tau-2} - \rho_{1,\tau-1} \rho_{1,\tau-2}}{1 - \rho_{1,\tau-2}^2} \end{aligned} \quad (15)$$

where:

$\rho_{1,\tau-1}$  is the first order autocorrelation coefficient between the values  $\varepsilon_{p,\tau}$  and  $\varepsilon_{p,\tau-1}$ ;

$\rho_{1,\tau-2}$  is the first order autocorrelation coefficient between the values  $\varepsilon_{p,\tau-1}$  and  $\varepsilon_{p,\tau-2}$ , and

$\rho_{2,\tau-2}$  is the second order autocorrelation coefficient between the values  $\varepsilon_{p,\tau}$  and  $\varepsilon_{p,\tau-2}$ .

### 3.4 Independent and stationary stochastic component

By relation (14), the independent and stationary stochastic component is determined as:

$$\xi_{p,\tau} = \frac{\varepsilon_{p,\tau} - \alpha_{1,\tau-1} \varepsilon_{p,\tau-1} - \alpha_{2,\tau-2} \varepsilon_{p,\tau-2}}{\sqrt{1 - (\alpha_{1,\tau-1}^2 + \alpha_{2,\tau-2}^2 + 2\alpha_{1,\tau-1} \alpha_{2,\tau-2} \rho_{1,\tau-1}^2)}}. \quad (16)$$

For the variable  $\xi$  one can choose a normal probability density function of the form:

$$f(\xi) = \frac{1}{\sqrt{2\pi}\sigma_\xi} e^{-\frac{1}{2} \left(\frac{\xi - \mu_\xi}{\sigma_\xi}\right)^2}. \quad (17)$$

Parameters  $\mu_\xi$  and  $\sigma_\xi$  of the proposed function can be determined by the maximum likelihood methods or the least squares method. Its adjustment to the values of the data set obtained by relation (16) is tested with the chi-square test. The statistic can be computed with the relation:

$$\chi^2 = \sum_{j=1}^K \frac{(n_j - n_j^*)^2}{n_j^*}, \quad (18)$$

where  $n_j$  is the number of observations in class  $j$  and  $n_j^*$  is the theoretical number of observations in class  $j$  corresponding to the function being tested. If the value of this estimator is lower than the standard normal distribution quantile for which the distribution function is  $(1-\alpha)$ , the normal function is accepted as adequate for the data sample being analyzed.

### 4 Results

The array of 132 analyzed values has the average 27.11% and the standard deviation 3.29%.

The coefficients of the regression equation, specified by a first order polynomial, were obtained by the least squares method and have the form:

$$T(t) = 27.787 - 0.1126 \cdot t, \tag{19}$$

where  $t$  specified the year.

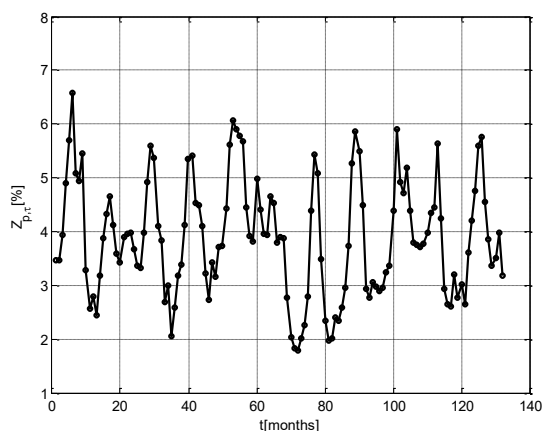
Transforming time from years into months and eliminating this trend from the original variable, the variable becomes  $Y$ , which has the same order of magnitude and the rate of the curve does not differ from that of the original variables (figure 1), because the average trend is insignificant.

The regression equation for the trend in the variance obtained from the values of the standard deviation for each year has the form:

$$T_s(t) = 6.737 - 0.041 \cdot t, \tag{20}$$

in which  $t$  is the year specification which also had to be converted into months.

By removing this trend from intermediate variable  $Y$  a new statistical process  $Z$  is obtained, which contains only components of periodicity and stochasticity. This variable is graphically represented in figure 2.



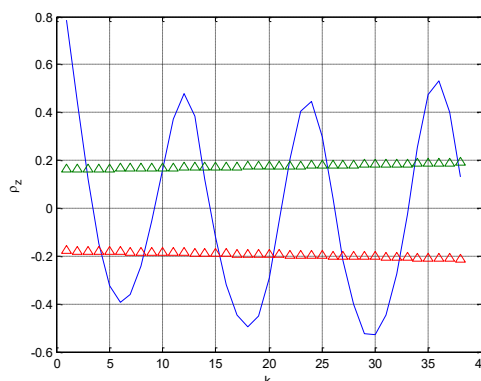
**Fig. 2.** Variable  $Z$ , obtained by removing the trend on average and in the variance from the original series.

While the original variable varies from 12 to 45%, figure 2 illustrates that the variable obtained after the elimination of both tendencies, on average and in the variant, is situated between 2 and 6.5%. The average is 3.88% and the standard deviation 1.07%.

The correlogram for variable  $Z$  is shown in figure 3. It is clear that the limits defined by the Anderson test are exceeded, which shows that variable  $Z$  has periodicity. This was to be expected, given the dependence between the hydro energy produced and the variation of the rivers flows which also show periodicity within the period  $T = 1$  year.

For periodicity modeling, the multi-year values of each month for average and variance are used for

calculating coefficients  $A$ ,  $B$ ,  $A'$  and  $B'$  of the 6 considered harmonics.



**Fig. 3.** Autocorrelation plot for variable  $Z$ .

**Table 1.** The coefficients of the 6 harmonics considered in the Fourier series to model the periodicity.

	$A$	$B$	$A'$	$B'$
1	-1.030	0.354	-0.002	-0.244
2	0.208	-0.261	0.041	0.093
3	0.040	0.140	0.053	0.027
4	-0.02	-0.062	0.078	-0.031
5	0.004	0.017	-0.058	-0.011
6	0.021	0	0.014	0

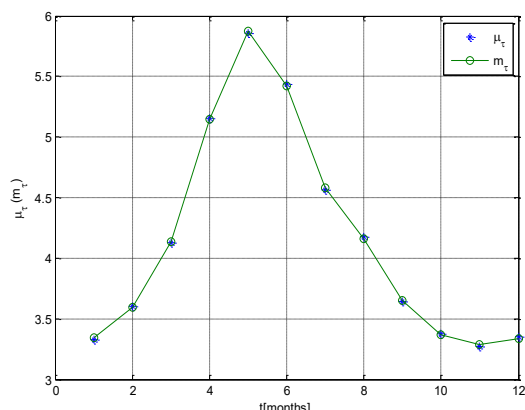
Periodic averages and standard deviations, calculated with Fourier series (6 harmonics) and for  $Z$  array respectively for every month, are very close, as can be seen in Table 2.

**Table 2.** Averages and standard deviations of  $Z$  values, respectively Fourier based calculated values.

	$\mu_{\tau}$ (Fourier)	$m_{\tau}$	$\sigma_{\tau}$ (Fourier)	$s_{\tau}$
1	3.130	3.140	0.685	0.692
2	3.380	3.370	0.493	0.486
3	3.870	3.880	0.458	0.465
4	4.832	4.821	0.444	0.437
5	5.480	5.491	0.482	0.489
6	5.083	5.072	0.857	0.850
7	4.269	4.280	0.790	0.797
8	3.903	3.893	1.062	1.056
9	3.407	3.418	1.021	1.027
10	3.142	3.132	0.682	0.675
11	3.039	3.050	0.780	0.787
12	3.110	3.100	0.842	0.835

The averages obtained by approximation with the Fourier series show a maximum deviation of 0.0106 as compared to the annual average of the sample and for the standard deviation the maximum difference is 0.0068.

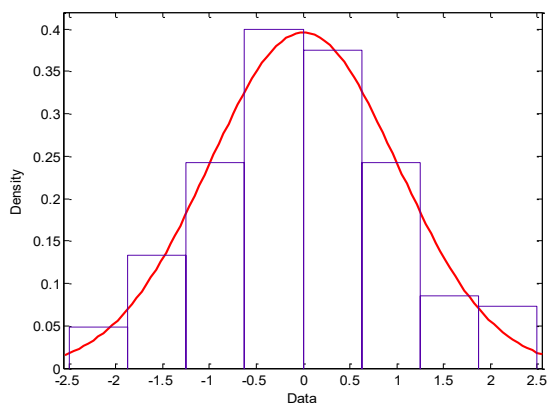
Figure 4 shows that the values calculated with the Fourier series are in good concordance with the sample values.



**Fig. 4.** Periodic average (continuous  $\mu_\tau$  and discrete  $m_\tau$ ).

The independent stochastic component is obtained by relation (16). For this variable, a classic statistical analysis can be performed [17].

The frequency histogram for classification in  $k = 8$  classes is shown in figure 5. One can see symmetry in the graph which suggests choosing a normal density function.



**Fig. 5.** Adjusting the normal probability density function to the values of the independent stochastic variable.

Note that the function of normal probability density, with parameters obtained for average  $\mu_\xi = 0.004$  and standard deviation  $\sigma_\xi = 1.0068$ , very well adjusts the set of data analyzed.

After the chi-square test, the estimator calculated by relation (18) was obtained as  $\chi^2 = 8.37$ , less than the standard normal distribution quantum for which the distribution function is  $(1-\alpha)$ . For a threshold of significance,  $\alpha = 0.05$ , the value of this critical estimator is  $\chi^2_{1-\alpha} = 11.05$ .

The normal function can be accepted as being appropriate for the variable representing the independent stochastic component.

## 5 Conclusions

The variation coefficient of the hydro energy percentage and the total energy produced in Romania over a period of 11 years shows an insignificant tendency of variation on average. This one faithfully follows the seasonal variations of river flows, the inflows in the hydropower developments reservoirs, respectively the turbine flow in hydropower plants. The period of the analyzed temporal series is of 1 year.

The random variable  $Z$  that was obtained after elimination of trend components can be used in future in a forecasting program by generating values for the stochastic variable according to a normal probability law that has been found to fit very well with the considered variable.

The authors gratefully acknowledge Professor Radu Popa, from the Department of Hydraulics, Hydraulic Machinery and Environmental Engineering, University POLITEHNICA of Bucharest, for valuable support and discussions.

This work has been funded by University Politehnica of Bucharest, through the "Excellence Research Grants" Program, Ctr. No. 8/25.09.2017 (SI - Acvi), UPB-GEX 2017.

## References

1. C. Tepedino, C. Guarnaccia, S. Iliev, S. Popova, J. Quartieri, *Recent Advances in Energy, Environment and Financial Planning*, 13-22 (2014)  
<http://www.wseas.us/e-library/conferences/2014/Florence/DEEE/DEEE-01.pdf>
2. M. B. Garcia-Morales, L. Dubus, *INT J CLIMATOL*, **27**, 1691-1705 (2007)  
<https://doi.org/10.1002/joc.1608>
3. R. M. Roken, M. A. Badri, *CURA The First Annual Student Research Symposium* (2007)  
<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.136.9509>
4. E. Almshaiei, H. Soltan, *Alexandria Engineering Journal*, **50**, 137-144 (2011)  
<https://doi.org/10.1016/j.aej.2011.01.015>
5. H.A. Amarawickrama, L. C. Hunt, *ENERGY*, **33** (5), 724-739 (2007)  
<https://doi.org/10.1016/j.energy.2007.12.008>
6. A. V. Dias, *Forecasting hourly prices in the Portuguese power market with ARIMA model* (thesis), ISCTE, Lisboa (2009)  
<https://repositorio.iscte-iul.pt/bitstream/10071/2040/1/Tese%20Final.pdf>
7. L. Tang, C. Wang, S. Wang, *Energy Time Series Data Analysis based on a Novel Integrated Data Characteristic Testing Approach*, *PROCEDIA COMPUT SCI*, **17**, 759-769 (2013)  
<https://doi.org/10.1016/j.procs.2013.05.098>

8. K. Dmitrieva, *Forecasting of a hydropower plant energy production*, Master's Thesis in Computer Science, HIOF (2015)  
[https://brage.bibsys.no/xmlui/bitstream/handle/11250/293868/15-00486-7%20Masteravhandling%20235126\\_1\\_1.pdf?sequence=1](https://brage.bibsys.no/xmlui/bitstream/handle/11250/293868/15-00486-7%20Masteravhandling%20235126_1_1.pdf?sequence=1)
9. A. Tealeb, H. Hefny, A. Badr, FCIJ, **8** (1), 39-47 (2017) <https://doi.org/10.1016/j.fcij.2017.05.001>
10. L. Tang, L. Yu, S. Wang, J. Li, S. Wang, APPL ENERG, **93**, 432-443 (2012)  
<https://doi.org/10.1016/j.apenergy.2011.12.030>
11. J. Wang, D. Chi, J. Wu, H-y. Lu, EXPERT SYST APPL, **38** (7), 8419-8429 (2011)  
<https://doi.org/10.1016/j.eswa.2011.01.037>
12. G. Năstase, A. Șerban, A. F. Năstase, G. Dragomir, A. I. Brezeanu, N. F. Iordan, RENEW SUST ENERG REV, **80**, 297-312 (2017)  
<https://doi.org/10.1016/j.rser.2017.05.209>
13. E. Gradinaru, L. Bucs, G. Brătucu, Bulletin of the Transilvania University of Brașov, **9** (58), 1, 335-346 (2016)  
[http://webbut.unitbv.ro/BU2015/Series%20V/2016/BULETIN%20I%20PDF/38\\_Gradinaru.pdf](http://webbut.unitbv.ro/BU2015/Series%20V/2016/BULETIN%20I%20PDF/38_Gradinaru.pdf)
14. Transelectrica, *Annual Report* (2017)
15. C.C. Lee, C.P. Chang, ENERG ECON, **27** (6), 857-872 (2005)  
<https://doi.org/10.1016/j.eneco.2005.08.003>
16. Hidroelectrica website, [www.hidroelectrica.ro](http://www.hidroelectrica.ro)
17. R. Popa, A. Neagoe, *Statistics and data processing* (in Romanian, Printech, Bucharest, 2008)