Change point detection in recent hydropower generation in Romania

A Neagoe\textsuperscript{1}, E-I Tică\textsuperscript{1}, F Popa\textsuperscript{2}, B Popa\textsuperscript{1}
\textsuperscript{1} Department of Hydraulic, Hydraulic Machineries and Environmental Engineering, University POLITEHNICA of Bucharest, Romania
\textsuperscript{2} Department of Electronics, Telecommunications and Energy Engineering, University VALAHIA of Targoviste, Romania

E-mail: bogdan.popa@upb.ro

Abstract. In Romania, about thirty percent of the installed capacity in National Power Grid and a third of the electricity generation is in hydropower plants. Thus, any aspect related to this form of renewable energy source is very valuable. Chronological records of yearly hydropower generation in Romania represent a time series that can be statistically analysed. The aim of this work is to determine the trend of this time series and, also, to locate the change points, where the time series has significant change. The period analysed is from 1980 to the present. To determine the change point of this time series, the Pettitt method was applied.

1. Introduction
Often, in the field of meteorology and hydrology the issue arises of identifying points where significant changes occurs in series of data over long periods for which their justification is sought. A major change in the natural regime of flows occurs when a dam is built on a river and, also, climate change is often the reason for changes in the series of meteorological parameters such as rainfall [1], [2], humidity [3], air temperature [4], [5], wind speed [6], etc. To determine the moments in which this change occur the statistical analysis of the recorded data set is used.

For each country, the trend for generation and consumption of electricity is an interesting subject. In Romania, about a third of the total energy generation is from hydropower plants [7], and therefore it is important to follow the evolution over time and trend of this variable.

In [8] was presented a review of evolution in time – between 1884 and 2015 – of hydropower plants. The aim of this paper is to determine the evolution of hydropower generation as time series from a statistical approach and to determine the change point of this time series.

2. Time series data
The analysed time series are composed by annual values of recorded data of energy generation in hydropower plants in Romania between 1980 and 2019, data published in annual report of Hidroelectrica [9] and Transelectrica website [10] regarding the National Power Grid (NPG).

In Romania, during 1950-1995, most of the important hydropower plants were built, which contributed to a major increase of energy production. For example, between 1980 and 1990, 47 new hydropower plants were put into operation, including Iron Gates II, Retezat, Şugag, Petreşti, Tismana, Brădişor, etc. This evolution easily indicates a continuous increase for hydropower production over this period.
The development over time of annual hydropower generation for the period 1990 – 2019 decrease, when the development of new hydropower plants was slowed down. The recorded data used for the hydropower generation, \( G \), for the analysis period are shown in figure 1.

![Figure 1. Time evolution of the annual hydropower generation in Romanian.](image)

From figure 1 it can be noticed that for 2 years, 2005 and 2010, the hydropower generation is very high. It is noted that the extreme values observed on the graph are not change points for the time series and the average of all 40 values is 14,804 GWh.

3. Pettitt’s method
The Pettitt method is commonly used to detect change point in the mean value of observed series [9].

If the series \( x_1, x_2, \ldots, x_N \) has a change at time \( t \), then \( x_1, x_2, \ldots, x_t \) and \( x_{t+1}, x_{t+2}, \ldots, x_N \) are two homogeneous subsets with different distribution function. The \( U_t \) statistics is:

\[
U_t = \sum_{i=1}^{t} \sum_{j=t+1}^{N} \text{sign}(x_i - x_j),
\]

\[
\text{sign}(x_i - x_j) = \begin{cases} 
1, & \text{if } (x_i - x_j) > 0 \\
0, & \text{if } (x_i - x_j) = 0 \\
-1, & \text{if } (x_i - x_j) < 0 
\end{cases}
\]

Its statistic \( K_t \) and the associated probabilities used in the significance testing are:

\[
K_t = \max \{|U_t|\},
\]

\[
p = \exp\left\{ -K_t / \left( N^2 + N^2 \right) \right\}.
\]

It can be considered that where the function \( U_t \) has a peak of, there is the moment when in the time series occurs a change.

The Pettitt method can be used if one change point is present, i.e., one abrupt change of the mean. To find if another change point exists, the original time series can be split into two sub-series and repeat the analysis for each sub-series [10].

4. Results
After applying Pettitt method for series between 1980-2019, it can be observed that the maximum of function \( U_t \) is obtained for the year 1994. The change in trend is obvious at that time and is since new
Hydropower plants were commissioned in Romania between 1980 and 1994 and it was slowed down afterwards.

For recorded values between 1994 and 2019 the Pettitt method was applied again and the function associated with this method is shown in figure 2.b. In this situation, because the length of the string is shorter, the main maximum is not so obvious that in figure 2.a, however, from this period the probability that one year would represent a change point is 2004.

![Figure 2. Temporal distributions of the test statistic $U_t$: a) 1980 – 2019; b) 1995 – 2019.](image)

In table 1 are presented two values of $K_t$ for first and second change points.

| Year          | $K_t$ |
|---------------|-------|
| First change point | 353   |
| Second change point | 68    |

For analysis, the time series will be split into three subsets: $S_1$ contains $N_1$ values for the period 1980-1994, $S_2$ contains $N_2$ values for the period 1995-2004 and $S_3$ contains $N_3$ values recorded in the period 2005-2019. The main statistical characteristics of these three subsets are given in table 2, with the specification of the year in which the minimum and maximum values of each period were recorded.

| $E[GWh]$ – $S_i$ (Year) | Average | Standard deviation | Min (Year) | Max (Year) |
|--------------------------|---------|--------------------|------------|------------|
| $S_1$ (1980-1994)        | 11975   | 1128               | 9930 (1983) | 14110 (1991) |
| $S_2$ (1995-2004)        | 16103   | 1685               | 13130 (2003) | 18690 (1998) |
| $S_3$ (2005-2019)        | 16767   | 2246               | 12200 (2012) | 20470 (2010) |

The three subsets will be analysed separately because they have different statistical characteristics and averages. Also, for the period 2005-2019 it is noted that the limits within which the recorded values are situated are higher, implying a higher average standard deviation. For each of the three subsets, a linear regression is found which best approximates the discrete values recorded.
For the period 1980-1994 shown in figure 3, the recorded data are well approximated by a regression equation, in the form of:

\[ y_1 = 11340 + 79 \cdot t, \quad t = 1, 2, \ldots, N_1. \]  

(5)

Figure 3. The linear regression of hydropower generation in 1980-1994.

The average trend for this period is upward.

For the period shown in figure 4, 1995-2004, a decrease in the energy produced in hydropower plants can be observed.

The slope of the regression line is strongly downward as shown by the regression equation found for the variation of hydropower generation over the period 1995-2004:

\[ y_2 = 17481 - 251 \cdot t, \quad t = 1, 2, \ldots, N_2. \]  

(6)
From figure 5, it can be observed that the trend of hydropower generation in the time period 2004-2019 is decreasing, but the slope of the regression equation is low, i.e. the downward trend is smaller than for the period 1994-2004:

\[ y_3 = 17734 - 121 \cdot t, \quad t = 1, 2, ..., N_3. \] (7)

The average values of the last analysed period are also the highest, around 16.7 TWh annually. The increase is most likely due to the upgrading of hydropower plants during this period.

The Pettitt method finds the moment where the statistical characteristics before and after this differ significantly. It can be observed that for the 3 subsets in which the original series was divided, both the averages and the variants are very different. If it relates to 1980, the time series can be considered to have an upward trend and an average of 14804 GWh. This value is well above the average of the last analysis period 2005 to 2019, which is 16767 GWh. When considering the series from 2004 onwards it can be concluded that although there are annual values with high standard deviation, they fluctuate around the average, with a slightly decreasing trend.

![Figure 5. The linear regression of hydropower generation for the period 1995-2004.](image)

5. Conclusions
After applying Pettitt method for time series represented by hydropower generation recorded between 1980-2019, it can be observed that the maximum of the function \( U_t \) is obtained for the year 1994. The change in trend at that time is justified since new hydropower plants were commissioned in Romania between 1980 and 1994 and the development of the sector was almost stopped afterwards.

Using the Pettitt method, the time series was fragmented, so the years of changes in the time series of the annual energy generation in hydropower plants in Romania over the last 40 years were determined. Using all the values from 1980 to 2019 it was obtained that 1994 represents the year when the change point is abrupt. Instead, if only the range of values from 1995 to 2019 are analysed, in 2004 the change was not so radical, however, the analysis can conclude that for the period 2004-2019 it was recorded in average an increase in hydropower generation compared to previous years.

References
[1] Liu Q, Yang Z F and Cui B S 2008 Spatial and temporal variability of annual precipitation during 1961-2006 in Yellow River Basin, China, *J. of Hydrology* 361 (3-4) pp 330–338
[2] Swanson K L and Tsonis A A 2009 Has the climate recently shifted? *Geophysical Research Letters* 36 (6) 06711
[3] Bryson C B Richard E C and Adrian W B 2012 Trend estimation and change point detection in individual climatic series using flexible regression methods, \textit{J of Geophysical research} \textbf{117} D16106

[4] Zhang P, Ren G, Xu Y, Wang X L, Qin Y, Sun X and Ren Y 2019 Observed changes in extreme temperature over the global land based on a newly developed station daily dataset, \textit{J of Climate} \textbf{32} (24) pp 8489-8509

[5] Jaiswal R K, Lohani A K and Tiwari H L 2015 Statistical analysis for change detection and trend assessment in climatological parameters, \textit{Environ Process} \textbf{2}, pp 729–749

[6] Birsan M, Marin L and Dumitrescu A 2013 Seasonal changes in wind speed in Romania, \textit{Romanian Reports in Physics} \textbf{65} (4) pp 1479–1484

[7] Neagoe A, Tică E-I, Diminescu M A, Popa B and Cismaş S C 2019 The analysis of hydropower ratio from total energy production in Romania \textit{E3S Web Conf.}, \textbf{85} 06009

[8] Năstase G, Șerban A, Năstase A F, Dragomir G, Brezeanu A I and Iordan N F 2017 Hydropower development in Romania. A review from its beginnings to the present, \textit{Renewable and Sustainable Energy Reviews}, \textbf{80} pp 297-312

[9] Pettitt A N 1979 A non-parametric approach to the change-point problem, \textit{Journal of the Royal Statistical Society. Series C (Applied Statistics)} \textbf{28} (2) pp 126–135

[10] Busuioc A and Von Storch H 1996 Changes in the winter precipitation in Romania and its relation to the large-scale circulation \textit{Tellus A} \textbf{48}(4) pp 538–552